If rain falls in India and no one reports it, are historical trends in monsoon extremes biased?

Marena Lin[1], Peter Huybers[1]

Marena Lin, lin8@fas.harvard.edu

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02139, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018GL079709

With two-thirds of the total Indian population employed by the agriculture sector, changes in Indian monsoon precipitation have widespread implications for human welfare. Increased extreme precipitation since 1950 has been widely reported for Central India. Major studies have relied upon the gridded daily precipitation observations provided by the India Meteorological Department (IMD), which assimilate observations from a variable network of weather stations. Replicating the IMD's interpolation method on satellite-based precipitation observations, however, indicates that temporal changes in the observing weather station network introduce a jump in the record toward more extreme rainfall after 1975. Trends evaluated across this jump are suspect, and trends evaluated subsequent to it are insignificant ($p > 0.1$). This positive bias may also mask declines in average monsoon rainfall. Greater accuracy in these trends may resolve distinctions between climate model simulations of future changes. Access to the underlying data from IMD rain-gauges would facilitate improved rainfall reconstruction.

Keypoints:

• A method is introduced to diagnose how changes in gauge locations bias rainfall estimates.

• Central Indian rainfall estimates from 1951-2016 are biased toward increasing extreme rainfall.

Plain Language Summary

Previously reported trends in daily monsoon rainfall since 1950 have been estimated using interpolated weather station observations released by the India Meteorological Department. The number of reporting weather stations changes over time, and poor coverage by weather stations can overlook extreme rainfall events. By applying the interpolation of this changing network to satellite-based rainfall data, we show that the changing coverage of weather stations in the Indian rainfall data leads to spurious increases in extreme rainfall. This suggests that previously reported trends of extreme rainfall are biased positive. Access to the raw weather station data would improve our ability to track changes in the Indian monsoon and assess modeled predictions given climate change.

1. Introduction

Extreme precipitation events in South Asia can translate to catastrophic loss of life and displacement of millions [e.g. *Sharma and Hussain*, 2017]. A number of studies have noted increases in the frequency, intensity, and spatial variability of extreme rainfall in Central India since 1950 [*Goswami et al.*, 2006; *Rajeevan et al.*, 2008; *Roxy et al.*, 2017; *Krishnamurthy et al.*, 2009; *Singh et al.*, 2014; *Ghosh et al.*, 2012; *Malik et al.*, 2016]. Studies agree on the apparent increase of spatially aggregated metrics of extreme rainfall, defined as the count of days with rainfall greater than 100 mm day^{-1} summed over gridboxes in Central India, but extreme rainfall trends taken over smaller domains or individual locations are heterogeneous and of varying statistical significance [*Krishnamurthy et al.*, 2009; *Ghosh et al.*, 2012; *Malik et al.*, 2016; *Vinnarasi and Dhanya*, 2016; *Pai et al.*, 2015]. Furthermore, observed increases in extremes are concurrent with weak declines in average daily monsoon rainfall [*Turner and Annamalai*, 2012; *Bollasina et al.*, 2011; *Singh et al.*, 2014].

Analysis of historical changes in extremes has only become possible with the publication of precipitation data collected by the India Meteorological Department (IMD), gridded at resolutions of $1° \times 1°$ and $0.25° \times 0.25°$ [*Rajeevan et al.*, 2006; *Pai et al.*, 2014]. Following the methods used to produce major global precipitation datasets, the IMD interpolate station data onto these grids using a distance-based weighted averaging of the nearest weather stations within a specified search radius [*Alexander et al.*, 2006; *Hofstra et al.*, 2008; *Donat et al.*, 2013; *Rajeevan et al.*, 2006]. Although of substantial utility, the gridded India data are nonetheless subject to similar issues of spatio-temporal inhomogeneities from differences and changes in collection methods that have been noted for other precipitation data-sets [*Adler et al.*, 2003; *Alexander et al.*, 2006; *Hofstra et al.*, 2008, 2010; *Donat et al.*, 2013]. Interpolation methods used are well-suited to taking an area-weighted mean but smooth the geographic variability of precipitation [*Hofstra et al.*, 2010; *King et al.*, 2013]. Furthermore, variations in collection interval, topographical influences on precipitation, and type of rain gauge used can introduce spatial heterogeneity, and changes in the weather station network through time can bias time trends in rainfall [*Alexander et al.*, 2006; *King et al.*, 2013].

Restricting interpolation to stations that have coverage for the majority of a time-interval helps guard against inhomogeneities from changing coverage [*Rajeevan et al.*, 2006; *Goswami et al.*, 2006; *Donat et al.*, 2013; *Harris et al.*, 2014]. Accordingly, the $1° \times 1°$ IMD product only assimilates data from stations that report for at least 90% of the 1951-2003 interval, but the number of stations used drops by more than half between 1990 and 2010, a proportional decline also seen in the 0.25° IMD data [*Rajeevan et al.*, 2006]. The potential for bias from changing coverage therefore remains and has been difficult to quantify without access to the raw precipitation data [*Alexander et al.*, 2006; *Hofstra et al.*, 2010; *Donat et al.*, 2013; *King et al.*, 2013]. In the following, we introduce a method for examining the influence of changes in the network of rainfall gauges on the inference of extremes by applying the IMD's interpolation approach to satellite rainfall data. We follow previous studies by focusing on the Central India domain and define extreme rainfall using a 100 mm day^{-1} threshold [e.g. *Goswami et al.*, 2006; *Rajeevan et al.*, 2008]. The 100 mm day^{-1} threshold is breached for less than 1% of all June through September days in Central India.

2. Data and Methods

2.1. Counting extreme rain

Extreme rainfall incidence has been typically presented as a single annual estimate over Central India, a region approximately defined as bounded by 74.5°E, 86.5°E and 16.5°N, 26.5°N [e.g. *Goswami et al.*, 2006; *Singh et al.*, 2014]. For the 0.25° × 0.25° IMD data, this region comprises 1845 gridboxes. The annual count of extreme rainfall events is the number of days that daily rainfall exceeds the 100 mm day^{-1} threshold, summed over these gridboxes for the 122 June-September days in a monsoon season. In keeping with past studies, we also consider a 150 mm day^{-1} threshold.

