Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products based on GPM DPR

Zhe Li[1], Daniel Wright[1], Samantha Hartke[1], Dalia Kirschbaum[2], Sana Khan[3], Viviana Maggioni[4], and Pierre-Emmanuel Kirstetter[5]

[1]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, WI, USA
[2]NASA
[3]NASA Goddard Space Flight Center
[4]George Mason University
[5]University of Oklahoma, National Severe Storms Laboratory

November 26, 2022

Abstract

The usefulness of satellite multisensor precipitation products such as NASA's 30-minute, 0.1° Integrated Multi-satellitE Retrievals for the Global Precipitation Mission (IMERG) is hindered by their associated errors. Reliable estimates of uncertainty would mitigate this limitation, especially in near-real time. Creating such estimates is challenging, however, due both to the complex discrete-continuous nature of satellite precipitation errors and to the lack of "ground truth" data precisely in the places—including complex terrain and developing countries—that could benefit most from satellite precipitation estimates. In this work, we use swath-based precipitation products from the Global Precipitation Mission (GPM) Dual-frequency Precipitation Radar (DPR) as an alternative to ground-based observations to facilitate IMERG uncertainty estimation. We compare the suitability of two DPR derived products, 2ADPR and 2BCMB, against higher-fidelity Ground Validation Multi-Radar Multi-Sensor (GV-MRMS) ground reference data over the contiguous United States. 2BCMB is selected to train mixed discrete-continuous error models based on Censored Shifted Gamma Distributions. Uncertainty estimates from these error models are compared against alternative models trained on GV-MRMS. Using information from NASA's Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis, we also demonstrate how IMERG uncertainty estimates can be further constrained using additional precipitation-related predictors. Though several critical issues remain unresolved, the proposed method shows promise for yielding robust uncertainty estimates in near-real time for IMERG and other similar precipitation products at their native resolution across the entire globe.

Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products based on GPM DPR

Zhe Li[1], Daniel B. Wright[1], Samantha H. Hartke[1], Dalia B. Kirschbaum[2], Sana Khan[2,3], Viviana Maggioni[4], Pierre-Emmanuel Kirstetter[5,6]

[1]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, WI, USA

[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA

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

[4]Civil, Infrastructure and Environmental Engineering Department, George Mason University, Fairfax, VA, USA

[5]School of Meteorology and School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA

[6]NOAA/National Severe Storms Laboratory, Norman, OK, USA

Version: JUNE 07, 2021

Corresponding author: Dr. Zhe Li (zli875@wisc.edu)

Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1221 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA

Key Points:

- We propose a globally-applicable uncertainty quantification framework for satellite precipitation products at their native resolution
- The framework performs well over the contiguous United States for according to both deterministic and probabilistic evaluation metrics
- The framework's uncertainty estimates can be further constrained using additional precipitation-related predictors

Abstract

The usefulness of satellite multisensor precipitation products such as NASA's 30-minute, 0.1° Integrated Multi-satellitE Retrievals for the Global Precipitation Mission (IMERG) is hindered by their associated errors. Reliable estimates of uncertainty would mitigate this limitation, especially in near-real time. Creating such estimates is challenging, however, due both to the complex discrete-continuous nature of satellite precipitation errors and to the lack of "ground truth" data precisely in the places—including complex terrain and developing countries—that could benefit most from satellite precipitation estimates. In this work, we use swath-based precipitation products from the Global Precipitation Mission (GPM) Dual-frequency Precipitation Radar (DPR) as an alternative to ground-based observations to facilitate IMERG uncertainty estimation. We compare the suitability of two DPR derived products, 2ADPR and 2BCMB, against higher-fidelity Ground Validation Multi-Radar Multi-Sensor (GV-MRMS) ground reference data over the contiguous United States. 2BCMB is selected to train mixed discrete-continuous error models based on Censored Shifted Gamma Distributions. Uncertainty estimates from these error models are compared against alternative models trained on GV-MRMS. Using information from NASA's Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis, we also demonstrate how IMERG uncertainty estimates can be further constrained using additional precipitation-related predictors. Though several critical issues remain unresolved, the proposed method shows promise for yielding robust uncertainty estimates in near-real time for IMERG and other similar precipitation products at their native resolution across the entire globe.

1 Introduction

The potential of satellite precipitation estimates to understand and predict global-to-regional water cycles has been recognized for decades (Kidd et al., 2020; Lettenmaier et al., 2015; Skofronick-Jackson et al., 2018). Due to the limited number and uneven distribution of rain gauges that accurately measure precipitation on the ground (e.g., Kidd et al., 2017), global satellite multi-sensor precipitation (SMP) products have been increasingly applied to support decision-making, particularly in data-sparse regions such as the oceans, mountainous areas, and developing countries (e.g., Kirschbaum et al., 2017; Wright, 2018).

SMP products generally merge measurements from passive microwave (PMW) and infrared (IR) sensors to create consistent high-resolution gridded precipitation estimates (Li et al., 2020; Maggioni et al., 2016; Sun et al., 2018). A number of global SMP products have been developed based on different merging techniques, including the NASA's Integrated Multisatellite Retrievals for Global Precipitation Measurements (IMERG; Huffman et al., 2019), and its predecessor—the Tropical Rainfall Measuring Mission Multisatellite Precipitation Analysis (TMPA; Huffman et al., 2007), the Climate Prediction Center morphing technique (CMORPH; Joyce et al., 2004; Joyce & Xie, 2011), and the Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) family (Nguyen et al., 2018; Sorooshian et al., 2000).

Despite continual improvements, the usefulness of SMP products remains limited due to their oftentimes poor accuracy (e.g., Foufoula-Georgiou et al., 2020; Massari & Maggioni, 2020). These errors stem from a variety of sources, including heterogeneous sensor properties (Guilloteau et al., 2017; Tan et al., 2016), retrieval algorithm deficiencies (Kirstetter et al., 2020), and insufficient spatial and temporal sampling (Behrangi & Wen, 2017; Kidd & Maggioni, 2020). The absolute and relative roles of these error sources can depend on season, precipitation intensity, storm type, geophysical features such as latitude and land surface type, and other factors (Ebert et al., 2007; Gebregiorgis & Hossain, 2014; Gebregiorgis et al., 2017; Kirstetter et al., 2018).

A large number of existing studies have presented empirical characterizations of SMP error using "ground truth", i.e., more reliable reference observations (typically rain gauges or gauge-corrected weather radar; see Kirstetter et al., 2012; Massari & Maggioni, 2020 for a discussion). Errors are often separated into systematic (i.e., bias) and random error components (AghaKouchak et al., 2012; Tang, 2020; Tian et al., 2013). These errors typically depend on precipitation magnitude via conditional bias (heteroscedasticity) in the case of systematic (random) error (e.g., Massari & Maggioni, 2020). Other approaches have considered additional terms to characterize errors in both detection and magnitude estimation, distinguishing between "false alarm" precipitation, missed precipitation, and hit bias (e.g., Tian et al., 2009). These studies have always been undertaken at local to regional scales due to the lack of sufficient ground reference globally (e.g., Beck et al., 2019; Li et al., 2013; O et al., 2017; Tang et al., 2016). Unfortunately, however, lessons learned in such studies cannot be easily transferred to other places due to the complexity of satellite precipitation uncertainties (Kidd & Maggioni, 2020; Tang & Hossain, 2012).

Furthermore, ex-post SMP error studies are not sufficient to meet uncertainty characterization requirements for applications, particularly those in near-realtime. Such a requirement has recently been prioritized by the IMERG development team—specifically, to provide uncertainty estimates at IMERG's native 30-minute, 0.1° resolution and at the time of creating IMERG data files (Huffman et al., 2019; Jackson Tan, personal communication, 30 December 2020).

A more limited number of studies have sought to develop so-called error models that attempt to characterize the uncertainty associated with any particular SMP product, generally expressed in the form of a probability distribution of "true precipitation" (e.g., Sarachi et al., 2015; Wright et al., 2017). Error model development is challenging due in part to the mixed discrete-continuous nature of intermittent precipitation, an issue that becomes increasingly important to address as SMP products advance to higher spatial and temporal resolutions. Some error models just ignore intermittency altogether to focus on hit biases and random errors (e.g., Sarachi et al., 2015; Tian et al., 2013), while others have attempted to address it (e.g., Gebremichael et al., 2011; Hossain & Anagnostou, 2006; Maggioni et al., 2014a), but arguably at the expense of relatively complicated formulations and limited flexibility (Wright et al., 2017). An alternative approach has been also proposed to characterize uncertainty as an integral part of SMP retrieval algorithms, and to subsequently yield probabilistic precipitation estimates (Kirstetter et al., 2018).

Regardless of the specific error model formulation, the availability of ground reference data to train these models has posed a fundamental limitation, since reference measurements are lacking precisely in the locations (i.e., data-sparse regions) that could benefit most from spaceborne remote sensing (e.g., Gebregiorgis & Hossain, 2014). It is thus highly desired to explore universal uncertainty quantification approaches that can perform anywhere, even in the total absence of local or regional ground reference observations.

In this study, we explore the idea that the Dual-frequency Precipitation Radar (DPR) on board the Global Precipitation Measurement (GPM) core observatory—the most accurate spaceborne precipitation measurement instrument to date—can be utilized in place of ground reference data. If valid, this facilitates the development of worldwide native-resolution error estimates of IMERG. Recent studies have explored the potential for DPR as an alternative reference to evaluate PMW-only precipitation estimates (e.g., Adhikari et al., 2019; You et al., 2020), and merged precipitation products (Khan et al., 2018). This study advances that concept to propose a prototype uncertainty quantification framework for IMERG. We use two DPR derived products and co-located IMERG estimates to train a parsimonious mixed discrete-continuous error model. These DPR-trained error models are evaluated against alternative models trained on ground reference observations over the contiguous United States (CONUS). The error model can also incorporate additional predictors. We examine whether a NASA reanalysis dataset can further constrain IMERG uncertainties. As far as we are aware, this is the first study to explore the feasibility of a globally-applicable prototype framework for quantification at the IMERG native resolution, though we leave global validation and several other important details to future work.

The datasets used in this study are described in Section 2. The data resampling and matching algorithm, the error model, and evaluation metrics are introduced in Section 3. Section 4 presents the results; discussion follows in Section 5. A summary and conclusions are provided in Section 6.

2 Data

We selected CONUS as the study area (Figure 1) for two reasons: firstly, it is covered by a high-quality, high-resolution NASA-sanctioned ground reference precipitation product that allows us to validate the proposed approach; secondly, its large geographic extent and climatic diversity allows a relatively comprehensive assessment of the approach's robustness. The study period is June 2014 to April 2019 (~ 5years). No attempt is made to address seasonally-varying uncertainty, nor to discriminate by precipitation phase. Prior studies have argued that the former may not be critical (Maggioni et al., 2014b; Wright et al., 2017), while the latter certainly is.

2.1 IMERG

IMERG merges all available PMW estimates with IR observations to produce 30-minute, 0.1° gridded precipitation estimates over the entire globe (Huffman et al., 2020; Tan et al., 2016). Three variants—Early (hereafter IMERG-E), Late and Final—address different user requirements for latency and accuracy. This study focuses on version 06B IMERG-E (Huffman et al., 2019), which is arguably the most useful for realtime applications due to its short latency (4 hours for IMERG-E, compared to 12 hours and 2 months for Late and Final, respectively) but features the largest errors due to the more limited availability of short-latency satellite and ground observations.

While the IMERG processing algorithm consists of many elements beyond the scope of this study, it is worth mentioning that it uses observations from the DPR and GPM Microwave Imager (GMI) on board the GPM core observatory. Microwave radiances from all partner constellation PMW sensors are calibrated to GMI for a bias-corrected, consistent radiometric dataset before retrieving precipitation rates (Hou et al., 2014). Then, the combined DPR and GMI data product from the GPM Combined Radar–Radiometer algorithm (CORRA; Grecu et al., 2016) contributes to IMERG in terms of its derived hydrometeor profiles and surface precipitation. The former is used to construct a-priori hydrometeor databases in the Goddard profiling algorithm (GPROF; Kummerow et al., 2015; Randel et al., 2020) to convert the calibrated PMW radiances into precipitation, while the latter is used to calibrate those PMW-only precipitation estimates on a rolling 45-day basis over ocean (the calibration is based on the Global Precipitation Climatology Project data over land; Huffman et al., 2019). We mention this because it constitutes a potential objection to the usage of DPR (and, as the reader will see, GMI) as the reference for uncertainty estimation due to a possible lack of independence between IMERG and those instruments. This issue is discussed further in Section 5.1.

