N. L. Sarda · P. S. Acharya · S. Sen *Editors*

Geospatial Infrastructure, Applications and Technologies: India Case Studies

Springer

Geospatial Infrastructure, Applications and Technologies: India Case Studies

N. L. Sarda · P. S. Acharya
S. Sen
Editors

Geospatial Infrastructure, Applications and Technologies: India Case Studies

Springer

Editors
N. L. Sarda
Indian Institute of Technology Bombay
Mumbai, Maharashtra, India

P. S. Acharya
Department of Science and Technology
Government of India
New Delhi, India

S. Sen
Department of Computer Science and Engineering, Advanced Lab on GISE
Indian Institute of Technology Bombay
Mumbai, Maharashtra, India

ISBN 978-981-13-2329-4 ISBN 978-981-13-2330-0 (eBook)
https://doi.org/10.1007/978-981-13-2330-0

Library of Congress Control Number: 2018952617

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

I am happy that Natural Resources Data Management System (NRDMS) and National Spatial Data Infrastructure (NSDI) of the Department of Science and Technology are promoting geospatial science and applications by providing support to the academic institutions, national laboratories, and various other stakeholders to undertake research and develop various applications. In the process, efforts have been made to carry out R&D and innovation ensuring better health and nutrition, imparting education and training of skill development, setting up of infrastructure and connectivity, and promoting e-governance. The outputs of research projects supported by NRDMS-NSDI over the last 5–7 years have been compiled and brought out as a special volume on *Geospatial Infrastructure, Applications and Technologies: India Case Studies*.

It has been noted that acquisition of geospatial data through various techniques like LiDAR, drone, and satellite images has made tremendous progress, particularly in data resolution and coverage. With this, it has become easier to analyze such datasets and develop various applications to facilitate implementation of developmental schemes at the village, block, district and state levels. Under NRDMS program, which is focusing on promoting geospatial science and applications, specific R&D projects are being supported to provide a required push to help implementation of various flagship programs of the government. The research outputs of the R&D projects of NRDMS and NSDI are helping state governments to effectively implement the developmental schemes ensuring spatial planning.

I am very happy that the present volume contains a glimpse of various R&D studies supported by DST. This is expected to serve as a reference material to the researchers and the planners who are working in the areas of geospatial science, applications, and solutions. Finally, it would also help in achieving the "India-2022" vision of the Government by providing relevant geospatial information and solutions for the benefit of the stakeholders.

New Delhi, India

Prof. Ashutosh Sharma
Secretary, Ministry of Science and Technology
Government of India

Preface

As geospatial technologies continue to converge with mainstream IT, research efforts in the domain have gained significant overlap with those in computational sciences. Newer spatial data acquisition techniques such as LiDAR, large-scale airborne imagery, geosensors, and automated vehicle tracking have motivated researchers to explore newer geospatial applications. Similarly, spatial data infrastructures are being increasingly used as open data portals, and setting up of spatial data infrastructures at different levels of governance has prompted new policy initiatives. All these initiatives are creating a continuous cycle of innovations in technologies and applications developments.

It is necessary to chronicle the efforts in geospatial technology, especially through multiple efforts by researchers in India. We invited contributions from leading researchers from India in the form of extended abstracts on different topics that they have contributed in recent years. We requested full-paper versions of the contributions that represented a wide spectrum of the efforts covering policies, standards, infrastructure developments and deployments, cutting-edge applications, and new technology developments. The full papers received were peer-reviewed and accepted for this book. These papers are original works and provide an overview of the studies from Indian researchers. Many of the papers resulted from research funding from the Department of Science and Technology, Government of India, through its Natural Resources Data Management System (NRDMS) program. This book represents an effort to provide a reference for researchers, practitioners, and users of geospatial technologies, particularly in India, to understand the policy initiatives, innovative techniques, and applications of this domain.

We thank all the authors for their enthusiastic efforts in completing the book. Many of the authors have also helped in reviewing other chapters of the book. We thank the Department of Science and Technology, IIT Bombay, and our colleagues for their support and help. Lastly, but very importantly, we thank our publishers for their help in making the book presentable.

Mumbai, India N. L. Sarda
New Delhi, India P. S. Acharya
Mumbai, India S. Sen

Contents

Editors and Contributors

About the Editors

N. L. Sarda is Professor Emeritus in the Department of Computer Science and Engineering. He has chaired the NSDI 2013 meeting and OGC TC meetings at IIT Bombay. He serves as Chair of NSDI's Steering Committee on State Spatial Data Infrastructures (SDIs) and its Applications. His research interests include database modelling, spatiotemporal, geospatial databases, and software engineering.

P. S. Acharya is CEO of the National Spatial Data Infrastructure of India and Senior Scientist in the Department of Science and Technology, Government of India. He has been associated with the NRDMS programme in the department and has contributed to policy and infrastructure development for the geospatial community in India, including the National Data Registry.

S. Sen is Sr Research Scientist, GISE Lab, Department of Computer Science and Engineering, IIT Bombay. His areas of interest include large spatial databases, road network data, semantic interoperability, geospatial visualization, and analytics.

Contributors

H. Allirani is Junior Research Fellow, Transportation Engineering Lab, Indian Institute of Science, Bangalore. She received her master's degree with a Specialization in transportation engineering and management from National Institute of Technology, Tiruchirappalli. Her work specifically focuses on the description and analysis of adaptation strategies with respect to the transportation sector for Bangalore Metropolitan Region.

R. Ansari is Assistant Professor at K. J. Somaiah College of Engineering, Mumbai. He received his Ph.D. from IIT Bombay and specializes in multi-resolution signal and image analysis, digital communication systems, etc.

N. Arora was Research Assistant in the GISE Lab in the Department of Computer Science and Engineering, IIT Bombay. She worked on ontology-based query processing.

P. V. Arun is pursuing his Ph.D. at the Centre of Studies in Resources Engineering, IIT Bombay. His research areas include deep learning, sub-pixel classification, high-performance computing, and super-resolution techniques.

V. Balasubramanian is Associate Professor in the Department of Computer Science and Engineering, Amrita University. She has worked in large-scale spatio temporal data management, information systems, and cyber-physical systems. She has previously contributed to the ITR Rescue Project at California Institute for Telecommunications and Information Technology. She worked on scalable multi-geography path planning algorithm in the context of designing integrated disaster response simulation test beds.

U. Bellur is Professor in the Department of Computer Science and Engineering, IIT Bombay, Mumbai. He has previously worked in the industry where he helped establish distributed object standards such as CORBA with OMG and J2EE with JCP. His research interests include spatial data processing, virtualization, and cloud computing—containers in virtualized cloud data centers, handling boot storms in data centers, energy-aware provisioning, VM migration, cloud simulation, and SDNs and NFV in cloud data centers.

M. Bhattacharya is SERB-National Post-doctoral Fellow at ICAR-Central Inland Fisheries Institute, Kolkata. His work focuses on computational biology platforms.

A. Bhushan received her Ph.D. from IIT Guwahati and is Member of the GISE Lab in the Department of Computer Science and Engineering, IIT Bombay. Her research interests include spatiotemporal data mining, evacuation planning, graph algorithms for massive datasets.

K. M. Buddhiraju is Professor at the Centre of Studies in Resources Engineering, IIT Bombay. His areas of interest include machine learning, remotely sensed image processing and analysis, and multimedia educational tool development.

B. Chakrabarti is Senior Scientist at CESCRA, ICAR-Indian Agricultural Research Institute. Her focus areas include Climate4Change (impact and adaption), carbon sequestration and simulation modelling.

D. Chakraborty is National Fellow at ICAR-Indian Agricultural Research Institute. He specializes in agricultural physics and has research interests in soil physics, soil water management, remote sensing, and GIS.

S. Chaudhary is Professor and Associate Dean at the School of Engineering and Applied Sciences, Ahmedabad University. He was previously Professor and Dean (Academic) at DA-IICT, Gandhinagar. His research areas are distributed computing, cloud computing, data analytics, and ICT applications in agriculture and rural development.

B. K. Das is Director of ICAR-Central Inland Fisheries Institute, Kolkata. His work has focused on fish health management, fish genetics, and microbial molecular analysis. His current interests cover aquaculture resource mapping, management, and spatial modelling.

R. Dhobal is Director General of Uttarakhand State Council for Science and Technology, and is Principal Investigator for Uttarakhand State SDI.

A. Dasgupta is Research Scholar in Indian Institute of Technology Kharagpur, India, where he received his MS degree. His research interests include volunteered geographic information (VGI) for disaster management, location-based services (LBSs) in mobile devices, spatial Web services orchestration, and spatial data mining.

A. Dashora is Assistant Professor at Indian Institute of Technology Guwahati. During his Ph.D., he worked in the area of airborne data acquisition and developed a flight planning system for airborne LiDAR. His research interests are mathematical modelling of historical images like CORONA, synthetic simulation for geoinformatics, sensor calibration, and applications of geoinformatics for solving social issues.

S. S. Dawn is Assistant Professor in the Department of Chemical Engineering, Sathyabama University, Chennai, and also Head of the Centre for Waste Management, Chennai. Her primary research interests include environmental impact assessment and mass transfer.

P. Dev is Research Scholar at Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University. An avid biker, her interests lie in diverse aspects of geospatial technology, especially big data.

P. Dixit is Senior Research Fellow of the Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute. Her areas of expertise include geoinformatics and its applications in agricultural sciences.

D. Dutta is Adviser and Scientist of the NRDMS Division in the Department of Science and Technology, Government of India. His areas of expertise and research interest are in the application of geoinformatics for natural resources management, watershed management technologies, geoinformation management, and capacity building for local-level planning and Interational S&T Cooperation in the context of sustainable development goals.

S. K. Ghosh is Professor in the Department of Computer Science and Engineering, IIT Kharagpur. Before joining IIT Kharagpur, he worked for the Indian Space Research Organisation. His research interests include spatial data science, spatial Web services, and cloud computing.

S. Ghosh is Associate Professor in the Department of Civil Engineering, MGM's Jawaharlal Nehru Engineering College. He specializes in the visualization of LiDAR data and pattern discovery from geospatial datasets. He is currently involved in the study of smart cities and the application of geospatial technologies to such ideas.

R. Gupta is Professor of civil engineering at Visvesvaraya National Institute of Technology, Nagpur. With over 20 years of academic and research experience, he has made significant contributions to water supply and wastewater collection systems, environmental system modelling, and optimization.

R. Gupta is Senior Professor in the Department of Civil Engineering, BITS Pilani. He has over 30 years of academic experience, and his research interests include construction planning, rainwater harvesting, GIS remote sensing, and soft computing.

R. C. Harit is Senior Technical Officer at CESCRA, ICAR-Indian Agricultural Research Institute. He has worked extensively on yield response of field crops.

H. Hemanth Kumar heads the Karnataka Natural Resources Data Management Systems (NRDMS). His areas of interest include natural resource management, non-organic solid wastes, Web portals and use of geoinformatics for decision making.

N. Jain is Principal Scientist at CESCRA, ICAR-Indian Agricultural Research Institute. She has made a significant contribution in the quantification of greenhouse gas (GHG) emission from different field crops, vegetables and different management practices (nutrient, water, NI, LCC) and development of inventories of methane, nitrous oxide and other air pollutants emission from crop residues burning in the agriculture sector. Her research interests include natural resource management, greenhouse gas emissions, and climate change adaptation.

R. S. Jatav is Senior Research Fellow at CESCRA, ICAR-Indian Agricultural Research Institute. He has previously worked in World Bank-funded projects on climate change adaptation. Agronomy and climate change adaptation are his key focus areas.

N. V. Joshi is Associate Professor at the Centre of Ecological Sciences and Member of the Energy and Wetland Research Group at IISc Bengaluru. His topics of interest include mathematical modelling, ecology, and evolutionary biology.

A. Kar is a Ph.D. candidate in the Department of Zoology, Vidyasagar University, West Bengal, and Research Fellow on a DST-sponsored project. His present research focuses on diversity, conservation, and management aspects of fishery resources.

R. Katiyar is Senior Research Fellow at CESCRA, ICAR-Indian Agricultural Research Institute. His areas of expertise include pollution monitoring and impacts of pollutants.

P. Khatavkar is Chief Operating Officer and Adjunct Faculty at the Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University. An

Erasmus Mundus scholar, an entrepreneur by profession and a wildlifer at heart, he has worked extensively on wildlife issues.

S. Kumar is Professor at the Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University, and Guest Professor at the University of Cologne, Germany, since 2013. A keen practitioner of interdisciplinary research, she has multifaceted research interests in the fields of landscape ecology, biodiversity conservation, geospatial technologies, geospatial big data, and education for sustainable development.

V. Kumar is a Ph.D. student at School of Engineering and Applied Sciences, Ahmedabad University, Ahmedabad, India. He is working on a DST-sponsored project as a JRF, and his areas of research interest in natural language processing (NLP), machine learning (ML), mathematical modelling, and data interoperability.

D. Lingote is Scientist at the EDM (IT) Division of the CSIR-National Environmental Engineering Research Institute, Nagpur.

B. Lohani is Professor at Indian Institute of Technology Kanpur. His area of interest is generating and processing high-resolution remote sensing data and their applications. He, in particular, focuses on terrestrial, mobile, and airborne LiDAR technologies and their application including simulating these data.

U. Mina is Associate Professor in the School of Environmental Sciences, JNU. Her areas of specialization include agroecology, air pollution, and environmental impact assessment. She has previously worked at ICAR-Indian Agricultural Research Institute as Senior Scientist.

J. Mohanty is pursuing his M.Tech. at the Centre of Studies in Resources Engineering, IIT Bombay. He is also Software Engineer in the Reliance Group of Companies in Mumbai. His areas of work include Web-based geospatial application development, and watermarking of vector datasets.

A. Najmi is a Ph.D. student in the Department of Computer Science and Engineering, IIT Bombay. Her interests include algorithms, optimization techniques, and machine learning.

S. Pal is an MS student in the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, India. His current research interests include machine learning, representation learning, deep neural networks, and spatial data mining.

S. Pandey is Scientist in the Department of Science and Technology, Government of India. She has been Post-doctoral Fellow at the University of Missouri Columbia, and independent Young Principal Investigator at IIT Delhi. Her areas of interests include geochemistry, sustainable development, natural resource development, and geospatial data policy.

H. Pathak is Director of ICAR-National Rice Research Institute, Cuttack, and has contributed significantly in the quantification of greenhouse gas (GHG) emission from soil, evaluation of the effects of management practices and biotic and abiotic factors on GHG emission, and development of inventories of methane and nitrous oxide emission from Indian agriculture. His main research interests include climate change, simulation modelling, and soil science.

B. C. Patra is Senior Professor of Zoology, Vidyasagar University, West Bengal. He has extensive R&D experience in fishery science and nutrition biology. His current interests include fishery data management, coastal fishery, and geoinformatics.

A. K. Prasad is Associate Professor in the Department of Applied Geology, IIT (ISM), Dhanbad. His areas of interest are digital image processing, remote sensing, and GIS with focus on multi-satellite earth observation and applications in geosciences.

J. R. Raj is Associate Professor at the School of Computing in the Department of Information Technology, Sathyabama University, Chennai. His research interests include big data Analytics, image processing, IoT, and spatial data processing.

T. V. Ramchandra is Coordinator of the Energy and Wetlands Research Group at IISc Bengaluru and Convener of Environmental Information System (ENVIS) at the Centre for Ecological Sciences at the Indian Institute of Sciences. His interests include integrated energy planning, conservation of terrestrial and aquatic ecosystems, biodiversity-ecology-hydrology linkages, and ecological modelling.

R. Rathna is an Assistant Professor in the Department of Information Technology in Sathyabama Institute of Science and Technology for the past 11 years. Her areas of interest are networking, wireless sensor networks, and Internet of things.

J. S. Rawat is Professor and Head in the Department of Geography, Kumaun University, Almora, and Director of COE NRDMS. He has a long record in teaching GIS and, more recently, spatial data infrastructure (SDI). He is Executive Committee Member of the Uttarakhand State SDI.

T. Sasiprabha is Pro-Vice-Chancellor at Sathyabama University and Faculty in the Department of Computer Science and Engineering. Her research aims to assist society through projects such as in GIS-based Web services for emergency management, detection of harmful radiation using wireless sensor networks, and LBS for everyday essentials.

S. Sahni is Senior, Applications at Oracle India Pvt. Ltd. She completed her M. Tech. in the Department of Computer Science and Engineering, IIT Bombay, where she worked on ontology-based query answering. She currently works on supply chain management.

B. Setturu is a Ph.D. student and Researcher at the Centre of Ecological Sciences, IISc Bengaluru. His interests include forest fragmentation analysis of central

Western Ghats, landscape dynamics, and remote sensing applications for environment and biodiversity.

J. Shah is Sr Research Scientist, GISE Lab, Department of Computer Science and Engineering, IIT Bombay. He was previously Professor in the Department of Mechanical Engineering, VJTI, Mumbai, and served several assignments with various departments of the Government of Maharashtra. His areas of interest include use of geoinformatics for societal change and empowerment.

A. Sharma is Scientist at the CSIR-National Environmental Engineering Research Institute, Nagpur. He is actively involved in the development of application software for environmental challenges and has a special interest in traffic noise modelling.

D. K. Sharma is Principal Scientist at CESCRA, ICAR-Indian Agricultural Research Institute. He specializes in ecosystem services and waste management.

A. Shinde is Associate Professor of Information Technology at the Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University.

P. K. Shit is Assistant Professor in the Department of Geography, Raja N L Khan Women's College, West Bengal. He has worked on applied geomorphology and environmental modelling and presently focuses on natural resource mapping and modelling.

B. Singh is Head of NRDMS in the Department of Science and Technology, Government of India. Apart from being a senior scientist in geospatial sciences, he has over 35 years of experience in geospatial policy matters. His main areas of interest include landslides, damage assessment mapping, health GIS, and village information systems.

S. Singh is Assistant Professor in the Department of Applied Geology, IIT (ISM), Dhanbad. Previously, he worked as Lecturer (1999–05) and Sr Lecturer (2006–10) at St. Xavier's College, Ranchi University. His areas of interest are the genesis of sulfide-hosted refractory gold mineralization using trace element systematics and S isotope studies in the Archean–Proterozoic sedimentary units of eastern Indian craton.

M. Sohoni is Professor in the Department of Computer Science and Engineering and Centre for Technology Alternatives for Rural Areas (CTARA), IIT Bombay. His areas of interest include combinatorial optimization, algorithms, mathematical programming, and water supply for rural areas.

J. Sreevalsan-Nair is Associate Professor at the International Institute of Information Technology, Bangalore. Her research interests are visual analytics and computational modelling. Her current focus is on the use of matrices and tensors in visual analytics of geospatial datasets and complex networks.

S. Srinivasa heads the Web Science Lab and is Dean (R&D) at IIIT Bangalore where he is a professor. He works in the area of Web science and models of the impact of the Web on humanity. He is in various technical committees including ICWSM, COMAD, and ODBASE, and the reviewer of VLDB Journal, IEEE Transactions on KDE, and IEEE Transactions on Cloud Computing.

M. D. Subash Chandran currently works at the Centre for Ecological Sciences, Indian Institute of Science. He does research in ecology, coastal ecology, and botany. His current focus is on 'All India Co-cordinated Project on Sacred Groves Ecosystem Services'.

A. Subramanian is an advanced Ph.D. scholar at the International Institute of Information Technology, Bangalore. She has previously worked in the IT industry, and her current research interests include knowledge representation and reasoning models, Semantic Web, machine learning, and graph networks.

N. M. Tara is Member of the Energy and Wetland Research Group at IISc Bengaluru. She specializes in spatial decision support systems and database management.

H. Tulapurkar received M.E. from University of Mumbai and is presently pursuing Ph.D. at the Centre of Studies in Resources Engineering, IIT Bombay. Her area of research is deep learning methods for dimensionality reduction of very high-dimensional image datasets such as hyperspectral images.

H. Vajjarapu is a Ph.D. candidate in the Department of Civil Engineering from Indian Institute of Science (IISc), Bengaluru. He has been working on the development of sustainable policies that helps in mitigating emissions from the transportation sector. He also works on the formulation of adaptation policies that helps the city's transportation system to cope with the urban floods.

P. Venkatachalam has retired as a Professor at the Centre of Studies in Resources Engineering, IIT Bombay. Her areas of research include geographical information system theory and algorithm development, indigenous geospatial technology development, and geostatistics.

A. Verma is Associate Professor of Transportation Systems Engg. at Dept. of Civil Engineering, Centre for infrastructure, Sustainable Transportation, and Urban Planning (CiSTUP), and Robert Bosch Centre for Cyber Physical Systems (RBCCPS) at Indian Institute of Science (IISc) Bengaluru, India. His research interests include sustainable transportation planning, integrated public transport planning and management, modelling and optimization of transportation systems, travel behavior, driver behavior and road safety, intelligent transportation system (ITS), traffic management, etc.

R. Vijay is Principal Scientist in the Centre for Strategic Urban Management (C-SUM) / Director's Technical Cell of CSIR-National Environmental Engineering Research Institute, Nagpur. His primary research interests include liquid waste management, urban noise pollution, and application of geospatial tools in environmental science and engineering.

S. Vinay is a Ph.D. student and Researcher at the Centre of Ecological Sciences, IISc Bengaluru. His interests include urban modelling, hydrology, and environmental flow.

U. V. S. Yadav was Research Scholar at the Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University, when his life was cut short in a tragic accident. He was a practicing conservationist and an environmental scientist.

S. Zope-Chaudhari received her Ph.D. from IIT Bombay specializing in algorithms for data security applied to raster and vector geospatial data. She is currently Professor and Head of the Department of Information Technology at A.C. Patil College of Engineering, Navi Mumbai. Her research interests include computer security and Internet of things.

Introduction

Geospatial technologies refer to a wide range of tools and techniques that contribute to mapping and analysis of earth-related natural and human activities, whether those relate to planning infrastructure or habitations or nature studies. These technologies have made tremendous progress in all aspects: data collection, processing, analysis, map preparation, and interactive visualizations. Spatial data have truly become a part of the everyday infrastructure as people use maps and mobile apps for navigation and various kinds of spatial collaborations.

Like most other countries, the use of spatial data in India has been sharply increasing in the last few years to cover infrastructure planning, monitoring of developmental works, and improving civic facilities and thrust on smart cities from governance viewpoints. The Government of India, through the Department of Science and Technology (DST), has initiated a number of developmental and research programs to encourage spatial applications and making spatial data available for citizens. A number of projects have been initiated to build spatial data infrastructures (SDIs) at national, state, and city levels. To facilitate the development of technologies, adopt standards for interoperability, and build proof-of-concept-advanced applications, DST has sponsored a large number of research projects. These initiatives and projects, in a way, represent the progress within India in the geospatial domain. The contributions from these projects can be broadly categorized as follows:

1. Development of the objectives, policies, and efforts toward building spatial data infrastructures (SDIs).
2. Research in and development of technologies that advances the capability to use geospatial information.
3. Nuanced applications of geospatial technologies.

The chapters in this book attempt to highlight the progress made in the context of each category. The book is organized into three parts as per the above categorization: spatial infrastructures, spatial technology research, and geospatial applications. As can be expected, some research works cut across this categorization. We give a brief overview of each part below.

Geospatial Infrastructure and Policy

Spatial technologies have quickly evolved as important tools for planning and governance at national and regional levels. Policies that facilitate the use of geospatial data and applications have been discussed in the first chapter of the section. The first chapter of the section discusses the policies that facilitate the use of geospatial data and applications. The proposed national data registry (NDR) is a major step toward data federation within NSDI. The third chapter discusses the development and demonstration of a prototype of such a registry. Spatial data infrastructures are important investments for state and regional governance, and this is the focus of the remaining chapters in this part. State SDIs are discussed in the context of Karnataka and Uttarakhand to highlight the approach, technical architectures, and envisaged applications. They also bring out the challenges encountered in building such infrastructures. The last chapter in this part outlines an approach to human resource development for geospatial data modeling and services for better realization of SDIs in the country.

Spatial Technology—Research Contributions

Innovation in technology has resulted in both expanding the reach of geospatial technology and increasing capabilities and applications. Indoor GIS has steadily evolved as a technique for applications such as emergency response and smart houses. The first chapter of the section discusses the policies that facilitate the use of geospatial data and applications. The authors report the development of a toolkit that facilitates querying of indoor spatial data and discuss different use cases. The ninth chapter discusses evacuation planning in the context of indoor spaces. The tenth chapter continues the theme of disaster management but sets it in the context of an urban environment and exploring the utility of volunteered geographic information (VGI) in SDIs. The eleventh chapter discusses the importance of location-based information retrieval that combines semantic processing. It discusses the role of such a service in answering queries of users about news and events relevant to the present location as well as information about everyday essential services in an urban setup. The next chapter in this section pertains to the urban landscape and discusses applying artificial intelligence techniques for urban growth simulation using large-scale data that include satellite imagery and data on road network. The thirteenth chapter discusses urban planning and smart cities in the context of facility management and proposes algorithms and multi-criteria decision support technique for location of facilities in a city. Thereafter, an attempt has been made in the subsequent chapter to gain a more generic understanding of the process of planning in a typical organizational context that discusses the mathematical basis of planning and the role of ontologies in a knowledge-based approach. The fifteenth chapter discusses ontologies in the domain of agriculture and the integration of spatial information for cotton farmers. The sixteenth chapter discusses the use of ontologies for cotton farming as well but with an objective of building a query response system. Such a system has wider applications in many other domains where schemaless data are preferable over structured datasets. The seventeenth chapter explores this aspect further and discusses the use of search algorithms in linked open data on the cloud. It discusses a knowledge representation framework that allows the thematic integration results to be organized into related domains.

The next three chapters relate to the development and use of algorithms related to spatial data applications in transportation. The chapter on evaluating climate change mitigation and adaptation policies in transportation explores GIS-based flood modeling techniques at the local level in an urban setup. The chapter on reliable trip planning discusses the planning of trips in typical transit systems and explores several algorithmic approaches to the problem. The follow up chapter reports progress in the identification of a swarm in moving object data. Identification of swarms, which could represent people or vehicles moving in groups in a transportation network, is important in many applications such as evacuation planning and traffic management.

The final four chapters in this part are devoted to techniques of handling high-resolution datasets of urban areas using LiDAR surveys, securing those using digital watermarks, and analyzing those for relevant GIS applications. The first of these reviews emerging standards in conducting LiDAR surveys useful in developing urban governance applications. The next chapter on digital watermarking reports effective algorithms to embed copyright information in both raster and vector datasets for securing those from wrongful copying and misuse. The other two chapters elaborate the use of geospatial data relating to the management of infrastructure facilities and provision of location-based services in typical urban setups.

Applications of Geospatial Technology

Applications of geospatial technologies are varied and have evolved over many years. Some recent and unique applications funded by DST have been included in this part. The twenty-fourth chapter discusses N-Dhwani, which is a point source-based noise pollution simulation module for studying point source noise pollution and suggesting suitable techniques for its mitigation. The twenty-fifth chapter discusses sensor-based monitoring and measurement techniques for studying greenhouse gas (GHG) emissions at dumpyards around a typical city and drawing up strategies for its prevention. The next chapter reports a technique for assessing the impact of development projects like that of construction of highways on the local environment including wildlife habitats. The twenty-seventh chapter discusses techniques developed for capturing the interrelationships among different interacting elements of a region for working out strategies at the local level for conservation of ecologically sensitive regions. Tools for geovisualization of local hotspots of biodiversity form a part of the discussion. The twenty-eighth chapter reports the use of geospatial technologies along with satellite imageries for development of aquaculture by assessing potential marine fishing zones. The subsequent chapter outlines and discusses an approach to assess footprints of water, carbon, and nitrogen on major crops useful in designing strategies for mitigating the impact of climate change and adapting to changes in agriculture. The last chapter of the book outlines a GIS-based technique for regional-scale delineation of mineral occurrence zones like gold prospective zones using multi-source datasets on relevant parameters.

This book covers different aspects of geospatial technology and their applications, ranging from policy and standards to developing new algorithms and techniques as well as emerging application areas. Studies in the geospatial domain included in this book serve as a reference to the progress made in the last few years and provide an overview of the span of geospatial research in the country toward evolving strategies for the future. The developing nature of technology that started as geographic information systems used for natural resources management has flourished as a mainstream technology with massive data and applications affecting governments, communities, and individuals.

We expect each chapter to serve as a reference work and will also appeal to a wide readership, including graduate students, scholars, and researchers. Some chapters hold high value for geographic information officers (GIOs) and policy makers, whereas others would stimulate further research. Case studies and projects discussed within this book have been carried out in leading institutes across India and thus provide a glimpse into the advances made in both the national and international contexts. With an accessible breadth of content, intuitively organized and indexed, this volume is considered as a useful reference in the field of GIS in the years to come.

N. L. Sarda
P. S. Acharya
S. Sen

Part I
Geospatial Infrastructure and Policy

Geospatial Policies in India and Their Implementation in Good Governance

B. Singh and S. Pandey

1 Introduction

During the First Five-Year Plan of the government, a S&T policy was announced in 1956 to give the boost of science, technology, application and development. Also, the government has opened various opportunities under various themes to promote multiple sectors towards the vision of inclusive growth in science. In the process, number of new institutions and academic institutions were set up to work exclusively in different areas. One of the important institutions, i.e. Survey of India was already in existence. The basic job of the Survey of India (SoI) was assigned to survey topographic maps on 1:50 K scale for the entire country. SoI has collected topographical details during 1965–1970 and finally printed the maps in 1980 for the use of different purposes by the government as well as by the general public. In order to distribute such maps, there was a policy under which topographical sheets were made available to the public with the written request.

With the advancement of satellite technology and also Indian capability in this area made possible to acquire large amount of topographical information by the satellite images and upscale the maps quickly. With this advancement, the availability of geospatial maps and data accessibility become easier and faster. In addition, there are number of other agencies in government like Geological Survey India (GSI), Forest Survey of India (FSI), National Informatics Centre (NIC), Directorate of Census and Ministry of Statistical and Programme Implementation have also collected huge data sets (spatial and non-spatial). In order to handle and use such voluminous data

B. Singh (✉) · S. Pandey
Natural Resource Data Management System Division, Department of Science and Technology, Ministry of Science and Technology, Technology Bhawan, New Mehrauli Road, New Delhi 110016, India
e-mail: bhoopsingh@nic.in

S. Pandey
e-mail: shubha.p@nic.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_1

for various applications, the need was felt at different levels to draw the policies to address various issues related to the data quality, standards, usage, security and civil applications. In this series, the following policies were announced by various departments of the government.

2 Existing Policies Dealing Geospatial and Non-spatial Data

Open availability of satellite data and digital forms of maps, especially when seen in the global context, has necessitated policymaking in India, over the last two decades [1, 2], such as:

- **National Map Policy (NMP), 2005**: It basically defines the scope, distribution and liberalised access of digital survey of India topographic maps to user groups without jeopardising national security. As per this policy, SOI has prepared open series topographical maps for the use of the civil purposes which were made available as per the policy.
- **The Civil Aviation Requirement (CAR), 2012**: This policy basically details procedures for issuance of flight clearances for agencies undertaking aerial photography, geophysical surveys and cloud seeding.
- **The Remote Sensing Data Policy (RSDP), 2011**: It defines the distribution process of satellite images to different category of users and their availability for attending various applications.
- **The National Data Sharing and Accessibility Policy (NDSAP), 2012**: This policy basically deals with providing an enabling provision and platform for proactive and open access to data generated through public funds available with various departments and organisations of Government of India. The Policy at present has made considerable progress to enrol about 106 Government departments to proactively sharing the data on the national portal, i.e. data.gov.in. Further efforts are on to enrol State Government agencies to adopt this policy and share their data for public use.
- **National Geospatial Policy (NGP)**: This policy is being coordinated by Department of Science and Technology and is at the drafting stage. Once, this policy is ready, it will empower people to access geospatial data for addressing their need using geospatial technologies without compromising national security. Also, it will play a vital interface between government and other agencies for doing ease of business.

3 Standards, Spatial Framework and Technologies

Open Geospatial Consortium (OGC) standards are immensely popular and adopted by government and industry and bring about a high level of focus on interoperability through open standards. OGC standards have been defined to systematise metadata, map services and Web services and host of applications services. The following key parameters are defined for standards development:

- Content standards
- Metadata standards
- Schema/data models
- Spatial framework
- Quality
- Image
- GIS services
- Portal standards.

In India, Department of Space has published NNRMS standards, 2005. Thereafter NSDI standards which were limited to metadata and SOI data exchange. Under Ministry of Urban Development, the national urban information system programme is initiated in 2004–06, adopted from NNMRS standards. As a part of this, 153 towns have been completed and verified.

4 Advancement in Technology for Data Acquisition

World over, satellite images have reached sub-meter accuracy levels operationally and with unprecedented coverage. As our capacity to handle massive spatial data sets grows, with higher bandwidth of networks, newer applications have evolved. Many more such applications that harness evolving technologies and standards need to be developed in various sectors like infrastructure development, agriculture and water resources, disaster management, land use and land management, developing state spatial data infrastructure for attending area specific problems using high resolution and accurate satellite images. This has made tremendous impact in the use of geospatial data and developing area specific problems. In the light of this, all the geospatial policies listed above have some component of supporting the technologies and applications and promoting and adopting to solve micro-level planning for good governance.

5 Other Advanced Technologies for Data Acquisition

In addition to the satellite images available in various resolutions and accuracy, the technologies unmanned vehicle, drone mapping and global positioning system (GPS)/ground penetrating radar (GPR) have made significant headway for data acquisition of high standard accuracy. For example, unmanned ariel vehicle (UAV) is of high potential and useful to support rapid mapping activities in terms of accuracy and flexibility. Its high resolution can be turned into proper higher accuracy by means of integration with existing geospatial data, i.e. airborne IFSAR data to map the disaster area immediately [3, 4]. In a similar light, unmanned surface vehicles (USVs) are useful in bathymetric surveys of reservoirs and other water bodies, towards creating data sets to help water resource planning or flood mitigation [5]. Therefore, conventional data available with various departments are considered not so much important due to the availability of such data sets. Therefore, all the data sharing policies should be facilitating the use and accessibility of geospatial data to the common person and its benefit to the society.

6 Future Scope of Geospatial Data and Policies

At present, the advancement in the development of sensors and resolution of data capturing and accuracy has made significant progress. Considering this in view, the development of infrastructure to deal such data set and attend application needs high level of expertise and computational facility which needs to be developed at various levels including institutional as well as operational level. The only issue remaining is the operationalization of geospatial data and its usage at various hierarchical level, i.e. state, district and subdistrict level. The geospatial technology for good governance at the micro level needs to be promoted to accrue maximum benefit.

References

1. Rao, M., Ramamurthy, V.S., Baldev, R.: Standards, Spatial Framework and Technologies for National GIS, National Institute of Advanced Studies (NIAS). IISc Campus, Bangalore (2015)
2. Shilpa, R., Chavan, V., Bandil, D.: Report on Design and Functional Evaluation of GIS Portals, National Institute of Advanced Studies (NIAS). IISc Campus, Bangalore (2015)
3. Tampubolon, W., Reinhardt, W.: UAV data processing for large scale topographical mapping. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **XL**-5, 565–572 (2014)
4. Tiwari, A., Dixit, A.: Unmanned aerial vehicle and geospatial technology pushing the limits of development. Am. J. Eng. Res. (AJER) **2**(1), 16–21 (2015)
5. Scerri, P., Kannan, B., Velagapudi, P., Macarthur, K., Stone, P., Taylor, M., Dolan, J., et al.: Flood disaster mitigation: a real-world challenge problem for multi-agent unmanned surface vehicles. In: International Conference on Autonomous Agents and Multi-agent Systems, pp. 252–269. Springer, Heidelberg (2011)

National Spatial Data Infrastructure India (NSDI-India)—Present Status and the Future Strategies

P. S. Acharya and S. Pandey

1 Background

India has been at the threshold of a major transformation with the adoption of a series of reforms by the government and with a positive future prospect for all-round development [1]. Priorities of the government have been set on 'Ease of Living or Doing Business', 'Inclusive Growth' and 'Good Governance' to ensure economic prosperity, environmental preservation and social welfare. Effective implementation of these priorities requires a sound integrated framework covering issues from all sectors of economy, environment and society for decision-making and governance. In order to remain sustainable, there is a need for making governance and decision-making processes more transparent and people-oriented while taking into account the local (district or subdistrict) natural resource endowment and needs. Drawing up such integrated strategies being complex and information-intensive requires utilization of various geospatial data and technologies at different levels of the planning and governance hierarchy. Use of geospatial data and technologies is essential for making sound decisions and preparing implementation plans along scientific lines towards successful delivery of benefits of various programmes and schemes. In support of implementation of these schemes and programmes, activities of various ministries/departments of the government and the states have been reoriented towards their use besides improving user's accessibility to and application of digital geospatial data sets and processes that provide the required underpinnings to decision-making.

In spite of the above, there have been several problems encountered by the end-user agencies in effective utilization of geospatial databases and technologies. Non-

P. S. Acharya (✉) · S. Pandey
Department of Science & Technology, NRDMS-NSDI,
Government of India, New Delhi, India
e-mail: psa@nic.in

S. Pandey
e-mail: shubha.p@nic.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_2

availability of data sets of right resolution and currency; inaccessibility to databases over the Web/Internet/mobile devices; lack of standardization and interoperability; inability of organizations in making their data sets amenable to quick retrieval, update and analysis; nonavailability of a sound technology platform (hardware, software, connectivity, decision support module, etc.) for collaborative map-making and processing; inadequate integration or alignment of geospatial (Geo-ICT) processes with workflows of end-user organizations/agencies; and inadequate capacity amongst the end-users have been the major bottlenecks. There is thus a need for developing and making accessible geospatial databases and requisite applications/products at national, state and local levels to support governance and decision-making.

Taking note of the requirement of the stakeholders like the government, private enterprises, academia; and the civil society organizations, National Spatial Data Infrastructure (NSDI) initiative has been jointly launched by Department of Science and Technology (DST) and Department of Space (DOS). Various agencies concerned with provision and consumption of geospatial data sets in the central and state governments have been developing geospatial databases and redesigning their geospatial data management processes for improving data sharing over the Web in the framework of the National Spatial Data Infrastructure (NSDI).

2 Introduction

To facilitate data sharing over the Web in SDI framework, various agencies have been adopting international spatial data and process standards from Open Geospatial Consortium (OGC); and International Standardization Organization (ISO). Interoperable spatial data services based on standards have been operationalized by the respective organizational data nodes/portals of various agencies for providing access to data services for automated access by applications/software utilities. Metadata (data about data) of the geospatial data sets to facilitate search and discovery of maps and data have been made accessible from the NSDI portal. At the state level, State Spatial Data Infrastructures (SSDIs) are being set up in collaborative mode between DST and the state governments for providing access to various geospatial data sets relevant to state's jurisdiction, e.g. cadastral plot boundaries, land ownership, groundwater, mining and geology.

3 Related Policies

With the launch of NSDI, efforts have been made to relax the restrictive policy regime. In 2005, the National Map Policy has been approved by the Union Cabinet providing for the preparation of Open Series Maps (OSM) for civilian use and Defence Series Maps (DSM) for exclusive use by the defence forces. NSDI initiative has been formally launched through a cabinet resolution in June 2006 with the objectives of improving access to organized collection of geospatial data, making those searchable and discoverable through their metadata and sensitizing end-users on the benefits of NSDI [2]. A two-tier coordinating structure as indicated in Sect. 7 with more details has been installed to facilitate standardization and interoperable sharing of geospatial data amongst the stakeholder agencies. A single window access mechanism of India geoportal has been made operational for providing access to metadata of various agencies.

In order to facilitate sharing of data sets and improving accessibility, a National Data Sharing and Accessibility Policy (NDSAP) has been notified by the government in February 2012. As per the provisions of the NDSAP, the National Open Government Data Portal (data.gov.in) has been implemented by the National Informatics Centre (NIC), New Delhi [3]. A National Remote Sensing Data Policy has been framed in May 2011 to govern sharing of and access to satellite data sets and images in the country. The Information Technology (IT) Act, 2000 of Ministry of Electronics and Information Technology (MeitY); the Right to Information (RTI) Act, 2005 of Department of Personnel & Training (DoPT); and various state-level data sharing and Accessibility Policies have been framed by various states to improve sharing and accessibility. Despite availability of the above policy framework, it has not been possible to search/discover geospatial data sets in an easy and user-friendly way; quickly interpret analysed results and use them in decision-making process; and seamlessly link data sets from various sources/agencies. As a part of the implementation of the NDSAP, Open Government Data Portal has been operationalized for providing open access to government data sets.

Towards overcoming the above problems, it has been decided to frame a more comprehensive National Geospatial Policy (NGP) covering geospatial data, products, services and applications. Formulation of a National Topographic Frame and a National Image Frame covering the National Foundation Spatial Data Sets, formation of a National Geospatial Authority (NGA); and user-friendly sharing of geospatial data sets for easy accessibility by various stakeholders have been envisaged in the proposed Policy.

4 Databases, Metadata and Architecture

Over the past years, efforts have been made by NSDI to compile and serve metadata of different partnering agencies and reengineer the feature data sets for improving their use. The NSDI partnering agencies have large amounts of feature data sets/images and are presently managed in file-based systems largely through quadrangle sheets/scenes without appropriate strategies for fully automated search, querying and processing. This is evidenced in the present ways to store and manage metadata sets as per the indices of quadrangle sheets/scenes of data providing organizations. Identification and maintenance of feature data sets, in particular, thus pose a major challenge to the stakeholders, thereby inhibiting appropriate use of the geospatial data in real-life applications.

Data providing agencies partnering with NSDI at the national and state levels have thus developed and operationalized relational databases for effective search, discovery and sharing of processable data over the Web. Development and maintenance of the India geoportal have been the single window gateway to provide standardized Web-accessible services of the data providing agencies like Survey of India (SOI), National Remote Sensing Centre (NRSC), Geological Survey of India (GSI), Forest Survey of India (FSI), Central Water Commission (CWC), Central Ground Water Board (CGWB), and Ministry of Statistics and Programme Implementation (MoSPI) and several state governments.

Besides the geospatial metadata accessible from the India geoportal in BIS 16439 [4]/ISO 19115 standards, a National Data Registry (NDR) prototype containing additional metadata has been demonstrated with registry and catalogue services to facilitate automated search and discovery of map, feature and coverage data services. The NDR contains metadata relating to the feature data definitions, application schemas and classification codes for uniquely identifying geospatial data from various data nodes and correct interpretation of results of geospatial analyses carried out using services from the data nodes of multiple organizations.

A standard operating procedure (SOP) has been prepared for Survey of India's 1:50,000 scale topographic data reengineering towards development of processable GIS data and provision of Web Feature Service (WFS) in OGC Geography Markup Language (GML) format [5]. Preprocessing and cleaning of the feature data sets have been carried out for the publication of the Web Map Services (WMS) with a countrywide coverage from the data node. Development of SOI's foundation data layers for Haryana State and publication of feature data sets for WFS/GML access from NSDI portal has been completed and demonstrated. SOI has been in the process of providing WFS/GML for all its 1:50,000 OSM data. Collaborative harmonization of administrative boundaries for the development of the National Foundation Spatial Data (NFSD) up to the revenue village level has been initiated with the involvement of the concerned State SDI set-up in one of the states as a demonstration for adoption of the procedure in other states.

Establishment of a formal linkage with Bureau of Indian Standards (BIS) has led to the development and publication of national standards. ISO standards critical to the development of SDIs at the national and state levels like conceptual schema language (ISO 19103), rules for application schema (ISO 19109), Web Map Service (WMS—ISO 19128), Web Feature Service (WFS—ISO 19142), Geography Markup Language (GML—ISO 19136) are being co-branded as national standards by the BIS. Data content standards on surface geological mapping, forests and soils have been prepared for publication of national standards through stakeholder consultation.

In addition to the national agencies above, state government agencies in Karnataka, West Bengal, Uttarakhand, Odisha, Jharkhand and Haryana holding spatial data sets have collaborated with NSDI as partners for improving their data management procedures. Under the collaborative State Spatial Data Infrastructure (SSDI) initiative with the concerned state governments, state geoportals have been operationalized for provision of Web-based data services. To ensure regular flow of data sets in the state governments from the lower area (villages, block and district) levels to the state level for updating the state geoportals, a coordination mechanism with a GIS set-up has been established at each district.

More effective management and use of the data/images requires their registration, storage and sharing so that those could be put to appropriate processes of governance in line with the provisions of NSDI Resolution, 2006. Registering spatial data sets involves capturing definition/description of each data item, making each item persistently and uniquely identifiable by assigning them unique identifiers, organizing/maintaining the associated classification systems or the code lists for reference by all and a suitable governance system along with procedures for managing the related registers. Moreover, provision of geospatial resources like data, metadata and catalogue is governed by various standards that evolve or change over time. For consistent processing and interpretation of the data and services, it is essential that these changes are kept track of and properly managed. Establishment of registers in the form of a registry with a suitable management structure is thus essential. ISO 19135 prescribes standard specifications on procedures for registration of geographical information items and establishing a management structure.

In line with the end-user requirement stated above and the developments in geospatial technology-related data and process standards, a proof of concept (PoC) of an NDR for NSDI was implemented and demonstrated using data services from Survey of India, National Remote Sensing Centre and the Ministry of Railways' IRCTC train running data over a week from the Northern Railway [6]. The PoC successfully demonstrated the utility of a sample registry, technical architecture and led to the development of an RFP for a real-life NDR implementation for engaging a system integrator. Potential system integrators have provided possible sample solutions around different open source and COTS-based products for the selection of a suitable partner for NSDI for the NDR development. This forms a part of a separate chapter (Chap. 3) in this volume.

NSDI uses service-oriented architecture (SoA) in the management of the data and related services both at the national and state levels. A typical SoA (publish-find-bind model) widely used in the provision of geospatial information services is given in Fig. 1.

National- and state-level data nodes use both COTS (e.g. Oracle Spatial) and open source products (e.g. PostGreSQL/PostGIS) for the management of the data sets on the geoportal. Data services are published in the NDR using the registry service accessible at the India geoportal that is accessed by the end-users and application utilities for finding and discovering the data sets. Once identified, the end-user or application utility binds itself with the data provider for getting the data sets or for meeting any requirement of update of the provider's database, if necessary, as per the service-level agreements.

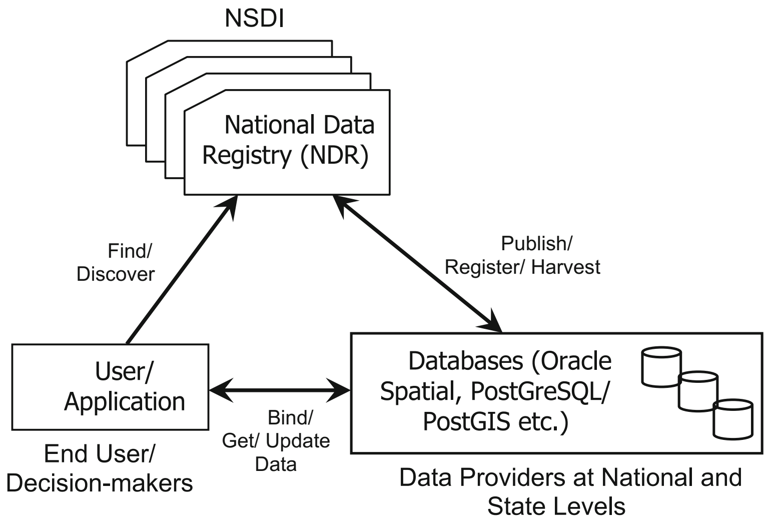


Fig. 1 Service-oriented architecture of NSDI-India—Publish-Find-Bind Model with National Data Registry

5 Geospatial Cloud

Efforts are currently being made by these nodal agencies to make interoperable Web Feature Services (WFS) accessible for download of feature data in standard encoding specifications (e.g. OGC's Geography Markup Language) for processing on a GIS platform. More national- and state-level geoportals are likely to be deployed in the near future for provision of feature and attribute data services over the Web. With Web-based data services gradually getting accessible from various agencies, emphasis in NSDI is shifting from enabling 'data' services to 'product' or 'application' services to support provision of GIS products or applications over the Web [7]. In the context of the above service-oriented framework and the GIS application requirement of the end-users and decision-makers, three of the cloud service modes of usage, i.e. Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), are proposed to be used in the development and implementation of geospatial products and applications. Some of the service models of geospatial cloud as a part of the SaaS proposed to be used include GIS as a Service (GaaS), Application as a Service (AaaS) and ready-to-use Data sets/Imagery as Service (DaaS or Iaas). An experimental geospatial cloud has been developed and demonstrated around the State SDI data sets from West Bengal to the representatives of NSDI and State SDI partnering agencies.

A proof of concept around the above cloud service models is thus under development to explore and demonstrate the utility of the cloud services using sample geospatial data sets prepared for the city of Varanasi (Uttar Pradesh) by Survey of

India under the National Urban Information (NUIS) project of Ministry of Urban Development, Government of India. In addition to the imageries, photographs and topographic sheets of the area, existing data sets are proposed to be made available/accessible for integration with the GIS-ready digital data is of 1:2000 scale on UTM projection and WGS 84 datum as per SOI NUIS data model. Additional 3D content up to LOD-3 (for selected areas only) with appropriate methods and techniques to meet specified quality and accuracy standards has been proposed to be used in the PoC of the geospatial cloud.

Some of the other proposed characteristics of the PoC include unilateral provisioning of services on the cloud by NSDI/SOI, use of appropriate communication bandwidth; providing for placing resources on a specific geographical zone as per the SOI/NSDI's requirement; near real-time service provisioning and deprovisioning, etc. The PoC is expected to throw light on the performance and sizing of an appropriate geospatial cloud facility for the automated processing and application/product development by the end-user.

6 R&D Activities

For effective utilization of the above Web-based data and application services by the stakeholders, efforts are being made to develop application prototypes and demonstrate applications with the involvement of user ministries/departments/state agencies and private enterprises. Assignment of unique IDs to plots/cadastres with inputs from the Department of Land Resources (DoLR), train running information and GIS enablement of the Railways with the Ministry of Railways; developing 3D SDI for urban governance; setting up of a GIS Cell in National Disaster Management Authority (NDMA) for consuming WMS/WFS/GML for disaster mitigation; development of GIS applications in watershed development for Watershed Directorate of the Government of Karnataka have been some of the initiatives taken in this direction. The NSDI 2.0 Mission is proposed to be launched for the development of GIS applications for decision support at different levels of governance. Some of the important priority areas for the future include establishment of Geospatial Research Hubs and conducting R&D in areas like urban governance; village information systems, health GIS; watershed management; disaster management; coastal hazards and urban flood modelling; land resource management; agriculture; geodesy; sustainable development goals; assessment of regional hydrological systems using space-borne gravity observations; heat flux of urban cities; landslide forewarning systems with specific reference to the north-eastern region; interoperability and 3D SDI; large-scale mapping at the level of gram panchayats or wards. A gram panchayat or a ward is invariably a functional area or an electoral unit at the local level comprising of a few revenue villages or households for developmental administration as a part of government's effort towards decentralizing governance in the rural or urban areas for real-life geospatial application development.

Research & Development in Geo-Information and Communication Technologies (Geo-ICT) domains like ontologies and reasoning, high-performance computing and cyberGIS; geodesy; geospatial interoperability and analytics; cartographic generalization; 3D SDIs; and volunteered geographic information will be pursued with the involvement of a network of academic and research institutions.

Over the past years, training modules have been developed for large-scale building of technical capacity amongst staff and personnel of central and state government agencies and scientists/faculties from research and academic institutions. Development of additional training modules/kits in emerging areas of various relevant domains including Geo-ICT as mentioned above has been an important priority for NSDI.

7 Management Structure

At the national level, NSDI is managed through a two-tier coordinating arrangement as per the provisions of the NSDI Resolution, 2006, approved by the Union Cabinet as indicated in Sect. 3.

The apex body, called the National Spatial Data Committee (NSDC), is chaired by Minister of Science and Technology with the concerned secretaries to Government of India as the members. NSDC provides the overall policy direction to the activities of NSDI and advises, guides, authorizes and approves its overall implementation, reviews performance and authorizes/approves any mid-course corrections—including resource prioritization/reallocation. Surveyor General of India and the Director, National Remote Sensing Centre, act as member secretaries of the NSDC. A few professional experts in the domain of geospatial technologies also serve as members.

At the lower tier, the NSDI Executive Committee (EC) is co-chaired by the Surveyor General of India and the Director, National Remote Sensing Centre, with the heads of different data providing agencies of the central government as its members. EC is entrusted with the task of guiding NSDI in implementing the policies/decisions of the NSDC and NSDI's day-to-day operations. Working Groups/Technical Committees have been constituted in different thematic domains like 'Geospatial Metadata', 'Data Content Standards', 'Ínteroperability', 'National Data Registry', 'Geospatial Cloud'; 'Administrative Boundary Harmonization' have been constituted with the participation of representatives from the partnering agencies from the national and state levels and leading experts from the IITs, universities and R&D institutions.

Invariably, two-tier coordination structures have also been in place for the State SDIs with the Steering Committee chaired by the chief secretaries with secretaries/principal secretaries of the concerned departments as members at the upper tier for policy-level guidance for the State SDI set-up. At the lower tier, the Executive Committee (EC) is chaired by the secretary/principal secretary, planning and coordination or science and technology with the directors of various state-level geospatial data handling agencies as its members for implementing the decisions of the Steering Committee.

Above coordinating structures are technically advised at the national level by a Steering Committee and several Technical Committees with members drawn from various leading R&D institutions, IITs, universities and laboratories. A Standing Finance Committee (SFC) under the chairmanship of Secretary, DST, with members from NSDI partnering agencies reviews allocation of funds to different activities.

8 Conclusions

Three major accomplishments of NSDI, so far, have been (i) the establishment of the NSDI mechanism with the enrolment of the stakeholders from national and state levels (ii) the institutionalization of geospatial data and process standards-making and adoption with the involvement of the Bureau of Indian Standards (BIS); and (iii) setting up the technical infrastructure. The technical infrastructure covers hardware, software; connectivity; data; metadata; manpower of NSDI and State SDI data nodes for provision of standards-based data and metadata services and applications [8]. Some of the other accomplishments of NSDI include formulation and adoption of a shared vision and goals by various partnering agencies for achieving the NSDI aspirations; framing and implementing relevant policies for a governance framework for data management; and training and capacity building of scientists, officials, staff and personnel of the partnering agencies. NSDI is now poised towards soon evolving into a national geospatial information infrastructure with its focus gradually shifting to building working linkages with the stakeholder; strengthening the development and sharing of standardized applications and solutions; framing supportive policies and acts for its more effective utilization. Setting up research hubs for pursuing R&D in the emerging areas of Geospatial Information Science and Technology for its continuous reorientation and upgradation; and nurturing international collaborative partnerships with Open Geospatial Consortium (OGC), International Standardization Organization (ISO), and United Nations Global Geospatial Information Management (UNGGIM)'s committee of experts have been set as priorities for NSDI in the coming years. All these are expected to help NSDI achieve its vision of sustaining the economic growth in the short term that in turn will usher in the era of sustainable future.

Acknowledgements The authors wish to sincerely thank Secretary, DST, for his overall guidance and support in carrying out the work reported in this article. All kinds of support and inputs from various Chairmen and Members of the NSDI-related Committees, Working Groups; NSDI and State SDI Partnering Agencies; Principal Investigators; and Experts received in the implementation of the NSDI and State SDI activities are thankfully acknowledged.

Bibliography

1. National Institution for Transforming India (NITI) Ayog, Government of India, 2017: Report on "India—Three Year Action Agenda—2017–18 to 2019–20", New Delhi. http://niti.gov.in/writereaddata/files/coop/India_ActionAgenda.pdf
2. Gazette of India, Jul 2001, 2006, DST, New Delhi, 13 June 2006, Resolution on "Constitution of National Spatial Data Infrastructure (NSDI)". https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/NSDIGazetteConstitution.pdf
3. Open Government Data Portal. https://data.gov.in (2018). Accessed 21 Mar 2018
4. Bureau of Indian Standards (BIS), 2016: "BIS IS 16439, 2016, Metadata Standard for Geospatial Information"
5. Department of Science & Technology (DST), 2015: NSDI Report on "Draft Standard Operating Procedure on Re-Engineering of 1: 50,000 Survey of India Topographic Data and Provision of GIS-Ready Data in ISO/OGC-compliant Web Feature Service (WFS)/Geography Markup Language (GML)"
6. Department of Science & Technology (DST) Technical Committee Report, 2015: Report on "National Data Registry (NDR) for NSDI" under the chairmanship of Prof. N. L. Sarda, Department of Computer Science & Engineering, IIT Bombay
7. Department of Science & Technology (DST) 2016: "Aspirational Document for NSDI, 2011": Consultancy Development Centre (CDC), Ministry of Science and Technology, Government of India, New Delhi. https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/aspiration.pdf
8. NSDI India Geo-portal. https://nsdiindia.gov.in/ (2018). Accessed 21 Mar 2018

National Data Registry (NDR) for the National Spatial Data Infrastructure (NSDI)

P. S. Acharya and N. L. Sarda

1 Introduction

Over the past years, efforts have been made by NSDI to compile and serve metadata of different partnering agencies and re-engineering the feature data sets for improving their use. The NSDI agencies have large amounts of feature data sets and are presently managed in file-based systems largely through quadrangle sheets without any sound strategies for their search and querying. This is evidenced in the present ways to store and manage metadata sets at the NSDI server as per the indices of the quadrangle sheets. Identification and maintenance of such feature data sets pose a major challenge to use of the data. More effective use of the data requires proper registration of the feature data sets so that the data could be put to appropriate processes of governance in line with the provisions of the NSDI Resolution, 2006 [1]. Registering spatial data sets involves capturing their definitions and descriptions, making them persistently and uniquely identifiable by assigning them identifiers and organizing/maintaining the associated code lists.

Data registries are in the process of development in different countries to facilitate governance of data sets. Infrastructure for Spatial Information in Europe (INSPIRE) has set up a registry to manage spatial data across various member states of the European Union [2]. The member states have set up national registries to support implementation of the INSPIRE initiative. Open Geospatial Consortium (OGC) has developed specifications for registry services for enabling data providers to register their data sets and services at a common central place [3]. The electronic business Registry Information Model (eb-RIM) specifications have been adopted in devel-

P. S. Acharya (✉)
Department of Science & Technology, NSDI Division, Government of India, New Delhi, India
e-mail: psa@nic.in

N. L. Sarda
Computer Science & Engineering, Indian Institute of Technology Bombay, Mumbai, India

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_3

oping the OGC specifications for the registry services [4, 5]. ISO 19135 provides specifications for registering data items in a register and related governance processes [6].

2 Background

Taking into account the problems in the management of spatial data sets in different NSDI partnering organizations and the absence of a central register for keeping track of the data sets, it has been decided by NSDI, DST to evaluate the concept of NDR and its utility in possible application development to address specific end-user needs. The job consists of working out the specific roles that NDR seeks to play, its scope and structure; develop a few sample but related domain models and the application schemas for implementation of Web-based applications; evaluate the service orientation aspects of the prototype implementation; study some portals playing the role of registry; and recommend steps for actions at the data source level, at the registry (or NSDI) level, and for an early realization of the NDR.

This article has been divided into sections on role of NDR, its scope and structure, an approach for its realization covering the technical specifications, results of the prototype implementation, and the recommendations.

3 National Data Registry for NSDI

NSDI has been working with national- and state-level data providers to facilitate sharing of spatial data among the stakeholders. This is to be achieved by (i) compiling and providing metadata sets to facilitate discovery and access of spatial data; (ii) setting up of the spatial data nodes for the provision of interoperable spatial data services (e.g. WMS, WFS); and (iii) establishing a single-window access mechanism to the metadata and spatial data sets through installation of a clearing house and geoportals. Metadata sets in BIS standard format [7] have been compiled from various Survey Agencies and State level Line Departments and made accessible from the geoportals (e.g. India Geo Portal at the National and State Geoportals at State Levels).

3.1 Problems with the Present Set-up

In the present set-up, metadata sets can be accessed by making queries/searches on the meta-database stored in NSDI or State SDI Data Nodes that are accessed through applications. The following are some of the limitations:

(i) Inadequate access to and maintenance of resources describing the data in order to allow for its correct processing and interpretation by applications for real-life decision support.
(ii) Non-availability of a mechanism to keep track of and manage changes in various rules and standards those govern the above resources for consistent processing and interpretation.
(iii) Even if feature metadata is to be provided by individual data nodes from the data providers, application schemas and code lists (domain value sets or enumerations with thesauri) should be made accessible from the geoportal for search and use. Presently, application schemas or code lists are not accessible.
(iv) Detailed specifications of coordinate reference systems used in data sets, units of measurement, identifier namespaces.
(v) Metadata sets are retrieved for each data unit of the provider (e.g. topographical or geological quadrangle sheets) instead of feature data-wise.
(vi) Inadequate updation of metadata corresponding to changes/updates in spatial data.
(vii) Spatial data sets are not appropriately organized on databases systems (but on file systems) causing major difficulties in accessing/searching data for decision support.
(viii) Although, there are services accessible from the geoportals, there is no metadata about the services to facilitate their search and discovery.
(ix) Catalogue service as per OGC standard is not accessible.
(x) Machine access to metadata requires XML implementation, but presently no metadata from the portal is in the machine-accessible form.

While these problems are continuously encountered by end-users, concepts, technologies and standards have further evolved and are gradually getting available to address some of the issues. Concepts like geospatial information infrastructure, service-oriented architecture, registry, technologies like geospatial cloud, geospatial platform, interoperability, geosemantics, and standards like catalogue service on web, procedure for registration of geographical information items (ISO 19135), geography markup language (ISO 19136) are thus required to be explored in this context, and prototypes demonstrated for possible adoption and improvement of the performance of Geospatial Decision Support Systems (Geo-DSS). The concept of "registry" in particular requires to be evaluated to overcome some of the above problems by studying a few existing portals acting as registries, available standards and tools, review of related literature, defining a role for the NDR, its scope and structure, and recommending a suitable approach for its development and realization for real-life applications.

3.2 The Concept of NDR

With several agencies geographically distributed and providing sharable geospatial data, there is a need for end-users—humans or machines—to search, discover, understand and access them over a distributed network for use in applications. Towards this, the data-providing agency must draw up and share a set of metadata consisting of the data type, meaning of the features and their associated attributes; underlying schema or structure; classification codes, used if any, in the attributes; location of the data or its bounding box; the coordinate reference system; time, accuracy or resolution of acquisition/survey; standards used in the data format; usage terms and conditions, etc. All these are required to be made accessible from a common catalogue, called a registry, in order to enable the end-user assess the fitness of the data sets in meeting his purpose or application requirements. On satisfactory assessment, the end-user gets access to the data sets from the provider with the help of Application Program Interfaces (APIs) to facilitate automated usage of the data in his/her software packages/products. Such a common and appropriately managed registry at the national level—called a National Data Registry (NDR)—is vital to successful GIS application development by the end users by utilizing geospatial data from multiple sources of data. In effect, NDR links the data providers with the end-users in a service-oriented architecture by providing access to end-users to a common up-to-date system of catalogue.

Interoperability specifications or standards in the context of the above provider–user link assume significance considering the heterogeneity of data sets made accessible from multiple sources in terms of their meaning or semantics, their thematic content and the underlying data formats/structures. Such standards are available from ISO, OGC, BIS or other standard bodies for use in the development of the NDR that are open, precise, unambiguous; comprehensive; extensible; widely acceptable; and consumable by both humans and machines.

NSDI's present metadata contains information regarding spatial data from various data-providing agencies as per Bureau of Indian Standards (BIS)' IS 16439:2016 Metadata Standards for Geospatial Information. IS 16439 standards are founded on the ISO 19115 specifications and have evolved out of NSDI Metadata Standards 2.0 [7]. These specifications provide for citation, quality, features, lineage, temporality, reference systems, extent, etc., and are expected to be served to users or their applications as catalogue services that offer the functionalities to store and serve metadata about the data sets. However, the current catalogue specifications do not define formal procedures for the modification of the stored information. The user does not have the ways and means to ascertain, for example, by whom and when some data sets or services have been modified, added or removed from the database or catalogue.

Since data, metadata and standards undergo changes due to data life cycle management in data-providing agencies and development/adoption of new standards or evolution of existing ones, there is a need to keep track of the changes. Keeping track of the changes facilitates end-user's sustained access to data and metadata and support correct interpretation of their analyses. NDR provides the framework for tracking such changes. Defining the roles for the NDR and its stakeholders or associated actors in this context assumes importance.

Moreover, provision of these resources like data, metadata and catalogue is governed by various standards that evolve or change over time. For consistent processing and interpretation of the data and services, it is essential that these changes are kept track of and properly managed with the support of an appropriate management structure. Establishment of registers and registries with a management structure is thus essential. These are required over and above the establishment and maintenance of data and service catalogues providing standardized catalogue services capable of getting consumed by end-users or applications.

ISO 19135 prescribes standard specifications on procedures for registration of geographical information items and establishing a management structure. As per the standard document, a "register" is defined as a set of files containing identifiers assigned to items with descriptions of the associated items. An identifier is a linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated. A "registry" is defined as an information system on which a register is maintained.

It is therefore essential to maintain registers of different geographical data items and keep track of the data item status for the benefit of the end-user. The registers should be managed through proper governance procedures. ISO 19135 prescribes a common structure for the registers like identifier, name, definition, item class and status. Some of the attributes like description, date of acceptance are treated as optional items.

3.3 Roles of the NDR and Its Stakeholders

The data item identifier is a positive integer that uniquely identifies the item in the containing register. The item class identifies a class of geographic information items specified in a technical standard. In the implementation described hereafter it also defines any additional attribute required for describing the items of that class. The item status clearly identifies whether the item is valid at the time it is queried: "valid" means that it can be used in the production of new data sets.

The content of a register can change over time since there might be the need to introduce new items, retire existing items, replace (supersede) existing items with other items, clarify an existing item to correct the phrasing without changing its semantics, etc.

Item status of an item is an indication at any moment in its life the status of the item like not valid, valid, superseded or retired. An item starts in the "Not Valid" status at the time of a data-providing organisation creating a proposal for a new item.

The standards defined in the ISO 19135 could be used to assign various roles to the NSDI that proposes to remain the nodal organization for managing the NDR, while the data-providing organizations could be defining and maintaining their metadata details, application schema, services, coordinate reference systems, code lists, etc., in their respective databases for access. At the minimum, they are expected to provide the WFS, WMS and the CSW. NDR should also consider allowing the private agencies to put up their data and metadata services in the registers.

ISO 19135 specifies the following roles involved in the management and use of a registry and the registers it contains. The definitions are mainly taken from ISO 19135.

Registry Manager
The registry manager is the entity that maintains the registry service operational.

Register Owner
The register owner is an organization that establishes a register and is responsible for the content of the register in terms of intellectual property and management. The register owner can delegate the management of the register to a register manager.

Register Manager
The register manager is responsible for maintaining the register. The register owner, if unable to play this role, can appoint a different organization.

Submitting Organization
An organization authorized by the register owner to propose changes to the content of a register.

Control Body
The control body of a register is a group of technical experts who make decisions regarding the content of the register.

Register User
A register user accesses a registry in order to use one or more of the registers held in that register.

4 Scope and Structure

NSDI uses service-oriented architecture (SoA) in the management of the data assets and related services. A typical SoA (publish–find–bind model) widely used in the provision of geospatial information services is given in Fig. 1 (page 12).

Based on the above service-oriented model, a detailed architecture is proposed for the implementation of registry and other associated components:

It is important to note that this architecture primarily separates any content in one of two categories—spatial objects representing phenomena in the real-world (the "data") and the other items of spatial information, i.e. data description, referenced by or related to the spatial objects, i.e. "metadata", but in a wider sense than geographic information metadata as understood in ISO 19115.

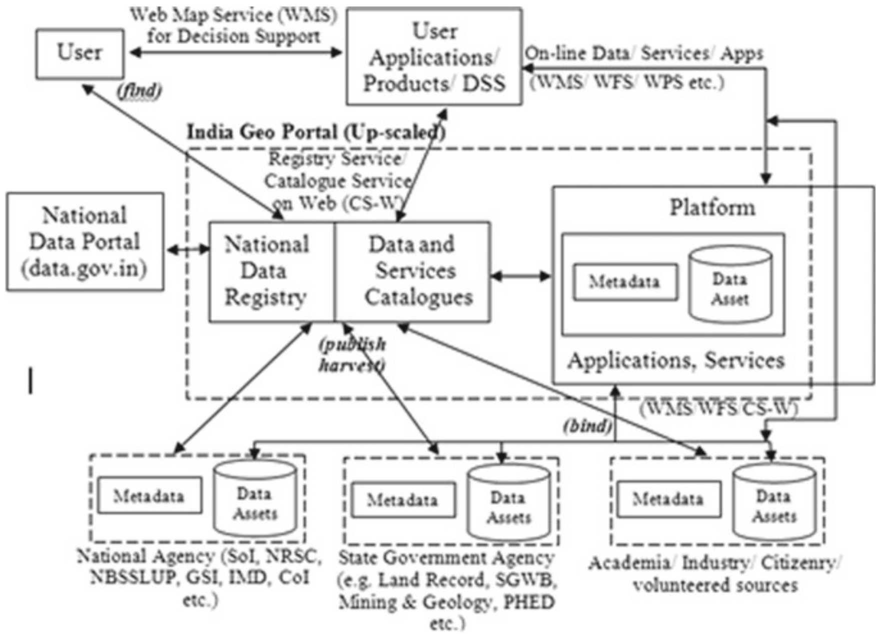


Fig. 1 Overview of NSDI technical architecture shown with National Data Registry and data/service catalogues [8]

Data Assets
Data Assets form the core component of the NSDI. These are GIS-processable data sets and could be accessed by Web Feature Services (WFS) by applications and services in the NSDI platform. The Data Assets belong to different national, state agencies or stakeholder groups in the academia, industry or non-governmental sectors. The Data Assets are maintained by the concerned agencies and made accessible as services.

Metadata
Agencies publishing their Data Assets' information in the data catalogue of the India Geo Portal must make their metadata accessible from their website or portal or the local registers. Setting up the local registers may be considered as an additional job to be completed by the agencies participating in the NDR initiative. Over the past years, various agencies have been establishing their NSDI node with a view to compiling and making accessible their metadata by compiling and uploading those to the India Geo Portal. Considering this additional component to the metadata, it is essential that metadata in respect of each of the Data Assets are maintained by the agencies themselves at their local registers of the data nodes for harvesting and automated updation at the NDR node.

Services
Web Map Services (WMS) provide access to ready-made Data Assets and products for use in applications. Web Feature Service (WFS) helps access to Data Assets for download in different formats including geography markup language (GML). The agencies may consider providing data or service catalogues to the end-users and applications to facilitate discovery and access.

India Geo Portal
Presently, India Geo Portal provides product metadata available for each quadrangle sheet for visualization. In addition, it provides WMS for a select group of survey agencies like SoI, NRSC and FSI or state government (e.g. Government of Karnataka) who have made their data sets accessible from their data nodes. WMS could be viewed concurrently from multiple sources at the India Geo Portal. The India Geo Portal is proposed to be upgraded with additional features like National Data Registry comprising of data/service catalogue and the Geospatial platform for provision of application services (i.e. processed products) for decision support.

National Data Registry is expected to consist of a set of register with items like data specifications indicating the semantics and the characteristics of the types of spatial objects in a data set. The spatial object types provide a classification of the spatial objects and determine among other information the properties that any spatial object may have (be they thematic, spatial, temporal, a coverage function, etc.) as well as known constraints (e.g. the coordinate reference systems that may be used in spatial data sets). This information is, in principle, captured in an application schema using a conceptual schema language standard, which is a part of the data specification. As a result, a data specification provides the necessary information to enable and facilitate the interpretation of spatial data by an application. However, presently, existing spatial data sets are not adequately documented. Thus, in the current proposal, it is proposed that the survey agencies participating in the NDR prototype experiment make concerted effort to document their data assets for building the registers.

With the present emphasis on the transition of NSDI from the development (enabling) phase to the implementation (performing) phase, the ability to consistently reference these resources in the registry through unique identifiers is a vital requirement. There is thus a need to collaborate in a long term on defining common specifications for metadata, data, network services as well as data and service sharing, monitoring and reporting. These are useful in consistently establishing interoperability at the organizational, semantic and technical levels in the NSDI infrastructure. Technical guidelines, data models and code lists are required to be hosted in document and UML repositories, and should be accessible through and maintained. In the long term, a separate dedicated agency may be assigned the task of maintaining these resources on behalf of NSDI.

5 Approach

There are several key characteristics associated with a register. Every item in the register is associated with a unique, unambiguous and permanent identifier. A register keeps track of changes so that data created in the past can still be interpreted completely and correctly. Superseded or retired register items are also maintained in the registry for future reference.

Worldwide, experiences suggest that the number of registers that eventually are required in an infrastructure is significant. A clear and sustainable operational model forms a key part of the set-up of the infrastructure. However, initially the following kinds of spatial information items could form registers that could be implemented:

Feature concept dictionary,
Feature catalogue,
Application schema,
Code list,
Coordinate reference systems and operations,
Units of measurements,
Spatial object identifier namespaces,
Glossary, etc.

The feature concept dictionary establishes a set of feature-related concepts that may be used to describe geographic information. A common feature concept dictionary is proposed to be maintained for all NSDI application schemas and feature catalogues. Conflicts in definitions and meanings of geographical features (e.g. water bodies versus lakes versus ponds) could be overcome by this register. Naming of features at the national and state levels differs significantly and needs to be addressed by setting up a common dictionary accessible to all.

A feature catalogue assumes significance at the time of data acquisition, processing and use. A catalogue containing definitions and descriptions of the spatial object types, their attributes and associated components occurring in one or more data sets, together with any operations that may be applied is essential for access by end-users for correct data processing and interpretation.

As per the ISO documentation, an application schema is a conceptual schema for data required by one or more applications. It is a part of data specification and specified in UML. Each data-providing agency at the national and state levels should publish their application schemas in UML/XSD formats for access by human and software users.

A code list is a dictionary describing the attribute value domains for selected property types in a feature catalogue/application schema. The value domain is not fixed in the feature catalogue/application schema, but is managed separately.

Coordinate reference systems and operations are managed in a dictionary of coordinate reference systems, datums, coordinate systems and coordinate operations which are used in data sets.

A dictionary of units of measurement, which are used in data sets, should be maintained as a register. For example, coordinates are presented as geographical latitude/longitude values or in metres.

Spatial object identifier namespaces bear significance considering the fact that similar spatial objects are maintained by different agencies or content providers for varying purposes. A mechanism is required to guarantee uniqueness of object identifiers across various content providers. Identifier namespaces are thus required to be managed.

A register of general terms and definitions used in NSDI data specifications should be prepared and maintained for access by all.

6 Prototype Implementation

A prototype of the complete system depicted in the above architecture has been implemented for a comprehensive demo of the utility of the registry [8]. The subsystems of the architecture proposed to be implemented (Fig. 1) include

(i) Database consisting of data assets and the metadata at individual nodes of data providers;
(ii) The registry;
(iii) The platform for developing GI applications/products and services;
(iv) The Geo Portal as a single-window access to data/services/applications;
(v) Decision Support Systems based on end-user workflows capable of consuming various data sets/services/applications to support decision-making by end-users.

(i) Database

The database of a data provider consists of feature data (Data Assets) along with its metadata. The data and metadata from the data provider's node should be accessible or shareable as WMS, WFS/GML, CS-W. Both feature metadata and service metadata should be stored as per ISO 19115 and ISO 19119 specifications. Data updation is an important component and may be implemented either through database protocols in a local area network from a GIS or using WFS-T from a remote surveyor in the field. Each node should implement a relational database of sample feature data sets on open source or commercial tools for demonstration of services like WMS, WFS/GML, WCS and CS-W.

(ii) The Registry

The registry should contain the description of the data in the form of feature concept dictionary, object catalogue, application schema and a set of procedures for registration of various items maintained in the registry. ISO 19135 standard specifications should be followed in the implementation of the registry.

(iii) The Geospatial Platform

The Geospatial Platform prototype should be developed and implemented to test the utility of the services from various data nodes, processing of the data sets and provision of application/products services to the end-users for decision support. The platform should be accessible as a portal for use by end-users for processing of the data sets for delivery of products/applications or access to ready-made applications from the web or mobile devices.

(iv) The India Geo Portal

The India Geo Portal should be upscaled to provide a single-window access to the data and product services from multiple data providers along with the Catalogue Service on Web (CS-W). A comprehensive catalogue of all the data sets, services, applications and products should be made accessible from the India Geo Portal.

(v) Decision Support Systems

Decision Support Systems based on end-user workflows capable of consuming various data sets/services/applications from either India Geo Portal or Geospatial Platform Portal should be established to support decision-making by end-users.

7 Implementation Result

In the context of the above framework, different components of the sub-systems have been developed. Experiences of a prototype implementation of the catalogue service along with features like metadata insertion/deletion/modification/harvest available at selected NSDI nodes have been made use of in the implementation of a use case relating to the Railways [8].

The use case involves locating a moving train between various stations of the Northern Railway on a topographic sheet of Survey of India or/and with Bhuvan imagery of National Remote Sensing Centre (NRSC) of Indian Space Research Organisation (ISRO) along with departure and arrival information at each station. Relevant data sets on departure and arrival of trains have been taken from the Northern Railway and the Indian Railway Catering and Tourism Corporation (IRCTC) of the Indian Railway. Existing data services from NSDI, Surveykshan and Bhuvan Portals have been made use of by a user on the platform of Karnataka Geo Portal. The GeoNetwork tool has been used in the implementation of the sample data registry along with metadata management and catalogue service for demonstration. IIT Bombay implemented the proof-of-concept registers on the GeoNetwork for demonstration. Application schema for SoI is available in UML/GML implementation and developed using Enterprise Architect tool for developing the schema register. Feature concept dictionaries (samples) on topography, surface geological mapping, soils, forests have been developed along with the feature catalogues and code lists for demonstration of registers. A sample screenshot of the prototype implementation of the application has been shown in Fig. 2. As per the use case of train

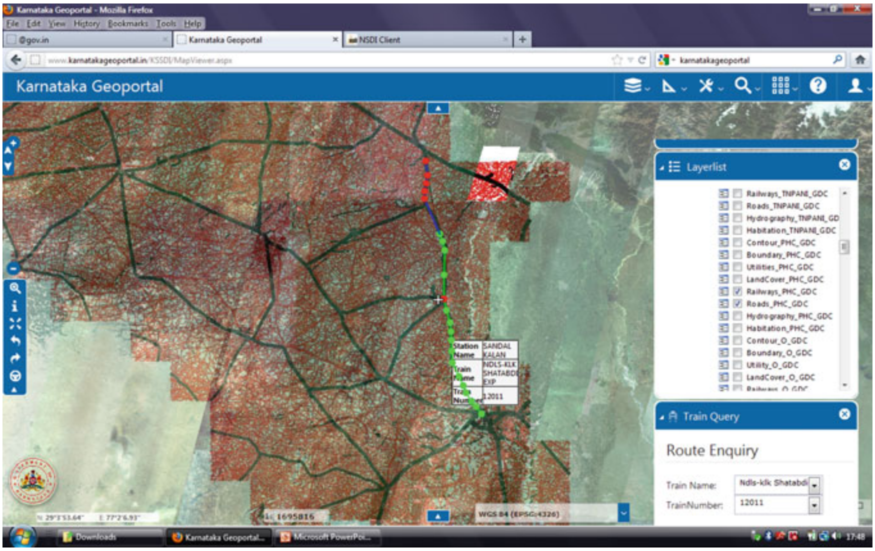


Fig. 2 Sample screenshot showing the train route and location on the SOI topographic sheet and NRSC's Bhuvan imagery coming from different sources

running information and route enquiry, the application shows the actual location of the selected train on the combined WMS display of topographic sheet of Survey of India and the satellite imagery of the National Remote Sensing Centre.

Implementation of the NDR prototype containing sample registers from SOI, GSI, FSI, the Indian Railways and demonstration of its utility in automated processing of spatial data for a selected use case have thus been successful. NDR has the potential to help overcome the issues relating to sound governance and utilisation of the spatial data sets by NSDI in India.

8 Recommendations

NDR is a useful concept and has the potential to overcome the problems of search and access and automated processing of spatial data sets from the NSDI agencies [8]. NDR platform should be operationalised by initially linking it up with a few (say 2–3) agencies to share their metadata and services and then gradually covering more agency nodes in an incremental manner in order that the implementation could be iteratively improved through the process of verification and validation. Considering that the ingredients of NDR and their utilities have been understood through demonstration of a proof-of-concept in the Indian context, professional agencies should be involved to operationally develop and implement the NDR. Harvesting should be a part of the NDR utility for automatic and periodic updation of metadata. In particular, focus of the implementation should be on development of domain-specific application schemas using ISO/OGC/BIS standards, existing data sets re-engineered and new data sets appropriately engineered using the standards-based feature model for ensuring interoperability; setting up spatial databases and application systems to provide WFS and WMS services; identifying use cases for provision of web processing services based on WPS and RESTful Service standards; setting up necessary system platform, hardware and software for operationalising registers and required processing services; demonstrating search on the NSDI partnering agency sources; and preparing a few case studies for demonstrating discovery and use of data and services from the NDR data nodes for training on integrating the web-based services with the end user's business workflows.

Acknowledgements The authors are grateful to Secretary, DST for his overall guidance and support in carrying out the work reported in this article. Data/image support and other inputs from IRCTC, SOI, GSI, NRSC, and several other organizations and experts in the implementation of the NDR prototype are thankfully acknowledged.

References

1. National Spatial Data Infrastructure (NSDI) Resolution 2006, Constitution of National Spatial Data Infrastructure (NSDI), 13 June 2006 published in the Gazette of India, July, 2006. https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/NSDIGazetteConstitution.pdf
2. Infrastructure for Spatial Information in Europe (INSPIRE) Knowledge Base. https://inspire.ec.europa.eu/about-inspire/563
3. Open Geospatial Consortium (OGC). http://www.opengeospatial.org/docs/is
4. OGC: CSW-ebRIM Registry Service - Part 1: ebRIM profile of CSW; OGC 07-110r4 Version: 1.0.1 (2008). http://www.opengeospatial.org/standards/cat
5. OGC: OpenGIS® Catalogue Services Specification; OGC 07-006r1 Version 2.0.2, Corrigendum 2 Release (2007). http://www.opengeospatial.org/standards/cat
6. ISO 19135-1: Draft International Standards on "Geographic information—Procedures for item registration—Part 1:Fundamentals"; ISO/TC 211 (2014). https://www.iso.org/committee/54904.html
7. Indian Standard (IS) 16439, 2016; Bureau of Indian Standards (BIS) on "Metadata Standard for Geospatial Information"; BIS/SC LITD 22; http://www.bis.gov.in
8. Department of Science & Technology (DST), Government of India; 2015; Report of the NSDI Technical Committee (TC) on National Data Registry (NDR) for the National Spatial Data Infrastructure (NSDI)

Development of State Spatial Data Infrastructure (SSDI): Indian Experience

D. Dutta and S. Pandey

1 Introduction

In its journey of planned development, the current emphasis in India is on local-level planning. The concept of spatial planning has been brought in to strengthen it. The constitutional support has come in the form of 73rd and 74th amendments, while Geographical Information and Communication Technology (Geo-ICT) has been adopted as the main tool for achieving it. The Government of India initiated a number of Geo-ICT-based programmes to support the local-level planning in 1980s, viz. Natural Resources Data Management System (NRDMS) of the Department of Science and Technology, National Natural Resources Management System (NNRMS) of the Department of Space (DOS) and Geographical Information System (GISNIC) and District Information System (DISNIC) of the National Informatics Centre (Ministry of Communication and Information Technology). These programmes had demonstrated the utility of spatial data in generating information required for efficient decision-making, as a result since early 90s there has been considerable awareness among local administrators about Geo-ICTs [1]. In order to facilitate and coordinate the availability and access of spatial data between stakeholders at all levels of government, National Spatial Data Infrastructure was established in 2006.

However, non-availability of and inaccessibility to spatial data of the desired resolution and currency, incompatible data sets and inadequate capacity among the officials of the local government departments (LGDs) remained the major bottlenecks to the utilization of National Spatial Data Infrastructure (NSDI) for local-level

D. Dutta (✉) · S. Pandey
Natural Resource Data Management System Division, Department of Science and Technology, Ministry of Science and Technology, Technology Bhawan, New Mehrauli Road, New Delhi 110016, India
e-mail: ddutta@nic.in

S. Pandey
e-mail: shubha.p@nic.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_4

planning. In order to obviate the same and create synergy between spatial infrastructure at national level and state level, decision was taken during the Chief Secretaries' Conference held on 4–5 February 2011 for the establishment of State Spatial Data Infrastructure (SSDI) by the Department of Science and Technology, Government of India. In October 2011, National Spatial Data Committee (NSDC), chaired by the Honourable Minister of Science & Technology—apex body for the formulation of Spatial Data Policy had approved the creation of SSDI.

Globally, the experience of development and operationalization of SSDI is very limited, especially in the developing world. In the developed world, there are some experiences in Australia, USA, Europe and Canada. In Australia, state governments play a significant role in policy development and the building and managing of spatial data infrastructures. Each of the Australian states and territories has established a coordination agency or group to reflect each of their mandates and state goals with respect to spatial data infrastructures. The state government agencies are active in pursuing SDI development with most activity generally focussed on delivering product outcomes. Each state has an overarching policy to facilitate SDI development which usually includes encouraging the active participation of the private sector in their SDI vision. In the USA, individual state coordination bodies have generally shouldered the majority of the SDI coordination responsibilities, e.g. Maryland, Illinois, California. In addition, the Federal Geographic Data Committee (FGDC) has encouraged the development of SDIs at sub-national levels through partnerships and collaboration [2]. Müller and Würriehausen [3] has reported a case study of local SDI implementation in the South Western German State of Rheinland-Pfalz it in the context of the evolution at higher SDI levels represented by the INSPIRE initiative at the European level and by the GDI-DE initiative at the German national level.

The Canadian Geospatial Data Infrastructure (CGDI) is developed by the Natural Resources Canada to provide open, secure and continually available access to comprehensive location-based information about Canada through the community-sustained CGDI in support of prosperity and well-being for all Canadians. In British Columbia, the provincial Spatial Data Infrastructure (SDI) is utilized by legislated and mandated lines of business such as resource development, water management, land tenure, transportation, public safety and emergency response [4].

The main objectives of SSDI in India are:

- To develop and operationalize a standard framework in states for the purpose of sharing and exchange of spatial data sets from various state-level line departments/agencies
- To utilize the spatial data sets from the above framework in GIS operations required for decision support by various stakeholder communities
- To create awareness and build the capacity of various stakeholders for carrying out various activities of SDIs.

2 Development of SSDI

In order to realize the potential of spatial data to be made available for planning processes at various levels, the Department of Science and Technology (DST) through its NSDI (National Spatial Data Infrastructure) initiatives has supported the establishment of State|Spatial Data Infrastructure in some states on competitive selection basis. However, this would require a synergizing effort between various entities/stakeholders involved towards creating a network of state data nodes for operationalization of State SDIs.

2.1 Key Components of SSDI

The key components required for successful operation of a SSDI are:

- People: Implementing agency, data and ICT infrastructure providers, policymakers and users.
- Processes: Data acquisition, data management, data availability and sharing.
- Technology: Hardware, software, geoportal services, and applications.
- Policy.

2.2 Resource Requirement

2.2.1 IT Infrastructure and Processes

The system architecture shown in Fig. 1 represents the hardware and software required in the development of a geoportal for SSDI. The geoportal has to be designed using Web-enabled GIS Software Supporting RDBMS (open source/proprietary) for delivering dynamic maps and related data and services via the Web. It should operate in a distributed environment that consists of both client-side and server-side components. Typically, the client will request information from an Internet server holding the data repository. Then, the server will process the request and send the information back to the client viewer. The server components, i.e., the Web server/application server and the data server, will form a part of the server architecture. Only authorized clients will have access to the server application and database. For a better performance, this architecture should also include two backup servers.

Open source/commercial off-the-shelf SDI software package interoperable to various OS platforms will be required to provide various Web services, viz. Catalogue Service on Web (CSW), Web Map Service (WMS), Web Feature Service (WFS/GML), Web Registry Service (WRS) and Web Coverage Service (WCS).

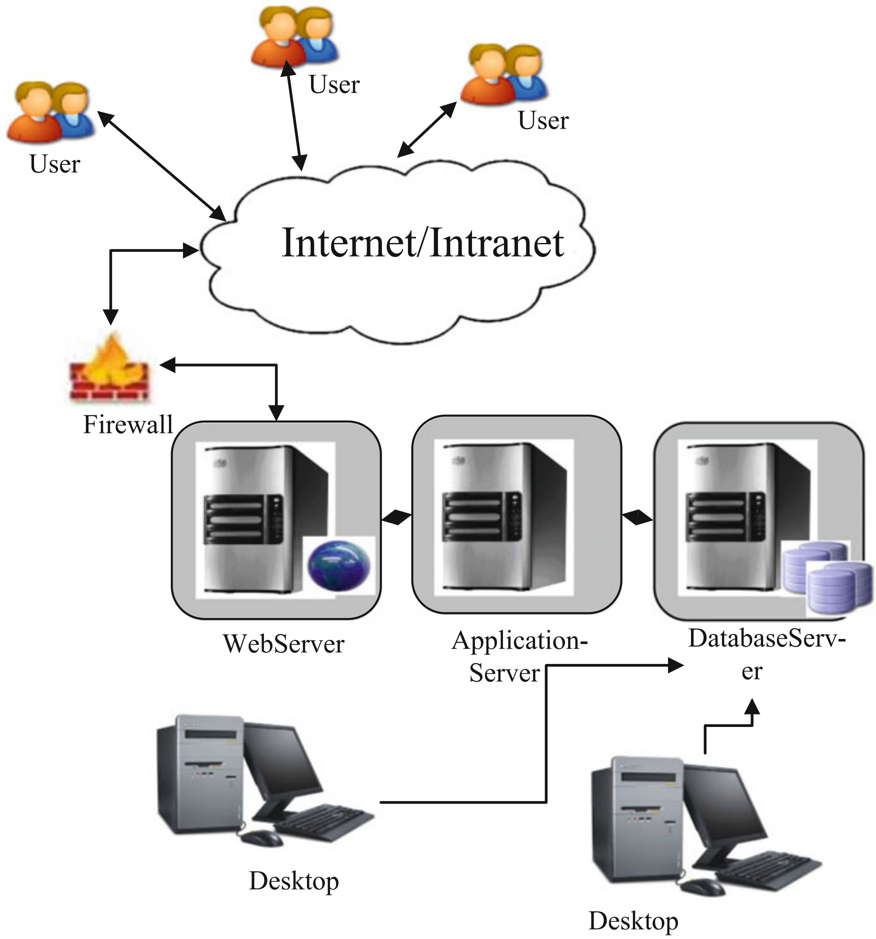


Fig. 1 Systematic representation of a common state's spatial database infrastructure (system architecture)

2.2.2 Skill Required

The indicative technical skill sets required for the development of a SSDI are GIS expertise; thematic expertise; IT expertise (website developers and managers, data/system administration software developers and programmers, security related, etc.); information risk management expertise; other expertise on need basis, etc.

2.3 Methodology

The following generic steps would be required to ensure the successful implementation of SSDIs (Fig. 2):

- **Selection of nodal agency**: The nodal agency for the implementation of SSDI is selected based upon their technical/organizational capacity in geospatial information management (GIM).
- **Selection of system integrator**: RFP document should be prepared by the nodal agency for the selection of the vendor for setting up the state geoportal/data clearing house as well as developing state, district and block data nodes.
- **Identification of stakeholders and need assessment**: Identifying the users/user groups already associated or to be potentially associated with the use of geospatial data in the context of developmental planning is an essential step. This could be done through a state-level workshop/meeting, etc. Interviews should be conducted with various line departments/agencies of the state government to assess their information needs.

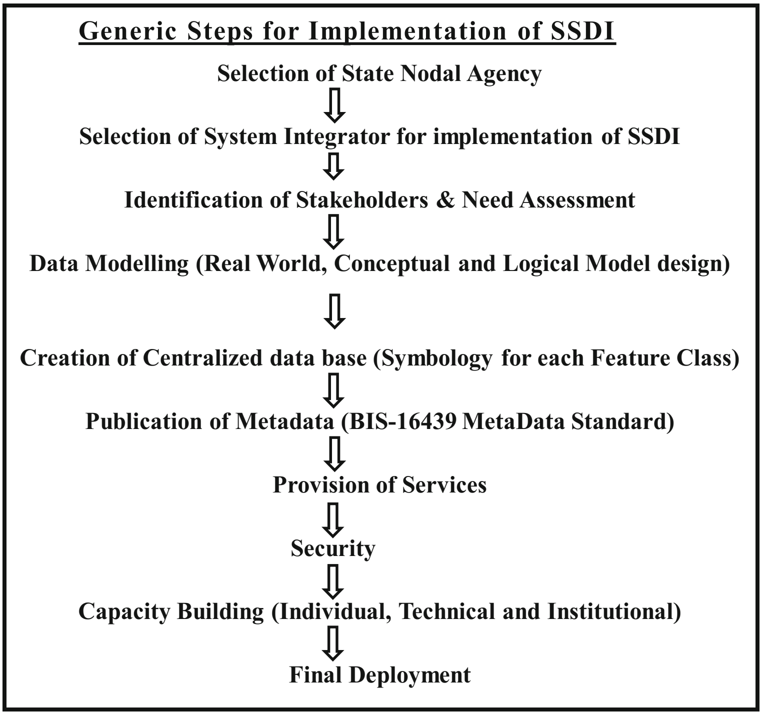


Fig. 2 Common steps for development of state geoportal/data clearing house

- **Data Modelling**
  - **Creation of real-world object catalogue**: Based on the need assessment, a detailed catalogue of underlying real-world objects clearly defining each object should be drawn up. A real-world object catalogue is useful in developing the data models (conceptual/physical) for the implementation.

Conceptual Data Model: A conceptual data model bringing out the objects and the associations is required to be developed and appropriately documented. The conceptual model should address software recommendations, refinement of physical system configuration (servers, network, storage requirements, etc.).

- **Creation of a centralized database**: Initially, a Web-enabled centralized database of various existing spatial, attribute or statistical data should be developed at the state node to store spatial, attribute or statistical data sets already available with the nodal agency and line departments. Proper indexing mechanism of data sets should be adopted to facilitate efficient query/access. Mechanism for updating the existing data sets on the database using OGC standards-based Web Feature Service (WFS) should be provided to facilitate upgradation/maintenance of the database from remote clients located in the individual line departments/agencies of the state government, GIS data centres/cells, viz. district NRDMS centres.
- **Publication of metadata**: In order to support easy discovery and access, various geospatial data providers in the state should publish their metadata sets into the geoportal/centralized database following metadata standards prescribed in BIS 16439.
- **Provision of services**: After following the specifications of OGC's geospatial portal reference architecture, a geoportal (a website for SSDI) should be developed, demonstrated and installed to provide portal services/portrayal services/data services, and catalogue services.
- **Security system to handle data, user identity and network**: In order to secure the geoportal/clearing house from unauthorized access, an appropriate security software module having provisions to safeguard data sets, manage user identity and transaction security, detect unauthorized intrusion, permit authorization/authentication, and non-repudiation-based access controls on resources is required.
- **Capacity building of stakeholders**: In order to build up individual, technical and organizational capacity, training workshops at the state and organizational levels should be organized.
- **Final deployment**: Based on the user feedback, the complete system should be integrated and operationalized for use. The complete system and its use should be presented in a series of workshops in different parts of the state having participants from groups like elected representatives, officials from different tiers of the governmental hierarchy from the state, district, tehsil, and village, other employees of different concerned organizations, and selected citizens.

Table 1 List of data layers required to make foundation of a SSDI

State/district data layers	State/district data layers
Agroclimatic zones	Drainage
Taluk/district boundary	Basin boundary
District headquarters	Catchment boundary
Division boundary	Subcatchment boundary
Division headquarters	Subdivision boundary
Geology	Subwatershed boundary
Meteorological station	Tanks
Parliamentary constituencies	National highway
Legislative assembly constituencies	State highway
Rail network	Major district road
Reservoir	Other district road
Rail gauge station	Village road
River	Airport

2.4 Data Layers

The data layers displayed in Table 1 should make the foundation of the SSDI. Additional thematic layers could be added depending on the availability of data on additional themes and future user needs.

2.5 Management Structure for Governance of SSDIs

The management arrangements may comprise of a three-tier organization structure consisting of the following committees:

- State Spatial Data Committee (apex body): A State Spatial Data Committee (apex body) should be constituted under the chairmanship of a senior official of the level of Chief/additional Chief Secretary or Secretary Planning with membership drawn from Secretaries of line departments concerned with spatial data in the state.
- SSDI Executive Committee: Committees similar to the Executive Committee may be constituted at the district level, headed by the district collectors with heads of various line departments in the districts to coordinate spatial data management.
- GIS cell: In order to coordinate among different line departments and office of the district collector/magistrate, a district GIS cell should be established at headquarters of each district. The cell should update its GIS data on daily, monthly and annual basis. Nodal officers should be identified in the line departments at state and district levels for coordinating with the GIS cell.

2.6 Enabling Policies

Geospatial policies governing spatial data sets and processes constitute an important component of Spatial Data Infrastructures (SDIs). Various policies in vogue regarding different aspects of data/information are as follows:

- **National Map Policy (2005)** defines the scope, distribution and liberalized access of digital Survey of India (SOI) topographic maps to user groups without jeopardizing national security.
- **Civil Aviation Requirement (CAR)** was issued in 2012 detailing the procedure for the issuance of flight clearances for agencies undertaking aerial photography, geophysical surveys, cloud seeding, etc.
- **Remote Sensing Data Policy (RSDP, 2001 & 2012)** defines the distribution process of satellite images to users.
- **National Data Sharing and Accessibility Policy (2012)** ensures proactive and open access to the data generated through public funds available with various departments/organizations of Government of India.
- There are also the Information Technology (IT) Act (2000); the Right to Information (RTI) Act (2005); and various State-Level Data Sharing & Accessibility Policies getting framed by various states to improve sharing and accessibility.

Despite the availability of the above policy framework, it has not been possible to discover and access geospatial data sets easily. Towards overcoming the above problems, a comprehensive National Geospatial Policy (NGP) is under formulations covering geospatial data, products, services and applications.

3 Best Practices in the Implementation of SSDI

Various state government agencies in states like Karnataka, West Bengal, Uttarakhand, Odisha, Jharkhand, J&K, Nagaland, Punjab and Haryana holding spatial data sets have collaborated with NSDI as partners for improving their data management procedures by developing GIS data assets (Table 2). Under various State Spatial Data Infrastructure (SSDI) collaborative projects implemented in cost-sharing mode between the DST Government of India and the concerned state governments, the agencies have either operationalized or demonstrated state geoportals/or their prototypes, worked on contract management for engaging system integrators, standardizing spatial data sets for the provision of Web-based data services in the framework of service-oriented architectures (SOA), and set up coordination mechanisms in the state governments. The best practices of various states in technological implementation, development of institutional mechanism and capacity building along with applications in governance are shared below.

Table 2 Representing status of spatial data infrastructure readiness in the various SSDI implementing states in the country

S. No.	States	Funding status	Geoportal status	Data sharing policy	Long-term sustenance mechanism	Institutional mechanism
1.	Karnataka	Rs. 253 lakh (Rs. 192 lakh/Rs. 61 lakhs)	Operational	Draft ready	Yes	Coordination structure
2.	West Bengal	Rs. 1178.98 lakhs (Rs. 477.38 lakhs/Rs. 701.60 lakhs)	Prototype	Draft ready	Yes	Coordination structure
3.	Jammu and Kashmir	Rs. 377.72 lakh (Rs. 264.40 lakh/Rs. 113.32 lakh)	Prototype	Yet to initiate	Initiated	Initiated
4.	Uttarakhand	Rs. 375.63 lakhs (Rs. 251.68 lakhs/Rs. 123.9 lakhs)	Operational	Draft ready for finalization	In the process	Well-established coordination structure
5.	Haryana	Rs. 203 lakhs (Rs. 125 lakh/Rs. 78 lakh)	Operational	Yet to initiate	Initiated	Process initiated
6.	Odisha	Rs. 650.02 lakh (Rs. 383.3 lakh/Rs. 266.72 lakh)	Operational	Odisha Spatial Data Policy in 2015	Yes	Coordination structure
7.	Jharkhand	Rs. 195.49 lakh (Rs. 181.35 lakh/Rs. 14.14 lakh	Operational	Draft ready	Yes	Coordination structure
8.	Nagaland	Rs. 184.95 lakh (Rs. 123.76 lakh/Rs. 61.19 lakh)	Data layers identified for upload	Initiated the process	Not yet	Not yet

3.1 Operationalization of the Geoportal: Karnataka Experience

The Karnataka State Council of Science and Technology (KSCST) with the support of state and central government have developed a geoportal (Fig. 3) and data clearing house to discover/access spatial data sets useful in the local-level planning. Karnataka Geoportal is a centralized single window access mechanism for all the spatial data acquired by various agencies. The council has institutionalized the geospatial technologies for informed decision-making by establishing district NRDMS centres in all the thirty districts of Karnataka with the support of Government of Karnataka. The available data is catalogued, and the search for the availability of any data set is made available through metadata.

The complete Karnataka Geoportal is developed using an OGC-compliant proprietary software with facilities to share the vector and raster data sets through WMS, WFS and WCS Web services. The geoportal has been kept operational and maintained by the council through funding support by the state.

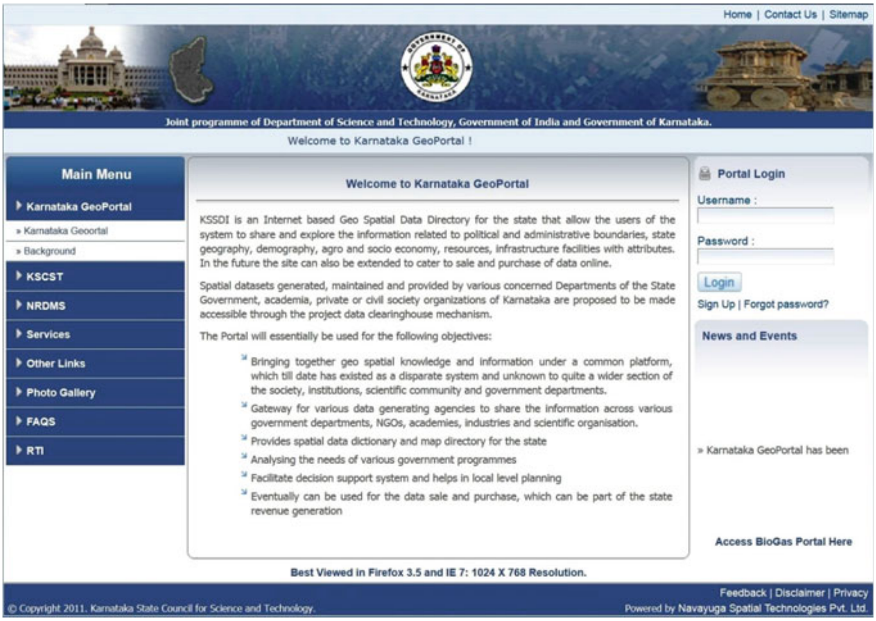


Fig. 3 Screenshot of the Karnataka Geoportal (http://www.karnatakageoportal.in)

3.2 Institutional Framework and Capacity Building

Apart from technologies, appropriate institutional framework and capacity building are essential for long-term sustainability of SSDIs.

Keeping this in view, many SDI implementing states such as Karnataka, Odisha, Madhya Pradesh Uttarakhand and Jharkhand have already published in the process of publishing State Spatial Data Policy to promote sharing and enable access to the government-owned data. For example, Government of Odisha has published Odisha Spatial Data Policy through its Gazette No. 1270, Dt. 29-08-2015, to facilitate easy access and sharing of government-owned data in open format for supporting sustainable and inclusive governance and effective planning, implementation and monitoring of developmental programmes, managing and mitigating disasters and scientific research aiding informed decisions, for public good.

Among SDI implementing states, Uttarakhand (UK) has shown a best example of capacity building and coordination mechanism. The Uttarakhand State has a coordinating structure in the state government with nodal officers identified at the state/district levels. For capacity building of the state government personnel and staff, various district-wise sensitization and training workshops have been organized where the primary focus has been on exposing them to application of Web-based geospatial technologies; providing hands-on training on GIS development and data capturing through GPS; information need assessment; and drawing up a road map for SDI development as per their own needs. The state government has also established a GIS cell at the headquarters of each district for coordinating the SSDI activities. The GIS cells at districts are integral to State SDIs to support geospatial data upgradation, publishing and sharing up-to-date data as WMS/WFS/GML, and utilization of GIS in various applications.

3.3 Applications in Governance

The utilization of geospatial data and services for a wide range of uses has seen steady growth in the requests for both data and services by planners and administrators. The NRDMS-NSDI programme of DST is continuously updating its data sets in diverse fields with an increasing set of application requirements. One of the successful examples is in watershed management.

Use of Geospatial Technologies for Watershed Management: The KSCST has showcased the utilization of geospatial technologies by developing geospatial applications on Karnataka Geoportal to evaluate watersheds in Belagavi Division of Karnataka. The council has integrated the KWDD's MIS with the Web-based SDI approach to provide standards-based online geospatial information services to end users in KWDD by utilizing the Karnataka State Geoportal already operational in the state. Earlier, KWDD does not have an easy and timely access to geospatial data of right resolution and currency for use in planning, implementation, monitoring, and evaluation processes. The geospatial services like multi-spatial-attribute queries; mobile compatibility; crowdsourcing; online editing of data and integration of MIS are the highlights of this development. Other portlets such as monitoring of biogas programme, health GIS, location of train services have also been demonstrated to the end users to support planning.

4 Future Directions

Rajabifard et al. [5] proposed that the future of SDI lies in addressing the balance between the competing foci in SDI development to encourage process-based facilitation for participant initiatives for spatial data sharing. While ongoing content development, a product-based approach, is essential to complement process-based facilitation, with the main aim to link existing and potential databases of the respective political/administrative levels of the community.

In order to maintain this balanced approach of Indian SSDI development, some of the considerations for the technological requirement are as follows:

New Techniques to handle Geospatial Data Streams
With the development of geo-sensor networks in various geo-domains (e.g. smart cities, agriculture, hydrology), human sensing, etc., geospatial streams are going to be one of the new modes of data. There are need to upgrade SDI to handle geostreams, interfaces to process and analyse these streams, development of domain-specific techniques/algorithms for events detection and monitoring and geospatial complex event processing (CEP) approaches.

Developing Approaches for Knowledge Representation of Geospatial Information
Geoinformation is highly diverse and needs approaches for reconciliation and integration at various syntactic, semantic levels. Geoinformation generated by various application projects needs to be converted to knowledge and represented formally. Thus, new ways of querying and reasoning on knowledge to obtain actionable intelligence are required. National-level geo-knowledge platform should be developed to provide various tools for knowledge representation, reasoning and visualization. There is need to develop controlled vocabularies, emergence of interdisciplinary programme on geo-ontology engineering.

Linked Geospatial Data
With the growth of multimedia, social networks, news, Twitter, user comments, etc., there is a need to develop approaches to link geospatial data with other forms of data available online to create various views of geospatial data in conjunction with other data. Thus, approaches need to be developed so that geospatial data conforms to the linked data concepts. For example, enabling RDF data available on the Web, and interconnect it with other data with the aim of increasing its value for user, enabling the development of geospatial linked data graphs, etc.

Spatial Data Fusion
New research directions in data fusion approaches should be explored to develop fusion models of multi-sensor, multi-source, multi-temporal geospatial data, which are also scale- and rotation-invariant. India specific fusion products should be identified, and standardized work flows for geospatial data fusion for various applications including near real-time applications should be published. This should be based on the type of product required for that application.

Big Data Analytics and Geo-Computational Systems
With the emergence of geospatial big data processing approaches, there is a need of development of geo-computational platforms integrated with SDI big data analytics techniques that are domain-specific.

Geospatial Approaches for Critical Infrastructures Protection (CIP)
Critical infrastructures are under constant threat from various man-made and natural disasters and other subversive activities. Understanding how resilient these infrastructures are after an event is critical. Also, there is a need to understand the interdependencies between the critical infrastructures. Thus, research is needed on deterministic and stochastic approaches for interdependency modelling, development of simulation environments, situational awareness, visualization.

3D Visualization
There will be an increase in the ways we visualize geospatial data, moving from a mapping environment view to a view that demonstrates how the user would experience a location as if their feet were on the ground, e.g. street views in Google Earth. Three-dimensional building models will continue to grow and help to create realistic views of our landscape. Geospatial data is critical in helping to supplement 3D views of other geo-coded information to provide 3D business intelligence. Three-dimensional visualization will be fed by new semi-structured data from mobile devices, geo-tagging of existing enterprise structured data, and tapping into new streams of location-aware unstructured data.

Location-based Business Intelligence
Solving business problems with geospatial data and cloud computing will become increasingly important in future. More and more geospatial solutions and 3D business intelligence services that provide customers with “answers” rather than “data” need to be developed. Increasing complexity of geospatial solutions, and the continued growth in the geospatial technology industry, will lead to both business consolidation as well as increasing partnerships.

Crowdsourcing
In an increasingly connected world, volunteered information will become more and more important. Information created in this way will be increasingly important for government and business and may radically challenge traditional approaches to geospatial data generation and maintenance.

References

1. Sahay, S., Welsham, G.: The use of geographical information systems in developing countries: social and management issues. In: Emerging technology Series. International Center for Science and High Technology and United Nations Industrial Development Organization, Vienna (1997)
2. McDougall, K.: A local-state government spatial data sharing partnership model to facilitate SDI development. Ph.D. Thesis: University of Melbourne (2006)
3. Hartmut Müller, H., Würriehausen, F.: A case study on local SDI implementation in Germany. In: Joint FIG Commission 3 and 7 Workshop and Commission 3 Annual Meeting Information and Land Management. A Decade after the Millennium, Sofia, Bulgaria (2010)
4. https://www2.gov.bc.ca
5. Rajabifard, A., Feeney, M., Williamson, I.P.: Future directions for the development of spatial data infrastructure. J. Int. Inst. Aerosp. Surv. Earth Sci. (ITC), The Netherlands (2001)

Web-Based Geospatial Technologies to Support Decision-Making

H. Hemanth Kumar

1 Introduction

Conceptual changes in the practice of planning were brought in, around late 70s, by adopting the decentralized or local-level planning to ensure that the development is sustainable and area-specific and takes into account the felt needs of the local people. The Constitutional (73rd and 74th Amendments, 1992 and 1993) amendments empowered the State governments to form the institutions of local self-governance i.e., rural local bodies (RLBs/Panchayats) and municipalities (ULBs) in rural and urban areas, respectively. KSCST, recognizing the need for both spatial and non-spatial data to local-level planning across Karnataka, established Karnataka natural resources data management system (NRDMS) programme in 1992 to develop a comprehensive spatial data management system for easy access of data and information to support RLBs/ULBs. The 73rd and 74th Amendments to the Constitution also emphasized the need for geospatial data for development planning to local-level governments or panchayats.

Karnataka NRDMS programme is a multi-disciplinary and multi-institutional programme aimed at developing methodologies for building and promoting the use of spatial data management and analysis in local area planning. The vision of the NRDMS programme during its inception was to provide S&T inputs for operationalizing the concept of decentralized planning of the country by developing computer compatible spatial databases on natural resources, socio-and agro-economic parameters to further the concept of area-specific decentralized planning.

The council established district NRDMS centres in each district over a period and established all the centres by 2008 with the funding support of both the Central and State governments. The district NRDMS centres now are being solely supported by the State government, and Karnataka government is the first State in India to

H. Hemanth Kumar (✉)
KSCST, IISc, Bangalore 560012, India
e-mail: hemanth@kscst.iisc.ernet.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_5

institutionalize spatial data for local-level planning. These centres are tasked with empowering local communities to make informed decisions in local-level planning initiatives using geospatial technologies.

The utilization of geospatial data and services for a wide range of uses has seen steady growth in the requests for both data and services by planners and administrators. The NRDMS programme is continuously updating its datasets in diverse fields with an increasing set of application requirements. Application domains include, for example, public health and education, environmental analysis and mapping, transportation, water quality/quantity, watershed management, elections, support to disaster management, administration and planning. NRDMS centres provide value-added information, spatial decision support systems, training and support to planners and administrators in local governments [3].

1.1 NRDMS Methodology

NRDMS programme aims to analyze the evolution of methodologies and techniques for formulating development strategies. In a scenario of large diversity of data sets, data users and data generating agencies, the programme aims at developing and demonstrating the use of spatial decision support tools for integrated planning and management of resources at the local level. The NRDMS programme uses both maps and textual data collected from various data generating agencies and remote sensing sources. Maps are digitized and stored on a computer as a series of thematic data layers. Depending on the real unit of planning, maps on 1: 250,000, 1: 50,000 and cadastral scales are utilized. Data on natural resources, demography, agro-and socio-economy and infrastructural facilities for a district form the core database. Those pertaining to specific problems of an area constitute the sectoral database. MIS data obtained from secondary sources are normally stored with village as a unit. Data gaps are filled in by limited primary surveys. GIS based decision support systems/modules help process the data and generate alternative scenarios for making optimal planning decisions. Outputs in the form of thematic maps, tables and reports based on spatial analysis are obtained from the database as per the user requirements (see Fig. 1) [3].

1.2 Major Activities/Achievements [3]

- NRDMS centres have been providing custom applications, value-added information, training and support to planners and administrators.
- Developed a standardized seamless spatial data for storing and sharing through Karnataka Geoportal to facilitate data access and discovery on Survey of India framework data.
- Seamless administrative and derived boundaries geospatial information for entire state of Karnataka is available, and data has been updated as of 2017.

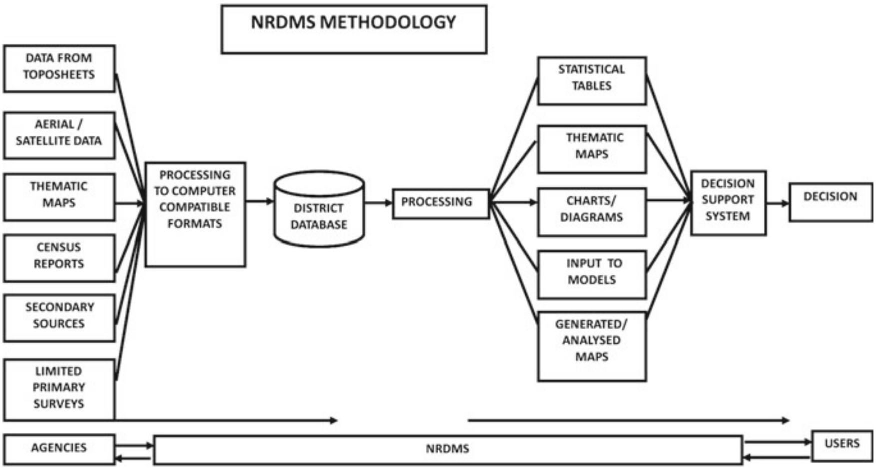


Fig. 1 NRDMS methodology

- Development of Web-based geospatial applications/decision support system to support informed decision-making.
- Initiated mapping of urban local bodies as well as resource mapping for Panchayati Raj institutions using latest geospatial technologies.
- NRDMS centres have been continuously updating information (spatial and non-spatial) for the benefit of user departments.
- Conducted district level workshops in all the districts to create awareness among end-users.

1.3 Typical Applications Developed Under NRDMS Programme [3]

- Election information and management system.
- Rural Health Facilities-Allocation/Location and its jurisdiction.
- Impact of stone crushers on environment.
- Crime mapping.
- Education Management Information system.
- Rural water supply and water quality for PRIs.
- Resource, Infrastructure and Habitation mapping for Gram Panchayats.
- Prioritization of MI tanks (see Appendix Fig. 6).
- Watershed and related applications [6].
- Digital Tourism ATLAS.
- Gram Panchayat ATLAS—2016 (see Appendix Fig. 7) [1].
- Planning ATLAS of Karnataka.

2 Development of Karnataka Geoportal

With increasing dependence on spatial datasets and extensive use of Web-based technologies in other domains, a need has been felt to allow the discovery of and access to up-to-date spatial datasets. The council with the support of State and Central government developed a geoportal and data clearing house to discover/access spatial datasets useful in the local-level planning process through Karnataka State Spatial Data Infrastructure (KSSDI) Project [7]. KSSDI also called as Karnataka Geoportal is a centralized single window access mechanism for all the spatial data acquired by various agencies in the Government of Karnataka using State's resources primarily to support planning activities of the Panchayati Raj institutions, district/state line departments, urban local bodies and civil society organizations in the State.

