Nonlin. Processes Geophys. Discuss., 1, 193–233, 2014
www.nonlin-processes-geophys-discuss.net/1/193/2014/
doi:10.5194/npgd-1-193-2014
© Author(s) 2014. CC Attribution 3.0 License.

Nonlinear Processes in Geophysics Discussions
Open Access

This discussion paper is/has been under review for the journal Nonlinear Processes in Geophysics (NPG). Please refer to the corresponding final paper in NPG if available.

Logit-normal mixed model for Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

University of Minnesota, School of Statistics, Minneapolis, Minnesota, USA

Received: 8 February 2014 – Accepted: 21 February 2014 – Published: 13 March 2014

Correspondence to: L. R. Dietz (diet0146@umn.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union & American Geophysical Union.

Abstract

Describing the nature and variability of Indian monsoon rainfall extremes is a topic of much debate in the current literature. We suggest the use of a *generalized linear mixed model* (GLMM), specifically, the logit-normal mixed model, to describe the underlying structure of this complex climatic event. Several GLMM algorithms are described and simulations are performed to vet these algorithms before applying them to the Indian precipitation data procured from the National Climatic Data Center. The logit-normal model was applied with fixed covariates of latitude, longitude, elevation, daily minimum and maximum temperatures with a random intercept by weather station. In general, the estimation methods concurred in their suggestion of a relationship between the El Niño Southern Oscillation (ENSO) and extreme rainfall variability estimates. This work provides a valuable starting point for extending GLMM to incorporate the intricate dependencies in extreme climate events.

1 Introduction

Indian rainfall extremes have been studied by many in the past few decades with varying published results. Goswami et al. (2006) used daily central Indian rainfall in a linear regression framework and noted significant rising trends in frequency and magnitude of extreme ($\geq 150\,\mathrm{mm\,day^{-1}}$) rain events. Ghosh et al. (2009) conducted a similar study at a finer spatial scale, using $1^\circ \times 1^\circ$ data vs. the 10° latitude $\times 12^\circ$ longitude (central India) used in Goswami et al. (2006). Their findings indicated more of mixture of increases and decreases of extreme rainfall events dependent on location. Ghosh et al. (2012) expanded on this result and added that there has been increasing spatial variability in observed Indian rainfall extremes. Methodology in this research included the use of extreme value theory (EVT) as an attempt to account for the modeling of extremes. Turner and Annamalai (2012) indicated even complex climate models have high uncertainty

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

for projections of monsoon rainfall due to linked variability on different temporal and spatial scales.

Regression techniques and EVT failed to adequately capture the broad statistical properties of the Indian summer monsoon precipitation. Several drivers of monsoons have been proposed in the literature including Himalayan/Eurasian snow extent (Kumar et al., 1999), Pacific trade winds (Li and Yanai, 1996), atmospheric CO_2 concentration (Prell and Kutzback, 1992), and most commonly sea surface temperatures (SST) attributed to El Niño-Southern Oscillation (ENSO) (Turner and Annamalai, 2012; Kumar et al., 1999; Li and Yanai, 1996; Prell and Kutzback, 1992). However, none have been conclusively shown to drive the monsoon rainfall which suggests an intricate relationship between some or all of these factors.

Based on the above context, we propose using a *generalized linear mixed model* (GLMM) as a potential broader framework for analysis of Indian monsoon precipitation data. A GLMM is a broader framework compared to the standard (linear, log-linear, logistic, or other) regression in that there are *random effects* involved. This implies part of the signal is random, and changes from one set of circumstances to another. In the current context, a GLMM may be a suitable model for capturing local, instantaneous variability. Such local variability may arise from cloud and other physical microproperties. In case there is no such local variability, an appropriate variance component in the GLMM would be zero, thus recovering any true underlying "fixed-effects" regression model. GLMMs are commonly used in epidemiological and other biostatistical areas when repeated measures are available for each "subject" (stations, states, etc). A GLMM essentially allows for some underlying forces to drive the observable data in a particular hierarchy. This could be key in climate applications as we may not sufficiently be able to attribute drivers of extreme rainfall to observable data. One issue in GLMMs that has been studied since their introduction in Stiratelli et al. (1984) is consistent estimation, i.e. with enough observations the estimates essentially represent the truth.

NPGD
1, 193–233, 2014

CC BY

This paper provides a glance on how to extend GLMMs to handle climate applications, where extremes and complicated temporal and spatial dependency patterns are important. We present a case study on Indian monsoon summer precipitation. One of our major findings is that the random effect variance component is non-negligible in the analysis of *extreme* precipitation due to Indian monsoons. It may be significant for light or moderate precipitation as well, although further studies are needed to verify this claim. The estimates of variability in the models for extreme rainfall indicated a meaningful correspondence to El Niño events over the time frame examined, which also deserves further analysis.

Section 2 provides a short background on GLMMs and in particular, elucidates the logit-normal model. Section 3 provides a background on the theory and application of several estimation methods for GLMMs. Section 4 provides the results of several simulations using these existing methods. Section 5 implements the application of these methods in their current state on precipitation data from India. Finally, Sect. 6 presents conclusions and future work in this area.

2 GLMM background

Assume we have data from an exponential family which are independent conditional on a vector of unobservable random effects where "subjects" IDs run over $i = 1, \ldots, m$ and the number of (possibly correlated) measurements per subject is from $j = 1, \ldots, n_j$. Following the notation of McCulloh and Searle (2010), a GLMM can be written as:

$$\mathbf{Y}_i|\boldsymbol{U} = \boldsymbol{u} \overset{\text{ind.}}{\sim} f_{\eta_i}(\boldsymbol{y}_i|\boldsymbol{u}) = \exp\left\{\frac{\boldsymbol{y}_i\eta_i - b(\eta_i)}{\tau^2} - c(\boldsymbol{y}_i, \tau)\right\}$$
$$\eta_i = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{u}$$
$$\boldsymbol{U} \sim f(\boldsymbol{u})$$

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

CC BY

Here $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\mathrm{T}$, $\boldsymbol{U}$ is the vector of random effects of suitable length, $\boldsymbol{x}_{ij}$ is a vector of covariates for the fixed effects of the ith subject at the jth time, and $\boldsymbol{z}_i$ is a vector of covariates for the random effects of ith subject.

A highly applicable version of this model is the logit-normal GLMM. A simple hierarchical form of the model is:

$$Y_{ij}|\boldsymbol{u} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = \eta_i = \boldsymbol{x}_{ij}^\mathrm{T}\boldsymbol{\beta} + u_i$$

$$U_i \overset{\text{ind.}}{\sim} N(0, \sigma^2)$$

This implies the density for a single observation Y_{ij} is

$$f_{\boldsymbol{\beta},\sigma}(y_{ij}|\boldsymbol{u}) = \exp\left\{y_{ij}\left(\boldsymbol{x}_{ij}^\mathrm{T}\boldsymbol{\beta} + u_i\right) - \log\left(1 + \exp\left(\boldsymbol{x}_{ij}^\mathrm{T}\boldsymbol{\beta} + u_i\right)\right)\right\}$$

Thus, the density for the ith subject is

$$f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}_i|\boldsymbol{u}) = \exp\left\{\sum_{j=1}^{n_i} y_{ij}\left(\boldsymbol{x}_{ij}^\mathrm{T}\boldsymbol{\beta} + u_i\right) - \sum_{j=1}^{n_i}\log\left(1 + \exp\left(\boldsymbol{x}_{ij}^\mathrm{T}\boldsymbol{\beta} + u_i\right)\right)\right\}$$

The assumption of conditional independence among subjects implies the joint density of the vectors $\boldsymbol{y}_i$ is

$$f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}|\boldsymbol{u}) = \prod_{i=1}^{m} f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}_i|\boldsymbol{u}) \tag{1}$$

The assumption of independent standard normal random effects implies the joint density of $(\mathbf{Y}, \boldsymbol{U})$ is

$$f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}, \boldsymbol{u}) = f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}|\boldsymbol{u})f(\boldsymbol{u}) = f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}|\boldsymbol{u})\prod_{i=1}^{m} f(\boldsymbol{u}_i) \tag{2}$$

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

CC BY

However, since the random effects are unobserved, to utilize the observed data likelihood, one must find the marginal distribution with respect to the observed data **Y** only.

$$f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y}) = \int f_{\boldsymbol{\beta},\sigma}(\boldsymbol{y},\boldsymbol{u})\mathrm{d}\boldsymbol{u} \quad (3)$$

This integral is almost never analytically tractable, thus, maximum likelihood estimation is very difficult, if not impossible. Many methods for inference have been proposed. Variants of some of the most popular methods are examined later in this paper.

3 Methods for estimating in GLMM

3.1 Penalized quasi-likelihood with Laplace approximation

Breslow and Clayton (1993) were the first to introduce a feasible method for computation in GLMMs. *Penalized quasi-likelihood* (PQL) approximates the high-dimensional integration using the well-known Laplace approximation (Tierney and Kadane, 1986) and the approximated likelihood function has that of a Gaussian distribution.

