7-27-2015

Evaluation of High-Resolution Satellite Precipitation Products in Hydrologic Simulations of Northern Latitude River Basins

Muhammet O. Dis
momerdis@engr.uconn.edu

Evaluation of High-Resolution Satellite Precipitation Products in Hydrologic Simulations of Northern Latitude River Basins

Muhammet Omer Dis, Ph.D.

University of Connecticut, 2015

The goal of this research is to develop a framework to improve satellite precipitation applications in hydrologic modeling, and is carried out in three parts. First, the multi-scale variability of streamflow in the Connecticut River Basin (CRB) (i.e. study area) is captured using the Coupled Routing and Excess STorage (CREST) model. This analysis presents a good opportunity to enhance and refine the estimation of hydrological processes via the CREST model and understand its calibration and validation procedures over northern-latitude basins. It introduces a large time frame and focuses on different time resolutions and multiple basin scales (191–25,000 km^2). Available observations are used to assess the accuracy of simulations, determine the spatial variability of calibration parameters, and understand the applicability of calibrated parameter values applied to interior catchments in order to predict flows at ungauged locations.

In the second research dimension, multiple high-resolution precipitation products from currently available quasi-global satellite algorithms (TRMM-3B42V7, CMORPH, and PERSIANN) are investigated in terms of their hydrologic applicability relative to the National Weather Service (Stage IV) gauge-adjusted radar-rainfall dataset, which is a novel analysis in terms of satellite error propagation issues. Basin-average satellite precipitation and resulting streamflow simulations are evaluated for a period of 9 years (2002 through 2010) using as reference the Stage IV rainfall and associated CREST-based flow simulations. Error metrics in basin-average rainfall and runoff simulations are determined for high values (>90^{th} percentile) and overall (>20^{th} percentile), warm vs. cold periods and two temporal aggregations (3 hourly and daily). Results show that the TRMM-3B42V7 product, which includes gauge adjustment, is

the most accurate in terms of capturing the high percentile values of precipitation and runoff over the study basins. The satellite-alone CMORPH and PERSIANN precipitation products consistently underestimated the radar precipitation and reference-driven streamflow simulations. The study demonstrated a basin scale dependency of the satellite-precipitation error metrics and the precipitation-to-runoff simulation propagation properties of these errors.

The third part was motivated by the above findings to evaluate the accuracy of return period flow estimations using a hydrologic model driven with high-resolution satellite precipitation data. Four quasi-global satellite products (TRMM-3B42V7, TRMM-3B42RT, CMORPH, and PERSIANN) at 3-hourly/0.25-deg resolution and the National Weather Service (Stage IV) gauge-adjusted radar rainfall dataset (representing the reference rainfall) are integrated in this study with the CREST model to simulate 3-hourly and daily temporal resolution annual peak flows during warm season (May-November) months in CRB. The log-Pearson type III frequency distribution applied on an 11 year (2002 to 2012) record of annual peak flow data is used to derive 1.01-yr, 5-yr, 10-yr, 20-yr and 50-yr return period flows. Evaluation against the Stage IV driven simulations shows that the TRMM-3B42V7 product has the highest correlation and a slight overestimation in terms of the derived annual maxima flows compared to the other satellite products. Each basin exhibits distinct flood frequency curves derived from reference precipitation data, which are well captured by the flood frequency analyses of the various satellite products driven simulations. With the increasing record length of high-resolution satellite products, results from this study can motivate future studies on flood frequency analysis of ungauged basins driven by remotely sensed precipitation.

Evaluation of High-Resolution Satellite Precipitation Products in Hydrologic Simulations of Northern Latitude River Basins

Muhammet Omer Dis

B.S., Cukurova University, 2007

M.S., University of Connecticut, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of Doctor of Philosophy

at the

University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

Evaluation of High-Resolution Satellite Precipitation Products in Hydrologic Simulations of Northern Latitude River Basins

by

Muhammet Omer Dis, B.S., M.S.

Major Advisor _____

Emmanouil N. Anagnostou

Associate Advisor _____

Amvrossios C. Bagtzoglou

Associate Advisor _____

Marina Astitha

Associate Advisor _____

Glenn Warner

Associate Advisor _____

Guiling Wang

University of Connecticut

2015

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful

Elhamdulillah, all praises to Allah for the strengths and His blessing in completing this dissertation.

Next, I owe my deepest appreciation to my parents, family and my wonderful wife for their endless and unconditional love, prayers and support throughout my life. My siblings deserve my wholehearted thanks as well. Their sacrifices have benefitted us in many ways, and our accomplishments in life would not have been possible without their love and dedication.

I would like to sincerely express my heartfelt gratefulness to my advisor Prof. Emmanouil N. Anagnostou for his guide and support. He is not only a supervisor believing in me from the very beginning and providing me with the opportunity to pursue my degree at UCONN, but also an elder brother that I learned from the best.

I would also like to thank my other doctoral committee members, Dr Bagtzoglou, Dr Astitha, Dr Warner, and Dr Wang for their valuable guidance and service in my PhD dissertation.

I consider myself very fortunate for having a chance to work with extraordinary colleagues and being wonderful friends whom we had lots of memories. I would like to acknowledge the support of the various members of the Prof. Anagnostou's research group, particularly Yiwen Mei for useful discussions and his contributions towards of this project.

Last but not least, I would like to thank Turkish Government for the Ministry of Education's State scholarship towards my education.

To My Wife and Parents

For their endless love, supports and encouragement throughout my life

Table of Contents

List of Figures

List of Tables

1. Executive Summary

Precipitation is arguably one of the most important parameters controlling the hydrologic cycle over land. Flood forecasting, water management and other hydrological applications rely principally on the accurate quantification of precipitation at different space-time scales. Given the lack of in situ measurements of precipitation over many regions on earth, remote sensing of precipitation is the only data source for quantifying spatiotemporal variations of precipitation at global scale. However, the uncertainty in satellite precipitation estimation at local scale hampers the data's hydrologic applicability. Previous studies have shown that validation of satellite precipitation and understudying their error propagation in hydrological simulations can provide a strong basis for motivating uses in hydrological applications (Seyyedi, et al., 2015; Behrangi, et al., 2014; Nikolopoulos, et al., 2013; Maggioni, et al., 2013; Bitew & Gebremichael, 2010; Shrestha, et al., 2008).

Arguably, the availability of high-resolution satellite precipitation products provides a remarkable advantage for studying flood hydrology and water resources over data poor areas. Near-real-time satellite-forced runoff simulations can help to mitigate the damages caused by devastating flood hazards, such as the loss of life and property, through early warning systems. In addition, contribution of satellite-derived streamflow simulations is an undeniable importance to improve hydrologic prediction capability for water management such as water supply and improving water quality (Apip, et al., 2010; Hong, et al., 2010; Hong, et al., 2007; Bastiaanssen, 1998; Kite & Pietroniro, 1996). High spatiotemporal resolution data is available globally as well as locally, and this yields to work with various basin sizes at different time scales (Mei, et al.,

2014; Gao & Liu, 2013; Yong, et al., 2012; Beighley, et al., 2011; Tang, et al., 2010; Pan, et al., 2010; Hirpa, et al., 2010; Artan, et al., 2007).

Numerous studies have been carried out in order to improve understanding about satellite-precipitation products and associated hydrologic applicability over different regions. Hirpa et al. (2010) compared the performance of three high-resolution satellite products (TRMM, CMORPH, and PERSIANN) in capturing the reference rainfall measurements at different elevations in the Ethiopian highland. They concluded that TRMM and CMORPH follow similar trend in elevation dependency while PERSIANN considerably underestimates rainfall in high-elevation areas. Nikolopoulos et al. (2013) forced a hydrologic model with three satellite products (3B42-V6, CMORPH, and PERSIANN-CCS) for a single case study over a mountainous basin in northeastern Italian Alps. They found that the satellite-driven streamflow simulations are not able to explain the observed hydrologic response during the flash flood event. There is significant bias in terms of volume and peak runoff. Moreover, Bitew & Gebremichael (2011) claimed that even though satellite precipitation resulting runoff simulations capture the tendency and shape of the streamflow observations, they underestimate the large flood peaks. They examined performance of satellite-driven flow simulations over two distinctive watersheds areas in Ethiopia. They found that increasing watershed area increases the simulation performance and dampening the random error in particularly CMORPH precipitation estimates due to additional averaging process in larger watersheds.

Seyyedi et al. (2014) used a distributed hydrologic model forced by TRMM and Global Land Data Assimilation reanalysis precipitation datasets to simulate flows over Susquehenna River Basin, in Northeast US. They found that TRMM Multisatellite Precipitation Analysis (TMPA) 3B42 version 7 (TRMM-3B42V7) is better in capturing the flow predictions and suggested the

consideration of multiple satellite products. Moreover, Khan et al. (2011) implemented the Coupled Routing and Excess Storage (CREST) model with TRMM-based TMPA 3B42 Real-Time rainfall product for flood inundation mapping in Lake Victoria basin in Africa. They showed that TRMM data can be applicable to calibrated hydrologic models with sufficient accuracy in predicting spatial flood extends. They suggested the use of ground-based observations and multiple remote sensing products to help for further improvements in flood monitoring.

Shrestha et al. (2008) applied a semi-distributed hydrologic model to obtain flow simulations with daily satellite rainfall estimates over a catchment located in Nepal. They found significant underestimation in the simulated discharges from coarse resolution data and emphasized the necessity of high resolution precipitation data in long-term hydrologic simulations. Gourley et al. (2011) implemented Hydrology Laboratory Research Distributed Hydrologic Model over a small basin (342-km^2) in Oklahoma. The calibrated hydrologic model was forced with combinations of radar, satellite products, and gauge rainfall data. In the study, they highlighted the importance of considering high resolution rainfall during model calibration for the same satellite product (ARS Micronet). They also concluded that the TRMM-based (0.25°) simulations are better than the PERSIANN-driven (0.04°) simulations.

Flood frequency analysis requires long data records of flows, which is not typically available in poorly gauged basins (Leclerc & Ouarda, 2007; Javelle, et al., 2002). Given the currently available long record of high-resolution satellite products we are interested to evaluate potential uses of these data in flood frequency analysis as well as in deriving threshold based flood early warning techniques. Boughten et al. (1986) presented regional flood frequency techniques for application in ungauged watersheds (drainage areas ranging from 0.01 to 10,000 km^2) in

southeastern Arizona. Awadallah et al. (2011) used jointly limited rainfall records (8-yr) and data from TRMM (11-yr) to generate Intensity Duration Frequency (IDF) curves over North-West Angola. Cunderlik & Ouarda (2007) examined converging flood duration frequency with the rainfall depth duration frequency based on relationships between statistical properties of flood and rainfall. This study built upon these studies to investigate uses of streamflow simulations by satellite rainfall data for deriving flood frequency maps for basins of various scales. To elucidate applicability limitations, the following questions aim to address in this study -

What is the accuracy of current high-resolution satellite precipitation products in hydrologic simulations of northern-latitude river basins?

What is the basin scale and event severity effect on the satellite-based runoff simulation uncertainty?

What is the skill of satellite-driven flood frequency relative to current radar-rainfall driven flood simulations?

The main objectives of this work are -

i. Use available observations to assess the accuracy of simulations, determine the spatial variability of calibration parameters, and understand the applicability of calibrated parameter values applied to interior catchments in order to predict flows at ungauged locations,

ii. Use the distributed hydrologic model to evaluate the error propagation from the precipitation and associated runoff simulations for high resolution multiple satellite-precipitation products (TRMM-3B42V7, TRMM-3B42RT, PERSIANN, and CMORPH), and evaluate the

error properties for different temporal resolutions and seasonal effects as well as basin scale dependency conditions,

iii. Understand how accurate the current high-resolution satellite precipitation products are for deriving flood frequency maps,

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

In chapter 2, calibration and validation of rainfall-runoff hydrologic model, the Coupled Routing and Excess Storage (CREST), are carried out over Connecticut River Basin (CRB) at various selected observation stream gauges. The auto-calibration routine based on Differential Evolution Adaptive Metropolis (DREAM) method is devised to calibrate CREST model over a number of watershed areas in CRB based on hourly stream flow data for the period 2005 to 2007. The algorithm runs to mitigate sum of squared residuals to estimate posterior parameters using multiple Markov Chain Monte Carlo chains from prior iterations (Vrugt, et al., 2009). The simulations are started with a warm-up period to avoid the melting season and reduce the initial condition effect. After optimization of parameters, CREST is validated for a four year period (2003-2004 & 2008-2009). The model performance evaluation is based on different metrics, including the Nash-Sutchliffe Coefficient of Efficiency (NSCE), Mean Relative Error (MRE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (PCC). Three nested subbasins are considered as ungauged interior catchments. After model is calibrated for the outlet of the parent basin, the parameters are fixed, and then CREST is simulated for interior catchments.

In chapter 3, three widely used satellite-precipitation products, TRMM-3B42V7, CMORPH, and PERSIANN, are investigated in terms of their use in simulating stream flows for different

sub-basins (scales ranging from 191 to 25,000 km^2) of CRB using the CREST model. Basin-average satellite precipitation and resulting stream flow simulations are evaluated for a period of 9 years (2002 through 2010) using as reference the National Weather Service (Stage IV) gauge-adjusted radar rainfall and associated CREST-based flow simulations. Error metrics in basin-average rainfall and runoff simulations determined for high (90^{th} percentile) and low values (20^{th} percentile) are compared for the different basin scales, warm/cold periods (Nov-April, May-Oct) and two temporal aggregations (3 hourly and daily).

In chapter 4, 11-year (2002-2012) satellite precipitation products-driven flow simulations are obtained from the calibrated model to derive flood frequencies over the CRB. Four quasi-global satellite products (TRMM-3B42V7, TRMM-3B42RT, CMORPH, and PERSIANN) and the Stage IV rainfall-driven simulations (reference) are integrated with the CREST model to estimate annual peak flows over the CRB. May through November period is assigned to assure the assumption that homogeneity for the generation mechanism of flood are only affected rainfall process. By fitting simulated flood events to a log-Pearson type III frequency distribution, streamflow simulations used to estimate peak flow magnitudes in this period for exceedance probabilities of 0.99, 0.50, 0.20, 0.10, 0.04, and 0.02 (1.01-yr, 2-yr, 5-yr, 10-yr, 25-yr, and 50-yr return periods, respectively). The statistical moment of the distribution namely skew coefficient adjusted for generalized skew coefficient and frequency estimates tested for outliers. Additionally, the approach is repeated to estimate frequencies for 5^{th}, 25^{th}, 75^{th}, and 95^{th} percentiles. The skill of satellite-driven peak flow simulations are evaluated by multiple error metrics and flood frequency schemes compared relative to reference for the two temporal resolutions over the study basins. Advantage of the study is that the approach relies on remote

sensing rainfall data, which are not only more abundant than in situ measurements but also applicable to ungauged catchments.

2. Modeling Warm Season Flows in the Connecticut River Basin using the Coupled Routing and Excess Storage Model

2.1. Introduction

Rainfall-runoff modeling has a long history; the first hydrologist that used rainfall-runoff model was an Irish engineer Thomas James Mulvaney (1822-1892) who published his work in 1851. During the last few decades, a number of conceptual and physically-based models have been developed and used for simulation of floods (Wang, et al., 2011; Odoni & Lane, 2010; Dingman, 2002; Beven, 2000; Chow, et al., 1988; Crawford & Linsley, 1966). The analysis of hydrological model simulations and their spatiotemporal fluctuations can be used as vital tools to support management activities such as flood risk management, water supply and improving water quality. According to Brakendridge et al. (2003), their study shows that an increasing population density has caused a greater risk regarding the natural processes, such as flooding, which makes it crucial to identify such risky areas. Dehotin & Braud (2008) mentioned the importance of distributed hydrologic models, indicating that they are valuable mechanisms to spatially and realistically study the prediction of water balance components. Hydrologic modeling can be employed to evaluate flood mitigation alternatives to flood and drought risks. They can be used to evaluate and study the impact of land use on water resources. Additionally, Moriasi et al. (2007) states that hydrological modeling is an essential tool for managing water quantity and quality. The spatial structure of distributed hydrological models can inspect and

evaluate the heterogeneities of watershed characteristics and its parameters (Dehotin & Braud, 2008; Refsgaard, 1997).

Even though distributed models capture sufficient details and realistic catchment characteristics, over parameterization can be a problem in calibration. Dehotin & Braud (2008) state the growing of concerns about optimum parameterization. They indicated that the major concern is the contrasting aspects of model complexity versus data availability. Rozalis et al. (2010) states that simplicity versus complexity of hydrological models is still a controversial subject for better representation of a catchment. In their study, a simpler model was used with minimum number of parameters. The smaller the number of parameters does not require calibration to decrease uncertainty over ungauged areas. Moreover, Bergstrom (1991) claimed an excessive increase in model complexity does not always improve the quality of the results. Brandt (1990) supported this statement and further elaborated in his study that going from complex to simpler model does not affect model performance. Thus, using a model with optimal complexity relative to data availability and resolution is the key to improving hydrologic predictability. Also, it is worth mentioning that one of the important parameterizations in the modeling are the initial conditions and associated parameters sets. The initial condition of a model mostly depends on catchment area, catchment topography, antecedent moisture condition, ground water table position and land use. Hence, for distributed hydrological modeling, parameterizations of these initial conditions and its spatial variability are the fundamental factors for runoff simulations, especially for extreme rainfall events (Noto, et al., 2008). The influence of initial conditions on the model output reduces with increasing the simulation time period (Vrugt, et al., 2009). Additionally, a spin-up or run-up period can be used to diminish the

sensitivity to initial conditions. In this study a one-year spin up period was used in the model simulation. Therefore, the number of initialization parameters was efficiently reduced.

Two other significant issues for hydrological modeling are related to calibration and validation methodology, specifically the metrics used to analyze and evaluate simulation results (Moriasi, et al., 2007; Boyle, et al., 2000; Refsgaard, 1997; Bergstrom, 1991; Beven, 1989). In this study, we evaluated the accuracy of a grid-based distributed hydrologic model, which was calibrated and validated in the Connecticut River Basin. The performance of the model related to calibration and verification was evaluated using the Nash-Sutchliffe Coefficients of Efficiency (NSCE), Mean Relative Error (MRE), Root Mean Square Error (RMSE), and the Pearson Correlation Coefficient (PCC), which compared supplemental normalized simulations with obtained hourly observations obtained for the basin. For this analysis, a continuous simulation was applied for the hydrological analysis in the Connecticut River Basin using the Coupled Routing and Excess Storage (CREST) Hydrological Model. CREST model was selected for this study because the model currently represents both a national (Hoedjes, et al., 2014) and global flood simulation system (Wang, et al., 2011). Similar to other distributed models, catchments are represented as grid cells to simulate spatio-temporal variation of water, energy fluxes, and storage.

One of the purposes of this work is to examine the performance of the CREST model with calibrated parameters over interior sub-basin stations. Calibrating a model at a larger scale basin and using these parameters to model ungauged locations is one of the popular ways in situations where no observations are available. According to Bingeman et al. (2006), a calibrated hydrological model based on streamflow data should be applicable on other sub-catchments without recalibration of the parameters set. State variables of basin characteristics related to land

cover and soil texture for a given set of watersheds can be applied to other watersheds that are hydrologically similar without extensive recalibration (Motovilov, et al., 1999). Kouwen et al. (1993) study also confirmed the fact that a parameter set can effectively transferred between watersheds to simulate peak flow without further calibration in southern Ontario. Additionally, Xie et al. (2007) transferred calibrated parameters from data-rich areas to data-sparse areas and the results showed promising in estimating daily runoff with transferred variables. Hence, CREST was calibrated against streamflow observations at sub-basin outlets and was used to simulate streamflows based on the same parameter set at its interior nested sub-basins without further calibration.

This case study aims to capture the multi-scale variability of streamflow in the Connecticut River Basin using the CREST model following the calibration procedures suggested in the literature. Potential impacts of these fluctuations are important for water supplies, during the peak seasons of water demand, and mitigating flooding risk associated with peak flows. This study presents an opportunity to enhance and refine the estimation of hydrological processes via the CREST model and understand its calibration, parameterization and validation procedures over a mid-latitude basin. The main objectives of this work are: (1) use available observations to assess the accuracy of simulations, (2) determine the spatial variability of calibration parameters, and (3) understand the applicability of calibrated parameter values applied to interior catchments in order to predict flows at ungauged basin locations. Based on the topography and USGS observation gauges, the Connecticut River Basin was divided into nine sub-basins for all CREST simulations. Knowledge of the fluctuations and the accuracy of the predictions can assist in performing simulations over ungauged locations. For this purpose, three of the sub-basins, which

have enough interior locations for ungauged analysis, were further subdivided into sub-watersheds in order to perform an analysis of parent basin calibration parameter runs.

The paper is organized as follows. Section 2.2 describes the study area, data, and the CREST hydrological model. Section 2.3 provides information regarding the model calibration, parameterization and validation with analysis of evaluated statistics. Section 2.4 provides a discussion of the results of the hydrological simulation using the CREST model for the Connecticut River Basin. The conclusions and future work are discussed in the section 2.5.

2.2.Study Region and Data

The Connecticut River Basin (CRB) is the study basin for this work and the CRB is a major river basin in New England. Runoff from the CRB discharges to the Connecticut River. The Connecticut River starts in Quebec Canada and runs through Connecticut (CT), Massachusetts (MA), New Hampshire (NH), and Vermont (VT) and empties into Long Island Sound in Old Lyme Connecticut. The total watershed area contributing to the Connecticut River is approximately 28,500 km^2. There are approximately 390 towns and cities located within the watershed with a total population of approximately 2.3 million people. The land uses that are within the watershed consist of forest, agriculture, residential and water. Approximately, 79% is covered by forested, 11% by agriculture and the remaining area is covered by residential and water. The CT River flows for about 660 km and provides hydroelectric power, is navigable up to Windsor Locks, used for irrigation, and is used for recreation (Marshall & Randhir, 2008).

Instantaneous records of river discharges obtained from nine United States Geological Survey (USGS) gauging stations within the Connecticut River Basin were analyzed in this study. The location of the gauging stations and their contributing sub-basins are illustrated in Figure

2.1. The black dots represent the streamflow gauges, which are labeled with USGS station numbers. The locations of USGS monitoring stations were used to determine the number of sub-basins based on the location and topography and their watersheds area. Given a particular gauging station along with the use of DEM data, we are able to determine the area contributing runoff at each station. The data from the nine stream gauges were used to calibrate and verify continuous simulations from the CREST model across watersheds with drainage areas ranging from 191 to 25,000 km^2.

Zanon et al. (2010) has argued that the flood response could be reasonably well reproduced when using high resolution rainfall observations. Even though continuous rain gauge data are available for the area, there are several gaps in these data records. As Hirpa et al. (2010) reported unequal time period data between observation gauges might affect the certainty of parameter estimations. Additionally, Du et al. (2009) have stated that there are many stations in the USA that do not have extensive flow records or have limited flow data. The gauge data measurements are available in 5-min, 15-min, and 30-min intervals, with 15-min intervals being the most common. Thus, the finer time scale gauge measurements were converted into hourly streamflow averages in units of m^3/s from January 2002 to December 2009. The input forcing data consists of gridded radar-rainfall (mm/h) and Potential Evapotranspiration (PET) (mm/3h) data. The precipitation data was extracted from the WSR-88D Stage IV product obtained from the North American Regional Reanalysis. The Stage IV radar rainfall fields were used to force the model at the hourly time step and 4 km spatial grid resolution (Lin & Mitchell, 2005), while PET was used to force the model at 3-hourly time steps and 32 km grid resolution (Mesinger, et al., 2005). In addition to the radar rainfall and the PET data, High Resolution Digital Elevation Model

(DEM) data was used to delineate the various watersheds and to perform river routing of the streamflow simulations.

2.3.Methods of Analysis

2.3.1. Model

A raster-based distributed hydrologic model, Coupled Routing and Excess Storage (CREST) was implemented over Connecticut River Basin. The CREST model is a hybrid modeling strategy, developed by the University of Oklahoma (http://hydro.ou.edu) and NASA SERVIR-an acronym meaning to "to serve" in Spanish- Project Team (Wang, et al., 2011). The model is spatially distributed rainfall-runoff model dedicated to simulate flow discharges at regional and global scales aimed to represent hydrological processes associated with floods and droughts. Distributed CREST model can be applicable to almost any kind of hydrological problems and provides important advantage over existing models under different land cover and soil type scenarios with user defined spatiotemporal resolutions. The CREST model provides water managers with information about streamflow amounts that can enable better decision-making regarding water resources, floods, and agriculture.

CREST model uses a combination of DEM, PET, and precipitation input data with different user defined spatiotemporal resolutions. The processes modeled include rainfall-runoff generation and capacity for cell-to-cell routing, canopy interception, infiltration, evaporation, recharge baseflow, sub-grid cell variability of soil moisture storage capacity and routing processes at the sub-grid scale. The CREST model has been discussed extensively in previous papers (Wang, et al., 2011; Khan, et al., 2011) and details of the model can be found therein.

CREST model contains several default parameters. These parameters have value ranges initially specified based on land cover and soil type data. The definitions of parameters with their value ranges for CRB are listed in Table 2.1. The model contains 14 parameters, and three of them (related to initial soil conditions) can be adjusted using warm-up period. The initial estimates are fine-tuned through calibration procedures utilizing using observational data at gauging stations (Crawford & Linsley, 1966).

2.3.2. Model Calibration and Validation

CREST model calibration and validation were carried out over Connecticut River Basin at various selected observation stream gauges (Figure 2.2). Bergstrom (1991) defined model calibration as a process that model parameters are arranged to make model results to meet the measurements. However, if the number of parameters used in the calibration is large, automatic calibration is a better option to reduce labor-intensive. Additionally, automatic approach for model calibration abbreviates the time with the advantage of speed and power of high performance computers as well as the approach eliminates the kinds of subjective human judgments (Moriasi, et al., 2007; Boyle, et al., 2000). The auto-calibration routine based on Differential Evolution Adaptive Metropolis (DREAM) method was devised to calibrate CREST over a number of watershed areas in Connecticut River Basin based on hourly stream flow data for the period 2005 to 2007. The algorithm runs to mitigate sum of squared residuals to estimate posterior parameters using multiple Markov Chain Monte Carlo chains from prior iterations (Vrugt, et al., 2009).

The numbers of parameters which were subject to calibration were limited to 14 as defined by Wang et al. (Wang, et al., 2011). Calibrated values are listed in Table 2.2 along with the

acceptable intervals (Table 2.1) used to constrain the parameters search. Each basin was calibrated and validated separately to capture the spatial variability of the parameters over the region. Even though all parameters are in the acceptable range, the parameter values vary across basins. For example, the slope flow speed multiplier “coem” varied between 10.39 and 74.33 with a mean of 34.93 and 0.61 coefficient of variation. The maximum soil water capacity “pwm” parameter has the maximum coefficient of variation, i.e 1.15.

The simulation starts with a warm-up period to reduce the initial condition effect. Specifically, the initial value of soil water "iwu", initial value of overland reservoir "iso", and initial value of interflow reservoir "isu" are model states that were adjusted automatically by running CREST with a warm-up period. Marshall et al. (2008) observed in their study using a 40-year study period over Connecticut River Basin that the maximum snow accumulation occurs in the month of January and decreases exponential through April. To avoid snowmelt effects in parameter calibration the period of January through the middle of March was excluded from the error analysis.

Finally, after parameter calibration, CREST was validated for a four year period. Refsgard (1997) defined model validation as a process of illustrating that the calibrated parameter set without further adjustments is capable to reproduce flows in different times other than the calibration period. This process was done again without any interruption during snow accumulation and melting span, but we excluded the snow process period (January 1-March 15) from the analysis.

2.3.3. Evaluation Indexes

Different error metrics have been used in hydrologic model validation exercises of past studies (Moriasi, et al., 2007; Boyle, et al., 2000; Legates & McCabe, 1999). Multi criteria error functions aid to explain various aspects of hydrographs. In this study, the model efficiency to predict streamflow at basin outlets was demonstrated qualitatively by plotting time-series of observed and simulated streamflows and determining error statistics based on hydrographs normalized by the corresponding catchment area. The error metrics are listed below:

Nash-Sutcliffe Coefficients of Efficiency (NSCE):

$$\mathrm{NSCE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{o,i} - \overline{Q_o}\right)^2} \qquad (2.1)$$

where $Q_{o,i}$ is measured streamflow of the time i^{th}; $Q_{s,i}$ is simulated streamflow of the time i^{th}; $\overline{Q_o}$ is the average of entire observed streamflow values; and n is the total number of time steps. The value of NSCE varies from -∞ to 1. On the one hand, a value of NSCE ≤ 0 represents that the model does not have capability to use the observed mean as a predictor; on the other hand, NSCE=1 indicates that simulation results are capturing the measurements perfectly.

Mean Relative Error (MRE):

$$\mathrm{MRE} = \frac{\sum_{i=1}^{n}\left(Q_{s,i} - Q_{o,i}\right)}{\sum_{i=1}^{n} Q_{o,i}} \qquad (2.2)$$

MRE gives an indication of how close predictions are relative to the observations. A value of MRE=0 shows that the simulated total amount of discharges is commensurate to measurements.

Root Mean Square Error (RMSE):

$$\text{RMSE}\ (\%) = \frac{100}{\overline{Q_o}} * \sqrt{\frac{\sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)^2}{n}} \qquad (2.3)$$

RMSE measures the magnitude of the differences between simulated and observed discharges relative to mean observed discharge value. Therefore, a low RMSE indicates better fit and the value of zero signifies the perfect fit.

Pearson Correlation Coefficient (PCC):

$$\text{PCC} = \frac{\sum_{i=1}^{n}\left(Q_{o,i} - \overline{Q_o}\right)\left(Q_{s,i} - \overline{Q_s}\right)}{\sqrt{\sum_{i=1}^{n}\left(Q_{o,i} - \overline{Q_o}\right)^2 * \sum_{i=1}^{n}\left(Q_{s,i} - \overline{Q_s}\right)^2}} \qquad (2.4)$$

where $\overline{Q_s}$ is the average of entire simulated streamflow values. PCC represents how well the linear relationship between measurements and predictions. This value ranges from -1 to 1.If PCC=0, then there is no relation between the two variables. The closer either -1 or 1 indicates stronger correlation between them.

Boyle et al. (2000) have shown that RMSE is sensitive to the peak flows and can strongly bias recession error characteristics. PCC, on the other hand, reflects the collinearity between simulations and observations and correlation-based measures have excessive sensitivity to peak flows (Legates & McCabe, 1999). Moreover, MRE measures the average tendency of simulated against observed data and explains this tendency with overestimations or underestimations (Moriasi, et al., 2007). Sevat & Dezetter (1991) pointed out that even though NSCE error metric shows some weakness with low flows, it is the best objective function to provide extensive information on hydrograph prediction accuracy.

2.4.Results

CREST model was applied for the period between 2002 and 2009 to represent the hydrologic response of the Connecticut River Basin and several of its sub-basins. The calibration of the model was carried out over nine sub-basins from 2005 to 2007. Then, CREST stream flow simulations were compared to the measured flows at the outlets of the nine sub-basins and interior locations of three of the catchments over the seven-year time period using 2002 as the model spin-up. Simulations for the various basin sizes are summarized below showing an overall good agreed with observations for different event magnitudes.

