Efficient Uncertainty Assessment for Satellite Rainfall Observations with Application to Flood Prediction

Faisal Hossain[1]*, Emmanouil N. Anagnostou[2]*, and Amvrossios C. Bagtzoglou[3]*

Abstract

With the advent of the Global Precipitation Measurement mission (GPM) in 2009, satellite rainfall measurements are expected to become globally available at space-time scales relevant for flood prediction of medium sized ungauged 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 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 or prohibitive. In this study, Latin Hypercube Sampling (LHS) was implemented on a satellite rainfall error model to explore the degree of computational efficiency that could be achieved with a complex hydrologic model. For assessment of errors in time to peak, peak runoff, and runoff volume it was observed that LHS offered no significant benefit over the MC sampling method. However, for deriving the runoff simulation uncertainty at confidence limits greater than 90%, LHS was found highly efficient. With only a small fraction (in the order of 5%) of MC runs, LHS was able to produce the same 90% confidence limits in runoff simulation.

Introduction

The GPM mission planned jointly by US, Japanese and European space agencies envisions providing global rainfall products from a constellation of Passive Microwave (PM) satellite sensors at time scales ranging from 3 to 6 hours and spatial resolutions of 100 km^2 (Yuter *et al.*, 2003). Nevertheless, satellite rainfall retrieval is

---

* **All at:** Department of Civil and Environmental Engineering, 261 Glenbrook Road,
Unit 2037, University of Connecticut, Storrs, CT 06269-2037, Fax: (860) 486-2298

[1] Graduate Research Assistant, faisal@engr.uconn.edu

[2] Associate Professor, Tel: (860) 486-6806, manos@engr.uconn.edu

[3] Associate Professor, Tel: (860) 486-4017, acb@engr.uconn.edu

subject to errors caused by various factors ranging from infrequent sampling to the 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.

Conventional uncertainty assessment of such space-based rainfall observations for flood prediction would therefore require the derivation of the probability distribution of runoff from a given probabilistically formulated satellite rainfall error model through the rainfall-runoff transformation mechanism using Monte Carlo (MC) simulations. For slow running hydrologic models (such as fully distributed physically based models), such MC assessment can be computationally prohibitive. This makes the hydrologic assessment of satellite rainfall data limited to mainly fast running conceptually lumped models. A broader uncertainty assessment of satellite rainfall observations across increasing levels of hydrologic model complexity therefore warrants the investigation of computationally more efficient sampling schemes. Latin Hypercube Sampling (LHS) is one such technique that offers promise in reducing the computational burden of MC uncertainty assessment. This study therefore aims to investigate 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 with MC sampling?*

The study is organized in the following manner. We first describe the watershed, data, and hydrologic model. This is followed by a brief description of the satellite error model and the LHS scheme. Then the simulation framework is presented and finally the last two sections discuss the results and conclusions of this study.

Watershed, Data and Hydrologic Model

The Posina watershed, located in northern Italy, is chosen for this study. It is close to the city of Venice (Figure 1, right panel). Posina has an area of 116 km^2 and altitude ranging from 2,230 to 390 meters at the outlet (Figure 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 rainfall runoff model TOPMODEL (Beven and Kirkby, 1979) was chosen to simulate the rainfall-runoff processes of floods in the Posina watershed. It 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. A storm event that took place in August 1987 was chosen for the flood simulation study. The storm lasted about 72 hours and registered a peak discharge of 54.4 m^3/s (Figure 2).

Satellite Rainfall Error Model

Conceptually, the satellite rainfall error model (SREM) works by accounting for the following possible outcomes for a sensor overpassing a watershed during a storm event (see Appendix 1):

1) It retrieves non-zero rainfall when it actually rains (successful rain detection modeled as P_1- probability of rain detection).
2) It retrieves zero rainfall when it actually rains (false no-rain detection).
3) It retrieves zero rainfall when the reference rainfall is zero (successful no-rain detection modeled as P_0 - probability of no rain detection).
4) It retrieves non-zero rainfall when the reference rainfall is zero (false rain detection modeled as D_{false}- the error distribution of false alarms).

The statistical features associated with each of the above outcomes were derived from the literature. Surface gauge rainfall was used as "reference rainfall" (R_{REF}). The successful detection of rain and no-rain events was modeled through Bernoulli trials (P_1 and P_0 respectively, Appendix 1). For more details about the model, the reader is referred to the works of Hossain *et al.*(2003), and Hossain and Anagnostou (2003).

