STREAM-Sat: A Novel Near-Realtime Quasi-global Satellite-Only Ensemble Precipitation Dataset

Kaidi Peng[1], Daniel B. Wright[1], Yagmur Derin[1], Samantha H. Hartke[2], Zhe Li[3], Jackson Tan[4,5]

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

[2]National Center for Atmospheric Research, Boulder, CO, USA.

[3]Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO, USA.

[4]NASA Goddard Space Flight Center, Greenbelt, MD USA.

[5]University of Maryland, Baltimore County, Baltimore, MD, USA.

Corresponding author: Kaidi Peng (kaidi.peng@wisc.edu)

Key Points:

- A quasi-global precipitation ensemble is derived entirely from satellite multisensor estimates, independent of ground-based measurements.
- The proposed method is also applicable to any fine temporal resolution gridded precipitation dataset and can work in near-realtime.
- The new dataset has an unprecedented combination of accuracy, resolution, and time latency compared to other global precipitation datasets.

Abstract

Satellite-based precipitation observations provide near-global coverage with high spatiotemporal resolution in near-realtime. Their utility, however, is hindered by oftentimes large errors that vary substantially in space and time. Since precipitation uncertainty is, by definition, a random process, probabilistic expression of satellite-based precipitation product uncertainty is needed to advance their operational applications. Ensemble methods, in which uncertainty is depicted via multiple realizations of precipitation fields, have been widely used in other contexts such as numerical weather prediction, but rarely in satellite contexts. Creating such an ensemble dataset is challenging due to the complexity of errors and the scarcity of "ground truth" to characterize it. This challenge is particularly pronounced in ungauged regions, where the benefits of satellite-based precipitation data could otherwise provide substantial benefits. In this study, we propose the first quasi-global (covering all continental land masses within 50°N-50°S) satellite-only ensemble precipitation dataset, derived entirely from NASA's Integrated Multi-SatellitE Retrievals for Global Precipitation Measurement (IMERG) and GPM's radar-radiometer combined precipitation product (2B-CMB). No ground-based measurements are used in this generation and it is suitable for near-realtime use, limited only by the latency of IMERG. We compare the results against several precipitation datasets of distinct classes, including global satellite-based, rain gauge-based, atmospheric reanalysis, and merged products. While our proposed approach faces some limitations and is not universally superior to the datasets it is compared to in all respects, it does hold relative advantages due to its combination of accuracy, resolution, latency, and utility in hydrologic and hazard applications.

1 Introduction

Accurate and timely precipitation measurements are crucial for monitoring and assessing hydrometeorological hazards (Liao et al., 2010; Liu, Guo, et al., 2018); this imperative is growing with continued climate warming and its impacts on rainfall and rainfall-related hazards (Fowler et al., 2021). These hazards include extreme rainfall (e.g., Ayat et al., 2022), floods (e.g., Wilhelm et al., 2022), rainfall-triggered landslides (e.g., Kirschbaum et al., 2020), debris flows (e.g., Pan et al., 2018), crop failures (e.g., Sloat et al., 2018) and waterborne disease outbreaks (e.g., Exum et al., 2018).

Due to its high variability in space and time, precipitation is difficult to measure, particularly at the global scale (Kidd et al., 2017; Wright, 2018). Dense ground-based rain gauge and radar networks can capture the variability of rainfall at very high resolution, but such networks are limited in coverage and are typically found only in wealthy countries and urban areas. Satellites, on the other hand, deploy a range of sensors operating across various channels to estimate different types of precipitation and precipitation-related processes. The relatively wide field of view of individual sensors, as well as the near-global coverage provided by the international "constellation" of relevant earth observing satellites, is particularly promising for observing highly heterogeneous rainfall patterns. However, the periodic nature of orbiting platforms' sampling, the heterogeneity in sensor characteristics, and the indirect nature of their measurements from passive microwave and infrared sensors lead to serious challenges in the creation of global precipitation datasets such as NASA's Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (IMERG; Huffman et al., 2019), while contributing to oftentimes large errors in the precipitation detection and quantification. Because of this, ground stations are often used to "post-process" satellite-based estimates and reduce errors (Funk et al., 2015; Huffman et al., 2020).

There are alternatives and supplements for satellite-based precipitation products for creating global precipitation datasets. One is global land-ocean-atmosphere weather forecast or reanalysis systems. A recent example of the latter is the fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA5; Hersbach et al., 2020). Such datasets typically assimilate precipitation-relevant satellite data, can provide comparable or sometimes better accuracy than satellite-only precipitation datasets, and may include an ensemble that reflects aspects of estimation uncertainty. On the other hand, reanalysis products tend to suffer from insufficient parameterization of key precipitation processes including convection. Their performance also relies on the assimilation of station data, which can pose latency challenges for real-time applications. Weather forecast systems, on the other hand, tend to lack the consistent long-term "hindcasts" necessary for many applications.

Additional sources of global precipitation data include gauge-based datasets, which tend to suffer from poor resolution, low accuracy where gauges are scarce, and high latency (Hartke & Wright, 2022; Kidd et al., 2017; Wright, 2018). In principle, the highest accuracy of precipitation estimates can be achieved through leveraging the strengths of multiple datasets via data merging (e.g., MSWEP; Beck, Pan, et al., 2019). This merging, however, complicates the diagnosis of error and its output is limited by the resolution and latency of its "slowest" input data source.

The degree of uncertainty present in global-scale precipitation data seriously hinders its uptake in real-world applications, since it propagates through hydrologic and other types of environmental models (Falck et al., 2015; Hartke et al., 2020; Hartke et al., 2023; Schreiner-McGraw et al., 2020). Consequently, there is a critical need to quantify this uncertainty in ways that are compatible with such applications. As with other inherently random processes, precipitation uncertainty is best described probabilistically (i.e., via probability distributions). Probabilistic methods can depict irreducible uncertainty arising from a lack of sufficient explanatory information. The deterministic input requirements of virtually all water prediction models, however, are at odds with such probabilistic depictions. Translating such depictions into an ensemble that conveys the uncertainty through a number of members thus constitutes the most direct—and perhaps the only—way to make uncertainty information "digestible" by such models (Hartke, Wright, et al., 2022).

Nonetheless, global ensemble precipitation datasets are rare. The 10-member ensemble component (3-hourly and 0.5°) of ERA5 reflects the relative and random uncertainty associated with the data assimilation process but does not necessarily provide a broader representation of overall uncertainty (Hersbach et al., 2020). The Ensemble Meteorological Dataset for Planet Earth (EM-Earth) merges ERA5 and the station-based Serially Complete Earth (SC-Earth) dataset to generate a 25-member global land precipitation ensemble dataset (Tang et al., 2022). Both systematic and random errors in ERA5 are included in EM-Earth. The number of ensemble precipitation datasets covering regional-to-continental domains is growing, including gridded Ensemble Precipitation Estimates for North America (Newman et al., 2015; Newman et al., 2019; Newman et al., 2020; Tang et al., 2021), a radar ensemble generator designed for usage in the European Alps (Germann et al., 2009), and the Europe-wide 100-member E-OBS precipitation dataset (Cornes et al., 2018). Spatially COherent Probabilistic Extended Climate dataset is a 25-member ensemble of 142-year daily high-resolution reconstructions of precipitation over France via stochastic downscaling (Caillouet et al., 2019). The 25-member 6-hour High-Resolution

Ensemble Precipitation Analysis (HREPA) covers Canada and the northern part of the contiguous United States (Khedhaouiria et al., 2020).

Most of the aforementioned datasets rely to some extent on rain gauges to constrain errors; such constraints are weak or nonexistent over ungauged regions such as the Global South or complex terrain. Even when gauges are present, a combination of technical and geopolitical constraints limits access—at least in real-time—to such rain gauge observations in many jurisdictions. Meanwhile, quantification of uncertainty in satellite-based precipitation products—and in particular how this uncertainty varies over a range of spatiotemporal scales and across storm systems—has proven elusive (Guilloteau et al., 2021, 2022; Hartke, Wright, et al., 2022; Huffman et al., 2020; Li et al., 2023), again largely because of a paucity of higher-fidelity "ground truth." As such, no "observation-oriented" (as opposed to numerically-forecasted) global ensemble precipitation dataset currently exists for near-realtime applications.

This study seeks to fill this gap by proposing a quasi-global (in this case, all continental land masses within 50°N-50°S, though our method could be employed over the oceans and in high-latitude regions, subject to challenges described below) satellite-only ensemble precipitation dataset that, while evaluated over retrospective period here, could be deployed in near-realtime applications, limited only by the latency of satellite inputs. Development of the dataset requires three key components: (1) a gridded input precipitation dataset—IMERG Early V06B in this case, but most subdaily global datasets would be suitable including high-resolution reanalysis such as ERA5; (2) a probabilistic depiction—typically referred to as an error model—of uncertainty for that input dataset at the individual grid cell scale; and (3) a means of "connecting" that uncertainty across grid cells and time steps to generate ensemble realizations of precipitation that reflect the uncertainty in the input dataset over large areas. More detailed explanations of how we approach components 2 and 3, and broader explanations of challenges and relevant past studies, are described in Li et al. (2023) and Hartke, Wright, et al. (2022), respectively.

While various error models have been proposed for quantifying grid cell-scale uncertainty in satellite-based precipitation products (e.g., AghaKouchak et al., 2012; Guilloteau et al., 2022; Maggioni et al., 2014; Sorooshian et al., 2015; Tan et al., 2016; Wright et al., 2017), these have generally relied on ground-based observations for parameterization, precluding them from global application. Li et al. (2023), in contrast, used observations from the GPM Core Observatory satellite and in particular its dual-frequency precipitation radar (DPR) in place of ground-based observations to parameterize an error model over the contiguous United States (CONUS), demonstrating reasonable performance while opening the door to global-scale uncertainty quantification. Likewise, while multiple studies have shown ways of connecting precipitation uncertainty structures across multiple grid cells and time steps (e.g., AghaKouchak et al., 2010; Caseri et al., 2016; Germann et al., 2009; Hossain et al., 2006; Leblois et al., 2013; Newman et al., 2015), Hartke, Wright, et al. (2022) made the important step of inferring such structures contemporaneously from the input dataset (IMERG in both that study and the present one) itself. This conceptual advance, and its stochastic implementation via the Space-Time Rainfall Error and Autocorrelation Model (STREAM), allowed for robust depictions of uncertainties across spatiotemporal scales, including nonstationary (i.e., location-dependent) and anisotropic (i.e., direction-dependent) features that typify precipitation uncertainty structures but that existing frameworks struggled to capture or ignored entirely.

This study unifies the advances of Li et al. (2023) and Hartke, Wright, et al. (2022) to create a quasi-global satellite-only ensemble precipitation dataset hereafter referred to as STREAM-Sat.

We first performed ground validation of STREAM-Sat against NCEP/EMC Stage IV (Du, 2011) over part of the contiguous United States (CONUS). Then we further compared STREAM-Sat with five other global precipitation datasets. These datasets are IMERG Early (which, as will be seen, is also an input to STREAM-Sat), Multi-Source Weighted-Ensemble Precipitation (MSWEP V2.8), The Climate Hazards group Infrared Precipitation with Stations (CHIRPS V2.0), ERA5, and EM-Earth. As described further below, each of these datasets is generally considered as state-of-the-art, while each represents a different class of dataset (i.e., satellite, reanalysis, merged, etc.). Each dataset features relative advantages and disadvantages in terms of accuracy, resolution, latency, and availability of uncertainty information (or lack thereof). A challenge when evaluating precipitation data at a global scale is the lack of adequate "ground truth." One option is to assemble gauge datasets from different parts of the world (e.g., Derin et al., 2016). This is laborious and still leaves vast land areas unexamined (Kidd et al., 2017). We instead opt to compare STREAM-Sat to other global alternatives. However, caution is needed with such comparisons, since validation metrics (e.g., mean absolute error) may reflect inadequacies from both datasets, rather than just one. The objective of this study, therefore, is to understand and establish the performance of STREAM-Sat relative to its peers—the other benchmark global precipitation datasets—rather than in an absolute sense. As will also be seen, while STREAM-Sat is not superior in all respects, it holds distinct and important advantages over the other datasets. This study will try to answer the following questions:

1. How does STREAM-Sat compare to IMERG Early and other global precipitation datasets?
2. What are the factors that influence STREAM-Sat performance over different regions?
3. Can STREAM-Sat capture patterns of precipitation structure at varying spatiotemporal resolutions?

The datasets used in this study are described in Section 2. The uncertainty estimation and ensemble generation methods, as well as evaluation metrics, are introduced in Section 3. Section 4 presents the results, following discussion in Section 5. Conclusions are provided in Section 6.

