Assessment of Precipitation Error Propagation in Discharge Simulations over the Contiguous United States

Nergui Nanding[1, 2], Huan Wu[1, 2, 3, *], Jing Tao[4], Viviana Maggioni[5], Hylke E. Beck[6], Naijun Zhou[7], Maoyi Huang[8], Zhijun Huang[1]

[1]Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Guangzhou, China

[2]Southern Marine Science and Engineering Laboratory, Zhuhai, China

[3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, MD 20740, USA

[4]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA,

[5]George Mason University, Fairfax, VA, USA

[6]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

[7]Department of Geographical Sciences, University of Maryland, College Park, Maryland, MD 20740, USA

[8]Office of Science and Technology Integration, National Weather Service, National Oceanic and Atmospheric Administration, Silver Spring, MD 20910

[*]Corresponding author address: Dr. Huan Wu, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, P. R. China

Email: wuhuan3@mail.sysu.edu.cn

Early Online Release: This preliminary version has been accepted for publication in *Journal of Hydrometeorology,* may be fully cited, and has been assigned DOI 10.1175/JHM-D-20-0213.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

This study characterizes precipitation error propagation through a distributed hydrological model based on the river basins across the Contiguous United States (CONUS), to better understand the relationship between errors in precipitation inputs and simulated discharge (i.e., P-Q error relationship). The NLDAS-2 precipitation and its simulated discharge are used as the reference to compare with TMPA-3B42 V7, TMPA-3B42RT V7, StageIV, CPC-U, MERRA-2, and MSWEP-2.2 for 1,548 well gauged river basins. The relative errors in multiple conventional precipitation products and their corresponding discharges are analysed for the period of 2002-2013. The results reveal positive linear P-Q error relationships at annual and monthly timescales, and the stronger linearity for larger temporal accumulations. Precipitation errors can be doubled in simulated annual accumulated discharge. Moreover, precipitation errors are strongly dampened in basins characterized by temperate and continental climate regimes, particularly for peak discharges, showing highly nonlinear relationships. Radar-based precipitation product consistently shows dampening effects on error propagation through discharge simulations at different accumulation timescales compared to the other precipitation products. Although basin size and topography also influence the P-Q error relationship and propagation of precipitation errors, their roles depend largely on precipitation products, seasons and climate regimes.

Keywords:

Precipitation; Error propagation; Discharge simulation; DRIVE model; Flood forecasting

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

1. Introduction

Difficulties in deriving accurate precipitation (P) estimation arise in many remote areas, particularly in complex terrain basins (Tao; Barros 2013; Tao et al. 2016). Consequently, satellite-based precipitation products have been increasingly developed to facilitate large-scale hydrological applications (Wu et al. 2012a; Wu et al. 2014). However, despite a wide range of hydrological studies applying satellite-based rainfall estimates for many years (Adler et al. 2000; Buarque et al. 2011; He et al. 2017; Huffman et al. 2001; Maggioni; Massari 2018; Meng et al. 2014; Nikolopoulos et al. 2013; Prakash et al. 2016; Stampoulis; Anagnostou 2012; Su et al. 2011; Wu et al. 2014; Yan et al. 2020; Zhong et al. 2019), the practical applications remain limited due to a number of error sources and uncertainties (Hossain; Anagnostou 2004; Sarachi et al. 2015). Reliable estimation of precipitation in space and time is highly desired for hydrological applications, as uncertainties in rainfall estimates can potentially lead to large errors in simulation outputs (Arnaud et al. 2002; Borga 2002; Courty et al. 2018; Morin et al. 2006; Nanding; Rico-Ramirez 2021; Rico-Ramirez et al. 2015; Smith et al. 2004; Tscheikner-Gratl et al. 2018; Wu et al. 2017; Younger et al. 2009; Zhang et al. 2018).

Many attempts have been conducted to quantify errors and uncertainties associated with satellite-based rainfall estimates to improve the understanding of precipitation physics (Behrangi et al. 2010; Hong et al. 2004; Liu; Fu 2010; Xu et al. 1999), retrieval algorithms (Bauer et al. 2001; Hossain et al. 2004; Moradkhani; Meskele 2010; Tong et al. 2014), measuring devices (e.g., infrared and microwave) (Hossain; Anagnostou 2004; Todd et al. 2001) and sampling frequencies (Iida et al. 2006; Nijssen; Lettenmaier 2004; Steiner et al. 2003). Since precipitation instruments have their own strengths and limitations in accuracy and spatial-temporal representativeness, the quality of satellite-based rainfall products were

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

evaluated for defining the best available precipitation product in specific regions (Bitew et al. 2012; Chen et al. 2020; Collischonn et al. 2008; Cui et al. 2019; Diem et al. 2014; Dinku et al. 2010; Dinku et al. 2008; Dos Reis et al. 2017; Feidas 2010; He et al. 2017; Le Coz; van de Giesen 2020; Meng et al. 2014; Monsieurs et al. 2018; Nicholson et al. 2019; Nikolopoulos et al. 2013; Stampoulis; Anagnostou 2012; Toté et al. 2015; Wu et al. 2017). These studies largely improve our understanding of the characteristics of errors and uncertainties in satellite rainfall estimates, which is crucial to improve future satellite rainfall products (Huffman et al. 2015) and hence their hydrological applications, such as global flood monitoring and forecasting (Wu et al. 2014).

Simulating the rainfall-runoff process using hydrological models can be useful for evaluating precipitation products at the river basin scale through comparison with an independent reference, e.g., discharge at river basin outlet (Beck et al. 2017b; Wu et al. 2017). Since rainfall is the most important driving component of the hydrologic cycle, the model performance depends heavily on the quality of precipitation inputs. However, due to the strong nonlinearity in hydrological processes, precipitation errors can be either amplified or dampened in simulated hydrological fluxes (mainly discharge) in different river basins (Artan et al. 2007; Bitew et al. 2012; Ehsan Bhuiyan et al. 2019; Gourley et al. 2011; Nikolopoulos et al. 2013). Therefore, propagation of precipitation errors through hydrological modelling has been identified as one of the critical issues in understanding the scale relationship between the errors in precipitation and in the corresponding hydrological simulations (Nikolopoulos et al. 2010).

The main characteristics of modeling experiments, datasets and key findings in recent literature on precipitation-discharge error relationships are summarized in Table 1. Typically, a linear relationship between precipitation error and hydrological simulation errors was

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

identified, and the errors in precipitation were translated into even larger errors in corresponding simulation outputs (Biemans et al. 2009; Decharme; Douville 2006; Kobold 2005; Maggioni et al. 2013; Nikolopoulos et al. 2010; Sharif et al. 2004; Wu et al. 2017). However, a few other studies demonstrated dampening effects of rainfall errors in discharge simulations (Falck et al. 2015; Mei et al. 2016). Moreover, propagation of precipitation errors through hydrological modelling varies with seasons (Biemans et al. 2009) and dampening or amplification of errors depends on rainfall products and basins sizes (Maggioni et al. 2013; Nijssen; Lettenmaier 2004; Nikolopoulos et al. 2010). Specifically, Nikolopoulos et al. (2010) demonstrated that the amplitude of dampening effect reduces with the increase of basin size. In contrast, Maggioni et al. (2013) reported a stronger dampening of precipitation errors for larger basins. However, there was no sign of dependency of error propagation on basin size in the study of Falck et al. (2015). Nikolopoulos et al. (2010) also reported that precipitation error propagation depends on the error metric, indicating that the propagation of rainfall errors into runoff volume result in completely different conclusions compared to the propagation from the same errors into runoff peaks.

Although various studies have investigated error propagation of precipitation inputs through hydrological modelling (Fekete et al. 2004; Nijssen; Lettenmaier 2004; Serpetzoglou et al. 2010; Vivoni et al. 2007), how precipitation error propagates through a hydrological model remains unclear and sometimes controversial. Therefore, a systematic investigation on how errors in precipitation (P) translate into errors in hydrological predictions, in particular on precipitation-discharge errors, is crucial for better interpretation and use of various P products to derive simulations for hydrological applications, such as water resource management, flow storage and flood control designs.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Specifically, building on the previous studies, such as Nikolopoulos et al. (2010) and Wu et al. (2017), this study aims to further examine the relationship between the errors in several widely used P products derived from different sources (i.e., gauge, ground-based radar and satellite) and their simulated discharge over an extensive number of sites across the CONUS. Moreover, this study assesses how precipitation error propagates (amplification or dampening) into hydrological simulations as a function of several factors, including the type of P product, temporal scale, climate, basin size and topography. It is worth mentioning that another motivation behind this study is that the P-Q error relations can be useful to estimate potential errors in flood predictions induced by precipitation errors when validation is not feasible. In particular, such relations can better inform the users of the Global Flood Monitoring System (GFMS) in their response to predicted flood events.

The rest of this paper is organized into the following sections. First, a description of the study basins and P datasets are presented in Section 2, including the criteria for final selection of study basins. Section 3 provides a detailed description of modeling framework based on a distributed hydrological model, methodology of basins classification and definition of error metrics. Section 4 analyses the characteristics of precipitation error propagation through discharge simulations and P-Q error relationship, and their potential influencing factors including P product, climate type, discharge magnitude, temporal accumulation scale, seasonality, and basins topography (size, elevation and slope). Finally, Section 4 provides a summary and concluding remarks based on the study results.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

2. Study domain and data

2.1. Study basins

The river basins are selected based on the availability of in-situ gauges from the United States Geological Survey (USGS) Geospatial Attributes of Gauges for Evaluating Streamflow (GAGES-II) database (Falcone 2011). The GAGES-II dataset consists of 9,322 streamflow gauges maintained by USGS over the CONUS with a large variety of geospatial and hydroclimate characteristics. A total of 1,548 candidate river basins are identified for this study based on the following criteria: (1) each gauge has sufficient historic streamflow observations (i.e., at least 10 years of records between 2002 and 2013) at daily scale, and (2) there is a good agreement (within ±10% difference) between the hydrological model calculated drainage area and USGS National Water Information System (NWIS) reported area, the latter of which is considered as the reference value (following Wu et al. (2014) and Alfieri et al. (2013)). The spatial distribution of the selected GAGES-II sites is shown in Figure 1b.

2.2. Precipitation products

Multiple spatially gridded precipitation products are available across the CONUS, including Version 7 of the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis for real-time (TMPA-3B42RT V7) and research (TMPA-3B42 V7) applications (Huffman et al. 2007), the NOAA/National Canters for Environmental Prediction (NCEP) Stage IV product (Lin; Mitchell 2005), the phase 2 of the North American Land Data Assimilation System (NLDAS-2) (Rui; Mocko 2013; Xia et al. 2012a), the CPC Unified (CPC-U) (Xie et al. 2007), the observation-corrected Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al. 2017; Reichle et al. 2016)

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

and the Multi-Source Weighted-Ensemble Precipitation, version 2.2 (MSWEP-2.2) (Beck et al. 2017a; Beck et al. 2019). These P products are mainly sourced from rain gauges (i.e., NLDAS-2 and CPC-U), ground-based radars (i.e., Stage IV), satellite-only (i.e., TMPA-3B42 V7 and TMPA-3B42RT V7) and reanalysis datasets (i.e., MERRA-2 and MSWEP-2.2). Precipitation products with higher temporal resolution (e.g., NLDAS-2, Stage IV, MERRA-2 and MSWEP-2.2) were simply aggregated to 3-hourly accumulations, while the CPC-U product at daily timescale was disaggregated based on the temporal distribution of 3-hourly TMPA-3B42 V7 product. A summary of the characteristics of these gridded P products, including their temporal and spatial resolutions, is provided in Table 2.

3. Methodology

3.1. Hydrological model

The Dominant river tracing-Routing Integrated with VIC Environment (DRIVE) model (Wu et al. 2014) is used to simulate river discharges in this study. The DRIVE model, as the core of the real-time Global Flood Monitoring System (GFMS), couples a widely used land surface model, the Variable Infiltration Capacity (VIC) model (Liang et al. 1996; Liang et al. 1994), with a hierarchical Dominant River Tracing-based Routing (DRTR) model (Wu et al. 2011; Wu et al. 2014). The VIC model solves full water and energy balances with good skill due to its characterization of both rainfall and snowmelt dominated runoff generation processes and soil frost dynamics (Christensen et al. 2004; Elsner et al. 2010; Hamlet et al. 2005). The DRTR is a gridded and physically based routing model, which allows the use of high-resolution sub-grid information aggregated for coarser resolution routing simulation and integrates the grid-level drainage network with numerical schemes based on the Strahler ordering system (Wu et al. 2014).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Since this study aims to provide insights on error propagation of precipitation for GFMS applications, the DRIVE model setup is adopted directly from the existing GFMS configuration that runs at spatial and temporal resolution of 1/8th degree and 3-hourly respectively. The VIC model runs in "Water Balance" mode. The hydrographic parameters (e.g., flow direction, drainage area, flow length, channel width, channel slope, overland slope, flow fraction, river order) for the DRTR runoff-routing scheme were derived by using the hierarchical Dominant River Tracing (DRT) river network upscaling algorithm (Wu et al. 2011; Wu et al. 2012b). The global soil and vegetation parameters were specified following Nijssen et al. (2001). Other atmospheric forcing data (i.e., air temperature and wind speed) were obtained from the NASA Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis (Rienecker et al. 2011). More detailed descriptions of the DRIVE model, forcing inputs, and model parameter setup are presented in previous studies (Huang et al. 2021; Wu et al. 2014; Yan et al. 2020). All forcing inputs are preprocessed at the same resolution in space and time that are consistent with model configuration.

To examine the error propagation of precipitation inputs through the hydrological model, the basic approach is to use the reference P and its simulated discharges to quantify errors in precipitation and their corresponding discharge simulations (Mei et al. 2016). Based on Xia et al. (2012b) and Wu et al. (2017), the NLDAS-2 and its DRIVE simulated discharges are defined as reference in this study. Note that the NLDAS-2 might not be among the best precipitation products over all the study basins. The absolute accuracy, however, does not impact the purpose of investigating the error propagation from precipitation through hydrological processes into discharge error, i.e., the relationship between errors in precipitation versus the errors in simulated discharge.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

The hydrological model setups (e.g., initial states, parameters, atmospheric forcings other than precipitation) are kept the same for different P products. The same initial condition for each P product based DRIVE simulation is provided by the continuous NLDAS-2 based simulations for the period 1997-1998 (with two years of spinning-up running before the start of the simulation of 1997). This removes the effects of differences in initial conditions among products and focusing only on rainfall-runoff process. Since the model simulation uncertainty results from the interactions between each source of uncertainty rather than individual ones, the decomposition of individual source of errors in model predictions remains challenging (Pianosi; Wagener 2016; Pianosi et al. 2016). Therefore, the design of this study is based on the assumption that the discharge errors are mainly due to precipitation errors and the interactions between different sources of errors and uncertainties are neglected. The DRIVE model runs between 1998-2013. Simulations during the period of 2002–2013 are analyzed, while the first four years are treated as the spin-up period. Note that the improvement of model performance per se is not within the scope of this study, and therefore further calibration of the DRIVE model is not conducted.