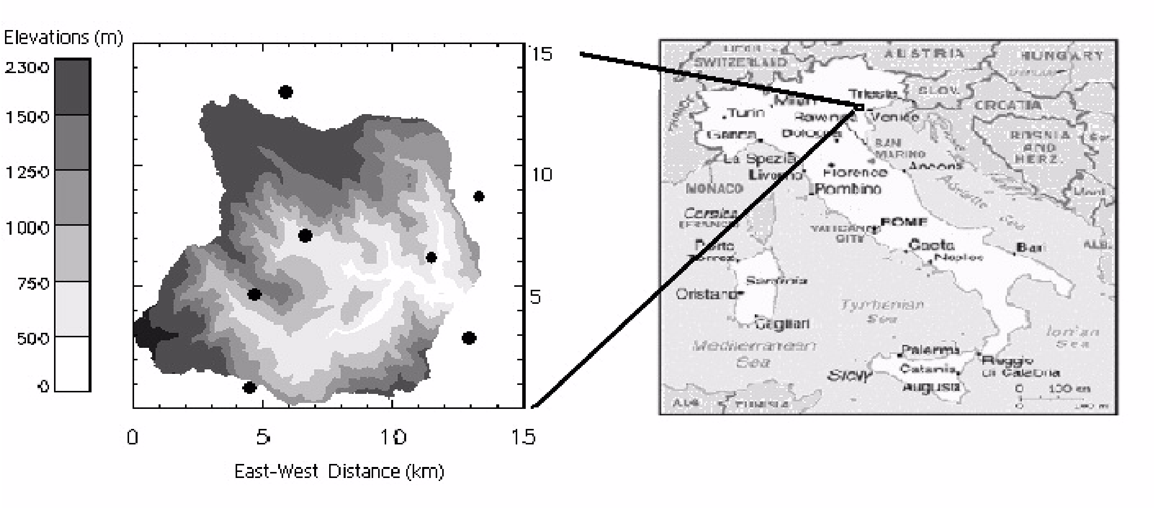

Figure 1. Geographic location of the Posina Watershed (right panel) and watershed elevation map (left panel) overlaid by the rain gauge network locations (solid circles).

The LHS Technique

The LHS technique is a constrained sampling technique whereby the input parameter range is divided into non-overlapping intervals of equal probability. This way, we try to explore the parameter space as completely 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 (say) intervals and perform sampling as follows,

$$P_m = U(0,1) \times [(A-B)/N] + (m-1) \times [(A-B)/N] \qquad m = 1, 2, 3, \ldots\ldots, N \qquad (1)$$

Here P_m is the cumulative probability value used with the inverse distribution to produce the specific parameter value to be used with LHS. For further details about the LHS technique the reader is referred to McKay *et al.* (1979), and Iman and Shortencarier (1985). We have employed the above concept of LHS in SREM for investigating efficient sampling of Bernoulli trials in order to detect rain and no-rain events.

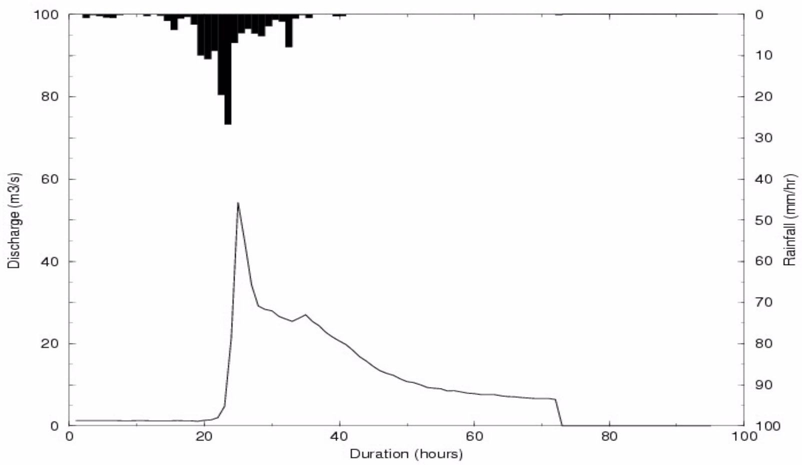

Figure 2. August 1987 storm event hydrograph (lower axis) and hyetograph (upper axis) in Posina, Italy.

