A New Parameter to Assess Impact of Rain Gauge Density on Uncertainty in the Estimate of Monthly Rainfall over India

SATYA PRAKASH

Divecha Centre for Climate Change, Indian Institute of Science, Bengaluru, India

ASHWIN SESHADRI

*Divecha Centre for Climate Change, and Centre for Atmospheric and Oceanic Sciences,
Indian Institute of Science, Bengaluru, India*

J. SRINIVASAN

Divecha Centre for Climate Change, Indian Institute of Science, Bengaluru, India

D. S. PAI

India Meteorological Department, Pune, India

ABSTRACT

Rain gauges are considered the most accurate method to estimate rainfall and are used as the "ground truth" for a wide variety of applications. The spatial density of rain gauges varies substantially and hence influences the accuracy of gridded gauge-based rainfall products. The temporal changes in rain gauge density over a region introduce considerable biases in the historical trends in mean rainfall and its extremes. An estimate of uncertainty in gauge-based rainfall estimates associated with the nonuniform layout and placement pattern of the rain gauge network is vital for national decisions and policy planning in India, which considers a rather tight threshold of rainfall anomaly. This study examines uncertainty in the estimation of monthly mean monsoon rainfall due to variations in gauge density across India. Since not all rain gauges provide measurements perpetually, we consider the ensemble uncertainty in spatial average estimation owing to randomly leaving out rain gauges from the estimate. A recently developed theoretical model shows that the uncertainty in the spatially averaged rainfall is directly proportional to the spatial standard deviation and inversely proportional to the square root of the total number of available gauges. On this basis, a new parameter called the "averaging error factor" has been proposed that identifies the regions with large ensemble uncertainties. Comparison of the theoretical model with Monte Carlo simulations at a monthly time scale using rain gauge observations shows good agreement with each other at all-India and subregional scales. The uncertainty in monthly mean rainfall estimates due to omission of rain gauges is largest for northeast India (~4% uncertainty for omission of 10% gauges) and smallest for central India. Estimates of spatial average rainfall should always be accompanied by a measure of uncertainty, and this paper provides such a measure for gauge-based monthly rainfall estimates. This study can be further extended to determine the minimum number of rain gauges necessary for any given region to estimate rainfall at a certain level of uncertainty.

1. Introduction

An accurate estimate of rainfall at various spatio-temporal resolutions is vital for several applications ranging from hydrology to climatology, and for verifying numerical model outputs. Rainfall is generally measured using rain gauges, radars, and satellites, and each have their respective advantages and limitations (Bell and Kundu 2003; Kucera et al. 2013; Sun et al. 2018). Ground-based rain gauges typically provide the most accurate observations of rainfall for a specific location and are used as reference for calibration and validation of remotely sensed rainfall measurements such as from satellites and weather radars (Collier 1986).

Corresponding author: Satya Prakash, satyaprakash@iisc.ac.in

The density of rain gauges is generally not sufficient everywhere, being limited by various constraints such as accessibility, maintenance, and logistics (Schneider et al. 2014; Kidd et al. 2017).

With the advent of Earth-observation satellites, remote sensing techniques have advanced quite rapidly in the recent three decades, contributing significantly to the monitoring of extreme weather events (Srinivasan and Joshi 2007; Sorooshian et al. 2011; Kucera et al. 2013; Gairola et al. 2014). Ground-based rain gauge observations play a crucial role in evaluating satellite-derived rainfall products as well as in improving satellite retrieval algorithms. Therefore, the density and distribution of gauges over any region of interest are important, and their effects must be carefully considered (Bell and Kundu 2003; Villarini et al. 2008; Piyush et al. 2012; Mishra 2013; Dai et al. 2017).

The determination of the optimal rain gauge density and its design remains challenging to hydrologists and meteorologists due to complexity of the problem. Several factors such as topography, accessibility, cost of installation and maintenance, regional climate, and purpose of the network are crucial for any rain gauge network design. The accuracy of gauge-based rainfall estimates generally reduces with the decrease in rain gauge density. Mishra (2013) studied the error characteristics in daily accumulated rainfall as a function of gauge spacing and reported an increase in absolute error of rainfall measurement from 15% to 64%, when the number of gauges was reduced from 7 to 1 in a $0.5° \times 0.5°$ grid box. However, this study was restricted to a small region of southern peninsular India. Such empirical results are very important but cannot be directly generalized for other regions due to differences in spatial variability of rainfall associated with distinct geography, topography, and nature of rainfall variability. Similarly, the effect of rain gauge network configurations on the spatial mean precipitation has been investigated at catchment scale (e.g., Xu et al. 2013; Girons Lopez et al. 2015). Recently, Dai et al. (2017) proposed a sophisticated method based on principal component analysis to determine the optimal rain gauge density, and also suggested a scheme based on cluster analysis for determination of potential locations for installation of gauges over southwestern England.

In contrast, this paper considers the effects of rain gauge density on uncertainty in monthly estimates of spatial average rainfall in the context of missing observations. The Indian subcontinent contains substantial heterogeneity in terms of geography, topography, and climate, and comes under the influence of two large-scale monsoon systems each year—the southwest and northeast monsoon systems, both of which exhibit considerable spatiotemporal rainfall variability (Gadgil 2003; Rajeevan et al. 2012). The India Meteorological Department (IMD) has a large network of rain gauges across the country. Rainfall observations from these gauges are used for several operational hydrometeorological and climatological applications. In addition, long-term daily gridded rainfall products at different spatial resolutions [e.g., 1° latitude–longitude (Rajeevan et al. 2006, 2008, 2010), 0.5° latitude–longitude (Rajeevan and Bhate 2009), 0.25° latitude–longitude (Pai et al. 2014)] have been developed using these gauge observations. These gridded datasets were developed by using an inverse distance weighted interpolation scheme with weighted average of observations (Shepard 1968). These gauge-based gridded products are widely used as reference for the evaluation of both satellite-based rainfall estimates and numerical model simulations over India (Mitra et al. 2013; Prakash et al. 2015, 2018). The number of reporting gauges generally varies from day to day (e.g., Fig. 1a). Furthermore, these gauges are not uniformly distributed across the country, or allocated in accordance to spatiotemporal variability of rainfall, and this could influence the quality of these gauge-based gridded products in some regions more than in others.

