Understanding the Dynamics of Transfer of Satellite Rainfall Error Metrics From Gauged to Ungauged Satellite Gridboxes Using Interpolation Methods

Ling Tang and Faisal Hossain

Abstract—**Knowledge of error characteristics of high resolution satellite rainfall data at different spatial/temporal scales is useful, especially when the scheduled Global Precipitation Mission (GPM) plans to provide High Resolution Precipitation Products (HRPPs) at global scales. Satellite rainfall data contain errors which need ground validation (GV) data for characterization, while satellite rainfall data will be most useful in the regions that are lacking in GV data. Therefore, a critical step to bridge this gap is to assess spatial interpolation schemes for transfer of the error characteristics from GV regions to non-GV regions. In this study, a comprehensive assessment of kriging methods for spatial transfer (interpolation) of error metrics is performed. Three kriging methods for spatial interpolation are compared, which are: ordinary kriging (OK), indicator kriging (IK) and disjunctive kriging (DK). Additional comparison with the simple inverse distance weighting (IDW) method is also performed to quantify the added benefit (if any) of using geostatistical methods. Four commonly used satellite rainfall error metrics are assessed for transfer to non-GV satellite gridboxes: Probability of Detection (POD) for rain, False Alarm Ratio (FAR), bias (BIAS), and Root Mean Squared Error (RMSE). Results show that performance of a kriging scheme is strongly sensitive to the timescale for which the errors are interpolated (monthly and weekly) wherein the extent of coverage by GV data plays an equally sensitive role. While most kriging techniques perform well according to correlation measure at climatologic timescales for a range of GV data coverage, only DK and OK appear to retain accuracy at the shorter timescales (monthly and weekly). However, scalar assessment metrics such as mean and standard deviation of error (i.e., difference between true and interpolated errors) reveal a completely different picture of accuracy of each interpolation method. In terms of such assessment measures, the overall performance ranking of the kriging methods is as follows: $\mathrm{OK} = \mathrm{DK} > \mathrm{IDW} > \mathrm{IK}$. Assessment of kriging methods also revealed that the transfer accuracy is sensitive to error metric type. The ranking of error metrics with highest accuracy in transfer is: $\mathrm{POD} > \mathrm{FAR} > \mathrm{RMSE} > \mathrm{BIAS}$. Overall, the assessment of kriging methods revealed that these best linear unbiased spatial estimators may not be appropriate transfer methods for transfer of satellite rainfall error metrics at time scales shorter than a week. It is worthwhile now to pursue more non-linear transfer methods (such as neural networks) and other kriging methods that use additional spatial information on the rainfall process (such as co-kriging) to further constrain the interpolation uncertainty.**

Manuscript received September 05, 2010; revised December 25, 2010; accepted March 22, 2011. This work was supported by a NASA New Investigator Program (NIP) grant (NNX08AR32G) awarded to F. Hossain. The work of L. Tang was supported by a NASA Earth System Science Fellowship.

The authors are with the Department of Civil and Environmental Engineering, Tennessee Technological University, Cookeville, TN 38505-0001 USA (e-mail: fhossain@tntech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2011.2135840

Index Terms—**Disjunctive kriging, GPM, indicator kriging, ordinary kriging, satellite rainfall error, spatial interpolation.**

I. Introduction

SPATIAL interpolation can bridge gaps in spatial distribution of variables that are of value for scientific or practical applications and yet are difficult or expensive to measure. Most spatial interpolation methods work on the principle of leveraging the spatial dependence (hereafter also referred to as "spatial structure") of the variable. One such class of techniques, that are best linear unbiased estimators (BLUE), is called kriging methods [6], [21]. These kriging methods (for a comprehensive list, readers are referred to [2]) provide an optimal estimate of the variable at interpolated (unsampled) locations wherein the interpolation error is minimum at sampled locations. As such, kriging methods have been applied to numerous spatial variables. For example, Hill *et al.* [8] have explored the effectiveness of ordinary kriging for estimation arsenic contamination of groundwater in Bangladesh for limited sampling data scenarios. Goovaerts [5] has applied kriging for assessing lung cancer mortality rates in the southeastern US. Krajewski [17] reported the use of co-kriging to optimally merge rainfall from radar and gauge networks. Hudson and Wackernagel [12] used kriging for spatial interpolation of temperature in Scotland. Ma *et al.* [20] have assessed kriging methods in the field of groundwater modeling in southcentral Kansas. Theodossiou and Latinopoulos [25] have demonstrated the use of kriging methods for interpolation of ground water levels. There are numerous other studies that report the use of kriging for spatial interpolation.

One area where kriging methods have not experienced any assessment, to the best of our knowledge, is the error of satellite estimated rainfall (hereafter "error" is also interchanged with "uncertainty"). Ever since rainfall products began to be developed using satellite infrared sensors on geostationary orbit three decades ago, satellite remote sensing of rainfall has experienced tremendous progress [4]. In recent times, satellite rainfall estimation has seen significant improvement in resolution both spatially and temporally [23]. From typical resolutions of degree-daily in the 1980s (such as the Global Precipitation Climatology Project (GPCP) [15]; see also [27]), current high resolution precipitation products (HRPP) now provide satellite rainfall estimates using a variety of sensors and platforms at typical scales of 25×25 km every 3 hours across the globe. A few examples of such products are CMORPH [16], Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) [14], and GSMap [26]. A few other products also

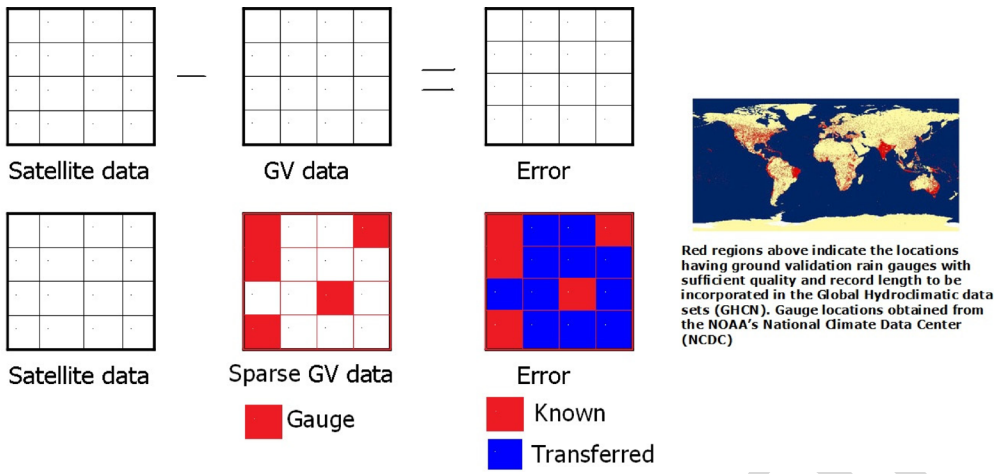


Fig. 1. The set of squares on the left side show the conceptual rendition of the idea of 'transfer' of satellite rainfall error information from a gauged (GV) location to an ungauged (non-GV) location (modified from [24]). Upper panels depicts the notion of 'error' of satellite rainfall data (in this case, the scalar deviation of magnitudes is termed 'error' although there can be many other types of error). Lower panel depicts how the known error (derived from GV sites shown in red in the middle panel) would be 'transferred' to the non-GV (ungauged) sites shown in blue (right most panel). To connect the concept of spatial transfer of satellite rainfall error globally, map with location of ground validation gauges is shown using similar colors.

provide routine rainfall data at much smaller resolutions of 30 min 4×4 km, such as PERSIANN [1].

NASA's planned Global Precipitation Measurement (GPM) mission, in collaboration with other international partners, represents the follow-on mission to the highly successful TRMM. GPM is a unique constellation of rain measuring satellites centered around a core satellite comprising a high-resolution, multi-channel passive microwave (PMW) rain radiometer known as the GPM Microwave Imager (GMI) and augmented by the Dual-frequency Precipitation Radar (DPR) [11]. GPM is currently scheduled for launch in 2013. GPM will seek to achieve measurements with a 3-hr average revisit time over 80% of the globe, and it is expected to provide global high-resolution precipitation products (HRPP) with temporal sampling rates ranging from 3 to 6 hours and spatial resolution of 25–100 km^2 (for more information, refer to http://gpm.gsfc.nasa.gov). If satellite rainfall data continues to witness the improvement it has experienced in recent times, then it is reasonable to expect within the next five years, global availability of more coherent satellite rainfall datasets at scales of interest to a variety of users. Hence, knowledge of the distribution of error will be important in dictating the proper use of satellite rainfall data [9].

