3B2v8.06a/w (Dec 5 2003).51c +model CAGEO : 1618 Prod.Type:FTP pp.1–17(col.fig.:NIL) ED:JanaganP. PAGN:Jay SCAN:Mouli

ARTICLE IN PRESS

ELSEVIER

Computers & Geosciences ▮ (▮▮▮▮) ▮▮▮–▮▮▮

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

On Latin Hypercube sampling for efficient uncertainty estimation of satellite rainfall observations in flood prediction

Faisal Hossain[a,*], Emmanouil N. Anagnostou[b], Amvrossios C. Bagtzoglou[b]

[a]*Department of Civil and Environmental Engineering, Tennessee Technological University, Cookeville, TN 38505-0001, USA*
[b]*Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA*

Received 2 April 2005; received in revised form 13 September 2005; accepted 15 October 2005

Abstract

With the advent of the Global Precipitation Measurement (GPM) in 2009, satellite rainfall measurements are expected to become globally available at space–time scales relevant for flood prediction of un-gauged watersheds. For uncertainty assessment of such retrievals in flood prediction, error models need to be developed that can characterize the satellite's retrieval error structure. A full-scale assessment would require a large number of Monte Carlo (MC) runs of the satellite error model realizations, each passed through a hydrologic model, in order to derive the probability distribution in runoff. However, for slow running hydrologic models this can be computationally expensive and sometimes prohibitive. In this study, Latin Hypercube Sampling (LHS) was implemented in a satellite rainfall error model to explore the degree of computational efficiency that could be achieved with a complex hydrologic model. It was found that the LHS method is particularly suited for storms with moderate rainfall. For assessment of errors in time to peak, peak runoff, and runoff volume no significant computational advantage of LHS over the MC method was observed. However, the LHS was able to produce the 80% and higher confidence limits in runoff simulation with the same degree of reliability as MC, but with almost two orders of magnitude fewer simulations. Results from this study indicate that a LHS constrained sampling scheme has the potential to achieve computational efficiency for hydrologic assessment of satellite rainfall retrievals involving: (1) slow running models (such as distributed hydrologic models and land surface models); (2) large study regions; and (3) long study periods; provided the assessment is confined to analysis of the large error bounds of the runoff distribution.

Keywords: Satellite rainfall estimation; Retrieval uncertainty; Hydrologic assessment; Monte Carlo simulation; Latin Hypercube Sampling

1. Introduction

The Global Precipitation Measurement (GPM), which is a mission to be launched by the international community by 2009, envisions a large constellation of Passive Microwave (PM) sensors to provide global rainfall products at scales ranging from 3 to 6 h, and spatial resolution of 100 km^2 (Smith, 2001; Bidwell et al., 2002; Flaming, 2002; Yuter et al., 2003). These resolutions offer tremendous opportunities to address the problem of flood prediction in un-gauged watersheds over the globe. Nevertheless, satellite rainfall retrieval is subject to errors caused by various factors ranging from

*Corresponding author. Tel.: +1 9313723257; fax: +1 9313726239.
E-mail address: fhossain@tntech.edu (F. Hossain).

0098-3004/$ - see front matter
doi:10.1016/j.cageo.2005.10.006

sampling error to high complexity and variability in the relationship of the measurement to precipitation parameters. The presence of such errors in remote sensing of rainfall can potentially lead to high uncertainties in runoff simulation (Winchell et al., 1998; Borga et al., 2000; Hossain et al., 2004a). Thus, it is important to assess the uncertainty in satellite rainfall observations in order to better evaluate the utility of the GPM for flood prediction.

Conventional uncertainty assessment of such space-based rainfall observations requires the derivation of the probability distribution of runoff by combining the following three components: (1) a probabilistically formulated satellite rainfall error model; (2) a deterministic or probabilistic hydrologic model for the rainfall-runoff transformation; and (3) Monte Carlo (MC) framework linking (1) and (2)—see papers by Hossain et al. (2004a,b) and Hossain and Anagnostou (2004) describing this problem. The MC sampling technique, due to absence of restrictive assumptions and completeness in sampling the input error structure, is generally considered the preferred method for uncertainty assessment (Beck, 1987; Kremer, 1983; Isukapalli and Georgopoulos, 1999). Recent satellite rainfall error studies in hydrologic prediction have utilized the MC technique to conduct random simulations on error propagation. Hossain et al. (2004b) and Hossain and Anagnostou (2004) have devised a MC technique on the basis of a Satellite Rainfall Error Model (SREM) and a topographically driven hydrologic model (TOPMODEL) to assess PM's retrieval and sampling error on flood prediction uncertainty. The SREM statistically characterized the sensor's success in discriminating rain from no-rain, and quantified the structure of the sensor's rainfall retrieval error at the sensor resolution using '*reference*' rainfall data from more definitive sources (see also, Hossain and Anagnostou, 2005a,b). This MC technique involving the SREM can work in conjunction with any deterministic hydrologic model without imposing on that model any structural or distributional assumptions. Another study by Nijssen and Lettenmaier (2004), which focused primarily on satellite rainfall sampling error, used an error model proposed by Steiner et al. (2003) and a macro-scale hydrologic model (Variable Infiltration Capacity, VIC model) within a MC-based random experiment to evaluate the GPM rain retrieval error propagation in hydrologic predictions.

However, definitive rules for determining the number of simulations required for convergence of the MC technique are not available (Melching, 1995) and are strongly dependent on the nature of the problem (Beck, 1987). Siddall (1983) suggested that MC simulations should require at least 5000–20,000 repetitive model runs. For slow running hydrologic models such as fully distributed and physically based models and macro-scale land surface models which simultaneously balance the water and energy budgets, such MC assessment can be computationally prohibitive. This makes the hydrologic assessment of satellite rainfall data limited to mainly fast running conceptually lumped hydrologic models or to regions that are either small in size ($<500\,\mathrm{km}^2$) or involve a short study period (<500 time steps). For example, the study by Nijssen and Lettenmaier (2004) was restricted to 1000 MC simulations of the VIC model at a large scale ($>500\,\mathrm{km}^2$) spanning 6 years at the daily time step (>500 time steps).

Therefore, a broader uncertainty assessment of satellite rainfall observations across increasing levels of hydrologic model complexity warrants the investigation of computationally more efficient sampling schemes. Such schemes could potentially achieve greater flexibility in the following: (1) design of simulation experiments to assess satellite rainfall retrievals; (2) choice of hydrologic and land surface models; and (3) choice of study regions and time period. A broad-based assessment of satellite rainfall observations may also have long-term implications for the well-known argument proposed by Krzysztofowicz (1999, 2001) that in short-term forecasting of floods, the principal source of uncertainty is the unknown future rainfall, which should therefore be treated as a random input. The recent methodologies developed for quantifying predictive uncertainty of remote sensing retrievals (Grecu and Krajewski, 2000; Seo et al., 2000, among others) now offer tremendous opportunities to explore the development of probabilistic forecasting schemes for surface hydrologic processes that have been argued as the way forward to reduce the inherent uncertainty in our geosystems. Because probabilistic schemes are usually based on MC model runs, any computationally efficient statistical sampling scheme for rainfall will always be in contention for incorporation into an operational probabilistic technique.

Latin Hypercube Sampling (LHS) is one such technique that offers promise in reducing the computational burden of MC uncertainty assessment. The LHS technique is a constrained sampling technique usually known to simulate uncertainty as accurately as a MC sampling method while using an order of magnitude fewer samples (McKay et al., 1979; Iman and Conover, 1980; Loh, 1996). LHS has therefore found application in a wide range of uncertainty assessment problems in environmental modeling. Isukapalli and Georgopoulos (1999) have summarized previous work on the application of LHS. However, like most complex sampling techniques, LHS is based on the assumption of monotonicity of model output in terms of input parameters, in order to be unconditionally guaranteed of accuracy with an order of magnitude fewer runs than MC sampling (McKay et al., 1979; Iman et al., 1981). Previous simulation studies by Hossain et al. (2004b) have clearly demonstrated that the response surface of the runoff simulation error (in terms of peak runoff, runoff volume and time to peak) is not always a monotonic function of the satellite's retrieval error parameters (such as bias or error variance). Thus, for any new application, such as flood prediction uncertainty based on satellite rainfall observations, LHS needs to be carefully verified of its effectiveness, before the method can be used confidently.

This study aims at investigating the use of LHS for efficient uncertainty analyses of satellite rainfall measurements for flood prediction. The specific question that this study seeks to answer is—*is it possible to infer similar uncertainty statistics in runoff using a LHS scheme as those derived with MC sampling but with fewer simulations*? The study is organized in the following manner. We first describe the watershed, data, and hydrologic model (Section 2). Section 3 describes the satellite error model and is followed by the description of the LHS scheme in Section 4. In Section 5, we present the simulation framework, while the last two sections (6 and 7) discuss the results and conclusions of this study.

