



Machine Learning–Based Blending of Satellite and Reanalysis Precipitation Datasets: A Multiregional Tropical Complex Terrain Evaluation

MD. ABUL EHSAN BHUIYAN, EFTHYMOS I. NIKOLOPOULOS, AND EMMANOUIL N. ANAGNOSTOU

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

(Manuscript received 2 April 2019, in final form 4 September 2019)

ABSTRACT

This study evaluates a machine learning–based precipitation ensemble technique (MLPET) over three mountainous tropical regions. The technique, based on quantile regression forests, integrates global satellite precipitation datasets from CMORPH, PERSIANN, GSMaP (V6), and 3B42 (V7) and an atmospheric reanalysis precipitation product (EI_GPCC) with daily soil moisture, specific humidity, and terrain elevation datasets. The complex terrain study areas include the Peruvian and Colombian Andes in South America and the Blue Nile in East Africa. Evaluation is performed at a daily time scale and 0.25° spatial resolution based on 13 years (2000–12) of reference rainfall data derived from dense in situ rain gauge networks. The technique is evaluated using *K*-fold, separately in each region, and leave-one-region-out validation experiments. Comparison of MLPET with the individual satellite and reanalysis precipitation datasets used for the blending and the recent Multi-Source Weighted-Ensemble Precipitation (MSWEP) global precipitation product exhibited improved systematic and random error statistics for all regions. In addition, it is shown that observations are encapsulated well within the ensemble envelope generated by the blending technique.

1. Introduction

At the global scale, precipitation estimation primarily relies on satellite-based observations (Smith et al. 2007; Hou et al. 2014; Huffman et al. 2015; Grecu et al. 2016). However, over complex terrain regions, satellite precipitation estimates can be associated with significant error (particularly underestimation of heavy precipitation), due to variability and uncertainty introduced by orographic effects (Roe 2005; Houze 2012; Mei et al. 2014; Derin et al. 2016). Another precipitation data source available at the global scale is from atmospheric reanalyses produced by different national and international organizations, including the National Centers for Environmental Prediction (NCEP; Kalnay et al. 1996), the European Centre for Medium-Range Weather Forecasts (ECMWF; Uppala et al. 2005; Bosilovich et al. 2008), and NASA's Goddard Space Flight Center (GSFC; Rodell et al. 2004). These products are affected by irregularly distributed observation stations and orographic effects (Gottschalck et al. 2005; Mooney et al. 2011; Peña-Arancibia et al. 2013; Seyyedi et al. 2015; Luo et al. 2019) that limit their use in

water resource applications. Recent studies have demonstrated significant uncertainties in reanalysis datasets throughout warm and cold seasons, in particular over complex terrain regions (Gao et al. 2012; Wang et al. 2015; Gao et al. 2017; Liu et al. 2018; Luo et al. 2019).

To overcome the limitations of individual satellite and reanalysis precipitation products, researchers have developed different merging techniques that combine these global precipitation products (from satellite and/or reanalysis) with in situ observations (Sinclair and Pegram 2005; Woldemeskel et al. 2013; Huffman et al. 2015; Beck et al. 2017a,b; Bhuiyan et al. 2017, 2018; Beck et al. 2019). These algorithms are based on the kriging interpolation method (Ly et al. 2011), nonparametric kernel smoothing (Li and Shao 2010), Bayesian regression (Todini 2001; Mazzetti and Todini 2004), and wavelet analysis (Heidinger et al. 2012), to name a few. Past studies have shown that merging precipitation products with rain gauge measurements has considerable advantages (Li and Shao 2010; Duque-Gardeazábal et al. 2018). The Integrated Multisatellite Retrievals for GPM (IMERG) is a blended high-resolution precipitation product developed by NASA, which merges multiple satellite precipitation estimates and ground-based precipitation measurements (Huffman et al. 2015). IMERG has

Corresponding author: Emmanouil Anagnostou, manos@uconn.edu

certainly proven promising in several studies (Sahlu et al. 2016; Dezfuli et al. 2017) but uncertainties are still present in IMERG precipitation product (Xu et al. 2016; Ma et al. 2016; Tang et al. 2016; Li et al. 2017; Ning et al. 2017). Another merging technique based on a double smoothing blending algorithm was recently developed by Duque-Gardeazábal et al. (2018) that combines remotely sensed and reanalysis data. The merging technique was shown to successfully reduce the error based on data from a local rain gauge network. However, the evaluation of the merging technique is constrained by the limited availability of validation data, that is, 3-yr record of daily rain gauge measurements from a specific region with low gauge coverage.

The most comprehensive and extensively validated blending technique to date is the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset that merges a wide range of gauge, satellite, and reanalysis precipitation datasets at the global scale (Beck et al. 2017a,b, 2019). MSWEP dataset is global, spanning a period of 1979–2017 (for version 2; Beck et al. 2019), and is designed to take advantage of the complementary strengths of gauge-, satellite-, and reanalysis-based data (Beck et al. 2019). MSWEP presents more complete global patterns (including ocean areas) in mean annual precipitation, percentiles, and annual number of dry days over topographically complex terrain compared to other precipitation datasets (Beck et al. 2019). In addition, to obtain the best possible precipitation estimates at global scale, MSWEP accounted a gauge-correction scheme that minimizes timing mismatches when applying the daily gauge corrections (Beck et al. 2019).

Bhuiyan et al. (2017) developed a machine learning-based multisource data blending technique and have used it to evaluate the impact of land surface conditions (e.g., vegetation cover and soil moisture) on passive microwave precipitation retrieval error. They successfully evaluated the technique based on high-resolution ground radar-rainfall estimates over the southern continental United States. The technique was then extended to integrate multiple global precipitation datasets for improving the accuracy of 3-hourly precipitation estimates and characterizing estimation uncertainty (Bhuiyan et al. 2018). The technique [hereafter named machine learning-based precipitation ensemble technique (MLPET)] combines daily soil moisture, terrain elevation, and atmospheric variables (temperature and humidity) with multiple global precipitation datasets (from satellite and reanalysis products) to statistically generate ensembles of precipitation fields. Its performance was demonstrated in Bhuiyan et al. (2018) over the Iberian Peninsula, which represents a combination of oceanic and semiarid climate, in terms of both precipitation estimates and hydrological

simulations. However, one vital question that pertains to such a regionally tuned approach is how “transferable” it is to other complex terrain regions representing varying topography and precipitation characteristics, and how the technique performs relative to other blended products (e.g., IMERG and MSWEP).

This study addresses this objective by evaluating the technique over three tropical complex terrain regions and comparing its performance against individual satellite precipitation datasets and the MSWEP (both V1 and V2) blended product. The study areas are the Peruvian and Colombian Andes in South America and the Blue Nile in East Africa. In this study, we used ground-reference precipitation derived from dense rain gauge networks. Through this multiregional/multiyear error analysis our study investigates the advantages of the proposed blending technique over individual precipitation datasets and its ability to characterize the estimation uncertainty through stochastic generation of precipitation fields. In this paper we will discuss two different validation experiments that represent scenarios of in situ tuning versus global transferability of the technique: *K*-fold conducted for each region separately and leave-one-region-out cross validation. The leave-one-region-out validation experiment is conducted to evaluate transferability of the technique to mountainous areas lacking adequate in situ gauge coverage.

The paper is organized as follows: in section 2, the study area and the datasets used in this study are briefly explained. The evaluation methodology is explained in section 3. Results and discussion are described in section 4. Finally, conclusions and recommendations for future research are summarized in section 5.

2. Study area and datasets

a. Study domains and rain gauge datasets

The Blue Nile, Peruvian Andes, and Colombian Andes study areas and the corresponding in situ gauge networks are shown in Fig. 1. The Blue Nile and the Colombian Andes are generally characterized by tropical climatic conditions. The Peruvian Andes is mostly influenced by tropical, equatorial, and monsoon climate conditions. Table 1 summarizes the climatic information for the selected study areas. Moreover, the updated world Köppen–Geiger climate map where details of the climatic zones for each region are presented can be found in the following URL (<https://people.eng.unimelb.edu.au/mpeel/koppen.html>; Peel et al. 2007).

