Geophysical Research Letters

Supporting Information for

Automated Quality Control Scheme for GPM Satellite Precipitation Products

Jackson Tan[1,2], George J. Huffman[1], and Yi Song[1]

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

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

Contents of this file

Text S1, S2
Figures S1 to S5
Tables S1, S2

Introduction

This document provides supporting information for the *Geophysical Research Letters* article entitled "Automated Quality Control Scheme for GPM Satellite Precipitation Products". It includes a description on the details of an experiment that is tangential to the main study, five additional figures that provide deeper insights into different aspects of the Satellite Precipitation Estimate Error Detector (SPEEDe) scheme, and one table listing the composition of the GPM Constellation.

Text S1.

In developing SPEEDe, variants of the model architecture were explored before the final architecture described in the study and Figure 2 was adopted. These variants were developed using an early set of known bad cases and a selection of good cases, which may not be as comprehensive as the final list used for the study. These variants tested different kernel sizes, number of filters, number of hidden layers, as well as using Conv2DTranspose layers instead of Conv2D layers in the decoding half of the model.

Table S2 enumerates the variants and summarizes the model settings for each variant, as well as three output metrics for assessing the variant's performance. Specifically, the three metrics are the lowest SPEEDe score of the set of bad cases, the 10^{th} percentile SPEEDe score of the set of bad cases, and the highest SPEEDe score of the set of good cases. "Good performance" is determined by how much lower the highest score of the good cases is relative to the distribution of the scores of the bad cases. As well, box-whisker diagrams of these scores were analyzed (not shown). We stress that, since the scores are dependent on the machine learning model settings, comparisons cannot be made between the absolute scores across different variants.

From this preliminary testing, we decided to use a machine learning model with three pairs of Conv2D and MaxPooling2D in the hidden encoding layer (and equivalently in the decoding layer), with 64 filters in the outermost hidden layers and 32 filters in the inner hidden layers, and using Conv2DTranspose for the hidden decoding layers. This is listed as v0.4.9 in Table S2.

In addition, a variant in which log-transformation was not applied to the precipitation rates (Sec. 2.3) was tested. Comparing this to the log-transformed variant (v0.6.0 vs. v0.4.9.1 in Table S2), the ratio of the maximum score of the good cases to the 10^{th} percentile score of the bad cases is 1.13 for the variant without log-transformation and 1.07 for the variant with log-transformation. Such a comparison revealed a marginal advantage with the log-transformation.

Text S2.

To quantify the deviation of the distribution of a half-hour IMERG precipitation field from the climatology, we use the normalized histogram counts by computing the counts in the 12 logarithmically spaced bins as shown in Figure 1 (including the extrema), then dividing them by the total counts such that all bin counts sum to one. We first compute the climatological histogram over all half-hours in 2014—a year with no known bad data—before doing the same for each half-hour and computing the mean squared error between the half-hour's histogram and the climatology's histogram.

