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

Simon Pfreundschuh[1, 3], Clément Guilloteau[2], Paula J. Brown[1], Christian D. Kummerow[1], and Patrick Eriksson[3]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, United States of America
[2]Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, California
[3]Department of Space, Earth and Environment, Chalmers University of Technology, 41296 Gothenburg, Sweden

Correspondence: Simon Pfreundschuh (simon.pfreundschuh@colostate.edu)

Abstract.

The Goddard Profiling Algorithm (GPROF) is used operationally for the retrieval of surface precipitation and hydrometeor profiles from the passive microwave (PMW) observations of the Global Precipitation Measurement (GPM) mission. Recent updates have led to GPROF V7, which has entered operational use in May 2022. In parallel, development is underway to

5   improve the retrieval by transitioning to a neural-network-based algorithm called GPROF-NN.

This study validates GPROF V7 and multiple configurations of the GPROF-NN retrieval against ground-based radar measurements over the conterminous United States (CONUS) and the tropical Pacific. GPROF retrievals from the GPM Microwave Imager (GMI) are validated over several years and their ability to reproduce regional precipitation characteristics and effective resolution is assessed. Moreover, the retrieval accuracy for several other sensors of the constellation is evaluated.

10   The validation of GPROF V7 indicates that the retrieval produces reliable precipitation estimates over CONUS. During all four assessed years, annual mean precipitation is within 8% of gauge-corrected radar measurements. 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. The effective resolution of GPROF V7 is found to be $51\,\mathrm{km}$ over CONUS and $18\,\mathrm{km}$ over the tropical Pacific. GPROF V7 produces robust precipitation estimates also for the other sensors of the GPM constellation.

15   The evaluation further shows that the GPROF-NN retrievals have the potential to significantly improve the GPROF precipitation retrievals. 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 by about twenty percent for GPROF GMI while the effective resolution is improved to $31\,\mathrm{km}$ over land and $15\,\mathrm{km}$ over oceans. The two GPROF-NN retrievals that are based on convolutional neural networks can further improve the accuracy up to the level of the combined radar/radiometer retrievals

20   from the GPM core observatory. However, these retrievals are found to overfit on the viewing geometry at the center of the swath, reducing their overall accuracy to that of GPROF-NN 1D. For the other sensors of the constellation, the GPROF-NN retrievals produce larger biases than GPROF V7 and only GPROF-NN 3D achieves consistent improvements compared to

GPROF V7 in terms of the other assessed error metrics. This points to shortcomings in the hydrometeor profiles or radiative
transfer simulations used in the training of the retrievals for the other sensors of the GPM constellation as a critical limitation
25   for improving GPM PMW retrievals.

1   Introduction

Satellite-derived, global measurements of precipitation are essential for a wide range of societal and scientific applications,
including the identification of climatic changes in the hydrological cycle (Hegerl et al., 2015), the validation of climate models
(Tapiador et al., 2017) as well as the management of droughts (Kirschbaum et al., 2017) and risks associated with extreme
30   weather (Skofronick-Jackson et al., 2017). The Global Precipitation Measurement (GPM, Hou et al., 2014) aims to provide
such measurements at high spatial and temporal resolution. The GPM core observatory, the core satellite of the mission, carries
the Dual-frequency Precipitation Radar (DPR) and the GPM Microwave Imager (GMI). The combined observations from
DPR and GMI are used to produce global reference retrievals of precipitation and hydrometeor profiles (Grecu et al., 2016).
These reference measurements are used to produce the a priori or training data for precipitation retrievals from a constellation
35   of passive microwave (PMW) sensors. Finally, retrieval results from the combined DPR/GMI and PMW observations are
combined with data from geostationary satellites and rain gauges to produce the level 3 GPM IMERG product (Huffman et al.,
2020), which provides global precipitation estimates at a nominal spatial resolution of $10\,\mathrm{km}$ and a temporal resolution of
$30\,\mathrm{min}$.

The Goddard Profiling Algorithm (GPROF, Kummerow et al., 2015) is the algorithm used for the operational processing
40   of the PMW precipitation retrievals from GPM. Since the PMW retrievals account for the bulk of the precipitation retrievals
integrated into GPM IMERG, GPROF plays a central role in the GPM processing pipeline. Recent algorithm development
has produced a new version of the retrieval, GPROF V7, which replaced GPROF V5 in the operational processing of the
GPM precipitation products in May 2022. In parallel, two novel, neural-network-based implementations of GPROF have been
developed. These implementations, GPROF-NN 1D and GPROF-NN 3D, have been assessed under idealized conditions in
45   Pfreundschuh et al. (2022) where it was found that they can significantly improve the accuracy of the GPM PMW precipita-
tion retrievals. An analysis of the spatial variability in the retrieved precipitation fields shows that the GPROF-NN retrievals
notably improve the effective resolution of the retrieval, i.e., increase its ability to reproduce the small-scale spatial variabil-
ity of precipitation. Given the considerable improvements afforded by the neural-network-based implementations of GPROF,
integrating these algorithms into the operational processing presents a simple but cost-effective opportunity to improve global
50   precipitation measurements.

The GPROF and GPROF-NN algorithms have in common that they are designed to reproduce precipitation estimates from a
dataset of reference precipitation measurements, the so-called *a priori database*. Climatologically, precipitation estimates from
GPROF and GPROF-NN will be similar to those of the a priori database. If the a priori database deviates from given validation
measurements, the resulting errors will be reproduced by the GPROF and GPROF-NN retrievals. Pfreundschuh et al. (2022)
55   evaluated the GPROF retrievals against held-out data from the a priori database. In this scenario, all retrieval errors are due

to the limited information content of the input observations and the limitations of the retrieval method. This study compares the GPROF retrievals to independent precipitation estimates and the error 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 is essential to

60   guide efforts to improve GPM PMW retrievals.

This study validates the newly-released operational version of GPROF and the GPROF-NN algorithms against independent, ground-based precipitation measurements. The questions that this study aims to address are:

1. What are the changes and improvements in GPROF V7 compared to GPROF V5?

2. To what extent contribute errors in the a priori database to GPROF retrieval errors?

65   3. Can the GPROF-NN retrievals improve precipitation estimates of the GPM PMW observations even when compared to independent measurements?

The validation is based on several years of collocations between different GPM sensors and ground-based precipitation radars over the conterminous United States (CONUS) and the Kwajalein atoll in the tropical Pacific. The first part of this study focuses on retrievals from GMI, for which a detailed evaluation of the retrieval accuracy from instantaneous to annual

70   time scales is performed. In addition, we analyze the effective resolution of the GMI retrievals to confirm the ability of the GPROF-NN retrievals to improve the effective resolution of GPROF, which has been found to be particularly low over land (Guilloteau et al., 2017). Finally, for the second part of the study, the retrieval accuracy of a selection of currently operational and historical sensors of the GPM constellation is assessed.

2   Data and methods

75 This investigation uses quantitative precipitation estimates (QPEs) from ground-based radars as principal means for validating the GPROF retrievals. While these measurements are themselves affected by non-negligible uncertainties, they offer the best compromise in terms of sensitivity to precipitation and ability to resolve the spatial structure of precipitation fields. Gauge-corrected radar estimates will be used where possible as they have been shown to reduce uncertainties (Kirstetter et al., 2012). Since the ground-based radar QPEs are used as independent measurements to validate the GPROF retrievals, we will refer to

80 them as *validation measurements*.

Due to the importance of the a-priori database for the GPROF retrievals, precipitation estimates from the a priori database and the GPM combined radar/PMW retrievals (upon which large parts of the a priori database are based) are included in the assessment of the GPROF retrievals. To distinguish these measurements from the validation measurements and since the GPROF retrievals aim to reproduce these estimates, they will be referred as *reference measurements*. Comparing the accuracy

85 of GPROF with respect to validation and reference measurements as well as the accuracy of the reference measurements with respect to the validation measurements will allow attributing retrieval errors to the observation system and retrieval method on one hand and the a priori database on the other.

2.1 Validation measurements

2.1.1 MRMS

90 Instantaneous 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 are produced specifically for GPM ground validation and are gauge-corrected to match monthly accumulations (Kirstetter et al., 2012). The processing includes a basic filtering that removes measurements with excessive gauge-correction factors. The data is provided on an approximately $0.01° \times 0.01°$ grid covering CONUS. The data is smoothed using a Gaussian average filter with a full-width at half-maximum (FWHM) of $5\,\mathrm{km}$
95 and interpolated to the $5\,\mathrm{km} \times 5\,\mathrm{km}$ collocation grid using nearest-neighbor interpolation.

The MRMS validation data includes a radar quality index (RQI), which quantifies the reliability of the MRMS measurements. It takes into account beam blocking, beam height and width, and the height of the melting layer. The RQI is mapped to the satellite-following grid using nearest neighbor interpolation. The precipitation mask is downsampled by applying a Gaussian averaging filter with a FWHM of $5\,\mathrm{km}$ to the binary mask of each class. The resulting fields are interpolated to the $5\,\mathrm{km} \times 5\,\mathrm{km}$
100 grid and each grid point is assigned the class with the highest mask value.

2.1.2 Kwajalein Polarimetric S-band Weather Radar

Validation data for the Pacific ocean stems from the Kwajalein Polarimetric S-band Weather Radar (KPOL). The lowest-elevation scan is used to estimate surface precipitation. A full $360°$ sweep consists of 360 scans, each of which comprises 779 radar bins with a range resolution of $200\,\mathrm{m}$. Rain rates are estimated using the CSU-HIDRO algorithm for polarimetric
105 radar observations (Cifelli et al., 2011). The radar has recently been calibrated using the self-consistency method proposed by Ryzhkov et al. (2005).

2.2 Reference measurements

2.2.1 GPM combined

Surface precipitation from version 7 of the GPM combined radar/radiometer product (GPM CMB, Grecu et al., 2016) accounts
110 for the major part of the surface precipitation in the GPROF a priori database over ocean and snow-free land surfaces. It is included in this validation as a proxy for the retrieval database in the years that are not covered by the a priori database.

2.2.2 A priori database

The GPROF retrievals rely on a common *a priori database*, which comprises PMW observations and corresponding hydrometeor profiles and precipitation rates. Over oceans and snow-free land surfaces, the database is derived from collocations of
115 GMI observations with corresponding surface precipitation and hydrometeor profiles from GPM CMB (Grecu et al., 2016). Although GPM CMB also incorporates PMW observations from GMI, the retrieval is primarily driven by measurements from the precipitation radar.

Over oceans, light precipitation from the currently operational Microwave Integrated Retrieval System (Boukabara et al., 2011) is added to pixels in which no precipitation is detected by the GPM CMB product. Over mountain regions, the pre-
120 cipitation from the combined product is increased to account for orographic effects. Over snow surfaces, the database uses collocations between GMI and MRMS, while precipitation over sea-ice is derived using collocations with the ERA5 reanalysis (Hersbach et al., 2020). The current GPROF retrieval database used by GPROF V7 and the GPROF-NN retrievals is based on observations from the water year 2019.

2.3 GPROF retrievals

125 2.3.1 GPROF V7

As of May 2022, the publicly available version of GPROF is V7. GPROF V7 adds several algorithm improvements over GPROF V5, the previous operational version. Most notably, new surface classes have been introduced to account for orographic effects. GPROF V5 is included in the validation as baseline retrieval against which GPROF V7 and the GPROF-NN algorithms will be assessed. Since GPROF V7 replaced GPROF V5 in the spring of 2022 as the operational version of the GPM PMW retrievals,
130 comparisons to GPROF V5 are only possible for the water years 2019, 2020 and 2021.

2.3.2 GPROF-NN

GPROF-NN (Pfreundschuh et al., 2022) is a novel, neural-network-based implementation of GPROF. The GPROF-NN retrieval has been designed as a drop-in replacement for GPROF V7 and is based on the same a priori database as GPROF V7. Two principal configurations of GPROF-NN exist, named GPROF-NN 1D and GPROF-NN 3D, which differ in their treatment of the
135 input observations. While GPROF-NN 1D retrieves precipitation independently for each observed pixel, GPROF-NN 3D uses all pixels in the swath simultaneously, thus enabling the retrieval to leverage structural information in the PMW observations. The implementation of both retrievals are available through a public code repository (Pfreundschuh, 2022).

2.3.3 GPROF-NN HR

All retrieval variables in the GPROF database are spatially smoothed to the expected resolution of the respective sensor. For
140 GMI, the expected resolution is $18.7\,\mathrm{km}$ in along track direction and $11\,\mathrm{km}$ in across track dimension, which corresponds to the FWHM of the $18.7\,\mathrm{GHz}$ channels. The higher frequency channels of GMI have smaller footprints and may thus allow for retrieving precipitation at higher resolution. However, the smoothing of the precipitation in the a priori limits on the resolution of the retrieved precipitation to that of the $18.7\,\mathrm{GHz}$ channels.

To explore the possibility of achieving higher resolution retrieval from GMI, we include an additional configuration of the
145 GPROF-NN retrieval, which is named GPROF-NN *high-resolution* (HR). GPROF-NN HR retrieves precipitation at a resolution of approximately $5\,\mathrm{km} \times 5\,\mathrm{km}$. It is trained directly on GPM CMB surface precipitation without any smoothing applied to the precipitation fields. Furthermore, it does not make use of the ancillary data used by GPROF V7 and the other two GPROF-NN versions described above. This has the advantage that the retrieval can be run directly on the level 1C-R observations from

GMI. The underlying neural network model uses the same architecture as the GPROF-NN 3D retrieval adapted to the reduced
150 number of inputs. The output resolution of $5\,\mathrm{km} \times 5\,\mathrm{km}$ is achieved through an additional upsampling block in the final part
of the decoder, which increases the resolution in the along-track direction by a factor of 3. The source code of the GPROF-NN
HR retrieval is published together with the other GPROF-NN retrievals (Pfreundschuh et al., 2022).

2.4 Collocations

For the comparison of validation and reference measurements and the GPROF retrievals, all data have to be collocated. To this
155 end, ground- and satellite-based results are remapped to a common, approximately equidistant $5\,\mathrm{km} \times 5\,\mathrm{km}$ grid that follows
the satellite swath. An equidistant grid is required for the analysis of the effective resolution. The grid resolution of $5\,\mathrm{km} \times 5\,\mathrm{km}$
was chosen finer than the expected resolution of the PMW-based GPROF retrievals in order to avoid interference of the spatial
sampling with the analysis of the effective resolution of the retrievals.