Table 1 Predefine applications under Karnataka Geoportal

Department	Application requirement	Application query	Data required
Department of Education (Akshaya Yojana Programme)	Opening of new primary school	Distance of a school from a village doesn't exceed 1.5 km	Scale 1:50 K: village point, school point data Scale <1:25 K: village polygon, school polygon/point data
(District Primary Education Programme)	Opening of new primary school	1. Population exceed 300 2. Village with Adult Education Centre	Village level population/census data with: Village Name, Village Panchayat Population, Existing school, Existing Adult Education centre
(Akshaya Yojana Programme)	Construction of new school rooms	Select schools for which additional rooms were constructed Population is in excess of 600 number of students in excess	Village name Village population School name Number of students Number of dropouts Current year Number of School Rooms Number of rooms added last 3 year
(Akshaya Yojana Programme)	Repair to school rooms	Select schools for which rooms were repaired	Village census/settlement data Rooms repaired attribute information Date of repair information

Spatial datasets generated, maintained and provided by various concerned line departments of the State government, academia, private or civil society organizations of Karnataka have been made accessible through the portal. The available data are catalogued and the search for availability of any dataset is made available through metadata. The complete Karnataka Geoportal is developed using an OGC compliant proprietary software with facilities to share the vector and raster datasets through WMS, WFS and WCS Web services along with the solutions and products from OEM for server hardware, storage, backup and retrieval systems, networking components and for RDBMS requirements. The portal was developed with the support of a turn-key solution provider through open tendering process [4].

2.1 Need Assessment of Karnataka Geoportal

The objectives of the need assessment were to determine how Karnataka Geoportal can leverage geospatial technologies, and what resources are required to support access/discovery of data and applications in terms of hardware, software, data and personnel. The need assessment of Web geoportal includes logical overview, data description, scope, potential applications, functional/non-functional requirements and map services. The need assessment provides the basis on which the conceptual system design is structured. The design is based on the information compiled in the needs assessment phase of this project and addresses mainly the functionalities of the portal [4].

KSCST organized several workshops at district/division level to assess the geospatial requirements of the user, and with the support and knowledge of district NRDMS centres, more than 500 GIS applications (typical query and analysis) were prepared. The applications are categorized based on the department scheme/programme or a search keyword. A sample list of predefined applications and application requirements is shown in Tables 1, 2, Figs. 2 and 3 and are available as part of geoportal modules.

2.2 Design of Real-World Object Catalogue

A layer or a table dataset is defined and expressed as a real-world object bringing out the feature properties of each and the attributes. Defining the layer/dataset in terms of a real-world object is defined as following (Table 3) [3].

- Real-world object—properties or feature properties
- Physical data model

The objective behind designing the real-world object catalogue is to:

Table 2 Sample need assessment (UML shown in Fig. 2.)

Name of the application	Seasonal depth to water level fluctuation
Name of the department	Mines and Geology department
Brief about the application	The water level fluctuation is mapped for assessment of drought and ground water recharging using interpolation method.
Application query	To estimate seasonal fluctuation of water
Data required	Administrative boundaries, observation wells (depth to water level of pre-and post-monsoon)
Description	Administrative boundaries: district boundary, village boundary DTW: fluctuation level range –0.5–5 mbgl GPS: observation well points
Conceptual design of the application	1. Display district and village boundary. 2. Calculate seasonal water level fluctuation from post-and pre-monsoon depth to water level by finding the difference. 3. Interpolation of seasonal fluctuation level values using inverse distance weighting method. 4. Final map shows distribution of interpolated seasonal level fluctuation with ranges of variation

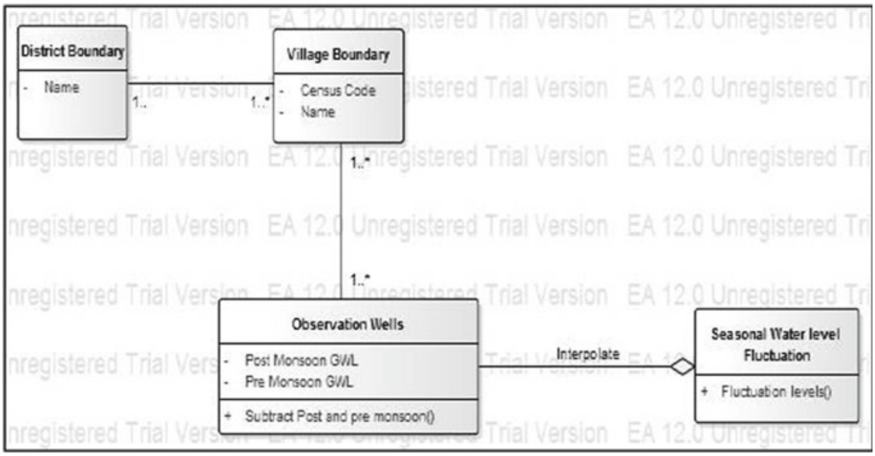


Fig. 2 UML class diagram model showing interpolation of seasonal water level fluctuation

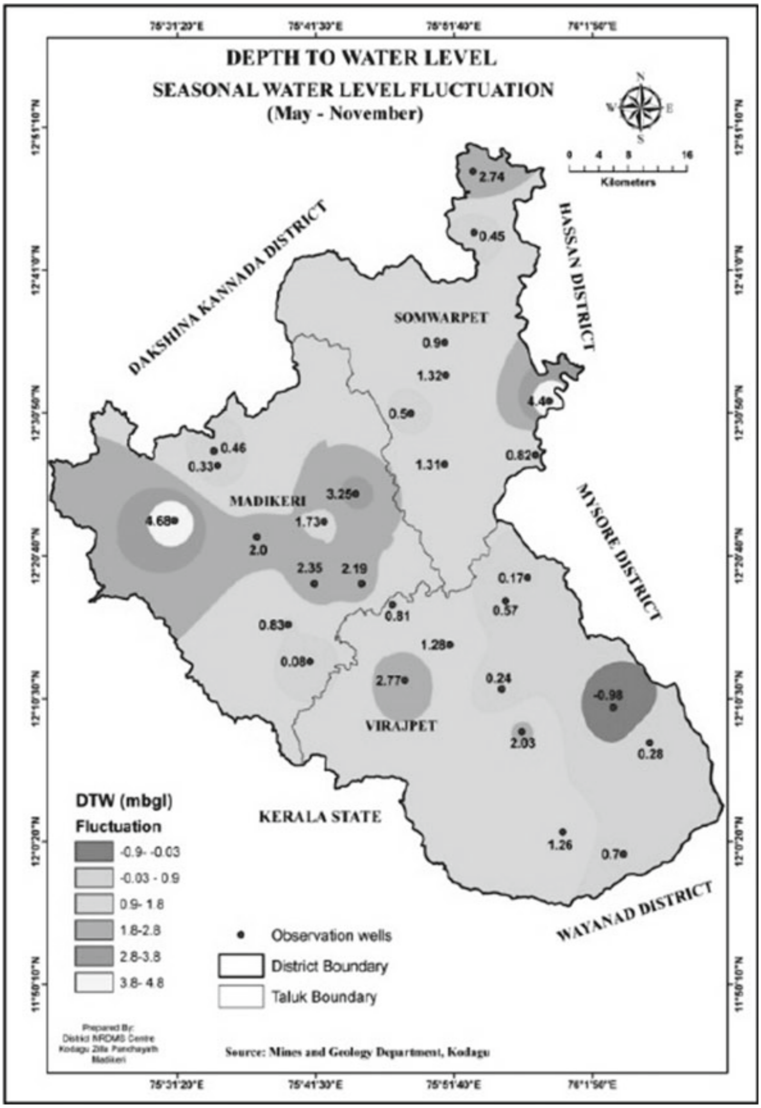


Fig. 3 Seasonal water level fluctuation map of Coorg District, Karnataka as represented by UML diagram in Fig. 2

- Listing of available real-world objects.
- Defining real-world objects and identifying their sources.
- Listing of features and attributes of each real-world object.
- Describing of each of the features or attributes identified.
- Comprehensive data dictionary of the real-world objects.

Table 3 Real-world object catalogues

Variables	Description
Layer code	A unique code given to a layer for its identification
Catalogue ID	Identification code given to the layer to link to the data dependencies at a user use case level
Layer title	Title of the layer
Layer Alias	Alias of the layer, is the most common name of identification which will label the layer while display
Category	Category identifier for the layer classification
Layer type	Geometry feature type of the layer, whether Polygon, Line or a Point
Source	Name of the data generating agency
Scale	Scale of the layer reflecting the feature details that are captured
Layer definition	A definition to the layer describing the layer
Definition source	Source of the layer

- Real-world objects classification by feature type.

2.2.1 Real-World Object Properties: The Real-World Objects Are Defined by the Following Properties

The procedure adopted for developing Karnataka Geoportal: is given below.

- Requirement analysis
- Real-world object catalogue
- Conceptual data model
- Procurement of hardware/software
- Centralized database
- Publishing metadata
- Creation of geoportal
- Customization of available software
- Testing of the geoportal/database/services
- User acceptance testing
- Final deployment of the complete system

2.3 Karnataka Geoportal System Architecture

The diagram (see Fig. 4) shows a schematic representation of the infrastructure and network requirements to host the geoportal and the clearing house. The diagram also shows required number of servers, and server types, and how these servers will communicate with existing infrastructure and partner systems.

The Karnataka geoportal has been developed using OGC compliant proprietary software to share the vector and raster datasets through WMS, WFS and WCS Web services along with the solutions. Products from OEM's have been procured for server hardware, storage, backup and retrieval systems, networking components and for RDBMS requirements.

The KSSDI portal passed preliminary stage of development and was released to users during October 2009 for portal functionality testing. The release allowed the testing of the portal to undergo usability testing with selected users who provided feedback to the project team. Identified malfunctions were rectified. The council officially launched Karnataka Geoportal during December 2009 at the ninth annual event of National spatial data infrastructure [2, 4].

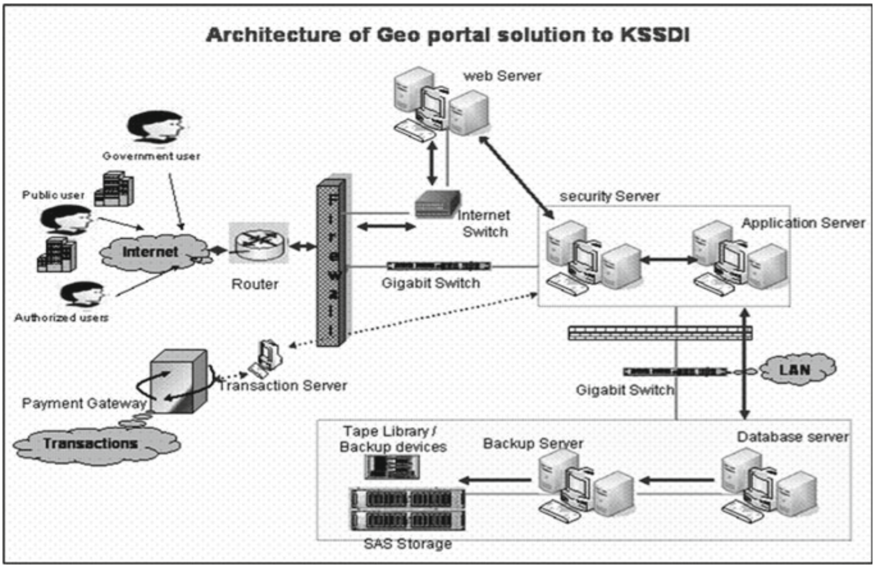


Fig. 4 Karnataka geoportal architecture

2.4 Services Under Karnataka Geoportal

Karnataka Geoportal envisaged as a centralized single window access mechanism for all spatial data held and acquired by various agencies/line departments of the state. The modules in the Karnataka Geoportal are categorized into [5]:

- Map viewer—Web map service (WMS)
- Product catalogue/metadata—catalogue service on Web (CS-W)
- Services specific service/feature data sets—Web feature service (WFS)
- Simple applications (query-based decision support)
- Coverage services/images—Web coverage service (WCS)
- Biogas portlet
- Help/support

On Karnataka Geoportal, few portlets were developed to support Ministry of New and Renewal energy, and Ministry of Human Resource Development supported Government of India to monitor the implementation of Biogas programme and to check the availability of schools in rural areas, respectively.

User acceptance testing (UAT): UAT is the last phase of the portal/software testing process. During UAT, actual portal was tested to make sure it can handle required tasks in real-world scenarios, according to specifications with the support of state and district NRDMS centre staff, experts from IIT—Bombay and IISc.

Following are the steps involved in in-house UAT:

- Planning—the UAT strategy was adopted by the advice of experts from IIT —Bombay.
- Designing test cases—test cases are designed to cover all the functional scenarios of the software in real-world usage. The access and discovery of data was an important bench marking for testing the portal. They were designed in a simple language and manner to make the test process easier for the testers.
- Selection of testing team—the testing team comprised of real-world end-users.
- Executing test cases and documenting—majority of the test cases were documented and tested randomly for bugs.
- Bug fixing—responding to the bugs found by the in-house testing team, the vendor made few adjustments to the code to make the portal relatively bug-free.

The geoportal has a centralized geospatial database on variety of spatial layers available to users for query and analysis. A real-world object catalogue was developed to indicate all the datasets with its data dictionary. However, the primary objective of providing geospatial services through Web geoportal remained elusive. Based on the learnings, the requirements for showcasing the utilization of geospatial technologies by user were defined and formed the basis for initiating the second phase of portal development.

3 State Geoportal for Watershed Management

Karnataka Watershed Development Department (KWDD)—an agency under Government of Karnataka is in-charge of running various watershed schemes/projects, formulating policies, planning, organizing, coordinating the activities and implementing the watershed programmes. KWDD uses datasets for planning, implementation, monitoring and evaluation of watershed programmes. Pradhan Mantri Krishi Sinchai Yojana (PMKSY) programme (earlier known as Integrated Watershed Management Programme—IWMP) is a flagship programme of KWDD with a financial outlay of above Rs. 400 Crores per annum. KWDD has multi-disciplinary technical experts drawn from agriculture, horticulture and forest departments who are pooling their experience, expertise and technology in the watershed programmes being financed under External, Central, State and District Sector schemes. The department currently has an in-house online management information system (MIS) for monitoring watershed programmes by its Head Office. It makes use of existing cadastral and thematic maps available with other departments, from planning to implementation stages, and tries to document all the activities with geo-tagged time-stamped photographs before, during and after the implementation of various programme-based activities. However, they do not have an easy and timely access to geospatial data of right resolution and currency for use in planning, implementation, monitoring and evaluation processes. KSCST felt the need for integrating KWDD's MIS with the Web-based SDI approach of NRDMS-NSDI/DST to provide standards-based online geospatial information services to end-users in KWDD by utilizing the Karnataka State geoportal already operational in the State. Towards achieving the above goal, the project "Development of Geospatial Web Applications on Karnataka Geoportal for G-governance" was launched by KSCST under the collaborative support of DST, Govt. of India and Govt. of Karnataka during April 2013 to March 2016 [2, 6].

3.1 Features of the Portal

On the launch of the above project, a user requirement analysis was taken up to understand and study the geospatial information needs of the KWDD to be accessed over the Web. KSCST showcased the utilization of geospatial technologies by developing geospatial applications on the Karnataka State geoportal to evaluate watersheds in Belagavi Division of Karnataka under the Phase I and Phase II projects of PMKSY. Development and deployment of the "State Geoportal for Watershed Management" technology enables the users to query on various geospatial and non-spatial (attributes) parameters at cadastral/beneficiary level for activities taken up under agriculture/horticulture/forestry sectors. The geoportal enlists a host of open geospatial consortium (OGC)/international standardization organization (ISO) Standards-based Web services for accessing datasets to support watershed management.

Crowd sourcing of data for data updating, mobile (smart phone) compatibility and online editing of datasets are some of the technology's major highlights [2].

3.2 Additional Features of Watershed Portal Information Services for Watershed Management Include the Following [2]

- Web processing service (WPS)
- Transactional Web feature service (WFS-T)
- Mobile mapping
- Multi-spatial and multi-attribute query and analysis (see Fig. 5)
- Multi-lingual support
- Online help/support to end user

Accessible on the State geoportal, above Web-based services could be used by the end-users in the field (watersheds) to capture relevant datasets from the watersheds and update the geospatial database at a central server over the Web. In this instance, the database is located at KSCST for access by all end-users across the Belagavi Division of Karnataka contains the cadastral map layers secured from the Survey Settlement and Land Records (SSLR) Department of the State's Revenue Department. Non-spatial information from the MIS of the Karnataka Watershed

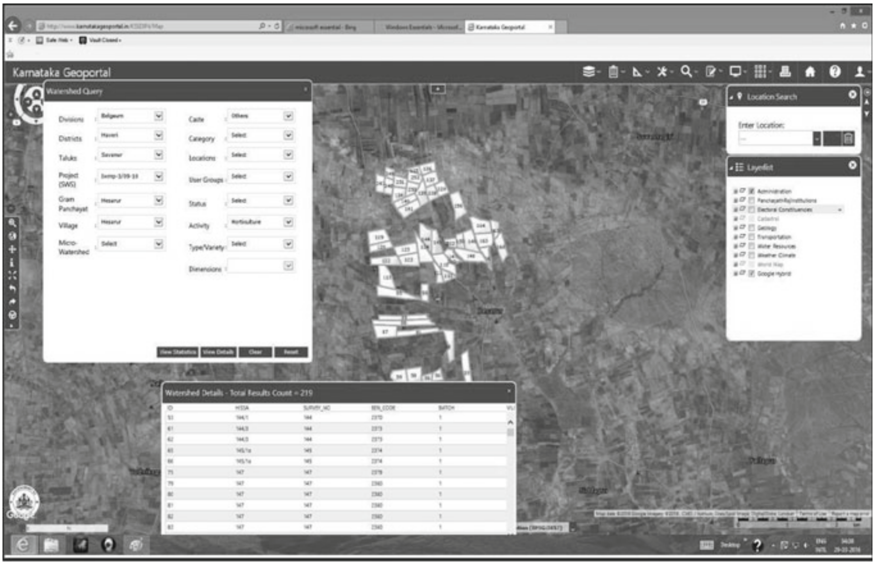


Fig. 5 Locating cadastral plots showing horticulture scheme implementation under IWMP

Development Department (KWDD) has been integrated with the cadastral (spatial) data layers to facilitate querying and analysis over the hitherto unavailable on integrated database and the Geoweb. Use of the services has been demonstrated to the potential end-users in various training workshops in Karnataka and several other States for adoption in State SDI projects of NRDMS-NSDI. The Karnataka District NRDMS Centres in the State have been trained on the use of the State Geoportal in watershed Management.

This portlet enables users to query and analyse on various geospatial and non-spatial (attributes) parameters at cadastral/beneficiary level for activities taken up under agriculture/horticulture/forestry sectors by integrating with the MIS developed by the Karnataka Watershed Development Department. This facility hitherto was unavailable on integrated database and the Geoweb. The geospatial services like multi-spatial-attribute queries, mobile compatibility, crowd sourcing, online editing of data and integration of MIS are the highlights of the programme [2].

4 Conclusions and Future Perspectives

The geoportal provides tools to the end-users to monitor the progress/implementation of projects/programmes of the government and allows adopting possible mid-course corrections through better and improved management practices through geospatial technologies. This, in turn, will contribute to the economy by saving governmental staff-time, improving quality of decision-making, ensuring better transparency and accountability, and bringing in better delivery of outputs of government schemes and programmes. We have envisaged and proposed following activities to enable Karnataka geoportal to fully support the informed decision-making.

- Creation of spatial data 1:10 k or below by adopting technical standards i.e., category, content, metadata, accuracy, services, etc.
- Mapping of all public/private assets (POI)
- Integration of MIS to GIS content near real time
- Identification of core application for each department
- Building of SDSS/applications
- Very good data connectivity including hardware and software
- Capacity building and spatial data policy for usage
- Smart city Applications (2D/3D)

Appendix

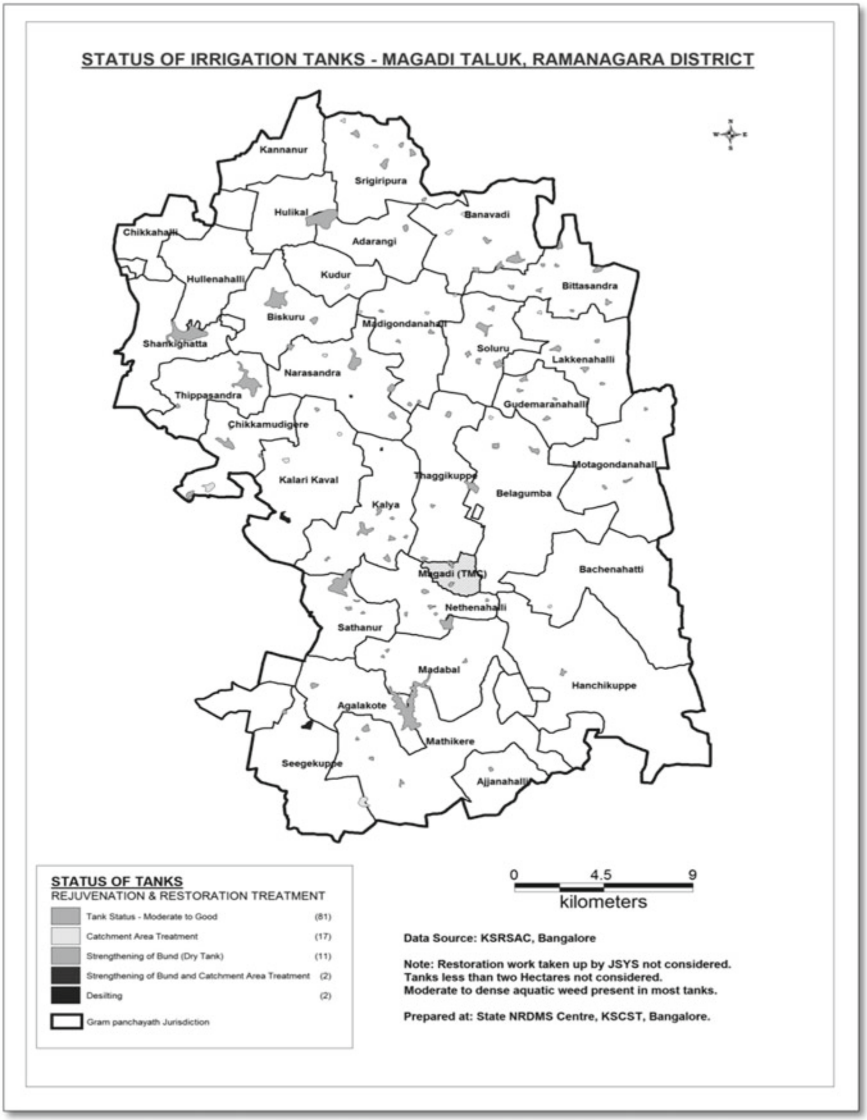


Fig. 6 Status of minor irrigation tanks of Magadi Taluk, Ramanagara District

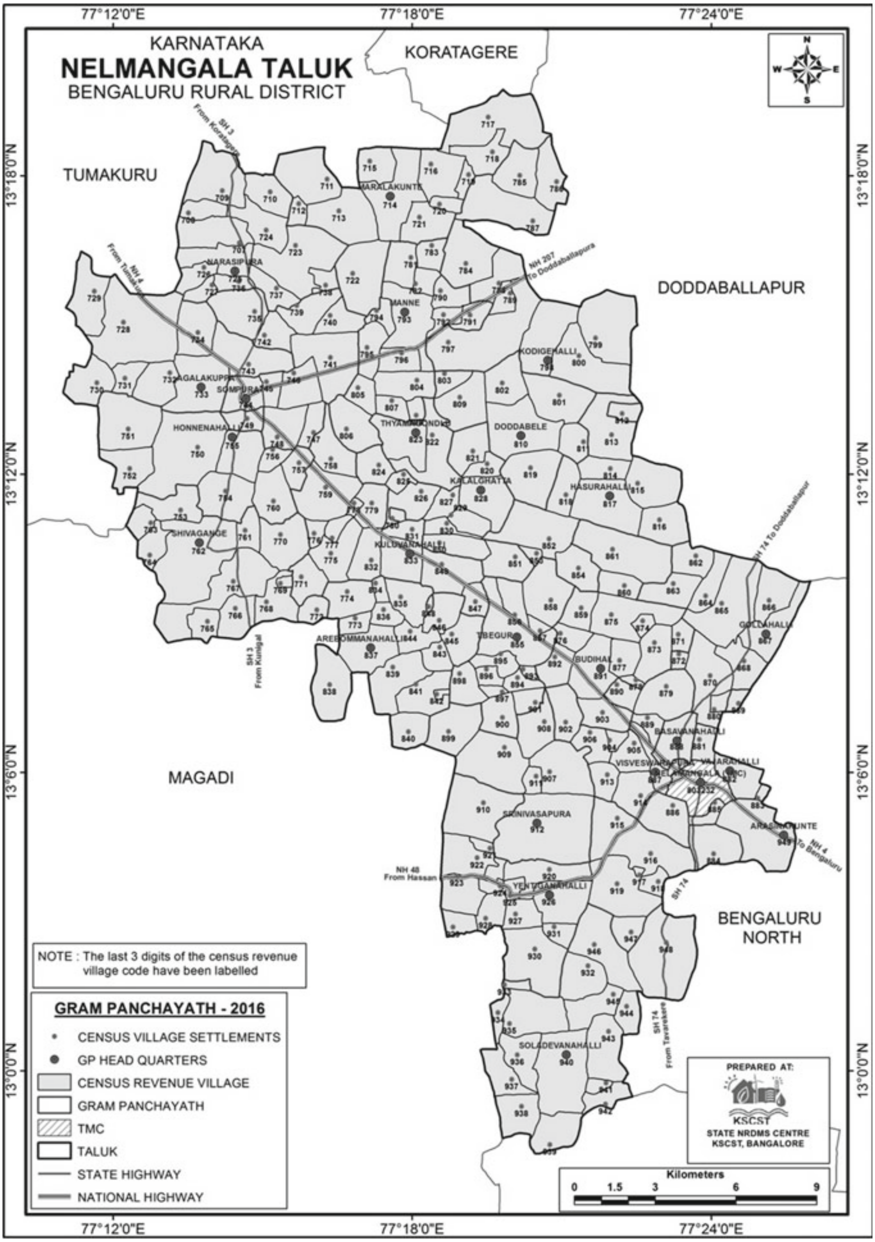


Fig. 7 Gram Panchayat map of Nelamangala taluk, Bengaluru rural district [1]

References

1. Hemanth Kumar, H.: Gram Panchayath ATLAS pp. 1–484 (2016). http://www.kscst.iisc.ernet.in/nrdms_files/reports/Gram_Panchayat_Atlas_2016.pdf. Accessed 20 Jan 2017
2. Hemanth Kumar, H.: A report on Development of Geospatial web based applications on Karnataka Geoportal for G-Governance (KSSDI– Phase II) pp. 1–72 (2017). http://www.kscst.iisc.ernet.in/kssdi/pdf/KSSDI_PhaseII_Completion_Report_9Aug2017. Accessed 5 May 2017
3. Hemanth Kumar H.: Technical Report on Geospatial Technologies for Informed Decision-Making—(Karnataka NRDMS Program) pp. 1–163 (2014). http://www.kscst.iisc.ernet.in/nrdms_files/reports/nrdms_technical_report_sep2014.pdf. Accessed 15 Feb 2017
4. Hemanth Kumar, H., Sondur, S.N., Vijay, U.T.: Need Assessment Report pp. 1–84 (2009). http://www.kscst.iisc.ernet.in/kssdi/KSSDI_PhaseI_Need_Assessment_Report.pdf. Accessed 15 Feb 2017
5. Hemanth Kumar, H., Sondur, S.N., Vijay, U.T.: A report on Karnataka Spatial Data Infrastructure—Phase I, pp. 1 -15 (2011). http://www.kscst.iisc.ernet.in/nrdms_files/reports/KSSDI_PhaseI_Completion_Report.pdf. Accessed 15 Feb 2017
6. Hemanth Kumar, H., Vijay, U.T.: Final Evaluation Reports IWMP Batch I—Projects pp. 1–168 (2017) (Unpublished)
7. Karnataka Geoportal (2009). http://www.karnatakageoportal.in. Accessed 20 Feb 2017

Uttarakhand State Geo-Portal for Decentralized Governance

J. S. Rawat and R. Dobhal

1 Introduction

In order to fulfil the objectives of Digital India Initiative of the Government, the present "e-governance" system is gradually being transformed into "g-governance" to improve transparency and decentralize the process of governance to lower area levels. In the recent years, there has been an increased need for more timely access to maps to facilitate effective planning and administration. In e-governance, we are getting information quickly through computers in the form of tables, text and analogue maps but in the g-governance system, one can access quickly digital or geo-stamped map showing geographic location and information of the targeted area/place for decision support. Thus, g-governance makes use of concepts and principles from the domain of Geographic Information Science [1] by utilizing IT, MIS and Communication Technologies. For developing the information system for g-governance, an enormous amount of geo-stamped data gathering, compilation, analysis and manipulation through the geospatial technologies is needed. The next decade is likely to witness complete switch over to g-governance in India with all the office records and maps accessible to planners and decision-makers in digital form. Under the Natural Resources Data Management System (NRDMS) Programme of the Department of Science and Technology, Government of India (DST, GOI), efforts are being made to promote geospatial technologies in the country by establishing about 72 district-level GIS Centres spread in different States of the country towards establishing State Spatial Data Infrastructures (SDI) for supporting implementation of the g-governance concept. Uttarakhand is one of the few states where all the districts

J. S. Rawat (✉)
Centre of Excellence for NRDMS in Uttarakhand, Department of Geography,
Kumaun University, SSJ Campus, Almora 263601, India
e-mail: jsrawat1955@gmail.com

R. Dobhal
Uttarakhand State Council of Science and Technology, Dehradun, India

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_6

have been equipped with district-level GIS Cells to support data management by conducting GIS training programmes for the staff and personnel of the line departments. One District GIS Cell initially set up in 2009 at the beginning of the DST Initiative in Almora district of the state and the co-located Centre for NRDMS established at the Department of Geography, Kumaun University, Almora have significantly contributed to this process of change in state-level geospatial data management towards the implementation of g-governance.

2 Background

2.1 District GIS Cell at Almora

In an earlier pilot phase of the related activities, through a research project funded by the DST GOI, a District Digital Data Base Centre Almora was established at the Department of Geography in the Kumaun University in 2002. At that time the Centre was the 54th district database centre of the country. Through this project the GIS infrastructure of Almora District was developed in association with the district line departments as per their respective geographic information needs. During this second phase (2006–2008) the GIS work of district Almora was updated and upgraded in web-accessible mode and developed and demonstrated as a template of a District Geo-Portal that could be successfully linked to the national level India Geo-Portal hosted at the NSDI, New Delhi. In the third phase beginning in 2008, considering the requirements of further technology customisation & development and intensive capacity building among the staff and personnel in Districts on Geospatial Technologies, a Centre of Excellence for NRDMS in Uttarakhand (COE NRDMS) was set up in 2009.

2.2 Centre for NRDMS at Kumaun University

With the mission to build professional education and applied research capacity in the latest geospatial technologies for various stakeholders, viz., academia, industry, governance and the civil society with special reference to the Uttarakhand State, the COE NRDMS has the following objectives:

(i) conduct education and research programmes in geospatial technologies, viz., Remote Sensing and Geographic Information Science;
(ii) develop spatial data infrastructure and GIS of the Uttarakhand state as per needs of the government departments; and
(iii) need-based capacity building of various stakeholders, viz. academia, industry, governance and civil society in application of geospatial technologies.

To replicate the Web-GIS template of District Almora in all the Districts of the State, a project on “Development of Standards-Based Uttarakhand Geo-Portal for Decentralised Governance” funded jointly by the DST, GOI and the DST, Government of Uttarakhand on 70% and 30% cost-sharing basis and collaboratively launched by the Uttarakhand State Council of Science and Technology (UCOST), Dehradun, and the COE NRDMS, Department of Geography, Kumaun University, Almora in November 2012. The project has ended in March 2018. With the help of this joint collaborative research project, Uttarakhand Geo-Portal has been developed for its application in planning and decision-making by the end-users. The specific characteristics of the Geo-Portal include the following:

(i) it is developed as per the needs of the users, i.e. different line departments;
(ii) it is developed in association with the line departments while sensitizing them on its utility and building their technical capacities;
(iii) it has provision of regular updating through the line departments;
(iv) it has a regular management system comprising of technically trained personnel at State, District and District line department levels; and
(v) it has also provision for developing its applications and upgradation as per the needs of the line departments.

The objective of this paper has been to provide the details of the method adopted in State-level spatial data management through the State Geo-Portal in the framework of the State SDI, its management structure, the architecture utilized in its implementation, the conceptual data model underlying the geospatial database that remains at its heart, provision of the standards-based data services for state-wide access by the end-users. Assessment of the geographic information needs of the end-users is vital to the development of this infrastructure to ensure better utilization.

3 End-User Need Assessment

The first two-day workshop on Information Need Assessment for Land and Water Management in District Almora was organized in May 2000 District planners/administrators/representatives of Municipal Board and Scientists of GB Pant Institute of Himalayan Environment and Development participated. Through this workshop, department-wise documentation of digital data needs for governance were assessed [2]. Based on these end-user information needs, the proposal was developed in association with the level planners and administrators for setting up a data management centre at Almora in 2002 and was supported by DST during 2002–2005.

To update the information needs for district planning and decision-making, the second workshop in the State on end-users need assessment was organized in August 2012 at Administrative Training Institution Nainital. In this workshop, the heads of line departments and administrators of Nainital district, participated and an updated digital information need of end-users was documented [3]. Based on these two end-users information need assessments, the district-wise GIS infrastructure of different line departments was developed in the state in association with the end-users by establishing District GIS Cell in each district.

4 Uttarakhand (UKH) Geo-Portal

4.1 Definition and Objectives

The UKH Geo-Portal is an Internet-based directory of geospatial data, metadata, and services for the State that allows data providers and users to share and explore geospatial information related to boundaries, demography, agro- and socio-economy, resources, infrastructure facilities with attributes. Geospatial data sets generated, maintained and provided by various concerned departments of the State Government, academia, business/industry or non-governmental organizations (NGOs) of Uttarakhand are proposed to be made accessible through a clearing house mechanism. To develop and maintain/update the data sets on Uttarakhand Geo-Portal, District GIS Cells have been established in all the districts of the State. Towards this end, Government of Uttarakhand has appointed a State-level Nodal Officer in each Department and a District-level Nodal Officer in each district of the State. For better coordination and management of data-related activities towards the development of the UKH SDI, the State Government has constituted a State SDI Steering Committee and a State SDI Executive Committee.

Thus, through the UKH Geo-Portal, a common platform of Web GIS at 1:50,000 scale has been developed for better managing all the spatial data sets of the State so that those could be updated and made sharable for use in GIS packages all over the State. Following the specifications of OGC's Geospatial Portal Reference Architecture, Ukh Geo-Portal development and updating has been in progress. It has provision to provide portal services, portrayal services, data services and catalogue services. The portal services shall be available to address the needs of the viewer clients, discovery clients, management clients, authentication and access control and exposed services. The portrayal services include map services, styling services, coverage services and the map context services. The data services include features, coverage and symbology management. The catalogue services include data discovery, service discovery, catalogue update and query language. The development which is still going on is being carried out by the users themselves with the technical support of GIS Cell established in each district/COE NRDMS Centre of Kumaun University.

The Geo-Portal is essentially being used for the following objectives:

1. Bringing together GIS-based knowledge and information of the Uttarakhand State under a common platform.
2. Providing a common gateway for various data generating agencies to share the information across various government departments, academia, business/industry and NGOs.
3. Making accessible a geographic data dictionary and map directory for the State.

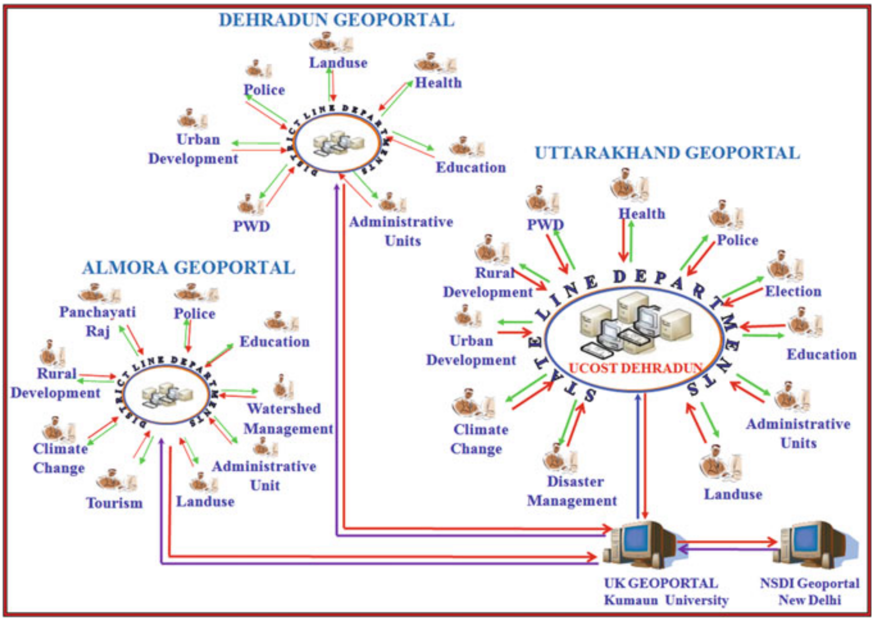


Fig. 1 High-level Architecture of Uttarakhand Geo-Portal for improving access to geospatial data at national, state and district levels

4. Analysing the needs of various government schemes/programmes.
5. Facilitating development of decision support systems and helping in local-level planning.

Figure 1 presents the architecture of the UKH Geo-Portal which reveals that it is being linked with all the district GIS Cells and also tagged with the India Geo-Portal.

4.2 Structure and Responsibility

The UKH Geo-Portal has seven major organizational components. These are

(i) Department of Planning, Government of Uttarakhand;
(ii) UCOST Dehradun;
(iii) NRDMS Centre, Department of Geography, Kumaun University;
(iv) Uttarakhand Space Application Centre (USAC) Dehradun;
(v) District GIS Cells;
(vi) District line departments; and
(vii) State line departments.

5 Capacity Building

For building the capacity of the State in development and application of GIS in decentralised and good governance, sector-wise training workshops have been organized. For example, Public Works Department (PWD) for development of GIS of Roads [4], Department of Revenue for modernization of land records [5], and Department of Police for police administration [6, 7], etc.

5.1 Setting-Up District GIS Cells

For developing GIS with the involvement of the end-users, i.e. line departments, the State Government has made provision of setting-up GIS Cell at each district headquarter, under the control of district magistrates. By conducting training workshops jointly by the COE NRDMS and the UCOST Dehradun on "GIS Development and Setting-up District GIS Cell" [8] at the headquarter of each district of the state, the GIS Cells have been established in all the 13 districts of Uttarakhand. The responsibility of the district GIS Cells has been to develop and regularly update GIS of different line departments and to build their technical capacity in application of GIS in governance.

5.2 Compilation and Publication of Metadata

A provision has been made in the system to permit various GIS data providers in the State to publish their metadata sets into the Geo-Portal/centralized database and get them registered directly at UCOST as per the metadata standards prescribed in BIS 16439. For data updation and sharing of up-to-date data through geo-web services, a metadata format as per the standard has been prepared for each layer of information.

6 Geographical Data Modelling

In geographic information systems, a data model is a mathematical construct for representing geographical objects in a formal and standardized way. In the design of the database for the UKH Geo-Portal, data models have been developed in various sectors like boundaries, physiography, hydrography, land use, geology, soils. Figure 2 provides overviews of the sector-wise data models used in the UKH Geo-Portal covering the end-user needs of 52 different sectors of the State and detailed feature-level classes in UML for the State Police Department have been shown in Fig. 3.

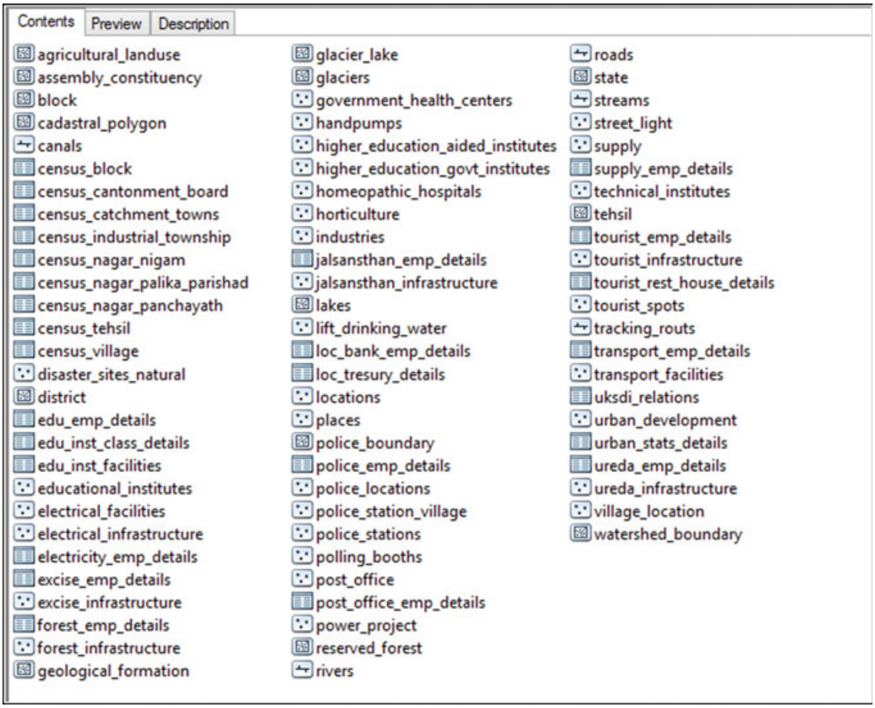


Fig. 2 Selected features from 52 different sectoral data layers of UKH Geo-Portal

6.1 Development of GIS Database

Based on the above data models, the GIS database of the State has been developed in association with district line departments and the district GIS Cells The database consists of sectoral data sets re-engineered in line with the Conceptual Schema Language specifications of the International Standardisation Organisation (ISO 19103) and implemented in the open source PostGreSQL/PostGIS. UCOST Dehradun hosts the UKH Geoportal along with the database.

6.2 Development of Web-GIS

To develop and demonstrate methods and techniques for better managing and utilising geospatial data sets in Uttarakhand State by the line departments, efforts are being made to make the database accessible along with Web GIS functionalities for processing. The spatial data sets in the Geo-Portal are also accessible in ESRI's shape format for the benefit of the ARC GIS users in the State. Associated metadata for the GIS data sets are appropriately included in the database to support discovery and

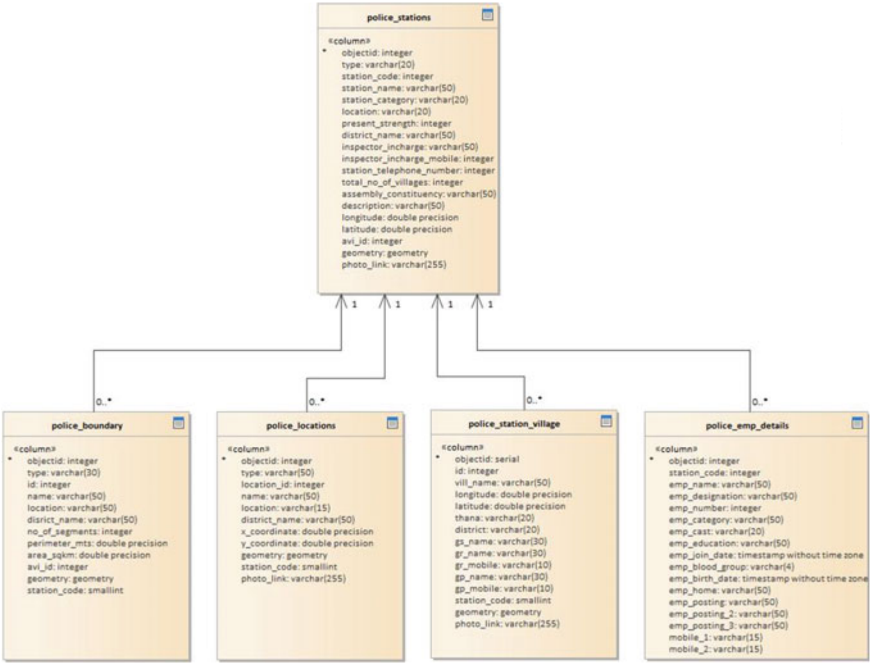


Fig. 3 UML Class diagram of different data items of the Department of Police, Government of Uttarakhand

access by concurrent users over the net. Mechanism for updation of existing data sets on the database using OGC standards-based services has been developed and implemented. Web Feature Service (WFS) and Web Map Service (WMS) have been provided to facilitate updation/maintenance of the database from users located in the individual line departments/agencies of the Government of Uttarakhand. In due course, depending on requirement, individual line departments/agencies of the government may be persuaded to develop and maintain their spatial/attribute data nodes as per their mandates in the interest of users/stakeholders of local-level planning for faster access to up-to-date data.

6.3 UKH Geo-Portal Servers

The UKH Geo-Portal hosted at the UCOST Dehradun consists of three servers—the database server, security server and the application server.

6.4 Security System

In order to secure the geo-portal/clearing house from unauthorized access, an appropriate security software module having provisions to safeguard data sets, manage user identity and transaction security; detect unauthorized intrusion, permit authorization/authentication, and non-repudiation-based access controls on resources have been installed. The security system has facilities to create federated access and single sign-on, on-the-fly creation of secured access and functionalities for user-friendly operations and access to data services from the Geo-Portal.

7 Provision of Standards-Based Data Services

UKH Geo-Portal provides services adhering to the ISO/OGC specifications like WMS, WFS, WCS, CS-W, WFS-T, and WPS.

A Web Map Service (WMS) is a standard protocol for serving georeferenced maps (spatial data with colours, styles, legend, annotation, etc.) for visualization of data sets by the end-users while taking decisions. Web Feature Service (WFS) provides an interface allowing requests for geographical features across the Web using platform-independent calls. WFS is the provision for upload/download of OGC's Geography Markup Language (GML) data (in GML 2.1.2 specification) for further processing in GIS package subject to authorization for download policy for sharing and download of geographical feature data by authorized user is currently being framed for approval by the State Cabinet. **Web Coverage Service Interface Standard (WCS**) provides an interface allowing requests for geographical coverage including map photographs or satellite images. **Web Processing Service (WPS)** is designed to standardize the way GIS calculations are made accessible to the end-users over the net. WPS can describe any calculation (i.e. process) including all of its inputs and outputs, and trigger its execution as a web service. **Catalogue Services on Web (CSW)** defines common interfaces to discover, browse and query metadata about data, services and other potential resources on UKH Geo-Portal's local register. The Uttarakhand local registers are proposed to be in turn linked to the proposed National Data Registry of NSDI for automatic harvesting and updation of the national registers for countrywide access and utilization.

8 Data Policy

The equitable, productive and sustainable developmental planning as well as good and inclusive governance need quality and authentic updated data and information. Different stakeholders of the State Government have been generating large quantities of different forms of data and information with public funds. But this activity, most of the times, creates redundant data generated utilizing public money, time and human resources because of lack of information about such data and openness in sharing the data sets between different government departments and stakeholders. General public also seems to be deprived of searching, discovering and accessing government-owned data from the concerned departments for participation in the governance process although most of such data may be non-sensitive in nature and shareable, because State Government departments do not generally publish catalogue or metadata of available data or even the data themselves. Sharing of relevant data is however essential to support effective and transparent governance. Further, due to lack of proper standards and interoperability of the data and information, inter-departmental sharing and integration of such data and information become difficult and ineffective, apart from its access and use by common citizens.

Hence, there is a need to evolve a protocol to facilitate easy accessibility, effective sharing and efficient utilization of the large amount of data generated and stored in disparate mode among the different entities of the State Government. There is a need for both independent dedicated data enterprise system in different government departments and need to share critical and relevant information across the departments with a central repository to make integrated data set available for informed decision-making. This calls for adoption of a data policy in the state which conceptualizes provisions for proper data management methods, defining data standards and processes for providing proactive and open access to such data.

The Draft State Data Policy has been formulated with a mandate of using government-owned data and information in the interest of concern citizens of the State. Under this policy, line departments will be mandated to share data and information with each other and other stakeholders including common citizen thus enhancing sustainable and productive use of the data sets in planning process of the State. The State Data Policy will provide greater scope for unified decision-making in the State. The basic objectives of Uttarakhand State Data Policy are:

(i) to support generation and storage of data and information in all forms at all State Government departments in a standard and interoperable format;
(ii) to implement a common protocol of updated data access and sharing in an integrated, proactive and periodic manner;
(iii) to promote use of all data and information including standardized geospatial data sets in planning, implementation and monitoring developmental programmes, disaster mitigation and scientific research; and
(iv) to facilitate accessibility of authentic data and information to by all the stakeholders through an electronic network within the framework of various related Policies, Acts and Rules of Government of Uttarakhand and Government of India made from time to time with a view to empowering the citizens.

Benefits of Uttarakhand State Data Policy are expected as follows:

(i) Avoiding data redundancy: By sharing data the practice of different State Government department/organizations collecting the same data will be avoided, resulting in significant cost and time savings.

(ii) Open information on data ownership: The preparation of data catalogue/metadata (data about data) by each government department will identify the ownership for the principal data sets to users and enable information about available data sets.
(iii) Maximizing use: Ready access to government-owned data using valuable public resource will enable more extensive use of data generated for the benefit of the citizens of the State.
(iv) Maximized integration: By adopting common standards for creation, generation, storage and transformation, the data can be made interoperable making integration of individual data sets feasible for holistic planning.
(v) Decision support systems generation: Interoperable data and information will support preparation of effective decision support systems needed in development planning, managing assets, improving socio-economic conditions, mitigating disasters, and maintaining internal security of the State, etc.
(vi) Equity of access: A more open and transparent data policy will ensure better access to data and information by all bonafide users in the State.

Using this policy, Government of Uttarakhand will strive to form one umbrella under which users can be able to access all data and information as per their choices. The scope and objectives of this policy may be expanded and enriched in future, as the data management technologies and procedures are change day by day.

For developing data policy and mechanism for data sharing among the users within and outside the State a Committee has been constituted under the Chairmanship of the Secretary, Science and Technology, Government of Uttarakhand.

9 Future Works

Thus, through the UKH Geo-Portal, a common platform of Web GIS at 1:50,000 scale has been developed for better managing all the spatial data sets of the State so that those could be updated and made sharable for use in GIS packages all over the State. The Uttarakhand State also has a coordinating structure with nodal officers at state/district levels. Now a road map for Uttarakhand State Geospatial Data Assets Management is a prerequisite and is proposed to be approved by the State Government for implementation by all the GIS users in the State. A data management (updation/maintenance/sharing) plan should be developed and implemented by UCOST, State Line Department Nodal Officers through a series of workshops. Regular workshops (3–4 per year) at State and District levels need to be organized for reviewing the progress of the management plan. The UKH Geo-Portal database needs to be continuously improved with not only data updation but with high-resolution data addition for its upgradation depending on GIS applications for different projects.

In view of the above necessity of smooth functioning and sustenance of the UKH Geo-Portal in order to provide uninterrupted access to data sets by end users, there is an urgent of need for development of a road map for formulation and publication of a framework and detailed guidelines for spatial data assets management and applications for the use of state.

10 Conclusions and Recommendation

Geo-Portals are becoming a major resource for gaining quicker access to geospatial data and services useful in decentralising governance on a Statewide scale. They are also contributing to sound decision-making, effectively implementing e-government applications and provision of better services to citizens. Geo-Portals are getting increasingly necessary for emergency preparedness and response, counterterrorism, monitoring of and response to pandemics, and environmental protection. In order for these transnational applications to be effective, Geo-Portal must interoperate. The interoperability can be achieved only through consistent and structured implementation of interface and encoding standards. The UKH Geo-Portal has successfully demonstrated that by building capacity of the end-users, i.e. line departments, they can convert their office records/data/document and maps into GIS environment and can also update GIS on regular basis. It is recommended that GIS should be developed in the country by capacity building of the end-users by which they could develop GIS as per their own need and could make application of GIS in governance, i.e. in planning, implementation, monitoring, management, surveying and research, and in administration for bringing transparent, good and decentralized governance by which the country could reach up to the logical culmination of the Digital India Mission.

Acknowledgements This paper constitutes parts of research projects on "Setting-up Digital Data base Centre District Almora", "Upgradation and Application of NRDMS Technologies in District Almora" and "Setting-up Centre of Excellence for NRDMS in Uttarakhand" sponsored by the NRDMS Division, DST New Delhi during 2002–2012; and project on "Development of Standard-based Uttarakhand Geo-Portal for Decentralised Governance" conducted during 2012–2018 jointly by the UCOST Dehradun and COE NRDMS Centre, Department of Geography, Kumaun University and collaboratively funded by the DST, Government of India and Government of Uttarakhand.

References

1. Wilson, J.P., Fotheringham, A.S.: The Handbook, of Geographic Information Science (Edited), pp. 1–620. Blackwell Publishing, USA, UK, Australia (2008)
2. Rawat, J.S.: Report on Two Day Workshop on Information Digital Information Need Assessment for land and water management in district Almora, sponsored by the NRDMS Division, DST New Delhi, organised during 18–19 May, 2000 at Department of Geography, Kumaun University, SSJ Campus Almora, pp. 1–37 (2000)
3. Rawat, J.S.: Report on Information Need Assessment Workshop on GIS Development, sponsored by the NRDMS Division, DST New Delhi, organised on 8th August, 2012 at Administrative Training Institute Nainital, sponsored by the NRDMS Division, DST New Delhi, pp. 1–34 (2012)

4. Rawat, J.S.: Kumaun Engineers' Meet 2013 for GIS Development, A Report, Centre Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University SSJ Campus Almora, pp. 1–11 (2013a)
5. Rawat, J.S.: Modernization of Land Records, A Report, Centre Excellence for NRDMS in Uttarakhand Department of Geography, Kumaun University SSJ Campus Almora, pp. 1–13 (2013b)
6. Rawat, J.S.: Application of GIS in Police Administration, A Report, Centre Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University SSJ Campus Almora, pp. 1–7 (2009)
7. Rawat, J.S.: Workshop on Development and Management of Police GIS in District Almora, A Report, Centre Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University SSJ Campus Almora, pp. 1–12 (2010)
8. Rawat, J.S.: GIS Development and Setting-up GIS Cell District Bageshwar, A report, COE, NRDMS, Department of Geography, Kumaun University, SSJ Campus Almora, pp. 1–56 (2016)

Human Resource Development in Geospatial Data Modeling and Services for Better Realization of National Spatial Data Infrastructure

S.K. Ghosh

1 Introduction

With the increasing availability of geospatial data, there is an urgent necessity to promote spatial data sharing, and to develop the ways for providing easy accessibility on such data so as to facilitate in spatial data discovery, evaluation, and various geospatial applications/services. However, the formats of these dispersed data, stored in distributed data repositories, may vary from organization to organization, making the process of data sharing a difficult and time-consuming task. In order to facilitate sharing of these diverse data sets, it is needed to model the individual databases using some standardized technique [2]. However, the domain knowledge needed for appropriate modeling often lies with the respective organizations. Thus, the domain experts of the organizations may help in effective modeling and vis-á-vis sharing of the data. This work aims at designing a training resource module to enrich the users of the line departments with the working principles of a standardized framework for geospatial data modeling and services, and thus, it can aid in further development of the national spatial data infrastructure (NSDI).

1.1 Objective

The key purpose of the work is to realize the interoperability among the various spatial and non-spatial data infrastructures provided at the national level as well as commercial level, and thereby to overcome the difficulty in coordinating data acquisition, storing, processing, and dissemination across distributed data repositories. The major objectives in this regard are summarized below:

S. K. Ghosh (✉)
Department of Computer Science and Engineering,Indian Institute of Technology Kharagpur,Kharagpur 721302, West Bengal, India
e-mail: skg@cse.iitkgp.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_7

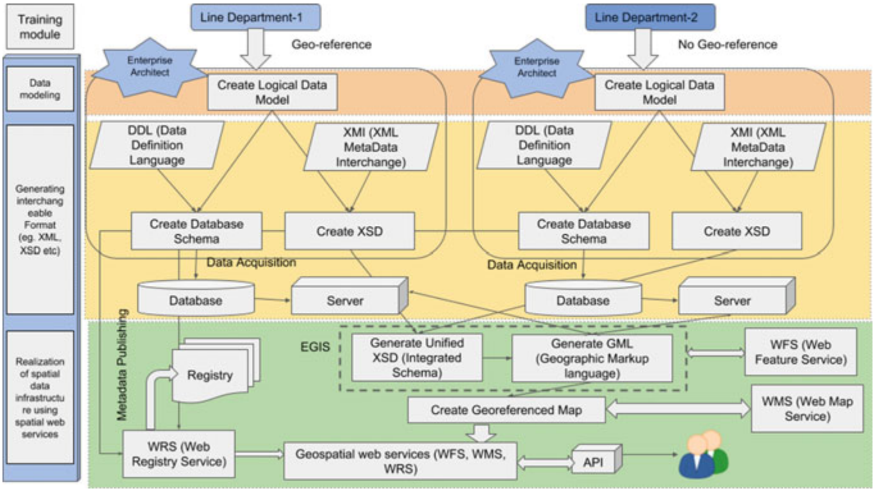


Fig. 1 Overall framework of the proposed module

- Modeling geospatial data
- Generating XMI, XSD, DDL, and GML from the geospatial data model
- Integrating the GML into E-GIS framework
- Introducing registry service for spatial data.

The rest of the chapter is organized as follows. Section 2 provides a detailed description of the proposed resource module. A case study on the overall standardized process is explained in Sect. 3. Finally, the concluding remarks have been made in Sect. 4.

2 Proposed Approach: An Overview

The overall framework of the proposed approach is depicted in Fig. 1. As shown in the figure, the module consists of three major components: (i) geospatial data modeling, (ii) generating interchangeable format (XMI, XSD, etc.) from the data model, and (iii) realization of SDI using spatial Web services. The details of each of these components are discussed in the subsequent subsections.

2.1 Geospatial Data Modeling

This component of the training module aims at providing a standardized process for modeling geospatial data available with various line departments. The objective is to represent the data in a simplified format using a minimum set of standards/specifications and thereby to facilitate in easy accessibility and sharing/distribution of the data. The proposed approach uses the unified modeling language (UML) class diagram [1] as the standardized technique for modeling geospatial data, and it employs the Enterprise Architect (EA) as the data modeling software [4]. The overall process of creating geospatial data model (in terms of UML class diagram) using Enterprise Architect is described below.

This component focuses on converting the data model into exchangeable/interchangeable format in terms of XMI and XSD representation. The data model represented in terms of UML class diagram may or may not be supported by other UML tools/packages used by other organizations. But, if every organization keeps their data in XMI format, the schema of the proprietary data becomes easily shareable. Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is human-readable as well as machine-readable, and XML Metadata Interchange (XMI) is a standard, proposed by Object Management Group (OMG), for exchanging metadata information via XML. On the other side, XSD is an XML schema definition language which can be used to express a set of rules to which an XML document must conform in order to be considered 'valid' according to that schema.

The entire data model can be developed by following the above-mentioned steps to represent the various entities and their relationships with each other. In a spatial data model, at least one class should contain spatial attribute(s).

Steps to create logical data model (Class Diagram) using Enterprise Architect:

Step 1: Open an existing project in Enterprise Architect, or create a new one.

Creating a new project: Start Enterprise Architect → Click the *File* option from the menu bar → Click on *New Project* option → Locate a suitable folder for your project and enter a distinctive name → Press the *Save* button to continue → Now, a *Model Wizard* is opened. In the left hand column ensure that the option, *Core Modelling* selected as model technology, and in the right hand side tick the *Class* check box as the model name → Click *OK*.

Step 2: Select a package or view in the Project Browser → Right click on the workspace → Select *New Element* or *Connector* → Select *Class*.

Step 3: Update the diagram name, if required.

Step 4: Select *UML Structural* as the category of diagrams in the left hand pane and *Class* as the diagram type from the right pane and click on the *OK* button. This creates a new diagram as a child of the currently selected package and it is opened in the *Diagram View* which is now ready for editing.

Step 5: ***Add the required classes in the diagram:*** Click on the *Class* option in toolbox → Drag it in the workspace → Double Click on the *Class* → Provide appropriate name to the class → Set stereotype as `Table' Click on OK
To modify or update the name, appearance, database etc. of a particular class, first select the class and then double click on it. As a result, a dialog box corresponding to the properties of the class is opened, where you can make the necessary changes.

Step 6: Add attributes to the classes: Right click on the class → Feature & Properties → Click on the option `Attributes'→ Provide *Name* and *Type* of the attribute → Add as many attributes as required → Click *Close* when all the attributes are added.

Step 7: In the same fashion, if required, the operations can also be added to the classes in following manner:
Right click on the class → Feature & Properties → Click on the option `Operations' → Provide name, parameters and return type of the operation → Add `Constraints' of the operations, if required →Add as many as operations →Click close when all the operations are added.

To update or modify any attribute (or operation) of a particular class, first select the class and then click on that particular attribute (or operation). As a result, the attribute (or operation) will be selected. Now right click on that selected attribute (or operation) and go to View Properties. This will open a dialog box corresponding to the properties of that attribute (or operation).
Now you can make the required changes in the relevant fields.

Step 8: Add connectors between the classes (if required).
Different types of connectors are used in UML representation of different logical data models. For example, Associate, Generalize, Aggregate etc. are some of the connectors used in the class diagram. The steps to add connectors between two classes have been demonstrated below.
Select connector as required from the toolbox→ Connect the required classes of your choice.

To add name and direction of the connector: Right click on the connector → Provide appropriate name → Select appropriate direction → Click on *OK*

To add multiplicity at the end of a connector: Double click at the end of the connector → Association properties will be shown → Select *Role(s)* → Select *Multiplicity* → Provide appropriate *Multiplicity* value both in source and destination/target class.

2.2 Generating XMI, XSD, and DDL from the Spatial Data Model

This component also introduces the process of converting the data model into data definition language or DDL format which can be executed to create the database schema maintaining the original data model. The DDL file is basically a structured query language (SQL) script file with a .sql extension. It contains the SQL statement which, when executed one by one, automates the process of schema generation. The generated schema eventually helps to publish metadata information in geospatial service registry, so that any client system can automatically bind with the associated geospatial service.

The various steps to convert the data model into XMI, XSD, and DDL format are described below.

2.2.1 Logical Data Model to XMI

Following are the different steps to convert a data model, represented through an UML diagram, into XMI format using Enterprise Architect (version 12.1):

Step 1 : Open the UML representation of the data model.
Step 2 : Go to Package option.
Step 3 : Select Import/Export option.
Step 4 : Choose Export Package to XMI.
Step 5 : Provide the location in the Filename text box to save the *.xmi* file.
Step 6 : Choose the Export Type.
Step 7 : Now click on the Export button. The *.xmi* file will be saved to the specified location.

The XMI representation can be viewed by clicking on the *View XMI* button in the *Export Package to XMI* dialog box.

2.2.2 Logical Data Model to XSD

In order to produce the XML schema (XSD) representation of logical data model, the '*Generate XML Schema*' feature of the Enterprise Architect tool is used. This feature helps to generate a *W3C XML Schema* (XSD) file corresponding to a given logical data model. The following are the various steps to convert a data model into XSD format by using Enterprise Architect tool. The steps are described below with respect to the Enterprise Architect tool version 12.1.

Step 1 : In the Project browser, right click on the root package (Class Model).
Step 2 : Go to Code Engineering.
Step 3 : Select Generate XML Schema. A dialog box, named Generate XML Schema will open.
Step 4 : Check the package Class Model.
Step 5 : Provide the location in the Filename text box to save the XSD in your preferred location.
Step 6 : Click on Generate button. The *.xsd* file will be saved to the specified location.

The XSD representation can be viewed by clicking on the *View Schema* button in the *Generate XML Schema* dialog box.

2.2.3 Logical Data Model to Database Schema

Enterprise Architect also offers facilities to generate DDL for the different elements/components in logical data model. SQL script can be generated for a single element (a Table) or for a whole package. Following are the steps to convert a data model into DDL format using Enterprise Architect (version 12.1):

Step 1 : In the Project browser, right click on the root package (Class Model).
Step 2 : Go to Code Engineering.
Step 3 : Select Generate DDL. A dialog box, named Generate Package DDL will open.
Step 4 : To recursively generate DDL, select the Include All Child Packages checkbox.
Step 5 : Click on Single File in the File Generation option. Browse your preferred location to save the DDL file.

The corresponding SQL can be viewed by clicking on the *View* button in the *Generate Package DDL* dialog box. Executing this *.sql* file in Oracle will create the relevant schema of the data in the database.

2.3 Realization of SDI Using Spatial Web Services

This part of the training module introduces the concept of geospatial *Web services* and *registry services*, for integrating heterogeneous geospatial data and facilitating the data reuse/sharing process.

2.3.1 Spatial Web Services (WMS and WFS)

- **Web Map Services (WMSs)**: A Web Map Service (WMS) is a standard for providing visualization of geospatial data over the Internet. It allows the users to retrieve information on a map over the Internet or to publish map/GIS layers onto the Web. WMS operations can be invoked using a standard Web browser by submitting requests in the form of URLs. The URL indicates the image format, the output image height and width, what information is to be displayed on the map, which part of the earth is to be mapped, the desired coordinate reference system, and so on. A Web Map Service is usually not invoked directly. Most of the time, it is invoked by a client application that provides the user with interactive controls. Additionally, WMS supports the creation and display of registered and superimposed maps that come simultaneously from multiple sources which are remote and heterogeneous.
- **Web Feature Services (WFSs)**: A Web Feature Service (WFS) is an interface allowing requests for geographical features across the Internet using platform-independent calls in the form of URLs. Geographical features can be thought of as the 'source code' behind a map, whereas the WMS interface returns only an image, which cannot be edited or spatially analyzed. The WFS specification defines interfaces for describing data manipulation operations of geographic features. The data manipulation operations include the ability to:
  - Get or query features based on spatial and non-spatial constraints
  - Create a new feature instance
  - Delete a feature instance
  - Update a feature instance.

The data is passed back and forth between a Web Feature Server and a client by means of the Geography Markup Language (GML) [3]. GML is an XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. Web Feature Service (WFS) defines operations to manipulate information regarding geographic features such as points, lines, and polygons. These operations allow to execute query and permit transactions, like create, update, or delete, on spatial data through the Web. The geometric descriptions of features in the WFS specification are encoded in GML.

2.3.2 Integration of Data Sources

The overall process of data integration is performed with the help of geoservices, and it is comprised of the following three steps: (i) accessing local WFS, (ii) publishing local WFS, and (iii) publishing remote WFS. Each of these steps is described below, with respect to the 'IITKGP GeoServices' and assuming the underlying database to be Oracle Spatial. The 'IITKGP GeoServices' is an Enterprise GIS, an OGC complaint geo-spatial web-services installed in a laboratory in IIT Kharagpur for geo-spatial research, framework that offers the main building blocks for spatial data infrastructures (SDIs).

- **Accessing local WFS**: In order to access the local WFS, it is necessary to collect the information regarding *IP and port of Oracle database system*, *SID name of database*, *Username*, and *Password* from the database administrator. Then, a XSD file is developed for accessing Oracle data. The same data is also presented in the form of GML.

  After creating the XSD file, it is saved in the following location: */Tomcat 7.0/webapps/iitkgp-geoservice/WEB-INF/conf/wms/featuretypes*

  Then, the Tomcat server is restarted and checked whether the XSD is working properly by using this URL: http://localhost:8080/iitkgp-geoservice/

  Now, in order to download the required data (in GML file format), one should use the following URL: http://10.14.81.6:8080/iitkgp-geoservice/client/client.html and enter the appropriate XML request. A typical interface for IITKGP GeoServices *Client* to access WFS is shown in Fig. 2.

- **Publishing Local WFS**: In order to publish a geospatial feature, the following information is required:
  - Spatial reference system (SRS)
  - Minimum and maximum values of latitude and longitude from downloaded GML file
  - Name of the <xsd:element name> as in .xsd file
  - Name of the geometry element in .xsd file.

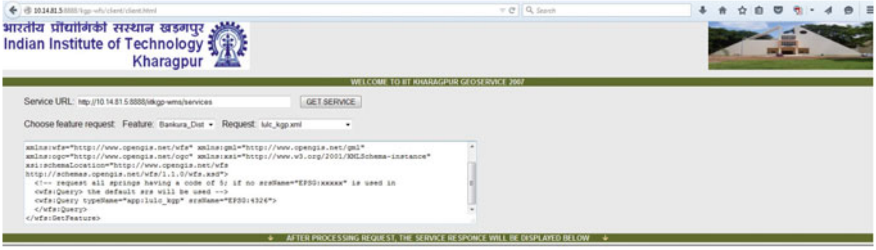


Fig. 2 IITKGP GeoServices client to access Web Feature Services (WFSs)

The file 'wms_configuration.xml' is updated with the above-mentioned information (relevant to the user requirement), and the local WFS layer is published. Now, the user can browse the map using appropriate URL.

- **Publishing Remote WFS**: In order to publish remote WFS, it is necessary to access WFS *GetCapability* service with remote URL. Again, *Describe Feature Service* is used to access attributes of the selected feature, because geometry attribute is essential to display the map.

  For example, in NSDI Web portal, the WFS GetCapability URL is as follows: https://nsdiindia.gov.in/erdas-apollo/vector/POPULATION?service=WFS&request=GetCapabilities

  The Describe Feature Service URL is as follows: https://nsdiindia.gov.in/erdas-apollo/vector/POPULATION?serviceWFS&version=1.1.0&request=DescribeFeatureType&typename=iwfs:POPULATION

2.3.3 Geospatial Registry Service

A geospatial registry service is a software component that supports the run-time discovery and evaluation of resources, such as services, data sets, application schema, and thereby facilitates the data reuse and sharing process. The primary utilities of the registry services can be summarized as follows:

- Enhancing interoperability among spatial services
- Facilitating communication for publishing and discovery of the services among providers and consumers
- Reducing the manual interventions toward service utilities
- Providing intercommunications among other registry services across the Web
- Offering mechanism to crawl the entire site of the services to automate the registration of its metadata.

There are *two* techniques that can be employed to publish the metadata through a geospatial registry service:

- **Pull-based Registration**: This technique provides an incremental harvesting model. Harvesting is the process of ingesting metadata from remote sources and storing it locally in the catalog for fast searching. It is a scheduled process, so local copy and remote metadata are kept aligned. The whole harvest process is mainly composed of three steps as described below.

> **Step 1:** Retrieving time and incremental items from remote servers
> **Step 2:** Processing fetched items with UUID (Universally Unique IDentifier)
> **Step 3:** Posting updates and logging the harvest time

- **Push-based Registration**: This technique involves the manual publishing of the service metadata by the providers. In this case, the published services need to be *pushed* into the corresponding registries. The steps of *push* registrations are as follows.

> **Step 1:** Send a request to the registry admin for creating an account within the registry and obtaining the authentication credentials for the same
>
> **Step 2:** Enter into the registry account through the provided user_id and password from the registry admin
>
> **Step 3:** Goto to the Contribute tab followed by the Add new record button
>
> **Step 4:** Select the necessary option among Dataset, Map, Service as relevant to the service that is to be provided
>
> **Step 5:** After completion of all the necessary details about the shared resources, the click the Save metadata to save the changes
>
> **Step 6:** All shared and saved metadata can viewed through Contribute tab by the service provider. Provider also view the publicly shared metadata from this tab. Service provider has provision to modify, delete and duplicate own metadata details from here

Once the metadata is published, it can be accessed by any client, through the search portal associated with the registry service. The client may be either machine or human user. For machine user, *get* or *post* services are allowed to retrieve the metadata information. This registry service portal also provides an user interface through which any data or map service can be searched manually.

3 Case Study: Accident Management System

This section illustrates the proposed training module with respect to a real-life case study on geospatial application for *Accident Management*.

The ‘Accident Management’ system is an end-to-end framework for collecting and reporting information about accidents, and storing them in a database in a seam-

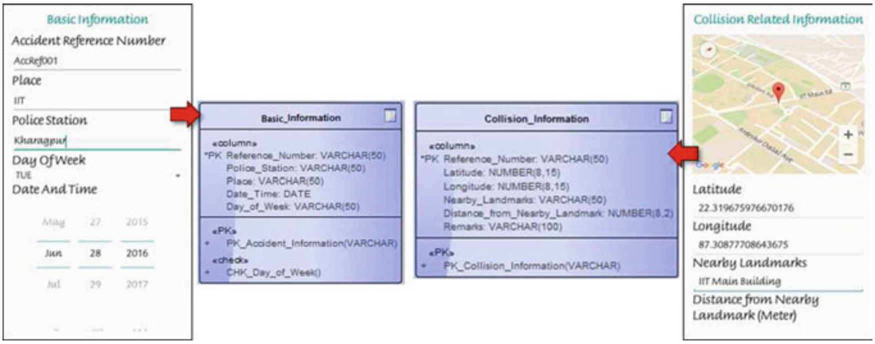


Fig. 3 Typical data acquisition forms in the 'Accident Management' application and the corresponding class representations in the logical data model

```
-- Create Tables
CREATE TABLE    Basic_Information
(
    Reference_Number   VARCHAR(50) NOT NULL,
    Police_Station     VARCHAR(50),
    Place              VARCHAR(50),
    Date_Time          DATE,
    Day_of_Week        VARCHAR(50)
)
;
....
....
CREATE TABLE Collision_Information
(
    Reference_Number                VARCHAR(50) NOT NULL,
    Latitude                        NUMBER(8,15),
    Longitude                       NUMBER(8,15),
    Nearby_Landmarks                VARCHAR(50),
    Distance_from_Nearby_Landmark   NUMBER(8,2),
    Remarks                         VARCHAR(100)
)
;
```

Fig. 4 DDL representation of the logical data model for the 'Accident Management' application

less manner. With the use of this system, the painstaking efforts of manually filling a form are replaced by a few taps on the phone. The application is developed based on the form provided by the Police Department. It has various fields for the different information about the accident, including basic information, accident-related factors, road-related factors, collision related factors, which the official needs to fill. The application then takes the GPS coordinates from the phone (or the official can manually select the location of the accident), and the data is sent to the server maintaining an accident information database.

Now, in order to ensure an easy accessibility and sharing of these data among various line departments, the overall process needs to comply with some standards/protocols/specifications, and the proposed standardized framework can successfully be followed in this context. Some typical portions of data acquisition form in Accident Management application and the corresponding data model (class diagram representation) prepared as per the proposed framework are depicted in Fig. 3. The XMI and DDL representation of the same data model is depicted in Fig. 5 and Fig. 4, respectively.

Further, in order to facilitate the discovery of relevant information about this application, the corresponding metadata or schema, in the form of XMI format, is uploaded or published in the registry service (Fig. 6). A typical view of the uploaded and shared schema of the 'Accident Report' in registry service is shown in Fig. 7.

```xml
<?xml version="1.0" encoding="windows-1252"?>
<XMI xmi.version="1.1" xmlns:UML="omg.org/UML1.3" timestamp="2016-06-25 23:07:24">
    <XMI.header>
        <XMI.documentation>
            <XMI.exporter>Enterprise Architect</XMI.exporter>
            <XMI.exporterVersion>2.5</XMI.exporterVersion>
        </XMI.documentation>
    </XMI.header>
    <XMI.content>
        <UML:Model name="EA Model" xmi.id="MX_EAID_0EA945B2_A12A_46c4_89B5_FFC66FBDCB5F">
            <UML:Namespace.ownedElement>
                <UML:Class name="EARootClass" xmi.id="EAID_11111111_5487_4080_A7F4_41526CB0AA00" isRoot="true" isLeaf='
                <UML:Package name="Class Model" xmi.id="EAPK_0EA945B2_A12A_46c4_89B5_FFC66FBDCB5F" isRoot="false" isLe
                    <UML:ModelElement.taggedValue>
                        <UML:TaggedValue tag="parent" value="EAPK_44C6D61F_E16E_4ca0_8C8B_A3CDC08B06BC"/>
                        <UML:TaggedValue tag="ea_package_id" value="2"/>
                        <UML:TaggedValue tag="created" value="2016-06-12 14:58:27"/>
                        <UML:TaggedValue tag="modified" value="2016-06-12 14:58:28"/>
                        <UML:TaggedValue tag="iscontrolled" value="FALSE"/>
                        <UML:TaggedValue tag="isnamespace" value="1"/>
                        <UML:TaggedValue tag="lastloaddate" value="2016-06-12 14:58:27"/>
                        <UML:TaggedValue tag="lastsavedate" value="2016-06-12 14:58:27"/>
                        <UML:TaggedValue tag="isprotected" value="FALSE"/>
                        <UML:TaggedValue tag="usedtd" value="FALSE"/>
                        <UML:TaggedValue tag="logxml" value="FALSE"/>
                        <UML:TaggedValue tag="tpos" value="6"/>
                        <UML:TaggedValue tag="packageFlags" value="isModel=1;VICON=3;CRC=0;"/>
                        <UML:TaggedValue tag="batchsave" value="0"/>
                        <UML:TaggedValue tag="batchload" value="0"/>
                        <UML:TaggedValue tag="phase" value="1.0"/>
                        <UML:TaggedValue tag="status" value="Proposed"/>
                        <UML:TaggedValue tag="complexity" value="1"/>
                        <UML:TaggedValue tag="ea_stype" value="Public"/>
                        <UML:TaggedValue tag="tpos" value="6"/>
                        <UML:TaggedValue tag="genfile" value="D:\Workshop_july_2016\Class Model.sql"/>
                    </UML:ModelElement.taggedValue>
                    <UML:Namespace.ownedElement>
                        <UML:Class name="Pedestrian" xmi.id="EAID_097FFF73_D250_47a7_9854_C69737F73423" visibility="pu
                            <UML:ModelElement.stereotype>
                                <UML:Stereotype name="table"/>
                            </UML:ModelElement.stereotype>
                            <UML:ModelElement.taggedValue>
```

Fig. 5 XMI representation of the logical data model for the ‘Accident Management’ application

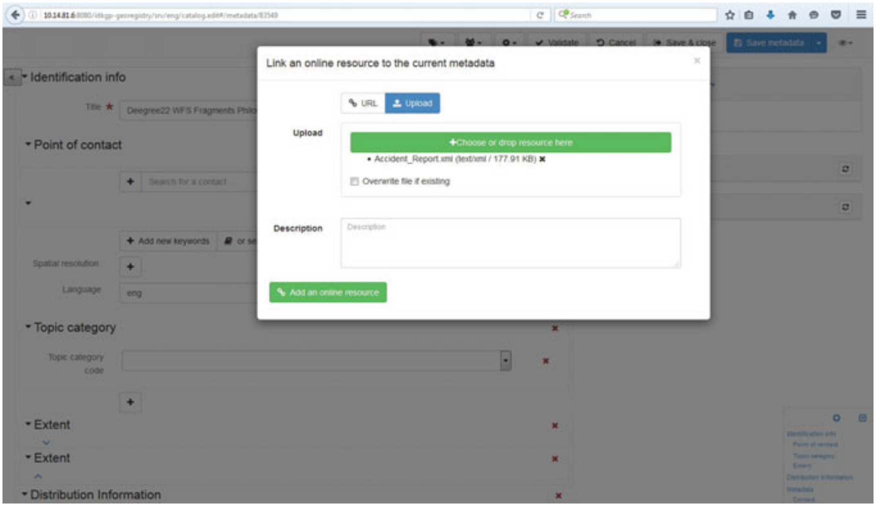


Fig. 6 Uploading ‘Accident_Report.xmi’ file in spatial Web registry

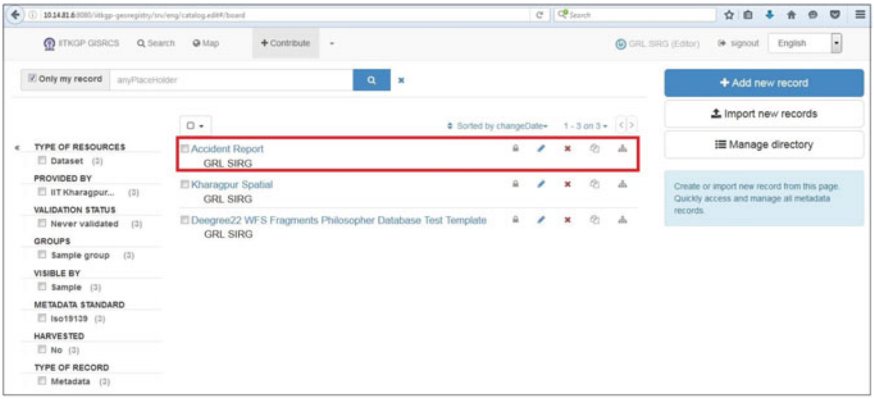


Fig. 7 View of 'Accident Report' entry in spatial Web registry

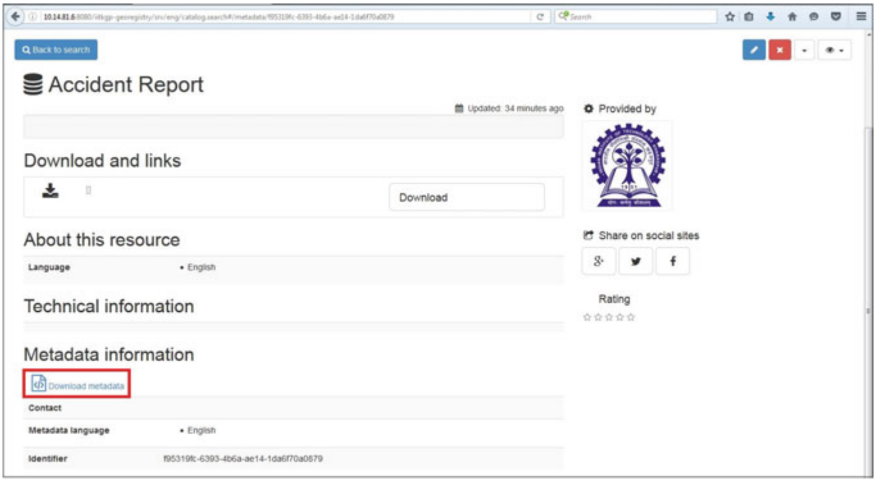
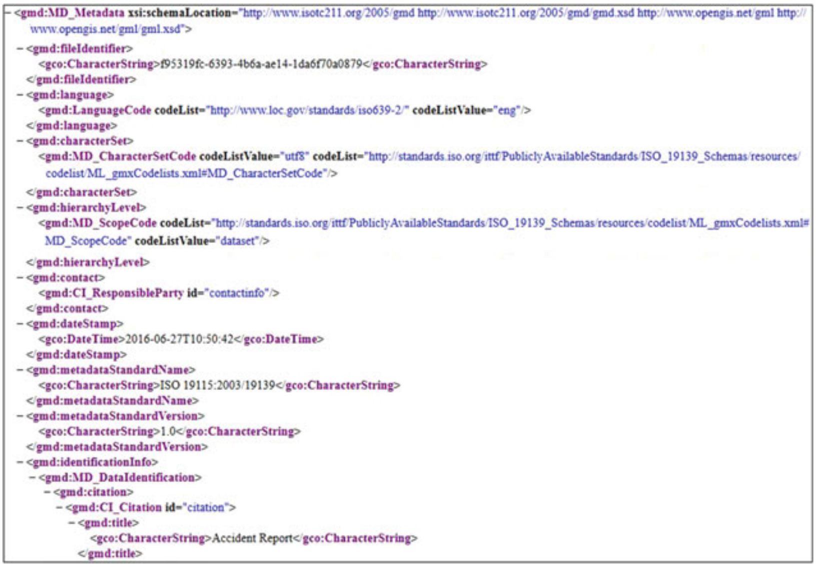


Fig. 8 Downloadable 'Accident Report' metadata link

After clicking on 'Accident Report' entry, metadata of this Accident Management application/service can be downloaded (Fig. 8). The detailed view of this metadata is shown in Figs. 9 and 10.



```
- <gmd:MD_Metadata xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://www.isotc211.org/2005/gmd/gmd.xsd http://www.opengis.net/gml http://www.opengis.net/gml/gml.xsd">
 - <gmd:fileIdentifier>
     <gco:CharacterString>f95319fc-6393-4b6a-ae14-1da6f70a0879</gco:CharacterString>
   </gmd:fileIdentifier>
 - <gmd:language>
     <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="eng"/>
   </gmd:language>
 - <gmd:characterSet>
     <gmd:MD_CharacterSetCode codeListValue="utf8" codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_CharacterSetCode"/>
   </gmd:characterSet>
 - <gmd:hierarchyLevel>
     <gmd:MD_ScopeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_ScopeCode" codeListValue="dataset"/>
   </gmd:hierarchyLevel>
 - <gmd:contact>
     <gmd:CI_ResponsibleParty id="contactinfo"/>
   </gmd:contact>
 - <gmd:dateStamp>
     <gco:DateTime>2016-06-27T10:50:42</gco:DateTime>
   </gmd:dateStamp>
 - <gmd:metadataStandardName>
     <gco:CharacterString>ISO 19115:2003/19139</gco:CharacterString>
   </gmd:metadataStandardName>
 - <gmd:metadataStandardVersion>
     <gco:CharacterString>1.0</gco:CharacterString>
   </gmd:metadataStandardVersion>
 - <gmd:identificationInfo>
   - <gmd:MD_DataIdentification>
     - <gmd:citation>
       - <gmd:CI_Citation id="citation">
         - <gmd:title>
             <gco:CharacterString>Accident Report</gco:CharacterString>
           </gmd:title>
```


Fig. 9 Metadata (in .xmi format [Part-1]) of 'Accident Management' application

```
         </gmd:CI_Citation>
       </gmd:citation>
       <gmd:abstract id="abstract"/>
     - <gmd:pointOfContact>
         <gmd:CI_ResponsibleParty id="contactinfo"/>
       </gmd:pointOfContact>
       <gmd:descriptiveKeywords id="keywords"/>
     - <gmd:language>
         <gco:CharacterString>eng</gco:CharacterString>
       </gmd:language>
     - <gmd:topicCategory>
         <gmd:MD_TopicCategoryCode/>
       </gmd:topicCategory>
     - <gmd:extent>
         <gmd:EX_Extent id="boundingbox"/>
       </gmd:extent>
     - <gmd:extent>
         <gmd:EX_Extent id="tempextent"/>
       </gmd:extent>
     </gmd:MD_DataIdentification>
   </gmd:identificationInfo>
 - <gmd:distributionInfo>
   - <gmd:MD_Distribution>
     - <gmd:transferOptions>
       - <gmd:MD_DigitalTransferOptions>
         - <gmd:onLine>
           - <gmd:CI_OnlineResource>
             - <gmd:linkage>
               - <gmd:URL>
                   http://localhost:8080/iitkgp-georegistry/srv/eng/resources.get?uuid=f95319fc-6393-4b6a-ae14-1da6f70a0879&fname=Accident_Report.xml&access=public
                 </gmd:URL>
               </gmd:linkage>
             - <gmd:protocol>
                 <gco:CharacterString>WWW:DOWNLOAD-1.0-http--download</gco:CharacterString>
               </gmd:protocol>
             - <gmd:name>
                 <gmx:MimeFileType type="text/xml">Accident_Report.xml</gmx:MimeFileType>
               </gmd:name>
             </gmd:CI_OnlineResource>
           </gmd:onLine>
         </gmd:MD_DigitalTransferOptions>
       </gmd:transferOptions>
     </gmd:MD_Distribution>
   </gmd:distributionInfo>
</gmd:MD_Metadata>
```


Fig. 10 Metadata (.xmi format [Part-2]) of 'Accident Management' application

Table: Basic_Information

Reference_Number	Police_Station	Place	Date_Time	Day_of_Week
AccRef001	Kharagpur	IIT	28-6-2016 10:55	TUE
AccRef2001	Kharagpur	Golbajar	23-6-2016 16:0	THU
accRef3001	Kharagpur	Inda	20-6-2016 17:51	MON

Table: Collision_Information

Reference_Number	Latitude	Longitude	Nearby_Landmarks	Distance_from_Nearby_Landmark	Remarks
AccRef001	22.319675976670176	87.30877708643675	IIT Main building	200	No traffic police
AccRef2001	22.354432494388654	87.26828511804342	Shiv Mandir	500	Damaged road
accRef3001	22.262861337286125	87.37437855452299	Water tank	100	busy road

Fig. 11 Visualization of the data regarding *basic accident information* and *collision information* stored in database

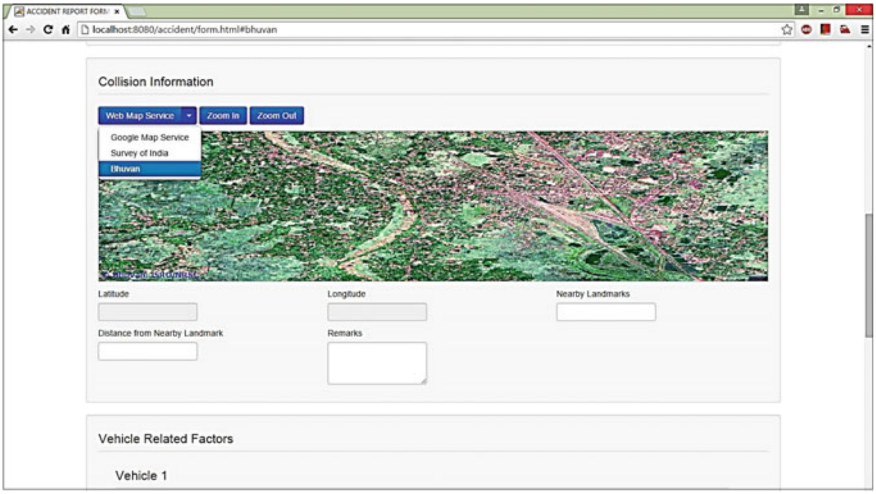


Fig. 12 ‘Accident Management’ application embedded with Bhuvan Web Map Service

A typical view of data stored in database and its access/visualization through Bhuvan WMS (Web Map Service) is shown in Fig. 11 and Fig. 12, respectively. It can be noted that the structure of the data stored in the database maintains the data model as depicted in Fig. 3.

4 Conclusion

The present work proposes a training module that throws light on the process of generating standardized, interoperable, value-added spatial data, and attempts to develop human resources for a better realization of the NSDI. However, the realization of NSDI is a time-consuming process and it is difficult to procure a single perfect model for easy implementation of the same. Hence, in order to establish an ideal model for this purpose, a long and continuous process of further refinement is necessary.

Acknowledgements This work is financially supported by the Department of Science and Technology (DST), Government of India, New Delhi. NRDMS Division, DST, and NSDI, New Delhi, have provided technical support for the work. IIT Kharagpur has provided necessary infrastructure to carry out the project.

References

1. Mall, R.: Fundamentals of Software Engineering. PHI Learning Pvt., Ltd (2014)
2. NSDI: National Spatial Data Infrastructure (NSDI). https://nsdiindia.gov.in/nsdi/nsdiportal/meetings/ReportsOftheWorkingGroups.pdf (2011). [Online; Jan-2018]
3. OGC: Geography Markup Language. http://www.opengeospatial.org/standards/gml (2017), [Online; Jan-2018]
4. Sparx-System: Enterprise-Architect. http://sparxsystems.com/products/ea/ (2017). [Online; Jan-2018]

Part II
Techniques and Technology Research

InGIST: A Queryable and Configurable IndoorGIS Toolkit

V. Balasubramanian

1 Introduction

The growth in pervasive and ubiquitous computing has opened up the floodgates for many exciting applications like asset and people tracking, crowd monitoring, smart health care, effective emergency response, energy-efficient building solutions, intelligent infrastructures, etc, and many of these applications operate indoors. The foundation for pervasive systems is spatial–temporal databases, specifically indoor information systems. One must be able to configure and deploy an indoor information system for a given indoor space so that pervasive applications can be configured and monitored effectively.

An indoor information system is defined as a system that gives information about the details of a building such as the different spaces within the building, exits, and their properties and also about other entities in the buildings like utilities, assets, and people. It not only allows the users to view the building layout but also enables end users to query the system for information like exits, paths, the location of rooms, etc.

There are, however, several challenges in designing such systems that are deployed over any indoor space. Most building plans are in CAD format, and recently, BIM and CityGML [7] formats are getting popular. One of the problems of such representations is that it is difficult to extract features of a building like rooms, doors, and exits. It is essential to obtain both building information and semantic data from such representations automatically, which is not straightforward [6]. Another challenge in these models is that such formats do not support querying over indoor space. Hence, there is a need to convert CAD data into query-friendly GIS formats and extract metadata from CAD maps to augment the GIS representations [6].

This work has been funded in part by DST(India) grant DyNo. 100/IFD/2764/2012-2013.

V. Balasubramanian (✉)
Department of Computer Science and Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: b_vidhya@cb.amrita.edu

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_8

Indoor Modeling Once the data is extracted, it should be represented in appropriate indoor models. However, existing indoor databases/models are primarily developed for specific applications like navigation systems [14], 3D visualization of buildings, etc. There are many indoor models, both hierarchical and integrated [3, 4, 10, 12] and standards like IndoorGML [11], and they provide the ability to model indoor data effectively for certain applications. If the building data is already available in IFC or CityGML [7] formats, supporting querying over these formats in the indoor context is still a challenge. Besides, the system must support indoor data at different abstractions like spatial, semantic, geometric, etc., [16] in an integrated manner, so that different application needs can be met. For instance, while topological models are suitable for navigation, a grid-based model is more effective for representing sensor data. Several hybrid models have been proposed specifically in the context of navigation [4, 13, 15]. For supporting heterogeneous applications, it is important that multiple models be integrated into a hierarchical extendible model [10, 16]. Standards like IndoorGML [11] can help provide a starting point for this hierarchy.

Need for Configurable and Queryable Indoor Information Systems Any generic indoor information system must cater to heterogeneous applications [8]. An indoor information system developed for navigation purposes need not work well for a system that monitors a smart building. Therefore, a generic indoor information system architecture is imperative, and it should support the following features

- Configurability: The system should allow for configuring an indoor database
- Flexibility: The system must support different models and standards that are most suitable for a host of different applications
- Queryability: The indoor data must be easily queryable, and query engines that operate over heterogeneous data models must be supported.
- Visualization of Indoor Information: Interfaces must support effective 2D and 3D interactions with the building, and support both query input and visualization of query results in the context of the building.

With that in view, this paper discusses the design and implementation details of a novel toolkit for indoor spaces (InGIST) that is queryable and configurable. InGIST supports a hybrid indoor model that effectively combines spatial, geometric, and topological models, over which a newly developed query language operates to support different types of indoor queries. The system also supports IndoorGML standards and can generate IndoorGML documents from different indoor representations. The overall framework and functionalities are explained. Using sample outputs and use cases, this chapter highlights the salient features of InGIST that addresses the above needs. The next section first describes the indoor modeling adopted in InGIST and then the query engine.

2 Indoor Modeling and Querying in InGIST

The foundation for designing and implementing an indoor information system is a comprehensive indoor data model. InGIST supports a hierarchical hybrid indoor model that can capture different application requirements, support multiple levels of abstractions, and be able to answer different types of queries. To support these requirements, the model effectively combines commonly used spatial representations, grid and topological representations, and semantic representations. The data model includes

1. indoor spatial data as the base layer which consists of the shapefiles of the building plan, and the layers representing the different components of the building like window features, door properties, room properties, etc.
2. a semantic representation which represents the indoor details and objects within the building like tables, shelves, boards, chairs, etc., which is part of an object-relational representation of the indoor space
3. a grid representation for providing the flexibility to represent sensor and object information as the next layer
4. a topological model based on IndoorGML for supporting navigation.

The spatial model and the object-relational representation in PostgreSQL help with the configurability of the system. The other two layers support localization, navigation, and mapping sensor information. To maintain consistency across layers, InGIST includes logical links across the layers. The shapefile features are linked to the corresponding semantic information, the set of grid cells contained within them, and the node representing them. These links are embedded implicitly or captured using relations in the object-relational representation in PostgreSQL. To create the different levels of abstractions and to standardize them, the following functionalities have been added to InGIST.

- The model represents the spatial aspects of the building using shapefiles (defined in different layers) derived from CAD maps
- Functions to automatically map the shapefiles to 2D grid representation for any resolution
- Functions to derive nodes for a topological graph from the grid model that takes into consideration the structural aspects of the building using a region filling algorithm and to create a topological graph using the detected nodes using a graph generation algorithm [2]
- Generation of different topological graphs based on the IndoorGML standards which include adjacency graphs and connectivity graphs.

2.1 Querying Using InGIST

One of the significant features of InGIST is its support for querying over indoor data, and the above model provides the flexibility to support it. A querying framework has been designed to operate on the indoor model in InGIST, which abstracts out the fundamental entities and operators, which are common to multiple models. This framework supports novel SQL-type text-based query operators and an SQL-type query language [1] that was developed as part of the project. The syntax shares similarities with most of the existing query languages. A visual querying component is added over this language to help the user construct queries with much ease and improved comprehension. For using the querying system over this hybrid model, translation modules are designed to translate the input queries to the general-purpose languages supported by the models. It provides support for SQL-type queries, querying over Neo4j graph representations and IndoorGML.

Since the indoor model provides flexibility, different types of queries are supported, which include attribute queries, spatial queries, geometric queries, and navigational queries. Figure 1 shows results of two sample queries. Attribute queries select the indoor entities based on some operations on their attributes. An "attribute query" finds all entities of the specified type whose attributes satisfy the conditions given in the query, and these queries are primarily answered by the data layer represented in PostgreSQL. Spatial queries deal with the spatial characteristics of the entities in the indoor space. Example spatial queries include range queries, nearest neighbor queries, and location queries. These queries are answered using the underlying spatial layer. Geometric queries deal with the geometric attributes of the indoor space entities, and they work based on the relationships between geometries. Example queries include volume queries. Geometric queries deal with the geometric attributes of the indoor space entities, and they work based on the relationships between geometries. Navigation queries help find paths between two indoor locations and are answered using the IndoorGML and grid layer.

The features of this query language and more details are available in [1]. Table 1 gives an overview of the different types of indoor queries and the descriptions of these queries. The goal of this language is to cover the indoor queries irrespective of the

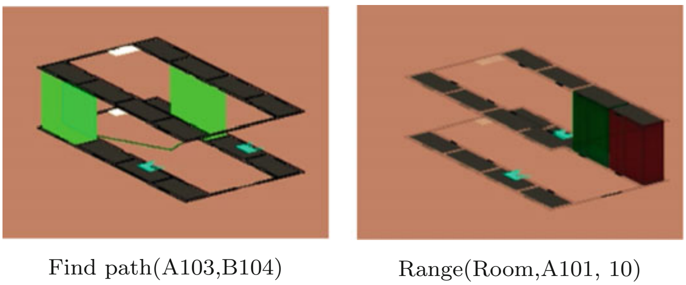


Fig. 1 Sample queries with results

Table 1 Indoor query definitions

Query category	Syntax	Definition	Description
Attribute query	Find E_i where $E_i.Attr_1$ = $value_1$ && $E_i.Attr_2$ = $value_2$	For every E_i , return E_i : { $E_i.Attr_1$ = $value_1$ && $E_i.Attr_2$ = $value_2$ }	Returns all entities of the specified type which satisfy the conditions specified on its attributes
Adjacency query	Adjacent (S_i) where *conditions*	For every space S_j, return S_j, : { S_i adjacent S_j is true && *conditions* met}	Two SPACE entities are adjacent if they share a common boundary (navigable or non navigable) and all the conditions met
Path query	Path (S_{start}, S_{end}/ T_{end}) [not] through E_i where *conditions*	Returns P = { S_{start}, S_2, . . , S_{end} } such that for every S_i, S_i linked NB_i linked S_{i+1}/ T_{end} is true && E_i [not] in P && *conditions* met	Returns the sequence of connected spaces or transitions between the start and end entities, with the conditions met
Range query	Range (Type of entity, S_{origin}, *range value*) where *conditions*	For every E_i , return E_i :{ p = path(S_{origin} to E_i) exists && *conditions* met && length(p) $\leq$ *range value* }	An entity of the type specified selected if there is an accessible path which falls within the specified range meeting the conditions specified
K-nearest neighbor query	Knn (Type of entity, S_{origin}, *k value*) where *conditions*	For every E_i , return E_i :{ p = path(S_{origin} to E_i) exists and E_i is not kth entity && *conditions* met}	Finds the first k entities of a given type at closest navigable distance from the queried entity and meet the conditions specified
Volume	Find Volume(E_i)	Returns Volume(E_i)	Returns the volume of the specified entity calculated based on the geometry

underlying models and to provide a user-friendly querying experience. It is designed so that a single query in this language can replace a set of multiple queries in a model-specific query language. The indoor query language is defined to have an SQL-like syntax, with clauses, predicates, and expressions, since SQL-like statements are more natural to express. Any query in this proposed query language has the format

```
'Find indoor entity where conditions'.
```

The '*Find*' clause contains the indoor domain-specific entities or items to be selected. This is followed by an optional '*Where*' clause in which one or more conditions can be specified similarly to an SQL query. Table 2 shows two queries using this framework and the query output that is visually depicted, thus demonstrating the effectiveness of the querying system. The extension of this query language for IndoorGML documents is explained in Sect. 4.

3 Architecture of InGIST

This section describes the overall architecture of InGIST. The system is designed so that the indoor models can be reflected in the database and back-end file system. The supporting services, which include data representation services and querying services, reside in a service layer which is a configurable plug and play layer. Above these resides the user interface which enables the user to configure the system. The system is designed using open-source tools (PostgreSQL as back-end supported by OSGEO and GDAL, Python and Java for middleware services, and Java 3D and OpenSceneGraph as front-end). The rest of the section explains the main modules of InGIST (Fig. 2).

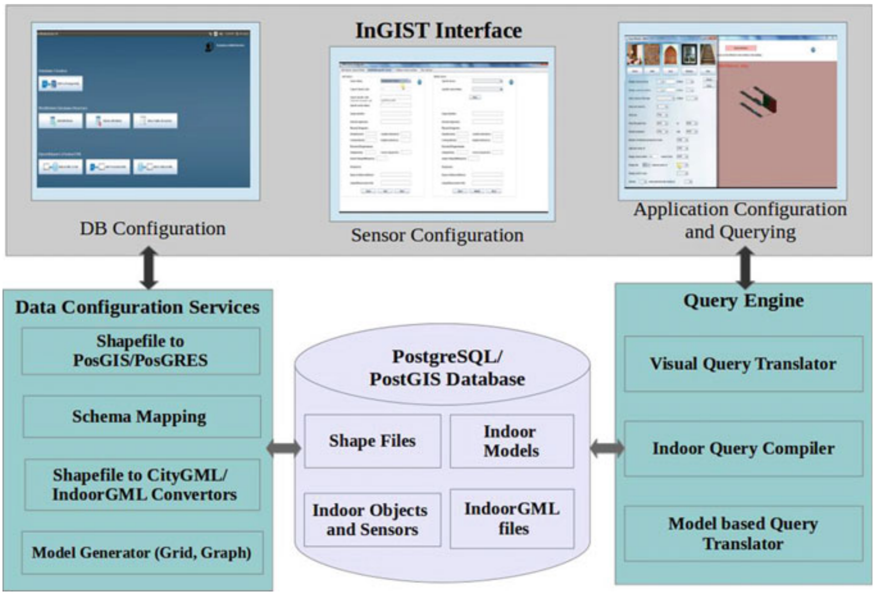


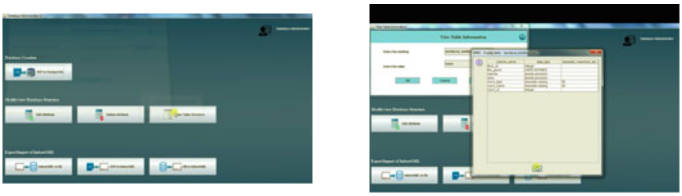
Fig. 2 InGIST architecture

3.1 Indoor Database

The indoor database is implemented over PostgreSQL with PostGIS extensions. This database supports spatial models of buildings (shapefiles of indoor structures like rooms, walls, etc.). In addition to the basic information derived from the CAD files, additional properties like wall materials, door types, room types, etc., have been added either manually using the front-end or using scripts. An object/asset database is also part of InGIST. Assets are classified using a hierarchical classification system, and the assets needed for the application are loaded accordingly. The system supports different types of queries like attribute queries, spatial queries, and geometric queries. The back-end database stores additional grid representation of the floors to support navigational queries.

3.2 Data Configuration Services

One of the primary goals of InGIST is to enable end users to configure any indoor-pervasive system. InGIST helps users to configure a new indoor-pervasive project and configure the database corresponding to the indoor space. InGIST supports data stored in PostGIS, CityGML, BIM/IFC, or IndoorGML. Functionalities for importing and converting between the different formats and standards are provided. InGIST also provides support for translation between different indoor models. For instance, one can generate equivalent topological or grid maps based on the spatial files provided. Grid/graph generators are added to InGIST to create network models for navigability. Figure 3 shows the configuration menus in InGIST, and a sample result after configuration.



Configuration Options

Snapshot of table after updation

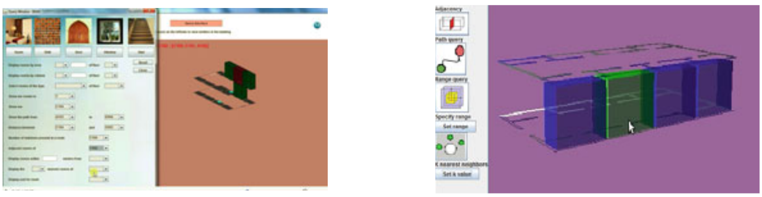
Fig. 3 Data configuration in InGIST

3.3 Query Engine

The query engine operates on a framework that abstracts out the details that are common to most used indoor models to construct a generic representation. Text-based and visual querying languages are designed based on this framework. The working of the system starts with the user building a visual query, which is converted to the text-based query explicitly defined for indoor spaces. This query is then converted to the corresponding query languages like SQL or cypher query language associated with the underlying database. Figure 4 shows the query interfaces in InGIST which has a comprehensive menu-based query interface that can help compose the different types of queries in the indoor context, and a visual query interface with the visual operators [1]. The latter shows the result of applying the knn operator.

3.4 User Interface

An essential goal of this toolkit is to help visualize the indoor space and the entities inside it, and toward this an effective interface is critical. The goal of the user interface is as follows: (1) to visualize the building in both 2D and 3D, (2) to allow the user to configure the building, objects, and their metadata and configure the database, (3) to support different types of queries using user-friendly interface design, and (4) to support effective visualization of the results from the different queries. InGIST supports all of the above and includes (1) a front-end to visualize a building in both 2D and 3D, (2) interface to query the building, (3) an administrator user interface which enables a building manager to configure this system for a building, consisting of simple functionalities, and (4) a barcode/rfid-based object registration interface to register objects in a building into the system. Thus, InGIST allows for different types of users, from database administrators to building managers and end users of various applications to conveniently use the system based on their needs.



Menu-driven Query System

Visual Querying

Fig. 4 Query interfaces

4 IndoorGML Support

IndoorGML [11] (Indoor Geographic Markup Language) is a standard for modeling indoor spaces. It focuses on topological modeling of indoor space for navigation. A key focus in developing InGIST is to contribute to the indoor data standardization process in OGC, specifically to IndoorGML. The key contribution to the IndoorGML standards adoption process has been the support for IndoorGML in InGIST. InGIST supports different aspects of configuring and editing IndoorGML files. To the author's best knowledge, there is only one other tool which is developed by authors of the IndoorGML Editor [9]. However, while the existing tool supports editing IndoorGML files, and configuring the graphs, InGIST supports more functionalities as given in the following paragraphs.

4.1 Creating and Configuring IndoorGML Files

It is important that any indoor representation is standardized based on the standards specified by IndoorGML. IndoorGML standard focuses on the modeling of indoor space, with primary importance to navigation. The standards document provided by OGC are used as reference for configuring the indoor space of any building. InGIST provides the ability to obtain the indoor data from the shapefiles and create the IndoorGML document. This document contains the details of the rooms, walls, and doors, conforming to the standards prescribed by the IndoorGML Standard Specification document. By providing this important functionality, any indoor building represented using shapefiles can be standardized to the IndoorGML specification.

4.2 Support Translation of Different Models to IndoorGML

To support different applications, InGIST supports different models. The data for any building is derived from multiple models, namely spatial model (from shapefiles), geometric model, and also from databases. Hence InGIST supports translation from different models to the IndoorGML specification. Translator modules that translate from shapefiles, PostgreSQL/PostGIS database, and CityGML to IndoorGML are integrated within the system. Any indoor data represented in any of these methods can be converted to IndoorGML. The generated IndoorGML has been validated using schema validation. Figure 3 shows the options provided in InGIST for IndoorGML support. InGIST also supports export of data in an IndoorGML format to any of these representations.

4.3 Topological Modeling (with Conformance to IndoorGML)

One of the most important requirements for navigation inside the building is the generation of the topological model. A topological model represents the various connections of an indoor space within the building over which navigation algorithms can be easily implemented for indoor space. The topological model generated must also conform to the IndoorGML specification. In the IndoorGML specification, there are different graphs that represent various relationships between the nodes such as “Adjacent,” “Connects,” “Overlaps,” “Contains,” and “Related to.” InGIST is designed to cater to this need for creating different topological representations of indoor space. Besides, the visualization of the graphs generated for IndoorGML Instance Document is also possible using the IndoorGML Configuration tool. The visualization allows the user to identify the connections within the building and thereby give the user the possibility of identifying the nearby exits and other critical points of navigation within the building, which can be used during emergency situations. All the different types of graphs are also visualized using the system. The tool can generate the following graph representations and visualize them.

- Adjacency Graph focuses on the adjacency relationship of entities in indoor space. For instance, two rooms are adjacent if they share a common wall or door.
- Connectivity Graph primarily focuses on the navigation relationship of entities. Here two rooms are connected if they have a door or corridor connecting them.
- Relationship Graph addresses the relationship between entities of indoor space, like a relationship between a room and its bounding walls. Also, the transitions between the different graphs are also generated by the tool. To improve the accuracy of the representation, subspacing of open spaces is done.

Figure 5 shows the visualization of these graphs for a sample building in InGIST. A 30-floor building was created to study the scalability of the IndoorGML standard, and the resultant document was tested for IndoorGML Instance Document creation and graph representation issues. We have observed that even large buildings can be

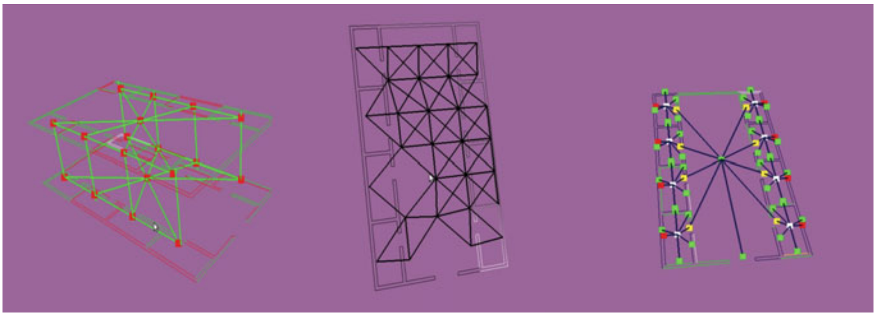


Fig. 5 Sample IndoorGML graphs in InGIST

represented efficiently using the IndoorGML representation. However, the complexity of storing different graphs must be explored and is part of ongoing work.

4.4 Support Querying over IndoorGML

An important feature of this tool is support for querying over IndoorGML format. Based on the IndoorGML document created, the design for a query tool over IndoorGML was conceived. A fully developed tool that performs a variety of queries like attribute, geometric, spatial queries have been developed. Since IndoorGML uses the topological model, graph representation and graph algorithms are used for the query processing. Our tool is designed to provide accurate results to these above-mentioned queries, and they are visualized over the topological model, thereby displaying the results over the graphs. The indoor query system as mentioned earlier supports querying over IndoorGML data. There are two ways to support this. One way is to store the topological representations in Neo4j and query over it as discussed in [1]. The indoor query language is extended to support querying over the IndoorGML document directly, and it has an XQuery-like syntax and is similar to GQuery [5]. It has the following syntax:

```
'For pathExpression Where conditionPath Return entity/function'.
```

The '*For*' clause contains the path expression which includes the path based on the IndoorGML schema. The '*Where*' clause is optional which contains one or more conditions to extract the result based on the given criteria. The '*return*' clause contains the indoor function or the path of the indoor entity to be returned. Table 2 shows the corresponding language to support querying over IndoorGML documents.

A few sample queries are provided to provide a better understanding of how this query language is defined. For example, the following query is used to display the adjacent rooms of 'B101':

Table 2 IndoorGML Query Syntax

Function	Indoor query syntax	IndoorGML query syntax
Adjacent	**find** adjacent(space)	**for** pathExpression **return** adjacent(space)
Knn	**find** knn(entity,space, kvalue)	**for** pathExpression **return** knn(space, kvalue)
Range	**find** range(entity, space, rangeValue)	**for** pathExpression **return** range(space, rangeValue)
Path	**find** path(entity, entity)	**for** pathExpression **return** path(entity, entity)

```
'For \$x in doc(IndoorGML.xml) /IndoorFeatures//CellSpace
Where \$x/gml:name=B101 Return adjacent(\$x)'.
```

The range query and k-nearest neighbor query are computed based on the navigable distances. For example, the following query computes the rooms that lie within the specified range.

```
'For \$x in doc(IndoorGML.xml) /IndoorFeatures//CellSpace
Where \$x/gml:name=B101 Return range(\$x,100)'.
```

By providing these functionalities, InGIST serves as a comprehensive framework for creating, editing, exporting, and visualizing IndoorGML documents.

5 Conclusion and Future Research Directions

This paper discussed InGIST, a novel queryable and configurable IndoorGIS framework. The paper discussed the indoor modeling concepts that aid both configurability and querying and a novel indoor query language that supports different types of indoor queries. Finally, it discussed how InGIST supports IndoorGML standardization by providing functionalities for creating, exporting to IndoorGML documents, visualizing them and querying over them. InGIST has been demonstrated to support different types of queries, and hence its robustness in supporting different applications is evident. The different query results and screenshots demonstrate its utility and application developers can use InGIST to configure indoor information systems for any building easily. As part of ongoing work, support for moving object modeling and querying over moving objects is being added to InGIST, which will help make InGIST suitable for real-time indoor-pervasive applications.

References

1. Amrutha, H., Balasubramanian, V.: A model independent and user-friendly querying system for indoor spaces. In: Proceedings of the 20th International Conference on Management of Data, COMAD '14, pp. 17–28. Computer Society of India, Mumbai, India (2014)
2. Balasubramanian, V.: Supporting scalable activity modeling in simulators. Ph.D. thesis, Irvine, CA, USA (2009). AAI3347827
3. Balasubramanian, V., Kalashnikov, D.V., Mehrotra, S., Venkatasubramanian, N.: Efficient and scalable multi-geography route planning. In: Proceedings of the 13th International Conference on Extending Database Technology, EDBT '10, pp. 394–405. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1739041.1739090
4. Bandera, A., Urdiales, C., Sandoval, F.: An hierarchical approach to grid-based and topological maps integration for autonomous indoor navigation. In: Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2001, vol. 2, pp. 883–888 (2001)

5. Boucelma, O., marie Colonna, F.: Gquery: a query language for gml. In: 24th Urban Data Management Symposium, UDMS 2004. Chioggia (Italie), 27–29 October 2004, pp. 23–32 (2004)
6. CAD to GIS: a step-by-step guide to converting .dwg CAD files to GIS shapefiles (2005)
7. Gröger, G., Kolbe, T.H., Czerwinski, A., Nagel, C.: Opengis® city geography markup language (citygml) encoding standard (2008). OGC 08-007r1
8. Hightower, J., Borriello, G.: Location systems for ubiquitous computing. Computer **34**, 57–66 (2001). https://doi.org/10.1109/2.940014
9. Hwang, J.R., Kang, H.Y., Choi, J.w.: Development of an editor and a viewer for indoorgml. In: Proceedings of the Fourth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, ISA '12, pp. 37–40. ACM, New York, NY, USA (2012). https://doi.org/10.1145/2442616.2442625
10. Imad Afyouni, C.R., Claramunt, C.: Spatial models for context-aware indoor navigation systems: A survey. J. Spat. Inf. Sci. **4**(1), 85–123 (2012)
11. Ki-Joune Li, J.L., S.Z.T.H.K.C.N.T.B.: OGC ® indoorgml (2013). OGC 13-nnnrx
12. Li, K.J., Lee, J.: Indoor spatial awareness initiative and standard for indoor spatial data. In: Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010) (2015)
13. Lorenz, B., Ohlbach, H.J., Stoffel, E.P.: A hybrid spatial model for representing indoor environments. In: Proceedings of the 6th International Conference on Web and Wireless Geographical Information Systems, W2GIS'06, pp. 102–112. Springer-Verlag, Berlin, Heidelberg (2006)
14. Sobika, S., Rajathilagam, B.: Intelligent System for Wayfinding through Unknown Complex Indoor Environment. Springer International Publishing (2016)
15. Thrun, S., Bü, A.: Integrating grid-based and topological maps for mobile robot navigation. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence, vol. 2, AAAI'96, pp. 944–950. AAAI Press (1996). http://dl.acm.org/citation.cfm?id=1864519.1864527
16. Worboys, M.: Modeling indoor space. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, ISA '11, pp. 1–6. ACM, New York, NY, USA (2011). https://doi.org/10.1145/2077357.2077358

Indoor Evacuation Planning

A. Bhushan and N. L. Sarda

1 Introduction

Evacuation path planning, one of the most important steps in evacuation planning of a building, schedules the movement of evacuees to move them out of a building in minimal time in case of a disaster such as fire or explosions. In the literature, the problem has been formulated as a network flow problem [2, 6] and has been solved using linear programming method. Various mathematical models with different objective functions for building evacuation are given in [8]. These models give the optimal solution but cannot be used for large sized networks such as shopping malls and airports due to the large time complexity. However, capacity-constrained routing planner (CCRP) [12, 17], one of the most popular heuristics, is suitable for finding efficient evacuation plans on large sized networks. The evacuation plans computed by CCRP are close to optimal. The CCRP has been modified further to improve its run time complexity to handle very large network [7, 9, 15, 16] and large number of evacuees [13]. The CCRP has also been used in solving other variants of evacuation path planning problem [4, 10, 14].