3.2. Basin classification

According to the Köppen climate classification (Beck et al. 2018; Köppen 1918; McKnight 2000), the 1,548 river basins over the CONUS are categorized into three main climate groups (continental, temperate and dry), as shown in Figure 1.c. A brief description of the nature of climate regimes, including precipitation, temperature, and hydrological characteristics of basins with dominant subclass are presented in Table 3. The description of the catchment hydrological characteristics are based on the assessment of regional patterns of seasonal water balance presented in Berghuijs et al. (2014).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Many studies have also identified catchment morphology, such as slope and flatness (Pena Arancibia et al. 2010; Uchida et al. 2005), as an influential factor on rainfall-runoff generation. In order to measure the combined impacts of elevation and slope on precipitation error propagation, the Multi Resolution index of Valley Bottom Flatness (MRVBF) (Gallant; Dowling 2003) is calculated based on the HydroSHEDS (Lehner et al. 2008) global hydrography dataset at 1-km resolution (as shown in Figure 1.a) using the Dominant River Tracing (DRT) algorithm, as in Wu et al. (2019). MRVBF identifies flat valley bottoms based on their topographic signature as flat low-lying areas at a range of scales. Flatness is measured by the inverse of slope, and lowness is measured by a ranking of elevation with respect to a circular surrounding area. The MRVBF indices are then grouped into seven general classes (Figure 1.d), where higher class numbers indicate cells with generally flatter and lower topography, consistent with Wu et al. (2019).

Hydroclimatic heterogeneity results in varying streamflow regimes across the CONUS (Berghuijs et al. 2014). In general, the total amount of annual precipitation decreases from east to west over the region (Figure 2). There is no significant seasonal difference in precipitation amount in the eastern CONUS, while the western and central CONUS show stronger seasonality in precipitation accumulations. Specifically, precipitation is higher (lower) during winter (summer) in the western part, while central CONUS shows the opposite. In higher-latitude and mountainous regions streamflow regimes are dominated by snowmelt, which means precipitation is accumulated as snow in winter and released as runoff in spring leading to large variations in seasonal discharge (Berghuijs et al. 2016). In humid regions soil moisture supply and atmospheric demand play an important role (Novick et al. 2016; Yuan et al. 2019).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

3.3. Error metrics

The following evaluation metrics are computed to quantify errors in basin-average precipitation and simulated discharges: Nash-Sutcliffe coefficient (Nash; Sutcliffe 1970) (NSE), relative absolute bias (r_ABIAS), and relative root mean square error (r_RMSE). r_ABIAS is used to measure the systematic error in both precipitation and discharge simulations, while the r_RMSE represents the random errors normalized with the references. They are defined as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{N}[R_i - M_i]^2}{\sum_{i=1}^{N}[R_i - \bar{R}_i]^2} \tag{1}$$

$$r\_ABIAS = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{M_i - R_i}{R_i}\right| \times 100\% \tag{2}$$

$$r\_RMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - R_i)^2}}{\bar{R}_i} \times 100\% \tag{3}$$

$$f = \frac{E_Q}{E_P} \tag{4}$$

where M_i and R_i represent the model simulated and reference discharges, respectively, at time step i; $\bar{R}_i$ is the mean reference discharges; N is the total number of time steps, and only M_i is considered for calculation when $R_i > 0$. The corresponding error metrics for precipitation can be obtained by replacing M_i and R_i with the target P products and NLDAS-2 respectively. As shown in Equation 4, a scale factor is defined to quantify how much error in precipitation translates into simulated discharge (Mei et al. 2016; Nikolopoulos et al. 2010), where E_Q and E_P represent the relative errors (e.g., r_ABIAS and r_RMSE) in simulated discharges and precipitation, respectively. A propagation factor greater (smaller) than one indicates the amplification (dampening) of errors in precipitation through discharge simulation; ratio of one indicates an equal translation of errors from precipitation to discharges. A linear regression model is applied to describe the relationship between the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

errors in precipitation estimates and theirs simulated discharges. The strength of linearity is identified if the coefficient of determination (R^2) is greater than 0.5 (with P-value < 0.05). For example, a given P-Q error relationship is considered significantly 'positive linear' only if the slope of linear regression is positive with P-value < 0.05. Otherwise, the error relationship is considered to be 'non-linear'. Note that the errors in both target P products and their estimated discharge are calculated based on pre-defined references and don't necessarily represent "true" errors.

4. Results and discussions

4.1. Evaluation of DRIVE-NLDAS-2 simulated discharge

The evaluation of simulated discharge driven by NLDAS-2 precipitation against USGS observed discharge is to gain a basic understanding of the efficiency of NLDAS-2 based simulations. Hydrological performances of the DRIVE-NLDAS-2 simulated discharge vary largely across the CONUS (Figure 2), reflecting the model efficiency in different regions. The result shows that the reference simulation generally represents a certain degree of observation fidelity in simulations with about 40% of gauge sites showing positive NSE scores with a mean (median) value of 0.47 (0.48) and maximum of 0.97 (USGS: 05474000). The spatial patterns of r_ABIAS and r_RMSE are similar, and the lower errors correspond to higher NSE values.

An overview of the distributions of different P products against the reference P (NLDAS-2) is shown in Figure 3. The gauge-involved P products (i.e., CPC-U, MERRA-2 and MSWEP) agree better with the reference P for most of the regions, but they show overestimations in annual precipitation accumulations over the West Coast of the CONUS. Whereas the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

satellite-only TMPA-3B42RT product clearly overestimate the reference rainfall, and the situation is improved effectively after gauge adjustment in TMPA-3B42. Radar-based Stage IV shows underestimations of reference rainfall over basins in the Western CONUS (Henn et al. 2018). This is because of difficulties in retrieving orographic precipitation due to beam blockage, and signal attenuation in mountainous regions (Bringi et al. 2011; Dai et al. 2015; Germann et al. 2006; Nanding; Rico-Ramirez 2021; Nanding et al. 2015; Rico-Ramirez 2012; Wang et al. 2015). The performances of these precipitation products vary between basins due to the strengths and limitations in their measuring devices. There is no single product outperforms the others in terms of both precipitation estimations and hydrological predictions for all study basins, including the reference product. In this study, the definition of reference for both precipitation and discharge is for comparing between the errors in precipitation and discharge directly in the model system.

The following results focus on (1) the scale relationship between errors in basin-averaged precipitation and those in simulated discharges (hereafter referred to as P-Q error relationship), (2) error propagation from precipitation to simulated discharges, and (3) dependency of those of characteristics on a variety of factors.

4.2. P-Q error relationship

A strong positive linear relationship between errors in precipitation and simulated discharges is observed regardless of the P products and accumulation timescales in terms of r_RMSE (Figure 4). The Stage IV product shows the strongest linearity in P-Q error relationship, followed by MSWEP, with the highest value of coefficient of determination. Remote sensing precipitation estimates (e.g., TMPA and Stage IV) tend to have larger errors in both precipitation and simulated discharges compared to gauge-based P products. Specifically,

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Stage IV mainly shows larger errors over basins in western regions of the CONUS (e.g., taking the 100th meridian west longitude as the approximate dividing line). This is due to the blockage in complex mountainous topography and the ground clutter contamination to radar signals. The lower skills were also reported in both precipitation estimation and discharge simulations (Maddox et al. 2002; Moreda et al. 2002; Nelson et al. 2016) reflecting the challenges in both data obtaining and modeling in complex terrain. However, although relative errors are much higher for basins on western edge, the P-Q error relationship of Stage IV shows strong ($r^2 > 0.5$) positive linear behaviors for both the western and eastern CONUS (Figure S1). Gauge-adjusted satellite P products present lower errors, suggesting that the gauge-correction procedure can effectively reduce errors in satellite rainfall estimates and lead to better hydrological simulations, which is consistent with previous studies (Su et al. 2011; Wu et al. 2014). As shown in Figure 4, the distributions of errors in precipitation and discharge tend to shift toward smaller errors at annual timescale (blue dots) compared to monthly timescales (grey dots), indicating that errors in precipitation are suppressed at larger timescale accumulation. Similar patterns in the P-Q error relationship are also observed for r_ABIAS (not shown). In general, linearity in the P-Q error relationship is stronger at annual timescale compared to monthly, in line with Wu et al. (2017). This is probably because of delays in the transformation of P to Q (due to channel routing and storage in snow, subsurface, lakes, etc.) are less important at annual than monthly time scales. Annual accumulations are more reliable to estimate hydrological water budget components (e.g., discharge) than monthly timescales (Berghuijs et al. 2014), assuming that changes in water storage (soil and surface) are negligible at annual accumulations for a closed watershed without significant streamflow diversions or impact of reservoirs (Adam et al. 2006).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

The relationship between the r_ABIAS in precipitation and NSE scores for discharge simulation tends to have negative linear behaviours with varying strength for different P products and accumulation timescales (Figure 5). This negative linear relationship indicates that precipitation products with less bias or errors tend to have better hydrological performances in terms of NSE scores. The linear relationship between r_ABIAS in precipitation and NSE scores for discharge simulation is weaker than aforementioned P-Q error relationship. In particular, the satellite-based products show a much weaker linear relationship ($r^2 < 0.25$) between errors in precipitation and NSE for discharge simulations though they have strong linearity ($r^2 > 0.5$) in the P-Q error relationship. This can be explained by equations (1) and (2) where NSE is more sensitive to larger errors, while the errors in precipitation tend to be translated into even larger errors in simulated discharge (Figure 4).

4.3. Error propagation

The relative errors in P products are generally translated into even larger relative errors in their simulated discharges (Figure 6). The propagation factors are higher for annual accumulations than for monthly accumulations in terms of their mean values. This could be due to the fact that the spatially distributed precipitation errors are filtered out by averaging them to mean areal precipitation, while the error in the discharge at the outlet of a river basin is resulted from a cascade of numerical solutions for equations of various no-linear processes with spatially distributed precipitation inputs containing the original errors. Although the numerical solutions deployed by the hydrological model derive reasonably stable discharge simulations, there is inherent error propagation from precipitation input to discharge output which depends much more on the error pass and propagation among the numerical solution schemes than on spatial scales and seasonal changes. Therefore, the canceling effects for the

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

precipitation error as a function of basin area and averaging time (Nijssen; Lettenmaier 2004) doesn’t apply to the discharge error. This explains that the error propagation ratio between the discharge errors and precipitation errors tends to be greater than one, in particular at annual scale for a river basin with a stable hydrological water budget, while it varies significantly at monthly and shorter time scales.

However, the amplification effects of precipitation-to-discharge errors vary across P products. Stage IV consistently shows lower values in propagation factors at different accumulation timescales. This could be due to the better performance of the higher resolution radar-based product in moderating the precipitation errors in hydrological processes, as presented in Nikolopoulos et al. (2010). The errors in discharges simulated by satellite-based P products are more than double the errors in precipitation. Similarly, the errors in gauge-based P products are at least doubled in annual discharge, although they show relatively low errors in precipitation estimations (Figure 4). Similar patterns are observed for the propagation of both r_RMSE and r_ABIAS.

Spatial patterns of error propagation factors also vary among P products, and accumulation timescales (Figures 7-8). There are clear differences in error propagation patterns between P products over the study basins. Specifically, at annual timescale, the errors in TMPA-3B42 V7 simulated discharges are more than triple the errors in precipitation for about 33% (the highest) of basins, while a number of basins with the lowest proportion of 6% is obtained by the Stage IV (Figure 7). The spatially distributed hydrological model used in this study takes into account the spatial variability of precipitation and calculates flow contributions from elementary grid areas. Therefore, the spatial variability in precipitation amounts and intensities between different P products in each river basin could be responsible for the variability in the distribution of propagation factors among different P products, which

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

explains the different error propagation rates even if P products have similar basin-averaged precipitation accumulations.

Compared to the annual scale, monthly accumulations show less amplification and more dampening effects of errors in most basins (Figure 7-8). For example, the number of basins with dampening effect is 51 for Stage IV at annual accumulations, while a number is 111 basins for monthly accumulations. This decrease in the propagation factor at monthly timescales might be related to the fact that soil infiltration removes some of the short-term rainfall fluctuations by generating subsurface runoff, which contributes to the river channel discharge through a slower routing process. The filtering effects can be more visible in precipitation with higher errors (e.g., satellite-based P) particularly at monthly timescales and when subsurface soil layers are not saturated. This also partially explains the stronger linearity in annual P-Q error relationship compared to the linear relationship at monthly timescale (Figures 4).

The sensitivity of precipitation error propagation patterns on reference data is also tested by using different P products and their simulated discharges as references, including satellite-based (TMPA-3B42RT V7 and TMPA-3B42 V7), radar-based (StageIV) and reanalysis (MSWEP-2.2) products. Results based on TMPA-3B42 V7 are shown in Figure S7-S8. According to Figure S7, there are strong linear relationship between errors in precipitation and simulated discharges, regardless of P products and accumulation timescales. This is similar to the findings when NLDAS-2 is used as the reference. Moreover, Figure S8 also shows that precipitation errors are less amplified at monthly timescale compared to annual timescale, which is also observed when using MSWEP-2.2 and StageIV as the references. Moreover, similar analysis has also been done by only using 584 basins for which the NLDAS-2 DRIVE simulations show positive NSE scores, which leads to similar results that

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

are obtained by using all basins (figures omitted). Therefore, the findings on changing patterns of precipitation error propagation are convincingly hold.

4.4. Impact of hydroclimatic factors

With regard to the impacts of climate regime on P-Q error relationship, the differences in the strength of linearity are much smaller among PQE products and seasons in basins situated in temperate and continental climate regimes compared to those with dry climates (Figure 9). Temperate and continental climate regimes have less seasonal streamflow variations, while large variations in storage and streamflow regimes over basins with dry climates are observed (Table 3). In dry climates, basins are mostly semi-arid and have distinct seasonality in precipitation. For instance, basins in mountainous region in western CONUS receive most of precipitation during winter, while streamflow in these basins is mainly generated from snow-melting (Berghuijs et al. 2014), suggesting a non-linear behavior in P-Q error relationship during the warm season, particularly during April-June (AMJ) (Figure 9).

