Title
A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products

Author
Sarachi, Sepideh

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,

IRVINE

A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products

DISSERTATION

Submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Sepideh Sarachi

Dissertation Committee:

Professor Soroosh Sorooshian, Chair

Professor Kuo-lin Hsu

Professor Padhraic Smyth

2015

DEDICATION

To

My Mom and Dad who enlighten my path and

My sister who witnesses my journey

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Acknowledgements

Now that this chapter of my life as a graduate student at UCI is coming to its end, I reflect on myself and see that I cannot be more blessed and thankful for my life. I owe this to the amazing group of people who helped me be where I am today. Words cannot describe how thankful I am to have Professor Sorooshian as my advisor as well as a very important part of my new family since I moved here. His dedication not only to the science community, also to help and nourish students is always an inspiration for me. I would also like to thank Dr. Hsu, who has always been patient and motivational throughout my research. I should thank my other committee member Dr. Smyth from whom I learnt some fundamentals of my work.

Among people, who without their help, this achievement could not be possible are Mrs. Diane Hohnbaum with her assistant and attention and Mr. Dan Braithwaite with his technical support.

During my study at CHRS, I could not ask for better colleagues and friends, specifically Nasrin Nasrollahi and Andrea Thorstensen who are my sisters by choice. Also Parisa Jalal Kamali and Parnian Arjmand who have been my best friends and my main support system.

My most sincere thank goes to my family, who despite the time difference, never made me feel their absence. My sister was always there to brighten my days and nights. My dad always reminded me how strong I am and my mom always reassured me that having a kind heart is the biggest goal in our lives.

Primary financial support for this study was provided by the NASA Precipitation Measurement Mission (grant NNX10AK07G and NNX13AN60G), the U.S. Army Research Office (grant W911NF-11-1-0422), and the NOAA NCDC (Prime Award NA09NES4400006, NCSU CICS Sub-Award 2009-1380-01).

CURRICULUM VITAE

SEPIDEH SARACHI

EDUCATION & LICENSURE

Ph.D. ~ CIVIL & ENVIRONMENTAL ENGINEERING (6/2015)

MASTER OF SCIENCE ~ CIVIL & ENVIRONMENTAL ENGINEERING (6/2010)

University of California, Irvine

CALIFORNIA ENGINEER-IN-TRAINING CERTIFICATE (2014)

BACHELOR OF SCIENCE ~ CIVIL & ENVIRONMENTAL ENGINEERING (6/2009)

Sharif University of Technology, Tehran, Iran

Honors, Awards, & Fellowships

- Travel Grant Recipient ~ Association of Graduate Students at University of California, Irvine (2014)
- Research and Travel Grant Recipient ~ Henry Samueli School of Engineering (2013)
- National Fellowship Incentive award (2013,2014)
- 2012 Public Impact Fellowship

RESEARCH & TEACHING EXPERIENCE

GRADUATE RESEARCH ASSISTANT, Center for Hydrometeorology & Remote Sensing 9/2011 - Present

Conducted statistical modeling and data analysis and applications in hydrology, remote sensing, civil engineering design and calculations, hydrologic/hydraulic modeling, groundwater modeling, and the study of climate variability/change and its effect on communities.

- Performed precipitation and drought estimation and analysis.
- Worked with IPCC future climate model scenarios for water resource management and drought analysis, and conducted sustainability assessments and data analysis.
- Wrote project report, presented project details in meetings and conferences, wrote grants and proposals, and prepared maps, charts and complicated analysis related to water resource systems.
- Conducted statistical and uncertainty analyses, risk assessment and risk management, machine learning; presented at two conferences and published at one scientific journal.

GRADUATE RESEARCH ASSISTANT, University of California, Irvine
9/2009 - 9/2011

Served as a member of the "Andes-Amazon Initiative" project focused on Amazon deforestation.

- Provided 100-year time series of climatological data; precipitation, temperature, wind, and solar radiation.
- Worked with IPCC future climate model scenarios for water resource management and drought analysis.
- Conducted water resource management, hydrology and hydrologic modeling; utilized Matlab and Fortran programming to model climate change, as well as hydrological and statistical concepts.
- Analyzed behavioral factors in climate change, agriculture, animal husbandry, and deforestation pattern recognition.

TEACHING ASSISTANT ~ Introduction to Hydrology, University of California-Irvine

Assisted with the instruction of classes to graduate and under graduate students, prepared lectures on different topics such as precipitation measurement, unit hydrograph, and routing. **9/2014 - 12/2014** **9/2013 - 12/2013** **9/2011 - 12/2011**

- Conducted office hours, led problem solving sessions, prepared and graded homework assignments and exams, and defined both class projects and various class activities.

INTERNHIP ~ Tehran Oil Refinery Expansion" project, Chagalesh Consulting Firm

Structural and foundation design, heavy foundation design and analysis, **9/2008 - 9/2009**

and heavy structural analysis.

• Soil sampling, field investigation, sample analysis, data base preparation and management
• Concrete structural and foundation design and reinforcement, working with design standards and codes
• Reviewing technical documents, writing repots

JOURNAL PUBLICATIONS & CONFERENCE PRESENTATIONS

Sarachi S., K. Hsu, S. Sorooshian, A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products., 2015, *J. Hydrometeorol*., In press, Ref.: JHM-D-15-0028

Moghim S., Bowen A. J., **Sarachi S**., Wang J., Retrieval of Hourly Records of Surface Hydrometeorological Variables Using Satellite Remote Sensing Data, 2015, *J. Hydrometeorol*., 16, 147-157, doi: 10.1175/JHM-D-13-0127.1

Conference Presentations

Sarachi, S., Hsu, K., Sorooshian, S., Hydrologic evaluation of a Generalized Statistical Uncertainty Model for Satellite Precipitation Products, 2014 American Geophysical Union Fall Meeting, December 14-19, 2014 San Francisco

Sarachi, S., Hsu, K., Sorooshian S., A Generalized Statistical Uncertainty Model for Satellite Precipitation Products , 2013 American Geophysical Union Fall Meeting, December 4-9, 2013 San Francisco

Sarachi, S., Moghim S., Famiglietti, J., Decreasing trend of groundwater in Turkey , 2010 American Geophysical Union Fall Meeting, December 14-19, 2010 San Francisco

Moghim, S., **Sarachi, S**., Wang, J., Bras, R.., Retrieval of Hourly records of Surface Hydrometeorological Variables using Satellite Remote Sensing data 2010 American Geophysical Union Fall Meeting, December 14-19, 2010 San Francisco

PROFESSIONAL AFFILIATIONS

American Geophysical Union (AGU)
International Association of Hydraulic Engineering and Research (IAHR)
American Society of Civil Engineers (ASCE)
Environmental & Water Resources Institute (EWRI)
International Association of Hydrological Sciences (IAHS)
American Meteorological Society (AMS)

TECHNICAL KNOWLEDGE

Engineering Software: AutoCAD, Civil3D, ArcMap, Arc View GIS, HEC-GeoHMS, HEC-GeoRAS, HEC-HMS, HEC-RAS, HEC-6, HEC-1, Gridded Surface Subsurface Hydrologic Analysis(GSSHA), TOPMODEL, HBV, WMS (Watershed Modeling System), GMS, Modflow, ASM, WRF, ELCOM (3Dhydrodynamic and thermodynamic Estuary and Lake Computer Model)

Project Management Software: MS project, Primavera

Operating Systems: Microsoft Windows Vista/XP/2000, Linux

Programming: FORTRAN, UNIX shell scripting, MATLAB and PYTHON

ABSTRACT OF THE DISSERTATION

A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products

by

Sepideh Sarachi

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2015

Earth observing satellites provide a method to measure precipitation from space with good spatial and temporal coverage. These estimates have a high degree of uncertainty associated with them. Understanding and quantifying the uncertainty of the satellite estimates can be very beneficial when using these precipitation products in hydrological applications given that these uncertainties will propagate throughout the hydrologic cycle.

In this study a generalized uncertainty distribution is introduced to model the probability distribution of the Stage IV Multi-sensor Precipitation Estimates (MPE) as the reference measurement given the PERSIANN satellite-based precipitation product (Precipitation estimation from remotely sensed information using Artificial Neural Network). The model is calibrated for an area of $5° \times 5°$, over the southeastern United States for both summer and winter seasons separately from 2004-2009. The uncertainty model parameters are further extended across various rainfall rates and spatial and temporal resolutions.

The method is evaluated for the period of 2006-2008 over the Illinois River watershed south of Siloam Springs, Arkansas. Results show that, using the proposed method, the estimation of the precipitation is improved in terms of percent bias and root mean squared error.

To further study the hydrological response of the satellite precipitation uncertainty; this uncertainty model is propagated as an input into the SAC-SMA (Sacramento Soil Moisture Accounting) hydrology model over the same case study watershed. The results shows that the proposed uncertainty model improves the simulated streamflow from the PERSIANN satellite precipitation product with regards to its, percent bias by more than 90 % and the root mean squared error by more than 30 %.

The uncertainty model is also applied to for the new GPM satellite precipitation product using the IMERG algorithm and results shows improvement in estimating the precipitation.

Chapter 1: Introduction

1.1 Satellite-precipitation products and their application

Precipitation is one of the most important components of the water budget and plays a key role in connecting water and energy cycles. Too much or too little precipitation can lead to potential disasters, such as floods and droughts, and result in thousands of dollars of damage to people's lives and properties. We can observe this in recent events around the world, e.g., the California drought or the Hagupit super-typhoon over the Philippines. Therefore, the task of providing reliable measurements of precipitation is crucial for a safer environment.

For precipitation measurement, rainfall gauges are the most accurate instruments at point scale, but the lack of a dense network of gauges, especially in remote areas, is a challenge. Radars provide more coverage from the gauges, but radars cannot detect the precipitation that occurs over the mountainous regions or oceans.

Recently, satellite instruments have become feasible for precipitation estimation at fine spatial and temporal scales. Unlike gauges and radars, satellite measurements can overcome limitations from ground sensors in terms of coverage and operation.

A number of high-resolution, satellite-based precipitation estimates (HRSPE) are available in near real-time (e.g., Hsu *et al*., 1997; Huffman *et al*., 2001; Sorooshian *et al*., 2000; Scofield *et al*., 2003; Xie *et al*., 2003; Joyce *et al*., 2004; Okotoma *et al*., 2005; Huffman *et al*., 2007). These satellite-precipitation estimates are used by hydrologists, meteorologists, and decision- makers to study floods, droughts, and climate change and to quantify how much water is available for communities to use.

1.2 Scientific study and problem statement

There is uncertainty associated with satellite-precipitation estimates that should be evaluated. Because satellites do not provide full continuous images at all times, the averaging of limited image samples in time also contributes to the error of the final precipitation product. In addition, the uncertainty of data is dependent on the spatial scale and time accumulation of the estimates. In general, the products in finer spatial and temporal scales are associated with higher uncertainty than products in coarser spatial and temporal scales (Steiner, 1996). There is also uncertainty associated with the satellite-precipitation retrieval algorithm and its calibration. These precipitation products should be studied to obtain a better understanding of how these uncertainties can affect the final product.

A number of studies related to precipitation error analysis were proposed (i.e., Krajewski, 2000; Steiner, 1996 and 1999; Li *et al*., 1998; Huffman, 1997; Hossain and Anagnostou, 2004; Hong, 2006 and 2009; Villarini, 2007; Gebregiorgis, 2011; Aghakouchak, 2011 and 2012; Maggioni, 2014).). An extensive discussion of the previous studies is provided in the next section (literature review).

Further understanding of the uncertainty of satellite estimates is important to hydrologists and operational meteorologists who use satellite-precipitation products for their water-resource management, as well as for applications such as rainfall-runoff modeling for their river-flow forecasting.

1.3 Literature review

Assuming that the true measurement ($R_{AT,}$ rainfall for temporal accumulation of T and spatial resolution of A) is not available, a reference data (R_{AT}^{ref}) is frequently used for the evaluation of model- or satellite-based estimates. A reference error can be assigned as:

$$(\hat{R}_{AT} - R_{AT}^{ref}) = (\hat{R}_{AT} - R_{AT}) - \left(R_{AT}^{ref} - R_{AT}\right) = (\varepsilon - \varepsilon^{ref}) \quad (1.1)$$

in which R_{AT}^{ref} is the available reference, such as radar data, and $\hat{R}_{AT}$ is the satellite estimate. ε^{ref} is the error of the reference, and ε is the error of the product that is not known because the true value is never available. The variance of estimate error (to the reference data) can be presented as:

$$Var[(\hat{R}_{AT} - R_{AT}^{ref})] = Var[(\hat{R}_{AT} - R_{AT})] + Var[\left(R_{AT}^{ref} - R_{AT}\right)] - 2Cov[(\hat{R}_{AT} - R_{AT}), ((R_{AT}^{ref} - R_{AT})] = Var[\varepsilon] + Var[\varepsilon^{ref}] - 2Cov[\varepsilon, \varepsilon^{ref}] \quad (1.2)$$

If the two (reference data and satellite estimates) are uncorrelated, the covariance of their errors is zero. Eq. (1.2) can be presented as:

$$Var[(\hat{R}_{AT} - R_{AT}^{ref})] = Var[\varepsilon] + Var[\varepsilon^{ref}] \quad (1.3)$$

Ciach and Krajewski (1999) and Anagnostou *et al*. (1999) attempted to separate the error into algorithm error and sampling error. Sampling error is the error associated with the infrequent passing of satellites over a region that causes the product to be a temporal average that is slightly different from its true value. Laughlin (1981) showed that the sampling error is a mean-squared error of the mean value of the precipitation:

$$\sigma_A^2 = Var\left(R_A(t)\right) \quad (1.4)$$

where $R_A(t)$ is the variance of the area-averaged rainfall rate. Laughlin studied the GATE (Global atmospheric research program Atlantic Tropical Experiment) data set over the Atlantic Ocean for

the summer of 1974. The sampling errors over the ocean, where there is the least amount of in-situ observation (no gauges and radar data are available) as a reference for satellite estimates, were estimated. Similar case studies have also been done by other researchers (Seed and Austin, 1990; Somen *et al*., 1995; Somen *et al*., 1996; Oki and Sumi 1994; Weng *et al*., 1994).

Bell and Kundu (2000) showed that the sampling error is a function of the mean monthly precipitation and also depends on the sampling space and the number of satellite visits over one month. They proposed the following equation to compute the sampling error:

$$\frac{\sigma_{samp}}{R} \approx c(\tau)\left(\frac{r}{R}\frac{a}{A}\frac{1}{S}\right)^{\frac{1}{2}} \quad where \quad c(\tau) = \left(1 - \frac{2\tau}{T/S}\right) \qquad (1.5)$$

where σ_{samp} is the sampling error, R is the mean rainfall rate over area A, r is the mean rainfall rate during an event over area a, S is the number of satellite visits,τ is the correlation time of the rainfall events, and T is the length of the sample time . Additionally, there are some studies, such as Kunsch (1989), that used moving-block bootstrapping, which is a non-parametric method based on the sampling experiment. Steiner and Bell (2003) came up with the same relationship as in Bell and Kundu (2000) for the data over the Rocky Mountains by using the Laughlin (1981) formula and the resampling method. Steiner and Bell concluded that this uncertainty is a statistical variable and should be defined in probabilistic terms.

Gebremichael and Krajewski (2004) compared parametric and non-parametric error estimation. They used Laughlin's (1981) formula for the parametric approach and moving-block bootstrapping methods for the nonparametric approach. Gebremichael and Krajewski (2005) further defined satellite-precipitation sampling error as asymmetric distribution, such as shifted Gamma and shifted Weibull. They found out that, for large sampling intervals such as 12 or 24 hours, the conditional distribution of error to rainfall rate is shifted Weibull; for smaller sampling intervals, such as 3 or 6 hour, the logistic distribution works better.

Steiner *et al*. (2003) proposed a relationship between radar rainfall estimates and several other factors; a similar relationship was used by Hong *et al*. (2006) to quantify the variance of the measurement error as a function of area coverage, time integration, sampling frequency, and space-time average rainfall rate. Several studies on the impact of precipitation uncertainty on flood prediction have been performed (Hossain and Anagnostou, 2004). In those studies, the effect of passive microwave and infrared-based satellite-product error on flood prediction using a probabilistic error model was demonstrated. In AghaKouchak *et al*. (2012), satellite-precipitation error is divided into systematic and random error, and their correlation to space and time accumulation is presented.

There is significant interest in the evaluation of available satellite-precipitation products. In a study by Maggioni *et al*. (2014), the joint model of satellite vs. reference precipitation is divided into four regions of hit (where both reference and satellite show precipitation), miss (where reference shows precipitation but satellite shows zero), false alarm (where satellite shows rainfall but reference shows zero), and correct no-precipitation (where both show zero). They modeled hit and missed precipitation using a Gamma function and used a constant probability for correct-zero precipitation and false alarms.

In most of the previous studies, the errors associated with satellite estimates are assumed to be Gaussian, where error variance is estimated (e.g., Ciach and Krajewski, 1999; Anagnostou *et al*., 1999). More recent studies (e.g., Gebremichael and Krajewski, 2005) have demonstrated that error distribution is significantly different from Gaussian distribution. Those studies showed that error distribution is relevant to the spatial and temporal resolution of estimates, and shifted Gamma, shifted Weibull, and shifted lognormal, logistic, and normal distribution are fitted to the error of estimate at various spatial and temporal scales.

In this study, a generalized distribution function is introduced to fit the probability distribution of reference data given the satellite-based estimates. The proposed model consists of four different parts: hit precipitation, false alarm, missed precipitation, and hit-zero precipitation. It is assumed that the different parameters of the precipitation product uncertainty model can vary with rainfall rate and product resolution in space and time, which are further fitted by simple mathematical functions.

1.4 Research motivation

Access to a global-precipitation product with high spatial and temporal resolution is a necessity for climate and hydrological studies, which are now made possible using the satellite-precipitation products. To be able to rely on these precipitation estimates, the uncertainty associated with them should be quantified and used in water-resources management. The answer to this need is a generalized-uncertainty model that is able to quantify different elements of the uncertainty with a good level of confidence. As a general approach, the model should be able to perform for the satellite-precipitation products at different spatial and temporal scales.

1.5 Objective of the study

In this study, a generalized-distribution function is introduced to fit the probability distribution of reference data given satellite-based estimates. This probability distribution will quantify the uncertainty of satellite-precipitation products for a given rainfall rate. This probability distribution is divided into 4 regions of hit precipitation, false alarm, missed precipitation, and hit-zero precipitation. For the hit precipitation, where both satellite and radar show non-zero precipitation, the proposed function is in the shape of a generalized normal distribution type II.

This distribution is flexible enough to provide non-symmetric probability functions considering bias and variance, as well as higher-order moments of uncertainty (skewness) beyond the first (mean) and the 2^{nd} (standard variation) moments, because the precipitation data are skewed towards smaller values and are not normally distributed.

The parameters of this generalized-statistical distribution are estimated as functions of spatial and temporal resolution and rainfall rate, which enables the user to model asymmetrical uncertainty distributions without having to choose from different distribution forms for different product resolutions. It is assumed that the distribution parameters of the precipitation-product uncertainty can vary through rainfall rate and product resolution in space and time. The generation of the parameter space is achieved by estimating the distribution parameters by aggregating the product at various spatial and temporal scales. For instance, the chosen data product (e.g., PERSIANN estimates) is processed from $0.25\,°$ and 3-hourly resolutions to coarser spatial resolutions, such as $0.5\,°$, $0.75\,°$, and $1\,°$, and temporal resolutions, such as 3-hourly, 6-hourly, 12-hourly, and daily resolutions.

For the false alarm percentage, where the satellite shows precipitation but radar shows zero, the proposed model is in the form of an exponential equation. Similar to the above, the parameters are estimated at different resolutions.

For missed precipitation, where the satellite shows zero precipitation but radar shows a value, the percentage is calculated at different resolutions, and the uncertainty distribution is modeled using a fitted Gamma distribution to the reference data.

All of the models are combined for a cohesive distribution of satellite and reference precipitation.

1.6 Scope of the research

In Chapter 2, the general form of the uncertainty model is presented first and then each part of the uncertainty is introduced. Furthermore, the uncertainty modeling in different spatio-temporal resolutions is explored, and the parameter space is introduced. In Chapter 3, the data used and the study area are described. Chapter 4 shows the results of the model calibration. In Chapter 5, the results of precipitation-level uncertainty analysis are presented over the validation region. With regard to the application, Chapter 6 presents a case study for uncertainty analysis of satellite precipitation products. Chapter 7 introduces the details about the hydrologic modeling and the propagation of precipitation uncertainty in streamflow simulation. The model is further calibrated and evaluated over the new GPM precipitation product using the IMERG algorithm in chapter 8. Conclusions and recommendations for future work are given in Chapter 9.

Chapter 2: Satellite-Precipitation Uncertainty Model

The uncertainty model presented in this study is a conditional distribution of radar data as the reference given satellite precipitation. The parameters of this model should be defined and calibrated. In this chapter, the general form of the uncertainty model is defined, and each component is then described separately.

2.1 Uncertainty model

The scatter plot of the satellite precipitation and radar data is divided into 4 different components based on the ability of the satellite to capture the rainfall event and its magnitude (Figure 2.1). Before processing the data, a threshold is considered, and any precipitation smaller than that threshold is assigned to zero. The threshold should be a small value close to zero (Liu et. al. , 2015). In this study, the value of 1 mm/day is used as the threshold for the precipitation data. The threshold would remove very small precipitation values that are generated as the noise by the precipitation estimation algorithm. It also removes the precipitation data that are smaller than the accuracy of the remote sensing device. In this study, we show the satellite precipitation with R_{sat}, the reference precipitation from radar with R_{ref} and the threshold with thr.

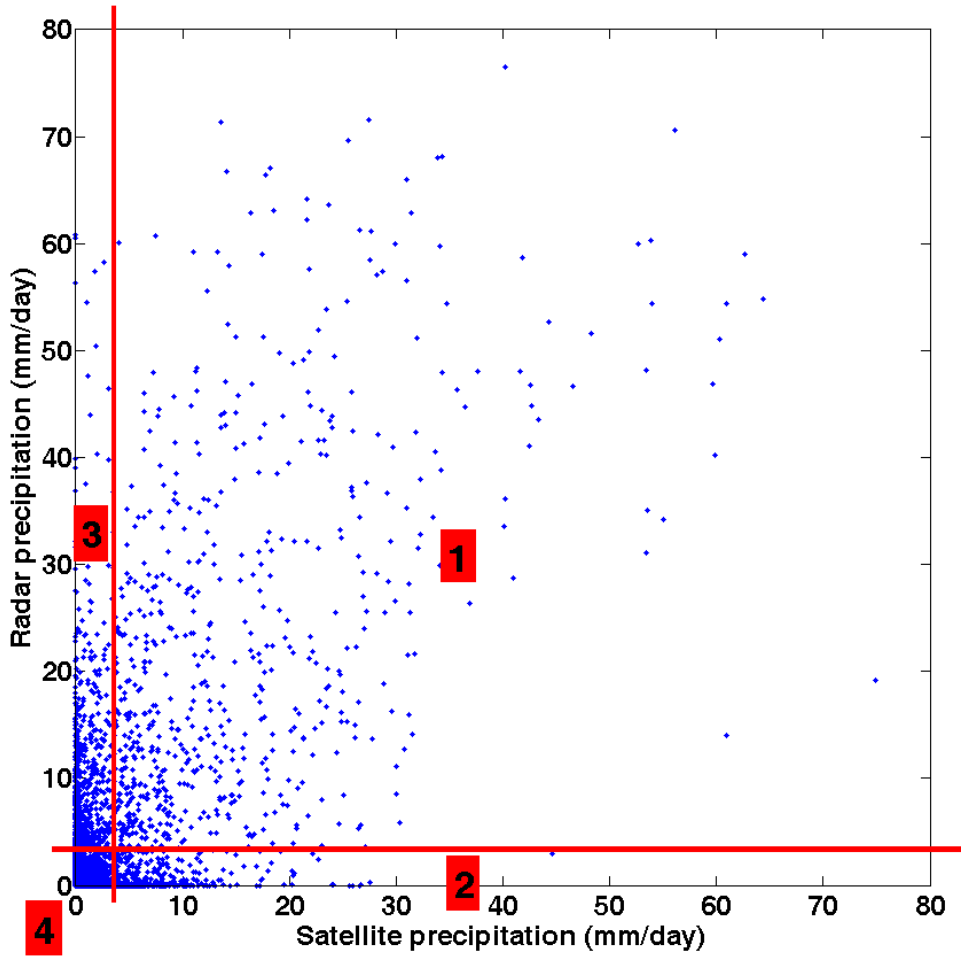

Fig. 2.1. Uncertainty decomposition of satellite and reference precipitation into 4 regions: 1: hit precipitation, 2: false alarm, 3: missed precipitation, and 4: hit-zero precipitation.

The plane of the satellite-precipitation product and the reference precipitation is divided into 4 parts. The first region ([1]), called hit precipitation, is where both satellite and reference show non-zero precipitation; this part also indicates that the satellite correctly detected the precipitation event. This biased value should be evaluated based on its corresponding reference precipitation.

The second region ([2]) of the product is false alarm, when the satellite shows precipitation, but the reference shows zero rainfall. The third region ([3]) shows missed precipitation, when

reference shows a precipitation event, but the satellite fails to detect it. The last region ([4]) is called hit or correct-zero precipitation, where both reference and satellite show zero precipitation.

In order to generate a full model to quantify precipitation uncertainty, one needs to consider these different situations.

When dealing with a time series of satellite precipitation, there are time steps in which the satellite shows precipitation, which might be hit precipitation or false alarm, and there are time steps in which the satellite shows zero precipitation, which can be either missed precipitation or correct-zero precipitation.

To assign probabilities to these different situations, 1 is chosen as rainfall and 0 is chosen as no rainfall. Here P_{10} is assigned to the conditional probability of reference showing no rainfall given satellite shows rainfall (false alarm) and P_{11} is the conditional probability of reference showing rainfall given satellite also shows rainfall (hit precipitation). By the law of total probability, $P_{10} + P_{11} = 1$.

In addition, if P_{01} is assigned to the conditional probability of reference showing rainfall given the satellite shows no rainfall (missed precipitation) and P_{00} is the conditional probability of reference showing no rainfall given satellite also detects no rainfall (hit zero-precipitation), then by the law of total probability, $P_{01} + P_{00} = 1$.

To model the complete uncertainty, each of these conditional probabilities must be defined and the densities within each region must be modeled.

2.2 Hit precipitation

The first part of the uncertainty addressed in this section is hit precipitation, which is the probability of reference data given the satellite estimate. A number of different probability distributions can be used as uncertainty model to model the conditional probability of the hit precipitation given the satellite precipitation is larger than the threshold. The most commonly used one is Gaussian distribution, which is a symmetric function with two parameters (mean and standard deviation); its skewness and kurtosis are zero. However, precipitation data are not shaped symmetrically (Gebremichael and Krajewski, 2004); rather, they are skewed with a larger occurrence of smaller rainfall rates. There are different distributions that have skewness other than zero, e.g., Gamma distribution, log-normal distribution (which has been shown to be unrealistic at high rainfall rates; Gebremichael and Krajewski, 2005), generalized-normal distribution (GND), etc. There are certain characteristics that make the generalized-normal distribution superior to the others. The GND can generate unimodal distributions, and its flexibility makes it possible to model the sharper peaks over the smaller values of rainfall rates when the data are highly skewed (Figure 2.2). The GND can also model data with smaller peaks at larger temporal and spatial accumulations. Furthermore, the distribution is bounded from the left when the shape parameter is negative, which is the case for precipitation data because they are skewed to the left. The GND can be modeled as a shifted distribution because it contains a shift parameter. This is not possible in two-parameter distributions, e.g., Gamma or Weibull. If the satellite precipitation is defined at R_{sat} the general density function of the GND model is as follows:

$$P_{GND}(x > thr \mid R_{sat} = R > thr\ ) = \frac{\Phi(y)}{\alpha - \kappa(x-\xi)} \qquad \alpha > 0 \qquad (2.1)$$

$$y = \begin{cases} -\frac{1}{\kappa} log\,[1 - \frac{\kappa(x-\xi)}{\alpha}] & \kappa \neq 0 \\ \frac{x-\xi}{\kappa} & \kappa = 0 \end{cases} \qquad (2.2)$$

$$x \in \left(\xi + \frac{\alpha}{\kappa}, +\infty\right) \; if\; \kappa < 0$$

$$x \in (-\infty, +\infty) \; if\; \kappa = 0$$

$$x \in \left(-\infty, \xi + \frac{\alpha}{\kappa}\right) \; if\; \kappa > 0$$

with α (positive and real) as scale, κ (real) as shape, and ζ (real) as the location parameters of the GND model and $\Phi(y)$ as the standard normal probability-distribution function.

