ON THE PERFORMANCE OF SATELLITE-BASED PRECIPITATION PRODUCTS FOR SIMULATING STREAM WATER QUALITY

by

Jennifer Solakian
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Civil, Environmental and Infrastructure Engineering

Committee:

Viviana Maggioni	Dr. Viviana Maggioni, Dissertation Director
Celso Moller Ferreira	Dr. Celso Ferreira, Committee Member
	Dr. Kuo Tian, Committee Member
Adil N. Godrej	Dr. Adil Godrej, Committee Member
D. Salem	Dr. Sam Salem, Department Chair
Kenneth S Ball	Dr. Kenneth S. Ball, Dean, Volgenau School of Engineering

Date: 07/27/2020

Summer Semester 2020
George Mason University
Fairfax, VA

On the Performance of Satellite-based Precipitation Products for Simulating Stream Water Quality

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Jennifer Solakian
Master of Science
University of Maryland, 2005
Bachelor of Science
Virginia Polytechnic Institute and State University, 2001

Director: Viviana Maggioni, Associate Professor
Civil, Environmental and Infrastructure Engineer, George Mason University

Summer Semester 2020
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone who has helped and supported me throughout my journey. Foremost, I want to express my deepest appreciation to my advisor, Dr. Viviana Maggioni, for all of her professional guidance and support throughout my graduate studies. Thank you for making my doctoral experience very enjoyable and all of your patience with me as a part-time student.

I would like to show my gratitude to committee member Dr. Adil Godrej from the VA Tech Occoquan Monitoring Laboratory. Thank you for your guidance and suggestions in developing my Ph.D. research. Without your collaboration, this research would not be possible. I also want to thank committee members Dr. Celso Ferreira and Dr. Kuo Tian for their support and interest in my research.

I would like to acknowledge my colleagues at George Mason University for their valuable support and assistance with processing the data included in this research.

To my parents – no words can describe my appreciation. Your continuous encouragement and support will never be forgotten.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND SYMBOLS

Ammonium-Nitrate ..... NH_4-N
Areal-Weighted and Segment-Aggregated ..... AWSA
Bias ..... B
Bias Propagation Factor ..... E_{bias}
Biological Oxygen Demand ..... BOD
Climate Prediction Center's Morphing Technique ..... CMORPH
Correlation Coefficient CC
Critical Success Index ..... CSI
Climate Prediction Center ..... CPC
Cloud Classification System ..... CCS
Degrees Celsius ..... °C
Dissolved Oxygen ..... DO
District of Columbia ..... D.C.
Environmental Protection Agency ..... EPA
European Centre for Medium Range Weather Forecasts ..... ECMWF
False Alarm Rate ..... FAR
Hydrologic Simulation Program FORTRAN ..... HSPF
Infrared ..... IR
Mean Relative Error ..... E_p
Meters Squared per Second ..... m^3/s
Milligrams per Liter ..... mg/l
Millimeter ..... mm
Modern Era Retrospective-analysis for Research and Applications ..... MERRA
Multi-Source Weighted-Ensemble Precipitation ..... MSWEP
Nash-Sutcliffe Efficiency ..... NSE
National Aeronautics and Space Administration ..... NASA
National Oceanic Atmospheric Administration ..... NOAA
Nitrate-Nitrogen ..... NO_3-N
Orthophosphate Phosphorus ..... OP
Passive Microwave ..... PMW
Precipitation, Daily AWSA ..... P
Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks ..... PERSIANN
Probability Density Function ..... PDF
Probability of Detection ..... POD

Relative Bias ..... rB
Relative Root-Mean-Square Error ..... rRMSE
Relative Root-Mean-Square Error Propagation Factor..... E_{rRMSE}
Root-Mean-Square Error ..... RMSE
Satellite Precipitation Product ..... SPP
Soil and Water Assessment Tool ..... SWAT
Square Kilometer ..... sq. km
Standard Deviation..... σ
Stream Temperature..... TW
Streamflow..... Q
Total Phosphorus ..... TP
Total Suspended Solids..... TSS
TRMM Multi-Satellite Precipitation Analysis..... TMPA
Tropical Rainfall Measuring Mission ..... TRMM
Two-Dimensional Hydrodynamic and Water Quality Model CE-QUAL-W2
Water Reclamation Facility ..... WRF

ABSTRACT

ON THE PERFORMANCE OF SATELLITE-BASED PRECIPITATION PRODUCTS FOR SIMULATING STREAM WATER QUALITY

Jennifer Solakian, Ph.D.

George Mason University, 2020

Dissertation Director: Dr. Viviana Maggioni

This dissertation presents an investigation on the use of satellite-based precipitation products (SPPs) in a hydrologic model to estimate water quality indicators in stream simulations. Three SPPs based on different retrieval algorithms are considered: the Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis, TMPA 3B42-V7; the Climate Prediction Center's CMORPH V1.0 product; and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System, PERSIANN-CCS. The three SPPs are compared to rain gauge-based records over a 5-year period across the Occoquan Watershed, a 1500 square kilometer area, located in the suburban Washington, D.C. area. The three SPPs and the gauge-based dataset are then used as input to the Hydrologic Simulation Program FORTRAN (HSPF) hydrology and water quality model. Each SPP-forced simulation is compared to the reference model simulation forced with the gauge-based observations, in terms of streamflow and several water quality indicators including stream temperature

(TW) and concentrations of total suspended solids (TSS), dissolved oxygen (DO), biochemical oxygen demand (BOD), orthophosphate phosphorus (OP), total phosphorus (TP), ammonium-nitrate (NH_4-N), and nitrate-nitrogen (NO_3-N). First, the skill of each SPP is evaluated on a continuous basis over the 5-year study period. Second, the propagation of errors from input SPPs to simulated streamflow and water quality indicators are evaluated Third, the model is evaluated during eight extreme hydrometeorological events in terms of simulated streamflow and water quality indicators. Results indicate that the spatiotemporal variability of SPPs, along with their algorithms to estimate precipitation, have a quantifiable impact on SPP-simulated streamflow and water quality indicators during both continuous and event-based modeling of extreme hydrometeorological events.

CHAPTER 1. INTRODUCTION

Precipitation is widely accepted as the most influential meteorological input in hydrological and water quality investigations. Reliable and consistent hydrologic predictions depend on the accurate measurement of precipitation including its intensity, duration, spatial patterns, and extent. It is well understood that these characteristics of precipitation have significant and direct impacts on runoff and streamflow. In turn, precipitation is also a major driver of water quality. Large scale hydrologic models often rely on remotely-sensed precipitation data from satellite sources in locations where ground-based monitoring systems and rain gauge networks are unavailable or unreliable. Satellite-based precipitation products (SPPs) offer a viable alternative to ground-based precipitation data through continuous monitoring of precipitation with greater spatial coverage.

A plethora of research has been conducted over the past two decades on the use of remote sensing products, including SPPs, for hydrologic model simulation and predictions intended for flood forecasting, riverine modeling, and climate studies. Conversely, a number of studies evaluating and quantifying predictions of stream water quality from ground-based sources have demonstrated that spatial-temporal variations of water quality is strongly dependent on the spatial and temporal scales of analysis. Where

current research falls short is the application of SPP data in hydrologic models specifically aimed at simulating and forecasting water quality.

It is the objective of this research to fill in this gap in the literature, by providing a proof-of-concept for evaluating the performance of a hydrologic model in estimating water quality indicators, forced with three SPPs based on different retrieval algorithms; the Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis, TMPA 3B42-V7; the Climate Prediction Center's CMORPH V1.0 product; and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System, PERSIANN-CCS. This research is carried out in the Occoquan Watershed, located in the northern Virginia suburbs of Washington, D.C., United States. The overarching purpose of this study is to compare the skill of SPPs to the data collected by a dense rain gauge network to determine their capability to (1) accurately characterize the spatial and temporal variability of precipitation within the study region, (2) simulate streamflow and water quality on a continuous basis during a 5-year study period, and (3) simulate event-based streamflow and water quality during hydrometeorological extremes. This work represents a first attempt to utilize SPPs for water quality modeling which could be of critical importance in areas of the world where rain-gauge networks or monitoring stations are either sparse or not available altogether.

Research Questions

To address the aforementioned overarching research objective, this work presents a methodology to comprehensively assess uncertainties associated with SPPs and

investigate their ability to simulate stream water quality indicators using a hydrologic model. This research aims to answer the following scientific questions:

1. How well do SPPs perform in estimating precipitation on a continuous basis and during extreme hydrometeorological events?
2. How well do SPPs perform in simulating streamflow and water quality indicators on a continuous basis and during extreme hydrometeorological events?
3. How does the performance of SPPs influence the propagation of error between input precipitation and simulated streamflow and water quality indicator output?

Thesis Organization

To accomplish this research, the following tasks have been identified:

Task 1: Investigate the efficiency of SPPs for simulating stream water quality indicators. Three SPPs with different spatial resolutions and based on different retrieval algorithms (TMPA 3B42-V7, CMORPH V1.0, PERSIANN-CCS) are compared to gauge-based records over a 5-year period across the study region. The three SPPs and the gauge-based dataset are used as input to the Hydrologic Simulation Program FORTRAN (HSPF) hydrology and water quality model developed for the Occoquan Watershed. Each SPP-forced simulation in the HSPF model is compared to the reference model simulation forced with the gauge-based observations, in terms of streamflow and water quality indicators, i.e., stream temperature (TW), total suspended solids (TSS), dissolved oxygen (DO), and biological oxygen demand (BOD). Task 1 includes an uncertainty analysis in terms of correlation coefficients (CC), root-mean-square error (RMSE), and bias (B) for

streamflow, TW, and concentrations of TSS, DO, and BOD to evaluate the performance of each SPP.

Task 2: Investigate the error propagation from input precipitation to output water quality indicators simulated by a hydrologic model. This task evaluates the propagation of error in the precipitation input to that of the simulated streamflow and water quality output from a hydrologic model. This work assesses error statistics and the performance of each SPP to gauge-based data for the entire watershed on a seasonal basis. A seasonal analysis is performed using CC, relative root-mean-square error (rRMSE), and B for daily precipitation input and model output of streamflow and water quality indicators to examine the error propagation from precipitation to model output (streamflow, TW, TSS, and DO).

Task 3: Investigate the performance of SPPs in simulating streamflow and water quality indicators during extreme hydrometeorological events. The intent of this task is to investigate the performance of SPPs in simulating streamflow and water quality during eight extreme hydrometeorological events. The variability of stream response to extreme precipitation events is compared to base conditions as well as the ability of each SPP simulation to capture the timing and magnitude of peaks. The skill of each SPP is assessed using CC, rRMSE, and relative bias (rB) in simulating streamflow and the following water quality indicators: TW, TSS, DO, BOD, orthophosphate phosphorus (OP), total phosphorus (TP), ammonium-nitrate (NH_4-N), and nitrate-nitrogen (NO_3-N).

The research undertaken to address the aforementioned tasks is presented in Chapters 2 through 5. Chapter 2 describes the study area, the hydrologic and water quality model, the precipitation data sets used in this study, and the methods used to process precipitation data for model simulations. Chapter 3 discusses Task 1, the methodology and results for investigating the efficiency of using SPPs for simulating water quality indicators on a continuous basis over a 5-year study period. Chapter 4 presents Task 2, an analysis of the error propagation from satellite-based input precipitation to water quality indictor output on a seasonal basis. Chapter 5 describes Task 3, an investigation of the performance of SPPs in simulating stream water quality during hydrometeorological extremes. Lastly, concluding remarks associated with this research are provided in Chapter 6.

Dissemination of Results

A variation of the work described herein is presented in the following peer-reviewed research articles:

Chapter 2: Solakian, J., Maggioni, V., Lodhi, A., & Godrej, A. (2019). Investigating the use of satellite-based precipitation products for monitoring water quality in the Occoquan Watershed. *Journal of Hydrology: Regional Studies, 26*, 100630. https://doi.org/10.1016/j.ejrh.2019.100630

Chapter 3: Solakian, J., Maggioni, V., & Godrej, A. (n.d.). Investigating the error propagation from satellite-based input precipitation to output water quality indicators simulated by a hydrologic model. *Advances in Water Resources,* under review.

Chapter 4: Solakian, J., Maggioni, V., & Godrej, A. (n.d.). On the performance of satellite-based precipitation products in simulating stream water quality during hydrometeorological extremes. *Frontiers in Environmental Science*, in preparation.

CHAPTER 2. STUDY AREA, HYDROLOGIC MODEL, AND DATA SETS

2.1 Study Area

The focus of this study is the Occoquan Watershed, a 1500 square kilometer (sq. km) area located in in the northern portion of Virginia, within suburban Washington, D.C. (Figure 1). The watershed supplies water to two drinking water reservoirs: Lake Manassas and the run-of-the-river Occoquan Reservoir, which are part of the drinking water supply to approximately two million people. The area is urbanized with a mix of suburban and urban land use and has a mild topographic variation. The Occoquan Watershed is part of the Potomac River Watershed, ultimately discharging into the Chesapeake Bay, a waterbody that has been the focus of immense restoration efforts over the past several decades to improve the water quality in the contributing watershed.

The Occoquan Watershed has also been a major focus of regulatory oversight for the past 40 years due to the unprecedented growth of the D.C. metropolitan area which resulted in a substantial increase in reclaimed domestic wastewater discharged to receiving waters in the watershed, the closure of eleven marginally-functioning water reclamation facilities (WRFs) and the construction of a single state-of-the-art WRF that currently receives wastewater from Prince William and Fairfax counties and the cities of Manassas and Manassas Park, and discharges to one of the main tributaries (Bull Run) of the Occoquan Reservoir. Since 1973 the Virginia Tech Occoquan Watershed Monitoring

Laboratory has monitored the watershed through data acquisition from a network of rain gauges, meteorological stations and stream monitoring stations. Precipitation is measured continuously throughout the watershed using automated rain gauges, whereas streamflow and several indicators of water quality are measured using automated stream monitoring stations. Figure 1 presents the study area, the Occoquan Watershed, including a delineation of seven catchments, topography, major streams, major waterbodies, and the locations of rain gauges and stream monitoring stations, shown as black triangles and grey squares, respectively. The locations of model evaluation points are also represented in Figure 1 as grey circles (S47, S86, and S79 coincide with stream monitoring stations).

The watershed is formed by three main stream systems, Broad Run, Cedar Run, and Bull Run, and is divided into seven distinct catchments, which are then partitioned into 87 segments used in the Occoquan Watershed hydrologic model (Section 2.2). Three of the catchments (Upper Bull Run, Upper Broad Run, and Cedar Run), represented in Figure 1 in green, are the focus of this study. These three catchments are chosen since they represent the headwaters of the watershed prior to entering a major waterbody and are each monitored by a steam monitoring station near the confluence of the catchment. To investigate the uncertainty by basin scale, a cascade of sub-basins with different drainage areas is assessed at six evaluation points ranging from almost 18 to 500 sq. km in size.

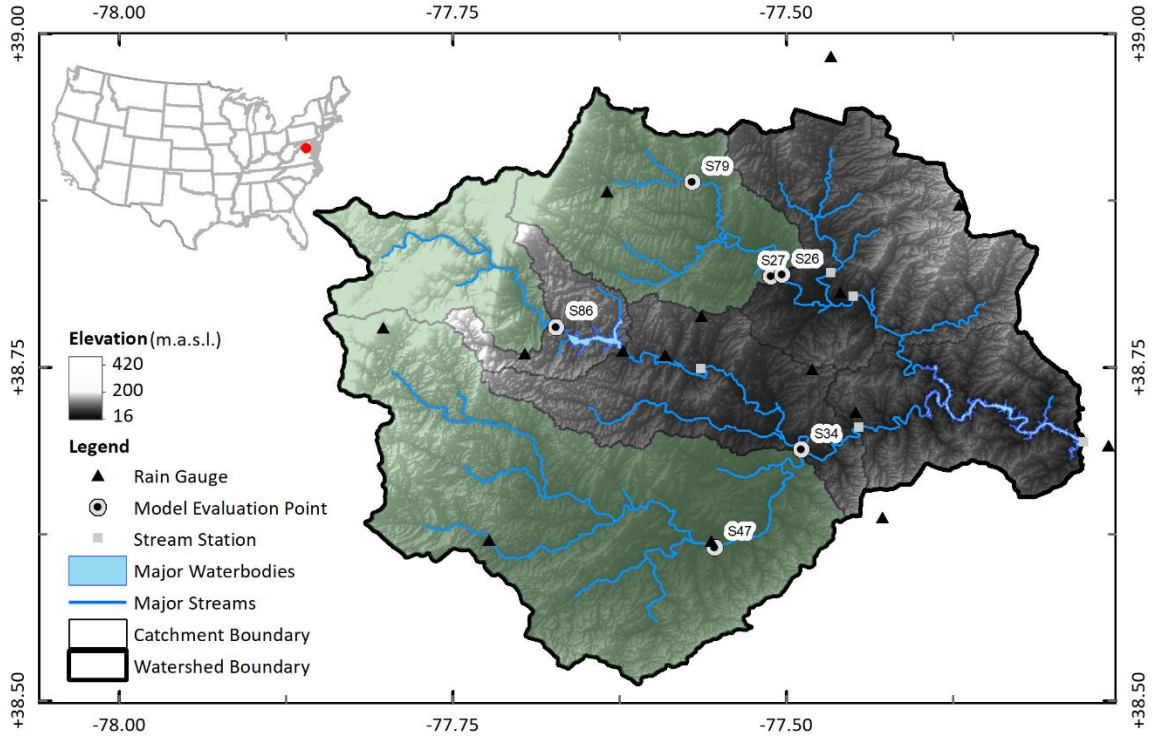

Figure 1. Occoquan Watershed study area

2.2 Hydrologic Model

The Occoquan Watershed Monitoring Laboratory uses the U.S. Environmental Protection Agency's (EPA) Hydrologic Simulation Program FORTRAN (HSPF) model to simulate the hydrology, nonpoint source runoff, and water quality of the watershed. This long-standing model with origins from the 1970s has been widely adopted in the U.S. for its ability to simulate complex watersheds within numerous land cover and climatic conditions along with various fate and transport processes (Albek et al., 2004; Duda et al., 2012; Li et al., 2015; Mishra et al., 2007). The Occoquan Watershed HSPF conceptional lumped hydrological model of the is designed to simulate various hydrological processes and associated water quality components in the Occoquan

Watershed (Xu et al., 2007). The model delineates 87 segments within the seven catchments of the Occoquan Watershed, as depicted in Figure 2. The two reservoirs—Lake Manassas and Occoquan Reservoirs—are simulated using the U.S. Army Corps of Engineers' developed two-dimensional hydrodynamic and water quality model CE-QUAL-W2, but are not included in this study as they contribute very little to the water budget except for the rain that falls directly on the reservoir water surfaces.

The period of 2008-2012 was chosen for analysis since at the start of this study the latest Occoquan Watershed model assesses that period. The model is updated every five years with current land use data and is modeled in 5-year increments with in-situ meteorological data: air temperature, cloud cover, dew point temperature, wind speed, solar radiation, potential evapotranspiration, and precipitation. Air temperature, cloud cover, dew point temperature, solar radiation, wind speed, and wind direction are all collected at the National Oceanic and Atmospheric Administration (NOAA) weather station located at the Washington Dulles International Airport approximately 27 km from the centroid of the Occoquan Watershed. Hourly readings from this station are applied to the entire watershed for those parameters. Since a direct measurement of potential evapotranspiration is not available from the Dulles station, potential evapotranspiration is calculated from the modified Penman Pan empirical method using air temperature, dew point temperature, wind movement, and solar radiation obtained from the Dulles station. The temporal and spatial resolution of the retrieved land use data and all meteorological data (i.e., air temperature, cloud cover, dew point temperature, wind speed, solar radiation, and potential evapotranspiration), aside from precipitation, are consistently

maintained and are not altered in any way during model simulations. An evaluation of the efficiency and a validation of the long-established HSPF hydrology model is not within this scope of study; however, a brief description of the model's setup is described herein. For additional details of model setup and execution may be found in Xu et al. (2005, 2007) and Maldonado and Moglen (2013).

Precipitation input is retrieved on an hourly basis for each segment from the nearest-neighbor rain gauge, while all other meteorological data used in the model (air temperature, cloud cover, dew point temperature, solar radiation, wind speed, and wind direction) are applied consistently throughout all 87 segments. During certain periods of this 5-year study, rain gauge observations are not recorded at various gauges. Missing precipitation records are interpolated since the HSPF model requires continuous precipitation input. For missing gauge records, a standardized procedure is employed to estimate the discrete hourly values based on the spatial and temporal infilling strategy outlined by Xu (2005). A similar infilling strategy is applied to missing precipitation data for satellite-based precipitation product (SPPs) during the study period. Using similar procedures outlined in multiple metrological infilling studies (Hema & Kant, 2017; Kim & Ryu, 2015; Lee & Kang, 2015) and recommended by the U.S. EPA (2000), meteorological data gaps are classified into categories based on gap duration. For gaps up to four hours, temporal linear interpolation is applied. For gaps greater than four hours, a combination of temporal linear interpolation and nearest-neighbor pixel value is utilized. Precipitation input from each satellite product is then aggregated for each segment (described in Section 2.4).

The watershed is comprised of 87 segments within distinct seven catchments. Each of the seven catchments has a separate HSPF model that is linked to create the overall watershed model. The HSPF model used in this analysis is deterministically calibrated and validated between simulated gauge-based results and observed data over a 5-year period from January 1, 2008 to December 31, 2012 (described in Section 2.2.2). While HSPF is able to produce output results as low as 15-minute increments, (which is the increment flow data are recorded and used for model input) the Occoquan HSPF model is set to output results using a daily scale. The daily scale is used to minimize the impacts of timing errors often found using lower increment outputs since HSPF is prone to timing errors, such as under- or over-predicting when an event starts and/or stops.

Table 1. Occoquan Watershed characteristics

Model Evaluation Point	Catchment	Drainage Area (sq. km)	Number of Segments	Contributing Rain Gauge(s)	Stream Monitoring Station
S27	Upper Bull Run	17.81	1	BLFD	-
S79	Upper Bull Run	63.83	5	EVGR	ST60
S26	Upper Bull Run	186.24	16	EVGR, BLVD	-
S47	Cedar Run	394.26	12	AIRL, RITC, CROK, CEDA	ST25
S34	Cedar Run	498.42	15	AIRL, RITC, CROK, CEDA	-
S86	Upper Broad Run	126.44	7	AIRL, EVGR, RITC	ST70

2.2.1 Stream Monitoring Stations

Streamflow, stream stage, and water quality are measured at eight stations across the watershed. Flow and stream stage are continuously measured by automated equipment including a Sutron datalogger and flow meter on an hourly basis provided there is no change. When there is a change in flow, such as during storm events, flow and

stage are recorded every 15 minutes until streamflow and stage are normalized (Xu et al., 2007). Water quality samples for baseflow conditions are typically obtained once each week, and storm samples are composited in the field using an automated sampler paced to sample equal aliquots for equal volume of flow that passes the sampling point, and adding these to a compositing bottle. Water quality parameters including temperature (TW), and concentrations of total suspended solids (TSS), dissolved oxygen (DO), biochemical oxygen demand (BOD), orthophosphate phosphorus (OP), ammonium nitrogen (NH_4-N), nitrate-nitrogen (NO_3-N), and total organic carbon (TOC), are obtained for all samples. Observed information at the eight stations is used to calibrate and validate the hydrologic and water quality model analyzing the Occoquan Watershed. While data from eight stations are used for model calibration and validation, this study analyzes results from three monitoring stations (ST25, ST60, and ST70) specifically for streamflow, TW, TSS, DO, and BOD.

By definition TSS are particles that are larger than 0.7 micrometers whereas anything smaller than 0.7 micrometers is typically considered a dissolved solid. Since TSS is a specific measurement of all suspended solids, organic and inorganic, by mass, it is a good representation of the total sedimentation rate of a watershed. TSS is both a visible and quantifiable indicator of overall water quality in a waterbody and therefore is selected for further evaluation in this study. DO is the measurement of gaseous oxygen dissolved in an aqueous solution which results from diffusion, by aeration, and as a product of photosynthesis. DO is an important indicator, determining the health and

quality of a waterbody for its ability to support life. Similarly, BOD is the amount of DO needed by aerobic organisms to break down organic matter.

Nitrogen-based nutrients, NH_4-N, and NO_3-N, are good indicators of water quality and, like their phosphorous-based counterpart, play a role in eutrophication, oxygen depletion, and biomass production. NH_4-N, or ammonium-nitrogen, is found in runoff from lawn care fertilizer or a result of industrial and wastewater discharge. NO_3-N, is the concentration of nitrogen due to nitrates in waterbody. Phosphorus in aquatic systems is found in both soluble and insoluble forms. Insoluble phosphorous concentrations are related to the total sediment yield while the soluble nutrients account for the contribution from rainfall, land use, and anthropogenic impacts. TP is a measurement of all forms of phosphorus including OP, soluble phosphate-phosphorus and organic phosphorus. Phosphorous is a limiting nutrient in the Occoquan watershed, which can play a role in the eutrophication, oxygen depletion, and biomass production if concentrations exceed need. In the Occoquan Watershed HSPF model, empirical relationships are built in between TP and sediment loads to simulate phosphorus concentrations in reaches, however, the spatial and temporal distribution of phosphorous is a limitation, as with most water quality models. To overcome this limitation, the Occoquan Watershed model is modeled with 87 unique segments to represent the temporal and spatial conditions of the watershed.

2.2.2 Model Calibration

Each catchment HSPF model is deterministically calibrated and validated over a 5-year period (January 1, 2008 to December 31, 2012) between simulated gauge-based

results and observed data prior to linking into the adjoining downstream model. Eight stream monitoring stations throughout the watershed that collect streamflow, stream stage, and water quality constituents are used to calibrate the HSPF model. The model is first calibrated on both daily and monthly streamflow by comparing simulated flows from gauge records with observed data from stream monitoring stations during the first two years of the study period (2008–2009). Results of the simulated model are then validated for an independent period (2010–2012). The model provides output for streamflow (Q), and water quality indicators including TW, TSS, OP, TP, NH_4-N, NO_3-N, TN, DO, BOD, and TOC. Next, the model is calibrated for TSS, TW, OP, NH_4-N, and NO_3-N based on monthly and yearly loads to that of observed in-stream data.

The HSPF model is used to simulate streamflow and select water quality indicators at six designated evaluation points in the Occoquan Watershed. These six locations are located within three distinct catchments (Cedar Run, Upper Broad Run, and Upper Bull Run), which capture runoff from a number of segments of varying size, land use, and topography (Table 1). Of the six model evaluation points, three points (S47, S79, and S86) are equipped with stream monitoring stations, which are also used in the original model calibration/validation process. The streamflow error analysis performed with respect to in-stream observations reveal a well calibrated model at each of the three observation locations (S47, S79, and S86). Error analyses of TW and DO at a daily scale and TSS at a monthly scale also show good model skills at the three locations (Table 2) in terms of Nash-Sutcliffe Efficiency (NSE), relative bias (rB), and root-mean-square

error (RMSE) (eqs. 1-3). Optimal values of NSE, rB, and RMSE are 1, 0, and 0, respectively.

Equation 1. Nash-Sutcliffe Efficiency

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(Q_{s_i} - Q_{o_i})^2}{\sum_{i=1}^{n}(Q_{o_i} - \overline{Q_s})^2}\right]$$

Equation 2. Relative Bias

$$rB = \left[\frac{\sum_{i=1}^{n}(Q_{s_i} - Q_{o_i})}{\sum_{i=1}^{n}(Q_{o_i})}\right] \text{ x } 100\%$$

Equation 3. Root-Mean-Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_{s_i} - Q_{o_i})2}{n}}$$

Where n is the total number of events, Q_{o_i} is the *i*th observed streamflow or water quality indicator value, Q_{s_i} is the *i*th simulated satellite-based streamflow or water quality indicator value, $\bar{Q}$ is the corresponding mean value and n is the number of values. NSE is dimensionless whereas RMSE is in provided in respective units (m^3/s, °C, mg/l) and rB is a percentage.

The model is next forced with precipitation data from each of the three SPPs to simulate streamflow and water quality at the six evaluation points. Table 2 presents the error statistics of calibrated individual catchment-based HSPF Models at three model evaluation points (S47, S86 and S79) equipped with stream monitoring stations for streamflow and select water quality indicators (TW, TSS, and DO).

Table 2. Error statistics of calibrated individual catchment-based HSPF models

	S47/ST25			S79/ST60			S86/ST70		
	NSE	**RMSE**	**rB (%)**	**NSE**	**RMSE**	**rB (%)**	**NSE**	**RMSE**	**rB (%)**
Daily Q (m^3/s)	0.70	0.55	-5.19	0.79	0.46	-6.14	0.77	0.48	-0.88
Monthly Q (m^3/s)	0.79	0.46	-4.94	0.84	0.40	-6.14	0.80	0.44	-3.31
Daily TW (°C)	0.92	0.28	-0.37	0.85	0.39	-0.73	-	-	-
Daily DO (mg/l)	-1.35	1.53	11.96	0.56	0.66	-2.03	0.61	0.63	2.26
Monthly TSS (mg/l)	0.62	0.61	10.81	0.74	0.51	-2.17	0.75	0.50	-5.85

HSPF has been widely accepted as a premier water quality model due to its ability to simulate complex watersheds within numerous land cover and climatic conditions along with various fate and transport processes (Albek et al., 2004; Duda et al., 2012; Li et al., 2015; Mishra et al., 2007). The HSPF model was proven to provide satisfactory performance both in terms of streamflow and water quality processes in several past studies (Hayashi et al., 2004, Huo et al., 2015; Li et al., 2015; Stern et al., 2016). Nevertheless, the accuracy of HSPF-modeled sediment transport predictions was shown to be i) limited by the inability of ground-based meteorological stations to adequately cover the spatial extents and density necessary to represent watershed precipitation (Stern et al., 2016); ii) influenced by storm magnitude and frequency (Stern et al., 2016); and iii) seasonally dependent (Hayashi et al., 2004). A few studies (Diaz-Ramirez et al., 2013; Wu et al., 2006; Young et al., 2000) evaluated the propagation of errors in an HSPF model from input to output. Diaz-Ramirez et al. (2013) suggested that streamflow uncertainty is significantly impacted by precipitation patterns and magnitude, but may also be impacted by several other parameters and variables (e.g., land use classification, slope, infiltration capacity, soil moisture, groundwater recharge, and interflow recession). Young et al. (2000) determined that the quality of precipitation input data has a

significant impact on the uncertainty associated not only with HSPF-simulated streamflow, but also with sediment transport loads and water quality constituents.