2 Data

2.1 IMERG Early V06B

The IMERG algorithm is designed to merge, intercalibrate, and interpolate all available satellite passive microwave (PMW) retrievals and microwave-calibrated infrared (IR) satellite estimates to produce 30-min, 0.1° gridded precipitation estimates over the majority of the Earth's surface. PMW-only precipitation estimates are retrieved using the Goddard profiling algorithm. Gaps between the instantaneous PMW observations are interpolated via a morphing technique using motion vectors calculated from total precipitable water vapor from MERRA-2 or GEOS-FP (Tan et al., 2019). The forward- and backward-propagated PMW estimates are composited via a Kalman filter, with PMW-calibrated IR precipitation estimates from the PERSIANN algorithm (Hong et al., 2004). IMERG runs three times under different time latency: Early (~4 hr after observation time; used in this study to emphasize the near-realtime capabilities of our approach), Late (~14 hr after observation time), and Final (~3.5 months after the observation month). IMERG Early only involves forward propagation of sensor observations and uses about 95% of the input data, on top of other minor calibration differences. At the time of writing, the latest version of IMERG Early Run is V06B. The data field "precipitationCal" was used.

2.2 GPM 2B-CMB V07A

The GPM Core Observatory satellite carries the DPR, which consists of Ku-band (13.6 GHz) and Ka-band (35.5 GHz) precipitation radar as well as the multi-channel microwave radiometer GPM Microwave Imager (GMI, 10 GHz to 183 GHz). These sensors serve as calibration standards for other members in the GPM satellite constellation (Huffman et al., 2020). To quantify grid cell-scale uncertainty in IMERG V06B, we followed Li et al. (2023) in using the GPM L2B "combined" (i.e. DPR and GMI) product 2B-CMB V07A Normal Scans (NS) from Combined Radar-Radiometer Algorithm (CORRA; Iguchi et al., 2018; Olson, 2022) with a native resolution of approximately 5 km footprint along the swath. Due to the 65° inclination angle of the GPM Core Observatory orbit, 2B-CMB covers up to a latitude range of around 67° N/S. The data field "estimSurfPrecipTotRate" was used.

2B-CMB was mapped onto the IMERG native grid by averaging all the DPR footprint estimates falling within a 0.1° grid cell, and then matched into the nearest 30-minute IMERG observation interval (Li et al., 2023). Due to the significant underestimation of snowfall in both IMERG and DPR products (Behrangi et al., 2018; Skofronick-Jackson et al., 2019), only regridded coincident data where the IMERG Early data field "probabilityLiquidPrecipitation" is greater than 90 (percent) were used. Data with "flagHail" from the GPM DPR Precipitation Profile L2A (2ADPR) dataset (Iguchi et al., 2021) equal to 1 were also excluded.

2.3 Benchmark Precipitation Datasets

Three deterministic global precipitation datasets—MSWEP V2.8, CHIRPS V2.0, and ERA5—and one ensemble dataset—EM-Earth—were used for comparison against STREAM-Sat globally. Information about these datasets is summarized in Table 1, which highlights their varied resolutions and latencies. These are commonly-used global precipitation datasets and generally considered as state-of-the-art. As a merged dataset, MSWEP has been shown to have the highest correlation with Stage IV, while ERA5 has the best performance among reanalysis-based datasets (Beck, Pan, et al., 2019). CHIRPS mainly relies on infrared sensors and has less overlap with IMERG in terms of input sources compared to other popular satellite-based precipitation products such as CMORPH (Xie et al., 2017) and GSMaP (Kubota et al., 2007). Also shown in Table 1 is Stage IV. Stage IV radar-gauge product has high resolution and high accuracy over the eastern CONUS relative to the global precipitation products considered here, which also has been previously used for evaluating the accuracy of satellite-based precipitation datasets (e.g., AghaKouchak et al., 2011; Li et al., 2020; Nelson et al., 2016).

The global benchmark precipitation datasets are henceforth referred to as "comparison datasets," while ground validation Stage IV is referred to as "validation reference." 2B-CMB is used to train our error model and is henceforth referred to as "training reference."

2.4 Ancillary Data

To understand the influence of physiographic factors on the performance of STREAM-Sat, the following datasets were also used: Global Bathymetry and Topography at 15 Arc Sec (SRTM15+; Tozer et al., 2019), 0.5° seasonal temperature covering all land areas (excluding Antarctica) from the Climate Research Unit (CRU TS v. 4.07; Harris et al., 2020) and a 0.5° Köppen–Geiger Climate Zone Classification (Rubel, 2010).

237 **Table 1.** *Precipitation Datasets Compared in this Study*

Name	Details	Resolution; coverage	Latency	Reference
Stage IV	Regional hourly/6-hourly multi-sensor (radar and/or gauges) precipitation analyses produced by the 12 River Forecast Centers with some manual quality control; mosaicked into a national/CONUS product at NCEP	4km, hourly/6-hourly; CONUS	The second week of the following month	Du (2011)
IMERG Early V06B	See Section 2.1	0.1°, half-hourly; global	4 hours	Huffman et al. (2020)
MSWEP V2.8	MSWEP merges global gauge observations, satellite estimates (IMERG and GridSat), and model output (ERA5). Historical MSWEP is used here, as opposed to MSWEP-NRT.	0.1°, 3-hourly; global	MSWEP-NRT has 1.5-4.5-hour latency, and it is progressively upgraded in 15 days. Historical MSWEP is available up to 2020.	Beck, Wood, et al. (2019)
CHIRPS V2.0	CHIRPS mainly relies on GridSat and CPC TIR and uses TMPA 3B42 pentadal precipitation for TIR observation calibration. A monthly precipitation climatology (CHPclim) and in-situ station data correction are also integrated.	0.05°, daily; 50°S-50°N, land	The third week of the following month.	Funk et al. (2015)
ERA5	ERA5 combines model predictions with observations via data assimilation.	0.25°, hourly; global	5 days	Hersbach et al. (2020)
EM-Earth	25-member global land precipitation ensemble merges ERA5 and the station-based Serially Complete Earth (SC-Earth) dataset.	0.1°, daily; global land except for the Antarctic	Available up to 2019	Tang et al. (2022)
STREAM-Sat	User-defined number of ensemble members conditioned on IMERG Early V06B and corrected by 2B-CMB (see Sections 3). 20 members were generated in this study.	0.1°, half-hourly; 50°S-50°N, land	4 hours	This paper, Hartke, Wright, et al. (2022) and Li et al. (2023)

3 Methods

3.1 The CSGD Error Model

The censored shifted gamma distribution (CSGD) satellite precipitation error modeling framework (Hartke et al., 2020; Li et al., 2023; Wright et al., 2017) was used to model pixel-scale uncertainty of IMERG Early. The CSGD error model gives the probability distribution of what the true rainfall might have been for every IMERG Early 0.1°, 30-min precipitation estimate over land. It does so by identifying and removing systematic bias and describing the random error via the CSGD.

CSGD modifies the conventional two-parameter gamma distribution with left-censoring. By doing so the error model can depict both precipitation occurrence and magnitude. This framework is flexible, capturing the probability of precipitation, central tendency (i.e. median, mean), and uncertainty using three parameters (Scheuerer et al., 2015): the mean μ, standard deviation σ, and shift δ. δ allows the model to describe the probability of both zero and positive precipitation, with the cumulative distribution function (CDF) evaluated at zero being equal to the probability of zero precipitation. The CSGD error model generates conditional distributions of rainfall via a nonlinear regression, whereby the three parameters can be conditioned on the observed satellite precipitation and other time-varying covariates, such as the Wetted Area Ratio (WAR; the percentage of pixels with positive precipitation in a box centered on each pixel) used in this study. In Hartke, Wright, et al. (2022), WAR was shown to improve the detection performance of CSGD. The same WAR radius (ten pixels) as in Hartke, Wright, et al. (2022) was used in this study. The CSGD-based error model is written as:

$$\mu(t) = \frac{\mu}{\alpha_1}\log\left\{\left(e^{\alpha_1}-1\right)\left[\alpha_2+\alpha_3\frac{P(t)}{\bar{P}}+\alpha_5\frac{C(t)}{\bar{C}}\right]+1\right\} \tag{1}$$

$$\sigma(t) = \alpha_4\sigma\sqrt{\frac{\mu(t)}{\mu}} \tag{2}$$

$$\delta(t) = \delta \tag{3}$$

where $\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5$ are regression parameters. $P(t)$ and $C(t)$ are IMERG precipitation and WAR at time t, respectively. $\bar{P}$ and $\bar{C}$ are corresponding climatological mean values at the same location. $\mu,\sigma,$ and δ are derived from a climatological fitting of a CSGD at each location. $\mu(t)$, $\sigma(t)$ and $\delta(t)$ define a distinct CSGD at a specific time t at each location conditioned on $P(t)$ and $C(t)$.

Coincident observations from IMERG and 2B-CMB from 2018-2021 were used to calibrate the error model. The calibration was implemented in each 1° by 1° box over land between 50°N and 50°S. If the data sample size within a box was less than 15,000 or if the hit fraction (i.e., when both IMERG and 2B-CMB observe positive precipitation) was lower than 1%, we expanded the box by 0.5° in all directions to expand the sample size prior to training the model. The error model primarily assesses the behavior of liquid precipitation as well as nonoccurrence of precipitation, as opposed to solid or mixed-phased precipitation (see Sections 2.2 and 5.2 for further details).

3.2 Space-Time Rainfall Error and Autocorrelation Model (STREAM)

While the CSGD error model provides the uncertainty of IMERG precipitation at its native resolution, generating usable ensemble information is not straightforward, since the error in IMERG and other datasets exhibits spatial and temporal autocorrelation. For example, if one were to generate a possible realization by sampling from CSGD at each grid box independently, the output would be unrealistic, as it does not account for correlation between neighboring grid boxes. Therefore, a method for sampling from the CSGDs at every location and time step that accounts for this autocorrelation of the error is needed.

STREAM combines uncalibrated, anisotropic, and spatially nonstationary modeling of satellite precipitation spatiotemporal correlation with the CSGD error model (Section 3.1) to stochastically generate precipitation ensembles that resemble "ground truth" precipitation (Hartke, Wright, et al., 2022). A flowchart of STREAM/STREAM-Sat is shown in Figure 1, and a highly abbreviated explanation is provided here. See Hartke, Wright, et al. (2022) for a more detailed explanation.

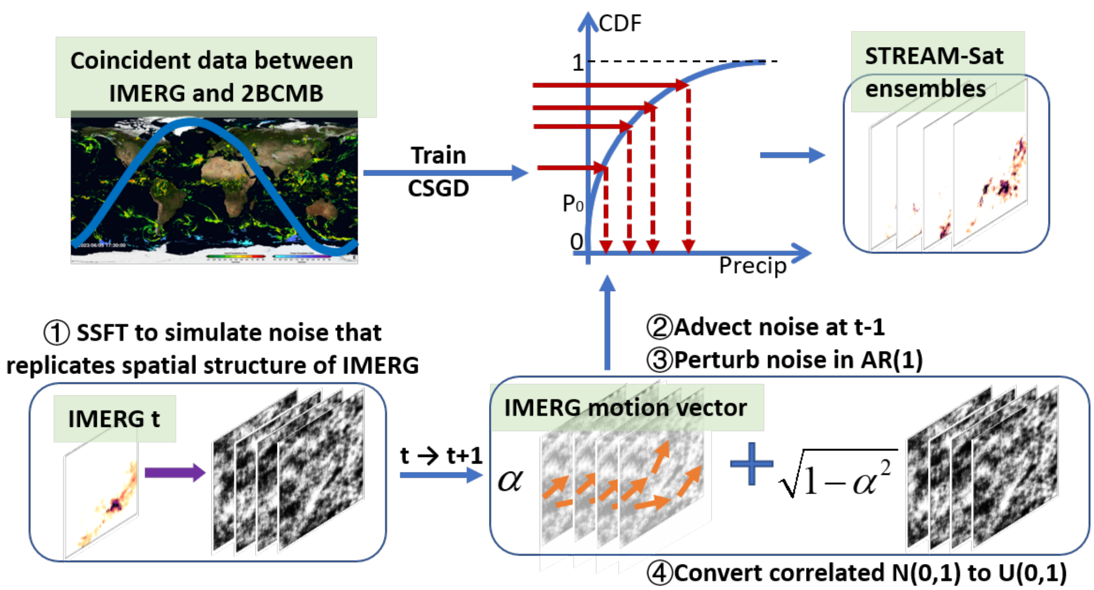


Figure 1. STREAM-Sat flowchart and its connection with the CSGD error model. Each step of STREAM is labeled in order. Green boxes highlight the input data and final output. α is the AR(1) coefficient. N(0,1) denotes the standard Gaussian normal distribution and U(0,1) denotes uniform distribution. CDF refers to the cumulative distribution function of the CSGD, which connects U(0,1) to corrected precipitation. P_0 is the probability of non-precipitation.

