GPROF V7 and beyond: Assessment of current and potential future versions of the GPROF passive microwave precipitation retrievals against ground radar measurements over the continental US and the Pacific Ocean

Response to reviewers

1 Response to reviewer 1

We thank the reviewer for investing their time to read our manuscript and provide constructive feedback to improve it. The principal changes that we have implemented based on comments from both reviewers are the following:

1. We repeated our analysis excluding all precipitation classified as frozen by MRMS, i.e., excluding not only snow but also hail. Significant effects of excluding hail were only observed in the SW region and we have updated the manuscript accordingly.
2. We have added a short section that analyzes the impact of excluding both frozen precipitation as well as precipitation of snow-covered and mountain surfaces and the behavior of the different retrievals over those surfaces.
3. We have revised all references to the a priori database and the MRMS validation data to make the distinction of the two clearer.

Finally, we have also updated that precision-recall curves to show the precision over the recall instead of the recall over the precision, which is the more common way to present these curves. During this, we also realized that the PR curves previously included samples over ocean, which we have now excluded. This lead to a minor change in the relative skill of the GPROF-NN HR retrieval.

In what follows, line and figure numbers are given with respect to the revised manuscript.

1.1 Specific comments

Reviewer comment 1

Abstract, Line 12-13: What does "retrieval reproduces the principal precipitation characteristics of each region" mean? Can you please elaborate?

Author response:

What we were referring to here were the different regional seasonal and diurnal cycles of precipitation, which were reliably reproduced by the GPROF retrievals. We have reformulated the sentence in the revised manuscript to make this clear.

Changes in manuscript:

> **Changes starting in line 13:**
>
> Although biases of up to 25 % are observed over sub-regions of the CONUS and the tropical Pacific, the retrieval ~~reproduces the principal precipitation characteristics of each region~~reliably reproduces each region's diurnal and seasonal precipitation characteristics.

Reviewer comment 2

Abstract, Line16-18: I appreciate that authors are providing this significant finding here at the abstract, however, can you please be more specific about the time resolution of this comparison? Meaning at what time resolution GPROF NN 1D is improving mean absolute error, corr etc.?

Author response:

All retrieval errors are computed with respect to instantaneous precipitation estimates at 5 km resolution. We have added this information to the sentence in question in the revised version of the manuscript.

Changes in manuscript:

> **Changes starting in line 18:**
>
> GPROF-NN 1D, the most basic neural network implementation of GPROF, improves the mean-squared error, mean absolute error, correlation and symmetric mean absolute percentage error of instantaneous precipitation estimates by about twenty percent for GPROF GMI while the effective resolution is improved to 31 km over land and 15 km over oceans.

Reviewer comment 3

Line 58-60: This is a very confusing sentence. Can you please reword it?

Author response:

We agree with the reviewer that the sentence is badly worded. We have reformulated it in the revised version of the manuscript.

Changes in manuscript

> **Changes starting in line 58:**
>
> This study compares ~~the~~ GPROF retrievals to independent ~~precipitation estimates and the error~~ validation data derived from ground-based precipitation radars. In

> this case, differences between a priori database and ~~validation data will thus lead to additional errors in the retrieval. Attributing whether retrieval errors found in the validation against independent measurements are due to the limitation of the input observations and the retrieval method or the a priori database~~ the validation data constitute a second source of errors that will increase the total retrieval error. These two sources of error are fundamentally different and reducing their impact requires different approaches. Therefore, quantifying the extent to which these sources contribute to the total retrieval error is essential to guide future efforts to improve GPM PMW retrievals.

Reviewer comment 4

Line 64: Can you please reword this question, something along the lines: “to what extent a priori database errors contribute to GPROF overall retrieval errors?”

Author response:

We have reworded the question in the revised version of the manuscript.

Changes in manuscript

> **Changes starting in line 66:**
>
> ~~To what extent contribute~~ What is the contribution of errors in the a priori database to ~~GPROF retrieval errors~~the GPROF retrieval error?

Reviewer comment 5

Line 65: I think it would be better to remove “even” from this question“... GPM PMW observations even when compared to ...”

Author response:

We have removed even in the revised version of the manuscript.

Changes in manuscript

> **Changes starting in line 67:**
>
> Can the GPROF-NN retrievals improve precipitation estimates of the GPM PMW observations ~~even~~ when compared to independent measurements?

Reviewer comment 6

Line 90: Authors mention that rain gauge corrected MRMS data are used. Can authors please be more specific which database they have used? Because the way it has been presented is slightly confusing. Gauge corrected MRMS precipitation magnitudes are accumulations. However, radar only MRMS data provides precipitation rates at 2 min temporal intervals. Did the authors conduct their own gauge correction to the radar only MRMS product?