2.3 Data Sets

2.3.1 Ground-based Rain Gauge Network

Ground-based precipitation observations are obtained from a network of 15 tipping bucket rain gauges located within or approximate to the Occoquan Watershed (Figure 2). The quality-controlled gauges collect precipitation measurements in increments of 0.254 mm (0.01 inches) by recording the time of occurrence of successive tips logged hourly throughout the 5-year temporal span of this study (2008 – 2012). While ground-based gauge observations are widely accepted as providing the most accurate method for precipitation measurement, tipping buckets have been known to experience significant levels of error at low rainfall intensities (Habib et al., 2012b) and may be sensitive to ground vibrations, wind effects, and orifice opening size. The density of gauges is also important when capturing variability of rainfall at fine temporal scales, for extreme events, or the variability of rainfall over complex terrain (Kidd et al., 2017). Additionally, many gauges are typically not able to measure snowfall accurately and therefore may underestimate its liquid precipitation equivalent (Rasmussen et al., 2012).

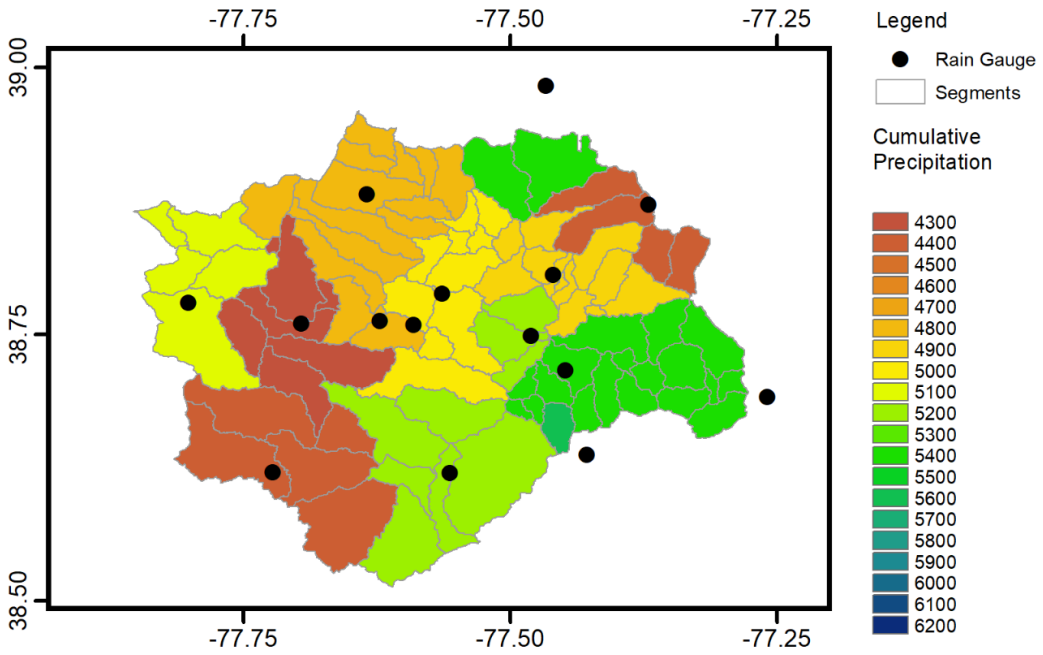

Figure 2. Watershed segment delineations and 5-year cumulative precipitation measured by the rain gauges and interpolated to each segment

Gauge observations are interpolated to each of the 87 segments in the watershed using the nearest neighbor method. Cumulative precipitation of the interpolated gauge data during the study period is shown in Figure 2 for each segment of the Occoquan Watershed. Fourteen of the 15 rain gauges across the study area provide hourly records; however, one gauge (LRTN 150), located at the downstream extents of the watershed, collects records at a daily scale. The observations from this gauge are used as precipitation input to 6 segments (S54, S60-S63, and S78) located in the Occoquan catchment. Daily records collected at gauge LRTN 150 are disaggregated to hourly values for input into the HSPF model. The percentage of missing records provided in Table 3 does not include missing hourly records for gauge LRTN 150.

2.3.2 Satellite-based Precipitation Products

SPPs are used to estimate precipitation quasi-globally using a combination of remotely sensed microwave and infrared sensors. Three of the most widely used satellite-based precipitation products: (1) the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) (Huffman, et al., 2010); (2) the U.S. National Oceanic Atmospheric Administration (NOAA) Climate Prediction Center's (CPC) morphing technique (CMORPH) (Joyce et al., 2004); and (3) the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)-Cloud Classification System (CCS) (Hsu et al., 2010) were adopted for this study. This study compares these three SPPs of varying spatial and temporal resolutions as forcing inputs of the hydrologic model to simulate water quality in the Occoquan Watershed.

TMPA. The TRMM satellite mission was launched in 1997 and ceased in 2015. During that time TRMM-based products provided rainfall rates over the tropics and subtropics between 50ºN-50ºS (Huffman et al., 2007, 2010). Specifically, the TMPA algorithm combines both infrared and microwave sensor data to produce precipitation products with a spatial resolution of 0.25° and a temporal resolution of three hours in both real-time and latent calibrated versions (3B42-RT and 3B42-V7, respectively). This study uses the bias-adjusted 3B42-V7 product, which will be herein referred to as TMPA.

CMORPH. The CMORPH algorithm produces global precipitation analyses at high-resolutions using precipitation estimates derived exclusively from low orbiter satellite microwave observations transported via spatial propagation information obtained

from geostationary satellite data (Joyce et al., 2004). CMORPH interpolates precipitation features between consecutive microwave sensor samples aboard multiple spacecrafts including TRMM. CMORPH V1.0, used in this study, is bias corrected by matching raw data with the CPC daily gauge analysis over land and is available at a spatial and temporal resolution of 0.07° at equator and 0.5 hours, respectively, between 60° N and 60° S (Joyce et al., 2004).

***PERSIANN*.** PERSIANN-CCS, first released in 2003, is an algorithm based on infrared brightness temperature that extracts cloud features from a number of geostationary satellites operated by several agencies covering an area between 60ºN and 60ºS (Hong et al., 2007; Hsu et al., 1997; Sorooshian et al., 2000). Precipitation estimates are assigned using infrared cloud images from segmented cloud features such as statistics, textures, and geometry at brightness temperature thresholds. This information is used to obtain a relationship between brightness temperature and rainfall rates. PERSIANN-CCS (hereafter called PERSIANN) is available at a spatial resolution of 0.04° and a 30-minute temporal frequency.

This study compares these three SPPs of varying spatial and temporal resolutions as forcing inputs of the hydrologic model to simulate water quality in the Occoquan Watershed. The main characteristics of precipitation products used in this study are summarized in Table 3 including the percentage of missing records during the study period (2008-2012).

Table 3. Characteristics of precipitation products used in study

	Spatial Resolution	Temporal Resolution	Domain Coverage	Pixel/ Point Coverage	Missing Records
TMPA 3B42V7	0.25°	3 hour	50°N – 50°S	6	0.01%
CMORPH V1.0	0.07°	30 min	60°N – 60°S	48	0.35%
PERSIANN-CCS	0.04°	30 min	60°N – 60°S	133	0.48%
Rain Gauge Network	Point-source	1 hour	Watershed	15	<0.01%

2.4 Processing of Satellite-based Precipitation Product Data

The SPPs are processed by spatially averaging all pixels falling within the boundaries of each watershed segment. Since each SPP is provided in a different temporal resolution, all satellite data are matched to the hourly temporal scale. The spatially interpolated gauge data and spatially averaged satellite products at the temporally processed to the hourly scale are areal-weighted and segment-aggregated (AWSA) as input to the HSPF model. Since the Occoquan Watershed HSPF model output is set to output at a daily scale, precipitation data is subsequently aggregated to daily values for comparison of the SPPs (input data and output results) to the gauge-based data. Figure 3 presents the spatial distribution of cumulative precipitation over the 5-year study period for each SPP at their original pixel resolutions (left panels) and their corresponding AWSA value for each segment in the watershed (right panels) for (a) TMPA, (b) CMORPH, and (c) PERSIANN. Foremost, Figure 3 highlights the degree of detail for which each precipitation product is available (i.e., a total of 6, 48, and 133 precipitation values for TMPA, CMORPH and PERSIANN, respectively). In terms of cumulative precipitation within the watershed during the 5-year study period, the gauges recorded 4909 mm on average, whereas TMPA, CMORPH, and PERSIANN resulted in

5298 mm, 5267 mm, and 5834 mm, respectively, therefore PERSIANN showing the largest overestimation with respect to the reference dataset. TMPA shows the overall lowest variability of precipitation values, which can be attributed to its coarse spatial resolution and consequent low representativeness of the local precipitation distribution (only six pixels covering the entire watershed).

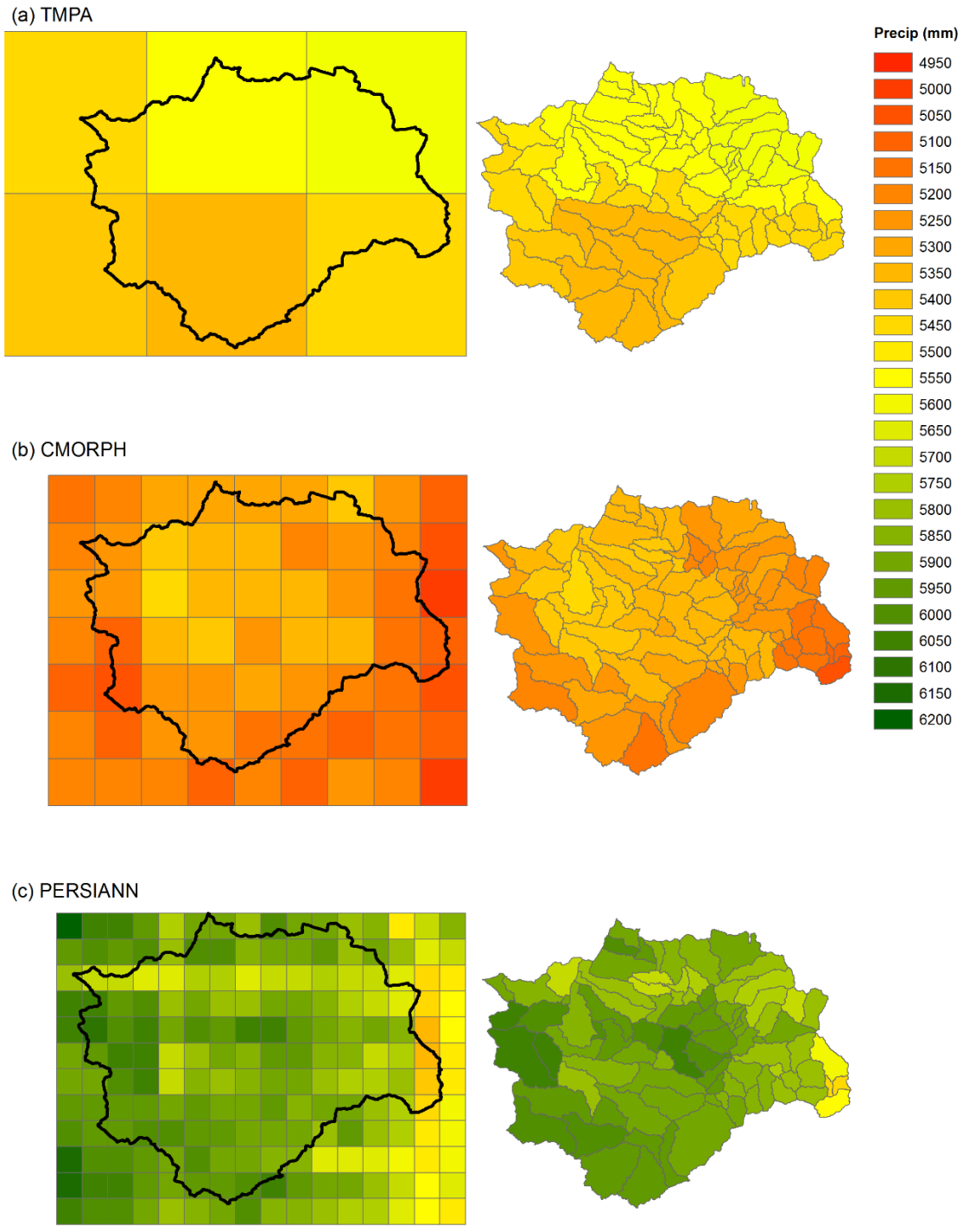

Figure 3. Five-year SPP (left panels) and aggregated segment (right panels) cumulative precipitation

CHAPTER 3. PERFORMANCE OF SATELLITE-BASED PRECIPITATION PRODUCTS IN SIMULATING WATER QUALITY INDICATORS

3.1 Introduction

The performance of both lumped and distributed hydrologic models is significantly influenced by a number of factors that contribute to uncertainty, e.g., observation errors, boundary or initial conditions errors, model or system errors, scale discrepancies, and unknown heterogeneity of parameters. Nevertheless, precipitation is widely accepted as the most influential meteorological input in hydrological and water quality investigations. Reliable and consistent hydrologic predictions for water quantity and quality depend on the accurate measurement of precipitation. The accuracy of precipitation data, including its intensity, duration, spatial patterns, and extent greatly influences the output of land surface and hydrologic models (Maggioni et al., 2013; Sorooshian et al., 2011; Zeng et al., 2018).

Due to the lack of ground-based monitoring systems and rain gauge networks, large scale hydrologic models often rely on remotely-sensed precipitation data from satellite sensors (Maggioni & Massari, 2018). Precipitation, whether from rain gauges or satellites, is characterized by spatial and temporal variability as well as measurement errors. Although ground-based instrument networks, including rain gauges and radars, represent the most direct surface rainfall measurements and often provide measurements with high temporal frequency, there are many limitations with these systems. Not only

are gauges restricted by point-scale observations, but they are also subject to false readings due to wind effects and evaporation. In addition to measurement errors, spatial interpolation of point-based observations adds uncertainty to the resulting gridded precipitation datasets (Habib et al., 2012a; Li et al., 2008; Maggioni et al., 2016; McMillan et al., 2012). The spatial distribution and density of gauges play important roles in the adequacy of measurement. A number of studies found that hydrologic model uncertainty is greatly impacted by sparse and irregular rain gauge networks and that uncertainty is reduced by increasing gauge density, or by optimizing the distribution pattern (Bardossy & Das, 2008; Girons Lopez et al., 2015; Moulin et al., 2009; Maggioni et al., 2017; Zeng et al., 2018). Conversely, ground-based radar networks often provide continuous spatial coverage with high spatial and temporal resolution, but they are also limited in accuracy by attenuation and extinction of signal, surface backscatter, brightband effects, and uncertainty of the reflectivity-rain-rate relationship (Anagnostou et al., 2010; Habib et al., 2012a; Porcacchia et al., 2017). In addition to being limited in accuracy, both ground-based gauges and radars are often inadequate due to sparse availability, especially outside of the United States.

Satellite-based precipitation products (SPPs) offer a viable alternative to ground-based data through continuous monitoring of precipitation with greater spatial coverage. Over the past twenty years several studies have compared SPPs to in-situ ground observations, evaluating the use of SPPs as a resource in hydrologic modeling. Nijssen and Lettenmaier (2004) noted that major sources of SPP uncertainty arise from algorithmic miscalculations, as well as instrumentation and sampling errors. Hydrological

modeling using SPPs as forcing input, particularly for simulations in large, complex watersheds involving heterogeneous land use (influencing soil moisture fluctuations and evaporation) and significant terrain changes, is not always adequate (Mei et al., 2014a, 2014b; Milewski et al., 2015; Seyyedi et al., 2014; Yang & Luo, 2014). For instance, SPPs often show an elevation-dependent bias with an underestimation of light precipitation at higher elevations (Maggioni et al., 2016). Mei et al. (2016a), Villarini et al. (2011), and Xu et al. (2014) found that the accuracy of hydrologic model simulations from SPPs is impacted not only by product resolution but also by storm severity and basin scale, where the most accurate simulation of hydrologic models using SPPs is when a combination of moderate precipitation magnitudes, finer product spatial resolutions, and larger basin scales are associated with the model.

Aside from the three principal measurement sources used to determine precipitation (ground-based gauge observations, ground-based radars, and satellite remote sensing), precipitation estimates can be obtained from atmospheric retrospective-analysis models (Beck et al., 2017). Two of the most widely accepted atmospheric reanalysis datasets with long-term coverage are (i) the European Centre for Medium Range Weather Forecasts (ECMWF) ERA-Interim, a dataset that uses a four-dimensional variational analysis, a revised humidity analysis, and variational bias correction for satellite data (Berrisford et al., 2011) and (ii) NASA's Modern Era Retrospective-analysis for Research and Applications (MERRA) (Rienecker et al., 2011). Moreover, there are products that combine a set of data from multiple methods (models and observations) and/or instruments (ground- and satellite-based). For instance, the Multi-Source

Weighted-Ensemble Precipitation (MSWEP) is a fully global dataset from 1979-2017 at 3-hour and 0.1° temporal and spatial scales, respectively. This dataset uses various gauge, satellite, and reanalysis data to provide precipitation estimates globally (Beck et al., 2019). Though re-analysis datasets and blended products have the potential for providing good quality precipitation estimates (Beck et al., 2017; Duque-Gardeazábal et al., 2018; Ma et al., 2018), these products were not considered in this study since the use of blended products would not allow for the differentiation of results among instruments and algorithms.

A plethora of research has been conducted over the past two decades on the use of remote sensing products for hydrologic model simulation and predictions intended for flood forecasting, riverine modeling, and climate studies in complex terrains (Gebregiorgis & Hossain, 2012; Hussain et al., 2018; Mei et al., 2016; Sharifi et al., 2018) and across various climatic regions (Bitew et al., 2012; Guo & Liu, 2016; Guo et al., 2015; Milewski et al., 2015; Verdin et al., 2016). Where current research falls short is the application of SPP data in hydrologic models specifically aimed at simulating and forecasting water quality.

It is well understood that the intensity, frequency, and duration of precipitation have significant and direct impacts on runoff quantity and thus streamflow. In turn, precipitation is also a major driver of water quality. A number of studies evaluating and quantifying predictions of stream water quality has demonstrated that spatial-temporal variations of water quality is strongly dependent on the spatial and temporal scales of analysis (Barakat et al., 2016; Bengraine & Marhaba, 2003; Chang, 2008; Jeznach et al.,

2017; Kang et al., 2010; Mei et al., 2014a, 2014b; Wang et al., 2011; Wunderlin et al., 2001; Xu et al., 2012). The relationship between water quality indicators and streamflow has been investigated using both parametric (distribution-dependent) and non-parametric (distribution-free) methods (Azhar et al., 2015; Barakat et al., 2016; Bu et al., 2010; Chang, 2008; Fovet et al., 2018; Johnson et al., 2012; Jung et al., 2016; Kisi & Ay, 2014; Nóbrega et al., 2018; Noori et al., 2010; Soler, 2017; von Freberg et al., 2017). All of these past studies demonstrated a direct impact in water quality as it relates to streamflow.

This study fills this gap in the literature, by evaluating the performance of a hydrologic model, forced with the SPPs, in estimating water quality indicators in the Occoquan Watershed, located in the northern Virginia suburbs of Washington D.C., United States. The overarching purpose of this study is to determine the suitability of SPPs as a substitute for rain gauge precipitation data for simulating selected water quality indicators. In this study, the satellite products are compared to the data collected by a dense rain gauge network to determine their capability of (1) accurately characterizing the spatial and temporal variability of precipitation within the study region, and (2) simulating water quantity and quality in the Occoquan Watershed. Section 3.2 introduces the methodology; Section 3.3 presents and discusses the results; and Section 3.4 highlights the main conclusions and limitations of this study.

3.2 Methodology

The first step in this study is to compare each SPP (TMPA, CMORPH, and PERSIANN) to the dense rain gauge network in the Occoquan Watershed in order to identify differences among the satellite products in the region. Then, the hydrologic model is forced with each satellite product and the traditional rain gauge dataset at the hourly scale to simulate streamflow and water quality indicators in the watershed during the 5-year study period (2008–2012). Model outputs are evaluated by comparing the results of daily simulated streamflow and selected water quality indicators from the three SPP simulations to that of the reference model simulation forced with rain gauge-based records.

3.2.1 Precipitation Analysis

As previously discussed, hourly precipitation data are areal-weighted and segment-aggregated (AWSA) for input into the model. The hourly data are then aggregated to the daily scale for comparison with the ground observations. Precipitation (P) is evaluated at the daily scale for consistency with the daily model output results. Several metrics are considered to quantify the performance of SPPs: root-mean-square error (RMSE), correlation coefficient (CC), standard deviation (σ), and bias (B). The RMSE, a measure of the differences between SPP and observed (gauge) values, combines the magnitude of errors in the estimation over time into a single measure of error. The CC is the average product of the deviation of variables from their respective means, divided by the standard deviation of those variables. Standard deviation is used to quantify the amount of variation in the data set. B measures the average tendency of satellite data to

either overestimate or underestimate the gauge measurements, whereas rB evaluates the relative difference in percentage between the estimated and the observed data. Optimum values of RMSE (eq. 3), CC, σ, and B are 1, 0, 0, and 0, respectively. The definitions of σ and B are provided in eqs. 4-6.

Equation 4. Correlation Coefficient

$$\boldsymbol{CC} = \frac{\sum_{i=1}^{n}(\boldsymbol{P_{o_i}} - \overline{\boldsymbol{P_o}})(\boldsymbol{P_{s_i}} - \overline{\boldsymbol{P_s}})}{\sqrt{\sum_{i=1}^{n}\left(\boldsymbol{P_{o_i}} - \overline{\boldsymbol{P_o}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(\boldsymbol{P_{s_i}} - \overline{\boldsymbol{P_s}}\right)^2}}$$

Equation 5. Standard Deviation

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - \bar{P})}$$

Equation 6. Bias

$$B = \frac{\sum_{i=1}^{n}(P_{s_i} - P_{o_i})}{n}$$

Where n is the total number of events, P_{o_i} is the ith observed rain gauge precipitation and P_{s_i} is the ith satellite-based precipitation value, $\bar{P}$ is the corresponding mean value. CC is dimensionless whereas σ and B are in mm/d. The RMSE, CC, and σ are summarized using a Taylor diagram (Taylor, 2001), which shows the overall skill associated with each SPP in relation to the gauge-based dataset (Section 3.3.1).

Probability of detection (POD), false alarm rate (FAR), and critical success index (CSI) are used to assess the detection capability of each SPP based on the AWSA daily

precipitation (P). POD corresponds to the ratio of correct detection of the SPP to the overall occurrence of P observed. Conversely, FAR indicates the number of cases when P is estimated by SPP; however, no P is observed. CSI is a measure of events successfully detected by the SPP to the total number of events observed that are either made or needed (Schaefer, 1990). These metrics all range from 0 to 1, with the accuracy of the SPP increasing as the FAR approaches 0 and the POD and CSI approach 1. Optimum values of POD, FAR, and CSI are 1, 0, and 1, respectively. For this study, the P threshold is set to 0.254 mm/day, which is the threshold of the rain gauges. POD, FAR, and CSI are calculated using eqs. 7-9, where hits, misses and false alarms represent the total number of SPP observations that either correctly detect an event, miss, an event, or incorrectly detect an event, respectively, in relation to the gauge observation (Table 4).

Equation 7. Probability of Detection

$$POD = \frac{Hits}{Hits+Misses}$$

Equation 8. False Alarm Ratio

$$FAR = \frac{False\ alarms}{Hits+False\ alarms}$$

Equation 9. Critical Success Index

$$CSI = \frac{Hits}{Hits+False\ alarms+Misses}$$

Table 4. Contingency table comparing hourly rain gauge observations and SPP values

		Forecast (satellite)	
		≥ 0.254 mm	**< 0.254mm**
Observed (gauge)	**≥ 0.254mm**	Hits	Misses
	< 0.254mm	False alarms	Correct negatives

Detection metrics are summarized using a performance diagram (Roebber, 2009) that presents the POD, success ratio (1-FAR), CSI and B associated with each SPP in relation to gauge-based data for the 5-year study period (Section 3.3.1).

3.2.2 Streamflow and Water Quality Indicator Analysis

It is well understood that there is a direct relationship between precipitation and streamflow. Research has also shown a significant and quantifiable correlation between atmospheric conditions and water quality indicator response (Chang et al., 2015; Gelca et al., 2016; Jeznach et al., 2017; Johnson et al., 2012; Murdoch et al., 2000; Soler et al., 2007; Thorne & Fenner, 2011).

Three statistical metrics (CC, RMSE, and B) are used to evaluate the performance of the model in simulating streamflow and the four water quality indicators. The model simulation forced with rain gauge observations is used as a benchmark to assess the performance of the simulations forced with the three different SPPs. Uncertainties are quantified for simulated streamflow and selected water quality indicators at six evaluation points (S26, S27, S34, S47, S79, and S86) of varying drainage area within the watershed, discussed in detail in Section 3.3.2.

3.3 Results and Discussion

3.3.1 Satellite-based Precipitation Products

The SPPs are compared to gauge observations for all 87 segments of the watershed. This comparison is based on the interpolated and aggregated AWSAs rather than a pixel-to-point comparison, since the main focus of this study is to assess the impact of using different precipitation forcing datasets on simulating streamflow and water quality indicators in the watershed. AWSA satellite-based precipitation records aggregated to daily values (P) are compared to daily aggregated gauge-based records. RMSE, CC, and σ are calculated and plotted in a Taylor diagram to assess the performance of each SPP with respect to the gauged-based data (Figure 4a). For the three products, correlation coefficients are concentrated between 0.26 and 0.60 and RMSEs range from 7.25 to 9.17 mm/day. The standard deviations of both TMPA and CMORPH are 8.08 mm/day, which is most likely due to the fact that these products use very similar input retrieval data in their algorithms. The standard deviation of PERSIANN, however, is lower (6.90 mm/day), which may be due to the fact that PERSIANN only uses one source (infrared) of observation data, whereas TMPA and CMORPH use both passive microwave and infrared observations. The best overall performance is from the CMORPH product, followed closely by TMPA. PERSIANN shows overall relatively inferior performance with low correlation values against gauge-based records. While both TMPA and CMORPH show good agreement with gauge-based records, both products overestimate P magnitude (rB=20.6% and 13.0%, respectively).

The Performance diagram (Figure 4b) suggests that both TMPA and CMORPH have lower false alarm rates (0.30 and 0.33, respectively) compared to PERSIANN, which exhibits a false alarm rate of 0.46. However, PERSIANN presents a higher probability of detection at 0.71 compared to TMPA (0.57) and CMORPH (0.68), respectively. B is relatively low for both TMPA and CMORPH at 0.51 and 0.39 mm/d, respectively, while PERSIANN has a slightly higher bias of 0.70 mm/d. TMPA and CMORPH exhibit similar performances, whereas simulation results derived from PERSIANN are quite different and show larger error metrics. This is most likely due to the fact that CMORPH and TMPA merge observations from passive microwave and infrared sensors, whereas PERSIANN only uses infrared observations.

Table 5. Statistical error and performance characteristics of AWSA SPPs compared to gauge-based observations

Product	PCC	SD (mm/d)	B (mm/d)	rB (%)	RMSE (mm/d)	POD	FAR	CSI
TMPA 3B42V7	0.54	8.08	0.51	20.60	8.20	0.57	0.30	0.46
CMORPH	0.60	8.08	0.39	13.01	7.25	0.68	0.33	0.51
PERSIANN-CCS	0.26	6.90	0.70	24.29	9.17	0.71	0.46	0.44

Results corroborate what was previously observed in the scientific literature. As summarized by Maggioni et al. (2016), mean errors and detection capabilities of satellite-based precipitation estimation vary across precipitation products and different regions of the world. A positive bias and season-dependent magnitude are reported over the contiguous United States for both CMORPH and PERSIANN-CSS. TMPA 3B42 is reported as having the lowest bias when compared to other satellite datasets.

Furthermore, TMPA 3B42 exhibits a low probability of detection and low false alarm rate, whereas CMORPH has a higher probability of detection and also a larger number of false alarms. PERSIANN has both the highest false alarm rate and highest probability of detection compared to other products (Ebert et al., 2007; Yang & Luo, 2014). Results of this study validate that PERSIANN has the highest false alarm rate and probability of detection compared to TMPA and CMORPH within this study region.

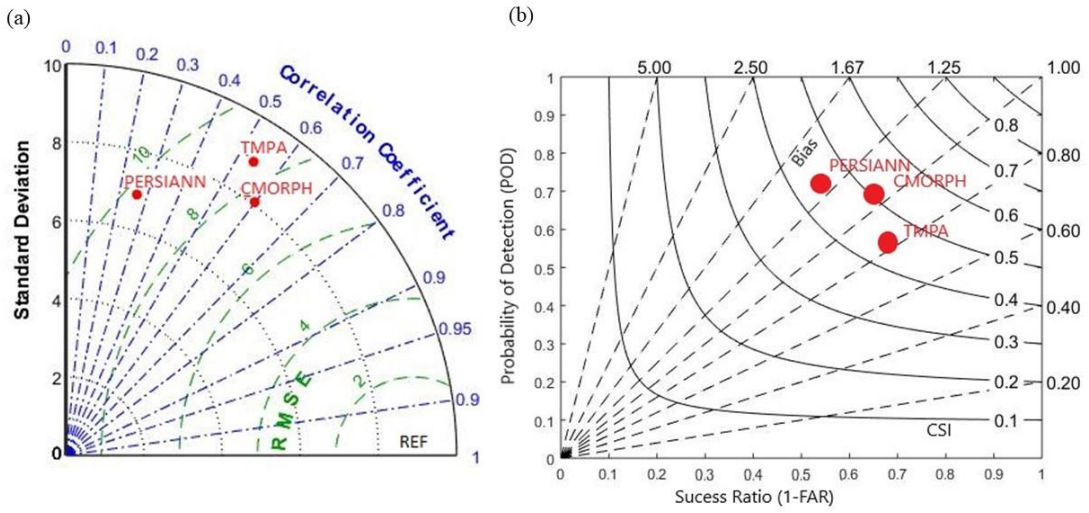

Figure 4. (a) Taylor diagram and (b) Performance diagram of segment-aggregated daily precipitation

3.3.2 Error Analysis of Simulated Streamflow and Water Quality Indicators

This section assesses the skills (CC, RMSE, and B) of the complex-linked HSPF model of the Occoquan Watershed in terms of both Q and water quality indicators (TW, TSS, DO, and BOD). Specifically, simulations forced with the three different SPPs

(TMPA, CMORPH, and PERSIANN) are evaluated with respect to the reference run forced with gauge-based P data. Aside from altering precipitation input, no other inputs, parameters, or model boundary conditions are modified in this experiment. Each of the seven catchments that comprise the watershed are fairly uniform in topography, soil characteristics, and land use conditions; therefore, it is expected that these conditions will not have a significant impact to the uncertainty analysis based on drainage area size between SPPs and gauge-based simulations. The goal of this analysis is to quantify uncertainties in simulated Q and selected water quality variables at the six evaluation points (S26, S27, S34, S47, S79, and S86) in the watershed. These locations are chosen based on the representative basin scale ranging between 17.8 and 498.8 sq. km (refer to Table 1).