What therefore makes the assessment of spatial interpolation techniques, such as kriging, for transfer of satellite rainfall error worthwhile is the general lack of ground validation (GV) rainfall information around the world to characterize error directly for every satellite gridbox. Fig. 1 (modified from [24]) conceptually demonstrates this concept for the spatial transfer of satellite rainfall error and places it in the context of the global land areas. Given the general sparseness of GV rainfall data around the world (as seen from Fig. 1, rightmost panel), most locations (satellite gridboxes) would need to have an "estimate" of the satellite rainfall error because of the impossibility to "measure" the error against information from ground sources. Hence, an effective spatial transfer scheme to interpolate the error at these

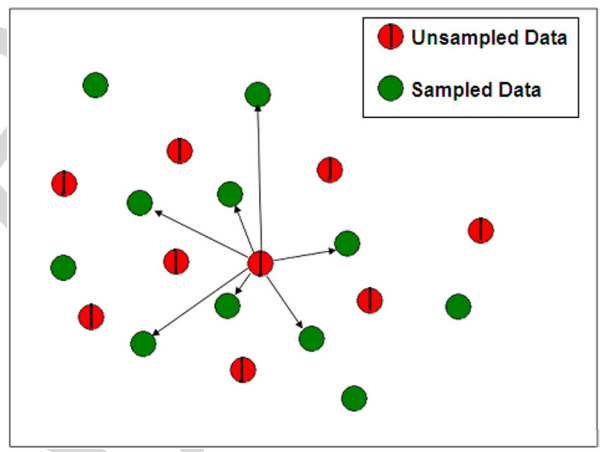


Fig. 2. The general concept of spatial interpolation at unsampled (non-GV) locations (shown in red) on the basis of sampled (known or GV) locations (shown in green). The estimate of error at the non-GV location would be a weighted function of the errors at the GV locations nearby (shown through arrows).

(unsampled) non-GV gridboxes (the white regions of Fig. 1) on the basis of known (sampled) error at nearby GV gridboxes (Fig. 2) would be useful for various users of satellite rainfall data.

But why is information of satellite rainfall estimation uncertainty useful and who may be the biggest beneficiaries of this information? Among the various uses, hydrologic application over land, water management and crop yield forecast will comprise a major avenue through which current and future (GPM) multi-sensor satellite precipitation products will be able to demonstrate tangible benefits to society. In particular, the global nature of coherent and more accurate satellite precipitation products offer hydrologists tremendous opportunities to improve flood monitoring in medium-to-large river basins where rainfall is abundant but in situ measurement networks are generally inadequate. Similarly, existing multi-sensor products such as the TRMM Multi-satellite Precipitation Analysis

(TMPA) product 3B42RT is already used in forecasting of crop yield [22] and famine outbreak [3]. A knowledge of uncertainty of a specific rainfall product can therefore guide the user towards optimal choice of scale and application. For example, a high probable bias would inform a user that the crop yield forecast is likely to be severely overestimated. Similarly, low probability of detection of rain is likely to underestimate and even miss the peak flow for a satellite rainfall based flood forecasting system.

This study builds on two earlier works of Tang and Hossain [24] and Tang *et al.* [28]. Tang and Hossain [24] demonstrated a proof of concept of the application of ordinary kriging for spatial transfer of satellite rainfall error metrics. However, the demonstration was for a rather optimistic scenario at the seasonal timescale using six-year climatologic average (where the spatial dependence and correlation lengths are highest) and assuming only 50% of a region as "gauged" (i.e., having access to GV data). The study also used correlation measure as the assessment metric. The follow-up study by Tang *et al.* [28] revealed that the use of correlation measure as an assessment metric for the accuracy of kriging may be inadequate. However, their study assessed only ordinary kriging.

In this study, various kriging methods are therefore assessed for transfer of error metrics for a range of GV data coverage (from 10% to 90% gauged) and timescales using non-correlation type measures. Three kriging spatial interpolation methods are compared: ordinary kriging (OK), indicator kriging (IK), and disjunctive kriging (DK). The simple and common method of inverse distance weighting (IDW) is also assessed as a "control" method to understand the added benefit (if any) of using such BLUE methods. Because each kriging method has strengths and weaknesses due to assumptions that are inherent, the goal is to understand how each method performs and identify if there exists a "best" technique for a particular scenario. It is also hypothesized that the spatial structure of an error metric may also be dependent on the temporal scale of averaging. For example the satellite bias averaged over a month will likely have much longer spatial correlation lengths as that for bias averaged over a week. Yet, the transfer of satellite rainfall error may need to be performed at various time scales depending on the needs of the user, distribution of GV data and the operational frequency of satellite rainfall data.

The paper is organized as follows. Section II presents study domain and data while Section III presents kriging methods. Section IV provides details on the experimental set-up used for assessment of the interpolation methods. The results and discussion of the application of the kriging methods and IDW are presented in Section V, followed by conclusions in Section VI. The Appendix provides the mathematical formulation of the error metrics used in this study.

II. Study Region and Data

The study region for assessing kriging methods was the Central United States (US) comprising the Great Plains and the Midwest. The geolocation of the four corners of this region are provided in Table I. We chose this region because of the availability of quality controlled GV rainfall data. In addition, the region

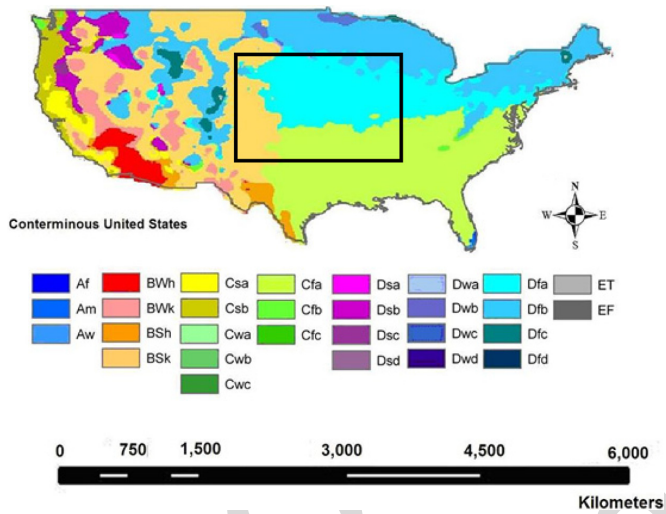


Fig. 3. Study region in the US comprising the Great Plains and the Midwest (shown inside the box). The background colors represent the climate classification according to the Koppen system.

TABLE I
Geolocation of the Four Corners of the Study Region Where Kriging Methods Were Assessed

	Longitude (West)	Latitude (North)
Upper left corner	-104.5	43.5
Upper right corner	-88.25	43.5
Lower left corner	-104.5	33.5
Lower right corner	-88.5	33.5

exhibits diverse climates and rainfall systems for testing spatial interpolation schemes (see Fig. 3 on Koppen climates of the study region). The convective season of the Plains ranges between May and September. Because of the contrasting air mass in this zone, it has frequent severe thunderstorms and tornado outbreaks during spring and summer. Maximum precipitation generally occurs in late spring and early summer. The GV rainfall data pertained to the ground radar network in the US. In order to minimize the error of the GV data in our investigation, we used the National Center for Environmental Prediction's (NCEP) 4 km Stage IV NEXRAD (ground radar) rainfall data that is adjusted to GV gauges over the US [19]. NASA's TMPA satellite rainfall data-product labeled as 3B42RT was used as the satellite rainfall data [14]. This product is globally available on a near real-time basis at 0.25 degree ($\sim 25 \times 25$ km near equator) and 3-hourly resolution from the world-wide web (see ftp://trmmopen.gsfc.nasa.gov). This satellite dataset uses raw radiances from two types of sensors (and orbits): passive microwave (PMW) sensors (low earth orbit) and infrared (IR) sensors (geostationary orbits). For recent upgrades in the 3B42RT algorithm, the reader is referred to Huffman *et al.* [13]. The data for GV and satellite rainfall data spanned the period of June-July-August of 2007 (three months). A point to note is that there also exists research-grade satellite product 3B42 (V6) that is produced by NASA retrospectively by adjusting the bias using gage rainfall. In this study, the focus was on testing the concept of spatial transfer in the operational mode using real-time (RT) products and hence 3B42V6 was not used.