Author response:

The MRMS estimates that are used in the study are instantaneous, gauge-corrected radar QPEs. This is a special product that is produced specifically for GPM ground validation. It uses gauge correction factors derived from hourly gauge-corrected and radar-only accumulations to correct instantaneous radar QPE's.

Changes in manuscript:

> **Changes starting in line 97:**
>
> ~~Instantaneous~~ The principal source of validation measurements for this study are instantaneous, gauge corrected precipitation estimates from the NOAA Multi-Radar Multi-Sensor System (MRMS)~~are used as the principal source of validation measurements in this study. The estimates used here~~. These estimates are produced specifically for GPM ground validation and are gauge-corrected to match monthly accumulations ~~(Kirstetter et al., 2012). The processing~~ following the approach described (Kirstetter et al., 2012). These estimates are not part of the operational MRMS processing suite but can be obtained from the GPM ground validation data archive (Wolff, 2023). The processing of the ground-validation data includes a basic filtering that removes measurements with excessive gauge-correction factors (Kirstetter et al., 2012). The data is provided on an approximately $0.01\,^\circ \times 0.01\,^\circ$ grid covering ~~CONUS. The~~ the CONUS. For the comparison against the satellite retrievals, the MRMS data is smoothed using a Gaussian average filter with a full-width at half-maximum (FWHM) of 5 km~~and interpolated to the 5 km × 5 km collocation grid~~. Following this, the mapping to the collocation grid is performed using nearest-neighbor interpolation.

Reviewer comment 7

Line 211-212: Can authors please clearly indicate whether the mountain surfaces are excluded or included with a correction.

Author response:

We have rewritten this section to clearly state that these pixels are excluded because of the correction applied to them.

Changes in manuscript

> **Changes starting in line 226:**
>
> ~~Snow-covered~~ Retrievals over snow-covered and mountain surfaces are excluded from the validation ~~. For the former, reference precipitation in the retrieval database~~ due to the uncertainties in both the satellite estimates as well as the validation data. In addition to this, the GPROF a priori database for snow-covered surfaces is derived from collocations with MRMS~~while, for the latter, a correction for orographic precipitation is applied. These modification~~ , while precipitation over mountains is obtained by scaling the GPM CMB precipitation to account for the orographic enhancement of precipitation. These two modifications aim to counteract known weaknesses ~~in~~ of the GPM CMB ~~reference data but they~~ retrievals, but would skew the comparison between GPM, GPM CMB, and MRMS. ~~Furthermore~~
>
> Similarly, precipitation that is identified as ~~snow~~ frozen by MRMS is excluded from the validation. The retrieval of frozen precipitation from both PMW and radar is particularly challenging due to its uncertain radiometric properties. Because of these increased uncertainties and the small contribution of frozen precipitation to the total precipitation in the validation data, the retrieval accuracy for frozen precipitation should be assessed in a dedicated study.

Reviewer comment 8

Line 228: Can authors please explain how they calculate the bias or what is the definition of the bias? And at what temporal resolution (I am assuming this is annual but it would be nice to indicate).

Author response:

We have added the requested information in the revised version of the manuscript.

Changes in manuscript

> **Changes starting in line 254:**
>
> Maps of ~~the retrieval biases~~annual mean retrieval biases, calculated as the annual average of the difference between retrieved precipitation and the precipitation in the a priori database or the MRMS validation data, are displayed in Fig. 3.

Reviewer comment 9

Line 239-242: To add to the explanation here (this is through a personal communication with a MRMS team member): "It is not documented on Iowa website, however, in Oct 2020, the gauge correction methodology and associated products are changed." This corresponds exactly to the 2021 water year that authors are using in this study.

Author response:

We would like to thank the reviewer for this useful information. We have included it in the revised version of the manuscript.

Changes in manuscript

> **Changes starting in line 274:**
>
> ~~A potential explanation for this change in the bias patterns is a version~~ As pointed out by one of the anonymous reviewers, it is likely that this is due a change in the ~~MRMS processing that occurred in February 2021 (NOAA National Severe Storms Laboratory, 2023)~~ gauge correction methodology that occurred around October 2020 (Anonymous Referee 2, 2023) .

1.1.1 Reviewer comment 10

Line 244: "... it is possible that the bias relative to MRMS is introduced by rain gauge correction" please include "by" in this sentence to make it clear.