There are a few implementations of PQL which were used within the simulations. The first function, `glmer{lme4}`, is from a very popular mixed modeling package in R. According to Bates (2010), they define the conditional mode, $\tilde{u}$ as the value that maximizes the unscaled conditional density given in Eq. (2). From there, they determine $\tilde{u}$ as the solution to a penalized nonlinear least squares (PNLS) problem solved by adapting iterative techniques, such as the Gauss–Newton method. Once the algorithm has converged, the method calculates the Laplace approximation to the deviance where the Cholesky factor and penalized residual sum of squares are both evaluated at the conditional mode, $\tilde{u}$.

The second function, `glmmPQL{MASS}` relies on multiple calls to `lme{nlme}` which is the previous version of the `lme4` package. PQL is reasonably accurate when the data are approximately normal and can be very fast depending on the algorithm used

NPGD
1, 193–233, 2014
Indian Monsoon rainfall extremes
L. R. Dietz and S. Chatterjee

for implementation. However, Lin and Breslow (1996) and others have criticized this method for it's bias in highly non-normal data.

3.2 Method of simulated moments

Jiang (1998) describes methodology known as the *method of simulated moments* (*MSIM*). The method first derives a set of sufficient statistics. Estimating equations are then obtained by equating sample moments of sufficient statistics to their expectations. An example of the system of equations to solve in the simple logit-normal intercept only case ($\boldsymbol{x}_i^T \boldsymbol{\beta} = \mu$) are

$$\frac{1}{m}\sum_{i=1}^{m} y_{i.} \overset{\text{set}}{=} E\left(\frac{1}{m}\sum_{i=1}^{m} y_{i.}\right) = E(Y_{1.})$$

$$\frac{1}{m}\sum_{i=1}^{m} y_{i.}^2 \overset{\text{set}}{=} E\left(\frac{1}{m}\sum_{i=1}^{m} y_{i.}^2\right) = E(Y_{1.}^2)$$

The expectations on the right hand side are generated by use of Monte Carlo simulations and the whole system can then be solved by the Newton–Raphson method or some asymptotically equivalent algorithm. We implemented this method in a newly created R package.

As shown in Jiang (1998), this method is consistent and computationally much less intensive than a Markov Chain Monte Carlo (MCMC) method. However, limitations include slow convergence and less than ideal small sample properties.

3.3 Data cloning

GLMM estimates can be produced in a traditional Bayesian framework; one must choose priors for the parameters of interest and calculate the posterior distribution by multiplying the prior densities by the likelihood in Eq. (3). One may then use MCMC

NPGD
1, 193–233, 2014

CC BY

to generate a dependent sample from the posterior distribution from which estimates can be derived based on strong laws.

Lele et al. (2010) derived a method called *data cloning* to be used in conjunction with MCMC. The basic idea of their algorithm can be summarized in the following three steps. First, create a k-cloned data set $\mathbf{Y}^{(k)} = (\mathbf{Y}, \mathbf{Y}, \ldots, \mathbf{Y})$ where observed data vector is repeated k times. Second, using an MCMC algorithm, generate a dependent sample from the posterior distribution $\pi_k(\theta|\mathbf{Y})$ which corresponds to the k-cloned data. Third, calculate the sample means and variances of the components of θ; the MLEs of θ correspond to the sample means and the approximate variances of the MLEs correspond to k times the posterior variance of the original data. This method was implemented using `dclone{dclone}` discussed in Solymos (2010). This method is computationally intensive as it involve MCMC, however, it may provide more accuracy than MSIM in small samples.

As an additional method, we proposed a hybrid method using the nonparametric bootstrap and the idea of the data cloning method (*Boot Dclone*). This was done by generating the data, running a nonparametric bootstrap on the rows of the data to create $(k-1)$ new data sets, and then appending the original and all the bootstrapped data into one large set. From there, the data were processed by `dclone` as if it were a single clone. In essence, this created an approximately k-cloned data set to be run through the appropriate MCMC algorithm.

4 Simulations

4.1 Simulation setup

Simulations of the intercept only logit-normal mixed model were conducted with all combinations of 10, 50, 200, and 1000 as the number of subjects (m) and 2, 10, 50, and 200 observations per subject (n). All 5 methods were tested using 100 random data

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

sets; means and standard errors of the 100 sets of estimates were then calculated. We set the true values of the parameters as $(\mu, \sigma^2) = (2, 1)$.

To describe the estimation discrepancy between μ and $\hat{\mu}_{m,n}$, we used squared error loss, $Q(\hat{\mu}_{m,n}) = (\hat{\mu}_{m,n} - \mu)^2$. Because squared error loss is criticized for a bounded parameter space, we used Stein's loss, $S(\hat{\sigma}^2_{m,n}) = \frac{\hat{\sigma}^2_{m,n}}{\sigma^2} - 1 - \log \frac{\hat{\sigma}^2_{m,n}}{\sigma^2}$, to measure how well σ^2 was estimated. A combined loss was then calculated as $G(\hat{\mu}_{m,n}, \hat{\sigma}^2_{m,n}) = Q(\hat{\mu}_{m,n}) + S(\hat{\sigma}^2_{m,n})$.

4.2 Simulation analysis

The results of the simulations are displayed in Tables 1–5. Unsurprisingly, all methods failed to reasonably estimate both μ and σ^2 in the smallest scenario with 10 subjects and 2 observations each. *MSIM*, `dclone`, and `glmer` estimated μ within 2 standard errors for all other settings. The methods also provided reasonable estimates of σ^2 for settings other than those that involved 10 subjects. In general, these three methods indicated a mixture of over and under estimates in each parameter showing no obvious bias.

Boot Dclone displayed an obvious positive bias for both parameters in all estimates; this was exceptionally noticeable in settings with only 2 observations. Reasonable estimates for μ were made in some of the 200 observations per subject settings; σ^2 was never estimated well.

`glmmPQL` did not converge toward true values and seems to display an underestimating bias in both the μ and σ parameters. There were also issues with the function being able to produce estimates for some of the 100 simulations.

4.3 Simulation speed comparison

4 of the 16 settings of data were subsequently tested to determine the speed of the different methods. These were the combinations of 50 and 200 subjects with 10 and

NPGD
1, 193–233, 2014

CC BY

200 observations. The `system.time()` command in R was used to record times. The simulations were each run on 4 computers in one R script on a single core. The computers used were:

- assawa: 2010 Frontier i7 8-core Intel i7 940 (2.93 GHz) with 3 GB of RAM
- geneva: 2011 Frontier i7 8-core Intel i7 950 (3.07 GHz) with 6 GB of RAM
- nokomis: 2012 Optiplex 7010 8-core Intel i7-3770 (3.40 GHz) with 8 GB of RAM
- tilde: 2013 Optiplex 7010 8-core Intel i7-3770 (3.40 GHz) with 8 GB of RAM

MSIM was implemented in two ways for the speed test. In the simplest intercept-only form of the logit-normal model, the code can be reduced to eliminate the use of matrices and can handle much larger data. This is the method used in the previous section. However, a more general form is needed for problems that included covariates. This form relies on matrices and cannot work with large data sets at this time. These methods are referred to *MSIM Fast* and *MSIM Slow* respectively. `dclone` and *Boot Dclone* methods were implemented with 5 clones each- a more reasonable setting than 100 clones for large data.

The results in Tables 6–9 indicated that `glmmPQL` was the fastest in the 50 subject cases and `glmer` was fastest in the 200 subject cases. It is not surprising that these two methods were the best in terms of speed due to the nature of the approximations that they make. `dclone` and *Boot Dclone* were slower. The fastest of either took around 4 min while the longest took nearly 25 min. *MSIM Fast* was faster than the MCMC methods and only slightly slower than the approximations methods at 3 to 6 s per run. The *MSIM Slow* method was much slower with the best case scenario of 1.5 min and the worst case of nearly 4 h. The case with 200 subjects 200 observations could not be handled because the size of matrices and vectors exceeded the storage capacity allowed by R.

NPGD
1, 193–233, 2014

CC BY

4.4 Simulation conclusions

In general, `glmer` and `glmmPQL` were the fastest, but `glmmPQL` did not give good estimates. `dclone` gave accurate estimation but was much slower than the approximation methods. *Boot Dclone* was very poor in both speed and accuracy. *MSIM* proved to be reasonably fast in the intercept-only implementation, however, was much slower than other methods in it's matrix version and failed when too many observations were used. It should be noted that tuning parameters within each of the methods, such as convergence criteria for *MSIM* or number of MCMC draws in `dclone` may impact the time significantly. Because the bootstrapping `dclone` method and glmmPQL showed poor accuracy in the simulations, neither will be used in the following application section.

5 Application in Indian rainfall extremes

5.1 Data

We applied the logit-normal mixed model to Indian rainfall data using fixed covariates of latitude (°), longitude (°), elevation (m) and daily minimum and maximum temperatures (0.1 of °C). Because it is reasonable to assume each weather station is fundamentally different, we fit a random intercept by station. Multiple datasets have been used in the literature of different temporal and spatial granularity, but initially the goal was to develop and test the methods on one widely and freely available dataset. Thus, the data were collected from the National Climatic Data Center (NCDC) in the National Oceanic and Atmospheric Administration (NOAA) from 1 January 1973 to 31 December 2013. Data were queried for all available Indian stations in the database. This produced a total of 149 weather stations with daily precipitation as well as all necessary covariates.