Figure 5 shows that, for Q, CMORPH outperforms TMPA and PERSIANN in terms of all skill metrics. When compared to the gauge-forced simulation, CMORPH presents the highest correlation at all evaluation points with CCs ranging from 0.47 to 0.75, nominally increasing as area increases, with a few exceptions. These exceptions are most likely due to differences in the hourly precipitation records of CMORPH, as well as TMPA and PERSIANN, in comparison to the rain gauge records in the Upper Broad Run catchment draining to S86/ST70. B and RMSE associated with Q also increase as area increases, with CMORPH marginally outperforming TMPA. PERSIANN shows the worst overall skill consistently for all basin scales. All three products overestimate daily Q in comparison to gauge-simulated Q, which is also observed in the precipitation analysis. The increase of B and RMSE based on basin scale is expected since the Q, both

simulated and observed at these points, increases and generates the potential for greater residuals.

In HSPF, most water quality variables are highly dependent on precipitation and its associated sediment and water flows, i.e., sediment loading and overland runoff flow (Duda et al., 2012). Satellite-based simulations of TSS showed the lowest skill among the water quality indicators evaluated in this study. Again, CMORPH showed the best performance among the SPPs, outperforming both TMPA and PERSIANN. CC ranged from 0.20 to 0.66, 0.29 to 0.75, and 0.04 to 0.40 for TMPA, CMORPH, and PERSIANN, respectively. The lowest CCs for all three products are at point S86, which also corresponds to the lower skill detected in the simulated Q. Each SPP overestimates TSS, with the exception of TMPA at point S47, though both TMPA and CMORPH, with a B of -2.4 to 5.8 m^3/s, and 0.46 to 4.9 m^3/s, respectively, have a considerably lower B than PERSIANN, which ranges between 2.3 and 15.8 m^3/s. For TSS, the basin extent seems to play a factor, showing generally decreasing CC and increasing B and RMSE trends, as the drainage area increases. HSPF predicts sediment loading rates based on channel processes of deposition, scour, and transport that in turn determine both the total sediment load and outflow sediment concentrations in streamflow (Duda et al., 2012). Land use and soil type in the model are the major drivers for sediment load, calculated as total sediment for simulation of in-stream processes. Since land use is constant between the gauge-based and satellite-based simulations, there is no uncertainty or error associated with that input parameter. Therefore, precipitation, and consequently Q, is the major driver for TSS concentration differences presented in this analysis.

Table 6. Statistical error characteristics of simulated output for streamflow

Model Eval. Location	Product	PCC	SD (m^3/s)	RMSE (m^3/s)	MSE (m^3/s)	rB (%)	B (m^3/s)
S27	TMPA	0.41	0.82	0.90	0.81	22.98	0.05
S27	CMORPH	0.47	0.75	0.77	0.60	9.29	0.02
S27	PERSIANN	0.27	0.79	0.96	0.91	50.74	0.10
S79	TMPA	0.55	2.68	2.65	7.00	35.22	0.23
S79	CMORPH	0.67	2.52	2.13	4.52	18.98	0.13
S790	PERSIANN	0.20	2.64	3.37	11.36	67.32	0.44
S86	TMPA	0.51	4.22	4.45	19.79	37.93	0.49
S86	CMORPH	0.63	3.91	3.58	12.83	25.41	0.33
S86	PERSIANN	0.20	3.84	4.96	24.58	74.50	0.95
S26	TMPA	0.58	7.27	6.93	48.03	32.06	0.65
S26	CMORPH	0.71	6.81	5.42	29.33	16.77	0.34
S26	PERSIANN	0.24	6.62	8.24	67.92	57.75	1.17
S47	TMPA	0.72	10.78	8.75	76.51	33.57	1.30
S47	CMORPH	0.75	10.35	7.72	59.63	22.11	0.86
S47	PERSIANN	0.32	11.45	13.85	191.77	79.10	3.07
S34	TMPA	0.69	13.81	11.35	128.81	28.47	1.45
S34	CMORPH	0.74	13.17	9.79	95.80	17.03	0.87
S34	PERSIANN	0.32	14.63	17.54	307.62	71.65	3.64

Simulated in-stream TW is well captured by all three SPP simulations with high correlations, ranging from 0.97 to 1.0, low biases, ranging from -0.21 to 0.24 ºC, and low RMSEs, ranging from 3.3 to 9.9 ºC. While stream temperature is impacted by precipitation, and thus Q, it is most closely correlated to ambient air temperature and

therefore shows the lowest variation and highest skills among all output variables evaluated in this study. Generally, all three SPPs slightly underestimate TW, with a B of -0.16 to -0.06 ºC for TMPA, 0.17 to 0.05 ºC for CMORPH, and -0.21 to 0.03 ºC for PERSIANN. Among all six evaluation points, only S86 overestimates TW consistently among all three products (0.00, 0.05 and 0.24 ºC, respectively for TMPA, CMORPH, and PERSIANN), though CMORPH also overestimates TW at S34 and PERSIANN at S47.

Both DO and BOD show a large spatial variability among the three catchments having separate but interlinked HSPF models. CC for DO is relatively high, ranging from 0.86 to 0.99, 0.83 to 0.99, and 0.78 to 0.98 for TMPA, CMORPH, and PERSIANN, respectively, in the Upper Bull Run and Upper Broad Run catchments. However, a decreasing skill for TMPA (0.60 and 0.78), CMORPH (0.62 and 0.75), and PERSIANN (0.63 and 0.74) in the Cedar Run catchment at locations (S47 and S34) is detected. Both B and RMSE show similar results of increased error with increasing drainage area in the Cedar Run catchment. B for all three products ranged from 0.02 to 0.68 mg/l, -0.01 to 0.76 mg/l, and 0.02 to 1.16 mg/l, for TMPA, CMORPH, and PERSIANN, respectively, while RMSE ranges from 0.71 to 2.57 mg/l, 0.69 to 2.60 mg/l, and 0.73 to 3.03 mg/l, for TMPA, CMORPH, and PERSIANN, respectively. In comparison to Q and TSS, the variability in the DO metrics is notably smaller, indicating that precipitation has moderate impacts on DO compared to other water quality indicators evaluated in this analysis. Similar to DO, BOD has a satisfactory CC in the Upper Bull Run catchment, but low CC in the Cedar Run and Upper Board Run catchments, especially when PERSIANN is used as input precipitation. The model performs well in terms of B and RMSE in the Upper

Bull Run and Upper Board Run catchments, but has low skill in the Cedar Run catchment (S47 and S34), with the exception of TMPA, which outperforms both CMORPH and PERSIANN in this catchment.

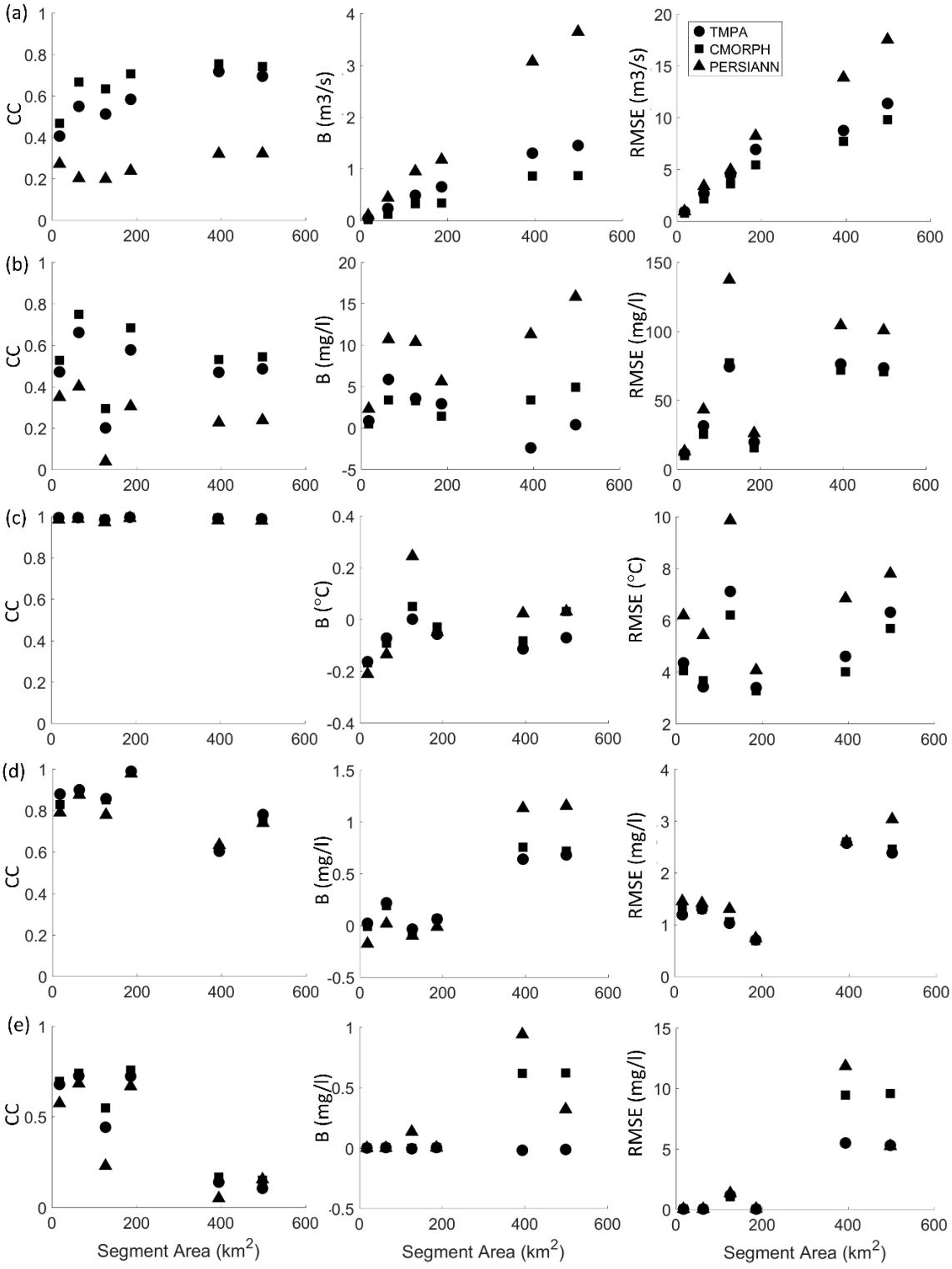

Figure 5. Error plots by evaluation point drainage area between daily gauge- and SPP-simulations for (a) Q, (b) TSS, (c) TW, (d) DO, and (e) BOD

3.3.3 Temporal Analysis of Simulated Streamflow and Water Quality Indicators

As shown in the previous section, the positive bias in the SPPs is propagated into simulated Q, but not in all the water quality indicators. As part of this study the temporal variability and seasonality of these variables is also analyzed. Figure 6 shows time series of daily P, Q, and water quality indicators (TSS, TW, DO, and BOD) at evaluation point S27 (the smallest drainage area among all areas considered in this study) for one year (2011) of the 5-year study. CMORPH and TMPA are able to decently capture the timing of the Q peaks, but tend to underestimate some peaks. While peaks in Q generally coincide with precipitation, it is evident that precipitation during the summer months (June through August) does not directly translate into Q. This may be due to a seasonal distribution within the model associated with an increased infiltration component. In HSPF, precipitation runoff is divided into surface runoff and infiltration. Infiltration then comprises of interflow, upper zone soil moisture storage, and percolation to lower zone soil moisture and groundwater storage. Within the HSPF model, increasing infiltration will reduce immediate surface runoff (including interflow) and increase the groundwater component, which is observed in Figure 6 during summer months. It is also evident from Figure 6 that PERSIANN overestimates precipitation in January and February, which is then rendered into an overestimation of both low and peak Q during cool months. Another factor impacting the observed decrease in Q during summer months may be due to the fact that higher evapotranspiration rates are typically seen in warmer months from increased ambient air temperature, which is sufficient to overcome the impact of

increased humidity since drier soil tends to have increased infiltration rates compared to moister soil.

There is a direct relationship between Q and TSS concentrations (mg/l), with peaks well represented in both TMPA and CMORPH simulations. As with Q, TSS tends to be low throughout summer months aside from a few instances where intense P is captured and translated into stream peak flow. Because PERSIANN overestimates Q in winter months, it in turn overestimates TSS concentrations during these months, since TSS is a flow-dependent variable, highly interrelated with Q. In-stream temperature is naturally dependent on ambient air temperature conditions, showing higher values in summer months and lower values in winter months following the same TW trends as the gauge-based simulation.

As expected, concentrations of DO follow an inverse pattern with respect to TW, lower during warmer seasons and higher during cooler seasons. Simulated DO concentrations for all SPPs are fairly consistent with gauge-based simulations. Similar to gauge-based simulations, a strong flux in concentrations of DO is seen in warmer months which overlaps with Q low-flow periods. Comparatively, as shown in Figure 6, it appears that PERSIANN displays the greatest number of daily fluctuations of BOD in comparison to TMPA and CMORPH. This fluctuation is especially true during January and February, which is likely due to the overestimation of precipitation during these months when compared to other products. Conversely, simulated BOD concentrations (mg/l) increase in warmer periods following patterns associated with TW, yet this indicator also appears to be influenced by TSS concentrations during peak events.

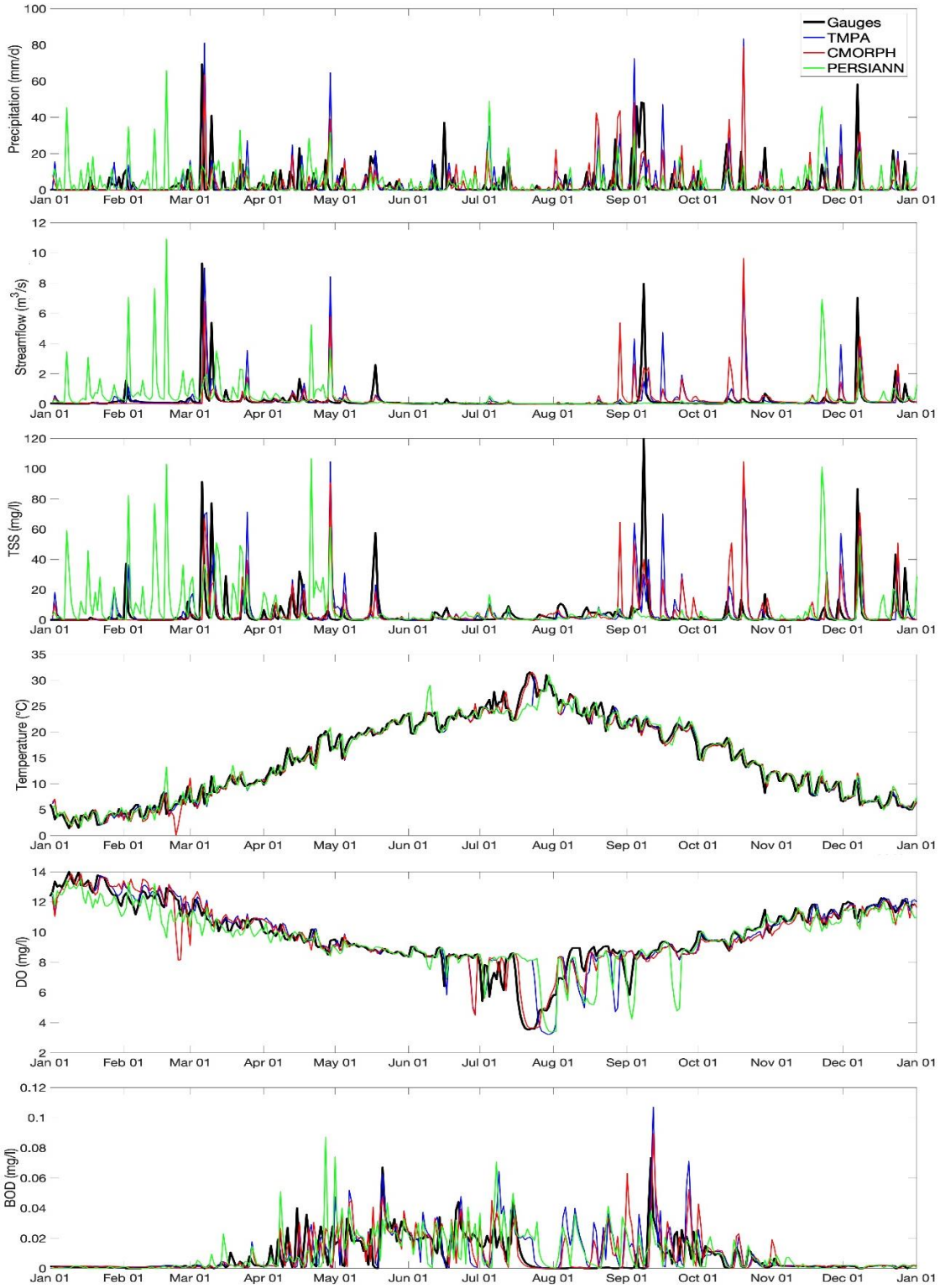

Figure 6. Example daily time series of AWSA P and simulated output for Q, TSS, TW, DO, and BOD

3.4 Conclusions

This work investigated for the first time the potential of using SPPs for water quality monitoring and predictions. Three SPPs (TMPA, CMORPH, and PERSIANN) of different spatial resolutions are compared to gauge-based records over a 5-year period (January 2008 – December 2012) across the Occoquan Watershed. This study provides a comprehensive analysis of the errors associated with Q and four water quality indicators (TW, TSS, DO, and BOD) simulated by forcing a hydrological model (the complexly-linked HSPF) with the three SPPs. These simulations are compared to a reference run, forced with gauge-based observations. The major findings of this research are summarized as follows:

- All three SPPs show moderate skill with respect to the daily gauge-based dataset. CMORPH shows the best overall performance, followed closely by TMPA. PERSIANN shows overall relatively inferior performance with low correlation with the gauge-based records. While both TMPA and CMORPH have good agreement with reference dataset, both products overestimate P magnitude (rB=20.6% and 13.0%, respectively). In terms of cumulative P during the 5-year study period, PERSIANN shows the largest overestimation with respect to the gauges. TMPA shows the overall lowest variability, which can be attributed to its coarse spatial resolution and consequent low representativeness of the local precipitation distribution (only six pixels covering the entire watershed).

- Both TMPA and CMORPH present lower false alarm rates compared to PERSIANN across the Occoquan Watershed. However, PERSIANN presents a higher probability of detection compared to the other two products, especially during winter months.
- Satellite-based simulations of Q show that CMORPH outperforms both TMPA, and PERSIANN when compared to the reference simulation forced with gauge observations. CMORPH has the highest skill for daily Q across all evaluation points with a CC ranging from 0.47 to 0.75, nominally increasing as drainage area increases. CMORPH and TMPA are able to capture the timing of the Q peaks well, but tend to underestimate some peaks. While peaks in Q generally coincide with precipitation, it is evident that P during the summer months (June through August) do not directly translate to Q.
- SPP simulations of TSS show the lowest skill among the water quality indicators evaluated in this study. CMORPH shows the best performance among the SPPs, which generally all overestimate TSS, with a few exceptions. The seasonal analysis confirms that peaks of TSS are fairly well represented for both TMPA and CMORPH simulations.
- Simulated TW is well captured by all three SPP simulations with high correlations, low biases, and low RMSEs. While TW is impacted by precipitation, and thus Q, it is most closely correlated to ambient air temperature and therefore presents the lowest variation and highest skills among all output variables evaluated in this study.
- Both DO and BOD show a large spatial variability among the three catchments in this study. Skills of all three SPPs are high for DO in the Upper Bull Run and Upper

Broad Run catchments. However, a decreasing skill for each product is detected in the Cedar Run catchment. In comparison to Q and TSS, the variability in the DO metrics is notably smaller, indicating that precipitation has moderate impact on DO than on other water quality indicators evaluated in this analysis. Simulated DO concentrations for all three SPPs are fairly consistent with gauge-based simulations. Similar to gauge-based simulations, a strong flux in concentrations is seen in warmer months, which also overlaps with Q low-flow periods.

- The SPPs show satisfactory skills for BOD in the Upper Bull Run catchment, but low CC in the Cedar Run and Upper Board Run catchments, especially when PERSIANN is used as input precipitation. The model performs well in terms of B and RMSE in the Upper Bull Run and Upper Board Run catchments, but has low skill in the Cedar Run catchment. Simulated BOD concentrations (mg/l) increase in warmer periods and drop during cooler periods following patterns associated with TW, yet this indicator also appears to be influenced by TSS concentrations during peak events.

Overall, results indicate that the spatiotemporal variability of the SPPs, along with the algorithms used by these products to estimate precipitation, have a quantifiable impact not only on streamflow, but also on water quality output from the hydrology model. However, it should be noted that there are limitations to this study. Foremost, this analysis was conducted in a single location, situated in the suburban Washington, D.C. area, characterized by a temperate climate and mild topographic variation. Additionally, this study only considered one hydrology/water quality model in this analysis, i.e., HSPF. While this model is well calibrated and has been validated to observation results, a

different model may respond differently to changes in streamflow and water quality constituents resulting from forcing precipitation inputs. Nevertheless, this work represents a first attempt to utilize SPPs for water quality modeling, which could be of critical importance in areas of the world where rain-gauge networks or monitoring stations are either sparse or not available altogether.

CHAPTER 4. ERROR PROPAGATION FROM SATELLITE-BASED PRECIPITATION TO SIMULATED WATER QUALITY INDICATORS

4.1 Introduction

It is well understood that precipitation is the most important forcing input in a hydrologic model, as it influences both watershed hydrology and water quality processes. Land surface and hydrologic models are greatly influenced by the accuracy of input precipitation data including its spatial and temporal distribution, intensity, and duration (Sorooshian et al., 2011; Zeng et al., 2018; Hazra et al., 2019). While the traditional approach has been to measure precipitation using ground-based rain gauges, the use of satellite-based precipitation products (SPPs) in hydrologic modeling has been gaining more popularity due to their continuous geographic coverage with high spatial and temporal resolution. Whilst these products offer a viable resource, seasonal precipitation patterns, storm type, resolution of measurement, and background surface all have an influence on the performance of SPPs and thus impact the output of hydrologic models (Maggioni & Massari, 2018). Major sources of uncertainty with SPPs emerge from inaccuracy of instrumentation, sampling errors, and algorithmic miscalculations (Nijssen & Lettenmaier, 2004). In addition to errors associated with precipitation input, uncertainty in hydrologic modeling may come from a number of other sources such as input parameter heterogeneity, model structure and algorithm errors, and boundary condition errors.

Understanding the uncertainty associated with individual SPPs is crucial when assessing their performance as input into a hydrologic model. SPPs combine information mainly from geostationary infrared (IR) and/or low orbiting passive microwave (PMW) satellites. Many studies have documented the performance of SPPs suggesting that seasonal patterns and types of precipitation can impact the quality of the retrieval. For instance, thermal IR algorithms may miss light stratiform precipitation events and snow cover (Hong et al., 2007; Maggioni & Massari, 2018). Ebert et al. (2007) reported that in temperate climates SPPs generally perform better with convective storms during the warm season and that IR-based SPPs decline in accuracy with stratiform precipitation. Conversely, PMW algorithms may underestimate heavy precipitation events often associated with convective storms.

Numerous past studies have discussed how to quantify uncertainty of SPPs and the relationship between forcing precipitation estimated by SPPs and their simulated streamflow results through a hydrologic model (Hossain & Anagnostou, 2004; Flack et al., 2015; Gebremichael et al., 2011; Guo et al., 2016; Hong et al., 2006; Maggioni et al., 2013; Mei et al., 2016b; Nikolopoulos et al., 2010, 2013; Seyyedi et al., 2015; Yu et al., 2016). Results documented that output simulations derived from SPPs can be affected by event intensity, product resolution, and basin scale, noting that moderate precipitation magnitudes, larger domain scales, longer time integration and finer spatial and temporal resolutions of SPPs generally improve the accuracy in simulating streamflow (Maggioni & Massari, 2018).

Several studies have also examined the propagation of error from SPPs to streamflow, indicating that the output error is highly dependent on the uncertainty associated with the forcing precipitation dataset. Errors generated from SPP resolution, both spatial and temporal, are known to have a significant impact on the error in streamflow. Guetter and Georgakakos (1996), Nijssen and Lettenmaier (2004), Flack et al. (2015), Hossain and Anagnostou, (2004), Nikolopoulos et al. (2013), Maggioni et al. (2013), Mei et al. (2016), Seyyedi et al. (2015), Vivoni et al. (2007), Yu et al. (2016), and others all found that errors associated with precipitation input can have an amplifying transformation on simulated streamflow uncertainty with relative errors magnified up to three times greater than input errors. Maggioni et al. (2013) reported that error metrics of ensemble precipitation for three SPPs in the Tar-Pamlico basin in the southeast United States (U.S.) had no dependency on basin scale for relative bias, although relative root-mean-square-error (RMSE) decreased as a function of scale. Their results also showed that biases doubled, while RMSE decreased as a function of basin scale between input and output simulations. Vergara et al. (2014) also reported an amplified bias and a dampened RMSE in the same basin. Another study investigated error corrections to SPPs in streamflow simulations through an ensemble error model in the Tocantins-Arguaia basin, Brazil (Falck et al., 2015). Results there indicated that mean errors and RMSEs decreased as a function of basin scale, but no scale dependence on the precipitation-to-streamflow error propagation was observed. Yu et al. (2016) quantified uncertainty propagation of rainfall forecasts into streamflow through a distributed hydrologic model in the Kii Peninsula of Japan. Their results coincided with prior studies demonstrating a

decrease in uncertainty by basin scale. Ehsan Bhuiyan et al. (2019) investigated the propagation of precipitation uncertainty to simulated surface runoff, subsurface runoff, and evapotranspiration using five SPPs, showing that uncertainties in modeled surface runoff were strongly sensitive to precipitation uncertainty. In addition to basin scale, other factors have been cited to impact error propagation of streamflow including complex terrain and elevation (Nikolopoulos et al., 2010; Mei et al., 2016b), hydrologic model type and complexity (Hostache et al., 2011; Zhu et al., 2013), and seasonality (Gebregiorgis & Hossain, 2013; Mei et al., 2016b; Vivoni et al., 2007).

Research associated with the use of satellite and reanalysis precipitation products for water quality modeling is very limited. Neal et al. (2012) found that soil-water chemistry has variability at the local scale, translating into a range of responses in the chemistry of localized runoff, and thus streamflow, further indicating that a high temporal frequency and spatial resolution is needed for modeling and simulating streamflow processes including sediment. Himanshu et al. (2017) evaluated the performance of TMPA 3B42V7 for predicting suspended sediment loads in two watersheds in south India using a machine learning technique and found moderate prediction efficiency. Stryker et al. (2017) used the North American Regional Reanalysis data for simulating suspended sediment loads and concentrations in the Mad River watershed located in Vermont, U.S.; however, the main objective of this study was to evaluate model performance for sediment simulations as opposed to uncertainty of precipitation inputs. A recent study by Ma et al. (2019) investigated the use of two SPPs and one reanalysis product in the Lancang River Basin in southwest China to assess their performance in simulating

streamflow and suspended sediment using the Soil and Water Assessment Tool (SWAT) model. They found that, at the monthly timestep, both SPPs were at better estimating precipitation than the reanalysis product and also both SPPs had a good capability of modeling monthly streamflow and sediment loads. While this study evaluated spatial and temporal sediment yield, it neither considered the spatial and temporal variability of SPPs nor evaluated the propagation of uncertainty between precipitation input and model output.

While there is substantial research investigating the association between precipitation and water quality response (e.g., Bezak et al., 2017; Cheng et al., 2015; Gelca et al., 2016; Hayashi et al., 2004; Jeznach et al., 2017; Johnson et al., 2012; Murdoch et al., 2000; Neal et al., 2012; Soler et al., 2007; Thorne & Fenner, 2011), only a few studies investigated how the spatial and temporal differences of SPPs may impact the simulation and forecasting of water quality (Ma et al., 2019; Solakian et al., 2019). Furthermore, there is a gap in the literature assessing the propagation of errors in input SPP (at different resolutions) to simulated water quality indicators. Initial research conducted by Solakian et al. (2019) showed that spatial and temporal differences in SPP resolutions, along with the algorithms used to estimate precipitation magnitude, had an impact on modeled streamflow and water quality indicators in the Occoquan Watershed, Virginia, U.S. It was also noted that the seasonality dissimilarities observed in SPPs may translate into seasonal differences of simulated streamflow and water quality indicators.

