Ensemble Representation of Satellite Precipitation Uncertainty using an Uncalibrated, Nonstationary, Anisotropic Autocorrelation Model

Samantha H. Hartke[1], Daniel Benjamin Wright[1], Zhe Li[2], Viviana Maggioni[3], Dalia Kirschbaum[4], and Sana Khan[5]

[1]University of Wisconsin-Madison
[2]Colorado State University
[3]George Mason University
[4]NASA Goddard Space Flight Center
[5]NASA Goddard Space Flight Center, and Earth System Science Interdisciplinary Center, University of Maryland

November 21, 2022

Abstract

The usefulness of satellite multi-sensor precipitation (SMP) and other satellite-informed precipitation products in water resources modeling can be hindered by substantial errors which vary considerably with spatiotemporal scale. One approach to cope with these errors is by combining SMPs with ensemble generation methods, such that each ensemble member reflects one plausible realization of the true—but unknown—precipitation. This requires replicating the spatiotemporal autocorrelation structure of SMP errors. The climatology of this structure is unknown for most locations due to a lack of ground reference observations, while the unique anisotropy and nonstationarity within any particular precipitation system limit the relevance of this climataology to the depiction of error in individual storm systems. Characterizing and simulating this autocorrelation across spatiotemporal scales has thus been called a grand challenge within the precipitation community. We introduce the Space-Time Rainfall Error and Autocorrelation Model (STREAM), which combines anisotropic and nonstationary SMP spatiotemporal correlation structures with a pixel-scale precipitation error model to stochastically generate ensemble precipitation fields that resemble "ground truth" precipitation. We generate STREAM precipitation ensembles at high resolution (1-hour, 0.1@) with minimal reliance on ground-reference data, and evaluate these ensembles at multiple scales. STREAM ensembles consistently "bracket" ground-truth observations and replicate the autocorrelation structure of ground-truth precipitation fields. STREAM is compatible with pixel-scale error/uncertainty formulations beyond those presented here, and could be applied globally to other precipitation sources such as numerical weather predictions or "blended" products. In combination with recent work in SMP uncertainty characterization, STREAM could be run without any ground data.

Hosted file

`stream.supportinginfo.docx` available at https://authorea.com/users/529616/articles/597144-ensemble-representation-of-satellite-precipitation-uncertainty-using-an-uncalibrated-nonstationary-anisotropic-autocorrelation-model

Ensemble Representation of Satellite Precipitation Uncertainty using an Uncalibrated, Nonstationary, Anisotropic Autocorrelation Model

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

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

[2]Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, Colorado

[3]Department of Civil, Environmental, and Infrastructure Engineering, George Mason University.

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

[5]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.

Corresponding author: Samantha Hartke (shartke@wisc.edu)

Key Points:

- High resolution precipitation ensemble fields are generated that represent the uncertainty range of error-prone precipitation products.
- The space-time correlation structure of satellite precipitation error is modeled without calibration and without ground-reference data.
- Precipitation ensembles demonstrate the ability to “bracket” ground-reference observations at multiple space-time scales.

Abstract

The usefulness of satellite multi-sensor precipitation (SMP) and other satellite-informed precipitation products in water resources modeling can be hindered by substantial errors which vary considerably with spatiotemporal scale. One approach to cope with these errors is by combining SMPs with ensemble generation methods, such that each ensemble member reflects one plausible realization of the true—but unknown—precipitation. This requires replicating the spatiotemporal autocorrelation structure of SMP errors. The climatology of this structure is unknown for most locations due to a lack of ground reference observations, while the unique anisotropy and nonstationarity within any particular precipitation system limit the relevance of this climataology to the depiction of error in individual storm systems. Characterizing and simulating this autocorrelation across spatiotemporal scales has thus been called a grand challenge within the precipitation community. We introduce the Space-Time Rainfall Error and Autocorrelation Model (STREAM), which combines anisotropic and nonstationary SMP spatiotemporal correlation structures with a pixel-scale precipitation error model to stochastically generate ensemble precipitation fields that resemble "ground truth" precipitation. We generate STREAM precipitation ensembles at high resolution (1-hour, 0.1˚) with minimal reliance on ground-reference data, and evaluate these ensembles at multiple scales. STREAM ensembles consistently "bracket" ground-truth observations and replicate the autocorrelation structure of ground-truth precipitation fields. STREAM is compatible with pixel-scale error/uncertainty formulations beyond those presented here, and could be applied globally to other precipitation sources such as numerical weather predictions or "blended" products. In combination with recent work in SMP uncertainty characterization, STREAM could be run without any ground data.

1 Introduction

Accurate, timely, high-resolution, and reliable precipitation data is critical for a range of water modeling contents including floods, droughts, crop yields, and landslide hazards. Interest in deploying such models at continental-to-global scales has grown in recent years. Examples include the Famine Early Warning System (FEWS; Funk et al., 2019), the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004), the Global Flood Monitoring System (GFMS; Wu et al., 2014), the Global Flood Awareness System (GloFAS; Alfieri et al., 2013), and the Landslide Hazard Assessment for Situational Awareness (LHASA; Kirschbaum & Stanley, 2018). This interest has been driven in part by increasing availability and accuracy of global precipitation datasets to "fill in" where no ground-based sensors (e.g., rain gages or weather radar) exist. These datasets include satellite multisensor precipitation (SMP) products, satellite-assimilating numerical weather models, and "blended" options that combine the prior two, oftentimes with rain gages (see Beck et al., 2017, Nogueira, 2020, and Sun et al., 2018 for recent reviews). While these datasets share a common set of advantages–namely, global coverage at increasingly high resolutions and ever lower latencies– and have improved in accuracy over time (Gebregiorgis et al., 2018; Maggioni et al., 2016; Tang et al., 2020), they also share a general tendency towards high systematic biases and random errors in both precipitation occurrence and rate (e.g., Nogueira, 2020; Tian & Peters-Lidard, 2010; Wright, 2018).

Errors in SMPs can arise from a variety of causes, including variable sensor accuracy and sampling error from infrequent satellite overpasses, and are modulated by retrieval conditions (e.g., Tan et al., 2016, 2018; Tian & Peters-Lidard, 2007). Validation studies have demonstrated that errors tend to grow with latitude, precipitation intensity, terrain complexity, and in frozen or mixed-phase precipitation conditions (e.g., Aghakouchak et al., 2011; Shige et al., 2013). Spatial and temporal autocorrelation among SMP errors exists because the retrieval conditions and sampling limits that impact a precipitation estimate at a given location and time tend to also impact estimates that are nearby in space or time. This autocorrelation means that error properties vary according to the level of spatial or temporal aggregation of the data (Quintero et al., 2016; Sarachi et al., 2015; Tang et al., 2016); specifically, errors tend to diminish with increasing aggregation as errors tend to cancel.

When used to force water prediction models, errors in precipitation products lead to errors in model estimates of key variables such as streamflow, soil moisture, and groundwater storage (e.g., Falck et al., 2015; Hossain et al., 2004; Maggioni et al., 2011; Schreiner-McGraw & Ajami, 2020; Serpetzoglou et al., 2010). Precipitation uncertainty and error also depend on spatial and temporal resolution, with random errors tending to diminish with aggregation in space or time (P. Kirstetter et al., 2018; Quintero et al., 2016; Sarachi et al., 2015). The same is true when erroneous precipitation is used to predict streamflow, since river networks serve to aggregate rainfall-runoff errors over spatial and temporal scales (Maggioni et al., 2013; Nikolopoulos et al., 2010). Because of these issues and the limits they impose on large-scale water modeling, characterizing the space-time autocorrelation structure of SMP error at arbitrary space-time scales has been called a “grand challenge” for the precipitation community (Huffman et al., 2019). This work takes aim at this grand challenge by attempting to model the space-time autocorrelation of SMP error; the proposed approach could be applied to precipitation estimates from satellite-assimilating numerical weather models or blended datasets due to the aforementioned broad similarities in their error/uncertainty characteristics.

A significant challenge in addressing the space-time correlation structure of SMP error is the nonstationarity and anisotropy of SMP error structures, which this study hypothesizes are closely linked to the nonstationarity and anisotropy of rainfall fields themselves. For example, the spatiotemporal structure of SMP error is likely very different during an elongated frontal storm than during an isolated convective event or a highly-coherent tropical cyclone. This suggests that it would likely prove very challenging to develop robust characterizations of these structures based on a climatology of past storms, at least in a way that could be used operationally to supply uncertainty information to end users. As will be seen, we avoid such an approach, diverging from previous attempts to address this challenge.

It should be noted that the findings from the numerous validation studies that have assessed SMP accuracy relative to ground-reference data (e.g., Asong et al., 2017; Gadelha et al., 2019; N. Li et al., 2016; Tian et al., 2009 to name just a few) are not directly useful for SMP-based water modeling applications. This is because the metrics they calculate—such as mean squared or absolute errors, biases, and probabilities of detection and false alarms—do not readily translate into “new” (i.e., better) precipitation fields that are needed as model inputs. They do, however, highlight the challenge of providing better inputs by showing the prevalence, complexity, and magnitudes of such errors. In recognition of this, Gebremichael et al. (2011) called for a shift in SMP error characterization work towards “converting deterministic satellite rainfall estimates to probabilistic form by overlaying an estimated error distribution around the deterministic rainfall estimate.” In addition to our work presented here, several earlier efforts—detailed in Section 2—have answered this call by introducing techniques that can generate distributions to characterize the uncertainty of a specific SMP estimate.

While precipitation uncertainty is a random variable that can be described probabilistically—typically via a probability distribution describing the possible "true" (but unknown) precipitation rate at a given location and time—virtually all water prediction models are formulated to ingest deterministic precipitation estimates. This disconnect between probabilistic precipitation uncertainty and the need for deterministic input can be bridged by ensemble methods, in which multiple realizations of possible precipitation can be generated which, in their totality, reflect the range of uncertainty. These can be used to force an ensemble of water model simulations that then hopefully provide useful estimates of hydrologic modeling uncertainty. Ensemble methods are well-developed in the numerical weather prediction community, since members can be created by perturbing the initial conditions, boundary conditions, or parameters of a numerical atmospheric model (Cuo et al., 2011). Ensemble methods are much less developed in the SMP community, for reasons that are difficult to summarize and beyond the scope of this work. Nonetheless, a relatively limited set of studies have used ensemble approaches to assess propagation of SMP error through hydrological and land surface models (Falck et al., 2015; Gottschalck et al., 2005; Hossain & Anagnostou, 2005; Nijssen & Lettenmaier, 2004; Serpetzoglou et al., 2010; Shrestha et al., 2020). These studies relied on ground reference data both to characterize SMP uncertainty and to simulate the space-time correlation structure of SMP error.

This work introduces the Space-Time Rainfall Error and Autocorrelation Model (STREAM), which combines the simulated nonstationary, anisotropic space-time autocorrelation structure of precipitation error with pixel-scale estimates of precipitation uncertainty (Figure 1). STREAM uses an ensemble-based approach, generating realizations of "reference-like" precipitation fields—that is, fields that individually represent plausible realizations of the true (unknown) precipitation based on satellite precipitation estimates, and together represent the range of possible true rainfall (Figure 1). While not demonstrated here, the output from STREAM can be ingested by hydrologic or land surface models without requiring any modification to these models' structures. Uniquely, STREAM's space-time autocorrelation component is calibration-free and requires no ground-reference data. This capability, demonstrated below, rests on the hypothesis—which appears to be confirmed by our results—that the known space-time structure of SMP fields themselves provides a useful approximation of the unknown space-time structure of SMP error fields. This paper is organized as follows: Past error modeling work is summarized in Section 2. Section 3 describes the study region and data. The methodologies for STREAM and a previous error modeling approach, SREM2D, are covered in Section 4. Model results are shown and discussed in Sections 5 and 6, respectively, and the contributions of this work are summarized in Section 7.

