

Incorporating IMERG Satellite Precipitation Uncertainty into Seasonal and Peak Streamflow Predictions using the Hillslope Link Hydrological Model

Samantha Hartke¹, Daniel B Wright¹, Felipe Quintero², and Aline S Falck³

¹University of Wisconsin-Madison

²University of Iowa

³National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN)

December 7, 2022

Abstract

In global applications and data sparse regions, which comprise most of the earth, hydrologic model-based flood monitoring relies on precipitation data from satellite multisensor precipitation products or numerical weather forecasts. However, these products often exhibit substantial errors during the meteorological conditions that lead to flooding, including extreme rainfall. The propagation of precipitation forcing errors to predicted runoff and streamflow is scale-dependent and requires an understanding of the autocorrelation structure of precipitation errors, since error autocorrelation impacts the accumulation of precipitation errors over space and time in hydrologic models. Previous efforts to account for satellite precipitation uncertainty in hydrologic models have demonstrated the potential for improving streamflow estimates; however, these efforts use satellite precipitation error models that rely heavily on ground reference data such as rain gages or weather radar and do not characterize the nonstationarity of precipitation error autocorrelation structures. This work evaluates a new method, the Space-Time Rainfall Error and Autocorrelation Model (STREAM), which stochastically generates possible true precipitation fields, as input to the Hillslope Link Model to generate ensemble streamflow estimates. Unlike previous error models, STREAM represents the nonstationary and anisotropic autocorrelation structure of satellite 2 precipitation error and does not use any ground reference to do so. Ensemble streamflow predictions are compared with streamflow generated using satellite precipitation fields as well as a radar-gage precipitation dataset during peak flow events. Results demonstrate that this approach to accounting for precipitation uncertainty effectively characterizes the uncertainty in streamflow estimates and reduces the error of predicted streamflow. Streamflow ensembles forced by STREAM improve streamflow prediction nearly to the level obtained using ground-reference forcing data across basin sizes.

1 Incorporating IMERG Satellite Precipitation Uncertainty into Seasonal 2 and Peak Streamflow Predictions using the Hillslope Link Hydrological 3 Model

4 Samantha H. Hartke¹, Daniel B. Wright¹, Felipe Quintero², Aline S. Falck³

5 ¹Department of Civil and Environmental Engineering, University of Wisconsin-Madison,
6 Madison, Wisconsin, USA.

7 ²IIHR—Hydroscience & Engineering, Iowa Flood Center, University of Iowa, Iowa City, USA

8 ³National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São
9 José dos Campos, SP, Brazil

10 Abstract

11 In global applications and data sparse regions, which comprise most of the earth, hydrologic
12 model-based flood monitoring relies on precipitation data from satellite multisensor precipitation
13 products or numerical weather forecasts. However, these products often exhibit substantial errors
14 during the meteorological conditions that lead to flooding, including extreme rainfall. The
15 propagation of precipitation forcing errors to predicted runoff and streamflow is scale-dependent
16 and requires an understanding of the autocorrelation structure of precipitation errors, since error
17 autocorrelation impacts the accumulation of precipitation errors over space and time in hydrologic
18 models. Previous efforts to account for satellite precipitation uncertainty in hydrologic models
19 have demonstrated the potential for improving streamflow estimates; however, these efforts use
20 satellite precipitation error models that rely heavily on ground reference data such as rain gages or
21 weather radar and do not characterize the nonstationarity of precipitation error autocorrelation
22 structures. This work evaluates a new method, the Space-Time Rainfall Error and Autocorrelation
23 Model (STREAM), which stochastically generates possible true precipitation fields, as input to the
24 Hillslope Link Model to generate ensemble streamflow estimates. Unlike previous error models,
25 STREAM represents the nonstationary and anisotropic autocorrelation structure of satellite

26 precipitation error and does not use any ground reference to do so. Ensemble streamflow
27 predictions are compared with streamflow generated using satellite precipitation fields as well as
28 a radar-gage precipitation dataset during peak flow events. Results demonstrate that this approach
29 to accounting for precipitation uncertainty effectively characterizes the uncertainty in streamflow
30 estimates and reduces the error of predicted streamflow. Streamflow ensembles forced by
31 STREAM improve streamflow prediction nearly to the level obtained using ground-reference
32 forcing data across basin sizes.

33 Keywords: satellite methods, precipitation, catchment dynamics, uncertainty

34 1. Introduction

35 1.1 Motivation

36 Hydrologic models are central to efforts to mitigate the devastating impacts of floods and are
37 used around the world to predict streamflow, with particular interest in high streamflow events
38 that cause flooding (e.g., Alfieri et al., 2013; Wu et al., 2014). Despite continual advances, these
39 models cannot perfectly predict streamflow for a number of reasons. In addition to model
40 uncertainty (i.e. imperfect representation of physical processes and parameter uncertainty), forcing
41 uncertainty due to errors in rainfall data—whether obtained from gauge data, radar, numerical
42 weather prediction model (NWP) forecasts, or satellite multi-sensor precipitation (SMP)
43 products—leads to errors in runoff and streamflow estimates (Hong et al., 2006; Sperna Weiland
44 et al., 2015). In global hydrologic modeling studies, precipitation forcing data has been found to
45 be the primary driver of predictive uncertainty, and using parameter calibration to compensate for
46 biases in meteorological forcing data can actually result in inconsistent performance between
47 calibration and evaluation periods, leading to calls for “improved characterization of global rainfall

48 amounts at spatial resolutions of 0.5° and smaller” (Fekete et al., 2004; Sperna Weiland et al.,
49 2015).

50 Precipitation uncertainty in basins in high latitudes and the tropics leads to hydrologic
51 uncertainty of the same or greater magnitudes due to the relatively moist soil conditions and
52 propensity of rainfall to become runoff in these regions (Biemans et al., 2009; Fekete et al., 2004).
53 Precipitation error propagation is more variable in semi-arid regions where the rainfall-runoff
54 generation processes are highly nonlinear (Fekete et al., 2004). Initial and simulated soil moisture
55 conditions in watersheds, which are also critical to accurate streamflow prediction, are dependent
56 on multiple previous precipitation events (Alvarez-Garreton et al., 2014; Trambly et al., 2010,
57 2011).

58 The issue of precipitation uncertainty in streamflow prediction is particularly prevalent in
59 regional-to-global-scale efforts, which due to lack of alternatives, typically must rely on either
60 precipitation from satellites or numerical weather models—both of which are prone to
61 substantially larger errors than in-situ gage measurements. Examples include the Global Flood
62 Monitoring System (GFMS; Wu et al., 2014) and the Global Flood Awareness System (GloFAS;
63 Alfieri et al., 2013). These use NASA’s Integrated MultisatellitE Retrieval for the Global
64 Measurement Mission (IMERG) and the ECMWF Integrated Forecast System (EFS) data,
65 respectively, to generate global streamflow predictions. Similar to other satellite multi-sensor
66 precipitation products, IMERG precipitation estimates exhibit considerable systematic bias and
67 random error, and IMERG often overestimates rainfall during extreme precipitation events (e.g.,
68 Anjum et al., 2018; Asong et al., 2017; Gilewski & Nawalany, 2018; Omranian et al., 2018; Wang
69 et al., 2017). A number of studies have demonstrated that IMERG can introduce large uncertainties
70 into streamflow predictions when compared with gauge-based or radar-gauge products, although

71 the extent to which IMERG-based results underperform varies (Amorim et al., 2020; Jiang &
72 Bauer-Gottwein, 2019; N. Li et al., 2016). For instance, IMERG-forced streamflow predictions
73 across China were shown to exhibit substantially more error in arid watersheds than humid
74 watersheds (Jiang et al., 2019).

75 For a thorough review of the use of satellite multi-sensor precipitation (SMP) products in
76 hydrologic modeling, see Quintero et al. (2016). Hydrologic modeling is not the only satellite
77 precipitation application that contends with satellite precipitation uncertainty; land surface
78 modeling, snow simulations, groundwater modeling, and landslide hazard assessment are also
79 susceptible to errors in output when ingesting erroneous precipitation data (e.g., Hartke et al., 2020;
80 Maggioni et al., 2011; Marc et al., 2022; Pradhan & Indu, 2021; Raleigh et al., 2015; Schreiner-
81 McGraw & Ajami, 2020; Serpetzoglou et al., 2010; Shrestha et al., 2020). Although this work
82 focuses on accounting for uncertainty in SMP data, numerical weather prediction model (NWP)
83 precipitation fields also exhibit similar levels of bias and random error, including high
84 uncertainties during extreme events, due to a highly dynamic atmosphere and the sensitivity of
85 model precipitation forecasts to a range of model parameters (e.g., Lowrey & Yang, 2008; Luitel
86 et al., 2018; Moosavi et al., 2021; Nasrollahi et al., 2012; Nogueira, 2020).

87 1.2 Precipitation Error Modeling Background

88 Within hydrologic models runoff is accumulated over a range of basin sizes, and precipitation
89 error propagation is dependent on both basin scale as well as the autocorrelation of precipitation
90 errors in space and time (Cunha et al., 2012; Nijssen & Lettenmaier, 2004; Nikolopoulos et al.,
91 2010). Error autocorrelation describes the intuitive fact that an overestimation of precipitation by
92 IMERG at a pixel in a given timestep likely corresponds to an overestimation by IMERG in

93 surrounding pixels and timesteps. While accumulation during rainfall-runoff generation processes
94 can serve to “average out” random errors in precipitation datasets, it can also serve to accumulate
95 a field of precipitation with correlated errors (i.e. precipitation being overestimated over an entire
96 storm system) and propagate these into streamflow predictions. Cunha et al. (2012) found “the
97 efficiency of the river basin in filtering out random errors to be highly sensitive to the presence of
98 spatial correlation in errors,” and that “when rainfall errors are correlated in space, the process of
99 aggregation and attenuation by the river network is not as effective in filtering out uncertainties.”
100 Correctly simulating the autocorrelation of precipitation error fields is paramount to capturing
101 precipitation uncertainty at coarser scales (e.g., basin scales; Hartke et al., 2022). Just like the
102 correlation structure of precipitation, the correlation structure of precipitation error fields is
103 anisotropic and nonstationary in time and space, changing depending on the observed precipitation
104 system and retrieval conditions (i.e., errors may exhibit lower spatial autocorrelation during a
105 scattered precipitation event than during a larger—and more highly spatially autocorrelated—
106 frontal precipitation event).

107 A number of SMP error models have been introduced that generate distributions to describe
108 the uncertainty surrounding a given SMP estimate at a single time and pixel (i.e., Kirstetter et al.,
109 2018; Maggioni et al., 2014; Wright et al., 2017; see Section 3.3). One of the major limitations of
110 these “pixel-scale” error formulations is that they do not relate the uncertainty of SMP estimates
111 across space and time; there is no intuitive way to combine the uncertainty distributions at every
112 pixel in a field to create an ensemble of precipitation error fields. Previous work has attempted to
113 link pixel-scale precipitation uncertainty estimates using calibrated correlation coefficients or
114 lengths to quantify the space-time autocorrelation structure of errors (e.g., Ciach et al., 2007;
115 Hossain & Anagnostou, 2006), but such approaches do not represent the nonstationarity of satellite

116 precipitation errors and rely on ground reference data for calibration. Previous studies on ensemble
117 SMP error correction in hydrologic modeling have utilized the Two-Dimensional Satellite Rainfall
118 Error Model (SREM2D; Hossain & Anagnostou, 2006), which does not account for the
119 nonstationarity and anisotropy of SMP error correlation structures (Falck et al., 2015, 2018, 2021;
120 Maggioni et al., 2013). These studies demonstrate that the inclusion of precipitation uncertainty
121 using ensemble methods does generally improve streamflow prediction, although not always in
122 large basins when an error model applies bias correction to entire basin areas (as shown with
123 SREM2D in Falck et al., 2015), but do so using an approach that is not viable for most of the
124 world, due to a scarcity of ground-based data which is unlikely to change.

125 1.3 Incorporating satellite precipitation uncertainty into hydrologic modeling

126 Accounting for precipitation uncertainty in hydrologic modeling applications poses a challenge
127 because of the probabilistic nature of precipitation uncertainty, the nonstationary correlation
128 structure of precipitation errors, and the need to calibrate precipitation uncertainty models with
129 little to no ground-reference data in many parts of the world. The Space-Time Rainfall Error and
130 Autocorrelation Model (STREAM) was developed with an eye to overcoming these challenges
131 (Hartke et al. 2022). STREAM stochastically generates a precipitation ensemble that represents
132 the possible true precipitation based on a satellite precipitation field (or a numerical weather
133 forecast). It consists of two pieces: 1) using an uncalibrated, nonstationary error autocorrelation
134 scheme that requires no ground reference data and 2) pixel-scale uncertainty estimates that can be
135 obtained from existing error model formulations. By not calibrating the autocorrelation scheme on
136 historical ground reference or SMP error fields and instead replicating the changing, anisotropic
137 autocorrelation structure of SMP fields, STREAM nimbly incorporates nonstationarity into its
138 representation of SMP uncertainty. Like previous ensemble-based approaches, STREAM

139 ensembles can be ingested by SMP applications, including hydrologic models, without any
140 modification to the model structure.

141 One strategy to address precipitation uncertainty in hydrologic modeling is bias correction
142 (e.g., Charles et al., 2020; Ciupak et al., 2019; Habib et al., 2014; Ji et al., 2020). However, when
143 applied in near real-time this approach only adjusts for systematic bias and does not consider or
144 account for the substantial random error in SMP and NWP datasets. This work employs a bias
145 correction scheme in order to compare such an approach with one that considers the full range of
146 precipitation uncertainty – both systematic bias and random error.