Most of these existing approaches consider the following assumptions: (i) All evacuees are fit to evacuate by themselves, and (ii) the existing exits are sufficient to evacuate evacuees from a building. However, some people may be injured due to fire or explosion, or they are physically unfit to evacuate by themselves. In such scenarios, rescuers are needed to help these injured evacuees in evacuating the building. In such case, unlike the unidirectional flow during evacuation time, there is a bidirectional flow inside the building where rescuers move inside toward the injured evacuees to rescue them and fit evacuees move outside toward exits. In such case, the above methods cannot be used to plan evacuation. Next, sometimes, emergency exits are blocked due to some reasons and the existing exits are not sufficient for evacuees to

A. Bhushan (✉) · N. L. Sarda
GISE Lab, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai, India
e-mail: alka.bhushan@yahoo.com

N. L. Sarda
e-mail: nls@cse.iitb.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_9

evacuate a building before it collapses. In such circumstances, use of ladders has been a standard practice to reduce evacuation time. However, ladders are often limited in number and thus need to be placed optimally to maximize the utilization of these ladders. Evacuation planning with ladders has been discussed previously in [3, 4].

There are various evacuation computer models being used in practice for simulating a building evacuation. A review of all these models is given in [11]. All of these models either require preconstructed network model of a building or require CAD files. None of these models provide the following functionalities in a single model: (i) create network model using floor layout images, (ii) run multiple evacuation planning algorithms, (iii) visualize evacuation paths between each source node and each destination node, and (iv) perform sensitivity analysis of each plan. The evacuation time of a building can be assessed by considering a set of different distributions of evacuees. Such type of analysis helps in setting lower bounds on evacuation egress time required to evacuate a building in a typical scenario and helps in both the re-design to improve evacuation time as well as creating more optimal allocations of building spaces.

In this chapter, our contributions are as follows: (i) A heuristic method based on CCRP to evacuate different types of people is presented, and (ii) a user-friendly interface is presented to build network using floor layout images, create different distributions, visualize the evacuation plans computed for different situations, and perform sensitivity analysis.

The chapter is organized as follows: Sect. 2 presents modeling of a building, and the problem statement is given in Sect. 3. The heuristic methods used in computing evacuation plans for different types of scenarios are given in Sect. 4, 5, and 6. Further, evacuation planner is described in Sect. 7. Finally, the chapter is concluded in Sect. 8.

2 Modeling of a Building

Similar to [3, 5], a building is modeled as a graph, where nodes represent rooms, corridors, lobbies, and intersection points, while edges represent stairs and hallways. Each *node i* has maximum capacity which is the maximum number of people who can stay at the node at any point of time. Each *edge* is associated with intake capacity and travel time where intake capacity is the maximum number of people who can enter the edge per unit of time and travel time is the time required by an evacuee to travel from one end of the edge to the other end. We consider the following three different types of people inside a building: (i) *fit evacuees* who are able to evacuate by themselves, (ii) *injured evacuees* who are not fit and need help from rescuers in moving out of a building, and (iii) *rescuers* who are assumed to be trained in evacuating people from a building. Without loss of generality, we assume that each injured evacuee requires one rescuer while evacuating a building. Every person inside a building, either a fit evacuee, or an injured evacuee or a rescuer, requires 1 unit of space inside the building.

For evacuation, evacuees need to be scheduled from their current location to an exit. However, a rescuer needs to be scheduled from his current location, i.e., an exit or a rescue location to a source location of an injured person. In such scenario, a path used by rescuers to move from an exit to a source location may be used by evacuees to move from their location to the exit. In such case, evacuees and rescuers may cross each other on an edge at some point of time. Instead of changing their speed in bidirectional movement, we assume that a person can travel with the same speed if proper space is given. Thus, an impedance factor α is introduced for bidirectional flow that will reduce the available capacity to use more space by people in the opposite direction. Thus, total intake capacity at any point of time during bidirectional movement on an edge must be at most α times the maximum intake capacity, where $0 < \alpha \leq 1$. Similar to previous work, we consider the FIFO network that follows first in, first out strategy.

A place like window or balcony where a ladder can be placed is marked as a ladder point. A ladder placed at a ladder point can be used in moving the evacuees out of the building. A ladder is considered as an edge connecting the ladder point to the ground. In addition to intake capacity and travel time, it has ladder load which is a constraint on how many people can be present on a ladder at any point of time. A ladder cannot be placed instantly and thus requires initial placement time. See [3] for more details on modeling of a ladder.

Evacuation Scenarios

An evacuation scenario is created by assigning fit and injured evacuees to a set of nodes, assigning rescuers to a rescuer node, and assigning few nodes as exit nodes where evacuees can stay for a long period of time.

3 Problem Statement

The following two different types of problems are being discussed in this chapter.

3.1 Evacuation Using Rescuers

For a given capacitated network of a building with node set N, edge set E, source nodes $S \subset N$, rescuer nodes $R \subset N$, and destination nodes $D \subset N$, the problem is to compute an evacuation plan that obtains a set of source–destination routes (indicating paths to be followed by the evacuees and the rescuers), and a schedule of evacuees and rescuers on each route. A schedule gives start time, end time, and the number of people on each path along with start time and wait time at each node of the path. The objective is to minimize the evacuation egress time. The source nodes contain either injured evacuees or fit evacuees.

3.2 Evacuation Using Ladders

For a given capacitated network of a building with node set N, edge set E, source nodes $S \subset N$, ladder points $P \subset N$, destination nodes $D \subset N$, and L number of available ladders, the problem is to select ladder points to place L ladders and compute an evacuation plan using the ladders placed at selected ladder points. The objective is to minimize the evacuation egress time. The source nodes contain only fit evacuees.

4 CCRP Algorithm [12]

The CCRP algorithm considers only fit evacuees. Each source node is connected to a supersource node by an edge of zero travel time and intake capacity equal to the initial occupancy of the source node. The occupancy of the supersource node is equal to the total number of evacuees in source nodes. Iteratively, the following steps are executed until no evacuee is left to get a path:

1. Compute a shortest path p with time schedule using a generalized shortest path algorithm given in [12] from the supersource node to an exit node such that the path has nonzero available capacity and the path is smallest among all the shortest paths of nonzero available capacity between a source node and a destination node.
2. Compute the minimum of maximum available capacity of the path p and the remaining number of evacuees.
3. Reserve the capacity of the path for these evacuees.

5 Extended CCRP for Computing Evacuation Plans Using Rescuers

The CCRP algorithm is modified to incorporate bidirectional movement on an edge and different types of flows in the graph. Modification for bidirectional movement is given next.

5.1 Generalized Shortest Path Algorithm for Bidirectional Movement on an Edge

The CCRP algorithm assumes that the flow is always from a source node to a destination node and thus does not consider any constraint for bidirectional movement on an edge. However, the bidirectional movement is possible when the flow is from both the directions between a pair of nodes. For example, if people equal to edge intake capacity are scheduled to travel on an edge at time instance t, then again the same number of people can be scheduled in the same direction to travel on the edge at time instance $t + 1$. But, these people cannot be scheduled to travel on the same edge in the opposite direction at time instance $t + 1$ until travel time of the edge is 1 because these evacuees will meet at some point of time on the edge and the total number of people at that time will exceed intake capacity of the edge. We add this constraint in the generalized shortest path algorithm as follows: A time series array is maintained for intake capacity of an edge. We add the direction along with the available capacity at each time instant and put a constraint that if an edge is used by x people at time t, then the edge in the opposite direction can be used by at most $\alpha(c - x)$ people between time interval $[t, t + \lambda - 1]$ where c is the edge intake capacity, α is the impedance factor, and λ is the edge travel time.

5.2 Method for Computing Evacuation Paths

Here, we extend the CCRP algorithm to incorporate injured evacuees and rescuers as follows: Let F and I be the set of source nodes of fit and injured evacuees, respectively, and R be the set of rescuer nodes. All source nodes of fit evacuees are connected to a supernode, denoted by *fit supernode*. The fit supernode contains the initial occupancy equal to the total number of initial occupancy of fit evacuees at the source nodes, and each edge connected to the supernode contains the travel time equal to 0 and intake capacity equal to the initial occupancy of fit evacuees at the adjacent source node. Similarly, all source nodes of injured evacuees are connected to a supernode, denoted by *injured supernode*, and the initial occupancy of the supernode is the total occupancy of injured evacuees at the source nodes.

Initially, rescuers are available at the rescuer nodes. But later, since rescuers will take the injured evacuees to an exit node, therefore rescuers will be available at the same exit node to help injured people. Thus, new rescuer nodes will be added as they escort injured evacuees to an exit node and travel inside again from that exit node to another node $i \in I$. The availability of a rescuer at a rescuer node and at a source node of injured evacuees will vary with time.

Our method iterates over the following steps until no evacuee is left in F and I for allocating a path:

1. If F contains a node of nonzero occupancy, then compute a shortest path $p1$ with time schedule using our generalized shortest path algorithm from the fit supernode to a destination node in D such that the path $p1$ has nonzero available capacity. The computed path $p1$ is smallest among all the shortest paths between all pair of nodes (i, j), where $i \in F$ and $j \in D$.
2. If a rescuer is available at node i in I at time t, then compute a shortest path $p2$ with time schedule from the node i to a destination node in D such that the path has nonzero available capacity and the path is smallest among all the shortest

paths computed between all pair of nodes (i, j), where $i \in I$, and node i has a rescuer and $j \in D$. The start time of the path $p2$ is greater than or equal to t.

3. If R is not empty, then compute a shortest path $p3$ with the time schedule from a node i in R to a source node in I such that the path $p3$ has nonzero available capacity. The path $p3$ is smallest among all the shortest paths between all pair of nodes (i, j), where $i \in R$ and $j \in I$, and node j has some injured evacuees.
4. Select the shortest path among $p1$, $p2$, $p3$.
5. If path $p1$ is selected, then
   - The number of evacuees who can be scheduled on the path is equal to the minimum of available path capacity and available number of evacuees at the corresponding source node.
   - Reserve the path for these evacuees by reducing the available capacity of the nodes and edges of the path at the corresponding time.
   - Reduce the initial occupancy of the source node and the supernode by these number of evacuees.
6. If path $p2$ is selected, then
   - The number of injured evacuees paired with the rescuers who can be scheduled on the path $p2$ is equal to the minimum of twice the number of rescuers available at the source node and the available path capacity.
   - Reserve the path for the number of these injured evacuees paired with the rescuers by reducing the available capacity of the nodes and edges of the path at the corresponding time.
   - Reduce the initial occupancy of the corresponding source node in I by the number of injured evacuees scheduled on this path.
   - If the destination of the path $p2$ is not in R, then add the node in R along with the number of rescuers available at the time equal to the reaching time of path $p2$.
7. If path $p3$ is selected, then
   - The number of rescuers who can be scheduled on the path $p3$ is equal to the minimum of available path capacity, the number of rescuers available at the source node of $p3$, and the number of injured evacuees present at the destination node of $p3$.
   - Reserve the path for these rescuers by reducing the available capacity of the nodes and edges of the path at the corresponding time.
   - Reduce the available number of rescuers at the source node of $p3$ at the corresponding time by the number of rescuers scheduled on this path.
   - Add these number of rescuers at the destination node of $p3$ with time equal to the reaching time of path $p3$.

In this method, we use greedy approach to allocate a rescuer to an injured person where a rescuer is allocated to a nearest injured person. The aim of the algorithm is to send the rescuers to the help locations as soon as possible without affecting much the movements of fit people.

Run Time Complexity:

In every iteration of the algorithm, shortest paths are computed. For each injured person, shortest paths are computed twice, where one path is for a rescuer to reach from its source node to the source node of the injured person, and another path is to reach an exit node from the source node of the injured person. Thus, one shortest path computation is required for each fit person and 2 shortest path computation is required for each injured person. One execution of shortest path algorithm takes $O(n \log n)$ time, where n is the number of nodes in a graph. For q number of total evacuees, the run time complexity of our algorithm is $O(qn \log n)$, which is the same as the complexity of the CCRP algorithm.

6 Extended CCRP for Computing Evacuation Plans Using Ladders

Since ladders are limited in number compared to available ladder points to place the ladders, therefore the first task is to select ladder points among all the ladder points to place the given number of ladders. Then, ladders are added as edges whose one end point is a selected ladder point and the other end point is an exit point. The selection of ladder points for ladder placement is based on the distribution of evacuees at the time of evacuation. For ladder point selection, we compute the density of each ladder point and the ladder points of largest density values are selected for placing ladders. The density of each ladder point is calculated using the following method given in [4]: (i) For each pair of source node and ladder point, shortest travel time is computed and multiplicative inverse of travel time is denoted to densityFactor, and (ii) for each ladder point, density is calculated as the sum of fraction of number of evacuees at each source node where fraction is equal to densityFactor between the source node and the ladder point divided by the total sum of densityFactor values between the source node and all ladder points.

Further, the CCRP algorithm is modified as follows to handle ladder edges in computing evacuation plans. The constraint of ladder load is implemented by adding a time window of length equal to its travel time. When evacuees enter into the ladder edge, a time window is open that allows evacuees to enter on the ladder edge during the window until number of evacuees are smaller than the load.

Run Time Complexity:

The run time complexity of the CCRP algorithm is $O(qn \log n)$, where q is the number of evacuees and n is the number of nodes in the graph. In our algorithm, density of each ladder point requires shortest travel time between each pair of source and ladder point which is computed by running shortest path algorithm from each source node. This step requires $O(qn \log n)$ time. The remaining steps in density computation require $O(qn)$ time. Thus, run time complexity of the algorithm is the same as the complexity of the CCRP algorithm.

7 Evacuation Planner

All evacuation planning algorithms including the above-mentioned algorithms require indoor network and scenarios to analyze a building for evacuation planning purposes. In this section, we present our evacuation planner that provides a user interface to model building evacuations with the following functionalities: (i) create building network using floor layout images, (ii) create multiple scenarios of a building, (iii) run the above-mentioned evacuation planning algorithms to compute evacuation plans and visualize the computed evacuation plans by seeing the summary of the output and paths along with the group sizes and schedule, and (iv) perform sensitivity analysis of a plan using simulation. We have used JSP-based GUI and Java to implement the evacuation planner. Our planner is available at [1]. Each user needs to create own id to access the planner. A logged in user can use the above-mentioned functionalities. The description of each functionality is given in subsequent sections.

7.1 Building Network

A user creates a building by entering name of the building, number of floors, a small description about the building, and an image of the building. To create network, floor images need to be added in the building. A graph on a floor can be added using mouse and click functions. Node capacity, node name, ladder point, and ladder characteristics can be given to each node by filling up a form available through a click on the node. If a node is a stair node, then the option is given to connect it to the respective stair node in the lower floor. Similarly, edge name, edge capacity, and edge length can be given by filling up a form available through clicking on an edge. A created graph on a floor image is shown in Fig. 1. In the figure, all nodes and edges are shown with red and black color, respectively. However, a node denoted as a ladder point is shown in blue color.

7.2 Scenarios

For a building, multiple scenarios can be created. A new scenario can be added by filling up the name of scenario, speed of people, number of available ladders, speed on ladder, and description of the scenario in a form available by clicking on a new scenario. After this, floorwise graphs are available to add evacuees and rescuers to nodes and to add exit points. All exit points are shown with green color, and source nodes are shown with purple color. Node information can be seen by hovering mouse over a node. Scenario graph of a floor is shown in Fig. 2.

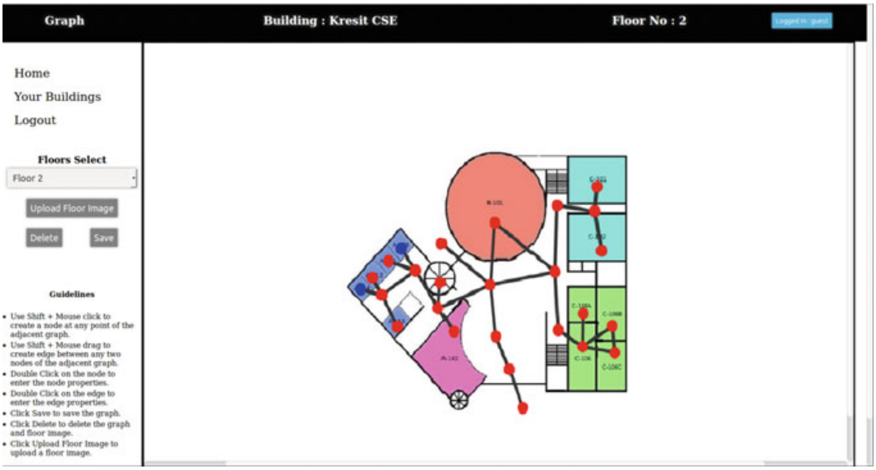


Fig. 1 Graph created for a floor

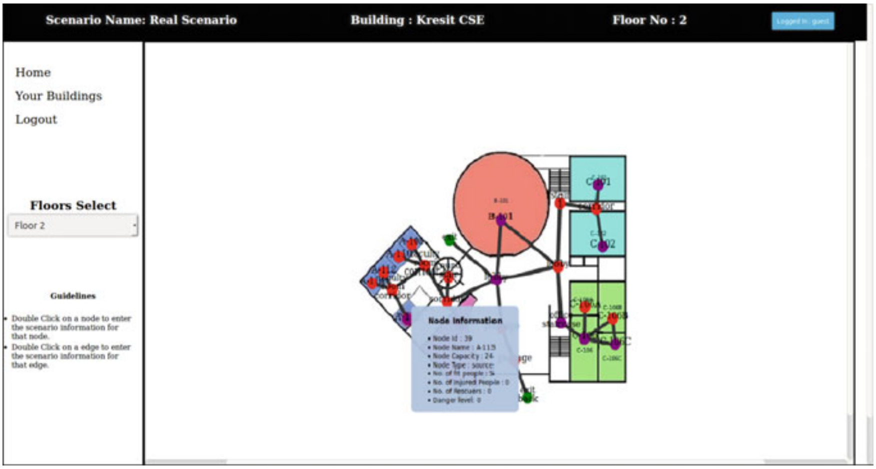


Fig. 2 View scenario of a floor. Purple colored nodes are source nodes, and green colored nodes are exits

7.3 Evacuation Plans

Our planner provides the following three algorithms to compute evacuation plans: (i) CCRP algorithm [12], (ii) evacuation plans using rescuers (Sect. 5), and (iii) evacuation using ladders [4] (Sect. 6). Based on a scenario, we can select any one of the algorithm to compute an evacuation plan. The following summary of the output is shown on page as given in Fig. 3: egress time of fit and injured people, average

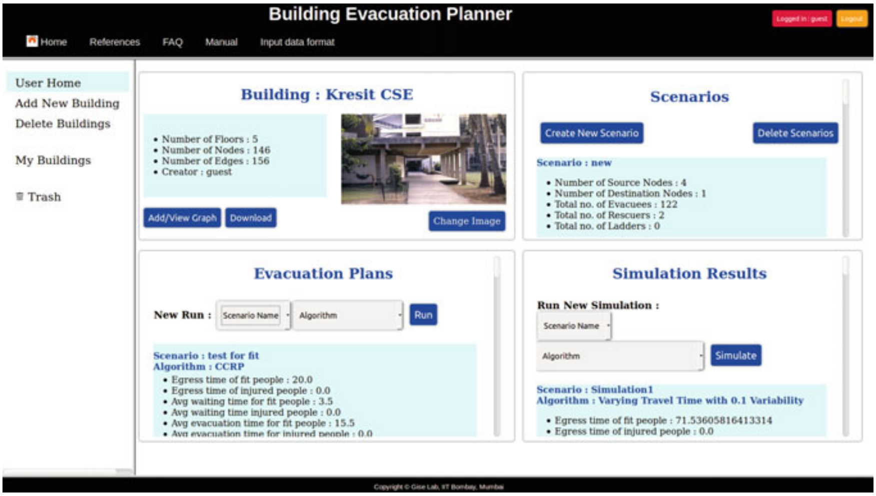


Fig. 3 Run evacuation plan

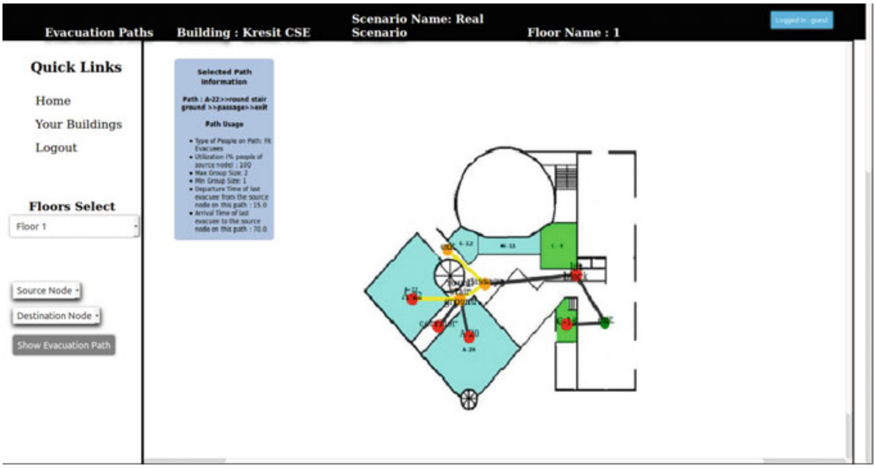


Fig. 4 View path between a source node and a destination node. Path is shown by yellow colored nodes and edges

waiting time and evacuation time of fit and injured people, and maximum waiting time and evacuation time of fit and injured people. By clicking on a scenario name, paths can be seen floorwise as shown in Fig. 4.

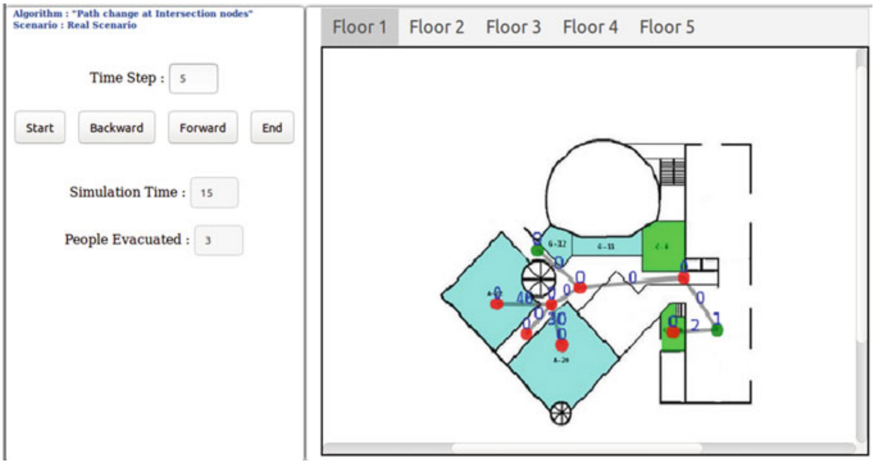


Fig. 5 View people on edges at each time step

7.4 Sensitivity of Evacuation Plans to Human Behavior

Our planner provides feature to analyze sensitivity of an evacuation plan to human behavior by computing the bottleneck nodes and edges, and change in evacuation time for the following human behavior: (i) All evacuees do not walk with the same speed but follow their paths, and (ii) all evacuees do not prefer to wait at the intersection and use an adjacent edge if the capacity is not full. Thus, the inputs for sensitivity analysis are evacuation plans generated by our planning algorithms and the sensitivity methods. The output contains the paths followed by each evacuee and the overcrowded nodes and links due to change in evacuees' behavior. SimJava library is used for the implementation of these methods.

The output summary contains the egress time of fit and injured people, average waiting time and evacuation time of fit and injured people, and maximum waiting time and evacuation time of fit and injured people. At each time unit, the evacuees on the nodes and edges can be seen on the graph as shown in Fig. 5.

8 Conclusion

Heuristic algorithms for evacuation planning have been proposed in the literature to compute the evacuation plans of large buildings in real time. Most of the existing algorithms assume that all evacuees can evacuate the buildings on their own. However, during disaster, some of the evacuees may get injured and may not be able to evacuate by themselves. In such scenarios, rescuers are needed to help injured evacuees. In this work, we modeled the evacuation planning problem for different types of evacuees inside a building. To solve this problem, an extension of CCRP was proposed to schedule the paths for fit evacuees, injured evacuees, and rescuers concurrently so as to achieve fastest evacuation for both fit and injured evacuees. Further, a user interface is provided to build network of a building using floor images and use the network for creating scenarios. Algorithms are provided to compute the evacuation plans of these scenarios. The summary of the results and paths can be seen by using our interface. Sensitivity of evacuation plans to human behavior can also be computed using our planner. Our Web interface is user-specific. A user first needs to create user id to access the interface. In this work, we assume that the number of evacuees at each node is counted correctly. However, the position of an evacuee may not be accurate. In such a case, including the uncertainty in the scenarios and designing a heuristic algorithm for computing evacuation plans could be an interesting future work.

References

1. http://www.gise.cse.iitb.ac.in/Evacuation/
2. Ahuja, R., Magnanti, T., Orlin, J.: Network Flows: Theory, Algorithms, and Applications. Prentice Hall (1993)
3. Bhushan, A., Sarda, N.: Modeling of builing evacuation using ladders. Fire Saf. J. **55**, 126–138 (2013)
4. Bhushan, A., Sarda, N., Reddy, P.: Evacuation planning of large buildings using ladders. In: Database Expert Systems Applications, vol. 7446, pp. 71–85. Springer (2012)
5. Chalmet, L., Francis, R., Saunders, P.: Network models for building evacuation. Manag. Sci. **18**(1), 86–105 (1982)
6. Ford, L., Fulkerson, D.: Flows in Network. Princeton University Press (1962)
7. Gupta, A., Sarda, N.: Efficient evacuation planning for large cities. In: Database Expert Systems Applications, vol. 8644, pp. 201–225. Springer (2014)
8. Hamacher, H., Tjandra, S.: Mathematical modeling of evacuation problems: a state of the art. In: Pedestrian and Evacuation, Dynamics, pp. 227–266 (2002)
9. Kim, S., George, B., Shekhar, S.: Evacuation route planning: scalable heuristics. In: 15th Annual ACM International Symposium on Advances in Geographic Information Systems, pp. 20:1–20:8. ACM (2007)
10. Kim, S., Shekhar, S., Min, M.: Contraflow transportation network reconfiguration for evacuation route planning. IEEE Trans. Knowl. Data Eng. **20**(8), 1115–1129 (2008)
11. Kuligowski, E., Peacock, R., Hoskins, B.: A Review of Building Evacuation Models, 2nd edn. Technical report, National Institute of Standards and Technology, NIST TN 1680 (2010)
12. Lu, Q., George, B., Shekhar, S.: Capacity constrained routing algorithms for evacuation planning: a summary of results. In: Advances in Spatial and Temporal Databases, vol. 3633, pp. 291–307. Springer (2005)
13. Min, M., Lee, J., Lim, S.: Effective evacuation route planning algorithms for updating earliest arrival time of multiple paths. In: Proceedings of the Third ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, pp. 8–17. ACM (2014)
14. Nallagalva, A., Sarda, N., Bhushan, A.: Indoor evacuation planning using a limited number of sensors. In: EM-GIS, ACM SIGSPATIAL (2015)

15. Yin, D.: A scalable heuristic for evacuation planning in large road network. In: IWCTS, pp. 19–24. ACM (2009)
16. Zeng, M., Wang, C.: Evacuation route planning algorithm: longer route preferential. In: ISNN, vol. 5551, pp. 1062–1071. Springer (2009)
17. Zhou, X., George, B., Kim, S., Wolff, J., Lu, Q., Shekhar, S.: Evacuation planning: a spatial network database approach. IEEE Data Eng. Bull. (Special Issue on Spatio-temporal databases), **33**(2), 26–31 (2010)

A Volunteered Geographic Information (VGI) Framework for Disaster Management Based on Mobile Services and SDI

A. Dasgupta and S. K. Ghosh

1 Introduction

The volunteered geographic information (VGI) is referred as such type information which is generated by the citizens without any expertise in geospatial domain [1]. It may be used as a precious information resource for emergency decision-making in disaster management. It can be obtained in real time from the local people of disaster area. When the traditional methods such as satellite imagery and hydrological sensors unable to provide information, then it can act as a viable supplement.

In developing country, it is a challenging task to get large-scale map comparatively for the small disaster regions, especially in rural area. The spatial data infrastructure (SDI) can be used as an authoritative data repositories. Since there is a need to integrate different heterogeneous datasets for emergency management, the SDI may act as a significant component for such integration and generation of real-time disaster maps. Therefore, the SDI can be used as a temporary repository to store data provided by the local volunteers. The Open Geospatial Consortium (OGC) provides a set of standards to access data from SDI [2]. It enhances the potentiality of developing real-time risk map for generating rescue route. In the present work, the SDI concept has been utilized to produce a flood risk map with different risk value by using data provided by local volunteers. The ant colony optimization (ACO) technique has been used to get best rescue route map.

A. Dasgupta
Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India
e-mail: adgkgp@gmail.com

S. K. Ghosh (✉)
Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India
e-mail: skg@cse.iitkgp.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_10

2 Related Works

In emergency decision-making, a large volume of updated information should be needed [3]. There are some research work has been carried out for disaster management with SDI. Mansourian et al. [4] showed the role of SDI in disaster management. A flexible and expandable architecture for flood forecasting with the SDI has been developed by Akinci et al. [5]. Koswatte et al. [6] have explored the issues of using crowdsourced information for disaster management. A technique has been developed by Liu et al. [8] for developing crisis map based on map mash-up. Schnebele et al. [7] developed a method for flood damage assessment in the New York City area following Hurricane Sandy. After the Haiti earthquake, [9] developed a system which helps to create crisis map based on the information provided by the volunteers. Erskine et al. [10] developed a mobile-based VGI framework for creating relevant map for the disaster area.

Recently, the swarm intelligence algorithms have been explored in various research to find suitable evacuation route. Zheng et al. [11] have used a multiobjective PSO approach to generate evacuation routes. In the work of Duan et al. [12], an application based on ACO algorithm has been developed for evacuation people from a large building.

3 Integration of VGI Platform with SDI

In this work, a framework has been developed which estimates the flood inundation extent by using the information from the local volunteers of the susceptible area. In Fig. 1, the components for integration of the VGI platform with the SDI have been illustrated.

3.1 VGI Data Collector

The purpose of this component is to communicate between the volunteers and the orchestration engine for flood data collection by using GeoSMS. The Open Geospatial Consortium (OGC) has proposed a standard for developing interface for GeoSMS [13]. It is a mobile application that enables the volunteers to enter their observed values regarding the current environment in the survey form.

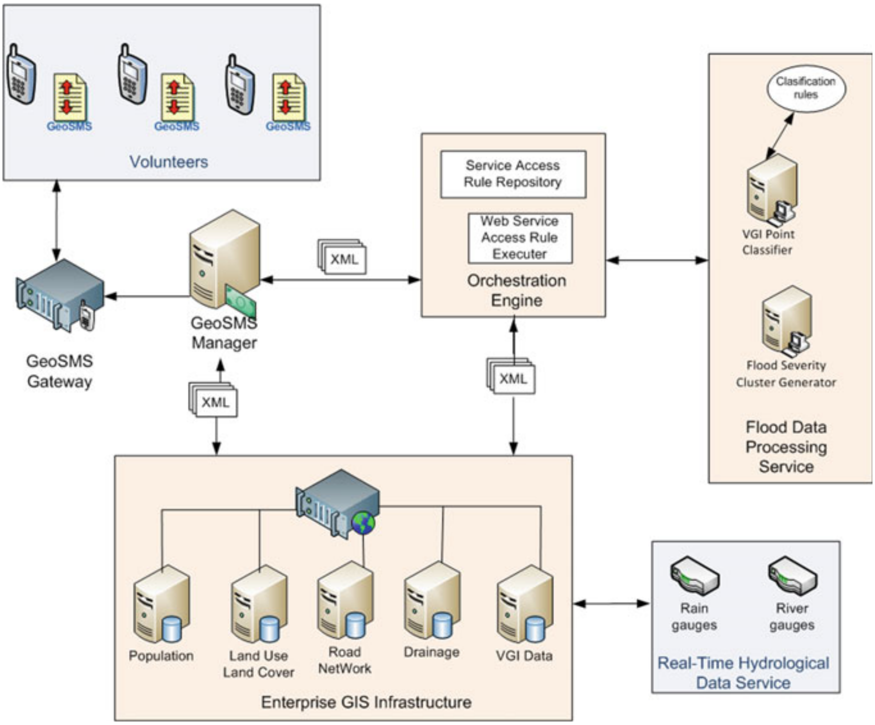


Fig. 1 Integration of VGI platform with SDI for disaster management

3.2 Spatial Data Sources

This component extracts authoritative from the SDI. In order to find the risk value each volunteered data, the authoritative spatial data resources such as "land use/land cover", "population", "elevation", "drainage network", and "hydrological data" have been considered.

3.3 Risk Map Generator

It consists of multiple geospatial data processing modules which implements algorithms to generate essential information regarding flood risk assessment. All data processing services use OGC Web processing services specifications [14]. Data processing rules are defined in the orchestration engine, and according to the requirement of geospatial information, the needed data processing services are accessed.

3.3.1 Decision Rules for Flood Risk Estimation

This module consists of a rule repository in a local database. It contains a set of "IF-THEN" rules considering the relevant attributes of geospatial data along with VGI attributes.

3.3.2 Generation Clusters of Risk Zones

This service creates the clusters in three separate layers such as "low", "moderate", "high". We use the DBSCAN clustering algorithm [15] to generate these clusters or zones. After creating the clusters for each risk value, the boundary points of each cluster has been found by applying "Graham's scan" [16] convex hull algorithm.

4 Evacuation Route Finding Methodology

The system generates different possible rescue routes from each safe zone and multiple combinations risk zones can be possible to generate a rescue route. After generating the optimized routes, each route has be assigned to a team of volunteers. There are *nr* number of risk zones, R $= \{r_1, r_2, \ldots, r_{nj}\}$ and *ns* number of safe zones, S $= \{s_1, s_2, \ldots, s_{ni}\}$ identified in the road network. In each safe zone, multiple teams of rescue volunteers have been created, and the team of volunteers set is denoted by K $= \{k_1, k_2, \ldots, k_{nk}\}$. The objective is to generate optimized rescue network from each safe zone with minimum rescue distance.

4.1 Risk Zones Assignment to a Safe Zone

The cyclic assignment technique has been used to link each risk zone to a particular safe zone. If there are *i* safe zones, then *i* number of lead cluster set has been created. Each lead cluster set l_i contains a safe zone and its most nearest risk zones. In order to store the distance between all risk zones, the risk distance matrix RDist[k][j] has been computed. The safe distance matrix, SDist[k][j] has been computed to store the distance between all safe zones to all risk zones.

4.2 Assignment of a Volunteer Team to a Route

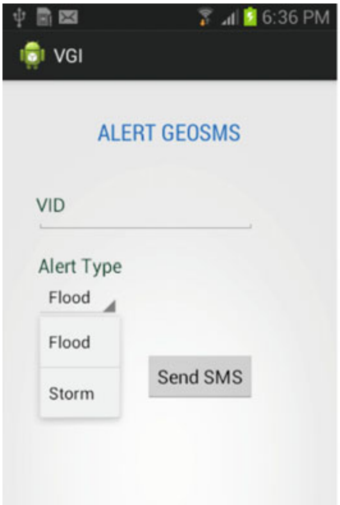
The rescue route should be created in such a way so that the effort requirement and distance traveled by a rescue team can be minimized. The Clarke and Wright savings algorithm has been adopted to get initial rescue route solution for each team of volunteers. A subset risk zones has been selected in a route which saves maximum distance for rescue operation.

4.3 Route Optimization of a Volunteer Team

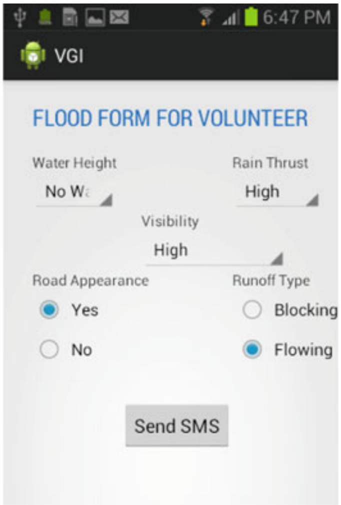
The routes are optimized by using ACO technique. It optimizes the rescue routes such that maximum route length should not exceed the rescue capacity of each team of volunteers. It constructs a complete tour for the first ant prior to the second ant starting its tour. Each ant generates a new route from each safe zone. There are a total of m solutions constructed sequentially if m artificial ants are used in one run of iteration. Before starting the tour, each ant is assigned to a randomly chosen risk zone as its first city to visit from the corresponding safe zones. Each risk zone r_j is associated Nr_j number of evacuees. For each route construction step, an ant k at current zone m will select the next zone n to visit from a feasible neighborhood N_m^k according to following probability distribution.

$$P_{mn}^k = \frac{[\tau_{mn}]^\alpha [\eta_{mn}]^\beta [\mu_{mn}]^\gamma}{\sum_{1 \in N_m^k} [\tau_{mn}]^\alpha [\eta_{mn}]^\beta [\mu_{mn}]^\gamma}, \text{if } n \in N_m^k$$

Here τ_{mn} denotes the pheromone between the current location m and next location n, and it indicates how much appropriate for the choice of the next zone. The η_{mn} is a heuristic value of desire for choosing the next zone, and it is defined as the inverse of the cost for rescue effort of a team of volunteers. The μ_{mn} measures the savings distance of route k. The parameters α, β and γ have been used to bias the pheromone concentration, and the set N_m^k denotes the list of candidate zones to be visited next. The candidate zone has been selected by considering maximum probability P_{mn}^k value.



(a) Alert Activation



(b) Flood Data Acquisition Form

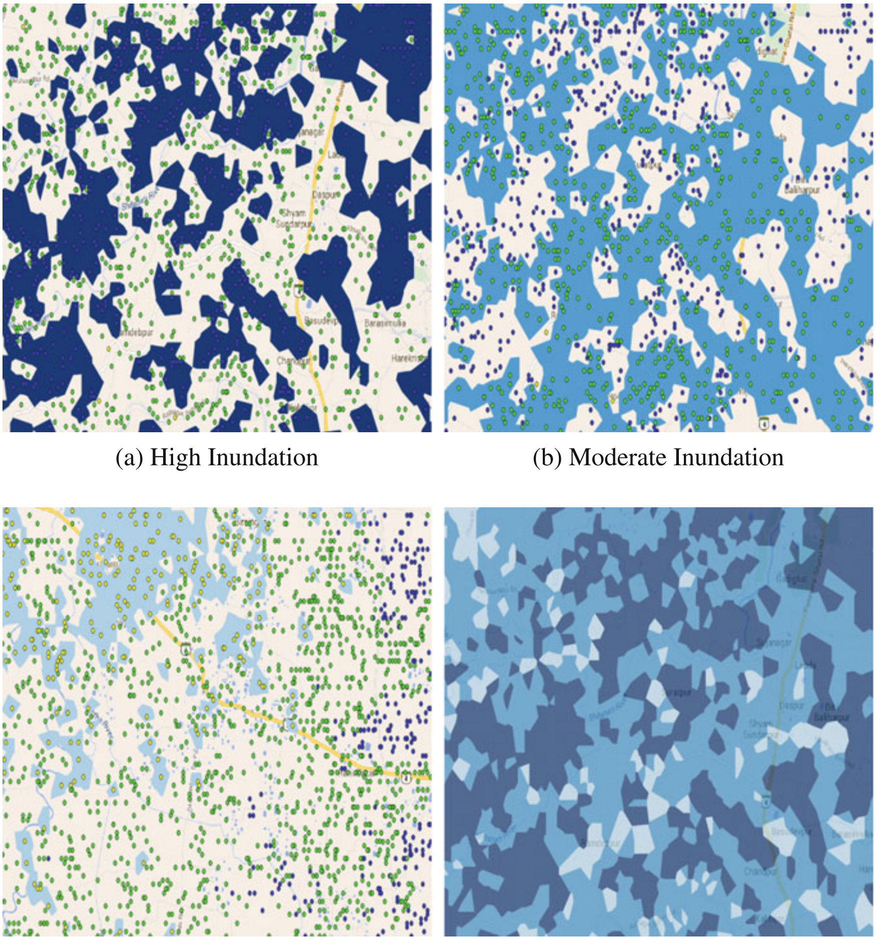
Fig. 2 GeoSMS application interface for volunteers

5 Case Study

In order to demonstrate this work, a prototype has been developed. The area under the catchment of Silabati River at Ghatal Block, Paschim Medinipur district, West Bengal, India, selected for deployment. A GeoSMS application has been developed by using Android platform to upload real-time disaster information around a volunteer. In Fig. 2, the GeoSMS application has been shown. According to the attributes of the survey form, the VGI database at the server end will be populated.

5.1 Development of Risk Zone Map

The DBSCAN clustering algorithm has been implemented to find risk map with different risk value. The service gets input from the orchestration engine, and it contains parameters EPS, MinPts, name of the VGI data service, and the decision value of a data provided by a volunteer. In this case study, the value of EPS 5KM and the value of MinPts 10 have been considered. Based on the specific risk values in an area, the boundary points are identified by computing the convex hull. In Fig. 3, the flood inundated zone based on deferent severity has been shown.



(a) High Inundation

(b) Moderate Inundation

(c) Low Inundation

(d) Over All Inundation

Fig. 3 Probable flood inundated zones with varying severity level

5.2 Rescue Route Planner

In this experiment, 42 risk zones and 14 safe zones have been identified. The system generates multiple team of volunteers who are associated with the different safe zones. A team of volunteers must visit a subset of risks zones with their route generated by the system. Each rescue route has been used by a single team of volunteers. The routes of the evacuation from all safe zones are shown in Fig. 4. The "rescue route optimizer" service based on ACO has been applied to obtain the best route for each team of volunteers.

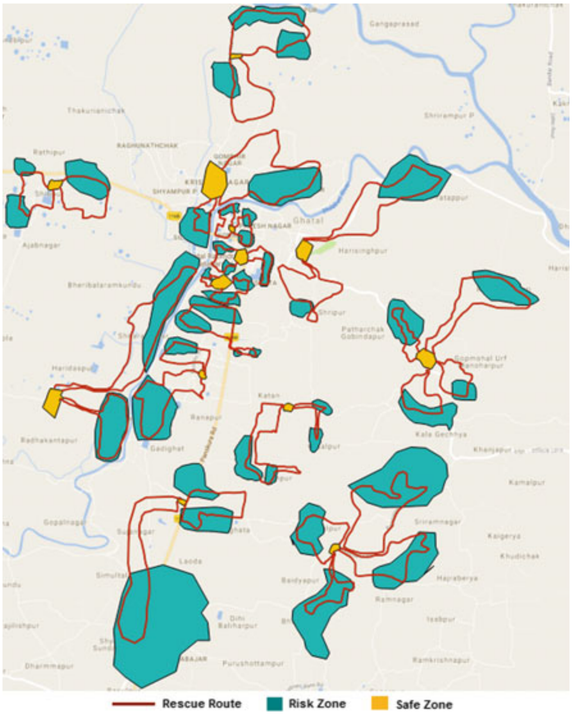


Fig. 4 Best rescue path map for all team of volunteers

6 Conclusion

In this work, local people of a disaster area have been utilized for collecting real-time information and rescue operation based on GeoSMS application. If mobile network has not been interrupted, then this framework can be utilized to develop a cost-effective rescue management system from the disaster area. The collected data has been integrated with SDI for risk map generation. This framework uses the OGC compliant Web services to access authorized geospatial data and produces the risk map along with rescue route in real time. The orchestration engine uses a rule-based classification algorithm to assess the different risk levels zones.

The rescue volunteers communicates the situational information through their mobile devices. If any team of volunteers finds that the route is not suitable, then the system modifies route for that team of volunteers. The route optimization has been done by ACO algorithm. The rescue teams have been routed in such a way so that all evacuees along the route can be rescued with violating rescue capacity and distanced constrained.

References

1. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**, 211–221 (2007)
2. Open Geospatial Consortium (2005). http://www.opengeospatial.org
3. Alvarez, M., Fernandez, T.D., Iglesias, R.C.: Social networks and Web 2.0 tools as a good complement to the local SDI. In: Proceedings of the GSDI12, Singapore (2012)
4. Mansourian, A., Zoej, M.J.V.: Expanding SDI hierarchy for countries with non-federated system: a case study of Iran. World Appl. Sci. J. **3**(4), 597–602 (2008)
5. Akinci, H., Erdogan, S.: Designing a flood forecasting and inundation-mapping system integrated with spatial data infrastructures for Turkey. Nat. Hazards **71**, 895–911 (2014)
6. Koswatte, S., McDougall, K., Liu, X.: SDI and crowdsourced spatial information management automation for disaster management. Surv. Rev. **47**, 307–315 (2015)
7. Schnebele, E., Cervone, G., Waters, N.: Road assessment after flood events using non-authoritative data. Nat. Hazards Earth Syst. Sci. **14**, 1007–1015 (2013)
8. Liu, S.B., Palen, L.: The new cartographers: crisis map mashups and the emergence of neogeographic practice. Cartogr. Geogr. Inf. Sci. **37**, 69–90 (2010)
9. Zook, M., Graham, M., Shelton, T., Gorman, S.: Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Med. Health Policy **2**, 7–33 (2010)
10. Erskine, M.A., Gregg, D.G.: Utilizing volunteered geographic information to develop a real-time disaster mapping tool: a prototype and research framework. In: Proceedings of the CONF-IRM12, Vienna (2012)
11. Zheng, Y.J., Ling, H.F., Xue, J.Y., Chen, S.Y.: Population classification in fire evacuation: a multiobjective particle swarm optimization approach. IEEE Trans. Evol. Comput. **18**, 70–81 (2014)
12. Duan, P., Xiong, S., Jiang, H.: Multi-objective optimization model based on heuristic ant colony algorithm for emergency evacuation. Natural hazards. In: Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, USA, pp. 1258–1262 (2012)
13. Open GeoSMS Standard Core 1.0. http://www.opengeospatial.org/standards/opengeosms (2012)
14. OpenGIS Web Processing Service Interface Standard 2.0. http://www.opengeospatial.org/standards/wps (2015)
15. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, USA, pp. 226–231 (1996)
16. Graham, R.: An efficient algorithm for determining the convex hull of a finite planar set. Inf. Process. Lett. **1**, 13–133 (1973)

Location-Based Services for Everyday Essentials

J. R. Raj and T. Sasipraba

1 Introduction

Advancement of technology ensures in delivering the right information at the right time to the varying requirement of the end-users. The paradigm shift of technology from object-oriented to service-oriented is used to simplify the design and development of software. Advent of Internet and proliferation of smartphones makes the user smarter to access and interact with any kind of Web-based application. Cloud infrastructure offers on-demand provision of computing power, memory and platform to the end user applications. Therefore, large amount of data is generated from these open systems. Data analytics defines a method or algorithm to make use of this huge volume of data and provides a suitable solution to the given problem. The key issue in this regard is to store, process and manage the data. The role of technology is to simplify the data processing and delivering the information in an understandable way to human and systems. Geographic information system (GIS) is a system empowers to process, store and present the spatial and attributed data to the end-users. Location-based services (LBS) are a technology used to navigate and process spatial data that can be visually represented into the map [1]. LBS spread across various categories which include navigation, routing, integration with social media. Proliferation of electronic devices like smartphone, sensors and other electronic gadgets plays vital role in human life. LBS is required to integrate with different kinds of applications that simplify the delivery of data in a user required format. The main challenge in LBS is to manage process and store the data [2]. Spatial data infrastructure (SDI) aims to share, exchange spatial data among stakeholders and provide better solution to the government and decision-makers. Childhood proposed volunteered geographic information (VGI) emphasize the user contributing source for the GIS data instead

J. R. Raj (✉) · T. Sasipraba
Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai 600119, India
e-mail: jebersonretnarajr@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_11

of traditional GIS system [3]. With the user's input, one can able to collect, modify, update and manage the GIS data. Furthermore, human acts as a sensor to collect the spatial information and making use of it for any GIS implementations [4]. The real challenge is accuracy of the recommended locations. Chennai is one of the densely populated cities in India which acts as a major technology hub for frontiers in automobiles, software and other allied services industries. Major software companies setting up their campus in Chennai and establish their operations. Citizens from various parts of the country and peoples around the world have coming to Chennai for medical aids, education, business, etc. It is very difficult for a layman who has limited knowledge about the city to find the locations of the essential services which includes banks and ATMs, RTO offices, electronics shops, malls and theatres, universities and colleges, schools, post offices, taluk office, municipal office, court. Though the existing systems provide the details of some of these services, they are limited in accuracy. They does not cover all these services provided by LBSEE. They focus on specific applications like transport, environment. Google provides location information based on user's user input. There is a possibility of false guidance which frustrates the user who supposes to visit the location. Therefore, there is a need of a system which provides the location details of all the essential services with more accurate and shows the route to the user. Numerous local search engines provide location information for different applications. However, these search engines intended for specific applications and none of them is being specific to the everyday essentials services. All of them are showing much irrelevant information which confuses the layman user. LBSEE bridges the gap of unavailable and inaccuracy of these specific services and processes all kinds of user's complex queries. The results are more accurate and show it to the user with advertisement free. Furthermore, LBSEE helps the user to locate these services based on their requirement, facilities available, and they can be visualized through map. The system receives the user query, processes it semantically, and the relevant spatial objects are retrieved and provided. The shortest route from the user's location to the required service is provided. The system is designed in such a way that it can be accessed via traditional Web-based and advanced mobile-based application.

1.1 Literature Survey

The term everyday essentials meant for a layman user who wants to know the locations of taluk office, post office, ATM, bank, Tahsildar office, RTO office, electronic shops, malls and theatre for a particular locality. The data used by the system is collected from volunteered input data sources. The aim of LBSEE is to develop a large-scale spatial information system of road network and all the essential services like schools and colleges, universities and government services. To achieve this, a novel algorithm is designed for automatic classification of unstructured input from the volunteered data. A GUI is designed to execute the user request and a data visualization mechanism to present the data to the user. The key challenge is the transformation of unstructured data into structured format and classification of these data. To achieve this, a novel Web news extraction and automatic classification method are introduced. With the help of novel knowledge discovery technique, the data can be extracted from the database. The main difference between other LBS systems and proposed LBSEE is the user can search for the essential service with facilities. For example, the complex queries like locating the government offices in a particular place, schools with facility, banks with locker facilities. The proposed system supports this user request and returned the results so that one can visualize it through map. The proposed system employs novel methods which include query processing algorithms, data mining algorithms, unstructured to structured conversion, Web news extraction algorithm. The scope of the system covers 600 km^2 of greater Chennai City. The user can access the required services like hospitals, police stations, schools, colleges and universities, and shopping malls. These are highly needed by the user, and the user can get the desired result within seconds. The location of the required service is recovered from the latitude and longitude (LatLng) values of the service, and the exact and accurate location is marked on the map. The route map details and directions are also shown to the user so that the user can reach his/her destination with ease. Numerous works have been done in location-based services and its applications. In [1], location-based mobile government services for emergency management are presented and analyses the social importance of these services. The acceptance and rejection of these services by various users are discussed. Geocrowd [5], a system has been introduced to collect spatial data from humans. Spatial task allocation scheme is used to collect the data from user. Samrat et al. proposed a map query interface used for interpreting user's complex query by the level of zooming the location on the map. User can zoom the particular location as a query instead of giving text input [6]. In [7], an automated news extraction system comprising of DOM and tag pattern generation method is proposed. They have used the method for extracting news items from online websites of Indian languages. In [8], an environment application is presented based on OGC enabled Web services. The Web services WFS, WMS and WPS are used for data processing, mapping and visualization. In general, SDI has limited support in analytical processing because it requires high computing and processing for inference the data. They used an environGRIDs Infrastructure for storage and computation of complex queries. A geospatial information (GI) retrieval system GeoFairy has been introduced by [9] for providing one-stop delivery of GI for disaster response and military applications. Sheng-Yuan et al. presented a Google map-based information system used as a tourist guide for visualizing the archaeological site in New Taipei City in Taiwan [10]. Sarah Spiekermann et al. designed a LBS-based administrative framework for supporting the interaction with clients. The framework employs a premise model which act as a vocabulary and ideas for supporting different classes of uses [11]. Check Strassman designed a LBS-based traveller desk application which provides guidance to the traveller for pointing of the locations [12]. In [13], a framework of location-based services in smart campus environments is presented. This android application provides end-users to recognize their locations and access the nearby services available to them. In [14], an efficient vehicle tracking system is designed and implemented for tracking the movement of vehicle from any location at any time. The proposed system made good use of a popular technology that combines a smartphone application with a microcontroller. A smart bus tracking system presented using Quick Response (QR) code. The passenger with a smartphone or mobile device with the QR code reader can scan QR codes placed at bus stops to view estimated bus arrival times, current locations of the bus and the bus routes are shown on the map [15]. Google Maps API provides a number of utilities for adding individual content to the Google map, and various Web map applications can be explored based on Google Maps API [16].

2 Elements of LBSEE

The architecture of the LBSEE is shown in Fig. 1. The service provider's details are collected from openly available volunteered input sources. The collected data is in the form of unstructured text format. It may be an online news item or text content. The unstructured data is converted into structured data. The classification algorithm based on the services label defined, classify it and updated into the GIS database. On the other hand, the user has to specify the request through the Web-based GUI or android smartphones. The spatial query processor analyses the request and appropriate stemming and tokenization process encountered. The appropriate call can be passed to the service discovery agent. The service discovery agent does the matchmaking process, and appropriate spatial objects are retrieved and provided to the client. The client can visualize it through the map.

The working methodology of the proposed system as follows:

- Collection of service providers details from the volunteered data sources.

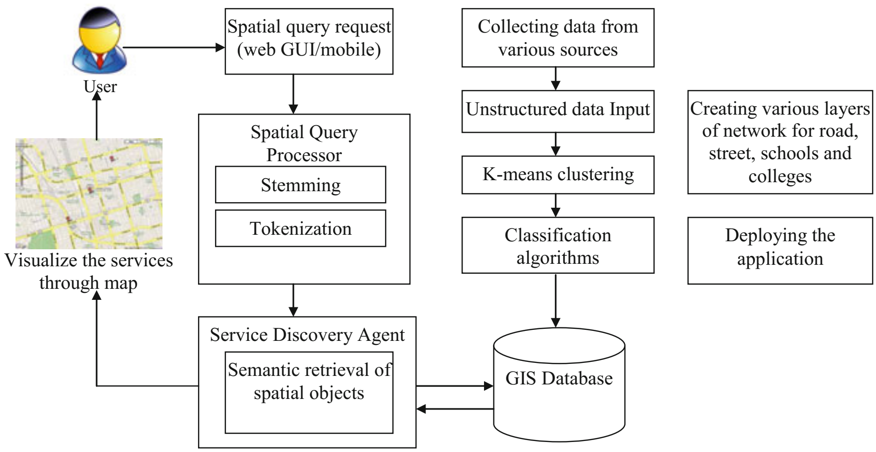


Fig. 1 Architecture of LBSEE

- Database designing.
- Conversion of unstructured input data into structured data.
- Designing of data mining algorithm for automatic classification.
- Query processing technique to execute the user request and development of knowledge discovery technique for extracting the data.
- Designing of GUI for data processing and visualization mechanism to present the data to the user
- Testing and validation on the communication ability, spatial query processing and response generation.

2.1 Data Collection

The data is collected from openly available sources. The details of all service providers are collected and stored into the GIS database. A total of 210 areas are identified for locating all these services. The database contains 2000 banks and ATMs, 970 schools, 700 police stations, 920 hospitals, 250 colleges, 150 post offices, 130 theatres, 30 fire stations and 20 RTO offices in the greater Chennai City. Each area is identified with area id of all essential services. The data is collected from various sources by verifying them throughout the updating process which in turn guarantees the relevant data to the user.

2.2 Database Designing

Database is a collection of information that is organized so that it can be easily accessed, managed and updated. Data is organized into rows, columns and tables, and it is indexed to make it easier to find relevant information. Data gets updated, expanded and deleted as new information is added. Databases process workloads to create and update themselves, and querying the data they contain and running applications against it. Database is made up of a set of tables with data that fits into a predefined category. Each table has at least one data category in a column, and each row has a certain data instance for the categories which are defined in the columns. Figure 2 shows the database view of the proposed system.

Each service providers consists of list of attributes. For example, areas contain area id, area name, stemmed name, latitude and longitude. The common attributes of all the providers contain id, name, address, area, phone number, latitude and longitude.

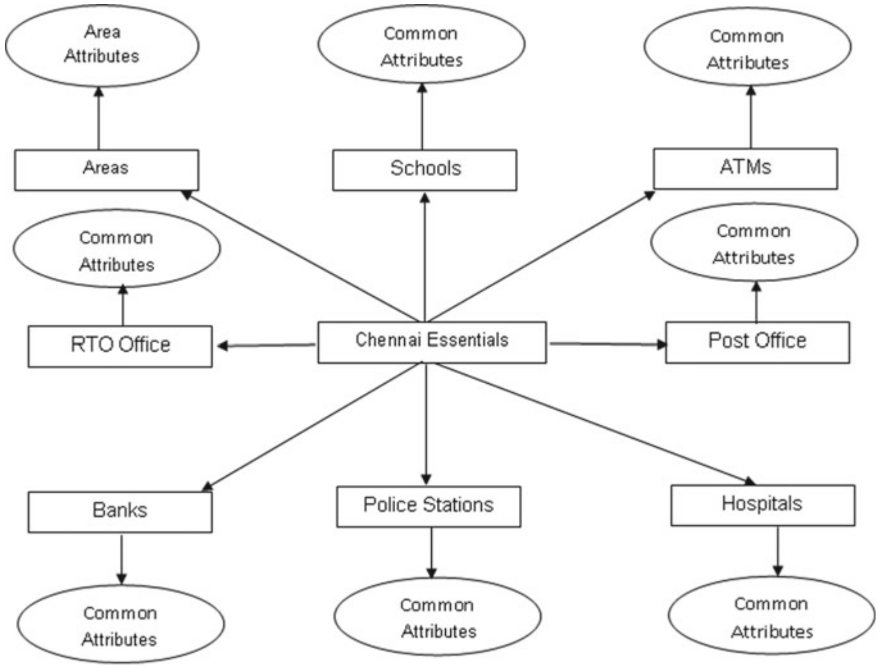


Fig. 2 LBSEE database architecture

2.3 Conversion of Unstructured Input Data

Conversion of unstructured data into structured data is an important phase of LBSEE. This can be achieved by Web news extraction method. The proposed framework considers a URL is the source of input. At that point, php cURL is utilized to extricate the information and is changed over into cluster by evacuating clamour and applying the stemming calculation to delineate words into their root words. This ensures to get the significant information from the insignificant information. The algorithm is described as follows:

- Take Input as URL.
- Extract data of URL using php cURL function and store in a string.
- Generate an array of words from a string.
- Apply stemming algorithm on these arrays of words.
- Apply string matching function to these Stemmed words.
- Store the relevant data.
- Get the structured formation of stored data.

Figure 3 shows the flow diagram of Web news extraction algorithm.

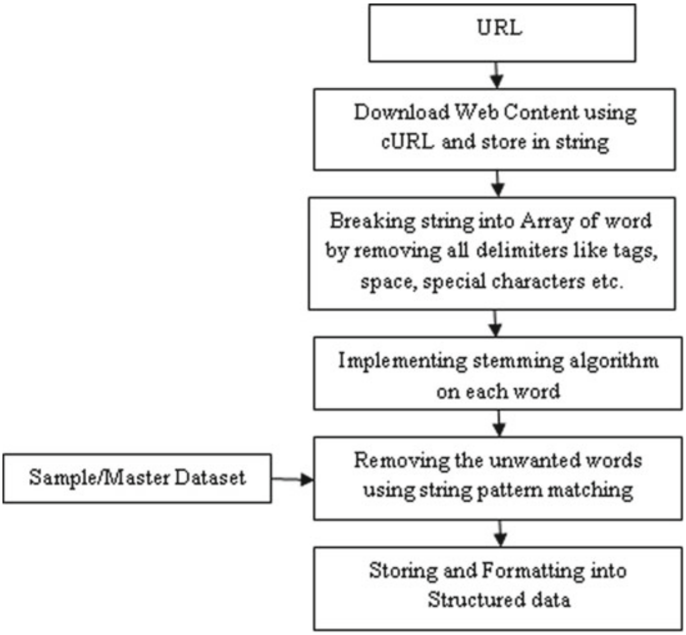


Fig. 3 Flow diagram of Web news extraction

2.3.1 Filtration of Extracted Content

The extracted information may contain a substance like labels, spaces, uncommon characters. This is changed over into crude information and put away in cluster for further execution. The information is removed, i.e. a string which was at first having an array here holds a word which was gotten by separating extricated Web substance is currently changed over into exhibit by evacuating all the delimiters.

2.3.2 Query Processing Technique to Execute the User Request

The user query can be processed semantically, and it involves several steps like tokenization, named entity extraction and parsing. The user query has to be preprocessed by removing the stop words, connectivity and making as a meaningful query. In the tokenization process, sentences are mapped into character strings and string of words. A tokenizer splits a stream of characters into a series of tokens. It involves set of rules such as case change, suffix numbers elimination, underscore separator. Name similarity can be applied only after the tokenization process, which produces the set of terms to be actually compared.

2.3.3 Stemming Algorithms

Stemming is the process of removing prefix, suffix and stop words. The output resulted in stemming (e.g. a, was, need, value, of, from, to, etc.). The importance of stemming at indexing time is for file compression and increases the search efficiency. The drawback is that the additional storage is required for storing the stemmed and unstemmed forms. Automatic approaches such as affix removal algorithms remove suffixes or prefixes from terms leaving a stem. These stemming algorithms sometimes also can transform the resultant stem.

Eg Input query—find RTO office near Adyar.

Stopwords = {“a”, “able”, “about”, “above”, “according”, “accordingly”, “across”, “actually”, “after”, “afterwards”…, I, …, need,………, etc.};

After tokenization and removal of stop words, the input query is split into RTO, office, Adyar. A simple example of an affix removal stemmer is the one that removes the plurals from terms.

2.3.4 K-Means Clustering Algorithms

The mean term for the similarity of list of terms is taken into an account. The terms relate with similar domain are clustered together based on the mean value. After the tokenization process, domain-specific mechanism will be implemented.

2.4 Designing of GUI for Data Visualization

The proposed framework utilizes format autonomous way to deal with concentrate GIS, GPS and Web news articles. There are two kinds of searches which can be done with the system. The basic service search feature helps the user to select a service and destination place and based on this the results are retrieved from database and are presented on map. The user clicks on the retrieved results to get the shortest and possible number of routes are shown to them. Figure 4 shows the user and the system interaction.

The process involved in the basic search is as follows:

- Getting input parameters as services and destination.
- Query formation based on the user selection.
- Retrieving the data from database.
- Displaying the contents on map.

In the advanced search, user can specify the query and the system understand the query and respond to the user. It is an advanced query processing search algorithm where a natural query is taken as input and is processed to get the results. Query from client is being separated, and filtration is performed to expel names, spaces, remarkable characters, etc.

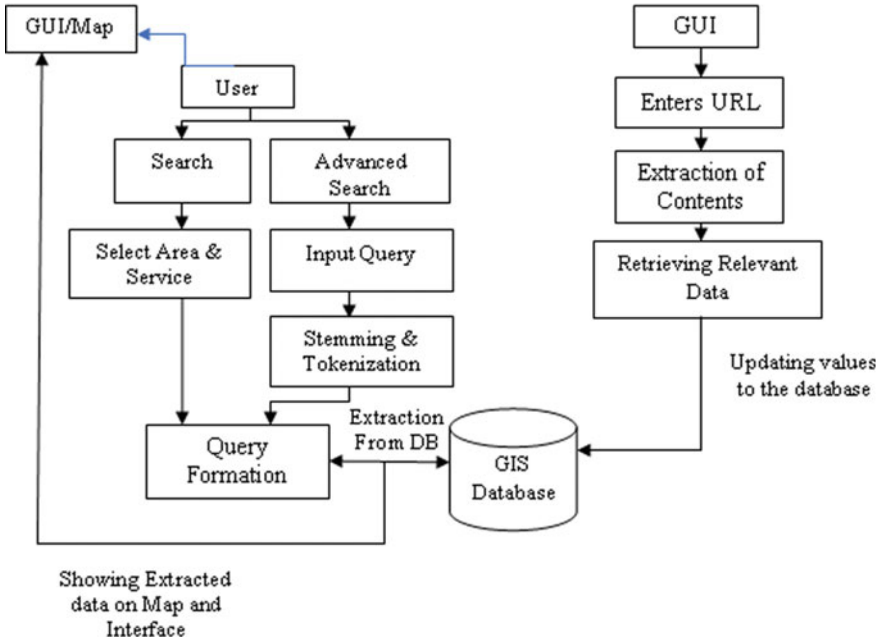


Fig. 4 User and system interaction

3 Experimental Results

Location-based services for everyday essentials (LBSEE) are an information system which provides spatial data to the end-user. The service providers' details are collected from volunteered data sources. The Web news extraction algorithm implemented to extract the relevant provider's information and updated into the database. Table 1 shows the results and time taken by the server to process the query and give the required and necessary output without any irrelevant data. The response time and accuracy are considered for testing the performance of the proposed system. Response time defines the elapsed time between a request and its response provided to the user. Accuracy defines the validity of the returned results. The system has been tested with variety of test data, and the results are shown in Table 1.

The user query is complex in nature, and the system has been designed in such a way that it can handle all kind of queries. The proposed system provides accurate results with a less throughput time with more accurate. The same will be applied in all the services. Table 2 shows some sample user queries and its average response time.

The precision of the proposed news extraction method is compared with other methods. The proposed method yields better results comparing with the existing system. Table 3 shows the comparison of the precision values with the LBSEE and existing DOM and tag-based content extraction method. Existing method extracts

Table 1 Response time of LBSEE news extraction

Website name (only news regarding new openings in Chennai)	Volume of data (Low/Moderate/High)	Maximum response time (ms)	Data accuracy and correctness
www.thehindu.com	Low	69299	Accurate
https://timesofindia.indiatimes.com/	High	51765 and session timeout	Partially accurate
http://www.indiapress.org/gen/news.php/Indian_Express/	Low	52147	Accurate

Table 2 Accuracy of LBSEE results

Input	No. of outputs	Maximum response time (ms)	Data accuracy and correctness
Banks in Sholinganallur	30+	1104	Accurate
State Bank of India in Sholinganallur	3	1430	Accurate

Currently, the Web-based LBSEE is available in http://www.chennaiessentials.co.in and android-based application is available in Google play store lbsee sathyabama

Table 3 Performance comparison

Method	Precision
DOM [7]	95.14
LBSEE	97.04

news from Indian regional languages websites. However, LBSEE is used the Web news extraction method from English online news websites.

The same algorithm implemented for English news extraction. The result shows that the LBSEE better precision, and result is more accurate comparing with the existing system.

LBSEE is helpful for a layman user who has limited knowledge about the Chennai City. It is readily available in the form of Web and mobile implementation so that one can easily access the services and find its locations. The system has its unique in arranging the services so that a user can use the click and get experience. Furthermore, a query interface is available to meet the user's complex and multiple queries. Through this advanced search mechanism, the user can specify the detailed query along with the services and additional facilities involved. The system is designed such a way that it can understand and relevant spatial response is provided. The system is extended its support to the android mobile users who can access the desired services.

4 Conclusion

Location-based services for everyday essentials (LBSEE) are successfully implemented, and it is available in a public domain. Web- and android-based applications are implemented and tested with various user queries. The system is successfully responded with accurate results. Web news extraction algorithm is successfully implemented, and the relevant service provider details are extracted and stored into the database. The query processing and stemming algorithm used to process and stemmed the user query for further processing. The user can search the service with the facility required. At the background, the system continuously processes the news items and the relevant provider details are updated. The android-based system handy is helpful to the user to search all kind of essential services. In future, the system can be extended for additional services like locating bus stops, tracking buses and smart parking. Using the available technology, the smart parking system can be implemented for Chennai City for locating parking places. The system can also be implemented for ios mobile phones.

Acknowledgement We thank NRDMS/Department of Science and Technology for funding this project DST No: NRDMS/11/1933/012.

References

1. Aloudat, A., Michael, K., Chen, X., Al-Debei, M.M.: Social acceptance of location-based mobile government services for emergency management. Telemat. Inf., 1–19. https://doi.org/10.1016/j.tele.2013.02.002
2. Gartner, G.: Advances in Location-Based Services, pp. 97–104 (2014). https://doi.org/10.1016/B978-0-444-62713-1.00007-6
3. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**(4), 211–221 (2007)
4. Elwood, S., Goodchild, M.F., Sui, D.Z.: Researching volunteered geographic information: spatial data, geographic research, and new social practice. Ann. Assoc. Am. Geogr. **102**(3) (2012)
5. Kazemi, L., Shahabi, C.: Geocrowd: enabling query answering with spatial crowdsourcing. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems, pp. 189–198. ACM (2012)
6. Samet, H., Sankaranarayanan, J., Lieberman, M. D., Adelfio, M.D., Fruin, B.C., Lotkowski, J.M., Teitler, B.E.: Reading news with maps by exploiting spatial synonyms. Commun. ACM **57**(10), 64–77 (2014)
7. Wanjari, Y.W., Mohod, V.D., Gaikwad, D.B., Deshmukh, S.N.: Automatic news extraction system for Indian online news papers. In: 2014 3rd International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions). IEEE (2014). ISBN: 978-1-4799-6896-1/14
8. Giuliani, G., Ray, N., Lehmann, A.: Grid-enabled spatial data infrastructure for environmental sciences: challenges and opportunities. Future Gener. Comput. Syst. **27**, 292–303 (2011). https://doi.org/10.1016/j.future.2010.09.011

9. Sun, Z., Di, L., Heo, G., Zhang, C., Fang, H., Yue, P., Jiang, L., Tan, X., Guo, L., Lin, L.: GeoFairy: towards a one-stop and location based service for geospatial information retrieval. Comput. Environ. Urban Syst. **62**, 156–167 (2017). https://doi.org/10.1016/j.compenvurbsys.2016.11.007
10. Yanga, S.-Y., Hsu, C.-L.: Allocation-based services and Google maps-based information master system for tour guiding. Comput. Electr. Eng., 1–19 (2015)
11. Spiekermann, S.: General Aspects of Location-Based Services. Humboldt-Universität zu Berlin, Location-Based Services, pp. 9–26 (2011)
12. Strassman, C., Collier, C.: Traveller information—an evolutionary perspective. In: The 3rd International Conference on Vehicle Navigation and Information Systems (1992)
13. Petcovici, A., Stroulia, E.: Location-based services on a smart campus: a system and a study. In: IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 94–99 (2016)
14. Lee, S., Tewolde, G., Kwon, J.: Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smartphone application. In: World Forum on Internet of Things (WF-IoT), pp. 353–358 (2014)
15. Eken, S., Sayar, A.: A smart bus tracking system based on location-aware services and QR codes. In: Proceedings of IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), pp. 299–303 (2014)
16. Li, H., Zhijian, L.: The study and implementation of mobile GPS navigation system based on Google Maps. In: International Conference on Computer and Information Application (ICCIA), pp. 87–90 (2010)

A Large-Scale Data-Oriented Intelligent System for Urban Growth Simulation

S. Pal and S. K. Ghosh

1 Introduction

According to the United Nations demographic survey, the world population is expected to rise from 7.3 billion in 2015 to 9.7 billion by 2050.[1] Furthermore, United Nations World Urbanization Prospects 2014 states that the current urban population of 54 percent is expected to increase to 66% by 2050.[2] This amounts to nearly an increase of 2.4 billion people in the urban areas, and 90% of this increase is expected to happen in *Asia* and *Africa*. All the above predictions indicate a huge shift of population from rural to urban areas, leading to massive urbanization in the cities which need to be monitored and controlled.

One of the major outcomes of massive urbanization is essentially urban growth [4]. Urban growth has various potentially devastating effects on the environment, climate, and biodiversity [23]. Rapid urban growth occurs at the cost of prime agricultural land, by transforming the environment and nonurban spaces, posing a threat to the flora and fauna of a region. This influences global environmental change [7] and increases in the levels of carbon dioxide in the atmosphere, which has considerable aftermath on biodiversity, human health, water quality, and climate change. Due to its acute impact on mankind, urban growth needs to be monitored and controlled in a systematic manner.

[1] http://www.un.org/en/development/desa/population/events/pdf/other/10/World_Population_Projections_Press_Release.pdf.
[2] http://www.un.org/en/development/desa/news/population/world-urbanization-prospects.html.

S. Pal · S. K. Ghosh (✉)
Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India
e-mail: skg@cse.iitkgp.ernet.in; skg@cse.iitkgp.ac.in

S. Pal
e-mail: pal.saptarshi@iitkgp.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_12

The studies focused on urbanizing regions have been conducted in various countries such as China [24], India [28], Nepal [6], Yemen [1], Belgium [21], Greece [14], Portugal [2], Slovenia [16], and Switzerland [26]. Despite having several studies in urban growth conducted in various cities of the world, there is still a requirement to carry forward more studies in this domain and develop systems which can monitor and predict growth. This is because a profound system of understanding the causes and development of new laws is required in order to prevent irresponsible growth.

As urban growth continues to emerge as a fundamental component of global environmental change, there is an immediate need to predict the evolution of urban areas in terms of magnitude and direction [17]. A systematic way to achieve this in a given area is to learn growth patterns from past time steps using machine learning/statistical techniques and project them into the future. Machine learning provides a set of tools based on statistics which can approximate functions from data. Recent advances in machine learning, neural networks, representation learning along with huge capacity GPU servers, and cloud infrastructure have allowed us to tackle large-scale problems. It has found applications in various domains, for instance, computer vision [12], recommendation systems [20], character recognition [11], bioinformatics [9], trajectory analysis [15]. With these technologies, we now have the ability to develop systems which can assist in monitoring and managing urban growth over large areas, thus providing sustainable solutions for our world.

In this work, we propose a large-scale system which can use two massively available data sources, namely remotely-sensed satellite imagery and road network to model urban growth. A distinct feature of the system is that it is dependent primarily on parameters and does not require significant human ingenuity to function. This means that the performance of urban growth can be improved by changing the parameters of the system and not any human-devised factors. With this system, we have modeled urban growth in the city of Mumbai, India and achieved superior accuracy than the existing learning-based methods.

2 Methodology

In this section, we describe the system components for modeling urban growth by integrating two distinct data sources, viz. remotely-sensed satellite images and road/transportation network. A pipeline of the modeling process is shown in Fig. 1 and is composed of two processing components, namely *automatic data representation* and *model design*.

- *Automatic data representation* is a set of techniques or algorithms which is used for representing unstructured data into a structured form in order to make it suitable for assembling and model design. This stage is designed using unsupervised learning methods.

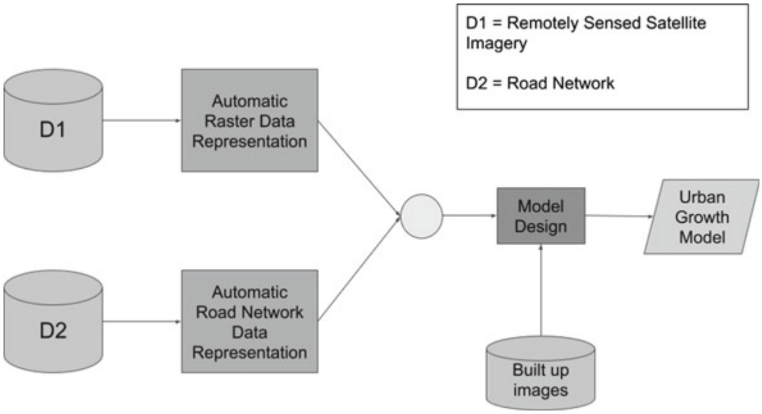


Fig. 1 Process pipeline of urban growth simulation

- The *Model design* stage matches the assembled dataset with an objective to build a model which can then be used to simulate urban growth. This is the supervised portion of our system which matches the objective of urban growth simulation with the generated representations.

2.1 Automatic Data Representation

Automatic data representation is a process of transforming/encoding unstructured data from a data source into a structured form. The reason behind this transformation is given as follows.

Structured data refers to information which can be easily stored in relational form, i.e., rows and columns format where rows represent features and columns represent attributes, for instance, spreadsheets. Due to the relational form of structured data, it is machine understandable but may not always be human readable. On the other hand, unstructured data cannot be easily stored in relational form and hence is not akin to machine language. For instance, road network and satellite imagery are unstructured data which cannot be easily converted into relational form, and hence, it is not machine understandable. Therefore, we use a set of unsupervised methods to first organize the unstructured data into a structured form.

This module is inspired by the idea of representation learning [3] which includes learning useful representations from unstructured data. We have proposed such a module in the system [18, 19] because we are learning representations from data in an unsupervised manner rather than manually devising them. This is a strength of our system which makes it more robust than the existing methods.

2.1.1 Automatic Raster Data Representation

The *automatic raster data representation* component transforms satellite remote sensing data into a *data matrix* with rows representing a cell and columns representing a feature. The *data matrix* can be formed from the raster using the following three steps.

- The neighborhood (discussed in Sect. 2.2) values of each pixel, including the pixel value, is extracted from each pixel in the raster image to form a matrix.
- An autoencoder is trained on the formed matrix to learn a representation in unsupervised fashion. An autoencoder [27] is a neural network which is composed of an encoder and decoder whose parameters are learned in an unsupervised manner from data.
- The trained encoder present in the autoencoder is then used to form encodings of fixed length. The encodings are assembled to form a *data matrix*.

Figure 2a gives a flow diagram of the procedure of creation of the data matrix using raster imagery.

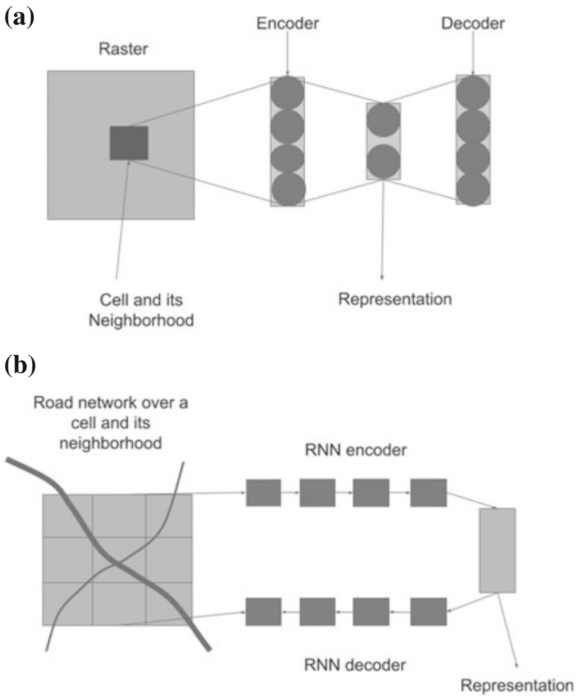


Fig. 2 Automatic data representation of **a** raster data **b** road network data

2.1.2 Automatic Road Network Data Representation

The *automatic road network data representation* component transforms road network data into a fixed length representation from road sequences extracted from the network. The process of forming the *fixed length representation* from road network data is as follows.

- Sequences are extracted from the road network according to the neighborhood criterion $N()$ as in Eq. 1 (discussed in Sect. 2.2).
- The sequences are used to train a recurrent neural network autoencoder to learn fixed length representation in an unsupervised form. A recurrent neural network autoencoder [25] is a neural network which is composed of an RNN encoder and decoder whose parameters are learned in an unsupervised manner from data.
- The trained encoder present in the autoencoder is then used to form encodings of fixed length. The encodings are assembled and concatenated to a *data matrix*.

A flow diagram for *automatic road network data representation* is given in Fig. 2b.

2.2 *Model Design*

The fundamental basis of our proposed urban growth simulation system is the theory of cellular automata (CA). A typical CA model is composed of an infinite array of cells having a finite number of states, which transforms at discrete time steps using certain transition rules. The transition rules of a CA model signify the relationship between a cell and its neighborhood which can be either Von Neumann (four neighbors) or Moore neighborhood (eight neighbors). The simulation process occurs when transition rules are iteratively applied to the cells over multiple time steps. The model definition of urban growth simulation is as follows.

We define a cell state at point p and time t as $S_p^t =< l_p^t, \tau_p^t, R_{raster}, R_{road} >$, where l_p^t is a binary label variable representing *built-up* and *non built-up*, τ_p^t is a transition indicator, R_{raster} is the fixed length raster representation and R_{road} is the fixed length road network representation. The transition indicator τ_p^t indicates whether the cell under consideration has undergone a transformation in the current time step. The transition indicator can take the values *non built-up* to *non built-up* (C_{NB}^{NB}), *built-up* to *built-up* (C_B^B), *non built-up* to *built-up* (C_{NB}^B), and *built-up* to *non built-up* (C_B^{NB}).

We define the update rule for the transition indicator as

$$\tau_p^{t+1} = f_T(l_p^t, N(l_p^t), R_{raster}, R_{road}), \quad (1)$$

where N is the neighborhood criterion and f_T is a update function which can be modeled using a supervised classification technique (for instance, random forests [5]). The transition function for update of the state S_p^t to S_p^{t+1} is given as,

$$S_p^{t+1} = \begin{cases} (S_B, \tau_p^{t+1}, R_{raster}, R_{road}) & \tau_p^{t+1} \in \{C_{NB}^{B}, C_{B}^{B}\} \\ (S_{NB}, \tau_p^{t+1}, R_{raster}, R_{road}) & \tau_p^{t+1} \in \{C_{NB}^{NB}, C_{B}^{NB}\} \end{cases} \tag{2}$$

The model design step comprises of the following steps.

- From the *data matrix* received from the *automatic data representation* stage, append the two factors $< l_p^t, N(l_p^t) >$ to it. This operation completes the *data matrix* with the parameter list of Eq. 1.
- Once we have the *data matrix* from the previous computations, a *label matrix* is formed from the built-up images datasource given in Fig. 1. The *label matrix* represents the transition indicator τ_p^t as given in Eq. 1.
- Finally, the *data and the label matrix* are fed to a classifier (say random forest) to build a model which can simulate urban growth.

2.3 Features of the Proposed System

The proposed system has the following features which distinguish it from the previously attempted studies on urban growth simulation.

- The key feature of the proposed system is the absence of a manual feature engineering process which enables us to scale the process of simulation to an arbitrary size of datasets.
- The *automatic data representation* module is unsupervised and utilizes machine intelligence to represent data into fixed length features. This immensely reduces the need for human ingenuity in the model design process.

However, these qualities come with a drawback because the models and representations are computational in nature, and this reduces human readability.

3 Case Study

In this section, we provide details of the experiments and inferences which have been carried out in the region of Mumbai (latitude: 19.0760° N and longitude: 72.8777° E). Mumbai is the capital city of Maharashtra and one of the major cities of India which has seen urban growth in last 20 years according to the Census of India. According to the reports by the Census, population of Mumbai has steadily risen from approximately 9 million in 1991 to more than 12 million in 2011.

The experiments have been conducted in a *Virtual Machine (VM)* running in an *Open Stack*-based cloud infrastructure. The VM consists of *8 VCPUs*, *16 GB RAM*, and *Ubuntu 14.04* as operating system.

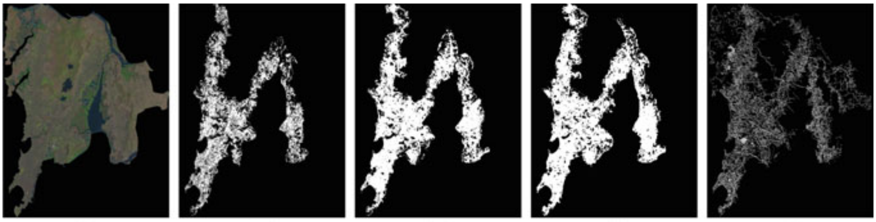


Fig. 3 Dataset of the Mumbai, India area, **a** Raster image Mumbai 1991, **b** Built-up 1991, **c** Built-up 2001, **d** Built-up 2011, **e** Road network of Mumbai

3.1 Data Collection and Preprocessing

For experimentation two kinds of datasets are required, namely, (i) raster images for previous years (1991, 2001, and 2011) and (ii) road network of the region. The raster images are downloaded from United States Geological Survey (USGS) Web site, and the road network data for the region of Mumbai have been collected from the open street maps data portal.[3] The extracted region of Mumbai from the satellite image consists of 972280 pixels. The satellite images are segmented semiautomatically to generate built-up maps for the years 1991, 2001 and 2011. The generated maps are verified using Google Timelapse,[4] Google Maps[5] as well as previously published works on Mumbai. The raster, built-up images, and road network for the region of Mumbai is shown in Fig. 3.

The data collection is followed by the *automatic data representation* stage where the data is converted into structured form. The neighborhood $N()$ considered is Moore neighborhood of radius 1. The sequences from the road network and the neighborhood information from the raster are extracted according to the neighborhood criterion to maintain consistency. In the data representation stage for raster images, we preprocess the neighborhood information using an autoencoder and encoding the data into a representation of size 10. For the road network, the extracted sequences are converted into fixed length representations of length 10 with the procedure discussed in Sect. 2.1.

3.2 Training and Representation

The generated sequences and the raster data matrix are concatenated to form feature vectors for model design. The total number of features generated from the prepro-

[3]https://www.openstreetmap.org/.

[4]https://earthengine.google.com/timelapse/.

[5]https://www.google.com/maps.

cessing is 972280, and the dimension of each feature vector is 46. From the data matrix and the label matrix, we have used a random forest classifier to build an urban growth model.

For validation of our model, we have used the validation criteria of urban growth (FoM, PA, UA, and OA) to quantitatively measure the representation quality. The four validation metrics (FoM, PA, UA, and OA) by [22] are dependent on five variables given as follows.

- A = Area of error due to observed change predicted as persistence.
- B = Area correct due to observed change predicted as change.
- C = Area of error due to observed change predicted in the wrong gaining category.
- D = Area of error due to observed persistence predicted as change.
- E = Area correct due to observed persistence predicted as persistence.

Figure of merit (FoM) provides us the amount of overlap between the observed and predicted change. Producer's accuracy (PA) gives the proportion of pixels that the model predicts accurately as change, given that the observed change is indicated by the reference maps. User's accuracy (UA) gives the proportion of pixels that the model predicts accurately as change, given that the model predicts change. The equations of the metrics are given as follows.

$$FoM = \frac{B}{A + B + C + D} \tag{3}$$

$$PA = \frac{B}{A + B + C} \tag{4}$$

$$UA = \frac{B}{B + C + D} \tag{5}$$

$$OA = \frac{B + E}{A + B + C + D + E} \tag{6}$$

3.3 Results and Discussion

The comparison of our framework with existing frameworks [8, 10, 13, 23] based on the four parameters (FoM, PA, UA, OA) as in Fig. 4 reveals that *end-to-end* learning performs better than the existing learning-based methods developed for urban growth prediction. We argue based on the results that this is possible due to the superior representation and hidden rule detection capability of our proposed *end-to-end* system. Our proposed framework provides an improvement of 18%, 17%, 2%, and 3% on FoM, PA, UA, and OA approximately, respectively, over the existing frameworks.

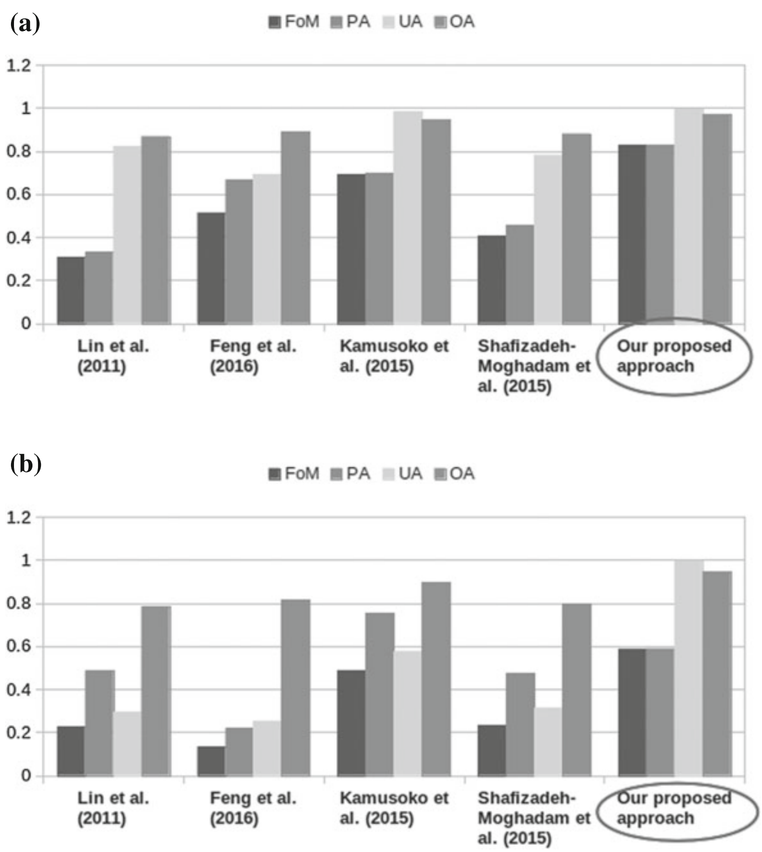


Fig. 4 Performance analysis of urban growth simulation with existing learning-based methods **a** 1991–2001 **b** 2001–2011

The improvement in performance of our system can be attributed mainly to the *automatic data representation* module, which is not present in the existing methodologies. By learning representations automatically from the data, we have replaced *manually designed features* that have been prevalently used in the existing methodologies [8, 10, 13, 23]. The issue with *manually designed features* is it limits the performance of machine learning models [3]. Moreover, increase in performance requires an addition of new features to the existing set of features, which requires human ingenuity. This is a major drawback in designing systems because constant human intervention during the functioning reduces the quality of service.

Another major problem which has not been addressed in the existing works is imbalanced dataset problem that occurs because the number of transition pixels (C_{NB}^{B}, C_{B}^{NB}) is generally much lesser than the number of persistent pixels (C_{NB}^{NB}, C_{B}^{B}). This also is a reason for the reduction in performance but has not been considered

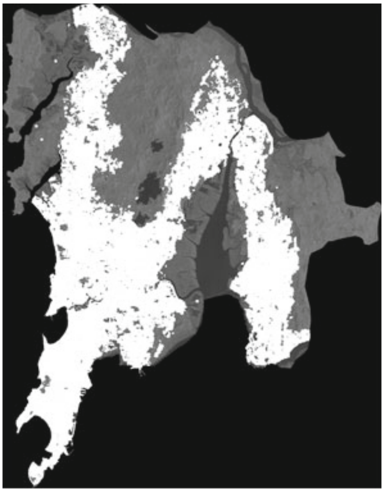


Fig. 5 Simulated urban conditions of Mumbai 2031

in the existing works. We have divided the classes into four labels in order to solve this problem, which is also a reason for the enhancement of performance.

From the simulated urban growth in Fig. 5, we can see that the major water bodies and swamps in the area are less affected by urban growth. This is because such patterns have occurred infrequently in the dataset, and even without any manual feature engineering, the system is able to capture them. On the contrary, the region of Mumbai became denser as seen from the dataset, and the simulation also provides similar results. This indicates that the automatic pattern recognition is in accordance with the urban maps provided in the dataset.

In order to make the system more robust, it is important to be able to add various data sources in an unsupervised fashion using concepts of *representation learning*. An important data source which determines urban growth is policy documents which may include economic policies, building policies, etc. However, it is in a completely different mode than other spatial or geographic data format. This makes the task more challenging than integrating spatial data sources into an inherently spatial–temporal problem.

4 Conclusion

We have proposed a large-scale data-oriented system using the concept of representation learning in urban growth prediction by using remote sensing and road network data. We have empirically verified our system on simulating urban growth for the region of Mumbai over a time frame of 20 years. Our proposed system has outperformed existing learning-based methodologies by providing a more efficient and generic structure. Future work on this system can be done by extending it to data sources other than raster and road network data. Furthermore, the approach needs to be verified on data obtained from other cities for measuring its effectiveness in simulating urban growth.

References

1. Al-shalabi, M., Billa, L., Pradhan, B., Mansor, S., Al-Sharif, A.A.A.: Modelling urban growth evolution and land-use changes using gis based cellular automata and sleuth models: the case of Sanaa metropolitan city, Yemen. Environ. Earth Sci. **70**(1), 425–437 (2013)
2. Araya, Y.H., Pedro, C.: Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal. Remote Sens. **2**(6), 1549–1563 (2010)
3. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans, Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
4. Bhatta, B.: Urban growth analysis and remote sensing: a case study of Kolkata, India 1980–2010. Springer Science & Business Media (2012)
5. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
6. Chitrakar, R.M., Baker, D.C., Guaralda, M.: Urban growth and development of contemporary neighbourhood public space in Kathmandu Valley, Nepal. Habitat Int. **53**, 30–38 (2016)
7. Deng, X., Huang, J., Rozelle, S., Uchida, E.: Economic growth and the expansion of urban land in China. Urban Stud. **47**(4), 813–843 (2010)
8. Feng, Y., Liu, Y., Batty, M.: Modeling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu-Songjiang area of Shanghai, China. Stoch. Environ. Res. Risk Assess. **30**(5), 1387–1400 (2016)
9. Frank, E., Hall, M., Trigg, L., Holmes, G., Witten, I.H.: Data mining in bioinformatics using weka. Bioinformatics **20**(15), 2479–2481 (2004)
10. Kamusoko, C., Gamba, J.: Simulating urban growth using a random forest-cellular automata (RF-CA) model. ISPRS Int. J. Geo-Inf. **4**(2), 447–470 (2015)
11. Kim, K.K., Kim, K.I., Kim, J.B., Kim, H.J.: Learning-based approach for license plate recognition. In: Neural Networks for Signal Processing X, 2000. Proceedings of the 2000 IEEE Signal Processing Society Workshop, vol. 2, pp. 614–623. IEEE (2000)
12. Lafferty, J., McCallum, A., Pereira, F., et al.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the Eighteenth International Conference on Machine Learning, ICML, vol. 1, pp. 282–289 (2001)
13. Lin, Y.-P., Chu, H.-J., Wu, C.-F., Verburg, P.H.: Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling–a case study. Int. J. Geogr. Inf. Sci. **25**(1), 65–87 (2011)
14. Louri, H.: Urban growth and productivity: the case of greece. Urban Stud. **25**(5), 433–438 (1988)
15. Morris, B.T., Trivedi, M.M.: A survey of vision-based trajectory learning and analysis for surveillance. IEEE Trans. Circuits Syst. Video Technol. **18**(8), 1114–1127 (2008)
16. Murray-Rust, D., Rieser, V., Robinson, D.T., Miličič, V., Rounsevell, M.: Agent-based modelling of land use dynamics and residential quality of life for future scenarios. Environ. Model. Softw. **46**, 75–89 (2013)
17. Musa, S.I., Hashim, M., Md Reba, M.N.: A review of geospatial-based urban growth models and modelling initiatives. Geocarto Int. 1–21 (2016)
18. Pal, S., Ghosh, S.K.: Learning representations from road network for end-to-end urban growth simulation (2017). arXiv:1712.06778
19. Pal, S., Ghosh, S.K.: Rule based end-to-end learning framework for urban growth prediction (2017). arXiv:1711.10801
20. Pazzani, M., Billsus, D.: Content-based recommendation systems. In: The Adaptive Web, pp. 325–341 (2007)

21. Poelmans, L., Van Rompaey, A.: Complexity and performance of urban expansion models. Comput. Environ. Urban Syst. **34**(1), 17–27 (2010)
22. Pontius, R.G., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K.: Comparing the input, output, and validation maps for several models of land change. Ann. Reg. Sci. **42**(1), 11–37 (2008)
23. Shafizadeh-Moghadam, H., Hagenauer, J., Farajzadeh, M., Helbich, M.: Performance analysis of radial basis function networks and multi-layer perceptron networks in modeling urban change: a case study. Int. J. Geograph. Inf. Sci. **29**(4), 606–623 (2015)
24. Shu, B., Bakker, M.M., Zhang, H., Li, Y., Qin, W., Carsjens, G.J.: Modeling urban expansion by using variable weights logistic cellular automata: a case study of Nanjing, China. Int. J. Geograph. Inf. Sci. 1–20 (2017)
25. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
26. Tannier, C., Thomas, I., Vuidel, G., Frankhauser, P.: A fractal approach to identifying urban boundaries. Geograph. Anal. **43**(2), 211–227 (2011)
27. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. **11**, 3371–3408 (2010)
28. Wakode, H.B., Baier, K., Jha, R., Azzam, R.: Analysis of urban growth using landsat TM/ETM data and GIS a case study of Hyderabad, India. Arabian J. Geosci. **7**(1), 109–121 (2014)

Spatial Facility Management: A Step to Design Smart City

R. Gupta

1 Previous Works

1.1 Genetic Algorithm

Evolutionary algorithms are a promising alternative for solution methodology where adaptivity to problems, complex computations, innovation and large-scale parallelism is required [1]. Genetic algorithm is a heuristic search algorithm belonging to the class of evolutionary algorithms [2]. Holland's book 'adaption in natural and artificial systems' presented the genetic algorithm as an abstraction of biological evolution [2]. It can be viewed as search method based on the process of natural selection and transformation. This algorithm mimics Darwin's theory of evolution by means of natural selection survival of the fittest [3–5]. In one line, we can say that genetic algorithm is an evolutionary algorithm which initiates by generating some individuals (solutions) randomly, performs some operations on these individuals so as to make them better and aims at terminating with a solution/individual of intended fitness. GA begins with encoding all the possible solutions into finite length array of bits. These strings called chromosomes and represent the potential solutions to our problem in the form of genes. These chromosomes form the initial population for the problem and are tested for fitness/quality. Fitness is the measure of how good that solution is. It is measured with fitness function; fitness function is defined according to the problem. After calculating, the fitness selection is done in such a way that chromosomes with greater fitness value have greater probability of getting selected [6, 7]. Then crossover and mutations operations are performed on these selected chromosomes, and a new population is formed by replacing the bad solutions with

R. Gupta (✉)
Civil Engineering Department, Birla Institute of Technology and Science, Pilani 333031, Rajasthan, India
e-mail: rajiv@pilani.bits-pilani.ac.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_13

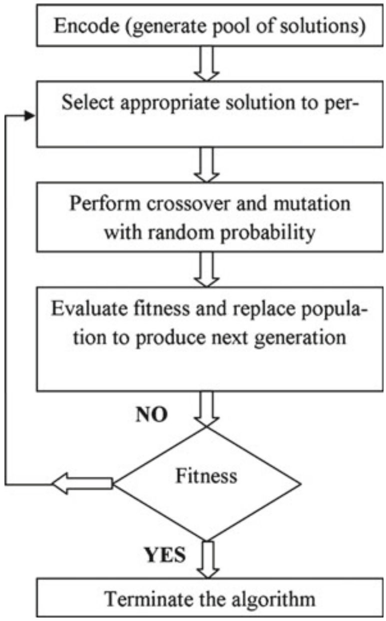


Fig. 1 Flow chart of genetic algorithm

the good ones. This process is repeated until a satisfactory fitness value is reached or in some cases other constraints bound us to terminate the process [8–10]. In genetic algorithm, binary strings are stored in computer memory and are transformed over time in a way that is similar to evolution of population under natural selection [11]. According to Ph. Preux and E.G. Talbi, there exist certain dynamics in the algorithm which is inherent to the process, independent of the solution representation or other manipulations by the operators [12]. Genetic algorithm provides an alternative method and outperforms many traditional methods for problem-solving [3, 4]. Flow chart in Fig. 1 summarizes the entire process.

1.2 2D Genetic Algorithm

Genetic algorithm is a kind of evolutionary algorithm that starts with random solutions and evolves to the final solution by following the concept of natural selection and survival of the fittest. It is applied generally when there is no or very less domain-specific information available and when a large set of combinations is there to deal with. There are many areas in computer science where genetic algorithm has been applied successfully [13]. Some of them are optimization problems like solving travelling salesman problem, robotics and code breaking.

The algorithm starts with encoding the chromosomes or initializing in 2D form. Although initialization is random, domain-specific knowledge or other information can easily be incorporated while initializing the population [14]. Choosing a representation scheme that correlates with the fitness function used can make the problem much easier [15]. There are several methods to denote the chromosomes in simple genetic algorithm like binary encoding, value encoding [16], etc.. Like [17] have used a problem-specific quantum bit representation we have also deployed matrix encoding (2D problem) for our problem. Every encoded solution is evaluated for fitness using a pre-defined fitness function. Then based on their fitness, they are selected for further operations.

After selecting from the pool of population, appropriate individual crossover and mutations are performed on them. Unlike evolutionary strategies, mutation is often the secondary operator in GAs, performed with a low probability [18, 19]. Crossover operator is considered to be the central component of genetic algorithm [20, 21]. Basic crossover in two-dimensional encoding has several problems. With traditional crossover techniques, there may be more than one instance of the same facility and in the other child that particular facility will be missed out entirely. One way is to discard this type of solutions and move on with the valid ones only. But it will lead to computational loss and convergence time will also significantly increase. Hence, we have come up with a new crossover method the algorithm for which is given below.

1.3 Algorithm for Modified Crossover

Step 1 Choose any two chromosomes randomly from the parent pool. Name them as matrix one and two.

Step 2 Generate two empty matrices each representing one child matrix.

Step 3 Select some facilities from parent1/matrix1. Place these selected facilities in child1 at the place they were present in matrix2.

Step 4 Try to place rest of the facilities at the same place as they were in matrix1. If no contiguous space is available, then place them randomly wherever continuous space is available.

Step 5 Repeat step 3 and step 4 for child 2 by replacing matrix1 with matrix2 and the other way also.

Step 6 Repeat step 1–5 for crossing over of other chromosomes.

The process is followed by evaluation of child chromosomes generated by crossover and replacement of some candidate solutions. The process of evaluation followed by crossover replacement is repeated until we reach at fitness value 1. This means all the facilities have been allocated space successfully while satisfying all the constraints.

1.4 Facility Management (FM)

Facility management is an interdisciplinary field which aims at best possible utilization of scarce resource like land in highly populated country like India. There are various definitions of facility management proposed by different researchers all around the world. There is no single definition which can be treated as standard one because people work and take facility management with different viewpoints. Facility management can be viewed as an integrated approach for maintaining, improving and adapting the buildings of an organization in order to create an environment that strongly supports the primary objective of that particular organization [22]. The above-stated definition is given by Barret with a perspective of handling an organization and its facilities. This is a reduction from town or city scale to organization. But the concept given by Becker can directly applied on large scale as well. We need to manage the facilities in any area in accordance with the objective already defined before starting the management process as different areas can be have different requirements. These requirements will influence the output directly. The major aim while managing the facilities in one area that is already overcrowded with population will definitely be different from other area which is less populated.

The demands hence vary with change in area and other factors like population, location. The scope of facility management hence covers all the aspects of property, space, environmental control, health, safety and support services. Facility management is responsible for coordinating all efforts related to planning, designing and managing buildings, hospitals and other facilities in that area to enhance the ability of the city or country to adapt to constantly changing world [23].

In early 1980s, International Facility Management Association was established to incorporate the associations dedicated to serve as facility management professional in North America. With the development in society and facility management as well, the association is serving in more than 60 countries across the world. In India, also a trend of planned cities has arrived in the previous years. If we follow the organizational approach to facility management, then we can say that there are three core factors to it [24]. First, facility management is a strategic discipline when it is clear that most of its practitioners are situated at an operational level in their respective organizations. Second, FM professes to want to be at centre of organizational development when clearly many of FM services are delivered either by external consultants or in-house teams set up as internal consultants. Third, FM professes to be proactive in managing change within organizations when quite clearly it is reactive in most cases. Facility management shortly called FM is an interdisciplinary field devoted to the coordination of space, infrastructure, people, places and organization. It is often associated with the administration of parks, post offices, other office blocks, arenas, schools, convention centres, shopping complexes, hospitals, hotels, etc. However, FM facilitates on a wider range of activities than just business services and these are referred to as non-core functions.

Many algorithms have been developed to deal with the problem of facility management effectively and efficiently. Some of these are branch and bound algorithms introduced by Kuenne and Soland in 1972, simulated annealing, tabu search, random descent method and many more. Each of these algorithms accepts input in the format specified and produces output based on certain constraints. Genetic algorithm is one algorithm that can be applied to any problem that involves a huge number of combinations and to problems for which no specified procedure till date is known to reach the output. They operate entirely on random basis. The general practice is to apply the algorithm in one dimension only with the help of arrays. But to preserve the information and produce more correct and accurate output, we can directly use genetic algorithm to solve the facility management problem in two dimensions instead of following the traditional approach.

2 Methodology

Genetic algorithm is a six-stage process with mutation or crossover being optional, and we now move towards the implementation of these stages of the algorithm. After selecting the study area as cities, town or even village, we calculate the area occupied by each facility. In view of set objectives, the first step is identification of study area. Preliminary data acquisition is carried out for various existing facilities. The province map with various specific details is acquired from the remote sensing agency which is used in GIS analysis. The spatial information is collected with the aid of global positioning system. Questioner is framed, and various attribute information are collected by the actual survey. Various targeted facilities are prioritized and used as input steps for the developing the tool with the aid of graphical user interface (GUI). These results are used for the actual implementation and increasing social awareness among end-users which in turn helps to increase the standard of living of the community. The generated database in GIS in the form of maps and tables is used to develop SDI and share the same on network which is useful for the replication of similar systems. SDI developed is tested on field based on the acquired inputs and the need of the settlement.

2.1 Overall Procedure

Starting from satellite images till developing the genetic algorithm, the entire methodology is described in terms of flow chart (Fig. 2) and steps are explained subsequently.

The genetic algorithm applied will be in two dimensions that will prevent loss of values from the maps to the algorithm mapping. These algorithms perform random searches through a domain of possible solutions, with the aim of finding the best alternative with respect to given definition of 'goodness'.

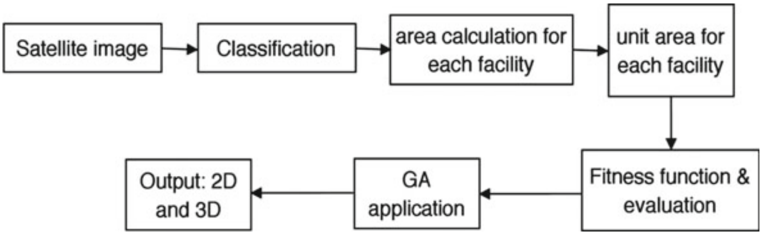


Fig. 2 Flow chart of complete methodology to be followed

Fig. 3 Satellite image of university campus

2.2 *Steps of Algorithm*

Use panchromatic satellite image of required resolution. You need to specify the longitude and latitude of the area under consideration. The image shown in Fig. 3 is an example of the kind of image that we can work on.

1. Input the number of facilities to be planned, like industrial, commercial, public facility, residential, agricultural, recreational, educational. A classified image is shown in Fig. 4.
2. Input the total area of all facilities.
3. Input the area of the region in which facilities are to be placed.
4. Input the number of localities planned to be developed.
5. Input the minimum distance required within all types of facilities.
6. Generation of population using random placement of all the facility units contiguously locality by locality such that facility units placed contiguously at once are total facility units divided by localities.
7. Generation of distance matrix for every solution after calculation of Euclidean distance between every pair of facility units.

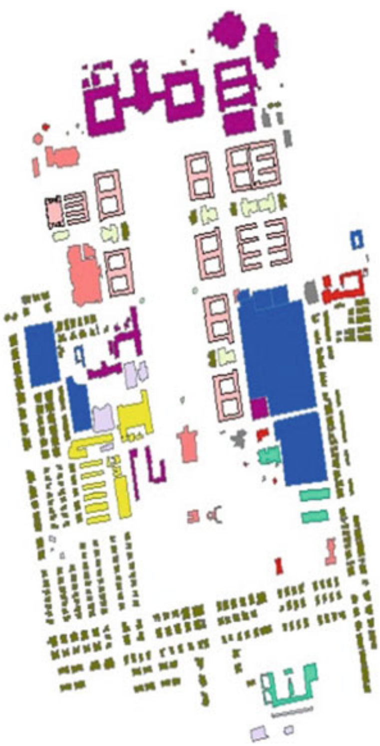


Fig. 4 Classified image of BITS campus (shown in Fig. 3)

8. Generation of an ideal matrix between every pair of facility units referring the input from point 5 above.
9. Comparing the values of ideal matrix and distance matrix of every solution of population such that all the entries in distance matrix should be greater than or equal to entries in ideal matrix.
10. If point 9 is not satisfied, four solutions are selected as child solutions randomly among the generated population.
11. Every child solution is used to generate one solution taking any four randomly selected facility types from itself and placing these four and rest facilities again in the same random fashion as place in point 6.
12. Again the fitness value is evaluated for the solutions generated as population and child solutions.
13. Two solutions with highest fitness and lowest fitness are stored for further use.
14. Solutions with best fitness equal in numbers to population generated in point 6 less 2 are selected and again the same process runs from point 9 to point 14 until we obtain the desired fitness value.
15. When the desired fitness value will be obtained, an image presenting the optimal facility placement according to the fitness function will be displayed.

Flowchart of the complete methodology is shown in Fig. 5.

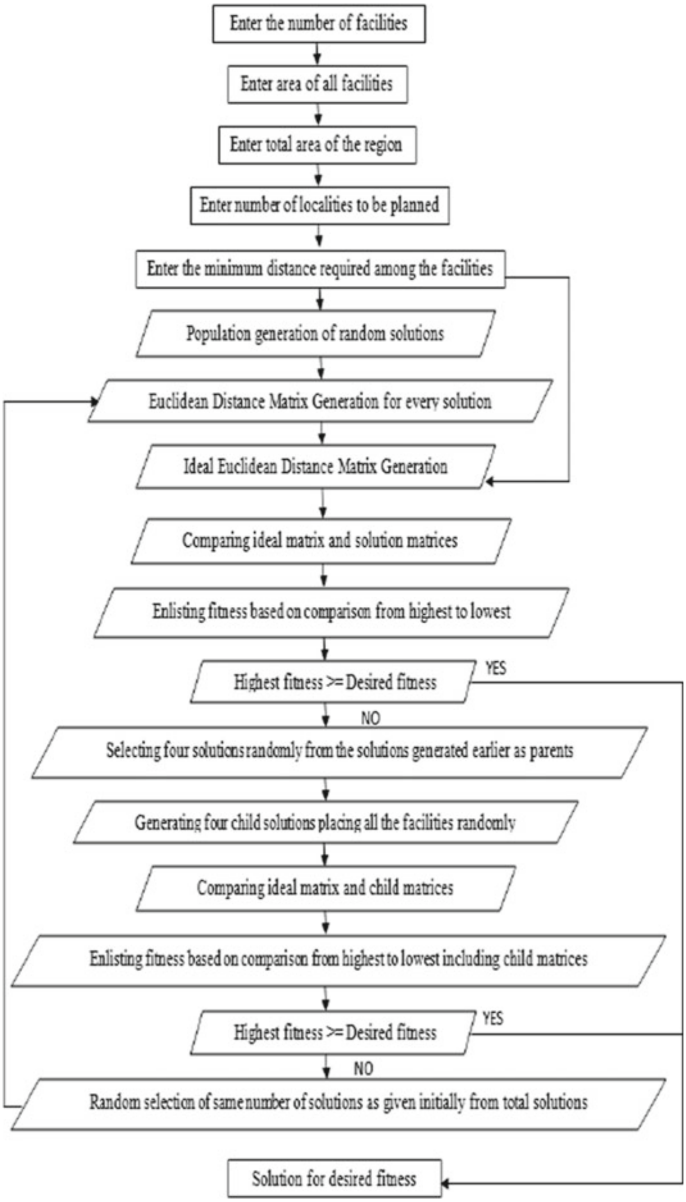


Fig. 5 Flowchart of the methodology developed

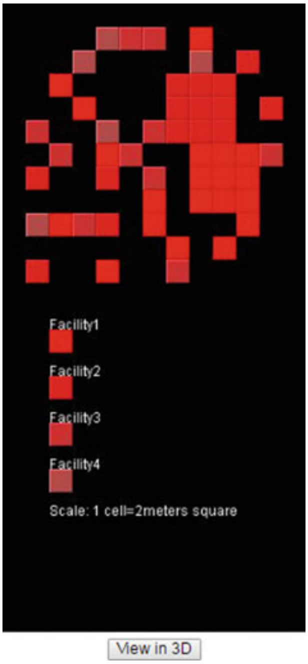


Fig. 6 Optimal 2D plan generated by GA

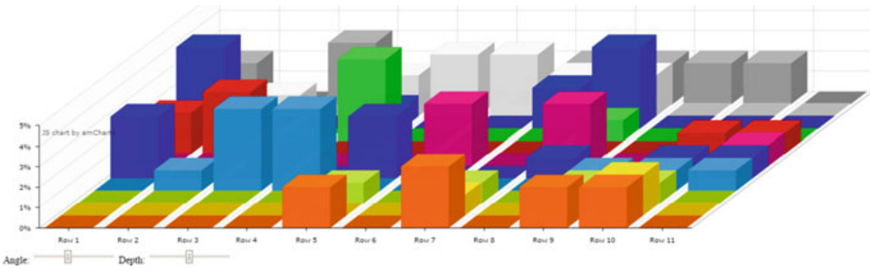


Fig. 7 Optimal 3D plan generated by GA

2.3 Development of Spatial Data Infrastructure (SDI)

Initially, a preliminary development of the methodology had been implemented in MATLAB. After being satisfied with its functionalities, it is again coded in Java. Later the code developed in Java has been used in Netbeans as scriptlets in web pages designed. This has led to the development of a spatial data infrastructure which can be accessed and used by any user to plan any area with respect to different facilities of interests. The main inputs for the program are (i) number of facilities, (ii) number of locations and (iii) specific area. Once these data are entered, depending on the information provided, one by one the area and distance among the facilities are asked. Figures 6 and 7 indicate the generated plan in 2D and 3D, respectively.

3 Results and Discussions

The algorithm developed can be applied to any study area that can be categorized in terms of facility groups. We have applied it on the campus image as it is stated in the above discussion. The output of the algorithm is the image of campus which satisfies all the distance criteria defined in the starting phase of the algorithm. Also certain random inputs are evaluated which generated proper justified plans with respect to the inputs provided. For example, a number of facilities to be planned in an area of 200 m^2 are 4 which must be distributed in 5 localities. Let the four localities have their respective area of 10, 20, 30 and 40 m^2. It is required to consider a buffer area of at least 60% of the total specified area to be planned. Required distance among the facilities is entered by the user within a specified limit with respect to the total area and such that the distance can be accommodated in the area given.

3.1 Future Enhancements

The proposed methodology can be used in future for the planning and development of various cities or villages that are not properly planned and well organized. If it can be implemented on open source, then any user will be able to access it and draw benefits from it.

In future, we can include the following in the list of facilities to be managed

- **Public transport routes**: In this initial algorithm, we have just implemented the management of discrete facilities only. Public transport routes are an integral part of any location. These can consider as a facility and can be fed as input to the same algorithm, but it will face a lot of difficulties in finding the output then. Some modifications to the basic algorithm generated can lead to inclusion of transport facilities as well.
- **Water supply lines**: Once it is made capable for handling transport routes, in a similar manner we can deal with the water supply lines. These lines are generally laid underground, and there are issues like lack of empty space of unconstructed field through which it could be passed and some more. These conditions can be defined as mandatory in the fitness function definition, and this way we can simultaneously mange these lines layout as well.
- **Protecting ecologically sensitive regions**: There is a lot of deforestation, and a concrete jungle is spreading on earth. In this situation, we need to keep in mind the protection of ecological regions more and more. Some enhancements in the

algorithm can also give us the output with all the facilities placed at right place and not destroying the ecological sensitive regions.

- **Developing resources like rainwater harvesting**: We have selected all the study areas for the proposed algorithm from Rajasthan. Rajasthan and many other areas suffer from water problems. Rainwater harvesting can be one solution to the major problem of water scarcity. To implement this on a very large-scale genetic algorithm can be a good platform.

These all were the enhancements related to encompassing more types of facilities into the algorithm and managing them. Another improvement can be providing the facilities desired shapes. Till now we are taking the facilities to be of any possible shape depending on wherever space is available for all the facility units of a facility continuously. But some facilities may require appearing in the desired shape or the shape in which they exist in the actual map/image. The output of the algorithm can hence be further refined if we change it such that it only generates specified shapes for those facilities which cannot be deformed into other shapes due to some reasons.