Simulation Framework

Using three different seeds for random number generators, we first found the equilibrium MC sample size (N_{MC}) beyond which error statistics in runoff stabilized. N_{MC} is therefore the number of SREM model runs required to be passed through TOPMODEL to be deemed adequate for assessment of the satellite's retrieval uncertainty in runoff simulation. N_{MC} was found to be approximately equal to 20,000 (Figure 3, left panel). We further assumed that this N_{MC} simulation yields the most accurate representation of the hydrologic uncertainty of satellite rainfall measurements. The hydrologic uncertainty was assessed in terms of three runoff error statistics, namely mean relative errors in peak runoff (PR), time to peak (TP), and runoff volume (RV). We define the relative error (ε) in each of the three runoff parameters as:

$$\varepsilon_X = \frac{X_{sim} - X_{ref}}{X_{ref}} \tag{2}$$

X being defined as one of the runoff parameters (i.e., RV, PR, TP). The gauge rainfall based simulated hydrograph was considered as “reference runoff”.

To evaluate the performance of the LHS scheme, various increasing levels of LHS simulation runs were considered, ranging from 10, 50, 100 up to 20,000. The LHS simulation runs involved the use of the LHS concept modified into the SREM with the TOPMODEL. By comparing the error statistics in runoff simulation between the LHS and MC schemes, a comparative analysis was then conducted.

Results and Discussion

Figure 3 compares the performance of the LHS scheme with the MC sampling scheme. Here the “% Change in Error” at a given simulation size N is defined as,

$$\% \text{ Change in Error} = (\text{Error}_{N_{MC}} - \text{Error}_{N})/\text{Error } N_{MC} \qquad (3)$$

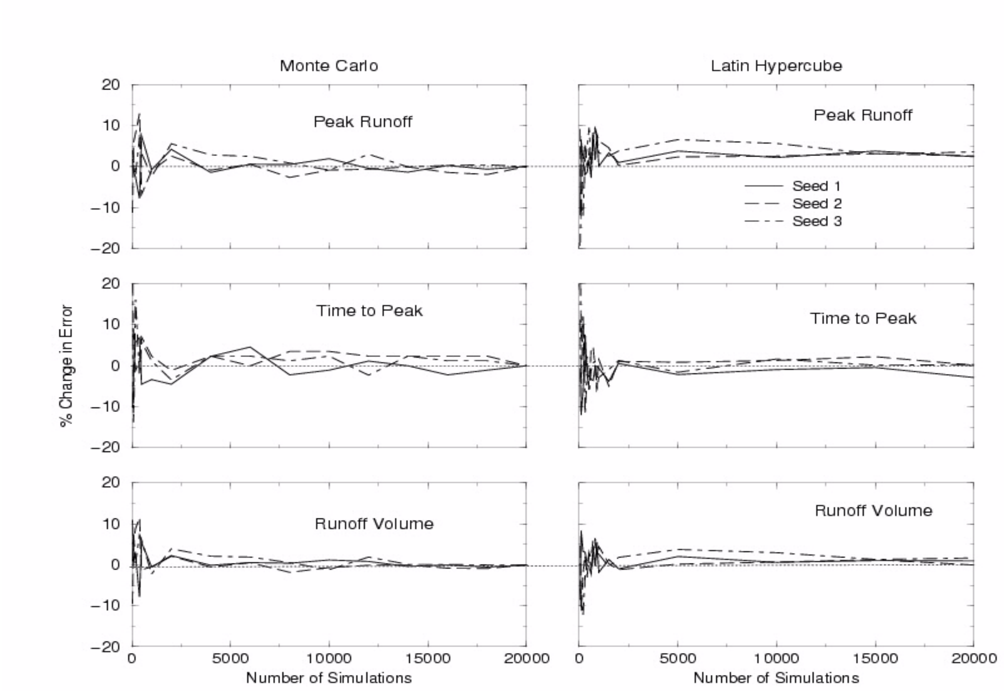

Figure 3. Comparison of LHS scheme (right panel) with MC scheme (left panel) for three hydrologic error parameters (peak runoff–uppermost panel; time to peak- middle panel; and runoff volume–lowermost panel). The dotted horizontal red line represents the line of zero % change error with respect to MC error at N_{MC} =20,000.