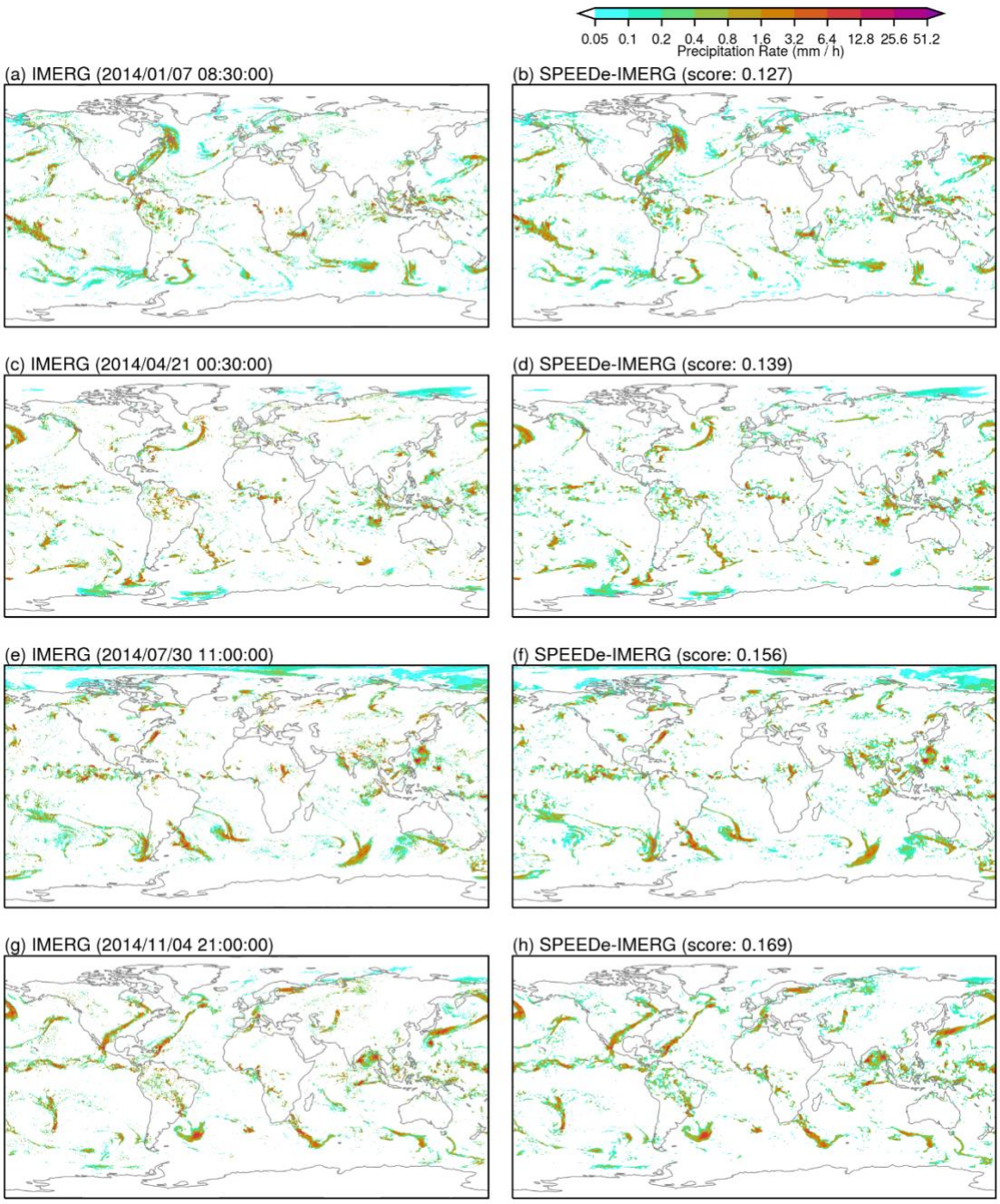


Figure S1. (a, c, e, g) Four randomly selected half-hourly IMERG V07A snapshots showing precipitation fields without the presence of bad orbits. (b, d, f, h) The SPEEDe-IMERG outputs corresponding to these snapshots.

