Error Uncertainty Analysis of GPCP Monthly Rainfall Products: A Data-Based Simulation Study

MEKONNEN GEBREMICHAEL AND WITOLD F. KRAJEWSKI

IIHR—Hydroscience and Engineering, The University of Iowa, Iowa City, Iowa

MARK MORRISSEY

Environmental Verification and Analysis Center, The University of Oklahoma, Norman, Oklahoma

DARIN LANGERUD

North Dakota Atmospheric Resource Board, Bismarck, North Dakota

GEORGE J. HUFFMAN

Science Systems and Applications, Inc., and NASA Goddard Space Flight Center, Greenbelt, Maryland

ROBERT ADLER

NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 17 January 2003, in final form 6 June 2003)

ABSTRACT

This paper focuses on estimating the error uncertainty of the monthly 2.5° × 2.5° rainfall products of the Global Precipitation Climatology Project (GPCP) using rain gauge observations. Two kinds of GPCP products are evaluated: the satellite-only (MS) product, and the satellite–gauge (SG) merged product. The error variance separation (EVS) method has been proposed previously as a means of estimating the error uncertainty of the GPCP products. In this paper, the accuracy of the EVS results is examined for a variety of gauge densities. Three validation sites—two in North Dakota and one in Thailand—all with a large number of rain gauges, were selected. The very high density of the selected sites justifies the assumption that the errors are negligible if all gauges are used. Monte Carlo simulation studies were performed to evaluate sampling uncertainty for selected rain gauge network densities. Results are presented in terms of EVS error uncertainty normalized by the true error uncertainty. These results show that the accuracy of the EVS error uncertainty estimates for the SG product differs from that of the MS product. The key factors that affect the errors of the EVS results, such as the gauge density, the gauge network, and the sample size, have been identified and their influence has been quantified. One major finding of this study is that 8–10 gauges, at the 2.5° scale, are required as a minimum to get good error uncertainty estimates for the SG products from the EVS method. For eight or more gauges, the normalized error uncertainty is about 0.86 ± 0.10 (North Dakota: box 1) and 0.95 ± 0.10 (North Dakota: box 2). Results show that, despite its error, the EVS method performs better than the root-mean-square error (rmse) approach that ignores the rain gauge sampling error. For the MS products, both the EVS method and the rmse approach give negligible bias. As expected, results show that the SG products give better rainfall estimates than the MS products, according to most of the criteria used.

1. Introduction and problem description

The Global Precipitation Climatology Project (GPCP) provides monthly global rainfall estimates over areas of 2.5° lat × 2.5° lon grids since 1979 (Adler et al. 2003). Quantitative evaluation of these rainfall estimates by comparison with independently obtained rainfall estimates is essential in assessing the accuracy of the GPCP rainfall estimates. The independent rainfall estimates are commonly based on rain gauge observations. The tremendous spatial scale difference between the GPCP grids and rain gauges makes the direct comparison of the two rainfall estimates problematic because of the sampling error. To compare the two rainfall estimates, the rain gauge point estimates are often interpolated to the spatial scale of GPCP and the sampling error is estimated statistically.

Corresponding author address: Dr. Witold F. Krajewski, IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA 52242.
E-mail: witold-krajewski@uiowa.edu

The sampling error of rain gauge networks has been extensively studied since the pioneering work of Huff (1970). Some of the studies were experimental and the results were believed to be representative for specific precipitation regimes (e.g., Huff 1970; Flitcroft et al. 1989). Others focused on developing theoretical formulations (e.g., Rodriguez-Iturbe and Meija 1974; North and Nakamato 1989; Valdes et al. 1994; Morrissey et al. 1995). The latter studies primarily focused on modeling of the spatial correlation structure of rainfall as being the main factor influencing the sampling error.

Barnston (1991) and Ciach and Krajewski (1999) proposed a method of estimating the uncertainty of remote sensing rainfall estimates that takes the gauge sampling error into account. We refer to the method as the error variance separation (EVS). A key concept in the EVS method is the decomposition of the variance of the remote sensing and rain gauge–rainfall difference into error of the remote sensor and the gauge sampling error. To achieve this, we assume that the gauge sampling error and the remote sensing rainfall error are uncorrelated. Krajewski et al. (2000) implemented the EVS method to estimate the uncertainty of the GPCP products over several sites across the United States. They reported that the results agreed well with the uncertainty estimated using the algorithm developed by Huffman (1997) that is based on the space–time sampling uncertainty of the various sensors involved.

The EVS method is attractive because it filters out the gauge sampling error from the GPCP–gauge difference. However, aside from the observation-based arguments raised by Barnston (1991) and Krajewski et al. (2000), there is yet no rigorous evidence for or against the validity of the assumptions involved in the EVS method. An accuracy assessment of the EVS method is, therefore, required before its practical implementation. Some important questions that arise in this context are as follows: ''How does the accuracy of the EVS method vary with the gauge density?'' ''To what degree does violation of the lack of correlation affect the results of the EVS application?'' We address these questions below.

The main objective of our study is to assess the accuracy of the EVS method for estimating the error uncertainty of the GPCP products. We used three validation sites—two in North Dakota and one in Thailand—all with a large number of rain gauges that allowed us to perform Monte Carlo resampling experiments. We carried out the assessment for a variety of gauge densities. To make the assessment independent of specific gauge networks, we used a resampling experiment to create an ensemble of about 100 gauge networks for each gauge density. We also evaluated the overall performance of the GPCP products using a set of error criteria.

We organized this paper as follows. In the first part we describe the EVS method, the criteria used to select the validation sites, and the characteristics, quality, and error uncertainty of the reference datasets (sections 1–4). In the second part we describe our method of analysis in a stepwise manner (section 5). In the third part we first evaluate the GPCP rainfall estimates by directly comparing them with the reference datasets. Then we present the results of applying a dimensionless error term that indicates the accuracy of the EVS error uncertainty estimates for a variety of gauge densities (section 6). After discussing the results, we close with conclusions and a recommendation.