An inadequate rain gauge density as well as choice of interpolation method could substantially impact the accuracy of gridded rainfall products (Bastin et al. 1984; Morrissey et al. 1995). Recently, Lin and Huybers (2019) replicated the interpolation method, used for the development of IMD gauge-based gridded rainfall datasets, at a multisatellite precipitation product, and reported that temporal changes in the gauge density would bias the historical trends in the southwest monsoon rainfall extremes over India. They also noticed that the use of a variable network of gauge stations would introduce bias in the mean monsoon rainfall. Moreover, rain gauge observations from the IMD network are used for several socioenvironmental purposes and for the national decisions. For instance, IMD rain gauge observations are utilized for the assessment of meteorological drought, and all-India drought is declared when the monsoon rainfall is below a rather tight threshold of 10% of the long-period mean rainfall. If the uncertainty in the gauge-based mean rainfall estimates is too high, a false deficit decision could be made. Hence, it is important to know the associated uncertainties in gauge-based rainfall estimates owing to the nonuniform layout and placement pattern of the rain gauge network. Reliable estimates of uncertainty in observational rainfall datasets are also critical for evaluating any numerical model output and satellite-derived precipitation product for agriculture,
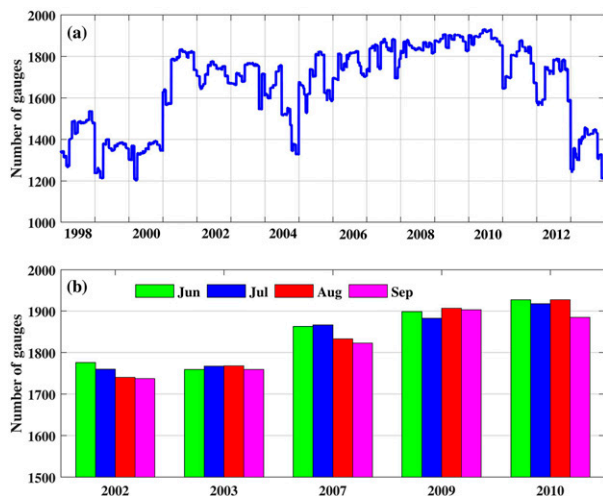
FIG. 1. (a) Daily variation in the number of IMD rain gauge observations from the fixed network over India for the period 1998–2013. (b) Variation in the number of IMD rain gauge observations from the fixed network over India for the selected southwest monsoon months.

water resources management, and precipitation-induced natural hazards preparedness (Krajewski 2007; Collins et al. 2013).

In view of the above, the key objectives of this study are to address the following questions:

1) What is the uncertainty in the spatially averaged estimates of monthly rainfall over a region due to temporal changes in rain gauge density?
2) Which regions need augmentation of the rain gauge network to attain similar uncertainties in gridded rainfall data over the entire country?

For addressing these issues, we consider the impact of gauge density on statistics of monthly mean monsoon rainfall at four homogeneous regions within the country in addition to the all-India scale. It is also important to develop a theoretical model for the uncertainty in estimating mean rainfall that can be applied to different situations. A recently developed theoretical model of uncertainty in estimating the spatial average owing to missing rain gauges (Seshadri 2018) has been applied, and sensitivity analyses have been reported based on rain gauge observations across the country during each month of the southwest monsoon season. The theoretical model is also verified with Monte Carlo simulations using rain gauge observations at a monthly time scale. Section 2 describes the dataset and method of analysis. Section 3 provides a brief description of the theoretical model. Results are presented and discussed in section 4, and major conclusions of this study are outlined in section 5.

2. Data and methods

We obtained daily rainfall observations from a fixed rain gauge network of IMD. A total of 2140 rain gauge stations across India were considered under this fixed gauge network and observations were used for the development of 1° latitude–longitude gridded long-term rainfall data (Rajeevan et al. 2010). The quality-controlled gauge observations for five southwest monsoon seasons (e.g., June to September months) of 2002, 2003, 2007, 2009, and 2010 were chosen for the analysis. According to the IMD monsoon reports, all-India seasonal monsoon rainfall was deficit during 2002 and 2009, whereas it was above normal during 2007 and normal during 2003 and 2010 with respect to long period average of the monsoon rainfall.

The analyses in this paper were performed using monthly means of daily rainfall from this gauge network. The variation in the rain gauge density is less than 10% across the nation for these selected 20 months (Fig. 1b). The spatial distribution of the quality-controlled rain gauges over India for June 2007 is shown in Fig. 2. The distribution of gauges is not uniform across the country, and the density of the gauge network is highest in southern peninsular India and lowest over northeast India. The boundaries of four selected relatively homogeneous regions of India are also shown in Fig. 2. These four homogeneous regions have distinct monsoon rainfall characteristics in terms of mean rainfall and its spatial variability.

The arithmetic means of monthly monsoon rainfall for the selected years are presented in Fig. 3 for all-India and four homogeneous regions. It can be seen from this figure that the deficit in all-India seasonal rainfall during 2002 was primarily due to a rainfall deficit in July, whereas the deficit in all-India seasonal rainfall during 2009 was due to a rainfall deficit in June and August. Even though northeast India receives higher rainfall than other three homogeneous regions during the monsoon season, its year-to-year variability is out of phase with that of rainfall over other parts of the country.

To assess the impact of rain gauge density on uncertainty in the estimates of spatial average rainfall, domain-mean monthly rainfall was computed for these homogeneous regions as well as for all-India after omitting individual rain gauges with a given probability. The choice of omitting individual rain gauges was based on generating statistically independent and identically distributed Bernoulli random variables. If the random variable corresponding to a gauge was equal to zero, the corresponding gauge was omitted from calculation. The Monte Carlo simulations involved the following steps:
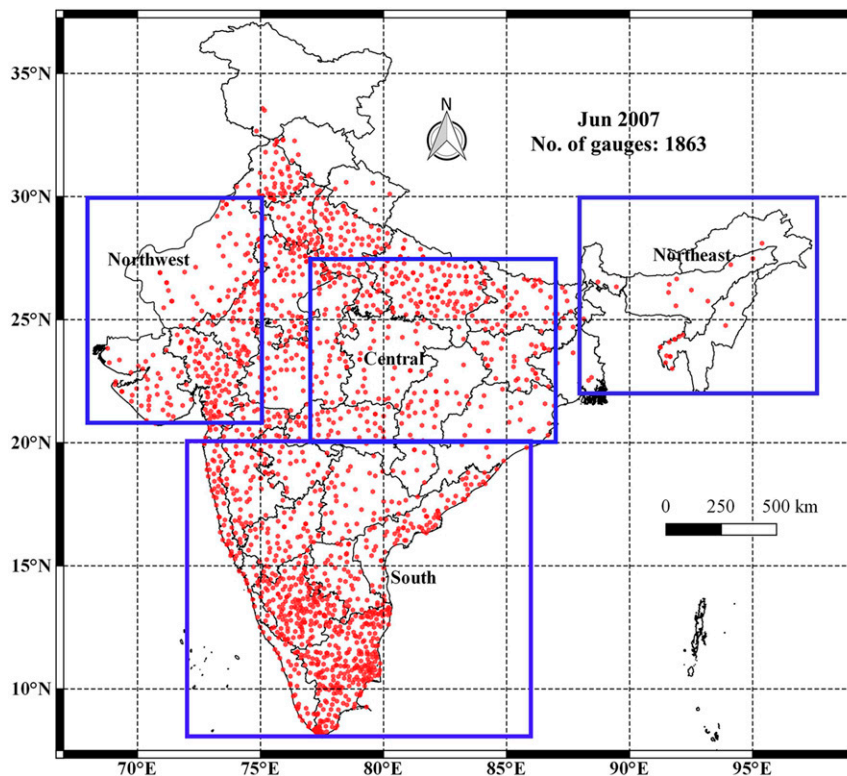
FIG. 2. Spatial distributions of IMD rain gauges over India in June 2007. Four homogenous
regions within India considered for the analysis are shown by rectangular boxes.