This study builds upon previous work, providing a comprehensive evaluation of the seasonal skill of three different SPPs, of varying native spatial and temporal

resolutions, compared against observations from a dense rain gauge network over the Occoquan Watershed. The three SPPs evaluated in this study are used as forcing input into a gauge-calibrated hydrologic and water quality model to simulate streamflow (Q) and three water quality indicators, i.e., total suspended solids (TSS), stream water temperature (TW), and dissolved oxygen (DO), at six locations within the watershed. The skill of the SPP-based model simulations is then compared to gauge-based simulations over a 5-year study period (2008-2012). The propagation of error from input precipitation to each of the three products is investigated by basin scale and on a seasonal basis. Section 4.2 discusses the methods employed in this analysis to assess the uncertainty of SPPs, simulated output, and the propagation of error. Section 4.3 presents, interprets and discusses the results of this study, while Section 4.4 provides conclusory remarks of notable findings.

4.2 Methodology

Firstly, the three SPPs are evaluated against reference precipitation measurements from the rain gauge network in the Occoquan Watershed. Secondly, the HSPF model is forced with the three SPPs to simulate output of Q, TSS, TW, and DO. Model output are evaluated at six evaluation points by comparing the three SPP-forced simulations to that forced with rain gauge-based records both temporally, by season, and spatially, by basin scale. Thirdly, the propagation of error from model input to simulated output is investigated.

Hourly precipitation data from SPPs are areal-weighted and segment-aggregated (AWSA) for input into the hydrologic model. To process the SPPs as AWSA input,

pixels overlaying segment boundaries are spatially aggregated. Next, SPPs are temporally matched to the hourly temporal resolution which is the resolution of the rain gauge data. Figure 7 presents the spatial distribution of rain gauges (7a) including cumulative precipitation of each watershed segment over the 5-year study period recorded by the gauge network. Figure 7 (b-d) highlights the cumulative 5-year AWSA precipitation measured for each segment for (b) TMPA, (c) CMORPH, and (d) PERSIANN including the spatial resolution of SPP pixels at 0.25°,0.07°, and 0.04°, respectively. The average cumulative AWSA precipitation estimated by the three SPPs over the watershed during the 5-year study period moderately overestimates the one recorded by the rain gauges (4909 mm) with values of 5298 mm, 5267 mm, and 5834 mm for TMPA, CMORPH, and PERSIANN, respectively.

The hydrologic model developed for the Occoquan Watershed processes precipitation input at the hourly resolution and simulates output at the daily resolution. Thus, hourly precipitation data are aggregated to the daily scale for a uniform comparison with the model output. Specifically, error and performance metrics of AWSA daily precipitation (P) are computed on a seasonal basis for all the 87 watershed segments. Seasons are defined as follows: December-January-February (winter), March-April-May (spring), June-July-August (summer), and September-October-November (fall).

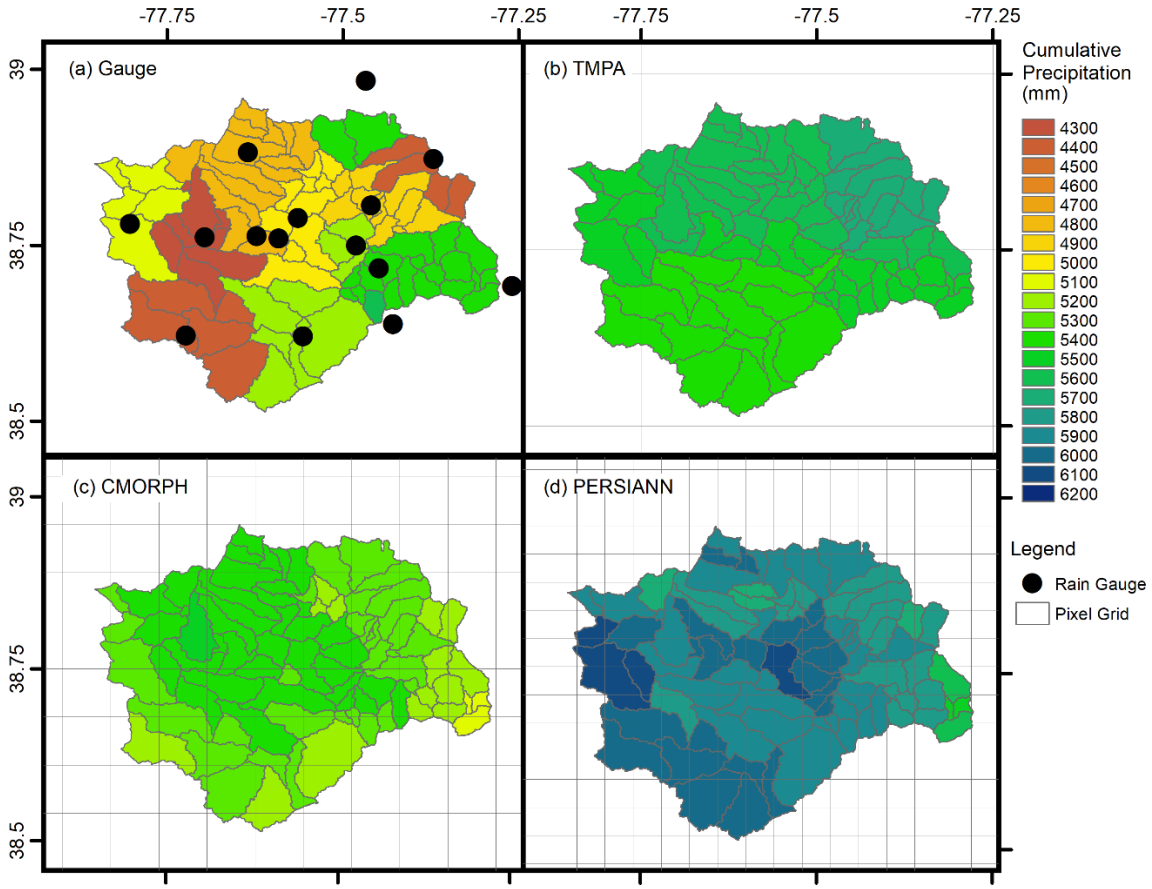

Figure 7. Cumulative 5-year AWSA precipitation for each segment in the Occoquan Watershed

Aside from altering precipitation input, no other input, parameters, or model boundary conditions are modified in this experiment. The goal of this analysis is to quantify uncertainties in simulated streamflow and select water quality indicators by season, and by basin scale at the six evaluation points (S26, S27, S34, S47, S79, and S86) in the watershed. These locations are chosen based on the representative basin scale (drainage area size) and model output locations. Stream water quality is not only influenced by precipitation, but also by many other factors including atmospheric deposition, rainfall chemistry, vadose zone leaching, groundwater chemistry, stormwater

runoff, streambank sediment transport, and anthropogenic sources such as wastewater discharges (Bezak et al., 2017; Neal et al., 2012). This study only investigates how precipitation uncertainty affects output uncertainty; however, precipitation may have tangential impacts to other influencers (e.g., groundwater chemistry, etc.) that are not part of this evaluation.

4.2.1 Precipitation Analysis

To begin, probability density functions (PDFs) of gauge data and SPPs are evaluated both over the 5-year study period and on a seasonal basis. This comparison is based on the interpolated and aggregated daily AWSA values of SPPs for all 87 watershed segments rather than a pixel-to-point comparison. PDFs reveal the inhomogeneity of the different products as well as the relationship between intensity and occurrence, as discussed in Section 4.3.1.

Second, the detection capability of daily AWSA SPPs in comparison to rain gauge observations is assessed through the following statistics: probability of detection (POD), false alarm rate (FAR), and critical success index (CSI). POD measures the ratio of correct detection of the SPP to observed occurrence of precipitation from the rain gauge. FAR measures the number of events when precipitation is detected by a SPP but no precipitation is observed by the reference gauge. CSI is a measure of successfully detected events to the total number of events observed (i.e., hits, false alarms, and missed event) (Schaefer, 1990). The total number observations by the SPP with respect to the gauge observation that either correctly detect an event, miss an event, or incorrectly detect an event, are hits, misses, and false alarms, respectively. The rain/no-rain threshold

is set to 0.254 mm/day, which corresponds to the minimum precipitation detectable by the rain gauges. Definitions used to calculate the POD, FAR, and CSI are provided as eqs. 7-9, respectively. POD, success ratio (1-FAR), CSI, and performance bias (ratio of the POD to the success ratio) are summarized by season using a performance diagram (Roebber, 2009).

Third, the skills of each SPP are quantified using four metrics: relative bias (rB), root mean square error (RMSE), correlation coefficient (CC), and standard deviation (σ) (eqs. 2-5, respectively). rB is a representation of the relative difference (in percentage) between estimated and observed data with positive and negative values indicating precipitation overestimation and underestimation, respectively. RMSE is a measure of the magnitude of errors between SPP and observed gauge values, whereas the CC provides a measure of the linear agreement between two variables. The amount of variation in the data set is quantified by the standard deviation. In this study, the overall skill associated with each of the three SPPs in relation to the gauge-based data is summarized on a seasonal basis using a Taylor diagram (Taylor, 2001).

4.2.2 Streamflow and Water Quality Indicator Analysis

The model performance in simulating Q and the three water quality indicators (i.e., TSS, TW, and DO) is evaluated through the absolute bias (B), the absolute value of B (eq. 6) and the relative RMSE (rRSME), provided as eq. 10. Absolute bias is the total measure of the systematic error of a dataset, whereas rRMSE is used to measure the random error between datasets. Error metrics are investigated as a function of spatial (i.e., basin size) and temporal scales (i.e., seasonality).

Equation 10. Relative Root-Mean-Square Error

$$rRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{s_i} - Q_{o_i}\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}(Q_{o_i})}$$

Where Q_{o_i} is the ith reference streamflow/water quality indicator measurement and Q_{s_i} is the ith simulated streamflow/water quality indicator value. $\bar{Q}$ is the corresponding mean value and n is the number of values (i.e., Q, TSS, TW, and DO).

To comprehensively evaluate and quantify the propagation of error from input precipitation to output streamflow and water quality indicators, two error metrics are adopted: bias propagation factor (Ebias) and rRMSE propagation factor (ErRMSE), defined as the ratio of error metric of output (i.e., Q, TSS, TW, or DO) to their respective error metric of input (i.e., P). The propagation factor is an indication of either dampening (less than 1) or amplification (greater than 1) of error as it is translated from precipitation to model output simulations.

Equation 11. Bias Propagation Factor

$$E_{bias} = \frac{B_{output}}{B_{input}}$$

Equation 12. Relative Root-Mean-Square Error Propagation Factor

$$E_{rRMSE} = \frac{rRMSE_{output}}{rRMSE_{input}}$$

4.3 Results and Discussion

This section presents and discusses the results of (1) the error and performance analysis of SPPs, (2) the seasonal error analysis of simulated model output, and (3) the propagation of error from model input to simulated output.

4.3.1 Seasonal Performance of Satellite-based Precipitation Products

The three SPPs (TMPA, CMORPH, and PERSIANN) are compared to gauge-based measurements by season for all 87 watershed segments based on the daily interpolated and aggregated AWSA values. AWSAs are compared rather than a pixel-to-point comparison to assess the error propagation of different precipitation forcing datasets to simulated streamflow and water quality indicators on a seasonal basis and by basin scale.

Over the 5-years, all three SPPs tend to under-detect the occurrence of low-intensity precipitation ($P \leq 1$ mm/day), over-detect moderate-intensity ($1 < P < 20$ mm/day) events, but are more agreeable with heavy-intensity ($P \leq 20$ mm/day) events (Figure 8). Overall, the PDF of CMORPH is closer to the one of rain gauge observations with respect to the other two products. This may be due to the fact that CMORPH (like TMPA) uses both PMW sensors, which tend to accurately detect heavy, convective precipitation events, as well as IR retrievals which better detect shallow, light precipitation events (Guo et al., 2016). The superiority of CMORPH over TMPA may be

related to the finer spatial resolutions and time-scales of CMORPH (0.07° and 30 minutes), which may be able to better detect isolated, short duration, and low-intensity precipitation events.

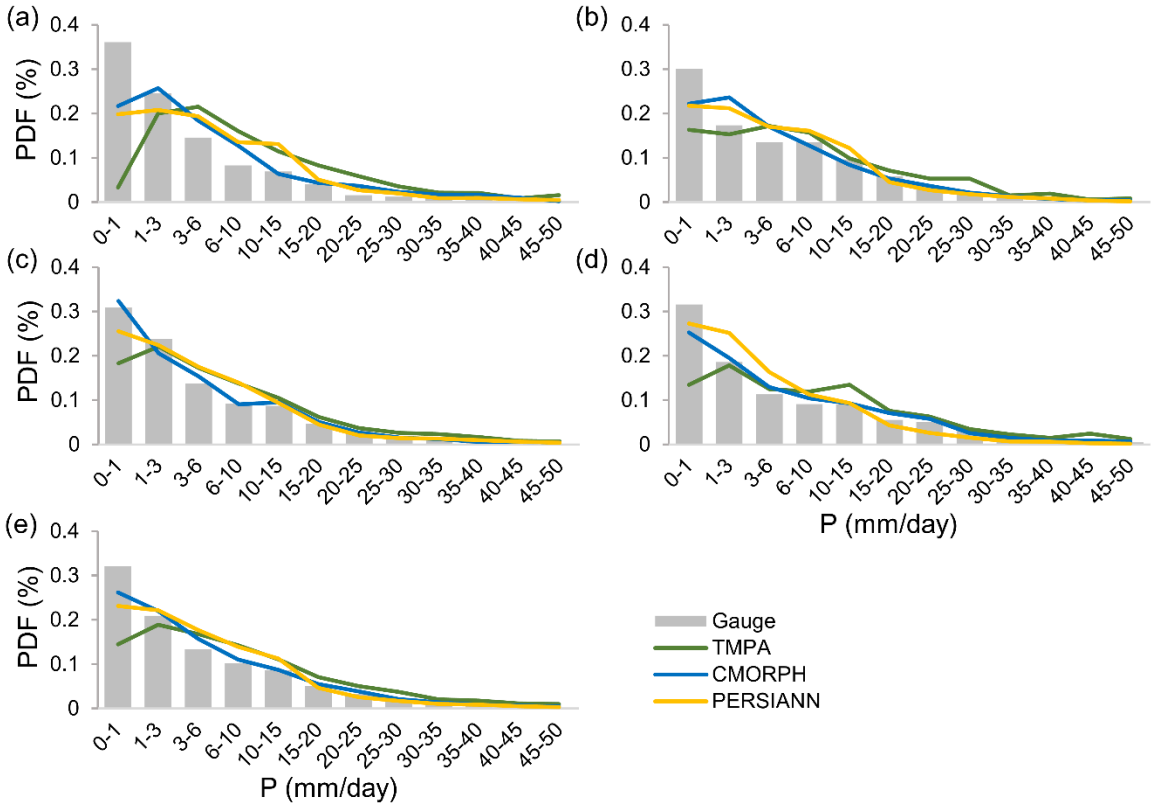

Figure 8. Probability density functions of daily AWSA precipitation from rain gauge observations and SPP estimates during (a) winter, (b) spring, (c) summer, (d) fall, and (e) 5-year period

Although SPPs tend to perform differently during different seasons, all three SPPs under-detect the occurrence of low-intensity P in comparison to rain gauge observations in all seasons. Moreover, in the spring PERSIANN tends to largely overestimate moderate-intensity P, whereas TMPA has a significant overestimation of high-intensity P

events. During summer, CMORPH outperforms TMPA and PERSIANN in capturing P of all intensities with very similar results to that of the gauges. In the fall, CMORPH has the best performance for all intensities whereas, as in summer, TMPA tends to underestimate low-intensity and over-estimate moderate- and high-intensity events. During winter, significant differences are observed with moderate-intensity P events during winter, with large overestimations of TMPA and PERSIANN. On the other hand, SPP estimates for high-intensity events are well matched with PERSIANN with a significant overestimation by TMPA. While these results suggest that TMPA tends to overestimate the magnitude of P during winter, it has a low POD and therefore misses a number of low-intensity events. This is likely related to the fact that PMW-based algorithms have difficulty estimating winter precipitation since they are influenced by snow and ice and are degraded by the presence of snow cover. Additionally, the overestimation of PERSIANN in winter may be associated with the imperfect screening of cold surfaces by IR sensors (Guo et al., 2016).

Over the 5-year study period, TMPA and CMORPH present lower FARs, but also lower PODs compared to PERSIANN (Figure 9a). On a seasonal basis, both TMPA and CMORPH exhibit comparable FARs, lowest in fall and winter (but when POD is lower) and highest in summer. On the other hand, PERSIANN demonstrates lowest FARs and highest POD in spring and winter. TMPA carries the lowest POD during all seasons, which may be attributed to the coarser spatial and temporal resolution of the product in comparison to CMORPH and PERSIANN. These results agree with Sun et al. (2018) indicating that TMPA has a lower POD and FAR than other products in both warm and

cold seasons over North America. The performance bias (the ratio of the POD to the success ratio) is relatively low in winter for both TMPA and CMORPH (Figure 9a). CMORPH and PERSIANN underestimate P during winter and fall, respectively. TMPA and CMORPH generally exhibit similar performances, whereas PERSIANN shows larger error metrics, particularly in winter, which is most likely due to the fact that PERSIANN only uses IR observations, affected by retrieval inaccuracy of stratiform precipitation and over snow during cooler months. Additionally, TMPA and CMORPH perform better during summer and fall since PWM/IR algorithms tend to better detect and estimate convective events common during warmer seasons, though they tend to overestimate P.

For the three SPPs, the error metrics shown in the Taylor diagram (Figure 9b) present a seasonal variation. Correlation coefficients vary by product and by season with CMORPH slightly outperforming TMPA and significantly outperforming PERSIANN. While overall correlations are concentrated between 0.26 and 0.60, both TMPA and CMORPH have the highest correlation during the fall, followed by winter. While inferior performance with low correlation values is observed with PERSIANN, it appears to perform best in winter and summer. The lowest RMSEs for TMPA and CMORPH are in winter, whereas the lowest RMSE for PERSIANN is in summer. TMPA and CMORPH have the lowest standard deviations during winter and the highest in fall, while PERSIANN has a lower standard deviation in fall and higher in winter. The notable differences found between SPPs analyzed in this study, along with their seasonal dependency, such as larger positive biases of TMPA and CMORPH during warmer

seasons and an overestimation of PERSIANN during winter, may lead to a seasonal impact on the hydrologic model output.

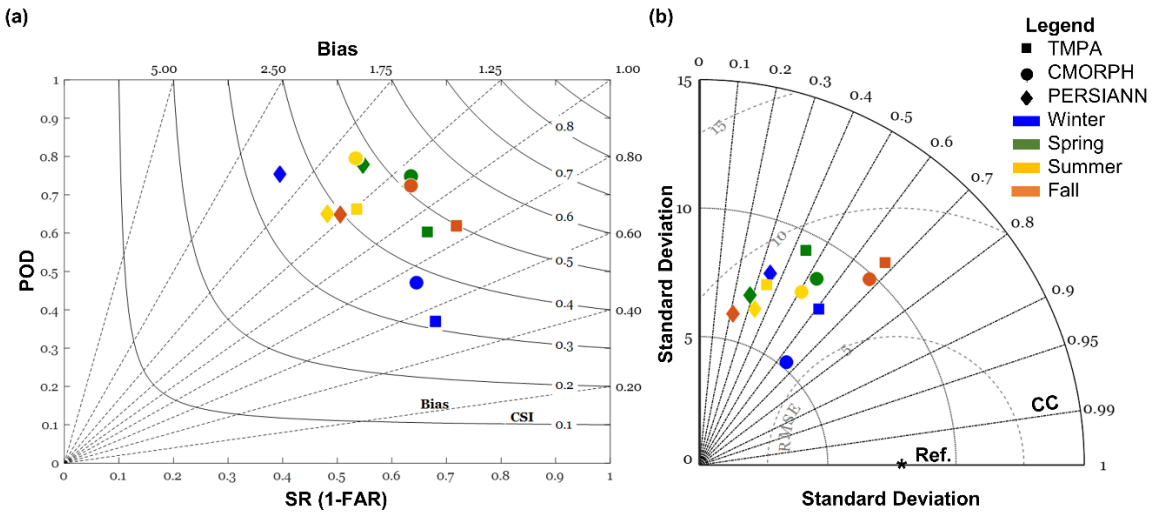

Figure 9. Seasonal (a) Performance diagram and (b) Taylor diagram of AWSA daily precipitation between gauge-based observations and SPPs

4.3.2 Error Analysis of Simulated Streamflow and Water Quality Indicators

Figure 10 shows density scatter plots (in the logarithmic scale) of daily SPPs against ground-based P observations and Q, TSS, TW, and DO simulated by the HSPF model forced with the SPPs against the corresponding output simulated by the model forced with gauge P observations. Overall statistics in terms of CC, B, and rRMSE are also shown on the scatterplots. As already concluded from the previous section, TMPA and CMORPH perform better than PERSIANN in terms of P with CMORPH showing an overall slightly better performance in reference to the gauge-based data. TMPA- and CMORPH-simulated Q have a moderate linear relationship to gauge-simulated Q values;

however, both SPPs tend to overpredict Q. PERSIANN underpredicts the reference Q and shows a large dispersion around the 1:1 line. Correlations of simulated Q are higher with respect to the P ones, although the B is also higher. Generally, the rRMSEs for the three SPPs are relatively close between Q and P. These results indicate that the HSPF model may have a dampening effect on the Q error for both TMPA and CMORPH; however, the poor quality of the PERSIANN product is actually amplified in the modeled Q.

Correlations of water quality indicators vary considerably with high values for TW (0.98-0.99) and DO (0.81-0.85), but lower values for TSS (0.26-0.56). TW has the strongest positive linear relationship of all water quality indicators with little dispersion. The B and rRMSE of both TW and DO are very low for all three SPPs. In general, all SPP-simulated values have a positive linear relationship with gauge-simulated values for DO, though all SPPs tend to overpredict lower DO concentrations and underpredict higher concentrations.

SPP-simulated TSS has the weakest relationship to gauge-simulated TSS for all water quality indicators. For TSS, TMPA and CMORPH tend to slightly underpredict TSS concentrations whereas PERSIANN significantly overpredict TSS concentrations. B and rRMSE are moderate for TMPA and CMORPH, but much higher for PERSIANN, which is expected since TSS is highly dependent on Q. These results corroborate the ones of Wu et al. (2006) that concluded that uncertainty in Q was the main source of variance in simulated TSS concentrations and nutrient loads.

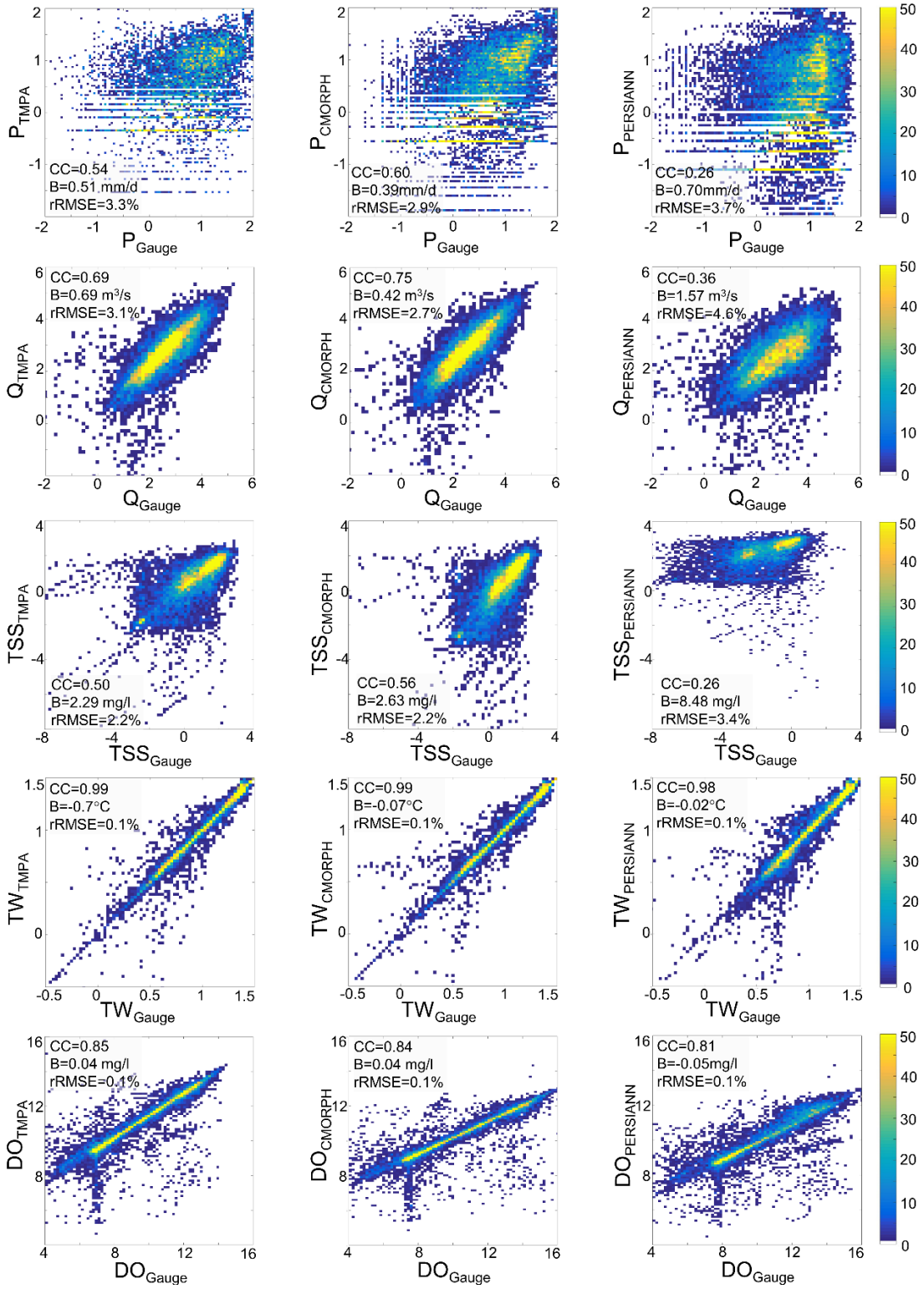

Figure 10. Logarithmic density scatter plots between the gauge-based reference and SPP-simulated model output for TMPA, CMORPH, and PERSIANN

To better understand the model's behavior between forced input and simulated output, error metrics are evaluated seasonally for P, Q, and the water quality indicators. Specifically, the simulations forced with the three different SPPs with respect to the one forced with gauge-based precipitation are evaluated in terms of both B and rRMSE (Figures 11 and 12). For Q, TMPA and CMORPH outperform PERSIANN, though the seasonal performance varies considerably by product. Generally, TMPA shows the lowest B during summer and the highest B in winter across all evaluation points. On the other hand, CMORPH-Q tends to have the lowest B during winter and highest during summer, with a few exceptions. B results between TMPA-Q and CMORPH-Q appear quite interesting considering that the B associated with their P counterpart does not vary much by season. Additionally, B tends to follow the same trends of P for CMORPH, with summer presenting the highest B for both TMPA and CMORPH and a similar skill amongst the other seasons. TMPA-Q and CMORPH-Q generally present similar rRMSE with the highest values during summer (when RMSEs of P are also high) and the lowest in spring for TMPA and winter for CMORPH. Low rRMSEs in winter may be due to less Q and thus lower residuals even though the B of P may be greater. Similar to TMPA, PERSIANN-Q is relatively inferior in the winter with similar Bs for other seasons. In general, a high FAR and B noted in the PERSIANN-P during the winter season leads to a notably larger B in PERSIANN-Q when compared to the TMPA and CMORPH products (i.e., almost ten times greater). It is also noteworthy to mention that all three SPPs overestimate daily Q in comparison to gauge-simulated Q, which may be attributed to the fact that Q increases with drainage area causing the potential for greater residuals. These

results coincide with other studies evaluating the performance of SPPs by basin scale (Maggioni et al., 2013; Falck et al., 2015; Nikolopoulos et al., 2010; Yu et al., 2016).

For TSS simulations in this study, TMPA and CMORPH generally outperform PERSIANN during all seasons, which is expected since similar results are noted from Q, and simulated TSS concentrations are highly dependent on Q. A large seasonal variation in B is observed for PERSIANN, but significantly lower for TMPA and CMORPH. Interestingly, in summer and fall both P and Q are overestimated, whereas TSS is underestimated by TMPA simulations. CMORPH often underestimates TSS during the winter and spring seasons. TSS is also generally underestimated across basin scales by PERSIANN, which tends to overestimate low intensity P that may contribute very little to Q and increased sediment transport. During winter, B of PERSIANN-TSS is approximately four times larger than for TMPA and CMORPH. Basin scale does not appear to impact B for TSS; however, rRMSE tends to decrease by basin scale for all products during all seasons. Generally, seasonal rRMSE for all three SPPs follows the same trends as the error associated with Q – higher in the summer and lower during spring. Poor HSPF model performance is noted in simulated TSS, which may be attributed to misprediction of intensity and the spatial scale of SPPs that may impact simulations of sediment loads generated from land use runoff, and to a lesser degree, stream scour.