However, as seen in Tables 11–13, 8 of these stations have no actual observations, 14 others have less than 3 observations, and another 22 have less than 1000 out of the possible days in the date range. As an example, on 8 August 2012 (a date within the

CC BY

rainy season for most of India), there were 33 stations with NA values, 12 stations with precipitation of 0 mm, and 31 stations with greater than 0 mm precipitation. This implies several stations were not included in the data for this day and in general, stations included in the data change over time. Figure 1 illuminates the rainfall on this date.

Attri and Tyagi (2010) in the Indian Meteorological Department Report defined 3 categories of rainfall: light rainfall ($0 < x < 64.4\,\text{mm}\,\text{day}^{-1}$), light to heavy rainfall ($64.4 \leq x < 124.4\,\text{mm}\,\text{day}^{-1}$), and extreme rainfall ($\geq 124.4\,\text{mm}\,\text{day}^{-1}$). All stations were were marked each day with indicators of these categories to be used in the modeling. Because of the magnitude of the data, only observations considered to be within monsoon season were used. This conservatively includes the time period from 1 May to 31 October (184 days) for each year. In order to understand the changing variability by station over time we fit models for each year from 1973–2013.

5.2 Results of GLMMs

5.2.1 Light and moderate rainfall results

First, note that all methods failed to produce estimates for the year 1975 because there were only 3 stations with observations in the data set. Fixed coefficient estimation for light rainfall are seen in Fig. 2. Within the light rainfall model, fixed coefficient estimates remained relatively constant over time in each method, but did show some movement around the year 2000 especially in the intercept, minimum temperature, and maximum temperature coefficients. All coefficients other than that of the intercept appear to be very close to 0 indicating that they may have little explanatory power in the case of light rainfall.

Moderate rainfall fixed coefficient estimates found in Fig. 3 showed more of a cyclic pattern over time. Latitude coefficients tended to increase over time as longitude coefficients decreased from slightly positive to slightly negative indicating higher incidence of moderate rainfall in the west. However, both of these coefficients were very close to 0 over time and may actually not significantly explain rainfall. Minimum temperature and

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

maximum temperature also showed mirroring behavior over time but again were very close to 0. In general, the minimum temperature coefficient decreased as the maximum temperature coefficient increased.

Variability estimates for light and moderate rainfall can be found in Fig. 4a and b respectively. Within light rainfall, variability was nearly 0 over time for each of the methods. Slight increases were seen in the 1998–1999 and 2002–2004 time periods. *MSIM* produced a much higher, and likely anomalous, estimate for 2006 than all other methods.

Moderate rainfall still indicated relatively low variability with more of a cyclic nature over time for each of the methods. Peaks were seen in 1976, 1990, 1998–1999, 2002–2004 time periods. Methods produced fairly similar estimates of variability save for the 2008–2012 period when *MSIM* estimated a slightly lower variability than the other methods.

5.2.2 Extreme rainfall results

Again, *MSIM*, `glmer`, and `glmmPQL` could not produce estimates for the year 1975. `dclone` produced reasonable estimates for the extreme rainfall thus, these were used in the figures for 1975.

The results of the fixed coefficient estimation for annual data are seen in Fig. 5. Fixed coefficient estimates in the extremes remained relatively constant over time in each method. `glmer` disagreed with the other methods in the coefficient estimate of maximum temperature but all methods agreed for each of the other fixed coefficients. Elevation and minimum temperature coefficients were effectively 0 in all methods over time. Latitude coefficients were mostly positive and longitude coefficients were mostly negative indicating a higher estimated probability of extreme rainfall in the north west.

Variability estimates for extreme rainfall in Fig. 6a indicated a cyclic pattern in variability over time for each of the methods. If we compare the output of the extreme variability to the monthly Niño–3.4 Sea Surface Temperature (SST) index in Fig. 6b

CC BY

provided by NOAA[1], a clear connection is noted. This provides a monthly average SST anomaly in the Niño region from 5° N–5° S, 120–170° W at each time point. Positive values of this index correspond to El Niño events and negative to La Niña events.

Looking only at the El Niño events classified as a SST anomaly of 0.5 or higher, one can see most of the higher variability estimates are following this type of event. The largest events in the early 1980's and late 1990's appear to have an effect on the variability estimates in the models in the following 5 years.

Visually, this index indicated a similar pattern to the variability estimates by the *MSIM* and `glmer` methods in the later 1980s and *MSIM* and `dclone` in the late 1990s to early 2000s. Correlations computed between the December SST Index value and each of the variability estimates by method can be seen in Table 10. *MSIM* variability estimates had the highest Pearson correlation of 0.27, while `dclone` had the highest Spearman correlation of 0.24. Kendall's τ was not larger than 0.20 for any of the methods. These correlations seem relatively weak, however, it is well known that ENSO is not the only relevant underlying cyclic activity to affect climate. It is also likely that there exists a random lag time between an ENSO anomaly and the corresponding effect on precipitation in the Indian monsoon.

6 Conclusions

We have shown that it is feasible, both theoretically and computationally, to use GLMM in the context of modeling the precipitation that accompanies the Indian monsoon. There appears to be a physical significance to the models based on the connection of the random effects in the extreme rainfall models to ENSO. This shows promise for modeling some of the unobservable physical features within the complicated network of interactions the underly extreme climate patterns while maintaining statistical

[1]http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/detrend.nino34.ascii.txt

NPGD
1, 193–233, 2014

integrity. The current methods of GLMM estimation explored in this article all exhibit some drawbacks, but even so, converged on similar answers in the application to Indian monsoons.

Since the relevance of using GLMM in this context has been established, climate model output, such as that of CMIP5, will be explored to gain deeper intuition of the nature of this random effect. Further work on GLMMs in this context may include additional analysis of ENSO and other proposed drivers of Indian monsoons in their contributions to fixed or random effects.

Providing improvements to the GLMM estimation methods presented here, especially in computing time for MSIM, is another open research area. One limitation of GLMM, as presented in this context, is the reliance of modeling random effects as normal. Expanding the possible distributions of random effects to include extreme value distributions would be a major breakthrough in mixed modeling. Addressing the issue of providing reasonable standard errors for the variance components in GLMM would also lead to a more conclusive test of significance.

Appendix A

Additional simulation specifications

MSIM Fast

- Number of Monte Carlo simulations: 100 000

MSIM Slow

- Number of Monte Carlo simulations: 100
- Convergence criterion for Newton's Method: Euclidean norm of change ≤ 0.01

Dclone

NPGD
1, 193–233, 2014
Indian Monsoon rainfall extremes
L. R. Dietz and S. Chatterjee

CC BY

- Clones: 100
- Prior for μ: $N(0, \frac{1}{0.0001})$
- Prior for $\frac{1}{\sigma^2}$: Gamma(0.01,0.01)
- Adaptation length: 100
- Markov chain length after adaptation: 10 000

Bootstrapping Dclone

- Prior for μ: $N(0, \frac{1}{0.0001})$
- Prior for $\frac{1}{\sigma^2}$: Gamma(0.01,0.01)
- Adaptation length: 100
- Markov chain length after adaptation: 10 000
- Number of bootstraps: 99

Appendix B

Application method specifications

MSIM Slow

- Number of Monte Carlo simulations: 100
- Convergence criterion for Newton's Method: Euclidean norm of change ≤ 0.01

Dclone

- Clones: 5

NPGD
1, 193–233, 2014

CC BY

– Priors for β_0–β_5: $N(0, \frac{1}{0.0001})$

– Prior for $\frac{1}{\sigma^2}$: Gamma(0.01,0.01)

– Adaptation length: 100

– Markov chain length after adaptation: 10 000

Acknowledgements. Thanks to Auroop Ganguly for input on the dataset. Funding support of this research for Ms. Dietz is provided by means of the U. Minnesota Eva O. Miller Fellowship. This research is partially supported by the National Science Foundation under grant #IIS-1029711 and #SES-0851705, and by grants from the Institute on the Environment (IonE), and College of Liberal Arts, U. Minnesota.