An example of the collocations used in this validation study is provided in Fig. 1. The scene displays an overpass of the GPM
160 core observatory over the mid-latitude cyclone associated with the 2020 Easter tornado outbreak (Kerr and Alsheimer, 2022).
The various panels show surface precipitation estimates from the validation and reference measurements, and the different
GPROF retrieval results.

Four water years (October through September) from 2019 until 2022 are used for the validation of the GPROF retrievals from
GMI. This time period was chosen because it includes the water year 2019, which is used to derive the a priori database for the
165 GPROF retrievals. The analysis is extended to include several years in order to allow assessment of the temporal variability of
the retrieval accuracy.

For the other sensors of the GPM constellation, the retrieval accuracy was assessed using a single year of collocations. The
collocation periods were chosen according to the availability of observations and ground-validation data. The periods chosen
for the different sensors are listed in Table 1.

Sensor	Platform	Collocation period
TMI	TRMM	2013-03-01 — 2014-03-01
SSMIS	F17	2015-01-01 — 2016-01-01
AMSR2	GCOM-W1	2018-10-01 — 2019-10-01
MHS	NOAA-19	2018-10-01 — 2019-10-01

Table 1. Other GPM sensors and corresponding collocation periods used to evaluate the GPROF retrievals.

2.5 Analysis of the effective resolution

The distance between neighboring pixels of the GMI swath is approximately $13.5\,\mathrm{km}$ in along-track direction and $5\,\mathrm{km}$ in
across-track directions at the center of the swath. Since GPROF traditionally retrieves precipitation at the same grid as the input
observations, the resolution of the retrieved precipitation fields cannot exceed the spatial sampling of the input observations. In
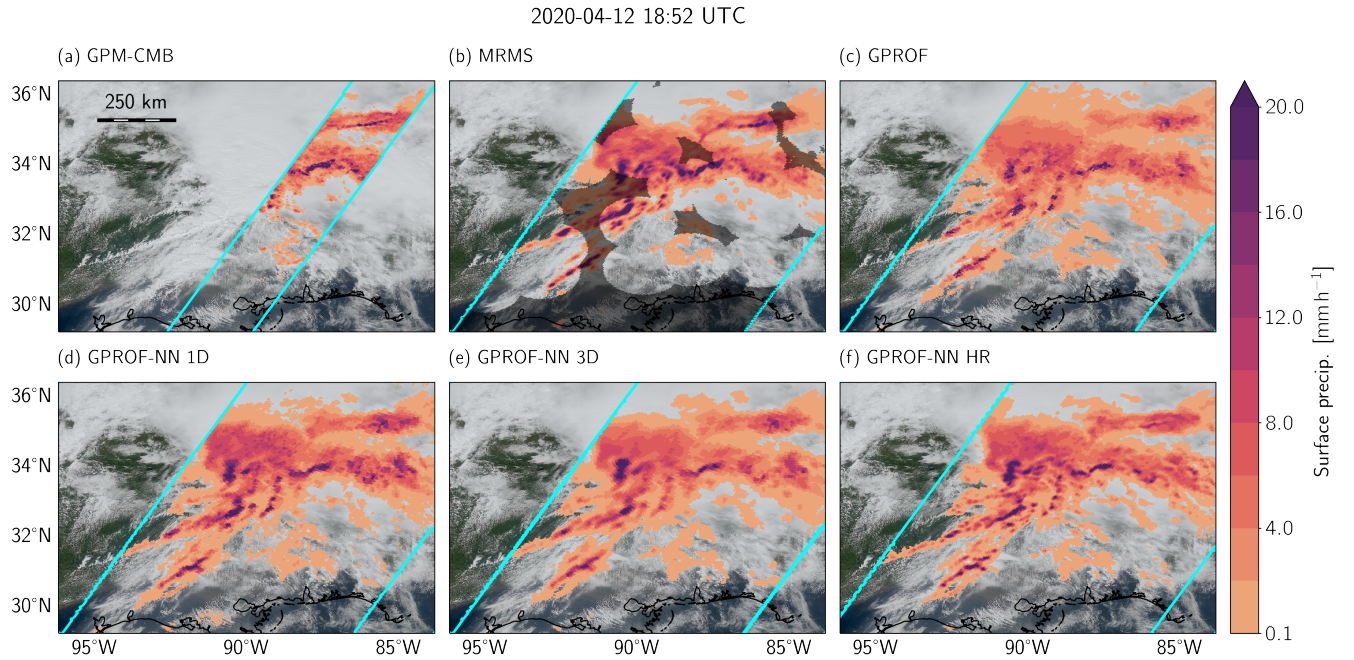
Figure 1. Sample scene from the collocation dataset used to validate the GPROF retrievals displaying an overpass of the GPM core observatory over the 2020 Easter Tornado outbreak. Results are shown for the GPM CMB product (Panel (a)), gauge-corrected MRMS QPEs from ground-based radars (Panel (b)), the GPROF V7 retrieval (Panel (c)), the GPROF-NN 1D retrieval (Panel (d)), the GPROF-NN 3D retrieval (Panel (e)) and the GPROF-NN HR retrieval (Panel (f)). Cyan lines in Panel (a) mark the boundaries of the DPR, while the cyan-colored lines in the remaining panels mark the boundaries of the GMI swath. Dark shading in Panel (b) marks measurements with a RQI lower than 0.8.

general, however, not all variability at scales smaller than the spatial sampling of the output data is truthfully represented in the
175   retrieval results. This motivates the definition of an effective resolution as a numerical measure of retrieval's ability to resolve small-scale spatial variability.

The consistency between retrieved and reference precipitation fields across different spatial scales can be quantified using the relative spatial Fourier spectral coherence between reference and retrieved fields. The spectral coherence can be interpreted as the complex correlation between the Fourier coefficients of two signals within a narrow frequency (or wavenumber) band.
180   In two dimensions, the Fourier spectral coherence between two fields $g(x, y)$ and $h(x, y)$ is defined as:

$$C_{g,h}(k_x, k_y) = \frac{P_{g,h}(k_x, k_y)}{\sqrt{P_g(k_x, k_y) \times P_h(k_x, k_y)}} \tag{1}$$

with k_x and k_y the wavenumbers in the two orthogonal directions x and y. $P_g(k_x, k_y)$ is the Fourier power spectral density (PSD) of $g(x, y)$, defined as:

$$P_g(k_x, k_y) = \hat{g}(k_x, k_y) \times \hat{g}^*(k_x, k_y) \tag{2}$$

with $\hat{g}^*$ denoting the complex conjugate of the Fourier coefficient field $\hat{g}$. $P_{g,h}(k_x, k_y)$ is the Fourier cross power spectral density (CPSD) between $g(x,y)$ and $h(x,y)$ defined as:

$$P_{g,h}(k_x, k_y) = \hat{g}(k_x, k_y)\hat{h}^*(k_x, k_y) \tag{3}$$

The computation of the PSD and CPSD is performed through a discrete cosine transform (DCT) and averaged over the evaluated retrieval scenes.

Passing into the polar coordinate system through the change of variables $k = \sqrt{(k_x^2 + k_y^2)}$ and $\theta = \text{atan}(\frac{k_x}{k_y})$, the spectral quantities can be analyzed as a function of the spatial wavenumber k and the azimuthal direction θ. Since we did not find any significant dependency on the azimuthal direction, we integrated the spectral quantities along θ to analyze the coherence as a function of k only.

The effective resolution of the retrieved fields can then be defined from the spectral coherence between retrieved and reference field as

$$\text{ER} = \frac{1}{2}\arg\max_k C_{g,h}(k) \geq \frac{1}{\sqrt{2}}, \tag{4}$$

i.e. half of the shortest wavelength for which the spectral coherence is larger than $\frac{1}{\sqrt{2}}$. Defined in this way, the effective resolution corresponds to the smallest spatial scale at which the signal to noise ratio of the retrieved precipitation fields exceeds 1.

3 Results

The assessment of the GPROF retrieval against ground-based radar is split into three parts. The first part analyzes the accuracy of the GMI-based precipitation estimates over CONUS and the Kwajalein site. The second part assesses the effective spatial resolution of these retrievals. Finally, the third part evaluates the retrieval accuracy for the sensors of the GPM constellation of PMW radiometers for which MRMS ground-validation data is available

3.1 Retrieval accuracy of GPROF for GMI

The validation of the GPROF GMI retrievals uses collocations from October 2018 until October 2022. The first year of this period coincides with the a priori database. For the comparison against MRMS data only data points with an RQI of at least 0.8 were used. The validation is performed at the $5\,\text{km}$ resolution of the collocations. We choose to evaluate the retrieval at this resolution instead of the footprint size of the PMW observations because it makes accuracy metrics between different sensors comparable.

Snow-covered and mountain surfaces are excluded from the validation. For the former, reference precipitation in the retrieval database is derived from collocations with MRMS while, for the latter, a correction for orographic precipitation is applied. These modification aim to counteract known weaknesses in the GPM CMB reference data but they would skew the comparison between GPM, GPM CMB and MRMS. Furthermore, precipitation that is identified as snow by MRMS is excluded from the validation.

3.1.1 CONUS

Distributions of retrieved and reference precipitation conditioned on the validation precipitation are displayed in Fig 2. The first column of panels shows the distribution of retrieved precipitation with respect to the GPROF a priori database. The spread in these distributions is due to the limitations of the retrieval method and the ill-posed character of the retrieval and they thus represent the best-case accuracy of each retrieval. Some spread is observed even between GPM CMB and the a priori database, which is due to the spatial smoothing applied to the precipitation in the a priori database.

The spread in all GPROF retrievals is significantly larger when compared to MRMS. However, even the GPM CMB retrieval shows significant deviations from the MRMS measurements. From the GPROF retrievals, V5 and V7 exhibit the largest spread, while there is less spread in the distributions of the GPROF-NN 1D and 3D retrievals. Compared to the 1D and 3D variants, the GPROF-NN HR retrieval further decreases the spread. The distributions corresponding to different validation periods are very similar for all retrievals. This indicates that, conditioned on the reference precipitation, the accuracy of each retrieval exhibits little inter-annual variability.

Maps of the retrieval biases are displayed in Fig. 3. When evaluated against the latest GPROF a priori database, GPROF V5 exhibits the largest biases. Since GPROF V5 is based on a different a priori database, whose climatology may deviate from the V7 a priori database, this is expected. However, even GPROF V7 exhibits noticeable dry biases along the east and west coasts and a weaker wet bias over the central CONUS. GPROF-NN 1D exhibits a weaker dry bias over western CONUS and the South-East and a wet bias in the North-East. The biases of the GPROF-NN 3D and GPROF-NN HR retrievals are even weaker but exhibit a spatially more consistent tendency to overestimate precipitation. Since the retrieval database is derived from the GPM CMB retrieval, the GPM CMB retrieval is practically bias-free compared to the retrieval database.

Significantly larger biases are observed when the precipitation is compared to MRMS measurements. For the year 2019, the GPM CMB retrieval exhibits a wet bias over the western Great Plains and a weak dry bias over the remaining CONUS. A similar pattern is observed for all the GPROF retrievals, indicating that the bias patterns of the GPROF retrievals result from the biases in the a priori database.

Dry biases in the South-East and North-East persist throughout all considered years. Over the western CONUS, however, the retrievals are biased dry during years 2019 and 2020 but biased wet in the years 2021 and 2022. A potential explanation for this change in the bias patterns is a version change in the MRMS processing that occurred in February 2021 (NOAA National Severe Storms Laboratory, 2023). The spatial distributions of retrieval biases for the GPROF V7 and GPROF-NN retrievals largely follow the biases of GPM CMB also for the years 2020, 2021 and 2022. Given that the CMB over land is largely a radar-derived product, it is possible that the bias relative to MRMS is introduced rain gauge correction applied to the validation data and caused by precipitation properties that may not be resolved by the radar observations.

