Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales

Faisal Hossain
Department of Civil and Environmental Engineering
Tennessee Technological University
Cookeville, TN 38505-0001
USA

And

George J. Huffman
Science Systems and Applications, Inc.
NASA Goddard Space Flight Center
Laboratory for Atmospheres
Greenbelt, MD 20771
USA

Submitted to:
Journal of Hydrometeorology

First Submitted:
April 2007
Revised September, 2007

Corresponding Author:
Faisal Hossain
Department of Civil and Environmental Engineering
Tennessee Technological University
1020 Stadium Drive
Cookeville, TN 38505-0001
USA

Tel: (931) 372 3257
Fax: (931) 372 6239
Email: fhossain@tntech.edu

ABSTRACT

This paper addresses the open question - *What is an ideal set of error metrics for satellite rainfall data that can advance the hydrologic application of the anticipated Global Precipitation Measurement (GPM) products overland?* Advancing the development of hydrologically relevant satellite rainfall algorithms over land requires the use of error metrics that are mutually interpretable by hydrologists (users) and algorithm developers (data producers). In this study, our primary aim is to initiate a framework for building such metrics that data producers can use to provide hydrologists with more insightful information on the quality of the satellite estimates. In addition, hydrologists can use the framework to develop a space-time stochastic error model for simulating equi-probable realizations of satellite estimates for quantification of the implication on hydrologic simulation uncertainty. First, we conceptualize the error metrics in three general dimensions: 1) temporal dimension (*how does the error vary in time?*); ii) spatial dimension (*how does the error vary in space?*) and iii) retrieval dimension (*how 'off' is each rainfall estimate from the true value over rainy areas?*). We suggest formulations for error metrics specific to each dimension, in addition to ones that are already widely used by the community. We then investigate the behavior of these metrics as a function of spatial scale ranging from 0.04 degree to 1.0 degree for the PERSIANN geostationary infrared-based algorithm. We observe that moving to finer space-time scales for satellite rainfall estimation requires explicitly probabilistic type of measures that are mathematically amenable for space-time stochastic simulation of satellite rainfall data. The probability of detection of rain as a function of ground validation rainfall, which can be modeled using Bernoulli trials, is found to be most sensitive to scale followed by the correlation length for detection of rain. The coefficient of correlation, frequency bias, false alarm ratio and the equitable threat score are found to be modestly sensitive to scales smaller than 0.24° lat./long. Error metrics that account for an algorithm's ability to capture rainfall intermittency as a function of space appear useful in identifying the optimal spatial scales of application for the hydrologist. We show that our proposed conceptual framework can identify seasonal and regional differences in reliability of four global satellite rainfall products over the United States that is otherwise not apparent with conventional metrics. Hence, we believe that the framework for building such error metrics can lay a foundation for better interaction between the data-producing community and hydrologists in shaping the next generation of satellite rainfall algorithm being developed for the planned Global Precipitation Measurement (GPM) mission.

Keywords: Satellite rainfall, Infrared, Geostationary, Uncertainty, Metrics, Scale, GPM and Hydrologic Modeling.

1.0 INTRODUCTION

Rainfall is a critical input for hydrologic models that predict the make-up of the hydrologic state over land. Because rainfall is intermittent, accurate modeling of the dynamic surface hydrologic state requires accurate rainfall data at the highest possible resolution. However, given the gradual and global decline of in situ networks for hydrologic measurements (Stokstad, 1999; Shikhlomanov et al., 2002), space-borne global observations provide the only means to promote our understanding of terrestrial hydrology over the vast regions that are ungauged (Hossain and Lettenmaier, 2006).

The global importance of satellite-derived rainfall has led to the development of an increasing number of satellite-based rainfall products to meet the needs of various users. Anagnostou (2004) provides a detailed synopsis of the evolution of current satellite estimation techniques over land, while Ebert et al. (2007) summarize several "high-resolution precipitation products" that are currently available via the internet. Recognizing the need for metrics of uncertainty, several recent studies have also been initiated to compare the accuracy of various satellite rainfall products over land. For example, the International Precipitation Working Group (IPWG) assessed six widely available satellite data products using an array of error metrics (Ebert et al., 2007). Hong et al. (2006) have evaluated an infrared satellite estimation technique for hydrologic applications using error conceptualizations initiated by North and Nakamoto (1989) and subsequently formalized by Steiner (2003). Other examples of evaluating satellite rainfall uncertainty include: McCollum et al. (2002) on the assessment of bias; Gebremichael and Krajewski (2005, 2004) on sampling errors; and Ali et al. (2005) on satellite error functions for the Sahel region.

While these and other studies of satellite rainfall uncertainty no doubt, have advanced the application in terrestrial hydrology to some extent, some issues continue to remain open. For example, many studies treat error as a uni-dimensional measure and use power law type relationships or models for estimating this aggregate error as a function of spatial and temporal sampling parameters (Moradkhani et al., 2006; Hong et al., 2006). Such frameworks may be acceptable for estimating the average error over an areal domain. However, they do not represent the space-time covariance structure of the estimation error that can have significant implications in terrestrial hydrology (Hossain and Anagnostou, 2005). Uni-dimensional measures also have limited power in articulating the strengths and weaknesses of satellite algorithms that are relevant to hydrologic modeling. Also, most studies, such as that of the IPWG (Ebert et al., 2007), have typically addressed uncertainty at daily or larger time scales and spatial resolutions greater than 20 km, which are somewhat coarse for resolving the evolution of the dynamic hydrologic state over land (e.g. for floods and soil moisture).

Generally, the satellite data and hydrologic communities tend to characterize the accuracy of rainfall data using metrics such as bias, correlation coefficient and standard deviation of “error”. Additional measures, such as Critical Success Index (CSI), Heidke Skill Score (HSS; Heidke, 1926), Equitable Threat Score (ETS), False Alarm Ratio (FAR; Ebert et al., 2007) have seen use in the meteorological community engaged in forecasting (for example, the National Weather Service or the European Center for Medium Range Weather Forecast). These measures have proved useful in assessing satellite rainfall algorithms at scales pertinent for climate modeling, weather prediction or even large-scale water management studies. However, with the planned Global

Precipitation Measurement (GPM) mission (Smith et al., 2006) and the continued shift toward hydrologically more relevant scales (5~10 km and hourly), there is an urgent need to investigate metrics that can more effectively advance the use of satellite algorithms for hydrology over land, among other uses (Huffman et al., 2004; Lee et al., 2004). Hossain and Lettenmaier (2006) have argued that a shift in paradigm is needed to properly assess estimates of rainfall from satellite sensors for modeling of dynamic hydrologic phenomenon such as flood prediction. The issue is that uncertainties in satellite-estimated rainfall cascade non-linearly through the simulation of the terrestrial hydrologic processes (Nijssen and Lettenmaier, 2004). This non-linear effect is difficult to model because of the prominent discontinuities of the rainfall process in space and time that are observed as scales become smaller.

In this paper, we therefore ask ourselves the open question - *What is an ideal set of error metrics for satellite rainfall data that can advance the hydrologic application of the anticipated Global Precipitation Measurement (GPM) products overland?* The satellite rainfall data producing community have long recognized that information on the reliability of satellite precipitation estimates is valuable to a wide range of users. Yet, the definition of 'optimal' use of satellite data is relative to the nature of application. Ebert et al. (2007; on page 49) provides a lucidly explained perspective on the diverse accuracy requirements. Our pursuit of an answer to the main question leads to a set of additional questions: *What should be the characteristics of error metrics at hydrologically relevant scales? How should they be designed so that they are conveniently interpretable by both data producing and hydrologic communities? How should these metrics be packaged into* *standard satellite data products for best use in hydrologic modeling and decision making?*

Clearly, these questions require error expressions that capture mean behavior accounting for space-time correlations and intermittency in the estimated precipitation fields. Hence, for the hydrologist, error should be defined in terms of the rainfall and tagged to a given space and time scale. We therefore conceptualize that the error metrics should be associated, at a minimum, with three general dimensions: 1) temporal dimension (*how does the error vary in time?*); ii) spatial dimension (*how does the error vary in space?*), and iii) retrieval dimension (*how 'off' is the rainfall estimate from the true value over rainy areas?*).

