Spatial interpolation of rainfall for Mudumalai Wildlife Sanctuary and Tiger Reserve Tamil Nadu, India

CES Technical Report
Number 130

2013

H.S. Dattaraja, Sandeep Pulla, Nandita Mondal, H.S. Suresh, C.M. Bharanaiah, R. Sukumar

Centre for Ecological Sciences
Indian Institute of Science
Bangalore 560012
India

Project Supervisor: Prof. R. Sukumar

Data collection and compilation: C.M. Bharanaiah, H. S. Dattaraja, H. S. Suresh, Meghana Kulkarni, Meera Baban Mane, Nachiketha Sharma, Nandita Mondal, Raghavendra, Rutuja ChitraTarak

Rainfall imputation in R: Sandeep Pulla

Spatial interpolation in ArcGIS: Nandita Mondal

Correspondence for further information and permissions for data usage:
Prof. R. Sukumar
e-mail: rsuku@ces.iisc.ernet.in

Suggested citation:
Dattaraja, H.S., Pulla, S., Mondal, N., Suresh, H. S., Bharanaiah, C. M. and Sukumar, R. 2013. Spatial interpolation of rainfall for Mudumalai Wildlife Sanctuary and Tiger Reserve, Tamil Nadu, India. CES technical report no. 130, Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India.

Contents

Introduction 2
Rationale for the project 2
Description of the general movement patterns of the monsoon over the Indian subcontinent 2
Study area 4
Description of local topographical influences on rainfall in the study area 4

Methods 6
Data 6
Imputation of missing rainfall data 6
Seasonality 6
Spatial interpolation 8

Results 10
Seasonality 10
Spatial variations in rainfall across years 10
Annual average rainfall 12
Standard error of the prediction using kriging, and cross-validation 12

References 13

Appendices 14

Appendix A: List of rain gauges used, their location (latitude and longitude in decimal degrees), elevation (extracted from ASTER GDEM, Fig. 2) and the source of data... 14

Appendix B: R-code used for the imputation of annual rainfall from linear regressions 15

Appendix D: Details of the interpolation procedure used for generating the annual average rainfall map 24

List of figures and tables

Figure 1: General movement of south-west monsoon over the Indian subcontinent 1

Figure 2: Location of the study area with respect to the Nilgiri massif.. 3

Figure 3: Spatial distribution of the sixteen rain gauges across and around Mudumalai.. 7

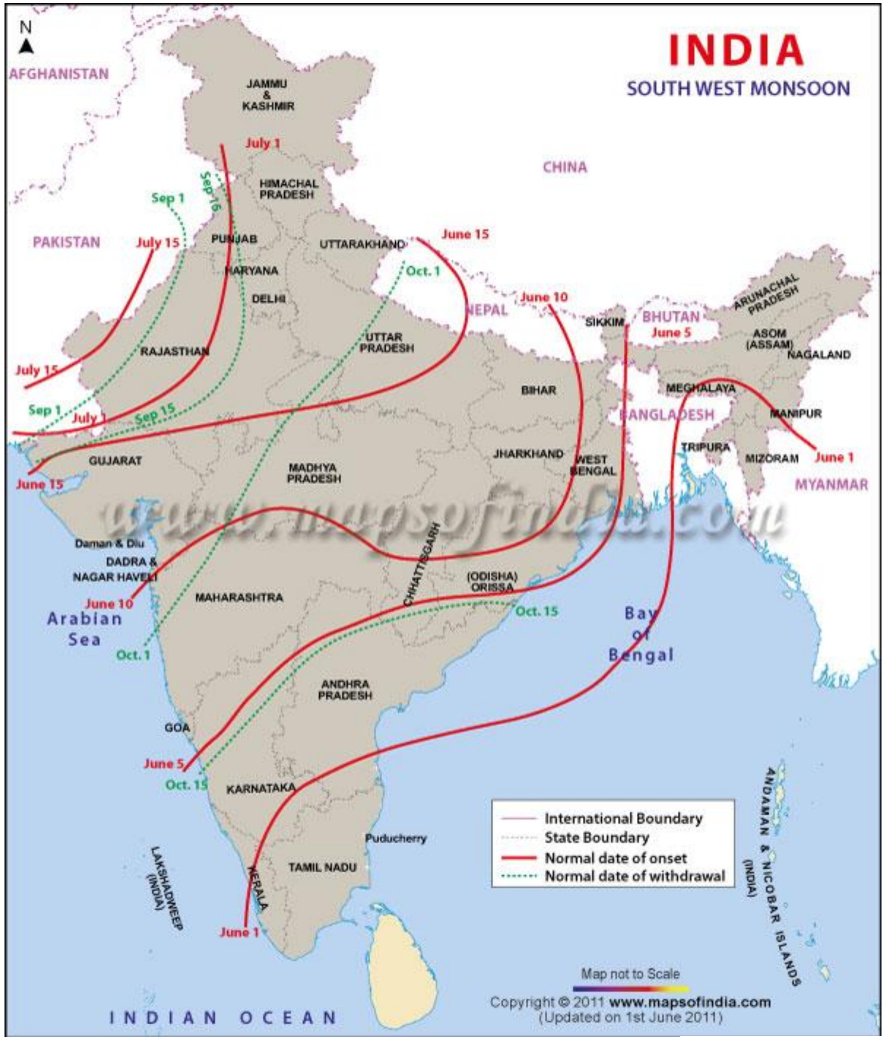
Figure 4: Average monthly rainfall 9

Figure 5: Variation in rainfall across 16 stations for each year from 1990 to 2010... 9

Figure 6: Maps of average annual rainfall and standard error of the prediction 11

Table 1: Annual rainfall for each of the sixteen stations used in imputation of missing values... 5

Figure 1: General movement of the south-west monsoon over the Indian subcontinent