2.2 Ground Reference: GV-MRMS

The Ground Validation Multi-radar/Multi-Sensor (GV-MRMS; Kirstetter et al., 2012, 2018) dataset is derived from the MRMS system that combines the polarimetric WSR-88D CONUS radar network with rain gauges and other auxiliary information to generate high-resolution quantitative precipitation estimates (QPE) over CONUS (Zhang et al., 2016). GV-MRMS QPE has been used as a ground reference for evaluation of various satellite precipitation products (Gebregiorgis et al., 2018; Kirstetter et al., 2012, 2014, 2020; O & Kirstetter, 2018). In this study, we use the Level-3 regridded GV-MRMS QPE product, which was created specifically to support GPM ground validation (Kirstetter et al., 2020). This product includes a 30-minute, 0.01° gauge-corrected precipitation rate (GCP) as well as a radar quality index (RQI) which ranges from 0 to 100, with 100 representing the best quality.

2.3 DPR-based Reference Datasets

We consider two recent (version 06) GPM Level-2 DPR products as potential alternatives to a ground-based reference: 2ADPR and 2BCMB. Both provide high-resolution (approximately 5 km DPR footprint diameter at nadir) precipitation estimates on an instantaneous field of view basis between 65ºN and 65ºS. 2ADPR is derived based on Ku (13.6 GHz) and Ka (35.5 GHz) band DPR measurements and it uses dual-frequency observations to infer precipitation phase and reconstruct three-dimensional hydrometeor and precipitation fields (Iguchi, 2020; Iguchi et al., 2018). This study uses the 2ADPR data field "precipRateESurface", which is extrapolated from the lowest clutter-free DPR bin to estimate surface precipitation rate (Petracca et al., 2018). 2BCMB, on the other hand, combines DPR reflectivities and GMI radiances using the CORRA algorithm to offer the highest-quality precipitation estimates from spaceborne sensors (Hou et al., 2014). We use the 2BCMB data field "surfPrecipTotRate" in the following analysis.

Both 2ADPR and 2BCMB data fields are obtained from normal scans (i.e., the widest swath scans from DPR; Iguchi et al., 2018) to maximize sample size. While 2BCMB and 2ADPR present different error structures (Gatlin et al., 2020), post-launch evaluations showed that DPR and GMI can detect precipitation rates down to 0.1 mm h^{-1} (e.g., Adhikari et al., 2019; Hamada & Takayabu, 2016). This precipitation rate was thus selected as the rain/no-rain detection threshold in this study for all datasets.

2.4 Additional Predictors: MERRA-2

We also examine the potential to further constrain the uncertainty estimates by incorporating additional predictors such as the total precipitable water vapor (TQV), topmost soil layer's ground wetness index ($GWET_{TOP}$), and 2-m air temperature (T2M) from NASA's MERRA-2 reanalysis product (Gelaro et al., 2017). To match with the above datasets, this study uses 0.5° (latitude) ×0.625° (longitude), hourly MERRA-2 outputs. $GWET_{TOP}$ is a dimensionless relative saturation index for the upper 5 cm of soil. Based on previous studies showing that soil moisture changes can enhance satellite precipitation estimation (e.g., Brocca et al., 2014; Crow et al., 2011), we also derive a variable we call $GWETD_{TOP}$, which is the difference between the current and preceding value of $GWET_{TOP}$. Negative values of $GWETD_{TOP}$ correspond to soil evaporation, while positive values indicate precipitation occurrence. We transform all the negative $GWETD_{TOP}$ values to zero before including it as a predictor in the uncertainty framework.

It should be emphasized here that our goal was not to identify the best possible additional predictors, but rather to simply illustrate that such predictors could be utilized to constrain IMERG uncertainty estimates. This issue is discussed further in Section 5.3.

3 Methodology

3.1 Matching and Preprocessing of Multiple Datasets

Following the approach of Khan et al. (2018), IMERG-E, GV-MRMS, DPR, and MERRA-2 data are matched in space and time to a consistent 0.1° 30-minute grid. GV-MRMS is upscaled by averaging all grid cells in a 10×10 window, provided that the RQI for at least 90% of these pixels is 100. The two DPR derived products are regridded by averaging all the DPR footprint scale (~5 km) estimates falling within a 0.1° grid cell, and then matched into the nearest 30-minute IMERG observation interval. MERRA-2 is mapped into the IMERG grid using nearest neighbor interpolation, and also matched to the nearest 30-minute IMERG observation interval. Figure 1 (upper panels) shows an example of the regridded coincident precipitation estimates from 2BCMB, GV-MRMS and IMERG-E.

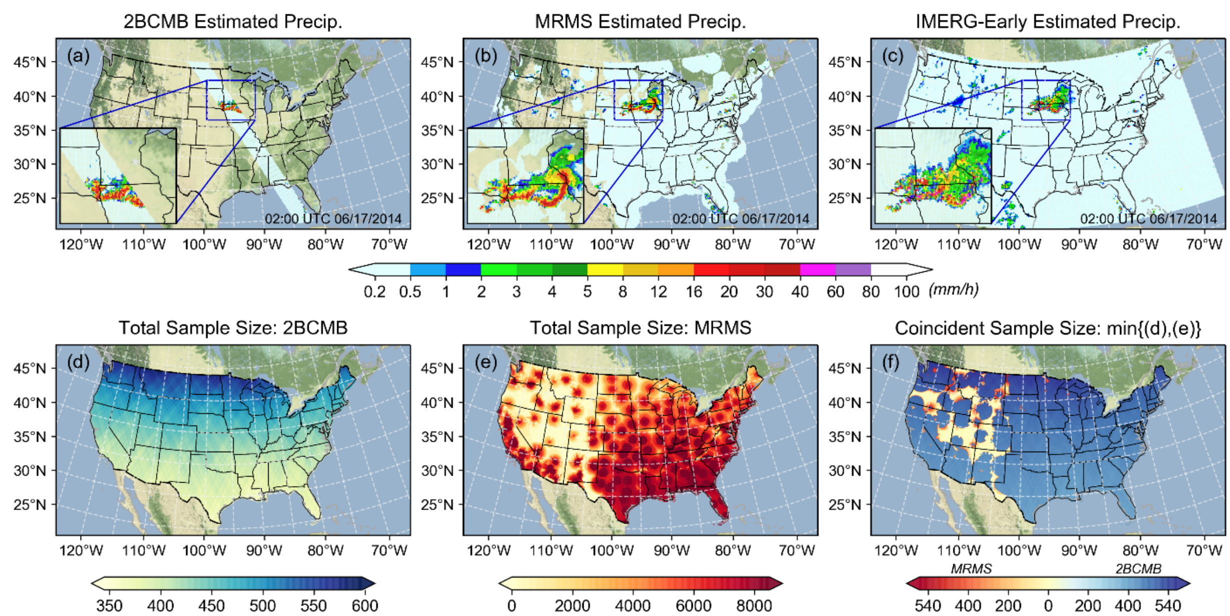

Figure 1. Coincident precipitation estimates from regridded (a) 2BCMB (2ADPR is similar; see Supplemental Figure S1), (b) GV-MRMS, and (c) IMERG-E for 02:00–03:00 UTC 17 June 2014, with the maps for the total sample size of (d) 2BCMB (also 2ADPR), (e) GV-MRMS, and (f) the coincident sample size—the minimum from (d) and (e)—within 0.1°×0.1° boxes during the study period.

The sample size of DPR products generally decreases from north to south due to the inclined orbit of GPM (Figure 1d), while GV-MRMS data is limited in western CONUS because of radar beam blockage (Figure 1e). The coincident data sample size thus depends on location and is generally less than 600 in each 0.1° grid cell (Figure 1f). To ensure a sufficiently large sample size, error models are trained and validated by pooling all coincident 0.1° data samples within 1°×1° spatial windows. In some parts of western CONUS this pooling is insufficient; in windows where the sample size is less than 5,000, we further pool data from the adjacent four windows in the east–west and north–south directions.

3.2 CSGD-based Uncertainty Quantification Framework

The uncertainty quantification framework selected in this study follows the censored shifted gamma distribution (CSGD) method developed by Scheuerer & Hamill (2015) for postprocessing ensemble numerical precipitation forecasts. It was adapted by Wright et al., (2017) to characterize the uncertainty for daily-scale satellite precipitation estimates. The CSGD is able to simultaneously depict precipitation occurrence and magnitude by introducing a “shift” parameter δ (δ<0) into the conventional two-parameter gamma distribution $F_{\mu,\sigma}$ (parameterized here by its mean μ and standard deviation σ, rather than shape and scale/rate parameters). The cumulative distribution function (CDF) of the CSGD is left-censored at zero:

$$F_{\mu,\sigma,\delta}\left(x\right)=\begin{cases}F_{\mu,\sigma}\left(x-\delta\right), & for\ x\geq 0\\ 0, & for\ x<0\end{cases} \tag{1}$$

where x is precipitation rate (mm h^{-1}). The vertical intercept $F_{\mu,\sigma,\delta}(0)$ is one minus the probability of precipitation (POP), and the CDF to the right of zero represents the nonexceedance probabilities associated with nonzero precipitation rates.

The CSGD-based error model consists of two main pieces: 1.) a “climatological CSGD” with parameters μ, σ, δ [i.e., Eqn. (1)]; and 2.) a regression system that comprises the error model. Once trained, this regression system can produce an estimated “conditional” CSGD with parameters $\mu(t)$,

$\sigma(t)$, and $\delta(t)$ that represents the possible true precipitation rate and POP conditioned on an IMERG retrieval at time t (and other optional predictors). This regression system can capture both conditional bias and heteroscedasticity, as well as the discrete-continuous nature of precipitation and associated errors. The most basic regression system lets $\mu(t)$ increase linearly with IMERG magnitude $P_I(t)$, and all models used here assume that $\sigma(t)$ is proportional to the square root of $\mu(t)$ (see Scheuerer & Hamill, 2015). We will refer to this most basic variant as the "linear model":

$$\mu(t) = \mu\left[\alpha_2 + \alpha_3 \frac{P_I(t)}{\overline{P}_I}\right] \tag{2}$$

$$\sigma(t) = \alpha_4 \sigma \sqrt{\frac{\mu(t)}{\mu}} \tag{3}$$

$$\delta(t) = \delta \tag{4}$$

where $\overline{P}_I$ denotes the climatological IMERG mean.

The linearity assumption can be further relaxed to account for nonlinear conditional bias. This version (hereafter "nonlinear model") replaces Eqn. (2) with:

$$\mu(t) = \frac{\mu}{\alpha_1} \mathrm{log1p}\left\{\mathrm{expm1}(\alpha_1)\left[\alpha_2 + \alpha_3 \frac{P_I(t)}{\overline{P}_I}\right]\right\} \tag{5}$$

where log1p(x) = log(1+x), and expm1(x) = exp(x) − 1.

Both the linear and nonlinear models can also accommodate extra time-varying predictors or covariates $C(t)$, potentially further constraining (i.e., narrowing) uncertainty estimates. To this end, Eqns. (2) and (5) can be replaced with:

$$\mu(t) = \mu\left[\alpha_2 + \alpha_3\frac{P_I(t)}{\bar{P}_I} + \alpha_5\frac{C(t)}{\bar{C}}\right] \tag{6}$$

$$\mu(t) = \frac{\mu}{\alpha_1}\operatorname{log1p}\left\{\operatorname{expm1}(\alpha_1)\left[\alpha_2 + \alpha_3\frac{P_I(t)}{\bar{P}_I} + \alpha_5\frac{C(t)}{\bar{C}}\right]\right\}, \tag{7}$$

respectively, where $\bar{C}$ is the climatological mean of the covariate. While multiple covariates can be used simultaneously (not depicted in Eqns. 6–7; Scheuerer & Hamill, 2015 and Wright et al., 2017), this study only considers covariates individually.

All of the above regression coefficients (α_1–α_5) as well as the three CSGD parameters are optimally estimated using the techniques detailed in Scheuerer & Hamill (2015), which minimize the continuous ranked probability score (CRPS) between empirical and theoretical CDFs.

3.3 Error Model Training and Validation

2ADPR and 2BCMB are first compared against coincident GV-MRMS observations over CONUS. This comparison considers the ability to correctly detect precipitation occurrence and to estimate precipitation rates of hit cases. To evaluate precipitation occurrence, we create contingency tables showing the numbers and rates of hits, misses, false alarms, and correct non-detects (Wilks, 2019) using 0.1 mm h^{-1} as the detection threshold (see Section 2.3). Precipitation rates for hits are then assessed for every 1°×1° spatial window in terms of three evaluation metrics: relative bias (RB), root mean squared error (RMSE), and Pearson's correlation coefficient (CC), which have been widely used in previous studies (e.g., Tan et al., 2018; Khan et al., 2018).

