3-20-2014

# Performance Assessment of Satellite Rainfall Products for Hydrologic Modeling

Hojjat Seyyedi
seyyedi@engr.uconn.edu

# Performance Assessment of Satellite Rainfall Products for Hydrologic Modeling

Hojjat Seyyedi, Ph.D.

University of Connecticut, 2014

The overarching goal of the research described in this study is to improve uses of satellite rainfall in hydrological applications. To address this goal, the study followed the following three research dimensions. First, we improved the characterization of rainfall retrieval uncertainty from instantaneous earth-orbiting platform passive microwave observations, which form the basis of the merged satellite products, by investigating the significance of conditioning an error model for NASA's Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI) rainfall algorithm (2A12) to near-surface soil moisture data derived from a land surface model. A two-dimensional satellite rainfall error model, SREM2D, was used to provide ensemble error representation of 2A12 rainfall using two different parameter calibration approaches: conditioning vs. not the SREM2D parameters to the surface soil wetness categories. The statistical analysis of model generated ensembles and associated error metrics showed better performance when surface wetness information is used in SREM2D.

The second research dimension was to investigate the hydrologic applicability of a quasi-global high-resolution satellite precipitation product relative to a global reanalysis product through comparison with rain gauge adjusted radar-rainfall estimates. The study presents error metrics for moderate and extreme precipitation events over the Susquehanna River Basin in the Northeast United States. Results show improvements of the statistical scores up to 50% with increasing basin size, particularly for the satellite product. Overall, the satellite product exhibits

better error statistics compared to the coarse resolution re-analysis product. For the simulated streamflow, the re-analysis precipitation product is shown to have up to seven times higher mean relative error compared to the corresponding high resolution satellite product for moderate streamflow values; three times higher for extreme streamflows. This significant divergence in the runoff simulation error statistics is attributed to differences between the two precipitation products in terms of the propagation of their error properties from precipitation to simulated streamflow.

The last research dimension was motivated by the above findings to develop an error correction and downscaling scheme to advance the use of global reanalysis precipitation products in hydrologic modeling. The error correction and spatial downscaling was based on the SREM2D model and the high-resolution satellite precipitation data of the previous studies. This study focused on 437 significant storm events from the past 10 years that occurred over the Susquehanna River Basin in the Northeast United States. We quantify improvements on the reanalysis datasets through the satellite-driven downscaling scheme by presenting error metrics in (1) rainfall, and (2) generated runoff as function of basin size and storm severity. Results show that the generated rainfall ensembles of the downscaled reanalysis products could encapsulate well the reference rainfall. The statistical analysis including frequency and quantile plots, mean relative error and root mean square error statistics demonstrated significant improvements on downscaled data relative to the original re-analysis data in terms of both precipitation and runoff simulations. For instance in fall the re-analysis precipitation product is shown to have up to three times higher mean relative error compared to corresponding high resolution satellite data, this ratio increases up to four times for the simulated streamflow. The proposed downscaling scheme is modular in design- it can be applied for any gridded dataset in any region in the world.

# Performance Assessment of Satellite Rainfall Products for Hydrologic Modeling

Hojjat Seyyedi

B.S. Irrigation Engineering, 2005
M.S. Hydraulic Structures, 2008
M.S. Geo-info Science for Environmental Modeling, 2010

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy at the

University of Connecticut

2014

# APPROVAL PAGE

Doctor of Philosophy Dissertation

**Performance Assessment of Satellite Rainfall Products for Hydrologic Modeling**

Presented by

Hojjat Seyyedi, B.S., M.S.

Major Advisor _____

Emmanouil N. Anagnostou

Associate Advisor _____

R. Edward Beighley

Associate Advisor _____

Guiling Wang

Associate Advisor _____

Amvrossios C. Bagtzoglou

Associate Advisor _____

Jeffrey McCollum

University of Connecticut

2014

# ACKNOWLEDGEMENTS

Foremost, I owe my deepest gratitude to my advisor Prof. Emmanouil Anagnostou (Manos). Thank you very much for your continuous support, enthusiasm, and immense knowledge. I was overwhelmed by his patience and the encouragement that he gave me from the beginning of Ph.D. research till production of this dissertation.

I would like to acknowledge the support of the various members of the Prof. Anagnostou's research group, especially Vivana and Xinxuan. I am especially indebted to Viviana, who has contributed ideas, assistance, and energy towards virtually every aspect of this project.

I would also like to acknowledge the support of several key people who contributed towards the development of this research. I particularly thank Prof. Beighly, and Dr. McCollum for useful discussions and contributions. Additionally, I would like to thank the rest of my doctoral committee, Prof. Wang, and Prof. Bagtzoglou for their guidance. In addition, and importantly, I would like to acknowledge valuable assistance from the FM Global especially Dr. Hosam Ali, head of natural hazards group, Dr. Yizhong Qu, and the other staff of the company who has been incredibly helpful.

I would especially like to thank my wonderful friend, Mahboobe. She is absolutely wonderful and amazing in so many ways.

Last but not least, I would like to thank my parents. My mom has instilled in me a love of science and thanks her for all she has done. My dad has shown me how to lead a life of honor and respect, and how to persevere in the face of difficulty. They both have always given everything of themselves such that my brother and sisters and I could have every possible opportunity to succeed. Their sacrifices have benefitted us in many ways, and our accomplishments in life would not have been possible without their love and dedication.

*To My Parents*
*For all your supports throughout my life*

# Table of Contents

# LIST OF FIGURES

# LIST OF TABLES

## 1. Executive Summary

Deriving global-coverage flood maps requires accurate characterization of precipitation variability at high spatio-temporal resolution and global scale. Rainfall measurements have been based on data from ground sources (weather radars and rain gages) as well as space based sensor observations (from both geostationary and earth orbiting satellite platforms). The traditional ground-based observations are suffering from time and regional coverage gaps, which limit our ability to model hydrologic processes over remote and ungauged areas. Recent quasi-global integrated satellite rainfall products (i.e. merged satellite infrared and passive microwave rainfall estimates) are expected to compensate some of the conventional ground based observational limitations. Depending on the product type and the objective for the product development, each product is a combination of information from satellite sensors onboard geostationary (e.g. visible/infrared) and low earth orbit (infrared and active/passive microwave) platforms. Despite the differences in the various retrieval techniques used for the rainfall estimation and data merging, all integrated satellite rainfall product techniques are aiming to improve the accuracy of precipitation data at spatio-temporal scales relevant to hydrological applications (i.e. sub-daily temporal resolutions and less than quarter degree spatial scales). However, the accuracy of the different estimation methods varies with temporal and spatial scale, event type and terrain complexity.The uncertainties in remotely sensed precipitation data limit the use of satellite rainfall products in hydrological modeling. To elucidate applicability limitations, the following questions need to be investigated:

*What is the accuracy of current high-resolution quasi-global satellite rainfall products at hydrologic-relevant spatio-temporal scales, and what is the uncertainty dependence on rainfall type and terrain characteristics?*

*How does rainfall uncertainty propagate through integrated hydrologic simulations (streamflow, soil moisture, etc.), and how can this uncertainty characterization be inverted to improve uses of satellite data in hydrologic applications?*

The main goal of this research is to improve our understanding of satellite rainfall estimation uncertainty over land and use this improved understanding to advance uses of these data in hydrological applications. To address this overarching goal, this focused focused on the following three research objectives:

i. Incorporate surface soil moisture information in error modeling of the TRMM Passive Microwave (2A12) rainfall algorithm;
ii. Evaluate the accuracy of two quasi-global high-resolution satellite rainfall products and a global reanalysis rainfall product over mixed land cover/terrain conditions;
iii. Apply a stochastic modeling framework to correct and downscale satellite rainfall and reanalysis datasets for basin flood modeling application.

For the first specific objective, the study area is located in Oklahoma (OK) in the south-central United States (ranging from 32N to 39N and 93W to 102W). The NOAA/NSSL National Mosaic and Quantitative Precipitation Estimation system (named Q2) radar fields, The TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product, 2A12 (version 7), the NASA Catchment Land Surface Model (CLSM) soil moisture simulations and a two dimensional stochastic error model(SREM2D) are employed.

For the second and third specific objectives, the proposed study area is Susquehanna River Basin in United States (ranging from 39N to 43N and 75W to 79W). There are 1,006 river reaches identified to study the effect of basin size. The NCEP stage IV precipitation radar data, NASA's TRMM2B42V7 precipitation product, NOAA's GLDAS rainfall fields, a distributed

hydrological model (HRR), and a two dimensional stochastic error model (SREM2D are employed.

Findings from this study have been organized in following three chapters:

Chapter2. Incorporating surface soil moisture information in TRMM 2A12 error modeling

We used a two-dimensional satellite rainfall error model, SREM2D, to provide ensemble error representation of 2A12 rainfall using two different parameter calibration approaches: conditioning vs. not the SREM2D parameters to the soil surface wetness categories. The surface soil moisture condition determined in this study from NASA's Community Land Surface Model simulations, and used to derive area-averaged volumetric soil moisture values within 2A12 overpasses. Overpasses associated with less (more) than the 25th (75th) quantile of the cumulative distribution are considered as dry (wet), whereas the remaining constitutes normal wetness conditions. SREM2D parameters are calculated with and without considering the above orbit-average surface soil moisture classification. It is often convenient to separate the error in two parts, systematic error and random error. The significance of accounting for the soil moisture is assessed by quantifying its impact on the average discrepancies between the rainfall ensemble and the reference (systematic error) and on the spread of the ensembles (random error). The systematic and random error quantification of simulated reference-like ensemble members indicate in this study the significance of considering surface soil moisture condition in the error modeling of 2A12.

Chapter3: Hydrologic Evaluation of the Satellite and Re-analysis Precipitation Datasets over Mid-Latitude Basins

The error analysis carried on based on assessing the combined effect of resolution and retrieval uncertainty for a range of basin scales.We force the HRR model with a quasi-global high-resolution multi-sensor-satellite rainfall product (TRMM3B42V7) and a reanalysis global rainfall product (GLDAS) data to generate surface runoff. Parallel to this step we do the reference runoff generation via forcing the HRR with the stage IV radar data at the original resolution. There are two statistical analysis steps associated with the methodology in order to quantify the error propagated from rainfall product type on hydrologic simulation. The first statistical analysis is between radar rainfall data at the original resolution (4km) and TRMM3B42V7, and GLDAS precipitation products. The second one is between generated runoff from radar at the 4km and TRMM.3B42V7, and GLDAS precipitation products at 25km and 100km resolution.

Chapter4: A Stochastic Technique for Error Correction and Spatial Downscaling of Global Gridded Precipitation Products for Hydrological Applications

We implement and assessed SREM2D as a technique for spatial downscaling (25 km and 3-hourly) and error correction of the gridded re-analysis retrievals evaluated in chapter 2 (GLDAS), for improving the accuracy for hydrologic modeling. Specifically, we used SREM2D to predict the downscaled retrieval uncertainty in the form of ensemble realizations. For this reason we used TRMM3B42V7, and GLDAS, to calculate the error factor (β) which is a ratio of TRMM3B42V7 precipitation fields to reanalysis precipitation fields over the time (u,t). We used mean and standard deviation of logarithmic β fields as a part of SREM2D input error parameters.

We calculated the remaining required error parameters for SREM2D with TRMM3B42V7 precipitation fields and GLDAS precipitation fields. Once all error parameters are calculated we perturb GLDAS precipitation data using SREM2D to generate an ensemble of TRMM3B42V7-like realizations at 25km resolution. The results from the proposed downscaling method then used to force the HRR model to generate surface runoff. The quality of generated ensemble precipitation and runoff assessed in comparison with reference radar rainfall fields and associated surface runoff values. This study allowed us to investigate the error propagation of the re-analysis retrieval error ensembles in hydrologic simulations for a range of basin scales and evaluate the need of downscaling vs. error correction alone.

# 2. Incorporating Surface Soil Moisture Information in Error Modeling of TRMM Passive Microwave Rainfall

## 2.1.Introduction

The critical role of precipitation on hydrological processes requires accurate estimates over land at the highest possible spatio-temporal resolution. The conventional ground based rainfall measurements exhibit significant limitations as well as error sources. Specifically, measurements from rain gauges and weather radars are associated with sampling issues due to uneven distribution of sensors and limited spatial coverage. As a consequence, time and coverage gaps, particularly over complex terrain and several developing regions of the globe, limit the representativeness of those measurements in hydrologic modeling(Tapiador et al., 2012). Scientists have also raised inquiries for radar measurements due to issues with radar calibration, variability in reflectivity-to-rainfall transformation, beam geometry and precipitation profile effects, among others (Anagnostou et al., 2001;Krajewski et al., 1996;Krajewski and Smith, 2002).

Since 1970s meteorological satellites have been operational providing valuable remote sensing observations of rainfall over remote and ungauged regions of the earth(Scofield and Kuligowski, 2003) . Depending on the type, i.e. geostationary vs. low earth orbit, can carry the range of visible, infrared, and microwave sensors (Ebert and Manton, 1998;Heinemann and J. Kerényi, 2003;Simpson et al., 1996). Passive microwave (PMW) frequencies have been used for rain retrieval from low earth orbiting sensors for about 25 years. The techniques that have been developed and refined in time rely on the signal emitted by raindrops over oceans at frequencies at or below 37 GHz, and on the scattering signal of ice particles in the precipitation layer over land at frequencies at or above 85 GHz (Lensky and Levizzani, 2008). Depending on the type of

satellite sensor and designated goals, different PMW precipitation retrieval algorithms have been developed. Each product has its own strengths and weaknesses, but none of them appears to be universally better than the other (Kummerow et al., 2007). A widely used algorithm for precipitation retrieval from PMW frequencies is the Goddard Profiling (GPROF) technique (Kummerow et al., 2001;Wang et al., 2009). GPROF is a Bayesian based approach to retrieve the instantaneous rainfall with matching observed and simulated brightness temperature database through a radiative transfer model. Despite the sturdy linkage of PMW precipitation retrieval algorithms with the physical processes of precipitation formation, GPROF exhibits acceptable performances in heavy rain convective conditions over the ocean. However, in over land surface Lensky and Levizzani (Lensky and Levizzani, 2008) have indicated that PMW retrieval algorithms are suffering from uncertainties in precipitation type classification and temporal resolution.

The precipitation radar (PR) on-board the Tropical Rainfall Measuring Mission (TRMM) satellite that samples the vertical profile of precipitation can provide higher accuracy measurements against other space-based products and, to this capacity, it contributes on evaluating and developing PMW retrieval algorithms, such as the TRMM Microwave Imager (TMI-hereafter named 2A12) overland rain estimation algorithm (Dinku and Anagnostou, 2005;Furuzawa and Nakamura, 2005;McCollum and Ferraro, 2003). 2A12 has different rainfall retrieval procedures for each surface type (ocean, land, and coastal regions) to account for differences in the microwave emission characteristics. Rainfall has high emissivity and is non-polarized at PMW frequencies; which makes it possible to distinguish it from the ocean surface that is associated with low emissivity and high polarization. The overland retrieval algorithm faces more complexity due to the land cover changes and variable emissivity surface

background, besides to the non-polarized nature of the land (Gopalan et al., 2010;Wang et al., 2009); and it is affected by large number of surface parameters such as soil moisture, vegetation properties, and snow cover (Aires et al., 2011). Due to the surface complexity, the 85-GHz brightness temperature (Tb) observations associated with ice scattering have been the primary data for overland rainfall estimation. However, a screening procedure is required for PMW overland rainfall retrievals to define rain vs. no-rain areas and to mask-out regions producing scattering signals similar to rainfall, such as deserts, semiarid land, and snow cover (Ferraro et al., 1998).

The 2A12 algorithm has been extensively evaluated over land trough comparisons with ground based gauges and radar rainfall data, aircraft measurements and other spaceborne products (Fiorino and Smith, 2006;Liu and Zipser, 2009;Wang et al., 2009). Several deficits associated with the 2A12 land algorithm have been identified and discussed in those studies. Overall, the algorithm tends to overestimate rainfall rates from convective systems and to underestimate the rainfall resulted from the droplet coalescence processes and warm rainfall due to the lack of 85-GHz depression (Liu and Zipser, 2009;Liu et al., 2008;Wang et al., 2009;Zipser et al., 2006). The sensitivity of the rain screening algorithm to snow background and deserts enhances false alarms of the algorithm over those regions (Seto et al., 2005;Wang et al., 2009). Recently, Kirstetter et al. (Kirstetter et al., 2012b) presented the joint influence of precipitation structure and surface conditions (soil wetness and vegetation cover) on the error structure of the 2A12 product over land. While more research is needed to quantify the impact of specific factors like precipitation horizontal and vertical structure or vegetation water content, the biases of the 2A12 algorithm appear to be significantly organized alongside the surface wetness. Specifically, 2A12 is shown to overestimate under both dry and wet conditions, with stronger overestimation

exhibited under dry surface conditions. The authors attributed the associated overestimation of 2A12 to biases in the convective precipitation fraction classification, which were shown to be sensitive to surface wetness conditions.

Anomalies associated with 2A12 show underestimation in warm rain and overestimation in deep convective events. This plays an indubitable role in water resources management, and hydropower system operation as a part of satellite rainfall products application. Thus improving the accuracy and the error characterization of 2A12 rainfall estimates is critical for a number of satellite rainfall products and applications (Wang et al., 2009), as uncertainties in the 2A12 algorithm propagate to TRMM-based global multi-satellite rainfall estimates.

After having identified potential (physical) error factors, their impact on satellite rainfall uncertainties need to be quantified. An empirical way to assess this impact is simulating error corrected rainfall fields with accounting, or not, for the error factor at stake. Hossain and Anagnostou (Hossain and Anagnostou, 2006a) have developed a multi-dimensional satellite rainfall error model (SREM2D), which has been used in several studies to produce ensemble rainfall fields representing realizations of error-adjusted satellite rainfall products (Hossain and Anagnostou, 2006b;Maggioni et al., 2012;Maggioni et al., 2011). Triggered by the Kirstetter et al. (Kirstetter et al., 2012b) preliminary findings on soil wetness effects in TMI rainfall, the present study aims to quantify the impact of accounting for soil wetness in the statistical error modeling of 2A12 rainfall retrieval through SREM2D.

The surface soil moisture condition was determined in this study from NASA's Community Land Surface Model simulations, and used to derive area-averaged volumetric soil moisture values within 2A12's orbit over Oklahoma. Overpasses associated with less (more) than the $25^{th}$ ($75^{th}$) quantile of the cumulative distribution are considered as dry (wet) conditions, whereas the

remaining constitutes normal wetness conditions. SREM2D parameters are calculated with and without considering the above orbit-average surface soil moisture classification. It is often convenient to separate the error in two components: the systematic error and the random error. The significance of accounting for soil wetness will be assessed by quantifying its impact on the average discrepancies between the rainfall ensemble and the reference (systematic error), as well as on the spread of the ensemble (random error). The systematic and random error quantification of simulated reference-like ensemble members will indicate the significance of considering area-average surface soil wetness condition in the error modeling of 2A12.

In the next section we present the study area and data sets. In section 3 we describe the 2A12 rainfall error analysis, and in section 4 the SREM2D and the parameter calibration, which is performed accounting - and not - for the soil wetness conditioning term. Finally, section 5 presents results and major findings from this research.

## 2.2 Study region and data

The study area is located in Oklahoma (OK) in the south-central United States (ranging from 32N to 39N and 93W to 102W, Fig. 1). The region has an elevation gradient from west to south-east, with the highest peak in northwestern corner and the lowest point in the southeastern corner respectively with 1515 and 88 meter above sea level. The study area is located in the tepid latitude of the globe, where cold and warm fronts interact frequently, yielding severe events and laying the region on the Tornado Alley. The high spatial gradient of soil moisture across the region makes it an ideal testbed for studying sensitivity of the rainfall retrievals for various soil moisture conditions. Established networks of in-situ sensors from the Oklahoma Mesonet and the good radar coverage from the NEXRAD network provide high-resolution information on the

land-surface state and on the rainfall in the Oklahoma region. The Q2-based radar-rainfall, 2A12 rainfall estimates and soil moisture simulations from the NASA-Catchment Land Surface Model utilized in this study are described next.

### 2.2.1. Q2 radar-based reference rainfall

All significant rain fields observed coincidentally by TRMM overpasses and the NEXRAD (http://www.roc.noaa.gov/WSR88D/) radar network over the study area from March to October 2009-2010 are collected. The NOAA/NSSL National Mosaic and Quantitative Precipitation Estimation system (named Q2; (Zhang et al., 2011)) combines information from all ground-based radars comprising the Weather Surveillance Radar – 1988 Doppler (WSR-88D) network (NEXRAD) to derive experimental radar-based products comprising high-resolution (0.01°, 5 min) instantaneous rainfall rate mosaics available over the Continental United States (Kitzmiller et al., 2011;Lakshmanan et al., 2007;Vasiloff et al., 2007;Zhang et al., 2005).

Many errors affect the estimation of rainfall from ground-based radars, such as non-weather echoes, non-uniform beam filling, range-dependency due to Vertical Profile of Reflectivity (VPR) variability, conversion of Z-to-R, and calibration of the radar signal (see (Villarini and Krajewski, 2010) for a review). While several procedures are already in place within the Q2 system to correct for these errors, additional post-processing steps were taken to refine the reference data set as much as possible.

The original Q2 products utilized in this study are (i) the radar-only instantaneous rain-rate national mosaic updated every 5min, (ii) the radar-only rain-rate national mosaic at hourly time step, and (iii) the hourly rain gauge-corrected national mosaic product. The reference rainfall is derived from an instantaneous bias-corrected Q2 product. Instantaneous Q2 products are adjusted

using co-located rain gauge observations to minimize the aforementioned errors: pixel-by-pixel ratios between the hourly gauge-adjusted and the hourly radar-only products are calculated and applied as multiplicative adjustment factors to the radar-only 5 min product. This adjustment is designed to minimize uncertainties with the Z-R relationship and calibration errors (see (Kirstetter et al., 2012a) for more details). Extreme adjustment factors (outside the [0.1-10] range) are discarded so the gauge-adjustment also serves as a data quality control procedure. Note that the rain gauge network in Oklahoma is dense and well suited to provide reliable hourly adjustments in the radar rainfall estimates. Moreover, working with data from summer season over a generally flat terrain avoids overestimation in the bright band and mitigates range dependency caused by VPR effects so the best measurements conditions (i.e., no beam blockage and radar beam below the melting level of rainfall) are retained. Although the quantitative interpretation of the weather radar signal in terms of rainfall may be complex, radars enable a reliable evaluation of area-averaged rainfall estimates.

The Q2 5-min products that encapsulate the TRMM satellite local overpass schedule time are used as the reference rainfall fields to the corresponding 2A12 rainfall estimates. Specifically, the reference rainfall $R_{ref}$ is computed from a block-Q2 rainfall pixel matching each land surface model pixel. All Q2 pixels (rainy and non-rainy) corresponding to a land surface model pixel are used to compute the pixel-average reference rainfall rate.

### 2.2.2. TMI 2A12 rainfall algorithm

The TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product, 2A12 (version 7) with horizontal resolution of ~5 km at 85.5 GHz is evaluated in this study. The algorithm is detecting rain vs. no-rain areas through establishing relationship between Rain Rate (RR),85 and

21 GHz vertically polarized channel using coincident TMI and TRMM precipitation radar (PR) data for both convective and stratiform data. Then the algorithm defines the percentage of convective and stratiform rainfall patterns by calculating convective probability of observed rainfall with accounting for TB10V, TB37V, TB85V, and TB85H, which are 10 and 37 GHz vertically and 85 GHz vertically (horizontally) polarized channels respectively (Gopalan et al., 2010;McCollum and Ferraro, 2003). The convective/stratiform pattern recognition reduces the associated uncertainty with possible rainfall rates for a given brightness temperature. Kirstetter et al. (Kirstetter et al., 2012b) performed a diagnostic error characterization on 2A12 rainfall estimates conditioned by the surface wetness, cloud structure and vegetation covers. During summer, better satellite performance was reported over regions with wetter soil moisture conditions. Regarding rainfall detection, the missed rain volume was shown to decrease for increasing surface wetness conditions. In terms of quantification, 2A12 algorithm exhibited greater overestimation (up to 55%) over dry surface conditions and lower (less than 10%) over wet surface conditions. This was attributed to overestimation in the convective rain fraction, which was shown to increase inverse proportionally to soil wetness. Therefore, soil moisture is identified as a potential factor for the 2A12 rainfall retrieval error, which will be modeled in section 3 using SREM2D.

### 2.2.3. Simulations of surface soil moisture

Surface conditions were assessed through surface soil moisture estimates from the NASA Catchment Land Surface Model (CLSM) (Ducharne et al., 2000;Koster et al., 2000). This model uses a catchment-based method to assess land surface processes, as the hydrological catchment defined by topography is the essential element of the land surface model. The variability of soil moisture is associated with three bulk soil moisture variables at the element scale. These

variables display equilibrium conditions and non-equilibrium conditions. In particular, equilibrium conditions are associated with the water table distribution and non-equilibrium conditions are found near the land surface.

CLMS utilizes several surface level meteorological variables (precipitation, air temperature and radiation among others), and pre-determined climatological vegetation parameters. CLMS is forced with precipitation data from the bias-corrected Q2-based rainfall dataset, and the residual meteorological variable data is the output from the NASA global atmospheric data assimilation system (Bloom et al., 2005), as part of the Global Land Data Assimilation Systems (GLDAS)(Rodell et al., 2004). The parameters utilized in this model are from the Goddard Earth Observing System Model, Version 5 (GEOS-5) system (Rienecker et al., 2008). Spin-up of the model is executed with Q2-based rainfall that is looped three times through three years of forcing data (2004-2006), which produces land-only model integrations. Studies such as Bowling et al.(Bowling et al., 2003), Nijssen et al. (Nijssen et al., 2003), and Boone et al. (Boone et al., 2004) have demonstrated the ability of CLSM in depicting soil moisture dynamics. The consistency and high correlation (0.54) between CLSM simulated surface soil moisture data utilized in this research and *in-situ* Oklahoma Mesonet observations have been demonstrated in (Maggioni et al., 2011).

### 2.2.4. Data matching

The Q2 radar-rainfall dataset, which represents the reference rainfall in this study, is matched to the 2A12 overpass times and spatially averaged to the Catchment model grid resolution (25 km). The total number of overpasses and pixels considered in this study are 867 and 190740, respectively. The Q2 pixels found within the land surface model pixel *A* are utilized to compute

the pixel-average reference rainfall rate $R_{ref}(A)$. The mean number of Q2 pixels (with native Q2 resolution being $1km^2$) associated with a model grid cell is around 620. Matched estimates of 2A12 and $R_{ref}(A)$ can only be produced at locations where actual observations were recorded by both the 2A12 and ground radars. On the other hand, the instantaneous satellite rainfall estimate $R_{sat}(A)$ is a block-2A12 rainfall pixel computed to match each land surface model pixel. The average number of 2A12 pixels (with native 2A12 resolution being around 8 km) associated with each model pixel is around 16. The 2A12 and Q2 data are dropped from comparison when there is more than 25% missing Q2 pixel values or when less than 8 2A12 pixels are available. By dropping these values, the block-Q2 values are more representative of the population that comprises the ground references, and the reliability of the block-2A12 values that comprises the satellite estimate is also increased. Statistics are applied on the matched Q2-2A12 data for comparing reference rainfall intensities to satellite-based estimates as a function of soil moisture conditions estimated by the Catchment model. Fig. 2 presents a sample of matched Q2 and 2A12 data for two representative cases of dry and wet surface conditions. In both cases the spatial distribution of rainfall is similar between Q2 and 2A12, although the Q2 data captures greater rainfall intensity than 2A12. The Fig. 2 represents a clear distinction between dry and wet condition in terms of intensity and pattern.

The average soil moisture values from CLSM over the entire 2A12 overpasses is determined to provide the mean wetness condition over the area. Overpasses with more than 50 recorded values are considered for calculating mean soil moisture values, therefore 20% (254 overpasses) of the matched data were discarded from the error analysis. Fig. 3 shows the cumulative distribution of the area-average volumetric soil moisture values from CLSM over considered overpasses. The data from overpasses associated with area-average soil moisture of less than the

25th quantile represent dry conditions. Similarly, data from overpasses associated with area-average soil moisture higher than the 75th quantile represent wet conditions, whereas the remaining overpasses represent normal conditions. The number of overpasses and associated data sample statistics associated with each soil wetness category is presented in Table 1. The rain fractions over the entire area and study period are 17%, 12%, and 9% for the wet, normal and dry conditions, respectively. Fig. 4 shows the distribution of the satellite retrieval relative error (defined as the difference between 2A12 and Q2 values divided by the total for the detected rain values higher than 0.1 mm/h.) computed at the pixel scale and instantaneous timescale for wet and dry conditions. It demonstrates a clear distinction in the probability density functions of satellite retrieval relative error under wet and dry conditions. Both distributions have their mode at 0.5 (overestimation of 2A12 compared to the reference) but the wet distribution is much narrower and more peaked than the dry distribution: the maximum relative frequency for wet condition is 0.48 compared to 0.31 in dry conditions. The 2A12 rainfall retrievals present more random error over dry conditions. The greatest difference between wet and dry conditions can be seen around the zero relative error value. Overall, the systematic error of 2A12 retrieval is shown to be lower over wet versus dry land surface conditions.

## 2.3. TMI rain rate error analysis

In order to characterize the precipitation, additional Q2-based products are considered in the present study such as the freezing level height and the radar echo top. As brightness temperature channels associated with the ice scattering in precipitating clouds significantly influence the 2A12 retrievals over land, equivalent ice water content and a simplified rain type classification were elaborated from the Vertically Integrated Liquid Content (VIL) provided at the original Q2

resolution (1 km, 5 min). The ice water content was evaluated from the 3D reflectivity mosaics between the freezing level height and the echo top. A two-step approach similar to Steiner et al. (Steiner et al., 1995) was applied to identify convective areas. First the centers of convective cells are identified from the VIL map using a threshold value (5.5 kg.m-2) above which it is assumed that precipitation can only result from convective processes. Then an associated convective region is identified from surrounding pixels with VIL values greater than 2 kg.m-2 at distances within 20 km. Pixels flagged as non-convective are designated as stratiform.

For each common grid domain pixel, *A*, a Convective Percent Index (CPI) is computed to quantify the contribution of convective rainfall to $R_{ref}(A)$ as follows:

$$CPI(A) = 100 \frac{\sum_{j=1}^{nconv} Q2(a_j)}{\sum_{i=1}^{n} Q2(a_i)} \quad (1)$$

Where *Q2* denotes the Q2 rain rate at the original data product resolution (1 km2) $a_i$; $n$ is the number of Q2 data points inside the common grid pixel *A*; and $nconv \leq n$ is the number of Q2 pixels flagged as convective inside the pixel *A*. Each 2A12 pixel is associated with a rainfall rate at ground and with an associated convective rainfall rate. As for Q2, a CPI is computed to quantify the contribution of convective rainfall to *R(A)*. We also consider the inputs of the retrieval algorithm with the brightness temperatures from the 1B11 product measured by TMI. They are matched in space and time with the 2A12 products, the Q2 rainfall, and the soil moisture outputs from the land surface model. In order to preserve the statistical characteristics of the information input for the 2A12 algorithm, the brightness temperatures are taken at their native resolution.

For the last sample, we define two extreme surface wetness conditions (see Table 2): “dry” that represents the 25% driest sub-sample (soil wetness values below 0.178) and “wet”

representing the 25% wettest sub-sample (soil wetness values above 0.309). The vegetation was defined with two classes of NDVI (see Table 2): the low vegetated areas that represent the 25% sub-sample with NDVI values below 0.411, and the remaining cases classified as "normal". These sub-samples differ in terms of rainfall regimes. Fig. 5 shows the averaged echo top, and reference convective contribution of the sampled rainy systems as functions of the ice water content $VIL_{ice}$ over the 6 different surface conditions. All variables consistently increase with the $VIL_{ice}$ as a characterization of more intense convective systems ($VIL_{ice}$ > 0.3 g.m-2) usually associated with high ice content.

The echo top and convective contribution ($CPI$) values are higher over dry surfaces than over wet surfaces and higher over vegetated areas. The relation between these cloud features and surface characteristics is likely associated through a seasonal effect. During the warm season, more vegetation and drier soils are associated with higher convective activity related to thunderstorms. Wintertime period is associated with less vegetation, wetter soils and lighter precipitation arising from large-scale frontal systems. The rain fraction increases with the $VIL_{ice}$ from less than 10% for $VIL_{ice}$ < 0.05 g.m-2 up to more than 40% for $VIL_{ice}$ > 0.5 g.m-2, and its values are generally higher over wet surfaces than over dry surfaces and lower over vegetated areas: isolated convective cells are likely associated with dry and vegetated surfaces during the warm season and more extended rain systems occur during the cold season over wet and less vegetated surfaces. The shifts of the rain rates towards higher values over wet and less vegetated areas are correlated to the rain fraction. While higher rain rates are generally associated with convective clouds, working at the 25-km resolution certainly explains why the differences of rain fraction are correlated to the differences of mean intensity over the various surfaces.

The features shown on Fig. 6 impact the brightness temperature (Tb) signatures of rain events from which the 2A12 algorithm infers rain rates at the ground. Fig. 6 brings insight with the conditional median of the Tb distributions from 10V, 37V, and 85H channels over the various surfaces as functions of $VIL_{ice}$. Recall that the Tb are taken at their native resolution. The distributions include Tb sampling raining and non-raining scenes because a given pixel is not necessary completely filled by rainfall (especially for $VIL_{ice}$ < 0.2 g.m-2, see Fig. 7). Differences at the lowest $VIL_{ice}$ values (associated with light rain and rain/no-rain transition) could be linked to the differences in surface emissivity between the regions. For all channels considered, the Tb distributions are shifted towards lower values over wet surfaces compared to dry surfaces (e.g. ~15K Tb drop from dry to wet soil conditions at 10GHz channel over low vegetated areas) and downward over less vegetated areas (e.g. ~10K Tb drop between for low vegetated areas at 10GHz channel). It is likely that surface (soil moisture and vegetation) affect the lower frequency Tb (10GHz and 21GHz channels) because the dielectric constant of water is higher than soil, so the surface emissivity (and upwelling Tb) is colder over a wet soil than a dry soil. Wintertime period associated with less vegetation is characterized by colder skin temperature that also decreases the upwelling Tb. Surface emissivity may have an impact at higher frequencies (85GHz channel) when the rain fraction is low, especially for $VIL_{ice}$ < 0.2 g.m-2. The higher frequency channels present consistently higher sensitivity to the $VIL_{ice}$. The surface is probably masked once precipitation develops an optically thick ice canopy at ~ 0.3 g.m-2. A difference of ice content, ice density or spatial homogeneity of the ice field may drive the differences in Tb at the 85GHz channel.

All of these characteristics present a significant diversity for assessing the 2A12 precipitation retrievals. Note that the seasonal cycle of the surface due to the time of the year is correlated to

the seasonal cycle of precipitation characteristics. It is difficult to distinguish uniquely the impact of the surface conditions on the 2A12 retrievals from the impact of the vertical structure of the precipitating clouds.

To investigate the rainfall quantification from 2A12, Fig. 7 shows the behavior of the Mean Relative Error (expressed in percentage) as a function of $VIL_{ice}$ for various conditions of soil moisture and vegetation. A rainy pixel is included in the statistics if both satellite and the reference are nonzero to emphasize the satellite sensor's ability to quantify precipitation when it is raining. The satellite algorithm significantly overestimates the reference mean values for dry and medium soil moisture conditions over the range of ice water content values. For wet conditions, 2A12 overestimates the reference values for $VIL_{ice}$ < 0.3 g.m-2 and underestimates the reference values for $VIL_{ice}$ > 1 g.m-2. The degree of overestimation significantly relates to the soil moisture with greater MRE values (up to 250%) over dry surface conditions and lower MRE values (up to 150%) over wet surface conditions. One can note that the impact of the vegetation on the bias seems less important than soil moisture. The overestimation decreases with increasing $VIL_{ice}$ with the strongest decrease being at $VIL_{ice}$ below 1 g.m-2. This pattern correlates well with the rapid drop of Tbs at 37V and 85H GHz channels (see Fig. 6).

The increase in satellite-based overestimation of rain rates from wet to dry soil moisture conditions is correlated with the corresponding increase in overestimation of convective detection and convective fraction from the same range of wetness conditions (shown in Fig 5). Because convective rainfall generally implies greater rainfall rates than stratiform rainfall (see (Gopalan et al., 2010), their Fig. 1 and 3), a weighting shifted towards positively biased convective fractions would result in overestimation of 2A12 rain rates. Therefore, this increase in

the positive bias relative to the reference from wet to dry regions is attributed to the rainfall misclassification of the 2A12 algorithm.