2. The EVS method

We begin by decomposing the variance of the GPCP–gauge rainfall difference into its different components as

$$\begin{aligned} V(\hat{R}_{\mathrm{GPCP}} - \hat{R}_G) &= V(\hat{R}_{\mathrm{GPCP}} - R_T) + V(\hat{R}_G - R_T) \\ &\quad - 2\,\mathrm{Cov}(\hat{R}_{\mathrm{GPCP}} - R_T, \hat{R}_G - R_T), \end{aligned} \tag{1}$$

where R_T is the true rainfall averaged over the GPCP box, $\hat{R}_{\mathrm{GPCP}}$ is the GPCP rainfall estimate, $\hat{R}_G$ is the sample average of gauges within the GPCP box, and the operators $V()$ and $\mathrm{Cov}()$ are the variance and covariance operators, respectively. The quantities $(\hat{R}_{\mathrm{GPCP}} - R_T)$ and $(\hat{R}_G - R_T)$ represent the GPCP estimation error and the gauge sampling error, respectively. If the error covariance term can be neglected, (1) reduces to

$$V(\hat{R}_{\mathrm{GPCP}} - R_T) = V(\hat{R}_{\mathrm{GPCP}} - \hat{R}_G) - V(\hat{R}_G - R_T). \tag{2}$$

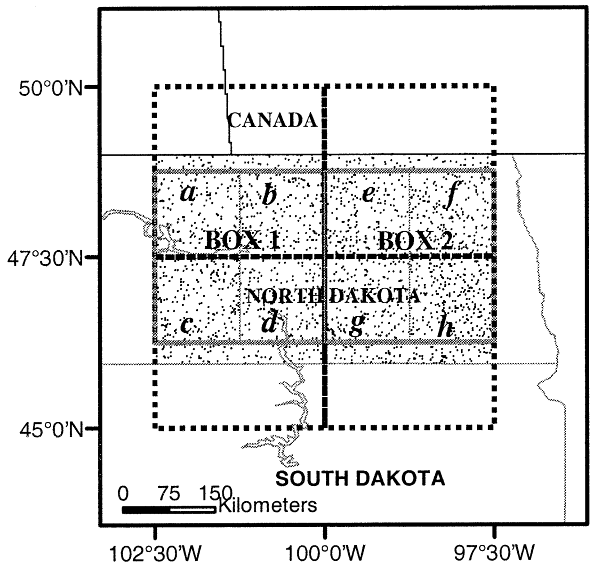
Equation (2) is referred to as the EVS method. The applicability of this method hinges on the assumption of negligible error covariance.

3. Super sites and datasets

Our strategy for addressing the question about accuracy of the EVS method is based on finding areas around the globe that are very densely instrumented. We refer to such sites as ''super sites.'' Availability of rain gauge data from super sites allows designing and conducting resampling experiments focused on specific questions. This strategy has been used successfully in previous studies by Rudolf et al. (1994) and McCollum and Krajewski (1998). We designed the following criteria for the selection of the validation sites:

1) the site must include at least about 100 gauges distributed over a 2.5° × 2.5° area so that adequate resampling experiment can be carried out and
2) the gauges must be uniformly distributed so that the true area-averaged rainfall can be estimated accurately.

It is not easy to find sites that satisfy both criteria, especially because data access in many countries is restricted. Nevertheless, based on the above criteria, we selected three validation sites—two in North Dakota and one in Thailand. The North Dakota rain gauge dataset



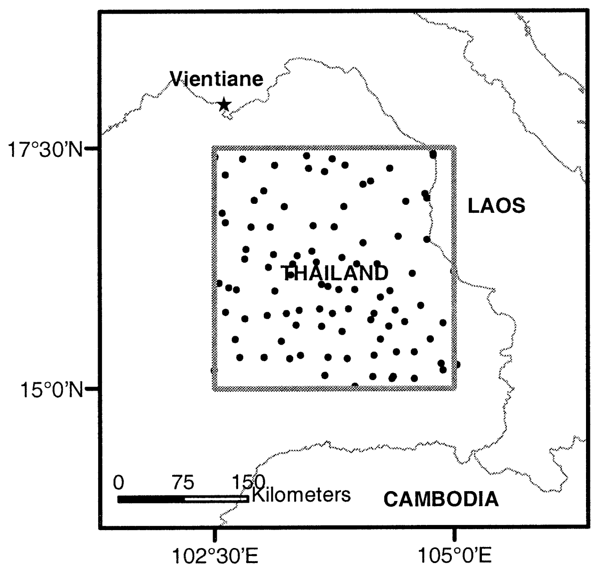


FIG. 1. (top) The two inner boxes (thick solid lines) show the two validation sites in North Dakota. The grids bounded by dotted lines represent the original GPCP grids. (bottom) The inner box shows the validation site in Thailand.

is provided by the Atmospheric Resource Board Cooperative Observer Network (ARBCON). This network is maintained by a number of volunteers who record daily rainfall during April–September using small, inexpensive plastic rain gauges. The Thailand dataset was provided by the Surface Reference Data Center (SRDC) of the GPCP.

The upper panel in Fig. 1 shows the validation sites in North Dakota. The areas bounded by dotted lines represent the GPCP grids. The upper two GPCP grids extend into Canada, and the lower two extend into South Dakota. The ARBCON spans the GPCP grid in such a

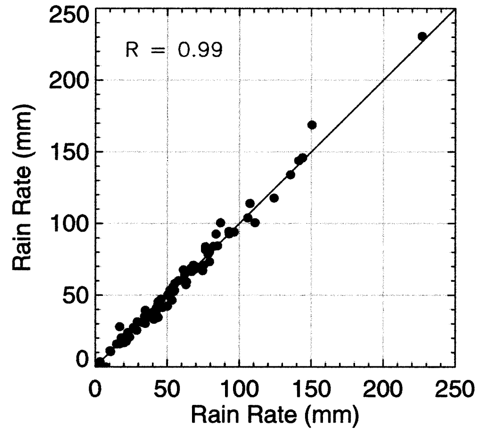


FIG. 2. Scatterplots of the ARBCON rainfall estimates averaged over (x axis) subboxes a, b, c, d, e, f, g, and h, and (y axis) subboxes b, d, e, and g. The identification letters represent the subboxes shown in Fig. 1.

way that no box is completely covered, in particular in the north–south direction. To avoid a situation in which a large portion of the GPCP would be with no rain gauges—which would lead to a significant increase in the sampling error and, therefore, undermine the supersite status of the ARBCON—we position our sites box 1 and box 2 straddling two GPCP grids. Because the overlap is about 50%, we used a simple average of the appropriate GPCP product values to conduct our investigation. Averaging two GPCP values for each box might be a potential source of error. On the one hand, such averaging should reduce the random error, but on the other hand it might introduce error because of rain in the ungauged parts of the region being different than in the gauged part of region. We examined this issue, except for a possible bias due to mean north–south gradients in precipitation. To test how a factor of 2 in the area at the $2.5° \times 2.5°$ scale affects the GPCP statistics, we averaged the ARBCON rainfall estimates for subboxes a, b, c, d, e, f, g, and h (shown in Fig. 1) and compared it with the average of the ARBCON rainfall estimates for the subboxes b, d, e, and g. Figure 2 displays the scatterplot of these two averages; Table 1 provides the corresponding first- and second-order moment statistics. The scatterplot indicates a good agreement between the two sets of values with high correlation coefficients ($R = 0.99$). The mean remains virtually the same, indicating a meteorologically homogeneous area in the east–west direction. Averaging reduces the var-

TABLE 1. Sample statistics for the ARBCON rainfall estimates averaged over subboxes a, b, c, d, e, f, g, and h, and subboxes b, d, e, and g.

