GPROF-NN: A neural network based implementation of the Goddard Profiling Algorithm

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

[1]Department of Space, Earth and Environment, Chalmers University of Technology, 41296 Gothenburg, Sweden
[2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523
[3]Previously Department of Space, Earth and Environment, Chalmers University of Technology, 41296 Gothenburg, Sweden

Correspondence: Simon Pfreundschuh (simon.pfreundschuh@chalmers.se)

Abstract. The Global Precipitation Measurement (GPM) mission aims to provide global measurements of precipitation with a temporal resolution of three hours in order to allow close monitoring of the global hydrological cycle. To achieve global coverage at such high temporal resolution, GPM combines observations from a constellation of passive microwave (PMW) sensors. The Goddard Profiling Algorithm (GPROF) is the operational retrieval for precipitation and hydrometeor profiles
5   from these sensors.

Aiming to investigate the effect of the retrieval algorithm on the accuracy of the PMW precipitation retrieval, we present two neural network based, probabilistic implementations of GPROF: GPROF-NN 1D, which processes individual pixels just as the current GPROF algorithm, and GPROF-NN 3D, which employs a convolutional neural network to incorporate structural information into the retrieval. To isolate the impact of the retrieval method from any issues in the training data, the GPROF
10   and GPROF-NN algorithms are evaluated on a separate test set that has the same statistical properties as the data used for the training of the GPROF-NN algorithms and the development of the current GPROF algorithm. Comparison of GPROF and the GPROF-NN 1D algorithm shows that by replacing GPROF with an identical neural network based retrieval the accuracy of the retrieved surface precipitation from the GPM Microwave Imager (GMI) can be improved by 10 to 25% in terms of absolute error, root mean squared error and symmetric mean absolute percentage error. Comparable improvements are observed for the
15   retrieved hydrometeor profiles and their column integrals. The improvements are consistent spatially as well as with respect to different surface types. The effective resolution in along track direction of the retrieved surface precipitation fields is increased from 23 km to 14 km. Similar, additional improvements are found for the GPROF-NN 3D retrieval over the performance of the GPROF-NN 1D retrieval, showing the added benefits of incorporating structural information into the retrieval. The effective resolution in along-track direction of the GPROF-NN 3D algorithm is reduced to 13.5 km, which is the upper limit imposed
20   by the along track separation of consecutive scan lines. Comparable improvements are found also when the algorithms are applied to synthetic observations from the cross track scanning Microwave Humidity Sounder (MHS) sensor. Application of the retrieval algorithm to real observations from the GMI and MHS sensors of Hurricane Harvey suggest that these improvements can be expect to carry over to operational application.

The novel GPROF-NN algorithms presented here were designed to be functionally equivalent to the current implementation,
25   which enables the neural network approach to replace the current Bayesian scheme in a future update. Despite their superior

retrieval accuracy, the single CPU core runtime required for the operational processing of an orbit of observations is lower than that of GPROF. The GPROF-NN algorithms thus promise to be a simple and cost efficient way to improve the accuracy of the PMW precipitation retrievals of the GPM constellation and thus help improve the monitoring of the global hydrological cycle.

1 Introduction

30 The Goddard Profiling Algorithm (GPROF, Kummerow et al. (2015)) is the operational precipitation retrieval algorithm for the passive microwave (PMW) observations from the constellation of satellites of the Global Precipitation Measurement (GPM, Hou et al. (2014)) mission, whose objective is to provide consistent global measurements of precipitation at a temporal resolution of 3 hours. In addition to being used directly by meteorologists and climate scientists, the precipitation that is retrieved using GPROF serves as input for GPM level 3 retrieval products. The algorithm thus constitutes an essential component of the

35 system of global observations that enables monitoring of the hydrological cycle for the benefit of science and society.