As with any modeling exercise, there is probably no unique way of representing error completely. But, we need to recognize that studies of uncertainty in hydrologic prediction have usually evolved independently of efforts to characterize uncertainty in remote sensing estimates of rainfall. In this study, our aim is therefore to initiate a common framework for building such error metrics. In particular, we are motivated to build such a framework of multi-dimensional metrics that can mathematically be transformed in to a model for simulating equi-probable realizations of satellite rainfall data for a given satellite rainfall algorithm. While there are several mathematical error models today (such as, Steiner, 1996; Steiner et al., 2003; Hong et al., 2006), there are none, to the best of our knowledge, that remain conceptually flexible enough to allow inclusion of additional, or deletion of redundant, error metrics for a given application. Hence, our formulation of such a framework is timely for the community.

We also emphasize that our framework should not be construed as a dismissal of the commonly used metrics (e.g. correlation, bias, RMSE etc.) that currently have widespread diagnostic use. Rather, our framework aims to minimize the limitations of such metrics for assessment of algorithms at hydrologically relevant scales. In section 2, we introduce a set of error metrics originally suggested by Hossain and Anagnostou (2006). Overviews of the study region and rainfall datasets (reference and satellite) are provided in section 3. In section 4, we present our error assessment across hydrologically relevant spatial scales ranging from 0.04° to 1.0° of lat./long. for a particular satellite-based set of precipitation estimates. The implications on data use are discussed along with and the challenges ahead in developing more robust metrics for operational data products. Finally, in section 5, we summarize the major findings and recommend future work.

2.0 ERROR METRICS FOR SATELLITE RAINFALL

In section 1 we hypothesized that error metrics should quantify, at a minimum, three specific dimensions related to rainfall intermittency. Hereafter, we follow up on that concept using the error modeling approach first outlined by Hossain and Anagnostou (2006). In order to identify the set of error metrics that may be required, we should first recognize that the error structure necessary to capture the rainfall intermittency at hydrologically relevant scales arises from the physical issues associated with satellite rainfall estimation. Satellite-derived estimates are typically instantaneous, area-averaged rainfall (Villarini and Krajewski, 2007). Since rainfall is an intermittent process in the continuum of space and time, each satellite gridbox will be classified by the algorithm as rainy or non-rainy. When compared to the corresponding ground validation (hereafter referred to as 'reference') rainfall data, a satellite estimate may fall into one of four possible outcomes:

1) Satellite successfully detects rain (successful rain detection, or "hit").
2) Satellite fails to detect rain (unsuccessful rain detection, or "miss").
3) Satellite successfully detects the no-rain case (successful no-rain detection).
4) Satellite fails to detect the no-rain case (unsuccessful no-rain detection, or "false alarm").

For the data producing community, there are already accepted metrics in use that can quantify these notions of 'hits', 'misses' and 'false alarms'. Some examples are, Frequency Bias (FB), False Alarm Ratio (FAR) and Equitable Threat Score (ETS). Ebert et al. (2007; page 52) provides an introductory background on the formulation of these metrics that are often tagged with the satellite estimates by the data producers during algorithm comparisons (see also Appendix 2). A limitation, however, associated with these metrics is the difficulty in mathematically modeling the property they represent for simulation of stochastic realizations of satellite rainfall data. For example, the metric FB is supposed indicate the tendency of an algorithm to overestimate or underestimate the aerial extent of rainy areas (>1 for overestimation; < 1 for underestimation). Yet, it is not clear how one would use the FB measure to simulate satellite-like rainy areas with coherent spatial structures. The same applies for FAR, ETS and correlation coefficient. Hence, many of the currently accepted metrics have diagnostic power but lack prognostic qualities for hydrologic error propagation experiments. Hence, for the satellite algorithm developers, these metrics could be considered as a starting point to bridge the currently adopted practice to what follows next in our proposed framework on hydrologically more relevant metrics.

In Figure 1, we outline our logical thought process in a step by step manner to conceptualize the error metrics. The gridboxes that are successfully detected as rainy may exhibit three additional properties. Each of these properties may be considered fully or partially representative of the three general dimensions outlined in section 1 (page 7). At this stage we are not certain how completely these properties represent a given dimension. For example, the temporal variation of error probably arises from a mixture of the true spatial and temporal correlations of the rain system in its Lagrangian (system-following) frame of reference, and the advection speed of that frame of reference. Yet, we address the temporal dimension (‘*how does error vary in time?*’) with a simple representation – assuming that only the mean field bias (systematic error) is correlated in time in an Eulerian (surface-based) frame of reference. Since we wish to build the framework first, we feel it is important that we first begin with simplest representation to understand the usefulness of the associated error metrics.

The successful rain or no-rain detection capability may exhibit a strong covariance structure (i.e., the probability of successful detection of a gridbox as rainy or non-rainy may be a function of the proximity to a successfully detected gridbox). For gridboxes that are detected as non-rainy, the algorithm can be characterized by a marginal probability of no-rain. However, for gridboxes that are detected as rainy, the probability of successful detection may depend on the magnitude of the rainfall rate. The functional dependency of probability of detection of rain may be tagged with reference (ground validation) or the estimated rain rate. For example, the hydrologist users would likely be interested in the probability of rain detection benchmarked with respect to ground data. On the other, the data producers may find it almost impossible to tag the probability of detection of the satellite estimates in a likewise manner for the hydrologist on an operational basis due to lack of global scale ground validation data and hence, choose to use satellite estimates instead.

Collecting all these components, and by following our logical modeling steps outlined in Figure 1, it appears that one possible set of error metrics is: (1) Probability of rain detection (and as a function of rainfall magnitude) - POD_{RAIN}; (2) Probability of no-rain detection - POD_{NORAIN}; (3) First and second order moments of the probability distribution during false alarms; (4) Correlation lengths for the detection of rain-CL_{RAIN}, and (5) no rain–CL_{NORAIN}; (6) Conditional systematic retrieval error or mean field bias (when reference rain > 0); (7) Conditional random retrieval error or error variance; (8) Correlation length for the retrieval error (conditional, when rain >0.0)—CL_{RET}; and finally, (9) Lag-one autocorrelation of the mean field bias. The mathematical formulation of each of these nine error metrics are reasonably straight-forward and are provided in Appendix 1. We also encourage readers to refer to Hossain and Anagnostou (2006) where the more complete description of the nine error metrics and how they are used in a mathematical error model to simulate realizations of satellite rainfall data are provided.

It is not clear to us if these nine metrics can completely describe the error structure of satellite rainfall estimation at hydrologically relevant scales. The needs of particular users and applications will necessarily drive the evolution to the best representation of these error structure parameters.

3.0 DATA AND STUDY REGION

We chose the US National Weather Service (NWS) Stage II rainfall data as the ground validation rainfall dataset for characterizing the nine error metrics. This dataset uses the NWS WSR-88D radar estimates with real-time adjustments based on mean-field radar-rain gauge hourly accumulation comparisons (Fulton et al. 1998; Seo et al. 1999, 2000).

For a representative satellite rainfall algorithm, we used data from a recently improved version of the **P**recipitation **E**stimation from **R**emotely **S**ensed **I**nformation Using **A**rtificial **N**eural **N**etworks (PERSIANN; Sooroshian et al., 2000). In its original version, PERSIANN is a satellite IR-based algorithm with calibration by passive microwave precipitation estimates in a neural net framework that produces global estimates of rainfall at 0.25° × 0.25°, half-hourly resolution (Hong et al., 2005a). In this study, we used an improved version that includes a self-organizing nonlinear output (SONO) neural network for cloud-patch–based rainfall estimation. This revised PERSIANN algorithm estimates 0.04^{o}, half-hourly rainfall and is available over the internet at http://hydis8.eng.uci.edu/GCCS/ (Hong et al., 2005b). This fine sub-microwave –footprint scale is achieved by using the Climate Prediction Center Merged IR Data Set (Janowiak et al., 2001) at full resolution and disaggregating the microwave estimates from the original (PERSIANN-produced) 0.12^{o} grids with guidance from the IR field before use in training the neural network scheme.

To minimize effects due to complex terrain and radar range effects the error computation exercise was performed over the region of Oklahoma bounded by 100^{o}W – 95^{o}W and 37^{o}N - 34^{o}N (Figure 2), which is relatively flat and well-covered by radars and the PERSIANN data. We selected a period of 1 month of data (May 1, 2002 to May 30, 2002; 720 hourly time steps).