References

Attri, S. D. and Tyagi, A.: Climate Profile of India, Tech. rep., Government of India, Ministry of Earth Sciences, India Meteorological Department, 2010. 204

Bates, D.: lme4: Mixed-Effects Modeling lme4: Mixed-Effects Modeling with R, Springer, available at: http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf, preprint, 2010. 198

Breslow, N. and Clayton, D.: Approximate inference in generalized linear mixed models, J. Am. Stat. Assoc.,88, 9–25, 1993. 198

Ghosh, S., Luniya, V., and Gupta, A.: Trend analysis of Indian summer monsoon rainfall at different spatial scales, Atmos. Sci. Lett., 10, 285–290, 2009. 194

Ghosh, S., Das, D., Kao, S.-C., and Ganguly, A. R.: Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes, Nat. Clim. Change, 2, 86–91, 2012. 194

Goswami, B., Venugopal, V., Sengupta, D., Madhusoodanan, M. S., and Xavier, P. K.: Increasing trend of extreme rain events over India in a warming environment, Science, 314, 1442–1445, 2006. 194

Jiang, J.: Consistent estimators in generalized linear mixed models, J. Am. Stat. Assoc., 93, 720–729, 1998. 199

Kumar, K. K., Rajagopalan, B., and Cane, M. A.: On the weakening relationship between the Indian Monsoon and ENSO, Science, 284, 2156–2159, 1999. 195

CC BY

Lele, S. R., Nadeem, K., and Schumuland, B.: Estimability and likelihood inference for generalized linear mixed models using data cloning, J. Am. Stat. Assoc., 105, 1617–1625, 2010. 200

Li, C. and Yanai, M.: The onset and interannual variability of the Asian summer monsoon in relation to land sea thermal contrast, J. Climate, 9, 358–375, 1996. 195

Lin, X. and Breslow, N. E.: Bias correction in generalized linear mixed models with multiple components of dispersion, J. Am. Stat. Assoc., 91, 1007–1016, 1996. 199

McCulloh, C. E. and Searle, S. R.: Generalized, Linear, and Mixed Models, Wiley Series in Probability and Statistics, 2 Edn., Wiley, 2010. 196

Prell, W. L. and Kutzback, J. E.: Sensitivity of the indian monsoon to forcing parameters and implications for its evolution, Nature, 360, 647–652, 1992. 195

Solymos, P.: dclone: data cloning in R, The R Journal, 2, 29–37, 2010. 200

Stiratelli, R., Laird, N., and Ware, J. H.: Random-effects models for serial observations with binary response, Biometrics, 40, 961–971, 1984. 195

Tierney, L. and Kadane, J. B.: Accurate approximations for posterior moments and marginal densities, J. Am. Stat. Assoc., 81, 82–86, 1986. 198

Turner, A. G. and Annamalai, H.: Climate change and the South Asian summer monsoon, Nat. Clim. Change, 2, 587–595, 2012. 194, 195

Table 1. MSIM fast simulation results $\mu = 2$, $\sigma^2 = 1$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	17.41 (4.38)	2.11 (0.07)	2.05 (0.03)	2.00 (0.01)
	50	2.08 (0.05)	1.98 (0.02)	2.02 (0.01)	2.00 (0.00)
	200	2.01 (0.03)	1.98 (0.02)	1.99 (0.01)	1.99 (0.00)
	1000	2.00 (0.04)	1.99 (0.01)	2.01 (0.01)	2.00 (0.00)
σ^2	10	741.99 (302.51)	1.71 (0.33)	1.16 (0.09)	0.97 (0.04)
	50	1.02 (0.10)	0.98 (0.05)	0.98 (0.02)	0.99 (0.01)
	200	0.87 (0.05)	0.97 (0.03)	0.98 (0.02)	0.99 (0.01)
	1000	0.92 (0.06)	0.99 (0.02)	1.00 (0.02)	1.00 (0.01)
Loss	10	2885.74 (1016.73)	4.32 (0.71)	1.14 (0.33)	0.09 (0.01)
	50	3.29 (0.58)	0.19 (0.02)	0.04 (0.01)	0.01 (0.00)
	200	0.33 (0.04)	0.07 (0.01)	0.02 (0.00)	0.00 (0.00)
	1000	0.50 (0.16)	0.05 (0.00)	0.02 (0.00)	0.00 (0.00)

NPGD

1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

Table 2. Data cloning simulation results $\mu = 2$, $\sigma^2 = 1$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	13.18 (2.65)	2.12 (0.05)	2.03 (0.03)	2.02 (0.04)
	50	2.11 (0.07)	1.99 (0.02)	1.99 (0.02)	1.99 (0.01)
	200	2.05 (0.03)	2.02 (0.01)	1.99 (0.01)	2.01 (0.01)
	1000	2.01 (0.01)	2.00 (0.00)	1.99 (0.00)	2.00 (0.00)
σ^2	10	7.79 (1.79)	1.18 (0.11)	0.95 (0.06)	0.98 (0.05)
	50	1.67 (0.33)	1.00 (0.05)	0.99 (0.03)	1.00 (0.02)
	200	1.16 (0.08)	0.99 (0.02)	0.98 (0.01)	1.00 (0.01)
	1000	0.98 (0.04)	0.99 (0.01)	1.00 (0.01)	0.99 (0.00)
Loss	10	131.65 (86.2)	1.26 (0.14)	0.32 (0.04)	0.31 (0.05)
	50	1.58 (0.44)	0.19 (0.02)	0.06 (0.01)	0.04 (0.00)
	200	0.40 (0.06)	0.04 (0.00)	0.02 (0.00)	0.01 (0.00)
	1000	0.09 (0.01)	0.01 (0.00)	0.00 (0.00)	0.00 (0.00)

NPGD

1, 193–233, 2014

L. R. Dietz and S. Chatterjee

CC BY

Table 3. Data cloning bootstrap simulation results $\mu = 2$, $\sigma^2 = 1$.

Par.	# of Sub.	Obs. per Subject 2	10	50	200
μ	10	14.32 (2.5)	2.50 (0.07)	2.12 (0.04)	2.05 (0.04)
	50	4.30 (0.13)	2.32 (0.03)	2.07 (0.02)	2.02 (0.01)
	200	4.23 (0.05)	2.34 (0.01)	2.07 (0.01)	2.04 (0.01)
	1000	4.13 (0.02)	2.32 (0.01)	2.07 (0.00)	2.02 (0.00)
σ^2	10	18.00 (2.08)	2.61 (0.21)	1.26 (0.07)	1.07 (0.06)
	50	14.45 (0.76)	2.27 (0.07)	1.29 (0.03)	1.10 (0.02)
	200	13.95 (0.36)	2.25 (0.03)	1.29 (0.02)	1.10 (0.01)
	1000	13.19 (0.18)	2.25 (0.01)	1.30 (0.01)	1.09 (0.00)
Loss	10	170.18 (103.37)	1.69 (0.35)	0.32 (0.04)	0.29 (0.03)
	50	16.90 (1.29)	0.68 (0.05)	0.10 (0.01)	0.05 (0.01)
	200	15.60 (0.56)	0.59 (0.02)	0.06 (0.01)	0.02 (0.00)
	1000	14.22 (0.25)	0.55 (0.01)	0.05 (0.00)	0.01 (0.00)

NPGD
1, 193–233, 2014

cc BY

Table 4. `glmer` simulation results $\mu = 2$, $\sigma^2 = 1$.

Par.	# of Sub.	Obs. per Subjects 2	10	50	200
μ	10	6.02 (0.70)	2.77 (0.18)	2.33 (0.09)	2.10 (0.02)
	50	2.18 (0.09)	1.99 (0.02)	2.02 (0.01)	2.00 (0.00)
	200	2.03 (0.04)	1.99 (0.02)	1.99 (0.01)	1.99 (0.00)
	1000	2.02 (0.04)	1.99 (0.01)	2.01 (0.01)	2.00 (0.00)
σ^2	10	198.73 (81.39)	7.48 (1.36)	3.07 (0.76)	1.19 (0.04)
	50	1.66 (0.54)	0.95 (0.05)	0.94 (0.02)	0.94 (0.01)
	200	0.93 (0.06)	0.97 (0.03)	0.97 (0.01)	0.98 (0.01)
	1000	0.97 (0.05)	1.00 (0.02)	1.00 (0.01)	0.99 (0.00)
Loss	10	270.68 (84.19)	13.35 (1.91)	3.79 (1.12)	0.11 (0.02)
	50	4.96 (1.15)	0.19 (0.02)	0.04 (0.00)	0.01 (0.00)
	200	0.32 (0.04)	0.06 (0.01)	0.02 (0.00)	0.00 (0.00)
	1000	0.47 (0.20)	0.04 (0.00)	0.01 (0.00)	0.00 (0.00)

NPGD
1, 193–233, 2014

CC BY

Table 5. `glmmPQL` simulation results $\mu = 2$, $\sigma^2 = 1$.

Par.	# of Sub.	Obs. per Subjects 2	10	50	200
	10	3.10 (0.17)	1.92 (0.16)	1.34 (0.14)	0.68 (0.08)
	50	1.83 (0.06)	1.61 (0.03)	1.60 (0.02)	1.54 (0.01)
	200	1.81 (0.04)	1.71 (0.02)	1.73 (0.01)	1.72 (0.01)
	1000	1.81 (0.04)	1.81 (0.02)	1.81 (0.01)	1.79 (0.01)
	10	1.71 (0.13)	1.26 (0.11)	0.81 (0.11)	0.26 (0.06)
	50	0.52 (0.06)	0.25 (0.04)	0.15 (0.04)	0.01 (0.01)
	200	0.51 (0.04)	0.54 (0.03)	0.67 (0.02)	0.68 (0.01)
	1000	0.48 (0.04)	0.72 (0.03)	0.75 (0.01)	0.74 (0.01)
	10	6.04 (0.50)	5.95 (0.47)	7.80 (0.50)	10.21 (0.47)
	50	4.81 (0.44)	5.61 (0.38)	6.26 (0.31)	7.89 (0.12)
	200	3.40 (0.48)	1.77 (0.33)	0.24 (0.06)	0.15 (0.01)
	1000	3.95 (0.47)	0.75 (0.25)	0.10 (0.01)	0.10 (0.00)

NPGD
1, 193–233, 2014

CC BY

Table 6. Total system time (in seconds) results for assawa.