1) Generate Bernoulli random variables with values
being either zero or one for each rain gauge, with f
being the probability of being zero. Only those rain
gauges for which the corresponding variable equals
one were considered for the calculation of spatial
average monthly rainfall.

2) Estimate the spatial average of the included gauges,
which were decided based on random variables gen-
erated in step 1.

3) Repeat steps 1 and 2 for an ensemble having 5000
different realizations, with each realization taking a
new set of the Bernoulli random variables indicat-
ing whether each gauge was present or not in the
computation of spatial average monthly rainfall.
Estimate mean rainfall for each realization.

4) Modify the probabilities of the Bernoulli random
variables taking the values of one and zero, and
repeat steps 1–3. This corresponds to the fraction of
omitted gauges, as described below.

For each realization in an ensemble, we estimated
mean rainfall and then computed statistics (e.g., mean
and standard deviation) across the ensemble. Differ-
ent ensembles were generated, one for each fraction
of omitted rain gauges. The probability of individual

gauges being included in the estimate of spatial average
for each realization was approximately equal to the
fraction of omitted gauges owing to the law of large
numbers. This result in statistics states that the sample
average of a collection of independent and identically
distributed random variables converges to the expected
value as the number of variables increases. The fraction
of omitted gauges in any given realization was sample
average of the Bernoulli random variables, whereas
the probability of gauges being included equals its ex-
pected value.

For a given ensemble, corresponding to a given
probability of omitted rain gauges, the uncertainty in
domain-mean rainfall was measured by standard de-
viation of the mean rainfall estimated for each of the
5000 realizations, normalized by mean rainfall calcu-
lated using all the gauges and expressed as percentage.
For regions where this uncertainty is large, not only do
unreported gauge measurements have a larger impact
but also augmenting the rain gauge network in that re-
gion would be more beneficial. An ensemble size of 5000
has been chosen for the Monte Carlo simulations in this
study. The differences in ensemble mean approach
closely to zero for 5000 or more simulations, when uni-
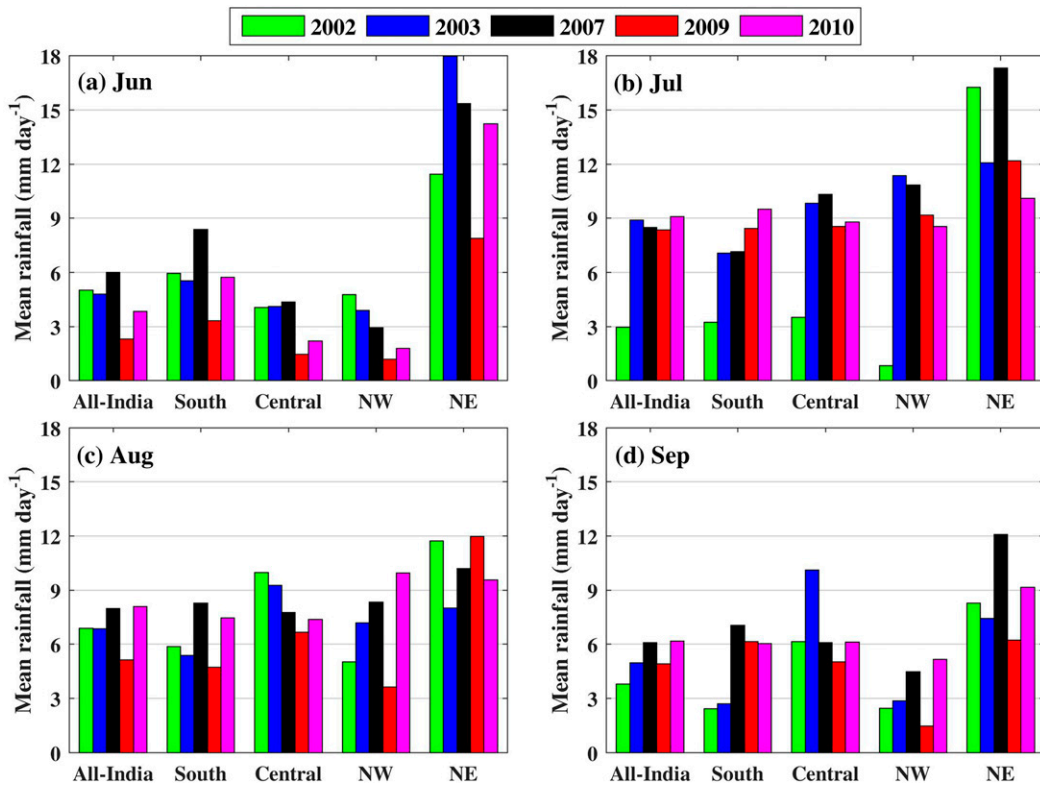form weights were chosen. This is close to the theoretical

FIG. 3. Mean monthly rainfall over India and four homogeneous regions for (a) June, (b) July, (c) August, and (d) September months of the selected years.

mean of zero for uniform weights, as discussed in the next section, and hence this choice was made for the ensemble size.

3. Theoretical model to account for effects of omitting rain gauges

We have used a statistical model proposed by Seshadri (2018) to assess the effects of changing rain gauge density on ensemble mean and standard deviation. He estimated variance and bias in the estimate of a spatially averaged quantity in the presence of missing data. The average was calculated only over those observations which were reported. There were four contributions to variance of the spatial average in the presence of missing data: 1) temporal variance, 2) temporal covariance between locations, 3) measurement error, and 4) effects of missing data itself. Assuming perfect measurements, using only a subset of available rain gauges is analogous to the missing data problem, and the variance in the spatial average owing to omitting a certain fraction of randomly chosen rain gauges can be deduced from the results in Seshadri (2018) as

$$\text{Var} = \frac{f}{1-f} \begin{bmatrix} \beta_1 & \cdots & \beta_N \end{bmatrix}$$

$$\times \begin{bmatrix} (\mu_r - r_1)^2 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & (\mu_r - r_N)^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}, \quad (1)$$

where f is the probability of any individual rain gauge being absent from the calculation (and also corresponding to the expected fraction of omitted gauges), $\beta_1, \beta_2, \ldots, \beta_N$ are the weights given to each of the N available gauges that could be used, $r_1, r_2, \ldots, r_N$ are the corresponding rainfall observations, and $\mu_r = \sum_{i=1}^{N} \beta_i r_i$ is the weighted average.

If we do not assume perfect measurements, then we must take into account the contribution of measurement error to the variance, which is $[f/(1-f)]\sigma_\varepsilon^2 \sum_{i=1}^{N} \beta_i^2$, where σ_ε^2 is the measurement error variance from each individual rain gauge, and the gauges have been assumed to be identical in this aspect with their measurement errors being statistically independent (Seshadri 2018). Since the interpolation scheme (e.g., Shepard 1968) used for the development of gridded gauge-based rainfall products involves linear weights of rain gauges that sum to

one, Eq. (1) can be used to estimate ensemble standard deviation, and hence uncertainty in the resulting gridded rainfall products.