At the $5\,\mathrm{km}$ resolution considered here, biases of GPROF V7 and the GPROF-NN retrievals are well correlated with those of GPM CMB and exhibit similar variability throughout the assessed years. A potential explanation for this are weather-regime-dependent errors in the GPM CMB retrieval with respect to the validation measurements. The GPROF retrievals reproduce these errors, which leads to similar responses to the inter-annual variability of the occurrence of these weather regimes.
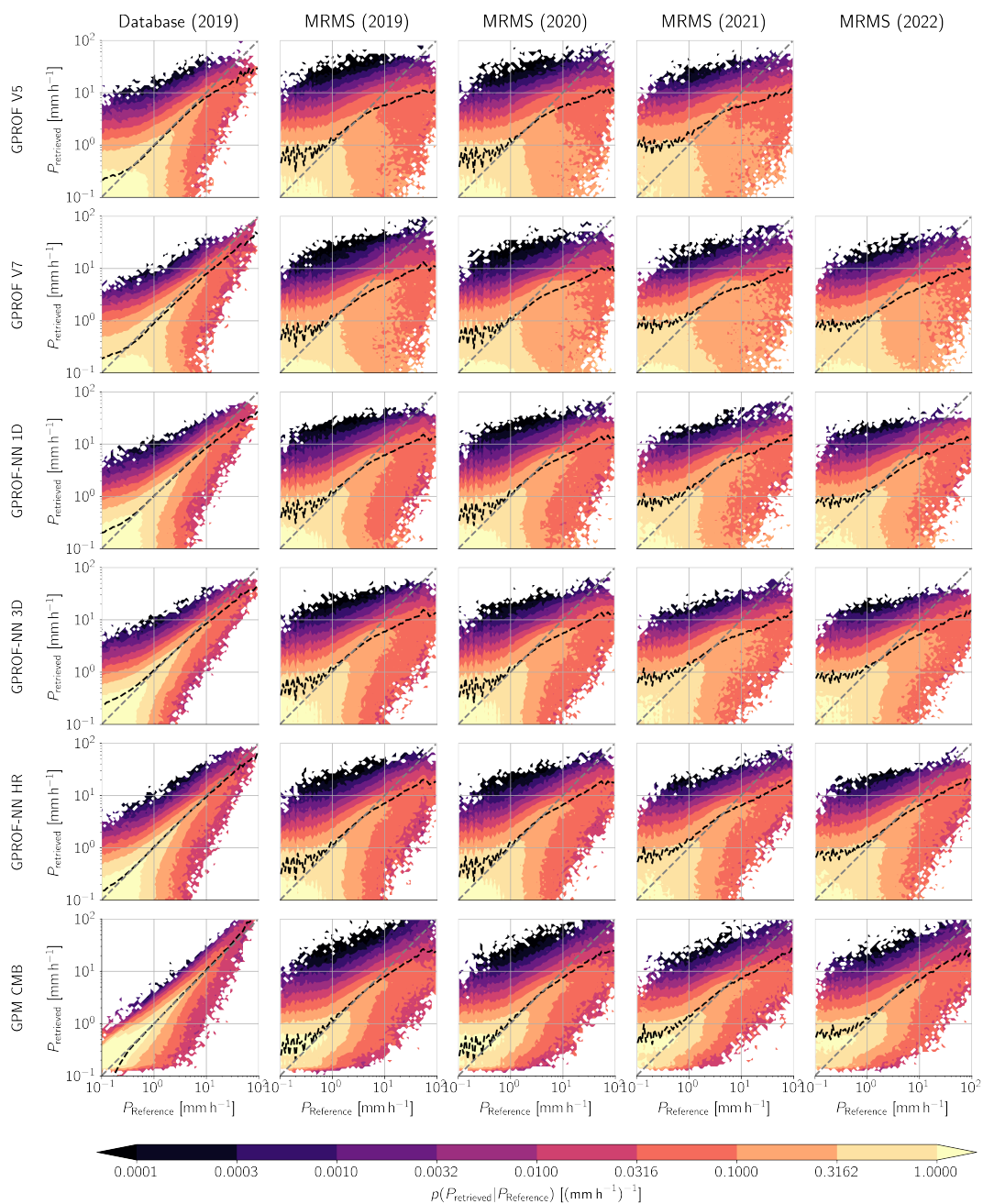
Figure 2. Distributions of retrieved precipitation conditioned on reference precipitation. Panels in the left-most column show the distributions for all retrievals assessed against the GPROF a priori database. The remaining columns show the respective distributions for the water years 2019, 2020, 2021, 2022 with respect to the MRMS validation QPEs. The black dashed line in each panel shows the conditional mean of the retrieved precipitation.
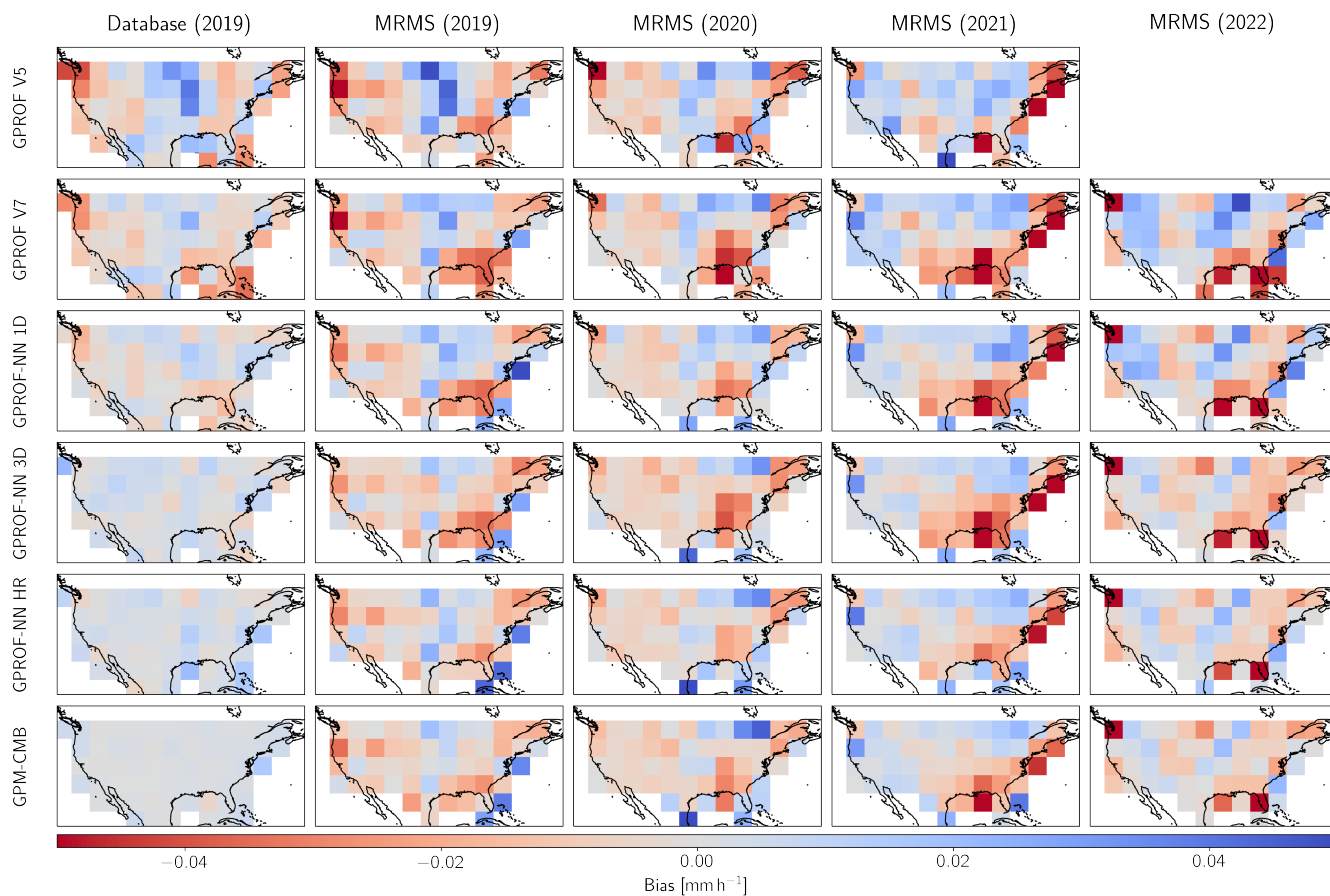
Figure 3. Spatial distribution of retrieval biases. Each panel shows the biases calculated over a uniform longitude-latitude grid with a grid size of $5°$. Rows show the results for the different retrieval algorithms. The left-most column shows the biases with respect to the a priori database. Remaining columns show the biases with respect to the MRMS validation data for each of the assessed water years.

250   Maps of correlations between reference and retrieved precipitation are displayed in Fig. 4. Since GPROF V5 is based on a different a priori database, it has the lowest correlation with respect to the current database. The correlation of GPROF V7 is highest in the western Central Plains and there even surpasses that of GPROF-NN 1D. However, GPROF-NN 1D achieves higher correlations than GPROF V7 in the remaining regions. Both GPROF-NN 3D and GPROF-NN HR achieve consistently higher correlations than GPROF-NN 1D. GPM CMB exhibits the highest correlation with the a priori database but remains

255   below 1.0 due to the smoothing applied to the surface precipitation in the a priori database.

    All retrievals have lower correlation with the MRMS QPEs than with the retrieval database. In part, this reduction is due to the higher resolution of the MRMS validation precipitation compared to the reference precipitation in the database. In addition to this, deviations of the a priori database from the MRMS validation data contribute to the reduced correlation. The reduction in correlation is strongest for the GPROF V7 retrievals which exhibit only slightly higher correlations than V5 when evaluated

against the ground-radar measurements. For the conventional GPROF retrievals correlations are between 0.4 and 0.6 over most of CONUS. The correlations of GPROF-NN 1D are consistently higher, ranging from 0.5 to 0.7 in most regions. The correlations of GPROF-NN 3D are, again, slightly higher than those of GPROF-NN 1D and GPROF V7.

Surprisingly, the correlations of the GPROF-NN HR retrieval exceed those of GPM CMB. This is certainly remarkable, given that the GPROF-NN HR retrieval is trained on GPM CMB data. A possible explanation for this are random errors in the GPM CMB retrieval that are smoothed out in the passive-only retrievals.
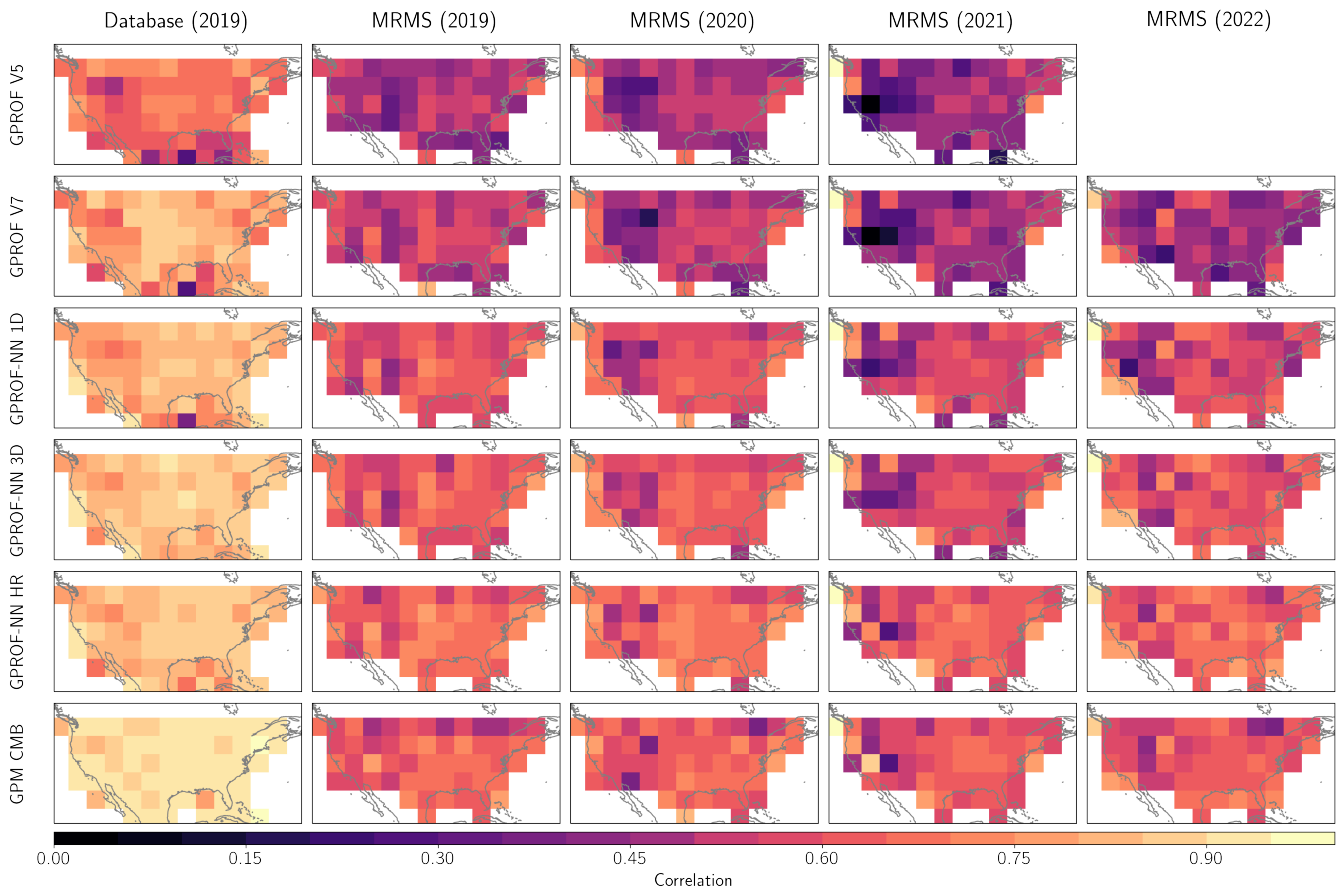
Figure 4. Like Fig. 3 but for the correlation coefficient between retrieved surface precipitation and validation data.

A quantitative analysis of the retrieval accuracy is presented in Fig. 5. The assessed metrics include the mean error (Bias), mean-squared error (MSE), mean absolute error, correlation coefficient and the symmetric mean absolute percentage error (SMAPE$_t$), which is defined as:

$$\text{SMAPE}_t(P, P_{\text{MRMS}}) = \frac{1}{N_{P_{\text{MRMS}} \geq t}} \sum_{P_{\text{MRMS}} \geq t} \frac{|P - P_{\text{MRMS}}|}{0.5(|P| + |P_{\text{MRMS}}|)} \} \tag{5}$$
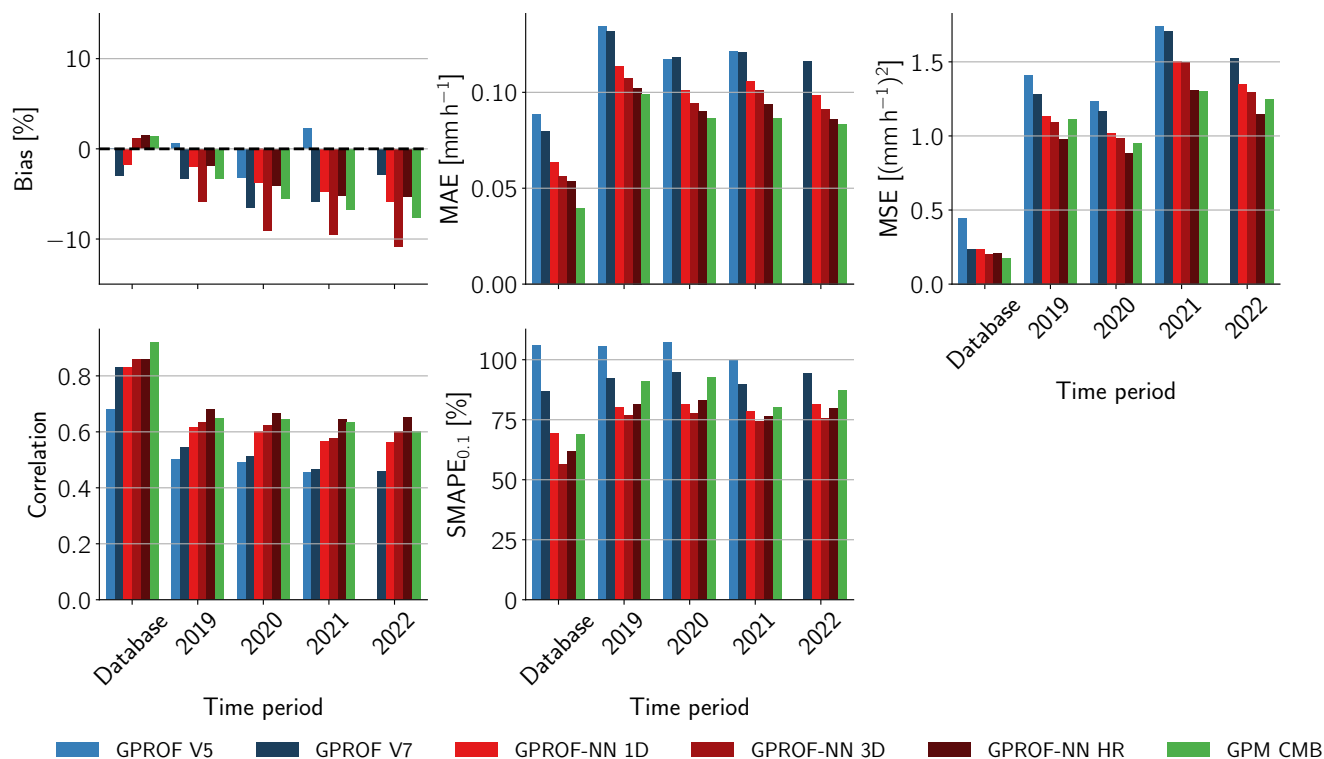
Figure 5. Accuracy metrics of the GPROF retrievals compared to the GPROF a priori database and MRMS.