Sites (subboxes)	Mean (mm month^{-1})	Variance (mm month^{-1})^2	Sample size
b, d, e, g	58.41	1592.82	79
a, b, c, d, e, f, g, h	58.47	1491.93	79

iance by only about 6%, a consequence of large positive correlation between boxes 1 and 2. We think that such a small reduction in the GPCP variance will not affect the results of this analysis to a significant degree (given the statistical nature of the GPCP algorithms).

We checked the quality of the ARBCON dataset by performing a comparison with independent datasets. Our attempt to look for collocated ''official'' rain gauge data from the standard size gauges was only partially successful. In lieu of many such situations we were forced to consider gauges located some distance away. In general, the disagreement increases as the separation distance increases, but this fact by itself was not sufficient to make convincing arguments for the sufficient quality of the ARBCON. This is because for point data and the daily scale the definition of ''day'' differs between the ARBCON and the other networks and daily summer rainfall displays considerable spatial variability. Integrating the point data to a longer (monthly) timescale improves the situation considerably. However, we still observed some worrisome disagreements. A thorough investigation of this is difficult because the overall confidence in even the official rain gauge data is low. The subsequent step in our analysis was spatial integration.

We used the unified rain gauge dataset (URD) (Higgins et al. 2000) to investigate the quality of the ARBCON network. The URD has been developed from multiple sources of rain gauge data by the Climate Prediction Center (CPC). The daily URD is gridded at a horizontal resolution of 0.25° lat × 0.25° lon over the domain 20°–60°N, 140°–60°W. The URD has passed through standard quality controls conducted by the CPC. In order to compare the ARBCON datasets with the URD, we had to choose an appropriate spatio–temporal scale. The larger the scale, the smaller the area-averaged field variability, and, hence, the better agreement between the two datasets assuming good quality datasets. For the purpose of this study, we selected a spatial scale of 1.25° and a temporal scale of 1 month. In Fig. 3, we show a comparison of the ARBCON and URD monthly rain gauge data aggregated over areas of 1.25° lat × 1.25° lon grids. There is no significant bias between the two gauge datasets. The correlation coefficients (R) and the associated uncertainties (ΔR) corresponding to a 95% confidence interval computed using Fisher's transformation (Stuart and Ord 1994) indicate a strong statistical relationship between the two datasets. More than 90% of the variance of the ARBCON datasets is explained by the URD variance. Therefore, we concluded that the quality of the ARBCON datasets suffices for the purpose of this study. We used the rain gauge data in the 6-month period from April through September over the 23-yr period of 1977–99. Over this period, the long-term monthly rainfall varies from 20 (April) to 100 (July) mm with a mean of 57 mm. The number of gauges available during the period 1977–86 was 364 in box 1 and 427 in box 2. During the period 1987–99, there were 490 gauges in box 1 and 486 in box 2. These numbers include gauges with data gaps.

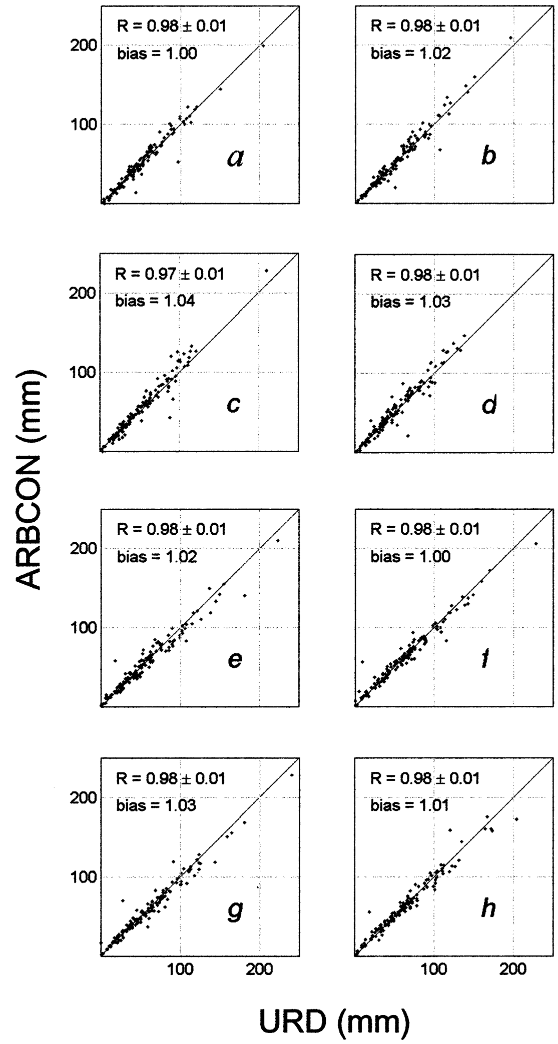


FIG. 3. Comparison of the monthly, 1.25° × 1.25° grids, rainfall estimates obtained from ARBCON and URD for the two validation sites in North Dakota. The identification letters represent the subboxes shown in Fig. 1.

The validation site in Thailand consists of uniformly distributed 95 gauges for the period 1987–91 (see the lower panel in Fig. 1). The rainy season lasts normally 6 months and stretches from May to October. During this season, the long-term monthly rainfall varies from 73 (October) to 276 (August, September) mm with a mean of 180 mm.

We used two types of GPCP products in our analyses, the satellite-only (MS) product and the satellite–gauge (SG) product. We used the GPCP products available during the period 1987–99 (over North Dakota) and 1987–91 (over Thailand). Over North Dakota, the MS product is formed based on microwave data only. However over Thailand, the MS product is formed based on microwave and geostationary IR data. The SG product is produced in two steps. First, the MS estimate is adjusted to the large-scale bias of the rain gauge data. Then, the gauge-adjusted satellite estimate and the gauge analyses, provided by the Global Precipitation Climatology Centre (Rudolf et al. 1992), are combined with inverse-error variance weighting. GPCP also provides the corresponding random error uncertainty estimates for these products based the heuristic approach proposed by Huffman (1997). We made sure that no gauges involved in the SG product are used in our validation.

4. Estimation of the reference data error uncertainty

a. Variance reduction factor

We assumed that the true area-averaged rainfall is estimated accurately over the validation sites so that we would be able to estimate the relevant terms involved in (1) fairly accurately. To achieve this, the sampling error of the reference data has to be practically insignificant. We estimated the true area-averaged rainfall over each validation site as the arithmetic mean of all the gauges within each site. The sampling error of such an estimate can be expressed in terms of the variance reduction factor (VRF), which can be calculated from the following approximation:

$$V(\hat{R}_G - R_T) = V(R_G)\mathrm{VRF}, \quad \text{for } E(\hat{R}_G) = E(R_T), \tag{3a}$$

with

$$\begin{aligned} \mathrm{VRF} = {} & \frac{2}{n^2}\sum_{i=1}^{K-1}\sum_{j=j+1}^{K}\rho(d_{i,j})\delta(i)\delta(j) \\ & - \frac{2}{Kn}\sum_{i=1}^{K}\sum_{j=1}^{K}\rho(d_{i,j})\delta(i) \\ & + \frac{1}{K} + \frac{2}{K^2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\rho(d_{i,j}), \end{aligned} \tag{3b}$$

where $V(R_G)$ is the variance of the gauge point value. The sampling domain is divided into K grid boxes, over which the rainfall is estimated as the arithmetic mean of n gauges. The Kronecker delta function, $\delta(i)$, denotes whether box i contains a rain gauge. The term $\rho(d_{i,j})$ represents the rainfall spatial correlation, where $d_{i,j}$ is the distance between box i and box j. A detailed description of (3) is given in Morrissey et al. (1995) and Krajewski et al. (2000). Because the above formula requires specifying a spatial correlation function, next we discuss its estimation for monthly rainfall for the three super-site locations.