147 The goal of this work is to understand how precipitation uncertainty impacts streamflow
148 estimates, with an emphasis on peak streamflow events, and assess whether STREAM presents a
149 suitable way to account for IMERG uncertainty and improve predicted streamflow. In this work,
150 IMERG uncertainty is incorporated into the Iowa Flood Center’s distributed hydrologic model, the
151 Hillslope Link Model (HLM), using precipitation ensemble fields generated by STREAM. The
152 resulting ensemble of streamflow estimates are compared against USGS observations as well as
153 streamflow estimates from HLM when forced with IMERG-Early, a bias-corrected version of
154 IMERG-Early, and a ground-reference product, NEXRAD Stage IV (Lin, 2011). Unlike previous
155 attempts to account for SMP uncertainty, namely by using SREM2D, STREAM ensembles model
156 the nonstationarity of SMP error correlation structures and do not rely on ground-reference
157 precipitation data to do so.

158 This paper is structured as follows: the study area in Iowa and precipitation and streamflow
159 datasets are described in Section 2. Section 3 details the methodology of the Hillslope Link Model
160 and STREAM and presents the chosen evaluation metrics for streamflow results. Section 4

161 presents HLM simulation results using a range of precipitation inputs. Results and implications of
162 this work are discussed in Section 5.

163 2. Study Area and Data

164 2.1 Study Area and Period

165 The study area comprises the state of Iowa (roughly 97°W to 90°W, 40°N to 44°N), including
166 the major river basins of the Cedar, Iowa, Skunk, and Des Moines rivers that drain into the
167 Mississippi River along the eastern border of the state (Figure 1c). This study area is classified as
168 humid continental on the Koppen climate scale, covers a range of watersheds and sub watersheds,
169 and is comparable to or larger than many study areas used to validate hydrologic performance of
170 SMP data (e.g., Amorim et al., 2020; Maggioni et al., 2013; Nikolopoulos et al., 2010). In the past
171 three decades, the state has experienced several widespread flooding events with substantial social
172 and economic tolls, with the most notable events occurring in 1993, 2008, and 2011 (Lott, 1993;
173 Mutel, 2010; Vennapusa & White, 2015). The study period covers 2008 – 2013, which includes
174 one of the largest flood events that Iowa has experienced in June 2008 (Budikova et al., 2010;
175 Smith et al., 2013).

176 2.2 Precipitation Data

177 The NEXRAD Stage IV radar-gauge product, available over the Continental U.S. (CONUS)
178 at an hourly, 1/24° resolution (Lin, 2011), is used as the ground-reference precipitation product in
179 this work. Although Stage IV is not exempt from errors, it has been considered to exhibit negligible
180 uncertainty relative to SMP data in previous SMP studies (e.g., Aghakouchak et al., 2011), and is
181 considered highly accurate in this study region (Quintero et al., 2020b).

182 NASA's IMERG satellite product is available globally from 2000 to present at a half-hourly,
183 0.1° resolution in three versions: IMERG-Early at a 4-hour latency, available in near real-time but
184 excluding remote sensing data following a satellite pass, IMERG-Late at a 12-hour latency, which
185 incorporates additional remote sensing retrievals, and IMERG-Final at an approximately 2.5
186 month latency, which assimilates gauge data to improve product accuracy (Huffman et al., 2019;
187 Tan et al., 2019). In this work, IMERG-Early is used because of its low latency and availability
188 for early warning systems. IMERG-Early is rescaled to an hourly resolution to match the temporal
189 resolution of Stage IV data during calibration of the SMP error model required for STREAM
190 (Section 3.2).

191 The threshold of detection for all precipitation data in this work is 0.1 mm/hour, consistent
192 with previous sub daily precipitation studies (e.g., Germann & Zawadzki, 2002; Li et al., 2021).

193 2.3 Streamflow Data

194 USGS gauge data throughout the state of Iowa was used to validate the streamflow estimated
195 by the Hillslope Link Model (Section 3.1). A total of 192 USGS stations are used. Periods with
196 missing streamflow data are excluded from analysis. Fourteen gauge sites which are heavily
197 influenced by reservoir operations or other anthropogenic diversions are also excluded from
198 analysis since the hydrologic model described in Section 3.1 does not account for these effects.
199 Roughly 20% of the resulting 178 gauge sites have an upstream area less than 100 km²; 30% of
200 sites have upstream areas between 100 and 1,000 km²; 15% of gauge sites have upstream areas
201 greater than 5,000 km² with the largest upstream area being 32,645 km² at the Des Moines River
202 at Keosauqua (Figure 1d).

203 3. Methods

204 3.1 Hillslope Link Model

205 The Iowa Flood Center’s Hillslope Link Model is a continuous rainfall-runoff model that
206 routes runoff calculated at individual ‘hillslopes’ (Figure 1b) into links which are connected via
207 channel routing (Figure 1a; Krajewski et al., 2017). The HLM was developed initially as part of
208 the Iowa Flood Information System (IFIS; Demir et al., 2018; Demir & Krajewski, 2013;
209 Krajewski et al., 2017) in response to the devastating 2008 flooding in eastern Iowa and began
210 operational use in 2012, providing distributed streamflow predictions and a flood potential index
211 (Quintero et al., 2020a) every 15 minutes at sites across the state (Figure 1c). This study uses the
212 same parameters as in the operational model and does not calibrate the HLM model for specific
213 precipitation datasets. Although the HLM is not a continental or global scale model that utilizes
214 SMP or NWP data, it shares the objective of other operational forecast models to provide
215 expedient, accurate data to prepare communities for potential high flow and flooding events.

216 In the HLM, the landscape is divided into individual channels and hillslopes based on USGS
217 digital elevation model (DEM) data with a 90-meter resolution. Rainfall input is partitioned into
218 soil moisture, drainage, and runoff fluxes. The uncertainty in rainfall input data is viewed as one
219 of the largest uncertainty sources for model predictions (Quintero et al., 2020b).

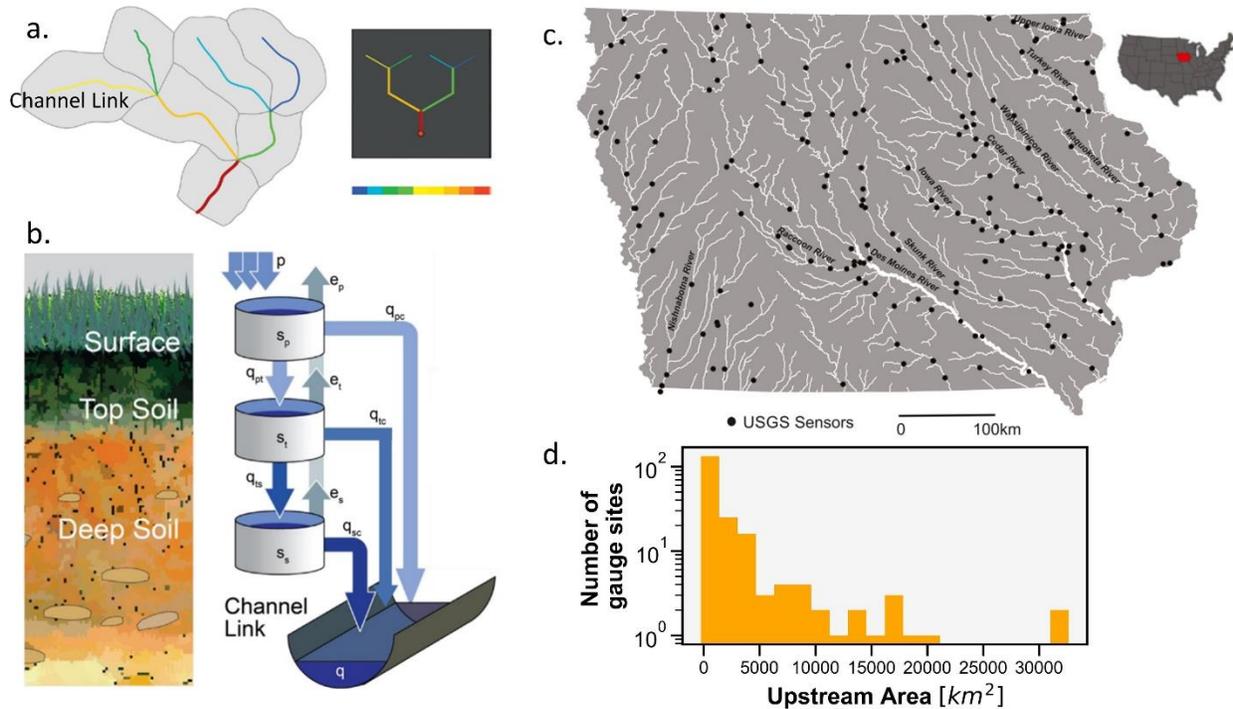


Figure 1. Methodology schematic of the Iowa Flood Center's Hillslope Link Model. (a) A channel routing program combines the (b) runoff calculated from hillslopes for individual channel links. (c) Model streamflow estimates are output and compared to observed streamflow at USGS gauge sites across the state of Iowa. (d) Histogram of upstream area at all sites. Figures from Quintero et al. (2020a) and Quintero et al. (2020b).

220 3.2 Space-Time Rainfall Error and Autocorrelation Model (STREAM)

221 The Space-Time Rainfall Error and Autocorrelation Model (STREAM) was developed to
 222 allow pixel-scale SMP uncertainty estimates to be combined in space and time by simulating the
 223 space-time autocorrelation structure of SMP error using minimal ground reference data (Hartke,
 224 Wright, Li, Maggioni, & Dalia, 2021). The space-time correlation structure of an SMP field – in
 225 this application, IMERG – is adopted as the space-time correlation structure of SMP error at each
 226 time step using the python package pySTEPS and a semi-Lagrangian advection scheme (Pulkkinen
 227 et al., 2019; Figure 2; see Hartke et al., (2022) for further details). IMERG uncertainty at each
 228 pixel and timestep is represented by a distribution that is conditional on an IMERG precipitation
 229 estimate; selecting any value randomly from this conditional distribution produces equally

230 probable values for the true precipitation. To generate a field of possible true precipitation values,
 231 however, requires consideration of the correlation between satellite precipitation errors. Error
 232 autocorrelation describes the intuitive fact that an overestimation of precipitation by IMERG at a
 233 pixel in a given timestep likely corresponds to an overestimation by IMERG in surrounding pixels
 234 and timesteps. By stochastically generating uniform noise fields that replicate the autocorrelation
 235 structure of IMERG fields, a proxy for the unknown autocorrelation structure of the IMERG error
 236 field, possible true precipitation values can be selected from conditional distributions to generate
 237 precipitation fields which incorporate autocorrelated precipitation errors (Figure 2). The Censored

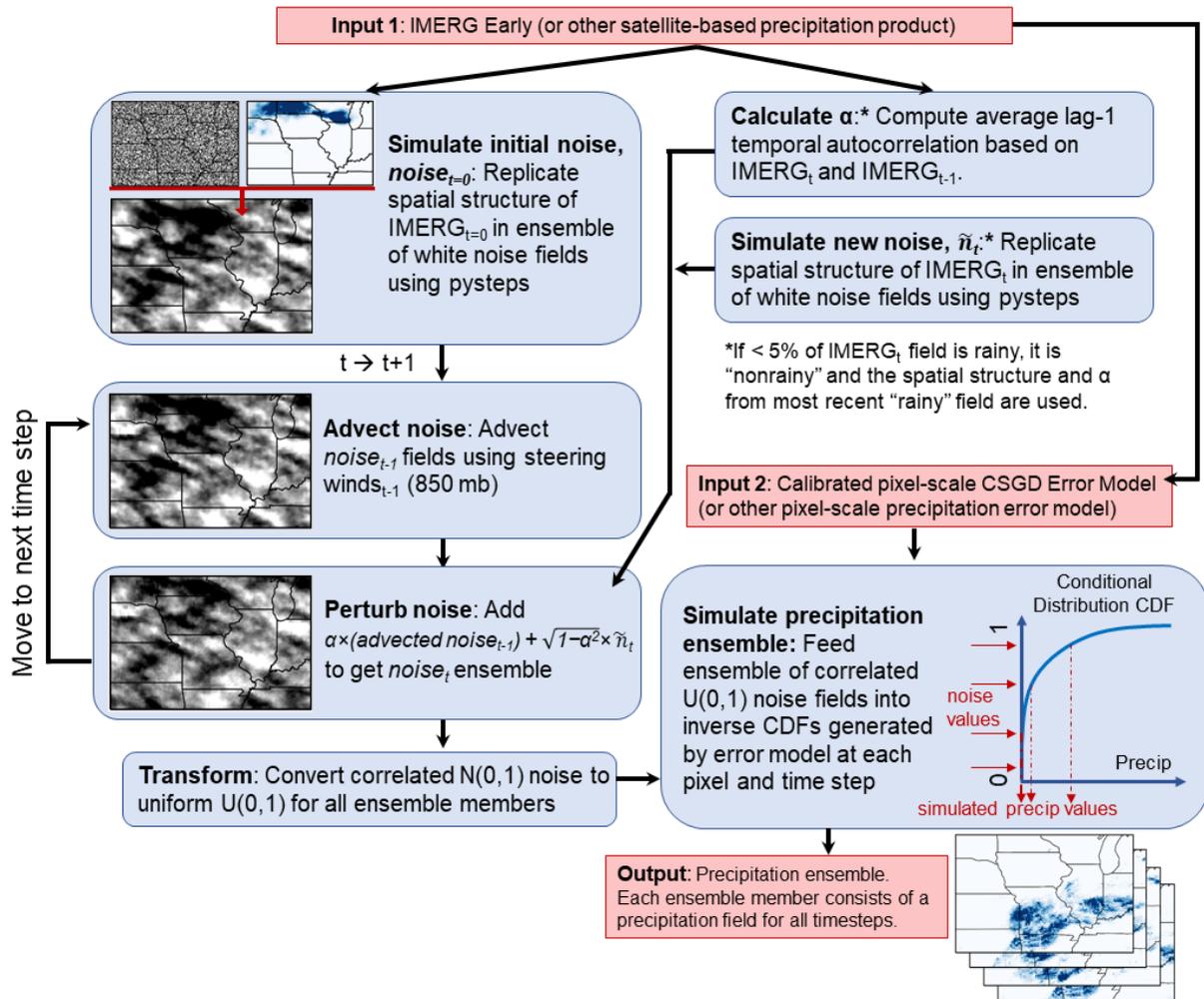


Figure 2. STREAM schematic (Hartke et al. 2022)

238 Shifted Gamma Distribution (CSGD) error model and the conditional distributions it generates are
239 further described in Section 3.3.