Source: Maps of India. URL:www. mapsofindia.com

Introduction

Rationale for the project

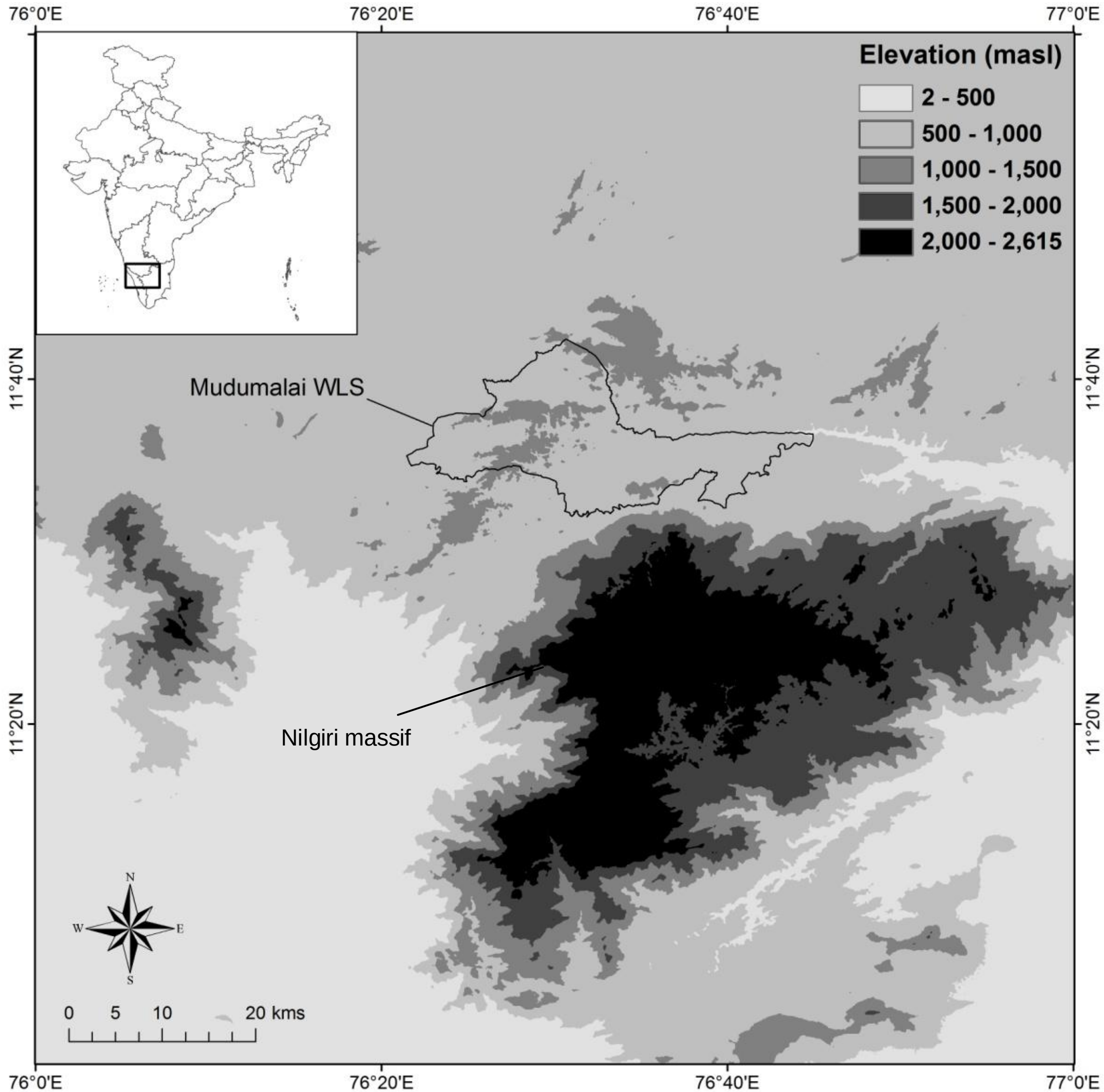
Within an area of 321 km^2, Mudumalai Wildlife Sanctuary (now Tiger Reserve) contains a wide variety of vegetation types that appear to more-or-less closely correspond to the observed spatial variation in precipitation across the landscape. An understanding of the spatial and temporal variation of precipitation is therefore necessary for studies of vegetation structure and dynamics within Mudumalai.

Here, we develop an average annual rainfall map for Mudumalai using a 21-year rainfall dataset from 16 rain gauges in and around Mudumalai. While our data spans less than the typical period of 30 years used by meteorologists worldwide (Arguez *et al*. 2012), it should still be representative of the long-term rainfall regime in the region and can serve as a basis for further refinement based on continuing rainfall monitoring efforts. The map and products derived from it are anticipated to serve as important inputs to models of ecosystem dynamics in the region.

Description of the general movement patterns of the monsoon over the Indian subcontinent

The rainy season typically begins with convectional rains in the months of April and May termed as the “inter-monsoon” or “pre-monsoon” period (Gunnell 1997). This is followed by the summer or south-west monsoon during June-September. The south-west monsoon is a consequence of the heating-up of the landmass in northern India, with winds being drawn in from the India Ocean in the south towards this low-pressure region in the north-east (Gunnel 1997; Attri and Tyagi 2010). The south-west monsoon contributes to approximately 75% of the annual rainfall in India (Attri and Tyagi 2010). However, the spatial distribution of rainfall over the subcontinent is not uniform during this season, as the Western Ghats effectively obstructs the winds from crossing over to the eastern region of the subcontinent. The eastern regions, however, receive a substantial amount of annual rainfall from the north-east monsoon. The winter or north-east monsoon sets in by October and may last until early December, and a majority of the districts of Andhra Pradesh, south interior Karnataka and Tamil Nadu (including where Mudumalai is located) receive approximately 35% of its annual rainfall in this time period (Attri and Tyagi 2010).

Figure 2: Location of the study area (Mudumalai) with respect to surrounding topography, and specifically the Nilgiris, in southern India. The gradient from lighter to darker shades of grey represents 500-m elevation increments (source: ASTER GDEM tile no. N11E076, downloaded from URL: http://gdem.ersdac.jspacesystems.or.jp/outline.jsp).



Study area

Mudumalai Wildlife Sanctuary (c. 321 km^2), also designated as a Tiger Reserve and a part of the larger Nilgiri Elephant Reserve and Nilgiri Biosphere Reserve, is located in the state of Tamil Nadu in the Western Ghats of southern India. The elevation within Mudumalai ranges from 460m to 1220m above mean sea level (asl) with 95% of the area lying between 800m and 1100m asl (Fig. 2). Average maximum temperature during 1990-2010 varied from 25.4 $\pm$ 0.5^{o}C (1 SE) in August to 31.0 $\pm$ 0.3^{o}C in April, and average minimum temperature from 13.9 $\pm$ 0.5^{o}C in January to 18.1 $\pm$ 0.6^{o}C in April (data from the weather station at Kargudi, centrally located in the sanctuary).

Description of local topographical influences on rainfall in the study area

The southern tail of the Western Ghats to the southwest (Gunnell 1997) and the Nilgiri massif to the southeast of Mudumalai (Fig. 2) form effective barriers to the moisture-laden equatorial westerlies (south-west monsoon) forcing orographic rainfall over the western portions of these barriers. The winds, after crossing the plateau and descending the eastern slopes, are warm and dry (Lengerke 1977). This results in a rainfall gradient across Mudumalai with high levels of rainfall in the west and lower levels in the east.