2.2. Shepard interpolation scheme

Daily precipitation data is available in the form of gridded data sets from the IMD, which assimilate daily weather station data onto evenly spaced gridpoints using a variant on the Shepard interpolation method [*Shepard*, 1968]. To produce the IMD $0.25^\circ \times 0.25^\circ$ gridded rainfall data-set, each gridded rainfall value on a given day is derived from a weighted average of up to the four nearest weather stations, n, within a search radius of 1.5° [*Rajeevan et al.*, 2005; *Pai et al.*, 2014]. The weight for each observation depends on the distance, r_k, of the station from the center of the gridbox. For each gridded rainfall value, an average interpolation distance between the gridbox center and the n stations meeting the search criteria is computed as, $h = \frac{1}{n}\sum_{k=1}^{n} r_k$. In regions with sparse weather station coverage, the nearest station to the center of a gridbox may be 1°, or 10^2 kilometers, away. The variant of Shepard interpolation employed to produce all IMD gridded precipitation data uses weights, w_k, which depend on a distance-based metric between the gridpoint center and the weather station. The distance metric, s_k, is defined in terms of the search radius, R, and the distance vector, r_k, between the gridbox center and station location,

$$s_k = \begin{cases} r_k^{-1}, & r_k < \frac{R}{3} \\ \frac{27}{4R}\left(\frac{r_k}{R} - 1\right)^2, & \frac{R}{3} \leq r_k < R. \end{cases} \tag{1}$$

The weight is computed as the product, $w_k = s_k^2(1+t_k)$, where t_k is a directional weighting cosine term. For each weather station used to estimate daily rainfall in a gridbox, a directional weighting term is applied, $t_k = \frac{\sum_j s_j\left(1-\frac{\vec{\mathbf{r}}_k\cdot\vec{\mathbf{r}}_j}{r_k r_j}\right)}{\sum_j s_j}$, where $\vec{\mathbf{r}}$ is a vector from the center of a gridbox to the station and j iterates across the n stations used in the estimate.

This interpolation scheme gives similar results in extreme event counts to the simpler weighting scheme, $w_k = {r_k}^{-2}$.

2.3. IMD rainfall and station network variability

Although the exact station locations are not provided, the IMD gridded data include daily counts of reporting stations within each gridbox. For the 1951-2016 interval of the 0.25° IMD data, 56% of Central India gridboxes on an average day contain no reporting stations, 32% have one station, and the remainder have multiple stations. Inferred interpolation distances show variability in annual averages ranging from 0.36° in 1971 to 0.24° in 1991 (Fig. 1c). In addition to the effects of its temporal variability, the average geographic placement of the station network has implications for how rainfall is mapped. The spatial correlation between average rainfall and average interpolation distance from weather stations is positive at $r = 0.44$ (Fig. 1). This implies that regions with heavier rainfall are more sparsely sampled, and that spatial interpolation will play a correspondingly larger role in regions more likely to experience extreme rainfall events.

In order to quantify changes in the observing network, we use a breakpoint analysis. A breakpoint of 1975 for the annual average interpolation distance from 1951-2016 (Fig. 1c) is found when minimizing the sum of residual squares of a two-mean model, one mean taken over values preceding and including the breakpoint and a second mean of values subsequent. Maps of the average interpolation distance for these two intervals show a decline in the number of interpolation distances exceeding 0.5° (Fig. 1de) and an overall increase in density. The percent of all interpolation distances greater than 0.5° transitions from 11% until 1975 to 7% over the more recent interval. These changes occur in the vicinity of where extreme events are climatologically expected, further indicating that changes in the network may influence observations of extremes. Of note is that the 1971 dramatic decline in reporting IMD stations in Central India (Fig. 1c) was preceded by the November 12, 1970 Bhola cyclone in today's Bangladesh. This cyclone swept away an estimated quarter-million people and may have been a factor in instigating the Bangladesh Liberation War [*Hossain*, 2018]; degradation in the rainfall observing network may have resulted in their immediate aftermath. We speculate that improved density of stations between 1972-1975 may be related to the 1972 monsoon failure and demographic consequences [*Drèze and Sen*, 1989].

2.4. Replicating interpolation effects through TRMM

To examine the implication of an inhomogeneous, changing network of rain gauges, it is useful to compare IMD estimates that bear the artifacts of interpolation against those from the spatially-complete observations of the Tropical Rainfall Measuring Mission (TRMM). We use the microwave-only subset of the 3-hour TRMM Multi-satellite Precipitation Analysis 3B42 v.7 data, which have not been adjusted for rain gauge observations, unlike other available TRMM data-sets. These data are provided at 3-hourly intervals centered at times starting at 0 UTC and gridded at $0.25° \times 0.25°$ from 1998-2015 within 50° of the equator [*Huffman et al.*, 2007]. Microwave-sensed rainfall has its limitations in accurately estimating high rainfall rates and may miss short-duration rainfall events, but TRMM has been shown to be broadly consistent with global gauge data in capturing extreme daily precipitation, particularly in South Asia [*Brown*, 2006; *Libertino et al.*, 2016]. Over Central India, the correlation in monsoon total rainfall between TRMM and IMD is $r = 0.45$ between 1998-2015, but increases to $r = 0.71$ when excluding the earliest two years of data.