4.0 METHODOLOGY AND RESULTS

We assessed the nine error metrics at seven spatial scales: (i) 0.04 degree (original); (ii) 0.08 degree; (iii) 0.12 degree; (iv) 0.16 degree; (v) 0.24 degree; (vi) 0.48 degree and (vii) 1.0 degree. The lower end of this range is considered more relevant to hydrologic modeling, while the higher end is typical of many long-term satellite precipitation products such as the GPCP products (Huffman et al., 2001) and IPWG (Ebert et al., 2007). Note, however, that a statistically significant sample for spatial correlation lengths is not possible at the two largest scales due to the size of the study region (5.0° X 3.0°) and hence, these values have not been reported. In addition, three other commonly used diagnostic metrics such as frequency bias (FB), false alarm ratio (FAR) and equitable threat score (ETS) were also evaluated (see Appendix 2 for their mathematical formulation).

Using a very simple cropping technique, the WSR-88D data were remapped to the 0.04° PERSIANN grid to allow consistent comparisons. We subsequently verified that the cropping-based interpolation had no effect on the statistics of the data. The temporal resolution was kept fixed at hourly. The spatial scales of aggregation are intentionally chosen to be integer multiples of the original 0.04°grid to avoid spatial interpolation errors, which was found to be problematic in our preliminary investigation. This choice allowed us to focus on the scaling behavior of error parameters purely as a function of aggregation, seeking to identify how each of the nine error metrics responds to spatial scaling and whether there exists some minimum scale at which some or most of the error parameters remain "acceptable" for the hydrologist user.

Ordinarily, one would expect the data producer to use error metrics that are robust to changes in scale for the sake of consistency. However, such an approach for evaluating error metrics may not provide the best insight into applying satellite rainfall data at hydrologically relevant scales. As an example, consider the case when the spatial scale for a data product decreases from 0.24 degree to 0.16 degree or 0.12 degree as part of algorithm enhancement involving spatial downscaling. The correlation coefficient or systematic bias may register a change with scale that is considerably more modest than changes in the algorithm's ability to correctly delineate the rainy or non-rainy areas, given the intermittency of the rainfall process. This is because marginal measures such as correlation coefficient are parameters that reflect essentially the aggregate effect of the algorithm's ability to retrieve rainfall over the study area (also discussed later). But the intermittency has important implications for hydrologic simulation of the terrestrial water cycle and must be considered in evaluating the use of satellite data.

In Table 1, we summarize results for the metrics correlation coefficient, RMSE, FB, FAR and ETS. The conditional correlation refers to the cases when both reference and satellite rain is non-zero. In general, we observe that the response to spatial scale is similar for all these conventional metrics, often appearing somewhat mild at scales smaller than 0.24 degree. For example, Figure 3 shows that the correlation coefficient behavior graphically as a function of scale. The sensitivity of correlation appears modest and yet it remains difficult a metric to be used in an error model. A similar assessment can be made for FAR, FB and ETS. Obviously, the hydrologist wishing to quantify the implication of each of these metrics on terrestrial simulation would expect to be able to use it mathematically for simulation of equi-probable realizations of satellite data and error propagation experiments.

An interesting pattern emerges in Figure 4 for our proposed metric on probability of rain detection – POD_{RAIN} as a function of the magnitude of rain rate. With spatial aggregation, both the maximum POD_{RAIN} (at reference rain rates greater than 15 mm/hr) and the gradient of probability detection as a function of reference rainfall rate increase noticeably, as expected (the highest being for 1.0 degree). Compared to the probability of no-rain detection, POD_{NORAIN} clearly exhibits stronger scale dependence (Figure 5). The probabilities for both rain and no-rain detection decrease definitively as spatial scales decrease below 0.48 degree. Hence, these two metrics, that account for rainfall intermittency, add to the existing value of traditional metrics for exploring satellite rainfall data application in hydrologic models.

In Figure 6, we show the spatio-temporal structure for metrics on conditional retrieval error (random; upper left panel), mean field bias (temporal autocorrelation function; upper right panel), and the spatial correlation functions for rain detection (lower left panel) and no-rain detection (lower right panel). Even though the distinction in spatial scaling for these spatio-temporal error parameters is weak at the scales considered, these metric can be used coherently in an error model to simulate realistic space-time covariance structure of satellite rainfall data (Hossain and Anagnostou, 2006). If correlation lengths in space and time are computed (assuming that an exponential model is appropriate; see Appendix 1), a more informative picture emerges (Figure 7). We observe a clear sensitivity of the correlation length to spatial scales, with the correlation length for rain detection being the most sensitive. The lag-one (hourly) temporal correlation of mean field bias however remains insensitive to scale as would be expected since the domain's area is the same for all grid sizes. These results demonstrate that the suggested error metrics on rain detection/delineation correlation lengths give insight into the useful spatial scales for applying satellite-based rainfall in hydrologic models. While the assessment of the direct implications of these metrics on hydrologic modeling is beyond the scope of this study, Hossain and Anagnostou (2004) have shown that an improvement in the probability of rain detection can yield substantial improvements in flood prediction at the 0.1^{o} scale for saturation-excess watersheds in the Alps. Intuitively, the same can be expected of Hortonian watersheds where spatial pattern of the rainy areas along with the rain rate and soil's infiltration capacity dictate the propensity of a region to produce direct runoff.

Finally, in order to demonstrate the value of our proposed error building framework for distinguishing the strengths and weaknesses of existing algorithms, we assessed four global satellite rainfall products. These were: 1) NASA's Infrared Rainfall (IR) product 3B41RT (Huffman et al., 2007); 2) NASA's Merged Microwave Rainfall product 3B42RT (Huffman et al., 2007); 3) NOAA CPC Passive Microwave Rainfall product CMORPH (Joyce et al., 2004), and 4) PERSIANN. Error analysis was performed for the year 2004 over two regions in the US known to have a distinct hydro-climatology – i) Mid-western US (a semi-arid zone), and ii) Florida (sub-tropical zone modulated by coastal effects). WSR-88D Stage-II rainfall data was used as reference for ground validation data. Tables 2a and 2b show a summary of correlation, standard deviation of error and the lag-one autocorrelation of mean field bias. In Figure 8, the variation of

POD_{RAIN} during the winter (January and February) and summer (June, July and August) is shown. It is very clear from the tables 2a and 2b that the correlation metric fails to highlight any major differences in quality among the algorithms as a function of region and season. On the other hand, the lag-one autocorrelation and the POD_{RAIN} show clear sensitivity across algorithms, regions and seasons.

5.0 CONCLUSION AND FUTURE NEEDS

Representing the error structure of satellite rainfall as a function of scale against quality-controlled ground validation datasets remains a critical research problem. Hydrologists and other users, to varying degrees, need to know the errors of the satellite rainfall data sets across the range of time/space scales over the whole domain of the data set. In this study, we investigated the behavior of a suggested set of error metrics that were linked primarily to rainfall intermittency for a microwave-calibrated geostationary infrared based algorithm. In general, the conventional error metrics such as correlation coefficient, frequency bias, false alarm ratio and equitable threat score appeared to have similar levels of sensitivity to scale. Yet, the use of these common metrics for simulating probabilistic realizations of satellite rainfall with realistic space-time covariance structures is not feasible. This limits the value of the metrics to the hydrologist who may choose to probabilistically quantify the implications of each metric on overland hydrologic simulations. The probability of detection of rain and its functional relationship to ground validation rainfall were found to be the most sensitive to scale followed by the correlation length for detection of rain. These specific error metrics appear informative for identifying the useful scales for data integration over land and can also be used in a stochastic error model.

The data sets envisioned for GPM pose significant opportunities for hydrologists, but with associated challenges. Specifically, effective assessment frameworks and metrics for satellite precipitation data must be developed that enhance GPM's utility for land surface hydrology and are jointly defined by the hydrologic and data-producing communities. While we have shown some tangible examples of error metrics and their potential value in gauging utility, more work is needed to define hydrologically relevant metrics that can connect more directly to the physics and geometries of satellite rainfall estimation.