240 As in Hartke et al. (2022), the CSGD error model is used here to generate pixel-scale
241 uncertainty distributions for IMERG estimates at every time step and pixel in the study area. For
242 further information on STREAM, see Hartke et al. (2022).

243 3.3 The Censored Shifted Gamma Distribution Error Model and a Bias-corrected 244 IMERG

245 The Censored Shifted Gamma Distribution error model was first introduced by Scheuerer &
246 Hamill (2015) for post-processing of NWP precipitation fields and was adapted by Wright et al.
247 (2017) to characterize the uncertainty of single satellite precipitation estimates across the
248 continental U.S. The CSGD which is used to describe the uncertainty around a precipitation
249 estimate is an adaptation of the two-parameter gamma distribution with an additional “shift”
250 parameter δ that shifts the probability density function (PDF) leftward. The density left of zero
251 represents the probability of zero precipitation, while the density at any value greater than zero
252 represents the likelihood of that amount of precipitation (Figure 3c). The reparameterized
253 distribution is then left-censored at zero, replacing all negative values with zero.

254 A regression model, described by error model parameters α_1 , α_2 , α_3 , and α_4 , is trained based
255 on contemporaneous co-located SMP and ground-truth observations. The calibrated model
256 generates unique “conditional” CSGD parameters $\mu(t)$, $\sigma(t)$, and $\delta(t)$ at any time t as a function of
257 the model parameters and the SMP estimate $R_s(t)$:

$$258 \quad \mu(t) = \frac{\mu_c}{\alpha_1} \log \left\{ 1 + (\exp(\alpha_1) - 1) \left[\alpha_2 + \alpha_3 \frac{R_s(t)}{\bar{R}} \right] \right\} \quad \text{Equation 1,}$$

259
$$\sigma(t) = \alpha_4 \sigma_c \sqrt{\frac{\mu(t)}{\mu_c}} \quad \text{Equation 2,}$$

260
$$\delta(t) = \delta_c \quad \text{Equation 3,}$$

261

262 Where μ_c , σ_c , and δ_c are the parameters of the climatological CSGD, calculated using SMP

263 data (Figure 3a and b), and \bar{R} is the mean of SMP estimates. $\mu(t)$, $\sigma(t)$, and $\delta(t)$ are the parameters

264 of a conditional CSGD describing the uncertainty around SMP estimate $R_s(t)$ (Figure 3c).

265 Equations 1-3 describing the CSGD are written in terms of its mean and standard deviation but

266 can be reparametrized in terms of shape and scale parameters. The error model parameters and

267 climatological CSGD can be trained for a single location using ground reference and SMP

268 timeseries from a single grid cell or can be regionalized by concatenating the timeseries from

269 multiple grid cells and using the resulting timeseries during calibration. In this application of the

270 STREAM (Section 3.2), the CSGD error model is trained regionally over 25 0.1° grid cells at a

271 time (5 grid cells wide and 5 grid cells high; Figure 3a). This approach ensures that sufficient data

272 is used during model training and that differences in IMERG error characteristics are represented

273 across the study region.

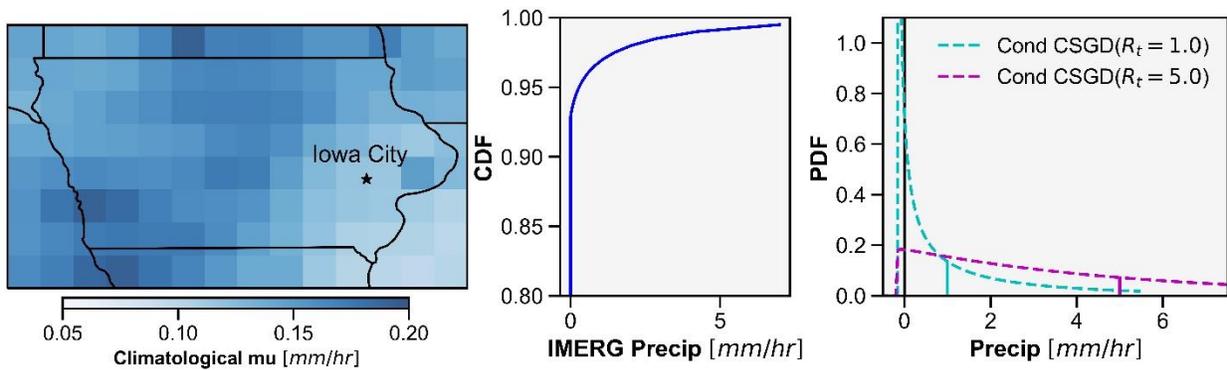


Figure 3. (Left) Climatological μ of 5 by 5 pixel “regional” CGSD error models in study area. (Middle) CDF of climatological CSGD for region over Iowa City. (Right) Conditional CSGDs generated using regional error model calibrated over Iowa City for IMERG estimates of 1 mm/hr (blue) and 5 mm/hr (pink).

274 An additional precipitation dataset is generated by selecting the mean value of the conditional
275 distribution generated for each IMERG estimate in space and time, adjusting the IMERG estimate
276 for systematic bias. This precipitation dataset is effectively a bias-corrected version of IMERG
277 and is referred to as bias-corrected IMERG in the remainder of this work.

278 3.4 Experimental Set Up

279 The HLM is run for April – October in 2008 – 2013 at a 1-hour timestep with the following
280 five precipitation inputs: 0.1° IMERG, 0.1° Stage IV, $1/24^\circ$ Stage IV, 0.1° STREAM ensemble
281 fields (Section 3.2), and a 0.1° bias-corrected IMERG field (Section 3.3). All precipitation inputs
282 produce deterministic streamflow estimates except the STREAM ensemble fields, which produce
283 an ensemble of streamflow estimates. Two resolutions of Stage IV precipitation are used so that
284 the coarser 0.1° product, bilinearly interpolated from the native $1/24^\circ$ resolution product, matches
285 the resolution of IMERG, and any differences between the performance of the two Stage IV
286 products can be attributed to the resolution of the precipitation data. The first month of the HLM
287 simulation in each year (April) is discarded as a spin-up period, and May – October is used for
288 evaluation.

289 3.5 Streamflow Peak Event Identification

290 In order to assess HLM streamflow performance during high flow events, we identify peak
291 events using the wavelet-based event identification method proposed in Towler & McCreight
292 (2021) and Liu et al. (2011). This method assesses the statistical significance of the spectral power
293 “jumps” that large streamflow events represent when the entire streamflow record is transformed
294 into spectral space using a wavelet transform, which creates a representation of the signal in both
295 the time and frequency domain (Figure 4 a-c). We assessed the power spectrum averaged over the

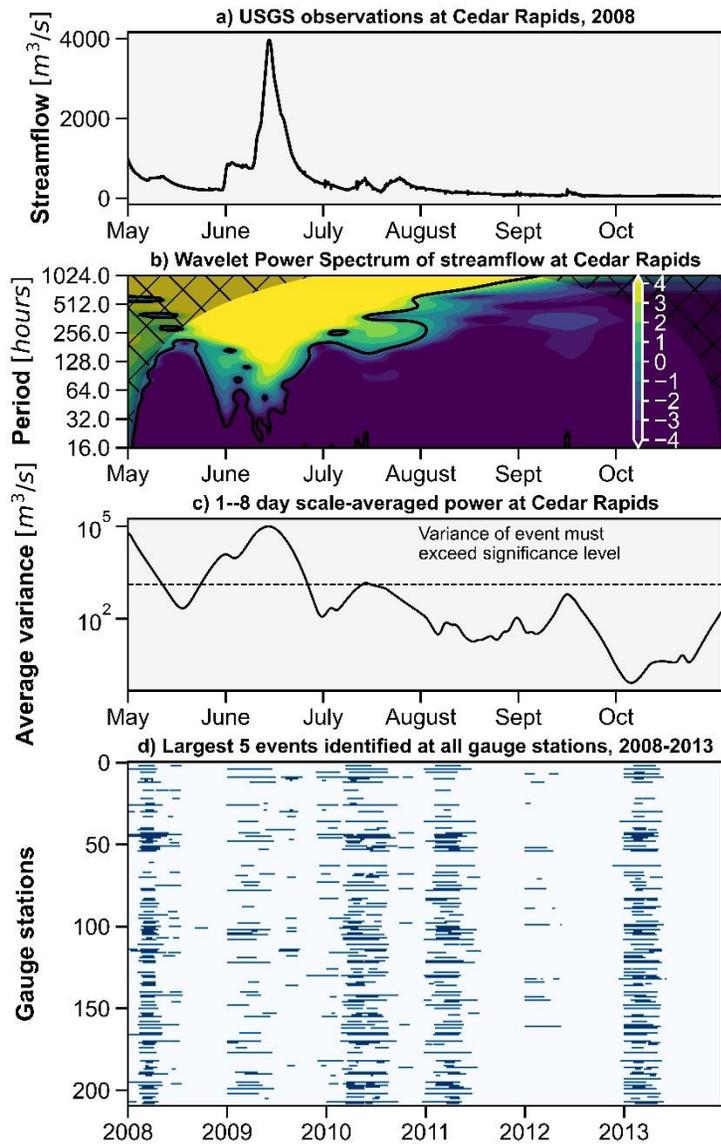


Figure 4. (a) Hydrograph, (b) Wavelet power spectrum of streamflow, and (c) Variance of the power spectrum averaged over 1-8 day (24-192 hour) period for 2008 at the Cedar River at Cedar Rapids. (d) The identified five largest events at each gauge site over 2008-2013 study period.

1-8 day period to find statistically significant events at this scale. Wavelet-based event identification is advantageous because it allows comparison of peak event characteristics (peak flow, flow volume, event duration) without the complication of timing errors.

Intermittent periods with missing data in USGS gauge records are linearly interpolated for the purposes of event identification and calculation of total event volume. After identifying events at each gauge site using observed USGS streamflow (Figure 4d), the peak event flow and total event

volume are calculated for the five largest events, as determined by event volume, using the observed streamflow and the simulated streamflow obtained

316 by forcing HLM with all precipitation datasets. At almost every gauge site, the top five identified
 317 events included events during June 2008 and June 2013. The mean absolute error (MAE) of
 318 IMERG and Stage IV simulations and mean continuous ranked probability score (MCRPS;

319 described in Section 3.6) of STREAM simulations are calculated for peak flow and event volume
 320 at these events.

321 3.6 Performance metrics

322 The normalized root mean squared error (NRMSE), also referred to as the relative RMSE
 323 (Falck et al., 2015), is calculated for IMERG-based and Stage IV-based streamflow as well as for
 324 the mean of the STREAM-based streamflow ensemble.

$$325 \quad NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{Q}_t - Q_t)^2}}{\frac{1}{N} \sum_{t=1}^N Q_t} \quad \text{Equation 4,}$$

326 Where Q_t is the observed streamflow and \hat{Q}_t is the predicted streamflow at timestep t . The
 327 mean absolute error (MAE) for deterministic streamflow estimates and the mean continuous
 328 ranked probability score (MCRPS) for streamflow ensembles are calculated. The continuous
 329 ranked probability score of an ensemble streamflow estimate is analogous to the absolute error of
 330 a single streamflow estimate and reduces to the absolute error as the ensemble size decreases to
 331 one; therefore, the MCRPS of a probabilistic streamflow timeseries is analogous to the MAE of a
 332 deterministic timeseries, allowing for comparison of performance between probabilistic and
 333 deterministic streamflow predictions.

$$334 \quad MAE = \frac{1}{N} \sum_{t=1}^N |\hat{Q}_t - Q_t| \quad \text{Equation 5,}$$

$$335 \quad MCRPS = \frac{1}{N} \sum_{t=1}^N \int_{\mathbb{R}} [F(\hat{Q}_t) - I(\hat{Q}_t \geq Q_t)]^2 d\hat{Q}_t \quad \text{Equation 6,}$$

336 Where $F(\hat{Q}_t)$ is the forecasted distribution of streamflow, in this case an ensemble rather than
337 an explicit distribution, and $I[\bullet]$ is an indicator function. The ideal value for NRMSE, MAE, and
338 MCRPS is 0, and higher values indicate worse performance.