III. Kriging Methods

A. Basic Concept

Because kriging methods have a long heritage of development and use, and because readers can refer to comprehensive geostatistical textbooks (such as [2] and [6]), a very brief introduction of the general kriging as a BLUE method is provided herein. This is followed by a concise description of each type of kriging method used by highlighting only the major difference or assumption behind its formulation.

Kriging is basically an estimator used to find the best linear unbiased estimate of a second-order stationary random field with an unknown constant mean as follows:

$$\hat{Z}(x_{\mathbf{0}}) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

where $\hat{Z}(x_0)$ is the kriging estimate at unsampled location x_0, $Z(x_i)$ is the sampled value at location x_i, and λ_i is the weighting factor for $Z(x_i)$.

The estimation error is

$$\hat{Z}(x_0) - Z(x_0) = R(x_0) = \sum_{i=1}^{n} \lambda Z(x_i) - Z(x_0) \tag{2}$$

where $Z(x_0)$ is the unknown true value at x_0 and $R(x_0)$ is the estimation error. For an unbiased estimator, the mean of the estimation error must equal zero. Therefore,

$$E\{R(x_0)\} = 0 \tag{3}$$

and

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{4}$$

As a BLUE technique, the estimator must also have minimum variance of estimation error. The minimization of the estimation error variance under the constraint of unbiasedness leads to a set of simultaneous linear algebraic equations for the weighting factors, λ_{i}, which can be solved by an optimization routine and the method of Lagrange multipliers. A major step in any kriging method is the estimation of the semi-variogram model parameters such as correlation length, sill and nugget variance. A semi-variogram represents the spatial dependence of the variable being interpolated. In this study, the semi-variogram was modeled as an exponential type function (discussed in Section IV).

B. Ordinary Kriging

The general method outlined in Section III-A is essentially for Ordinary Kriging (OK) and hence is not repeated here.

C. Indicator Kriging

Indicator Kriging (IK) makes no assumption of normality (unlike OK) and is essentially a non-parametric counterpart to OK. However, like OK, the correlation between data points determines weights λ in (1) with the help of the semi-variogram. Instead of working directly with the variable, IK works with "indicator" variables based on threshold values. These threshold values, referred as IK cutoffs, are used to numerically build the distribution of the estimation point. For each IK cutoff, data in the neighborhood are transformed into 0s and 1s: 0s if the data are above the threshold, and 1s if they are below. IK then estimates the probability that the estimation point is less than the threshold value, using this transformed data and a variogram model of the IK cutoff correlation structure. The final output from IK is a spatial map indicating exceedance probability for a threshold at each point/gridbox (which can then be converted back to the variable value within a given range of thresholds). Median IK uses median cutoff (such as 0.5) for indicator covariance. In this paper, we used nine cutoffs (from 0.1 to 0.9). This was based on preliminary sensitivity studies to identify the optimal number of cutoffs that led to the highest accuracy in interpolation.

D. Disjunctive Kriging

The main highlight of Disjunctive Kriging (DK) is that it is a nonlinear procedure in which the original dataset is transformed using a series of additive functions, typically Hermite polynomials, using the semi-variogram of the Gaussian transformed values. DK is also designed for kriging of non-normal data. When the weight function is linear and the random function is multivariate normal, the DK method is the same as the OK method. OK can be considered a special case of the more general DK method.

E. Inverse Distance Weighting

The Inverse Distance Weighting (IDW) is not a BLUE or a geostatistical method. Essentially, the weights λ in (1) are derived as proportional to the inverse of the squared distance between the sampled and unsampled point. In this study, the same set of gridboxes used for kriging methods was used for the IDW method.

IV. Assessment Set-Up

A. Selection of Satellite Rainfall Error Metrics

In general, satellite rainfall error can serve two purposes: 1) to aid algorithm developers to improve their products and understand how individual products could be merged, and 2) to aid users in understanding if the particular product is right for them. This study addressed the latter purpose. Because different users have naturally different needs, it is important to assess a range of error metrics. Furthermore, the spatial structure of each error metric is expected to be unique at a given space-time scale. For example, according to previous research [7], [10], [29], hydrologist users engaged in flash flood or monsoonal flood forecasting benefit more from the knowledge on error metrics such as probability of detection for rain (POD-rain) to understand the accuracy in estimating peak flow, false alarm ratio (FAR) to understand the probable frequency of false alarms in flood warnings and bias (BIAS) to minimize under/over estimation in river stage. On the other hand, crop yield and famine forecasters focus more on the monthly rainfall estimation bias during the growing season as the important indicator of reliability to forecast crop growth (from personal communication with Dr. Chris Funk of the Famine Early Warning System). Hence, the error

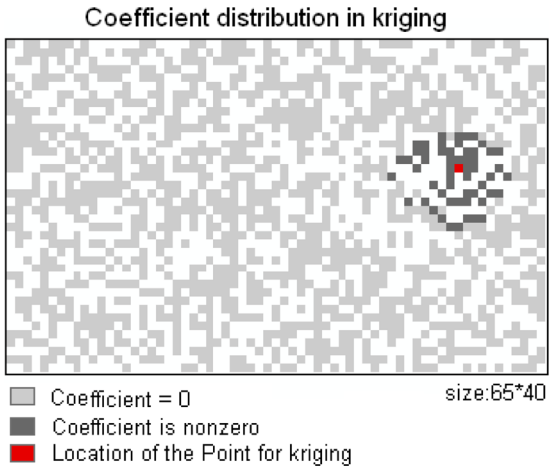


Fig. 4. Map showing kriging weight coefficients [i.e., weighting factors λ in (1)] for interpolation of monthly BIAS at the red (non-GV) satellite grid box. Here each grid box is 0.25° and the correlation length for monthly BIAS is about five grid boxes (~125 km). The entire study domain (comprising 65 × 40 grid-boxes) was used for kriging assuming 50% of the region was gauged (sampled). Thus, almost all the non-zero kriging weighting coefficients are located within two correlation length radius around the non-GV gridbox.

metrics chosen for assessment in this study are: 1) BIAS, 2) root mean squared error (RMSE), 3) POD-rain, and 4) FAR. The mathematical formulation of each error metric is provided in the Appendix. Hereafter the term POD will refer to POD-rain.

B. Kriging Application

Spatial correlograms were derived for each error metric and the correlation length (CL), where the autocorrelation dropped to 1/e (e-folding distance), was then computed. Next, the empirical semi-variograms were derived and then idealized as exponential semi-variogram functions prior to the kriging interpolation as follows:

$$\gamma(h) = c_0 + c(1 - e^{-h/a}) \tag{5}$$

where $\gamma(h)$ is the semi-variance at spatial lag h; c_0 represents the nugget variance (i.e., the minimum variability observed or the noise level at the smallest separating distance equals 0; c is the sill variance—when spatial lag is infinite; and a is the correlation length.

To keep the matrix computations for kriging efficient, the spatial interpolation was performed using a smaller square-sized window around the ungauged satellite grid box rather than the entire collection of gauged gridboxes spanning the whole study domain. The sides of this squared window were set to two correlation length of the error metric being transferred. Preliminary analyses has shown that such a moving window based kriging is justified as the interpolation weights due to gridboxes farther than two correlation lengths from the unsampled gridbox are negligible (see Fig. 4 for monthly timescale using OK to transfer BIAS).

C. Experimental Set-Up to Assess Kriging Methods

Two key aspects of kriging were assessed in this study: 1) spatial aspect and 2) temporal aspect. In the spatial aspect, the goal was to investigate the effectiveness of kriging as a function of relative gaugedness of a region (in other words, the percent of gridboxes having access to GV data). It is expected that the accuracy of kriging at unsampled gridboxes would be lower for a region that was less relatively gauged than one with higher percentage of GV gridboxes. However, this relationship needs to be quantified for each error metric. In the temporal aspect, the goal was to understand the effectiveness of kriging as a function of the timescale of averaging for the error metrics. For example, error metrics were averaged for periods comprising monthly and weekly. As timescales shortens, it is intuitive to expect more randomness in the spatial organization of the error metric (i.e., lower correlation length). A monthly averaged BIAS (for the summer season) will probably have much smoother error field than a weekly averaged BIAS field for the same summer month. It is important therefore that the effectiveness of kriging is understood also as a function of the timescale of averaging of the error metric. As GPM becomes operational, different users will have needs at different timescales, and hence the temporal aspect of the effectiveness of kriging is useful.