The practicality of the approach presented in this paper can be questioned because the details likely vary by region and season, and because many regions lack the necessary ground validation data to develop region-specific error representations. However, it appears conceptually feasible to build on work already accomplished on global classification of precipitation systems (Petersen and Rutledge, 2002). In addition, it is possible to use the TRMM Precipitation Radar (PR) as the reference for the spatial domain (e.g., Hossain and Anagnastou, 2004), and apply a recent approach suggested by Bellerby and Sun (2005) based on transfer of probability distribution functions for the temporal domain. Another approach could be the use of geostatistical simulation techniques such as ordinary or indicator kriging to transfer error metrics from a ground validation site to ungauged regions under the assumption of stationarity of the metric values.

A challenge that remains is to choose a small set of error parameters that enable practical use of the uncertainty information and capture the time/space structure of the uncertainties. At this level of complexity it might be best to establish functional forms for the error metrics and supply coefficients for data- and algorithm-driven choices for large regions and seasons. In some cases average or "climatological" coefficients might suffice, while in other cases routine updates as a function of time might be required. Given this information a user could easily estimate the errors that correspond to the time/space scale of their application. In particular, a hydrologist could identify the necessary scale of aggregation of satellite rainfall data to achieve a specified level of accuracy that would minimize error propagation in a hydrologic model (Harris and Hossain, 2007).

6.0 ACKNOWLEDGEMENTS: The authors wish to thank Dr. Soroosh Sorooshian of University of California-Irvine, for providing the PERSIANN satellite rainfall data. The authors also gratefully acknowledge the comments received from three anonymous reviewers which substantially improved the quality of the manuscript. The first author was supported by a research initiation grant from the Tennessee Technological University. Additional support from the Center for Management, Utilization and Protection of Water Resources of Tennessee Technological University is also acknowledged.

7.0 REFERENCES

Ali, A., T. Lebhl, A. Amani, 2005: Rainfall Estimation in the Sahel. Part I: Error Function. *J. Appl. Meteor.*, **44**, 1691-1706.

Anagnostou, E.N., 2004: Overview of overland satellite rainfall estimation for hydro-meteorological applications. *Surveys in Geophysics,* **25**, 511–537.

Astin I., 1997: A survey of studies into errors in large scale space-time averages of rainfall, cloud cover, sea surface processes and the earth's radiation budget as derived from low orbit satellite instruments because of their incomplete temporal and spatial coverage. *Surveys in Geophysics,* **18**, 385-403.

Bellerby, T., and J. Sun, 2006: Probabilistic and ensemble representations of the uncertainty in IR/Microwave precipitation product, *Journal of Hydrometeorology,* **6**.

Beven, K.J., A.M. Binley, 1992: The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes,* **6**, 279–298.

Dai, Y., X. Zeng, R.E. Dickinson, I. Baker, G.B. Bonan, M.G. Bosilovich, S. Denning, P. Dirmeyer, P.R. Houser, G. Niu, K.W. Oleson, A. Schlosser, Z.-L. Yang, 2003: The common land model. *Bull. Amer. Meteor. Soc.*, **84**, 1013–1023.

Ebert, E.E., J.E. Jonowiak, C. Kidd, 2007: Comparison of near real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, (To appear).

Fulton, R.A., J.P. Breidenbach, D.-J. Seo, D.A. Miller, T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Weather and Forecasting,* **13**, 377–395.

Gebremichael, M. W.F. Krajewski, 2004: Characterization of the temporal sampling error in space-time averaged rainfall estimates from satellites. *J. Geophys. Res. (Atmos)*, **109**, D1110 (doi: 10.1029/2004JD).

Gebremichael, M., W.F. Krajewski, 2005: Modeling distribution of temporal sampling errors in area-time averaged satellite rainfall estimates. *Atmospheric Research,* **73**, 243-259.

Harris, A., and F. Hossain, 2007: Investigating the Optimal Configuration of Conceptual Hydrologic Models for Satellite Rainfall-based Flood Prediction in the Upper Cumberland River. *Journal of Hydrometeorology* (In review; copy available: http://iweb.tntech.edu/fhossain/papers/JHM_Amanda.pdf).

Heidke, P., 1926: Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst (In German). *Geogr. Ann*. **8**, 301-349.

Hong, Y., K.-L. Hsu, S. Sorooshian, X. Gao, 2005a: Self-organizing nonlinear output (SONO): A neural network suitable for cloud patch–based rainfall estimation from satellite imagery at small scales. *Water Resources Research*, **41**(W03008), (doi:10.1029/2004WR003142).

Hong, Y., K.-L. Hsu, H. Moradkhani, S. Sorooshian, 2006: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resources Research*, **42**(8), (W08421, doi.10.1029/2005WR004398).

Hong, Y., K.-L. Hsu, S. Sorooshian, X. Gao, 2005b: Improved representation of diurnal variability of rainfall from the Tropical Rainfall Measuring Mission Imager adjusted Precipitation Estimation from Remotely-sensed Information using

Artificial Neural Networks (PERSIANN) system. *J. Geophys. Res.*, **110**(D06102) (doi: 10.1029/JD2004JD005301).

Hossain, F., and E.N. Anagnostou. 2005: Using a multi-dimensional satellite rainfall error model to characterize uncertainty in soil moisture fields simulated by an offline land surface model. *Geophys. Res. Lett,* **32**, L15402 (DOI.10.1029/2005GL023122).

Hossain, F., D.P. Lettenmaier, 2006: Flood prediction in the future: recognizing hydrologic issues in anticipation of the global precipitation measurement mission - Opinion Paper. *Water Resources Research.* **44** (doi:10.1029/2006WR005202)

Hossain, F., E.N. Anagnostou, 2006: A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Rem. Sens.,* **44** 1511-1522 (doi: 10.1109/TGRS.2005.863866).

Hossain, F., E.N. Anagnostou, 2004a: Assessment of current passive microwave and infrared based satellite rainfall remote sensing for flood prediction. *J. Geophys. Res. (Atmos.)*, **109**:D07102.

Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B. McGavock, J. Susskind, 2001: Global precipitation at one-degree daily resolution from multi-satellite observations. *J. Hydrometeor*., **2**, 36-50.

Huffman, G.J., R.F. Adler, D.T. Bolvin, E.J. Nelkin, 2004: The status of the multi-satellite precipitation analysis and insights gained from adding new data sources. *Second International Precipitation Working Group Workshop, 25-28* October 2004, Monterey, California, 123-127.

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, D.B. Wolff, 2007: The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeor.*, **8**, 38-55.

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

Janowiak, J., R.J. Joyce, Y. Yarosh, 2001: A real-time global half-hourly pixel-resolution infrared datasets and its application. *Bull. Amer. Meteor. Soc.,* **82**, 205-217.

Lee, K.H., E.N. Anagnostou, 2004: Investigation of the nonlinear hydrologic response to precipitation forcing in physically based land surface modeling. *Canadian J. Rem. Sens.*, **30**, 706-716.

Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian, 2006: Investigating the Impact of Remotely Sensed Precipitation and Hydrologic Model Uncertainties on the Ensemble Streamflow Forecasting. *Geophys. Res. Lett.*, 33(12), L12401, (DOI.10.1029/2006GL026855).

McCollum, J.R., W.F. Krajewski, RR. Ferraro, M.B. Ba, 2002: Evaluation of biases of satellite rainfall estimation algorithm over continental US. *J. Appl. Meteor.,* **41** 1065-1080.

Nijssen, B., and D.P. Lettenmaier: 2004: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *J. Geophys Res.* **109**, D02103.

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

Petersen, W.A., S.A. Rutledge, 2002: Regional variability in tropical convection: observations from TRMM. *J. Climate,* **14**, 3566-3586.

Seo, D-J., J.P. Breidenbach, R. Fulton, D. Miller and T. O'Bannon, 2000: Real-time adjustment of range-dependent biases in WSR-88D rainfall estimates due to nonuniform vertical profile of reflectivity. *J. Hydrometeor.*, **1**, 222-240.

Seo, D-J, J.P. Breidenbach, E.R. Johnson, 1999: Real-time estimation of mean field bias in radar rainfall data. *J. Hydrol.*, **223**, 131-147.

Shiklomanov, A.I., R.B. Lammers, C.J. Vörösmarty, 2002: Widespread decline in hydrological monitoring threatens pan-arctic research. *EOS Trans.* **83**, 16–17.