The different moments of the distribution can be defined in terms of its parameters, as follows:

$$Mean = \xi - \frac{\alpha}{\kappa}(e^{\frac{\kappa^2}{2}} - 1) \qquad (2.3)$$

$$Variance = \frac{\alpha^2}{\kappa^2} e^{k^2}(e^{k^2} - 1) \qquad (2.4)$$

$$Skewness = \frac{3e^{k^2} - e^{3k^2} - 2}{\left(e^{k^2} - 1\right)^{\frac{3}{2}}} sign(\kappa) \qquad (2.5)$$

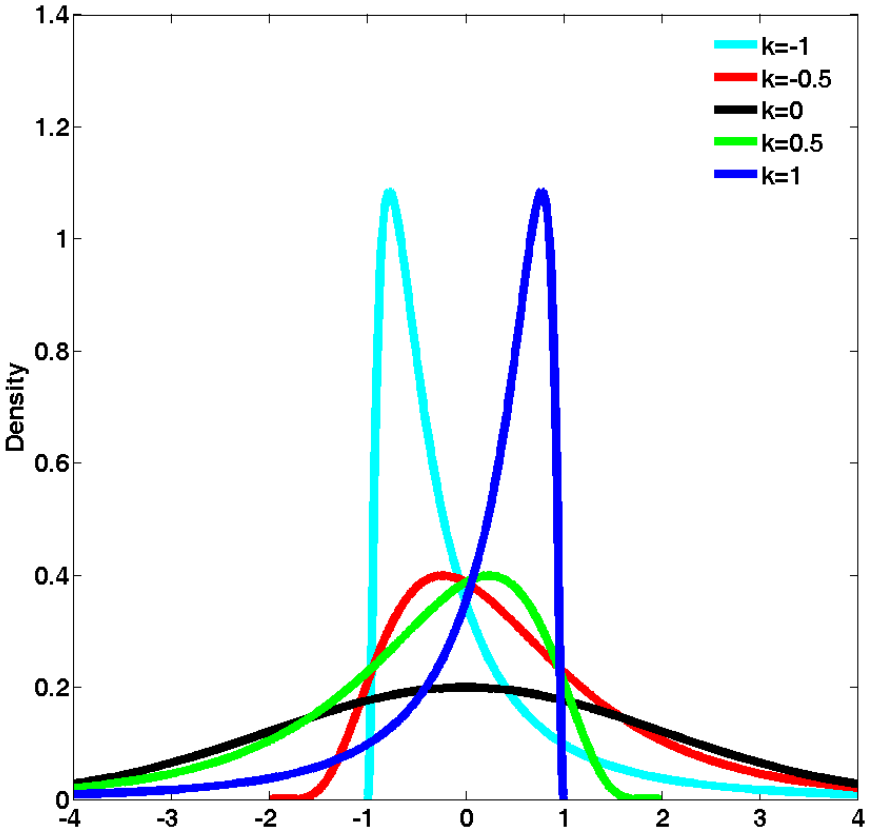

Fig. 2.2. General form of generalized-normal distribution (GND).

If at time step t, R_s is the satellite precipitation estimate and thr is the threshold for zero precipitation, GND ($x > thr | R_{sat} = R > thr$) shows the distribution of the reference precipitation given the satellite precipitation estimate equals R.

2.3 False alarm

One of the largest uncertainties of satellite-precipitation products comes in the form of false alarm, where the satellite detects false precipitation. False alarm contributes to a large amount of precipitation overestimation. For example, false alarm is detected in IR-based satellite precipitation products where the estimation is based on the cloud-top temperature. During the

winter months, when high, non-precipitating cirrus clouds are formed; the satellite often detects false precipitation. Many different studies have investigated false alarm (e.g., AghaKouchak *et al*., 2011; Tian *et al*., 2009).

To reduce this false detection, Nasrollahi (2015) attempted to improve the PERSIANN algorithm by adding better classification of clouds, which is made possible using cloud profiles when available. To model this component of uncertainty, Maggioni *et al*. (2014) used an exponential model to quantify the percentage of false alarm. An exponential decay model is proven suitable (Hossain and Anagoustou, 2006) for the false alarm probability. In this work, false alarm probability is modeled using an exponential function of the satellite rain rate. Here, we define the false alarm ratio (P_{10} = P (R_{ref} < thr | R_{sat} > thr)) as:

$$P_{10} = b \times exp\,(-r \times R_{sat}) \qquad (2.6)$$

In equation (2.6), r and b are the parameters of the exponential decay function. If the probability of false alarm (where reference detects no precipitation given the satellite shows precipitation detects precipitation larger than the threshold) is called P_{10}, and the probability of hit precipitation (where reference detects precipitation given the satellite shows precipitation) is called P_{11}, and we can conclude that:

$$P_{10} + P_{11} = 1 \qquad (2.7)$$

In the case of satellite detecting precipitation, the uncertainty distribution of the reference given the satellite precipitation is in the form of the mixture model:

$$P\left(R_{ref} < x \mid R_{sat} = R\right) = P_{10} + P_{11} \times GND_{cdf}\left(R_{ref} < x \mid R_{sat} = R > thr\right) \qquad (2.8)$$

GND_{cdf} is the CDF of the generalized normal distribution in the form of :

. $$GND_{cdf}\left(R_{ref} < x \mid R_{sat} = R > thr\right) = \Phi(y) \qquad (2.9)$$

Where Φ is the standard normal CDF and y is as described in Eq. (2.2).

In this work, the expected value of this equation is used as a representative of the uncertainty range. This expected value is calculated as in Eq. (2.10):

$$E(x \mid R_{sat} = R > trh) = P_{10} \times 0 + P_{11} \times E_{GND}(x \mid R_{sat} = R > thr) = (1 - P_{10}) \times E_{GND}(x \mid R_{sat} = R > thr) \quad (2.10)$$

E_{GND} is the expected value of the generalized normal distribution.

To make a more accurate model, the temporal correlation of satellite-precipitation products to model the probability of false alarm is investigated in this study. To do so, we try to determine if the probability of false alarm at a specific time step has any relationship to the precipitation estimate at the previous time step.

Two different states are considered in this study: (1) when the satellite is showing non-zero precipitation at time *t* and *t*- Δt (previous time step), and (2) when the satellite is showing non-zero precipitation at time *t* but zero precipitation at time *t*- Δt**.** It is expected that, when Δt is increasing, the effect of precipitation at *t*- Δt becomes less distinct, especially when there is the condition for short-term precipitation.

2.4 Missed precipitation

One of the largest uncertainties of satellite-precipitation products comes in the form of false alarm, where the satellite detects false precipitation. False alarm contributes to a large amount of precipitation overestimation. For example, false alarm is detected in IR-based satellite precipitation products where the estimation is based on the cloud-top temperature. During the winter months, when high, non-precipitating cirrus clouds are formed; the satellite often detects

false precipitation. Many different studies have investigated false alarm (e.g., AghaKouchak *et al*., 2011; Tian *et al*., 2009).

To reduce this false detection, Nasrollahi (2015) attempted to improve the PERSIANN algorithm by adding better classification of clouds, which is made possible using cloud profiles when available. To model this component of uncertainty, Maggioni *et al*. (2014) used an exponential model to quantify the percentage of false alarm. An exponential decay model is proven to be suitable (Hossain and Anagoustou, 2006) for the false alarm probability. In this work, false alarm probability is modeled using an exponential function of the satellite rain rate. We define the false alarm ratio (P_{10} = P (R_{ref} < thr | R_{sat} > thr)) as:

$$P_{10} = b \times exp\,(-r \times R_{sat}) \qquad (2.6)$$

In equation (2.6), r and b are the parameters of the exponential decay function. If the probability of false alarm (where reference detects no precipitation given the satellite shows precipitation detects precipitation larger than the threshold) is called P_{10}, and the probability of hit precipitation (where reference detects precipitation given the satellite shows precipitation) is called P_{11}, and we can conclude that:

$$P_{10} + P_{11} = 1 \qquad (2.7)$$

In the case of satellite detecting precipitation, the uncertainty distribution of the reference given the satellite precipitation is in the form of the mixture model:

$$P(R_{ref} < x \mid R_{sat} = R) = P_{10} + P_{11} \times GND_{cdf}\left(R_{ref} < x \mid R_{sat} = R > thr\right) \qquad (2.8)$$

GND_{cdf} is the CDF of the generalized normal distribution in the form of :

. $$GND_{cdf}(R_{ref} < x \mid R_{sat} = R > thr) = \Phi(y) \qquad (2.9)$$

Where Φ is the standard normal CDF and y is as described in Eq. (2.2).

In this work, the expected value of this equation is used as a representative of the uncertainty range. This expected value is calculated as in Eq. (2.10):

$$E(x \mid R_{sat} = R > trh) = P_{10} \times 0 + P_{11} \times E_{GND}(x \mid R_{sat} = R > thr) = (1 - P_{10}) \times E_{GND}(x \mid R_{sat} = R > thr) \quad (2.10)$$

E_{GND} is the expected value of the generalized normal distribution.

To make a more accurate model, the temporal correlation of satellite-precipitation products to model the probability of false alarm is investigated in this study. To do so, we try to determine if the probability of false alarm at a specific time step has any relationship to the precipitation estimate at the previous time step.

Two different states are considered in this study: (1) when the satellite is showing non-zero precipitation at time *t* and *t*- Δt (previous time step), and (2) when the satellite is showing non-zero precipitation at time *t* but zero precipitation at time *t*- Δt**.** It is expected that, when Δt is increasing, the effect of precipitation at *t*- Δt becomes less distinct, especially when there is the condition for short-term precipitation.

2.6 The complete uncertainty model

For a complete model, all of the different components must be combined. Given a time series of satellite precipitation, if the satellite detects a rainfall rate larger than the threshold at each time step, there is the possibility of hit precipitation or false alarm at that time step. The uncertainty of that point is in the form of Eq. (2.8). If the satellite detects zero precipitation, at that time step, there exists the possibility of missed precipitation or hit-zero precipitation. From Eq. (2.15), the uncertainty distribution of that detection can be obtained.

2.7 Parameter space at various spatiotemporal resolutions

After introducing the general form of the model, the goal is to calibrate the model so that it provides the ability to capture the uncertainty of satellite-precipitation products at different spatial and temporal resolutions. The resolution of the satellite-precipitation data directly affects their uncertainty. When data are provided in high resolution, there is larger uncertainty associated with the estimates, and their distribution is highly skewed. On the other hand, when the observations are estimated at a coarser resolution, the associated uncertainties are less skewed, and their distribution is closer to a random process represented by a normal distribution.

To study this uncertainty at different resolutions, both the satellite and the reference data are aggregated into different spatial and temporal resolutions, and the uncertainty model is calibrated for each of them separately. The result is used to generate a parameter space for the parameters of the uncertainty distribution at each of those different resolutions.

The ultimate goal is that, at any spatial and temporal resolution, the parameters of the uncertainty model can be quantified based on those resolutions, and the uncertainty model is generated.

The temporal resolutions used in this study are 3-hourly, 6-hourly, 12-hourly, and daily, and the spatial resolutions are 0.25 °, 0.5 °, 0.75 °, and 1 °. The model is calibrated for 16 different pairs of temporal and spatial resolutions. For cases where the desired resolution falls between these resolutions, a linear interpolation to find the model parameters is used. It is recommended that, for all of the other resolutions, extrapolation of the parameters be done with care.

Chapter 3: Study Area and Data

Information on the study area and the data products that are used in this study are provided in this chapter. First, the satellite-precipitation product is introduced, followed by describing the radar product and the study domain.

3.1 Satellite-precipitation data

In this study, PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) is used at the satellite precipitation estimation. PERSIANN uses an artificial neural-network algorithm to estimate the rainfall rate from the infrared images of the Geostationary Environmental Satellite (GOES) (Hsu et al., 1997, Figure 3.1). The reliable but less-frequent instantaneous precipitation rates from microwaves sensors are used in an algorithm that estimates rainfall rates from the IR GEO satellite images. The input to this artificial-neural network is the set of features that are extracted from the long-wave IR images (10.2-11.2 μm). The process to obtain the rainfall rate is as follows: (1) calculate the mean and variance of the brightness temperature of the pixels of the IR image as the features of the image, (2) A neural network–based classification scheme (self-organizing feature map algorithm) is used to classify these features into different groups associated with different cloud surface characteristics and (3) Using a linear multivariate function, those classes are related to the surface rainfall rate using the available rainfall. The parameters of this neural network are updated whenever passive microwave-based rainfall is available. These data are produced in 0.25° resolution every 30 minutes and are then integrated to different spatial and temporal resolutions. The PERSIANN algorithm is running in two modes. The first one is rainfall estimation using only the GEO

infrared images every 30 minutes, and the second mode is the updated version using the PMW instantaneous rainfall rates whenever they are available. The IR images are from GOES-8, GOES-10, GMS-5, and MeteoSat-6&7, and the PMW rainfall is from the information provided by TRMM, NOAA-15, -16, -17, DMSP F-13, F-14, and F-15. There are multiple factors upon which the accuracy of the PERSIANN product is dependent. Among those factors are the error and frequency of the PMW data to update, the error of the input and output functions, or the feature detection and classification algorithm.

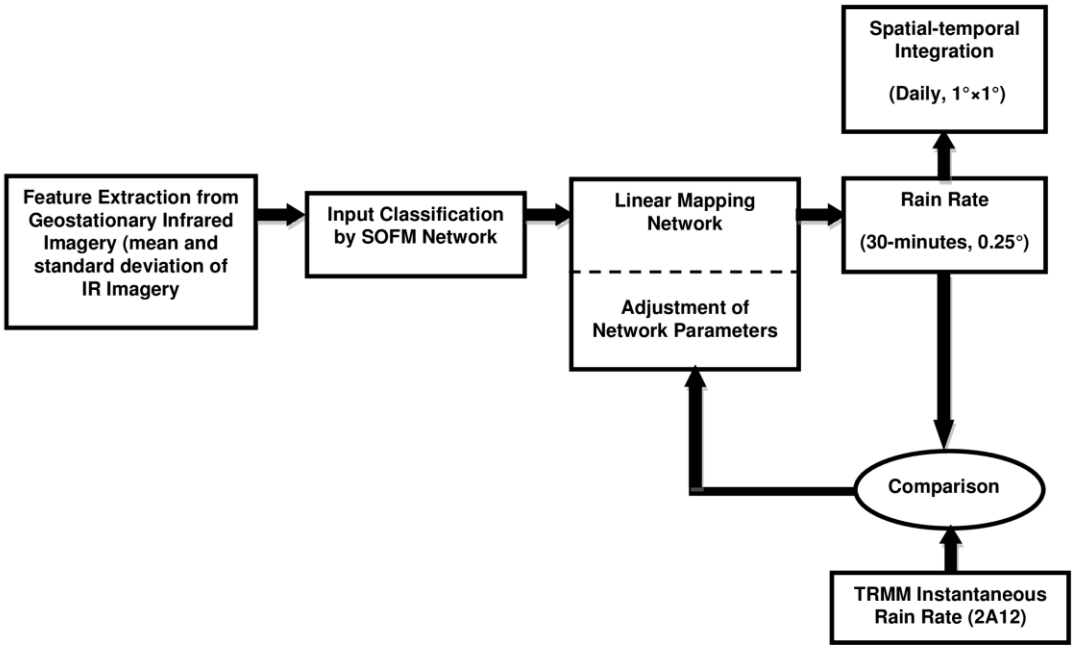

Fig. 3.1. Rainfall estimation from the PERSIANN system using GEO and LEO satellite information (Sorooshian et. al, 2000).

3.2 Reference-precipitation data

The Next-Generation radar (NEXRAD), a network of 178 Doppler weather radars (WRS-88D) operated by the National Weather Service (NWS) across the United States, measures precipitation at a high-spatial resolution. To measure precipitation, a relationship between reflectivity and the rainfall rate (Z-R relationship) is used, which is calibrated differently for different types of precipitation (Rinehart, 2004). To create the NEXRAD stage-IV data, the gauge observations and their conditioning are also added to the data using the Multisensor Precipitation Estimation (MPE), where the values of the rainfall gauges are taken from the weather service stations in the NWS Hydrometeorological Automated Data System network. These data are in a grid format of 4-km resolution pixels, and they are most accurate after 2002 (Westcott *et al.,* 2008). These radar-rainfall data are adjusted for various biases using rainfall-gauge measurements *(*Lin and Mitchell, 2005) and can be considered as the best radar-rainfall data available (AghaKouchak *et al*., 2010*;* AghaKouchak *et al*., 2010 c,d), but these data also have their own uncertainty (Smalley *et al., 2014).* However, in this study, the product is considered as the reference and thus is assumed perfect.

The radar data are re-gridded into 0.25 ° and averaged into 3-hourly resolutions to be comparable to the PERSIANN data.

3.3 Study area

The study area to calibrate the precipitation uncertainty model is a $5° \times 5°$ region over the southeast United States bounded between 30 ºN-35 ºN and 85 ºW-90 ºW, as shown in Figure 3.2. Data include the months of June, July, and August for the summer season and December, January, and February for the winter season for a 5-year period from 2005-2009.

Both satellite and radar data are available in binary format and are converted to text format. All of the data from both satellite and radar that are smaller than the threshold of 1 mm/day are assigned to zero rainfall pixels, which construct the false alarm, missed precipitation, and hit zero precipitation parts.

Both data sets are averaged temporally into 3-, 6-, 12-, and 24-hourly resolutions and spatially into 0.25 ° × 0.25 °, 0.5 ° × 0.5 °, 0.75 ° × 0.75 °, and 1 ° × 1 ° resolutions to construct the parameter space.

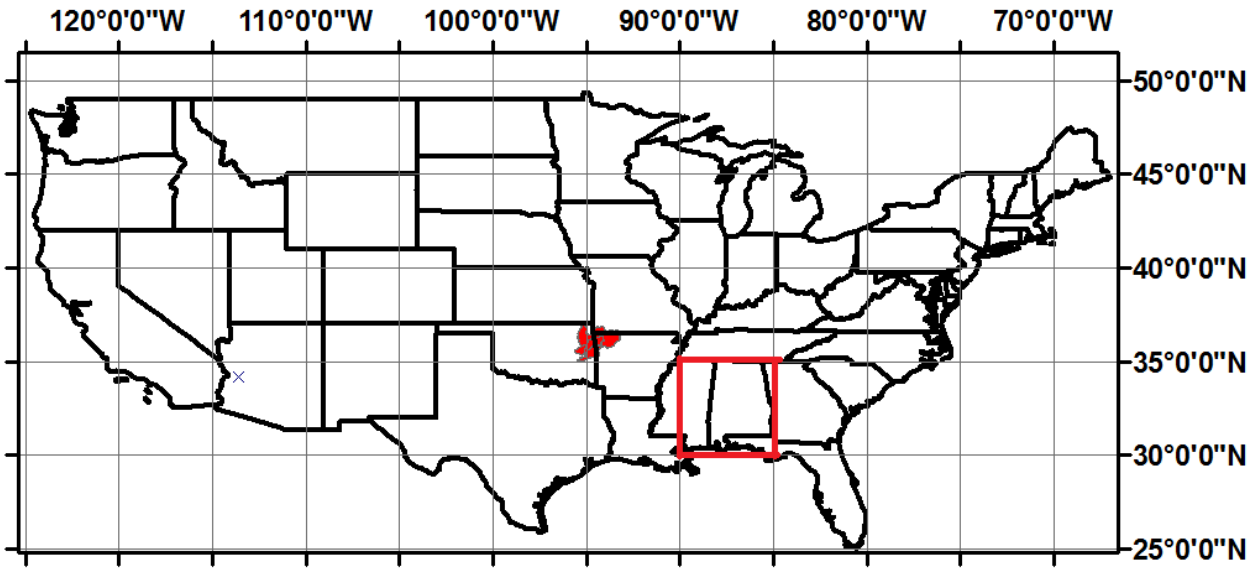

Fig. 3.2. Model-calibration domain (red box) and DMIP2 study watershed's location (red area).

3.4 Seasonality of the uncertainty

The study region has a humid subtropical climate, and precipitation patterns vary greatly between the summer and winter seasons. During the summer, precipitation is almost entirely convective and caused by mesoscale complexes and thunderstorms. Wintertime precipitation is mostly stratiform and tied to synoptic-scale systems. Days with measurable snowfall are very rare in the region, but the warm moist air coming from the Gulf of Mexico during the winter could cause frontal freezing rainfall, resulting in ice cover, which usually lasts for several days. These different precipitation patterns result in the difference in the satellite-precipitation data for the two seasons.

Investigating the precipitation products shows that these data are statistically different in the winter vs. the summer (Anagnostou et al., 2010). During the summer, data are more scattered, and satellite images show a wider range of values, whereas in the winter, satellite images show a narrower range of change in rainfall values, as seen in Figure 3.3.

The seasonal differences are also apparent in the three first moments of the data (expected value, standard deviation, and skewness).

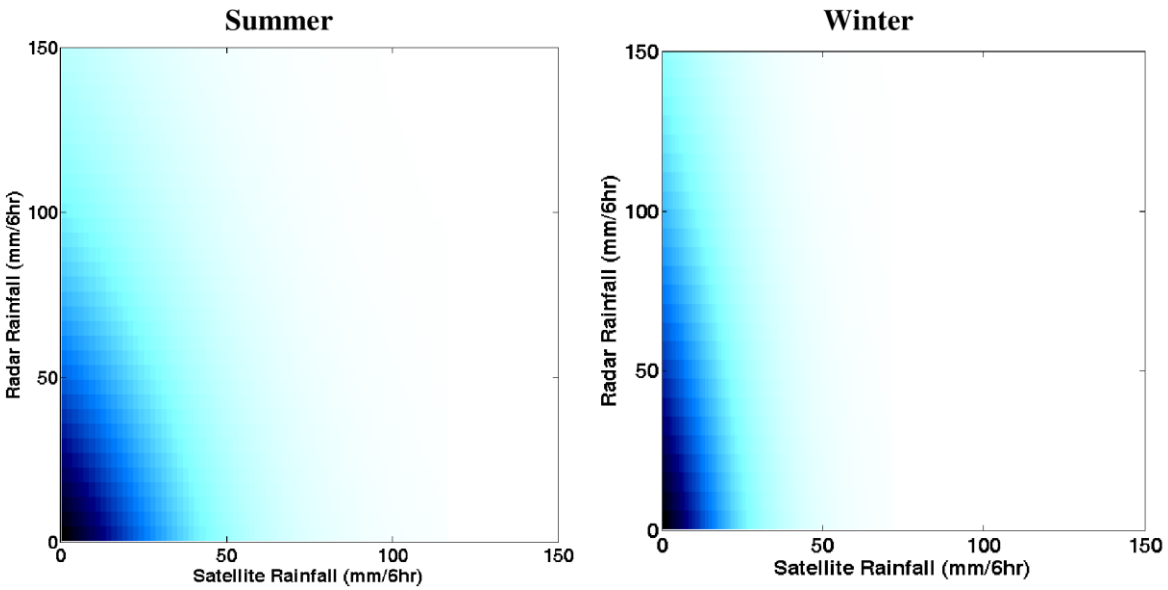

Fig. 3.3. Seasonal differences in distribution of satellite-precipitation data with respect to the corresponding radar data for summer (left panel) and winter (right panel) from the calibration domain with temporal resolution of 6-hourly and spatial resolution of 0.25 °.

From the graphs of the mean of the data (Figure 3.4, left column), we can see that, for winter, the mean is closer to a 1-1 linear function and, for summer, it looks like a power function, and the satellite data are overestimating the amount of precipitation. This shows that treating summer and winter data separately would result in a more accurate estimation of uncertainty.

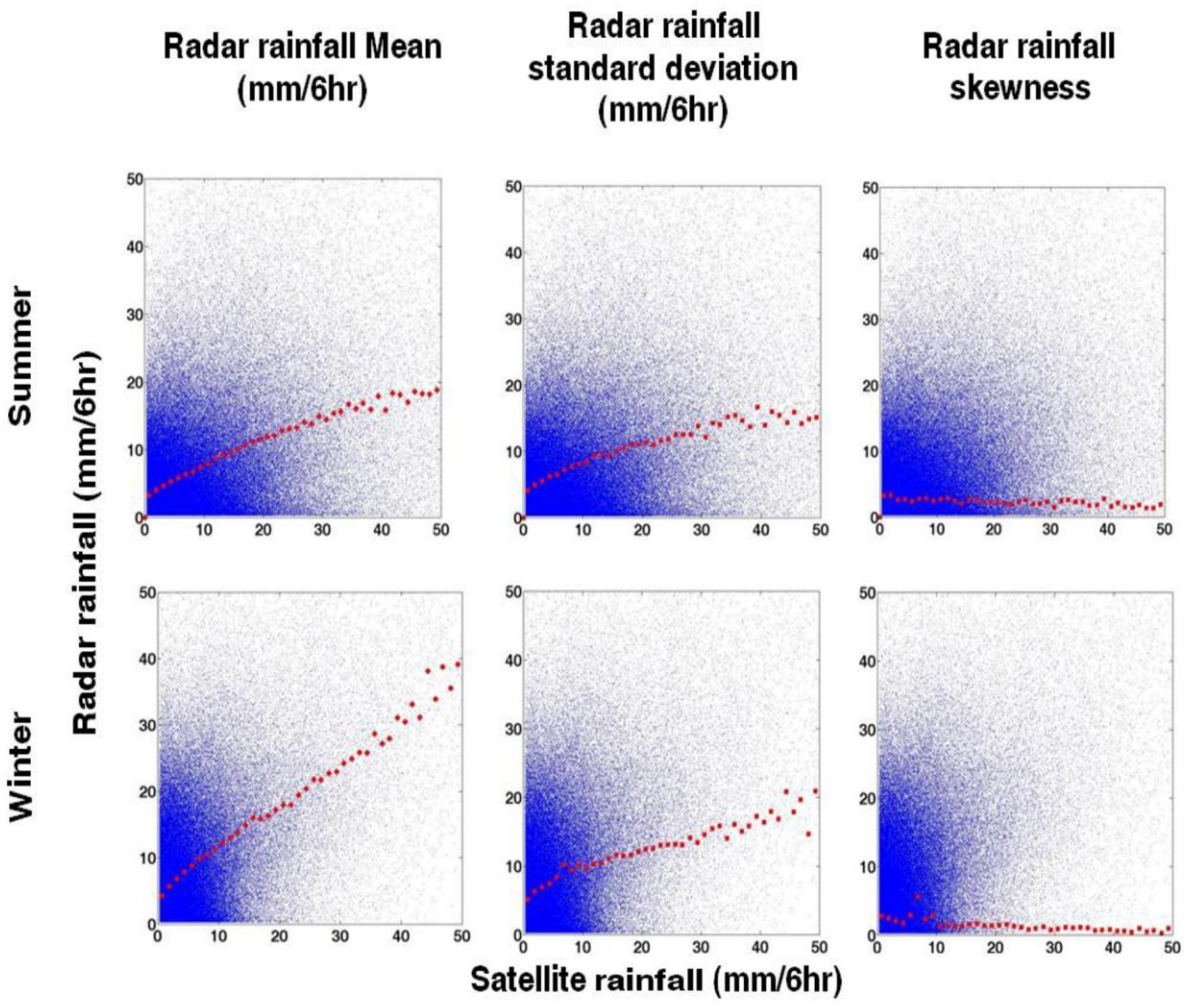

FIG. 3.4. Comparison of the three moments of radar data vs. the mean satellite rainfall rate temporal resolution of 6 hourly and spatial resolution of 0.25 ᵒ. The scatter plot in blue is satellite rainfall rate (x-axis) vs. radar rain rate (y-axis) for June, July, and August 2005-2009 in the upper panel and December, January, and February for 2005-2009 in the lower panel. The dotted lines in red from the left column to the right are, consecutively, the mean (mm/6hour), standard deviation (mm/6hour), and skewness (no unit) of radar rainfall rate (y-axis) vs. the mean of the satellite precipitation bins (x-axis).