The weights must sum to one, in the case where all rain gauges are used, so that in case of uniform weights involving simply the arithmetic average of the rain gauges, these weights are $\beta_i = 1/N$ and weighted average becomes $\mu_r = (1/N)\sum_{i=1}^{N} r_i = r'$, the arithmetic mean of all rain gauge observations. Then Eq. (1) simplifies to

$$\text{Var} = \frac{1}{N^2}\frac{f}{1-f}\sum_{i=1}^{N}(r_i - r')^2, \qquad (2)$$

and using $(1/N)\sum_{i=1}^{N}(r_i - r')^2 = \sigma_s^2$, the spatial variance estimated using all the rain gauges, the ensemble standard deviation of estimation due to omitting fraction f of rain gauges is

$$\sigma_f = \sqrt{\text{Var}} = \sqrt{\frac{1}{N}\frac{f}{1-f}}\sigma_s, \qquad (3)$$

which increases proportionally to spatial standard deviation σ_s. For a given period of time, the spatial standard deviation is assumed to be known for a specific region and estimated using all available rain gauge observations. Therefore, changing the omitted fraction of gauges changes only the factor $\sqrt{f/(1-f)}$. For small f, we can approximate using first-order Taylor series $1/(1-f) \cong 1+f$ so that, to first order in f, $\sigma_f \cong \sqrt{(1/N)f}\sigma_s$, and the ensemble standard deviation increases with square root of the omitted fraction of rain gauges. Because of proportionality in spatial standard deviation σ_s within the region, the uncertainty in measuring spatially averaged rainfall σ_f is larger where this quantity is higher. This effect can be compensated by higher density of coverage owing to the dependence on square root of N in Eq. (3).

For uniform weights with $\beta_i = 1/N$, the contribution of measurement error to the variance becomes $[f/(1-f)]\sigma_\varepsilon^2\sum_{i=1}^{N}(1/N)^2 = (1/N)[f/(1-f)]\sigma_\varepsilon^2$. Combining the results with Eq. (3), total variance in the presence of measurement error becomes

$$\sigma_f^2 = \frac{1}{N}\frac{f}{1-f}(\sigma_s^2 + \sigma_\varepsilon^2). \qquad (4)$$

The contribution of measurement error can be neglected if the error variance from each individual gauge is much lower than the spatial variance over the region, that is, for $\sigma_\varepsilon^2 \ll \sigma_s^2$. This is generally the case for the regions considered in this study, and hence we have omitted the term arising from measurement error in our calculations.

Omitting rain gauges also contributes to bias in estimates of the spatial average. However, for the special case of arithmetic averaging this bias is zero. From Seshadri (2018), the contribution of missing or omitted observations to bias in estimating the spatial average is given by

$$\text{bias} = \frac{f}{1-f}\left(\mu_r\sum_{i=1}^{N}\beta_i^2 - \sum_{i=1}^{N}\beta_i^2 r_i\right), \qquad (5)$$

with $\mu_r = \sum_{i=1}^{N}\beta_i r_i$ as noted previously, and with $\beta_i = 1/N$, we obtain

$$\text{bias} = \frac{f}{1-f}\left(\frac{1}{N^2}\sum_{i=1}^{N}r_i - \sum_{i=1}^{n}\frac{1}{N^2}r_i\right) = 0, \qquad (6)$$

so that with arithmetic averaging, the bias vanishes. This property is unique to arithmetic averaging schemes, and in case of more general nonuniform weights, the bias due to missing data does not necessarily vanish. However, missing observations do not alter the ensemble mean with the use of arithmetic averaging to estimate the spatial average. In other words, the bias associated with leaving some of the gauges would theoretically be zero as all gauges are treated equally.

4. Results and discussion

For demonstrating the relevance and implications of this type of analysis, we first considered gauge-based daily mean rainfall for the month of June 2007 over a $1.5° \times 1.5°$ grid box bounded by 20°–21.5°N and 75°–76.5°E. This grid box has 22 gauge observations for this period (Fig. 4), and is located in the northern plains of India (or central India) where the spatial variability of the seasonal monsoon rainfall is rather small (Gadgil 2003). The domain-mean daily rainfall estimates over the month following random omission of successive gauges are presented in Fig. 4. Although the ensemble mean rainfall does not vary considerably, the error in daily mean rainfall over the month increases consistently when the number of gauge observations is reduced. For instance, about 10% error in daily mean rainfall (as a fraction of the spatial average) is evident following the omission of only three gauges from this analysis. However, the error increases more rapidly if we choose the grid box over the regions having larger spatial variability of the monsoon rainfall. In general, such errors posed by finite sampling and missing data can be significant for rainfall analysis as well as development of gauge-based gridded rainfall products that are based on averaging rain gauges over small grid boxes, even smaller than 1.5° latitude × 1.5° longitude. The resulting
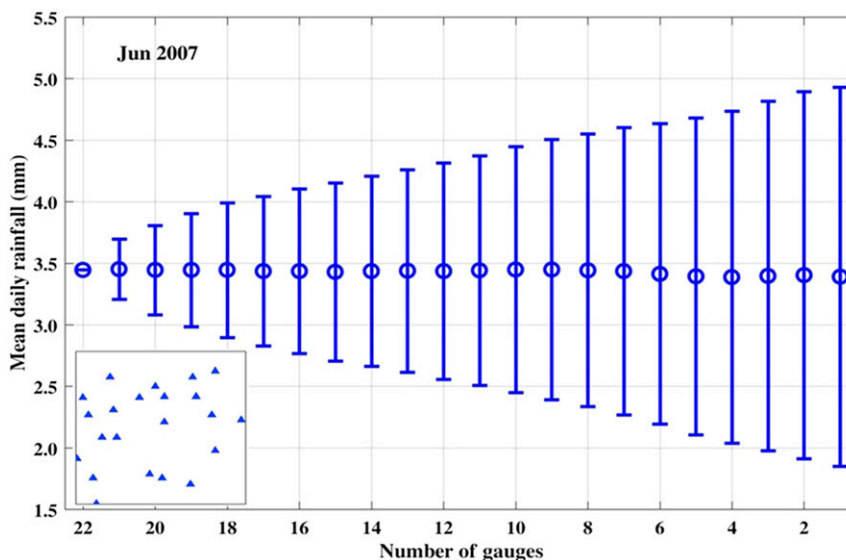
FIG. 4. Mean daily rainfall over a 1.5° × 1.5° grid box in India and associated errors with reduction in randomly chosen gauge observations for June 2007. The distribution of 22 rain gauges over the selected grid box bounded by 20°–21.5°N and 75°–76.5°E is shown at the bottom left of the figure.

uncertainties could be higher over orographic regions because of a larger spatial variation of rainfall. The number of daily gauge observations over India shows large variations even from the fixed network (Fig. 1a), which could have a significant influence on spatial averages. Hence, time series of the spatial average rainfall should be interpreted with caution, especially when estimates of uncertainty have not been provided.