Author response:

We have fixed this in the updated version of the manuscript.

Changes in manuscript:

> **Changes starting in line 270:**
>
> Given that the CMB over land is largely a radar-derived product, it is possible that the bias relative to MRMS is introduced by the rain gauge correction applied to the validation data and caused by precipitation properties that may not be resolved by the radar observations.

Reviewer comment 10

Line 266-268: Can authors please describe why they decide to use mean error, mean-squared error and mean absolute error all together? What do they explain differently and why did authors needed all of them together? Moreover, can authors please describe symmetric mean absolute percentage error in more detail i.e., what does this score mean, what are the max and min values etc.

Author response:

We have decided to include multiple error metrics in our analysis because our ultimate aim is to improve precipitation estimates and not just tune them to optimize a single error metric. Both MSE and MAE are fairly common error metrics and providing them can provide a reference for other retrievals. However, MSE and, to a lesser extent, MAE are dominated by heavy precipitation.

A relative error such as the mean absolute percentage error (MAPE) is more sensitive retrieval errors for light precipitation estimates (c.f. Fig. 1.2). However, the issue with the MAPE is that it penalizes overestimation heavier than underestimation. The symmetric mean absolute percentage error corrects this shortcoming by modifying the MAPE to be symmetric in its arguments.

We added an appendix to the revised manuscript that motivates our choice of metrics, states their formulas, and illustrates the characteristics of MSE, MAE, MAPE, and SMAPE. The appendix contains the figure shown in Fig. 1.1, which displayes the different behavior of the error functions underlying MAE, MSE, MAPE and SMAPE. It also shows the asymmetry of the MAPE, which motivated our choice of the SMAPE over MAPE.

Finally, the appendix also contains the figure shown in Fig. 1.2, which shows the relative contribution of different precipitation intensities to the final value of the metric. This figure clearly shows the complementary behavior of MSE, MAE, and SMAPE in terms of sensitivity to different precipitation intensities.

Changes in manuscript:

1. We have rewritten the paragraph introducing the metrics:

> **Changes starting in line 297:**
>
> The assessed metrics include the mean error (Bias), mean-squared error (MSE), mean absolute error (MAE), correlation coefficient and the symmetric mean absolute percentage error (SMAPE~~$_t$), which is defined as:~~
>
> $$\mathrm{SMAPE}_t(P, P_{\mathrm{MRMS}}) = \frac{1}{N_{P_{\mathrm{MRMS}} \geq t}} \sum_{P_{\mathrm{MRMS}} \geq t} \frac{|P - P_{\mathrm{MRMS}}|}{0.5(|P| + |P_{\mathrm{MRMS}}|)} \}$$
>
>). Several metrics are used to assess the retrieval accuracy in order ensure a comprehensive assessment of the quality of each algorithm's precipitation estimates. Definitions, basic characteristics and a motivation for the choosing those metrics is provided in Appendix B.

2. We have added an appendix discussing our choice of error metrics.

> **Changes starting in line 678:**
>
> Ranking the quality of precipitation estimates is a non-trivial problem because what constitutes a good estimate depends heavily on the downstream application. The ultimate motivation for this work is to improve the precipitation estimates from the PMW sensors of the GPM constellation in a way that benefits all possible downstream applications. To ensure that we are working towards an

actual improvement of these estimates instead of simply tuning the results to improve a single error metric, we use a selection of error metrics to evaluate the retrievals.

The quantitative error metrics that we use in this study are listed together with their formulas and valid range in table 1.1. The behavior of the error functions of mean squared error (MSE), mean absolute error (MAE), and symmyetric mean percentage error (SMAPE) is illustrated in Fig. 1.1. Since both MSE and MAE depend on directly on the absolute difference between estimate and reference value, the largest errors occur in the regions where either the estimate or the reference precipitation is heavy. This effect is exacerbated by the quadratic nature of the MSE.

Relative errors such as MAPE and SMAPE increase sensitivity to deviation at light precipitation rates by normalizing the error. However, since the MAPE uses only the reference precipitation for normalization it is unsymmetric and will thus favor estimates that underestimate the reference value. Since this would bias the evaluation towards retrievals that underestimate precipitation, we use the SMAPE in this study, which uses a symmetric normalization term.