A short space Fourier transform (SSFT) is applied to replicate space autocorrelation of IMERG within a moving spatial window (128 by 128 pixels in this study) using a normal Gaussian noise field (Nerini et al., 2017; Pulkkinen et al., 2019), while a first-order autoregression model (AR (1)) introduces a temporally correlated "shock term." A semi-Lagrangian advection of the noise field based on IMERG motion vectors further connects the noise across space and time, consistent with the structure of the original IMERG fields. The AR(1) coefficient is calculated as the linear correlation between two consecutive IMERG observations within the same moving spatial window used for the SSFT. Standard normal and CSGD quantile functions convert the autocorrelated normal Gaussian noise field to precipitation. Additional ensembles are created by reseeding the initial noise and shock terms.

In this study, STREAM was used to create twenty ensemble members with the same resolution as IMERG (i.e., 0.1° and half-hour) for one year (2017). Though any number of members could be created, twenty is chosen in this paper to conserve storage space and since prior work has suggested that this setting is adequate, at least for certain applications (e.g., Hartke et al., 2023). As previously mentioned, the resulting ensemble dataset is referred to as STREAM-Sat.

3.3 Performance Metrics

The continuous ranked probability skill score (CRPS; Thorarinsdottir et al., 2013) has been used widely in probabilistic weather forecasting. CRPS considers both the expected value of the absolute error and the sharpness of the probabilistic prediction. The discrete expression of CRPS is used to evaluate the precipitation ensembles. It can be written as

$$CRPS\left(F_{ens},x\right)=\frac{1}{M}\sum_{m=1}^{M}\left|x_m-x\right|-\frac{1}{2M^2}\sum_{m=1}^{M}\sum_{n=1}^{M}\left|x_m-x_n\right| \tag{4}$$

where F_{ens} is ensemble distribution with size M (Grimit et al., 2006). x_m and x_n are the individual ensembles and x is the reference deterministic value. When the ensemble size is one, CRPS reduces to the well-known deterministic metric mean absolute error (MAE). This useful feature allows comparison between probabilistic and deterministic estimates via CRPS and MAE. Lower CRPS (or MAE) indicates better performance; their values can range from zero to positive infinity.

The containing ratio (CR) is another commonly used metric in ensemble evaluation. It calculates the percentage of instances in which a deterministic reference (i.e. "ground truth") falls within the ensemble spread:

$$CR=\frac{1}{T}\sum_{t=1}^{T}I_t\left(x_{\min}\le P_{ref}\le x_{\max}\right) \tag{5}$$

where $x_{\max}$ and $x_{\min}$ are the largest and smallest value of the predicted ensembles, and P_{ref} is the reference precipitation. I_t is an indicator that equals one when the specific criterion in the bracket is fulfilled at time t. It is zero under any other condition. T is the total number of time steps to be evaluated at a location. The range for CR is zero to one, with the latter being optimal.

For evaluation, data were classified into four categories: (1) Hits, in which both the reference (e.g., Stage IV) and the estimate (e.g., IMERG) report the precipitation; (2) Misses, in which reference reports the precipitation but estimate does not; (3) False alarms, in which reference does not report precipitation but the estimate does; and (4) Correct Non-detects, in which both the reference and estimate do not report the precipitation. The classification for STREAM-Sat is based on IMERG (i.e., if IMERG reports a hit, STREAM-Sat is classified as a hit) to evaluate STREAM-Sat conditioned on different types of IMERG error.

The method proposed by Guilloteau et al. (2021) is used here to examine the spatial anisotropy at different scales for the reference and STREAM-Sat. Fourier power spectral density (PSD) from a three-dimensional Fourier transform allows spectral space–time analysis. Two-dimensional spatial PSD value exhibits the energy associated with different spatial sampling distances, while the preferred directionality reveals the anisotropy of the precipitation fields.

4 Results

4.1 Evaluation against Stage IV

We examined the performance of STREAM-Sat and other benchmark precipitation products against Stage IV over 40°N-30°N, 105°W-85°W (i.e., the eastern/southeastern United States where Stage IV is relatively reliable). Specifically, we estimated CRPS for STREAM-Sat and EM-Earth and MAE for the deterministic datasets (Figure 2). All datasets were regridded to daily and 0.25° (the finest common resolution among the datasets) for this comparison. MSWEP's standout performance is probably attributed to its incorporation of gauge observations, which are plentiful in this region. STREAM-Sat's total CRPS is 19% higher than MSWEP's (compared to 100%, 90% and 135% for CHIRPS, ERA5, and IMERG, respectively), albeit without the benefit of ground-based observations. STREAM-Sat demonstrates an improvement over IMERG in both precipitation intensity and correct detection, as evidenced by the reduced CRPS by 53% in hits, 27% in misses, and 30% in false alarms. STREAM-Sat is close to, if not better than, other benchmark global precipitation datasets in terms of CRPS vs. MAE. Compared to the STREAM-Sat ensemble mean, which is the result of systematic bias removal, the full STREAM-Sat ensemble shows a further reduction of total CRPS by 26%. This is similar to EM-Earth, in which total CRPS from the ensemble is reduced by 25% compared with the ensemble mean. These results highlight the importance of probabilistic ensemble-based representations of precipitation uncertainty.

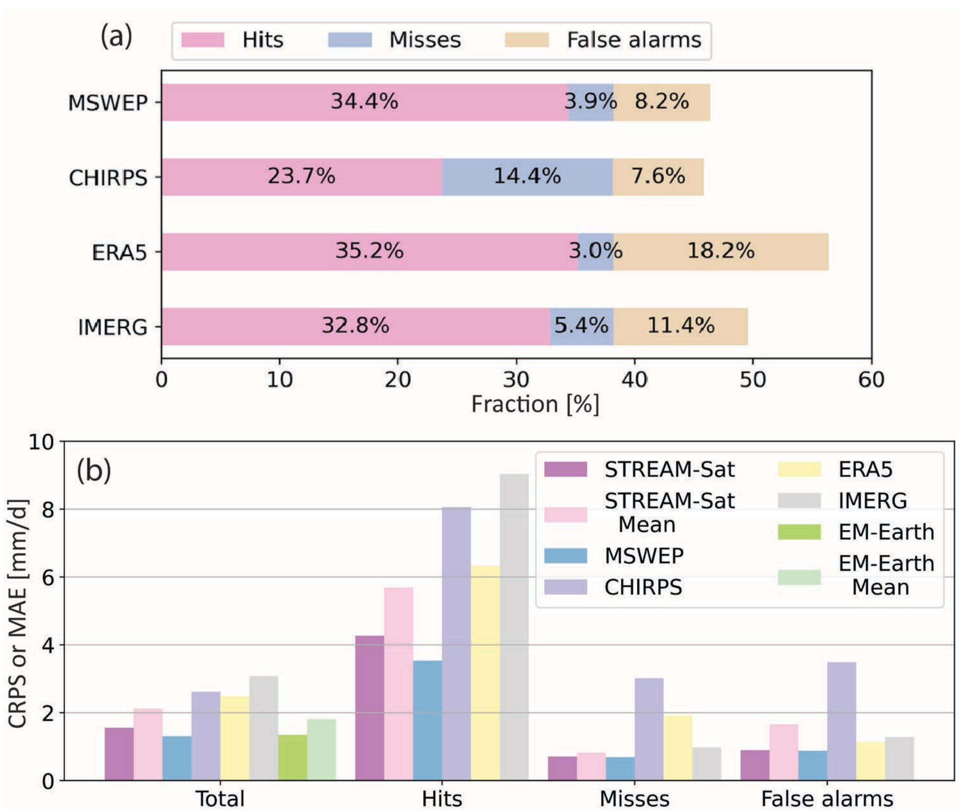


Figure 2. Validation of global precipitation datasets against Stage IV for 40°N-30°N, 105°W-85°W. (a) The fraction of hits, misses, and false alarms (the remainder of cases being correct non-detects) in deterministic datasets. (b) CRPS for STREAM-Sat and EM-Earth (total only) and MAE

for other deterministic datasets. The MAE for STREAM-Sat and EM-Earth ensemble mean are also shown. All data were regridded to daily 0.25° resolution using mass conserving interpolation.

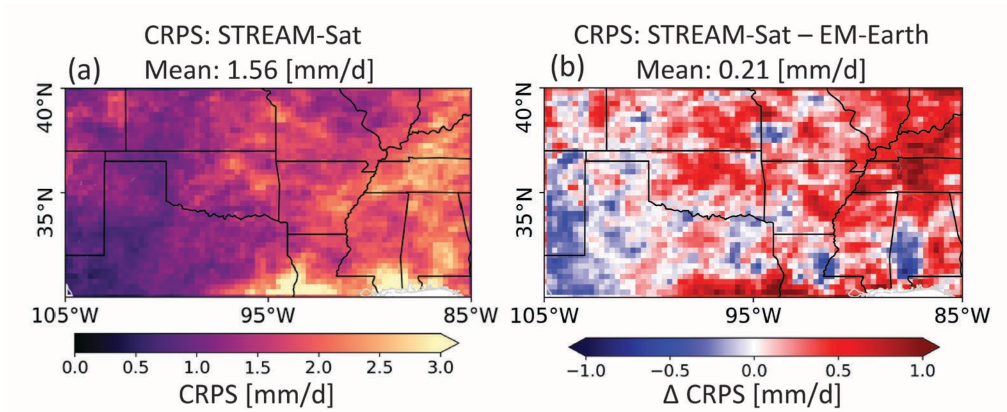


Figure 3. Comparison of STREAM-Sat and EM-Earth against Stage IV for 40°N-30°N, 105°W-85°W. (a) CRPS for STREAM-Sat ensembles. (b) CRPS of STREAM-Sat minus CRPS of EM-Earth.