The regridded coincident datasets are randomly divided with 80% of observations used for CSGD-based error model training and 20% for validating model performance. A range of error model complexities are explored: linear models [Eqns. (1)–(4)], nonlinear models [Eqns. (1), (3)–(5)], linear models with a single covariate [Eqns. (1), (3), (4), (6)], and nonlinear models with a single covariate [Eqns. (1), (3), (4), (7)]. All the error models are trained using both the selected DPR data and GV-MRMS, while the latter are used as performance benchmarks to evaluate if the proposed DPR-based model appears reasonable.

DPR- and GV-MRMS-trained error models are then applied to the validation dataset, and their conditional CSGD estimates of reference precipitation are evaluated against GV-MRMS observations from that dataset using both deterministic and probabilistic metrics. To examine if the model can effectively characterize the central tendency of IMERG error, we compare the conditional CSGD median with GV-MRMS using mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - o_t\right| \qquad (8)$$

where y_t is either the CSGD median or IMERG, o_t is the coincident GV-MRMS observation at time t, and n is the number of (y_t, o_t) pairs.

Similar to other deterministic evaluation metrics (e.g., RB, RMSE, CC), MAE is insufficient for fully characterizing the predicted (probabilistic) distributions from CSGD error models. CRPS, on the other hand, measures the dispersion of these distributions around a GV-MRMS observation. CRPS thus offers a probabilistic performance measure of the error models:

$$\mathrm{CRPS}(F_{\mu(t),\sigma(t),\delta(t)}, o_t) = \int_{-\infty}^{\infty}\left[F_{\mu(t),\sigma(t),\delta(t)}(x) - \mathrm{I}(o_t \le x)\right]^2 dx \qquad (9)$$

where $F_{\mu(t),\sigma(t),\delta(t)}$ denotes the CDF of the conditional CSGD model at time t, and $\mathrm{I}(\cdot)$ is a step function that takes the value of 1 if $x \geq o_t$ (i.e., GV-MRMS observation at time t) and 0 elsewhere. Low CRPS indicates that the predicted CSGD's density is concentrated relatively close to the reference, while high CRPS implies either a very "wide" distribution or one that is heavily biased. Note that CRPS is mathematically identical to MAE for deterministic—as opposed to probabilistic—predictions.

Heteroscedasticity in IMERG errors means that we should not simply compare or combine CRPS values across various locations, since, like MAE or RMSE, CRPS will tend to be larger for heavier precipitation regimes. This has three implications for our model validation. First of all, we apply "reduction CRPS" (RCRPS; Trinh et al., 2013) for comparing model performance across different locations (i.e., 1°×1° boxes). It is normalized by the standard deviation of GV-MRMS observations at that location (denoted as σ_M) and thus is dimensionless:

$$\mathrm{RCRPS} = \frac{\mathrm{CRPS}}{\sigma_M} \tag{10}$$

Second, the validation dataset is then grouped into four categories: hits, misses, false alarms, and correct non-detects, by comparing the coincident IMERG and GV-MRMS data (using the same 0.1 mm h^{-1} threshold). CRPS is then calculated for each group to evaluate model performance under different cases. In addition, the calculated CRPSs of hit cases are further grouped by the magnitude of IMERG, to investigate the magnitude-dependent performance of the uncertainty estimates from different error models.

Finally, we further examine the performance of the error models' probabilistic estimates of precipitation events given a number of thresholds using reliability diagrams (see Wilks, 2019 for details). Considering GV-MRMS observation o_t and the CDF of CSGD error model $F_{\mu(t),\sigma(t),\delta(t)}$ at time t, the observed and predicted probability that an "event" occurred can be defined:

$$X_{e,t} = \begin{cases} 1, & for\ o_t > TH \\ 0, & for\ o_t \leq TH \end{cases} \tag{11}$$

$$Y_{e,t} = 1 - F_{\mu(t),\sigma(t),\delta(t)}(TH) \tag{12}$$

where TH is the threshold (mm h^{-1}). $X_{e,t}$ denotes the observed probability of threshold exceedance (either 0 or 1), while $Y_{e,t}$ is the predicted probability of threshold exceedance (between 0 and 1) from the error model. Following previous studies (e.g., Clark & Slater, 2006; Ghazvinian et al., 2020), we sort and group all predicted probabilities $Y_{e,t}$ (e.g., t =1, 2, … N for validation dataset) into ten equally-sized bins (0–10%, 10%–20%, …, 90%–100%). For each group, both the average predicted probability and the average observed probability are calculated. In a reliability diagram, a perfect prediction model would yield results that fall along the 1:1 line. For example, when $Y_{e,t}$ = 0.90, we expect the event to occur 90% of the time in reality. All coincident samples across CONUS are pooled for this analysis.

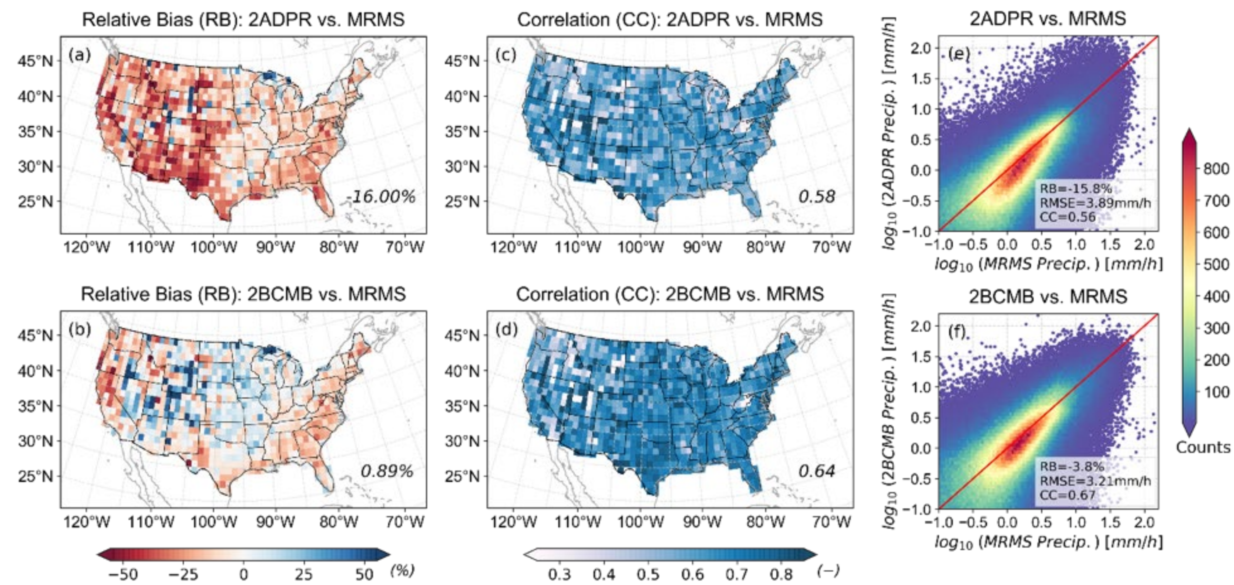

Figure 2. The spatial maps of (a-b) relative bias, (c-d) correlation coefficient, and (e-f) density scatterplots by comparing the coincident precipitation estimates from 2ADPR and 2BCMB versus GV-MRMS during the study period. Only precipitation estimates greater than 0.1 mm h^{-1} are considered. Inset values in (a)-(d) are the mean across all grid boxes (1°×1°) over CONUS.

4 Results

4.1 DPR Products as Reference Precipitation

Figure 2 shows CONUS-wide evaluation “hits only” results of DPR derived products against coincident GV-MRMS observations. 2ADPR underestimates precipitation almost everywhere, particularly in the western parts of the country leading to a CONUS-wide average underestimation of 16% (Figure 2a). 2BCMB, on the other hand, varies geographically with overestimation (e.g., the Rockies and Great Plains) and modest underestimation (e.g., the West and East Coasts) leading to a CONUS-wide average within 1% of GV-MRMS (Figure 2b). Moreover, 2BCMB is better correlated with GV-MRMS observations over most of CONUS (Figures 2c–d). Scatterplots and three summary statistics (RB, RMSE, and CC) again indicate that 2BCMB generally outperforms

2ADPR (Figures 2e–f). 2ADPR, however, shows somewhat better detection skills—lower numbers of false alarms and missed precipitation (Table 1). From Table 1, however, it can be seen that hits are more common than false alarms and missed precipitation. They are also probably more important in the context of applications, which tend to focus on medium-to-heavy rainfall in which hits are prevalent. Prioritizing "hit-relevant" performance such as bias and correlation, we have elected to focus on 2BCMB for the remainder of this study. This issue deserves further attention, however, as the relative performance of 2ADPR and 2BCMB can be expected to vary geographically, seasonally, and with precipitation microphysics (Skofronick-Jackson et al., 2017; Skofronick-Jackson et al., 2018).

Table 1. The contingency tables of 2ADPR and 2BCMB, benchmarking against the ground reference GV-MRMS. For each pair of estimates, hits (top left), false alarms (top right), misses (bottom left), and correct non-detects (bottom right) are shown. The total paired data sample size over CONUS is 20,986,107.

	$P_{GV\text{-}MRMS} \geq 0.1$mm h^{-1}	$P_{GV\text{-}MRMS} < 0.1$mm h^{-1}
$P_{2ADPR} \geq 0.1$mm h^{-1}	931,165 (4.4%)	52,187 (0.2%)
$P_{2ADPR} < 0.1$mm h^{-1}	44,407 (0.2%)	19,958,348 (95.1%)
$P_{2BCMB} \geq 0.1$mm h^{-1}	833,803 (4.0%)	324,859 (1.5%)
$P_{2BCMB} < 0.1$mm h^{-1}	141,769 (0.7%)	19,685,676 (93.8%)

In addition to the absolute precipitation estimation performance, another key consideration is the need for approximate (if not strict) independence from the SMP product being evaluated, if DPR derived products are to be used as alternative references. Khan et al. (2018) and You et al., (2020) argue that the independence need can be approximately met as numerous processing steps and assumptions stand between the DPR/GMI observations and their manifestation within IMERG (as highlighted in Section 2.1). Nonetheless, we examined this by comparing the accuracy of

IMERG (for hits only) relative to both 2BCMB and GV-MRMS (Figure 3). The RBs between IMERG and the two reference datasets are similar (31% in the case of GV-MRMS, and 36% for 2BCMB; while visual inspection shows different conditional bias features). RMSEs are very similar (5.08 versus 5.13 mm h^{-1}), while Pearson correlation CC with 2BCMB is slightly higher (0.45) than with GV-MRMS (0.41). Contingency tables corresponding to Figure 3 are shown in Supplemental Table S1 and reveal similar detection skills of IMERG relative to 2BCMB and GV-MRMS. Taken together, these results suggest that there is indeed approximate independence between IMERG and 2BCMB, confirming the latter's potential to evaluate the former. This issue is discussed further in Section 5.1.

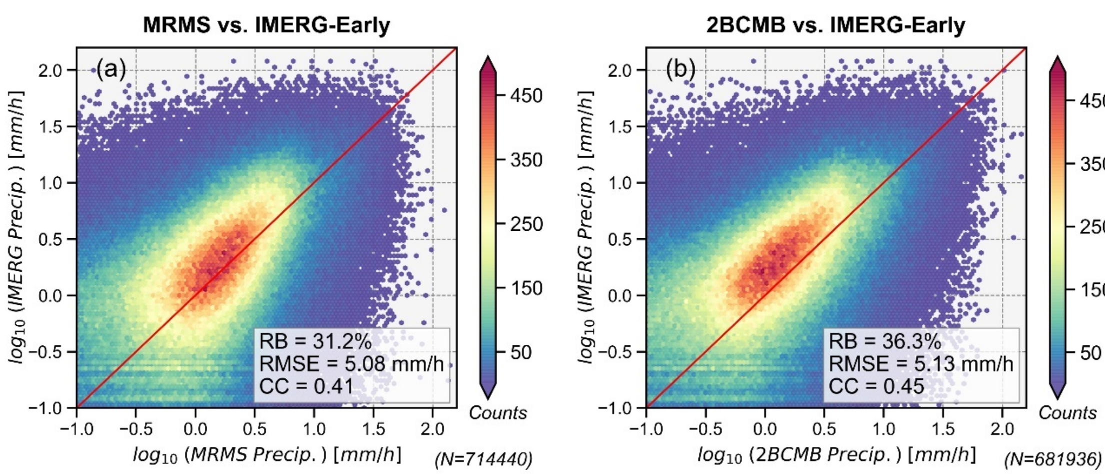

Figure 3. Density scatterplots of coincident precipitation estimates from GV-MRMS and 2BCMB versus IMERG. Only precipitation estimates greater than 0.1 mm h^{-1} are considered, including all the data samples over the CONUS during the study period.