References

1. Melanie, M.: An Introduction to Genetic Algorithm. MIT Press (1996)
2. Neubauer, A.: Theory of the simple genetic algorithm with α-selection. In: GECCO 2008: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation. ACM (2008)
3. Bajpai, P., Kumar, M.: Genetic algorithm-an approach to solve global optimization problems. Indian J. Comput. Sci. Eng. **1** (2009)
4. Srinivas, M., Patnaik, L.M.: Genetic algorithms: a survey. IEEE J. Comput. **27**(6) (1994)
5. Goldberg, D.E.: Genetic and evolutionary algorithms come of age. Commun. ACM **37**(3) (1994)
6. Jones, G.: Encyclopedia of Computational Chemistry-Genetic and Evolutionary Algorithms (2002)
7. Thede, S.M.: An introduction to genetic algorithms. J. Comput. Sci. Coll. **20**(1) (2004)
8. Forrest, S.: Genetic algorithms: principles of natural selection applied to computation. Science **261**(5123), 872–878 (1993)
9. Chaudhry, S.S., He, S., Chaudhry, P.E.: Solving a class of facility location problems using genetic algorithms. Expert Syst. **20**(1) (2003)
10. Alba, E., Troya, J.M.: A survey of parallel distributed genetic algorithm. Complexity **4**(4) (1999)
11. Oliveto, P.S., Witt, C.: On the analysis of the simple genetic algorithm. In: GECCO 2012: Proceedings of the Fourteenth International Conference on Genetic and Evolutionary Computation Conference. ACM (2012)
12. Preux, P., Talbi, E.-G.: Towards hybrid evolutionary algorithms. Int. Trans. Oper. Res. **6**, 557–570 (1999)
13. Mitchell, M.: Genetic algorithms and Artificial Life. Sci. Am. 114–116 (1993)
14. Sastry, G., Kendall, G.: Search Methodologies-Introductory Tutorials in Optimization and Decision Support Techniques, 1st edn. Springer (2005)
15. Clark, A., Thornton, C.: Trading spaces: computation, representation and limits of uninformed learning. Behav. Brain Sci. **20**, 57–90 (1997)

16. Pandey, H.M., Dixit, A., Mehrotra, D.: Genetic algorithms: concepts, issues and a case study of grammar induction. In: Proceedings of the CUBE International Information Technology Conference. ACM (2012)
17. Silveira, L.R., Tanscheit, R., Vellasco, M.: Quantum-inspired genetic algorithms applied to ordering combinatorial optimization problems. In: WCCI 2012 IEEE World Congress on Computational Intelligence, Brisbane, Australia (2012)
18. Reeves, C.: Genetic algorithms and neighborhood search, evolutionary computing, AISB Workshop. In: Fogarty, T.C. (ed.) Lecture Notes in Computer Science, vol. 865, pp. 115–130. Springer (1994)
19. Črepinšek, M., Liu, S.-H., Mernik, M.: Exploration and exploitation in evolutionary algorithms: a survey. Comput. Surv. (CSUR), ACM **45**(3) (2013)
20. Whitley, D.: An overview of evolutionary algorithms: practical issues and common pitfalls. Inf. Softw. Technol. **43**, 817–831 (2001)
21. Goldberg, D.E., Holland, J.H.: Genetic Algorithms and Machine Learning. Springer (1988)
22. Barret, P.: Facilities Management: Towards Best Practice. Blackwell Science Ltd., Oxford (1995)
23. Becker, F.: The Total Workplace. Van Nostrand Reinhold, New York, NY (1990)
24. Hinks, J.: The creation of a management-by-variance tool for facilities management performance assessment. Facilities **17**(1/2), 31–53 (1999)

Knowledge-Based Approach to Planning: A Case Study-Based Approach

S. Sen, J. Shah and M. Sohoni

1 Introduction

1.1 Background

Planning is a key to the process of development in a society and can be seen through multiple perspectives. While the part of society that is directly affected (and presumably benefited) from the plan, such as a village, plans, owns and is a key stakeholder of the planning project. However, there are multiple players in the process including all citizens, the government machinery, financial agencies, engineering, and other technical participants. The consensus on economic development as put by Barbier [1] pivots on the "process whereby real per capita income of a society increases over a long period of time" subject to stipulations of a non-rising poverty line or inequality of income distribution. Such development (i) is characterized as being only one part of the development of a society, (ii) is quantitatively associated with economic accumulation, (iii) is qualitatively associated with technological and institutional change through innovation, (iv) and should ideally be measureable. A revision of the worldview on development with the emergence of sustainability as a key issue has led to characterizing development goals within the social system (SS), economic system (ES), and the biological and resource system (BS) [2] represented in (Fig. 1). The Millennium Development Goals proposed by the United Nations, which have been adopted by multiple countries, serve as recognition of the need to maximize the com-

S. Sen (✉) · J. Shah
GISE Lab, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai, India
e-mail: sumitssen@gmail.com

M. Sohoni
Department of Computer Science and Engineering, Centre for Technology Alternatives for Rural Development (CTARA), Indian Institute of Technology Bombay, Mumbai, India

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_14

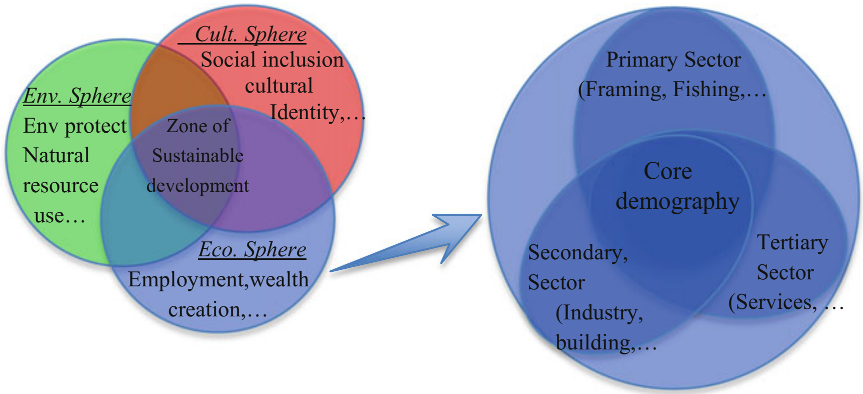


Fig. 1 Concept of economic development and its expanded notion in the context of sustainability across social, economic, and biological systems

mon goals across the different systems through sustainable development strategies [3]. Those ensure cyclical and interactive process of planning, community participation, and action such that concerted activities of communities, governments, public, private agencies, and other partners collectively work toward identified and targeted goals.

Planning is evidentially central to the government's efforts in development and should be seen as a complex and cyclical process (often interactive but not necessarily). Hayek [4] suggested a long time ago that

> … we describe by the word "planning" the complex of interrelated decisions about the allocation of our available resources. All economic activity is in this sense planning; and in any society in which many people collaborate, this planning, whoever does it, will in some measure have to be based on knowledge which, in the first instance, is not given to the planner but to somebody else, which somehow will have to be conveyed to the planner.

He argued that knowledge to every participant in planning is partial and further adds that

> It suggests rather that there is something fundamentally wrong with an approach which habitually disregards an essential part of the phenomena with which we have to deal: the unavoidable imperfection of man's knowledge and the consequent need for a process by which knowledge is constantly communicated and acquired. Any approach, such as that of much of mathematical economics with its simultaneous equations, which in effect starts from the assumption that people's knowledge corresponds with the objective facts of the situation, systematically leaves out what is our main task to explain.

Evidently, the objective facts of the situation used in planning can be sourced from data (mostly spatially grounded) and its contextual understanding.

In this paper, we argue that a knowledge-based approach to planning, especially in the cases where spatial facts are of consequence, is critical for achieving developmental goals. We review background literature on the topic of spatial planning and the use of geospatial information for development in rural communities. Using a case study on water resource planning in Maharashtra, we outline a scientific approach and information sources necessary therein. Thereafter, we discuss knowledge modeling needs to ensure a holistic outlook toward planning. We argue how geospatial ontologies can be leveraged to promote knowledge-based approach and greater information sharing that remains critical to achieving development goals across the three systems in different communities.

2 Developmental Planning Theories

Development planning, as argued by Friedmann in his seminal paper "Planning in the Public Domain: Discourse and Praxis," is an inherently moral practice and affects the way we live [5]. He derives at his popular definition of planning as the link between knowledge and action. Planning theorists refer to the task of planning without a qualification of the level or profession to which it is applied: urban or rural, regional or national, economic or transportation, land use or community.

Spatial planning practices are considered an important part of planning theory. The ideal object of spatial planning is the totality of human and social activities in space, to create desirable built and natural environments [6]. Spatial planning is one of the critical areas of governance, in general, and development planning, in particular. It is carried out by governments at local, district, state, and central levels, and sharing such plans is a key to concerted efforts for both urban and rural development objectives. In another example, the identification of floodplain boundaries in the urban district has been identified as a key item on disaster mitigation, especially in the monsoon. Interestingly, such boundaries are also key to optimizing public good in irrigation planning as well as drinking water supply, which is critical in the drier periods.

From operative research to technical applications, organization and sharing of knowledge is a crucial step. In the field of "planning," intended not only as a theory but mainly as a practice, scientists and technicians develop their rational exercise on a multisectorial knowledge framework. In the process, it always includes several active bodies, with different functions and responsibilities. Such an inclusion is mainly increased by the application of participative techniques (based on Internet and ICT e-government tools), and the role of communication, in planning process, has considerably increased during last decade. Communication requires a sharing of ontologies between communicating parties [7], and it also needs new tools in order to facilitate a bottom-up participation process.

The theme of knowledge management has assumed relevant importance in planning processes, especially due to the contribution of GIS, spatial data infrastructures (SDIs), and WEB-GIS. These technologies have added a new dimension toward information-driven approach to planning as opposed to selective data-driven approaches. This posits the possibility of a coordinated approaches and a more evolved knowledge-based approach. Such starting considerations also lay the foundations for the role of an ontological approach in developing planning theories.

2.1 Information Structures for Spatial Planning

Planning processes requires multiple sources of information, and spatial data-based systems such as geographic information systems (GISs) and spatial decision support systems (SDSSs) are central to such infrastructure. Examples of spatial expert systems in numerous domains have been discussed by Demetriou et al. [8]. It is evident that most systems develop with a base layer of maps, which includes cadastral maps and demographics. Further layers including layers about infrastructure and domain-relevant information are added. A schematic of such information sources is shown in Fig. 2.

Integration of expert systems and GIS has been employed in planning projects such as land consolidation planning. The central task of such systems is to guide policy makers on the implications of a development plan as well as make suggestions to the plan, given the objectives and constraints [8]. We shall see how such constraint-based optimization requirements help us to formulate a mathematical model of planning tasks. It is also important to point out at this stage that every planning process is unique in nature but has a similar structure to other planning tasks. While rules and heuristics used for a particular planning task in a particular location may differ, the logic of planning remains consistent. These tend to be based focused on theories like central place theory and thus tend to be similar, speaking in terms of the concepts and approaches.

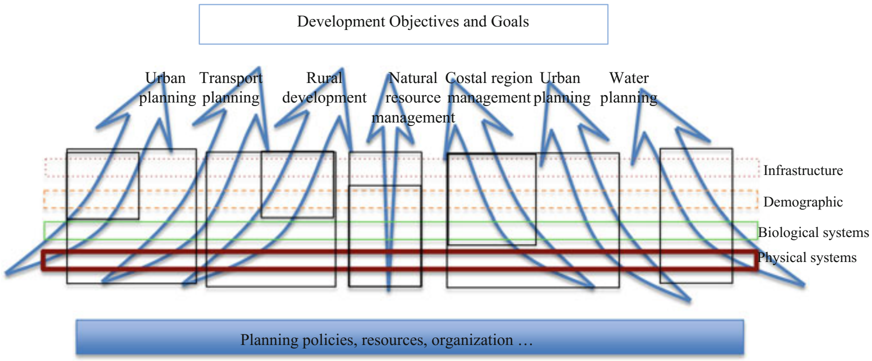


Fig. 2 Using multiple layers of information in spatial planning based on application requirements. Blocks across layers represent systems that combine information from multiple sources that are often geospatially referenced. Expert systems that employ intelligent processing and real-time systems, are part of such systems and help to achieve a planning goal

2.2 *Spatial Data as a Central Link*

While all planning data pertains to a given region of interest, it is linked to several other actions that need to be linked to develop, implement, or evaluate a plan. Such data can only be linked through the spatial context. For example, land use planning and water resource planning can be linked based on water supply reaching every land parcel (by pipelines or canals). Similarly, connectivity and mobility between villages can be established on the basis of transportation networks available at the location. Several classical theories had been proposed to elaborate the spatial link in planning tasks. Such geographical theories model dynamics related to location of resources.

Location theory has focused primarily on developing formal mathematical models of the optimal location of a facility given the costs of transporting input materials and final products or services.

Spatial Competition theory argues that spatial proximity gives facilities reachability power, because people living nearby would have a marked preference for their services given that additional transportation costs can be avoided [9].

Central Place theory pivots around the basic idea that the relative size of a facility's influence area, defined as the territory over which it provides its services, is determined by the combined influence of scale economies and transportation costs. It postulates that if scale economies are strongly relative to transportation costs, services provided from a single location would appear more feasible [10].

In more recent work, Friedmann's **center-periphery model** [11], Porter's geographic clustering theories [12], Krugman's **core-periphery** model [13], and **new economic geography model** of Fan et al. [14] are examples of how location serves as the basis of modeling economic development.

Invariably, the spatial nature of planning tasks requires the use of spatial data, which is increasingly served through spatial data infrastructures (SDIs) and public participation GIS (PPGIS). Simulation and modeling combined with advanced analytics using such data serves as the information backbone of planning projects. Data and applications used in such projects are

(1) Share across multiple planning projects (in most cases) and impose interoperability requirements.
(2) Combine both real-time and historical data in a manner where the place names and boundaries do not easily synchronize.
(3) Lessons learnt from pervious planning processes in the same region have significant value even if they are functionally different.

Very often, planning projects are burdened with higher costs of information because of the inability to share and reuse data. The maturity of planning processes in terms of their use of information as modeled on the capability maturity model (CMM) [15] is shown in Table 1 and illustrates the path to more evolved systems.

Table 1 Capability maturity model for planning processes

Maturity level	Characteristics	Tools and techniques
Data-driven (definition stage)	Data sources identified and secured Systemic approach to decision making Some reports are available and transparency	Desktop applications, flat files, folder-based storage
Information-driven (Managed stage)	Database management in place, some analytics available, data sharing begins but interoperability restricted, data models available, some metadata exists	DBMS, enterprise applications but limited integration. Analytic tools
Knowledge-driven	Data sharing, optimal analytics, DB models reused, learning from previous projects, continuous system evolution	Ontology-based data sharing, enterprise wide applications, advanced analytics

2.3 Geospatial Ontologies in Support of Planning Systems

Geospatial ontologies provide a specification of the entities and action in society and thus help to resolve semantic ambiguities [16]. They are often derived from textual descriptions and database models. They sometimes serve as statements of facts or a knowledge base and can be stored in multiple ways such as description logic, controlled natural language or based on standards such as OWL[1] and RDF.[2]

While ontologies can be employed for synergistic use of disparate geospatial information [17], their utility beyond interoperability requirements can be seen in deriving taxonomies [18], question answering, and intelligent searches. Furthermore, their utility in ascertaining factual anomalies within a system using ontology reasoning as well as reducing redundancies by inferring equivalent classes has already been shown [19]. We summarize the possible applications of geospatial ontologies in planning processes in Table 2.

Ontologies for spatial planning have been proposed as a solution for problems in knowledge management tasks related to development plans [20] and achieving interoperability within other systems of the geospatial cyberinfrastructure. Ontologies are used to verify technical plans in construction projects by representing plans as knowledge and applying reasoning to ensure compliance, reuse of previously defined concepts, and pruning inconsistencies [21]. Murgante and Garramone [22] discuss the opportunities that enhanced spatial planning using knowledge management especially using geospatial ontologies especially in linguistic diversity. They argue that

[1] https://www.w3.org/OWL/.

[2] https://www.w3.org/RDF/

Table 2 Some application areas of ontologies in spatial planning

Area	Body of knowledge/scientific theories involved	Top potential use
Land use planning	Zoning, polyrationality, guidelines of Food and Agriculture Organization of UN	Creating standardized and communicable plans across agencies
Transportation planning	Rational planning, transit-oriented development, accessibility, TPP's Portfolio of Technical Knowledge	Inter-regional interoperability of transit systems
Forest management and planning	Sustainable forestry, activity theory, guidelines of Food and Agriculture Organization of UN	Sharing sustainable practices of indigenous societies
Urban planning	Blueprint planning, advocacy planning, participatory/synoptic planning, communicative approach, UCGIS Body of Knowledge	Participatory urban planning
Agriculture planning	Location theory of agriculture, climate-resilient agriculture, Indian Council of Agricultural Research knowledge base	Info sharing and automated question answering for farmers
Coastal area planning	Strategic collaborative planning, Integrated Coastal Zone Management, sustainable fishing, Global Water Partnership guidelines	Interoperability of data across marine, land use, demography
Disaster management	Contingency theory, risk mitigation, actuarial Science, National Disaster Management Authority guidelines	Integration of multiple data sources through ontology-based mediation
Water supply planning	Network design, hydrology models, World Health Organization guidelines	Evaluation of completeness, utility, inconsistencies of information used

ontology-based tools can better analyze the potential of a plan, and they can support regulatory dialogue between decision makers and citizens. Lancura et al. [23] have recently reported how linguistic barrier, different visual representations, and data schemas have been overcome by using semantically enriched geospatial map data collected in the context of a multidisciplinary study focused on a coastal area.

2.4 Mapping Ontologies Across Information Communities

Planners interact with different sections of society. The needs of these sections are often expressed in terms that are diverse and specific to the society itself. For example, agrarian communities associate provisioning of water with crop cycles as opposed to the provisioning of water around the year for drinking water. Contrastingly, engineers and domain specialists such as irrigation experts and/or water supply consultants prefer to use technical terms associated with their sector, which often have *disjunctures*. It has been reported that for regional planning to succeed in critical areas such as drinking water and climate-resilient agriculture, it is necessary to bring to the table both the demand-side issues of scarcity and coverage, with various supply-side alternatives such as large surface water schemes and single village groundwater schemes [24]. Given the multiple ontologies of water [25] and the challenges that such plurality presents in water governance [26], it is inevitable that strategies to map terminologies and needs of different communities are developed.

While participatory decision making across information communities may be a complex exercise [27], some level of coherence and integration can be achieved with the spatiotemporal grounding of real-world entities and spaces [28].

3 Case Study of Water Resource Planning

Water resources are among the most critical resource of any society, and planning of water resource use is often the center of many economic development projects. Planning, developing, and managing water resources to ensure adequate, inexpensive, and sustainable supplies and qualities of water for both humans and natural ecosystems can succeed if we recognize and address the causal socioeconomic factors, such as inadequate education, corruption, population pressures, and poverty.

Natural water resources such as groundwater and surface water sources have been augmented with engineering accomplishments such as canals, pipelines, desalination plants, reservoirs, and even supply as bulk cargo. Global and water balance studies that report on the availability of water at different time horizons are closely linked to efficient management of water resources. We discuss in this section a redesign of water pipe network for a rural part of Maharashtra that had faced drinking water shortage. The objective of the study was to derive at a design of the network that ensures reliable drinking water supply at lower cost and under given constraints [29]. We shall outline the knowledge gaps in the planning process using the case study and list the advantages of using geospatial ontologies in such a planning project.

3.1 Redesigning Khardi Rural Piped Water Network Scheme for Sustainability

Khardi is a small village with a population of about 5000 in Shahapur taluka of Thane district in the state of Maharashtra at a distance of about 90 kms from Mumbai in the vicinity of Bhatsa reservoir. A water supply scheme that depended on the tailwater of the Bhatsa dam existed for Khardi and nearby villages, for many years but needed to be replaced as the pipelines failed and these villages became depended on tankers (bulk transport) for drinking water.

The schematic of a network solution is as shown in Fig. 3a which shows how water flows from the source to a hamlet. A water treatment plant supplies water to a mass balance reservoir (MBR), which ensures balancing of water supply to an elevated supply reservoir (ESR) or a ground supply reservoir (GSR) based on demand around the year. A replacement scheme that used Bhatsa as the source was proposed without detailed study. However, this scheme required a high capital and operating costs. It would also have kept nearby villages dependent on tankers. Finally, using spatial data available for the region, it was seen that water from an alternate source at a higher elevation as shown in Fig. 3b was available at Kundan Dam which was a much smaller dam and mainly used for agricultural purposes. This enabled the reduction of both the capital cost of the redesign and the energy and operation/maintenance costs of the network. A complete description of this study can be found in the published report [29].

3.2 Knowledge Gaps in Planning Processes of Water Pipeline Networks

In the Khardi redesign case, it is evident that in the earlier two planning processes data was incomplete and inadequate. The availability of alternate water resources was not explored, and the feasibility of such other sources of water, which satisfy governmental protocols for a water supply scheme, was not analyzed. Figure 4 illustrates the protocol for designing a multivillage water supply scheme with annotations of different concepts used and a reference to where such data could be sourced.

Thus, the Khardi example shows that

1. Multiple concepts were used in planning, and they need to be specified accurately to ensure that the data used in the process is adequate.
2. When the search for data is incomplete, the plans are suboptimal and can lead to higher costs.
3. Planning processes are composed of rules, some of which have empirical basis and tend to be revised. The protocol itself may evolve, and hence, the need to update the definition of concepts such as MBR, WTP, demand, and ESR/GSR is subject to updates.
4. It is understandable that such projects are required and conducted at many places and possibly for most villages in the vicinity. While software for a similar study

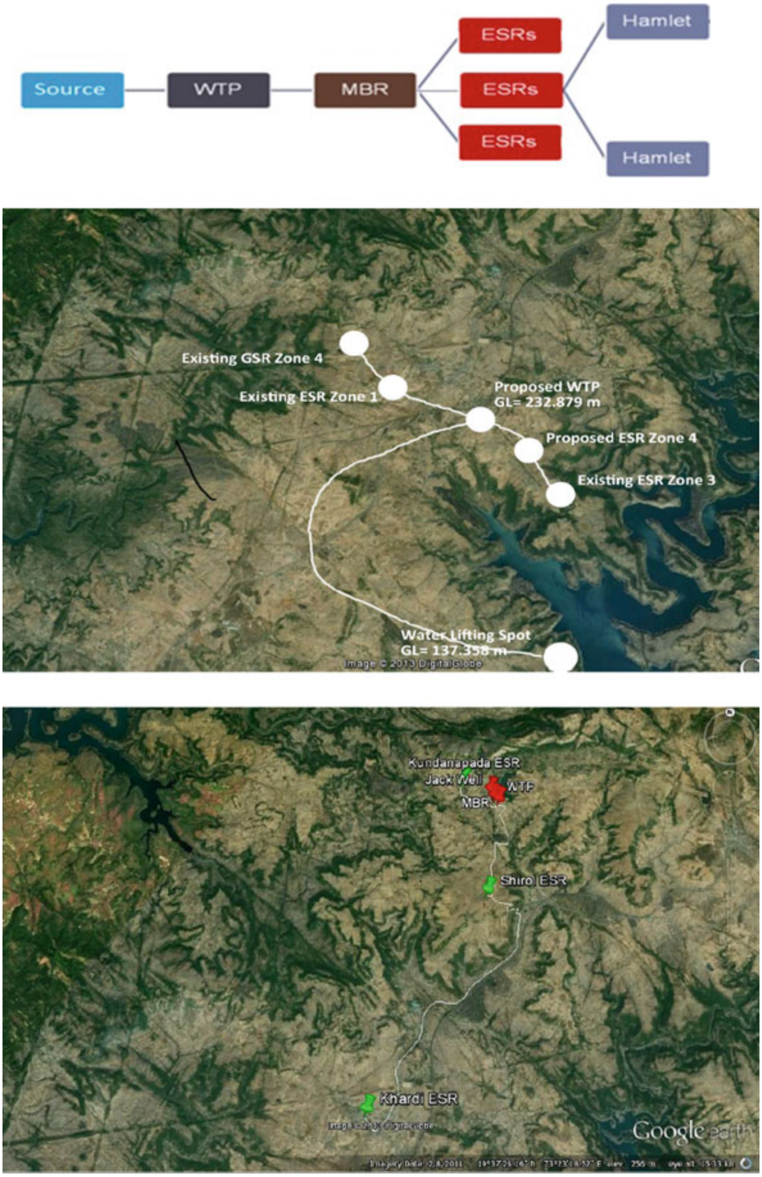


Fig. 3 Redesign of Khardi multivillage water supply scheme. **a** Schematic display of a water supply network components according to the planning protocol. **b** Replacement network and **c** alternate solution using Kundan dam as the source

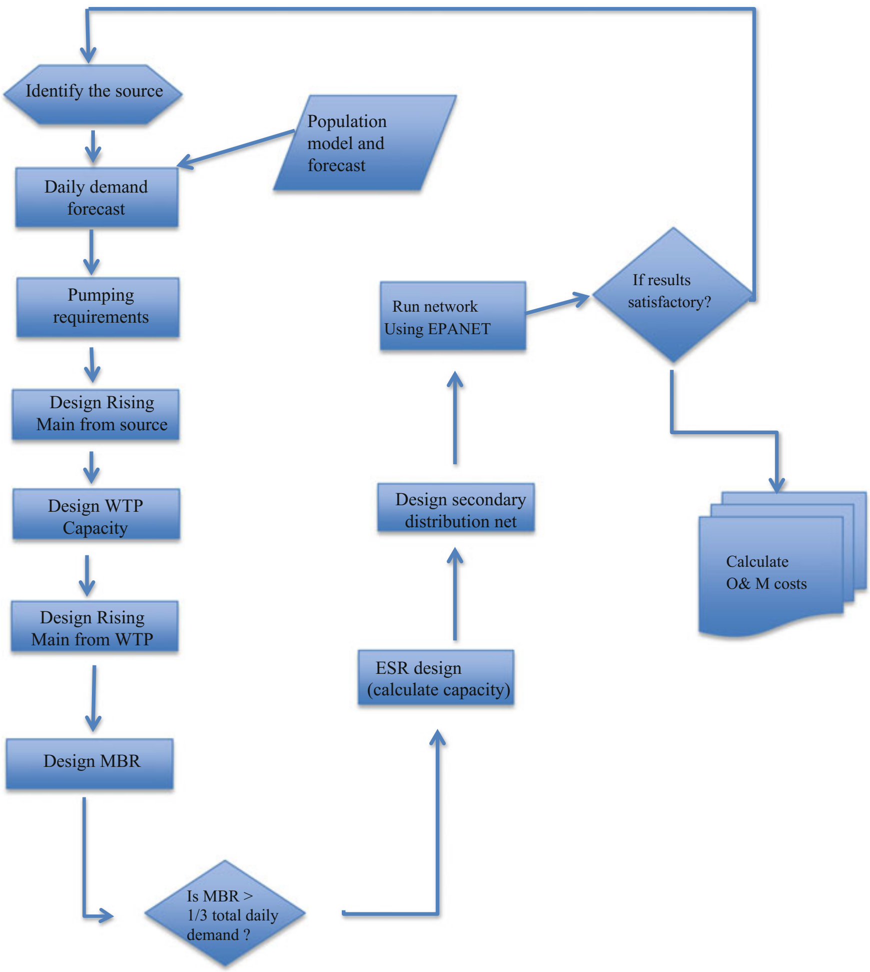


Fig. 4 Structure of the (Maharashtra Jeevan Pradhikaran) protocol for multivillage water supply scheme

is available [30], it is possible that the knowledge required to conduct a similar study would still require domain expertise and thumb rules that could be easily specified using informal ontologies.

5. It is evident that the location data along with elevation required to run an optimization routine for the redesign required integration of the spatial information with an optimization tool and also hydrological network planning or simulation such as BRANCH and EPANET.

4 Knowledge Gaps in Planning

Planning processes rely on multiple sources of information, and spatial information systems (such as GIS) need to be integrated with statistical, network design, optimization, and other tools; such integration is often a challenge. However, a bigger challenge is the integration of data from heterogeneous sources. Davoudi [4] proposes that planning is a practice of knowing, which is both provisional and situated in space and time given the success of *evidence-based* practices. The knowledge, however, should be understood in the context of knowing what, how, and to what end in combination with the knowledge of doing it as well. It is conceivable that planning requires pragmatic judgment as well as interactive problem solving. We argue that *knowing what* and *knowing how* are related to the specification of the concepts used in the planning processes and therefore essential to any planning project. Furthermore, we believe that such knowledge is also critical in the communication of the plan to partners and stakeholders.

We believe that theoretical framework of planning should be specified as ontologies along with commonly used upper-level concepts that could be used to derive specific project-related concepts.

4.1 A Mathematical Model of Planning

Planning processes can be seen as a resource allocation model or a risk mitigation or distribution exercise based on the goals of the task at hand. For example, in the water resources domain, the maximization of allocation of resources (water) as per the demand of consumers (based on socioeconomic factors) is required subject to technical and environmental feasibility (engineering and biophysical factors). This can be modeled as a general function

$$\text{Maximize}\, g(x) \tag{1}$$

subject to

$$f_i(x) \leq b_i \forall i \tag{2}$$

$$\text{x conditions} \tag{3}$$

where x is a vector of unknown decision variables, x $= [x_1, x_2, x_3, \ldots, x_n]$, g() and f_i() are functions of x, and b_i is a coefficient that limits the value of each associated function f_i(). The conditions on x (Eq. 3) typically consist of some combination of real, integer, or binary requirements; nonnegativity stipulations; or both. They may be reflective of statutory or legal factors, which restrict governance and policies in addition to the techno-feasibility factors that represent supply-side constraints.

What makes an optimization problem (Eqs. 1, 2, and 3) spatial is the specification(s) of decision variables, coefficients, functions, and constraining conditions. Because geography and space become part of the model structure by design, this makes the resulting optimization problem unique. It is important to note that geographic variables can be continuous, involving an infinite number of potential locations, or discrete, associated with a finite number of possibilities. Much of the variables used in spatial planning are difficult to observe or calculate. For example, allocation of new schools in rural areas often requires to optimize travel times for students but the computation of travel times requires information about roads and pedestrian pathways that are not easily obtained. Similarly, in the context of allocation of water for drinking or agriculture, it is difficult (almost impossible) to ascertain weights to different needs in a society. As discussed by Storme and Witlox [31], location allocation models are heavily dependent on the quality and availability of input data.

A similar mathematical model of planning can be seen in risk mitigation or distribution such that no region is subjected to a higher risk [32]. For example, in a flood risk mitigation plan, it is necessary to minimize risks and hence the objective function (1) changes to

$$\text{Minimize}\, h(\mathrm{x}) \tag{4}$$

subject to the constraints similar to those expressed in Eqs. (2) and (3). Risk mitigation models are more difficult to formulate because input data in such cases is often based on perception, and obtaining a shared value of perceived risk can be impossible. There are further difficulties in the interpretation of risks and the formulation of the objective function. For example, in a urban flood risk mitigation model, it is extremely difficult to formulate a common definition of flood risk and the determination of blue lines (where inundation has been reported in past 25 years and with 71% chance of reoccurrence in 30 years) and red lines (where inundation has been reported in past 25 years and with 26% chance of reoccurrence in 30 years) [33].

4.2 *Sharing of Data and Evolution of Knowledge*

Du et al. [17] have discussed that with rapid development of crowd-sourced geospatial data, there is a stronger need to match formal and informal real-life geospatial ontologies, at both terminology level and instance level, ensuring that overall information is logically coherent and consistent. However, success in ontology matching is only limited to the availability of ontologies for every dataset and a completely automated matching is deemed impractical leading to various semi-automated methods [19]. While partial intervention in the matching of terms across domains and data sources is feasible and understandable, large-scale translation and mediation remain a challenge [34] and also limited by our practical capability to access cross-domain datasets.

It is important to note that generation of ontologies and the ontology-matching problem faces the dilemma of using either a top-down or a bottom-up approach [35]. Given the role of the planner and the mathematical model of planning that we have discussed, it is justified to develop a partial top-down approach that could be matched based on similarity to bottom-up ontologies derived from specific tasks and projects.

Ontologies can be best described as a set

$$O := \{C, I, R, A\} \quad (5)$$

where

O is an ontology for a domain of interest;
C is a set of concepts/classes in a domain;
I is a set of individuals/instances in a domain;
R is a set of relations among concepts, relations, and objects;
A is a set of axioms holding among concepts, relations, and individual; axioms provide explicit logical assertions about these three elements.

Although the set of concepts may remain relatively stable over time if they are adequately defined, the set of axioms and relations may undergo revision over time [36]. As discussed in the case study, knowledge evolves over time and newer axioms are likely to replace earlier ones.

Some initial approaches to building ontologies for the water resources domain can be seen in the work of Elag and Goodall [37]. A related ontology of flood forecasting has been discussed by Agresta et al. [38] and can be utilized for creating a more comprehensive knowledge base.

5 Knowledge-Based Approach to Planning Tasks

Planning is shown to be based on knowledge that the planner can obtain either through information systems or through their own acumen and pragmatic understating from earlier plans. Information systems such as GIS and expert systems have evolved greatly and are now interfaced with big data and analytics. Using pooled computational resources, on-demand services, and elasticity of processing, large-scale tasks in sustainable development such as obtaining land use and land cover from satellite imagery and simulation of natural phenomena can now be accessed using the cloud [39]. Although the nature of information used in planning and underlying techniques remain unaltered, the opportunities for rapid conversion of data into information and translation of information to knowledge have grown immensely.

We propose to set up a knowledge-based framework for planning processes based on

1. Enumerate concepts or evolving a shared vocabulary (i.e., the key concepts, ways of obtaining, validating, interrelationships) in a particular projects.

2. Mediate or integrate workflows and data between cross-functional projects.
3. Compare the conceptual framework for completeness or novelty of planning projects and detect missing components based on similar previous projects.
4. Comprehend the evolution of new concepts in spatial planning and the semantic evolution over time.

5.1 Spatial Basis for Planning Ontologies

The centrality of spatial relations and axioms in planning theory as discussed over many decades has already been discussed. However, the consensus on upper-level concepts in geospatial sciences that can be easily used across domains and planning activities is still evolving [40]. Problems of traversing geographic scales and domains are pragmatic challenges that build on fundamental issues of geospatial information sciences [41].

It is reasonable to hypothesize that specification of geospatial facts (as set of axioms) [28] is possible by

1. Identifying upper-level concepts (corresponding to C in Eq. 5);
2. Specifying relations between such concepts; and
3. Specifying axioms known about such concepts and relations (in description logic).

While such knowledge would be machine readable and amenable to automated reasoning, they would also be translatable to readable text and vice versa [42]. Planning artifacts such as protocols, handbooks, guidelines, and previous plans provide a wealth of information for creation of project-level knowledge base, which can be integrated with other projects and upper-level concepts of planning. An example of such an approach for a limited planning domain can be seen in the work of Elag and Goodall [37].

The task of creating planning ontologies can be done by taking up use cases. A use case-based approach in this regard would not only create the ontologies of the particular case but also ensure integration of the knowledge infrastructure. It would also serve as a learning model for future projects and assist in the documentation of knowledge evolution over time. Such knowledge evolution is exemplified by the addition of new concepts and phasing out of older ones.

5.2 Tool for Ontology-Based Integration of Data

Planning artifacts are expected to contain development planning concepts that typically would align with a subset of the entire planning ontology. We propose to develop software based on ontological reasoning [36] WordNet-based similarity approaches [43] to assess the conceptual relatedness of an artifact will be employed. The tool would thus help to evaluate for missing/skipped/overlooked concepts in a document. It would be useful to evaluate similarity of documents based on their conceptual structure. In a broader view to comprehend the development of the knowledgescape in planning, the tool would help to analyze newly evolved concepts and unused ones over time. Most importantly, the tool will attempt to provide planner with planning concepts used in other functional groups (or locations). It is expected that while the tool will enrich the planning process itself based on its suggestions, the tool will also enrich the ontologies as further concepts are integrated, thereby expanding the domain knowledge.

6 Future Perspectives

We have discussed the need for a knowledge-based approach in the context of an evolving notion of developmental planning. We have demonstrated the competitive nature of theories, and using a case study, we have demonstrated that knowledge gaps can cause impediments in achieving optimal conditions in planning processes.

We believe that spatial planning projects should follow a knowledge-based approach that is both scientific and mature. It should be able to harness the power of advanced spatial information systems while being able to utilize spatial theories in planning that continue to serve the modern day planner. An ontology tool, which captures multifarious concepts used in planning based on formal and informal specifications, provides an invaluable help to a knowledge-based approach in spatial planning that is advocated in this paper.

References

1. Barbier, E.B.: The concept of sustainable economic development. Environ. Conserv. **14**(2), 101–110 (1987)
2. Lintz, G.: The aspect of space in the concept of sustainable development: overview and consequences for research. In: ERSA Conference Papers. European Regional Science Association (1998)
3. Persson, Å., Weitz, N., Nilsson, M.: Follow-up and review of the sustainable development goals: alignment versus internalization. Rev. Eur. Comp. Int. Environ. Law **25**(1), 59–68 (2016)
4. Davoudi, S.: Planning as practice of knowing. Plan. Theor. **14**(3), 316–331 (2015)
5. Friedmann, J.: Planning in the Public Domain: From Knowledge to Action. Princeton University Press (1987)
6. Alexander, E.R.: There is no planning—only planning practices: notes for spatial planning theories. Plan. Theor. **15**(1), 91–103 (2016)
7. Mark, D.M., Smith, B., Tversky, B.: Ontology and geographic objects: an empirical study of cognitive categorization. In: International Conference on Spatial Information Theory. Springer (1999)
8. Demetriou, D., See, L., Stillwell, J.: Expert systems for planning and spatial decision support. In: GeoComputation. CRC Press (2014)
9. Hotelling, H.: Stability in competition. Econ. J. **39**(153), 41–57 (1929)

10. Lösch, A.: The Economics of Location: Translated from the Second rev. German ed. by William H. Woglom with the Assistance of Wolfgang F. Stolper. Yale University Press (1954)
11. Friedmann, J.: A general theory of polarized development. Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL) (1967)
12. Porter, M.E.: Clusters and the New Economics of Competition, vol. 76, Harvard Business Review Boston (1998)
13. Krugman, P.: The Self-Organizing Economy (1996)
14. Fan, W., Treyz, F., Treyz, G.: An evolutionary new economic geography model. J. Reg. Sci. **40**(4), 671–695 (2000)
15. Paulk, M.C., et al.: Capability maturity model, version 1.1. IEEE Softw. **10**(4), 18–27 (1993)
16. Sen, S.: Representing Geospatial Concepts: Activities or Entities? Universal Ontology of Geographic Space: Semantic Enrichment for Spatial Data: Semantic Enrichment for Spatial Data, p. 124 (2012)
17. Du, H., et al.: A method for matching crowd-sourced and authoritative geospatial data. Trans. GIS **21**(2), 406–427 (2017)
18. Tambassi, T.: From a geographical perspective: spatial turn, taxonomies and geo-ontologies. In: The Philosophy of Geo-Ontologies, pp. 27–36. Springer (2018)
19. Sen, S., et al.: Framework of semantic interoperability using geospatial ontologies. J. Geomatics, India, 71–76 (2007)
20. Murgante, B., Scorza, F.: Ontology and spatial planning. In: International Conference on Computational Science and Its Applications. Springer (2011)
21. Zhong, B., et al.: An ontological approach for technical plan definition and verification in construction. Autom. Constr. **55**, 47–57 (2015)
22. Murgante, B., Garramone, V.: Web 3.0 and knowledge management: opportunities for spatial planning and decision making. In: International Conference on Computational Science and Its Applications. Springer (2013)
23. Lanucara, S., et al.: Interoperable sharing and visualization of geological data and instruments: a proof of concept. In: International Conference on Computational Science and Its Applications. Springer (2017)
24. Prasad, P., Mishra, V., Sohoni, M.: Reforming rural drinking water schemes. Econ. Political Wkl **49**(19), 59 (2014)
25. Barnes, J., Alatout, S.: Water worlds: introduction to the special issue of social studies of science. Soc. Stud. Sci. **42**(4), 483–488 (2012)
26. Yates, J.S., Harris, L.M., Wilson, N.J.: Multiple ontologies of water: politics, conflict and implications for governance. Environ. Plan. D Soc. Space **35**(5), 797–815 (2017)
27. Honkalaskar, V., Sohoni, M., Bhandarkar, U.: Selection of development agenda with the community by the generation of a shared understanding. J. Rural Commun. Dev. **12**(1) (2017)
28. Bittner, T., Donnelly, M., Smith, B.: A spatio-temporal ontology for geographic information integration. Int. J. Geogr. Inf. Sci. **23**(6), 765–798 (2009)
29. Choudhary, V., et al.: Redesigning khardi rural piped water network scheme for sustainability. Technical Report No. TR-CSE-2013-56 (2013). Department of Computer Science and Engineering, IIT Bombay
30. Hooda, N., Damani, O.: A system for optimal design of pressure constrained branched piped water networks. Proced. Eng. **186**, 349–356 (2017)
31. Storme, T., Witlox, F.: Location-allocation models. The international encyclopedia of geography
32. Jabareen, Y.: Planning the resilient city: concepts and strategies for coping with climate change and environmental risk. Cities **31**, 220–229 (2013)
33. Aeron, A., et al.: Advancements in flood risk mapping systems: a brief review. i-Manager's J. Civil Eng. **2**(1), 40 (2011)
34. Horrocks, I., et al.: Using semantic technology to tame the data variety challenge. IEEE Int. Comput. **20**(6), 62–66 (2016)
35. Sen, S., et al.: Schema matching as a step towards interoperability: experiments in the Indian NSDI. In: Beyond Spatial Data Infrastructures International Workshop at the GIScience (2006)

36. Baader, F., et al.: Introduction to Description Logic. Cambridge University Press (2017)
37. Elag, M., Goodall, J.L.: An ontology for component-based models of water resource systems. Water Resour. Res. **49**(8), 5077–5091 (2013)
38. Agresta, A., et al.: An ontology framework for flooding forecasting. In: International Conference on Computational Science and Its Applications. Springer (2014)
39. Yang, C., et al.: Utilizing cloud computing to address big geospatial data challenges. Comput. Environ. Urban Syst. **61**, 120–128 (2017)
40. Milz, D., et al.: Reconsidering scale: using geographic information systems to support spatial planning conversations. Plan. Pract. Res., 1–18 (2017)
41. Mark, D., et al.: Ontological foundations for geographic information science. Res. Chall. Geogr. Inf. Sci., 335–350 (2004)
42. Gyawali, B., et al.: Mapping natural language to description logic. In: European Semantic Web Conference. Springer (2017)
43. Wan, S., Angryk, R.A.: Measuring semantic similarity using wordnet-based context vectors. In: IEEE International Conference on Systems, Man and Cybernetics, ISIC. IEEE (2007)

Service Oriented Architecture and Spatial Data Integration for Agro Advisory Systems

S. Chaudhary and V. Kumar

1 Introduction

Economic growth of a country highly depend on crop production and affects a very large portion of world's population and is a major cause of poverty and deceleration of development. Main reason for deceleration in agriculture growth is declining investment in agricultural research and development combined with inefficiency of institutions providing inputs including rural credit and extension services [1]. Other factors affecting economical and ecological cries are lack of supply chain, rural infrastructure and poor policies. The decisions made on the basis of efficient spatial data analytics and precision agriculture can increase profit in comparison to traditional framing practices. Target of agriculture growth and poverty alleviation can be achieved by efficient utilization of rain-fed and other less endowed areas. A system is required to process farmer's queries and which should be able to generate location specific recommendations. Agricultural research will increasingly be required to address location specific problems faced by the communities.

We are experiencing tremendous growth of data now a days due to increasing use of Internet and mobile applications such as social media, GPS based weather forecast and search engines. Typically spatial data has different features and geographic information associations. Spatial data captured through diverse sources and mobile applications (i.e. GPS, sensors, satellite etc.) can be utilized in a wide range of applications for decision making. The development of expert information system is possible only if the data is integrated and analyzed in a scientific manner. The most challenging task in spatial data processing is integration, feature extraction and

S. Chaudhary (✉)
School of Engineering and Applied Science, Ahmedabad University,
Ahmedabad, India
e-mail: sanjay.chaudhary@ahduni.edu.in

V. Kumar
e-mail: vikas.tyagi@ahduni.edu.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_15

interpretation. The data captured through Global Positioning Systems (GPS), Sensors and Satellites have heterogeneous formats (XML, GML, CSV, map etc.) and due to involvement of many organization it becomes challenging to integrate and process. Spatial data integration is an essential and challenging task for developing cross domain applications. Service Oriented Architecture provide technical support for development of cross domain applications in a heterogeneous environment following well accepted and stable standards. In our project, we have used W3C and OGC recommended standards for developing RESTful web service, data integration and Information dissemination. We have developed android application for communicating our generated recommendations to the end users. Our focus in this work is to develop Service Oriented Architecture for Spatial Data Integration so that useful recommendations can be generated for farmers by using the data captured from diverse sources.

1.1 Motivation

Over the past few years, huge volume of data is being generated by different kinds of devices using interactive mobile applications. Information consumption and dissemination is very convenient now a days using mobile devices. Although the increasing use of Internet in this information age accelerated the development of mobile applications; there is no potential decision support system application for agriculture and rural development which is highly dependent on other domains i.e. soil, weather, geology and environment science. Spatial data is very important for taking significant valuable decisions in farming practices. There are heterogeneous sources like agriculture universities, meteorological, soil, geological, irrigation and water resource departments, gathering data in different formats so it becomes challenging to integrate the spatial data and obtain interoperability. Once we achieve data integration, it becomes feasible to develop inter-operable decision support systems for agriculture. The decisions and recommendations can be disseminated to the end users through mobile applications. Thus, spatial data integration is essential for agriculture and rural development.

1.2 Service Oriented Architecture

Web service is a mechanism to enable access of one or more capabilities of software components using a prescribed interface and consistent with constraints and policies as specified by description. Web services are the software components which can be invoked using XML artifacts and facilitates the described functionalities to the user over Internet. The development of cross domain applications using geo-spatial data is possible through web services.

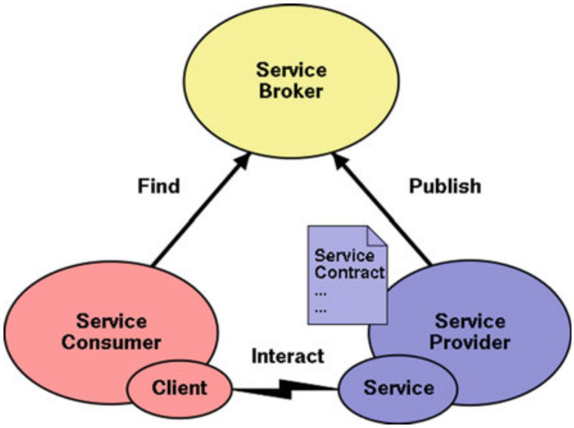


Fig. 1 Service Oriented Architecture (SOA) [2]

Service Oriented Architecture (SOA) is a software design paradigm, where services are developed to implement the components of an application which can communicate using protocols over a network. The basic principles of service oriented architecture are independent of vendors, products and technologies. Web service development is based on WS standards by word wide web consortium. A service is a discrete unit of functionality that can be accessed remotely, acted upon and updated independently, such as retrieving a cheapest flight on-line. A web service has four essential properties according to one of many definitions of SOA:

- **Atomicity** A web service logically represents a business activity with a specified outcome.
- **Containment** A web service is self contained with getter and setter methods.
- **Transparency** A web service is a black box for its consumer or user.
- **Composition** A web service may consist of other underlying services.

Different components of SOA are shown in Fig. 1. Essentially there are three components named service provider, consumer and broker. The responsibility of these components are to publish, consume and register the services.

1.3 Geographic Information System (GIS)

GIS is a combination of hardware, software and geographic data. I It is used for capturing, managing, analyzing, and displaying all forms of geographically referenced information. Intelligent decisions can be made by processing spatial data using GIS [3]. In agriculture, GIS can be used for storing data such as soil type, nutrient levels, temperature and administrative boundaries.The information is maintained in form of layers and assign that information to the specific geo-referenced field location. A fully functional GIS can be used to analyze characteristics between layers to develop application maps. Spatial analysis, the study of geographic features, and the relationships that exist among them can be applied to many areas of the agricultural practices. GIS has capabilities to collect, manage, analyze, and share huge amount of agricultural data to aid in discovering and establishing sustainable agriculture practices.

1.4 Knowledge Base

Knowledge of a particular domain can be represented in form of ontologies and hence, ontology repositories can be used for development of agricultural knowledge base. Unfortunately, the repositories for plant and crop production are neither well organized nor properly utilized. At present, there are research efforts to develop ontology by automatic and semi-automatic approaches, in order to decrease cost of development [4]. We are working on ontology learning techniques to develop heavy weight application level ontologies with minimal human intervention. Agrovoc [5] is used as a domain ontology for food and agriculture domain. In our work, we have developed a cotton ontology by extending the available ontology [6], which contains the knowledge of cotton cultivation practices at every stage in detail. Ontologies are used to process the complex requests of a farmer and generate recommendations.

2 Related Work

During our literature survey, we found interesting work done for developing innovative applications based on GIS and knowledge base technologies. A system is implemented for Maharashtra state in India for managing information in a sugarcane cooperative with an SMS-based mobile application [7], which sends the sugarcane crop related recommendations on mobile phone. Significance and use of ontology for dynamic data gathering of Geographic Information Service, knowledge representation, semantic resolution of geo-spatial query is discussed for development of intelligent decision support system in [8–11]. A farmer centric system which integrates relevant web services like soil, plant weather, pesticide and fungicides information is proposed in [12]. Although the system proposed in [12] is UDDI based SOA for plant yet unable to generate expert recommendations. aAQUA [13] is an innovative tool for answering farmer's questions. Spatial data analysis is helpful generating recommendations for agriculture crop production [14]. eSagu [15] is personalized agro-advisory System to resolve crop related queries of the farmers. Our work is motivated by the existing systems described above. There is a need of semantic web based system which is capable enough to respond complex crop related queries based on spatial data integration and reasoning. In this work, we have developed cotton crop ontology and implemented web processing services (WPS), generating recommendations for various stages of cotton crop cultivation.

3 System Architecture

Spatial data captured through diverge sources enable opportunities to integrate and support complex applications such as agro advisory and flood forecast system. Geo-spatial data and services can become extremely useful if distributed over large networks. A common and inter-operable solution for discovery, storage and access of geo-spatial data and information has become the necessity. We aim to evaluate and test our proposed architecture for agricultural recommendation system specifically in the context of climate change and its impact on agricultural practices. We have used an approach which integrates GIS and ontology, to work together by making use of services and generate fertile recommendations for the farmers. The primary objective of the system is to generate recommendations for different types of queries using two different technologies, GIS and Semantic Web.

The system receives input in the form of geographical data regarding climate parameters such as temperature, humidity, rainfall, administrative boundary data etc. Such spatial data is passed to the GIS based spatial analysis module for further processing. The geo-spatial algorithms analyze the data and generate response. These results are passed to the reasoning engine, which uses cotton ontology along with the results obtained from the geo-spatial analysis. Based on integration of responses from both the modules, relevant recommendations like disease prevention, prescription, use of fertilizer and pest control are generated and communicated to farmers through mobile phones using RESTful web services.

3.1 GIS Based Query and Information Representation

In our application, spatial data is stored into PostgreSQL database and is retrieved through RESTful web services for each query. Diversity of geospatial data is one of the main concerns during integration. When a farmer enters inputs for the query, services will be invoked through RESTful URI and processing will be done on geospatial data stored in PostgreSQL.

3.2 Knowledge Processing

Knowledge represented in the form of ontologies can be extracted by executing SPARQL queries. Ontology is a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for particular topic. Knowledge models in agropedia [16, 17] can be extended for application level ontology development. We are developing cotton crop ontology using Protégé [18, 19] by referring standard agrovoc ontology developed by Food and Agriculture Organizations of United Nations (FAO). Ontology contains intensive knowledge

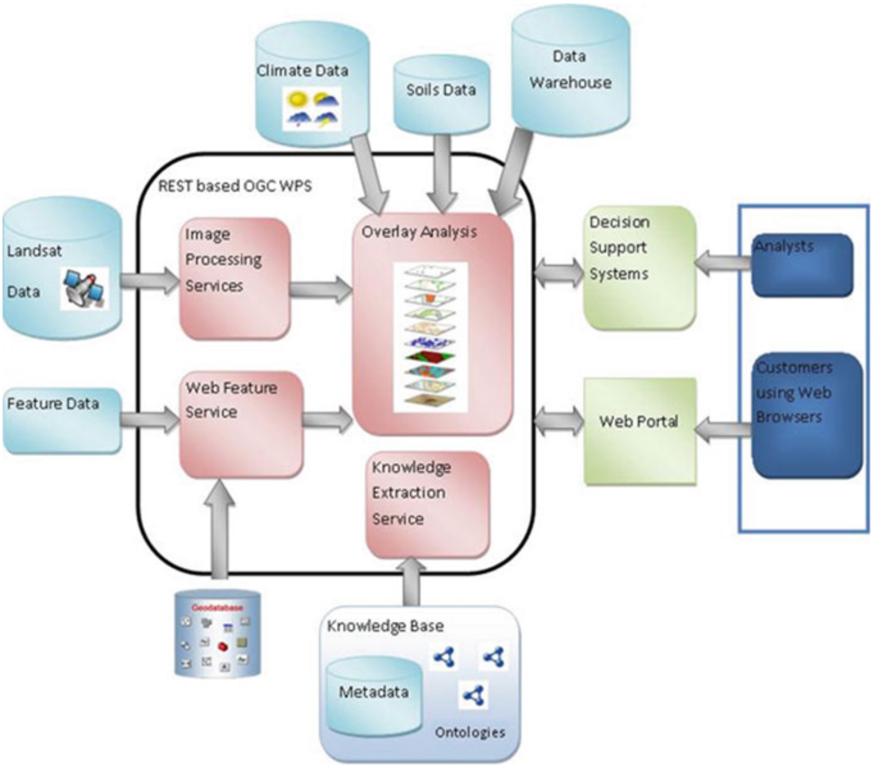


Fig. 2 RESTful web processing service architecture [18]

of cotton cultivation practices at all stages, starting from land preparation to post harvesting. For example, when a farmer inputs the observation, required knowledge is extracted by executing corresponding SPARQL query.

3.3 An Approach for RESTful Web Processing Service

Our system named 'Krishimantra' is implemented by developing RESTful web services for knowledge extraction from ontology and spatial data base. The detailed architecture of the REST based services is shown in Fig. 2. Our developed system can be interfaced with existing decision support systems using JSON[1] and GeoJSON messages. System allows user to post requests i.e. recommended fertilizer, pesticide, weed control applications for their crop. The system identifies location and based

[1] www.json.org.

on agro meteorological parameters, it generates relevant recommendations or alerts regarding requests from the farmers. Key characteristics of the developed system are:

- Sophisticated processing of spatial data using WPS which provides graphical context to the data as well as analyzing the data sets to produce location specific recommendations.
- Use of Open Geospatial Consortium (OGC) based services to enable data discovery based on relevant domain parameters.
- Use of RESTful Web and GeoWeb services which provide simple and cleaner interfaces, thereby reducing the complexity of the system.

4 Implementation

This section will emphasize mainly on the implementation and processing of different modules discussed above.

4.1 RESTful Web Services for GIS

RESTful web services are generated using JAVA API for RESTful Web Services (JAX-RS) [18]. Each service described below responds to a specific query.

- **Registration**: This RESTful web service is responsible to fetch the user's personal information.
- **Soil Helth Card**: This RESTful web service is developed to fetch the soil health card information related with the farm's unique ID.
- **Crop Recommendation**: This RESTful web service generate the recommendation for crops variety to be cultivated based on farmer's locations and current Agro-Ecological Zone (AEZ).
- **Prescription Recommendation**: A RESTful web service is created for generating recommendations regarding taking necessary action on having some disease, pest or weed observation.
- **Local Facilities Finder**: This RESTful web service is developed to find the local facilities for farmer based on their current location.

4.2 RESTful Web Services for Knowledge Extraction

We have developed RESTful Web Services to be invoked for specific complex query in order to generate recommendation, e.g. "Finding recommendations for pest prevention based on observations". User is expected to input the observations about cotton crop and recommendations are generated based on the available knowledge in crop ontology. When a user is providing observation, service is invoked and it finds the pest symptoms based on the observation captured. In subsequent steps, pest and its related prevention is searched and respective description is returned in the form of recommendation.

Query: If there is a bore in the bolls of cotton crop what should be Pest Prevention? The service will fetch the text "Bore in to the bolls" from users selected observation from the observations drop down list. Using this text the request of farmer is converted in to SPARQL query as shown below:

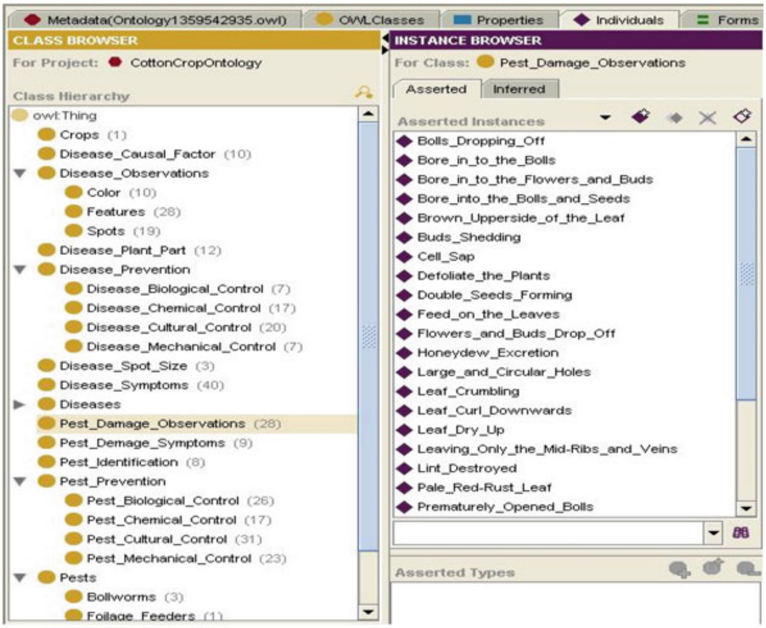
```
SELECT ?Description
{ <http://www.owl-ontologies.com/Ontology1359542935.owl
#Bore_in_to_the_Bolls >
cotton:is_Observation_Of ?Pest_Damage_Symptoms.
?Pests cotton:Cause_Damage ?Pest_Damage_Symptoms.
?Pests cotton:is_Prevented_By ?Pest_Prevention.
?Pest_Prevention cotton:has_Description ?Description.}
```

4.3 Knowledge Base Development for Cotton Crop

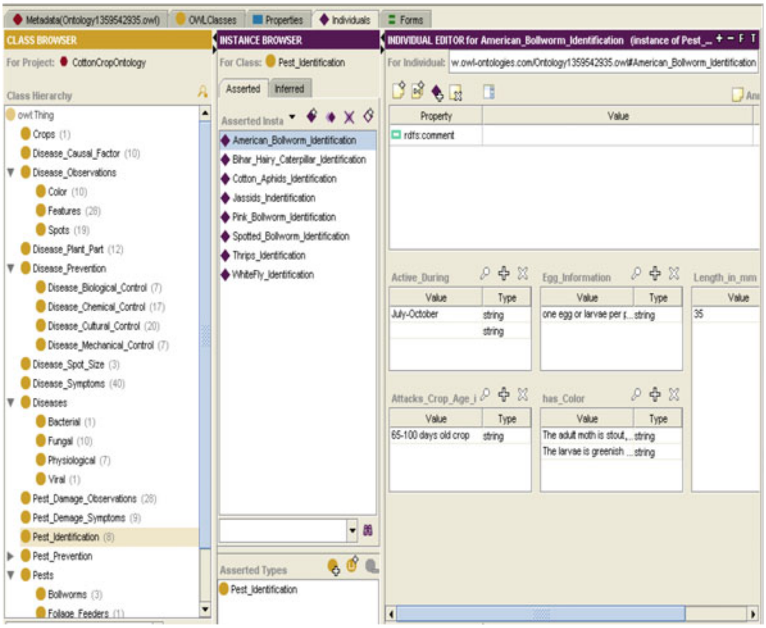
Knowledge of a particular domain can be represented through ontologies. In our experimental setup, the objective was to perform reasoning and inferencing regarding cotton crop. In order to capture knowledge related to various crop cultivation stages, we needed cotton crop knowledge. Concepts of disease, pest, weed, climate and soil are populated for the cotton crop using Protégé ontology editor tool (Fig. 3). We have used agrovoc as the base ontology in order to expand the complexity of SPARQL queries. Developed RESTful services interact with the user through mobile application and generate recommendations whenever any specific service is invoked.

4.4 Android Application Development

We have developed an android application for information dissemination which is capable enough to fetch the user's profile and providing the interface. We have named this application as "Krishimantra". This is a user friendly interface using which a user can supply input to the system without even much literacy. The RESTful services work on the back end of interface and parse the information obtained in JSON and GeoJSON format. Google Map's JavaScript API v3 is used which is facilitating,

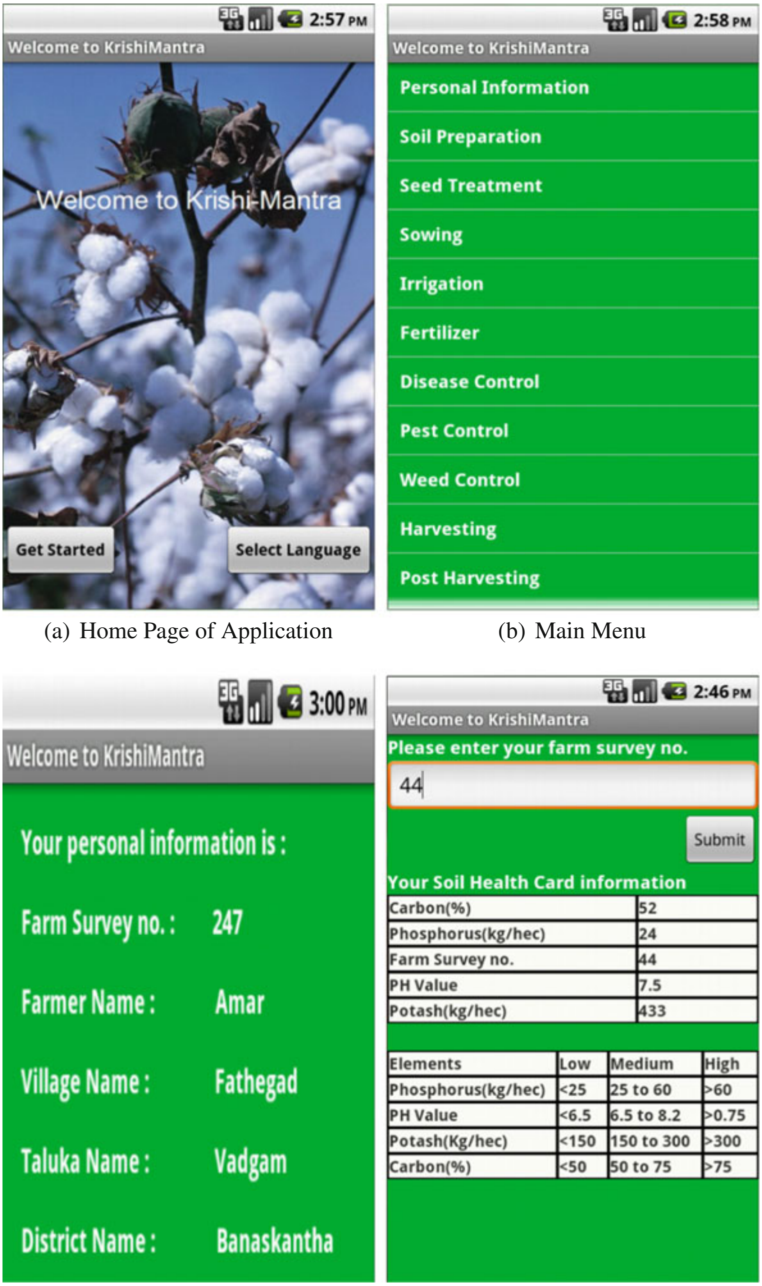


(a) Pest Observation



(b) Pest Identification

Fig. 3 Populated cotton ontology [18, 20]



(a) Home Page of Application (b) Main Menu

(c) Farmer's Personal Information (d) Soil Health Card Information

Fig. 4 Set of figures depicting the application interface driving directions to the nearest entity based on the farmer’s current location are processed and displayed on a separate widget. Main menu of an application is shown in Fig. 4b. Main menu contains the concepts related to current cultivation practices. The user can navigate in any sub menu and can query through this main menu.

5 Results

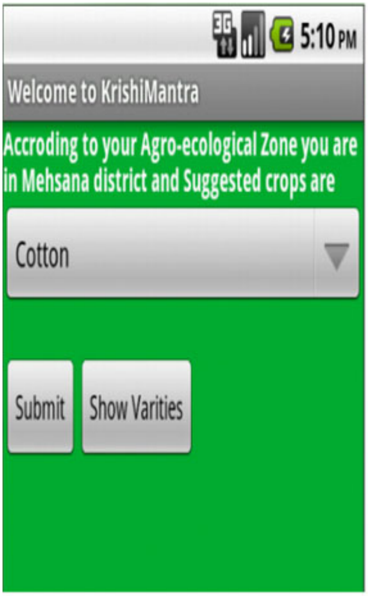
This section discusses a list of queries we have populated and their output on android based mobile device [18]. The home page of an application is shown in Fig. 4a. A farmer can select the preferable language by clicking “Select Language” button.

QUERY 1: Extraction of farmer’s personal information.
When a farmer starts an application, device’s IMEI (International Mobile Equipment Identity) number is extracted and sent to GIS server for information extraction. As a result, his/her personal information is extracted and displayed on android device. As show in Fig. 4c, details such as farmer’s name, his/her farm survey number, village name, taluka name and district name are displayed.

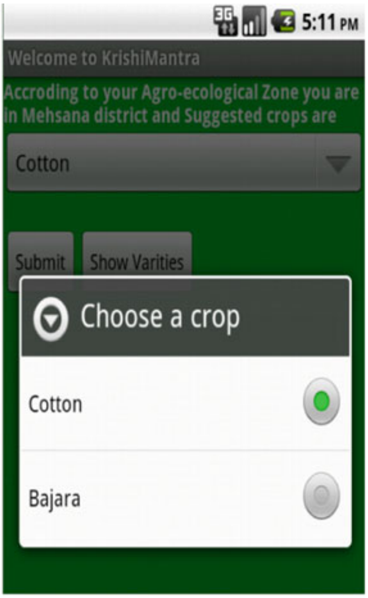
QUERY 2: Extraction of Soil Health Card (SHC) information of a farmer’s farm.
In this query, farmers enter their farm survey number as an input. Based on this farm survey number, their farm’s soil properties are extracted in the form of soil health card and displayed on screen. Soil health card also contains the range values for each element in order to provide basic understanding of soil quality. As shown in Fig. 4d, farm survey number (i.e. 44) is entered and elements of specific farm are displayed with its range values.

QUERY 3: Recommendation related to dynamic generation of crop and crop varieties based on Agro-ecological Zone (AEZ).
In this query, recommendations regarding varieties of crop to be grown are generated using user’s location and its mapping to the available agro-ecological data in database. For processing this query, Gujarat’s map is digitized which contains agro-ecological zones (AEZ). Process of geo-referencing is done followed by the overlaying of AEZs and adding attribute data. When a farmer selects the query, RESTful service is invoked which is responsible for fetching the current location of the farmer and identification of the AEZ. Based on AEZ, the farmer receives a dynamic recommendation related to the crop and their varieties that should be grown. As shown in Fig. 5a, Mehsana district is identified based on farmer’s current location. The system dynamically displays the suggested crop based on AEZ of Mehsana district (i.e. Cotton and Bajra) as shown in Fig. 5b. When a farmer selects the particular crop, suggested varieties are displayed dynamically on the mobile device as shown in Fig. 5c.

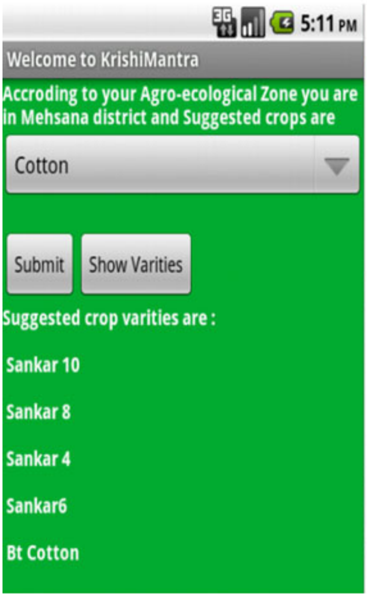
QUERY 4: Recommendation for pest prevention based on observed symptoms.
Knowledge base contains the static knowledge related to current cultivation practices. This query returns the recommendations steps for pest prevention based on



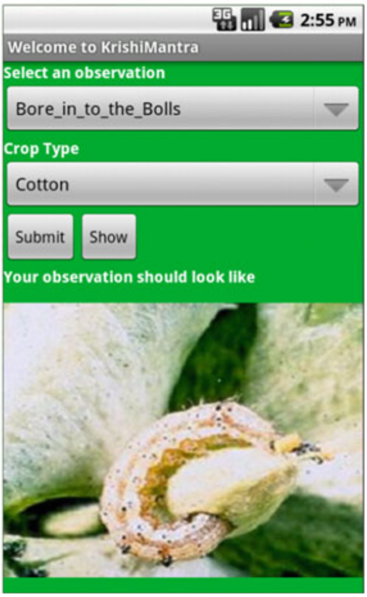
(a) Zone Identified Based on AEZ



(b) Crops Suggested Based on Identified AEZ



(c) Crop Varieties Suggested Based on Crop Selection



(d) Pest Observation Selection to get Preventive Actions

Fig. 5 Set of figures depicting the results of query using application interface

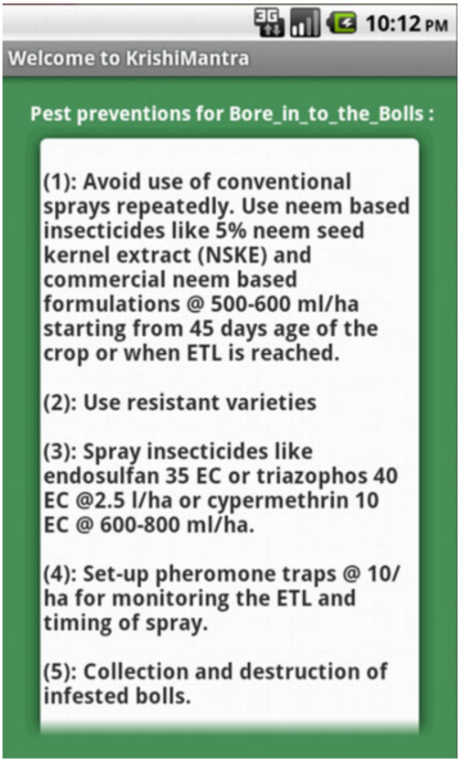


Fig. 6 Recommendation for pest prevention

the observations. Recommendation describes use of biological, cultural, chemical and mechanical control for pest prevention practices. As shown in Fig. 5d, farmer selects an observation and crop type from provided drop down list. At the bottom of the screen, picture is displayed dynamically based on the selected observation. This feature ensures the correctness of the observation to be entered. As a result, steps related to pest prevention are displayed on mobile device as shown in Fig. 6.

6 Conclusion and Future Work

In this chapter, we have shown the results of complex processing of queries, using GIS information and cotton ontology. Results for queries are tested on Android based mobile phones. The cotton ontology is developed and proof of concept is tested using the current state of ontology; crop ontology will require updates by domain experts in order to make an intensive knowledge base. We have performed experiments for North Gujarat region. Future work involves developing user friendly interface in regional language (Gujarati) and generate recommendations. Context aware alerts and notifications will be sent to the user. RESTful web services will be hosted on the cloud environment. There is a need to generalize our proposed architecture and further seamless interoperability among spatial data sources needs to be tested and implemented.

Acknowledgements This work was a part of research project on "*Service-Oriented Architecture for Spatial Data Integration and Spatial Reasoning*", funded by The NRDMS and NSDI, Department of Science and Technology, Govt. of India.

References

1. Retna Raj, T.J., Sasipraba, R.: Disaster management system based on gis web services. In: Recent Advances in Space Technology Services and Climate Change (RSTSCC) (2010)
2. https://www.w3.org/2003/Talks/0521-hh-wsa/slide5-0.html
3. Lu, X.: An investigation on service-oriented architecture for constructing distributed web gis application. In: 2005 IEEE International Conference on Services Computing, vol. I (2005). https://doi.org/10.1109/SCC.2005.27
4. Thunkijjanukij, A.: Ontology development for agricultural research knowledge management: a case study for thai rice. Ph.D. thesis, Graduate School, Kasetsart University (2009)
5. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S.: Reengineering thesauri for new applications: the agrovoc example. J. Digit. Inf. **4**(4) (2006)
6. Sarda, N.L.: Building crop ontology for farmers. Dissertation, Banasthali University (2012)
7. Veeraraghavan, R., Yasodhar, N., Toyama, K.: Warana unwired: replacing PCs with mobile phones in a rural sugarcane cooperative. In: 2007 International Conference on Information and Communication Technologies and Development, vol. 5, no. 1, pp. 1–10 (2007)
8. Bhattacharjee, S., Prasad, R.R., Dwivedi, A., Dasgupta, A., Ghosh, S.K.: Ontology based framework for semantic resolution of geospatial query. In: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 437–442. IEEE (2012)
9. Chen, R.: Ontology-based dynamic data gathering of geographic information services. In: 2010 International Conference on Future Information Technology and Management Engineering (FITME), vol. 1, pp. 181–184 (2010)
10. Mao, X., Li, Q.: Ontology-based web spatial decision support system. In: 2011 19th International Conference on Geoinformatics, vol. 08193, pp. 1–4 (2011)
11. Yuanyuan, W., Rujing, W., Xue, W., Yimin, H.: Ontology-based knowledge representation for agricultural intelligent information systems. In: 2010 International Conference on Management and Service Science MASS, pp. 9–12 (2010)
12. Shyamaladevi, K., Mirnalinee, T.T., Trueman, T.E., Kaladevi, R.: Design of ontology based ubiquitous web for agriculturea farmer helping system. In: 2012 International Conference on Computing, Communication and Applications (ICCCA), pp. 1–6. IEEE (2012)
13. Ramamritham, K., Sahni, S., Baru, M., Bahuman, C., Chandran, A., Joshi, M., Bahuman, A.: The aaqua approach: innovative web 2.0 tools for developing countries. Int. Comput. IEEE **12**(2), 62–70 (2008)
14. Kumar, V., Jain, Y., Sharma, A., Chaudhary, S.: Spatial analysis for generating recommendations for agricultural crop production. In: India Conference on Geospatial Technologies and Applications (ICGTA-12) (2012)
15. Krishna Reddy, P., Syamasundar Reddy, G., Reddy, B.: Overview of eSagu and future plan. In: Presented at IEEE Conference on Technologies for Humanitarian Challenges (2009)
16. Pappu, N., Sarkar, R., Prabhakar, T.V.: Agropedia: humanization of agricultural knowledge. IEEE Int. Computi. **14**(5), 57–59 (2010)
17. Sini, M., Yadav, V., Singh, J., Awasthi, V.: Knowledge models in agropedia. In: Indica (2009)

18. Kumar, V., Dave, V., Nagrani, R., Chaudhary, S., Bhise, M.: Crop cultivation information system on mobile devices. In: Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), pp. 196–202. IEEE (2013)
19. Noy, N.F., et al.: Creating semantic web contents with protege-2000. IEEE Intell. Syst. **16**(2), 60–71 (2001)
20. Kumar, V., Dave, V., Bhadauriya, R., Chaudhary, S.: Krishimantra: agricultural recommendation system. In: ACM DEV, p. 45 (2013)

OntoAQ: Ontology-Based Flexible Querying System for Farmers

S. Sahni, N. Arora, S. Sen and N. L. Sarda

1 Introduction

In many domains, information can be modeled and stored by employing unary and binary relations alone. Indeed, graph databases, which can be seen as databases where relations are limited to such arity, have gained much attention. Domain knowledge modeled as ontologies can be used to provide a knowledge-based approach to tasks such as query answering and data sharing. Application of ontologies to the agricultural domain has been previously proposed and implemented [1], but owing to the specialized nature of knowledge for different farming practices, the need of a knowledge-based approach, in general, and the use of ontologies, in particular, can benefit the farming community.

1.1 Background

While formal querying involves syntax and rules that untrained users may not be familiar with, natural language queries suffer from ambiguities and logical complications that render such interfaces problematic. Natural language interfaces supported by backend ontologies offer end users a familiar and convenient option for querying

S. Sahni
Oracle India Pvt. Ltd, Mumbai, India

S. Sahni · N. L. Sarda
Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai, India

N. Arora · S. Sen (✉)
GISE Lab, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai, India
e-mail: sumitssen@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_16

ontology-based knowledge bases. They hide the formal requirements of ontologies and query languages from end users by offering them a familiar and intuitive way of formulating queries.

Many approaches have been suggested for the natural language interfaces to ontologies, which convert keyword queries to formal query languages. Some of them are based on either syntactic input [2] or keyword or unrestricted input [3, 4]. Some of the other approaches work on constructing and operating subgraphs from RDF graphs [5, 6].

Querix [2] is an ontology-based question answering system that translates generic natural language queries into SPARQL. In case of ambiguities, Querix relies on clarification dialogues with users. In this process, users need to disambiguate the sense from the system-provided suggestions.

NLP-Reduce [4] is a naive and completely domain-independent natural language interface for querying Semantic Web knowledge bases. The major advantage of the system is that it is robust to deficient, ungrammatical, and fragmentary input. SemSearch [3] is a concept-based system with a Google-like query interface. It deals with keyword queries in semantic search engines. Here, keywords are mapped to elements of triple patterns of predefined query templates. These templates fix the structure of the resulting queries beforehand. However, only such templates can capture some but not all interpretations of the keywords.

In [5], a more generic graph-based approach has been proposed to explore all possible connections between nodes that correspond to keywords in the query. This way, all interpretations that can be derived from the RDF graph can be computed. The approach used in Q2Sematic [6] search support keywords that match literals and concepts contained in the RDF. For example, Kolia et al. [7] have reported algorithms and a prototype implementation of SPARQL query answering over OWL ontologies that have been widely used.

1.2 Motivating Example

Farming generally relies on techniques to promote cultivation and maintain the lands suitable for their species. Farmers have many questions regarding the type of soil/climate for a particular crop, type of pests within crop, different diseases, timelines associated with each activity related to crop, etc. The farmers do farming more by convention than by scientific knowledge.

Farmers express their queries in a natural language, which are usually answered by human experts. It is desired to enable the data or knowledge captured in the system to understand the query as exactly as farmers see and ask questions. Hence, a need arises for a context-based knowledge-driven advisory solution that has a very simple keyword interface and that can give both information and recommendations for farmers. The initial idea came from the work on the proposal of agro-advisory system [8], which is a query-answering support for farmers. It is nothing but an ontology-based knowledge system. Knowledge acquisition is done with the aid of agro-experts.

2 Knowledge Requirements for Cotton Farming

Cotton farming is a rewarding occupation, but several factors influence productivity, and experts advise farmers on both general advisories pertinent to certain places and times and specific advice on problems they may face such as plant diseases and the use of pesticides.

2.1 Agro-Advisory Systems

Agro-advisory systems have been developed in the past, which help in advising the farmer for their crop-related queries. Initial work was carried out at IIT Bombay for developing a discussion forum for farmers, namely aAQUA [8], which is an online, multilingual, multimedia question and answer-based community forum. In a typical aAQUA thread, a farmer submits a problem, and agriculture experts or other farmers provide solutions to it. eSagu [9] is another advisory system where agricultural expert delivers the expert advice at regular intervals (once in one or two weeks) to each farm by getting the crop status in the form of digital photographs and other information. It is a query-less system and provides agro-advice even without the farmer asking a question by following a proactive approach and avoiding problematic situations. mKRISHI [10] is a mobile agro-advisory system which provides the farmer with audio–video facilities on a mobile phone to express their queries to experts with minimal use of text. It adapts event-based experiential approaches to analyze data from different sources and develop techniques to store and visualize all this multimedia data, which holds relevant to the context. The drawbacks behind each of these approaches are the presence of an agro-expert for looking into the farmer's problems. The approach presented in this paper mainly concerns with developing an independent interface where farmer can query and obtain results for the query specific to their context. Crop information is stored in a knowledge model known as ontology, and queries are answered by searching over these ontology graphs.

2.2 Agro Ontologies

2.2.1 Agropedia Indica

Agropedia [8] is an online knowledge repository for information related to agriculture in India. Knowledge models (KMs) are mainly used to navigate agricultural knowledge and to organize and search agricultural content. Such KMs are the structural representation of knowledge by using symbols to represent concepts and relationships between them, which can be used to connect seamlessly to the knowledge base in *agropedia* using semantic tools. Agropedia contains top-level generic map for crops (which is called the "Foundational Agricultural Crop Ontology"), and many specialized maps based on specific topics, such as rice, diseases, and pesticides. The repository contains KMs for nine selected crops—chickpea, groundnut, litchi, pigeon pea, rice, sorghum, sugarcane, vegetable pea, and wheat [8].

2.2.2 Ontology-Driven Context-Based Flexible Querying System

OntoAQ has been designed to handle farming-related queries without the need of agro-expert handling them at the other end. The end users (farmers) who can post their observations about the crop to the system are advised about specific action to be taken against the problem or provide information on the query.

Comprehensive information about the crop is stored as the knowledge given by the agro-expert needs to be acquired by the system for making it independent of answering queries without the aid of an agro-expert. Along with this, the system should also be capable of capturing context information. This can be information about farmer's context, context information on weather, diseases history, and history about the insecticides.

Knowledge of agro-experts is acquired as knowledge models known as ontologies, which are capable of storing the complete information related to the crop. Knowledge seen and exhibited by an agro-expert can be equally depicted by the ontology. It depends upon how well it is captured. This encourages the existence of a system independent of an agro-expert competent of satisfying the end users. Apart from the information related to the crop stored as a knowledge model, contextual information is also stored in the database. One is the information of the farmer's context, which details about the various activities performed by the farmer time to time. This can be captured only when the farmer registers about his activity done on the crop to the system. Another context information captured is the weather prediction, which forecasts the climatic conditions in the region and helps in suggesting the farmer better.

Hence arises the need on designing an ontology-based querying system which maps the information asked by the user on the knowledge stored in the ontology, and with the assistance of the contextual information stored in the database about the crop, weather, geographic details, etc., advice or suggestion is made to the user. So we designed a system named OntoAQ, which handles both context- and keyword-based queries. We discuss the architecture of the system as below.

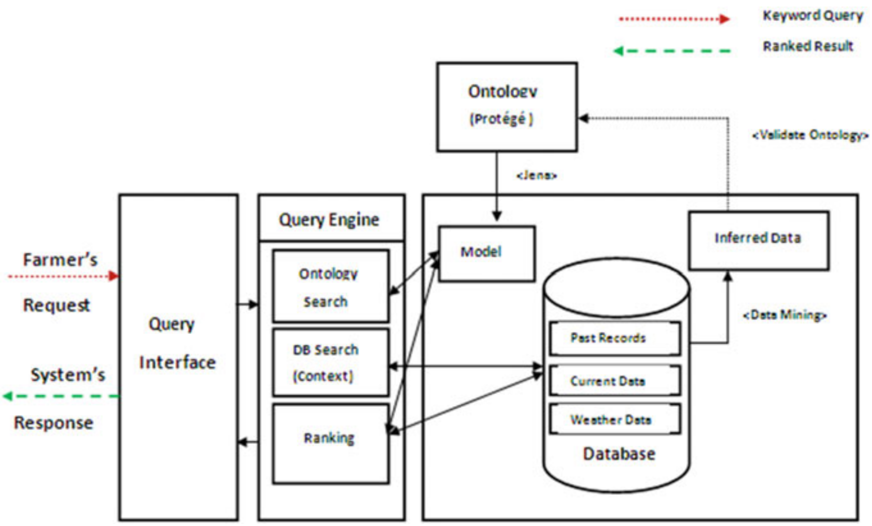


Fig. 1 OntoAQ architecture

2.3 System Architecture

System consists of the following components explained below (see Fig. 1):

2.3.1 Query Interface

This is an interface wherein the farmer logs in and posts her/his farming-related query. Query can be an observation made by him, information which he wants to seek, or an action/suggestion-oriented query. The query made by the farmer can be an attribute–value pairs, or it may be a keyword-based query. This query input is parsed, tokenized, and mapped to concepts within the ontology.

2.3.2 Query Engine

This is the main component of the system, which handles farmer query. A given query of the farmer can be of the form attribute–value pairs, or it can be a keyword-based query. Query is mapped to the ontology resources, which can be any of classes, instances, object/datatype property, or literals. Based on the selected resources, information is extracted and also marked on the ontology. Secondly, farmer's current activities are queried from the database, and suggestions are made for what action should be taken by him depending upon the weather forecast and past activities performed, if any. Finally, the engine ranks retrieved results it according to priority.

2.3.3 Database

The database consists of the repository, which holds complete information about a given crop in the form of ontology. This ontology is stored in the form of a graph within the database. Also, past records of the farmers are present as relational data, which give us the farming practices done by them. Current farming operations are also stored in separate tables. Weather forecast for the next 5 days is gathered from metrological department and captured in the database. Based on these generic rules, observations are made which help in validating the ontology. Advice is generated by the system based on the current farming, which helps him to take precautionary measures in his practice. The query made by the farmer is saved as a log in the database, which stores the query keywords, crop sown, observations seen, actions or suggestions made.

2.3.4 Cotton-Specific Ontologies

Knowledge about the cotton crop is stored as ontologies. Concepts which are critical in the context of cotton farming are soil and climatic conditions for cotton, recommended varieties of cotton which can be location-specific, disease affecting the crop, reasons for their occurrence, their symptoms shown, and cure for those diseases; similarly information on pests attacking the crop, their precautions and cure. In addition, the cotton ontology requires concepts related to the various activities of farming, like, hoeing, sowing, irrigation, fertilizing, spraying, and harvesting, along with their timelines.

We have defined a specific ontology for cotton farming based on a generic crop ontology as shown in Figs. 2a, b. In addition, disease ontologies about the crop were developed that specified both the ***stage*** and the ***control measures*** related to the disease. Actionable knowledge about the disease as well as general information that is helpful to farmers in regard to the particular disease is also stored in the disease ontology.

Ontologies are stored as RDF triples in database. Jena provides with APIs for modeling it to a graph with nodes and edges. Ontology is hence expressed as a graph with subject, object as nodes, and predicates as edges. Traversing this graph and establishing relations between nodes would result in a path of meaningful information for the crop ontology.

2.3.5 Contextual Data of the Farmer

Data from various villages of Punjab have been collected and is then used for analysis. The data has been collected from three districts of Punjab. It contains information regarding farmers and their usual farming practices in five different years. Analysis of this data is performed in various aspects to validate details and find patterns in farming practices of farmers. Such analysis serves as the key to answering queries

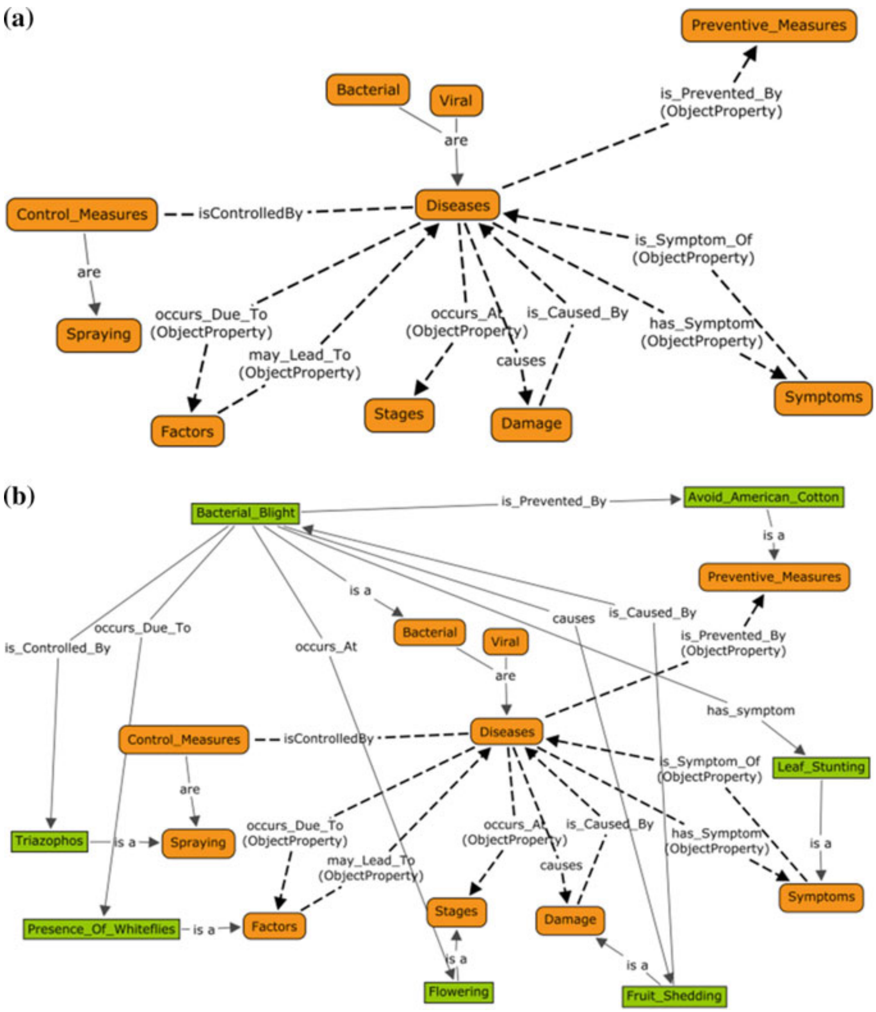


Fig. 2 **a** Crop ontology (generic) and **b** Cotton ontology (specific)

raised by farmers for their specific crops. Observations made for related past records can be verified and queries be answered.

Current practices of the farmer are also noted regularly and are monitored periodically. Activities of the farmers at various stages are captured, like when the crop was sown, its variety and location stored in a database as shown in Fig. 3. Based on the past practices, farmer may be advised best ways to manage the current crop cycle.

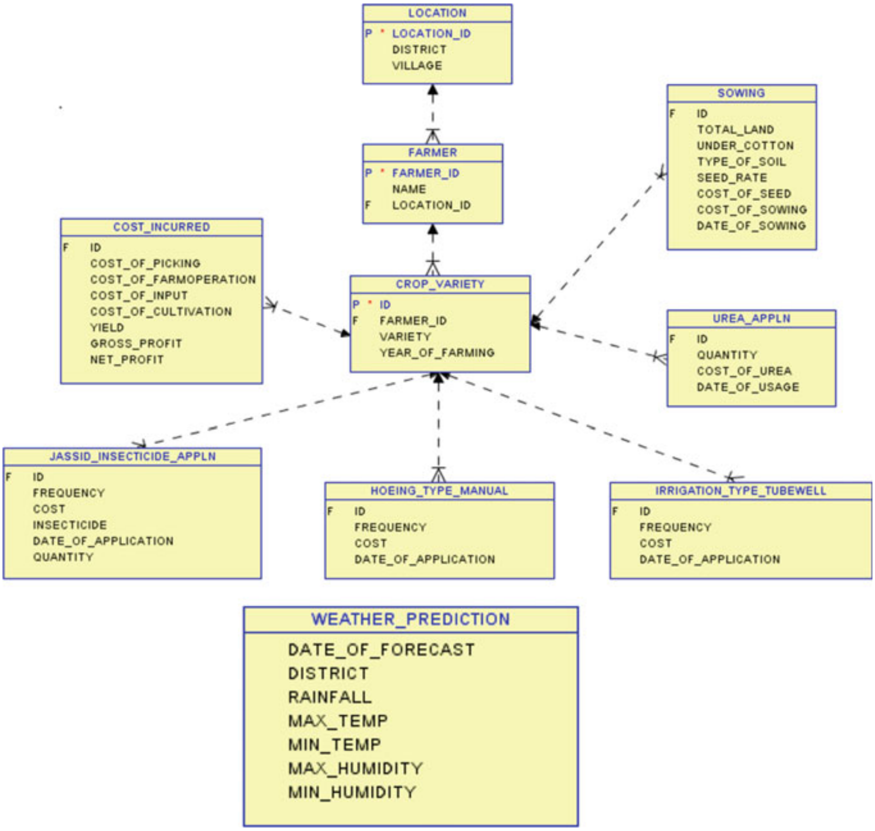


Fig. 3 Database schema for farmer activity and weather prediction

2.3.6 Weather Predictions

Weather forecast for the next 5 days is gathered from metrological department and captured in the database. Mainly data about rainfall, humidity, and temperature are captured.

3 Querying OntoAQ

Query engine handles two types of queries and forms the central component of the system. It is designed to enable efficient querying and thus provide an answer graph based on the type of the query asked.

Keyword-based Query: Ontologies are stored as a graph structure, and to answer a query, keywords of the user are mapped over this graph structure and returning a subgraph, which satisfies these selected keywords. The algorithm for the keyword-based query thus obtains the best subgraph for a given set of keywords. This approach is similar to the approach adopted by many researchers [11, 12].

Context-based Query: While answering to a farmer based on the observations made by him in the query interface, analysis is done on his current activities, i.e., when did he sow his latest crop, his crop is at what stage, what all activities he has performed, and what next is to be done by him. This is supported by the observations from the weather predictions on rainfall, temperature, and humidity for the next 5 days captured which also help in suggesting the farmer for what needs to be done and care to be taken. Suggestions are made to him based on the activity information stored in ontology and weather predictions made.

3.1 Search Algorithm

Farmer posts his query over the query interface, which is handled as a graph-based search. Nature of the query could be set of observations made by him related to the crop. Simplest form of presenting his observations could be set of <attribute-value> pairs, where an attribute could be the context and value is the observation made by him. Farmer may or may not be aware of the predefined attribute names. Hence, we offer him to perform keyword query where the system captures the observations made by the farmer and best maps it to the attributes it corresponds to. Some example queries posted by farmer could be:

Leaf turning yellow to red,

Varieties of cotton.

First query is a kind of observation-based query where farmer posts his field observations. Also it is an action query where a farmer would seek an advice from the system. The second query is an information-based query where the farmer wishes to know some general information about the crop.

Ontology is stored as a graph in the database. Generic ontology O_G defines a graph, which consists of classes, their object properties, and datatype properties. Classes form the nodes of a graph, and edges are comprised of object property and/or datatype property. Specific ontology O_S consists of generic ontology with instances of classes and relations (object and datatype properties) between them. Here, nodes of the graph are classes, instances, and literal values, whereas the edges are object property and/or datatype property.

The various steps (see Fig. 4) involved in answering the queries are described below. Before performing query search, preprocessing is done for resources in O_S by categorizing them into sets of *classes*, *instances*, *objectProp*, *dataProp,* and *literals*.

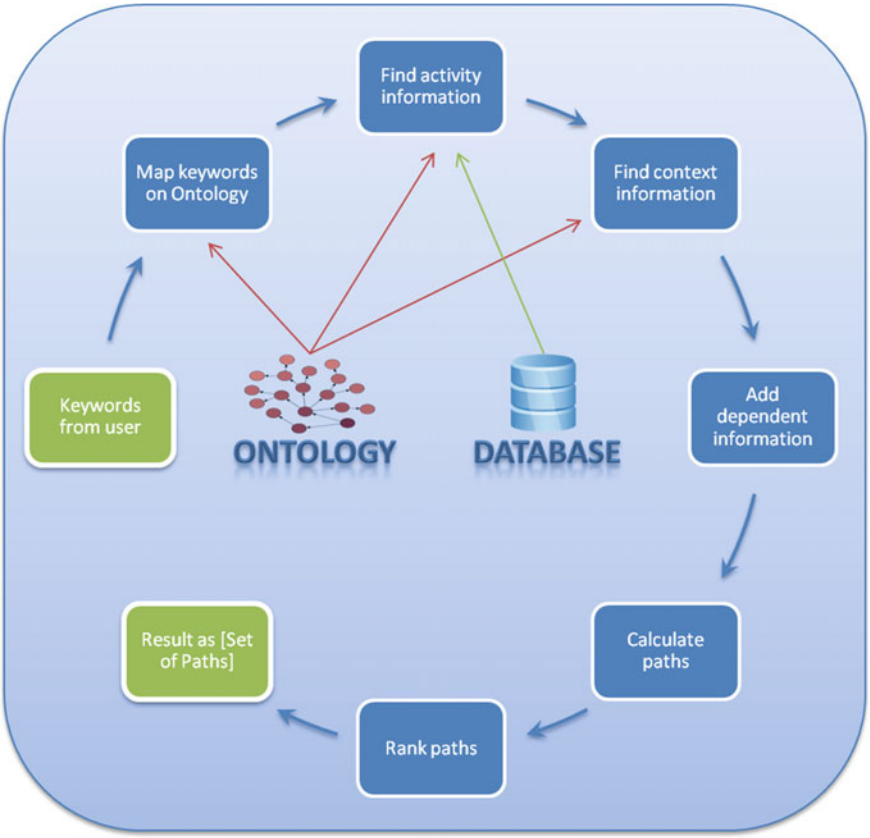


Fig. 4 Overview of search algorithm: beginning with (i) identification of keywords, (ii) classification of the keywords as nodes and edges, (iii) inferred contextual information, (iv) inferred farmer activity from integrated database, (v) calculate paths from ontology graphs as potential answers, (vi) ranking of paths based on *similarity count* and *actionable information.* For further information and complete details of the system, please refer [13]

3.2 Performance

Query performed over the ontology can be purely keyword-based which is independent of farmer who is posting the query. Or query can be context-based which depends on various factors like farmer posting the query, time/stage of crop when query is posted. Below are some of the sample results.

Case1: Ontology-Based Search
Example query#1: Varieties of Cotton
Results: *Cotton has_Variety RCH_134*

Cotton has_Variety MRC_70

Cotton has_Variety LD_694

This is an informative query where different varieties of cotton are returned to the user. This does not depend upon the context of the user.

Case 2: Ontology query—Context-Based
Example query#2: Boll Shedding
Results:

The results for this query depend upon the user querying. An ontology-based search would return all possibilities of the query, whereas context-based would check for the state of farmer crop and return results specific to him. Different possible results for the query are:

Boll_Shedding is_Caused_By American_Bollworm is_Controlled_By Acephate
Boll_Shedding is_Caused_By Tobacco_Caterpillar is_Controlled_By Thiodicarb
Boll_Shedding is_Caused_By Whitefly is_Controlled_By Triazophos
Boll_Shedding is_Caused_By Lack_of_Irrigation may_Lead_To Tirak is_Controlled_By

Proper_Fertilization

These results depend upon the stage of a farmer's crop and current season and hence vary. For example, considering two different cases with different context:

1. If season is June and there is no irrigation and nitrogen for crop. The results for this context are:

Boll_Shedding is_Caused_By Lack_of_Irrigation may_Lead_To Tirak is_Controlled_By

Proper_Fertilization

Boll_Shedding is_Caused_By Nitrogen_Deficiency is_Controlled_By Application_of_Urea
Boll_Shedding is_Caused_By Lack_of_Irrigation may_Lead_To Tirak has_Symptom

Yellow_to_Red_Leaf_color

Black_Boll_Color is_Symptom_Of Tirak occurs_Due_To Lack_Of_Irrigation causes Boll_Shedding
Bad_Boll_Shedding is_Symptom_Of Tirak occurs_Due_To Lack_Of_Irrigation causes
Boll_Shedding
Boll_Shedding is_Caused_By Spotted_Bollworm

2. If season is July and there is no nitrogen, potassium, and phosphorus. The results for this context are:

Boll_Shedding is_Caused_By Nitrogen_Deficiency is_Controlled_By

Application_of_Urea

Table 1 Results from queries. The columns represent the queries executed, the precision of the queries, and the recall of the results

Query	Precision [0–1]	Recall (0–1)
Leaf turning yellow	0.5	1.0
Boll shedding	0.8	0.8
Varieties of cotton	1.0	1.0

Tobacco_Caterpillar causes Boll_Shedding
Boll_Shedding is_Caused_By Spotted_Bollworm
Whitefly causes Boll_Shedding
Mealy_Bug causes Boll_Shedding

The results obtained are shown in Table 1, and the columns represent the queries executed, the precision of the queries, and the recall of the results.

3.3 Ontology Validation

Ontology validation involves the practical experiences of the farmer being validated against the information given in the ontology. Ontologies are often validated with human interfacing [14], and this approach ensures that the knowledge used in OntoAQ remains accurate. Information stored about the farmer in database mainly holds the activities performed by him, time of their occurrence, quantity or amount applied, if any, and the cost incurred in performing those activities. Parallel to this, we have similar information stored in the ontology, which conveys all of these parameters. Knowledge stored in the ontology is from the agro-experts, which may or may not be the best for the crop. Hence, we validate this with the actual experiences of the farmer.

The practices followed by the farmers whose yields have been better in their areas are analyzed and compared with the activity information given in the ontology. Differences between the two are pointed out and seen whether maximum number of farmers follow a practice which is in accordance with the one suggested by the ontology or maximum perform a practice which violates the ontology information. In the former case, it is proved that ontology has accurate information and farmers who do not follow that can be alerted. The latter case shows that ontology contains some information, which is not followed by most farmers, though getting best yields. Hence in that case, the ontology is to be updated.

The steps followed to perform ontology validation are as below:

1. Select crop variety for which validation is to be performed. Possible types are year-wise, district-wise, village-wise.
2. For each of these varieties, calculate parameters such as

    a. No. of times activity is performed
    b. Start time of the activity

   c. Period/interval of the activity
   d. Quantity

3. Find the average for all the crop varieties.
4. Compare this with the similar parameters calculated from ontology.
5. In case of differences, find whether maximum farmers are above the threshold calculated from crop varieties.
6. If so, modify the ontology. Else, indicate the farmers who do not follow the practice.

4 Conclusions and Future Work

The designed system handles farmer's crop-related query that is posted based on observations about the crop. OntoAQ handles the query by searching a knowledge base ontology stored as a graph in database performing a graph-based search. The search over ontology graph is aided by filtering information based on farmer's context as well as weather predictions made. Finally, user is presented with set of paths from the ontology, which best answer his query.

We believe that OntoAQ is a unique ontology-based approach for assisting farmers that utilizes graph matching on domain ontologies and also enables specification of context. Its use in the context of cotton farming is demonstrably advantageous. The use of ontologies toward developing a knowledge-based approach to agriculture is been increasingly seen as the way forward [15], and OntoAQ demonstrates the capability of such an approach.

4.1 Context-Based Spatial Search

Question answering for farming-related queries has gained much traction in recent years, and multiple systems have been proposed [16, 17]. While the traditional ontology-based question answering has served many domains [18], contextual information in the agricultural domain is critical for their pragmatic use. Answering the user query by referring to farmer's contextual information currently deals with the temporal aspect of the activities performed by him. Suggestions are made to the farmer by verifying the stage of the crop, the time for performing certain activities, time of attack of pests, time of occurrence of a disease, time of performing control measures. Apart from looking into the temporal part, spatial aspect of it can also be considered. This necessitates finding relatedness between the farmer's crop and the one's sown in his nearby villages. Some of the scenarios like a disease affecting a crop would lead farmers in that location to be warned prior the occurrence of the disease as it can spread in their fields too. Also other cases would be that if a farmer has his field near a tube well, he can make use of it or we could suggest him a way of irrigation using such assets. It is also important to utilize upper-level ontologies of space and time [19] in developing the spatiotemporal context in the system

4.2 Extensions of OntoAQ

The system designed answers the query better by referring to farmer's contextual information. The current system takes care of this as each of the activity performed by the farmer is validated against the knowledge in the ontology, and the farmer is recommended or warned suitably. It can be further extended by validating the farmer's current activity with the activity performed for the same variety, at the same location and in the same season last year, which gave the best yield. This would assist in confirming whether the farmer is following the practice, which is best in his area. Crop knowledge evolves from time to time by the new observations seen by farmers, impact of specific local conditions, experimentation done in consultation with agriculture research institutes, etc. Such knowledge is not captured in the system and requires ontology to be extended for conveying such knowledge.

References

1. Kawtrakul, A.: Ontology engineering and knowledge services for agriculture domain. J. Integr. Agric. **11**(5), 741–751 (2012)
2. Kaufmann, E., Bernstein, A., Zumstein, R.: Querix: a natural language interface to query ontologies based on clarification dialogs. In: 5th International Semantic Web Conference (ISWC 2006). Springer (2006)
3. Lei, Y., Uren, V., Motta, E.: Semsearch: a search engine for the semantic web. In: International Conference on Knowledge Engineering and Knowledge Management. Springer (2006)
4. Kaufmann, E., Bernstein, A., Fischer, L.: NLP-Reduce: a naive but domain independent natural language interface for querying ontologies. In: 4th European Semantic Web Conference ESWC (2007)
5. Tran, T., et al.: Ontology-based interpretation of keywords for semantic search. In: The Semantic Web, pp. 523–536. Springer (2007)
6. Wang, H., et al.: Q2semantic: a lightweight keyword interface to semantic search. In: European Semantic Web Conference. Springer (2008)
7. Kollia, I., Glimm, B., Horrocks, I.: SPARQL query answering over OWL ontologies. In: Extended Semantic Web Conference. Springer (2011)
8. Sini, M., et al.: Knowledge Models in Agropedia Indica (2009)
9. Ratnam, B., Reddy, P.K., Reddy, G.: eSagu 1: An IT based personalized agricultural extension system prototype-analysis of 51 Farmers' case studies. Int. J. Educ. Dev. Using Inf. Commun. Technol. **2**(1), 79 (2006)
10. Pande, A.K., Jagyasi, B.G., Jain, R.: mKRISHI: A Mobile Multimedia Agro Advisory System for Remote Rural Farmers. IEEE (2011)
11. Unger, C., et al.: Template-based question answering over RDF data. In: Proceedings of the 21st international conference on World Wide Web. ACM (2012)
12. Vargas-Vera, M., Motta, E., Domingue, J.: AQUA: an ontology-driven question answering system. In: New Directions in Question Answering (2003)

13. Sahni, S.: Ontology Based Agro Advisory System. Department of Computer Science and Engineering, lIT Mumbai, M. Tech. thesis (2012)
14. Sen, S.: Human perspectives in semantic translations—role of entity functions. In: GI Forum. Salzburg: Springer-Wichmann (2007)
15. Abrahão, E., Hirakawa, A.R.: Task ontology modeling for technical knowledge representation in agriculture field operations domain. In: 2017 Second International Conference on Information Systems Engineering (ICISE). IEEE (2017)
16. Devi, M., Dua, M.: ADANS: an agriculture domain question answering system using ontologies. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). IEEE (2017)
17. Walisadeera, A.I., Wikramanayake, G.N., Ginige, A.: An ontological approach to meet information needs of farmers in Sri Lanka. In: International Conference on Computational Science and its Applications. Springer (2013)
18. Konys, A.: Knowledge-based approach to question answering system selection. In: Computational Collective Intelligence, pp. 361–370. Springer (2015)
19. Bittner, T., Smith, B.: Formal ontologies for space and time. IFOMIS, Department of Philosophy. Leipzig, Buffalo, University of Leipzig, University at Buffalo and NCGIA, vol. 17 (2003)

Semantic Interpretation and Integration of Open Data Tables

A. Subramanian and S. Srinivasa

1 Introduction

A large chunk of open data is made available through open government initiatives, such as data.gov,[1] data.gov.uk,[2] data.gov.in,[3] mostly in the form of csv files containing tabular data. These tables normally contain information regarding government administration, policies and indicators in diverse sectors like ‘Agriculture’, ‘Health and Family Welfare’ or ‘Environment’. Since open data is generated with no preconceived data models, there is no overarching data model or purpose that guides the consumption of such tables. Each table, however, represents a set of *subjects* or *themes* for which related data has been captured. Tables from the ‘Agriculture’ and ‘Industry’ domain describing say *crop production*, *yield*, *nutritional value* and *export value* can be integrated along the **State or District or Market** perspective that produced the crops or from the **Crop** perspective indicating the countries that it was exported to or its nutritional content. Data related to a particular theme could be collected across various tables or a single table could contain information regarding multiple themes. A given set of tables can be aggregated along several perspectives—each of which may have different extents of ‘footprint’ on the data. Not all such characterization are equally descriptive of the data; nevertheless, it is desirable to create multiple such perspectives to obtain richer understanding of the data. This is a non-trivial task, and we address this problem by the process of *Thematic*

[1]US Government Open Data: https://www.data.gov/.
[2]UK Government Open Data: https://data.gov.uk/.
[3]Open Government Data (OGD) Platform India: https://data.gov.in/.

A. Subramanian (✉) · S. Srinivasa
International Institute of Information Technology, 26/C, Hosur Road, Electronics City Phase 1, Electronic City, Bengaluru 560100, Karnataka, India
e-mail: asha.subramanian@iiitb.org

S. Srinivasa
e-mail: sri@iiitb.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_17

Integration of tables. Thematic Integration infers the various themes or perspectives from a collection of tables and consequently using a formal model, extracts the footprint of each table contributing to the various themes.

Multiple research efforts have addressed the problem of semantic annotation of web tables[4] and csv tables,[5] which mainly involves interpreting tabular data by linking them to relevant vocabularies (e.g., [2, 8, 9, 15]); however, they have not focused on the problem of **Thematic Integration** of tables.

Semantic Annotation annotates each table in a collection independently using vocabularies to determine each column type, cell entries and the relationships between each pair of columns if any. **Thematic Integration**, on the other hand, observes a collection of tables and extracts characteristics that bind the collection in order to facilitate interesting inferences.

Our contributions include (1) a suite of algorithms for the extraction of themes given a collection of tables and the schematic characterization of each table in the collection using the themes. Both the theme extraction and schematic characterization use the LOD[6] as reference vocabularies. (2) The 'Many Worlds on a Frame' (MWF) knowledge representation framework to organize the knowledge from the semantic integration. (3) Implementation of MWF using RDF N-Quads[7] such that the resultant 'Open Knowledge Graph' can be queried by themes or schematic associations.

2 Motivating Example

The example described in Fig. 1 illustrates the utility and effectiveness of Thematic Integration for a collection of tables. The goal is to use LOD vocabularies such as DBpedia,[8] YAGO [13] and custom vocabularies if available, to find one or more overarching themes that can be used to describe a given collection of tables.

Figure 1 shows a set of tables obtained from data.gov.in, namely *RiceProduction.csv*, *NutrientContent.csv* and *IndianStates.csv*. They depict market-wise rice prices in various Indian states and districts, nutrient content against various parameters in Indian food crops, and geographical information regarding various Indian states, respectively.

The common themes that we could extract for this collection are namely 'States and Territories of India', 'City' and 'Rice'. Hence, when viewed from the perspective of an Indian state, this aggregation can reveal data about production of different crops including 'Rice' and geographical information for that state. Similarly, when viewed from the perspective of the crop 'Rice', this aggregation can reveal rice

[4]Web Tables: http://webdatacommons.org/webtables/.

[5]csv: https://en.wikipedia.org/wiki/Comma-separated_values.

[6]The Linking Open Data cloud diagram: http://lod-cloud.net/.

[7]N-Quads: https://www.w3.org/TR/n-quads/.

[8]DBpedia: http://wiki.dbpedia.org/.

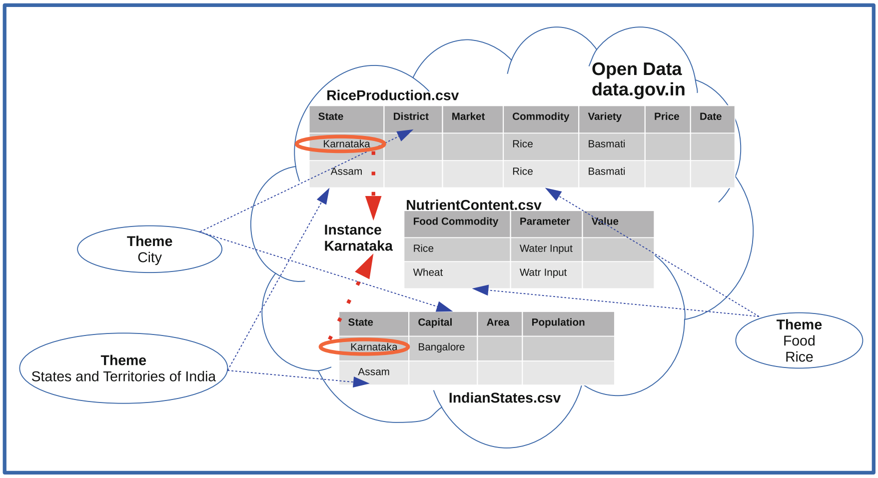


Fig. 1 Motivating example illustrating Thematic Integration of open data tables

production data across different states and its nutrient content. The result of several such aggregations is a theme driven Open Knowledge Graph enriched with LOD classes and properties to facilitate query and logical inferences.

This form of semantic integration when applied to large number of tables allows for: (1) extraction of tables by specific themes; (2) extraction of multiple themes that are applicable to a single table; (3) integrated view of instance data across the various tables; and (4) extracting the relations (schematic associations) that the themes drive across various tables.

3 Thematic Integration Model

We present a model that addresses the Thematic Integration in two stages: *Theme Identification* and *Schematic Characterization*. We define *Themes* and *Anchoring Columns* for a collection of tables as follows. Our model works on tabular data in csv format. Data in other formats such as JSON or XML is converted into csv format as part of a pre-processing routine.

Definition 1 Themes: Classes or concepts from a vocabulary that best explain topics or contexts for a collection of tables and aligns with one or more columns from tables such that the intent of the class not only subsumes the majority of instances or data values in those columns, but also has enough support from the rest of the columns in its respective table to explain the complete semantics of the table.

Definition 2 Anchoring Column(s): A combination of columns (single or multiple) that drives the theme or subject of a table and forms relationships with the other non-subject columns in the table, thus explaining the semantics of the complete table from the perspective of the theme.

The *Theme Identification* process extracts the **dominant themes** and the respective **anchoring columns** that drive the dominant themes for a collection of tables. The *Schematic Characterization* process decodes each individual table in the light of the extracted dominant themes for the collection, by mining the relations that the non-subject or non-anchoring columns form with the anchoring columns of the table.

3.1 Preliminaries

Let T be a collection of tables (csv format) where for each $t_i \in T$, the term c_{ij} denotes the jth column of the table t_i. Let K represent a the set of LOD vocabularies or custom vocabularies used to interpret T. Let $d_r(ij)$ denote the data in the rth row for the ith table in the jth column, $label(c_{ij})$ denote the attribute name or column name of column c_{ij}.

Candidate Concepts
The first step is to determine a set of candidate concepts for each c_{ij}, given $t_i \in T$. Each c_{ij} could be of two types, one where $d_r(ij)$ across r have a matching resource in K or c_{ij} is a literal-value column that contains arbitrary numbers, text or dates. For the latter, we use basic types such as *xsd:string*, *xsd:date* or *xsd:number* to classify them. For vocabulary match-able c_{ij}, we use two techniques that searches K for a class or a category that maximally subsumes the entities found in the column across $d_r(ij)$. This includes an existing well- known entity disambiguation engine: DBpedia Spotlight [4] and an internally developed search and matching algorithm using SPARQL.

Associations
Associations are relations between a pair of columns in a table, and we use three different techniques to determine them:

1. For every pair of columns that have a non-empty set of candidate concepts, we look for properties from K in either direction using query tuples of the form $(d_r(ij), ?p_1, d_r(ik))$ and $(d_r(ik), ?p_2, d_r(ij))$. Here $?p_1$ and $?p_2$ denote the rdf:property that links the data values between columns c_{ij} and c_{ik} and vice versa. $d_r(ij)$ and $d_r(ik)$ are data elements from the respective columns c_{ij} and c_{ik} across r. For the property $?p_1$, c_{ij} denotes the *subject column* and c_{ik} denotes the *object column*.

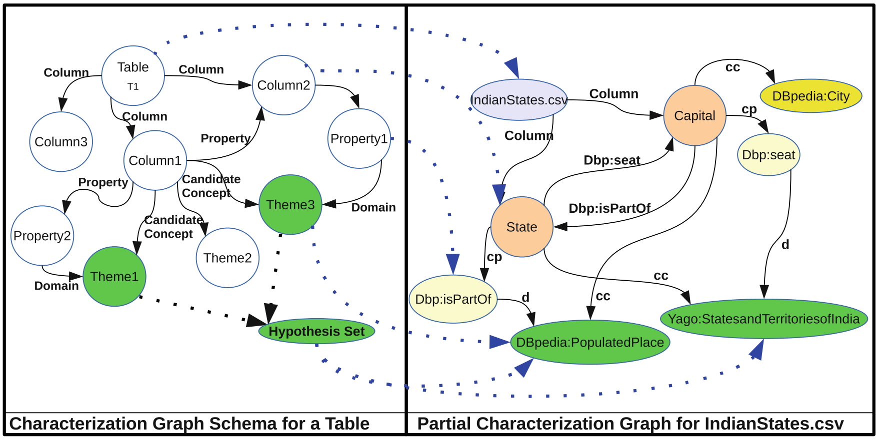


Fig. 2 Characterization graph (Schema) and populated characterization graph for the table IndianStates.csv mentioned in Fig. 1. The dotted lines explain the population of the characterization graph from the schema

2. A pair of columns could be related through a category hierarchy relationship; we use this observation too to mine possible relations between columns. Specifically, we use the SKOS vocabulary to determine such associations.
3. For literal-value columns, we generate synthetic properties based on the class from K that the values across the column represent.