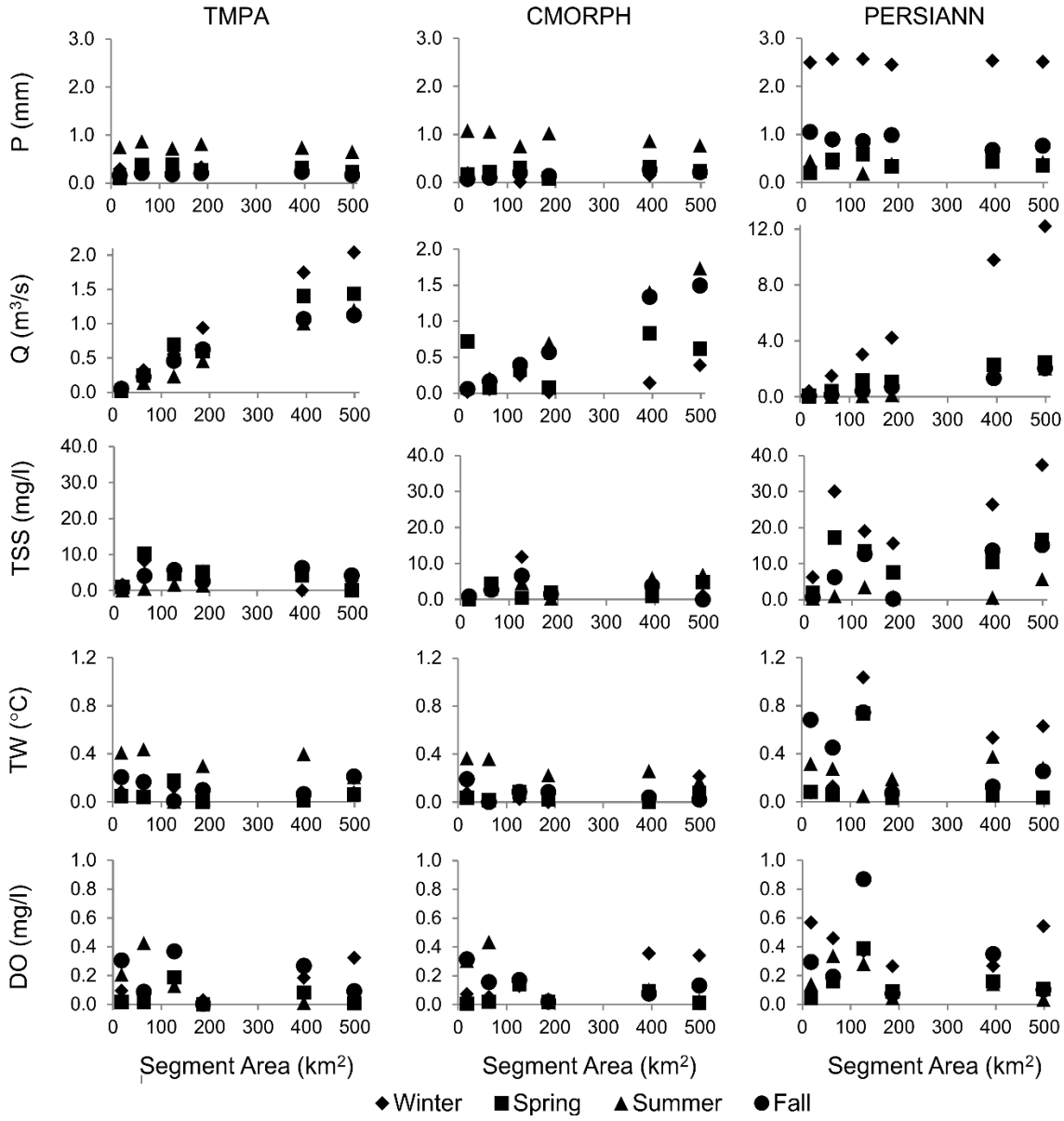

Figure 11. Seasonal analysis of absolute B as a function of basin scale for P, Q, TSS, TW, and DO between gauge-based records and SPPs

TMPA and CMORPH also generally outperform PERSIANN in terms of TW during all seasons. Highest B is found in summer, as expected, since there is a larger fluctuation in ambient air temperature during this season. B associated with TMPA and

CMORPH for all other seasons is below 0.22°C, but larger Bs are found during winter for PERSIANN. B for PERSIANN-TW presents a large variance among seasons, which is not as evident in the other two simulations. Winter presents significantly higher rRMSE for TW in comparison to other seasons.

During warmer seasons, high water temperatures generally increase the rate of biological activity and chemical reactions, which in turn decrease the solubility and concentration of saturated DO in a waterbody (Cox and Whitehead, 2009; Irby et al., 2018). Seasonal precipitation patterns also play a role in DO concentrations with higher moderate- and high-intensity events found in spring and fall: increasing streamflow allows for more aeration of the water, which also increases DO concentrations. Results of this study indicate a cyclical pattern with SPP-simulated DO concentrations higher in the cooler months and lower in the warmer months. Errors of simulated DO indicate more conformance among the three SPPs, with TMPA and CMORPH marginally outperforming PERSIANN and with all three SPPs underestimating DO, especially in the fall. The basin scale does not seem to play a factor and there does not appear to be a clear indication of how seasonality impacts B. The lowest rRMSEs are generally found in winter, followed closely by spring and fall, with the highest errors during summer. The seasonality trends of rRMSE are interesting since the DO concentrations are highest in winter, which would lend for greater residuals though this is not seen in the results of rRMSE indicating that other influencers built into the model may have an impact on random errors. A recent study by Moreno-Rodenas et al. (2019) evaluating the uncertainty and propagation of error between input variables and DO simulated output

found that precipitation uncertainty accounted for approximately 20% of the variance in DO. This same trend appears among all evaluation points and may coincide with the skill among SPPs presented for Q.

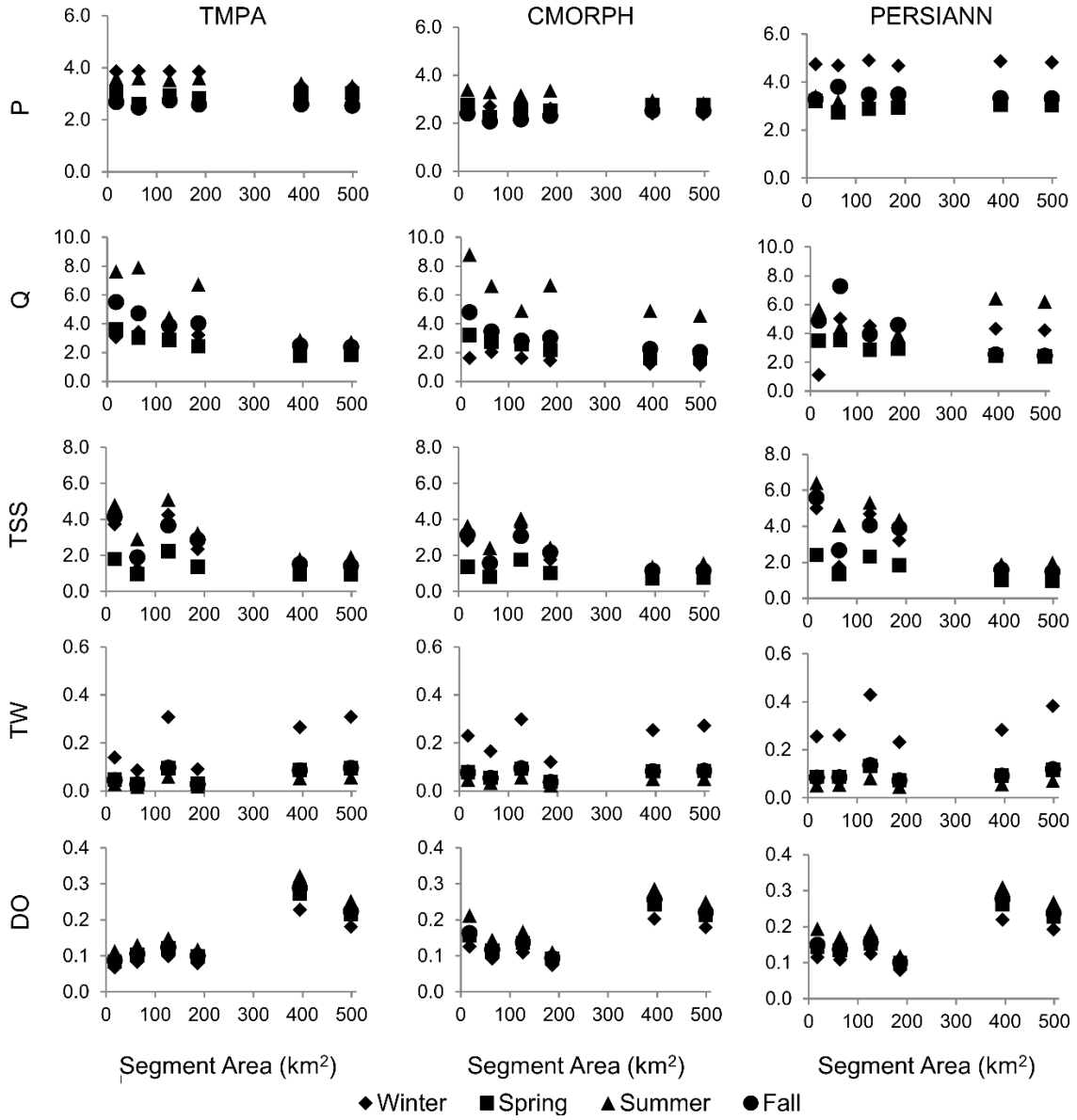

Figure 12. Seasonal analysis of rRMSE as a function of basin scale for P, Q, TSS, TW, and DO between gauge-based records and SPPs

4.3.2 Error Propagation

As shown in the previous section, the positive bias in SPPs is propagated into simulated Q, but not in all the water quality indicators. The E_{bias} and E_{rRMSE} of Q and water quality indicators are investigated on a seasonal basis and by basin scale (Figures 13 and 14, respectively). No single product outperforms the others in terms of E_{rRMSE}.

For Q, E_{bias} ranges between almost zero and 8.5, linearly increasing with basin scale. The largest error propagation is seen for CMORPH at evaluation point 86 during the winter season, which is caused by a very low absolute bias of P (0.01 mm/day). While the bias associated with Q and water quality indicators for CMORPH at evaluation point 86 is not particularly high, E_{bias} is magnified due to the low B in P. There do not appear to be any distinct seasonal trends associated with the E_{bias} for Q, but when investigating E_{rRMSE}, basin scale has an impact across all seasons, similar to results discussed above for rRMSE. E_{rRMSEs} close to 1 are found in winter for all three SPPs, indicating a higher dampening effect when compared to other seasons. These results are consistent with those found by others (Guetter & Georgakakos, 1996; Flack et al.; 2015; Hossain & Anagnostou, 2004; Maggioni et al., 2013; Mei et al., 2016b; Nijssen & Lettenmaier, 2004; Nikolopoulos et al., 2013; Seyyedi et al., 2015; Vivoni et al.; 2007; Yu et al., 2016), indicating that error associated with precipitation input is often amplified when translated into streamflow error.

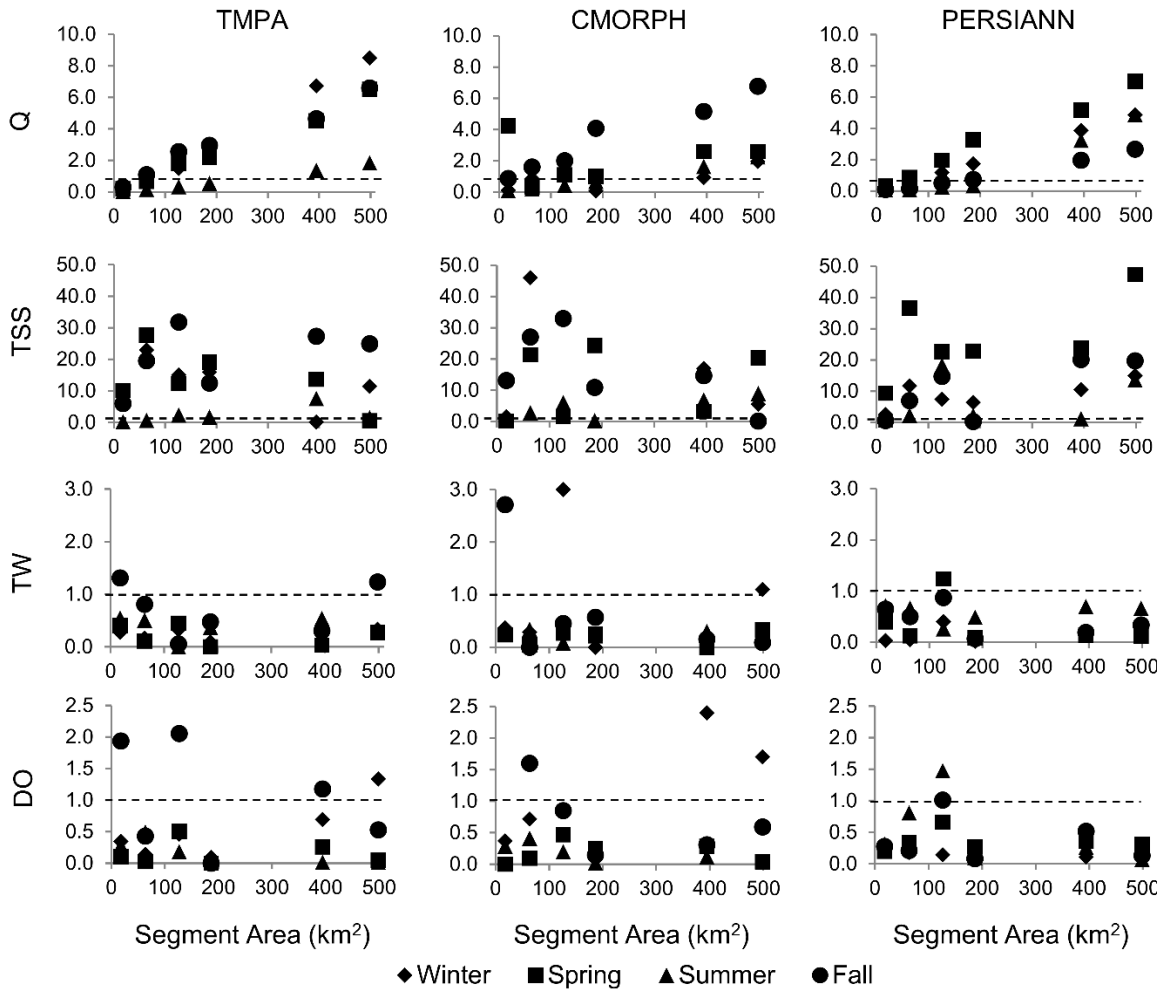

Figure 13. Seasonal analysis of B error propagation analysis as a function of basin scale for Q, TW, TSS, and DO between gauge-based records and SPPs

There is no notable scale dependency on E_{bias} for TSS and there is no clear indication of seasonality either, although a lower E_{bias} is generally seen during summer for the three SPPs. It is worthwhile to mention that while larger TSS errors are generally found in summer, this seasonal difference is not propagated for TSS and in some cases a dampening affect (0.23- 0.91) is noted during the summer. Aside from a few instances, E_{bias} is significantly greater than 1 and TSS error is amplified up to almost 50 times

greater than that found in the input P. E_{bias} is much greater for TSS than for Q or other water quality indicators, indicating that TSS simulations in the HSPF model are highly impacted by forcing P. E_{rRMSE} does show a dependence on the basin scale though results also vary considerably by season. E_{rRMSE} is smaller during spring/winter for TMPA, spring for CMORPH (similarly to rRMSE), and winter for PERSIANN. Conversely, higher E_{rRMSE} values are found during fall for TMPA and CMORPH and summer for PERSIANN. In general, the majority of E_{rRMSE} values (0.3 to 1.8) are below one indicating a dampening effect of the P random error when translated onto TSS, especially in larger basins.

Uncertainty propagation factors of TW is generally below 1 aside from a few instances. While larger B values of TW are found during summer in comparison to other seasons, there is no seasonal consistency for TW E_{bias}, reinforcing the conclusion that ambient air temperature is the greatest influencer on in-stream temperature. As expected, it does not appear that basin scale has any influence on TW E_{bias} or E_{rRMSE}. The E_{rRMSE} of TW appears to be rather consistent among all three SPPs, well below 1, with larger E_{rRMSEs} during the winter and lowest E_{rRMSEs} during the summer, which is in line with results of rRMSE for TW.

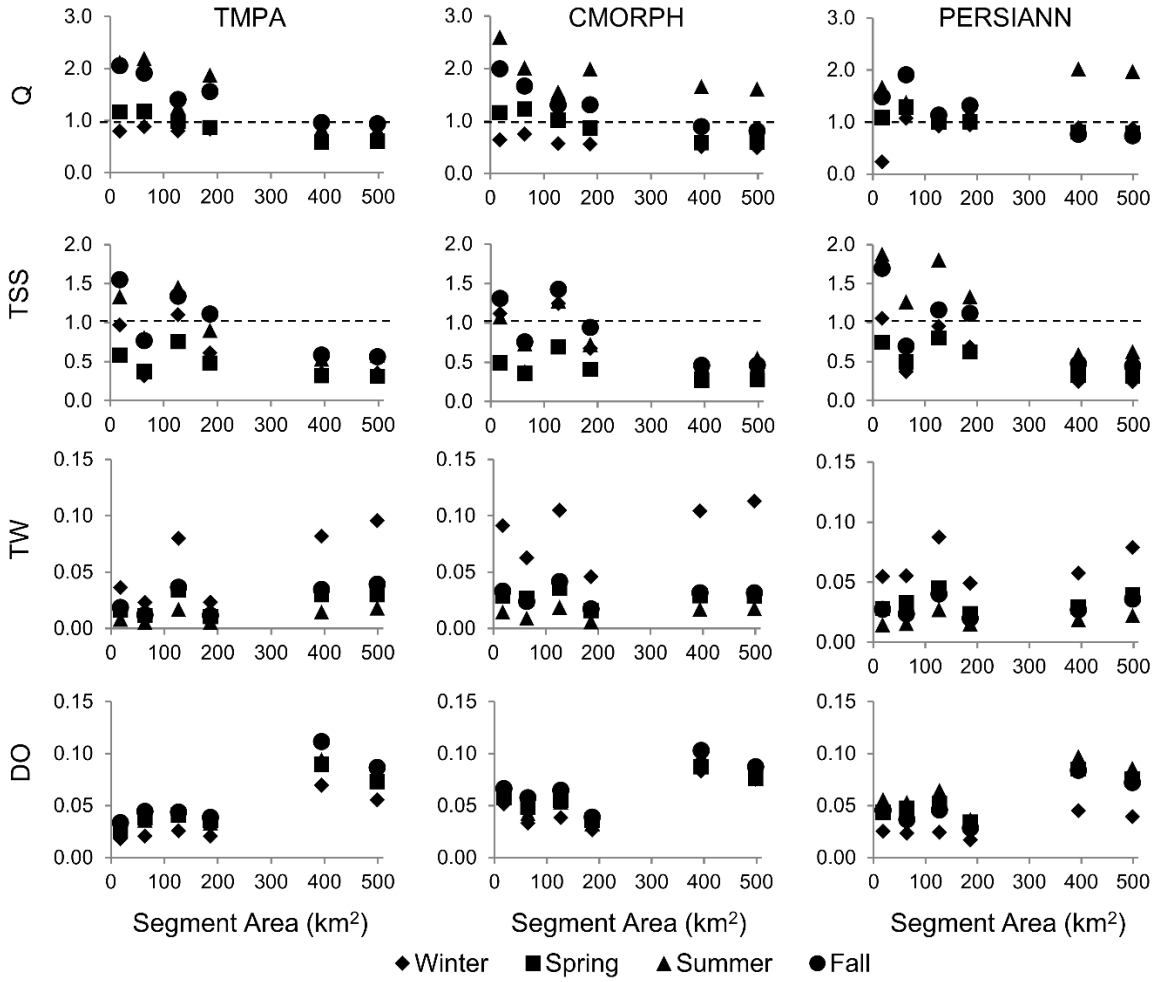

Figure 14. Seasonal analysis of rRMSE error propagation as a function of basin scale for Q, TW, TSS, and DO between gauge-based records and SPPs

Propagation of uncertainty for DO varies by season with a comparatively lower E_{bias} during spring and summer, though most values are below 1 (as low as 0.0), indicating a dampening effect of the error in P. Larger E_{bias} values are found during fall and winter for TMPA and CMORPH, while there is no consistent seasonal dependence in PERSIANN. Propagation of rRMSE for DO is consistent with results found for rRMSE, with basin scale having some marginal impact on the results. All E_{rRMSE} values are below

one (0.0 to 0.1), showing a dampening of the random error of P (when translated into DO) as well. Typically, colder waters found in cooler seasons (winter) tend to have higher concentrations of DO than in warmer periods (summer). One explanation for the increased E_{rRMSEs} presented in fall and winter for TMPA and CMORPH may be due to the seasonal fluxes of DO concentrations since overall DO concentrations are typically higher during those seasons creating greater residuals.

4.4 Conclusions

This work investigates the potential of using SPPs as forcing input into a hydrologic model for simulating and predicting streamflow and water quality. Three SPPs of different spatial and temporal resolutions (TMPA, CMORPH, and PERSIANN) are compared to gauge-based records in the Occoquan Watershed over a 5-year study period. The seasonal error of simulated model output is investigated along with the propagation of error from the SPP-forced input to simulated output. The major findings of this study are summarized as follows:

- SPPs show mixed performance skills with a seasonal dependence. Substantial differences are observed with moderate-intensity precipitation events, especially during winter, with both TMPA and PERSIANN grossly overestimating moderate-intensity precipitation. On the other hand, SPP estimates for high-intensity events are well predicted by PERSIANN but overestimated by TMPA. In the spring, all three SPPs tend to under-detect low-intensity precipitation. TMPA tends to under-detect low-intensity events and over detect moderate- and high-intensity events; however, it is much better able to detect low intensity events in warmer seasons.

- Correlations between SPP-simulated Q and reference Q (simulated by forcing the model with rain gauge observations) are higher with respect to the precipitation ones, although biases are also higher. Generally, random errors for the three SPPs are relatively close between Q and precipitation. These results indicate that the HSPF model may have a dampening effect on the Q error for both TMPA and CMORPH; however, due to the poor quality of the PERSIANN product, its error is actually amplified through the model, with a positive dependency on basin scale.
- For TSS, TMPA and CMORPH appear to generally outperform PERSIANN during all seasons, which is expected since similar results are noted from Q and simulated TSS concentrations are highly dependent on Q. Although the systematic error in P is amplified in TSS, the random error is not (and it is actually dampened by the model). It is worthy to mention that while larger TSS errors are generally found in summer for all three SPPs, this seasonal difference is not propagated for TSS.
- The model shows good performance when simulating TW, with TMPA and CMORPH generally outperforming PERSIANN. The error in precipitation is dampened when translated into TW and, as expected, basin scale has no influence on the precipitation-to-TW error propagation.
- Satisfactory model performance is also shown in simulated DO, with TMPA and CMORPH marginally outperforming PERSIANN. The propagation of systematic and random errors from precipitation to DO vary by season, with larger dampening effects in spring and summer for all three SPPs.

This study demonstrated that the spatiotemporal variability of SPPs, along with their different algorithms, have a quantifiable impact on water quality simulations. However, results shown here are limited by several factors. Firstly, this study was conducted using only one hydrologic model. While the HSPF model developed for the Occoquan Watershed is widely used and well calibrated and validated, other hydrologic models may propagate errors in input precipitation differently. Secondly, the model was calibrated based on rain gauge data, which may not reflect the actual distribution, extents, or magnitude of precipitation in the watershed. Thirdly, a single watershed, located in an area characterized by a temperate climate, mild topographic variation, and moderate precipitation intensity, was the focus of this study. Since it is well documented that different SPPs perform differently by climate, topography, and geographic regions, results of this study may not be translatable to other locations. Nonetheless, this work suggests that SPPs may be used to monitor and forecast water quality of a hydrologic network.

CHAPTER 5. PERFORMANCE OF SATELLITE-BASED PRECIPITATION PRODUCTS IN SIMULATING STREAM WATER QUALITY INDICATORS DURING HYDROMETEOROLOGICAL EXTREMES

5.1 Introduction

Understanding the spatiotemporal behavior of hydrometeorological events is of critical importance for water resource management including flood mitigation and response, ecosystem restoration, river and water supply reservoir recharge, and water quality impacts. Evaluating how hydrometeorological extremes have historically behaved, including variations in intensity, duration, and frequency is of upmost importance not only for current water resource management, but also to understand long-term climate impacts and provide accurate predictions of future behavior (Alexander et al., 2019; Maggioni & Massari, 2019; Mahbod et al., 2018; Tongal, 2019).

While there is no widely used definition for a hydrometeorological extreme, which is regionally specific, indices based on daily precipitation data are typically used, such as annual maxima or arbitrary thresholds (e.g., 95^{th}, 99^{th} and 99.9^{th} percentiles). Extreme events are often also classified by localized intensity-duration-frequency curves of representative return periods (e.g., 100-year event) or as a named storm event (e.g., hurricanes, tropical storms, etc.). More recent research, especially with long-term and climate change studies, have shifted to the use of standardized indices to allow for consistency between studies. These indices include the Expert Team on Climate Change

Detection and Indices (ETCCDI), the Standardized Precipitation Index (SPI), the Standardized Precipitation and Evapotranspiration Index (SPEI), and the Palmer Drought Severity Index (PDSI) which measure aspects of frequency (e.g., days above fixed thresholds), intensity (e.g., wettest day, average daily intensity), and duration (e.g., consecutive wet and dry days) based on daily precipitation measurements from in situ, satellite, and/or reanalysis datasets (Alexander et al., 2019; Qin et al., 2019).

Spatiotemporal variations of hydrometeorological extremes and the subsequent influence on land surface hydrology and streamflow have been extensively investigated using precipitation measurements from ground-based systems (i.e., rain gauges and radars). While the most accurate precipitation measurements are obtained from ground-based observations, they typically lack the spatial representativeness often needed in large-scale studies. Thus, the use of satellite-based precipitation products (SPPs) in hydrologic modeling is a good alternative due to their continuous geographic coverage with high spatial and temporal resolution. A number of past studies have evaluated the uncertainty of SPPs specific to extreme precipitation events both regionally and globally. Previous studies have shown precipitation measurement uncertainty of SPPs is associated with intensity, duration, and scale, with a decrease in uncertainty during higher rainfall rates, larger domains, and longer time integration (Maggioni & Massari, 2018).

Large scale studies by Bharti et al. (2016), Katiraie-Boroujerdy et al. (2017), Chen et al. (2020), Demirdjian et al. (2018), Li et al. (2013), Lockhoff et al. (2014), Meng et al. (2014), Mehran and AghaKouchak (2014), Nastos et al. (2013), and Pombo and de Oliveira (2015) all found that SPPs tend to underestimate extreme precipitation in

comparison to gauge-based observations. AghaKouchak et al. (2011) evaluated SPP precipitation rate retrieval during extreme events for three products across the central U.S. and concluded that the skill of all three products is reduced with higher intensity events. Habib et al. (2009) evaluated six extreme hydrometeorological events in Louisiana, U.S. and found that TMPA products tend to underestimate high intensity and overestimate low intensity observations. Derin et al. (2019) investigated the ability of six SPPs to estimate extreme precipitation at nine mountainous locations with dense gauge networks, globally. This study showed a constant underestimation of extreme precipitation values consistent with other studies (Derin et al., 2016, Kubota et al., 2009; Kwon et al., 2008, and Maggioni et al., 2017). Derin et al. (2019) attributed the underestimation of extreme precipitation to the warm rain process resulting in the occurrence of shallow, but high accumulation precipitation. Mehran and AghaKouchak (2014) investigated the capability of SPPs in detecting intense precipitation rates over different temporal resolutions (3-24 hours) and found that the detection and skill of the SPPs evaluated improve with increasing temporal resolution further suggesting that integrating finer (e.g., 3 hourly) temporal resolution data into hydrological models may lead to significantly biased results.

Well-developed, physically-based distributed hydrological models are vital tools for simulating hydrological processes, in particular for forecasting and monitoring flood hydrographs. These models can characterize hydrological processes in watersheds by using spatialized variables and parameters (Su et al., 2017). However, the accuracy of input precipitation data including its spatial and temporal distribution, intensity, and

duration significantly impact hydrologic models (Hazra et al., 2019; Sorooshian et al., 2011; Zeng et al., 2018). Maggioni and Massari (2018) suggest that a significant error shift may present in runoff prediction due to non-linearity of the hydrological processes. This error may rise since SPPs tend to better detect higher intensity and miss lower intensity observations which in turn may lead to a higher probability of underestimating or overestimating streamflow magnitude. This could be especially true since most hydrological models are calibrated on a continuous period of data rather than event-based calibration. Event-based modeling evaluates discrete rainfall-runoff events in isolation, as opposed to continuous modeling, which contains integrated responses by synthesizing hydrologic processes over a long period of hydroclimatic conditions. For instance, Xie et al. (2019a) showed that model performance decreased as precipitation intensity increased in event-based modeling, when compared to continuous modeling of streamflow, which they attributed to the fact that hydrologic models are often calibrated on continuous flow and one set of parameters that may not be appropriate for event-based modeling.

There has been a number of studies that evaluate the performance of SPPs and their streamflow response during hydrometeorological extreme events (Chintalapudi et al., 2014; Gourley et al., 2011; Huang et al., 2013; Shanshan Jiang et al., 2017; Shanhu Jiang et al., 2018; Mehran & AghaKouchak, 2014; Maggioni et al., 2013; Mei et al., 2016; Nikolopoulos et al. 2013, 2015; Seyyedi et al., 2015; Shah & Mishra, 2016; Su et al., 2017; Sun et al., 2016; Yang et al., 2017; Zhang et al., 2015; Q. Zhu et al., 2016; Y. Zhu et al., 2017; B. Zhu et al., 2019). Su et al. (2017) concluded that while four different SPPs generally captured the spatial distribution of precipitation over the Upper Yellow

River Basins in China, mixed results were found when simulating high peak discharges and flow events. Mei et al. (2016a) investigated the performance of eight SPPs in simulating 128 flood events in the Eastern Italian Alps and found that though timing of the precipitation event dispersion exhibited good agreement with the reference data, the resulting hydrograph had a dampening effect of both systematic and random error relative to the SPP hyetograph. Shanhu Jiang et al. (2018) evaluated six SPPs in capturing 13 extreme precipitation events and simulating resulting streamflow over the Xixian Basin in China. They concluded that gauge-adjusted SPPs perform better than their real time counterparts in simulating daily streamflow extremes, although all six SPPs exhibited a deviation of peak magnitude and timing inconsistency when compared to observed data.

Several studies have evaluated hydrometeorological extremes and resulting event-based streamflow and water quality, though the majority of studies utilize ground-based observations (Ahn & Kim, 2016; Jeznach et al., 2017; de Oliveira et al., 2019; Rue et al., 2017; Qiu et al., 2018; Xie et al., 2019b). Jeznach et al. (2017) investigated methods to quantify potential impacts of extreme precipitation on water quality and found that extreme events are a major driver for the export of terrigenous organic-bound nutrients directly linked to erosion and sediment transport during large events. Rue et al. (2017) investigated the relationship between water quality and streamflow during an extreme hydrometeorological event and found a consistent increase/decrease in solutes during flood/flood recession, yet noted a disproportionate decrease in concentrations due to a seasonal flushing of streams. Ma et al. (2019) assessed the performance of two SPPs in simulating streamflow and suspended sediment at the monthly timestep in the Lancang

River Basin in southwest China. They found both SPPs show good capability of estimating monthly sediment loads. Stern et al. (2016) found streamflow and sediment supply predictions using a hydrologic model of the Sacramento River Basin, California improved with better spatial representation of watershed precipitation. To date there are only a few studies that have evaluated the simulation and forecasting of water quality based on spatial and temporal differences of SPPs (Ma et al., 2019; Solakian et al., 2019). Moreover, while there has been much research assessing the impact of hydrogeological extremes captured by SPPs (at different resolutions) to simulated streamflow response, there is a notable gap in literature associated with simulating and forecasting water quality during extreme events using SPPs.