Several distinct features seem to emerge from Figure 3. For assessing uncertainty in satellite-derived prediction of time to peak, peak runoff and runoff volume, LHS offers no clear computational efficiency over MC sampling. Also, LHS seems to give a bias in predicting the mean error in peak runoff. We observe noisy behavior by the LHS scheme up to the 1000 run range (5% of MC runs) for predicting mean errors in runoff (Figure 3, right panels). Beyond 1000 runs (>5% of MC runs), LHS appears to yield much smoother uncertainty output with less variability than the MC sampling scheme.

However, upon comparison with the 90% confidence limits derived from MC sampling and LHS schemes, we observed a drastically different scenario (Figure 4). For assessing the range of uncertainty (error bars/tails) in runoff, LHS was found indeed very useful. The confidence limit range of 90% seems to be predicted by LHS with minimum error with about 40 to 20 times less runs when compared to the MC sampling (Figure 4, lower three panels on the right).

Conclusions

Results from this study have provided brief insights on the assessment of space-based rainfall estimates for hydrological forecasting. LHS was found to be an efficient sampling method for an approximate uncertainty analysis by satellite rainfall retrievals with wide confidence limits in the runoff simulation. However, for assessment of simulation uncertainty of runoff volume, time to peak and peak runoff, LHS offered no clear computational benefit over the MC sampling method. This study has provided an objective framework to improve the problem definition for scientific questions such as, *what are the trade-offs between the reduction achieved in computational burden by LHS and the decrease in the accuracy of the uncertainty assessment of satellite rainfall data for flood prediction?* However, more work is required to answer the above question coherently. Such work is currently in progress and we hope to be able to report it soon.

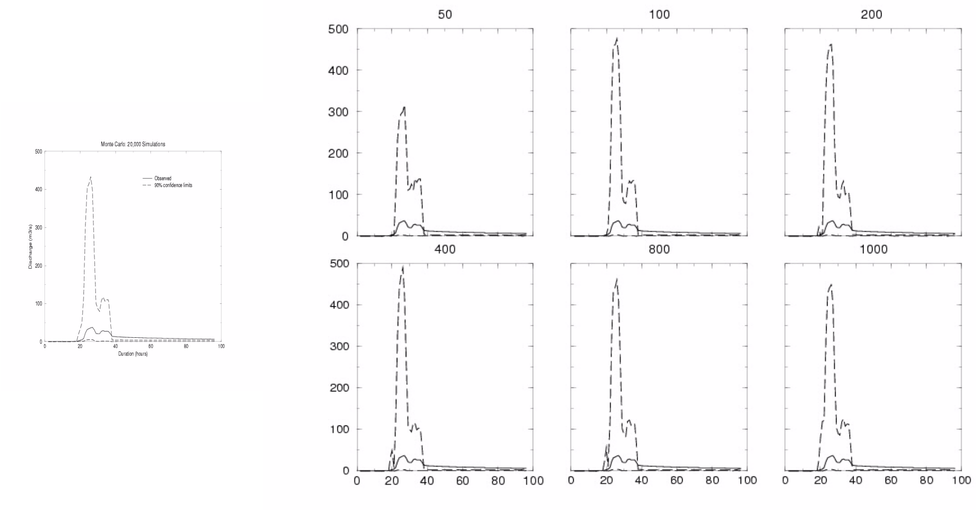

Figure 4. Comparison of the 90% confidence limits generated by MC sampling (leftmost panel) with N_{MC} (20,000 runs) and LHS (right panels, 3 upper and 3 lower). Only the LHS confidence limits obtained with 50 runs was found to be clearly different from MC sampling. All others (>100 runs, >0.5% of MC runs) appeared to yield a very accurate 90% confidence limit similar to that of MC at 20,000 runs.

References

Beven K.J., and M.J. Kirkby (1979). “A physically-based variable contributing area model of basin hydrology.” *Hydrol. Sciences Journal*, *24*(1), 43-69.

Hossain, F. E.N. Anagnostou, and T. Dinku (2003). “Sensitivity analyses of passive microwave and IR retrieval and sampling on flood prediction uncertainty for a medium sized watershed.” *IEEE Trans. Geosci. Remote Sens.* (in press).

Hossain, F. and E.N. Anagnostou (2003). “Assessment of current Passive Microwave and Infra-red based satellite rainfall remote sensing for flood prediction.” *J. Geophys. Res.* (accepted).

McKay, M.D., R. J. Beckman and W.J. Conover (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.

Iman, R.L. and M. J. Shortencarier (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 (Prepared by Sandia National Laboratories).

Yuter, S., M-J. Kim, R. Wood, and S. Bidwell (2003). “Error and Uncertainty in Precipitation measurements.” *GPM Monitor*, February.