Climatological CSGDs fitted to GV-MRMS and 2BCMB share similar spatial patterns of μ and POP (Supplemental Figures S2a–d; σ and δ were also investigated but are not shown). 2BCMB tends to slightly underestimate μ and POP relative to GV-MRMS, likely reflecting its imperfect detection and quantification. Fitted CDFs for climatological CSGDs are illustrated for a 1°×1° box in the state of Tennessee in the Southeastern CONUS and a 1°×1° box in New Mexico in the Southwest (Figures S2e–f), which are randomly selected to represent the locations characterized by different climates. Although these CSGDs closely match the empirical CDFs over the more humid Southeastern box, 2BCMB exhibits a higher probability of zero precipitation and relatively large differences from GV-MRMS for light precipitation rates less than 1 mm h^{-1} (Figure S2e). In the drier Southwest, a small negative bias in estimated POP is evident (Figure S2f), consistent with previous studies and related to the CRPS minimization scheme (Ghazvinian et al., 2020).

4.2 CSGD Error Model Visual Inspection and Deterministic Performance

Linear and nonlinear versions of the CSGD error models trained by GV-MRMS and 2BCMB are further compared over the 1°×1° boxes in the Southeast and Southwest CONUS (Figure 4; see Figure S3 for identical results plotted on linear rather than log-log scales). For these selected boxes and other locations in the CONUS, IMERG is prone to overestimate precipitation at half-an-hour scale, particularly at higher precipitation rates. This is consistent with previous regional studies (e.g., Tan et al., 2017; Moazami and Najafi, 2021) and CONUS-wide analysis (Gebregiorgis et al., 2018).

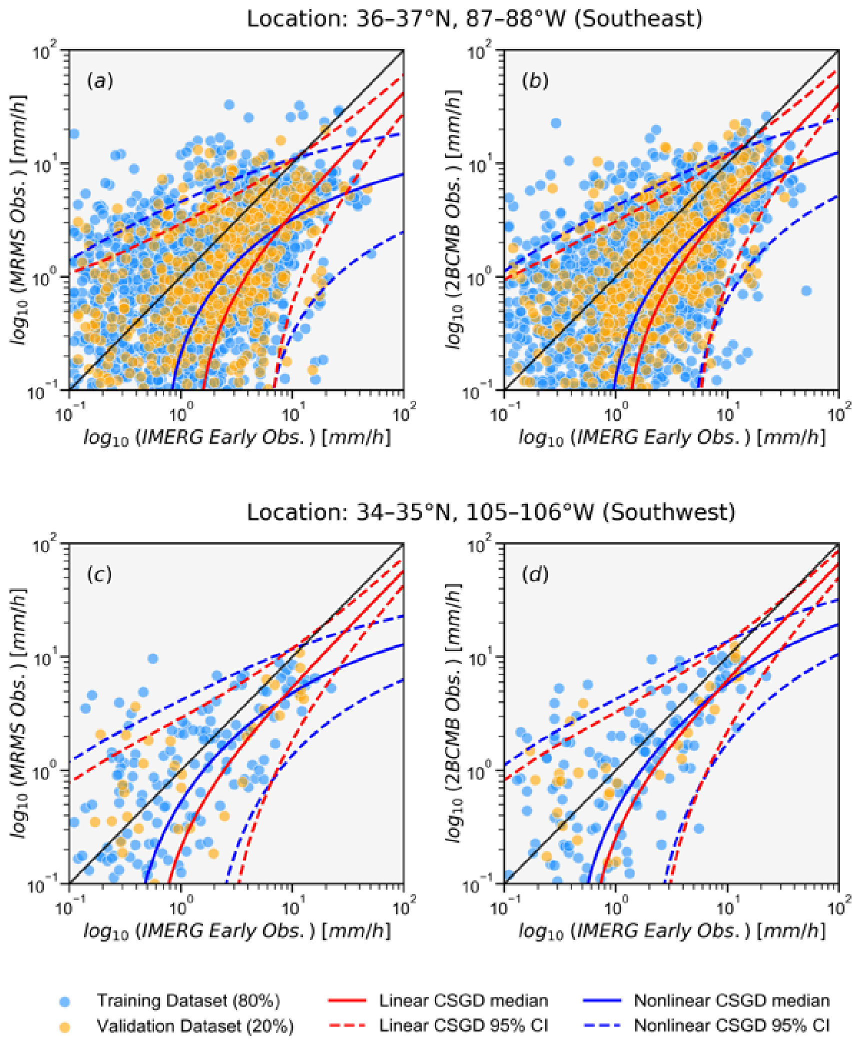

Figure 4. Linear (red lines) and nonlinear (blue lines) conditional CSGD models for (a, b) the Southeast 1°×1° box and (c, d) Southwest 1°×1° box, trained and compared against GV-MRMS (left panels) and 2BCMB (right panels). See Figure S2 for identical results, but plotted on linear scales.

The nonlinear models in Figure 4 perform better than linear models for capturing the conditional bias that is evident at high precipitation rates (e.g., >10 mm h^{-1}). Visual inspection suggests that the 2BCMB-trained models have similar features to the GV-MRMS-trained models, though the nonlinear versions show slightly weaker systematic bias and a wider uncertainty bound. Both GV-MRMS- and 2BCMB-trained CSGD models provide reasonable fits to the validation dataset, showing the robustness of this uncertainty quantification framework. The sample size of coincident data for the more arid southwestern box (Figures 4c–d) is limited due to lower POP, while IMERG systematic bias is somewhat lower than in the Southeast (Figures 4a–b). These results highlight the relative flexibility of this CSGD-based uncertainty quantification method under very different climatic conditions.

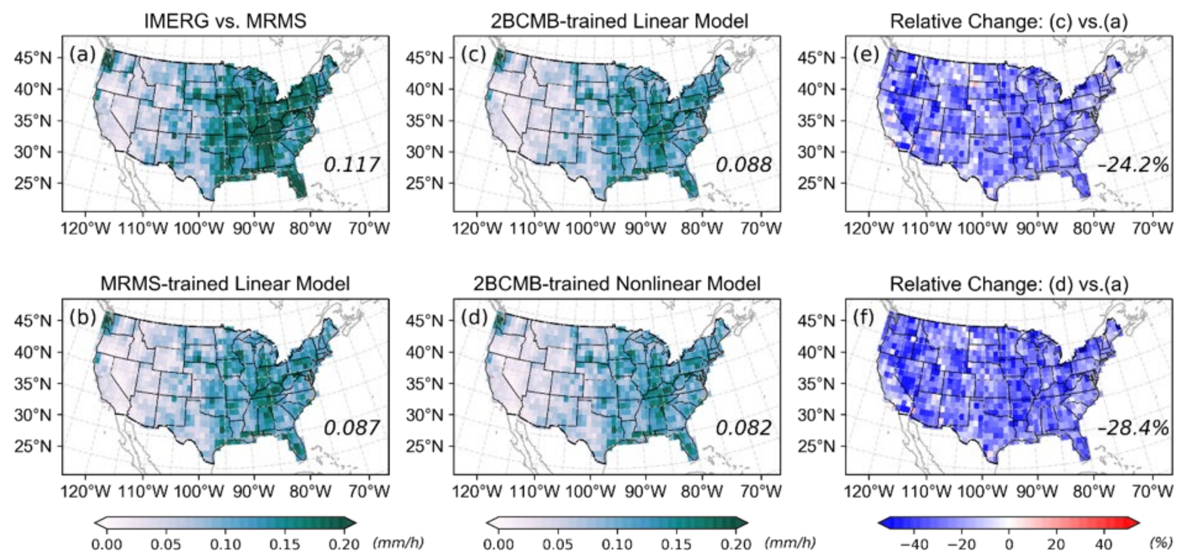

Figure 5. MAE calculated relative to GV-MRMS: (a) IMERG, and the median predicted by (b) GV-MRMS-trained linear model, (c) 2BCMB-trained linear model, and (d) 2BCMB-trained nonlinear model. Relative percentage change of MAE relative to IMERG results in (a): (e) 2BCMB-trained linear model, and (f) 2BCMB-trained nonlinear model. All the results are calculated by validation data samples, and inset values are the means of all 1°×1° boxes in the CONUS.

The central tendency (i.e., means or medians) predicted by the CSGD error model represent the reducible IMERG error (i.e., bias; see Wright et al., 2017). We compare the CSGD medians predicted by the GV-MRMS- and 2BCMB-trained error models over CONUS against GV-MRMS observations using MAE (Figures 5a–d). The 2BCMB-trained linear (nonlinear) models can isolate bias in IMERG—reducing MAE by 24.2% (28.4%) on average (Figures 5e–f). The GV-

MRMS-trained model performs similarly (see Figure 5d for the linear model; nonlinear model not shown). There is no obvious change in MAE incorporating $GWTD_{TOP}$ (inset statistics in Figure 6) or other MERRA-2 predictors (results not shown), indicating that these variables have limited roles in explaining IMERG systematic error.

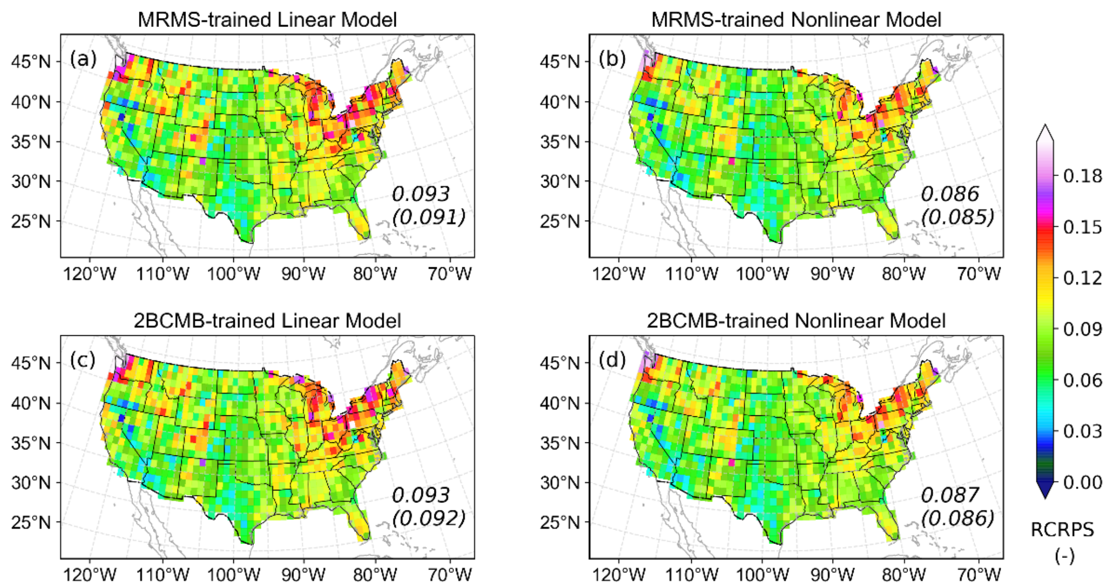

Figure 6. The comparison of mean RCRPS maps for (a) GV-MRMS-trained linear model, (b) GV-MRMS-trained nonlinear model; (c) 2BCMB-trained linear model, and (d) 2BCMB-trained nonlinear model. Inset values are the means of all 1°×1° boxes in the CONUS, and the results for nonlinear models with $GWTD_{TOP}$ are in parentheses.

4.3 CONUS-wide Probabilistic Evaluation

Figure 6 presents the mean RCRPS maps for different error models over CONUS. The results show somewhat lower CONUS-averaged RCRPS for the nonlinear GV-MRMS- and 2BCMB-trained models (0.086 and 0.087, respectively) compared to the linear models (0.093 for both the GV-MRMS- and 2BCMB-based models). This highlights both the potential of 2BCMB as a reference product and confirms the superiority of the nonlinear model. Interestingly, the areas with highest RCRPS—e.g., Northeast CONUS, the Rockies, the Great Lakes states, and the Pacific

Northwest—still show relatively high RCRPS values after considering nonlinear bias. This reflects region-dependent IMERG uncertainties, likely associated with relatively high precipitation totals combined with "complicating factors" such as orography, lake-effect snow, and high fractions of annual precipitation falling as snow.