b. Empirical estimation of correlation functions

To estimate the spatial correlation function first we have to compute the sample correlation coefficient for all of the gauge pairs. We estimated the true correlation coefficient as the product-moment sample correlation coefficient. To avoid problems with a very small sample size, we calculated the correlation coefficient only when both pairs contain at least 30 months of data. A practical concern with the product-moment correlation is that it is not robust with respect to nonnormality of the data (Kowalski 1972; Wilcox 1997; Habib et al. 2001). For each gauge, we performed a test of normality based on the sample correlation coefficient between the data and the corresponding normal scores. The results and the critical values are displayed in Table 2. Only 29% of the gauges over Thailand, and none over North Dakota, passed the normality test at the 5% significance level. All of the datasets exhibit some positive skewness. Log transformation of the data did not improve the results (not shown here). The results of a sensitivity analysis of the error estimate to the correlation structure (described in the following section) indicated that the error in the product-moment correlation does not affect the estimate of the reference data error uncertainty significantly.

TABLE 2. Percentages of gauges with different sample correlation coefficient and skewness values. The correlations are between the gauge data and the corresponding normal scores. The 5% significant level critical values for the normal scores correlation normality test appear in boldface.

Site	No. of gauges (%)	Sample correlation coef	Sample skewness
North Dakota	1	<0.900	2.66–2.88
	31	0.900–0.945	1.48–2.23
	68	0.950–0.986	0.86–1.12
	0	>**0.986**	—
Thailand	4	<0.900	1.82–2.22
	28	0.900–0.945	0.57–1.57
	39	0.950–0.972	0.17–1.14
	29	>**0.972**	0.11–0.59

The result of the empirical analysis of the correlation coefficients is a scatterplot of correlations. To identify the correlation function, we assumed the true correlation structure to be of the form

$$r = c_1 \exp[-(d/c_2)^{c_3}], \tag{4}$$

where d is the separation distance between two gauges, and c_1, c_2, and c_3 are the parameters. This function has the required property to be positive definite. Figure 4 shows the spatial correlation functions along with the 95% confidence intervals of the sampling uncertainty estimated using Fisher's transformation. The fitted functions are $r = \exp[-(d/580)^{0.52}]$ (North Dakota) and $r = 0.93 \exp[-(d/595)^{0.62}]$ (Thailand). A value of c_1 less than unity (0.93 over Thailand) may have originated from small-scale variability, the random measurement error, or simply as an artifact of the fitting procedure. The decorrelation distance, the distance at which the decorrelation decreases to $1/e$, is often used for a simple characterization of the correlation function (e.g., Bras and Rodriguez-Iturbe 1993). In our case the decorre-

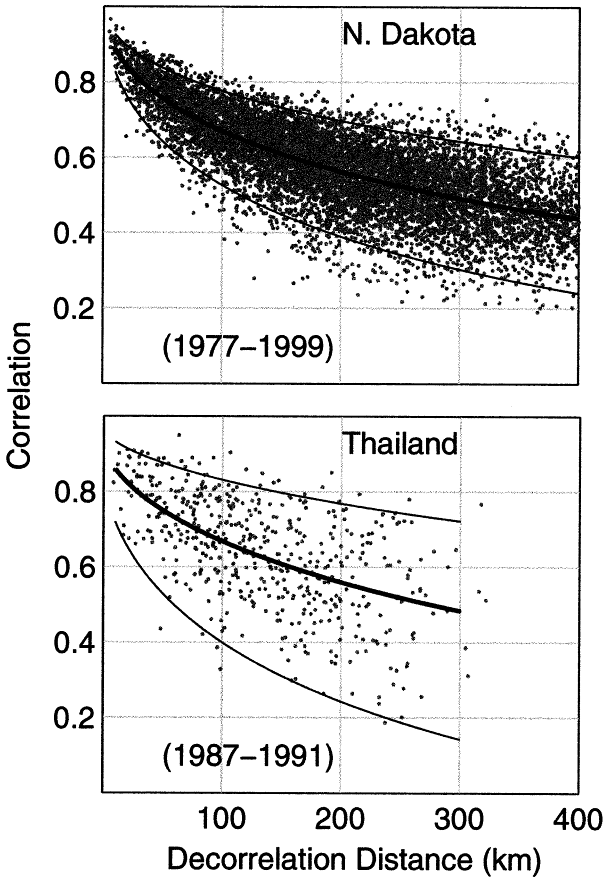


FIG. 4. Correlation coefficients estimated from multiple realizations with 95% confidence interval and fitted correlation function. Temporal resolution: 1 month; model domain: 2.5° × 2.5°.

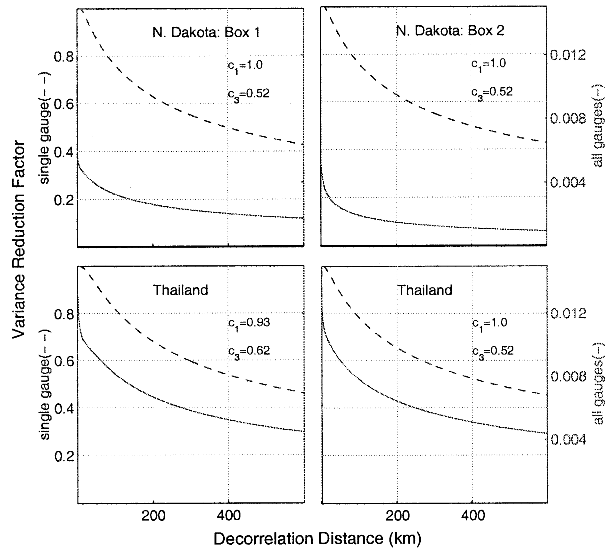


FIG. 5. Sensitivity of the variance reduction factor to decorrelation distance for different rainfall spatial correlation models. The variance reduction factors are calculated for all the network gauges (solid line) and a single gauge close to the center of the GPCP box (dashed line).

lation distances are 580 km for North Dakota and 527 km for Thailand.

c. Error uncertainty of the reference data

Based on the above correlation functions, we calculated the VRF using all the gauges active since 1987 (Table 3). In the same Table, we also showed the VRF for a single gauge located close to the center of the GPCP box. We found that the VRF of each network is very small and the error variance of each network is less than 2% of the single gauge error variance.