Integrated to match the IMD daily collection period ending at 3 UTC [*Rajeevan et al.*, 2005], this TRMM product reports systematically lower daily rainfall totals relative to IMD. For example, area-averaged mean daily rainfall estimated by TRMM during the summer season over our region of interest is 5.2 mm day^{-1}, compared to 7.4 mm day^{-1} in IMD data (Fig. 1a). In order to accommodate this offset, TRMM extremes are estimated according to the matching percentile, where a daily rainfall value of 100 mm day^{-1} corresponds to the 99.6$^{\text{th}}$ percentile in the IMD dataset, when aggregating over Central India. The estimated extremes time-series in the IMD data-set is very similar if a spatially variable 99.6$^{\text{th}}$ percentile is instead used for detecting extremes, having a correlation of $r = 0.94$ with the time-series derived using a fixed 100 mm day^{-1} threshold. *Roxy et al.* [2017] finds a similarly high $r = 0.97$ using the 150 mm day^{-1} threshold. We thus set the threshold for counting extreme precipitation to the 99.6$^{\text{th}}$ percentile rainfall value of each Central India monsoon season of TRMM data. Although the absolute precipitation amounts estimated from TRMM are generally lower than rain gauge observations, satellite data preserve spatio-temporal variation [*Libertino et al.*, 2016].

The sparse and changing IMD station network is examined relative to the complete and constant spatial coverage of the TRMM dataset using a technique similar to that employed to estimate overlooked hurricanes [*Vecchi and Knutson*, 2010]. Specifically, a given monsoon season of TRMM precipitation is repeated across the 66 monsoon seasons of the 1951-2016 IMD data. For gridboxes where the IMD data have no stations, the values from the spatially-complete TRMM data are replaced by interpolated values from the nearest gridboxes having stations using Shepard interpolation (Section 2.2). Station distance is taken as the length between gridbox centers. In Fig. 2, the August 7, 2007 TRMM rainfall is interpolated using the IMD station networks corresponding to August 7, 1971 and August 7, 1993. The count of extreme events for each of these interpolations is markedly different. The rainfall collected on August 7, 2007 corresponded to a monsoon deep depression [*India Meteorological Department*, 2008] and was part of a South Asian monsoon season that had already displaced 20 million people [*CNN*, 2007]. The original TRMM data of this deep depression showed 60 extreme events; interpolation according to the 1971 network shows only 33 extreme events; and interpolation using the denser 1993 network captures 52 events. This example illustrates the tendency of a sparse network to underestimate the true count of extreme events. Interannual changes to extreme rainfall in the resulting time-series are exclusively a function of the interpolation of the station network. Eighteen such time-series are realized, one for each year of available TRMM data from 1998-2015.

3. Results and Discussion

A monsoon season of interpolated TRMM data always underestimates the count of extreme rainfall events of the original TRMM data, omitting 17.8% of extreme events across all years of interpolated TRMM data. Such underestimation is not strictly necessary, with 6.5% of individual days actually showing higher counts, but is strongly favored because the smoothing inherent to IMD's interpolation method suppresses the occurrence of extremes.

Anomalies in the occurrence of extreme rainfall found in the surrogate TRMM time series show a breakpoint (Fig. 3a) of 1975, identical to the breakpoint found for the annual average interpolation distance (Fig. 1c). Although there is variability across these anomaly time-series depending on the year of TRMM data used, a general shift toward higher counts is apparent across 1975. The cause of the apparent increase is that station densities increase after 1975, east of 78°, a region with high daily mean rainfall (Figs. 1cde). The low density of weather stations in this region may be tied to its history of colonial land tenure systems and the related limited delivery of public goods, relative to other parts of India [*Banerjee and Iyer*, 2008]. Interpolation of the varying station network, which shows a net rise in density of stations, imparts a positive bias to estimates of extreme rainfall trends across the mid-1970s, absent actual interannual changes to the South Asian monsoon. Note that the effects of interpolation cannot be deconvolved from true rainfall variability, in general, even given knowledge of the exact station locations, because the smoothing inherent to IMD's interpolation approach suppresses variability from individual stations.

It is useful to revisit the significance of trends in extreme rainfall in light of the presence of observational bias. We rely on a bootstrapping technique to evaluate the statistical significance of linear trends in the IMD data because climate extremes do not obey Gaussian statistics [*Efron and Tibshirani*, 1993]. For each of these IMD extreme event time-series (Fig. 4), a null hypothesis of zero trend is tested by sampling these values with replacement and equal probability to form 10^4 surrogate time-series. The 10^4 simple linear regression trends of these time-series then provide a sample distribution for our null.

Consistent with major studies of extreme rainfall over Central India using IMD data, the positive extreme rainfall trends since 1950 would appear significant, absent accounting for any breakpoints ($p < 0.01$) [*Goswami et al.*, 2006; *Roxy et al.*, 2017]. However, the 1973-2016 time-trends of an increase of 4.67 extreme events per year using the 100 mm day^{-1} threshold and 2.12 extreme events per year using the 150 mm day^{-1} threshold are insignificant at the 90% confidence level ($p > 0.1$, two-sided test, Fig. 4). While the 1974-2016 trend with the 100 mm day^{-1} threshold appears significant at $p=0.09$, the fact that the trend for the corresponding 150 mm day^{-1} threshold and trends for both thresholds beginning at 1975 and 1976 are not significant ($p > 0.1$) raises concerns of the suitability of a trend metric that is so sensitive to start year. A large change in station network is apparent in the 2000s, and it would also be possible to exclude data after 2000 on the basis of further changes in station network density, but the resulting interval also provides no significant evidence for a trend, as can be anticipated given the omission of the high extreme event counts in the 2000s.

A concern is that the observed biases are peculiar to the TRMM precipitation data. To test this, we apply the same interpolation to each monsoon season of the gridded IMD data, assuming the counterfactual that each day of these rainfall data is an accurate and complete spatial field. The range of the anomaly extreme rainfall counts from these IMD interpolations is presented in Fig. 3a by the gray shading. A breakpoint of 1973 is obtained for the average interpolated IMD time-series. The increase across the 1973 breakpoint is as high as 91 extreme events (> 100 mm day^{-1}) using the re-interpolated 2007 IMD data. For the TRMM data, the maximum offset at 1975 is 71 events for 99.6$^{\text{th}}$ percentile events when using the re-interpolated 2011 monsoon season. The simulated network changes are only those that can be directly inferred from the daily gridded station counts included in the IMD data. More detailed information, such as station locations and reporting intervals, could yield larger changes, as could shifts in instrumentation or measurement practice that may accompany changes in network coverage.