Figure 7. Boxplots of the percentage decrease in CRPS relative to the linear model with no predictors for (a-b) GV-MRMS-trained models, and (c-d) 2BCMB-trained models with various model complexities and predictors. The results are calculated based on training data samples, including all the 1°×1° boxes in the CONUS. The covariates TQV, T2M, $GWET_{TOP}$ represent the total precipitable water vapor, 2-m air temperature and topmost soil layer's ground wetness index from MERRA-2 respectively, and $GWETD_{TOP}$ is the ground wetness change indicator that is derived from $GWET_{TOP}$.

4.4 Model Complexity and Conditional Performance

We compared CONUS-wide performance improvements for a range of GV-MRMS- and 2BCMB-based error models with varied model complexities, measured by the percentage decrease of CRPS relative to the linear model without any covariate (Figure 7). Consistent with Figure 4, nonlinearity is the most critical model feature for constraining uncertainty (i.e., improving CRPS). Moreover, the improvement gained via the nonlinear formulation tends to be larger for the GV-MRMS-based model than for the 2BCMB-based model (a mean of 6% vs. 4%, respectively), consistent with the larger conditional bias "detected" by GV-MRMS as shown in Figures 4 and S2. Figure 7 also shows that the most informative covariate we evaluated is $GWTD_{TOP}$, which can provide modest improvements (generally 0.5-1.5% reduction in CRPS; ranging as high as 5% in some locations) to both the linear and nonlinear models. On average, the other covariates—TQV, T2M, $GWTD_{TOP}$—provide more limited benefits. Therefore, only $GWTD_{TOP}$ is shown elsewhere in this study.

We further group the validation dataset and corresponding model predictions into four cases: hits, misses, false alarms and correct non-detects. Table 2 summarizes CONUS-wide mean CRPS values of the four groups. Since the number of instances of the groups differ widely, we also show an overall "weighted mean" for each model. In general, the 2BCMB-based model shows similar performance as GV-MRMS-trained model in characterizing both the overall uncertainty and the errors associated with these four cases. The results highlight that the role of model nonlinearity and $GWTD_{TOP}$ may vary among different cases. The nonlinear model shows improved performance in both hits and correct non-detects cases, but performs worse than the linear model for misses and false alarms. Because the correct non-detects and hits dominate in the coincident samples (accounting for 92.6% and 3.4% of the validation dataset, respectively), the weighted mean CRPS of the nonlinear model outperforms that of the linear version.

Table 2. The CONUS-wide mean CRPS for different cases: hits (3.4% of validation data), misses (2.0%), false alarms (2.0%) and correct non-detects (92.6%), and their weighted mean according to the percentage of different cases in validation dataset.

Cases[1]	Mean CRPS for different model complexities (mm h^{-1})			
	GV-MRMS-trained Linear Model		GV-MRMS-trained Nonlinear Model	
	No Covariate	With $GWTD_{TOP}$	No Covariate	With $GWTD_{TOP}$
Hits	1.484	1.449	1.282	1.275
Correct Non-detects	0.0009	0.0013	0.0003	0.0007
Misses	0.913	0.889	0.924	0.903
False Alarms	0.132	0.131	0.219	0.214
Weighted Mean	0.072	0.071	0.067	0.066
	2BCMB-trained Linear Model		2BCMB-trained Nonlinear Model	
	No Covariate	With $GWTD_{TOP}$	No Covariate	With $GWTD_{TOP}$
Hits	1.497	1.470	1.330	1.325
Correct Non-detects	0.0004	0.0007	0.0001	0.0003
Misses	0.922	0.904	0.931	0.916
False Alarms	0.133	0.132	0.190	0.186
Weighted Mean	0.072	0.071	0.067	0.067

[1]: The four cases are divided based on GV-MRMS and the coincident IMERG observations. The percentages of each case are similar to those in Table S1, but only the validation dataset is considered here.

At the same time, $GWTD_{TOP}$ improves the uncertainty estimates for hits and misses, and the latter particularly. $GWTD_{TOP}$ can indicate precipitation occurrence using changes in land surface wetness conditions, and it thus helps in quantifying the "missed" precipitation that is difficult to address in most SMP products, even with gauge corrections (Li et al., 2015; Tian et al., 2009). However, the inclusion of $GWTD_{TOP}$ also increases CRPS in cases of correct non-detects.

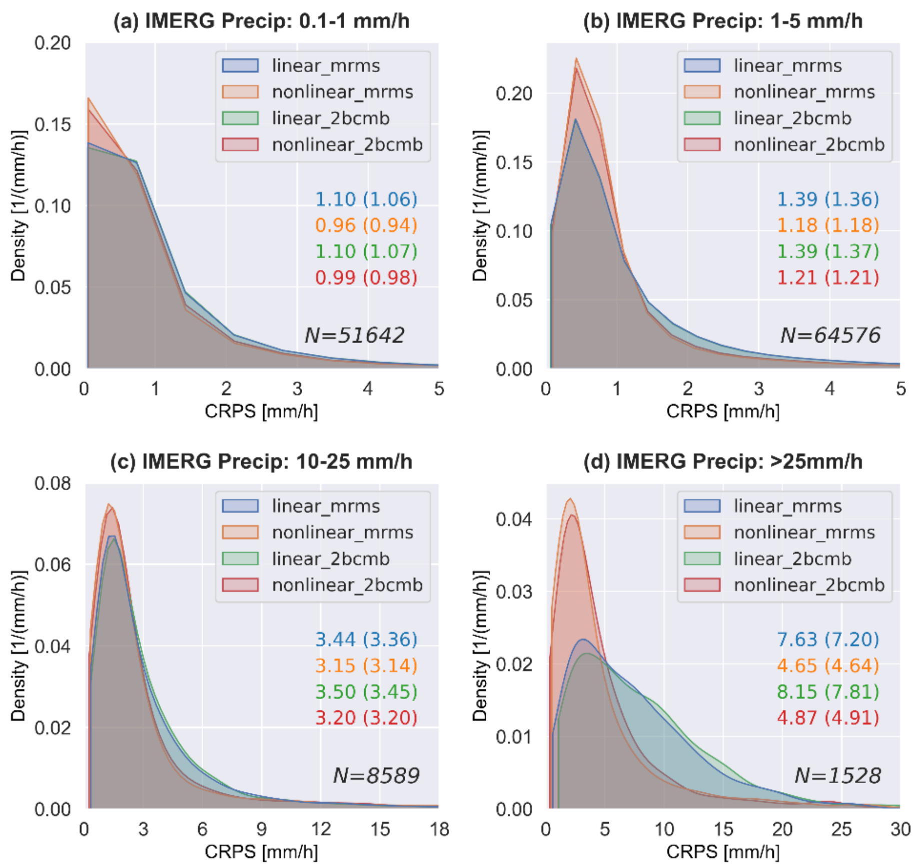

Figure 8. Distributions of CRPS for GV-MRMS- and 2BCMB-trained models for different intervals of IMERG precipitation rates (validation data samples of "hits" only; N is sample size). Inset values are the mean CRPS [mm h^{-1}] for the different models. Values inside (outside) the parentheses include (don't include) $GWTD_{TOP}$ as a predictor.

To better characterize the magnitude-dependent performance of the uncertainty estimates for hit cases, we consider 0.1, 1, 5, 10, and 25 mm h^{-1} thresholds, which roughly correspond to the 0.97, 0.99, 0.995, and 0.998 quantiles of IMERG climatology. Figure 8 shows kernel-based density functions of CRPS for IMERG data between these thresholds. For IMERG-detected precipitation events, the nonlinear model outperforms the linear model, as the distributions of CRPS of the former are more concentrated towards zero and thus have a smaller mean. For larger IMERG estimates, differences become more apparent and the nonlinear models yield smaller CRPS scores than the linear models (Figures. 8c–d). This is unsurprising, since IMERG shows substantial nonlinear conditional bias at high precipitation rates (e.g., Figure 4). This analysis also shows that the 2BCMB-trained models perform similarly to the GV-MRMS-trained model at different precipitation intensities, showing slight degradation for very heavy precipitation (>25 mm h^{-1}), likely attributable to attenuation. The benefits of adding $GWTD_{TOP}$ is not visually obvious in these CRPS-based analyses; it does offer very modest benefits, however, particularly in the linear models at higher precipitation rates (inset statistics in Figure 8).

Reliability diagrams for all the error models with thresholds of 0.1 mm h^{-1}, 5 mm h^{-1}, and 10 mm h^{-1} are compared in Figure 9. The results clearly shows that the nonlinear model always offers more reliable estimates than the linear model (i.e., it always falls closer to the 1:1 line). The 2BCMB-based nonlinear model presents similar skill to the GV-MRMS-trained model, except for events larger than 10 mm h^{-1}. The model reliability varies at different event thresholds. For precipitation occurrence (i.e., 0.1 mm h^{-1}), all error models tend to be consistently above the diagonal, particularly at low-medium forecast probability categories. This feature indicates that the observed frequency in each category exceeds the model estimated frequency (i.e., a “dry bias”; Wilks, 2019). This is similar to the findings of an earlier study on CSGD models (Ghazvinian et al., 2020) that relate this dry bias to the underestimation of POP by climatological CSGDs (which can be seen in Figure. S2f). This dry bias is reduced by incorporating $GWTD_{TOP}$, which is unsurprising since it serves as an “indicator” of precipitation occurrence. For heavy events, on the other hand, all error models fall below the diagonal at high forecast probability categories, indicating the model tends to overestimate occurrence frequency. As discussed in Scheuerer et al.

(2017), however, the uncertainty in observed frequency increases at higher precipitation rates due to limited sample sizes.

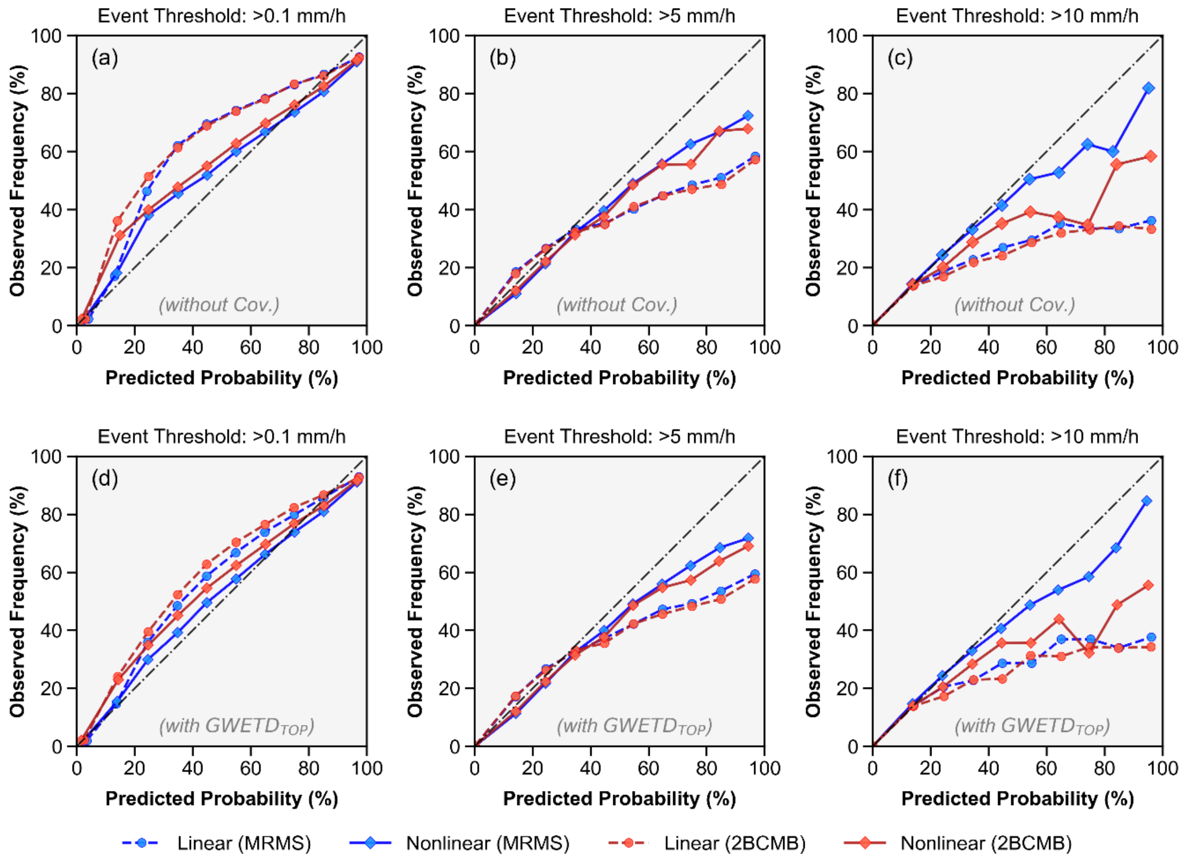

Figure 9. Reliability diagrams for GV-MRMS- and 2BCMB-trained models with various event thresholds. The upper pannels are for models without covariates, and lower pannels are for models with $GWTD_{TOP}$.