A stronger linearity in the P-Q error relationship is also observed when considering all discharge magnitudes rather than considering only peak discharges, i.e., $Q \geq 99^{th}$ percentile (Figure 9). Stronger dampening effects are further observed when considering peak discharges compared to all discharge magnitudes at monthly timescale (Figure 10), regardless of climate regimes and PQE products. The error in precipitation is calculated based on river basin average, while the translation of such precipitation error into discharge is based on a chain of rainfall-runoff generation and routing processes. That said, the precipitation error is temporally distributed along the whole hydrograph and the error in flood peak is associated to only part of the precipitation error. Stronger attenuation effects are observed for discharge peaks in basins with temperate and continental climates compared to dry climate basins

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

(Figure 10). In addition to precipitation, antecedent wetness conditions and water storage also play an important role in runoff generation in temperate and continental climates (Berghuijs et al. 2016). Although all P products tend to have more challenges in deriving accurate precipitation during extreme events, the error contributions to flood peak are from both the flood-causing precipitation and the antecedent wetness of the river basin. Therefore, the error dampening in the propagation can be caused by the reasonable model performance in continuous hydrological simulation of the wetness dynamics and filtering effects of the complex non-linear rainfall-runoff processes.

In terms of seasonality, there is no clear difference between the mean propagation factors for peak discharges during seasons of AMJ and OND in continental and temperate basins (Figure 11). For arid/semi-arid basins, however, the precipitation errors are clearly translated into even larger errors in discharges due to the high variability in precipitation during AMJ, whereas precipitation errors tend to be translated equally into discharge errors during OND. This is because the large water deficit during the dry period filters out the discharge errors and somehow similar to precipitation errors.

4.5. Impact of basin size and topography

The basin size also plays an important role in the P-Q error relationship and precipitation error propagation. Valley bottom regions (higher MRVBF class) are relatively larger in the plains of the eastern CONUS with less seasonal variations in precipitation and streamflow, except for a few basins in the north-eastern and south-eastern CONUS with strong seasonality. Basins at high elevations (lower MRVBF class) are mainly clustered in arid/semi-arid regions with strong seasonality in streamflow. Thus, the role of topography in

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

the P-Q error relationship and error propagation is studied together with the aforementioned hydroclimate classes (Figure S2-S4).

The impact of basin size on the strength of P-Q error relationship varies between P products, seasons, and basin classes (Figure 12 and S2). For example, with the Stage IV, large basins show strong positive linear P-Q error relationship for dry climate regime compared to small basins, while the basin size show no clear impact on the strength of P-Q error relationship for basins with temperate and continental climate regimes. With MSWEP-2.2, large basins show strong linearity in P-Q error relationship for temperate and continental climate regimes particularly for warm seasons, while small basins show relatively stronger P-Q error relationship than large basins for dry climate regime. Moreover, for temperate climate regimes (or MRVBF classes ≥ 5), small basins also show less variations in the strength of P-Q error relationship between seasons and P products compared to large basins. These results indicate that the impacts of basin size on the P-Q error relationship also depends on P product, season and climate regime.

Figure 13 and Figure 14 demonstrate the propagation factors of r_RMSE for basins with different size in each MRVBF class during seasons of OND and AMJ respectively when considering peak discharges. In terms of mean values, similar propagation factors are observed between small and large basins of each MRVBF class for each P product during the OND season (Figure 13). For example, precipitation errors are translated into similar magnitude of discharge errors for both small and large basins with lower MRVBF class (steeper and higher elevated basins). With the increase of MRVBF class (lower and flatter basins), the mean of propagation factors increases for both small and large basins, particularly for reanalysis P products (e.g., MERRA-2 and MSWEP-2.2). In contrast, during the AMJ season, larger basins show amplification effects on precipitation errors propagation

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

regardless of P products, MRVBF class and climate regimes, except the Stage IV product (Figure 14 and S4). Moreover, large basins also show greater variations in error propagation between P products and seasons which in line with the variations in P-Q error relationship. Meanwhile, the larger uncertainty in propagation factors with wider distributions (shape of violin plot) for each P is observed for large basins compared to small basins, which indicates that factors like basin size other than P products, seasons, and climate regimes also control the magnitude of error propagation for basins with different climate regimes.

5. Conclusions

This study investigates the characteristics of precipitation error propagation and the P-Q error relationship by applying a distributed hydrological model to 1,548 river basins across the CONUS. The NLDAS-2 precipitation and its DRIVE simulated discharge are used as the reference to quantify the relative errors (e.g., $RMSE$ and $ABIAS$) in several P products and their corresponding discharge simulations for the period of 2002-2013. The analysis focus on the dependency of precipitation error propagation and the P-Q error relationship on a variety of factors including P product, temporal scale, climate regime, and basin topography.

The results show the positive linear behaviours in P-Q error relationship at annual and monthly accumulations, and the linearity is stronger at larger accumulation timescale, suggesting that it is more reliable to estimate potential errors in hydrological simulation outputs due to precipitation errors at larger time scales. Precipitation errors are at least doubled in simulated discharge for annual accumulations, while dampening effects are more common in peak discharges. Moreover, the patterns of P-Q error relationship and error propagation are seasonal, and sensitive to the climate regimes of basins particularly for larger basins. The differences of linearity in P-Q error relationship are much smaller between

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

seasons and PQE products for the basins in temperate and continental climate regimes compare to the basins with dry climates. Regarding to the role of basin topography, the role of basin size on precipitation error propagation and P-Q error relationship largely depends on climate regimes and seasons. Generally, during the OND season precipitation errors are translated into similar magnitude of discharge errors for both small and large basins in each climate regime and MRVBF class in terms of mean propagation factor, while during the AMJ season larger basins show amplification effects on precipitation errors propagation regardless of P products, seasons and climate regimes. Large basins also show greater variations in error propagation and P-Q error relationship between P products and seasons.

This study quantifies the P-Q error relationship at annual and monthly scales based on existing P products, and investigates the precipitation error propagation in discharge simulations and its links to hydroclimate, topographic characteristics of river basins. It helps in understanding the quality (bias and uncertainty) of flood simulation outputs and their relation to precipitation inputs, thus also shed light on potential ways to improve precipitation estimation products. However, floods are also highly linked to daily and finer time scales, and such P-Q error relationships at finer time scales warrant further investigation with more complicated processes considered in future studies. It is worth noting that the current findings are only based on a given hydrological model, and therefore the results might change with different models with different numerical schemes for runoff-routing processes. However, this study advances the understanding of the P-Q error relationship and its propagation in hydrological models that are similar to the DRIVE model and the possible uncertainty that is associated to the global flood monitoring and forecasting systems such as the GFMS.

Furthermore, despite the findings revealed in this study fairly hold, the shortcoming is that the climate regimes are broadly classified into three main groups based on Köppen

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

classification, but the diversity of climate and hydrological regimes remains considerable within each class resulting in varying patterns of P-Q error relationship and error propagation within each class. Therefore, the changing patterns of error propagation from precipitation to hydrological simulation outputs require further detailed study of sub-clusters or regions based on both climate and catchment characteristics including seasonal water balances.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (Grants: 41905101, 41861144014, 41775106, U1811464, and 41975113), National Key R&D Program of China (Grant: 2017YFA0604300), and partially by the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (Grant: 2017ZT07X355). The authors would like to thank the anonymous reviewers for their constructive comments and suggestions.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

References

Adam, J. C., E. A. Clark, D. P. Lettenmaier, and E. F. Wood, 2006: Correction of global precipitation products for orographic effects. *Journal of Climate*, **19,** 15-38.

Adler, R. F., G. J. Huffman, D. T. Bolvin, S. Curtis, and E. J. Nelkin, 2000: Tropical rainfall distributions determined using TRMM combined with other satellite and rain gauge information. *Journal of Applied Meteorology*, **39,** 2007-2023.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, 2013: GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, **17,** 1161.

Arnaud, P., C. Bouvier, L. Cisneros, and R. Dominguez, 2002: Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology,* **260,** 216-230.

Artan, G., H. Gadain, J. L. Smith, K. Asante, C. J. Bandaragoda, and J. P. Verdin, 2007: Adequacy of satellite derived rainfall data for stream flow modeling. *Natural Hazards,* **43,** 167-185.

Bauer, P., P. Amayenc, C. D. Kummerow, and E. A. Smith, 2001: Over-ocean rainfall retrieval from multisensor data of the Tropical Rainfall Measuring Mission. Part II: Algorithm implementation. *Journal of Atmospheric and Oceanic Technology*, **18,** 1838-1855.

Beck, H. E., N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, 2018: Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data,* **5,** 180214.

Beck, H. E., A. I. J. M. Van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, and A. De Roo, 2017a: MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology & Earth System Sciences,* **21,** 1-38.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Beck, H. E., and Coauthors, 2019: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, **100,** 473-500.

Beck, H. E., and Coauthors, 2017b: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, **21,** 6201.

Behrangi, A., K. Hsu, B. Imam, and S. Sorooshian, 2010: Daytime precipitation estimation using bispectral cloud classification system. *Journal of Applied Meteorology and Climatology*, **49,** 1015-1031.

Berghuijs, W. R., M. Sivapalan, R. A. Woods, and H. H. Savenije, 2014: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, **50,** 5638-5661.

Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan, 2016: Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, **43,** 4382-4390.

Biemans, H., R. Hutjes, P. Kabat, B. Strengers, D. Gerten, and S. Rost, 2009: Effects of precipitation uncertainty on discharge calculations for main river basins. *Journal of Hydrometeorology*, **10,** 1011-1025.

Bitew, M. M., M. Gebremichael, L. T. Ghebremichael, and Y. A. Bayissa, 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,** 338-350.

Borga, M., 2002: Accuracy of radar rainfall estimates for streamflow simulation. *Journal of Hydrology*, **267,** 26-39.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Bringi, V., M. Rico-Ramirez, and M. Thurai, 2011: Rainfall Estimation with an Operational Polarimetric C-Band Radar in the United Kingdom: Comparison with a Gauge Network and Error Analysis. *Journal of Hydrometeorology*, **12,** 935-954.

Buarque, D. C., R. C. D. de Paiva, R. T. Clarke, and C. A. B. Mendes, 2011: A comparison of Amazon rainfall characteristics derived from TRMM, CMORPH and the Brazilian national rain gauge network. *Journal of Geophysical Research: Atmospheres*, **116**.

Chen, C., Z. Li, Y. Song, Z. Duan, K. Mo, Z. Wang, and Q. Chen, 2020: Performance of multiple satellite precipitation estimates over a typical arid mountainous area of China: Spatiotemporal patterns and extremes. *Journal of Hydrometeorology*, **21,** 533-550.

Christensen, N. S., A. W. Wood, N. Voisin, D. P. Lettenmaier, and R. N. Palmer, 2004: The effects of climate change on the hydrology and water resources of the Colorado River basin. *Clim. Change*, **62,** 337-363.

Collischonn, B., W. Collischonn, and C. E. M. Tucci, 2008: Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, **360,** 207-216.

Courty, L. G., M. Á. Rico-Ramirez, and A. Pedrozo-Acuña, 2018: The significance of the spatial variability of rainfall on the numerical simulation of urban floods. *Water*, **10,** 207.

Cui, W., X. Dong, B. Xi, Z. Feng, and J. Fan, 2019: Can the GPM IMERG Final product accurately represent MCSs' precipitation characteristics over the Central and Eastern United States? *Journal of Hydrometeorology*.

Dai, Q., D. Han, M. A. Rico-Ramirez, L. Zhuo, N. Nanding, and T. Islam, 2015: Radar rainfall uncertainty modelling influenced by wind. *Hydrological Processes*, **29,** 1704-1716.

Decharme, B., and H. Douville, 2006: Uncertainties in the GSWP-2 precipitation forcing and their impacts on regional and global hydrological simulations. *Climate Dynamics*, **27,** 695-713.

Diem, J. E., J. Hartter, S. J. Ryan, and M. W. Palace, 2014: Validation of satellite rainfall products for western Uganda. *Journal of Hydrometeorology*, **15,** 2030-2038.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Dinku, T., F. Ruiz, S. J. Connor, and P. Ceccato, 2010: Validation and intercomparison of satellite rainfall estimates over Colombia. *Journal of Applied Meteorology and Climatology*, **49,** 1004-1014.

Dinku, T., S. Chidzambwa, P. Ceccato, S. Connor, and C. Ropelewski, 2008: Validation of high‐resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing*, **29,** 4097-4110.

Dos Reis, J. B. C., C. D. Rennó, and E. S. S. Lopes, 2017: Validation of satellite rainfall products over a mountainous watershed in a humid subtropical climate region of Brazil. *Remote Sensing*, **9,** 1240.

Ehsan Bhuiyan, M. A., and Coauthors, 2019: Assessment of precipitation error propagation in multi-model global water resource reanalysis. *Hydrology and Earth System Sciences*, **23,** 1973-1994.

Elsner, M. M., and Coauthors, 2010: Implications of 21st century climate change for the hydrology of Washington State. *Clim. Change,* **102,** 225-260.

Falck, A. S., V. Maggioni, J. Tomasella, D. A. Vila, and F. L. Diniz, 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.

Falcone, J. A., 2011: GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow. *Center for Integrated Data Analytics Wisconsin Science Center*.

Feidas, H., 2010: Validation of satellite rainfall products over Greece. *Theoretical and Applied climatology*, **99,** 193-216.

Fekete, B. M., C. J. Vörösmarty, J. O. Roads, and C. J. Willmott, 2004: Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, **17,** 294-304.

Gallant, J. C., and T. I. Dowling, 2003: A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, **39**.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, **30**.

Germann, U., G. Galli, M. Boscacci, and M. Bolliger, 2006: Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society*, **132,** 1669-1692.

Gourley, J. J., Y. Hong, Z. L. Flamig, J. Wang, H. Vergara, and E. N. Anagnostou, 2011: Hydrologic evaluation of rainfall estimates from radar, satellite, gauge, and combinations on Ft. Cobb basin, Oklahoma. *Journal of Hydrometeorology*, **12,** 973-988.

Hamlet, A. F., P. W. Mote, M. P. Clark, and D. P. Lettenmaier, 2005: Effects of temperature and precipitation variability on snowpack trends in the western United States. *Journal of Climate*, **18,** 4545-4561.