In Figure 2.3, observed and simulated normalized hourly discharges were plotted along with measured rainfall for the period 2004 and 2005 at multiple stations. For brevity, three of the spatially various CRB sub-basins associated with different sizes are illustrated. Results, in the figure, from first and second columns belong to validation and calibration periods, respectively. Three different size basins (small, middle, and large scale catchments), namely B0128500 (673 km^2), B01122500 (1,046 km^2), and B01184000 (25,019 km^2) catchments, are visualized from top to bottom. As it can be seen from the hydrographs, overall, model performed well over CRB and the variability depicted in the flow simulation is close agreement with observed flows. Even though the timing of the peaks is estimated well, CREST simulations tend to slightly underestimate some of the peak flows. As it can be seen from the hydrographs in April 2005 simulations underestimated observations associated with small amounts of rainfall. However, it would be expected that light rainfall produce smaller runoff and this mismatch in terms of peak discharges can be explained by uncertainties in either rainfall or streamflow observational data. However, although slight underestimation is captured in qualitative stream flow comparisons, the majority of the simulated discharges were attained close to the measurements.

The model outputs from the hydrological simulations are summarized in Table 2.3 (A) and (B) for calibration and validation periods, respectively. Basin areas, observed and simulated mean discharges, error metrics results (NSCE, MRE, RMSE, and PCC), and percentage of unavailable observations during the analysis are illustrated. During calibration period, as it can be seen from Table 2.3, NSCE values vary between 0.31 and 0.68 for the nine-basins while MRE results range from -6% to 13%. Additionally, 0.60 is the smallest value with hourly time resolution for PCC metric. The demonstrated NSCE, MRE, and PCC values show that the model simulations have good agreement with measurements in the calibration data period. In terms of RMSE error metric, however, at some stations, it is difficult to appreciate the quality of the results.

In the validation period, on the other hand, Nash is lower (ranging between 0.12 and 0.58) for hourly resolution, while PCC values ranged between 0.42 and 0.77 for the nine-basins. It is interesting enough, during validation, overall MRE exhibited underestimation in the range of 4% to 26%. However RMSE values dropped between 2 and 20% in the basins, which indicate improved random error.

Moriasi et al. (2007) have shown that hydrologic model performance is better for longer time steps (i.e. annual versus monthly) and claimed that simulation statistics improve as a function of time resolution. Fernandez et al. (2005) supported this statement and reported that NSCE values during calibration in their study in the range of 0.36 and 0.66 for daily and monthly time resolutions, respectively. Hence, daily error metrics were calculated and illustrated in Table 2.3 (C).

The study illustrates a better agreement between estimated and measured flows at daily scale and model results become quite realistic. For instance, in B01205500, NSCE values increase from 0.31 to 0.47 during calibration and from 0.16 to 0.30 in validation period for hourly and daily resolutions, respectively. While PCC values improve from 0.60 to 0.70 (calibration period) and 0.46 to 0.59 (validation period) as a function of time resolution in the same catchment, no change is observed in terms of the MRE values, which is expected given that resolution primarily affects the random component of error. Similar improvements are also captured for the other basins (Table 2.3.C). Moreover, RMSE values drop significantly and become less than 100% during both validation and calibration in the basins with the exception of B01193500 (103.91 %).

Analysis of the error metrics with their annual fluctuations were, additionally, examined in Figure 2.4. The figure provides an illustration of simulations versus observation streamflow variability. Results reported in Figure 2.4 A, B, C and D show the hourly calculated NSCE, MRE, RMSE, and PCC statistics, respectively for each year in the 7-year time period. Multiple transparent dots represent the basin statistic and darkness of the dots is increased as a function of watershed size. Analysis of the results acquired for 7-year gives more reasonable amount for smaller basins. The majority of NSCE, MRE, PCC values ranged in acceptable levels during both calibration and validation. As expected, NSCE and PCC values for calibration are better than the validation periods and increase with longer time steps. Moreover, model outputs are quite realistic in between 2004 to 2008, while results diverged in the year 2003 and 2009. Thus, that can yield a bias into the optimization process with low NSCE, RMSE, and PCC.

Three nested subbasins were considered as ungauged interior catchments namely B01122500 (Figure 2.5), B0112700 (Figure 2.6), and B01205500 (Figure 2.7). After model was calibrated for the outlet of the parent basin, the parameters were fixed, and then CREST was used to

simulate the interior catchments. Figure 2.5, Figure 2.6, and Figure 2.7 show the comparison between model outputs and measurements via statistical error functions with annual variability. The catchments were ordered with increasing area. Differences in model responses increase with increasing size of the catchment. Overall, results show that there is good agreement between the predictions and the measurements.

In Figure 2.5, statistical metrics of NSCE, MRE, RMSE, and PCC presented for different years and different interior subbasins of parent basin B01122500. The results show that the highest NSCE values are obtained in the larger basin (B01122500) while the small basin (B0119500) exhibits the poorest agreement in terms of the error metric. Basin size dependency is also observed in RMSE metric values with the highest value reported for the smaller basins. MRE results, on the other hand, are within +/- 0.5, and the overall performance of the model in terms of PCC is above 0.5.

To test calibration, validation periods, and resolution effect on the error metrics, Table 2.4 statistics were calculated and showed the results as a summary. In Table 2.4, basin statistics are visualized at hourly and daily resolution during calibration and validation periods for ungauged interior subbasins of parent B01122500 basin. As it can be seen from table, number of time steps in the observed data is decreased as a function of catchment area. In addition, it is important to note that discharge point of small basin is at the furthest location from the calibrated downstream outlet. It is expected that the simulation performance is decreased as a function of catchment area. Additionally, slight improvements are observed changing time resolution from hourly to daily.

Table 2.5 shows calculated basin statistics for interior subbasin of parent B01127000 at hourly time resolution during calibration period in section (A) and during validation period in section (B). Additionally, their daily statistics are reported in part (C). It is observed that, overall NSCE results, in B01124000, are close to each other during calibration and validation periods. MRE value reduces from 0.41 to 0.23 during calibration and validation, respectively. While RMSE drops 20% from calibration to validation period, slight change is obtained in PCC with the value being around 0.7. Slight changes are also obtained as function of the time resolution for all the metrics in calibration an validation periods.

On the other hand, in Figure 2.6, variability is noted with respect to year for the same parent basin (B01127000) and nested catchment. The same trend is seen in the error metrics results of two basins spanning a 7-yr period with higher accuracy for the large basin (B01127000). This suggests that, overall the model provides a reasonably good description of the fluctuation.

Finally, calibrated largest parent basin (B01205500) is examined with interior nested basins in Figure 2.7. The smallest basin (01199000) tends to overestimate in terms of MRE results in addition to negative bias for the other two basins. NSCE results are below 0.5 while PCC values are above 0.3 in the 7-yr period. RMSE values improve as function of basin size.

The model efficiency is summarized in Table 2.6 during validation (section A) and calibration (section B) periods (hourly) as well as daily resolution (section C) for B01199000, B01200500 (considered as internal ungauged basins) and B01205500 (parent basin). MRE results show that the values are in the range of -0.02 and 0.21 for calibration and -0.15 and 0.07 for validation. Correlation is around 0.6 during calibration while it drops to 0.46 for the parent basin and improves to 0.65 for the medium basin (B01200500) in validation period. Specifically

for the daily simulations, all RMSE values are below 100%. Examination of these statistics reported in the table shows that observations of the interior basins are well represented with CREST simulations.

Figure 2.8 and Figure 2.9 illustrate the comparison of the simulations in terms of quantiles at the basins outlets and interior ungauged locations, respectively. MRE and RMSE error metrics were calculated at 0.2, 05, 07, 0.9 and 0.95 quantiles and shown with each spatially distributed station with transparency dots. Reduction of the RMSE and increment of the MRE tendency are observed from low to high flows. The poorer performances of the model in terms of MRE and RMSE are obtained at the higher quantiles and considered to be related to inadequate representation of the peak events. Figures confirm the tendency of underestimations at peak discharges and this consistency supports the argument of using frequency analysis with large flow data to improve model results.

2.5. Conclusions

This case study has demonstrated the implementation of the CREST model, and its calibration and validation processes at high temporal resolution with spatial distributed gauged observations as well as ungauged/poorly gauged catchments over Connecticut River Basin. The primary objective of this study is calibration and validation processes of distributed hydrological model based on streamflow observations and provide explicitly statistical analysis of the simulations via different objective functions. Another objective of this study is producing flow data for insufficient historical records at gauging stations using calibrated CREST model and predicting flows over ungauged locations of the Connecticut River Basin. In this manner, the basin was partition into 9 subbasins to represent distributed parameters sets via CREST model

based on the USGS monitoring stations and their locations, and the model was evaluated with several important statistics with supplemental hydrographs.

CREST model was calibrated for three years and validated in four years during the period 2002-2009. Calibration of the model was done over the period 2005 to 2007 and validations were carried out over two distinct periods: 2003 to 2004 and 2008 to 2009.The model was calibrated against 14 parameters within specified limited boundaries while a few of these parameters were adjusted using 2002 as spin up period for plausible initial conditions so that model predicts rainfall-runoff responses closely to the flow measurements and reflects the behavior of watershed system in a various manner. The figures and tables demonstrated the model's accuracy for the simulation of the rainfall-runoff transformation at 9 stations and ungauged interior locations.

In this study, several statistical indicators and supplemental graphical illustrations were applied for evaluation of the CREST model performance. Using multiple statistics helps to cover a different aspect of the hydrographs. Based on the error metrics and graphical comparisons, CREST is considered to provide accurate rainfall-runoff simulations during calibration and validation period. The results also indicate satisfactory in model performance over Connecticut River Basin.

The CREST hydrological model estimates the hourly flow at spatially various catchments well, but with relatively large errors at peak quantiles. Overall, when we take into account statistical metrics and normalized hydrographs of the model results, it can be concluded that the CREST is capable of reproducing continuous hourly streamflows in Connecticut River Basin, which allows model use for flood management applications in both gauged and ungauged basins. Results with CREST illustrated that model performance was satisfactory in representing

amplitude and timing of the flow peaks, and that the models performs better in predicting the entire hydrograph. The study also confirms the need for further work on frequency analysis because none of the parameter sets were able to reproduce the exact peak discharges in any of the catchments. The model application and its calibrated parameters can be used for future work such as comparing to satellites-driven flow, flash flood predictions, flood frequency, and future climatic analysis on local and national scale as well.

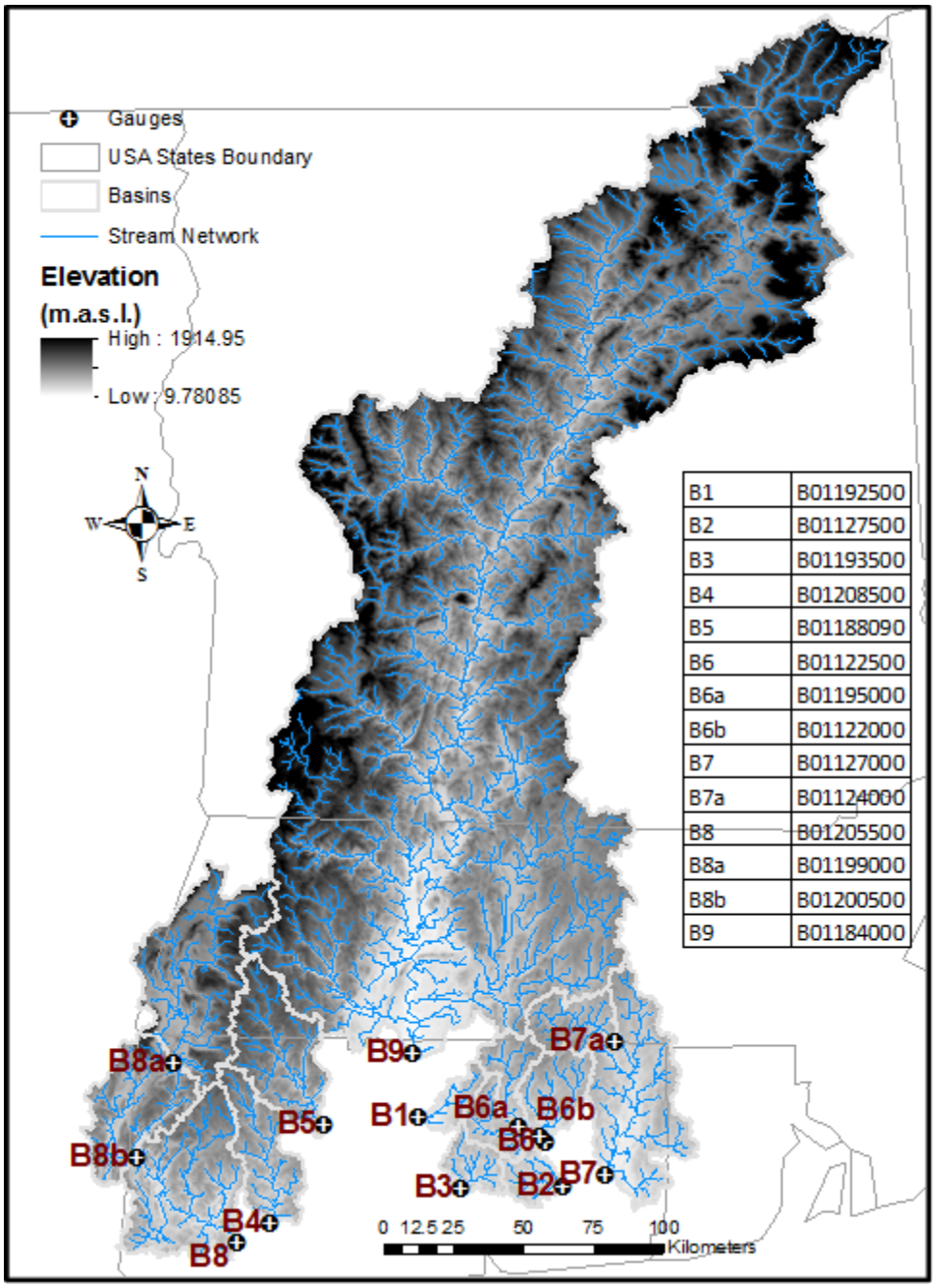


Figure 2.1. Map of Connecticut River Basin showing the drainage areas of the subbasins used in this study. The black dots represent streamflow gauges labeled with USGS station numbers.

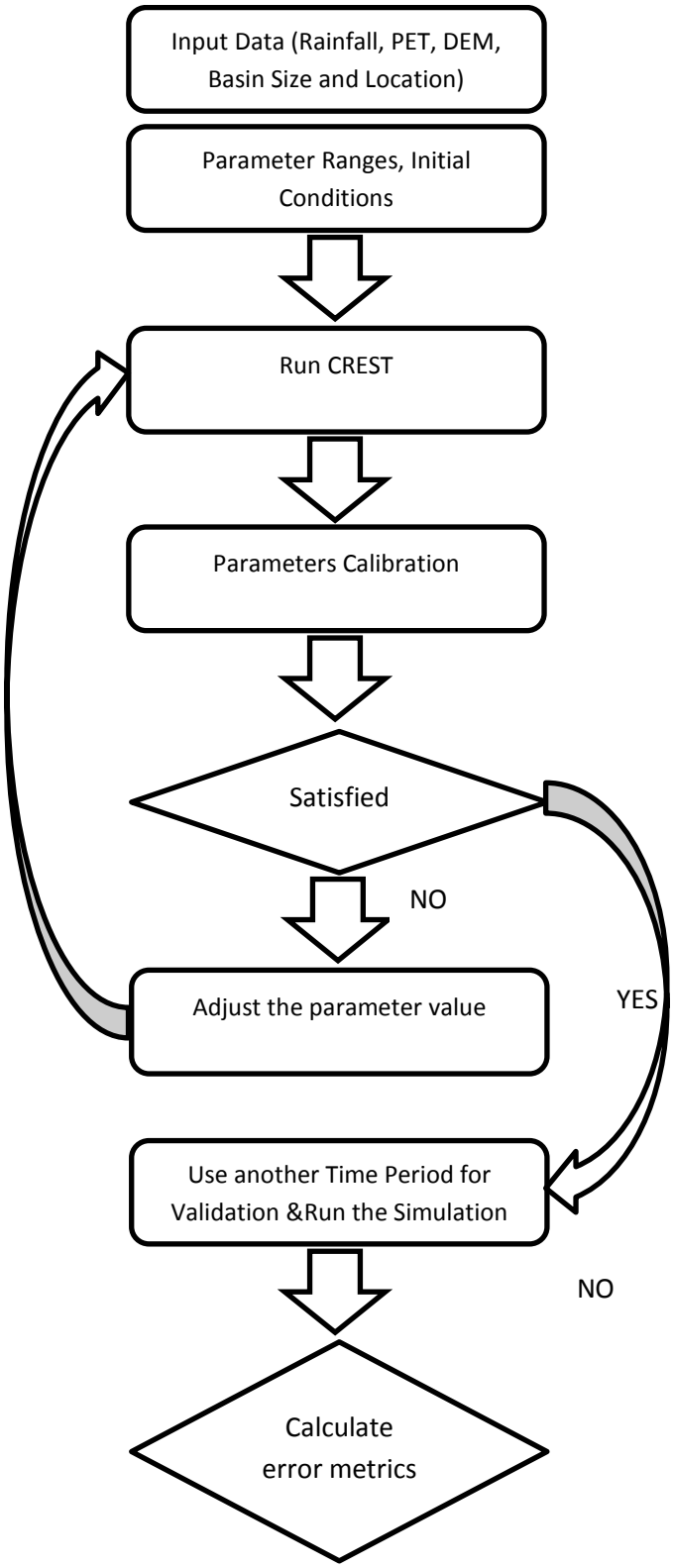


Figure 2.2. Calibration-Validation procedure

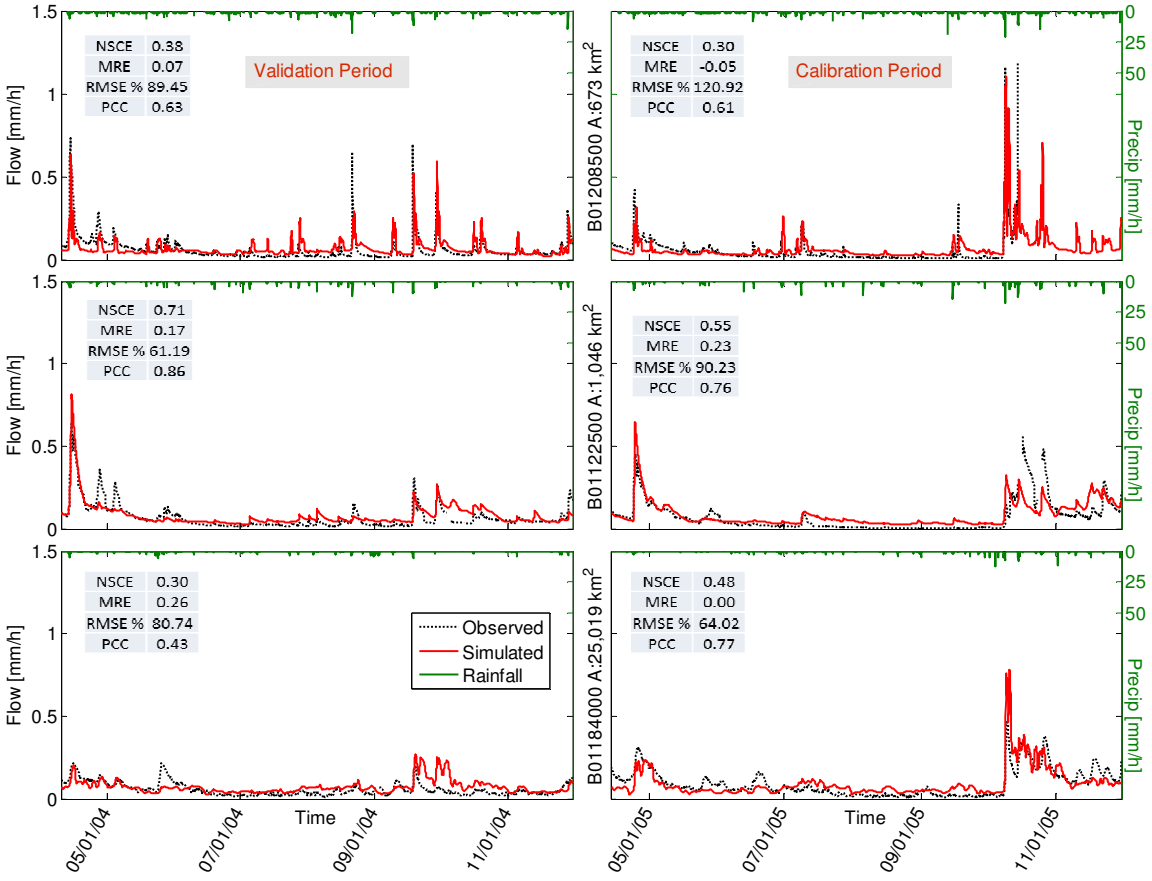


Figure 2.3. Normalized hydrographs of the model results and observations streamflow along with precipitation; partially calibration and validation period is zoomed in.

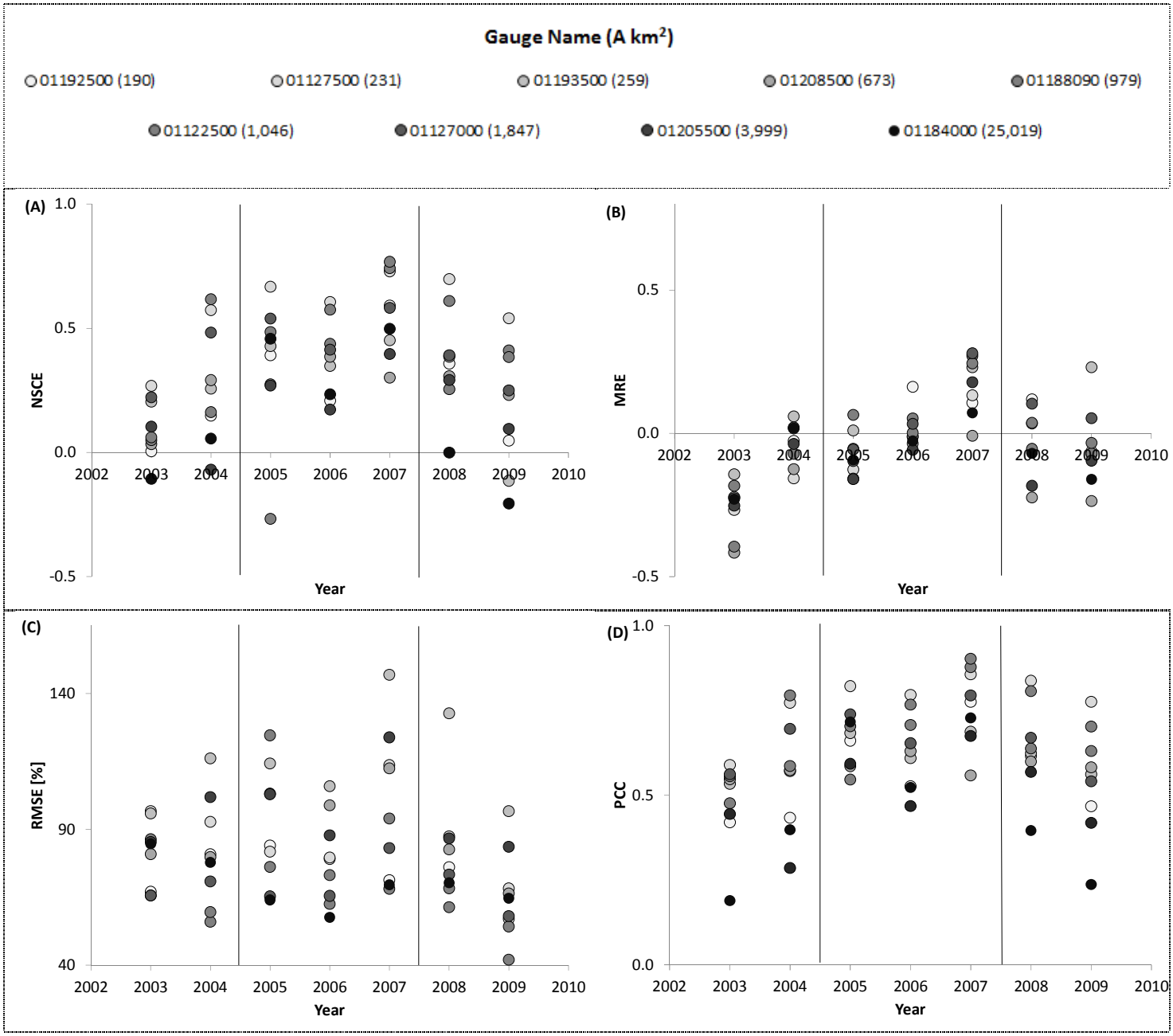


Figure 2.4. Statistical metrics of (A) NSCE, (B) MRE, (C) RMSE, and (D) PCC presented for different years and basin scales.

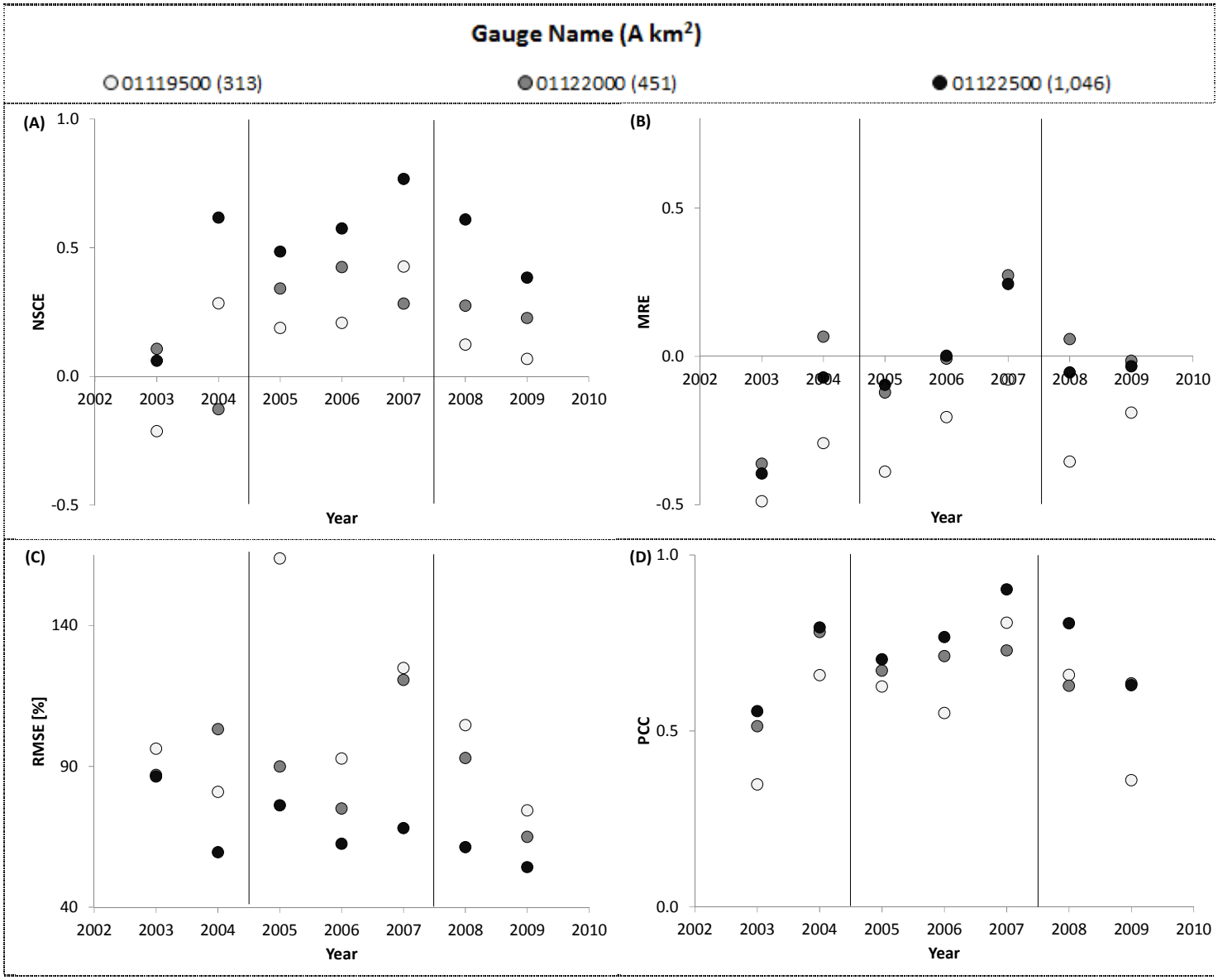


Figure 2.5. Statistical metrics of (A) NSCE, (B) MRE, (C) RMSE, (D) PCC presented for different years and different interior subbasins of parent basin 01122500.

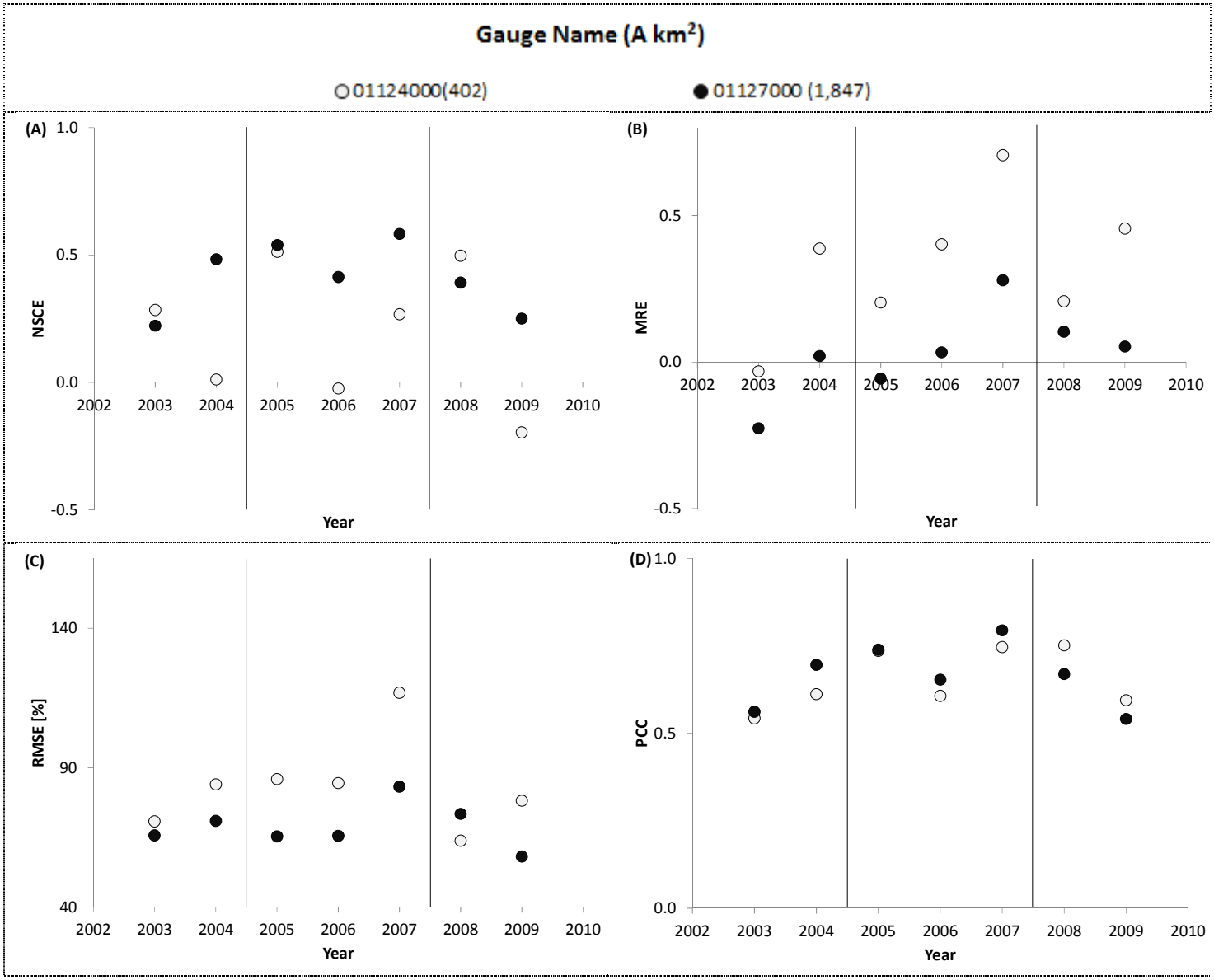


Figure 2.6. Same as in Figure 2.5 but for parent basin 01127000.