We build a characterization graph for each table $t_i \in T$ (Fig. 2). A node of type *table* is created for each table in the collection and is connected with its respective *column* nodes with a *column* link. Column nodes are connected to respective candidate concept nodes with links labelled either *cc* or *sc* depending on whether a class or a category matched the column from K. A *property* node represents an association connecting a subject column and an object column. Object columns are connected to *property* nodes using a *cp* link. A *property* node is associated with a *class* node with a *d* link, representing that the said class forms the domain of the property. Ideally, these domains coincide with a subset of the candidate concepts for some column. In cases where a property does not have a rdfs:domain defined, we look at the subject column and link the *property* node to the class(es) from K that link to this subject column with an existing *cc* link. In cases, where the *class* node is a category from K defined in the SKOS vocabulary, we use the concept of 'broader' to interpret the domain of a property. In such cases, an *sb* link is used to connect the property node with the *class* node. The object column in such cases connects to the property node with a *sc* link. Finally, the *property* label for an edge refers to the actual property URI from the K.

We build the consolidated characterization graph for the entire collection of tables connecting the individual characterization graphs for each table by the common candidate concepts that they represent. Figure 2 shows a partial characterization graph for the table *IndianStates.csv*.

Let $\mathscr{C}$ be the set of candidate concepts for the collection of tables T. The *Hypothesis Set* $\mathscr{H} \subseteq \mathscr{C}$ is the set of all candidate concepts—*class* nodes which have an incoming *cc* as well as *d* or *sb* links. The concepts belonging to the *Hypothesis Set* not only describe columns, but also have enough properties associated with them, so as to potentially contain a theory to describe the collection of tables. Each concept in the *Hypothesis Set* represents a potential theme with which the collection of tables can be characterized.

Every concept $h \in \mathscr{H}$ in the *Hypothesis Set* is associated with an *anchoring column*, which is the central entity of inquiry for this perspective. The anchoring column is in turn associated with a set of properties and their object columns, which collectively form a characterization. We are interested in scoring the members of the *Hypothesis Set* to narrow down on a set of themes which summarize the collection of tables.

3.2 Formalization

Formally, for every candidate concept γ_k mined from K, we compute its data support (Concept Score) for the column c_{ij} as follows:

$$f_k(c_{ij}) = \frac{|\gamma_k|}{|c_{ij}|} \tag{1}$$

Here, $|\gamma_k|$ is the number of data elements from column c_{ij} which identified the candidate concept γ_k and $|c_{ij}|$ represents the total number of data elements in column c_{ij} for the table t_i.

Similarly, we determine the support for each property p, PS (Property Score), that is associated with pairs of columns c_{ij} and c_{ik} for a table as follows:

$$PS(p) = \frac{|\gamma_{jk}|}{\max(|c_{ij}|, |c_{ik}|)} \tag{2}$$

Here, γ_{jk} is the set of all query tuples of the form $(d_r(ij), ?p, d_r(ik))$ matching property $?p$ across r. The Property Score for a property captures how well the property subscribes to a relationship between two columns in a particular direction. Given a collection of tables T and a collection of vocabularies K, we are interested in finding a set of themes $(\hat{h})$ and its anchoring columns $(\hat{c})$ and deriving the corresponding schematic characterization such that:

$$\mathscr{S}(T) = \mathscr{S}(\hat{h}, \hat{c}) = \bigcup_{t_i \in T} \hat{Tup_i}(\hat{c}, \hat{P}; h) \tag{3}$$

where $\mathscr{S}(T)$ represents the schematic characterization of T and $\hat{Tup_i}(\hat{c}, \hat{P}; h)$ are the associated tuples for a table $t_i \in T$ given its anchoring columns $\hat{c}$ and its associated set of properties $\hat{P}$ for a given theme h.

The subsequent sections propose a solution for the above problem in the form of a mathematical model using probabilistic techniques, information theory and diffusing computations.

3.3 Proposed Model

We propose a maximum likelihood equation for identifying the themes, its anchoring columns and associated properties in line with the definitions of themes and anchoring columns using the Concept Score and Property Score as follows:

For a given concept $h \in \mathscr{H}$ from the Hypothesis Set, its anchoring column $(\hat{c})$ and its associated set of properties $\hat{P}$ are determined as those columns $(\hat{c})$ that not only have a maximum Concept Score (Eq. 1), but also have maximum Property Scores (Eq. 2) from the properties $\hat{P}$ that link the other non-subject columns with the anchoring columns $(\hat{c})$. Formally, this is modelled as the following optimization problem:

$$(\hat{c}, \hat{P}) = \arg\max_{c,P} f_h(c) \sum_{\forall p \in P} PS(p) \tag{4}$$

Here, $f_h(c)$ denotes the Concept Score for concept h, for column c and $PS(p)$ is the Property Score for property p.

3.3.1 Scoring the Schematic Characterization

We build a decision table to score $\mathscr{S}(\hat{h}, \hat{c})$ against the collection of tables.

Table 1 shows the schema of the decision table. The columns are as follows: Dominant Concept belonging to Hypothesis Set (h_k), Property p_{kj} linking to h_k with a d or sb link, Property Score of property p_{kj} (Eq. 2), anchoring column c_{ik} from table t_i linking to h_k, Object Column of property p_{kj} say c_{ij} and Concept Score for

Table 1 Schema of the decision table

Concept	Property	Property Score	Anchoring column	Object column	Support
h_k	p_{kj}	$PS(p_{kj})$	c_{ik}	c_{ij}	f_k

the anchoring column f_k (Eq. 1). This table is used to score each of the concepts in the Hypothesis Set.

Let (c_{ik}, p_{kj}, c_{ij}) denote the set of all tuples linking to concept h_k where c_{ik} denotes the anchoring column, c_{ij} the object column and p_{kj} the candidate association between c_{ik} and c_{ij} that has the maximum likelihood score $(PS(p_{kj}))$ based on Eq. (4).

1. The *tabular support* score for a concept h_k in a table t_i defines the coverage of the concept for the table calculated as the number of distinct columns from the (c_{ik}, p_{kj}, c_{ij}) tuples that link to h_k, for a given anchoring column c_{ik} normalized by the total number of columns in t_i. Let N_{kj} denote the number of distinct columns from the table that link to h_k via anchoring column c_{ik}. Let the term $cscols(t_i)$ indicate the number of columns in table t_i. The tabular support score for concept h_k is given by:
$$tabsupport(h, t_i.c_{ik}) = \frac{N_{kj}}{cscols(t_i)} \tag{5}$$

2. Confidence for concept h_k in a table t_i is determined by the strength of the associations linking to h_k from the various columns in the table t taking into account the Concept Score of the anchoring column.
$$confidence(h_k, t_i.c_{ik}) = \frac{\sum\limits_{(c_{ik}, p_{kj}, c_{ij})} f_k(c_{ik}) \cdot PS(p_{kj})}{N_{kj}} \tag{6}$$
Since f_k and PS both range from 0 to 1, confidence is a measure in [0, 1].
3. Overall cluster score $OCS(h_k, t_i.c_{ik})$ is a consolidated score that measures the applicability of a hypothesis h_k (concept, its anchoring column and associations) to a cluster of columns in a table t_i. It is calculated as:
$$OCS(h_k, t_i.c_{ik}) = \|cv(h_k, t_i.c_{ik})\|_2 \cdot CoSim(cv(h_k, t_i.c_{ik}), [1, 1]) \tag{7}$$
where: $cv(h_k, t_i.c_{ik})$ is the vector $[tabsupport(h_k, t_i.c_{ik}), confidence(h_k, t_i.c_{ik})]$ and $\|cv(\ldots)\|_2$ is the L2 norm of the vector represented by $cv(h_k, t_i.c_{ik})$. $CoSim(cv(h_k, t_i.c_{ik}), [1, 1])$ is calculated as
$$CoSim([a, b], [1, 1]) = \frac{a + b}{\sqrt{2} \cdot \sqrt{a^2 + b^2}} \tag{8}$$

Thus, the overall cluster score determines the likelihood of a Dominant Concept's suitability for a table by comparing it with the 'ideal' distribution ([1, 1]) of tabular support and confidence scores.

Extracting T-Box (Terminological facts) and A-Box (Assertional facts)
The combination of a theme, its anchoring column and associations represents a *schematic association* for a table, and we use the overall cluster score to determine

the best schematic associations for a table. These schematic associations form the T-BOX. The individual instances across the rows and the mined relations between them represent the A-Box. For tables characterized by multiple themes and hence multiple schematic associations, a combination of ('parent theme', 'child theme') and ('parent anchoring column', 'child anchoring column') is used to represent the T-BOX.

3.3.2 Identifying the Best Themes and Anchoring Columns

The set of themes in the Hypothesis Set and their associated set of anchoring columns form a bipartite graph, where each theme is pointed to by one or more anchoring columns and each anchoring column points to one or more themes. We formulate a recursive definition of what constitutes a good anchoring column and a good theme. A 'good' anchoring column is defined as one that is associated with a 'good' theme and vice versa. The above definition leads to the formulation of a diffusing computation, derived from the hyperlink-induced topic search (HITS) algorithm [6].

HITS algorithm uses an initial set of weights for each of the nodes in the bipartite graph. We start with an initial weight vector of **1** for the set of all anchoring columns. For the theme, we use a score capturing the Information Content, or the specificity of the concept as its initial weight. This is calculated as:

$$I(h) = \log(\frac{N}{N(h)}) \tag{9}$$

where: $N(h)$ is the number of resources in K of type(rdf:type) h and N is the total number of resources of any type in K.

We then iteratively calculate the two sets of weights alternatively until the weights for the anchoring columns and concepts reach a fixed point. The two update operations on weights translate to:

$$\begin{aligned} v^i &= A^\tau \cdot u^{i-1} \\ u^i &= A \cdot v^{i-1} \end{aligned} \tag{10}$$

Here, A is the matrix of overall cluster scores—OCS(h_k,$t_i.c_{ik}$) (Eq. 7), for each theme in the Hypothesis Set and each anchoring column. u is the vector of weights for the anchoring columns and v is the vector of weights for the themes.

The superscript represents the iteration number in a power-iteration method for computing the scores. The iterations are said to converge when the relative ordering of the scores become stable across iterations.

3.3.3 Enhancing the Theme Identification

This module ensures that the model generates pertinent themes by filtering out the themes such as `owl:Thing`, `http://dbpedia.org/class/yago/Yago-`

GeoEntity, http://dbpedia.org/class/yago/YagoLegalActor which are less informative. The model also ensures that the theme extraction process presents unique themes for a collection rather than displaying different flavours of the same class from other vocabularies. Consider class 'http://dbpedia.org/ontology/City': classes 'http://schema.org/City' and 'http://www.wikidata.org/entity/Q515' have similar intent albeit from other ontologies such as 'schema.org' and 'wikidata.org'. It is enough to extract the theme for **city** once from the DBpedia ontology. We handle this by using the `owl:equivalentClass` property.

3.3.4 Handling Numeric Literal-Value Columns

Numeric literal-value columns are columns in tables that contain arbitrary numbers whose values cannot be associated with distinct URIs from known vocabularies. We use the characteristics of the data distributions of numeric columns, coupled with the combination of a k-NN classifier (bottom-up approach) and mining the LOD using the column label (top-down approach) to learn the semantics. Using this approach, we have implemented a domain specific algorithm to learn semantics of numeric literal-value columns for a set of select labels. This approach can be extended for more domains and labels depending on the availability of training data in the respective domains. For columns that represent concepts such as 'http://dbpedia.org/resource/Unique_identifier' for serial number columns and others where an equivalent property (of type rdf:property) is unavailable in LOD, we create synthetic properties derived from the concepts that the column data represent to generate the appropriate semantics for the numeric literal-value columns.

4 Semantic Integration Using Many Worlds on a Frame (MWF)

MWF—**Many Worlds on a Frame** [12] is an intuitive knowledge representation framework loosely modelled on *Kripke* semantics [7]. It allows for facts to be represented, grouped and related across many interconnected worlds. Each world is considered a concept and concepts are organized in hierarchies, represented as rooted, acyclic graphs. Every world has a type and a location, both of which refer to other worlds. This gives rise to the is-a (concept hierarchy) and is-in (containment hierarchy) hierarchies. The concept hierarchy is used to inherit properties and associations, and the containment hierarchy is used to inherit 'access pathways' and manage visibility. The root of the concept hierarchy is a concept called **Concept**, and the root of the containment hierarchy is a concept called **Universe of Discourse (UoD)**. Each concept in a MWF system acts as a local **context world** and hosts a set of knowledge fragments in the form of **associations** across concepts. **Associations** are triples of the form (source, predicate, target). Here source and target are concepts in some target world say C_w and predicate is a label describing the association. In any association contained in world C_w, if the target concept is the world C_w itself, such associations are called **Roles**. The source concept is said to be playing a role defined by the predicate label in C_w. A concept that cannot be subclassed using the 'is-a' relation is called an **Instance World**. The 'context worlds', 'instance worlds', their properties, interdependencies in the form of associations, roles, the 'is-a' and 'is-in' hierarchies between these worlds make up the 'Frame'. We use three tables from data.gov.in to explain the semantic integration using MWF namely *AgmarkRice2012.csv*, *NutrientContent.csv* and *IndianStates.csv*. These datasets contain market-wise rice prices in various Indian states and districts, nutrient content against various parameters in Indian food crops and geographical information regarding various Indian states, respectively. Here, 'Yago/YagoPermanentlyLocatedEntity' and 'dbo:Food' are the most pertinent themes produced by the 'Thematic Integration' model for the collection of tables (*AgmarkRice2012.csv*, *NutrientContent.csv* and *IndianStates.csv*). These themes translate into 'context worlds' in MWF. The schematic associations from the model translate into associations in MWF. We have illustrated the components of two 'context worlds' in Fig. 3. Note that table *AgmarkRice2012.csv* constitutes a complex subject determined by the columns—*State* and *Commodity* and has been consumed in two contexts. The context 'dbo:Food' shows the commodities that the various states sold while the context 'YagoPermanentlyLocatedEntity' shows the same table from the states perspective. The parent and child **associations** depicted by *(State, Commodity)* and *(Commodity, State)* in their respective contexts hold the complete semantics of the table *AgmarkRice2012.csv*.

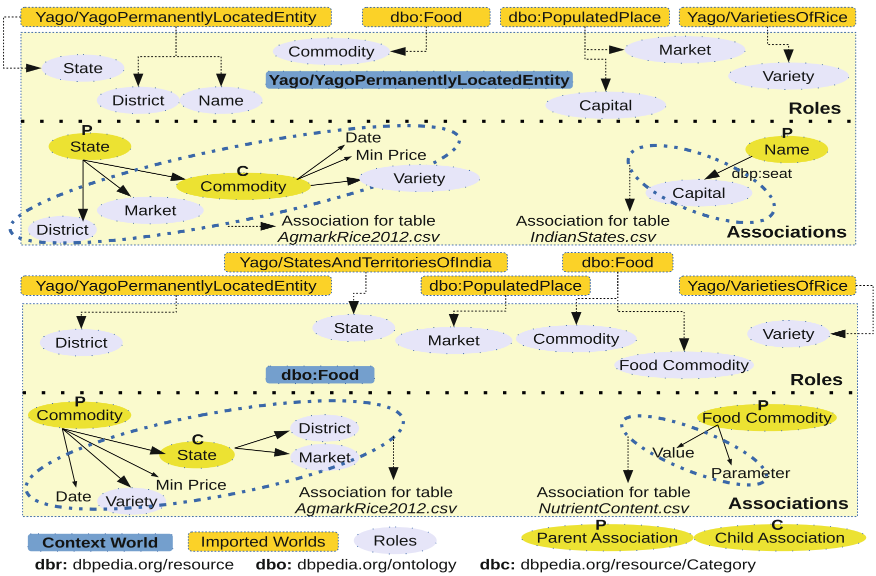


Fig. 3 Illustration of the working model of Many Worlds in a Frame (MWF) using open data tables

Similarly, table *IndianStates.csv* using the association *Name*, associates with the context 'Yago/YagoPermanentlyLocatedEntity'. The table *IndianStates.csv* constitutes a simple subject with the anchoring column *Name* explaining all the columns of this table. The semantics of this table is explained by the association *Name*.

Knowledge Representation—The Open Knowledge Graph
We generate an Open Knowledge Graph from the T-Box and A-Box generated by the 'Thematic Integration' model in the form of N-Quads, using the 'context' in the N-Quads notation to store the graph name that each triple belongs. The MWF framework dictates the knowledge representation semantics used to generate the N-Quads.

The metadata for the graph name stores the themes, table and anchoring columns associated with the context. This allows us to SPARQL the knowledge graph associated with a collection of tables by themes, table properties such as 'Domain', 'Source' or schematic associations.

5 Empirical Evaluation

We empirically study the effectiveness of the outputs of 'Theme Identification' and 'Schematic Characterization' for a collection of tables, generated by our model. In particular, we assess the capability of our model to extract pertinent themes given a collection of tables and the accuracy of the model to map these themes and its associated properties with individual tables in the collection. Our evaluation tables consisted of over 100 files, downloaded from Open Government Data (OGD) Platform India[9] from various sectors such as 'Agriculture', 'Health and Family Welfare', 'Environment' and tables used by Limaye et al.[10]

5.1 Implementation

Our knowledge base is a locally installed Virtuoso database comprising of vocabularies and ontologies from the LOD cloud, namely DBpedia (2015-04 version), YAGO (2015-11-04 version), Wikidata (2015-10-26 version), Schema.org (2015-11-04 version), Linkedgeodata (2014-09-09 version) and Umbel (2015-04 version). Our knowledge base contains approximately 1 billion RDF triples.

[9]OGD India: https://data.gov.in/.

[10]Limaye Datasets: https://www.cse.iitb.ac.in/~sunita/wwt/#data.

Table 2 Evaluation sheet per table for each ranked theme

Theme	Child theme	Table	Anchoring column	Connected column	Property

5.2 Experimental Method

We used human evaluators as the gold standard for the evaluation of the *Thematic Integration* outputs. For assessing **Theme Identification**, we presented groups of tables to evaluators and recorded their inputs on pertinent themes. We then matched the closest LOD class/concept relating to these themes (Set A). The model results from 'Theme Identification'—ranked themes for a collection, where then presented to the evaluators and the users marked each theme as 'applicable/not applicable'. The themes marked by the evaluators as 'applicable' were consolidated as Set B. $\frac{\mathbf{card}(A \cap B)}{\mathbf{card}(A)}$ denotes the ability of the model to **recall** the themes identified by the evaluators and $\frac{\mathbf{card}(B - A)}{\mathbf{card}(A)}$ captures the **novelty** of the model to suggest novel themes not previously identified by the evaluators.

For evaluating **Schematic Characterization**, we generated the following evaluation sheet (Table 2) for each table using the schematic characterization outputs generated by the model.

The model outputs were evaluated for: (1) **applicability of themes** to the individual table in the collection and (2) **validation of properties**—verifying the mined relationships between the anchoring columns and the connected columns in the table. Human evaluators validated evaluation sheets for each table in two separate evaluation outputs: (1) for applicability of themes and (2) validation of properties, using scores that indicate 0 for disagreement with the model output, 1 for moderate agreement with the model output and 2 for total agreement with the model output. Each evaluation sheet was validated by a minimum of 2 evaluators.

5.3 Metrics

We report evaluation results for **Theme Identification** on Recall and Novelty of the model. We report results from the evaluation of **Schematic Characterization** on (1) Error, (2) Accuracy, (3) Precision, (4) Recall and (5) F1-score. Further, we also analysed inter rater agreement and report results using Cohen's kappa[11] and Fleiss' kappa[12] coefficients. A confusion matrix (Table 3) captures the comparison of model generated output and human evaluation. From the confusion matrix entries, Error is calculated as $\frac{\textit{False Positives} + \textit{False Negatives}}{\textit{Total}}$, Accuracy is calcu-

[11] https://en.wikipedia.org/wiki/Cohen%27s_kappa.

[12] https://en.wikipedia.org/wiki/Fleiss%27_kappa.

Table 3 Confusion matrix for 'applicability of themes' and 'validation of properties'

		Model reports →	
		Theme applicable	Theme not applicable
User	Confirms theme	Both evaluators rate 1 or 2 (True Positives)	At least one evaluator rates 0 (False Negatives)
	Rejects theme	At least one evaluator rates 0 (False Positives)	Both evaluators rate 1 or 2 (True Negatives)

Table 4 Evaluation Results. Table 4a shows the evaluation summary for Schematic Characterization outputs. Value (I) are the evaluation results for 'Applicability of Themes' and Value (II) are the evaluation results for 'Validation of Properties'. The Cohen's kappa and Fleiss' kappa coefficients for measuring the inter rater agreements have also been illustrated for each of the experiments. Table 4b shows the evaluation for Theme Identification, reporting results on Recall and Novelty capabilities of the model

(a) *Schematic characterization*

Evaluation parameter	Value (I)	Value (II)
Cohen's kappa/Fleiss' kappa	0.502	0.486
Error	0.147	0.102
Accuracy	0.852	0.897
Precision	0.931	0.914
Recall	0.909	0.979
F1-score	0.919	0.945

(b) *Theme identification*

Scores	Recall	Novelty
Average	0.97	0.51

lated as $\frac{True\,Positives+True\,Negatives}{Total}$, Precision is calculated as $\frac{True\,Positives}{True\,Positives+False\,Positives}$, Recall is calculated as $\frac{True\,Positives}{True\,Positives+False\,Negatives}$ and finally F1-score is calculated as $\frac{2\times Precision\times Recall}{Precision+Recall}$.

5.4 Discussions

The experiments suggests that the themes extracted from the 'Thematic Integration' process explain the tables in the collection quite effectively. The precision, recall and F1 scores are above 90% for both the 'Applicability of Themes' and 'Validation of Properties'. The inter rater agreements in the range of 0.45–0.50 indicate a 'moderate' to 'substantial' agreement among the evaluators. To summarize, the observed results emphasize that the intuition and formalization of the model to automatically learn the themes and anchoring columns given a collection of tales is sound. Secondly, each individual table contributes to a footprint of themes inferred for a collection of table.

This footprint can be extracted by identifying the contributing anchoring columns and its properties. Evaluation results for Schematic Characterization is backed by this intuition implemented by the maximum likelihood equation (Eq. 4).

6 Related Work

A number of research efforts have addressed the *Semantic Annotation* problem for structured datasets using LOD and custom vocabularies. They focus on interpreting a single table, by using knowledge from the LOD. In contrast, our work addresses the problem of *Thematic Integration* of a collection of tables. However, it is indeed pertinent to discuss the state-of-the-art efforts in this area and contrast with our model. Semantic annotation research focuses largely on web tables and csv tables, [1, 2, 5, 8, 10, 14] to name a few. Mulwad [10] presents a comprehensive approach to determining the meaning of a table and uses a parameterized graphical model to represent a table. The core module performs joint inference to infer the meaning of the table by using a semantic message-passing scheme. Zhang [15] uses a similar probabilistic graphical model to interpret web tables, but argues that the data from the complete table need not be used to infer semantics. Additionally information about a table, including text outside the tables and pre-defined semantic markups within the web pages is utilized as valuable clues towards the semantic interpretation. Limaye's work [8] use a machine-learning and probabilistic graphical model to semantically annotate web tables; multiple potential functions are utilized to determine entities, classes and relationships that best characterize the cell values, column headers and associations between columns. Chandrashekar et al. [2] focus on *Entity Linking* task for values in web tables using a supervised learning approach which disambiguates each cell value in the web table to known entities from a knowledge base(YAGO). Their approach moves away from relying on a knowledge base for entities, type classifications and property definitions. It is instead, based on the co-occurence of a pair of entities in Wikipedia documents and tables. Braunschweig et al. [3] extract column specific information from the context of web tables. They mine this information using directly related context and indirectly related context, the former using matched instances of phrases in the column header in the surrounding text and the latter using a taxonomy from an external knowledge base. Ritze et al. [11] prepare a T2D gold standard, a collection of schema level and entity level correspondences between the Web Tables Corpus and DBpedia. They use this gold standard to evaluate the performance of their iterative matching technique T2K Match. Zwicklbauer et al. [16] figure out table header information from types of cell columns and also provide insights on the quality of ontology and the number of data points required to extract the information.

Our work emphasizes on the objective of Thematic Integration and works on a collection of tables, generating common themes that best relate to a collection of tables. This approach is specifically useful when dealing with open government data since multiple tables capture information regarding common topics, thus allowing the consumer to view various perspectives of the tables in the light of different themes.

7 Conclusions and Future Work

Since most of the open data especially data from open government data portals constitute georeferenced data (places/locations) along with data on various indicators such as health, climate, education, our model can be successfully applied to semantically integrate such open data with minimal human intervention. We present 'Sandesh'—the semantic data mesh over Indian Open Government Data as a demonstration of our model.[13] 'Sandesh' semantically integrates a collection of open data csv files enabling exploration through intuitively traversable thematic perspectives. While other related work mostly confine to inferring semantics for a single table, our research focuses on collections of tables to infer themes that best categorize a collection. Further, we have also presented an intuitive knowledge representation model that allows the Thematic Integration results to be organized into related worlds. The resultant implementation is an Open Knowledge Graph facilitating a semantic mesh of interconnected tables using the themes that they represent.

The model however does need improvement in filtering out incorrect assignments from the knowledge base by using the context information better. Introducing the *human in the loop* might be an effective way to address this limitation. The Open Knowledge Graph generated by our model is currently used to query the collection for retrieval of related information based on themes, tables or associations. It will be an interesting exercise to integrate OWL semantics and explore reasoning on the tables by using standard reasoners such as Apache Jena.

Acknowledgements We would like to thank the members of the Web Science Laboratory, IIIT Bangalore, for their support in the meticulous evaluation of the model results for over 100 tables and also help prepare the ground truth for comparison against the baseline.

References

1. Auer, S., Dietzold, S., Lehmann, J., Hellmann, S., Aumueller, D.: Triplify: light-weight linked data publication from relational databases. In: Proceedings of the 18th International Conference on World Wide Web, pp. 621–630. ACM (2009)
2. Bhagavatula, Sekhar, C., Noraset, T., Downey, D.: TabEL: entity linking in web tables. In: International Semantic Web Conference, pp. 425–441. Springer, Cham (2015)
3. Braunschweig, K., Thiele, M., Eberius, J., Lehner, W.: Column-specific context extraction for web tables. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp. 1072–1077. ACM (2015)

[13]Demo: http://wsl.iiitb.ac.in/sandesh-web Video: https://www.youtube.com/watch?v=pt1j2k1M97o.

4. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving efficiency and accuracy in multilingual entity extraction. In: Proceedings of the 9th International Conference on Semantic Systems, pp. 121–124. ACM (2013)
5. Ermilov, I., Auer, S., Stadler, C.: CSV2RDF: user-driven CSV to RDF mass conversion framework. Proc. ISEM **13**, 4–6 (2013)
6. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM (JACM) **46**(5), 604–632 (1999)
7. Kripke, S.: Semantical Considerations of the Modal Logic (2007)
8. Limaye, G., Sarawagi, S., Chakrabarti, S.: Annotating and searching web tables using entities, types and relationships. Proc. VLDB Endow. **3**(1–2), 1338–1347 (2010)
9. Mulwad, V.V.: Tabel-a Domain Independent and Extensible Framework for Inferring the Semantics of Tables. University of Maryland, Baltimore (2015)
10. Mulwad, V., Finin, T., Joshi, A.: Semantic message passing for generating linked data from tables. In: International Semantic Web Conference, pp. 363–378. Springer, Berlin, Heidelberg (2013)
11. Ritze, D., Lehmberg, O., Bizer, C.: Matching HTML tables to DBpedia. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, p. 10. ACM (2015)
12. Srinivasa, S., Agrawal, S.V., Jog, C., Deshmukh, J.: Characterizing utilitarian aggregation of open knowledge. In: Proceedings of the 1st IKDD Conference on Data Sciences, pp. 1–11. ACM (2014)
13. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: a core of semantic knowledge. In: Proceedings of the 16th International Conference on World Wide Web, pp. 697–706. ACM (2007)
14. Unbehauen, J., Hellmann, S., Auer, S., Stadler, C.: Knowledge extraction from structured sources. In: SeCO Book, vol. 7538, pp. 34–52 (2012)
15. Zhang, Z.: Start small, build complete: effective and efficient semantic table interpretation using tableminer. Under Transp. Rev. Semant. Web J. (2014)
16. Zwicklbauer, S., Einsiedler, C., Granitzer, M., Seifert, C.: Towards disambiguating web tables. In: Proceedings of the 12th International Semantic Web Conference (Posters and Demonstrations Track), vol. 1035, pp. 205–208. CEUR-WS.org (2013)

Evaluating Mitigation and Adaptation Policies for Transport Sector Using GIS—Case Study of Bangalore

H. Vajjarapu, A. Verma and A. R. Hemanthini

1 Introduction

A swift growth in the economy has resulted in urbanization causing urban sprawl across the length and breadth of India increasing the travel distances. This has caused an exponential growth of private vehicle trips, which further led to congestion, reduction in travel speeds, increased travel times, vehicle kilometres travelled (VKT) and increased pollution. An effective way to solve these problems is to develop proper mitigation policies to reduce emissions from transport sector [5]. Mitigation policies concentrate on developing the existing transport infrastructures like bicycle lanes, heavy occupancy vehicle lanes, footpaths that can reduce the usage of private vehicles. Although transportation is one of the major causes for GHG emissions, it, being a critical sector that contributes to the smooth functioning of societies and fosters the economic growth of a nation, is under a constant threat of climate change [4, 5]. India is one of the world's most vulnerable countries to climate change [6]. Transport sector needs to cope up with climate change as such it is important to anticipate the impacts of climate change on the transportation system. Transport sector should also prepare for it on time owing to the key role of transportation system in ensuring economic stability and growth of a nation. Adaptive measures are seen as a tool to reduce the vulnerability to the potential negative impacts of climate change and strengthen the inherent capacity of a system to undertake defensive as well as protective actions

H. Vajjarapu (✉) · A. Verma (✉) · A. R. Hemanthini (✉)
Department of Civil Engineering, Indian Institute of Science, Bengaluru, India
e-mail: vajjarapu06@gmail.com

A. Verma
e-mail: ashishv@iisc.ac.in

A. R. Hemanthini
e-mail: hemumalu@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_18

that help to avoid loss and facilitate recovery from any impact by increasing the resilience of the entire system.

India has a vast road network of 5.23 million kilometres with 100475 km of national highways and expressways, 154522 km of state highways, 2577396 km of major district roads and 1433577 km of rural and other roads [2]. In case of any upgradation of the existing network as a part of a new policy, accessing information related to the roads at a broader level becomes difficult. Adopting new technologies like Geographic Information System in Transportation (GIS-T) helps to improve decision-making process by effectively managing the data across the country. In short, it is possible to state unequivocally that GIS-T has ‘arrived’ and now represents one of the most important application areas of GIS technology [3]. GIS helps the users in storing wide varieties of information related to road length, width, land use, population densities, topography, provision of drainage facilities, etc. Planners and engineers can more easily visualize data inputs (e.g. socio-economic variables) using the GIS colour graphic displays. The collections of data used to support a particular model scenario can also be more efficiently stored in a GIS than in a tabular environment [9].

Bangalore Metropolitan Region (BMR) has approximately 6000 km of road length for an area of 8005 km^2 [10]. BMR is intercepted with 2 national expressways, 3 national highways and 12 state highways connecting major towns and cities within BMR and beyond. The radial road network in the BMR converges into the core and contains centre–periphery traffic as well as the transit traffic, which chokes the city centre. In this paper, the authors discuss the usage of GIS in mitigation and adaptation policies designed for the Bangalore Metropolitan Region to reduce emissions from transport sector and adapt to climate change. This study is done for two scenarios, which are business as usual scenario (BAU) for the base year (2008) and policy scenarios for the target years 2030 and 2050.

2 Mitigation

2.1 Business as Usual Scenario

The study area is Bangalore Metropolitan Region, which is about 8005 km^2. The whole study area is divided into 384 traffic analysis zones for transport modelling. The road network map which is used to model the routes and estimate the volumes on each is overlaid on the zonal map. Information about trip productions, trip attractions and origin–destination matrix will go into the model as inputs, and the vehicle kilometres travelled and volumes on links will be the outputs. The population and employment variables for the base year (2008) have been considered to generate the trip end equations for productions and attractions, respectively, using a linear regression model. These trip end equations are used to forecast the future year productions

Fig. 1 Road network of BMR

and attractions using the future year's population and employment and are developed separately for private and public transport (Figs. 1 and 2).

In order to validate these trip end equations, the flows are assigned onto the GIS road network separately for private and public transport. Screen line analysis was performed at different locations, and it was found that the errors are within permissible limits. The model will be termed as calibrated once the traffic loadings on the GIS map are matching the 44 selected checkpoints (screen line locations) on the road network. It was found that the errors are within the acceptable range, which validates the trip generation model. An MNL model is developed to estimate the share of each mode in the trips. The OD matrices are loaded on the GIS road network of BMR region, and trips are assigned between 384 zones to obtain flows on the road network.

Using the trip end equations, the base year trips are forecasted for the years 2030 and 2050 that are then loaded on to the BMR network. Figures 3 and 4 correspond to the volumes of private and public transport on BMR road network for the years 2008, 2030 and 2050, respectively. These figures will aid the modeller in understanding the vehicle flows across the network and design appropriate policies to reduce congestion. The advantage of working on GIS map is the feasibility of changing the road network for the future years by adding approved upcoming projects and the projects in progress

Fig. 2 Zonal map of BMR

to the road network as additional links at the desired locations. It is observed from the above figures that in business as usual scenario, the volumes of vehicles on the roads are way over the capacity. The black lines in the figures show high volume/capacity (V/C), whereas green colour represents links with low V/C ratio. The total estimated vehicle kilometres travelled for 2008, 2030 and 2050 under the BAU scenario is 31 million, 47 million and 72 million, respectively.

2.2 Evaluating Mitigation Policy Bundles Using GIS

From the earlier sections, it is seen that the effective way to reduce emissions is to draw proper mitigation plans. Emissions are estimated from vehicle kilometres travelled using emission factors. Therefore, the main mitigation policy bundles are aimed to reduce the vehicle kilometres travelled from private vehicles (Table 1).

The above shown bundles are carefully formulated with a mixture of different policy instruments. Bundle 1 is a combination of planning and regulatory instruments, Bundle 2 is a combination of economic and regulatory instruments, Bundle 3 is a mixture of planning, regulatory and economic instruments, and Bundle 4 has an additional technological instrument along with planning, regulatory and economic

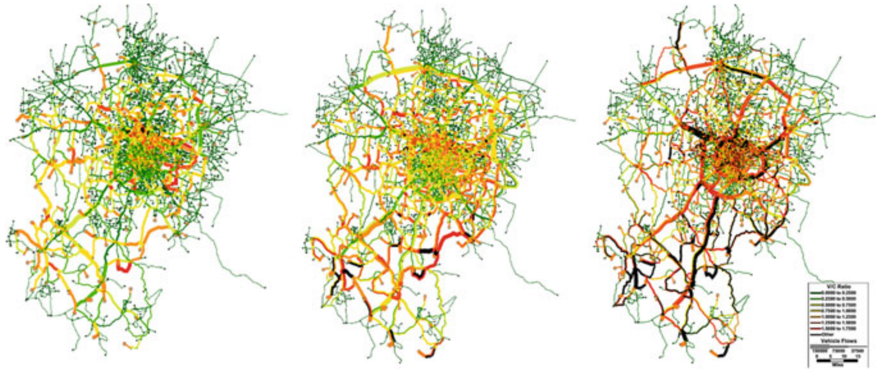


Fig. 3 Private transport vehicle flows on the road network for 2008, 2030, 2050 (left to right)

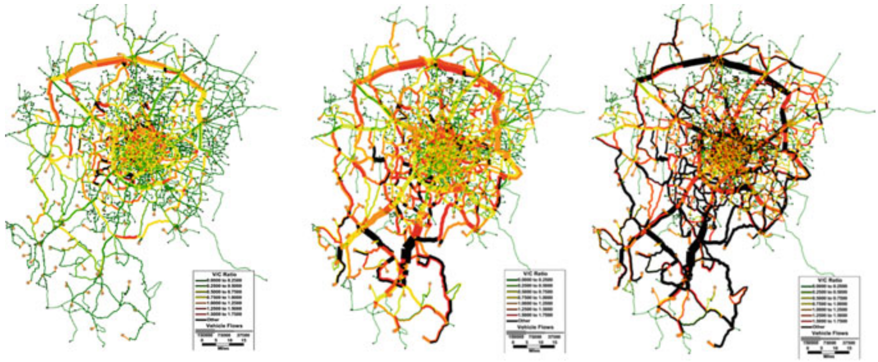


Fig. 4 Public transport vehicle flows on the road network for 2008, 2030, 2050 (left to right)

policy instruments. It is assumed that Bundle 4 has the same VKT and mode share as of Bundle 3, but the emissions vary because of electrification. Some of the above-mentioned policies are evaluated with the help of GIS map.

GIS has a major role in evaluating the policies like increase in public transportation road network, defining car-restricted zones, congestion pricing, park and ride, cycling and walking infrastructure.

Increase in public transportation road network:
The aim of this policy is to improve the connectivity of the public transportation system, to reduce the last mile connectivity and to reduce the walking time required to reach the bus stop. Since public transport follows a designated path unlike private transport, it cannot choose another route to reduce the congestion levels on a link. The potential locations of implementing this policy are identified using the GIS flow maps from Fig. 4. Traffic analysis zones that have more productions and more congested public transport links (black lines) are supported by adding more public transport routes close to them and reduce the burden on congested links. The newly added public transport road links are shown in Fig. 6. The addition of the new links

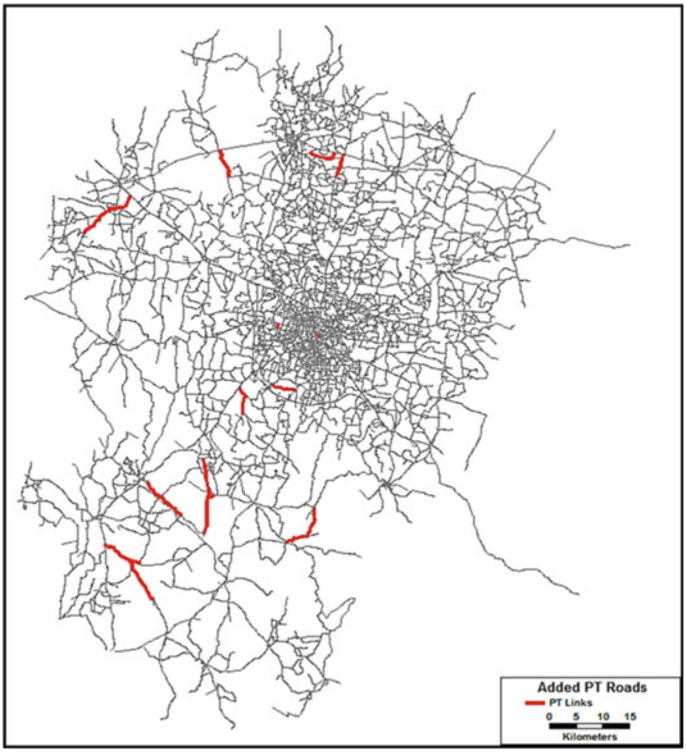


Fig. 5 Added new public transport road links

will cause a change in travel time, which are estimated and fed into the mode choice model to obtain mode shares. The new mode share values are used to estimate the new OD matrix which is then modelled and assigned onto the road network (Figs. 5 and 6).

Defining car-restricted roads:
The aim of this policy is to reduce the usage of car at certain locations, where the pedestrian movement is high. The advantage of using a GIS tool is that the links can be enabled or disabled depending on the model requirement. Heavily congested links on private transport road network are identified from the GIS flow map as shown in Fig. 3. This policy is modelled in four main steps.

Step 1: Identification of car-restricted roads and disabling those links from GIS road network in the model.
Step 2: Assigning all car trips on to the network without the car-restricted roads.
Step 3: Assigning all non-car trips with car trips as preloads and enabling the car-restricted roads for the model.

Table 1 Mitigation policy bundles

*Policies under **Bundle 1***
Increasing network coverage of public transit
Cycling and walking infrastructure
Additional tax on purchasing vehicles
*Policies under **Bundle 2***
Additional tax on purchasing vehicles
Strict vehicles inspection/improvement in standards for vehicle emission
Increase in fuel cost
*Policies under **Bundle 3***
Increasing network coverage of public transit
Defining car-restricted roads
Congestion pricing
Park and ride
Cycling and walking infrastructure
Encouraging carpooling and high occupancy lanes
High-density mix building use along the main transport corridors
*Policies under **Bundle 4***
All policies in Bundle 3 + All buses and cars running on electricity

Fig. 6 Car-restricted roads

Step 4: Assigning public transport trips with all private vehicle trips as preloads.

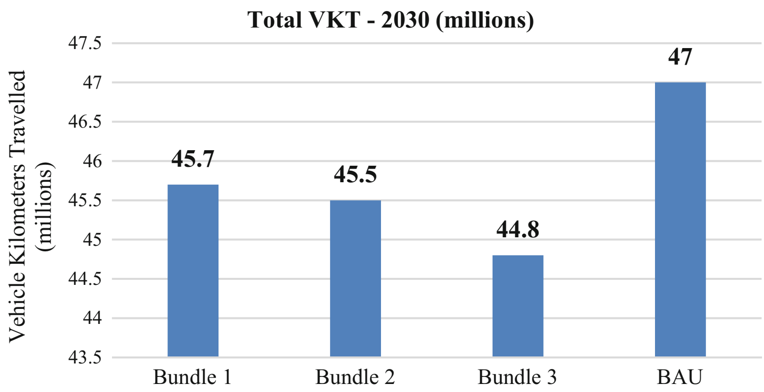


Fig. 7 Vehicle kilometres travelled—2030

This way the car trips are assigned on the private transport road network with no car entering the car-restricted roads, whereas every other mode can access those roads.

Cycling and walking infrastructure:
The aim of this policy is to encourage non-motorized transport (NMT) by providing proper cycling and walking infrastructure thereby improving accessibility to public transport. Bangalore is known to be the Silicon Valley of India, and this policy is targeted towards the IT professionals. A spatial analysis of the Bangalore city is done, and it was found that most of the IT industries are located around the outer ring road. This policy is tested within outer ring road limits to encourage NMT. To evaluate this policy, an acceptable trip distance of 0.75 km for walking and 1.66 km for cycling is considered [8]. It is assumed that the motor vehicle trips within acceptable trip distance of cycle and walk shift to NMT. New mode shares are calculated, new OD matrix is generated which is fed to the model, and the trips are assigned to the road network.

Similarly, other policies were developed based on proper assumptions and predicted values from proper sources. The impact of a single policy to mitigate emissions is less compared to a policy bundle implementation. For example, congestion pricing for car and two-wheelers when coupled with reduction in bus fare will yield better results since we are penalizing the private vehicle users and incentivizing the public transport users. Thus, a policy bundle is implemented rather than a single policy.

These mitigation bundles have been evaluated using GIS technology for the target years 2030 and 2050. It is observed that the mitigation policy bundles reduce the vehicle kilometres travelled when compared to the BAU scenario for 2030 and 2050 as shown in Figs. 7 and 8. Policy Bundle 3, which is a proper amalgamation of planning, regulatory and economic instruments, is showing a greater reduction in VKT, thus leading to emission reduction.

Bundle 4, which is an addition to Bundle 3 by electrifying all buses and cars, will give lesser emissions than Bundle 3 due to low emission factors. Figures 9 and 10

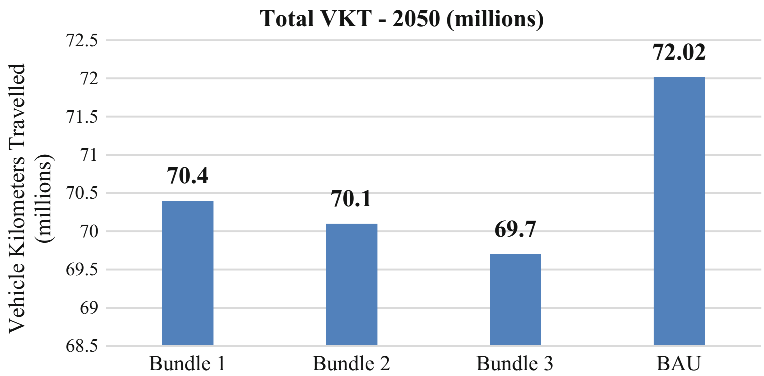


Fig. 8 Vehicle kilometres travelled—2050

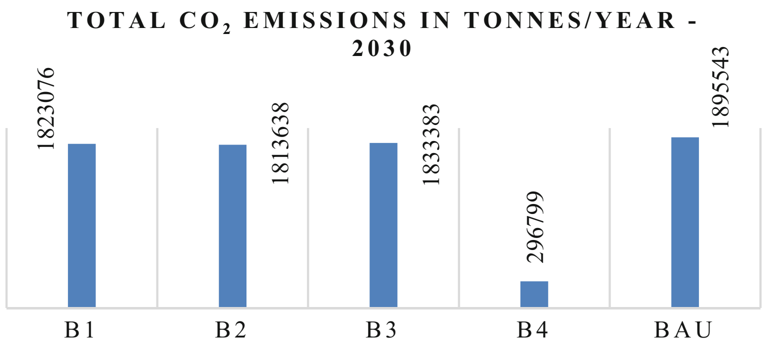


Fig. 9 Total CO_2 emissions—2030

clearly show the drastic reduction in emissions upon electrification of buses and cars.

Also, Bundle 3 shows higher CO_2 levels as shown in Fig. 9 compared to bundles 1 and 3 which is because of more shift towards public transportation. When the total emissions are converted to per capita emissions, the same Bundle 3 shows lesser emissions per person compared to other bundles as shown in Figs. 11 and 12.

3 Adaptation

3.1 BAU Scenario for Adaptation

Urban floods are the main focus for the adaptation part of this work. Hence, most of the policies have been formulated focusing on urban flood. Climate change is inevitable;

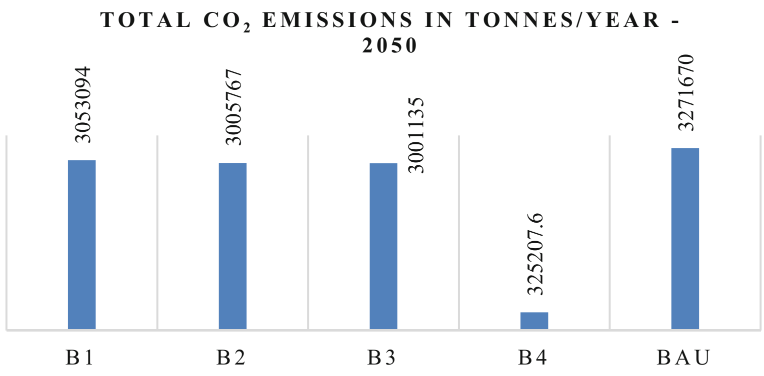


Fig. 10 Total CO_2 emissions 2050

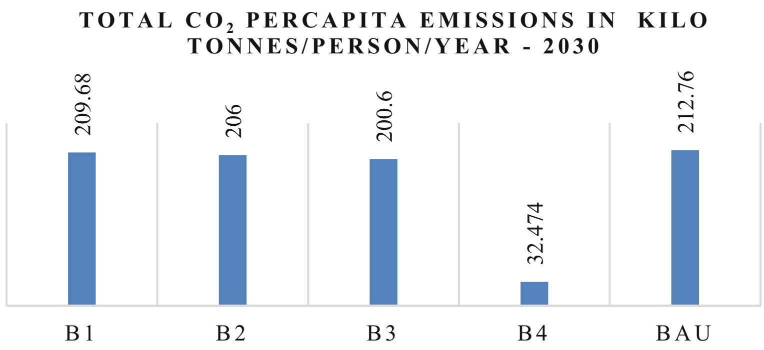


Fig. 11 Total CO_2 per capita emissions—2030

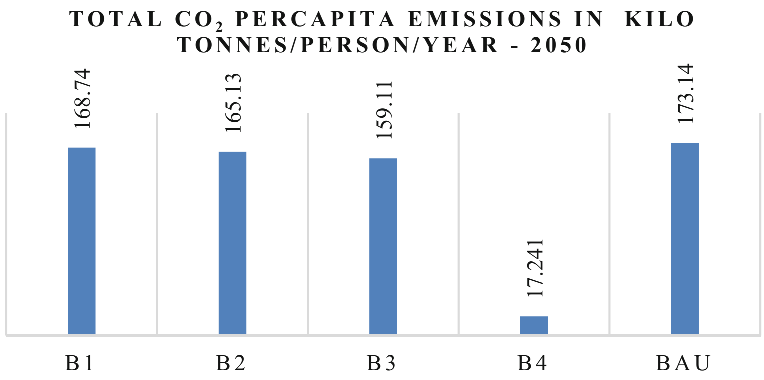


Fig. 12 Total CO_2 per capita emissions—2050

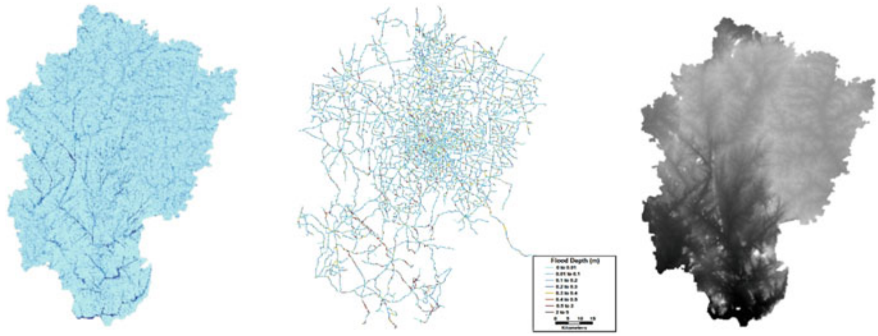


Fig. 13 From left to right in order—flood map of BMR, road network map with flood levels, digital elevation model (DEM) map

Table 2 Relation between flood depth and travel speed reduction

Flood depth (m)	Travel speed reduction
0–0.1	No change
0.1–0.2	0–25%
0.2–0.3	25–50%
0.3–0.5	50–75%
Above 0.5	75–99%
	Road closed

Source Pregnolato et al. [7]

however, adaptive strategies help in strengthening the road network system and act as resilient measures against urban floods.

The rainfall event considered for the study is of 3 November 2015 with a total rainfall of 266 mm over a duration of 4 h and 10 min and return period of 100 years. In Figure 13 from left, the first map shows the flood map of BMR, the second map is the road network with flood, and the third shows the elevation map.

For business as usual scenario, the base road network which was used for mitigation is considered. Flood map is overlaid over the road network to find out the vulnerable locations. Flooding reduces the vehicle speeds, and the percentage change in speed with flood depth is shown in Table 2.

For the BAU network, each road link is divided into multiple small links depending on the level of flood in that particular link. Links which carry a flood depth above than 0.5 m are not considered in the BAU model. Roads that are heavily flooded and the zones that do not have redundant roads for the trip to happen are considered to have no trips from those zones. There are about 7 such zones in which the trips are cancelled.

Bangalore has about 180–200 km length of primary and secondary storm water drainages, and nothing was added in the BAU model. BAU modelling is done for the base year, i.e. 2008, and for the future years 2030 and 2050 for both private and public transport networks. Since this is not a frequent event and not a usual scenario,

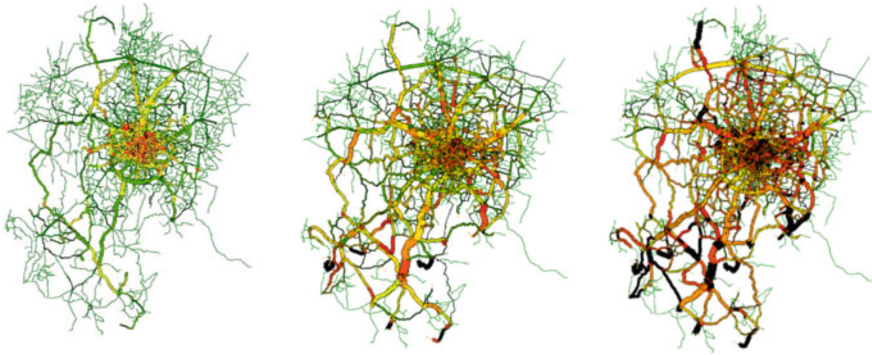


Fig. 14 Private transport vehicle flows under flood scenario for the years 2008, 2030, 2050 (from left to right)

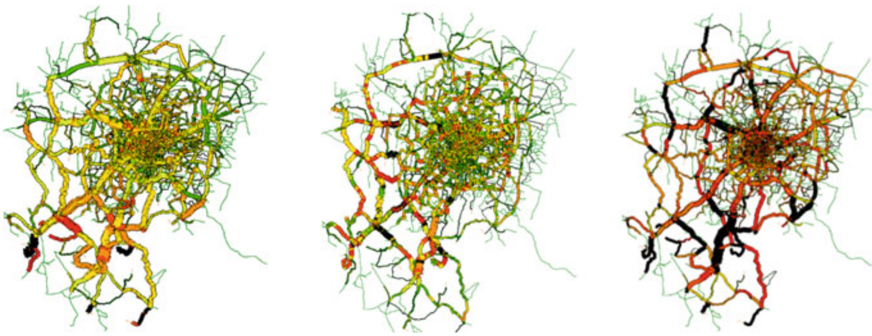


Fig. 15 Public transport vehicle flows under flood scenario for the years 2008, 2030, 2050 (from left to right)

the mode shift is kept the same even though trip lengths change. Trips are assigned to the road network, and the flow figures are shown in Figs. 14 and 15.

3.2 BAU Adaptation Results

Vehicle kilometres travelled is compared between no flooding and flooding scenarios for 2008, 2030 and 2050.

Trips are usually assigned through the shortest paths between origin and destination. From Figs. 16, it is observed that the vehicle kilometres travelled is increasing due to the fact that the shortest paths are flooded and the commuters are taking another path which is longer than the usual path. From Table 2, it is observed that as the level of flood increases, the vehicle speeds decrease. Using this relation, new

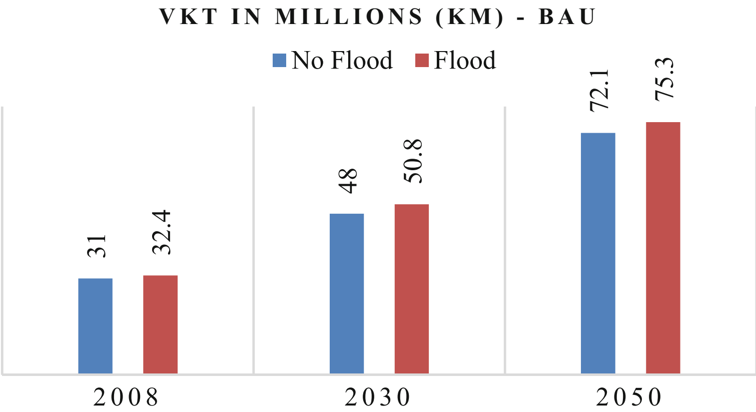


Fig. 16 Total VKT comparison for flood and non-flood scenario

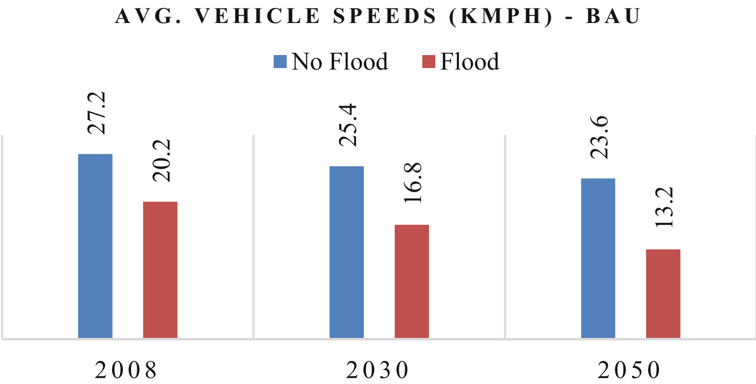


Fig. 17 Average vehicle speed comparison for flood and non-flood scenario

Table 3 Cancelled trips for BAU no flooding and BAU flooding case for BAU

Cancelled trips			
BAU—No flooding	BAU—Flooding	Total trips assigned	Percentage of trips cancelled
0	160786	6 million	2.8

vehicle speeds and travel times are estimated which are then fed into model for assigning trips to the road network (Fig. 17).

It is assumed that the zones which are flooded and have no access to other zones are considered to have cancelled their trips. Table 3 shows the total trips cancelled due to flooding for the BAU scenario.

Table 4 Policy bundles for adaptation

Adaptation policy bundles
Bundle 1
Replacement of impermeable road surface with permeable material in vulnerable areas
Slum relocation and rehabilitation
Providing proper drainage facilities at vulnerable areas
Construction of redundant infrastructure
Bundle 2
Rerouting people during flooding
Restricting development in low-lying or vulnerable areas
Slum relocation and rehabilitation
Bundle 3
Replacement of impermeable surfaces with permeable material in vulnerable areas
Providing proper drainage facilities at vulnerable areas
Rerouting people during flooding

3.3 Evaluating Adaptation Policy Bundles Using GIS

The main aim of the adaptation policies is to create a flood-resilient transportation infrastructure. Each policy bundle has a certain number of policies amalgamated together to bring out specific results after implementation for the years 2030 and 2050. The policy bundles for adaptation are listed in Table 4.

Replacement of impermeable road surface with permeable material in vulnerable areas:
Most of the Bangalore road networks have impermeable pavement systems. If there is improper drainage system, then even less intense rainfall accumulates certain depth of water on the surface of the pavement. This causes congestion on the road, resulting in higher vehicle hours travelled and higher vehicle kilometres travelled. This policy aims at replacing impermeable road surface by permeable material in the high flooding road links. A spatial analysis of the flooded road network is done to identify the vulnerable links. Flood map is overlaid on the road network, and the flood information on all links is extracted and added to the road network data. Links are considered vulnerable if the flood depth is between 0.3 and 0.5 m. Figure 18 shows the vulnerable locations selected for policy implementation. According to the literature, in comparison with conventional asphalts, permeable and porous pavements provide effective peak flow reductions up to 42% and longer discharging times [1]. Thus keeping other parameters constant, the flood depth reduces by 42%. The reduced flood depths will result in new vehicle speeds on the links, and accordingly, the trip patterns change.

Fig. 18 Locations of impermeable roads replaced with permeable material

Slum relocation and rehabilitation:
Slums are usually found in the low-lying areas because these areas are cheaper or no development exists there. When urban floods occur, their neighbourhood gets flooded with water that further blocks the roads and their mobility is hindered. Slums that are located below the elevation of Bangalore (920 m) are considered to be low-lying. Slums in the low-lying areas are identified by overlaying DEM map and flood map on the zonal map of BMR using ArcGIS. Slums that are common in both low-lying and flooded criteria are considered for the policy implementation. Since the slums will be relocated to a new place, this will affect the number of trips being produced and attracted to zones. This policy thus impacts the productions and attractions to each zone in the demand model. A sum of 34 slums were identified for incorporating in the model which are shown in Fig. 19.

Construction of redundant infrastructure:
Due to events like flooding, if a certain section of people are connected with only one single road and it gets flooded, then that particular section is cut off from their usual activities. In such situations, it is always good to have another road link that can connect them to a location where there is no flooding. Flood maps are overlaid on the road network to identify the links with flood level above 0.5 m and which have only one entry and one exit. It is to be noted that links with flood level above 0.5 m

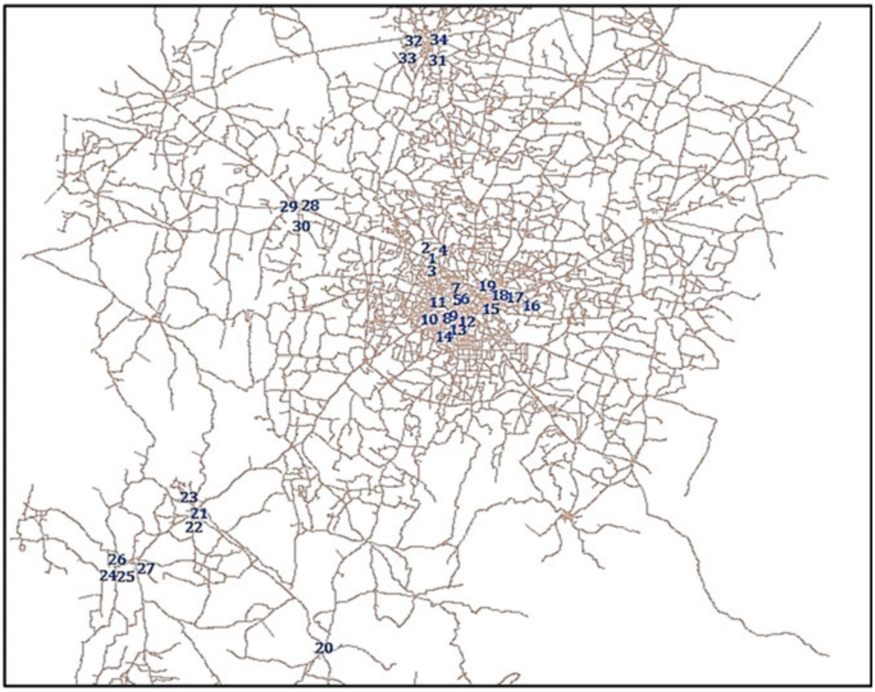


Fig. 19 Slums in low-lying areas (34 locations)

are not considered for modelling. The policy will be evaluated by adding extra links to the roads (which bypass the flooded links). These extra links which are added to public transport network are essentially picked from private transport road network. This helps in reducing the vehicle hours travelled during flooding and also reduces the cancellation of trips. This policy is tested at 10 locations across the Bangalore city, which is shown in Fig. 20.

Providing proper drainage facilities at vulnerable areas:
The storm water drainage system helps the water to flow from the tertiary pipelines to the trunk line. In case the drains are not properly provided, the water might stop flowing in the pipelines and thus lead to ineffectiveness of the drainage system to seep water off the roads. This policy helps in lowering the level of flood on the road surface which reduces the Vehicle Hours Traveled, cancelled trips, congestion, damages to road and also vehicles. A typical length and width of a storm water drainage is 1 m × 1.5 m for a metre length. The capacity of the drain would be 1.5 m^3. Assuming an upper cap of about 0.2 m, the holding capacity of the drainage is 1.3 m^3. If this drainage is provided (Fig. 21) on both sides of the road, then the total capacity of the drainage will be 2.6 m^3. So, for a road width of 3.5 m the reduction in depth of flood will be 74.3%, for 5.5 m road width the reduction in flood depth on road will be 47.2% and for a 7 m road width flood depth will reduce by 37.1%. Land use maps

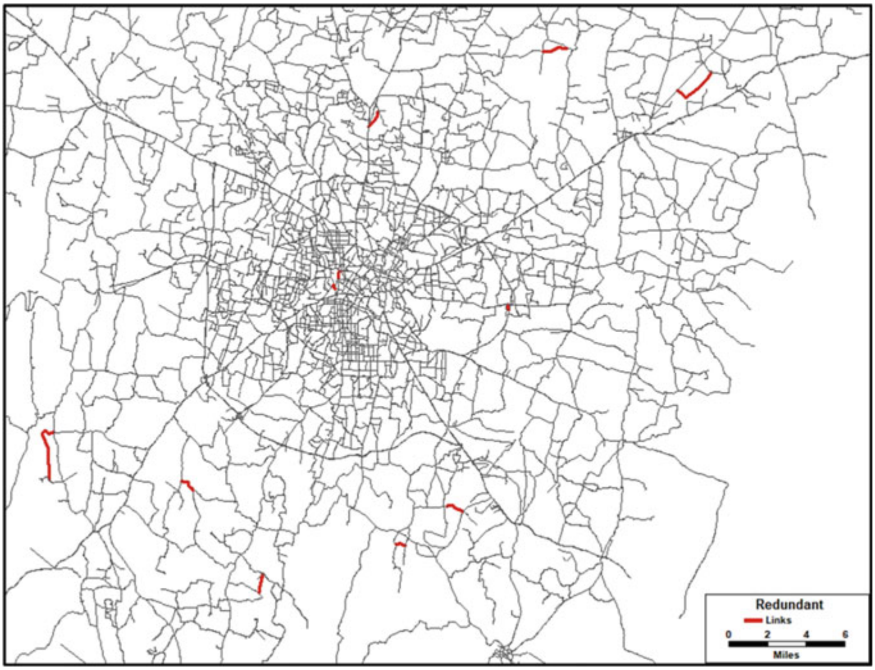


Fig. 20 Provision of redundant infrastructure (red lines)

of Bangalore are used to identify water bodies like lakes and tanks that can store the storm water. Road links that are flooded and close to a water body are provided with storm water drains and extended till the nearest water body as shown in Fig. 20. In the selected locations, these reductions in flood depth were considered for modelling.

Similarly, GIS has been extensively used in designing various adaptation policies. For example, spatial analysis using DEM map, flood map and zonal map is done to identify low-lying areas for the policy ‘restricting development in low-lying areas’. Low-lying areas that are flooded are identified by overlaying DEM maps, flood maps and zonal maps. Since development is restricted in those zones, it is assumed that there are no productions or attractions from those zones. A shortest distance analysis is done to shift these trips to the nearest zone that is not flooded and at a higher elevation. Also, rerouting in case of flooding is helpful for the commuters to choose the best path to reach the destinations, where GIS plays a major role in choosing alternate paths.

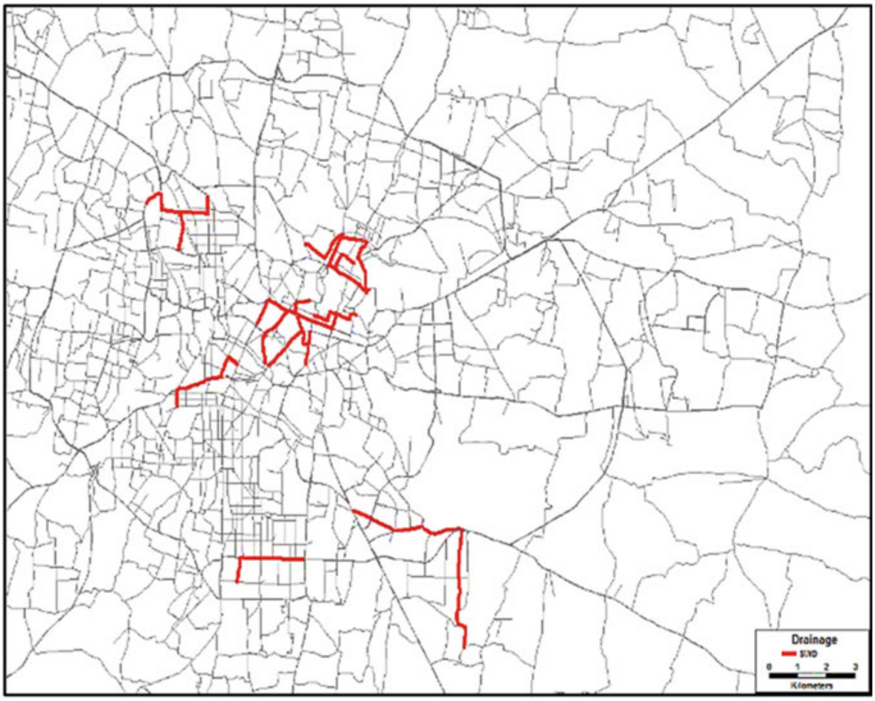


Fig. 21 Roads provided with drainage (red lines)

3.4 Adaptation Policy Bundles' Evaluation Results

The above-mentioned policies are combined as bundle as given in Table 4 and evaluated. From Figs. 22 and 23, it is seen that the vehicle kilometres travelled is increasing in BAU flooding scenario when compared to BAU no flooding scenario. The implementation of policy Bundle 1 (B1) reduces the VKT by a greater amount compared to Bundle 2 (B2) and Bundle 3 (B3) for the years 2030 and 2050.

As mentioned in the earlier sections, the aim of the adaptation policy bundles is to create a flood-resilient transport infrastructure. By implementing policy Bundle 1, which is a blend of land use and infrastructure policy instruments, there are no cancelled trips. A comparison of cancelled trips for the three adaptation policies is shown in Fig. 24. This means that even in a flood situation, with proper land use and infrastructure management, commuters can make their trips.

4 Conclusions

This paper contains the formulation of mitigation policies aiming to reduce VKTs and emissions, which have been evaluated using GIS technology for the years 2030

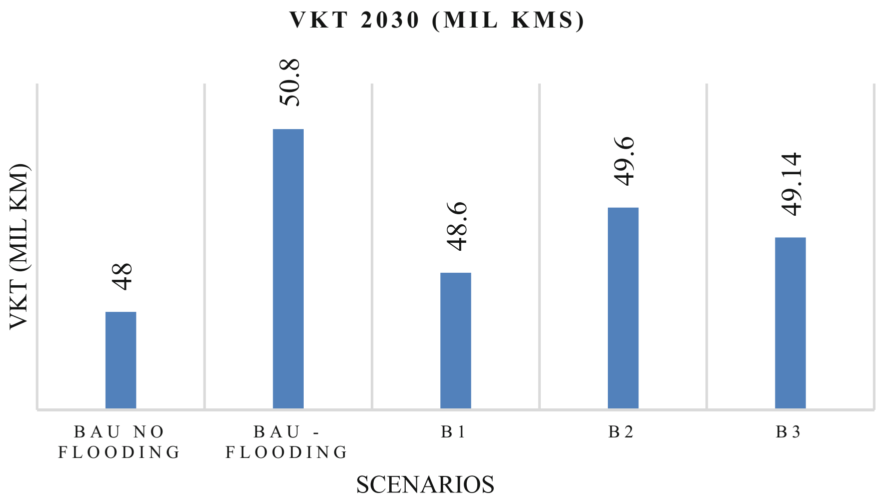


Fig. 22 Vehicle kilometres travelled—2030

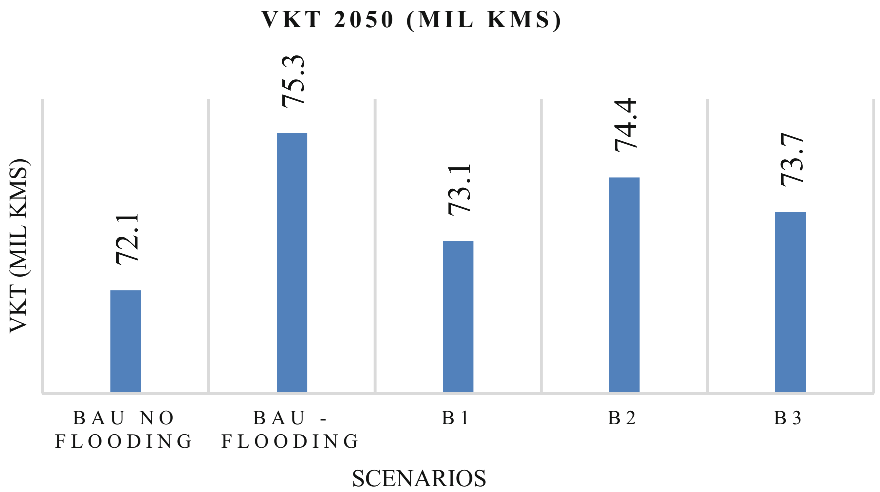


Fig. 23 Vehicle kilometres travelled—2050

and 2050. Four mitigation policy bundles have been formulated, and all the policies lead to a significant reduction in the total VKTs travelled when compared to the BAU scenario. GIS technology has played a major role in evaluating these policy bundles. Spatial analysis is done to identify the locations for infrastructure development, adding public transport road links to reduce congestion, etc. Bundle 3 which is a comprehensive mixture of 7 policies gives the best results with respect to VKT reduction, improved public transport share and reduction in CO_2 per capita emissions.

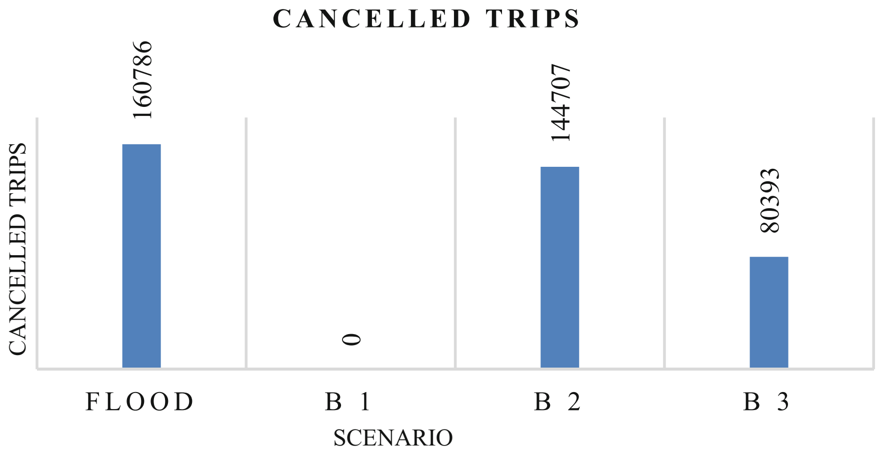


Fig. 24 Comparison of cancelled trips

The substantial reduction in emissions is observed with the implementation of Bundle 4 which includes the electrification of buses and cars. Electrification of buses and cars, if done on a priority basis, can reduce the emissions by almost 84% and 90% for 2030 and 2050, respectively. GIS is also used widely in drafting and evaluation of the adaptation policies to make a flood-resilient transport infrastructure. Spatial analysis using flood maps and DEM maps helped in identifying the vulnerable locations for flooding and designing location-specific policies. Three adaptation policy bundles are created and evaluated at various locations that are vulnerable. It is found that upon implementation, Bundle 1 is effective in reducing the vehicle hours travelled and increases vehicle speeds. Proper land use and infrastructure-related policies can nullify the cancelled trips due to flooding as seen in Bundle 1.

Acknowledgements This research work is supported through Indo-Norway research project 'CLIMATRANS: Coping with Climate: Assessing Transport Sector Strategies for Climate Change Adaptation and Mitigation in Indian Cities' funded by the Research Council of Norway.

References

1. Abbott, C.L., Comino-Mateos, L.: In-situ hydraulic performance of a permeable pavement sustainable urban drainage system. Water Environ. J. **17**(3), 187–190 (2003)
2. Annual Report of National Highways Authority of India (NHAI), Ministry of Road Transport and Highways (MoRTH), GOI, F.Y. 2015–16
3. Maguire, D.J., Smith, R., Jones, S.: GIS on the move: some transportation applications of GIS. In: Proceedings Thirteenth Annual ESRI User Conference, vol. 3, pp. 39–46 (1993)
4. NAPCC National Action Plan on Climate Change (2008). http://www.mma.gob.cl/1304/w3-article-49744.html

5. Parry, M., Canziani, O., Palutikof, J., Van der Linder, P., Hanson, C.E. (eds.): Climate Change 2007: Impacts, Adaptation and Vulnerability, Contribution of Working Group II to the 4th Assessment Report of IPCC. Cambridge University Press, Cambridge, UK and New York, USA (2007)
6. Paul, R.K., Birthal, P.S., Khokhar, A.: Structural breaks in mean temperature over agroclimatic Zones in India. Sci. World J. **2014** (2014)
7. Pregnolato, M., et al.: Assessing urban strategies for reducing the impacts of extreme weather on infrastructure networks. Royal Soc. Open Sci. **3**(5), 160023 (2016)
8. Rahul, T.M., Verma, A.: A study of acceptable trip distances using walking and cycling in Bangalore. J. Trans. Geogr. **38**, 106–113 (2014)
9. Souleyrette, R.R., et al.: Applications for interactive transportation planning GIS (TPGIS). In: Center for Transportation Research and Education Iowa State University. Final Report/User's Manual May (1996)
10. Wilbur Smith Associates: Comprehensive Traffic and Transportation Study for Bangalore Metropolitan Region (CTTS), June 2010

Alternatives for Reliable Trip Planning

S. Sen and A. Najmi

1 Introduction

With advances in location-sensing and wireless technologies, public transport systems are increasingly investing in obtaining bus location data to improve customer experiences. It is important to understand the informational value of such data in trip planning, especially in a multimodal scenario. To facilitate multimodal traveling in cities, trip-planning services are extremely important. For example, if a traveler decides to drive to a metro station from home and then take the metro, a journey planner should suggest a metro station that has an earlier service, instead of simply suggesting the nearest metro station.

Multimodal trip planning has been the focus of a significant body of work [1–3]. Interestingly, there are important differences in approaches to trip planning based on the nature of data used and also in respect to the availability of dynamic location data or historical data of arrival times. In typical systems where only routes and planned frequency of bus departure times are known, a frequency-based model simplifies the transit network by regarding each line as having many runs and the headway of runs as having a mean value. While the common bus lines approach [4] has been the basis of many algorithms in trip planning like Mumbai Navigator [5, 6], there are other approaches such as the schedule-based model [7] that has been adapted for options where the transit components reliably run on a pre-defined schedule and has been discussed and extended by [8]. The availability of AVL and near real-time location data of buses and other transit components has prompted incorporation of real-time time at different levels such as the one implemented in South Bay of San Francisco Bay area [9].

S. Sen (✉) · A. Najmi
Indian Institute of Technology Bombay, Mumbai, India
e-mail: sumitssen@gmail.com

A. Najmi
e-mail: alefiya.lightwala90@gmail.com

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_19

This paper attempts to review different alternatives available in trip planning from a computational perspective. In the first section of the paper, we introduce the need for reliable trip-planning infrastructure as a significant component of a public transit system. Thereafter, we describe computational problems in trip planning based on conventional approaches and recent standards and techniques. In the third section, we focus on comparing the different approach for the city of Mumbai and the different trip-planning alternatives that are available. Finally, we discuss the challenges that remain unaddressed and draw our conclusions.

1.1 *Reliability of Public Transit Systems*

Reliable public transit systems are considered important components of urban planning, and investment in transport infrastructure has increased focus as urban population in developing countries in South Asia have crossed 30% and expected to grow further [10]. In the context of transit systems, reliability is conventionally associated with the non-existence of "unplanned passenger waiting time at the stop." Sustainable urban growth has often been linked to "New mobility" as well as "Integrated intelligent transport management" [11]. Such intelligent transport entails the use of "feature-enhanced passenger information, real-time bus arrival displays" as well as "automatic vehicle location and driver instruction systems to reduce bus bunching and improve reliability." Despite such investments, unplanned waiting time at transit stops can easily go up with poor trip-planning tools available to the commuter. Improving the quality of trip planning in a transit system is therefore an area of significant interest for public transit providers. As reported by Diab et al. [12], transport agencies in North America are attempting to use higher accuracy of AVL in order to achieve closer connect between passenger expectations and actual performance, thus enabling transit agencies to achieve better service reliability that is positively perceived by the public.

Providing reliable travel time for better trust

Travel time reliability is an important indicator of public transportation service quality and strongly influences user's attitudes toward transit systems. Traffic conditions and multiple other factors introduce variability in arrival and dwell times of transit system components, which turn navigational planning into challenging problems. For example, in a metropolitan area, it is possible to navigate the city by bus, suburban trains, metro trains as well as shared rides, bike-shares in addition to walking and cab rides. While assuming a maximum of three transfers within a one-hour transit, any trip planner could potentially provide multiple trip plans. For the user, however, it is important that the plan she chooses is reliable. Any informational assistance in terms of the estimated time of arrivals (ETAs) at every transfer is critical to reduce the wait-time windows as shown in Fig. 1. In addition, if any of the transit options are too crowded to accommodate her, she is likely to have a longer travel time than she was

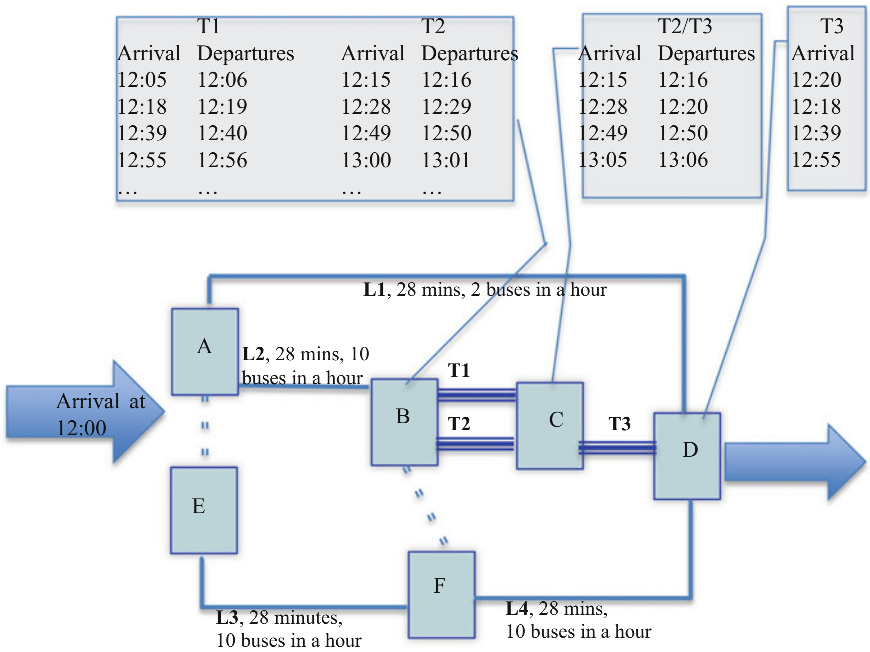


Fig. 1 Conceptual layout of the general trip-planning problem based on the Spiess–Florian network. This transit network represents different modes of transit, namely buses (L1, L2, L3, and L4), trains (T1, T2, and T3 with schedules as indicated above), and non-transit connections (between A–E and B–F) which can be assumed as walking distances. Based on an arrival at 12:00, a commuter has fixed time windows to make transit connections at every transit station. These time windows are altered by any delays or early arrivals and are more likely on the bus lines, which share their path with the general traffic

promised. Hence, travel time predictability is not the only measure for transit system reliability but the availability of space to board the bus also accounts as major reliability predictor. In the context of most South Asian urban localities, with the travel time and cost of personal transport (at least with two-wheelers) being comparable if not better with what public transit can offer, it is important to ensure better reliability and an overall positive outlook about the performance of the transit system.

1.2 Role of Trip-Planning Tools in Public Transit Systems

Trip-planning tools have become has become pervasive and for the urban traveler, with the availability of General Transit Feed Specification (GTFS) feeds and Open Trip Planner (OTP) [13]. Transit planning is available alongside navigation applications including Google Maps^{TM}, Bing Maps^{TM}, and HERE 360^{TM}. Many transit providers have designed and provided apps with trip-planning options within the area of operation. Such apps provide reliable trip plans based on published schedules as well as real-time alerts about delays and cancelations.

OneBusAway or OBA (http://onebusaway.org), a suite of transit traveler information tools developed at the University of Washington, provides real-time arrival information, a trip planner, a schedule and route browser, and a transit-friendly destination finder for Seattle area bus riders. Users of such applications have reported a set of important positive outcomes: strongly increased overall satisfaction with public transit, decreased waiting time, increased transit trips per week, increased feelings of safety, and even a health benefit in terms of increased distance walked when using the transit service [14]. This is consistent with an assumption that accurate real-time predictions can facilitate adaptive decision-making for the traveler. Thus, regular walking for certain distance may be a healthier alternative than waiting at the bus stop for the next connection under anxiety. It is important to note that information from OBA has not been integrated to trip-planning algorithms to generate plans but is provided as traveler information.

2 Trip Planning as a Computational Problem

Journey planners or trip planners are specialized search engines that help find optimal travel itineraries typically based on cost, time or transfer, and some given constraints. Key to the computation of trip plans is the available data of routes, the schedule or frequency of the transit components, and cost matrix (often as travel times).

2.1 Static and Real-Time Data

Most transit trip-planning systems are based on static schedules and generate trips that are not dynamically responding to delays in transit operation, caused by traffic congestion or accidents. Schedules are independent of the real-time traffic conditions on road and may not be followed (resulting in cancelations and delays). Such fixed schedule-based multimodal trip planning often assumes that every transit component (like a bus or a metro) will reach intermediate stops at a predetermined schedule based on historical travel time between stops. When maximum speeds are very low (such as in most congested cities), delays or early arrivals are averaged. Furthermore, in case of slow-moving times, the relative variance of travel time is understated, thereby ensuring a reasonable performance of static schedule-based trip planning.

Examples of static data used in GTFS feed (Fig. 2) include a connectivity graph of each route with distance between each stop and schedules. Static route data combined with schedule information can be built into a simple but effective trip-planning tool as without using GTFS as well [15]. In the case of a frequency-based data, such schedules can be computed by deriving appropriate headway between departures.

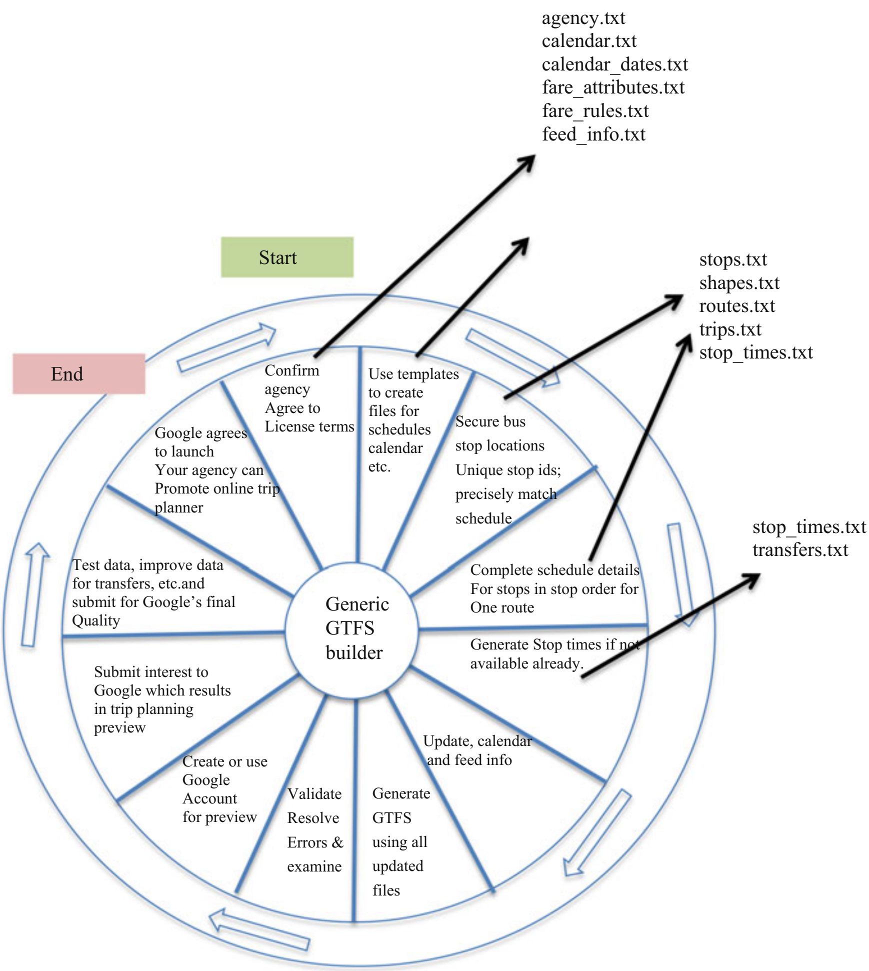


Fig. 2 Schematic representation of GTFS feed components (files) and steps to generate them based on a generic GTFS generation cycle. Many worksheet-based GTFS builders are available to create a feed

If the nodes of the graph are geo-referenced such as in GTFS, the routes could be displayed on a map. In a stage-based system (mostly followed by bus transit), multiple bus stops along a route are grouped into a 'stages' and the fare rises only after crossing over to the next stage in the route. Thus, in stage-based systems, it is important to include information about which stops are included in each stage.

Using dynamic or real-time data in Trip planning can be fairly complex. Real-time GTFS is only provided for trip updates that contain trip updates such as delays at particular stops, which need to be propagated to subsequent stops. Such updates alter the set of feasible trip plans that would be generated in usual scenarios. When a bus is late (the information is updated in real time), the set of pre-computed feasible paths is re-calculated. This update removes all paths that cannot be reached because of the delay and add a cost to all the paths that were affected by the delay of the bus.

2.2 Algorithms Used in for Multimodal Transit Planning

Algorithms for trip planning discussed earlier [1–8] have mainly focused on computer-generated trip plan. In addition to such algorithms, there are other techniques that focus on traveler information systems, which enable users to plan their own trips. The problem of finding shortest paths in transit networks is quite difficult because of waiting times and common bus line issue [4]. A traveler in a transit network has choice of traveling by multiple buses at particular node to reach destination. The problem becomes more complicated because of randomness in travel times and bus arrival rates. Because of these factors, researchers are motivated to develop adaptive routing algorithms for the transit.

The estimated time of arrival (ETA) such as those used in the times.txt file of the GTFS is usually expressed as the time of the day. In the absence of a fixed schedule, ETAs at every node of a transportation network are calculated and are important for trip planning. The calculation requires an estimation of the travel time (τ) as well as wait times (ω). Ni et al. [16] models expected travel time as a log-normal distribution using average travel times (μ) and standard deviation (σ) for every travel between i and k during interval j,

$$\tau_{i,k}^{j} \sim logN\left(\mu_{i,k}^{j}, \sigma_{i,k}^{j}\right)$$

Wait times are modeled using frequency (f) of bus arrivals during the time interval j at every stop i.

$$\omega_i^j = \frac{1}{f_i^j}$$

It is important to note that the waiting time is simply the combined frequency which is simply based on the often used assumption of periodic arrivals [4–14, 16–18]. The two components can be modeled as a network graph in a variety of ways, and different techniques of shortest path search can be applied [19]. When one is interested in the earliest possible arrival given a specified departure time (or, symmetrically, the latest departure), one can model this as the time-dependent shortest path problem, which assigns travel time functions to (some of) the edges, representing how long it takes to traverse them at each time of the day. By using both mean travel times and standard deviation, Ni et al. [16] have developed a stochastic time-dependent model. Interestingly, Zografos et al. [2] have reported that despite the computational complexity of trip planning in a time-dependent network, trip itineraries can be generated for most practical time windows with worst-case scenarios computed within 9s. Botea et al. [20] have shown that when the distributions of stochastic edge costs vary with time, minimizing the maximal travel time is NP-hard. They employed heuristics and state-dominance pruning rules to achieve solutions in feasible time although not fast enough for real-time use for a realistic urban scenario.

While most transit trip-planning systems are based on static schedules and generate trips that are not dynamically responding to delays in transit operation, caused by traffic congestion or accidents, there is a growing interest to use real-time traffic feeds and bus AVL data to generate accurate trip plans [9–14, 17].

2.3 Beyond Plans—Reporting ETAs and Occupancy

Commuters often check the reliability of their trip plans based on information about the next bus and occupancy levels. Any travel plan would be of no consequence if it is not possible for the traveler to board the bus or train which is often the case in over-stressed transit systems [21]. Estimation of occupancy levels (based on historical ridership time series data) adds additional level of complexities to the requirement of trip planning. In spite of a large availability of AVL data, occupancy prediction models are largely overlooked in the past. The transit assignment problem is analyzed by [22] where traveler boards the first bus available from the set of buses only if capacity is not exceeded. Wood et al. [23] present regression models to predict individual bus occupancies at future stops using the AVL data.

The enhancing estimation of travel time and wait times based on historical data and predictive analytics has been shown by Amirian et al. [24]. A range of prediction models for forecasting traffic states such as travel time and traffic flow including historical based models, time series, Kalman filtering technique, machine learning (using artificial neural networks), and support vector machines (SVM) have been developed [25]. The Mean Absolute Percentage Error (MAPE) for all such techniques was shown to be lowered to 13–10% by using hybrid approaches, although every method performs differently for different datasets. Trip-planning algorithms pessimistically prune out shorter trip plans based on the uncertainty of arrival time uncertainty of the first leg of the trip. A reduction in the uncertainty and hence an increase in the accuracy of the arrival time can therefore potentially reduce travel costs.

3 Comparing Alternatives for Trip Planning in Mumbai

As a running example, we use the case of trip planning in the Mumbai metropolitan area. The area is home to almost 20 million people and is served by multiple transit authorities such as BEST Undertaking (Bus services), suburban rail services (Western, Central, and Harbour lines), Mumbai metro and the Mumbai monorail. Taxis and auto-rickshaws are alternatives to such public transport in addition to private automobiles and two-wheelers. Although private transportation costs within

Mumbai are fairly low (INR 6 per km), parking constraints and slow traffic are major deterrents for use of private vehicles. Public transit agencies, on the other hand, have little or no coordination and in spite of a centralized urban planning, no systemic inter-coordination of services, ticketing, or trip-planning services are provided.

Transit planning in Mumbai is available through Mumbai Navigator [5], and GTFS-based online trip planners such as Google Transit, in addition to ETA-based Travel information apps such as M-Indicator^{TM1} and Transitpedia^{TM2}.

3.1 Transportation Networks in Mumbai

Mumbai has a fairly large transportation network as highlighted by the following points

- It has six major transit authorities operating across the metropolitan area.
- BEST Undertaking is the sole public bus authority although Navi Mumbai Municipal Transport (NMMT) buses do operate within the city. It operates about 3000 buses along more than 500 routes carrying almost 3 million passengers daily.
- The three suburban rail lines which carry over 7.5 million passengers daily across 141 train stations serves as the lifeline of the city and connects the north and south. Different types of services running on the same lines connect stations accounting for almost 30 different patterns.
- Recently started metro and monorail lines provide east–west linkages and have steadily picked up a ridership of 0.4 million.
- Over 0.3 million auto-rickshaws and 0.1 million taxis operate (sometimes as shared rides) in the metro region apart from 1.7 million two-wheelers and a million private automobiles.
- Walking on busy streets is common, and commuters are required to walk certain distances before catching their next ride. Walking speeds can vary based on location.

While some costs of transport remain heavily subsidized, they account for about 10% of the income of an average person living in Mumbai [26, 27]. It is evident that the requirement of multimodal transport in Mumbai is mainly for the regular commuter. Thus, the reliability of public transport trip plans in comparison to alternate (private) transport plans remains key to the choice of the more sustainable option.

3.2 Trip-Planning Alternatives

We analyzed trip-planning alternatives in Mumbai based on three parameters, viz., performance of the plan, inclusion of different modes, and ease of use. We did not choose to compare between applications that use the same data sources (such as GTFS feeds).

M-Indicator App^{TM}

The application provides local train and bus timetables based on location and time, and auto-rickshaw and taxi fares information to travelers. It regularly posts updates on delays, and cancelation intuitively provides chat forums related to traveling in Mumbai. Designed mainly for the suburban train traveler, it does not attempt trip planning by itself. It only provides the next scheduled transit details. However, in combination with real-time updates, it serves as the most reliable and usable tool for commuters in Mumbai. It covers schedules of suburban trains, buses, metro rail, monorail, and ferries.

Google Transit^{TM}

Based on unofficial static GTFS feeds of suburban trains and BEST buses, this application provides a familiar interface for trip planning on Mumbai's public transit. We evaluated the performance of generated trip plans and realized 100% variance in actual travel time (observed on 5 days of a week on the same 4 routes) with a mean-time variance of 62% of the total travel time. Since each route had three alternate plans, on most days the alternate plans performed better. It integrates train, bus, and metro routes in its plans.

Mumbai Navigator

Created 20 years earlier, this tool uses BEST and railway schedules to generate multimodal trip plans on the transit route network of Mumbai. The reliability of the trip plans evidenced from the previous users of the system was reported to be satisfactory [5]. The schedule and routes in Mumbai Navigator have been recently updated.

4 Discussion

It is evident that trip planning in Mumbai presents unique challenges in terms of both algorithmic complexity and scale. M-Indicator^{TM} with over 10 million users does not attempt to provide trip plans and thus makes it easier to use. GTFS-based trip planners provide plans that are not reliable but, in the absence of an alternative, remain the only option for trip planning. A comparison of the relative variance of actual travel time vs predicted travel time of the plans from a GTFS-based trip plan and an updated Mumbai Navigator would be useful.

Occupancy

It is important to note in the context of the case of Mumbai (like any other in congested cities) trip planning needs to account for occupancy as criteria failing, which

Parameters	M-Indicator^{TM}	Google Transit^{TM}	Mumbai Navigator
Trip-planning algorithms	No trip planning	Schedule-based	Frequency-based
Reliability of plans	Users make their own plans based on schedule information. Schedule information is accurate and real-time updates are provided	Provides the same plans during peak and non-peak traffic hours if the schedule remains the same, resulting in large variance in reported travel time	Also suffers from fixed travel time assumptions (which could be changed)
Inclusion of modes of transport	Includes bus (BEST, NMMT, TMT), suburban trains, metro, monorail, ferry, and private transport alternatives	Includes BEST bus, suburban trains, metro and private transport alternatives	Includes only BEST and suburban trains. Requires regular updates about schedules and routes
Ease of use	Can be used only on a mobile app but is easy to use and helpful when offline	Easy to use either as an app or online. Cannot be used offline unless cached	Can only be used online and not as an app

the commuter is forced to look for the *next transit*. In such scenarios, trip planners are reduced to schedule browsers. Also, it cannot be missed that trip plans in regions with multiple uncoordinated transit system components can easily have occupancy bottlenecks, which translate into failed travel plans and longer travel times. Although occupancy-based studies have been used in planning of transit systems, trip planners that incorporate occupancy levels (even at a coarse level such as *full* and *not-full*) are not available.

Data and algorithms at a city scale

With the prevalence of GTFS feed as a standard format for sharing transit schedules, it is important to bridge the data requirements of trip planners that use a frequency-based system as opposed to fixed schedules. Although schedule-based systems [7] do perform well where transit runs on schedule (such as trains and metros), frequency-based systems are able to model variances in waiting and travel times in a manner that is more appropriate to dense traffic and frequent delays. Recent attempts to model frequency- and schedule-based transit information jointly are promising but preliminary [28]. The performances of both methods are comparable if stochastic nature of travel or wait times is not considered. Benefits of using a *routing policy* give better results than travel plans when deterministic times are not available in a city-wide transit network [29]. Brczi et al. point out algorithmic issues and performance challenges in devising a trip planner for even moderately complex transit system and large number of commuters [29]. Time-sensitive trip planning for an urban center with multiple transit systems can be seen both as a scientific as well as an engineering problem.

First- and last-mile connectivity

In most public transit systems, the first–last-mile connectivity is responsible for the highest travel cost and which may have highest cost components possibly involving longer walking [30]. An important component here is the *access* and *egress* part of the transfer which, like in any transfer, can be a longer than normal walk times. Typical challenges in providing transit assistance, beyond normal trip planning, for multimodal transit have been discussed by [31]. Such information may be critical for any transit planning, and transit planning algorithms need to integrate multiple options for last-mile connectivity such as walking, rideshares as well as taxi services.

4.1 Value of Trip-Planning Information and Relevance of Accurate ETA

While we have discussed a perspective in transit planning that focuses on reporting the ETA more accurately and the need for integrating trip planners with such information, it is important that applications such as M-Indicator and OneBusAway are examples of location-aware tools that indicate time of arrival and are used extensively by commuters and improve public transit usability [32]. Firmani et al. argue that timetable-based routing may not be the best for route planning [33].

Cats and Gkioulou [34] have shown that the provision of real-time information could potentially reduce travel uncertainty; its impacts depend on the underlying service reliability, the performance of the prognosis scheme, and its perceived credibility. They argue that passenger expectations of transit system uncertainty, beyond the prior knowledge of headways and schedules, are formed from accumulated experience and real-time information. In the context of a trip planner, the real-time information of trip being planned is only available to a certain extent (the first transit connection and possible information about next connections at the time of planning). However, with the large-scale availability of AVL history data, it is also possible to analyze mean arrival time and variance over time and to model ETAs based on historical trends [35]. The possibility of trip planning that uses multiple strategies based on availability of data is promising and possibly the way forward for the next generation of trip-planning applications. Delling et al. [36] have shown that higher reliability of trip plans can be achieved by exploiting GPS data using novel algorithms.

Travel times using transit

It is important to note that in most congested urban clusters where the speed traffic on the roads is slow, the lure of reliable public transit is greater. Interestingly, with low speed of traffic, the variance in ETA reduces considerably and the value of real-time systems (at least in bus transit) becomes lower. However, this may not be a desirable condition for any city as it reduces the mobility in a city.

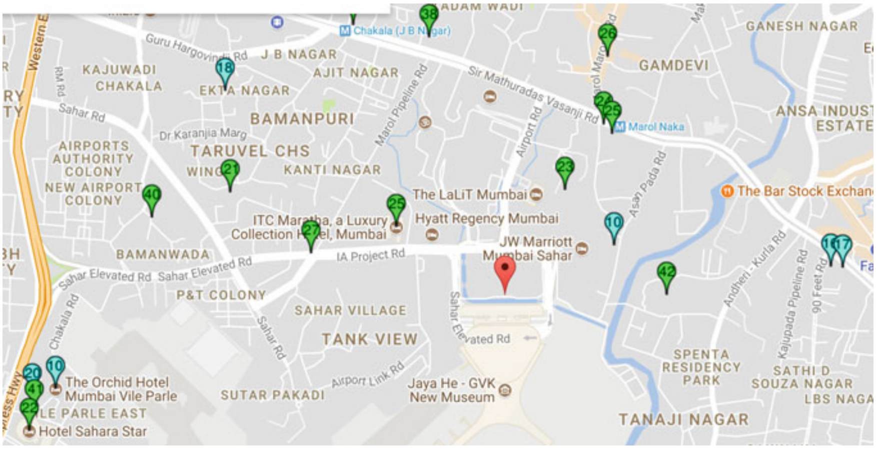


Fig. 3 Isochrones based on transit times around Mumbai airport using Google Transit APIs. The figure shows markers indicating travel times using public transit (buses) from the airport at a particular time. The red marker indicates the origin, blue markers are less than 20 min away, and green markers are more than 20 min away. It is important to note that unlike Isochrones by driving distances. Naturally, closest banks or restaurants for city commuters dependent on public transport can be found using such Isochrones which are irregular and time-sensitive

Travel time of cities using transit systems at different times of the day and GTFS enables a quick plot of Isochrones based on distances covered at equal intervals of time using public transits at a particular time [37]. Figure 3 shows Isochrone points (restaurants) around the Mumbai airport showing that some are accessible faster than others although they are nearer by network distances (at a particular time).

It is important to note these isochrone points are generated using GTFS schedules alone and do not consider combining available shared rides and longer walking options. However, the availability of such travel time maps in an urban landscape is very intuitive and can be used as an interface for trip planning.

5 Conclusion and Perspectives

We have reviewed multiple aspects of trip planning in public transit systems and have discussed different approaches used. It is evident that GTFS has steadily evolved as a widely used standard, and OTP-based trip planning has become central to transit trip planning in most cities. However, even with real-time GTFS feeds, variance of travel and wait times is not modeled in such systems. ETA-based applications that use real-time location of transit units are often more accurate and helpful for commuters.

The availability of reliable trip plans is directly related to the positive outlook about the transit system, and the reliability of such plans is directly related to the accuracy and sophistication of data and algorithms used to generate them.

Planning with stochastic travel and wait times with coarse occupancy levels

It can be seen from the comparison of the alternatives and discussion that stochastic travel and wait times are practical requirements for trip planning that aims to achieve higher reliability in a dense urban landscape. Integration of both schedule-based (such as those using GTFS feeds) and frequency-based approaches (such as those using frequencies and headways) is being increasingly seen to be beneficial. Personalized trip planning that accounts for the travelers' level of optimism in making transfers and occupancy on the transit systems is also necessary for a commuter in cities in most parts of South Asia. Services that find shortest as well as reliable trip plans on a fairly large time-dependent transit network with such additional constraints present an algorithmic challenge. Our ability to incrementally resolve such challenges can alleviate the problems that public transit systems face in highly populated cities.

References

1. Delling, D., Sanders, P., Schultes, D., Wagner, D.: Engineering route planning algorithms. In: Algorithmics of Large and Complex Networks, pp. 117–139 (2009)
2. Zografos, K.G., Androutsopoulos, K.N.: Algorithms for itinerary planning in multimodal transportation networks. IEEE Trans. Intell. Transp. Syst. **9**(1), 175–184 (2008)
3. Bast, H., Delling, D., Goldberg, A., Mller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., Werneck, R.F.: Route Planning in Transportation Networks, pp. 19–80. Springer (2016)
4. Chriqui, C., Robillard, P.: Common bus lines. Transp. Sci. **9**(2), 115–121 (1975)
5. Ranade, A., Srikrishna, M., Tilak, K., Datar, M.: Mumbai naviagator. Indian J. Transp. Manag. Jan–Mar 2005
6. Zhang, Y., Tang, J., Chen, L.: Itinerary planning with deadline in headway-based bus networks minimizing lateness level. In: 2014 11th World Congress on Intelligent Control and Automation (WCICA), pp. 443–448. IEEE (2014)
7. Tong, C.O., Richardson, A.J.: A computer model for finding the timedependent minimum path in a transit system with fixed schedules. J. Adv. Transp. **18**(2), 145–161 (1984)
8. Xie, J., Wong, S., Lo, S.: Three extensions of tong and richardsons algorithm for finding the optimal path in schedule-based railway networks. J. Adv. Transp. **2017** (2017)
9. Li, J.-Q., Zhou, K., Zhang, L., Zhang, W.-B.: A multimodal trip planning system incorporating the park-and-ride mode and real-time traffic/transit information. In: Proceedings ITS World Congress, vol. 25, pp. 65–76
10. Sarker, S., Khan, A., Mannan, M.M.: Urban population and economic growth: South Asia perspective 2016, vol. 5, no. 1, p. 12 (2016)
11. Goldman, T., Gorham, R.: Sustainable urban transport: four innovative directions. Technol. Soc. **28**(1), 261–273 (2006)
12. Diab, E.I., Badami, M.G., El-Geneidy, A.M.: Bus transit service reliability and improvement strategies: integrating the perspectives of passengers and transit agencies in North America. Transp. Rev. **35**(3), 292–328 (2015)
13. McHugh, B.: Pioneering open data standards: the gtfs story. In: Beyond transparency: open data and the future of civic innovation, pp. 125–135 (2013)
14. Ferris, B., Watkins, K., Borning, A.: Onebusaway: results from providing real-time arrival information for public transit. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1807–1816. ACM
15. Ferris, B., Watkins, K., Borning, A.: Data structures to represent public transport systems for small towns case study of Dharwad (2017)

16. Ni, P., Vo, H.T., Dahlmeier, D., Cai, W., Ivanchev, J., Aydt, H.: Depart: dynamic route planning in stochastic time-dependent public transit networks. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), pp. 1672–1677. IEEE (2015)
17. Borole, N., Rout, D., Goel, N., Vedagiri, P., Mathew, T.V.: Multimodal public transit trip planner with real-time transit data. Procedia-Soc. Behav. Sci. **104**, 775–784 (2013)
18. Datar, M., Ranade, A.: Commuting with delay prone buses. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, pp. 22–29
19. Lpez, D., Lozano, A.: Techniques in multimodal shortest path in public transport systems. Transp. Res. Procedia **3**, 886–894 (2014)
20. Botea, A., Nikolova, E., Berlingerio, M.: Multi-modal journey planning in the presence of uncertainty. In: ICAPS
21. Liu, Y., Bunker, J., Ferreira, L.: Transit users routechoice modelling in transit assignment: a review. Transp. Rev. **30**(6), 753–769 (2010)
22. Cea, J., Fernndez, E.: Transit assignment for congested public transport systems: an equilibrium model. 27, 133–147 (1993)
23. Wood, J., Yu, Z., Gayah, V.V.: Regression models for real-time bus occupancy prediction. In: Transportation Research Board 95th Annual Meeting (2016)
24. Amirian, P., Basiri, A., Morley, J.: Predictive analytics for enhancing travel time estimation in navigation apps of Apple, Google, and Microsoft. In: Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science, pp. 31–36. ACM
25. Khetarpaul, S., Gupta, S., Malhotra, S., Subramaniam, L.V.: Bus arrival time prediction using a modified amalgamation of fuzzy clustering and neural network on spatio-temporal data. In: Australasian Database Conference, pp. 142–154. Springer
26. Cropper, M., Bhattacharya, S.: Public transport subsidies and affordability in Mumbai, India. Urban Stud. Res. **2012** (2012)
27. Shirgaokar, M.: Employment centers and travel behavior: exploring the work commute of mumbais rapidly motorizing middle class. J. Transp. Geogr. **41**, 249–258 (2014)
28. Eltved, M., Nielsen, O.A., Rasmussen, T.K., Frederiksen, R.: Joint modeling of schedule-and frequency-based services in public transport assignment models. In: Trafikdage 2017 (2017)
29. Brczi, K., Jttner, A., Laumanns, M., Szab, J.: Stochastic route planning in public transport. Transp. Res. Procedia **27**, 1080–1087 (2017)
30. Clifton, K., Muhs, C.: Capturing and representing multimodal trips in travel surveys: review of the practice. Transp. Res. Rec. J. Transp. Res. Board **2285**, 74–83 (2012)
31. Rehrl, K., Bruntsch, S., Mentz, H.-J.: Assisting multimodal travelers: design and prototypical implementation of a personal travel companion. IEEE Trans. Intell. Transp. Syst. **8**(1), 31–42 (2007)
32. Ferris, B., Watkins, K., Borning, A.: Location-aware tools for improving public transit usability. IEEE Pervasive Comput. **9**(1), 13–19 (2010)
33. Firmani, D., Italiano, G.F., Laura, L., Santaroni, F.: Is timetabling routing always reliable for public transportl. In: ATMOS-13th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems-2013, vol. 33, pp. 15–26. Schloss DagstuhlLeibniz-Zentrum fuer Informatik (2013)
34. Cats, O., Gkioulou, Z.: Modelling the impacts of public transport reliability and travel information on passengers waiting time uncertainty. EURO J. Transp. Logist. (2014)
35. Moreira-Matias, L., Mendes-Moreira, J., de Sousa, J.F., Gama, J.: Improving mass transit operations by using avl-based systems: a survey. IEEE Trans. Intell. Transp. Syst. **16**(4), 1636–1653 (2015)
36. Delling, D., Italiano, G.F., Pajor, T., Santaroni, F.: Better transit routing by exploiting vehicle gps data. In: Proceedings of the 7th ACM SIGSPATIAL International Workshop on Computational Transportation Science, pp. 31–40. ACM (2014)
37. Sen, S.: Time sensitive isochrones maps using gtfs. In: Proceedings of the GIS in Transit Conference, Washington, DC, URISA (2015)

Mining Swarms from Moving Object Data Streams

A. Bhushan and U. Bellur

1 Introduction

Recent advancements in positioning technology, sensor networks, and online social media enables to capture the spatio-temporal data streams of various moving objects such as animals, product items, people and vehicles as sensor tags, RFID records, location-based services, and GPS logs. These spatio-temporal data streams contain the positions of each moving object captured at discrete point of time. The main characteristics of these streams are: unbounded length and probably fast arrival rate [1]. Many emerging applications require processing of these high volume streams in real time. Examples of such applications include traffic management, surveillance system, and recommender system [2–5].

A mining problem arising from the domain of evacuation planning is to detect groups of people moving together during evacuation. It is likely that the people in such groups know each other and want to stay together during the evacuation. Since, in urban emergency response systems, following the schedules given with the evacuation paths are very important to avoid chaos and congestion [6, 7], early detection of moving groups and using this information to compute the evacuation schedules, and monitoring the evacuation schedules during evacuation can improve the efficacy of the evacuation. Another application of mining moving groups is mobile phone sensing where different type of data is being sensed by mobile devices such as velocity and CO_2 emission [8]. The limited storage available on mobile devices demands that this data be uploaded to a server frequently. Recognizing a moving group among this set will aid in improving the effectiveness (power consumption

A. Bhushan (✉) · U. Bellur
GISE Lab, Department of Computer Science and Engineering,
Indian Institute of Technology Bombay, Mumbai, India
e-mail: alka.bhushan@yahoo.com; abhushan@cse.iitb.ac.in

U. Bellur
e-mail: umesh@cse.iitb.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_20

and data congestion) of scheduling updates from members of this set. Since we would only need data from one member of a closely knit moving group (given the similarity of sensor readings in a small geographic area), this will allow scheduling of more number of uploads on a server. In traffic management, early detection of trucks moving together can help in avoiding traffic jams by diverting their routes or controlling the traffic signals.

As many of the applications shown above illustrate, groups identified from the most recent data about their geographic position are often more useful than groups detected out of all the available data. Storing the groups of past data will unnecessarily increase the memory requirement and computation time required for storing all the groups and computing new groups, respectively. Moreover, occurrences of these groups will be counted from the beginning which may result of reporting many anomalous groups. A solution to remove old data is to use the sliding window technique which include the data of recent last L timestamps [9, 10] and remove the old data. The sliding window technique is popular to process streaming data and has been used widely in querying moving object data streams [3, 11].

While there are many different type of patterns defined in the literature which have been used in detecting moving groups such as flock [4, 12], convoy [13], and swarm [14, 15], the applications that we have presented earlier are best served by the generalized swarm pattern where a group of objects of size greater than a specified threshold is defined as swarm if all the objects of the group have been together for a sufficient amount of time. However, the existing methods for computing swarms either require all the data of a window to be present in advance where computational time increases exponentially with the number of objects, or do not prune groups that exist outside of the window [14, 16]. Thus, existing methods of computing swarms are not suitable for sliding windows.

To overcome these drawbacks of existing methods, in this paper, we make the following contribution: (i) we propose an efficient *incremental* method for computing swarms over sliding windows for data streams where the groups computed from the old data are discarded and new groups are added as soon as any new data arrives, (ii) we present results from two real datasets to compare the performance of our method with the previous methods. The results show that our method performs efficiently over sliding windows. In particular, our method is 5–13 times faster than the previous incremental method [16] with 2–6 times memory overhead over sliding windows. A poster of this work has been published in [17] where results are shown for one dataset only.

2 Related Work

Flocks [12], convoys [13], closed swarm [14], and gathering Patterns [18] are the most popular moving patterns that have been defined in the literature. These patterns differ from each other with respect to spatial proximity and temporal behavior.

In particular, flock is defined as a set of objects that stay together within a specified disk for at least k consecutive timestamps whereas convoy allows objects to stay together in arbitrary shape clusters. Closed swarm defined in [14] relaxes the constraint of groups to stay together at consecutive timestamps. Further, gathering pattern defined in [18] captures large congregations of objects which form large and stable areas of high density. Among these patterns, the time complexity of computing closed swarm is the largest for static datasets. Thus, distributed algorithms have been proposed for computing swarms for large datasets [19, 20]. However, the method presented in [19] does not compute closed swarms. Incremental methods for computing these patterns from the streaming data have also been proposed in the literature [4, 16, 21]. These methods use the data collected from the beginning.

Further, two other moving groups have been defined in the literature for streaming data [2, 5]. In particular, the method presented in [2] reports the top-k group patterns instead of reporting all moving patterns, and the method presented in [5] computes the travelling companions which are group of objects who are moving together and can disperse for a limited time period. The travelling companion method is used to monitor the groups which appeared at the beginning.

The method given in [14] cannot be used for computing swarm over sliding windows because it requires all the data to be present in advance, and its runtime increases exponentially with the number of objects. The experimental results in [19] supports the argument. Another method presented in [16] adds the data incrementally. However, it does not remove the old data.

Most of these methods compute clusters from each snapshot arriving at discrete time interval. Type of clusters depends upon an application [14]. For example, density-based methods to cluster cars moving on highways and k-means method to cluster birds in 3-D space [14].

3 Problem Definition

3.1 Definitions

Spatio-temporal Data Stream Let $S_0, S_1, \ldots, S_{t-1}, S_t$ be a series of snapshots, where S_0 and S_t are the first and current snapshot, respectively, and snapshot S_i contains the position of each moving object at timestamp t_i. We assume that position of each object is reported at the same time. Similar to [10], let $S_{t-L+1}, S_{t-L+2}, \ldots, S_t$ be the set of snapshots in the current window $[t - L + 1, t]$ where L is the length of the window and the window slides by 1 snapshot. A snapshot is called old if it is not in current time window and timestamp of the snapshot is smaller than the first timestamp of the current time window.