2. Watershed, data and hydrologic model

The watershed chosen for this study (the Posina Watershed) is located in northern Italy, close to the city of Venice (Fig. 1, right panel). Posina has an area of 116 km^2 and altitudes ranging from 2230 to 390 m at the outlet (Fig. 1, left panel). Within a radius of 10 km from the center of the watershed there exists a network of 7 rain gauges providing representative estimates of the basin-averaged hourly rainfall (hereafter referred to as '*reference rainfall*'). The estimation of basin-averaged rainfall was based on an inverse distance weighting techni-

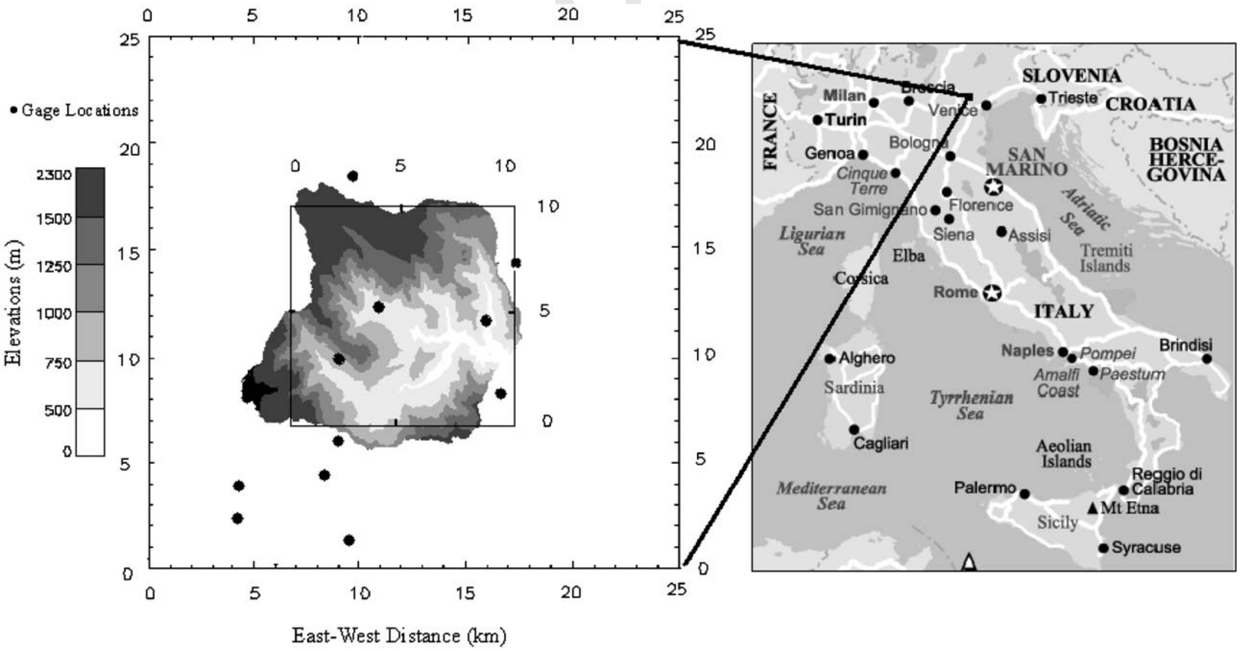

Fig. 1. Geographic location of Posina Watershed (right panel), and watershed elevation map (left panel) overlaid by rain gauge network locations (in solid circles) within 25 and 10 km grids that are equivalent to a typical satellite footprint.

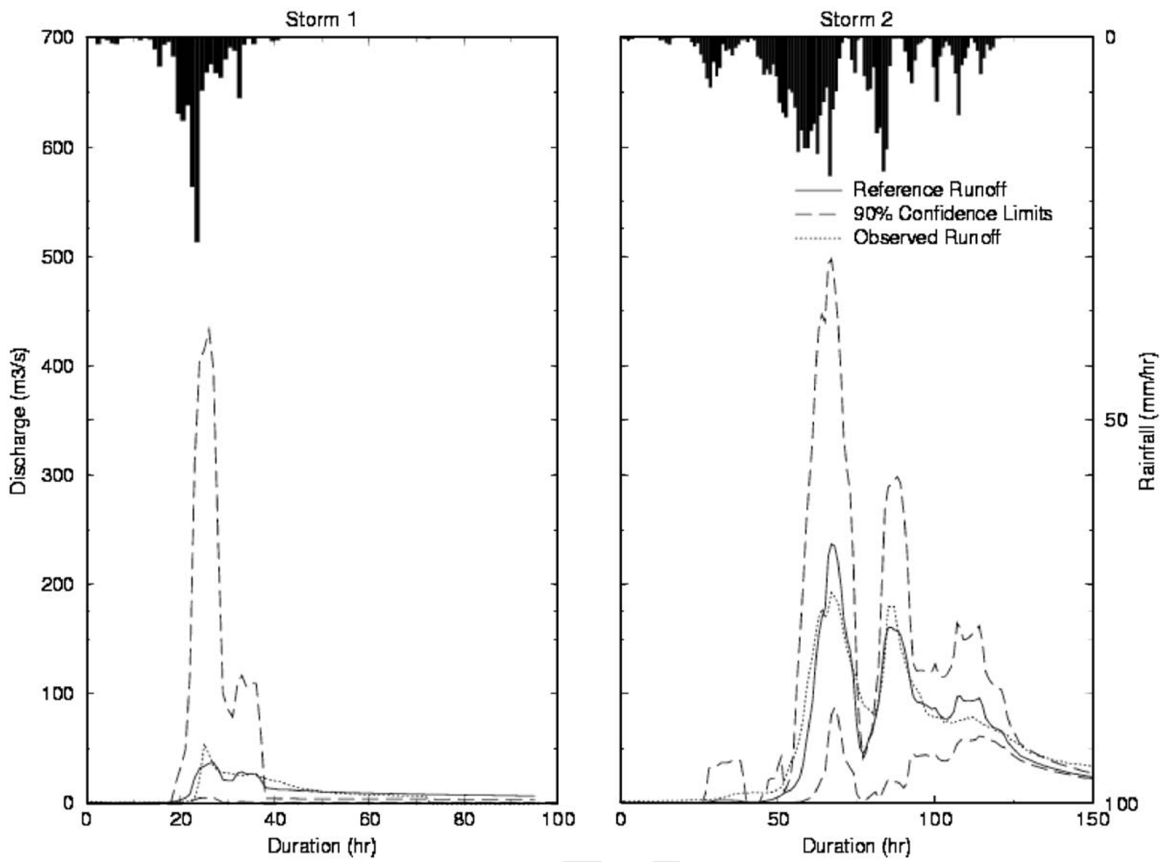

Fig. 2. Two selected storm cases for LHS study. Left panel—Storm 1 (August 1987); right panel—Storm 2 (October 1992). Upper axis represents basin averaged rainfall hyetograph measured by gauge network. Reference runoff (solid line) represents the simulated hydrograph using reference rainfall (gauge). Long dashed lines represent runoff uncertainty at 90% confidence limits using 20,000 MC simulations of SREM passed through TOPMODEL.

Table 1
Morphological summary of two selected storm cases over Posina Watershed

Storm no.	Date	Duration (h)	Total rainfall (mm)	Maximum rainrate (mm/h)	Fraction of rain (%)	Peak discharge (m^3/s)
1	August 1987	72	127.4	26.7	34.0	54.4
2	October 1992	120	440.3	18.0	86.7	192.5

que that had earlier proved to be a reliable method for similar hydrologic studies over Posina (Borga et al., 2000; Dinku et al., 2002; Hossain et al., 2004a). The annual precipitation accumulation is estimated to be in the range of 1600–1800 mm. The Posina Watershed is 68% forested, thereby rendering saturation-excess the main rainfall-runoff generation mechanism of the basin.

Two storm events of contrasting morphological properties were chosen for this study (Fig. 2). The first storm (referred to as Storm 1) represents a mild event that took place in August 1987 and produced moderate flooding (peak discharge was 54.4 m^3/s). It was associated with an isolated precipitation pattern where the basin witnessed rain during 34% of the total hours (referred to as % *Rainy* in Table 1). The second storm (referred to as Storm 2) was a major storm event that took place in October 1992 and was associated with catastrophic flooding (peak discharge was 192.5 m^3/s). It was associated with a widespread precipitation pattern (% *Rainy*—86.7). Fig. 2 shows the storm hydrographs (lower axis) and the corresponding hourly basin-averaged gauge rainfall hyetograph (upper axis). The gauge-derived runoff simulation (referred to as *reference runoff*) is shown in solid lines. In Table 1, we summarize the morphological features of the two storm events. Further details about the study area, including its terrain characteristics and rain climatology can be found in Borga et al. (2000) and Bacchi et al. (1996).