Table 1: Annual rainfall (in mm) for each of the sixteen stations used in imputation of missing values. Highlighted table cells are imputed values (details in *Methods*)

year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Ambalavayal	1680	2301	2319	2128	2692	2288	2181	2151	1728	1578	1749	1435	1198	1540	1965	2562	2227	2303	1848	2012	1871
Wentworth	1994	2338	2960	1864	2902	2337	2543	2854	3105	2582	2295	2147	2257	2149	2582	3993	3716	4925	2997	3265	2675
Biderkadu	1759	2234	2042	1573	1751	2009	1700	1426	1450	1605	1140	927	765	637	1579	1535	1196	1370	1084	1177	1367
Woodbriar	1256	2215	2679	1918	2107	1590	1996	1885	1653	1623	1482	1653	1356	1275	1809	2539	2116	2181	1649	1852	1686
Thorapalli	792	1769	2427	2552	3395	1991	1735	1541	2292	1363	1578	1537	1162	1141	1864	1863	1568	2179	1489	1749	1532
Moolehole	800	1237	1208	1286	1445	1154	1430	1314	990	988	1013	811	679	519	1216	1434	1170	1252	1060	1251	1401
Gamehut	1024	1402	1226	979	1410	1019	1008	1178	1213	1170	1472	1021	976	913	1221	1675	1390	1624	1124	1513	1460
Kargudi	1062	1453	1455	1151	1085	914	1840	1342	1334	1436	967	764	834	1002	1456	1648	1396	1635	1262	1449	1079
Kekkanahalla	872	1201	951	733	850	1179	1038	998	1183	1012	1302	865	884	830	1044	1533	1182	1468	938	1091	1079
Mukhahalli	649	1150	917	752	1121	563	865	1284	830	571	808	643	564	514	760	749	663	679	577	770	811
Bandipur	745	1530	996	1075	1232	790	1129	1109	952	713	1160	751	560	515	1189	1469	834	1131	1023	942	881
Masinagudi	561	934	1092	730	1019	698	841	889	865	820	827	562	576	449	1059	1231	768	902	887	1071	899
Gundalpet	397	899	747	716	1070	585	771	961	723	651	808	683	439	426	903	1013	541	1188	1173	832	812
Odagaramarigudi	399	801	786	452	725	464	564	675	610	651	657	367	376	246	882	1192	837	764	782	889	712
Kundakere	625	1040	857	654	1010	551	774	641	690	710	1171	626	522	456	806	1207	577	677	825	813	987
Terakanambi	320	839	820	457	735	772	842	812	1085	661	808	509	598	513	751	1043	576	868	831	751	949

Methods

Data

Rainfall data are available over a number of rain gauges across the Nilgiri Biosphere Reserve from various sources: coffee and tea estates, Tamil Nadu Electricity Board (TNEB), Karnataka Bureau of Statistics and Economics, and research stations. In this study we used rainfall data for the period 1990 to 2010 from 16 rain gauges in and around Mudumalai (Appendix A). Data from these rain gauges comprised the most complete dataset available for the study period in the proximity of Mudumalai.

Imputation of missing rainfall data

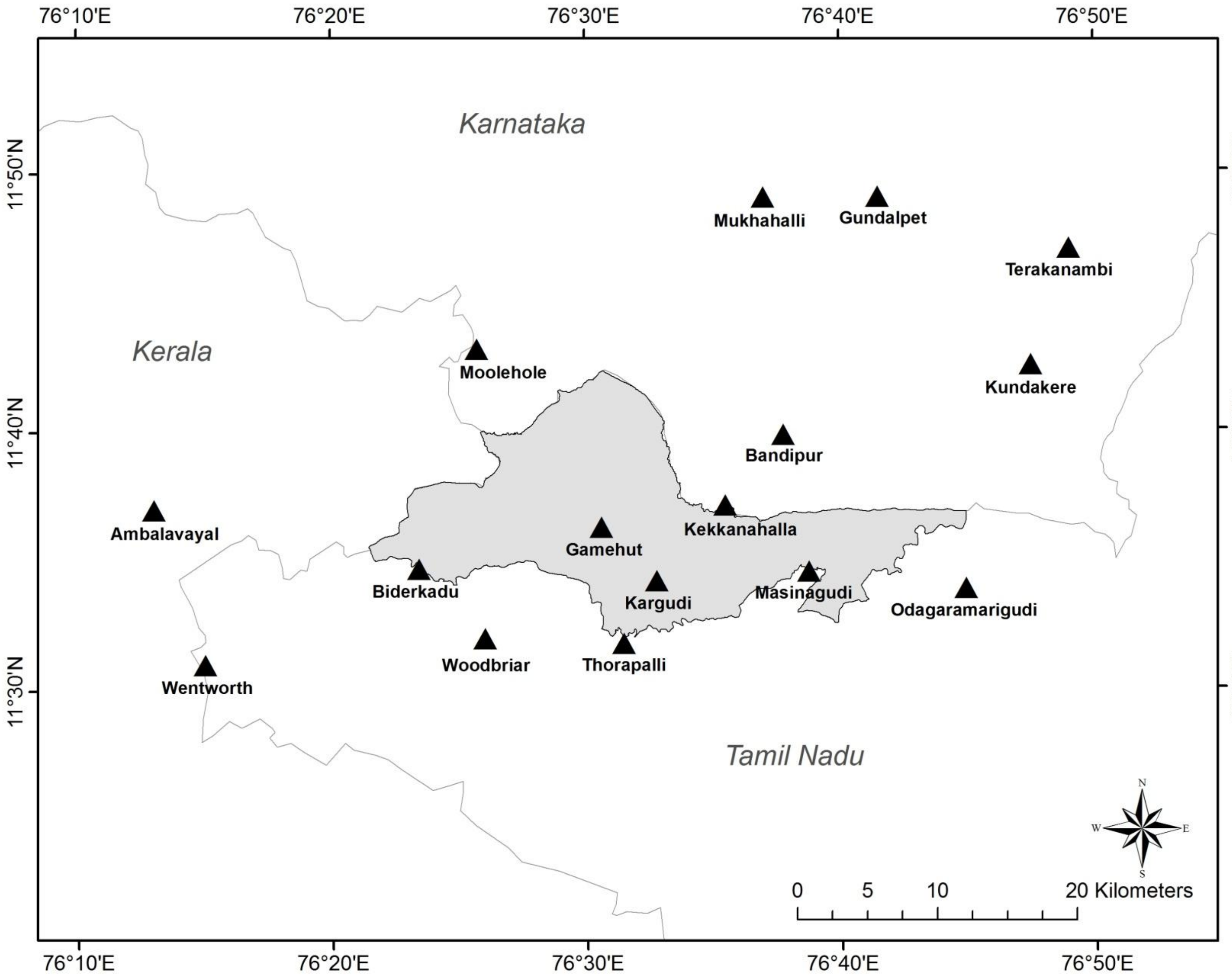
Rainfall data were available for most years at these 16 stations. However, there were a few years with rainfall data missing for some stations. Missing rainfall data were imputed at an annual scale by linearly regressing the focal weather station's annual rainfall against the annual rainfall of one or more neighbouring stations for non-missing years. Sixteen stations were used in the imputation of rainfall data (Table 1). All analyses were performed in *R* version 2.14.0 (R Core Team 2013). Pair-wise correlations between stations were used to determine which stations should be included in the linear regressions. Stations with rainfall highly correlated with the focal station (Pearson's correlation coefficient > 0.5) were selected as predictors. This included up to 5 stations for each regression. Since collinearity of predictors makes regression coefficients unstable (the coefficient of a predictor may change erratically depending on whether or not another is present), we checked for collinearity using variance inflation factors (VIF, Zuur et al. 2007) using the R package "car" (Fox and Weisberg 2011). In this method, each predictor is linearly regressed against all the remaining predictors in turn. The VIF for a given predictor is $1/(1 - R^2)$, where R^2 is the coefficient of determination. VIF values close to 1 indicate non-collinear predictors (variances not inflated) whereas large VIF values indicate collinearity. Stations with VIF greater than 5 were removed from the regression and VIF re-assessed. Predictive accuracy of the regressions was assessed by calculated k-fold cross-validation prediction errors using the R package "boot" (Canty and Ripley 2013). Regressions with the highest adjusted r^2 and smallest k-fold cross-validation prediction error for the set of stations chosen were used for data imputation (Appendix C).