Closed Swarm [14] A group g is a pair $\{c, T_c\}\}$, where c contains the moving objects who appeared in same cluster at every timestamp in T_c, and T_c is a subset of timestamps of the current window $[t - L + 1, t]$. The group g is called a swarm for a given size threshold $sizeTh$, and time threshold $timeTh$ if it satisfies the followings conditions:

- $|c| \geq sizeTh$. Number of objects in c is at least equal to the size threshold.
- $|T_c| \geq timeTh$. Number of timestamps in T_c is at least equal to the time threshold.
- All the objects in c appear in the same cluster at each timestamp $t_i \in T_c$.

A swarm $g_i = \{c_i, T_{c_i}\}$ is called a *closed* swarm, if there is no other group $g_j = \{c_j, T_{c_j}\}$ such that either $c_i \subset c_j$ and $T_{c_j} = T_{c_i}$, or $c_i = c_j$ and $T_{c_i} \subset T_{c_j}$.

Swarm Candidate Since snapshots arrive continuously, a group $g_i = \{c_i, T_{c_i}\}$ such that $|c_i| \geq sizeTh$ and $|T_{c_i}| < timeTh$ can become a swarm later. We denote a group $g_i = \{c_i, T_{c_i}\}$ as a *swarm candidate*, if $|c_i| \geq sizeTh$ and $|T_{c_i}| \geq 1$.

3.2 Problem Statement

Let $S_0, S_1, \ldots, S_{t-1}, S_t$ be the series of snapshots containing the positions of each moving object at every timestamp and t be the current timestamp. The problem is to compute the closed swarms from the recent L snapshots, i.e., $S_{t-L+1}, S_{t-L+2}, \ldots, S_t$.

4 Our Method

Our method for computing swarms comprises of the following five steps: (i) compute clusters from the recent snapshot, (ii) compute new swarm candidates from these clusters and previous swarm candidates, and store them, (iii) update the timestamp list of relevant swarm candidates,(iv) report the closed swarms, and (v) remove all the swarm candidates computed from old snapshots.

First, we compute density-based clusters from the resent snapshot S_t using the grid-based DBSCAN algorithm given in [22] due to its simplicity and better performance than other algorithms.

Next, new swarm candidates are computed from the clusters extracted from the snapshot S_t and the swarm candidates computed from previous snapshots $S_{t-L+1}, S_{t-L+2}, \ldots, S_{t-1}$. All the swarm candidates computed from previous snapshots are stored in a graph where each node represents a swarm candidate and two nodes i and j are connected by a directed edge (i, j), if $c_j \subset c_i$, and there does not exist any other node k such that $c_k \subset c_i$ and $c_j \subset c_k$.

4.1 Computing Swarm Candidates

A cluster is added as a swarm candidate if the size of its object set is greater than or equal to the size threshold. If the cluster is already stored as a swarm candidate, then its timestamp list is updated by adding the timestamp t. Otherwise, a new swarm candidate is created whose timestamp list contains timestamp t. There may exist a swarm candidate in the stored candidates whose object set contains the objects of the cluster. In such case, timestamps from the timestamp list of swarm candidate is added in the timestamp list of new swarm candidate. There may also be a case where some objects are common between object set of the cluster and the object set of a swarm candidate, and number of these common objects is greater than or equal to the size threshold. In such case, if there is no swarm candidate in the stored candidates whose object set is same, a new swarm candidate is created for these common objects whose timestamp list contains the timestamps of the swarm candidate and the timestamp of the cluster. Otherwise, timestamp t is added in the timestamp list of the corresponding swarm candidate. Since computing swarm candidates from a cluster will require multiple scans of stored candidates without using any data structure, we use graph for storing the swarm candidates to reduce the search space.

4.1.1 Timestamp List

Over the period of time, each node may build up a long list of timestamps. We can reduce the space by storing the first and last timestamp of each interval, where each interval is a set of consecutive timestamps.

4.2 Computing Swarms

If size of timestamp list of a current visited node is greater than or equal to $timeTh$, then the corresponding swarm candidate is reported as a swarm.

4.3 Removing Old Data

When window slides by 1 snapshot, first timestamp of previous window is removed from the timestamp list of each node in the graph. In such case, timestamp list of some nodes becomes empty and some nodes may have timestamp list equal to the timestamp list of their parents. We detect these non-closed swarm candidates by finding a node whose timestamp list is equal to its parents timestamp list. Thus, a single scan of the nodes and edges is sufficient to find all non-closed swarm candidates.

5 Illustrative Example

We consider the example of eight moving objects to describe our method. Their movements in four snapshots are shown in Fig. 1. Three clusters are computed from each snapshot and are shown either by a circle or by an ellipse. The window size is three, and thus, there are two windows [1, 3] and [2, 4].

The corresponding graph computed at each snapshot is given in Fig. 2. For window [1, 17], at snapshot 1, three clusters $c1$, $c2$, $c3$ are computed. Each cluster is initialized to a candidate, and isolated nodes are created in graph G_1. At snapshot 2, cluster $c4$ and $c5$ are computed. $c4$ is compared with each node of graph G_1 where $c4$ is superset of $c1$. Node of $c1$ becomes the child of node of $c4$, and timestamp list of $c4$ is added in the timestamp list of $c1$. Cluster $c5$ is a subset of $c2$ and thus, becomes the child node of $c2$, and timestamp list of $c2$ is added in the timestamp list of $c5$. Timestamp list of $c3$ is updated by adding 2 in the list as the cluster appears again in timestamp 2. Groups $g1$, $g3$, and $g5$ are reported as swarms because their timestamp list size is at least equal to $timeTh$, i.e., 2.

At snapshot 3, cluster $c6$ and $c3$ are computed, where $c6$ is a superset of $c4$. Thus, node of $c6$ becomes the parent of node $c4$. Timestamp list of $c6$ is added in the timestamp list of $c4$ and $c1$. The intersection of cluster $c6$ and $c2$ is of size greater than the threshold value, and thus, a new node for cluster $c7 = c2 \cap c6$ is created. Timestamp of $c7$ is union of timestamp list of $c2$ and $c6$. Timestamp of $c3$ is updated by adding timestamp 3 in the list. At timestamp 3, in addition to $g1$, $g3$ and $g5$, new swarms $g4$ and $g7$ are reported. Since window 1 ends at timestamp 3 and

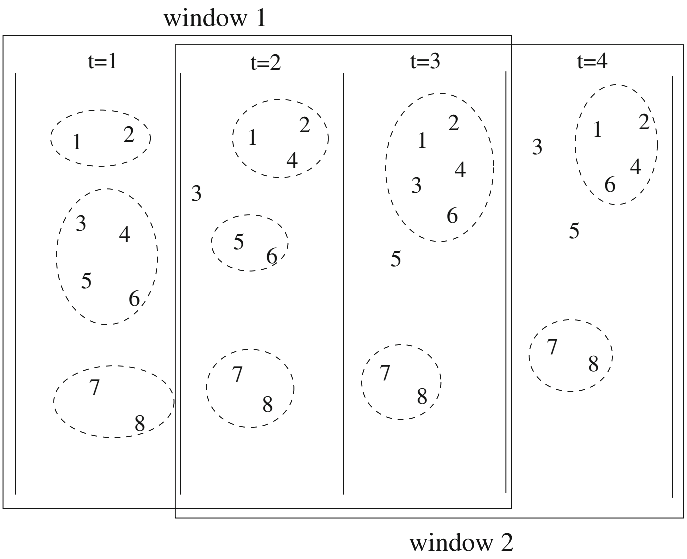


Fig. 1 Example

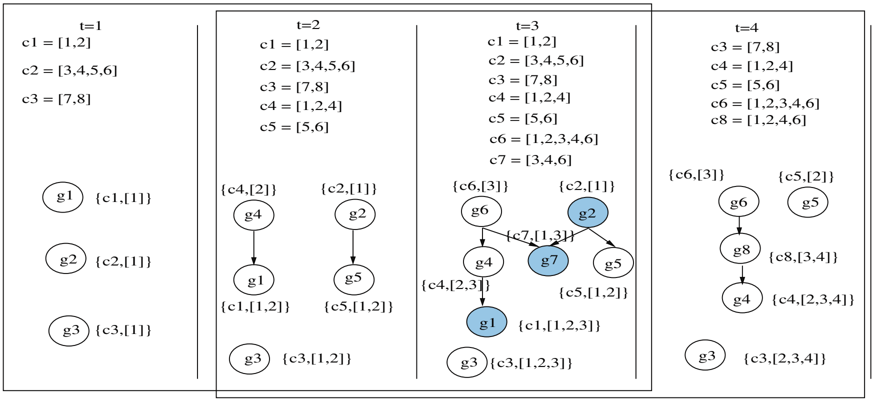


Fig. 2 Computed graphs of the example at each snapshot

new window starts at timestamp 2, data of timestamp 1 needs to be removed from the graph G_3. For this, at timestamp 3, timestamp 1 is removed from the timestamp list of each node. In such case, $g2$ has the empty timestamp list, and 1 is removed from the timestamp list of nodes $g1$, $g3$, $g5$ and $g7$. Timestamp list of node $g1$ is equal to the timestamp list of node $g4$. Thus, $g1$ is not a closed swarm in window 2. Similarly timestamp list of $c7$ is equal to timestamp list of $c6$ and is not closed swarm in window 2. Thus, nodes $g2$, $g1$ and $g7$ are removed from the graph. Node $g5$ does not have any parent node and thus becomes the root node. In snapshot 4, cluster $c8$ and $c3$ are computed. Cluster $c8$ is subset of cluster $c6$ and superset of cluster $c4$. Thus, it is added as child of node of $g6$ and parent of node $g4$. Timestamp 4 is added in the timestamp list of node $g4$ and $g3$. $g3$, $g4$ and $g8$ are reported as swarms.

5.1 Correctness Proof

The following lemmas prove that our method computes all closed swarm candidates present in the current window. Thus, swarms reported by our method must be equal to the swarms computed by previous swarm methods using the snapshots of the current window. Due to lack of space, proves are omitted.

Lemma 1 *For $t \geq 1$, for each new node $i \in N$, if there exists a node $j \in V_{t-1}$ such that $c_k = c_i \cap c_j$, $|c_k| \geq sizeTh$, then the group g_k must be present in G_t and must be in G_t^i and G_t^j before removing old data for next window $[t - L + 2, t + 1]$.*

Lemma 2 *For $t \geq 1$ and nonintersecting clusters at each timestamp t, every node in the graph G_t contains an object closed swarm candidate, i.e., for each node $i \in V_t$ and $j \in G_t^i$, $c_j \subset c_i$ and $T_i \subset T_j$. If $T_j = T_i$ for node j, then all its parents are* *intersecting clusters appeared at the same timestamp, and the object set of node j is the intersection of the object set of its parents.*

5.2 Runtime and Space Complexity

5.2.1 Runtime Complexity

The worst case time complexity for computing clusters from a snapshot is $O(n \log n)$ time [22], and $O(n_t + m_t)$ time is required to insert all the new swarm candidates computed from a cluster, where n_t and m_t are the number of nodes and edges, respectively. Thus, $O(n(n_t + m_t))$ time is required to compute all swarm candidates from the current snapshot at timestamp t.

To compute timestamps of each swarm candidate, a node is visited whenever its parent is visited to add the timestamps of its parents. Thus time complexity for computing swarms at timestamp t is $O(m_t \log n_t)$, where n_t is the number of nodes and m_t is the number of edges in the graph at timestamp t. The time complexity of computing new swarm candidates from the clusters is more than the runtime complexity of other operations.

5.2.2 Space Complexity

Nonincremental algorithm requires $O(nT)$ space to store object ids and locations, where n is the number of moving objects and T is the total number of snapshots in a window. We assume that there are no missing locations of any object. However, our incremental algorithm requires $O(n_t(n + T) + m_t)$ space, where n_t is the number of nodes (swarm candidates) computed from the dataset and m_t is the number of edges in the graph.

The runtime complexity and space requirement of our method along with previous methods for T timestamps are given in Table 1.

Table 1 Runtime complexity and space requirement for T snapshots of n moving objects. n_t represents the number of swarm candidates at timestamp t and $|m_t|$ represents the total number of edges in the graph

Method	Runtime	Space
Swarm [14]	$O(2^n nT)$	$O(nT)$
Swarm stream [16]	$O(n \sum_{t=1}^{T} n_t^2)$	$O(\max\{n_1(n+T), \ldots, n_T(n+T)\})$
Our method	$O(n \sum_{t=1}^{T} \|G_t\|)$	$O(\max\{n_1(n+T)+m_1, \ldots, n_t(n+T)+m_t\})$

6 Experiments

6.1 Datasets and Parameter Settings

We demonstrate our results on two real datasets which are related to movements of vehicles. The details are given below:

- **T-Drive Dataset**[1]: The T-Drive trajectory dataset consists of 1 week trajectories of 10,357 taxis moving in Beijing, China. The average sampling rate of the data is 177 s. It has been used previously in [23, 24]. Since the data is not synchronized, linear interpolation method is used in computing the snapshots at the interval of 3 min. The dataset contains 2964 snapshots. This dataset also contains some null trajectories which we have ignored in computing snapshots. Thus, number of objects in the dataset is 8,911, and each snapshot contains 8,911 locations.
- **Geolife Trajectory Dataset**[2]: The geolife trajectory dataset consists of 17,621 trajectories of 182 users moving mainly in Beijing, China. The data was collected at 5 s interval. It has been used previously in [25–27]. Since the data is not synchronized and the data is very sparse, we have ignored the trajectories which contained very few locations. We computed the snapshots using linear interpolation method at 1 min interval. The total number of objects and snapshots are 10,443 and 1441, respectively.

6.1.1 Clusters

Clusters are computed using the DBSCAN algorithm given in [22]. There are following two threshold parameters required for clustering: (i) density threshold (μ) that denotes to a minimum number of objects required in a neighborhood of an object to be included in the cluster and (ii) distance threshold (ϵ) that denotes to maximum distance required between two objects to get one object to be included in the neighborhood of another object. We set $\epsilon = 0.0009$ and $\mu = 2$ for geolife datasets, $\epsilon = \{0.004, 0.005, 0.008, 0.1\}$ and $\mu = 3$ for T-Drive dataset. We have considered different thresholds for T-Drive dataset to ensure that the number of clusters and average size of the clusters at every timestamp remain approximately same. All the clusters from each dataset are of size at most 60, and approximately 22% objects were included in the clusters at each timestamp. The statistics of each dataset is given in Table 2.

[1] https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/.

[2] https://www.microsoft.com/en-us/download/details.aspx?id=52367.

Table 2 Statistics of the datasets

Attributes	T-Drive		Geolife	
	10%	100%	10%	100%
# objects	891	8,911	1044	10,443
# snapshots	2,964	2,964	1,441	1,441
Avg # clusters	21	245	29	207
Avg size of a cluster	9	9	11	10

6.1.2 Parameters

The performance of our method depends upon the size threshold, number of objects in the dataset, and size of a sliding window because each of these parameters affects the size of the graph. The values of these parameters are given in Table 3. The default values of these parameters are given in bold. Since the previous methods are not scalable and the considered datasets are large, experiments are done while considering 10% of objects of T-DriSve and geolife datasets. Thus, statistics of both the datasets for 10% objects is also given along with the statistics of complete dataset in Table 2.

6.1.3 Base Cases

The performance comparison between swarm method and swarm stream has been presented in [15] who demonstrated that swarm stream method is significantly faster than the swarm method. Thus in this work, we compare our method with the swarm stream method only.

6.2 Results

All the experiments were performed in Java using Eclipse. The experiments are conducted on a PC with 3.60 GHz Intel core i7-4790 processor with 16 GB memory and Ubuntu 14.04 LTS operating system. For a sliding window model, the method

Table 3 Parameters

Parameters	Values
Sliding window size	**30**, 50, 70, 90, 110
Size threshold	5, 7, **9** , 11, 13
Ratio of objects	**10** % , 40 %, 70 %, 100 %
Time threshold	**10 %** of window size

should be memory and time efficient. We show the average runtime and average memory requirement per window for various values of parameters listed in Table 3. These averages were computed based on performance over 1000 sliding windows.

6.3 Comparison with Swarm Stream Method

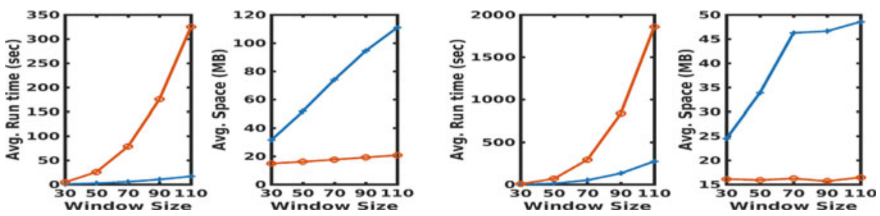
In this Section, we evaluate the performance of our method on two datasets and compare it with the swarm stream method (swarm stream) for various values of parameters.

6.3.1 Varying Sliding Window Size

Size of the sliding window is varied from 30 to 110 while keeping default values of other parameters. A swarm candidate can stay longer with large window size and can contribute in creating more new swarm candidates. Thus, as window size increases, the runtime is expected to increase due to increase in number of swarm candidates. The results shown in Fig. 3 for two datasets support the argument and show that the rate of increase in runtime of swarm stream method is more than our method. As compared to swarm stream method, our method is approximately 13 times and six times faster for T-Drive dataset and geolife dataset, respectively. However, our method requires approximately four times more memory for T-Drive dataset and two times more memory for geolife dataset.

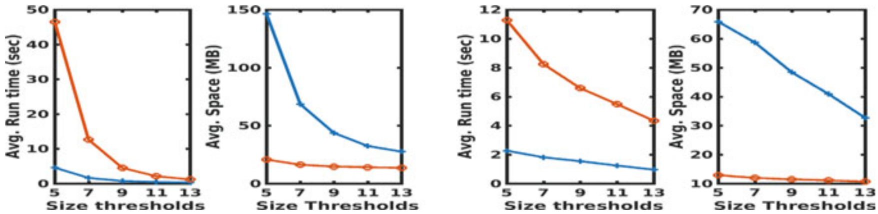
6.3.2 Varying Size Threshold

In this experiment, the size threshold is varied from 5 to 13 while keeping default values of other parameters.



(a) TDrive dataset (b) Geolife Dataset

Fig. 3 Results for varying sliding window size where average runtime is the average of total runtime in each window, and average memory is the average of total memory used in each window. In figure, −o− denotes swarm stream method and −+− denotes our method



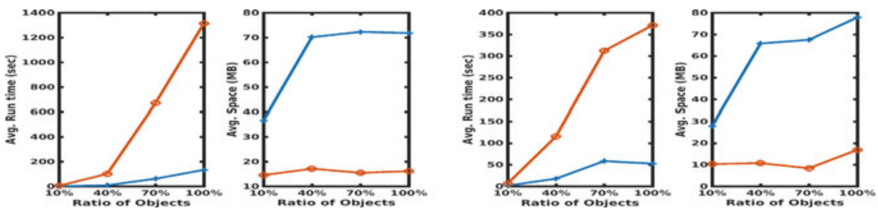
(a) TDrive Dataset (b) Geolife Dataset

Fig. 4 Results for varying size thresholds where average runtime is the average of total runtime in each window and average memory is the average of total memory used in each window. In figure, –o– denotes swarm stream method and –+– denotes our method

The results given in Fig. 4 show that the rate of decrease in runtime of both the methods decreases more with the increase in size threshold for T-Drive dataset than geolife dataset because T-Drive dataset contains large number of small-sized clusters which are included as swarm candidates for smaller sizeTh. As compared to swarm stream method, our method is approximately seven times and five times faster for T-Drive dataset and geolife dataset, respectively. However, as compared to swarm stream method, our method requires approximately three times and four times more memory for T-Drive dataset and geolife dataset, respectively.

6.3.3 Varying Number of Objects

In this experiment, we compare the performance of our method with swarm stream method for varying number of objects of each dataset while keeping default values of other parameters. It is expected that the runtime will increase with the increase in number of objects. The results given in Fig. 5 supports the argument and shows that the rate of increase in runtime of swarm stream is more than rate of increase in runtime of our method with the increase in number of objects. In particular our method is



(a) TDrive dataset (b) Geolife dataset

Fig. 5 Results for varying number of objects where average runtime is the average of total runtime in each window, and average memory is the average of total memory used in each window. In figure, –o– denotes swarm stream method and –+– denotes our method approximately eight times and five times faster than the swarm stream method for T-Drive and geolife dataset, respectively. However, as shown in Fig. 5, our method requires approximately five times more memory than the swarm stream method for T-Drive dataset and approximately six times more memory than the swarm stream method for geolife dataset.

7 Conclusion

In this work, a graph-based method is proposed for computing swarms from the recent data in real time. The runtime complexity of our method is significantly smaller than the previous swarm stream method, and the experimental results on real datasets show that our method performs 5–13 times faster than the previous swarm stream method. However, the memory requirement is 2–6 times more than the swarm stream method.

References

1. Babcock, B., Babu, S., Datar, M., Motwani, R., Widom, J.: Models and issues in data stream systems. In: Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2002, pp. 1–16. ACM (2002)
2. Li, X., Čeikutė, V., Jensen, C., Tan, K.: Effective online group discovery in trajectory databases. IEEE Trans. Knowl. Data Eng. **25**(12), 2752–2766 (2013)
3. Patroumpas, K., Kefallinou, E., Sellis, T.: Monitoring continuous queries over streaming locations. In: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2008, pp. 81:1–81:2. ACM (2008)
4. Vieira, M., Bakalov, P., Tsotras, V.: On-line discovery of flock patterns in spatio-temporal data. In: ACM-GIS, pp. 286–295 (2009)
5. Yu, Y., Cao, L., Wang, Q.: Detecting moving object outliers in massive-scale trajectory streams. In: KDD, pp. 422–431 (2014)
6. Shekhar, S., Yang, K., Gunturi, M., Manikonda, L., Oliver, D., Zhou, X., George, B., Kim, S., Wolff, M., Lu, Q.: Experiences with evacuation route planning algorithms. Int. J. Geogr. Inf. Sci. **26**(12), 2253–2265 (2012)
7. Song, X., Zhang, Q., Sekimoto, Y., Shibasaki, R.: Intelligent system for urban emergency management during large-scale disaster. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2014, pp. 458–464. AAAI Press (2014)
8. Rajalakshmi, T.M.P., Desai, U.: Novel sampling algorithm for human mobility-based mobile phone sensing. IEEE Internet Things J. **2**(3), 210–220 (2015)
9. Datar, M., Gionis, A., Indyk, P., Motwani, R.: Maintaining stream statistics over sliding windows. SIAM J. Comput. **31**(6), 1794–1813 (2002)
10. Datar, M., Motwani, R.: The Sliding-Window Computation Model and Results, pp. 149–167. Springer, Boston (2007)
11. Mouratidis, K., Papadias, D.: Continuous nearest neighbor queries over sliding windows. TKDE **19**(6), 789–803 (2007)
12. Benkert, M., Gudmundsson, J., Hubner, F., Wolle, T.: Reporting flock patterns. Comput. Geom. **41**, 111–125 (2008)

13. Jeung, H., Yiu, M.L., Zhou, X., Jensen, C.S., Shen, H.T.: Discovery of convoys in trajectory databases. Proc. VLDB Endow. **1**(1), 1068–1080 (2008)
14. Li, Z., Ding, B., Han, J., Kays, R.: Swarm: mining relaxed temporal moving object clusters. Proc. VLDB Endow. **3**(1–2), 723–734 (2010)
15. Li, Z., Han, J., Ji, M., Tang, L.-A., Yu, Y., Ding, B., Lee, J.-G.: Movemine: mining moving object data for discovery of animal movement patterns. ACM Trans. Intell. Syst. Technol. **2**(4), 37:1–37:32 (2011)
16. Yu, Y., Wang, Q., Kuang, J., He, J.: Tgcr: an efficient algorithm for mining swarm in trajectory databases. In: Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, pp. 90–95 (2011)
17. Bhushan, A., Bellur, U., Sharma, K., Deshpande, S., Sarda, N.L.: Mining swarm patterns in sliding windows over moving object data streams. In: SIGSPATIAL 2017, no. 60, pp. 1–4 (2017)
18. Zheng, K., Zheng, Y., Yuan, N., Shang, S.: On discovery of gathering patterns from trajectories. In: ICDE, pp. 186–197 (2013)
19. Fan, Q., Zhang, D., Wu, H., Tan, K.-L.: A general and parallel platform for mining co-movement patterns over large-scale trajectories. Proc. VLDB Endow. **10**(4), 313–324 (2016)
20. Yu, Y., Qi, J., Lu, Y., Zhang, Y., Liu, Z.: MR-Swarm: Mining Swarms from Big Spatio-Temporal Trajectories Using MapReduce, pp. 568–575. Springer International Publishing (2016)
21. Zheng, K., Zheng, Y., Yuan, N., Shang, S., Zhou, X.: Online discovery of gathering patterns over trajectories. IEEE Trans. Knowl. Data Eng. **26**(8) (2014)
22. Gunawan, A.: A faster algorithm for DBSCAN. Master thesis, Technische Universiteit Eindhoven (2013)
23. Yuan, J., Zheng, Y., Xie, X., Sun, G.: Driving with knowledge from the physical world. In: The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, pp. 316–324. ACM (2011)
24. Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G., Huang, Y.: T-drive: driving directions based on taxi trajectories. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2010, pp. 99–108. ACM (2010)
25. Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.: Understanding mobility based on gps data. In: Proceedings of ACM Conference on Ubiquitous Computing (UbiComp 2008), pp. 312–321. ACM (2008)
26. Zheng, Y., Xie, X., Ma, W.: Geolife: a collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull. **33**(2), 32–40 (2010)
27. Zheng, Y., Zhang, L., Xie, X., Ma, W.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of International conference on World Wild Web, WWW 2009, pp. 791–800. ACM (2009)

Digital Watermarking of Geospatial Data

H. Tulapurkar, S. Z. Choudhary, R. Ansari, P. V. Arun, J. Mohanty, P. Venkatachalam and K. M. Buddhiraju

Abbreviations

MRA	Multi-resolution Analysis
CPSNR	Cumulative PSNR
CMSE	Cumulative MSE
DCT	Discrete cosine transform
DCTG1	Discrete Curvelet Transform Generation 1
DCTG2	Discrete Curvelet Transform Generation 2
DFT	Discrete Fourier Transform
FDCT	Fast Discrete Curvelet Transform
FFT	Fast Fourier Transform
MSE	Mean Square error
PSNR	Peak signal to Noise ratio
SURF	Speeded-Up Robust Features
USFFT	Unequally Spaced Fast Fourier Transform

The rapid development of technology has made archival and transmission of multimedia information such as music, image, and video very convenient. However, it has also introduced new challenges related to privacy and security of data. Therefore, there is a compelling need for ensuring of authenticity and protection of ownership. For providing security of digital data, various techniques are used like encryption, decryption, cryptography, and digital watermarking. Digital watermarking is a technique

H. Tulapurkar (✉) · P. V. Arun · J. Mohanty · P. Venkatachalam · K. M. Buddhiraju
Centre of Studies in Resources Engineering, IIT Bombay, Mumbai, India
e-mail: harshulat@yahoo.com

S. Z. Choudhary
Department of Information Technology, A. C. Patil College of Engineering, Navi Mumbai, India

R. Ansari
K. J. Somaiah College of Engineering, Mumbai, India

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_21

of embedding selected user information into the digital content like image, video, speech, music. The watermarking algorithm should be able to detect intentional tampering of the original data and retain its integrity within the content even after various manipulation attacks like compression, enhancement, cropping, filtering.

1 Introduction to Watermarking

A digital image is a collection of two-dimensional data as a finite set of digital values called picture elements or pixels [1] organized in a matrix of M rows and N columns. Processing a digital image by using a digital computer is called digital image processing. For providing the security of digital data, various techniques are used like encryption, decryption, cryptography, steganography, and digital watermarking. The problem of copyright infringement has brought new opportunity for the effective protection of intellectual property. New techniques have been invented to embed the company logo, specific digital identifier, and other information into the multimedia files for the sake of identification of ownership. Such a technique is called digital watermarking, which facilitates embedding visible or invisible watermarks of user's choice in image, speech, or video data.

The basic requirements of watermarking are:

(1) Transparency
Watermarking information is embedded in a digital media host, such that it is imperceptible and without causing degradation to the original media.
(2) Robustness
Watermarking must be immune to common attacks like brightness, contrast, saturation, tint adjustments, low-pass filtering, JPEG compression attack, Gaussian noise attack and Laplacian filtering, symmetrical and asymmetrical image cropping, geometric attacks like scaling and rotation.
(3) Verifiability
Watermark must be giving reliable and complete information about the ownership or copyright protection data of the product. It should determine whether or not the object is protected or not and whether it should monitor its spread. Also, identify its authenticity and control the spreading of illegal copies.
(4) Security
Watermark data should have the unique and correct information to identify and this detection should only be possible by an authorized person who can legally detect this watermark and also make necessary modification if needed to preserve the copyright.

2 Digital Watermarking Basic Architecture

Every digital watermarking technique includes two algorithms: one as the embedding algorithm and another as the detecting algorithm. These two processes are same for all the type of watermarking (Fig. 1).

2.1 Embedding Stage

The embedding stage is the first stage in which the watermark is embedded in the original image by using the embedding algorithm and the secret key. Then, the watermarked image is generated. This watermarked image is transmitted over the network.

2.2 Detection/Retrieval Stage

In the detection stage, the watermark is detected or extracted by the detector from the watermarked image by applying some detection algorithm and by using secret key. Robustness of the system depends on how it handles common attacks like brightness, contrast, saturation, tint adjustments, low-pass filtering, JPEG compression attack, Gaussian noise attack and Laplacian filtering, symmetrical and asymmetrical image cropping, geometric attacks like scaling and rotation.

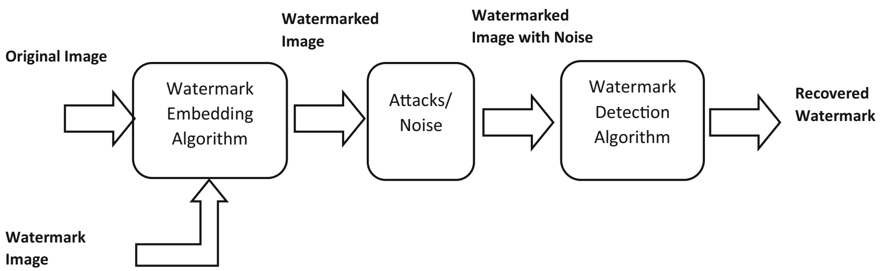


Fig. 1 Digital watermarking basic architecture [2]

3 Classification of Digital Watermarking Algorithm

Digital watermarking algorithms are classified based on working domain, type of document, human perception, or application area [1]. Embedding watermarks in the least-significant bits is an easy way to detect any attempts to tamper with the *fragile* watermark, but it offers virtually no protection against attacks. Embedding the watermark in higher-order bits makes it visible, if the location of embedding is not carefully chosen. It is easy to understand that near constant intensity areas in the image shows up any change in pixel values, the same in textured areas where bright, medium bright, and dark gray levels occur together as neighbors are better options as sites for *robust* watermark embedding. Embedding the watermark in higher-order bits in textured areas makes it withstand attacks better since the change in brightness level of a pixel is less likely to change by a significant amount due to simple image processing operations.

Transforming the image and embedding the watermark in the transform coefficients makes it even harder to locate since the attacker may not know which transform was employed to embed the watermark. In the next section, two multi-resolution approaches, viz. wavelet and curvelet decompositions of the image, are evaluated for embedding the watermark in the transform domain. A number of attacks are simulated and applied on the watermarked image, and the retrieved watermark is used to evaluate the robustness of the watermarking technique. It is found as will be seen that the curvelet-based watermarking is quite robust compared to many other schemes.

4 Multi-resolution Image Processing Transforms

Multi-resolution analysis (MRA), as implied by its name, analyses the signal at different frequencies with different resolutions. Multi-resolution offers a natural, hierarchical view of information.

Transform theory plays a fundamental role in image processing, as working with the transform of an image instead of the image itself may give us more insight into the properties of the image. Transform offers

- Conceptual insights into spatial frequency information;
- Fast computation;
- Alternative representation and sensing;
- Efficient storage and transmission.

Wavelet and **curvelet** transforms overcome this shortcoming of the Fourier transform by giving a time–frequency joint representation. **Wavelet** transform [3] provides a time–frequency joint representation by cutting the signal of interest into several parts and then analysing each of the parts separately. It is clear that analyzing a signal in this way will give more information about the when and where of different frequency components. Wavelet transforms are based on small waves called wavelets of varying frequency and limited duration. In wavelet analysis, a fully scalable modulated window is used to cut the signal. The window is shifted along the signal, and for every position, the spectrum is calculated. Then, this process is repeated many times with a slightly shorter (or longer) window for every new cycle. The wavelet coefficients measure how closely correlated the wavelet is with each section of the signal. In the end, the result will be a collection of time–frequency representations of the signal, all with different resolutions. Several families of wavelets that have proven to be useful include Haar, Daubechies, Morlet, Mexican Hat, Meyer. For compact representation, choose a wavelet that matches the shape of the image components—example: Haar wavelet for black and white drawings.

Curvelet transform [4] is a special members of the family of multiscale orientation-selective transforms and was developed as an answer to the weakness of the separable Wavelet transform in sparsely representing what appears to be simple building block atoms in an image, i.e., lines, curves, and edges. Ridgelet transform involves taking a wavelet transform (1-D WT) along the radial variable in the Radon domain. The curvelet transform, like the wavelet transform, is a multiscale transform, with frame elements indexed by scale and location parameters. It preserves the same time–frequency localization property as for wavelets, and at the same time, with their elongated support in the Fourier domain, curvelet becomes directional. It acts like a bandpass filter. In addition, anisotropic scaling principle, which is quite different from the isotropic scaling of wavelets, helps in sparse representation. The elements obey a special scaling law, where the length of the support of a frame elements and the width of the support are linked by the relation width $\approx$ length^2. Figure 2 shows the nonlinear, parabolic approximation.

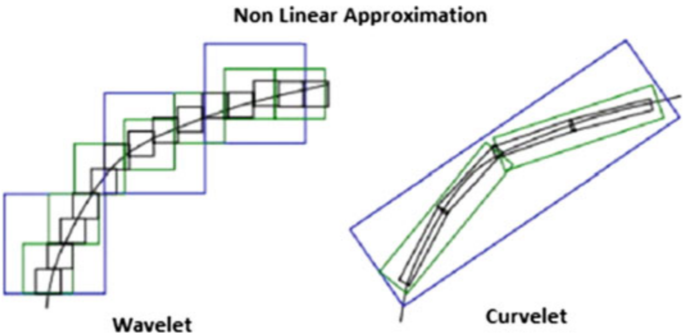


Fig. 2 Wavelet and curvelet representations of a curvilinear feature

4.1 Watermark Embedding Algorithm in Multi-resolution Framework

A robust image-watermarking scheme in curvelet domain is proposed. The second-generation curvelet transform, fast discrete curvelet transform (FDCT), which is proposed, by Candes et al. [5] is the basis for this algorithm. Candes et al. [5] suggested two strategies, namely unequally spaced fast Fourier transform (USFFT) and frequency wrapping. In this algorithm, we have the wrapping DCT method is used. Curvelab2.1.3 software package [6] implements the above FDCT_WRAPPING algorithm. This has been used to get curvelet transform of any given Image. The output of a 4 scale, 8 angle curvelet transform results is

Curvelet Transform output = $\{1 \times 1\}\{1 \times 8\}\{1 \times 16\}\{1 \times 16\}$
$\{1 \times 1\}$—Represents the low-frequency components.
$\{1 \times 8\}$—Scale 2. Each of the 8 columns represents data for that angle (wedge).
$\{1 \times 16\}$—Scale 3. Each of the 16 columns represents data for that angle (wedge).
{1x16}—Scale 4. Each of the 16 columns represents data for that angle (wedge).

4.2 Embedding Algorithm

This algorithm incorporates semi-blind robust watermark extraction, wherein the original image is not available for extraction; hence, a key containing detail about embedding is attached to the watermarked image. Watermark is embedded in the third band of the image. Figure 3 presents the flowchart depicting the embedding algorithm and Fig. 5 presents the Host image with embedded watermarks.

(1) Resize the Host Image to MXN such that M and N are power of 2
(2) Choose a watermark image (A $\times$ B) and convert it into binary. Serialize the watermark Ws. Ws = {Wk = 1, 2, 3, …, A $\times$ B; Wk $\{-1,1\}$}
(3) Split the Host image into smaller non-overlapping blocks (e.g. image size 2048 $\times$ 2048 block size is 512 $\times$ 512; image size 512 $\times$ 512 block size is 64 $\times$ 64)
(4) Extract co-occurrence texture features for each block. Mark blocks that have high texture feature (as specified by user) for embedding. Deselect 1 row and 1 column of blocks from all the four edges to ensure that no watermarks are lost when borders are cropped
(5) For each marked block do the following
    (a) Apply Wrapping Curvelet Transform (fdct_wrapping) with finest level as wavelet, 4 scales and 8 angles. C denotes this Curvelet Transform
    (b) Implement LOCEDGES logic that selects locations and orientation for embedding watermarks in Scale 3 Curvelet coefficients (section III A). arr_max contains orientation and location for embedding Wk = 1 and arr_min contains orientation and location for Wk = -1

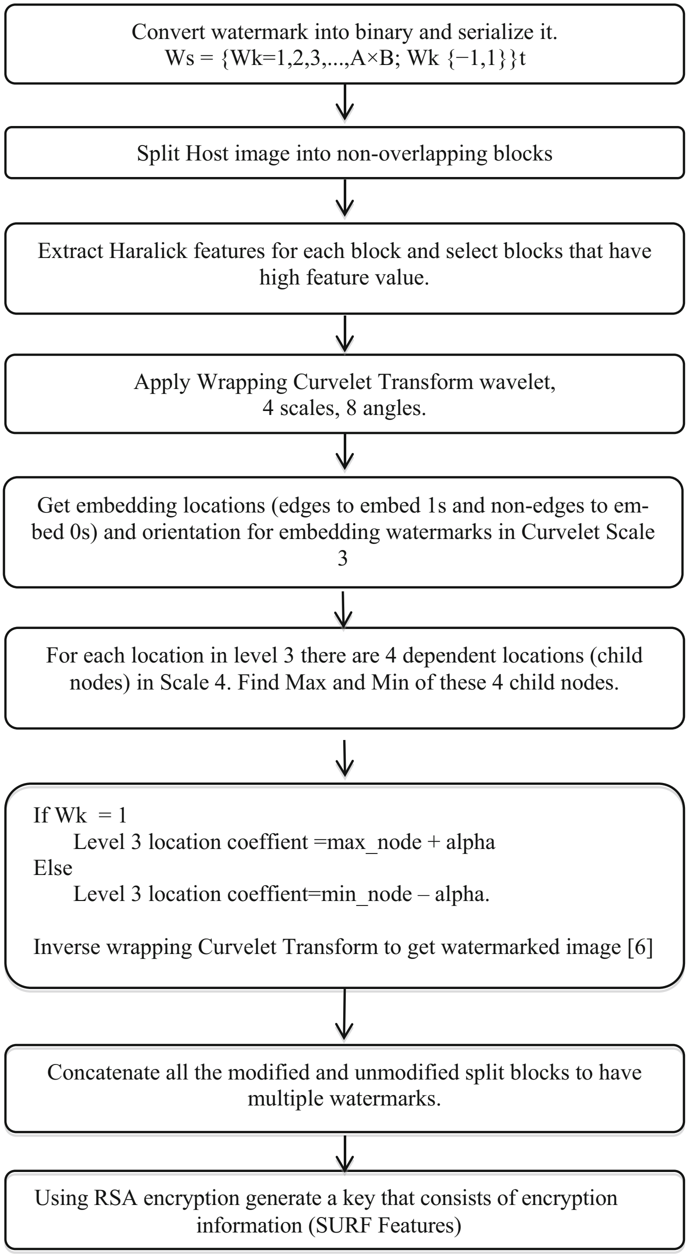


Fig. 3 Flowchart of algorithm for embedding invisible watermark using curvelet transform

(c) For each location in level 3, S3O (i, j), O indicates orientation; there are 4 dependent locations in Scale 4, S4O. (2i, 2j; 2i, 2j-1; 2i-1, 2j; 2i-1, 2j-1). These are called child nodes. Find Max (max_node) and Min (min_node) of these 4 child nodes
(d) If Wk = 1 select location (i, j) and orientation (O) from arr_max C {1,3} {1, O} (i, j) = max_node + alpha else select location (i,j) and orientation (O) from arr_max C {1,3} {1, O} (i,j)= min_node - alpha End if The chosen value of alpha = 160, which can be adjusted to change the strength of invisibility
(e) Apply Inverse wrapping Curvelet transform to get watermarked image

(6) Concatenate all the split blocks (modified and unmodified) to form Host image with multiple Generate a key by using RSA encryption to encrypt following data

(a) SURF features of the watermarked image. This includes features and valid points for each descriptor. This is useful for visualizing the descriptor orientation
(b) Watermarked location, orientation and original coefficient value of each watermarked block
(c) Blocks that are watermarked
(d) Original size of the Host Image and watermark image

4.2.1 LOCEDGES

Logic for selecting orientation and locations for embedding watermarks This algorithm chooses edges (high value coefficients) for embedding 1's in watermark and low value coefficients for embedding -1's in watermark.

(1) Scan the entire Curvelet coefficients across all orientations in Level 3 and create an array arr_level_max that holds value of maximum coefficient in each orientation
(2) Select the orientation from arr_level_max with maximum coefficient value max_value
(3) Set threshold T_{max}=0.5*max_value. Create an array arr_max that holds orientation and location of all coefficient values greater than T_{max}
(4) If no. of locations found is less than total no of 1s in watermark repeat step 3 with next highest value from arr_level_max
(5) For identifying locations to embed -1s set T_{min}=0 and choose all locations in a given orientation which are less than T_{min}

4.3 Key Features of the Algorithm

- Use of Haralick Features [7] to identify areas in the image for watermarking.
- Use of Speeded-Up Robust Features (SURF) [8] to detect any cropping and scaling attacks.
- Watermarks embedded along the curves ensure robustness against geometric attacks like scaling and rotation, compression, contrast enhancement, and Gaussian noise attacks.
- Watermarks embedded along the smooth areas ensure robustness against average and low-pass filtering.

4.4 Watermark Extraction Algorithm

Figure 4 is the flowchart that describes the extraction algorithm.

Algorithm first checks for any geometrical attacks. It involves

1. Get the SURF features of the original embedded image which are saved during embedding.
2. Find the SURF features of the altered embedded image.
3. Find the matching points between the original and altered images.
4. Restore the image. (This takes care of the size and rotation attacks.)

Another technique for detecting geometric attacks is using Radon transform where the maximum value in the Radon transform can be used to detect scale change and the rotation can be detected by location of the maximum value. However, this cannot detect any asymmetrical cropping of the image, and hence, above algorithm provides a robust solution to detecting geometric attacks and asymmetrical cropping (Fig. 5).

4.5 Results of Image Watermarking Using Multi-resolution Methods

A satellite image SAT5 (1856×1404) was used as host image and watermark image was of size 7×22. Robustness of algorithm for a variety of watermarking attacks was tested. Same embedding logic was incorporated in wavelet transform as well as curvelet transform and then the performance was compared. Figures 6 and 7 indicate the host image and watermark image, respectively.

Haralick texture feature 'Contrast' was used to identify watermarking area. Response to six attacks, viz. compression, Gaussian noise, geometric attack, contrast adjustment, smoothing, is tabulated below (Tables 1, 2, 3, 4, 5, 6). Table 7 shows samples of retrieved watermarks in various attacks. Figures 7 and 8 show the difference

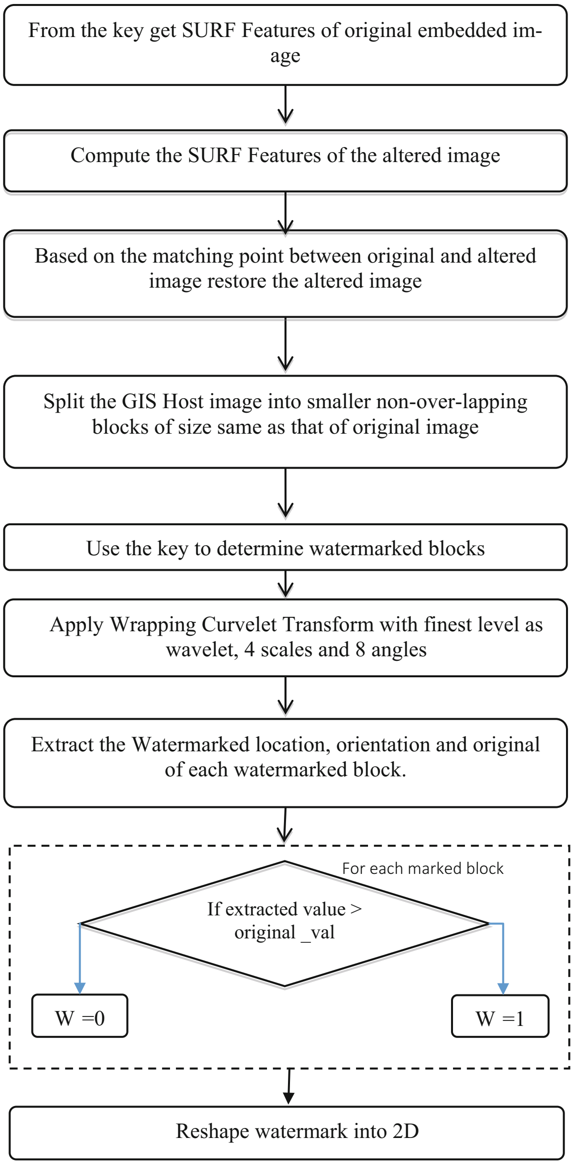


Fig. 4 Flowchart for extraction of invisible watermark embedded using curvelet transform

Fig. 5 Host image SAT5 (1856 × 1404)

Table 1 Comparative analysis of Gaussian attack—curvelet versus wavelet watermarking

% Noise in dB	Curvelet watermarking				Wavelet watermarking			
	#W embedded	#W extracted	CPSNR	CMSE	#W embedded	#W extracted	CPSNR	CMSE
10	8	0	65.36	4.7	11	0	61.74	11.45
15	8	8	Infinity	0	11	0	63.13	10.12
20	8	8	Infinity	0	11	0	64.14	10.38
25	8	8	Infinity	0	11	0	64.34	5.3
30	8	8	Infinity	0	11	0	65.36	4.4

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

between watermarked image and host image using curvelet and wavelet transform, respectively.

From the above observation, it is evident that the curvelet watermarking algorithm fails the low-pass filtering and averaging attacks. This was expected because the algorithm chose edges for embedding the watermarking. Hence when filtering is done, the edges are smoothened and watermarks are lost. Therefore, a hybrid-watermarking algorithm is introduced. In this, 80% of the watermarks are embedded

Fig. 6 Watermark image (7 × 22)

Table 2 Comparative analysis of geometric attack—curvelet versus wavelet watermarking

Scaling (S), Rotation (R) factor		Curvelet watermarking				Wavelet watermarking			
S	R	#W embedded	#W extracted	CPSNR	CMSE	#W embedded	#W extracted	CPSNR	CMSE
1	2	8	7	70.06	1.20	11	1	63.07	5.07
1	4	8	6	67.62	3.89	11	0	63.12	6.49
1	6	8	4	63.11	4.20	11	2	63.14	5.54
1	10	8	4	64.17	4.80	11	0	64.62	6.84
1.2	2	8	6	60.45	1.56	11	0	62.69	7.32
1.2	4	8	6	58.46	4.10	11	0	61.50	4.40
1.2	6	8	4	56.12	4.60	11	0	63.34	5.19
0.21	10	8	4	55.43	5.10	11	0	63.98	6.96

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

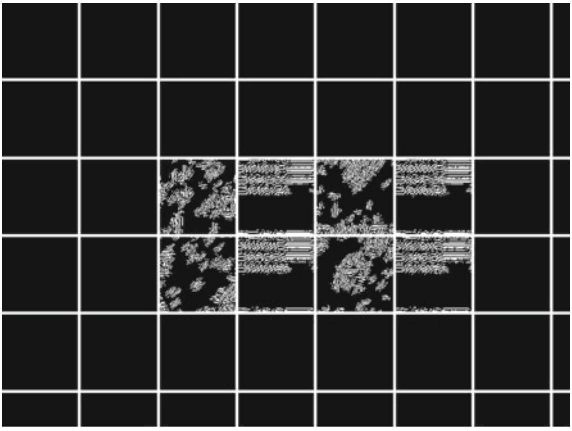


Fig. 7 Difference between the host image and curvelet watermarked image. (Eight watermarked blocks and the dashes indicate watermarking)

Table 3 Comparative analysis of contrast enhancement attack—curvelet versus wavelet watermarking

Technique	Curvelet watermarking				Wavelet watermarking			
	#W embedded	#W extracted	CPSNR	CMSE	#W embedded	#W extracted	CPSNR	CMSE
Histogram equalization	8	5	70.06	1.6	11	0	55.12	30.46
Adjusting image intensity to increase contrast	8	8	Infinity	0	11	0	60.06	8.50
Contrast-limited adaptive histogram equalization	8	6	66.96	0.81	11	0	58.29	22.43

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

using the above logic, whereas for the 20% watermarks locations are chosen where there is not much variation; thus, they can with stand averaging and low-pass filtering attacks. Embedding in high-frequency coefficients offers better imperceptibility, while low-frequency coefficients provide high robustness against filtering attacks.

Table 4 Comparative analysis of low-pass filtering attack—curvelet versus wavelet watermarking

Filter	Curvelet watermarking				Wavelet watermarking			
	#W embedded	#W extracted	CPSNR	CMSE	#W embedded	#W extracted	CPSNR	CMSE
Averaging	8	0	57.14	101	11	0	68.32	1.89
Gaussian LPF (f = 10)	8	0	Infinity	0	11	1	70.06	1.89
Gaussian LPF (f = 100)	8	0	57.18	112	11	1	70.06	1.89
Gaussian LPF (f = 1 k)	8	0	57.19	115	11	1	70.06	1.89
Gaussian LPF (f = 10 k)	8	0	57.19	116	11	1	70.06	1.89

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

Table 5 Comparative analysis of cropping attack—curvelet versus wavelet watermarking

Cropping	Curvelet watermarking				Wavelet watermarking			
	#W embedded	#W extracted	CPSNR	CMSE	#W embedded	#W extracted	CPSNR	CMSE
10% Left	8	8	Infinity	0	11	11	Infinity	0
10% Right	8	8	Infinity	0	11	11	Infinity	0
10% Top	8	8	Infinity	0	11	11	Infinity	0
10% Bottom	8	8	Infinity	0	11	11	Infinity	0

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

4.6 Conclusions on Image Watermarking

This new approach for embedding invisible watermarking using curvelet transform shows improved performance over wavelet transform when embedding logic is same. Robustness against variety of attacks is due to use of texture features to select blocks combined with selection of appropriate locations for embedding. Use of SURF features serves in synchronizing embedding location, which helps in detecting and recovering from geometric attacks resulting in negligible MSE. The proposed technique can also be used to watermarked multiband images.

Table 6 Results for various attacking for hybrid curvelet watermarking algorithm

Filter	#W embedded	#W extracted	CPSNR	CMSE
Averaging	1	0	61.55	9.09
Gaussian LPF (f = 10)	1	0	62.22	7.79
Gaussian LPF (f = 100)	1	0	70.06	2.28
Gaussian LPF (f = 1 k)	1	0	60.90	10.3
Gaussian LPF (f = 10 k)	1	0	62.22	7.79

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

Table 7 Samples of retrieved watermarks in various attacks

Attack	Curvelet watermarking				Wavelet watermarking			
	#W embedded	#W extracted	Max MSE	Max error watermark	#W embedded	#W extracted	Max MSE	Max error watermark
20% Compres-sion	8	0	7.79		11	0	44.15	
10 dB Gaussian noise	8	0	3.9		11	0	11.45	
Average filter*	8	0	9.09		11	0	1.89	
Contrast adjust-ment	8	5	1.6		11	0	48.12	
Scaling and rotation	8	7	1.2		11	1	9.09	

Notations used in table
#W embedded no. of watermarks embedded
#W retrieved no. of watermarks extracted correctly
CPSNR Cumulative PSNR (for all watermarks)
CMSE Cumulative MSE (for all watermarks)

It has been observed that the algorithm is not capable of handling low-pass filtering and average filtering attack. This is as expected because edges in the images are used for embedding watermark. To overcome this drawback, a hybrid embedding logic is incorporated where 20% watermarks are embedded in locations that do not have sharp edges. Eighty percent of watermarks are embedded in edges; this ensures robustness against filtering attacks.

Table 8 Summary of existing vector watermarking approaches

Research article	Domain	Robustness								Fidelity		
		NO	GT	VIR	VR	CR	MS	ST	FC	ERR	SD	TR
Ohbuchi et al. [10]	Spatial	√	√	√	√	×	–	–	–	Yes	No	No
Voig and Busch [11]	Spatial	–	√	–	–	–	√	–	–	Yes	No	No
Wang et al. [12]	Spatial	√	√	√	√	–	√	–	–	Yes	No	No
Shujun et al. [13]	Spatial	–	√	√	–	√	–	–	–	Yes	No	No
Schulz and Voigt [14]	Spatial	√	√	–	–	√	√	–	–	Yes	No	No
Marques et al. [15]	Spatial	√	–	–	√	–	–	√	–	No	No	No
Cao et al. [16]	Spatial	–	–	–	–	–	–	–	–	Yes	No	No
Huo et al. [17]	Spatial	√	√	√	√	√	–	–	–	No	No	No
Kim [9]	Spatial	√	√	√	–	–	–	–	–	Yes	No	Yes
Men et al. [18]	Spatial	–	√	–	–	–	×	–	–	No	No	No
Yan et al. [19]	Spatial	√	–	√	–	–	–	√	√	No	No	No
Wang et al. [20]	Spatial	–	√	–	–	×	×	–	–	Yes	No	No
Li et al. [21]	Spatial	–	√	–	–	√	–	–	–	No	No	No
Solachidis and Pitas [22]	DFT	√	√	×	–	×	–	–	–	No	No	No

(continued)

Table 8 (continued)

Research article	Domain	Robustness								Fidelity		
		NO	GT	VIR	VR	CR	MS	ST	FC	ERR	SD	TR
Kitamura et al. [23]	DFT	√	–	–	√	–	–	–	–	Yes	No	No
Tao et al. [24]	DFT	–	√	–	–	–	–	√	√	Yes	No	No
Xu and Wang [25]	DFT	–	√	√	–	–	–	–	√	Yes	No	No
Voigt et al. [26]	DCT	–	–	×	–	–	×	–	–	Yes	No	No
Wang et al. [27]	DCT	√	√	√	√	√	–	–	–	Yes	No	No
Liang et al. [28]	DCT	–	√	–	–	–	√	–	–	No	No	No
Lianquan and Qi-hong [29]	DCT	√	√	–	–	√	–	–	–	No	No	No
Li and Xu [30]	DWT	√	√	–	–	–	–	–	–	Yes	No	No
Zhu et al. [31]	DWT	√	–	√	–	–	√	–	√	Yes	No	No
Liu and Lv[32]	DCT	–	√	√	–	√	–	–	–	Yes	No	No

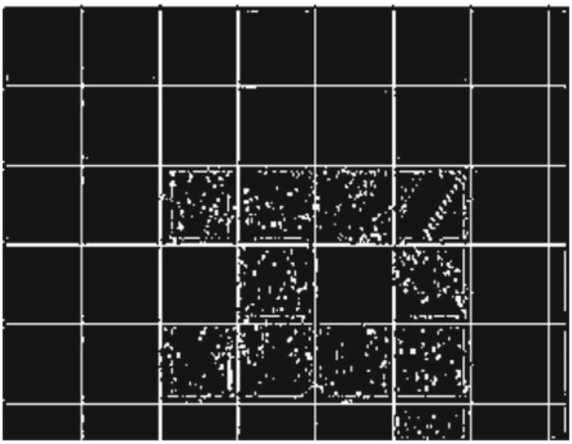


Fig. 8 Difference between the host image and wavelet watermarked image. (11 watermarked blocks and the dashes indicate watermarking)

5 Watermarking of Vector Datasets

The watermarking of vector datasets such as digital maps, CAD drawings are relatively less explored when compared to that of images and videos. The reason is obvious that there is a lot of redundancy in raster data like images and videos or time-dependent signals like speech and audio, while the vector datasets are inherently compact, with very little redundancy. In such cases, it is very difficult to find locations to embed the watermark data. It must be noted that a vector dataset, essentially, is a collection of (x,y) pairs or point locations. These locations have different meanings in different contexts. For example, in case of point data, an (x,y) pair may indicate location of a bus stop or a restaurant. In case of line segment data, an (x,y) pair may indicate a road junction. In case of polygon data, an (x,y) pair may indicate a point object located inside the polygon, like an electric transformer in a residential block. When a watermark point is embedded in vector data, the locations of the (x,y) pairs are displaced, even if the embedding takes place in 4th or 5th decimal place. In such cases, the criteria for choosing a vector host to embed the watermark will be very different compared to raster data. Having said that a map is much more valuable compared to a satellite image since the map is readily deployable information while the image is still data that needs to be interpreted to extract the usable information. Therefore, there have been recent attempts to embed watermark in the vector datasets as well. Watermarks may be directly embedded or encrypted so that even if the algorithm is known, the watermark may not be retrievable without knowledge of the encryption keys. Table 8 presents a comparative summary of vector data watermarking algorithms in spatial and transform domains based on: (1) robustness against attacks like noise (NO), geometric transformations (GT), vertex insertion and removal (VIR), vertex reordering (VR), cropping (CR), map simplification (MS), similarity transformation (ST), and format conversion (FC); (2) fidelity criteria such as error (ERR), shape distortion (SD), and topological relationship (TR). Shortcomings and unresolved issues in previous studies have been clearly highlighted. Very few algorithms are robust against all the attacks. None of the algorithm is checked against the important fidelity criteria like shape distortion and the topological relationship integrity. Kim [9] has claimed integrity of topological relationship using intersection test, but it fails to check all topological relations.

As in the case of raster watermarking, the embedding of watermarks in the transform coefficients may offer greater protection when compared to the direct embedding in the (x,y) coordinate data. Transforms such as wavelet and curvelet, which are widely used for raster watermarking, are currently being investigated for vector watermarking as well. In this chapter, some of our work in this direction is presented. The watermark too, instead of a raster image chip, is a small vector dataset.

References

1. Gonzalez, R.C., Woods, R.E.: Digital Image Processing (2002)
2. Rao, C.R.S., Prasad, M.V.: Digital Watermarking Techniques in Curvelet and Ridgelet Domain. Springer International Publishing (2016)
3. The Wavelet Tutorial Part I, Rowan University, College of Engineering Web Servers, 15
4. Starck, J.L., Murtagh, F., Fadili, J.M.: Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity. Cambridge University Press (2010)
5. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. Multiscale Model. Simul. **5**(3), 861–899 (2006)
6. http://www.curvelet.org/download-secure.php?file=CurveLab-2.1.3.tar.gz
7. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **6**, 610–621 (1973)
8. Chincha, R., Tian, Y.: Finding objects for blind people based on SURF features. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), pp. 526–527. IEEE (2011)
9. Kim, J.: Robust vector digital watermarking using angles and a random table. AISS **2**(4), 79–90 (2010)
10. Ohbuchi, R., Ueda, H., Endoh, S.: Robust watermarking of vector digital maps. In: Proceedings IEEE International Conference on Multimedia and Expo, ICME 2002, vol. 1, pp. 577–580. IEEE (2002)
11. Voigt, M., Busch, C.: Feature-based watermarking of 2D vector data. In: Electronic Imaging, International Society for Optics and Photonics, pp. 359–366 (2003)
12. Wang, C., Wang, W., Wu, B., Qin, Q.: A watermarking algorithm for vector data based on spatial domain. In: 2009 1st International Conference on Information Science and Engineering (ICISE), pp. 1959–1962 (2009)
13. Shujun, D., Liang, L., Shujun, D., Sen, C.: Research on a digital watermarking algorithm suitable to vector map. In: 2007 IEEE International Conference on Automation and Logistics, pp. 1236–1240 (2007)
14. Schulz, G., Voigt, M.: A high capacity watermarking system for digital maps. In: Proceedings of the 2004 Workshop on Multimedia and Security, pp. 180–186. ACM (2004)
15. Marques, K., Dahab, R.: SB-RAWVec—A semi-blind watermarking method for vector maps. In: 2009 IEEE International Conference on Communications, Dresden, pp. 1–6 (2009)
16. Cao, L., Men, C., Li, X.: Iterative embedding-based reversible watermarking for 2D-vector maps. In: 2010 17th IEEE International Conference on Image Processing (ICIP), pp. 3685–3688. IEEE (2010)

17. Huo, X.-J., Seung, T.-Y., Jang, B.-J., Lee, S.-H., Kwon, S.-G.: A watermarking scheme using polyline and polygon characteristic of shapefile. In: 2010 3rd International Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 649–652. IEEE (2010)
18. Men, C., Cao, L., Li, X., Wang, N.: Global characteristic-based lossless watermarking for 2D-vector maps. In: 2010 International Conference on Mechatronics and Automation (ICMA), pp. 276–281. IEEE (2010)
19. Yan, H., Li, J., Wen, H.: A key points-based blind watermarking approach for vector geo-spatial data. Comput. Environ. Urban Syst. **35**(6), 485–492 (2011). ISSN 0198-9715
20. Wang, C., Peng, Z., Peng, Y., Yu, L., Wang, J., Zhao, Q.: Watermarking geographical data on spatial topological relations. Multimed. Tools Appl. **57**(1), 67–89 (2012). ISSN 1380-7501
21. Li, A., Lin, B.X., Chen, Y., Lu, G.: Study on copyright authentication of GIS vector data based on zero-watermarking. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 38, pp. 1783–1786 (2008)
22. Solachidis, V., Pitas, I.: Watermarking polygonal lines using Fourier descriptors. IEEE Comput. Graph. Appl. **24**(3), 44–51 (2004). ISSN 0272-1716
23. Kitamura, I., Kanai, S., Kishinami, T.: Copyright protection of vector map using digital watermarking method based on discrete Fourier transform. In: IEEE 2001 International Geoscience and Remote Sensing Symposium, 2001, IGARSS 2001, vol. 3, pp. 1191–1193 (2001)
24. Tao, S., Dehe, X., Chengming, L., Jianguo, S.: Watermarking GIS data for digital map copyright protection. In Proceedings of 24th International Cartographic Conference, pp. 15–21 (2009)
25. Xu, D., Wang, Q.: The study of watermarking algorithm for vector geospatial data based on the phase of DFT. In: 2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), pp. 625–629 (2010)
26. Voigt, M., Yang, B., Busch, C.: Reversible watermarking of 2D-vector data. In: Proceedings of the 2004 Workshop on Multimedia and Security, MM 38 Sec 2004, pp. 160–165. ACM, New York, USA (2004). ISBN 1-58113-854-7
27. Wang, X., Huang, D.J., Zhang, Z.Y.: A DCT-based blind watermarking algorithm for vector digital maps. Adv. Mater. Res. **179**, 1053–1058 (2011)
28. Liang, B., Rong, J., Wang, C.: A vector maps watermarking algorithm based on DCT domain. ISPRS Congr. **38**(3) (2010)
29. Lianquan, M., Qihong, Y.: A digital map watermarking algorithm based on discrete cosine transform. J. Comput. Appl. Softw. 146–148 (2007)
30. Li, Y., Xu, L.: A blind watermarking of vector graphics images. In: International Conference on Computational Intelligence and Multimedia Applications, pp. 424–424. IEEE Computer Society (2003)
31. Zhu, C., Yang, C.S., Wang, Q.: A watermarking algorithm for vector geo-spatial data based on integer wavelet transform. ISPRS Congr. **37**, 15–18 (2008)
32. Liu, A., Lv, G.: Copyright protection of DEM by means of DCT-based digital watermarking. Geo Inf. Sci. **10**(2), 214–223 (2008)

A Review of Standards for Airborne LiDAR Data Acquisition, Processing, QA/QC, and Delivery

B. Lohani, S. Ghosh and A. Dashora

1 Introduction

Light detection and ranging (LiDAR) is an industry standard technique for collecting three-dimensional topographic data. The acquisition process and the resulting data from LiDAR technology are quite different from the data collected through other modes like radar interferometry, photogrammetry, land surveying, satellite imaging, and therefore, it carries different specifications for acquisition, processing, quality assessment and quality control (QA/QC), and delivery.

Topographic data collection through LiDAR technology is possible in three different modes, viz. terrestrial (including mobile and stationary laser sensors), airborne, and spaceborne [1]. In this paper, the discussion is limited to airborne LiDAR technology. Several airborne laser sensors have been made available in the market since the inception of LiDAR technology since 1996. To name a few, these sensors are available from Leica, Optech Inc., Riegl, Trimble, Velodyne, etc. [2]. Data can be captured either in point format, i.e., with x, y, and z coordinates, with other attributes, or in waveform mode. Mostly, the sensors available in the market have their own proprietary formats, which can be processed on proprietary software, typically for tasks like data visualization, quality assessment, classification, and storage. As in most cases, these data are captured by a vendor for a user or a user organization, it necessitates the need of an understanding between them regarding data capture procedure, data processing approach, and data delivery including the stages of different mile-

B. Lohani (✉)
Indian Institute of Technology Kanpur, Kanpur, India
e-mail: blohani@iitk.ac.in

S. Ghosh
MGM's Jawaharalal Nehru Engineering College, Aurangabad, India
e-mail: suddhasheel@gmail.com

A. Dashora
Indian Institute of Technology Guwahati, Guwahati, India
e-mail: abd@iitg.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_22

stones in this process. Further, in several cases, like the 3DEP (3D Elevation Program) in the USA [3], the data generated through public funding must meet a minimum standard for the use of these data in a nationwide data repository. Therefore, there is a need to establish a standard for various aspects, namely data procurement, data processing, QA/QC, data storage formats, and delivery with metadata.

Several standards have been proposed for each of these aspects through various national and international agencies or organizations. This paper reviews these existing standards and proposes a generic set of parameters relevant to a user in order to specify requirements for LiDAR datasets. It is considered that the availability of such generic list of parameters would help a user organization to assess their need and then adopt a subset of parameters for their respective projects.

2 Review of Standards

An exposition on the specifications of LiDAR data procurement was made by Martin Flood [12] in 2002, where the author highlighted several angles in terms of research

Table 1 Names of standards/guidelines proposed by various organisations/agencies

	Name	Agency	Version	Country	Year
1	Guidelines and specifications for flood hazard mapping partners [4]	Federal Emergency Management Agency (FEMA)	–	USA	2003
2	Light detection and ranging (LiDAR) requirements [5]	National Oceanic and Atmospheric Administration (NOAA)	–	USA	2009
3	ICSM LiDAR late [6]	Intergovernmental Committee on Surveying and Mapping (ICSM)	1	Australia	2010
4	Specifications for LiDAR [7]	GeoBC, British Columbia	2	Canada	2014
5	LiDAR base specifications [8]	United States Geological Survey (USGS)	1.2	USA	2014
6	New Zealand National Aerial LiDAR base specification [9]	Land Information New Zealand (LINZ)	–	New Zealand	2016
7	Federal airborne LiDAR data acquisition guideline [10]	Natural Resources, Canada (NRCAN)	1.1	Canada	2017
8	NDOR LiDAR mapping guidelines [11]	Nebraska Department of Roads (NDOR)	–	USA	2017

and applications. Several standards and guidelines for LiDAR data have been put forth since then, which are summarized in Table 1.

The forthcoming part of the document will review and discuss the various standards/guidelines, with reference to the aspects mentioned earlier. It is felt that the NDOR guidelines [11] are not comprehensive and therefore are not considered in the following discussions.

3 Comprehensive LiDAR Specifications

The three-dimensional topographic data, acquired through LiDAR technology, is in the form of points with x, y, and z coordinates and several other attributes as specified from time to time by ASPRS [13–17]. In general, LiDAR specifications can clearly be divided into the following areas: (a) data acquisition, (b) data processing, (c) QA/QC, (d) delivery including metadata. The general LiDAR specifications have been dealt in most detail, by the ASPRS, ICSM, and USGS [6, 8, 18]. Several terms appearing in this paper are not being defined here as it would unnecessarily increase the length of the paper while these definitions are generally known to LiDAR community or can easily be seen in the references cited.

3.1 Data Acquisition Specifications

Data acquisition specifications include the information about area of interest (AOI), collection conditions, data requirements, sensor requirements, pre-project activities, and parameters for scanning operations.

The AOI is usually supplied as an ESRI shapefile [8] or Microstation DGN file or Google KML file, as seen from the specifications. However, emerging formats, namely Geography Markup Language (GML), are also possible, which the LiDAR community can think of, as it is interoperable. The collection conditions are specific to the condition of the atmosphere, weather, ground, and water. When simultaneous photographic data acquisition is required, i.e., LiDAR as well aerial photography, the sun angle will also become an additional requirement. In addition, the camera and the laser scanner should be compatible, as mentioned by Dashora et al. [19].

It is observed that all of the specifications are negligent about the wind speed, however, Piel et al. [20] recommends that the wind speed should not exceed 20 knots.

Data requirements involve accuracy, overlap, data density, and constraint on across and along flight point spacing. These requirements are specific to the application for which LiDAR data are being acquired. It is observed that overlap and data density vary in range, respectively, from 10 to 15% [6, 10], 1–8 points per m^2 [21, 22]. On the other hand, maximum value of ratio of across-track to along-track data spacing is restricted to 2/3 [6]. Acquisition specifications also specify the requirement of GNSS reference stations with redundancy and maximum base length [6–8, 23].

Accuracy of LiDAR data is defined by horizontal accuracy (HA) and vertical accuracy, which is further subdivided into fundamental vertical accuracy (FVA), supplemental vertical accuracy (SVA), and consolidated vertical accuracy (CVA) [18]. As per the new definitions, the vertical accuracies are being reported as non-vegetated vertical accuracy (NVA) and vegetated vertical accuracy (VVA) [8]. Determination of horizontal accuracy of LiDAR is difficult and is not specifically stated in standards. Indirect means are suggested based on comparison of objects for this purpose.

3.2 Data Processing

Data processing deals with classification and respective accuracies, levels of classification, and derivables from LiDAR data. For classification and respective accuracies, one may refer to the documentation by ICSM [5], where it is mentioned that classification of LiDAR data can be performed in four levels, viz: (a) Level 1: automated and semiautomated classification, (b) Level 2: ground surface improvement, (c) Level 3: ground correction, and (d) Level 4: detailed classification and correction.

The derivables from LiDAR data include DSM, nDSM, DEM (Digital Elevation Model), CHM (Canopy Height Model), and foliage cover model (FCM) (as raster models) [6] with each having its respective specifications.

3.3 Quality Assessment and Quality Control (QA/QC)

QA/QC deals with the quantitative estimation of data density, uniform spatial distribution of data, overlap in different swaths, type of accuracies, and their assessment on the field.

Data density is verified using the nominal point spacing (NPS) and nominal pulse density (NPD), which have long been an industry standard. The assessments of the NPS and NPD are done using only the first return, a single instrument, and a single swath. However, lately the practice has been to cover the entire area using multiple coverages, and therefore the concepts of ANPS and ANPD have evolved [8].

The acquired data should be uniformly distributed across the swath and should not contain any voids. This can be ensured by creating a grid of size twice the ANPS, and each grid thus created should contain at least one LiDAR data point.

The types of accuracies have already been mentioned in Sect. 3.1. These definitions of accuracy are according to the type of terrain. The horizontal accuracy should also be assessed on the field [18]. For assessment of these accuracy measures, a minimum number of check/control points with their desired spatial accuracy are suggested [18]. Also it is important to specify the method that should be adopted for comparing LiDAR data with reference data.

3.4 Deliverables

For a standard project requiring LiDAR data, the deliverables are classified into two parts: (a) pre-acquisition derivables, and (b) post-acquisition deliverables.

Pre-acquisition deliverables comprise of specifications and calibration reports of the sensors, the flight planning and layout specifications, locations of GNSS stations, and reports on calibration survey. These deliverables should be promising to acquire data of desired quality.

Post-acquisition deliverables include quality assessment reports, project extents, data and derived data, breaklines (if applicable and needed), and product metadata for each of the products delivered. The data should include trajectory and smoothed best estimate of trajectory (SBET), LiDAR point cloud (unclassified and classified as desired), and the derivables, e.g., intensity image, DEM, DSM, nDSM, CHM, FCM. Tiling these products to files of smaller sizes have been recommended by some specifications, but since emergent 64-bit file systems and graphical processing units have enabled the handling of large files, tiling may also be avoided for smaller projects.

3.5 Data Storage

The ASPRS has recommended the LAS storage format from 1.0 in 2003 to 4.0 in 2011 [13–17]. These activities have been recorded on a web page [24]. Most specifications, which have been drawn up till date, have adhered to the ASPRS LAS storage format.

In this context, it is important to mention here that Martin Isenberg [25] developed the LASZip format for the compressed storage of laser point data, whereas ESRI designed the optimized LAS format [26]. To avoid confusion and keeping the data format open, the Open Source Geospatial Foundation (OSGeo) has supported the use of LAS format for data storage and delivery [27].

Although it is not specified in the available specifications, in case of simultaneous photographic data acquisition, orthorectified images with specific GSD should be delivered.

4 Generic Standard

This document has reviewed multiple specifications in the domain of airborne LiDAR, and dealt with the aspects of acquisition, QA/QC, processing and handling, and delivery specifications. Although many parameters may seem different, this document has brought out the factors common to most of the specifications and guidelines. Consequently, this document also proposes the following generic list of requirements for airborne LiDAR data acquisition, which can be referred while preparing a new standard by an organization or a country. Alternatively, this list would also help while a tender document is being prepared. The paper has intentionally not provided the values of any of the parameters listed below as these are very much application and location dependent and are available in existing standards. Also, specific values can be determined by users based on their application and consultation with experts.

1. Data acquisition
   a. Units of reference
   b. Horizontal datum and projection
   c. Vertical datum global and local
   d. Area of interest (AOI) including buffer
   e. Collection conditions: ground, water, atmosphere, weather, wind speed, sun angle, time of day
   f. Data requirements: accuracy, overlap, data density, comparable point spacing
   g. Sensor requirements: range accuracy, multiple returns, intensity information, waveform
   h. Sensor information: information, calibration reports and certification for GNSS base stations, IMU, laser scanner, and aerial camera, if applicable
   i. Flight plan: flying and scanning parameters should include FOV, scanning frequency, flying altitude, aerial platform speed, azimuth of flying direction, and PRF. Flight plan should specify the type and make of aerial vehicle, type of turning, starting line, line interval for turning, maximum banking angle, cushion period, and GDOP and PDOP variation map, and flight line layout.
   j. Locations of GNSS base stations and maximum base length
   k. Calibration survey and its reporting
2. QA/QC parameters
   a. Accuracies
      i. HA, FVA, SVA, CVA or NVA, and VVA for entire data
      ii. Relative accuracy for individual swath and adjacent swaths
   b. Data uniformity requirements (ANPS, ANPD, spatial distribution of data).
   c. Checkpoints (control points): minimum number, distribution, location, and spatial accuracy for evaluation of HA, FVA, SVA and CVA or NVA, and VVA
3. Data processing and handling
   a. Classification: classes, accuracy, and levels
   b. Derivables from the acquired data: Intensity image, DEM, DSM, nDSM, CHM, etc. with respective format and tiling requirements
   c. Hydro flattening
   d. Orthomaps (orthorectified images) in case simultaneous aerial photography is being carried out.

4. Data delivery

   a. Metadata: as per standard specifications taken up by a country or region including file naming conventions
   b. Pre-acquisition deliverables: flying and scanning parameters and flight plan (see 1(h))
   c. Post-acquisition deliverables: quality assessment reports, project extents and tiles, breaklines.
   d. Individual project metadata as per standard specifications
   e. Trajectory information
   f. LiDAR point cloud (unclassified or classified as required) and its format
   g. Classified model key points
   h. Intensity image
   i. Aerial photographs and derived products.

5 Conclusion

This paper has reviewed existing standards for airborne LiDAR data. On the basis of the existing standards the paper has arrived at a set of parameters that form a generic standard. This set can help LiDAR community in arriving at their own specifications by selecting the relevant entries as per their application. LiDAR point cloud accuracy and data density along with GSD of image are the most important parameters. Quantitative assessment of these parameters, prior to data capture, is possible for the chosen sensors and selected flight parameters. In order to ensure comprehensiveness of a standard, this quantitative assessment should also form a part of the standard, though not being done at present. As Dashora et al. [19] suggest that in case of simultaneous LiDAR and photographic data acquisition, it is possible to determine the LiDAR sensor FOV for the given camera FOV, and similarly, for given image GSD, orthoimage GSD can also be determined. So, specifications for both these aspects should also be developed. Furthermore, authors also recommend that criterions for number of checkpoints and their distribution according to size of the AOI should also be developed and included.

References

1. McCormick, M.P.: Airborne and spaceborne LiDAR. In: Weitkamp, C. (ed.) LiDAR. Springer, New York (2005)
2. Petrie, G.: Airborne topographic laser scanners. Geoinformatics **14**, 35–44 (2011)
3. USGS: 3D Elevation Program (3DEP). https://nationalmap.gov/3DEP/. Last accessed 22 Jan 2018
4. FEMA: Guidelines and Specifications for Flood Hazard Mapping Partners. http://kam-az.com/index_htm_files/FEMA%20Standards.pdf. Last accessed 03 Jan 2018

5. NOAA: Light Detection and Ranging (LiDAR) Requirements. https://www.ngs.noaa.gov/RSD/LIDAR_SOW_NGSDec2009.pdf. Last accessed 03 Jan 2018
6. ICSM: ICSM LiDAR Acquisition Specifications and Tender Template. http://www.icsm.gov.au/elevation/LiDAR_Specifications_and_Tender_Template.pdf. Last accessed 03 Jan 2018
7. GeoBC: Specifications for LiDAR. https://www.for.gov.bc.ca/ftp/hts/external/!publish/yuan/LEFI_Workshop2016/GeoBC_LiDARSpecifications_v2.0_FINAL.pdf. Last accessed 02 Jan 2018
8. Hiederman, H.K.: LiDAR Base Specification Version 1.2. https://pubs.usgs.gov/tm/11b4/pdf/tm11-B4.pdf. Last accessed 03 Jan 2018
9. LINZ: New Zealand National LiDAR Base Specifications. https://www.linz.govt.nz/system/files_force/media/doc/loci_nz-lidar-base-specification-20161220.pdf. Last accessed 04 Jan 2018
10. NRCAN: Federal Airborne LiDAR Data Acquisition Guideline version 1.0 (Draft). http://publications.gc.ca/collections/collection_2017/rncan-nrcan/M113-1-117-eng.pdf. Last accessed 03 Jan 2018
11. NDOR: NDOR Airborne LiDAR Mapping Guidelines. http://dot.nebraska.gov/media/6770/consult-lidar.pdf. Last accessed 04 Jan 2018
12. Flood, M.: Product definitions and guidelines for use in specifying LiDAR deliverables. Photogramm. Eng. Remote Sens. **68**, 1230–1234 (2002)
13. ASPRS: ASPRS LAS Format Standard 1.4 (2011). http://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf
14. ASPRS: ASPRS LAS Format Standard 1.3 (2010). http://www.asprs.org/wp-content/uploads/2010/12/LAS_1_3_r11.pdf
15. ASPRS: ASPRS LAS Format Standard 1.2 (2008). http://www.asprs.org/wp-content/uploads/2010/12/asprs_las_format_v12.pdf
16. ASPRS: ASPRS LAS Format Standard 1.1 (2005). http://www.asprs.org/wp-content/uploads/2010/12/asprs_las_format_v11.pdf
17. ASPRS: ASPRS LAS Format Standard 1.0 (2003). http://www.asprs.org/wp-content/uploads/2010/12/asprs_las_format_v10.pdf
18. ASPRS: ASPRS positional accuracy standards for digital geospatial data. Photogramm. Eng. Remote Sens. **81**, A1–A26 (2015)
19. Dashora, A., Lohani, B., Deb, K.: Two-step procedure of optimization for solving the flight planning problem for airborne LiDAR data acquisition. Int. J. Math. Model. Numer. Optim. **4**, 323–350 (2013). https://doi.org/10.1504/IJMMNO.2013.059194
20. Piel, S., Populus, J., Chust, G., Ibon, G.: Recommended Operating Guidelines (ROG) for Lidar Surveys. http://www.emodnet-seabedhabitats.eu/pdf/MeshA_ROG_Lidar_Survey_v4.0.pdf. Last accessed 14 Jan 2018
21. Caldwell, J., Agarwal, S.: LiDAR best practices. Earth Imaging J. (2013)
22. Pradhan, B., Mezaal, M.R.: Optimized rule sets for automatic landslide characteristic detection in highly vegetated forests. In: Pradhan, B. (eds) Laser Scanning Applications in Landslide Assessment. Springer (2017)
23. GeoBC: Specifications for LiDAR for Province of British Columbia. https://www2.gov.bc.ca/assets/gov/data/geographic/digital-imagery/geobc_lidar_specifications.pdf. Last accessed 02 Jan 2018
24. ASPRS: LAS File Format Exchange Activities (2017). https://www.asprs.org/committee-general/laser-las-file-format-exchange-activities.html. Last accessed 15 June 2017
25. Isenburg, M.: LASzip: Lossless Compression of LiDAR Data. https://www.cs.unc.edu/~isenburg/lastools/download/laszip.pdf. Last accessed 02 Jan 2018
26. van Maren, G.: ESRI Introduces Optimized LAS. https://blogs.esri.com/esri/arcgis/2014/01/10/esri-introduces-optimized-las/. Last accessed 02 Jan 2018
27. OSGEO: LiDAR Format Letter. http://wiki.osgeo.org/wiki/LIDAR_Format_Letter. Last accessed 02 Jan 2018

Visual Analytics of Three-Dimensional Airborne LiDAR Point Clouds in Urban Regions

J. Sreevalsan-Nair

1 Introduction

Data science workflows [10] involve its four steps, namely data preparation, analysis, reflection, and dissemination. Guo [10] has quoted [21]—"*Scientific computing is more than number crunching.*"—to elucidate how data organization is a bottleneck in performing *substantive analysis*. In this paper, we revisit the processing of airborne LiDAR (light detection and ranging), to which some of the findings seen in modern data science workflows can be applied. While analysis of LiDAR imagery has matured over time, the interest in working with geometry-aware three-dimensional (3D) point clouds is more recent [29]. The LiDAR point cloud acquisition is an outcome of advent in sensor technology in the instruments. Specifically in urban regions, extensive study is ongoing on building detection and reconstruction and road extraction from LiDAR imagery as well as point clouds. Rottensteiner has discussed how LiDAR point clouds are significant, as the 3D geometric data in combination with imagery are useful in combating issues of occlusions, shadows, and nondetection of building smaller than 30 m^2.

Given this premise, we discuss how we have incorporated visualization in the data science workflow to process airborne LiDAR point clouds in urban regions. In two different methodologies proposed by our research group in [19] and in [33, 34], visual analytics is the key aspect of the data science workflow. Keim et al. [13] have defined *visual analytics* as an intermix of conventional data mining and interactive visualizations in a data science/analytic workflow, which unifies sense-making, inferential understanding, and decision-making support for big data. *Big data* itself is characterized by the five V's—volume, variety, veracity, velocity, and value.

Conventionally, processing of airborne LiDAR point clouds starts with identifying local neighborhood for each point. This is followed by the computation of covariance

J. Sreevalsan-Nair (✉)
Graphics-Visualization-Computing Lab, International Institute of Information Technology Bangalore (IIIT-B), Bengaluru 560100, India
e-mail: jnair@iiitb.ac.in

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_23

matrix $C(p) = \sum_{y \in N(p)} (y-p)^T . (y-p)$, at each point $p \in \mathbb{R}^3$. The covariance matrix is referred to as the *structure tensor*, but it is broadly a local geometric descriptor of the point. Most of the data mining for classification include semantic classification, where features at each point are extracted and used in supervising learning techniques, such as conditional random forest classifier and support vector machines. In this background, we have attempted to answer two questions

1. How *can* we classify points if we do not have training dataset to execute supervised or semisupervised learning algorithms?
2. How *good* is the geometric description provided by the local geometric descriptor?

In order to answer these questions, we have used visual analytics in combination with appropriate data modeling. Kumari and Sreevalsan-Nair [19] have used hierarchical (divisive) clustering of the point cloud to answer the issue on unsupervised classification technique. Sreevalsan-Nair and Kumari [34] have proposed a novel local geometric descriptor for LiDAR point clouds using tensor voting [11]. They have proposed processing local geometric descriptors in its form of positive semidefinite symmetric second-order tensors. Sreevalsan-Nair and Jindal [33] have further improved on the novel descriptor using gradient information. Tensor voting is a voting scheme [11] for detection and classification of feature points using proximity and continuity principles of Gestalt psychology for vote propagation.

2 Literature Survey

In this section, we describe the relevant literature on the key topics of semantic classification as well as local geometric descriptors. Semantic classification of LiDAR data has been widely studied. We discuss some of the work which is relevant to the visual analytic framework, proposed in [19]. Further, we describe relevant work on local geometric descriptors, its tensor representations, and its uses in LiDAR research community.

Semantic (or Object-Based or Contextual) Point Classification: Song et al. [32] have determined the effectiveness of using LiDAR intensity data for land cover classification, where a uniform grid derived from point cloud is used. Chehata et al. [2] have given a classification of parameters used for semantic classification and results from using multiple classifiers using random forest classifiers. Niemeyer et al. [22] have performed supervised classification using conditional random fields (CRFs), using geometrical features and an intensity value. These results have been improved by using random forests with the CRFs in [23]. Niemeyer et al. [24] have proposed inclusion of context of spatial locality, as an additional cue to the supervised classification.

Other Similar Methodologies: Ramiya et al. [28] have used curvature and colorimetric distance for segmenting colored LiDAR data. Other unsupervised classification techniques, such as [8, 20], exist, which use density-based clustering and graph cut-based methods. We have found applications of interactive agglomerative clustering, which is bottom-up approach of merging clusters. Preiner et al. [27] have used hierarchical agglomerative expectation maximization (EM) clustering for surface reconstruction from point cloud data, unlike the divisive clustering method used in [19].

Covariance Analysis: Covariance analysis of local neighborhoods based on centroid is a robust method for normal estimation [17], but not necessarily for finding the shape of the neighborhoods. Tombari et al. [35] have made the argument of lack of repeatability of sign of a local reference frame when using the covariance matrix and have proposed a weighted covariance matrix based on the point itself instead of the centroid, for surface matching. Local tensor-based techniques are a trade-off between computational complexity and accuracy in feature detection; e.g., use of tensor voting [25, 36] for feature classification.

Tensor Representation of Local Geometric Descriptor: Knutsson [15] has defined a structure tensor as a tensor computed using differential calculus of functions. Structure tensor has been used as a descriptor in 3D space. Knutsson et al. [16] have discussed different descriptors used for images, with a potential for extension to 3D point clouds.

Structural Classification: Structural (or geometric or feature) classification of point clouds has been less explored [25, 37]. Structural classification is implicitly used in semantic classification through the use of eigenvalue-based features obtained from the local geometric descriptor, such as covariance matrix in LiDAR point clouds [5, 19].

Multiscale Classification: Pauly et al. [26] have proposed the use of multiscale surface variation, estimated using covariance matrix of local neighborhood. There, surface variation at a user-defined scale gives feature weights, which upon appropriate thresholding gives features. Keller et al. [14] have used a similar multiscale approach, for LiDAR point clouds, in determining feature weights from covariance matrix of local neighborhoods. However, the difference between the methods in [14, 26] is that a single adaptive scale and averages across multiple scales have been used, respectively. Algorithms for finding optimal neighborhood size or scale have been of interest to the LiDAR community [5, 9, 41]. Blomley et al. [1] have used multiscale approach using shape distribution features for point classification, as opposed to covariance features, proposed by Keller et al. [14]. Park et al. [25] have used tensor voting and surface variation to classify and detect line features in point clouds, where the surface variation function is computed using a multiscale method. An approach based on anisotropic diffusion is used in [36], where anisotropic diffusion is performed after tensor voting for feature classification and extraction in polygonal mesh data, and subsequent mesh segmentation.

3 Augmented Semantic Classification

Semantic classification of LiDAR point clouds has been extensively studied. As discussed in Sect. 2, many of the recent methods rely on supervised learning methods [24, 38, 39]. Supervised learning methods require training datasets, which are conventionally generated by a domain expert, data from the site measured using complementary processes, or other publicly available alternative (albeit partial) sources, e.g., OpenStreetMap [12] and Google Earth.[1] In order to have a facility to visually explore new airborne LiDAR datasets with a preliminary semantic classification, Kumari and Sreevalsan-Nair [19] have proposed a visual analytics framework using unsupervised clustering for semantic classification. The novelties of the method proposed in [19] are: (a) the concept of *augmented semantic classification*, (b) interactive setting of parameters in the visualization of the tree data structure, representing hierarchical clustering method, and (c) the use of visualizations to guide *choosing* the features for clustering at each level of hierarchy. They have demonstrated a prototypical tool implementing the proposed visual analytics method, referred to as the *tree visualizer* (Fig. 1).[2] The experiments have been done on the Vaihingen dataset [3], provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF).[3]

Unsupervised Semantic Classification Using Tree Visualizer: With the recent success seen with supervised learning, one cannot discount the value of such a class of methods for semantic classification. However, as discussed earlier, the state-of-the-art methods develop training models separately for different datasets. The reason for this could be that this work is fairly recent, and additionally, it could take several experiments to derive a generic training model(s) owing to inherent differences in built environments worldwide. Hence, for first-cut and quick exploration of new datasets, Kumari and Sreevalsan-Nair [19] have proposed the use of unsupervised methods, in an adaptive manner to accommodate the high variability in environmental data in airborne LiDAR point clouds. The idea is to provide several iterations of classifications, and adaptiveness of the method is provided by the choice of parameters or features to be used for an iteration of classification in different regions. Given these requirements, hierarchical (divisive) clustering is the most appropriate unsupervised method that can be used for such an unsupervised classification problem.

Kumari and Sreevalsan-Nair have narrowed the use of hierarchical clustering specifically to *hierarchical EM clustering*. In this proposed method, the classification is agnostic of the spatial locality. Hence, the spatial locality information is eventually added in a postprocessing step. This step involves a region growing algorithm [30], which is used for correcting the labels of points, based on the majority *vote* of the label of the points in its local neighborhood.

---

[1] Google Earth at https://www.google.com/earth/. Last retrieved on January 09, 2018.

[2] The color version of the figure is available in the online version of this paper as well as in [19].

[3] DGPF at http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html.

Choice of Feature Vector for Hierarchical Clustering: The supervised learning methods for semantic classification have shown the set of features that are required in the feature vector [2]. Kumari and Sreevalsan-Nair have chosen the commonly used parameters or features, such as height, intensity, height variance, and height range. It also includes a set of features which are derived from the covariance matrix of the local neighborhood of each point. The covariance matrix is the local geometric descriptor [22], which we discuss further in more detail in Sect. 4. The colormaps or heat maps of the different features are then used for providing colors to the points with corresponding feature values. These visualizations allow the user to make decisions on which parameter, according to their visual perception, gives the best binary *clustering* of the concerned subset of points. The process entails the entire point cloud being clustered in the leaf nodes of the clustering hierarchy. Since the number of leaf nodes in a binary tree, which need not be balanced, exceeds the number of semantic classes, some of the clusters are appropriately merged to give the exact number of semantic classes.

The Proposed Augmented Semantic Classification: There are two types of classifications possible for the airborne LiDAR cloud, namely geometric (or structural) and object-based (or semantic). Semantic classification [24, 38, 39] is widely studied than structural classification [14, 18]. Structural classification labels the points as belonging to line-, surface-, or (critical) point-type features. The latter includes features like junctions. Conventionally, structural classification is obtained using the eigenvalue analysis of the local geometric descriptors. Thus, structural classification is used as an intermediate step in semantic classification, e.g., linear and areal anisotropies, which are indicative of structural classification. Kumari and Sreevalsan-Nair [19] have proposed to preserve the structural classification by introducing a tuple of labels for each point in the point cloud, instead of a singleton label. The tuple of labels includes both structural and semantic labels. The use of tuple of labels for each point in the point cloud has been called *augmented semantic classification*. The benefit of the augmented classification [19] has been in the improved rendering, as shown in Fig. 1, where the line-type features make the rendering of the point cloud sharper.

Results: A domain expert user study was done on the results of unsupervised semantic classification of Vaihingen benchmark dataset, in the absence of ground truth [19]. Visually, the domain expert has determined an overall accuracy of 80–85% in the classification. The point rendering of the LiDAR dataset using color scheme based on augmented semantic classification is visibly sharper than that of semantic classification, as shown in Fig. 1.

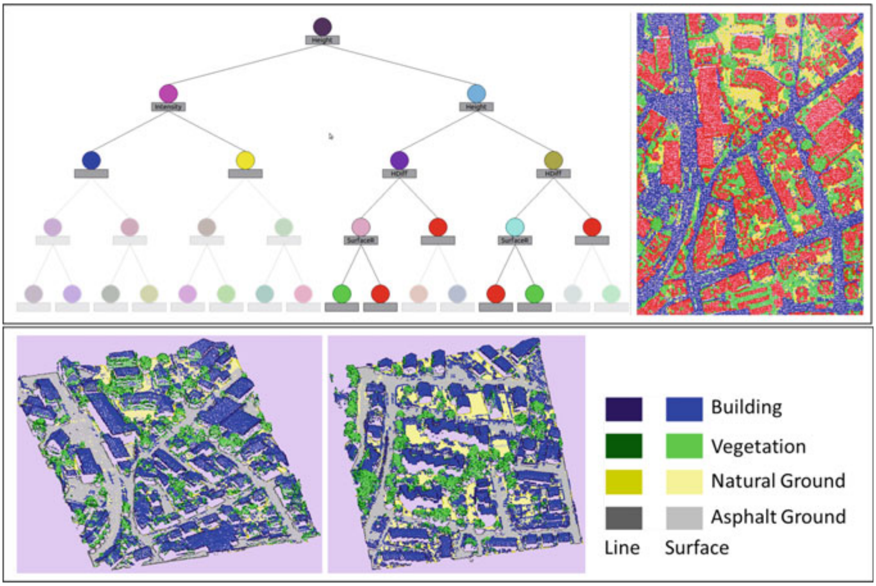


Fig. 1 (Top) The graphical representation of hierarchical divisive clustering of airborne LiDAR point clouds is based on its tree data structure (top left). The active leaf nodes of the tree, which are not dimmed/transparent, belong to four different object classes. The points of a class can belong to multiple leaf nodes, e.g., red nodes correspond to building. The point rendering shows the state of the active leaf nodes of the tree (top right). (Bottom) The results of augmented semantic classification of Area-1 (bottom left) and Area-2 (bottom middle) of Vaihingen dataset, as per the legend showing tuple of labels in a matrix (bottom right). The bottom image has been modified from an image in [19]

4 Local Geometric Descriptors

The decisive role of local geometric descriptors in both structural as well as semantic classifications begs for extensive research on them. This area of research has been underrepresented in the LiDAR community, unlike the computer graphics or geometric modeling communities [7]. Hence, we showcase work in deriving new local geometric descriptors in airborne LiDAR point clouds and its comparative analysis in [33, 34]. While covariance matrix is used ubiquitously for classification, it has been found to not identify the sharp line-type features, e.g., gable line in the roofs. Hence, the motivation is to identify a local geometric descriptor that behaves similar to the covariance matrix C and, at the same time, highlights sharp features.

Definition: A *local geometric descriptor* of a point, p, is defined as the data entity that captures the shape of the local neighborhood of a point p, $\mathcal{N}(p)$. This is a descriptor, because it defines the type of geometric feature to which the point itself belongs. Thus, a point with cylindrical neighborhood belongs to a line-type feature with disk-shaped neighborhood, to a surface-type feature and with spherical neighborhood, to a critical point-type feature [14]. This shows us how local geometric descriptors are significant for structural classification. In order to determine the shape of the local neighborhood, one has to find the eigenvalue decomposition of the descriptor. Thus, the eigenvalues of the local geometric descriptor are required for both structural as well as semantic classifications. The local neighborhood itself can be defined either in terms of points within a specific Euclidean distance, which gives a spherical neighborhood or in terms of the number of nearest points, e.g., k-nearest neighbors, for $k \in \mathbb{Z}_{>0}$.

Tensor Voting for Local Geometric Descriptor: Inspired by the work of Wang et al. [36] on the use of tensor voting [11] for extracting sharp (line) features in triangular meshes, Sreevalsan-Nair and Kumari [34] have used tensor voting to define the local geometric descriptor for LiDAR point clouds. Comparing the tensor voting descriptor, V, computed for unoriented points as discussed in [25], with the conventionally used covariance matrix, C, it has been determined that C and V are both positive semidefinite and symmetric second-order tensors. Hence, this property of same type makes the local geometric descriptors comparable and substitutable. However, it must be noted that, that even though the two descriptors are of the *same type*, there are differences based on their provenance. One of the noteworthy differences between C and V is that they are generated in the tangent and normal space, respectively.

Visual Analytics: Visualization has been used for qualitatively comparing different descriptors in [34]. Two visual encodings or channels are required for encoding the eigenvalues and eigenvectors, namely color and geometry. The eigenvalues give the structural classification, and the eigenvectors give the orientation of the tensor, which is a representative of the local geometric descriptor. The chosen colormap is based on the saliency maps of the point. The saliency map gives the likelihood of the point to belong to any of the three structural classes. When considering geometry channel, both point rendering (Fig. 2)[4], as well as superquadric tensor glyphs [31], are applicable. The shape of the superquadric tensor glyphs and the color encoding based on saliency map redundantly give the perception of the structural classification of the points [34]. The redundancy is required for confirming cognition, in cases of disambiguations in perception.

Anisotropic Diffusion and Local Reference Frame Alignment: The qualitative comparison of the two local geometric descriptors, C and V, has revealed that the structural classification given by them are not equivalent. This difference has been attributed to the generation of C in the tangent space as opposed to V in normal space. At the same time, we see that Wang et al. have used *anisotropic diffusion*, using Gaussian weights and reciprocal to *flip* the sorted order of eigenvalues. Thus, we see that anisotropic diffusion performs two actions: (a) "corrects" the structural classification given by V to be similar to C and (b) slows down diffusion across sharp features and speeds it up along sharp features. The overall outcome of the

[4]The color version of the figure is available in the online version of this paper as well as in [34].

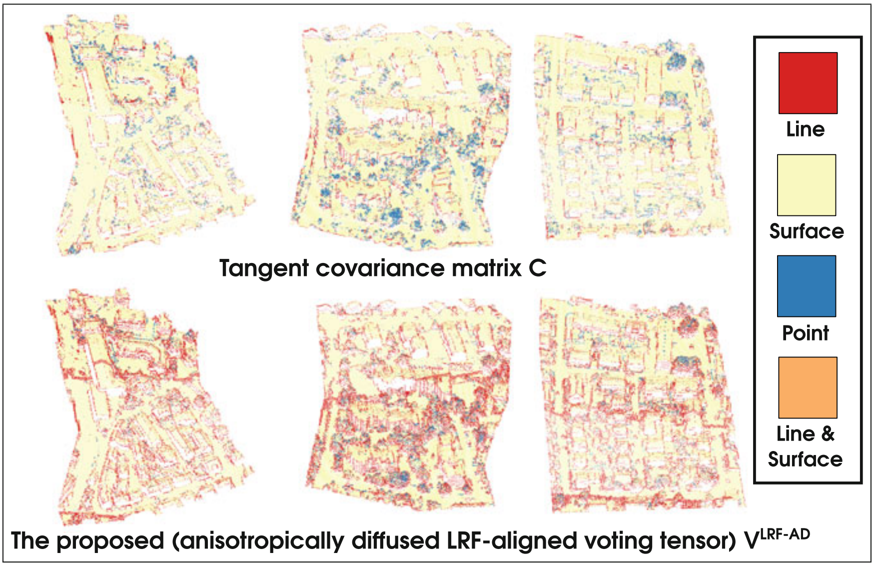


Fig. 2 The point rendering of (left-to-right) Area-1, Area-2, and Area-3 of Vaihingen benchmark dataset shows differences in structural classification based on the use of conventional (tangent) covariance matrix (top) with respect to the anisotropically diffused tensor voting-based local geometric descriptor (bottom). This image is a modified version of an image in [34]

anisotropically diffused tensor voting-based local geometric descriptor, V_{AD}, matches with that of C.

Multiscale Aggregation: The LiDAR point cloud analysis is generally performed on multiple scales. Multiple scales are assigned based on the definition of the local neighborhood. There are two ways of combining the analysis across multiple scales. One is to use optimal scale, based on optimizing a property, e.g., entropy [40] or aggregation using saliency maps [14].

Further, Sreevalsan-Nair and Kumari have found that the orientations of the local geometric descriptors based on their eigenvectors need to be aligned in order to compare the descriptors effectively. Hence, they align the eigenvectors, thus, aligning the local reference frame, for comparing two different local geometric descriptors. The descriptors with local reference frames aligned are referred to as C^{LRF} and V^{LRF-AD}, respectively. Sreevalsan-Nair and Kumari have observed that the use of V^{LRF-AD} enables identification of gable roofs, by strengthening line-type features which are found to be weak, when using covariance matrix. However, this method reduces the number of points belonging to critical point-type features. The presence of both line- and point-type features are essential for effective geometric reconstruction.

Improvement using Gradient Energy Tensor: To improve the identification of point-type features, Sreevalsan-Nair and Jindal [33] have proposed the use of gradients. The gradients, from first- to third-order derivatives, are computed for each point to give the gradient energy tensor (GET) [6]. The use of GET has been proposed owing to two reasons: (a) it detects points of interest, which are essentially junction points, and (b) in order to find derivatives, geometry of the point cloud must be described as a function or a map. This is possible for LiDAR point clouds, as height is a function of (x, y), thus exploiting the 2.5-dimensional nature of the point cloud data.

The usage of GET tensor, however, requires postprocessing, as it is a 2D tensor (which need not exist as positive semidefinite second-order tensor but can be used as positive semidefinite second-order tensor), which is to be used with a 3D tensor, V^{LRF-AD}. Exclusively for the points of interest detected by the 2D tensor, a mapping of the eigenvalues and eigenvectors between the 2D and 3D tensors is performed to complete the correction. Testing this new tensor on the Vaihingen dataset shows improved results in detecting point-type features.

Results: For Area-1 of Vaihingen dataset, the local geometric descriptors classify the following percentages of points as (line, surface, point) type features: covariance matrix C gives (24, 60, 16%); the proposed anisotropically diffused tensor voting-based local geometric descriptor without GET correction gives (29, 65, 6%) and with GET correction gives (27, 63, 10%). We observe that, as per design, V^{LRF-AD} identifies more line-type features than conventionally used C, as shown by the higher red tinge in datasets with V^{LRF-AD} in Fig. 2. We also see that the point-type features reduce drastically. Further, correction using GET tensor improves the percentage of point-type features.

5 The Role of Visual Analytics in Point Cloud Processing

In Sects. 3 and 4, we have demonstrated two different applications of visual analytics in processing airborne LiDAR point clouds. One has been in using visualizations to guide unsupervised (semantic) classification [19], and the other has been in using the same in comparing local geometric descriptors [33, 34], which can potentially impact both structural, and, subsequently, semantic classifications. These works go to show that data analytics can be conducted in an iterative and interactive manner, by introducing the feedback loop that is present in visual analytics workflow. These works also go on to show how using visual analytics can bring further insights into the dataset. Using visualization for exploration allows the viewer to choose appropriate analytic processes for making sense of the data. A visual analytic framework as proposed in [19] allows the user to flexibly move across different actions in a data science workflow.

These works have the potential to drive further research. The tree visualizer-based unsupervised classification is yet to be analyzed in terms of accuracy as well as adaptability. A richer analysis may be done based on cluster shapes which the human-in-the-loop can perceive before choosing the parameters for clustering. Similarly, the use of new local geometric descriptors, such as the one based on tensor voting, is yet to be tested for both classification as well as geometric reconstruction.

LiDAR point clouds can be considered as **spatial big data** [4] given the *volume* and *variety* in the data. In a specific instance, Cugler et al. have discussed how spatial big data may use computational models to generate hypotheses, which can be further used for answering questions for which the hypothesis applies. In [19, 33, 34], we observe that different visualizations have been used to create hypotheses for questions on which features reduce the uncertainty of semantic classification of points, can intensity be used as a reliable feature for classification in a given dataset, and which points are the best to be used for extracting footprints of buildings. While the results of visualization are qualitative, the outcomes become quantitative as we move through the data science workflow. Nonetheless, the quality of the outcomes is limited by the the ability of the viewer to make *relevant* sense of the data.

In conventional LiDAR point cloud processing, visualizations are limited to rendering points based on height or semantic class. In this paper, we have discussed two different applications where visualizations are used for visual exploration of datasets. This is particularly useful in order to use the knowledge discovered by perception for further analysis of the data. This is a thread which can be further pursued for research on what are different data models,[5] like the local geometric descriptors, which can have two-fold benefits, namely: (a) the data model influences appropriate visual representation of the dataset and (b) the data model can give insights of the data or enable big data analytics, in its own right. Appropriate choice of data structures helps in determining apt data models. An example of a data model is the feature graph by Keller et al. [14], which was instrumental in reducing the point cloud without losing significant features. Representing each point in the point cloud as a second-order tensor using local geometric descriptor is another example of data model.

In order to tackle the aspect of *volume* of the LiDAR point clouds, spatial data structures, such as octree and k-d tree, are used for storing the data and for speeding up search operations to compute local neighborhood in [19, 33, 34]. In addition to this, the tree visualizer in [19] iteratively subdivide the point cloud as the hierarchical EM clustering is executed at each node of the clustering tree. Thus, the leaf nodes of the tree visualizer correspond to subset of points with a specific label, and by design, a point belonging to a leaf node belongs to its parents up all the levels till the root node. The user can choose an action of binary clustering at a leaf node, using the user-defined parameter(s). Thus, the tree visualizer is designed for the user to interactively change the choice of feature/parameter for clustering, but this action does not process the entire point cloud unless the root node is activated for modification. By design, tree visualizer implements a *divide and conquer* approach.

[5] By data model, we refer to appropriate representations of the data, in its entirety or partial, where the representation does not distort the overall understanding of the data.

6 Conclusions

In this paper, we have summarized the combined significance of the works done in our research group in [19, 33, 34] in visual analytics of airborne LiDAR point clouds. We have discussed how visualizations can guide decision-making for improving unsupervised classification in [19]. The new local geometric descriptor is introduced in [34] using tensor voting, and its improvement using gradients [33] shows how visualization can enable the *substitutability* of local geometric descriptors. Local geometric descriptors are conventionally used for both structural as well as semantic classifications. We have shown how the research on processing point clouds acquired by airborne LiDAR is a topic considered in spatial big data analytics. We have briefly described how visualizations can be used beyond summarizations to make sense of the data. This is just the tip of the iceberg on how the use of visualizations of the 3D LiDAR data can usher in multidisciplinary approaches for data analytics.

Acknowledgements This work has been partially supported by NRDMS programme as well as Early Career Research Award to the author by the Science and Engineering Research Board, where both are from the Department of Science and Technology, Government of India. Different portions of this work have been conducted as part of the Masters study of Beena Kumari and Akshay Jindal at IIIT-B.

References

1. Blomley, R., Weinmann, M., Leitloff, J., Jutzi, B.: Shape distribution features for point cloud analysis-a geometric histogram approach on multiple scales. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **1**, 9–16 (2014)
2. Chehata, N., Guo, L., Mallet, C.: Airborne lidar feature selection for urban classification using random forests. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **38**(Part 3), W8 (2009)
3. Cramer, M.: The DGPF-test on digital airborne camera evaluation–overview and test design. Photogramm. Fernerkund. Geoinf. **2010**(2), 73–82 (2010)
4. Cugler, D.C., Oliver, D., Evans, M.R., Shekhar, S., Medeiros, C.B.: Spatial big data: platforms, analytics, and science. GeoJournal (2013)
5. Demantké, J., Mallet, C., David, N., Vallet, B.: Dimensionality based Scale selection in 3D LiDAR point clouds. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **38**(Part 5), W12 (2011)
6. Felsberg, M., Köthe, U.: GET: the connection between monogenic scale-space and gaussian derivatives. In: International Conference on Scale-Space Theories in Computer Vision, pp. 192–203. Springer (2005)
7. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: Proceedings of the Third Eurographics Symposium on Geometry Processing, SGP '05. Eurographics, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland. http://dl.acm.org/citation.cfm?id=1281920.1281953 (2005)
8. Ghosh, S., Lohani, B.: Heuristical feature extraction from LiDAR data and their visualization. In: Proceedings of the ISPRS Workshop on Laser Scanning 2011, vol. 38 (2011)
9. Gressin, A., Mallet, C., Demantké, J., David, N.: Towards 3D LiDAR point cloud registration improvement using optimal neighborhood knowledge. ISPRS J. Photogramm. Remote Sens. **79**, 240–251 (2013)

10. Guo, P.: Data science workflow: overview and challenges. blog@ CACM. Commun. ACM (2013)
11. Guy, G., Medioni, G.: Inference of surfaces, 3D curves, and junctions from sparse, noisy, 3D data. Pattern Anal. Mach. Intell. IEEE Trans. **19**(11), 1265–1277 (1997)
12. Haklay, M., Weber, P.: Openstreetmap: user-generated street maps. IEEE Pervasive Comput. **7**(4), 12–18 (2008)
13. Keim, D., Andrienko, G., Fekete, J.D., Gorg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. Lecture Notes Computer Science, vol. 4950, pp. 154–176 (2008)
14. Keller, P., Kreylos, O., Vanco, M., Hering-Bertram, M., Cowgill, E.S., Kellogg, L.H., Hamann, B., Hagen, H.: Extracting and visualizing structural features in environmental point cloud LiDaR data sets. In: Topological Methods in Data Analysis and Visualization, pp. 179–192. Springer (2011)
15. Knutsson, H.: Representing local structure using tensors. In: 6th Scandinavian Conference on Image Analysis, Oulu, Finland, pp. 244–251. Linköping University Electronic Press (1989)
16. Knutsson, H., Westin, C.F., Andersson, M.: Representing local structure using tensors II. In: Image Analysis, pp. 545–556. Springer (2011)
17. Kobbelt, L., Botsch, M.: A survey of point-based techniques in computer graphics. Comput. Graphics **28**(6), 801–814 (2004)
18. Kumari, B., Ashe, A., Sreevalsan-Nair, J.: Remote interactive visualization of parallel implementation of structural feature extraction of three-dimensional lidar point cloud. In: Big Data Analytics, pp. 129–132. Springer (2014)
19. Kumari, B., Sreevalsan-Nair, J.: An interactive visual analytic tool for semantic classification of 3D urban lidar point cloud. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, p. 73. ACM (2015)
20. Lafarge, F., Mallet, C.: Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. Int. J. Comput. Vis. **99**(1), 69–85 (2012)
21. Langtangen, H.P.: Python Scripting for Computational Science, vol. 3. Springer (2006)
22. Niemeyer, J., Rottensteiner, F., Soergel, U.: Conditional random fields for lidar point cloud classification in complex urban areas. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **1**(3), 263–268 (2012)
23. Niemeyer, J., Rottensteiner, F., Soergel, U.: Classification of urban LiDAR data using conditional random field and random forests. In: Urban Remote Sensing Event (JURSE), Joint, pp. 139–142 (2013)
24. Niemeyer, J., Rottensteiner, F., Soergel, U.: Contextual classification of lidar data and building object detection in urban areas. ISPRS J. Photogramm. Remote Sens. **87**, 152–165 (2014)
25. Park, M.K., Lee, S.J., Lee, K.H.: Multi-scale Tensor voting for feature extraction from unstructured point clouds. Graph. Models **74**(4), 197–208 (2012)
26. Pauly, M., Keiser, R., Gross, M.: Multi-scale feature extraction on point-sampled surfaces. Comput. Graphics Forum **22**(3), 281–289 (2003)
27. Preiner, R., Mattausch, O., Arikan, M., Pajarola, R., Wimmer, M.: Continuous projection for fast l-1 reconstruction. ACM Trans. Graphics (TOG) **33**(4), 47 (2014)
28. Ramiya, A., Nidamanuri, R.R., Krishnan, R.: Semantic labelling of urban point cloud data. ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **1**, 907–911 (2014)
29. Rottensteiner, F.: Status and further prospects of object extraction from image and laser data. In: 2009 Joint Urban Remote Sensing Event, pp. 1–10. IEEE (2009)
30. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China (2011)
31. Schultz, T., Kindlmann, G.L.: Superquadric glyphs for symmetric second-order tensors. IEEE Trans. Vis. Comput. Graph. **16**(6), 1595–1604 (2010)
32. Song, J.H., Han, S.H., Yu, K., Kim, Y.I.: Assessing the possibility of land-cover classification using lidar intensity data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **34**(3/B), 259–262 (2002)

33. Sreevalsan-Nair, J., Jindal, A.: Using gradients and tensor voting in 3D local geometric descriptors for feature detection in airborne LiDAR points in urban regions. In: Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1–4 (2017) (to appear). http://www.iiitb.ac.in/GVCL/pubs.html
34. Sreevalsan-Nair, J., Kumari, B.: Local Geometric Descriptors for Multi-scale Probabilistic Point Classification of Airborne LiDAR Point Clouds. Mathematics and Visualization, pp. 175–200. Springer, Cham (2017)
35. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: Computer Vision–ECCV 2010, pp. 356–369. Springer (2010)
36. Wang, S., Hou, T., Li, S., Su, Z., Qin, H.: Anisotropic elliptic PDEs for feature classification. Vis. Comput. Graph. IEEE Trans. **19**(10), 1606–1618 (2013)
37. Weber, C., Hahmann, S., Hagen, H.: Methods for feature detection in point clouds. In: OASIcs-OpenAccess Series in Informatics, vol. 19. Schlöss Dagstuhl-Leibniz-Zentrum fuer Informatik (2011)
38. Weinmann, M., Jutzi, B., Hinz, S., Mallet, C.: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogramm. Remote Sens. **105**, 286–304 (2015)
39. Weinmann, M., Jutzi, B., Mallet, C.: Feature relevance assessment for the semantic interpretation of 3D point cloud data. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **5**, W2 (2013)
40. Weinmann, M., Jutzi, B., Mallet, C.: Semantic 3D scene interpretation: a framework combining optimal neighborhood size selection with relevant features. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. **2**(3), 181 (2014)
41. Yang, B., Dong, Z.: A shape-based segmentation method for mobile laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. **81**, 19–30 (2013)

Part III
Geo-spatial Applications

Development of GIS-Based Noise Simulation Module for Point Sources: N-Dhwani

R. Vijay, D. Lingote, A. Sharma and R. Gupta

1 Introduction

Most of the available noise models and software are platform dependent and are not open geospatial consortia (OGC) compliant. They are integrated with geographical information system (GIS) but are not fully developed on GIS platform. Further, they do not include cumulative impact of point source (single or multiple) and background noises. Based on these, the objective of the present study is to develop a user-friendly OGC compliant GIS-based noise simulation software for points sources (N-Dhwani) for single and multiple point sources, which acts as a decision-making tool for better planning and management of noise environment. N-Dhwani serves as a pre- and post-processor for noise simulation and attenuation in the region based on noise sources, geographical information, meteorological parameters and position of receptors [1].

2 Development of N-Dhwani

GIS-based noise simulation software for point source [N-Dhwani] has been designed and developed using open-source tool and technology. For development of N-

R. Vijay (✉)
Centre for Strategic Urban Management, CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur 440020, India
e-mail: r_vijay@neeri.res.in

R. Gupta
Civil Engineering Department, Visvesvaraya National Institute of Technology, Nagpur 440001, India

D. Lingote · A. Sharma
CTM Division, CSIR-National Environmental Engineering Research Institute, Nehru Marg, Nagpur 440020, India

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_24

Dhwani, quantum GIS (Q-GIS), an open-source software and 'python' an open-source programming language were used. This is a user-friendly, stand-alone application, platform independent and complying with the open geospatial consortia (OGC) standards. The modules of N-Dhwani are defined as the following.

2.1 Point Module

Point module simulates the impact of industrial and construction activities in terms of fixed noise sources in the vicinity. For this, inputs required are location of source and receptor, position of barriers and background noise at receptor end. The module calculates the noise using the following equation [2].

$$L_p = L_w - 20 \times \log(d) - A_e - 8 \tag{2.1}$$

where L_p is sound level at receptor in dB(A), L_w is sound level of point source in dB(A), d is distance between source and receptors in m, and A_e is attenuation in dB(A).

2.2 Attenuation Module

There are many factors responsible for attenuation of noise at receptor end. These are termed as attenuation due to atmosphere (A_{abs}), vegetation ($A_{vegetation}$) and physical barriers ($A_{barrier}$) [3]. This module considers the total attenuation as per the following equation.

$$A_e = A_{abs} + A_{vegetation} + A_{barrier} \tag{2.2}$$

2.2.1 Atmospheric Attenuation

The mechanism for atmospheric absorption has been extensively studied, empirically quantified and codified into an international standard for calculation [4, 5]. For a standard pressure of one atmosphere, the absorption coefficient α (in dB/100 m) can be calculated as a function of frequency f (Hz), temperature T (degrees Kelvin) and molar concentration of water vapour h (%).