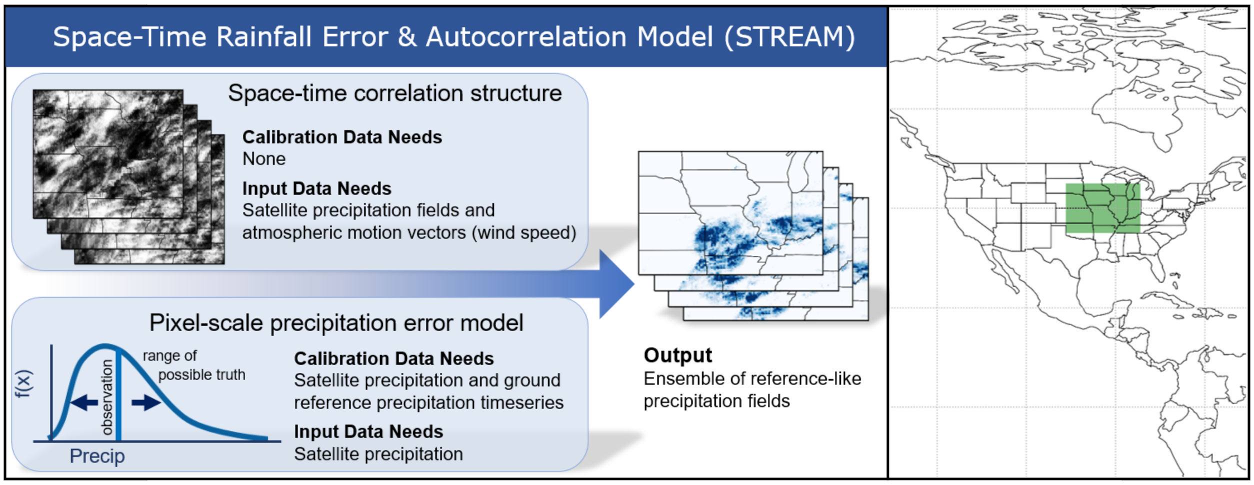


Figure 1. (left) Simple STREAM schematic and (right) study area highlighted in green in the central Continental United States (CONUS).

2 Background—Satellite Precipitation Error Modeling

Although the terms "error" and "uncertainty" are sometimes used interchangeably in the literature, in this paper we use error to refer to quantifiable differences between specific precipitation estimates and higher accuracy "ground truth" precipitation estimates, while using uncertainty to refer to the distribution of the possible true values relative to a precipitation estimate. For instance, the error for a given precipitation estimate is a deterministic value which can be calculated provided that high-quality ground truth data is available. In the absence of ground truth, this error is unknowable, and thus the best we can hope for is to know the uncertainty for that estimate—e.g. a range or distribution of plausible values which could be estimated through a variety of methods including those reviewed here. Regardless of our preferred terminology, the past literature uses the term "error model" to describe a method that provides an estimated distribution or range of possible true values based on an SMP observation. We keep with that terminological convention throughout this study.

Error models for SMP data can be placed in two categories: 1) pixel-scale error models, which characterize the SMP uncertainty associated with a single SMP estimate for a single control volume (invariably a grid cell) and time-step but do not consider the space-time autocorrelation structures between times and control volumes; and 2) space-time error models, which attempt to model the autocorrelation of SMP error. Both types, and the latter one in particular, have relied on extensive ground reference data for calibration. Additionally, space-time models have thus far neglected the nonstationarity and anisotropy in SMP error fields. Both categories face the challenge of representing the diversity of possible SMP errors—namely false alarms, missed precipitation, and hit errors (when a SMP estimate correctly detects rainfall but incorrectly estimates the magnitude). Some error models have focused entirely on hit cases while neglecting false alarms and missed cases (Reichle et al., 2007; Sarachi et al., 2015), while others handle rainfall detection and magnitude separately, resulting in either disjointed or overly complex model frameworks (Maggioni et al., 2014).

In pixel-scale error models, the uncertainty associated with a specific SMP estimate is described by a probability of precipitation and distribution of nonzero precipitation values which are conditional on the value of a particular SMP observation. It is worthy to note that some pixel-scale error models consider only hit cases and neglect the probability of precipitation component. Pixel-scale error models in literature include the Censored Shifted Gamma Distribution (CSGD; Wright et al., 2017), Precipitation Uncertainties for Satellite Hydrology (PUSH; Maggioni et al., 2014), and Probabilistic QPE using InfraRed Satellite Observations (PIRSO; Kirstetter et al., 2018), among others (Gebremichael et al., 2011). Sarachi et al. (2015) utilized a generalized normal distribution to model SMP uncertainty across scales by interpolating pixel-scale model parameters across various space-time resolutions. This approach considered hit cases only and required calibration at several scales. Pixel scale error models are advantageous in that they are trained using co-located timeseries of SMP and ground reference data and are therefore well suited to calibration using available rainfall records from sparse rain gage networks. Pixel-scale error models can also be “regionalized” by pooling together available training data from across a region to calibrate a regional error model (Hartke et al., 2020; Khan & Maggioni, 2020; Li et al., 2021). However, pixel-scale error models have no depiction of space-time autocorrelation; i.e. no way to relate the uncertainty of an SMP estimate in one pixel to the uncertainty in nearby pixels in space and time.

Space-time error models thus far have used calibration to characterize the climatological autocorrelation structure of precipitation error. The Two-Dimensional Satellite Rainfall Error Model (SREM2D) was developed by Hossain & Anagnostou (2006) in order to generate ensembles of SMP-like rainfall fields which preserve the error characteristics of SMP fields. Though SREM2D models the spatial correlation structure of SMP error fields as isotropic, these error fields often exhibit substantial anisotropy, reflecting the anisotropy inherent in real storm structures (Niemi et al., 2014; Zawadzki, 1973). Furthermore, SREM2D was not designed to represent differences in spatial autocorrelation of SMP error across a study area (i.e., spatial nonstationarity) and assumes that the average spatial correlation length calculated for a study region is representative of that region for all locations and time steps. Since the spatial correlation structure of SMP and SMP error can vary greatly at regional scales, this precludes SREM2D from application to large (i.e. subcontinental-to-global) scales. Because SREM2D relies on a climatological depiction of error autocorrelation, the model training and calibration process requires a gridded (or at least spatially extensive) ground-based precipitation dataset. Such datasets are lacking in many parts of the world (Kidd et al., 2017), further limiting is general applicability. Though applied to radar rainfall rather than SMP, Villarini et al. (2009) introduced an error-driven generator to stochastically perturb radar fields while accounting for the spatial correlation of multiplicative error. However, that error model considered hit cases only, neglected temporal correlation and anisotropy in error correlation structures, and used a computationally intensive method to generate Gaussian noise (Villarini et al., 2009). Space-time error models that rely on climatologically-calibrated parameters to simulate space-time correlation are not designed to simulate the unique correlation structure – i.e. varying degrees of anisotropy and correlation distances in space and time – of precipitation error that is associated with each new storm system.

The STREAM framework introduced in this article utilizes a calibration-free approach to modeling the space-time autocorrelation structure of precipitation error and provides a way to leverage pixel-scale estimates of precipitation uncertainty in space and time. Although this work utilizes a subcontinental study area, STREAM’s approach of reproducing the local spatial autocorrelation structures of SMP fields enables continental- to global-scale application.

3 Study Region and Data

3.1 Study Region

The study area covers the central U.S. (Figure 1; 100° to 85° W, 35° to 45° N), a region known for high agricultural production (Prince et al., 2001) and also marked by flood events often caused by heavy, long-lasting precipitation that severely impact local communities (e.g. the 1993 Mississippi River and 2008 Iowa flood events; Budikova et al., 2010; Najibi et al., 2016; Nakamura et al., 2013; Smith et al., 2013). Intense events provide a significant portion of the region's annual precipitation total, and convective storm systems are frequent during the warm summer period. The topography of this region is fairly uniform (Andresen et al., 2012).

3.2 Rainfall Data

The NASA Integrated MultisatellitE Retrievals for Global Precipitation Measurement (IMERG) Version 06 product is available globally at a 30-minute, 0.1˚ resolution and consists of three latency options (Huffman et al., 2019): IMERG Early (4-hour latency; lacks some data sources and data processing elements of longer latencies), IMERG Late (12-hour latency), and IMERG Final product (approximately 2.5-month latency; includes a gage-based correction). IMERG precipitation estimates are calculated by merging data from passive microwave (PMW) sensors, intercalibrating PMW estimates with a dual-frequency precipitation radar aboard the Global Precipitation Measurement (GPM) Core Observatory satellite, and interpolating (or "morphing") the resulting estimates in time using water vapor motion vectors from MERRA-2 and GEOS-5 (see Huffman et al., 2019; Tan et al., 2016 for more details). This study uses IMERG Early, aggregated to the hourly scale to match the radar-gage ground reference product; the approach could be readily applied to other IMERG latencies, as well as to other SMP products.

The NEXRAD Stage IV radar-gage product, available over CONUS at an hourly, roughly 1/24˚ resolution (Lin, 2011), is used as the ground reference in this study. Although NEXRAD's Stage IV product contains errors stemming from issues such as beam blockage and range from the nearest radar, we assume that the errors in this product are infrequent and negligible relative to IMERG, consistent with previous SMP studies (e.g. Aghakouchak et al., 2011) and consistent with our own prior experience using the dataset in this region. We upscaled Stage IV to IMERG's native 0.1˚ resolution using bilinear interpolation.

IMERG-Early (hereinafter IMERG) and Stage IV data from 2005-2007 were used for calibration of all models, while data from 2008-2013 for validation. To minimize issues related to frozen precipitation and maintain an accurate ground-reference during model calibration and validation, Stage IV and IMERG data were used only for March through October, excluding months with greater likelihood of frozen precipitation in the study area (November-February). This is admittedly a limitation of our study that should be addressed in the future. For both Stage IV and IMERG, the threshold for precipitation detection was set to 0.1 mm/hr, below which all hourly estimates were set to zero. This detection threshold is consistent with previous SMP studies (Germann & Zawadzki, 2002; Li et al., 2021).

3.3 Wind Data

As an approximation of the "steering winds" that govern the motion of storm systems, 850 mb wind fields were retrieved from the global MERRA-2 reanalysis product (Gelaro et al., 2017) at a hourly 0.5˚ by 0.625˚ resolution. These were regridded to 0.1˚. These wind fields were used together with IMERG fields to simulate the temporal evolution and autocorrelation structure of SMP error in STREAM (described in Section 4.2). Other sources of motion vectors could be used, including potentially those used in IMERG's aforementioned "morphing" space-time interpolation scheme. Those motion vectors are not publically available, however, so were not considered here. This point is discussed further in Section 6.4.