339 The containing ratio (CR) is the proportion of observed streamflow data which lies within the
340 range of the streamflow ensemble generated using STREAM precipitation ensembles. This
341 probabilistic performance metric, and its counterpart, the exceedance ratio (ER), have been used
342 by previous work in the satellite precipitation, hydrologic modeling, and forecast verification
343 communities to assess ensemble performance (e.g. Franz & Hogue, 2011; Hartke et al., 2021;
344 Maggioni et al., 2013; Xiong & O'Connor, 2008)

$$345 \quad CR = \frac{1}{n} \sum_{t=1}^n I[Q_t] \quad \text{Equation 7,}$$

346 Where the indicator function $I[\bullet]$ is equal to one when Q_t is bracketed by the minimum and
347 maximum values of the predicted streamflow ensemble at timestep t and is equal to zero when Q_t
348 lies outside the ensemble. The ideal value for CR is 1, indicating that the ensemble brackets
349 observed streamflow in every instance, while the poorest performance would be indicated by a
350 value of 0.

351 The uncertainty ratio (UR) is the ratio between ensemble spread and the corresponding
352 observed streamflow at each timestep and has been used in previous studies of probabilistic
353 streamflow and soil moisture (Falck et al., 2015; Maggioni et al., 2011).

$$354 \quad UR = \frac{\sum_{t=1}^n (\hat{Q}_{tmax} - \hat{Q}_{tmin})}{\sum_{t=1}^n Q_t} \quad \text{Equation 8,}$$

355 The UR can vary between 0 and infinity, with lower values indicating lower ensemble spread
356 relative to observed streamflow. A high CR and low UR are ideal for ensemble predictions because

357 this indicates that an ensemble is consistently bracketing observations while maintaining a small
358 ensemble spread relative to observed streamflow values. Periods with missing data in USGS gauge
359 records are excluded when calculating the above performance metrics.

360 4. Results

361 4.1 Performance of HLM streamflow estimates

362 Figure 5 shows the containing ratio (CR) of the streamflow ensemble generated using
363 STREAM ensemble precipitation input and the percent reduction in MAE of the IMERG-based
364 streamflow when compared to the MCRPS of the STREAM-based streamflow ensemble. The CR
365 of the STREAM-based streamflow ensemble varies across the study area but is consistently high
366 in watersheds in eastern and western Iowa. Lower CR values in the Des Moines Lobe landform in
367 central Iowa correspond to an area where the HLM configuration has low performance, mainly
368 because the model does not account for changes in infiltration and flow due to agricultural tile
369 drainage systems (Quintero et al., 2020). The MCRPS of the STREAM-based streamflow

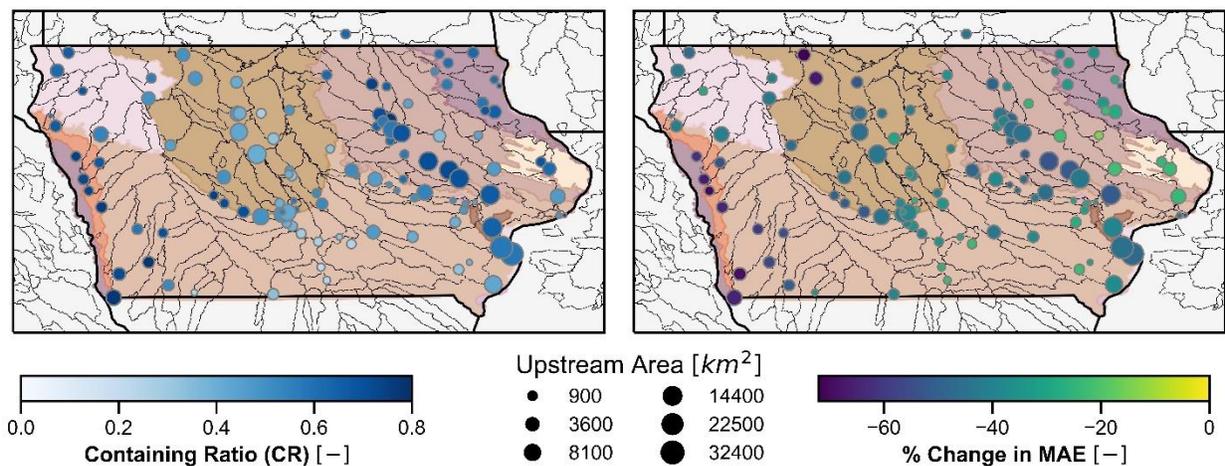


Figure 5. (Left) Map showing containing ratio (CR) of streamflow ensemble generated by HLM using STREAM ensemble as input and (Right) Percent difference between MAE of IMERG-based streamflow and MCRPS of STREAM-based streamflow ensemble. The STREAM-based streamflow ensemble has a MCRPS that is 20 - 60% lower than the MAE of IMERG-based streamflow at most gauge sites in the study area. Different geomorphological landforms across the state are shaded.

370 ensemble is lower than the MAE of IMERG-based streamflow across all watersheds, reducing the
371 MAE by up to 80% in some locations.

372 Figure 6 plots the NRMSE, MAE and MCRPS of IMERG-, Stage IV-, and STREAM-based
373 streamflow estimates. The CR and UR of the STREAM-based ensemble are also presented in
374 panels c and d. Stage IV-based streamflow generally exhibits the lowest error; however, the
375 STREAM-based streamflow exhibits errors nearly as low as that of Stage IV-based streamflow
376 and much lower than that of the IMERG-based simulations. MAE and MCRPS increase with
377 upstream area, which is expected since streamflow magnitudes are greater in bigger basins. The
378 reduction in MCRPS relative to MAE by STREAM-based streamflow estimates increases
379 substantially with basin size; the STREAM-based streamflow ensemble exhibits a MCRPS 50%
380 lower than the MAE of IMERG-based streamflow estimates at gauge sites with upstream areas
381 greater than 10,000 km². Except for basins < 100 km², the CR of STREAM simulations generally
382 increases with upstream area when basins < 100 km² (the size of a single IMERG pixel) are
383 excluded, while the UR of the STREAM ensemble consistently decreases with upstream area. The

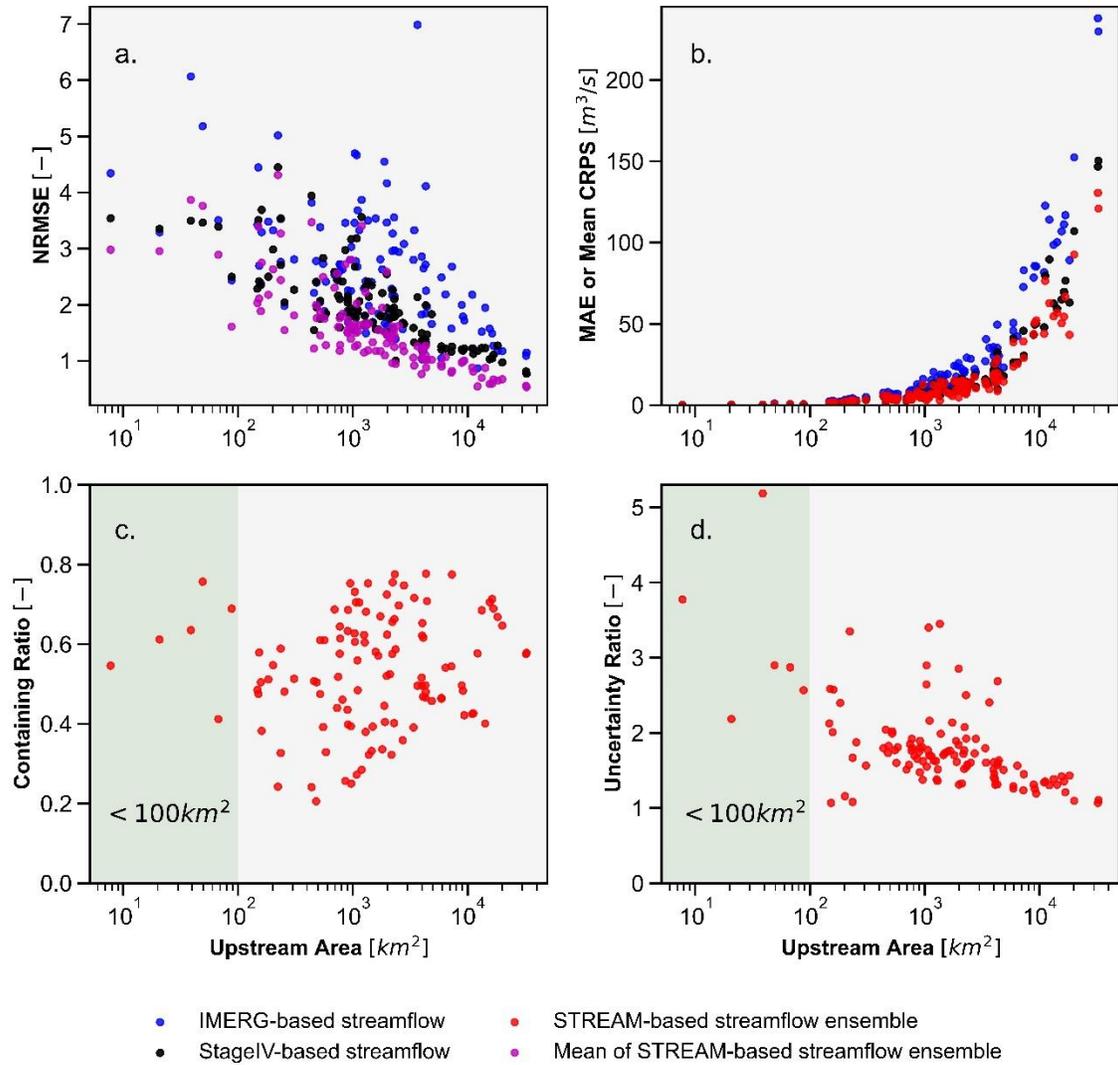


Figure 6. (a) NRMSE and (b) MAE and MCRPS calculated for all streamflow estimates. (c) Containing Ratio, and (d) Uncertainty Ratio calculated for STREAM-based streamflow estimates April-October 2008-2013 at USGS gauge sites across Iowa as a function of upstream area. The green areas in (c) and (d) highlight results from gauges with upstream areas less than 100 km², an area smaller than that of a single IMERG grid cell.

384 large ensemble spreads in small basins reflected by high UR may be due to the limitations of
 385 hydrological models in simulating fast hydrological responses in small basins, when remote
 386 sensing products such as IMERG are too coarse to capture (Nimmo et al., 2021; Széles et al.,
 387 2020).

388 Figure 7 displays the MCRPS and CR of STREAM-based streamflow ensembles as a function
 389 of ensemble size and upstream area. Ensemble members were chosen at random to generate this
 390 plot. MCRPS generally decreases and CR generally increases with increasing ensemble size. As
 391 shown previously in Figure 5, MCRPS increases with upstream area. MCRPS decreases with
 392 ensemble size in basins less than 4500 km² in size until an ensemble size of 10 is reached, after
 393 which the MCRPS appears to stabilize. In basins larger than 4500 km², a larger ensemble appears
 394 to improve MCRPS until an ensemble size of 20 is reached. The CR demonstrates greater change
 395 with ensemble size; for all basin sizes, the CR continuously increases with ensemble size, and
 396 basins less than 100 km² (greater than 2000 km²) in size eventually level out at a CR of
 397 approximately 0.6 (0.55). The upstream area range that consistently has the lowest CR is 100 –
 398 750 km².

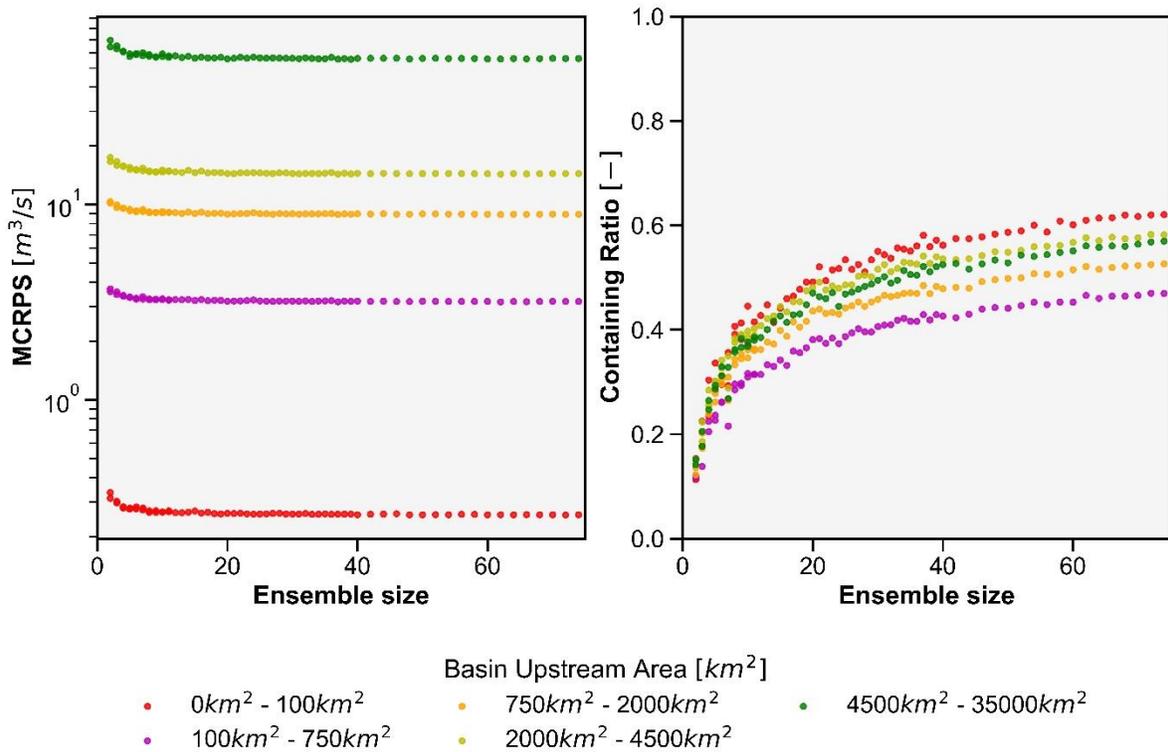


Figure 7. (Left) MCRPS and (Right) Containing Ratio of STREAM-based streamflow ensemble as a function of ensemble size and basin area. The MCRPS decreases with increasing ensemble size and is lowest in basins with the smallest upstream areas. CR increases with ensemble size.