The upper Blue Nile basin is located in the mountain range of the Great Horn of Africa. The elevation ranges from approximately 1000 to 3000 m MSL. A total of 70

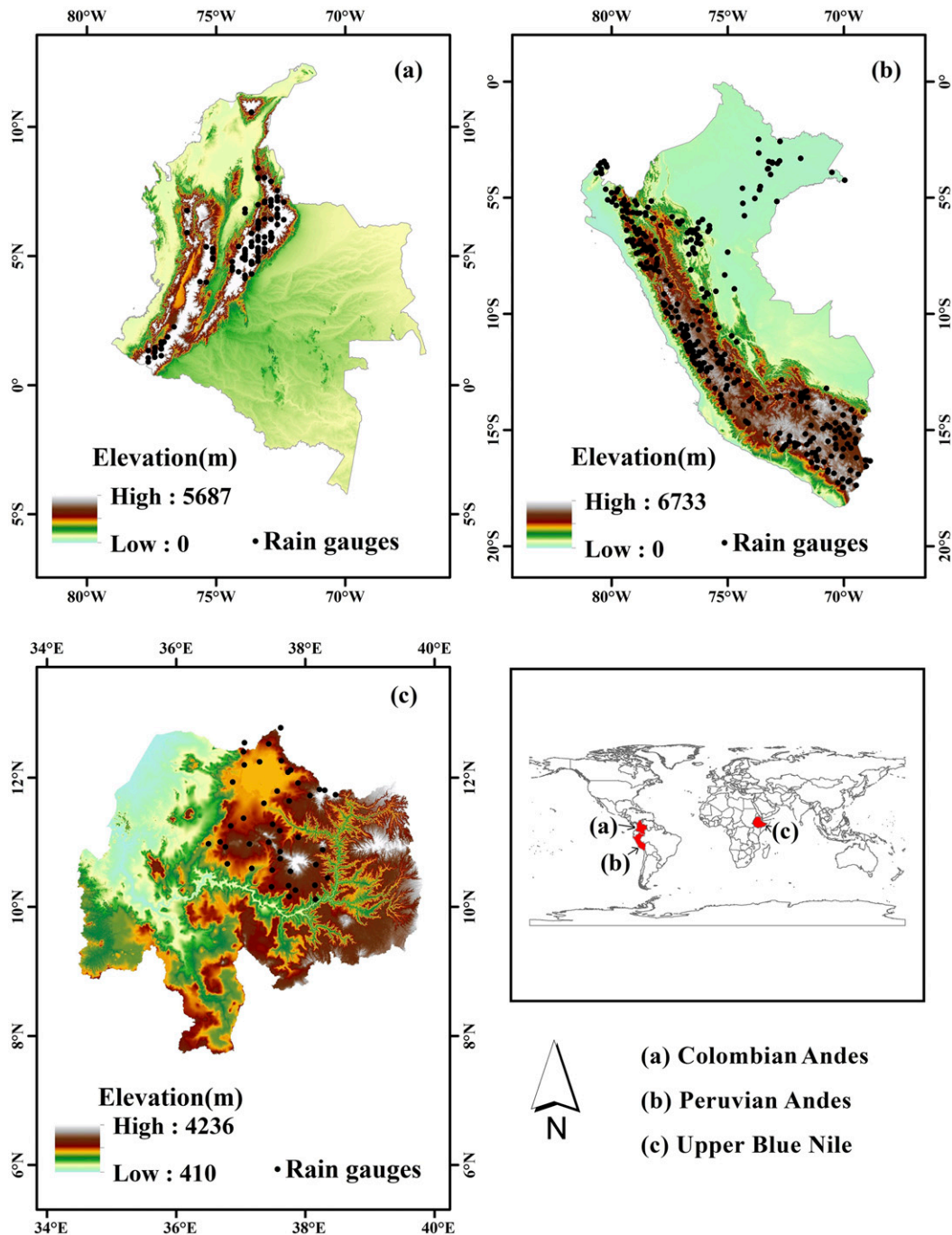


FIG. 1. Geographic locations and terrain elevation map of the study areas. Solid circles represent gauge locations.

daily rain gauge stations are available for this region for a period of 13 years (2000–12). Basin average annual precipitation is 1100 mm and the maximum basin average daily precipitation is 312 mm. The climate varies from cool highlands to hot deserts with the majority of precipitation falling from June to September (Cattani et al. 2016, 2018).

The study area in Peru has 323 rain gauges and elevations range from sea level to 5020 m MSL. Area average annual precipitation is 713 mm and maximum area average daily precipitation is 540 mm. The semiarid Pacific coastline experiences less than 100 mm precipitation per year due to the cold von Humboldt Current (Manz et al. 2016). Precipitation increases toward the Andes driven

TABLE 1. Köppen–Geiger climatic zones in the study areas.

Blue Nile	Colombian Andes	Peruvian Andes
1. Tropical Savanna climate (Aw)	1. Temperate, dry summer, warm summer (Csb)	1. Equatorial climate (Af)
2. Warm desert climate (BWh)	2. Temperate, dry summer, cold summer (Csc)	2. Monsoon climate (Am)
3. Warm semiarid climate (BSh)	3. Temperate, no dry season, warm summer (Cfb)	3. Tropical savanna climate (Aw)
4. Humid subtropical climate (Cwa)		4. Warm semiarid climate (BWh)
		5. Warm semiarid climate (Bsh)
		6. Cold semiarid climate (Bsk)
		7. Cold desert climate (BWk)
		8. Humid subtropical climate (Cwa)
		9. Humid subtropical climate/Subtropical oceanic highland climate (Cwb)
		10. Temperate oceanic climate (Cfb)

by strong topographic gradients (Espinoza et al. 2015; Espinoza Villar et al. 2009; Bookhagen and Strecker 2008).

The Colombian Andes study area has 104 rain gauges and elevation ranges from approximately 1000 to 3000 m MSL. The area average annual precipitation and maximum area average daily precipitation are 1600 and 439 mm, respectively. Regional hydroclimatic variability is caused by the intra-Andean valleys. Central Colombia has two rainy seasons from April to May and from September to November (Álvarez-Villa et al. 2011).

The reference precipitation datasets, derived from the above rain gauge networks, spanned a period of 13 years (2000–12). All datasets used in this study were mapped at the same spatial (0.25°) grid resolution. Gauges were interpolated at the 0.25° grid cells by averaging gauge measurements within common precipitation grid cells.

b. Satellite-based precipitation product

Four gauge-adjusted quasi-global satellite precipitation products CMORPH, PERSIANN, 3B42 (V7), and GSMaP (V6) were used in this study. The Climate Prediction Center morphing technique (CMORPH) of the National Oceanic and Atmospheric Administration (NOAA) depends on passive microwave (PMW) satellite precipitation fields propagated by motion vectors calculated from infrared (IR) observations (Joyce et al. 2004). Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) connects IR observations to PMW rainfall estimates through a neural network (Sorooshian et al. 2000). Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA), or 3B42 (V7), is gauge adjusted data that merges IR and PMW precipitation products from NASA and is available in both near-real time and post-real time (Huffman et al. 2010). Global Satellite Mapping (GSMaP) is calculated from PMW precipitation data and propagated using IR estimates GSMaP-MVK and gauge-adjustment GSMaP (V6). The Earth Observation Research Center (EORC) of the

Japan Aerospace Exploration Agency (JAXA) produces these datasets (Kubota et al. 2007; Ushio et al. 2013). The newest GSMaP algorithm includes a technique for correcting orographic rainfall effects (Yamamoto and Shige 2015; Yamamoto et al. 2017). The precipitation products used in this study have a spatial resolution of 0.25° and daily temporal resolution covering the period of 2000–12.

c. Reanalysis product

The reanalysis data were based on the original ERA-Interim data (Dee et al. 2011) that were used in ERA-Interim/Land (Balsamo et al. 2015) after rescaling based on the Global Precipitation Climatology Center (GPCC) dataset. The reanalysis precipitation dataset (EI_GPCC) that we used in this study was further downscaled from 0.5° to 0.25° based on the Climate Hazards Group's Precipitation Climatology (CHPclim). The WATCH (Water and Global Change FP7 project) Forcing Dataset ERA-Interim (hereafter WFDEI; Weedon et al. 2014) is based on ERA-Interim with a geographical resolution of 0.5° and bias corrections using gridded rain gauge datasets. We chose specific humidity among WFDEI variables as a predictor for the blending technique. Moreover, the soil moisture product (V1.0) was selected from Earth2Observe Tier-1 datasets, performed by ECMWF land surface scheme, and forced by the WFDEI datasets (details in Schellekens et al. 2017).

d. MSWEP

The MSWEP V1 and V2 datasets use information from rain gauge measurements, satellite precipitation datasets, and estimates from atmospheric models (Beck et al. 2017a,b, 2019). MSWEP (V1) product is available at 0.25° at the daily temporal scale. Weights given to the satellite/reanalysis products are determined from their comparative performance at the surrounding gauges, which creates improved precipitation at each grid (Beck et al. 2017a). Moreover, MSWEP (V2) is a high spatial

resolution product available at 3-hourly temporal and 0.1° spatial resolutions (Beck et al. 2019). The product spans a period from 1979 to 2017. We used nearest neighbor interpolation method to match the 0.25° spatial resolution of the different precipitation datasets. MSWEP (V2) considered several factors such as cumulative distribution function and precipitation frequency corrections, inclusion of ocean areas; daily (rather than monthly) gauge corrections; gauge reporting times; and correction of systematic terrestrial precipitation biases that were not considered in MSWEP (V1). MSWEP (V2) has been validated at global scale using a large number of gauge observations showing that it performs substantially better than other widely used satellite, ground-based, and reanalysis precipitation datasets (Beck et al. 2019). Therefore, in this study we consider that the MSWEP datasets represent a baseline performance for current state-of-the-art blended products.

e. MLPET technique

The herein evaluated product is based on the MLPET technique described in Bhuiyan et al. (2018). Specifically, the MLPET technique (Bhuiyan et al. 2018) combines multiple satellite and reanalysis precipitation datasets along with dynamic and static land surface variables to generate stochastically a precipitation ensemble based on, quantile regression forests (QRF) algorithm (Meinshausen 2006). In addition to optimally merging satellite and reanalysis precipitation datasets, MLPET technique is used to characterize the uncertainty of the blended product (Bhuiyan et al. 2018). As demonstrated by Bhuiyan et al. (2018), in the Iberian Peninsula the MLPET performed significantly better than the individual satellite and reanalysis precipitation datasets. In this study, the MLPET technique is applied over multiple complex terrain study areas using as input the same satellite precipitation products used in Bhuiyan et al. (2018): CMORPH, PERSIANN, GSMaP (V6), 3B42 (V7), and the EI_GPCC atmospheric reanalysis precipitation product; the daily soil moisture and specific humidity are derived from WFDEI and a terrain elevation dataset. The soil moisture and specific humidity variables were interpolated in space using the nearest neighbor interpolation technique to match the 0.25° spatial resolution of the different precipitation datasets. Finally, all daily data were mapped to the 0.25° grid chosen to be the final spatial grid for the blended product.