Firstly, for any scenario, the true field of error metrics at each grid box was derived using actual satellite and GV data, which was then used as the reference for assessment of kriged estimate of error metrics. For the spatial scenario, it was assumed that only X% of the region's gridboxes had access to GV rainfall data (i.e., "gauged"). The semi-variogram was modeled on the basis of these X% of grid boxes where the true error metric was known (or measured) *a priori*. Kriging was then implemented to estimate error metrics over the remaining (100X)% of the ungauged gridboxes (lacking in GV data). This was similar to a data withholding exercise. Selection of the gauged gridboxes was random and hence each kriging realization was repeated 10 times for each X% of GV data in a Monte Carlo (MC) fashion to derive an average assessment. Kriging was assessed for each "gaugedness" scenario by having X% systematically increased from 10% to 90% and the kriging experiment repeated. As noted earlier, that the true field for an error metric was known *a priori* for the entire study domain, and hence the accuracy of kriging could be directly assessed at the "unsampled" gridboxes. The assessment of accuracy used measures such as correlation, mean error, and standard deviation of error. Assessment error was defined for each non-GV gridbox where kriging was applied as follows:

$$\begin{aligned}\text{Assessment Error} = {} & (\text{Interpolated Error Metric} \\ & - \text{True Error Metric})/(\text{True Error Metric}).\end{aligned} \tag{6}$$

Next, this assessment error was averaged for each scenario (% gauged)

$$\begin{aligned}&\text{Mean Assessment Error} \\ &\quad = \text{Mean of Assessment Error} \\ &\quad (\text{as defined above in Eqn 6}) \text{over all non} - \text{GV} \\ &\quad \text{gridboxes where interpolation was applied}\end{aligned} \tag{7}$$

$$\begin{aligned}&\text{Std. Dev. of Assessment Error} \\ &\quad = \text{Standard Deviation of Assessment Error} \\ &\quad (\text{as defined in Eqn. 6}) \text{ over all non-GV gridboxes} \\ &\quad \text{where interpolation was applied.}\end{aligned} \tag{8}$$

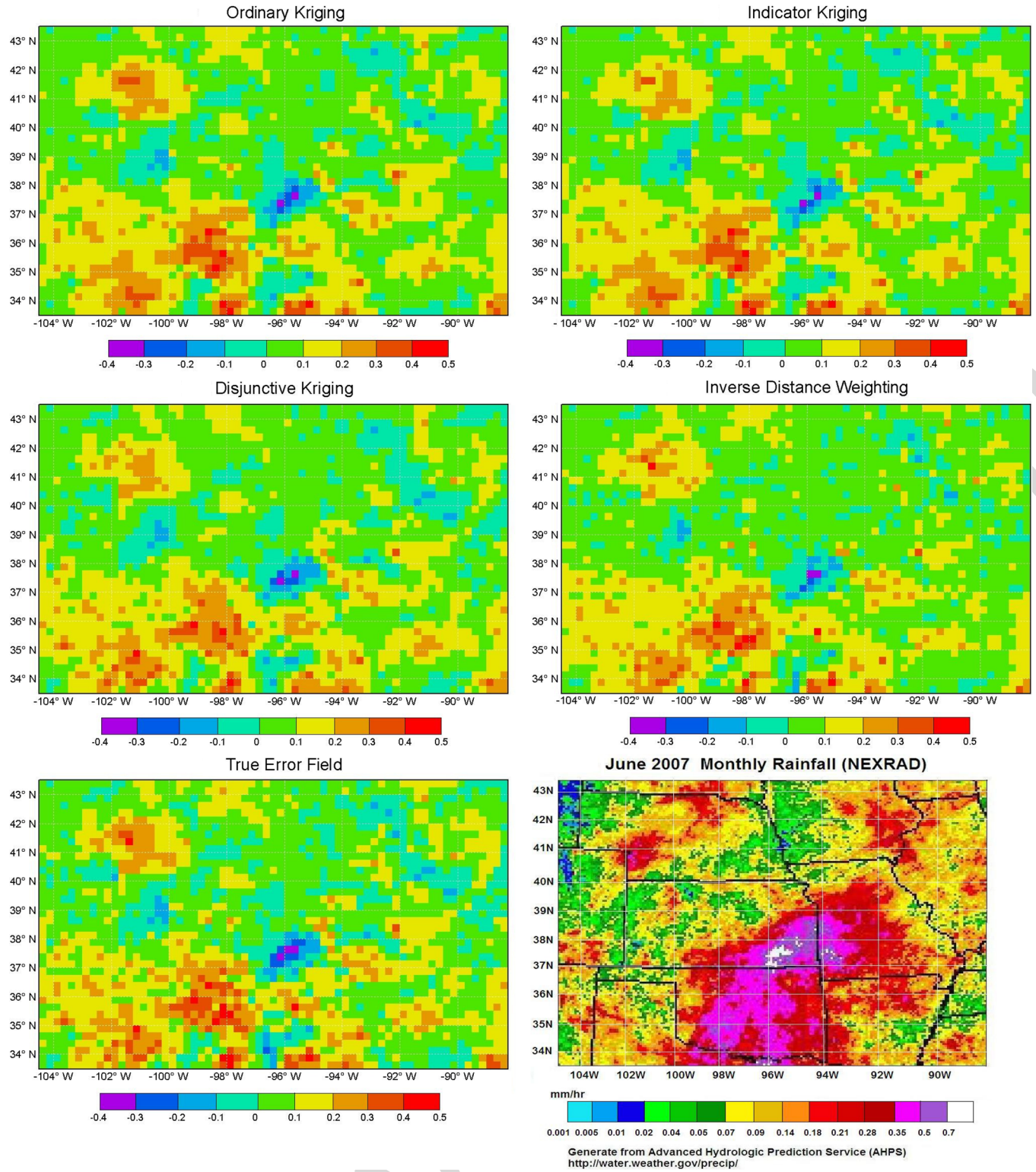


Fig. 5. Example of interpolated error field for the error metric BIAS at the monthly timescale (June 2007) assuming 50% of the study domain as gauged. Upper left panel: ordinary kriging; upper right panel: indicator kriging; middle left panel: disjunctive kriging; middle right panel: inverse distance weighting; lower left panel: true error field (derived from NEXRAD GV data); lower right panel: monthly averaged rainfall field for June 2007 according to TRMM monthly data. Units are in mm/hr.

V. Results and Discussion

Fig. 5 shows an example of application of kriging methods to estimate BIAS for a 50% gauged scenario (i.e., study domain having 50% of gridboxes access to GV rainfall data) at the monthly timescale. Comparing the estimated error field with that of true error and rainfall field reveals that all the kriging as well as IDW methods show promise in transferring BIAS from GV to non-GV regions. Qualitatively, it seems that these spatial interpolation methods are more effective over areas where rainfall varies smoothly. The statistical analysis for each method is