Chapter 4: Calibration of the Model

Each part of the uncertainty model should be calibrated separately. For calibration, the data are grouped in 16 pairs of spatial and temporal resolutions, and the results of the parameter calibrations are presented in the form of a function of the resolutions.

We begin by calibrating the parameters of the GND model for modeling the reference rainfall given satellite is larger than the threshold (hit precipitation, region [1] in figure 2.1) and find its parameters as a function of rainfall rate and spatial and temporal resolution. The results are presented in the form of a parameter space. Then, moving on to false alarm, the probability of false alarm as a function of rainfall rate for different resolutions is calibrated. Finally, for missed precipitation, both the probability and the rainfall rate in the form of a Gamma function are calibrated. These processes are performed on both summer and winter data separately.

4.1 Calibration of hit precipitation

4.1.1 Fitting the GND model to the data at a specific spatial and temporal resolution

After choosing a general-distribution model to fit to the data, the maximum-likelihood method is often the logical choice for the estimation of the distribution parameters. However, this may not be the case for distributions with a threshold that is a function of the parameters. In that case, the likelihood function may have multiple modes or reach an infinite value when the estimated parameter values are no longer suitable for data. In this work, the distribution is being fitted to the data using least-squares estimates of the cumulative-distribution functions. The idea is that the scatterplot of the empirical cumulative distribution function (CDF) of the data and the CDF of the fitted distribution fall along the 1:1 line from zero to one.

The CDF of the GND is in the form of:

$$P_{GND}\left(R_{ref} < x \middle| R_{sat} = R\right) = \Phi\ (y) \tag{4.1}$$

$$y = \begin{cases} -\frac{1}{\kappa} log\ [1 - \frac{\kappa(x-\xi)}{\alpha}] & \kappa \neq 0 \\ \frac{x-\xi}{\kappa} & \kappa = 0 \end{cases} \tag{4.2}$$

where Φ(y) is the standard normal CDF defined as:

$$\Phi\ (y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \tag{4.3}$$

The empirical CDF of the data is in the form of:

$$CDF_{emp}\left(R_{ref} < x \middle| R_{sat} = R\right) = \frac{1}{n} \sum_{i=1}^{n} 1\{x_i \leq x\} \tag{4.4}$$

To obtain this fitted distribution, the objective function of the sum of the squared differences between those two CDFs must be minimized. To compensate for the variance of the fitted functions, higher weights are given to the tails and lower weights to the center. The objective function is defined in Eq.(4.5):

$$F_{obj} = \sum_{i=1}^{n} (w_i (P_{GND}\left(R_{ref} < x_i \middle| R_{sat} = R\right) - CDF_{emp}\left(R_{ref} < x_i \middle| R_{sat} = R\right))^2 \tag{4.5}$$

$$w_i = \frac{1}{\sqrt{CDF_{emp}(x_i) \times (1 - CDF_{emp}(x_i))}} \tag{4.6}$$

To find the parameters for each specific pair of spatial and temporal resolutions, the data are divided into different bins (with the same bin size) with respect to their corresponding satellite-estimated rainfall rates, and a distribution is fitted to each bin of data. For example, if the resolution of interest is $1° \times 1°$ and 3-hourly accumulated rainfall and assuming the range of satellite estimates from 0-50 mm/day, data can be divided into 10 groups with an interval of 5 mm/day. As an example, we assign all of the satellite precipitation between 5 mm/day-10 mm/day to bin (2). Then, using Eq. (4.5), a GND is fitted to the radar rainfall corresponding to

the satellite precipitation of the bin (2) ($P_{GND}\left(R_{ref} < x \middle| 5\frac{mm}{day} < R_{sat} < 10\frac{mm}{day}\right)$). Same process is applied to the radar rainfall of all the other bins (Figure 4.1).

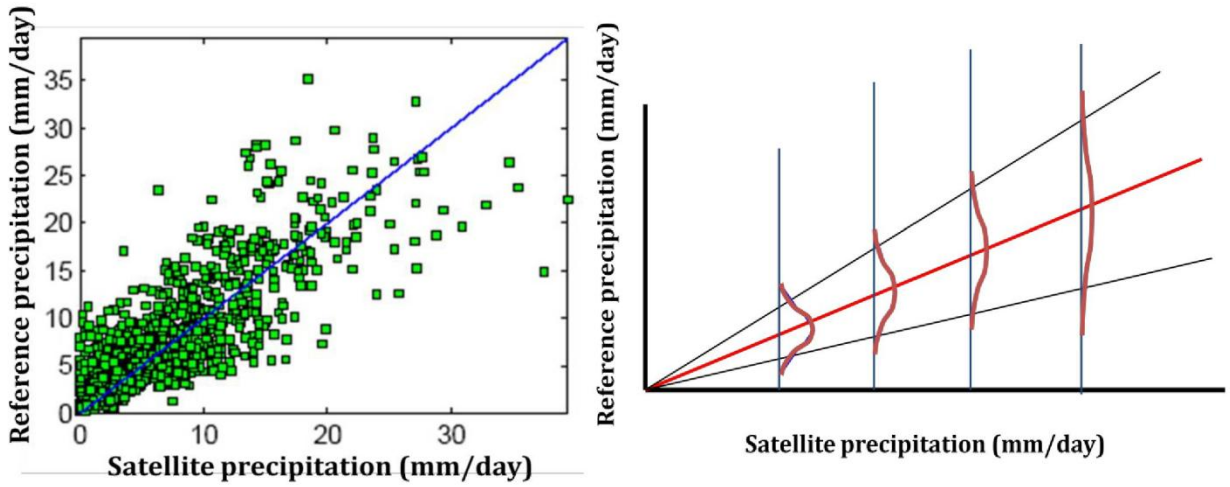

Fig. 4.1. Left panel: Daily satellite estimate vs. daily radar-rainfall observation. Right panel: Fitted distribution for a specific spatial and temporal distribution over different bins of satellite-estimated rainfall rates (schematic view).

4.1.2 Distribution Parameters at Various Spatiotemporal Scales

The three estimated parameters from each of the bins are modeled as a function of the satellite rainfall rate (mean bin value) using simple functions. For the shift and shape parameters, a linear function is used; for the scale parameter, a power function is used.

$$\xi(\Delta t, \Delta s) = a_{\Delta t,\Delta s} \times \bar{R}_{sat} + b_{\Delta t,\Delta s} \quad (4.7)$$

$$\kappa(\Delta t, \Delta s) = c_{\Delta t,\Delta s} \times \bar{R}_{sat} + d_{\Delta t,\Delta s} \quad (4.8)$$

$$\alpha(\Delta t, \Delta s) = e_{\Delta t,\Delta s} \times \bar{R}_{sat}^{\ f_{\Delta t,\Delta s}} \quad (4.9)$$

where $\bar{R}_{sat}$ (mm/day) is the mean value of the satellite precipitation at each bin of the specific time accumulation Δt (e.g., 3-hr) and spatial resolution Δs (e.g., $0.25° \times 0.25°$ latitude-longitude scales). Using these three functions, the parameters of the uncertainty distribution at each rainfall rate are calculated, and the distribution can be formed.

In order to expand this method for different spatial and temporal resolutions, our data must be aggregated into those resolutions. The satellite-precipitation product and its corresponding radar data are accumulated into lower-temporal resolutions from 3-hourly, 6-hourly, 12-hourly, and daily resolutions by obtaining the sum of the data for each interval. For spatial resolutions, the average of the rainfall rates is calculated from $0.25° \times 0.25°$, $0.5° \times 0.5°$, $0.75° \times 0.75°$, and $1° \times 1°$. For each of the 16 different pairs of spatial and temporal resolutions, the distribution is fitted to the bins of the precipitation, and the distribution parameters as a function of rainfall rate are modeled as $\zeta(\Delta t,\Delta s)$, $\alpha(\Delta t,\Delta s)$ and $\kappa(\Delta t,\Delta s)$ (as in Eqs. (4.7), (4.8), and (4.9)). The six parameters from Eqs. (4.7), (4.8), and (4.9) construct parameter space for different spatial and temporal resolutions.

Previously, it was stated that Eqs. (4.7), (4.8), and (4.9) represent the parameters of the selected generalized-probability distribution as a function of rainfall rate and that each equation contains two distinct parameters (a total of six), which need to be identified. Figures 4.2-4.3 represent the 3-dimensional parameter spaces for each of the six parameters (a, b, c, d, e, and f) as a function of spatial and temporal resolutions. For better visualization, each of the six panels is plotted so that the front center shows the lowest point in the space and gradually increases towards the far back center. For the sake of clarity, if, for example, one chooses a specific spatial and temporal resolution (e.g., 3-hourly and 0.25°), the corresponding six parameters are obtained from each of these planes for summer and winter separately. By inserting those parameters into Eqs. (4.7),

(4.8), and (4.9), we have the three parameters of the product's uncertainty distribution as a function of the rainfall rate, which now allows for the construction of the uncertainty model. For any other desired spatial and temporal resolution, these parameters can be interpolated linearly in the 3-dimensional planes. A word of caution: Any extrapolation of the parameters beyond the spatial and temporal resolutions used in this study should be evaluated further.

Figures 4.4-4.5 show the joint probability of the PERSIANN product and the reference data in 16 different pairs of spatial and temporal resolutions for summer and winter, respectively. These plots are generated using the uncertainty model presented in this work. In both plots, the upper-left corner displays the highest spatial and temporal resolution, and the lower-right corner represents the coarsest resolution. The joint probability shows that products with the highest resolution have the highest amount of randomness because the distributions are scattered in the plane with low probability over almost the entire space. Moving from right to left in the lower panels of Figures 4.4-4.5, more concentration on a line in the plane can be observed, which shows a higher peak and a more pronounced mode of the distribution and less randomness. Additionally, as mentioned before in the literature (Hong *et al*., 2006), the spread of the uncertainty increases in each plot along with the rainfall rate, which shows that standard deviation is an increasing function of rainfall rate.

In the distributions for summer, there is an obvious bias where the peak of the distribution is tilted towards the lower radar value, which shows that the satellite overestimates during the summer. The distributions are less biased during the winter, when the darker points with higher probabilities are closer to a 1:1 line.

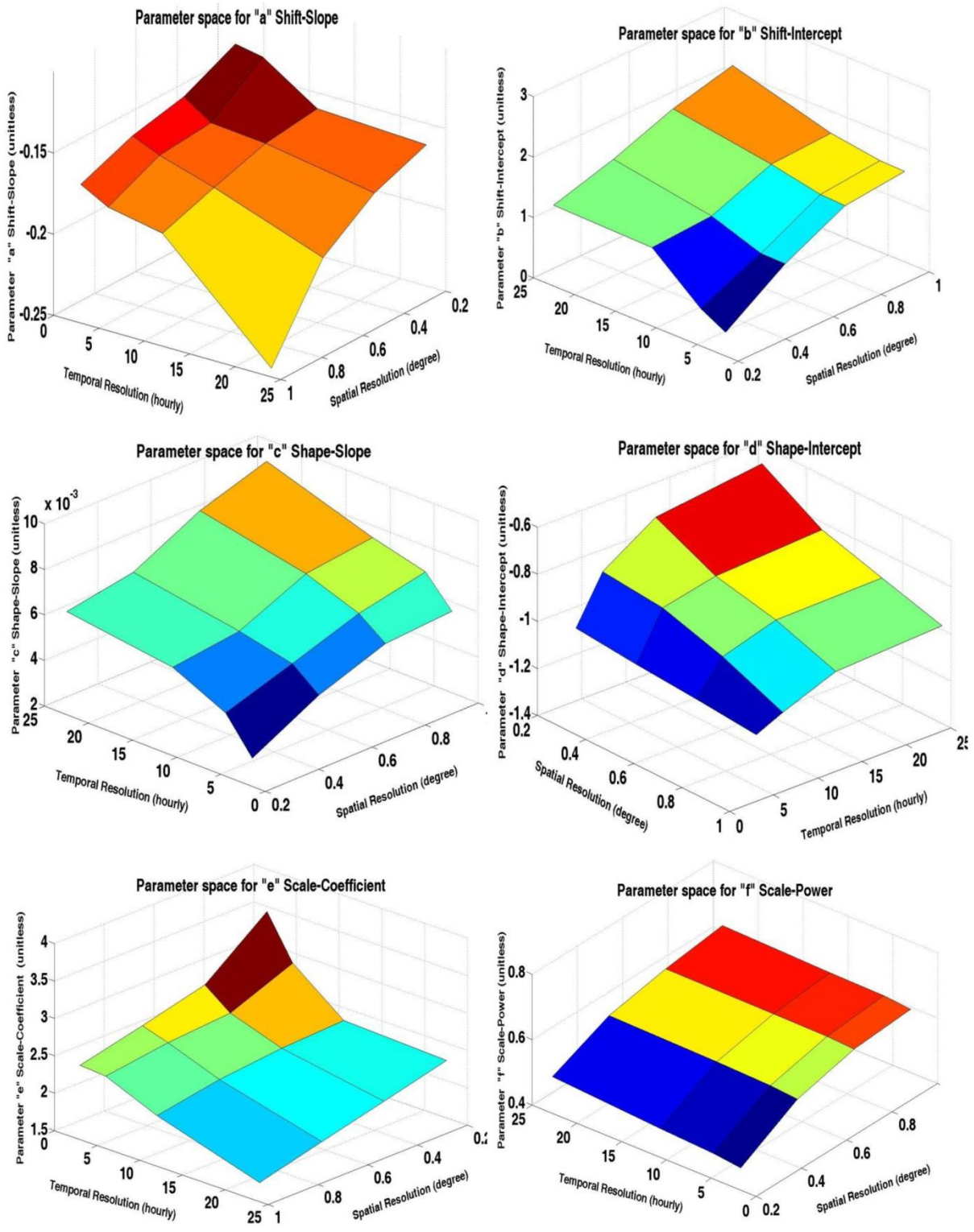

Fig. 4.2. GND parameter space for summer data.

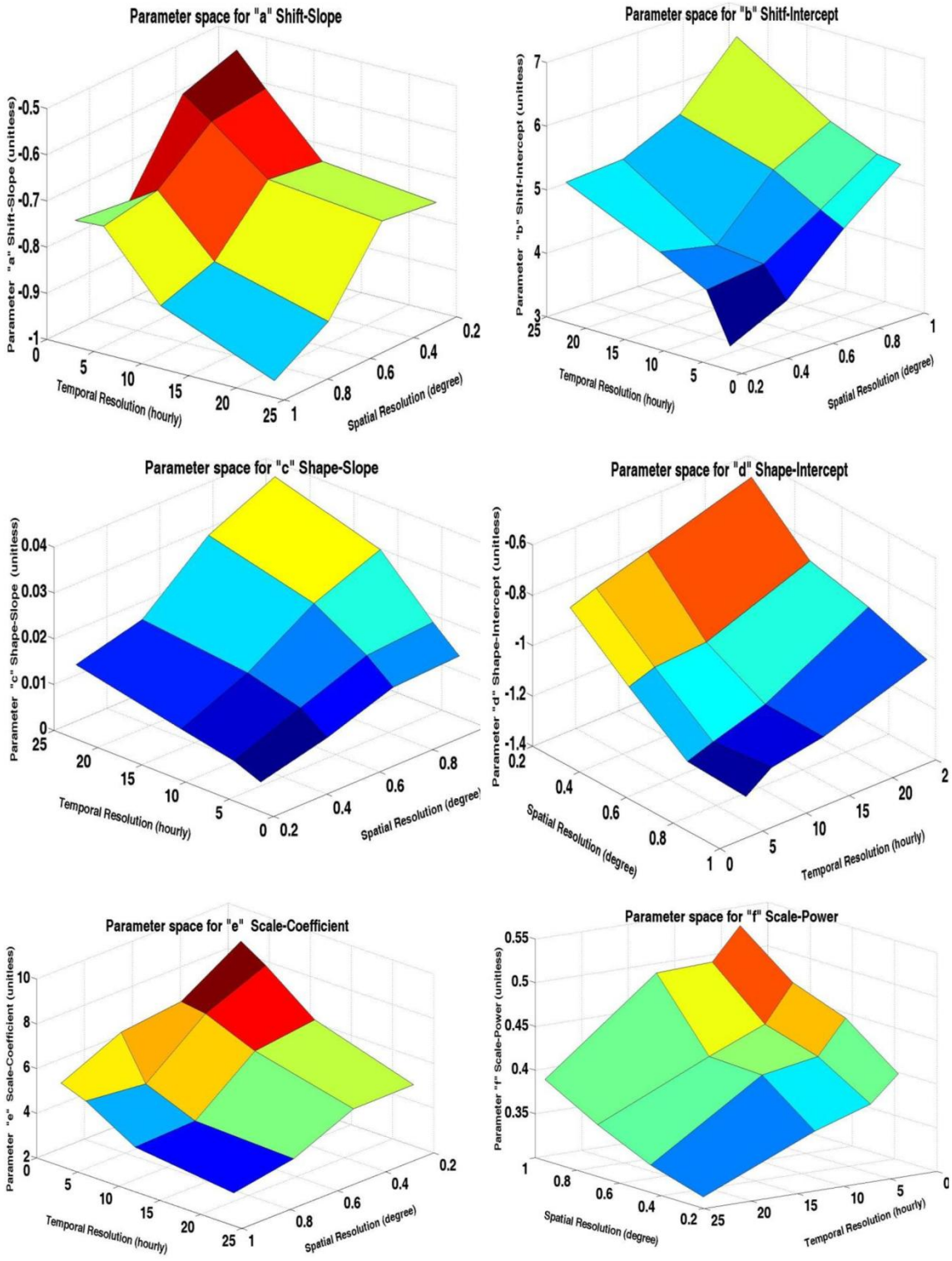

Fig. 4.3. GND parameter space for winter data.

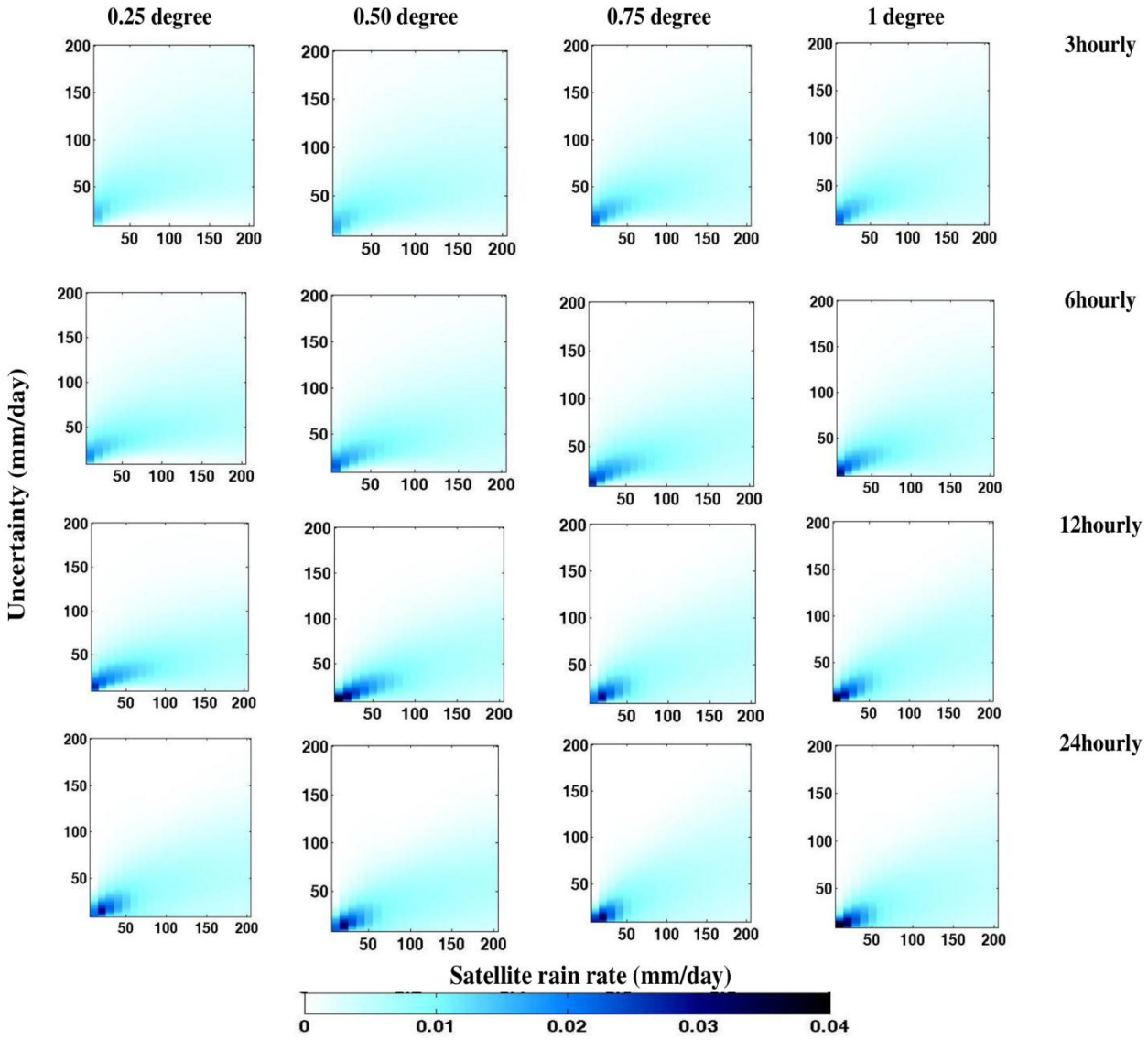

Fig. 4.4. Joint probability of radar rainfall and satellite data for summer.

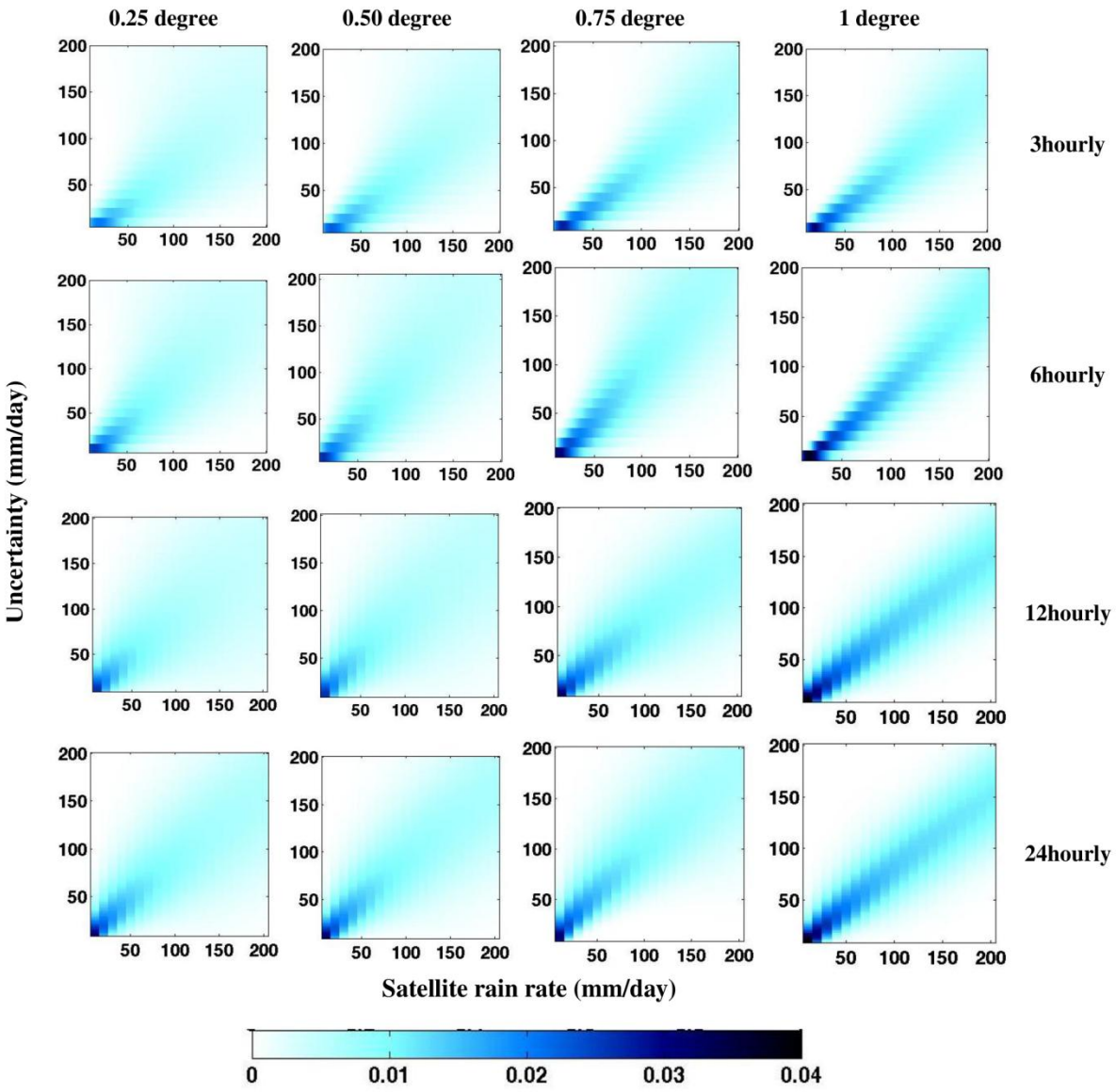

Fig. 4.5. Joint probability of radar rainfall and satellite data for winter.

4.2 Calibration of false alarm

To calibrate the model for the false alarm ratio, we refer to that section of the calibration data where the satellite precipitation is larger than the threshold (1 mm/day) and its corresponding radar precipitation is smaller than the threshold (see region [**2**] in figure 2.1).

The false alarm (FA) ratio is the conditional probability of radar showing zero precipitation when the satellite shows precipitation amounts larger than the threshold.

$$P(FA) = P\left(R_{ref} < thr \mid R_{sat} > thr\right) \tag{4.10}$$

Here, this probability is defined as the number of false alarm samples to the number of samples where the satellite detects non-zero precipitation:

$$\frac{\#\ samples\ (R_{ref} < thr\ \&\ R_{sat} > thr)}{\#\ of\ samples\ (R_{sat} > thr)} \tag{4.11}$$

To better model the false alarm, the data are divided into two different categories: one is when the satellite precipitation from the previous time step is also larger than the threshold; the other category is when the satellite precipitation from the previous time step is smaller than the threshold.

Condition 1:

$$\frac{\#\ samples\ (R_{ref}(t) < thr\ \&\ R_{sat}(t) > thr\ \&\ R_{sat}(t-1) > thr)}{\#\ samples\ (R_{sat}(t) > thr\ \&\ R_{sat}(t-1) > thr)} \tag{4.12}$$

Condition 2:

$$\frac{\#\ samples\ (R_{ref}(t) < thr\ \&\ R_{sat}(t) > thr\ \&\ R_{sat}(t-1) < thr)}{\#\ samples\ (R_{sat}(t) > thr\ \&\ R_{sat}(t-1) < thr)} \tag{4.13}$$

The FA is the highest for smaller-satellite rainfall rates and it decreases quickly. Eq.(2.6) is used to model the ratio of the false alarm, which is the conditional probability of the reference detecting precipitation smaller than the threshold given the satellite is showing a rain rate larger

than the threshold. Similar to the calibration of the GND model, for false alarm ratio we also bin the satellite data into different intervals of 5 mm/day. For each bin, we find the FA ratio from the Eqs. (4.12) and (4.13) for P(R_{rad} (t) < thr | R_{sat}(t) > thr & R_{sat}(t-1) > thr) and P(R_{rad}(t) < thr | R_{sat} (t) > thr & R_{sat} (t-1) < thr) consecutively. Then using the maximum likelihood method, we fit Eq. (2.6) to the FA ratios as a function of the mean satellite rain rate of each bin. We are careful that the parameter b of Eq. (2.6) is always smaller than one. Investigating the effect of the temporal correlation of satellite precipitation with the precipitation from the previous time step on the FA ratio shows that, during the summer, the FA ratio is higher when the previous time step also shows rainfall; during the winter, this pattern is reversed. This can be concluded from different precipitation types during the summer and winter. In the summer, when short, convective precipitation ceases, clouds remain in the sky, which results in a false detection of precipitation by the satellite, but this is not the case during the winter. Results show that the FA ratio difference between the two cases is less distinct than for larger time steps. Similar to the parameter spaces for modeling the hit precipitation presented in the previous section, the exponential model is fitted to data with different spatial and temporal resolutions, and the parameter spaces are generated for the two parameters of the model.