## 2.4. The rainfall error modeling (SREM2D)

SREM2D is a two-dimensional satellite rainfall error model developed for simulating satellite rainfall ensembles by perturbing a reference rainfall field (Hossain and Anagnostou, 2006a, b). However in this study we run the model in inverse mode, which works in the opposite way: namely, we force the model with the "satellite" dataset to obtain "reference"-like (i.e. error corrected) rainfall ensembles. Specifically, SREM2D was forced here with Q2 radar estimates to produce 2A12 (satellite) - like rain fields. SREM2D can model both the spatial and temporal variability of the satellite retrieval error as well as the spatial structure of the successful delineation of rainy and non-rainy areas. This aspect of SREM2D is innovative with respect to standard rainfall error models, which usually assume perfect delineation of rain areas and simply scale the input precipitation forcing with a multiplicative perturbation (Reichle et al., 2007). Previous studies have demonstrated that rainfall ensembles produced by SREM2D provide better encapsulation of the reference radar precipitation and they better reproduce satellite rainfall error statistics (such as bias and RMSE) than simpler rainfall error models (Hossain and Anagnostou, 2006b;Maggioni et al., 2011).

In this study, we perturbed 2A12 satellite data through SREM2D to generate an ensemble of 50 Q2-like realizations. The input error parameters evaluated for 2A12 with respect to the high-quality Q2 product along the study period (i.e., the two warm periods of 2009 and 2010) are: 1) probability of rain detection (POD); 2) mean of the logarithmic error (mu-Gaussian of log-error), defined as the log-error, where error is the multiplicative factor 'e' as in Rsensor=Rreference*e ;

3) non-detected mean rain rate; 4) probability of no-rain detection ($POD_{no\text{-}rain}$); 5) Correlation length for the retrieval error ($CL_{ret}$); 6) Correlation length for successful delineation of rain ($CL_{rain\ det}$); 7) Correlation length for successful delineation of no-rain ($CL_{no\ rain\ det}$). For more details about the model parameters, we refer the reader to (Hossain and Anagnostou, 2006a).

SREM2D model parameters are determined based on two different approaches, hereafter named conditional and unconditional. In the conditional approach, the model parameters are computed for each area-average soil wetness category (defined in Section 3) separately; in the unconditional approach, the model parameters are calibrated based on the entire matched database independent of soil wetness condition. The model parameters are summarized in Table 3. Results indicate that SREM2D calibration parameters are different for the different wetness categories. In terms of spatial patterns, the correlation lengths of rain detection, no-rain detection and retrieval error for the unconditional approach are less than 150 km and consistently fall between resulted values for wet and normal categories. The lower correlation length can be translated to the lower dependence between variables in space. Regarding the random error, the range of standard deviation of logarithmic multiplicative error is between 1.12 (wet) to 1.15 (dry) and equals 1.14 in the unconditional approach. The values represent higher magnitude of error between reference and sensor data in dry condition relative to the wet and unconditional approaches. The $POD_{no\text{-}rain}$ for the unconditional approach is equal to 0.97, and it is the same when computed for wet and normal conditions while for dry condition it is slightly lower (0.93). The non-detected mean rain rate is 0.14 (0.10) in wet (dry) condition, whereas its value for the unconditional approach is 0.12. Wet conditions are related to higher rain rates, which results in higher non detected rain rates from 2A12.

Furthermore, some SREM2D parameters are presented as a function of satellite rain rate thresholds (Fig.s 8 and 9). In fact, preliminary calculations indicated that parameters such as the probability of rain detection and the mean error are function of rain rate. For instance, the mean error shows positive values in low rain rates and negative values in high rain rates. This is consistent with Kirstetter et al. (Kirstetter et al., 2013) results demonstrated over West Africa for a similar overland PMW algorithm. In terms of POD, the dry surface soil moisture levels have the lowest values whereas wet conditions have the highest POD values. The surface soil moisture levels in the unconditional approach converge at higher levels of satellite rain rate. Thus, a dynamic characterization of these error parameters is chosen in place of a single value, representing a novelty with respect to the original version of SREM2D.

After calibration, SREM2D is used to generate 50 rainfall ensembles from overpasses in each soil wetness category based on the two calibration approaches (conditional and unconditional). Fig. 10 shows time series of cumulative rainfall simulated by SREM2D for each ensemble member using the two calibration approaches. The range of ensembles in the unconditional versus conditional SREM2D parameter calibration approach shows differences in encapsulation of the actual rainfall time series. The conditional approach exhibits a much smaller ensemble range than the unconditional approach. Overall, the Fig. indicates better convergence of SREM2D ensemble members in the conditional approach.

## 2.5. Results

The ensemble-based error modeling results are presented here for the two SREM2D calibration approaches, where parameters are conditioned or not to the simulated orbit-average soil moisture values. Specifically, in Section 4.1 we provide error metrics (relative standard

deviation and efficiency coefficient) to evaluate the improvement of conditional (surface moisture wetness) error modeling relative to unconditional error modeling, whereas in Section 5.2 we perform an ensemble verification of the SREM2D rainfall ensembles produced using the different calibration approaches to further quantify the improvement due to the use of soil wetness information.

### 2.5.1. Error metrics

Two error metrics are presented in this section to evaluate the improvement of applying error modeling conditional to the near-surface soil moisture vs. error modeling without use of soil moisture information. The first error metric is the ratio of standard deviation (RSTD) of differences between SREM2D-generated ensemble mean values and Q2 rainfall values to the mean Q2 radar rainfall values:

$$RSTD(U_j) = \frac{\sqrt[2]{\frac{\sum_i^n \left( \left( R^i_{sat_{mean}}(U_j) - R^i_{ref_{mean}} \right) - \overline{\left( R^i_{sat_{mean}}(U_J) - R^i_{ref_{mean}} \right)} \right)^2}{N}}}{\bar{R}_{ref_{mean}}} \qquad (2)$$

$R^i_{sat_{mean}}$ is the SREM2D-generated orbit-average satellite rainfall ensemble, $R^i_{ref_{mean}}$ is the orbit-average reference radar-rainfall and $\bar{R}_{ref_{mean}}$ is the mean radar rainfall over all orbits. $U_j$ and *i* represent the ensemble number and orbit number, respectively, while *N* is the total number of orbits. The second error metric is the Nash-Sutcliffe Efficiency (McCuen et al., 2006;Nash and Sutcliffe, 1970), which is presented to evaluate the goodness of fit of the SREM2D-generated ensembles and radar Q2 data:

$$NSE(U_j) = 1 - \frac{\sum_1^N \left(R^i_{sat_{mean}}(U_j) - R^i_{ref_{mean}}\right)^2}{\sum_1^N \left(R^i_{ref_{mean}} - \bar{R}^i_{ref_{mean}}\right)^2} \quad (3)$$

NSE equal to one indicates perfect model performance, while zero indicates model performance equal to climatology. Negative NSE values indicate that the model prediction introduces more error variability than the variability of the retrieved process. The two error metrics are computed for ensembles generated with the two different SREM2D calibration approaches; namely, calibrating the SREM2D parameters conditional and unconditional to the soil moisture.

The results for RSTD and NSE are shown in the boxplots of Fig.s 11 and 12, respectively. For the RSTD plots we note that the wet category's unconditional "box and whisker" plot has a relatively large interquartile range (IQR) compared to the conditional plots. Both conditional and unconditional plots are skewed to the lower value, which means that the concentration of values is closer to the 25$^{th}$ percentile. The 25$^{th}$ percentile value is 1.55 (1.7) for the conditional (unconditional) approach. The overall range of RSTD values is much larger in the unconditional simulation compared to the conditional simulation. Both approaches under normal wetness conditions show similar IQRs, similar overall range, and symmetrical distributions. Given all similarities, the 25$^{th}$ percentile for the conditional simulation is 1.75 compared to 1.9 for unconditional. For dry condition, the IQR for the unconditional simulation is larger than the conditional simulation, which corresponds to higher statistical dispersion in the unconditional simulation. Both are skewed to the lower value, as the data is concentrated closer to the 25$^{th}$ percentile, which is 1.7 and 1.75 for conditional and unconditional calibrations, respectively. Overall, the IQRs of the conditional simulations are smaller than the unconditional ones, which is a good indicator of lower variability in the generated ensembles.

Similar to the RSTD error metric, under wet conditions, the unconditional error model ensemble exhibits a much larger IQR for the NSE values than the conditional error model ensemble. The 25$^{th}$ percentile for the conditional case is 0.28 and -0.56 for the unconditional one. Both cases are skewed to the higher end, meaning that values are concentrated around the 75$^{th}$ percentile. In normal wetness conditions, the IQRs are similar to each other. The 25$^{th}$ percentiles are 0.55 and 0.45 for conditional and unconditional error model ensembles, respectively. The dry condition plots are both skewed to the higher value but their IQRs are largely different. The IQR of the unconditional error model ensemble is much greater than the IQR of the conditional one. The 25$^{th}$ percentiles are 0.1 and 0.58 for the unconditional and conditional error model ensembles, respectively. Overall, the conditional simulations perform better than the unconditional: this is evident in that the conditional plots have higher concentration of NSE values closer to 1, indicating that the SREM2D ensemble corrections bring satellite data closer to the reference (Q2 radar data).

### 2.5.2. Ensemble verification

We used three ensemble verification methods to assess the accuracy of the SREM2D-derived ensembles in terms of characterizing the variability of the satellite rainfall retrieval error: exceedance probability, uncertainty ratio, and rank histograms. The exceedance probability (EP) and the uncertainty ratio (UR) have been suggested by previous studies as a way to evaluate probabilistic prediction performance (Hossain and Anagnostou, 2005;Hossain et al., 2004;Maggioni et al., 2011). These two metrics evaluate two contradictory aspects of error modeling: if the ensemble limits are too narrow, i.e. the EP is high, then the benchmark

uncertainty is underestimated, while if the limits are too wide, i.e. the UR is high, the benchmark uncertainty is overestimated. In both cases, the model would have a poor predictive ability.

The exceedance probability (EP) is used to assess the ability of SREM2D-generated ensembles in encapsulating the reference data:

$$EP = \frac{N_{exceedance}}{N_t} \quad (4)$$

where $N_{exceedance}$ is the number of times the reference rainfall value falls outside the ensemble envelope and $N_t$ is the total number of data samples. A score of EP equal to zero indicates a perfect encapsulation of the reference within the ensemble envelope. On the other hand, if EP is equal to 1, that reference always falls outside the ensemble bounds. Results for EP values are shown in Fig. 13. The conditional simulation has a lower exceedence probability across all conditions. The difference between the resulted EP values of the conditional and unconditional error model ensembles is larger under dry soil condition (32%), which means that the number of times that Q2 falls outside the ensemble envelope is decreasing if the conditional approach is applied. The unconditional and conditional results show the same trend, dry > normal > wet. The magnitude of improvement from unconditional to conditional is increases also from dry to wet. Because the random error is greater for dry conditions (i.e. the spread is greater, see Fig. 4), it is harder for the reference rainfall to fall outside the ensemble envelope.

The mean uncertainty ratio (UR) is next used to evaluate the accuracy of ensemble envelope width. UR is computed as the ratio of aggregate ensemble widths divided by the 'benchmark' uncertainty, defined as difference between the ensemble mean and the reference rainfall:

$$UR = \frac{\sum_1^N \left(R^i_{Sat_{upper}} - R^i_{Sat_{lower}}\right)}{2 \times \sum_1^N \left|\hat{R}^i_{sat} - R^i_{ref}\right|} \quad (5)$$

where $R^i_{Sat_{upper}}$ and $R^i_{Sat_{lower}}$ represent the upper and lower bounds of simulated ensemble respectively, $\hat{R}^i_{sat}$ represents the mean of generated ensemble over orbit number *i*, $R^i_{ref}$ the radar rainfall field over orbit number *i* , *i* is the orbit number, and *N* is the total number of orbits in each category. A UR value of 1 refers to a perfect delineation of Q2. A UR value greater than 1 corresponds to an overestimation of the actual uncertainty; whereas a UR value less than 1 means that the ensemble underestimates the benchmark uncertainty.

Fig. 14 shows mean UR calculated for SREM2D ensembles generated by conditioning - or not - the parameters to the surface soil moisture. In all cases the ensemble spread overestimated the actual error variability. Evidently, the ensemble generated with conditional parameters shows values of UR closer to 1, which corresponds to a better estimate of the actual uncertainty. The improvement is relatively higher in wet and normal soil wetness conditions.

Finally, rank histograms are used to evaluate the SREM2D ensemble predictions. If the model is able to correctly reproduce the prediction uncertainty, the ensemble members are equally likely to occur as the reference simulation is (Siegert et al., 2012). Namely, the number of times that the reference falls within any two adjacent ensemble members should be independent of the position of the members in the ordered ensemble. Thus, the rank histogram, which is the histogram of the rate at which the reference falls into each interval, should be flat (Hamill, 2001;Hamill and Colucci, 1997). Two shapes of the rank histogram, a U-shape and a sloped shape, are both indicators of poor ensemble predictions. Specifically, if a U-shaped histogram represents a lack of variability in the ensemble, a sloped histogram represents a constant bias in the ensemble prediction.

Results are shown in Fig. 15. For the unconditional approach (right panels), the U-shape rank histogram in the wet and normal conditions indicates the ensemble suffered from the lack of variability while the dry condition histogram exhibits an uneven pattern, which may result from the lack of variability and a bias in the ensemble prediction. Flatter rank histograms are observed for all ensembles obtained by conditioning the SREM2D parameters to the surface moisture level and can be interpreted as equally likely occurrence of ensemble members as the reference simulation is.

## 2.6. Conclusions

In this study we assessed the impact of conditioning a stochastic error model for 2A12 rainfall estimates on near-surface soil moisture information derived from land surface simulations. The study used reference rainfall fields from Q2 algorithm over Oklahoma for the warm season months of 2009 and 2010. TRMM orbit-average surface wetness conditions (wet, normal, dry) over Oklahoma were defined based on surface soil moisture simulations from the NASA Catchment land surface model forced with Q2 rainfall fields. SREM2D, a two-dimensional satellite rainfall error model, was used to generate ensembles of satellite rainfall error with parameters that were conditioned or not to the three surface wetness conditions.

The error modeling improvement assessment was performed using different metrics: random error quantification (RSTD, NSE), uncertainty quantification (EP, UR), and ensemble verification (rank histograms). Significantly better performances in terms of all analyzed statistics are reported for the ensemble realizations generated using the conditional approach for estimating the SREM2D parameters. The interquartile range for boxplots of RSTD and NSE is larger for ensembles generated using the unconditional approach. This can be attributed to the

statistical dispersion of ensemble realizations generated disregarding soil wetness conditions. The random ensemble error quantification exhibits lower EP and UR for the ensemble simulations conditioned to the soil wetness. This results in better encapsulation of the surface rainfall (represented by reference rainfall data) and more accurate estimation of the retrieval error variability on the basis of the SREM2D-generated ensembles.

The study relies upon the quality and specification of 2A12, Q2, and NASA Catchment land surface model simulations. Future studies should investigate ways to directly incorporate soil moisture as a parameter of the error model, as we demonstrated how some error statistics are highly dependent on it. As a first attempt to investigate this question we focused on conditions ensuring good reference estimates. Performing the comparison during the warm season over Oklahoma offers the best measurement conditions related to Vertical Profile of Reflectivity (i.e., radar beam below the melting layer) and beam blockage effects. Applying the same approach over various areas to assess the influence of other potential error factors (e.g. shallow convection over complex terrain) or during other seasons (with different soil moisture and vegetation conditions) will be the subject of future studies.

To generalize findings of this study it requires extending this error modeling exercise over different regions, longer periods of time, and including additional surface characteristics (e.g. vegetation cover). The framework is applicable for the other microwave imagers onboard low-earth orbiting satellites and the result can be useful for the GPM mission.

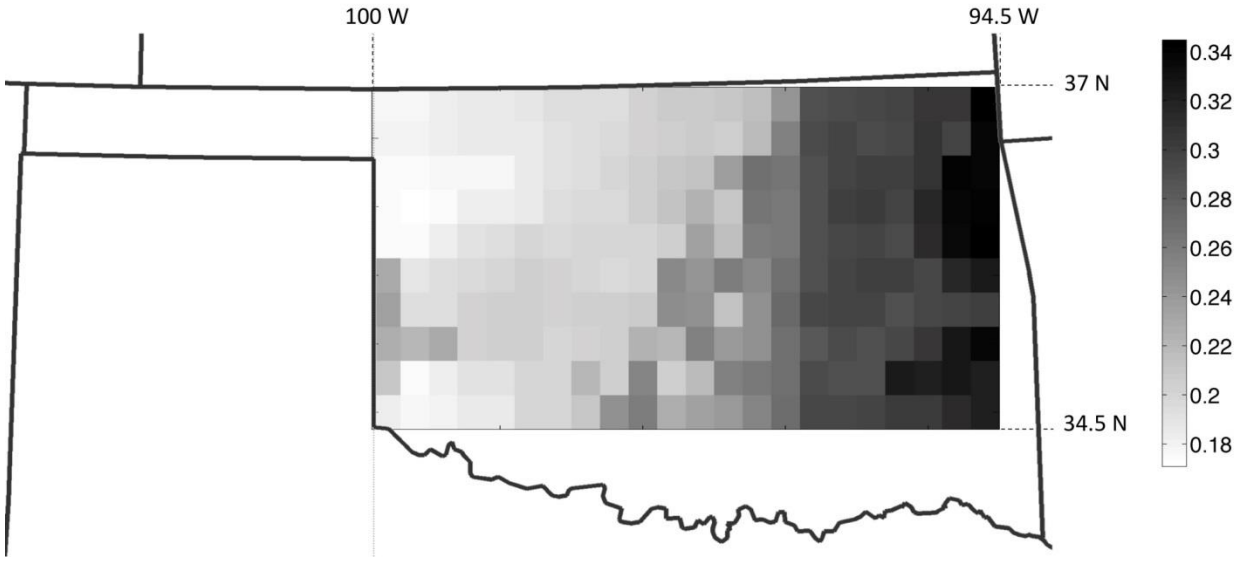


Fig.2.1: Map of 7-year (2004-2010) average surface soil moisture ($m^3/m^3$) simulated by CLSM, overlaid by the 25-km grid of the study domain.

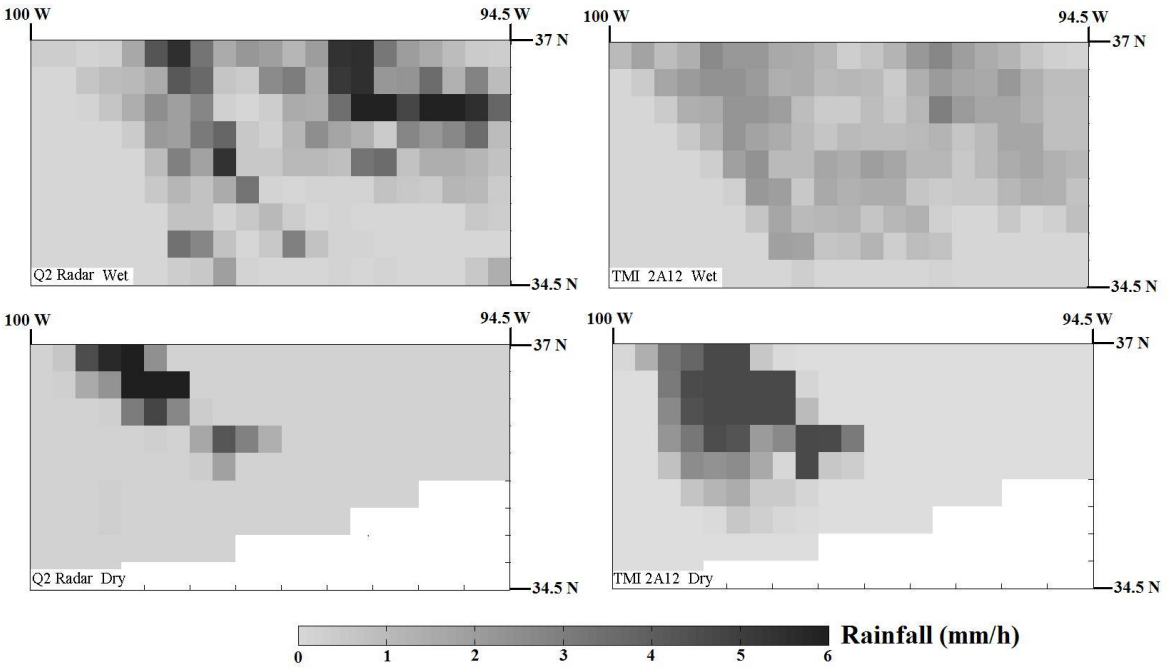


Fig. 2.2: Precipitation maps from 2a12 and Q2 radar - a) Wet soil condition, 12 Sep 2009, 21:45 hrs (top), b) Dry soil condition, 17 July 2009, 03:55 hrs (bottom), dashed lines indicate the 2A12 orbit over the study domain.

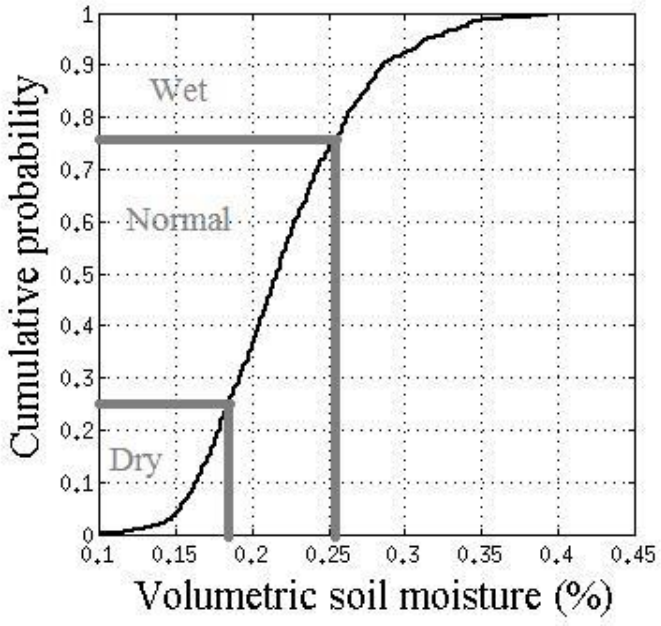


Fig. 2.3: Cumulative probability of the 2A12-orbit average volumetric soil moisture for all 2A12 orbits in March to October of 2009 and 2010.

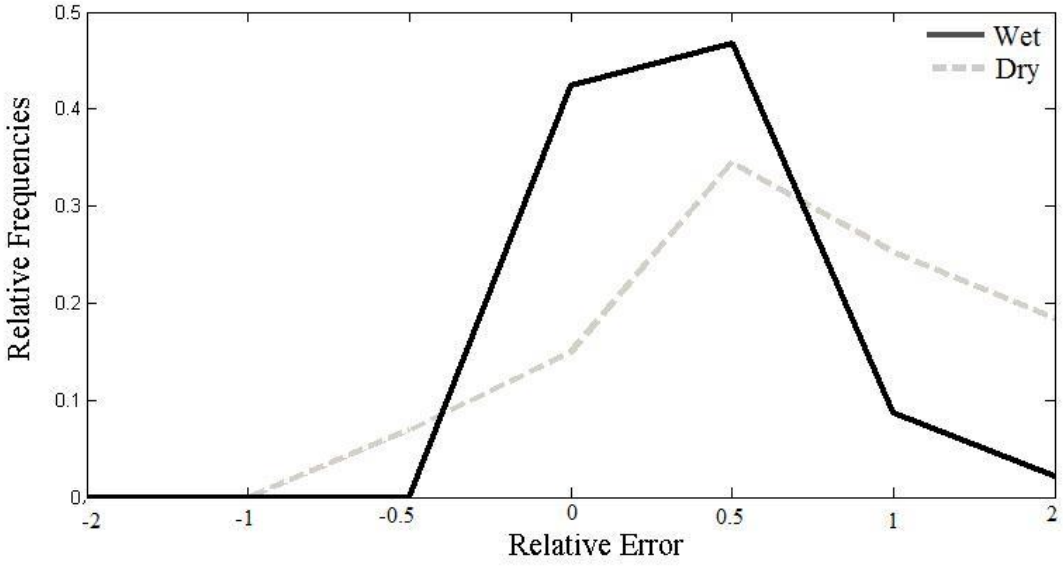


Fig. 2.4: Probability density function of relative error for all orbits.

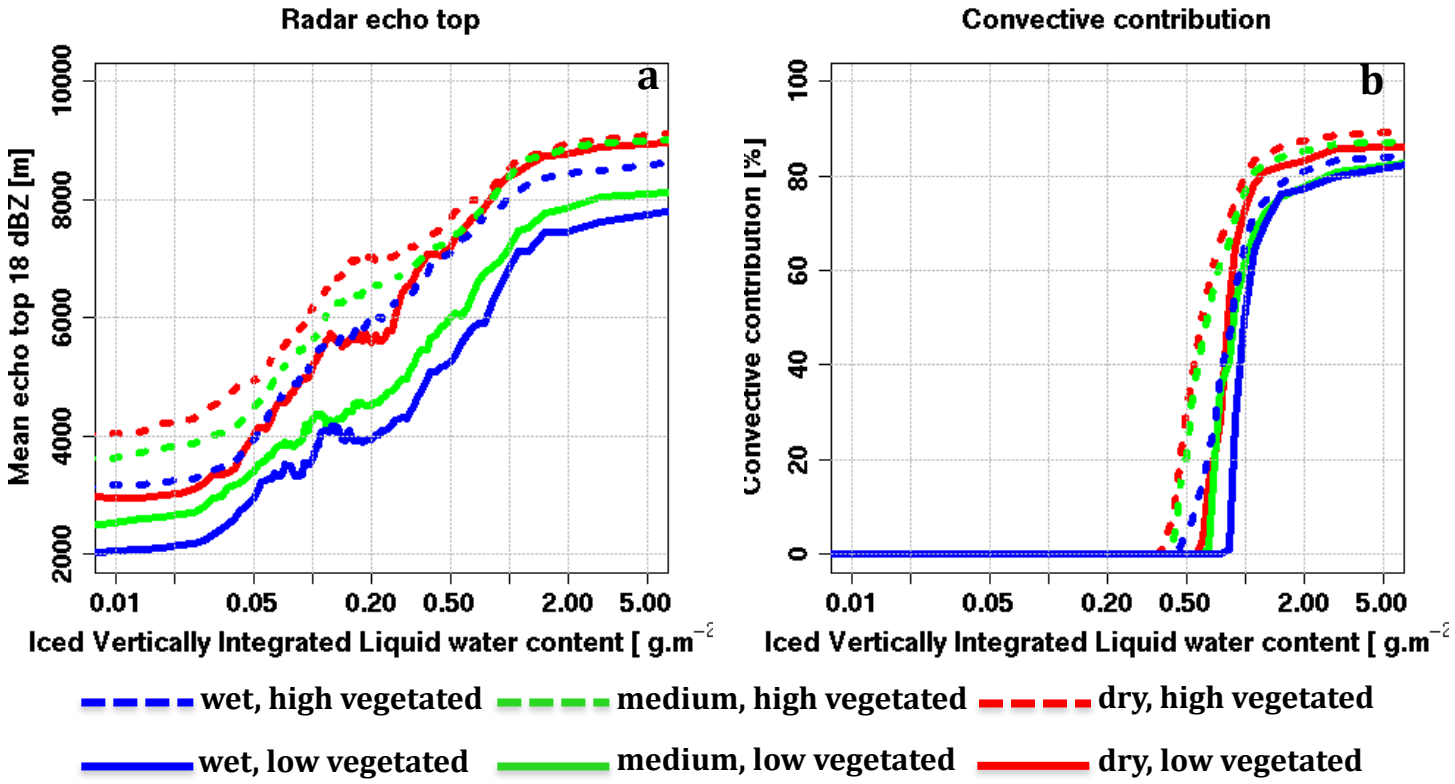


Fig. 2.5: Averaged radar echo top (a), and convective volume contribution (b), as functions of ice water content for various soil moisture and vegetation conditions.

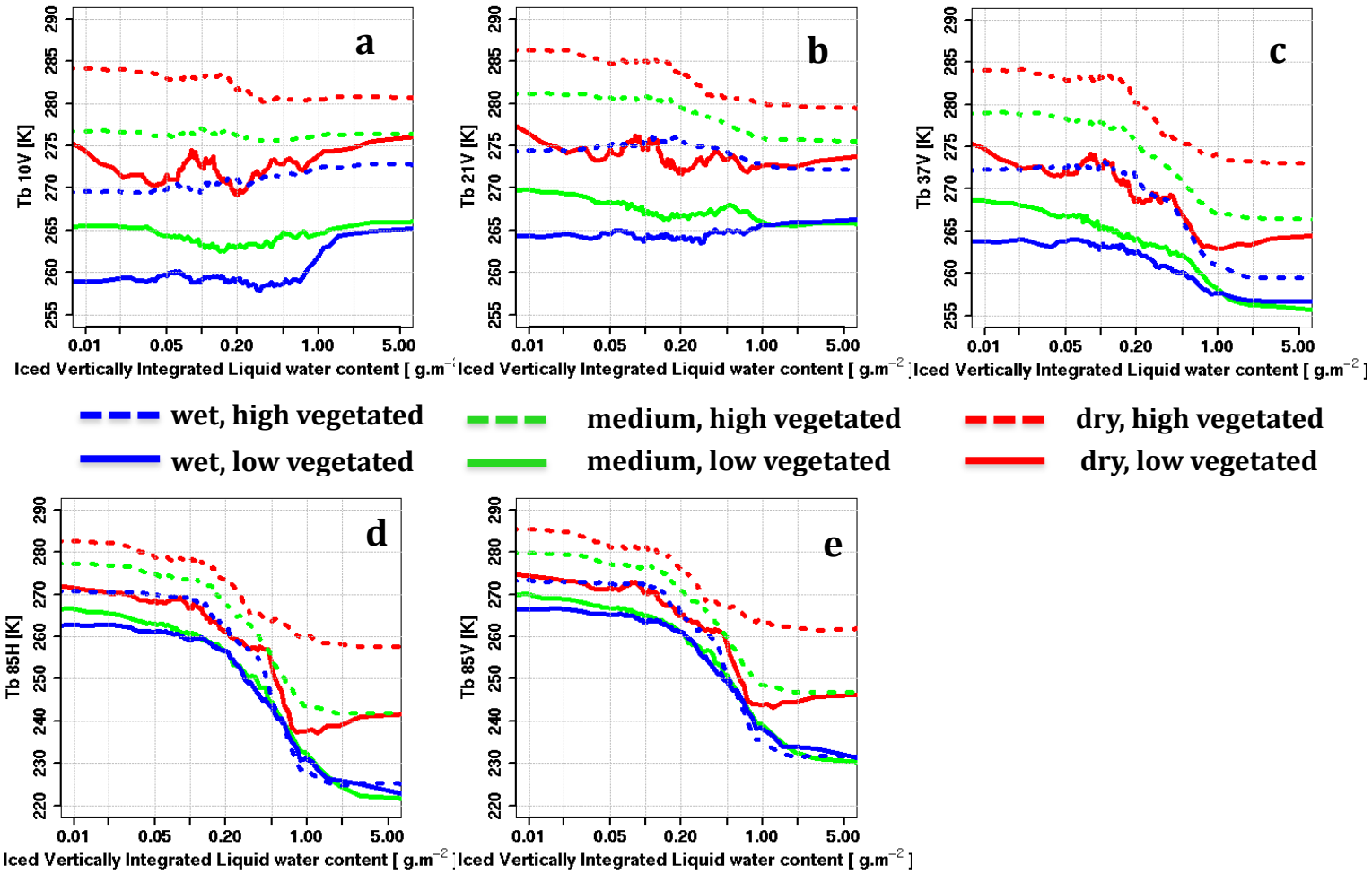


Fig. 2.6: Averaged brightness temperatures at 10V (a), 21V (b), 37V (c), 85H (d) and 85V (e) channels as functions of ice water content for various soil moisture and vegetation conditions.

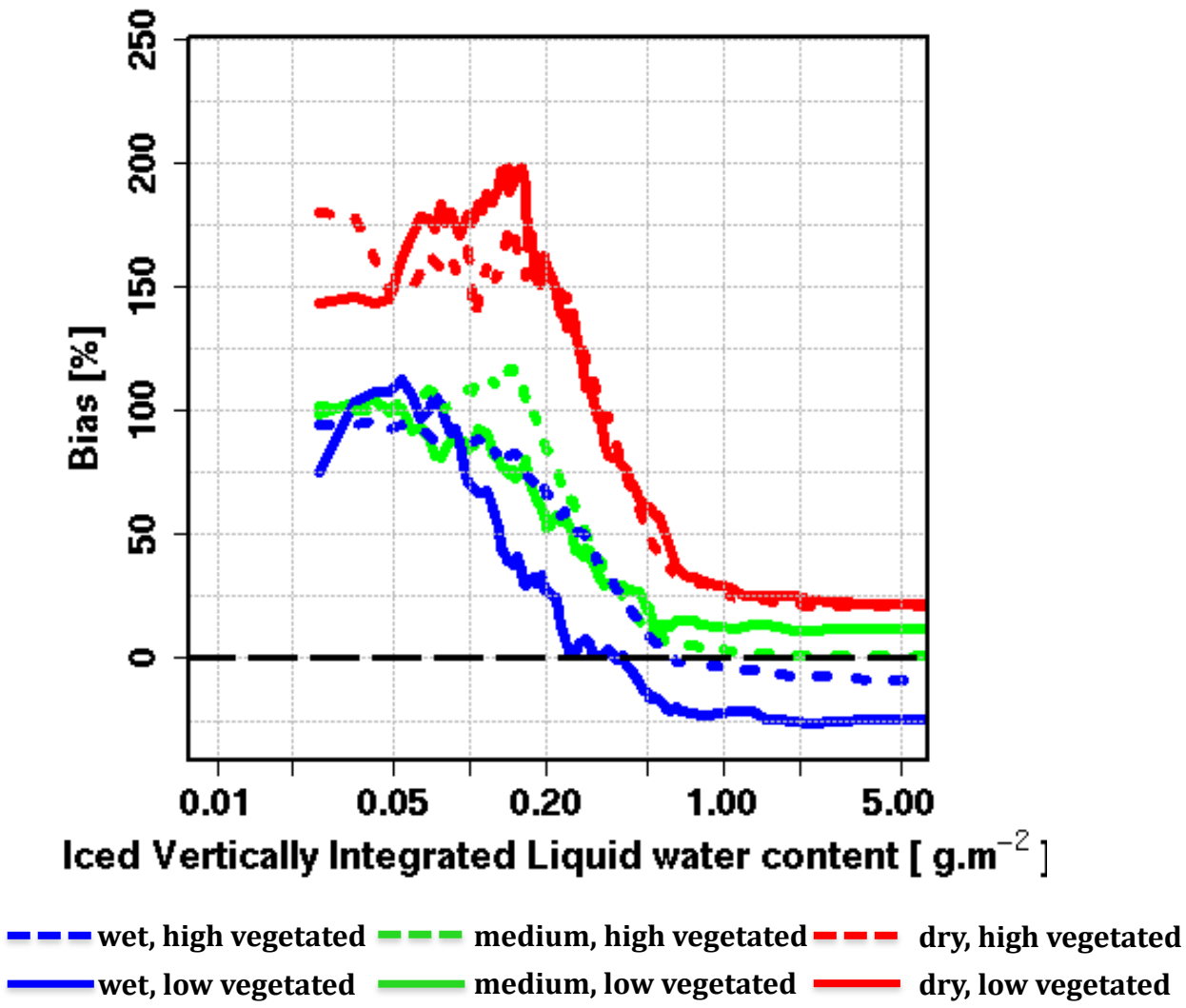


Fig. 2.7: Mean Relative Error of 2A12 versus reference as functions of ice water content for various soil moisture and vegetation conditions.