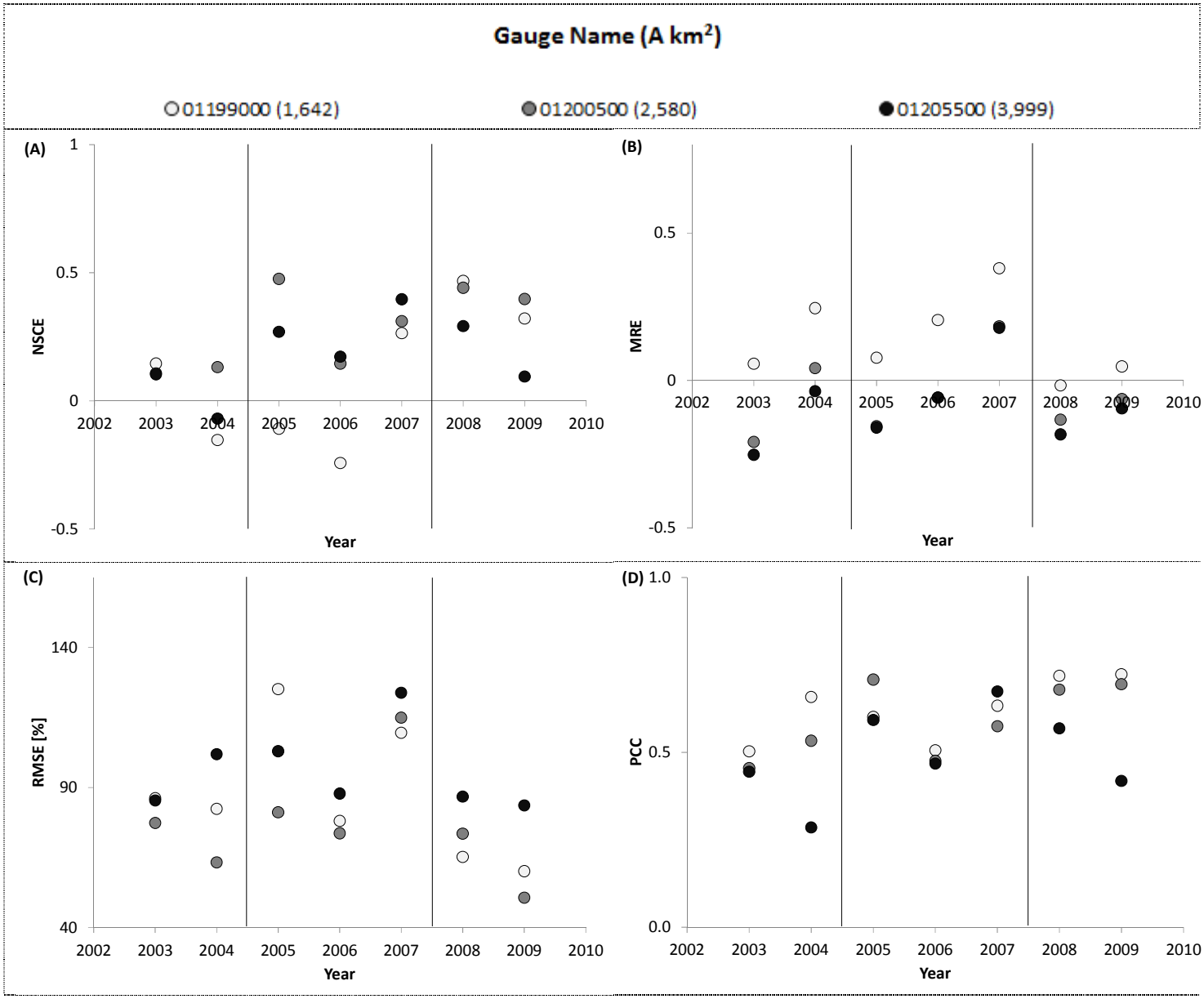


Figure 2.7. Same as in Figure 2.5 but for parent basin 01205500.

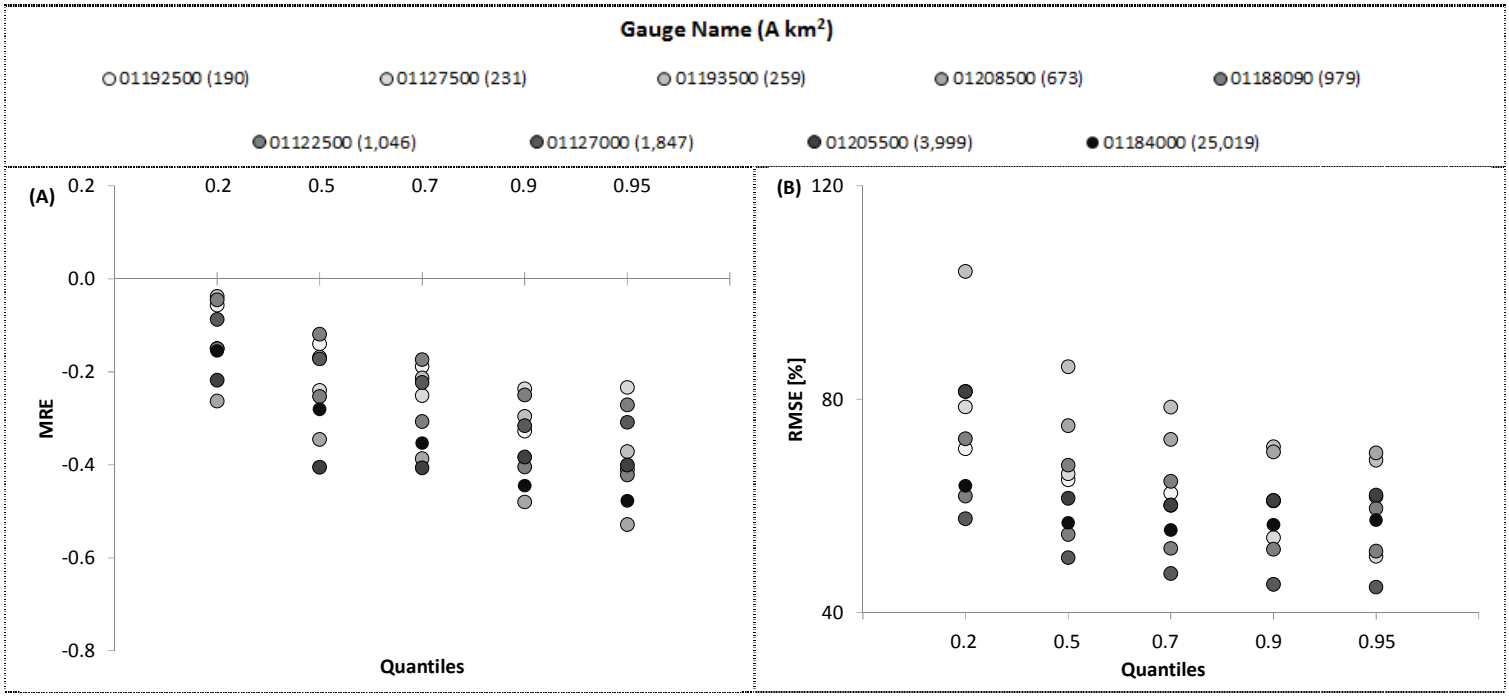


Figure 2.8. A) MRE and B) RMSE error metrics determined for different flow quantiles at the gauged subbasin outlets.

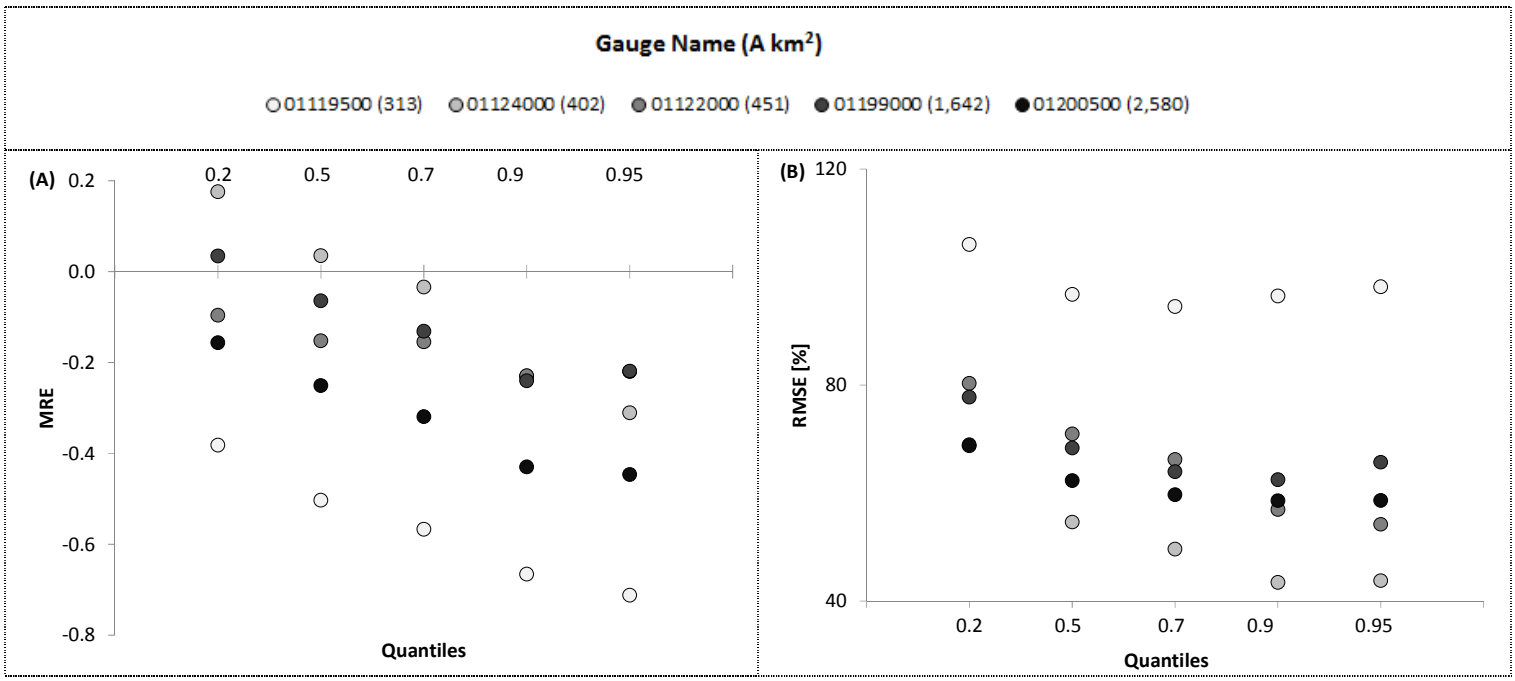


Figure 2.9. Same as in Figure 2.8 but for the ungauged interior subbasins.

Table 2.1. CREST model parameters with their value ranges

Parameter Name	Description	Suggested Range
coem	The slope flow speed multiplier	3.33, 100.0
river	The channel flow speed multiplier	3.33,100.0
under	The horizontal velocity under the ground (Generally hydraulic conductivity used for this velocity)	0.01,51.0
leako	The overland reservoir discharge multiplier	0.01,1.0
leaki	The interflow reservoir discharge multiplier	0.01,1.0
th	The flow accumulation needed for a cell to be marked as a channel cell. If a cells flow accumulation is greater than th then the cells slope flow speed is multiplied by river	0.01,30.0
pwm	The maximum soil water capacity (depth integrated pore space) of the soil layer	0.01,250.0
pb	The exponent of the variable infiltration curve.	0.01,4.0
pim	The impervious area ratio	0.00,100.0
pke	The multiplier to convert between input PET and local actual ET	0.01,1.0
pfc	The soil saturated hydraulic conductivity	0.01,51.0
iwu	The initial value of soil water. (This is a percentage of the pwm)	0.0,100.0
iso	The initial value of overland reservoir	0.01,100.0
isu	The initial value of interflow reservoir	0.01,100.0

Table 2.2. Parameters values determined from calibration for each gauged subbasin.

Parameter Name	01192500	01127500	01193500	01208500	01188090	01122500	01127000	01205500	01184000
coem	74.33	43.52	53.54	50.06	14.55	12.30	18.07	10.39	37.63
river	99.19	80.89	99.06	86.67	79.61	98.53	80.02	90.45	7.16
under	33.69	7.25	38.96	0.90	50.43	4.16	35.30	13.64	17.79
leako	0.02	0.36	0.74	0.40	0.08	0.02	0.10	0.23	0.10
leaki	0.94	0.72	0.84	0.60	0.73	0.53	0.52	0.42	0.23
th	24.76	16.81	15.29	26.82	21.85	13.32	11.62	16.72	13.69
pwm	32.89	3.03	7.59	1.51	2.52	1.91	3.02	7.56	14.30
pb	2.06	0.49	2.45	2.46	0.79	1.31	2.31	1.79	0.38
pim	1.42	0.93	0.99	0.04	0.85	1.08	1.50	1.05	0.69
pke	0.44	0.95	0.99	0.84	0.79	0.17	0.96	0.84	0.72
pfc	44.88	3.48	38.37	50.49	45.68	37.24	35.36	39.72	46.45
iwu	71.33	0.59	91.16	75.92	7.86	91.28	70.48	42.60	89.78
iso	9.17	15.14	2.11	4.77	92.52	40.60	59.05	65.20	80.21
isu	1.45	51.77	12.90	81.45	13.06	48.05	98.21	26.71	85.11

Table 2.3. Gauged basin statistics determined at hourly time resolution during calibration period in (A) and during validation period in (B). Additionally, their daily statistics are reported in (C).

(A) CALIBRATION

Gauge Name	Basin Area [km^2]	Q_O[m^3/s] μ	Q_O[m^3/s] σ	Q_S [m^3/s] μ	Q_S [m^3/s] σ	NaN Data [%]	NSCE	MRE	RMSE [%]	PCC
01192500	190.11	4.38	4.74	4.85	3.89	0.11	0.41	0.11	83.28	0.67
01127500	231.29	5.31	8.55	5.42	7.64	5.29	0.68	0.02	90.59	0.83
01193500	259.00	6.70	10.62	7.11	8.58	4.90	0.42	0.06	121.25	0.66
01208500	673.40	14.03	18.98	13.79	14.20	17.13	0.35	-0.02	109.36	0.61
01188090	979.02	22.32	29.14	25.28	28.85	3.92	0.40	0.13	101.32	0.70
01122500	1,046.36	23.09	26.85	24.75	18.98	1.77	0.64	0.07	70.03	0.81
01127000	1,846.66	41.74	44.58	45.85	34.66	0.10	0.54	0.10	72.34	0.74
01205500	3,998.94	90.83	114.55	89.73	94.74	1.08	0.31	-0.01	104.98	0.60
01184000	25,019.28	690.49	531.64	648.82	452.39	1.22	0.37	-0.06	61.18	0.65

(B) VALIDATION

Gauge Name	Basin Area [km^2]	Q_O[m^3/s] μ	Q_O[m^3/s] σ	Q_S [m^3/s] μ	Q_S [m^3/s] σ	NaN Data [%]	NSCE	MRE	RMSE [%]	PCC
01192500	190.11	4.20	3.35	4.00	2.40	1.79	0.21	-0.05	71.12	0.50
01127500	231.29	5.84	7.87	5.03	6.54	12.00	0.58	-0.14	87.64	0.77
01193500	259.00	6.63	9.08	6.86	7.10	13.21	0.29	0.04	115.00	0.58
01208500	673.40	17.88	16.33	13.15	9.76	4.24	0.24	-0.26	79.68	0.57
01188090	979.02	24.06	16.83	22.85	16.02	3.64	0.24	-0.05	61.18	0.60
01122500	1,046.36	26.05	23.91	22.31	15.62	5.35	0.44	-0.14	68.54	0.68
01127000	1,846.66	43.90	37.72	43.43	29.58	4.06	0.37	-0.01	68.19	0.63
01205500	3,998.94	97.80	95.21	82.89	57.22	0.02	0.16	-0.15	89.12	0.46
01184000	25,019.28	578.45	455.76	529.10	295.32	3.12	0.12	-0.09	74.09	0.42

(C) DAILY RESOLUTION

Gauge Name	Basin Area [km2]	CALIBRATION NSCE	CALIBRATION MRE	CALIBRATION RMSE	CALIBRATION PCC	VALIDATION NSCE	VALIDATION MRE	VALIDATION RMSE	VALIDATION PCC
01192500	190.11	0.44	0.11	76.75	0.69	0.21	-0.05	64.82	0.53
01127500	231.29	0.72	0.02	79.38	0.85	0.60	-0.14	79.17	0.78
01193500	259.00	0.50	0.06	103.91	0.72	0.35	0.04	97.85	0.62
01208500	673.40	0.47	-0.01	91.83	0.69	0.34	-0.26	69.81	0.67
01188090	979.02	0.44	0.13	93.16	0.73	0.25	-0.05	57.58	0.61
01122500	1,046.36	0.65	0.07	68.51	0.81	0.44	-0.14	67.07	0.69
01127000	1,846.66	0.57	0.10	68.68	0.76	0.41	-0.01	64.84	0.65
01205500	3,998.94	0.47	-0.01	83.77	0.70	0.30	-0.15	65.96	0.59
01184000	25,019.28	0.38	-0.06	59.89	0.65	0.12	-0.09	73.27	0.43

Table 2.4. Same as in Table 2.3, but for the ungauged interior subbasins of parent 01122500 basin.

(A) CALIBRATION

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	NSCE	MRE	RMSE [%]	PCC
01119500	313.39	7.21	12.08	5.60	3.25	10.88	0.26	-0.22	144.00	0.66
01122000	450.66	10.43	12.61	11.18	13.31	2.16	0.39	0.07	94.36	0.71
01122500	1,046.36	23.09	26.85	24.75	18.98	1.77	0.64	0.07	70.03	0.81

(B) VALIDATION

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	NSCE	MRE	RMSE [%]	PCC
01119500	313.39	7.69	7.71	5.03	2.11	16.02	0.09	-0.35	95.40	0.53
01122000	450.66	10.47	10.10	9.80	11.18	2.68	0.15	-0.06	89.14	0.62
01122500	1,046.36	26.05	23.91	22.31	15.62	5.35	0.44	-0.14	68.54	0.68

(C) DAILY RESOLUTION

Gauge Name	Basin Area [km2]	CALIBRATION NSCE	CALIBRATION MRE	CALIBRATION RMSE	CALIBRATION PCC	VALIDATION NSCE	VALIDATION MRE	VALIDATION RMSE	VALIDATION PCC
01119500	313.39	0.28	-0.22	135.32	0.67	0.09	-0.35	91.47	0.53
01122000	450.66	0.44	0.07	88.54	0.74	0.18	-0.06	84.68	0.64
01122500	1,046.36	0.65	0.07	68.51	0.81	0.44	-0.14	67.07	0.69

Table 2.5. Same as in Table 2.3, but for the ungauged interior subbasin of parent 01127000 basin.

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	(A) CALIBRATION NSCE	MRE	RMSE [%]	PCC
01124000	401.45	9.32	11.04	13.18	8.97	4.70	0.35	0.41	95.20	0.70
01127000	1,846.66	41.74	44.58	45.85	34.66	0.10	0.54	0.10	72.34	0.74

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	(B) VALIDATION NSCE	MRE	RMSE [%]	PCC
01124000	401.45	9.75	8.72	11.99	7.57	9.25	0.34	0.23	72.72	0.67
01127000	1,846.66	43.90	37.72	43.43	29.58	4.06	0.37	-0.01	68.19	0.63

Gauge Name	Basin Area [km2]	CALIBRATION NSCE	CALIBRATION MRE	CALIBRATION RMSE	CALIBRATION PCC	VALIDATION NSCE	VALIDATION MRE	(C) DAILY RESOLUTION RMSE	PCC
01124000	401.45	0.38	0.42	92.52	0.71	0.36	0.23	69.86	0.68
01127000	1,846.66	0.57	0.10	68.68	0.76	0.41	-0.01	64.84	0.65

Table 2.6. Same as in Table 2.3, but for the ungauged interior subbasins of parent 01205500 basin.

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	(A) CALIBRATION NSCE	MRE	RMSE [%]	PCC
01199000	1,642.05	35.87	39.35	43.37	44.76	6.37	0.00	0.21	109.91	0.58
01200500	2,579.63	57.99	65.09	56.90	42.27	2.36	0.36	-0.02	89.78	0.60
01205500	3,998.94	90.83	114.55	89.73	94.74	1.08	0.31	-0.01	104.98	0.60

Gauge Name	Basin Area [km2]	QO[m3/s] μ	QO[m3/s] σ	QS [m3/s] μ	QS [m3/s] σ	NaN Data [%]	(B) VALIDATION NSCE	MRE	RMSE [%]	PCC
01199000	1,642.05	40.75	35.37	43.46	35.54	2.81	0.30	0.07	72.67	0.65
01200500	2,579.63	65.17	54.80	58.52	41.43	8.77	0.32	-0.10	69.13	0.60
01205500	3,998.94	97.80	95.21	82.89	57.22	0.02	0.16	-0.15	89.12	0.46

Gauge Name	Basin Area [km2]	CALIBRATION NSCE	CALIBRATION MRE	CALIBRATION RMSE	CALIBRATION PCC	VALIDATION NSCE	VALIDATION MRE	(C) DAILY RESOLUTION RMSE	PCC
01199000	1,642.05	0.17	0.21	99.55	0.62	0.36	0.07	68.88	0.67
01200500	2,579.63	0.39	-0.02	86.88	0.63	0.37	-0.10	66.33	0.63
01205500	3,998.94	0.47	-0.01	83.77	0.70	0.30	-0.15	65.96	0.59

3. Evaluation of Satellite Precipitation Products in Hydrologic Simulations of a Northern Latitude River Basin

3.1.Introduction

Precipitation is arguably the most important parameters controlling the hydrologic cycle over land. Therefore, flood forecasting, water management and other hydrological applications rely principally on the accurate quantification of precipitation at different space-time scales. Given the limited in situ precipitation measurements at global scale, satellite remote sensing is becoming a viable data source for monitoring spatiotemporal variations of precipitation over remote and ungauged areas, but uncertainty in these estimates hampers the hydrologic potential of the satellite products (Seyyedi, et al., 2015; Behrangi, et al., 2014; Nikolopoulos, et al., 2013; Maggioni, et al., 2013; Bitew & Gebremichael, 2011; Shrestha, et al., 2008). Several studies, though, have argued about the potential of using these products over data poor regions for studying flood hydrology and water resources (Mei, et al., 2014; Gao & Liu, 2013; Yong, et al., 2012; Beighley, et al., 2011; Tang, et al., 2010; Pan, et al., 2010; Hirpa, et al., 2010; Artan, et al., 2007).

As a result several past investigations have focused on the understanding of satellite precipitation uncertainty and the associated errors in flow simulations over different regions of the globe. According to Yong et al. (2012), the Tropical Rainfall Measuring Mission (TRMM) based product has a good potential to simulate basin streamflows. In their study over China, higher correlation and lower false alarm rates for the TRMM product were observed in the summer months, while the performance dropped in winter. In addition, the post-real-time satellite products have been shown to provide more accurate estimates of precipitation compared

to the near-real-time products due to bias corrections applied based on monthly rain gauge accumulations (Yong, et al., 2012; Gourley, et al., 2011; Bitew & Gebremichael, 2011; Hirpa, et al., 2010). According to Pan et al. (2010) gauge adjusted satellite products exhibit significant improvements in skill of estimating the streamflow.

According to Aghakouchak et al. (2011), however, satellite-based estimates are not fully capable to capture extreme rainfall rates, which are a finding also supported by Gao & Liu (2013) and Mei, et al. (2014) who have shown that satellite products tend to overestimate light rainfall and significantly underestimate heavy rain rates. Hirpa et al. (2010) compared the performance of three high-resolution satellite products (TRMM-3B42RT, CMORPH, and PERSIANN) in capturing ground reference rainfall measurements at different elevations in the Ethiopian highland. They concluded that TRMM-3B42RT and CMORPH follow similar trend in elevation dependency while PERSIANN considerably underestimates rainfall in high-elevation areas.

Nikolopoulos et al. (2013) forced a hydrologic model with three satellite products (TRMM-3B42V6, CMORPH, and PERSIANN-CCS) for a single case study over a mountainous basin in northeastern Italian Alps. They found that the satellite-driven streamflow simulations are not able to explain the basin's hydrologic response during the flash flood event. They showed significant bias in terms of both volume and peak runoff. Similarly, Bitew & Gebremichael (2011) claimed that even though satellite precipitation driven runoff simulations capture shape of the streamflow observations, they underestimate the large flood peaks. They also found that the larger watershed gives better simulation performance in terms of dampening the random error in precipitation estimates due to additional averaging process in larger area. Moreover, Khan et al. (2011) implemented the Coupled Routing and Excess Storage (CREST) model with TRMM-

3B42RT rainfall product for flood inundation mapping in Lake Victoria basin in Africa. They exhibited an adequate skill in the satellite driven simulations for predicting spatial flood extends. They suggested the use of ground-based observations and multiple remote sensing products to help for further improvements in flood monitoring.

Shrestha et al. (2008) applied a semi-distributed hydrologic model to obtain flow simulations with daily satellite rainfall estimates over a catchment located in Nepal. They found significant underestimation in the simulated discharges from the coarse resolution data and emphasized the necessity of high resolution precipitation data for long-term hydrologic simulations. Gourley et al. (2011), who investigated satellite rainfall error propagation in hydrologic simulations of a small basin (342 km^2) in Oklahoma, highlighted the importance of high resolution rainfall for accurate distributed hydrologic modeling applications. They also concluded that although coarser in resolution TRMM-3B42RT (0.25°) simulations were better than the higher resolution PERSIANN-CCS-driven (0.04°) simulations.

This study continues the line of research on error propagation of satellite precipitation in flow simulations focusing on a mid-latitude hydro-climatic regime and a range of basin scales (191 – 25,000 km^2) constituting the Connecticut River Basin (CRB). The study is based on a distributed hydrologic model that was calibrated based on reference rainfall (i.e. gauge-adjusted radar-rainfall) and observed streamflow time series. The hydrologic model was then forced with the different satellite precipitation products and the reference rainfall to simulate streamflows (radar-driven hydrologic simulations are used as reference runoff). Rainfall and runoff simulations aggregated at two temporal resolutions (3-hourly and daily) are examined for a period of nine years (2002-2010) and based on two seasons (May-Oct and Nov-April) representing different precipitation systems. Additionally, error metrics are evaluated separately for low (20^{th}) and high

(90^{th}) quantile thresholds of precipitation and associated flows. This paper is organized as follows: Section 3.2 provides a description of datasets, study area and methodology; section 3.3 presents the comparison between satellite products and their resulting flow simulations and describes the results while conclusions are given in section 3.4.

3.2.Data and Methods

3.2.1. Study Area

The study area of this work is the Connecticut River Basin (CRB) in the northeast United States (Figure 3.1). The Connecticut River flows through New Hampshire (NH), Vermont (VT), Massachusetts (MA), and Connecticut (CT). The basin has an elevation gradient decreasing from north to southeast and the elevation varies from 10 to 1915 Mean Sea Level (M.S.L). The locations of USGS monitoring gauging stations used for the hydrologic model calibration are also presented in the figure. Blue lines and light gray rectangular in the figure represent the river network and 0.25^{o} satellite grids, respectively.

In this study, we considered nine basins with areas ranging from 191 to 25,019 km^2. The basins are classified into five scale ranges according to basin area. To evaluate error metric dependency on basin scales, we considered basins with drainage areas less than 300 km^2 as the smallest scale basin and denoted by B1 (B1a, B1b and B1c with areas of 191 km^2, 231 km^2 and 259 km^2). B2 group is located in the southwest of CRB and the basin areas are between 600 and 1000 km^2 (B2a at 673 km^2 and B2b at 979 km^2). B3a (1,046 km^2) and B3b (1,847 km^2) are grouped as B3 representing mid-size basins. Finally, B4a (4,000 km^2) and B5a (25,000 km^2) are classified as B4 and B5, representing separate basin scales.

3.2.2. Data

The TRMM-3B42V7, CMORPH, and PERSIANN over CRB spanning a record of 9-years (2002 to 2010) are used as a representative example of global-scale satellite precipitation products. The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis product 3B42 version 7 (here after named TRMM-3B42V7) is a gauge adjusted product available with one month latency. It uses microwave (MW)-calibrated infrared (IR) precipitation estimates to fill in MW coverage gaps and monthly gauge precipitation data to adjust the merged MW/IR rainfall estimates (Huffman, et al., 2007). Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN) uses a neural network approach to derive relationships between IR and various MW collections, and then estimations are generated directly from IR data (Sorooshian, et al., 2000). Climate Prediction Center Morphing Technique (CMORPH) uses precipitation estimates from MW data and produces spatiotemporal precipitation fields by interpolating between MW overpasses using cloud motion fields derived from IR observations (Joyce, et al., 2004). These aforementioned three satellite precipitation products are available at a 0.25^{o} and 3-hourly spatiotemporal resolution.

Gridded radar-rainfall data (~4km grid resolution and hourly temporal resolution) from the National Center for Environmental Prediction WSR-88D Stage IV product (Lin & Mitchell, 2005) was applied to represent the reference rainfall fields. The data represent a National mosaic of precipitation estimates 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 for each radar is merged with other radars in the WSR-88D network using an inverse-distance weighting (IDW) scheme. The stage IV precipitation product benefits from the 12 Continental United States River Forecast

Centers (RFCs) manual quality-control (QC), beside the various bias adjustments applied on the basis of the hourly gauge measurements (Lin & Mitchell, 2005). Potential evapotranspiration data were obtained at 3-hourly and 32 km resolution from the North American Regional Reanalysis (Mesinger, et al., 2005).

Instantaneous records of river discharge were obtained from measurements of USGS gauging stations in the CRB for the period of 2002 to 2010. The data records are available in 5-min, 15-min, and 30-min intervals with 15-min intervals being the most common. Any measurement within a specific hourly interval was averaged to represent hourly streamflows.

Basin-average satellite and Stage IV radar precipitation time series were determined for each basin of the study area; the Stage IV radar precipitation data are considered hereafter as the reference rainfall upon which we evaluated the satellite products. All forcing datasets and associated resulting streamflow simulations were aggregated to 3-hourly and daily temporal resolutions. Time periods where radar and all satellite products agreed of no precipitation occurrence within each basin were excluded from the respective error analysis. Reference basin-average precipitation and runoff simulations for each basin were ranked to define threshold values associated with the 20^{th} and 90^{th} quantiles; these threshold values were used to group datasets used in error analysis. The 90^{th} quantile precipitation and flow threshold values are visualized in Table 3.1. In the table, precipitation threshold values reported as mm/h for two time resolutions (3-Hourly and daily) and seasons (cold and warm) along with five basin scales. There is clear basin dependency of the thresholds, and the thresholds values are little higher in warm season compare to the cold one. In terms of time resolution, daily thresholds are smaller than values in high resolution, and the resolution effects are more significant in small scales. In

terms of the flow thresholds no significant effect is observed with time resolution change. The flow thresholds values in cold season are higher than in warm season.