APPENDIX 1: Flow Chart for Satellite Rainfall Error Model (SREM)

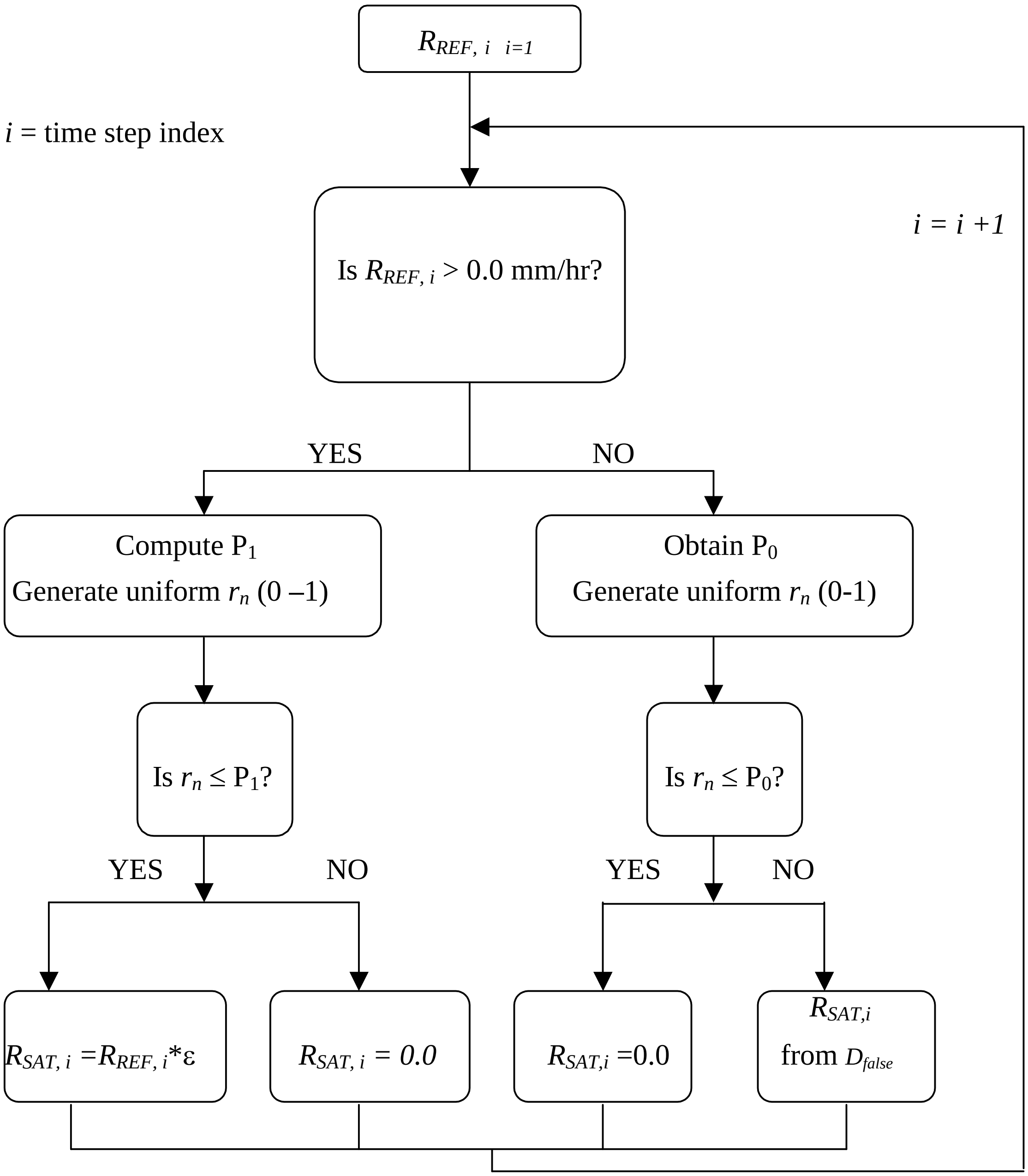

Figure A1. Satellite Rainfall Error Model (SREM) algorithmic structure (after Hossain and Anagnostou, 2003). The r_n is a randomly generated number from a Uniform [0-1] probability distribution. R_{SAT} is the satellite-retrieved rainfall with a multiplicative error ε. P_1 is computed as a function R_{REF} while P_0 is obtained from the literature.