This study provides a comprehensive evaluation of three different SPPs, of varying native spatial and temporal resolutions, during eight extreme hydrometeorological events with respect to observations from a dense rain gauge network over the Occoquan Watershed, located in Northern Virginia, US. The three SPPs evaluated are then used as forcing input into a hydrologic and water quality model to simulate streamflow and multiple water quality indicators at six locations within the watershed. The skill of the SPP-based model simulations is then compared to gauge-based simulations for the eight extreme hydrogeological events occurring within a 5-year study period (2008-2012). Section 5.2 presents the methods used in this study, Section 5.3 presents, interprets, and discusses the results, and Section 5.4 offers conclusory remarks.

5.2 Methodology

5.2.1 Event Selection

Eight extreme hydrometeorological events are evaluated in this study. These events are chosen according to the 95th percentile of daily precipitation recorded by rain gauges over the 5-year study period. To determine 95th percentile events, an empirical cumulative density function (CDF) of maximum daily intensity (mm/d) derived from rain gauge observations is evaluated over the 5-year study period (Figure 15). Precipitation events are then grouped according to the maximum daily intensity 95th percentile. Event durations span from 24 to 120 hours. Seven of the eight events in this study occur in spring (March – May) and fall (September – November), the other event, Event 8, is characterized as a convective storm during a period of unseasonably warm weather in December.

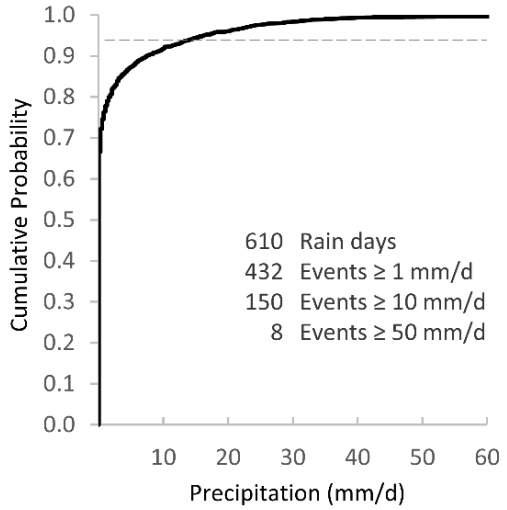

Figure 15. Cumulative density plot of daily precipitation events at representative rain gauge EVGR over the 5-year study period

Table 7 presents the characteristics of selected hydrometeorological events. The total average accumulation is the average amount of rainfall accumulated (mm) at all six gauges over the duration of the event. Maximum daily intensity (mm/d) is the greatest daily rainfall amount accumulated over one day during the event. Events 4, 7, and 8 have a maximum daily intensity (mm/d) greater than the average accumulation over the duration of the event. Events 7 and 8 have a duration of 24 hours, therefore the maximum daily intensity represents the event duration. For Event 4, a 96-hour event, the maximum daily intensity is higher than the average event accumulation due to the disproportional spatial distribution of rainfall over the rain gauge network, where one gauge measured a higher value during a one-day period, comparatively. Maximum hourly intensity (mm/hr) is the largest rain rate over a one-hour period during the event recorded at a gauge.

Table 7. Characteristics of selected hydrometeorological events

Event No.	Rainfall Duration (hr)	Start Date	Event Name	Average Event Accumulation (mm)	Maximum Daily Intensity (mm/d)	Maximum Hourly Intensity (mm/hr)
1	120	September 26, 2010	Event No. 1	122.47	97.28	15.24
2	120	March 6, 2011	Event No. 2	99.57	69.85	24.13
3	108	May 8, 2008	Event No. 3	159.55	90.93	49.78
4	96	May 21, 2012	Tropical Storm Alberto	117.14	145.80	70.36
5	84	September 5, 2011	Tropical Storm Lee	114.68	57.40	33.78
6	60	October 28, 2012	Hurricane Sandy	105.58	103.89	11.68
7	24	September 5, 2008	Hurricane Hanna	94.36	115.57	27.43
8	24	December 7, 2011	Winter Rain Storm	61.51	75.44	10.41

Since each of the six evaluation points have unique watershed conditions, there is no standard lag time or peak magnitude between locations, and therefore spatial

characteristic of the study area are also considered when evaluating performance. The eight hydrometeorological events are not only evaluated based on their precipitation intensity, but also based on their streamflow response. During three events, Events, 6-8, precipitation is continuous over the duration of the event hyetograph and the resulting event hydrograph are clearly visible. For multi-day intermittent precipitation, encountered with Events 1-5, streamflow hydrographs are examined to determine if flow returns to baseflow between precipitation periods since there is often a lag in watershed response between rainfall and runoff. To match the precipitation hyetograph and the resulting streamflow hydrograph, this study employs a linear regression baseflow separation method (Blume et al., 2007) to determine start and end times of events in the streamflow record. For these events, streamflow does not return to baseflow between precipitation periods, therefore the period is considered one event.

5.2.2 Statistical Metrics

First, to analyze the performance of SPPs in comparison to gauge precipitation observations during extreme events, each SPP pixel overlaying a representative rain gauge location (e.g., pixel-to-point) is compared. Precipitation performance is evaluated based on average event accumulation (mm), mean precipitation (mm), and maximum hourly intensity (mm/hr) for each of the eight extreme hydrometeorological events, as discussed in Section 3.1. The average accumulation is the amount of total rainfall accumulated (mm) over the duration of the event. This amount represents the average accumulation over the six rain gauge locations, with σ being the standard deviation of measurements from the six locations. Mean precipitation is defined as the total

accumulation divided by the duration in time of the event (mm/duration). Maximum hourly intensity (mm/hr) is the greatest rainfall experienced at a rain gauge during the event over a one-hour period.

Second, the HSPF model is forced with the three SPPs to simulate output of streamflow and water quality indicators using processed AWSA precipitation input. Model output are evaluated at six evaluation points by comparing the three SPP-forced simulations to that forced with rain gauge-based records for each of the eight hydrometeorological events. The mean relative error of peak streamflow (E_p) is evaluated between SPP- and gauge-simulated streamflow (eq. 13) for each of the eight events. Ep is defined as:

Equation 13. Mean Relative Error of Peak Streamflow

$$E_p = \frac{Q_s - Q_g}{Q_g}$$

Where Q_g is the peak gauge-simulated and Q_s is the peak SPP-simulated streamflow value (m^3/s) over the duration of the event.

The performance of simulated model output including Q and water quality indicators are then comparatively evaluated using the following verification metrics: correlation coefficient (CC), relative bias (rB), and relative root mean-square error (rRMSE) represented by eqs. 4, 2, and 10, respectively. CC is a measure of the linear argument between gauge-simulated and SPP-simulated output over the reference period with a perfect value of 1. rB is defined as the difference between the gauge-simulated and

SPP-simulated output, normalized by the gauge-simulated value (in %). Positive (negative) values indicate SPP-simulated output overestimation (underestimation), with a perfect value of 0%. rRMSE is a measure of random error, quantifying the error between the SPP-simulated output and the gauge-simulated output with a perfect value of 0%.

5.3 Results and Discussion

5.3.1 Extreme Hydrometeorological Events

TMPA, CMORPH, and PERSIANN products are evaluated against gauge-based records for the eight extreme hydrometeorological events described in Table 7. These eight events range in duration from 24 to 120 hours. Figure 16 shows the average event accumulation (mm), mean precipitation (mm/duration), and maximum hourly intensity (mm/hr) measured from rain gauges, TMPA, CMORPH, and PERSIANN for the eight events. In general, longer events (i.e., 1, 2, and 3) present a higher accumulation than shorter events (Events 7 and 8). The variation around the mean accumulation (σ) at the six locations varies not only by product, but also by event. Gauge-based measurements have the greatest σ amongst all of the products, which is attributed to the nature (point) of the measurement. Interestingly, both TMPA and CMORPH tend to overestimate event-based rainfall accumulation for longer events (Events 1 and 3) and underestimate accumulation during shorter-duration events, whereas PERSIANN consistently underestimates total event accumulation. This may be attributed to the fact that PERSIANN is based on a thermal infrared (IR) algorithm which has a tendency of missing light stratiform precipitation (Hong et al., 2007; Maggioni & Massari, 2018). Ebert et al. (2007) reported that in temperate climates IR-based SPPs generally are better

able to detect heavy precipitation during warm season convective storms and decline in accuracy with stratiform precipitation. On the other hand, low orbiting passive microwave (PMW) satellite-based algorithms, such as TMPA and CMORPH, are known to underestimate heavy precipitation events associated with convective storms.

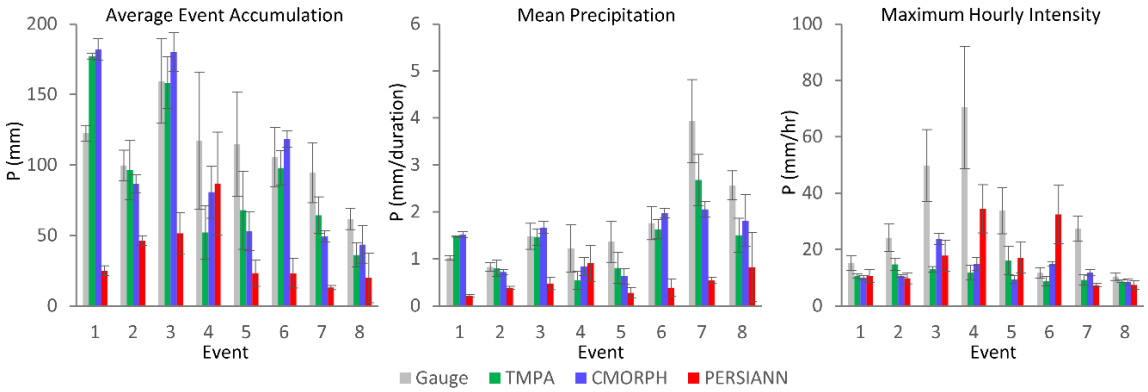

Figure 16. Average event accumulation, mean precipitation, and maximum hourly intensity for rain gauges, TMPA, CMORPH, and PERSIANN for eight extreme hydrometeorological events

Mean precipitation is inversely proportional to the event duration (Figure 16, middle panel). This is due to the fact that longer duration events (Events 1-3) are characterized by a mix of intermittent heavy and stratiform precipitation, whereas shorter events are typically of higher intensity. The same trend is noted for event accumulation, with PERSIANN grossly underestimating values (which is expected since mean precipitation is based on event accumulation). Maximum hourly intensity illustrates the dissimilarity of each product in the ability to measure heavy precipitation.

Aside from one event (Event 6), gauge measurements far exceed all SPPs. However, PERSIANN, with a few exceptions, tends to better detect heavy precipitation

in comparison to TMPA and CMORPH. Again, these results can be attributed to the ability of IR-based algorithms to better detect the temporal variability and intensity of heavy precipitation compared to PMW-based algorithms. The standard deviations of maximum hourly intensity (gray error bars), especially for Events 3 and 4, are significantly higher than other products indicating the variability of rainfall detected at different gauge locations.

5.3.2 Simulated Streamflow Peak Magnitude and Timing

The capability of SPPs to simulate streamflow during extreme hydrometeorological events is evaluated by investigating hyetographs and hydrographs for each event (Figure 17). Overall, in terms of observed precipitation and gauge-simulated Q, TMPA and CMORPH simulations tend to well capture the magnitude and timing of the peak events, however, there are a few exceptions. For Event 1, both TMPA and CMORPH grossly overestimate peak Q but are able to appropriately predict the timing of the peak. PERSIANN completely misses the Q peak in magnitude and timing resulting from its inability to accurately capture the accumulation of rainfall from intermittent stratiform precipitation between heavy intensity measurements during Event 1. For Event 2, TMPA and CMORPH capture both the timing and peak magnitude of Q, though gauge-based Q presented a second but smaller peak that is largely undetected by both TMPA and CMORPH. Based on precipitation estimates during Event 2, TMPA and CMORPH captured rainfall intensity at the beginning of the event, however, did not well capture rainfall during the second precipitation occurrence, thus leading to an underestimation of the second peak. Similar results are found with Events 3, 5, and 6

where TMPA and CMORPH are able to capture the peak magnitude of Q, though timing is delayed for Events 3 and 5.

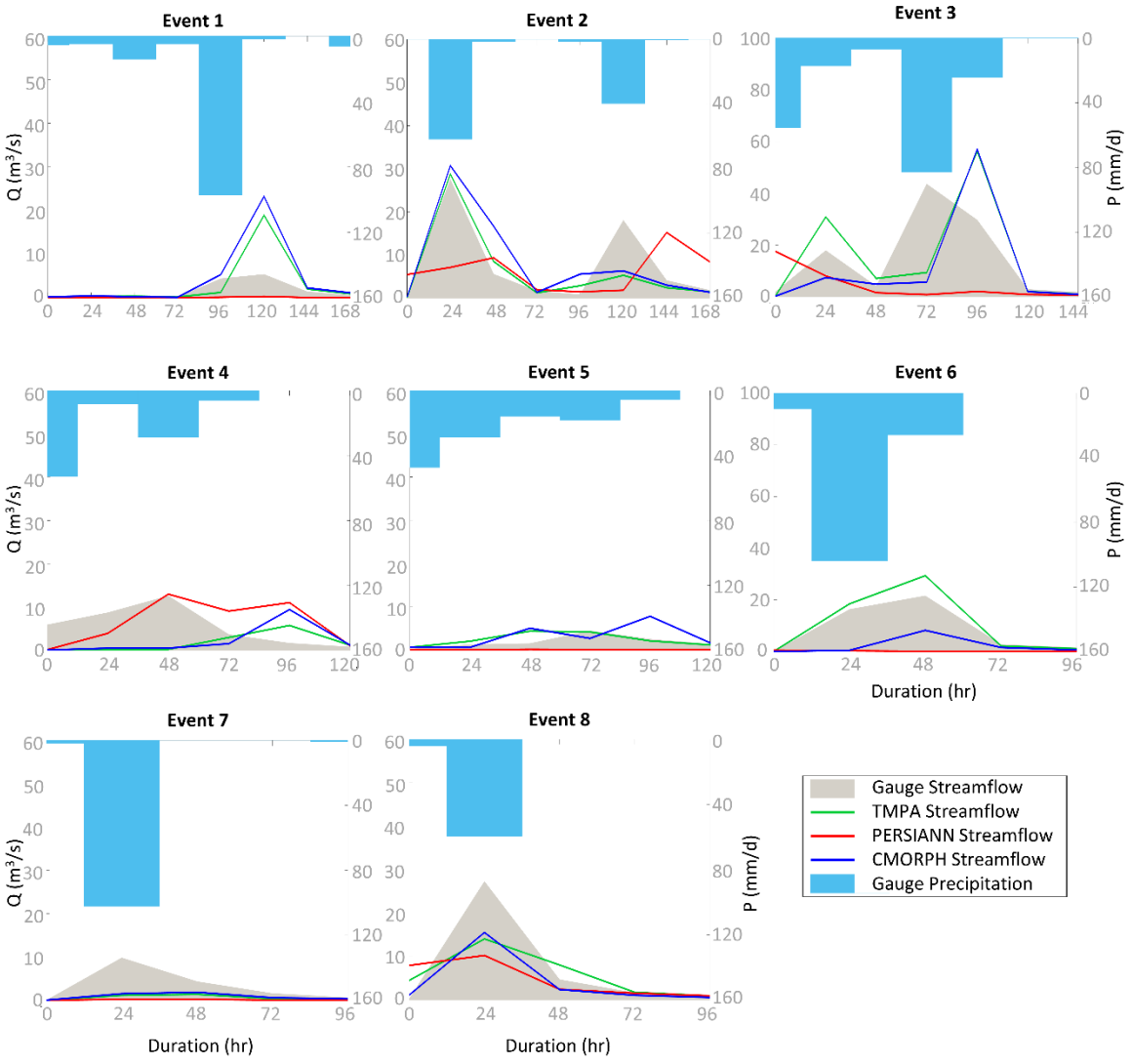

Figure 17. Hyetograph of the eight hydrological extreme events based on observations of daily precipitation intensity and corresponding hydrographs of simulated streamflow

PERSIANN exhibited the best performance for Event 4, able to capture the peak magnitude and timing of Q. During Event 4, PERSIANN captures the intermittent intense

precipitation over the duration of the event, whereas TMPA, and to a lesser degree CMORPH, cannot match hourly intensities observed from gauges early on in the event, and overestimate precipitation in later stages of the event's duration. All three SPPs are unable to capture precipitation intensity during Event 7 and thus significantly underpredict peak Q. Though Event 7 and Event 8 are classified as 24-hour events, and the average event accumulation (mm) and daily intensity (mm/d) are higher for Event 7, the peak Q simulated for Event 8 is greater. This is mostly likely due to the seasonal performance built into the HSPF model. Event 8 occurs in December whereas Event 7 occurs in September. HSPF inherently generates greater runoff from the watershed from reduced infiltration rates during cooler months (December – March), thus produces higher simulated Q than in warmer seasons.

To quantify the error associated with the peak magnitude of Q for each event, E_p is determined and averaged among all six evaluation points. Figure 18 displays the mean error of peak SPP-simulated Q in comparison to gauge-simulated Q. The most significant error in peak is associated with Event 1 where both TMPA and CMORPH significantly overestimate the peak magnitude at several evaluation points. The overestimation of the peak magnitude Q results from TMPA and CMORPH both estimating a higher precipitation accumulation during Event 1 than observed at the rain gauges. Except for Event 4, PERSIANN tends to underestimate peak magnitude of Q for all events. Aside from Event 6, results from TMPA and CMORPH are fairly consistent which is expected since they are based on similar input satellite retrievals to estimate precipitation. The inconsistency in peak magnitude of Q between TMPA and CMORPH for Event 6 is

attributed to the temporal resolution of the two products, 3-hour and 0.5 hour, respectively. In the three-hour window of TMPA, the duration of precipitation intensity is overestimated for TMPA when compared to CMORPH which then resulted in a higher simulated peak.

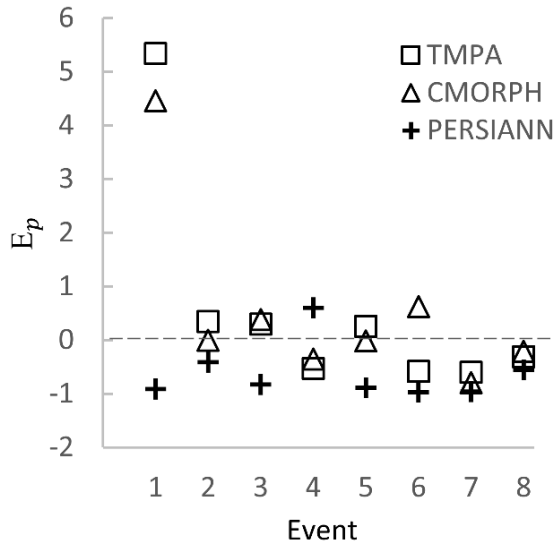

Figure 18. Mean peak error between gauge-simulated and SPP-simulated streamflow

5.3.3 Error Analysis of Simulated Streamflow and Water Quality Indicators

To assess model performance in simulating Q and water quality indicators between gauge-simulated output and SPP-simulated output, three error metrics are used: CC, rB, and rRMSE (Figures 19-21, respectively). Error is measured at the daily timestep and evaluated at six evaluation points over the duration of the defined event for Q, TW, and concentrations of TSS, DO, BOD, and nutrients OP, TP, NH_4-N, NO_3-N. Results indicate fairly good agreement between gauge- and SPP-simulated Q for TMPA and

CMORPH (CCs between 0.8-1.0) aside from Event 4, which is missed by all SPPs. The CC of PERSIANN-simulated Q is generally lowest among SPPs, due to its inability to accurately measure stratiform precipitation between intense periods of precipitation during an event. rB varies significantly for Q indicating SPPs either under- (negative) or over- (positive) predict Q. Highest rB errors for TMPA and CMORPH are 225% and 150%, respectively, and -100% for PERSIANN which all occur with Event 1. Events 2 and 5 have good agreement between gauge- and SPP-simulated Q. rRMSEs are somewhat consistent between SPPs for each event, however, high rRMSEs for TMPA and CMORPH are noted for Event 1, whereas the rRMSE of PERSIANN is fairly consistent between events. The high rRMSE values associated with Event 1 for TMPA and CMORPH are likely due the SPP's ability, as a gridded product with larger spatial resolution, to pick up localized precipitation, where some rain gauges did not record the same precipitation intensity during the event.

Correlations of water quality indicators vary considerably, however, TW has the strongest positive linear relationship compared to other indicators evaluated in this study. Strong correlations are attributed to in-stream temperature being naturally dependent on ambient air temperature conditions with precipitation, and thus Q, having a lesser impact of TW. Correlations fall between 0.8 to 1.0 for TMPA and CMORPH for all events except for Event 4. PERSIANN-simulated TW shows weaker correlations, especially for Events 5, 6, and 7. PERSIANN shows poor skill with simulating Q during these three events which indirectly impacts TW. rB and rRMSE for TW are also lowest (better

agreement) among all simulated output indicating that TW is more impacted by ambient air temperature than precipitation.

SPP-simulated TSS, a flow-dependent variable, has the weakest relationship to gauge-simulated TSS for all water quality indicators, with CMORPH performing slightly better than TMPA and PERSIANN. PERSIANN presents negative correlations during Events 1, 4, and 6 and generally performs worse than both TMPA and CMORPH, aside from two events (Events 5 and 7). Event 5 is characterized by a multi-day duration and lower-intensity precipitation than other events evaluated in this study. While PERSIANN misses the magnitude of peak TSS, CMORPH and TMPA both overpredict it. Event 8 presents a high correlation, and low rB and rRMSE for SPP-simulated TSS indicating that all three SPPs well capture TSS concentrations, likely due to the short duration of the event (24 hours) and the high skill of SPP-simulated Q, which directly influences TSS, especially during early stages of a storm event. In a study across the same watershed, Solakian et al. (2019) found a direct relationship between simulated Q and TSS concentrations, with peaks well represented in both TMPA and CMORPH simulations; however, TSS concentrations tend to be low throughout warmer months aside from a few instances where intense precipitation is captured and translated into peak Q.

Correlations of DO between SPP- and gauge-simulations during a 5-year continuous period are 0.81-0.85 (Solakian et al., 2019) indicating good agreement on a continuous basis. Correlations of event-based simulations suggest that model skill varies by event with majority of correlations above 0.8 for both TMPA and CMORPH, with a few exceptions. PERSIANN exhibits overall inferior performance for simulated DO

concentrations aside from Event 4. Both TMPA and CMORPH missed the peak Q for Event 4 whereas PERSIANN was able to well capture the timing and peak magnitude of Q, thus translating into a higher correlation of DO for Event 4 when compared to the other SPPs. Both rB and rRMSE for DO are much lower compared to other water quality indicators, with the exception of TW. This relatively small change in DO concentrations is not associated with precipitation, but rather DO is more temperature-dependent which is evident from seasonal fluctuations of DO concentrations.

Correlations of BOD appear to mimic TSS, though rB and rRMSE of SPP-simulated BOD vary considerably. CMORPH outperforms TMPA, and both CMORPH and TMPA significantly outperform PERSIANN aside from Event 5, where PERSIANN outperforms other SPPs. From a seasonal perspective, simulated BOD concentrations increase in warmer periods and drop during cooler periods following patterns associated with TW, yet this indicator also appears to be influenced by TSS concentrations during peak events.

Nitrogen-based nutrients, NH_4-N, and NO_3-N, are investigated based on a comparison of SPP-simulated to gauge-simulated concentrations. SPP-simulated NH_4-N and NO_3-N present similar correlations to TSS, as well as consistent rB and rRMSE values, aside for rB values for NH_4-N for Event 1 which present an inverse relationship. SPP-simulated NO_3-N correlations, rBs, and rRMSEs present similar results among the three SPPs indicating that the spatial and temporal differences of SPPs have little impact on simulation skill.

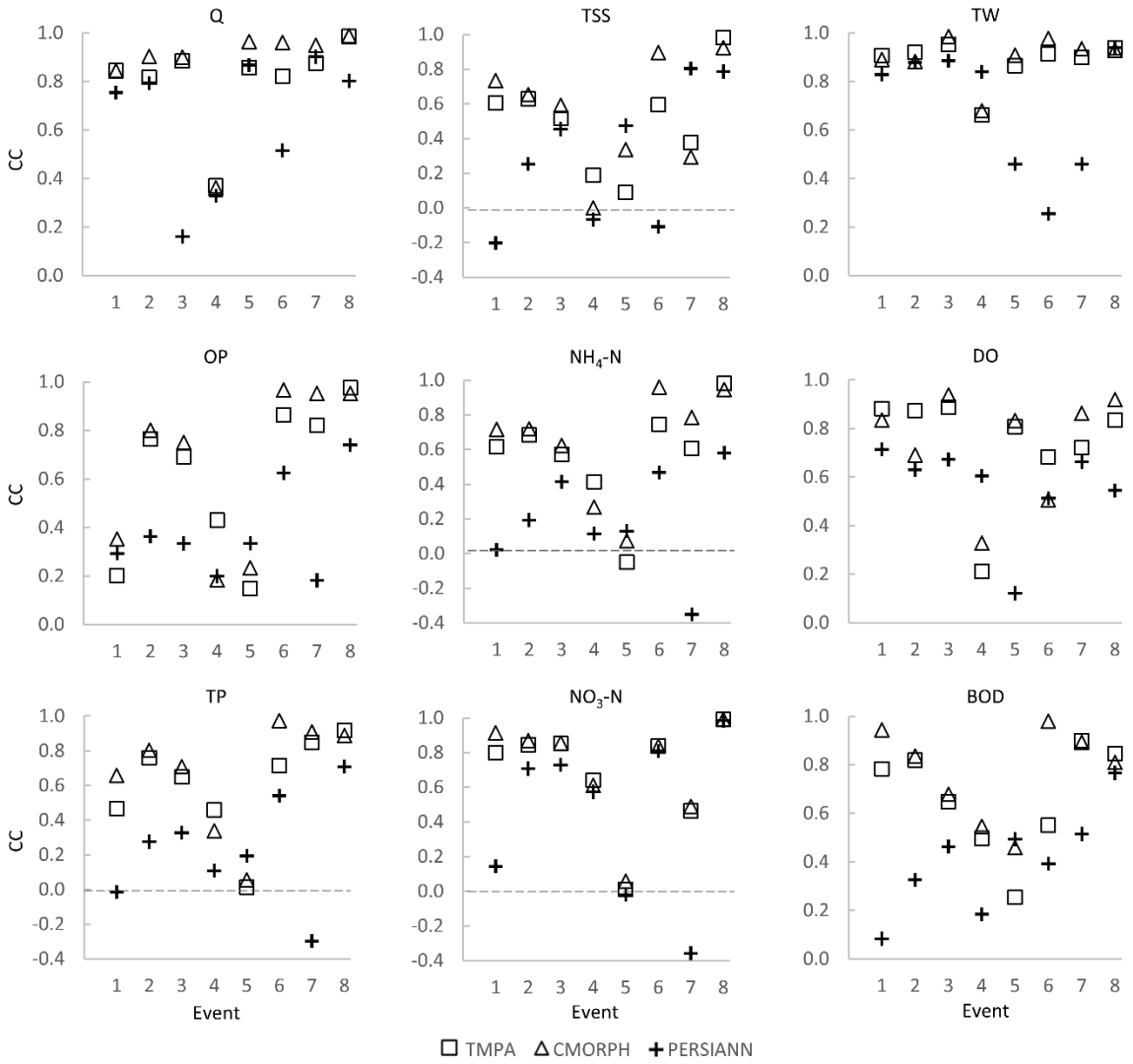

Figure 19. Correlation coefficients of streamflow and water quality indicators between gauge-simulated and SPP-simulated output

Two phosphorus-based compartments are investigated in this study: TP and OP. The Occoquan Watershed model, built with precipitation input at 87 individual segments, is able to well capture changes of simulated nutrient loads due to changes in precipitation input from the four data sources: rain gauges, TMPA, CMORPH, and PERSIANN. The behavior of TP and OP concentrations follow similar agreement between SPP- and

gauge-simulations as with TSS, and have similar correlations. rB of TP and OP are inversely proportional to TSS which may be due to the influence of both insoluble (sediment-dependent) and soluble (precipitation-dependent) forms of phosphorous included in the model.