He, Z., L. Yang, F. Tian, G. Ni, A. Hou, and H. Lu, 2017: Intercomparisons of rainfall estimates from TRMM and GPM multisatellite products over the Upper Mekong River Basin. *Journal of Hydrometeorology,* **18,** 413-430.

Henn, B., A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, 2018: An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of hydrology*, **556,** 1205-1219.

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

Hossain, F., and E. N. Anagnostou, 2004: Assessment of current passive‐microwave‐ and infrared‐based satellite rainfall remote sensing for flood prediction. *Journal of Geophysical Research: Atmospheres*, **109**.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Hossain, F., E. N. Anagnostou, and T. Dinku, 2004: Sensitivity analyses of satellite rainfall retrieval and sampling error on flood prediction uncertainty. *IEEE transactions on geoscience and remote sensing*, **42,** 130-139.

Huang, Z., H. Wu, R. F. Adler, G. Schumann, J. J. Gourley, A. Kettner, and N. Nanding, 2021: Multi-Sourced Flood Inventories over the Contiguous United States for Actual and Natural Conditions. *Bulletin of the American Meteorological Society***,** 1-47.

Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, P. Xie, and S.-H. Yoo, 2015: NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG). *Algorithm Theoretical Basis Document (ATBD) Version*, **4,** 26.

Huffman, G. J., and Coauthors, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *Journal of Hydrometeorology*, **2,** 36-50.

Huffman, G. J., and Coauthors, 2007: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, **8,** 38-55.

Iida, Y., K. i. Okamoto, T. Ushio, and R. Oki, 2006: Simulation of sampling error of average rainfall rates in space and time by five satellites using radar‐AMeDAS composites. *Geophysical research letters*, **33**.

Kobold, M., 2005: Precipitation forecasts and their uncertainty as input into hydrological models. *Hydrology and Earth System Sciences Discussions*, **9,** 322-332.

Köppen, W., 1918: Klassification der klimate nach temperatur, niederschlag und jahreslauf. Petermanns Geographische Mitteilungen. *Gotha*, **64**.

Le Coz, C., and N. van de Giesen, 2020: Comparison of Rainfall Products over Sub-Saharan Africa. *Journal of Hydrometeorology*, **21,** 553-596.

Lehner, B., K. Verdin, and A. Jarvis, 2008: New Global Hydrography Derived From Spaceborne Elevation Data. *Eos Transactions American Geophysical Union*, **89**.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Liang, X., E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change,* **13,** 195-206.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres,* **99,** 14415-14428.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. Hydrology*, Citeseer.

Liu, Q., and Y. Fu, 2010: Comparison of radiative signals between precipitating and non-precipitating clouds in frontal and typhoon domains over East Asia. *Atmospheric Research*, **96,** 436-446.

Maddox, R. A., J. Zhang, J. J. Gourley, and K. W. Howard, 2002: Weather radar coverage over the contiguous United States. *Weather and forecasting*, **17,** 927-934.

Maggioni, V., and C. Massari, 2018: On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology*, **558,** 214-224.

Maggioni, V., H. J. Vergara, E. N. Anagnostou, J. J. Gourley, Y. Hong, and D. Stampoulis, 2013: Investigating the applicability of error correction ensembles of satellite rainfall products in river flow simulations. *Journal of Hydrometeorology*, **14,** 1194-1211.

McKnight, T. L., 2000: Climate zones and types. *Physical geography: a landscape appreciation*.

Mei, Y., E. I. Nikolopoulos, E. N. Anagnostou, and M. Borga, 2016: Evaluating Satellite Precipitation Error Propagation in Runoff Simulations of Mountainous Basins. *Journal of Hydrometeorology*, **17,** 1407-1423.

Meng, J., L. Li, Z. Hao, J. Wang, and Q. Shao, 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.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Monsieurs, E., and Coauthors, 2018: Evaluating TMPA rainfall over the sparsely gauged East African Rift. *Journal of Hydrometeorology*, **19,** 1507-1528.

Moradkhani, H., and T. T. Meskele, 2010: Probabilistic assessment of the satellite rainfall retrieval error translation to hydrologic response. *Satellite Rainfall Applications for Surface Hydrology*, Springer, 229-242.

Moreda, F., S. Cong, D. Seo, Z. Zhang, S. Reed, V. Koren, and M. Smith, 2002: Evaluation of NEXRAD Rainfall Data in a Lumped and Distributed Hydrologic Models for a Mountainous Watershed. *AGUSM*, **2002,** H51B-05.

Morin, E., D. C. Goodrich, R. A. Maddox, X. Gao, H. V. Gupta, and S. Sorooshian, 2006: Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response. *Advances in Water Resources*, **29,** 843-860.

Nanding, N., and M. A. Rico-Ramirez, 2021: Precipitation Measurement with Weather Radars. *ICT for Smart Water Systems: Measurements and Data Science,* A. Scozzari, S. Mounce, D. Han, F. Soldovieri, and D. Solomatine, Eds., Springer International Publishing, 235-258.

Nanding, N., M. A. Rico-Ramirez, and D. Han, 2015: Comparison of different radar-raingauge rainfall merging techniques. *Journal of Hydroinformatics*, **17,** 422-445.

Nash, J., and J. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, **10,** 282-290.

Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Weather and Forecasting*, **31,** 371-394.

Nicholson, S., D. Klotter, L. Zhou, and W. Hua, 2019: Validation of satellite precipitation estimates over the Congo Basin. *Journal of Hydrometeorology*, **20,** 631-656.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

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

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *Journal of Climate*, **14,** 3307-3323.

Nikolopoulos, E. I., E. N. Anagnostou, and M. Borga, 2013: Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. *Journal of Hydrometeorology*, **14,** 171-185.

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

Novick, K. A., and Coauthors, 2016: The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nature Climate Change*, **6**.

Pena Arancibia, J., A. Van Dijk, M. Mulligan, and L. A. Bruijnzeel, 2010: The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrology & Earth System Sciences*, **14,** 2193-2205.

Pianosi, F., and T. Wagener, 2016: Understanding the time‐varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener, 2016: Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, **79,** 214-232.

Prakash, S., A. K. Mitra, D. S. Pai, and A. AghaKouchak, 2016: From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, **88,** 1-7.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. P. Mahanama, and G. S. Partyka, 2016: Land surface precipitation in MERRA-2. *Journal of Climate*, **30,** JCLI-D-16-0570.0571.

Rico-Ramirez, M., S. Liguori, and A. Schellart, 2015: Quantifying radar-rainfall uncertainties in urban drainage flow modelling. *Journal of hydrology*, **528,** 17-28.

Rico-Ramirez, M. A., 2012: Adaptive Attenuation Correction Techniques for C-Band Polarimetric Weather Radars. *IEEE transactions on geoscience and remote sensing*, **50,** 5061-5071.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate*, **24,** 3624-3648.

Rui, H., and D. Mocko, 2013: Readme Document for North America Land Data Assimilation System Phase 2 (NLDAS-2) Products. *The National Aeronautics and Space Administration: Greenbelt, MD, USA.*

Sarachi, S., K.-l. Hsu, and S. Sorooshian, 2015: A statistical model for the uncertainty analysis of satellite precipitation products. *Journal of Hydrometeorology*, **16,** 2101-2117.

Serpetzoglou, E., E. N. Anagnostou, A. Papadopoulos, E. I. Nikolopoulos, and V. Maggioni, 2010: Error Propagation of Remote Sensing Rainfall Estimates in Soil Moisture Prediction from a Land Surface Model. *Journal of Hydrometeorology,* **11,** 705-720.

Sharif, H. O., F. L. Ogden, W. F. Krajewski, and M. Xue, 2004: Statistical Analysis of Radar Rainfall Error Propagation. *Journal of Hydrometeorology*, **5,** 199-212.

Smith, M. B., V. I. Koren, Z. Zhang, S. M. Reed, J. J. Pan, and F. Moreda, 2004: Runoff response to spatial variability in precipitation: an analysis of observed data. *Journal of Hydrology*, **298,** 267-286.

Stampoulis, D., and E. N. Anagnostou, 2012: Evaluation of global satellite rainfall products over continental Europe. *Journal of Hydrometeorology*, **13,** 588-603.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

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

Su, F., H. Gao, G. J. Huffman, and D. P. Lettenmaier, 2011: Potential utility of the real-time TMPA-RT precipitation estimates in streamflow prediction. *Journal of Hydrometeorology*, **12,** 444-455.

Tao, J., and A. P. Barros, 2013: Prospects for flash flood forecasting in mountainous regions - An investigation of Tropical Storm Fay in the Southern Appalachians. *Journal of Hydrology*, **506,** 69-89.

Tao, J., D. Wu, J. Gourley, S. Q. Zhang, W. Crow, C. Peters-Lidard, and A. P. Barros, 2016: Operational hydrological forecasting during the IPHEx-IOP campaign–Meet the challenge. *Journal of hydrology*, **541,** 434-456.

Todd, M. C., C. Kidd, D. Kniveton, and T. J. Bellerby, 2001: A combined satellite infrared and passive microwave technique for estimation of small-scale rainfall. *Journal of Atmospheric and Oceanic Technology*, **18,** 742-755.

Tong, K., F. Su, D. Yang, and Z. Hao, 2014: Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau. *Journal of Hydrology*, **519,** 423-437.

Toté, C., D. Patricio, H. Boogaard, R. Van der Wijngaart, E. Tarnavsky, and C. Funk, 2015: Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, **7,** 1758-1776.

Tscheikner-Gratl, F., and Coauthors, 2018: Recent insights on uncertainties present in integrated catchment water quality modelling. *Water research*.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Uchida, T., I. T. Meerveld, and J. J. McDonnell, 2005: The role of lateral pipe flow in hillslope runoff response: an intercomparison of non-linear hillslope response. *Journal of Hydrology*, **311,** 117-133.

Vivoni, E. R., D. Entekhabi, and R. N. Hoffman, 2007: Error propagation of radar rainfall nowcasting fields through a fully distributed flood forecasting model. *Journal of applied meteorology and climatology*, **46,** 932-940.

Wang, L., S. Ochoa-Rodriguez, C. Onof, and P. Willems, 2015: Singularity-sensitive gauge-based radar rainfall adjustment methods for urban hydrological applications. *Hydrology and Earth System Sciences*, **19,** 4001-4021.

Wu, H., J. S. Kimball, N. Mantua, and J. Stanford, 2011: Automated upscaling of river networks for macroscale hydrological modeling. *Water Resources Research*, **47**.

Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, 2012a: Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *Journal of Hydrometeorology,* **13,** 1268-1284.

Wu, H., R. F. Adler, Y. Tian, G. Gu, and G. J. Huffman, 2017: Evaluation of quantitative precipitation estimations through hydrological modeling in IFloodS river basins. *Journal of Hydrometeorology*, **18,** 529-553.

Wu, H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, and R. F. Adler, 2012b: A new global river network database for macroscale hydrologic modeling. *Water resources research*, **48**.

Wu, H., R. F. Adler, Y. Tian, G. J. Huffman, H. Li, and J. Wang, 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research,* **50,** 2693-2717.

Wu, H., J. S. Kimball, N. Zhou, L. Alfieri, L. Luo, J. Du, and Z. Huang, 2019: Evaluation of real-time global flood modeling with satellite surface inundation observations from SMAP. *Remote Sensing of Environment*, **233,** 111360.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Xia, Y., and Coauthors, 2012a: Continental - scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS - 2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, **117**.

Xia, Y., and Coauthors, 2012b: Continental - scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS - 2): 2. Validation of model - simulated streamflow. *Journal of Geophysical Research: Atmospheres*, **117**.

Xie, P., M. Chen, S. Yang, A. Yatagai, T. Hayasaka, Y. Fukushima, and C. Liu, 2007: A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology*, **8,** 607-626.

Xu, L., S. Sorooshian, X. Gao, and H. V. Gupta, 1999: A cloud-patch technique for identification and removal of no-rain clouds from satellite infrared imagery. *Journal of Applied Meteorology*, **38,** 1170-1181.

Yan, Y., and Coauthors, 2020: Climatology and Interannual Variability of Floods during the TRMM Era (1998–2013). *Journal of Climate*, **33,** 3289-3305.

Younger, P. M., J. E. Freer, and K. J. Beven, 2009: Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. *Hydrological Processes*, **23,** 1988-2003.

Yuan, W., and Coauthors, 2019: Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Science Advances*, **5**.

Zhang, A., H. Shi, T. Li, and X. Fu, 2018: Analysis of the influence of rainfall spatial uncertainty on hydrological simulations using the Bootstrap method. *Atmosphere*, **9,** 71.

Zhong, R., X. Chen, C. Lai, Z. Wang, Y. Lian, H. Yu, and X. Wu, 2019: Drought monitoring utility of satellite-based precipitation products across mainland China. *Journal of Hydrology*, **568,** 343-359.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

List of Tables

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Table 1. A summary of relevant previous studies in terms of their study locations, adopted methods and key findings (the current paper is included for comparison).