Smith E., G. Asrar, Y. Furuhama, A. Ginati, C. Kummerow, V. Levizzani, A. Mugnai, K. Nakamura, R.Adler, V. Casse, M. Cleave, M. Debois, J. Durning, J. Entin, P. Houser, T. Iguchi, R. Kakar, J. Kaye, M.Kojima, D.P. Lettenmaier, M. Luther, A. Mehta, P. Morel, T. Nakazawa, S. Neeck, K. Okamoto, R. Oki, G. Raju, M. Shepherd, E. Stocker, J. Testud, E.F. Wood, 2006: The international Global Precipitation Measurement (GPM) program and mission: An overview. *Measuring Precipitation from Space: EURAINSAT and the Future,* V. Levizzani and F.J. Turk, Editors, Kluwer Academic Publishers (In Press).

Sorooshian, S., K.-L. Hsu, X. Gao, H.V. Gupta, B. Imam, D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.

Steiner, M. 1996: Uncertainty of estimates of monthly areal rainfall for temporally sparse remote observations. *Wat. Resour. Res.* **32**, 373–388.

Steiner, M., T.L. Bell, Y. Zhang, and E.F. Wood, 2003: Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar dataset. *J. Climate* **16**, 3759–3778.

Stokstad, E., 1999: Scarcity of rain, stream gages threatens forecasts. *Science*, **285**, 1199.

Villarini, G., and W.F. Krajewski, 2007: Evaluation of the research version of the TMPA three hourly 0.25^{o} X 0.25^{o} rainfall estimates over Oklahoma. *Geophysical Research Letters*, vol. 34(LO5402).

APPENDIX 1

Formulation of Error Metrics

Consider first, the following confusion matrix for hits and misses associated with satellite rainfall estimates:

Truth/Reference

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

Probability of Detection for Rain (POD_{RAIN}): $$\frac{N_A}{N_A + N_B} \qquad (1)$$

Probability of Detection for No Rain (POD_{NORAIN}): $$\frac{N_D}{N_D + N_C} \qquad (2)$$

We also define the (successful) rain detection probability, POD_{RAIN}, as a function of rainfall magnitude of either the reference rainfall or satellite estimate. The functional form is usually identified through calibration with actual data (see Hossain and Anagnostou, 2006). The POD_{NORAIN}, is the unitary probability that satellite retrieval is zero when reference rainfall is zero, which is also determined on the basis of actual data.

A probability density function (D_{false}) is defined to characterize the probability distribution of the satellite estimates when there are misses over non-rainy areas. This function is also identified through calibration on the basis of actual sensor data. Hossain and Anagnostou (2006) have reported that this D_{false} probability density function typically tends to appear exponential. Hence, both the moments (first and second) can be defined using only one parameter of the distribution, λ. This can be computed using the chi-squared or maximum likelihood method.

To identify the correlation lengths of error (i.e., *how does the error vary in space*) a simple exponential type auto-covariance function can be assumed. The correlation length (the separation distance at which correlation $= \frac{1}{e} = 0.3678$) is then determined on the basis of calibration with actual data over a large domain (the size of Oklahoma in this study). For identifying the spatial correlation length of rain detection, CL_{RAIN} (or, no-rain detection - CL_{NORAIN}) from data, all successfully detected rainy (non-rainy) pixels are assigned a value of 1.0 while the rest has a value of 0.0. The empirical semi-variogram is then computed as follows:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \left(z(x_i) - z(x_i + h) \right)^2 \qquad (3)$$

where $z(x_i)$ and $z(x_i+h)$ are the binary pixel values (0 or 1) at distance x_i and x_i+h, respectively and h is the lag in km. n represents the number of data points at a separation distance of h. The term $\gamma(h)$ is the semi-variance at separation distance h. Assuming that the empirical variogram is best represented by an exponential model, the functional parameters describing the spatial variability can be fitted as follows,

$$\gamma(h) = c_0 + c\,(1 - e^{-h/CL}) \qquad (4)$$

where c_0 represents the nugget variance, c is the sill variance and CL is the distance parameter known as "correlation length". Conversely, the correlation function is modeled as, C = EXP(-h/CL), where C is the correlation.

For identifying the correlation length for retrieval error, CL_{RET}, a similar set of steps are adopted as above for rain/no rain detection, with the exception that the binary values (0-1) are no longer pertinent. Instead, one computes the correlation length in terms of retrieval error defined as the difference between reference and satellite estimate as described below.

The conditional (i.e., reference rainfall >0 unit) non-zero satellite rain rates, R_{SAT}, can be statistically related to corresponding conditional reference rain rates, R_{REF}, as,

$$R_{SAT} = R_{REF-INST} . \varepsilon_S \quad (5)$$

where the satellite retrieval error parameter, ε_s, is assumed to be log-normally distributed. It is up to the data producers to verify the assumption of log-normality. The advantage of such an assumption is that a log transformation [log(R_{SAT})-log($R_{REF-INST}$)] of Equation 5 transforms the ε_s to a Gaussian $N(\mu,\sigma)$ deviate, ε, where μ and σ are the mean and standard deviation of retrieval error, respectively.

The retrieval error parameter ε is both spatially and temporally auto-correlated. The spatial aspect has already been discussed earlier in this appendix. For temporal correlation, a lag-one autocorrelation function is used to identify the temporal variability of μ (i.e., conditional satellite rainfall bias).

APPENDIX 2

Formulation of some common error metrics

Using the terminology adopted by Ebert et al. (2007), a gridbox can be classified as a hit (*H,* observed rain is correctly detected), miss (*M,* observed rain is not detected), false alarm (*F,* rain detected but not observed), or null (no rain observed or detected).

The frequency bias (FB) is defined as,

$$\text{FB}=\frac{(H+F)}{(H+M)} \tag{6}$$

The false alarm ratio (FAR) is defined as,

$$\text{FAR}=\frac{F}{(H+F)} \tag{7}$$

The equitable threat score (ETS) is defined as,

$$\text{ETS}=\frac{H-H_e}{(H+M+F-H_e)} \tag{8}$$

Where, $H_e=(H+M)(H+F)/N$ and N=the total number of gridboxes

8.0 TABLES

Table 1. Some commonly used error metrics as a function of spatial scales degrees of lat./long.).

	0.04°	0.08°	0.12°	0.16°	0.24°	0.48°	1.0°
Correlation (Unconditional)	0.386	0.383	0.393	0.401	0.418	0.469	0.569
Correlation (Conditional)	0.272	0.298	0.319	0.334	0.361	0.437	0.547
Root Mean Squared Error (mm hr^{-1})	4.708	3.933	3.530	3.230	2.776	1.98	1.25
Frequency Bias (FB)	1.524	1.405	1.423	1.419	1.460	1.548	1.677
False Alarm Ratio (FAR)	0.686	0.634	0.619	0.601	0.5804	0.5370	0.5066
Equitable Threat Score (ETS)	0.205	0.235	0.245	0.255	0.271	0.302	0.306

Table 2a. Summary of some error metrics over Oklahoma during winter and summer of 2004.

Winter

	Correlation (conditional)	Error Std Dev (conditional) (mm/hr)	lag-one correlation
CMORPH	0.226	2.19	0.540323
3B41	0.106	1.89	0.855313
3B42	0.18	1.95	0.467897
PERSIANN	0.268	1.58	

Summer

	Correlation (conditional)	Error Std Dev (conditional) (mm/hr)	lag-one correlation
CMORPH	0.251	2.47	0.641784
3B41	0.274	2.31	0.911418
3B42	0.216	2.32	0.491956
PERSIANN	0.303	1.72	

Table 2b. Summary of some error metrics over Florida during winter and summer of 2004.