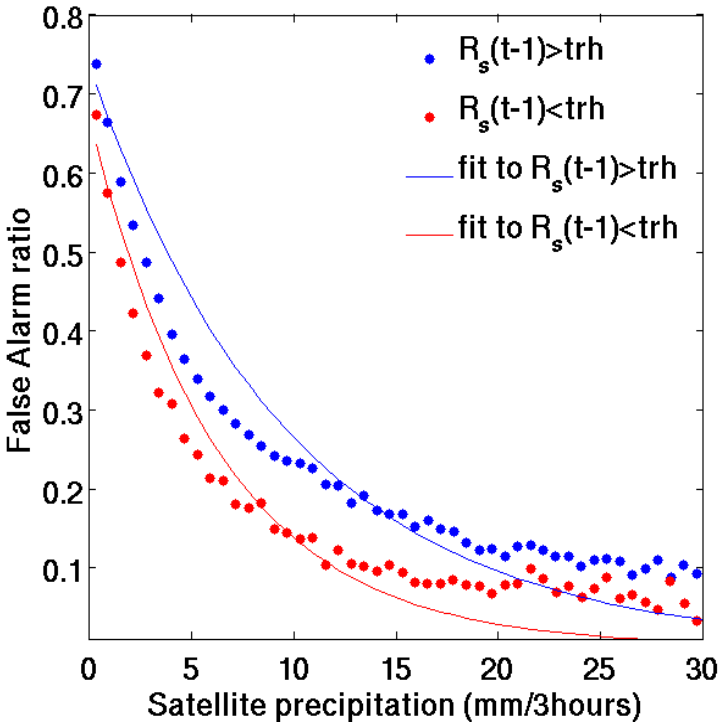

Fig. 4.6. False alarm ratio for summer data over the calibration domain for 2005-2009, in 3-hourly and 0.25 °resolution.

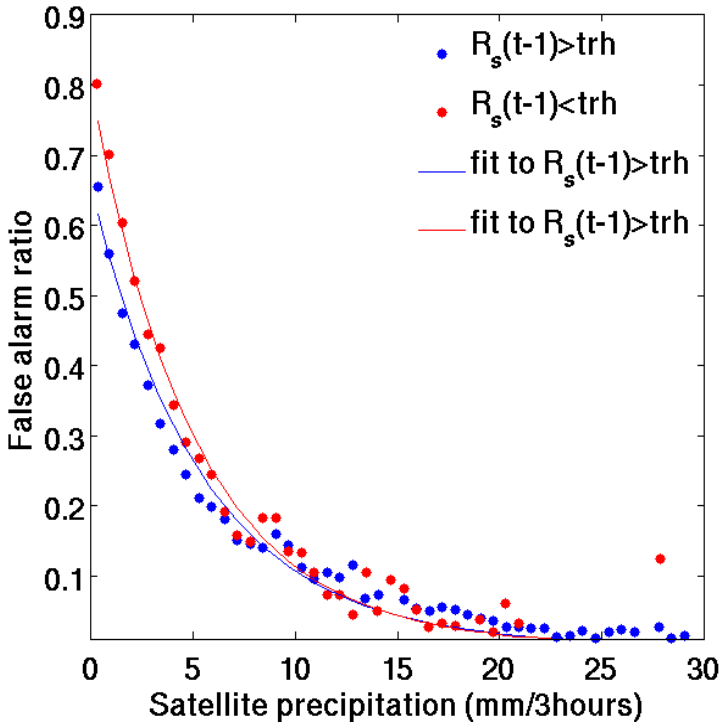

Fig. 4.7. False alarm ratio for winter data over the calibration domain for 2005-2009, in 3-hourly and 0.25 °resolution.

The data from Figures 4.6-4.7 show how the false alarm is different in the summer and winter in that it decreases faster during the winter. Additionally, the false alarm dependence on the satellite precipitation from the previous time step is more distinct during the summer, whereas there is very little difference in the winter each figure contains the FA ratio and the fitted function in the form of the exponential decay.

These data are modeled using an exponential formula in the form of the Eq. (2.6):

$$FA_{\Delta t,\Delta s} = b_{\Delta t,\Delta s} \times exp\,(-r_{\Delta t,\Delta s} \times R_{sat}\,) \qquad (4.14)$$

This model is fitted to summer and winter separately using the least squared error method. It is also fitted separately in each season to the situations when the previous time step detects rainfall or detects zero precipitation. The results of these fitted parameters in space are shown in Figures 4.8-4.9 below:

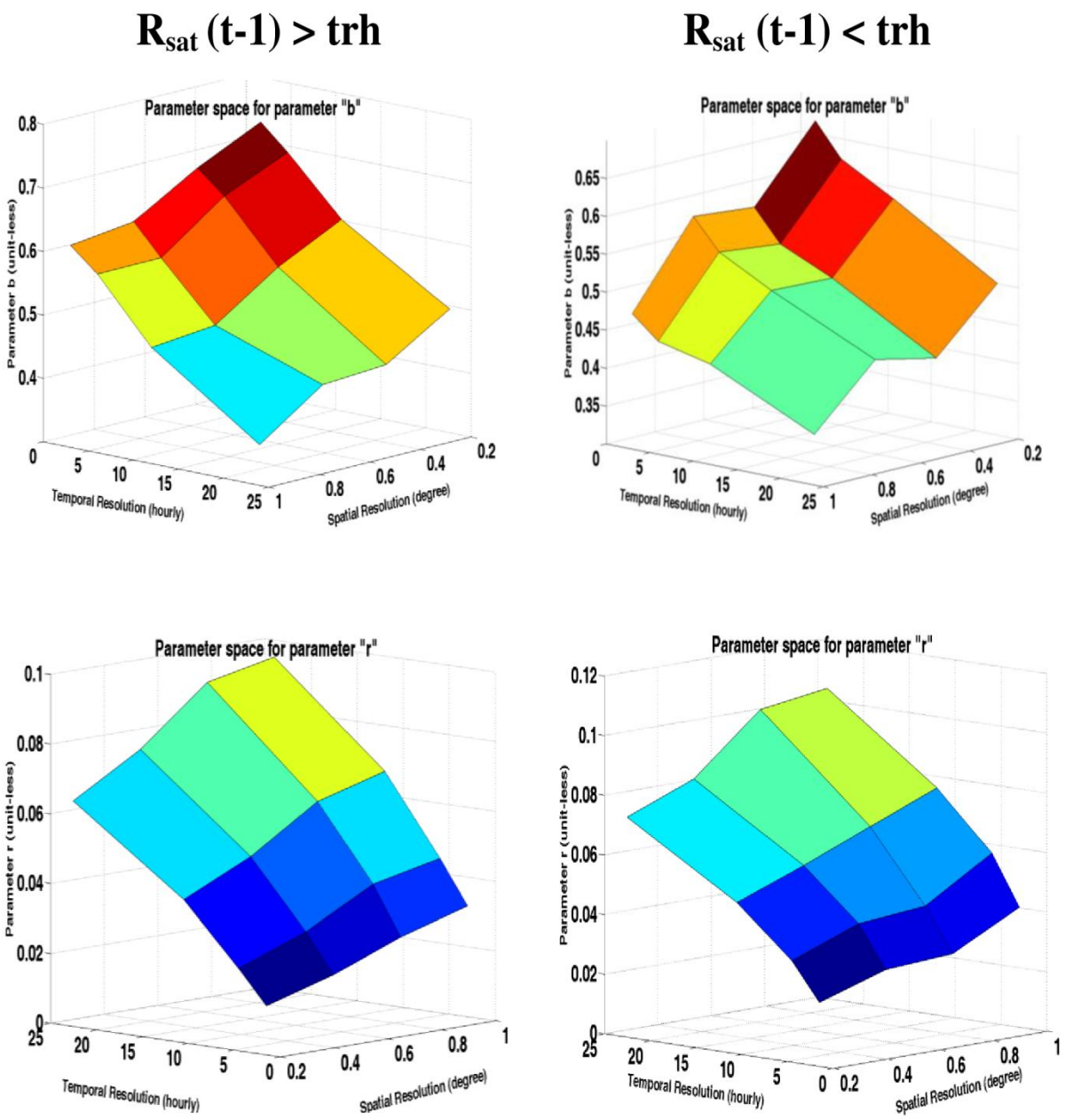

Fig. 4.8. Parameter space for parameters r and b from Eq. (4.14) for summer data for Rsat (t-1) > thr (left panel) and Rsat (t-1) < thr (right panel).

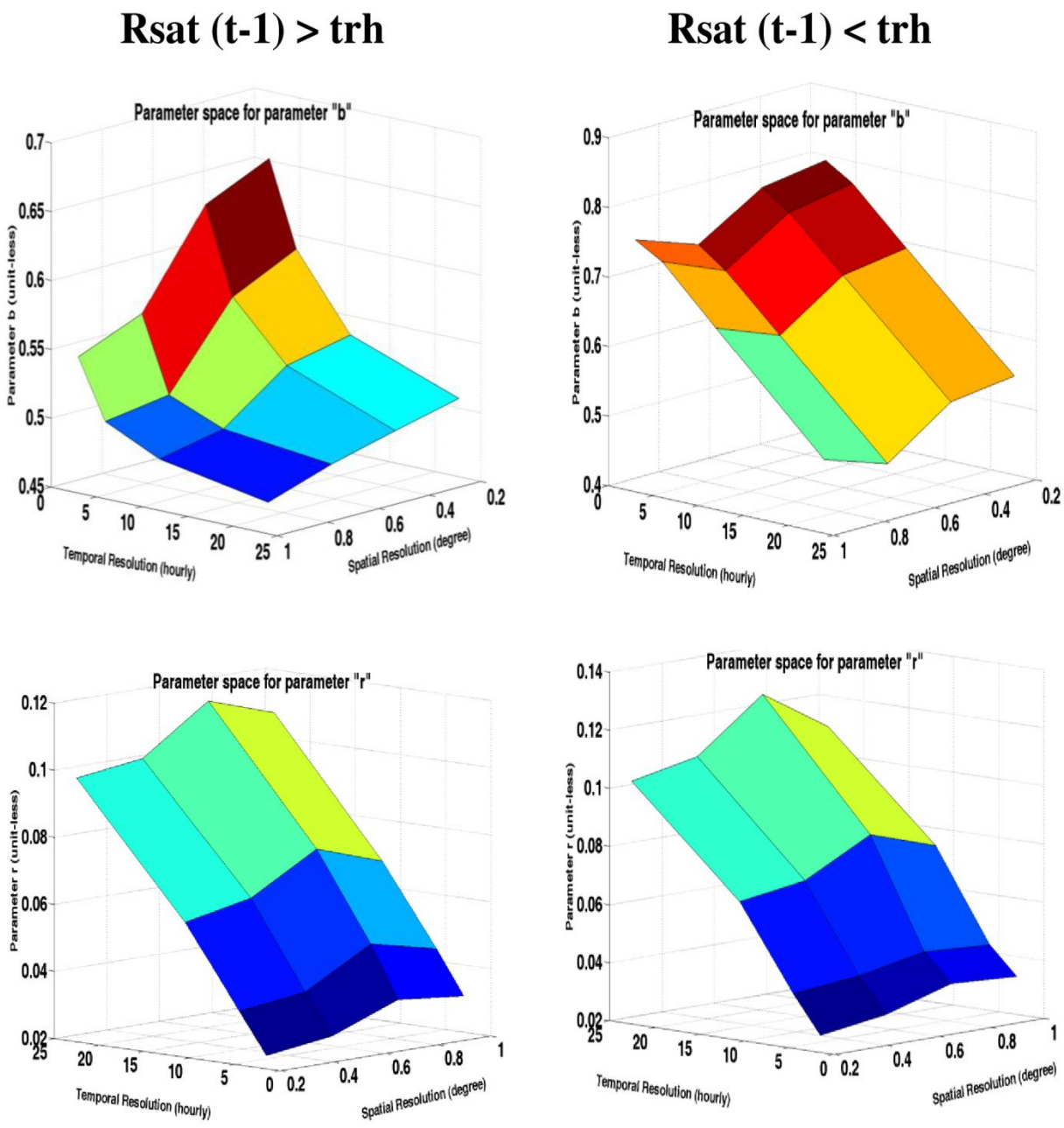

Fig. 4.9. Parameter space for parameters r and b from Eq. (4.14) for winter data for Rsat (t-1) > thr (left panel) and Rsat (t-1) < thr (right panel).

4.3 Calibration of missed precipitation

To calibrate the model for missed precipitation (see region [3] in Figure 2.1), we used data when the satellite precipitation detects rainfall smaller than the threshold and its corresponding reference precipitation shows rainfall larger than the threshold. In this case, because the only information available is zero rainfall for the product at a specific season and a specific spatial and temporal resolution, a single value for probability is assigned to the percent of missed precipitation, which is:

$$P\,(missed\ precipitation) = P\left(R_{ref} > thr \,\middle|\, R_{sat} < thr\right) \tag{4.15}$$

In this study, this probability is defined as:

$$\frac{\#\ samples\ (\ R_{ref} > thr\ \&\ R_{sat} < thr)}{\#\ of\ samples\ (R_{sat} < thr)} \tag{4.16}$$

In the case of missed precipitation, a Gamma distribution is used to model the distribution of the missed radar rainfall given the satellite rain rate is smaller than the threshold as shown in Eq. (2.10).

Here, we also use satellite precipitation from the previous time step as another source of information and define the probability of missed precipitation as:

Condition 1:

$$\frac{\#\ samples\ (R_{ref}(t) > thr\ \&\ R_{sat}(t) < thr\ \&\ R_{sat}(t-1) < thr)}{\#\ samples\ (R_{sat}\ (t) < thr\ \&\ R_{sat}\ (t-1) < thr)} \tag{4.17}$$

Condition 2:

$$\frac{\#\ samples\ (R_{ref}\ (t) > thr\ \&\ R_{sat}(t) < thr\ \&\ R_{sat}\ (t-1) > thr)}{\#\ samples\ (R_{sat}(t) < thr\ \&\ R_{sat}(t-1) > thr)} \tag{4.18}$$

Data show that, for both summer and winter, the missed precipitation ratio is 10% or less when the satellite from the previous time step shows zero precipitation. Because of the very low

percentage of missed rainfall, we consider these cases as hit-zero precipitation. When the satellite from the previous time step shows rainfall, uncertainty is sampled from the fitted Gamma distribution. Figures 4.10-4.11 show the parameter space for the missed precipitation ratio for summer and winter, respectively.

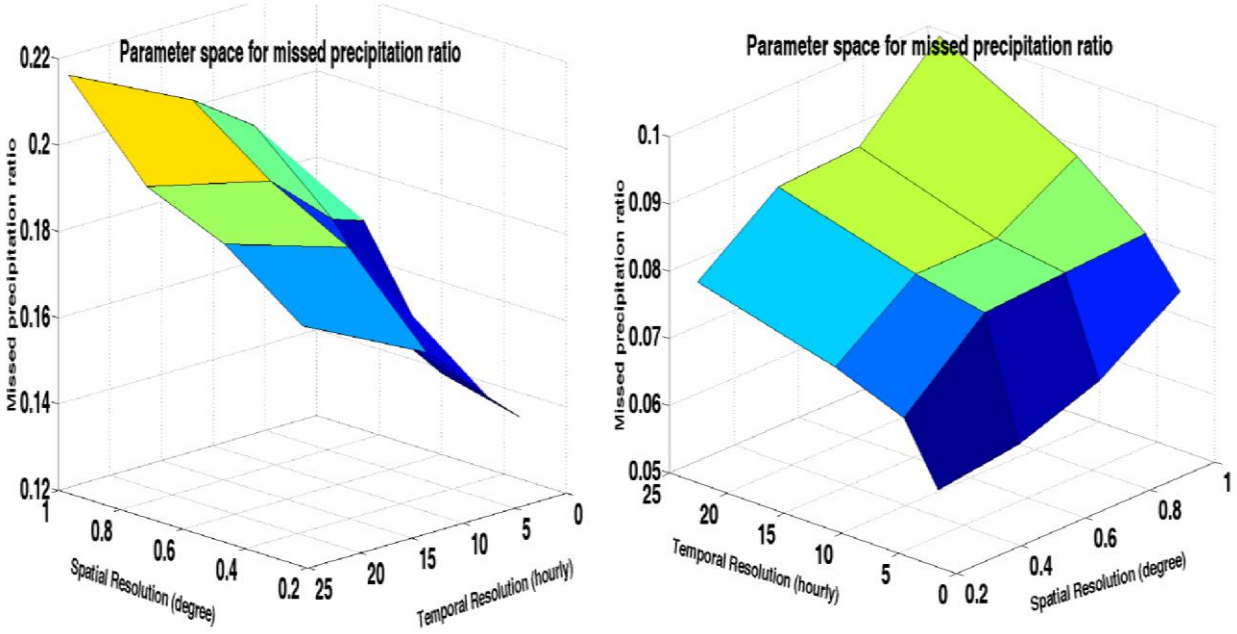

Fig. 4.10. Missed precipitation ratio over the region for 2005-2009 for summer. Left panel: when precipitation at the previous time step is not zero. Right panel: when precipitation at the previous time step is zero.

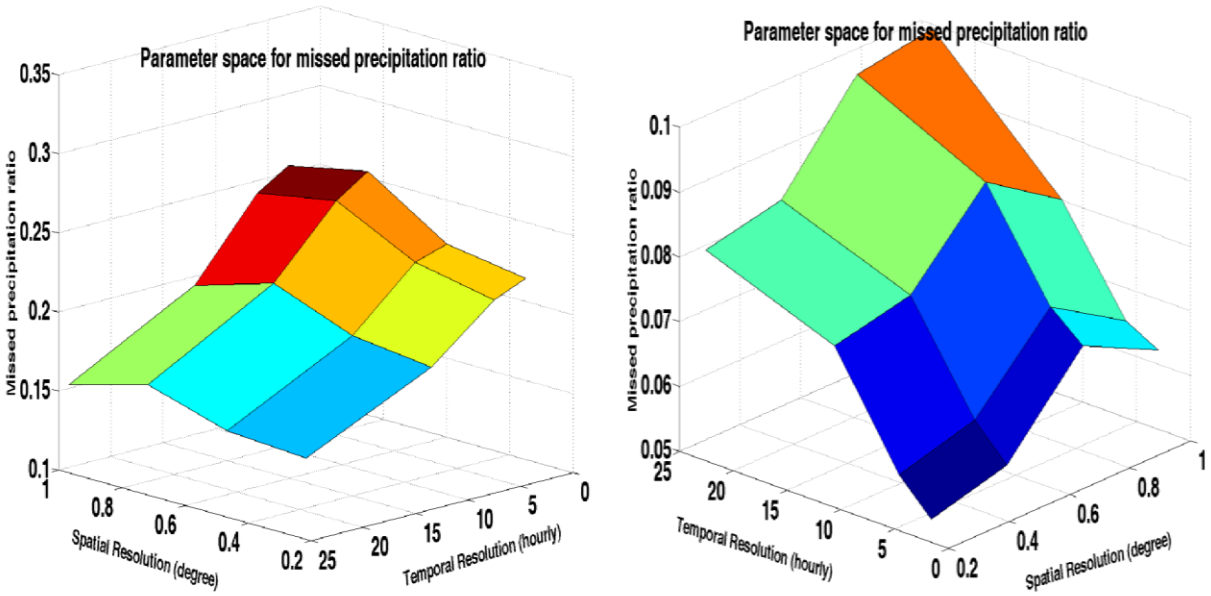

Fig. 4.11. Missed precipitation ratio over the region for 2005-2009 for winter. Left panel: when precipitation at the previous time step is not zero. Right panel: when precipitation at the previous time step is zero.

The two examples given in Figures 4.12-4.13 show how the missed radar data are distributed. To model these data, at each pair of spatial and temporal resolutions, a Gamma function is fitted to the data, and its two parameters are estimated. The parameter spaces for summer and winter are shown in Figures 4.14-4.15, respectively.

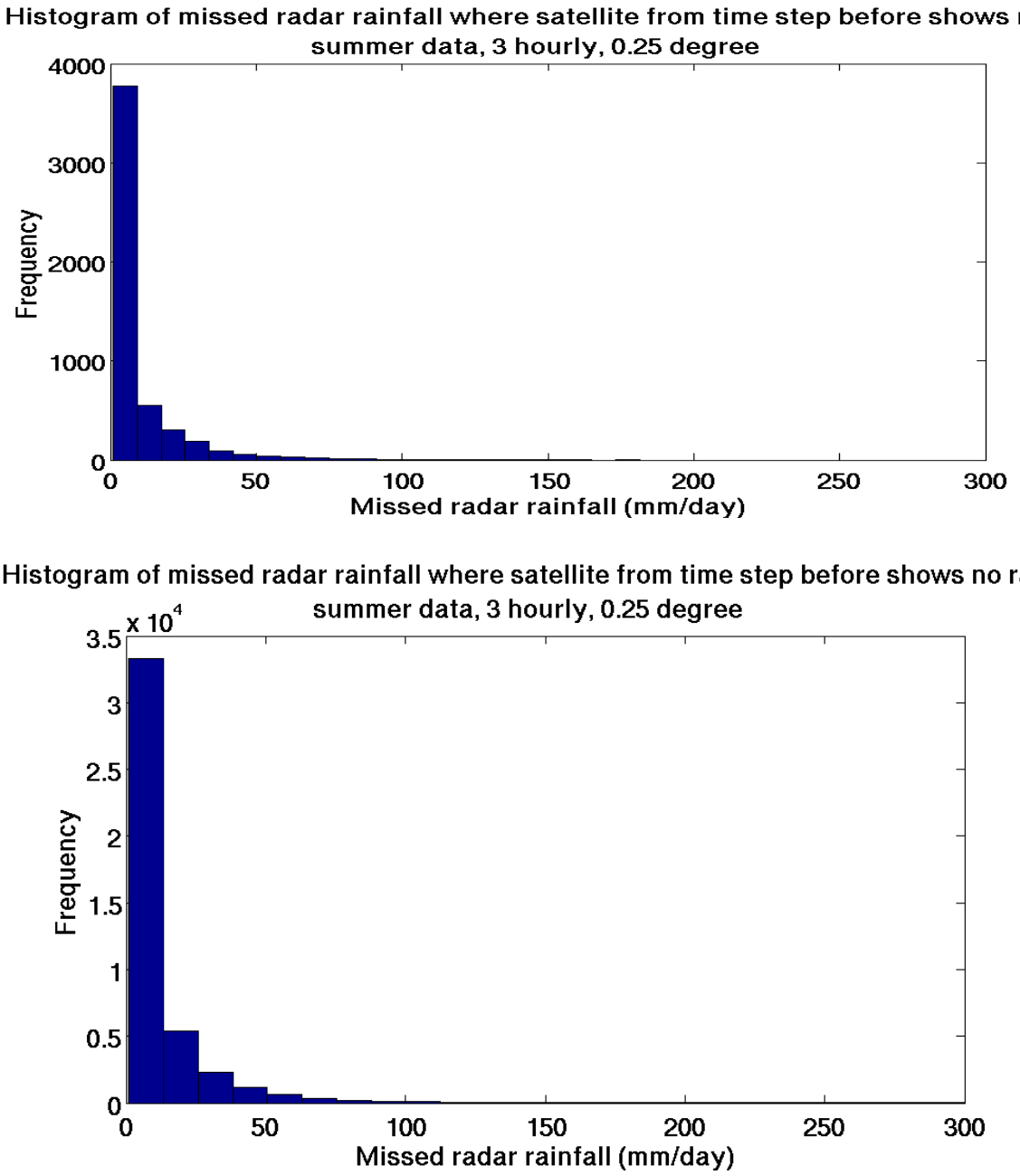

Fig. 4.12. Histogram of missed radar rainfall over the study domain for the summer for 2005-2009 in 3-hourly and 0.25 °resolution. Top panel: The satellite from (t-1) shows rainfall. Bottom panel: The satellite from (t-1) shows no rainfall.

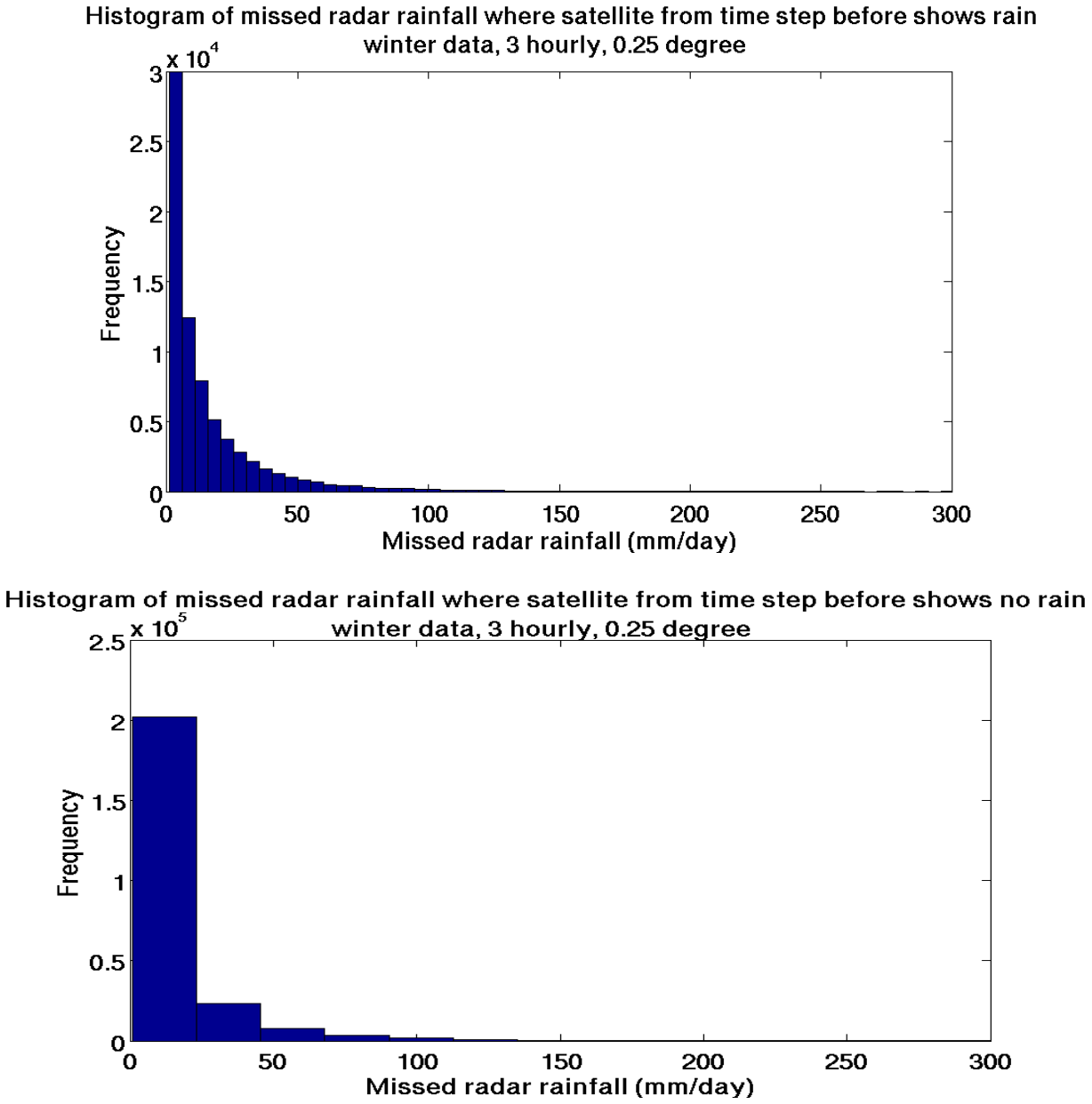

Fig. 4.13. Histogram of missed radar rainfall over the study domain for the winter for 2005-2009 in 3-hourly and 0.25 °resolution. Top panel: The satellite from (t-1) shows rainfall. Bottom panel: The satellite from (t-1) shows no rainfall.