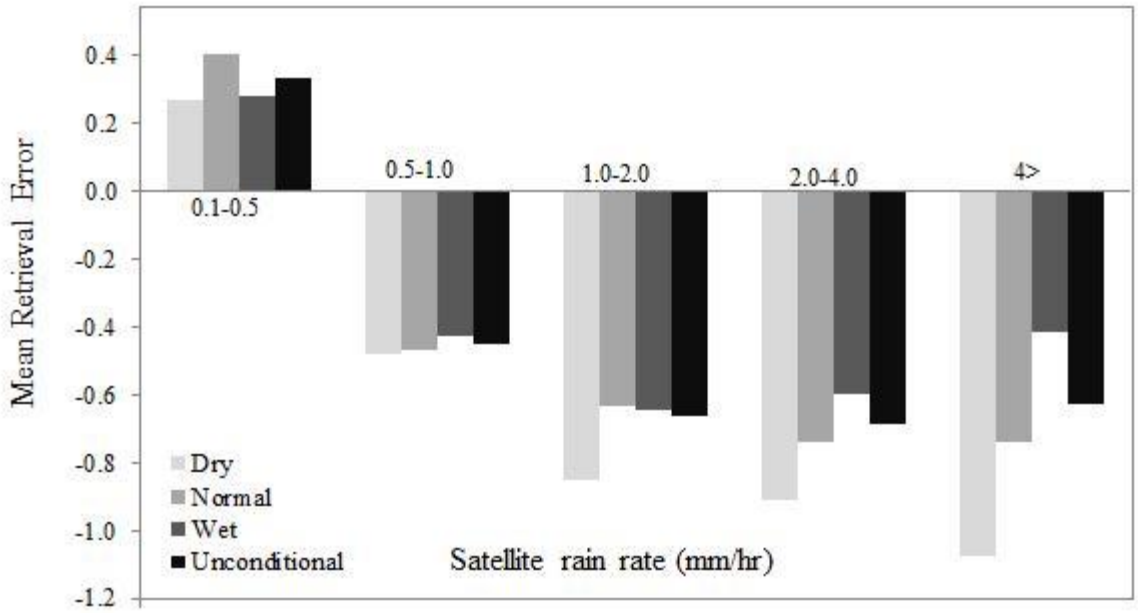


Fig. 2.8: Mean of 2A12 retrieval error vs. rainfall threshold for wet, normal and dry soils and unconditional.

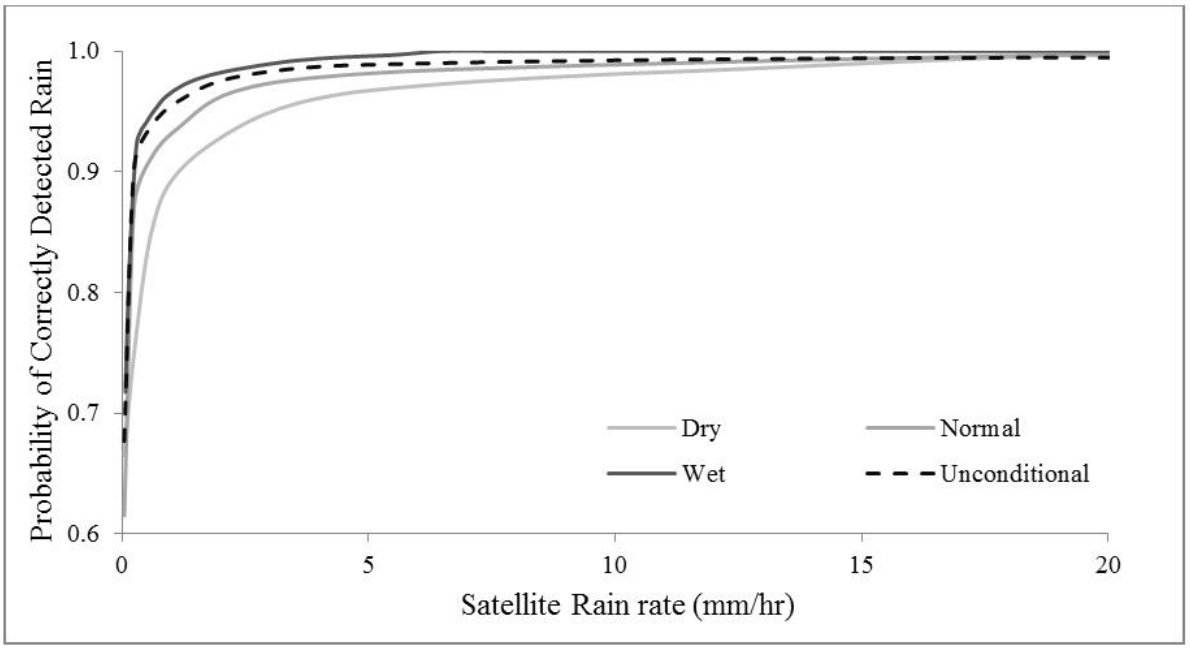


Fig. 2.9 : Probability of correctly detected rain vs. rainfall threshold for wet, normal and dry soils and unconditional.

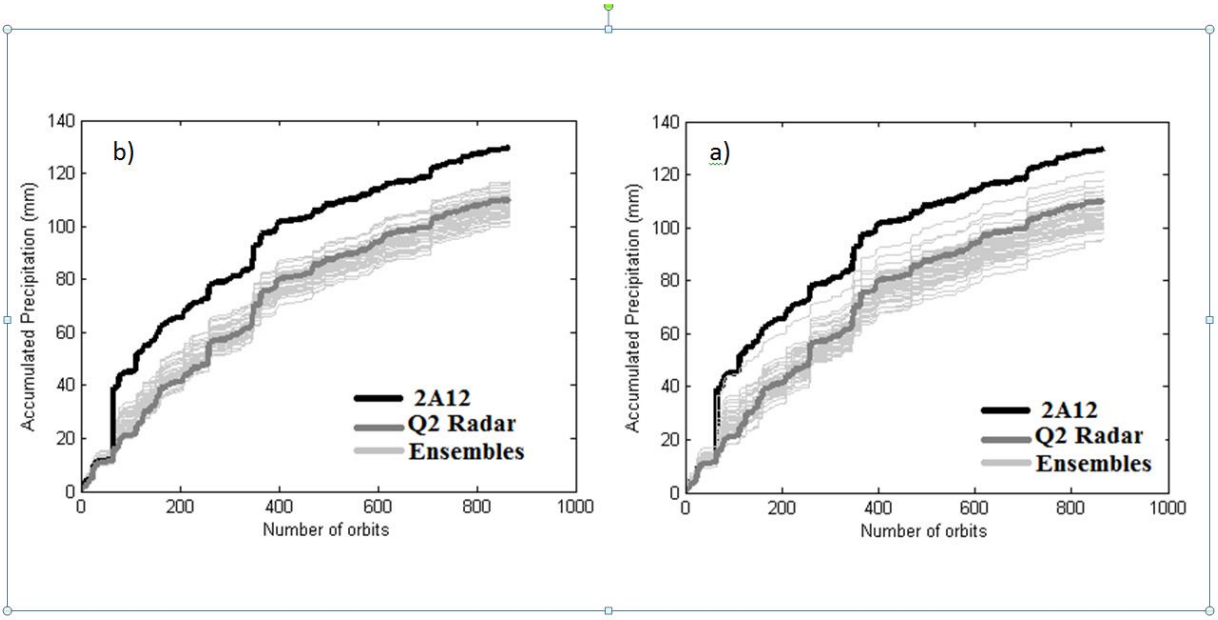


Fig. 2.10 Time series of cumulative rainfall of Q2, 2A12 and SREM2D ensemble members. In panel (a) SREM2D parameters are calibrated with the conditional approach and in panel (b) SREM2D parameters are calibrated with the unconditional approach.

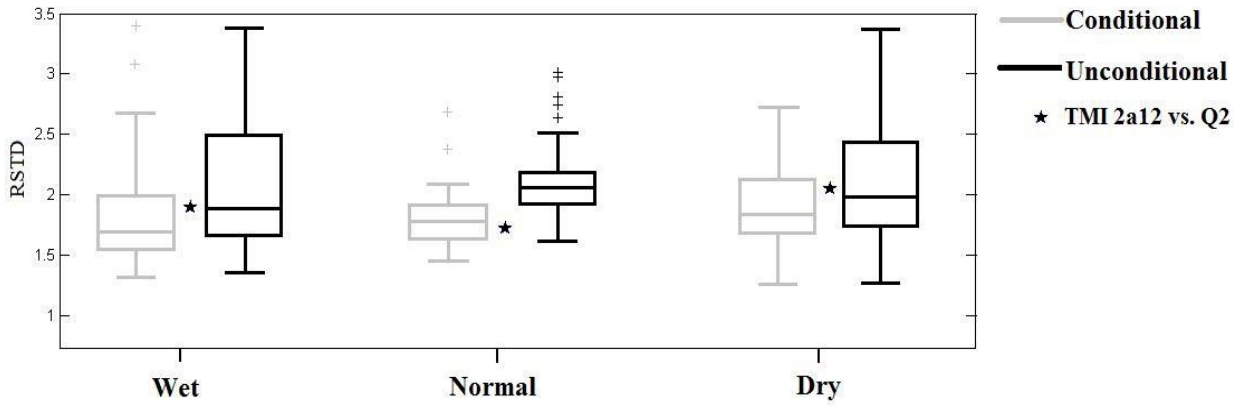


Fig. 2.11: Relative standard deviation (RSTD, unitless) of actual and SREM2D generated ensemble members for 2A12 as function of wet, normal and dry soil conditions.

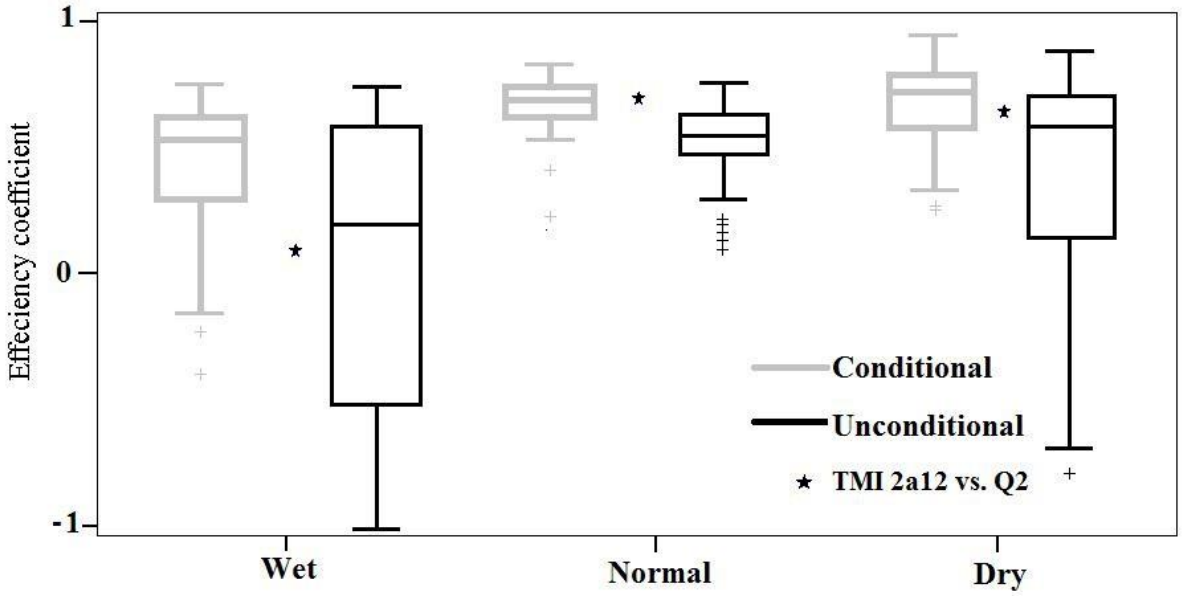


Fig. 2.12: Nash-Sutcliffe efficiency coefficient of actual and SREM2D-generated ensemble members for 2A12 in wet, normal and dry soil conditions.

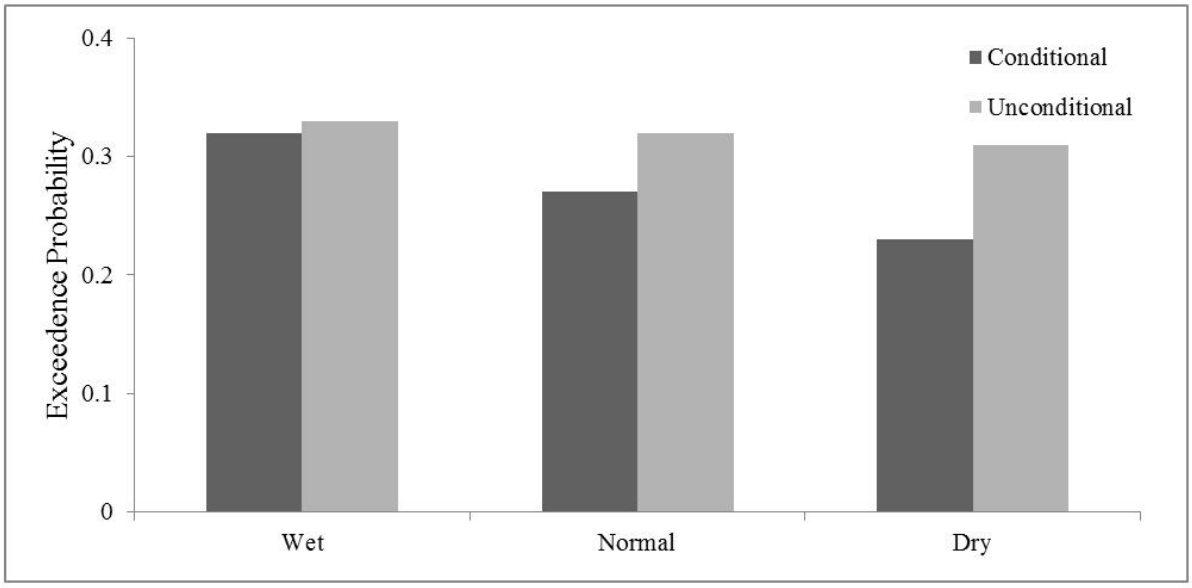


Fig. 2.13 : Exceedance probability of generated ensemble members of 2A12 in wet, normal and dry soil conditions based on conditional and unconditional approaches.

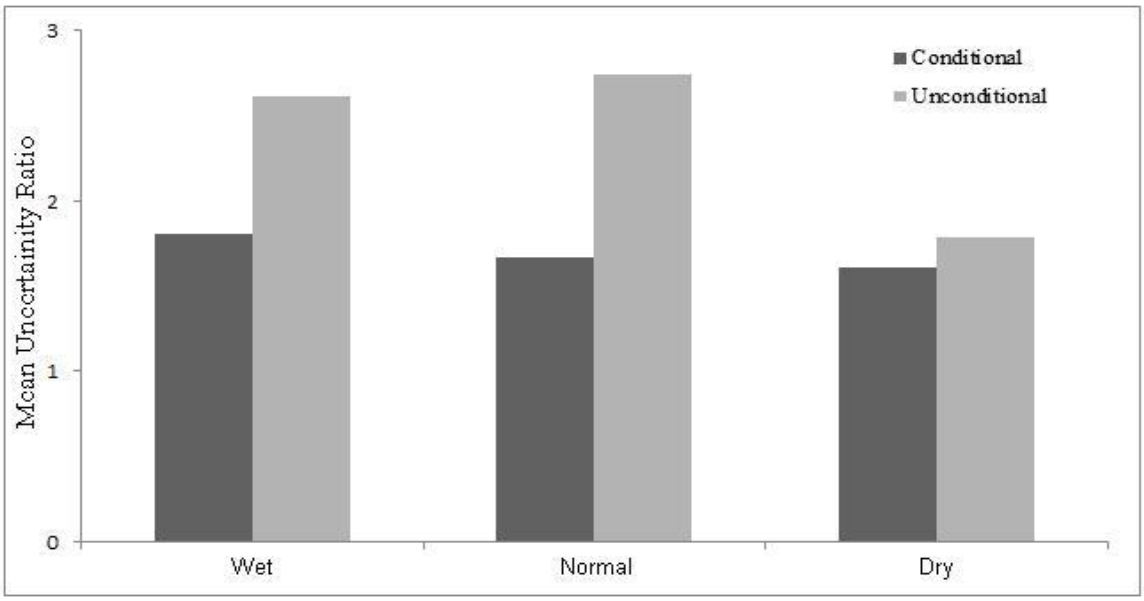


Fig. 2.14: Mean uncertainty ratio of generated ensembles of 2A12rainfall for wet, normal and dry soil conditions based on conditional and unconditional SREM2D parameter calibration.

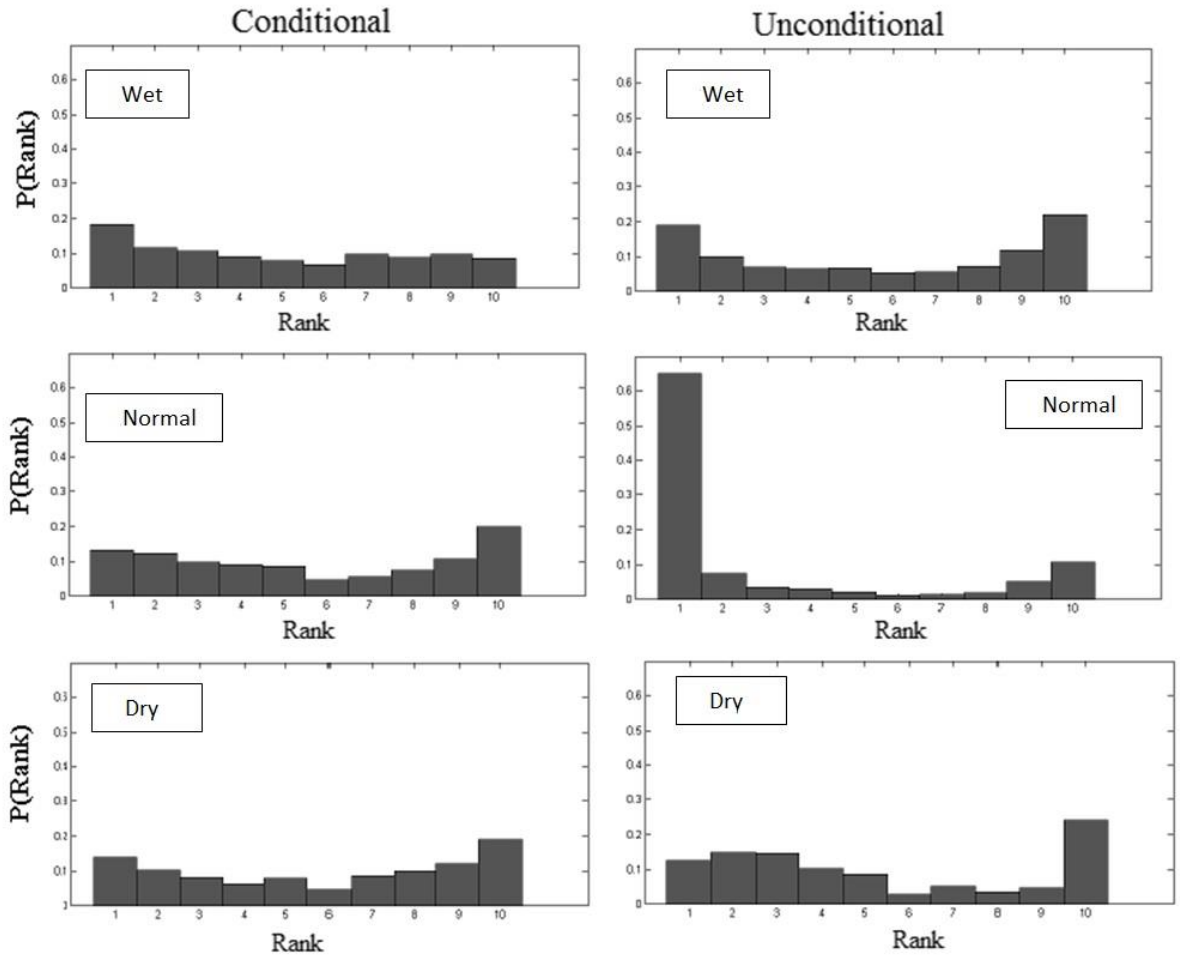


Fig. 15: Rank coefficient of generated ensemble members of 2A12 in a) wet, b) normal and c) dry soil conditions for conditional and unconditional approach.

# 3. Hydrologic Evaluation of the Satellite and Re-analysis Precipitation Datasets over Mid-Latitude Basins

## 3.1. Introduction

Developing global hydrologic modeling systems that can support flood warning and flood risk analysis studies has been an aim of research efforts in hydrology (Hong et al., 2007;Shrestha et al., 2008;Wu et al., 2012). Modeling a basin's flood response requires an accurate spatio-temporal characterization of precipitation variability within the basin. Estimates of precipitation at sub-basin scales are typically based on weather radar observations and/or gauge network measurements (Ogden et al., 2000;Vivoni et al., 2006). However, recent advances in multi-satellite rainfall retrievals have allowed uses of high-resolution satellite rainfall products in flood modeling applications (Bitew and Gebremichael, 2011a;Yong et al., 2012). The latter is particularly important by the fact that traditional ground-based observations have significant spatial coverage gaps over remote and ungauged regions of earth (Asadullah et al., 2008;Dai et al., 2007;Hossain and Huffman, 2008;Sapiano and Arkin, 2009). Some of the high resolution, near-global scale, multi-satellite-sensor rainfall products include TRMM3B42V7, TRMM3B42RT (Huffman et al., 2007), CMORPH (Joyce et al., 2004), PERSIANN (Sorooshian et al., 2000), Hydro-estimator (Scofield and Kuligowski, 2003), NRLBLD (Turk and Miller, 2005), and GSMaP (Ushio and Kachi, 2009). These products use different combinations of information from geostationary infrared (IR) and low earth orbit satellite microwave (MW) observations. The satellite rainfall product used in this study is the TRMM3B42V7 that is available at a 3 hourly temporal resolution and approximately 25-km spatial resolution. The TRMM3B42 algorithm suite uses more accurate, but infrequent, MW rainfall estimates to calibrate a rainfall algorithm applied on the less direct, but frequent, IR observations. The

merging technique then uses the IR-based rainfall estimates to fill in gaps where MW data are not available. The TRMM3B42RT product is available in near real-time, while the TRMM3B42V7 is a gauge adjusted product available with one month latency.

On the other hand, global retrospective analyses (reanalysis) products can provide long-term hydrologic datasets that can support global frequency analyses of hydrologic extremes (e.g. floods, droughts). Widely used reanalysis products include a 44-year reanalysis from the Global Data Assimilation System (GDAS) from the National Center for Environmental Prediction (NCEP) (Kalnay et al., 1996), a 40-yr re-analysis (ERA-40) from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Bosilovich et al., 2008;Uppala et al., 2005), ERA-Interim (Dee et al., 2011), and a 35-year reanalysis from the Global Land Data Assimilation System (GLDAS) of NASA's Goddard Space Flight Center (GSFC) (Rodell et al., 2004). A typical reanalysis system consists of two main components, the forecast system and the data assimilation system. The role of data assimilation system is merging available observational data with the forecast model simulations. The GLDAS precipitation data used in this study consist of outputs from a regional climate model plus spatially and temporally integrated datasets from radar, rain gauges and satellite observations. They are available at hourly temporal scale and approximately 100 km spatial grid resolution.

Numerous satellite rainfall validation studies have been carried out to provide a better and deeper understanding about the uncertainties associated with the remotely sensed precipitation products over different regions (Adler et al., 2001;AghaKouchak et al., 2009;Brown, 2006;Dinku et al., 2007;Ebert et al., 2007;Gottschalck et al., 2005;Krajewski et al., 2000;McCollum et al., 2002;Stampoulis et al., 2013;Su et al., 2008;Tang et al., 2010). Studies have shown that the accuracy of satellite rainfall products depends on the rainfall type (e.g.

convective vs. stratiform), topography, and climatological factors; for instance CMORPH has been shown to underestimate precipitation during hurricane Wilma (Turk et al., 2006), and overestimate precipitation over both South (Demaria et al., 2011) and North America (AghaKouchak et al., 2011).

In addition to the rainfall error analysis, studies have investigated accuracies of hydrological modeling forced with satellite rainfall data. (Behrangi et al., 2011) investigated the feasibility of simulating streamflows for mid-size basins by forcing a hydrologic model with different satellite based precipitation products (TRMM3B42V7 and RT, CMORPH, PERSIANN). They concluded that bias-adjustment of satellite products has a significant impact on capturing streamflow patterns and magnitude. (Beighley et al., 2011) predicted streamflows by forcing the Hillslope River Routing (HRR) hydrologic model with three satellite derived precipitation datasets (TRMM3B42V6, CMORPH, PERSIANN) over the Congo Basin. They argued that all three satellite products are unreasonably overestimating over equatorial regions. However, the TRMM3B42V6 product exhibited the best performance in terms of rainfall data quality and simulated streamflows. (Su et al., 2008) evaluated the skill of streamflow simulations from a semi-distributed hydrology model driven with TRMM3B42V6 rainfall data vs. simulations driven with rain gauge rainfall measurements. Their research demonstrated a good agreement between reference and satellite precipitation data for streamflow simulation at seasonal and inter-annual time scales, although there was an overestimation at the daily time scale. They recommended TRMM3B42V6 for hydrological simulations in ungauged areas. (Gourley et al., 2011), on the other hand, evaluated rainfall estimates from TRMM3B42V6 and PERSIANN-CCS in comparison to radar rainfall estimates for hydrological simulations over Ft. Cobb basin (342-$km^2$). The study highlighted the importance of considering rainfall products

resolution on hydrologic model calibration. They also demonstrated TRMM3B42V6 has relatively better performance than PERSIANN-CCS. However the study was carried out over a small watershed as well as short period of time (3 months). (Vergara et al., 2013) underlined the effect of rainfall data resolution and basin scale on hydrological simulations using TRMM3B42RT and MPE (Multi-sensor Precipitation Estimator) radar data over 5 medium size sub-basins. The study was limited to only one product and a small to medium range of basin scales (500-5000 $km^2$). (Maggioni et al., 2013) investigated the error propagation from TRMM3B42 (V7 and RT), CMORPH, and PERSIANN-CCS precipitation estimates on runoff simulations. The study was over a 2 year period and focused on small to medium size basin scales (500-5000 $km^2$). The results demonstrated significant dependency of error propagation to catchment area. (Nikolopoulos et al., 2012) evaluated TRMM3B42V6, CMORPH, and PERSIANN-CSS precipitation estimates for a major flash flood event simulation through forcing a distributed hydrologic model. They demonstrated that the examined products do not perform satisfactorily on capturing the flood peak from this complex terrain heavy precipitation event. (Bitew and Gebremichael, 2011b) in an earlier study, also evaluated various global satellite precipitation products for stream flow simulations over two small basins in Ethiopia through forcing a semi-distributed hydrologic model. They demonstrated that TRMM3B42RT and CMORPH are performing better than TRMM3B42V6 and PERSIANN. They argued that gauge adjustment in TRMM3B42V6 made the data set worse than the real-time (unadjusted) product. In general the product resolution, catchment size and hydrologic model calibration are influential parameters on quality of generated runoff through gridded precipitation products. Most of the above studies have been limited to small or mid-size basins and short simulation periods, while only few of them have considered heavy precipitation events. More in-depth analysis of the

precipitation error propagation in streamflow simulations accounting for more comprehensive ranges of basin scales and event severities is needed to demonstrate utility of satellite precipitation product in flood modeling.

In this study, the TRMM3B42V7 satellite rainfall product and the GLDAS reanalysis precipitation datasets are evaluated against multi-year (2001-2010) high-quality and resolution radar-rainfall data over multi-scale (39 – 70,000 $km^2$) mid-latitudes basins in the Northeast United States to assess their applicability in basin flow simulations. Streamflow simulations are based on the Hillslope River Routing (HRR) model, which has been used in several regional hydrologic modeling applications of satellite rainfall (Beighley et al., 2009;Beighley et al., 2011). The study provides a new insight on the combined effect of precipitation product (spatial resolution and accuracy) and basin scale in the error propagation from precipitation to flood prediction. The focus of the error analysis is on moderate to heavy precipitation events, which are defined as events falling between the $75^{th}$ and $90^{th}$ and greater than $90^{th}$ percentiles of events that occurred over the study area from 2001-2010, respectively. Comparisons between the satellite dataset (TRMM3B42V7) and the GLDAS reanalysis provide an insight of the expected benefits from using current near-global-scale satellite products, relative to the longer-term but coarser resolution global reanalysis precipitation datasets, in terms of hydrologic simulations.

In the next section we present the study area and data, while section 3 presents the implemented data processing and matching techniques. Section 4 describes the hydrologic modeling framework. Finally, section 5 presents the error analysis and hydrologic error propagation results. Conclusions and recommendations are presented in section 6.

## 3.2. Study area and data

The study area is the Susquehanna River Basin in United States (ranging from 39N to 43N and 75W to 79W, Figure 1). The region has an elevation gradient from north to south-east, with the highest peak in northwestern corner and the lowest point in the southeastern corner. The total area of Susquehanna River Basin is 71,000 $km^2$ of which 76% is in Pennsylvania, 23% in New York, and 1% in Maryland. In this study, catchment land areas draining laterally to individual river reaches range from 0.1 to 475 $km^2$ with a mean catchment area of 70.6 $km^2$. Cumulating the drainage areas along the river network at the outlet of each individual catchment provides 1006 unique watersheds with drainage areas ranging from 39 to 71,000 $km^2$. The Susquehanna River Basin is one of the most flood prone areas in the entire country, experiencing a major devastating flood on average every 14 years. The basin's average annual flood damage is $150 million dollars (Susquehanna River Basin Commission, http://www.srbc.net/). The time frame of this study is January 2002 to December 2011. To evaluate the dependency of statistical metrics to the basin size, the identified sub basins (1006) were grouped into six categories (Table 1) from very small to very large. Below we discuss the datasets used in this study, which include the reference radar rainfall data, the satellite rainfall products and the reanalysis dataset.

### 3.2.1. Radar Rainfall

The NCEP stage IV radar-rainfall product (Lopez, 2011) was used to represent the high-accuracy reference rainfall dataset. The data is originally in polar-stereographic grid with 4×4km spatial resolution. It combines estimates of precipitation from about 150 Weather Surveillance Radar-1988 Doppler (WSR-88D) radars and approximately 5500 hourly rain gauge rainfall measurements over the CONUS (Contiguous United States). The hourly accumulated rainfall

estimated through the Z-R relationship for each radar is merged with other radars in WSR-88D network using an inverse-distance weighting (IDW) scheme. The analysis of precipitation at each NCEP stage IV gridded data begins 35 minutes after each hourly collection period ends. The updates may continue for subsequent hours using data from the 12 CONUS RFCs (River Forecast Centers). The stage IV data benefits from the RFCs manually quality-control (QC).

### 3.2.2. TRMM3B42V7

TRMM3B42V7 is a 3-hourly combined microwave-infrared precipitation product (Huffman et al., 2007) at the spatial resolution of 25×25 km and 3-hourly time interval. There are a couple of updates for TRMM3B42V7 products and the one utilized in this study is the most updated version. The TRMM3B42V7 blending scheme uses a set of passive microwave (PMW) low earth orbit (LEO) data including TMI (the Microwave Imager on TRMM), SSM/I (Special Sensor Microwave Imager) on DMSP (Defense Meteorological Satellite Program) satellites, AMSR (Advanced Microwave Scanning Radiometer-Earth Observing System) on Aqua, AMSU-B (the Advanced Microwave Sounding Unit-B) on NOAA (the National Oceanic and Atmospheric Administration) satellite series, and MHS (Microwave Humidity Sounders) on later NOAA-series satellites and the European Operational Meteorological (MetOp) satellite (Kidd et al., 2011). The TRMM3B42V7 combination scheme uses the Goddard profiling (GPROF) algorithm (Kidd et al., 2011;Kummerow et al., 2001;Kummerow et al., 1996;Olson et al., 1999;Wang et al., 2009;Gopalan et al., 2010) for estimating precipitation from PMW imagers (TMI, SSM/I, and AMSR). Over land and coastal surface areas the algorithm reduces to a scattering-type procedure using only the higher-frequency channels. (Lensky and Levizzani, 2008) indicate that PMW retrieval algorithms are suffering from uncertainties in precipitation

type classification and temporal resolution over land surfaces. Precipitation estimates from PMW sounders (AMSU-B, and MHS) are derived using the IWP (ice water path) technique (Weng et al., 2003;Zhao and Weng, 2002).

The PMW calibrated infrared (IR) precipitation products (Janowiak et al., 2001) from GEO (Geosynchronous Earth Orbit) satellites are used to fill in the PMW gaps in three-hourly time intervals. Specifically, the algorithm takes the value of the PMW-calibrated IR precipitation products when the PMW is not available in a 3-hourly product time step. The TRMM3B42V7 uses gridded monthly ground based precipitation gauge data for bias removal and calibration. Since IR estimates are considered highly uncertain outside 50° N-S, the TRMM3B42V7 data extent is restricted to 50N-50S.

### 3.2.3. GLDAS

The GLDAS rainfall dataset is available at 1 deg spatial resolution and 3 hourly time intervals. GLDAS precipitation data exhibit global coverage and long record (since 1979), which is suitable for flood risk analysis studies. The data is "observation based", which come from a combination of NCEP's GDAS (Global Data Assimilation System) model outputs, CPC's (NOAA Climate Prediction Center's) CMAP (Merged Analysis of Precipitation) precipitation, and Air Force's AGRMET (AGRicultural METeorological modeling system) radiation datasets (Rodell et al., 2004). GDAS assimilates global in situ meteorological observations and satellite observations. CMAP is merged satellite (IR and MW) and rain gauge analysis (Xie and Arkin, 1997).AGRMET radiation fields are satellite observation based. GLDAS therefore represents merged, spatially and temporally interpolated fields of GDAS, CMAP, and AGREMET radiation fields.

## 3.3. Data processing

To make the datasets comparable, all data were aggregated to the TRMM3B42V7 temporal resolution (3hr), which is the lowest compared to radar and GLDAS. The next step was to extract precipitation values for each sub-basin based on the different precipitation products. For this purpose the gridded precipitation data from different sources were overlaid with the sub-basins (see figure 2) and corresponding product cell IDs to each sub-basin were identified. Besides the cell IDs, the percentage of coverage (α) for each cell within a given sub-basin was also determined using GIS. The area averaged precipitation value over each sub-basin, $P_k$, was calculated based on the following equation:

$$P_k = \sum_{i=1}^{n} \alpha_{k,i} P_i \quad (1)$$

where "$P_i$" refers to the precipitation value from a given product (e.g. TRMM3B42V7) for cell $ID_{k,i}$, $\alpha_{k,i}$ is the percentage of land area within sub-basin *k* covered by grid cell $ID_{k,i}$. Index "i" is the cell number and "n" is the number of cells that overlaid the sub-basin. The $P_k$ values are used as precipitation forcing in the HRR model and are hereafter called "basin averaged" values. The resulting "basin averaged" values will be used for error analysis between the two precipitation products and reference.

## 3.4. Hydrological Modeling

The Hillslope River Routing model (Beighley et al., 2009;Beighley et al., 2011) is used in this study for streamflow simulations. HRR integrates a water balance model for the vertical fluxes and a routing model for the horizontal routing of the surface/subsurface runoff. For each model unit, the landscape is approximated as an open book with two planes draining laterally to a main channel. Flow routing is performed using variants of the kinematic wave method for the

hillslopes on both the surface and shallow subsurface, and diffusion wave methodologies (i.e., Muskingum Cunge) for channels and floodplains.

In this application, the river network and corresponding catchments were derived from SRTM elevations using ArcGIS. A threshold area of 50 $km^2$ was used to define the river network (i.e., only rivers draining 50 $km^2$ or more are simulated). Figure 1 shows the river network used based on the above area threshold. For each catchment, the hillslope length was determined by approximating the catchment as two equally sized rectangular planes draining laterally to the river segment, where hillslope length is half the catchment area divided by its channel length. The slopes of both the channel and planes were determined by averaging the individual pixel slopes within the catchment or along the river reach.

To separate rainfall into surface runoff and infiltration, the Green-Ampt model was used. For the vertical water balance, mean monthly Evapotranspiration (ET) rates from GLDAS (Rodell et al., 2004) averaged over each catchment were used. For runoff generation and vertical water balance, the required parameters were estimated based on available datasets and then averaged within each model unit. For example, saturated hydraulic conductivity, Ksat, and the soil suction head, $\Psi$, were estimated using soil particles sizes (FAO, 2012) and lookup tables (Rawls W.J. et al., 1982;Saxton et al., 1986) to relate soil particle sizes (percent sand, silt and clay) with hydrologic quantities (Ksat and $\Psi$). Impervious areas within each basin were estimated from the 2006 National Land Cover Database available from the USGS (Fry et al., 2011).