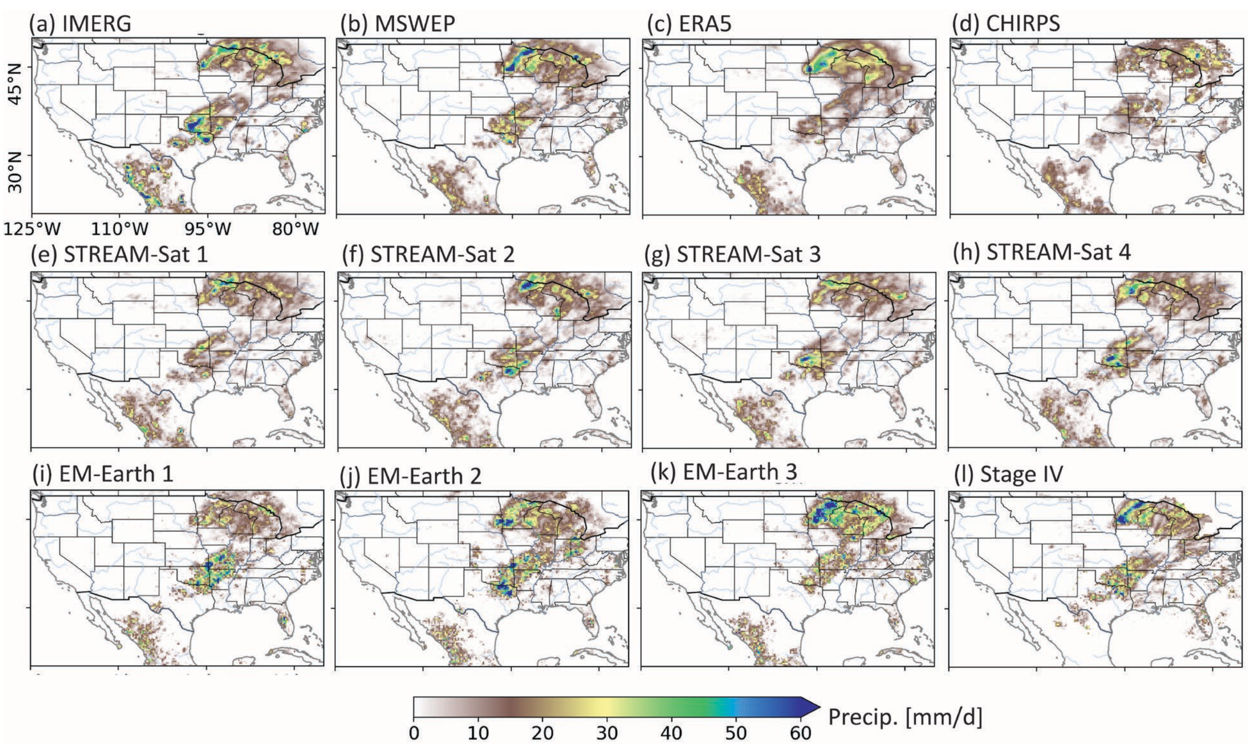
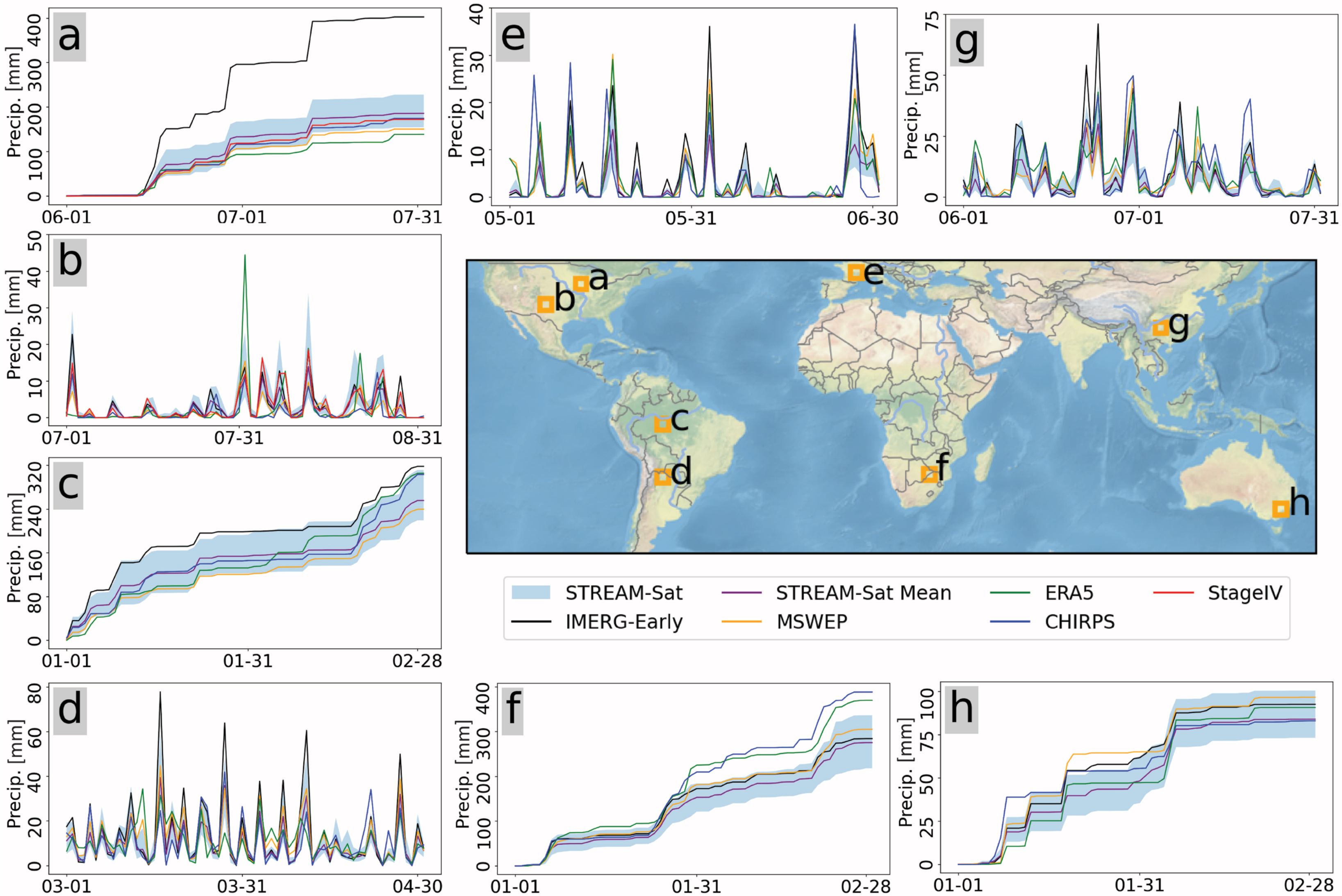


Figure 4. Precipitation over part of North America on 17 August 2017 estimated by (a) IMERG Early, (b) MSWEP, (c) ERA5, and (d) CHIRPS. Four STREAM-Sat ensembles are shown in (e)-(h). Three EM-Earth ensembles are shown in (i)-(k). (l) Stage IV; note that Stage IV doesn't cover the entire pictured area. All data were regridded to daily and 0.25°.

EM-Earth's CRPS is lower than that of STREAM-Sat over roughly two-thirds of the study region (Figure 3). Nonetheless, that leaves one-third of the region (e.g., Texas panhandle and New Mexico) where STREAM-Sat outperforms EM-Earth (which is gauge corrected); STREAM-Sat has on average 16% higher CRPS than EM-Earth over the whole map. It should be noted that such comparison over this particular study region should be expected to favor EM-Earth (and MSWEP) due to the high gauge density, which both datasets ingest. Likewise, CHIRPS benefits from sufficient station observations (Funk et al., 2015), and ERA5 assimilates Stage IV in this region (Hersbach et al., 2020; Lopez, 2011). The reduced performance of EM-Earth in the west region of Figure 3b is likely attributed to sparser gauge density compared to farther east. It is safe to summarize that in regions with fewer high-quality ground-based observations (i.e., the vast majority of the global land surface), the accuracy of precipitation datasets that use ground-based data would be degraded relative to datasets such as STREAM-Sat that don't rely on such measurements.

A one-day "snapshot" (Figure 4) of total rainfall accumulation from all the benchmark precipitation datasets over part of North America shows general consistency in the location of precipitation systems, but high variability in their amount and structure. Some STREAM-Sat ensembles correct the overestimate in IMERG around Oklahoma, where CHIRPS and ERA5 exhibit significant underestimation. STREAM-Sat generally captures the southwest-to-northeast observed spatial anisotropy better than EM-Earth.

These results show that the various global precipitation datasets (e.g., ERA5, CHIRPS, and MSWEP) have very different errors. We cannot rely on one alone to understand and evaluate the global performance of STREAM-Sat. In the following sections, they will be used together for comparison. We reiterate that comparisons of CRPS, MAE, and other metrics must be interpreted with caution, since they may reflect inadequacies from both comparison datasets and IMERG or STREAM-Sat, rather than just one.



404

405 **Figure 5.** Time series of precipitation depth and accumulated depth for different selected locations around the globe. Four global
406 precipitation products as well as the STREAM-Sat ensemble mean and spread are shown in all time series panels. Stage IV is only
407 shown in (a) and (b). Locations are indicated on the central map. Each time series is aggregated over a 1° by 1° box.

4.2 Global Comparison of STREAM-Sat over Different Spatiotemporal Resolutions

To understand whether STREAM-Sat captures precipitation dynamics in diverse settings, it was compared against the other precipitation datasets over eight unique locations (Figure 5). These locations cover equatorial, temperate, and arid climate zones with different elevation regimes. For example, Figure 5a is classified as a snow zone based on the Koppen-Geiger climate classification with annual precipitation around 400 mm, while Figure 5d is located in the Amazon rainforest with annual precipitation above 2,000 mm. Figure 5b has an elevation of around 1,200 m. Figure 5c, 5f & 5h are located in an arid climate zone with annual precipitation around 500 mm but with elevations of roughly 300 m, 900 m and 130 m, respectively. Figure 5e is located in a warm and fully humid climate zone. Figure 5g is located in the Yungui Plateau with about 1,000m elevation and annual precipitation of 1,300mm. In Figure 5, deterministic datasets vary substantially in precipitation intensity and timing, while the uncertainty information provided by STREAM-Sat is evident. STREAM-Sat usually envelops the other products, but there are exceptions when substantial differences in timing (e.g., CHIRPS in Figure 5e & 5g) or magnitude (e.g., ERA5 in Figure 5b) occur in a single precipitation dataset. The main tendency for any particular dataset varies geographically, though IMERG Early generally has higher rainfall amounts, consistent with documented high biases (e.g., Huffman et al., 2023; Li et al., 2021; Li et al., 2022). Stage IV data shown in Figure 5a & 5b are well captured by the STREAM-Sat ensemble. Figure S1 plots the EM-Earth ensemble in the same areas as in Figure 5, and EM-Earth generally exhibits a larger ensemble spread.

Multiple values of CR and CRPS for STREAM-Sat and MAE for IMERG were calculated for all continents except Antarctica, using MSWEP, CHIRPS, and ERA5 as the comparison dataset, respectively (Figure 6). These calculations were done at the "lowest common" spatial and temporal resolution of each pair of datasets (e.g., 3-hourly and 0.1° for MSWEP vs. IMERG; daily and 0.1° for CHIRPS vs. STREAM-Sat; refer to Table 1 for resolution details). STREAM-Sat is substantially closer to comparison datasets than IMERG in terms of CRPS vs. MAE for all continents (Figure 6a, 6c & 6e). For hits, the reduction of CRPS is 52%, 40%, and 47% on average with respect to MSWEP, CHIRPS, and ERA5, respectively. When IMERG gives a false alarm, the probability that the spread of the corresponding STREAM-Sat ensemble includes zero precipitation is 96%, 64%, and 87% on average with respect to MSWEP, CHIRPS, and ERA5, respectively. Although MSWEP is assumed to be higher-fidelity in North America and Europe where gauge networks are dense, STREAM-Sat does not appear to suffer relative to MSWEP in these regions (see Figures 6a-6b). The seeming inferior performance of STREAM-Sat when compared to CHIRPS in Figures 6c-6d is likely due to the complex terrain in central Asia, with both STREAM-Sat and CHIRPS being adversely affected in ways not possible to disentangle with presently available data. This result highlights the difficulty of benchmarking new global precipitation datasets in regions with limited ground-based observations. The consistency between STREAM-Sat and ERA5 in Europe, evinced by Figures 6e-6f, reflects well when considering the relatively skillful prediction of ERA5 in the extratropics (Lavers et al., 2022).

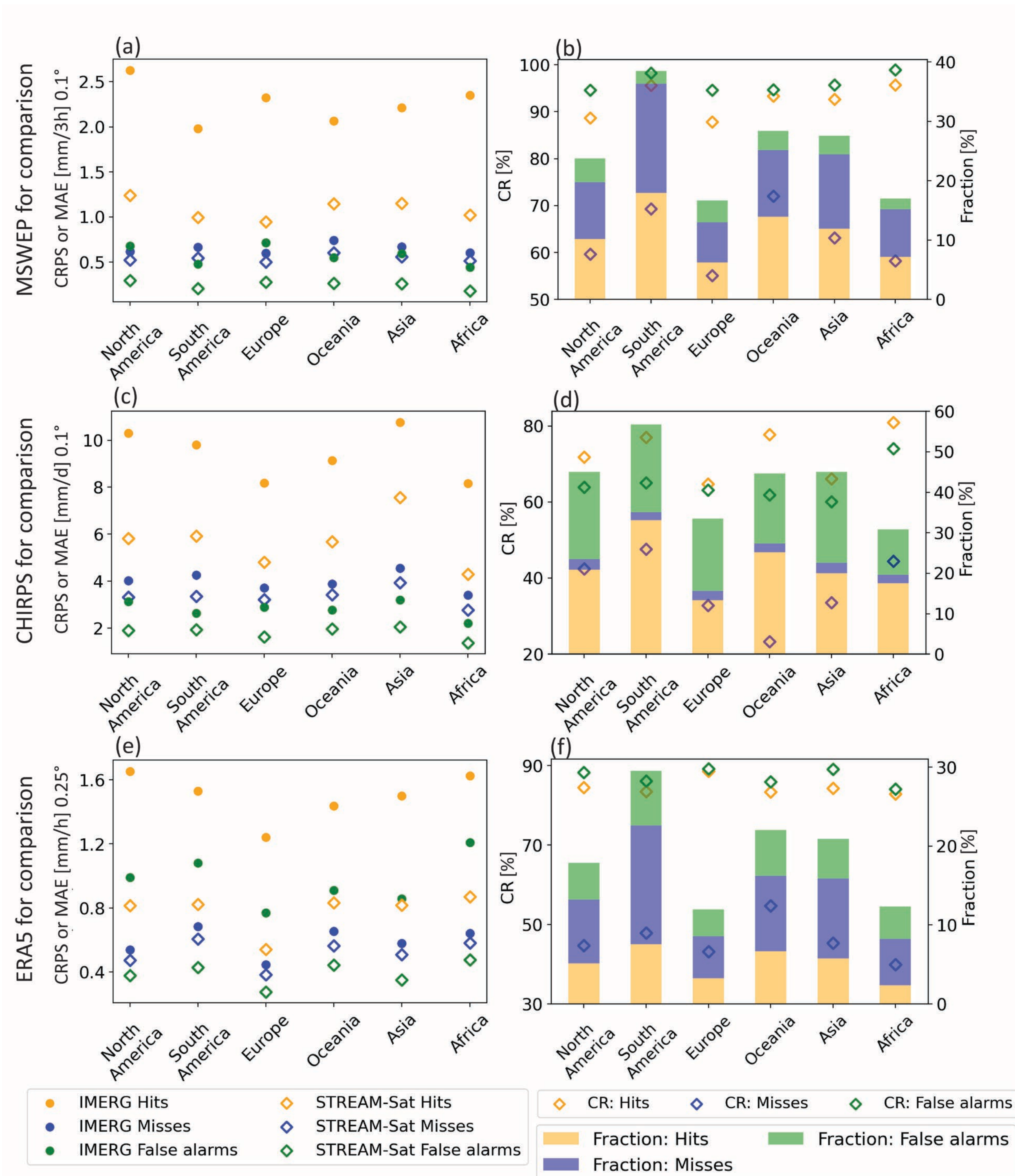


Figure 6. (a, c, e) Comparisons of CRPS (for STREAM-Sat) and MAE (for IMERG) against other global deterministic datasets. (b, d, f) Fraction of total IMERG cases for hits, misses, and false alarms categories (percentage of correct non-detects is not shown) and their corresponding CR against other global deterministic datasets.