The theoretical development in section 3 suggests that spatial variance and number of gauges across a region are the two main factors controlling uncertainty in estimates of spatial average rainfall. The relevant dimensionless quantity describing effects of changing gauge density is the spatial standard deviation divided by the spatial average rainfall, that is, σ_s/μ_s, termed as the spatial coefficient of variation. We define the "averaging error factor" as $\sigma_s/(\sqrt{N}\mu_s)$, following Eq. (3). This averaging error factor indicates the combined influence of spatial variability and rain gauge density, and a higher magnitude indicates larger uncertainty in the rainfall estimate and suggests the need for more gauge observations in that region. Multiplying the averaging error factor by $\sqrt{f/(1-f)}$, with f being the fraction of omitted gauges, would yield the uncertainty in the spatial average rainfall estimates.

We examined the uncertainty in monthly mean rainfall, after omitting a fraction of gauge observations, using the theoretical model as well as Monte Carlo simulations. The uncertainty in all-India mean rainfall

for each month of the southwest monsoon season is presented in Fig. 5 as a function of the fraction of omitted rain gauges. Monte Carlo simulations described in section 2 are shown together with results from Eq. (3), which explains simulations rather well. As expected, the uncertainty in spatially averaged monthly rainfall increases with the omitted fraction of rain gauges. However, the uncertainty increases more rapidly with omission of gauges during the month of June as compared to other monsoon months. For all-India rainfall, omission of about 10% rain gauges lead to 1% uncertainty in the estimate of the spatial averaged rainfall for June. This is due to large spatial variability in June rainfall associated with high rainfall over the southern peninsular and northeast India than that in central and northwest India (e.g., Fig. 3) associated with successive progress of the southwest monsoon. Similarly, the uncertainty shows anomalous increase with omission in gauges in July during deficit monsoon years of 2002 and 2009 due to relatively larger spatial variability of rainfall. These differences in uncertainties at intraseasonal and interannual scales are primarily due to differences in the spatial variance in rainfall, because the gauge density is almost invariable during these periods. This analysis is repeated with 16 years (1998–2013) of monsoon datasets, and similar characteristics of uncertainty were evident. As compared to uncertainties during the normal monsoon months, dry monsoon months have about 10% more uncertainties and the wet monsoon months have about 15% less uncertainties.
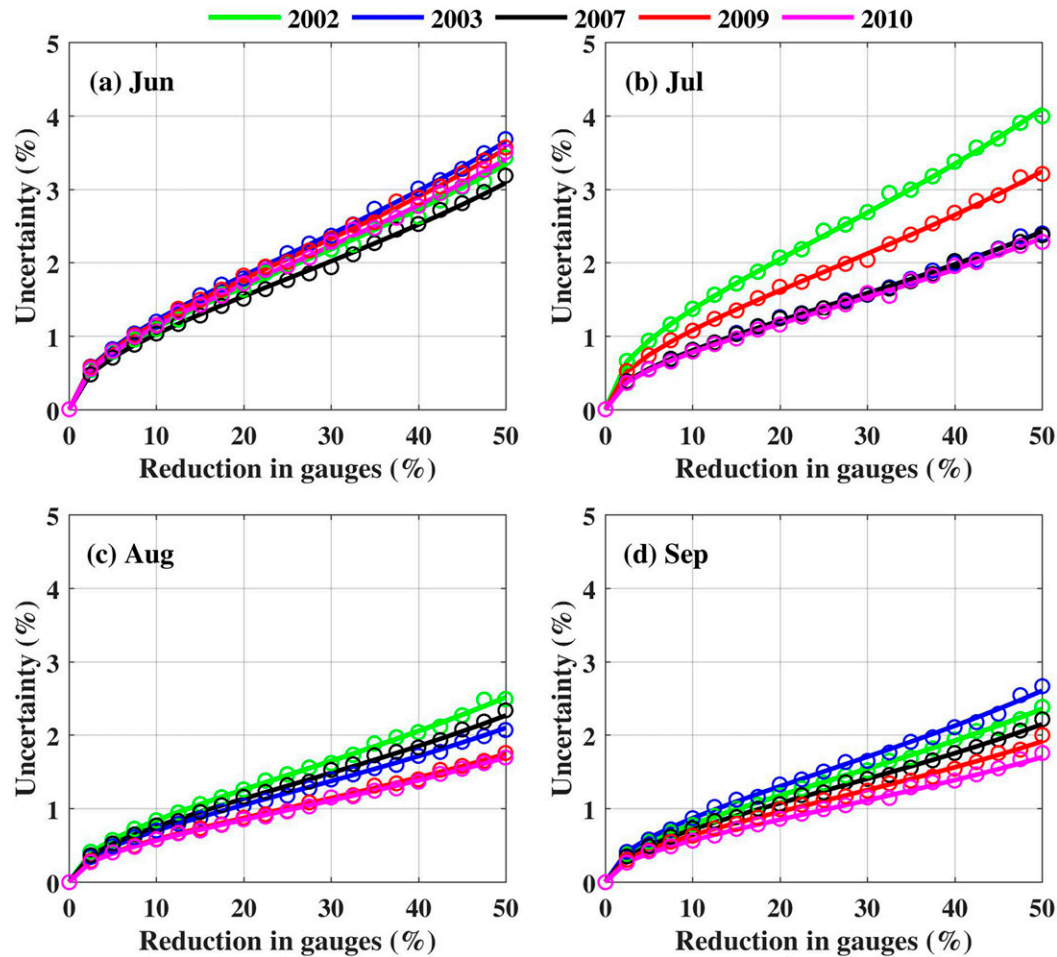
FIG. 5. Uncertainty in mean monthly rainfall over India with reduction in gauge density from the theoretical model (solid lines) and Monte Carlo simulations using gauge observations (open circles) for (a) June, (b) July, (c) August, and (d) September of 2002, 2003, 2007, 2009, and 2010.

It is to be noted that limiting uncertainty in the estimates of all-India mean monthly rainfall to around 1% is important for several applications such as for studying interannual variability of all-India monsoon rainfall adequately. For instance, IMD declares an all-India deficit monsoon if seasonal rainfall across the country is 10% below the long period average, an above normal rainfall if seasonal rainfall is 10% above the long period average, and a normal monsoon rainfall otherwise. For a given year, if the all-India monsoon rainfall is 9% below the long period average as per gauge-based rainfall estimates, to infer this as a normal monsoon year would require the uncertainty to be less than 1% in the all-India rainfall estimate. If the uncertainty in the mean rainfall estimate is too high, a false deficit decision could be made. This suggests a rather tightly bound threshold for an acceptable level of uncertainty. Furthermore, an estimate of uncertainty accompanied with

the spatial mean rainfall is crucial for the verification of precipitation forecasts from numerical models and for the assessment of reliable historical trends in mean rainfall and extremes.