The Topographic Index Model (TOPMODEL) (Beven and Kirkby, 1979) was chosen to simulate the rainfall–runoff processes of the Posina watershed. TOPMODEL is a semi-distributed watershed model that can simulate the variable source area mechanism of storm-runoff generation and incorporates the effect of topography on flow paths. TOPMODEL makes a number of simplifying assumptions about the runoff generation processes that are thought to be reasonably valid in this wet, humid watershed. The model is premised on the following two assumptions: (1) the dynamics of the saturated zone can be approximated by successive steady state representations; and (2) the hydraulic gradient of the saturated zone can be approximated by the local surface topographic slope. The topographic index $\ln(a/\tan\beta)$ is used as an index of hydrologic similarity, where a is the area draining through a point, and $\tan\beta$ is the local surface slope. The use of this form of topographic index implies an effective transmissivity profile that declines exponentially with increasing storage deficits. In this study, the derivation of the topographic index from a 20 m grid size catchment digital terrain model utilized the multiple flow direction algorithm by Quinn et al. (1991, 1995). For the case of unsaturated zone drainage, a simple gravity-controlled approach is adopted in the TOPMODEL version used here, where a vertical drainage flux is calculated for each topographic index class using a time delay based on local storage deficit. The generated runoff is routed to the main channel using an overland flow delay function. The main channel routing effects are considered using an approach based on an average flood wave velocity for the channel network. The model was run at hourly time steps for the rainfall–runoff transformation. The model has been applied in the study region by previous work of Borga et al. (2000). Model parameters were calibrated using reference rainfall and the optimization routine of Duan et al. (1992). Detailed background information of the model and applications can be found in Beven et al. (1995). The model has been successfully applied in the study region as demonstrated by previous hydrologic studies (Borga et al., 2000; Hossain et al., 2004a,b).

3. Satellite rainfall error model

The motivation for the formulation of a SREM comes from the need to fully characterize the retrieval error of satellite sensors at high resolutions so that it can be linked to a hydrologic model to assess the retrieval error propagation in runoff. In this study we adopted a probabilistic error model originally developed by Hossain et al. (2004b) and subsequently applied by Hossain and Anagnostou (2004) for point (1-D, lumped in space) error propagation studies. Very recently, Hossain and Anagnostou (2005a) formulated a fully two-dimensional (2D) Space–Time Error Model for SREM (called SREM2D) for distributed (spatial) error propagation studies (e.g., Hossain and Anagnostou, 2005b). The 1D error model is schematically presented as a flow chart in Fig. 3 and details are discussed below.

The approach is to simulate equally likely statistical realizations of satellite rainfall (PM) retrievals by corrupting a more accurate measurement of rainfall. In this study, the most accurate measurement of rainfall constituted the basin-averaged hourly rainfall rate derived from a dense network of rain gauges in the vicinity of the Posina basin (earlier labeled as *reference rainfall*). At any time during a storm event a satellite rain retrieval may exhibit the following possible outcomes: it can be zero (false no-rain detection) or non-zero (successful rain detection) when it actually rains, whereas when it does not rain the satellite retrieval can be zero (successful no-rain detection) or non-zero (false rain detection).

We define the successful rain detection probability, P_1, as a function of the *reference rainfall*. Therefore, the false no-rain detection is $1-P_1$. The successful no-rain detection, P_0, is the unitary probability that satellite retrieval is zero when *reference rainfall* is zero. The false rain detection probability is then $1-P_0$. A probability density function (D_{false}) is introduced to characterize the probability distribution of the satellite rain rate retrieval in false rain detection. The study reported by Hossain and Anagnostou (2004, 2005a) to characterize the error structure of PM and Infrared

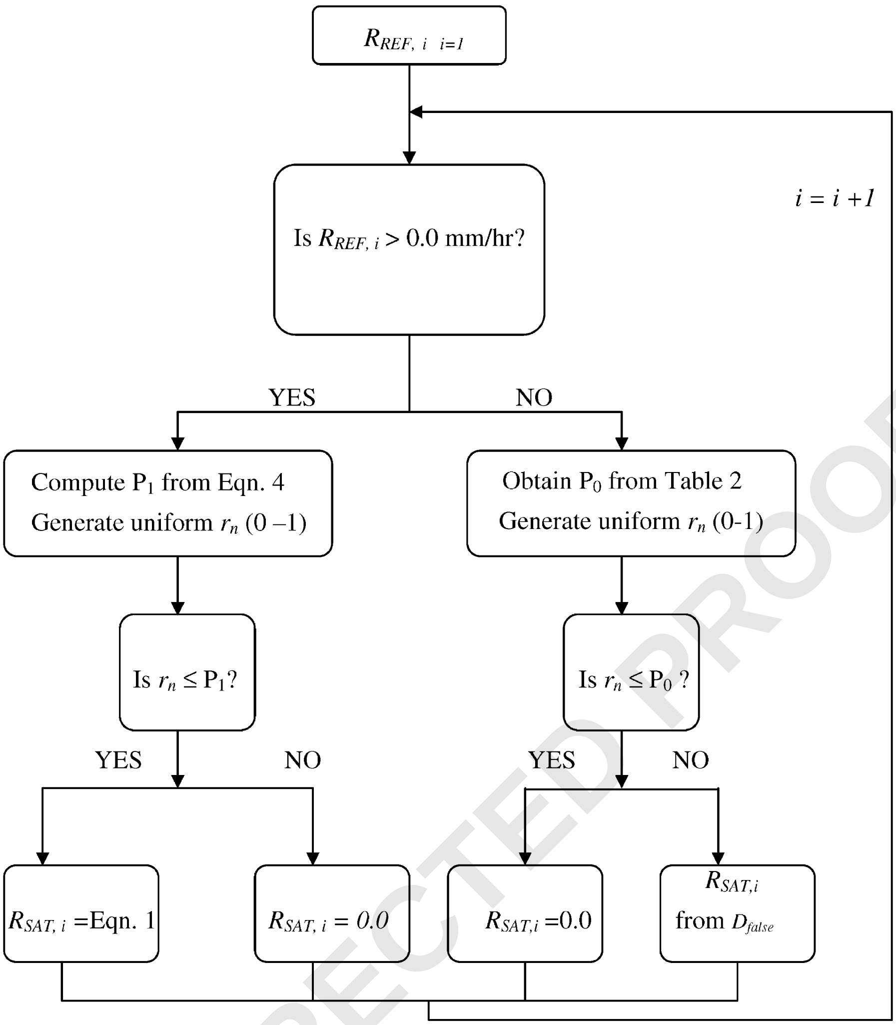

Fig. 3. Satellite Rainfall Error Model (SREM) algorithmic structure (after Hossain et al., 2004b). r_n is a randomly generated number from a uniform [0–1] probability distribution.

(IR) sensors based on real sensor data, had found D_{false} to be exponentially distributed. Hence, D_{false} was modeled as an exponential probability distribution function, $D_{false}(R_{SAT}) = \lambda\ \exp\{-\lambda R_{SAT}\}$.

The non-zero satellite rain retrieval, R_{SAT}, is statistically related to the corresponding non-zero *reference rainfall*, R_{REF}, by

$$R_{SAT} = R_{REF}\varepsilon_S, \tag{1}$$

where the multiplicative satellite error parameter, ε_s, is assumed log-normally distributed. A multiplicative error model is used on the basis of the assumption that the retrieval error variance varies as a function of the rainfall rate. Such an assumption has been found to be representative of the retrieval error analyses reported earlier by Hossain and Anagnostou (2004, 2005a). The log-normality of the distribution is suggested by the non-negative property of ε_s (Hossain and Anagnostou, 2005a). A logarithmic transformation of the $\log(R_{SAT})$–$\log(R_{REF})$ statistical relationship transforms the error ε_s to a Gaussian deviate ε with $N(\mu,\sigma)$ statistics, where μ and σ are the mean and standard deviation, respectively. To determine the multiplicative mean (mu) and standard deviation (S) of ε_s the following conversion is used in terms of μ and σ,

$$\mathrm{mu} = \exp(\mu + 0.5\sigma^2), \tag{2}$$

$$S^2 = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2). \tag{3}$$

The error parameter ε (hereafter also referred to as 'log-error') can be spatially and temporally autocorrelated. Only temporal autocorrelation is considered in this study because the watershed scale is represented by a single satellite retrieval pixel ($\approx 100\,\text{km}^2$). The reader is referred to the study of Hossain et al. (2004b) for further details on the mathematical formulation and the hydrologic assessment of temporal autocorrelation of precipitation retrieval error on runoff simulation uncertainty.

The SREM operation is summarized in the flow chart of Fig. 3. When at a certain time (hour) the *reference rainfall* is non-zero ($R_{REF}>0.0$) the model decides as to whether the satellite rainfall is non-zero or zero through Bernoulli trials as follows. First, a uniformly distributed random number, r_n, is generated from $U[0–1]$. If r_n is less than P_1 (which is determined as function of R_{REF}) then the satellite retrieval R_{SAT} is non-zero and modeled through Eq. 1. Otherwise, R_{SAT} is assigned a zero value. Similarly, at a non-rainy time ($R_{REF}=0.0$) a Bernoulli trial is used again to decide whether the satellite rainfall will be zero or non-zero. If the uniformly distributed random deviate r_n is less than P_0, then R_{SAT} is assigned a zero value. Otherwise, the non-zero satellite rainfall value is determined through random sampling on the basis of the false alarm probability density (D_{false}) function.