The analysis thus far was based on the product-moment correlation. As discussed in the preceding section, our datasets are not normally distributed, and, hence, we may be in a situation in which the product-moment correlation is a poor measure of the true correlation coefficient. Lai et al. (1999) and Habib et al. (2001) showed that the product-moment correlation overestimates the true correlation coefficient for a positively skewed data. Based on this, we believe that the true decorrelation distances are shorter than what we reported above. In order to determine how important it is to put effort into the estimate of the correlation function, we investigated the sensitivity of the reference data error uncertainty with respect to the decorrelation distance. Another concern regarding the Thailand network is the fact that the correlation model was obtained using a small sample size (5 years of data), and consequently the uncertainty of the correlation model is very large (Fig. 4). Therefore, for the Thailand network, we carried out the sensitivity analysis using the correlation models obtained over both Thailand and North Dakota sites. Figure 5 presents the VRF estimates for the networks and a single gauge close to the center of the GPCP box

TABLE 3. Variance reduction factors calculated for the reference gauges and a single gauge close to the center of the GPCP box.

Site	Decorrelation distance (km)	Variance reduction factor: All gauges active since 1987 (1)	Variance reduction factor: One gauge located close to the center of the GPCP box (2)	$^{(1)}/_{(2)} \times 100\%$
North Dakota: box 1	580	0.00171	0.23518	0.7
North Dakota: box 2	580	0.00095	0.23518	0.4
Thailand	527	0.00448	0.28675	1.6

as a function of the decorrelation distance. The results show that the VRF estimates are very small for all the decorrelation distances. This is also true for the Thailand site, although it exhibits relatively larger VRF estimates. It can also be seen that for distances larger than 100 km over North Dakota, the slope of the VRF curve is very small, indicating that the VRF is weakly sensitive to the decorrelation distance over this range. The sensitivity is relatively stronger for the Thailand network, but since the VRF estimates are very small over the entire range of decorrelation distances, the degree of sensitivity does not affect the reference data error uncertainty significantly. For the Thailand site, the results also do not change significantly by adopting the other correlation model parameters obtained for the North Dakota site. Clearly, within the broad range of uncertainty of specifying the correct decorrelation distance, the VRF is not very sensitive and the error variance of the network remains within 2% of the single gauge error variance. Therefore, we concluded that the reference data yields a good approximation of the true area-averaged rainfall over all three validation sites.

5. Method of analysis

We carried out a Monte Carlo resampling analysis to assess the accuracy of the EVS error uncertainty estimates for a variety of gauge density. Our strategy assumes that using all the network gauges results in negligible errors. Sampling a small number of gauges from the entire network leads to an error. Also, we can calculate the GPCP error either directly by using all the available data or by using the EVS framework. Repeating the procedure many times will provide a basis for statistical characterization of the results.

There were six steps involved in this process:

1) We estimated the true area-averaged rainfall R_T over the validation site as the arithmetic mean of all the gauges within the site.
2) We computed the sample estimate of the true GPCP error variance, $V(\hat{R}_{\mathrm{GPCP}} - R_T)$.
3) We performed a Monte Carlo experiment to select the gauge networks with a gauge density of 1, 2, 4, 6, 8, 10, 15, 20, 25, and 30 gauges. To avoid having the results dependent on a specific gauge network, we selected an ensemble of 96 (Thailand) and 100 (North Dakota) gauge networks for each gauge density. We assumed random gauge networks.
4) For each gauge density and gauge network obtained in step 3, we estimated R_T as the arithmetic mean of these gauges. Then we computed the sample estimates of the following terms: $V(\hat{R}_{\mathrm{GPCP}} - \hat{R}_G)$, $V(\hat{R}_G - R_T)$, and $V(\hat{R}_{\mathrm{GPCP}} - R_T)$.
5) We computed the sample estimate of the normalized error (ε_r):

$$\varepsilon_r = \frac{\sqrt{V(\hat{R}_{\mathrm{GPCP}} - \hat{R}_G) - V(\hat{R}_G - R_T)}}{\sqrt{V(\hat{R}_{\mathrm{GPCP}} - R_T)}}. \quad (5)$$

   The ε_r represents the EVS-estimated standard deviation of the GPCP error normalized by the true standard deviation of the GPCP error.
6) We computed the mean and standard deviation of ε_r over the ensemble.

In step 4 above, the term $V(\hat{R}_{\mathrm{GPCP}} - \hat{R}_G)$, is often interpreted as a measure of the uncertainty of the satellite rainfall estimation. Clearly, the validity of such interpretation depends on the accuracy of the reference (i.e., gauge) value. When the ground estimate is based on one or only few gauges it is subject to a significant sampling error. The EVS attempts to correct for this error at the variance level. The two terms that involve the ''true'' rainfall can be calculated only in studies such as this one. When validation sites with small number of gauges are used these terms cannot be evaluated.

6. Results and discussion

a. Evaluation of the GPCP estimates

First, we provide the quantitative evaluation of the GPCP rainfall estimates. Then we proceed to the results of the EVS calculation. We used a set of performance criteria to assess the strength of the statistical relationship between the GPCP rainfall estimates and the reference. These criteria are (a) the bias, (b) the mean absolute error (MAE), (c) the root-mean-square error (rmse), (d) the correlation coefficient (R), and (e) the resolved variance statistic (β). The bias is defined as the ratio of the mean GPCP estimate to the mean surface reference. The nondimensional resolved variance statistic β (Murphy and Epstein 1989; Murphy 1995) is expressed as

$$\beta = 1 - \mathrm{rmse}^2/V(\hat{R}_G). \quad (6)$$

The skill score β is a rigorous measure of the similarity between two variables: it measures their correspondence not only in terms of the relative departures from their means, but also in terms of the means and absolute variances of the two series. A trivial ''climatological'' estimate results in a skill of $\beta = 0$. Estimates with $\beta > 0$ can be considered skillful, with $\beta = 1$ for a perfect estimate.

We compared the surface reference rainfall to the MS and SG rainfall estimates, producing the summary statistics in Table 4. Scatterplots for the individual matchups are displayed in Fig. 6. Let us first discuss the MS rainfall estimates. The bias characteristics show that the MS overestimates rainfall over North Dakota by 29% (box 1) and 25% (box 2). However, it underestimates the rainfall over Thailand by a smaller amount (6%). The MS estimate shows very large rmse, which amounts to more than 100% of the standard deviation of the surface rainfall. Weak correlation (<0.5) exists between

TABLE 4. Summary statistics comparing the MS and SG rainfall estimates to the surface reference.