5 Discussion

5.1 DPR Data as Reference

Our findings point to the promise of DPR products—and particularly 2BCMB—to serve as a reference in the proposed global IMERG uncertainty estimation framework. It must be recognized, however, that while 2BCMB generally outperforms 2ADPR over CONUS (Figure 2; though not in terms of precipitation detection; see Table 1), this conclusion may not hold in other places around the globe. The rejection of 2ADPR as a reference here is not inconsistent with Khan et al., (2018), who showed that conditional biases in IMERG and 2ADPR are similar.

Despite the similar error modeling results between 2BCMB- and GV-MRMS-trained error models, the two references differ in important ways: 2BCMB estimates instantaneous precipitation rate, while GV-MRMS offers precipitation estimates aggregated into 30-minute intervals. This scale mismatch inevitably will introduce additional uncertainties, which are likely to manifest in the form of random errors both in precipitation occurrence and magnitude. This seems unlikely to influence systematic errors. It should be noted that 2BCMB underestimates high precipitation rates, probably due to attenuation of radar signals. This likely explains the somewhat poor performance of the 2BCMB-based error models at high precipitation rates, relative to performance at lower intensities.

As mentioned in Sections 2.1 and 4.1, an objection can be raised to the use of DPR-derived datasets as reference for IMERG or other GPM precipitation data products, owing to the (generally indirect) inclusion of DPR (and GMI, in the case of 2BCMB) in those products and thereby the potential for lack of independence between them and the DPR-based reference. Those results in Section 4.1 and the arguments of Khan et al. (2018) and You et al. (2020) suggest that this objection should not be overly concerning. The most direct contribution from DPR and GMI combined data in IMERG algorithm is the 45-day probability matching intercalibration of PMW-only precipitation retrievals, before the morphing and PMW-IR merging procedures that derive the ultimate gridded IMERG estimates (Huffman et al., 2020). The actual difference in IMERG estimates before and after this intercalibration is found to typically be less than 10% (Jackson Tan, personal communication, 7 April 2021). Nonetheless, recent work by Kirstetter et al. (2020) shows that systematic biases such as those associated with precipitation typology display similar features across Level-2 and Level-3 GPM products (i.e., QPE from DPR, GPROF-GMI, and IMERG), suggesting that uncertainty can propagate from DPR-based precipitation estimates into the SMP product.

5.2 CSGD-based Error Model

This study demonstrates that the CSGD error model can reasonably characterize IMERG uncertainty over CONUS. Consistent with previous studies (Hartke et al., 2020; Wright et al., 2017), model fitting was relatively robust to small sample sizes (Figure 5). This property is crucial if the framework is to be expanded to the entire globe, as GPM DPR sampling frequency decreases at lower latitudes (Figure1d, and Figure 1 in You et al., 2020).

The inclusion of additional predictors into the model provides a way of improving uncertainty estimates. As demonstrated in this study, incorporating $GWTD_{TOP}$ derived from the MERRA-2 reanalysis product modestly improved the model's performance in characterizing precipitation occurrence (Table 2, and Figure 9). Expectation of greater gains using MERRA-2 is unwarranted given its coarse temporal and particularly spatial resolution. It isn't clear that other satellite-based products could be much help here—consider, for example, that remotely-sensed soil moisture data likely lack the combination of global coverage and short latency needed to inform near real-time IMERG uncertainty estimates. High-resolution global-scale numerical weather forecasts such as those from NASA's Goddard Earth Observing System Forward Processing products (GEOS-FP; Molod et al., 2012), on the other hand, seem to offer potential as they are available on a consistent global basis in realtime. Another direction which is being explored in separate work is using simple metrics derived from the IMERG precipitation fields themselves as predictors in the error model.

Finally, while we explored several deterministic and probabilistic performance measures in this study, it is doubtful that we have fully explored all relevant aspects of model skill. Future global validation efforts will continue exploring this topic using more candidate evaluation metrics (e.g., Massari & Maggioni, 2020; Wilks, 2019) and in varied environmental settings.

5.3 IMERG Uncertainty Beyond the DPR Swath

This study's central premise—that DPR measurements on board the GPM core observatory can serve as an alternative reference for estimating IMERG uncertainty—carries a key limitation: since spatial and temporal coincidence between IMERG and DPR is needed to train the error model, this can only occur within the DPR swath. Of course, DPR and GMI are co-located on the GPM core observatory. The result is that the uncertainty estimates presented here primarily reflect the relationships between DPR and GMI-influenced IMERG. Because GMI is the most accurate PMW precipitation radiometer currently in space (Skofronick-Jackson et al., 2018), our analyses probably provide the "best scenario" (i.e., lowest uncertainty); the real uncertainty associated with IMERG estimates that are dominated by other PMW or IR sensors or morphing is likely greater (e.g., Tan et al., 2016, and Li et al., 2020; also see Figure S3) than our models would predict.

To overcome this limitation, more work is needed to examine how DPR-based uncertainty estimates can be "inflated" to better reflect the properties of other PMW and IR sensors as well as IMERG's morphing scheme. The work of You et al. (2020)—who compared 2ADPR precipitation estimates against those of different PMW sensors within GPROF—provides a possible blueprint for addressing this.

6 Summary and Conclusions

This study proposes a prototype uncertainty quantification framework for satellite precipitation products, in which GPM DPR-derived observations are used in place of ground-based measurements. Though we focus our study on the IMERG-Early dataset at its native 30-minute, 0.1° resolution over CONUS, the quasi-global availability of DPR measurements allows this framework to be applied across the globe to any satellite precipitation dataset whose spatiotemporal resolution is similar to that of IMERG, such as CMORPH and PERSIANN. Uncertainty is modeled using a flexible and parsimonious three-parameter censored shifted gamma distribution (CSGD) error model which can characterize the probability of precipitation as well as intensity-dependent systematic bias and the potential range of random error. Uncertainty estimates from the CSGD error model trained on the 2BCMB reference are compared against high-quality ground observations from the GV-MRMS dataset, as well as uncertainties inferred from GV-MRMS-based versions of the error model. We believe this is the first study to generate IMERG uncertainty estimates using this general approach.

Our CONUS-wide assessment suggests that the combined (DPR and GMI) product 2BCMB outperforms 2ADPR (which uses DPR exclusively) in terms of precipitation intensity statistics. 2ADPR has somewhat better detection properties, however. Due to substantial intensity-dependent error behaviors in IMERG, the rejection of 2ADPR as a reference by Khan et al. (2018), and the supposition that better uncertainty estimates during precipitating periods would be preferable to better estimates of detection uncertainty, we focused our error modeling analysis on the 2BCMB product.

Multiple CSGD-based IMERG error models of varying complexity were trained using both GV-MRMS and 2BCMB. We find that the precipitation climatology characterized by GV-MRMS- and 2BCMB-based models yield similar properties and comparable performance throughout CONUS, though the 2BCMB-based model has slightly lower mean and POP values, likely attributable to its imperfect detection skill. 2BCMB-based model performance also suffers at high precipitation rates compared with GV-MRMS-based models. Evaluation using CRPS indicates that IMERG uncertainty, for both models, is relatively high at the Northeast CONUS, the Rockies, and the Pacific Northwest, where a number of “complicating factors” such as orography, lake-effect snow, as well as high fractions of annual precipitation falling as snow may complicate IMERG errors.

Relatively weak error model performance can be ameliorated by incorporating additional predictors to further constrain uncertainty estimates. We illustrate this by incorporating predictors from NASA’s MERRA-2 reanalyses, including a derived variable representing positive temporal deviations in near-surface soil moisture that can be associated with precipitation occurrence and that improves uncertainty estimates, albeit modestly. Higher-resolution numerical weather predictions—particularly short-range high-resolution global forecasts—could be potentially useful for informing uncertainty estimates for near-realtime versions of IMERG. In addition, variables derived from IMERG's ancillary data or the spatial structure of the IMERG fields themselves offer further promise to constrain IMERG uncertainty.

Despite the promising performance of this uncertainty framework over CONUS, its flexibility and robustness in other parts of the globe remain untested. The accuracy of IMERG as well as the relative performance of the 2ADPR and 2BCMB products are influenced by climate, land surface conditions, and atmospheric and precipitation properties (e.g., Khan et al., 2018; Tang et al., 2016). DPR's sampling frequency also varies with latitude due to its inclined orbit. Future work will focus on validating uncertainty estimates in other locations and conditions.

Notwithstanding these remaining challenges, the IMERG uncertainty estimates provided by this error modeling framework can benefit end-user applications (Kirschbaum et al., 2017). As illustrated in a recent study, the incorporation of CSGD-based satellite precipitation uncertainty can improve regional landslide hazard nowcasting, even when a CSGD error model is trained by very limited data (Hartke et al., 2020). It should be noted that the maximum benefits of the IMERG error modeling framework only can be achieved when regional and global environmental modeling systems and workflows are adapted to "ingest" such uncertainty information. Proof-of-concept efforts in that direction are equally important as further validation of our uncertainty estimation approach.

Acknowledgements:

Z. Li's and D.B. Wright's contributions were funded by NASA Precipitation Measurement Mission Grant 80NSSC19K0951. S. H. Hartke was supported by the NASA Earth and Space Science Fellowship Program (Grant 80NSSC18K1321). IMERG, 2ADPR and 2BCMB were downloaded from the NASA GES DISC website https://disc.gsfc.nasa.gov/. The authors would like to thank GPM Ground Validation team for preparing the regridded GV-MRMS dataset, which is available at https://pmm-gv.gsfc.nasa.gov/.

REFERENCES:

Adhikari, A., Liu, C., & Hayden, L. (2019). Uncertainties of GPM Microwave Imager Precipitation Estimates Related to Precipitation System Size and Intensity. *Journal of Hydrometeorology*, *20*(9), 1907–1923. https://doi.org/10.1175/JHM-D-19-0038.1

AghaKouchak, A., Mehran, A., Norouzi, H., & Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets: SYSTEMATIC AND RANDOM ERROR COMPONENTS. *Geophysical Research Letters*, *39*(9), n/a-n/a. https://doi.org/10.1029/2012GL051592

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., et al. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*(1), 207–224. https://doi.org/10.5194/hess-23-207-2019

Behrangi, A., & Wen, Y. (2017). On the Spatial and Temporal Sampling Errors of Remotely Sensed Precipitation Products. *Remote Sensing*, *9*(11), 1127. https://doi.org/10.3390/rs9111127

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., et al. (2014). Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data: Using the soil as a natural raingauge. *Journal of Geophysical Research: Atmospheres*, *119*(9), 5128–5141. https://doi.org/10.1002/2014JD021489

Clark, M. P., & Slater, A. G. (2006). Probabilistic Quantitative Precipitation Estimation in Complex Terrain. *Journal of Hydrometeorology*, *7*(1), 3–22. https://doi.org/10.1175/JHM474.1

Crow, W. T., van den Berg, M. J., Huffman, G. J., & Pellarin, T. (2011). Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART): SOIL MOISTURE ANALYSIS RAINFALL TOOL. *Water Resources Research*, *47*(8). https://doi.org/10.1029/2011WR010576

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

Foufoula-Georgiou, E., Guilloteau, C., Nguyen, P., Aghakouchak, A., Hsu, K.-L., Busalacchi, A., et al. (2020). Advancing Precipitation Estimation, Prediction, and Impact Studies. *Bulletin of the American Meteorological Society*, *101*(9), E1584–E1592. https://doi.org/10.1175/BAMS-D-20-0014.1

Gatlin, P. N., Petersen, W. A., Pippitt, J. L., Berendes, T. A., Wolff, D. B., & Tokay, A. (2020). The GPM Validation Network and Evaluation of Satellite-Based Retrievals of the Rain Drop Size Distribution. *Atmosphere*, *11*(9), 1010. https://doi.org/10.3390/atmos11091010