4.3 Performance for Climatology and Heavy Rainfall

In this section, we examine the performance of STREAM-Sat compared to IMERG from the perspectives of climatology and heavy daily rainfall in the land area between 50°N and 50°S. Probability of daily rainfall (using a precipitation detection threshold of 0.5 mm/day), the 90th percentile of daily precipitation, and annual total precipitation from IMERG and from one randomly-selected STREAM-Sat ensemble were compared against MSWEP (Figure 7). Both STREAM-Sat and IMERG have a higher probability of precipitation in dry regions—where the probability of daily precipitation is less than 0.2—compared to MSWEP. Where daily precipitation probability exceeds 0.3, there is good (poor) agreement between STREAM-Sat (IMERG) and MSWEP. IMERG exhibits a value of 90th percentile of daily precipitation almost 20% higher than MSWEP; STREAM-Sat is much closer (only 1.6% difference). STREAM-Sat reports smaller RMSE and higher CC (319 mm, 0.93) relative to IMERG (379 mm, and 0.91) for annual totals.

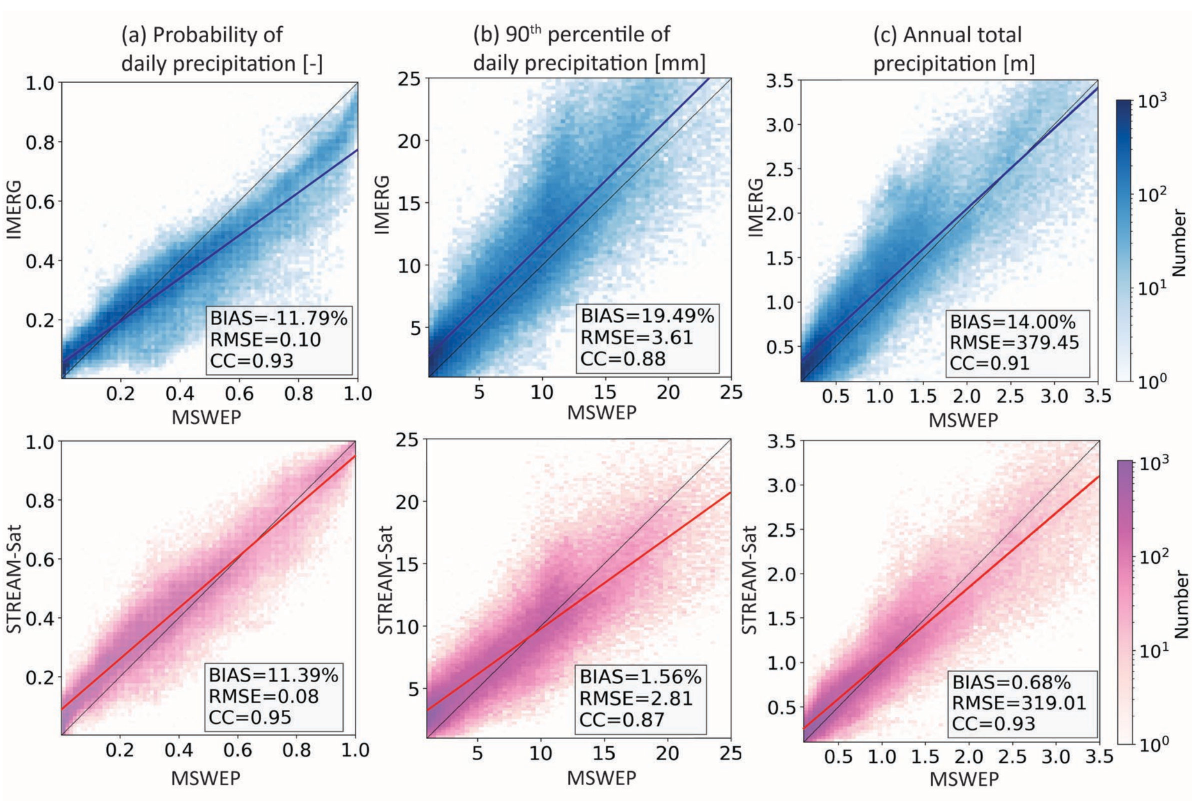


Figure 7. (a) Probability of daily precipitation (detection threshold of 0.5 mm/day). (b) 90th percentile daily precipitation. (c) Annual total precipitation. The first row is the comparison between IMERG and MSWEP. The second row is the comparison between one randomly-selected STREAM-Sat member and MSWEP. Bias, RMSE and correlation coefficient (CC) are provided in each panel. Data is from the whole study area and for the year 2017.

Figure 8 shows the average difference between IMERG (or STREAM-Sat) and three global precipitation datasets over the one-year study period for times and locations where pixel-scale IMERG exceeds its 90th percentile. Positive values mean that IMERG or STREAM-Sat has a higher value than the comparison dataset. The STREAM-Sat member has a nearly symmetric distribution of mean differences against the comparison datasets (i.e., it is unbiased), whereas

IMERG tends to be positively skewed. When moving from IMERG to STREAM-Sat, the mean difference was reduced from 7.9 mm/day to 1.1 mm/day, 6.7 mm/day to -0.37 mm/day, and 9.8 mm/day to 3.0 mm/day for MSWEP, CHIRPS, and ERA5 as comparison datasets, respectively.

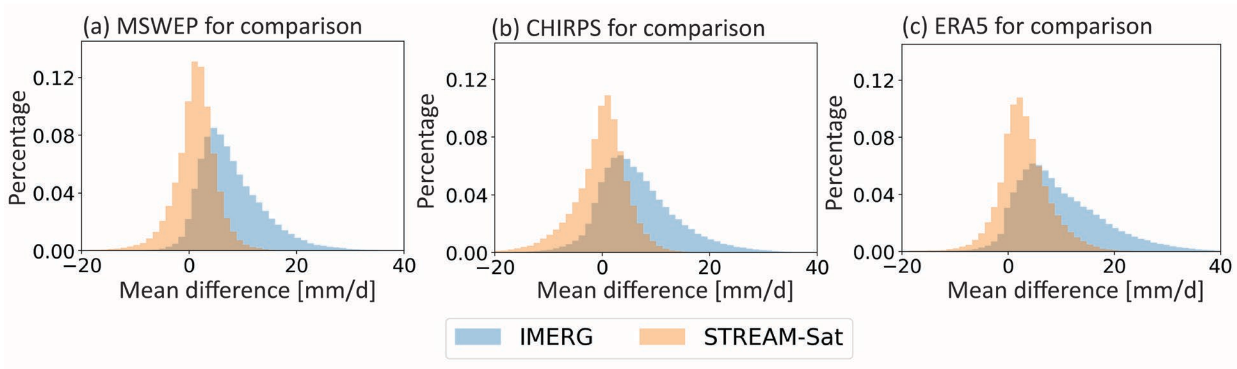
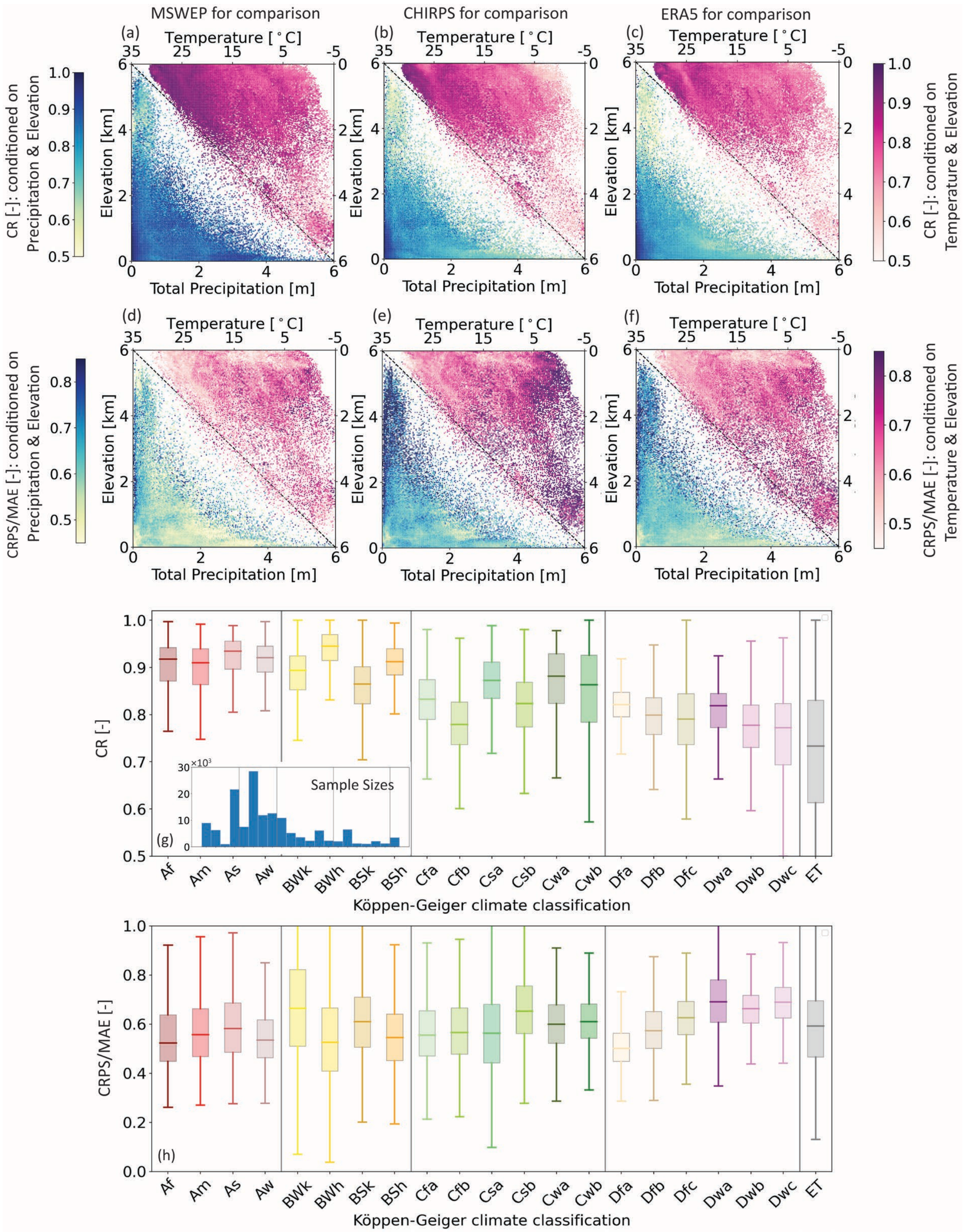


Figure 8. Mean differences of >90th percentile 0.25° daily IMERG precipitation and one random-selected STREAM-Sat ensemble against (a) MSWEP, (b) CHIRPS, and (c) ERA5.