3. Evaluation methodology

a. Variable importance

To generate the blended product, variable selection was based on recent research, which demonstrated that

soil moisture, satellite precipitation estimates, reanalysis precipitation products, and elevation are potential significant predictors in the error modeling of rainfall estimates (Seyyedi et al. 2014; Mei et al. 2016; Bhuiyan et al. 2017, 2018). In addition, we included a daily specific humidity dataset with global coverage at 0.5° spatial resolution. The specific humidity has been identified in past studies as a forcing factor in controlling orographic precipitation characteristics of tropical regions (Cau et al. 2007; Anders and Nesbitt 2015). After choosing these predictor variables, the variable importance methodology (Breiman 2001) was applied to confirm the relative importance of those variables at the different regions, similar to the process described in Bhuiyan et al. (2018).

b. Cross-validation experiments

In this analysis, the MLPET technique was evaluated using two cross validation experiments: K-fold, applied separately in each region, and leave-one-region-out cross-validation experiments. Cross validation is a useful procedure for evaluating the effectiveness of a machine learning model, specifically to determine the hyper parameters by minimizing test error of the machine learning model.

Initially, the K-fold ($K = 5$) cross-validation technique was used to develop the model in QRF application. In the fivefold cross-validation experiment, the dataset was randomly split into five partitions; model is trained on the training set and each holdout sample is treated as validation dataset. For each region, a five times random holdout was applied using 80% of the data as training and 20% of the data as validation.

The leave-one-region-out cross-validation procedure leaves one region out of the three study regions. The two regions are used to train the model and the hold-out region is used for validation. The holdout method is repeated three times, such that each time, one of the regions is used for validation and the other two regions are combined to form a training dataset.

A key benefit of these cross-validation experiments is that they give better estimates of the actual model performance because they treat each data point as an independent holdout sample. However, complex topography and regional climate characteristics play a vital role in controlling the spatial pattern of precipitation (Block and Rajagopalan 2007; Enyew and Steeneveld 2014). Moreover, accurate estimate of precipitation can be a challenging task when the quality of the validation dataset is low or the site is ungauged. The leave-one-region-out experiment is evaluating the model under this condition by testing the transferability of the technique's calibration from other complex terrain regions.

c. Performance evaluation error metrics

Several error metrics are used in this study and are summarized below.

The normalized centered root-mean-square error (NCRMSE) is a statistical metric used to measure the magnitude of the random error and defined as

$$\text{NCRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \left[\hat{y}_i - y_i - \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \right]^2}}{\frac{1}{n} \sum_{i=1}^n y_i}. \quad (1)$$

Here, y_i is the observed precipitation, \hat{y}_i is estimated rainfall for blended technique, and n is the total number of data points used in the calculation. NCRMSE ranges from 0 to positive infinity with 0 indicating no random error.

The bias ratio (BR) is the mean of the ratio of estimated rainfall to reference rainfall and is defined as

$$\text{BR} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i}{y_i} \right). \quad (2)$$

For an unbiased model, the BR would be 1.

Mean relative difference (MRD) is the mean of the relative percentage error, calculated by the normalized average:

$$\text{MRD} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right). \quad (3)$$

MRD represents the magnitude and direction of error with positive value referring to overestimation while negative value referring to underestimation.

To evaluate the relative error metric difference (Δ_{error} ; %) between MLPET and individual precipitation products we devised the following formula:

$$\Delta_{\text{error}} = \frac{\Delta_b - \Delta_i}{\Delta_i}, \quad (4)$$

where Δ_b indicates the error metric for MLPET and Δ_i represents the error metric for individual precipitation products. To calculate the relative random error reduction, NCRMSE error metric is used in Eq. (4). Similarly, to calculate the relative systematic error reduction, we used the mean relative difference in Eq. (4).

To demonstrate the performance of the individual precipitation products, the Taylor diagram is used in this study to examine how closely the precipitation products matched the reference precipitation (Taylor 2001). The normalized Taylor diagram presents the normalized

root-mean-square difference (NRMSD), correlation coefficient (CC), and standard deviation between estimates and observations, for evaluating the relative accuracy of various precipitation products (details in Bhuiyan et al. 2019).

Furthermore, the Kling–Gupta efficiency (KGE; Gupta et al. 2009) metric was used to show the overall model performance, which is widely adopted for global hydrological application (Formetta et al. 2011; Beck et al. 2016). KGE provides three distinct components representing the correlation, the bias, and a measure of relative variability in the simulated and observed values. KGE represents the difference between unity and the Euclidian distance (ED) from the ideal point in the three-dimensional criteria space:

$$\text{KGE} = 1 - \text{ED}, \quad (5)$$

$$\text{ED} = \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}. \quad (6)$$

KGE values range from negative infinity to 1. Here, r represents Pearson's correlation coefficient. Moreover, α and β indicate the variability error and the bias error, respectively. The defined parameters are expressed as follows:

$$\alpha = \frac{\sigma_m}{\sigma_o}, \quad (7)$$

$$\beta = \frac{\mu_m}{\mu_o}, \quad (8)$$

where σ_m = standard deviation of model, σ_o = standard deviation of reference, μ_m = mean of model, and μ_o = mean of reference.

Finally, to evaluate the technique's ensemble characterization of precipitation estimation uncertainty we used three error metrics. The first metric is exceedance probability (EP), which is the frequency that the reference value exceeds the prediction interval:

$$\text{EP} = 1 - \frac{1}{n} \sum_{i=1}^n 1_{\{Q_{\text{lower}_i} < y_i < Q_{\text{upper}_i}\}}. \quad (9)$$

Here, Q_{lower} and Q_{upper} denote lower and upper boundaries of prediction interval, respectively. An EP value of zero indicates perfect encapsulation of the reference within the ensemble envelope.

The second metric is uncertainty ratio (UR), which quantifies the prediction interval width (Q_{lower} , Q_{upper}) relative to the magnitude of the predicted variable and is defined as

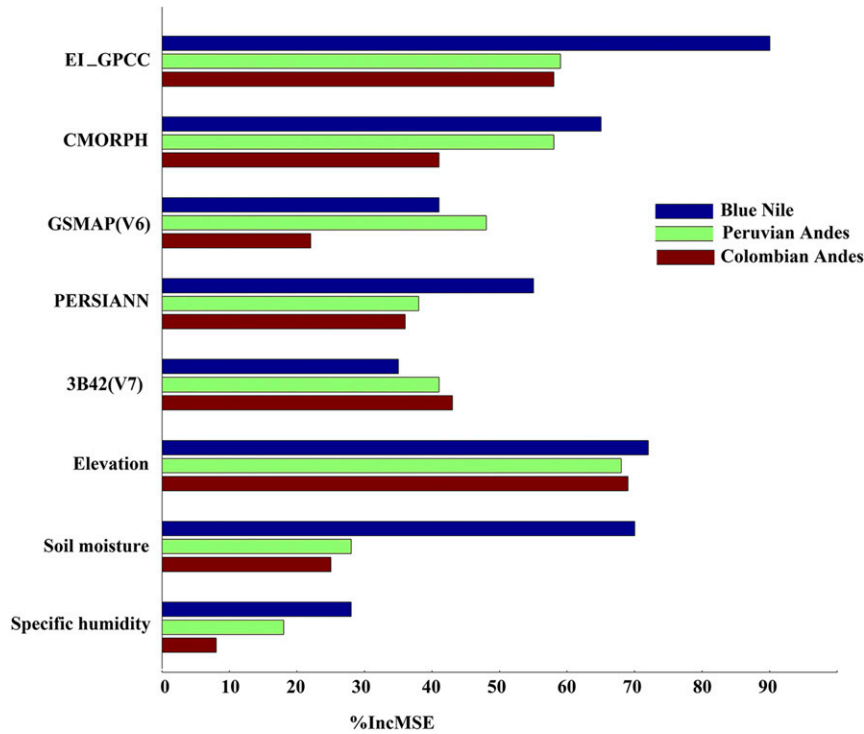


FIG. 2. Variable importance plot, where %IncMSE is the percentage increase in mean square error.

$$UR = \frac{\sum_{i=1}^n (Q_{upper} - Q_{lower})}{\sum_{i=1}^n y_i}. \quad (10)$$

A higher UR value indicates broader range of prediction intervals or low confidence in the ensemble mean. An accurate prediction is represented by low UR and low EP values.