Study	Site location	Methods & Datasets	Key findings
Sharif et al. (2004)	United States	a) Single watershed (21.2 sq.km) b) Radar rainfall estimates c) Ensemble of 500 events d) CASC2D model	a) Errors in rainfall volume are amplified in runoff predictions b) Positive linear relationship between the rainfall volume, runoff volume and peak discharge errors
Kobold (2005)	Slovenia	a) 2 river basins (87 and 1,848 sq.km) b) 2 rainfall estimates c) 3 storm events d) HEC-1 model	a) Deviation of runoff is much greater than that of the rainfall b) Errors in rainfall lead to 1.6 times greater error in peak discharge
Decharme; Douville (2006)	Europe	a) Single river basin (98,000 sq.km) b) SAFRAN rainfall estimates c) Annual and monthly scale d) ISBA land surface model	a) Quasi-linear relationship between relative errors in annual rainfall and annual discharges b) The errors in rainfall is translated at least the same or greater errors in total runoff
Biemans et al. (2009)	Worldwide	a) 294 major river basins b) 7 global rainfall products c) Annual and seasonal scale d) LPJmL model	a) Relative uncertainty in precipitation is amplified in discharge b) Propagation of uncertainty in precipitation show seasonality
Nikolopoulos et al. (2010)	Italy	a) Complex terrain basins (100 - 1200 sq.km) b) 4 rainfall products c) Single flood event d) tRIBS model	a) Propagation of rainfall errors shows linear behaviour b) Dampening or amplification of errors depends on error metric, rainfall products and basin size (i.e., strong dampening effects for smaller basins)
Maggioni et al. (2013)	United States	a) 5 sub-basins b) 3 satellite rainfall products c) SREM2D rainfall error model d) HL-RDHM model	a) Relative bias doubles from rainfall to runoff b) CMORPH shows more ensemble variability in bias propagation c) Strong dampening of RMSE for larger basins
Falck et al. (2015)	Brazil	a) 19 sub-basins (5,230 – 764,000 sq.km) b) 4 satellite rainfall products c) MHD-INP grid-based model d) SREM2D rainfall error model	a) Errors in rainfall are mostly dampened in simulated streamflow b) Propagation of errors in rainfall ensembles shows no dependency on basin size
Mei et al. (2016)	Italy	a) 16 mountainous basins (255 – 6,967 sq.km) b) 6 rainfall products c) ICHYMOD model d) Warm and cold months	a) Systematic errors in rainfall are mostly dampened for CMORPH and PERSIANN b) Random errors in rainfall are dampened strongly in simulated discharge for all products c) Temporal correlation in simulated discharge decreases with the increase of basin sizes, and for cold seasons
Wu et al. (2017)	United States	a) Single basin (32,381 sq.km) b) 9 rainfall products c) DRIVE model d) Annual, monthly and flood event at daily scale	a) Strong linear relationship between bias in rainfall and discharge simulation at annual and monthly scale b) Precipitation with less bias or errors tends to have higher NSC scores c) Good correlation between antecedent precipitation bias and streamflow bias at daily scale
This study	United States	a) 1548 river basins (55 – 50,642 sq.km) b) 6 global rainfall products c) Annual, monthly and seasonal scales d) Basin classification based on climates, topography, basin size e) Different error metrics f) DRIVE physically-based hydrological model	

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Table 2. Overview of the precipitation products used in this study.

Products	Spatial Res.	Temporal Res.	Coverage	Main Source
NLDAS-2	0.125°	hourly	North America	Rain-gauges
TMPA-3B42RT V7	0.25°	3–hourly	50°S–50°N	Satellites
TMPA-3B42 V7	0.25°	3–hourly	50°S–50°N	Satellites
Stage IV	4-km	hourly	CONUS	Radars
CPC-U V1.0/RT	0.25°	daily	Global Land	Rain-gauges
MERRA-2	0.625°	hourly	Global	Reanalysis
MSWEP-2.2	0.07°	30-min	Global	Reanalysis

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

Table 3. Characteristics of Köppen climate regimes in precipitation, temperature, and hydrology.

Köppen Class	Precipitation	Temperature	Hydrological Description
Dry (cold semi-arid)	Average annual precipitation is around 400 mm, and strong seasonality in precipitation	Average annual temperature above 18 °C	Snow storage causes a delay in the streamflow; large storage variations over the year
Temperate (humid subtropical)	Average annual precipitation is around 1000 mm; no significant precipitation difference between seasons; no dry months in the summer	At least one month's average temperature above 22 °C; and at least four months averaging above 10 °C	Catchment have soil water storage variations and a slightly seasonal streamflow regime with low flows during summer
Continental (humid continental)	Average annual precipitation is higher than 1000 mm; no significant precipitation difference between seasons	Combination of hot summers and snowy winters; the warmest month of greater than 22 °C, the coldest month of below 0 °C	Catchments have small soil water storage variations and a fairly constant seasonal streamflow regime

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

List of Figures

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

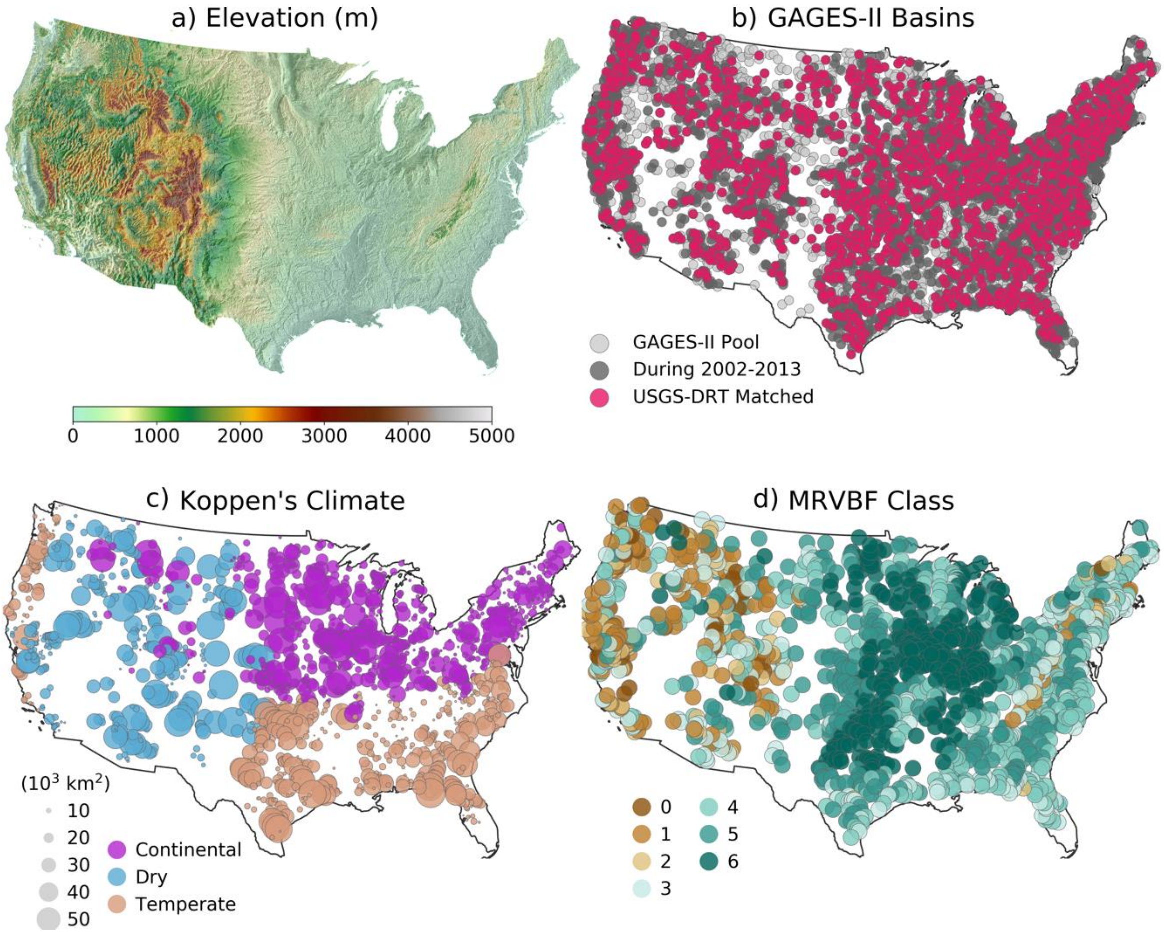

Figure 1. Maps of (a) elevation across the CONUS, (b) spatial distributions of GAGES-II sites matching with selecting criteria, (c) the Köppen's climate type and (d) MRVBF class of final selected 1,548 river basins.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

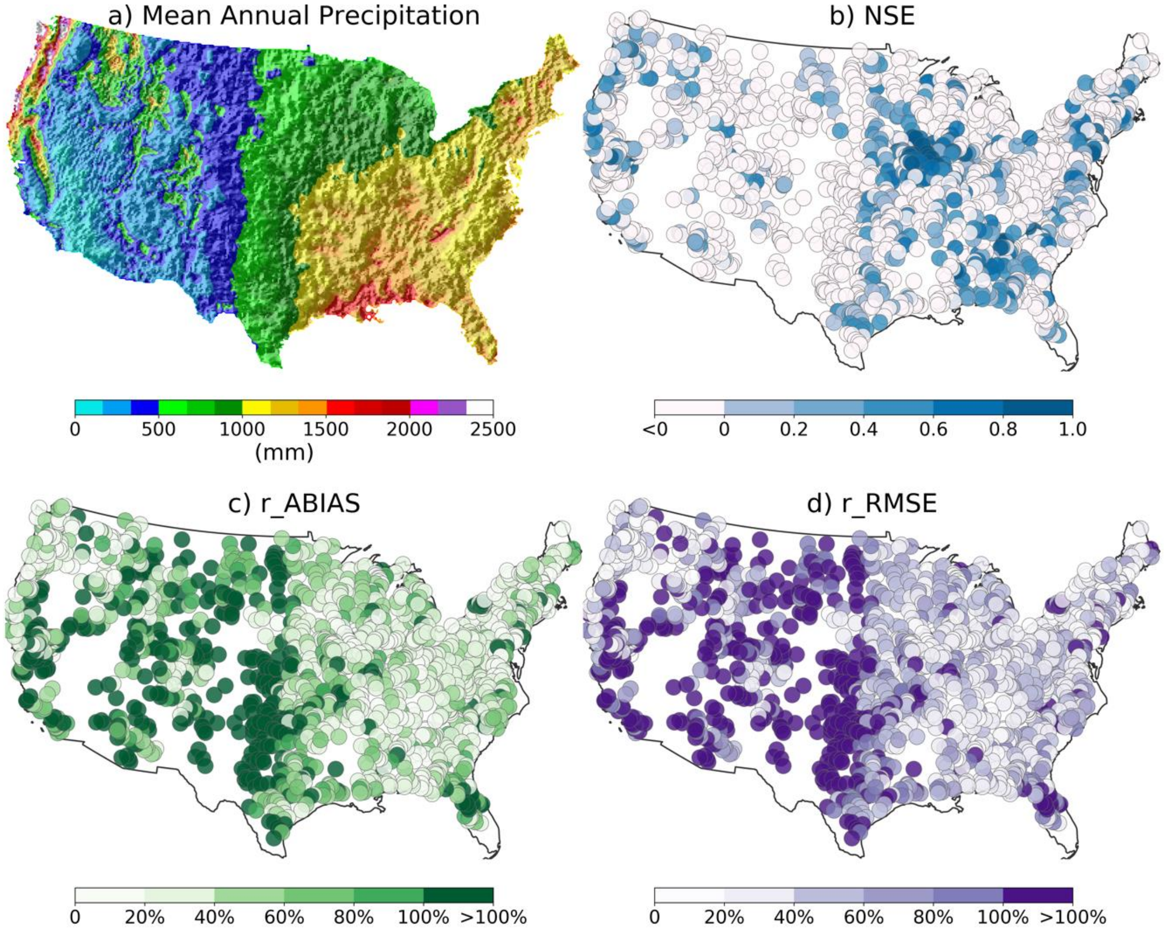

Figure 2. Maps of (a) mean annual precipitation (2002-2013) estimated by NLDAS-2 and its hydrological performance with (b) NSE, (c) relative ABIAS (r_ABIAS), and (d) relative RMSE (r_RMSE) against the USGS observations at annual timescale.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

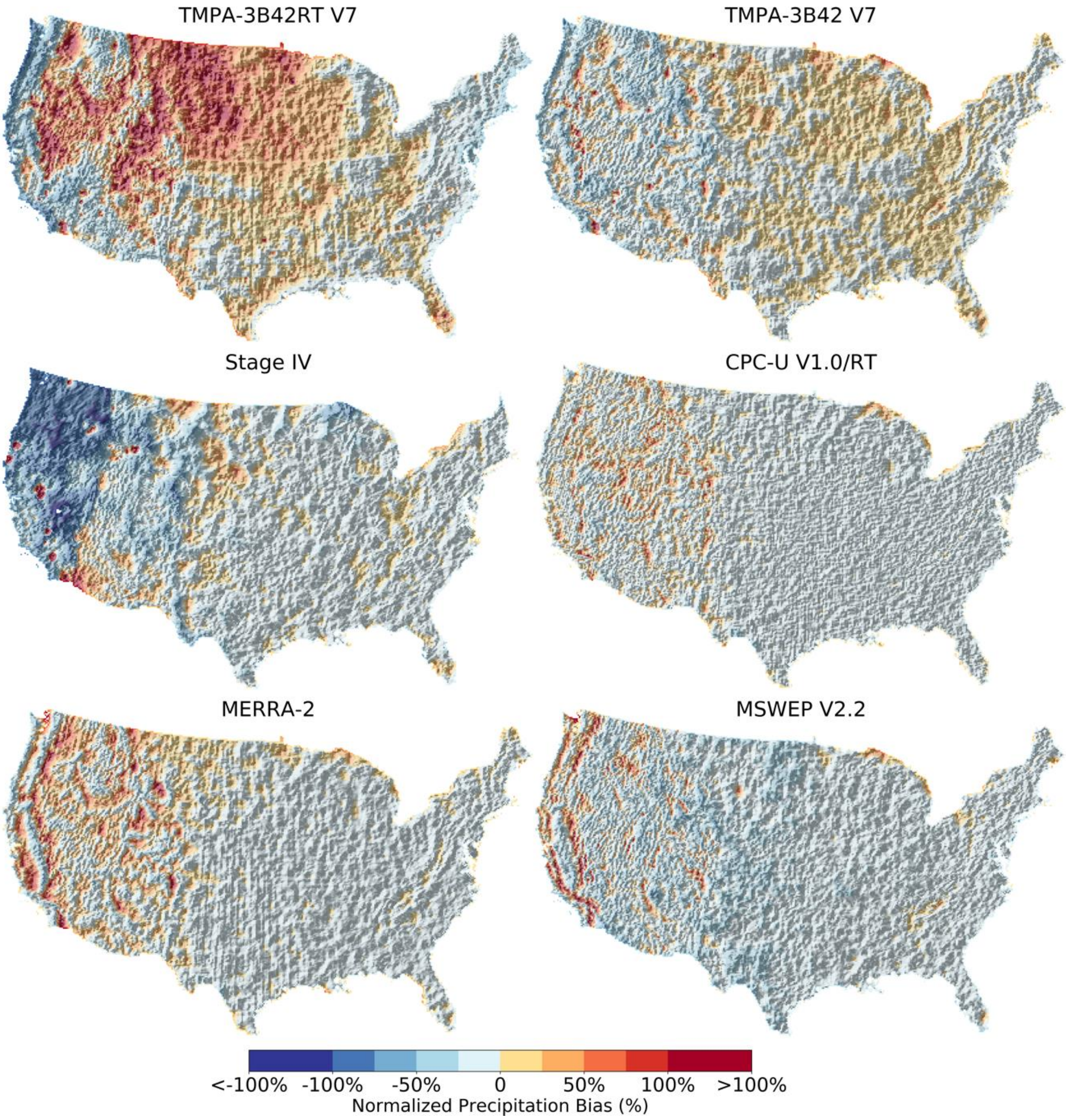

Figure 3. Normalized precipitation bias (%) in mean annual precipitation estimated by different PQEs with respect to the reference QPE (NLDAS-2) over the CONUS.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