4.4 Physiographic Factors Influencing STREAM-Sat Performance

The performance of IMERG and other satellite-based precipitation datasets is influenced by physiographic factors such as terrain, mean climate, and seasonal precipitation characteristics (Derin et al., 2019; Li et al., 2021; Yu et al., 2021). Here, we evaluate STREAM-Sat performance and compare it with IMERG (Figure 9), in part to understand to what extent the former "inherits" physiographically-linked performance from the latter and how much this performance is modulated by the error modeling and ensemble generation described in Sections 3.1-3.2. Physiographic factors (temperature and elevation) effects on STREAM-Sat performance are analyzed in Figure 9. This plot shows the conditional performance of STREAM-Sat against comparison datasets. Scatter plots in Figures 9a-9f represent two different conditioning analyses: (1) the bottom left of each panel shows STREAM-Sat performance conditioned on elevation (left y-axis) and total precipitation (bottom x-axis); and (2) the top right of each panel represents STREAM-Sat performance conditioned on temperature (top x-axis) and elevation (right y-axis). CR of STREAM-Sat is shown in Figures 9a-9c & 9g, while the ratio between MAE and CRPS (i.e., CRPS/MAE, where lower values mean closer agreement between STREAM-Sat than IMERG to comparison datasets) is shown in Figures 9d-9f & 9h. The three comparison datasets generally show similar patterns but different magnitudes. STREAM-Sat demonstrates the most similar performance to MSWEP in terms of CR and CRPS. STREAM-Sat shows greater discrepancy with comparison datasets—and less change from IMERG—with higher elevation, especially above 3 km. STREAM-Sat seems to demonstrate higher CR and greater improvement over IMERG for warmer temperatures, while perhaps demonstrating slightly less added value in cold regions, likely due to less mixed and solid phase precipitation and better overall satellite retrieval performance in warm conditions (Ebert et al., 2007). No obvious trend is identified relative to total annual precipitation. The three main climates (i.e., B: arid, C: warm temperate and D: snow) generally show lower agreement between STREAM-Sat and comparison datasets and smaller deviations from IMERG in cold regions compared to warm ones (Figure 9g & 9h, e.g., Csa to Csb, or Dfa to Dfc). The performance and improvement in the equatorial zones seem to be the most significant. Polar tundra (denoted as ET in Köppen-Geiger climate classification), dominated by mountains in

Tibet and the Andes, show the largest discrepancies with the comparison dataset, likely due to the combined challenges of precipitation estimation over complex terrain and snow-covered surfaces.



Main climates: A: equatorial, B: arid, C: warm temperate, D: snow, E: polar

Precipitation: W: desert, S: steppe, f: fully humid, s: summer dry, w: winter dry, m: monsoonal
Temperature: a: hot summer, b: warm summer, c: cool summer, d: extremely continental, h: hot arid, k: cold arid, F: polar frost, T: polar tundra

Figure 9. (a-c) CR of STREAM-Sat conditioned by elevation above mean sea level, total annual precipitation (from IMERG Early) and mean annual temperature, calculated against (a) MSWEP, (b) CHIRPS, and (c) ERA5. Same order for d, e and f, but for CRPS/MAE metrics. (g) CR of STREAM-Sat conditioned by Koppen-Geiger climate zone, taking MSWEP as a comparison dataset. (h) Same as (g), but for CRPS/MAE metrics. Climate zones with less than one thousand samples are ignored. Boxes show the first to the third quartiles of the data, with a line at the median. Whiskers extend from the box by 1.5 times the interquartile range. Sample sizes in each climate zone are also given in (g). All the data were upscaled to daily and 0.25°.

4.5 Precipitation Structure

Precipitation spatiotemporal structure is a key determinant of hydrologic response and hazards. The differences in lag-1 temporal autocorrelation between Stage IV and IMERG or STREAM-Sat are shown in Figure 10. At all scales, lag-1 temporal autocorrelation in STREAM-Sat is closer to Stage IV, at least for the US east of 105° W where Stage IV is considered most reliable. This improvement is most significant at the hourly 0.25° scale (Figures 10d-10f), where IMERG overestimates the temporal correlation. Figures S2-S4 extend this autocorrelation analysis to the global land surface. There, MSWEP and ERA5 both exhibit notably higher temporal autocorrelation than both IMERG and STREAM-Sat. Given the generally good correspondence of lag-1 autocorrelation with Stage IV in the eastern US shown in Figure 10, this global result suggests that both MSWEP and ERA5 may have unrealistic temporal autocorrelations; diagnosing the reasons for this is beyond the scope of our study but may be related to data weighting (MSWEP) and assimilation (ERA5). At the 0.25° hourly scale (Figure S3), the autocorrelation of STREAM-Sat is lower than IMERG, but is probably closer to reality if our CONUS results in Figures 10d-10f are indicative of global performance. Aggregation over time (Figure S4, daily and 0.1°) causes slightly higher autocorrelation in STREAM-Sat than IMERG, while STREAM-Sat generally shows the same spatial patterns of temporal autocorrelation as CHIRPS.

We examined spatial anisotropy by power spectrum density (PSD) at different scales for the southeast US region (41°N-31°N, 101°W-91°W) using the method of Guilloteau et al. (2021) (Figure S5). Higher PSD corresponds to lower spatial autocorrelation. IMERG and the STREAM-Sat ensemble show similarly reduced anisotropy and lower spectral power compared with Stage IV, especially at small scales (indicated by the darker color and narrower contour lines). The similarities between STREAM-Sat and IMERG are unsurprising given that the spatial autocorrelation in STREAM-Sat is "inherited" with minimal change from the IMERG field via the SSFT procedure (Section 3.2). The lower spectral density and thus higher spatial autocorrelation in IMERG and STREAM-Sat than in Stage IV is likely due to the morphing and the low effective spatial resolution of satellite-based PMW sensors (Guilloteau et al., 2017).

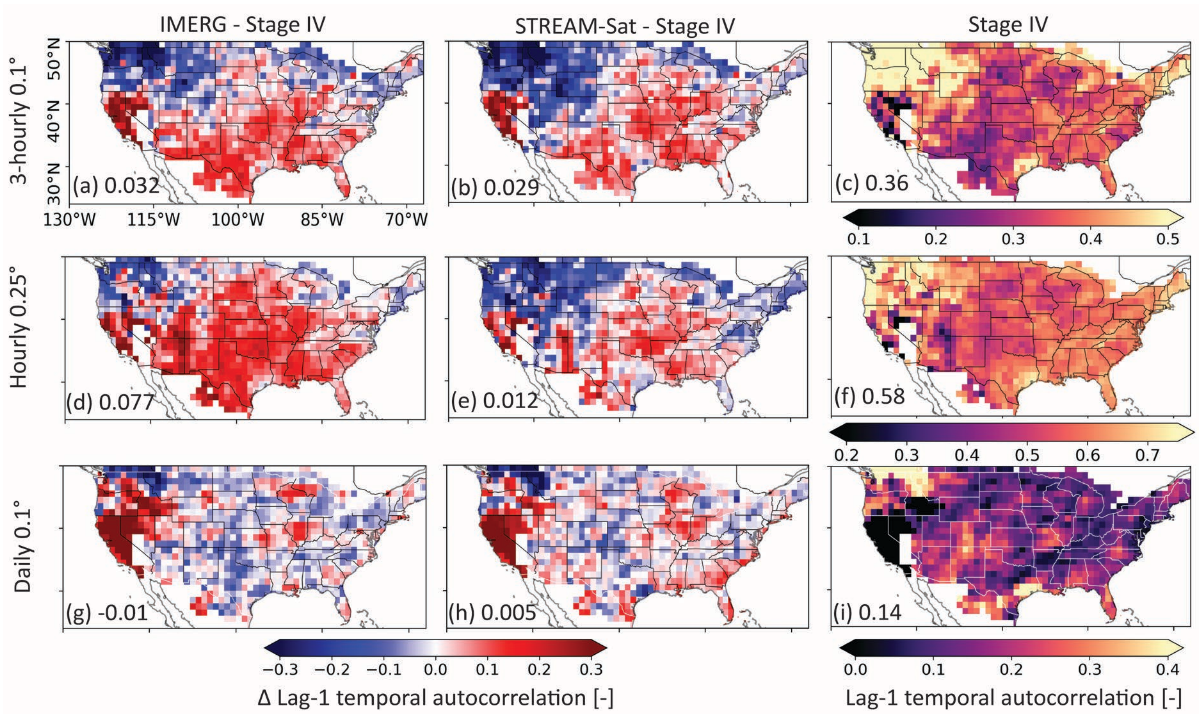


Figure 10. Lag-1 temporal autocorrelation at three spatiotemporal scales. (a)-(c) 3-hourly and 0.1°. (d)-(f) hourly and 0.25°. (g)-(i) daily and 0.1°. The first column is the difference between IMERG and Stage IV. The second column is the difference between one randomly-selected STREAM-Sat ensemble and Stage IV. Positive values mean higher autocorrelation of IMERG or STREAM-Sat than Stage IV. The third column is Stage IV. Only the coverage of Stage IV is shown. The mean value of east of 105° W (where Stage IV is most reliable) is given alongside the panel identifier. Temporal autocorrelation was averaged in each 1° by 1° box.

5 Discussion

5.1 Comparison with EM-Earth

As the two global ensemble precipitation datasets considered in this study, it is worth further investigating the differences between EM-Earth and STREAM-Sat. We highlight one key difference here, beyond differences in basic features (i.e., STREAM-Sat's higher temporal resolution and much lower latency) evident from Table 1. EM-Earth is highly reliant on rain gauge stations, which means that its uncertainty should be narrow in regions with dense station observations and wide where such observations are sparse. STREAM-Sat, in contrast, is satellite-only; its ensemble spread is independent of gauge density. This effect is evident in Figure 11: large coefficients of variation (the standard deviation of the ensemble divided by its mean) in EM-Earth that indicate high ensemble spread coincide with locations of low gauge density (e.g., sub-Saharan Africa and Bolivia) as well as locations of low precipitation. STREAM-Sat, in contrast, is more consistent with no obvious dependence between CV and gauge density.

Indeed, Figure 3 shows that EM-Earth generally but not decisively outperforms STREAM-Sat in the rain gauge-rich eastern United States. Because EM-Earth's CRPS can be expected, however, to be larger in regions with lower rain gauge densities (all else being equal), it is reasonable to surmise that STREAM-Sat would compare more favorably to EM-Earth in such regions. This fact, combined with STREAM's advantages in latency and temporal resolution, may be preferable in certain applications. However, we caution against interpreting lower CV values in STREAM-Sat as solid evidence of superiority over EM-Earth; see limitations of our approach in Section 5.2.

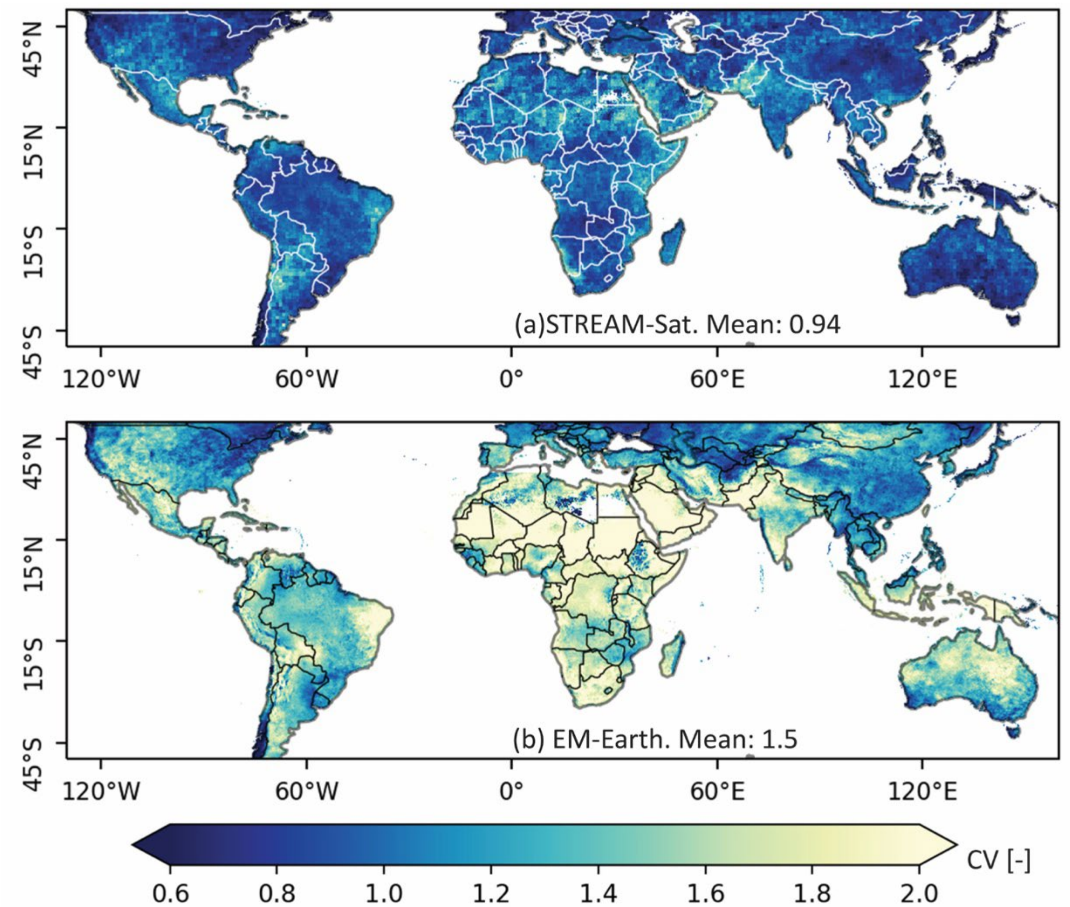


Figure 11. Coefficient of variation (CV; the standard deviation divided by mean) for the ensembles over land. (a) STREAM-Sat. (b) EM-Earth.

