AGU PUBLICATIONS

Water Resources Research

Supporting Information for

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

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

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

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

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

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

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

Contents of this file

Introduction

The supporting information in this file provides figure of EM-Earth time series to be compared with STREAM-Sat, temporal and spatial autocorrelation of STREAM-Sat compared to other global precipitation datasets, and evaluation of STREAM-Sat conditioned on different IMERG components.

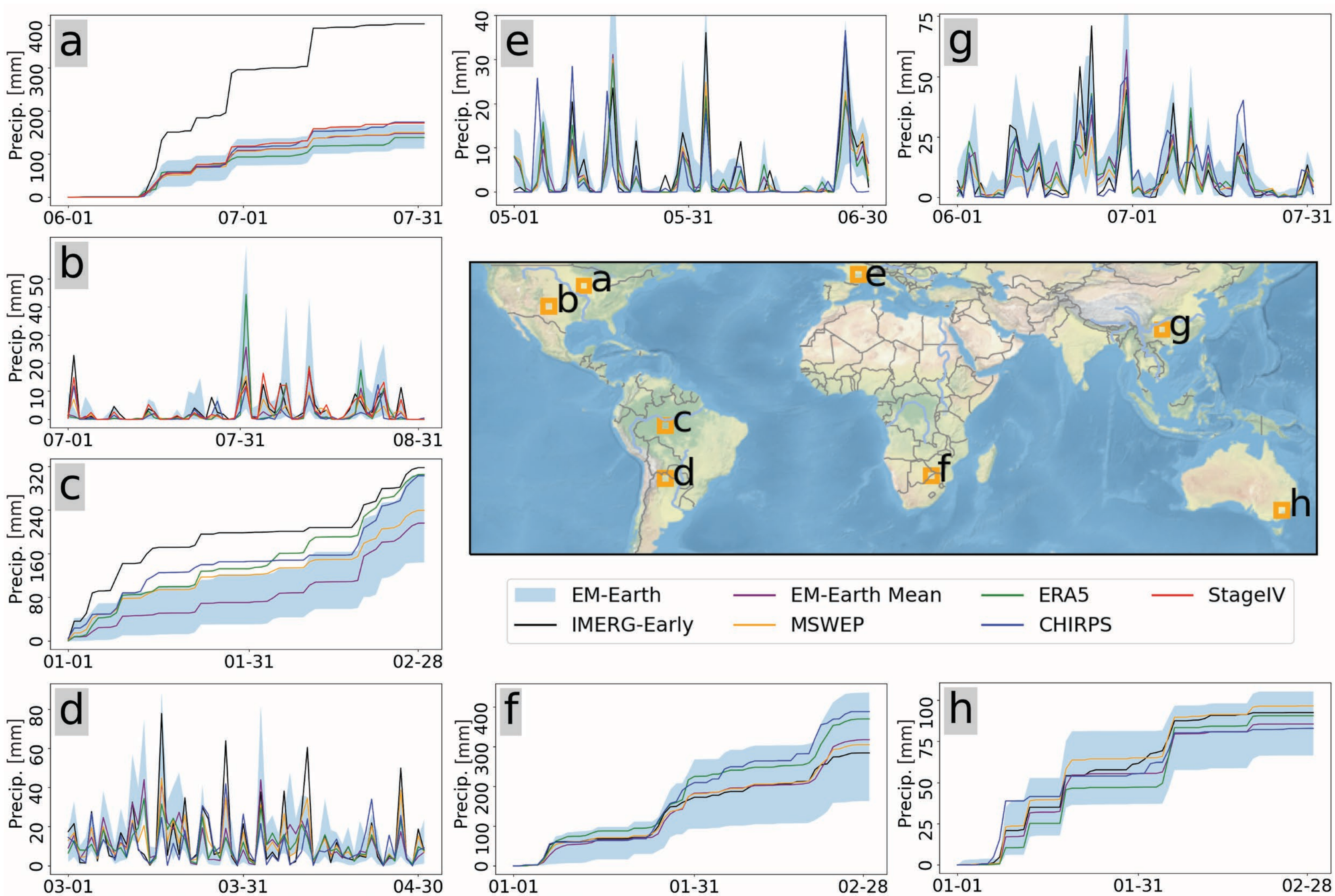

Figure S1. The same as Figure 5, but for EM-Earth.