4. Results

a. Sensitivity analysis

Results from the variable importance test of the eight predictor variables for the three regions are presented in Fig. 2. Results show that all variables used in this technique contribute significant information in the modeling process, which is consistent to the finding of Bhuiyan et al. (2018). According to a previous study over the Iberian Peninsula (Bhuiyan et al. 2018), satellite precipitation, reanalysis, and soil moisture, datasets were ranked as the most important predictor variables showing their strong impact in model prediction by producing higher values of percentage increase in mean square error, %IncMSE (0.3–0.6). In this study the four

satellite precipitation and soil moisture datasets have also high %IncMSE magnitude (0.4–0.6), which indicates the high level of influence of those predictors in the model prediction. Moreover, elevation and the reanalysis precipitation datasets were ranked as the most important predictor variables by showing higher %IncMSE (0.6–0.8), presenting a strong impact in model prediction. From the variable importance results the level of significance varies marginally among the different variables for the Peruvian Andes and Colombian Andes. Over the Blue Nile, higher %IncMSE is observed for almost all the predictor variables except GSMAP (V6) and 3B42(V7) compared to the other two regions. Based on the above significance results we included all variables in the MLPET blended precipitation product.

b. Error analysis

In this section, we used two error metrics (NCRMSE and BR) to evaluate model performances for the two validation experiments: *K*-fold applied at each region and leave-one-region-out. The NCRMSE is calculated for five reference precipitation ranges: <25th, 25th–50th, 50th–75th, 75th–95th, and >95th percentile values. Results are summarized in Fig. 3. In terms of NCRMSE values, MLPET exhibited considerably improved performances

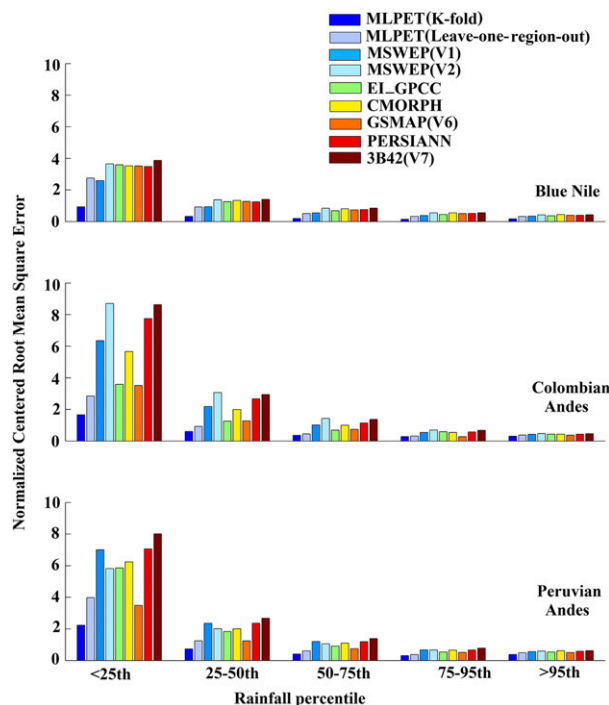


FIG. 3. Normalized centered root-mean-square error for the three study regions.

than other precipitation products over the three study regions. Especially, the magnitude of random error for MLPET is noticeably lower than individual precipitation products over the Blue Nile. Similarly, MLPET and GSMaP (V6) products exhibited very low NCRMSE values (0.25–0.4) for >75th percentile over the Colombian Andes and Peruvian Andes. It is noted that the gauge-adjusted 3B42 (V7) performance is worse than the rest of precipitation datasets.

To demonstrate the performance of MLPET in comparison to the individual precipitation datasets, the relative error metric difference (Δ_{error}) applied on NCRMSE was calculated for the different precipitation ranges (Table 2). The significant relative random error reduction (18%–81%) of MLPET with respect to individual satellite and reanalysis precipitation datasets is noted over the three regions for *K*-fold validation. Particularly, MLPET had significant reduction in random error (58%–78%) with respect to individual precipitation datasets over the Blue Nile. Over the Colombian and Peruvian Andes, the performance of MLPET generated almost similar relative random error reduction (18%–81%). The random error reduction of MLPET relative to MSWEP was also substantial for all regions: 29%–74% relative to MSWEP (V1) and 36%–81% relative to MSWEP (V2). It is noted that the MSWEP (V2) product performance is worse than the MSWEP (V1)

TABLE 2. Relative reduction of MLPET NCRMSE with respect to individual precipitation NCRMSE. Results are presented for different precipitation ranges and regions.

Study area	Rainfall percentile	K-fold									Leave-one-region-out								
		MSWEP			EL_GPCC			CMORPH			GSMAP			PERSIANN			3B42 (V7)		
		(V1)	(V2)	(V7)	(V1)	(V2)	(V7)	(V1)	(V2)	(V7)	(V1)	(V2)	(V7)	(V1)	(V2)	(V7)	(V1)	(V2)	(V7)
Blue Nile	<25th	64%	74%	76%	74%	74%	73%	73%	73%	76%	73%	74%	73%	73%	74%	74%	76%	77%	29%
	25–50th	65%	76%	77%	74%	74%	75%	74%	74%	77%	74%	74%	74%	74%	74%	74%	77%	34%	34%
	50–75th	65%	77%	77%	72%	72%	76%	76%	74%	77%	74%	74%	74%	74%	74%	74%	77%	40%	40%
	75–95th	61%	73%	73%	67%	67%	73%	73%	71%	74%	71%	71%	71%	72%	71%	71%	74%	40%	40%
	>95th	51%	60%	61%	54%	54%	62%	62%	58%	61%	58%	58%	58%	58%	58%	58%	61%	26%	26%
Colombian Andes	<25th	74%	81%	81%	54%	54%	71%	71%	71%	81%	53%	52%	52%	79%	79%	79%	81%	67%	67%
	25–50th	72%	80%	80%	52%	52%	70%	70%	70%	79%	52%	52%	52%	77%	77%	77%	79%	69%	69%
	50–75th	65%	75%	75%	48%	48%	64%	64%	64%	74%	52%	52%	52%	69%	69%	69%	74%	67%	67%
	75–95th	48%	59%	59%	53%	53%	48%	48%	48%	59%	–3%	–3%	–3%	51%	51%	51%	59%	54%	54%
	>95th	29%	36%	36%	30%	30%	31%	31%	31%	34%	18%	18%	18%	29%	29%	29%	34%	19%	19%
Peruvian Andes	<25th	68%	62%	62%	62%	62%	64%	64%	64%	72%	36%	36%	36%	68%	68%	68%	72%	44%	44%
	25–50th	68%	63%	63%	60%	60%	63%	63%	63%	72%	41%	41%	41%	69%	69%	69%	72%	53%	53%
	50–75th	65%	61%	61%	55%	55%	62%	62%	62%	70%	45%	45%	45%	65%	65%	65%	70%	56%	56%
	75–95th	54%	53%	53%	43%	43%	53%	53%	53%	61%	41%	41%	41%	54%	54%	54%	61%	52%	52%
	>95th	33%	36%	36%	30%	30%	38%	38%	38%	40%	25%	25%	25%	35%	35%	35%	40%	22%	22%

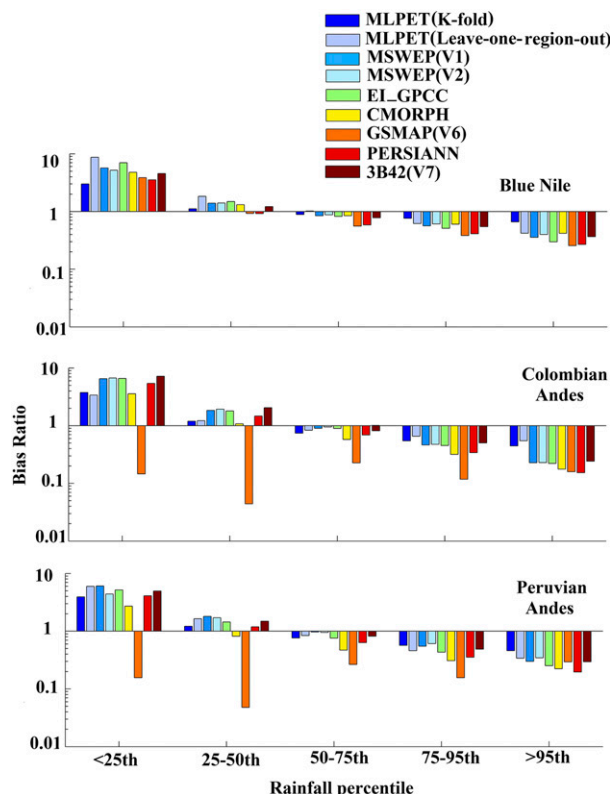


FIG. 4. As in Fig. 3, but for bias ratio.

product over the three study regions in terms of random error.