270   When compared to the database, all retrievals exhibit weak biases of the order of a few percent. When compared to MRMS, all retrievals except GPROF V5 exhibit a dry bias. The dry bias of GPM CMB increases from around $3\,\%$ for the water year 2019 to around $8\,\%$ in 2022. The GPROF-NN retrievals follow the tendency of the biases of GPM CMB with the 3D variant exhibiting a slightly stronger dry bias while the 1D and HR variants exhibit a slightly weaker dry bias. The biases of GPROF V7 follow the biases of GPM CMB for the years 2019, 2020, 2021 but decrease in 2022. GPROF V5 is closer to MRMS than
275   the other retrievals for the years 2019, 2020, and 2021.

    In terms of MAE, a consistent ranking of the retrievals can be observed. GPROF V5 has the largest MAE, closely followed by GPROF V7. GPROF-NN 1D, 3D, and HR yield consecutively lower MAEs and GPM CMB the lowest MAEs of all retrievals. Although the MAEs increase significantly when the retrievals are compared to MRMS measurements, the relative accuracy of the retrievals remains the same. Similar results are found for the MSE with the exception that the GPM CMB retrieval has
280   higher MSE than the GPROF-NN HR retrieval.

    The results for the correlation reflect the general findings from Fig. 4. With respect to the a priori database, GPROF V5 has the lowest correlation. The correlations of GPROF V7 and GPROF-NN 1D are almost identical and exceed those of GPROF V5 by about 0.15. Both GPROF-NN 3D and GPROF-NN HR have slightly higher correlations and GPM CMB achieves the highest

correlations with the database precipitation. When compared to MRMS, the correlation of GPROF V7 decreases noticeably

285 making it more comparable to the correlations of GPROF V5. GPROF-NN 1D and GPROF-NN 3D remain fairly close in terms of correlation. GPROF-NN HR yields the highest correlation with the MRMS validation data even surpassing GPM CMB.

In terms of SMAPE, the accuracy of the GPROF-NN HR and GPM CMB retrievals is lower compared to the other retrievals. For GPM CMB, the SMAPE is higher than any of the GPROF-NN retrievals while the SMAPE of the GPROF-NN HR retrieval is higher than that of the GPROF-NN 3D retrieval. The reduced retrieval accuracy of GPM CMB and GPROF-NN HR in terms

290 of SMAPE may be caused by the SMAPE excessively penalizing the misplacement of precipitation, which may put retrievals with a higher resolution at a disadvantage.

Finally, we evaluate the skills of the GPROF retrievals in detecting precipitation. To this end, precision-recall curves for the detection of precipitation from the a priori database and MRMS validation data are displayed in Fig. 6. The curves display the trade-off between precision, i.e., the fraction of confirmed raining pixels among those retrieved as raining, and the recall,

295 i.e., the fraction of confirmed raining pixels that are detected by the retrieval. By varying the detection threshold applied to the retrieved probability of precipitation, a curve of precision and recall values is obtained that fully characterizes the detection skill of each retrieval. As was to be expected from the previous results, the detection skill decreases markedly when the retrievals are evaluated against MRMS. Nonetheless, also for the detection of precipitation consistent improvements between the conventional GPROF retrieval, GPROF-NN 1D, and GPROF-NN 3D can be observed. The GPROF-NN HR retrieval yields

300 worse detection skill than GPROF-NN 3D on the database, which is likely a consequence of the reduced spatial resolution of the precipitation in the a priori database. When compared to MRMS measurements at $5\,\mathrm{km}$ resolution, the HR variant clearly outperforms GPROF-NN 3D.

3.1.2 Regional precipitation characteristics

This section assesses the retrieval accuracy over five regions within CONUS and the K-POL radar in the tropical Pacific in

305 order to assess the ability of the retrievals to represent regional variations in precipitation patterns. The location and extent of the five CONUS regions is displayed in Fig. 7. The five regions encompass the Pacific North-West (NW), the South-West (SW), central CONUS (C), the South-East (SE) and the North-East (NE).

The error metrics for the retrievals in the six regions are shown in Fig. 8. For both the database and the MRMS collocations the regional biases are generally larger in magnitude than they are for the full CONUS. With the exception of GPROF V5, the

310 biases of the GPROF retrievals remain within $\pm 25\,\%$ indicating that they reproduce the regional climates reasonably well. The biases of the GPM CMB product are of the same sign and similar magnitude as those of the GPROF retrievals, indicating that the GPROF biases are inherited from the a priori database.

For the central region, the GPM CMB retrieval exhibits the smallest biases ($< 5\,\%$). While GPROF V5 exhibits a consistent wet bias, the other retrievals exhibit biases close to that of GPM CMB. For the other CONUS regions, GPM CMB exhibits dry

315 biases of up to $15\,\%$. Also here, except for GPROF V5, the other GPROF retrievals mostly follow the GPM CMB results. The biases are persistent over the four evaluation years with the exception of the year 2021 in which the dry biases in the GPM CMB retrieval disappear. Over the Kwajalein site, the GPM CMB retrieval exhibits considerable wet biases for the years 2020,
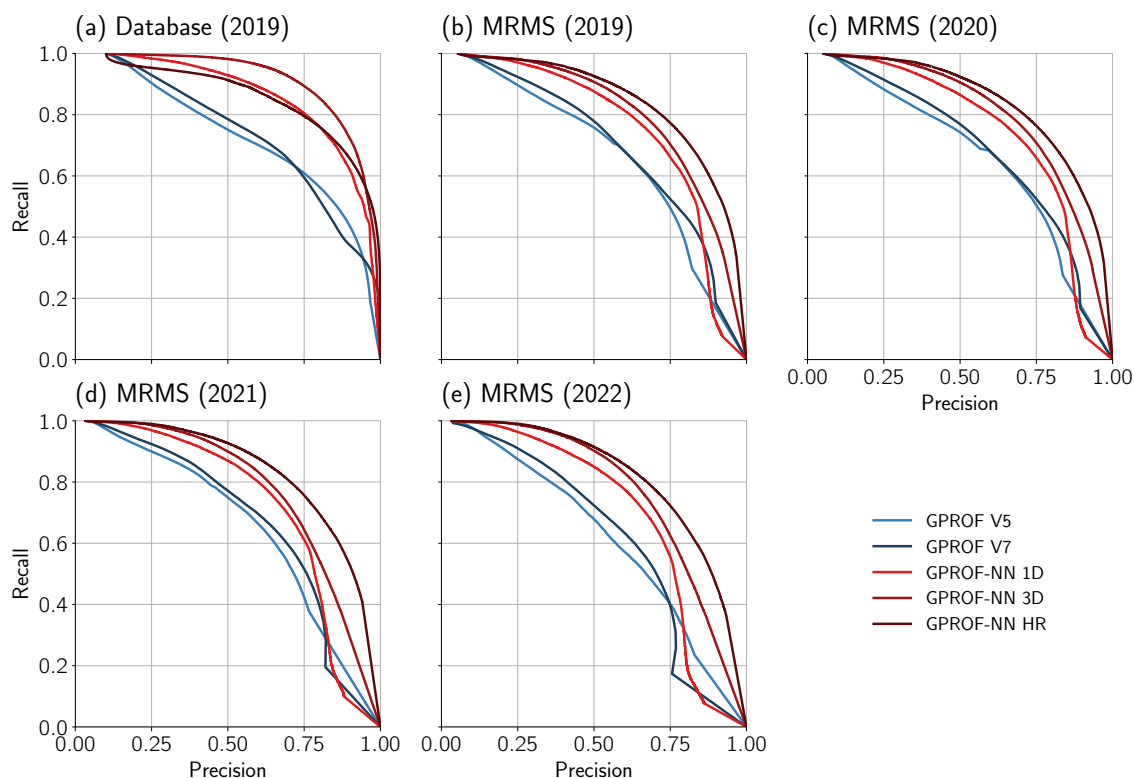
Figure 6. Precision-recall curves for the GPROF retrievals. Panel (a) shows the detection skill for the database collocations. Remaining panels show the results with respect to the MRMS measurements for the years 2019, 2020, 2021, and 2022.

2021, and 2022. The GPROF retrievals exhibit comparable biases with the exception of GPROF V5, which yields the smallest biases compared to the validation data. GPROF V5 being drier than GPROF V7 and the GPROF-NN retrievals is consistent

320 with a marked increase in of ocean precipitation in GPM CMB V7 compared the previous version upon which the GPROF V5 database was based.

Both MAE and MSE vary significantly in magnitude between the different regions with larger errors occurring in regions with more precipitation. Nonetheless, the relative accuracy of the different retrievals is similar to the results from the whole CONUS (Fig. 5): GPROF V7 is slightly more accurate than GPROF V5, GPROF-NN 1D is more accurate than GPROF V7,

325 GPROF-NN 3D more accurate than GPROF-NN 1D, and GPROF-NN HR more accurate than GPROF-NN 3D. This ranking is also reflected in the correlation and SMAPE.
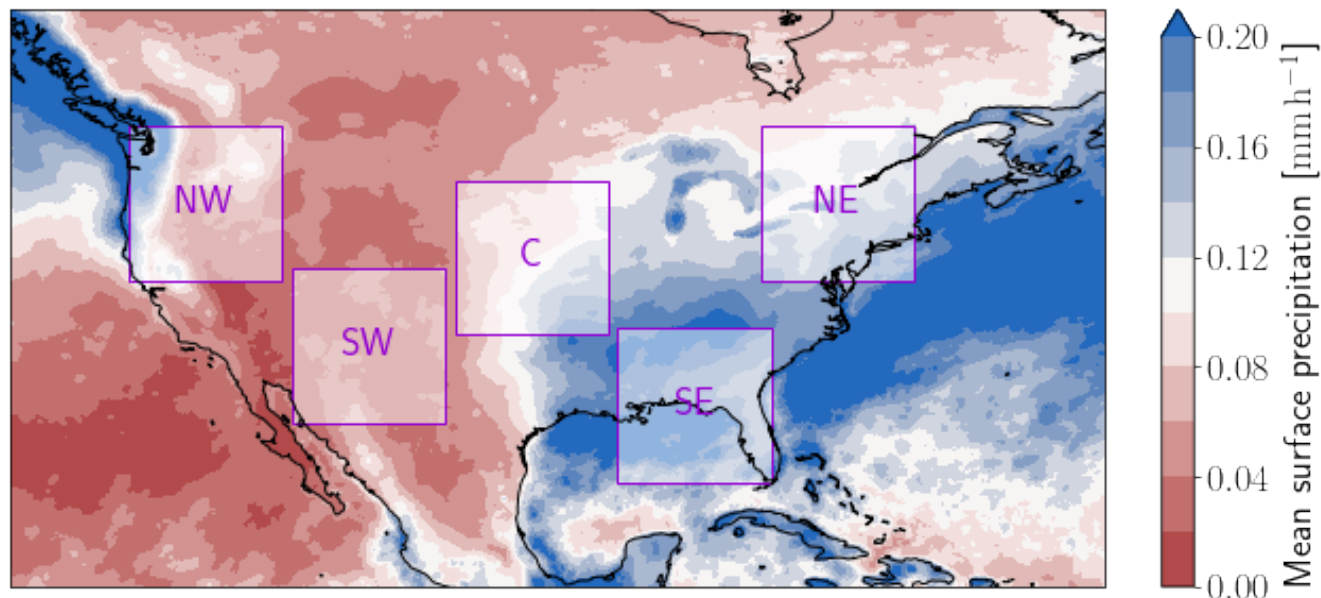
Figure 7. IMERG mean precipitation rate over the validation period Oct. 2018 - Sep. 2021. The purple rectangles mark the five regions used for the evaluation of the regional retrieval accuracy.

Seasonal cycles of precipitation are displayed in Fig. 9. The seasonal cycles were calculated using all collocations from the years 2020, 2021. Collocations from the year 2019 were excluded because it is part of the retrieval database while the year 2022 was excluded because no GPROF V5 data is available for that year.

GPM CMB and GPROF generally capture the seasonal cycles of the regions accurately. Notable deviations from the MRMS reference precipitation are an underestimation of stratiform winter precipitation in the North-West as well as an underestimation of convective summer precipitation in the South-East.

The largest differences between GPROF V5 and V7 are found for warm-season precipitation over the central CONUS, where GPROF V5 exhibits a strong wet bias, and cold-season precipitation in the NW, where GPROF V5 exhibits a dry bias. Weaker deviations are observed in the South-East, where GPROF V5 captures the warm-season precipitation peak better, and the tropical Pacific, where GPROF V5 retrieves less precipitation during the second half of the year, which puts it closer to the validation measurements.