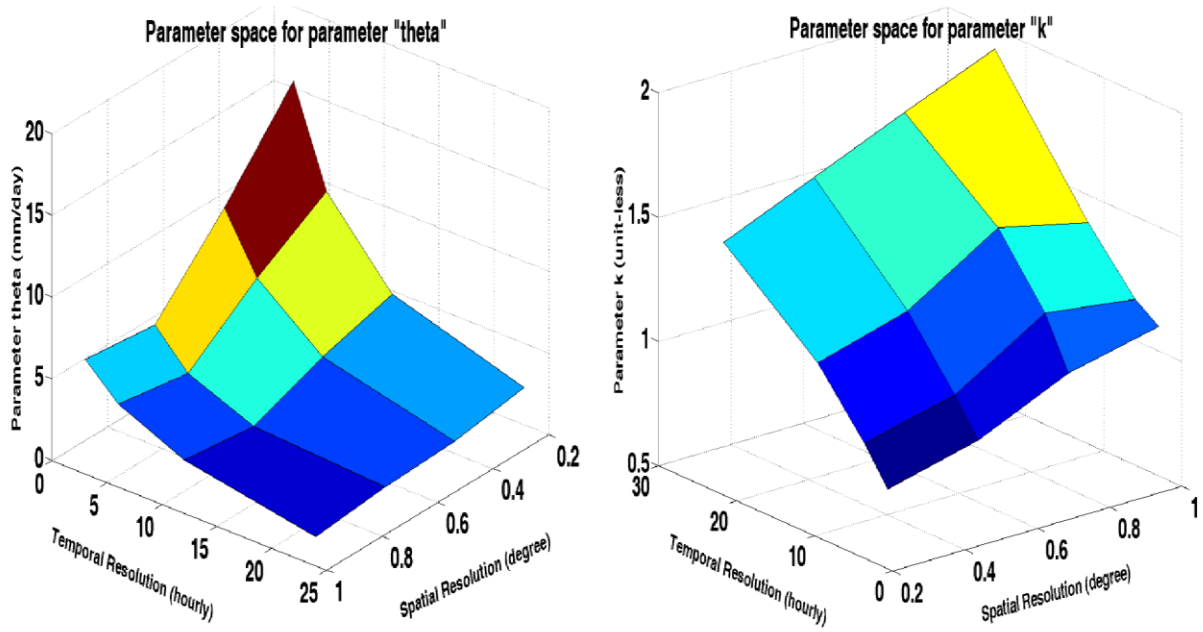

Fig. 4.14. Parameter space for Gamma distribution fitted to the missed radar precipitation over the region for 2005-2009 for summer, when the satellite from the previous time step shows rainfall. Left panel: Parameter Theta. Right panel: Parameter k.

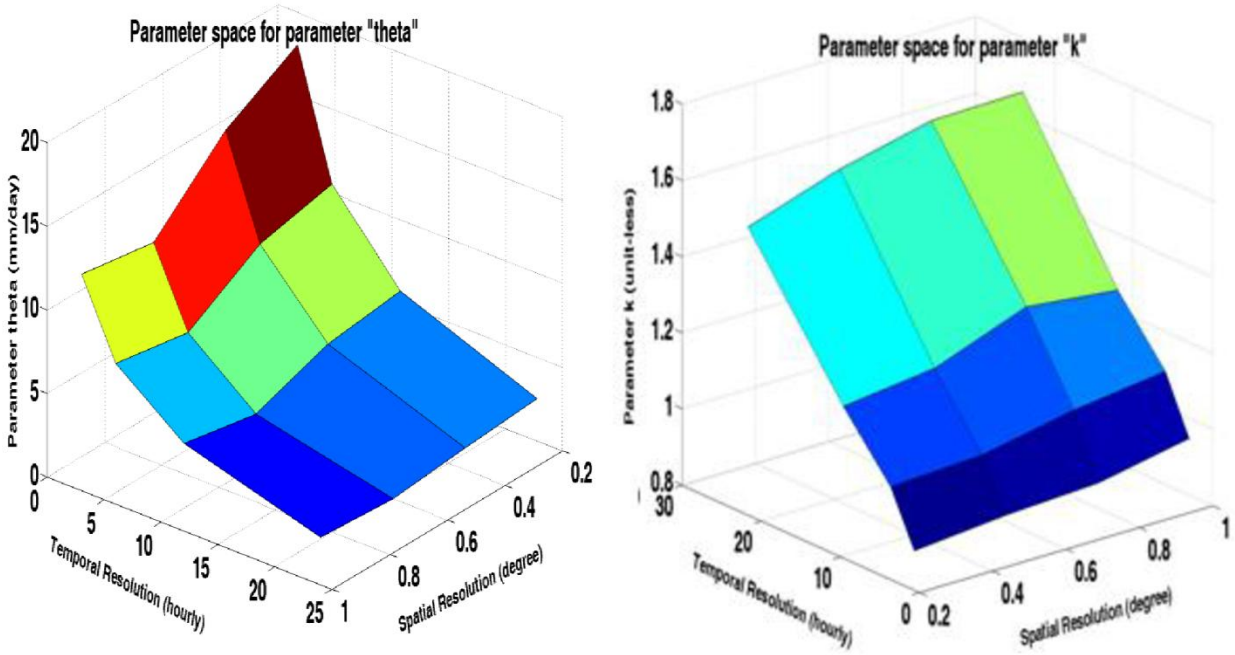

Fig. 4.15. Parameter space for Gamma distribution fitted to the missed radar precipitation over the region for 2005-2009 for winter, when the satellite from the previous time step shows rainfall. Left panel: Parameter Theta. Right panel: Parameter k.

Chapter 5: Model Evaluation over the Southeastern United States

An evaluation of this uncertainty model is presented in two different cases. The first case only evaluates the performance of the GND model to model the hit precipitation. In this case, it is assumed that, if the satellite shows a non-zero rainfall rate, it definitely rained, and the only source of uncertainty is bias in the value of the rainfall rate. The second case considers the probability of false alarm and missed precipitation.

5.1 GND model evaluation

To evaluate the GND model, the comparison of root mean squared error (RMSE) of fitting the uncertainty of PERSIANN and the corresponding Stage-IV data for different pairs of spatial and temporal resolutions is shown in Figure 5.1. The RMSE is calculated at each specific resolution as:

$$\boldsymbol{RMSE}\left(\bar{\boldsymbol{R}}_{\boldsymbol{sat}}, \Delta \boldsymbol{t}, \Delta \boldsymbol{s}\right) = \left(\sum \boldsymbol{CDF}_{\boldsymbol{x}}\left(\boldsymbol{R}_{\boldsymbol{ref}} | \bar{\boldsymbol{R}}_{\boldsymbol{sat}}\right) - \boldsymbol{CDF}_{\boldsymbol{emp}}\left(\boldsymbol{R}_{\boldsymbol{ref}} | \bar{\boldsymbol{R}}_{\boldsymbol{sat}}\right)\right)^{\mathbf{2}} \quad (5.1)$$

Where $\bar{\boldsymbol{R}}_{\boldsymbol{sat}}$ is the mean rainfall of each bin of the satellite precipitation product, x is any of the 5 different distributions that are mentioned below. $\boldsymbol{CDF}_{\boldsymbol{x}}\left(\boldsymbol{R}_{\boldsymbol{ref}} | \bar{\boldsymbol{R}}_{\boldsymbol{sat}}\right)$ is the CDF of the fitted distribution to the reference data corresponding the each bin and $\boldsymbol{CDF}_{\boldsymbol{emp}}\left(\boldsymbol{R}_{\boldsymbol{ref}} | \bar{\boldsymbol{R}}_{\boldsymbol{sat}}\right)$ is the empirical CDF of the reference data. Here, several types of distributions are used, including GND, normal, Gamma, log-normal, and Weibull distributions. The results show that the GND model fits better to those data than others. GND distribution is also able to model the skewness of the joint distribution associated with satellite- and Stage-IV precipitation products that cannot be estimated by symmetrical distribution, such as Gaussian.

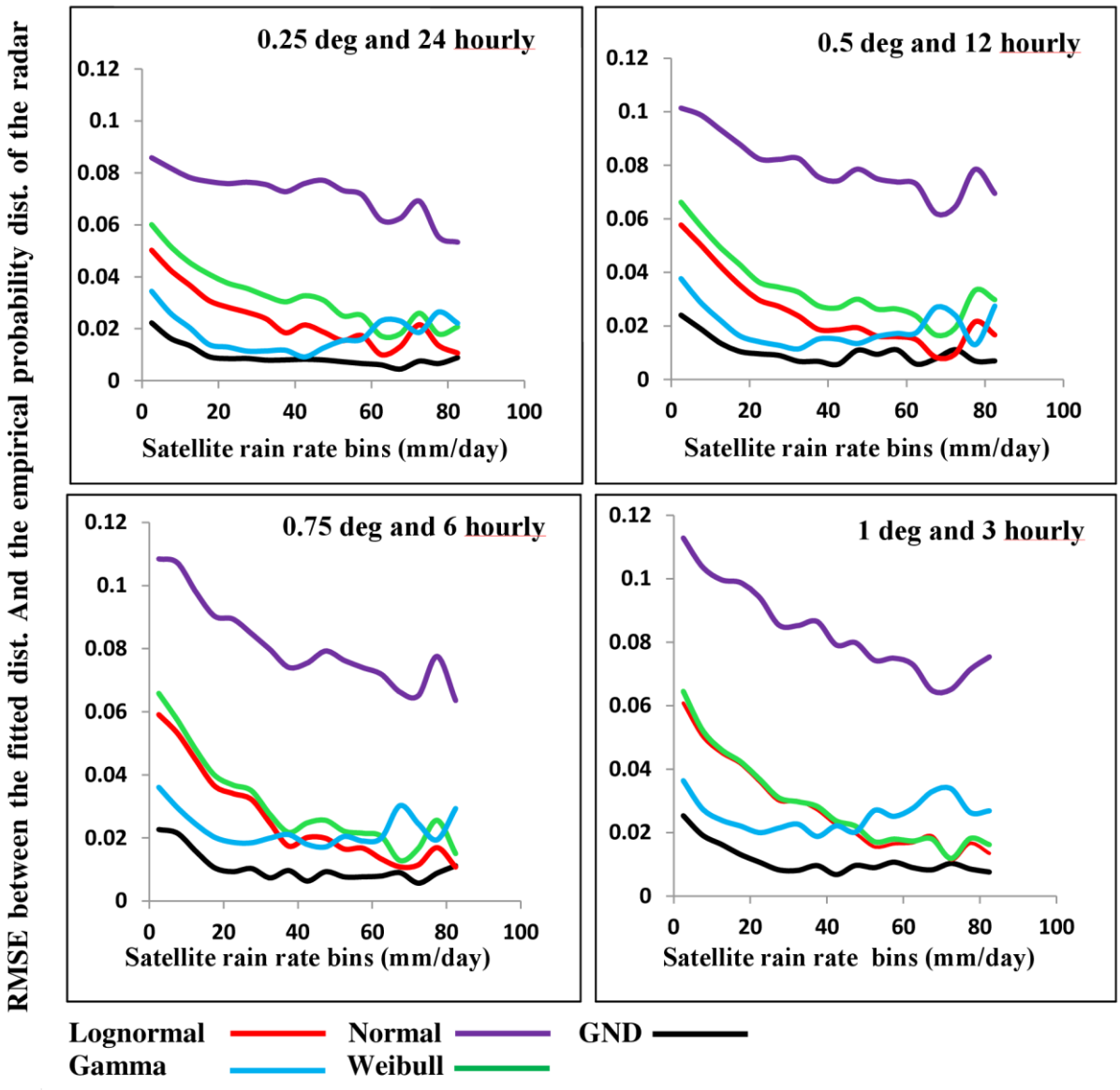

Fig. 5.1. Comparison between the RMSE of fitting the proposed distribution and Gamma, normal, log-normal, and Weibull distributions for the resolutions of 0.25 °, 24-hourly (upper-left), 0.5 °, 12-hourly (upper-right), 0.75 °, 6-hourly (lower left), and 1 °, 3-hourly (lower right) for the joint probability of PERSIANN and Stage-IV radar data during the summer and winter of 2004-2006. Black: GND, Blue: Gamma distribution, Purple: normal distribution, Red: log-

normal distribution, Green: Weibull distribution.

From Figure 5.1, we conclude that the Gamma distribution is also a very good fit to the data in competition with the GND. One might argue that, because the latter has one more parameter than the Gamma distribution, it possesses more flexibility to fit to the data. To further compare the goodness-of-fit between these two models, we looked at the Bayesian Information Criterion (BIC). The BIC will examine if the GND is a better fit compared to the Gamma distribution by reducing the effect of the number of parameters. BIC is calculated from Eq. (5.2):

$$BIC = -2 \times \ln(\hat{L}) + k \times \ln(n) \tag{5.2}$$

In this equation, $\hat{L}$ is the likelihood function of the fitted distribution at observation points, k is the number of parameters, and n is the number of observation points (here, the number of data in each bin). The model with a smaller BIC value is a better fit to the data. Here, similar to Figure 5.1, we compared the BIC values for the GND and the Gamma models at each bin of data for different combinations of spatial and temporal resolutions.

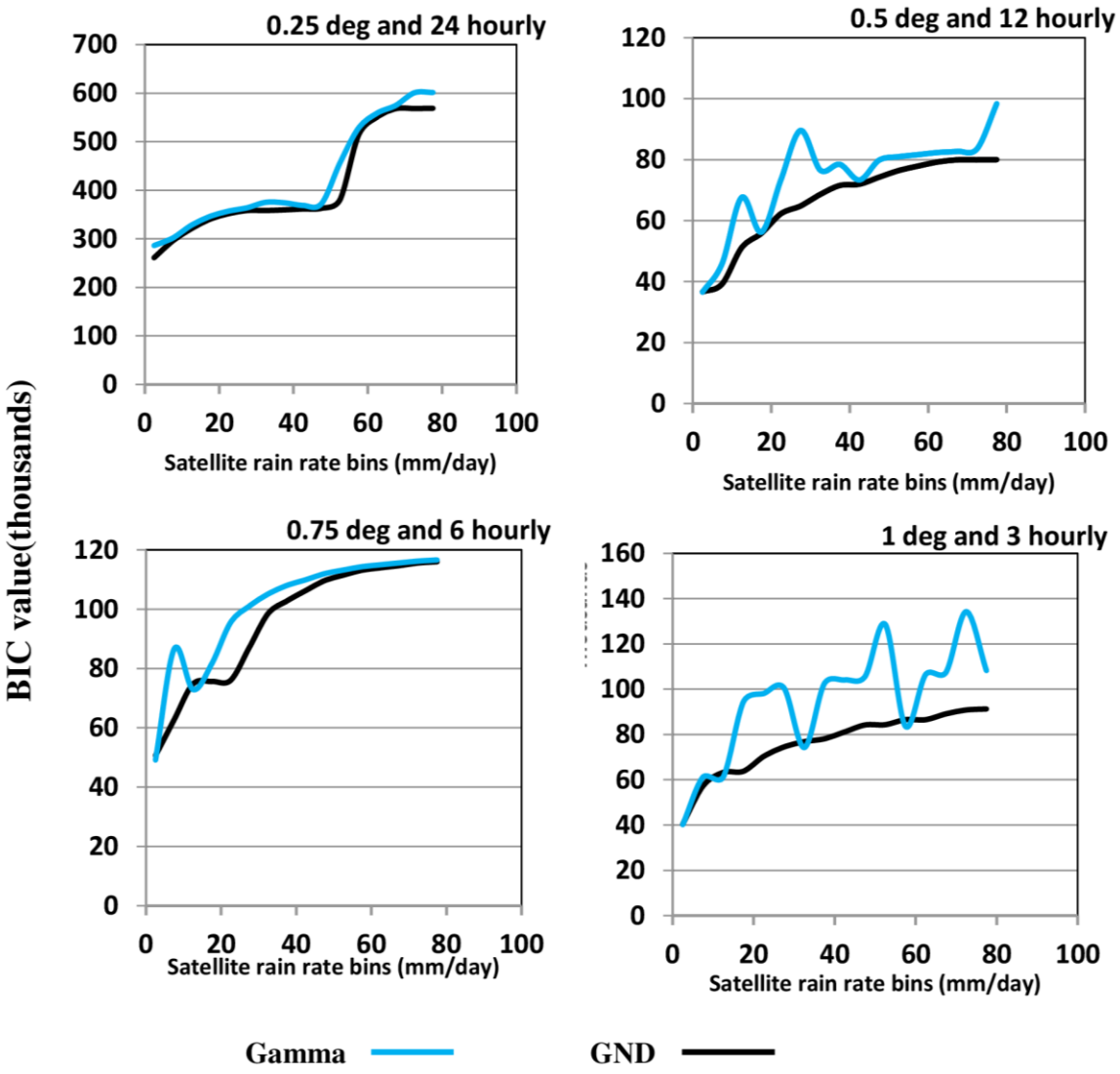

Fig. 5.2. Comparison between the BIC of fitting the proposed distribution and Gamma distribution for the resolutions of 0.25 °, 24-hourly (upper- left), 0.5 °, 12-hourly (upper-right), 0.75 °, 6-hourly (lower left), and 1 °, 3-hourly (lower right) for the joint probability of PERSIANN and Stage-IV radar data during the summer and winter of 2004-2006. Black: GND, Blue: Gamma distribution.

Figure 5.2 shows the comparison of the BIC value for fitting the GND and the Gamma distributions at different spatial and temporal resolutions. From the figure, we conclude that the GND always has a smaller BIC value compared to the Gamma distribution, which shows its superiority in modeling the conditional probability of the reference given the satellite-precipitation rainfall rate.

The proposed approach models the non-Gaussian joint distribution of the satellite and radar estimates when both are larger than the threshold, which enables us to estimate the range of the uncertainty of the satellite estimates (e.g., upper and lower bounds of the 90% interval). To further evaluate the uncertainty range obtained from the model, for the year 2010, the 80% and 90% uncertainty intervals (10%-90% and 5%-95% uncertainty ranges) are calculated for a domain from 30 °N-40 °N and 85 °W-95 °W. This domain also contains our calibration domain. The data are at 0.25 ° and 3- hourly resolutions and, for all of the year 2010 that equals 2,920 images of 40 × 40 pixels. For each pixel, using Δt= 3-hourly and Δs= 0.25 ° and the rainfall rate of that pixel, the parameters of the GND model are estimated, and the uncertainty model is generated. Using the inverse CDF of the distribution, the 10% and 90% rainfall rates and the 5% and 95% rainfall rates are estimated. For each pixel, the percentage of the images in which the radar falls in the uncertainty range is calculated, and the result is illustrated below. Figure 5.3 shows the percentage of detection for and the year 2010. From the figure, we see how the uncertainty range of the model is able to capture the radar rainfall rate for more than 75% (average of the percentages of all the pixels) of the time for the 90% uncertainty ranges and 65% (average of the percentages of all the pixels) of the time for the 80% uncertainty ranges. This difference between the uncertainty range and the percent range of the data detected is because at this stage, the model does not cover the the missed precipitation and false alarms. These parts

will be added to the model in section 5.2.

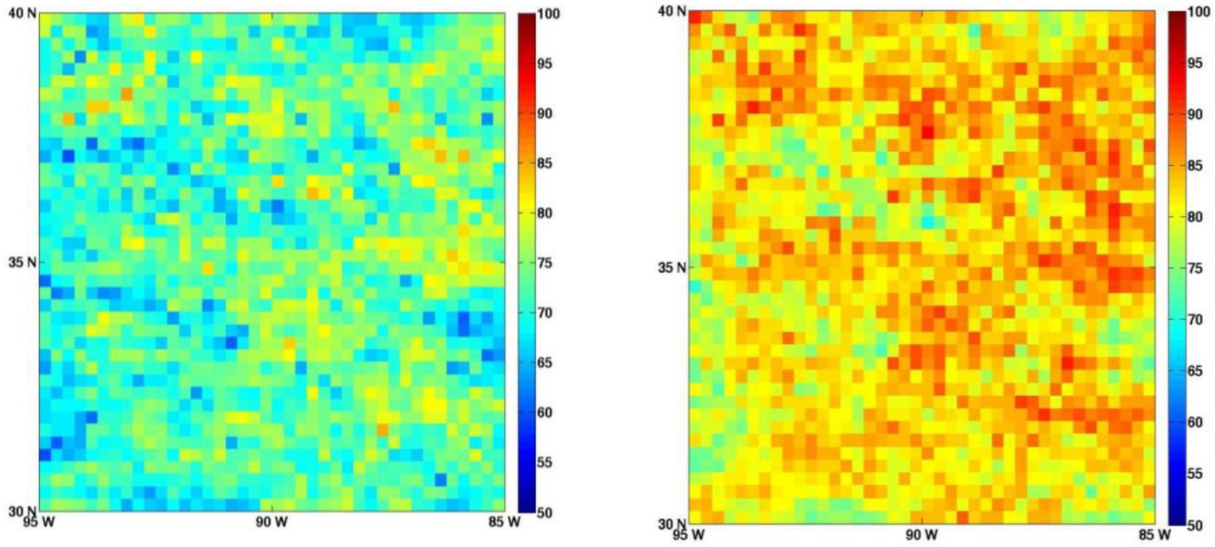

Fig. 5.3. Percentage of radar data which falls into 80% of the uncertainty range (left) and 90% of the uncertainty range (right) for 0.25 and 3-hourly data for 2010.

5.2 Complete model evaluation

Similar to the previous section, the complete model is evaluated over the calibration domain (30 ℕ-40 ℕ and 85 ℃-95 ℃) for 2010, and the 10% and 90% rainfall rates and the 5% and 95% rainfall rates are estimated. The percentage of the images in which the radar falls in the uncertainty range is calculated, and the result is illustrated in Figure 5.4. From these figures, we see how the uncertainty range of the model is able to capture the radar rainfall rate for more than 85% of the time for the 90% uncertainty ranges and 80% of the time for the 80% uncertainty ranges. This improvement in the detection of the correct rainfall is because of the addition of other parts of the uncertainty to the model.

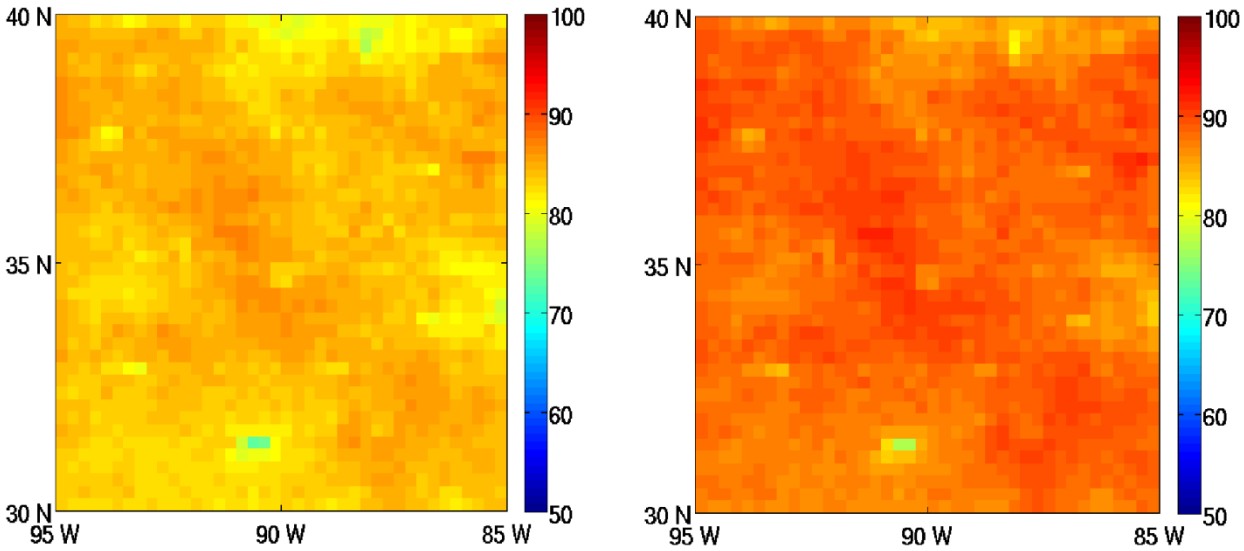

Fig. 5.4. Percentage of radar data which falls into 80% of the uncertainty range (left) and 90% of the uncertainty range (right) for 0.25 °and 3-hourly data for 2010.

To further investigate the results, we also looked at the pattern of the radar rainfall over the year 2010. To do so, we got the mean value of all the 3 hourly radar images over the region and presented in Figure 5.4.

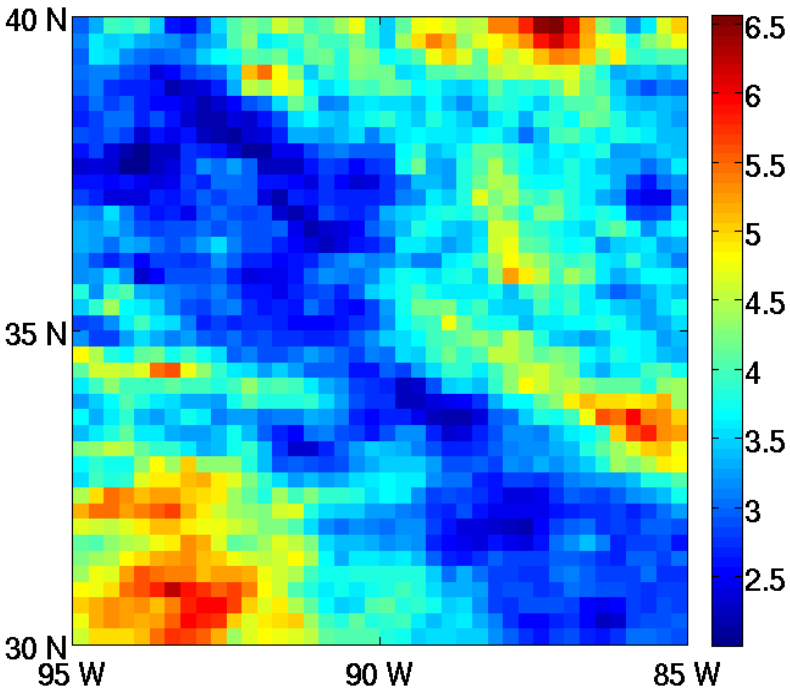

Fig. 5.5. Annual average of the 3 hourly Stage IV radar rainfall over the valibration domain for 0.25 °and 3-hourly data for 2010.

When we compare Figures 5.4 and 5.5, we can conclude that the detection of the radar by the uncertainty model is dependent of the pattern of the rainfall. Those parts of the region given in Figure 5.5 with smaller amounts of mean precipitation have a higher percentage of detection from those regions shown in Figure 5.4. This pattern is not recognizable from Figure 5.4, which does not take into account the false alarm ratio and the missed precipitation parts of the uncertainty. The results confirm that adding that information will better model the uncertainty of the satellite precipitation and better capture the radar precipitation as the reference point.

Chapter 6: Model Evaluation over a Case Study Watershed: Illinois River Basin, South of Siloam Springs, AR

After calibrating the model and constructing the parameter space, we evaluated it over a watershed as a case study.

6.1 Case study

The model is evaluated for a case study using 3 years of data from 2006-2008 over the Illinois River basin located upstream of the U.S. Geological Survey (USGS) gauging station (07195430) south of Siloam Springs, AR (Figure 6.1). The watershed has been used as a test basin for the Distributed Modeling Intercomparison Project (DMIP). The size of the Siloam watershed is 1,489 km^2. The elevation ranges from 285 m at the outlet to 590 m at the highest, and the basin's land cover can be described as uniform, with approximately 90% of the basin area being covered by deciduous broadleaf forest and the remainder being mostly wood. The dominant soil types in the basin are silty clay (SIC), silty clay loam (SICL), and silty loam (SIL). The average annual rainfall and runoff of the basin are about 1,200 and 300 mm/year, respectively (Smith *et al*., 2004). The Illinois River basin is free of major complications, such as orographic influences, significant snow accumulation, and stream regulations (Smith *et al*., 2004).

For this experiment, the PERSIANN data over the domain are used as the precipitation data, and Stage-IV radar data are used as the reference data. The data are averaged over the domain to determine the mean areal precipitation and aggregated from 3- to 6-hourly. The area is 1,480 km^2, which is approximately $0.35° \times 0.35°$ (each degree is considered to be about 111 km). For each point of the PERSIANN time series, considering its temporal resolution (6-hourly), spatial

resolution (0.35), and rainfall rate, the uncertainty-distribution parameters are calculated using the parameter spaces introduced in this study. The season of the precipitation should also be considered here, namely events from May-October for the summer precipitation and November-April for the winter precipitation.