In this study, we considered only PM sensors as they will comprise the major backbone of the GPM plan. Furthermore, we did not consider any PM sampling error and assumed that satellite overpasses are available every hour over the Posina Watershed during a storm event. This may be an optimistic assumption, but is deemed acceptable, as the main purpose of this study is to assess the performance of the LHS scheme in runoff error simulation. The relevant PM retrieval error parameters used in this study were obtained from the calibration exercise reported by Hossain and Anagnostou (2004) that used the Tropical Rainfall Measuring Mission's Microwave Image (TRMM-TMI) as the primary PM sensor for GPM. The probability of rain detection, P_1, was modeled as a sigmoidal function of R_{REF} as follows:

$$P_1(R_{REF}) = \frac{1}{A+\exp(-BR_{REF})}. \tag{4}$$

Table 2 summarizes the PM sensor's retrieval error parameters.

Table 2
Mean error model parameters calibrated for PM sensor retrievals on basis of coincident TRMM precipitation radar rainfall fields (after Hossain and Anagnostou, 2004)

Retrieval error parameter	Value
A	1.0
B	3.5
λ	0.9
Bias (mu)	1.27
Std. dev of log-error (S)	0.94
No rain detection probability (P_0)	0.93

4. The LHS technique

4.1. LHS formulation on the SREM

The LHS technique is a constrained sampling technique whereby the input parameter range is divided into equi-probable non-overlapping intervals. This way, we try to explore the parameter space as completely and with as few samples as possible. For example, if a parameter is uniformly distributed $U[A–B]$, we could divide its range (from A to B) into N intervals and perform a LHS as follows:

$$\begin{aligned} P_m = {} & [U(0,1)\times((A-B)/N)] \\ & +[(m-1)\times((A-B)/N)], \\ & m=1,2,3,\ldots,N. \end{aligned} \tag{5}$$

Here P_m is the cumulative probability value used with the inverse distribution to produce the specific parameter value to be used with LHS. Fig. 4 shows an example of such constrained sampling for $N=5$ along with a comparison with the Monte Carlo (simple random) sampling. The LHS technique can handle a wide variety of complexity such as parameter correlation and random pairing of parameter sets. For further details about the LHS technique the reader is referred to McKay et al. (1979), Iman and Shortencarier (1984), Stein (1987), and Isukapalli and Georgopoulos (1999).

We first explored the use of LHS at all possible sampling instances in the SREM algorithm where MC random sampling was used. These instances included (see flow-chart Fig. 3): (1) modeling the probabilities of successful detection for rain and no-rain events by random Bernoulli trials; (2) sampling from false alarm distribution during false alarms (false no-rain detection); (3) sampling from satellite rainfall retrieval error distribution (Eqs. (2) and (3)) during successful rain detection. However, given the

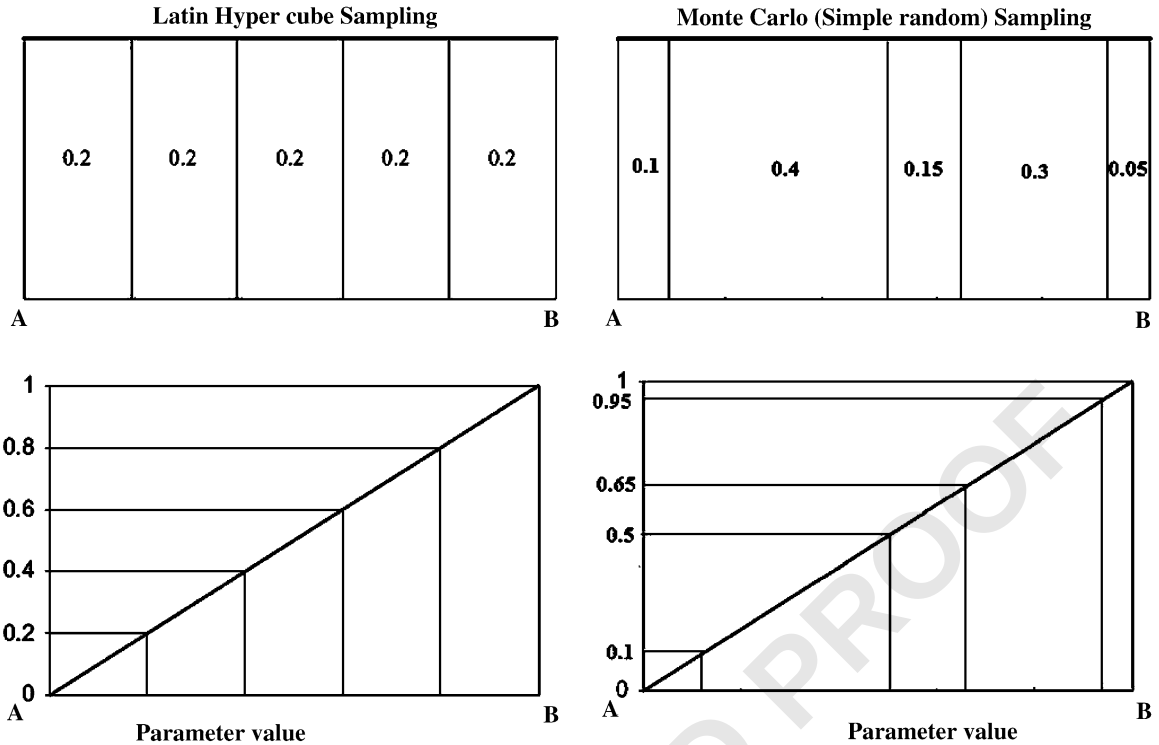

Fig. 4. Intervals used with a Latin Hypercube sample (left panels) and Monte Carlo (simple random) sample of size $N = 5$ in terms of the density function (upper panels) and cumulative distribution function (lower panels) for a uniform random variable $U[A–B]$.

complex nature of the SREM algorithm for which the probability of rain detection varies temporally according to the reference rainfall rate, our preliminary investigations have revealed that constrained sampling of the Bernoulli trials by the LHS technique was most notable aspect that yielded results consistent with the MC simulations. At this stage of our work, we do not fully know the reasons for LHS failing at other instances. While this limited use of LHS may raise concerns, which are understandable, we would also like to emphasize that there is no convincing reason not to explore the efficacy of LHS in this manner that has not been attempted before in literature concerning satellite-based hydrology to the best of our knowledge. Consequently, we used the concept of constrained sampling by LHS (Eq. (5)) in SREM for efficient modeling of rain and no-rain detection (see Fig. 3 for SREM flowchart). For rain detection ($R_{REF}>0.0$), the uniformly distributed $U[0–1]$ random number, r_n, was divided into non-overlapping intervals equal to the number of rainy hours (i.e., 25 and 104 for Storms 1 and 2, respectively). Similarly, for no rain detection ($R_{REF}=0.0$), the number of non-overlapping intervals within the $U[0-1]$ r_n was equal to the number of non-rainy hours (i.e., 47 and 16 for Storms 1 and 2, respectively). Given that the number of times Bernoulli trials are conducted for modeling rain and no-rain events are equal to the number of rainy and non-rainy hours, respectively, such a discretization ensures the complete sampling of the $U[0–1]$ space within each satellite realization of a storm event (refer to Fig. 4 for a conceptual elaboration).

4.2. LHS technique validation

Using 30 independent sets of seed numbers for random number generation, we compared the performance of the LHS technique with MC sampling on the rainfall retrieval aspect (i.e., excluding the use of hydrologic model). The use of a large number of seeds allowed us to analyze the variability in simulations and defuse any bias due to the choice of a specific set of seeds. The number of SREM runs varied from 10 to 20,000 in regular increments. For a given number of SREM realizations, the retrieval error parameters were derived inversely for each sampling method (i.e., MC and LHS). This means that the ensembles of satellite-like rainfall observations generated by MC or LHS embedded SREM realizations were used to compute *independently* the retrieval error parameters using the *reference rainfall* as the truth. Because the mean and standard deviation of an exponential distribution have value (1/λ), inverse derivation of the mean and standard deviation was considered sufficient to test for the physical nature of the distribution of false alarm rates. Figs. 5a and b present the performance of the LHS and MC technique in their ability to preserve the satellite retrieval error structure (Table 2) for Storms 1 and 2, respectively as a function of simulation runs. We observe a less degree of variability in modeling the probabilities of rain and no-rain detection for the LHS technique compared to MC (upper two panels, Figs. 5a and b). This is expected as LHS was applied in the constrained sampling of the Bernoulli trials for detecting rain and no rain events. Fig. 5 demonstrates this point further by comparing the coefficient of variation (CV) for each sampling method (see uppermost two panels). For modeling the probabilities of rain and no rain, the LHS technique achieves a very low CV two orders faster than the MC scheme. There is a significant reduction in CV of the LHS technique compared to the MC scheme at simulation runs ranging from orders 1 to 3 (i.e., 10–1000). For other satellite error parameters, we observe that the CV of the LHS technique as a function of sample size is almost equal to that of the MC scheme (Fig. 6, lower three panels). Thus, the LHS technique is able to preserve the satellite rainfall error structure as accurately as the MC.