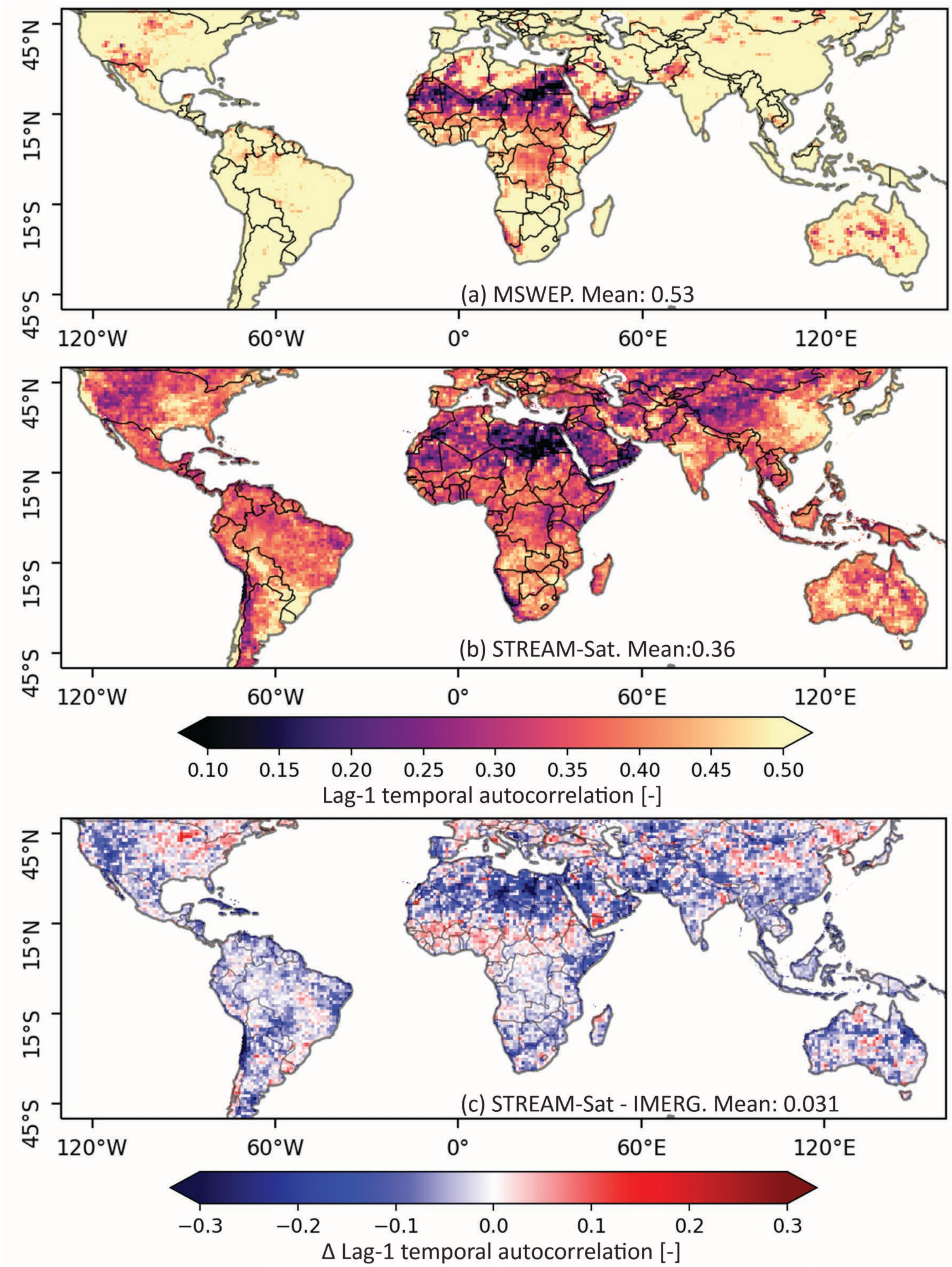

Figure S2. Lag-1 temporal autocorrelation at 3-hourly and 0.1°. (a) MSWEP. (b) STREAM-Sat. (c) Difference between STREAM-Sat and IMERG. Negtive value means STREAM-Sat has lower autocorrelation. Global average value is provided for each panel. IMERG and STREAM-Sat were aggregated.