The MSE, MAE, and SMAPE are evaulated by calculating their sample mean over all validation samples. Their final values are thus the combined result of the error function and the joint occurrence of retrieved and reference precipitation values. To illustrate the different characteristics of the three error metrics, Fig. 1.2 shows the relative contributions from different reference precipitation intensities to the error metrics calculated using all collocations between MRMS and GPROF V7 in the water years 2021 and 2022. As the three curves show, the three error metrics have very different contribution profiles across the spectrum of reference precipitation. While the SMAPE is most sensitive to errors at light precipitation, the MAE has a fairly flat contribution profile with a peak at moderate precipitation, and the MSE is dominated by errors at heavy reference precipitation values.

Table 1.1: Quantitative accuracy metrics used in this study. We use over bar to denote the sample mean and σ the sample standard deviation taken over all valid measurements.

Name	Formula	Lower bound	Upper bound	Optimal value
Bias	$\overline{P_\text{Retrieved} - P_\text{True}}$	$-\infty$	$+\infty$	0
Mean squared error (MSE)	$\overline{(P_\text{Retrieved} - P_\text{True})^2}$	0	∞	0
Mean absolute error (MAE)	$\overline{\lvert P_\text{Retrieved} - P_\text{True}\rvert}$	0	∞	0
Symmetric mean absolute percentage error with threshold t (SMAPE$_t$)	$\overline{\left(\frac{\lvert P_\text{Retrieved}-P_\text{True}\rvert}{\frac{1}{2}(\lvert P_\text{Retrieved}+P_\text{True}\rvert)}\right)}$ calculated only over samples with $P_\text{True} \geq t$	0	200%	0
Correlation coefficient	$\frac{\overline{(P_\text{Retrieved}-\overline{P_\text{Retrieved}})(P_\text{True}-\overline{P_\text{True}})}}{\sigma(P_\text{Retrieved})\sigma(P_\text{True})}$	0	1	1

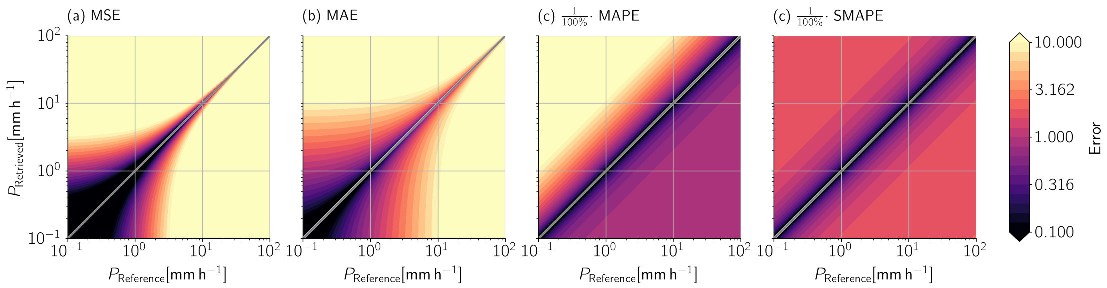

Figure 1.1: Error functions for evaluating precipitation retrievals. Panel (a) shows the value of the MSE for different combinations of retrieved and reference values. Panel (b), (c), and (d) show the according behavior for the MAE, MAPE, and SMAPE.

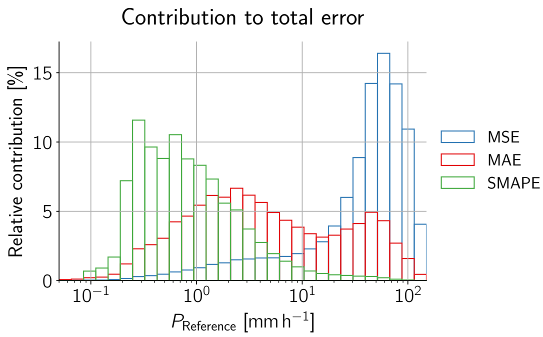

Figure 1.2: Relative contribution of different precipitation rates to total value of each error metrics. The three bar plots show the relative contribution from each corresponding bin the total value of each error statistic. The contributions were calculated for the GPROV V7 retrieval using all validation samples from the water years 2020 and 2021.

Bibliography

Anonymous Referee 2 (2023). Comment on egusphere-2023-1310. https://doi.org/10.5194/egusphere-2023-1310-RC2. Accessed: 2023-10-08.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W., and Amitai, E. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with noaa/nssl ground radar-based national mosaic qpe. *Journal of Hydrometeorology*, 13(4):1285 – 1300.

NOAA National Severe Storms Laboratory (2023). Mrms qpe code updates. https://inside.nssl.noaa.gov/mrms/code-updates. Accessed: 2023-06-13.

Wolff, D. B. (2023). Precipitation measurements missions ground validation data. https://pmm-gv.gsfc.nasa.gov/. Accessed: 2023-04-01.