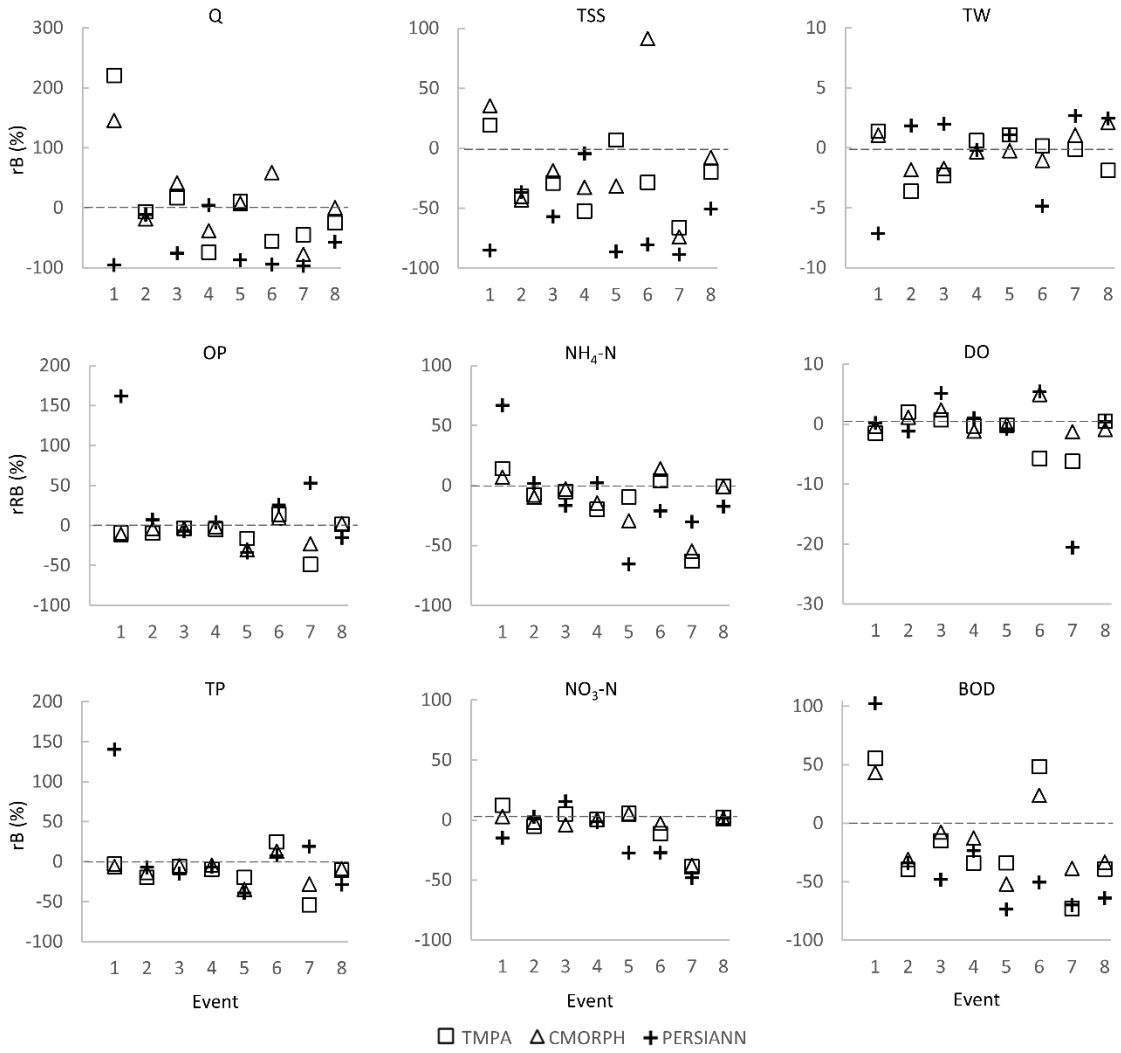

Figure 20. rBs of streamflow and water quality indicators between gauge-simulated and SPP-simulated output

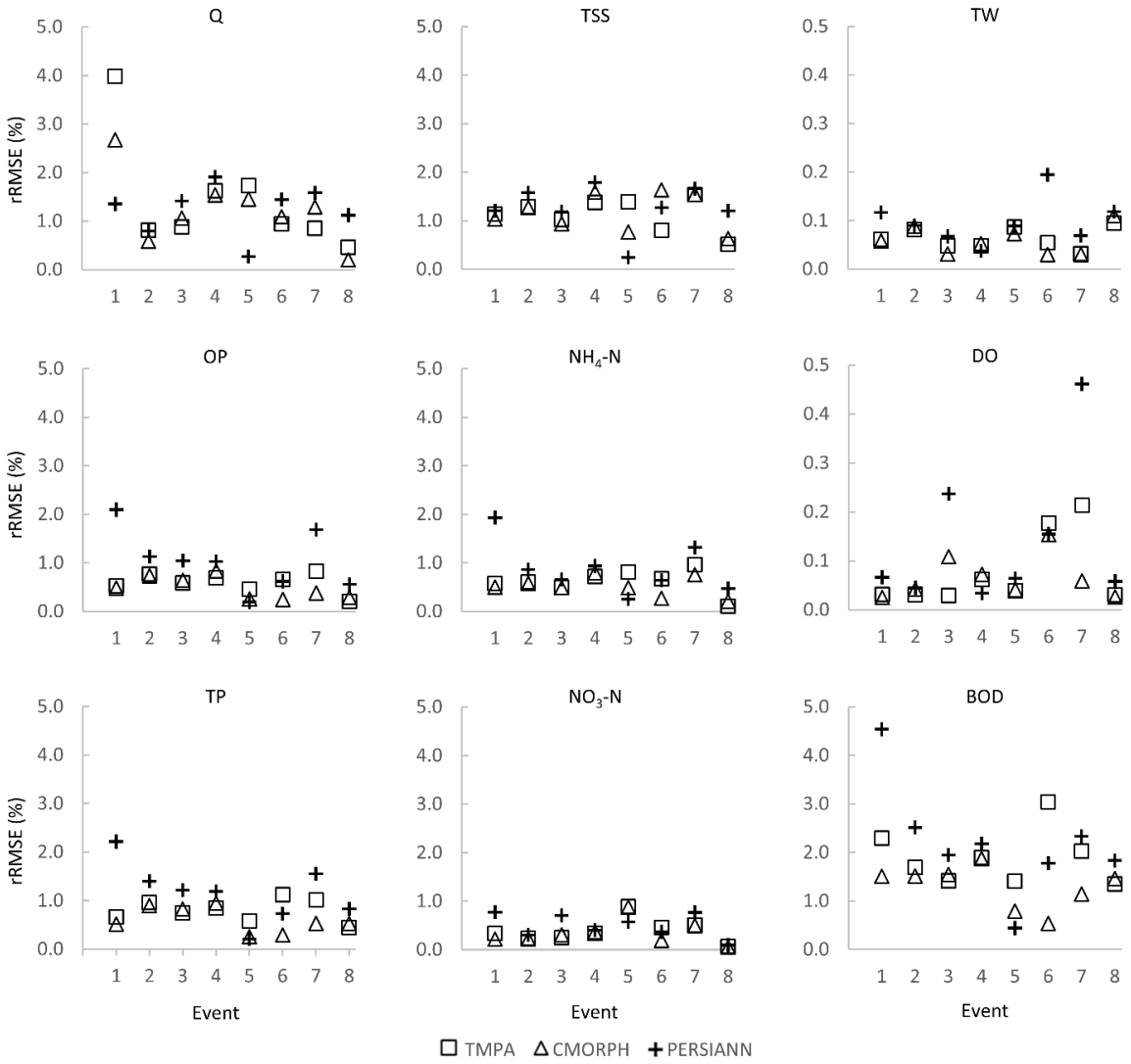

Figure 21. rRMSEs of streamflow and water quality indicators between gauge-simulated and SPP-simulated output

Overall, all three SPP-simulated TP and OP outputs are fairly consistent with no clear indication of one outperforming another. The one exception to consistence in performance is PERSIANN-simulated TP and OP (rB = +140% and +160%, respectively) for Event 1. These results suggest that, while TP and OP may be sediment-dependent indicators most likely tied to land use conditions in the watershed and the

spatial and temporal differences in precipitation have a low impact on simulated TP and OP concentrations.

5.4 Conclusions

This work investigates the performance of three SPPs of different spatial and temporal resolutions (TMPA, CMORPH, and PERSIANN) during extreme hydrometeorological events. First, the three SPPs are evaluated on their ability to estimate rainfall intensity and its temporal and spatial characteristics during eight extreme events. Next, a hydrological model, calibrated using a dense rain gauge network, is forced with the SPPs to simulate streamflow and water quality in the Occoquan Watershed during 5 years. Eight extreme events within the study period are identified according to the 95^{th} percentile of daily precipitation recorded by rain gauges during the study period. Event durations span from 24 to 120 hours. Event-based error is measured at the daily timestep and evaluated at six evaluation points for Q, TW, and concentrations of TSS, DO, BOD, and nutrients OP, TP, NH_4-N, NO_3-N.

Results indicate fairly good agreement between gauge- and SPP-simulated Q for TMPA, whereas PERSIANN-simulated Q is generally lowest among SPPs, due to its inability to accurately measure stratiform precipitation between intense periods of precipitation during an event. The skill of SPP- to gauge-simulated water quality indicators vary considerably. However, TW has the strongest agreement compared to other indicators evaluated in this study. SPP-simulated TSS, a flow-dependent variable, has the weakest relationship to gauge-simulated TSS among all water quality indicators, with CMORPH performing slightly better than TMPA and PERSIANN. Strong

agreement of TW simulations is attributed to in-stream temperature being naturally dependent on ambient air temperature conditions with precipitation, and thus Q, having a lesser impact. For both DO and BOD, event-based simulations suggest that model skill varies by event though SPP-simulated output shows overall good agreement to gauge-simulated output, with a few exceptions. PERSIANN exhibits overall inferior performance for simulated DO and BOD concentrations. For phosphorus-based nutrient simulations of TP and OP, all three SPP-simulated TP and OP outputs are fairly consistent with no clear indication of one outperforming another suggesting that while TP and OP may be sediment-dependent indicators most likely tied to land use conditions in the watershed and the spatial and temporal differences in precipitation have a low impact on simulated TP and OP concentrations. Nitrogen-based nutrients, NH_4-N, and NO_3-N, present similar results among the three SPPs indicating that the spatial and temporal differences of SPPs have less of an impact on simulation skill that other water quality indicators.

Overall, this study demonstrates that the spatiotemporal variability of SPPs, along with their different algorithms, are capable of predicting the characteristics of streamflow and water quality simulations during hydrometeorological extreme events. However, there are limitations to this study. Foremost, the study area, suburban Washington, D.C., is situated in a region characterized by a temperate climate and mild topographic variation with moderate precipitation intensity. Both climate and topography may have a significant impact on SPP performance. Secondly, this analysis was conducted in a single location utilizing only one hydrology/water quality model, i.e., HSPF. While this model

is well calibrated and has been validated to observation results, another model may respond in a different way to changes in streamflow and water quality indicators resulting from forcing precipitation inputs. Thirdly, the model was calibrated based on rain gauge data, which may not reflect the actual distribution, extents, or magnitude of precipitation in the watershed. Lastly, while this study provides a comprehensive evaluation of three of the most widely used SPPs, TMPA, CMORPH, and PERSIANN, it does not consider an exhaustive list of precipitation sources such as other SPPs, blended and reanalysis products, or radar data. Nonetheless, this work represents a novel approach to utilizing SPPs data for water quality modeling during extreme hydrometeorological events, which could be of critical importance in areas of the world where rain-gauge networks or monitoring stations are either sparse or not available altogether.

CHAPTER 6. CONCLUDING REMARKS

This work presents a comprehensive framework for evaluating the performance of satellite-based precipitation products (SPPs) in a hydrologic model to simulate and estimate water quality indicators. This study investigates three SPPs based on different retrieval algorithms (the Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis, TMPA 3B42-V7; the Climate Prediction Center's CMORPH V1.0 product; and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System, PERSIANN-CCS) using the Hydrologic Simulation Program FORTRAN (HSPF) hydrology and water quality model developed for the Occoquan Watershed, located in the northern Virginia suburbs of Washington, D.C., United States. This work represents a first attempt to utilize SPPs for water quality modeling, which could be of critical importance in areas of the world where rain-gauge networks or monitoring stations are unavailable.

The research presented herein addresses the scientific questions posed at the onset of this study, as follows:

1. *How well do SPPs perform in estimating precipitation on a continuous basis and during extreme hydrometeorological events?*

The three SPPs, with different spatial resolutions and based on different retrieval algorithms are compared to gauge-based records over a 5-year period across the study

region. On a continuous basis, all three satellite products present moderate agreements with the reference precipitation; CMORPH presenting the best overall performance followed closely by TMPA, and PERSIANN presenting a comparatively inferior performance. SPPs show mixed performance skills with a seasonal dependence. Substantial differences are observed with moderate-intensity precipitation events, especially during winter, with both TMPA and PERSIANN grossly overestimating moderate-intensity precipitation. On the other hand, SPP estimates for high-intensity events are well predicted by PERSIANN but overestimated by TMPA. In the spring, all three SPPs tend to under-detect low-intensity precipitation. TMPA tends to under-detect low-intensity events and over detect moderate- and high-intensity events; however, it is much better able to detect low intensity events in warmer seasons. TMPA, CMORPH, and PERSIANN products are evaluated against gauge-based records for the eight extreme hydrometeorological events. Gauge-based measurements have the greatest measurement variation amongst all of the products, which is attributed to the nature (point) of the measurement. Interestingly, both TMPA and CMORPH tend to overestimate event-based rainfall accumulation for longer events and underestimate accumulation during shorter-duration events, whereas PERSIANN consistently underestimates total event accumulation. In terms of maximum hourly intensity, gauge measurements far exceed all SPPs. However, PERSIANN, with a few exceptions, tends to better detect heavy precipitation in comparison to TMPA and CMORPH.

2. *How well do SPPs perform in simulating streamflow and water quality indicators on a continuous basis and during extreme hydrometeorological events?*

The performance of SPPs in simulating streamflow and water quality is investigated on both a continuous basis and during eight extreme hydrometeorological events. SPP simulations of streamflow show that CMORPH outperforms both TMPA, and PERSIANN when compared to the reference simulation forced with gauge observations. CMORPH has the highest skill for daily streamflow across all evaluation points, nominally increasing as drainage area increases. SPP simulations of TSS show the lowest skill among the water quality indicators evaluated in this study. CMORPH shows the best performance among the SPPs, which generally all overestimate TSS, with a few exceptions. Simulated in-stream TW is well captured by all three SPP simulations with low error metrics. Both DO and BOD show a large spatial variability among the evaluation points of this study. Skills of all three SPPs are high for DO, however, a decreasing skill for each product is detected in the at some locations. In comparison to streamflow and TSS, the variability in the DO metrics is notably smaller, indicating that precipitation has moderate impact on DO than on other water quality indicators evaluated in this analysis. Similar to gauge-based simulations, a strong flux in concentrations is seen in warmer months, which also overlaps with streamflow low-flow periods. The SPPs show satisfactory skills for BOD. BOD concentrations increase in warmer periods and drop during cooler periods following patterns associated with TW, yet this indicator also appears to be influenced by TSS concentrations during peak events. Overall, all three SPP-simulated TP and OP outputs are fairly consistent with no clear indication of one

outperforming another. SPP-simulated NH_4-N and NO_3-N present similar correlations to TSS, indicating that the spatial and temporal differences of SPPs have impact on simulation skill.

3. *How does the performance of SPPs influence the propagation of error between input precipitation and simulated streamflow and water quality indicator output?*

This work evaluates the propagation of error in the precipitation input to that of the simulated water quality output from the Occoquan Watershed HSPF model. Results indicate that the HSPF model may have a dampening effect on the streamflow error for both TMPA and CMORPH; however, due to the poor quality of the PERSIANN product, its error is actually amplified through the model, with a positive dependency on basin scale. The model shows good performance when simulating TW, with TMPA and CMORPH generally outperforming PERSIANN. The error in precipitation is dampened when translated into TW and, as expected, basin scale has no influence on the precipitation-to-TW error propagation. Satisfactory model performance is also shown in simulated DO, with TMPA and CMORPH marginally outperforming PERSIANN. The propagation of systematic and random errors from precipitation to DO vary by season, with larger dampening effects in spring and summer for all three SPPs.

This work, outlining a novel concept to utilize SPPs for water quality modeling, has the potential to be utilized in areas of the world where reference data is unavailable. Future research should evaluate the applicability of SPPs for simulating water quality in different regions, climates, and other hydrometeorological extremes such as droughts and long-term hydroclimatic changes. Additionally, different hydrologic and water quality

models may be used to simulate results, as well as using other precipitation products, such as re-analysis and blended products.

REFERENCES

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research*, *116*(D2), D02115. https://doi.org/10.1029/2010JD014741

Ahn, S.-R., & Kim, S.-J. (2016). The Effect of Rice Straw Mulching and No-Tillage Practice in Upland Crop Areas on Nonpoint-Source Pollution Loads Based on HSPF. *Water*, *8*(3), 106. https://doi.org/10.3390/w8030106

Albek, M., Bakır Öğütveren, Ü., & Albek, E. (2004). Hydrological modeling of Seydi Suyu watershed (Turkey) with HSPF. *Journal of Hydrology*, *285*(1–4), 260–271. https://doi.org/10.1016/j.jhydrol.2003.09.002

Alexander, L. V., Fowler, H. J., Bador, M., Behrangi, A., Donat, M. G., Dunn, R., Funk, C., Goldie, J., Lewis, E., Rogé, M., Seneviratne, S. I., & Venugopal, V. (2019). On the use of indices to study extreme precipitation on sub-daily and daily timescales. *Environmental Research Letters*, *14*(12), 125008. https://doi.org/10.1088/1748-9326/ab51b6

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., & Papadopoulos, A. (2010). Benchmarking High-Resolution Global Satellite Rainfall Products to Radar and Rain-Gauge Rainfall Estimates. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(4), 1667–1683. https://doi.org/10.1109/TGRS.2009.2034736

Azhar, S. C., Aris, A. Z., Yusoff, M. K., Ramli, M. F., & Juahir, H. (2015). Classification of River Water Quality Using Multivariate Analysis. *Procedia Environmental Sciences*, *30*, 79–84. https://doi.org/10.1016/j.proenv.2015.10.014

Barakat, A., El Baghdadi, M., Rais, J., Aghezzaf, B., & Slassi, M. (2016). Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *International Soil and Water Conservation Research*, *4*(4), 284–292. https://doi.org/10.1016/j.iswcr.2016.11.002

Bardossy, A., & Das, T. (2008). Influence of rainfall observation network on model calibration and application. *Hydrol. Earth Syst. Sci.*, 13. https://doi.org/10.5194/hess-12-77-2008

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation

(1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences Discussions*, 1–38. https://doi.org/10.5194/hess-2016-236

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R., & Adler, R. F. (2019). MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

Bengraïne, K., & Marhaba, T. F. (2003). Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials*, *100*(1–3), 179–195. https://doi.org/10.1016/S0304-3894(03)00104-3

Berrisford, P., Dee, D., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S., Uppala, S., & Simmons, A. (2011). *The ERA-Interim Archive, ERA Report Series No.1 Version 2.0*. European Centre for Medium Range Weather Forecasts. http://www.ecmwf.int/publications/

Bezak, N., Munier, S., Fijavž, M. K., Mikoš, M., & Šraj, M. (2017). Estimation of Suspended Sediment Loads Using Copula Functions. *Water*, *9*(8), 628. https://doi.org/10.3390/w9080628

Bharti, V., Singh, C., Ettema, J., & Turkington, T. A. R. (2016). Spatiotemporal characteristics of extreme rainfall events over the Northwest Himalaya using satellite data: SPATIOTEMPORAL CHARACTERISTICS OF EXTREME RAINFALL EVENTS. *International Journal of Climatology*, *36*(12), 3949–3962. https://doi.org/10.1002/joc.4605

Bhuiyan, M. A. E., Nikolopoulos, E. I., Anagnostou, E. N., Quintana-Seguí, P., & Barella-Ortiz, A. (2018). A nonparametric statistical technique for combining global precipitation datasets: Development and hydrological evaluation over the Iberian Peninsula. *Hydrology and Earth System Sciences*, *22*(2), 1371–1389. https://doi.org/10.5194/hess-22-1371-2018

Bitew, M. M., Gebremichael, M., Ghebremichael, L. T., & Bayissa, Y. A. (2012). Evaluation of High-Resolution Satellite Rainfall Products through Streamflow Simulation in a Hydrological Modeling of a Small Mountainous Watershed in Ethiopia. *Journal of Hydrometeorology*, *13*(1), 338–350. https://doi.org/10.1175/2011JHM1292.1

Blume, T., Zehe, E., & Bronstert, A. (2007). Rainfall—Runoff response, event-based runoff coefficients and hydrograph separation. *Hydrological Sciences Journal*, *52*(5), 843–862. https://doi.org/10.1623/hysj.52.5.843

Bu, H., Tan, X., Li, S., & Zhang, Q. (2010). Temporal and spatial variations of water quality in the Jinshui River of the South Qinling Mts., China. *Ecotoxicology and Environmental Safety*, *73*(5), 907–913. https://doi.org/10.1016/j.ecoenv.2009.11.007

Chang, C.-H., Cai, L.-Y., Lin, T.-F., Chung, C.-L., van der Linden, L., & Burch, M. (2015). Assessment of the Impacts of Climate Change on the Water Quality of a Small Deep Reservoir in a Humid-Subtropical Climatic Region. *Water*, *7*(12), 1687–1711. https://doi.org/10.3390/w7041687

Chang, H. (2008). Spatial analysis of water quality trends in the Han River basin, South Korea. *Water Research*, *42*(13), 3285–3304. https://doi.org/10.1016/j.watres.2008.04.006

Chen, S., Liu, B., Tan, X., & Wu, Y. (2020). Inter-comparison of spatiotemporal features of precipitation extremes within six daily precipitation products. *Climate Dynamics*, *54*(1–2), 1057–1076. https://doi.org/10.1007/s00382-019-05045-z

Chintalapudi, S., Sharif, H., & Xie, H. (2014). Sensitivity of Distributed Hydrologic Simulations to Ground and Satellite Based Rainfall Products. *Water*, *6*(5), 1221–1245. https://doi.org/10.3390/w6051221

Cox, B. A., & Whitehead, P. G. (2009). Impacts of climate change scenarios on dissolved oxygen in the River Thames, UK. *Hydrology Research*, *40*(2–3), 138–152. https://doi.org/10.2166/nh.2009.096

de Oliveira, V. A., de Mello, C. R., Beskow, S., Viola, M. R., & Srinivasan, R. (2019). Modeling the effects of climate change on hydrology and sediment load in a headwater basin in the Brazilian Cerrado biome. *Ecological Engineering*, *133*, 20–31. https://doi.org/10.1016/j.ecoleng.2019.04.021

Demirdjian, L., Zhou, Y., & Huffman, G. J. (2018). Statistical Modeling of Extreme Precipitation with TRMM Data. *Journal of Applied Meteorology and Climatology*, *57*(1), 15–30. https://doi.org/10.1175/JAMC-D-17-0023.1

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., Chang, C.-H., Delrieu, G., Hong, Y., Hsu, Y. C., Lavado-Casimiro, W., Manz, B., Moges, S., Nikolopoulos, E. I., Sahlu, D., Salerno, F., Rodríguez-Sánchez, J.-P., Vergara, H. J., & Yilmaz, K. K. (2016). Multiregional Satellite Precipitation Products Evaluation over Complex Terrain. *Journal of Hydrometeorology*, *17*(6), 1817–1836. https://doi.org/10.1175/JHM-D-15-0197.1

Derin, Y., Nikolopoulos, E., & Anagnostou, E. N. (2019). Chapter Seven -Estimating extreme precipitation using multiple satellite-based precipitation products. In V. Maggioni & C. Massari (Eds.), *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing*

Environment (pp. 163–190). Elsevier, Inc. https://doi.org/10.1016/B978-0-12-814899-0.00007-9

Diaz-Ramirez, J., Johnson, B., McAnally, W., Martin, J., & Camacho, R. (2013). Estimation and Propagation of Parameter Uncertainty in Lumped Hydrological Models: A Case Study of HSPF Model Applied to Luxapallila Creek Watershed in Southeast USA. *Journal of Hydrogeology and Hydrologic Engineering*, *02*(01). https://doi.org/10.4172/2325-9647.1000105

Duda, P., Hummel, P., Donigian Jr., A. S., & Imhoff, A. C. (2012). BASINS/HSPF: Model Use, Calibration, and Validation. *Transactions of the ASABE*, *55*(4), 1523–1547. https://doi.org/10.13031/2013.42261

Duque-Gardeazábal, N., Zamora, D., & Rodríguez, E. (2018). *Analysis of the Kernel Bandwidth Influence in the Double Smoothing Merging Algorithm to Improve Rainfall Fields in Poorly Gauged Basins*. 635–626. https://doi.org/10.29007/2xp6

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

Ehsan Bhuiyan, M. A., Nikolopoulos, E. I., Anagnostou, E. N., Polcher, J., Albergel, C., Dutra, E., Fink, G., Martínez-de la Torre, A., & Munier, S. (2019). Assessment of precipitation error propagation in multi-model global water resource reanalysis. *Hydrology and Earth System Sciences*, *23*(4), 1973–1994. https://doi.org/10.5194/hess-23-1973-2019

Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A., & Diniz, F. L. R. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins–Araguaia basin in Brazil. *Journal of Hydrology*, *527*, 943–957. https://doi.org/10.1016/j.jhydrol.2015.05.042

Fovet, O., Humbert, G., Dupas, R., Gascuel-Odoux, C., Gruau, G., Jaffrezic, A., Thelusma, G., Faucheux, M., Gilliet, N., Hamon, Y., & Grimaldi, C. (2018). Seasonal variability of stream water quality response to storm events captured using high-frequency and multi-parameter data. *Journal of Hydrology*, *559*, 282–293. https://doi.org/10.1016/j.jhydrol.2018.02.040

Gebregiorgis, A. S., & Hossain, F. (2013). Understanding the Dependence of Satellite Rainfall Uncertainty on Topography and Climate for Hydrologic Model Simulation. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 704–718. https://doi.org/10.1109/TGRS.2012.2196282

Gebremichael, M., Liao, G.-Y., & Yan, J. (2011). Nonparametric error model for a high resolution satellite rainfall product: NONPARAMETRIC ERROR MODEL. *Water Resources Research*, *47*(7). https://doi.org/10.1029/2010WR009667

Gelca, R., Hayhoe, K., Scott-Fleming, I., Crow, C., Dawson, D., & Patiño, R. (2016). Climate-water quality relationships in Texas reservoirs: Climate-Water Quality Relationships in Texas Reservoirs. *Hydrological Processes*, *30*(1), 12–29. https://doi.org/10.1002/hyp.10545

Girons Lopez, M., Wennerström, H., Nordén, L., & Seibert, J. (2015). Location and density of rain gauges for the estimation of spatial varying precipitation. *Geografiska Annaler: Series A, Physical Geography*, *97*(1), 167–179. https://doi.org/10.1111/geoa.12094

Gourley, J. J., Hong, Y., Flamig, Z. L., Wang, J., Vergara, H., & Anagnostou, E. N. (2011). Hydrologic Evaluation of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations on Ft. Cobb Basin, Oklahoma. *Journal of Hydrometeorology*, *12*(5), 973–988. https://doi.org/10.1175/2011JHM1287.1

Guo, H., Chen, S., Bao, A., Hu, J., Yang, B., & Stepanian, P. (2015). Comprehensive Evaluation of High-Resolution Satellite-Based Precipitation Products over China. *Atmosphere*, *7*(1), 6. https://doi.org/10.3390/atmos7010006

Guo, H., Chen, S., Bao, A., Hu, J., Yang, B., & Stepanian, P. (2016). Comprehensive Evaluation of High-Resolution Satellite-Based Precipitation Products over China. *Atmosphere*, *7*(1), 6. https://doi.org/10.3390/atmos7010006

Guo, R., & Liu, Y. (2016). Evaluation of Satellite Precipitation Products with Rain Gauge Data at Different Scales: Implications for Hydrological Applications. *Water*, *8*(7), 281. https://doi.org/10.3390/w8070281

Habib, E., ElSaadani, M., & Haile, A. T. (2012). Climatology-Focused Evaluation of CMORPH and TMPA Satellite Rainfall Products over the Nile Basin. *Journal of Applied Meteorology and Climatology*, *51*(12), 2105–2121. https://doi.org/10.1175/JAMC-D-11-0252.1

Habib, E., Haile, A. T., Tian, Y., & Joyce, R. J. (2012). Evaluation of the High-Resolution CMORPH Satellite Rainfall Product Using Dense Rain Gauge Observations and Radar-Based Estimates. *Journal of Hydrometeorology*, *13*(6), 1784–1798. https://doi.org/10.1175/JHM-D-12-017.1

Habib, E., Henschke, A., & Adler, R. F. (2009). Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmospheric Research*, *94*(3), 373–388. https://doi.org/10.1016/j.atmosres.2009.06.015

Hayashi, S., Murakami, S., Watanabe, M., & Bao-Hua, X. (2004). HSPF Simulation of Runoff and Sediment Loads in the Upper Changjiang River Basin, China. *Journal of Environmental Engineering*, *130*(7), 801–815. https://doi.org/10.1061/(ASCE)0733-9372(2004)130:7(801)

Hazra, A., Maggioni, V., Houser, P., Antil, H., & Noonan, M. (2019). A Monte Carlo-based multi-objective optimization approach to merge different precipitation estimates for land surface modeling. *Journal of Hydrology*, *570*, 454–462. https://doi.org/10.1016/j.jhydrol.2018.12.039

Hema, N., & Kant, K. (2017). Reconstructing missing hourly real-time precipitation data using a novel intermittent sliding window period technique for automatic weather station data. *Journal of Meteorological Research*, *31*(4), 774–790. https://doi.org/10.1007/s13351-017-6084-8

Himanshu, S. K., Pandey, A., & Yadav, B. (2017). Assessing the applicability of TMPA-3B42V7 precipitation dataset in wavelet-support vector machine approach for suspended sediment load prediction. *Journal of Hydrology*, *550*, 103–117. https://doi.org/10.1016/j.jhydrol.2017.04.051

Hong, Y., Gochis, D., Cheng, J., Hsu, K., & Sorooshian, S. (2007). Evaluation of PERSIANN-CCS Rainfall Measurement Using the NAME Event Rain Gauge Network. *Journal of Hydrometeorology*, *8*(3), 469–482. https://doi.org/10.1175/JHM574.1

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

Hossain, F., & Anagnostou, E. (2004). Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. *Journal of Geophysical Research*, *109*(D7), D07102. https://doi.org/10.1029/2003JD003986

Hostache, R., Matgen, P., Montanari, A., Montanari, M., Hoffmann, L., & Pfister, L. (2011). Propagation of uncertainties in coupled hydro-meteorological forecasting systems: A stochastic approach for the assessment of the total predictive uncertainty. *Atmospheric Research*, *100*(2–3), 263–274. https://doi.org/10.1016/j.atmosres.2010.09.014

Hsu, K., Behrangi, A., Imam, B., & Sorooshian, S. (2010). Extreme Precipitation Estimation Using Satellite-Based PERSIANN-CCS Algorithm. In M Gebremichael & F. Hossain (Eds.), *Satellite Rainfall Applications for Surface Hydrology*. Springer. https://doi.org/10.1007/978-90-481-2915-7_4