The leave-one-region-out validation experiment also demonstrated improvements of the MLPET against the different individual satellite and reanalysis products, with reduction in random error ranging between 3% and 69% for all precipitation ranges over the three regions. In addition, the reduction of random error of MLPET relative to MSWEP (V2) (18%–69%) was also prominent. Overall, among all individual precipitation products, the magnitude of relative random error reduction for the blended product was noticeably decreased for all regions.

Results on the evaluation of the systematic error, defined by BR, are shown in Fig. 4. The results indicate that the QRF technique applied in MLPET is able to reduce significantly the systematic error in moderate to high rain rates (>50th percentile). Specifically, for the very high rain rate category (>95th percentile), MLPET exhibited BR values closer to one than the individual precipitation datasets and MSWEP. The error metric difference (Δ_{error}) for the different precipitation datasets is presented in Table 3. For the K -fold experiment, the relative reduction of the systematic error by MLPET with respect to individual precipitation products was

TABLE 3. As in Table 2, but for mean relative difference.

Study area	Rainfall percentile	K -fold									Leave-one-region-out								
		MSWEP (V1)	MSWEP (V2)	EL_GPCC	CMORPH	GSMAP	PERSIANN	3B42 (V7)	MSWEP (V1)	MSWEP (V2)	EL_GPCC	CMORPH	GSMAP	PERSIANN	3B42 (V7)	MSWEP (V1)	MSWEP (V2)	EL_GPCC	CMORPH
Blue Nile	<25th	57%	52%	66%	47%	30%	21%	43%	64%	84%	28%	91%	97%	99%	93%	64%	84%	28%	91%
	25–50th	71%	71%	76%	63%	22%	18%	47%	11%	12%	25%	19%	39%	36%	31%	11%	12%	25%	19%
	50–75th	26%	5%	36%	26%	74%	72%	48%	87%	83%	88%	87%	95%	95%	91%	87%	83%	88%	87%
	75–95th	45%	39%	51%	40%	61%	60%	47%	13%	3%	22%	4%	38%	35%	16%	13%	3%	22%	4%
	>95th	47%	43%	51%	41%	54%	53%	46%	10%	3%	17%	1%	22%	21%	8%	10%	3%	17%	1%
Colombian Andes	<25th	50%	52%	51%	–6%	–22%	38%	55%	57%	58%	57%	8%	79%	46%	61%	57%	58%	57%	8%
	25–50th	77%	79%	76%	–15%	80%	58%	82%	74%	76%	73%	–71%	77%	53%	79%	74%	76%	73%	–71%
	50–75th	–16%	–43%	–49%	39%	67%	17%	9%	–65%	–23%	–57%	61%	79%	48%	18%	–65%	–23%	–57%	61%
	75–95th	15%	14%	17%	34%	49%	31%	27%	36%	35%	37%	50%	61%	47%	31%	36%	35%	37%	50%
	>95th	28%	28%	29%	33%	34%	35%	27%	41%	41%	42%	45%	46%	47%	40%	41%	41%	42%	45%
Peruvian Andes	<25th	42%	43%	30%	–72%	–95%	5%	26%	2%	3%	19%	–82%	–90%	–61%	–25%	2%	3%	19%	–82%
	25–50th	73%	68%	51%	–20%	77%	14%	55%	19%	17%	–45%	–67%	32%	–65%	–33%	19%	17%	–45%	–67%
	50–75th	–61%	–39%	1%	55%	68%	35%	–31%	–37%	–23%	33%	70%	78%	56%	12%	–37%	–23%	33%	70%
	75–95th	4%	2%	24%	38%	49%	33%	16%	–20%	–22%	5%	22%	36%	17%	–5%	–20%	–22%	5%	22%
	>95th	23%	17%	28%	30%	24%	33%	23%	6%	3%	12%	15%	7%	18%	6%	6%	3%	12%	15%

prominent for all three regions (1%–82%). The systematic error reduction relative to MSWEP products ranged between 5% and 71% over the Blue Nile and 2%–79% over the Peruvian and Colombian Andes, except for the 50th–75th percentile rain rates.

Similarly, for the leave-one-region-out experiment, systematic error reduction by MLPET was exhibited relative to individual precipitation products and MSWEP, except for the low rain rates (<50th percentile) over Blue Nile. All these results highlighted the usefulness of optimally combining satellite and reanalysis precipitation datasets through the QRF machine learning technique, and the significance of utilizing in situ data.

c. Evaluation of precipitation products

The normalized Taylor diagrams summarized the results for the different precipitation products. Figure 5 shows the results for the daily scale for all three regions. Results for K -fold based MLPET estimates are more consistent among the various precipitation products exhibiting low normalized root-mean-square difference (NRMSD) values, in the range of 0.25–0.75, for all the regions. Specifically, for the Blue Nile region, NRMSD reduced considerably ($\text{NRMSD} < 0.4$), showing a very high degree of agreement with reference precipitation. In addition, MLPET estimates based on the leave-one-region-out experiment, also performed comparatively better than other precipitation products including MSWEP by reducing NRMSD values (0.8–1.2) with normalized standard deviations in the range of 0.25–0.75. As shown in Fig. 5, the points for the satellite and reanalysis precipitation products were always the farthest from the reference ($\text{NRMSD} > 1$) with low correlation coefficient (0.1–0.3), which means satellite and reanalysis precipitation had comparatively poorer performances than MLPET estimates. Overall, results indicate that the MLPET performance is better than that of the other precipitation products, by exhibiting lower NRMSD and highest correlation coefficient (0.75–0.95) for all three tropical complex terrain regions. On the other hand, MLPET produced lower variability relative to the individual satellite, reanalysis and MSWEP datasets as shown by the lower normalized standard deviations (0.25–0.75), which indicates that, the method underdispersed precipitation variability in the three study regions.

In addition, KGE values for the different precipitation products are shown in Table 4. KGE considerably varied (from -0.64 to 0.60) for the different precipitation products as a function of magnitude. Specifically, MLPET based on the K -fold experiment is more prominent among the various precipitation products, exhibiting high KGE values in the range of 0.35–0.6 for all three tropical complex terrain regions. The other precipitation products

showed relatively lower performance with relatively low (<0.25) KGE values for the study areas.

d. Ensemble verification

Results for EP values are shown in Fig. 6. For lower rain rates (<50th percentile), small EP values (<0.1) were found for all three regions. Moreover, for the high rain rates (>95th percentile), relatively low exceedance probability values (<0.40) are reported, which suggests that reference precipitation is captured well within the ensemble envelope. It is noted that the leave-one-region-out validation experiment exhibited substantially reduced EP values (<0.25) for rain rates above the 95th percentile. This is attributed to the fact that the model was trained by datasets having larger size and longer dynamic range of values relative to the K -fold training that is based on local data. These results indicate the impact of training sample size on MLPET performances over the study areas. For example, over the Blue Nile, for the leave-one-region-out experiment, QRF was trained by 1 457 943 samples, which is 8 times greater than the K -fold trained sample sizes.

UR was used to assess the ensemble envelope width. Analysis of UR shows that the QRF ensemble envelope is associated with wider prediction intervals for lower rain rates (<50th percentile) (Fig. 7). An UR closer to one was exhibited for higher than the 50th percentile rain rates, demonstrating that the ensemble envelope represented the benchmark variability. The K -fold experiment exhibited marginally lower UR values than the leave-one-region-out experiment, demonstrating less variability in the ensemble envelop, explained by the sample size distribution of the K -fold training datasets. Specifically, in the QRF technique, the distribution of proper sample size with lower variability plays an important role in the empirical distribution function, which shows the measure of predictive performance as a function of a varying number of training samples.

Overall, our results show both low EP and UR magnitudes for moderate to high rain rates (>50th percentile), which represent an acceptable error characterization of the technique estimates over the three complex terrain areas. These ensemble verification statistics indicate that the model is capable of generating stochastic realizations that provide a useful estimation of the prediction uncertainty as it encapsulates well the reference precipitation while at the same time is characterized by an ensemble width that is low enough to maintain skill in the prediction.

e. Discussion

This study demonstrated that using the QRF technique for combining multiple satellite and reanalysis