GPCP product	Validation site	$R \pm \Delta R$	Bias	Rmse (mm)	MAE (mm)	β
SG	North Dakota: box 1	0.91 ± 0.05	1.02	15.48	12.05	0.83
	North Dakota: box 2	0.89 ± 0.06	1.04	17.70	13.91	0.79
	Thailand	0.96 ± 0.03	1.29	52.65	38.00	0.64
MS	North Dakota: box 1	0.39 ± 0.22	1.29	64.15	46.59	−1.87
	North Dakota: box 2	0.45 ± 0.21	1.25	74.82	51.64	−2.72
	Thailand	0.47 ± 0.27	0.94	87.69	63.70	0.01

the MS estimate and the surface reference. The MS rainfall estimates have very poor skill ($\beta \leq 0$), which indicates that the variance of the surface reference is not sufficient to explain the difference between these rainfall estimates and the surface reference.

The SG estimate over North Dakota does not exhibit any significant bias. However over Thailand, it overestimates rainfall by 29%. The number of gauges involved in the SG estimate over Thailand is eight. The significant bias observed over Thailand should be viewed critically because the inclusion of eight gauges introduced significant bias. Certainly, if there is any bias in the gauge-based analyses, it will also remain in the SG estimate. The correlation between the SG and the surface reference is in excess of 0.85 over all sites. The SG rainfall estimates exhibit a very good skill. The rmse of the SG rainfall estimates amounts to 40%–60% of the standard deviation of the surface reference. Except the bias noticed in the SG estimate over Thailand, all the other performance criteria confirm that the SG rainfall estimates are much better than the MS rainfall estimates.

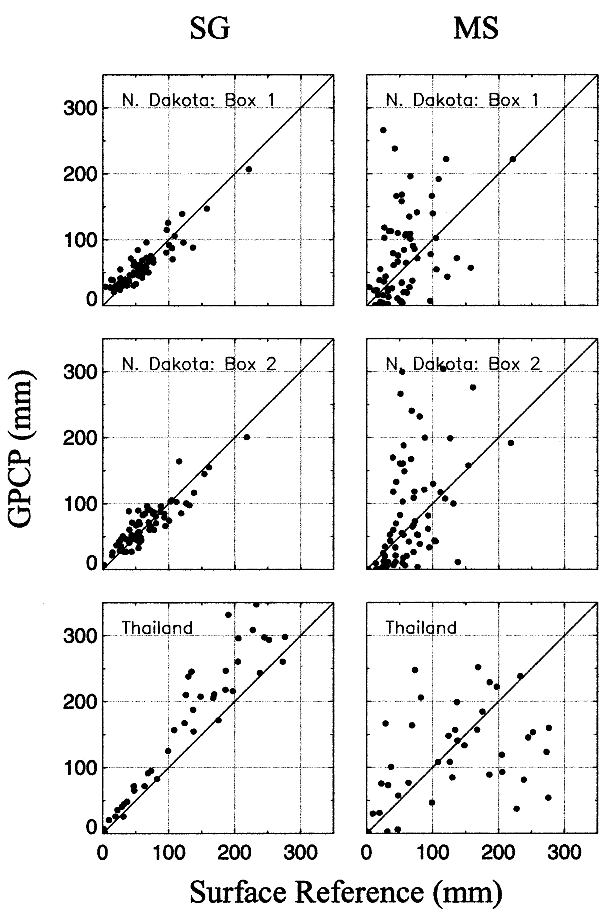


FIG. 6. Scatterplots of the GPCP (left) SG and (right) MS rainfall estimates vs the surface reference rainfall.

As mentioned in section 3, the GPCP also provides the corresponding error uncertainty estimates computed using Huffman's (1997) method. These error uncertainty estimates are given in the form of standard error for each month. Based on the monthly time series of the surface reference data and the GPCP rainfall estimates, we also computed a single standard error value for each box. It can be seen from Table 5 that for the SG products over North Dakota the standard error computed using the surface reference data falls within the 25th and 75th percentile of the GPCP error uncertainty estimates. However, for the SG product over Thailand and all of the MS products, the standard errors provided by GPCP are underestimated. We recomputed the standard errors after adjusting for the bias in the SG and MS rainfall estimates, however, the results did not change the conclusion drawn above.

b. *Error variance separation method results*

As already mentioned in section 5, we used the normalized error ε_r, as defined in (5), to evaluate the accuracy of the EVS results. Figure 7 shows the estimates of ε_r, taken to be the sample mean of about 100 ensemble members, for each gauge density. The uncertainty of ε_r estimates, taken to be the standard deviation of the ensemble members, is also shown on the same figure. Let us first analyze the results for the SG rainfall estimates over North Dakota. The EVS method generally gives biased error uncertainty estimates. For eight and more gauges at the 2.5° scale, the bias and the uncertainty appear to change very little with gauge density. For a network in North Dakota with eight or more gauges, the ε_r is about 0.86 ± 0.10 (box 1) and 0.95 ± 0.10 (box 2). The source of the bias is attributed to the neglected error covariance term in the EVS method, and the uncertainty arises from the variation in the relative location of the gauges within the GPCP grid. For

TABLE 5. Comparison of the standard error computed using Huffman's method (GPCP) and the surface reference.

GPCP product	Site	GPCP			Surface reference	
		25th percentile	50th percentile	75th percentile	Using original GPCP	Using bias-adjusted GPCP
SG	North Dakota: box 1	12.67	16.38	21.20	15.48	15.42
	North Dakota: box 2	11.39	15.67	20.33	17.70	17.52
	Thailand	6.41	17.08	26.74	52.65	37.50
MS	North Dakota: box 1	19.62	27.86	44.16	64.15	61.47
	North Dakota: box 2	20.04	35.27	56.90	74.82	72.19
	Thailand	9.59	31.87	46.14	87.69	87.33

the MS rainfall estimates over North Dakota, the EVS results give ε_r of 0.97 $\pm$ 0.05 or less for all of the gauge densities considered in this study. This indicates that the error covariance is very small relative to the large true standard error of the MS rainfall estimates (see the rmse results for the MS rainfall estimates in Table 4).

Next, we compared the EVS results with the rmse. The rmse, a measure often used to evaluate rainfall products, ignores the gauge sampling error. In Fig. 8, we plotted the normalized rmse expressed as

$$\text{rmse} = \frac{\sqrt{V(\hat{R}_{\text{GPCP}} - \hat{R}_G)}}{\sqrt{V(\hat{R}_{\text{GPCP}} - R_T)}}. \tag{7}$$

For the SG products, the rmse causes large systematic errors. As expected, the EVS method performs better than the rmse for the SG products. However, for the MS products, the rmse gives almost unbiased error uncertainty estimates. This indicates that the gauge sampling error is very small relative to the large true standard error of the MS rainfall estimates.

For regional differences, we also looked at the EVS results over Thailand. As can be seen in Fig. 7, the EVS results over Thailand generally exhibit larger uncertainty of the ε_r for both the SG and the MS products. To investigate whether this variability is caused by regional or sample size differences, we performed a sampling experiment using the data of North Dakota on a network configuration close to that of the true Thailand Over North Dakota (box 2), we extracted 95 gauges that resemble best the Thailand gauge network geometrically, and then we extracted the corresponding dataset for the period 1987–91. Using this dataset as a reference, we applied our method of analysis and computed the mean and standard deviation of the ε_r (bottom panels in Fig. 7). Comparison of the ε_r distribution over North Dakota (box 2) with the ε_r distribution over the simulated network reveals that the sample size indeed affects the uncertainty of the ε_r significantly. Therefore, the uncertainty of the ε_r computed over Thailand does not exclusively show the systematic error of the EVS method, because it is contaminated by the small sample size problem.