	(# of Subjects, Obs. per Subject)			
Method	(50,10)	(50,200)	(200,10)	(200,200)
glmer	0.135	0.076	0.125	0.096
PQL	0.398	0.336	0.554	0.580
MSIM Fast	4.315	5.545	4.026	4.017
Dclone	15.500	17.790	59.519	65.684
Boot Dclone	16.170	18.393	60.257	69.793
MSIM Slow	150.641	14 912.237	1631.360	–

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

Table 7. Total system time (in seconds) results for geneva.

Method	(# of Subjects, Obs. per Subject) (50,10)	(50,200)	(200,10)	(200,200)
glmer	0.148	0.079	0.134	0.101
PQL	0.420	0.351	0.564	0.565
MSIM Fast	4.299	5.489	3.936	3.934
Dclone	15.265	17.262	59.059	63.735
Boot Dclone	15.025	16.837	56.941	66.439
MSIM Slow	156.357	13 965.834	1550.543	–

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

CC BY

Table 8. Total system time (in seconds) results for nokomis.

	(# of Subjects, Obs. per Subject)			
Method	(50,10)	(50,200)	(200,10)	(200,200)
glmer	0.089	0.048	0.080	0.071
PQL	0.286	0.234	0.384	0.394
MSIM Fast	2.576	3.419	2.479	2.483
Dclone	10.028	11.355	38.069	40.004
Boot Dclone	9.920	10.433	37.474	41.712
MSIM Slow	94.729	9363.849	1069.468	–

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

Table 9. Total system time (in seconds) results for tilde.

	(# of Subjects, Obs. per Subject)			
Method	(50,10)	(50,200)	(200,10)	(200,200)
glmer	0.104	0.046	0.078	0.060
PQL	0.291	0.225	0.372	0.382
MSIM Fast	2.875	3.415	2.475	2.479
Dclone	10.095	10.989	39.086	39.268
Boot Dclone	9.979	10.551	37.743	41.640
MSIM Slow	93.239	9241.401	1057.157	–

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

CC BY

Table 10. Correlations between variability and Nino–34 SST index.

	Correlation Type		
Method	Pearson	Spearman	Kendall
MSIM	0.27	0.16	0.13
Dclone	0.13	0.24	0.15
glmer	0.11	0.15	0.11
PQL	0.06	0.10	0.06

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

Table 11. Summary statistics for Indian rainfall (1973–2013) in mm, Part I.

Station	Min	1st *Q*	Median	Mean	3rd *Q*	Max	NAs	#Obs.	#Actual Obs.
BOMBAY SANTACRUZ IN	0.00	0.00	0.00	7.73	3.00	461.00	4063	14 660	10 597
THIRUVANANTHAPURAM IN	0.00	0.00	0.00	5.42	4.10	361.90	4037	14 221	10 184
JAIPUR SANGANER IN	0.00	0.00	0.00	3.24	0.00	828.00	4456	14 619	10 163
POONA IN	0.00	0.00	0.00	2.75	0.50	471.90	4480	14 630	10 150
MANGALORE BAJPE IN	0.00	0.00	0.00	11.41	9.90	404.10	3897	14 037	10 140
MADRAS MINAMBAKKAM IN	0.00	0.00	0.00	4.79	0.80	294.90	4423	14 542	10 119
GOA PANJIM IN	0.00	0.00	0.00	11.75	4.10	614.90	4579	14 685	10 106
CALCUTTA DUM DUM IN	0.00	0.00	0.00	5.35	2.00	383.00	4500	14 458	9958
RATNAGIRI IN	0.00	0.00	0.00	10.29	6.10	389.90	4007	13 952	9945
GAUHATI IN	0.00	0.00	0.00	5.81	4.10	462.00	4491	14 211	9720
BANGALORE IN	0.00	0.00	0.00	3.37	0.80	306.10	4501	14 125	9624
NEW DELHI SAFDARJUN IN	0.00	0.00	0.00	2.46	0.00	262.90	5136	14 681	9545
NAGPUR SONEGAON IN	0.00	0.00	0.00	3.93	0.50	400.10	4580	14 119	9539
AHMADABAD IN	0.00	0.00	0.00	2.82	0.00	400.10	4842	14 147	9305
BHUBANESWAR IN	0.00	0.00	0.00	5.49	1.80	470.90	4763	14 050	9287
PATNA IN	0.00	0.00	0.00	4.07	0.30	462.00	4889	14 102	9213
AURANGABAD CHIKALTH IN	0.00	0.00	0.00	2.57	0.00	361.90	5209	14 271	9062
BHOPAL BAIRAGARH IN	0.00	0.00	0.00	3.86	0.00	470.90	5048	14 092	9044
MACHILIPATNAM IN	0.00	0.00	0.00	3.75	0.30	462.00	4144	13 181	9037
LUCKNOW AMAUSI IN	0.00	0.00	0.00	3.33	0.00	470.90	5278	14 230	8952
PBO ANANTAPUR IN	0.00	0.00	0.00	2.08	0.00	329.90	5130	14 079	8949
TIRUCHCHIRAPALLI IN	0.00	0.00	0.00	2.84	0.00	310.90	4983	13 931	8948
INDORE IN	0.00	0.00	0.00	3.28	0.00	400.10	5047	13 938	8891
SURAT IN	0.00	0.00	0.00	4.19	0.00	405.90	4798	13 689	8891
AGARTALA IN	0.00	0.00	0.00	6.45	4.10	380.00	4376	13 241	8865
COIMBATORE PEELAMED IN	0.00	0.00	0.00	2.09	0.00	371.10	5106	13 920	8814
RAJKOT IN	0.00	0.00	0.00	2.31	0.00	391.90	5356	14 036	8680
CHITRADURGA IN	0.00	0.00	0.00	2.37	0.50	321.10	4308	12 937	8629
SHOLAPUR IN	0.00	0.00	0.00	2.14	0.00	230.10	5916	14 489	8573
GADAG IN	0.00	0.00	0.00	1.98	0.30	370.80	5081	13 643	8562
BELGAUM SAMBRA IN	0.00	0.00	0.00	3.05	1.00	470.90	4559	13 094	8535
JABALPUR IN	0.00	0.00	0.00	4.60	0.50	443.00	5041	13 523	8482
CUDDALORE IN	0.00	0.00	0.00	3.78	0.00	380.00	5436	13 879	8443
NELLORE IN	0.00	0.00	0.00	2.95	0.00	370.10	5682	14 002	8320

NPGD
1, 193–233, 2014

CC BY

Table 11. Continued.

Station	Min	1st Q	Median	Mean	3rd Q	Max	NAs	#Obs.	#Actual Obs.
GWALIOR IN	0.00	0.00	0.00	2.71	0.00	377.40	5273	13 591	8318
JAGDALPUR IN	0.00	0.00	0.00	4.97	2.00	350.00	4847	13 102	8255
BALASORE IN	0.00	0.00	0.00	5.14	2.00	336.00	4460	12 641	8181
KAKINADA IN	0.00	0.00	0.00	3.02	0.00	300.00	5272	13 420	8148
BIKANER IN	0.00	0.00	0.00	1.18	0.00	363.00	5989	14 133	8144
PATIALA IN	0.00	0.00	0.00	4.04	0.00	494.30	4662	12 753	8091
KOZHIKODE IN	0.00	0.00	0.00	7.83	4.10	317.80	5192	13 156	7964
BHUJ RUDRAMATA IN	0.00	0.00	0.00	1.24	0.00	196.10	5751	13 647	7896
DIBRUGARH MOHANBAR IN	0.00	0.00	0.80	8.83	9.90	470.40	3617	11 328	7711
AKOLA IN	0.00	0.00	0.00	2.79	0.00	333.00	5111	12 680	7569
PORT BLAIR IN	0.00	0.00	1.00	9.56	9.90	297.90	3679	11 185	7506
JODHPUR IN	0.00	0.00	0.00	1.47	0.00	490.00	5613	13 056	7443
HISSAR IN	0.00	0.00	0.00	1.47	0.00	224.00	6363	13 783	7420
KURNOOL IN	0.00	0.00	0.00	1.94	0.00	251.00	5415	12 738	7323
PENDRA ROAD IN	0.00	0.00	0.00	4.94	2.80	462.00	4995	12 136	7141
GUNA IN	0.00	0.00	0.00	3.21	0.00	337.10	4822	11 212	6390
KOTA AERODROME IN	0.00	0.00	0.00	2.14	0.00	180.10	5732	11 922	6190
SRINAGAR IN	0.00	0.00	0.00	2.21	0.50	199.90	5981	12 017	6036
RAIPUR IN	0.00	0.00	0.00	3.21	0.00	300.00	1702	7506	5804
M.O. RANCHI IN	0.00	0.00	0.00	5.45	3.30	403.10	4866	10 541	5675
GAYA IN	0.00	0.00	0.00	4.03	1.00	271.00	5139	10 524	5385
GORAKHPUR IN	0.00	0.00	0.00	5.44	1.00	371.10	5001	10 317	5316
TEZPUR IN	0.00	0.00	0.50	6.83	7.10	300.00	3169	8477	5308
RAMGUNDAM IN	0.00	0.00	0.00	3.03	0.00	272.00	5750	10 961	5211
DEHRADUN IN	0.00	0.00	0.00	8.82	7.10	221.70	3579	8618	5039
JHARSUGUDA IN	0.00	0.00	0.00	6.02	4.10	224.00	5368	10 389	5021
JAISALMER IN	0.00	0.00	0.00	1.07	0.00	463.00	6798	11 549	4751
IMPHAL IN	0.00	0.00	0.50	5.48	5.80	461.80	3479	8178	4699

NPGD

1, 193–233, 2014

CC BY

Table 12. Summary statistics for Indian rainfall (1973–2013) in mm, Part II.