In addition to the base model parameters obtained from soils and land cover data, three parameters were calibrated: horizontal conductivity, Kh, for the subsurface routing, overland roughness, N, for surface routing, and Manning's roughness, n, for channel routing. These parameters are scale dependent in that they capture both the hydraulic features (river reach and

hillslope lengths) defined for a given model unit as well as all sub-model unit features not represented at the defined model scale (e.g., all tributaries not explicitly represented in the defined river network). The calibration was performed by systematically adjusting the three parameters (Kh, N, n) to achieve zero mean relative error (ME, m3/s) for the annual maximum peak discharges at the nine streamflow gauging stations shown in Figure 1:

$$ME = \frac{1}{m}\sum_{n=1}^{m}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{Q_{s,i,n} - Q_{g,i,n}}{Q_{g,i,n}}\right) \quad (1)$$

where Qs,i,n and Qg,i,n (m3/s) are the $i^{th}$ simulated and measured, respectively, annual maximum peak discharge at gauge location n; k is the number of annual maximum peak discharges is the study period (k=9: 2003-2011, as 2002 was used for model spin-up), and m is the number of gauge locations being considered (m = 9).

The mean (averaged over the 1006 basin units) calibrated model parameters are: Kh = 0.004 m/s, N = 17.3, n = 0.045. Note that, the magnitudes of calibrated parameters illustrate their scale dependence. For subsurface routing, the model assumes that water remains below ground for the entire hillslope length. However, it is likely that the subsurface water discharges to a tributary channel and travels the remaining distance as channel flow. Thus, Kh must be much larger than typical conductivity values. For surface runoff, the idealized planes do not account for flow sinuosity, surface storage or downslope infiltration, which all act to significantly attenuate the surface response. For example, a sinuosity value of 10 for surface runoff traveling from a drainage divide to a main channel, would yield an effective N value of 1.7, which when considering the scale of the model units and lack of surface storage seems reasonable for surface runoff from largely undeveloped mountainous hillslopes. Similarly, for channel roughness, considering a river sinuosity of 1.5 would yield a roughness value of 0.03, which would be more

typical for river channels in the SRB. Future research is planned for the HRR model to quantify the scaling relationships for these parameters to guide the calibration required.

In terms of overall model performance, Figure 3 shows the calibrated model results at each gauge as a function of gauge drainage area (Table 2). The overall ME is zero with mean relative errors for individual gauges ranging from -16% to 23%, and errors for individual events ranging from -62% to 224%. Note that, the largest error is from the gauge draining the smallest basin area (1,155 $km^2$) in 2011 during Tropical Storm Lee, which caused significant flooding especially in the northern SRB. The overestimated discharge at this gauge may have been a result of its relatively small area and the lack of flood control infrastructure simulated in HRR (i.e., no reservoirs). Overall, 86% of the errors are within ± 50%, and approximately half are within ±25%. In addition to annual peak discharges, Figure 4 shows select periods for several gauges ranging from the smallest to largest drainage areas to illustrate the generally favorable model performance.

For this study, the goal of the calibration process was to produce reasonable discharges for assessing the effects of precipitation datasets. The annual maximum discharges were selected for the calibration process because of our focus on event rainfall. The resulting ME of zero with minimal variations in error across basin scales and from year-to-year (Figures 3 and 4) suggests that the calibrated model forced with radar rainfall performs well and provides a reasonable framework from which to evaluate different precipitation datasets.

## 3.5. Error Analysis Framework

The experiment setup for the study is presented in Figure 5.The framework consists of two main components, reference driven simulation and product driven simulation. The time series of

basin averaged rainfall from TRMM3B42V7 (25km, 3hr) and GLDAS (100km, 1hr) are used to force HRR for the product driven simulations. Likewise the time series of basin averaged rainfall from stage IV (4km, 1hr) is used to force HRR for the reference driven simulations. The error analysis consists of two components: one for the precipitation products against the reference precipitation data and the second for the precipitation product-driven simulated runoff time series against the reference precipitation data-driven simulated runoff time series. The error metrics used in this study are (i) mean scale quantile relative error (QRE) and (ii) the quantile root mean square of error relative to the mean of reference (QRMSE). To evaluate the dependence of error metrics to the magnitude, we determined the above statistics for matched product/reference datasets categorized in two groups according to the exceedance probability of the rainfall and runoff reference values. The ranges of values for the two groups are defined according to the 75$^{th}$ to 90$^{th}$ percentile and above 90$^{th}$ percentile representing moderate and extreme precipitation and flood events.

Following the above, QRE is defined as the ratio of sum of differences between reference and sensor values (precipitation or runoff) to the sum of reference values determined over the sub-basins for each quantile range:

$$QRE = \frac{\sum_{i=1}^{n}((P_{Sensor}^{s}\left| \boldsymbol{t_j \le P_{ref}^{s} < t_{j+1}} \right.) - (P_{ref}^{s}\left| \boldsymbol{t_j \le P_{ref}^{s} < t_{j+1}} \right.))}{\sum_{i=1}^{n}(P_{ref}^{s}\left| \boldsymbol{t_j \le P_{ref}^{s} < t_{j+1}} \right.)} \qquad (2)$$

where $P_{Sensor}^{s}$ is the sensor "basin averaged" precipitation/runoff value, $P_{ref}^{s}$ is the reference "basin averaged" precipitation/runoff value over the sub-basin, *t* is the threshold value which is based on the reference data quantiles, *j* is the quantile index, and *n* is total number of value in a particular scale and quantile range. The perfect value for this metric is zero, which means there is

no difference between reference and the sensor values. Negative QRE value means the sensor is underestimating and the positive value means overestimating.

QRMSE is the root mean square of the differences between reference and sensor; it is normalized to the mean of reference values.

$$QRMSE = \frac{\sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}\left((P^{s}_{Sensor}\middle|\, t_j \le P^{s}_{ref} < t_{j+1}) - (P^{s}_{ref}\middle|\, t_j \le P^{s}_{ref} < t_{j+1})\right)^2}}{mean\,(P^{s}_{ref}\middle|\, t_j \le P^{s}_{ref} < t_{j+1})} \qquad (3)$$

where $P^{s}_{Sensor}$ is the sensor "basin averaged" precipitation/runoff value, $P^{s}_{ref}$ is the reference "basin averaged" precipitation/ runoff value over the sub-basin, *t* is the threshold value which is based on the reference data quantiles, *j* is the quantile index, and *n* is total number of value in a particular scale and quantile range. QRMSE quantifies the spread between sensor and reference data points.

## 3.6. Results

In this section we discuss results from the evaluation of the satellite based (TRMM3B42V7) and re-analysis (GLDAS) global precipitation datasets based on the data and error metrics discussed earlier. In the following sub-section we provide the quantitative statistics of QRE and QRMSE aimed to evaluate the performance of precipitation datasets (satellite and re-analysis) and the generated runoff values relative to the high-resolution reference radar-rainfall data. In sub-section 6.2 a frequency distribution analysis is presented to demonstrate the value of the above products in determining the frequency of occurrence of rainfall and runoff extremes for the different basin scales.

### 3.6.1. Quantitative Statistics

QRE as a function of basin scale for TRMM3B42V7 and GLDAS precipitation (top)/runoff (bottom) values are presented in Figure 6. Both products exhibit precipitation underestimation for all basin scales. The underestimation is stronger for the GLDAS product and for heavy precipitation (values greater than the $90^{th}$ percentile). There is less underestimation with increasing basin scale for both moderate (values between $75^{th}$ and $90^{th}$ percentile) and heavy precipitation. The trend is stronger TRMM3B42V7, particularly for the larger basins (exceeding 10,000 $km^2$). A comparison of the precipitation products' mean relative errors is summarized in Table 3 as the ratio of the GLDAS to the TRMM3B42V7 QRE metrics for three basin scale categories and for moderate or heavy precipitation values. It is shown that the error metric ratio is systematically greater than one across all table categories, indicating greater mean relative error for the GLDAS product. This ratio increases with increasing basin scale; it is between 1.3 and 2.86 for the moderate precipitation, and 1.15 to 1.90 for the heavy precipitation. In terms of the corresponding QRE of the simulated runoff (Figure 6b), the basin scale trend is not apparent as it was in the precipitation. Only in the TRMM3B42V7 simulated runoff we can see a slight reduction of the QRE (reaching almost unbiased simulations) for the largest basins. The main observation from the runoff simulation QRE metric in Figure 6b is the significant difference between the values of the two products. Specifically, the GLDAS runoff simulation QRE values are consistently below -0.5, while the TRMM3B42V7 QRE values are mostly above -0.1 for the moderate (between the $75^{th}$ and $90^{th}$ percentiles) flows and above -0.3 for the high (greater than the $90^{th}$ percentile) flows. This aspect is also presented in Table 3, which summarizes the simulated runoff QRE ratios between GLDAS and TRMM3B42V7. As noted in the table, the

QRE ratios are greater than 7 for moderate flows at all basin scales, and between 3 and 7 for the high flows.

For the random component of the precipitation error (QRMSE- Figure 7) the observations are similar to what we noted for the QRE metric. The random error decreases for both products as a function of basin scale, decreasing more strongly for the heavy precipitation events. Overall, TRMM has lower random error than GLDAS (particularly for moderate precipitation values) but for heavy precipitation the difference is very small. Table 4 presents the ratios of GLDAS to TRMM3B42V7 precipitation products' QRMSE error metrics. Again the ratio values are consistently greater than one indicating higher random error associated with the GLDAS product. The QRMSE ratios are greater for moderate precipitation values and range between 1.18 and 1.28, while the corresponding range of ration values for heavy precipitation is 1.02 to 1.09. In terms of the simulated runoff values (Figure 7b), the differences in runoff between TRMM3B42V7 and GLDAS are more significant than in precipitation, where TRMM3B42V7 simulated runoff exhibits stronger reduction of the random error as a function of basin scale. Conversely, the QRMSE trend for the GLDAS runoff simulations is very weak. For both products, random error does not depend on runoff magnitude. The runoff simulation QRMSE ratios between the two products are summarized in Table 4. Again, ratios are consistently greater than one, while values tend to be greater for the largest basins, namely 1.71 for moderate runoff values and ~3 for high runoff values.

### 3.6.2.Frequency Distribution

The cumulative probability of precipitation and simulated runoff values from GLDAS, TRMM3B42V7 and stage IV radar data are presented in Figure 8. Specifically, x-axis is

representing the precipitation or normalized runoff (runoff normalized to basin area) values in $mmhr^{-1}$. The Cumulative Density Function (CDF) plots are generalized on 3 main groups based on the basin scale: small, medium and large. The cumulative probability for precipitation data shows that TRMM3B42V7 CDF is getting closer to the reference radar data CDF going from small to large basin scales. However the strong underestimation from GLDAS does not show any significant improvement with basin scale. The simulated normalized-runoff CDFs are similar to the precipitation data CDFs, where TRMM3B42V7 is close to the reference radar data and GLDAS exhibits strong underestimation. The corresponding quantiles derived from the CDF values are presented in Figure 9. The horizontal axis represents the sensor data quantiles, while the vertical axis represents the radar data quantiles in $mmhr^{-1}$. The Q-Q plots for precipitation shows that TRMM3B42V7 is much closer to the 45-degree line compared to GLDAS, which is consistent at all basin scales. This agreement is more pronounced for the TRMM3B42V7 runoff simulated data. GLDAS on the other hand exhibits strong underestimation for all quantiles and basin ranges.

### 3.6.3.Error Propagation

The higher GLDAS to TRMM3B42V7 error metrics ratios in simulated runoff compared to precipitation are investigated next by analyzing the precipitation-to-runoff error propagation of the two products. Tables 5 and 6 present the ratios of runoff to precipitation QRE and QRMSE error metrics, respectively, for the two products, two quantile categories, and three basin scale ranges. The QRE ratio comparison of the two products in Table 5 indicate significant dampening of the mean relative error from precipitation to simulated runoff for TRMM3B42V7 and the opposite (increase) for GLDAS. This difference is greatest for the largest basin scales (> 31,500

km$^2$) and moderate runoff values (75$^{th}$ to 90$^{th}$ percentiles). The largest basins exhibited the greatest magnification of the GLDAS product's mean relative error with values between 1.88 (for moderate flows) and 1.4 (for extreme flows). The corresponding mean relative error ratios for TRMM3B42V7 are 0.03 and 0.31, which indicates significant difference in the behavior of the two products error propagation in runoff simulations. Vergara et al. (2013) has attributed this difference to the products' resolutions, which affects the way the hydrologic model partitions precipitation to infiltration and runoff. In terms of QRMSE (Table 6), the behavior of the two products is similar for moderate flows and small to medium scale basins. The largest basins show significant deviations between the two products' random error metric ratios; namely, error dampening of 0.6 (TRMM3B42V7) versus 0.8 (GLDAS) for moderate flows, and dampening of 0.5 (TRMM3B42V7) versus error increase of 1.33 (GLDAS) for the extreme flows. The overall observation from this error propagation analysis is that the strong underestimation of heavy rain rates and the resolution effect associated with the GLDAS precipitation products further amplifies the error properties of this dataset when used in river flow simulations. On the other hand, the bias-adjusted TRMM3B42V7 satellite precipitation product has error properties that are mainly smoothed through a distributed hydrologic model in simulating medium to large-scale basin flows (>3,000 km$^2$).

## 3.7. Conclusions

This study was aimed at providing a comprehensive investigation of two contrasting precipitation datasets (a gauge-adjusted satellite product versus a data-driven model reanalysis dataset) in terms of the combined effects of product error properties, basin scale and event severity in the error propagation from precipitation to basin flow simulations. The TRMM multi-

satellite and gauge-adjusted precipitation product and the GLDAS reanalysis precipitation dataset were used to force a distributed hydrologic model (HRR) for moderate and heavy precipitation events (namely, values in the range of $75^{th}$ to $90^{th}$ and greater than the $90^{th}$ percentiles of a 10-year record) over the Susquehanna River basin in Northeast US. High-resolution gauge-adjusted radar-rainfall estimates (Stage IV) were used to derive the basin-average reference precipitation and to simulate the reference basin flows. The two products were compared in terms of two error metrics (mean relative error and relative root mean square error) and frequency distribution analysis, based on matched product vs. reference data grouped in two quantile categories (Q75-Q90 and >Q90) and different basin scales. A precipitation-to-simulated runoff error propagation analysis for the two products was conducted to explain differences of the two products' error metrics.

The results showed that TRMM3B42V7 exhibits significantly better error statistics than the GLDAS re-analysis product in terms of basin average rainfall and simulated runoff values with respect to the reference radar-rainfall dataset. The differences in the error statistics are more significant in the basin runoff simulations than basin average precipitation. Specifically, the ratios of quantile relative error (QRE) in simulated runoff through GLDAS to the simulated runoff through TRMM3B42V7 increased to about 7 and 3 times for moderate ($75^{th}$ to $90^{th}$ quantile) and high flow values (above $90^{th}$ quantile), respectively. The comparison of cumulative probabilities of rainfall and normalized-runoff supports these results, showing significant underestimation of the GLDAS rainfall and runoff values (relative to the radar and satellite cumulative distributions) at all basin scales. The error propagation analysis showed that the GLDAS product exhibits significant increase of the mean relative error going from precipitation to runoff, while the opposite is true for TRMM3B42V7. The GLDAS random error component

propagation also increases from precipitation to runoff simulations, particularly for the largest basins of this study. The overall conclusion is that coarse resolution reanalysis products can exhibit strong biases when used for flood simulations. The gauge-adjusted satellite rainfall product evaluated in this study, although uncertain, results in more accurate statistical distributions of the basin response to precipitation.

This study highlights the significant uncertainty associated with coarse resolution re-analysis products for the hydrological simulation purposes. However, the availability of re-analysis products over long time frames (e.g. GLDAS, since 1979) makes these datasets desirable for deriving precipitation and flood frequency analyses for data poor areas. Error corrections and downscaling is needed to make these products useful for flood simulations. Given the improved accuracy, and error propagation properties of current high-resolution satellite precipitation products (such as TRMM3B42V7), combination with re-analysis datasets in data poor areas could lead to improved uses of the these datasets for water resource reanalysis studies.

Since TRMM3B42V7 is a gauge adjusted satellite-rainfall product, it is expected to have a better quality over areas with dense rain gauge networks. The study area of this work represents one of the denser networks of ground based measurements worldwide. To make robust conclusions regarding the accuracies and error propagation of TRMM3B42V7 product, further studies are needed to investigate areas with varying rain gauge data coverage and different geomorphology (e.g. complex terrain) and hydro-climatic conditions. Furthermore, this study was limited to one satellite product, while future work should also consider other global scale and possibly non-gauge-adjusted precipitation products, such as CMORPH and PERSIANN that were discussed in the introduction.

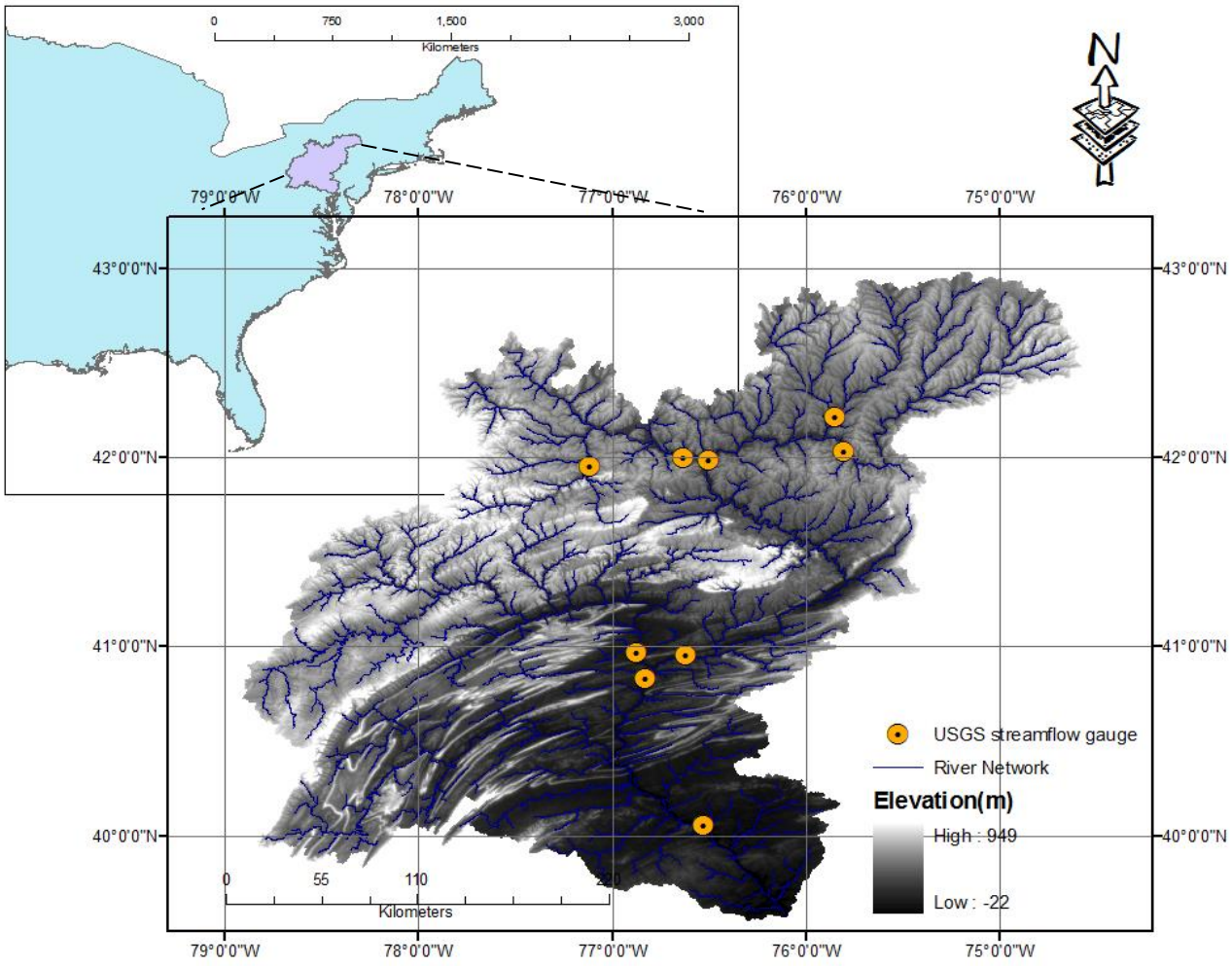


**Figure 3.1.** Map of the Susquehanna River basin. The location of USGS streamflow gauges

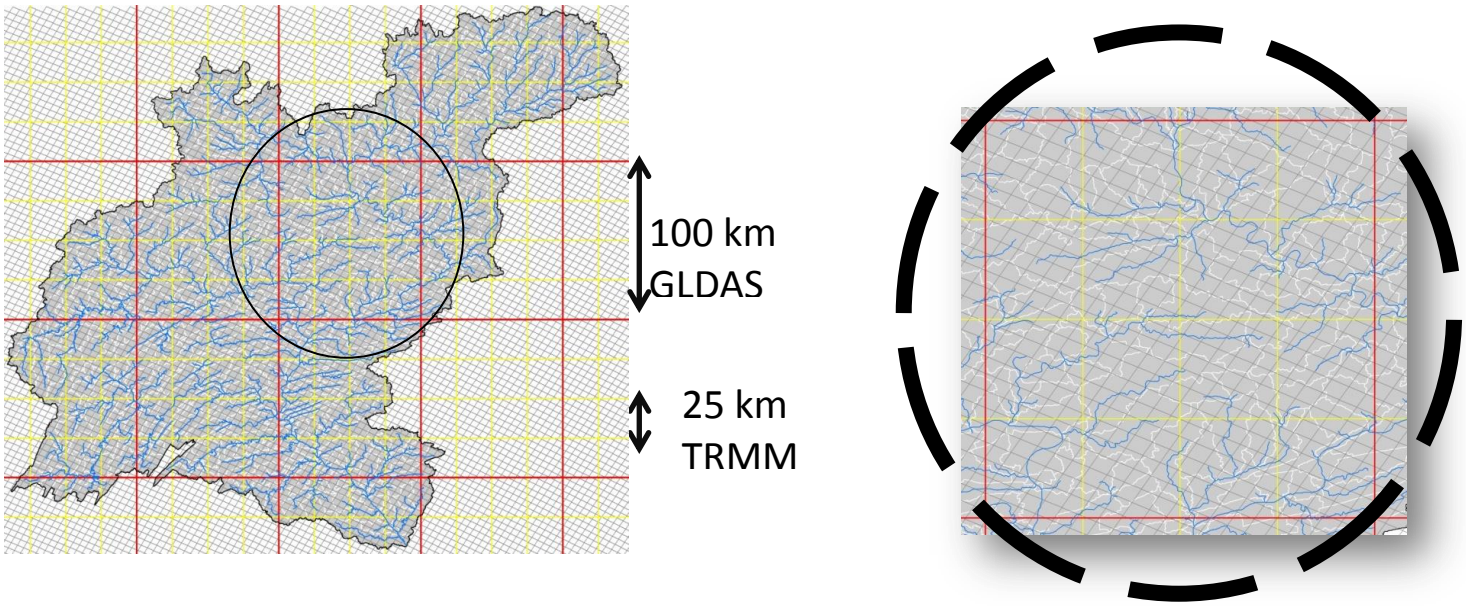


**Figure 3.2.** grid sizes and coverage over Susquehanna River Basin. Red is the GLDAS at 100 km resolution. Yellow is TRMM at 25km resolution. The black lines are stage IV radar data at 4km.

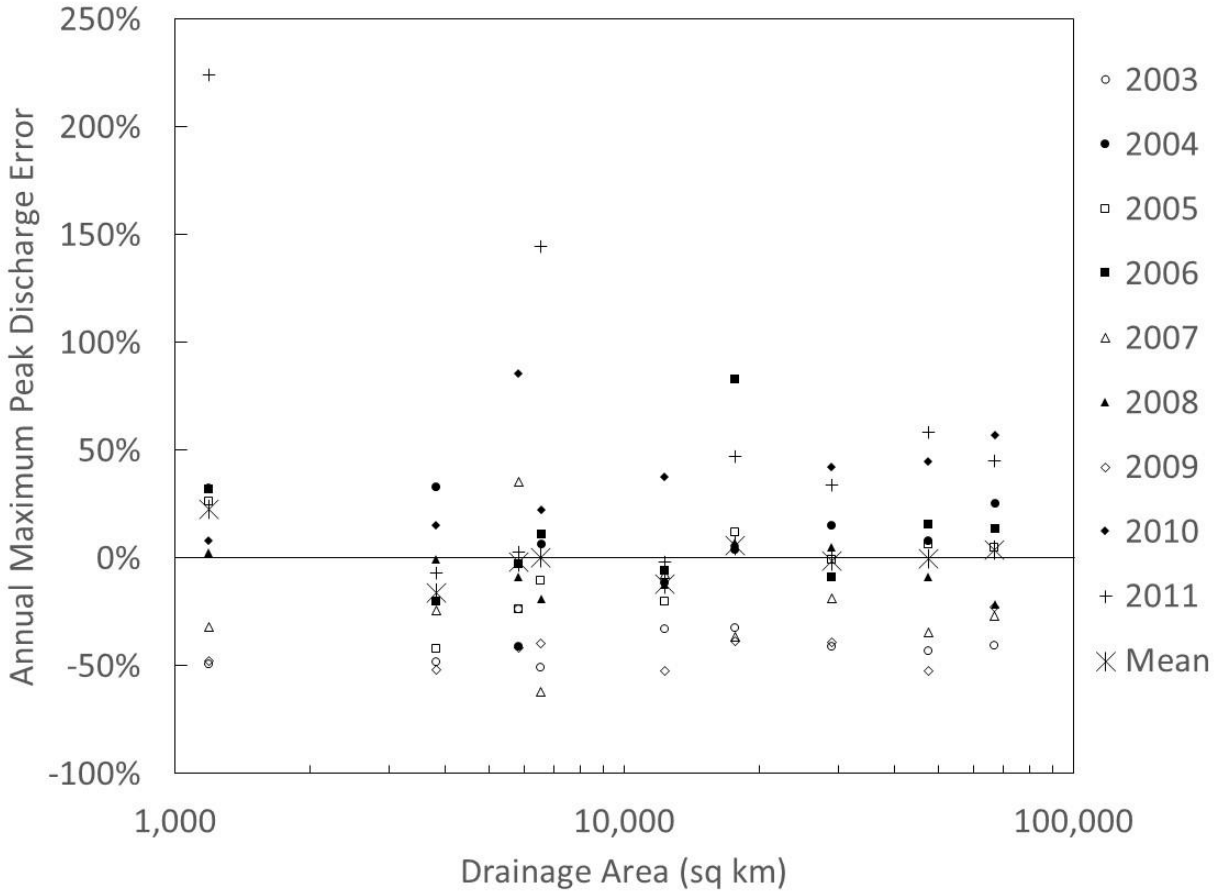


**Figure 3.3.** Simulated annual maximum peak discharges errors at USGS streamflow gauges shown in Figure 1 and listed in Table 2.

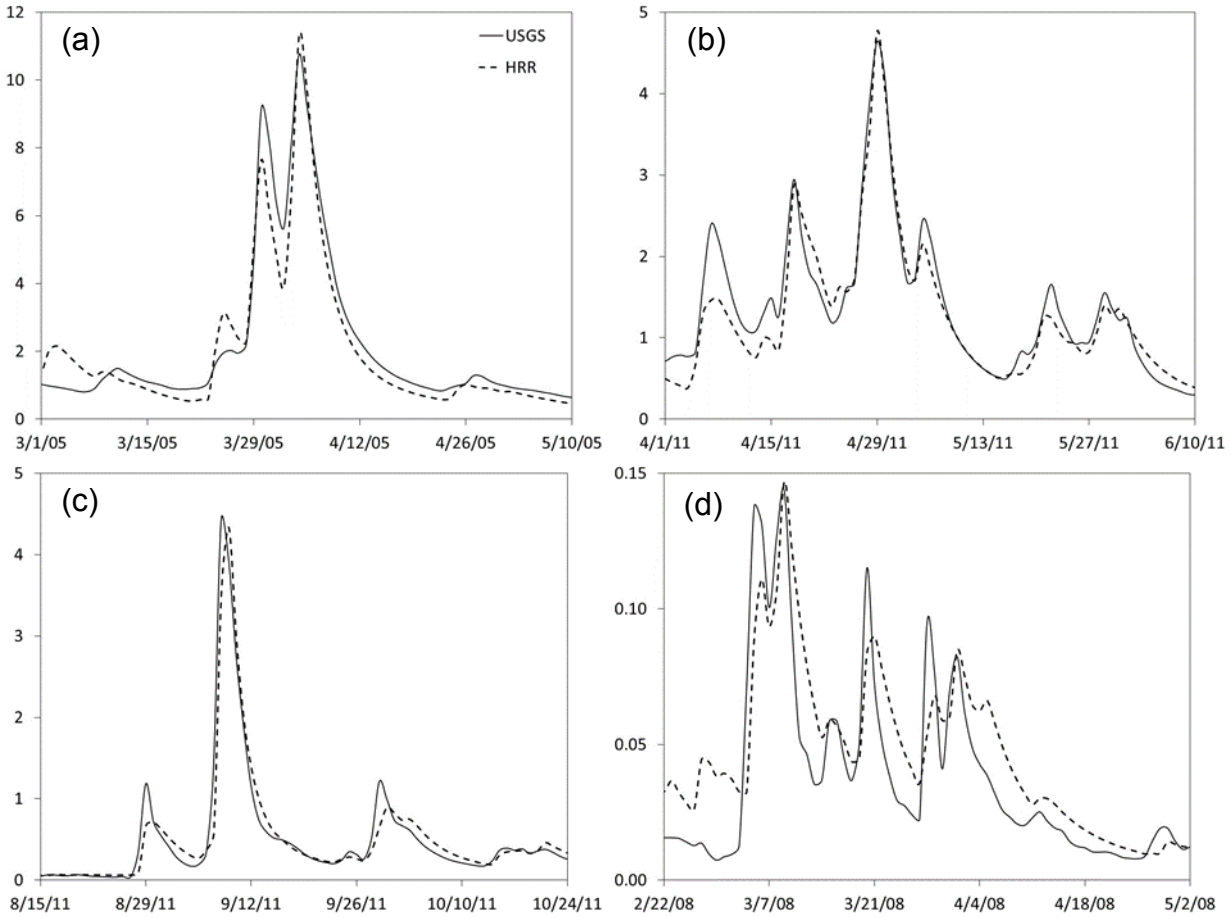


**Figure 3.4.** Comparison between simulated and measured mean daily discharge for four USGS gauges: (a) 01518700, (b) 01515000, (c) 01540500, (d) 01576000; y-axis is Discharge with units of 1,000 $m^3/s$ (see Map ID's 1,5,7,9 listed in Table 2 and shown in Figure 1).

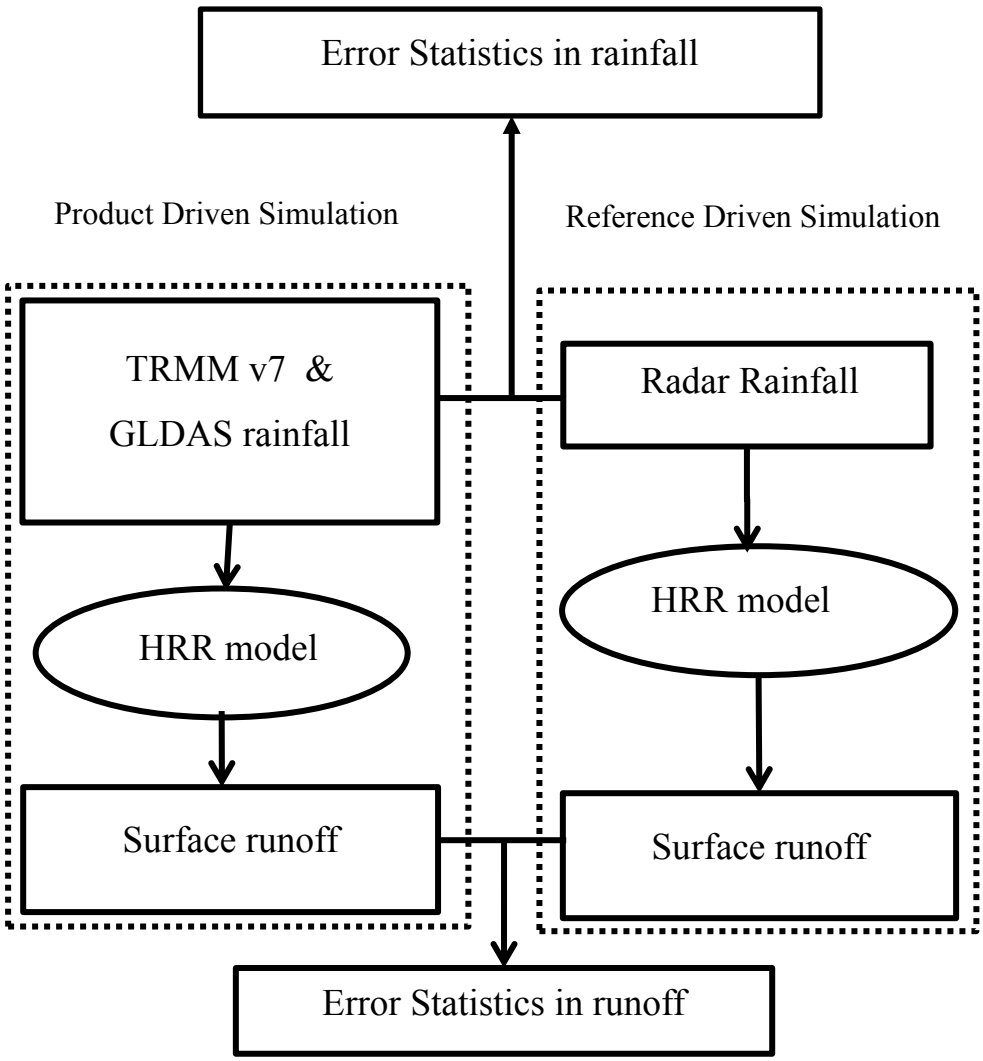


**Figure 3.5.** Experiment setup for assessing the effect of resolution plus retrieval on runoff generation

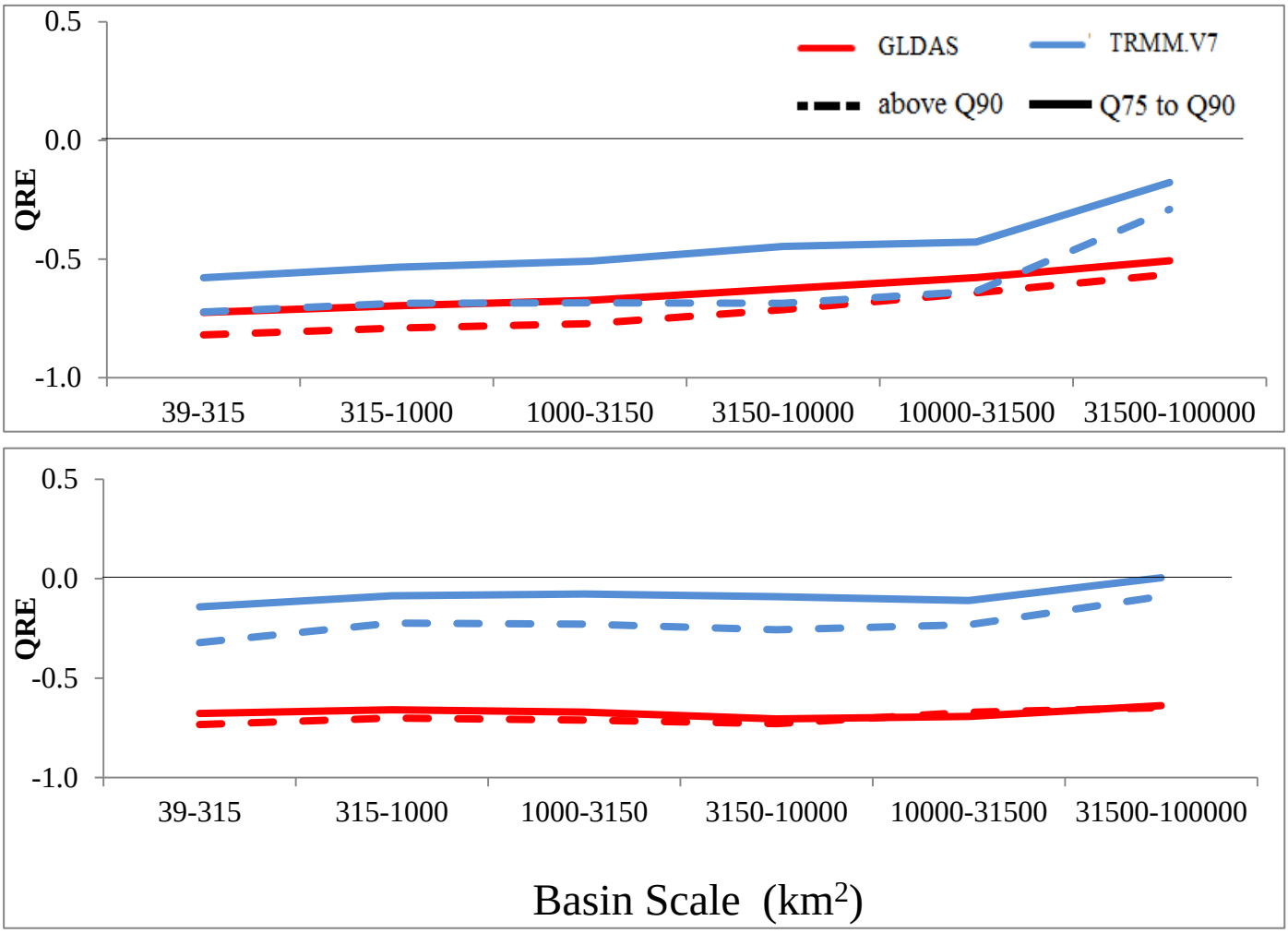


**Figure 3.6.** Mean Scale Quantile Relative Error (QRE) versus basin scale range for GLDAS and TRMM.v7, precipitation (top) and discharge (bottom). Different lines show different quantile ranges. The dashed lined are for values above Q90 and the solid line is for quantiles between 75 and 90.