7. Conclusions

In this paper, we examined the accuracy of the EVS method for estimating the error uncertainty of the GPCP $2.5° \times 2.5°$ monthly rainfall estimates. In particular, we concentrated our efforts on determining the effect of the gauge density on the accuracy of the EVS results. We found that the accuracy of the EVS error uncertainty estimates for the SG product differs from that of the MS product. Let us begin with the SG product.

We have identified the key factors that affect the errors of the EVS results, such as the gauge density, the gauge network, and the sample size, and we have quantified their influence. One major finding of this study is that 8–10 gauges, at the 2.5° scale, are required as a minimum to get good error uncertainty estimates from the EVS method. For eight or more gauges, the normalized error is about 0.86 $\pm$ 0.10 (North Dakota: box 1) and 0.95 $\pm$ 0.10 (North Dakota: box 2). We also demonstrated that the sample size affects the bias of the EVS results significantly. We showed that, despite its errors, the EVS method gives better error uncertainty estimates of the standard errors of the SG products than the rmse, therefore, the gauge sampling error should be accounted for appropriately as in the EVS method. For the MS products, we showed that both the EVS method and the rmse approach give almost unbiased error uncertainty estimates for all the gauge densities considered. Our results showed that this is mainly because the gauge sampling error and the error covariance are very small relative to the large true standard error of the MS products. The above results confirm the validity of the GPCP performance evaluation by Krajewski et al. (2000), who used sites with about 20 gauges.

We also studied the overall performance of the GPCP products using a set of error criteria. Results showed that the SG products give better rainfall estimates than the MS products according to most of the criteria used. However, the SG estimate over Thailand shows a more significant bias than the corresponding MS estimate. Therefore, we recommend that the gauge dataset used in the SG rainfall estimates be examined to detect the source of the bias. We also assessed the performance of the error uncertainty estimates provided by GPCP. For the SG products over North Dakota, the standard error estimates provided by GPCP agree well with those estimates calculated using the surface reference data. However, for the SG product over Thailand and all the MS products, the GPCP error uncertainty estimates are underestimated.

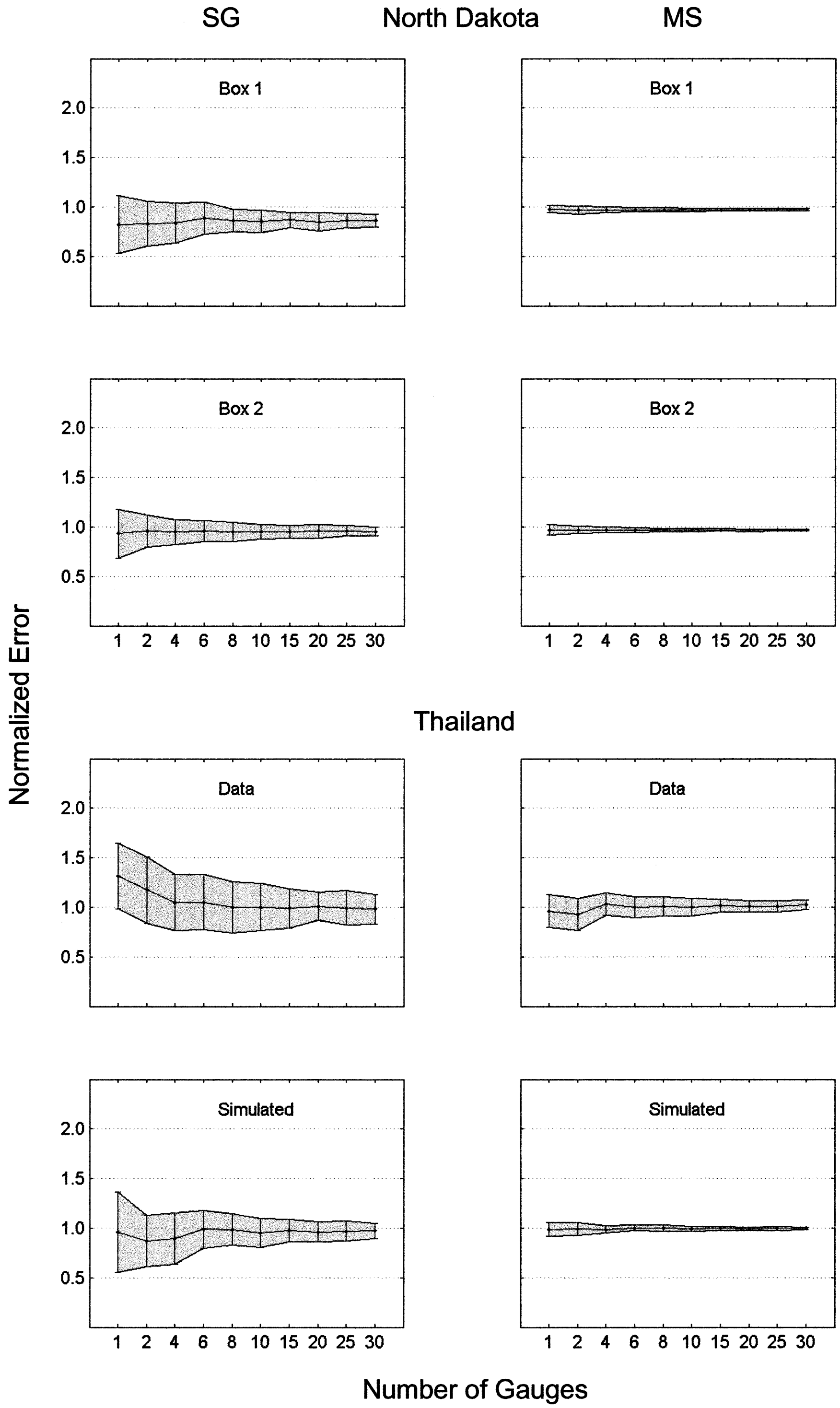


FIG. 7. Sample estimates of normalized error (EVS estimated standard errors normalized by the true standard errors of the GPCP rainfall products) for a variety of gauge densities over the validation sites. The middle curve in each panel represents an ensemble average of about 100 different gauge networks. The uncertainty bars (±1 std dev) represent the uncertainty in the mean of the ensemble. The bottom panels labeled "simulated" refer to the simulation performed using the data of the North Dakota network (to get a longer sample size) on a network configuration close to that of the true Thailand.