4 Methods

4.1 Censored Shifted Gamma Distribution Error Model

The Censored Shifted Gamma Distribution Error (CSGD) model framework was introduced by Scheuerer and Hamill (2015) to model uncertainty in numerical weather forecasts, and was adapted in Wright et al. (2017) to characterize pixel-scale SMP error across CONUS. The CSGD is an adaptation of the two-parameter gamma distribution (here written in terms of its mean and standard deviation, but which can be reparametrized in terms of shape and scale parameters) with an additional "shift" parameter δ that shifts the probability density function (PDF) leftward (Figure 2a). The density left of zero represents the probability of zero precipitation, while the density at any value greater than zero represents the likelihood of that amount of precipitation (Figure 2a, 2b). The shifted distribution is then left-censored at zero, replacing all negative values with zero. While previous precipitation error models either focused only on hit errors or required separate components to model rainfall occurrence and magnitude (see Section 2), the CSGD error model characterizes both the discrete and continuous components of satellite precipitation error using this single distribution. A regression model is trained based on contemporaneous co-located SMP and ground-truth observations to produce model parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, ... and, at any time t, unique "conditional" CSGD parameters $\mu(t)$, $\sigma(t)$, and $\delta(t)$ as a function of those parameters and the SMP estimate $R_s(t)$:

$$\mu(t) = \frac{\mu_c}{\alpha_1} \mathrm{log1p}\left\{\mathrm{expm1}(\alpha_1)\left[\alpha_2 + \alpha_3 \frac{R_s(t)}{\bar{R}}\right]\right\} \quad \text{Eq. 1}$$

$$\sigma(t) = \alpha_4 \sigma_c \sqrt{\frac{\mu(t)}{\mu_c}} \quad \text{Eq. 2}$$

$$\delta(t) = \delta_c \quad \text{Eq. 3}$$

where $\bar{R}$ is the mean of the SMP timeseries during the training period and $(\mu_c, \sigma_c, \delta_c)$ are the parameters of the climatological CSGD, a CSGD fit to the SMP time series. The regression model defined by Equations 1-3 allows the model to capture nonlinear behavior of SMP error across increasing precipitation rates. A simpler linear regression system can also be used in the CSGD error model framework by altering Eq. 1 (not shown; Scheuerer et al., 2015; Wright et al., 2017). The regression framework can also incorporate additional contemporaneous covariates $C_1(t)$, $C_2(t)$,.., $C_n(t)$, such as temperature or precipitable water, that could help to further characterize SMP uncertainty. These covariates are incorporated into the regression framework using an adjusted version of Eq. 1:

$$\mu(t) = \frac{\mu_c}{\alpha_1} \mathrm{log1p}\left\{\mathrm{expm1}(\alpha_1)\left[\alpha_2 + \alpha_3 \frac{R_s(t)}{\bar{R}} + \alpha_5 \frac{C_1(t)}{\bar{C_1}} + \alpha_6 \frac{C_2(t)}{\bar{C_2}} + \cdots\right]\right\} \quad \text{Eq. 4}$$

For more information on the CSGD error model framework, see Scheuerer and Hamill (2015) and Wright et al. (2017).

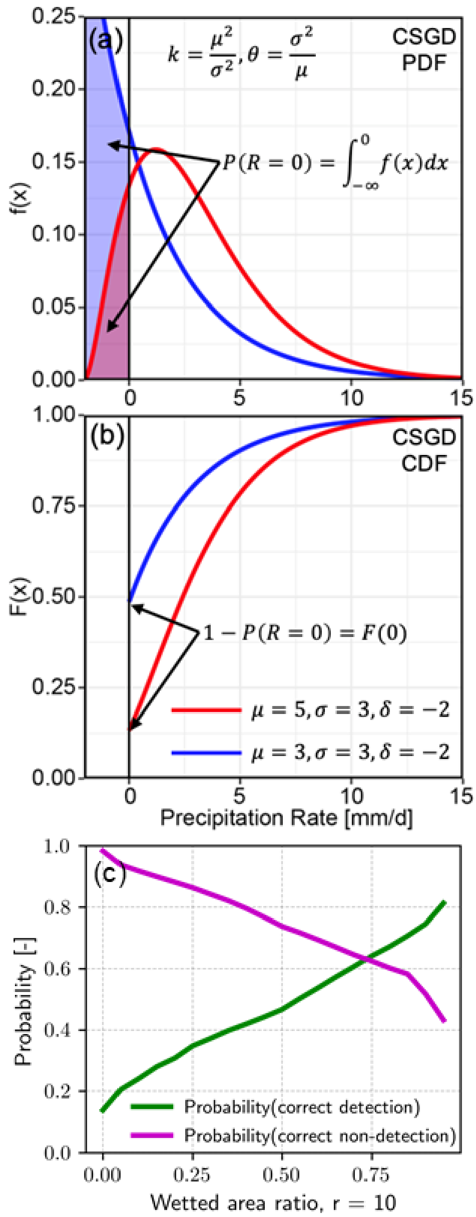


Figure 2. (a) Probability density function (PDF) and (b) cumulative density function (CDF) of two hypothetical censored shifted gamma distributions (CSGDs). (c) Observed probability of IMERG correct detection of nonzero rainfall (green) and probability of IMERG correct non-detection of rainfall (purple) as a function of the wetted area ratio (WAR) covariate. (c) uses data from entire study area for the period 2005-2007.

In this study, we use wetted area ratio (WAR) for the first time as a covariate in the CSGD error model. WAR for any IMERG estimate $R_s(t)$ at a given pixel is the proportion of pixels within a distance of r pixels that record nonzero rainfall at time t. WAR ranges from a value of 0 when no pixels within radius r have a nonzero precipitation rate, to 1, when all pixels within radius r have precipitation. Because WAR captures the spatial "context" of an IMERG observation, it is a useful covariate for predicting detection/non-detection performance within the CSGD framework. Figure 2c demonstrates that the probability of an IMERG estimate of nonzero rainfall being a correct detection is much greater if the associated WAR is high (i.e. close to 1.0) than if it is low. Likewise, the probability of IMERG correctly not detecting rainfall is highest when WAR is close to 0 (Figure 2c). A radius of r = 10 pixels was used to calculate WAR in this work; higher and lower values of r did not significantly alter CSGD error model performance (results not shown).

In this study, CSGD error model parameters are trained using timeseries "pooled" together from 25 co-located IMERG and Stage IV pixels (i.e. a 0.5˚×0.5˚ area). CSGD error model training for each 0.5˚×0.5˚ window in the study area is performed using the regression system defined in Equations 1-3 with the WAR covariate. The parameter estimation is completed via mean continuous ranked probability score (CRPS) minimization methods described in Scheuerer et al. (2015). Using timeseries from multiple pixels reduces sampling error and generates a more robust error model than model training using timeseries from a single IMERG pixel (not shown). This approach is suitable for the relatively homogenous terrain in the study area but may not be appropriate in more complex terrain where IMERG error characteristics are more closely tied to terrain heterogeneity.

4.2 The Space-Time Rainfall Error and Autocorrelation Model (STREAM)

4.2.1 Nonstationary anisotropic stochastic noise from pySTEPS

Nerini et al. (2017) introduced a non-stationary stochastic generator for radar precipitation fields using the short-space Fourier transform (SSFT). The Fourier power spectrum of a precipitation field (e.g. from weather radar or SMP) is convolved with Gaussian white noise to generate correlated Gaussian noise fields and ultimately produce an ensemble of precipitation forecasts which maintain the anisotropy and spatial correlation structure of observed radar rainfall fields. This methodology reproduces both the global and local power spectra of radar fields by using a moving window scheme. This moving window can thus capture spatial nonstationarity in field properties, since at any particular location the correlated noise is based on properties within the window. This SSFT-based non-stationary noise generator has since been incorporated into the pySTEPS Python library for short-range probabilistic precipitation forecasting, as a tool for generating ensemble nowcasts (Pulkkinen et al., 2019). While Nerini et al. (2019) used this tool to generate stochastic precipitation fields that replicate the local spatial correlation structure of observed radar rainfall fields, the authors emphasized that it could be applied to other applications involving complex non-stationary fields. Notably, this approach requires no calibration against ground truth measurements or parameterization of long-term precipitation behavior.

4.2.2 Correlated noise ensemble generation

In the first step of STREAM, the pysteps noise generator described in Section 4.2.1 is applied to stochastically generate Gaussian noise that replicates the local spatial correlation structure of an IMERG field, including anisotropy (Figure 3). After the initial noise field has been created for each ensemble member, each noise field is advected at an hourly time step via steering winds (described in Section 3.3) using a semi-Lagrangian scheme. In such a scheme, a time derivative (in this application, 850 mb wind vectors) is used to calculate where the value arriving at a grid cell, termed the arrival point, originated from in the previous time step (Lauritzen et al., 2011; Staniforth & Cote, 1991). This semi-Lagrangian scheme is advantageous over a strictly Lagrangian one because it does not allow individual parcels (in our case, noise values) to all advect into a single region and leave some regions without parcels. Our semi-Lagrangian scheme also incorporates a new instance of correlated noise, or a "shock term" (Nerini et al., 2017) which is the second term on the righthand side of Equation 5:

$$n_{t,i,j} = \alpha n_{t-1,i-v_t,\mathrm{j}-u_t} + \sqrt{1-\alpha^2}\,\tilde{n}_{t,i,j} \qquad \text{Eq. 5,}$$

where $n_{t,i,j}$ is a noise value to be calculated at time *t* and position (i,j) in the field and $n_{t-1,i-v_t,\mathrm{j}-u_t}$ is the noise value that has been advected by north-south and east-west wind vectors v_t and u_t from position $(i-v_t, \mathrm{j}-u_t)$ at time step $t-1$ to position (i,j) at time step *t*. v_t and u_t are obtained by multiplying MERRA2 wind vectors, originally in units of *m/s*, by 3600 seconds and dividing by 11,000 m, the approximate width of an IMERG pixel, to obtain units of 0.1° *pixel* hr^{-1}. $\tilde{n}_t$ is a new correlated Gaussian noise field based on the structure of IMERG at time *t*. The shock term is used to perturb the noise field and to incorporate the current IMERG spatial correlation structure, $\tilde{n}_t$, into the noise field at each time step. This allows the noise field to evolve over time and to reflect the nonstationary spatial correlation structure of IMERG and IMERG error. We assume that the error fields are first order autoregressive in time, calculating α as the Pearson correlation coefficient between IMERG fields at time t and $t-1$ (Figure 3). Analysis of the temporal