We repeat the interpolations of the TRMM data using the station network inferred from the 1951-2015 1° IMD data. These data use 1,803 stations, in contrast to the 6,955 used in the 0.25° data [*Rajeevan et al.*, 2005; *Pai et al.*, 2014]. For the daily 1951-2015 1° gridded IMD station counts, daily geographic maps of station locations are estimated by randomly sampling, without replacement, the 0.25° TRMM boxes contained within each 1° IMD gridbox to match the listed count of stations. Rainfall at these 1° gridbox centers are then interpolated according to these daily maps, in effect treating the TRMM 0.25° gridboxes as stations. The expected count of extremes, again using the 99.6th percentile as a threshold, is taken from the nearest-neighbor interpolation of the 0.25° TRMM data onto the IMD 1° grid. This is analogous to the presence of a reporting weather station at each 1° gridbox center. The combination of a lower-resolution grid and a sparser network yields even greater shortfalls in estimated extremes. Relative to the 1° nearest-neighbor TRMM rainfall dataset described above, the 1° TRMM dataset interpolated according to the 1° IMD stations captures 45% fewer extreme events. As a fraction of expected extremes, these time-series show greater interannual variance than those from the interpolated 0.25° TRMM data. The largest positive 1951-2015 trend in extremes in these 1° interpolated data is derived from the 2007 TRMM season, with an increase of 5.12 extreme events per fifty years, which is over a quarter of the trend of 18.4 extreme events per fifty years in the 1° IMD rainfall data.

Results are therefore qualitatively consistent between the 1° and 0.25° data, and we focus on the 0.25° results, which caution against taking trends of extreme rainfall using the gridded IMD data across the mid-1970s. Foregoing analyses of extreme rainfall that depend on IMD datasets [*Singh et al.*, 2014; *Rajeevan et al.*, 2008; *Goswami et al.*, 2006; *Ghosh et al.*, 2012; *Roxy et al.*, 2017] are presumably subject to the biases that we infer.

Although to a lesser extent than for extreme values, mean precipitation is also susceptible to biases from interpolation as a result of the long-tailed distribution of precipitation. Average daily monsoon rainfall is at the 80th percentile of all monsoon days. Recent studies indicate weak declines in mean summer rainfall in Central India [*Turner and Annamalai*, 2012; *Singh et al.*, 2014]. We estimate a decline of 0.29 mm day^{-1} per fifty years for the Central India area-averaged daily monsoon rainfall for the 1951-2016 IMD data, but the interpolated TRMM data suggest that this may be underestimated. Scaling for the 43% greater value of IMD mean daily rainfall over that of TRMM, an equivalent positive trend of 0.06 mm day^{-1} per fifty years is obtained from interpolating the 2008 TRMM data with the IMD network. On average, interpolated TRMM data show a positive trend of 0.02 mm day^{-1} per fifty years. The TRMM data interpolated according to the 1951-2015 1° IMD station network suggests an even larger positive bias in the trend of area-averaged mean daily monsoon rainfall, as a fraction of the observed 1° IMD trend. The interpolation of 2007 TRMM data gives a positive 1951-2016 trend that is 28% of the magnitude of observed drying of area-averaged mean daily monsoon rainfall in the 1°

IMD data. The average positive bias across these years of TRMM data interpolated to the IMD 1° station network is 14% of this magnitude.

In addition to affecting statistics of daily rainfall, the sparse station network may have implications for inferred changes in the incidence and spatial characteristics of low pressure systems, both due to the high proportion of total monsoon rainfall for which these systems are responsible and because the station network is least dense near the eastern coast of India, where synoptic activity is highest (Fig. 1b and see Fig. 2a of *Ajayamohan et al.* [2010]). For example, 2006 boasts the highest number of extreme rainfall events across the IMD record, but over a third of these events are accounted for by two depressions that made landfall in East India, one occurring from July 2-5 and the second from August 2-5 (black bars in Fig. 4) [*India Meteorological Department*, 2007]. Sparse networks can cause errors in estimates of the extent of depressions (i.e. Fig. 2), and because the frequency and magnitude of such depressions varies across India, the potential for missing extreme rainfall events also varies. Raw weather station data and averaged precipitation metrics over smaller homogeneous regions may be less susceptible to the biases we identify [*Krishnamurthy et al.*, 2009; *Guhathakurta and Rajeevan*, 2008; *Malik et al.*, 2016].

Estimation of the historical variability of monsoon precipitation has long motivated important studies on linked global atmospheric phenomena and the local influence of anthropogenic warming and aerosol emissions [*Walker*, 1923; *Jayasankar et al.*, 2015; *Mukherjee et al.*, 2018; *Bollasina et al.*, 2011]. The effects of anthropogenic warming on the summer Indian monsoon are uncertain [*Allen and Ingram*, 2002; *O'Gorman and Schneider*, 2009; *Jayasankar et al.*, 2015; *Mukherjee et al.*, 2018]. Quantifying the biases induced by a changing observational network is important for reconciling and attributing observed trends with modeled precipitation estimates under different emissions scenarios. Access to raw weather station observations would allow for more accurate estimates of change in the South Asian monsoon phenomena and their uncertainties. Like the Indian data, other precipitation measurements from around the globe are often available only at the discretion of state restrictions on data publication. International collaborations facilitating the analysis of raw rain gauge data seem critical for accurate evaluation of historical changes in precipitation [*Alexander et al.*, 2006; *Donat et al.*, 2013; *Yatagai et al.*, 2012].