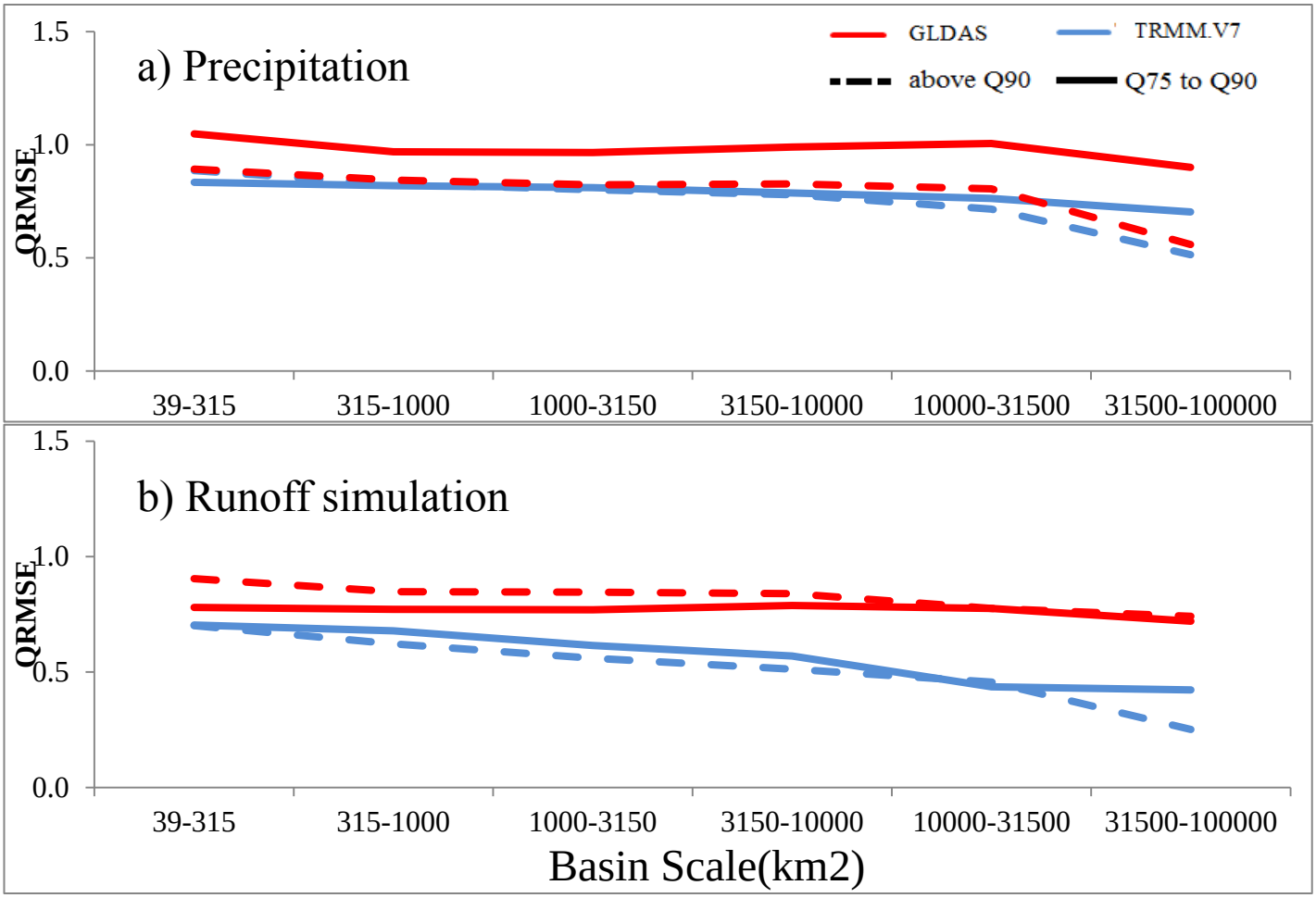


**Figure 3.7.** Quantile Root Mean Square Error versus basin scale range for GLDAS and TRMM.v7, precipitation (top) and discharge (bottom). Different lines show different quantile ranges. The dashed lined are for values above Q90 and the solid line is for quantiles between 75 and 90

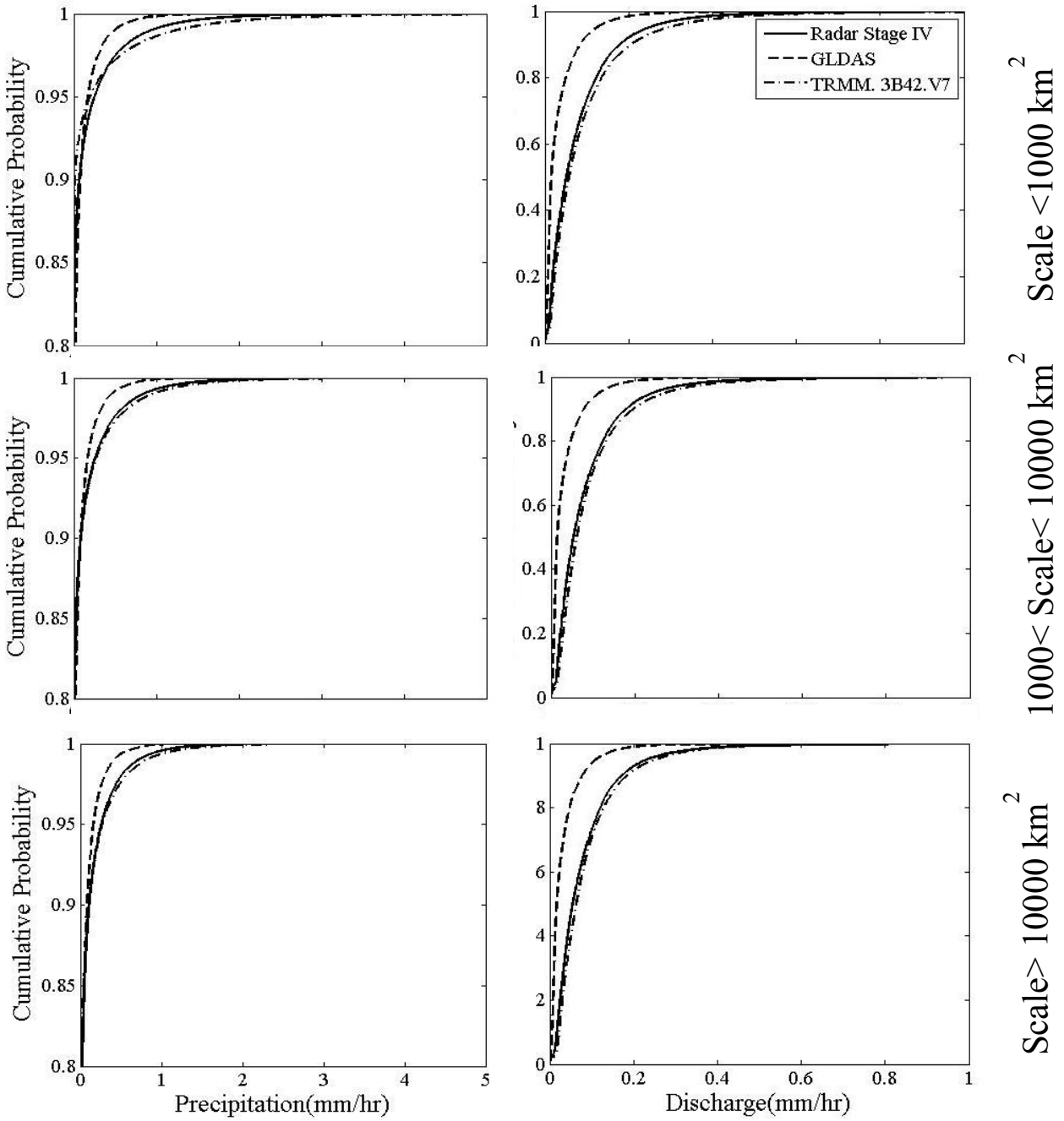


**Figure3. 8.** Cumulative probability for precipitation and run off data. The top row is small size basin (less than 1000km$^2$), the middle row is medium size basin (1000 to 10000 km$^2$) and the bottom row is the large size basin (over 10000 km$^2$).

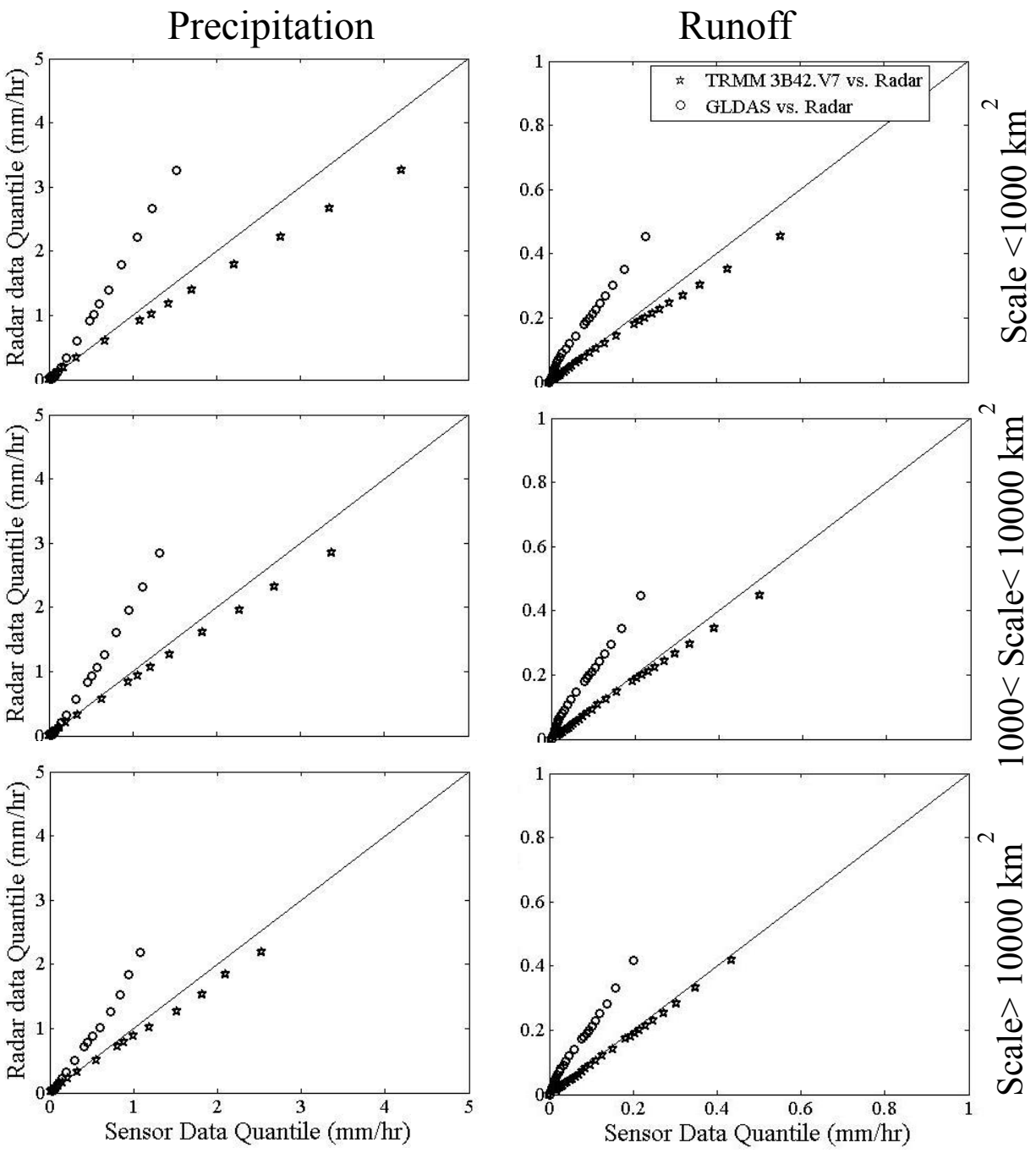


**Figure3. 9.** quantile-quantile plot for precipitation and run off data. The top row is small size basin (less than 1000km$^2$), the middle row is medium size basin (1000 to 10000 km$^2$) and the bottom row is the large size basin (over 10000 km$^2$).

**Table3.1.** Basin scale category

| Scale Category # | Range ($km^2$) | # of basins in scale |
|---|---|---|
| **1** | 39-315 | 633 |
| **2** | 315-1000 | 154 |
| **3** | 1000-3150 | 77 |
| **4** | 3150-10000 | 51 |
| **5** | 10000-31500 | 53 |
| **6** | 31500-100000 | 38 |

**Table 3.2.** Summary model errors statistics for annual maximum peak discharges for the period (2003-2011) at USGS stream flow gauge locations shown in Figure 1

| Map ID | USGS No. | USGS Name | Drainage Area ($km^2$) | Mean Error (%) | Error Range (%) |
|---|---|---|---|---|---|
| **1** | 01518700 | Tioga River at Tioga Junction, PA | 1155 | 23 | -49 to 224 |
| **2** | 01512500 | Chenango River near Chenango Forks, NY | 3841 | -16 | -52 to 33 |
| **3** | 01503000 | Susquehanna River at Coonklin, NY | 5781 | -1.9 | -42 to 86 |
| **4** | 01531000 | Chemung River at Chemung, NY | 6491 | 0.3 | -62 to 145 |
| **5** | 01515000 | Susquehanna River near Waverly, NY | 12362 | -12 | -52 to 38 |
| **6** | 01553500 | West Branch Susq. River at Lewisburg, PA | 17734 | 5.7 | -38 to 83 |
| **7** | 01540500 | Susquehanna River at Danville, PA | 29060 | -1.4 | -41 to 42 |
| **8** | 01554000 | Susquehanna River at Sunbury, PA | 47397 | -0.5 | -52 to 58 |
| **9** | 01576000 | Susquehanna River at Marietta, PA | 67314 | 3.9 | -40 to 57 |

**Table 3.3.** The ratio of QRE in GLDAS to QRE in TRMM3B42V7

| Scale | <1000 km$^2$ | | 1000 to 10000 km$^2$ | | >10000 km$^2$ | |
|---|---|---|---|---|---|---|
| | Precipitation | Runoff | Precipitation | Runoff | Precipitation | Runoff |
| **Q90>t>Q75** | 1.3 | 7.65 | 1.39 | 7.82 | 2.86 | 7.20 |
| **t>Q90** | 1.15 | 3.13 | 1.20 | 2.84 | 1.90 | 7.32 |

**Table 3.4.** The ratio of QRMSE in GLDAS to QRMSE in TRMM3B42V7

| Scale | <1000 km$^2$ | | 1000 to 10000 km$^2$ | | >10000 km$^2$ | |
|---|---|---|---|---|---|---|
| | Precipitation | Runoff | Precipitation | Runoff | Precipitation | Runoff |
| **Q90>t>Q75** | 1.18 | 1.38 | 1.26 | 1.38 | 1.28 | 1.71 |
| **t>Q90** | 1.02 | 1.36 | 1.06 | 1.64 | 1.09 | 2.95 |

**Table 3.5.** The ratio of QRE in runoff to QRE in precipitation for GLDAS and TRMM.v7

| **Scale** | **<1000 km$^2$** | | **1000 to 10000 km$^2$** | | **>10000 km$^2$** | |
|---|---|---|---|---|---|---|
| | TRMM v7 | GLDAS | TRMM v7 | GLDAS | TRMM v7 | GLDAS |
| **Q90>t>Q75** | 0.16 | 1.01 | 0.20 | 1.21 | 0.03 | 1.88 |
| **t>Q90** | 0.33 | 0.93 | 0.37 | 0.95 | 0.31 | 1.40 |

**Table 3.6.** The ratio of QRMSE in runoff to QRMSE in precipitation for GLDAS and TRMM3B42V7

| **Scale** | **<1000 km$^2$** | | **1000 to 10000 km$^2$** | | **>10000 km$^2$** | |
|---|---|---|---|---|---|---|
| | TRMM v7 | GLDAS | TRMM v7 | GLDAS | TRMM v7 | GLDAS |
| **Q90>t>Q75** | 0.73 | 0.80 | 0.72 | 0.80 | 0.60 | 0.80 |
| **t>Q90** | 0.75 | 1.00 | 0.66 | 1.02 | 0.49 | 1.33 |

# 4. Satellite-Driven Downscaling of Global Reanalysis Precipitation Products for Hydrological Applications

## 4.1. Introduction

Flooding is one of the costliest natural hazards, occurring repeatedly around the globe. Flood vulnerability analysis provides essential information to support decisions for policy and preparedness against catastrophic flood consequences and for quantifying risk for coping with this hazard (Sampson et al., 2014). However, flood frequency maps are not available for most regions around the world (Hagen and Lu, 2011) due to limited economic resources to support long-term observations; this results in lack of knowledge and data (e.g. ground based rain gauge measurements). Developing global scale flood maps (Porter and Demeritt, 2012) is of increasing interest in the scientific community with great applicability in the (re)insurance industry. Global gridded precipitation datasets from satellites and reanalysis datasets derived from data assimilation systems are two main sources for deriving global flood hazard maps (Cloke et al., 2013;Kappes et al., 2012).

Global reanalysis products can provide long-term precipitation datasets for frequency analyses of hydrologic extremes (e.g. floods, droughts). Widely used reanalysis products include the JRA-25 (Onogi et al., 2005), ERA-40 (Bosilovich et al., 2008;Uppala et al., 2005), ERA-Interim (Dee et al., 2011), GLDAS (Rodell et al., 2004), and GDAS (Kalnay et al., 1996). These reanalysis products are available at coarse spatial resolution (100 to 250 $km^2$) and suffer from model biases (Gottschalck et al., 2005;Peña-Arancibia et al., 2013). On the other hand, satellite-based rainfall products are associated with relatively higher spatial resolutions (10-25 $km^2$) and improved accuracy, but significantly shorter temporal coverage. The most widely used high

resolution, near-global, multi-sensor precipitation products include the 8 km/half-hourly National Oceanic and Atmospheric Administration CMORPH dataset (Joyce et al., 2004), the 25km/3-hourly U.S. Naval Research Laboratory NRLBLD (Turk and Miller, 2005), the 4km/15-min National Environmental Satellite, Data, and Information Service Hydro-estimator (Scofield and Kuligowski, 2003), the 10km/1-hourly Japan Aerospace Exploration Agency GSMaP (Ushio and Kachi, 2009;Kubota et al., 2007), the 25km/3-hourly Center for Hydrometeorology and Remote Sensing, University of California-Irvine PERSIANN (Sorooshian et al., 2000), and the 25km/3-hourly NASA TRMM3B42 (Huffman et al., 2007). Depending on the algorithm and the purpose of the product, each dataset utilizes various combinations of information from microwave (MW) and geostationary infrared (IR) sensors.

Characterizing the uncertainty in existing global gridded precipitation products is vital for the purpose of hydrological applications. Syed et al. (2004) have shown that rainfall is responsible for nearly 70%-80% of the variability in the land surface hydrology. Therefore precipitation uncertainty would critically affect the predicted variability in hydrologic simulations. Several validation studies have investigated uncertainties related to satellite rainfall remote sensing over diverse geographic and hydro-climatic regimes (Adler et al., 2001;AghaKouchak et al., 2009;Brown, 2006;Dinku et al., 2007;Ebert et al., 2007;Krajewski et al., 2000;McCollum et al., 2002;Seyyedi et al., 2014;Stampoulis et al., 2013;Su et al., 2008;Tang et al., 2010). These studies have shown that the precision of satellite rainfall products depends on precipitation type (e.g. deep convection vs. shallow convection), as well as terrain and climatological factors (AghaKouchak et al., 2011;Demaria et al., 2011;Turk and Miller, 2005; Seyyedi et. al., 2014). Gottschalck et al., (2005) evaluated precipitation products from global models, satellite and radar data against ground based gauge measurements over CONUS for a

period of 14 months. They demonstrated that some of the reanalysis precipitation products (ECMWF, GEOS and GDAS) can generally perform better than satellite precipitation datasets (TRMM3B42RT and PERSIANN). Peña-Arancibia et al., (2013) assessed daily detection and accuracy metrics for reanalysis and satellite precipitation datasets against gauge data. They argued that no product could demonstrate superior performance relative to the other, e.g. the ERA-Interim is better at southern and Northern Australia, JRA-25 performs better in South and East Asia, while TRMM3B42 and CMORPH are better during monsoon periods. Therefore, combined use of different datasets (including satellite and reanalysis) is expected to perform better than any single product, especially for hydrological applications.

Substantial efforts have been devoted to assessing the feasibility of utilizing global-scale precipitation datasets derived from satellite or models on land surface hydrological modeling (Behrangi et al., 2011;Beighley et al., 2011;Hong et al., 2007;Hong et al., 2006;Hossain and Anagnostou, 2004, 2005;Nijssen and Lettenmaier, 2004;Su et al., 2008;Bitew and Gebremichael, 2011). Some of these studies have highlighted the effect of product resolution (Gourley et al., 2011) and catchment size (Vergara et al., 2013) on the precipitation error propagation in hydrological simulations. Seyyedi et al., (2014) have recently utilized gridded precipitation datasets from TRMM3B42V7 (25km, 3hr) and GLDAS reanalysis (100km, 3hr) to conduct a more in depth assessment of the effect of resolution and data type (satellite vs. reanalysis product) on streamflow simulations at sub-daily scale. The study was based on a multi-year (2002-2011) and multi-scale approach considering 1006 sub-basins (36 – 71,000 $km^2$) of the Susquehanna River basin in the northeast United States. They demonstrated that statistical scores in both rainfall and runoff simulations improve with increasing basin size. However, the satellite dataset (TRMM3B42V7) was shown to perform significantly better than the reanalysis

(GLDAS) in the simulated runoff values. The mean relative error in runoff simulations based on GLDAS was up to seven times higher than that of TRMM3B42V7, which was attributed to the product resolution and associated underestimation of heavy precipitation. Results from that study necessitate the use of downscaling and error correction for the GLDAS reanalysis precipitation dataset before implementing it for runoff simulations. Bastola and Misra, (2014) have also evaluated two reanalysis precipitation datasets (ERA-40 and NCEP-R2) for hydrologic simulations over 28 small to midsize basins in southeastern United States. Their results demonstrated that ER-40 tends to underestimate while NCEP-R2 tends to overestimate relative to the reference data. They also concluded that downscaling the reanalysis precipitation products would significantly increase their performance in terms of runoff simulations.

The critical role of high resolution gridded rainfall datasets for hydrological simulations has led to the development of several rainfall disaggregation algorithms (Brussolo et al., 2008;Ferraris et al., 2003;Fowler et al., 2007;Frei et al., 2006;Maraun et al., 2010;Ning et al., 2011;Park, 2013;Rahman et al., 2009;Ramírez et al., 2006;Tao and Barros, 2010, among others). The main assumption for some recently developed downscaling methods for satellite based products is the relationship between spatial variability of rainfall and environmental factors such as topography and land surface conditions. Immerzeel et al., (2009) improved average annual TRMM3B43 from 25km to 1km grid resolution by establishing an exponential relationship between TRMM3B43 and Normalized Difference Vegetation Index (NDVI). (Jia et al., 2011) developed a statistical downscaling scheme based on the relationship between rainfall, terrain elevation and NDVI. They disaggregated TRMM3B43 from 25 km to 1 km grid resolution by forming a multiple linear regression model between parameters. The final products of both aforementioned methods are mean annual rainfall values. Duethmann et al., (2013) downscaled

ERA-40 with a horizontal resolution of 100 km to 12 km grid size using RCM Weather Research and Forecasting Model (WRF) over central Asia for hydrological modeling. The downscaled data exhibited significant modeling improvement. Haas and Born, (2011) introduced a two-step probabilistic downscaling method for disaggregating ERA-Interim using ground based gauge data over a complex terrain in southeastern Morocco. The technique is a combination of CDF transformation based on probability mapping and a multi linear regression model to extrapolate observation data to a high resolution grid using DEM data. The resulted downscaled high resolution precipitation data substantially outperformed the original ERA-Interim data. Although downscaling is a widely studied topic, no studies have implemented satellite precipitation datasets for downscaling reanalysis precipitation datasets. Moreover, most downscaling schemes for reanalysis datasets are in the context of regional climate focusing on producing consistent statistics for downscaled precipitation values and corresponding generated runoff. This work examines the value of satellite data for improving use of reanalysis precipitation datasets for flood simulations and flood frequency analysis. The reason for using satellite datasets is that a great deal of effort has been devoted to improving the accuracy and resolution of satellite retrievals, which is paired with the recent advent of satellite missions on precipitation (Hou et al., 2013). Moreover, satellite products are globally available, which leads to a globally consistent downscaling scheme for reanalysis products that can be particularly useful over areas lacking long-term ground based observations.

The herein study is motivated by the challenges of precipitation applications due to the nonlinear error propagation from rainfall to hydrological simulations and the vital need for high resolution and long-term gridded rainfall data for deriving flood frequency maps. Specifically, we examine the hydrologic impact of using the higher resolution and accuracy TRMM3B42V7

satellite precipitation product to derive finer scale and error-corrected precipitation maps from the GLDAS reanalysis product. The methodology developed for the satellite-driven error correction and downscaling of GLDAS rainfall data is based on a stochastic error model, which was originally developed for modeling the satellite retrieval uncertainty and its error propagation in hydrological applications (Hossain and Anagnostou 2004; Maggioni et al. 2012, 2013). The methodology is independent of ground based measurements, which makes it applicable over data poor areas of the globe. Since GLDAS and other reanalysis datasets are available over a relatively long period of time (35-50 years), developing a proper methodology for downscaling and improving the accuracy of these products could lead to an invaluable global gridded precipitation data source for water resources assessment and flood frequency studies.

This paper is organized into six sections. After a brief introduction, the study area and datasets are described, including the model used for hydrological simulations. The third section introduces the downscaling and error correction scheme, including the experiment setup and parameter calibration. The fourth section presents the error analysis methodology. The fifth section describes the results of the error analysis in rainfall and simulated runoff values. The conclusions section discusses the main findings of this research and provides recommendations for future studies.

## 4.2. Study area, datasets and models

The study area is the Susquehanna River Basin (39N to 43N and 75W to 79W, Figure 1) located in the northeast United States. The highest peak (949m above sea level) is in the northwestern corner and the lowest point (22m below sea level) is in the southeastern corner with a general elevation gradient from north to southeast. The total area of Susquehanna River Basin

is 71,000 $km^2$ of which 76% is in Pennsylvania, 23% in New York, and 1% in Maryland. The Susquehanna River Basin is subject to major devastating floods occurring once every 14 years with an average annual flood damage on the order of $150 million dollars (Susquehanna River Basin Commission, http://www.srbc.net/). Cumulating the drainage areas along the river network at the outlet of each individual catchment provides 373 unique watersheds with drainage areas ranging from 315 to 71,000 $km^2$. The identified sub basins were divided into five basin size categories (see Table 1) to study the effect of basin scale on the precipitation and runoff simulation error.

The study focuses on 437 flood-inducing rainfall events that occurred between 2002 and 2011. To investigate the effect of seasonality, the events were grouped by season. The number of events per season is reported in Table 2. 60% of the events in each season were used for the downscaling model calibration and the remaining 40% were kept for determining error statistics (results presented in this study). Figure 2 shows the cumulative probabilities (CDF) of the events randomly selected for inclusion in the calibration and validation datasets per season. The figure indicates that the probability distributions of calibration and validation rainfall rates are very close to each other, which indicates that the calibration and validation periods have similar statistical properties in terms of rainfall rates. It is noted that the study is based on time-series of catchment average precipitation values from each dataset. Catchment average stands for the weighted average of each dataset's cells that are within a catchment's boundary. The weights are function of the fraction of the catchment covered by each cell.

#### 4.2.1. Stage IV, Radar data

The radar based NCEP stage IV precipitation data (Lopez, 2011) are utilized as high accuracy, reference observation data in this study. The data are originally on a 4×4 km grid and in polar-stereographic projection. Stage IV combines estimates of precipitation from a network of 150 Doppler NEXRAD (Next Generation Weather Radar) with approximately 5500 hourly rain gauge measurements over the CONUS (Continental United States). NEXRAD is technically similar to the Weather Surveillance Radar-1988 Doppler (WSR-88D). The data benefit from the 12 CONUS RFCs (River Forecast Centers) manual quality-control (QC).

### 4.2.2. TRMM3B42V7

TRMM3B42V7 is a combined microwave-infrared precipitation product (Huffman et al., 2007) with 25 km spatial resolution and 3-hour temporal resolution. This study used the most updated version of TRMM3B42V7. The TRMM3B42V7 blending algorithm uses passive microwave (PMW) low earth orbit (LEO) data including the Microwave Imager on TRMM (TMI), Special Sensor Microwave/Imager (SSM/I) on Defense Meteorological Satellite Program (DMSP) satellites, Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR) on Aqua, the Advanced Microwave Sounding Unit-B (AMSU-B) on the National Oceanic and Atmospheric Administration (NOAA) satellite series, Microwave Humidity Sounders (MHS) on later NOAA-series satellites, and the European Operational Meteorological (MetOp) satellite (Kidd et al., 2011). The TRMM3B42V7 combination scheme is based on the Goddard profiling (GPROF) algorithm (Kidd et al., 2011;Kummerow et al., 2001;Kummerow et al., 1996;Olson et al., 1999;Wang et al., 2009;Gopalan et al., 2010) for rainfall estimation from PMW imagers (TMI, SSM/I, and AMSR). The PMW calibrated infrared (IR) precipitation products (Janowiak et al., 2001) from Geosynchronous Earth Orbit (GEO) satellites are used to

fill in the PMW gaps. Specifically, the algorithm takes the value of the PMW-calibrated IR precipitation products when the PMW is not available in a 3-hourly time step. The algorithm uses monthly ground precipitation gauge data extending between 50 N to 50S for bias removal and calibration.

### 4.2.3. GLDAS

The reanalysis precipitation dataset is from GLDAS and has 100 km spatial and 3-hourly temporal resolution. The reasons for selecting GLDAS are its global coverage, relatively high temporal resolution and long data record (since 1979). The data are "observations based", coming from a combination of reanalysis data from the Global Data Assimilation System (GDAS) from the National Center for Environmental Prediction (NCEP), NOAA Climate Prediction Center's CMAP (CPC Merged Analysis of Precipitation) precipitation(Xie and Arkin, 1997), and radiation datasets from Air Force's AGRicultural METeorological modeling system (AGRMET) (Rodell et al., 2004). GDAS assimilates global meteorological observations. CMAP is merged satellite based IR and MW observations with rain gauge analysis. AGRMET radiation fields are satellite observation based. GLDAS therefore represents merged, spatially and temporally interpolated fields of GDAS, CMAP, and AGREMET fields.

### 4.2.4. Hydrologic model simulations

Hillslope River Routing (HRR) (Beighley et al. 2009; 2011) is the modeling framework used in this study. HRR integrates a water balance model for the vertical fluxes and a routing model for the horizontal fluxes of the surface and subsurface runoff. For each model unit, the landscape is

approximated as an open book with two planes draining laterally to a main channel. Water and energy balance is used to simulate the vertical fluxes and storages of water in and through the canopy and soil layers on each plane. Flow routing is then performed using variants of the kinematic wave method for the hillslopes on both the surface and shallow subsurface, and diffusion wave methodologies (i.e., Muskingum Cunge) for channels and floodplains.

Seyyedi et al. (2014) provided details about the model implementation in the Susquehanna River Basin and reported model specifications, parameter calibration and performance results. In addition to the base model parameters (e.g. vertical hydraulic conductivity, suction head, and soil depth) three additional parameters were calibrated in Seyyedi et al. (2014) based on soil and land cover data: horizontal conductivity, Kh, for the subsurface routing, overland roughness, N, for surface routing, and Manning's roughness, n, for channel routing. These parameters are scale dependent in that they capture both the hydraulic features (river reach and hillslope lengths) defined for a given model unit as well as all sub-model unit features not represented at the defined model scale (e.g., all tributaries not explicitly represented in the defined river network). The calibration was performed by systematically adjusting the three parameters (Kh, N, n) to achieve zero mean error (ME, $m^3/s$) for the annual maximum peak discharges at nine streamflow gauging stations shown in Figure 1. As reported in Seyyedi et al. (2014), model performance after calibration includes zero mean error for the entire basin, while mean relative errors for individual gauges ranged between -16% to 23%, and errors for individual events ranged between -62% to 224%. The largest error is from the gauge draining one of the smallest basin area (1,155 $km^2$) during 2011 Tropical Storm Lee, which caused significant flooding especially in the northern Susquehanna River Basin. Overall, 86% of the errors are within ± 50%, and approximately half are within ±25%.

## 4.3. Error correction and downscaling scheme

A stochastic downscaling and error correction method based on the SREM2D satellite rainfall error model is described in this section. The aim of the scheme is to disaggregate GLDAS The stochastic space-time error model of Hossain and Anagnostou (2006), originally developed for satellite rainfall error modeling (hereafter named SREM2D), was devised in this study to disaggregate and error correct GLDAS precipitation datasets using reference data from the TRMM3B42V7 satellite precipitation product. Specifically, SREM2D was applied on the coarse (100 km) grid resolution GLDAS precipitation fields to generate 20-member ensembles of error-adjusted precipitation fields at 25 km grid resolution. Figure 3 illustrates the framework for the stochastic downscaling and error correction. First SREM2D parameters are determined for each season using TRMM3B42V7 and GLDAS data from the calibration datasets of each season. The SREM2D was applied to the GLDAS data during the validation period and evaluated against the reference Stage IV gauge adjusted radar-rainfall fields. Details about the SREM2D model are provided in Hossain and Anagnostou (2006), while below we describe the model calibration results for the different seasons.

SREM2D parameters calibrated in this study are listed as: 1) probability of rain detection (POD) (see Figure 4 panel f); 2) mean of the log-transformed multiplicative error, where error is the multiplicative factor "e=Rsensor/Rreference", this parameter is represented in 2-D spatial fields for each season on Figure 4, panels (a-d) ; 3) mean value of missed rain rate; 4) probability of no-rain detection (POD no rain); 5) correlation length for the retrieval error (CLret); 6) correlation length for the successful delineation of rain (CLrain det); and 7) correlation length for

the successful delineation of no rain (CLno rain det). The calculated values for parameters 3 to 7 are presented in Table 3 for the selected calibration events in each season.

In terms of spatial patterns, the correlation lengths of rain detection, no-rain detection and downscaled rain for all seasons are less than 83 km. The lower correlation length indicates lower dependence between variables in space. Regarding the random error, the range of standard deviation of logarithmic multiplicative errors is between 1.2 (fall) and 1.65 (winter). The values represent higher magnitude of error between reference and sensor data in winter relative to the other seasons. The $POD_{no\text{-}rain}$ takes its maximum value during the summer season (0.98) while it drops to 0.85 for the winter season. The maximum (minimum) mean rain rate of non-detected value is 0.82 for the summer; almost half (0.39) is the corresponding value for the winter. Summer events are associated with higher rain rates, which results in higher non-detected rain rates from GLDAS.

The mean of the log-transformed multiplicative error for each season is presented in 2-D spatial fields (Figure 4,a,b,c,d). The negative mean logarithmic error indicates that the GLDAS is underestimating relative to the TRMM3B42V7. As we see in Figure 4(a,b,c,d) the GLDAS is underestimating almost everywhere and for all seasons. The magnitude of underestimation in the summer is relatively higher than in the other seasons. The probability of rain detection is presented as a function of GLDAS rain rate (Figure 4,e). The summer events exhibit the lowest values whereas fall and spring have higher POD values.