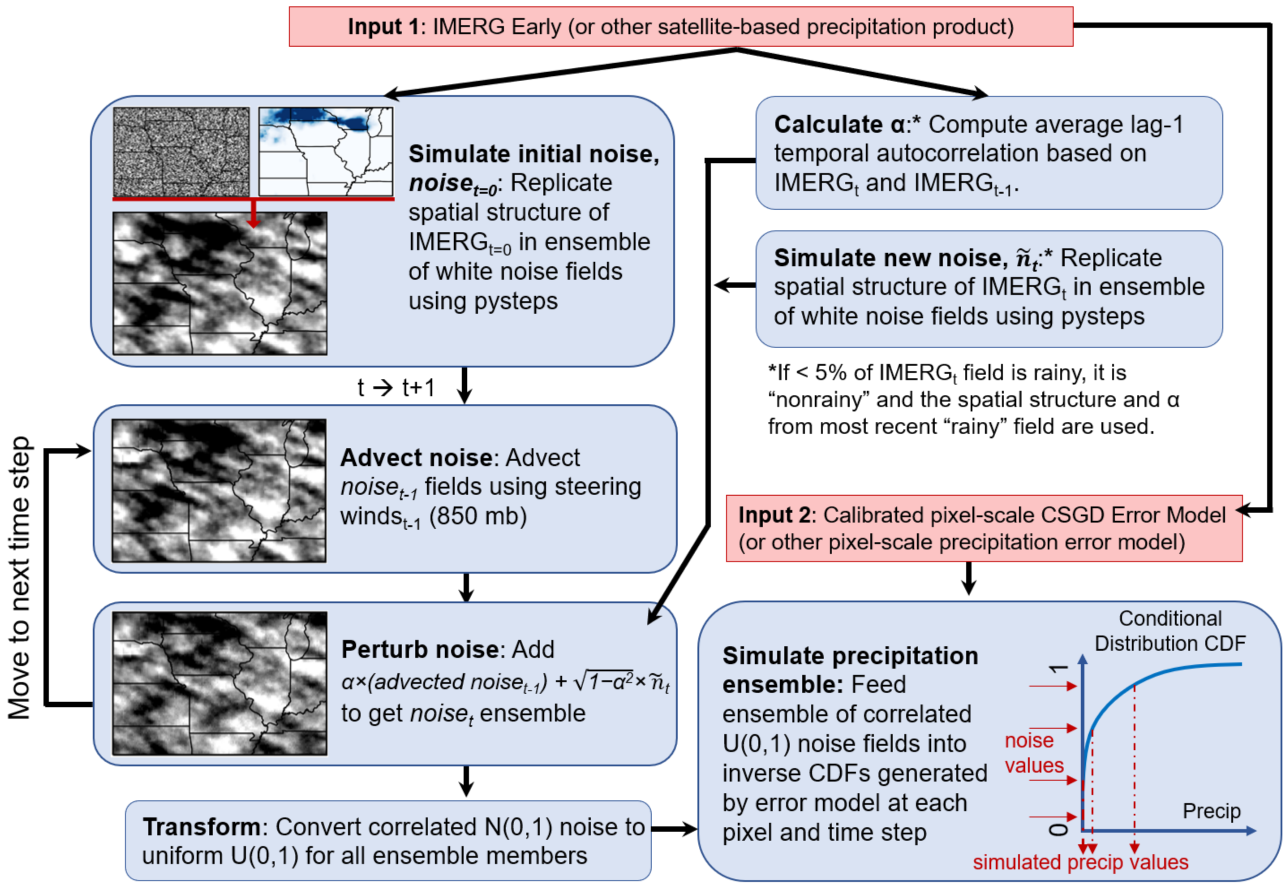


Figure 3. Schematic of STREAM methodology

autocorrelation function of IMERG error fields supports this autoregressive assumption (results not shown). After the noise ensemble has been generated for all time steps in the study period, the correlated Gaussian noise ensemble $N(0,1)$ is transformed to uniform noise $U(0,1)$ using the error function:

$$n_{uniform} = 0.5\left[1 + \text{erf}\left(\frac{n_{gaussian}}{\sqrt{2}}\right)\right] \qquad \text{Eq. 6.}$$

where $n_{gaussian}$ is the noise field described in Equation 5.

Note that correlated noise fields rn_t and temporal coefficient α are only calculated based on IMERG at time *t* when the IMERG field is "rainy," defined as when at least 5% of the study area registers rainfall (Figure 3). During time steps with non-rainy fields, which are frequent at the hourly scale, the spatial correlation structure from the most recent rainy field is used to generate rn_t. In either case, no parameters depend on a long-term climatology.

4.2.3 Precipitation ensemble generation

In the final step of STREAM, the correlated uniformly-distributed noise ensemble is combined with the CSGD error model. The CSGD model and training scheme methodology were briefly described in Section 4.1. The standard uniform noise values from the semi-Lagrangian scheme (Section 4.2.2) are inputted to the inverse CDF of the conditional CSGD generated at each time step and pixel. Thus, each noise value corresponds to a value of possible true precipitation conditional on a given IMERG estimate and associated WAR (Section 4.1), correlated with surrounding pixels. The uniform noise ensemble is censored at 0.995 to guard against unrealistically extreme precipitation values generated when very high noise values are used to select a precipitation value from conditional CSGDs with long tails. The output of STREAM consists of an ensemble of three-dimensional (north-south, east-west, time) precipitation fields, with each ensemble member representing one realization of the possible true precipitation across the study region for all time steps in the study period.

We also generated "uncorrelated" precipitation ensembles by using white (uncorrelated) noise as input to the inverse CDF of conditional CSGDs, thus neglecting spatial and temporal correlation of errors. The precipitation ensemble generated in this way is henceforth referred to as the uncorrelated ensemble, though they are not strictly uncorrelated since the resulting precipitation fields will inevitably exhibit some autocorrelation stemming from the IMERG precipitation rates (albeit much weaker than that of the ground-reference, IMERG, or autocorrelated noise fields).

4.3 SREM2D

The SREM2D error model was designed to generate ensembles of "satellite-like" fields that replicate the error properties of an SMP dataset relative to a ground-reference (Hossain et al., 2006). SREM2D separately accounts for the spatial correlation of detection errors and precipitation rate errors, and uses the Turning Bands algorithm (Mantoglou & Wilson, 1982) to generate 2-D Gaussian noise with correlation lengths matching that of the conditional error of SMP fields. In this work, SREM2D is run "in reverse" to generate reference-like rainfall fields that are closer to the ground-reference by replicating the error properties of Stage IV relative to IMERG Early. SREM2D has been used in this fashion previously in Falck et al. (2015) and Maggioni et al. (2013) to improve model-simulated streamflow estimates compared against hydrographs from SMPs. SREM2D parameters are trained using Stage IV and IMERG data for the 2005-2007 training period detailed in Section 3.2 and are listed in Table S1. Consistent with earlier SREM2D studies, additional trial-and-error calibration is needed (specifically, adjustment of the mean parameter) to minimize bias in SREM2D-perturbed fields. Note that these error parameters are calculated for the reference, Stage IV, relative to IMERG, highlighting SREM2D's need for ground reference data to characterize not only pixel-scale errors (akin to the CSGD approach) but also the spatiotemporal autocorrelation process (unlike STREAM, which doesn't require ground reference for this purpose).

4.4 Ensemble performance metrics

STREAM, SREM2D, and the uncorrelated ensembles were run at an hourly time step with an ensemble size of 50 for the evaluation period 2008-2013, excluding winter months (November–February). The spatial autocorrelation function (ACF), temporal ACF, probability of detection (POD), probability of false alarm (POFA), root mean square error (RMSE), and Containing Ratio were used to evaluate the performance of IMERG, the STREAM ensemble, the uncorrelated ensemble, and the SREM2D ensemble. The Containing Ratio (CR) is the proportion of observed data bracketed by the range of an ensemble, and has been used within the forecast verification and runoff modeling community to assess ensemble accuracy (Franz & Hogue, 2011; Xiong & O'Connor, 2008).

$$CR = \frac{1}{n}\sum_{t=1}^{n} I[R_{obs}(t)] \qquad \text{Eq. 7.}$$

where $I[\cdot]$ is an indicator function that equals 1 when the observed rainfall $R_{obs}(t)$ falls between the lowest and highest values of the ensemble at time *t* and that equals 0 when the observation falls outside ensemble bounds. For deterministic evaluation metrics, including RMSE, POD, and POFA, the mean of the ensemble was evaluated.

Spatial and temporal linear autocorrelation functions were calculated for each ensemble member to assess the ability of STREAM to generate reference-like precipitation fields in space and time. We note that assessing the space-time correlation structure of precipitation ensemble fields is not equivalent to assessing the space-time correlation structure of the SMP error introduced to create these fields; however, the correlation structures of SMP error fields can vary depending on the specific mathematical definition of SMP error. Since precipitation fields that resemble a ground-reference are the ultimate objective of an ensemble-based SMP error model, we chose to evaluate the ability of STREAM ensemble members to replicate the space-time correlation structures of Stage IV.

The above metrics were calculated for all precipitation datasets and ensembles at four space-time resolutions: 1-hour 0.1º, 1-hour 0.25º, 24-hour 0.1º, and 24-hour, 0.25º. Precipitation fields were regridded to coarser spatial resolutions using bilinear interpolation.

5 Results

Figure 4 shows IMERG, Stage IV, and outputs from the uncorrelated ensemble, STREAM autocorrelated noise and ensemble, and SREM2D ensemble for a six-hour period during a storm event in 2008 that led to heavy flooding in Cedar Rapids and Iowa City. Ensemble members shown in Figure 4 were chosen at random. While the uncorrelated ensemble fields do not resemble precipitation structures observed by Stage IV, STREAM and SREM2D fields visually resemble realistic precipitation structures from Stage IV, and STREAM also reproduces the observed anisotropy. The spatial correlation features generated in the STREAM noise fields clearly translate to similar spatial correlation features in STREAM precipitation fields. SREM2D fields exhibit less fine-scale anisotropic detail than STREAM, presumably due to its isotropic formulation.

Figure 5 provides additional event-scale analysis of STREAM, showing cumulative hourly precipitation and daily precipitation rates for a heavy rainfall event in June 2013 in southcentral Wisconsin. Area-averaged precipitation was calculated for the inset area in Figure 5a. The spatial autocorrelation function of precipitation within the inset area was also calculated to assess STREAM ensemble performance during this event, confirming STREAM ensemble members' ability to replicate the spatial structure of Stage IV rainfall (Figure 5c). The STREAM ensemble