Winter

	Correlation (conditional)	Error Std Dev (conditional) (mm/hr)	lag-one correlation
CMORPH	0.195	2.41	0.6277
3B41RT	0.102	2.18	0.9603
3B42RT	0.121	2.18	0.5635
PERSIANN	0.159	1.79	0.7210

Summer

	Correlation (conditional)	Error Std Dev	lag-one correlation
CMORPH	0.153	2.52	0.4282
3B41RT	0.219	2.38	0.8959
3B42RT	0.220	2.29	0.3712
PERSIANN	0.248	1.84	0.7105

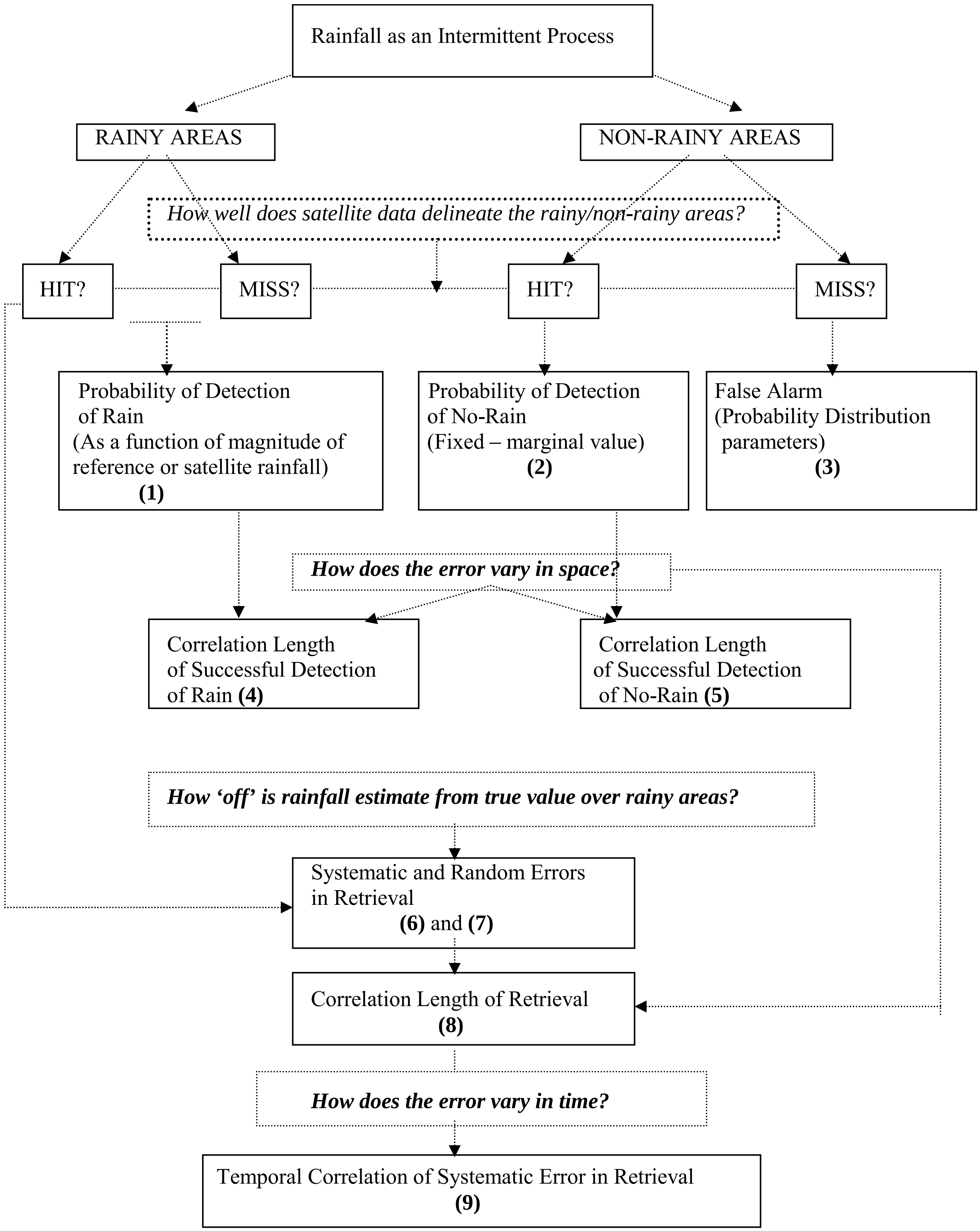

Figure 1. The logical thought process to building the conceptual framework for

hydrologically relevant error metrics. Numbers in parentheses denote the metrics numbered in section 2.

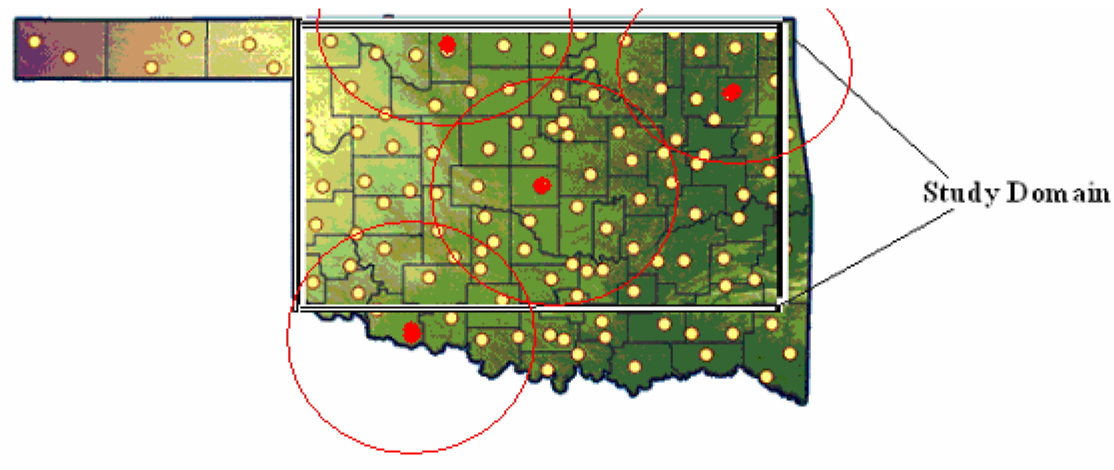

Figure 2. Study region over Oklahoma bounded by 100^{o}W – 95^{o}W and 37^{o}N - 34^{o}N. [The yellow dots show the location of the Oklahoma Meso-network meteorological stations with gauge rainfall data for WSR-88D radar calibration. Red dots indicate the location of WSR-88D radars inside Oklahoma; circles approximately indicate coverage with 100km radius].

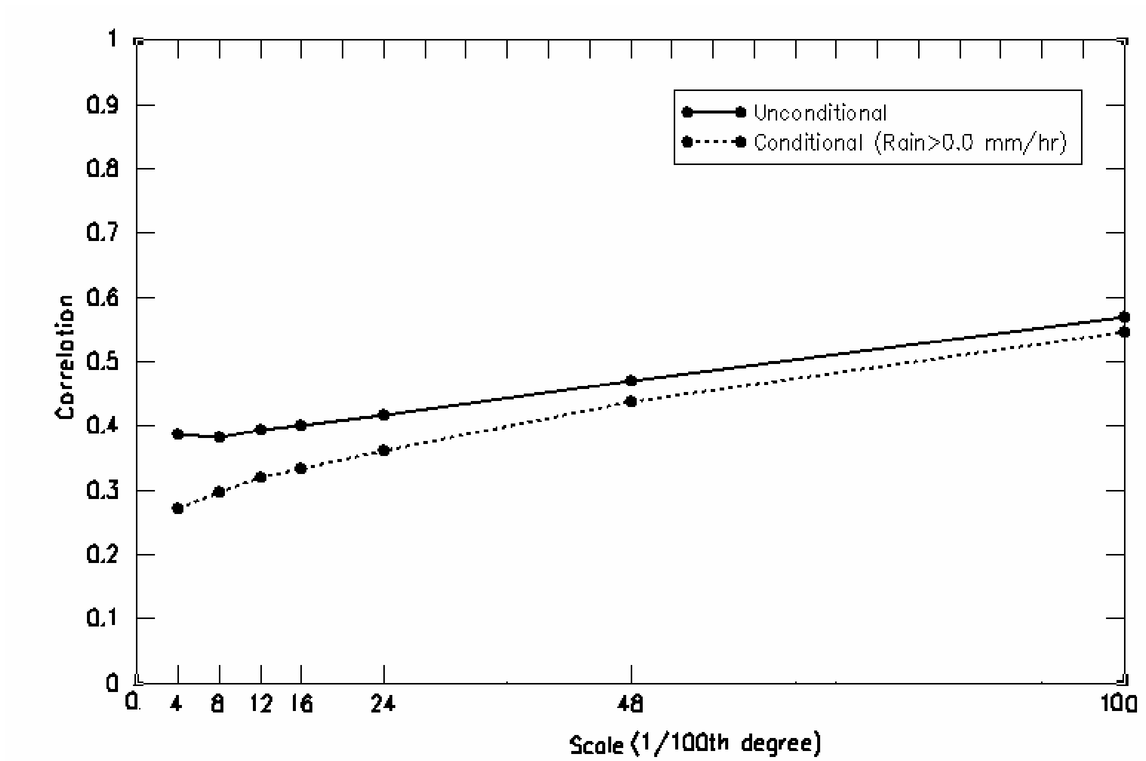

Figure 3. Correlation coefficient of satellite rainfall data with WSR-88D radar rainfall as a function of spatial scale.