It is noted that the strong non-linearities in the surface hydrological process can cause the input errors to be amplified or dampened depending on the specific nature of the error structure. For example, Hossain et al. (2004a,b) have reported runoff simulation error to be more sensitive to overestimation of rainfall (bias greater than 1.0) than underestimation (bias less than 1.0). Fig. 7 presents the mean rainfall volume (as simulated by the PM sensor) as a function of SREM simulation runs for both MC and LHS techniques. The purpose was to identify whether LHS had any significant dependence (bias) on storm morphological properties and to understand the propagation of this bias in runoff simulation error. While no clear bias was observed for Storm 1, the LHS underestimated the rainfall volume (compared to the MC) for Storm 2 by 2.3%. This is attributable to the higher percentage of rain coverage (% Rainy, Table 1) in Storm 2 (87.6%) which meant that a significantly higher number of Bernoulli trials were conducted for Storm 2 for modeling rain detection P_1, than Storm 1. In the constrained and systematic sampling framework of the Latin Hypercube scheme, this consequently resulted in an equally higher number of non-overlapping intervals where the r_n uniform deviate value sampled was higher than P_1 (i.e., unsuccessful Bernoulli trials to detect rain). This 2.3% underestimation in rainfall volume by LHS needs to be assessed in runoff simulation error before any clear conclusion can be drawn of its significance (which is discussed next).

5. Simulation framework

We now assess the performance of the LHS scheme (versus MC) in terms of runoff simulation error by passing each of the LHS (and MC)-based SREM realizations through TOPMODEL using the same set of 30 independent seeds. The hydrologic uncertainty was assessed in terms of three runoff error statistics: mean relative errors in peak runoff (PR), time to peak (TP), and runoff volume (RV). We define relative error (ε_X) as

$$\varepsilon_X = \frac{X_{sim} - X_{ref}}{X_{ref}}, \qquad (6)$$

where X is defined as one of the simulated runoff parameters (RV, PR, TP). The subscript '*ref*' refers to the runoff parameter derived from *reference runoff*. The gauge rainfall-based simulated hydrograph was considered *reference runoff*, which offers acceptable fit to the observed runoff data of both storm cases (see Fig. 2). For a given simulation size, N, the mean of ε_X is calculated as the arithmetic average over the N SREM realizations, each passed through TOPMODEL. Two approaches are followed to evaluate the LHS scheme's performance (in comparison with the MC technique) for various increasing levels of simulation sample sizes, in a way similar to that described in Section 4.2.

In the first approach we employ a statistic named '% *Change in Error*' to evaluate LHS and MC adequacy for different simulation sample sizes using as reference the MC simulation results derived on the basis of a large sample size. From Figs. 5 and 6 we note that a sample size of 20,000 runs could adequately represent the rainfall error structure, so it is chosen as the reference sample size for runoff

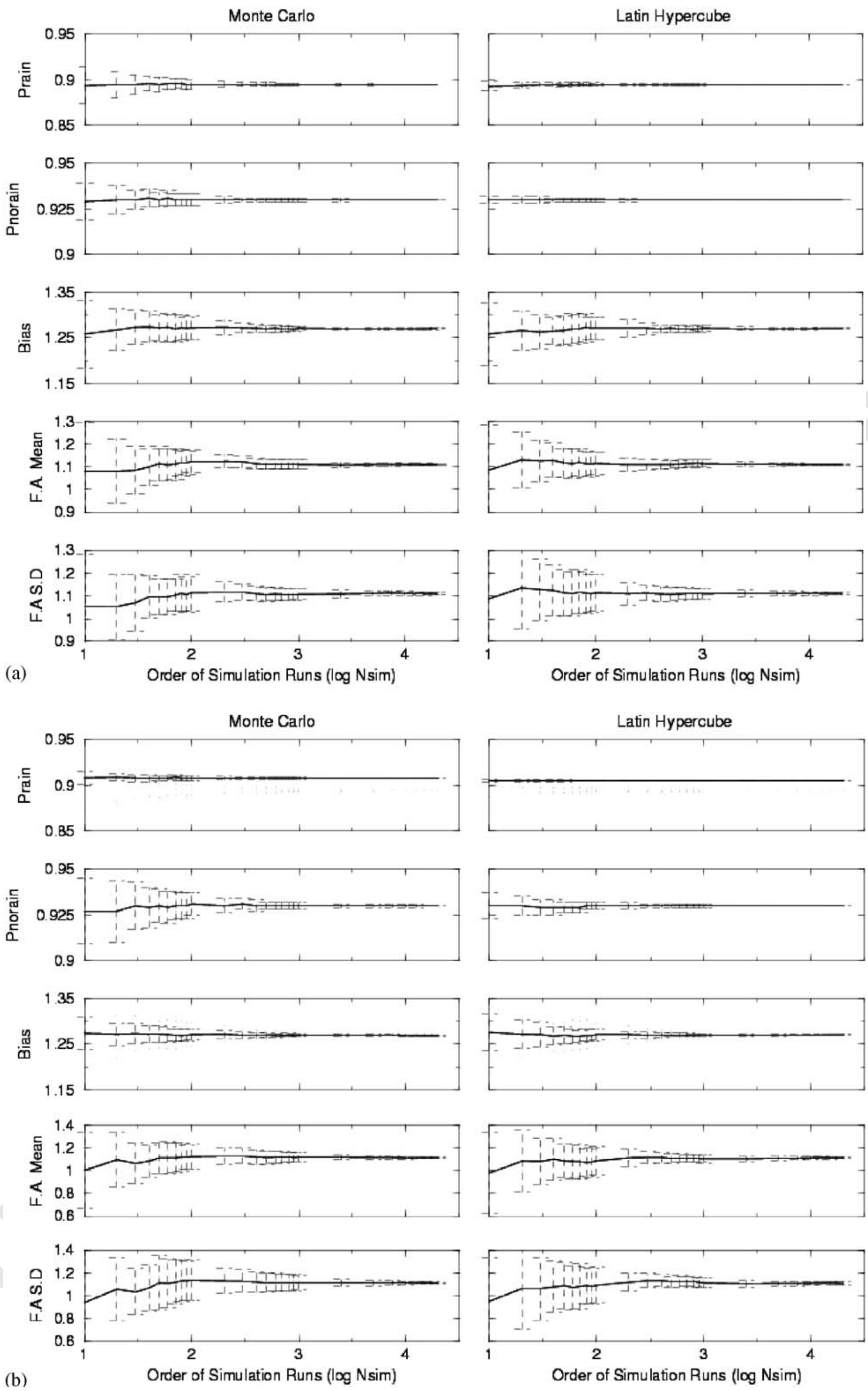

Fig. 5. (a) Comparison of MC (left panels) and LHS (right panels) techniques in preserving properties of rainfall retrieval error structure as a function of SREM simulation runs for Storm 1. Error bars represent one standard deviation of variability among 30 sets of independent seeds used. Bias is multiplicative as in Table 2. FA—false alarms; SD—standard deviation; P_{rain} is probability of rain detection averaged over storm duration (similar to P_1 in Table 2); P_{norain} is probability of no rain detection (P_0 in Table 2). (b) Same as Fig. 5a for Storm 2.

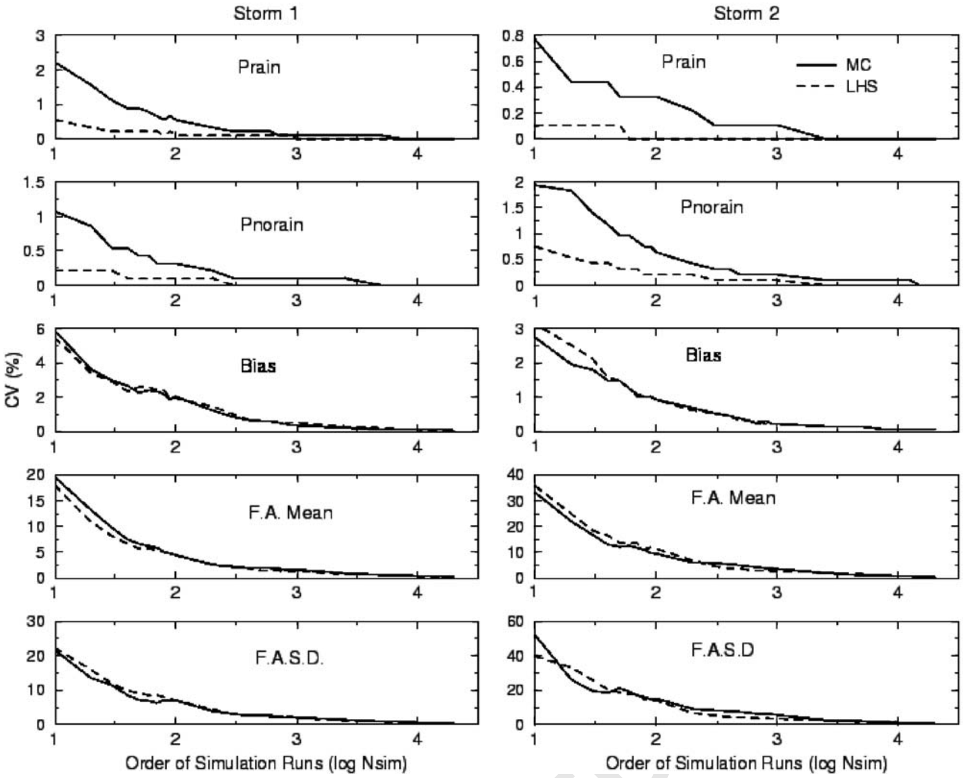

Fig. 6. Comparison of MC (solid) and LHS (dashed) schemes in modeling error structure of satellite rainfall as a function of simulation runs for Storm 1 (left panels) and Storm 2 (right panels). Coefficient of variation (CV) for a given sampling scheme is expressed as error parameter standard deviation of 30 seeds normalized by mean (true) error parameter in Table 2.