Seasonality

Rainfall was averaged across the 21-year dataset for each month and for each station.

Figure 3: Spatial distribution of the sixteen rain gauges across and around Mudumalai.

Note: All sixteen gauges were used for the imputation of missing rainfall data for a few years in some stations (see 'Methods' for details). Thirteen rain gauges (excluding Mukhahalli, Gundalpet amd Terakanambi) were used in the spatial interpolation of average annual rainfall for Mudumalai (grey shaded portion in the map).



Spatial interpolation

Rainfall data from 13 rain gauges were used in the interpolation of rainfall across Mudumalai (Appendix A and Fig. 3). This set of rainfall stations were selected to ensure an even spatial distribution of stations across Mudumalai for interpolation.

Average annual rainfall was calculated for each station for the period 1990 to 2010. These were interpolated using universal kriging with linear drift (Burrough 1986, Cressie 1991) to generate an average annual rainfall (mm yr^{-1}) raster at a 100 metre resolution using the Geostatistical Analyst extension (Johnston *et al.* 2001) in ArcGIS 9.2 (ESRI, California, USA) for the landscape of Mudumalai. Kriging was used as the interpolation procedure since it uses information on the relationship between all the data points for the final interpolation of values between points. Basically, points that are spatially closer to each other would be more similar in their properties than points that are spatially further apart (also called Tobler's first law of geography). This is visualised through a plot of the semi-variance for pairs of points at different spatial distances (specified as the 'lag'), called a semi-variogram. Semivariance γ is calculated as

$$\gamma(h) = 1/2n\sum_{i=1}^{n}\{z(x_i) - z(x_i + h)\}^2$$

where *n* is the number of data sample pairs separated by a distance *h*, $z(x_i)$ are sample values at location x_i, $z(x_i + h)$ are sample values at a distance *h* from x_i. Weights for interpolation are derived from a model of best-fit to the resulting semi-variogram. In the current case, several semi-variogram models were tested, and the one that generated the least prediction errors through cross-validation was used for the final interpolation. Cross validation is a procedure where-in repeated interpolations are conducted with the selected parameters used for generating the main map, but in each interpolation a point (in this case, station) is removed. The values of the predicted and the actual rainfall observations are then compared.

Universal kriging is a specific kriging procedure that "detrends" the data before analysing semi-variances. In the current case, there was a linear trend in rainfall, with higher rainfall in the west to lower rainfall in the east. Hence, "Universal Kriging with linear drift" accounts for a linear trend in the dataset of sample points.

The other advantage of kriging is that it will compute the errors for the predictions made through the chosen model. Standard errors of the prediction from kriging were calculated based on algorithms available in ArcGIS 9.2 (Johnston *et al.*, 2001).

Figure 4: Average monthly rainfall (mm month^{-1}) for the 16 stations. Red lines indicate those stations located towards the west of the sanctuary, blue lines are centrally located stations and yellow lines are those stations towards the east.