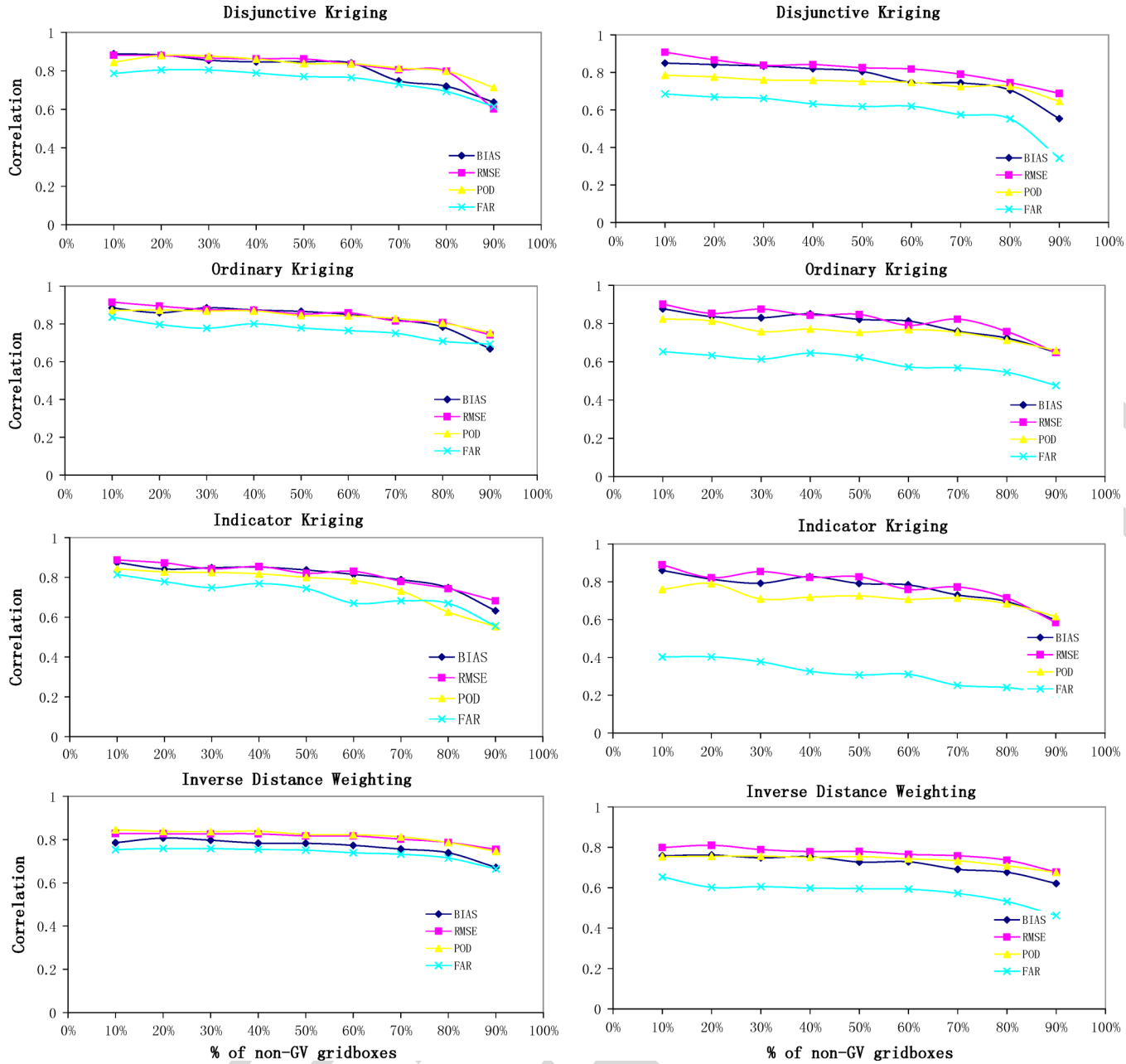


Fig. 6. Performance of kriging methods as a function of error metrics (different colored lines), timescales (left panel: monthly averaged error for June 2007; right panel: weekly averaged error for the first week of June 2007) and % of non-GV grid boxes (shown in x-axis).

provided in the following to identify the relative performance of each method.

Fig. 6 shows the performance of kriging methods as a function of error metrics (different colored lines), timescales (left panel: monthly averaged error; right panel: weekly averaged error) and % of non-GV grid boxes (shown in x-axis). The month of June 2007 is used for demonstration of this performance. For each timescale scenario, the error metrics POD, FAR, BIAS, and RMSE were first averaged accordingly (monthly average for June 2007 or weekly average for the first week of June 2007). The true error field was known *a priori* for the entire study domain since all the gridboxes in reality had access to GV rainfall data. Next, a data withholding exercise was carried out assuming X% of the study region's gridboxes as being non-GV, where X was increased from 10% to 90% in increments of 10% (shown in x-axis). For each X% scenario, ten MC kriging realizations based on ten random sampling of (100X) % of GV gridboxes were carried. The y-axis represents the average (of the 10 MC realizations) correlation measure between error estimated by kriging and the true error at the X% of gridboxes. As expected, the accuracy of interpolation decreased as the % of the domain being "ungauged" increased.

The general trend evident from Fig. 6 is that, using correlation as the assessment measure, there are no significant differences in performance observed across various interpolation methods. IDW performance is as similar as the kriging methods and the added benefit of using a more complex geostatistical method is also not apparent. At the weekly timescale, interpola-

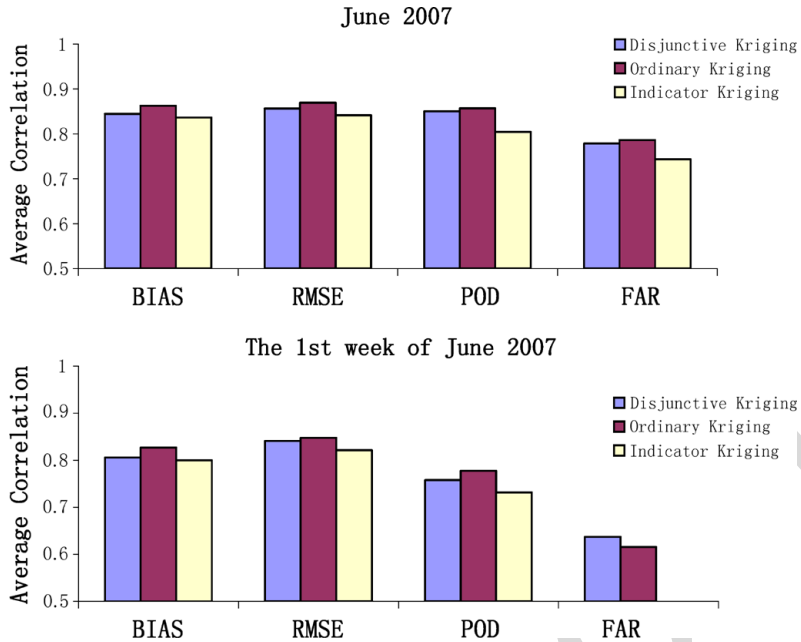


Fig. 7. Summary of performance of kriging methods as a function of error metric and timescale averaged over 10% to 90% of the region having GV gridboxes.

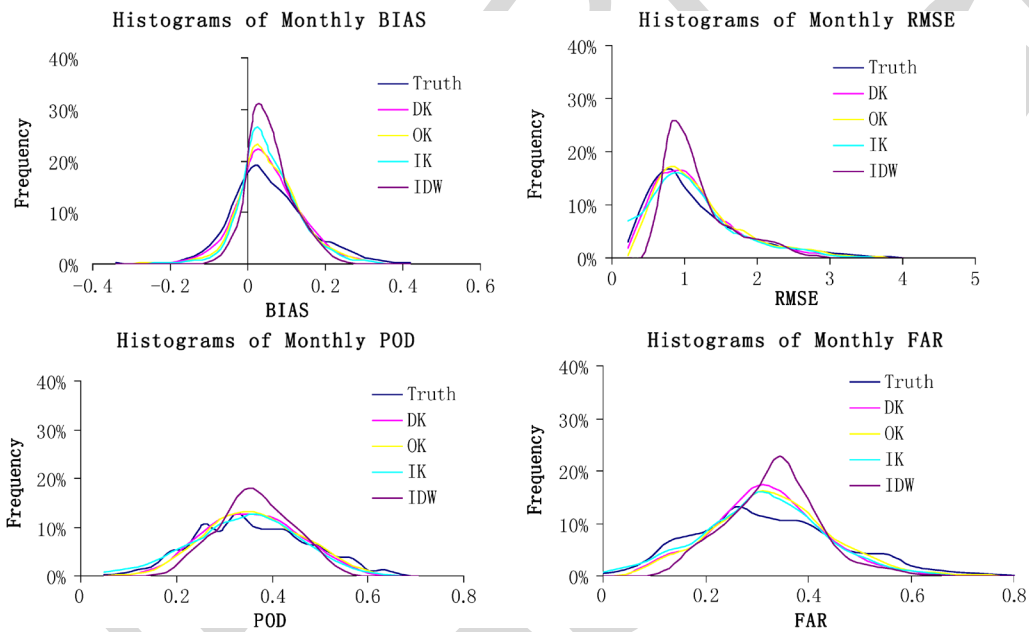


Fig. 8. Histogram estimated errors (from interpolated methods) and true error (measured from actual GV data) at the monthly timescale for the month of June 2007.

tion methods register only slightly lower correlation compared to monthly averaged error metrics. However, a clearer trend is observed when performance is scrutinized as a function of error metrics. The metric FAR appears to have the least accuracy in transfer followed by POD. When summarized over the range of 10%–90% of the domain as shown in Fig. 7, it is seen that the ranking of performance of kriging methods according to correlation measure is as follows: OK = DK > IK. At the weekly timescale, the difference in performance is significantly diminished.