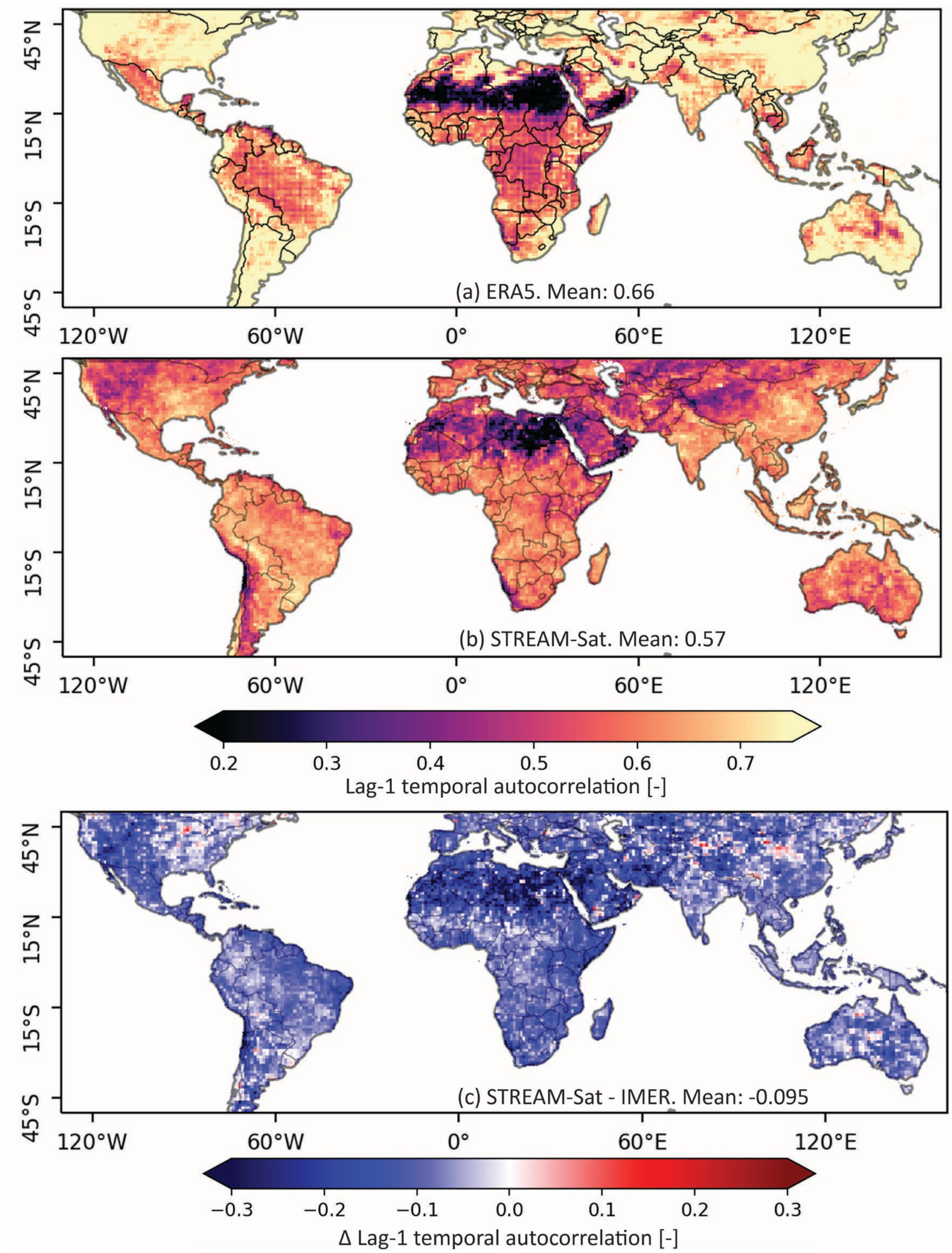

Figure S3. The same as Figure S2, but at hourly and 0.25°.