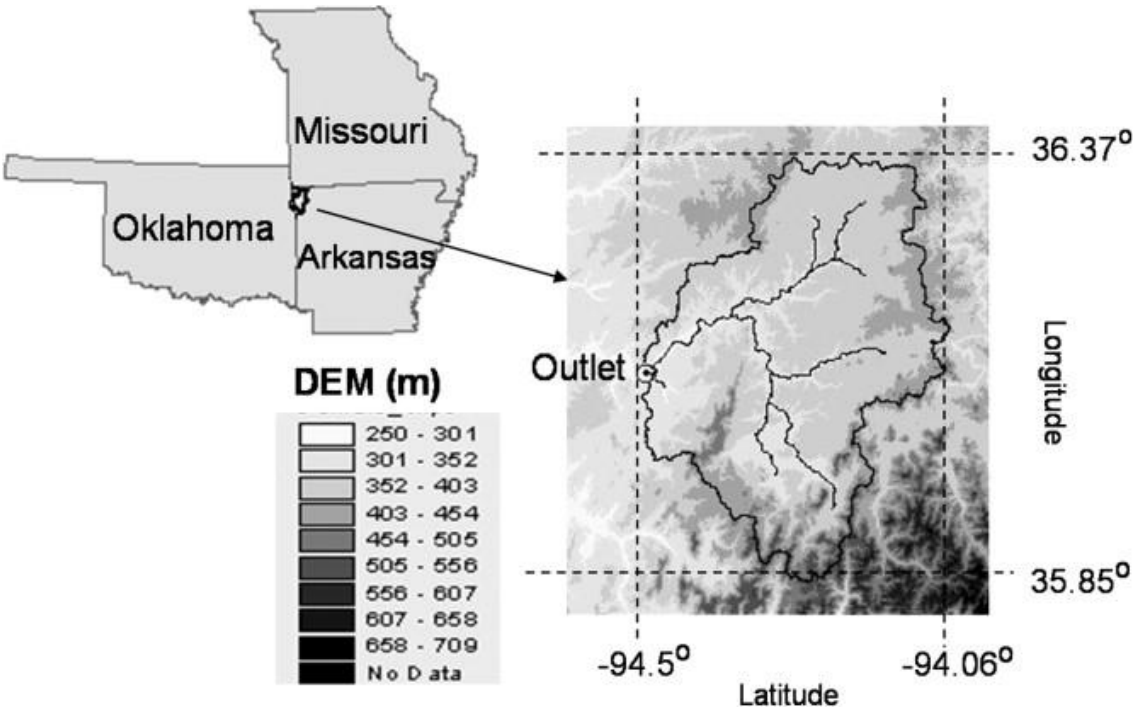

Fig. 6.1. The Illinois River basin south of Siloam Springs, AR (Behrangi *et al*., 2011).

6.2 Uncertainty analysis and results

The uncertainty of satellite precipitation over this watershed is analyzed in two forms. First, the hit precipitation was studied separately because this is the topic of most of the precipitation-uncertainty studies. Then, in the next section, the complete model is examined, considering also the missed precipitation and false alarm.

6.2.1 Uncertainty analysis for hit precipitation

Using the spatial and temporal resolutions and the rainfall rate, the uncertainty distributions are formed to model the hit precipitation. From each distribution, 1,000 samples are drawn randomly to estimate the uncertainty range of each satellite-precipitation point using a Monte Carlo approach.

For the period of November-April, the parameters are linearly interpolated from Figure 4.3 (winter) and, for the remainder of the year, the parameters are calculated from Figure 4.2 (summer). For winter, the parameters are: a = -0.7011, b = 1.7992, c = 0.0085, d = -0.9032, e = 7.7719, and f = 0.3979; for summer, the parameters are: a = 0.1867, b = 0.912, c = 0.0031, d = -0.9904, e = 3.1023, and f = 0.5446. The parameters are then incorporated into Eqs. (4.7), (4.8), and (4.9) to estimate the three parameters (shape, scale, and shift) of the uncertainty distribution. From each time step of the PERSIANN time series over the region, the three parameters of the uncertainty distribution are calculated based on the rainfall rate at the given time. This will result in the final form of the uncertainty distribution.

To evaluate the uncertainty model, the 90% uncertainty range is calculated using the CDF of the distribution. The rainfall rates corresponding to the 5^{th} and 95^{th} percentiles of the range are estimated and used as the lower and upper bounds of the uncertainty range, respectively. The mean of the satellite-precipitation uncertainty distribution is calculated using Eq. (8). The 5^{th} and 95^{th} percentiles of the uncertainty range are plotted for winter and summer in Figure 6.2, along with the scatterplot of the PERSIANN and Stage-IV radar rainfall over the watershed. Furthermore, the mean of the uncertainty model vs. the satellite-rainfall rate is plotted in black. The calculated results show that 68% of the summer reference-precipitation data and 70% of the winter reference precipitation data (blue-scattered dots) fall into the range, which means that the

calibrations for both seasons are efficient by covering the range of the variability in the data. In both seasons, the spread of the distributions increases with an increase in the rainfall rate. The summer season shows a more skewed distribution because the mean of the distribution is close to the lower bound; for winter, the mean is almost in the middle of the lower and upper bounds.

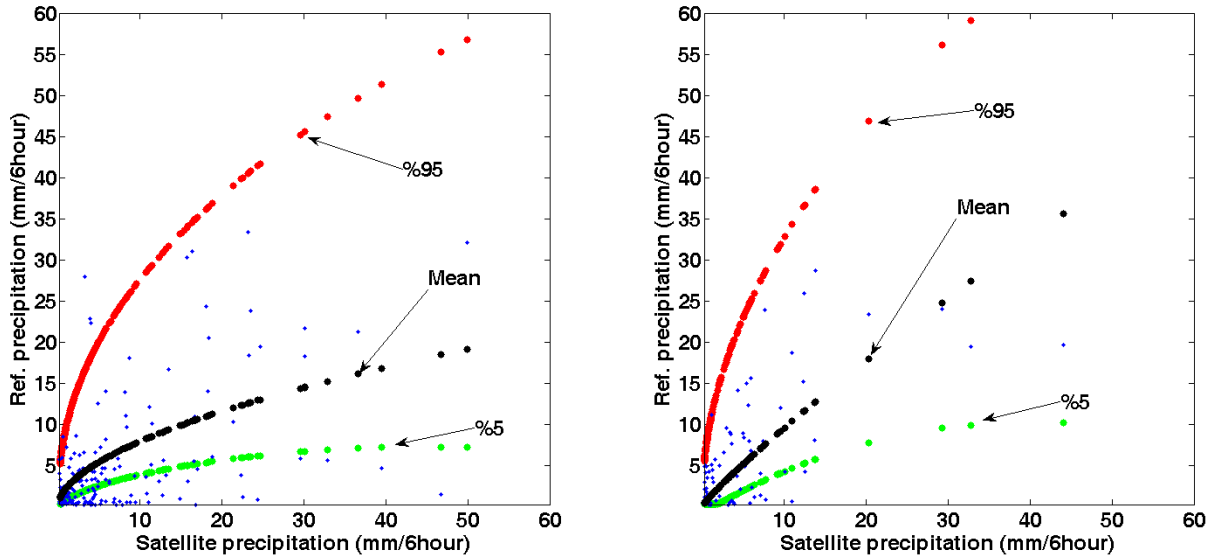

Fig. 6.2. Uncertainty range and the model mean for summer (right panel) and winter (left panel). The blue scatterplot shows the Stage-IV radar data (reference) vs. the PERSIANN rainfall data over the Illinois River basin south of Siloam Springs, AR, for 2006-2008. The green-scattered line represents the 5% of the uncertainty range, the red line gives the 95% of the uncertainty range, and the black line indicates the mean of the uncertainty distribution vs. the satellite-rainfall rate.

In Figure 6.3, the 6-hourly time series of the Stage-IV rainfall (top panel), PERSIANN (middle panel), and the mean of the uncertainty model (bottom panel) for 2006 are presented. The uncertainty model performs very well in reducing the overestimation from the PERSIANN product while keeping its pattern. In this work, the focus is to introduce a method to estimate the uncertainty range of the satellite-precipitation product. Specifically, the proposed model not only provides a range for the uncertainty, it also serves as a bias-correction method. In Table 1, the mean of the distribution is compared to the Stage-IV radar 6-hourly time series in terms of RMSE, %bias, and the correlation coefficient. If in each case, we call the Stage-IV radar as R_{rad} and the other timeseries as R_{sat}, the RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(R_{rad}(i)-R_{sat}(i))^2}{n}} \tag{6.1}$$

Similarly the %bias and the correlation coefficient are calculated as below:

$$\%bias = \frac{\sum_{i=1}^{n} R_{rad}(i)-R_{sat}(i)}{\sum_{i=1}^{n} R_{rad}(i)} \tag{6.2}$$

$$CORR = \frac{\sum_{i=1}^{n}(R_{rad}(i)-\bar{R}_{rad})(R_{sat}(i)-\bar{R}_{sat})}{\sqrt{\sum_{i=1}^{n}(R_{rad}(i)-\bar{R}_{rad})^2 \sum_{i=1}^{n}(R_{sat}(i)-\bar{R}_{sat})^2}} \tag{6.3}$$

The same statistics are calculated for the PERSIANN satellite estimates and radar data. By comparing the statistics for the summer and winter periods, respectively, as well as for the entire 3-year period, the mean of the uncertainty improves the satellite-precipitation estimates. In all three cases, the RMSE and %bias are improved. The improvement is more distinct in %bias when the two seasons are analyzed separately. Generally speaking, satellite-precipitation products have a larger bias in the summer (AghaKouchak *et al*., 2012), and the proposed method decreases this bias by 47%. For winter, the bias is decreased by 23%. The correlation coefficient remained the same in most of the cases because the mean is a function of satellite precipitation, and the transformation would not significantly change the correlation coefficient. In all three

cases for the respective summer and winter seasons, as well as the entire 3-year period, the RMSE of the satellite vs. radar rainfall improved. This improvement is 23% for the summer data, 6% for the winter data, and 18% for the entire 3-year period.

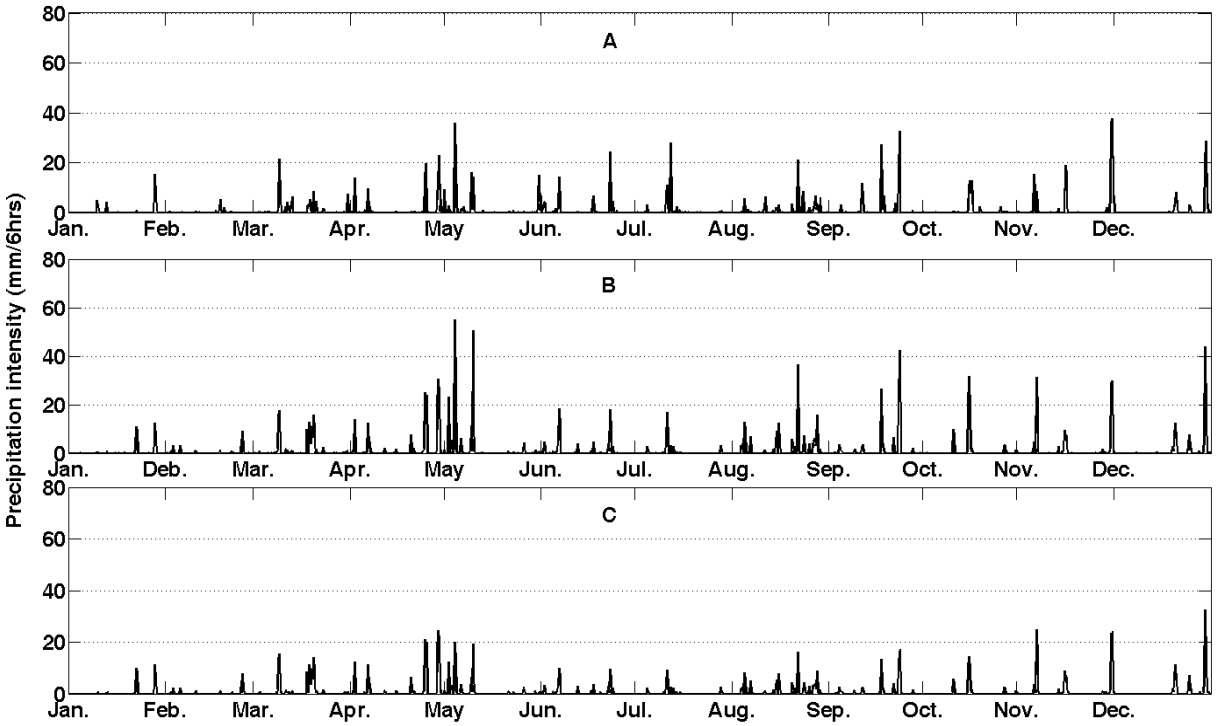

Fig 6.3. Six-hour, basin-average precipitation intensity for (a) Stage-IV radar data, (b) PERSIANN, and (c) mean of the uncertainty model over the Illinois River basin south of Siloam Springs, AR, for 2006.

TABLE 1. Evaluation of the mean of the GND uncertainty model compared to the original satellite product

Precipitation Comparison (mm/6hr)	**RMSE (mm/6hr)**	**Bias (%)**	**Correlation Coefficient**
Radar vs. satellite (PERSIANN) estimation	**12.89**	**24.6**	**0.70**
Radar vs. mean of uncertainty distribution	**10.47**	**16.01**	**0.70**
Summer Comparison			
Radar vs. satellite (PERSIANN) estimation	**14.15**	**-45.8**	**0.64**
Radar vs. mean of uncertainty distribution	**10.8**	**-23.9**	**0.64**
Winter Comparison			
Radar vs. satellite (PERSIANN) estimation	**7.22**	**4.2**	**0.7**
Radar vs. mean of uncertainty distribution	**6.77**	**3.22**	**0.7**

6.2.2 Uncertainty analysis for the complete model

To implement the complete model to our case study, we need to calculate all of the parameters of the different components of the model and combine them together as a complete model using the spatial and temporal resolutions of the satellite-precipitation time series and the rainfall rates.

At each time step: (1) If the satellite shows rainfall, then the uncertainty comes from Eq. (2.8), and (2) If the satellite shows no rainfall, then the uncertainty comes from Eq. (2.15), that is: If the time step shows rainfall, then the GND distribution is formed., Then, using the rainfall rate, and considering if the previous time step also shows rainfall or no rainfall the false alarm ratio is calculated from Eq. (4.14), and its parameters are interpolated from Figures 4.8-4.9 for summer and winter, respectively. Then by constructing the Eq. (2.8), 1000 samples is randomly drawn from the inverse of the function.

If the current time step shows no rainfall, then the previous time step is checked. If the previous time step also says no rainfall, then zero is assigned to the rainfall; if the previous time step shows rainfall, then the missed precipitation ratio is defined using the spatial and temporal resolutions from Figures 4.10-4.11 and also parameters of the Gamma distribution that the missing radar precipitation will be drawn from using the Gamma parameters drawn from Figures 4.14-4.15. Then by constructing the Eq. (2.15), 1000 samples are randomly drawn from the inverse of the function.

To demonstrate how well the model captures the uncertainty of the PERSIANN estimates over the case study watershed, the uncertainty model is generated for the 6-hourly, basin-averaged time series of precipitation over the Illinois River basin.

The temporal resolution of the data is 6-hourly, and the spatial resolution is approximately $0.35^\circ \times 0.35^\circ$ (each degree is considered to be about 111 km). Parameters of the GND model are the

same as mentioned in the previous section, where only the uncertainty of the hit precipitation was modeled. In order for Eq. (4.14) to estimate the false alarm ratio, for the summer data and the case where the satellite from the previous time step shows no precipitation, b = 0.6 and r = 0.035. When the satellite at the previous time step shows rainfall, b = 0.68 and r = 0.024. The same parameters for the winter season are, respectively: b = 0.77, r = 0.38, b = 0.58 and r = 0.0356. For the summer data, the missed-precipitation ratio is 0.13 and, for the winter data, it is 0.22. The parameters of the Gamma distribution for summer are: k = 0.94 and theta = 9.44; for winter, the parameters are: k = 1.01 and theta =1 1.04. Using the satellite-rainfall rate at each time step, the uncertainty distribution for that time step is quantified.

Similar to the evaluation of the GND model, here also for evaluation of the uncertainty model, the 90% uncertainty range (between 5%-95%) is calculated. The results are plotted for summer and winter separately in Figure 6.4, along with the scatterplot of the PERSIANN and Stage-IV radar rainfall over the watershed. In each panel of the figure, there are two sets of plots. One set is in a square shape, which shows the condition when the rainfall rate from the previous time step is not zero (in a darker shade of the colors). Also shown in Figure 6.4 are the circles for when the rainfall rate from the previous time step is zero (in a lighter shade of the colors). Furthermore, the mean of the uncertainty model vs. the satellite- rainfall rate is plotted in black. Calculations show that this 90% of the uncertainty in both seasons covers more than 85% of the data, which is an improvement compared to the results described in the previous chapter, where only the hit precipitation was corrected.

In Figure 6.5, the 6-hourly time series for the Stage-IV rainfall (top panel), PERSIANN (middle panel), and the mean of the uncertainty model (bottom panel) for 2006 are presented. The uncertainty distribution reduces the overestimation of the PERSIANN by reducing the %bias of

the hit precipitation and false alarm and adding the missed precipitation.

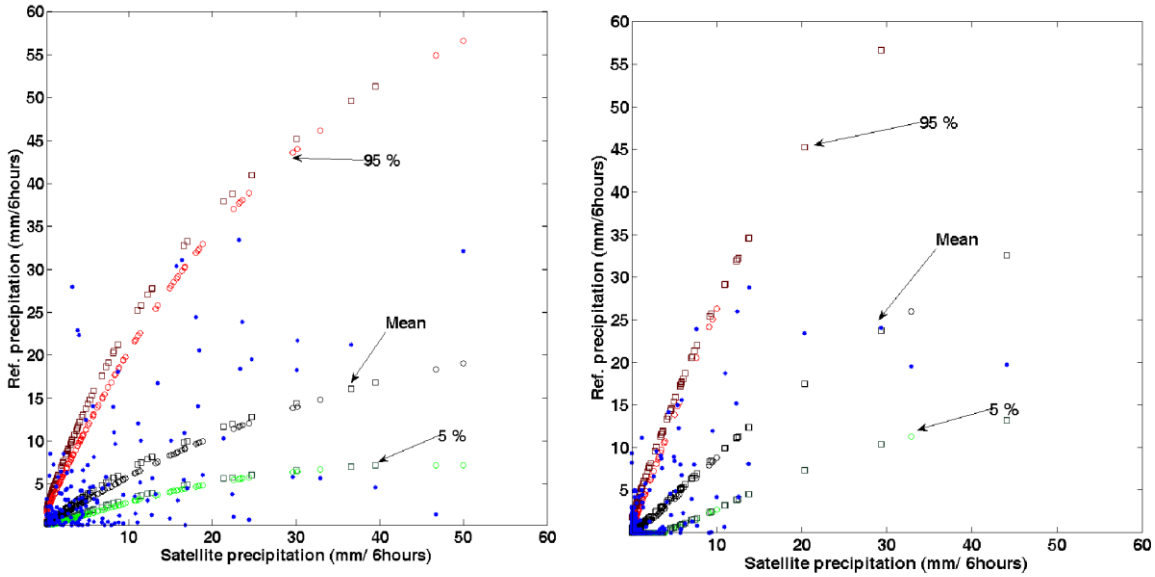

Fig. 6.4. Uncertainty range and the model mean for summer (right panel) and winter (left panel). The blue scatterplot shows the Stage-IV radar data (reference) vs. the PERSIANN rainfall data over the Illinois River basin south of Siloam Springs, AR, for 2006-2008. The green-scattered line represents the 5% of the uncertainty range, the red line gives the 95% of the uncertainty range, and the black line indicates the mean of the uncertainty distribution vs. the satellite-rainfall rate. The graphs in the square and in a darker shade of the colors represent the condition when the rainfall rate from the previous time step is not zero; the circles in a lighter shade of the colors indicate when the rainfall rate from the previous time step is zero.

We further compared the time series of the PERSIANN, Stage IV, and mean of the uncertainty model for 2006 using the correlation coefficient, RMSE, and %bias for 2006-2008 for summer and winter separately, as well as for the whole period. In terms of the mean of the distribution, the complete model improves the statistics slightly, because most of the modifications occur at the small rainfall-rate values (FA decreases exponentially with an increase in the rainfall rate, and the mean of the missed precipitation Gamma distribution is very small).

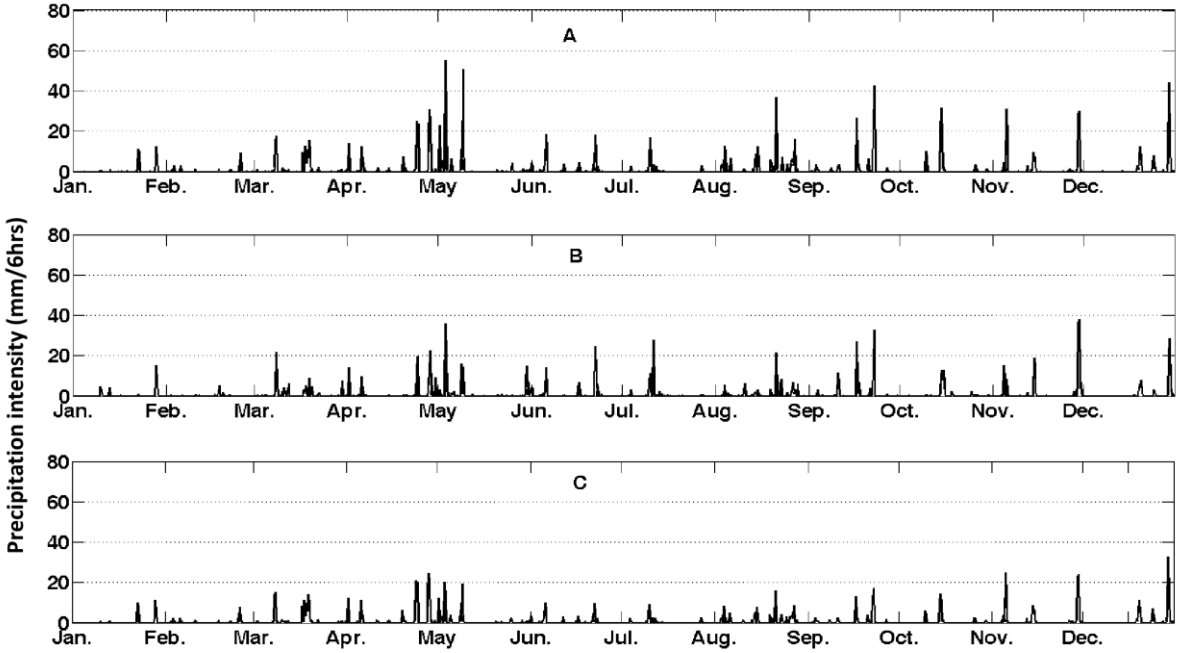

Fig 6.5. Six-hour, basin-average precipitation intensity for (a) Stage-IV radar data, (b) PERSIANN, and (c) mean of the uncertainty model over the Illinois River basin south of Siloam Springs, AR, for 2006.

TABLE 2. Evaluation of the mean of the uncertainty model compared to the original satellite product

Precipitation Comparison (mm/6hr)	**RMSE (mm/6hr)**	**Bias (%)**	**Correlation Coefficient**
Radar vs. satellite (PERSIANN) estimation	**12.89**	**24.6**	**0.70**
Radar vs. mean of uncertainty distribution	**10.2**	**20.01**	**0.70**
Summer Comparison			
Radar vs. satellite (PERSIANN) estimation	**14.15**	**-45.8**	**0.64**
Radar vs. mean of uncertainty distribution	**10.4**	**10.9**	**0.65**
Winter Comparison			
Radar vs. satellite (PERSIANN) estimation	**7.22**	**4.2**	**0.7**
Radar vs. mean of uncertainty distribution	**6.2**	**1.22**	**0.75**

Chapter 7: Application in Hydrologic Modeling

Development of high-resolution satellite-precipitation products that are available globally opened a new realm of possibility for hydrometeorological applications. Access to reliable precipitation data as the most important input for hydrological modeling and flood prediction is very crucial for scientists and the decision-makers because, in several populated countries around the world, e.g., developing countries, access to accurate ground-precipitation measurements is very limited or does not exist.

To ensure the reliability of hydrological modeling, the uncertainty inherent in the satellite-precipitation products should be taken into the account; hence, there are a large number of studies on the assessment of the uncertainty of these products and how this uncertainty is propagated into the hydrological modeling.

If the physics of the hydrology model and the calibration of its parameters are assumed to be perfect, the variability from the streamflow simulation can be attributed to their input data, e.g., precipitation.

In several studies, satellite-precipitation products are used for hydrological modeling (Hong *et al*., 2006; Yilmaz *et al*., 2005). In 1999, Tsintikidis *et al*. used the mean areal precipitation from satellites for hydrologic modeling over northern Africa. Grimes and Diop (2003) showed that, using the precipitation data from Meteosat improves streamflow prediction. The error of these remotely sensed precipitation data in streamflow simulation also has been investigated in several studies, e.g., Nijssen and Lettenmaier (2004) studied how the sampling error of precipitation estimates from satellites can affect the streamflow simulation, and Behrangi *et al*. (2011) evaluated the uncertainty in streamflow simulation from different satellite-based precipitation

products.

In this chapter, we investigate how the uncertainty in the PERSIANN product will result in the variability in streamflow simulation. To study this effect, the satellite-precipitation uncertainty model presented here is propagated through the SACramento Soil Moisture Accounting model, and the results are presented in the following sections.

7.1 Hydrologic model and calibration

The SACramento Soil Moisture Accounting (SAC-SMA) model (Burnash *et al*., 1973; Burnash, 1995) is used as the hydrology model to model the rainfall-ruffoff process. The SAC-SMA model (Figure 7.1) is a lumped conceptual model used at the National Weather Service River Forecasting System (NWSRFS) for hydrological modeling at the basin scale. The input to the model is the mean areal precipitation and potential evapotranspiration. This model has an upper zone representing the upper soil layer and a lower zone for the deeper layers of the soil profile, and each zone has a tension storage and a free-water storage. The deficiency of moisture in the lower zone and the free-water storage in the upper zone determine the percolation rate from the upper to the lower region. The model has 13 parameters, which should be calibrated to generate five response components, including: (1) direct runoff resulting from precipitation, (2) surface runoff when the precipitation rate is larger than the percolation rate, (3) interflow (the lateral outflow from the upper,-zone free-water storage), (4) supplementary baseflow, which is the lateral drainage from the lower-zone supplementary free-water storage, and (5) primary baseflow, which is the lateral drainage from the lower-zone primary free-water storage. The summation of these five runoffs is convolved with the unit hydrograph of the basin to result in the streamflow at the basin's outlet.