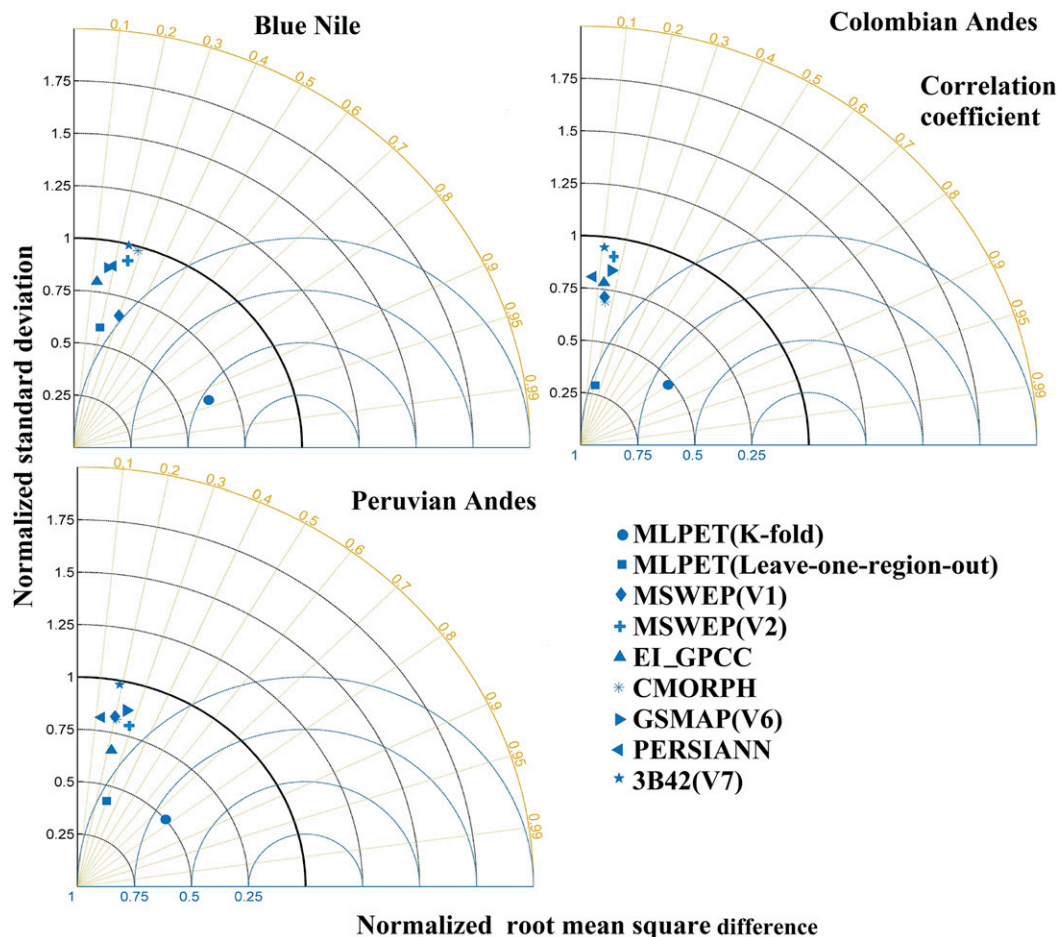


FIG. 5. Normalized Taylor diagrams for daily precipitation for the three study regions.

precipitation datasets over complex terrain along with other auxiliary variables (soil moisture, humidity and elevation) results in considerably lower random errors than those of the individual global precipitation datasets used in the technique. Over the study areas (Figs. 3 and 4), the performance of satellite precipitation was consistent with findings by other studies (Dinku et al. 2007, 2008; Derin et al. 2016). It is noted that the technique based on the leave-one-region-out cross validation (i.e., without using local data for calibration) did not exhibit improvements in terms of systematic error of low precipitation rates over the Blue Nile and Peruvian Andes. Although, in the leave-one-region-out experiment, QRF model was trained by a larger dataset, this was not well representative in terms of the distribution of zeros and low precipitation rates for these two study areas. On the other hand, the K -fold validation method (which uses local in situ station data) reduced greatly both the random and systematic error for rainfall rates exceeding the 75th percentile across all regions, which is vital for the quantification of flood inducing rain rates.

This indicates that local data help the method in terms of high and low values.

In addition, Fig. 5 showed that MLPET generated lower variability relative to the individual precipitation datasets. These results are consistent with findings from a recent study over the Iberian Peninsula (Bhuiyan et al. 2019) that investigated the propagation of precipitation uncertainty from MLPET and other precipitation products in hydrological simulations (surface and subsurface runoff and evapotranspiration), based on a number of global hydrological and land surface models. Results from that study showed that even though MLPET product underdispersed precipitation variability provided a significant improvement in hydrologic simulations compared to the other satellite and reanalysis precipitation datasets (Bhuiyan et al. 2019).

The model considered elevation as a control parameter, which displayed its ability to decrease the systematic and random error noticeably at complex terrain. Over the Iberian Peninsula, which has precipitation and climate variability due to complex orography influenced by both

TABLE 4. Results of KGE are presented for different regions, where all data points from all stations are used in the calculation.

Study area	MLPET (K-fold)	MLPET (Leave-one-region-out)	EI_GPCC	CMORPH	GSMAP	PERSIANN	3B42 (V7)	MSWEP (V1)	MSWEP (V2)
Blue Nile	0.60	0.10	0.08	0.26	0.05	0.08	0.20	0.20	0.24
Colombian Andes	0.35	0.07	0.06	−0.05	−0.64	−0.05	0.08	0.06	0.07
Peruvian Andes	0.38	0.05	0.08	0.01	−0.60	−0.01	0.13	0.15	0.23

Atlantic and Mediterranean climates, we had shown that MLPET provided significant improvements against individual precipitation products (Bhuiyan et al. 2018). Collective evidence from the Bhuiyan et al. (2018) and the herein multiregional evaluation study show that the QRF technique is reliable and robust for combining multiple precipitation estimates over complex terrain areas.

In terms of uncertainty quantification, we found low UR and EP values, which indicate acceptable characterization of estimation uncertainty. More specially, for the highest percentiles (>95th), EP values were relatively higher, which means that QRF ensemble predictions could not fully capture the reference precipitation at these value ranges. Sample size plays a significant role in empirical distribution function. Therefore, larger training datasets are required for the predictions of low and extremely high values to quantify the rate of convergence to the underlying cumulative distribution function, which was also highlighted in the reduction of EP from the K-fold to the leave-one-region-out cross-validation experiments.

5. Conclusions

The MLPET blended precipitation product evaluated in this study is based on the application of a non-parametric statistical technique, quantile regression forests (QRF), for blending multiple satellite and re-analysis precipitation dataset for complex terrain applications. To evaluate the regional accuracy of MLPET, the proposed framework was demonstrated over three mountainous tropical regions (Blue Nile, Peruvian Andes, and Colombian Andes) characterized by varying hydroclimatic characteristics. The study showed that MLPET significantly reduced random error, and this reduction exhibited rainfall magnitude dependence. Similarly, the systematic error of MLPET product varied from over to underestimation as rain rate increased over the three regions; however, this magnitude dependence was less prominent than the individual satellite precipitation products and MSWEP.

Overall, the estimation of high rain rates is challenging because of the complex terrain. It was shown that MLPET can significantly reduce both random and systematic errors of the highest rain rate category

(>95th percentile) for all three mountainous regions. Moreover, MLPET was shown to outperform the benchmark MSWEP products (V1 and V2) over the three tropical mountainous regions in terms of both random and systematic error metrics. Our overall finding is that the proposed framework offers a new way of combining globally available precipitation datasets, generating an improved ensemble precipitation product over complex terrain areas.

Accurate estimation of rainfall in ungauged areas is an essential component for efficient water management and future planning of water resources systems. The presented leave-one-region-out experiment has demonstrated that one can develop MLPET technique over regions where ground reference data exist and apply it over similar hydroclimatic regions that are data poor or ungauged. This leave-one-region-out experiment needs to be extended to include more regions for testing the transferability of the technique's calibration from other complex terrain regions. Such independent validation experiment could demonstrate the robustness of the framework, where complex topography and regional climate characteristics play a vital role in controlling the spatial pattern of precipitation.

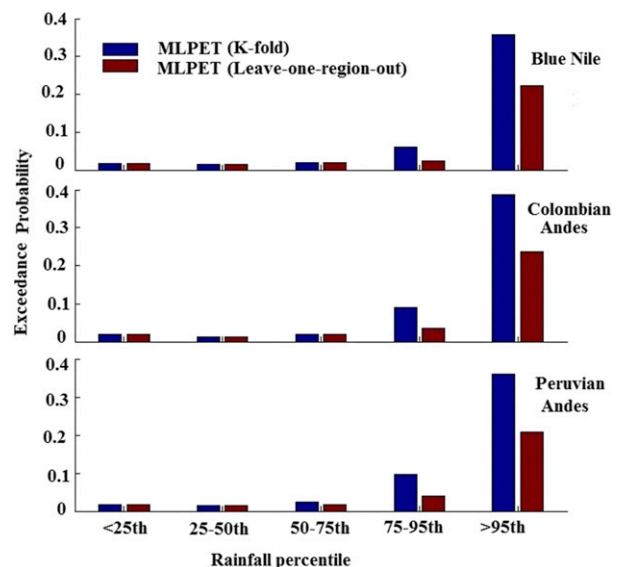


FIG. 6. Exceedance probability determined for different precipitation ranges.