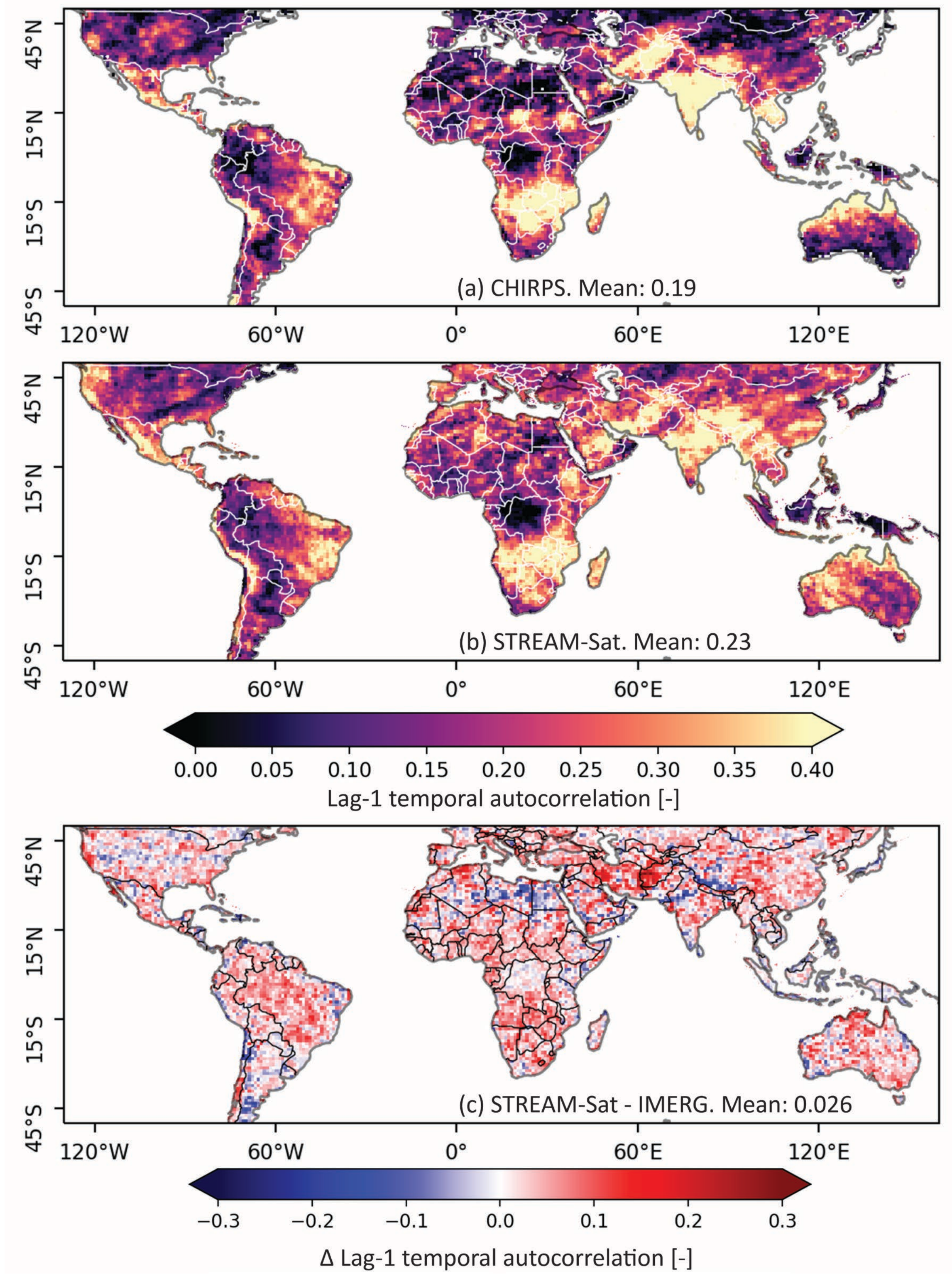

Figure S4. The same as Figure S2, but at daily and 0.1°.