Fig. 8 shows the histogram (pdfs) for various kriging methods and IDW at the monthly timescale for the month of June 2007 as an example. The histograms are averaged over the range of 10%–90% of the study domain missing in GV data. The unbiased nature of the kriging methods as well IDW is apparent from this figure. However, IDW is seen to have the histogram that deviates most from that of the true error, while other kriging methods show a much closer match. For BIAS, the performance of interpolation methods can be clearly ranked as follows: DK > OK > IK > IDW.

TABLE II(a)
ASSESSMENT OF THE TRANSFER OF ERROR METRICS AT MONTHLY TIME SCALES (FOR SUMMER MONTH OF JULY 2007)

Method	% of region lacking GV data	BIAS		RMSE		POD_{rain}		FAR	
OK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	0.54	0.71	0.20	0.20	0.16	0.15	0.32	0.43
	20	0.64	0.98	0.22	0.22	0.17	0.18	0.27	0.31
	30	0.67	0.98	0.22	0.21	0.16	0.19	0.28	0.32
	40	0.61	0.85	0.23	0.24	0.18	0.22	0.31	0.41
	50	0.67	1.01	0.25	0.24	0.18	0.21	0.31	0.39
	60	0.72	1.09	0.26	0.27	0.17	0.19	0.32	0.40
	70	0.80	1.13	0.26	0.24	0.19	0.24	0.33	0.44
	80	0.87	1.32	0.30	0.34	0.21	0.25	0.35	0.48
	90	0.99	1.46	0.33	0.34	0.24	0.30	0.39	0.51
DK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	0.57	0.79	0.19	0.20	0.17	0.19	0.31	0.40
	20	0.61	0.79	0.22	0.19	0.17	0.17	0.30	0.38
	30	0.68	0.92	0.24	0.24	0.18	0.20	0.31	0.38
	40	0.69	0.88	0.22	0.22	0.18	0.21	0.29	0.38
	50	0.71	1.05	0.22	0.21	0.18	0.21	0.30	0.37
	60	0.71	1.04	0.27	0.27	0.19	0.24	0.34	0.46
	70	0.75	1.08	0.29	0.30	0.20	0.21	0.33	0.42
	80	0.91	1.42	0.31	0.33	0.21	0.25	0.33	0.43
	90	2.46	7.82	0.36	0.35	0.26	0.32	0.41	0.55
IK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	0.57	0.75	0.24	0.23	0.17	0.16	0.34	0.44
	20	0.69	1.04	0.26	0.25	0.17	0.17	0.29	0.33
	30	0.70	1.00	0.24	0.21	0.17	0.19	0.29	0.33
	40	0.66	0.89	0.26	0.25	0.19	0.22	0.32	0.40
	50	0.72	1.05	0.27	0.25	0.19	0.21	0.32	0.38
	60	0.77	1.18	0.28	0.27	0.18	0.19	0.33	0.38
	70	0.84	1.15	0.28	0.25	0.20	0.22	0.34	0.40
	80	0.90	1.33	0.31	0.30	0.21	0.23	0.36	0.45
	90	1.05	1.51	0.35	0.28	0.24	0.24	0.38	0.43

In order to demonstrate a more rigorous level of accuracy of the interpolation methods beyond the correlation measure, Tables II(a) and II(b) show the mean assessment error and standard deviation of assessment error [(7) and (8), respectively] for the month of July 2007. Unlike correlation measure, the mean and standard deviation of error indicate a more revealing picture on the accuracy of the kriging methods. The error metric BIAS has the lowest accuracy for both monthly and weekly time scales. On the other hand, POD followed by RMSE and FAR have the highest accuracy for transfer of error metrics at both timescales according the mean assessment error measure. As expected, the precision of the kriging based transfer scheme degrades at shorter timescales. At very low GV coverage ($<20\%$), the standard deviation of transfer error becomes high ($>100\%$), indicating low reliability in the kriging methods regardless of the timescale at which the uncertainty metrics are transferred. Across methods, OK seems to be the best of the kriging methods where performance accuracy does not degrade as much as other methods when access to GV data decreases. The DK method, on the other hand, appears very unstable when GV coverage is very low with mean and standard deviation of assessment error doubling from 80% to 90% missing GV data scenario [Tables II(a) and II(b)].

The previous results were examples for a specific month (June or July) or a specific week (first week of June or July). Fig. 9(a) and (b) provide a generalization of the findings by presenting the performance of the interpolation methods averaged over the entire study period spanning June-July-August of 2007. These figures corroborate the major conclusions made so far, but present a much clearer picture. For example, both Fig. 9(a) and (b) show that the error metric POD has the least assessment error followed by FAR, RMSE, and BIAS. POD also has the least "spread" in estimation of error. This indicates that POD can be spatially transferred with the highest level of reliability and precision when compared to other common error metrics. The two figures also clearly show the added benefit of

TABLE II(b)
ASSESSMENT OF THE TRANSFER OF ERROR METRICS AT MONTHLY TIME SCALES (FOR THE FIRST WEEK OF JULY 2007)

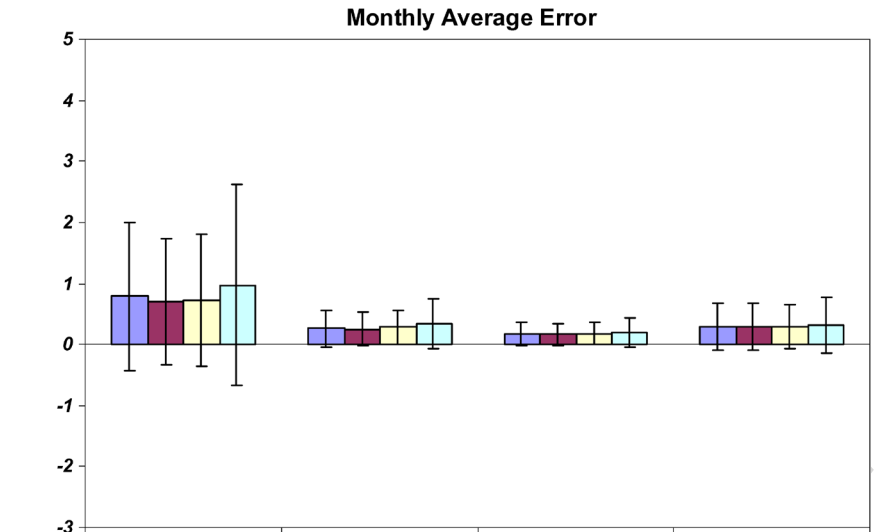
Method	% of region lacking GV data	BIAS		RMSE		POD_{rain}		FAR	
OK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	0.75	1.14	0.62	1.05	0.27	0.26	0.32	0.30
	20	0.81	1.15	0.54	0.90	0.29	0.27	0.32	0.27
	30	0.78	1.22	0.60	1.09	0.31	0.31	0.37	0.34
	40	0.87	1.27	0.63	1.09	0.31	0.32	0.36	0.35
	50	0.84	1.17	0.64	1.05	0.30	0.30	0.38	0.36
	60	0.85	1.11	0.71	1.24	0.31	0.31	0.37	0.38
	70	1.03	1.40	0.71	1.22	0.33	0.33	0.40	0.34
	80	0.89	1.10	0.80	1.25	0.34	0.36	0.39	0.37
	90	1.05	1.30	0.80	1.25	0.37	0.40	0.42	0.41
DK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	1.06	1.56	0.85	1.98	0.28	0.28	0.34	0.26
	20	1.15	1.71	0.86	1.96	0.29	0.29	0.36	0.32
	30	1.18	1.74	0.97	2.39	0.29	0.30	0.36	0.31
	40	1.29	2.08	1.02	2.28	0.29	0.28	0.36	0.31
	50	1.36	2.08	1.04	2.47	0.30	0.32	0.38	0.32
	60	1.38	2.14	1.05	2.84	0.32	0.34	0.39	0.36
	70	1.44	2.19	1.06	3.19	0.32	0.32	0.41	0.43
	80	1.61	2.34	1.66	5.88	0.37	0.39	0.43	0.39
	90	1.62	2.66	2.37	8.69	0.39	0.39	0.43	0.38
IK		Mean Error	Std. Dev of Error	Mean Error	Std. Dev Of Error	Mean Error	Std. Dev of Error	Mean Error	Std. Dev of Error
	10	0.79	1.11	0.62	0.89	0.35	0.29	0.49	0.35
	20	0.86	1.18	0.61	0.78	0.35	0.30	0.46	0.32
	30	0.85	1.15	0.62	0.83	0.36	0.31	0.50	0.36
	40	0.91	1.22	0.70	1.00	0.36	0.32	0.48	0.38
	50	0.90	1.17	0.67	0.90	0.36	0.32	0.51	0.38
	60	0.93	1.14	0.69	1.01	0.37	0.31	0.50	0.37
	70	1.06	1.38	0.70	1.00	0.38	0.30	0.51	0.34
	80	0.93	1.09	0.76	1.06	0.40	0.33	0.48	0.34
	90	1.07	1.26	0.73	1.06	0.44	0.37	0.57	0.37