Figure 5 presents the accumulated values based on all validation events for the different precipitation products and the 20-member SREM2D-generate ensembles of GLDAS downscaled precipitation, depicted by the shadow area on the plot. GLDAS rainfall significantly underestimates the other two precipitation datasets, especially in spring, fall and winter seasons,

while the SREM2D-generated ensemble envelops encapsulate well the TRMM3B42V7, and in most cases, the ground based reference accumulated rainfall. This indicates that the disaggregated GLDAS precipitation data are in agreement with the TRMM3B42V7 and the corresponding ground based radar rainfall data.

## 4.4. Error analysis methodology

The error analysis devised in this study, aimed to demonstrate the degree of improvement due to downscaling, consists of three main hydrologic components (Figure 6): reference simulation, observation simulation and downscaled and error corrected simulation. Reference simulation is based on generating runoff values through forcing HRR with the reference radar rainfall data. Observation simulation stands for forcing HRR with GLDAS or TRMM3B42V7 at the product resolution. Downscaled and error corrected simulation refers to forcing HRR with the ensemble mean of the SREM2D-downscaled GLDAS precipitation fields. There are two error analysis steps associated with the three main components: the rainfall error analysis and simulated surface runoff error analysis. Each error analysis component consists of three statistical metrics: quantile-quantile (Q-Q) plots, mean scale quantile relative error (QRE), and the quantile root mean square of error relative to the mean of reference (QRMSE).

The Q-Q plots are used to compare basin-average quantile rainfall and runoff values from the various data sources (GLDAS at 100 km, TRMM3B42 at 25 km, mean GLDAS downscaling ensemble at 25km) against the reference data source (radar at 4 km). The QRE is defined as the ratio of sum of differences between reference and sensor values (precipitation or runoff) to the sum of reference values determined over the sub-basins for each quantile range:

$$QRE = \frac{\sum_{i=1}^{n}((P^{S}_{Sensor}\left| \boldsymbol{t_j \le P^{S}_{ref} < t_{j+1}} \right.) - (P^{S}_{ref}\left| \boldsymbol{t_j \le P^{S}_{ref} < t_{j+1}} \right.))}{\sum_{i=1}^{n}(P^{S}_{ref}\left| \boldsymbol{t_j \le P^{S}_{ref} < t_{j+1}} \right.)} \quad (1)$$

where $P^{S}_{Sensor}$ is the sensor "basin averaged" precipitation/runoff value, $P^{S}_{ref}$ is the reference "basin averaged" precipitation/runoff value over the sub-basin, *t* is the threshold value which is based on the reference data quantiles, *j* is the quantile index, and *n* is total number of value in a particular scale and quantile range. The perfect value for this metric is zero, which means there is no difference between reference and the sensor values. Negative QRE value means the sensor is underestimating and the positive value means overestimating.

QRMSE is the root mean square of the differences between reference and sensor; it is normalized to the mean of reference values.

$$QRMSE = \frac{\sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}\left((P^{S}_{Sensor}\left| t_j \le P^{S}_{ref} < t_{j+1} \right.) - (P^{S}_{ref}\left| t_j \le P^{S}_{ref} < t_{j+1} \right.)\right)^2}}{mean\,(P^{S}_{ref}\left| t_j \le P^{S}_{ref} < t_{j+1} \right.)} \quad (2)$$

where $P^{S}_{Sensor}$ is the sensor "basin averaged" precipitation/runoff value, $P^{S}_{ref}$ is the reference "basin averaged" precipitation/runoff value over the sub-basin, *t* is the threshold value which is based on the reference data quantiles, *j* is the quantile index, and *n* is total number of value in a particular scale and quantile range. QRMSE quantifies the spread between sensor and reference data points.

To determine dependence of the error metrics on storm severity, QRE and QRMSE statistics are categorized in two groups according to the quantile values of rainfall and runoff, namely, values between 75$^{th}$ and 90$^{th}$ percentile and greater than the 90$^{th}$ percentile that represent moderate and extreme events, respectively. To investigate the effect of seasonality and basin scale statistics, Q-Q plots are presented for the four seasons and different basin scales.

## 4.5. Results

### 4.5.1. Rainfall Error Analysis

As mentioned above, the rainfall error analysis is divided into two categories: frequency distribution and quantitative statistics. The frequency distribution uses the quantile-quantile (Q-Q) plots, and the quantitative statistics include the QRE, and QRMSE error metrics. These are discussed next.

#### 4.5.1.1. Frequency Distribution

To assess the correspondence between sensor and reference rainfall data we plotted the quantile values from TRMM3B42V7, GLDAS and downscaled ensemble-mean GLDAS (sensor) against the corresponding quantile values of the reference radar rainfall (Figure 7). In each Q-Q plot the x-axis represents sensor values and the y-axis represents radar values in mm/hr. We show significant changes in the Q-Q plots for the different basin scales (small to large) and seasons.

GLDAS show a systematic underestimation at all seasons and basin scales. The underestimation is most severe at the smallest basin scales (top panels). During the summer convective rainfall season, the underestimation reduces significantly for medium to large basin scales, and it turns to slight overestimation for the small quantile values (< 1mm/hr). On the other hand, the GLDAS downscaled ensemble-mean data exhibit much better agreement with the reference radar rainfall data. The best agreement is observed during the fall and summer seasons, while good agreement is also depicted during the spring season. The winter season exhibits a strong underestimation (overestimation) of the low (high) quantile values. Overall, the downscaled GLDAS precipitation dataset exhibits similar performance to the TRMM3B42V7 product in the fall, summer and

spring seasons, while in the winter the downscaled GLDAS shows stronger underestimation than TRMM3B42V7 for the low quantile values.

#### 4.5.1.2. Quantitative Statistics

The seasonal variation of the mean relative error and relative root mean square error statistics versus basin scale for GLDAS, TRMM3B42V7, and the downscaled ensemble-mean GLDAS are presented in Figures 8 and 9, respectively. These statistics are based on precipitation values that exceed the 90th percentile. The main point to note is that no datasets show significant changes with basin scale. In spring the GLDAS is significantly underestimating while TRMM3B42V7 is almost unbiased, while the downscaled GLDAS is slightly overestimating for all basin size categories. In summer all datasets are underestimating. The magnitude of underestimation in GLDAS is significantly higher than that of TRMM3B42V7 or the downscaled ensemble-mean GLDAS. In fall GLDAS is significantly underestimating while the downscaled ensemble-mean GLDAS is almost unbiased, in contrast to the TRMM3B42V7, which is slightly overestimating. In winter GLDAS is underestimating while TRMM3B42V7 and downscaled ensemble-mean GLDAS are overestimating. The magnitude of overestimation in the downscaled ensemble-mean GLDAS is lower than the underestimation in GLDAS. For the random component of precipitation error (relative RMSE), the three precipitation datasets are performing similarly, with scores very close in the summer and fall seasons (scores ranging between 0.9 and 1.05). Overall, GLDAS exhibits lower relative RMSE values than the other two precipitation datasets, with this different becoming more significant (range between 0.8 and 1.4) during winter and spring seasons.

### 4.5.2. Simulated Runoff Error Analysis

Time series of the simulated runoff for the entire basin derived from forcing the HRR model with GLDAS (observation simulation), TRMM3B42V7 (product simulation), downscaled ensemble-mean GLDAS (downscaled and error corrected simulation), and radar-rainfall data (reference simulation) for the validation data sample of each season are presented in Figure 10. As shown in the time series plot, GLDAS is systematically underestimating runoff relative to the other datasets, and particularly during the major hurricane events in the fall. The downscaled ensemble-mean GLDAS performs significantly better and is shown to be able to capture the events and the overall flow patterns. In the case of the high flow fall events (associated with two hurricanes) the downscaled ensemble-mean GLDAS simulated runoff seems to be between TRMM3B42V7 and reference data. Below we discuss quantile-quantile (Q-Q) plots and QRE, and QRMSE error metrics for the runoff simulations.

#### 4.5.2.1. Frequency Distribution

The quantile-quantile (Q-Q) plots of the simulated runoff values from the three datasets (i.e. TRMM3B42V7, GLDAS, and downscaled ensemble-mean GLDAS) against the reference simulations are presented in Figure 11. Similar to Figure 7, GLDAS exhibits a strong underestimation of runoff at all seasons and basin scales. The underestimation is shown to the more significant in the fall, spring and winter seasons, while it reduces significantly during the summer events. The ensemble-mean downscaled GLDAS on the other hand exhibits very good agreement with the reference values, particularly during fall and spring seasons. This agreement is very similar to the one exhibited for the TRMM3B42V7 dataset, indicating that downscaling

makes GLDAS perform similarly as the corresponding TRMM3B42V7 dataset, which was used in the calibration of the stochastic model parameters.

#### 4.5.2.2. Quantitative Statistics

Figures 12 and 13 show the two error metrics (QRE and QRMSE) determined for the validation sample reference runoff simulation values exceeding the $90^{th}$ percentile value for the different seasons. As shown in the QRE plots of Figure 12, GLDAS underestimates significantly in all seasons (Figure 12). The magnitude of underestimation is the strongest in summer and fall seasons and the lowest in spring season. Winter season underestimation reduces with increasing basin scale. The ensemble-mean downscaled GLDAS QRE values exhibit significant bias reduction in runoff simulations, particularly in the fall and winter seasons. In spring the downscaled GLDAS exhibits overestimation, which is still lower in absolute magnitude than the underestimation of the original GLDAS runoff simulations? The QRE values of the TRMM3B42V7 product are consistently low, showing a positive bias of <10%.

For the random error component, downscaling consistently improves the QRMSE statistic at all basin scales and for all seasons. The greatest reduction on QRMSE is in the summer and winter seasons, while spring exhibits the least effect. The satellite product (TRMM3B42V7) shows consistently lower QRMSE values than both GLDAS and downscaled GLDAS products for all basin scales and seasons. The greatest difference is in the summer and fall seasons that are associated with more organized convective system and less snow/mixed phase precipitation. Spring season also exhibits a slight basin scale dependence on QRMSE for the downscaled GLDAS and TRMM3B42V7 product driven runoff simulations; no significant basin scale dependence is presented for the other seasons or products.

The above findings are in contrast with the increased random error component shown in the downscaled GLDAS precipitation product (Figure 9). To understand this aspect, we present in Table 4 the QRMSE ratios between runoff and precipitation (error propagation) for the two products, seasons and basin scales. The downscaled GLDAS exhibits dampening of the random error component from precipitation to runoff simulations; this dampening seems to be less dependent on basin scale and more related to season. For example, winter and spring seasons exhibit the strongest dampening of random error (ratios around 0.5), while in the summer the ratio is around one (i.e., no change), and in the fall the ratio is around 0.8 with a slight basin scale dependence (i.e., ranging from 0.86 for basins below 1000 $km^2$ to 0.79 for basins greater than 10,000 $km^2$). On the other hand the original GLDAS product shows either an increase in the random error component from precipitation to runoff simulations during summer and fall seasons, or a weaker (about half) dampening, compared to the downscaled product, in winter and spring seasons. These differences in precipitation to runoff error propagation convert the slightly increased random error of the downscaled GLDAS product in precipitation to a significantly lower random error in runoff simulations, which is consistent with our aim of improving the hydrologic use of GLDAS products in flood modeling.

## 4.6. Conclusions

The aim of the study was to evaluate a stochastic downscaling and error correction approach for improving the use of a global reanalysis precipitation dataset (GLDAS) in flood simulations. GLDAS is available over a long time frame (since 1979), which provides a good source of precipitation data for hydrological analysis and flood hazard mapping. However, it has been shown that the resolution and biases of this product (Seyyedi et al., 2014) introduce significant

runoff simulation errors, which limit its applicability for flood modeling. In this study we proposed utilizing a two dimensional stochastic error model (SREM2D) to downscale and adjust GLDAS precipitation data using as reference a higher resolution satellite precipitation product (the TRMM3B42V7). The study focused on a large basin (Susquehanna River Basin) in the northeast United States and 437 events selected over a 10 year period (2002-2011) and grouped in four seasons. The hydrologic simulations were performed with HRR, which was locally calibrated using radar-rainfall and streamflow observation from nine sub-basins in the region.

The improvements from downscaling and adjusting the GLDAS precipitation were evaluated in terms of both rainfall and runoff simulations using frequency distributions and quantitative error metrics. The effect of basin scale and seasonality were considered in this analysis. For the precipitation error analysis, the quantile-quantile plots indicated that GLDAS is significantly underestimating for all seasons and all basin scales, while the satellite-driven downscaled GLDAS ensembles reduced significantly that bias, reaching a performance similar to the TRMM3B42V7 precipitation product. This was confirmed by the mean relative error statistic, where downscaled GLDAS shows significant reduction of the strong underestimation exhibited in the original GLDAS product. The error analysis in simulated runoff values gave similar bias patterns as those in the precipitation products. The downscaled ensemble-mean GLDAS product has significantly reduced bias compared to the original GLDAS product. There is a slight basin scale effect on the evaluated statistics, with slightly better runoff simulations for large basin sizes. The random error in the simulated runoff values reduces significantly for the downscaled ensemble-mean GLDAS product relative to the original GLDAS. This was explained by the properties of the random error propagation from precipitation to runoff simulations, where for the original GLDAS the random error is either increasing (summer and fall seasons) or slightly

decreasing (winter and spring). On the other hand the downscaled GLDAS product showed a significant dampening (0.5-0.8) of the random error from precipitation to runoff simulations. This can be attributed to hydrologic processes (infiltration and runoff generation) that can average out the random precipitation error component of the high-resolution products (e.g. downscaled GLDAS), but make discharge errors worse for the strongly underestimated GLDAS rainfall rates within the basin.

Overall, results presented in this study indicate that the proposed satellite precipitation based downscaling and error correction method has the potential to improve the hydrological use of GLDAS precipitation reanalysis datasets. The main advantage of this approach is that it uses high-resolution global precipitation products from multi-sensor satellite observations, which makes it flexible to implement over areas with limited ground based measurements. Furthermore, the downscaling scheme is modular in design-it can be applied for any gridded dataset in any region in the world.

The proposed scheme was demonstrated over northeast United States, which is a data rich area. As stated in the study area section, TRMM3B42V7 technique uses regional ground based precipitation measurements from rain gauges to adjust the precipitation retrieval. Although this approach is consistently applied globally, many areas of the world do not have the gauge density of the US rain gauge network. As argued in studies reported in this paper, rain gauge adjustments in data poor areas may worsen the accuracy of TRMM3B42V7 product. Therefore, future research should evaluate this scheme on the basis of other satellite products that do not use rain gauge based adjustments to represent more accurately the conditions of data poor areas. Another extension of this research is to apply the SREM2D downscaling scheme on the entire (35 year) record of GLDAS precipitation data to derive multi-year downscaled GLDAS reanalysis

ensembles, through which we can derived flood return periods for the basins examined in this study. Finally, extending the downscaling methodology to GLDAS and other reanalysis products, such as ERA-40 and ERA-interim, at global scale in conjunction with multi-year (1998-2014) high-resolution precipitation products from satellite-only techniques (CMORPH, PERSIANN) would allow derivation of a global satellite-driven water resources (re)analysis. Such products could be used in many engineering and scientific applications, including flood and drought frequency analysis, design of structures, reservoir design and operation, etc.

***Acknowledgments:*** This study was partially supported by a NASA Precipitation Measurement Mission grant (PI Anagnostou). The first author (Hojjat Seyyedi) was also supported by a FM Global fellowship.

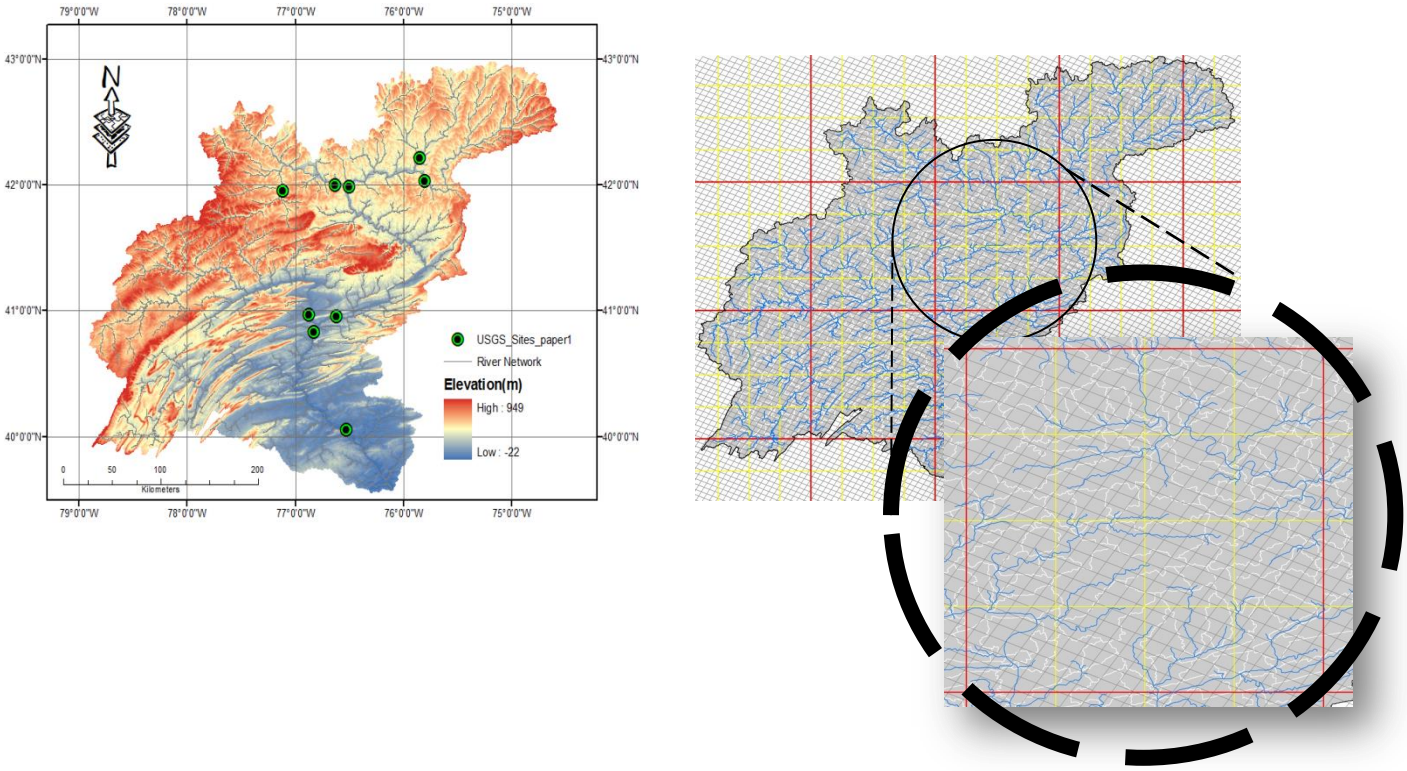


**Figure 4.1:** Study area (left panel) and precipitation product grids (right panel) over the Susquehanna River Basin. Red gird is indicating GLDAS (100km), yellow is indicating TRMM3B42V7 (25km) and black is indicating the stage IV radar rainfall product (4km).

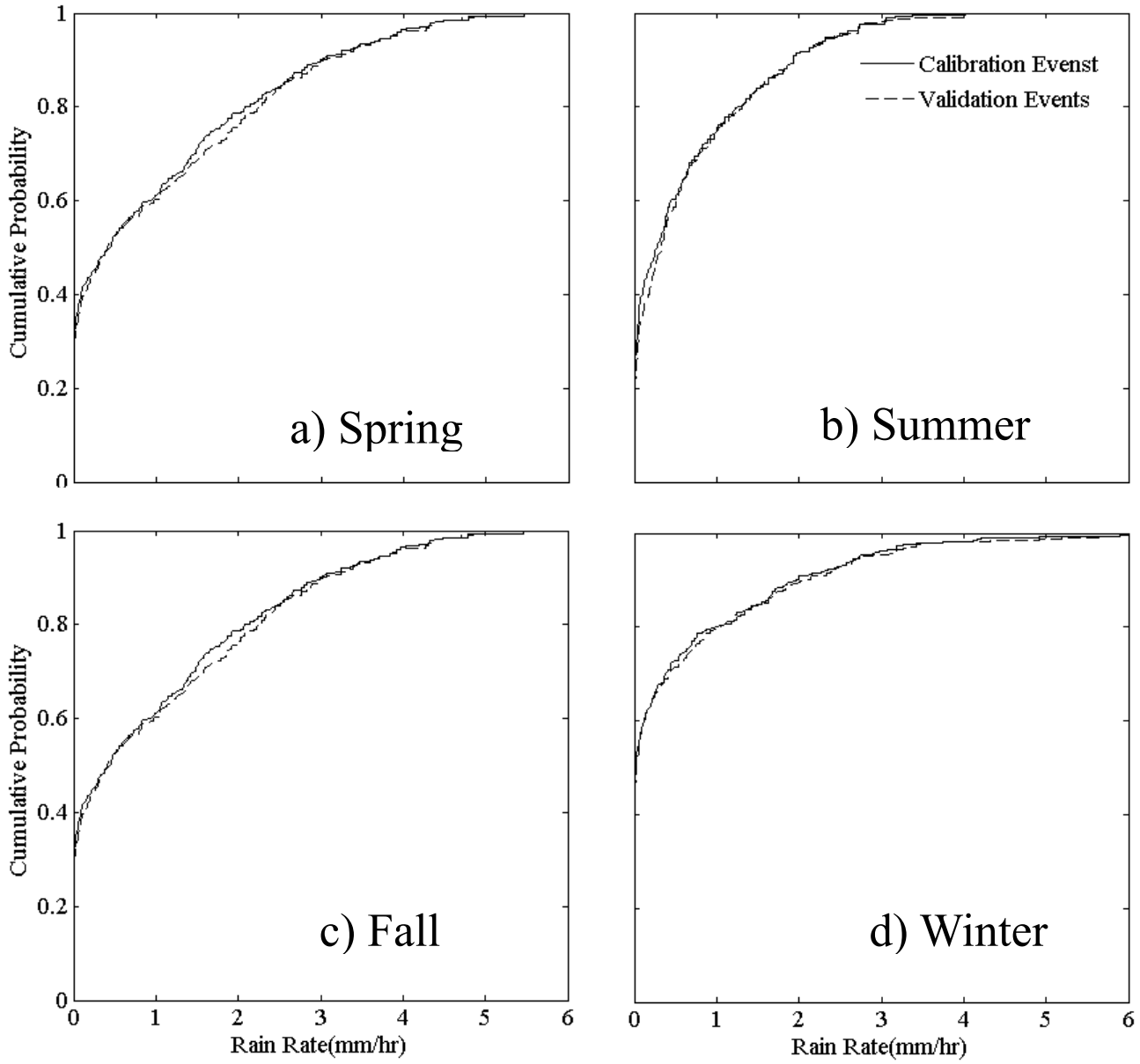


**Figure 4.2:** Cumulative probability of radar rainfall rain rates in the calibration and validation events selected for each season.

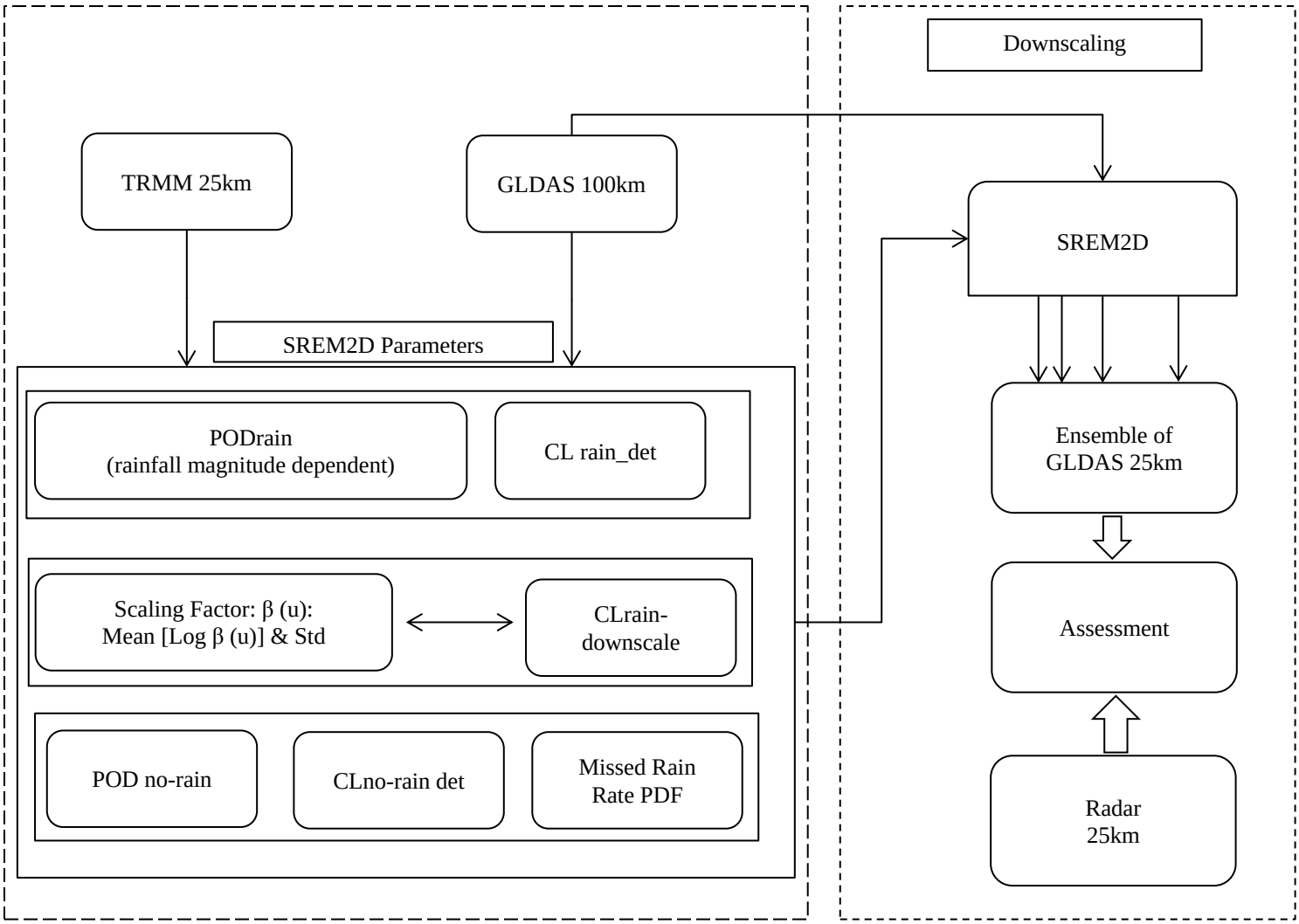


**Figure 4.3:** Stochastic downscaling framework. It consists of two main parts; the left side is indicating required SREM2D parameters and the right side shows GLDAS ensemble generating and quality assessment with the absolute reference data (Stage IV radar data).

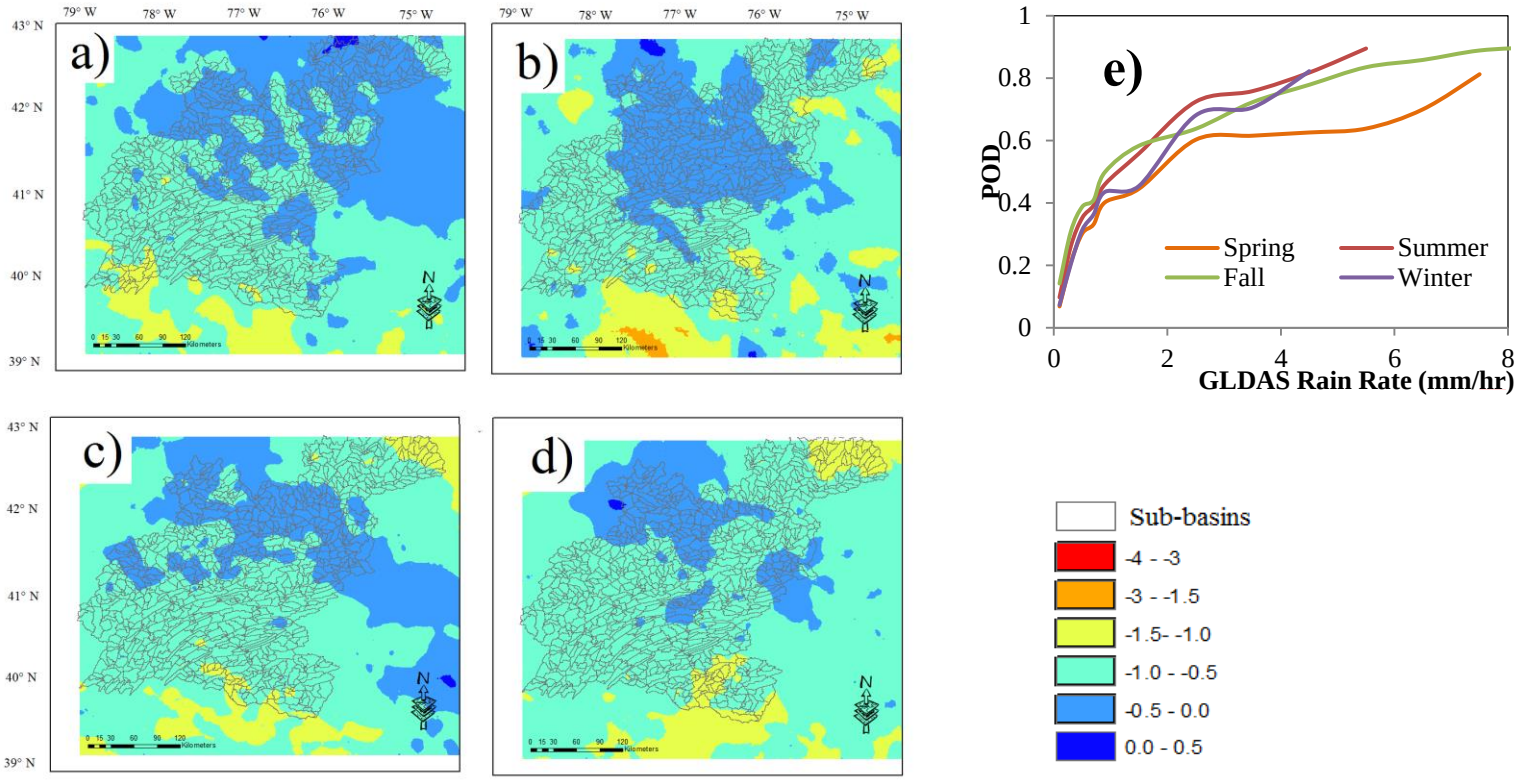


**Figure 4.4:** SREM2D parameters, 2-D spatial mean of logarithmic error 'e' for each season, a) Spring, b) Summer, c) Fall, d) Winter; Probability of rain detection as function of GLDAS rain rate (e).

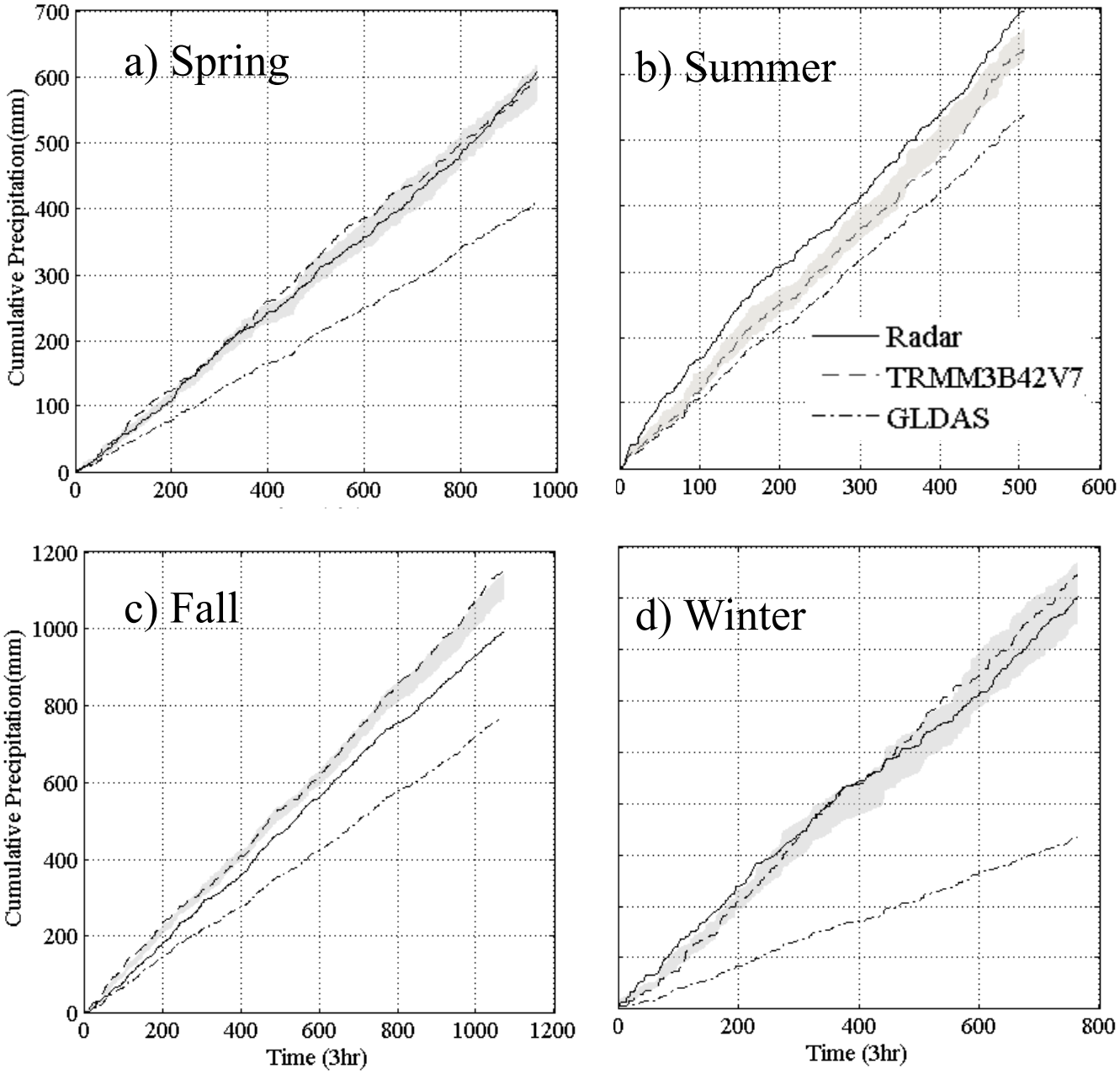


**Figure 4.5:** Cumulative precipitation values of the validation events in the three seasons; the shaded area indicates the 20 ensemble members of downscaled and error corrected GLDAS data.