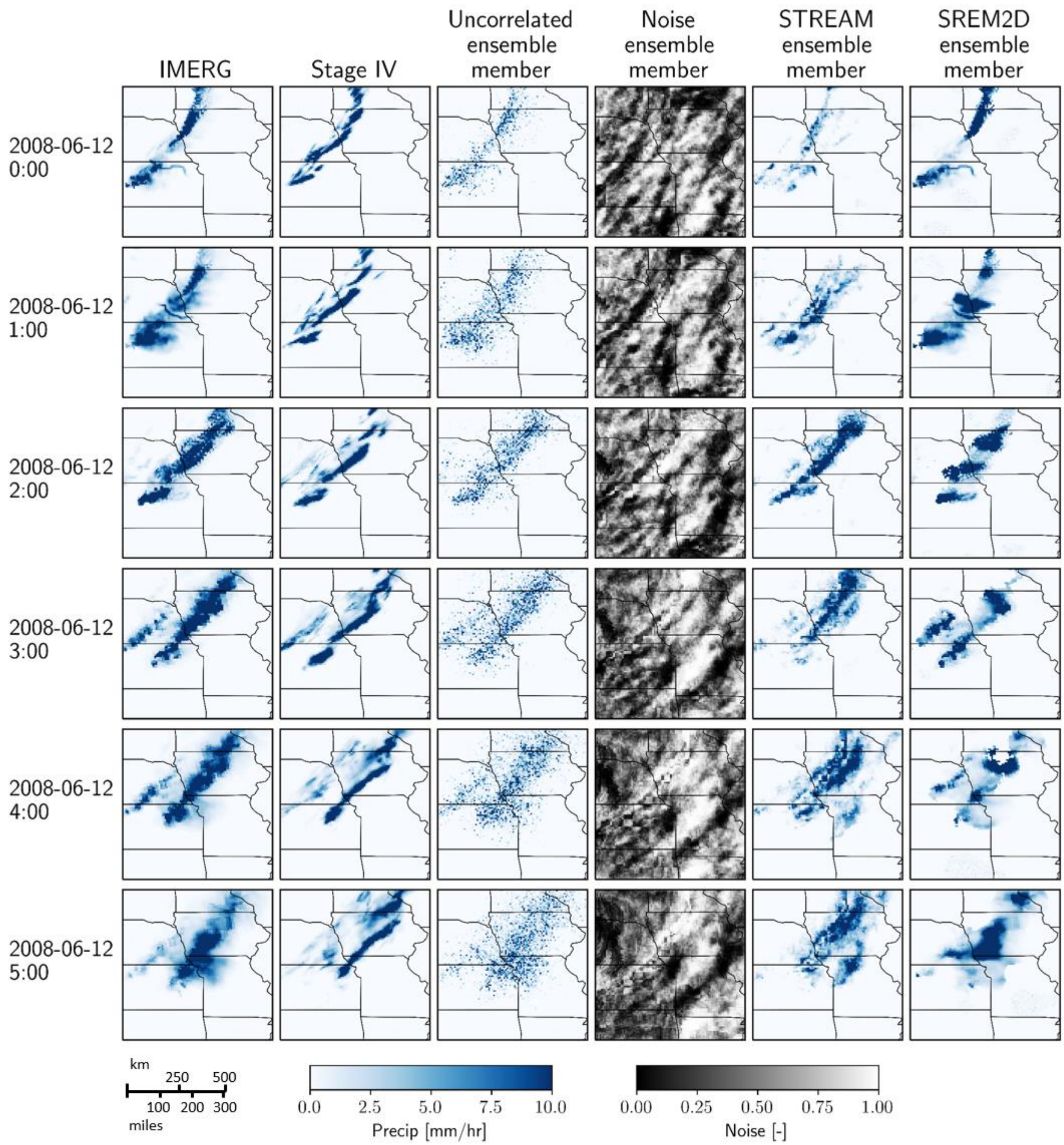


Figure 4. Example output of STREAM and other error modeling approaches. From left column to right column: IMERG, Stage IV, uncorrelated ensemble member, correlated noise ensemble member generated by STREAM, STREAM precipitation ensemble member, and SREM2D ensemble member during heavy rainfall event in study area on June 12, 2008. Ensemble members were chosen at random.

brackets the observed cumulative precipitation over the course of the event, reducing IMERG's stark overestimation (Figure 5b), and generally brackets observed precipitation rates at the daily scale, with the exception of two days (Figure 5d). Note that the uncertainty described by the range of the STREAM ensemble is small on days with low IMERG estimates, but widens when IMERG

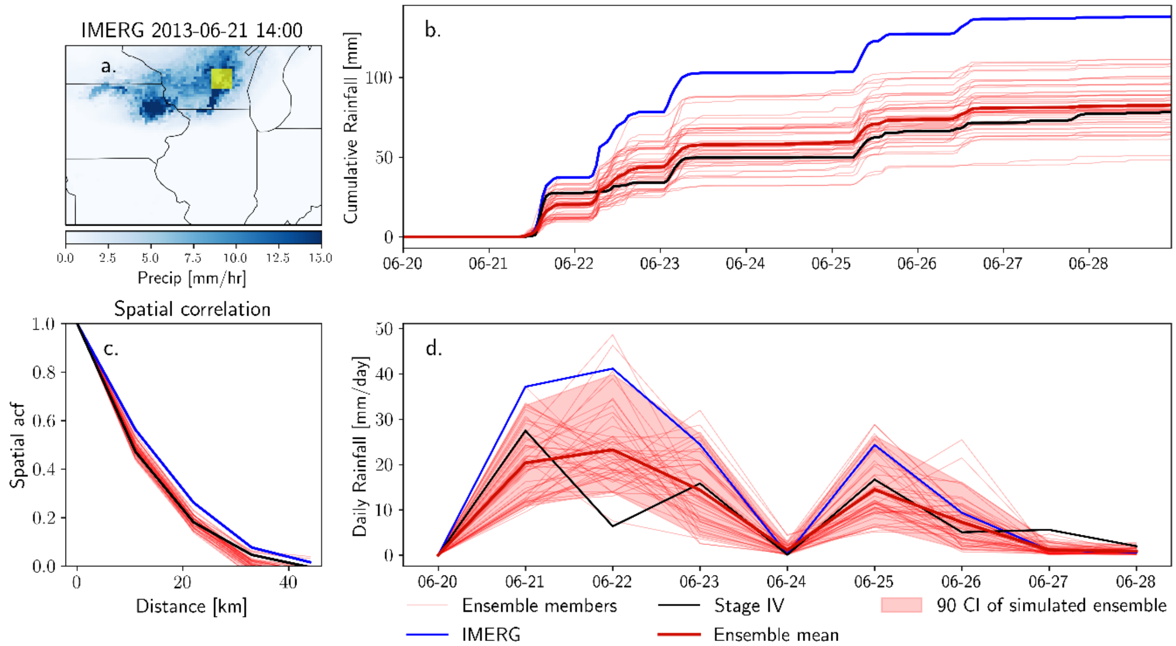


Figure 5. STREAM ensemble performance during 2013 flooding event in southcentral Wisconsin. (a) Area average precipitation is calculated over inset area in southcentral Wisconsin, denoted by yellow box (b) Hourly cumulative precipitation over course of event (c) Spatial autocorrelation function (ACF) calculated for precipitation in inset area over course of event (d) Daily precipitation rate over course of event.

observes nonzero rainfall, reflecting the greater range of random error in nonzero IMERG estimates (Figure 5d).

Figure 6 presents a seasonal-scale analysis of STREAM results, showing cumulative area-averaged spring precipitation (March–May) over eastern Iowa for all years in the validation period. The ensemble spread brackets the cumulative precipitation at the end of May in all years, regardless of whether IMERG over- or underestimates spring cumulative precipitation, except 2008, a year in which IMERG significantly underestimated cumulative precipitation. Precipitation in 2008 was well above the climatological average for all months shown in Figure 6, due in part to unprecedented rainfall occurring in the end of May and early June —conditions that likely pose a particular challenge for error modeling.

Figure 7 presents RMSE, POD, and POFA calculated over the entire study area and validation period for IMERG and all ensemble products at four space-time resolutions. The RMSE of IMERG and all ensemble means increases sharply for extreme hourly rainfall rates (> 8 mm/hr). The STREAM ensemble mean and uncorrelated ensemble mean exhibit reduced RMSE at all scales and across all rain rates, with the exception of heavy rain rates at an hourly scale. The SREM2D ensemble mean has a very similar RMSE to IMERG at all scales. The higher RMSE of the SREM2D ensemble mean relative to the STREAM ensemble aligns with results from Maggioni et al. (2011), who found that the relative RMSE of SREM2D-perturbed rainfall was slightly greater than that of the original satellite product.

The STREAM ensemble mean and uncorrelated ensemble mean exhibit higher POD across all space-time scales. Notably, the STREAM ensemble mean and uncorrelated ensemble mean are

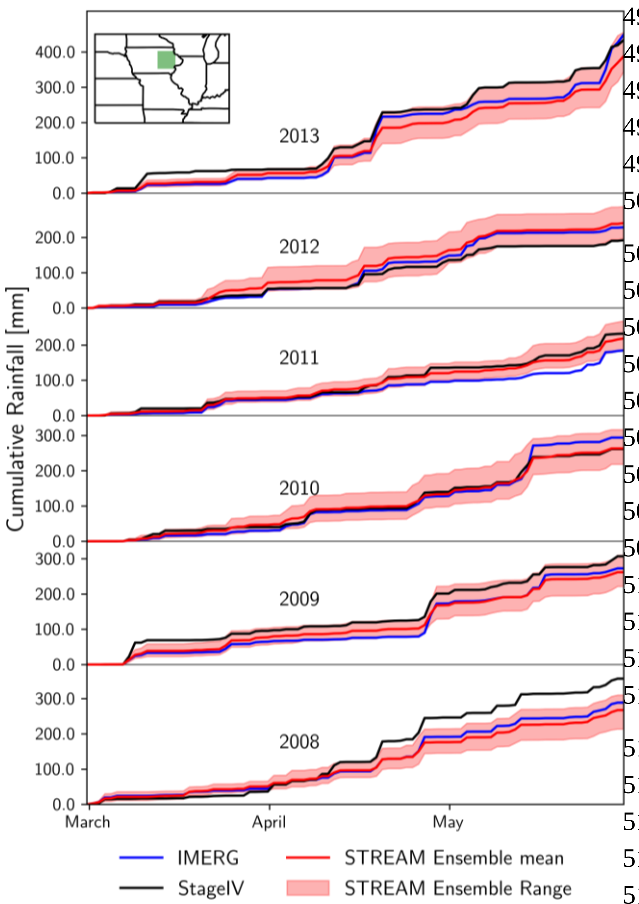


Figure 6. Cumulative area average rainfall over eastern Iowa subregion (green box in upper left inset map) estimated by IMERG (blue), Stage IV (black) and the STREAM ensemble (red).

able to simultaneously increase the POD while reducing the POFA at the hourly scale for precipitation rates greater than 1 mm/hr. The POFA of the STREAM and uncorrelated ensemble means are slightly higher than IMERG at the daily scale.

The spatial autocorrelation functions in the *x*- and *y*-directions (east-west and north-south, respectively) and the temporal autocorrelation function of IMERG, Stage IV, and ensemble fields are shown in Figure 8. Only ten members of each ensemble from STREAM, SREM2D, and the uncorrelated ensemble are displayed for clarity; since the ACFs are calculated over a long validation period, the ACFs of individual members within each error modeling method are nearly identical.

The correlation structure of STREAM ensemble fields nearly matches that of Stage IV at every scale (Figure 8), although the spatial ACF of ensemble fields—both in the *x*- and *y*-directions—is slightly lower than the spatial ACF of Stage IV. The uncorrelated ensemble members exhibit much lower spatial and temporal autocorrelation than Stage IV at the hourly scale, with the greatest difference at the finest spatial resolution. Once ensemble fields are aggregated to a coarser resolution (24-hr, 0.25°), all error model ensembles roughly replicate the average spatial and temporal autocorrelation functions of Stage IV. SREM2D ensemble members exhibit lower temporal autocorrelation than Stage IV at the hourly scale.

The Containing Ratios (CR) of the STREAM ensemble, SREM2D ensemble, and uncorrelated ensemble as a function of precipitation rate across four resolutions are presented in Figure 9. The STREAM ensemble consistently maintains a high CR (generally >0.8) across scales, though it dips at extreme rain rates. The STREAM ensemble brackets approximately 50% (70%) of the instances when ground-reference rainfall is greater than 8 mm/hr (35 mm/day) at an hourly (daily) resolution. SREM2D has a high containing ratio for Stage IV observations of zero precipitation but experiences a sharp decrease for nonzero values. The SREM2D ensemble fails to capture many of the nonzero ground-reference observations that the STREAM ensemble and uncorrelated ensemble successfully bracket. The performance of the uncorrelated ensemble degrades with increasing scale; at a daily, 0.25° scale, the uncorrelated ensemble fails to capture over 60% of the instances when the ground-reference observes rain rates greater than 30 mm/day.