using a more complex kriging method, as IDW is observed to have the highest interpolation uncertainty both in terms of the systematic (mean assessment error) and random nature (i.e., the standard deviation of assessment error shown as vertical error bars) at non-GV gridboxes.

VI. CONCLUSION

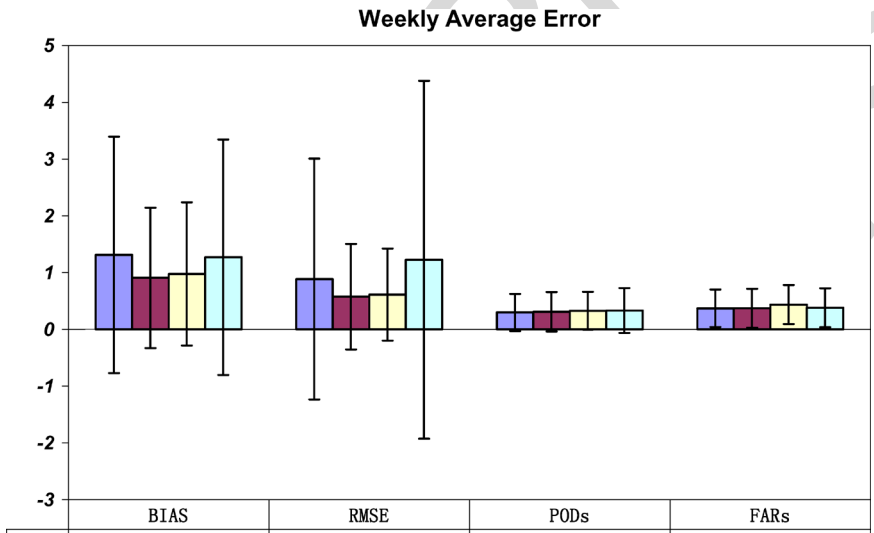
Although this study has presented several assessments, the key findings can be summarized as follows:

1) Correlation measure as an assessment metric for quantifying accuracy of interpolation methods for satellite rainfall error is largely inadequate. Any spatial transfer method should make use of scalar assessment measures such as mean and standard deviation of the difference between true and interpolated error metric.
2) There exists added benefit of using the more complex geostatistical methods of kriging as inverse distance weighting method yielded the highest mean and standard deviation of assessment error.
3) Although performance of disjunctive and ordinary kriging is similar, ordinary kriging is seen to be the better method across a range of scenarios of missing GV data and timescales. In general, indicator kriging (IK) is found to be the least effective of the three kriging methods studied.
4) Spatial transfer is strongly sensitive to the extent of gaugedness of the application domain. At weekly timescales, the assessment error is found to be almost twice as that of monthly timescale for the same percentage of a region having access to GV gridboxes.
5) Spatial transfer is found to be strongly sensitive to the error metric type with the highest accuracy in transfer observed for POD, followed by FAR, RMSE, and lastly BIAS. This indicates that in an operational setting, BIAS should be avoided for transfer from GV gridboxes, while POD can



	BIAS	RMSE	PODs	FARs
DK	0.7869	0.2625	0.1692	0.2886
OK	0.6933	0.2542	0.1644	0.2863
IK	0.7330	0.2839	0.1817	0.2923
IDW	0.9670	0.3373	0.1926	0.3171

(a)



	BIAS	RMSE	PODs	FARs
DK	1.3122	0.8863	0.2985	0.3723
OK	0.9086	0.5760	0.3094	0.3734
IK	0.9761	0.6133	0.3287	0.4373
IDW	1.2711	1.2247	0.3331	0.3808

(b)

Fig. 9. (a) Statistical average of mean and standard deviation of assessment errors over the entire study period spanning the 3 months of June-July-August of 2007 at the monthly timescale. The error bars indicate the standard deviation of assessment error averaged over the 3 months. (b) Statistical average of mean and standard deviation of assessment errors over the entire study period spanning the 12 weeks of June-July-August of 2007 at the weekly timescale. The error bars indicate the standard deviation of assessment error averaged over the 12 weeks.

probably be estimated at non-GV gridboxes with high accuracy.

The sobering finding from this study is that best linear unbiased spatial estimators may not be appropriate transfer methods for satellite rainfall error metrics at timescales of a week or shorter due to high interpolation errors. A more flexible kriging based scheme proposed by Jolly *et al.* [18], for efficient spatial interpolation of meteorological variables, may be worth pursuing as a future extension of this work. It is also worthwhile now to pursue more nonlinear transfer methods (such as neural networks) and other kriging methods that use additional spatial information (such as co-kriging) on the rainfall process and terrain to further constrain the prediction uncertainty [17].

TABLE III
CONTINGENCY TABLE
(A HIT IS DEFINED WHEN BOTH SATELLITE AND GV RAINFALL DATA AGREE ON THE TYPE OF EVENT DETECTED; A MISS IS WHEN THERE IS DISAGREEMENT BETWEEN SATELLITE AND GV DETECTED EVENTS)

Satellite Estimates \ Truth/Reference	*Rainy Gridboxes*	*Non-rainy Gridboxes*
	N_A (HIT)	N_B (MISS)
	N_C (MISS)	N_D (HIT)

In closing, given our current state of understanding, we believe that the operational scheme for "spatial transfer" of error metrics from gauged to ungauged regions should be able to perform the following two types of transfer: 1) for regions which have large voids of fixed-point GV data such as the tropics, transfer error on the basis of dynamic updates using more accurate satellite sensor data such as TRMM Precipitation Radar (PR) scans as a proxy for GV; 2) for regions having smaller voids in the fixed-point GV data such as in higher latitudes, transfer error on the basis of the stationary GV data (such as NEXRAD Stage IV for the US). The transfer scheme will need to automatically select the most appropriate technique and yield an error field comprising the calculated error over GV regions and the transferred error over non-GV regions. Users around the world would prefer to have a clear understanding of the pros and cons of applying satellite rainfall data for terrestrial applications at a given scale. Although spatial interpolation methods such as kriging have merit for such an effort, much work needs to be done to develop a transfer method that can work effectively at timescales shorter than a week in an operational setting for the Global Precipitation Measurement (GPM) mission.

APPENDIX
FORMULATION OF UNCERTAINTY METRICS

Consider the 2×2 contingency table, Table III, of hits and misses associated with satellite rainfall estimates.

Then the error metrics of probability of detection for rain (POD-rain) and False Alarm Ratio (FAR) are defined as follows:

$$\text{POD-rain}: \frac{N_A}{N_A + N_B} \tag{A.1}$$

$$\text{FAR}: \frac{N_B}{N_B + N_A}. \tag{A.2}$$

Bias is computed as the average of errors in a monthly or weekly period over the study domain as follows:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i) \tag{A.3}$$

where Y_i is satellite rainfall and X_i is the corresponding GV observation of rainfall. N is the number of data in the time period (in a month or a week).