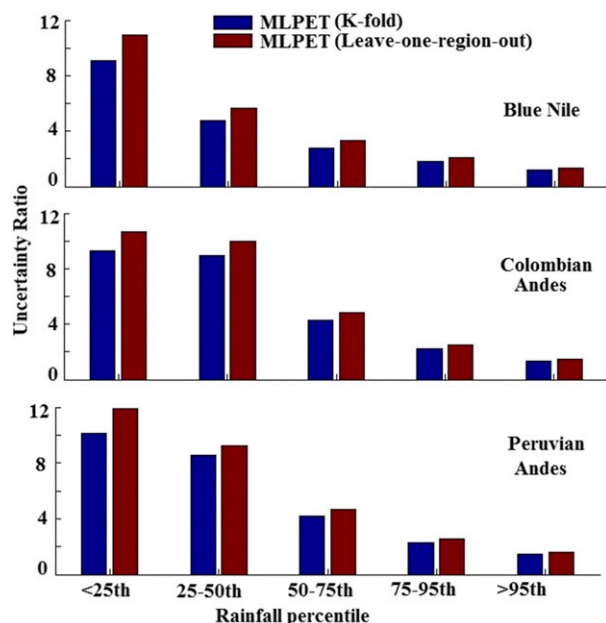


FIG. 7. As in Fig. 6, but for uncertainty ratio.

This research has demonstrated the premise of nonparametric statistical techniques in global precipitation estimation, but also highlighted challenges that need to be addressed. Specific future directions of the MLPET technique will include the newly available IMERG products and other recent satellite precipitation datasets. In that respect, it is needless to say that MLPET or any other blending technique depends on the characteristics of the individual precipitation products and thus such a technique is only meant to provide a way of optimally integrating the existing information in the various individual precipitation products. Given the great challenge of precipitation estimation over steep orography, a future study could focus on deriving a global mountainous MLPET dataset to support hydrologic modeling and water resource reanalysis studies in complex terrain areas.

Acknowledgments. This study was supported by the Eversource Energy Center at the University of Connecticut. The gauge derived rainfall dataset from Blue Nile is provided by Dejene Sahlu and Semu Moges. The people who contributed both Colombian and Peruvian datasets are Bastian Manz, Waldo Lavado, Wouter Buytaert, Vergara Arrieta, Humberto J. Vergara, and Juan-Pablo Rodríguez-Sánchez.

REFERENCES

Álvarez-Villa, O. D. A., J. I. Velez, and G. Poveda, 2011: Improved long-term mean annual rainfall fields for Colombia. *Int. J. Climatol.*, **31**, 2194–2212, <https://doi.org/10.1002/joc.2232>.

- Anders, A. M., and S. W. Nesbitt, 2015: Altitudinal precipitation gradients in the tropics from Tropical Rainfall Measuring Mission (TRMM) precipitation radar. *J. Hydrometeorol.*, **16**, 441–448, <https://doi.org/10.1175/JHM-D-14-0178.1>.
- Balsamo, G., and Coauthors, 2015: ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.*, **19**, 389–407, <https://doi.org/10.5194/hess-19-389-2015>.
- Beck, H. E., A. I. J. M. van Dijk, A. de Roo, D. G. Miralles, T. R. McVicar, J. Schellekens, and L. A. Bruijnzeel, 2016: Global-scale regionalization of hydrologic model parameters. *Water Resour. Res.*, **30**, 791–815, <https://doi.org/10.1002/2015WR018247>.
- , and Coauthors, 2017a: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.*, **21**, 6201–6217, <https://doi.org/10.5194/hess-21-6201-2017>.
- , A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, and A. de Roo, 2017b: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2014) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, **21**, 589–615, <https://doi.org/10.5194/hess-21-589-2017>.
- , E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- Bhuiyan, M. A. E., E. N. Anagnostou, and P. E. Kieswetter, 2017: A nonparametric statistical technique for modeling overland TMI (2A12) rainfall retrieval error. *IEEE Geosci. Remote Sens. Lett.*, **14**, 1898–1902, <https://doi.org/10.1109/LGRS.2017.2728658>.
- , E. I. Nikolopoulos, E. N. Anagnostou, P. Quintana-Seguí, and A. Barella-Ortiz, 2018: A nonparametric statistical technique for combining global precipitation datasets: development and hydrological evaluation over the Iberian Peninsula. *Hydrol. Earth Syst. Sci.*, **22**, 1371–1389, <https://doi.org/10.5194/hess-22-1371-2018>.
- , and Coauthors, 2019: Assessment of precipitation error propagation in multi-model global water resource reanalysis. *Hydrol. Earth Syst. Sci.*, **23**, 1973–1994, <https://doi.org/10.5194/hess-23-1973-2019>.
- Block, P., and B. Rajagopalan, 2007: Interannual variability and ensemble forecast of Upper Blue Nile Basin Kiremt season precipitation. *J. Hydrometeorol.*, **8**, 327–343, <https://doi.org/10.1175/JHM580.1>.
- Bookhagen, B., and M. R. Strecker, 2008: Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes. *Geophys. Res. Lett.*, **35**, L06403, <https://doi.org/10.1029/2007GL032011>.
- Bosilovich, M. G., J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.*, **47**, 2279–2299, <https://doi.org/10.1175/2008JAMC1921.1>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Cattani, E., A. Merino, and V. Levizzani, 2016: Evaluation of monthly satellite-derived precipitation products over East Africa. *J. Hydrometeorol.*, **17**, 2555–2573, <https://doi.org/10.1175/JHM-D-15-0042.1>.
- , —, J. A. Gujarró, and V. Levizzani, 2018: East Africa rainfall trends and variability 1983–2015 using three long-term satellite products. *Remote Sens.*, **10**, 931, <https://doi.org/10.3390/rs10060931>.

- Cau, P., J. Methven, and B. Hoskins, 2007: Origins of dry air in the tropics and subtropics. *J. Climate*, **20**, 2745–2759, <https://doi.org/10.1175/JCLI4176.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Derin, Y., and Coauthors, 2016: Multiregional satellite precipitation products evaluation over complex terrain. *J. Hydrometeorol.*, **17**, 1817–1836, <https://doi.org/10.1175/JHM-D-15-0197.1>.
- Dezfuli, A. K., C. M. Ichoku, G. J. Huffman, K. I. Mohr, J. S. Selker, N. van de Giesen, R. Hochreutener, and F. O. Annor, 2017: Validation of IMERG precipitation in Africa. *J. Hydrometeorol.*, **18**, 2817–2825, <https://doi.org/10.1175/JHM-D-17-0139.1>.
- Dinku, T., P. Ceccato, E. Grover-Kopce, M. Lemma, S. J. Connor, and C. F. Ropelewski, 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, **28**, 1503–1526, <https://doi.org/10.1080/01431160600954688>.
- , S. Chidzambwa, P. Ceccato, S. J. Connor, and C. F. Ropelewski, 2008: Validation of high-resolution satellite rainfall products over complete terrain. *Int. J. Remote Sens.*, **29**, 4097–4110, <https://doi.org/10.1080/01431160701772526>.
- Duque-Gardeazabal, N., D. Zamora, and E. Rodríguez, 2018: Analysis of the kernel bandwidth influence in the double smoothing merging algorithm to improve rainfall fields in poorly gauged basins. *HIC 2018: 13th Int. Conf. on Hydroinformatics*, Vol. 3, Palermo, Italy, University of Palermo, 635–642.
- Enyew, B. D., and G. J. Steeneveld, 2014: Analysing the impact of topography on precipitation and flooding on the Ethiopian highlands. *J. Geol. Geosci.*, **3**, 2, <https://doi.org/10.4172/2329-6755.1000173>.
- Espinoza, J. C., S. Chavez, J. Ronchail, C. Junquas, K. Takahashi, and W. Lavado, 2015: Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation. *Water Resour. Res.*, **51**, 3459–3475, <https://doi.org/10.1002/2014WR016273>.
- Espinoza Villar, J. C., and Coauthors, 2009: Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *Int. J. Climatol.*, **29**, 1574–1594, <https://doi.org/10.1002/joc.1791>.
- Formetta, G., R. Mantilla, S. Franceschi, A. Antonello, and R. Rigon, 2011: The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: Models of flow generation and propagation/routing. *Geosci. Model Dev.*, **4**, 943–955, <https://doi.org/10.5194/gmd-4-943-2011>.
- Gao, L., M. Bernhardt, and K. Schulz, 2012: Elevation correction of ERA-Interim temperature data in complex terrain. *Hydrol. Earth Syst. Sci.*, **16**, 4661–4673, <https://doi.org/10.5194/hess-16-4661-2012>.
- , —, —, and X. W. Chen, 2017: Elevation correction of ERA-Interim temperature data in the Tibetan Plateau. *Int. J. Climatol.*, **37**, 3540–3552, <https://doi.org/10.1002/joc.4935>.
- Gottschalk, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation products and preliminary assessment of their impact on Global Land Data Assimilation System land surface states. *J. Hydrometeorol.*, **6**, 573–598, <https://doi.org/10.1175/JHM437.1>.
- Greco, M., W. S. Olson, S. J. Munchak, S. Ringerud, L. Liao, Z. Haddad, B. L. Kelley, and S. F. McLaughlin, 2016: The GPM combined algorithm. *J. Atmos. Oceanic Technol.*, **33**, 2225–2245, <https://doi.org/10.1175/JTECH-D-16-0019.1>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Heidinger, H., C. Yarlequé, A. Posadas, and R. Quiroz, 2012: TRMM rainfall correction over the Andean Plateau using wavelet multi-resolution analysis. *Int. J. Remote Sens.*, **33**, 4583–4602, <https://doi.org/10.1080/01431161.2011.652315>.
- Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Houze, R. A., Jr., 2012: Orographic effects on precipitating clouds. *Rev. Geophys.*, **50**, RG1001, <https://doi.org/10.1029/2011RG000365>.
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 3–22.
- , D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.5, 30 pp., http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kubota, T., and Coauthors, 2007: Global precipitation map using satellite-borne microwave radiometers by the GSMap project: Production and validation. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2259–2275, <https://doi.org/10.1109/TGRS.2007.895337>.
- Li, M., and Q. Shao, 2010: An improved statistical approach to merge satellite rainfall estimates and raingauge data. *J. Hydrol.*, **385**, 51–64, <https://doi.org/10.1016/j.jhydrol.2010.01.023>.
- Li, N., G. Tang, P. Zhao, Y. Hong, Y. Gou, and K. Yang, 2017: Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin. *Atmos. Res.*, **183**, 212–223, <https://doi.org/10.1016/j.atmosres.2016.07.020>.
- Liu, Z. J., Y. S. Liu, S. S. Wang, X. J. Yang, L. C. Wang, M. H. A. Baig, W. F. Chi, and Z. S. Wang, 2018: Evaluation of spatial and temporal performances of ERA-Interim precipitation and temperature in Mainland China. *J. Climate*, **31**, 4347–4365, <https://doi.org/10.1175/JCLI-D-17-0212.1>.
- Luo, H., F. Ge, K. Yang, S. Zhu, T. Peng, W. Cai, X. Liu, and W. Tang, 2019: Assessment of ECMWF reanalysis data in complex terrain: can the CERA-20C and ERA-Interim datasets replicate the variation in surface air temperatures over Sichuan, China? *Int. J. Climatol.*, **2019**, 1–16, <https://doi.org/10.1002/joc.6175>.
- Ly, S., C. Charles, and A. Degré, 2011: Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrol. Earth Syst. Sci.*, **15**, 2259–2274, <https://doi.org/10.5194/hess-15-2259-2011>.
- Ma, Y., G. Tang, D. Long, B. Yong, L. Zhong, W. Wan, and Y. Hong, 2016: Similarity and error intercomparison of the GPM and its predecessor—TRMM multisatellite precipitation analysis using the best available hourly gauge network over the