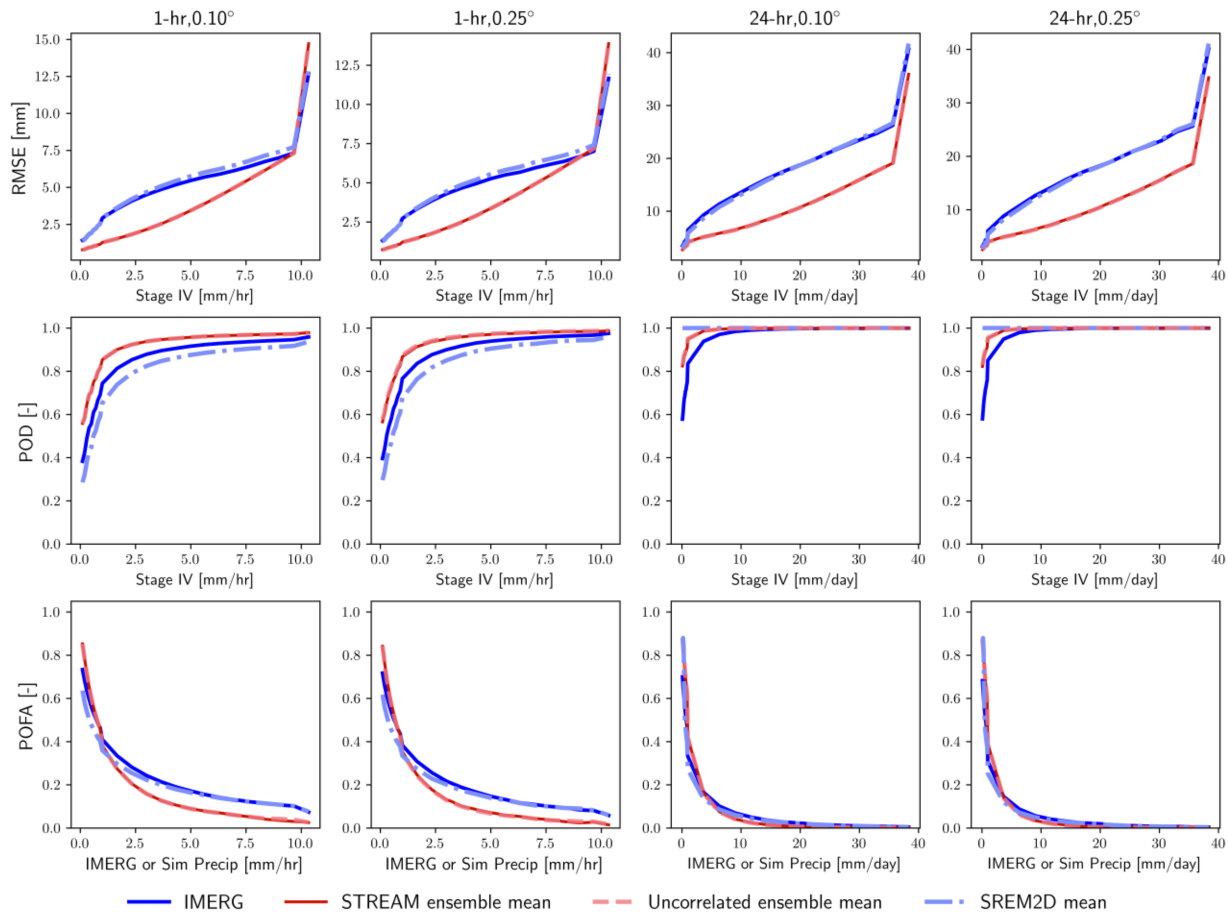


Figure 7. RMSE (top row), probability of detection (POD; middle row), and probability of false alarm (bottom row) for IMERG (blue), mean of STREAM ensemble (red), mean of uncorrelated ensemble (dashed pink), and mean of SREM2D ensemble (light blue) across four space-time resolutions. Metrics are calculated using study area-wide data for 2008-2013. STREAM ensemble and uncorrelated ensemble means are essentially identical (and are therefore difficult to distinguish from one another in the above plots).

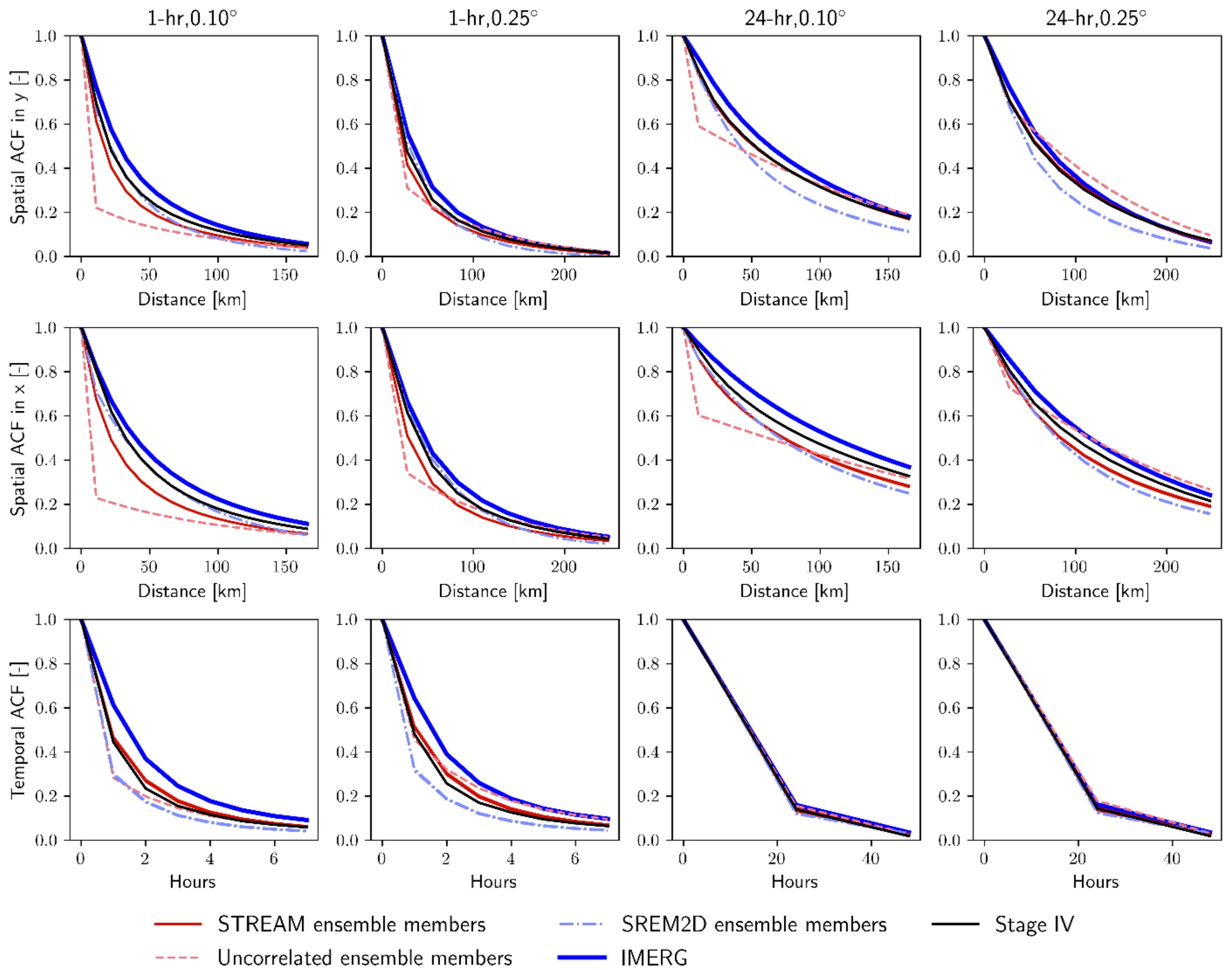
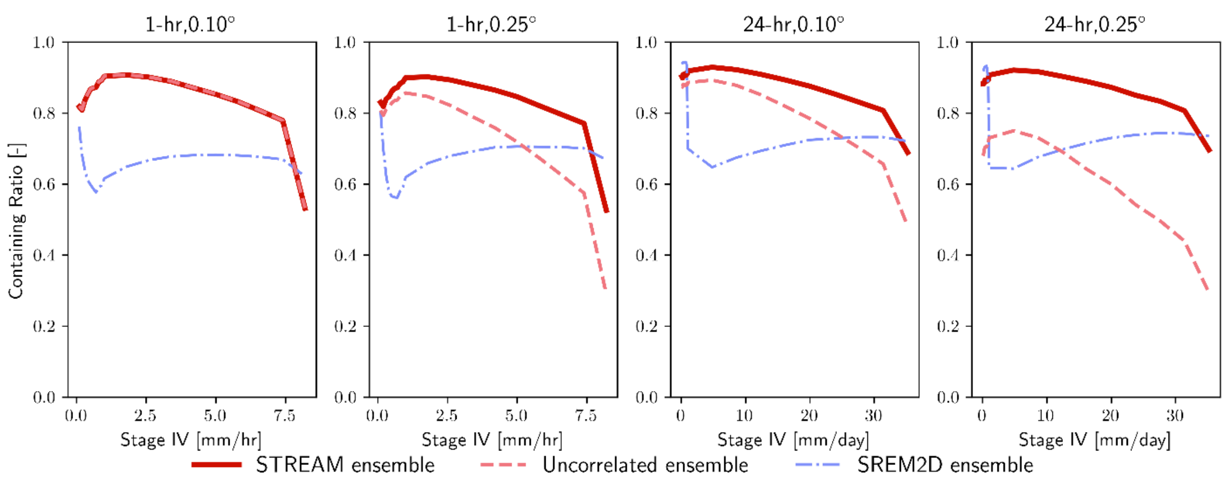


Figure 8. (Top) Spatial autocorrelation function (ACF) in the y-direction, (Middle) Spatial ACF in the x-direction, and (Bottom) Temporal ACF calculated for IMERG, Stage IV and 10 simulated rainfall ensemble members each from STREAM, SREM2D, and the uncorrelated ensemble at four resolutions.



549 **Figure 9**. Containing ratio (CR) of simulated ensembles at four resolutions.

6 Discussion

6.1 Performance of STREAM across multiple spatio-temporal scales

While the full STREAM ensemble at any given point in time and space represents the range of random error associated with IMERG, the ensemble mean represents the IMERG estimate adjusted only for systematic bias. Therefore, the mean of the ensemble will outperform individual ensemble members in terms of RMSE by strictly addressing systematic bias, but cannot capture the range of IMERG uncertainty on its own. The mean of the STREAM ensemble consistently reduces RMSE relative to IMERG across scales and rain rates except for hourly intensities greater than 10 mm/hr (Figure 7). This is likely due to the difficulty of predicting missed cases associated with heavy ground-reference rainfall. The STREAM ensemble mean has a higher or similar probability of detection relative to IMERG across rain rates and scales, with the greatest improvements achieved at lower precipitation rates (Figure 7). A portion of this improvement is due to the incorporation of the wetted area ratio (WAR) in the pixel-scale CSGD error model, which helps predict IMERG missed cases based on the presence or absence of nearby IMERG rainfall (Supplemental Figure S1). The probability of false alarm is slightly higher for the STREAM ensemble mean relative to IMERG at rates below 1 mm/hr at the hourly scale and below 10 mm/day at the daily scale. At the hourly scale (at both 0.1° and 0.25° resolutions), the probability of false alarm is significantly lower for the STREAM ensemble mean than for IMERG (Figure 7). At the hourly scale, the STREAM ensemble mean shows both a higher POD and lower POFA due to the use of the WAR covariate in the pixel-scale CSGD error model (Supplemental Figure S1). By removing the censoring of uniform noise greater than 0.995, the POD of STREAM can be slightly increased for low rain rates at the expense of a slight increase in POFA (Supplemental Figure 2). However, removal of this censoring component in STREAM can lead to 'INF' values in the precipitation ensemble when extremely high noise values are ingested by the inverse CDF of conditional distributions. The STREAM ensemble's ability to bracket ground-reference observations at event and seasonal scales (Figures 5 and 6) suggests that STREAM would be well-suited to creating inputs to hydrologic, land surface, or drought monitoring models—a direction that will be pursued in follow-on work.