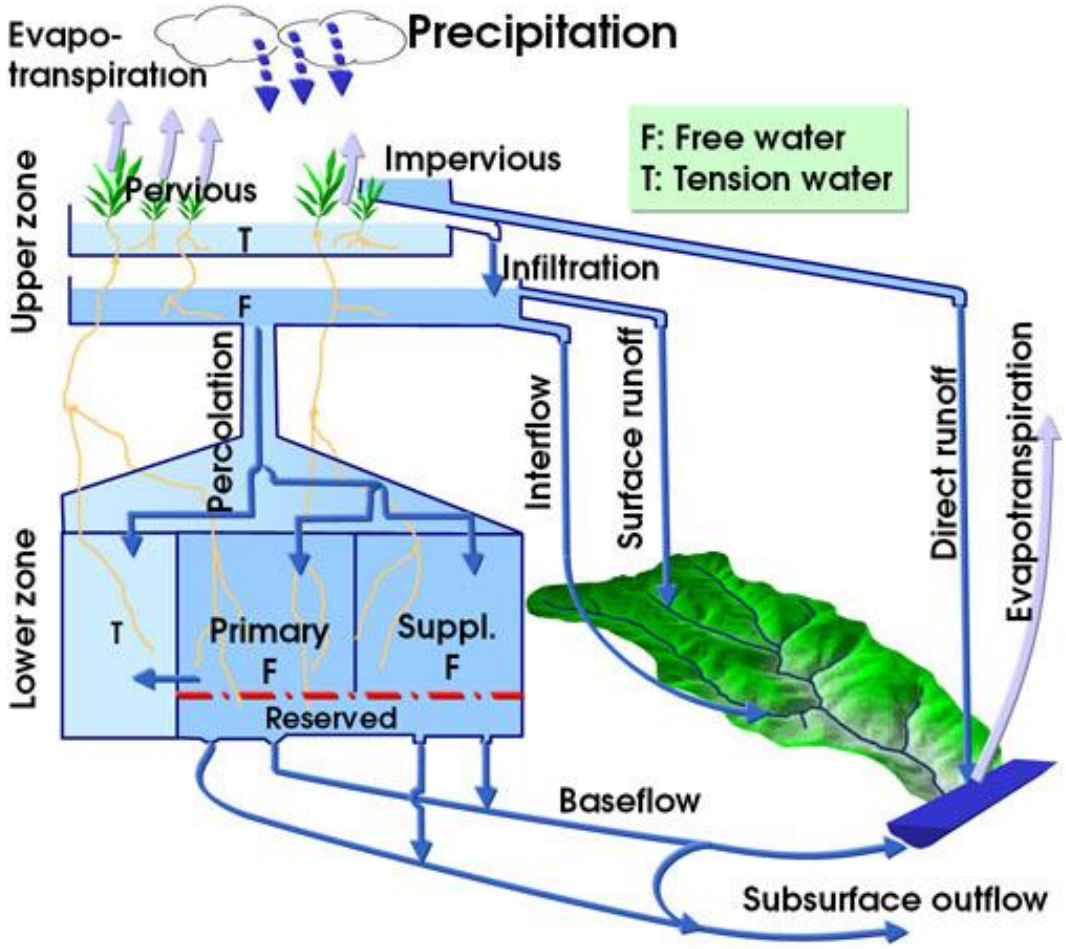

Fig. 7.1. Schematic view of the SAC-SMA model.

The SAC-SMA model should be calibrated for accuracy before streamflow simulation. In this study, the model is calibrated using the 6-hourly, Stage-IV precipitation data averaged over the basin. The monthly potential-evaporation rates and the unit hydrograph of the basin are obtained from the DMIP II study website. To calibrate the parameter, a global-search algorithm, termed the Shuffled Complex Evolution from University of Arizona (SCE-UA) (Duan *et al.*, 1992) is used to find the minimum value of the objective function below:

$$RMSE = (\sum(Q_{sim,t} - Q_{obs,t})^2)^{0.5} \quad (7.1)$$

where $Q_{sim,t}$ is the simulated, and $Q_{obs,t}$ is the observed streamflow at time step t.

This method is an effective and efficient search algorithm that combines the non-linear simplex method of Nelder and Mead (1965), a random-search procedure, and complex shuffling (Duan *et al*., 1992) to direct the parameter space towards the global optimal.

7.2 Uncertainty propagation and results

The hydrologic evaluation of the proposed uncertainty model is investigated two-fold. Similar to Chapter 6, only the uncertainty of the PERSIANN product is considered initially due to the bias in the hit precipitation in streamflow simulation. This part is modeled using a GND model. For the second part, the complete uncertainty model, considering all four parts of hit, false alarm, missed, and hit-zero precipitation is used to study the variability in the streamflow simulation due to the uncertainty in the satellite-precipitation product

The precipitation-uncertainty model is propagated in the hydrological modeling by drawing random time series from the uncertainty distributions to be used as the input to the hydrology model. The simulated streamflow from using the mean of the satellite-precipitation uncertainty model is compared to the mean from the original PERSIANN product.

In all of the streamflow simulations, the calibrated parameters from the Stage-IV radar data were used because that product is used as the reference in this study.

Figure 7.2 shows the simulated streamflow using the Stage-IV radar data and the simulated streamflow using the PERSIANN product as input to the SAC-SMA model. The difference between the two streamflows is due to the uncertainty in the PERSIANN product. The figure also displays how the overestimation in the PERSIANN product contributes to the very large

peaks and inability to capture the low flows.

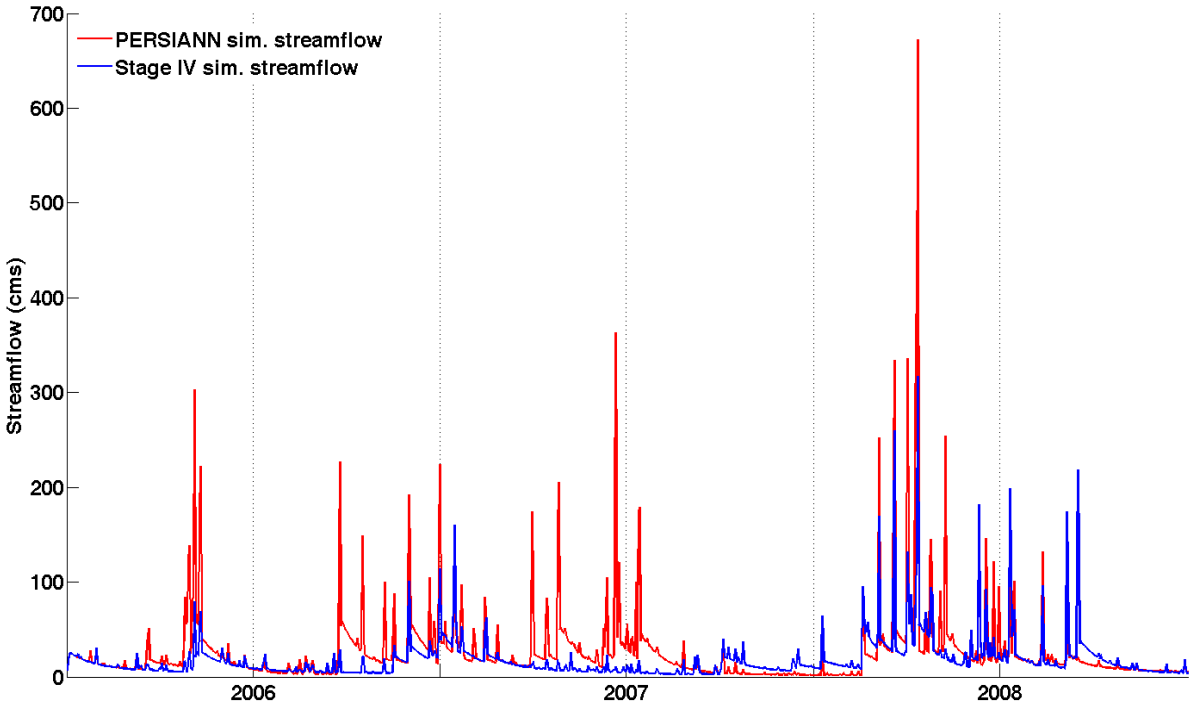

Fig. 7.2. Comparison of simulated streamflow between Stage-IV radar and PERSIANN satellite precipitation data.

7.2.1 Hydrologic evaluation of the satellite uncertainty model for hit precipitation

In order to estimate the streamflow variability due to the satellite-precipitation uncertainty, the GND uncertainty model must be propagated into the hydrologic model. This will result in a range where the streamflow simulation would change due to the precipitation uncertainty. To calculate this variability the following steps should be taken:

1. Given a precipitation time series with specific temporal and spatial resolutions of Δt and Δs respectively, for each point with a specific rainfall rate, there is an uncertainty

distribution associated with that precipitation estimate. First, we must determine if the time step (in this study, 6-hourly) falls in the summer (May-October) or winter (November-March). Then, from Figures 4.2-4.3, the parameters for estimating the GND parameters are obtained for Δt and Δs By using the rainfall rate at that time step, the parameters of the uncertainty distribution and, eventually the PDF and CDF of the distribution, become available.

2. From each uncertainty distribution at each time step, N (in this study N = 1,000) random rainfall rate is generated using the inverse of the distribution CDF, which will result in N-precipitation time series.
3. The N-precipitation time series are used separately as input to the SAC-SMA model, and the simulated streamflow is calculated, which results in N-streamflow time series.
4. The variability of the streamflow simulation is studied using the characteristics of the N-streamflow simulation at each output time step (in this study, daily).

Using the 1,000 precipitation samples as input for the hydrologic model will result in a range of streamflows at each time step. To quantify this variability, the 5% and 95% of each time step of the streamflow simulation are chosen for the 90% uncertainty range. This uncertainty range is compared to the simulated streamflow using the reference precipitation to investigate if the streamflow simulated from the uncertainty model can capture its variability.

Figure 7.3 shows the uncertainty ranges of 90% of the streamflow simulation from the GND model and how they compare to the Stage-IV simulated streamflow. For clarity, we only show the results from 2008. From this figure, we also see how the streamflow uncertainty model offers a range of uncertainty that can capture the variability of streamflow.

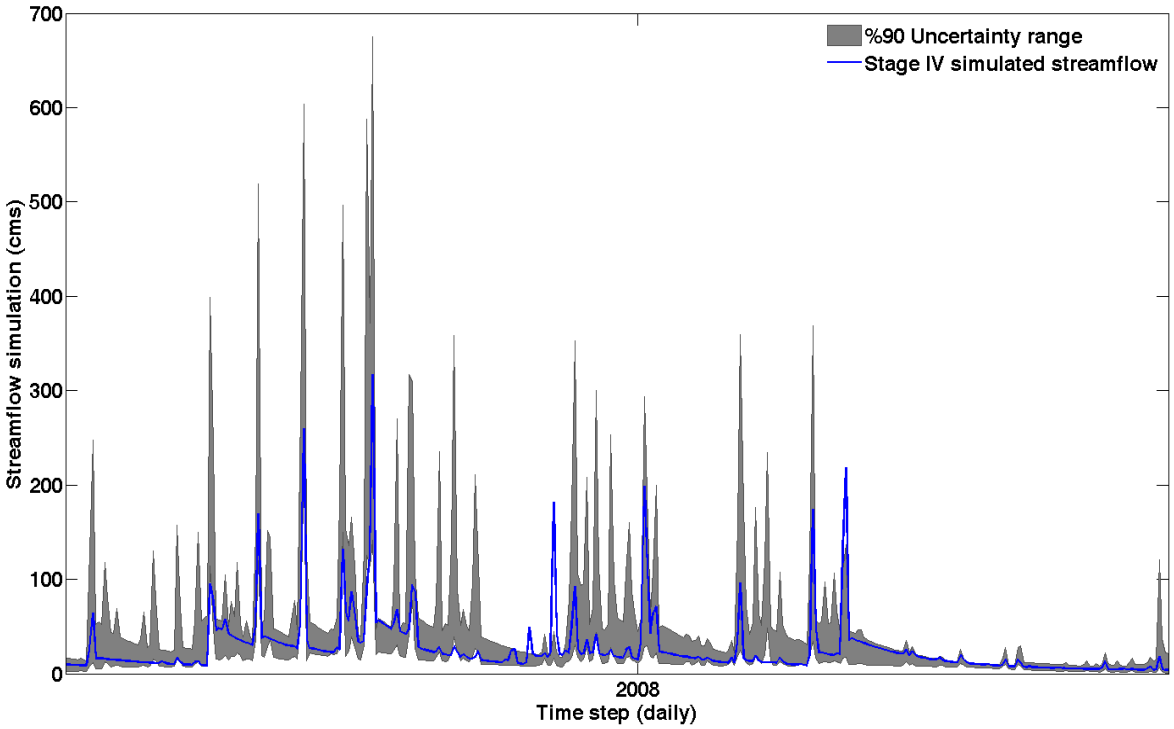

Fig. 7.3. Ninety percent of the uncertainty range of the streamflow simulation using the satellite- precipitation uncertainty model (grey-shaded areas) and the Stage-IV simulated streamflow (blue area) for 2008 over the case study watershed.

In Figure 7.4, the streamflow simulated from PERSIANN, Stage-IV, and the mean of the uncertainty model is presented. To enable a more clear comparison, we plotted the streamflows separately compared to the Stage-IV simulated streamflow (Figure 7.5). Further, the performance of the mean of the uncertainty distribution is compared to the original satellite- precipitation product in terms of correlation coefficient, RMSE, and %bias (Table 3). From both the plots and the statistics, the good performance of the model is very clear. The uncertainty model drastically improved the bias of the streamflow simulation. This improvement is 52% for the whole period and 60% and 88% for summer and winter and cannot change the correlation coefficient

significantly because the model is still based on the satellite precipitation and only changes its values. The RMSE also improved for the whole period, as well as for the summer and winter separately, with 20% for the first two and 10% for winter.

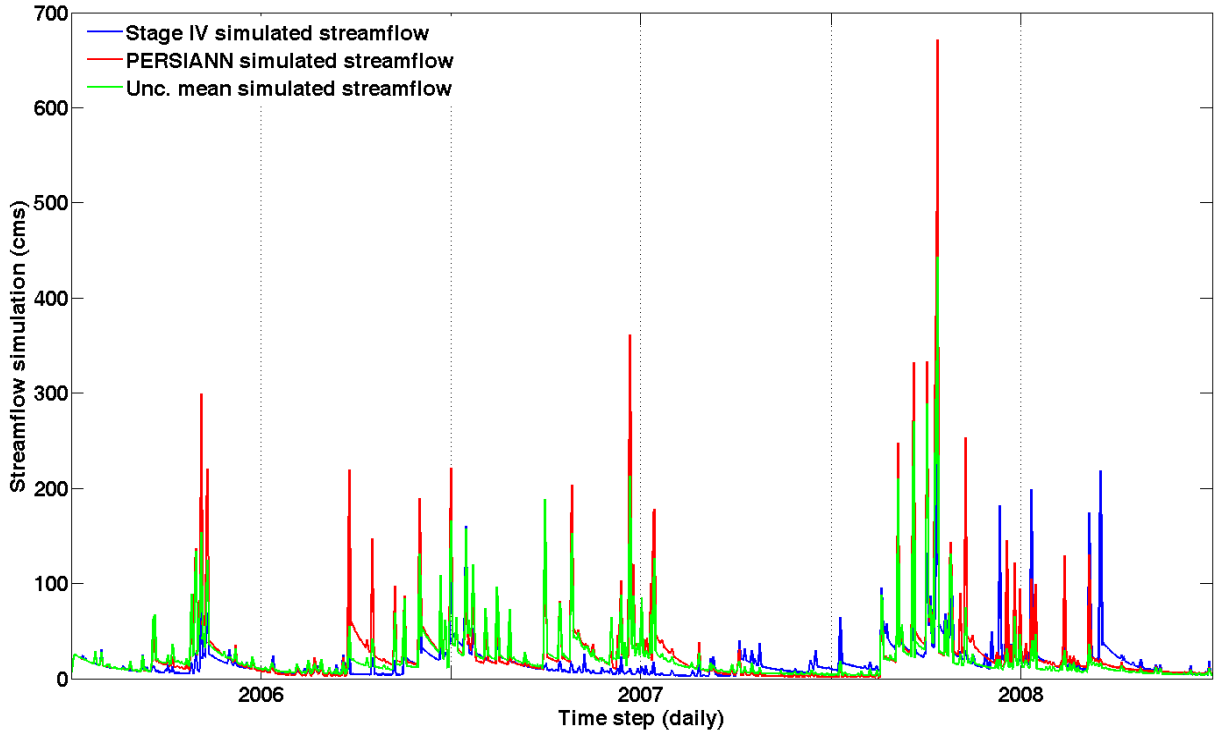

Fig. 7.4. Streamflow simulation from Stage-IV radar (blue), PERSIANN (red,) and the mean of the uncertainty model (green) for 2006-2008 over the case study watershed.

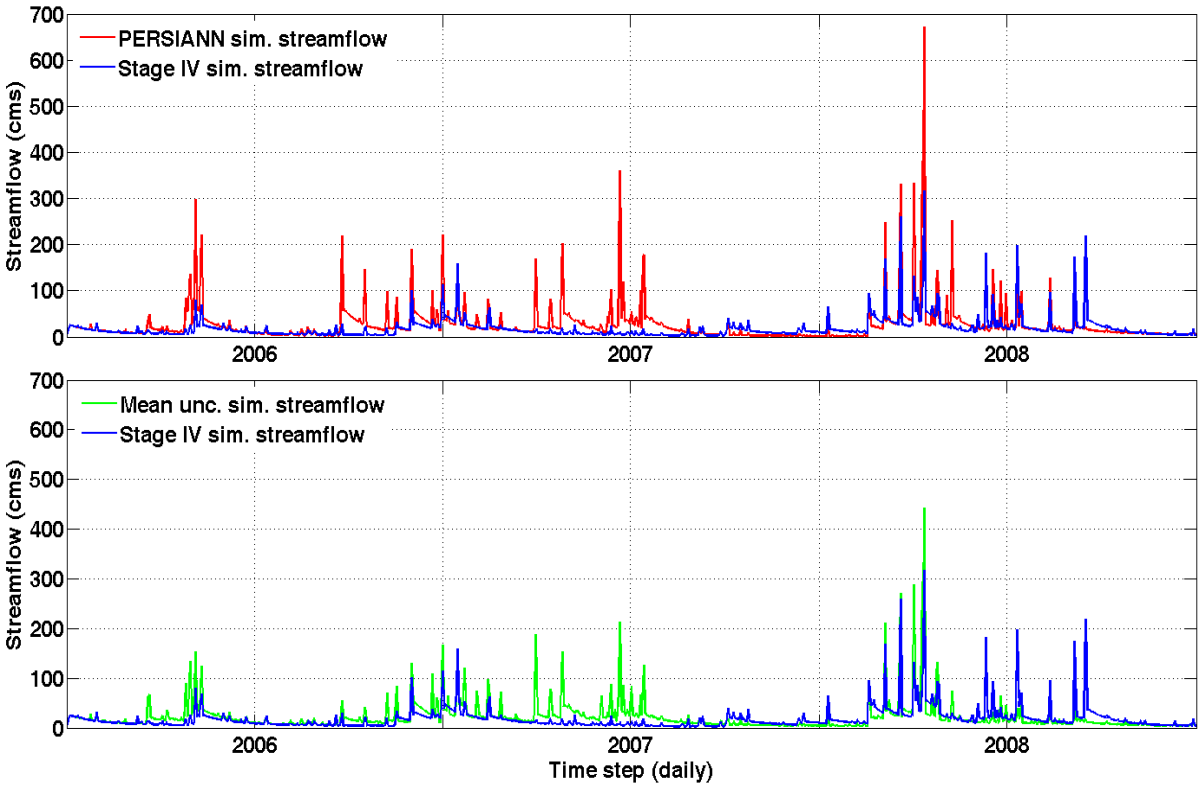

Fig. 7.5. Comparison of Stage-IV simulated streamflow with PERSIANN simulated streamflow (upper panel) and mean of the uncertainty model simulated streamflow (lower panel).

TABLE 3. Evaluation of the streamflow simulated from the mean of the GND uncertainty model compared to the streamflow simulated from the original PERSIANN product

Stream flow Comparison (cms)	**RMSE (cms)**	**Bias (%)**	**Correlation Coefficient**
Radar vs. satellite (PERSIANN) simulated stream flow	**33.1**	**-42.9**	**0.63**
Radar vs. mean of uncertainty distribution simulated stream flow	**23.7**	**-20.74**	**0.63**
Summer Comparison			
Radar vs. satellite (PERSIANN) simulated stream flow	**34.4**	**-69.9**	**0.27**
Radar vs. mean of uncertainty distribution simulated stream flow	**26.8**	**-28.9**	**0.32**
Winter Comparison			
Radar vs. satellite (PERSIANN) simulated stream flow	**15.19**	**5.9**	**0.82**
Radar vs. mean of uncertainty distribution simulated stream flow	**13.6**	**-0.65**	**0.83**

7.2.2 Hydrologic evaluation of the satellite uncertainty model for hit precipitation

As in the previous section, the same evaluation is performed but this time using the complete precipitation-uncertainty model. In this case, the random time series come from the complete-uncertainty model, considering the hit precipitation, false alarm, missed precipitation, and hit-zero precipitation.

Similar to Section 7.2.1, the 90% streamflow uncertainty range for 2008 is calculated and compared to the Stage-IV simulated streamflow provided in Figure 7.6.

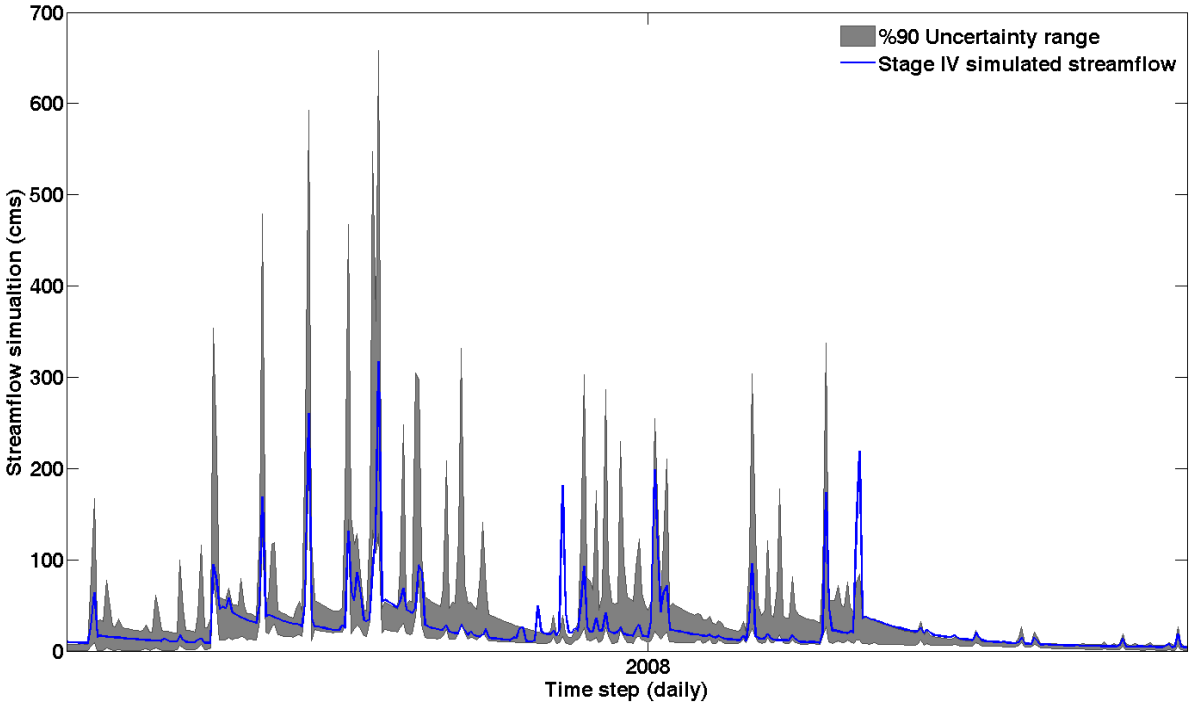

Fig. 7.6. Ninety percent of the uncertainty range of the streamflow simulation using the satellite- precipitation uncertainty model (grey-shaded areas) and Stage-IV simulated streamflow in (blue areas) for 2008 over the case study watershed.

Streamflow simulated using the mean of the uncertainty model is compared to the PERSIANN simulated streamflow and the Stage-IV simulated streamflow. The results are given in Figures 7.-7.8 and Table 4. Adding the false alarm and missed-precipitation probabilities helped produce a more complete model. A very important point to mention is that, in the modeling of the missed precipitation, adding the information from the previous time step is critical, especially in streamflow simulation. If all of the zero-precipitation time steps from the satellite estimates are assigned with the probable-missed precipitation values, non-zero precipitation at all of the time steps would remain, which is unrealistic. This also results in overestimation in streamflow simulation.

Also similar to the previous section, in this case, the correlation coefficient of the streamflow is slightly improved, but the improvement is more distinct in RMSE and %bias.

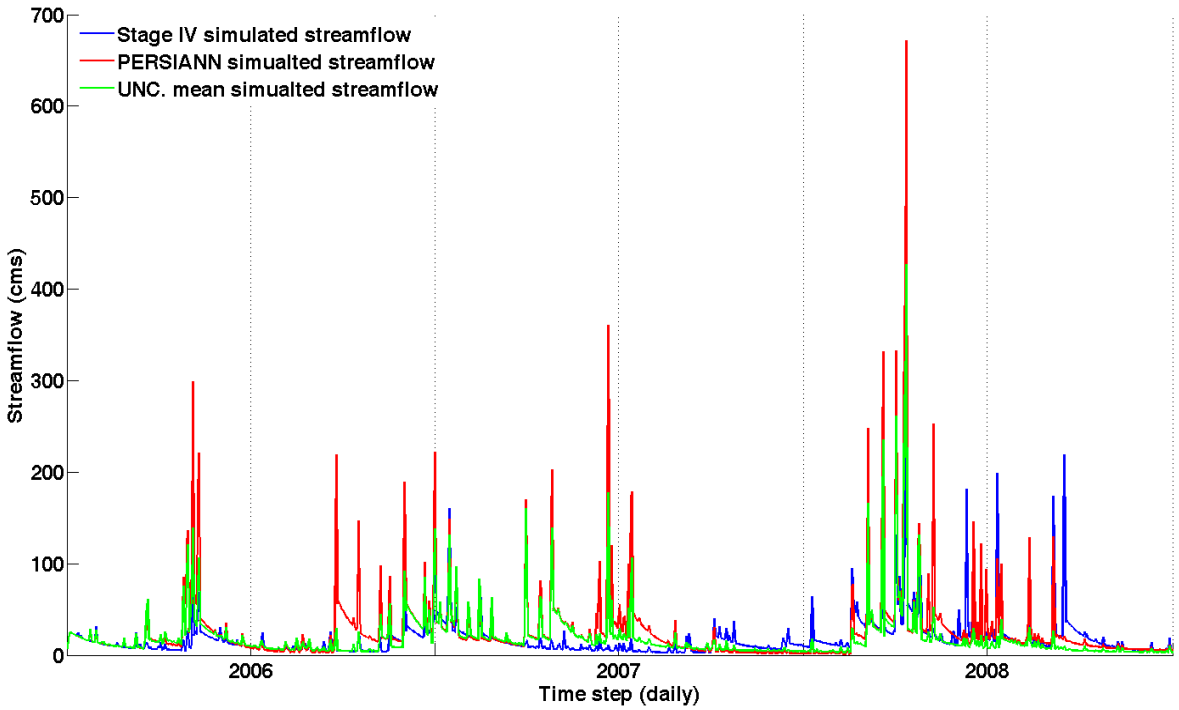

Fig. 7.7. Streamflow simulation from Stage-IV radar (blue), PERSIANN (red), and the mean of the uncertainty model (green) for 2006-2008 over the case study watershed.

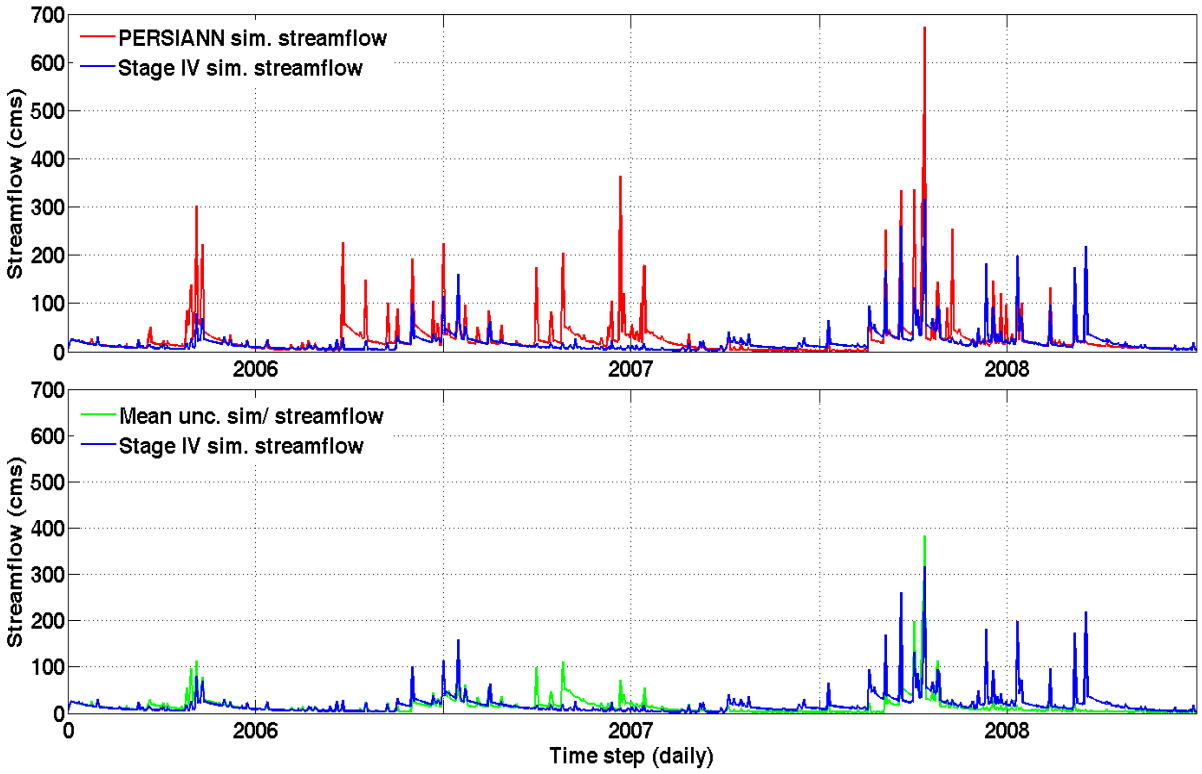

Fig. 7.8. Comparison of Stage-IV simulated streamflow with PERSIANN-simulated streamflow (upper panel) and mean of the uncertainty model-simulated streamflow (lower panel).