Figure 6 presents the uncertainty in monthly mean rainfall over northeast India for each month of the southwest monsoon season as a function of the fraction of omitted rain gauges. Northeast India shows a larger increase in uncertainty with decrease in gauge observations as compared to all-India. Omission of 10% of rain gauges leads to approximately 4% uncertainty in the estimate of the spatial average rainfall. There are few years between 1998 and 2013 (e.g., 1998, 1999, 2000), when the number of reporting rain gauges decreased by about 40% of the total number of available gauges of the fixed network. Hence, the uncertainty estimates associated with spatial mean rainfall are crucial for its integration in any application. The theoretical and
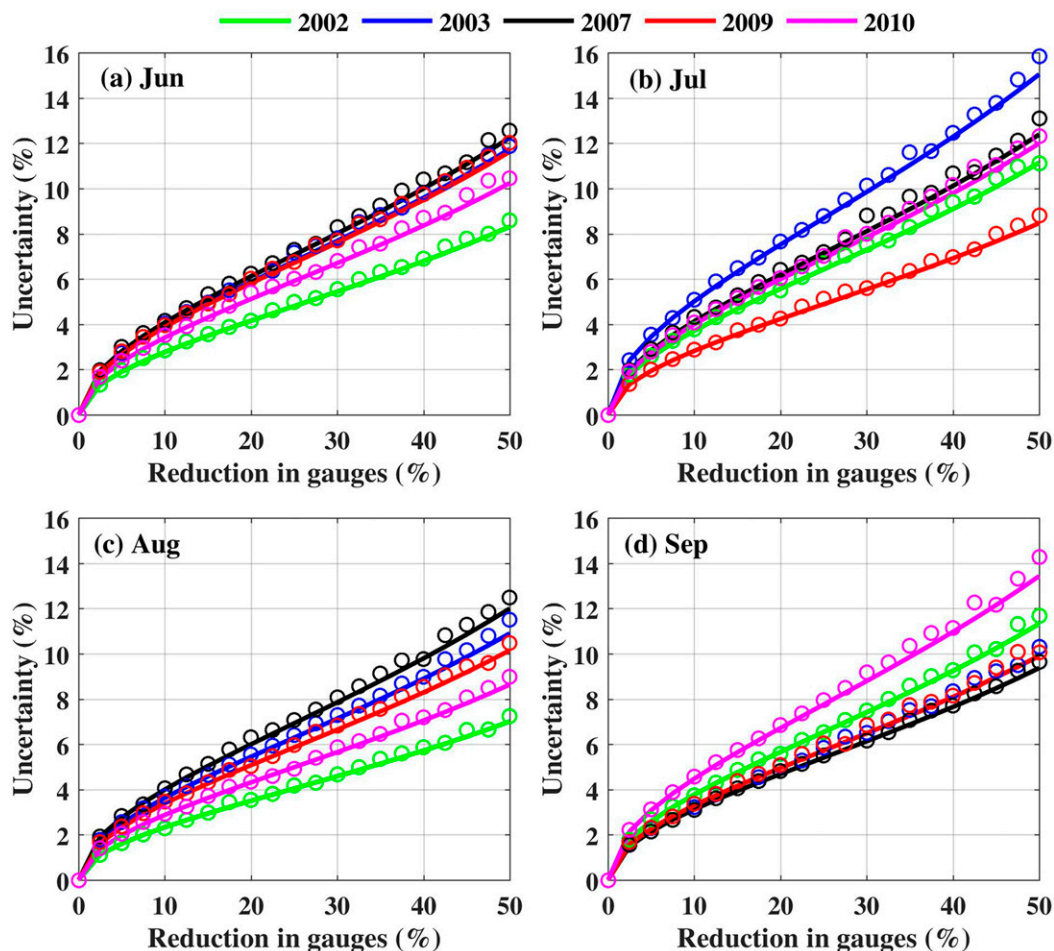
FIG. 6. Uncertainty in mean monthly rainfall over northeast India with reduction in gauge density from the theoretical model (solid lines) and Monte Carlo simulations using gauge observations (open circles) for (a) June, (b) July, (c) August, and (d) September of 2002, 2003, 2007, 2009, and 2010.

simulated uncertainties generally agree well with each other. However, there are some differences between them when the omission of gauges is more than 30% of the total gauges. One of the reasons behind this discrepancy might be a smaller number of gauges over the northeast region (Fig. 2). Additionally, the theoretical model is also an approximation, and variance is derived by truncating the Taylor series of a ratio and considering suitable expectations known as the "delta method" (Seshadri 2018). Generally, the uncertainty from omitting rain gauges would be larger at smaller spatial (e.g., statewise, meteorological subdivisional scale) and temporal (e.g., weekly, daily, and subdaily) scales. It is important to undertake analysis and develop theory for a wide range of relevant scales. Even though four homogeneous regions within India showed similar behavior of increase in uncertainty due to decrease in gauges, the uncertainty in monthly rainfall estimates due to changes in gauge density is the largest for northeast

India and smallest for central India among four homogeneous regions. Southern India showed larger range of uncertainty than northwest India.

Figure 7 illustrates the effects of spatial coefficient of variation of available rain gauges on the averaging error factor for each month of the monsoon season. This analysis would be useful for deciding which regions require augmentation of the rain gauge networks. Although spatial variability of rainfall is rather small over northeast India, it shows the largest averaging error factor due to the smaller number of gauges, and hence the analysis suggests that this region should be given priority for the installation of additional gauges. For instance, about 200 gauges are required over northeast India to attain a 5% averaging error factor at 70% spatial coefficient of variation. In case the fraction of missing gauges is limited to 10% so that $f = 0.10$, the factor $\sqrt{f/(1-f)} = 0.33$, and the uncertainty in the spatial mean with about 200 gauges becomes around
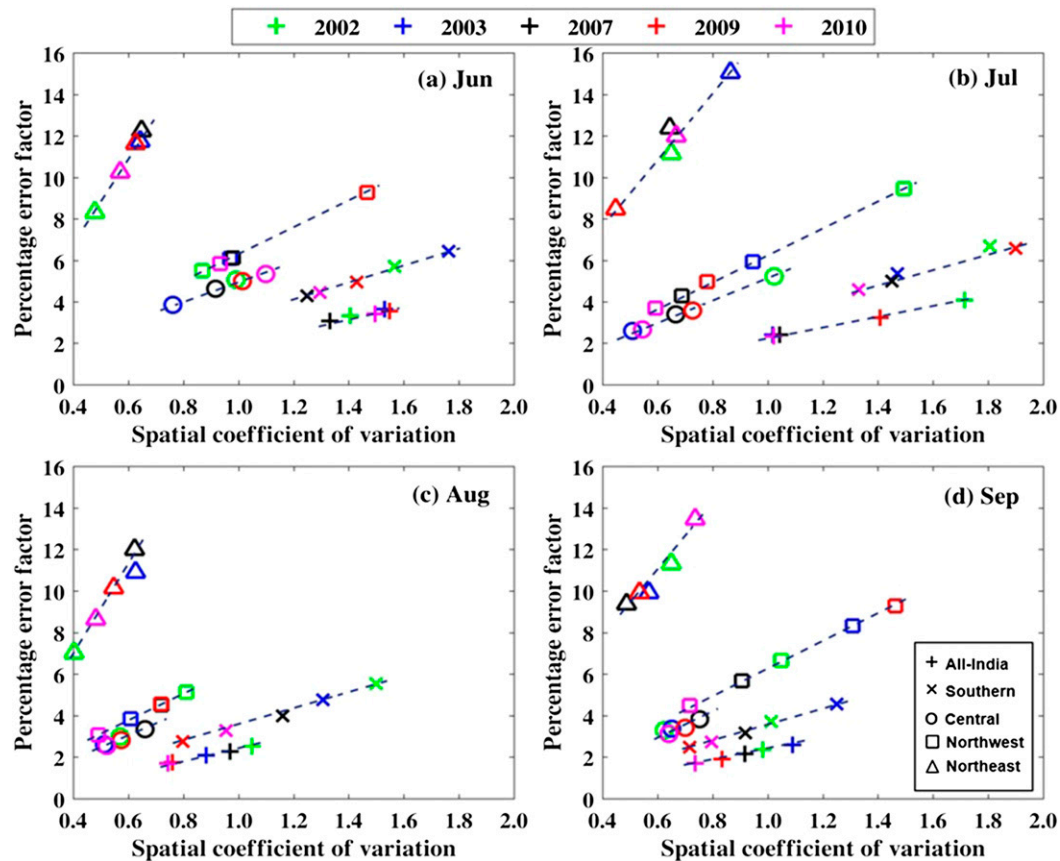
FIG. 7. Percentage averaging error factor $[\sigma_s/(\sqrt{N}\mu_s) \times 100]$ in mean monthly rainfall as a function of spatial coefficient of variation (σ_s/μ_s) for all-India and four homogeneous subregions for (a) June, (b) July, (c) August, and (d) September months of the selected years.