399 4.2 Performance of HLM streamflow estimates during extreme flow events

400 Figure 8 illustrates observed and simulated HLM streamflow during identified peak flow
401 events at three gauges: the Cedar River at Cedar Rapids (6,492 km²), the Iowa River at Lone Tree
402 (4,291 km²), and East Nishnabotna River at Red Oak (897 km²). Visual comparison of observed
403 and simulated streamflow reveals that Stage IV-based estimates generally match observations
404 leading up to and during high flow events; however, even though Stage IV precipitation data is
405 relatively accurate, Stage IV-based streamflow peaks often lag behind observed peaks, indicating
406 a timing error in the hydrologic model. Further, Stage IV-based streamflow does not agree with
407 observed streamflow during the recession period (e.g. late June 2008 at the Cedar River in Figure
408 8). In almost all high flow events illustrated in Figure 8, IMERG-based streamflow overestimates
409 the actual streamflow, sometimes by a considerable margin. This is consistent with findings that
410 IMERG-Early tends to overestimate extreme precipitation in this part of the world (Li et al., 2022).
411 The STREAM simulation mean generally matches Stage IV and observations more than IMERG,
412 and the 90% confidence interval brackets the observed streamflow at almost all timesteps during
413 increasing and peak flows. The STREAM 90% CI is widest during periods when IMERG
414 uncertainty is high and IMERG-based streamflow estimates correspondingly exhibit high error
415 (e.g., the 2010 event in Figure 8). All HLM simulations exhibit a tendency to underestimate
416 streamflow during the recession period following peak events.

417 The MAE of IMERG- and Stage IV-based streamflow and MCRPS of STREAM-based
418 streamflow ensembles when estimating the peak flow, event volume, and event duration calculated
419 for the largest ten events during the study period at each gauge are shown in Figure 9a-c. As with
420 the MAE and MCRPS for estimating streamflow, the MAE and MCRPS for estimated peak flow
421 and event volume increase with upstream basin area. The MCRPS of STREAM-based peak flow

422 estimates is substantially reduced relative to the MAE of IMERG-based peak flow at gauge
 423 locations with upstream areas greater than 1,000 km². The accuracy of event volume predictions
 424 does not improve as much as the accuracy of peak flow when predicted using the STREAM
 425 ensemble instead of IMERG. The absolute error in estimated peak flow, event volume, and event
 426 duration for identified events at 97 gauge stations in June 2008 and 101 gauge stations in June

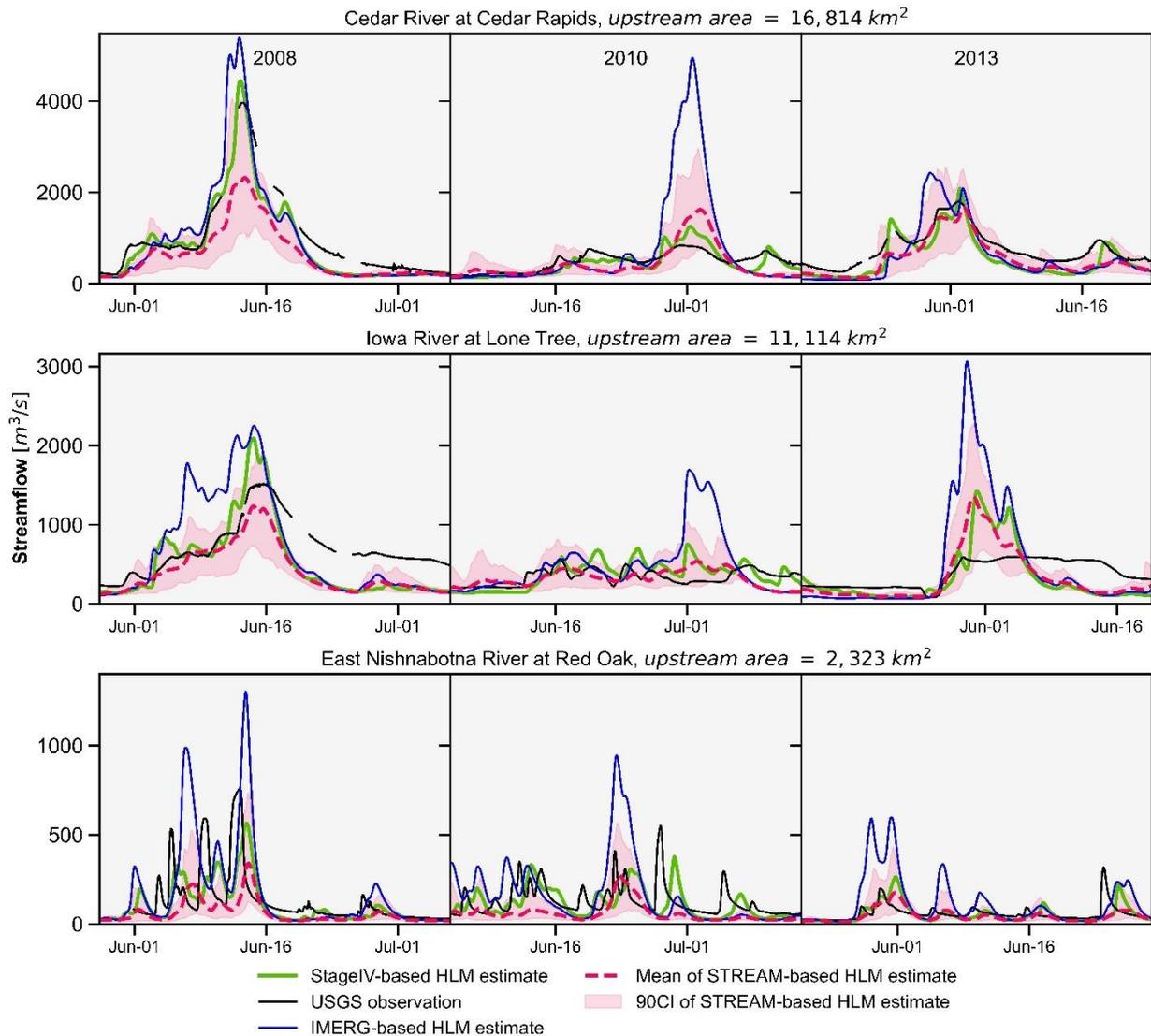


Figure 8. Example of identified peak streamflow events and performance of HLM streamflow estimates at three USGS gauge sites. Observed streamflow from USGS gauges (black) and predicted streamflow is shown for HLM results based on Stage IV (green), IMERG (blue), and the STREAM ensemble (red).

427 2013 are shown in Figure 9d-f and Figure 9g-i, respectively. As found in Quintero et al. (2020b),
 428 errors in peak flow estimation increase with upstream area (Figure 9d, 9g). The absolute error for
 429 STREAM simulations is shown as a 90% confidence interval, representing the range of STREAM-
 430 based estimates for peak flow, event volume, and event duration, which may be interpreted as the
 431 range of uncertainty surrounding these event metrics due to precipitation error. IMERG

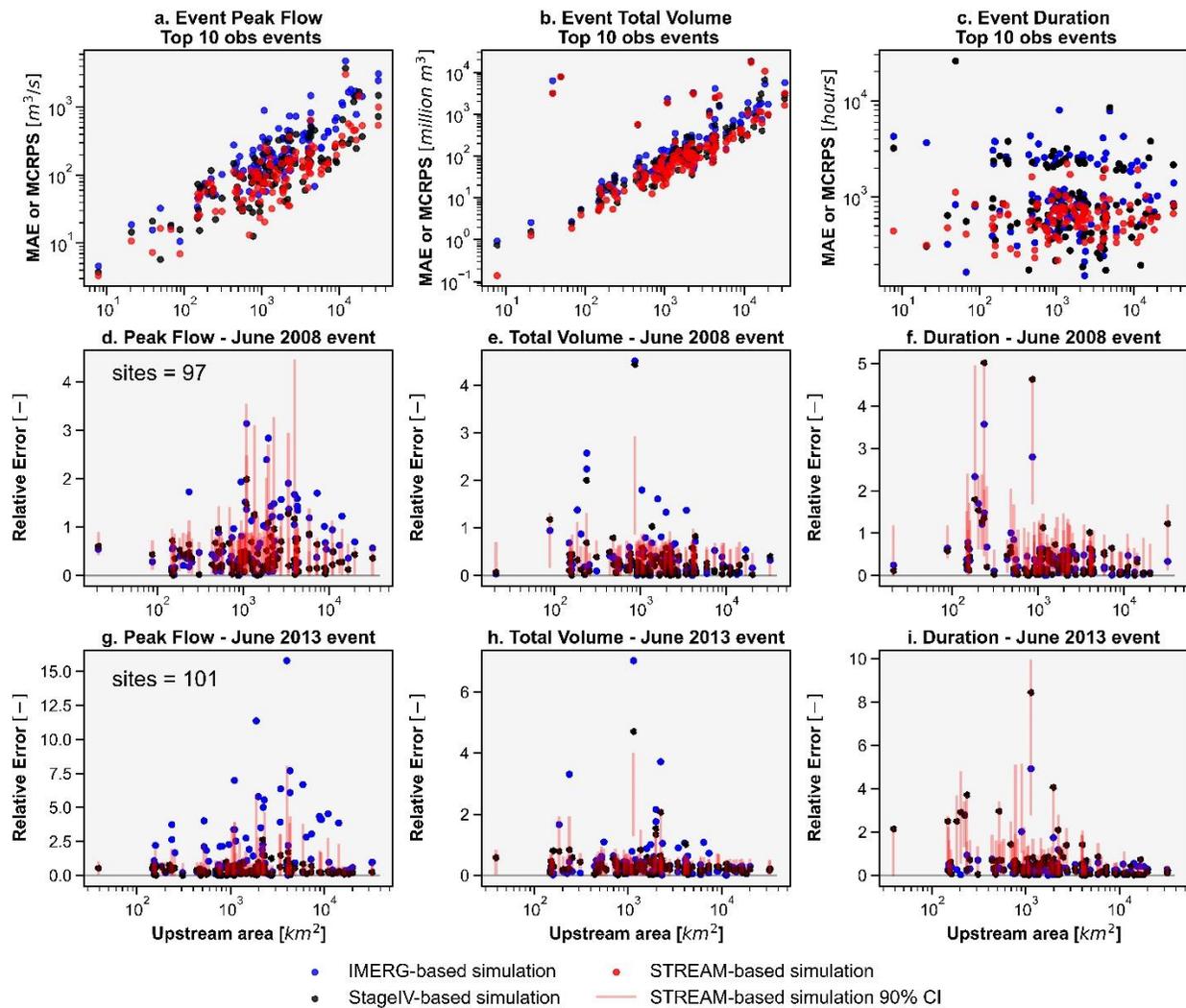


Figure 9. (Top) The MAE of IMERG-based and Stage IV-based estimates and MCRPS of STREAM-based estimates of (a) event maximum streamflow and (b) event total volume for the ten largest events at each gauge site. (Middle) The absolute error of simulated estimates of (d) peak flow, (e) event volume, and (f) event duration during the June 2008 events detected at 97 gauge sites. (Bottom) Absolute error of simulated estimates of (g) peak flow, (h) event volume, and (i) event duration during the June 2013 events detected at 101 gauge sites

432 consistently overestimates peak flow and volume across all sites during the 2008 event while Stage
 433 IV estimates are much closer to observed values. The 90% CI of the STREAM-based simulation
 434 consistently brackets observed peak flow and duration but fails to capture event volume in several
 435 sites with high upstream areas during the 2013 event. The performance of all simulations and range
 436 of uncertainty surrounding event duration in 2008 and 2013 are comparable across basin sizes.