We reiterate here that the representation of uncertainty via ensemble methods in both STREAM-Sat and EM-Earth is important. This significance can be observed in Figure 2 for both STREAM-Sat and EM-Earth. The total CRPS for STREAM-Sat (EM-Earth) is 26% (25%) lower than the MAE of the corresponding ensemble mean. This additional benefit is the result of the probabilistic representation of random errors. It is important to note that this probabilistic representation of random error has been shown to be particularly important in hazard and water prediction applications, where removal of systematic bias alone is at best insufficient and at worst leads to degraded predictions (Habib et al., 2014; Hartke et al., 2020; Hartke et al., 2023).

5.2 Limitations of the Framework

Despite the strengths of our approach, a number of limitations remain, mostly stemming from the nature of the input data. GPM's low-earth orbit means that continuous DPR observation over a single location is impossible; the user is instead limited to near-instantaneous "snapshots" of precipitation at a particular location. Thus, if DPR observations (2B-CMB included) are used to quantify uncertainty in a gridded dataset such as is done here with IMERG, the mismatch between the near-instantaneous nature of DPR and the accumulated nature of the gridded dataset introduces additional uncertainty. This uncertainty can be mitigated by deploying the method using gridded datasets with fine temporal resolutions—e.g., sub-hourly to perhaps an hour—but it precludes the application of our approach to datasets with coarser temporal resolutions such as daily (e.g., precluding application to CHIRPS).

In addition, though likely the best space-based precipitation sensor, DPR still suffers from errors resulting from imperfectly understood backscattering and other challenges, especially over snow-covered surfaces. Reliance on DPR for training also raises concern around the lack of independence between 2B-CMB and IMERG—essentially, DPR and GMI observations are used in both datasets, which can be problematic for using 2B-CMB to quantify uncertainty in IMERG. However, Li et al. (2023) used a high-quality ground-based reference to show that the correlation between 2B-CMB and IMERG is not concerningly high, helping to justify 2B-CMB as training reference. We experimented with different IMERG components (Figure S6), revealing that the 2B-CMB-trained model appears to adequately depict IMERG in times and locations where PMW estimates were available, but not when IR observations alone were used. But this phenomenon also exists to a degree when Stage IV was used as a training reference, implying that it is partly attributed to problematic IR retrievals and partly due to the training data source. Since we are likely to fit the error model using "best-case" (i.e., GMI), there is a potential risk of underestimating the uncertainty. More detailed analysis can be found in Text S1 and Li et al. (2023).

The model used in this paper has eight free parameters, with the freedom to add more covariates (and additional parameters in equal measure) to potentially explain additional IMERG error. Adding additional covariates increases the challenge of estimating these parameters, currently handled via a two-step nonlinear optimization (see Scheuerer et al. (2015) for more details). Identifying realistic parameter values requires relatively large samples of coincident IMERG and 2B-CMB estimates. In principle, distinct error models for each PMW sensor that contributes to IMERG could be developed but this would be made difficult by oftentimes limited coincident data samples for model calibration (See Table S1). Also, it might not be necessary, since our results in Figure S6 show that STREAM-Sat generally has comparable performance regardless of which PMW sensor contributed to the original IMERG observation—as long as such a sensor was available. A somewhat different issue arises if one attempts to fit an error model specifically to IR-based IMERG observations. The poor performance of IR-only IMERG retrievals means that correlation between such observations and 2B-CMB is poor and thus certain parameters cannot be adequately identified via optimization (results not shown). It is possible that this last problem will be ameliorated in future versions of IMERG, which will feature more advanced and more accurate IR retrieval schemes.

Finally, this study focuses on liquid precipitation (as indicated by IMERG's "probabilityLiquidPrecipitation" flag, see Section 2.2). Both IMERG and DPR-based products tend to underestimate snowfall (Behrangi et al., 2018; Casella et al., 2017; Song et al., 2020). However, previous research has proposed a dataset built by coincident observations between DPR and the 94 GHz CloudSat Cloud Profiling Radar (CPR) to serve as a snowfall training reference (Behrangi et al., 2020; Liu, Adhikari, et al., 2018). CPR is able to measure light snowfall, while DPR offers advantages for heavier snowfall rates. Validation of this is outside the scope of this study.

6 Conclusions

The potential of satellite-based precipitation products and other large-scale precipitation datasets is limited by their inherent uncertainties, including mischaracterization of both precipitation occurrence and intensity. In principle, this uncertainty can be quantified by using ensemble generation techniques that produce multiple plausible realizations of the unknown true precipitation field. In practice, however, the creation of such ensemble datasets has proven difficult due to the limited quality and quantity of available ground truth data, the complexity of satellite precipitation algorithms, and the difficulty of "connecting" such data and their uncertainties over space and time.

In this paper, we propose STREAM-Sat, a method for producing ensemble precipitation fields using only satellite data. The method can be applied globally, in near-realtime and at high resolution (here, 0.1°, half-hourly over land). We validated the performance of STREAM-Sat against Stage IV over CONUS and compared it with several other state-of-the-art benchmark precipitation datasets globally. STREAM-Sat shows clear improvement in detection ability and intensity compared to IMERG Early. STREAM-Sat corrects the high bias of IMERG and exhibits climatology statistics closer to comparison datasets. The accuracy of STREAM-Sat is close to, if not better than, these global precipitation datasets to which it is compared to. In contrast with precipitation datasets relying on ground-based gauge networks, STREAM-Sat shows spatially consistent ensemble spread and performance metrics, and our results indicate that STREAM-Sat could be preferable in gauge-limited regions. The ensemble representation of random errors adds a roughly 25% improvement on top of the ensemble mean, highlighting the importance of ensemble methods for large scale precipitation estimation and its application. Similar to other satellite-derived precipitation products, elevation and temperature are key factors that influence the performance of STREAM-Sat, with high-elevation and snow-covered areas showing poorer performance. STREAM-Sat generally maintains a similar precipitation structure to IMERG at varying spatiotemporal resolutions, which appears to be more realistic than at least some comparison datasets, which appear overly smooth.

While STREAM-Sat's accuracy (e.g., in terms of CRPS) is not superior in all respects over all other comparison datasets, its unique combination of traits—i.e., it is probabilistic via ensemble members, it has low bias, it can be produced in near-realtime, and it has high resolution—gives it certain advantages over the other benchmark global precipitation datasets. The proposed method is particularly valuable in ungauged regions with large and unknown meteorological uncertainties that merit ensemble approaches, while its near-realtime potential offers notable benefits in water-related disaster early warning systems. Future work will focus on evaluating the usefulness of STREAM-Sat in a range of water resources and hazard applications, as well as understanding its absolute accuracy through comparison with dense rain gauge networks, as opposed to its relative accuracy via the comparison with "peer" datasets pursued here.

Acknowledgments

K. Peng, D.B. Wright, and J. Tan were supported by the NASA Precipitation Measurement Mission (Grant Number 80NSSC22K0600). Y. Derin was supported by NASA Global Precipitation Measurement Ground Validation program (Grant Number NNX16AL23G). The authors thank the Center for High Throughput Computing (2006) at the University of Wisconsin-Madison for computational resources and support.

Open Research

IMERG Early, 2A-DPR, and 2B-CMB were downloaded from the NASA GES DISC website https://disc.gsfc.nasa.gov/. The IMERG motion vectors used in STREAM-Sat were provided by the IMERG development team. NEXRAD Stage IV data is available from the Earth Observing Laboratory data archive (https://data.eol.ucar.edu/dataset/21.093). ERA5 data is available from Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://cds.climate.copernicus.eu/. CHIRPS data was downloaded from https://data.chc.ucsb.edu/products/CHIRPS-2.0/. MSWEP is available from https://www.gloh2o.org/mswep/. EM-Earth is available from the Federated Research Data Repository https://doi.org/10.20383/102.0547. STREAM-Sat codes and data samples can be found at https://github.com/KaidiWisc/STREAM-Sat.git.

References

AghaKouchak, A., Bárdossy, A., & Habib, E. (2010). Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. *Advances in Water Resources, 33*(6), 624-634. doi:10.1016/j.advwatres.2010.02.010

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

AghaKouchak, A., Mehran, A., Norouzi, H., & Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters, 39*(9). doi:10.1029/2012gl051592

Ayat, H., Evans, J. P., Sherwood, S. C., & Soderholm, J. (2022). Intensification of subhourly heavy rainfall. *Science, 378*(6620), 655-659. doi:10.1126/science.abn8657

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., . . . Wood, E. F. (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. doi:10.5194/hess-23-207-2019

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

Behrangi, A., Bormann, K. J., & Painter, T. H. (2018). Using the Airborne Snow Observatory to Assess Remotely Sensed Snowfall Products in the California Sierra Nevada. *Water Resources Research, 54*(10), 7331-7346. doi:10.1029/2018wr023108

Behrangi, A., & Song, Y. (2020). A new estimate for oceanic precipitation amount and distribution using complementary precipitation observations from space and comparison with GPCP. *Environmental Research Letters, 15*(12). doi:10.1088/1748-9326/abc6d1

Caillouet, L., Vidal, J. P., Sauquet, E., Graff, B., & Soubeyroux, J. M. (2019). SCOPE Climate: a 142-year daily high-resolution ensemble meteorological reconstruction dataset over France. *Earth Syst. Sci. Data, 11*(1), 241-260. doi:10.5194/essd-11-241-2019

Casella, D., Panegrossi, G., Sanò, P., Marra, A. C., Dietrich, S., Johnson, B. T., & Kulie, M. S. (2017). Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat-CPR. *Atmospheric Research, 197*, 64-75. doi:10.1016/j.atmosres.2017.06.018

Caseri, A., Javelle, P., Ramos, M. H., & Leblois, E. (2016). Generating precipitation ensembles for flood alert and risk management. *Journal of Flood Risk Management, 9*(4), 402-415. doi:10.1111/jfr3.12203

Center for High Throughput Computing. (2006). Center for High Throughput Computing. doi:10.21231/GNT1-HW21.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres, 123*(17), 9391-9409. doi:10.1029/2017jd028200

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., . . . Yilmaz, K. K. (2019). Evaluation of GPM-era Global Satellite Precipitation Products over Multiple Complex Terrain Regions. *Remote Sensing, 11*(24). doi:10.3390/rs11242936

Derin, Y., Anagnostou, E., Berne, A., Borga, M., Boudevillain, B., Buytaert, W., . . . Yilmaz, K. K. (2016). Multiregional Satellite Precipitation Products Evaluation over Complex Terrain. *Journal of Hydrometeorology, 17*(6), 1817-1836. doi:https://doi.org/10.1175/JHM-D-15-0197.1

Du, J. (2011). *NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6PG1QDD.*

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. doi:https://doi.org/10.1175/BAMS-88-1-47

Exum, N. G., Betanzo, E., Schwab, K. J., Chen, T. Y. J., Guikema, S., & Harvey, D. E. (2018). Extreme Precipitation, Public Health Emergencies, and Safe Drinking Water in the USA. *Current Environmental Health Reports, 5*(2), 305-315. doi:10.1007/s40572-018-0200-5