The results of GPM CMB and the GPROF retrieval are very similar for most regions and months with the exception of summer precipitation in the South-West. In this region, GPROF V7 exhibits the strongest dry bias with respect to both the reference data and GPM CMB retrieval. The GPROF-NN 1D and 3D retrievals yield results closer to the reference precipitation, while the GPROF-NN HR retrieval slightly overestimates the precipitation. This may indicate that the neural-network-based retrievals exhibit greater skill in estimating summer-time convective precipitation in this region.
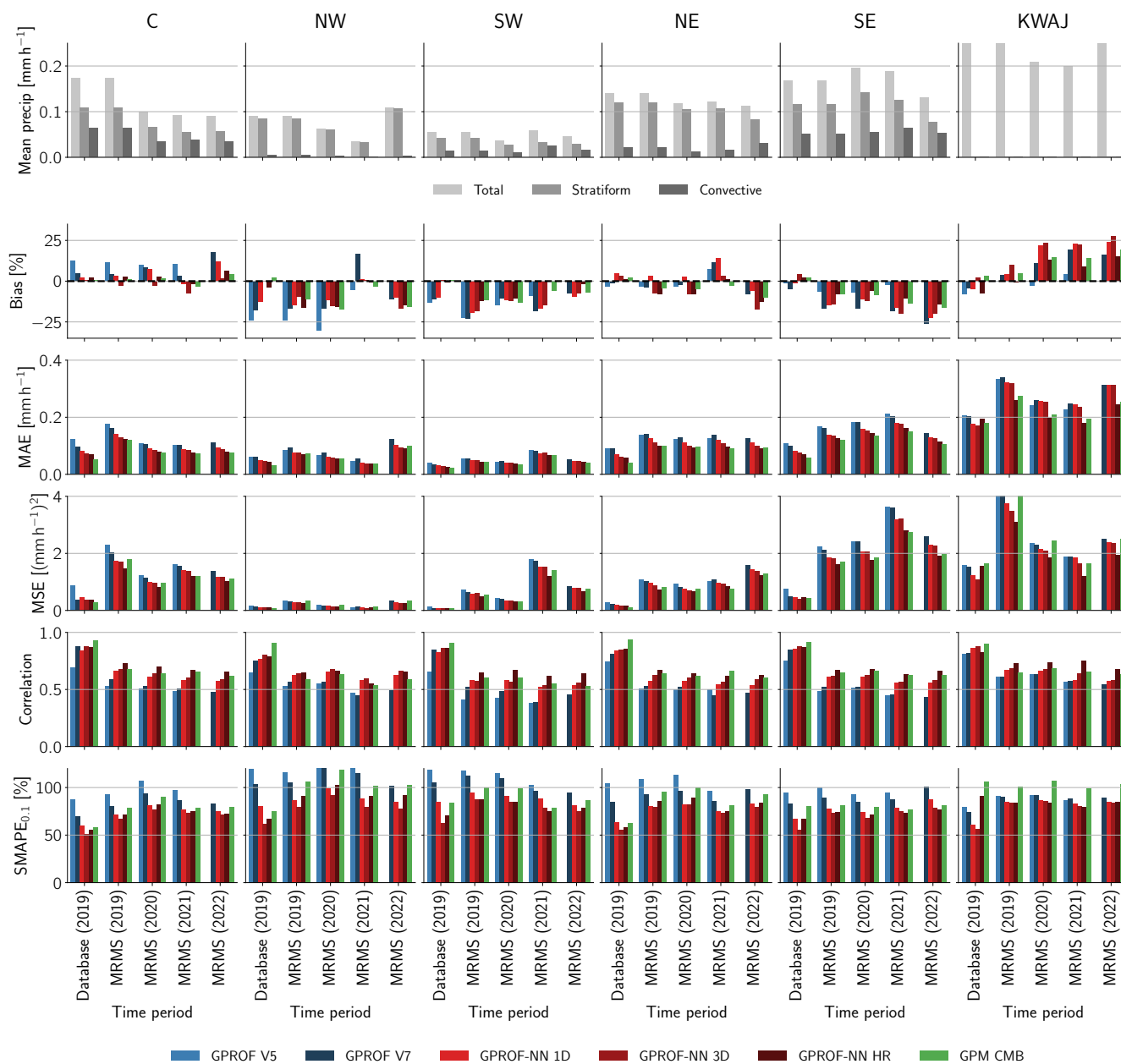
Figure 8. Retrieval accuracy metrics for the assessed climate regions over CONUS and the tropical Pacific. The first row shows the mean precipitation rate of the reference data and, where such a classification was available, the contributions from convective and stratiform precipitation. Remaining rows show accuracy metrics for the comparison of the retrieval results against the reference data. Columns from left to right show the results for the for the five considered regions.
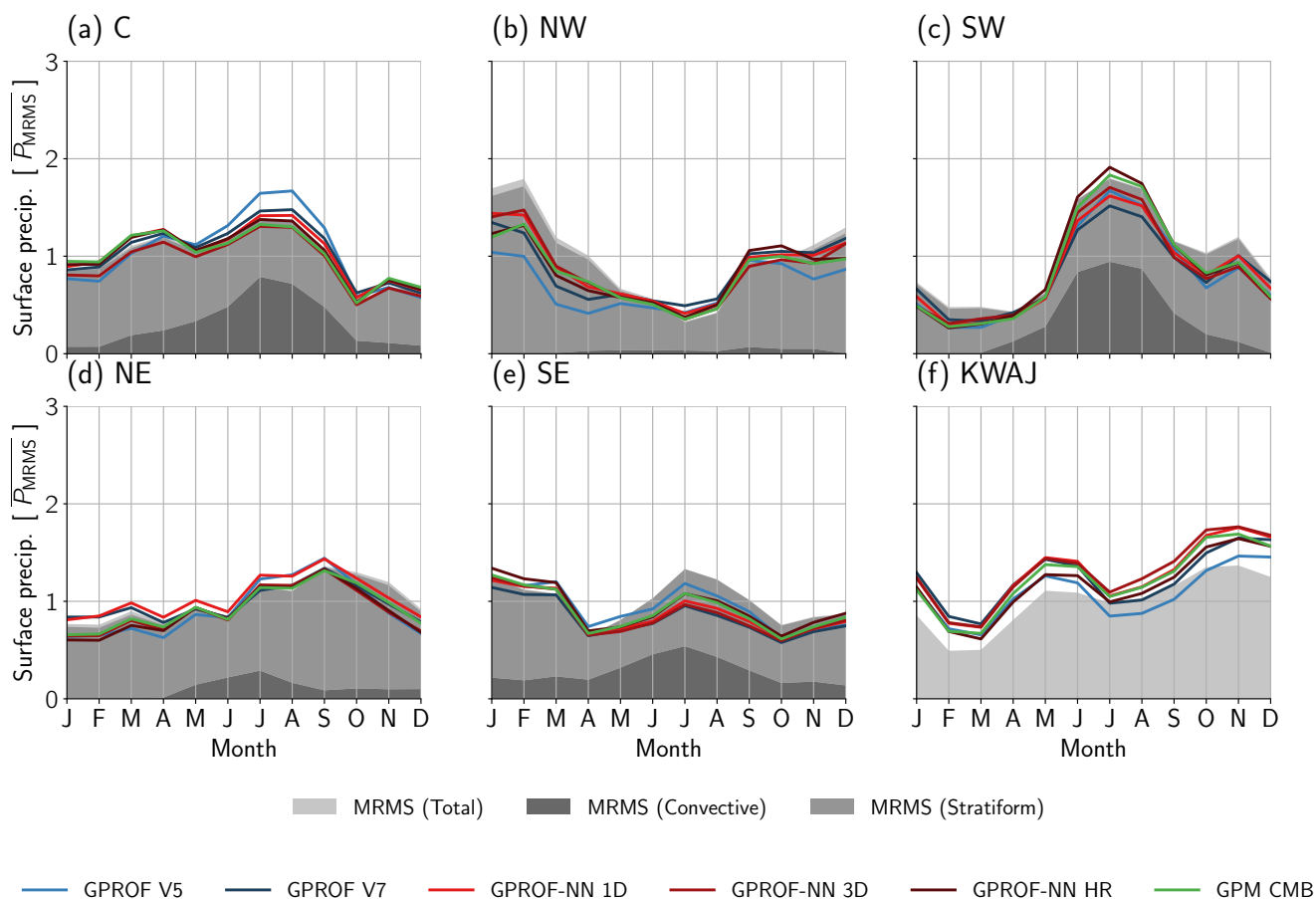
Figure 9. Regional seasonal cycles of precipitation. Each panel shows the seasonal cycles derived from GPROF retrievals, GPM CMB and the validation data for one of the considered regions. The seasonal cycles are normalized by the corresponding mean precipitation rate of the validation precipitation. The seasonal cycles were smoothed using a sliding window average with a width of three months. Only collocations from the water years 2020 and 2021 are used.

Diurnal cycles of precipitation shown in Fig. 10 reveal the diurnal variations of the regional deviations observed above. The underestimation of precipitation in the North-West is found to be related to underestimation of the precipitation during early

345 morning and afternoon. In the South-East, the underestimation is related to convective afternoon precipitation.

The deviations between the GPROF retrievals for summer precipitation in the South-West can also be observed in the diurnal cycles. In the South-West, GPROF V7 most strongly underestimates precipitation during afternoon peaks. GPROF V7, but also GPROF-NN 1D, underestimate precipitation most strongly during its onset in the afternoon but catch up to the other retrievals later during the day.

350 The diurnal cycles reveal further differences between the retrievals for convective precipitation in the South-East. Here, GPROF V7 underestimates the convective precipitation most-strongly but also GPROF-NN 1D and 3D retrieve less precipitation than GPM CMB. In contrast to the South-West, the underestimation is more homogeneous across the duration of the precipitation peak.
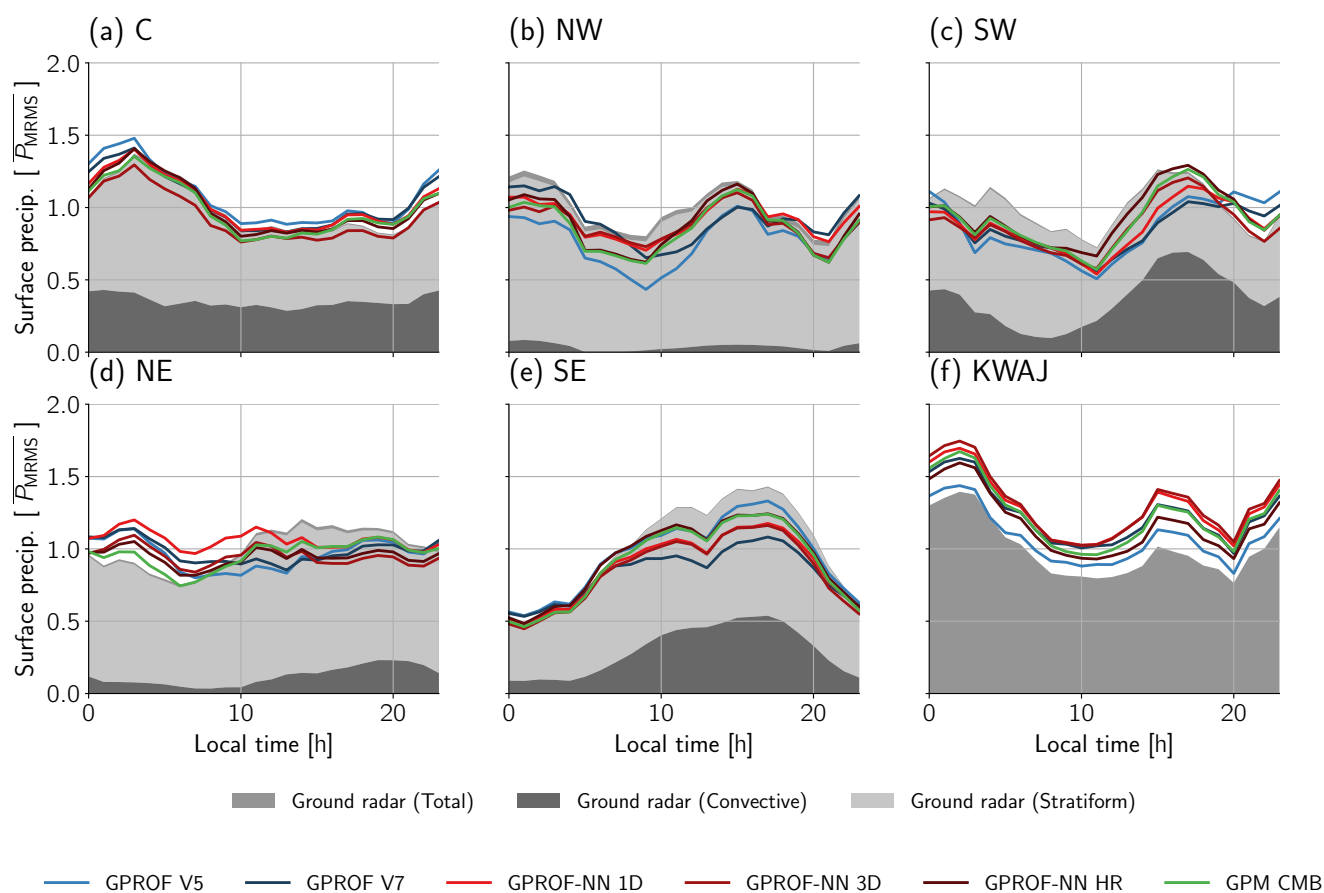
Figure 10. Like Fig. 9 but for diurnal cycles of precipitation. The curves were smoothed using sliding window average with a width of 3 hours.

3.2 Effective resolution

355   For the analysis of the effective resolution of the retrievals, the discrete cosine transform of the retrieved and reference precipitation fields were calculated over non-overlapping windows of size $160\,\mathrm{km} \times 160\,\mathrm{km}$ in which neither any of the GPROF retrievals nor the ground radar measurements had more than $2\,\%$ of missing values. Over CONUS only windows with a minimum radar-quality index of 0.5 were considered. Only collocations from the water years 2020 and 2021 were used in order to exclude scenes which are also present in the GPROF a priori database. The coherence was computed once over the full GMI
360   swath, and once only over the central part of the swath in order to allow comparison to the GPM CMB product.

The spectral coherence between the GPROF retrievals and the reference radar data as a function of the wavelength is displayed in Fig. 11. Over the CONUS, GPROF V7 has slightly higher spectral coherence than GPROF V5. The GPROF-NN retrievals achieve considerably higher coherence than the conventional GPROF retrievals across all wavelengths. The coherence curves of the GPROF-NN retrievals are similar up to spatial scales of about $15\,\mathrm{km}$, above which GPROF-NN 3D achieves
365   slightly higher coherence than GPROF-NN 1D. The coherence curves of the conventional GPROF retrievals and the GPROF-NN 1D and 3D retrievals are not significantly affected by the swath coverage.

The GPROF-NN HR retrieval achieves highest coherence of the PMW retrievals, even matching that of the GPM CMB retrieval, when it is evaluated only over the swath center. However, when evaluated over the full swath, the coherence decreases sharply and falls below that of the GPROF-NN 1D retrieval. This indicates that the retrieval overfits on the viewing geometry
370   of the central part of the GMI swath, which is the only part of the swath for which reference measurements of precipitation are available. Although the training of the GPROF-NN HR retrieval applies data augmentation that simulates the effect of perspective distortion across the swath, it does not account for the parallax effect, i.e., the displacement between observed precipitation signal and the actual precipitation caused by the slant viewing angle of the sensor. Since the parallax effect changes direction towards the edges of the swath, it can potentially explain the decrease in accuracy observed for the GPROF-
375   NN HR retrieval.