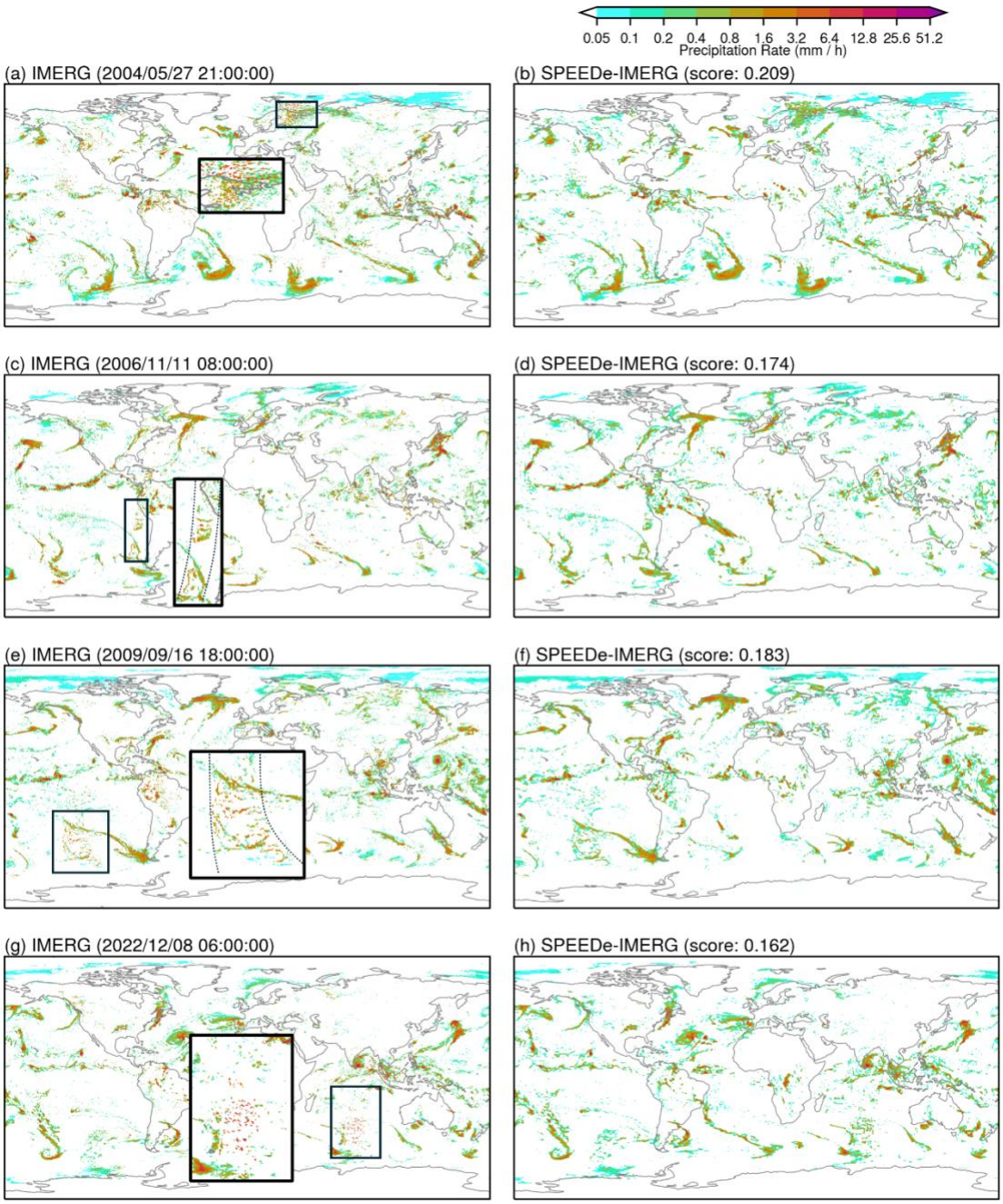


Figure S2. (a, c, e, g) Four randomly selected half-hourly IMERG V07A snapshots showing precipitation fields from marginally bad cases. (b, d, f, h) The SPEEDe-IMERG outputs corresponding to these snapshots. Artifacts are present (a) east of Finland; (c) over the oceans west of South America and north and south of western Australia; (e) in the southern Pacific Ocean; and (g) in the southern Indian Ocean. The black boxes indicate inset figures that enlarge the regions with artifacts; (c,e) satellite orbit swaths are indicated with black dotted lines to denote the increased precipitation occurrence within the swath.

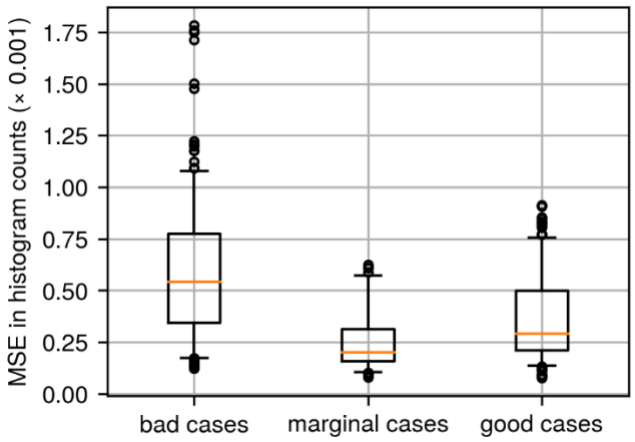


Figure S3. Distributions of the mean squared error in histogram counts for known bad cases, known marginal cases, and the same 200 normal cases in Figure 3a.