Acknowledgments. We thank Eric Stansifer for providing key contributions to experimental design and for feedback on the analysis and manuscript. We thank Duo Chan, Sunil Amrith, and Zhiming Kuang for helpful discussions. We are grateful to the India Meteorological Department (IMD) staff for curating rainfall data and fielding questions on their provenance. This work was supported in part by NSF GRFP grants DGE1144152 and DGE1745303, the Ray Goldberg Fellowship in Global Food Systems, and the Harvard Global Institute. Data analysis was performed on the Odyssey cluster supported by FAS Science Division Research Computing Group at Harvard University. The IMD gridded daily rainfall data are available through purchase from the IMD. The interpolations of TRMM and IMD data and code for all analyses are publicly available: https://github.com/marenalin/Lin_Huybers_2018

References

Adler, R., G. Huffman, A. Chang, R. Ferraro, P. Xie, J. J. B. Rudolf, U. Schneider, S. Curtis, D. B. an A Gruber, J. Susskind, and P. Arkin (2003), The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present), *J. Hydrometeor.*, *4*, 1147–1167.

Ajayamohan, R., M. Merryfield, and V. Kharin (2010), Increasing Trend of Synoptic Activity and Its Relationship with Extreme Rain Events over Central India, *Journal of Climate*, doi:10.1175/2009JCLI2918.1.

Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. K. Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. R. K. andJ Revadekar, G. Griffiths, L. Vincent, D. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. Vasquez-Aguirre (2006), Global observed changes in daily climate extremes of temperature and precipitation, *JGR*, doi:10.1029/2005JD006290.

Allen, M., and W. Ingram (2002), Constraints on future changes in climate and the hydrologic cycle, *Nature*, doi:10.1038/nature01092.

Banerjee, A., and L. Iyer (2008), History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India, *The American Economic Review*, *95*.

Bollasina, M., Y. Ming, and V. Ramaswamy (2011), Anthropogenic Aerosols and the Weakening of the South Asian Summer Monsoon, *Science*, *334*.

Brown, J. E. (2006), An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions, *Remote Sensing of the Environment*, *101*.

CNN (2007), Millions flee 'worst ever' floods, *CNN*.

Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, R. Dunn, K. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. K. Tank, A. Kruger, J. Marengo, T. Peterson, M. Renom, C. O. Rojas, M. Rusticucci, J. Salinger, A. Elrayah, S. Sekele, A. Srivastava, B. Trewin, C. Villarroel, L. Vincent, P. Zhai, X. Zhang, and S. Kitching (2013), Consistency of Temperature and Precipitation Extremes across Various Global Gridded In Situ and Reanalysis Datasets, *JGR*, doi:10.1002/jgrd.50150, 2013.

Drèze, J., and A. Sen (1989), A Case-Study: The Maharashtra Drought of 1970-1973, in *Hunger and Public Action*, chap. 8, pp. 126–133, Oxford University Press, Oxford.

Efron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall/CRC, New York.

Ghosh, S., D. Das, S.-C. Kao, and A. Ganguly (2012), Lack of uniform trends but increasing spatial variability in observed Indian rainfall extremes, *Nature Climate Change*, *2*, doi:10.1038/NCLIMATE1327.

Goswami, B., V. Venugopal, D. Sengupta, M. Madhusoodanan, and P. Xavier (2006), Increasing Trend of Extreme Rain Events Over India in a Warming Environment, *Nature*, *314*.

Guhathakurta, P., and M. Rajeevan (2008), Trends in the rainfall pattern over India, *International Journal of Climatology*, *28*, doi:10.1002/joc.1640.

Harris, I., P. Jones, T. Osborn, and D. Lister (2014), Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, *International Journal of Climatology*, doi:10.1002/joc.3711.

Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei (2008), Comparison of six methods for the interpolation of daily European climate data, *Journal of Geophysical Research*, *113*.

Hofstra, N., M. New, and C. McSweeney (2010), The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data, *Climate Dynamics*, *35*.

Hossain, N. (2018), The 1970 Bhola cyclone, nationalist politics, and the subsistence crisis contract in Bangladesh, *Disasters*, *42*, doi:10.1111/disa.12235.

Huffman, G., R. Adler, D. Bolvin, G. Gu, E. Nelkin, K. Bowman, Y. Hong, E. Stocker, and D. Wolff (2007), The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, *Journal of Hydrometeorology*, *8*, doi:10.1175/JHM560.1.

India Meteorological Department (2007), Report on Cyclonic Disturbances Over North Indian Ocean During 2006, *RSMC-TROPICAL CYCLONES, NEW DELHI*.

India Meteorological Department (2008), Monsoon 2007: A Report.

Jayasankar, C., S. Surendran, and K. Rajendran (2015), Robust signals of future projections of Indian summer monsoon rainfall by IPCC AR5 climate models: Role of seasonal cycle and interannual variability, *Geophysical Research Letters*, *42*(9), 3513–3520.

King, A., L. Alexander, and M. Donat (2013), The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia, *International Journal of Climatology*, *33*, doi:10.1002/joc.3588.

Krishnamurthy, C., U. Lall, and H.-H. Kwan (2009), Changing Frequency and Intensity of Rainfall Extremes over India from 1951 to 2003, *Journal of Climate*, *22*, doi: 10.1175/2009JCLI2896.1.

Libertino, A., A. Sharma, V. Lakshmi, and P. Claps (2016), A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design, *Environmental Research Letters*, *11*, doi:10.1088/1748-9326/11/5/054003.

Malik, N., B. Bookhagen, and P. Mucha (2016), Spatiotemporal patterns and trends of Indian monsoonal rainfall extremes, *Geophysical Research Letters*, *43*, doi: 10.1002/2016GL067841.

Mukherjee, S., S. Aadhar, D. Stone, and V. Mishra (2018), Increase in extreme precipitation events under anthropogenic warming in India, *Weather and Climate Extremes*.