1.7% of the spatial average. However, this estimate of number of required gauges can vary due to variability in the spatial standard deviation of rainfall, and further analysis should account for both interannual variability in the spatial standard deviation and dependence of its estimation on the available rain gauge density. Preliminary analysis suggests that the spatial standard deviation of rainfall is generally smaller than the observed value if the number of gauges increases substantially over a data-sparse region. However, the objective of the present study is not to estimate the optimal number of gauges over India, but to verify the recently proposed theoretical model of uncertainty with Monte Carlo simulations and quantify the uncertainty in monthly mean rainfall due to changes in gauge density.

From Fig. 7, it can be seen that although the number of gauges is larger over southern India than central India, larger spatial variability of rainfall across southern India associated with the complex topography of the Western Ghats leads to a larger averaging error factor in southern India when compared to central India.

Northwest India shows larger spatial variability during July and September, and hence a higher averaging error factor. In general, larger spatial variability of rainfall during the deficit months as compared to normal rainfall months leads to larger averaging error factors. Furthermore, the averaging error factor shows considerable interannual and intraseasonal variations due to remarkable changes in the spatial variability of monsoon rainfall, even while the number of gauges is almost invariable.

Another source of uncertainty in estimates of spatial mean rainfall is the uncertainty in the gauge observations themselves. However, this kind of uncertainty is small or negligible in the case of a good network of rain gauges, especially at larger spatial or temporal scales of integration (Seshadri 2018; Krajewski 2007). As shown in section 3, the contribution of the measurement error to uncertainty in the spatial mean can be neglected if the error variance of individual gauges is small compared to the spatial variance over the region, obeying $\sigma_\varepsilon^2 \ll \sigma_s^2$. Even assuming a conservative error variance

from individual gauges of about $1 \, \text{mm day}^{-1}$, this relation is generally obeyed for the regions that we have considered, and hence our calculations have omitted the contribution of measurement error to uncertainties in the estimates of spatial average. However, the uncertainty from measurement error can become pronounced at smaller spatial scales, for which the spatial standard deviation of spatial rainfall is comparable with the standard deviation of rain gauge errors.

5. Conclusions

The uncertainty in monthly mean monsoon rainfall due to changes in rain gauge density over India was quantitatively assessed based on a theoretical model and Monte Carlo simulations. The theoretical model considered ensemble standard deviation in the spatial average of rainfall estimated using rain gauge observations, if gauges were randomly omitted from the calculation. The theoretical model agreed reasonably well with Monte Carlo simulations using monthly rain gauge observations. The analysis using gauge-based rainfall observations from a fixed network across India revealed an uncertainty of about 4% in monthly mean rainfall over northeast India upon omission of about 10% of the available gauges. Northeast India showed substantial variations in rain gauge density between 1998 and 2013, which could lead to considerable uncertainty in monthly mean rainfall estimates. The increase in uncertainty due to decrease in rain gauges was largest for northeast India and smallest for central India among the four homogeneous regions. Such uncertainty estimates are vital for accurate inference about time series analysis of spatially averaged rainfall. However, there were considerable intraseasonal and interannual variations in uncertainties for all-India and four homogeneous regions.

The large uncertainty in northeast India was mainly due to a small number of rain gauges, whereas the larger spatial variance of the monsoon rainfall was responsible for higher uncertainty in monthly rainfall estimates over southern India. There was a substantial variation in the number of reporting rain gauges across India, and hence it is important to consider the effects of missing rain gauges on the uncertainty in estimates of regional mean rainfall. A new parameter called the ''averaging error factor,'' based on the spatial variance of rainfall in relation to its spatial mean as well as the current number of gauges, was proposed to identify the regions requiring augmentation of rain gauges. The averaging error factor can be large if either the spatial variance is high or the number of rain gauges is small. The analysis showed that northeast India needs augmentation of the rain gauge network (about 200 gauges to attain a 5% averaging error factor at 70% spatial variability) to study the interannual variability of monsoon rainfall adequately by limiting the uncertainty in the spatial average to an acceptable range.

The present study has focused on India due to availability of station data. However, results of this study would be valid for other parts of the globe as well, because the proposed averaging error factor depends primarily on the spatial variability of rainfall and the number of rain gauge observations. The optimal gauge density would vary based on specific applications, due to variability in the spatial statistics at intraseasonal and interannual scales, as well as the spatial and temporal resolutions of interest. With subsequent advancements in remote sensing technologies, several multisatellite rainfall products at finer spatial and temporal resolutions are available to end-users. Gauge-based rainfall observations play a critical role in evaluation of satellite-based estimates. However, the density of gauges and uncertainty in gauge-based rainfall estimates should be cautiously considered along with other topographical and geographical factors for the assessment of gridded rainfall products. As noted earlier, meteorological observations are usually made with the goal of supporting societal and policy decisions. To support decision-making, the statistics of rainfall must be accompanied by uncertainty estimates. While the level of precision that is required can vary by situation, this paper provides an approach for quantifying the uncertainty in rainfall estimates from rain gauge observations. This study can be extended to examine the minimum density of rain gauges essential for estimating rainfall over any given region to a desired level of accuracy.

REFERENCES

Bastin, G., B. Lorent, C. Duque, and M. Gevers, 1984: Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resour. Res.*, **20**, 463–470, https://doi.org/10.1029/WR020i004p00463.