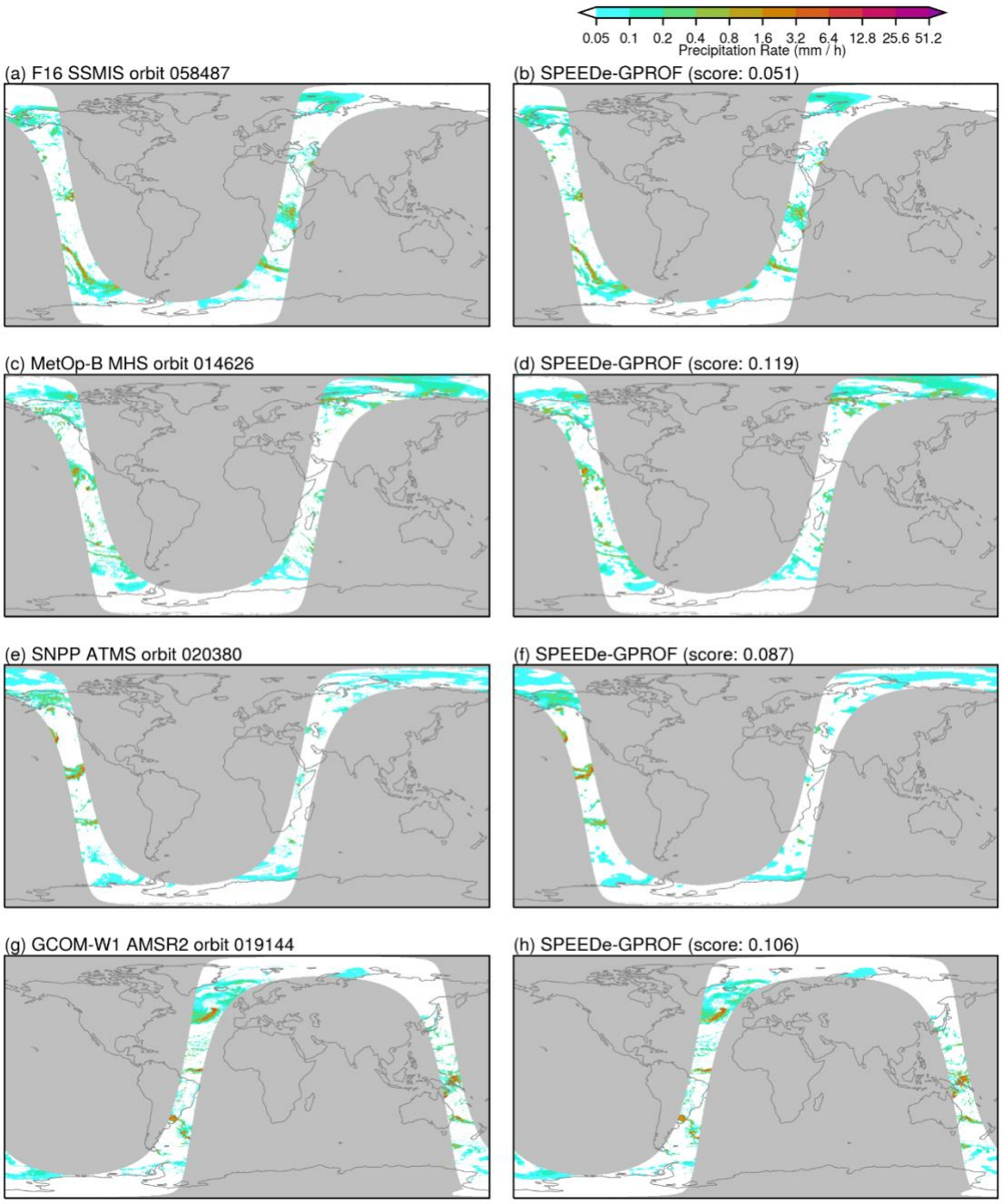


Figure S4. Examples of good orbit granules from (a) F16 SSMIS, (c) MetOp-B MHS, (e) SNPP ATMS, and (g) GCOM-W1 AMSR2, with (b, d, f, h) their corresponding SPEEDe-GPROF outputs.