437 4.3 Effect of precipitation uncertainty, precipitation resolution, and bias-correction

438 Figure 10 presents the proportion of streamflow error (here in terms of MAE) that can be
 439 attributed to different precipitation sources, sorted by upstream area. Figure 10a shows the percent
 440 reduction in MAE from IMERG-based streamflow to streamflow forced by more accurate Stage
 441 IV data (at 0.1° resolution and at native resolution). The percentage of streamflow MAE
 442 attributable to IMERG uncertainty increases with upstream area and ranges from 20 to 40% of
 443 MAE on average across basin sizes. In some basins, this percentage is as high as 60. Figure 10b

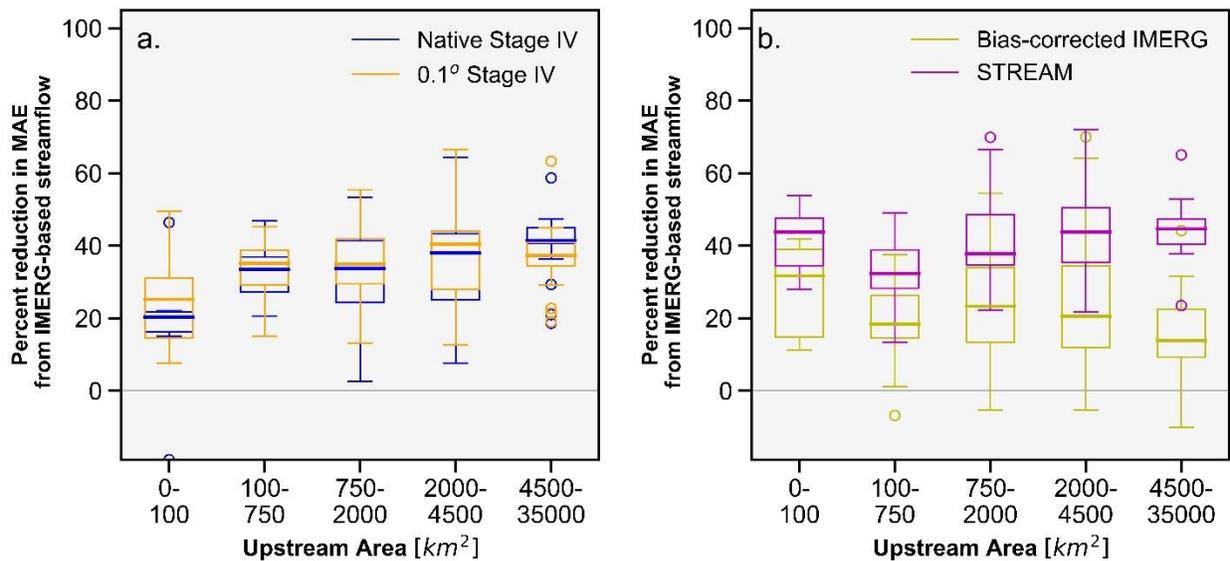


Figure 10. (a) Percent reduction in streamflow MAE from IMERG-based streamflow to native resolution and 0.1° Stage IV-based streamflow. (b) The reduction in streamflow MAE when using bias-corrected IMERG and the STREAM ensemble instead of IMERG as precipitation input to the Hillslope Link Model. The MCRPS is used to calculate the reduction in streamflow MAE for the STREAM ensemble.

444 compares the percent reduction of IMERG-based streamflow MAE when compared to bias-
445 corrected IMERG (yellow; described in Section 3.3) and the STREAM ensemble simulations
446 (magenta). The reduction in streamflow MAE is consistently higher across basin sizes when using
447 the STREAM ensemble instead of bias-corrected IMERG to force the HLM. In some basins, using
448 bias-corrected IMERG to predict streamflow actually increases the MAE

449 of predicted streamflow relative to IMERG. Unlike the STREAM-based streamflow estimates,
450 streamflow based on bias-corrected IMERG appears to perform worse relative to IMERG as the
451 upstream area increases at gauge sites. Although Figure 10b is calculated using all streamflow
452 estimates, Supplemental Figure 2 shows the same analysis for the highest 10% of streamflow
453 observations at each gauge site.

454 5. Discussion

455 5.1 Performance of HLM streamflow estimates overall and during peak flow events

456 Across study results, use of the STREAM ensemble to force the HLM hydrologic model
457 improves streamflow performance to nearly the same degree as use of ground reference data,
458 relative to the satellite multisensor product IMERG. Results that demonstrate greater relative
459 improvement in MAE in large basins when using STREAM ensemble fields to force the HLM
460 (Figures 6,10) correspond to results in Falck et al. (2018), which found greater reduction in random
461 error at large basin scales in Brazil. The notably greater reduction in MCRPS relative to MAE in
462 large basins demonstrates the importance of accounting for precipitation uncertainty in hydrologic
463 models for basins greater than 10,000 km² in size (Figures 5 and 6). While the error in streamflow
464 estimates in small basins may not greatly improve when precipitation uncertainty is accounted for,
465 accounting for precipitation uncertainty in large basins noticeably reduces error in streamflow

466 estimates. That STREAM-based streamflow ensembles exhibit NRMSE and MCRPS nearly as
467 low as that of Stage IV-based streamflow ensembles across all gauge sites overall and during peak
468 events (Figures 5 and 9) demonstrates that STREAM effectively captures the range of true
469 precipitation and is effective at substantially reducing the effects of precipitation uncertainty on
470 HLM output. Notably, this is contingent on both the ability of STREAM to simulate the space-
471 time autocorrelation structure of errors as well as the ability of the CSGD error model to provide
472 apt uncertainty characterizations for every IMERG estimate. Improved streamflow prediction
473 using the STREAM ensemble supports previous work demonstrating the advantages of ensemble-
474 based, probabilistic flood forecasting (e.g., Cloke & Pappenberger, 2009; Verbunt et al., 2007).

475 The STREAM ensemble outperforms bias-corrected IMERG because firstly, the bias-
476 corrected product cannot simulate extreme rain rates while it is correcting for the systematic
477 tendency of IMERG to overestimate moderate to high precipitation events and, furthermore, it
478 cannot simulate high rain rates over multiple pixels (e.g., autocorrelated errors) in a watershed,
479 which leads to heavy underestimation of streamflow in instances when high rain rates are, in fact,
480 occurring across a watershed.

481 It is also interesting that IMERG-based streamflow does not always fall within the STREAM
482 ensemble range, especially during peak flows (see Figure 8, Supplemental Figure 1). This occurs
483 when IMERG overestimates precipitation over multiple pixels and the STREAM methodology
484 generates precipitation fields that may contain some high rain rates, or a similar area of
485 precipitation, but that overall reduce the amount of IMERG-estimated precipitation across a
486 region. Essentially, during process of correcting for systematic bias, simulating the range of
487 random error, and simulating error autocorrelation structures, STREAM recognizes that such a
488 large area of extreme rainfall observed by IMERG is statistically very unlikely to correspond to

489 such extreme ‘true’ rainfall; thus, no ensemble members are generated that predict this high
490 regional precipitation as the possible true rainfall. The resulting streamflow forced by the
491 STREAM ensemble does not bracket the IMERG-based simulation, but it is clear that STREAM
492 was correct in generally reducing precipitation rates and the STREAM ensemble replicates the
493 Stage IV simulation and observations more than IMERG (Figure 8, Supplemental Figure 1).

494 The decreasing UR and increasing CR of the STREAM-based streamflow ensemble with
495 upstream area may be due to smaller relative uncertainties (UC) in larger basins (as in Falck et al.,
496 2015; Maggioni et al., 2013). These UR and CR trends also demonstrate that the higher CR in
497 larger basins is not due to a significantly larger ensemble spread relative to the average streamflow.
498 These results are consistent with previous findings by Falck et al. (2015), which also found UR to
499 decrease with increasing upstream area in Brazil. Additionally, it could be expected that basins
500 with upstream areas $< 100 \text{ km}^2$ perform differently than larger basins (e.g., Figure 6c) because the
501 runoff in these basins is based on a single IMERG estimate and does not need to account for the
502 autocorrelation of errors in multiple IMERG estimates across a watershed. Decreasing
503 performance of bias-corrected IMERG predictions with upstream area (Figure 9b, Supplemental
504 Figure 2b) indicates that basins in this study area are not effectively filtering out precipitation
505 random error. Otherwise, larger basin areas should correspond to better performance of streamflow
506 predictions generated by IMERG and bias-corrected IMERG. That STREAM-based predictions
507 meanwhile improve in performance with upstream area (Figure 9b, Supplemental Figure 2b)
508 indicates that (1) autocorrelation of random errors has a substantial impact on the propagation of
509 precipitation error in the HLM and (2) STREAM effectively simulates such autocorrelation
510 structures.

511 The containing ratio and MCRPS show consistent improvement with increasing ensemble size
512 until approximately an ensemble size of approximately 30 (Figure 7). In small basins, fewer
513 ensemble members are required, which is intuitive given the limited variety of potential true
514 precipitation fields that an ensemble generates for basins covered by only a handful of IMERG
515 pixels. However, in large basins, where there are more possible combinations of precipitation error
516 and error autocorrelation structures, more realizations of the possible true precipitation fields are
517 required to capture observed streamflow.

518 The Hillslope Link Model was designed principally to monitor high streamflow conditions that
519 may endanger communities and infrastructure in Iowa. It is therefore not unexpected that its
520 streamflow estimates during low flow conditions do not consistently match USGS observations
521 (see Figure 8; Quintero, et al., 2020). Given the inaccuracy in HLM streamflow estimates during
522 low flow conditions or recessions after peak events even when the high-quality Stage IV product
523 is used (e.g. Figure 8) provides reasonable grounds to conclude that poor streamflow accuracy at
524 low flows is due to structural or parameter error within HLM rather than problems with the
525 precipitation. Visual inspection of hydrographs also reveals timing errors in predicted peak
526 streamflow events that can be attributed to either the HLM or insufficient temporal resolution of
527 precipitation forcing data, but not to precipitation uncertainty since even the high accuracy ground
528 reference precipitation dataset fails to generate correct timing of peak streamflow in some
529 instances (Figure 8, bottom left panel).

530 5.2 Effect of precipitation uncertainty, precipitation resolution, and bias-correction

531 A sizeable percentage—between 20 and 40% on average across basin sizes—of the error in
532 predicted streamflow using IMERG data can be attributed to errors in this precipitation data

533 product (Figure 10a). The minimal difference in reduction in predicted streamflow MAE when
534 using 1/24° and 0.1° resolution Stage IV demonstrates that uncertainty in precipitation data makes
535 a larger contribution towards the error in simulated streamflow than does the resolution of the
536 precipitation data. Coarser resolution Stage IV improves predicted streamflow accuracy slightly
537 more than native resolution Stage IV, except in the largest basins (Figure 10a). When only the top
538 10% of streamflow observations are considered, however, native resolution Stage IV outperforms
539 the coarser resolution product in all basin size categories except those less than 100 km²
540 (Supplemental Figure 2). The ability of coarser resolution ground reference data to produce
541 streamflow estimates with a slightly lower MAE than 1/24° resolution ground reference data is
542 most likely due to sampling error; several of the USGS gauge sites for basins in this study only
543 have available data for two or three of the years in the six years study period.

544 Correcting for systematic bias does not account for random error, which is a substantial
545 component of overall IMERG and other SMP product uncertainty (Figure 10b). During all flows
546 and particularly during the highest subset of streamflow across gauge sites, streamflow error is
547 consistently reduced more when the STREAM ensemble is used as input to the HLM rather than
548 the bias-corrected IMERG dataset. This demonstrates the inadequacy of bias correction as an
549 approach to addressing precipitation uncertainty in hydrologic models like the HLM and supports
550 similar findings in (Habib et al., 2014) and (REFERENCE) Hartke et al (2020) regarding SMP
551 uncertainty in runoff simulations and landslide hazard applications, respectively.

552 6. Summary and Conclusions

553 Although satellite multisensor precipitation products and numerical weather model fields have
554 made possible near real-time hydrologic modeling on a continental to global scale in recent

555 decades, streamflow estimates are hampered by the uncertainty in precipitation data from these
556 sources. Overestimation, underestimation, and incorrect detection of precipitation all lead to less
557 accurate estimates of runoff and streamflow in hydrologic models. Although hydrologic model
558 calibration may reduce the propagation of forcing data uncertainty, it is no substitute for explicitly
559 representing the uncertainty in forcing precipitation data, which includes substantial random error
560 in addition to systematic bias. However, this effort is complicated by the need to simulate the
561 autocorrelation structure of precipitation errors in order to capture basin-scale precipitation
562 uncertainty, while minimizing reliance on ground references and representing precipitation
563 uncertainty in a way that hydrologic models or other applications with deterministic inputs can
564 readily ingest.

565 In this paper, precipitation ensembles generated by the Space-Time Rainfall Error and
566 Autocorrelation Model (STREAM) are used to force the Hillslope Link Model and simulate
567 streamflow ensembles which reflect the uncertainty in precipitation forcing from IMERG. Using
568 STREAM ensembles substantially reduces the error in streamflow estimates overall and during
569 peak flow events. Results show that a substantial portion of the error between predicted and
570 observed streamflow—as much as 60%—is due to the error in IMERG. Correcting for systematic
571 bias while neglecting the range of random error is shown to be an insufficient approach to
572 addressing precipitation uncertainty and can actually worsen streamflow predictions. Results also
573 reemphasize the results of Cunha et al. (2012)—that the spatial autocorrelation of precipitation
574 errors heavily impacts the ability of basins to filter out precipitation uncertainty. Although
575 STREAM is applied to SMP data in this work and in Hartke et al. (2022), it is also applicable to
576 numerical weather prediction (NWP) model fields.

577 Although the error characterization in this work utilizes a ground reference product for
578 calibration, recent work by Li et al. (2021) demonstrates a method for characterizing satellite
579 precipitation uncertainty using dual-precipitation radar data; this will allow STREAM to generate
580 precipitation ensembles to represent satellite precipitation uncertainty anywhere in the world
581 without ground reference data. Future work will use a similar methodology as demonstrated in this
582 work to account for satellite precipitation uncertainty in hydrologic modeling on a continental to
583 global scale without reliance on any ground-based precipitation data.

584 7. Acknowledgements

585 S. H. Hartke’s contributions were funded by the Grainger Wisconsin Distinguished Graduate
586 Fellowship program. D. B. Wright’s contributions were funded by NASA Precipitation
587 Measurement Mission Grant 80NSSC19K0951. We thank the Iowa Flood Center at the University
588 of Iowa for the use of the Hillslope Link Model and associated computing resources. We thank
589 Dr. Zhe Li for his matplotlib formatting template.