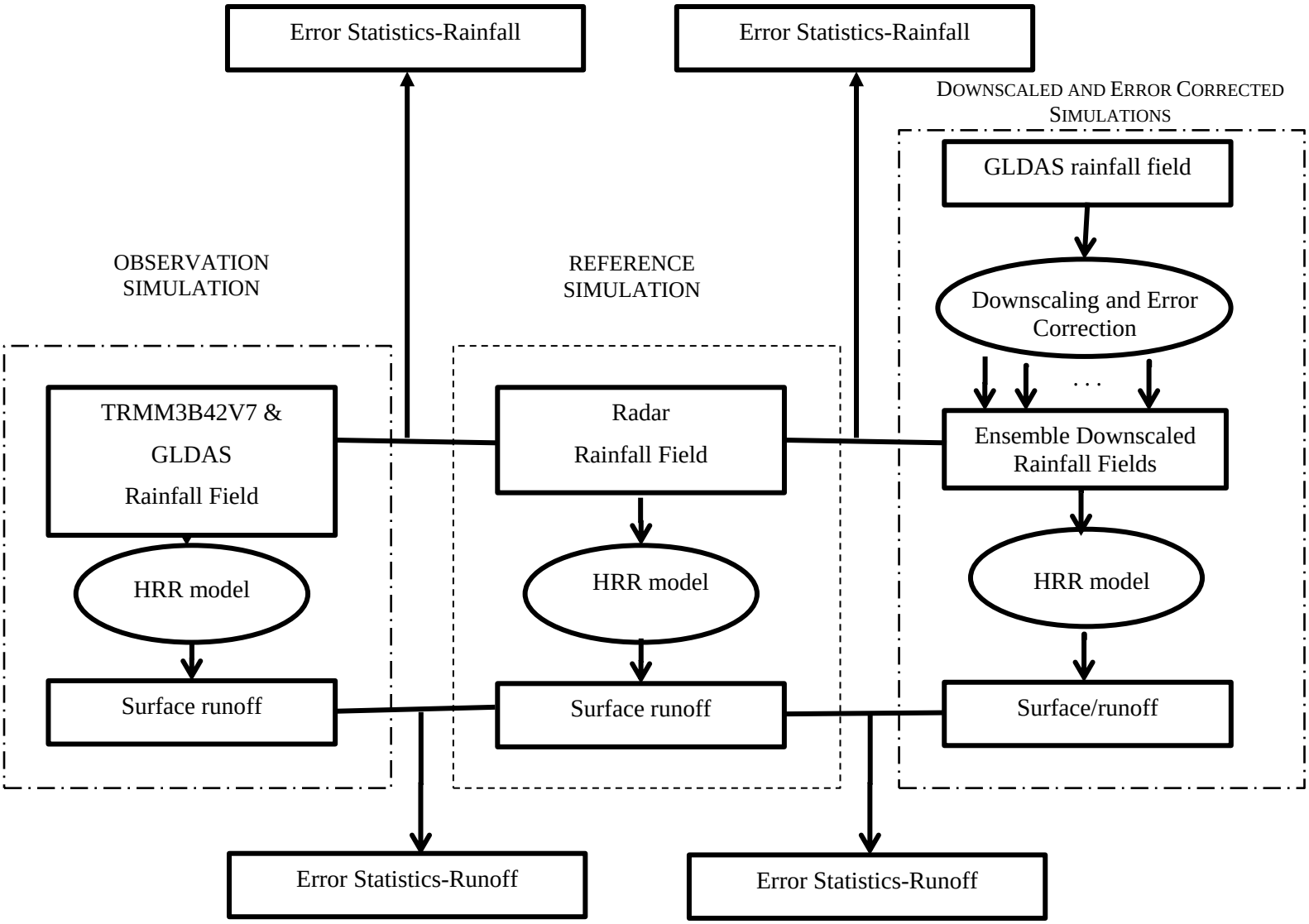


**Figure 4.6:** Flow diagram for the error analysis methodology

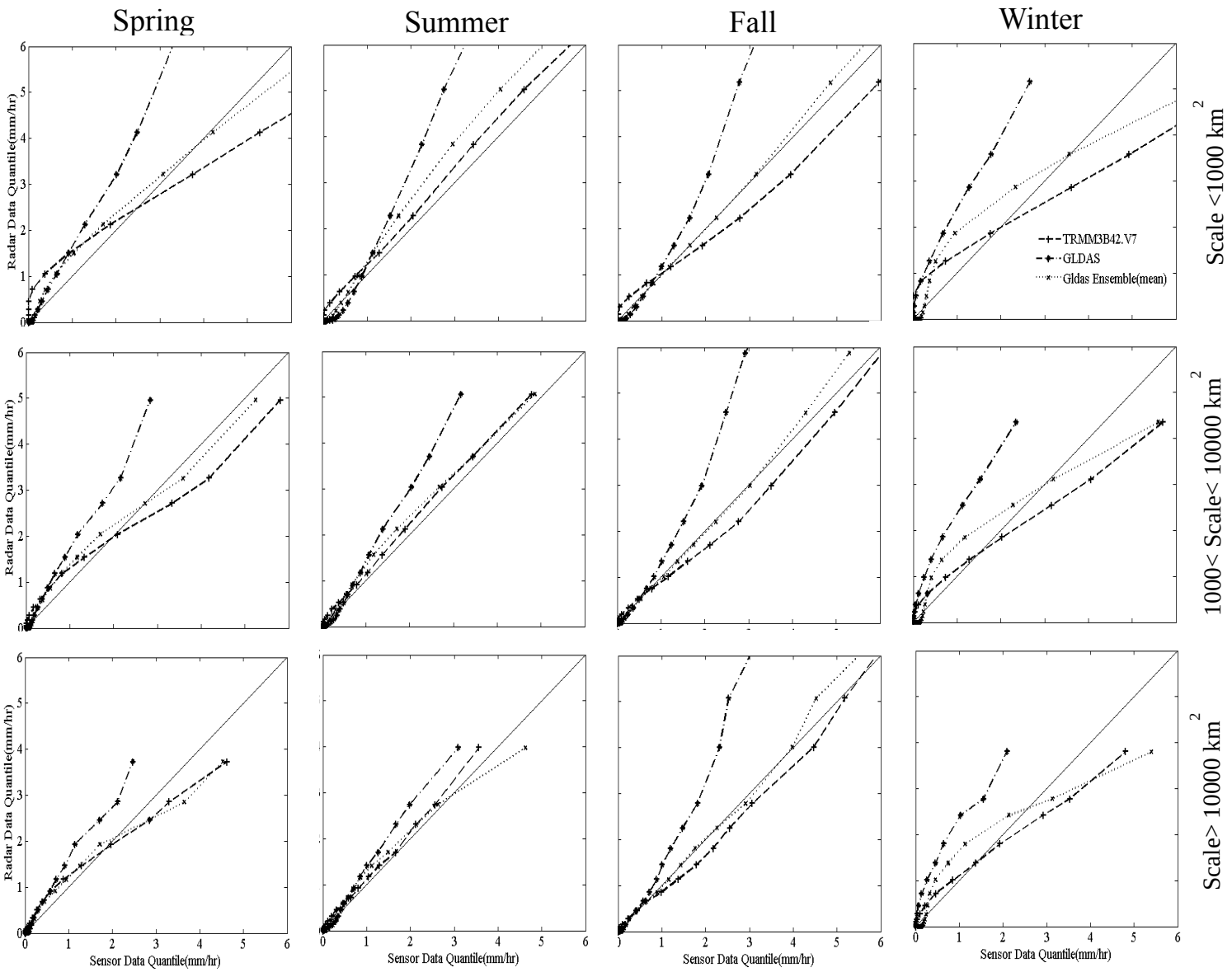


**Figure 4.7:** Quantile-Quantile plots between sensor (TRMM3B42V7, GLDAS and downscaled GLDAS) and radar precipitation values. Rows are indicating different basin scale categories. Columns are representing the different seasons.

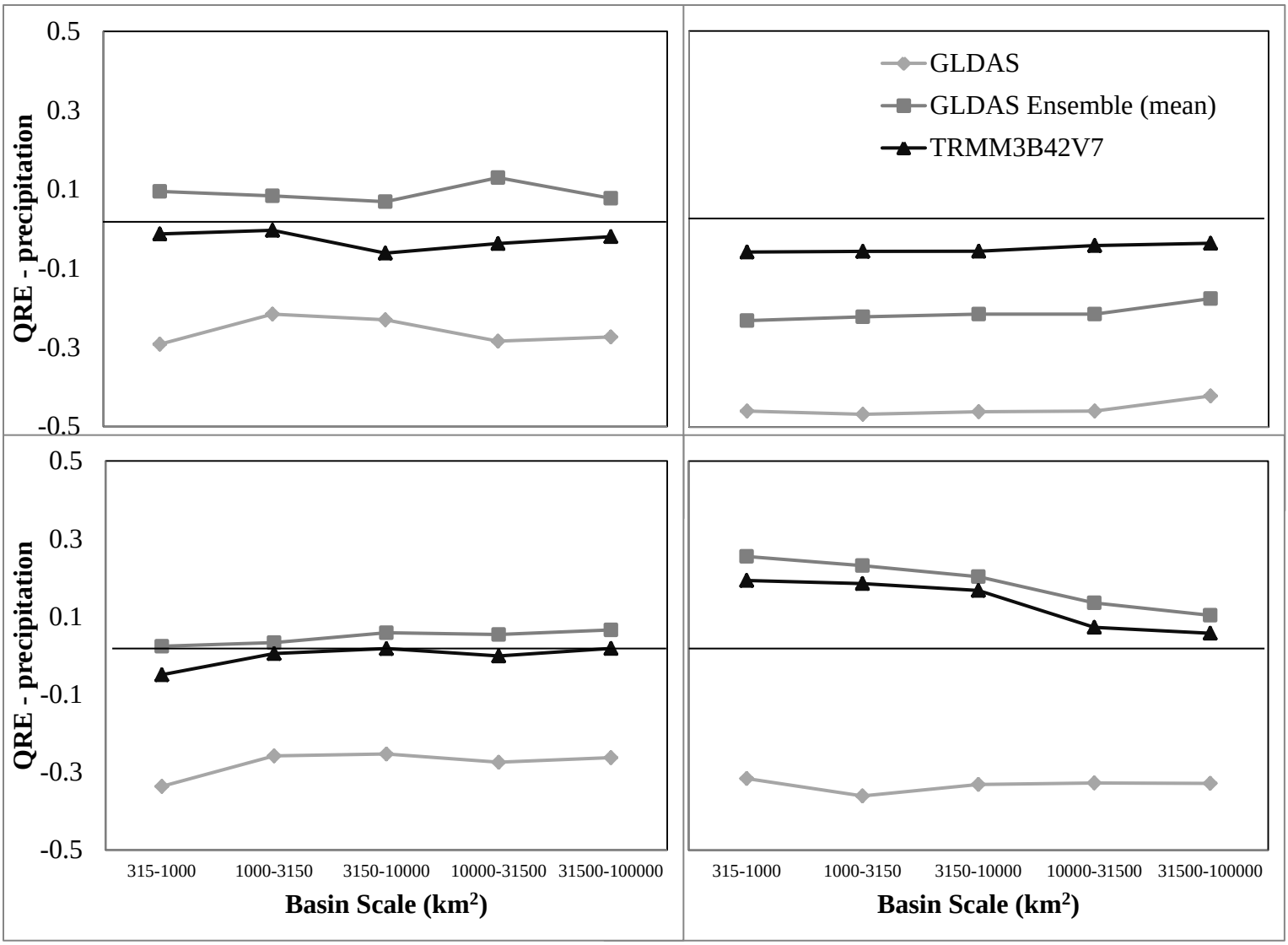


**Figure 4.8**: QRE error metric determined conditional to reference precipitation values exceeding their 90th percentile. The horizontal axis indicates basin scale categories presented in Table 2. Results are presented for spring (upper left panel), summer (upper right), fall (lower left) and winter (lower right) seasons.

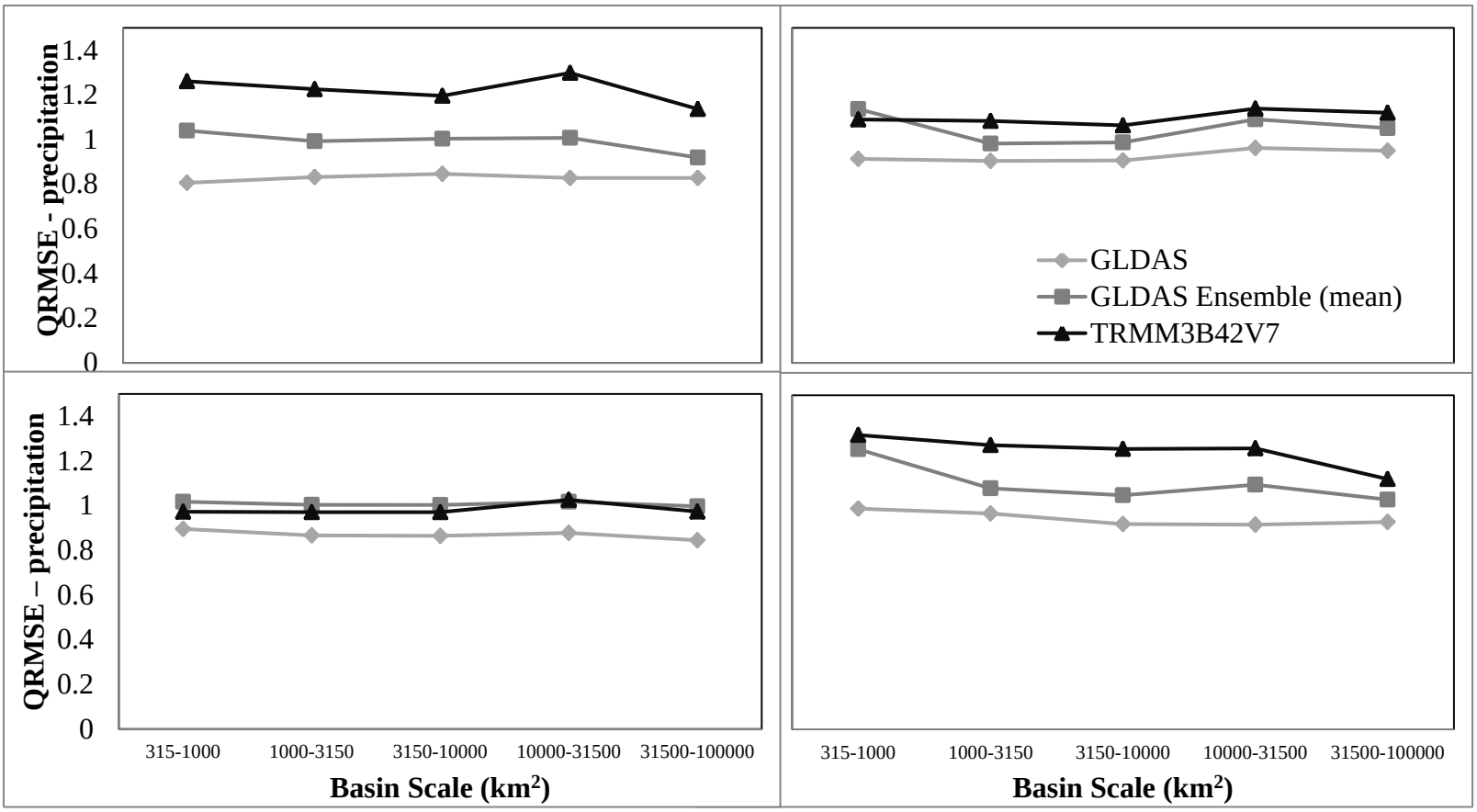


**Figure 4.9**: QRMSE error metric determined conditional to reference precipitation values exceeding their 90th percentile. The horizontal axis indicates basin scale categories presented in Table 2. Results are presented for spring (upper left panel), summer (upper right), fall (lower left) and winter (lower right) seasons.

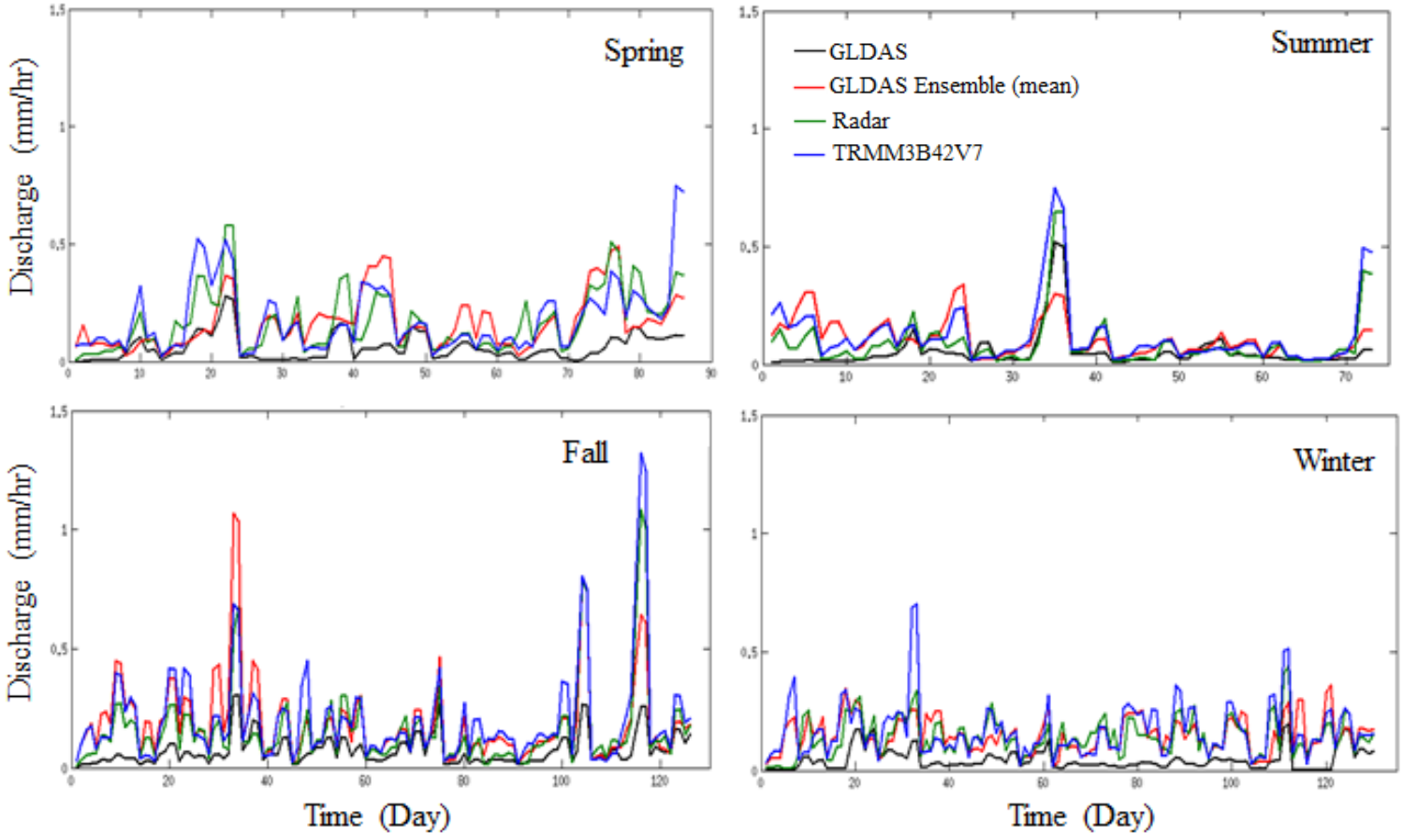


**Figure 4.10:** Runoff time series driven by the different precipitation product over the basin indicated in Figure 1 and consisting of the selected validation events of each season.

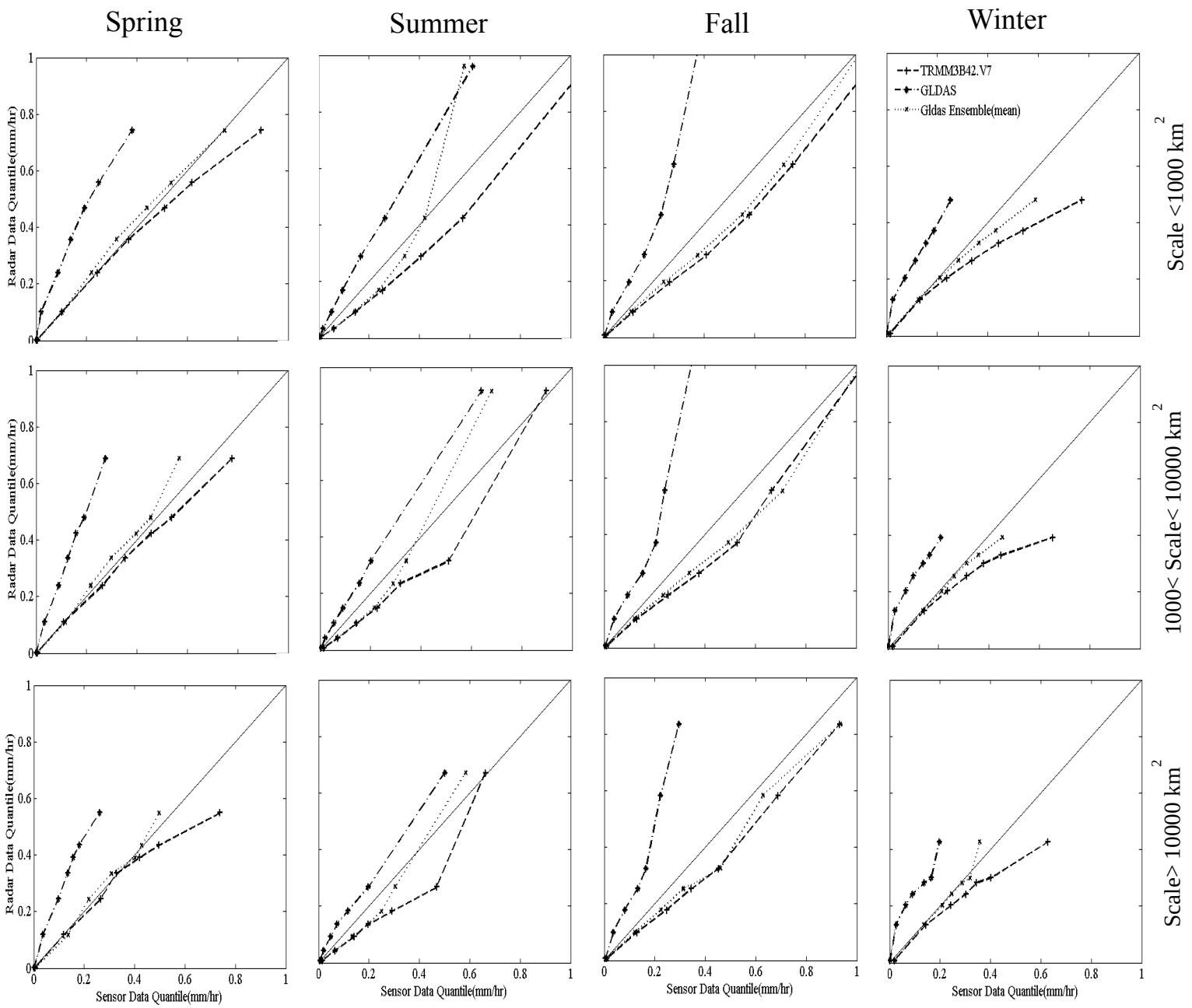


**Figure 4.11:** Quantile-Quantile plots between sensor (TRMM3B42V7, GLDAS and downscaled GLDAS) and radar rainfall-driven runoff simulations. Rows are indicating different basin scale categories. Columns are representing the different seasons.

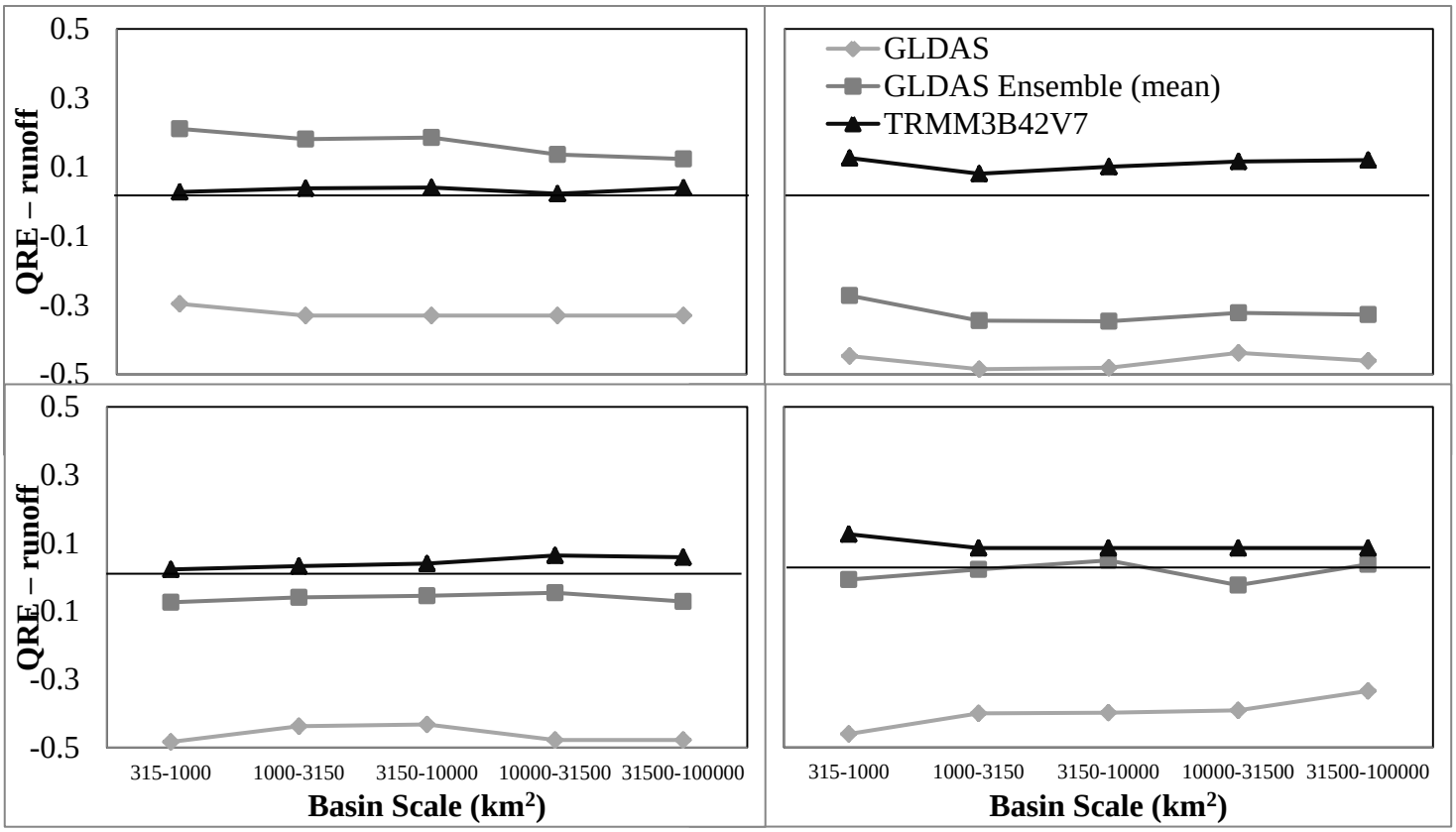


**Figure 4.12:** QRE error metric determined conditional to reference runoff values exceeding their 90$^{th}$ percentile. The horizontal axis indicates basin scale categories presented in Table 2. Results are presented for spring (upper left panel), summer (upper right), fall (lower left) and winter (lower right) seasons.

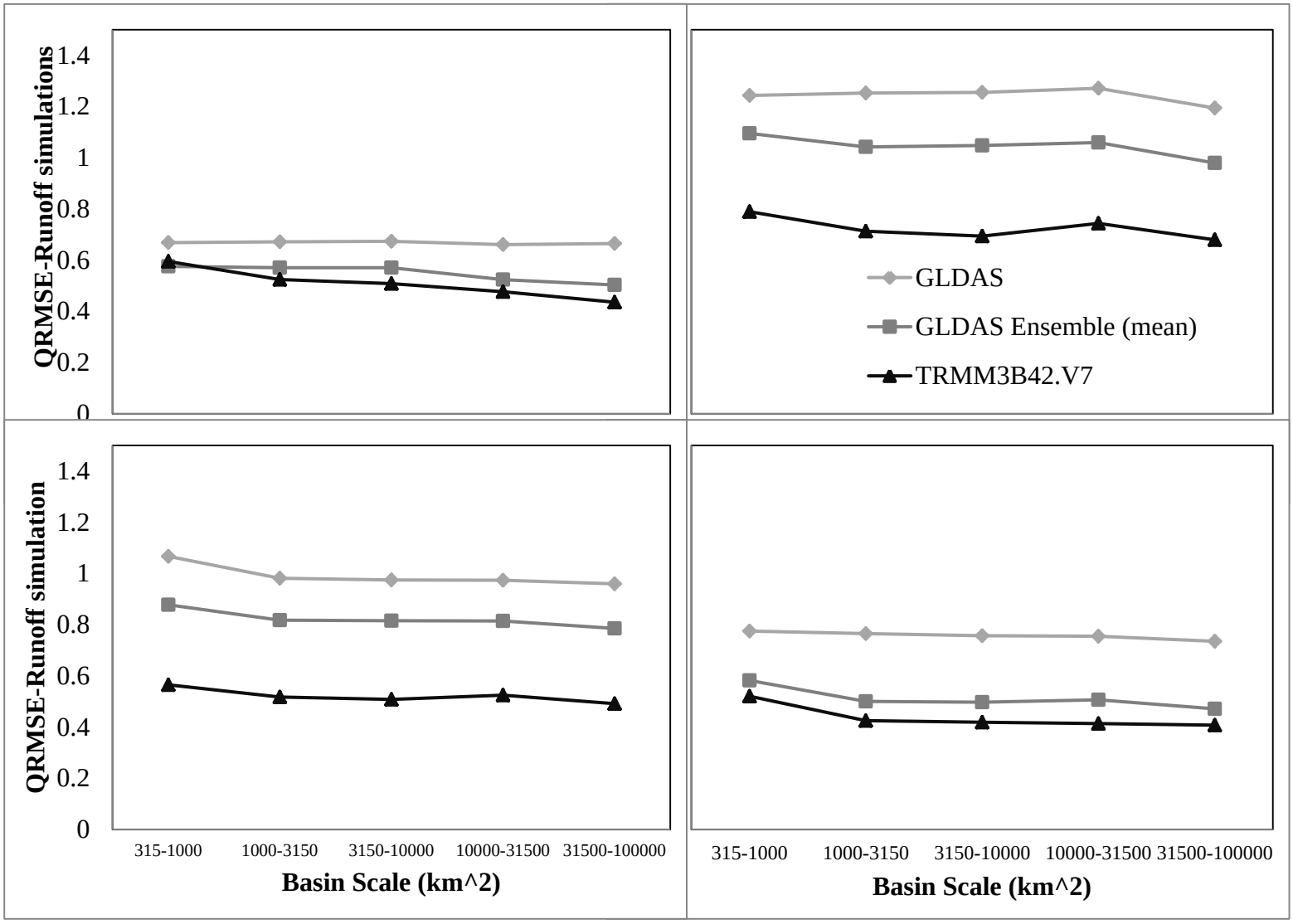


**Figure 4.13**: QRMSE error metric determined conditional to reference runoff values exceeding their 90th percentile. The horizontal axis indicates basin scale categories presented in Table 2. Results are presented for spring (upper left panel), summer (upper right), fall (lower left) and winter (lower right) seasons.

**Table 4.1.**: Number of basins for each basin scale category.

| Basin Scales ($km^2$) | # of basins |
|---|---|
| **315-1000** | 154 |
| **1000-3150** | 77 |
| **3150-10000** | 51 |
| **10000-31500** | 53 |
| **31500-100000** | 38 |

**Table 4.2**: Number of flood events selected for each season between January to December 2002-1011.

| Events | Number of Events | Percentage out of the total number of events(%) |
|---|---|---|
| Spring | 94 | 21 |
| Summer | 74 | 17 |
| Fall | 157 | 36 |
| Winter | 112 | 26 |
| Total | 437 | 100 |

**Table 4.3:** SREM2D parameters determined for GLDAS downscaling for the four seasons using the calibration events.

| SREM2D Parameter | Spring | Summer | Fall | Winter |
|---|---|---|---|---|
| Mean (mu-Gaussian of log-error) | Figure 4, a,b,c,d | Figure 4, a,b,c,d | Figure 4, a,b,c,d | Figure 4, a,b,c,d |
| PODrain | Figure 4, e | Figure 4, e | Figure 4, e | Figure 4, e |
| Sigma (std.dev Gaussian of log-error) | 1.48 | 1.35 | 1.20 | 1.65 |
| Missed mean rain rate (mm/hr) | 0.58 | 0.82 | 0.48 | 0.39 |
| POD no-rain | 0.88 | 0.98 | 0.93 | 0.85 |
| CLrain-downscale (Km) | 22.00 | 20.00 | 18.00 | 21.00 |
| CLrain det (Km) | 43.00 | 40.00 | 48.00 | 40.00 |
| CLno rain det (Km) | 49.00 | 15.00 | 67.00 | 83.00 |

**Table 4.4:** The ratio of QRMSE in runoff to QRMSE in precipitation for GLDAS and ensemble-mean downscaled GLDAS data.

| **Scale** | **<1000 km$^2$** | | **1000 to 10000 km$^2$** | | **>10000 km$^2$** | |
|---|---|---|---|---|---|---|
| | GLDAS | downscaled GLDAS | GLDAS | downscaled GLDAS | GLDAS | downscaled GLDAS |
| **Spring** | 0.83 | 0.55 | 0.80 | 0.57 | 0.80 | 0.53 |
| **Summer** | 1.36 | 0.96 | 1.39 | 1.06 | 1.29 | 0.95 |
| **Fall** | 1.19 | 0.86 | 1.13 | 0.81 | 1.12 | 0.79 |
| **Winter** | 0.78 | 0.46 | 0.81 | 0.47 | 0.81 | 0.46 |

## 5. Conclusion remarks

The major conclusions for this study are summarized as below:

- Significantly better performances in terms of all analyzed statistics are reported for the ensemble realizations generated using the conditional approach for estimating the SREM2D parameters.
- The interquartile range for boxplots of RSTD and NSE is larger for ensembles generated using the unconditional approach. This can be attributed to the statistical dispersion of ensemble realizations generated disregarding soil wetness conditions.
- The random ensemble error quantification exhibits lower EP and UR for the ensemble simulations conditioned to the soil wetness. This results in better encapsulation of the surface rainfall (represented by reference rainfall data) and more accurate estimation of the retrieval error variability on the basis of the SREM2D-generated ensembles.
- The reduction in the error characteristics and improvement in ensemble verification indicated the strong significance of considering surface soil wetness in conditioning the error modeling of passive microwave retrievals. The outcome is valuable for linking two major EO missions: Global Precipitation Measurement and Soil Moisture Active Passive.
- TRMM3B42V7 exhibits significantly better error statistics than the GLDAS re-analysis product in terms of basin average rainfall and simulated runoff values with respect to the reference radar-rainfall dataset.
- The differences in the error statistics are more significant in the basin runoff simulations than basin average precipitation. Specifically, the ratios of quantile relative error (QRE) in simulated runoff through GLDAS to the simulated runoff through TRMM3B42V7

increased to about 7 and 3 times for moderate (75th to 90th quantile) and high flow values (above 90th quantile), respectively.

- The comparison of cumulative probabilities of rainfall and normalized-runoff is showing significant underestimation of the GLDAS rainfall and runoff values (relative to the radar and satellite cumulative distributions) at all basin scales.
- The error propagation analysis showed that the GLDAS product exhibits significant increase of the mean relative error going from precipitation to runoff, while the opposite is true for TRMM3B42V7.
- The GLDAS random error component propagation increases from precipitation to runoff simulations, particularly for the largest basins of this study.
- Coarse resolution reanalysis products can exhibit strong biases when used for flood simulations. The gauge-adjusted satellite rainfall product evaluated in this study, although uncertain, results in more accurate statistical distributions of the basin response to precipitation.
- There is a critical need to develop a downscaling scheme to improve the accuracy of global precipitation products from re-analysis sources for global flood modeling projects.
- It is confirmed by frequency distribution and quantitative statistics that mean of downscaled GLDAS ensemble is outperforms significantly better than the original GLDAS in terms of the quality of precipitation data and generated runoff values.
- The proposed downscaling scheme is independent from the ground based measurements, because the method technically conditioned to the other global product with higher accuracy (i.e. multi-sensor-satellite precipitation product). This makes the downscaling

scheme quite flexible to implement even in areas with very limited ground based measurements.

- Satellite rainfall-driven downscaling of the GLDAS **reanalysis** rainfall datasets improves both the systematic and random component of the runoff simulation error, thus enhancing the use of this product for flood modeling applications (e.g. flood frequency analysis)
- The downscaling scheme is modular in design- it can be applied for any gridded dataset in any region in the world.

The framework developed in first objective is applicable for the other microwave imagers onboard low-earth orbiting satellites and the result can be useful for forthcoming missions on precipitation (Global Precipitation Measurement) and soil moisture (Soil Moisture Active Passive). We know the study relies upon the quality and specification of 2A12, Q2, and NASA Catchment land surface model simulations. Future studies should investigate ways to directly incorporate soil moisture as a parameter of the error model, as we demonstrated how some error statistics are highly dependent on it. As a first attempt to investigate this question we focused on conditions ensuring good reference estimates. Performing the comparison during the warm season over Oklahoma offers the best measurement conditions related to Vertical Profile of Reflectivity (i.e., radar beam below the melting layer) and beam blockage effects. Applying the same approach over various areas to assess the influence of other potential error factors (e.g. shallow convection over complex terrain) or during other seasons (with different soil moisture and vegetation conditions) will be the subject of future studies. To generalize findings of this study it requires extending this error modeling exercise over different regions, longer periods of time, and including additional surface characteristics (e.g. vegetation cover). The framework is

applicable for the other microwave imagers onboard low-earth orbiting satellites and the result can be useful for the GPM mission.