Station	Min	1st Q	Median	Mean	3rd Q	Max	NAs	#Obs.	#Actual Obs.
SATNA IN	0.00	0.00	0.00	3.27	0.00	302.00	5521	10 156	4635
PANAGARH IN	0.00	0.00	0.00	8.95	2.00	826.30	569	5168	4599
BAREILLY IN	0.00	0.00	0.00	5.36	0.80	470.90	5381	9417	4036
KARAIKAL IN	0.00	0.00	0.00	10.64	2.50	730.00	2120	6151	4031
THANJAVUR IN	0.00	0.00	0.00	7.12	0.00	520.70	0	3584	3584
KAPURTHALA OBSERVATORY IN	0.00	0.00	0.00	5.49	0.00	348.50	0	3472	3472
TUMKUR OBSERVATORY IN	0.00	0.00	0.00	5.54	0.00	448.60	0	3418	3418
CHICKMAGALUR OBSERVATORY IN	0.00	0.00	0.00	6.55	1.80	203.50	0	3315	3315
CHITTORGARH OBSERVATORY IN	0.00	0.00	0.00	5.53	0.00	366.30	0	3275	3275
KHAJURAHO OBSERVATORY IN	0.00	0.00	0.00	8.56	0.00	824.50	0	3271	3271
NORTH LAKHIMPUR IN	0.00	0.00	2.00	11.53	14.00	270.00	4083	7354	3271
KOTTAYAM OBSERVATORY IN	0.00	0.00	0.00	20.37	23.40	397.80	0	3216	3216
MANDYA IN	0.00	0.00	0.00	4.65	0.00	294.60	0	3183	3183
PONDICHERY OBSERVATORY IN	0.00	0.00	0.00	8.69	0.00	810.80	0	3140	3140
ERINPURA OBSERVATORY SR IN	0.00	0.00	0.00	3.79	0.00	428.80	0	3039	3039
KARAIKUDI IN	0.00	0.00	0.00	6.62	0.00	319.50	0	2964	2964
DHAR IN	0.00	0.00	0.00	7.51	0.00	407.90	0	2907	2907
AMRELI IN	0.00	0.00	0.00	1.95	0.00	190.50	0	2872	2872
CHAMBAL OBSERVATORY SR IN	0.00	0.00	0.00	5.17	0.00	769.60	0	2849	2849
SHIRALI IN	0.00	0.00	0.00	26.98	17.35	492.80	0	2692	2692
AMBALA IN	0.00	0.00	0.00	5.07	0.00	397.80	606	3277	2671
PANNA OBSERVATORY IN	0.00	0.00	0.00	11.93	0.00	960.90	0	2662	2662
MATHURA OBSERVATORY IN	0.00	0.00	0.00	5.08	0.00	315.70	0	2657	2657
K. PARAMATHY IN	0.00	0.00	0.00	6.79	0.00	933.50	0	2647	2647
VARANASI B.H.U. OBSERVATORY IN	0.00	0.00	0.00	7.05	0.00	337.80	0	2643	2643
DIU OBSERVATORY IN	0.00	0.00	0.00	5.58	0.00	969.30	0	2630	2630
DHARMAPURI IN	0.00	0.00	0.00	5.69	0.00	187.50	0	2555	2555

NPGD
1, 193–233, 2014

CC BY

Table 12. Continued.

Station	Min	1st Q	Median	Mean	3rd Q	Max	NAs	#Obs.	#Actual Obs.
PASIGHAT AERO OBSERVATORY SR IN	0.00	0.00	0.00	28.85	21.10	912.40	0	2526	2526
ARIYALUR IN	0.00	0.00	0.00	6.30	0.00	327.70	0	2522	2522
SAWAI MADHOPUR OBSERVATORY IN	0.00	0.00	0.00	5.50	0.00	558.80	0	2498	2498
HAZARIBAGH IN	0.00	0.00	0.00	9.45	0.00	337.30	0	2293	2293
SILIGURI IN	0.00	0.00	0.00	11.02	7.90	213.10	1799	4036	2237
HOSHIARPUR OBSERVATORY IN	0.00	0.00	0.00	6.42	0.00	344.20	0	2208	2208
PRADEEP CWR OBSERVATORY IN	0.00	0.00	0.00	8.56	0.80	669.50	0	2084	2084
DAMAN IN	0.00	0.00	0.00	12.64	1.00	510.50	0	2082	2082
BHARATPUR OBSERVATORY IN	0.00	0.00	0.00	4.96	0.00	330.20	0	2028	2028
A.F.CHABUA OBSERVATORY IN	0.00	0.00	0.00	18.38	19.30	337.30	0	1980	1980
RAXAUL IN	0.00	0.00	0.00	10.45	0.00	569.70	0	1949	1949
VISHAKHAPATNAM IN	0.00	0.00	0.00	2.67	0.00	279.90	1250	3055	1805
BAPATALA OBSERVATORY IN	0.00	0.00	0.00	5.05	0.00	246.40	0	1763	1763
GHAZIPUR OBSERVATORY IN	0.00	0.00	0.00	6.21	0.00	360.20	0	1746	1746
HONAVAR IN	0.00	0.80	9.90	23.41	34.00	221.00	1708	3162	1454
BARMER IN	0.00	0.00	0.00	0.26	0.00	39.90	2973	3981	1008
GANGANAGAR IN	0.00	0.00	0.00	0.33	0.00	46.00	3115	4085	970
VIJAYAWADA GANNAVA IN	0.00	0.00	0.00	6.30	5.10	132.10	1445	2263	818
UDAIPUR DABOK IN	0.00	0.00	0.00	0.85	0.00	89.90	2735	3432	697
KOLHAPUR IN	0.00	0.00	0.00	3.09	2.00	199.90	548	1199	651
BARODA IN	0.00	0.00	0.00	2.52	0.00	110.00	388	907	519
DALTONGANJ IN	0.00	0.00	2.00	10.00	13.00	115.10	1682	2192	510
BHAGALPUR IN	0.00	0.00	0.00	1.36	0.00	461.50	2771	3274	503
VELLORE IN	0.00	0.00	0.00	3.13	0.00	80.00	181	558	377
NAGAPPATTINAM IN	0.00	0.00	3.00	15.18	19.10	177.00	1072	1448	376
NEW DELHI PALAM IN	0.00	0.00	0.00	0.00	0.00	0.00	4	378	374
SALEM IN	0.00	0.00	0.00	2.94	1.00	53.10	151	476	325
BAHRAICH IN	0.00	0.00	0.00	2.12	0.00	80.00	170	461	291
PAMBAN IN	0.00	0.00	3.00	13.41	18.05	158.00	1158	1449	291
UDAIPUR IN	0.00	0.00	0.00	2.76	0.00	199.90	168	437	269
JAMNAGAR IN	0.00	0.00	0.00	0.50	0.00	41.90	225	472	247
GOPALPUR IN	0.00	0.00	0.00	2.63	0.00	119.90	124	361	237
VARANASI BABATPUR IN	0.00	0.00	0.00	1.15	0.00	36.10	3246	3468	222
NASIK OZAR IN	0.00	0.00	0.00	5.84	3.00	109.00	103	305	202
JAMMU IN AFB IN	0.00	0.00	0.00	3.42	0.00	300.00	136	288	152

NPGD
1, 193–233, 2014

CC BY

Table 13. Summary statistics for Indian rainfall (1973–2013) in mm, Part III.