We reran ensemble generation and analysis using Stage IV in place of IMERG in the correlated noise generation scheme (Figure 3) to understand if applying the ground-reference spatial correlation structure significantly improves ensemble performance; it does not (results not shown). This indicates that IMERG, although imperfect, provides valuable information about error correlation structures, on par with the information available through a ground-reference product.

STREAM was run for a 50-member ensemble in this work. Although performance metrics at our data's native pixel resolution (1-hr, 0.1°) are not impacted by an increase in ensemble size past 25, performance metrics at coarser resolutions (24-hr, 0.25°) improve with increasing ensemble size until a size of roughly 50 (results not shown). This reflects the increasing number of permutations of native resolution errors and error correlation structures that are combined during rescaling to coarser resolutions, leading to a greater range of possible precipitation estimates at coarse resolutions; the implications of this for water resources modeling are unclear and will be explored in future work. It is likely that at resolutions coarser than 24-hr and 0.25°, a larger STREAM ensemble could be beneficial.

Although an ensemble-based approach is currently the most feasible way to incorporate precipitation uncertainty into applications that ingest deterministic data, a large number of ensemble members may be required to accurately represent precipitation uncertainty. This may require prohibitive computing resources for the storage of precipitation outputs and the computational demands of hydrologic or land surface models. Although this study does not address this challenge, we note that very little work has been done in attempting to adapt the structure of environmental models to probabilistic precipitation inputs. As summarized in Nogueira (2020), large-scale precipitation estimates involve substantial uncertainties; thus, the adaption of models to ingest probabilistic precipitation data is an appropriate way to account for precipitation uncertainty (e.g. Hartke et al., 2020).

6.2 Comparison with SREM2D model

The STREAM ensemble meets or exceeds the performance of the SREM2D ensemble at all resolutions and rain rates except for the most extreme hourly rain rates (>10 mm/hr) when SREM2D exhibits a slightly higher containing ratio (Figure 9). SREM2D shows a particularly low containing ratio for light rainfall rates, meaning that SREM2D-perturbed IMERG fields often fail to bracket observed light rainfall rates. Visually, SREM2D fields exhibit more isotropic structure than IMERG, Stage IV, or STREAM ensemble fields (Figure 4). The noticeable drop in CR that occurs when observed rain rate shifts from zero to nonzero (Figure 9) is likely due to the separate handling of rainfall occurrence and hit errors in SREM2D. Even in the presence of plentiful ground data, a climatologically-trained approach to space-time correlation modeling, such as that used in SREM2D, is potentially problematic: each storm system is unique, so properties will deviate from a climatological training. The STREAM approach, in contrast, infers properties directly from each storm and thus foregoes the need for calibration or ground-reference data. STREAM's ability to outperform SREM2D suggests that the use of observed SMP space-time correlation is an attractive and practical alternative to the calibration-based simulation of error correlation.

6.3 Comparison with uncorrelated error modeling approach

The briefest visual analysis of the uncorrelated ensemble fields reveals that they do not resemble real precipitation, instead exhibiting scattered precipitation and little structure (Figure 4). The mean of the uncorrelated ensemble performs identically to the STREAM ensemble mean (Figure 7) because both ensemble means reflect a bias-corrected version of IMERG, but the range of the STREAM ensemble at coarser resolutions is much greater (compare Supplemental Figure S3 to Figure 5). At coarser space-time scales, the STREAM ensemble incorporates error correlation structures which allow ensemble members to simulate regional over- and underestimation by IMERG, ensuring greater variability among ensemble members. Meanwhile, the uncorrelated ensemble aggregates adjacent pixels with randomly simulated under and over-estimation, averaging out random errors and preventing any simulation of regional over- or underestimation. The uncorrelated ensemble's ability to bracket observed precipitation rates in fact worsens as the ensemble is aggregated to coarser resolutions (Figure 9). The improved performance of STREAM relative to the uncorrelated ensemble emphasizes the central importance of simulating the space-time correlation structure of precipitation error.

6.4 STREAM Future Adaptions

Although the demonstration of STREAM in this work uses the CSGD error model, other pixel-scale error models, such as PUSH (Maggioni et al., 2014) or PIRSO (Kirstetter et al., 2018) could likely be used within STREAM to represent IMERG uncertainty across arbitrary space-time scales. The CSGD error model is uniquely useful within STREAM, however, due to its ability to incorporate an arbitrary number of covariates to constrain pixel-scale uncertainty estimates.

The 850 mb steering winds from MERRA2 that are used here have a latency of several weeks. These data were chosen for illustrative purposes only; steering wind data could be obtained from lower-latency datasets such as from data-assimilating numerical weather forecasts or from the motion vectors used in the IMERG morphing scheme (Tan et al., 2019). This latter option would increase the consistency between how errors propagate over space and time within IMERG and how the correlated noise is propagated in STREAM's semi-Lagrangian advection scheme. This option was not pursued here since the IMERG motion vectors are not publically available; this may be pursued in future work.

7 Conclusions

The potential of satellite multi-sensor precipitation (SMP) products—and other large-scale precipitation sources with similar error/uncertainty properties, such as satellite-assimilating numerical weather models (NWM) and "blended" datasets that combine NWM and SMP data—in water resources modeling is limited by their uncertainties, which can mischaracterize both precipitation occurrence and intensity. Uncertainty during extreme precipitation events is particularly problematic for applications which assess hazards such as flooding or landsliding (e.g. Hartke et al., 2020; Jia et al., 2020; Prakash et al., 2016). Precipitation uncertainty and error vary according to spatial and temporal resolution, with random errors tending to "cancel out" when aggregated in space and time. SMP errors are autocorrelated in space and time, however, leading to regional (i.e. watershed scale) over- or underestimation by satellite-based products. This problem can be remedied using ensemble generation techniques that produce multiple plausible realizations of the unknown true precipitation field conditioned on the SMP observations. To incorporate precipitation uncertainty into applications which consider accumulated precipitation, such as flood prediction or drought monitoring, ensemble members must replicate the space-time correlation structure of precipitation error. This has been called a grand challenge within the precipitation community (Huffman et al., 2019), while the usability of other large-scale precipitation datasets would benefit from breakthroughs.

The Space-Time Rainfall Error and Autocorrelation Model (STREAM) combines space-time correlation structures with a pixel scale precipitation error model to generate precipitation ensembles that can "bracket" the magnitude and replicate the correlation structure of higher-accuracy "ground truth" rainfall fields. SMP-based STREAM ensembles are generated at high resolution (1-hour, 0.1°) and are shown to outperform the satellite product IMERG at several spatiotemporal scales. STREAM requires no ground-reference data to run and relies minimally on ground-reference data during calibration. Specifically, the approach taken to model spacetime correlation does not require any ground data and does not even require a "training period," since all necessary properties are inferred from IMERG and wind fields. STREAM ensembles generated at a high resolution can be aggregated to arbitrary space-time scales for use in hydrologic or land surface models while preserving the characteristics of real precipitation at these scales. The ensemble output of STREAM can be ingested in water modeling applications with no modification to those models' structures. This enables water resource predictions that reflect input precipitation uncertainty, though the computational demands of ensemble simulations may become burdensome.

Pixel-scale uncertainty (i.e. the probabilistic uncertainty surrounding a satellite-based precipitation estimate at a single pixel and time step) is the most feasible way to characterize SMP uncertainty around the world. In data-limited regions, pixel-scale precipitation error models can leverage available ground-reference data (i.e. sparse rain gage records), and pixel-scale uncertainty estimates can also be obtained via other approaches (i.e. Kirstetter et al., 2018a; Li et al., 2021). Taken alone, however, pixel scale uncertainty is of limited value in water resources applications because it offers no help in connecting or extending uncertainty estimates to nearby locations in space and time. STREAM allows users to combine pixel-scale precipitation uncertainty in space and time while accounting for nonstationary SMP error correlation structures. While not explored here, it appears that any pixel-scale uncertainty model—and not just the CSGD approach used here—can fit into the STREAM framework.

To be applicable to continental-to-global scale applications, a space-time SMP error model must rely minimally or not at all on ground-reference data. STREAM is shown to outperform a previous rainfall error model (SREM2D), which utilized extensive gridded ground-reference data for training SMP error and correlation properties. This work demonstrates that the anisotropic, nonstationary space-time correlation structure of SMP errors can be modeled using only SMP fields and atmospheric motion vectors. Meanwhile, ongoing work has demonstrated that the GPM Dual Precipitation Radar (DPR) instrument, which is quite accurate relative to other space-based microwave and infrared sensors, can be used to train pixel-scale error models (Khan et al., 2018; Li et al., 2021). Combining that approach with STREAM would completely eliminate the need for ground reference data, providing tools that could be used anywhere around the globe—though not without some shortcomings (Z. Li et al., 2021). In addition, the nonstationary and computationally efficient nature of the STREAM ensemble generation means that it could be applied at a global scale. Thus, while challenges remain, we believe that this work constitutes a meaningful step toward solving the grand challenge of characterizing precipitation error across arbitrary space-time scales.

Acknowledgments and Data

S.H. Hartke was supported by the NASA Earth and Space Science Fellowship Program (Award Number 80NSSC18K1321). Z. Li, D.B. Wright, D.B. Kirschbaum, and S. Khan were supported by the NASA Precipitation Measurement Mission (Award Number 80NSSC19K0951). The authors thank Dave Lorenz for providing guidance on space-time correlation structures and semi-Lagrangian schemes, and Danielle Nerini for his guidance in using pysteps noise generation functions. The authors thank Aline Falck for sharing SREM2D calibration and code materials.

IMERG-Early and MERRA-2 data is available from NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC) (https://disc.gsfc.nasa.gov/). NEXRAD Stage IV data is available from the Earth Observing Laboratory data archive (https://data.eol.ucar.edu/dataset/21.093). STREAM code is available via github repository (https://github.com/sam-hartke/STREAM).

References

Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research Atmospheres*, *116*(2), 1–11. https://doi.org/10.1029/2010JD014741

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Andresen, J., Hilberg, S., Kunkel, K., Winkler, J., Andresen, J., Hatfield, J., Bidwell, D., & Brown, D. (2012). *Historical Climate and Climate Trends in the Midwestern USA National Climate Assessment Midwest Technical Input Report* (Issue U.S. National Climate Assessment Midwest Technical Input Report). http://glisa.msu.edu/docs/NCA/MTIT_Historical.pdf.