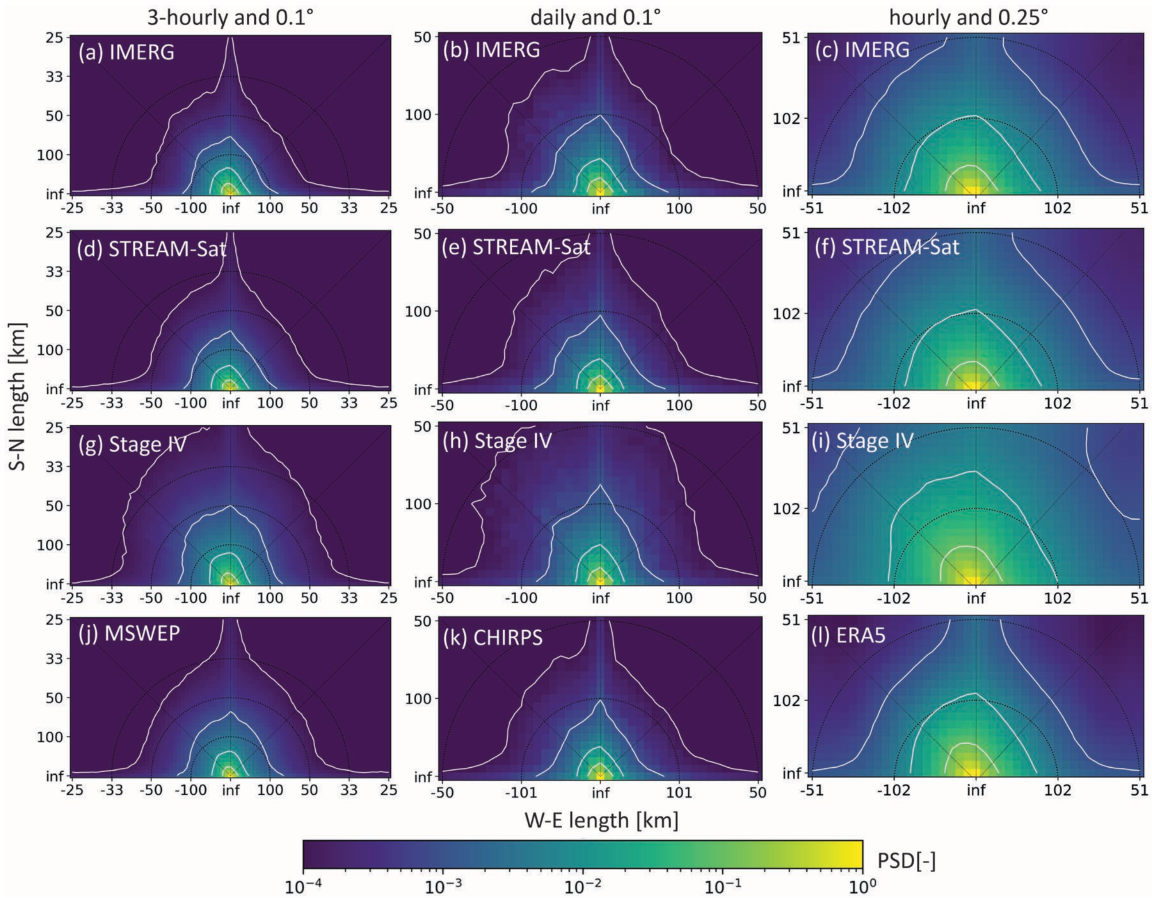

Figure S5. Fourier Power spectrum density normalized by zero frequency at three different scales in the sample region 41°N-31°N, 101°W-91°W. (a)-(c) IMERG. (d)-(f) STREAM-Sat. (g)-(i) Stage IV. (j)-(l) three Comparison datasets. Contour lines are at 1e-4, 1e-3, 1e-2, 1e-1. For the third column, 1e-4 is out of the range.

Text S1. IMERG components in error model training

Our approach to training the CSGD error model faces one important limitation. In relying on observations from DPR as "training reference," we are only able to train on spatially and temporally coincident data from IMERG and DPR. Because IMERG relies heavily on GMI when and where it is available, and because GMI and DPR are collocated on the GPM satellite, all training data in IMERG is associated with GMI observations, while other PMW and infrared sensors, as well as morphed interpolation components barely contribute to this training. Since GMI is generally the best PMW sensor currently flying (Ayat, Evans, & Behrangi, 2021), we will tend to understate the uncertainty in IMERG for times and places where GMI is unavailable (see Li et al., 2023 for additional details).

The central hypothesis we validated here is whether DPR-based products have sufficient reliability to parameterize the IMERG error model. To better understand this issue, Stage IV in the southeast US (40°N-30°N, 105°W-85°W) was used as a validation reference to evaluate STREAM-Sat conditioned on different IMERG components. A similar approach as STREAM-Sat but relying on Stage IV as the training reference in error model training was also tested (refer as STREAM-Stage IV). The comparison between the two can separate effects of training samples on the results.