Station	Min	1st Q	Median	Mean	3rd Q	Max	NAs	#Obs.	#Actual Obs.
MADURAI IN	0.00	0.00	0.00	2.34	0.00	55.10	46	150	104
CHANDIGARH IAFB IN	0.00	0.00	0.00	0.15	0.00	7.10	34	80	46
KOTA IN RAJASTHAN IN	0.00	0.00	0.00	2.38	0.50	17.00	61	86	25
BANGALORE HINDUSTAN IN	0.00	0.00	0.00	2.47	0.50	20.10	755	770	15
BILASPUR IN	0.00	0.00	0.00	0.00	0.00	0.00	1	3	2
MAINPURI IN	0.00	0.00	0.00	0.00	0.00	0.00	0	2	2
MALDA IN	0.00	0.75	1.50	1.50	2.25	3.00	0	2	2
BETUL IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
CHAIBASA IN	3.00	3.00	3.00	3.00	3.00	3.00	0	1	1
DHOLPUR IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
DUMKA IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
JHALAWAR IN	0.00	0.00	0.00	0.00	0.00	0.00	1	2	1
KHANDWA IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
KHERI IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
MIDNAPORE IN	1.00	1.00	1.00	1.00	1.00	1.00	0	1	1
NALIYA IN	0.00	0.00	0.00	0.00	0.00	0.00	0	1	1
RANCHI IN	2.00	2.00	2.00	2.00	2.00	2.00	2	3	1
AMBIKAPUR IN	NA	NA	NA	NA	NA	NA	1	1	0
KALLAKKURICHCHI IN	NA	NA	NA	NA	NA	NA	1	1	0
NIZAMABAD IN	NA	NA	NA	NA	NA	NA	1	1	0
PACHMARHI IN	NA	NA	NA	NA	NA	NA	1	1	0
SAGAR IN	NA	NA	NA	NA	NA	NA	1	1	0
SANGLI IN	NA	NA	NA	NA	NA	NA	1	1	0
SHIMLA IN	NA	NA	NA	NA	NA	NA	1	1	0
UMARIA IN	NA	NA	NA	NA	NA	NA	1	1	0

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

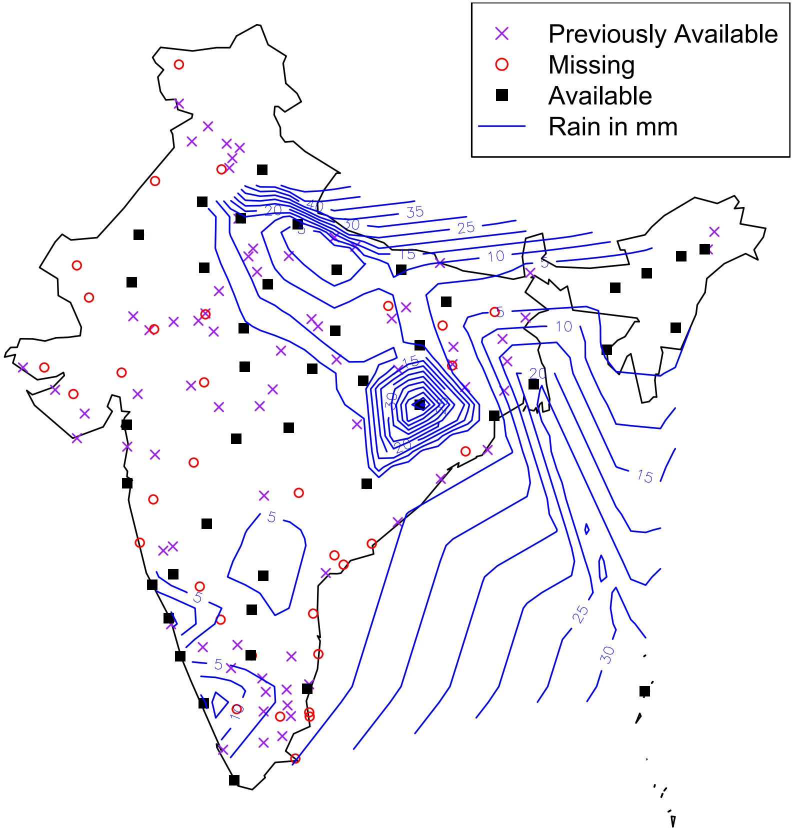

Fig. 1. Observed Indian rainfall (in mm) on 25 August 2012, shown in contours. Markers indicate data status of individual stations.

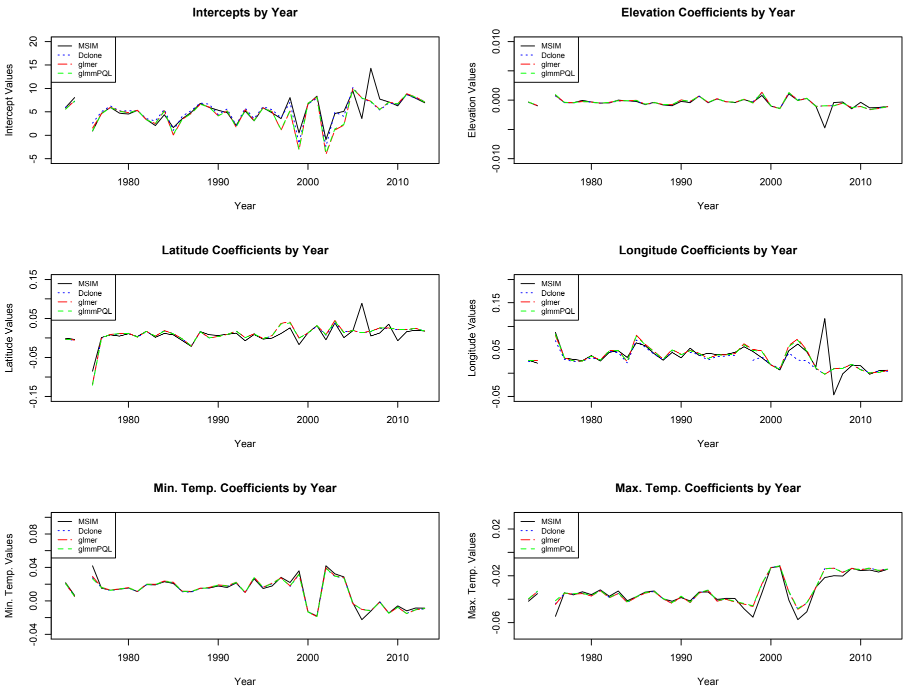

Fig. 2. Fixed coefficient estimates for logit-normal models with light Indian rainfall ($0 < x < 64.4\,\mathrm{mm\,day^{-1}}$) as the response from 1973–2013.

CC BY

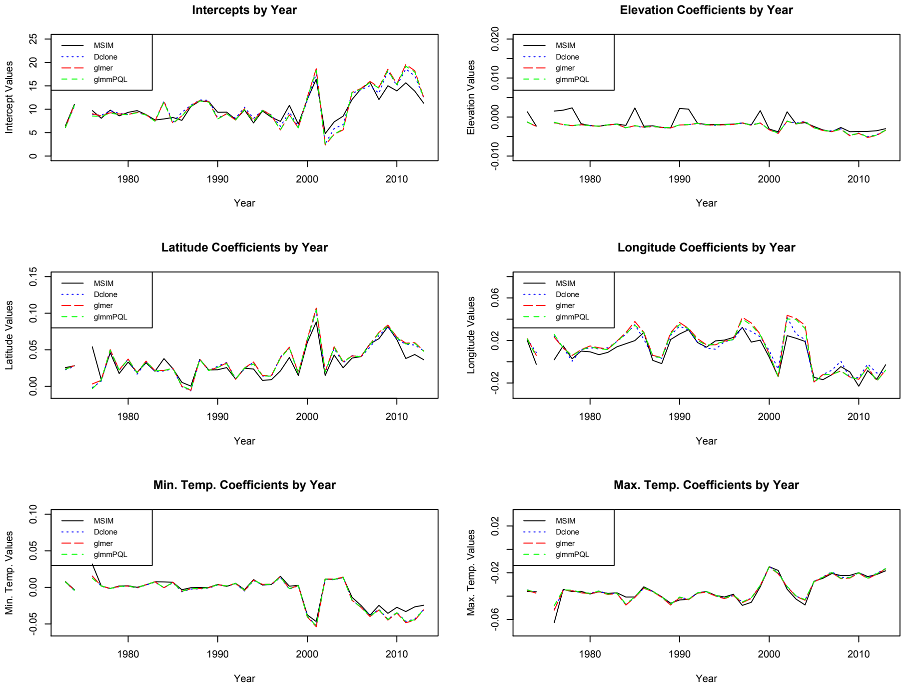

Fig. 3. Fixed coefficient estimates for logit-normal models with moderate Indian rainfall ($64.4 \leq x < 124.4\,\mathrm{mm\,day^{-1}}$) as the response from 1973–2013.

CC BY

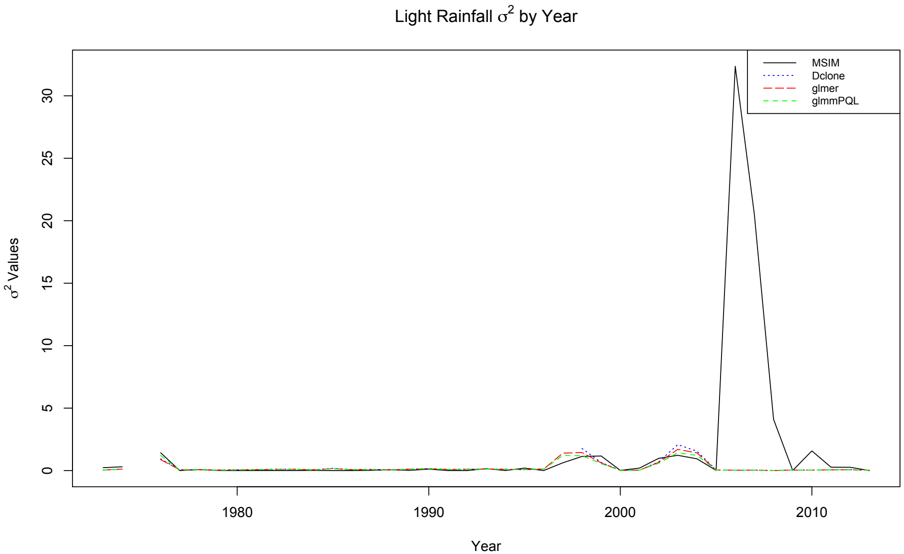

Fig. 4a. Weather station variance component estimates for logit-normal models with light Indian rainfall ($0 < x \leq 64.4\,\text{mm}\,\text{day}^{-1}$) as the response from 1973–2013. Estimates over time indicate variability near 0; the MSIM estimate in 2006 appears anomalous.