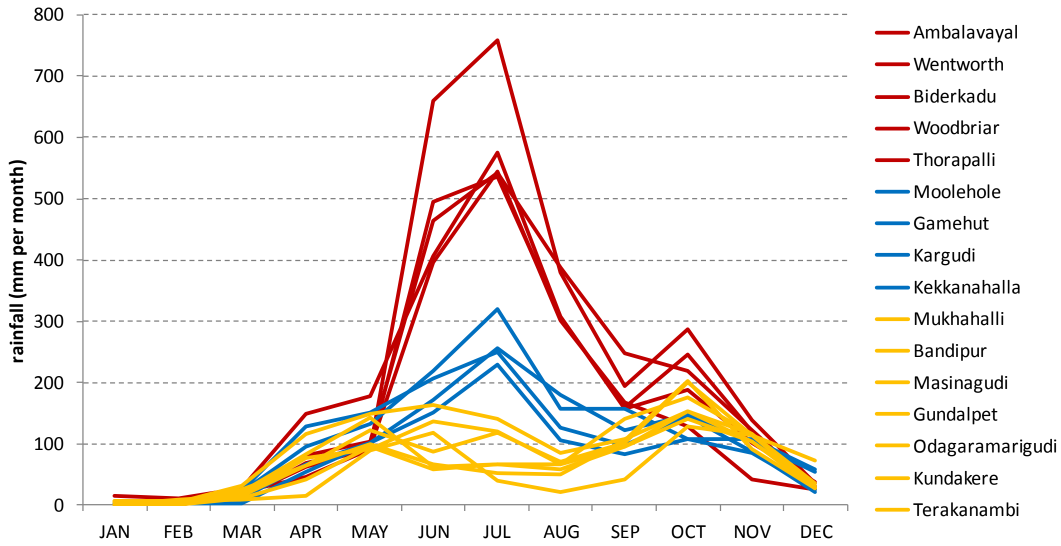
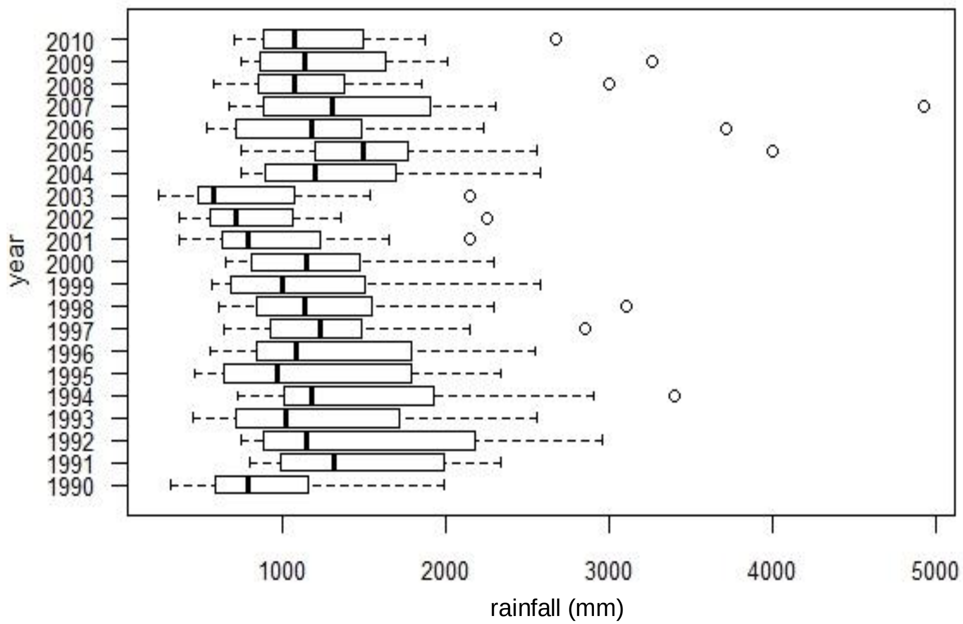


Figure 5: Variation in rainfall across 16 stations for each year from 1990 to 2010. Each box-and-whisker plot can be visualised as a representation of rainfall across the landscape of Mudumalai from east (left of the graph) to the west (right of the graph).

The vertical bold line is the median rainfall for each year. The boxes are the first (left) and third (right) quartiles showing the range of the middle 50% of the data. The horizontal dashed lines represent rainfall that is 1.5 times the inter-quartile range (difference between the first and third quartiles) of the data. Hollow circles are 'outliers' that are greater than 1.5 times the inter-quartile range above the third quartile (Crawley 2007, pg. 155).