Error analysis is performed separately for cold and warm season months that represent different precipitation types. According to Lins (1997) study cold season is best defined during November through January in the region while similar characteristic is also observed through April. Therefore, the month of April is considered as conjunction with the timing of the region's snowmelt in his study. Additionally, Marshall & Randhir (2008) showed in their study over CRB that snowfall starts in October, maximum snow occurs in January and the amount typically decreases through April. Therefore, in this study, the months between May and October were considered as warm period while November through April were selected as the cold season months.

3.2.3. Hydrologic Model

The Coupled Routing and Excess Storage (CREST) model used in this study is a hydrologic modeling system developed by the University of Oklahoma and NASA SERVIR (www.servir.net) (Wang, et al., 2011). CREST is a spatially distributed rainfall-runoff model designed to simulate flow discharges over regional and global scale. The distributed CREST model is applicable globally based on geomorphologic, land cover and soil type datasets. It has a total of 14 calibration parameters that are estimated with an automatic global optimization algorithm (Differential Evolution Adaptive Metropolis, DREAM). DREAM runs to mitigate sum of squared residuals to estimate posterior parameters using multiple Markov Chain Monte Carlo chains from prior iterations (Vrugt, et al., 2009). Five of the model parameters, i.e. the initial soil condition, infiltration capacity of soil layer, flow speed multiplier over slope and channel,

evapotranspiration multiplier were calibrated in this study with DREAM using streamflow data at gauging stations over CRB, while the rest of the parameters were defined based on physical criteria relating to land cover and soil types.

For the parameter calibration, a 3-year period from 2005 to 2007 was selected and the first two and a half months (January to March 15th) were used as warm-up period for soil moisture initialization. Two periods were employed for model validation (2003 to 2004 and 2008 to 2009) after the parameter calibration. Evaluation was performed at nine streamflow gauging stations. The Nash-Sutchliffe Coefficient of Efficiency (NSCE), Mean Relative Error (MRE), Centered Root Mean Square Error (CRMSE), and Pearson Correlation Coefficient (PCC) values of hourly flow simulations varied from 0.31 to 0.68, -0.06 to 0.13, 61 to 121 (%) and 0.60 to 0.83, respectively. The performance metrics indicates that CREST has sufficient accuracy for evaluating uses of satellite data in flow simulations.

3.2.4. Error Metrics

CREST simulations were performed using the satellite and radar datasets as input in the calibrated hydrologic model for the period 2002 to 2010. The runoff simulations were compared to the observed streamflow at the USGS gauging stations to evaluate the consistency of simulated flows from the different forcing datasets. The error analysis methodology is visualized in Figure 3.2. The three parts of the flowchart are satellite-based simulations, radar-based simulations and the observations at the nine basin outlets. The error analysis is conducted in terms of two aspects. First, we evaluate the disagreement between basin-average satellite precipitation values to the basin-average radar precipitation for the nine basins that constitute our

case study. Second, we evaluate the differences of the satellite-driven flow simulations to the radar-driven simulations and the actual streamflow observations.

The error metrics used in this analysis are the Mean Relative Error (MRE), Centered Root Mean Square Error (CRMSE), and Pearson Correlation Coefficient (PCC). These error metrics are listed below:

$$\mathrm{MRE} = \frac{\sum_{i=1}^{n}\left(P_{s,i} - P_{r,i}\right)}{\sum_{i=1}^{n} P_{r,i}} \tag{3.1}$$

where P stands for the basin-average precipitation [mm/h] or simulated flow [m^3/s] time series. The subscripts r and s indicate reference and simulated (radar or satellite) time series, respectively. Note that the reference of flow time series can be the radar-driven flow simulations or the observation. The index i represents the i^{th} time step and n is the total number of time steps. MRE gives indications on error direction (under/over-estimation) and on average how close the estimates are to the references.

$$\mathrm{CRMSE} = \frac{1}{\overline{P_r}}\sqrt{\frac{\sum_{i=1}^{n}\left[P_{r,i} - P_{s,i} - (\overline{P_r} - \overline{P_s})\right]^2}{n}} \times 100\% \tag{3.2}$$

$\overline{P_r}$ and $\overline{P_s}$ stand for the mean of the basin-average precipitation (or flow) values. CRMSE is used to evaluate the random error between simulations and reference values relative to the mean reference value. Lower CRMSE indicates better consistency and the value of zero signifies no random error between the time series.

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}\left(P_{r,i} - \overline{P_r}\right)\left(P_{s,i} - \overline{P_s}\right)}{\sqrt{\sum_{i=1}^{n}\left(P_{r,i} - \overline{P_r}\right)^2 \sum_{i=1}^{n}\left(P_{s,i} - \overline{P_s}\right)^2}} \tag{3.3}$$

PCC measures the strength of linear dependency between reference and satellite rainfall or simulated streamflows.

In addition to the three error metrics, the associated Error Metric Ratios (EM_{Ratio}) were analyzed. The ratio was calculated to measure the skill of satellite-driven simulations relative to the hydrologic model performance evaluated using the streamflow observations as a reference for radar-driven runoff simulations. EM_{Ratio} measures the relative significance of satellite precipitation uncertainty to the modeling uncertainty in terms of flow simulations. It is calculated as follows:

$$\mathrm{EM_{Ratio}} = \left| \frac{\mathrm{EM_{s,Q}}}{\mathrm{EM_{m,Q}}} \right| \tag{3.4}$$

where EM represents error metric (here use the MRE and CRMSE), while subscripts "s" and "m" indicate that those metrics were evaluated for satellite-driven runoff simulations using radar-rainfall simulations as reference, and radar-driven simulations using streamflow observations as reference. The EM_{Ratio} evaluates the significance of satellite precipitation uncertainty in runoff simulation relative to the modeling uncertainty considering that the Stage IV radar-rainfall product represent an accurate estimate of basin-average precipitation. Additionally, the Error Statistic (ES_{Ratio}) ratio was determined between runoff and precipitation CRMSE for each satellite product. The ES_{Ratio} was applied to explore how the random component of basin-average precipitation uncertainty changes through streamflow simulations. It is calculated as follows:

$$\mathrm{ES_{Ratio}} = \left| \frac{\mathrm{CRMSE}_{s,Q}}{\mathrm{CRMSE}_{s,P}} \right| \tag{3.5}$$

Since rainfall detection is an important component of satellite precipitation uncertainty, which affects water budget calculations, two relevant contingency table statistics, namely False

Alarm Volume Ratio (FAVR) and Missed Rain Volume Ratio (MRVR), were derived using the Stage IV radar-rainfall as reference. These two contingency table statistics have a range of 0 (the perfect performance) to 1, and they are given by:

$$\mathrm{FAVR} = \frac{\sum_{t=1}^{n}(P_{Sat} \mid P_{Rad} = 0 \;\&\; P_{Sat} > 0)}{\sum(P_{Rad})} \tag{3.6}$$

$$\mathrm{MRVR} = \frac{\sum_{t=1}^{n}(P_{Rad} \mid P_{Rad} > 0 \;\&\; P_{Sat} = 0)}{\sum(P_{Rad})} \tag{3.7}$$

MRVR defines the relative volume of satellite missed precipitation over the basin, while FAVR defines the relative volume of false detected satellite rainfall relative to the reference rainfall. In continuous hydrologic simulations MRVR and FAVR are important parameters that affect soil wetness and evapotranspiration with consequential effects on runoff simulations (Yong, et al., 2012; Tang, et al., 2010; Hossain, et al., 2009).

Finally, the contingency table-based binary score, Critical Success Index (CSI), was examined. The CSI value ranges from 0 to 1, with the perfect performance in detection represented by one. Classification of stream flow estimates constructed to the contingency matrix and CSI are calculated as below:

	$Q_s > Q_s^{Threshold}$	$Q_s \leq Q_s^{Threshold}$
$Q_r > Q_r^{Threshold}$	A	B
$Q_r \leq Q_r^{Threshold}$	C	D

$$\mathrm{CSI} = \frac{N_A}{N_A + N_B + N_C} \tag{3.8}$$

where N_A, N_B, N_C, N_D represent the total number of occurrence for A,B,C, and D situations, respectively. A point to note is that the CSI is sensitive to data sample misbalance in the

contingency table. Therefore, multiple (10,000 random iterations) balanced samples representing equal number of hits and non-hits were produced based on which we report the mean of CSI (Chen, et al., 2004). The threshold values used in the CSI contingency table were represented for each basin by the 20^{th} and 90^{th} quantiles of the reference ($Q_r^{Threshold}$) and satellite-driven ($Q_s^{Threshold}$) runoff values.

3.3. Results

3.3.1. Assessment of Model Performance

Figure 3.3 shows the basin-average precipitation accumulation time series (left column) for basins in B1, B3, and B5 scale categories and for the warm season months. In the precipitation estimations, results show significant differences among the three satellite products in the three basin scales examined. The TRMM-3B42V7 slightly overestimates the reference precipitation, but captures the temporal correlation relatively well. This satellite algorithm produces the closest estimates relative to the reference rainfall, and this is attributed to the gauge adjustment applied in this product. Besides, the CMORPH and PERSIANN products consistently underestimated the Stage IV radar precipitation in the three basins. PERSIANN and CMORPH give similar trend of underestimation for the small and medium size basins while their performance improves with basin size. Additionally, normalized hydrographs are visualized in cumulative time series to compare the different satellite-derived runoff simulations against the radar-derived simulations and observed streamflows for the same period (right column in Figure 3.3). We note overestimation of TRMM-3B42V7 and radar-driven flow simulations relative to the observations. CMORPH and PERSIANN-driven simulations on the other hand exhibit lower biases relative to the actual observations, except for the medium size basin where biases are

similar to radar-driven simulations. This exemplifies the complex interaction of errors in hydrologic modeling and precipitation estimation.

Figure 3.4 visualizes the error ratios (EM_{Ratio}) in terms of the relative significance of satellite precipitation uncertainty error propagation in runoff simulation to the hydrologic modeling uncertainty. We present EM_{Ratio} for MRE the two seasons and two thresholds (20^{th} and 90^{th} quantiles). The multiple gray scale circles are used to represent the different basin scales. The increasing distance of the ratios from one indicates higher biases in either the satellite or radar-driven flow simulation relative to streamflow observations. It is shown that in most of the cases the mean relative error ratio of satellite to modeling uncertainty increases with decreasing basin scale, and this basin scale dependency is amplified in warm season months. In cold season months all ratios are consistently greater than one indicating higher runoff bias due to satellite precipitation error-alone relative to the modeling uncertainty alone. This tendency is the smallest for TRMM-3B42V7 and largest for CMORPH with PERSIANN being in between the two products. In the warm season months, on the other hand, the ratios are further than one (either above or below one) in CMORPH and PERSIANN relative to TRMM-3B42V7. Using the low threshold (20^{th} quantile) in the analysis the two products (CMORPH and PERSIANN) exhibit ratios lower than one with stronger basin scale dependency indicating that the contribution of modeling uncertainty in the streamflow simulation bias is more significant than the satellite precipitation uncertainty-alone, which is attributed to improved estimation of the precipitation volume by the two products in combination with increased systematic error in the simulation of low flows by the model.

Figure 3.5 visualizes the CRMSE error metric ratios of the satellite-driven runoff simulations to the hydrologic model CRMSE. These ratios show that the random component of

the satellite-driven runoff simulations is stronger than the hydrologic model uncertainty alone in most of the cases examined. The ratios increase inverse proportionally to basin scale especially during warm season months. In addition, change in the temporal resolution (3-hourly to daily) does not yield significant differences in the CRMSE ratio values or the scale dependence of these values. Although PERSIANN product exhibits slightly higher ratios during cold season months, these ratios drop significantly during warm season months, particularly for the 90^{th} percentile values. Moreover, in most of the warm season cases, CMORPH produces the highest ratio values indicating the highest contribution of precipitation product uncertainty in runoff simulation error. A point to note is the stronger basin scale dependence of the random error ratios for the 90^{th} percentile. Specifically, the figure shows that for the 90^{th} percentile values the modeling error component becomes more significant than the error in satellite-driven simulations as the basin scale increases, which is attributed to an overall improved performance of CMORPH product in the warm season precipitation estimation. The satellite precipitation error analysis is discussed next.

3.3.2. Precipitation Error Analysis

In Figure 3.6, MRE_P values determined for the satellite precipitation products with respect to the radar are shown over the different basin scales and two thresholds: 90^{th} (P_{90}) and 20^{th} (P_{20}) percentile. MRE_P values for P_{20} show better performance compared to P_{90} due to the high spatiotemporal variability in the higher threshold. Results show that MRE_P values, overall, increase as a function of basin scale with the exception of CMORPH during cold season months exhibiting few increases with basin scales. At both cold and warm season months for higher precipitation rates (P_{90}), the satellite products tend to underestimate rainfall, and this trend does

not change for CMORPH and PERSIANN when considering data from P_{20}. On the other hand, MRE_P values for TRMM-3B42V7 yield the least biases for the basin-average precipitation estimates.

The centered root mean square error statistics for precipitation ($CRMSE_P$) are demonstrated for the three satellite products in Figure 3.7. It is shown that CRMSEP values vary from 40% through 100% for precipitation values above P_{90}, and between 100% and 300% when considering values exceeding P_{20}. Overall, $CRMSE_P$ is basin scale dependent, particularly apparent for the TRMM-3B42V7 product. We note significant drop in $CRMSE_P$ values moving from 3-hourly to daily time resolution, especially when considering data from P_{20}, pointing to the fact that random error is reduced due to the temporal aggregation. Other than the stronger basin scale dependence in TRMM-3B42V7, no significant difference is observed between the examined satellite products in terms of $CRMSE_P$. From the perspective of seasonal effect, a slight reduction is noted from cold to warm season.

Pearson Correlation Coefficient for precipitation (PCC_P) is plotted in Figure 3.8 for the various comparison dimensions. Basin scale dependency is observed in correlation values; this is more apparent for the daily scale results. For the 3-hourly rainfall time series data, it follows that larger scale basins produce higher PCC_P values (greater than 0.5 in some cases), referring to the more significant dampening effects on rainfall error from larger basins. Smaller scale basins exhibit lower correlation values. When it comes to high values (P_{90}), PCC_P are reduced on all three satellite products and both seasons. Overall, TRMM-3B42V7 gave the highest PCC_P, while PERSIANN product is designated with the lowest PCC_P values for most of the cases. Besides, from the perspective of temporal correlation, PCC_P values improve significantly from 3-hourly to

daily temporal resolution, indicating that the temporal similarity is higher when aggregating the data to a coarser resolution.

Figure 3.9 shows FAVR and MRVR results, and their ratios for the various comparison dimensions. The first and last two columns differentiate the cold and warm period months, respectively. The accuracy of satellite rain detection is higher at daily temporal resolution compared to the 3-hourly temporal resolution as indicated for the three satellite products by the lower FAVR and MRVR values in both seasons. These indices become nearly zero at monthly temporal resolution (results not shown). Results show that CMORPH and PERSIANN are characterized by relatively high MRVR and FAVR, respectively, during the cold period months. In the warm season, PERSIANN is characterized by the highest FAVR among the three products. CMORPH yields the smallest FAVR values in general. Basin scale dependency is shown for the statistics where both FAVR and MRVR are decreasing with basin scale for two of the products (CMORPH is an exception). The ratios of FAVR to MRVR are generally close to one, indicating balanced missed and false detection rain volumes, with the exception of CMORPH during cold season where the ratio drops below 0.1. CMORPH is dominated by MRVR in cold season, indicating the well-known susceptibility of this technique to missed precipitation in passive microwave retrievals over frozen or snow-covered surface conditions.

Below we analyze the errors resulting from the above precipitation products in flow simulations using as streamflow reference the Stage IV radar driven runoff simulations.

3.3.3. Flow Simulation Error Analysis

The error of satellite precipitation products in runoff simulation is evaluated over the CRB at 3-hourly and daily temporal resolutions for the different basin scales and two runoff

thresholds: 20th (Q_{20}) and 90th (Q_{90}) percentile values. Figure 3.10 shows MRE values for flow (MRE_Q). MRE_Q values range from -0.8 (underestimation) to 0.2 (overestimation). Underestimation is generally observed in all scenarios, with the exception of TRMM-3B42V7 in warm season. CMORPH exhibits the strongest biases in flow simulations, which is exemplified during cold period months due to the strong precipitation underestimation shown in the previous section. In general, MRE_Q for all satellite products reduces as basin scale increases, indicating a dampening error effect through the rainfall-runoff transformation process. In terms of seasonal effects, it is noted that MRE_Q values are lower in the warm season. All satellite-driven flow simulations exhibit strong negative biases relative to the radar-based runoff in high runoff values (Q_{90}); this effect is especially noted during cold period months.

$CRMSE_Q$ is presented as a percentage in Figure 3.11. $CRMSE_Q$ values range from 35% to 100% in all scenarios with the exception of the Q_{20} group for the small scale basins in cold season that exceed 100%. $CRMSE_Q$ is shown to decrease with increase in basin scale indicating the higher degree of buffering effect due to rainfall-runoff processes. The higher $CRMSE_Q$ values over the small scale basins (B1) is also attributed to the fact that satellite products cannot capture the relatively high rainfall variability at those spatial scales, resulting in higher degree of randomness in runoff simulations. Lower values of random error are shown for the warm season months relative to the cold season, which points to the potential issue in snow screening process of the satellite algorithms. The $CRMSE_Q$ values at the daily time scale drop to lower values due to the temporal aggregation effect in buffering the spatiotemporal variability of rainfall. The runoff to rainfall CRMSE ratio defined in Eq. (3.5) is reported in Table 3.2 (20th percentile) and Table 3.3 (90th percentile). Left (right) panels show 3-hourly (daily) temporal resolution results. A point to note is that the ratios are lower than one indicating dampening effect of the random

error component from rainfall to runoff. The second point to note is the basin scale dependence of CRMSE ratios. The larger scale basins exhibit stronger dampening (lower ratio values) for errors in basin-average precipitation to that in runoff simulations, which is consistent to results presented by Maggioni, et al. (2013). In addition, the daily time scale and warm period months show a slightly weaker dampening of the random error. The ratio values are lower for the 20^{th} quantiles relative to the 90^{th} quantiles.

In Figure 3.12, the Pearson Correlation Coefficient (PCCQ) values between the different satellite products and radar driven flow simulations are presented. The figure shows that the PCC_Q values are slightly lower for the 3-hourly time series when comparing to the daily temporal resolution. Additionally, larger basins are characterized by lower PCC_Q during cold season. The results also demonstrate that simulations from the TRMM-3B42V7 product is associated with the highest PCC_Q value in general with respect to the other two satellite-driven runoff simulations regardless of the seasons and basin scales (some exception exist for the high flow simulation values). It is noted that the weakest PERSIANN-based PCC_P values (in Figure 3.8) for the P_{20} yield the weakest PCC_Q values in flow simulations for the Q_{20} threshold.

The Critical Success Index (CSI) of satellite-driven simulations capturing flows exceeding the 20^{th} and 90^{th} percentiles is visualized in Figure 3.13. A first observation from Figure 3.13 is that CSI patterns for the cold season months are different to those for the warm season for the Q_{90} group, while those for the Q_{20} group are similar. The CSI of determining flows exceeding Q_{20} are above 0.7, with TRMM-3B42V7 based simulation exhibiting the highest performance (nearly 0.8 CSI). The skill of satellite driven simulations, on the other hand, reduces when considering flows exceeding Q_{90}, where CSI is in a range of 0.2-0.65. The highest CSI values are shown for the TRMM-3B42V7-simulated flow in the largest scale basin.

3.4. Conclusions

In this study, three satellite products, representing a gauge-adjusted (TRMM-3B42V7) and two satellite-only (CMORPH, and PERSIANN) estimates, were examined in terms of their performance in capturing basin-average precipitation and for simulating streamflows over different sub-basins of the Connecticut River Basin (CRB). Hydrologic simulations were performed with a distributed hydrologic model (CREST) calibrated based on nine years of streamflow observations from the case study basins. Temporal resolution (3 hourly vs. daily), seasonal effects (warm vs. cold period months) and basin scale dependency (191 to 25,000 km^2) were studied for high and low flow conditions. Gridded radar-rainfall (mm/h) from the WSR-88D stage IV product was used to represent the basin-average reference precipitation. USGS streamflow readings were employed to verify CREST simulations and evaluate the model simulation uncertainty. Graphical methods and quantitative error statistics, namely MRE, CRMSE, PCC for precipitation and flows along with their associated ratios and error propagation ratios were analyzed.

Results showed that TRMM-3B42V7 precipitation estimates, overall, can better capture precipitation and runoff time series over the mid-latitude case study basin compared to the other two satellite products, and this is attributed to the product's gauge adjustment. On the other hand, the non-gauge adjusted CMORPH and PERSIANN products consistently underestimated reference precipitation, and this suggests that a bias adjustment of the products would potentially improve the representation of basin rainfall estimates. The performance of CMORPH in warm season is significantly better than in the cold season where it exhibits strong underestimation of precipitation. PERSIANN, overall, showed more realistic estimates compared to CMORPH, and the performance increased as a function of basin area. However, PERSIANN exhibited more

frequent false precipitation detection than the other two products, and CMORPH is dominant for missed rainfall volume rates. In addition, we noted reduction in random error moving from 3-hourly to daily temporal resolution for both cold and warm season basin-average precipitation time series.

Satellite driven runoff time series were compared to radar-driven runoff simulations and streamflow observations at gauging stations. CMORPH-based runoff simulations considerably underestimated in both seasons, which is attributed to the underestimation exhibited in the precipitation error analysis. Underestimation tendency of CMORPH becomes more apparent for the high flow values (> 90^{th} percentile). TRMM-3B42V7 based runoff simulations gave reasonable results relative to other two satellite products. TRMM-3B42V7 driven simulations captured best the flow dynamics. Similar to the precipitation error analysis, daily values gave more accurate results than the 3-hourly. In addition to temporal resolution effect, significant basin dependency was captured in the results. It is observed that the skill of modeling the streamflow increases with basin scale. Finally, the error propagation was calculated as the ratio between runoff error metrics (MRE and CRMSE) to the associated precipitation error metrics. Similar to the basin dependency of runoff CRMSE, error dampening from precipitation to runoff is shown to increase proportionally to basin size for all satellite products.

While the results from this study are based on a single basin and a specific hydrologic model, many of the findings can generally apply. Overall, the conclusions of this work provide useful information about the application of satellite precipitation estimates for runoff simulations and evaluation of flood risk with a distributed model. Further research is required to enhance hydrologic applications, such as utilizing various satellite precipitation products for flood modeling or water resource assessment, and evaluate how accurate the current products are in

deriving flood frequency analyses and for issuing precipitation or runoff threshold based early warnings of floods. Reliable assessment of both flood magnitudes and their return periods can provide a more complete description of basins' flood regime. In the Global Precipitation Measurement (Hou, et al., 2014) era precipitation products can ultimately be used to estimate future floods and for optimal water resources management, such as designing flood related structures and regulation of flood plains.

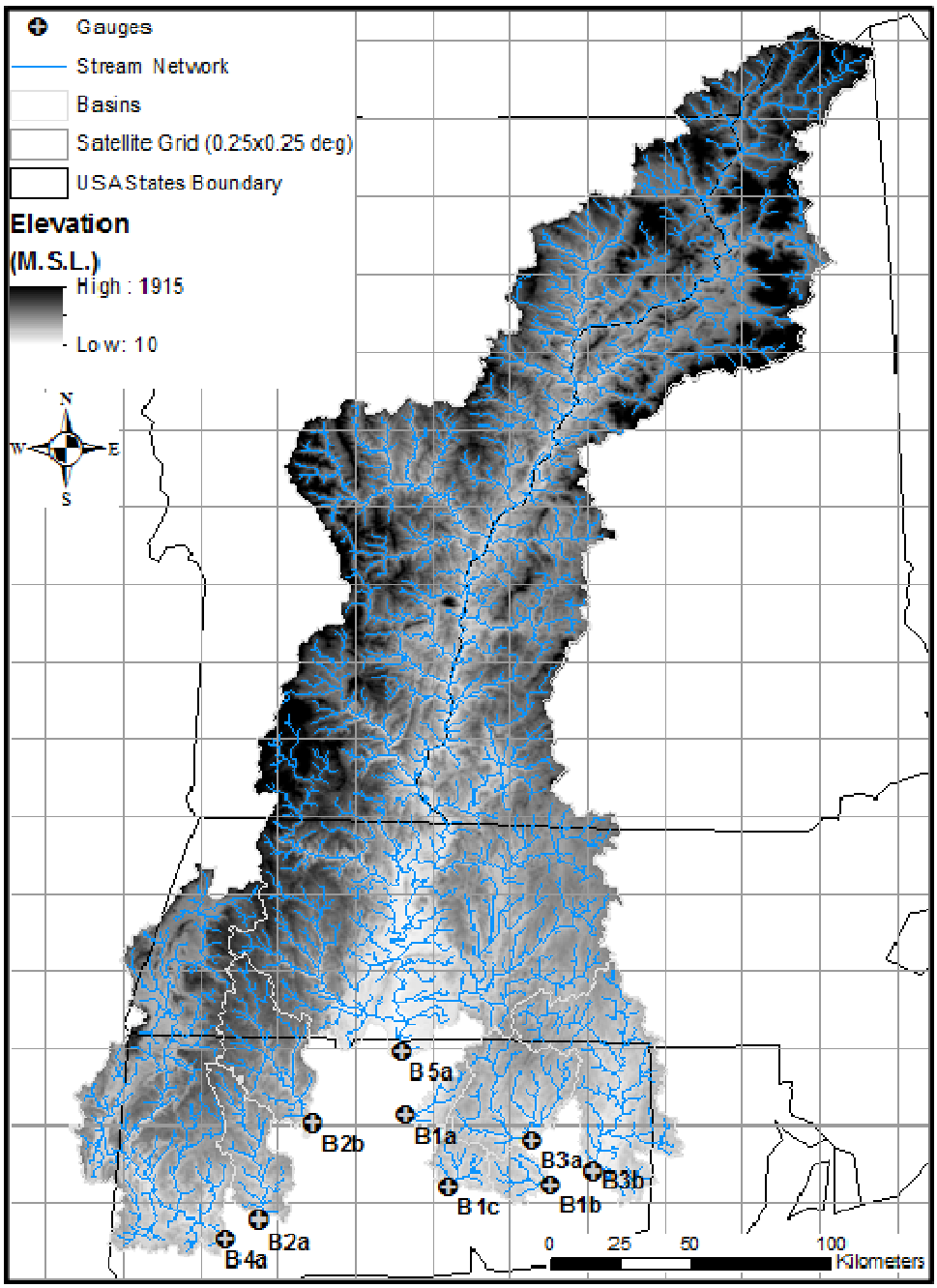


Figure 3.1. Map of Connecticut River Basin (CRB) and the location of USGS streamflow gauges.