Over the Kwajalein site, all GPROF retrievals yield significantly improved coherence curves due to the direct relationship between microwave signals and the liquid water content in the clouds. Also here, the GPROF-NN 1D and 3D retrievals yield higher coherence than the conventional GPROF retrievals albeit by a smaller margin. The GPROF-NN HR retrieval exhibits similar overfitting as over the CONUS. Evaluated over only the central part of the swath, the coherence curve of the GPROF-
380   NN HR is higher than that of every other retrieval. Evaluated over the full swath, however, the accuracy decreases below that of the GPROF-NN 1D retrieval. Interestingly, the coherence curve of the GPM CMB retrieval remains at the level of the GPROF-NN HR retrieval evaluated across the full swath and falls below the curves of the GPROF-NN retrievals for scales exceeding $12\,\mathrm{km}$. This is an unexpected result since the GPROF-NN HR retrieval is designed to reproduce the results from the GPM CMB retrieval. A potential explanation could be that the GPM CMB retrieval systematically misses precipitation from shallow
385   clouds while the PMW retrieval is able to recover some of that precipitation because the signal in the PMW observations is less dependent on the depth of the observed clouds.
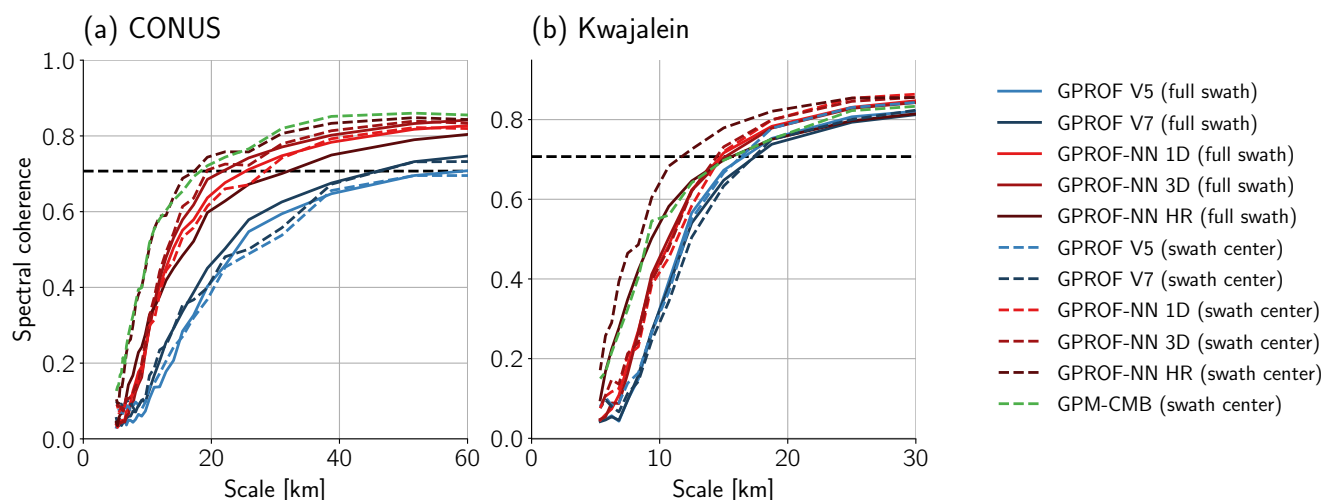
Figure 11. Spectral coherence curves for the GPROF retrievals over CONUS (Panel (a)) and the Kwajalein Atoll in the tropical Pacific (Panel (b)). Solid lines show the coherence curve for the evaluation covering the center of the GMI swath. Dashed lines show the coherence curves calculated over the full GMI swath.

The resulting effective resolutions of the retrievals are listed in Tab. 2. The effective resolution for the conventional GPROF retrieval over CONUS are around 45 and 60 km for the GPROF V7 and the GPROF V5 retrieval, respectively. The effective resolution of the GPROF V5 retrieval is significantly lower when calculated only over the swath center, but this is caused by the curve falling slightly below the threshold for the calculation of the effective resolution and does not correspond to a significant difference in the coherence curves. With effective resolution of about 29 and 21 km, respectively, the GPROF-NN 1D and 3D retrievals significantly improve the effective resolution over the CONUS. The effective resolution also clearly reflects the overfitting behavior of the GPROF-NN HR retrieval, which achieves an effective resolution of 17 km over the central part of the swath but only 31 km across the full swath. The effective resolution of the GPM CMB retrieval is 18 km over the CONUS.

Over the Kwajalein site, the effective resolution of the GPROF V5 retrieval of 16 km is slightly better than that of the GPROF V7 retrieval with 17 km. Both GPROF-NN retrievals achieve an effective resolution of around 15 km. The effective resolution of the GPROF-NN HR retrieval is 11 km evaluated at the swath center but decreases to 15 km over the full swath. The effective resolution of the GPM CMB retrieval is only 15 km.

3.3 Retrieval accuracy of GPROF for other GPM sensors

400   This section evaluates the accuracy of GPROF the other sensor types of the GPM constellation for which MRMS validation data is available. One year of collocations with MRMS and K-POL precipitation estimates were extracted for one instance of each sensor type. The corresponding platforms and collocation periods are listed in Table 1. The resulting retrieval accuracy metrics over CONUS and Kwajalein are displayed in Fig. 12 and Fig. 13, respectively.

| Retrieval | Effective resolution [km] | | | |
| | CONUS | | Kwajalein | |
	Full swath	Swath center	Full swath	Swath center
GPROF V5	59.38	86.24	16.6	16.36
GPROF V7	45.96	46.3	17.46	17.37
GPROF-NN 1D	29.49	28.48	14.75	14.77
GPROF-NN 3D	21.83	19.3	15.14	14.5
GPROF-NN HR	32.03	17.36	15.61	11.72
GPM CMB	-	18.32	-	15.65

Table 2. Effective resolution of the GPROF and GPM CMB retrieval evaluated against ground radar measurements.

The biases of the GPROF V7 retrieval are well within 10% over both the CONUS and the Kwajalein site. The exception are
405 MHS and ATMS, for which the bias of Kwajalein exceeds 30%. MSE, MAE and SMAPE are larger than for GMI, which is
expected considering that the other sensors of the GPM constellation have lower spatial resolution, fewer channels or both.

Compared to GPROF V5, the retrieval biases of V7 tend to be smaller than those of V5. GPROF V7 tends to exhibit larger
MAE but reduced MSE and SMAPE. A possible explanation for this may be the treatment of pixels with low probability of
precipitation, which were explicitly set to $0.0\,\mathrm{mmh}^{-1}$ in GPROF V5. The clearest difference between GPROF V7 and GPROF
410 V5 are observed for SSMIS, for which all accuracy metrics are improved.

The GPROF-NN retrievals exhibit significant spread in biases. Over the CONUS, the retrieval have a tendency to be biased
low with biases reaching up to 20%. Over Kwajalein, the retrieval for the conical scanners (TMI, SSMIS, AMSR2) are biased
low whereas the retrievals for the cross-track scanners exhibit wet biases of up 70% for MHS. GPROF-NN 1D yields similar or
slightly more accurate results in terms of MAE, MSE and correlation for the conical scanners but yields less accurate results for
415 the assessed cross-track-scanning sensors, MHS and ATMS. For GPROF-NN 3D the improvements are more robust. In terms
of the SMAPE, the results are not conclusive with the GPROF-NN retrievals yielding higher errors for some of the sensors.

The results obtained for the GPROF-NN retrievals contrast with the results for GMI, for which very clear improvements
over the conventional GPROF retrievals were observed. A likely cause for this is that the training of the GPROF-NN retrievals
is based on radiative transfer simulations, which introduce an additional error source into the retrieval. Especially the large
420 biases over the Kwajalein site in the retrievals from the cross-track scanners, whose simulations use a different code than the
simulations for the conical scanners, point to problems with the simulations.

4 Discussion

4.1 Limitations of this validation study

The validation excluded snow-covered land surfaces and mountainous terrain because of the enhanced uncertainties in the
425 reference data and because the GPROF retrieval database employs enhancement factors to account for orographic enhancement
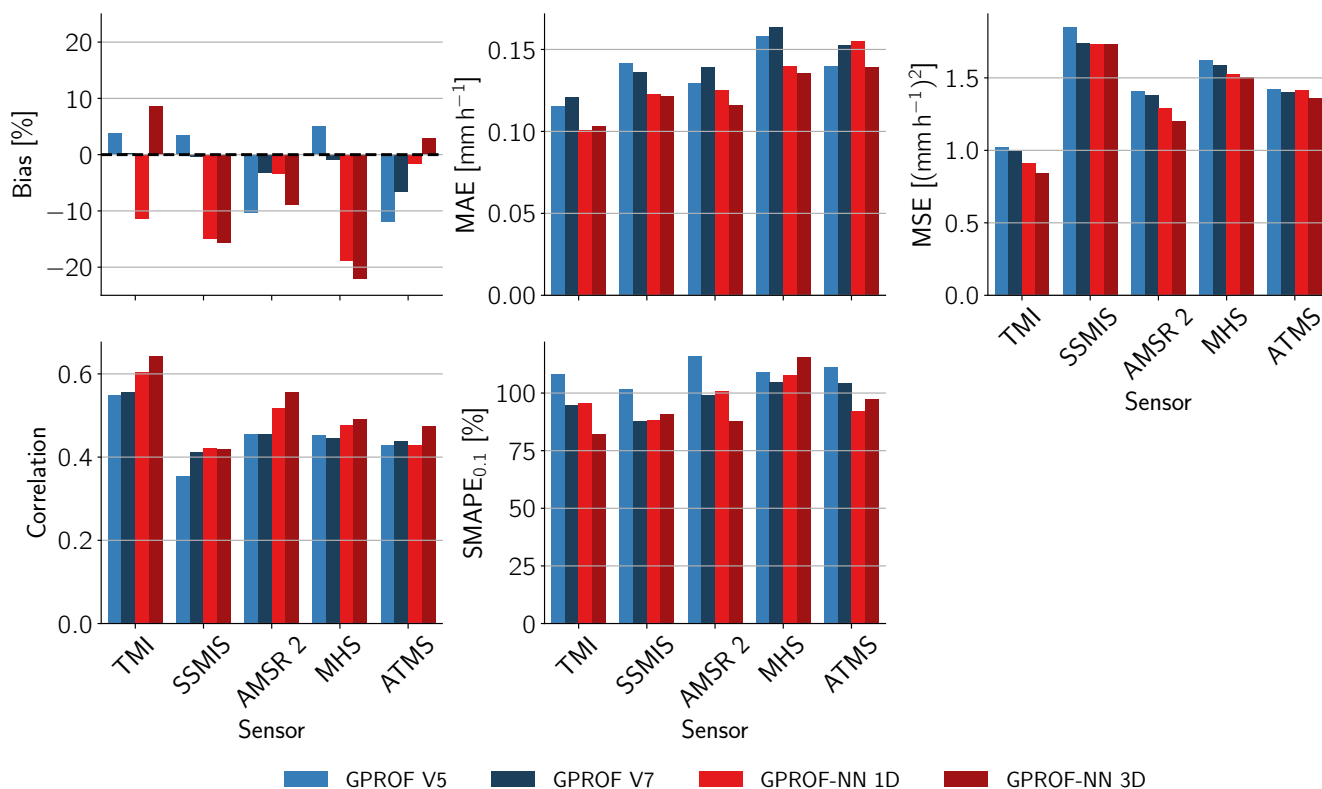
Figure 12. Accuracy metrics for a selection of other sensors of the GPM constellation evaluated against gauge corrected MRMS measurements of CONUS.

of precipitation. These enhancement factors would skew the comparison to the GPM CMB product and thus complicate the attribution of retrieval errors. Moreover, precipitation estimates in these scenarios remain so uncertain that their validation requires careful, specialized evaluation, which is outside the scope this study.

4.2 From GPROF V5 to V7

430 In terms of retrieval accuracy for GMI, GPROF V7 performs similar to the previous version of the algorithm, GPROF V5. While the biases of GPROF V7 compared to the ground-radar validation data are larger than those of V5, they are closer to GPM CMB, which the retrieval aims to reproduce. Small improvements are found in terms of MAE, MSE, Correlation and SMAPE. The effective resolution of the two retrievals are practically identical.

GPROF V7 was found to be about 5% drier than V5 over the whole CONUS. On regional scales, the most significant
435 changes are a reduction in the overestimation of warm-season precipitation over the central CONUS, reduced underestimation of cold-season precipitation in the North-West, a slightly more pronounced underestimation of warm-season precipitation in the South-East and a more pronounced overestimation of precipitation in the tropical Pacific. Without gauges to calibrate the
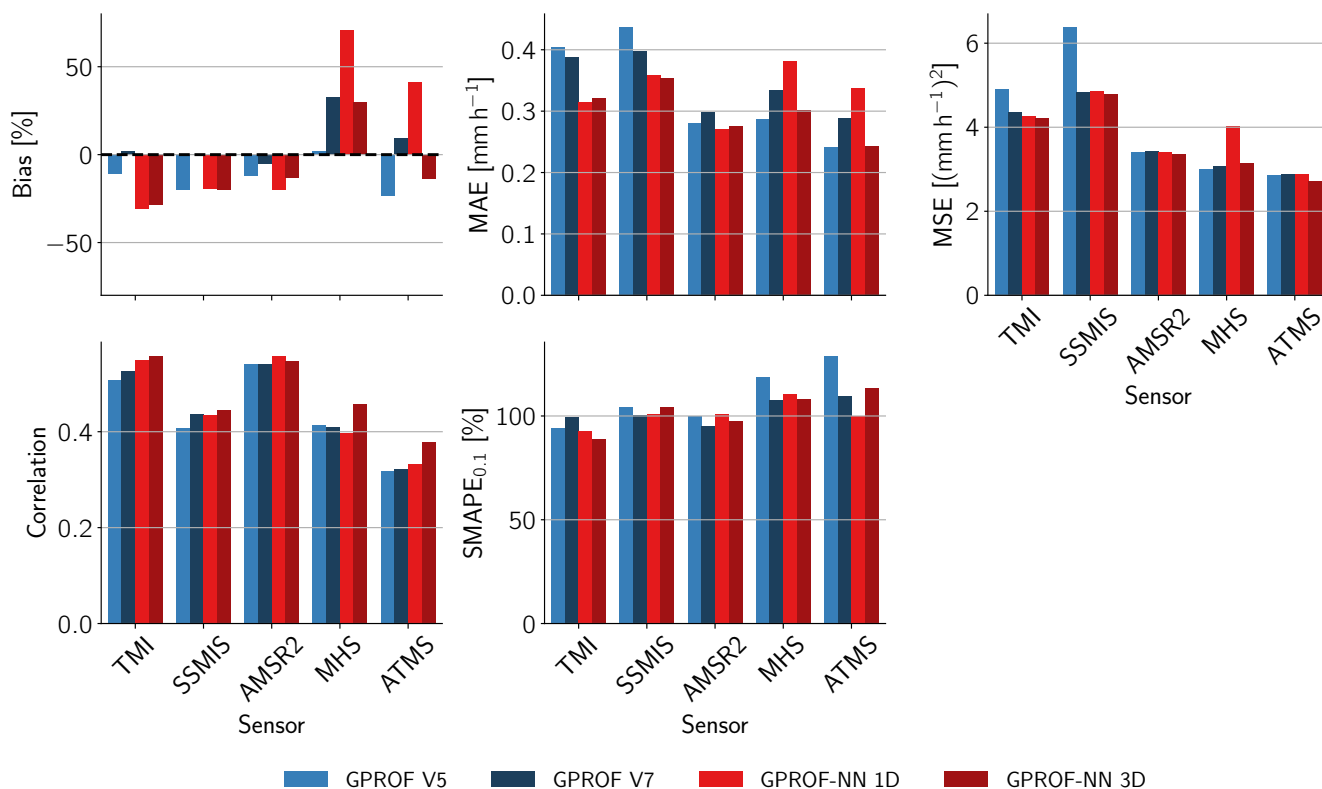
Figure 13. Accuracy metrics for a selection of other sensors of the GPM constellation evaluated against the K-POL radar in the tropical Pacific.

rainfall product, the Kwajalein site may, however, by more uncertain than the CONUS estimates. For the other sensors of the GPM constellation assessed here, GPROF V7 exhibited smaller biases compared to the ground-based reference data.