590 References

- 591 Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of
592 satellite-retrieved extreme precipitation rates across the central United States. *Journal of*
593 *Geophysical Research Atmospheres*, *116*(2), 1–11. <https://doi.org/10.1029/2010JD014741>
- 594 Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F.
595 (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology*
596 *and Earth System Sciences*, *17*(3), 1161–1175. <https://doi.org/10.5194/hess-17-1161-2013>
- 597 Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The
598 impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid
599 catchment. *Journal of Hydrology*, *519*(PD), 2763–2774.
600 <https://doi.org/10.1016/j.jhydrol.2014.07.041>
- 601 Amorim, J. da S., Viola, M. R., Junqueira, R., de Oliveira, V. A., & de Mello, C. R. (2020).
602 Evaluation of satellite precipitation products for hydrological modeling in the brazilian
603 cerrado biome. *Water (Switzerland)*, *12*(9). <https://doi.org/10.3390/W12092571>
- 604 Anjum, M. N., Ding, Y., Shangguan, D., Ahmad, I., Wajid Ijaz, M., Farid, H. U., Yagoub, Y. E.,
605 Zaman, M., & Adnan, M. (2018). *Performance evaluation of latest integrated multi-satellite*
606 *retrievals for Global Precipitation Measurement (IMERG) over the northern highlands of*
607 *Pakistan*. <https://doi.org/10.1016/j.atmosres.2018.02.010>
- 608 Asong, Z. E., Razavi, S., Wheeler, H. S., & Wong, J. S. (2017). Evaluation of integrated
609 multisatellite retrievals for GPM (IMERG) over Southern Canada against ground
610 precipitation observations: A preliminary assessment. *Journal of Hydrometeorology*, *18*(4),
611 1033–1050. <https://doi.org/10.1175/JHM-D-16-0187.1>
- 612 Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., & Rost, S. (2009). Effects
613 of precipitation uncertainty on discharge calculations for main river basins. *Journal of*
614 *Hydrometeorology*, *10*(4), 1011–1025. <https://doi.org/10.1175/2008JHM1067.1>
- 615 Budikova, D., Coleman, J. S. M. M., Strobe, S. A., & Austin, A. (2010). Hydroclimatology of the

- 616 2008 Midwest floods. *Water Resources Research*, 46(12).
617 <https://doi.org/10.1029/2010WR009206>
- 618 Charles, S. P., Chiew, F. H. S., Potter, N. J., Zheng, H., Fu, G., & Zhang, L. (2020). Impact of
619 downscaled rainfall biases on projected runoff changes. *Hydrology and Earth System
620 Sciences*, 24(6), 2981–2997. <https://doi.org/10.5194/hess-24-2981-2020>
- 621 Ciach, G. J., Krajewski, W. F., & Villarini, G. (2007). Product-Error-Driven Uncertainty Model
622 for Probabilistic Quantitative Precipitation Estimation with NEXRAD Data. *Journal of
623 Hydrometeorology*, 8(6), 1325–1347. <https://doi.org/10.1175/2007jhm814.1>
- 624 Ciupak, M., Ozga-Zielinski, B., Adamowski, J., Deo, R. C., & Kochanek, K. (2019). Correcting
625 satellite precipitation data and assimilating satellite-derived soil moisture data to generate
626 ensemble hydrological forecasts within the HBV rainfall-runoff model. *Water (Switzerland)*,
627 11(10). <https://doi.org/10.3390/w11102138>
- 628 Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. In *Journal of
629 Hydrology* (Vol. 375, Issues 3–4, pp. 613–626).
630 <https://doi.org/10.1016/j.jhydrol.2009.06.005>
- 631 Cunha, L. K., Mandapaka, P. V., Krajewski, W. F., Mantilla, R., & Bradley, A. A. (2012). Impact
632 of radar-rainfall error structure on estimated flood magnitude across scales: An investigation
633 based on a parsimonious distributed hydrological model. *Water Resources Research*, 48(10).
634 <https://doi.org/10.1029/2012WR012138>
- 635 Demir, I., & Krajewski, W. F. (2013). Towards an integrated Flood Information System:
636 Centralized data access, analysis, and visualization. *Environmental Modelling and Software*,
637 50, 77–84. <https://doi.org/10.1016/j.envsoft.2013.08.009>
- 638 Demir, I., Yildirim, E., Sermet, Y., Muhammed, & Sit, A., & Sit, M. A. (2018). FLOODSS: Iowa
639 flood information system as a generalized flood cyberinfrastructure. *International Journal of
640 River Basin Management*, 16(3), 393–400. <https://doi.org/10.1080/15715124.2017.1411927>
- 641 Falck, A. S., Maggioni, V., Tomasella, J., Diniz, F. L. R., Mei, Y., Beneti, C. A., Herdies, D. L.,
642 Neundorff, R., Caram, R. O., & Rodriguez, D. A. (2018). Improving the use of ground-based

- 643 radar rainfall data for monitoring and predicting floods in the Iguazu river basin. *Journal of*
644 *Hydrology*, 567, 626–636. <https://doi.org/10.1016/j.jhydrol.2018.10.046>
- 645 Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A., & Diniz, F. L. R. (2015). Propagation of
646 satellite precipitation uncertainties through a distributed hydrologic model: A case study in
647 the Tocantins-Araguaia basin in Brazil. *Journal of Hydrology*, 527, 943–957.
648 <https://doi.org/10.1016/j.jhydrol.2015.05.042>
- 649 Falck, A. S., Tomasella, J., Diniz, F. L. R., & Maggioni, V. (2021). Applying a precipitation error
650 model to numerical weather predictions for probabilistic flood forecasts. *Journal of*
651 *Hydrology*, 598, 126374. <https://doi.org/10.1016/j.jhydrol.2021.126374>
- 652 Fekete, B. M., Vörösmarty, C. J., Roads, J. O., & Willmott, C. J. (2004). Uncertainties in
653 precipitation and their impacts on runoff estimates. *Journal of Climate*, 17(2), 294–304.
654 [https://doi.org/10.1175/1520-0442\(2004\)017<0294:UIPATI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2)
- 655 Franz, K. J., & Hogue, T. S. (2011). Evaluating uncertainty estimates in hydrologic models:
656 Borrowing measures from the forecast verification community. *Hydrology and Earth System*
657 *Sciences*, 15(11), 3367–3382. <https://doi.org/10.5194/hess-15-3367-2011>
- 658 Germann, U., & Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from
659 continental radar images. Part I: Description of the methodology. *Monthly Weather Review*,
660 130(12), 2859–2873. [https://doi.org/10.1175/1520-0493\(2002\)130<2859:SDOTPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2)
- 662 Gilewski, P. P., & Nawalany, M. (2018). Inter-comparison of Rain-Gauge, Radar, and Satellite
663 (IMERG GPM) precipitation estimates performance for rainfall-runoff modeling in a
664 mountainous catchment in Poland. *Water (Switzerland)*, 10(11).
665 <https://doi.org/10.3390/w10111665>
- 666 Habib, E., Haile, A. T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of bias correction of
667 satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. *Remote*
668 *Sensing*, 6(7), 6688–6708. <https://doi.org/10.3390/rs6076688>
- 669 Hartke, S. H., Wright, D. B., Kirschbaum, D. B., Stanley, T. A., & Li, Z. (2020). Incorporation of

670 satellite precipitation uncertainty in a landslide hazard nowcasting system. *Journal of*
671 *Hydrometeorology*, 21(8), 1741–1759. <https://doi.org/10.1175/JHM-D-19-0295.1>

672 Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., & Dalia, B. (2021). Ensemble Representation
673 of Satellite Precipitation Uncertainty using an Uncalibrated , Nonstationary , Anisotropic
674 Autocorrelation Model. *Water Resources Research*. <https://doi.org/10.1029/2021WR031650>

675 Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., & Dalia, B. (2022). Ensemble Representation
676 of Satellite Precipitation Uncertainty using a Nonstationary, Anisotropic Autocorrelation
677 Model. *Water Resources Research*, 58(8). <https://doi.org/10.1029/2021WR031650>

678 Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D. B., & Khan, S. (2021).
679 Ensemble Representation of Satellite Precipitation Uncertainty using an Uncalibrated,
680 Nonstationary, Anisotropic Autocorrelation Model. *Earth and Space Science Open Archive*.
681 <https://doi.org/10.1002/essoar.10508893.1>

682 Hong, Y., Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of
683 satellite precipitation estimation and Monte Carlo assessment of the error propagation into
684 hydrologic response. *Water Resources Research*, 42(7), 1–15.
685 <https://doi.org/10.1029/2005WR004398>

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

689 Huffman, G., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian,
690 S., Tan, J., & Xie, P. (2019). NASA Global Precipitation Measurement (GPM) Integrated
691 Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation
692 Measurement (GPM) National Aeronautics and Space Administration (NASA). In *Algorithm*
693 *Theoretical Basis Document (ATBD) Version 06* (Issue March).
694 https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg

695 Ji, X., Li, Y., Luo, X., He, D., Guo, R., Wang, J., Bai, Y., Yue, C., & Liu, C. (2020). Evaluation
696 of bias correction methods for APHRODITE data to improve hydrologic simulation in a large

- 697 Himalayan basin. *Atmospheric Research*, 242(February), 104964.
698 <https://doi.org/10.1016/j.atmosres.2020.104964>
- 699 Jiang, L., & Bauer-Gottwein, P. (2019). How do GPM IMERG precipitation estimates perform as
700 hydrological model forcing? Evaluation for 300 catchments across Mainland China. *Journal*
701 *of Hydrology*, 572(March), 486–500. <https://doi.org/10.1016/j.jhydrol.2019.03.042>
- 702 Kirstetter, P. E., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate
703 estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological*
704 *Society*, 144(December 2017), 191–205. <https://doi.org/10.1002/qj.3243>
- 705 Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilllla, R.,
706 Niemeier, J., Quintero, F., Seo, B. C., Smallll, S. J., Weber, L. J., & Young, N. C. (2017).
707 Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the*
708 *American Meteorological Society*, 98(3), 539–554. [https://doi.org/10.1175/BAMS-D-15-](https://doi.org/10.1175/BAMS-D-15-00243.1)
709 [00243.1](https://doi.org/10.1175/BAMS-D-15-00243.1)
- 710 Li, N., Tang, G., Zhao, P., Hong, Y., Gou, Y., & Yang, K. (2016). *Statistical assessment and*
711 *hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang*
712 *River basin*. <https://doi.org/10.1016/j.atmosres.2016.07.020>
- 713 Li, Z., Tang, G., Kirstetter, P., Gao, S., Li, J.-L. L. F., Wen, Y., & Hong, Y. (2022). Evaluation of
714 GPM IMERG and its constellations in extreme events over the conterminous united states.
715 *Journal of Hydrology*, 606, 127357. <https://doi.org/10.1016/j.jhydrol.2021.127357>
- 716 Li, Z., Wright, D., Hartke, S., Kirschbaum, D., Khan, S., Maggioni, V., & Kirstetter, P.-E. (2021).
717 Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite
718 Multisensor Precipitation Products based on GPM DPR. *Earth and Space Science Open*
719 *Archive*. <https://doi.org/10.1002/ESSOAR.10507263.1>
- 720 Lin, Y. (2011). *GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage*
721 *IV Data. Version 1.0*. <https://data.eol.ucar.edu/dataset/21.093>
- 722 Liu, Y., Brown, J., Demargne, J., & Seo, D. J. (2011). A wavelet-based approach to assessing
723 timing errors in hydrologic predictions. *Journal of Hydrology*, 397(3–4), 210–224.

- 724 <https://doi.org/10.1016/j.jhydrol.2010.11.040>
- 725 Lott, N. (1993). *Research Customer Service Group Technical Report 93-04 the Summer of 1993 :*
726 *Flooding in the Midwest and Drought in the Southeast.* 93(4).
- 727 Lowrey, M. R. K., & Yang, Z. L. (2008). Assessing the capability of a regional-scale weather
728 model to simulate extreme precipitation patterns and flooding in central Texas. *Weather and*
729 *Forecasting*, 23(6), 1102–1126. <https://doi.org/10.1175/2008WAF2006082.1>
- 730 Luitel, B., Villarini, G., & Vecchi, G. A. (2018). Verification of the skill of numerical weather
731 prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *Journal of*
732 *Hydrology*, 556, 1026–1037. <https://doi.org/10.1016/j.jhydrol.2016.09.019>
- 733 Maggioni, V., Reichle, R. H., & Anagnostou, E. N. (2011). The effect of satellite rainfall error
734 modeling on soil moisture prediction uncertainty. *Journal of Hydrometeorology*, 12(3), 413–
735 428. <https://doi.org/10.1175/2011JHM1355.1>
- 736 Maggioni, V., Sapiano, M. R. P., Adler, R. F., Tian, Y., & Huffman, G. J. (2014). An Error Model
737 for Uncertainty Quantification in High-Time-Resolution Precipitation Products. *Journal of*
738 *Hydrometeorology*, 15(3), 1274–1292. <https://doi.org/10.1175/jhm-d-13-0112.1>
- 739 Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., & Stampoulis, D.
740 (2013). Investigating the Applicability of Error Correction Ensembles of Satellite Rainfall
741 Products in River Flow Simulations. *Journal of Hydrometeorology*, 14(4), 1194–1211.
742 <https://doi.org/10.1175/jhm-d-12-074.1>
- 743 Marc, O., Jucá Oliveira, R. A., Gosset, M., Emberson, R., & Malet, J.-P. (2022). Global assessment
744 of the capability of satellite precipitation products to retrieve landslide-triggering extreme
745 rainfall events. *Earth Interactions*, 1–42. <https://doi.org/10.1175/ei-d-21-0022.1>
- 746 Moosavi, A., Rao, V., & Sandu, A. (2021). Machine learning based algorithms for uncertainty
747 quantification in numerical weather prediction models. *Journal of Computational Science*,
748 50, 101295. <https://doi.org/10.1016/j.jocs.2020.101295>
- 749 Mutel, C. F. (2010). *A watershed year: anatomy of the Iowa floods of 2008*. University of Iowa