2.2.2 Vegetation Attenuation

Vegetation and foliage provide a small amount of attenuation, but only if it is sufficiently dense to fully block the view along the propagation path between source and receptor. The attenuation may be due to vegetation close to the source, close to the receptor or both [6].

2.2.3 Barrier Attenuation

When the line of sight between a source and receptor is obstructed by a rigid, non-porous wall or building, appreciable noise reductions can occur. Sound waves must diffract around the obstacle in order to reach the receptor. This phenomenon is a great advantage used in the attenuation of highway noise by barriers in congested urban areas [7]. An observer in the vicinity of a rigid, infinitely long barrier, for sound from a point source, will experience an excess attenuation as developed by Maekawa in 1968 [8].

2.3 Cumulative Module

Cumulative module helps in simulation of combined effect of point source and background noise at the receptor level. This module is important where impact of background noise is significant and it is difficult to assess the impact at receptor end.

2.4 Help Module

The information about the different modules in the software is defined and discussed in the help module. The details about every single entity of each module can be fetched through contents, index and search functions of the module. This help file is generally useful for the end users, who will be using this software.

3 Description of N-Dhwani

N-Dhwani has been developed considering point sources of noise, receptor positions as per user choice or uniformly distributed in the study area, impact of meteorological parameters like temperature and humidity and geographical features like building, barrier and vegetative cover. Attenuation due to atmosphere, barrier and vegetation is also considered in the software. The main screen of N-Dhwani is shown in Fig. 1 with modules of 'point', 'barrier width and height analysis', 'cumulative' module for impact assessment due to combined effect of point and mobile sources, 'split line', draw perpendicular' and 'help' modules. The main screen consists of 'layer' panel on left side, 'map window' in centre and 'excess attenuation' on right side of the main software window.

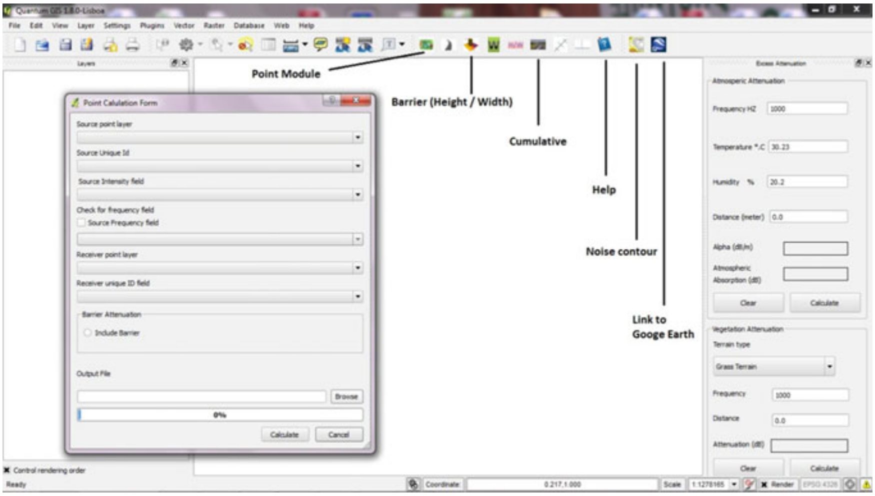


Fig. 1 Input requirement of N-Dhwani (point module)

4 Application of N-Dhwani

Noise due to construction activity is simulated in the premises of NEERI, Nagpur, where building construction work was under progress, using modules of N-Dhwani considering point module, barrier module and background noise at receptor levels. Continuous operation of concrete mixer pouring machine is considered as a noise source in the vicinity, and noise levels are simulated at the receptor end under scenarios without and with barrier/buildings.

4.1 Scenario 1—Without Barrier/Building

The simulation scenario of noise due to construction is carried out without barrier as explained in Fig. 2. S1 is the noise source with sound intensity of 90 dB(A) attributed in the source file, and R1 to R8 are the receptor locations. Background noise is considered in the simulation. Monitored background noise levels are attributed to receptor in the GIS database of receptor file. For distance between source and receptors, distance matrix algorithm is used to calculate the distance for attenuation due to distance. Based on the algorithm, calculated distance between source and receptor is shown in Fig. 2.

Based on the point model equation as described in Sect. 4.2, it computes the noise levels at the receptors considering source and distance between source and receptor (Fig. 3). Impact of construction activity can be noticed at receptor location R3 due

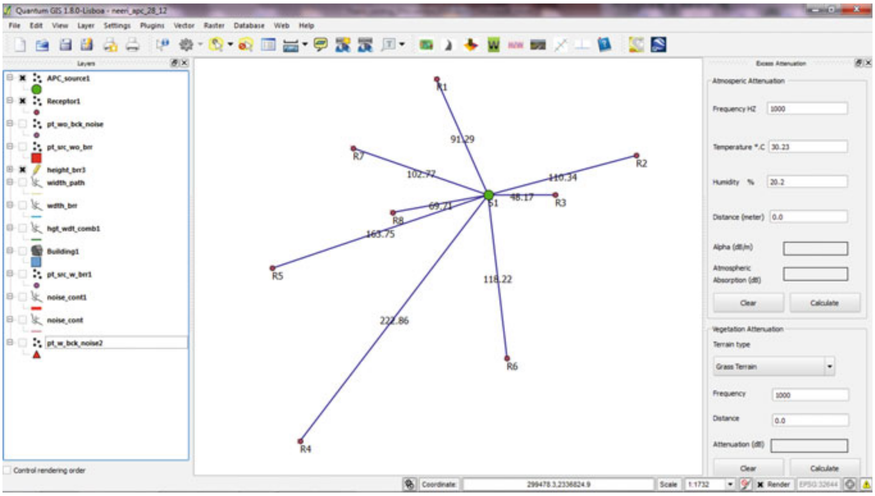


Fig. 2 Location and distance calculation between source and receptors

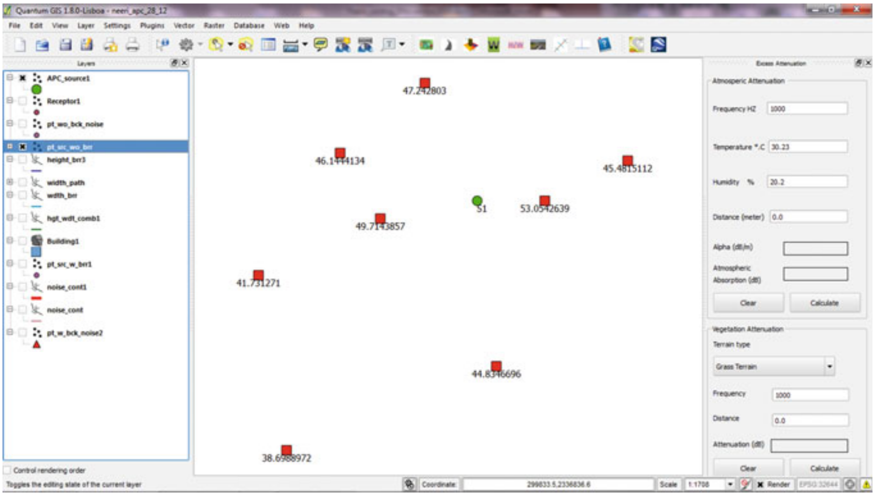


Fig. 3 Noise simulation due to construction activity at receptors considering without barrier

to its closest proximity to maximum noise level [53 dB(A)]. The minimum impact is observed at R4 with noise level 38.7 dB(A).

A difference can be seen in the simulated noise levels at receptor end considering without and with background noise as presented graphically in Fig. 4. Impact of construction activity and effect of background noise are significant at S3 and S8 as compared to other locations. While at other receptor locations, background noise is

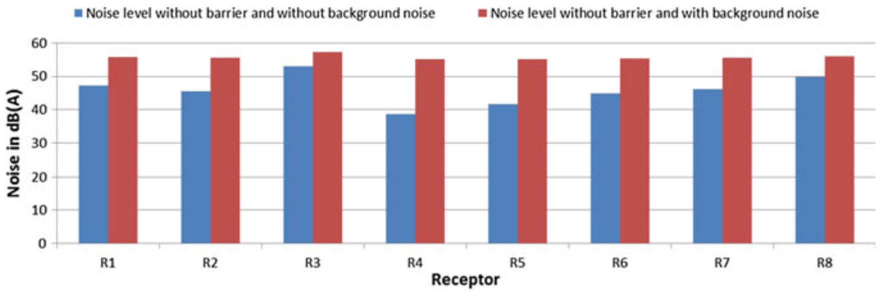


Fig. 4 Comparison of noise level at receptor due to point source and without barrier considering without and with background noise

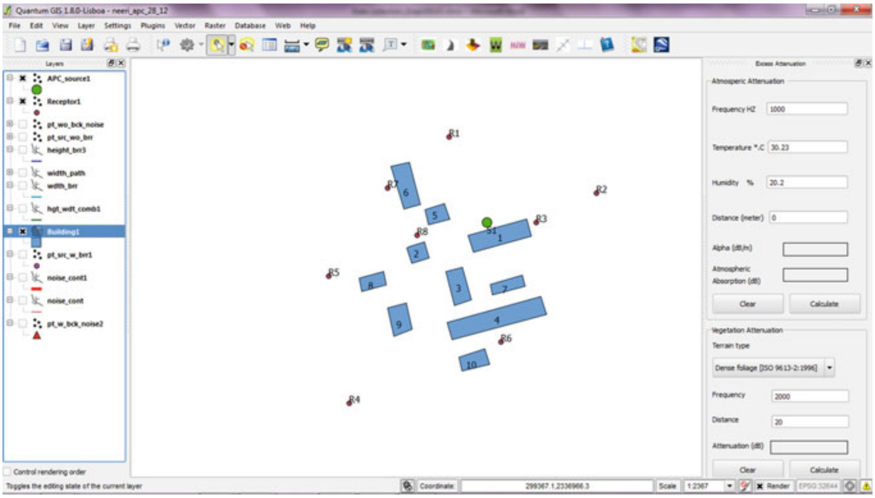


Fig. 5 Location of source (S1), receptors (R1–R8) and buildings (1–10)

prominent as compared to simulated noise due to construction activity. Therefore, final simulated noise levels considering construction activity and background noise are more or less equal to background noise except at R3 and R8.

4.2 *Scenario 2—With Barrier/Building*

In this scenario, apart from source and receptors existing barriers in terms of building are considered (Fig. 5) in the noise simulation. There are ten buildings present in the transmission path between source and receptors. Since building information is stored in GIS database in terms of its geometry and height, information is stored as their attribute.

Table 1 Comparison of barrier attenuation as per height and width analysis

(a) **Height wise**

Attribute table - height_brr3 :: 0 / 8 feature(s) selected

	sr_id	actual_len	barr_len	attenuation
0	S1R1	91.2866714401	91.2866714401	0
1	S1R2	110.34215049	110.34215049	0
2	S1R3	48.1725845576	54.5006400397	21.3487927894
3	S1R4	222.861213967	226.65037635	19.1428007215
4	S1R5	163.753690554	163.976577713	7.61159646281
5	S1R6	118.225907638	126.221100624	22.3577362826
6	S1R7	102.771576735	104.274267091	15.2058274332
7	S1R8	69.7133484613	69.7133484613	0

(b) **Width wise**

Attribute table - wdth_brr :: 0 / 8 feature(s) selected

	sr_id	actual_len	barr_len	attenuation
0	S1R1	91.2866714401	91.2866714401	0
1	S1R2	110.34215049	110.34215049	0
2	S1R3	48.1725845576	48.690390711	10.8210474292
3	S1R4	222.861213967	224.750453847	16.1731213973
4	S1R5	163.753690554	163.78122242	2.011736219
5	S1R6	118.225907638	194.033685054	32.1039516065
6	S1R7	102.771576735	111.643263719	22.8070149811
7	S1R8	69.7133484613	69.7133484613	0

(c) **Final as per height/width**

Attribute table - hgt_wdt_comb1 :: 0 / 8 feature(s) selected

	sr_id	actual_len	barr_len	attenuation
0	S1R1	91.2866714401	91.2866714401	0
1	S1R2	110.34215049	110.34215049	0
2	S1R3	48.1725845576	48.690390711	10.8210474292
3	S1R4	222.861213967	224.750453847	16.1731213973
4	S1R5	163.753690554	163.78122242	2.011736219
5	S1R6	118.225907638	126.221100624	22.3577362826
6	S1R7	102.771576735	104.274267091	15.2058274332
7	S1R8	69.7133484613	69.7133484613	0

Impact of barrier is required to be calculated before computing noise level at receptor due to noise source and its further consideration in noise simulation. For this, attenuation due to barrier is calculated based on the barrier height and width. Table 1a explains the calculation of the attenuation due to barrier height and width, respectively. It computes the actual distance between source and receptor and distance due to barrier in the path of source and receptor. A difference can be seen in the actual and barrier distance in the attribute table of the output file at the paths S1R3, S1R4, S1R5, S1R6 and S1R7 due to barrier obstruction. No difference is observed at the paths S1R1, S1R2 and S1R8 between actual and barrier distances, since no barriers exist in these paths. The minimum and maximum path differences are observed at S1R5 (0.23 m) and S1R6 (8.0 m), respectively. Based on the path difference, attenuation is calculated based on Fresnel number as given in the attenuation attribute of the output file. Therefore, attenuation due to barrier height at these two paths S1R5 and S1R6 is calculated as 7.6 dB(A) and 22.3 dB(A), respectively.

Similarly, attenuation is calculated due to barrier width in the path of source and receptor. First, it generates the probable path from the corners of building facing to source and reaching to receptor based on the building geometry; then, it selects

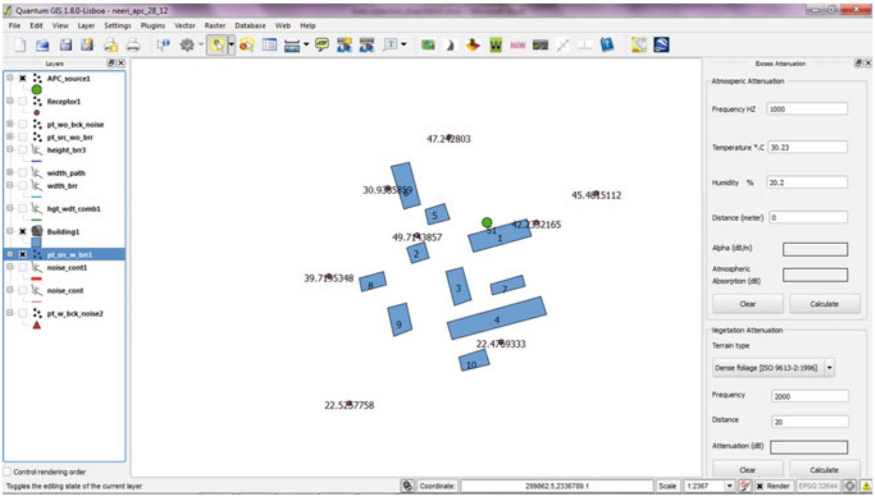


Fig. 6 Noise levels at receptor due to source and effect of building/barrier (without background noise)

the path having minimum distance from the source and receptor to the corner of the building. Table 1b provides the final path due to barrier width analysis. The minimum and maximum path differences are observed at S1R5 (0.03 m) and S1R6 (75.8 m), respectively. Therefore, attenuation due to barrier width at these two paths S1R5 and S1R6 is calculated as 2 dB(A) and 32.1 dB(A), respectively.

Based on the attenuation due to barrier height and width, final attenuation is calculated at the receptor based on the barrier combined analysis. This analysis requires the output of both barrier height and width results with attenuation values and computes the final attenuation at receptor ends. It considers the minimum path difference due to barrier between source and receptor means noise indicating more impact of source on receptor. This provides minimum and maximum path differences at S1R5 (0.03 m) and S1R6 (8.0 m), respectively, and corresponding attenuation as 2 dB(A) and 22.3 dB(A), respectively (Table 1c). This attenuation calculation due to barrier geometry helps in further simulation of noise from the construction activity at the receptor end.

Based on the barrier calculation, attenuation values of noise at receptor end were used to compute the simulated noise level due to point noise source (construction equipment). The simulated values of noise due to source and barrier are presented in Fig. 6 without considering the effect of background noise. Receptors in the back side of barrier have less noise levels due to construction activity as compared to those which are the in the direct side of source. Minimum and maximum noise levels are attributing to R6 [22.5 dB(A)] and R8 [49.7 dB(A)] based on their proximity to source (S1) and impact of barriers.

A comparison between simulated noise with and without barrier is carried out to assess the impact of barrier between the propagation path of source and recep-

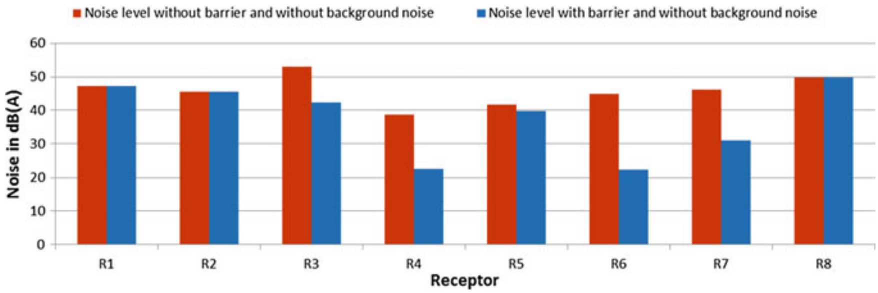


Fig. 7 Comparison of noise level at receptor due to point source considering without and with barrier effect

tor. Figure 7 presents the noise levels at receptor due to point source (construction activity) considering without and with barriers between source and receptor. No difference in the noise levels was observed in both the cases at R1, R2 and R8 since no barrier existed between source and receptor. A significant difference is observed in the noise levels at R3, R4, R6 and R7 from 10 to 20 dB(A) due to barrier attenuation. A slight difference in the noise levels is observed at R5. This may be due to position of barrier which is not intersecting the propagation path and passing through one of the sides.

The simulation result of the point module of N-Dhwani software is validated with the observed noise levels monitored at receptor ends. For this, scenario 2 is considered for validation where existing barriers are present between source and receptor. The observed and simulated noise levels indicate a strong relationship with correlation coefficient of 0.97.

5 Summary and Conclusion

GIS-based noise simulation software (N-Dhwani) has been designed and developed for point sources of noise as well as impact of geographical feature and meteorological parameters on noise calculation. Salient features of the N-Dhwani and its application are explained briefly in the following points.

- N-Dhwani is a user-friendly, stand-alone and platform-independent software. Since N-Dhwani is developed using open-source technology, it allows user to modify algorithms for noise calculation and estimation.
- Multiple point sources can be modelled with the help of point module of N-Dhwani as well as combined effect of point and mobile sources on receptor.
- Based on the capabilities of N-Dhwani, it helps in designing noise barriers and green belt to check the propagation of noise due to point sources including loudspeaker, generator sets, blasts in mine and any new development and construction activities for present and future noise predictions.

- N-Dhwani allows simulation results in terms of noise contour and grid for its display on Web and Google Earth for better visualization and understanding.
- N-Dhwani also assesses the noise pollution in a specified area considering Indian standards as per simulated and observed levels.
- N-Dhwani provides solution for abatement of noise pollution and acts as a decision-making tool in environmental impact assessment (EIA) studies as well as in environmental planning and management.

Acknowledgements Authors are thankful to the Director, CSIR-NEERI, Nagpur, for providing encouragement and necessary infrastructural support. The Department of Science and Technology, New Delhi, is also acknowledged for providing the financial support to develop this software.

References

1. Vijay, R., Sharma, A., Kumar, M., Shende, V., Chakrabarti, T., Gupta, R.: GIS based noise simulation open source software: N-GNOIS, Fluct. Noise Lett. **14**(1), 1550005 (12 p) (2015). https://doi.org/10.1142/s0219477515500054
2. Rau, J.G., Wooten, D.C.: Environmental Impact Analysis Handbook, University of California at Irvine. McGraw-Hill, New York (1980)
3. Bies, D.A., Hansen, C.H., Engineering Noise Control, Theory and Practice, 4 edn. Spon Press, London (2009)
4. ANSI S1-26-1995: Method for the calculation of the absorption of sound by the atmosphere
5. ISO 9613-1:1996: Acoustics-attenuation of sound during propagation outdoors-Part 1: Calculation of the absorption of sound by the atmosphere
6. ISO 9613-2:1996: Acoustics-attenuation of sound during propagation outdoors-Part 2: General method of calculation
7. Piercy, J.E., Embleton, T., Sutherland, L.: Review of noise propagation in the atmosphere. J. Acoust. Soc. Am. (JASA) **61**(6), 1403–1418 (1977)
8. Maekawa, Z.: Noise reduction by screens. Appl. Acoust. **1**, 157–173 (1968)

Detection of Harmful Radiations in the Environment by Deploying Wireless Sensor Networks

S. S. Dawn, R. Rathna and T. Sasipraba

1 Introduction

1.1 Proposed Work and Methodology

The proposed work is to detect harmful radiation in air using wireless sensor networks that helps in real-time radiation monitoring. In designing such a system, this project takes an insight into the problems related to node arrangement, data transfer, and energy consumption and comes up with an algorithmic procedure.

The proposed system is developed to overcome the disadvantages of existing system [1]. It is based on two functions. One is clustering, and another one is scheduling. Initially, the nodes are arranged in the random manner. Based on the hierarchical clustering, the clusters are formed. The hierarchical clustering is based on the max-min-d clustering. For each node, the random value is assumed. And it is named as alpha. The random value ranges from zero to one. The highest alpha is selected as an initial cluster head (ICH). The round one takes place to transfer the data based on the sorted alpha value in each cluster. The cluster head sends data to the sink node or base station.

The lower layer wireless sensor nodes sample and collect the heterogeneous data from the sensors placed inside and transmit the data packets to the upper layer. The upper layer (cluster heads) aggregates the data and forwards it to the sink node (gateway) kept at the deployment site. The TelosB motes inside the cluster can be

S. S. Dawn · R. Rathna · T. Sasipraba (✉)
Sathyabama Institute of Science and Technology, Jeppiaar Nagar, Chennai 600119, India
e-mail: tsasipraba@yahoo.com

S. S. Dawn
e-mail: dawnsudha@yahoo.com

R. Rathna
e-mail: rathna0809@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_25

placed 25 m apart. Inter-cluster nodes can be placed 60 m apart, to avoid interference between the clusters. Each node is interfaced with a GM counter. The data collected from the sensor nodes at the deployment site, through the cluster heads, reach the gateway node. The GPRS facility is present in the Stargate NetBridge (gateway node) for uploading the converted data from the field to the sensor Web using SOS of SWE.

2 Design

2.1 Mote Design

The mote is designed using a board for the microcontroller ATmega1281, 8 MHz frequency, 4 KB EEPROM, a small flash memory, a SD card, and a clock. This baseboard weighs 20 gms. The Zigbee radio which is used for communication is placed in a separate board, and it is integrated with the baseboard. The 'Geiger–Muller' (GM) tube is used for the detection of ionizing radiation. It is interfaced with the baseboard.

The ionizing effect is used in the GM tube as a means of detecting the radiation. It is a hollow cylinder filled with a gas at low pressure. It has a thin window made of mica at one end. An electrode is present in the center of the GM tube. A voltage supply is connected across the tube and the electrode. When an alpha particle or beta particle or gamma radiation enters the tube, it will ionize the gas. These newly created ions make the tube to conduct. For a short time, current is produced in the tube. As a result of this, a voltage pulse will be created. That pulse exactly corresponds with

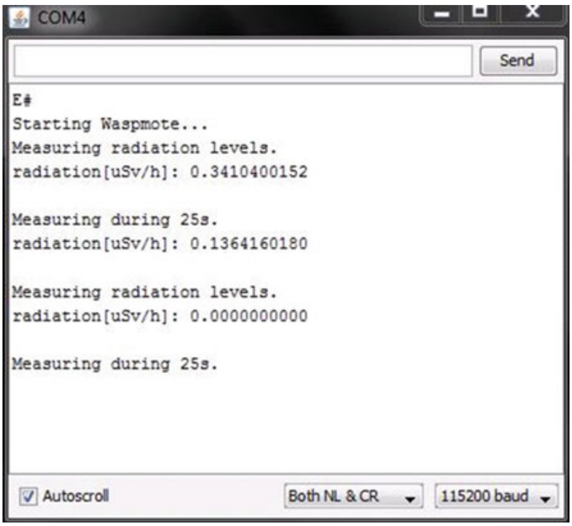


Fig. 1 Recorded values in software

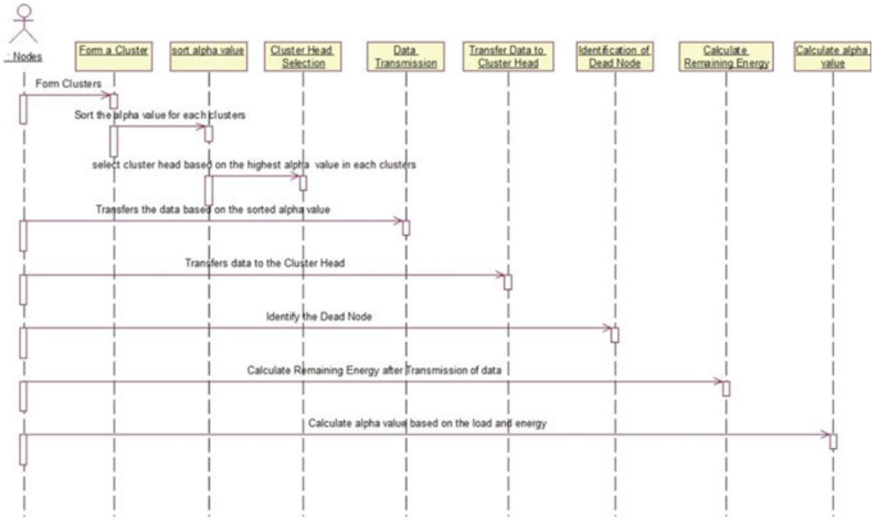


Fig. 2 Sequence diagram of nodes in the network

the amount of radiation entering the tube. So, the count of such pulses is recorded to get the exact amount of radiations in sievert (Sv).

On the deployment site, this mote is placed. This set of board with sensor and Zigbee board with antenna is referred as 'mote'. All the separate modules are shown in the following diagrams. They together make up the mote. The base sensor board with ATmega1281 microcontroller has been programmed to record the radiation dosage for every 25 s initially. Waspmote IDE is the software used to program the board in 'C'.

The data (ionizing radiation amount in sievert) should reach the access point for monitoring. This count sent from the mote through the Zigbee radio is received by the Zigbee radio (with antenna) interfaced to the server system. Accordingly, that is programmed. After finishing the programming phase, the sensor board starts detecting the radiation even in an enclosed safe place. The output obtained as a result of newly programmed sensor board in this work using Waspmote IDE is shown in Fig. 1. That figure shows the dynamic recording of natural radiations received by the GM tube when a normal human being came into contact with it.

2.2 Overall Network Design

As given in Sect. 2.1, several motes are configured and placed in the deployment site as a network of wireless nodes [2]. The overall design can be better understood by the following UML diagrams. The first two diagrams (Figs. 2 and 3) describe the design of all the nodes in the network.

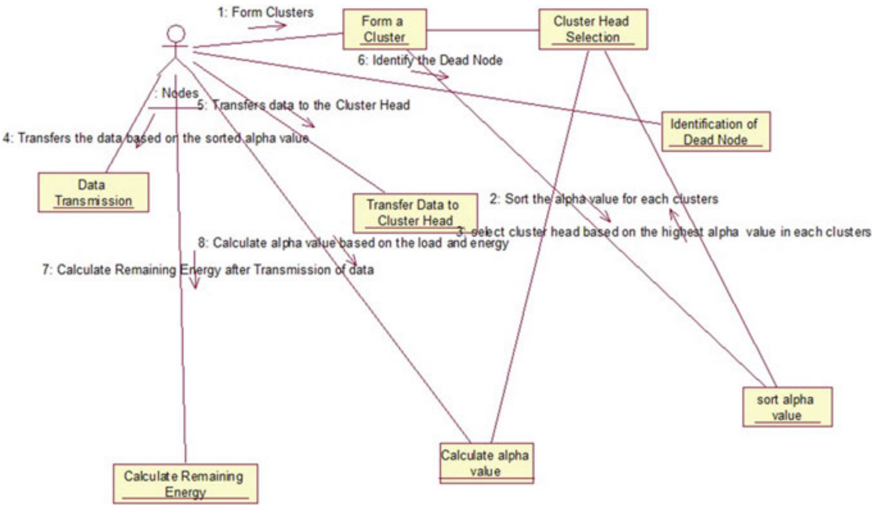


Fig. 3 Collaboration diagram of the nodes in the network

The cluster head selection based on the alpha values is explained clearly in the following two diagrams (Figs. 4 and 5).

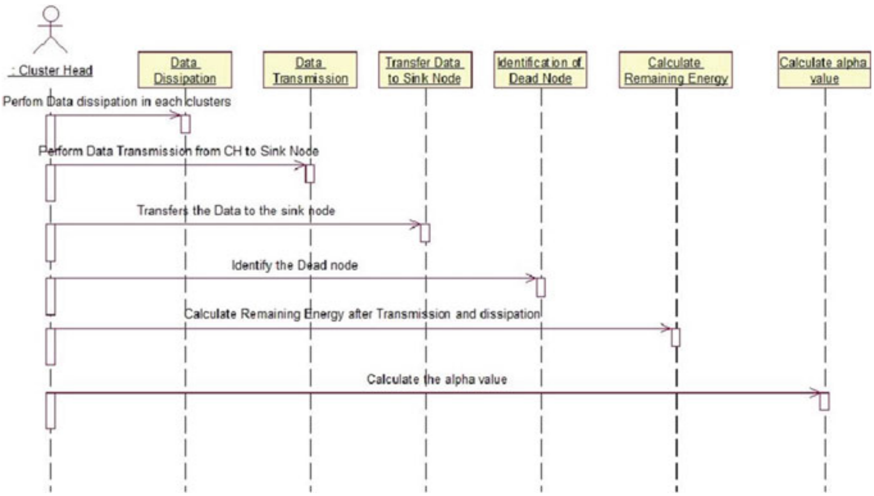


Fig. 4 Sequence diagram of the cluster head

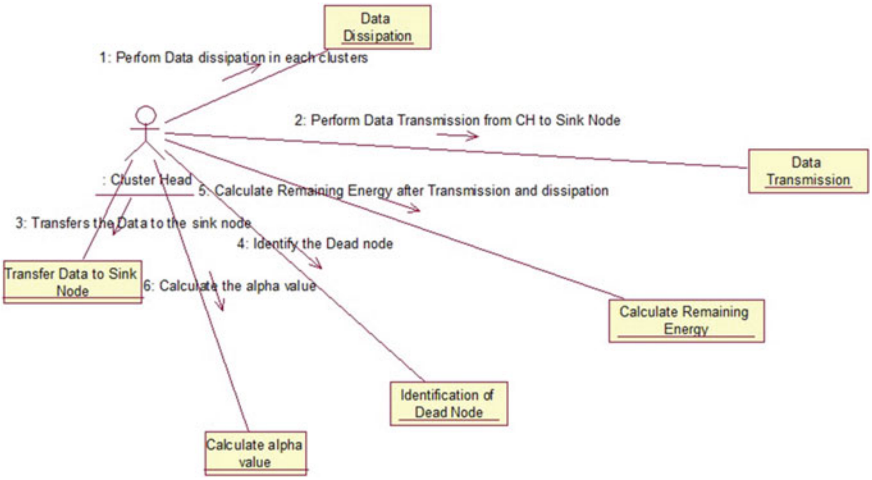


Fig. 5 Collaboration diagram of the cluster head

3 Implementation Details

Radiation sensor board is based on a Geiger–Muller tube, which consists of a tube filled with a low-pressure (~0.1 Atm) inert gas such as helium, neon or argon (usually neon), in some cases in a penning mixture, and an organic vapor or a halogen gas. The tube contains electrodes, between which there is a potential difference of several hundred volts, but no current flowing. The walls of the tube either are entirely metal or have their inside surface coated with a conductor to form the cathode, while the anode is a wire passing up the center of the tube.

When ionizing radiation passes through the tube, some of the gas molecules are ionized, creating positively charged ions and electrons. The strong electric field created by the tube's electrodes accelerates the ions toward the cathode and the electrons toward the anode. The ion pairs gain sufficient energy to ionize further gas molecules through collisions on the way, creating an avalanche of charged particles. This results in a short, intense pulse of current which passes (or cascades) from the negative electrode to the positive electrode and is measured or counted.

XBee-PRO ZB embedded RF module is used in the mote to provide wireless connectivity to other devices with Zigbee antenna (Fig. 6).

The gateway node (Fig. 7) allows to collect data from the sensor network and passes into a PC with a standard USB port acting as server. This gateway node acts as a 'data bridge or access point' between the sensor network and the server or application. The server is responsible for storing and using the data received depending on the specific needs of the Web application.

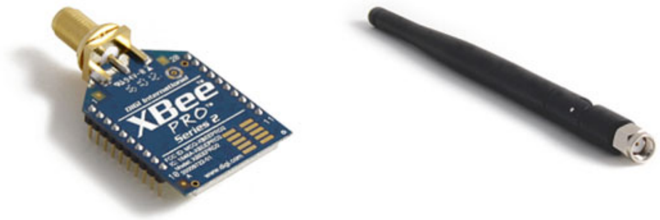


Fig. 6 XBee-PRO ZB Antenna

Fig. 7 Gateway node

3.1 Procedure Inside the Network

The above description gives the details of a single module of sensor node and its receiver. If 50 such modules are used in a particular area, then the radiation level in that area can be continuously kept under surveillance for the presence of harmful radiation. In that case, the clustering [3] and the corresponding scheduling algorithm will be used as given below.

Based on the clustering and scheduling, the LECSA is implemented. There are various clustering algorithms used to form a cluster. Here, hierarchical clustering and max-min-d clustering are used to form a cluster. The scheduling is based on the TDMA. The scheduling is used to avoid collision in the data transfer. The scheduling can schedule the data transfer based on the node weighting parameter.

Clustering. Clustering is the procedure used to group a number of similar things in order to achieve a common goal. With clustering, each node in the network assumes a certain role and along with that role some privileges in order to achieve the common goal. These roles can be easily distinguished and are: Leader or cluster head [4] is the node that is responsible for each one of the clusters. Simple nodes are those that have no special role in the clusters. Gateway nodes are located on the outskirts of the clusters and connect them to the rest of the network.

Scheduling. The sensor nodes can send the data based on the node weighting parameter [5]. The node weighting parameter is sorted in ascending order to all the clusters. Based on the sorted values of alpha, the nodes send data to the cluster head. The maximum alpha value of the node is considered as a cluster head (CH). In each cluster, except the maximum value of alpha the other nodes can transfer the data to the cluster heads. And all the cluster heads can communicate to the sink nodes. In each round, different cluster heads can communicate with the sink node.

4 Result and Outcomes

4.1 Simulation

The nodes are grouped together and form a cluster. The first node is selected and calculates the distance between the adjacent nodes. It starts generating the clusters by taking the first node by their ID and the nodes which are present around over 25 m. The alpha values are generated randomly for each node. And the highest alpha value of each cluster is selected as initial cluster head. For first round, based on the sorted values of alpha, the data can be transferred from each node to cluster head. The cluster head transfers the data to the sink node. For further rounds, the alpha value is calculated based on the load and remaining energy value. Same as previous procedure, the data can be transferred from each node to cluster head and cluster head to sink node.

Initially, the algorithm is implemented at simulation level. The LECSA [6] consumes less energy to transfer data. In each round, the cluster head dynamically changes from one node to another node. Therefore, the energy consumption of each round is reduced. Then, the lifetime of the network is increased.

The LECSA algorithm was simulated in MATLAB. Initially, all the nodes are randomly placed using random function based on MATLAB function. The nodes are uniquely distinguished by using their node ID. The clusters are formed based on the Euclidean distance. And the clusters are differentiated using different color codings using MATLAB function.

For the first round, the alpha value is generated to all the nodes. Based on the alpha value, the cluster head is selected and data transfer takes place. For further rounds, the alpha value is calculated based on the load and energy of each node. Same process takes place for all the remaining rounds [7]. Finally, the first node dies at 1096th round (Table 1).

Table 1 Comparison of the proposed LECSA with existing algorithms

Algorithm	Round in which first node dies
LEACH	996
DECSA	1050
LECSA	1096

4.2 Test Bed Implementation

A field study has been done in the surrounding areas of IGCAR. As given in the description in Chap. 3, some motes are deployed at four different places around the power plant and the values received were analyzed.

The radiation exposure values recorded by the GM tube in microsievert per hour have to be converted to dose and adjusted with factors (millisieverts per year) to estimate the health impact (if any). If the detected value is greater than or equal to 1.0 uSv/hr, then it is dangerous. So, the threshold value is set as 0.8. So when the measured values exceed this threshold, those values will be immediately sent to the concerned authorities through the app connected with the server. This test bed was checked for a period of 6 months, and the values never exceeded 0.8.

Our detection system aims for the systematic development of a Web application which can be used for accessing sensors observing nuclear radiation data. The system uses the ad hoc wireless sensor networks and spatial data infrastructure to deploy an interoperable Web app. Methods of data access, communication, and visualization were implemented using OGC norms and SWE specifications.

5 Conclusion

In the present-day scenario, the environmental monitoring does play a very important role and is the need of the hour. With the increasing radiations of various sorts at different levels, it has become a tedious process to protect the ecosystem from these evolving power sources.

Though there have been a lot of devices and systems employed in the domain, there is still a problem that exists with the efficiency, throughput, energy consumption, etc. Wireless sensor networks are the most applied technology among all that exists. However, under the belt there exist fewer and indeed grave problems that hold back the monitoring under a lot of parameters.

It has been observed that all the crucial factors which are responsible for energy wastage like the load of the node, its remaining power, and the distance from other nodes for communication have been taken into account in LECSA. And the implementation has proven that LECSA is best suitable for environmental monitoring applications.

References

1. Reyes, A., Pastor, E., Barrado, C., Gasull, M.: Monitoring radiological incidents through an opportunistic network. Lat. Am. Trans. IEEE (Rev. IEEE Am. Lat.) **13**(1), 54–61 (2015)
2. Ergen, S.C., Varaiya, P.: TDMA scheduling algorithms for wireless sensor networks. Wirel. Netw. **16**(4), 985–997 (2010)

3. Zeng, B., Dong, Y.: A collaboration-based distributed TDMA scheduling algorithm for data collection in wireless sensor networks. J. Netw. **9**(9), 2319–2327 (2014)
4. Yong, Z., Pei, Q.: A energy-efficient clustering routing algorithm based on distance and residual energy for wireless sensor networks. SciVerse ScienceDirect Proc. Eng. **29**, 1882–1888 (2012)
5. Shafiullah, G.M., Azad, S.A.Shawkat, Ali, M.: Energy-efficient wireless MAC protocols for railway monitoring applications. IEEE Trans. Intell. Trans. Syst. **14**(2), 649–659 (2013)
6. Rathna, R., et al.: Wireless sensor network for harmful radiation monitoring system and web enablement using OGC interface. Res. J. Pharm. Biol. Chem. Sci. **7**(2), 1096–1101 (2016)
7. Rathna, R.: Simple clustering for wireless sensor networks. Cybern. Inf. Technol. **16**(1), 57–72 (2016)

Modeling the Impact of Highways on the Impedance of Movement and Mortality Locations of Wildlife: A Case Study of the NH6 in Maharashtra, India

S. Kumar, P. Khatavkar, P. Dev, U. V. S. Yadav and A. Shinde

1 Introduction

Roads are critical components of human life and play a major role in the economic development of any country. Numerous studies prove the causal link between road networks and economic development [1–4]. India today has a major focus on the development of its road network through its ambitious flagship programs on building and upgrading road infrastructure across the country. With the world's second largest road network of around 70,000 km in length, the highways sector in India is a key focus for modernization and is being touted as a major opportunity for investors, with 2015 seeing projects worth an estimated INR 125 trillion earmarked for future development [5, 6]. However, very often, the development of road infrastructure becomes controversial due to its short-term and long-term environmental impacts and moreover lack of effective mitigative measures. They are thus viewed critically and very often challenge political interests as seen in the recent case of highway development along the National Highway NH7 and NH6 in Maharashtra, India [7].

2 Background

Enormous documentation is available on the effects of roads and road development on wildlife. Over the years, starting from the early 1970s, there have been several studies that began with documenting the effects of roads on wildlife. It is well known that

S. Kumar (✉) · P. Khatavkar · P. Dev · U. V. S. Yadav · A. Shinde
Institute of Environment Education and Research, Bharati Vidyapeeth Deemed University, Pune, India
e-mail: shamita@bvieer.edu.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_26

roads are barriers to animal movements [8–18]. They cause mortality from vehicle collisions [11, 19–32]. Roads are also known to bring about modification of behavior [13, 33–38] and act as conduits for invasive species, [39–43]. They enhance edge effects, [44–48] disrupt ecosystem structure and function, cause fragmentation, loss of connectivity, sedimentation, soil erosion, air and water pollution, increase in noise levels and bring about changes in land use and land cover [47–50].

As India embarks on an ambitious road development program connecting the corners of the country, these new roads will infringe on critical wildlife habitat causing fragmentation of corridors. However, careful planning, design, and implementation of appropriate environmental safeguards that are available today can minimize these impacts [51]. Geoinformatics is a key pillar of such technology development.

3 Modeling for Prioritizing Mitigation Measures

While there exists an enormous body of research that focusses on the effects of roads on environment, ecology, and wildlife, it is only recently that research has focused on identifying mortality risk posed by roads for wildlife and hotspot analysis based on wildlife mortality for setting up mitigation measures in the form of underpasses and overpasses [52, 53], Eslinger and Morgan (2017). Moreover, recent studies have proved that hotspot analysis of roadkills is not the key determinant in planning mitigation measures. Landscape structure and porosity also affect the rates of wildlife-vehicle collisions [54–57]. Most of these methods require extensive wildlife datasets, complex simulation models, and are oriented for single species. It is not always possible to acquire such extensive wildlife data sets. Moreover, the models very often only predict mortality risk posed by a road and do not prioritize wildlife mortality locations [58]. When dealing with limited datasets, understanding the interaction of wildlife with different land cover land use types is important, as animal movements are dependent on the land cover type along the road and not solely on the road itself. Each animal interacts differently with different land cover types. Thus, human settlements, normally avoided by most species, indicate higher resistance while grassland and forested areas may have lower resistance. This resistance is interpreted as impedance developed by a particular land cover type. If the impedance is high, the animal will not be able to cross and vice versa. A model that uses impedance and traversability to predict roadkills has the potential to be a useful tool in impact assessment of highway projects facilitating the building of under and overpasses wherever necessary, avoiding the long litigation process and contributing to sustainable development.

4 Objectives

The aim of this study is to develop a model for prioritizing locations for mitigation measures by combining the impedance and mortality risk posed by a selected section of NH6 in Maharashtra state. Considering availability of data, this was done for Chital deer (*Axis axis*) and Gray Langur (*Semnopithecus*). The main objectives were (1) to model the mortality risk posed by the NH6 highway for selected species and (2) to identify zones of relative mortality of indicator species for developing wildlife mitigation measures such wildlife underpass or overpass.

5 Materials and Methods

5.1 Study Area

The study area is along a selected section of the NH6 highway comprising 84 kms and passing through the Navegaon-Nagzira wildlife corridor, Maharashtra state. The National Highway Authority of India (NHAI) approached the Wildlife Institute of India seeking expert advice for building underpasses on this newly expanded road. The reporting of a large number of roadkills from this area following the expansion of the old highway to a wide four-lane highway with a divider justifies its selection as an ideal study site.

5.2 Wildlife Mortality Data

As accurate wildlife mortality data for this region was not readily available in the open domain, field data were collected by the team through a vehicle survey of the selected stretch of the highway traveling at the speed of 20 km/hr. The data collection form was customized using ArcPad software. The detailed form contains data fields relating to name of surveyor, date, time, and location of survey (automatic field), carnivore species (drop-down box), herbivore species (drop-down box), amphibian species (drop-down box), bird species (drop-down box), reptile species (drop-down box), angle of transverse, image Id (geotagged image), land cover land use at the point of roadkill, road condition, and field notes. This digital form with geotagged photographs eliminated the need for paper forms and separate equipment such as GPS, camera. It also eliminated human errors in data collection and provided accurate locations with geotagged photographs.

The surveys were conducted twice a day (6 am and 6 pm) over three days every month in the 29-month period. The entire length of the road was traversed by the vehicle with the researchers walking intermittent distances perpendicular to the road into the forests as very often bigger mammals wander off into the adjacent vegetation, especially if they are just hurt by the oncoming vehicles. Hair samples were collected in case of non-identification due to extensive injury.

5.3 Land Cover Land Use Data

Each land cover land use class offers varied impedance values for wildlife. Consideration of this parameter is thus crucial to such studies. Using a supervised classification approach, the land cover land use map was prepared using LISS IV images for the period December 2011, acquired from the National Remote Sensing Centre (NRSC). The data were projected in UTM 44 N, and the land cover land use map thus prepared consisted of eight classes, viz. water, forest, shrubland, open scrub, barren, agriculture, and built up. Validation of the classification through intensive ground truth resulted in an overall accuracy of 87.82% and a kappa statistics of 84.33.

5.4 Preparation of Impedance Map Using Expert Knowledge

In the absence of individual observations, we used expert knowledge to prepare the impedance maps and expert weights to evaluate the impedance of animal movements through the landscape. They constituted personnel from the local forest department, local wildlife observers, and conservation scientists. Each expert assigned an impedance value between 0 and 10 for each of the species, i.e., Chital and Gray langur with respect to each land cover type, where the value of zero represents no impedance and the value of 10 represents an absolute barrier for wildlife movements. The expert's knowledge about the species determined the weight given to the expert, e.g., a large body of expertise on the indicator species merited a higher weight to the expert and vice versa.

Impedance value for each individual land cover class $= \Sigma(VE_n \times WA_n)$.
where VE = Value given by each expert for every land cover category,
WA = Weight Assigned to each expert.

To enable prediction of the relative mortality locations, the land cover land use map was re-classified into four classes, viz. no impedance, low impedance, medium impedance, and high impedance using the impedance given by the experts. Points were then placed at a distance of 500 meters along the NH6 highway, and a circular buffer of 250-meter radius was developed around the points with the highway in the centre. (Buffer radius was determined by road width and heterogeneity of classes inside the circle using a thumb rule). Zonal statistics were used to determine the impedance class values.

5.5 Traffic Intensity Data and Traffic Characteristics

This dataset is an input for determining animal mortality. We monitored vehicular intensity data along the NH6 highway at four points over an eight-day period. The data were converted to weekly and further monthly and yearly averages and then into the average yearly number of vehicles per second as an input into the traversability equation for predicting animal mortality locations [58, 59].

5.6 Estimating the Probability for Mortality of Chital and Gray Langur

The entire road length does not pose a risk to animal mortality. Some parts pose a greater risk than the others. Wildlife mortality is directly dependent on the vehicular intensity and vehicular speed [60]. To determine the mortality risk on the NH6 highway the "*Traversability Equation for Animal Movements*" developed by [59] was used. The assumption of the traversability equation for animal movement is as follows:

- The collision between car and animal occurs when a car arrives at a spot before the animal leaves the spot
- The animal traverses the highway at a constant speed
- The animal dies if it collides with the vehicle
- The animal can successfully cross the road through the gaps of two vehicles at a given location.
- To estimate the probability of animal deaths with relation to traffic intensity, Poisson distribution has been used. In the traffic modeling, the Poisson distribution P(x) is used to estimate the probability of vehicles "x" coming in direction of road in given period of time T (in seconds) with equation

$$P(x) = \frac{(\lambda T)^x e^{-\lambda T}}{X!} \quad [47] \tag{1}$$

where λ = traffic volume in vehicles s^{-1}

If the animal has to successfully cross the road without getting killed, then x should be equal to "0". The equation then changes to

$$\mathrm{P}(0) = e^{-\lambda T}$$

This means P(0) probability of successful animal crossing depends if the front end of the next car will not arrive in time period T second considering the traffic flow on average vehicles λ s^{-1}. The probability of the animal being hit by a car will depend upon vehicular intensity, average vehicular length, average vehicular width, species characteristics and traffic speed.

$$P(a) = e^{-\lambda\left(\frac{\frac{W_c+L_a}{\cos(\alpha)}}{V_a} + \frac{L_c+W_a\cos(\alpha)}{V_c}\right)} \quad [47] \tag{2}$$

where W_c = average vehicle width in meter
L_a = average animal length in meter
V_a = average animal traverse
Cos(α) = angle of animal traverse
L_c = average vehicular speed
W_a = average animal width
V_c = car length
In equation (2), Cos(α) = 0

i.e., perpendicular traverse, the angle then equation would be

$$P(a) = e^{-\lambda\left(\frac{W_c+L_a}{V_a} + \frac{L_c+W_a\cos(\alpha)}{V_c}\right)} \quad [47] \tag{3}$$

To estimate risk for animal mortality, probability during traversing (Da)

$$D_a = (1 - P_a) \quad [47] \tag{4}$$

For the study, Eqs. 2 and 4 have been used for estimating the probability of the animal colliding with the vehicle. If data on animal crossings are available, then from Eq. 4, an estimate of the accurate number of animal deaths, which can occur due vehicular collision, is calculated by

$$D_a = (1 - P_a)K_{a,t} \quad [47] \tag{5}$$

where Da becomes actual number of roadkills due to vehicular collisions,
$K_{a,t}$ = Number of attempt made by animal to cross the road in time "t" from actual field data.

However, $K_{a,t}$ is almost impossible to predict as conditions change and the crossings will depend on diverse factors.

Angle of Traversing (cos α)—It is not possible to predict the angle at which animal may cross the road with respect to the oncoming vehicle. However, the probability of animal death at each traverse angle with respect to the vehicle can be calculated. As this is not feasible in the current research, an estimation of animal mortality due to vehicular collision is done at traverse angles of 0, 30, 60, 90, 270, 300, and 330 degrees. The mean calculated from all the estimated probability values of angle of traverse has been assigned as the estimated probability value for the given road segment.

Based on traffic monitoring stations, the NH6 highway was segmented into four sections. Using the traversability equation, the probability of mortality was calculated for each of these sections and was further divided into three classes, viz. low, medium, and high probability.

Table 1 Estimated probability classes using traversability equation

Impedance classes for animal movement			
→	Estimated probability classes using traversability equation		
↓	Low probability	Medium probability	High probability
Low impedance	Low risk	High risk	High risk
Medium impedance	Low risk	Medium risk	High risk
High impedance	Low risk	Low risk	Medium risk

5.7 Prediction of Mortality Risk Locations

For prediction of mortality zones along the NH6 highway, the map produced from the results of the traversability equation and a circular buffer map was overlaid with the impedance raster map.

Considering the situation, where the traversability equation predicts a high mortality of a particular species on a highway segment, where the mean impedance values are high, the number of wildlife crossing the highway would be less, resulting in fewer roadkills. However, if the traversability equation predicts a moderate mortality zone along a given part the highway, where the impedance is low or zero, then the wildlife crossing will be high, increasing the possibility of wildlife-vehicle collisions. The animal mortality risk zone is developed using (Table 1).

5.8 Validation of the Model

The model output is compared with the observed roadkills of Chital and Gray Langur through consideration of the percentage of roadkills observed in high, moderate, and low mortality risk zones. Similarly, present known wildlife crossings are used to compare their presence in high, medium, and low mortality zones. This is done as crossings will keep changing and wildlife very often may not die immediately after a collision with vehicles, but a little distance from the highway at some other spot.

6 Results

6.1 Wildlife Mortality

Over a two-year period, 499 roadkills were observed on the NH6 highway of which 238 were reptiles. Of these, the common krait, rat snake, stripped keelback were seen commonly. Sixty-three of the roadkills were herbivores, of which Gray langur comprised the bulk, 57 were small carnivores, namely civets and jungle cat, while 92 were birds and 49 were amphibians. Herbivore and carnivore roadkills were relatively higher in the summer months, while amphibian and reptile roadkills were significantly higher in the monsoon (Fig. 1).

6.1.1 Land Cover Land Use Classification

For the present research, the 1.5 km stretch on either side of NH6 is extracted for analysis as the maximum impact zone for a highway varies from 100 to 1500 m on either side [60, 61]. The clipped NH6 highway encompassed eight land use land cover classes with an area of 176.30 km^2. This includes 29% of agriculture, 29% forest land, 11% built up, 17% shrubland, 4% of barren land, and 6% of open scrub. Water bodies cover a very small portion land at 2%.

6.2 Impedance Maps

The impedance maps for the Gray langur along the NH6 highway depict that 46% of the zones (circular buffer of 250 m radius) are of high impedance for the langur movement, while 29% of the zones are of low impedance allowing for high mobility. The higher impedance is observed near deep water bodies and burnt land while the medium impedance is observed in 25% of zones, which mainly consist of built-up and agricultural patches.

The impedance maps for the Chital along the NH6 highway depict that 62% of the zones (circular buffer of 250 m radius) are of high impedance for the movement of the Chital and mainly consist of built-up areas, burnt land, and human settlements. Medium impedance is observed in 38% of zones, which mainly consist of shrubland, barren land, and agricultural patches.

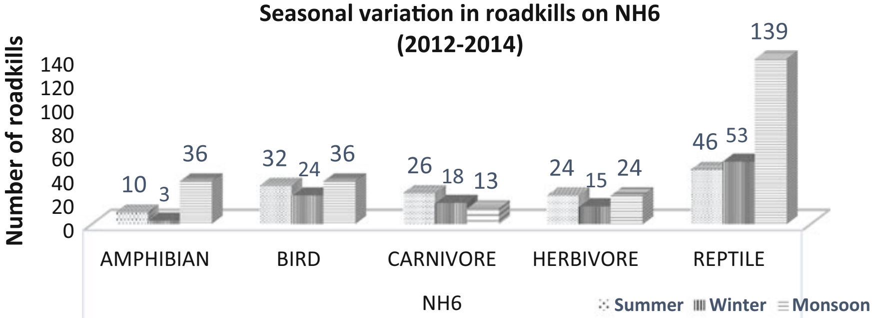


Fig. 1 Seasonal variation in roadkills along NH6

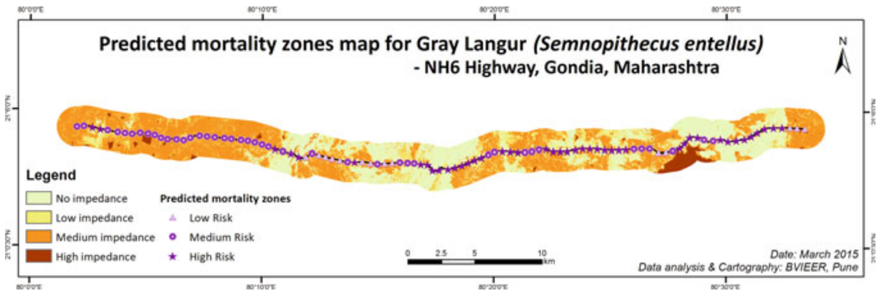


Fig. 2 Predicted mortality zones for Gray Langur on NH6

6.3 Results of the Traversability Equation

The traversability equation for animal movement includes traffic intensity flow, estimated probability for mortality at different traverse angles and predicting the probability associated with the highway segment for mortality of the Chital and the Langur. The probability was estimated for different traverse angle from 0 to 320 degrees. The highest estimated probability for Gray Langur and Chital mortality is observed for a transverse angle of 300 degrees. Given the unpredictable nature of animal movements, it is difficult to calculate the actual angle of traverse for wildlife crossing the road. Thus, an average estimated probability of all the angles is considered to predict the threat of mortality posed by a segment of the highway.

6.4 Traffic Pattern Analysis

The average traffic intensity per hour is observed to be a maximum of 1775 vehicles/hour and minimum of 800 vehicles/hour vehicles, which is extremely high traffic movement for a road that passes through key wildlife areas.

6.5 Prediction of Mortality Zones for Gray Langur

Along NH6 highway, 44% of the predicted mortality zones for Gray langur were high risk, while 14% of the zones were in a low risk and the residual 42% are in the medium risk category. The predicted Langur mortality varies from a maximum 0.77 to 0.44. Considering the probability rates produced through the traversability equation, NH6 highway poses a medium and high risk for Langur mortality. This is corroborated by field observations, where a large number of langur carcasses could be seen being eaten by dogs (Fig. 2).

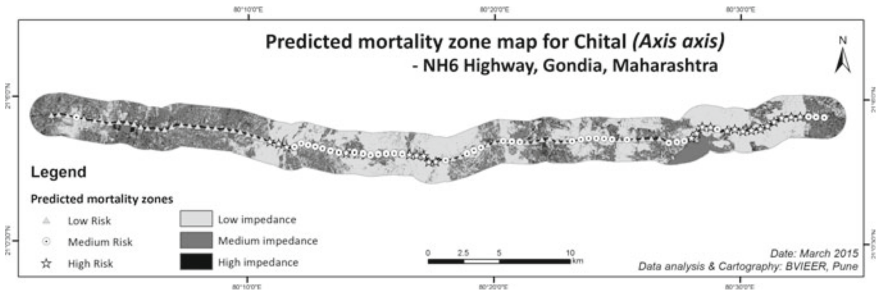


Fig. 3 Predicted mortality zones for Chital on NH6

6.6 Prediction of Mortality Zones for Chital

Chital has a considerably lower percentage of high risk zones, with 23% as high risk zone, 36% as medium risk, and 43% as low risk mortality zone. This finding also corroborates with field observations, where langurs were impatient and wanted to try to the cross the wide highway in one go resulting in their deaths (Fig. 3).

6.7 Validation of Relative Mortality Location for Gray Langur

The model is validated using the mortality data collected from NH6 for the period 2012–2014. The table below shows that the model is able to accurately predict the high risk zones. This can form an important input into impact assessment studies for roads and for planning exact locations of underpasses/overpasses (Table 2).

Table 2 Actual mortality data with predicted relative mortality zone for Gray Langur and Chital

	Predicted relative mortality zone for Gray Langur			Predicted relative mortality zone for Chital		
Predicted zones	Total number of predicted zones	Observed kills	Percentage	Total number of predicted zones	Observed kills	Percentage
Low risk zone	14	2	0.07	41	0	0
Medium risk zone	42	3	11	36	2	22
High risk zone	44	21	80	23	7	77
Total	100	26	100	100	9	100

7 Discussion

Spatial patterns of roadkills seen on the NH6 showed variations according to the taxonomic group, seasonal pattern, and landscape factors. Reptiles formed the largest group that were killed on this newly expanded highway especially during the monsoons, a finding that is in sync with other studies from India, mainly attributed to seasonal migration and juvenile dispersal [62]. Similarly, bird and small mammal mortality peaked during the summer months [11, 19, 23, 63, 24, 29]. The widening of the highway to include four lanes has increased the distance that needs to be traversed by wildlife and the heavy traffic density on this sensitive road does not enable the wildlife to reach the other side. During the field survey, we observed Rhesus Macaque imitating human behavior of watching out for oncoming traffic after reaching the divider! However, sadly other species do not have this trait.

The impedance effect for the two selected species differed on NH6. Both are shy animals, with the Chital keeping more toward the interiors and the Langur preferring higher branches of tall trees that are now restricted to either side of the NH6 highway. The Chital had a greater area of high impedance on this highway compared to the Langur. Development of eateries and other establishments, as well as land sale arising from speculation, will only increase around along the highway and this may increase the impedance for the Chital and associated species. Habitat fragmentation, which is inevitable as part of highway development, poses a greater risk! Rhesus Macaque generally takes over this niche and may end up flattened due to the feeding by travelers [27, 64].

As can be seen from the impedance map, the entire highway does not have a barrier effect. It is essential that conduits for crossing over be constructed if not already done. The study has differentiated between roadkills and impedance as key factors when discussing wildlife mortality along highways. Impedance alone cannot predict roadkills, as animal movements cannot be predicted. However, areas with high impedance may have lesser roadkills. Roadkills are not indicative of impedance, as the kill shows that the particular section of the highway offers less impedance! The animal death due to collision represents the probability of a roadkill.

This study leans on expert knowledge of impedance. Experts may not be knowledgeable about a large number of species. Thus, expert knowledge can be used for a group of similar species and impedance maps can then be created and overlaid along the highway to better decide the location of mitigation measures.

Jaarsma et al. [59] developed the traversability equation for predicting the probability of badger deaths along different local roads in the Netherlands. Khatavkar [58] used it for predicting mortality of roe deer and badger along the A 73 highway in the Netherlands effectively. This study too uses this equation for predicting mortality along the NH6 highway. The traversability equation uses traffic density, average vehicle speed, and species characteristics to predict the probability of a road kill. In this study, it is seen that Chital and Langur have the highest estimated mortality at an angle of 300^0. At this angle, the probability of road kill is at the rear wheels of vehicles. Increased traffic density reduces the gap between two vehicles, which thus provides very less time, and space for an animal to traverse the highway successfully.

Development of mitigation measures in the form of overpasses and underpasses is highly expensive and often comprise 5–7% of the entire cost of the road. Thus, it is important to predict the probability of roadkills along different sections of the highway. The high cost of overpasses and underpasses generally reduces the number of mitigation measures as seen in the NH7 case [7, 65]. However, they are the only solutions to balance the needs of humans and wildlife. Well-designed and well-placed structures can help mitigate the problem.

While most researchers use hotspot analysis to frequently identify areas of mitigation measures, mortality zones have rarely been predicted. This study uses the outputs of the traversability analysis and impedance maps to identify mortality zones. The validation shows that the method is effective in the planning process of highways.

8 Conclusion

Environment impact assessments underestimate the wildlife roadkill phenomena while planning for highways and usually build one or two token underpasses without understanding the species habitat usage. This paper has demonstrated the effectiveness of impedance modeling in planning for mitigation structures on highways. The model uses presence-only data combined with other readily available parameters. The effectiveness of this method can be improved if data of wildlife crossings is readily available along with sensitivity analysis. Toward this end, the development of a Web-based/app-based citizen science data collection system may assist in the development of roadkills database across the country. This open access data can be used effectively for the planning of mitigation measures thus aiding India's massive highway development program. Policies for controlling the linear development along highways especially in the vicinity of protected areas can go a long way in reducing fragmentation of forests along the highways.

Acknowledgements We express our gratitude to the Natural Resources Data Management System, Department of Science and Technology, Government of India for funding this study. The authors would like to especially thank Ujjwal Vir Singh Yadav, a key member of our team. We would also like to thank Vivek Tumsare, Swapnil Girade, Kranti Yardi, Shivam Trivedi, and Aseem Mujawar who provided much-needed support during the field observations.

References

1. Sahoo, P., Dash, R.K.: Infrastructure development and economic growth in India. J. Asia Pac. Econ. **14**, 351–365 (2009)
2. Alder, S.: Chinese roads in India: the effect of transport infrastructure on economic development. Working Paper. University of North Carolina, Chapel Hill (2014)

3. Asher, S., Novosad, P.: Rural Roads and Local Economic Development (2018)
4. Maparu, T.S., Mazumder, T.N.: Transport infrastructure, economic development and urbanization in India (1990–2011): is there any causal relationship? Trans. Res. Part A Policy Pract. **100**, 319–336 (2017)
5. Ghani, E., Goswami, A.G., Kerr, W.R.: Highway to success: the impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. Econ. J. **126**, 317–357 (2016)
6. Wang, R., Zhu, C.: Annual Report on the Development of the Indian Ocean Region (2016): Modi's India. Springer (2017)
7. Habib, B., Saxena, A., Mondal, I., Rajvanshi, A., Mathur, V.B., Negi, H.S.: Proposed mitigation measures for maintaining habitat contiguity and reducing wild animal mortality on NH 6 & 7 in the Central Indian Landscape. Technical Report, Wildlife Institute of India, Dehradun and National Tiger Conservation Authority, Government of India, New Delhi (2015)
8. Bischof, R., Steyaert, S.M., Kindberg, J.: Caught in the mesh: roads and their network-scale impediment to animal movement. Ecography **40**, 1369–1380 (2017)
9. Brody, A.J., Pelton, M.R.: Effects of roads on black bear movements in western North Carolina. Wildl. Soc. Bull. **1973–2006**(17), 5–10 (1989)
10. Chhangani, A.K.: Frequency of avian road-kills in Kumbhalgarh Wildlife Sanctuary, Rajasthan, India. Forktail **20**, 110–111 (2004)
11. Das, A., Ahmed, M.F., Lahkar, B.P., Sharma, P.: A preliminary report of reptilian mortality on road due to vehicular movement near Kaziranga National Park, Assam, India. Zoos' Print J. **22**, 2742–2744 (2007)
12. Kumara, H.N., Sharma, A.K., Kumar, A., Singh, M.: Roadkills of wild fauna in Indira Gandhi wildlife sanctuary, Western Ghats, India: implications for management. Biosph. Conserv. Nat. Wildl. Hum. **3**, 41–47 (2000)
13. Rico, A., KindlmAnn, P., Sedlácek, F.: Barrier effects of roads on movements of small mammals. Folia Zool. **56**, 1 (2007)
14. Fellows, S., Sharma, G.D., Fellows, A., Khan, I.: Impact of existing national and state highways on wild animals of Pench and Satpura Tiger Reserve. Entomol. Ornithol. Herpetol. **4**, 1 (2015)
15. Forman, R.T., Alexander, L.E.: Roads and their major ecological effects. Annu. Rev. Ecol. Syst. **29**, 207–231 (1998)
16. Hartmann, P.A., Hartmann, M.T., Martins, M.: Snake road mortality in a protected area in the Atlantic Forest of southeastern Brazil. S. Am. J. Herpetol. **6**, 35–42 (2011)
17. Hels, T., Buchwald, E.: The effect of roadkills on amphibian populations. Biol. Cons. **99**, 331–340 (2001)
18. Hibbitts, T.J., Fitzgerald, L.A., Walkup, D.K., Ryberg, W.A.: Why didn't the lizard cross the road? Dunes sagebrush lizards exhibit road-avoidance behaviour. Wildl. Res. **44**, 194–199 (2017)
19. Bhupathy, S., Srinivas, G., Kumar, N.S., Karthik, T., Madhivanan, A.: Herpetofaunal mortality due to vehicular traffic in the Western Ghats, India: a case study. Herpetotropicos **5** (2009)
20. Fahrig, L., Rytwinski, T.: Effects of roads on animal abundance: an empirical review and synthesis. Ecol. Soc. **14** (2009)
21. Gibbs, J.P., Shriver, W.G.: Estimating the effects of road mortality on turtle populations. Conserv. Biol. **16**, 1647–1652 (2002)
22. Lode, T.: Effect of a motorway on mortality and isolation of wildlife populations. AMBIO J. Hum. Environ. **29**, 163–166 (2000)
23. Seshadri, K.S., Yadav, A., Gururaja, K.V.: Roadkills of amphibians in different land use areas from Sharavathi river basin, central Western Ghats, India. J. Threatened Taxa **1**, 549–552 (2009)
24. Baskaran, N., Boominathan, D.: Road kill of animals by highway traffic in the tropical forests of Mudumalai Tiger Reserve, southern India. J. Threatened Taxa **2**, 753–759 (2010)
25. Lesbarreres, D., Fahrig, L.: Measures to reduce population fragmentation by roads: what has worked and how do we know? Trends Ecol. Evol. **27**, 374–380 (2012)
26. Loss, S.R., Will, T., Marra, P.P.: Estimation of bird-vehicle collision mortality on US roads. J. Wildl. Manag. **78**, 763–771 (2014)

27. Pragatheesh, A.: Effect of human feeding on the road mortality of Rhesus Macaques on National Highway-7 routed along Pench Tiger Reserve, Madhya Pradesh, India. J. Threatened Taxa **3**, 1656–1662 (2011)
28. Selvan, K.M.: Observation of roadkills on Kambam-Kumily Road (NH 220) in Tamil Nadu. Zoo's Print **26**, 25–26 (2011)
29. Selvan, K.M., Sridharan, N., John, S.: Roadkill animals on national highways of Karnataka, India. J. Ecol. Nat. Environ. **4**, 363–365 (2012)
30. Seshadri, K.S., Ganesh, T.: Faunal mortality on roads due to religious tourism across time and space in protected areas: a case study from south India. For. Ecol. Manage. **262**, 1713–1721 (2011)
31. Joshi, R., Dixit, A.: Wildlife mortality on national highway 72 and 74 across the Rajaji national park and the Haridwar conservation area, North India. Int. J. Conserv. Sci. **3** (2012)
32. Laurance, W.F., Goosem, M., Laurance, S.G.: Impacts of roads and linear clearings on tropical forests. Trends Ecol. Evol. **24**, 659–669 (2009)
33. Bowne, D.R., Peles, J.D., Barrett, G.W.: Effects of landscape spatial structure on movement patterns of the hispid cotton rat (Sigmodon hispidus). Landsc. Ecol. **14**, 53–65 (1999)
34. dos Santos, C.L.A., Silva, A.P., dos Santos, S.B., Pardini, R., Cassano, C.R.: Dog invasion in agroforests: the importance of households, roads and dog population size in the surroundings. Perspect. Ecol. Conserv. **15**, 221–226 (2017)
35. Jackson, C.R., Fossøy, F., Røskaft, E., May, R.: Do impala really avoid tourist roads? A response to Mulero-Pázmány, D'Amico & González-Suárez (2016). J. Zool. **302**, 219–223 (2017)
36. Laurance, S.G., Stouffer, P.C., Laurance, W.F.: Effects of road clearings on movement patterns of understory rainforest birds in central Amazonia. Conserv. Biol. **18**, 1099–1109 (2004)
37. McGregor, R.L., Bender, D.J., Fahrig, L.: Do small mammals avoid roads because of the traffic? J. Appl. Ecol. **45**, 117–123 (2008)
38. Popp, J.N.: Railways offer grain on a silver platter to wildlife, but at what cost? Anim. Conserv. **20**, 403–404 (2017)
39. Barbosa, N.P., Fernandes, G.W., Carneiro, M.A., Júnior, L.A.: Distribution of non-native invasive species and soil properties in proximity to paved roads and unpaved roads in a quartzitic mountainous grassland of southeastern Brazil (rupestrian fields). Biol. Invasions **12**, 3745–3755 (2010)
40. Gelbard, J.L., Belnap, J.: Roads as conduits for exotic plant invasions in a semiarid landscape. Conserv. Biol. **17**, 420–432 (2003)
41. Joly, M., Bertrand, P., Gbangou, R.Y., White, M.-C., Dubé, J., Lavoie, C.: Paving the way for invasive species: road type and the spread of common ragweed (Ambrosia artemisiifolia). Environ. Manage. **48**, 514–522 (2011)
42. Christen, D.C., Matlack, G.R.: The habitat and conduit functions of roads in the spread of three invasive plant species. Biol. Invasions **11**, 453–465 (2009)
43. Mortensen, D.A., Rauschert, E.S., Nord, A.N., Jones, B.P.: Forest roads facilitate the spread of invasive plants. Invasive Plant Sci. Manage. **2**, 191–199 (2009)
44. Bennett, V.J.: Effects of road density and pattern on the conservation of species and biodiversity. Curr. Landsc. Ecol. Rep. **2**, 1–11 (2017)
45. Brodie, J.F., Giordano, A.J., Ambu, L.: Differential responses of large mammals to logging and edge effects. Mammalian Biology-Zeitschrift für Säugetierkunde **80**, 7–13 (2015)
46. Cerboncini, R.A., Roper, J.J., Passos, F.C.: Edge effects without habitat fragmentation? Small mammals and a railway in the Atlantic Forest of southern Brazil. Oryx **50**, 460–467 (2016)
47. Eldegard, K., Totland, Ø., Moe, S.R.: Edge effects on plant communities along power line clearings. J. Appl. Ecol. **52**, 871–880 (2015)
48. Seiler, A.: Ecological Effects of Roads: A Review. Swedish University of Agricultural Sciences Uppsala (2001)
49. Van der Ree, R., Smith, D.J., Grilo, C.: Handbook of Road Ecology. Wiley (2015)
50. Dutta, T., Sharma, S., McRae, B.H., Roy, P.S., DeFries, R.: Connecting the dots: mapping habitat connectivity for tigers in central India. Reg. Environ. Change **16**, 53–67 (2016)

51. Polak, T., Rhodes, J.R., Jones, D., Possingham, H.P.: Optimal planning for mitigating the impacts of roads on wildlife. J. Appl. Ecol. **51**, 726–734 (2014). https://doi.org/10.1111/1365-2664.12243
52. Matos, C., Sillero, N., Argaña, E.: Spatial analysis of amphibian road mortality levels in northern Portugal country roads. Amphibi. Reptil. **33**, 469–483 (2012)
53. Wilson, D.D.: Hotspot analysis of roadkill in Southern California: a GIS approach. PhD Thesis. California State University, Northridge (2012)
54. da Rosa, C.A., Bager, A.: Seasonality and habitat types affect roadkill of neotropical birds. J. Environ. Manage. **97**, 1–5 (2012)
55. Pragatheesh, A., Rajvanshi, A.: Spatial patterns and factors influencing the mortality of snakes on the National Highway-7 along Pench Tiger Reserve, Madhya Pradesh, India. Oecol. Aust. **17**, 20–35 (2013)
56. Seok, S., Lee, J.: A study on the correlation between road-kill hotspot and habitat patches. J. Environ. Impact Assess. **24**, 233–243 (2015)
57. Zimmermann Teixeira, F., Kindel, A., Hartz, S.M., Mitchell, S., Fahrig, L.: When road-kill hotspots do not indicate the best sites for road-kill mitigation. J. Appl. Ecol. **54**, 1544–1551 (2017)
58. Khatavkar, P.P.: Modelling the Impact of Highways on Impedance of Movement and Mortality Locations of Badger, Meles Meles, and Roe Deer, Capreolus Capreolus: A Case Study of the A73 in The Netherlands. ITC (2009)
59. Jaarsma, C.F., van Langevelde, F., Botma, H.: Flattened fauna and mitigation: traffic victims related to road, traffic, vehicle, and species characteristics. Trans. Res. Part D Trans. Environ. **11**, 264–276 (2006)
60. Coffin, A.W.: From roadkill to road ecology: a review of the ecological effects of roads. J. Transp. Geogr. **15**, 396–406 (2007)
61. Rajvanshi, A., Mathur, V.B., Teleki, G.C., Mukherjee, S.K.: Roads, Sensitive Habitats And Wildlife: Environmental Guideline for India and South Asia. Wildlife Intstitute of India (2001)
62. Shine, R., Lemaster, M., Wall, M., Langkilde, T., Mason, R.: Why did the snake cross the road? Effects of roads on movement and location of mates by garter snakes (Thamnophis sirtalis parietalis). Ecol. Soc. **9** (2004)
63. Vijayakumar, S.P., Vasudevan, K., Ishwar, N.M.: Herpetofaunal mortality on roads in the Anamalai Hills, southern Western Ghats. Hamadryad-Madras- **26**, 253–260 (2001)
64. Menon, V., Daniel, J.C.: Field guide to Indian mammals. Dorling Kindersley, India in association with Penguin Book, India (2003)
65. Habib, B., Rajvanshi, A., Mathur, V.B., Saxena, A.: Corridors at crossroads: linear development-induced ecological triage as a conservation opportunity. Front. Ecol. Evolut. **4**, 132 (2016)
66. Shepard, D.B., Kuhns, A.R., Dreslik, M.J., Phillips, C.A.: Roads as barriers to animal movement in fragmented landscapes. Anim. Conserv. **11**, 288–296 (2008)
67. van der Ree, R., Jaeger, J.A., van der Grift, E.A., Clevenger, A.P.: Effects of roads and traffic on wildlife populations and landscape function: road ecology is moving toward larger scales. Ecol. Soc. **16** (2011)

Conservation and Sustainable Management of Local Hotspots of Biodiversity

T. V. Ramachandra, B. Setturu, S. Vinay, N. M. Tara, M. D. Subashchandran and N. V. Joshi

1 Introduction

Ecosystems are the distinct biological entities that sustain the biosphere and are characterized by a range of functions: nutrient cycling, bio-geochemical cycle, hydrologic cycling. The conservation and sustainable management of ecosystems are the vital components in the pursuit of ecologically sound, economically viable and socially acceptable development goals. This requires an understanding of the complex functioning of ecosystems, diversity of resources, values, ecological services and their significant ability in influencing climate at local as well as global scale. In this regard, an integrated holistic approach considering all components and functions of the ecosystems is quintessential for the developmental planning. Ecosystem conservation has become a challenging task in the face of increasing human pressures due to unplanned activities [1]. Large-scale land-cover transformations have resulted in the enhanced instances of human–animal conflicts, conversion of perennial streams to seasonal streams and affected the livelihood of dependent population with the impaired biological and economic productivities [2]. Decision-making based on the biophysical, economic and socio-cultural information provides an opportunity to overcome these constraints while ensuring the sustainability of natural resources [3]. Sustainable landscape planning aims for stability in ecological, physical and social systems (cultural, economic functions) by maintaining the sustainability of natural resources with intergeneration equity [1]. Prioritization of sensitive regions for conservation [4] through a multidisciplinary approach is widely accepted norm to identify hotspots of biodiversity. Ecologically sensitive region (ESR) is a bio-climatic unit (as demarcated by entire landscape) wherein human impacts may cause

T. V. Ramachandra (✉) · B. Setturu · S. Vinay · N. M. Tara · M. D. Subashchandran · N. V. Joshi
Energy & Wetland Research Group, CES TE 15, Centre for Ecological Sciences,
Indian Institute of Science, 560012 Bengaluru, India
e-mail: tvr@iisc.ac.in; energy.ces@iisc.ac.in
URL: http://ces.iisc.ernet.in/energy

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_27

irreversible changes in the structure of biological communities (as evident in number/composition of species and their relative abundances) and their natural habitats. A range of conservation actions being practiced includes protecting altitudinal gradients, sacred patches of forests, riverine corridors [5, 6] and participatory or incentive-based instruments at the local scale [7, 8]. In addition, the local conservation endeavours involving effective strategic landscape planning processes help in mitigating the impacts of climate changes [9, 10].

The spatial conservation planning considers ESR based on both ecological and cultural dimensions. Ecological dimension refers to the natural environment such as ecosystems and ecological processes, while cultural dimension refers to the political, social, technological and economic aspects. In India, section 5(1) of Environment Protection Act 1986 (EPA), the Ministry of Environment, Forests and Climate Change (MoEFCC) stipulates the location of industries or implementation of developmental projects based on the ecological sensitivity or fragility of a region considering permanent and irreparable loss of extant life forms, or significant damage to the natural processes of evolution and speciation [11]. *E*co-sensitive regions have to be determined based on biological, economic, socio-cultural values depending upon the context and the area or location for conservation [12]. ESRs are the 'unique' areas of ecological and economic importance, vulnerable to even mild disturbances, irreplaceable if destroyed and hence demand conservation [13]. Geo-informatics fortified with free and open-source software (FOSS) has gained significance in recent times due to the contribution to spatial conservation planning of a region by providing a consistent spatial analytical visualization and modelling abilities for an understanding of ecological systems [14, http://wgbis.ces.iisc.ernet.in/grass/; http://wgbis.ces.iisc.ernet.in/foss/]. Spatial decision support tools are playing an important role in increasing accountability and transparency of the planning process and leading to more economically efficient conservation actions [15]. The objective of the current endeavour is to develop spatial decision system to prioritize ecologically sensitive regions based on ecological, biological, social and geo-climatic attributes. This involved (i) demarcation of local hotspots of biodiversity for conservation based on biotic, abiotic and social criteria with an integrated biodiversity database and management prescriptions to beneficiaries at every level from the village communities to the Government; (ii) compilation of primary data related to biodiversity, ecology, energy, hydrology and social aspects and (iii). Development of a comprehensive management framework with measures to mitigate forest loss and attain sustainable growth and support to preserve biodiversity.

2 Materials and Method

2.1 Study Area

The Western Ghats, a rare repository of endemic flora and fauna is one of the 35 hotspots of global biodiversity and a home to diverse social, religious and linguistic group. Uttara Kannada district located in the central Western Ghats (Fig. 1) lies between 13.769°–15.732°N and 74.124°–75.169°E covering approximately an area of 10, 291 km^2. The region has the distinction of having highest forest area (80.48%) in Karnataka State, India, and has been undergoing severe anthropogenic pressures impacting biogeochemistry, hydrology, food security, climate and socio-economic systems. The district has varied geographical features with thick forest, perennial rivers and abundant flora, fauna. It has the unique distinction of having 3 agro-climatic zones and for the regional administrative purpose, the district is further divided into 11 taluks (also known as tehsil or mandal is an agglomeration of villages). The coastal region, which has hot humid climate and rainfall, varies between 3000 and 4500 mm. The Sahyadri interior region of the Western Ghats (500–1000 m high)is very humid to the south (rainfall varies from 4000 to 5500 mm). The plains are regions of transition, which are drier (rainfall varies between 1500 and 2000 mm).

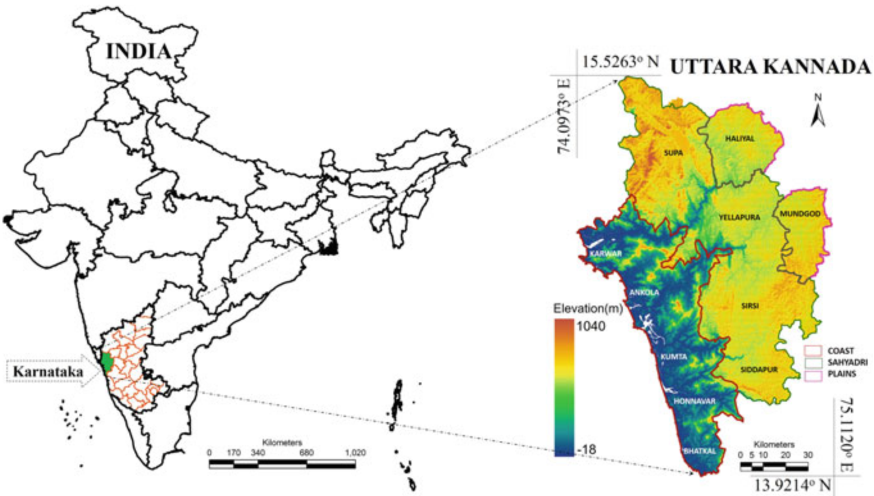


Fig. 1 Study area: Uttara Kannada district, Karnataka state, Central Western Ghats

2.2 Method

Local hotspots of biodiversity or ESRs in the district are prioritized considering biological (terrestrial and aquatic flora and fauna, estuarine biodiversity), ecological (diversity, endemism, conservation reserve), geo-climatic (altitude, slope, rainfall), renewable energy prospects (bio, solar, wind), social (population, forest-dwelling communities) as outlined in Fig. 2. The study area has been divided into $5' \times 5'$ equal area grids (168) covering approximately 9×9 km^2 (Fig. 3) for prioritizing ESR.

Table 1 lists the weightages assigned to each variable of various themes considering the minimal impact on the landscape and also to prioritize conservation regions for future planning. The weightages were assigned iteratively across the landscape with varied themes for a development solution and monitoring.

Developing a weightage metric score analysis requires knowledge of multidisciplines [16], and planning integrates the present and future needs in the landscape [17], and weightage is given by

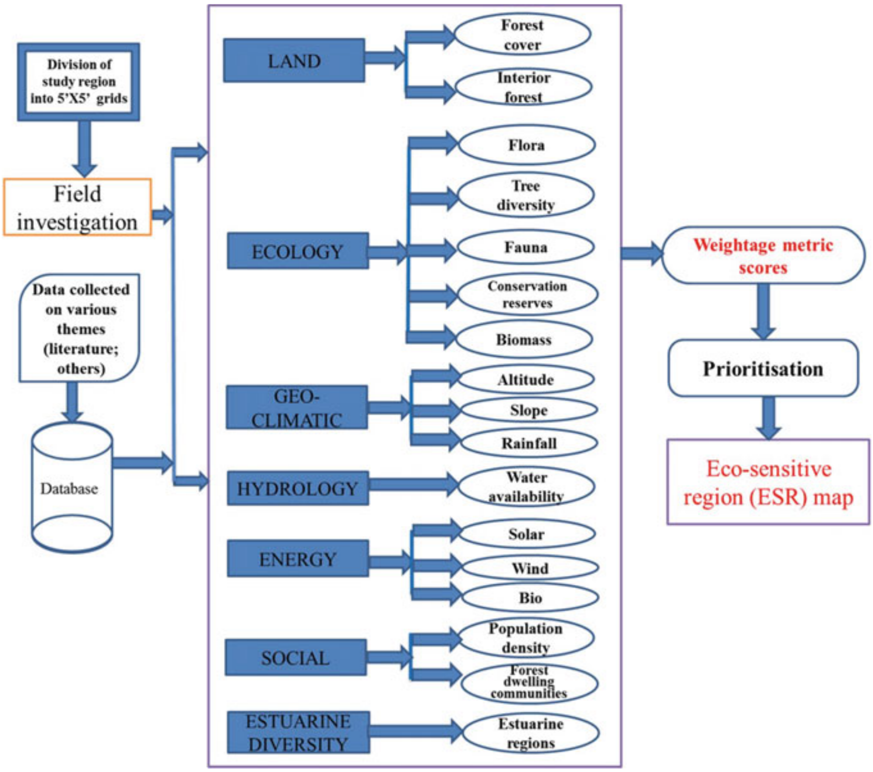


Fig. 2 Weightage metric criteria for prioritizing ESR

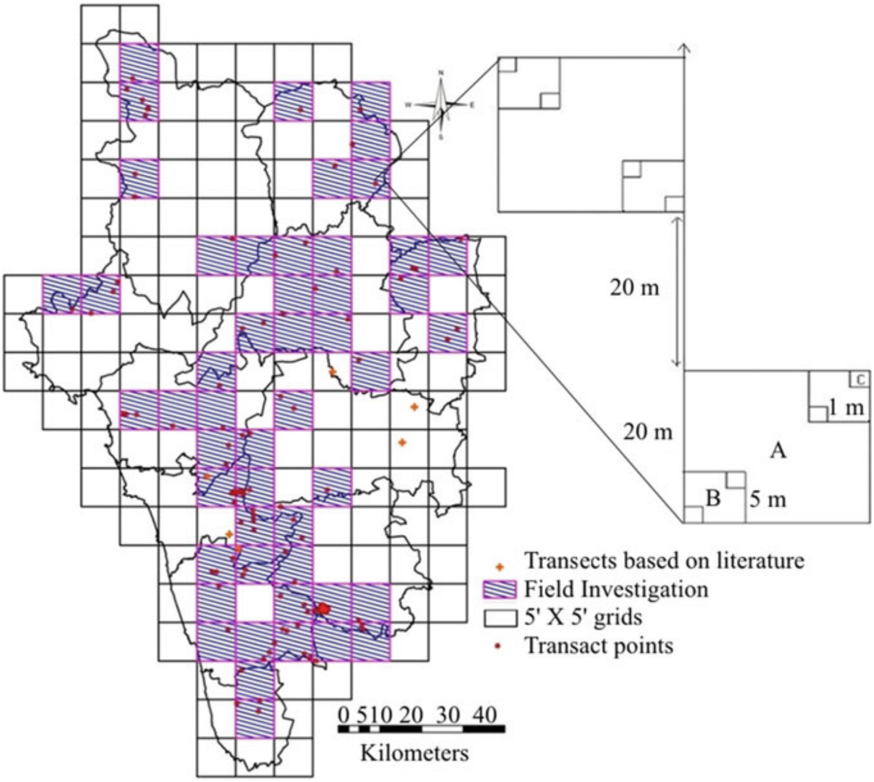


Fig. 3 Grids with the distribution of transects and transect cum quadrats (2 of 5 quadrats of 20 × 20 m only shown). Shaded grids are the representative grids chosen based on agro-climatic zones for field data collection

$$Weightage = \sum_{i=1}^{n} W_i V_i \quad (1)$$

where n is the number of data sets (variables), V_i is the value associated with criterion i, and W_i is the weight associated with that criterion. Table 1 expresses the theme-wise decision variable considered with their level of significance, ranked between 1 and 10. Value 10 corresponds to highest priority for conservation, whereas 7, 5 and 3 correspond to high, moderate and low levels of prioritization. Assigning weightages based on individual proxy based extensively on GIS techniques has proved to be the most effective for prioritizing ESR. Visualization of levels of ESR helps the decision-makers in opting eco-friendly development measures. A detailed database has been created for various themes covering all aspects from land to estuarine ecosystem. The theme-wise description given below highlights the consideration of variables for study and their significance in conservation priority.

Table 1 Various themes considered and their weightages

S. no	Themes	Weightages/ranking					Theme
		1	3	5	7	10	
1.	Land-use	FC < 20%	20 < FC < 40%	40 < FC < 60%	60 < FC < 80%	FC > 80%	Land
	Interior forest	IF < 20%	20 < IF < 40%	40 < IF < 60%	60 < IF < 80%	IF > 80%	
2.	Flora	NEND	END < 30%	30 < END < 50%	50 < END < 70%	END > 70%	Ecology
	Tree diversity	SHD < 2	2 < SHD < 2.5	2.5 < SHD < 2.7	2.7 < SHD < 3	SHD > 3	
	Fauna	–	NEND	–	–	END	
	Conservation reserves (CR)	–	–	–	–	National parks, wildlife reserves, Myristica swamps, Sanctuaries	
	Biomass (Gg)	BM < 250	250 < BM < 500	500 < BM < 750	750 < BM < 1000	BM > 1000	
3.	Altitude slope	–	–	–	Slope > 20%	Slope > 30%	Geo-climatic
	Precipitation	–	1000 > RF > 2000 mm	2000 > RF > 3000 mm	3000 > RF > 2000 mm	RF > 4000 mm	
4.	Stream flow	WA < 4	4 < WA < 6	6 < WA < 9	9 < WA < 12	WA = 12	Hydrology
5.	Solar	–	–	<5 KWh/m^2/day	5–6 KWh/m^2/day	6–6.5 KWh/m^2/day	Energy
	Wind	–	–	2.4–2.55 m/s	2.5–2.6 m/s	2.6–2.7 m/s	
	Bio	SD < 1	SD > 1	1 > SD < 2	2 < SD < 3	SD > 3	
6.	Population density (PD)	PD > 200	100 < PD < 200	100 < PD < 150	50 < PD < 100	PD < 50	Social
	Forest-dwelling communities (tribes)	–		Tribes are present then assigned 10; if no tribal population exists, then assigned as 0			
7.	Estuarine regions	–	Low	Moderate	High	Very high	Estuarine diversity

FC–forest cover; IF–interior forest cover; END–endemic; NEND–non-endemic; BM–biomass; SD–supply to demand ratio; WA–water availability

2.2.1 Land

Land uses based on the analysis of remote sensing data were considered, and grids were prioritized based on the proportion forest cover [18]. Forest fragmentation statistics is computed as per the standard protocol [18, 19]. The interior forest cover refers to the undisturbed core forest patches that aid in preserving the structure of the ecosystem while enhancing its functional aspects.

2.2.2 Ecology

Field investigations were carried out in 116 sample transects (Fig. 4) for data on the plant species diversity, basal area, biomass, estimates of carbon sequestration, percentage of evergreenness and Western Ghats endemism and about the distribution of threatened species, etc. Along a transect length ranging up to 180 m, quadrats each of 20 × 20 m were laid alternatively on the right and left, for tree study (minimum girth of 30 cm at girth at breast height (GBH) or 130 cm height from the ground), keeping intervals of 20 m length between successive quadrats. A number of quadrats per transect depended on species–area curve, and most transects had a maximum of five quadrats. Within each tree quadrat, at two diagonal corners, two sub-quadrats of 5 m × 5 m were laid for shrubs and tree saplings (<30 cm girth). Within each of these two herb layer quadrats, 1 m^2 area each was also laid down for herbs and tree seedlings. Supplementary data were compiled through the review of published literature, unpublished datasets and ground-based surveys other than transects. Approaches adopted in documenting flora and fauna are outlined by earlier studies [20].

The health of ecosystem and its significance is derived based on the key variables—endemism, floral diversity, evergreenness, etc., for evolving the composite conservation index. Tree species diversity is another measure calculated using Shannon's diversity index (H′). H′ is given by Eq. 2 [21, 22].

$$(H)' = -\sum_{i=1}^{n} (p_i) \ln p_i \qquad (2)$$

where i is the proportion of the species relative to the total number of species (p_i) multiplied by the natural logarithm of this proportion (ln p_i) and the final product

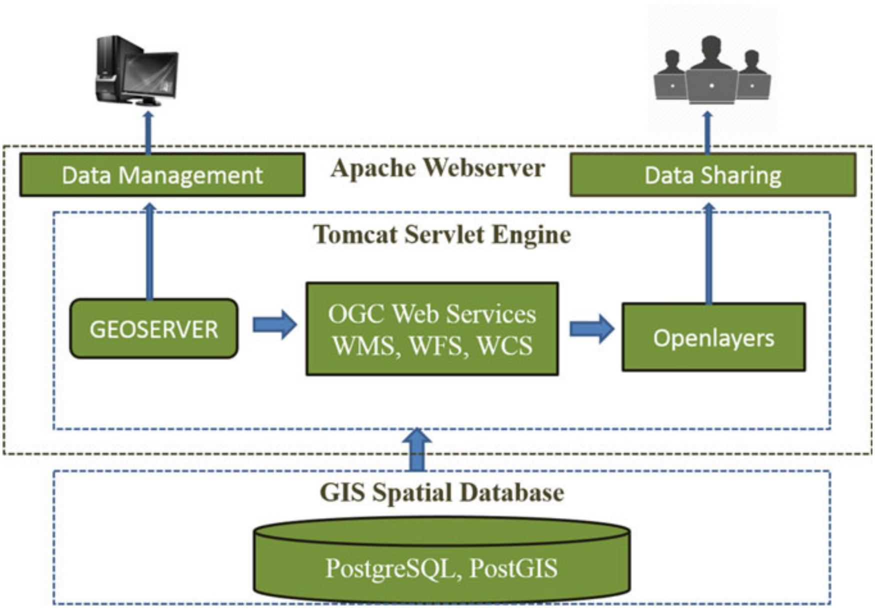


Fig. 4 Framework of the SDSS

multiplied by −1. The Shannon index ranges typically from 1.5 to 3.5 and rarely reaches 4.5. Higher diversity range was assigned higher weightage for conservation.

Faunal diversity is another surrogate variable used to assess the eco-sensitivity of a region. The region is storehouse of endemic fauna, in which occurrence of endemic species increases in the undulating terrains of upper Ghats. Species richness and endemism are two key attributes of biodiversity that reflect the complexity and uniqueness of natural ecosystems [4, 23]. The setting of regional conservation priorities based on combinations of modelling individual endemic species' distributions, evaluating regional concentrations of species richness and using complementarity of areas by maximizing inclusion of species in the overall system is most appropriate [24]. A set of criteria for prioritizing the regions has been prepared based on field investigation, interaction with stakeholders (researchers working in this region, forest officials, local people, subject experts).

Mammals are well represented in this chain of mountains, and many endemic birds are found in all other places of the district. Conservation Reserves (CR) are being established under the framework of Protected Areas (PA) under the Wildlife (Protection) Amendment Act of 2002. CRs are typically buffer zones or connectors and migration corridors between National Parks, Wildlife Sanctuaries and reserved protected forests in the district. These reserves protect habitats that are under private ownership also, through active stakeholder participation. The biological diversity in these zones like National parks, Sanctuaries (Anshi Dandeli Tiger reserve (ADTR)), botanical gardens (Shalmala Riparian Ecosystem Conservation Reserve, Aghanashini LTM Conservation Reserve, Hornbill Conservation Reserve, Attiveri Bird Sanctuary), zoological gardens hosts threatened (rare, vulnerable, endangered) flora/fauna. Higher weightage is assigned for CRs.

2.2.3 Biomass

Biomass is another important indicator of forest health and reveals its role in a global carbon sink. Most of the Uttara Kannada district is located in the high rainfall zone, except Mundgod and eastern parts of Haliyal and Yellapura support trees with higher biomass. Details of biomass quantification, flora and fauna diversity are available in the literature [16, http://wgbis.ces.iisc.ernet.in/biodiversity/database_new/]. The analysis has calculated total standing biomass of forest's vegetation [25, 26] based on field data and remote sensing data. Transect-wise basal area per hectare was estimated using allometric equations.

2.2.4 Geo-Climatic Variables

Geo-climate plays a major role in determining the speed of recovery (lag-time) of a landscape (and the ecosystem that governs it), and the studies reveal that variables such as altitude (elevation, slope, rainfall), easterly aspect, steepness and longer dry seasons have a significant role in local ecology [27]. The patterns of altitude, slope and rainfall bring about the sensitivity, heterogeneity, complexity of climate, soil, vegetation, land use, land cover in connection with socio-economic interactions [28, 29]. The elevation map is generated using Cartosat DEM of 1 arc-second resolution. Areas with steep slopes and high altitudes are likely to be eroded more easily, and hence vulnerable to natural erosion or landslides, need to be considered as least resilient and hence environmentally sensitive zones areas. The analysis has considered that the slopes and altitudes can be normalized within each grid from 0 (least average slope or lowest average altitude) to 10 (high slope and high altitude) and assigned to the grids. The slope map is generated from DEM dataset using Geographical Resources Analysis Support System (GRASS)—http://wgbis.ces.iisc.ernet.in/grass/index.html)—free and open-source tool.

Hydrology provides a fundamental basis for understanding material flows, environmental quality and stream ecosystem in a basin [30]. Conservation of high biodiversity forest landscapes is justified on the basis of hydrological benefits—in particular, reduction of flooding hazards for downstream floodplain populations [31]. Point-based daily rainfall data from various rain gauge stations in and around the study area between 1901 and 2010 were considered for analysis of rainfall [20, 32]. The rainfall data used for the study were obtained from Department of Statistics, Government of Karnataka; Indian metrological data (IMD), Government of India. Rainfall trend analysis was done for selected rain gauge stations to assess the variability of rainfall at different locations in the study area. Monthly monitoring of hydrological parameters reveals that streams in the catchments with undisturbed primary forest (evergreen to semi-evergreen and moist deciduous forests with spatial extent >60% in the respective catchment) cover have reduced run-off as compared to catchments with disturbed/altered forest covers. Run-off and thus erosion from monoculture plantation forests were higher from that of natural forests. Forested catchment has higher rates of infiltration as soil is more permeable due to enhanced microbial activities with higher amounts of organic matter in the forest floor. Streams with undisturbed forest cover (vegetation of native species) in the catchment showed a good amount of dry season flow. Native forests in the catchment aid as sponge retaining the water, while allowing infiltration during monsoon, which are steadily released during the lean season. Compared to this, streams in the catchment dominated by agricultural and monoculture plantations (of Eucalyptus sp. and *Acacia auriculiformis*) are seasonal with water availability ranging between 4 and 6 months. The grids where water is available during all months in a year (perennial flow) are assigned higher values.

2.2.5 Energy

Dependence on the conventional energy resources for electricity generation is eroding the natural resources at faster rate by causing a significant adverse effect on ecology by producing enormous quantities of by products including nuclear waste and carbon dioxide. Improving energy efficiency, switch over to renewable sources of energy and de-linking economic development from energy consumption (particularly of fossil fuels) is essential for sustainable development of a region. Potential of renewable energy sources is assessed (Solar, Wind, Bioenergy) month-wise and captured the variations [33–35]. The solar energy datasets are derived based on NASA's Surface Meteorology and Solar Energy (SSE) methodology. The solar energy is available greater than 10 months with higher potential. The availability of wind energy and its characteristics of Uttara Kannada District have been analysed based on primary data collected from India Meteorological Department (IMD) observatories. Wind energy conversion systems would be most effective during the period May to August. Energy pattern factor (EPF) and power densities are computed which show that the coastal taluks such as Karwar, Ankola and Kumta have good wind potential [34]. The households' survey carried out to understand the spatio-temporal patterns in the domestic fuelwood consumption reveals that 82–90% of the households still depend on fuelwood and agro-residues. Analyses of sector-wise contribution in the energy surplus zones show that horticulture residues contribute in the central dry zone, southern transition zone and the coastal zone, while in the hilly zone, forests contribute more towards the available bioenergy [35]. The adaptation of green technologies would aid in cutting down carbon footprint. Weightages are assigned based on the level and quantum of availability of energy from renewable resources.

2.2.6 Social Aspects

Forest Rights Act 2006, Government of India, seeks to recognize and vest the forest rights and occupation in forest land in forest-dwelling Scheduled Tribes and other traditional forest dwellers who have been residing in forests for generations but whose rights could not be recorded. Forest-dwelling communities (tribes) of the district are mapped at village level, and the grids with tribal population are assigned higher weightage. In the regional planning, demographic aspect is essential to many applications across the science and policy domains including assessment of human vulnerability to environmental changes. Land degradation is due to population pressure which leads to intense land-use conversions without proper management practices. Increase in population density will lead to the increasing exploitation of natural resources and the resulting loss of species and ecosystem richness, nature conservation [36]. Village-wise population density is computed considering 2011 population census data (http://censusindia.gov.in). Population density per sq. km is considered as one of the influencing social factors for prioritization, and the grids with lower population density are assigned higher weightage. The need for combining nature conservation with social aspect is to emphasize receiving a livelihood from natural resources and participation in enriching biodiversity.

2.2.7 Estuarine Diversity

Estuarine ecosystems are biologically productive, socio-economically vital and aesthetically attractive, while providing food and shelter for many vital biotic species, and some are commercially very important [37]. West coast estuaries of the district were assessed based on productivity, biodiversity and human pressure [38–40]. The analysis has identified the mangroves at species level using remote sensing data with field-based measurements. Estuarine productivity based on goods and services of the district [38] brings out the disparity in productivity and diversity between the neighbouring estuaries due to major human intervention in the form of construction of hydroelectric projects in upstream. Estuaries were given weightages based on the productivity and diversity.

2.2.8 SDSS Framework

SDSS integrates the scientific data in addressing the problems and provides appropriate solutions for sustainable utilization of the resources. OGC provide standardized interface specifications to support geospatial data sharing and interoperability among Web-based GIS systems (Fig. 4). SDSS Server GIS framework (i) is used remotely as data management done by the researchers and administrators, (ii) provides access to functionality via Web protocols such as the OGC Web Processing Service and (iii) allows the users to access the data and enter input parameters. Web-mapping API such as Openstreet Maps which is one of the popular Web-mapping application programming interfaces (APIs) and Bhuvan (http://bhuvan.nrsc.gov.in), WMS layers are used in our framework, and other examples are Google Maps (http://maps.google.co.in), Yahoo! Maps (http://maps.yahoo.com) or Bing Maps (http://bing.com/maps). The online user will be able to access the graphical user interface (http://10.58.20.79/ol3/ukwms.html) and choose the different map layers [41]. When the request is received, the Web server communicates with the GeoServer to retrieve the map layers as a service which will be fetched from the database and a response is sent to the user through the GUI. The user will be able to visualize the information on the maps.

3 Results and Discussion

ESRs in the district are prioritized considering biological, ecological, geo-climatic, renewable energy and social prospects. Weightages are assigned to the grids for prioritizing eco-sensitiveness based on the relative significance of themes based on the aggregate metric score as ESR 1 (regions of highest sensitivity), ESR2 (regions of higher sensitivity), ESR3 (regions of high sensitivity) and ESR4 (regions of moderate sensitivity), respectively. Land use of 2013 is assessed using remote sensing data of Landsat ETM+ sensor 30 m resolution. Land-use analysis reveals that the region has about 32.08% under evergreen–semi-evergreen forests (Table 2), and higher forest cover (>80%) is confined to the grids in Sahyadri region (Supa, Yellapura, Ankola, Sirsi taluks). The coastal taluks are having forest cover in the range 60–80% towards eastern part, whereas western side totally degraded due to higher pressure. The plains show least cover (<20%) reflecting higher degradation, and the natural forest cover in the district is only 542,475 Ha. The land clearing and subsequent

Table 2 Land use and fragmentation of forests in Uttara Kannada

Category	Land-use analysis		Fragment type	Spatial extent	
	Ha	%		Ha	%
Built-up	31589	**3.07**	Transitional	59435	5.78
Water	28113	2.73	Perforated	8909	0.87
Cropland	145395	14.13			
Open fields	37660	3.66	Patch	30618	2.98
Moist deciduous forest	**161996**	**15.74**			
Evergreen to semi-evergreen forest	**330204**	**32.08**	Edge	179870	17.48
Scrub/grass	40402	3.93			
Acacia/Eucalyptus/hardwood plantations	122927	**11.94**	Interior	**263643**	**25.62**
Teak/bamboo/softwood plantations	67111	**6.52**			
Coconut/arecanut/cashew nut plantations	53993	**5.25**	Non-forest area	**486611**	**47.3**
Dry deciduous forest	9873	0.96			
Total area (Ha)	**1029086**				

agricultural expansion, exotic plantations resulted in the degradation of large forest patches at temporal scale. Weightages were assigned to the grids based on the extent of forest cover, and grids in Sahyadri region have highest ranking (10) compared to plains (1). Fragmentation analysis considering the spatial extent of forests reveals that contiguous forests (interior forests) cover only 25.62%, and land use under non-forest categories (cropland, plantations, built-up, etc.) covers 47.29% of the landscape across coast, Sahyadri and plains.

Flora and fauna of terrestrial and aquatic ecosystems have been studied through field investigations and compilation of information from published literature. These strategies helped in documenting 1068 species of flowering plants, representing 138 families. Grids in Honnavar, Kumta, Sirsi, Bhatkal, Siddapur are with higher weights, and Mundgod and Haliyal show least endemism [29].

Analysis of faunal distribution shows that tiger (*Panthera tigris*), leopard, wild dog (dhole) and sloth bear are the main predators. The district is a paradise for birds; 272 birds are listed in the Dandeli, out of which 19 are considered to be endemic [36]. The distribution of freshwater fishes is highly correlated to terrestrial landscape elements, of which quantity and quality of evergreen forests are more important [2]. Higher weightages (10) are assigned to the grids with endemic species, and least (3) were assigned for grids with non-endemic fauna.

Biomass is estimated grid-wise, based on the spatial extent of forest and per hectare basal area. The total biomass of the district is **113823.58** Gg, with Sahyadrian taluks such as Supa, Sirsi and Yellapura having greater biomass (>1200 Gg) followed by the coastal taluks (Karwar, Ankola, Kumta, Honnavar). Grids with higher standing biomass regions were assigned higher weightages [26, 39], as these regions help in maintaining global carbon through sequestration. Tree diversity is computed through the Shannon diversity index which shows that most evergreen to semi-evergreen forests with diversity values ranging between 3 and 4. Uttara Kannada district has two important protected areas, namely Anshi National Park and Dandeli Wildlife Sanctuary, which are assigned higher weights as they are key eco-sensitive regions with diverse biodiversity [40].

Geo-climatic variables such as altitude, slope and rainfall are analysed to identify sensitive zones. Highest elevation is 758 m in Supa taluk. Grids with elevations >600 m as higher priority for conservation and > 400 m is moderate and rest is of least concern. Rainfall pattern shows that the district falls in the high rainfall zone, except Mundgod and eastern parts of Haliyal, Yellapura. Grids are assigned weights based on the quantum and duration of rainfall [40]. The sub-basin-wise field investigations were carried out to account perennial, seasonal flows of the region. Hydrological regime analysis reveals the existence of perennial streams in the catchment dominated by diverse forests with native vegetation (>60% cover) compared to the streams in the catchments of either degraded forests or dominated by monoculture plantations [32]. Grids in Sahyadri regions show 12-month water availability in the streams and were assigned higher weightages. Streams in Haliyal, Mundgod, eastern part of Yellapura have flow of only 4 months due to scarce rainfall and monoculture plantations.

Environmentally sound alternative sources of energy resources (solar, wind, io) potential were considered for prioritization [26, 33–36]. The region receives an average solar insolation of 5.42 kWh/m^2/day annually and has more than 300 clear sunny days. Wind resource assessment shows wind speed varies from 1.9 m/s (6.84 km/hr.) to 3.93 m/s (14.15 km/hr.) throughout the year with a minimum in October and maximum in June and July. Bioresource availability is computed based on the compilation of data on the area and productivity of agriculture and horticulture crops, forests and plantations. Sector-wise energy demand is computed based on a primary household survey of 2500 households, the National Sample Survey Organization (NSSO study) data and the information compiled from the literature. The supply/demand ratio in the district ranges from less than 0.5 to greater than 2. Sirsi, Siddapur, Yellapur, Supa and eastern hilly areas of Kumta, Honnavar and Ankola are fuelwood surplus regions. Hybridizing wind energy systems with other locally available resources (solar, bioenergy) would assure the reliable energy supply to meet the energy demand at decentralized levels, and weights were assigned based on the availability [33–36]. The location of forest-dwelling communities such as *Kunbis, Siddis, Goulis, Gondas* was spatially mapped, and the respective grids were assigned highest weights, because these people are directly and indirectly dependent on forest resources and have been protecting forests. Grid-wise population is computed by aggregating villages in the respective grid for 2011. Population density is computed for each grid and weightages were assigned. Grids with the lowest population density (<50 persons) were assigned higher weight (considering the likely lower anthropogenic stress) and vice versa [40, 41].

The four major estuaries, viz. Kali, Gangavali, Aghanashini, and Sharavathi, are rich in mangrove species diversity and vital for fishery and cultivation of Kagga rice (salt tolerant) varieties. The biological diversity analysis shows Aghanashini and Ganagavali estuaries have higher fish diversity and mangrove species due to the absence of major anthropogenic activities (dam or hydro projects). Estuaries such as Sharavathi and Kali are severely disturbed with unplanned developmental activities [38, 39], which has affected the productivity of livelihood resources (fish, bivalves, etc.). Coastal grids were assigned weightages based on the biological diversity and productivity (considering provisional goods—fish, bivalves, sand and salt).

Aggregation of these spatial layers corresponding to biological, ecological, geo-climatic, renewable energy and social variables aided in prioritising the grids as ESR 1, ESR 2, ESR 3 and ESR 4, respectively, (Fig. 5a) based on the composite metric score. Spatially, 52.38% of the district represents ESR 1, 14.29% of area represents ESR 2, 13.1% of area represents ESR 3 and about 20.23% of the district is in ESR 4. Figure 5b depicts ESR with taluk and gram panchayat (decentralized administrative units with a cluster of few villages) boundaries. Uttara Kannada district has 11 taluks and 209 panchayats. ESR analyses reveal that ESR 1 consists of mainly Supa, Yellapura, Ankola, Sirsi, Siddapura, Honnavar and Kumta taluks. Considering Panchayat-level analyses, 102 panchayats are in ESR 1, while 37 panchayats in ESR 2, 33 panchayats in ESR 3 and 37 panchayats in ESR 4. Sahyadri and eastern part of coastal regions represent highest ecological sensitiveness. ESR 2 is as good as ESR 1, except degradation of forest patches in some localities. ESR 3 represents moderate conservation region, and only regulated development is allowed in these areas. ESR 4 represents less sensitiveness.

The visualization is implemented through open layers by adding the WMS layer. Figure 6 visualizes layer of Western Ghats boundary, Western Ghats states and districts, Uttara Kannada Panchayats boundary on the backend layer of OpenStreetsMap and also the land-use WMS layer of Bhuvan. The user can choose different layers using the checkbox option and view accordingly. This information contributes to analyzing and utilizing the resources in an efficient way, which helps the decision-maker or the concerned citizen to use the data to make better plans and policies. SDSS aids users to visualize diverse themes of land, ecology, energy, socio, hydro and estuarine variables of rich biodiversity hotspots and also provides an opportunity to integrate ecological and socio-economic aspects in decision-making. The 73rd amendment to the constitution (1992) empowers local governing bodies to make relevant plans for the socio-economic development of a region. Inclusive growth enhances social capital for the public can be achieved by ensuring the active and effective participation of all sections of society at every level of governance. The implementation of SDSS at local levels would help in realizing the vision of Biodiversity act, 2002, which empowers Biodiversity Management Committees (BMC) at panchayat with the knowledge of local biodiversity richness with ecological status to take decisions towards the prudent use of natural resources.

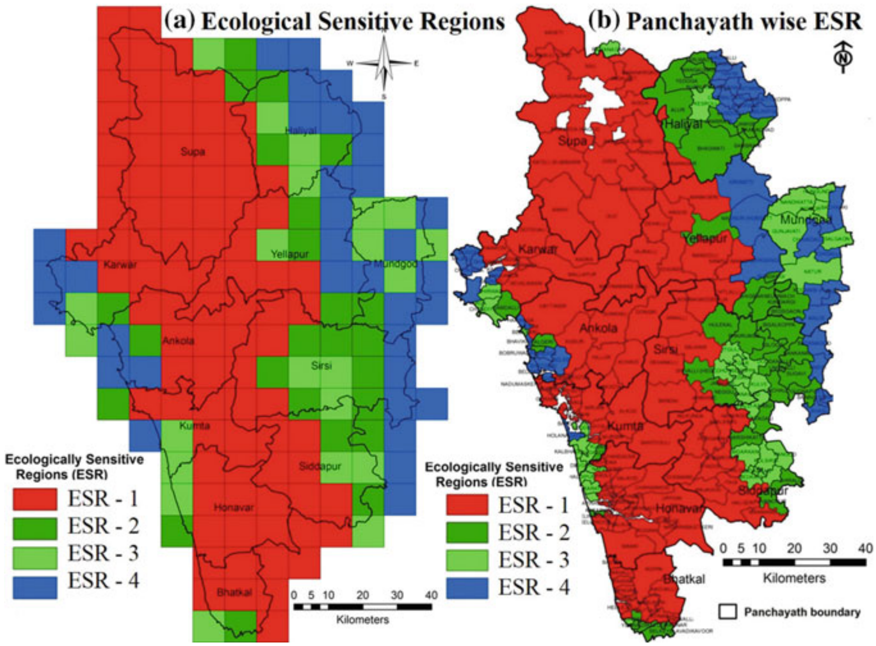
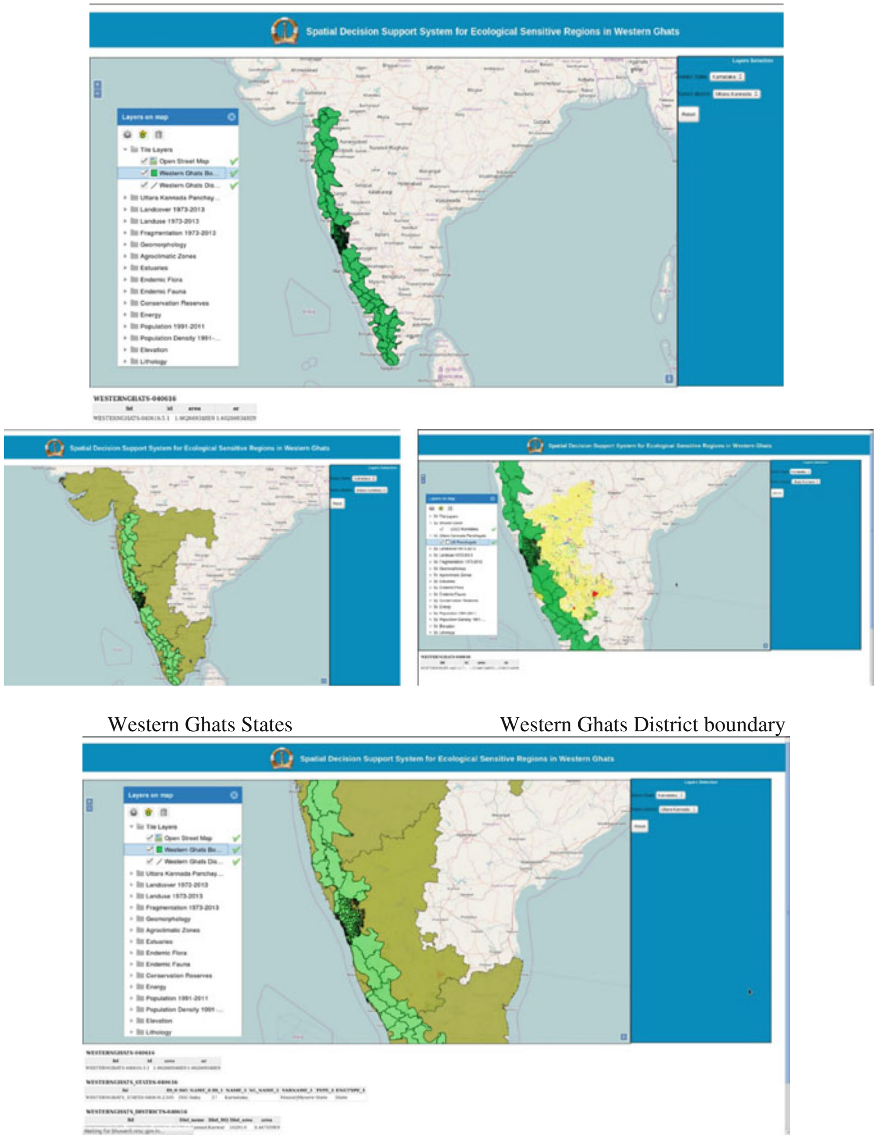


Fig. 5 Ecological sensitive regions of Uttara Kannada at panchayat level

4 Conclusion

ESRs are the eco-clusters or the 'ecological units' that may be easily affected or harmed. The ESR prioritization (ESR 1–4) via varied themes (biological, geo-climatic, social, etc.) at panchayat level is a major step towards an ecological audit that eventually results in the conservation and sustainable use of biodiversity. Spatially, 52.38% of the district represents ESR 1, while 14.29% of area represents ESR 2, 13.1% of area represents ESR 3 and about 20.23% of the district is in ESR 4. Regions under ESR 1 and ESR 2 are "no-go area" for any developmental activities involving large-scale land-cover changes. ESR 2 has eco-sensitiveness similar to ESR 1 and has scope to attain the status of ESR 1 with eco-restoration measures. The persistence of the endemic (rare, threatened, etc.) species in ESR 1 and ESR 2 calls for serious attention from conservationists and decision-makers to initiate programs immediately for conservation. The fact that watercourse forests have not only rare species but also high biomass and greater carbon sequestration potential also calls for revision of forest management policies, as the innumerable stream courses of Western Ghats offer tremendous potential for carbon stocking per unit area while also bettering the hydrology of these mountains, which form the main watershed for the entire Indian peninsula. Rendering such service for mitigating global climatic change can also, same time, serve well the cause of eco-sensitive



Western Ghats States Western Ghats District boundary

Overlay of Western Ghats States, District Boundaries and Uttara Kannada Panchayats

Fig. 6 Visualization spatial layers of Western Ghats

regions in an otherwise much-impacted biodiversity hotspot. Conservation and sustainable management of natural resources in the Western Ghats will ensure food and water security in the peninsular India and intergeneration equity. Hence, the premium should be on conservation of the remaining ecologically sensitive regions, which are vital for the water security (perenniality of streams), mitigation of global carbon and food security (sustenance of biodiversity). There still exists a chance to restore the lost natural evergreen to semi-evergreen forests in the Western Ghats region through appropriate conservation and management practices. The management of biodiversity hotspot regions should focus on the conservation as well as socio-economic developmental aspects. This approach aids in the conservation of ecology, biodiversity, water resources, culture and traditions while paving way for location-specific economic development, primarily aimed at elevating levels of livelihood security. The outcomes visualize an ongoing process and integrate ecological and environmental considerations into administration in the biodiversity-rich regions. It is shown that eco-clusters are crucial for a sustainable development and thus need political commitment and incentives for the development of eco-industry sector (based on the local renewable natural resources). Thus, eco-clusters will aid as catalysts in a well-ordered decision-making process through stake holder's active participation with the priorities for sustainable livelihood.