4.3 Attribution of retrieval errors

In order for this validation to inform the future development of GPROF, it is essential to determine whether the observed deviations from the validation data are due to the reference data used by GPROF or the inherent uncertainties of the retrieval and the applied method. Figure 3 revealed a high degree of similarity between biases in GPROF and GPM CMB in both space and time. This indicates that the GPROF retrieval inherits systematic errors from the GPM CMB retrieval. Further evidence for a notable contribution of GPM CMB to systematic errors in GPROF was found in the analysis of regional biases of GPROF, which tend to follow those of GPM CMB (Fig. 8, Fig. 9, Fig. 10).

The attribution of the errors can be formalized by investigating the coefficient of determination between the errors in GPROF and GPM CMB with respect to the validation data. The coefficient of determination represents the fraction of variance of the GPROF errors that can be explained by the GPM CMB errors and thus quantifies the contribution of the GPM CMB errors to

the MSE of GPROF. Table 3 lists the coefficients of determination of errors between GPM CMB and GPROF and validation

data for instantaneous estimates as well as the annual means at $5°$ resolution shown in Fig. 3. For the GPROF V7 retrievals, the

coefficient of determination for instantaneous error varies between 0.17 and 0.3 over the four assessed years. This means that

between 17 and 30% of the mean squared error can be explained by the error in the GPM CMB retrieval. For the GPROF-NN

retrievals this rate is consistently higher and reaches up to between 30 and 46% for the GPROF-NN HR retrieval. This indicates

455 that GPM CMB error has a notable influence of the accuracy of GPROF even for instantaneous precipitation estimates.

All rates of explained variability increase drastically when annual accumulations over $5°$ boxes are considered. For GPROF

V7 they range from 36 to 67% while they reach between 79 and 93% for GPROF-NN 3D. While the fraction of explained

variability in the GPROF V7 biases is around 50% over the 4-year period, it increases to about 75% for GPROF-NN 1D,

80% for GPROF-NN HR and 85% for GPROF-NN 3D. This confirms that systematic regional errors in the GPM CMB data

460 contributes significantly to regional biases in GPROF.

Retrieval	Instantaneous				$5°$ annual mean			
	2019	2020	2021	2022	2019	2020	2021	2022
GPROF V7	0.17	0.19	0.30	0.30	0.36	0.37	0.67	0.67
GPROF-NN 1D	0.20	0.25	0.34	0.37	0.76	0.73	0.69	0.79
GPROF-NN 3D	0.22	0.26	0.36	0.39	0.79	0.87	0.84	0.93
GPROF-NN HR	0.31	0.33	0.42	0.46	0.88	0.85	0.52	0.95

Table 3. Coefficient of determination between errors in GPM CMB and GPROF retrieval errors.

The role of GPM CMB for the systematic error in GPROF becomes even clearer when the regional biases of GPM CMB

are subtracted from the biases of GPROF. The resulting maps of reference-data-corrected GPROF biases is shown in Fig. 14.

Subtracting the GPM CMB biases reduces the overall magnitude of the biases and their temporal and spatial variability. The

impact is more pronounced for the GPROF-NN retrievals indicating that their higher accuracy puts increased weight on the

465 reference data.

4.4 Validation of the GPROF-NN retrievals

The assessment of the GPROF-NN retrievals against ground-radar measurements confirmed the significantly higher retrieval

accuracy of the neural-network based implementation for GMI. Although the biases of GPROF-NN 1D are similar to those of

GPROF V7, consistent improvements of the order of $10-20\%$ are observed in all other assessed accuracy metrics. Significant

470 improvements were also found in detection of precipitation and the effective resolution of the retrieval. The effective resolution

of the GPROF-NN retrievals over land is more than twice as high as that of GPROF V7. This is a very encouraging result,

considering that these improvements are comparable to those between sensor generations. The fact that a simple algorithm

update can unlock these improvements, highlights the importance of improving the retrieval methodology for current and

historical space-borne precipitation estimates.
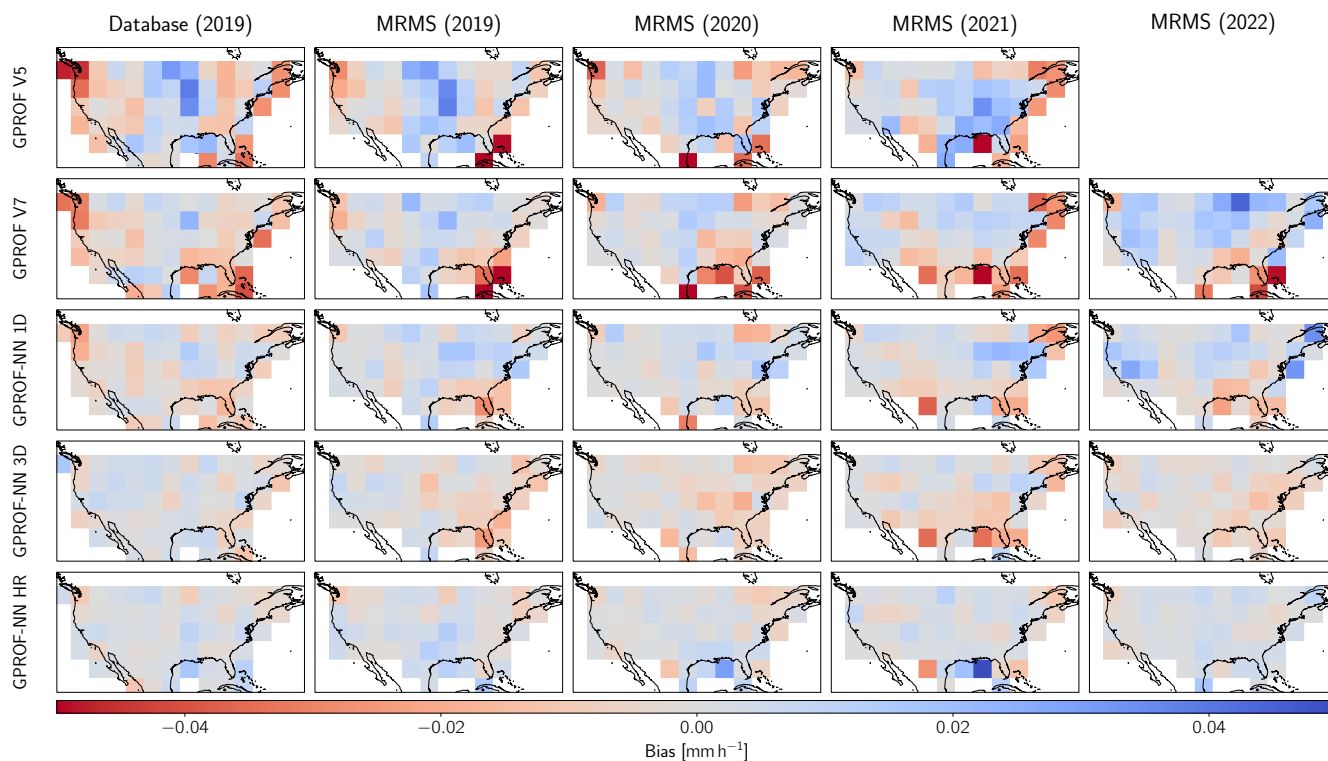
Figure 14. Spatial distribution of retrieval biases corrected for biases in GPM CMB. Each panel shows the biases calculated over a uniform longitude-latitude grid with a grid size of $5\,^\circ$. The corresponding biases of GPM CMB have been subtracted from each bias field. Rows show the results for the different retrieval algorithms. Columns show the results for different reference data and time periods.

475   Even higher accuracy is achieved by GPROF-NN 3D and the GPROF-NN HR retrievals. However, the analysis of the effective resolution presented in Sec. 3.2 indicates that the retrieval is overfitting to the viewing geometry at the center of the swath. It seems likely that the decrease in retrieval accuracy towards the edges of the swath is due to the parallax effect, which changes direction towards the edges of the swath. Since reference precipitation measurements are only available at the center of the swath, the neural network cannot learn that the direction of the parallax effect changes across the swath. Fully realizing

480   the potential of the GMI sensor may therefore require incorporating this effect into the training process. Furthermore, the overfitting of the GPROF-NN 3D and GPROF-NN HR retrievals points to a general limitation of using data based on the GPM DPR to train and evaluate GMI retrievals that are based on convolutional neural networks. When only data from the center of the swath is used to train and evaluate a retrieval, its accuracy will likely be overestimated.

Nonetheless, the results of the GPROF-NN 3D and especially the GPROF-NN HR retrievals are worth highlighting. The

485   GPROF-NN HR retrieval demonstrates that precipitation retrievals from GMI can achieve retrieval accuracy that is on par with that of the GPM CMB retrievals and significantly higher than that of the current GPROF retrievals for GMI. The analysis of the contribution of the GPM CMB retrieval errors indicates that with a suitable neural-network-based retrieval systematic errors

at regional scales are limited by the accuracy of the reference data which accounts to up to 37% of the MSE of instantaneous precipitation estimates and up to 95% of the variability in annual accumulations at regional scales.

490   Unfortunately the improvements observed for GMI did not carry over to the other sensors of the constellation. The principal difference for those sensors is that the a priori database uses simulated brightness temperatures instead of real observations. Errors in the simulated brightness temperatures thus constitute an additional error source that incurs a generalization error when the retrieval is applied to real observations. Pfreundschuh et al. (2022) found clear improvements in retrieval accuracy of the GPROF-NN retrieval also for the MHS sensor when the retrieval is evaluated against the a priori database. This indicates

495 that the simulation errors are to blame for the reduced accuracy of the GPROF-NN retrievals for the other GPM sensors. Nonetheless, the GPROF-NN 3D retrieval improves most error metrics compared to the GPROF V7 retrieval. This indicates that the incorporation of structural information helps to make the retrieval more robust.

     The largest retrieval biases were found for the GPROF-NN 1D retrieval for the cross-track scanners over the Kwajalein site where wet biases up to 70% were found. This may indicate a particular problem with the simulation of the observations from

500 cross-track scanners over ocean. While not as extreme, the GPROF-NN retrievals exhibit higher biases than the GPROF V7 retrieval over CONUS.

     It should be noted, however, that the same simulations are used by GPROF V7, which appears to be less affected by the suspected simulation errors. The improved robustness of the Bayesian scheme used by GPROF V7 is likely due to the hard-coded channel uncertainties, which take into account the forward simulation error of each sensor. In contrast to that, the

505 neural-network retrievals learn the observation uncertainties from the training data.

     Another notable difference between the GPROF and the GPROF-NN retrievals is that GPROF applies an additional calibration of the precipitation that ensures that the annual mean precipitation matches that of GMI. Such a calibration has not been included in the GPROF-NN retrievals. Furthermore, it should be noted that the measurements from the K-POL radar can be as much as 20% lower than the GPM CMB retrievals, which highlights the difficulty of finding reliable validation measurements

510 over the ocean.

4.5   Future improvements of GPROF

The results presented in this study show that observations from GMI can be used to produce precipitation estimates at essentially the same accuracy as the GPM CMB retrievals with respect to gauge-corrected ground-based radar measurements. However, achieving this level of accuracy across the full swath will require including reference data from the outer edges of the

515 GMI swath in the training data. In principle, this can be achieved through synthetic transformations of the existing training data that account for the parallax effect. However, a technically simpler approach is including reference data from other sources in the training. Given that these improvements can be retroactively applied to the full record of GMI observations, we consider this worthy of further development work.

     The largest impediment to an operational adoption of the GPROF-NN retrievals is the retrieval bias observed for other

520 sensors. The reason for this are likely errors in the assumed hydrometeor profiles and/or radiative transfer simulations used to generate the a priori databases for these sensors. We thus identify the radiative transfer simulations used for the GPROF a

priori database as a critical limitation for the accuracy of GPM precipitation estimates and suggest that future research should systematically assess and improve these simulations. Alternatively, collocations between GPM CMB and the other sensors may be incorporated in the training process to improve the retrievals from these sensors.

525   However, our results also show that systematic retrieval errors of all GPROF retrievals largely follow the biases of the GPM CMB retrieval. In the longer term, improvements of global precipitation measurements will therefore require improvements in what are today considered reference measurements of precipitation, i.e. (potentially gauge-corrected) ground-radar measurements and space-borne combined radar/PMW radiometer measurements. Convergence of these measurements is a prerequisite to better constrain global precipitation measurements from meteorological satellites.

530   5   Summary and conclusions

This study validated the most recent version of the GPROF PMW precipitation retrieval, GPROF V7, its predecessor, GPROF V5, and the experimental GPROF-NN retrievals, which are candidate algorithms for the next version of GPROF. The assessment was based on multiple years of collocations with gauge-corrected ground radar QPEs over CONUS as well as non-gauge-corrected ground-radar QPEs from the Kwajalein atoll. The validation data included collocations between the GPROF retrieval

535   database and the ground-based validation data as well as the GPM CMB product and the validation data in order to separate the contributions of the reference data used by the GPROF retrieval and the retrieval method.