simulation error. Then, the error statistic is as follows:

$$\begin{aligned}&\%\text{ Change in error at } N \text{ for LHS}\\ &\quad = \frac{\varepsilon^{N_{MC}}_{X\ MC} - \varepsilon^{N}_{X\ LHS}}{\varepsilon^{N_{MC}}_{X\ MC}},\\ &\%\text{ Change in error at } N \text{ for MC}\\ &\quad = \frac{\varepsilon^{N_{MC}}_{X\ MC} - \varepsilon^{N}_{X\ MC}}{\varepsilon^{N_{MC}}_{X\ MC}}.\end{aligned} \tag{7}$$

Subscripts *MC* and *LHS* refer to the MC and LHS techniques, while superscripts N and N_{MC} refer to the varying sample size and the reference sample size (20,000 runs), respectively.

The second approach is to compare the tails of the distribution of runoff simulations derived from the LHS and MC random experiments for varying sample sizes. The tails are evaluated using the confidence levels ranging from 90% (5% upper, 95% lower) to 60% (20% upper, 80% lower) in 10% increments. Fig. 2 shows an example of the runoff simulation quantiles at the 90% confidence limits based on the reference MC simulation experiment (20,000 runs). It has been previously reported (Stein, 1987, for example) that LHS gives an estimator with lower variance than MC when the number of simulations is larger than the number of input parameters provided that the condition of monotonicity holds. Because the retrieval error in SREM is multiplicative (i.e., error variance is proportional to rainfall rate according to Eq. (1)), higher error in the retrieval (arising from either higher or lower magnitudes of error parameters such as bias, random error variance, false alarm rates and probabilities of successful rain/no-rain detection) would cause elongated tails in runoff uncertainty distribution. Thus, the runoff uncertainty (tails) at strict confidence levels ($>60\%$ error quantiles) form an ideal candidate for assessing the reliability of the LHS technique, because of its expected monotonic relationship with retrieval error parameters. It is noted, however, that the same may not apply unconditionally for the calculated runoff

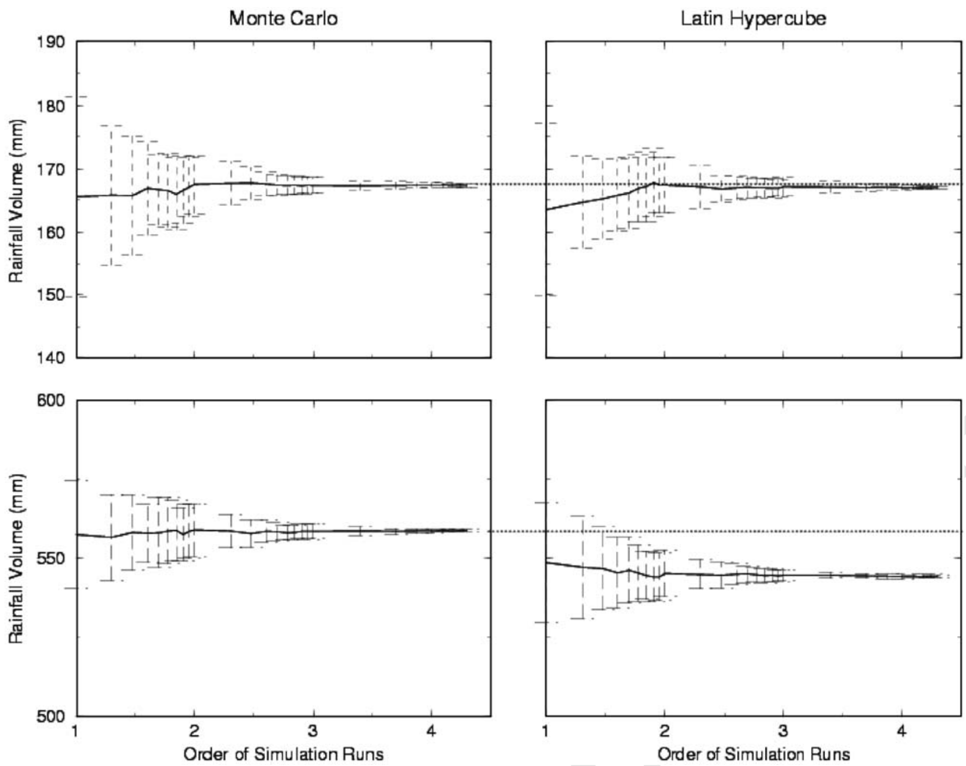

Fig. 7. Comparison of MC (left panel) with LHS (right panel) in simulating sensor retrieved rainfall volume as a function of SREM simulation runs. Storms 1 and 2 are represented in upper and lower panels, respectively. Dashed red line represents retrieved rainfall volume by MC at 20,000 runs (i.e., reference sensor retrieved rainfall volume).

parameters RV, PR and TP, which are evaluated using the first approach. We quantify the runoff uncertainty at a selected confidence level as the distance between the upper and lower quantiles integrated over the whole storm-runoff duration. This is hereafter referred to as Runoff Uncertainty Volume (*RUV*). Thus, wider confidence levels would be associated with higher RUV values, and vice versa.

6. Results and discussion

Figs. 8a and b present performance comparisons between the MC and LHS techniques in terms of runoff simulation uncertainty for PR, RV and TP for Storms 1 and 2, respectively. For Storm 1, the mean of % *Change in error* statistic for LHS appears similar to that for MC. However, the variability among seeds (i.e., the evaluated standard deviation) is clearly smaller for LHS. The variability across seeds is an indication of how reliable the uncertainty estimation for a given technique is, since in a real world application only one set of seeds will be typically employed. However, due to the non-linearities in the runoff transformation process, this variability does not converge for PR and TP. For RV, we observe a clear convergence at 20,000 simulations (Figs. 8a and b). This is expected since RV is a hydrologic parameter integrated over time and therefore random retrieval errors that propagate in runoff transformation can balance each other. For Storm 2, we observe a similar pattern, but with less apparent differences between LHS and MC. Also, there is a distinct negative bias (underestimation) observed in error simulation for PR and TP for the LHS scheme (Fig. 8b, uppermost and lowermost right panels). We attribute this to the negative bias in simulating the retrieved rainfall volume by the LHS for Storm 2 (Fig. 7, lower right panel). The 2.3% negative bias in retrieval of rainfall volume of the LHS technique has propagated to a slightly higher (3.0%) negative bias in Peak Runoff with respect to the MC scheme. It appears that this bias in rainfall volume has a strong effect on PR and TP runoff simulation error, but a negligible effect on RV. The overall picture emer-