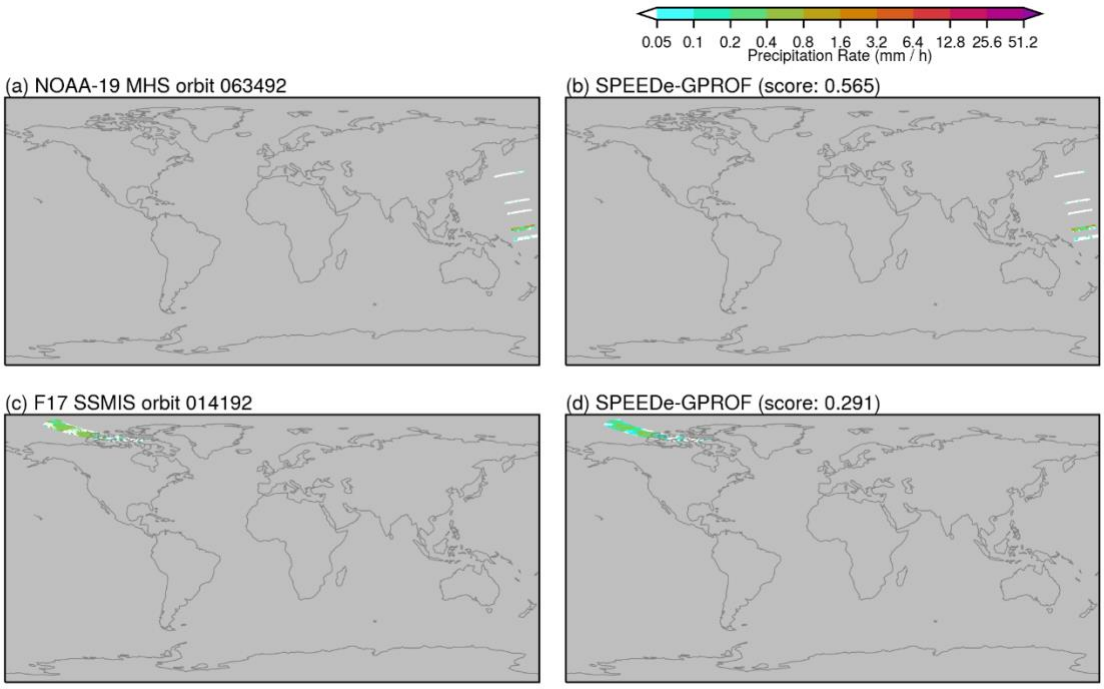


Figure S5. Examples of orbit granules with limited non-missing values from (a) F16 SSMIS and (c) GCOM-W1 AMSR2. (b, d) Their corresponding SPEEDe-GPROF outputs have a high score despite the absence of bad data.

Sensor	Abbreviation	Satellite(s)	Agency(s)
GPM Microwave Imager	GMI	GPM Core Observatory	NASA, JAXA
TRMM Microwave Imager	TMI	TRMM	NASA, JAXA
Advanced Microwave Scanning Radiometer 2	AMSR2	GCOM-W1	JAXA
Advanced Microwave Scanning Radiometer for EOS	AMSR-E	Aqua	JAXA
Advanced Microwave Sounding Unit–B	AMSU-B	NOAA-15,16,17	NOAA
Advanced Technology Microwave Sounder	ATMS	Suomi NPP; NOAA-20,21	NASA, NOAA
Microwave Humidity Sounder	MHS	NOAA-18,19; MetOp-A,B,C	NOAA
Special Sensor Microwave Imager	SSMI	DMSP F10,11,13,14,15	DoD
Special Sensor Microwave Imager / Sounder	SSMIS	DMSP F16,17,18,19	DoD

Table S1. The GPM Constellation of PMW sensors. See https://gpm.nasa.gov/missions/GPM/constellation for the periods of operation.

Version	Kernel size / layers	Filters	Min. score of bad cases	10th percentile score of bad cases	Max. score of good cases
v0.4.1	5 (for Conv2D); 5, 3, 2 (for MaxPooling2D)	32, 16	0.279	0.372	0.364
v0.4.2	5; 5, 3, 2	64, 32	0.237	0.322	0.298
v0.4.3	5; 5, 3, 2	16, 8	0.286	0.382	0.375
v0.4.4	5, 3, 2 (for MaxPooling2D and Conv2D)	32, 16	0.273	0.365	0.356
v0.4.5	5, 3, 2, 2	32, 16	0.377	0.497	0.521
v0.4.6	5, 5, 3	32, 16	0.374	0.498	0.532
v0.4.7.1	5, 3, 2	32, 16	0.187	0.253	0.229
v0.4.7.2	5, 2	32, 16	0.077	0.101	0.091
v0.4.8	Same as v0.4.4 but using Conv2DTranspose in decoding hidden layers	32, 16	0.261	0.350	0.337
v0.4.9	(Combining v0.4.2, v0.4.7.1, and v0.4.8)	64, 32	0.169	0.226	0.204
v0.4.9.1	(Using a revised list of good cases)	64, 32	0.168	0.214	0.230
v0.6.0	(Dropping log-transformation)	64, 32	8.468	11.943	13.477

Table S2. Variants of the convolutional autoencoder architecture explored in preliminary testing. Kernel size refers to the $n \times n$ kernel size for the 2D layers and shows only the encoding half of the architecture (the decoding half is the mirror image of the encoding half); for v0.4.1 to v0.4.3, the paired Conv2D and MaxPooling2D layers have different kernel sizes, but for v0.4.4 onwards the pairs of Conv2D and MaxPooling2D have the same kernel size. The filter size refers to the number of filters (or neurons) for the

outermost hidden layers and the inner hidden layers respectively. “Good performance” is determined by how much lower the highest score of the good cases is relative to the distribution of the scores of the bad cases. Version 0.4.9 of SPEEDe is the final version emerging from this set of preliminary testing.