The validation of GPROF V7 surface precipitation revealed the following characteristics:

1. Over CONUS, GPROF V7 retrievals from GMI show a weak tendency to underestimate precipitation with dry biases varying between 2 and 8% during the 4 surveyed years. At a resolution of $5\,\mathrm{km}$, retrieval errors range from 0.11 to

540    $0.13\ \mathrm{mm\,h^{-1}}$ in terms of MAE, 1.1 and $1.6\,(\mathrm{mmh^{-1}})^2$ in terms of MSE, and 80 to 90% in terms of $\mathrm{SMAPE}_{0.1}$. The correlation ranges from 0.45 to 0.55.

2. Over the five surveyed sub-regions over CONUS, GPROF V7 GMI was biased dry in four of them. The strongest dry biases were observed in the South-West, where they reached up to 25%. The central CONUS was the only region where consistent wet biases between 2 and 25% were observed.

545   3. Over the tropical Pacific GPROF V7 GMI exhibits a wet bias ranging from 2 to 18% compared to measurements from the K-POL radar. However, the retrieval remains within 10% of the GPM CMB product.

4. GPROF V7 GMI systematically underestimates cold-season precipitation in the North-West and warm-season, afternoon convective precipitation in the South-West and South-East.

5. Annual biases at $5°$ resolution are moderately correlated with GPM CMB indicating that differences between GPM

550   CMB and the ground-radar measurements contribute to the observed biases.

6. Compared to GPROF V5, GPROF V7 GMI is about 5% dryer than GPROF V5 over CONUS and $10-20\%$ wetter over the tropical Pacific. Comparable tendencies were observed for TMI, SSMIS, AMSR2, MHS and ATMS. MAE, MSE, SMAPE and correlation were found to be mostly identical.

Based on these findings, we conclude that GPROF V7 reliably captures principal precipitation characteristics from continental and regional scales. However, significant spatial and temporal variability in the retrieval errors make it difficult to generalize the validation results to a global context. It should be noted that our assessment ignored snow-covered surfaces and mountain environments, which will need to be assessed in future research.

Furthermore, our results confirm the potential of the neural-network-based implementations of GPROF to improve the GPM PMW retrievals. Consistent improvements in the accuracy of instantaneous precipitation estimates are observed for GMI. In particular, we find a doubling of the effective resolution of the retrievals over land. Even larger improvements can be achieved through the use of retrievals that exploit spatial patterns in the input observations. In particular, we have shown that a GMI super-resolution retrieval can achieve similar retrieval accuracy and effective resolution as the GPM CMB product. However, these were found to overfit on the GMI viewing geometry at the center of the swath and will thus require further development before they can achieve high accuracy across the full swath. In addition to that, the accuracy of the hydrometeor profiles and/or radiative transfer simulations of the GPROF a priori database were identified as a principal impediment for unlocking the improvements found for GMI for the other sensors of the GPM constellation. The presented results demonstrate that the current observational record of precipitation could be improved significantly by adopting the neural-network based retrieval for the operational processing of the observations from the GPM constellation but that further development is needed to harness the full potential of the PMW observations from the GPM constellation.

Code and data availability. The implementation of all GPROF-NN retrievals as well as the code used to extract the validation collocations analyze the results is available from Pfreundschuh (2022). GPROF level 2A retrieval results are available from GPM Science Team (2022c, b, d, e, a). GPM CMB retrieval are available from Olson (2022). The gauge-corrected MRMS measurements for overpasses of GPM sensors over CONUS as well as the measurements from the K-POL radar are available from Wolff (2023). The validation collocations created for this study are not publicly available due to the significant storage requirements. We are, however, happy to share them upon request.

Author contributions. SP has designed and implemented the study and lead the writing of the manuscript. CG has evaluated the effective resolution of the retrievals and contributed to writing the manuscript. PB helped gathering the data required for the validation. All authors have contributed to the writing process through discussion and feedback.

Competing interests. No competing interests are present.

Acknowledgements. We would like to acknowledge Pierre Kirstetter for the provision of the MRMS radar measurements for GPM ground
580 validation and advice on the use of the data. Furthermore, we would like to acknowledge David B. Wolff and Jason L. Pippit for providing
guidance on the use of the data from the K-POL radar.

The computations for this study were performed using several freely available programming languages and software packages, most
prominently the Python language (The Python Language Foundation, 2018), the IPython computing environment (Perez and Granger, 2007),
the numpy package for numerical computing (van der Walt et al., 2011), xarray (Hoyer and Hamman, 2017) and satpy (Raspaud et al., 2021)
585 for the processing of satellite data and matplotlib (Hunter, 2007) and cartopy (Met Office, 2010 - 2015) for generating figures.

References

Boukabara, S.-A., Garrett, K., Chen, W., Iturbide-Sanchez, F., Grassotti, C., Kongoli, C., Chen, R., Liu, Q., Yan, B., Weng, F., Ferraro, R., Kleespies, T. J., and Meng, H.: MiRS: An All-Weather 1DVAR Satellite Data Assimilation and Retrieval System, IEEE Transactions on Geoscience and Remote Sensing, 49, 3249–3272, https://doi.org/10.1109/TGRS.2011.2158438, 2011.

590 Cifelli, R., Chandrasekar, V., Lim, S., Kennedy, P. C., Wang, Y., and Rutledge, S. A.: A New Dual-Polarization Radar Rainfall Algorithm: Application in Colorado Precipitation Events, Journal of Atmospheric and Oceanic Technology, 28, 352 – 364, https://doi.org/10.1175/2010JTECHA1488.1, 2011.

GPM Science Team: GPM SSMIS on F17 (GPROF) Climate-based Radiometer Precipitation Profiling 1.5 hours 12 km V07, https://disc.gsfc.nasa.gov/datasets/GPM_2AGPROFF17SSMIS_CLIM_07/summary?keywords=GPM_2AGPROFF17SSMIS_CLIM_07,

595 https://doi.org/10.5067/GPM/SSMIS/F17/GPROFCLIM/2A/07, accessed: 2023-04-01, 2022a.

GPM Science Team: GPM AMSR-2 on GCOM-W1 (GPROF) Climate-based Radiometer Precipitation Profiling L2A 1.5 hours 10 km V07, https://disc.gsfc.nasa.gov/datasets/GPM_2AGPROFGCOMW1AMSR2_CLIM_07/summary?keywords=GPM_2AGPROFGCOMW1AMSR2_CLIM_07, https://doi.org/10.5067/GPM/AMSR2/GCOMW1/GPROFCLIM/2A/07, accessed: 2023-04-01, 2022b.

600 GPM Science Team: GPM GMI (GPROF) Climate-based Radiometer Precipitation Profiling L2A 1.5 hours 4 km x 4 km V07, https://disc.gsfc.nasa.gov/datasets/GPM_2AGPROFGPMGMI_CLIM_07, https://doi.org/10.5067/GPM/GMI/GPROFCLIM/2A/07, accessed: 2023-04-01, 2022c.

GPM Science Team: GPM MHS on NOAA-19 (GPROF) Radiometer Precipitation Profiling L2A 1.5 hours 17 km V07, https://disc.gsfc.nasa.gov/datasets/GPM_2AGPROFNOAA19MHS_CLIM_07/summary?keywords=GPM_2AGPROFNOAA19MHS_CLIM_07,

605 https://doi.org/10.5067/GPM/MHS/NOAA19/GPROFCLIM/2A/07, accessed: 2023-04-01, 2022d.

GPM Science Team: GPM ATMS on SUOMI-NPP (GPROF) Radiometer Precipitation Profiling L2 1.5 hours 16 km V07, https://disc.gsfc.nasa.gov/datasets/GPM_2AGPROFNPPATMS_CLIM_07/summary?keywords=GPM_2AGPROFNPPATMS_CLIM_07, https://doi.org/10.5067/GPM/ATMS/NPP/GPROFCLIM/2A/07, accessed: 2023-04-01, 2022e.

Grecu, M., Olson, W. S., Munchak, S. J., Ringerud, S., Liao, L., Haddad, Z., Kelley, B. L., and McLaughlin, S. F.: The GPM Combined

610 Algorithm, J. Atmos. Oceanic Technol., 33, 2225–2245, https://doi.org/10.1175/JTECH-D-16-0019.1, 2016.

Guilloteau, C., Foufoula-Georgiou, E., and Kummerow, C. D.: Global multiscale evaluation of satellite passive microwave retrieval of precipitation during the TRMM and GPM eras: Effective resolution and regional diagnostics for future algorithm development, Journal of Hydrometeorology, 18, 3051 – 3070, https://doi.org/10.1175/JHM-D-17-0087.1, 2017.

Hegerl, G. C., Black, E., Allan, R. P., Ingram, W. J., Polson, D., Trenberth, K. E., Chadwick, R. S., Arkin, P. A., Sarojini, B. B., Becker, A.,

615 Dai, A., Durack, P. J., Easterling, D., Fowler, H. J., Kendon, E. J., Huffman, G. J., Liu, C., Marsh, R., New, M., Osborn, T. J., Skliris, N., Stott, P. A., Vidale, P.-L., Wijffels, S. E., Wilcox, L. J., Willett, K. M., and Zhang, X.: Challenges in Quantifying Changes in the Global Water Cycle, Bulletin of the American Meteorological Society, 96, 1097 – 1115, https://doi.org/10.1175/BAMS-D-13-00212.1, 2015.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren,

620 P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S.,

and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The Global
625   Precipitation Measurement Mission, Bull. Amer. Met. Soc., 95, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1, 2014.

Hoyer, S. and Hamman, J.: xarray: N-D labeled arrays and datasets in Python, Journal of Open Research Software, 5, https://doi.org/10.5334/jors.148, 2017.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff,
       D. B., and Xie, P.: Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG), pp.
630   343–353, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-24568-9_19, 2020.

Hunter, J. D.: Matplotlib: A 2D graphics environment, Comput. Sce. Eng., 9, 90–95, https://doi.org/10.1109/MCSE.2007.55, 2007.

Kerr, C. A. and Alsheimer, F.: Storm-Scale Predictability and Analysis of the 13 April 2020 Central Savannah River Area Tornado
       Outbreak, Weather and Forecasting, 37, 901 – 913, https://doi.org/10.1175/WAF-D-21-0185.1, 2022.

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., McNally, A., Skofronick-Jackson, G., Stocker, E., Wu, H.,
635   and Zaitchik, B. F.: NASA's Remotely Sensed Precipitation: A Reservoir for Applications Users, Bulletin of the American
       Meteorological Society, 98, 1169 – 1184, https://doi.org/10.1175/BAMS-D-15-00296.1, 2017.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W., and Amitai, E.: Toward a Framework
       for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar-Based National Mosaic QPE,
       Journal of Hydrometeorology, 13, 1285 – 1300, https://doi.org/10.1175/JHM-D-11-0139.1, 2012.

640 Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Joseph Munchak, S., and Petkovic, V.: The Evolution of the Goddard
       Profiling Algorithm to a Fully Parametric Scheme, J. Atmos. Oceanic Technol., 32, 2265–2280,
       https://doi.org/10.1175/JTECH-D-15-0039.1, 2015.

Met Office: Cartopy: a cartographic python library with a matplotlib interface, Exeter, Devon, http://scitools.org.uk/cartopy, 2010 - 2015.

NOAA National Severe Storms Laboratory: MRMS QPE Code Updates, https://inside.nssl.noaa.gov/mrms/code-updates, accessed:
645   2023-06-13, 2023.

Olson, W.: GPM DPR and GMI Combined Precipitation L2B 1.5 hours 5 km V07,
       https://disc.gsfc.nasa.gov/datasets/GPM_2BCMB_07/summary?keywords=gpm%20cmb%20corra,
       https://doi.org/10.5067/GPM/DPRGMI/CMB/2B/07, accessed: 2023-04-01, 2022.

Perez, F. and Granger, B. E.: IPython: A System for Interactive Scientific Computing, Computing in Science Engineering, 9, 21–29,
650   https://doi.org/10.1109/MCSE.2007.53, 2007.

Pfreundschuh, S.: GPROF-NN: A neural network based implementation of the Goddard Profiling Algorithm,
       https://doi.org/10.5281/zenodo.5819297, 2022.

Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., and Norrestad, T.: GPROF-NN: A neural network based implementation of
       the Goddard Profiling Algorithm, Atmospheric Measurement Techniques Discussions, 2022, 1–37, https://doi.org/10.5194/amt-2022-15,
655   2022.

Raspaud, M., Hoese, D., Lahtinen, P., Finkensieper, S., Holl, G., Proud, S., Dybbroe, A., Meraner, A., Feltz, J., Zhang, X., Joro, S., Roberts,
       W., Ørum Rasmussen, L., strandgren, BENR0, Méndez, J. H. B., Zhu, Y., Daruwala, R., Jasmin, T., mherbertson, Kliche, C., Barnie, T.,
       Sigurðsson, E., R.K.Garcia, Leppelt, T., TT, ColinDuff, Egede, U., LTMeyer, and Itkin, M.: pytroll/satpy: Version 0.33.1,
       https://doi.org/10.5281/zenodo.5789830, 2021.

660 Ryzhkov, A. V., Giangrande, S. E., Melnikov, V. M., and Schuur, T. J.: Calibration Issues of Dual-Polarization Radar Measurements, Journal of Atmospheric and Oceanic Technology, 22, 1138 – 1155, https://doi.org/10.1175/JTECH1772.1, 2005.

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., and Wilheit, T.: The Global Precipitation Measurement (GPM) Mission for Science and Society, Bulletin of the American Meteorological Society, 98, 1679 –

665 1695, https://doi.org/10.1175/BAMS-D-15-00306.1, 2017.

Tapiador, F., Navarro, A., Levizzani, V., GarcÃa-Ortega, E., Huffman, G., Kidd, C., Kucera, P., Kummerow, C., Masunaga, H., Petersen, W., Roca, R., Sánchez, J.-L., Tao, W.-K., and Turk, F.: Global precipitation measurements for validating climate models, Atmospheric Research, 197, 1–20, https://doi.org/10.1016/j.atmosres.2017.06.021, 2017.

The Python Language Foundation: The Python Language Reference, https://docs.python.org/3/reference/index.html, 2018.

670 van der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science Engineering, 13, 22–30, https://doi.org/10.1109/MCSE.2011.37, 2011.

Wolff, D. B.: Precipitation Measurements Missions Ground Validation Data, https://pmm-gv.gsfc.nasa.gov/, https://doi.org/10.5067/GPM/DPRGMI/CMB/2B/07, accessed: 2023-04-01, 2023.