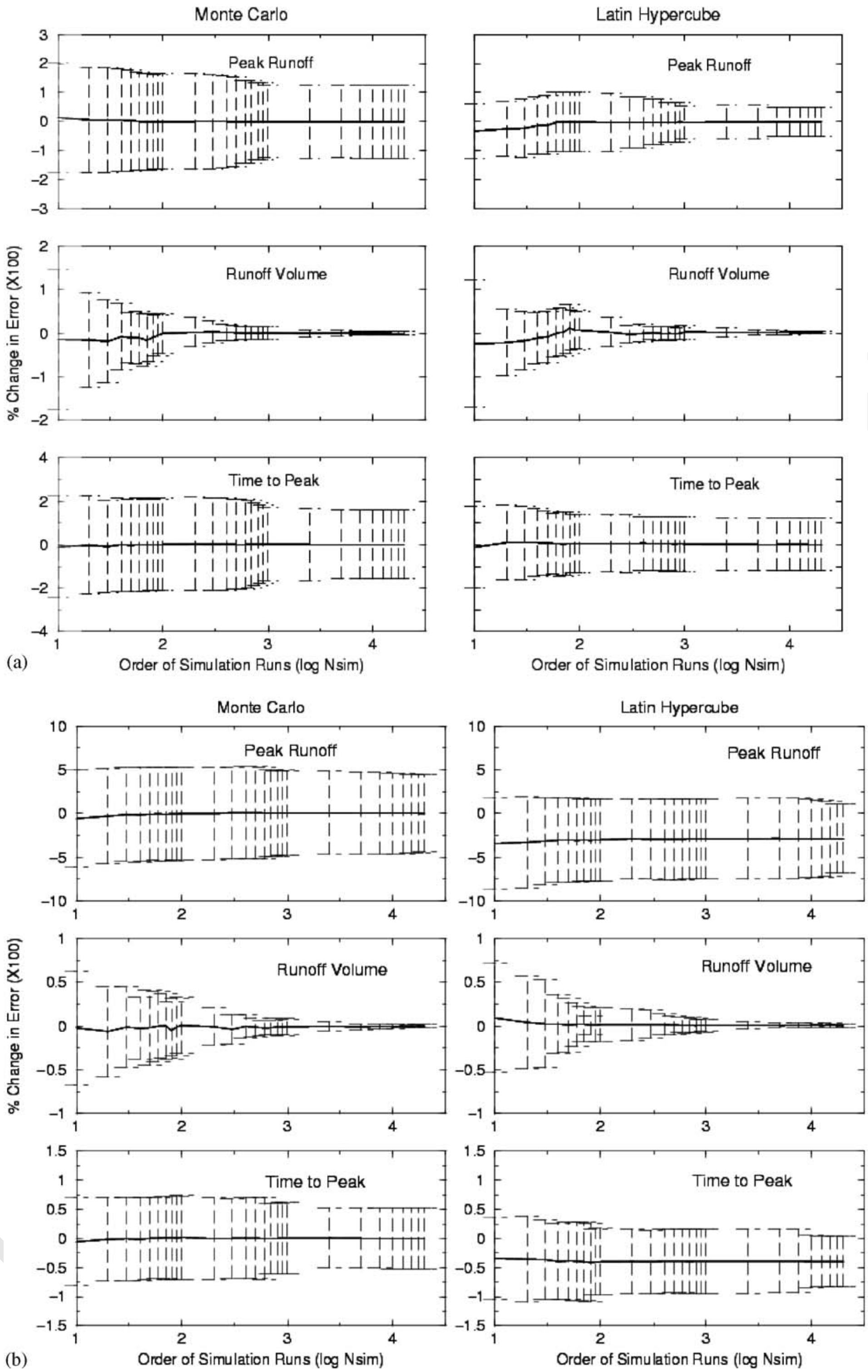

Fig. 8. (a) Comparison of MC (left panels) with LHS (right panels) in simulation uncertainty of Peak Runoff (uppermost panels), Runoff Volume (middle panels) and Time to Peak (lowermost panels) as a function of simulation runs for Storm 1. % Change in Error represents relative change in error with respect to reference runoff error (MC at 20,000, Eq. (7)). Solid line and error bars represent mean and one standard deviation for 30 realization/seeds. (b) Same as Fig. 8a, for Storm 2.

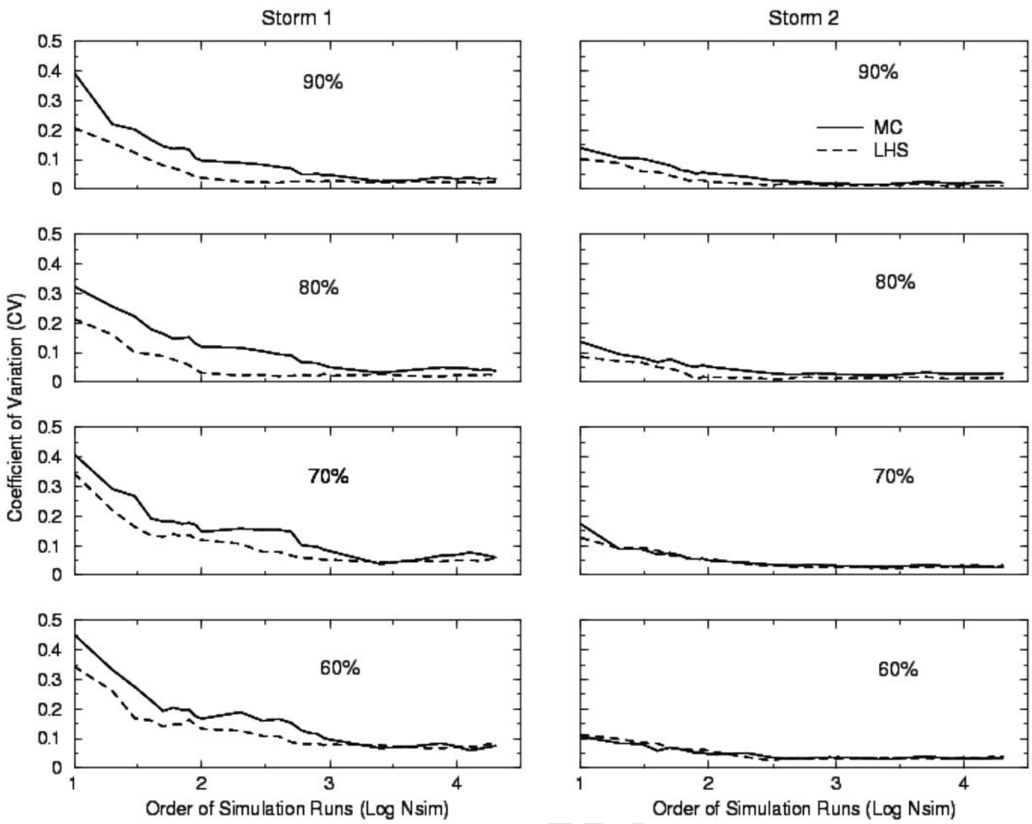

Fig. 9. Comparison of coefficient of variation (CV) of simulation runoff error bars (tails, quantified as runoff uncertainty volume, RUV) at given confidence limits for MC (solid) and LHS (dashed) as a function of simulation runs for Storm 1 (left panels) and Storm 2 (right panels). CV is computed with 30 sets of independent seeds.

ging from these two figures (8a and b) is that LHS offers moderate computational benefit in comparison with the MC technique for assessing errors in bulk PR, RV and TP runoff parameters. Furthermore, the LHS performance, as applied in the context of accelerating Bernoulli trials for rain and no rain discrimination, seems to be sensitive to the storm morphological properties. Storms with widespread rainfall patterns appear to be inducing an underestimation in rainfall volume by LHS and this bias persists in its propagation as bias in PR and TP runoff parameters.

In Fig. 9, we present the CV of the RUV at specified confidence levels ranging from 90% (5% upper and 95% lower) to 60% (20% upper and 80% lower). The mean values of the RUV derived from LHS were found to match accurately those from MC sampling for any given simulation size. The mean value of RUV was found to show insignificant sensitivity to simulation runs exceeding sample size of 100. Hence these results are not reported herein. We further observed that, for Storm 1, LHS yields significantly lower variability in RUV values for 80% and higher confidence levels. In fact, LHS attains the level of low CV (<0.05) at almost 2 orders of magnitude of simulation runs less than what MC requires (Fig. 9 first two upper left panels). This indicates that LHS can estimate the runoff simulation uncertainty bounds for high ($>80\%$) confidence levels with the same degree of reliability as MC, but for almost two orders of magnitude of fewer runs. For lower confidence levels we observe no clear computational benefit of LHS. This is probably because as confidence levels in runoff simulation become narrower, the condition of monotonicity is gradually violated. In fact, one can argue that as confidence levels narrow significantly, the ensemble of simulated runoff realizations converges towards an ideal single hydrograph realization, whereby the conditions of monotonicity cannot be expected to hold. For Storm 2, we still observe LHS to be an effective estimator of RUV, but the computational advantage over MC is not exemplified as for Storm 1. Again, this demonstrates an inherent sensitivity of the LHS scheme to the type of storm, with storms with widespread and heavy rainfall patterns making the LHS technique to not be as accurate as MC sampling.

It is appropriate to highlight, at this stage, a few words of caution on the use of LHS. While the LHS technique generally never performs worse than the MC technique in terms of computational efficiency, there are circumstances where the opposite may be observed (see Figs. 6 and 9). This is particularly so when the number of simulations is less than 100. Also, such situations seem to be more pronounced for Storm 2 than Storm 1 (mild storm event). This raises an interesting open-ended question as to whether it is due to the potential ''limitation'' of the LHS technique when the number of simulations is less or it is due to the type of ''data'' (storm) or it is due to combination of both (?). We stress that seeking an answer to this question in future research endeavours is important because it may have much wider implication as to the usefulness of the LHS technique in error propagation studies on satellite-based hydrology.