O'Gorman, P., and T. Schneider (2009), Scaling of precipitation extremes over a wide range of climates simulated with an idealized GCM, *Journal of Climate*, *22*, doi: 10.1175/2009JCLI2701.1.

Pai, D., L. Sridhar, M. Rajeevan, O. Sreejith, N. Satbhai, and B. Mukhopadhyay (2014), Development of a new high spatial resolution ($0.25° \times 0.25°$) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region, *Mausam*, *65*.

Pai, D., L. Sridhar, M. Badwaik, and M. Rajeevan (2015), Analysis of the daily rainfall events over India using a new long period (19012010) high resolution ($0.25° \times 0.25°$) gridded rainfall data set, *Climate Dynamics*, *45*, 755776, doi:10.1007/s00382-014-2307-1.

Rajeevan, M., J. Bhate, J. Kale, and B. Lal (2005), Met. Monograph Climatology: Development of a high-resolution daily gridded rainfall data for the Indian Region, *Government of India, India Meteorological Department*, (22).

Rajeevan, M., J. Bhate, J. Kale, and B. Lal (2006), High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells, *Current Science*, *91*(3).

Rajeevan, M., J. Bhate, and A. Jaswal (2008), Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data, *Geophysical Research Letters*, *35*, doi:10.1029/2008GL035143.

Roxy, M., S. Ghosh, A. Pathak, R. Athulya, M. Mujumdar, R. Murtugudde, P. Terray, and M. Rajeevan (2017), A threefold rise in widespread extreme rain events over central India, *Nature Communications*, *8*, doi:10.1038/s41467-017-00744-9.

Sharma, G., and Z. Hussain (2017), Aid workers struggle as South Asia floods affect more than 16 million, *Reuters*.

Shepard, D. (1968), A two-dimensional interpolation function for irregularly spaced data, *ACM '68 Proceedings of the 1968 23rd ACM national conference*, pp. 517–524.

Singh, D., M. Tsiang, B. Rajaratnam, and N. S. Diffenbaugh (2014), Observed changes in extreme wet and dry spells during the South Asian summer monsoon season, *Nature Climate Change*, *4*, doi:10.1038/NCLIMATE2208.

Turner, A., and H. Annamalai (2012), Climate change and the South Asian summer monsoon, *Nature Climate Change*, doi:10.1038/NCLIMATE1495.

Vecchi, G., and T. Knutson (2010), Estimating Annual Numbers of Atlantic Hurricanes Missing from the HURDAT Database (18781965) Using Ship Track Density, *Journal of Climate*, doi:10.1175/2010JCLI3810.1.

Vinnarasi, R., and C. Dhanya (2016), Changing characteristics of extreme wet and dry spells of Indian monsoon rainfall, *Journal of Geophysical Research: Atmospheres*, *121*, doi:10.1002/2015JD024310.

Walker, G. (1923), Correlation in seasonal variations of weather, VIII : a preliminary study of world weather, *India Meteorological Office, Poona*.

Yatagai, A., K. Yamaguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh (2012), APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges, *Bulletin of the American Meteorological Society*, doi:http://dx.doi.org/10.1175/BAMS-D-11-00122.1.

Table 1. Trends and p values for null hypothesis of no trend in extreme rainfall events estimated from $0.25° \times 0.25°$ IMD gridded precipitation data, using 10^4 surrogate time-series.

	100 mm day^{-1}		150 mm day^{-1}	
Time period	events yr^{-1}	p	events yr^{-1}	p
1951 - 2016	6.14	< 0.01	2.88	< 0.01
1951 - 2000	6.11	< 0.01	2.76	< 0.01
1973 - 2016	4.67	0.17	2.12	0.14
1974 - 2016	5.96	0.09	2.44	0.11
1975 - 2016	5.40	0.15	2.40	0.13
1976 - 2016	5.65	0.14	2.71	0.11

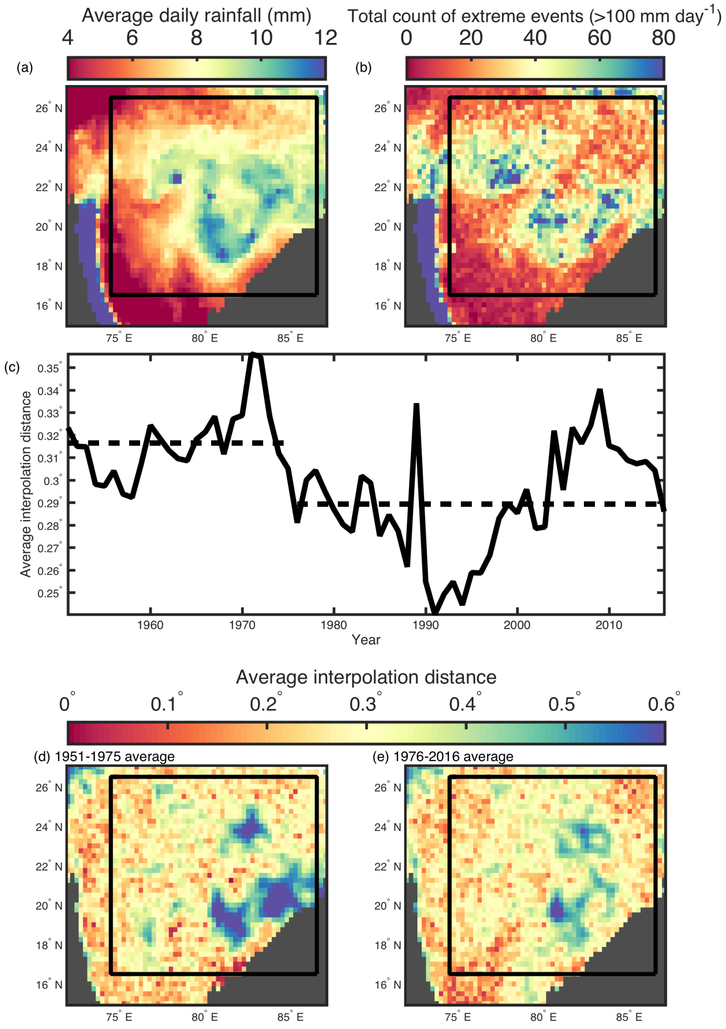

Figure 1. 1951-2016 rainfall climatology and distances to nearest stations. Geography of **(a)** the daily-average monsoon rainfall (June-September) and **(b)** the count of extreme events exceeding 100 mm day^{-1} over 1951-2016. **(c)** A breakpoint analysis identifies 1975 as the largest change in the average interpolation distance from each gridbox center to up to the nearest four stations over Central India. Geography of average interpolation distances for 1951-1975 **(d)** and 1976-2016 **(e)**.