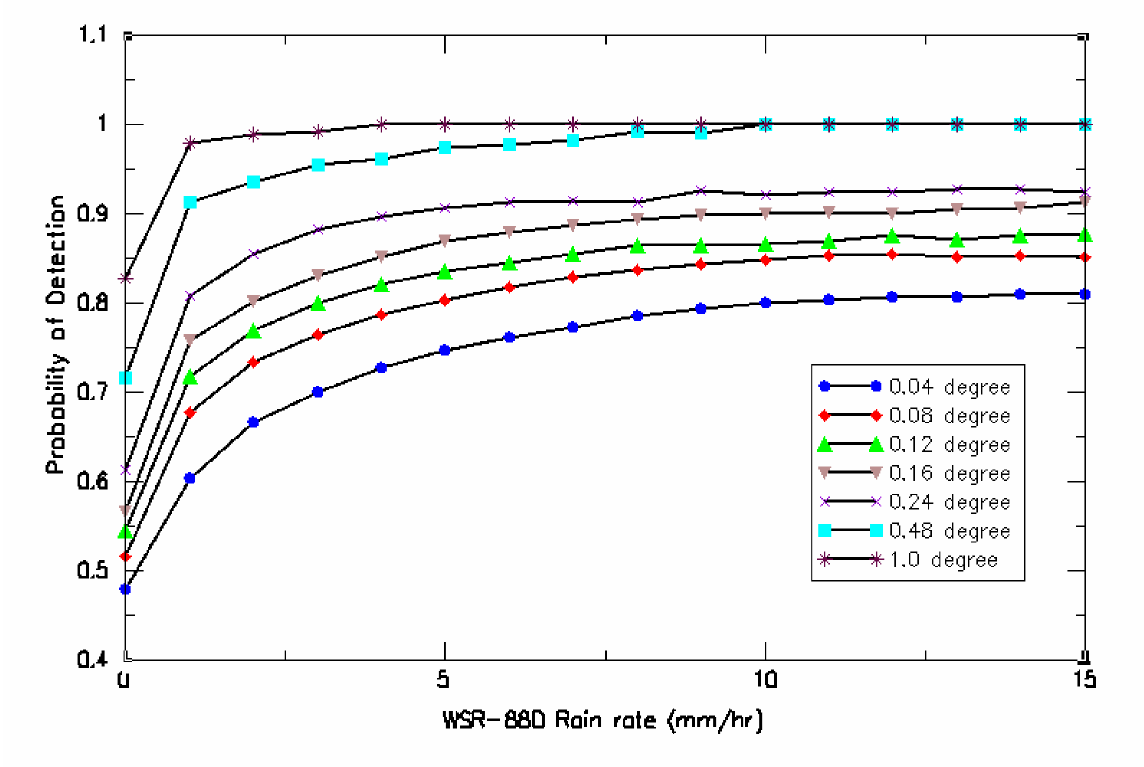

Figure 4. Probability of rain detection as a function of reference (WSR-88D) rainfall rate and spatial scales.

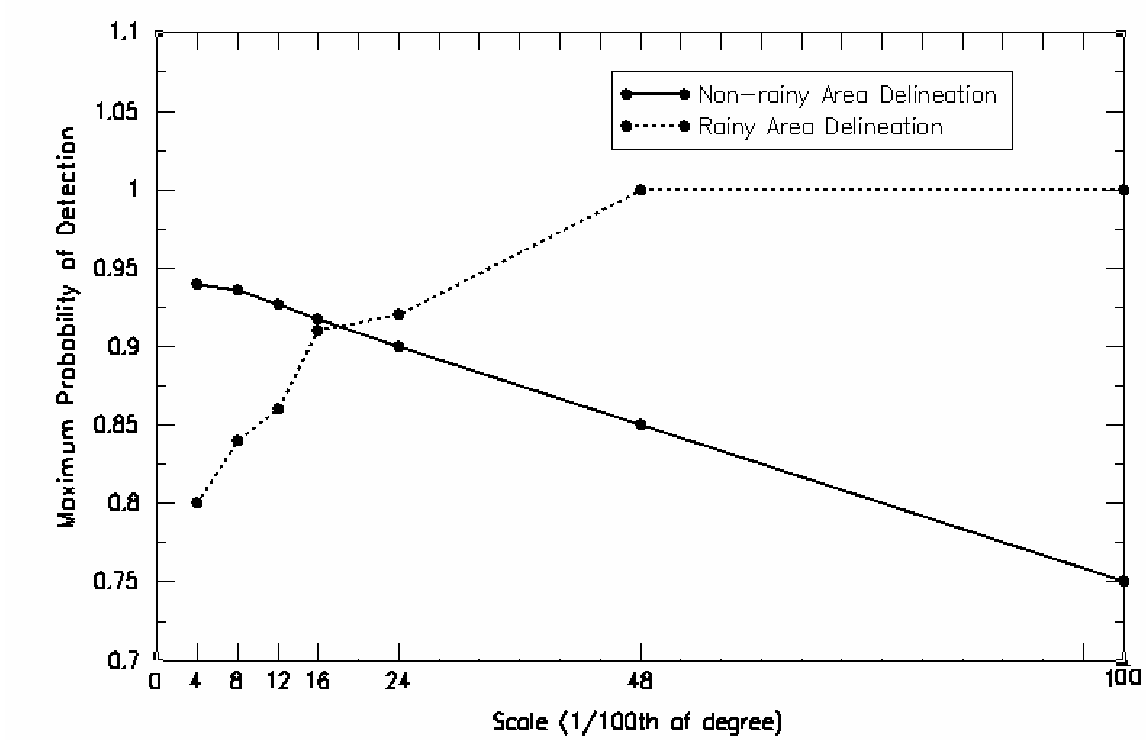

Figure 5. Comparison of scaling behavior of maximum probability for rain (at rain rates >15 mm/hr) and marginal probability of no-rain detection.