Gebregiorgis, A., & Hossain, F. (2014). Making Satellite Precipitation Data Work for the Developing World. *IEEE Geoscience and Remote Sensing Magazine*, *2*(2), 24–36. https://doi.org/10.1109/MGRS.2014.2317561

Gebregiorgis, Abebe S., Kirstetter, P., Hong, Y. E., Gourley, J. J., Huffman, G. J., Petersen, W. A., et al. (2018). To What Extent is the Day 1 GPM IMERG Satellite Precipitation Estimate Improved as Compared to TRMM TMPA-RT? *Journal of Geophysical Research: Atmospheres*, *123*(3), 1694–1707. https://doi.org/10.1002/2017JD027606

Gebregiorgis, Abebe Sine, Kirstetter, P.-E., Hong, Y. E., Carr, N. J., Gourley, J. J., Petersen, W., & Zheng, Y. (2017). Understanding Overland Multisensor Satellite Precipitation Error in TMPA-RT Products. *Journal of Hydrometeorology*, *18*(2), 285–306. https://doi.org/10.1175/JHM-D-15-0207.1

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

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Ghazvinian, M., Zhang, Y., & Seo, D.-J. (2020). A Nonhomogeneous Regression-Based Statistical Postprocessing Scheme for Generating Probabilistic Quantitative Precipitation Forecast. *Journal of Hydrometeorology*, *21*(10), 2275–2291. https://doi.org/10.1175/JHM-D-20-0019.1

Grecu, M., Olson, W. S., Munchak, S. J., Ringerud, S., Liao, L., Haddad, Z., et al. (2016). The GPM Combined Algorithm. *Journal of Atmospheric and Oceanic Technology*, *33*(10), 2225–2245. https://doi.org/10.1175/JTECH-D-16-0019.1

Guilloteau, C., Foufoula-Georgiou, E., & Kummerow, C. D. (2017). Global Multiscale Evaluation of Satellite Passive Microwave Retrieval of Precipitation during the TRMM and GPM Eras: Effective Resolution and Regional Diagnostics for Future Algorithm Development. *Journal of Hydrometeorology*, *18*(11), 3051–3070. https://doi.org/10.1175/JHM-D-17-0087.1

Hamada, A., & Takayabu, Y. N. (2016). Improvements in Detection of Light Precipitation with the Global Precipitation Measurement Dual-Frequency Precipitation Radar (GPM DPR). *Journal of Atmospheric and Oceanic Technology*, *33*(4), 653–667. https://doi.org/10.1175/JTECH-D-15-0097.1

Hartke, S. H., Wright, D. B., Kirschbaum, D. B., Stanley, T. A., & Li, Z. (2020). Incorporation of Satellite Precipitation Uncertainty in a Landslide Hazard Nowcasting System. *Journal of Hydrometeorology*, *21*(8), 1741–1759. https://doi.org/10.1175/JHM-D-19-0295.1

Hossain, F., & Anagnostou, E. N. (2006). A two-dimensional satellite rainfall error model. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1511–1522. https://doi.org/10.1109/TGRS.2005.863866

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor

Precipitation Estimates at Fine Scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., et al. (2019, March 13). Algorithm Theoretical Basis Document (ATBD) Version 06, NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG).

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., et al. (2020). Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 67, pp. 343–353). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_19

Iguchi, T. (2020). Dual-Frequency Precipitation Radar (DPR) on the Global Precipitation Measurement (GPM) Mission's Core Observatory. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 67, pp. 183–192). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_11

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., Le, M., et al. (2018, October). GPM/DPR Level-2 Algorithm Theoretical Basis Document.

Joyce, R. J., & Xie, P. (2011). Kalman Filter–Based CMORPH. *Journal of Hydrometeorology*, *12*(6), 1547–1563. https://doi.org/10.1175/JHM-D-11-022.1

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal

Resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Khan, S., Maggioni, V., & Kirstetter, P.-E. (2018). Investigating the Potential of Using Satellite-Based Precipitation Radars as Reference for Evaluating Multisatellite Merged Products. *Journal of Geophysical Research: Atmospheres*, *123*(16), 8646–8660. https://doi.org/10.1029/2018JD028584

Kidd, Chris, & Maggioni, V. (2020). *Quantifying errors and uncertainties in satellite precipitation estimates* (other). display. https://doi.org/10.5194/egusphere-egu2020-5857

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

Kidd, Christopher, Takayabu, Y. N., Skofronick-Jackson, G. M., Huffman, G. J., Braun, S. A., Kubota, T., & Turk, F. J. (2020). The Global Precipitation Measurement (GPM) Mission. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 67, pp. 3–23). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_1

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., et al. (2017). NASA's Remotely Sensed Precipitation: A Reservoir for Applications Users. *Bulletin of the American Meteorological Society*, *98*(6), 1169–1184. https://doi.org/10.1175/BAMS-D-15-00296.1

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., et al. (2012). Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–

Based National Mosaic QPE. *Journal of Hydrometeorology*, *13*(4), 1285–1300. https://doi.org/10.1175/JHM-D-11-0139.1

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Cao, Q., Schwaller, M., & Petersen, W. (2014). Research Framework to Bridge from the Global Precipitation Measurement Mission Core Satellite to the Constellation Sensors Using Ground-Radar-Based National Mosaic QPE. In V. Lakshmi, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, et al. (Eds.), *Geophysical Monograph Series* (pp. 61–79). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118872086.ch4

Kirstetter, P.-E., Petersen, W. A., & Gourley, J. J. (2018). GPM Ground Validation Multi-Radar/Multi-Sensor (MRMS) Precipitation Reanalysis for Satellite Validation Product. NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama, U.S.A. Retrieved from https://ghrc.nsstc.nasa.gov/hydro/details/gpmmrms

Kirstetter, P.-E., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*, *144*, 191–205. https://doi.org/10.1002/qj.3243

Kirstetter, P.-E., Petersen, W. A., Kummerow, C. D., & Wolff, D. B. (2020). Integrated Multi-satellite Evaluation for the Global Precipitation Measurement: Impact of Precipitation Types on Spaceborne Precipitation Estimation. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 69, pp. 583–608). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-35798-6_7

Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Joseph Munchak, S., & Petkovic, V. (2015). The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *Journal of*

Atmospheric and Oceanic Technology, *32*(12), 2265–2280. https://doi.org/10.1175/JTECH-D-15-0039.1

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., & Wood, E. F. (2015). Inroads of remote sensing into hydrologic science during the WRR era: REMOTE SENSING. *Water Resources Research*, *51*(9), 7309–7342. https://doi.org/10.1002/2015WR017616

Li, Z., Yang, D., & Hong, Y. (2013). Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the Yangtze River. *Journal of Hydrology*, *500*, 157–169. https://doi.org/10.1016/j.jhydrol.2013.07.023

Li, Z., Yang, D., Gao, B., Jiao, Y., Hong, Y., & Xu, T. (2015). Multiscale Hydrologic Applications of the Latest Satellite Precipitation Products in the Yangtze River Basin using a Distributed Hydrologic Model. *Journal of Hydrometeorology*, *16*(1), 407–426. https://doi.org/10.1175/JHM-D-14-0105.1

Li, Z., Wright, D. B., Zhang, S. Q., Kirschbaum, D. B., & Hartke, S. H. (2020). Object-Based Comparison of Data-Driven and Physics-Driven Satellite Estimates of Extreme Rainfall. *Journal of Hydrometeorology*, *21*(12), 2759–2776. https://doi.org/10.1175/JHM-D-20-0041.1

Maggioni, V., Sapiano, M. R. P., Adler, R. F., Tian, Y., & Huffman, G. J. (2014a). An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products. *Journal of Hydrometeorology*, *15*(3), 1274–1292. https://doi.org/10.1175/JHM-D-13-0112.1

Maggioni, V., Sapiano, M. R. P., Adler, R. F., Tian, Y., & Huffman, G. J. (2014b). An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products. *Journal of Hydrometeorology*, *15*(3), 1274–1292. https://doi.org/10.1175/JHM-D-13-0112.1

Maggioni, V., Meyers, P. C., & Robinson, M. D. (2016). A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *Journal of Hydrometeorology*, *17*(4), 1101–1117. https://doi.org/10.1175/JHM-D-15-0190.1

Massari, C., & Maggioni, V. (2020). Error and Uncertainty Characterization. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 69, pp. 515–532). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-35798-6_4

Moazami, S., & Najafi, M. R. (2021). A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada. *Journal of Hydrology*, *594*, 125929. https://doi.org/10.1016/j.jhydrol.2020.125929

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., & Eichmann, A. (2012). The GEOS-5 atmospheric general circulation model: Mean climate and development from MERRA to Fortuna.

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D., et al. (2018). The PERSIANN family of global satellite precipitation data: a review and evaluation of products. *Hydrology and Earth System Sciences*, *22*(11), 5801–5816. https://doi.org/10.5194/hess-22-5801-2018

O, S., & Kirstetter, P. (2018). Evaluation of diurnal variation of GPM IMERG-derived summer precipitation over the contiguous US using MRMS data. *Quarterly Journal of the Royal Meteorological Society*, *144*(S1), 270–281. https://doi.org/10.1002/qj.3218

O, S., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., & Petersen, W. A. (2017). Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrology and Earth System Sciences*, *21*(12), 6559–6572. https://doi.org/10.5194/hess-21-6559-2017

Petracca, M., D'Adderio, L. P., Porcù, F., Vulpiani, G., Sebastianelli, S., & Puca, S. (2018). Validation of GPM Dual-Frequency Precipitation Radar (DPR) Rainfall Products over Italy. *Journal of Hydrometeorology*, *19*(5), 907–925. https://doi.org/10.1175/JHM-D-17-0144.1

Pierre-Emmanuel Kirstetter, Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., et al. (2012). Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar–Based National Mosaic QPE. *Journal of Hydrometeorology*, *13*(4), 1285–1300. https://doi.org/10.1175/JHM-D-11-0139.1

Randel, D. L., Kummerow, C. D., & Ringerud, S. (2020). The Goddard Profiling (GPROF) Precipitation Retrieval Algorithm. In V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, & F. J. Turk (Eds.), *Satellite Precipitation Measurement* (Vol. 67, pp. 141–152). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-24568-9_8

Sarachi, S., Hsu, K., & Sorooshian, S. (2015). A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products. *Journal of Hydrometeorology*, *16*(5), 2101–2117. https://doi.org/10.1175/JHM-D-15-0028.1

Scheuerer, M., & Hamill, T. M. (2015). Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions. *Monthly Weather Review*, *143*(11), 4578–4596. https://doi.org/10.1175/MWR-D-15-0061.1

Scheuerer, M., Hamill, T. M., Whitin, B., He, M., & Henkel, A. (2017). A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation: SELECTING DATES IN THE SCHAAKE SHUFFLE. *Water Resources Research*, *53*(4), 3029–3046. https://doi.org/10.1002/2016WR020133

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., et al. (2017). The Global Precipitation Measurement (GPM) Mission for Science and Society. *Bulletin of the American Meteorological Society*, *98*(8), 1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

Skofronick-Jackson, G., Kirschbaum, D., Petersen, W., Huffman, G., Kidd, C., Stocker, E., & Kakar, R. (2018). The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: reviewing four years of advanced rain and snow observations. *Quarterly Journal of the Royal Meteorological Society*, *144*(S1), 27–48. https://doi.org/10.1002/qj.3313

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., & Braithwaite, D. (2000). Evaluation of PERSIANN System Satellite–Based Estimates of Tropical Rainfall. *Bulletin of the American Meteorological Society*, *81*(9), 2035–2046. https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2

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

Tan, J., Petersen, W. A., & Tokay, A. (2016). A Novel Approach to Identify Sources of Errors in IMERG for GPM Ground Validation. *Journal of Hydrometeorology*, *17*(9), 2477–2491. https://doi.org/10.1175/JHM-D-16-0079.1

Tan, J., Petersen, W. A., Kirstetter, P.-E., & Tian, Y. (2017). Performance of IMERG as a Function of Spatiotemporal Scale. *Journal of Hydrometeorology*, *18*(2), 307–319. https://doi.org/10.1175/JHM-D-16-0174.1

Tan, J., Petersen, W. A., Kirchengast, G., Goodrich, D. C., & Wolff, D. B. (2018). Evaluation of Global Precipitation Measurement Rainfall Estimates against Three Dense Gauge Networks. *Journal of Hydrometeorology*, *19*(3), 517–532. https://doi.org/10.1175/JHM-D-17-0174.1