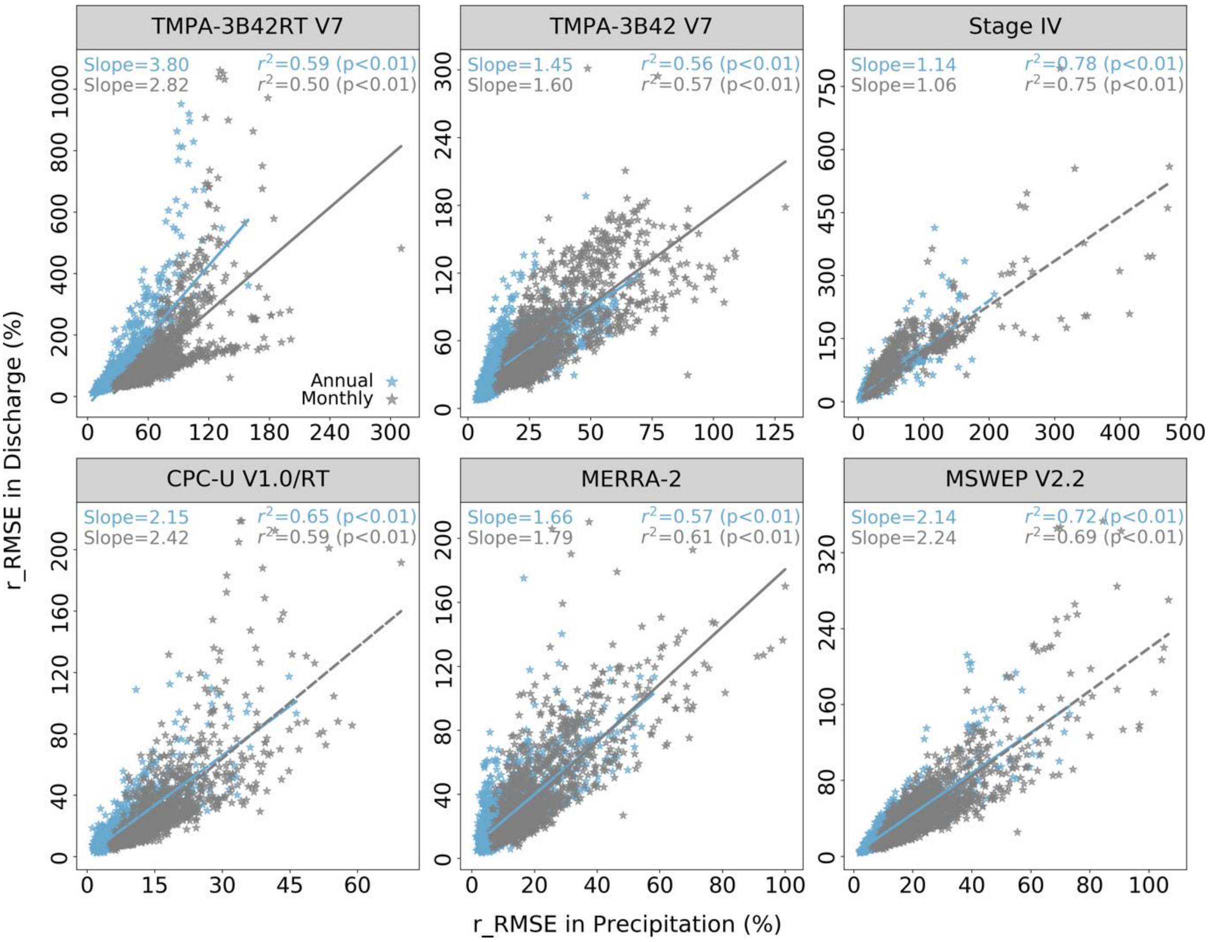

Figure 4. Scatter plot of the relative RMSE (r_RMSE) in precipitation against the r_RMSE in estimated discharges at annual (blue) and monthly (grey) timescales, respectively.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

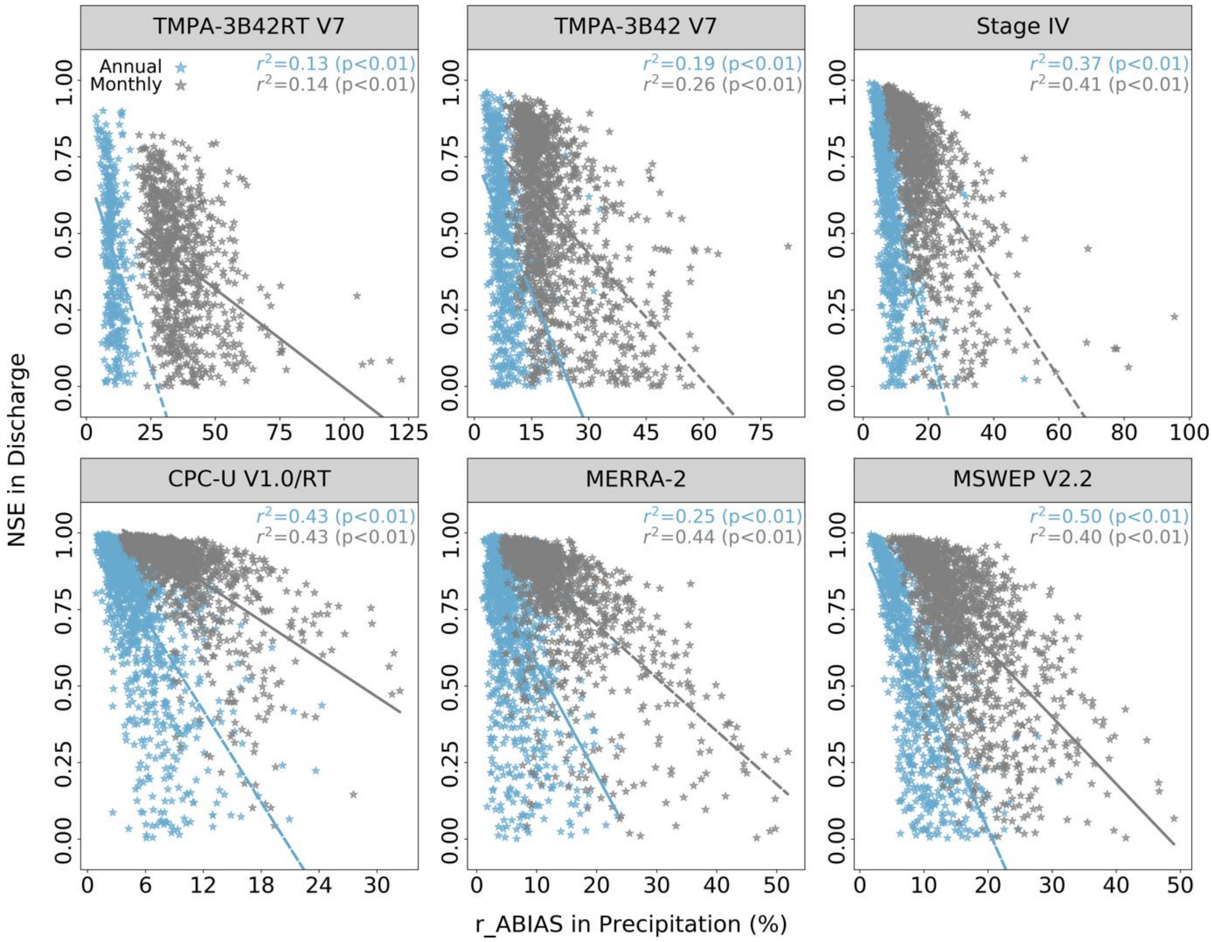

Figure 5. Scatter plot of the relative ABIAS (r_ABIAS) in precipitation against the NSE (ranging between 0 and 1) in estimated discharges at annual and monthly timescales, respectively.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

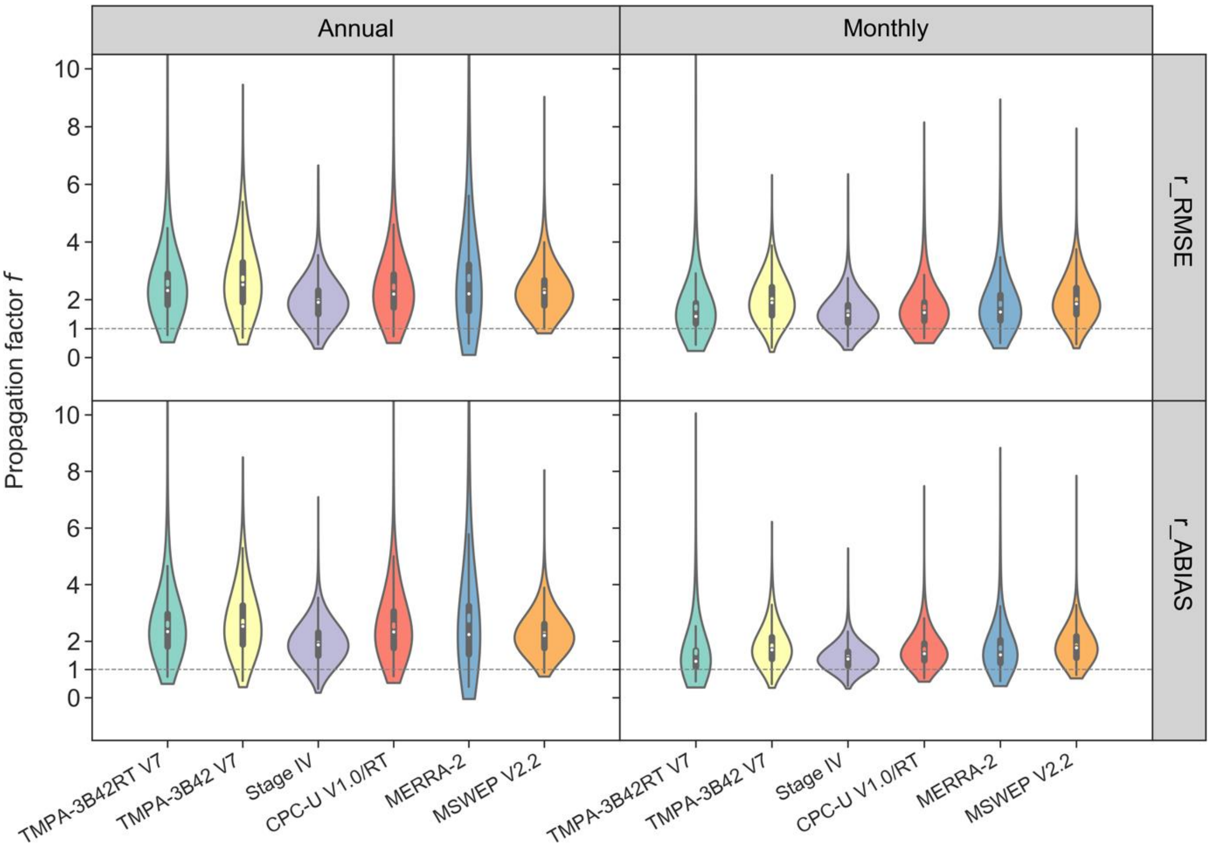

Figure 6. Violinplot of propagation factors for r_RMSE and r_ABIAS at annual and monthly timescale, dots in colors represent the mean of propagation factors for each PQE.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

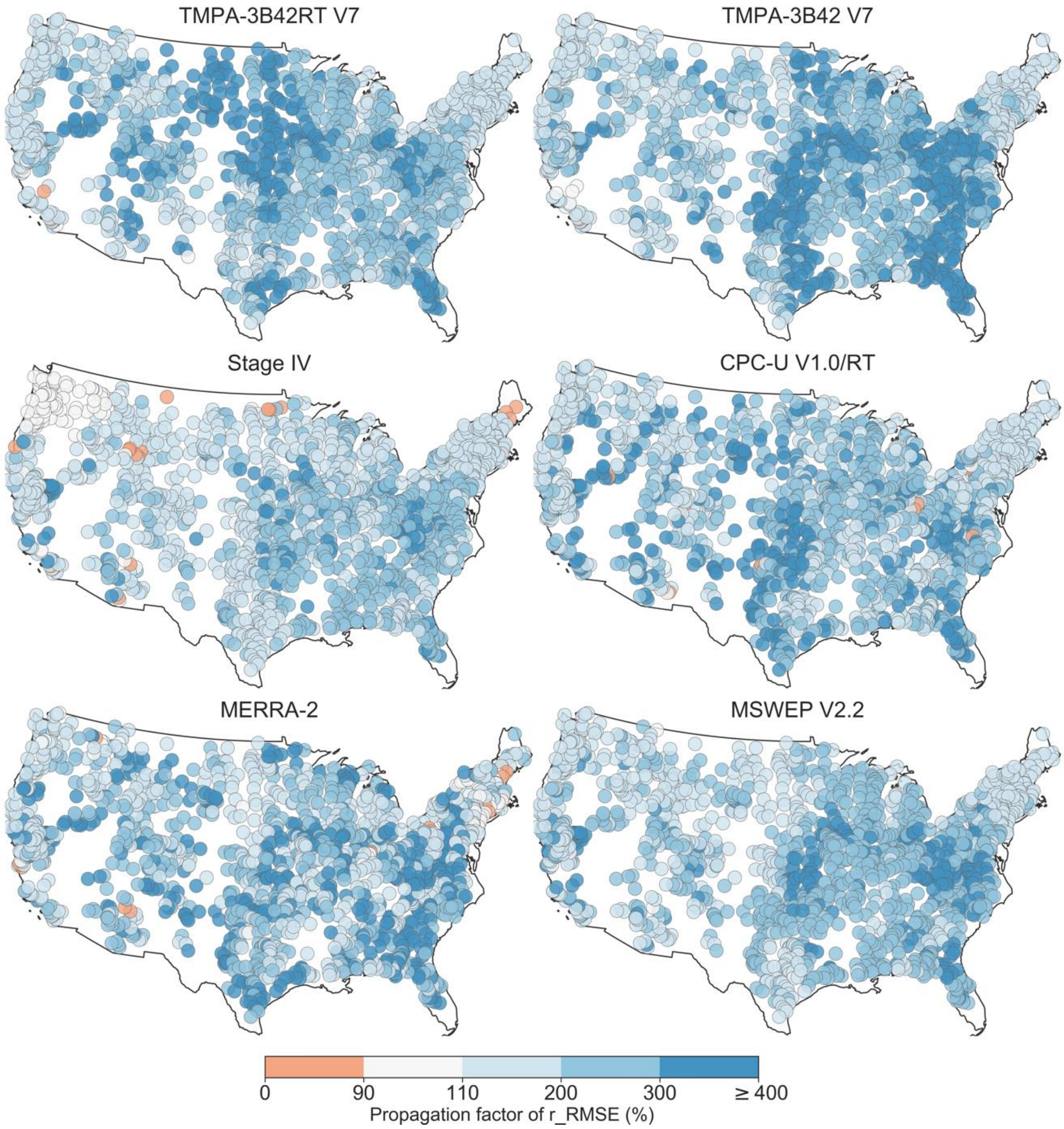

Figure 7. Spatial pattern of propagation factor of r_RMSE at annual timescale over 1,548 river basins.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