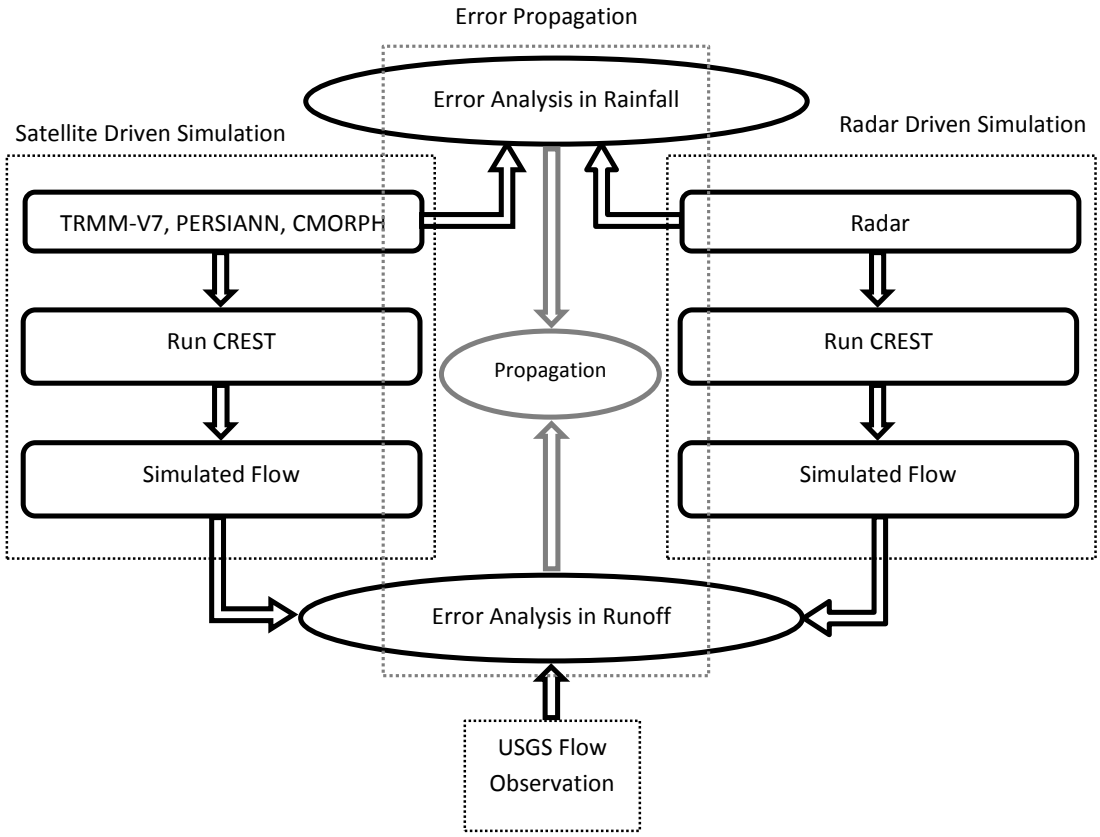


Figure 3.2. The error analysis methodology flow diagram

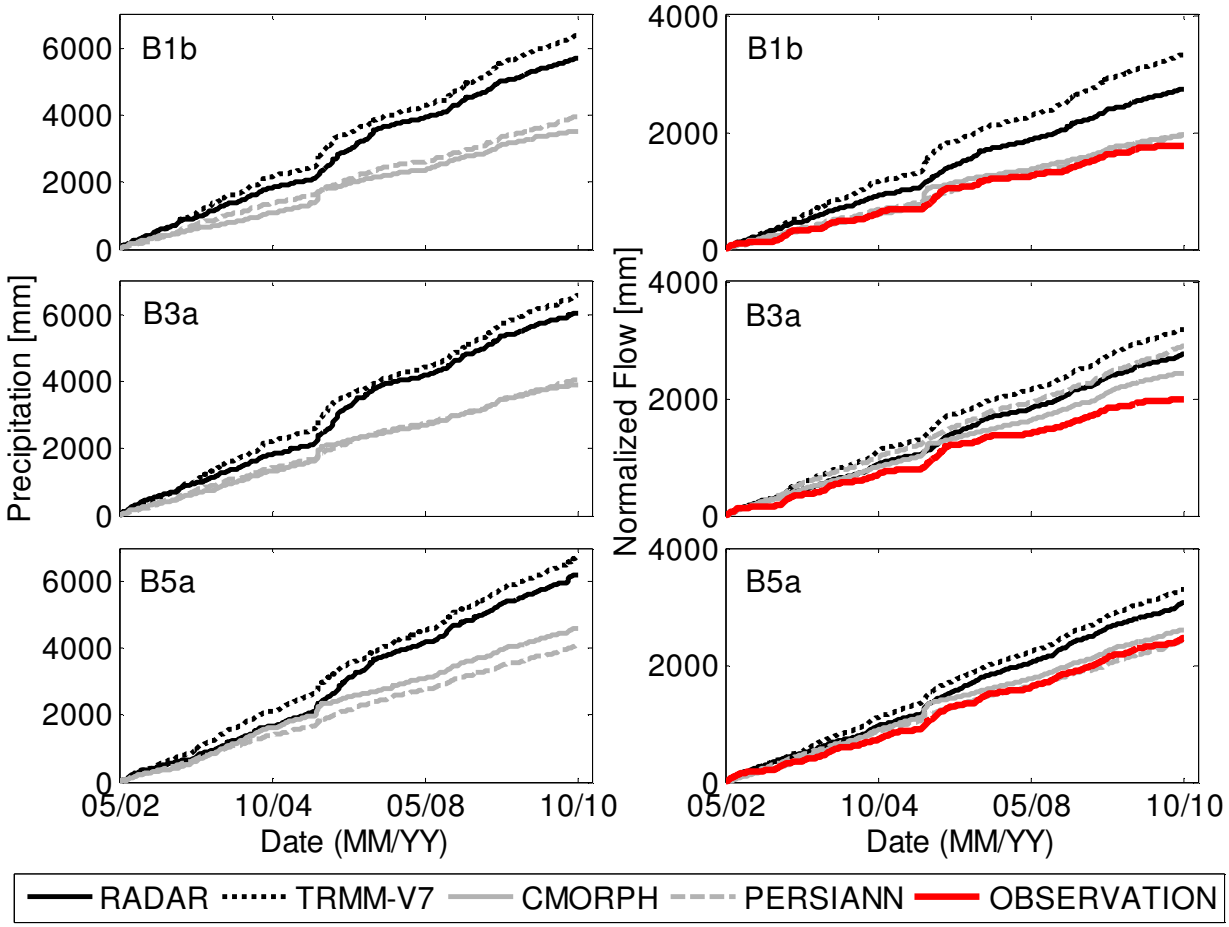


Figure 3.3. Cumulative comparisons between satellite products and radar precipitation data sets (left column) with normalized discharges (right column) for three basin scales B1b (small scale), B3a (medium scale), B5a (very large scale) over warm season

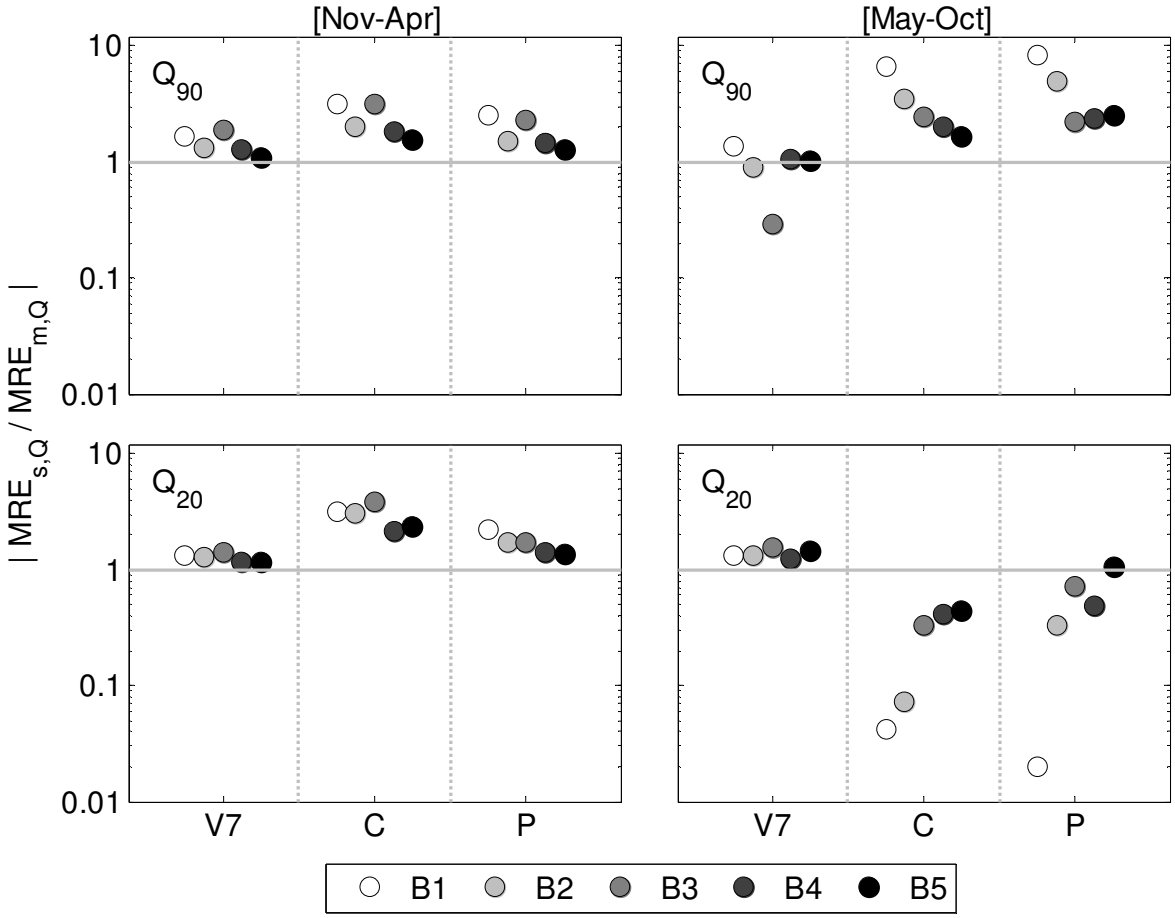


Figure 3.4. MRE_{Ratio} values determined for the different satellite products (V7 is for TRMM-3B42V7; C is for CMORPH; P is for PERSIANN), seasons, basin scales and thresholds (Q_{20} and Q_{90}).

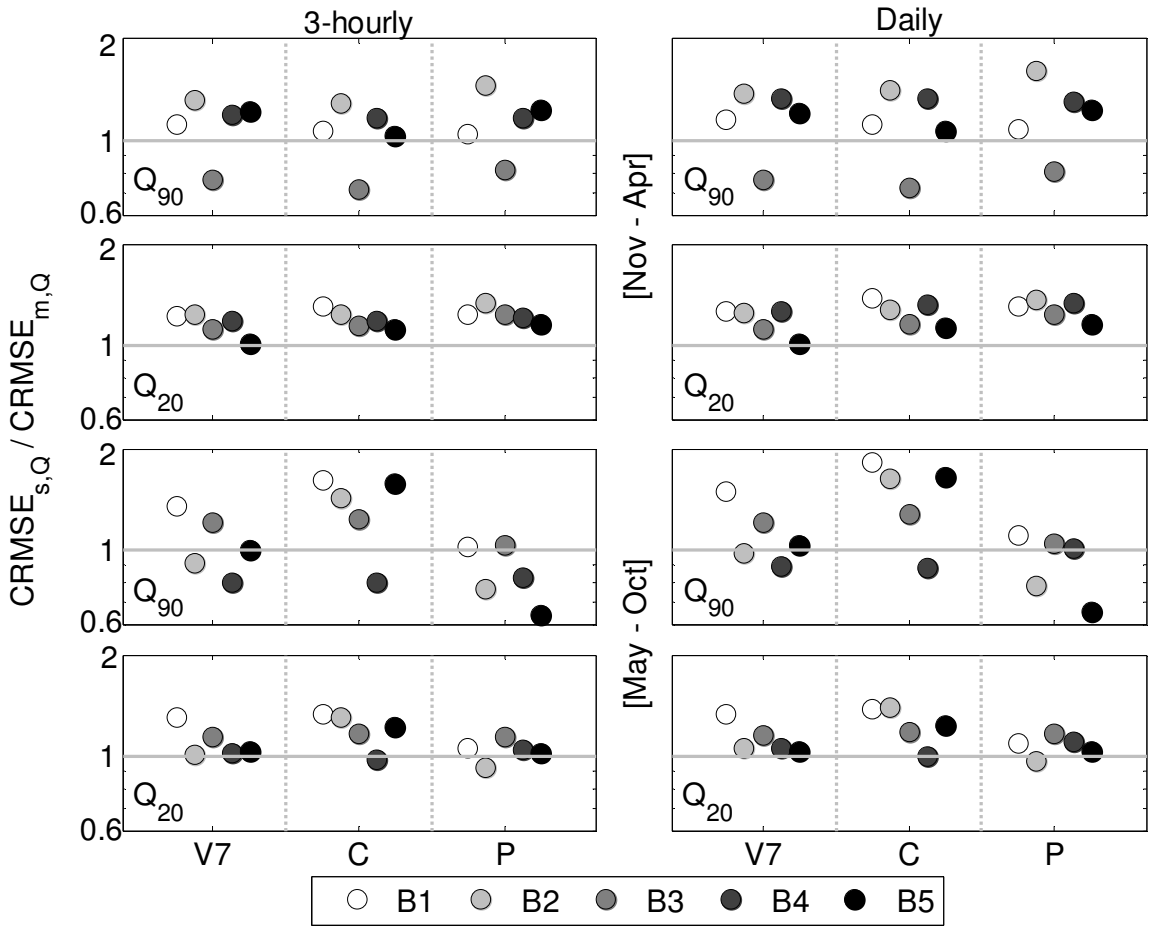


Figure 3.5. Same as in Figure 3.4, but for the $CRMSE_{Ratio}$.

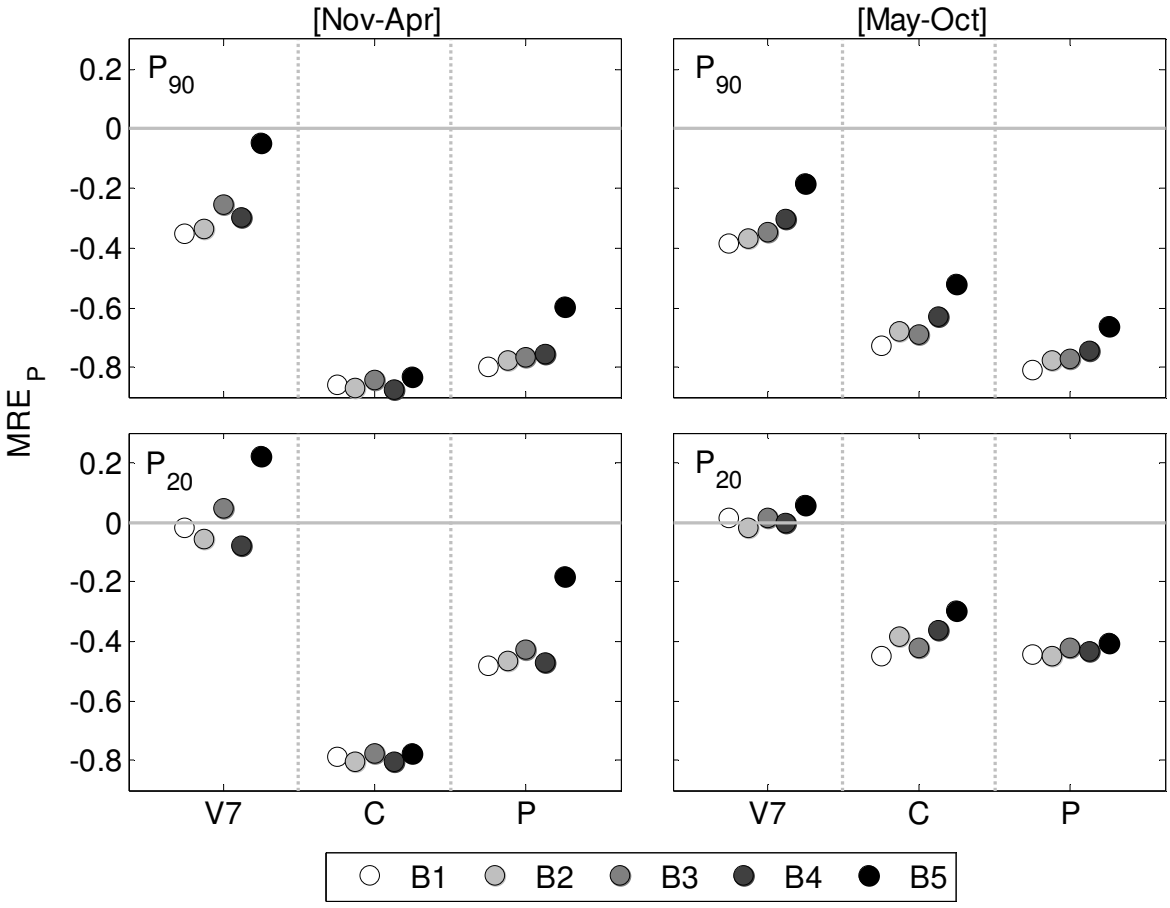


Figure 3.6. Mean Relative Error of precipitation (MRE_P) determined for the different satellite products (V7 is for TRMM-3B42V7; C is for CMORPH; P is for PERSIANN), seasons, basin scales and thresholds (Q_{20} and Q_{90}).

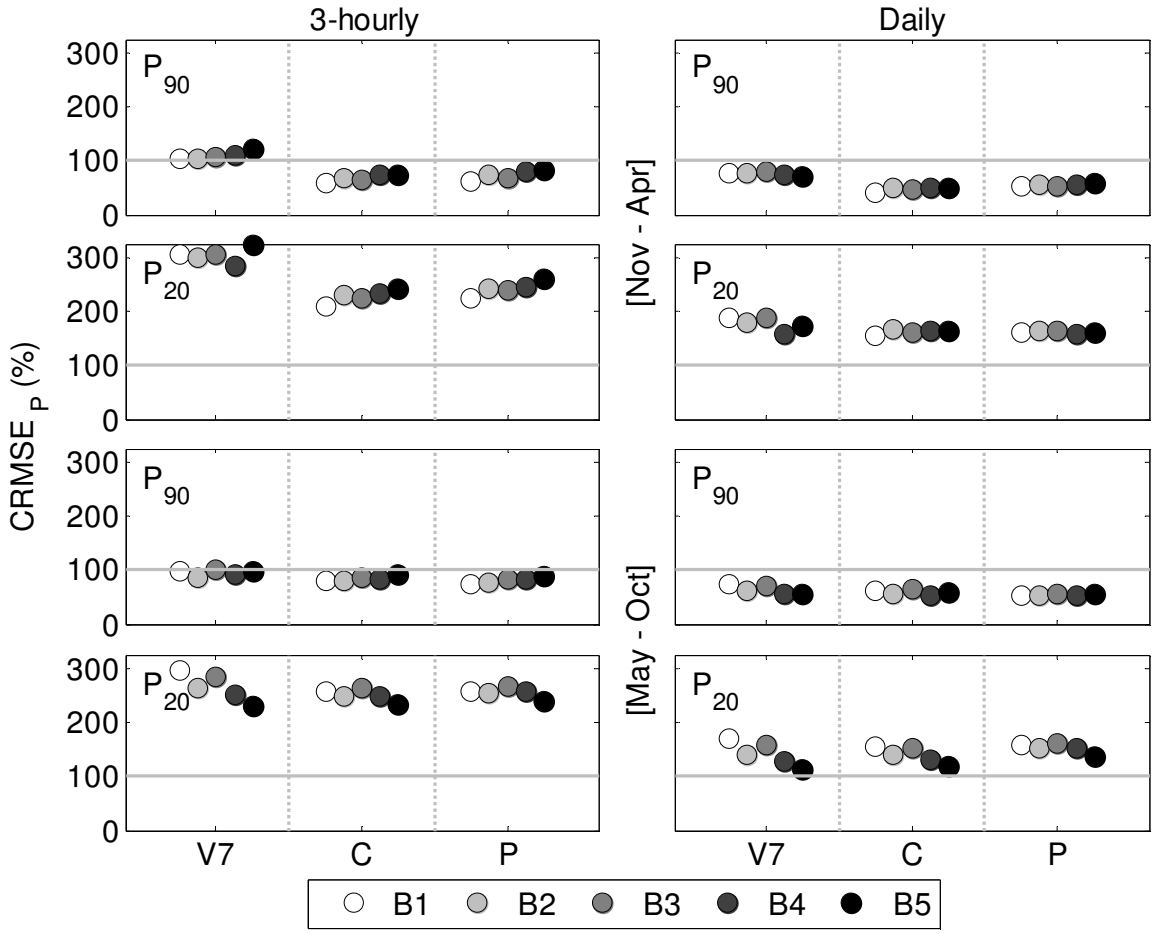


Figure 3.7. Same as in Figure 3.6, but for the Centered Root Mean Square Error of precipitation ($CRMSE_P$)

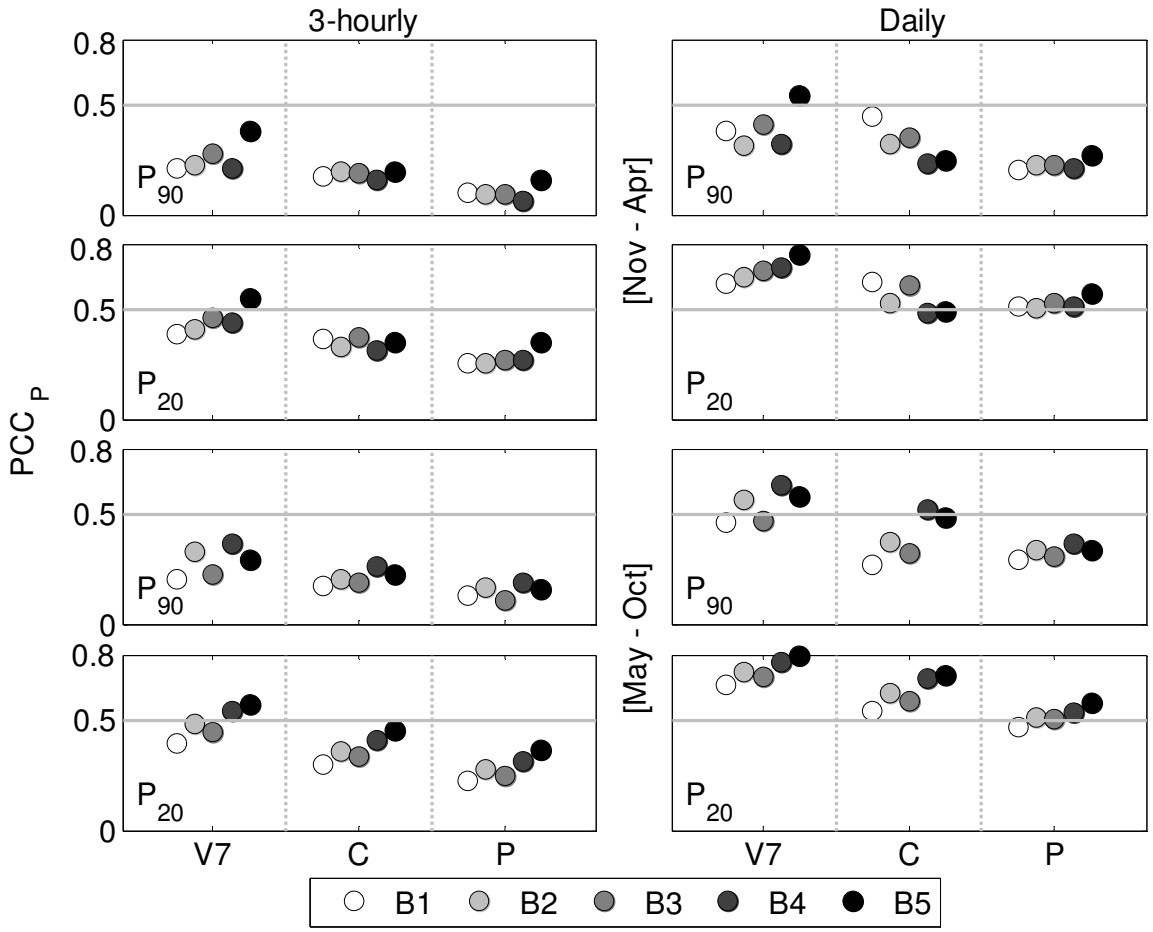


Figure 3.8. Same as in Figure 3.6, but for the Pearson Correlation Coefficients of precipitation (PCC_P)

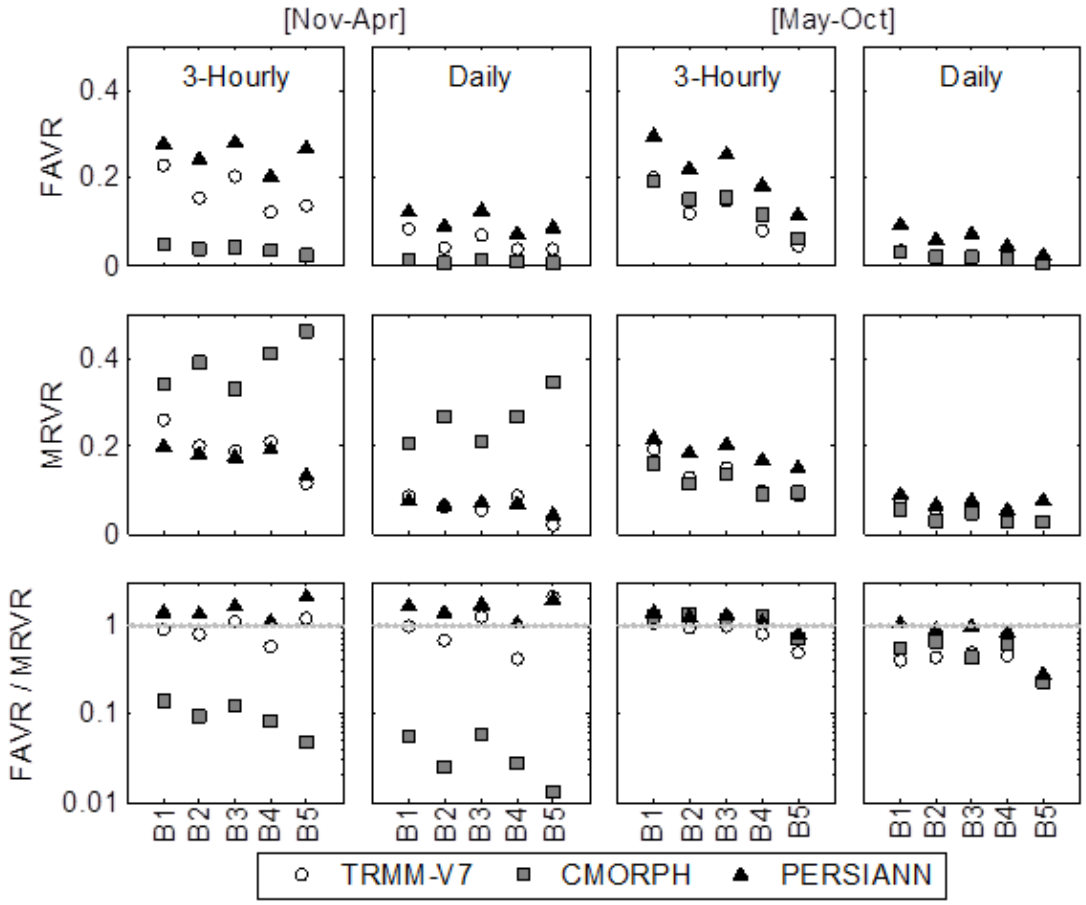


Figure 3.9. False Alarm Volume Ratio (FAVR), Missed Rain Volume Ratio (MRVR) and their ratios determined for the different basin scales and two temporal resolutions (3-hourly and daily).

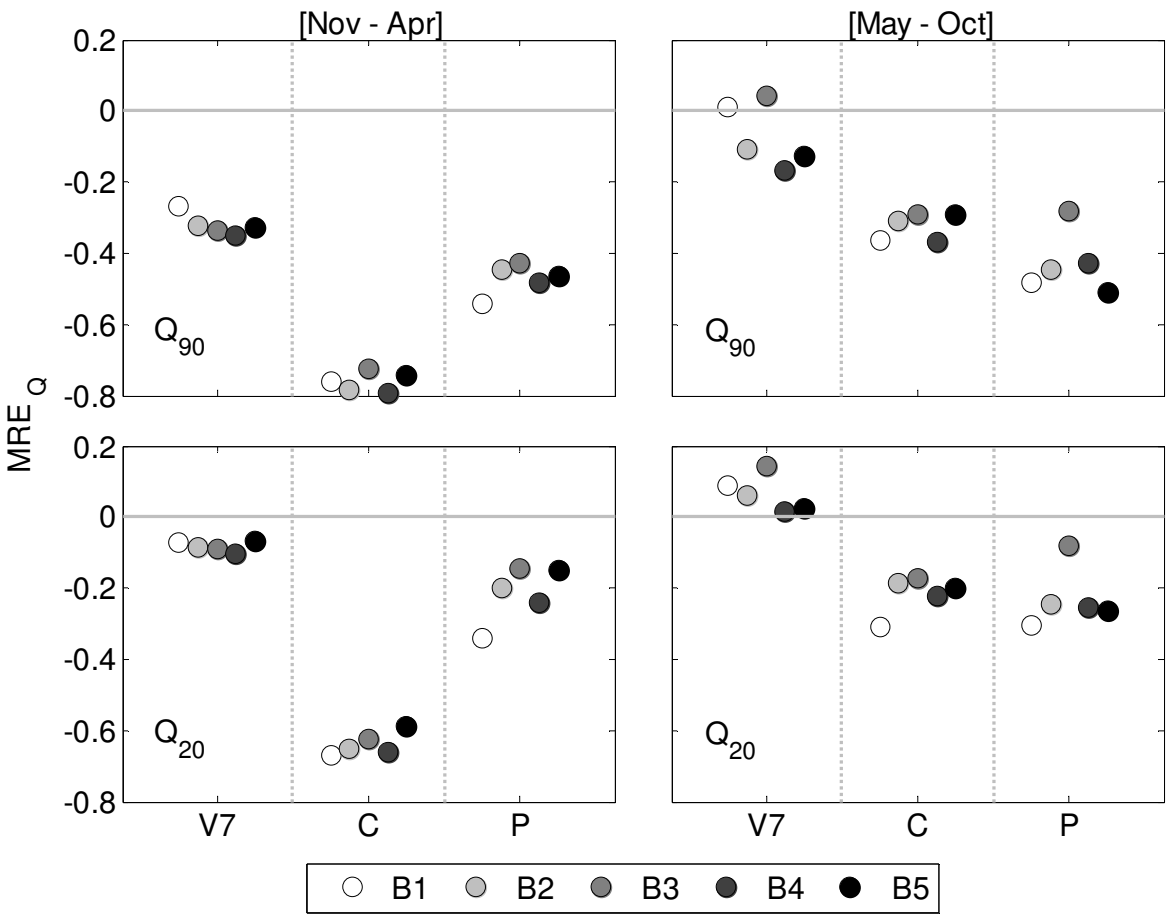


Figure 3.10. Same as in Figure 3.6, but for the Mean Relative Error in simulated flow (MRE_Q).