750 Press.

751 Nasrollahi, N., Aghakouchak, A., Li, J., Gao, X., Hsu, K., & Sorooshian, S. (2012). Assessing the
752 impacts of different WRF precipitation physics in hurricane simulations. *Weather and*
753 *Forecasting*, 27(4), 1003–1016. <https://doi.org/10.1175/WAF-D-10-05000.1>

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

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

762 Nimmo, J. R., Perkins, K. S., Plampin, M. R., Walvoord, M. A., Ebel, B. A., & Mirus, B. B. (2021).
763 Rapid-Response Unsaturated Zone Hydrology: Small-Scale Data, Small-Scale Theory, Big
764 Problems. *Frontiers in Earth Science*, 9. <https://doi.org/10.3389/feart.2021.613564>

765 Nogueira, M. (2020). Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last
766 40 years: Process-based analysis of systematic and random differences. *Journal of Hydrology*,
767 583. <https://doi.org/10.1016/j.jhydrol.2020.124632>

768 Omranian, E., Sharif, H. O., & Tavakoly, A. A. (2018). How well can Global Precipitation
769 Measurement (GPM) capture hurricanes? Case study: Hurricane harvey. *Remote Sensing*,
770 10(7). <https://doi.org/10.3390/rs10071150>

771 Pradhan, A., & Indu, J. (2021). Assessment of SM2RAIN derived and IMERG based precipitation
772 products for hydrological simulation. *Journal of Hydrology*, 603(PD), 127191.
773 <https://doi.org/10.1016/j.jhydrol.2021.127191>

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

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

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

781 Quintero, F., Krajewski, W. F., & Rojas, M. (2020). A flood potential index for effective
782 communication of streamflow forecasts at ungauged communities. *Journal of*
783 *Hydrometeorology*, 21(4), 807–814. <https://doi.org/10.1175/JHM-D-19-0212.1>

784 Quintero, F., Krajewski, W. F., Seo, B. C., & Mantilla, R. (2020). Improvement and evaluation of
785 the Iowa Flood Center Hillslope Link Model (HLM) by calibration-free approach. *Journal of*
786 *Hydrology*, 584(November 2019), 124686. <https://doi.org/10.1016/j.jhydrol.2020.124686>

787 Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error
788 characteristics on physically based snow simulations within a global sensitivity analysis
789 framework. *Hydrology and Earth System Sciences*, 19(7), 3153–3179.
790 <https://doi.org/10.5194/hess-19-3153-2015>

791 Scheuerer, M., & Hamill, T. M. (2015). Statistical Postprocessing of Ensemble Precipitation
792 Forecasts by Fitting Censored, Shifted Gamma Distributions*. *Monthly Weather Review*.
793 <https://doi.org/10.1175/MWR-D-15-0061.1>

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

797 Serpetzoglou, E., Anagnostou, E. N., Papadopoulos, A., Nikolopoulos, E. I., & Maggioni, V.
798 (2010). Error propagation of remote sensing rainfall estimates in soil moisture prediction from
799 a land surface model. *Journal of Hydrometeorology*, 11(3), 705–720.
800 <https://doi.org/10.1175/2009JHM1166.1>

801 Shrestha, A., Nair, A. S., & Indu, J. (2020). Role of precipitation forcing on the uncertainty of land
802 surface model simulated soil moisture estimates. *Journal of Hydrology*, 580(February 2019),
803 124264. <https://doi.org/10.1016/j.jhydrol.2019.124264>

- 804 Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B., & Krajewski, W. (2013). Extreme flood
805 response: The June 2008 flooding in Iowa. *Journal of Hydrometeorology*, 14(6), 1810–1825.
806 <https://doi.org/10.1175/JHM-D-12-0191.1>
- 807 Sperna Weiland, F. C., Vrugt, J. A., van Beek, R. L. P. H., Weerts, A. H., & Bierkens, M. F.
808 (2015). Significant uncertainty in global scale hydrological modeling from precipitation data
809 errors. *Journal of Hydrology*, 529, 1095–1115.
810 <http://dx.doi.org/10.1016/j.jhydrol.2015.08.061>
- 811 Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., & Blöschl, G. (2020). The
812 Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a
813 Small Catchment. *Water Resources Research*, 56(10).
814 <https://doi.org/10.1029/2019WR026153>
- 815 Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). IMERG V06: Changes to the
816 morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, 36(12), 2471–2482.
817 <https://doi.org/10.1175/JTECH-D-19-0114.1>
- 818 Towler, E., & McCreight, J. L. (2021). A wavelet-based approach to streamflow event
819 identification and modeled timing error evaluation. *Hydrology and Earth System Sciences*,
820 25(5), 2599–2615. <https://doi.org/10.5194/hess-25-2599-2021>
- 821 Trambly, Y., Bouvier, C., Ayral, P. A., & Marchandise, A. (2011). Impact of rainfall spatial
822 distribution on rainfall-runoff modelling efficiency and initial soil moisture conditions
823 estimation. *Natural Hazards and Earth System Science*, 11(1), 157–170.
824 <https://doi.org/10.5194/nhess-11-157-2011>
- 825 Trambly, Y., Bouvier, C., Martin, C., Didon-Lescot, J. F., Todorovik, D., & Domergue, J. M.
826 (2010). Assessment of initial soil moisture conditions for event-based rainfall-runoff
827 modelling. *Journal of Hydrology*, 387(3–4), 176–187.
828 <https://doi.org/10.1016/j.jhydrol.2010.04.006>
- 829 Vennapusa, P. K. R., & White, D. J. (2015). Performance Assessment of Secondary-Roadway
830 Infrastructure in Iowa after 2011 Missouri River Flooding. *Journal of Infrastructure Systems*,

831 21(4), 1–11. [https://doi.org/10.1061/\(asce\)is.1943-555x.0000255](https://doi.org/10.1061/(asce)is.1943-555x.0000255)

832 Verbunt, M., Walser, A., Gurtz, J., Montani, A., & Schär, C. (2007). Probabilistic flood forecasting
833 with a limited-area ensemble prediction system: Selected case studies. *Journal of*
834 *Hydrometeorology*, 8(4), 897–909. <https://doi.org/10.1175/JHM594.1>

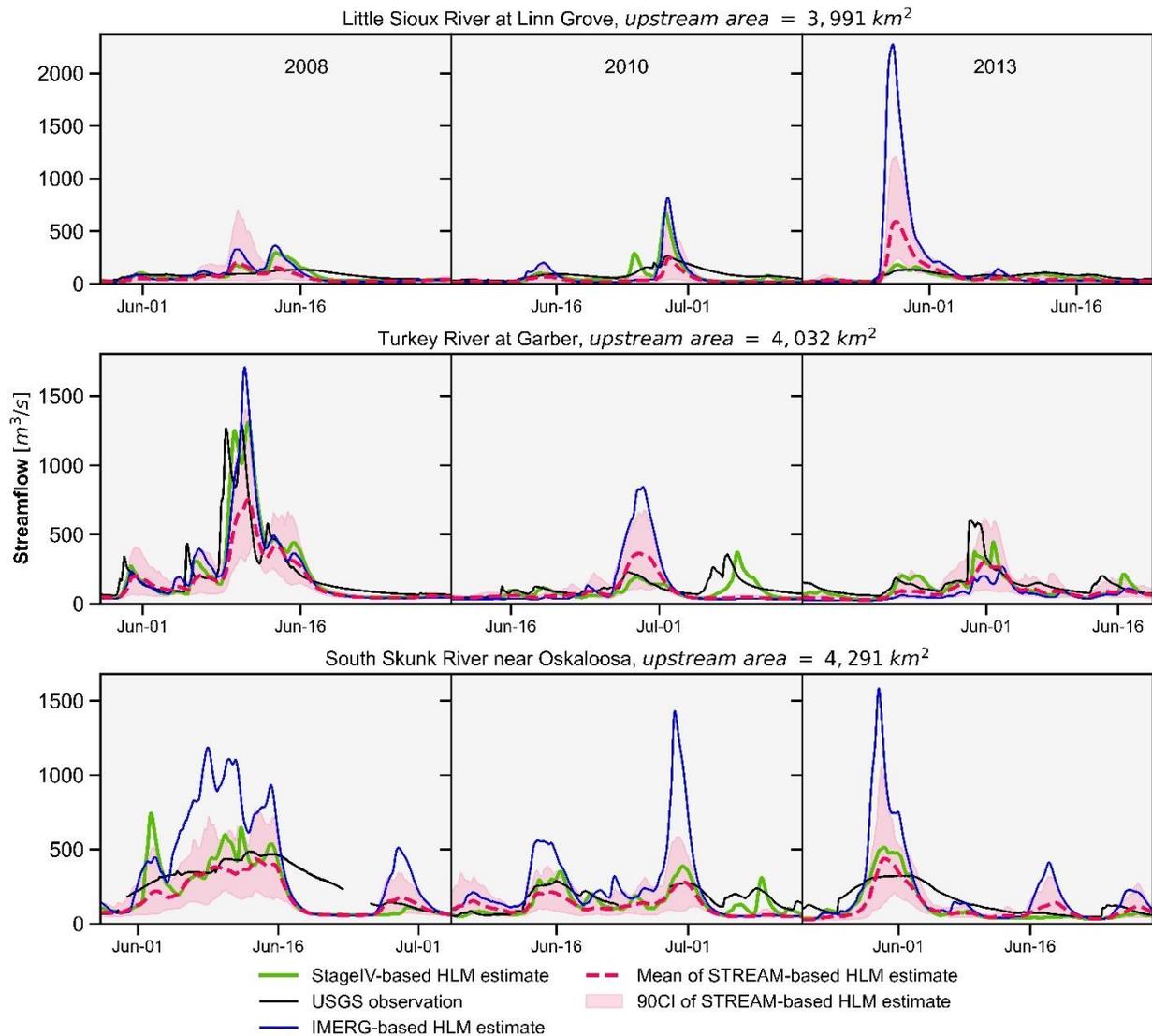
835 Wang, Z., Zhong, R., Lai, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based
836 precipitation products and the hydrological utility. *Atmospheric Research*, 196, 151–163.
837 <https://doi.org/10.1016/j.atmosres.2017.06.020>

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

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

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

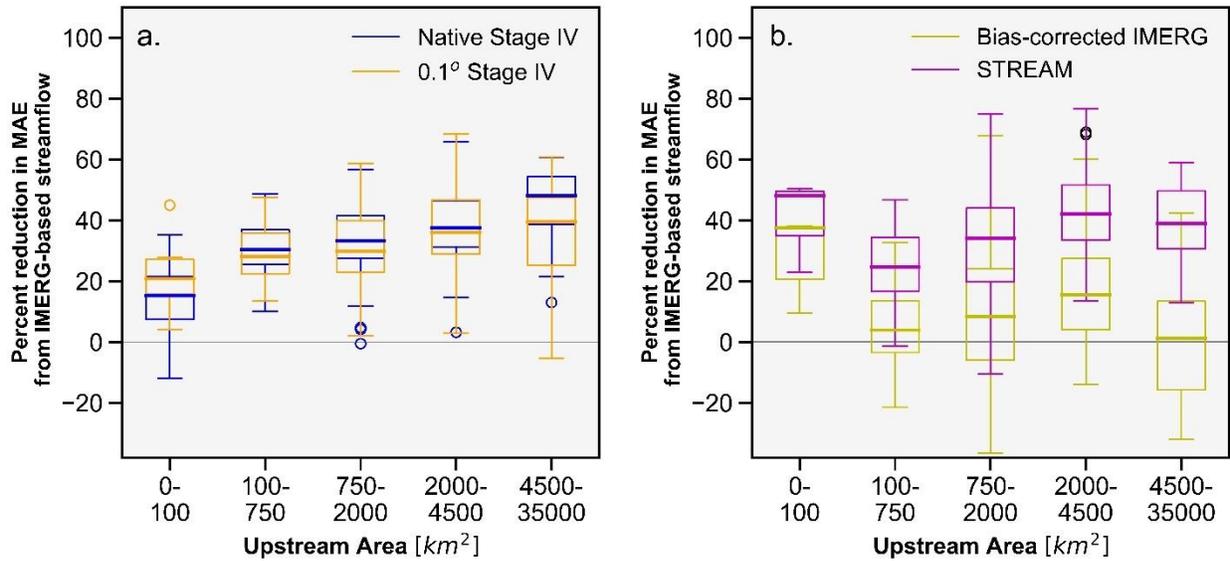
848



Supplemental Figure 1. As in Figure 8, example of identified peak streamflow events and performance of HLM streamflow estimates at three additional USGS gauge sites. Observed streamflow from USGS gauges (black) and predicted streamflow is shown for HLM results based on Stage IV (green), IMERG (blue), and the STREAM ensemble (red).

850

851



Supplemental Figure 2. As in Figure 10, but for top 10% of USGS streamflow observations. (a) Percent reduction in streamflow MAE from IMERG-based streamflow to native resolution and 0.1° Stage IV-based streamflow. This represents the percentage of streamflow MAE that is attributable to IMERG uncertainty and demonstrates the relatively small portion of streamflow MAE that is attributable to the resolution difference in precipitation data. (b) The reduction in streamflow MAE when using bias-corrected IMERG and the STREAM ensemble instead of IMERG as precipitation input to the Hillslope Link Model. The MCRPS is used to calculate the reduction in streamflow MAE for the STREAM ensemble.