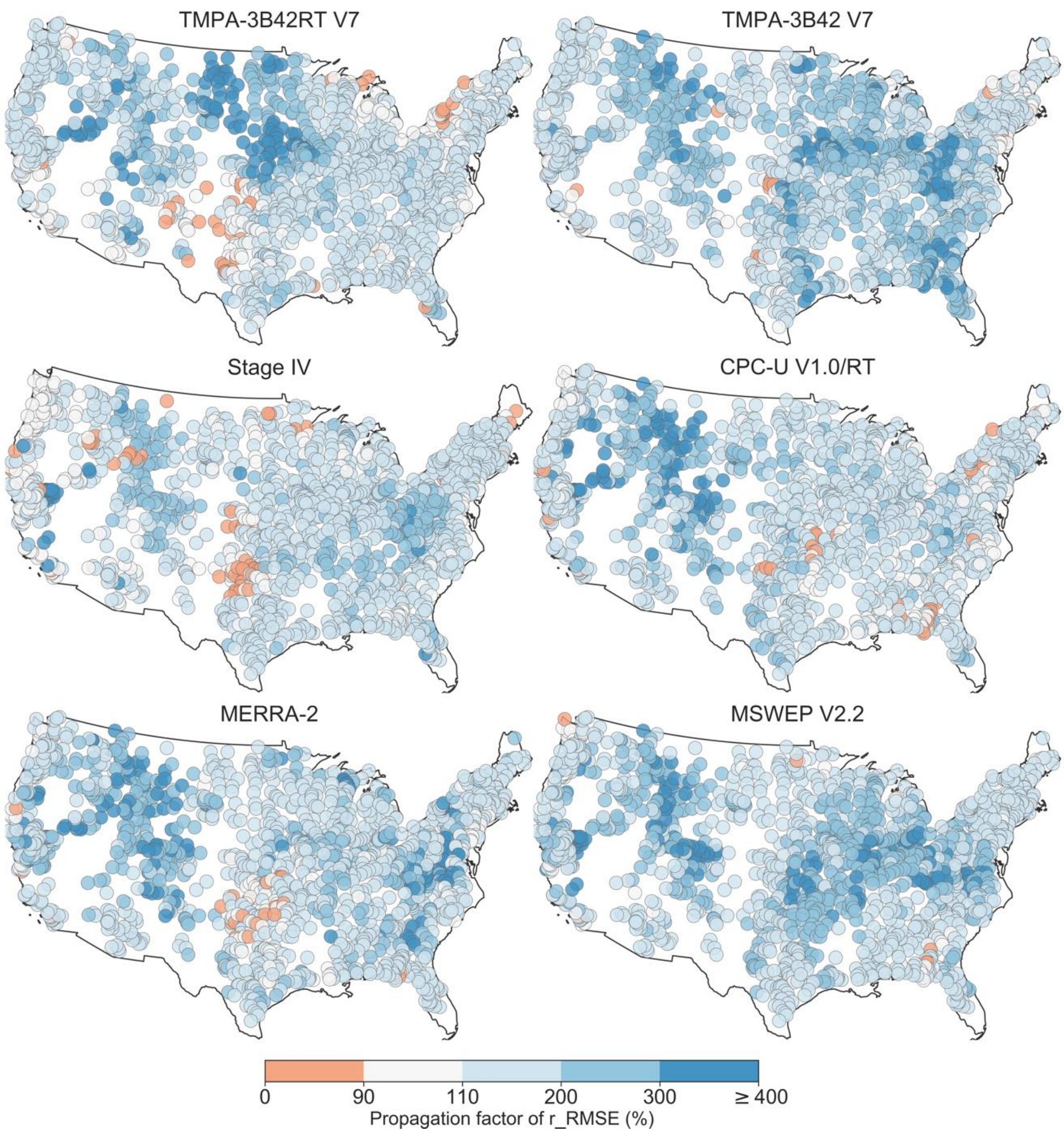

Figure 8. The same as Figure 7, but for monthly timescale.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

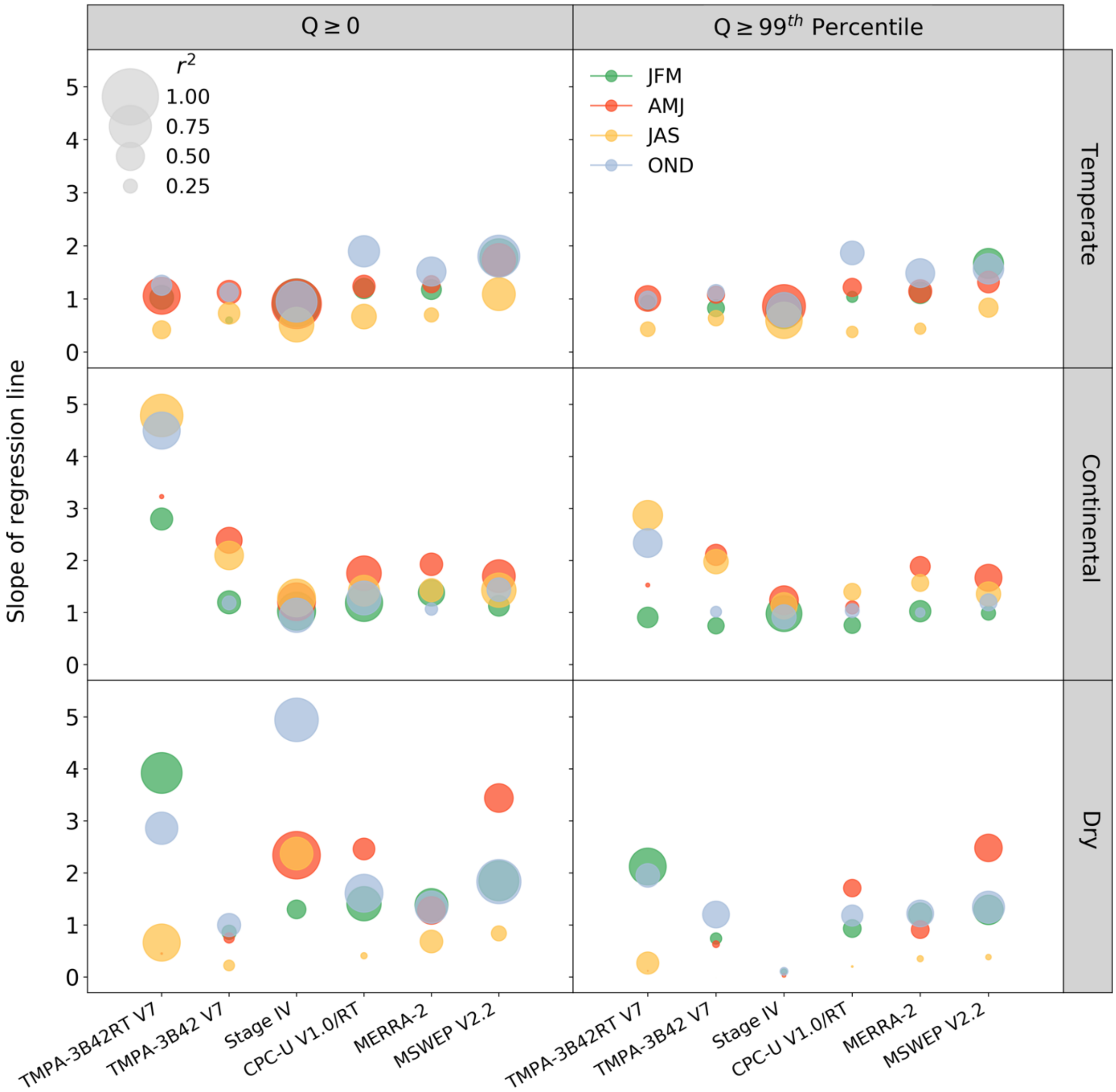

Figure 9. Seasonality in slope and coefficient of determination (r^2) of fitted regression line for the P-Q error relationship over the basins with different climate types from the perspective of r_RMSE. Different seasons are defined as January-March (JFM), April-June (AMJ), July-September (JAS) and October-December (OND).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

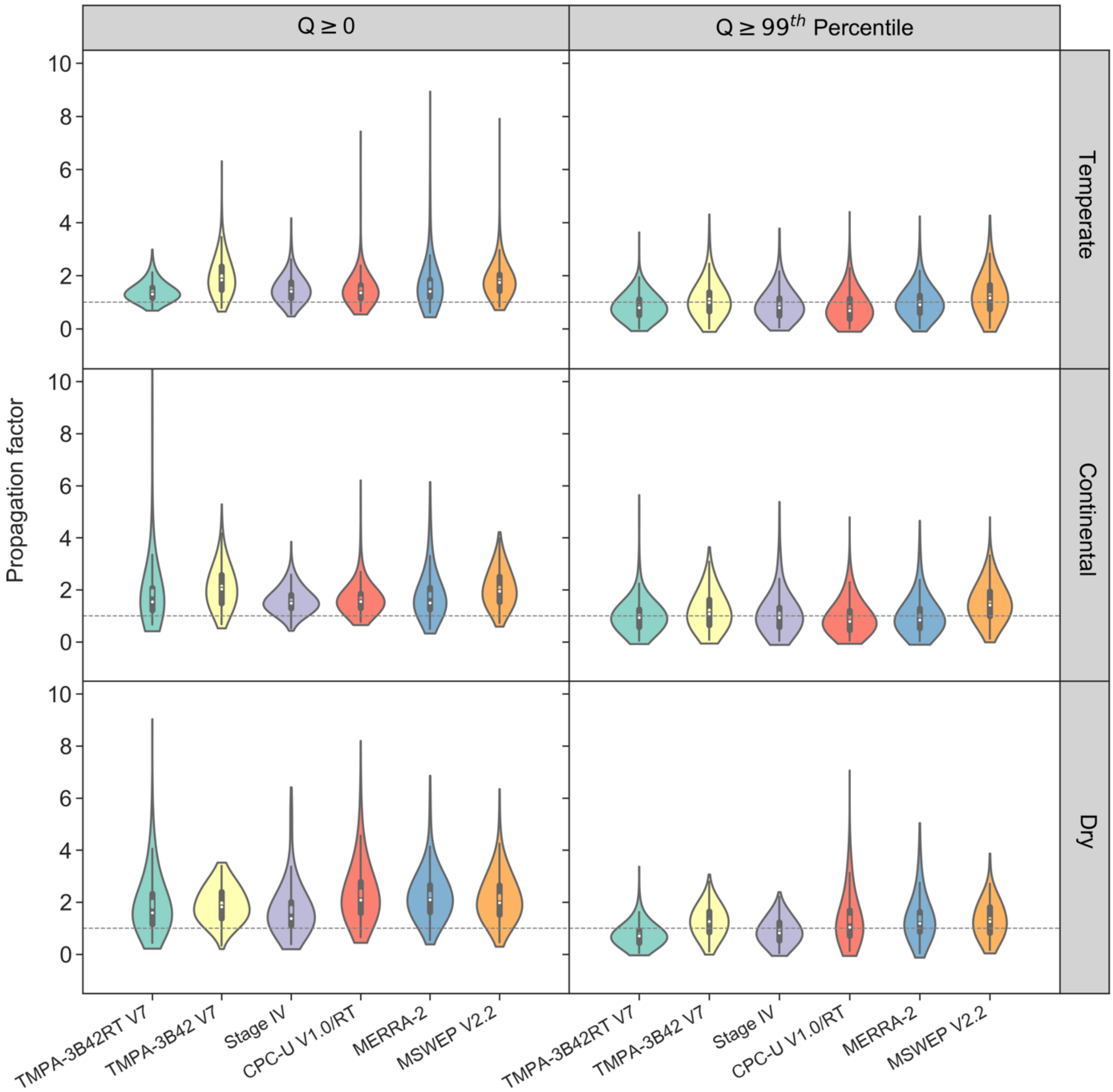

Figure 10. Propagation factors of r_RMSE for basins with different climate regimes at monthly timescale when considering different discharge magnitudes.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