Results

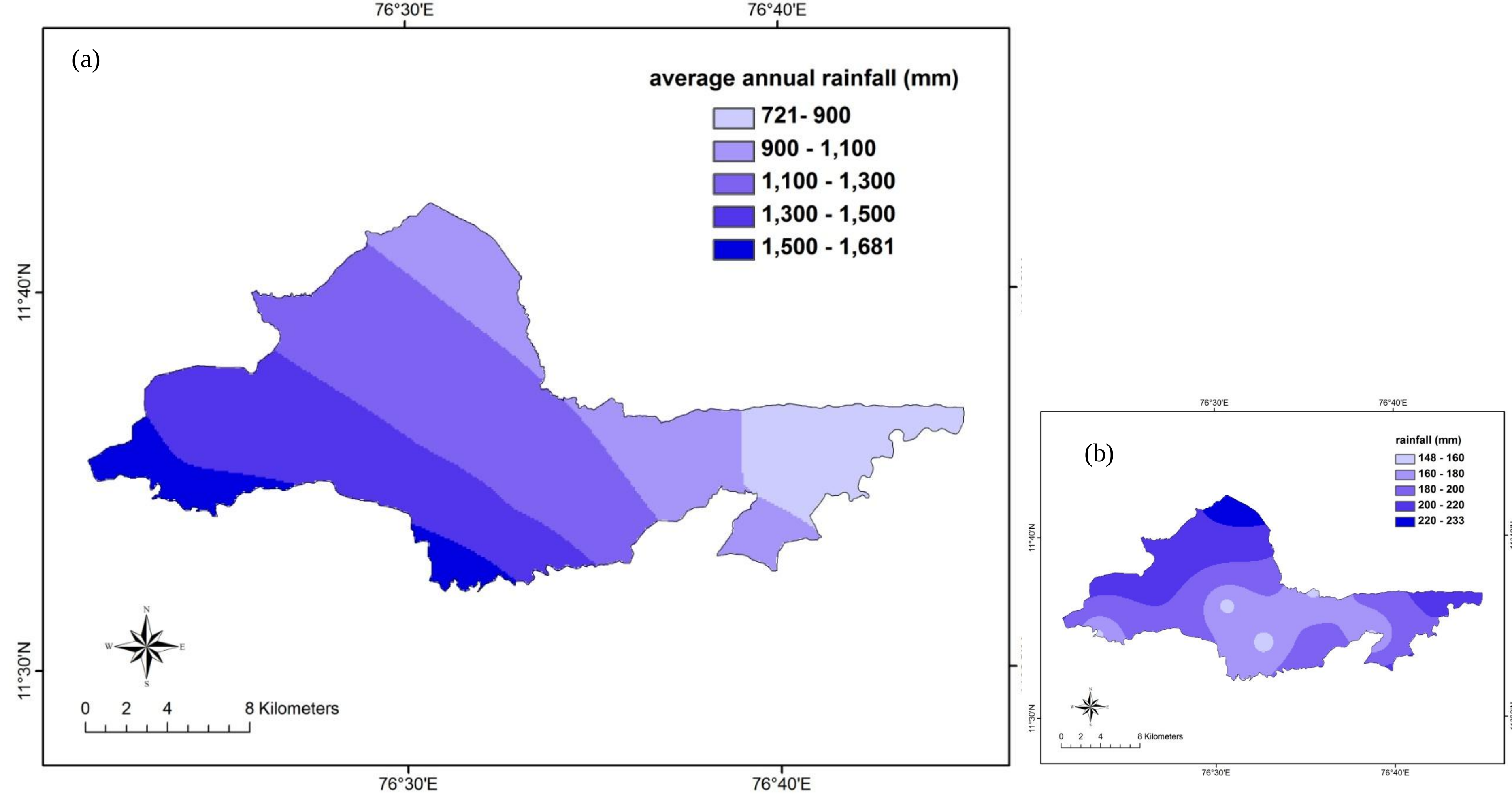
Seasonality

Stations could be distinctly categorized into three groups based on the relative distribution of rainfall across the year (Fig. 4). These were stations that were roughly located in the western, central and eastern portions of the landscape. In general, two peaks in rainfall were observed, one corresponding to the pre-monsoon or south-west monsoon and the other to the north-east monsoon. Eastern stations received the bulk of their rainfall in October with a second, relatively smaller pre-monsoon rainfall peak during May. Stations in the central and western portions of the landscape on the other hand received the bulk of their precipitation in July, with a smaller relative contribution during October, and with the July peak being much more distinct in the western stations. The landscape as a whole consistently experienced a continuous four-month dry season from December of one year to March of the following year during which rainfall was less than 100mm month^{-1} on average across all 16 stations considered in this study. Seven stations located towards the east also experienced a dry month in August (rainfall ranged between 21.4 and 85.7 mm across seven stations). Stations towards the extreme eastern limit in this study (Odagaramarigudi, Terakanambi and Kundakere) rarely received rainfall greater than 100 mm month^{-1} in the months of June and July as well (the exception being Odagaramarigudi which received 118.2 mm of average rainfall in June). Hence, the extreme eastern portions of the sanctuary and areas bordering it can be considered to have a second dry season from June to August.

Spatial variations in rainfall across years

Variation in annual rainfall across 16 stations is plotted as a box and whisker plot (Fig. 5). Considerable spatial rainfall variability was observed across years. Some years, such as 1990, 2001, 2002 and 2003, were drought years where the landscape as a whole received very low levels of rainfall. Other years, such as 2005, had above-average rainfall across the landscape. The 2001-2003 drought at Mudumalai is fairly consistent with the drought experienced by the entire country during this period; India experienced a 6% deficit in rainfall in 2001 (deficits are from the average calculated over 1990-2003), and a larger deficit of 14% in 2002, although rainfall was 3% above average for the country in 2003 (data from Guhathakurta and Rajeevan 2006).

Figure 6: Maps of (a) average annual rainfall (yr^{-1}) and (b) standard error of the prediction



Annual average rainfall

Average annual rainfall varied from 721mm yr^{-1} in the east to 1681mm yr^{-1} in the west (Fig 6a).

Standard error of the prediction using kriging, and cross-validation

Prediction standard errors ranged from 148mm yr^{-1} to 233mm yr^{-1} within Mudumalai (Fig 6b). The highest prediction errors were around the north of the study area, the least towards the centre. This could be because of the scarcity of rain gauges to interpolate from in the northern part of the sanctuary compared to a well-distributed set of rain gauges in the central part.

Cross-validation results were as follows:

Mean: -7.906 mm yr^{-1}
Root-Mean-Square: 313 mm yr^{-1}
Average Standard Error: 258.7 mm yr^{-1}
Mean Standardized Error: 0.036
Root-Mean-Square Standardized: 1.13

The mean error (that is, the average difference between the actual rainfall at a station and what is predicted) was approximately -8mm yr^{-1}, with a range of -576mm yr^{-1} to 389mm yr^{-1}. The mean standardized error (which is mean of the prediction errors divided by their respective prediction standard errors) was 0.036, with a range of -1.9 to 1.7 standard errors.

References

Arguez, A., Durre, I., Applequist, S., Vose, R. S., Squires, M. F., Yin, X., Heim, R. R. Jr. and Owen. T. W. 2012. NOAA s 1981–2010 U.S. Climate Normals: An Overview. *Bulletin of the American Meteorological Society*.

Attri, S.D. and Tyagi, A. 2010. Climate profile of India. Met Monograph No. Environment Meteorology-01/2010, URL: http://www.imd.gov.in/

Burrough, P.A. 1986. Principles of geographical information systems for land resources assessment, New York, Oxford University Press.

Canty, A. and Ripley, B. 2013. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-9.

Crawley, M. J. 2007. The R book. John Wiley and Sons Ltd., UK.

Cressie, N.A.C. 1991. Statistics for spatial data. John Wiley and Sons, Inc.

Fox, J. and Weisberg, S. 2011. An R Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Guhathakurta, P. and Rajeevan, M. 2006. Trends in the rainfall pattern over India. National Climate Centre Research Report No. 2/2006. India Meteorological Department, Pune, India. URL: http://www.imdpune.gov.in/ncc_rept/RESEARCH%20REPORT%202.pdf

Gunnell Y. 1997. Relief and climate in south Asia: the influence of the Western Ghats on the current climate pattern of peninsular India. *International Journal of Climatology* 17: 1169-1182.

Johnston, K., Ver Hoef, J.M., Krivoruchko, K. and Lucas, N., 2001, Using ArcGIS Geostatistical Analyst, ESRI, USA

Lengerke, H. J. von. 1977. The Nilgiris: weather and climate of a mountain area in south India. Weisbaden: Steiner

R Core Team .2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Zuur, A. F., Ieno, E. N. and Smith, G. M. 2007. Analysing Ecological Data. Springer Science + Business Media, LLC.

Appendices

Appendix A: List of rain gauges used, their location (latitude and longitude in decimal degrees), elevation (extracted from ASTER GDEM, Fig. 2) and the source of data

sr.no.	Station	Latitude	Longitude	Elevation (m)	Source
1	Ambalavayal	11.616670	76.216670	894	Kerala Agricultural Univeristy Field Research Centre
2	Wentworth	11.516670	76.250000	822	Wentworth Estate
3	Biderkadu	11.578139	76.390306	945	Biderkadu Estate
4	Woodbriar	11.533330	76.433330	974	Woodbriar Estate
5	Thorapalli	11.529694	76.524250	964	Thorapalli Estate
6	Moolehole	11.719671	76.428714	839	Karnataka Bureau of Statistics and Economics; Indo-French Cell, Indian Insitute of Science
7	Gamehut	11.604556	76.509972	989	Tamil Nadu Electricity Board
8	Kargudi	11.570111	76.546056	959	Centre for Ecological Sciences, Indian Insitute of Science
9	Kekkanahalla	11.618546	76.590841	894	Tamil Nadu Electricity Board
10	Mukhahalli	11.816670	76.616670	852	Karnataka Bureau of Statistics and Economics
11	Bandipur	11.663889	76.629167	967	Karnataka Bureau of Statistics and Economics
12	Masinagudi	11.575600	76.645601	948	Centre for Ecological Sciences, Indian Insitute of Science
13	Gundalpet	11.816667	76.691667	772	Karnataka Bureau of Statistics and Economics
14	Odagaramarigudi	11.564593	76.748474	799	Tamil Nadu Electricity Board
15	Kundakere	11.708330	76.791670	887	Karnataka Bureau of Statistics and Economics
16	Terakanambi	11.783333	76.816667	928	Karnataka Bureau of Statistics and Economics