NPGD
1, 193–233, 2014

CC BY

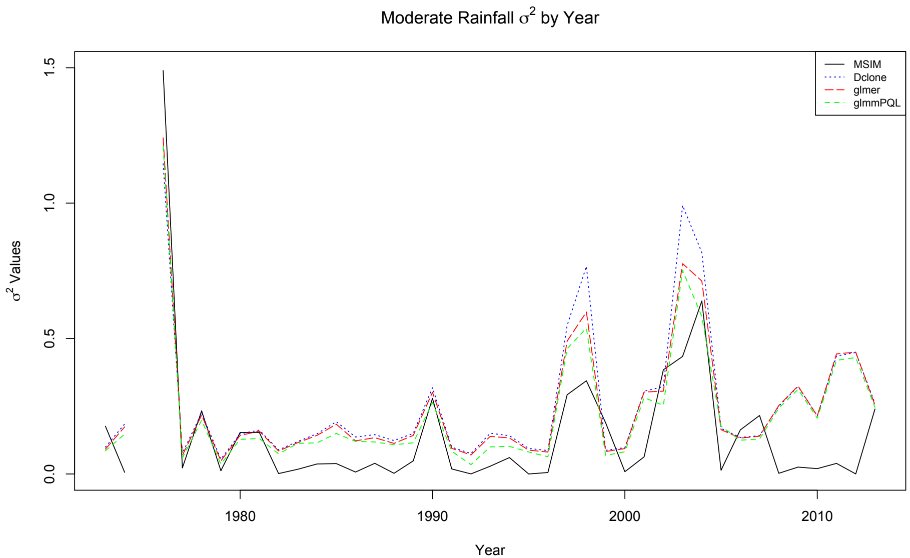

Fig. 4b. Weather station variance component estimates for logit-normal models with moderate Indian rainfall ($64.4 \leq x < 124.4\,\mathrm{mm\,day^{-1}}$) as the response from 1973–2013. Estimates over time show cyclic behavior with slight increases in peaks in the late 1990s and early 2000s.

NPGD
1, 193–233, 2014

L. R. Dietz and S. Chatterjee

CC BY

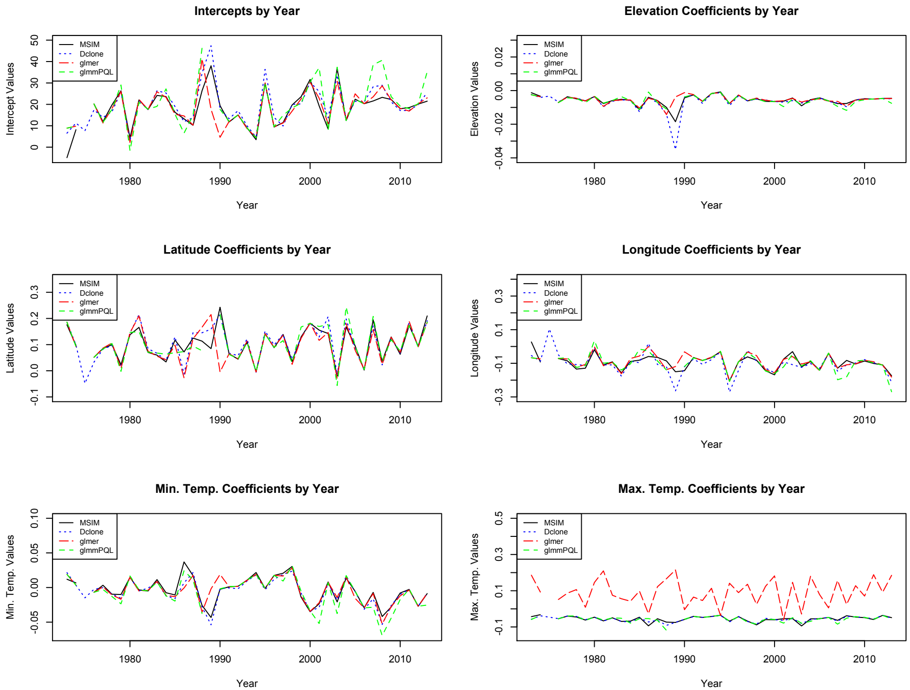

Fig. 5. Fixed coefficient estimates for logit-normal models with extreme Indian rainfall (≥ 124.4 mm day^{-1}) as the response from 1973–2013.

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

CC BY

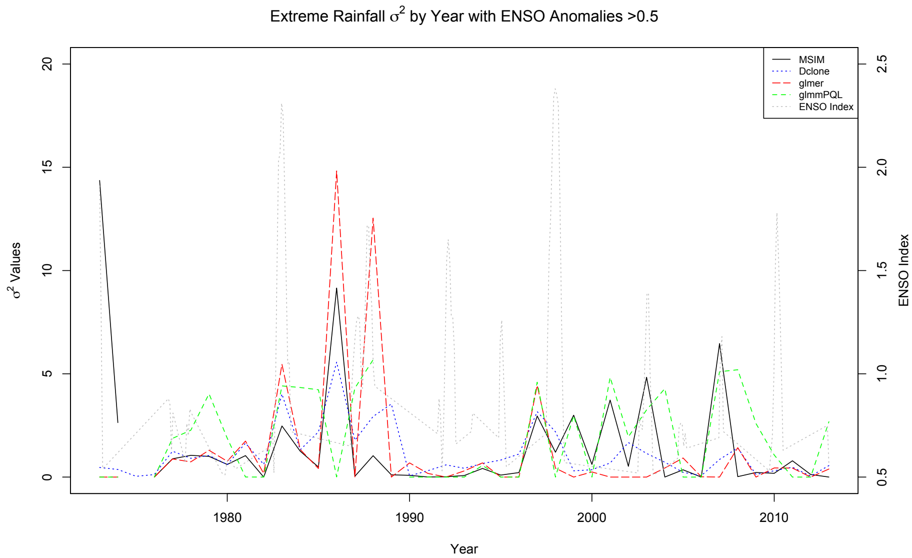

Fig. 6a. Weather station variance component estimates for logit-normal models with extreme Indian rainfall ($\geq 124.4\,\mathrm{mm\,day^{-1}}$) as the response from 1973–2013. Estimates are overlaid with ENSO Index anomalies > 0.5 which indicates larger variance estimates lagging notable El Niño events.

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀▌ ▶▌
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

cc BY

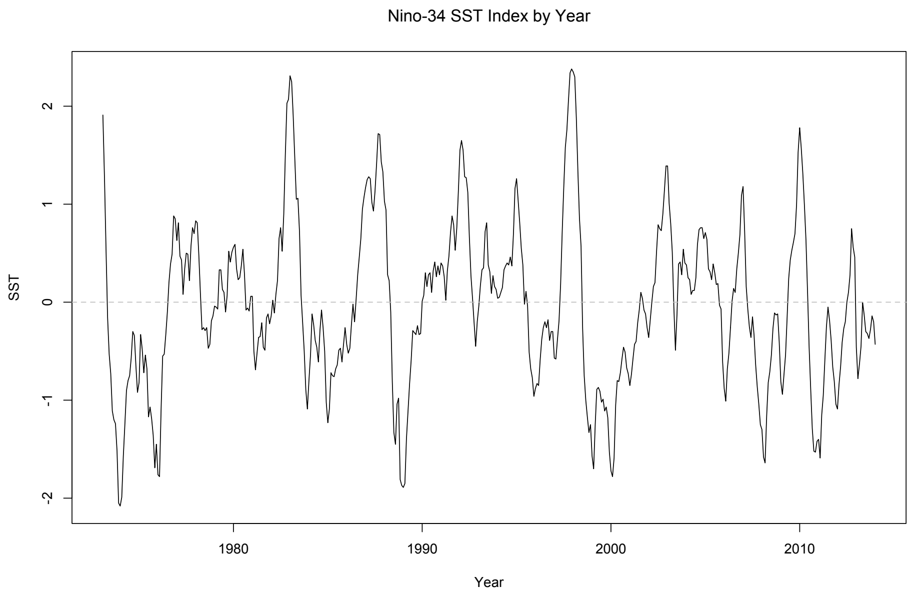

Fig. 6b. ENSO Index from 1973–2013. SST above 0 corresponds to El Niño events; SST below 0 corresponds to La Niña events.

NPGD
1, 193–233, 2014

Indian Monsoon rainfall extremes

L. R. Dietz and S. Chatterjee

CC BY