TABLE 4. Evaluation of the streamflow simulated from the mean of the uncertainty model compared to the streamflow simulated from the original PERSIANN product

Stream flow Comparison (cms)	**RMSE (cms)**	**Bias (%)**	**Correlation Coefficient**
Radar vs. satellite (PERSIANN) simulated stream flow	**33.1**	**-42.9**	**0.63**
Radar vs. mean of uncertainty distribution simulated stream flow	**21.5**	**-3.56**	**0.63**
Summer Comparison			
Radar vs. satellite (PERSIANN) simulated stream flow	**34.4**	**-69.9**	**0.27**
Radar vs. mean of uncertainty distribution simulated stream flow	**23.2**	**-2.16**	**0.31**
Winter Comparison			
Radar vs. satellite (PERSIANN) simulated stream flow	**15.19**	**5.9**	**0.82**
Radar vs. mean of uncertainty distribution simulated stream flow	**12.8**	**2.3**	**0.81**

Chapter 8: Uncertainty Analysis of the New GPM Precipitation Product Using the IMERG Algorithm

To extend this work, the model is calibrated and evaluated on NASA's Integrated Multi-Satellite Retrieval for GPM (IMERG) data product, which is a part of their Global Precipitation Mission (GPM). Because this product is very new and in development, it has limited availability. Here, we show some preliminary results of the uncertainty quantification of the product and are hoping that, in the future, we can extend this work when more data are available.

8.1 IMERG product

The GPM core observatory was launched in 2014 as collaboration between NASA and JAXA and consists of a core satellite and a network of 12 different satellites to measure precipitation on a global scale. IMERG combines the data from all 12 satellites into one uniform product. Some of these satellites are passive-microwave satellites, and some are geostationary.

The algorithm is as follows: different satellite PMW precipitations are combined into half-hourly resolutions and are used as input to both the CPC Morphing-Kalman Filter (CMORTH-KF; Joyce and Xie, 2011) Lagrangian time interpolation scheme and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS; Hong *et al*., 2004) re-calibration scheme. Parallel to that, CPC will correct the zenith angle and intercalibrate the geo-IR fields and send them for use in CMORPH-KF and PERSIANN-CCS. Then, PERSIANN-CCS estimates are computed and sent to the CMORPH-KF Lagrangian time interpolation to use the PMW and IR estimates to create the final product.

This product will be available at this level with 4-hour and 12-hour time latency. Then, the estimates are bias-adjusted using GPCC monthly gauge data, and the calibrated product is available with 2-month time latency. The resolution of the data is in $0.1°$ and half-hourly. The calibrated product is also available in monthly and $0.1°$ resolutions.

This product uses state-of-the-art technology and science in satellite-precipitation estimation and can revolutionize precipitation-data availability in terms of global coverage and real-time products.

8.2 Uncertainty analysis of the IMERG product

Because the real-time data with 18 hours of time latency have been available only recently (less than 10 days), for this study, we used the data with a 2-month time latency, which is available from April 2014-November 2014. Because the available data are very limited, the calibration is done using the months of June, July, and August and, for evaluation, the months of September, October and November are used. These data are available in both GPCC-calibrated and non-calibrated forms. Because the ultimate goal is to use the model for the real-time product when gage measurements are not available.

For consistency with the research that is already done in this area, the data were re-gridded from $0.1°$ to $0.25°$ and from half-hourly to 3-hourly resolutions. For reference data, Stage-IV radar data are used. The study domain is the same as the one used in the uncertainty analysis of PERSIANN product.

The results of the uncertainty analyses are presented in two different forms in Figures 8.1-8.2. The first figure compares the satellite precipitation and the mean of the uncertainty model with the Stage-IV radar data. This comparison is in the form of RMSE and %bias in the 3-hourly time

series for September, October, and November 2014.

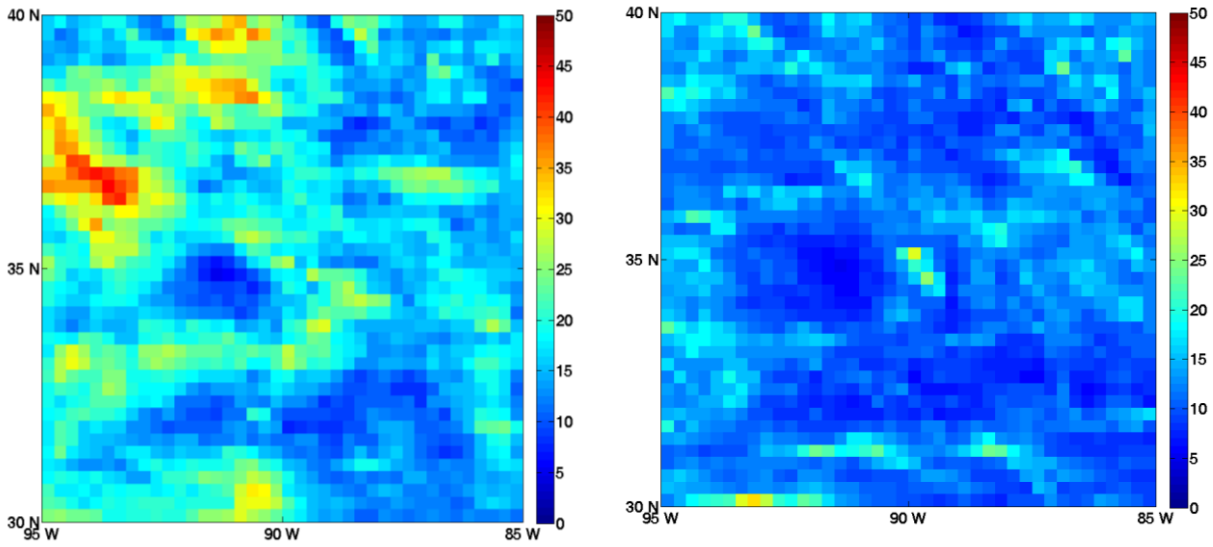

Fig. 8.1. Left panel: RMSE between IMERG and Stage-IV radar data; right panel: RMSE between the mean of the uncertainty distribution and the Stage-IV radar data for 3-hourly and 0.25 °time series for September, October, and November 2014.

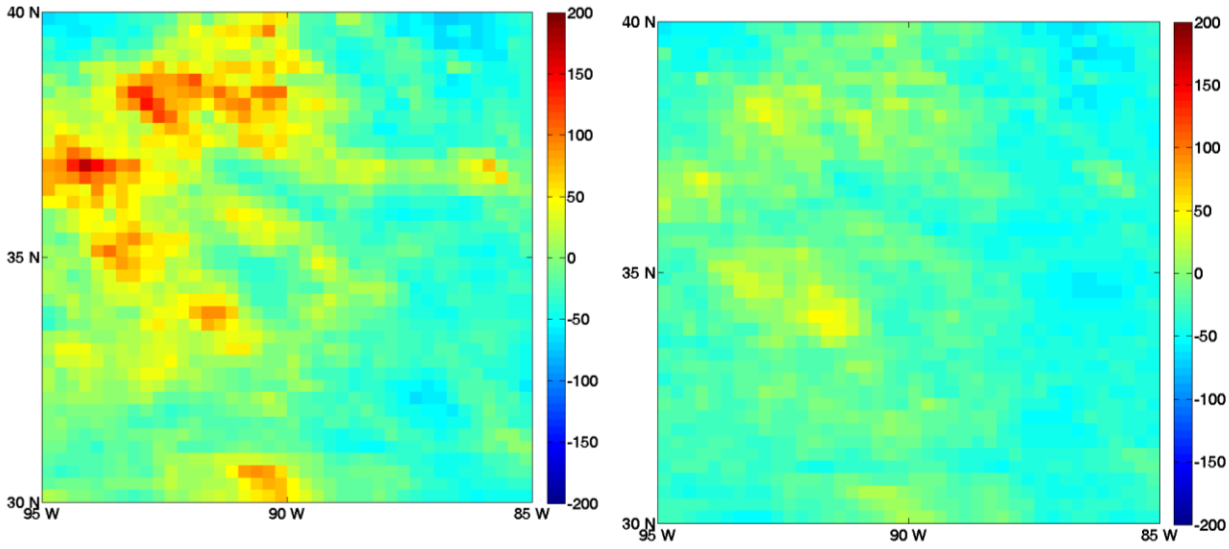

Fig. 8.2. Left panel: %bias between IMERG and Stage-IV radar data; right panel: %bias between the mean of the uncertainty distribution and the Stage-IV radar data for 3-hourly and 0.25 °time series for September, October, and November 2014.

From both figures, we can see that the mean of the uncertainty model can be used as a bias correction for the satellite precipitation because it reduces both the %bias and the RMSE of the satellite-precipitation estimates.

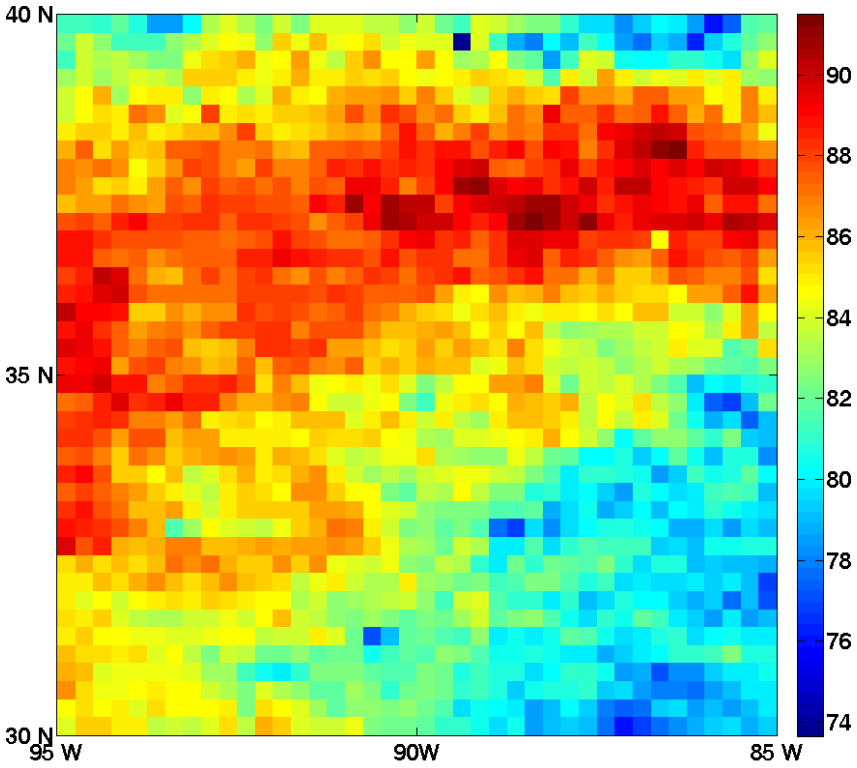

Fig. 8.3. Percentage of the time steps from for the 3-hourly and 0.25 °time series for September, October, and November 2014, which falls into the 90% uncertainty range of the model.

Figure 8.3 shows that more than 75% of the time steps during the evaluation period will fall into the 90% uncertainty range for the whole evaluation domain.

The IMERG product opens a new realm in satellite-precipitation products by producing global, near real-time precipitation data, and it is very important to evaluate the uncertainty of this product. The results so far show that the proposed method can measure this uncertainty and bias and adjust the product to a good extent.

Chapter 9: Summary, Conclusions, and Future Work

9.1 Summary and Conclusions

With the increasing emphasis on development and improvement of satellite-precipitation estimates for various applications, it is of paramount importance that the uncertainties associated with such products be evaluated and quantified carefully. This research aims to develop and validate an uncertainty model to answer this need. This model is in the form of a conditional probability distribution of the Stage-IV radar rainfall (used as the reference) given the satellite-precipitation product and for various spatial and temporal resolutions and rainfall rates. An application of this uncertainty model in hydrological modeling is tested over a mid-sized basin by propagating the uncertainty distribution into the SAC-SMA model to investigate the variability in the simulated streamflow resulting from the satellite-precipitation data.

The main conclusions drawn from the testing and comparison studies reported in this dissertation are as follows:

1. A generalized-uncertainty model is developed to model the probability of the radar precipitation as the reference given the satellite-precipitation product. This model consists of four different parts for modeling the hit precipitation: false alarm, missed precipitation, and hit-zero precipitation in the form of a mixed conditional-probability distribution.
2. The Generalized Normal Distribution is a skewed version of normal distribution, which is able to better model the hit-precipitation uncertainty. The model is evaluated in terms of the goodness-of-fit compared to the normal, Gamma, log-normal, and Weibull distributions. It is shown that, over different resolutions, this model always produces a

better fit.

3. The parameters of the uncertainty model are calibrated so that they cover a range of spatial and temporal resolutions, which will remove the confusion of probability distribution choice when dealing with the uncertainty of satellite-precipitation products at different resolutions.
4. The uncertainty model is evaluated over a $10° \times 10°$ region in the southeastern United States for 2010 and is shown in two different forms:

    (a) First, it is assumed that the uncertainty of the satellite-precipitation product is only in the form of the bias of the hit precipitation and that the GND model is evaluated. The results show that the uncertainty range of 90% covers more than 75% of the pixels, and the uncertainty range of 80% covers more than 65% of the pixels. This result supports the suggestion that the model can simulate the range of uncertainty to a good degree.

    (b) The complete uncertainty model is then evaluated considering all four components of the uncertainty. The results show that the uncertainty range of 90% covers more than 85% of the pixels, and the uncertainty range of 80% covers more than 80% of the pixels. Furthermore, this result indicates the importance of the analysis of the false alarm and missed precipitation in the study of the uncertainty of satellite-precipitation products.
5. The uncertainty model is evaluated over the Illinois River watershed, south of Siloam Springs, AR. The precipitation data are in 6-hourly and $0.35°$ resolution. The results show that the uncertainty model reduced the %bias of the precipitation estimation by 18% and the RMSE by 20%.

6. In addition, the effect of the uncertainty of the satellite-precipitation estimation is evaluated in hydrological modeling. The proposed uncertainty model can reduce the %bias in streamflow simulation by 90% and the RMSE by 35%.

The general framework of the proposed uncertainty model allows for its application to other satellite-precipitation products with the proper calibration of the model parameters, as was done for the PERSIANN data.

To extend the model to other satellite-precipitation products, the model is calibrated for the IMERG product, and the preliminary results show improvement in the uncertainty analysis of this new precipitation product.

9.2 Future Work

Although the initial development, validation, and implementation of this precipitation-uncertainty model are completed in this research, there is still an ongoing effort to further test the applicability of this model in different regions, climates, and applications.

1. In the future, the model should be tested on different basins with different climatology to investigate how different components of the uncertainty change in response to changes in the climate.

2. The emphasis should be on investigating the limitation of the model calibration while reference data are limited because there are several regions of the world where a complete network of ground observation or precipitation data is not available.

3. The effect of the seasonality of the model should be evaluated in regions with different seasons than the study area, e.g., tropical regions, the Southern Hemisphere.

4. The uncertainty model should be further calibrated and evaluated using satellite-precipitation

data other than PERSIANN. As shown in Chapter 8, the model is calibrated for the IMERG product with limited data availability. The evaluation of the model for the IMERG product should be investigated when more data become available in the near future.

REFERENCES

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States, *J. Geophys. Res*., **116**, D02115, doi:10.1029/2010JD014741.

——, A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets, Geophys. Res. Lett., 39, L09406, doi:10.1029/2012GL051592.

Anagnostou, E. N., A .J. Negri, and R. F. Adler, 1999: A satellite infrared technique for diurnal rainfall variability studies, *J. Geophys. Res*., **104(D24)**, 31477–31488, doi:10.1029/1999JD900157.

Anagnostou, E. N., V. Maggioni, E. I. Nikolopoulos, T. Meskele, F. Hossain, and A. Papadopoulos, 2010: Benchmarking highresolution global satellite rainfall products to radar and rain gauge rainfall estimates. *IEEE Trans. Geosci. Remote Sens*., **48**, 1667–1683, doi:10.1109/TGRS.2009.2034736.

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 earth orbit satellite instruments because of their incomplete temporal and spatial coverage, Surv. Geophys, 18, 385-403, doi:10.1023/A:1006512715662.

Austin, P. M., 1987: Relation between measured radar reflectivity and surface rainfall, *Mon. Wea. Rev.*, **115**, 1053-1070, doi:10.1175/1520-0493(1987)115<1053:RBMRRA>2.0.CO;2.

Behrangi, A., B. Khakbaz, T.C. Jaw, A. Aghakouchak, and K-.-L. Hsu, 2011: Hydrology evaluation of satellite precipitation products over a mid-size basin, *J. Hydrol*., **397**, 225-237, doi:10.1016/j.jhydrol.2010.11.043.

Bell, T. L., and P. K. Kundu, 2000: Dependence of satellite sampling error on monthly averaged rain rates: Comparison of simple models and recent studies, J. Climate, 13, 449-462, doi:10.1175/1520-0442(2000)0132.0.CO;2.

Burnash, R. J. C., R. L. Ferral, r. A. McQuire, 1973: A generalized streamflow simulation system, Conceptual modeling for Digital Computers, U.S. National Weather Service, Sacramento, Calif..P. 204.

Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance, *Adv. Water Resour.*, **22**, 585-595, doi:10.1016/S0309-1708(98)00043-8.

Duan, Q., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res*., **28(4)**, 1015–1031, doi:10.1029/91WR02985.

——, V. Gupta and S. Sorooshian, 1993: Shuffled complex evolution approach for effective and efficient global minimization, *J. Optimization Theory and Appl*, **76(3)**, 501-521, doi: 10.1007/BF00939380.

Gebregiorgis, A. and F. Hossain, 2011: How Much Can A Priori Hydrologic Model Predictability Help in Optimal Merging of Satellite Precipitation Products?. *J. Hydrometeor*, **12**, 1287–1298,doi: http://dx.doi.org/10.1175/JHM-D-10-05023.1

Gebremichael, M.,and W. F. Krajewski, 2004: Characterization of the temporal sampling error in space-time-averaged rainfall estimates from satellites, *J. Geophys. Res.*, **109**, D11110, doi: 10.1029/2004JD004509.

——, W. F. Krajewski, 2005: Modeling distribution of temporal sampling errors in area-time-averaged rainfall estimates, *Atmos. Res.*, **73**, 243-259, doi:10.1016/j.atmosres.2004.11.004.

Gupta, V., S. Sorooshian, 1983: Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined, *Water Resour. Res.*, **19(1)**, 269-276, doi: 10.1029/WR019i001p00269.

Hogue, T. S., S. Sorooshian, H. Gupta, 2000: A Multistep Automatic Calibration Scheme for River Forecasting Models, J. Hydrometeor, 1, 524-542, doi: 10.1175/1525-7541(2000)001.

Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system, *J. Appl. Meteor.*, **43**, 1834–1853.

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

Hossain, F., E. N. Anagnostou, 2004: Assessment of current passive -microwave and infrared-based satellite rainfall remote sensing for flood prediction, *J. Geophys. Res.*, **109**, D07102, doi:10.1029/2003JD003986.

——, and ——, 2006: Assessment of a multi-dimensional satellite rainfall errormodel for ensemble generation of satellite rainfall data. *Geosci. Remote Sens. Lett.*, **3**, 419–423, doi:10.1109/LGRS.2006.873686.

Hsu, K-.-L., X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteor*., **36**, 1176-1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

Huffman, G. J., 1997: Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteor*., **36**, 1191-1201, doi: 10.1175/1520-045(1997)036<1191: EORMSR>2.0.CO; 2.

——, 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, doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.

——, and Co-authors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, *J. Hydrometeor*., 8, 38-55, doi:10.1175/JHM560.1 .

Joyce, R. J., J. E. Janowiak, P. A. Arkin, 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, doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

——, P. Xie, 2011: Kalman filter-base CMORPH, *J. Hydrometeor*, **12**, 1547-1563, doi: 10.1175/JHM-D-11-022.1.

Khakbaz, B., Hsu, K., Imam, B., Sorooshian, S., 2009: From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. *J. Hydrol*., **418-419**, 61-77, doi: 10.1016/j.jhydrol.2009.02.021.

Krajewski, W. F., V. Lakshmi, K. P. Georgakakos, S. C. Jain, 1991: A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resour. Res.,* **27 (1)**, 119-128, doi:10.1029/90WR01977.

——, and J. A. Smith, 2002: Radar hydrology: Rainfall estimation, *Adv. Water Resour.*, **25**, 1387-1394, doi:10.1016/S0309-1708(02)00062-3.

Kunsch, H. R., 1989: The jackknife and the bootstrap for general stationary observation, Ann. Stat., 17, 1217-1241, doi:10.1214/aos/1176347265.

Laughlin, C. R., 1981: On the effect of temporal sampling on the observation of mean rainfall, *Precipitation Measurements from space*, Workshop report, Atlas, D. & O. Thielle, Eds., NASA Publication, D-59-D-66.

Li, Q., R. Ferraro, and N. C. Grody, 1998: Detailed analysis of the error associated with the rainfall retrieved by the NOAA/NESDIS SSM/I Rainfall Algorithm: Part I. Tropical oceanic rainfall, *J. Geophys. Res.*, **103**, 11 419-11 427, doi:10.1029/2001JD001172.

Lin, Y. and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Pre-prints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at http://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.].

Liu, H., S. Sorooshian & X. Gao, 2015, Assessment of the Spatial and Seasonal Variation of Error-Intensity Relationship in Satellite Precipitation using an Adaptive Parametric Model. *J. Hydrometeor,*.In press.

Maggioni, V., R. P. M. Sapiano, R. F. Adler, Y. Tian, G. J. Huffman, 2014: An error model for uncertainty quantification in high-time resolution precipitation products, *J. Hydrometeor.*, **15(3)**, 1274–1292, doi:10.1175/JHM-D-13-0112.1.

Nasrollahi, N., 2015: Improving infrared-based precipitation retrieval algorithm using multi-spectral satellite imagery, Springer these, doi: 10.1007/978-3-319-12081-2_2.

Nelder, J. A., R. Mean, 1965: A simplex method for function minimization, *Computational J.*, **7**, 308-313, doi:10.1098/comjnl/7.4.308.

Oki, R., A Sumi, 1994: Sampling simulation of TRMM rainfall estimation using RadaAMeDAS composites, *J. Appl. Meteor.*, **33**, 1597–1608, doi:10.1175/1520-0450(1994)033<1597:SSOTRE>2.0.CO;2.

K. Okamoto, T. Iguchi, N. Takahashi, K. Iwanami and T. Ushio, 2005: The Global Satellite Mapping of Precipitation (GSMaP) project, 25th IGARSS Proceedings, pp. 3414-3416.

Rinehart, R. E., 2004: Radar for Meteorologists, 4th ed. Rinehard, 484 pp.

Sapiano, M. R. P., and P. A. Arkin, 2009: An inter-comparison and validation of high resolution satellite precipitation estimates with three-hourly gauge data. *J. Hydrometeor.*, **10**, 149-166, doi: 10.1175/2008JHM1052.1.

Scofield, R. A., 1987: The NESDIS operational convective precipitation estimation technique, Mon. Wea. Rev., 115, 1773-1792.

Scofield, R. A., and R. J. Kuligowski, 2003: Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. Mon. Wea. Rev., 18, 1037-1051.

Seed, A. W., G. L. Austin, 1990: Sampling errors for raingauge-derived mean areal daily and monthly rainfall, J. Hydrol., 118(1-4), 163-173, doi:10.1016/0022-1694(90)90256-W.

Smalley, M., T. L'Ecuyer, M. Lebsock, and J. Haynes, 2014: A comparison of precipitation occurrence from the NCEP Stage IV QPE product and the CloudSat Cloud Profiling Radar, *J. Hydrometeor*, **15**, 444–458, doi: http://dx.doi.org/10.1175/JHM-D-13-048.1.

Smith, M. B., D. Seo, V. I. Koren, S. M. Reed, Z. Zhang, Q. Duan, F. Moreda, S. Cong, 2004: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol*., **298 (1–4)**, 4–26, doi:10.1016/j.jhydrol.2004.03.040.

Somen, V. V., J. B. Valdes, and G. R. North, 1995: Satellite sampling and the diurnal cycle statistics of Darwin rainfall data, *J., Appl. Meteor*., **34**, 2481-2490, doi:10.1175/1520-0450(1995)0342.0.CO;2.

——, and J., B., Valdes, 1996: Estimation of sampling errors and scale parameters using two- and three-dimensional rainfall data analyses, *J. Geophys. Res*., **101D**, 26453-26460, doi:10.1029/96JD01387.

Sorooshian, S., K-.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Amer. Meteor. Soc*., **81**, 2035- 2046, doi:10.1175/1520-0477081.

Steiner M., 1996: Uncertainty of estimates of monthly areal rainfall for temporally sparse remote observations, *Water Resour. Res.*, **32(2)**, 373 – 388, doi:10.1029/95WR03396.

——, J. A. Smith, S. J. Burges, C. V. Alonso, and R. W. Darden, 1999: Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation, *Water Resour. Res.*, **35**, 2487-2503, doi:10.1029/1999WR900142.

——, 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 large radar dataset, *J. Climate*, **16**, 3759-3778, doi:10.1175/1520-0442(2003)016<3759:COTMFE>2.0.CO;2.

Tian, Y., and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates, *J. Geophys. Res.*, **114**, D24101, doi:10.1029/2009JD011949.

Villarini, G., and W. F. Krajewski, 2007: Evaluation of the research-version TMPA three-hourly 0.25 °× 0.25 ° rainfall estimates over Oklahoma, *Geophys. Res. Lett*., **34**, L05402, doi: 10.1029/2006GL029147.

——, and W. F. Krajewski, 2008: Empirically based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauge, *Adv. Water Resour.*, **31**, 1015-1023, doi:10.1016/jadvwaters.2008.04.007.

——, P. V. Mandapaka, W.F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective, *Geophys. Res*., **113**, D11102, doi:10.1029/2007JD009214.

——, and W. F. Krajewski, 2009: New paradigm for statistical validation of satellite precipitation estimates: Application to a large sample of the TMPA 0.25 ° 3-hourly estimates over Oklahoma, *J. Geophys. Res*., **114**, D12106, doi:10.1029/2008JD011475.

Yan, J., and M. Gebremichael, 2009: Estimating actual rainfall from satellite rainfall products, *Atmos. Res*., **92**, 481-488, doi:10.1016/j.atmosres2009.02.004

Weng, F., R. R. Ferrero, and N. C. Grody, 1994: Global precipitation estimation using defense Meteorological Satellite program F10 and F11 special sensor microwave imager data, *J. Geophys. Res*., **99D**, 14493-14502, doi: 10.1029/94JD00961.

Westcoot, N. E., H. V. Knapp, S. D. Hilberg, 2008: Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United states, *J. Hydrol*., **351(1-2)**, 1-12, doi:10.1016/j.jhydrol.2007.10.057.

Xie, P., J. E. Janowiak, P. A. Arkin, R. F. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate*, **16**, 2197–2214, doi: http://dx.doi.org/10.1175/2769.1.

Yilmaz, K. K., T. S. Hogue, K-.-L. Hsu, S. Sorooshian, H. V. Gupta, T. Wagener, 2005: Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting, *J. Hydrometeor.* , **6**, 497–517, doi:10.1175/JHM431.1.