Acknowledgements We are grateful to (i) NRDMS Division, the Ministry of Science and Technology (Grant: DST/CES/TVR/1045—Geo-visualization of local hotspots of biodiversity, REF: NRDMS/11/1157/09, 28 Feb 2011—study during 2013–2017), Government of India and (ii) Indian Institute of Science (IISc/R1011) for data and the sustained financial and infrastructure support for ecological and energy research. We acknowledge the support of Karnataka Forest Department, Western Ghats Task Force, Karnataka Biodiversity Board, Government of Karnataka for giving necessary permissions to undertake ecological research in Central Western Ghats.

References

1. Ramachandra, T.V., Chandran, M.D.S., Gururaja, K.V., Sreekantha: Cumulative Environmental Impact Assessment. Nova Science Publishers, New York (2007)
2. Berkres, F., Davidson-Hunt, I.J.: Biodiversity, traditional management systems, and cultural landscapes: examples from the boreal forest of Canada. Int. Soc. Sci. J. **58**, 35–47 (2006)
3. Watson, J.E.M., Grantham, H., Wilson, K.A., Possingham, H.P.: Systematic conservation planning: past, present and future. In: Whittaker, R., Ladle, R. (eds.) Conservation Biogeography, pp. 136–160. Wiley-Blackwell, Oxford (2011)
4. Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A., Kent, J.: Biodiversity hotspots for conservation priorities. Nature **403**(6772), 853–858 (2000)
5. Watson, J.E.M., Cross, M., Rowland, E., Joseph, L.N., Rao, M., Seimon, A.: Planning for species conservation in a time of climate change. In: Climate Change, Research and Technology for Climate Change Adaptation and Mitigation, vol. 3, pp. 379–402. InTech Publishers (2011). ISBN 979-953-307-278-3
6. Boyd, C., Brooks, T.M., Butchart, S.H., Edgar, G.J., Da Fonseca, G.A., Hawkins, F., Hoffmann, M., Sechrest, W., Stuart, S.N., Van Dijk, P.P.: Spatial scale and the conservation of threatened species. Conserv. Lett. **1**(1), 37–43 (2008)
7. Rouget, M., Cowling, R.M., Lombard, A.T., Knight, A.T., Kerley, G.I.: Designing large-scale conservation corridors for pattern and process. Conserv. Biol. **20**(2), 54 (2006)
8. Levin, N., Watson, J.E., Joseph, L.N., Grantham, H.S., Hadar, L., Apel, N., Perevolotsky, A., DeMalach, N., Possingham, H.P., Kark, S.: A framework for systematic conservation planning and management of Mediterranean landscapes. Biol. Cons. **158**, 371–383 (2013)
9. Pressey, R.L., Cabeza, M., Watts, M.E., Cowling, R.M., Wilson, K.A.: Conservation planning in a changing world. Trends Ecol. Evol. **22**(11), 583–592 (2007)

10. Blicharska, M., Orlikowska, E.H., Roberge, J.M., Grodzinska-Jurczak, M.: Contribution of social science to large scale biodiversity conservation: a review of research about the Natura 2000 network. Biol. Cons. **199**, 110–122 (2016)
11. Sen, P.: Report of the Committee on Identifying Parameters for Designating Ecologically Sensitive Areas in India. New Delhi: Ministry of Environment and Forests, Government of India (2000)
12. Gadgil, M., Daniels, R.J.R., Ganeshaiah, K.N., Prasad, S.N., Murthy, M.S.R., Jha, C.S., Ramesh, B.R., Subramaniam, K.A.: Mapping ecologically sensitive, significant and salient areas of Western Ghats: proposed protocol and methodology. Curr. Sci. **100**(2), 175–182 (2011)
13. Margules, C., Pressey, R.: Systematic conservation planning. Nature **405**, 243–253 (2000)
14. Wang, X., Zhong, X., Gao, P.: A GIS-based decision support system for regional eco-security assessment and its application on the Tibetan Plateau. J. Environ. Manage. **91**, 1981–1990 (2010)
15. Knight, A.T., Cowling, R.M., Campbell, B.M.: An operational model for implementing conservation action. Conserv. Biol. **20**, 408–419 (2006)
16. Termorshuizen, J.W., Opdam, P.: Landscape services as a bridge between landscape ecology and sustainable development. Landsc. Ecol. **24**(8), 1037–1052 (2009)
17. Beinat, E.: Value Functions for Environmental Management, p. 241. Kluwer Academic, Boston, MA (1997)
18. Ramachandra, T.V., Setturu, B., Chandran, S.: Geospatial analysis of forest fragmentation in Uttara Kannada District, India. For. Ecosyst. **3**(1), 10 (2016)
19. Riitters, K.H., Wickham, J.D., Coulston, J.W.: A preliminary assessment of Montréal process indicators of forest fragmentation for the United States. Environ. Monit. Assess. **91**, 257–276 (2004)
20. Ramachandra, T.V., Chandran, M.D.S., Rao, G.R., Vishnu, M., Joshi, N.V.: Floristic diversity in Uttara Kannada district, Karnataka. In: Pullaiah, T., Rani, S. (eds.) Biodiversity in India, vol. 8, pp. 1–87. Regency publications, New Delhi (2015)
21. Brose, U., Martinez, N.D., Williams, R.J.: Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. Ecology **84**, 2364–2377 (2003)
22. Lou, J.: Entropy and diversity. Oikos **113**(2), 363–375 (2006)
23. de Lima, R.F., Dallimer, M., Atkinson, P.W., Barlow, J.: Biodiversity and land-use change: understanding the complex responses of an endemic-rich bird assemblage. Divers. Distrib. **19**(4), 411–422 (2013)
24. Peterson, A.T., Egbert, S.L., Cordero, V.S., Price, K.P.: Geographic analysis of conservation priority: endemic birds and mammals in Veracruz, Mexico. Biol. Cons. **93**(1), 85–94 (2000)
25. Brown, S.: Estimating biomass and biomass change of tropical forests: a primer, FAO Forestry Paper, 134 (1997)
26. Ramachandra, T.V., Joshi, N.V., Subramanian, D.K.: Present and prospective role of bio-energy in regional energy system. Renew. Sustain. Energy Rev. **4**, 375–430 (2000)
27. Daniels, R.J.R., Vencatesan, J.: Western Ghats: Biodiversity, People, Conservation. Rupa & Co., New Delhi (2008)
28. Wondie, M., Schneider, W., Melesse, A.M., Teketay, D.: Spatial and temporal land cover changes in the Simen Mountains National Park, a world heritage site in Northwestern Ethiopia. Remote Sens. **3**(4), 752–766 (2011)
29. Wasige, J.E., Thomas, A.G., Smaling, E., Victor, J.: Monitoring basin-scale land cover changes in Kagera Basin of Lake Victoria using ancillary data and remote sensing. Int. J. Appl. Earth Obs. Geoinf. **21**, 32–42 (2013)
30. Nagasaka, A., Futoshi, N.: The influences of land-use changes on hydrology and riparian environment in a northern Japanese landscape. Landsc. Ecol. **14**, 543–556 (1999)
31. Calder, I.R.: Forests and hydrological services: reconciling public and science perceptions. Land Use Wat. Resour. Res. **2**(2), 2.1–2.12 (2012)
32. Vinay, S., Bharath, S., Bharath, H.A., Ramachandra, T.V.: Hydrologic model with landscape dynamics for drought monitoring. In: Proceeding of the Joint International Workshop of ISPRS WG VIII/1 and WG IV/4 on Geospatial Data for Disaster and Risk Reduction, Hyderabad, November 2013, pp. 21–22 (2013)

33. Ramachandra, T.V., Hegde, G., Das, G.K.: Scope of solar energy in Uttara Kannada, Karnataka State, India: roof top PV for domestic electricity and standalone systems for irrigation. Productivity **55**(1), 100 (2014)
34. Ramachandra, T.V., Hegde, G., Krishnadas, G.: Potential assessment and decentralized applications of wind energy in Uttara Kannada, Karnataka. Int. J. Renew. Energy Res. **4**(1), 1 (2014)
35. Ramachandra, T.V., Hegde, G., Setturu, B., Krishnadas, G.: Bioenergy: a sustainable energy option for rural India. Adv. For. Lett. (AFL) **3**(1), 1–5 (2014)
36. Paloniemi, R., Tikka, P.M.: Ecological and social aspects of biodiversity conservation on private lands. Environ. Sci. Policy **11**(4), 336–346 (2008)
37. Zhang, Q., Shuzhen, S.: The mangrove wetland resources and their conservation in China. J. Nat. Res. **16**(1), 28–36 (2001)
38. Mesta, P.N., Setturu, B., Chandran, M.D.S., Rajan, K.S., Ramachandra, T.V.: Inventorying, mapping and monitoring of mangroves towards sustainable management of west coast, India. J. Geophys. Remote Sens. **3**, 130 (2014). https://doi.org/10.4172/2169-0049.1000130
39. Ramachandra, T.V., Bharath, S.: Geoinformatics based valuation of forest landscape dynamics in central western ghats, India. J. Remote Sens. GIS **7**, 1 (2018). https://doi.org/10.4172/2469-4134.1000227
40. Ramachandra, T.V., Bharath, S., Subash Chandran, M.D., Joshi, N.V.: Salient ecological sensitive regions of central western ghats, India. Earth Syst. Environ. (2018). https://doi.org/10.1007/s41748-018-0040-3
41. Ramachandra, T.V., Tara, N.M., Setturu, B.: Web based spatial decision support system for sustenance of western ghats biodiversity, ecology and hydrology. In: Sharma, A., Rajeswaran, J. (eds.) Creativity and Cognition in Arts and Design, pp. 58–70. Bloomsbury Publishing India Pvt. Ltd., New Delhi (2017)

Assessment of Potential Marine Fishing Zone Using Geospatial Technologies at the Coastal Stretch of West Bengal, India

B. C. Patra, M. Bhattacharya, A. Kar, B. K. Das and P. K. Shit

Glossary of Abbreviation

PFZ	Potential Fishing Zone
SST	Sea Surface Temperature
Chl-a	Chlorophyll-a
GIS	Geographic Information System
GPS	Global Positioning System
CZCS	Coastal Zone Colour Scanner
NOAA	National Oceanographic and Aerospace Administration
AVHRR	Advanced Very High-Resolution Radiometer
OCM	Ocean Colour Monitor
IRS	Indian Remote Sensing Satellite
INCOIS	Indian National Centre for Ocean Information Services
MODIS	Moderate Resolution Imaging Spectroradiometer
IUCN	International Union for Conservation of Nature
WORMS	World Register of Marine Species
OLI	Operational Land Imager
TIRS	Thermal Infrared Sensor
FLAASH	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube
ERDAS	Earth Resource Development Assessment System
L8SR	Landsat 8 Surface Reflectance

B. C. Patra (✉) · A. Kar · P. K. Shit
Centre for Aquaculture Research, Extension & Livelihood, Department of Aquaculture Management & Technology, Vidyasagar University, Midnapore 721 102, West Bengal, India
e-mail: patrabidhan1962@gmail.com

M. Bhattacharya · B. K. Das (✉)
ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata 700 120, West Bengal, India
e-mail: basantakumard@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_28

UTM	Universal Transverse Mercator
WGS	World Geodetic System
AOI	Automated Optical Inspection
CPUE	Catch Per Unit Effort

1 Introduction

Marine fish is a renewable resource in India, and it has a potential contribution in agro-economic development, employment generation, supply of animal protein and foreign exchange earnings for both economic values and human sustenance and livelihood generation [1]. Fishery resources in India are one of the largest and diversified natural resources of the world with respect to the abundance of fish species. Potential fishing resource mapping and forecasting is an important task to minimize the fishing cost, efforts and to maximize the profit of fisherman [2–4]. The fluctuation in seasonal abundance of stocks in different areas assumes great significance in the expansion and management of fisheries. Maximum fish catch can be achieved by directing fishermen to most profitable fishing grounds through fishery forecast [3]. Remote Sensing and Geographic Information System (GIS) technique are useful as modern tools for identifying potential marine fishing zone and forecasting [3–5]. In view of biological knowledge, phytoplankton is at the bottom of the marine food chain and it plays a vital role as primary producer in marine environment. Rich sources of chlorophyll concentration (CC) in the marine water indicate high primary productivity that in turn controls the growth and nourishment of the successive higher trophic level living organisms as zooplankton and small marine fishes. Researcher also reported that Sea Surface Temperature (SST) information is essential for detecting thermal fronts, upwelling zones, currents and large-scale oceanic eddies to aid the fishing industry on an operational basis [6]. It was observed in previous studies that by and large the pelagic species were found to concentrate at sharp horizontal thermal boundaries. Many investigators reported the mapping of Sea Surface Temperature (SST) and chlorophyll concentration using satellite remote sensing imagery as a tool to study the distribution of potential fishing zones [3, 7, 8]. Potential Fishing Zone (PFZ) advisory was started with an aim to minimize the fishing cost and efforts and to maximize the profit [2, 3]. Remote sensing as a tool was utilized by researchers for detecting potential fishing zones and for wise utilization of the marine fishery resources [5]. Historically, satellite remote sensing for predicting fishing trends was applied for the first time in USA in the year 1971 [9]. Initially, Sea Surface Temperature (SST) was used as the only oceanic feature for generating fisheries forecast [10]. Soon after [11] used ocean colour from Coastal Zone Colour Scanner (CZCS) along with SST from National Oceanographic and Aerospace Administration (NOAA) and Advanced Very High Resolution Radiometer (AVHRR) to enhance the quality of forecast. Presently, the research becomes radically changed and used various instruments like radar, sonar, several satellites to find out and predict the sites where fish schools aggregate in ocean boundary [10].

In India, PFZ forecasting started in the year 1989–90 which were developed using NOAA-AVHRR-derived SST data [12]. Subsequently, chlorophyll (Chl) concentration data derived from Ocean Colour Monitor (OCM) was also used along with the SST data after successful launching of Indian Remote Sensing Satellite (IRS-P4) on 26th May 1999 through which several geobiological, biochemical and physical processes find out the ocean colour [7, 13, 14]. Indian National Centre for Ocean Information Services (INCOIS) generates and disseminates PFZs advisories for the benefit of those involved in fishing. INCOIS develops several PFZs through satellite-derived chlorophyll and SST images. However, it is found that the local fish catch data in the recent past represents that there are many other high fishing zones that were observed, and these were not within the INCOIS predicted PFZs [15]. In the present study some additional comparatively high fishing zones, other than the INCOIS predicted PFZs have been demarcated by developing a spatial model derived PFZ lying adjacent to the West Bengal coast. In order to develop this model, Aqua MODIS (Moderate Resolution Imaging Spectroradiometer) plotted chlorophyll and SST images along with their in situ data were used. PFZ advisories of INCOIS were also used as an auxiliary data. The objectives of the present study are to (i) prepare a local spatial model derived PFZ in this study area, (ii) validate the INCOIS predicted PFZs with respect to the groundfish catch data. The coastline of India extends up to 7517 km and employs about 14.5 million people [16]. Fishing is a regular practice not only by fisherman but also by coastal people irrespective of cast and creed, throughout the year [17]. Out of all, around 314 marine fish species diversity found in the 158 km coastline at north eastern part of Bay of Bengal in West Bengal region [18]. It is also found that, in West Bengal, every year more than thousand fisherman have maintained their livelihood generation through directly or indirectly marine fish catch and fish by-products. Besides this, they have faced many problems like financial deficiency, lack of technical skill, fishing zones identification, bathymetric properties of PFZs and forecasting. Therefore, the present study deals with the superior estimation and assessment of the potential fishing zones using remote sensing and GIS techniques at entire West Bengal coast.

2 Description of the Study Area

Bay of Bengal is the largest bay in the world, which comprises waters flowing straight out of the Himalayas region, and it occupies a total area of 2,172,000 Km^2 (en.wikipedia.org). The Ganga River of India flows from west to east before flowing down into the Bay of Bengal. The continental shelf lying adjacent to the Ganga delta extends to almost 100 km off the coastline and has a very shallow bathymetry [19]. About 70–80% of annual rainfall occurs during the summer monsoon (south-west monsoon) resulting in high river discharge (ranging between 2,952 and 11,897 m^3 s^{-1}) which is found to decline steadily during non-monsoonal months (varying from 900 to 1,500 m^3 s^{-1}) [20]. Our study was conducted in the northernmost part of the Bay of Bengal having geographic extension between 20° 40′ N to 21° 30′ N and

87° 20′ E to 89° E comprising a large marine ecosystem with dynamic estuarine network base coastal area of almost 158.5 km. However, we selected nine landing stations which are namely (1) New Digha, (2) Digha Mohana, (3) Shankarpur, (4) Petuaghat, (5) Sagar, (6) Frezergunge, (7) Namkhana, (8) Kakdwip and (9) Diamond Harbour for marine fish resources collection, identification, in situ data compilation and monitoring in regular mode (Fig. 1).

3 Materials and Methods

3.1 In Situ *Sampling of Specimen from Different Potential Fishing Zone*

The specified fish landing centres were visited and surveyed at least once in each month during the survey period of 2 years (January2014 to December2016), and fish specimens were scientifically assessed. In the present study, two regular mechanized fishing boats were hired for in situ data collection. These cylinder mechanized boats were used which carried out fishing by operating the gill nets of mesh size 90–95 mm as per the standard protocol laid by Department of Fisheries, Government of West Bengal. Approximately, more than 10,000 hectors regular marine fish catch zone was considered for the research study and scientific observation area covered by trawlers, fishing boats, country boats, etc (Fig. 1). In several cases, some specimens were preserved in 4–6% formaldehyde and brought to the laboratory for proper identification up to species level. The specimens were normally varied 2 cm to 1 m (or more) in length as these were captured using gill net, cast net, drag net, scoop net, etc. Due attention was paid to record the local name of the fish species, their IUCN status, threat to humans and nature of utilization by humans in course of preparation of an organic checklist of fish fauna. The IUCN status and relevant information provided by WORMS (2017), and IUCN Red List of Threatened species version (2016-1) were taken into consideration [21].

3.2 In Situ *Sampling of Chlorophyll, SST and Other Different Water Parameter*

SST was measured using a Multikit (WTW Multi 340i Set; Merck, Germany) fitted with the probe WTW Tetracon 325. Chlorophyll-a was measured by standard spectrophotometric methods. For chlorophyll-a analysis, samples were collected and preserved at 4 °C in the ice box in dark and sent to the laboratory within 24 h of sampling. Two litres of each sample was passed through GF/F filter paper, extracted in 90% acetone and measured with a spectrophotometer [22]. All the in situ data was

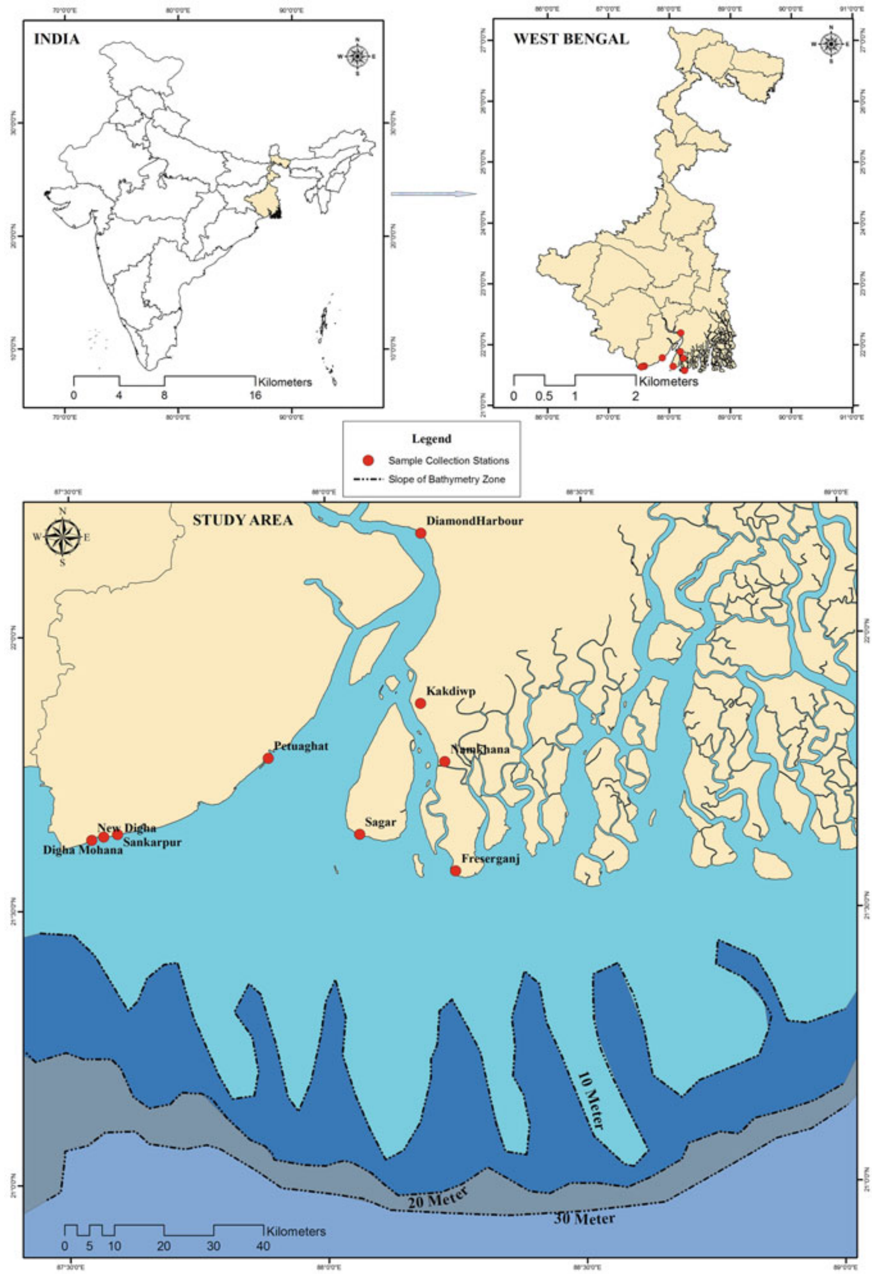


Fig. 1 Ground landing stations, horizontal movement of trawlers, fishing boats at PFZs of West Bengal coastal stretch

collected during the time of satellite pass. Dissolved oxygen (DO) was estimated by modified Winkler's method [23].

3.3 Image Analysis for Chlorophyll and SST

The study area was located in the OLI (Operational Land Imager)/Landsat-8 images, path/row 138/46 and 139/45 dated 14th April 2017. After a preliminary analysis of the images, only those without cloud cover were selected. In this research Landsat-8 OLI /TIRS (Thermal Infrared Sensor), data was collected from United States Geological Survey global website to understand the relation between Sea Surface Temperature (SST) and Chlorophyll concentration associated with different biological and physical processes for identification of Potential Fishing Zone (PFZ). The empirical models for estimating Chl-a concentration were fitted to be applied on multispectral images from OLI/Landsat-8, freely available from the United States Geological Survey (Watanabe, 2015). Most of the OLI bands have a terrestrial application. However, the bands with applications on vegetation and photosynthetic pigments can be suitable for estimating phytoplankton pigment concentrations in highly productive water bodies. The advantage of using OLI/Landsat-8 images is the acquisition of data every 16 days and their free availability.

The radiometric calibration was conducted to convert digital numbers into top-of-atmosphere radiance, using the metadata released with the images. The retrieval of the at-surface reflectance was accomplished using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH), an atmospheric correction module, implemented in the ERDAS software. SST was calculated from Landsat-8 OLI Image in the current year (2017) following Eq. 1.

Calculation of Sea Surface Temperature from Landsat imagery

$$Bt/1 + w \times (BT/p) \times ln\,(e) \tag{1}$$

Where

BT = satellite temperature; w = wavelength of emitted radiance (11.5 μm); p = h × c/s (1.438×10^{-2} m K); h = Planck's constant (60626×10^{-34} Js); s = Boltzmann constant (1.38×10^{-23} J/K); c = velocity of light (2.998×10^{8} m/s); e = 0.004Pv + 0.986; Pv = proportion of vegetation; p = 14380

Calculation of DN to Radiance

$$\mathbf{L_{\lambda}} = \frac{\mathbf{(LMAX_{\lambda} - LMIN_{\lambda})}}{\mathbf{QCALMAX - QCALMIN}} \times \mathbf{(DN - QCALMIN) + LMIN_{\lambda}} \tag{2}$$

where

LMIN and LMAX are the spectral radiances for each band at digital numbers 1 and 255.

DN is digital number value of pixel.
QCAL is the quantized calibrated pixel value in DN.
QCALMIN = the minimum quantized calibrated pixel value (corresponding to LMIN $_{\lambda}$) in DN = 1 (LPGS products) = 0 (NLAPS products)
QCALMAX = the maximum quantized calibrated pixel value (corresponding to LMNX $_{\lambda}$) in DN = 255
λ is wavelength.

Calculation of Radiance to Brightness Temperature

$$T = \frac{\mathbf{K_2}}{\ln\left(\frac{\mathbf{K_1}}{L_\lambda} + 1\right)} \tag{3}$$

where K_1 and K_2 become a coefficient determined by effective wavelength of a satellite sensor.

Chlorophyll Mapping The actual fieldworks were carried out to concur with OLI/Landsat-8 images in order to map the Chl-a using a model calibrated for the bands of the OLI sensor. However, a unique OLI/Landsat-8 image was acquired on 14 April 2017 in suitable conditions (without clouds). Before algorithm application, two products were tested: Landsat-8 surface reflectance (L8SR) (USGS, http://landsatLandsat.usgs.gov/CDR_LSR.phphttp://landsatLandsat.usgs.gov/CDR_LSR.php) and at-surface reflectance (Rsup) computed by Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) [24]. The comparison was carried out from five measurements collected precisely on the day of satellite overpass. Finally, the best models calibrated for OLI bands were applied to the corrected image.

3.4 *Potential Fishing Zone Mapping*

In this research Landsat-8 OLI (Operational Land Imager)/TIRS (Thermal Infrared Sensor), data was collected from United States Geological Survey global website to understand relation between Sea Surface Temperature (SST) and Chlorophyll concentration associated with different biological and physical processes for identification of Potential Fishing Zone (PFZ). The satellite images of the respective dates were spectrally subset to select the band of interest in ArcGIS 10.2 followed by the subset with an AOI of West Bengal coastline in ERDAS imagine. Subset images for both the chlorophyll and SST were re-projected to WGS 84 UTM zone 45. The upper and lower limits of chlorophyll and SST images were set, based on the observed maximum and minimum in situ data. SST lying between 19 and 32 °C (maximum) was used for the present modelling, while for chlorophyll it was 2.0 to 4.24 mg/m^3 (maximum). After extraction of the SST and chlorophyll data within the determined range, spatial variation maps were generated by density raster generation. Predicted PFZs were overlaid on the raster images of SST and chlorophyll to

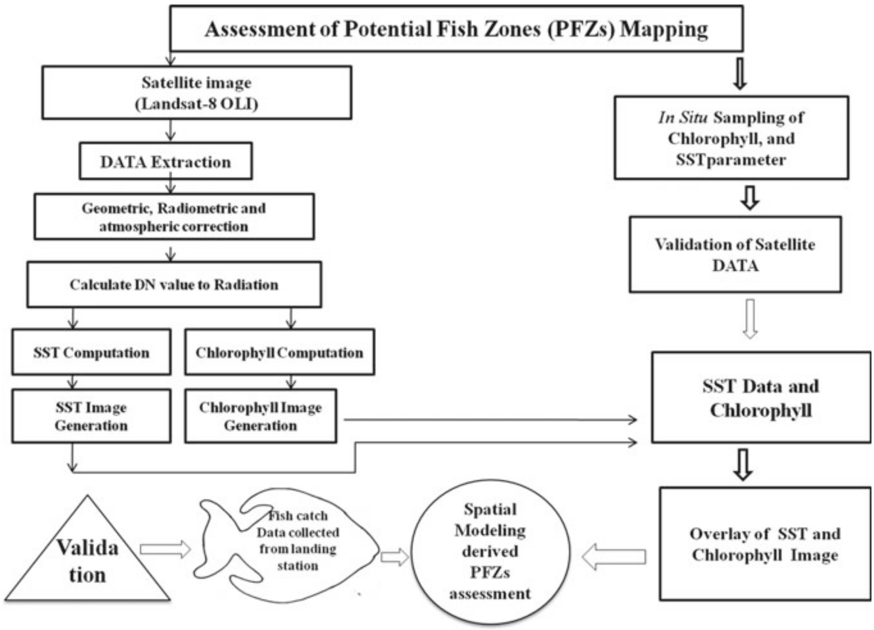


Fig. 2 Flowchart of the adopted methodologies to predict and identify the PFZs

extract the corresponding pixel values. Detailed methodology adopted for superior prediction and identification of PFZs is represented in Fig. 2.

4 Results and Discussion

Monthly Variability of In Situ SST and Chlorophyll In situ data of SST and chlorophyll were monitored from October to February for 2 consecutive years (20 October 2014 to February 2015 and October 2015 to February 2016). This phase of the year was deliberately chosen for the study because of favourable atmospheric conditions allowing for acquiring of ground data in substantial quantity. Moreover, the sky during this time of the year remains cloud-free and that aids in the processing of MODIS data. The mean of the respective months (of the 2 years) was considered to illustrate the results. As the LANDSAT satellite passed through the present study area during 1400 h, in situ data was taken close to this time in order to increase the accuracy of the model derived. Mean SST was found to be maximum in the summer season (29.05 ± 1.23 °C). In the winter months that followed the SST declined by 5 °C till January (24.51 ± 1.44 °C). The range of SST varied from 23.18 to 30.45 °C during the quarterly sampling in the consecutive 2 years. A maximum chlorophyll

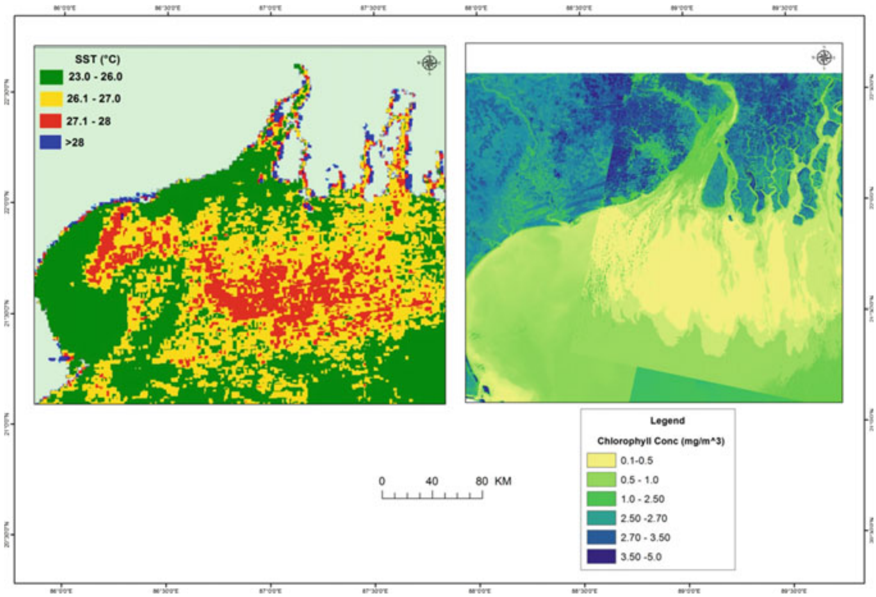


Fig. 3 Water mass of northern Bay of Bengal showing SST (°C) and chlorophyll concentration (mg/m^{-3}) in a particular day time in the coastal region of West Bengal, India

content of 4.245 mg m^{-3} was observed in the time of post-monsoon, and the least chlorophyll concentration of 2.012 mg m^{-3} was observed in the season of winter.

LANDSAT-Derived SST and Chlorophyll data According to LANDSAT-derived data, SST showed a similar trend as the in situ data in the present study area. A steady decreasing trend in SST was observed from winter to summer season. The magnitude of the LANDSAT-derived SST was in fair parity with our in situ data obtained onboard. Spatially, SST was more or less uniform throughout the study area. A negligible decrease in the SST by a margin of 0.3–0.5 °C was observed from the coastline towards offshore, and SST scores are interpreted (Fig. 4). This could be due to the cut-off land effect on the adjoining coastal waters which slowly dissipates towards offshore [25]. Similarly, LANDSAT-derived chlorophyll values also showed a similar trend to the in situ observed chlorophyll concentration. Unlike SST, chlorophyll showed a significant trend in spatial variability. On moving from the coastline towards the offshore a substantial decrease in chlorophyll concentration values by a margin of 1.2 to 1.8 mg m^{-3} was observed and noted to digitized images (Fig. 3).

Analysis of fishing operations Remotely sensed PFZs satellite-derived chlorophyll concentration provides a measure of the areas of enhanced biological production. Remotely sensed SST characterizes the oceanic environment suitable for enhanced biological production [2]. The zooplankton and fish population are known to accumulate for feeding and spawning at oceanic features like fronts, eddies, rings, meanders and upwelling. The use of both variables would explain the oceanic

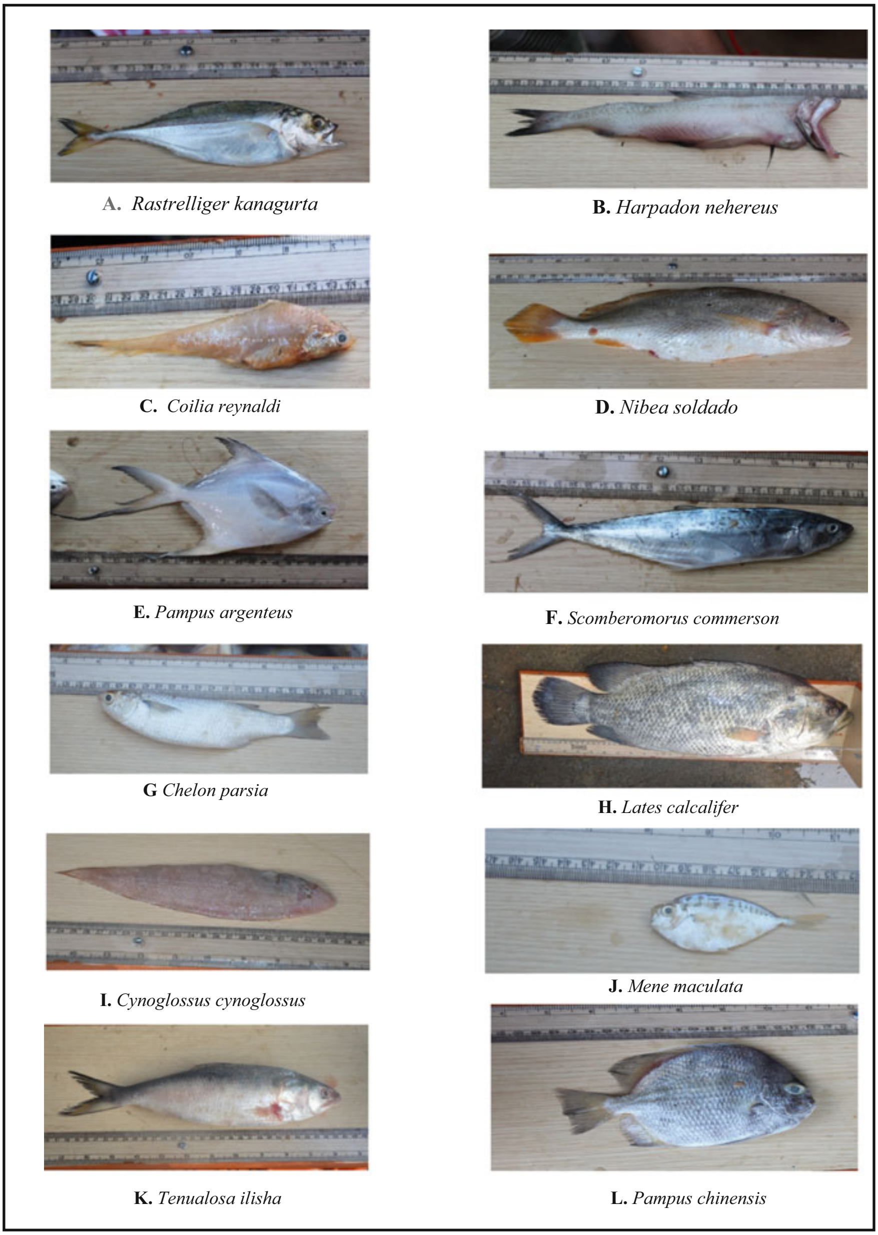


Fig. 4 Dominant marine fishes (localized) of entire coastal stretch of West Bengal state

Table 1 Sampling sites, GPS location and dominant marine fishes (localized) of entire coastal stretch of West Bengal state

Sl. no.	Name of the location for in situ sampling	GPS location		Dominant marine fishes
		Latitude data (N)	Longitude data (E)	
1	Digha site I	21°30′56.78″	87°31′42.38″	A, B, D, H, K
2	Digha site II	21°26′15.00″	87°33′38.93″	C, D, H, K, L.
3	Digha Mohana site I	21°35′57.52″	87°33′32.25″	A, B, D, E, F, G, I, K, L
4	Digha Mohana site II	21°31′1.57″	87°38′39.97″	A, B, C, D, E, F, G, I, J, K, L
5	Digha Mohana site III	21°22′26.88″	87°39′8.33″	A, B, C, D, E, F, H, I, J, K, L
6	Digha Mohana site IV	21°24′41.83″	87°38′35.88″	A, B, C, D, E, F, G, I, J, K, L
7	Mandarmoni site I	21°29′0.56″	87°43′56.62″	A, B, J, K
8	Mandarmoni site II	21°22′9.64″	87°45′20.36″	A, B, C, D, H
9	Sankarpur I	21°35′43.38″	87°47′4.11″	A, B, C, D, E, F, G, H, I, J, K, L
10	Sankarpur II	21°31′52.44″	87°52′4.18″	A, B, C, D, E, F, G, H, I, J, K, L
11	Sankarpur III	21°24′29.89″	87°48′39.95″	B, C, D, E, F, G, H, K, L
12	Petuaghat site I	21°43′25.21″	87°57′25.51″	A, B, C, D, G, H, I, J, K, L
13	Sagar site I	21°30′21.37″	88°3′59.20″	E, C, K, H
14	Sagar site II	21°36′16.38″	88°0′18.10″	B, E, G, H
15	Namkhnana site I	21°41′51.37″	8°10′27.49″	G, H, K
16	Bakkhali site I	21°24′56.37″	88°18′5.25″	C, E, G, H, K, L
17	Bakkhali site II	21°16′4.07″	88°13′8.57″	A, C, E, G, H, K, L
18	Bakkhali site III	21°26′11.13″	88°19′5.78″	B, C, E, G, H, K, L
19	Frsergunj site I	21°34′12.27″	88°12′54.05″	A, B, C, D, G, H, K, L
20	Kakdwip site I	21°53′27.02″	88°5′10.18″	A, B, C, D, G, H, K, L

environment and food resource availability in an ecosystem for exploring fishery resources [26]. A typical output generated using integration of the chlorophyll concentration and SST image is shown. SST contours overlaid on the chlorophyll concentration image indicate some matching features in images derived from both optical and thermal infrared sensors. Cool water is an indicator of nutrient-rich waters. In such areas, the probability of enhanced production will be greater than stratified warm waters. This indicates that there is an inverse relationship between chlorophyll concentration and SST [27]. Areas of matching features were selected as PFZs and suggested for experimental fishing. The coincidence of chlorophyll and SST features at some locations indicates that physical and biochemical processes are closely coupled at these locations. High catch points were observed in the vicinity of thermal as well as colour persistent features [9, 27]. Thus, coincidence of ocean colour and thermal features can be utilized for exploring fishery resources and PFZs identification [28]. The selection of features and their relevance to fishery resource distributions have been discussed by [7].

In order to authenticate that the places demarcated by the spatial model were truly PFZs, ground catch data from these points were obtained from the local fisherman [29]. A good way to assess the quality of a predicted PFZ is to analyse the data of catch per unit effort (CPUE) from the respective areas on the days for which the map is prepared. Catch per unit effort (CPUE) is the relative measurement of the abundance of a target species; i.e. the amount of a target species caught with unit effort. Change in the CPUE indicates the change in the magnitude of the target species available. In the perspective of the present study, higher CPUE indicates that more the aggregation of fish in a particular area, the more fish catch in unit time. In the West Bengal coast, some pelagic fishes like Sardine (*Sardinella fimbriata, Sardinella gibbosa*), Pomfret (*Pampus argenteus, P. chinensis*), Indian mackerel (*Rastrelliger kanagurta*), Hilsa (*Tenualosa ilisha*), Bombay duck (*Harpadon nehereus*), Bhola-bhetki (*Lates calcarifer*), Parse (*Chelon parsia*), Topse (*Polynemus paradiseus*), Amadi (*Coilia reynaldi*), Moon fish (*Mene maculate*) and Indo-Pacific King mackerel (*Scomberomorus guttatus*) contribute to dominant position than other marine fishes in the post-monsoon or winter catch (Fig. 4). Numbers of dominant group of marine fishes are varied according to the preferences of suitable habitat which inferred by in situ sampling survey (Table 1).

5 Conclusion

Analysing the consequences from the present research study, it can be concluded and summarized that both SST and chlorophyll concentration act as key factors in determining the spatial distribution of fishes. It is evident that fishes of the West Bengal coastal belt in substantial quantity are caught from regions with optimum SST and chlorophyll, especially in areas where a sharp gradient in these parameters are observed. In this study, spatial model was used to observe whether there are any other areas where substantial fish catches are found apart from the areas where INCOIS provided their PFZ predictions. Apart from, as survey out results also confirms the undisturbed, natural breeding zones of marine fish species of the specified part of the West Bengal coastal stretch showing higher numbers of species. In this regard, some areas were detected where CPUE was quite higher than the absolutely non-PFZ areas. This observation not only supports the precision of the model derived in this study but also validates the PFZ advisory of INCOIS. In showing the less availability of fishes even the supportive SST and chlorophyll concentration operates some location, due to the marginal confluence of undesirable pollutant components, soil erosion and other man-made factors. Thus, assessment of PFZs is beneficial to artisanal, motorized and small mechanized sector fishermen engaged in pelagic fishing activities such as ring seining, gill netting. However, PFZ advisories assist the local fisherman and concerned bodies in the significant reduction of search time and fuel cost, yielding the better catch with less human effort.

6 Acknowledgements

This research work was supported by the research grant from the Natural Resources Data Management System (NRDMS), Department of Science and Technology (DST), Government of India (Project Ref. No. NRDMS/11/1966/012), and Science and Engineering Research Board, Department of Science and Technology, Government of India, for financial assistance to carry out the research work (Project file No Ref. No.PDF/2016/001776).

References

1. Saitoh, S.-I., Chassot, E., Dwivedi, R., Fonteneau, A., Kiyofuji, H., Kumari, B., et al.: Remote sensing applications to fish harvesting. Remote Sens. Fish. Aquac., 57–76 (2009)
2. Nayak, S., Solanki, H., Dwivedi, R.: Utilization of IRS P4 ocean colour data for potential fishing zone—a cost benefit analysis (2003)
3. Choudhury, S., Jena, B., Rao, M., Rao, K., Somvanshi, V., Gulati, D., et al.: Validation of integrated potential fishing zone (IPFZ) forecast using satellite based chlorophyll and sea surface temperature along the east coast of India. Int. J. Remote Sens. **28**(12), 2683–2693 (2007)
4. Giri, S., Manna, S., Chanda, A., Chowdhury, A., Mukhopadhyay, A., Chakraborty, S., et al.: Implementing a Spatial model to derive Potential Fishing Zones in the Northern Bay of Bengal lying adjacent to West Bengal Coast, India. J. Indian Soc. Remote Sens. **44**(1), 59–66 (2016)
5. Mansor, S., Tan, C., Ibrahim, H., Shariff, A.R.M.: Satellite fish forecasting in south China sea. In: Paper presented at the 22nd Asian Conference on Remote Sensing, vol. 5, p. 9 (2001)
6. Deekshatulu, B., Nath, A., Rao, M.: Satellite derived sea surface temperature applications for Indian Ocean resources development. In: Scientific Organizing Committee of PORSEC, Proceedings of the PORSEC 1992 in Okinawa: Conference for Pacific Ocean Environments and Probing, vol. 1 (1992)

7. Solanki, H., Dwivedi, R., Nayak, S., Naik, S., John, M., Somvanshi, V.: Cover: application of remotely sensed closely coupled biological and physical processes for marine fishery resources exploration. Int. J. Remote Sens. **26**(10), 2029–2034 (2005)
8. Nammalwar, P., Satheesh, S., Ramesh, R.: Applications of remote sensing in the validations of potential fishing zones (PFZ) along the coast of North Tamil Nadu. India Indian J. Geo-Mar. Sci. **42**(3), 283–292 (2013)
9. Laurs, R.M.: Fishery-advisory information available to tropical Pacific tuna fleet via radio facsimile broadcast. Comm. Fish. Rev. **33**, 40–42 (1971)
10. Chandran, R.V., Solanki, H., Dwivedi, R., Nayak, S., Jeyaram, A., Adiga, S.: Studies on the drift of ocean colour features using satellite-derived sea surface wind for updating potential fishing zone. Indian J. Mar. Sci. **33**(2), 122–128 (2004)
11. Mueller, J.L., LaViolette, P.E.: Color and temperature signatures of ocean fronts observed with the Nimbus-7 CZCS. In: Oceanography from Space, pp. 295–302. Springer (1981)
12. Solanki, H., Raman, M., Kumari, B., Dwivedi, R., Narain, A.: Seasonal trends in the fishery resources off Gujarat: salient observation using NOAA AVHRR (2016)
13. Solanki, H., Dwivedi, R., Nayak, S., Somvanshi, V., Gulati, D., Pattnayak, S.: Fishery forecast using OCM chlorophyll concentration and AVHRR SST: validation results off Gujarat coast. India Int. J. Remote Sens. **24**(18), 3691–3699 (2003)
14. Wilson, C., Morales, J., Nayak, S., Asanuma, I., Feldman, G.: Ocean colour radiometry and fisheries. Why Ocean Colour, 47–58 (2008)
15. Butler, M.: The application of remote sensing technology to marine fisheries: an introductory manual (Vol. 295): Food & Agriculture Org (1988)
16. Vinu Chandran, R., Jeyaram, A., Jayaraman, V., Manoj, S., Rajitha, K., Mukherjee, C.: Prioritization of satellite-derived potential fishery grounds: an analytical hierarchical approach-based model using spatial and non-spatial data. Int. J. Remote Sens. **30**(17), 4479–4491 (2009)
17. Khare, V., Nema, S., Baredar, P.: Status of solar wind renewable energy in India. Renew. Sustain. Energy Rev. **27**, 1–10 (2013)
18. Kar, A., Raut, S., Bhattacharya, M., Patra, S., Das, B.K., Patra, B.C.: Marine fishes of West Bengal coast, India: diversity and conservation preclusion. Reg. Stud. Mar. Sci. **16**, 56–66 (2017)
19. Saha, P., Saha, B., Hazra, S., Sinha Roy, S.: Geomorphology and sediment character of Bakkhali-Frasergunj coastal belt and adjoining inner-continental shelf, Bay of Bengal. Indian J. Geosci. **65**(3), 195–210 (2011)
20. Mukhopadhyay, S., Biswas, H., De, T., Jana, T.: Fluxes of nutrients from the tropical River Hooghly at the land–ocean boundary of Sundarbans, NE Coast of Bay of Bengal. India J. Mar. Syst. **62**(1), 9–21 (2006)
21. World Register of Marine Species (WoRMS) (2017). WoRMS Editorial Board. http://www.marinespecies.org Accessed 18 December 2017
22. Parsons, T.R.: A manual of Chemical & Biological Methods for Seawater Analysis. Elsevier (2013)
23. Strickland, J.D., Parsons, T.R.: A practical Handbook of Seawater Analysis (1972)
24. Adler-Golden, S., Berk, A., Bernstein, L., Richtsmeier, S., Acharya, P., Matthew, M., et al.: FLAASH, A MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In: Proceedings 7th Ann. JPL Airborne Earth Science Workshop, vol. 97, pp. 9–14. JPL Publication Pasadena, CA (1998)
25. Robinson, I.S.: Shelf seas, estuaries, and coasts. Discov. Ocean from Space, 485–538 (2010)
26. Horning, N.: Remote Sensing for Ecology and Conservation: A Handbook of Techniques. Oxford University Press (2010)
27. Solanlki, H., Dwivedi, R., Nayak, S., Jadeja, J., Thakar, D., Dave, H., et al.: Application of Ocean Colour Monitor chlorophyll and AVHRR SST for fishery forecast: preliminary validation results off Gujarat coast, northwest coast of India (2001)
28. Clerici, M., Hoepffner, N., Diop, M., Ka, A., Kirugara, D., Ndungu, J.: SST derivation from MSG for PUMA pilot projects in fisheries. Int. J. Remote Sens. **30**(8), 1941–1959 (2009)

29. Claasen, V.R.: The application of digital remote sensing techniques in coral reef, oceanographic and estuarine studies, report on a regional Unesco/COMAR/GBRMPA Workshop; Townville, Australia, August 1985. UNESCO Reports in Marine Science = Rapports de l'Unesco sur les sciences de la mer (1986)

Water, Carbon and Nitrogen Footprints of Major Crops in Indo-Gangetic Plains

B. Chakrabarti, U. Mina, D. Chakraborty, H. Pathak, D. K. Sharma, N. Jain, R. S. Jatav, P. Dixit, R. Katiyar and R. C. Harit

1 Introduction

Water, carbon and nitrogen cycles hold centre stage in global change studies due to its role in influencing weather and determining food, fibre and wood supply for human use through plant productivity. In recent times, overuse of inputs in agricultural sector is leading to water scarcity, groundwater and soil pollution and enhanced emission of greenhouse gases (GHG). Agricultural fields emit GHGs like methane (CH_4) and nitrous oxide (N_2O), and leaching of nitrate (NO_3^-) from soil causes groundwater pollution. Overuse of groundwater for irrigation purpose is causing depletion of groundwater table. Emission of GHGs has increased by 70% from 1970 to 2004 and is projected to further increase by 25 to 95% by 2030 [21]. The Intergovernmental Panel on Climate Change (IPCC) warned of dire consequences of changing climate on agriculture, and natural resources, if measures are not taken to curb the ill-effects of global warming. In India, agriculture sector contributes to 18% of the total GHG emission. Enteric fermentation contributes to 63.4% of GHG emission from agriculture, while rice cultivation accounts for 20.9%, agricultural soils 13%, manure management 2.4% and crop residue burning 2% (INCCA [13]. The global food system is using almost half the planet's ecological capacity and is degrading the natural resources [6]. Intensive cultivation practices, monoculture, excessive use of agrochemicals as well as water for irrigation purposes are leading to nutrient mining, water scarcity and pollution. Hence, new technologies need to be developed to reduce GHG emission as well as for better utilization of natural resources in the agricultural sector.

B. Chakrabarti (✉) · U. Mina · D. Chakraborty · H. Pathak · D. K. Sharma · N. Jain · R. S. Jatav
P. Dixit · R. Katiyar · R. C. Harit
Centre for Environment Science and Climate Resilient Agriculture, Indian Agricultural Research Institute, New Delhi 110012, India
e-mail: bidisha2@yahoo.com; bidisha_env@iari.res.in

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_29

1.1 Water Footprint

The concept of water footprint (WFP) was first introduced by [9] and subsequently elaborated by [10]. Water footprint refers to the total volume of direct and indirect freshwater used, consumed and polluted during all processes involved in formation of a product. The water footprint of a product is expressed in volume of water per unit of product (m^3 ton^{-1}). The equation to calculate WFP of crops is:

$$\mathrm{WFP} = \mathrm{CWU}/\mathrm{Y}$$

where CWU is crop water use and Y is yield of crop.

Water footprint is a useful indicator of consumption and pollution of water [2]. Water footprint has three components; i.e. blue water footprint, green water footprint and grey water footprint. The blue water footprint refer to consumption of blue water resources, i.e. surface and groundwater, while the green water footprint refers to consumption of green water resources, i.e. rainwater stored in the soil as soil moisture. Greywater footprint is the indicator of water polluted due to a process and is defined as the volume of freshwater that is required to assimilate the pollutant load based on existing ambient water quality standards. The water footprint for India has been estimated by [4] as 980 m^3 $capita^{-1}$ yr^{-1}. Consumption of agriculture goods contributes 94%, domestic water contributes 3.9%, and industrial goods contribute the remaining 2.1%. The largest portions of these volumes are attributable to growing the plants. The global water footprint of rice production is estimated to be 784 billion m^3 yr^{-1} [3]. In countries like India, Indonesia, Vietnam, Thailand, Myanmar and the Philippines, the green water fraction is substantially larger than the blue water fraction. In the USA, however, the blue water fraction is 3.7 times the green water fraction.

1.2 Carbon Footprint

Carbon footprint (CFP) is originated as a subset of "ecological footprint" given by [22]. Wiedmann and Minx [23] introduced carbon footprint for quantifying the impact of GHG emission on environmental sustainability. At present time, carbon footprint has become a widely used term dealing with the responsibility and abatement actions against the global climate change. According to Pathak et al. [18], CFP is the total set of greenhouse gas (GHG) emission caused by a product and is expressed in terms of carbon dioxide (CO_2) equivalent. Assessment of carbon footprint has been widely accepted for identifying and developing low carbon options in different sectors. In recent times, there is a growing interest in reducing the CFP of agricultural products (Williams and Wikstrom [24]. CFP of a product can be quantified by estimating GHG emissions at all stages in its life cycle like field preparation, fertilizer and pesticide application, harvesting of crop, storage, processing, packaging, transport and finally consumption (Chakrabarti et al. [1]). Carbon footprinting helps in quantifying and comparing the GHG emission of different crops and identification of strategies to reduce the footprint. Carbon footprint of China's crop production was assessed by [5] using national statistical data available for the period of 1993–2007. Fertilizer use, electricity for irrigation and mechanical operations accounted for 89% to the total CFP. The Intergovernmental Panel on Climate Change (IPCC) indicated that lifestyle changes like: reducing animal protein consumption can help in lowering GHG emission. Quantification of carbon footprints of Indian food items revealed that GHG emission from mutton is 11.9 times more than that of milk, 12.1 times fish, 12.9 times rice and 36.5 times chapatti [18].

1.3 Nitrogen Footprint

Nitrogen is the most limiting nutrient controlling the primary production of all agricultural systems. Use efficiency of external N supply is as low as 33% [14]. Unused nitrogen can be immobilized, denitrified, washed into surface water or leached into groundwater (Huang et al. 1992). Agricultural activities have greatly altered the global nitrogen cycle and produced nitrogenous gases of environmental significance. A major consequence of this human-driven change in global nitrogen cycle is the increased emission of N-based trace gases, such as nitrous oxide (N_2O) and NO_x ($NO + NO_2$) and ammonia that impacted regional and global atmospheric chemistry [16]. Nitrogen footprint is defined as the total amount of reactive nitrogen (all N species except N_2) released to the environment as a result of an entity's consumption patterns [8]. The first country-specific nitrogen footprints were calculated for USA and the Netherlands using the N-calculator tool [15]. This Web-based tool helps individuals in estimating their contribution to N losses through their consumption, energy use and purchase of goods and services.

2 Addressing the Problem

Sustainability of agricultural system depends on its water, carbon and nitrogen footprint. Quantification of these footprints can help in assessing the impact of conventional cultivation practices on water use as well as emission of GHGs and pollution load caused due to the agricultural operations. Henceforth, alternate strategies can be identified to reduce the footprints and thereby improve the environmental quality. Water, carbon and nitrogen footprinting tools will help in providing a framework to make decisions about use of natural resources and identify options to decrease the footprints, thereby improve the environmental quality and improve resource use efficiency.

The Indo-Gangetic Plain (IGP) is the backbone of food security of the country due to its higher productivity. But higher production in this region is accompanied with

Table 1 Physico-chemical properties of soil in Taraori village

Parameters	Values
Sand (%)	47%
Silt (%)	24%
Clay (%)	29%
Texture class	Sandy loam
Bulk density (gm cm^{-3})	1.45
pH (1:2::soil: water)	7.7
Electrical conductivity (d S m^{-1})	0.42
Organic carbon (%)	0.55
Total Nitrogen (%)	0.063
Available N (kg ha^{-1})	291.6
Available P (kg ha^{-1})	23.9
Available K (kg ha^{-1})	383.4

heavy usage of agricultural inputs like irrigation, fertilizer, electricity and diesel. For the present study, four locations, i.e. Karnal, Delhi, Varanasi and Kalyani, have been selected which represent different transects of IGP.

2.1 Collection of Primary and Secondary Data

Rice and wheat crop are the major crops grown in the regions. Farmers' fields were selected in Taraori village of Karnal, Haryana. Farmers are practicing conventional and conservation agricultural practices in the region. Soil samples were collected during the rice–wheat season. Samples were analysed for their different physico-chemical properties (Table 1). During harvesting of rice and wheat, grains as well as straw yields were quantified. Rice yield varied from 3.9 to 4.1 Mg ha^{-1}, while wheat yield was 4.9 and 5.4 Mg ha^{-1}.

Secondary data on climate, soil type, major crops and input use was collected for the locations. Daily weather data was collected for 5 years (2005–2009) which was used for quantification of water footprint of the crops. Rainfall received during the kharif season was highest in Kalyani (1225 mm), while Karnal received minimum rainfall (423 mm) during this period (Table 2). Survey was conducted in farmer's fields, and data related to crops grown, management practices and crop yield was collected for calculation of water, C and N footprint.

Table 2 Weather parameters of selected sites

Site	Mean temperature (kharif) (°C)	Mean temperature (rabi) (°C)	Kharif rainfall (mm)	Rabi rainfall (mm)
Karnal	28.3	16.2	423	105
Delhi	28.8	16.9	529	49
Varanasi	28.6	19.8	783	77
Kalyani	29	22.5	1225	130

2.2 Methodology for Quantification of Water, C and N Footprints of Crops

InfoRCT model was selected for calculation of C and N footprint in the study region. The Information on Use of Resource-Conserving Technologies (InfoRCT) is a model, developed for simulating GHG emissions, C and N fluxes [19]. The model integrates information on climate, crop, soil type and inputs used to calculate GHGs emission. The inputs and outputs are calculated on a season basis using the target-oriented approach [17, 20]. GHGs emission was calculated based on the amount of input used and its related soil-plant-atmospheric processes. The model is programmed in Microsoft Excel containing different parameters organized in different worksheets. The worksheets can be amended easily for assessing different technologies. A schematic diagram showing the inputs used, driving variables and outputs in the model is shown in Fig. 1.

The model requires input data on climate, soil type, crops and input use. Among climate data, annual temperature and rainfall as well as seasonal temperature and rainfall were collected. Soil properties like soil organic carbon (SOC), bulk density (BD), soil moisture, N, P and K content were used as inputs in the model. Management inputs like amount of fertilizer used, amount of irrigation, human labour, machine labour, biocide used were collected during survey.

Water footprint was calculated following methodology given by Hoekstra et al. [11]. Weather data of the location and crop management data collected from field were used to calculate water footprint. Crop water use was quantified using the CROPWAT 8.0 model of Food and Agriculture Organization [7]. CROPWAT is a programme that uses the Penman–Monteith methods for calculating reference crop evapotranspiration. The crop module of the model requires data on crop coefficient (kc), duration of crop growth stages, critical depletion fraction, yield response factor and crop height which were obtained from FAO 56 paper and also from the experimental data carried out in Indian Agricultural Research Institute. Soil parameters required for running the model were obtained from the experiments carried out in the institute and used for calibration of the model. Management data related to irrigation

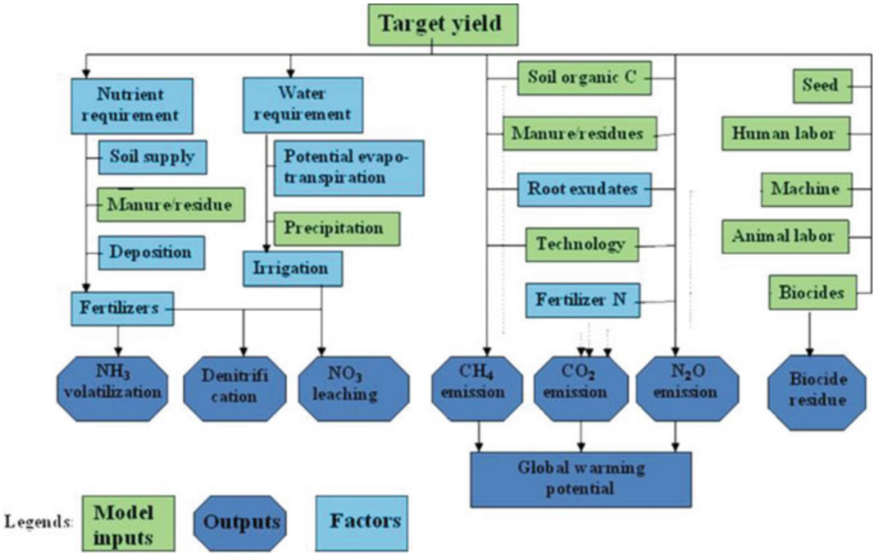


Fig. 1 Schematic overview of the inputs, driving variables and outputs in InfoRCT model (Pathak and Wassman [17]

used was incorporated in the model to finally calculate crop water use. A green water ET was obtained from the minimum values between total crop ET and effective precipitation (Peff). Blue water ET was calculated as the difference between total ET and Peff [12].

3 Results and Discussion

3.1 Carbon Footprint of Rice and Wheat Crop

In the present study, CFP of rice and wheat crops was quantified by taking into account the total amount of GHG emission during crop cultivation. Emissions of GHGs from soil, fertilizer application and farm operations were estimated. Nitrous oxide emission occurred from soils and fertilizer application in both rice and wheat crop, while CH_4 was emitted only in rice crop. On-farm operations like use of farm machineries led to CO_2 emission.

CFP of rice production is higher than that of wheat irrespective of the locations due to the fact that during rice cultivation both methane (CH_4) and nitrous oxide (N_2O) gases are emitted, while in wheat only N_2O is released (Table 3). Production of rice and wheat crop is associated with higher GHG emission in Karnal and Delhi as compared to Varanasi and Kalyani due to application of more amounts of inorganic

Table 3 Carbon footprint (kg CO_2 eq. kg^{-1}) of rice and wheat crop grown in different locations of Indo-Gangetic Plains (IGPs)

	Rice	Wheat
Karnal	1.09	0.16
Delhi	1.26	0.12
Varanasi	1.03	0.21
Kalyani	0.80	0.16

Table 4 Water footprint (m^3 kg^{-1}) of rice and wheat crop grown in different locations of Indo-Gangetic Plains (IGPs)

	Rice	Wheat
Karnal	2.66	0.53
Delhi	2.68	0.63
Varanasi	2.66	1.34
Kalyani	1.92	1.27

fertilizers, irrigation water and use of farm machineries. This has led to higher CFP of rice in these locations. In case of wheat, GHG emission was maximum in Karnal and minimum in Kalyani region. Lower use of nitrogenous fertilizer in Kalyani has led to lower N_2O emission, while higher dose of nitrogen fertilizer in Karnal led to more emission of N_2O.

3.2 Water Footprint of Rice and Wheat Crop

WFP of rice production was higher in Karnal, Delhi and Varanasi as compared to Kalyani region (Table 4). This is mainly attributed to the fact that evapotranspiration in these three locations was higher than that of Kalyani resulting in higher WFP. In case of wheat crop, WFP was less in Delhi and Karnal than Varanasi and Kalyani due to higher yield of wheat crop as well as lower evapo transpiration in these locations (Table 4).

In Karnal and Delhi, blue water footprint was 62% while green water footprint was 38% in rice crop (Fig. 2). Less rainfall and application of more amount of irrigation water have resulted in higher blue WFP of rice in these locations. On the other hand, blue water footprint was less than green water footprint in Kalyani region. Higher amount of rainfall and less application of irrigation water have caused high green water footprint in this region.

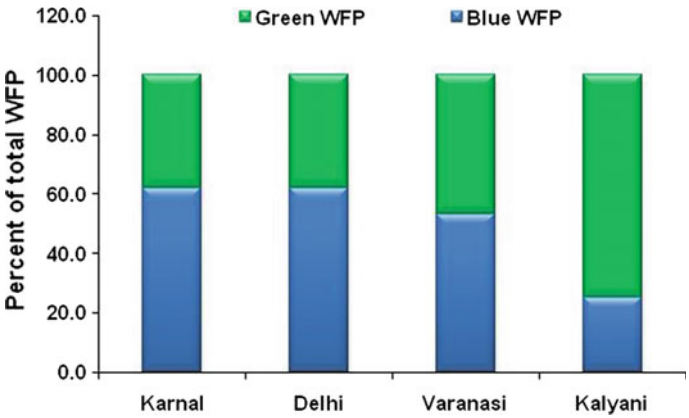


Fig. 2 Percent share of blue and green water footprint in rice crop in different locations of IGP

3.3 N Footprint of Rice and Wheat Crop

The InfoRCT model was used to calculate nitrogen loss from crop fields in Taraori village, Karnal, Haryana. N loss as nitrous oxide (N_2O) emission, nitrate leaching (NO_3^-) and N loss through ammonia (NH_3) volatilization were taken into account. Data related to fertilizer use was collected by survey from farmers' field. Soil nitrogen was used as input in the model. Crop, soil, weather and input use data were incorporated in the model for calculation of N loss from rice and wheat crop fields. The model is being further calibrated for estimating N loss using field data and data collected from the literature.

3.4 Geospatial Mapping of C Footprint

Secondary data related to weather, crop, soil properties and management practices was collected. District-wise yield data of rice and wheat crop and fertilizer application rates were collected from India agristat website (www.indiagristat.com). Carbon footprint values were calculated for all districts of IGP using InfoRCT model. The CFP maps were generated using QGIS software.

CFP of wheat in different districts ranged from 0.11 to 0.28 kg CO_2 eq•kg^{-1} produce (Fig. 3). Higher yield of wheat crop in Punjab and Haryana has resulted in lower CFP. In some places CFP values were. 0.2 to 0.25 kg CO_2 eq • kg^{-1} was observed which is mainly attributed to less yield of wheat crop in those districts.

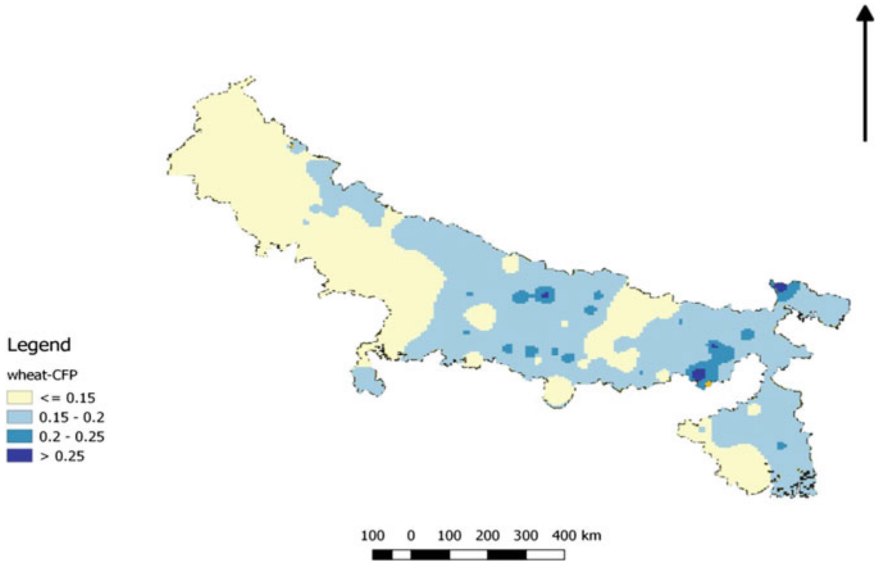


Fig. 3 Carbon footprint of wheat production in Indo-Gangetic Plains

4 Conclusions

In pursuit of sustainable agricultural system, there is a need that researchers as well as developmental agencies become more aware of the importance of developing and implementing less input-intensive technologies in agricultural operations. In future, technologies developed for large-scale implementation might also be assessed for their footprints on the natural environment. Quantification of water, C and N footprints of any technology developed will be beneficial for assessing its economic as well as ecological implications. Although research has been going on for estimating water, carbon and nitrogen footprints of different agricultural products, still further work is needed for refining the estimates. Climatic condition, soil type and management practices followed by farmers are very much diverse. Hence, the footprints of agricultural products will be different for different locations. Region-specific emission factors are required to quantify the CFP and NFP. Hence, more research work is needed in this field in order to reduce the uncertainties and generate footprints of crops for various regions. Limitation of data poses difficulty in estimating the footprints during the whole life cycle of agricultural products. Besides this, quantification in larger spatial scale is also very limited. In the following study, an attempt was made to quantify water, carbon and nitrogen footprints of rice and wheat crop. Footprint maps have been prepared for the IGP region. However, these were calculated on the basis of a limited number of data sets. Hence, the values may not be taken as final. More information is being collected, and more experimental data is being generated to further fine-tune these values.

References

1. Bidisha, C., Kumar, S.N., Pathak, H.: Carbon footprint of agricultural products. In: Muthu, S.S. (ed.) The Carbon Footprint Handbook, pp. 431–449. CRC Press andTaylor & Francis Group (2015)
2. Chakrabarti, B., Mina, U., Pramanik, P., Sharma, D.K.: Water footprint of transplanted and direct seeded rice. In: Mishra, G.C. (ed.) Environmental Sustainability: Concepts, Principles, Evidences and Innovations, pp. 75–79 (2014)
3. Chapagain, A.K., Hoekstra, A.Y.: The green, blue and grey water footprint of rice from both a production and consumption perspective. Value of Water, Research Report Series No. 40 UNESCO-IHE Institute for Water Education, Delft, the Netherlands, pp. 1–48 (2010)
4. Chapagain, A.K., Hoekstra, A.Y.: Virtual water flows between nations in relation to trade in livestock and livestock products. Value of Water Research Report Series No. 13, UNESCO-IHE, Delft, the Netherlands (2003)
5. Cheng, K., Pan, G., Smith, P., Luo, T., Li, L., Zheng, J., Zhang, X., Han, X., Yan, M.: Carbon footprint of China's crop production—an estimation using agro-statisticsdata over 1993–2007. Ag. Ecosys. Environ. **142**, 231–237 (2011)
6. Deumling, D., Wackernagel, M., Monfreda, C.: Eating up the earth: how sustainable food systems shrink our ecological footprint. Agriculture Footprint Brief, July 2003. Redefining Progress. http://www.redefiningprogress.org/newpubs/2003/ag_food_0703.pdf (2003)
7. FAO: CROPWAT 8.0 model, Food and Agriculture Organization, Rome, Italy, www.fao.org/nr/water/infores_databases_cropwat.html (2009)
8. Galloway, J.N., Winiwarter, W., Leip, A., Leach, A.M., Bleeker, A., Erisman, J.W.: Nitrogen footprints: past, present and future. Environ. Res. Lett. **9(**115003), 11 (2014)
9. Hoekstra, A.Y.: Virtual water trade: proceedings of the international expert meeting on virtual water trade. Value of Water Research Report Series No. 12, UNESCO-IHE, Delft, The Netherlands (2003)
10. Hoekstra, A.Y., Chapagain, A.K.: Globalization of Water: Sharing the Planet's Freshwater Resources, p. 232. Blackwell Publishing, Oxford, UK (2008)
11. Hoekstra, A.Y., Chapagain, A.K., Aldayal, M.M., Mekonnen, M.M.: Water Footprint Manual—State of the Art 2009. http://www.waterfootprint.org/downloads/WaterFootprintManual2009.pdf (2009)
12. Holcomb, M.K.: Quantifying the water footprint: growing crops sustainably in Northwest India. In: Paper submitted in Partial Requirements of the Undergraduate Honors Program, The Ohio State University, pp. 1–34 (2010)
13. INCCA (Indian Network for Climate Change Assessment): Assessment of the Greenhouse Gas Emission: 2007, The Ministry of Environment & Forests, Govt. of India, New Delhi (2010)
14. Ladha, J.K., Pathak, H., Krupnik, T.J., Six, J., Van Kessel, C.: Efficiency of fertilizer nitrogen in cereal production: retrospect and prospects. Adv. Agron. **87**, 85–156 (2005)
15. Leach, A.M., Galloway, J.N., Bleeker, A., Erisman, J.W., Kohn, R., Kitzes, J.: A nitrogen footprint model to help consumers understand their role in nitrogen losses to the environment Environ. Dev. **1**, 40–66 (2012)
16. Mosier, A.R., Wassmann, R., Verchot, L., Palm, C.: Greenhouse gas emissions from tropical agriculture: sources, sinks and mechanisms. Environ. Dev. Sustain. **6**, 11–49 (2004)
17. Pathak, H., Wassmann, R.: Introducing greenhouse gas mitigation as a development objective in rice-based agriculture: I Generation of technical coefficients. Agril. Syst. **94**, 807–825 (2007)
18. Pathak, H., Jain, N., Bhatia, A., Patel, J., Aggarwal, P.K.: Carbon footprints of Indian food items. Ag. Ecosys. Environ. **139**, 66–73 (2010)
19. Pathak, H., Saharawat, Y.S., Gathala, M., Ladha, J.K.: Impact of resource-conserving technologies on productivity and greenhouse gas emission in rice-wheat system. Greenh. Gas Sci. Technol. **1**, 261–277 (2011)
20. Ponsioen, T.C., Hengsdijk, H., Wolf, J., Van Ittersum, M.K., Roetter, R.P., Son, T.T., Laborte, A.G.: TechnoGIN—a tool for exploring and evaluating resource use efficiency of cropping systems in East and Southeast Asia. Agril. Syst. **87**, 80–100 (2006)

21. Rose, S.K., McCarl, B.A.: Greenhouse gas emissions, stabilization and the inevitability of adaptation: challenges for U.S. agriculture. Choices: The Magazine of Food, Farm & Resource Issues 23:23–18 (2008)
22. Wackernagel, M., Rees, W.: Our Ecological Footprint: Reducing Human Impact on the Earth New Society Publishers. Gabriola Island, Canada, BC (1996)
23. Wiedmann, T., Minx, J.: A Definition of 'Carbon Footprint'. In: Pertsova, C.C. (ed.), Ecological Economics Research Trends: Chapter 1, pp. 1–11. Nova Science Publishers, Hauppauge NY, USA (2008)
24. Williams, H., Wikstrom, F.: Environmental impact of packaging and food losses in a life cycle perspective: a comparative analysis of five food items. J. Clean. Prod. **19**, 43–48 (2011)

Delineation of Regional Scale Gold Prospective Zones within the North Singhbhum Mobile Belt, Eastern India: An Integrated Approach Through Geospatial Technologies

S. Singh and A. K. Prasad

1 Introduction

The recent development in hyperspectral remote sensing methods and its evolution as a mature technology, during the last 35 years, has gained widespread acceptance in various geological applications like soil mapping, lithological characterization, lineament delineation, minerals identification, and the exploration of economic minerals. Sabins [1] extensively used multispectral Landsat data and the hyperspectral imagery to identify potentially mineralized host rocks by utilizing signatures of hydrothermally altered and unaltered rocks through space-based satellite images. However, multispectral satellite images have limitations with respect to their spectral resolution and the number of bands available to the user for identification of specific minerals. The invention of sensors, having high spectral resolution, led to the development of a new technique to collect data related to earth surface composition. Hyperspectral imagery provides a huge array of bands at high spectral resolution that enables identification of individual minerals. For instance, instead of the identification of only iron-bearing minerals through multispectral imagery, the distribution of individual minerals such as hematite, goethite, jarosite, and pyrite can be identified using hyperspectral imaging. The presence of specific minerals can be used to determine the metal leachability and pH condition of the study area. Similarly, the information regarding the individual clay minerals, i.e., illite, montmorillonite, and kaolinite content in soil can be obtained from hyperspectral imagery and hyperspectral library. These information/data can be verified later on with ground-based instrument, which can collect spectra of rocks and minerals present in the area. After

S. Singh (✉) · A. K. Prasad
Department of Applied Geology, Indian Institute of Technology (Indian School of Mines), Dhanbad, India
e-mail: sahendra@iitism.ac.in; sahendrasingh02@gmail.com

A. K. Prasad
e-mail: anup@iitism.ac.in; anupiitk@gmail.com

N. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0_30

the ground validation of the hyperspectral response of rock types through field visits in the study area followed by geochemical analysis, the work can further be extended to the entire study area to identify the potential zones for future analysis.

The recent finding of gold prospects at various locations within the Chandil Formation of North Singhbhum mobile belt (NSMB) has generated a debate over the exploration policy. This policy needs to be revived to achieve considerable targets in terms of the finding of new gold deposits. The higher resolution geological information is unavailable for the North Singhbhum mobile belt. This is further complicated by the inaccessible nature of the terrain. In such a scenario, remote sensing technology is an effective tool to find the gold prospects. Generation of a GIS geodatabase of the study area, i.e., lithological mapping, surface structural pattern, and hydrothermal alteration zones by using hyperspectral imagery would greatly help in exploration work of the study area. Our work aims to use the integrated applications of remote sensing and GIS technology to evaluate the regional scale parameters related to geological/structural control for further simulation and delineation of the prospects related to gold, base metal, and PGE resources.

1.1 Remote Sensing for Gold Exploration

The gold prospect of any area is exclusively related to the mineralizing ore system. Since these ore systems are manifested on the surface in the form of hydrothermal alteration zones, the mapping of these alteration zones is the key to find the gold prospects. Similar types of investigations have been carried out in Carlin-type ore system [2], orogenic lode gold [3], skarns, calcic skarns [4–6], and volcanogenic massive sulfide (VMS) ore deposits [7]. The manifestation of the hydrothermal system as studied for the earth surface has also been used to study the hydrothermal system of Mars [8–12].

The component of hyperspectral remote sensing has been included in this work to quantify the lithology and mineralogy of the earth surface from calibrated spectra, i.e., radiance, reflectance, and emissivity, which will be acquired as images in several narrow and contiguous bands. For hyperspectral data classification, two categories of analytical techniques are preferred—spectral matching techniques and subpixel methods [13]. The spectral matching technique is based on the spectral similarity of the reference (library or field spectra of known minerals) and image spectra. The common method is the spectral fitting approach, which looks for the similarity between the reference spectra and the pixel spectra of the hyperspectral images [14]. The other method is subpixel method, which is the technique of unmixing the hyperspectral images with the objective of finding relative abundance of minerals within a pixel. This method has the advantage of adopting knowledge-based approach. There are also other statistical techniques to identify the end members [15–17]. Another popular technique is the pixel purity index (PPI) [18–21].

Such types of gold exploration projects were successively carried out in Patagonia, Argentina, where in the initial phase, the regional scale gold prospects were delineated on the basis of alteration zones, which were mapped using multispectral Landsat Thematic Mapper (TM) data. During the second phase, advanced spaceborne thermal emission and reflection radiometer (ASTER) data were used to characterize the alteration mineral assemblages of the prospective area. In this phase, principle component analysis (PCA) was used to analyze the ASTER multispectral data. The results were checked against the field spectroscopic data on selected sites representing end members. During the third phase, hyperspectral approach was used for the processing of ASTER images for the spatial distribution of the alteration mineral assemblages [22]. The results obtained were outstanding, and several gold targets were identified over an area of 60,000 km^2 in northern Patagonia in Argentina. Hyperspectral remote sensing methods were also used extensively for gold mineralizing systems, especially epithermal gold system [23–29], where the altered minerals were used as a vector to delineate the ore and provided the basis to identify gold prospects.

2 Geological Overview of the Study Area

The Palaeo-to Mesoproterozoic North Singhbhum mobile belt (NSMB) is a 200 km east–west trending arcuate belt, lying between the Archean Singhbhum craton, i.e., Archean cratonic core region (ACCR > 2.4 Ga) in the south, and the Meso-Neoproterozoic (0.9–1.7 Ga) Chotanagpur gneissic complex (CGC) in the north (Fig. 1). The mobile belt has been further subdivided into five broad litho-tectonic domains. The study area belongs to the fourth and fifth domain of the North Singhbhum mobile belt, referred as the Dalma volcano-sedimentary belt and Chandil Formation, respectively.

Gold occurrences from this part of the NSMB have been reported from the meta-volcano-sedimentary units at several locations; however, most of the prospects are confined to the north of the Dalma volcano-sedimentary belt (Fig. 2). Gold mineralization occurs mostly in association with sulfides such as pyrite, arsenopyrite, pyrrhotite, chalcopyrite, and sphalerite. Arsenopyrite and pyrite are closely associated with the gold occurrences in the area. Free-milling gold ore and the sulfide-hosted refractory gold occurrences have been reported from the carbonaceous phyllite/reef quartzite in and around Rudia area [30]. The rock formations lie at the eastern most extension of east–west trending carbonaceous metasedimentary units (study area 4) lying along the northern fringe of Dalma volcano-sedimentary belt (Fig. 2). The present work aims to explore the refractory gold potential of the various gold prospects, i.e., Parasi-Sindauri (study area 2), Babaikundi-Birgaon (study area 1), Lawa-Mayasera (study area 1), Rudia (study area 3), and Tamar–Porapahar Shear Zone (TPSZ), i.e., South Purulia Shear Zone(SPSZ) (study area 1 and 2).

Gold panning has been reported along the river Subarnarekha from pre-independence period and it is still being carried out by local tribes in east Singhbhum region (Fig. 3a and b). Samples for the present study were collected from different parts of the NSMB, but mostly from the northern fringe of the Dalma volcano-sedimentary belt. Petrographic and ore microscopic studies were carried out for the

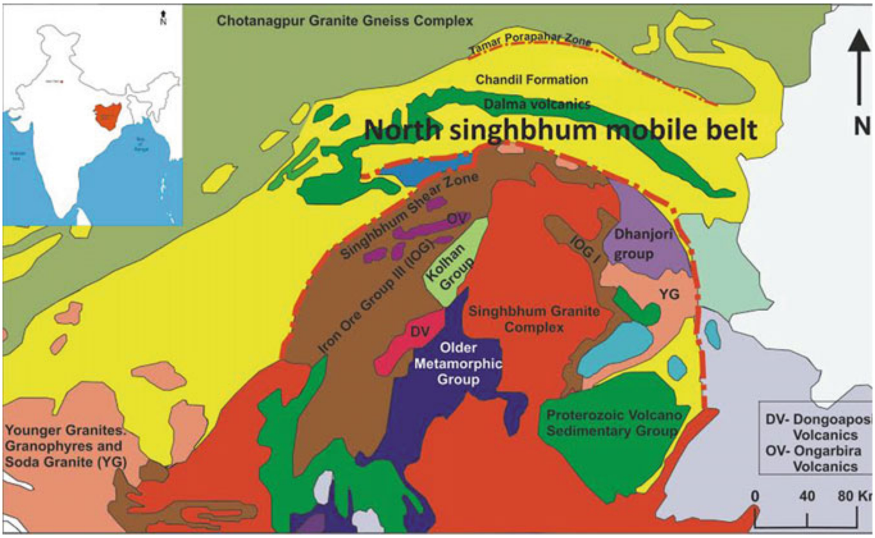


Fig. 1 Regional geological setting of North Singhbhum mobile belt (NSMB) (modified after Saha, 1994)

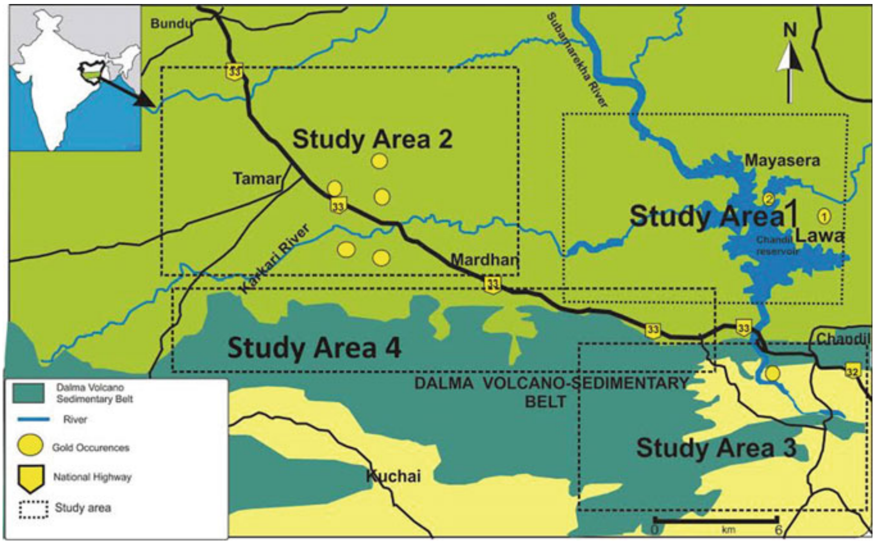


Fig. 2 Gold prospective zones within the NSMB

mineralogical characterization, i.e., identification of oxides and sulfides (Fig. 3c–f). These samples were further analyzed by SEM-EDX, electron probe microanalysis (EPMA), and LA-ICP-MS for their trace/REE and gold content. In Rudia area, shear controlled gold prospective regions lie close to the northern boundary of Dalma belt.

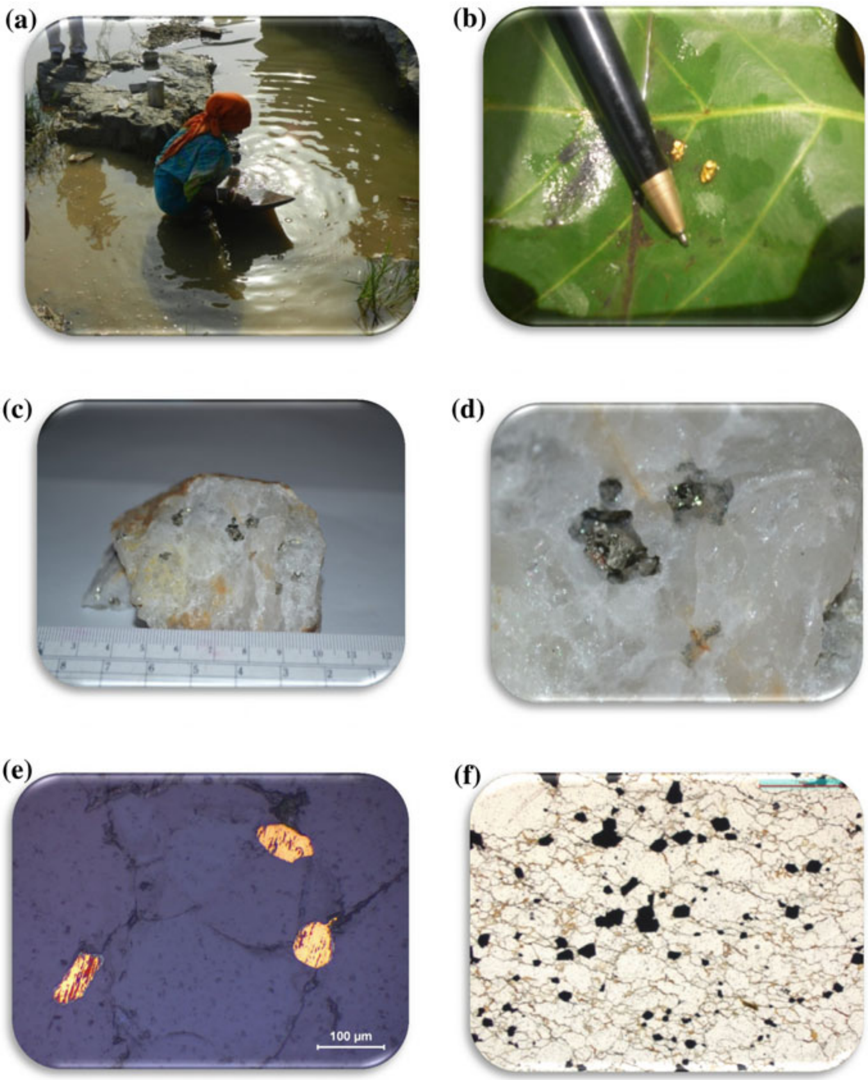


Fig. 3 **a** Gold panning along Subarnrekha River near Chandil, Jamshedpur, **b** Panned gold grain from Subarnrekha River near Chandil, Jamshedpur, **c** Pyrite grains within the reef quartz of Babaikundi, Tamar block, Ranchi district, **d** Pyrite grains within the reef quartz of Babaikundi, (enlarged view), **e** Gold grains within the reef quartz of Babaikundi, Tamar block, Ranchi [31], f. Photomicrograph of pyrite grain from Lawa-Mayasera area, Saraikela-Kharsawan district

Table 1 Electron probe microanalysis of pyrite with the metasedimentary units of Rudia area near Jamshedpur, East Singhbhum District

Mineral phase (wt%)	S	Fe	Zn	Cu	Co	Ni	As	Ag	Pb	Au	Bi	Total
Pyrite	53.87	46.70	0.59	0.25	0.29	−0.05	0.09	0.34	2.52	0.31	0.50	105.84
Pyrite	47.53	46.71	0.43	0.33	0.37	0.08	0.00	0.22	2.78	0.18	0.38	99.15
Pyrite	54.13	46.18	0.59	0.38	0.26	0.15	0.08	0.25	1.93	0.26	0.46	105.05
Pyrite	53.15	46.58	0.42	0.25	0.35	−0.10	0.23	0.25	2.52	0.22	0.38	104.54
Pyrite	53.65	46.16	0.40	0.18	0.25	−0.12	0.19	0.28	2.37	0.29	0.59	104.58
Pyrite	45.79	42.42	0.38	0.23	0.28	0.20	0.03	0.18	2.01	0.23	0.40	92.44
Pyrite	53.65	46.00	0.47	0.51	0.42	−0.10	0.27	0.27	2.61	0.22	0.38	105.11
Pyrite	53.85	45.65	0.70	0.22	0.21	0.10	0.03	0.35	2.98	0.25	0.36	104.94
Pyrite	53.25	45.89	0.51	0.29	0.34	−0.10	0.02	0.27	2.45	0.22	0.43	104.06
Pyrite	54.30	46.46	0.53	0.37	0.20	0.17	0.19	0.23	2.68	0.15	0.42	106.02
Pyrite	53.07	44.69	0.73	0.32	0.33	0.05	0.09	0.34	4.67	0.31	0.41	105.44
Pyrite	44.94	45.92	0.68	0.20	0.17	0.31	0.18	0.31	−0.13	0.15	0.53	93.40
Pyrite	53.47	45.69	0.49	0.41	0.35	0.05	0.31	0.26	1.86	0.27	0.46	104.08
Pyrite	53.90	45.49	0.65	0.09	0.32	0.10	0.13	0.26	2.50	0.19	0.45	104.54
Pyrite	54.28	46.39	0.61	0.41	0.31	−0.05	0.08	0.31	2.21	0.18	0.51	105.55

The mineralization is mainly hosted by phyllite and quartzite. A substantial amount of invisible gold with its concentration varying from 200 to 1000 ppm (spot analysis) has been reported from the quartz-reef-hosted sulfides in and around Babaikundi-Birgaon axis [31]. Pyrite-hosted gold content varying from 7.0 g/t to 18 g/t has been reported from Lawa-Mayasera area [32]. EPMA studies indicate significant concentration of invisible/refractory gold within the pyrites (Table 1) in and around Rudia area (Fig. 4a and b) and gold concentration up to 3000 ppm (spot analysis) within the pyrites and arsenopyrite in the SPSZ (Fig. 4c and d) [30, 33].

3 Data and Methodology

In this work, we propose the use of Landsat (TM) and ASTER data for the regional alteration mapping followed by the hyperspectral analysis for target selection at deposit scale. Landsat TM data were used by the geoscience communities for various thematic applications, which include lithological, structural, and other issues related to mineral exploration [1, 34–40]. Landsat TM data were also used and integrated with other airborne gravity, magnetic and gamma-ray survey, and spaceborne remote sensing data for integrated mapping approaches. The advanced spaceborne thermal emission and reflection radiometer, ASTER, [41–43] is considered suitable for the mineral-mapping capabilities. Multispectral images were used to study signatures

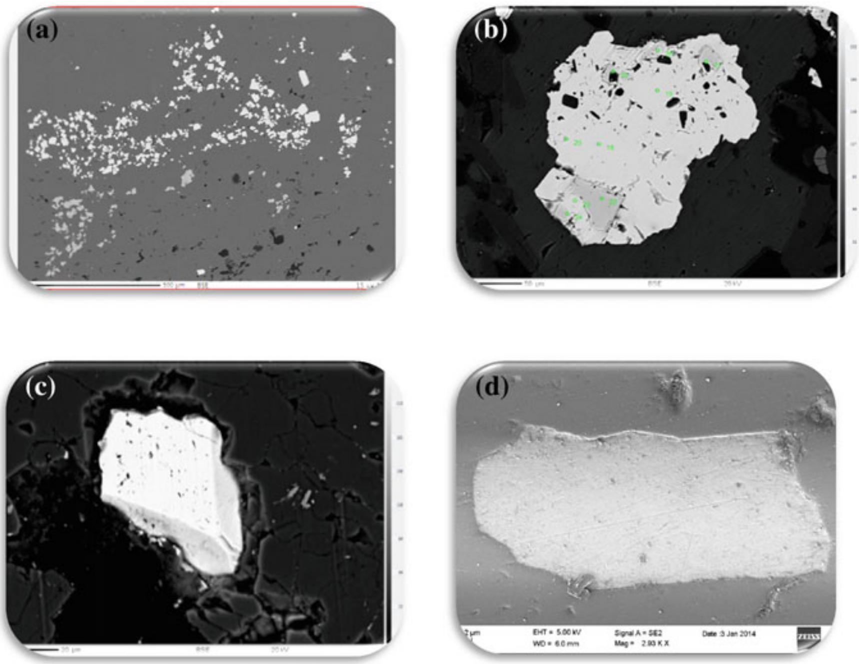


Fig. 4 **a** BSE image of disseminated pyrite grain within quartzite from Rudia area, Chandil Jamshedpur, East Singhbhum, **b** BSE image of pyrite and pyrrhotite grain within quartzite from Rudia area, Chandil, Jamshedpur, East Singhbhum, **c** BSE image of Arsenopyrite within the phyllite South Purulia Shear Zone (SPSZ), d. BSE image of gold grains from Babaikundi area, Tamar Block, Ranchi district [31]

of various iron-bearings, sulfide-bearings, clay-bearing minerals, rocks, and soils by several researchers in the past [1].

Spectral reflectance of transition metal ions (e.g., Fe, Cr, and Ni) is confined to the VNIR wavelength as the characteristics absorption feature may be attributed to the electronic processes. SWIR is appropriate for the determination of the presence or absence of water and hydroxyl, carbonate, and sulfate. The characteristic absorption features also depend upon the grain size as the scattering and absorption phenomena is dependent upon the grain size. In general, absorption band is correlated with the abundance of materials. It has been observed that the abundance of kaolinite can be derived with an accuracy of about 2 wt% and organic carbon can be quantified with an accuracy of about 2 wt% [44]. The role of organic carbon in the refractory gold mineralization within the carbonaceous metasedimentary rocks of Rudia area has been discussed by previous workers [30].

The current study aims to derive the distribution of clay minerals, oxidized iron-bearings, sulfides, and other ferrous/iron-bearing minerals using multispectral Landsat-8/Operational Land Imager (OLI) imagery dated 2014.09.27. The Landsat-8 data (Fig. 6a) have been obtained from USGS Landsat repository (Earth Explorer

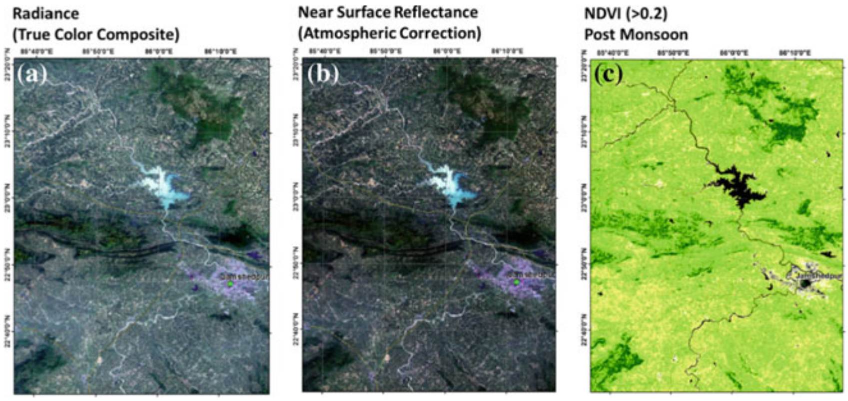


Fig. 5 True color composite as derived from Landsat-8 imagery dated 2014.09.27 around Chandil, Tamar, and Jamshedpur region using **a** Radiance, and **b** Near-surface reflectance (with atmospheric correction). **c** High vegetation density and shrub areas (NDVI > 0.2, shown in green color) over the study region masks signature of minerals from the rocks and sediments. Atmospheric correction, Fig. 5b) obtained from Landsat-8/OLI imagery dated 2014.09.27 around Chandil, Tamar, and Jamshedpur. The deep green color represents dense vegetation (NDVI > 0.6) and black color represents water bodies (lake, streams, rivers, and ponds) around the study region

https://earthexplorer.usgs.gov/). The Landsat imagery was corrected for atmospheric effects using dark object subtraction (DOS) model. Top-of-the-atmosphere (TOA) reflectance (not shown here) and near-surface reflectance (Fig. 6b) have been derived to explore the distribution of group of minerals using well-known band-ratio techniques. The derivation of normalized difference vegetation index (NDVI) over the study region shows high values (NDVI > 0.2) that indicates the presence of dense shrubs and trees around the study region (Fig. 5c). The presence of dense shrubs and trees tends to mask the mineralogical characteristics of the surface. However, open/barren regions show strong signatures for the presence of specific group of minerals. Mineralogical mapping (Fig. 7) has been performed using the following band-ratio techniques:

Clay minerals (clay minerals ratio): Relative distribution of clay minerals has been obtained using band ratio of shortwave-infrared (SWIR 1 and SWIR 2 bands). This ratio highlights the presence of clay and alunite in the hydrothermally altered rocks [45].

Oxidized iron-bearing sulfides (iron-oxide ratio): Relative distribution of iron-oxide bearing minerals has been obtained using band ratio of red and blue bands. The hydrothermally altered rocks, particularly oxidized iron-bearing sulfides, is highlighted by band ratio of red and blue bands [45, 46].

Other ferrous/iron-bearing minerals (ferrous minerals ratio) have been highlighted by band ratio of SWIR and near infrared (NIR) bands. This is one of the commonly used indexes using multispectral satellites [45, 46].

4 Results and Conclusion

The mineralizing ore system of the North Singhbhum mobile is similar to the other major gold producing orogenic metamorphic belts and has all the critical fundamental controls, which are required for the formation of gold deposits at regional scale. The ore system is characterized by the presence of a crustal scale shear zone, which provides the conduit for the progressive/repetitive transfer of gold from the deeper part to the all important shallow level seismogenic regime (brittle ductile regime) and has a primary control on the gold enrichment process. The locations with known gold occurrences show the presence of arsenopyrite, pyrite, pyrrhotite, and other sulfide minerals, which is manifested in satellite imagery (band-ratio images) in the form of strong signatures for oxidized iron-bearing sulfides (iron-oxides ratio) and other ferrous/iron-bearing minerals surrounded by the presence of altered rocks (clay-bearing minerals) (Fig. 6a, b, c). For instance, the hills to the east of Chandil reservoir, especially around open/barren regions, clearly show zones of enriched oxidized iron-bearing sulfides and other ferrous/iron-bearing minerals bounded by

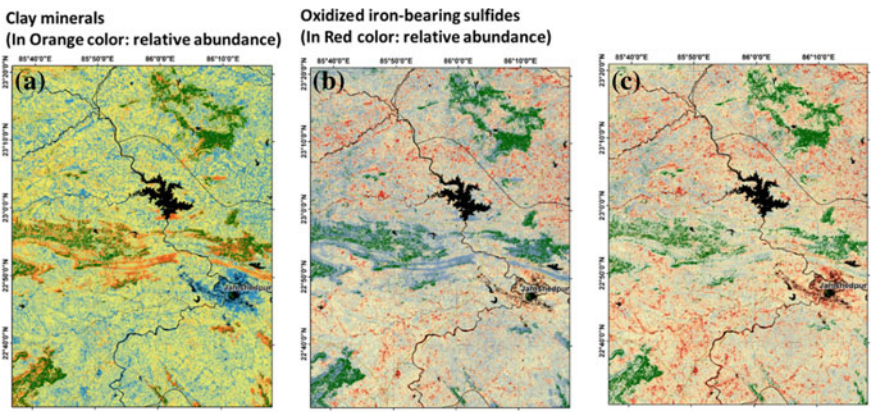


Fig. 6 Mineralogical mapping and relative abundance of a group of minerals. **a** Clay group of minerals (highlighted in the orange color), **b** Oxidized iron-bearing sulfides (Iron-oxides ratio) (highlighted in the red color), and **c** Other ferrous/iron-bearing minerals (highlighted in the red color), as derived from near-surface reflectance

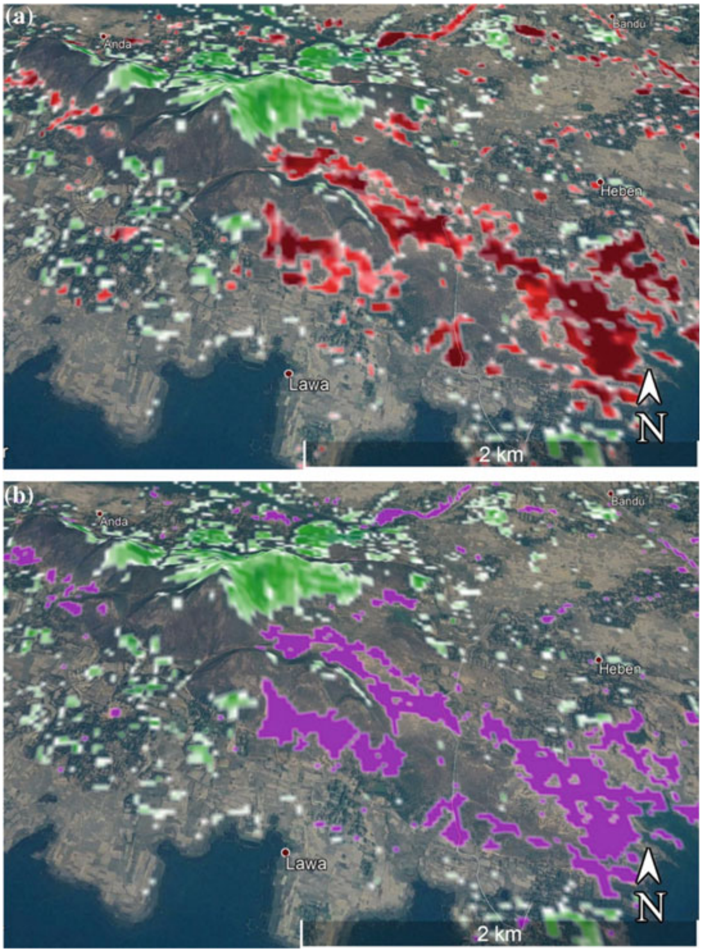


Fig. 7 The southern slope of hills, which is situated in the north of Lawa and east of Chandil reservoir, is one of the known sites with gold occurrences (Fig. 3a and b). The barren/open area around the hills shows conspicuous signatures and anomalies with respect to iron oxides and other iron-bearing minerals derived using Landsat-8/Operational Land Imager (OLI)-based mineralogical mapping. **a** Iron oxides (in red color), and **b** other iron-bearing minerals (in purple color) show the presence and distribution of prospective host rocks in this region which is surrounded by clay minerals (in green color)

clay-bearing minerals (Fig. 7). The distribution pattern of enriched zones around Chandil reservoir clearly shows strong lithological and structural controls. Analysis of hyperspectral dataset (ground and spaceborne) is underway over the study region that holds the key to not only detect individual minerals and alteration zones, but also the lithological and structural controls of prospective gold-bearing zones.

Acknowledgements This research work is supported by Department of Science and Technology (DST), New Delhi, and we gratefully acknowledge the research grant vide project No NRDMS (/01/07/014). The authors would like to thank the Director, Indian Institute of Technology (Indian School of Mines), Dhanbad for providing necessary facilities; IIT (ISM) Central Research Facility, for electron probe micro-analysis. Author is thankful to. Prof. N. V. Chalapati Rao and Dr. Dinesh Pandit, Banaras Hindu University for EPMA analysis and Atomic Minerals Directorate for Exploration and Research, Jamshedpur for their help in field work. SS is thankful to Ms. Mousoma Khatun, Ms. Subhashree Majumdar, and Mr Rajarshi Chakravarti, Research Fellow, Department of Applied Geology, Indian Institute of Technology (Indian School of Mines), Dhanbad for their help in field and laboratory work related to project work.

References

1. Sabins, F.F.: remote sensing for mineral exploration. Ore Geol. Rev. **14**(3–4), 157–183 (1999)
2. Rockwell, B.W., Hofstra, A.H.: Identification of quartz and carbonate minerals across northern Nevada using ASTER thermal infrared emissivity data-implications for geologic mapping and mineral resource investigations in well-studied and frontier areas. Geosphere **4**(1), 218–246 (2008)
3. Bierwirth, P., Huston, D., Blewett, R.: Hyperspectral mapping of mineral assemblages associated with gold mineralization in the Central Pilbara, Western Australia. Econ. Geol. **97**(4), 819–826 (2002)
4. Bedini, E.: Mapping lithology of the Sarfartoq carbonatite complex, southern West Greenland, using HyMap imaging spectrometer data. Remote Sens. Environ. **113**(6), 1208–1219 (2009)
5. Kozak, P.K., Duke, E.F., Roselle, G.T.: Mineral distribution in contact-metamorphosed siliceous dolomite at Ubehebe Peak, California, based on airborne imaging spectrometer data. Am. Miner. **89**(5–6), 701–713 (2004)
6. Rowan, L.C., Mars, J.C.: Lithologic mapping in the Mountain Pass, California area using advanced spaceborne thermal emission and reflection radiometer (ASTER) data. Remote Sens. Environ. **84**(3), 350–366 (2003)
7. Berger, B.R., King, T.V.V., Morath, L.C., Phillips, J.D.: Utility of high-altitude infrared spectral data in mineral exploration: application to northern Patagonia Mountains, Arizona. Econ. Geol. **98**(5), 1003–1018 (2003)
8. Aubrey, A., James Cleaves, H., Chalmers, J.H., et al.: Sulfate minerals and organic compounds on Mars. Geology **34**(5), 357–360 (2006)
9. Ehlmann, B.L., Mustard, J.F., Swayze, G.A., et al.: Identification of hydrated silicate minerals on Mars using MRO-CRISM: geologic context near Nili Fossae and implications for aqueous alteration. J. Geophys. Res. Planets **114**(E2) 2009
10. Farmer, J.D.: Hydrothermal systems on mars: an assessment of present evidence. In: Evolution of Hydrothermal Ecosystems on Earth and Mars, pp. 273–299 (1996)
11. Loizeau, D., Mangold, N., Poulet, F., et al.: Phyllosilicates in the Mawrth Vallis region of Mars. J. Geophys. Res. Planets **112**(E8) (2007)
12. Mangold, N., Gendrin, A., Gondet, B., et al.: Spectral and geological study of the sulfate-rich region of West Candor Chasma, Mars. Icarus **194**(2), 519–543 (2008)
13. van der Meer, F.: The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. **8**(1), 3–17 (2006)
14. Ning, X.U., Bin, L.E.I., et al.: Mineral information extraction for hyperspectral image based on modified spectral feature fitting algorithm. Spectrosc. Spectral Anal. **31**(6), 1639–1643 (2011)
15. Tompkins, S., Mustard, J.F., Pieters, C.M., Forsyth, D.W.: Optimization of endmembers for spectral mixture analysis. Remote Sens. Environ. **59**(3), 472–489 (1997)
16. Van Der Meer, F.: Iterative spectral unmixing (ISU). Int. J. Remote Sens. **20**(17), 3431–3436 (1999)

17. Zortea, M., Plaza, A.: A quantitative and comparative analysis of different implementations of N-FINDR: a fast endmember extraction algorithm. IEEE Geosci. Remote Sens. Lett. **6**(4), 787–791 (2009)
18. Chang, C.-I., Plaza, A.: A fast iterative algorithm for implementation of pixel purity index. IEEE Geosci. Remote Sens. Lett. **3**(1), 63–67 (2006)
19. Harsanyi, J.C., Chang, C.-I.: Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. IEEE Trans. Geosci. Remote Sens. **32**(4), 779–785 (1994)
20. Settle, J.J., Drake, N.A.: Linear mixing and the estimation of ground cover proportions. Int. J. Remote Sens. **14**(6), 1159–1177 (1993)
21. Shimabukuro, Y.E., Smith, J.A.: The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. IEEE Trans. Geosci. Remote Sens. **29**(1), 16–20 (1991)
22. Boardman, J.W., Kruse, F.A., Green, R.O.: Mapping Target Signatures Via Partial Unmixing of AVIRIS Data (1995)
23. Bedini, E., Van Der Meer, F., Van Ruitenbeek, F.: Use of HyMap imaging spectrometer data to map mineralogy in the Rodalquilar caldera, southeast Spain. Int. J. Remote Sens. **30**(2), 327–348 (2009)
24. Chen, X., Warner, T.A., Campagn, D.J.: Integrating visible, near-infrared and short wave infrared hyperspectral and multispectral thermal imagery for geologic mapping: simulated data. Int. J. Remote Sens. **28**(11), 2415–2430 (2007)
25. Crosta, A.P., Sabine, C., Taranik, J.V.: Hydrothermal alteration mapping at Bodie, California, using AVIRIS hyperspectral data. Remote Sens. Environ. **65**(3), 309–319 (1998)
26. Gersman, R., Ben-Dor, E., Beyth, M., et al.: Mapping of hydrothermally altered rocks by the EO-1 Hyperion sensor, Northern Danakil Depression, Eritrea. Int. J. Remote Sens. **29**(13), 3911–3936 (2008)
27. Kruse, F.A., Perry, S.L., Caballero, A.: District-level mineral survey using airborne hyperspectral data, Los Menucos, Argentina. Ann. Geophys. **49**(1) (2006)
28. Rowan, L.C., Crowley, J.K., Schmidt, R.G., Ager, C.M., Mars, J.C.: Mapping hydrothermally altered rocks by analyzing hyperspectral image (AVIRIS) data of forested areas in the southeastern United States. J. Geochem. Explor. **68**(3), 145–166 (2000)
29. van der Meer, F.: Indicator kriging applied to absorption band analysis in hyperspectral imagery: a case study from the Rodalquilar epithermal gold mining area, SE Spain. Int. J. Appl. Earth Obs. Geoinf. **8**(1), 61–72 (2006)
30. Khatun, M., Singh, S.: Petrographic and ore microscopic study of carbonaceous host rocks and associated gold mineralization within dalma volcano-sedimentary basin, North Singhbhum Mobile Belt, Eastern Indian Craton. J. Geosci. Res. **1**, 35–41 (2017)
31. Jha, V., Singh, S., Venkatesh, A.S.: Invisible gold occurrence within the quartz reef pyrite of Babaikundi area, North Singhbhum fold-and-thrust belt, Eastern Indian Shield: evidence from petrographic, SEM and EPMA studies. Ore Geol. Rev. **65**, 426–432 (2015)
32. Chandan, K.K., Jha, V., Roy, S., et al.: Ore microscopic study of the gold mineralization within chandil formation, North Singhbhum Mobile Belt, eastern India. Int. J. Earth Sci. Eng **6**, 213–222 (2014)
33. Majumdar, S., Singh, S., Sahoo, P.R.: Petrographic characterisation of host rocks and ore mineralisation from palaeo to mesoproterozoic South Purulia Shear Zone, Eastern Indian Craton. J. Geosci. Res. **1**, 43–50 (2017)
34. Abdelsalam, M.G., Stern, R.J., Berhane, W.G.: Mapping gossans in arid regions with Landsat TM and SIR-C images: the Beddaho Alteration Zone in northern Eritrea. J. Afr. Earth Sci. **30**(4), 903–916 (2000)
35. Boccaletti, M., Bonini, M., Mazzuoli, R., et al.: Quaternary oblique extensional tectonics in the Ethiopian Rift (Horn of Africa). Tectonophysics **287**(1–4), 97–116 (1998)
36. Ferrier, G., White, K., Griffiths, G., Bryant, R., Stefouli, M.: The mapping of hydrothermal alteration zones on the Island of Lesvos, Greece using an integrated remote sensing dataset. Int. J. Remote Sens. **23**(2), 341–356 (2002)

37. Fraser, A., Huggins, P., Rees, J., Cleverly, P.: A satellite remote sensing technique for geological structure horizon mapping. Int. J. Remote Sens. **18**(7), 1607–1615 (1997)
38. Gad, S., Kusky, T.: Lithological mapping in the eastern Desert of Egypt, the Barramiya area, using Landsat thematic mapper (TM). J. Afr. Earth Sc. **44**(2), 196–202 (2006)
39. Schetselaar, E.M., Chung, C.-J.F., Kim, K.E.: Integration of Landsat TM, gamma-ray, magnetic, and field data to discriminate lithological units in vegetated. Remote Sens. Environ. **71**(1), 89–105 (2000)
40. Yesou, H., Besnus, Y., Rolet, J.: Extraction of spectral information from Landsat TM Data and Merger with SPOT panchromatic imagery- a contribution to the study of geological structures. ISPRS J. Photogramm. Remote Sens. **48**(5), 23–36 (1993)
41. Abrams, M.: The advanced spaceborne thermal emission and reflection radiometer (ASTER): data products for the high spatial resolution imager on NASA's Terra platform. Int. J. Remote Sens. **21**(5), 847–859 (2000)
42. Abrams, M., Hook, S.J.: Simulated ASTER data for geologic studies. IEEE Trans. Geosci. Remote Sens. **33**(3), 692–699 (1995)
43. Yamaguchi, Y., Kahle, A.B., Tsu, H., Kawakami, T., Pniel, M.: Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER). IEEE Trans. Geosci. Remote Sens. **36**(4), 1062–1071 (1998)
44. Krüger, G., Erzinger, J., Kaufmann, H.: Laboratory and airborne reflectance spectroscopic analyses of lignite overburden dumps. J. Geochem. Explor. **64**(1–3), 47–65 (1998)
45. Drury, S.A., Stephen, A.: Image Interpretation in Geology. Allen and Unwin, London, Boston (1987). http://swbplus.bsz-bw.de/bsz012751383cov.htm
46. Segal, D.B.: Use of landsat multispectral scanner data for the definition of limonitic exposures in heavily vegetated areas. Econ. Geol. **78**(4), 711–722 (1983)

Index

N. L. Sarda et al. (eds.), *Geospatial Infrastructure, Applications and Technologies: India Case Studies*, https://doi.org/10.1007/978-981-13-2330-0