Bell, T. L., and P. K. Kundu, 2003: Comparing satellite rainfall estimates with rain gauge data: Optimal strategies suggested by a spectral model. *J. Geophys. Res.*, **108**, 4121, https://doi.org/10.1029/2002JD002641.

Collier, C. G., 1986: Accuracy of rainfall estimates by radar, part 1: Calibration by telemetering raingauges. *J. Hydrol.*, **83**, 207–223, https://doi.org/10.1016/0022-1694(86)90152-6.

Collins, M., K. AchutaRao, K. Ashok, S. Bhandari, A. K. Mitra, S. Prakash, R. Srivastava, and A. Turner, 2013: Observational challenges in evaluating climate models. *Nat. Climate Change*, **3**, 940–941, https://doi.org/10.1038/nclimate2012.

Dai, Q., M. Bray, L. Zhuo, T. Islam, and D. Han, 2017: A scheme for rain gauge network design based on remotely sensed rainfall measurements. *J. Hydrometeor.*, **18**, 363–379, https://doi.org/10.1175/JHM-D-16-0136.1.

Gadgil, S., 2003: The Indian monsoon and its variability. *Annu. Rev. Earth Planet. Sci.*, **31**, 429–467, https://doi.org/10.1146/annurev.earth.31.100901.141251.

Gairola, R. M., S. Prakash, M. T. Bushair, and P. K. Pal, 2014: Rainfall estimation from Kalpana-1 satellite data over Indian land and oceanic regions. *Curr. Sci.*, **107**, 1275–1282.

Girons Lopez, M., H. Wennerstrom, L. Norden, and J. Seibert, 2015: Location and density of rain gauges for the estimation of spatial varying precipitation. *Geogr. Ann.*, **97A**, 167–179, https://doi.org/10.1111/geoa.12094.

Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, https://doi.org/10.1175/BAMS-D-14-00283.1.

Krajewski, W. F., 2007: Ground networks: Are we doing the right thing? *Measuring Precipitation from Space: EURAINSAT and the Future*, V. Levizzani, P. Pauer, and F. J. Turk, Eds., Springer, 403–417.

Kucera, P. A., E. E. Ebert, F. J. Turk, V. Levizzani, D. Kirschbaum, F. J. Tapiador, A. Loew, and M. Borsche, 2013: Precipitation from space: Advancing earth system science. *Bull. Amer. Meteor. Soc.*, **94**, 365–375, https://doi.org/10.1175/BAMS-D-11-00171.1.

Lin, M., and P. Huybers, 2019: If rain falls in India and no one reports it, are historical trends in monsoon extremes biased? *Geophys. Res. Lett.*, **46**, 1681–1689, https://doi.org/10.1029/2018GL079709.

Mishra, A. K., 2013: Effect of rain gauge density over the accuracy of rainfall: A case study over Bangalore, India. *SpringerPlus*, **2**, 311, https://doi.org/10.1186/2193-1801-2-311.

Mitra, A. K., I. M. Momin, E. N. Rajagopal, S. Basu, M. N. Rajeevan, and T. N. Krishnamurti, 2013: Gridded daily Indian monsoon rainfall for 14 seasons: Merged TRMM and IMD gauge analyzed values. *J. Earth Syst. Sci.*, **122**, 1173–1182, https://doi.org/10.1007/s12040-013-0338-3.

Morrissey, M. L., J. A. Maliekal, J. S. Greebe, and J. Wang, 1995: The uncertainty of simple averages using rain gauge networks. *Water Resour. Res.*, **31**, 2011–2017, https://doi.org/10.1029/95WR01232.

Pai, D. S., L. Sridhar, M. Rajeevan, O. P. Sreejith, N. S. Satbhai, and B. Mukhopadhyay, 2014: Development of a new high spatial resolution (0.25° × 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** (1), 1–18.

Piyush, D. N., A. K. Varma, P. K. Pal, and G. Liu, 2012: An analysis of rainfall measurements over different spatiotemporal scales and potential implications for uncertainty in satellite data validation. *J. Meteor. Soc. Japan*, **90**, 439–448, https://doi.org/10.2151/jmsj.2012-401.

Prakash, S., A. K. Mitra, I. M. Momin, R. M. Gairola, D. S. Pai, E. N. Rajagopal, and S. Basu, 2015: A review of recent evaluations of TRMM Multisatellite Precipitation Analysis (TMPA) research products against ground-based observations over Indian land and oceanic regions. *Mausam*, **66** (3), 355–366.

——, ——, R. M. Gairola, H. Norouzi, and D. S. Pai, 2018: Status of high-resolution multisatellite precipitation products across India. *Remote Sensing of Aerosols, Clouds, and Precipitation*, T. Islam et al., Eds., Elsevier, 301–314, https://doi.org/10.1016/B978-0-12-810437-8.00014-1.

Rajeevan, M., and J. Bhate, 2009: A high resolution daily gridded rainfall dataset (1971–2005) for mesoscale meteorological studies. *Curr. Sci.*, **96** (4), 558–562.

——, ——, J. D. Kale, and B. Lal, 2006: High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells. *Curr. Sci.*, **91** (3), 296–306.

——, ——, and A. K. Jaswal, 2008: Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.*, **35**, L18707, https://doi.org/10.1029/2008GL035143.

——, S. Gadgil, and J. Bhate, 2010: Active and break spells of the Indian summer monsoon. *J. Earth Syst. Sci.*, **119**, 229–247, https://doi.org/10.1007/s12040-010-0019-4.

——, C. K. Unnikrishnan, J. Bhate, K. N. Kumar, and P. P. Sreekala, 2012: Northeast monsoon over India: Variability and prediction. *Meteor. Appl.*, **19**, 226–236, https://doi.org/10.1002/met.1322.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.*, **115**, 15–40, https://doi.org/10.1007/s00704-013-0860-x.

Seshadri, A. K., 2018: Statistics of spatial averages and optimal averaging in the presence of missing data. *Spat. Stat.*, **25**, 1–21, https://doi.org/10.1016/j.spasta.2018.04.002.

Shepard, D., 1968: A two-dimensional interpolation function for irregularly spaced data. *Proc. 23rd ACM National Conf.*, New York, NY, ACM, 517–524, https://doi.org/10.1145/800186.810616.

Sorooshian, S., and Coauthors, 2011: Advanced concepts on remote sensing of precipitation at multiple scales. *Bull. Amer. Meteor. Soc.*, **92**, 1353–1357, https://doi.org/10.1175/2011BAMS3158.1.

Srinivasan, J., and P. C. Joshi, 2007: What have we learned about the Indian monsoon from satellite data? *Curr. Sci.*, **93** (2), 165–172.

Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, and K.-L. Hsu, 2018: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.*, **56**, 79–107, https://doi.org/10.1002/2017RG000574.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, D11102, https://doi.org/10.1029/2007JD009214.

Xu, H., C.-Y. Xu, H. Chen, Z. Zhang, and L. Li, 2013: Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *J. Hydrol.*, **505**, 1–12, https://doi.org/10.1016/j.jhydrol.2013.09.004.