Tang, G. (2020). Characterization of the Systematic and Random Errors in Satellite Precipitation Using the Multiplicative Error Model. *IEEE Transactions on Geoscience and Remote Sensing*, 1–10. https://doi.org/10.1109/TGRS.2020.3028525

Tang, G., Ma, Y., Long, D., Zhong, L., & Hong, Y. (2016). Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *Journal of Hydrology*, *533*, 152–167. https://doi.org/10.1016/j.jhydrol.2015.12.008

Tang, L., & Hossain, F. (2012). Investigating the similarity of satellite rainfall error metrics as a function of Köppen climate classification. *Atmospheric Research*, *104–105*, 182–192. https://doi.org/10.1016/j.atmosres.2011.10.006

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., et al. (2009). Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research*, *114*(D24). https://doi.org/10.1029/2009JD011949

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., & Wu, H. (2013). Modeling errors in daily precipitation measurements: Additive or multiplicative?: MODELING ERRORS IN PRECIPITATION DATA. *Geophysical Research Letters*, *40*(10), 2060–2065. https://doi.org/10.1002/grl.50320

Trinh, B. N., Thielen-del Pozo, J., & Thirel, G. (2013). The reduction continuous rank probability score for evaluating discharge forecasts from hydrological ensemble prediction systems: Reduction continuous rank probability score for HEPS. *Atmospheric Science Letters*, *14*(2), 61–65. https://doi.org/10.1002/asl2.417

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., et al. (2009). A Kalman Filter Approach to the Global Satellite Mapping of Precipitation (GSMaP) from Combined Passive Microwave and Infrared Radiometric Data. *Journal of the Meteorological Society of Japan*, *87A*(November 2008), 137–151. https://doi.org/10.2151/jmsj.87A.137

Wilks, D. S. (2019). Forecast Verification. In *Statistical Methods in the Atmospheric Sciences* (pp. 369–483). Elsevier. https://doi.org/10.1016/B978-0-12-815823-4.00009-2

Wright, D. B. (2018). Rainfall Information for Global Flood Modeling. In G. J.-P. Schumann, P. D. Bates, H. Apel, & G. T. Aronica (Eds.), *Geophysical Monograph Series* (pp. 17–42). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119217886.ch2

Wright, D. B., Kirschbaum, D. B., & Yatheendradas, S. (2017). Satellite Precipitation Characterization, Error Modeling, and Error Correction Using Censored Shifted Gamma Distributions. *Journal of Hydrometeorology*, *18*(10), 2801–2815. https://doi.org/10.1175/JHM-D-17-0060.1

You, Y., Petkovic, V., Tan, J., Kroodsma, R., Berg, W., Kidd, C., & Peters-Lidard, C. (2020). Evaluation of V05 Precipitation Estimates from GPM Constellation Radiometers Using KuPR as the Reference. *Journal of Hydrometeorology*, *21*(4), 705–728. https://doi.org/10.1175/JHM-D-19-0144.1

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., et al. (2016). Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1

AGU PUBLICATIONS

Journal of Geophysical Research: Atmospheres

Supporting Information for

Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products based on GPM DPR

Zhe Li[1], Daniel B. Wright[1], Samantha H. Hartke[1], Dalia B. Kirschbaum[2], Sana Khan[2,3], Viviana Maggioni[4], Pierre-Emmanuel Kirstetter[5,6]

[1]Department of Civil and Environmental Engineering, University of Wisconsin–Madison, Madison, WI, USA

[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA

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

[4]Civil, Infrastructure and Environmental Engineering Department, George Mason University, Fairfax, VA, USA

[5]School of Meteorology and School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK, USA

[6]NOAA/National Severe Storms Laboratory, Norman, OK, USA

Version: JUNE 07, 2021

Corresponding author: Dr. Zhe Li (zli875@wisc.edu)

Contents of this file

Table S1. The contingency tables for IMERG, benchmarked against GV-MRMS and 2BCMB. For each pair of the estimates, hits (top left), false alarms (top right), misses (bottom left), and correct non-detects (bottom right) are shown. The total paired data sample size over CONUS is 20,986,107.

	$P_{IMERG} \geq 0.1\text{mm h}^{-1}$	$P_{IMERG} < 0.1\text{mm h}^{-1}$
$P_{GV\text{-}MRMS} \geq 0.1\text{mm h}^{-1}$	714,440 (3.4%)[a]	444,222 (2.1%)
$P_{GV\text{-}MRMS} < 0.1\text{mm h}^{-1}$	437,897 (2.1%)	19,389,548 (92.4%)
$P_{2BCMB} \geq 0.1\text{mm h}^{-1}$	681,936 (3.2%)	301,416 (1.4%)
$P_{2BCMB} < 0.1\text{mm h}^{-1}$	470,401 (2.2%)	19,532,354 (93.1%)

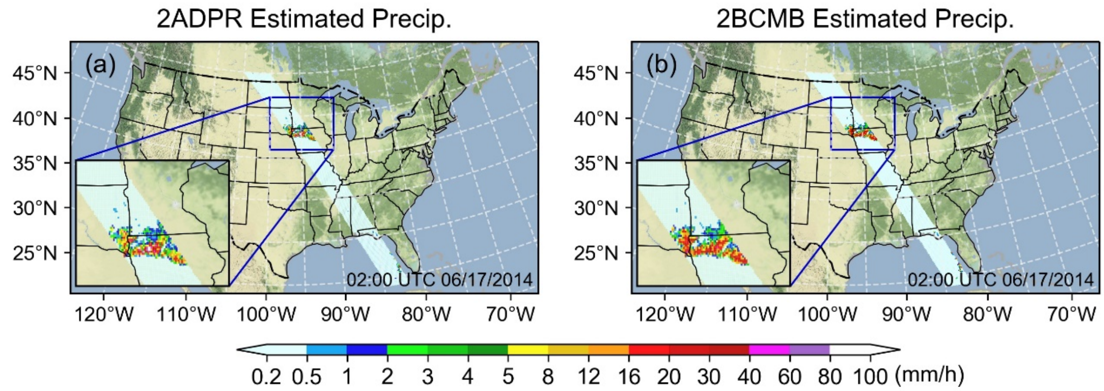

Figure S1. Coincident precipitation estimates from regridded (a) 2ADPR, and (b) 2BCMB for 02:00–03:00 UTC 17 June 2014. The swath coverage and retrieved precipitation spatial pattern of the two DPR derived products are similar, though 2BCMB shows enhanced precipitation intensities (which are closer to GV-MRMS observations, as shown in Figure 1b in the main article).

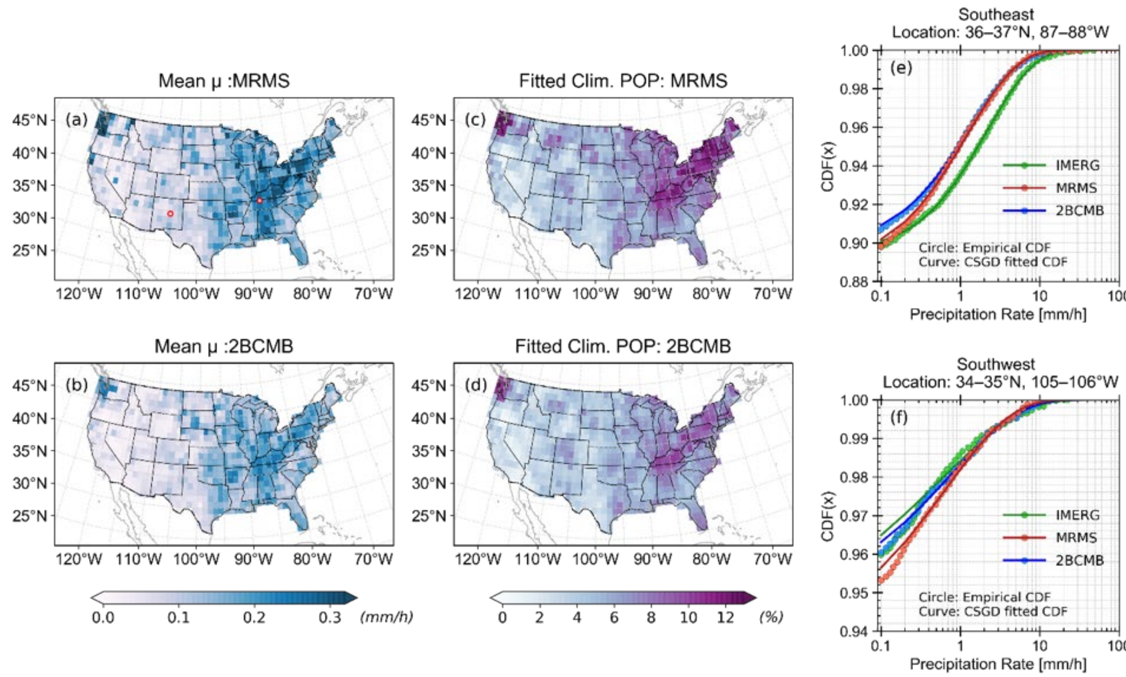

Figure S2. The spatial maps of fitted (a-b) climatological CSGD mean parameter μ and (c-d) POP, and (e-f) the comparison of empirical CDFs (markers) and climatological CSGD fitted CDFs (lines) based on the coincident IMERG, GV-MRMS, and 2BCMB training data samples within the 1°×1° boxes from the Southeast and Southwest CONUS, respectively. The locations of the two boxes are indicated by red circles in (a).

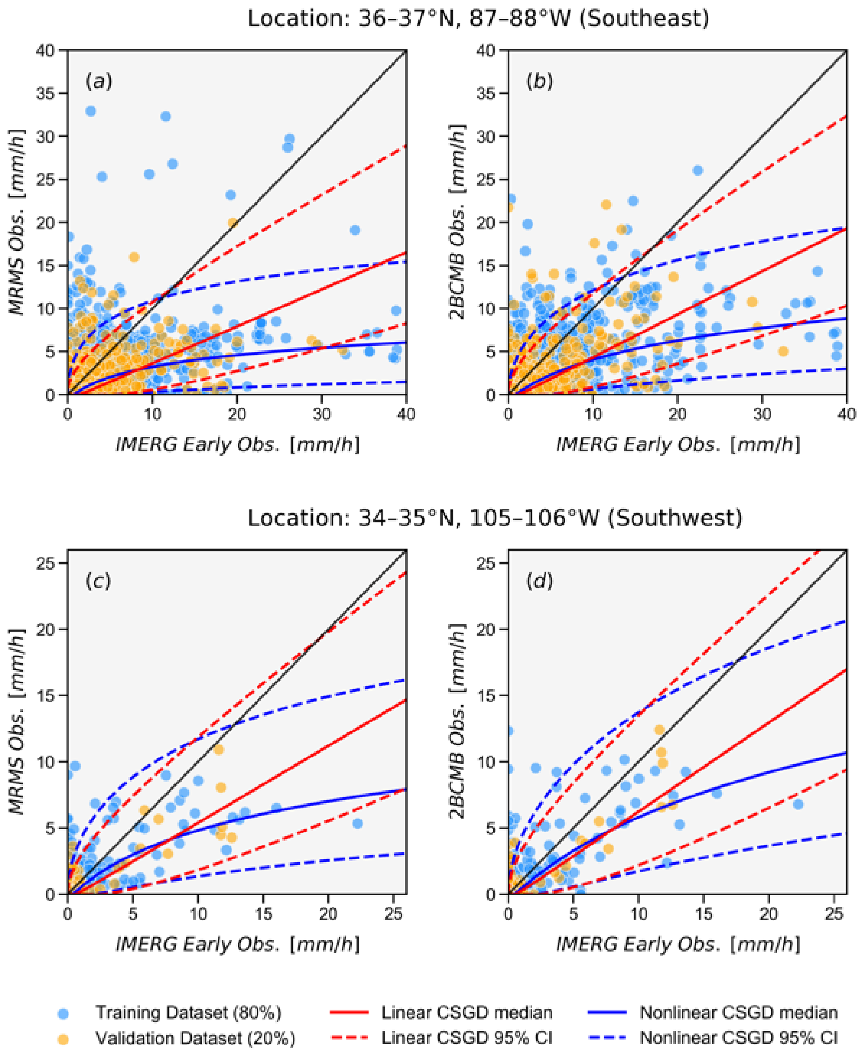

Figure S3. Linear (red lines) and nonlinear (blue lines) conditional CSGD models for (a, b) the Southeast 1°×1° box, (c, d) Southwest 1°×1° box, trained and compared against GV-MRMS (left panels) and 2BCMB (right panels). See Fig. 4 for identical results, but plotted on log scales.