Appendix B: R-code used for the imputation of annual rainfall from linear regressions

```
# ---------------------------------------------------------------------------
# Missing precipitation data estimation.
# Author: Sandeep Pulla
# Created: 28 November 2011
# Last updated: 22 August 2012.
# Instructions: (1) Install packages 'boot' and 'car' from CRAN. (2) Set the
# 'data_folder' variable below to the folder that contains the files
# 'mudumalai_precipitation.csv' and 'mudumalai_precipitation_models.csv'.
# (3) Source this file.
# ---------------------------------------------------------------------------

library(boot); library(car)

# Operator for concatenating objects as text
"%+%" = function(obj_1, obj_2) {
  paste(obj_1, obj_2, sep = "")
}

# cat() that appends a newline
catn = function(..., file = "", sep = " ", fill = FALSE, labels = NULL,
  append = FALSE){
  cat(..., "\n", file = file, sep = sep, fill = fill, labels = labels,
    append = append)
}

# Returns a character vector that can be used to print a horizontal 'line' made of a
# single repeating character.
# ch - Character to use. Defaults to '-'.
```

```
# length - Number of times to repeat the character, i.e., the number of text
#   columns the horizontal line should span.
# returns - Character vector representing horizontal line.
hr = function(ch = '-', length = 79) {
  "\n" %+% paste(rep(ch, length), sep = "", collapse = "") %+% "\n"
}

# Appends an item to a container (list, vector). Doesn't squash items like the
# base::append() does.
# container - List or vector.
# item - Item to be added.
# name - Optional item name (should be unique!).
# returns - Modified 'container'.
push_back = function(container, item, name = NULL) {
  name = if(is.null(name)) { length(container) + 1 } else { as.character(name) }
  container[[name]] = item
  return(container)
}

# Displays relevant information about a linear model (an "lm"-derived object).
# model - An "lm"-derived object.
# digits - Number of significant digits to print after the decimal point (applies
#   only to numbers related to response (e.g. sd and deviance). See print().
# p - Whether or not to print p-values and significance.
# units - Optional units to display alongside values like standard deviations.
disp = function(model, digits = 2, p = T, units = "") {
  c = class(model)
  n = length(model$residuals)
  k = length(model$coefficients)
```

```
catn("n = " %+% n %+% ", k = " %+% k)
if(n < k) {
  warning("number of observations (n) < number of parameters (k)")
  sigma_hat_squared = NA
} else {
  sigma_hat_squared = sum(model$residuals ^ 2)/(n - k)
}
sigma_squared = var(model$residuals + model$fitted.values)
catn("sd = " %+% round(sqrt(sigma_squared), digits = digits) %+% units %+%
  ", resid sd = " %+% round(sqrt(sigma_hat_squared), digits = digits) %+% units %+%
  ", r-squared = " %+%
  round(1 - (sigma_hat_squared/sigma_squared), digits = 2) %+%
  " adj / " %+%
  round(1 - ((sigma_hat_squared * (n - k))/(sigma_squared * (n - 1))),
    digits = 2) %+% " multiple")
if("glm" %in% c) {
  catn("deviance: " %+%
    "null " %+% format(round(model$null.deviance, digits = digits),
    big.mark = ",") %+%
    " resid " %+% format(round(model$deviance, digits = digits),
    big.mark = ",") %+%
    " diff " %+% format(round(model$null.deviance - model$deviance, digits = digits),
    big.mark = ","))
}
c = coef(model)
s = summary(model)
p = s$coefficients[, "Pr(>|t|)"]
signif = significance(p)
d = data.frame(est = round(c, digits = digits), p = formatC(p, format = "g"),
```

```
    signif = signif)
  print(d)
}

# Returns significance symbols for p-values.
# p - Vector of p values.
# levels - Significance levels. One asterisk is added per level.
# returns - Vector of significance symbols for values in 'p'.
significance = function(p, levels = c(0.05, 0.01, 0.001)) {
  return(sapply(p, function(x) {
    if(is.na(x)) {
      s = NA
    } else {
      s = ""
      for(l in levels) { if(x < l) s = s %+% "*" }
    }
    return(s)
  }))
}

# Estimates missing annual precipitation for a focal weather station by
# linearly regressing that station's annual precipitation against the annual precipitation
# of one or more neighbouring stations for non-missing years.
# K-fold cross validation is performed to assess predictive ability of the regression.
# Variance Inflation Factors (VIF) are used to assess collinearity of predictors.
# missing_stn - Focal weather station name.
# neighbor_stns - Neighbor weather station names.
# data - Data frame containing annual precipitation data.
# vif_warning - VIF value above which a warning should be printed.
```

```
# returns - Imputed annual precipitation data.
est_precip = function(missing_stn, neighbor_stns, data, vif_warning = 5) {
  stns = c(missing_stn, neighbor_stns)
  complete_cases = subset(data, subset = complete.cases(data[, stns]), select = stns)
  model_str = missing_stn %+% "~" %+% paste(neighbor_stns, collapse = "+")
  f = as.formula(model_str)
  m = glm(f, data = complete_cases)
  catn(hr())
  catn(model_str)
  disp(m, units = "mm")
  par(mfrow = c(2, 3))
  for(n in neighbor_stns)
  {
    plot(complete_cases[, missing_stn] ~ complete_cases[, n],
      xlab = n, ylab = missing_stn, asp = 1)
    mini_model = lm(complete_cases[, missing_stn] ~ complete_cases[, n])
    x = range(complete_cases[, n])
    lines(x, x * coef(mini_model)[2] + coef(mini_model)[1], lty = "dotted")
  }
  par(mfrow = c(2, 4))
  plot(m, ask = F)
  cverr = sqrt(cv.glm(data = complete_cases, glmfit = m)$delta[2])
  catn("K-fold cross-validation prediction error: " %+%
    format(round(cverr), big.mark = ",") %+% "mm")
  if(length(coef(m)) > 2) {
    catn("Variance inflaction factors:")
    v = vif(m)
    sapply(names(v), function(x) {
      catn(x %+% " " %+% round(v[x], 2) %+% (if(v[x] > vif_warning) "[WARNING]" else "")) })
```

```
  }
  missing = subset(data, subset = is.na(data[, missing_stn]), select = stns)
  predicted = predict(m, newdata = missing)
  return(predicted)
}


data_folder = ""
focal_yrs = 1990:2010
catn("\nFocal years: " %+% paste(focal_yrs, collapse = ", "))
precip = read.csv(file.path(data_folder, "mudumalai_precipitation.csv"),
  header = T, row.names = 1, check.names = F)
# Drop non-focal years and non-precipitation columns
d = data.frame(subset(t(precip), subset = colnames(precip) %in%
  as.character(focal_yrs)))

pdf(file.path(data_folder, "mudumalai_precipitation.pdf"), width = 11.5, height = 8)

pairs(d, pch = 20, cex = 0.6, gap = 0.5)
plot(range(rownames(d)), range(d, na.rm = T), xlab = "year",
  ylab = "annual precipitation (mm)", main = "Annual precipitation", type = "n")
cols = rainbow(ncol(d))
ltys = rep(1:2, ncol(d)/2)
for(c in 1:ncol(d)) {
  lines(as.numeric(rownames(d)), d[, c], col = cols[c], lty = ltys[c])
}
legend(x = "topright", legend = colnames(d),
  col = cols, lty = ltys, cex = 0.8, text.col = grey(0.3))
par(mfrow = c(3, 3))
```

```
for(c in colnames(d)) {
  hist(d[, c], main = c %+% " annual precipitation",
    xlab = "annual precipitation (mm)", ylab = "frequency")
}

dfilled = d

# Uncomment to write out template file for models
# write.csv(matrix(F, nrow = ncol(d), ncol = ncol(d),
#   dimnames = list(sort(names(d)), sort(names(d)))),
#   file = file.path(data_folder, "mudumalai_precipitation_models.csv"))

# Notes:
# - bidarkad and thorapalli often show the "wrong" sign
# - Can't use gamehut/moolehole/odagara for kekkanahalla because some missing years
#   are common
# - Can't use kekkanahalla for gamehut because some missing years are common
# - Can't use woodbriar for ambalavayal because missing years are common
m = as.matrix(
  read.csv(file = file.path(data_folder, "mudumalai_precipitation_models.csv"),
  row.names = 1))
models = list()
add_full_models = T # change as needed
complete_data_stns = sort(colnames(d)[! is.na(colSums(d))])
for(r in 1:nrow(m)) {
  if(any(m[r, ])) {
    models = push_back(models, list(missing_stn = rownames(m)[r],
      neighbor_stns = colnames(m)[m[r, ]]))
    if(add_full_models) {
```

```
      # "Full" models -- these use *every* other station that has complete data as a
      # a predictor, which can be useful for comparison
      models = push_back(models, list(missing_stn = rownames(m)[r],
        neighbor_stns = complete_data_stns))
    }
  }
}

for(model in models) {
  pred = est_precip(missing_stn = model$missing_stn, model$neighbor_stns, data = d)
  catn("\nPredicted for " %+% model$missing_stn)
  print(pred)
  dfilled[names(pred), model$missing_stn] = pred

  non_missing_yrs = rownames(d)[! is.na(d[, model$missing_stn])]
  mean1 = mean(d[non_missing_yrs, model$missing_stn])

  all_yrs = rownames(dfilled)[! is.na(dfilled[, model$missing_stn])]
  mean2 = mean(dfilled[all_yrs, model$missing_stn])

  catn("\n" %+% model$missing_stn %+% " original mean (" %+%
    paste(non_missing_yrs, collapse = ", ") %+% "):\n" %+% round(mean1) %+% "mm")
  catn(model$missing_stn %+% " mean with missing years estimated (" %+%
    paste(all_yrs, collapse = ", ") %+% "):\n" %+% round(mean2) %+% "mm")
  catn("Difference in means: " %+% round(mean2 - mean1) %+% "mm")
}
write.csv(dfilled, file =
  file.path(data_folder, "mudumalai_precipitation_filled.csv"))
eat = dev.off()
```