- Tibetan Plateau. *Remote Sens.*, **8**, 569, <https://doi.org/10.3390/rs8070569>.
- Manz, B., W. Buytaert, Z. Zulkafli, W. Lavado, B. Willems, L. A. Robles, and J.-P. Rodríguez-Sánchez, 2016: High-resolution satellite-gauge merged precipitation climatologies of the tropical Andes. *J. Geophys. Res. Atmos.*, **121**, 1190–1207, <https://doi.org/10.1002/2015JD023788>.
- Mazzetti, C., and E. Todini, 2004: Combining raingages and radar precipitation measurements using a Bayesian approach. *geo-ENV IV—Geostatistics for Environmental Applications*, X. Sanchez-Vila, J. Carrera, and J. J. Gómez-Hernández, Eds., Quantitative Geology and Geostatistics, Vol. 13, Springer, 401–412, https://doi.org/10.1007/1-4020-2115-1_34.
- Mei, Y., E. N. Anagnostou, E. I. Nikolopoulos, and M. Borga, 2014: Error analysis of satellite rainfall products in mountainous basins. *J. Hydrometeorol.*, **15**, 1778–1793, <https://doi.org/10.1175/JHM-D-13-0194.1>.
- , —, —, and —, 2016: Evaluating satellite precipitation error propagation in runoff simulations of mountainous basins. *J. Hydrometeorol.*, **17**, 1407–1423, <https://doi.org/10.1175/JHM-D-15-0081.1>.
- Meinshausen, N., 2006: Quantile regression forests. *J. Mach. Learn. Res.*, **7**, 983–999.
- Mooney, P. A., F. J. Mulligan, and R. Fealya, 2011: Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *Int. J. Climatol.*, **31**, 545–557, <https://doi.org/10.1002/joc.2098>.
- Ning, S., F. Song, P. Udmale, J. Jin, B. R. Thapa, and H. Ishidaira, 2017: Error analysis and evaluation of the latest GSMaP and IMERG precipitation products over Eastern China. *Adv. Meteorol.*, **2017**, 1803492, <https://doi.org/10.1155/2017/1803492>.
- Peel, M. C., B. L. Finlayson, and T. A. McMahon, 2007: Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, **11**, 1633–1644, <https://doi.org/10.5194/hess-11-1633-2007>.
- Peña-Arancibia, J. L., A. I. van Dijk, L. J. Renzullo, and M. Mulligan, 2013: Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and South and East Asia. *J. Hydrometeorol.*, **14**, 1323–1333, <https://doi.org/10.1175/JHM-D-12-0132.1>.
- Rodell M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Roe, G. H., 2005: Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, **33**, 645–671, <https://doi.org/10.1146/annurev.earth.33.092203.122541>.
- Sahlu, D., E. I. Nikolopoulos, S. A. Moges, E. N. Anagnostou, and D. Hailu, 2016: First evaluation of the Day-1 IMERG over the upper Blue Nile basin. *J. Hydrometeorol.*, **17**, 2875–2882, <https://doi.org/10.1175/JHM-D-15-0230.1>.
- Schellekens, J., and Coauthors, 2017: A global water resources ensemble of hydrological models: The earth2Observe Tier-1 dataset. *Earth Syst. Sci. Data*, **9**, 389–413, <https://doi.org/10.5194/essd-9-389-2017>.
- Seyyedi, H., E. N. Anagnostou, P. E. Kirstetter, V. Maggioni, Y. Hong, and J. J. Gourley, 2014: Incorporating surface soil moisture information in error modeling of TRMM passive Microwave rainfall. *IEEE Trans. Geosci. Remote Sens.*, **52**, 6226–6240, <https://doi.org/10.1109/TGRS.2013.2295795>.
- , —, E. Beighley, and J. McCollum, 2015: Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. *Atmos. Res.*, **164–165**, 37–48, <https://doi.org/10.1016/j.atmosres.2015.03.019>.
- Sinclair, S., and G. Pegram, 2005: Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Lett.*, **6**, 19–22, <https://doi.org/10.1002/asl.85>.
- Smith, E. A., and Coauthors, 2007: International Global Precipitation Measurement (GPM) program and mission: An overview. *Measuring Precipitation from Space*, V. Levizzani, P. Bauer, and F. J. Turk, Eds., Springer, 611–653, https://doi.org/10.1007/978-1-4020-5835-6_48.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSS>2.3.CO;2).
- Tang, G., Y. Ma, D. Long, L. Zhong, and Y. Hong, 2016: Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *J. Hydrol.*, **533**, 152–167, <https://doi.org/10.1016/j.jhydrol.2015.12.008>.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, <https://doi.org/10.1029/2000JD900719>.
- Todini, E., 2001: A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrol. Earth Syst. Sci.*, **5**, 187–199, <https://doi.org/10.5194/hess-5-187-2001>.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, <https://doi.org/10.1256/qj.04.176>.
- Ushio, T., T. Matsuda, T. Tashima, T. Kubota, M. Kachi, and S. Yoshida, 2013: Gauge adjusted global satellite mapping of precipitation (GSMaP_Gauge). *Proc. 29th Int. Symp. on Space Technology and Science*, Nagoya, Japan, JAXA, 48.
- Wang, S. J., M. J. Zhang, M. P. Sun, B. L. Wang, X. Y. Huang, Q. Wang, and F. Feng, 2015: Comparison of surface air temperature derived from NCEP/DOE R2, ERA-Interim, and observations in the arid northwestern China: A consideration of altitude errors. *Theor. Appl. Climatol.*, **119**, 99–111, <https://doi.org/10.1007/s00704-014-1107-1>.
- Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo, 2014: The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.*, **50**, 7505–7514, <https://doi.org/10.1002/2014WR015638>.
- Woldemeskel, F. M., B. Sivakumar, and A. Sharma, 2013: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *J. Hydrol.*, **499**, 167–176, <https://doi.org/10.1016/j.jhydrol.2013.06.039>.
- Xu, S., Y. Shen, and Z. Du, 2016: Tracing the source of the errors in hourly IMERG using a decomposition evaluation scheme. *Atmosphere*, **7**, 161, <https://doi.org/10.3390/atmos7120161>.
- Yamamoto, M. K., and S. Shige, 2015: Implementation of an orographic/nonorographic rainfall classification scheme in the GSMaP algorithm for microwave radiometers. *Atmos. Res.*, **163**, 36–47, <https://doi.org/10.1016/j.atmosres.2014.07.024>.
- , —, C.-K. Yu, and L.-W. Cheng, 2017: Further improvement of the heavy orographic rainfall retrievals in the GSMaP algorithm for microwave radiometers. *J. Appl. Meteor. Climatol.*, **56**, 2607–2619, <https://doi.org/10.1175/JAMC-D-16-0332.1>.