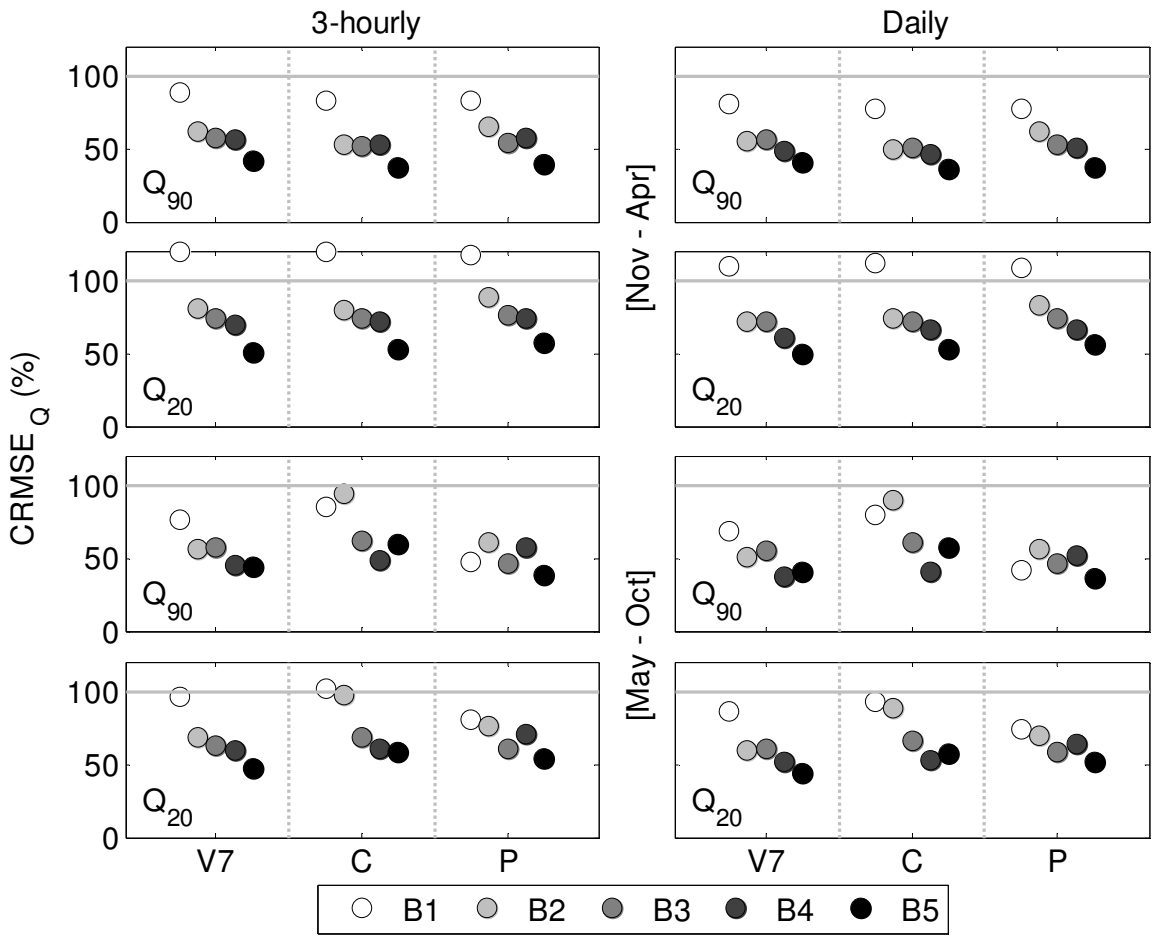


Figure 3.11. Same as Figure 3.7, but for the Centered Root Mean Square Error in simulated flow ($CRMSE_Q$)

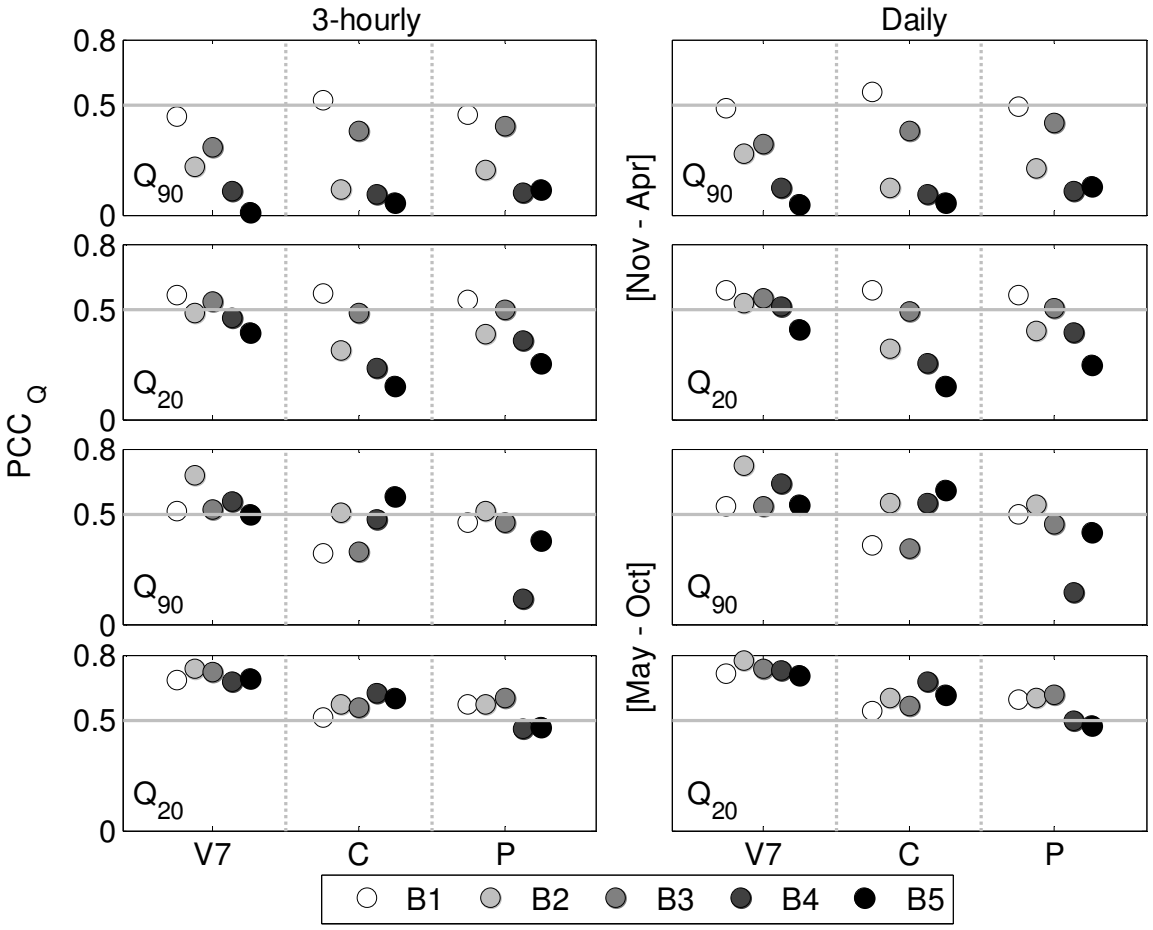


Figure 3.12. Same as in Figure 3.8, but for the Pearson Correlation Coefficient in simulated flow (PCC_Q)

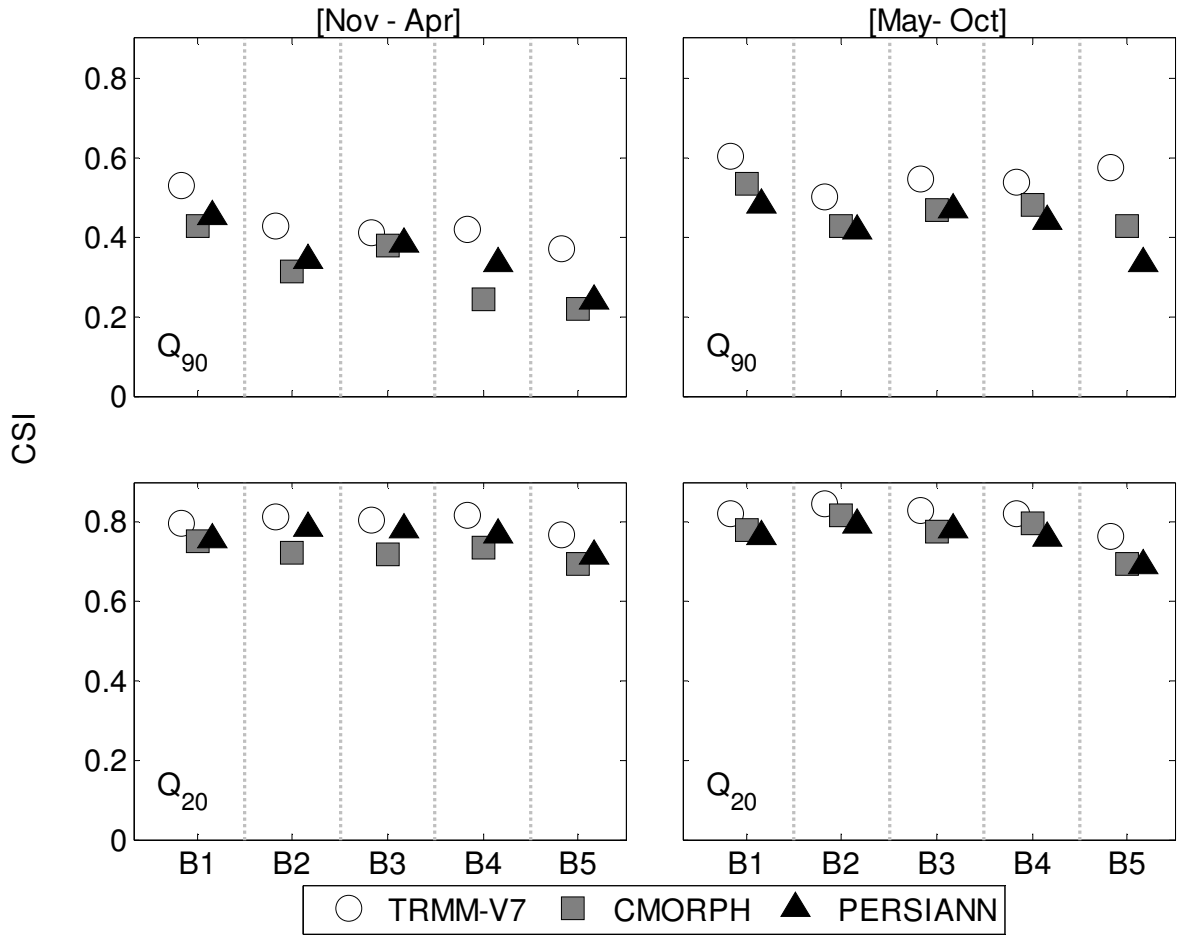


Figure 3.13. Critical Success Index (CSI) of detecting flows exceeding the 20th (Q_{20}) and 90th (Q_{90}) percentiles determined for the different basin scales and seasons.

Table 3.1. The 90th percentile values for precipitation and simulated streamflows.

Basin Scale	Precipitation [mm/h]				Runoff [m^3/s]			
	3-Hourly		Daily		3-Hourly		Daily	
	Cold	Warm	Cold	Warm	Cold	Warm	Cold	Warm
B1	1.10	1.14	0.64	0.62	12	10	12	10
B2	0.90	1.01	0.55	0.60	35	30	34	30
B3	0.87	0.93	0.57	0.58	72	55	72	55
B4	0.76	0.83	0.53	0.54	169	152	166	158
B5	0.39	0.65	0.34	0.48	875	837	874	840

Table 3.2. Ratios of the flow-to-precipitation Centered Root Mean Square Error ($CRMSE_{s,Q}$/ $CRMSE_{s,P}$) determined for the 20th percentile threshold values.

Basin Scale	3-Hourly						Daily					
	Cold			Warm			Cold			Warm		
	V7	C	P	V7	C	P	V7	C	P	V7	C	P
B1	0.4	0.6	0.5	0.3	0.4	0.3	0.6	0.7	0.7	0.5	0.6	0.5
B2	0.3	0.3	0.4	0.3	0.4	0.3	0.4	0.4	0.5	0.4	0.6	0.5
B3	0.2	0.3	0.3	0.2	0.3	0.2	0.4	0.4	0.4	0.4	0.4	0.4
B4	0.2	0.3	0.3	0.2	0.2	0.3	0.4	0.4	0.4	0.4	0.4	0.4
B5	0.2	0.2	0.2	0.2	0.3	0.2	0.3	0.3	0.4	0.4	0.5	0.4

Table 3.3. Same as in Table 3.2, but for the 90th threshold values.

Basin Scale	3-Hourly						Daily					
	Cold			Warm			Cold			Warm		
	V7	C	P	V7	C	P	V7	C	P	V7	C	P
B1	0.8	1.4	1.3	0.8	1.1	0.6	1.1	1.9	1.5	0.9	1.3	0.8
B2	0.6	0.8	0.9	0.6	1.2	0.8	0.7	1.0	1.1	0.8	1.6	1.1
B3	0.5	0.8	0.8	0.6	0.7	0.6	0.7	1.1	1.0	0.8	0.9	0.8
B4	0.5	0.7	0.7	0.5	0.6	0.7	0.7	0.9	0.9	0.7	0.8	1.0
B5	0.3	0.5	0.5	0.5	0.7	0.4	0.6	0.7	0.6	0.7	1.0	0.6

4. Using High-Resolution Satellite Precipitation Products for Flood Frequency Analysis: Case Study over the Connecticut River Basin

4.1.Introduction

Flooding is ranking as the most devastating natural disaster due to its causality to high economic loss and portion of fatalities (Borga, et al., 2011). International organizations have indicated an increasing mindfulness on the occurrence of floods/flash floods and associated inundation hazards (Leclerc & Ouarda, 2007; Brocca, et al., 2011). Therefore, studies on characterizing the frequency of occurrence of extreme flood events are critical for building societies and infrastructure resilient to hydro-climatic extremes. Some examples include developing flood control structures, flood plain maps, optimal reservoir operations, and dam designs.

Assessment of hydro-climatic extremes (i.e. floods or droughts) is typically based on long records of streamflow observations. Over the past half century in-situ data have been used as the source for flood frequency assessments (Keast & Ellison, 2013; Urias, et al., 2007; Boughten, et al., 1987; Langbein, 1949). However, for the sparsely gauged or ungauged catchments lacking sufficient in-situ measurements, running hydrologic models forced with observed or model-derived precipitation is a common way to simulate streamflows and thus analyze its frequency characteristics over different basin scales (Khan, et al., 2012; Brocca, et al., 2011; Blazkova & Beven, 2009; Moretti & Montanari, 2008). Brocca et al. (2011) implemented a semi-distributed model to estimate the design floods in small-medium basin (100-200 km^2) and flood forecasting for a large basin (>500 km^2) over the Upper Tiber River, Italy. In addition, Moretti & Montanari

(2008) used a distributed rainfall-runoff model to estimate peak streamflows for ungauged river sections in Riarbero River, Italy. Their results indicate that the peak flow estimation from hydrologic model is consistent with that from the regional empirical method. Aronica & Candela (2007) applied a simple stochastic rainfall-runoff model to stochastically simulate floods and consequently conducted frequency analysis of peak flows over Mediterranean catchments. They compared flood frequency distributions derived based on the simulated flows and the observed, and concluded that the model generated peak flows can reasonably reproduce the trend on the observed annual maxima peak flows.

The atmospheric forcing of the hydrologic models, i.e. precipitation, can either come from reanalysis or remote sensing datasets in case of data poor regions or absence of high quality ground observations. As the high resolution quasi-global satellite precipitation data now exceed 15 years of record, they become a valuable data source to support flood frequency analyses over regions lacking ground observations. Awadallah et al. (2011) used maximum daily rainfall depth from jointly ground gauges for eight years (1943-1950) and gauge adjusted TRMM data sets for 11 years (1998-2008) to develop intensity duration frequencies over the northwest Angola region, in Africa. In the study, mean and variance of the TRMM products were shown to be of similar magnitude to the ones derived from the ground gauge data. The study presented that over data poor regions, satellite datasets can be used to overcome the limited data from ground-based rainfall measurement. Khan et al. (2012) calibrated a distributed hydrologic model driven by the quasi-real-time Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis product (3B42) (TRMM-3B42RT) (Huffman, et al., 2007) with the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E)-detected water surface signal over a poorly gauged basins in South Arica. Regional flood frequency analysis based on

non-stationary approach used to evaluate flood risks at poorly gauged basins (Leclerc & Ouarda, 2007). Shrestha et al. (2008) calibrated a semi-distributed model and used it to simulate daily streamflows with gauge observed rainfall and satellite rainfall estimates over Bagmati basin in Nepal. They found that even though satellite precipitation product was able to detect rainfall events, the coarse resolution satellite-driven flow simulations underestimated the observed rainfall-driven flow simulations. The study indicated the necessity of high resolution precipitation data for reasonable flood predictions with longer time period.

In this study, a novel framework is developed by integrating multiple satellite remote sensing precipitation products into a distributed hydrologic model for deriving flood frequency distributions for basins within the Connecticut River Basin (CRB). This study evaluates the use of high-resolution satellite precipitation data as a basis to derive flood frequency curves at basin scale. Several quasi global satellite products are investigated in terms of their accuracy to generate annual flow maxima and flood frequency distributions via the hydrologic simulations. This paper is organized as follows: Section 4.2 provides a description of study area, datasets and the hydrologic model. The methodology is presented in section 4.3. Comparison between satellite and reference-driven peak flow simulations as well as the flood frequency results are discussed in section 4.4. Finally, we discuss conclusions and future research directions in section 4.5.

4.2.Study Area, Data and Hydrologic Model

4.2.1. Study Area

The study area is the Connecticut River Basin (CRB) located in the northeast United States (Figure 4.1). CRB is a major river basin in New England region and discharges to the Connecticut River. The river starts in Canada and flows for about 660 km through four states of the USA (New Hampshire, NH; Vermont, VT; Massachusetts, MA; and Connecticut, CT). The total area of the CRB is approximately 28,500 km^2, and 390 towns and cities are located within it. Approximately, 79% is covered by forested, 11% by agriculture and the remaining area is covered by residential and water (Marshall & Randhir, 2008).

In Figure 4.1 blue lines and light gray rectangular represent the river network and 0.25^o satellite grids, respectively. The study area is divided into nine sub-basins with drainage areas ranging from 190 to 25,019 km^2. The highest peak (1915m M.S.L., Mean Sea Level) is in the northeastern corner and the lowest point (10m M.S.L.) is in the southeast with general elevation gradient dropping from north to southeast.

In New England region, cold season is defined during November through April where the maximum snow occurs in January and the amount typically decreases through April (Marshall & Randhir, 2008; Lins, 1997). This definition is similar with the one used in Hodgkins et al. (2003). Therefore, in this study, the time period from May to November was defined as the warm season where flooding is considered to be driven by rain storms.

In the past decade, numerous severe storms occurred in the New England region during June through October months including Storm Bonnie (August, 2004), Hurricane Jeanne (September, 2004), Hurricane Katrina (August, 2005), Tropical Storm Barry (June, 2007), Hurricane Kyle

(September, 2008), Hurricane Irene (August, 2011) and Hurricane Sandy (October 2012). Hurricane Irene was one of the most devastating events causing $6.5 billion of estimated flood damages in the region (National Oceanic and Atmospheric Administration, 2012). In addition, flooding and power outages from Hurricane Sandy (October, 2012) lasted several days effecting roughly nine million people, and causing an estimated $50 Billion in damage in the US (Blake, et al., 2013). Overall, heavy rains in the study area result in extensive flooding and cause significant damages.

4.2.2. Data

The National Center for Environmental Prediction WSR-88D Stage IV precipitation dataset (Lin & Mitchell, 2005) (4-km/hourly spatiotemporal resolution) is used as the reference precipitation (hereafter named Radar). Stage IV combines precipitation estimations from a network of 150 Doppler Next Generation Weather Radars (NEXRAD) with gauge precipitation measurements (approximately 5500 rain gauges) over the Contiguous United States (CONUS) (Lopez, 2011). The accumulated precipitation estimates (hourly) for each radar is combined with other radars in the WSR-88D network using an inverse-distance weighting (IDW) scheme. The data benefit from 12 CONUS River Forecast Centers (RFCs) manual quality control (QC) applying multiple bias adjustments on the gauge observations (Lin & Mitchell, 2005).

Four quasi-global satellite products at 0.25^{o} spatial resolution and 3-hour temporal resolution are evaluated in this study. The quasi-real-time Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis product (3B42) (here after named TRMM-3B42RT), which uses microwave (MW)-calibrated infrared (IR) precipitation estimates to fill in MW coverage gaps to provide precipitation estimates at 3-hourly intervals. A monthly gauge-adjusted version

of the TRMM-3B42RT available with one month latency (named TRMM-3B42V7) is also used (Huffman, et al., 2007). Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN) is a MW-calibrated infrared retrieval technique. The precipitation product uses a neural network approach to derive relationships between IR and various MW collections, and then estimations are generated directly from IR data (Sorooshian, et al., 2000). The Climate Prediction Center Morphing Technique (CMORPH) is the fourth product developed by the National Oceanic and Atmospheric Administration (NOAA). The algorithm uses multi-satellite based MW precipitation estimates integrated based on spatiotemporal cloud motion fields derived from IR observations to estimate 3-hourly precipitation over the globe (Joyce, et al., 2004).

Potential evapotranspiration (PET) data based on the North American Regional Reanalysis (NARR) available at 32 km spatial and 3-hourly temporal resolution (Mesinger, et al., 2005) were used as forcing variable for the hydrological model.

The aforementioned precipitation and PET datasets were used to force a distributed-hydrologic model for the period 2002 to 2012, based on which we simulated streamflows for the different basins of this study. The resulting runoff simulations were evaluated at two temporal resolutions (3-hourly and daily) to investigate time resolution effect. The general assumption of the flood frequency analysis in our study is that the peak flows are generated by homogenous rainfall-runoff process (i.e. generated purely from rainfall). Previous studies have highlighted the fact that avoiding different hydrological processes (by generated rainfall-driven, mixed rainfall/snowmelt-driven and snowmelt-driven floods) is important to meet the flood homogeneity assumption (Leclerc & Ouarda, 2007; Javelle, et al., 2003). Therefore this study focuses on peak runoff generated by rainfall during the warm season (May-Nov).

4.2.3. Hydrologic Model

The Coupled Routing and Excess Storage (CREST) model is the modeling framework implemented in this study. The CREST model is developed by the University of Oklahoma and NASA SERVIR (www.servir.net) (Wang, et al., 2011). The distributed CREST model is a hybrid modeling, devoted to generate rainfall-runoff simulations over flexible spatiotemporal scales. It provides important advantages over existing models under different land cover and soil type scenarios with user defined spatiotemporal resolutions. The modeled processes include rainfall-runoff generation and capacity for cell-to-cell routing, canopy interception, infiltration, evaporation, recharge base flow, sub-grid cell variability of soil moisture storage capacity and routing processes at the sub-grid scale. The model controls the maximum storage of the infiltrating water and yield surface runoff generation with connected layers within the soil profile. Cell-to-cell flow routing of direct surface runoff is applied using a kinematic wave assumption. Besides, coupling between the flow simulation and routing component via feedback mechanisms provides realistic applications of the hydrologic variables (i.e. soil moisture) (Wang, et al., 2011; Khan, et al., 2011).

With 14 calibration parameters, CREST is adaptable globally based on geomorphologic, land cover and soil type datasets. The parameters can be calibrated with an automatic global optimization algorithm (Differential Evolution Adaptive Metropolis, DREAM). The algorithm runs to mitigate sum of squared residuals to estimate posterior parameters using multiple Markov Chain Monte Carlo chains from prior iterations (Vrugt, et al., 2009). Dis et al. (2015) provides details about the model implementation in the CRB and reported model specifications, parameter calibration and performance results. They evaluated the model performance with a calibration of 3-year period (from 2005 to 2007). Four model performance metrics are reported: including the

Nash-Sutchliffe Coefficient of Efficiency (NSCE) ranged from 0.31 to 0.68; Mean Relative Error (MRE) varied between -0.06 and 0.13; Centered Root Mean Square Error (CRMSE) varied between 61 and 121 (%); and the Pearson Correlation Coefficient (PCC) ranged between 0.60 and 0.83.

4.3.Methodology

4.3.1. Flood Frequency Analysis

The log-Pearson Type III statistical distribution technique is the standard method to predict the design flood for a river and it is also recommended in Bulletin #17B (Ahearn, 2003; U.S. Water Resources Council, 1981). The three statistical moments used to describe the distribution are the mean, standard deviation, and skew coefficient. Chow (1951) has shown that many probability distributions can be generalized to Eq. (4.1) in frequency analysis. Flow rates (Q, in m^3/s) can be computed for a given return periods as:

$$\log Q = \overline{Q_l} + K_T S_l \tag{4.1}$$

where $\overline{Q_l}$ is the mean of the (base 10) logarithms of the annual peak discharges. K_T is the frequency factor of annual exceedance probability for return period (T) and it is tabulated for log-Pearson Type III with skew coefficient (C_S) in Table 7.7 of Haan (1977) while S_l is the standard deviation of logarithms of the annual peak discharges; C_S can be calculated with Eq.(4.2):

$$C_S = \frac{n \sum_{i=1}^{n} \left(Q_{l_i} - \overline{Q_l}\right)^3}{(n-1)(n-2)\, {S_l}^3} \tag{4.2}$$

where i and n are time steps and the number of values in the datasets, respectively. After calculating the flood magnitude with given recurrence interval, data is tested for high (Q_{High}) and low (Q_{Low}) outliers. Outliers are defined as data points that depart significantly from the trend of the remaining data (U.S. Water Resources Council, 1981). The modification or deletion of these outliers can significantly affect the magnitude of the data-based statistical parameters (i.e. skew coefficient) in particular for small samples (Chow, et al., 1988). Eq. (4.3) is used to detect high/low outliers (U.S. Water Resources Council, 1981):

$$\log Q_{High(Low)} = \overline{Q_l} \pm K_n S_l \tag{4.3}$$

K_n is outlier test frequency values and tabulated in a table with a given sample size (U.S. Water Resources Council, 1981). The values are used in one-sided tests that detect outliers at the 10-percent level of significance in normally distributed data. If the historic peaks in a sample are greater (smaller) than Q_{High} (Q_{Low}) then they are considered high (low) outliers. High (low) outliers were deleted from the sample and the frequency analysis is calculated based on the U.S. Water Resources Council (WRC) method.

Sample size is important and effective on the skew coefficient used in fitting the Log-Pearson Type III distribution. This coefficient is very sensitive especially for small samples. Thus, U.S. WRC recommended using weighted skew coefficient (C_w) to improve the accuracy of estimated skew coefficient. Eq. (4.4) is used to calculate C_w by combining the computed station skew coefficient (C_S) using the sample data and generalized map skew coefficient (C_m, can be read from the Plate I map in Bulletin #17B of U.S. Water Resources Council, (1981)) (Tasker, 1978):

$$C_w = \frac{V(C_m)C_S + V(C_S)C_m}{V(C_m) + V(C_S)} \tag{4.4}$$

where $V(C_m)$ and $V(C_S)$ represent variances of the associated (generalized and station) skew coefficients. WRC recommended $V(C_m)$ value of 0.302 for the United States when generalized skews are extracted from Plate I. According to the results of Monte Carlo experiments by Wallis et al. (1974), the variance of station skew $V(C_S)$ for Log-Pearson Type III random variables can be approximated with sufficient accuracy as (Chow, et al., 1988):

$$V(C_S) = 10^{A - B \log_{10}(\frac{n}{10})} \tag{4.5}$$

where

$$A = \begin{cases} -0.33 + 0.08|Cs| & \text{if } |Cs| \leq 0.90 \\ -0.52 + 0.30|Cs| & |Cs| > 0.90 \end{cases} \tag{4.6}$$

and

$$B = \begin{cases} 0.94 - 0.26|Cs| & \text{if } |Cs| \leq 1.50 \\ 0.55 & |Cs| > 1.50 \end{cases} \tag{4.7}$$

Statistical estimates are often presented with confidence intervals that depend on the confidence level (β) or significance level (α= (1-β)/2). The 90- and 50-percent confidence limits for the 1.01, 2, 5, 10, 25, and 50-yr return period discharges for CRB were determined as follows:

$$U(L)_{T,\alpha} = \overline{Q_l} \pm K_{T,\alpha}^{U(L)} S_l \tag{4.8}$$

where $K_{T,\alpha}^{U}$ ($K_{T,\alpha}^{L}$) stands for upper (lower) confidence limit factor to calculate upper (lower) limit logarithmic discharge $U_{T,\alpha}$ ($L_{T,\alpha}$). These factors' values can be approximated with Eqs. (4.9) and (4.10) (U.S. Water Resources Council, 1981; Natrella, 1963):

$$K_{T,\alpha}^{U(L)} = \frac{K_T \pm \sqrt{{K_T}^2 - ab}}{a} \tag{4.9}$$

where

$$a = 1 - \frac{{z_\alpha}^2}{2(n-1)} \quad \text{and} \quad b = {K_T}^2 - \frac{{z_\alpha}^2}{n} \tag{4.10}$$

where z_α is the standard normal variable with exceedance probability α.

4.3.2. Error Analysis

Maximum satellite products-driven runoff values over the warm period months are evaluated against the reference (i.e. Stage IV) driven runoff values via Mean Relative Error (MRE), Pearson Correlation Coefficient (PCC), and Centered Root Mean Square Error (CRMSE) statistics. MRE gives indications on overall mean error magnitude and the error direction (under/over-estimation), while CRMSE quantifies the random error. Correlation measures the strength of linear dependency between reference and satellite-driven streamflows. These error metrics are calculated as:

$$\mathrm{MRE} = \frac{\sum_{i=1}^{n}(Q_{s,i} - Q_{r,i})}{\sum_{i=1}^{n} Q_{r,i}} \tag{4.11}$$

$$\mathrm{PCC} = \frac{\sum_{i=1}^{n}(Q_{r,i} - \overline{Q}_r)(Q_{s,i} - \overline{Q}_s)}{\sqrt{\sum_{i=1}^{n}(Q_{r,i} - \overline{Q_r})^2 \sum_{i=1}^{n}(Q_{s,i} - \overline{Q_s})^2}} \tag{4.12}$$

$$\mathrm{CRMSE} = \frac{1}{\overline{Q_r}} \sqrt{\frac{\sum_{i=1}^{n}\left[Q_{r,i} - Q_{s,i} - (\overline{Q_r} - \overline{Q_s})\right]^2}{n}} \times 100\% \tag{4.13}$$

where Q stands for the annual maximum from the simulated flow [m^3/s] time series over the time period of May 1st to Nov 30th. $\overline{Q}$ stands for the mean of the annual maximum flow values for the 11-year data record. The subscripts r and s indicate radar and satellite-driven runoff simulations, respectively.

After deriving the flow rates of different return periods (1.01-yr, 2-yr, 5-yr, 10-yr, 25-yr, and 50-yr) (Figure 4.2), two binary metrics are calculated for flows at two temporal resolutions (3-hourly and daily). These metrics are named Bias Score (BIASs) and Critical Success Index (CSI). To calculate these metrics, a set of runoff thresholds was created for each catchment. Then, the scores were determined based on the following occurrences H, F, M and NH:

H: satellite-driven flow > satellite threshold and radar-driven flow > radar threshold

F: satellite-driven flow > satellite threshold and radar-driven flow < radar threshold

M: satellite-driven flow < satellite threshold and radar-driven flow > radar threshold

NH: satellite-driven flow < satellite threshold and radar-driven flow < radar threshold

BIASs is defined as the ratio of the number of occurrences that estimated flow exceeded a certain threshold relative to the occurrences that the reference flow exceeded the same threshold. The value of one indicates that the number of predicted occurrences is equal to the total number of reference occurrences even though the score of one does not necessarily mean a perfect skill of the estimator. CSI takes into account both false alarms and missed events, and its range varies from 0 to 1, with a value of 1 indicating a perfect detection. The two metrics are given with Eqs. (4.14) and (4.15):

$$\mathrm{BIASs} = \frac{\mathrm{H} + \mathrm{F}}{\mathrm{H} + \mathrm{M}} \tag{4.14}$$

$$\mathrm{CSI} = \frac{\mathrm{H}}{\mathrm{H} + \mathrm{F} + \mathrm{M}} \tag{4.15}$$

The contingency table statistics are sensitive to data sample misbalance. Imbalanced data in this study are corrected following the balanced sampling technique (Chen, et al., 2004). Namely, multiple random iterations were computed to create 10,000 balanced samples that represent

equal number of hits and non-hits. The mean of BIASs and CSI values from the 10,000 iterations are reported for the different catchments of the study.

4.4.Results

To assess the correspondence between satellite-driven and radar-driven streamflow annual maxima we reported the MRE, PCC, and CRMSE statistics for the 3-hourly (Table 4.1) and daily (Table 4.2) temporal resolutions. The unadjusted (satellite-only) satellite products-driven runoff peak flow values show in general underestimation, while peak flow simulations from TRMM-3B42V7 (i.e. the gauge-adjusted satellite product) yield mostly overestimation of the radar-driven peak flow values. B5 and B8 basins consistently demonstrated negative MRE values regardless of product and temporal resolution. The MRE metric is shown to improve for the daily scale relative to the 3-hourly scale flows. In addition, the PCC values are reported in the second column of the tables. The results show that the highest PCC value between annual peak flows (0.97) is obtained in the TRMM-3B42V7-driven simulation. High PCC values are also obtained for the TRMM-3B42RT and CMORPH-driven maximum streamflow cases with some degradation. The lowest PCC values are mostly found for the peak flows of the PERSIANN-driven runoff simulations. Temporal resolution did not affect the correlation metric. CRMSE is reported in the last column of the table as a percentage. The values range from 23% to 68% for TRMM-3B42V7, 24% to 59% for TRMM-3B42RT, 23% to 60% for PERSIANN, and 31% to 136% for CMORH-driven annual maximum runoff values. The highest (lowest) CRMSE values are depicted in B2 (B4) basin regardless of satellite product and temporal resolution. Similar to PCC values, TRMM-3B42V7 exhibits the best consistency in terms of the CRMSE values.

The flood frequency curve plots of the simulated runoff from the five datasets (i.e. Radar, TRMM-3B42V7, TRMM-3B42RT, PERSIANN, and CMORPH) are presented against different return periods for B6 basin (Figure 4.3). The flood frequency curves are reported for the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} quantiles, while the Universal Weibull formula (Weibull, 1939) is used for plotting the position of annual peaks. As it can be seen from the figure, the Stage IV dataset and gauge-adjusted satellite product (TRMM-3B42V7) exhibit the best agreement between the fitted frequency curves and the Universal Weibull formula's annual runoff maxima frequency values. However, the unadjusted satellite-driven maxima spread over the 5% and 95% percentile range; one peak flow value is above the 95% percentile in the CMORPH-driven runoff flood frequency. That specific peak is excluded by the outlier test and is not considered in the frequency analysis.

To assess all the satellite derived frequency curves against the reference one, box plots based on the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles for the 1.01-yr, 2-yr, 5-yr, 10-yr, 25-yr, and 50-yr return period are presented over the study basins at 3-hourly (Figure 4.4) and daily (Figure 4.5) temporal scales. A main point to note is that each basin shows unique frequency curves and the curves of using the satellite products-driven runoff simulations mostly captures the radar-driven simulations at the 25^{th}-75^{th} percentile envelope. A basin-wide comparison shows that the satellite simulations for B5 basin have stronger bias relative to the reference regardless of satellite product, temporal resolution, and exceedance probability (an exception exists at 1.01-yr return period). In addition, the magnitude of underestimation is the highest in the 50-yr return period with the CMORPH-derived flood frequency. Normalized flow values for daily temporal resolution are smaller than those for the 3-hourly temporal scale. The variability in flow

magnitudes are narrower in daily scale compared to the 3-hourly, which is particularly recognized for the B2 and B8 basins.