Hsu, K., Gao, X., Sorooshian, S., & Gupta, H. V. (1997). Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *Journal of Applied Meteorology*, *36*(9), 1176–1190. https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2

Huang, C., Chen, X., Li, Y., Yang, H., Sun, D., Li, J., Le, C., Zhou, L., Zhang, M., & Xu, L. (2015). Specific inherent optical properties of highly turbid productive water for retrieval of water quality after optical classification. *Environmental Earth Sciences*, *73*(5), 1961–1973. https://doi.org/10.1007/s12665-014-3548-3

Huang, Y., Chen, S., Cao, Q., Hong, Y., Wu, B., Huang, M., Qiao, L., Zhang, Z., Li, Z., Li, W., & Yang, X. (2013). Evaluation of Version-7 TRMM Multi-Satellite Precipitation Analysis Product during the Beijing Extreme Heavy Rainfall Event of 21 July 2012. *Water*, *6*(1), 32–44. https://doi.org/10.3390/w6010032

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2018). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG)* (p. 31). National Aeronautics and Space Administration. https://pmm.nasa.gov/category/keywords/imerg

Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2010). The TRMM Multi-Satellite Precipitation Analysis (TMPA). In M Gebremichael & F. Hossain (Eds.), *Satellite Rainfall Applications for Surface Hydrology* (pp. 3–22). Springer. https://doi.org/10.1007/978-90-481-2915-7_1

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

Huo, S.-C., Lo, S.-L., Chiu, C.-H., Chiueh, P.-T., & Yang, C.-S. (2015). Assessing a fuzzy model and HSPF to supplement rainfall data for nonpoint source water quality in the Feitsui reservoir watershed. *Environmental Modelling & Software*, *72*, 110–116. https://doi.org/10.1016/j.envsoft.2015.07.002

Hussain, Y., Satgé, F., Hussain, M. B., Martinez-Carvajal, H., Bonnet, M.-P., Cárdenas-Soto, M., Roig, H. L., & Akhter, G. (2018). Performance of CMORPH, TMPA, and PERSIANN rainfall datasets over plain, mountainous, and glacial regions of Pakistan. *Theoretical and Applied Climatology*, *131*(3–4), 1119–1132. https://doi.org/10.1007/s00704-016-2027-z

Irby, I. D., Friedrichs, M. A. M., Da, F., & Hinson, K. E. (2018). The competing impacts of climate change and nutrient reductions on dissolved oxygen in Chesapeake Bay. *Biogeosciences*, *15*(9), 2649–2668. https://doi.org/10.5194/bg-15-2649-2018

Jeznach, L. C., Hagemann, M., Park, M.-H., & Tobiason, J. E. (2017). Proactive modeling of water quality impacts of extreme precipitation events in a drinking water reservoir. *Journal of Environmental Management*, *201*, 241–251. https://doi.org/10.1016/j.jenvman.2017.06.047

Jiang, Shanhu, Liu, S., Ren, L., Yong, B., Zhang, L., Wang, M., Lu, Y., & He, Y. (2017). Hydrologic Evaluation of Six High Resolution Satellite Precipitation Products in Capturing Extreme Precipitation and Streamflow over a Medium-Sized Basin in China. *Water*, *10*(1), 25. https://doi.org/10.3390/w10010025

Jiang, Shanshan, Zhang, Z., Huang, Y., Chen, X., & Chen, S. (2017). Evaluating the TRMM Multisatellite Precipitation Analysis for Extreme Precipitation and Streamflow in Ganjiang River Basin, China. *Advances in Meteorology*, *2017*, 1–11. https://doi.org/10.1155/2017/2902493

Johnson, T. E., Butcher, J. B., Parker, A., & Weaver, C. P. (2012). Investigating the Sensitivity of U.S. Streamflow and Water Quality to Climate Change: U.S. EPA Global Change Research Program's 20 Watersheds Project. *Journal of Water Resources Planning and Management*, *138*(5), 453–464. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000175

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*, 487–503. https://doi.org/10.1175/1525-7541(2004)005%3C0487:CAMTPG%3E2.0.CO;2

Jung, K. Y., Lee, K.-L., Im, T. H., Lee, I. J., Kim, S., Han, K.-Y., & Ahn, J. M. (2016). Evaluation of water quality for the Nakdong River watershed using multivariate analysis. *Environmental Technology & Innovation*, *5*, 67–82. https://doi.org/10.1016/j.eti.2015.12.001

Kang, J.-H., Lee, S. W., Cho, K. H., Ki, S. J., Cha, S. M., & Kim, J. H. (2010). Linking land-use type and stream water quality using spatial data of fecal indicator bacteria and heavy metals in the Yeongsan river basin. *Water Research*, *44*(14), 4143–4157. https://doi.org/10.1016/j.watres.2010.05.009

Katiraie-Boroujerdy, P.-S., Ashouri, H., Hsu, K., & Sorooshian, S. (2017). Trends of precipitation extreme indices over a subtropical semi-arid area using PERSIANN-CDR. *Theoretical and Applied Climatology*, *130*(1–2), 249–260. https://doi.org/10.1007/s00704-016-1884-9

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, How Much of the Earth's Surface is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kim, J., & Ryu, J. H. (2015). Quantifying a Threshold of Missing Values for Gap Filling Processes in Daily Precipitation Series. *Water Resources Management*, *29*(11), 4173–4184. https://doi.org/10.1007/s11269-015-1052-5

Kisi, O., & Ay, M. (2014). Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey. *Journal of Hydrology*, *513*, 362–375. https://doi.org/10.1016/j.jhydrol.2014.03.005

Kubota, T., Ushio, T., Shige, S., Kida, S., Kachi, M., & Okamoto, K. (2009). Verification of High-Resolution Satellite-Based Rainfall Estimates around Japan Using a Gauge-Calibrated Ground-Radar Dataset. *Journal of the Meteorological Society of Japan*, *87A*, 203–222. https://doi.org/10.2151/jmsj.87A.203

Kwon, E.-H., Sohn, B.-J., Chang, D.-E., Ahn, M.-H., & Yang, S. (2008). Use of Numerical Forecasts for Improving TMI Rain Retrievals over the Mountainous Area in Korea. *Journal of Applied Meteorology and Climatology*, *47*(7), 1995–2007. https://doi.org/10.1175/2007JAMC1857.1

Lee, H., & Kang, K. (2015). Interpolation of Missing Precipitation Data Using Kernel Estimations for Hydrologic Modeling. *Advances in Meteorology*, *2015*, 1–12. https://doi.org/10.1155/2015/935868

Li, J., Sorooshian, S., Higgins, W., Gao, X., Imam, B., & Hsu, K. (2008). Influence of Spatial Resolution on Diurnal Variability during the North American Monsoon. *Journal of Climate*, *21*(16), 3967–3988. https://doi.org/10.1175/2008JCLI2022.1

Li, X., Zhang, Q., & Ye, X. (2013). Dry/Wet Conditions Monitoring Based on TRMM Rainfall Data and Its Reliability Validation over Poyang Lake Basin, China. *Water*, *5*(4), 1848–1864. https://doi.org/10.3390/w5041848

Li, Z., Liu, H., Luo, C., Li, Y., Li, H., Pan, J., Jiang, X., Zhou, Q., & Xiong, Z. (2015). Simulation of runoff and nutrient export from a typical small watershed in China using the Hydrological Simulation Program–Fortran. *Environmental Science and Pollution Research*, *22*(10), 7954–7966. https://doi.org/10.1007/s11356-014-3960-y

Lockhoff, M., Zolina, O., Simmer, C., & Schulz, J. (2014). Evaluation of Satellite-Retrieved Extreme Precipitation over Europe using Gauge Observations. *Journal of Climate*, *27*(2), 607–623. https://doi.org/10.1175/JCLI-D-13-00194.1

Ma, D., Xu, Y.-P., Gu, H., Zhu, Q., Sun, Z., & Xuan, W. (2019). Role of satellite and reanalysis precipitation products in streamflow and sediment modeling over a typical alpine and gorge region in Southwest China. *Science of The Total Environment*, *685*, 934–950. https://doi.org/10.1016/j.scitotenv.2019.06.183

Maggioni, V., & Massari, C. (2018). On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology*, *558*, 214–224. https://doi.org/10.1016/j.jhydrol.2018.01.039

Maggioni, V., & Massari, C. (Eds.). (2019). *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment A Remote Sensing Approach* (1st ed.). Elsevier, Inc. https://doi.org/10.1016/C2017-0-02344-3

Maggioni, V., Meyers, P. C., & Robinson, M. D. (2016). A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *Journal of Hydrometeorology*, *00*, 17. https://doi.org/DOI: 10.1175/JHM-D-15-0190.1

Maggioni, V., Nikolopoulos, E. I., Anagnostou, E. N., & Borga, M. (2017). Modeling Satellite Precipitation Errors Over Mountainous Terrain: The Influence of Gauge Density, Seasonality, and Temporal Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 4130–4140. https://doi.org/10.1109/TGRS.2017.2688998

Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., & Stampoulis, D. (2013). Investigating the Applicability of Error Correction Ensembles of Satellite Rainfall Products in River Flow Simulations. *Journal of Hydrometeorology*, *14*(4), 1194–1211. https://doi.org/10.1175/JHM-D-12-074.1

Mahbod, M., Shirvani, A., & Veronesi, F. (2019). A comparative analysis of the precipitation extremes obtained from tropical rainfall-measuring mission satellite and rain gauges datasets over a semiarid region. *International Journal of Climatology*, *39*(1), 495–515. https://doi.org/10.1002/joc.5824

Maldonado, P. P., & Moglen, G. E. (2012). Low-flow variations in source water supply for the Occoquan reservoir system based on a 100-year climate forecast. *Journal of Hydrologic Engineering*, *18*(7), 787–796. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000623

McMillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, *400*(1–2), 83–94. https://doi.org/10.1016/j.jhydrol.2011.01.026

Mehran, A., & AghaKouchak, A. (2014). Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations: capabilities of

satellite data to estimate heavy precipitation rates. *Hydrological Processes*, *28*(4), 2262–2270. https://doi.org/10.1002/hyp.9779

Mei, K., Liao, L., Zhu, Y., Lu, P., Wang, Z., Dahlgren, R. A., & Zhang, M. (2014). Evaluation of spatial-temporal variations and trends in surface water quality across a rural-suburban-urban interface. *Environmental Science and Pollution Research*, *21*(13), 8036–8051. https://doi.org/10.1007/s11356-014-2716-z

Mei, Y., Anagnostou, E. N., Nikolopoulos, E. I., & Borga, M. (2014). Error Analysis of Satellite Precipitation Products in Mountainous Basins. *Journal of Hydrometeorology*, *15*(5), 1778–1793. https://doi.org/10.1175/JHM-D-13-0194.1

Mei, Y., Nikolopoulos, E., Anagnostou, E., Zoccatelli, D., & Borga, M. (2016). Error Analysis of Satellite Precipitation-Driven Modeling of Flood Events in Complex Alpine Terrain. *Remote Sensing*, *8*(4), 293. https://doi.org/10.3390/rs8040293

Mei, Y., Nikolopoulos, E. I., Anagnostou, E. N., & Borga, M. (2016). Evaluating Satellite Precipitation Error Propagation in Runoff Simulations of Mountainous Basins. *Journal of Hydrometeorology*, *17*(5), 1407–1423. https://doi.org/10.1175/JHM-D-15-0081.1

Meng, J., Li, L., Hao, Z., Wang, J., & Shao, Q. (2014). Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River. *Journal of Hydrology*, *509*, 320–332. https://doi.org/10.1016/j.jhydrol.2013.11.049

Milewski, A., Elkadiri, R., & Durham, M. (2015). Assessment and Comparison of TMPA Satellite Precipitation Products in Varying Climatic and Topographic Regimes in Morocco. *Remote Sensing*, *7*(5), 5697–5717. https://doi.org/10.3390/rs70505697

Mishra, A., Kar, S., & Singh, V. P. (2007). Determination of runoff and sediment yield from a small watershed in sub-humid subtropics using the HSPF model. *Hydrological Processes*, *21*(22), 3035–3045. https://doi.org/10.1002/hyp.6514

Moreno-Rodenas, A. M., Tscheikner-Gratl, F., Langeveld, J. G., & Clemens, F. H. L. R. (2019). Uncertainty analysis in a large-scale water quality integrated catchment modelling study. *Water Research*, *158*, 46–60. https://doi.org/10.1016/j.watres.2019.04.016

Moulin, L., Gaume, E., Obled, C., & Paris-Est, U. (2009). Uncertainties on mean areal precipitation: Assessment and impact on streamflow simulations. *Hydrol. Earth Syst. Sci.*, 18. https://hal-insu.archives-ouvertes.fr/insu-00413244/

Murdoch, P. S., Baron, J. S., & Miller, T. L. (2000). Potential Effects of Climate Change on Surface-Water Quality in North America. *Journal of the American Water Resources Association*, *36*(2), 347–366. https://doi.org/doi:10.1111/j.1752-1688.2000.tb04273.x

Nastos, P. T., Kapsomenakis, J., & Douvis, K. C. (2013). Analysis of precipitation extremes based on satellite and high-resolution gridded data set over Mediterranean basin. *Atmospheric Research*, *131*, 46–59. https://doi.org/10.1016/j.atmosres.2013.04.009

Neal, C., Reynolds, B., Rowland, P., Norris, D., Kirchner, J. W., Neal, M., Sleep, D., Lawlor, A., Woods, C., Thacker, S., Guyatt, H., Vincent, C., Hockenhull, K., Wickham, H., Harman, S., & Armstrong, L. (2012). High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge. *Science of The Total Environment*, *434*, 3–12. https://doi.org/10.1016/j.scitotenv.2011.10.072

Nijssen, B., & Lettenmaier, D. P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research*, *109*(D2). https://doi.org/10.1029/2003JD003497

Nikolopoulos, E. I., Bartsotas, N. S., Anagnostou, E. N., & Kallos, G. (2015). Using High-Resolution Numerical Weather Forecasts to Improve Remotely Sensed Rainfall Estimates: The Case of the 2013 Colorado Flash Flood. *Journal of Hydrometeorology*, *16*(4), 1742–1751. https://doi.org/10.1175/JHM-D-14-0207.1

Nikolopoulos, Efthymios I., Anagnostou, E. N., & Borga, M. (2013). Using High-Resolution Satellite Rainfall Products to Simulate a Major Flash Flood Event in Northern Italy. *Journal of Hydrometeorology*, *14*(1), 171–185. https://doi.org/10.1175/JHM-D-12-09.1

Nikolopoulos, Efthymios I., Anagnostou, E. N., Hossain, F., Gebremichael, M., & Borga, M. (2010). Understanding the Scale Relationships of Uncertainty Propagation of Satellite Rainfall through a Distributed Hydrologic Model. *Journal of Hydrometeorology*, *11*(2), 520–532. https://doi.org/10.1175/2009JHM1169.1

Noori, R., Sabahi, M. S., Karbassi, A. R., Baghvand, A., & Taati Zadeh, H. (2010). Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination*, *260*(1–3), 129–136. https://doi.org/10.1016/j.desal.2010.04.053

Pombo, S., & de Oliveira, R. P. (2015). Evaluation of extreme precipitation estimates from TRMM in Angola. *Journal of Hydrology*, *523*, 663–679. https://doi.org/10.1016/j.jhydrol.2015.02.014

Porcacchia, L., Kirstetter, P. E., Gourley, J. J., Maggioni, V., Cheong, B. L., & Anagnostou, M. N. (2017). Toward a Polarimetric Radar Classification Scheme for Coalescence-Dominant Precipitation: Application to Complex Terrain. *Journal of Hydrometeorology*, *18*(12), 3199–3215. https://doi.org/10.1175/JHM-D-17-0016.1

Qin, Z., Peng, T., Singh, V. P., & Chen, M. (2019). Spatio-temporal variations of precipitation extremes in Hanjiang River Basin, China, during 1960–2015. *Theoretical and Applied Climatology*, *138*(3–4), 1767–1783. https://doi.org/10.1007/s00704-019-02932-7

Qiu, J., Shen, Z., Wei, G., Wang, G., Xie, H., & Lv, G. (2018). A systematic assessment of watershed-scale nonpoint source pollution during rainfall-runoff events in the Miyun Reservoir watershed. *Environmental Science and Pollution Research*, *25*(7), 6514–6531. https://doi.org/10.1007/s11356-017-0946-6

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., Black, J., Thériault, J. M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., & Gutmann, E. (2012). How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bulletin of the American Meteorological Society*, *93*(6), 811–829. https://doi.org/10.1175/BAMS-D-11-00052.1

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., … Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Roebber, P. J. (2009). Visualizing Multiple Measures of Forecast Quality. *Weather and Forecasting*, *24*(2), 601–608. https://doi.org/10.1175/2008WAF2222159.1

Rue, G. P., Rock, N. D., Gabor, R. S., Pitlick, J., Tfaily, M., & McKnight, D. M. (2017). Concentration-discharge relationships during an extreme event: Contrasting behavior of solutes and changes to chemical quality of dissolved organic material in the Boulder Creek Watershed during the September 2013 flood: SOLUTE FLUX IN A FLOOD EVENT. *Water Resources Research*, *53*(7), 5276–5297. https://doi.org/10.1002/2016WR019708

Schaefer, J. T. (1990). The Critical Success Index as an Indicator of Warning Skill. *Weather and Forecasing*, *570–575*. https://doi.org/10.1175/1520-0434(1990)005%3C0570:TCSIAA%3E2.0.CO;2

Seyyedi, H., Anagnostou, E. N., Beighley, E., & McCollum, J. (2015). Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. *Atmospheric Research*, *164–165*, 37–48. https://doi.org/10.1016/j.atmosres.2015.03.019

Shah, H. L., & Mishra, V. (2016). Uncertainty and Bias in Satellite-Based Precipitation Estimates over Indian Subcontinental Basins: Implications for Real-Time Streamflow Simulation and Flood Prediction. *Journal of Hydrometeorology*, *17*(2), 615–636. https://doi.org/10.1175/JHM-D-15-0115.1

Sharifi, E., Steinacker, R., & Saghafian, B. (2018). Multi time-scale evaluation of high-resolution satellite-based precipitation products over northeast of Austria. *Atmospheric Research*, *206*, 46–63. https://doi.org/10.1016/j.atmosres.2018.02.020

Solakian, J., Maggioni, V., Lodhi, A., & Godrej, A. (2019). Investigating the use of satellite-based precipitation products for monitoring water quality in the Occoquan Watershed. *Journal of Hydrology: Regional Studies*, *26*, 100630. https://doi.org/10.1016/j.ejrh.2019.100630

Soler, M., Regüés, D., Latron, J., & Gallart, F. (2007). Frequency–magnitude relationships for precipitation, stream flow and sediment load events in a small Mediterranean basin (Vallcebre basin, Eastern Pyrenees). *CATENA*, *71*(1), 164–171. https://doi.org/10.1016/j.catena.2006.06.009

Sorooshian, S., AghaKouchak, A., Arkin, P., Eylander, J., Foufoula-Georgiou, E., Harmon, R., Hendrickx, J. M. H., Imam, B., Kuligowski, R., Skahill, B., & Skofronick-Jackson, G. (2011). Advanced Concepts on Remote Sensing of Precipitation at Multiple Scales. *Bulletin of the American Meteorological Society*, *92*(10), 1353–1357. https://doi.org/10.1175/2011BAMS3158.1

Sorooshian, S., Hsu, K., Gao, X., Gupta, H. V., Imam, B., & Braithwaite, D. (2000). Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall. *Bulletin of the American Meteorogogical Society*, *81*(9), 2035–2046. https://doi.org/10.1175/1520-0477(2000)081%3C2035:EOPSSE%3E2.3.CO;2

Stern, M., Flint, L., Minear, J., Flint, A., & Wright, S. (2016). Characterizing Changes in Streamflow and Sediment Supply in the Sacramento River Basin, California, Using Hydrological Simulation Program—FORTRAN (HSPF). *Water*, *8*(10), 432. https://doi.org/10.3390/w8100432

Stryker, J., Wemple, B., & Bomblies, A. (2017). Modeling sediment mobilization using a distributed hydrological model coupled with a bank stability model. *Water Resources Research*, *53*(3), 2051–2073. https://doi.org/10.1002/2016WR019143

Su, J., Lü, H., Wang, J., Sadeghi, A., & Zhu, Y. (2017). Evaluating the Applicability of Four Latest Satellite–Gauge Combined Precipitation Estimates for Extreme Precipitation and Streamflow Predictions over the Upper Yellow River Basins in China. *Remote Sensing*, *9*(11), 1176. https://doi.org/10.3390/rs9111176

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Sun, R., Yuan, H., Liu, X., & Jiang, X. (2016). Evaluation of the latest satellite–gauge precipitation products and their hydrologic applications over the Huaihe River basin. *Journal of Hydrology*, *536*, 302–319. https://doi.org/10.1016/j.jhydrol.2016.02.054

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Thorne, O., & Fenner, R. A. (2011). The impact of climate change on reservoir water quality and water treatment plant operations: A UK case study: The impact of climate change on reservoir water quality and WTP operations. *Water and Environment Journal*, *25*(1), 74–87. https://doi.org/10.1111/j.1747-6593.2009.00194.x

Tongal, H. (2019). Spatiotemporal analysis of precipitation and extreme indices in the Antalya Basin, Turkey. *Theoretical and Applied Climatology*, *138*(3–4), 1735–1754. https://doi.org/10.1007/s00704-019-02927-4

U.S. Environmental Protection Agency. (2000). *Meteorological Monitoring Guidance for Regulatory Modeling Applications* (Guidance Document EPA-454/R-99-005; p. 171). U.S. Environmental Protection Agency. https://www3.epa.gov/scram001/guidance/met/mmgrma.pdf

Verdin, A., Funk, C., Rajagopalan, B., & Kleiber, W. (2016). Kriging and Local Polynomial Methods for Blending Satellite-Derived and Gauge Precipitation Estimates to Support Hydrologic Early Warning Systems. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2552–2562. https://doi.org/10.1109/TGRS.2015.2502956

Vergara, H., Hong, Y., Gourley, J. J., Anagnostou, E. N., Maggioni, V., Stampoulis, D., & Kirstetter, P.-E. (2014). Effects of Resolution of Satellite-Based Rainfall Estimates on Hydrologic Modeling Skill at Different Scales. *Journal of Hydrometeorology*, *15*(2), 593–613. https://doi.org/10.1175/JHM-D-12-0113.1

Villarini, G., Smith, J. A., Serinaldi, F., & Ntelekos, A. A. (2011). Analyses of seasonal and annual maximum daily discharge records for central Europe. *Journal of Hydrology*, *399*(3–4), 299–312. https://doi.org/10.1016/j.jhydrol.2011.01.007

Vivoni, E. R., Entekhabi, D., & Hoffman, R. N. (2007). Error Propagation of Radar Rainfall Nowcasting Fields through a Fully Distributed Flood Forecasting Model. *Journal of Applied Meteorology and Climatology*, *46*(6), 932–940. https://doi.org/10.1175/JAM2506.1

von Freyberg, J., Studer, B., & Kirchner, J. W. (2017). A lab in the field: High-frequency analysis of water quality and stable isotopes in stream water and precipitation. *Hydrology*

and Earth System Sciences, *21*(3), 1721–1739. https://doi.org/10.5194/hess-21-1721-2017

Wang, L., Wang, Y., Xu, C., An, Z., & Wang, S. (2011). Analysis and evaluation of the source of heavy metals in water of the River Changjiang. *Environmental Monitoring and Assessment*, *173*(1–4), 301–313. https://doi.org/10.1007/s10661-010-1388-5

Wu, J., Yu, S. L., & Zou, R. (2006). A WATER QUALITY-BASED APPROACH FOR WATERSHED WIDE BMP STRATEGIES1. *JAWRA Journal of the American Water Resources Association*, *42*(5), 1193–1204. https://doi.org/10.1111/j.1752-1688.2006.tb05294.x

Wunderlin, D. A., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A., & de los Ángeles, B. M. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquía River Basin (Córdoba–Argentina). *Water Research*, *35*(12), 2881–2894. https://doi.org/10.1016/S0043-1354(00)00592-3

Xie, H., Shen, Z., Chen, L., Lai, X., Qiu, J., Wei, G., Dong, J., Peng, Y., & Chen, X. (2019). Parameter Estimation and Uncertainty Analysis: A Comparison between Continuous and Event-Based Modeling of Streamflow Based on the Hydrological Simulation Program–Fortran (HSPF) Model. *Water*, *11*(1), 171. https://doi.org/10.3390/w11010171

Xie, H., Wei, G., Shen, Z., Dong, J., Peng, Y., & Chen, X. (2019). Event-based uncertainty assessment of sediment modeling in a data-scarce catchment. *CATENA*, *173*, 162–174. https://doi.org/10.1016/j.catena.2018.10.008

Xu, H. S., Xu, Z. X., Wu, W., & Tang, F. F. (2012). Assessment and Spatiotemporal Variation Analysis of Water Quality in the Zhangweinan River Basin, China. *Procedia Environmental Sciences*, *13*, 1641–1652. https://doi.org/10.1016/j.proenv.2012.01.157

Xu, X., Li, J., & Tolson, B. A. (2014). Progress in integrating remote sensing data and hydrologic modeling. *Progress in Physical Geography*, *38*(4), 464–498. https://doi.org/10.1177/0309133314536583

Xu, Z. (2005). *A Complex, Linked Watershed-Reservoir Hydrology and Water Quality Model Application for the Occoquan Watershed, Virginia* [Dissertation, Virginia Polytechnic Institute and State University]. https://vtechworks.lib.vt.edu/bitstream/handle/10919/37186/dissertation_zhongyan.pdf

Xu, Z., Godrej, A. N., & Grizzard, T. J. (2007). The hydrological calibration and validation of a complexly-linked watershed–reservoir model for the Occoquan watershed, Virginia. *Journal of Hydrology*, *345*(3–4), 167–183. https://doi.org/10.1016/j.jhydrol.2007.07.015

Yang, N., Zhang, K., Hong, Y., Zhao, Q., Huang, Q., Xu, Y., Xue, X., & Chen, S. (2017). Evaluation of the TRMM multisatellite precipitation analysis and its applicability in supporting reservoir operation and water resources management in Hanjiang basin, China. *Journal of Hydrology*, *549*, 313–325. https://doi.org/10.1016/j.jhydrol.2017.04.006

Yang, Yanfen, & Luo, Y. (2014). Evaluating the performance of remote sensing precipitation products CMORPH, PERSIANN, and TMPA, in the arid region of northwest China. *Theoretical and Applied Climatology*, *118*(3), 429–445. https://doi.org/10.1007/s00704-013-1072-0

Yang, Yumeng, Du, J., Cheng, L., & Xu, W. (2017). Applicability of TRMM satellite precipitation in driving hydrological model for identifying flood events: A case study in the Xiangjiang River Basin, China. *Natural Hazards*, *87*(3), 1489–1505. https://doi.org/10.1007/s11069-017-2836-0

Young, C. B., Bradley, A. A., Krajewski, W. F., Kruger, A., & Morrissey, M. L. (2000). Evaluating NEXRAD Multisensor Precipitation Estimates for Operational Hydrologic Forecasting. *JOURNAL OF HYDROMETEOROLOGY*, *1*, 14. https://doi.org/10.1175/1525-7541(2000)001<0241:ENMPEF>2.0.CO;2

Yu, W., Nakakita, E., Kim, S., & Yamaguchi, K. (2016). Impact Assessment of Uncertainty Propagation of Ensemble NWP Rainfall to Flood Forecasting with Catchment Scale. *Advances in Meteorology*, *2016*, 1–17. https://doi.org/10.1155/2016/1384302

Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., Chen, J., Guo, S.-L., & Liu, J. (2018). The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. *Journal of Hydrology*, *563*, 106–122. https://doi.org/10.1016/j.jhydrol.2018.05.058

Zhang, Y., Hong, Y., Wang, X., Gourley, J. J., Xue, X., Saharia, M., Ni, G., Wang, G., Huang, Y., Chen, S., & Tang, G. (2015). Hydrometeorological Analysis and Remote Sensing of Extremes: Was the July 2012 Beijing Flood Event Detectable and Predictable by Global Satellite Observing and Global Weather Modeling Systems? *Journal of Hydrometeorology*, *16*(1), 381–395. https://doi.org/10.1175/JHM-D-14-0048.1

Zhu, B., Chen, J., & Chen, H. (2019). Performance of multiple probability distributions in generating daily precipitation for the simulation of hydrological extremes. *Stochastic Environmental Research and Risk Assessment*, *33*(8–9), 1581–1592. https://doi.org/10.1007/s00477-019-01720-z

Zhu, D., Peng, D. Z., & Cluckie, I. D. (2013). Statistical analysis of error propagation from radar rainfall to hydrological models. *Hydrology and Earth System Sciences*, *17*(4), 1445–1453. https://doi.org/10.5194/hess-17-1445-2013

Zhu, Q., Xuan, W., Liu, L., & Xu, Y.-P. (2016). Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR, TRMM 3B42V7, and NCEP-CFSR over humid regions in China: Evaluation and Hydrological Application of Precipitation Estimates. *Hydrological Processes*, *30*(17), 3061–3083. https://doi.org/10.1002/hyp.10846

Zhu, Y., Lin, Z., Zhao, Y., Li, H., He, F., Zhai, J., Wang, L., & Wang, Q. (2017). Flood Simulations and Uncertainty Analysis for the Pearl River Basin Using the Coupled Land Surface and Hydrological Model System. *Water*, *9*(6), 391. https://doi.org/10.3390/w9060391

BIOGRAPHY

Jennifer Solakian received her Bachelor of Science in Biological Systems Engineering from Virginia Polytechnic Institute and State University (VA Tech) in 2001 and her Master of Science in Environmental Management from the University of Maryland in 2005. Since 2001, she has been employed as a water resources engineer. Her professional experience comprises of various projects involving hydrologic and non-point source pollution modeling, water quality, and watershed management. She joined the Ph.D. program in Civil, Environmental and Infrastructure Engineering at George Mason University in 2014 with an interest in linking the fields of water quality and remote sensing.