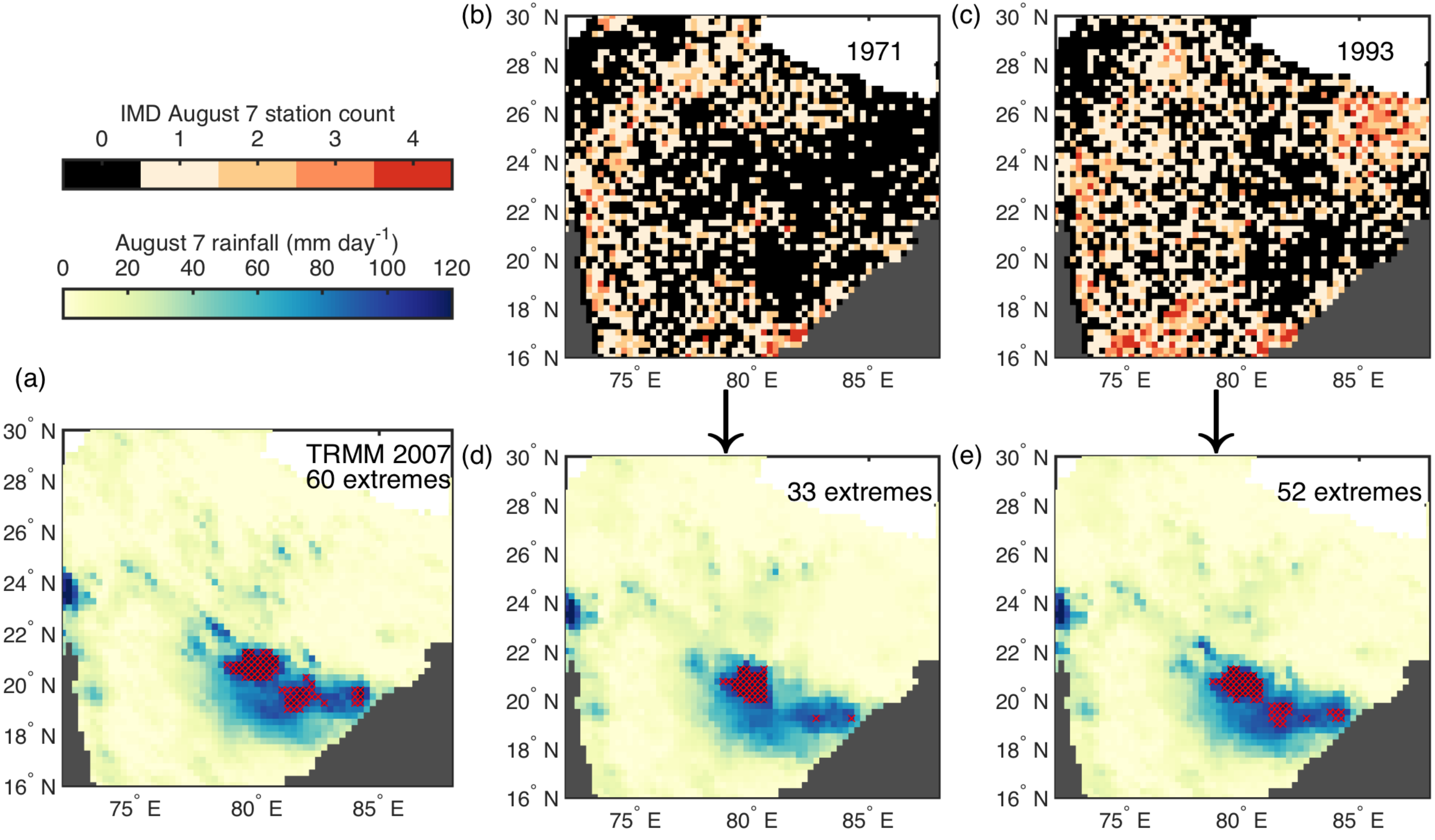

Figure 2. **Interpolation example of TRMM rainfall according to the IMD station network.** The August 7, 2007 rainfall estimate from TRMM **(a)** is interpolated according to the IMD station network on August 7, 1971 **(b)** and August 7, 1993 **(c)**. While the original TRMM estimate shows 60 gridboxes with rainfall greater than 100 mm day^{-1}, interpolating values at gridboxes without stations leads to 33 for the 1971 station distribution **(d)** and 52 for the 1993 distribution **(e)**. Red hatching in a, d, and e mark gridboxes with greater than 100 mm day^{-1} of precipitation.