Asong, Z. E., Razavi, S., Wheater, H. S., & Wong, J. S. (2017). Evaluation of integrated multisatellite retrievals for GPM (IMERG) over Southern Canada against ground precipitation observations: A preliminary assessment. *Journal of Hydrometeorology*, *18*(4), 1033–1050. https://doi.org/10.1175/JHM-D-16-0187.1

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., & Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Budikova, D., Coleman, J. S. M. M., Strope, S. A., & Austin, A. (2010). Hydroclimatology of the 2008 Midwest floods. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2010WR009206

Cuo, L., Pagano, T. C., & Wang, Q. J. (2011). A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. *Journal of Hydrometeorology*, *12*(5), 713–728. https://doi.org/10.1175/2011JHM1347.1

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

Franz, K. J., & Hogue, T. S. (2011). Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community. *Hydrology and Earth System Sciences*, *15*(11), 3367–3382. https://doi.org/10.5194/hess-15-3367-2011

Funk, C., Shukla, S., Thiaw, W. M., Rowland, J., Hoell, A., Mcnally, A., Husak, G., Novella, N., Budde, M., Peters-Lidard, C., Adoum, A., Galu, G., Korecha, D., Magadzire, T., Rodriguez, M., Robjhon, M., Bekele, E., Arsenault, K., Peterson, P., … Verdin, J. (2019). Recognizing the famine early warning systems network over 30 years of drought early warning science advances and partnerships promoting global food security. *Bulletin of the American Meteorological Society*, *100*(6), 1011–1027. https://doi.org/10.1175/BAMS-D-17-0233.1

Gadelha, A. N., Coelho, V. H. R., Xavier, A. C., Barbosa, L. R., Melo, D. C. D., Xuan, Y., Huffman, G. J., Petersen, W. A., & Almeida, C. das N. (2019). Grid box-level evaluation of IMERG over Brazil at various space and time scales. *Atmospheric Research*, *218*(October 2018), 231–244. https://doi.org/10.1016/j.atmosres.2018.12.001

Gebregiorgis, A. S., Kirstetter, P. E., Hong, Y. E., Gourley, J. J., Huffman, G. J., Petersen, W. A., Xue, X., & Schwaller, M. R. (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

Gebremichael, M., Liao, G. Y., & Yan, J. (2011). Nonparametric error model for a high resolution satellite rainfall product. *Water Resources Research*, *47*(7), 7504. https://doi.org/10.1029/2010WR009667

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., … Zhao, B. (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

Germann, U., & Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Monthly Weather Review*, *130*(12), 2859–2873. https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2

Gottschalck, J., Meng, J., Rodell, M., & Houser, P. (2005). Analysis of multiple precipitation products and preliminary assessment of their impact on Global Land Data Assimilation System land surface states. *Journal of Hydrometeorology*, *6*(5), 573–598. https://doi.org/10.1175/JHM437.1

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. (2005). Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture. *Advances in Water Resources*, *28*(12), 1336–1350. https://doi.org/10.1016/j.advwatres.2005.03.013

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

Hossain, F., Anagnostou, E. N., & Dinku, T. (2004). Sensitivity analyses of satellite rainfall retrieval and sampling error on flood prediction uncertainty. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(1), 130–139. https://doi.org/10.1109/TGRS.2003.818341

Huffman, G., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2019). NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA). In

Algorithm Theoretical Basis Document (ATBD) Version 06 (Issue March). https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg

Jia, G., Tang, Q., & Xu, I. X. (2020). Evaluating the performances of satellite-based rainfall data for global rainfall-induced landslide warnings. *Landslides*, *17*, 283. https://doi.org/10.1007/s10346-019-01277-6

Khan, S., & Maggioni, V. (2020). Evaluating the Applicability of the PUSH Framework to Quasi-Global Infrared Precipitation Retrievals at 0.5°/Daily Spatial/Temporal Resolution. *Asia-Pacific Journal of Atmospheric Sciences*. https://doi.org/10.1007/s13143-020-00185-3

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, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kirschbaum, D., & Stanley, T. (2018). Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. *Earth's Future*, *6*(3), 505–523. https://doi.org/10.1002/2017EF000715

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*(December 2017), 191–205. https://doi.org/10.1002/qj.3243

Kirstetter, P., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*, *144*(S1), 191–205. https://doi.org/10.1002/qj.3243

Lauritzen, P. H., Jablonowski, C., Taylor, M. A., & Nai, R. D. (2011). Atmospheric Transport Schemes: Desirable Properties and a Semi-Lagrangian View on Finite-Volume Discretizations. In *Numerical Techniques for Global Atmospheric Models* (pp. 185–250). Springer Berlin / Heidelberg.

Li, N., Tang, G., Zhao, P., Hong, Y., Gou, Y., & Yang, K. (2016). *Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin*. https://doi.org/10.1016/j.atmosres.2016.07.020

Li, Z., Wright, D., Hartke, S., Kirschbaum, D., Khan, S., Maggioni, V., & Kirstetter, P.-E. (2021). Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products based on GPM DPR. *Earth and Space Science Open Archive*. https://doi.org/10.1002/ESSOAR.10507263.1

Lin, Y., *GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0*, (2011) (testimony of Y. Lin). https://data.eol.ucar.edu/dataset/21.093

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. In *Journal of Hydrometeorology* (Vol. 17, Issue 4, pp. 1101–1117).

https://doi.org/10.1175/JHM-D-15-0190.1

Maggioni, V., Reichle, R. H., & Anagnostou, E. N. (2011). The effect of satellite rainfall error modeling on soil moisture prediction uncertainty. *Journal of Hydrometeorology*, *12*(3), 413–428. https://doi.org/10.1175/2011JHM1355.1

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. https://doi.org/10.1175/jhm-d-13-0112.1

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

Mantoglou, A., & Wilson, J. L. (1982). The Turning Bands Method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, *18*(5), 1379–1394. https://doi.org/10.1029/WR018i005p01379

Najibi, N., Devineni, N., & Lu, M. (2016). *Hydroclimate Drivers and Atmospheric Teleconnections of Long Duration Floods: An Application to Large Reservoirs in the Missouri River Basin*. http://www.elsevier.com/open-access/userlicense/1.0/2

Nakamura, J., Lall, U., Kushnir, Y., Robertson, A. W., & Seager, R. (2013). Dynamical structure of extreme floods in the U.S. Midwest and the United Kingdom. *Journal of Hydrometeorology*, *14*(2), 485–504. https://doi.org/10.1175/JHM-D-12-059.1

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. *Hydrology and Earth System Sciences*, *21*(6), 2777–2797. https://doi.org/10.5194/hess-21-2777-2017

Niemi, T. J., Kokkonen, T., & Seed, A. W. (2014). A simple and effective method for quantifying spatial anisotropy of time series of precipitation fields. *Water Resources Research*, *50*(7), 5906–5925. https://doi.org/10.1002/2013WR015190

Nijssen, B., & Lettenmaier, D. P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research D: Atmospheres*, *109*(2), 1–15. https://doi.org/10.1029/2003jd003497

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., & Borga, M. (2010). Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. *Journal of Hydrometeorology*, *11*(2), 520–532. https://doi.org/10.1175/2009JHM1169.1

Nogueira, M. (2020). Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences. *Journal of Hydrology*, *583*. https://doi.org/10.1016/j.jhydrol.2020.124632

Prakash, S., Mitra, A. K., Pai, D. S., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, *88*(December 2014), 1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Prince, S. D., Haskett, J., Steininger, M., Strand, H., & Wright, R. (2001). Net primary production of U.S. midwest croplands from agricultural harvest yield data. *Ecological Applications*, *11*(4), 1194–1205. https://doi.org/10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2

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. https://doi.org/10.5194/gmd-12-4185-2019

Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., & Seo, B. C. (2016). A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, *17*(8), 2137–2154. https://doi.org/10.1175/JHM-D-15-0195.1

Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research Atmospheres*, *112*(9). https://doi.org/10.1029/2006JD008033

Rodell, B. Y. M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., & Toll, D. (2004). THE GLOBAL LAND DATA ASSIMILATION SYSTEM. *Bull. Amer. Meteor. Soc.*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

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*. 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). https://doi.org/10.1029/2020WR027639

Serpetzoglou, E., Anagnostou, E. N., Papadopoulos, A., Nikolopoulos, E. I., & Maggioni, V. (2010). Error propagation of remote sensing rainfall estimates in soil moisture prediction from a land surface model. *Journal of Hydrometeorology*, *11*(3), 705–720. https://doi.org/10.1175/2009JHM1166.1

Shige, S., Kida, S., Ashiwake, H., Kubota, T., & Aonashi, K. (2013). Improvement of TMI rain retrievals in mountainous areas. *Journal of Applied Meteorology and Climatology*, *52*(1), 242–254. https://doi.org/10.1175/JAMC-D-12-074.1

Shrestha, A., Nair, A. S., & Indu, J. (2020). Role of precipitation forcing on the uncertainty of land surface model simulated soil moisture estimates. *Journal of Hydrology*, *580*(February 2019), 124264. https://doi.org/10.1016/j.jhydrol.2019.124264

Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B., & Krajewski, W. (2013). Extreme flood response: The june 2008 flooding in Iowa. *Journal of Hydrometeorology*, *14*(6), 1810–1825. https://doi.org/10.1175/JHM-D-12-0191.1

Staniforth, A., & Cote, J. (1991). Semi-Lagrangian integration schemes for atmospheric models - a review. *Monthly Weather Review*, *119*(9), 2206–2223. https://doi.org/10.1175/1520-0493(1991)119<2206:SLISFA>2.0.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., 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. https://doi.org/10.1175/JTECH-D-19-0114.1

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

Tan, J., Petersen, W., & 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.s1

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., & Hong, Y. (2020). Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sensing of Environment*, *240*(January). https://doi.org/10.1016/j.rse.2020.111697

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

Tian, Y., & Peters-Lidard, C. D. (2007). Systematic anomalies over inland water bodies in satellite-based precipitation estimates. *Geophysical Research Letters*, *34*(14), 1–5. https://doi.org/10.1029/2007GL030787

Tian, Y., & Peters-Lidard, C. D. (2010). A global map of uncertainties in satellite-based precipitation measurements. *Geophysical Research Letters*, *37*(24), 1–6. https://doi.org/10.1029/2010GL046008

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., Hsu, K. L., Turk, F. J., Garcia, M., & Zeng, J. (2009). Component analysis of errors in Satellite-based precipitation estimates. *Journal of Geophysical Research Atmospheres*, *114*(24), 1–15. https://doi.org/10.1029/2009JD011949

Villarini, G., Krajewski, W. F., Ciach, G. J., & Zimmerman, D. L. (2009). Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates. *Water Resources Research*, *45*(1), 1–11. https://doi.org/10.1029/2008WR006946

Winiwarter, V., Haidvogl, G., & Bürkner, M. (2016). The rise and fall of Munich’s early modern water network: a tale of prowess and power. *Water History*, *8*(3), 277–299. https://doi.org/10.1007/s12685-016-0173-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. *Journal of Hydrometeorology*, *18*(10), 2801–2815. https://doi.org/10.1175/JHM-D-17-0060.1

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research*, *50*(3), 2693–2717. https://doi.org/10.1002/2013WR014710

Xiong, L., & O'Connor, K. M. (2008). An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling. *Journal of Hydrology*, *349*(1–2), 115–124. https://doi.org/10.1016/j.jhydrol.2007.10.029

Zawadzki, I. I. (1973). Statistical Properties of Precipitation Patterns. *Journal of Applied Meteorology and Climatology*, *12*(3), 459–472. https://doi.org/https://doi.org/10.1175/1520-0450(1973)012<0459:SPOPP>2.0.CO;2