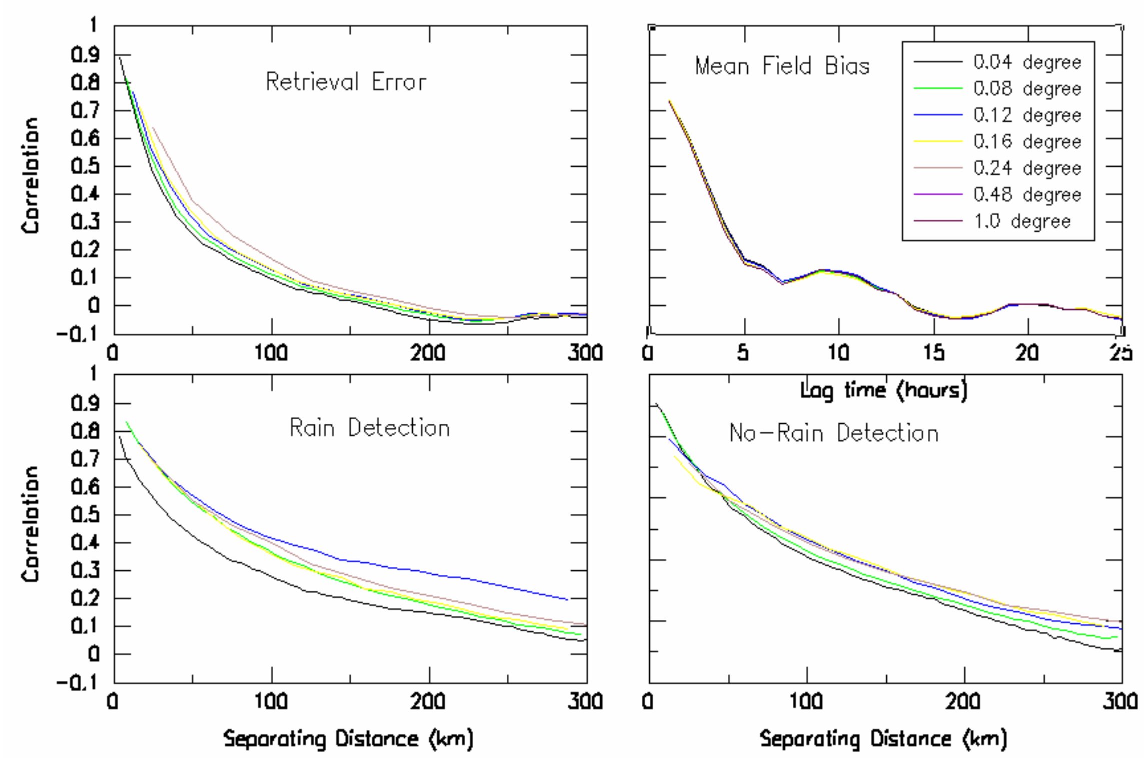

Figure 6. Spatio-temporal metrics as a function of spatial scales. Upper left panel – spatial correlation function for conditional retrieval error (rain> 0); Upper right panel – temporal correlation function for mean field bias; Lower left panel – spatial correlation function for rain detection and Lower right panel – spatial correlation function for no-rain detection. Note: spatial correlation functions are not reported at scales 0.48 degree and 1.0 degree due to the small spatial sample available over the study region. The separating distance is reported in km by making an approximate assumption that 1 degree is equivalent to 100 km.

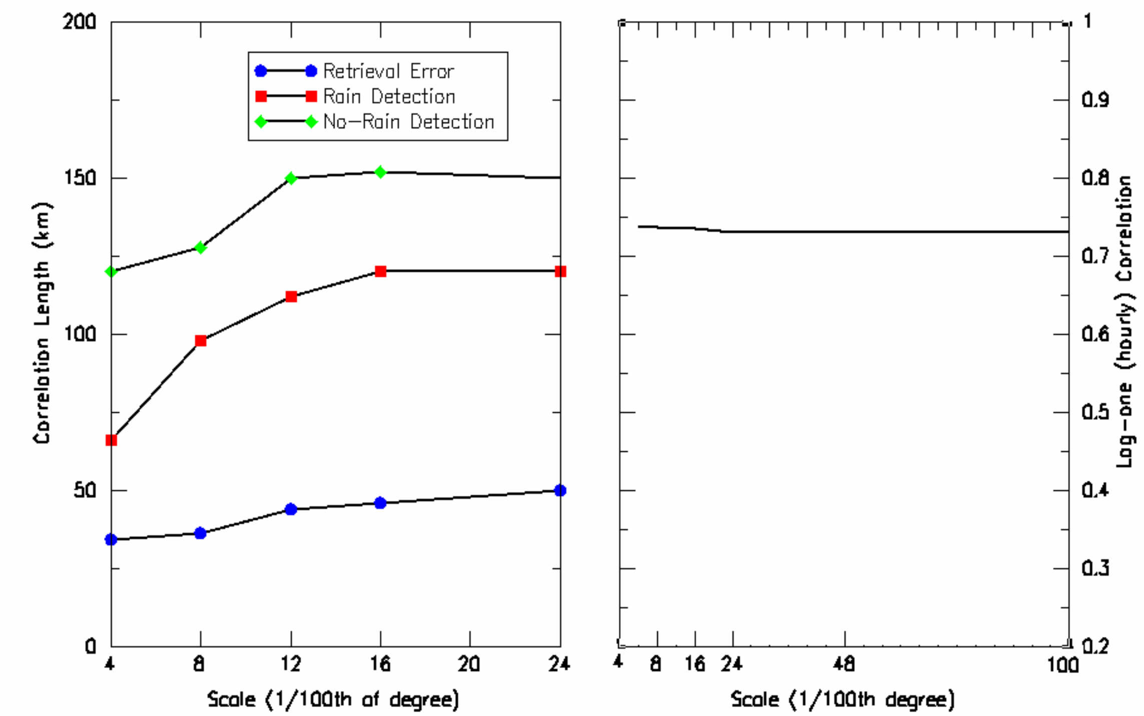

Figure 7. Correlation lengths and lag-one autocorrelation as a function of scale (assuming an exponential model is appropriate to describe the correlation function in space). [Note: spatial correlation lengths are not reported at scales 0.48 degree and 1.0 degree due to the small spatial sample available for the study region.]

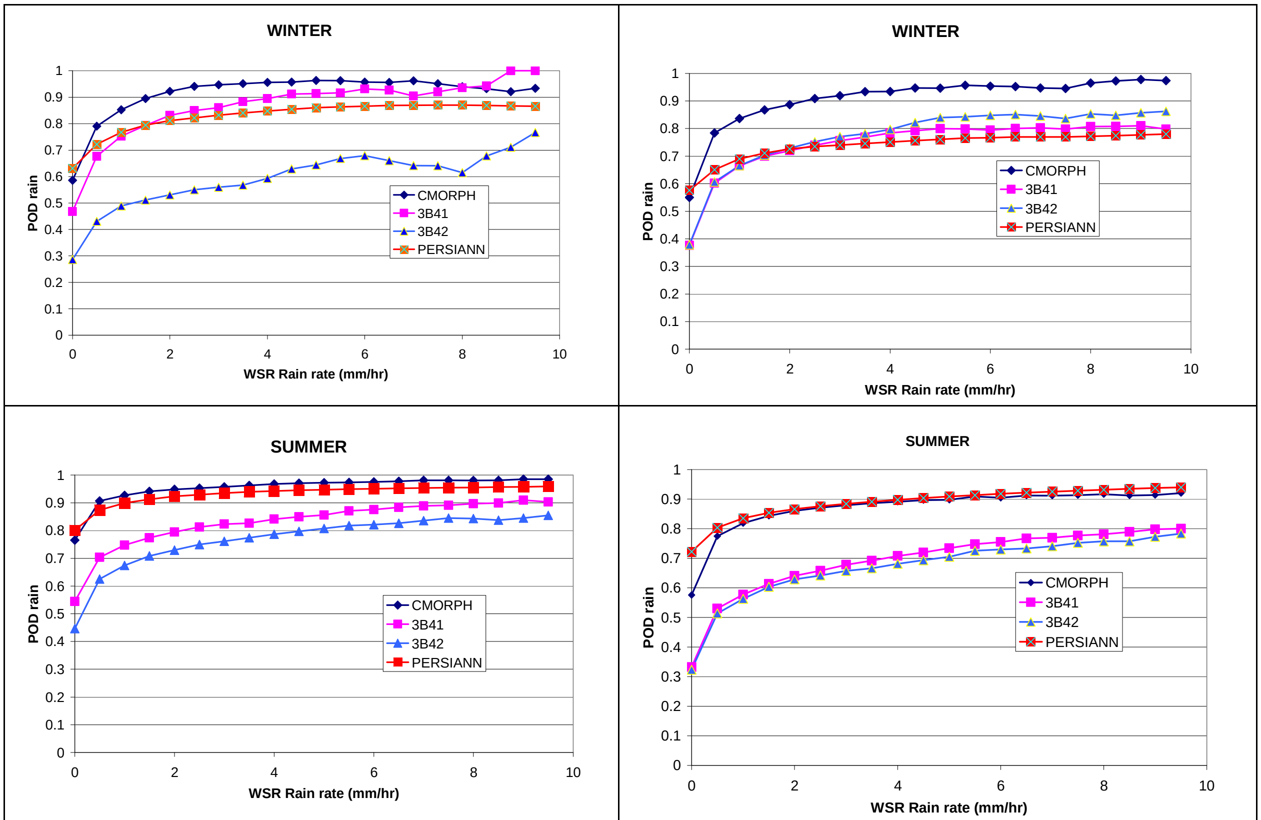

Figure 8. Probability of detection of rain (***POD$_{RAIN}$***) as a function of season, region, algorithm and reference (GV) rain rate. Left panels – Oklahoma (semi-arid hydrology); Right panels – Florida (coastal hydrology).