7. Conclusion

This study presented an assessment of Latin Hypercube Sampling for uncertainty estimation of satellite rainfall observations in flood prediction for two storm cases of contrasting morphological properties. A Satellite Rainfall Error Model (SREM), calibrated for Passive Microwave sensors, was linked with a hydrologic model in a Monte Carlo framework to study and understand the error propagation of retrieval error in runoff simulation. The concept of LHS was applied in the constrained sampling of Bernoulli trials to achieve higher computational efficiency in modeling sensor's rain and no rain detection probabilities. It was observed that LHS offered no additional computational benefit over MC in assessing runoff simulation error. Furthermore, LHS appeared sensitive to storm morphology, namely its accuracy was undermined for storms with widespread rainfall patterns. However, the LHS was able to predict the 80% and higher confidence limits in runoff simulation with the same degree of reliability as MC with almost two orders of magnitude fewer simulations.

Results from this LHS assessment study have implications for wide scale assessment of satellite rainfall retrievals for flood prediction and other land surface processes. While LHS, as applied in the present context of discrimination of rain and no rain events, offers no computational advantage for assessing simulation errors in peak runoff, time to peak and runoff volume, it can serve as a very useful tool for assessing the bounds of the runoff simulation distribution at large confidence limits. This knowledge therefore allows the efficient use of a LHS modified sampling scheme on a satellite rainfall error model (such as SREM) involving: (1) slow running models (such as distributed hydrologic models and land surface models); (2) larger regions; and (3) longer study periods; provided the study is confined to analysis of bounds of the simulated runoff distribution.

Findings from this study also indicate that LHS could be potentially useful for a small-scale prototype end-to-end probabilistic flood prediction system that was implemented by the National Weather Service, known as the Ensemble Stream flow Prediction (ESP) system (Day, 1985; Schaake et al., 2001). The ESP system amounts to generating a number of ensembles of traces of future precipitation on the basis of an uncertainty framework and running them numerically through a hydrologic model. From the resulting multiple hydrographs, several probabilistic statements are then drawn about the future river stage (Schaake and Larson, 1998). However, Krzysztofowicz (1999) argued that the ESP typically under-represents uncertainty estimation in river stage (runoff) as it does not incorporate hydrologic prediction uncertainty as a random process. Thus, a better application of ESP would require an enhanced ensemble simulation (MC) framework that combines both precipitation uncertainty and hydrologic prediction uncertainty. As this combined input-prediction uncertainty would require computationally prohibitive runs for the MC simulation, LHS can play an important role, at least in principle, in reducing the number of ensemble runs of precipitation inputs by a few orders of magnitude. This consequently frees up computational power to incorporate a sufficient ensemble of hydrologic prediction scenarios to be combined with the precipitation input realizations. Work is currently in progress to assess how useful LHS can be in the runoff simulation uncertainty assessment that accounts for both input uncertainty and hydrologic prediction uncertainty.

References

Bacchi, B., Ranzi, R., Borga, M., 1996. Statistical characterization of spatial patterns of rainfall cells in extratropical cyclones. Journal of Geophysical Research 101 (D21), 26277–26286.

Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water Resources Research 23 (8), 1393–1442.

Beven, K.J., Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrological Sciences Journal 24 (1), 43–69.

Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., Freer, J., 1995. TOPMODEL. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO, pp. 627–668.

Bidwell, S., Turk, J., Flaming, M., Mendelsohn, C., Everett, D., Adams, J., Smith, E.A., 2002. Calibration plans for the global precipitation measurement. In: Proceedings Joint Second International Microwave Radiometer Calibration Workshop and CEOS Working Group on Calibration and Validation, Barcelona, Spain, October 9–11, 2002.

Borga, M., Anagnostou, E.N., Frank, E., 2000. On the use of real-time radar rainfall estimates for flood prediction in mountainous basins. Journal of Geophysical Research 105 (D2), 2269–2280.

Day, G.N., 1985. Extended stream flow forecasting using NWSRFS. Journal of Water Resources and Planning Management 111 (2), 157–170.

Dinku, T., Anagnostou, E.N., Borga, M., 2002. Improving radar-based estimation of rainfall over complex terrain. Journal of Applied Meteorology 41, 1163–1178.

Duan, Q., Sorooshian, S., Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28, 1015–1031.

Flaming, M., 2002. Requirements of the Global Precipitation Mission. Proceedings of IGARSS, 2002 (International Geoscience and Remote Sensing Symposium), Toronto, Canada, June 24–28.

Grecu, M., Krajewski, W.F., 2000. Simulation study of the effects of model uncertainty in variational assimilation of radar data on rainfall forecasting. Journal of Hydrology 239 (1–4), 85–96.

Hossain, F., Anagnostou, E.N., 2004. Assessment of current passive microwave and infra-red based satellite rainfall remote sensing for flood prediction. Journal of Geophysical Research 109 (D7) April D07102.

Hossain, F., Anagnostou, E.N. 2005a. A two-dimensional satellite rainfall error model. IEEE Transactions in Geosciences and Remote Sensing, in press.

Hossain, F., Anagnostou, E.N., 2005b. Using a multi-dimensional satellite rainfall error model to characterize uncertainty in soil moisture fields simulated by an offline land surface model. Geophysical Research Letters 32.

Hossain, F., Anagnostou, E.N., Borga, M., Dinku, T., 2004a. Hydrologic model sensitivity to parameter and radar rainfall estimation uncertainty. Hydrological Processes 18 (17), 3277–3299.

Hossain, F., Anagnostou, E.N., Dinku, T., 2004b. Sensitivity analyses of passive microwave and IR retrieval and sampling on flood prediction uncertainty for a medium sized watershed. IEEE Transactions on Geosciences and Remote Sensing 42 (1), 130–139.

Iman, R.L., Conover, W.J., 1980. Small sample sensitivity analysis techniques for computer models, with application to risk assessment. Communication in Statistics, Part A/ Theory and Methods 17, 1749–1842.

Iman, R.L., Shortencarier, M.J., 1984. A FORTRAN 77 Program and user's guide for the generation of Latin Hypercube and random samples for use with computer models. NUREG/CR-3624, SAND83-2365. Report, Sandia National Laboratories, Albuquerque, NM.

Iman, R.L., Helton, J.C., Campbell, J.E., 1981. An approach to sensitivity analysis of computer models: Part I—Introduction, input, variable selection and preliminary variable assessment. Journal of Quality Technology 13 (3), 174–183.

Isukapalli, S.S., Georgopoulos, P.G., 1999. Computational methods for the efficient sensitivity and uncertainty analysis of models for environmental and biological systems. Technical Report CCL/EDMAS-03, Rutgers, State University of New Jersey.

Kremer, J.N., 1983. Ecological implications of parameter uncertainty in stochastic simulation. Ecological Modelling 18, 187–207.

Krzyzstofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35 (9), 2739–2750.

Krzyzstofowicz, R., 2001. The case for probabilistic forecasting in hydrology. Journal of Hydrology 249, 2–9.

Loh, W.-L., 1996. On Latin Hypercube sampling. Annals of Statistics 24 (5), 2058–2080.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Melching, C.S., 1995. Reliability estimation. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO, pp. 69–118.

Nijssen, B., Lettenmaier, D.P., 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: anticipating the global precipitation measurement satellites. Journal of Geophysical Research 109 (D2) DOI: D02103, 0.1029/2003JD003497.

Quinn, P.F., Beven, K.J., Chevallier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes 5, 59–79.

Quinn, P.F., Beven, K.J., Lamb, R., 1995. The ln(a/tan b) index: how to calculate it and how to use it in the TOPMODEL framework. Hydrological Processes 9, 161–182.

Schaake, J.C., Larson, L., 1998. Ensemble streamflow prediction (ESP): progress and research needs. In: Preprints of Special Symposium on Hydrology, J19–J24. American Meteorological Society, Boston, MA.

Schaake, J.C., Welles, E., Graziano, T., 2001. Comment on 'Bayesian theory of probabilistic forecasting via deterministic model by Roman Krzyzstofowicz'. Water Resources Research 37 (2), 439.

Seo, D.-J., Perica, S., Welles, E., Schaake, J.C., 2000. Simulation of precipitation fields from probabilistic quantitative precipitation forecast. Journal of Hydrology 239 (1–4), 203–229.

Siddall, J.N., 1983. Probabilistic analysis. In: Probabilistic Engineering Design. Marcel Dekkar, New York, NY, pp. 145–236.

Smith, E.A., 2001. Satellites, orbits and coverages. In: Proceedings of IGARSS 2001 (International Geoscience and Remote Sensing Symposium), Sydney, Australia, July 9–13.

Stein, M., 1987. Large sample properties of simulations using Latin Hypercube sampling. Technometrics 29 (2), 143–151.

Steiner, M., Bell, T.L., Zhang, Y., Wood, E.F., 2003. Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar data set. Journal of Climate 16 (22), 3758–3777.

Winchell, M., Gupta, H.V., Sorooshian, S., 1998. On the simulation of infiltration- and saturation excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation. Water Resources Research 34 (10), 2655–2670.

Yuter, S., Kim, M.-J., Wood, R., Bidwell, S., 2003. Error and uncertainty in precipitation measurements. GPM Monitor, February (URL: http://gpm.gsfc.nasa.gov/Newsletter/february03/calibration.htm; last accessed September 9, 2005).