Appendix C: Linear regression equations for each station with r^2 of the regression

ambalavayal~wentworth+woodbriar n = 13, k = 3 sd = 426.23mm, resid sd = 324.54mm, r-squared = 0.42 adj / 0.52 multiple
bidarkad~kundakere+mukhahalli+woodbriar n = 19, k = 4 sd = 426.05mm, resid sd = 344.95mm, r-squared = 0.34 adj / 0.45 multiple
gamehut~kundakere+wentworth n = 17, k = 3 sd = 222.89mm, resid sd = 102.23mm, r-squared = 0.79 adj / 0.82 multiple
kekkanahalla~bandipur+kundakere+terakanambi+wentworth n = 17, k = 5 sd = 211.08mm, resid sd = 158.67mm, r-squared = 0.43 adj / 0.58 multiple
masinagudi~bandipur+kargudi+odagaramarigudi+thorapalli+woodbriar n = 14, k = 6 sd = 142.88mm, resid sd = 76.86mm, r-squared = 0.71 adj / 0.82 multiple
moolehole~bandipur+bidarkad+kargudi+thorapalli+woodbriar n = 15, k = 6 sd = 248.86mm, resid sd = 150.21mm, r-squared = 0.64 adj / 0.77 multiple
mukhahalli~bandipur n = 20, k = 2 sd = 209.28mm, resid sd = 163.46mm, r-squared = 0.39 adj / 0.42 multiple
odagaramarigudi~bandipur+masinagudi n = 14, k = 3 sd = 183.13mm, resid sd = 115.94mm, r-squared = 0.6 adj / 0.66 multiple
thorapalli~bidarkad+gundalpet+mukhahalli+woodbriar n = 18, k = 5 sd = 555.49mm, resid sd = 491.2mm, r-squared = 0.22 adj / 0.4 multiple

Appendix D: Details of the interpolation procedure used for generating the annual average rainfall map

Method: Universal Kriging

Trend type: First order polynomial (Linear)

Trend removal: Global Polynomial Interpolation

Search neighbourhood: Standard
- Neighbours to include: 13
- Include at least: 10
- Sector type: Full
- Angle: 0
- Minor and major semiaxis: 61430

Semi-variogram parameters:
- Number of lags: 12
- Lag size: 5220m
- Nugget: 13330

Semi-variogram model: Exponential
- Range: 61430
- Anistropy: None

Partial sill: 77503