We separated IMERG data into four parts (Tan, Petersen, & Tokay, 2016): 1) PMW: data with an identified PMW instrument and a "IRkalmanFilterWeight" of zero, 2) Morph only: data without PMW instrument assigned and a "IRkalmanFilterWeight" of zero. 3) IR only: "IRkalmanFilterWeight" equal to 100%, 4) IR+morph: "IRkalmanFilterWeight" between 0 and 100%. The GMI component was also evaluated separately. Table S1 collects the contribution of each component in a 1° by 1° training sample for both STREAM-Sat and STREAM-Stage IV. In STREAM-Sat, 80% of data samples are from GMI, and 5% are from other passive microwave platforms. IR accounts for 8%. About 4% of data is from morph only or IR+morph. But in STREAM-Stage IV, 70% of data samples are from morph only or IR+morph, with only 20% from PMW and 5% from IR. The evaluation in terms of CR for each component is shown in Figure S6. STREAM-Sat's representation of uncertainty is relatively good when PMW and morphing are used; the metrics in Figure S6 are quite similar to those when GMI is used. A similar conclusion can also be drawn when Stage IV is used for training. IR in STREAM-Sat exhibits much poorer uncertainty representation, with higher probabilities of misses and false alarms. Results when Stage IV is used for training also show reduced performance in IR components, albeit with some mitigation, implying that poor uncertainty representation in IR is partly attributable to problematic IR retrievals and partly due to the training data source. It is not surprising that ensemble results when Stage IV was used as training has higher CR metrics than STREAM-Sat, especially in misses class.