The second study highlights the significant uncertainty associated with coarse resolution re-analysis products for the hydrological simulation purposes. However, the availability of re-analysis products over long time frames (e.g. GLDAS, since 1979) makes these datasets desirable for deriving precipitation and flood frequency analyses for data poor areas. Since TRMM3B42V7 is a gauge adjusted satellite-rainfall product, it is expected to have a better quality over areas with dense rain gauge networks. The study area of this work represents one of the denser networks of ground based measurements worldwide. To make robust conclusions regarding the accuracies and error propagation of TRMM3B42V7 product, further studies are needed to investigate areas with varying rain gauge data coverage and different geomorphology (e.g. complex terrain) and hydro-climatic conditions. Furthermore, this study was limited to one satellite product, while future work should also consider other global scale and possibly non-gauge-adjusted precipitation products, such as CMORPH and PERSIANN that were discussed in the introduction.

The third study proposed an error corrections and downscaling scheme to make reanalysis products useful for flood simulations. Given the improved accuracy, and error propagation properties of current high-resolution satellite precipitation products (such as TRMM3B42V7), combination with re-analysis datasets in data poor areas could lead to improved uses of the these datasets for water resource reanalysis studies. The proposed downscaling scheme was demonstrated over Northeast United States. We believe results presented in this study indicate that the proposed satellite precipitation based downscaling and error correction method has the potential to improve the hydrological use of GLDAS precipitation reanalysis datasets. The main

advantage of this approach is that it uses high-resolution global precipitation products from multi-sensor satellite observations, which makes it flexible to implement over areas with limited ground based measurements. Furthermore, the downscaling scheme is modular in design-it can be applied for any gridded dataset in any region in the world. We know there are further studies are required to finalize the qualifications of the proposed downscaling scheme. In local scale (for the proposed study area on this research) one can apply SREM2D downscaling scheme on 30 years GLDAS data with existing data setting to demonstrate the impact on flood frequency analysis derived from re-analysis products. Then the differences in the derived return-period flows for the different basin scales can be evaluated. In global scale for selected regions over the globe one can use multiple reanalysis products including ERA-40 (6-hourly, 1957-2001,~1.125 deg), ERA-interim (6-hourly, 1979-2010, ~0.75 degrees), GLDAS (3-hourly, 30-year), CPC in conjunction with 16-years (1998-2014) of TRMM-corrected high-resolution CMORPH (8km/hourly) to calibrate SREM2D downscaling for the different reanalysis products. Finally the resulted high resolution long term reanalysis product would be applicable to derive IDF curves for use in flood frequency analysis.

## References

Adler, R. F., Kidd, C., Petty, G., Morissey, M., and Goodman, H. M.: Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP–3), Bulletin of the American Meteorological Society, 82, 1377-1396, 10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2, 2001.

AghaKouchak, A., Nasrollahi, N., and Habib, E.: Accounting for Uncertainties of the TRMM Satellite Estimates, Remote Sensing, 1, 606-619, 2009.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., and Amitai, E.: Evaluation of satellite-retrieved extreme precipitation rates across the central United States, Journal of Geophysical Research: Atmospheres, 116, D02115, 10.1029/2010JD014741, 2011.

Aires, F., Prigent, C., Bernardo, F., Jiménez, C., Saunders, R., and Brunel, P.: A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 137, 690-699, 10.1002/qj.803, 2011.

Anagnostou, E. N., C. A. Morales, and T. Dinku: The Use of TRMM Precipitation Radar Observations in Determining Ground Radar Calibration Biases., J. Of Atmospheric and Oceanic Technology, 18, 618-628, 2001.

Asadullah, A., McIntyre, N., and Kigobe, M. A. X.: Evaluation of five satellite products for estimation of rainfall over Uganda / Evaluation de cinq produits satellitaires pour l'estimation des précipitations en Ouganda, Hydrological Sciences Journal, 53, 1137-1150, 10.1623/hysj.53.6.1137, 2008.

Bastola, S., and Misra, V.: Evaluation of dynamically downscaled reanalysis precipitation data for hydrological application, Hydrological Processes, 28, 1989-2002, 10.1002/hyp.9734, 2014.

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K., and Sorooshian, S.: Hydrologic evaluation of satellite precipitation products over a mid-size basin, Journal of Hydrology, 397, 225-237, http://dx.doi.org/10.1016/j.jhydrol.2010.11.043, 2011.

Beighley, R. E., Eggert, K. G., Dunne, T., He, Y., Gummadi, V., and Verdin, K. L.: Simulating hydrologic and hydraulic processes throughout the Amazon River Basin, Hydrological Processes, 23, 1221-1235, 10.1002/hyp.7252, 2009.

Beighley, R. E., Ray, R. L., He, Y., Lee, H., Schaller, L., Andreadis, K. M., Durand, M., Alsdorf, D. E., and Shum, C. K.: Comparing satellite derived precipitation datasets using the Hillslope River Routing (HRR) model in the Congo River Basin, Hydrological Processes, 25, 3216-3229, 10.1002/hyp.8045, 2011.

Bellerby, T. J., and Sun, J.: Probabilistic and Ensemble Representations of the Uncertainty in an IR/Microwave Satellite Precipitation Product, Journal of Hydrometeorology, 6, 1032-1044, 10.1175/JHM454.1, 2005.

Bellerby, T. J.: Ensemble Representation of Uncertainty in Lagrangian Satellite Rainfall Estimates, Journal of Hydrometeorology, 14, 1483-1499, 10.1175/JHM-D-12-0121.1, 2013.

Bitew, M. M., and Gebremichael, M.: Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model, Water Resources Research, 47, W06526, 10.1029/2010WR009917, 2011a.

Bitew, M. M., and Gebremichael, M.: Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands, Hydrol. Earth Syst. Sci., 15, 1147-1155, 10.5194/hess-15-1147-2011, 2011b.

Bloom, S., A. da Silva, D. Dee, M. Bosilovich, J. D. Chern, S. Pawson, S. Schubert, M. Sienkiewicz, I. Stajner, W. W. Tan, and M. L. Wu: Documentation and Validation of the Goddard Earth Observing System (GEOS) Data Assimilation System, Version 4 Technical Report NASA/TM-2005?104606/ VOL26/VER4; Rept- 2005-01264-0/VOL26/VER4, 2005.

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., Mahanama, S., Mitchell, K., Nasonova, O., Niu, G. Y., Pitman, A., Polcher, J., Shmakin, A. B., Tanaka, K., van den Hurk, B., Vérant, S., Verseghy, D., Viterbo, P., and Yang, Z. L.: The Rhône-Aggregation Land Surface Scheme Intercomparison Project: An Overview, Journal of Climate, 17, 187-208, 10.1175/1520-0442(2004)017<0187:TRLSSI>2.0.CO;2, 2004.

Bosilovich, M. G., Chen, J., Robertson, F. R., and Adler, R. F.: Evaluation of Global Precipitation in Reanalyses, Journal of Applied Meteorology and Climatology, 47, 2279-2299, 10.1175/2008JAMC1921.1, 2008.

Bowling, L. C., Lettenmaier, D. P., Nijssen, B., Graham, L. P., Clark, D. B., El Maayar, M., Essery, R., Goers, S., Gusev, Y. M., Habets, F., van den Hurk, B., Jin, J., Kahan, D., Lohmann, D., Ma, X., Mahanama, S., Mocko, D., Nasonova, O., Niu, G.-Y., Samuelsson, P., Shmakin, A. B., Takata, K., Verseghy, D., Viterbo, P., Xia, Y., Xue, Y., and Yang, Z.-L.: Simulation of high-latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 1: Experiment description and summary intercomparisons, Global and Planetary Change, 38, 1-30, http://dx.doi.org/10.1016/S0921-8181(03)00003-1, 2003.

Brown, J. E. M.: An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions, Remote Sensing of Environment, 101, 63-81, http://dx.doi.org/10.1016/j.rse.2005.12.005, 2006.

Brussolo, E., von Hardenberg, J., Ferraris, L., Rebora, N., and Provenzale, A.: Verification of Quantitative Precipitation Forecasts via Stochastic Downscaling, Journal of Hydrometeorology, 9, 1084-1094, 10.1175/2008JHM994.1, 2008.

Cloke, H. L., Wetterhall, F., He, Y., Freer, J. E., and Pappenberger, F.: Modelling climate impact on floods with ensemble climate projections, Quarterly Journal of the Royal Meteorological Society, 139, 282-297, 10.1002/qj.1998, 2013.

Dai, A., Lin, X., and Hsu, K.-L.: The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes, Clim Dyn, 29, 727-744, 10.1007/s00382-007-0260-y, 2007.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553-597, 10.1002/qj.828, 2011.

Demaria, E. M. C., Rodriguez, D. A., Ebert, E. E., Salio, P., Su, F., and Valdes, J. B.: Evaluation of mesoscale convective systems in South America using multiple satellite products

and an object-based approach, Journal of Geophysical Research: Atmospheres, 116, D08103, 10.1029/2010JD015157, 2011.

Dinku, T., and Anagnostou, E. N.: Regional Differences in Overland Rainfall Estimation from PR-Calibrated TMI Algorithm, Journal of Applied Meteorology, 44, 189-205, 10.1175/JAM2186.1, 2005.

Dinku, T., Ceccato, P., Grover- Kopec, E., Lemma, M., Connor, S., and Ropelewski, C.: Validation of satellite rainfall products over East Africa's complex topography, International Journal of Remote Sensing, 28, 1503-1526, 2007.

Ducharne, A., Koster, R. D., Suarez, M. J., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration, Journal of Geophysical Research: Atmospheres, 105, 24823-24838, 10.1029/2000JD900328, 2000.

Duethmann, D., Zimmer, J., Gafurov, A., Güntner, A., Kriegel, D., Merz, B., and Vorogushyn, S.: Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, Hydrol. Earth Syst. Sci., 17, 2415-2434, 10.5194/hess-17-2415-2013, 2013.

Ebert, E. E., and Manton, M. J.: Performance of Satellite Rainfall Estimation Algorithms during TOGA COARE, Journal of the Atmospheric Sciences, 55, 1537-1557, 10.1175/1520-0469(1998)055<1537:POSREA>2.0.CO;2, 1998.

Ebert, E. E., Janowiak, J. E., and Kidd, C.: Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models, Bulletin of the American Meteorological Society, 88, 47-64, 10.1175/BAMS-88-1-47, 2007.

FAO: Harmonized World Soil Database (version 1.2), FAO/IIASA/ISRIC/ISSCAS/JRC, FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Ferraris, L., Gabellani, S., Rebora, N., and Provenzale, A.: A comparison of stochastic models for spatial rainfall downscaling, Water Resources Research, 39, 1368, 10.1029/2003WR002504, 2003.

Ferraro, R. R., Smith, E. A., Berg, W., and Huffman, G. J.: A Screening Methodology for Passive Microwave Precipitation Retrieval Algorithms, Journal of the Atmospheric Sciences, 55, 1583-1600, 10.1175/1520-0469(1998)055<1583:ASMFPM>2.0.CO;2, 1998.

Fiorino, S. T., and Smith, E. A.: Critical Assessment of Microphysical Assumptions within TRMM Radiometer Rain Profile Algorithm Using Satellite, Aircraft, and Surface Datasets from KWAJEX, Journal of Applied Meteorology and Climatology, 45, 754-786, 10.1175/JAM2336.1, 2006.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C.: Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, International Journal of Climatology, 27, 1547-1578, 10.1002/joc.1556, 2007.

Frei, C., Schöll, R., Fukutome, S., Schmidli, J., and Vidale, P. L.: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models, Journal of Geophysical Research: Atmospheres, 111, D06105, 10.1029/2005JD005965, 2006.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J.: Completion of the 2006 National Land Cover Database for the Conterminous United States, Photogrammetric Engineering & Remote Sensing 77, 858-864, 2011.

Furuzawa, F. A., and Nakamura, K.: Differences of Rainfall Estimates over Land by Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) and TRMM Microwave

Imager (TMI)—Dependence on Storm Height, Journal of Applied Meteorology, 44, 367-383, 10.1175/JAM-2200.1, 2005.

Gopalan, K., Wang, N.-Y., Ferraro, R., and Liu, C.: Status of the TRMM 2A12 Land Precipitation Algorithm, Journal of Atmospheric and Oceanic Technology, 27, 1343-1354, 10.1175/2010JTECHA1454.1, 2010.

Gottschalck, J., Meng, J., Rodell, M., and Houser, P.: Analysis of Multiple Precipitation Products and Preliminary Assessment of Their Impact on Global Land Data Assimilation System Land Surface States, Journal of Hydrometeorology, 6, 573-598, 10.1175/JHM437.1, 2005.

Gourley, J. J., Hong, Y., Flamig, Z. L., Wang, J., Vergara, H., and Anagnostou, E. N.: Hydrologic Evaluation of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations on Ft. Cobb Basin, Oklahoma, Journal of Hydrometeorology, 12, 973-988, 10.1175/2011JHM1287.1, 2011.

Haas, R., and Born, K.: Probabilistic downscaling of precipitation data in a subtropical mountain area: a two-step approach, Nonlin. Processes Geophys., 18, 223-234, 10.5194/npg-18-223-2011, 2011.

Hagen, E., and Lu, X. X.: Let us create flood hazard maps for developing countries, Nat Hazards, 58, 841-843, 10.1007/s11069-011-9750-7, 2011.

Hamill, T. M., and Colucci, S. J.: Verification of Eta–RSM Short-Range Ensemble Forecasts, Monthly Weather Review, 125, 1312-1327, 10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2, 1997.

Hamill, T. M.: Interpretation of Rank Histograms for Verifying Ensemble Forecasts, Monthly Weather Review, 129, 550-560, 10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2, 2001.

Heinemann, T., and J. Kerényi: The EUMETSAT Multi Sensor Precipitation Estimate (MPE): Concept and Validation, Proceedings of the EUMETSAT users conference, Weimar, Germany, 2003.

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

Hong, Y., Adler, R. F., Hossain, F., Curtis, S., and Huffman, G. J.: A first approach to global runoff simulation using satellite rainfall estimation, Water Resources Research, 43, W08502, 10.1029/2006WR005739, 2007.

Hossain, F., and Anagnostou, E. N.: Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction, Journal of Geophysical Research: Atmospheres, 109, D07102, 10.1029/2003JD003986, 2004.

Hossain, F., Anagnostou, E. N., Dinku, T., and Borga, M.: Hydrological model sensitivity to parameter and radar rainfall estimation uncertainty, Hydrological Processes, 18, 3277-3291, 10.1002/hyp.5659, 2004.

Hossain, F., and Anagnostou, E. N.: Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture, Advances in Water Resources, 28, 1336-1350, http://dx.doi.org/10.1016/j.advwatres.2005.03.013, 2005.

Hossain, F., and Anagnostou, E. N.: A two-dimensional satellite rainfall error model, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1511-1522, 10.1109/TGRS.2005.863866, 2006a.

Hossain, F., and Anagnostou, E. N.: Assessment of a multidimensional satellite rainfall error model for ensemble generation of satellite rainfall data, Geoscience and Remote Sensing Letters, IEEE, 3, 419-423, 10.1109/LGRS.2006.873686, 2006b.

Hossain, F., and Huffman, G. J.: Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales, Journal of Hydrometeorology, 9, 563-575, 10.1175/2007JHM925.1, 2008.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, Journal of Hydrometeorology, 8, 38-55, 10.1175/JHM560.1, 2007.

Immerzeel, W. W., Rutten, M. M., and Droogers, P.: Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula, Remote Sensing of Environment, 113, 362-370, http://dx.doi.org/10.1016/j.rse.2008.10.004, 2009.

Janowiak, J. E., Joyce, R. J., and Yarosh, Y.: A Real–Time Global Half–Hourly Pixel–Resolution Infrared Dataset and Its Applications, Bulletin of the American Meteorological Society, 82, 205-217, 10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2, 2001.

Jia, S., Zhu, W., Lű, A., and Yan, T.: A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China, Remote Sensing of Environment, 115, 3069-3079, http://dx.doi.org/10.1016/j.rse.2011.06.009, 2011.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution, Journal of Hydrometeorology, 5, 487-503, 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2, 2004.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437-471, 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kappes, M. S., Gruber, K., Frigerio, S., Bell, R., Keiler, M., and Glade, T.: The MultiRISK platform: The technical concept and application of a regional-scale multihazard exposure analysis tool, Geomorphology, 151–152, 139-155, http://dx.doi.org/10.1016/j.geomorph.2012.01.024, 2012.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K. L., and Braithwaite, D.: Intercomparison of High-Resolution Precipitation Products over Northwest Europe, Journal of Hydrometeorology, 13, 67-83, 10.1175/JHM-D-11-042.1, 2011.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W., and Amitai, E.: Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–Based National Mosaic QPE, Journal of Hydrometeorology, 13, 1285-1300, 10.1175/JHM-D-11-0139.1, 2012a.

Kirstetter, P.-E., Viltard, N., and Gosset, M.: An error model for instantaneous satellite rainfall estimates: evaluation of BRAIN-TMI over West Africa, Quarterly Journal of the Royal Meteorological Society, 139, 894-911, 10.1002/qj.1964, 2013.

Kirstetter, P. E., Y. Hong, J.J. Gourley, Q. Cao, M. Schwaller, W. Petersen, J. Zhang, E. Anagnostou, V. Maggioni, H. Seyyedi, and S. Chen: Toward A Research Framework to Bridge Cross-platform Error Characterization of Spaceborne Passive/Active Sensors using

NOAA/NSSL Ground Radar-based National Mosaic QPE Products over CONUS., Oral presentation at AGU Fall Meeting, San Francisco, CA, U.S., Dec. 3-7, 2012, 2012b.

Kitzmiller, D., Van Cooten, S., Ding, F., Howard, K., Langston, C., Zhang, J., Moser, H., Zhang, Y., Gourley, J. J., Kim, D., and Riley, D.: Evolving Multisensor Precipitation Estimation Methods: Their Impacts on Flow Prediction Using a Distributed Hydrologic Model, Journal of Hydrometeorology, 12, 1414-1431, 10.1175/JHM-D-10-05038.1, 2011.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, Journal of Geophysical Research: Atmospheres, 105, 24809-24822, 10.1029/2000JD900327, 2000.

Krajewski, W. F., Anagnostou, E. N., and Ciach, G. J.: Effects of the radar observation process on inferred rainfall statistics, Journal of Geophysical Research: Atmospheres, 101, 26493-26502, 10.1029/96JD01616, 1996.

Krajewski, W. F., Ciach, G. J., McCollum, J. R., and Bacotiu, C.: Initial Validation of the Global Precipitation Climatology Project Monthly Rainfall over the United States, Journal of Applied Meteorology, 39, 1071-1086, 10.1175/1520-0450(2000)039<1071:IVOTGP>2.0.CO;2, 2000.

Krajewski, W. F., and Smith, J. A.: Radar hydrology: rainfall estimation, Advances in Water Resources, 25, 1387-1394, http://dx.doi.org/10.1016/S0309-1708(02)00062-3, 2002.

Kummerow, C., Olson, W. S., and Giglio, L.: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors, Geoscience and Remote Sensing, IEEE Transactions on, 34, 1213-1232, 10.1109/36.536538, 1996.

Kummerow, C., Hong, Y., Olson, W. S., Yang, S., Adler, R. F., McCollum, J., Ferraro, R., Petty, G., Shin, D. B., and Wilheit, T. T.: The Evolution of the Goddard Profiling Algorithm (GPROF) for Rainfall Estimation from Passive Microwave Sensors, Journal of Applied Meteorology, 40, 1801-1820, 10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2, 2001.

Kummerow, C., Masunaga, H., and Bauer, P.: A Next-generation Microwave Rainfall Retrieval Algorithm for use by TRMM and GPM, in: Measuring Precipitation From Space, edited by: Levizzani, V., Bauer, P., and Turk, F. J., Advances In Global Change Research, Springer Netherlands, 235-252, 2007.

Lakshmanan, V., Fritz, A., Smith, T., Hondl, K., and Stumpf, G.: An Automated Technique to Quality Control Radar Reflectivity Data, Journal of Applied Meteorology and Climatology, 46, 288-305, 10.1175/JAM2460.1, 2007.

Lensky, I., and Levizzani, V.: Estimation of precipitation from space-based platforms, in: Precipitation: Advances in Measurement, Estimation and Prediction, edited by: Michaelides, S., Springer Berlin Heidelberg, 195-217, 2008.

Liu, C., Zipser, E. J., Cecil, D. J., Nesbitt, S. W., and Sherwood, S.: A Cloud and Precipitation Feature Database from Nine Years of TRMM Observations, Journal of Applied Meteorology and Climatology, 47, 2712-2728, 10.1175/2008JAMC1890.1, 2008.

Liu, C., and Zipser, E. J.: "Warm Rain" in the Tropics: Seasonal and Regional Distributions Based on 9 yr of TRMM Data, Journal of Climate, 22, 767-779, 10.1175/2008JCLI2641.1, 2009.

Lopez, P.: Direct 4D-Var Assimilation of NCEP Stage IV Radar and Gauge Precipitation Data at ECMWF, Monthly Weather Review, 139, 2098-2116, 10.1175/2010MWR3565.1, 2011.

Maggioni, V., Reichle, R. H., and Anagnostou, E. N.: The Effect of Satellite Rainfall Error Modeling on Soil Moisture Prediction Uncertainty, Journal of Hydrometeorology, 12, 413-428, 10.1175/2011JHM1355.1, 2011.

Maggioni, V., Anagnostou, E. N., and Reichle, R. H.: The impact of model and rainfall forcing errors on characterizing soil moisture uncertainty in land surface modeling, Hydrol. Earth Syst. Sci., 16, 3499-3515, 10.5194/hess-16-3499-2012, 2012.

Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., and Stampoulis, D.: Investigating the Applicability of Error Correction Ensembles of Satellite Rainfall Products in River Flow Simulations, Journal of Hydrometeorology, 14, 1194-1211, 10.1175/JHM-D-12-074.1, 2013.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Reviews of Geophysics, 48, RG3003, 10.1029/2009RG000314, 2010.

McCollum, J. R., Krajewski, W. F., Ferraro, R. R., and Ba, M. B.: Evaluation of Biases of Satellite Rainfall Estimation Algorithms over the Continental United States, Journal of Applied Meteorology, 41, 1065-1080, 10.1175/1520-0450(2002)041<1065:EOBOSR>2.0.CO;2, 2002.

McCollum, J. R., and Ferraro, R. R.: Next generation of NOAA/NESDIS TMI, SSM/I, and AMSR-E microwave land rainfall algorithms, Journal of Geophysical Research: Atmospheres, 108, 8382, 10.1029/2001JD001512, 2003.

McCuen, R., Knight, Z., and Cutter, A.: Evaluation of the Nash–Sutcliffe Efficiency Index, Journal of Hydrologic Engineering, 11, 597-602, 10.1061/(ASCE)1084-0699(2006)11:6(597), 2006.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282-290, http://dx.doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nijssen, B., Bowling, L. C., Lettenmaier, D. P., Clark, D. B., El Maayar, M., Essery, R., Goers, S., Gusev, Y. M., Habets, F., van den Hurk, B., Jin, J., Kahan, D., Lohmann, D., Ma, X., Mahanama, S., Mocko, D., Nasonova, O., Niu, G.-Y., Samuelsson, P., Shmakin, A. B., Takata, K., Verseghy, D., Viterbo, P., Xia, Y., Xue, Y., and Yang, Z.-L.: Simulation of high latitude hydrological processes in the Torne–Kalix basin: PILPS Phase 2(e): 2: Comparison of model results with observations, Global and Planetary Change, 38, 31-53, http://dx.doi.org/10.1016/S0921-8181(03)00004-3, 2003.

Nijssen, B., and Lettenmaier, D. P.: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites, Journal of Geophysical Research: Atmospheres, 109, D02103, 10.1029/2003JD003497, 2004.

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., and Borga, M.: Understanding the Scale Relationships of Uncertainty Propagation of Satellite Rainfall through a Distributed Hydrologic Model, Journal of Hydrometeorology, 11, 520-532, 10.1175/2009JHM1169.1, 2010.

Nikolopoulos, E. I., Anagnostou, E. N., and Borga, M.: Using High-Resolution Satellite Rainfall Products to Simulate a Major Flash Flood Event in Northern Italy, Journal of Hydrometeorology, 14, 171-185, 10.1175/JHM-D-12-09.1, 2012.

Ning, L., Mann, M. E., Crane, R., and Wagener, T.: Probabilistic Projections of Climate Change for the Mid-Atlantic Region of the United States: Validation of Precipitation Downscaling during the Historical Era*, Journal of Climate, 25, 509-526, 10.1175/2011JCLI4091.1, 2011.

Ogden, F. L., Sharif, H. O., Senarath, S. U. S., Smith, J. A., Baeck, M. L., and Richardson, J. R.: Hydrologic analysis of the Fort Collins, Colorado, flash flood of 1997, Journal of Hydrology, 228, 82-100, http://dx.doi.org/10.1016/S0022-1694(00)00146-3, 2000.

Olson, W. S., Kummerow, C. D., Hong, Y., and Tao, W.-K.: Atmospheric Latent Heating Distributions in the Tropics Derived from Satellite Passive Microwave Radiometer Measurements, Journal of Applied Meteorology, 38, 633-664, 10.1175/1520-0450(1999)038<0633:ALHDIT>2.0.CO;2, 1999.

Onogi, K., Koide, H., Sakamoto, M., Kobayashi, S., Tsutsui, J., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kato, K., Oyama, R., Ose, T., Kadokura, S., and Wada, K.: JRA-25: Japanese 25-year re-analysis project—progress and status, Quarterly Journal of the Royal Meteorological Society, 131, 3259-3268, 10.1256/qj.05.88, 2005.

Park, N.-W.: Spatial Downscaling of TRMM Precipitation Using Geostatistics and Fine Scale Environmental Variables, Advances in Meteorology, 2013, 9, 10.1155/2013/237126, 2013.

Peña-Arancibia, J. L., van Dijk, A. I. J. M., Renzullo, L. J., and Mulligan, M.: Evaluation of Precipitation Estimation Accuracy in Reanalyses, Satellite Products, and an Ensemble Method for Regions in Australia and South and East Asia, Journal of Hydrometeorology, 14, 1323-1333, 10.1175/JHM-D-12-0132.1, 2013.

Porter, J., and Demeritt, D.: Flood-risk management, mapping, and planning: theÿinstitutional politics of decision support in England, Environment and Planning A, 44, 2359-2378, 2012.

Rahman, S., Bagtzoglou, A. C., Hossain, F., Tang, L., Yarbrough, L. D., and Easson, G.: Investigating Spatial Downscaling of Satellite Rainfall Data for Streamflow Simulation in a Medium-Sized Basin, Journal of Hydrometeorology, 10, 1063-1079, 10.1175/2009JHM1072.1, 2009.

Ramírez, M. C., Ferreira, N. J., and Velho, H. F. C.: Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil, Weather and Forecasting, 21, 969-989, 10.1175/WAF981.1, 2006.

Rawls W.J., Brakensiek, D. L., and Saxton, K. E.: Estimation of soil water properties, Transactions of the ASAE 25, 1316-1320, 1982.

Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., and Owe, M.: Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR), Journal of Geophysical Research: Atmospheres, 112, D09108, 10.1029/2006JD008033, 2007.

Rienecker, M. M., M.J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, M. Sienkiewicz, R.D. Koster, R. Gelaro, I. Stajner, and J.E. Nielsen: The GEOS-5 Data Assimilation System-Documentation of Versions 5.0.1, 5.1.0, and 5.2.0, NASA/TM, 2008.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J. K., Walker, J. P.,

Lohmann, D., and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381-394, 10.1175/BAMS-85-3-381, 2004.

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. B., Freer, J. E., and Cloke, H. L.: The impact of uncertain precipitation data on insurance loss estimates using a Flood Catastrophe Model, Hydrol. Earth Syst. Sci. Discuss., 11, 31-81, 10.5194/hessd-11-31-2014, 2014.

Sapiano, M. R. P., and Arkin, P. A.: An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data, Journal of Hydrometeorology, 10, 149-166, 10.1175/2008JHM1052.1, 2009.

Saxton, K. E., Rawls, W. J., Romberger, J. S., and Papendick, R. I.: Estimating Generalized Soil-water Characteristics from Texture1, Soil Sci. Soc. Am. J., 1031-1036, 1986.

Scofield, R. A., and Kuligowski, R. J.: Status and Outlook of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events, Weather and Forecasting, 18, 1037-1051, 10.1175/1520-0434(2003)018<1037:SAOOOS>2.0.CO;2, 2003.

Seto, S., Takahashi, N., and Iguchi, T.: Rain/No-Rain Classification Methods for Microwave Radiometer Observations over Land Using Statistical Information for Brightness Temperatures under No-Rain Conditions, Journal of Applied Meteorology, 44, 1243-1259, 10.1175/JAM2263.1, 2005.

Seyyedi, H., Anagnostou, E. N., Kirstetter, P. E., Maggioni, V., Hong, Y., and Gourley, J. J.: Incorporating Surface Soil Moisture Information in Error Modeling of TRMM Passive Microwave Rainfall, Geoscience and Remote Sensing, IEEE Transactions on, PP, 1-16, 10.1109/TGRS.2013.2295795, 2014.

Shrestha, M. S., Artan, G. A., Bajracharya, S. R., and Sharma, R. R.: Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin, Journal of Flood Risk Management, 1, 89-99, 10.1111/j.1753-318X.2008.00011.x, 2008.

Siegert, S., Bröcker, J., and Kantz, H.: Rank Histograms of Stratified Monte Carlo Ensembles, Monthly Weather Review, 140, 1558-1571, 10.1175/MWR-D-11-00302.1, 2012.

Simpson, J., Kummerow, C., Tao, W. K., and Adler, R. F.: On the Tropical Rainfall Measuring Mission (TRMM), Meteorl. Atmos. Phys., 60, 19-36, 10.1007/BF01029783, 1996.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN System Satellite–Based Estimates of Tropical Rainfall, Bulletin of the American Meteorological Society, 81, 2035-2046, 10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2, 2000.

Stampoulis, D., Anagnostou, E. N., and Nikolopoulos, E. I.: Assessment of High-Resolution Satellite-Based Rainfall Estimates over the Mediterranean during Heavy Precipitation Events, Journal of Hydrometeorology, 14, 1500-1514, 10.1175/JHM-D-12-0167.1, 2013.

Steiner, M., Houze, R. A., and Yuter, S. E.: Climatological Characterization of Three-Dimensional Storm Structure from Operational Radar and Rain Gauge Data, Journal of Applied Meteorology, 34, 1978-2007, 10.1175/1520-0450(1995)034<1978:CCOTDS>2.0.CO;2, 1995.

Su, F., Hong, Y., and Lettenmaier, D. P.: Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin, Journal of Hydrometeorology, 9, 622-640, 10.1175/2007JHM944.1, 2008.

Syed, T. H., Lakshmi, V., Paleologos, E., Lohmann, D., Mitchell, K., and Famiglietti, J. S.: Analysis of process controls in land surface hydrological cycle over the continental United

States, Journal of Geophysical Research: Atmospheres, 109, D22105, 10.1029/2004JD004640, 2004.

Tang, L., Hossain, F., and Huffman, G. J.: Transfer of Satellite Rainfall Uncertainty from Gauged to Ungauged Regions at Regional and Seasonal Time Scales, Journal of Hydrometeorology, 11, 1263-1274, 10.1175/2010JHM1296.1, 2010.

Tao, K., and Barros, A. P.: Using Fractal Downscaling of Satellite Precipitation Products for Hydrometeorological Applications, Journal of Atmospheric and Oceanic Technology, 27, 409-427, 10.1175/2009JTECHA1219.1, 2010.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., and de Castro, M.: Global precipitation measurement: Methods, datasets and applications, Atmospheric Research, 104–105, 70-97, http://dx.doi.org/10.1016/j.atmosres.2011.10.021, 2012.

Teo, C.-K.: Application of satellite-based rainfall estimates to crop yield forecasting in Africa, Ph.D. thesis, University of Reading, 242 pp., 2006.

Turk, F. J., and Miller, S. D.: Toward improved characterization of remotely sensed precipitation regimes with MODIS/AMSR-E blended data techniques, Geoscience and Remote Sensing, IEEE Transactions on, 43, 1059-1069, 10.1109/TGRS.2004.841627, 2005.

Turk, F. J., Huffman, G. J., Joyce, R., Kidd, C., and Kuligowski, R.: Evaluation of satellite-based estimates of precipitation in the Yucatan region during Hurricane Wilma 27th Conf. on Hurricanes and Tropical Meteorology Monterey, CA, 2006.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L.

V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society, 131, 2961-3012, 10.1256/qj.04.176, 2005.

Ushio, T., and Kachi, M.: Kalman filtering applications for global satellite mapping of precipitation (GSMaP), in: Satellite Rainfall Applications for Surface Hydrology, edited by: Gebremichael, M., and Hossain, F., Springer-Verlag, New York, 2009.

Vasiloff, S. V., Howard, K. W., Rabin, R. M., Brooks, H. E., Seo, D.-J., Zhang, J., Kitzmiller, D. H., Mullusky, M. G., Krajewski, W. F., Brandes, E. A., Brown, B. G., Berkowitz, D. S., McGinley, J. A., and Kuligowski, R. J.: Improving QPE and Very Short Term QPF: An Initiative for a Community-Wide Integrated Approach, Bulletin of the American Meteorological Society, 88, 1899-1911, 10.1175/BAMS-88-12-1899, 2007.

Vergara, H., Hong, Y., Gourley, J. J., Anagnostou, E. N., Maggioni, V., Stampoulis, D., and Kirstetter, P.-E.: Effects of Resolution of Satellite-based Rainfall Estimates on Hydrologic Modeling Skill at Different Scales, Journal of Hydrometeorology, 10.1175/JHM-D-12-0113.1, 2013.

Villarini, G., and Krajewski, W. F.: Sensitivity Studies of the Models of Radar-Rainfall Uncertainties, Journal of Applied Meteorology and Climatology, 49, 288-309, 10.1175/2009JAMC2188.1, 2010.

Vivoni, E. R., Entekhabi, D., Bras, R. L., Ivanov, V. Y., Van Horne, M. P., Grassotti, C., and Hoffman, R. N.: Extending the Predictability of Hydrometeorological Flood Events Using

Radar Rainfall Nowcasting, Journal of Hydrometeorology, 7, 660-677, 10.1175/JHM514.1, 2006.

Wang, N.-Y., Liu, C., Ferraro, R., Wolff, D., Zipser, E., and Kummerow, C.: TRMM 2A12 Land Precipitation Product - Status and Future Plans, Journal of the Meteorological Society of Japan. Ser. II, 87A, 237-253, 2009.

Weng, F., Zhao, L., Ferraro, R. R., Poe, G., Li, X., and Grody, N. C.: Advanced microwave sounding unit cloud and precipitation algorithms, Radio Science, 38, 8068, 10.1029/2002RS002679, 2003.

Wu, H., Adler, R. F., Hong, Y., Tian, Y., and Policelli, F.: Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model, Journal of Hydrometeorology, 13, 1268-1284, 10.1175/JHM-D-11-087.1, 2012.

Xie, P., and Arkin, P. A.: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs, Bulletin of the American Meteorological Society, 78, 2539-2558, 10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2, 1997.

Yong, B., Hong, Y., Ren, L.-L., Gourley, J. J., Huffman, G. J., Chen, X., Wang, W., and Khan, S. I.: Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin, Journal of Geophysical Research: Atmospheres, 117, D09108, 10.1029/2011JD017069, 2012.

Zeweldi, D. A., and Gebremichael, M.: Evaluation of CMORPH Precipitation Products at Fine Space–Time Scales, Journal of Hydrometeorology, 10, 300-307, 10.1175/2008JHM1041.1, 2009.

Zhang, J., Howard, K., and Gourley, J. J.: Constructing Three-Dimensional Multiple-Radar Reflectivity Mosaics: Examples of Convective Storms and Stratiform Rain Echoes, Journal of Atmospheric and Oceanic Technology, 22, 30-42, 10.1175/JTECH-1689.1, 2005.

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., Van Cooten, S., Kelleher, K., Kitzmiller, D., Ding, F., Seo, D.-J., Wells, E., and Dempsey, C.: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, Results, and Future Plans, Bulletin of the American Meteorological Society, 92, 1321-1338, 10.1175/2011BAMS-D-11-00047.1, 2011.

Zhao, L., and Weng, F.: Retrieval of Ice Cloud Parameters Using the Advanced Microwave Sounding Unit, Journal of Applied Meteorology, 41, 384-395, 10.1175/1520-0450(2002)041<0384:ROICPU>2.0.CO;2, 2002.

Zipser, E. J., Liu, C., Cecil, D. J., Nesbitt, S. W., and Yorty, D. P.: WHERE ARE THE MOST INTENSE THUNDERSTORMS ON EARTH?, Bulletin of the American Meteorological Society, 87, 1057-1071, 10.1175/BAMS-87-8-1057, 2006.