RMSE measures the magnitude of error, giving greater weight to the larger error:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i)^2}. \tag{A.4}$$

REFERENCES

[1] A. Behrangi, K. Hsu, B. Imam, S. Sorooshian, G. J. Huffman, and R. J. Kuligowski, "PERSIANN-MSA: A precipitation estimation method from satellite-based multispectral analysis," *J. Hydrometeorol.*, vol. 10, pp. 1414–1429, 2009.

[2] C. Deutsch and A. Journel, *GSLIB: Geostatistical Software Library and User's Guide*. Oxford, U.K.: Oxford Univ. Press, 1998, 978-0195100150, 340 pp.

[3] C. Funk and J. Verdin, "Real-time decision support systems: The famine early warning system network," in *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds. New York: Springer, 2010, 978-90-481-2914-0.

[4] M. Gebremichael and F. Hossain, *Satellite Rainfall Applications for Surface Hydrology*. Heidelburg, Germany: Springer, 2009, 327pp.

[5] P. Goovaerts, "Geostatistical analysis of county-level lung cancer mortality rates in the southeastern US," *Geographical Analysis*, vol. 42, pp. 32–52, 2010.

[6] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. New York: Oxford Univ. Press, 1997, 483 pp.

[7] A. Harris and F. Hossain, "Investigating optimal configuration of conceptual hydrologic models for satellite rainfall-based flood prediction for a small watershed," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, Jul. 2008.

[8] F. Hossain, A. J. Hill, and A. C. Bagtzoglou, "Geostatistically-based management of arsenic contaminated ground water in shallow wells of Bangladesh," *Water Resources Management*, vol. 21, pp. 1245–1261, 2007, doi: 10.1007/s11269-006-9079-2.

[9] F. Hossain and D. P. Lettenmaier, "Flood prediction in the future: Recognizing hydrologic issues in anticipation of the global precipitation measurement mission," *Water Resources Research*, vol. 44, 2006, doi: 10.1029/2006WR005202.

[10] F. Hossain and E. N. Anagnostou, "Assessment of current passive microwave and infra-red based satellite rainfall remote sensing for flood prediction," *J. Geophysical Research*, vol. 109, no. D7, p. D07102, Apr. 2004, doi: 10.1029/2003JD003986.

[11] A. Hou, C. S. Jackson, C. Kummerow, and C. M. Shepherd, "Global precipitation measurement," in *Precipitation: Advances in Measurement, Estimation, and Prediction*, S. Michaelides, Ed. New York: Springer, 2008, pp. 1–39.

[12] G. Hudson and H. Wackernagel, "Mapping temperature using kriging with external drift: Theory and an example from Scotland," *Int. J. Climatol.*, vol. 14, no. 1, pp. 77–91, 1994.

[13] G. J. Huffman, R. F. Adler, D. T. Bolvin, and E. J. Nelkin, "The TRMM Multi-satellite Precipitation Analysis (TMPA)," in *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds. New York: Springer, 2010, 978-90-481-2914-0.

[14] G. J. Huffman, R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff, "The TRMM multi-satellite precipitation analysis: Quasi-global, multi-year, combined sensor precipitation estimates at fine scales," *J. Hydrometeorol.*, vol. 8, pp. 28–55, 2007.

[15] G. J. Huffman, R. F. Adler, and M. M. Morrissey *et al.*, "Global precipitation at one-degree daily resolution from multisatellite observations," *J. Hydrometeorol.*, vol. 2, pp. 36–50, 2001.

[16] R. L. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, "CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *J. Hydrometeorol.*, vol. 5, pp. 487–503, 2004.

[17] W. F. Krajewski, "Cokriging radar-rainfall and rain gage data," *J. Geophysical Research*, vol. 92, no. D8, pp. 9571–9580, 1987, doi: 10.1029/JD092iD08p09571.

[18] W. M. Jolly, J. M. Graham, A. Michaelis, R. Nemani, and S. W. Running, "A flexible, integrated system for generating meteorological surfaces derived from point sources across multiple geographic scales," *Environmental Modelling & Software*, vol. 20, no. 7, pp. 873–882, 2004.

[19] Y. Lin and K. Mitchell, "The NCEP Stage II/IV hourly precipitation analyses: Development and applications," presented at the 19th AMS Conf. Hydrology, San Diego, CA, 2005.

[20] T.-S. Ma, M. Sophocleous, and Y.-S. Yu, "Geostatistical applications in ground-water modeling in south-central Kansas," *ASCE J. Hydrol. Eng.*, vol. 4, no. 1, pp. 57–64, 1999.
[21] G. Matheron, "The Theory of Regionalized Variables and Its Applications," (in French) Centre de Morphologie Mathématique, Ecole des Mines de Paris, 5, Fontainebleau, France, 1971.
[22] C. Reynolds, "Real-time hydrology operations at USDA for monitoring global soil moisture and auditing national crop yield estimates," in *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds. New York: Springer, 2010, 978-90-481-2914-0.
[23] R. A. Scofield and R. J. Kuligowski, "Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events," *Weather and Forecasting*, vol. 18, pp. 1037–1051, 2003.
[24] L. Tang and F. Hossain, "Transfer of satellite rainfall error from gauged to ungauged locations: How realistic will it be for the global precipitation mission?," *Geophys. Res. Lett.*, vol. 36, 2009, doi: 10.1029/2009GL037965.
[25] N. Theodossiou and P. Latinopoulos, "Evaluation and optimisation of groundwater observation networks using the Kriging methodology," *Environmental Modelling and Software*, vol. 21, no. 7, pp. 991–1000, 2006.
[26] T. Ushio, K. Sasashige, T. Kubota, S. Shige, K. Okamoto, K. Aonashi, T. Inoue, N. Takahashi, T. Iguichi, M. Kachi, R. Oki, T. Moromoto, and Z.-I. Kawasaki, "A Kalman filter approach to the global satellite mapping of precipitation (GSMap) from combined passive microwave and infrared radiometric data," *J. Meteorol. Soc. Japan*, vol. 87A, pp. 137–151, 2009.
[27] P. Xie and P. A. Arkin, "Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs," *Bull. Amer. Meteorol. Soc.*, vol. 78, pp. 2539–2558, 1997.
[28] L. Tang, F. Hossain, and G. J. Huffman, "Transfer of satellite rainfall error from gauged to ungauged regions at regional and seasonal timescales," *J. Hydrometeorol.*, vol. 11, pp. 1263–1274, 2010, doi: 10.1175/2010JHM1296.1.
[29] F. Hossain and G. J. Huffman, "Investigating error metrics for satellite rainfall at hydrologically relevant scales," *J. Hydrometeorol.*, vol. 9, no. 3, pp. 563–575, 2008.

Ling Tang received the B.S. degree in geomatics from the Geotechnical Engineering and Surveying Institute, Wuhan University, China, in 2003. She received the M.S. degree in photogrammetry and remote sensing from the same university in 2006. Currently, she is pursuing the Ph.D. degree in civil engineering at Tennessee Technological University, Cookeville, TN.

Ms. Tang is the recipient of the 2008 NASA Earth System Science Fellowship.

Faisal Hossain received the B.S. degree from the Indian Institute of Technology, Varanasi, India, and the M.S. degree from the National University of Singapore. He received the Ph.D. degree from the University of Connecticut, Storrs, CT, in 2004.

He is currently an Associate Professor in the Department of Civil and Environmental Engineering at Tennessee Technological University, Cookeville, TN. His research interests include hydrologic remote sensing, uncertainty modeling of water cycle variables, human modification of extreme hydro-climatology, sustainable water resources engineering, space-borne transboundary water resources management, engineering education and public outreach.

Dr. Hossain is the recipient of such awards as the NASA Earth System Science Fellowship (2002), Outstanding Ph.D. Thesis Award (2005), NASA New Investigator Award (2008), American Society of Engineering Education (ASEE) Outstanding Research Award (2009), and National Association of Environmental Professionals Education Excellence Award (2010). He has published over 70 peer-reviewed journal articles, authored an undergraduate-level textbook, edited one book volume and contributed five book chapters. Currently he serves as the Water volume editor for Elsevier Sciences five-volume reference series on Climate Vulnerability: Understanding and Addressing Threats to Essential Resources.