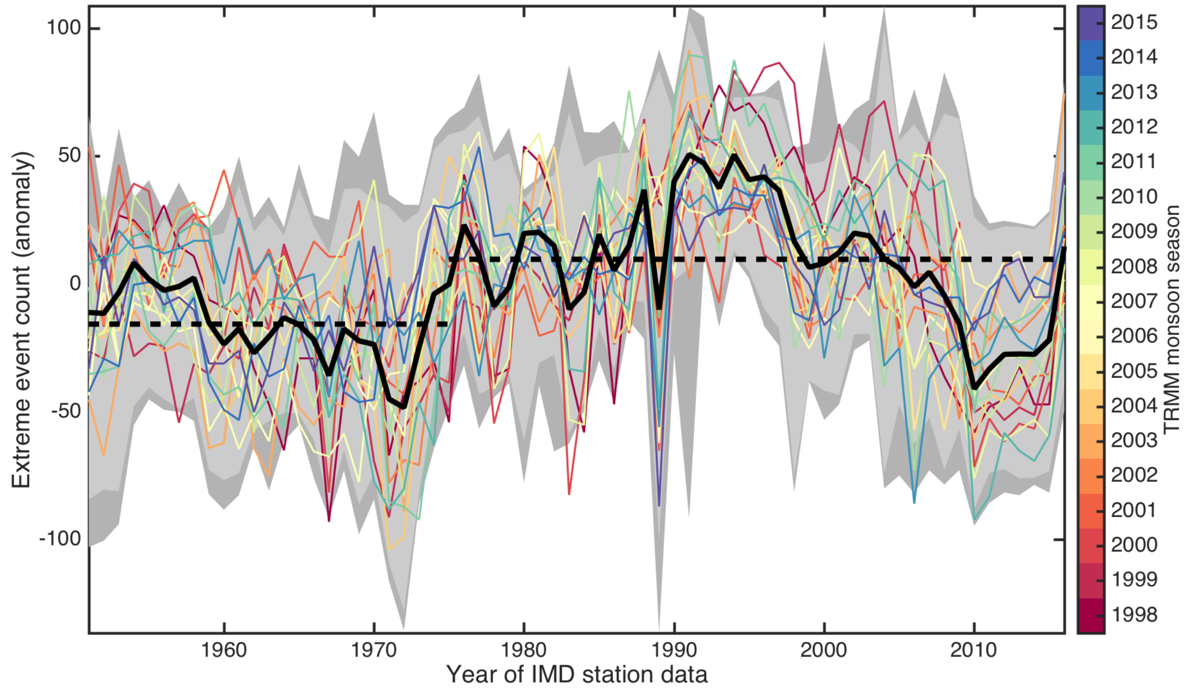

Figure 3. **Station network variability effects on extreme rainfall counts:** Counts of extreme rainfall events exceeding the 99.6$^{\mathrm{th}}$ percentile for the June-September of a Central India TRMM monson season are assessed from that TRMM season interpolated according to the 1951-2016 IMD station network. Each of the 18 years of 1998-2015 TRMM data yields a 1951-2016 time-series of extreme event counts. Each time-series is expessed as an anomaly from its mean, with the average anomaly indicated by the solid black line. A breakpoint is estimated at 1975, with the mean anomalies before and after indicated (horizontal dashed lines). The distribution of extreme event counts when re-interpolating IMD data, as opposed to interpolating TRMM data, is also shown in terms of the 95% interval (light gray shading) and the full range (dark gray shading).

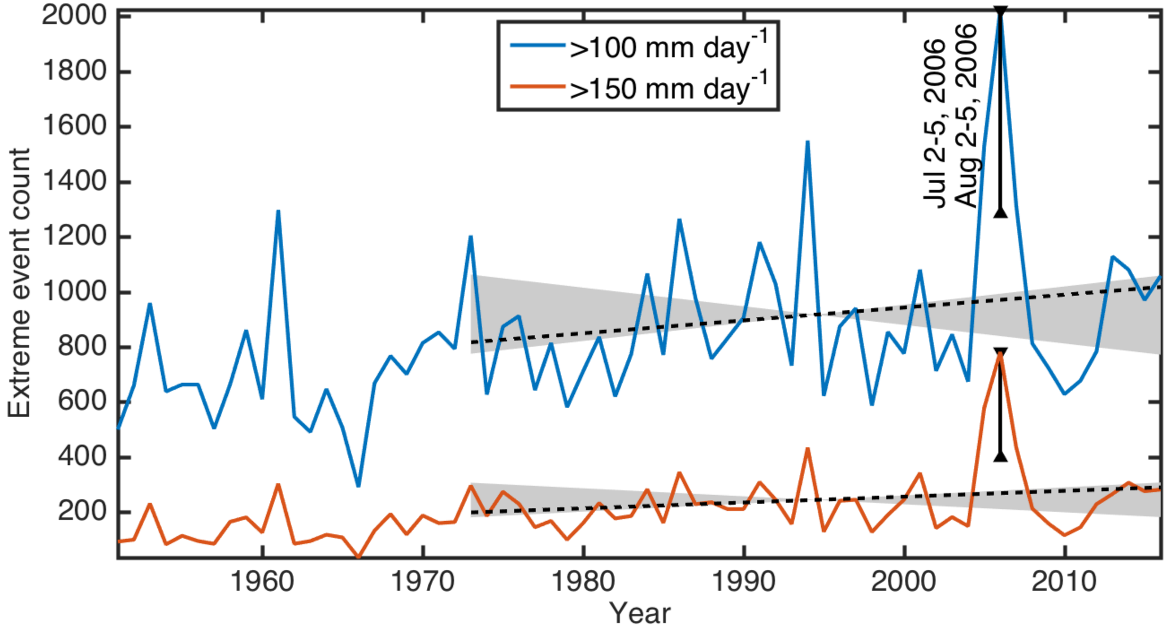

Figure 4. Monsoon extreme precipitation counts over Central India: Counts of days with greater than 100 mm precipitation over June-September summed across Central India at $0.25° \times 0.25°$ resolution (blue), with corresponding counts for rainfall exceeding 150 mm in a day (red). Least-squares trends of 4.67 events per year and 2.12 events per year (dashed lines) are, respectively, fit over 1973-2016. Each trend is within the 95% confidence interval for a zero trend (gray shading). Two monsoon deep depressions on July 2-5 and August 2-5 of 2006 account for 742 and 383 events, respectively, for the 100 mm day^{-1} and 150 mm day^{-1} thresholds (vertical black bars).