The magnitude of overestimation is generally the strongest in TRMM-3B42V7 generated flood frequencies over the B1, B2, and B6, basins. However, the skill of capturing the radar derived flood frequency for this product overlaps within the 25^{th}-75^{th} quantile range in B3, B4 and B7 basins especially for the higher return periods. Underestimation is exhibited in B8 and B9 basins for this product. TRMM-3B42RT exhibits better consistency than the TRMM-3B42V7 for B1, B2, B3, B6, and B7 basins while the gauge-adjusted counterpart (TRMM-3B42V7) shows higher degree of agreement compared to the TRMM-3B42RT for the B4 and B5. The magnitude of the underestimation is the strongest in the B8 and B9 basins at the lowest exceedance probability for this product. The CMORPH-driven simulations underestimated flood frequencies in higher exceedance probabilities for all the basins and the performance consistently improved with decreasing exceedance probabilities. CMORPH flow rates was found to be the best product for capturing significantly the value range of radar-driven flow rates at higher return periods (i.e. 25-yr, 50-yr) over all basins. Finally, the performance of PERSIANN is generally in between TRMM-3B42RT and CMORPH-derived flood frequencies with the exception of B2 and B3 basins where the magnitude of the underestimation is the strongest for the product at the higher return periods (i.e. 10-yr, 25-yr, and 50-yr).

The assessment of satellite precipitation products in deriving flood frequencies is evaluated based on the contingency error metrics at 3-hourly and daily temporal resolutions for the study catchments and three return period thresholds: 1.01-yr, 2-yr, and 5-yr (very few data points exist beyond the 5-year return period). Figure 4.6 shows sample streamflow time series associated with a short time period over B1 basin. This figure is used as an example to demonstrate the

occurrences of the forecast and reference used to derive the contingency table statistics. Each satellite-driven runoff dataset has different threshold values associated with the 1.01-yr, 2-yr, and 5-yr return periods. If both (reference and satellite) of the flow values are greater (smaller) than their thresholds, it is counted as hit (correctly forecasted non-hit). If the satellite exceeds its own threshold but reference does not, it is reported as false detection, and the reverse condition is counted as missing. These four groups of detections are visualized in the figure.

The two binary error metrics Bias Score (BIASs) and Critical Success Index (CSI) are reported in Figure 4.7 and Figure 4.8, respectively. Mean values of the BIASs for each basin are illustrated in Figure 4.7. A first observation from Figure 4.7 is that BIASs values are more spread in the 3-hourly scale compared to the daily scale for all three thresholds. Furthermore, zero score is obtained from some of the basins at the 5-yr return period threshold. The highest scores were retrieved in PERSIANN at all thresholds and two temporal scales, while the score of CMORPH is the lowest. The CSI (generated by the balanced sampling procedure) is demonstrated in Figure 4.8. It is noted that CSI-based skill of the satellite-driven flow detection diminishes with increase of the flow thresholds. In the case of the 5-yr return period flow threshold, CSI values become zero for some of the basins (i.e. no predictive skill). The PERSIANN product outperformed the other satellite products at both temporal scales. The gauge adjusted TRMM-3B42V7 product displayed higher CSI values relative to the TRMM-3B42RT and CMORPH products. Additionally, the CSI values at the 3-hourly time scale are lower than those determined for the daily scale flows.

4.5. Conclusions

The focus of this study was to understand the potential of using satellite precipitation products in estimating the flood frequencies of selected basins in the Connecticut River Basin. The proposed analysis framework assessed satellite data in flood frequency applications, and evaluated how satellite precipitation products can be used in predicting event severities, which is the standard approach of early flood warning systems.

Four satellite precipitation products, TRMM-3B42V7, TRMM-3B42RT, CMORPH, and PERSIANN were investigated by deriving flood frequencies at 3-hourly and daily temporal resolutions using the Coupled Routing and Excess Storage distributed hydrologic model. Gridded radar-rainfall from the WSR-88D stage IV product and corresponding flow simulations were used as reference. The evaluation was reported for nine catchments of the Connecticut River Basin.

The log-Pearson type III distribution was applied on the annual maxima flow values of each precipitation dataset-driven simulations to derive the flood frequencies. The skew coefficient for the distribution was improved using the generalized skew coefficient, and flood frequency outliers were adjusted based on WRC method. Flood frequency values are compared for the 1.01-yr, 2-yr, 5-yr, 10-yr, 25-yr, and 50-yr return periods and two temporal scales of aggregation (3-hourly and daily) over the different basin scales (190 to 25,000 km^2). The error statistics (BIASs and CSI) for the evaluation of the event severity predictions were illustrated for 1.01-yr, 2-yr, and 5-yr return period flows.

MRE values illustrated underestimation of annual peak flows from the unadjusted satellite product-driven runoff simulations, while simulations from TRMM-3B42V7 gave slight overestimation of the peak flow to the reference values. The best PCC of the estimated annual

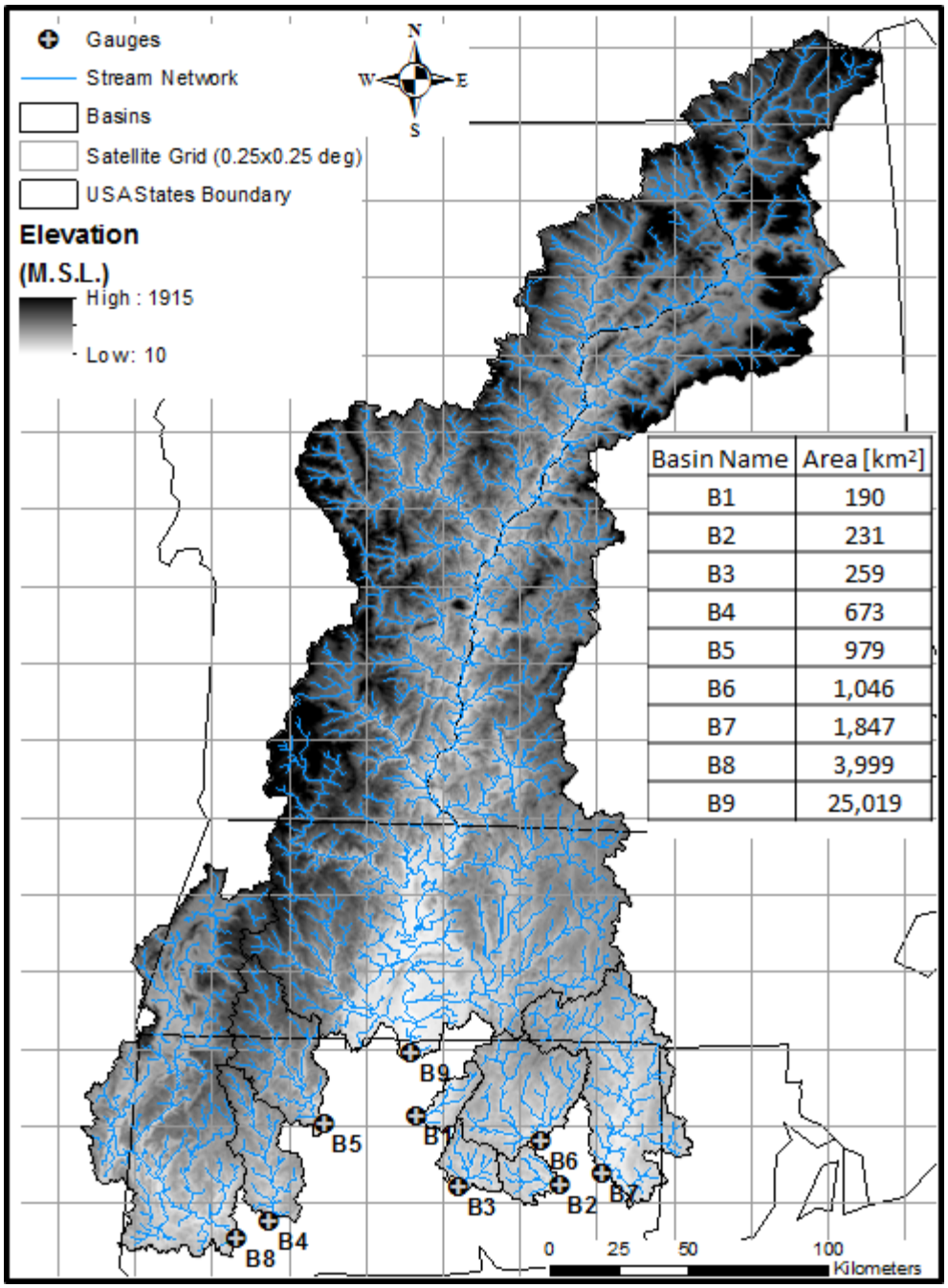
maxima flows was obtained with the TRMM-3B42V7-driven simulations, and ranged between 0.53 and 0.97 for the different basins examined in this study. Correlation values from the TRMM-3B42RT and CMORPH-driven simulations were lower than those of TRMM-3B42V7, while PERSIANN exhibited the lowest. The minimum (maximum) CRMSE values evaluated for TRMM-3B42V7, TRMM-3B42RT, PERSIANN, and CMORPH based simulated annual maximum records were 23% (68%), 24% (59%), 23% (60%), and 31% (136%), respectively.

Overall, the study demonstrated a good skill of satellite-driven flood frequencies relative to flood frequencies derived based on the Stage IV precipitation data. Each basin exhibited unique frequency curves in terms of the return period flows. Satellite derived flood frequency curves mostly encapsulated the Stage IV-derived return period flows based on their 25^{th}-75^{th} quantile ranges. Flood frequency curves derived from daily flow simulations showed better consistency between satellite and Stage IV precipitation data to those derived from 3-hourly flow simulations.

The CSI and BIASs statistical scores identified PERSIANN (CMORPH) satellite product having the best (worst) performance in assessing event severity based on the flood frequency thresholds. The gauge adjusted TRMM-3B42V7 product yielded higher values compare to TRMM-3B42RT. Finally, the CSI values at daily time scale were higher than those determined for the 3-hourly flow simulations.

Future extensions of this study will aim at extending data length to reduce the prediction uncertainty of satellite precipitation-derived flood frequencies. Such work could also include other gauge-adjusted satellite products in the flood frequency analysis to investigate the gauge adjustment effect over different gauge-density regions. Another subject for future research is application of satellite precipitation products in flood frequency studies of other hydro-climatic

regimes (representing different flow generation processes). Generally, results from this study combined with the increasing record length of satellite precipitation products from the Global Precipitation Measurement mission (Hou, et al., 2014) can motivate future studies on global flood frequency analysis driven by remotely sensed precipitation.



Basin Name	Area [km²]
B1	190
B2	231
B3	259
B4	673
B5	979
B6	1,046
B7	1,847
B8	3,999
B9	25,019

Figure 4.1. Map of Connecticut River Basin (CRB) and the location of USGS streamflow gauges with their coverage area.

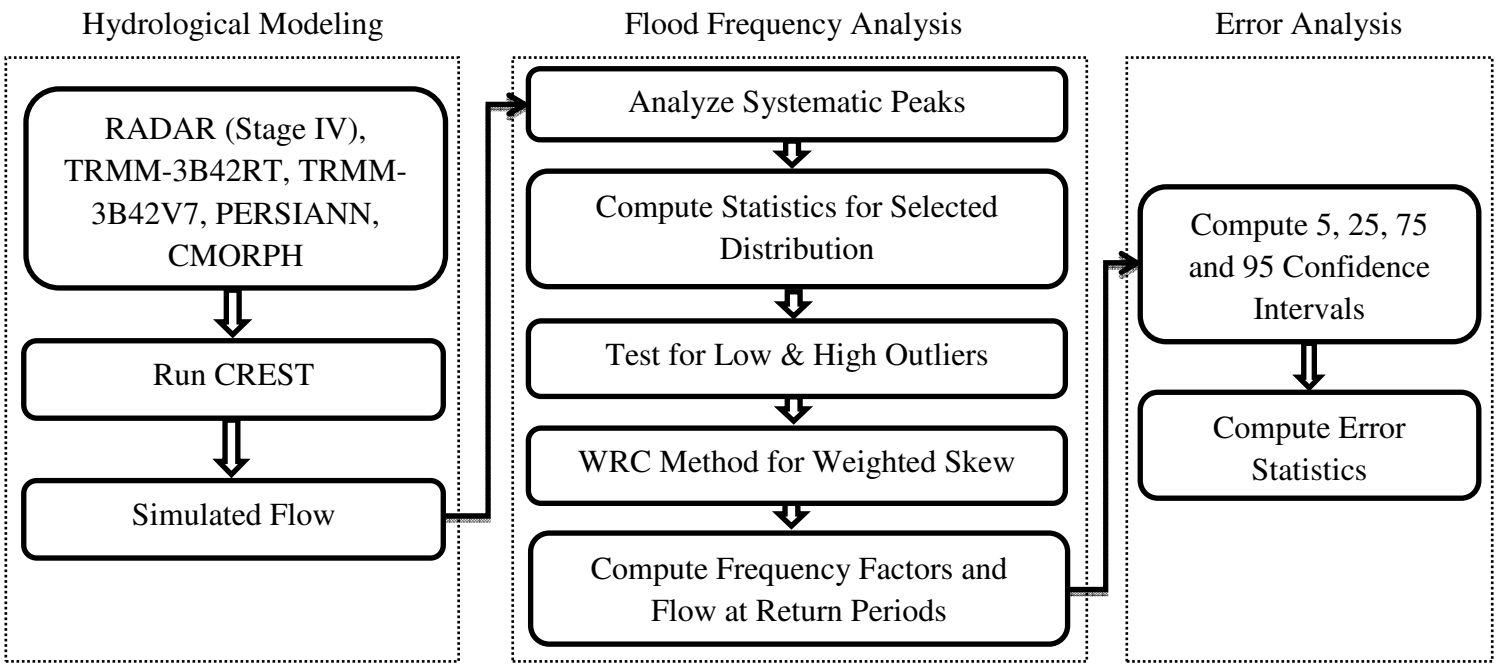


Figure 4.2. Flow diagram of the flood frequency analysis methodology

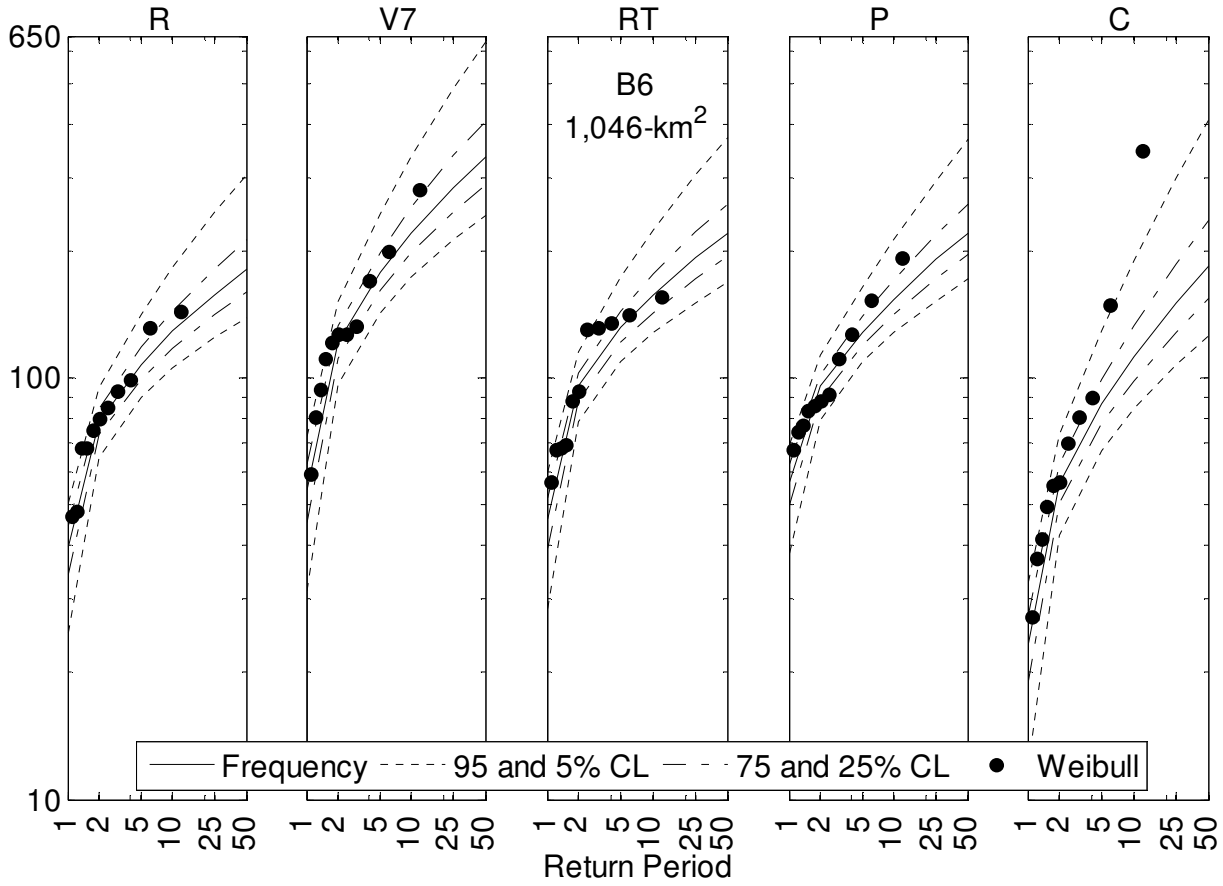


Figure 4.3. Flood frequency curve (mean and 5%, 25%, 75%, and 95% confidence intervals) for B6 (1,046-km^2) basin, Weibull formula, (number of years+1)/rank), used for plotting the annual peaks in the frequency graphs

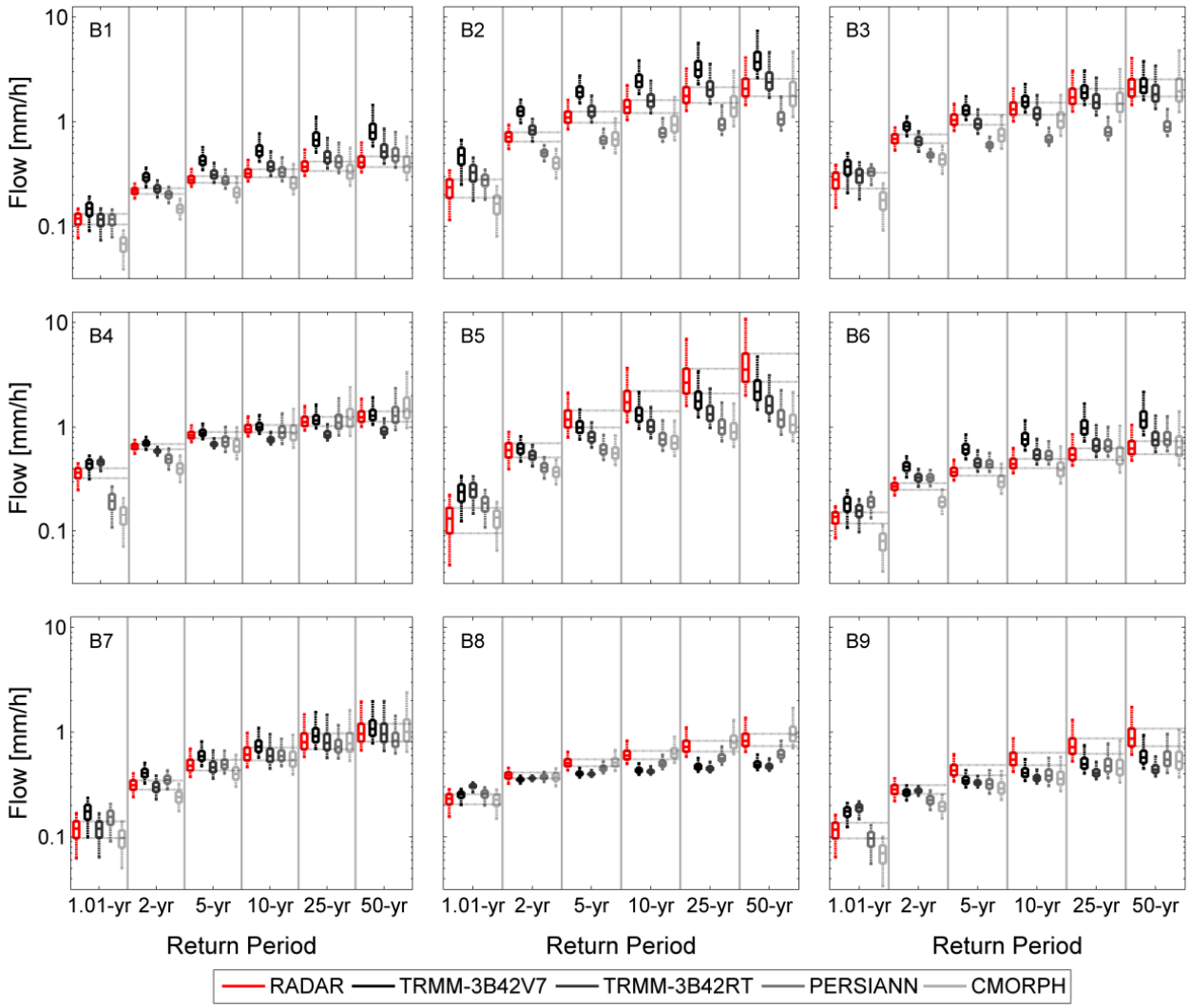


Figure 4.4. Box plots of Flood Frequency Curves derived from the different satellite products and reference (Stage IV) driven flow simulations at 3-hourly temporal resolution.

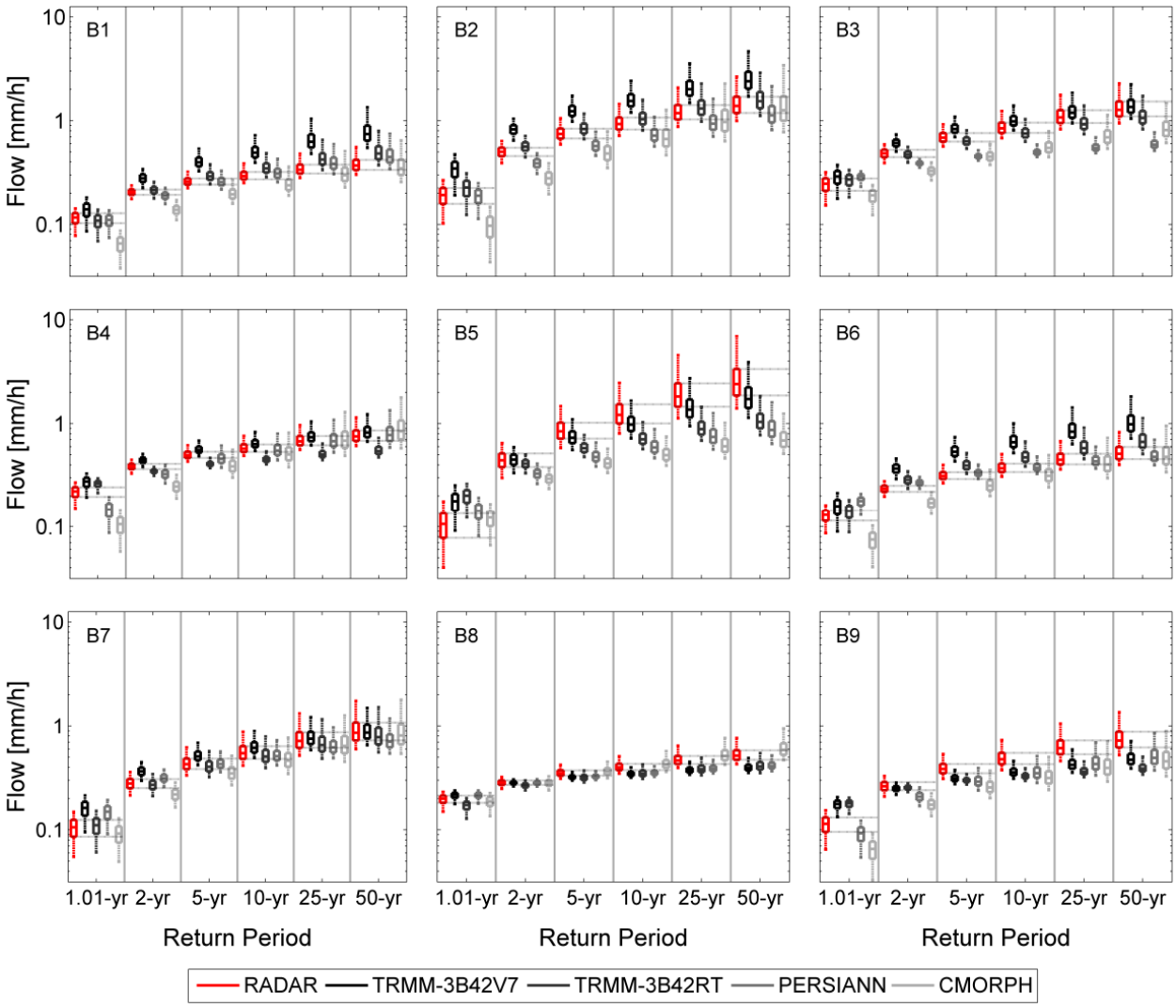


Figure 4.5. Same as in Figure 4.4, but for the daily temporal resolution

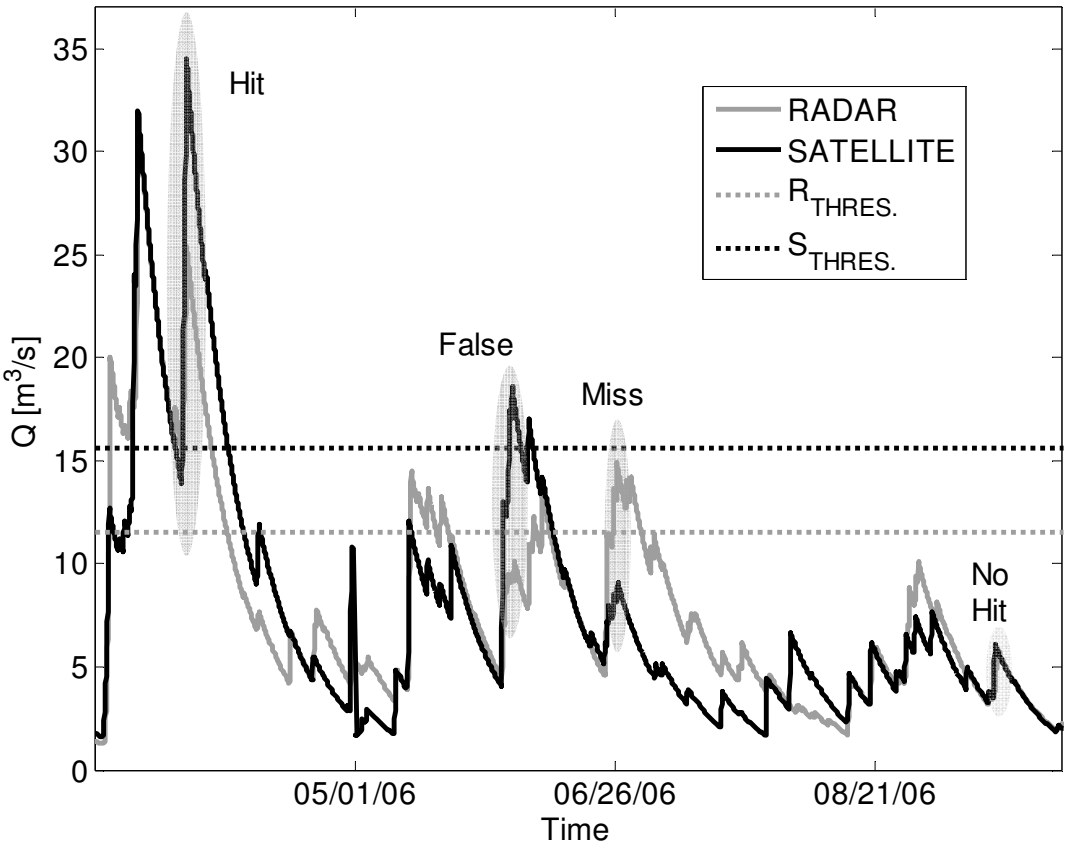


Figure 4.6. Sample streamflow time series for B1 basin (190-km^2) showing hits (H), misses (M), false alarm (F), and correctly classified none occurrences (NH) for dichotomous satellite and reference (radar) driven runoff.