Sensor type	GMI	AMSR-2	SSMIS	MHS	ATMS	IR Only	Only Morph	Morph +IR
STREAM-Sat	36,439	712	1456	142	46	3,824	768	901
STREAM-Stage IV	6741	7842	24475	27463	10829	23299	73223	176809

Table S1. Sample Size of Each IMERG Component in 1° by 1° Training Samples.

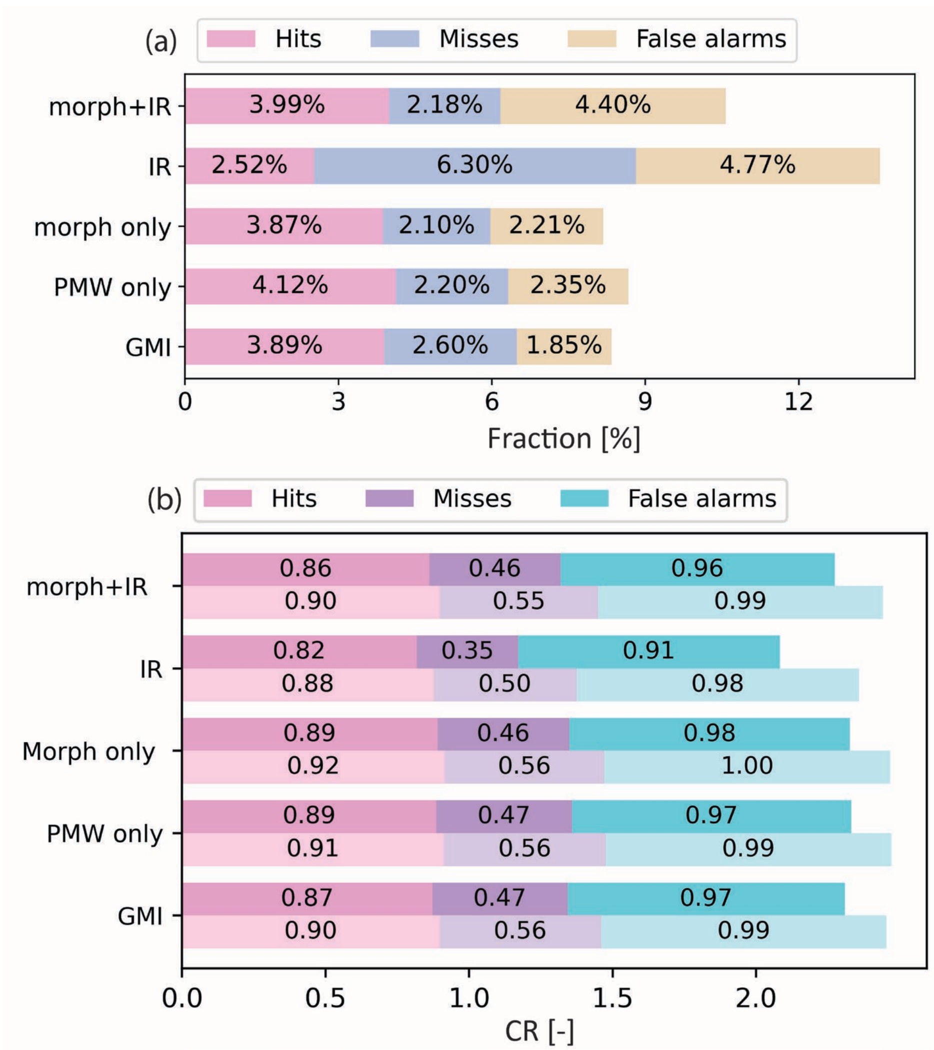

Figure S6. Evaluation of STREAM-Sat and STREAM-Stage IV for different IMERG components. (a) Fraction of three classes for five components. (b) CR of three classes for five components. Results of STREAM-Sat and STREAM-Stage IV are shown. Stage IV was used as validation reference in this example. The remaining (unshown) fraction is for correct non-detects. All data are aggregated at hourly and 0.1°.