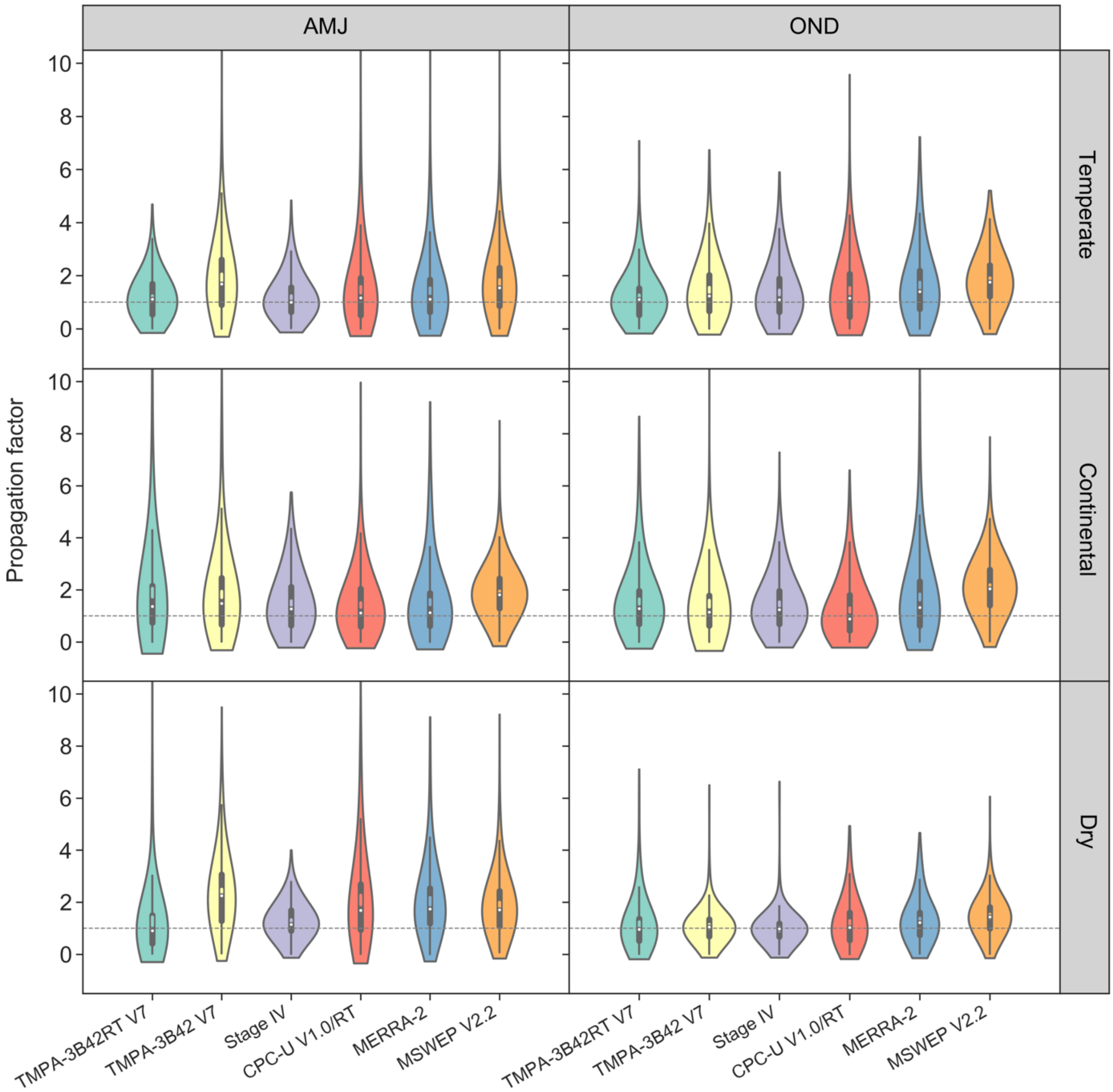

Figure 11. Propagation factors of r_RMSE for basins with different climate regimes for seasons of April-June (AMJ) and October-December (OND) when considering peak discharges ($Q \geq 99$thpercentile).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

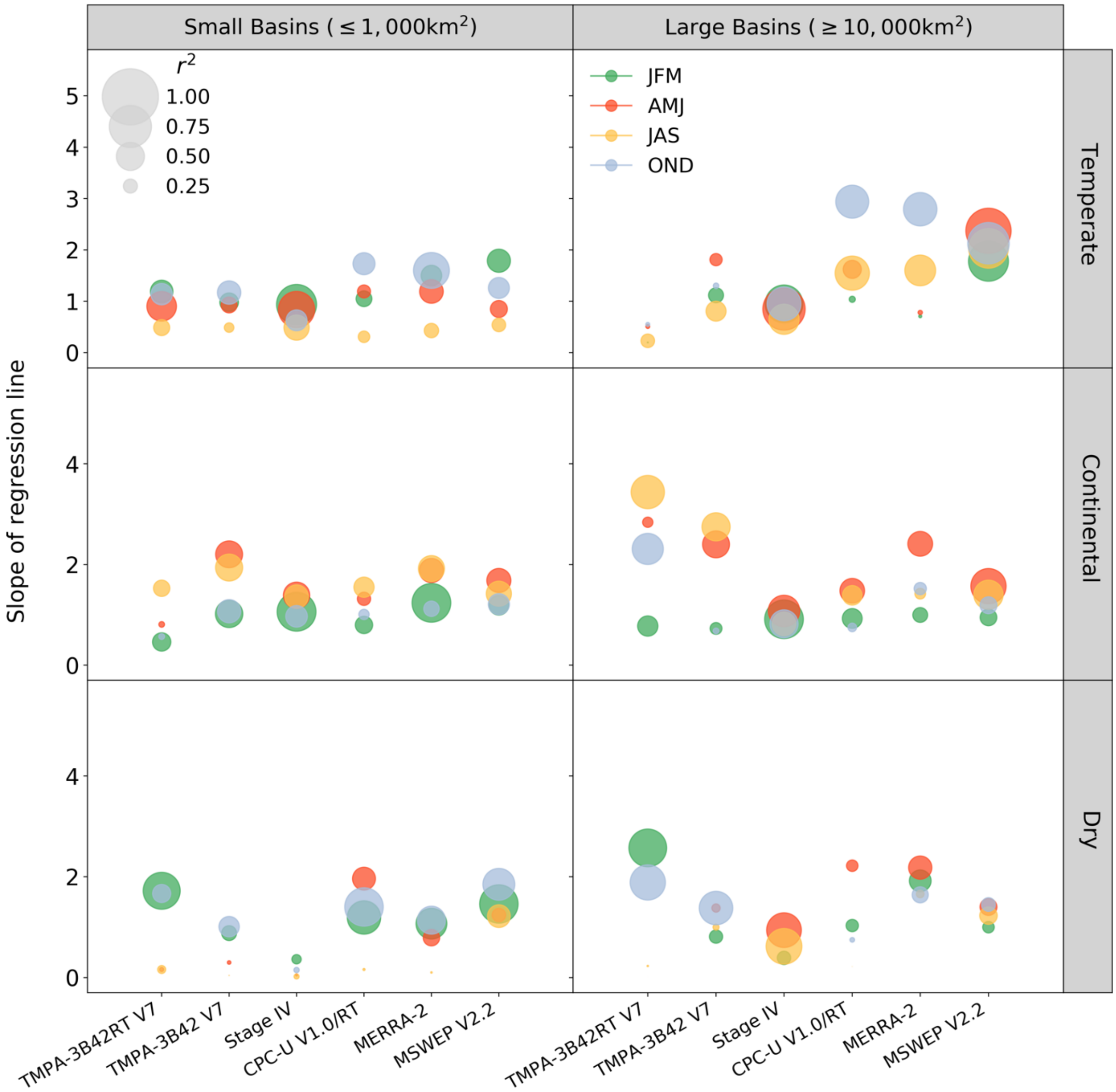

Figure 12. Seasonality in slope and coefficient of determination (r^2) of fitted regression line for the P-Q error relationship over the basins with different climate types and sizes from the perspective of r_RMSE, when considering only peak discharges ($Q \geq 99^{th}$ percentile). Seasons are defined as January-March (JFM), April-June (AMJ), July-September (JAS) and October-December (OND).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

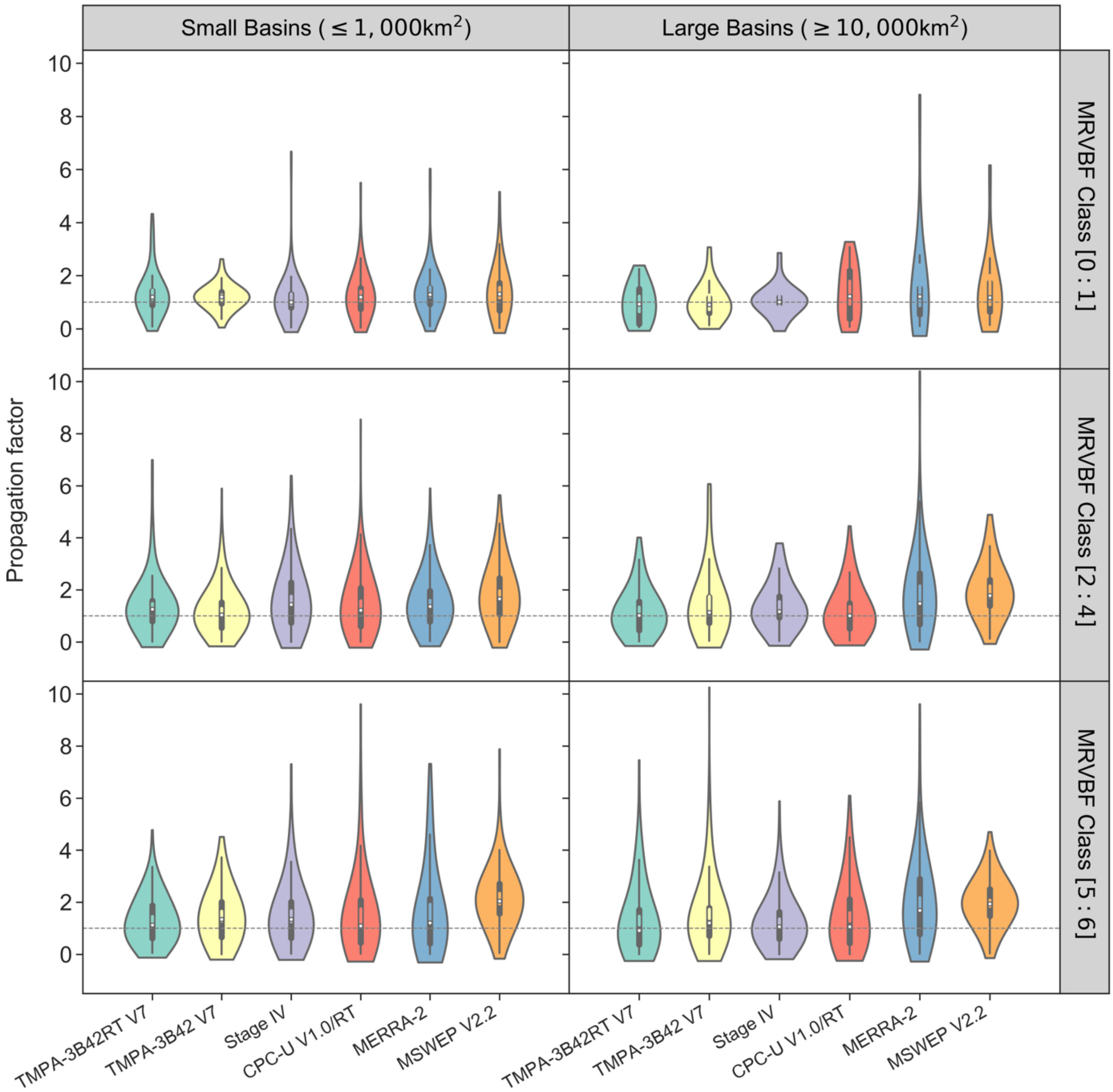

Figure 13. Propagation factors of r_RMSE for basins with different size and MRVBF classes for peak discharges ($Q \geq 99^{th}$ percentile) during the OND season. The horizontal line indicates the propagation factors of one, while white dots represent the mean value.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.

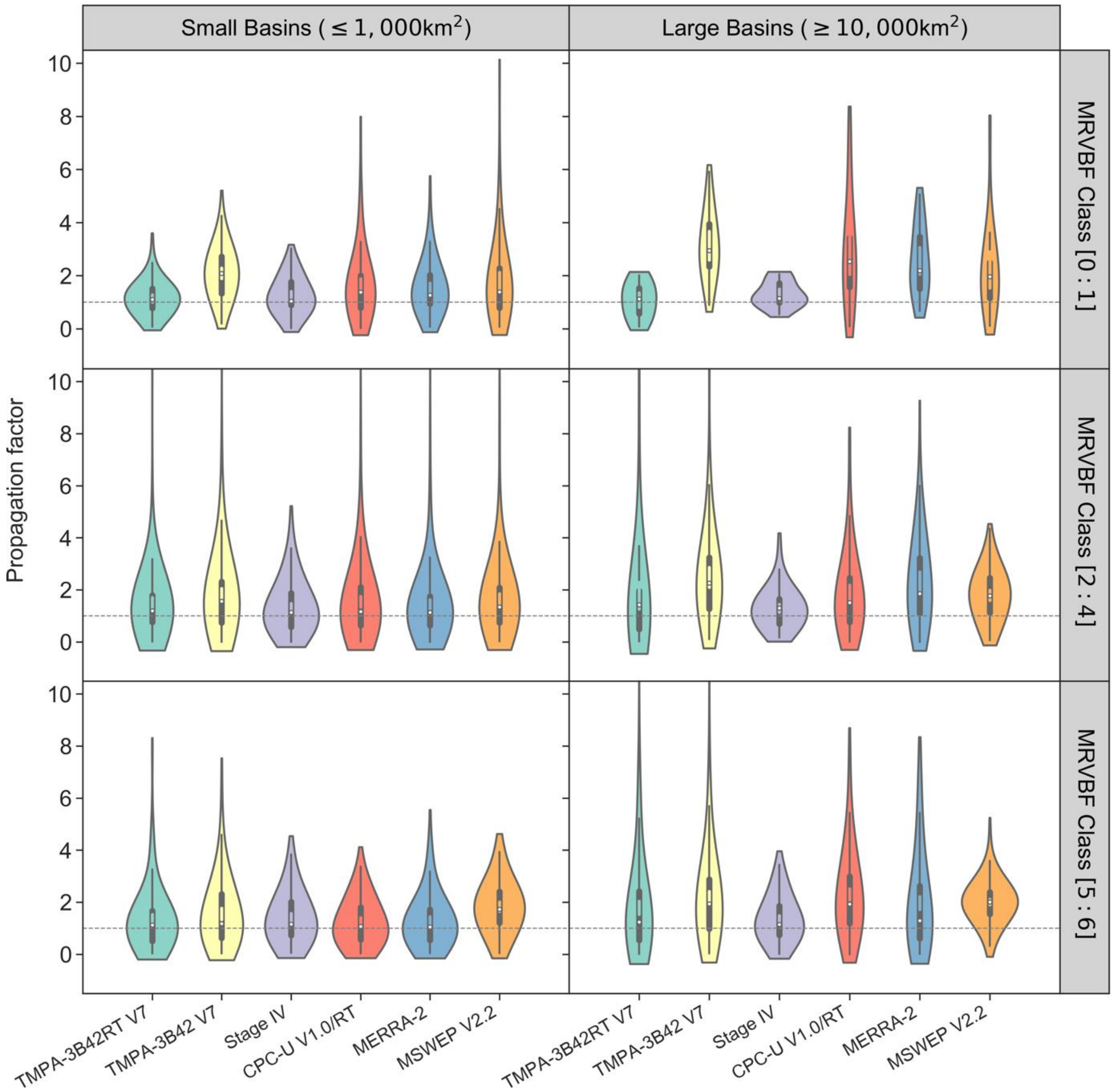

Figure 14. The same as Figure 13, but for the AMJ season.

975

976

977

978

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0213.1.