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

Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., . . . Zhang, X. (2021). Anthropogenic intensification of short-duration rainfall extremes. *Nature Reviews Earth & Environment, 2*(2), 107-122. doi:10.1038/s43017-020-00128-6
Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., . . . Michaelsen, J. (2015). The climate hazards infrared precipitation with stations--a new environmental record for monitoring extremes. *Sci Data, 2*, 150066. doi:10.1038/sdata.2015.66
Germann, U., Berenguer, M., Sempere-Torres, D., & Zappa, M. (2009). REAL-Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quarterly Journal of the Royal Meteorological Society, 135*(639), 445-456. doi:10.1002/qj.375
Grimit, E. P., Gneiting, T., Berrocal, V. J., & Johnson, N. A. (2006). The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quarterly Journal of the Royal Meteorological Society, 132*(621C), 2925-2942. doi:10.1256/qj.05.235
Guilloteau, C., Foufoula-Georgiou, E., Kirstetter, P., Tan, J., & Huffman, G. J. (2021). How well do multi-satellite products capture the space-time dynamics of precipitation? Part I: five products assessed via a wavenumber-frequency decomposition. *Journal of Hydrometeorology*. doi:10.1175/jhm-d-21-0075.1
Guilloteau, C., Foufoula-Georgiou, E., Kirstetter, P., Tan, J., & Huffman, G. J. (2022). How Well Do Multisatellite Products Capture the Space–Time Dynamics of Precipitation? Part II: Building an Error Model through Spectral System Identification. *Journal of Hydrometeorology, 23*(9), 1383-1399. doi:https://doi.org/10.1175/JHM-D-22-0041.1
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. doi:https://doi.org/10.1175/JHM-D-17-0087.1
Habib, E., Haile, A. T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of Bias Correction of Satellite-Rainfall Estimates on Runoff Simulations at the Source of the Upper Blue Nile. *Remote Sensing, 6*(7), 6688-6708. doi:10.3390/rs6076688
Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data, 7*(1), 109. doi:10.1038/s41597-020-0453-3
Hartke, S. H., & Wright, D. B. (2022). Where Can IMERG Provide a Better Precipitation Estimate than Interpolated Gauge Data? *Remote Sensing, 14*(21). doi:10.3390/rs14215563
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. doi:https://doi.org/10.1175/JHM-D-19-0295.1
Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D. B., & Khan, S. (2022). Ensemble Representation of Satellite Precipitation Uncertainty Using a Nonstationary, Anisotropic Autocorrelation Model. *Water Resources Research, 58*(8). doi:10.1029/2021wr031650
Hartke, S. H., Wright, D. B., Quintero, F., & Falck, A. S. (2023). Incorporating IMERG satellite precipitation uncertainty into seasonal and peak streamflow predictions using the Hillslope Link hydrological model. *Journal of Hydrology X, 18*. doi:10.1016/j.hydroa.2023.100148
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999-2049. doi:10.1002/qj.3803
Hong, Y., Hsu, K.-L., Sorooshian, S., & Gao, X. (2004). Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *Journal of Applied Meteorology, 43*(12), 1834-1853. doi:https://doi.org/10.1175/JAM2173.1
Hossain, F., & Anagnostou, E. N. (2006). A two-dimensional satellite rainfall error model. *IEEE Transactions on Geoscience and Remote Sensing, 44*(6), 1511-1522. doi:10.1109/tgrs.2005.863866
Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., . . . Xie, P. (2020). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD)*. Retrieved from https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf:
Huffman, G. J., Bolvin, D. T., Joyce, R., Kelley, O. A., Nelkin, E. J., Portier, A., . . . West, B. J. (2023). IMERG V07 Release Notes.
Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). GPM IMERG Early Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06. *Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC)*. doi:10.5067/GPM/IMERG/3B-HH-E/06

Iguchi, T., & Meneghini, R. (2021). GPM DPR Precipitation Profile L2A 1.5 hours 5 km V07. *Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC)*. doi:10.5067/GPM/DPR/GPM/2A/07

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., Le, M., . . . Kubota, T. (2018). *GPM / DPR Level-2 Algorithm Theoretical Basis Document (ATBD)*. Retrieved from https://gpm.nasa.gov/sites/default/files/2019-05/ATBD_DPR_201811_with_Appendix3b.pdf:

Khedhaouiria, D., Bélair, S., Fortin, V., Roy, G., & Lespinas, F. (2020). High-Resolution (2.5 km) Ensemble Precipitation Analysis across Canada. *J Journal of Hydrometeorology, 21*(9), 2023-2039. doi:https://doi.org/10.1175/JHM-D-19-0282.1

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

Kirschbaum, D., Kapnick, S. B., Stanley, T., & Pascale, S. (2020). Changes in Extreme Precipitation and Landslides Over High Mountain Asia. *Geophysical Research Letters, 47*(4), e2019GL085347. doi:https://doi.org/10.1029/2019GL085347

Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., . . . Okamoto, K. (2007). Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation. *IEEE Transactions on Geoscience and Remote Sensing, 45*(7), 2259-2275. doi:10.1109/TGRS.2007.895337

Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (2022). An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society, 148*(748), 3152-3165. doi:10.1002/qj.4351

Leblois, E., & Creutin, J.-D. (2013). Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method. *Water Resources Research, 49*(6), 3375-3387. doi:10.1002/wrcr.20190

Li, Z., Tang, G., Hong, Z., Chen, M., Gao, S., Kirstetter, P., . . . Hong, Y. (2021). Two-decades of GPM IMERG early and final run products intercomparison: Similarity and difference in climatology, rates, and extremes. *Journal of Hydrology, 594*, 125975. doi:https://doi.org/10.1016/j.jhydrol.2021.125975

Li, Z., Tang, G., Kirstetter, P., Gao, S., Li, J. L. F., Wen, Y., & Hong, Y. (2022). Evaluation of GPM IMERG and its constellations in extreme events over the conterminous united states. *Journal of Hydrology, 606*, 127357. doi:https://doi.org/10.1016/j.jhydrol.2021.127357

Li, Z., Wright, D. B., Hartke, S. H., Kirschbaum, D. B., Khan, S., Maggioni, V., & Kirstetter, P.-E. (2023). Toward a Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products Based on GPM DPR. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1-15. doi:10.1109/tgrs.2023.3235270

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. doi:https://doi.org/10.1175/JHM-D-20-0041.1

Liao, Z., Hong, Y., Wang, J., Fukuoka, H., Sassa, K., Karnawati, D., & Fathani, F. (2010). Prototyping an experimental early warning system for rainfall-induced landslides in Indonesia using satellite remote sensing and geospatial datasets. *Landslides, 7*(3), 317-324. doi:10.1007/s10346-010-0219-7

Liu, C., Adhikari, A., & Kulie, M. S. (2018). Global Distribution of Snow Precipitation Features and Their Properties from 3 Years of GPM Observations. *Journal of Climate, 31*(10), 3731-3754. doi:10.1175/jcli-d-17-0012.1

Liu, C., Guo, L., Ye, L., Zhang, S., Zhao, Y., & Song, T. (2018). A review of advances in China's flash flood early-warning system. *Natural Hazards, 92*(2), 619-634. doi:10.1007/s11069-018-3173-7

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

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

Nelson, B. R., Prat, O. P., Seo, D. J., & Habib, E. (2016). Assessment and Implications of NCEP Stage IV Quantitative Precipitation Estimates for Product Intercomparisons. *Weather and Forecasting, 31*(2), 371-394. doi:https://doi.org/10.1175/WAF-D-14-00112.1

Nerini, D., Besic, N., Sideris, I., Germann, U., & Foresti, L. (2017). A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space Fourier transform. *Hydrol. Earth Syst. Sci., 21*(6), 2777-2797. doi:10.5194/hess-21-2777-2017

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., . . . Arnold, J. R. (2015). Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States. *Journal of Hydrometeorology, 16*(6), 2481-2500. doi:https://doi.org/10.1175/JHM-D-15-0026.1

Newman, A. J., Clark, M. P., Longman, R. J., Gilleland, E., Giambelluca, T. W., & Arnold, J. R. (2019). Use of Daily Station Observations to Produce High-Resolution Gridded Probabilistic Precipitation and Temperature Time Series for the Hawaiian Islands. *Journal of Hydrometeorology, 20*(3), 509-529. doi:https://doi.org/10.1175/JHM-D-18-0113.1

Newman, A. J., Clark, M. P., Wood, A. W., & Arnold, J. R. (2020). Probabilistic Spatial Meteorological Estimates for Alaska and the Yukon. *Journal of Geophysical Research: Atmospheres, 125*(22), e2020JD032696. doi:https://doi.org/10.1029/2020JD032696

Olson, W. (2022). GPM DPR and GMI Combined Precipitation L2B 1.5 hours 5 km V07. *Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC)*. doi:10.5067/GPM/DPRGMI/CMB/2B/07

Pan, H.-L., Jiang, Y.-J., Wang, J., & Ou, G.-Q. (2018). Rainfall threshold calculation for debris flow early warning in areas with scarcity of data. *Natural Hazards and Earth System Sciences, 18*(5), 1395-1409. doi:10.5194/nhess-18-1395-2018

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., & Foresti, L. (2019). Pysteps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development, 12*(10), 4185-4219. doi:10.5194/gmd-12-4185-2019

Rubel, F., and M. Kottek. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift, 19*(2), 135-141. doi:10.1127/0941-2948/2010/0430

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. doi:https://doi.org/10.1175/MWR-D-15-0061.1

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of Uncertainty in Precipitation Forcing Data Sets on the Hydrologic Budget of an Integrated Hydrologic Model in Mountainous Terrain. *Water Resources Research, 56*(12), e2020WR027639. doi:https://doi.org/10.1029/2020WR027639

Skofronick-Jackson, G., Kulie, M., Milani, L., Munchak, S. J., Wood, N. B., & Levizzani, V. (2019). Satellite Estimation of Falling Snow: A Global Precipitation Measurement (GPM) Core Observatory Perspective. *J Appl Meteorol Climatol, 58*(7), 1429-1448. doi:10.1175/JAMC-D-18-0124.1

Sloat, L. L., Gerber, J. S., Samberg, L. H., Smith, W. K., Herrero, M., Ferreira, L. G., . . . West, P. C. (2018). Increasing importance of precipitation variability on global livestock grazing lands. *Nature Climate Change, 8*(3), 214-218. doi:10.1038/s41558-018-0081-5

Song, Y., Behrangi, A., & Blanchard-Wrigglesworth, E. (2020). Assessment of Satellite and Reanalysis Cold Season Snowfall Estimates Over Arctic Sea Ice. *Geophysical Research Letters, 47*(16). doi:10.1029/2020gl088970

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

Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). IMERG V06: Changes to the Morphing Algorithm. *Journal of Atmospheric and Oceanic Technology, 36*(12), 2471-2482. doi:10.1175/jtech-d-19-0114.1

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. doi:10.1175/jhm-d-16-0079.1

Tang, G., Clark, M. P., & Papalexiou, S. M. (2022). EM-Earth: The Ensemble Meteorological Dataset for Planet Earth. *Bulletin of the American Meteorological Society, 103*(4), E996-E1018. doi:10.1175/bams-d-21-0106.1

Tang, G., Clark, M. P., Papalexiou, S. M., Newman, A. J., Wood, A. W., Brunet, D., & Whitfield, P. H. (2021). EMDNA: an Ensemble Meteorological Dataset for North America. *Earth System Science Data, 13*(7), 3337-3362. doi:10.5194/essd-13-3337-2021

Thorarinsdottir, T. L., Gneiting, T., & Gissibl, N. (2013). Using Proper Divergence Functions to Evaluate Climate Models. *SIAM/ASA Journal on Uncertainty Quantification, 1*(1), 522-534. doi:10.1137/130907550

Tozer, B., Sandwell, D. T., Smith, W. H. F., Olson, C., Beale, J. R., & Wessel, P. (2019). Global Bathymetry and Topography at 15 Arc Sec: SRTM15+. *6*(10), 1847-1864. doi:https://doi.org/10.1029/2019EA000658

Wilhelm, B., Rapuc, W., Amann, B., Anselmetti, F. S., Arnaud, F., Blanchet, J., . . . Wirth, S. B. (2022). Impact of warmer climate periods on flood hazard in the European Alps. *Nature Geoscience, 15*(2), 118-123. doi:10.1038/s41561-021-00878-y

Wright, D. B. (2018). Rainfall Information for Global Flood Modeling. In G. Schumann, P. D. Bates, H. Apel, & G. T. Aronica (Eds.), *Global flood hazard*: *Applications in modeling, mapping, and forecasting* (pp. 19–42). John Wiley & Sons, Inc. and American Geophysical Union.

Wright, D. B., Kirschbaum, D. B., & Yatheendradas, S. (2017). Satellite Precipitation Characterization, Error Modeling, and Error Correction Using Censored Shifted Gamma Distributions. *J Hydrometeorol, Volume 18*(No 10), 2801-2815. doi:10.1175/JHM-D-17-0060.1

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F., & Lin, R. (2017). Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *Journal of Hydrometeorology, 18*(6), 1617-1641. doi:https://doi.org/10.1175/JHM-D-16-0168.1

Yu, L., Leng, G., Python, A., & Peng, J. (2021). A Comprehensive Evaluation of Latest GPM IMERG V06 Early, Late and Final Precipitation Products across China. *Remote Sensing, 13*(6), 1208.

References from the Supporting Information

Ayat, H., Evans, J. P., & Behrangi, A. (2021). How do different sensors impact IMERG precipitation estimates during hurricane days? *Remote Sensing of Environment, 259*, 112417. doi:https://doi.org/10.1016/j.rse.2021.112417

Li, Z., Wright, D. B., Hartke, S. H., Kirschbaum, D. B., Khan, S., Maggioni, V., & Kirstetter, P.-E. (2023). Toward a Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products Based on GPM DPR. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1-15. doi:10.1109/tgrs.2023.3235270

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. doi:10.1175/jhm-d-16-0079.1