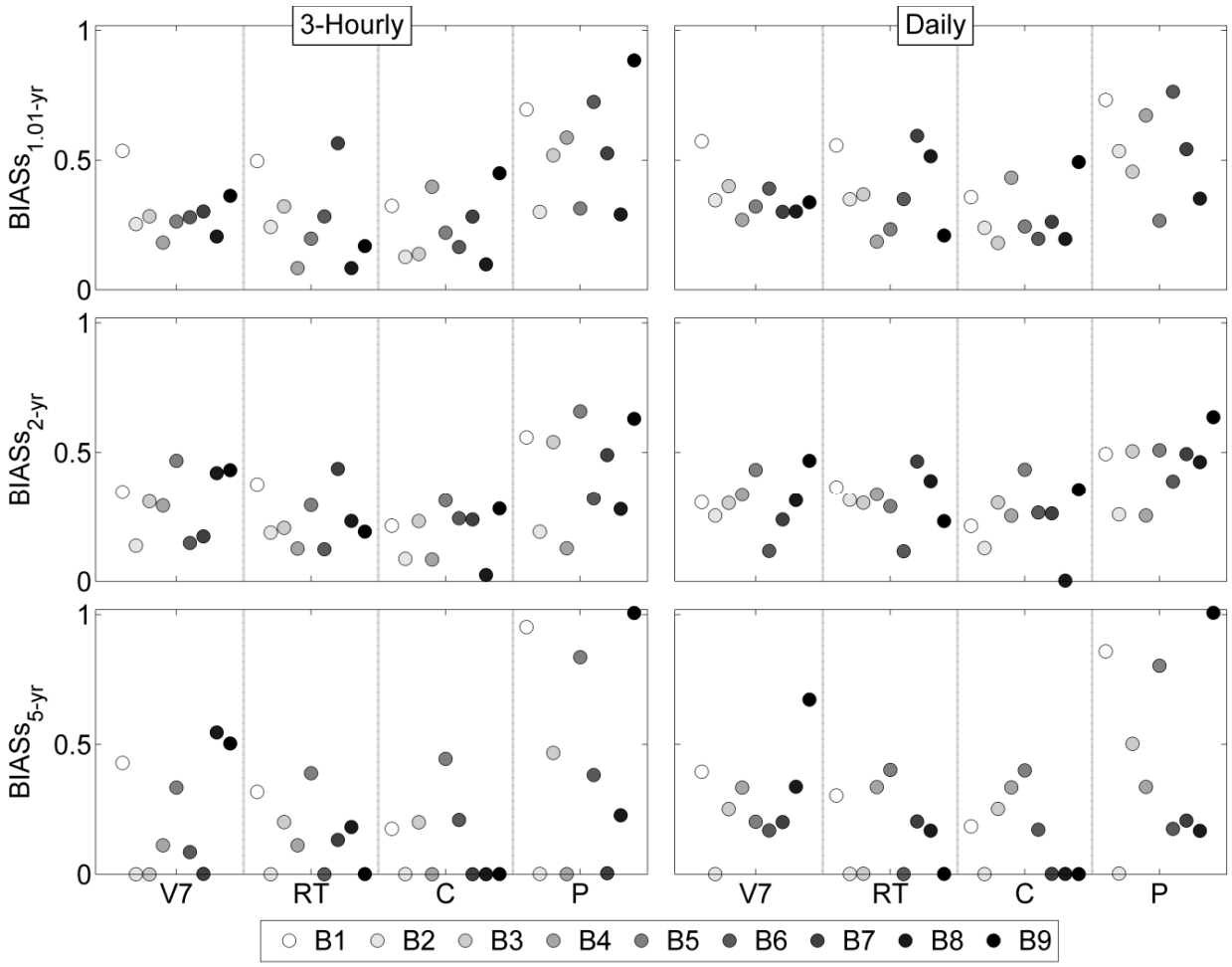


Figure 4.7. Bias score for thresholds of different return periods at the two temporal resolutions

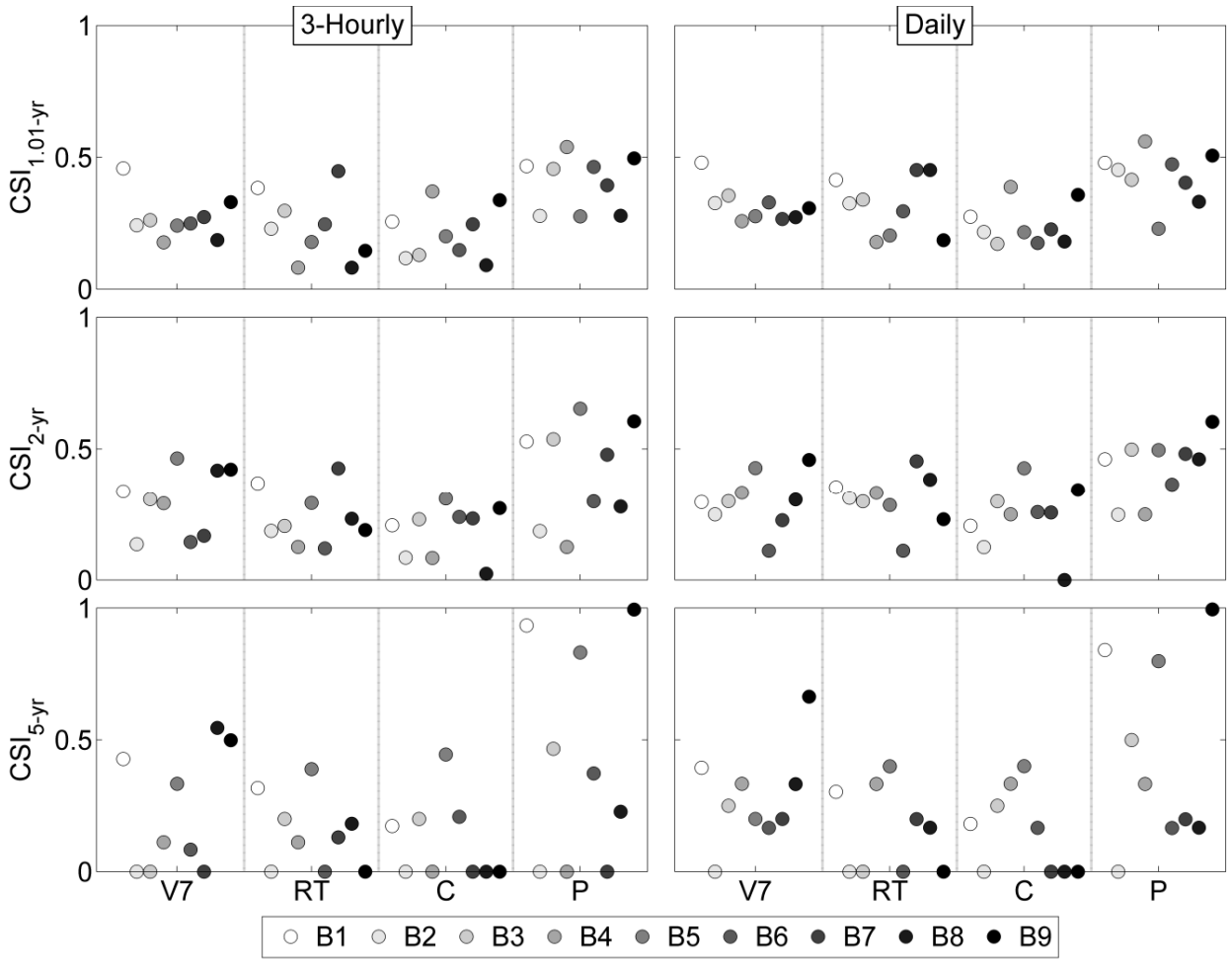


Figure 4.8. Same as in Figure 4.7, but for the Critical Success Index (CSI)

Table 4.1. MRE, PCC, and CRMSE statistics calculated for annual maxima attained from 3-hourly temporal resolution simulated streamflow.

Basin Name	MRE				PCC				CRMSE [%]			
	V7	RT	P	C	V7	RT	P	C	V7	RT	P	C
B1	0.27	-0.05	-0.15	-0.05	0.90	0.78	0.46	0.91	23	29	42	67
B2	0.79	0.18	-0.23	0.07	0.53	0.70	0.17	0.27	68	35	48	136
B3	0.25	-0.06	-0.27	-0.28	0.66	0.49	0.29	0.61	35	42	45	41
B4	0.02	-0.15	-0.20	-0.30	0.70	0.52	0.65	0.55	23	24	23	32
B5	-0.09	-0.24	-0.43	-0.17	0.97	0.56	0.79	0.70	28	59	54	74
B6	0.60	0.21	0.23	0.07	0.53	0.19	0.37	0.33	58	47	44	96
B7	0.26	-0.03	0.06	-0.16	0.66	0.14	0.00	0.28	39	59	57	59
B8	-0.16	-0.25	-0.23	-0.22	0.91	0.90	0.23	0.91	28	45	60	31
B9	-0.01	-0.13	-0.25	-0.08	0.78	0.39	0.26	0.90	32	46	50	47

Table 4.2. Same as in Table 4.1, but for daily temporal resolution.

Basin Name	MRE				PCC				CRMSE [%]			
	V7	RT	P	C	V7	RT	P	C	V7	RT	P	C
B1	0.28	-0.05	-0.14	-0.05	0.90	0.80	0.46	0.91	23	27	41	68
B2	0.68	0.13	-0.20	-0.01	0.45	0.53	0.04	0.22	69	41	53	117
B3	0.21	-0.05	-0.18	-0.21	0.45	0.23	0.09	0.47	41	45	47	43
B4	0.08	-0.16	-0.13	-0.28	0.61	0.42	0.71	0.58	27	26	22	31
B5	-0.07	-0.23	-0.38	-0.10	0.90	0.35	0.74	0.78	31	65	53	75
B6	0.62	0.23	0.22	0.08	0.52	0.05	0.19	0.20	57	51	47	105
B7	0.24	-0.04	0.06	-0.17	0.69	0.17	0.01	0.29	35	54	56	54
B8	-0.06	-0.18	-0.19	-0.20	0.92	0.79	0.07	0.83	22	39	55	34
B9	0.00	-0.11	-0.24	-0.11	0.81	0.44	0.30	0.89	26	40	44	47

5. Conclusion remarks

The major conclusions for this study are summarized below:

- During calibration period, NSCE, MRE, RMSE, and PCC values of hourly flow simulations varied from 0.31 to 0.68, -0.06 to 0.13, 61 to 121 (mm) and 0.60 to 0.83, respectively. At daily time scale the performance metrics exhibited improved values indicating that CREST has sufficient accuracy for long term hydrologic simulations of the catchment.
- Based on the error metrics and graphical comparisons, CREST is considered to provide accurate rainfall-runoff simulations performance at high temporal resolution ungauged/poorly gauged catchments over CRB.
- Performance of TRMM-3B42V7 product, which includes gauge adjustment, is the most accurate in terms of capturing the high percentile values of precipitation and runoff over the study basins. The satellite-alone CMORPH and PERSIANN precipitation products consistently underestimated the radar precipitation and reference-driven streamflow simulations and this suggests that a bias adjustment of the products would potentially improve the representation of basin rainfall estimates
- PERSIANN exhibited more frequent false precipitation detection than the CMORPH and TRMM-3B42V7 products, and CMORPH is dominant for missed rainfall volume rates. In addition, we noted reduction in random error moving from 3-hourly to daily temporal resolution for both cold and warm season basin-average precipitation time series.
- CMORPH-based runoff simulations considerably underestimated in both seasons, which is attributed to the underestimation exhibited in the precipitation error analysis.

Underestimation tendency of CMORPH becomes more apparent for the high flow values (> 90^{th} percentile).

- The study demonstrated a basin scale dependency of the satellite-precipitation error metrics and the precipitation-to-runoff simulation propagation properties of these errors. It is observed that the skill of modeling the streamflow increases with basin scale. The error propagation was calculated as the ratio between runoff error metrics (MRE and CRMSE) to the associated precipitation error metrics. Similar to the basin dependency of runoff CRMSE, error dampening from precipitation to runoff is shown to increase proportionally to basin size for all satellite products.
- It is confirmed by flood frequency results that the TRMM-3B42V7 product-driven annual peak flows are characterized with the highest correlation, and has a range from 0.53 to 0.9. Correlation values from TRMM-3B42RT and CMORPH-driven annual maximum runoff values are lower than those from the TRMM-3B42V7 product while PERSIANN is the lowest. The minimum (maximum) CRMSE is 23% (68%), 24% (59%), 23% (60%), and 31% (136%) for TRMM-3B42V7, TRMM-3B42RT, PERSIANN, and CMORH-simulated annual maximum records.
- MRE values illustrated underestimation from the unadjusted satellite products-driven annual maximum runoff relative to reference while flood simulations from TRMM-3B42V7 slightly overestimated the reference in general. The annual maximum values estimated from daily resolution show better consistency to those from the 3-hourly resolution.
- The study showed a good skill of satellite-driven flood frequency relative to flood frequencies derived based on the Stage IV-driven flow simulations. Each basin shows

unique frequency curves that mostly capture the radar-driven simulations within the range of uncertainty 25th and 75th quantiles of the reference.

- The CSI and BIASs results showed that PERSIANN scores are the highest of the three satellite products, while CMORPH gave the lowest scores for both temporal scales. The gauge adjusted TRMM-3B42V7 product yielded higher scores compared to TRMM-3B42RT. Finally, the CSI values at daily time scale are higher than those determined at the 3-hourly scale.
- Based on the results from the study, assessment of both flood magnitudes and their return periods from satellite precipitation data can provide a more complete description of basins' flood regime.

The CREST modeling results of the first study can be useful for future work such as comparing to flash flood predictions, and future climatic analysis of basins at regional and national scales. For this purpose future studies should investigate ways to include snow process in the model including additional surface characteristics (e.g. vegetation cover, solar radiation, snowpack, temperature, wind speed) and consider for changing land surface conditions.

The second study presented in chapter 3, evaluated the satellite precipitation products in hydrologic simulations of northern latitude river basin. While results from this study are based on a single basin and a specific hydrologic model (i.e. CREST), many of the findings can generally apply. Overall, the conclusions of this work provide useful information about the application of satellite precipitation estimates for runoff simulations and evaluation of flood risk with a distributed model over the study area. This study was limited to one gauge adjusted satellite product, while further research is required to enhance hydrologic applications, such as utilizing various gauge-adjusted satellite precipitation products for water resource assessment,

and evaluate how accurate the current products are for issuing precipitation or runoff threshold based early warnings of floods. In the Global Precipitation Measurement (Hou, et al., 2014) era precipitation products can ultimately be used to estimate future floods and for optimal water resources management, such as designing flood related structures and regulation of flood plains.

The framework developed in the last study aimed at understanding the accuracy of estimating flood frequencies and assessing event severities based on satellite precipitation-driven hydrologic simulations. Future studies related to this work will focus on extending data length to reduce the prediction uncertainty of satellite precipitation-derived flood frequencies and include other satellite precipitation products in the flood frequency analysis. Another subject for future research is evaluation of satellite precipitation products in flood frequency analysis of other hydroclimatic regions (representing different flow generation processes). Generally, results from this study combined with the increasing record length of satellite precipitation products from the Global Precipitation Measurement mission (Hou et al. 2014) can motivate such future studies on flood frequency analysis of data poor basins driven by remotely sensed precipitation.

References

AghaKouchak, A. et al., 2011. Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research,* 116(D2), p. DO2115.

Ahearn, E., 2003. *Peak-Flow Frequency Estimates for U.S. Geological Survey Streamflow-Gaging Stations in Connecticut,* East Hartford, CT: U.S. Geological Survey Water-Resources Investigations Report 03-4196.

Ahearn, E. A., 2004. *Regression equations for estimating flood flows for the 2-,10-, 25-, 100-, and 500-year recurrence intervals in Connecticut,* Virginia: U.S. Geological Survey .

Ahearn, E. A., 2010. *Regional regression equations to estimate flow-duration statistics at ungaged stream sites in Connecticut,* Virginia: U.S. Geological Survey.

Antigha, R. & Ogarekpe, N., 2013. Development of intensity duration frequency curves for Calabar Metropolis, South-South, Nigeria. *The International Journal of Engineering And Science,* pp. 39-42.

Apip, et al., 2010. A distributed hydrological–geotechnical model using satellite-derived rainfall estimates for shallow landslide prediction system at a catchment scale. *Landslides,* 7(3), pp. 237-258.

Aronica, G. & Candela, A., 2007. Derivation of flood frequency curves in poorly gauged Mediterranean catchments using a simple stochastic hydrological rainfall-runoff model. *Journal of Hydrology,* 347(1-2), pp. 132-142.

Artan, G. et al., 2007. Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards,* 43(2), pp. 167-185.

Awadallah, A. G., ElGamal, M., ElMostafa, A. & ElBadry, H., 2011. Developing intensity-duration-frequency curves in scarce data region: An approach using regional analysis and satellite data. *Engineering,* 3(3), pp. 215-226.

Bastiaanssen, W., 1998. *Remote Sensing in Water Resources Management: The State of the Art,* Colombo, Sri Lanka: International Water Management Institute.

Behrangi, A. et al., 2014. Satellite-based precipitation estimation and its application for streamflow prediction over mountainous western U.S. basins. *Journal of Applied meteorology and climatology,* Volume 53, pp. 2823-2842.

Beighley, R. E. et al., 2011. Comparing satellite derived precipitation datasets using the Hillslope River Routing (HRR) model in the Congo River Basin. *Hydrological Processes,* 25(20), pp. 3216-3229.

Bergstrom, S., 1991. Principles and confidence in hydrological modeling. *Nordic Hydrology,* Volume 22, pp. 123-136.

Bernard, M., 1932. Formulas for rainfall intensities of long duration. *Transactions,* pp. 592-624.

Beven, K., 1989. Changing ideas in hydrology-The case of physically-based models. *Journal of Hydrology,* 105(1-2), pp. 157-172.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. *Advances in Water Resources,* Volume 16, pp. 41-51.

Beven, K. J., 2000. *Rainfall-Runoff Modeling.* Chichester: John Wiley and sons ltd.

Bingeman, A., Kouwen, N. & Soulis, E., 2006. Validation of the hydrological processes in a hydrological model. *Journal of Hydrologic Engineering,* 11(5), pp. 451-463.

Bitew, M. & Gebremichael, M., 2010. Spatial variability of daily summer rainfall at a local-scale in mountainous terrain and humid tropical region. *Atmospheric Research,* 98(2-4), pp. 347-352.

Bitew, M. M. & Gebremichael, M., 2011. Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands. *Hydrology and Earth System Sciences,* Volume 15, pp. Bitew, M. M. and Gebremichael, M., 2011: Assessment of satellite rainfall products for streamflow simulation in medium watersheds of the Ethiopian highlands, Hydrology and Earth System Sciences, 15, 1147-1155.

Blake, E. et al., 2013. *Tropical cyclone Report Hurricane Sandy,* s.l.: National Hurricane Center.

Blazkova, S. & Beven, K., 2009. A Limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka Catchment, Czech Republic. *Water Resources Research,* Volume 45, p. W00B16.

Borga, M., Anagnostou, E., Bloschl, G. & Creutin, J., 2011. Flash flood forecasting, warning and risk management: the HYDRATE project. *Environmental Science and Policy,* 14(7), pp. 834-844.

Boughten, W. C., Renard, K. G. & Stone, J. J., 1987. Flood frequency estimates in Southeastern Arizona. *Journal of Irrigation and Drainage Engineering,* 113(4), pp. 469-478.

Boyle, D., Gupta, H. & Sorooshian, S., 2000. Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods. *Water Resources Research,* 36(12), pp. 3663-3674.

Brakenridge, G. et al., 2003. *Flood warnings, flood disaster assessments, and flood hazard reduction: the roles of orbital remote sensing.* Pasadena, CA, Jet Propulsion Laboratory, National Aeronautics and Space Administration.

Brandt, M., 1990. Simulation of runoff and nitrate transport from mixed basins in Sweden. *Nordic Hydrology,* Volume 21, pp. 13-34.

Braud, I. et al., 2010. The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes. *Journal of Hydrology,* Volume 394, pp. 162-18.

Brocca, L., Melone, F. & Moramarco, T., 2011. Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting. *Hydrological Processes,* pp. 2801-2813.

Brooks, S., 1958. A discussion of random methods for seeking maxima. *Operations Research,* 6(2), pp. 244-251.

Chen, C., Liaw, A. & Breiman, L., 2004. *Using Random Forest to Learn Imbalanced Data,* s.l.: Department of Statistics, University of California Berkeley, Technical Report, 666.

Chow, V., Maidment, D. & Mays, L., 1988. *Applied Hydrology.* International Editions ed. Singapore: McGraw-Hill.

Chow, V. T., 1951. A general formula for hydrologic frequency analysis. *Transactions American Geophysical Union,* 32(2), pp. 231-237.

Crawford, N. & Linsley, R., 1966. *Digital simulation in hydrology: Stanford Watershed Model IV. Technical Report 39,* Palo Alto, CA: Stanford University.

Cunderlik, J. M. & Ouarda, T. B., 2007. Regional flood-rainfall duration-frequency modeling at small ungaged sites. *Journal of Hydrology,* pp. 61-69.

Dawdy, D. & Thompson, T., 1967. Digital computer simulation in hydrology. *Journal American Water Works Association,* 59(6), pp. 685-688.

Dehotin, J. & Braud, I., 2008. Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments. *Hydrology and Earth System Sciences,* Volume 12, pp. 769-796.

Dingman, S., 2002. *Physical hydrology.* 2nd ed. Upper Saddle River, NJ: Prentice-Hall Inc.

Dis, M. et al., 2015. Evaluating Multi-sclae Flow Predictions for the Connecticut River Basin Submitted.

Duan, Q., Gupta, V. & Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Application,* 76(3), pp. 501-521.

Du, B., Ji, X., Harmel, R. & Hauck, L., 2009. Evaluation of a watershed model for estimating daily flow using limited flow measurements. *Journal of the American Water Resources Association,* 45(2), pp. 475-484.

Fernandez, G., Chescheir, G., Skaggs, R. & Amatya, D., 2005. Development and testing of watershed-scale models for poorly drained soils. *Transactions of the American Society of Agricultural Engineers,* 48(2), pp. 639-652.

Gao, Y. C. & Liu, M. F., 2013. Evaluation of high-resolution satellite precipitation products using rain gauge observations over the Tibetan Plateau. *Hydrology abd Earth System Sciences,* Volume 17, pp. 837-849.

Gourley, J. J. et al., 2011. Hydrologic Evaluation of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations on Ft. Cobb Basin, Oklahoma. *Journal of Hydrometeorology,* 12(5), pp. 397-988.

Haan, C., 1977. *Statistical methods in Hydrology.* Ames, Iowa, USA: Iowa State University Press.

Hirpa, F., Gebremichael, M. & Hopson, T., 2010. Evaluation of High-Resolution Satellite Precipitation Products over Very Complex Terrain in Ethiopia. *Journal of Applied Meteorology and Climatology,* 49(5), pp. 1044-1051.

Hirpa, F., Gebremichael, M. & Over, T., 2010. River fluctuation analysis: effect of watershed area. *Water Resources Research,* Volume 46, p. W12529.

Hodgkins, G., Dudley, R. & Huntington, T., 2003. Changes in the timing of high river flows in New England over the 20th century. *Journal of Hydrology,* Volume 278, pp. 244-252.

Hoedjes, J. et al., 2014. A Conceptual Flash Flood Early Warning System for Africa, Based on Terrestrial Microwave Links and Flash Flood Guidance. *ISPRS International Journal of Geo-Information,* 3(2), pp. 584-598.

Hong, Y. et al., 2007. A first approach to global runoff simulation using satellite rainfall estimation. *Water Resources Research,* 43(8), p. W08502.

Hong, Y., Adler, R., Huffman, G. & Pierce, H., 2010. Applications of TRMM-Based Multi-Satellite Precipitation Estimation for Global Runoff Prediction: Prototyping a Global Flood Modeling System. In: M. Gebremichael & F. Hossain, eds. *Satellite Rainfall Applications for Surface Hydrology.* Netherlands: Springer, pp. 245-265.

Hossain, F., Tang, L., Anagnostou, E. N. & Nikolopoulos, E., 2009. A practical guide to a space-time stochastic error model for simulation of high resolution satellite rainfall data. *Satellite Rainfall Applications for Surface Hydrology, Springer Science and Business Media B.V.,* pp. 145-167.

Hou, A. Y. et al., 2014. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society,* 95(5), pp. 701-722.

Huffman, G. J. et al., 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology,* pp. 38-55.

Javelle, P., Ouarda, T. B. & Bobee, B., 2003. Spring flood analysis using the flood-duration–frequency approach: application to the provinces of Quebec and Ontario, Canada. *Hydrological Processes,* Volume 17, pp. 3717-3736.

Javelle, P. et al., 2002. Development of regional flood-duration-frequency curves based on the index-flood method. *Journal of Hydrology,* pp. 249-259.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P., 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology,* 5(3), pp. 487-503.

Keast, D. & Ellison, J., 2013. Magnitude Frequency Analysis of Small Floods Using the Annual and Partial Series. *Water,* pp. 1816-1829.

Khan, S. et al., 2011. Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: implications for hydrologic prediction in ungauged basins. *IEEE Transactions on Geoscience and Remote Sensing,* 49(1), pp. 85-95.

Khan, S. I. et al., 2012. Microwave Satellite Data for Hydrologic Modeling in Ungauged Basins. *IEEE Geoscience and Remote Sensing Letters,* 9(4), pp. 663-667.

Kite, G. & Pietroniro, A., 1996. Remote Sensing Applications in Hydrological Modelling. *Hydrological Sciences Journal,* 41(4), pp. 563-591.

Kouwen, N. et al., 1993. Grouped response units for distributed hydrologic modelling. *Journal of Water Resources Planning and Management,* 119(3), p. 289–305.

Langbein, W., 1949. Annual Floods and the Partial Duration Flood Series. *Transactions American Geophysical Union,* 30(6), pp. 879-891.

Leclerc, M. & Ouarda, T. B., 2007. Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology,* 343(3-4), pp. 254-265.

Legates, D. & McCabe, G., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research,* 35(1), pp. 233-241.

Lins, H. F., 1997. Regional streamflow regimes and hydroclimatology of the United States. *Water Resources Research,* 33(7), pp. 1655-1667.

Lin, Y. & Mitchell, K. E., 2005. *The NCEP stage II/IV hourly precipitation analyses: Development and applications,* San Diego, CA: American Meteorological Society 1.2..

Lopez, P., 2011. Direct 4D-Var Assimilation of NCEP Stage IV Radar and Gauge Precipitation Data at ECMWF. *monthly Weather Review,* Volume 139, pp. 2098-2116.

Maggioni, V., Reichle, R. H. & Anagnostou, E. N., 2011. The Effect of Satellite Rainfall Error Modeling on Soil Moisture Prediction Uncertainty. *Journal of Hydrometeorology,* 12(3), pp. 413-428.

Maggioni, V. et al., 2013. Investigating the applicability of error correction ensembles of satellite rainfall products in river flow simulations. *Journal of Hydrometeorology,* 14(4), pp. 1194-1211.

Marshall, E. & Randhir, T., 2008. Effect of climate change on watershed system: a regional analysis. *Climatic Change,* 89(3-4), pp. 263-280.

Mei, Y., Anagnostou, E. N., Nikolopoulos, E. & Borga, M., 2014. Error Analysis of Satellite Rainfall Products in Mountainous Basins. *Journal of Hydrometeorology,* 15(5), pp. 1778-1793.

Mei, Y. et al., 2014. Rainfall organization control on the flood response of mild-slope basins. *Journal of Hydrology,* Volume 510, pp. 565-577.

Mesinger, F. et al., 2005. North American Regional Reanalysis. *Bulletin of the American Meteorological Society,* pp. 343-360.

Moretti, G. & Montanari, A., 2008. Inferring the flood frequency distribution for an ungauged basin using a spatially distributed rainfall-runoff model. *Hydrology and Earth System Sciences,* Volume 12, pp. 1141-1152.

Moriasi, D. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers ,* 50(3), pp. 885-900.

Motovilov, Y., Gottschalk, L., Engeland, K. & Rodhe, A., 1999. Validation of a distributed hydrological model against spatial observations. *Agricultural and Forest Meteorology,* Volume 98–99, p. 257–277.

Nash, J. & Sutcliffe, J., 1970. River Flow forecasting through conceptual models part-I-A discussion of principles. *Journal of Hydrology,* Volume 10, pp. 282-290.

National Oceanic and Atmospheric Administration, 2012. *Hurricane Irene, August 21-30,2011,* Maryland: U.S. DEPARTMENT OF COMMERCE.

Natrella, M., 1963. *Experimental Statistics,* s.l.: National Bureau of Standards Handbook 91.

Nikolopoulos, E., Anagnostou, E. & Borga, M., 2013. Using High-Resolution Satellite Rainfall Products to Simulate a Major Flash Flood Event in Northern Italy. *Journal of Hydrometeorology,* 14(1), pp. 171-185.

Noto, L., Ivanov, V., Bras, R. & Vivoni, E., 2008. Effects of initialization on response of a fully distributed hydrological model. *Journal of Hydrology,* 352(1-2), pp. 107-125.

Odoni, N. & Lane, S., 2010. Knowledge-theoretic models in hydrology. *Progress in Physical Geography ,* 34(2), pp. 151-171.

Pan, M., Li, H. & Wood, E., 2010. Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resources Research,* 46(9), p. W09535.

Refsgaard, J., 1997. Parameterization, calibration and validation of distributed hydrological models. *Journal of Hydrology,* 198(1-4), pp. 69-97.

Reich, B. M. & Renard, K. G., 1981. Application of advances in flood frequency analysis. *American Water Resources Association ,* pp. 67-74.

Rozalis, S., Morin, E., Yair, Y. & Price, C., 2010. Flash flood prediction using an uncalibrated hydrological model and radar rainfall data in a Mediterranean watershed under changing hydrological conditions. *Journal of Hydrology ,* 394(1-2), pp. 245-255.

Servat, E. & Dezetter, A., 1991. Selection of calibration objective functions in the context of rainfall-runoff modeling in a Sudanese Savannah area. *Hydrological Sciences Journal,* 36(4), pp. 307-330.

Seyyedi, H., Anagnostou, E., Beighley, E. & McCollum, J., 2015. Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. *Atmospheric Research,* Volume 164-165, pp. 37-48.

Seyyedi, H., Anagnostou, E. N., Beighley, E. & McCollum, J., 2014. Satellite-driven downscaling of global reanalysis precipitation products for hydrological applications. *Hydrology and Earth System Science,* Volume 18, pp. 5077-5091.

Shrestha, M. S., Artan, G., Bajracharya, S. & Sharma, R., 2008. Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin. *Journal of Flood Risk Management,* Volume 1, pp. 89-99.

Sorooshian, S. et al., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society,* 81(9), pp. 2035-2046.

Su, F., Hong, Y. & Lettenmaier, D., 2008. Evaluation of TRMM multisatellite precipitation analysis (TMPA) and its utility in hydrologic prediction in the La Plata basin. *Journal of Hydrometeorology ,* 9(4), p. 622–640.

Tang, L., Hossain, F. & Huffman, G. J., 2010. Transfer of Satellite Rainfall Error from Gauged to Ungauged Regions at Regional and Seasonal Timescales. *Journal of Hydrometeorology,* 11(6), pp. 1263-1274.

Tasker, G. D., 1978. Flood Frequency Analyis with a Generalized Skew Coefficient. *Water Resources Research,* 14(2), pp. 373-376.

U.S. Water Resources Council, (. C. o. W., 1981. *Guidelines for determining flood flow frequency, bulletin 17B,* Reston, VA: U.S. Geological Survey.

Urias, H. Q., Garcia, H. & Mendoza, J. S. P., 2007. *Determination of the relationship between precipitation and return periods to assess flood risks in the city of Juarez, Mexico.* Carbondale, Southern Illinois University Carbondale, p. 9.

Vergara, H. et al., 2013. Effects of Resolution of Satellite-based Rainfall Estimates on Hydrologic Modeling Skill at Different Scales. *Journal of Hydrometeorology,* 15(2), pp. 593-613.

Vincendon, B. et al., 2010. Benefit of coupling the ISBA land surface model with a TOPMODEL hydrological model version dedicated to Mediterranean flash-floods. *Journal of Hydrology,* Volume 394, pp. 256-266.

Vrugt, J. et al., 2009. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation,* 10(3), pp. 273-290.

Vrugt, J., ter Braak, C., Gupta, H. & Robinson, B., 2008. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?. *Stochastic Environmental Research and Risk Assessment,* 23(7), pp. 1011-1026.

Wallis, J., Matalas, N. & Slack, J., 1974. Just a moment. *Water Resources Research,* 10(2), pp. 211-219.

Wang, D., Wang, G. & Anagnostou, E., 2007. Evaluation of canopy interception schemes in land surface models. *Journal of Hydrology,* Volume 347, pp. 308-318.

Wang, J. et al., 2011. The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrological Science Journal,* 56(1), pp. 84-98.

Weibull, W., 1939. *A Statistical Theory of the Strength of Materials.* Stockholm: Generalstabens litografiska anstalts förlag.

Wu, H. et al., 2012. Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model. *Journal of Hydrometeorology,* 13(4), p. 1268–1284.

Xie, Z. et al., 2007. Regional Parameter Estimation of the VIC Land Surface Model: Methodology and Application to River Basins in China. *Journal of Hydrometeorology,* 8(3), pp. 447-468.

Yong, B. et al., 2012. 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,* 117(D9), p. D09108.

Zanon, F. et al., 2010. Hydrological analysis of a flash flood across a climatic and geologic gradient: The September 18, 2007 event in Western Slovenia. *Journal of Hydrology,* 394(1-2), pp. 182-197.

Zhang, X. et al., 2013. Using NWP Simulations in Satellite Rainfall Estimation of Heavy Precipitation Events over Mountainous Areas. *Journal of Hydrometeorology,* Volume 14, pp. 1844-1858.

Zipser, E. et al., 2006. Where ate the most intense thunderstorms on earth?. *Bulletin of the American Meteorological Society,* 87(8), pp. 1057-1071.

URL References:

Connecticut River Watershed Council: http://www.ctriver.org/

Sistema Regional de Visualización y Monitoreo de Mesoamérica: www.servir.net

U.S. Geological Survey: www.usgs.gov