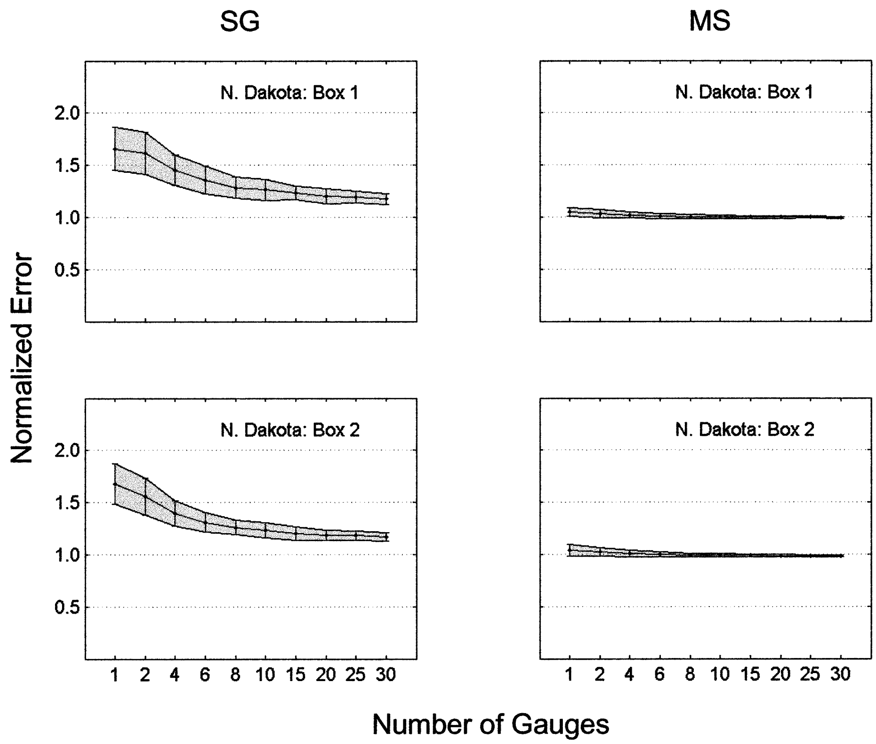


FIG. 8. Rmse normalized by the true standard errors of the GPCP rainfall products for a variety of gauge densities. The mean and the uncertainty bars (± 1 std dev) are computed from an ensemble of 100 different gauge networks.

As a closing remark, the assumption that the SG and gauge errors are uncorrelated leads to errors in the EVS error uncertainty estimates. For eight and more gauges at the 2.5° scale, these errors are small and might be perhaps tolerable, depending on the accuracy sought. However, we believe that there is a need to develop a rigorous validation methodology that can properly and explicitly account for the error covariance. We also point out that although this study was carried out using data collected from a few sites, we believe that the results of this study are valid for climatological regimes similar to those of the validation sites. However for areas characterized by different climatological regimes, our results may not be applicable.

The EVS method provides only one aspect of the GPCP uncertainty characterization. It does not answer questions such as ''What is the probability distribution of the errors?,'' ''Are the errors correlated in space and time?,'' or ''How are the errors related to the characteristics of rainfall such as the amount and frequency of rainfall?'' Addressing these questions requires additional data and research. Some such investigations are under way, and we will report on the results in the future.

Acknowledgments. This research was supported by the NOAA Office of Global Programs through Grant NA57WHO517 to the second author. Aaron Gilstad provided the ARBCON dataset and the staff of the Environmental Verification and Analysis Center provided the SRDC data for Thailand. Their help is gratefully appreciated.

REFERENCES

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Barnston, A. G., 1991: An empirical method of estimating raingage and radar rainfall measurement bias and resolution. *J. Appl. Meteor.,* **30,** 282–296.

Bras, R. L., and I. Rodriguez-Iturbe, 1993: *Random Functions and Hydrology.* Dover, 559 pp.

Ciach, J. G., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.,* **22,** 585–595.

Flitcroft, I. D., J. R. Milford, and G. Dugdale, 1989: Relating point to area average rainfall in semiarid west Africa and the implications for rainfall estimations derived from satellite data. *J. Appl. Meteor.,* **28,** 252–266.

Habib, E., W. F. Krajewski, and G. J. Ciach, 2001: Estimation of rainfall interstation correlation. *J. Hydrometeor.,* **2,** 621–629.

Higgins, R. W., W. Shi, and E. Yarosh, 2000: Improved United States precipitation quality control system and analysis. NCEP/Climate Prediction Center Atlas 7, 40 pp.

Huff, F. A., 1970: Sampling errors in measurement of mean precipitation. *J. Appl. Meteor.,* **9,** 35–44.

Huffman, G. J., 1997: Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteor.,* **36,** 1191–1201.

Kowalski, C. J., 1972: On the effect of non-normality on the distribution of the sample product-moment correlation function. *Appl. Stat.,* **27,** 1–12.

Krajewski, W. F., G. J. Ciach, J. R. McCollum, and C. Bacotiu, 2000: Initial validation of the Global Precipitation Climatology Project monthly rainfall over the United States. *J. Appl. Meteor.,* **39,** 1071–1086.

Lai, C. D., J. C. W. Rayner, and T. P. Hutchinson, 1999: Robustness of the sample correlation—The bivariate lognormal case. *J. Appl. Math. Decis. Sci.,* **3,** 7–19.

McCollum, J. R., and W. F. Krajewski, 1998: Uncertainty of monthly rainfall estimates from rain gauges in the Global Precipitation Climatology Project. *Water Resour. Res.,* **34,** 2647–2654.

Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang, 1995: The uncertainty in simple spatial averages using raingage networks. *Water Resour. Res.,* **31,** 2011–2017.

Murphy, A. H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting,* **10,** 681–688.

——, and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.,* **117,** 572–581.

North, G. R., and S. Nakamoto, 1989: Formalism for comparing rain estimation designs. *J. Atmos. Oceanic Technol.,* **6,** 985–992.

Rodriguez-Iturbe, I., and J. M. Mejia, 1974: The design of rainfall networks in time and space. *Water Resourc. Res.,* **10,** 713–728.

Rudolf, B., H. Hauschild, M. Reiss, and U. Schneider, 1992: Beiträge zum weltzentrum für Niederschlags-klimatologie—Contributions to the Global Precipitation Climatology Centre (English abstract). *Meteor. Z. N. F.,* **1,** 7–84.

——, ——, W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. *Global Precipitations and Climate Change,* M. Desbois and F. Desalmand, Eds., NATO ASI Series I: Global Environmental Change, Vol. 26, Springer Verlag, 173–186.

Stuart, A., and J. K. Ord, 1994: *Kendall's Advanced Theory of Statistics.* Vol. 1, *Distribution Theory,* 6th ed. Edward Arnold, 676 pp.

Valdes, J. B., E. Ha, C. Yoo, and G. R. North, 1994: Stochastic characterization of space-time precipitation: Implications for remote sensing. *Adv. Water Res.,* **17,** 47–59.

Wilcox, R. R., 1997: *Introduction to Robust Estimation and Hypothesis Testing.* Academic Press, 296 pp.