The development of GPROF was originally motivated by the Tropical Rainfall Measurement Mission (TRMM, Simpson et al. (1996)), the precursor of the GPM mission, and thus dates back almost 30 years (Kummerow and Giglio, 1994b, c). Due to the conceptual and computational complexity of simulating PMW observations of clouds and precipitation, the algorithm was and remains based on a retrieval database consisting of observations and corresponding profiles of hydrometeors and

40 precipitation rates. Nonetheless, the algorithm has undergone several updates since its conception: Methodologically, the most fundamental modification was the introduction of the Bayesian retrieval scheme in Kummerow et al. (1996), which is used in the algorithm until today. Following this, algorithm updates were mostly focused on improving the retrieval database and the incorporation of ancillary data into the retrieval. While the first version of GPROF still used hand crafted hydrometeor profiles to generate the retrieval database, these were soon replaced by profiles from a meso-scale weather model (Kummerow et al.,

45 1996). An important improvement was the replacement of the model-derived database by an observationally-generated database for the GPROF 2010 algorithm (Kummerow et al., 2011, 2015), which helped reduce errors caused by misrepresentation of atmospheric states in the database. The 2014 version of GPROF (Kummerow et al., 2015) introduced the first fully-parametric version of the algorithm, which was designed to be applicable to all sensors of the GPM constellation. This version of GPROF became the operational PMW precipitation retrieval of the GPM mission.

50 This study focuses on the computational method that is used to produce the retrieval results from the retrieval database used by GPROF. Since its introduction in Kummerow et al. (1996), the currently used Bayesian method has not received much consideration, mainly because the database and the incorporation of ancillary data were deemed to be more relevant for improving the accuracy of the retrieval. However, two disadvantages of the current retrieval method have become apparent with the introduction of the much larger, observationally-generated retrieval databases into the algorithm (Elsaesser and Kummerow,

55 2015): Firstly, the retrieval database must be compressed into self-similar clusters to make calculation of the retrieval results sufficiently fast. This lossy compression may limit the extent to which the current algorithm can benefit from the size and representativeness of observationally generated retrieval databases. This is expected to affect retrievals of high rain rates in particular due to their scarcity in the retrieval database. Secondly, the accuracy of the retrieval results depends on the uncertainties

assigned to the database observations. Since there is no principled way to calculate these uncertainties, they need to be tuned
60   heuristically for each sensor.

This study investigates the benefits of using the GPROF retrieval database to train a neural network to retrieve precipitation and hydrometeor profiles. Since the retrieval database has grown to a size of several hundred million entries it is perfectly suited for the application of deep neural networks, which scale very well to large amounts of data and are capable of learning complex relationships from them. In addition to this, a neural network based implementation has the advantage of allowing
65   the integration of spatial information into the retrieval, which is not as easily feasible with the current method. This study thus aims to answer the following two questions:

1. Can a deep learning based retrieval method with identical inputs improve the accuracy of retrieved surface precipitation and vertical hydrometeor profiles?

2. Can the incorporation of spatial information into the retrieval help to improve the retrievals?

70   To answer these questions, this study presents two novel, neural network based implementations of the GPROF algorithm:

GPROF-NN 1D: This algorithm uses a fully-connected neural network to retrieve single column hydrometeor profiles and rain rates based on the observed brightness temperature vector. Its inputs and outputs are thus identical to those of the standard GPROF algorithm.

GPROF-NN 3D: This algorithm extends the GPROF-NN 1D algorithm by incorporating spatial information into the retrieval.
75   It produces the same output as GPROF and GPROF-NN 1D but processes all observations simultaneously, thus allowing the algorithm to combine information from pixels across the swath.

The retrieval performance of the two GPROF-NN implementations is assessed on a held-out part of the retrieval database and compared to that of the upcoming version of GPROF. This new version of GPROF will be designated as GPROF 2021 or GPROF V7 in the GPM processing system (NASA, 2021). The retrieval performance is assessed for two sensors of the
80   GPM constellation: The GPM Microwave Imager (GMI) and the Microwave Humidity Sounder (MHS, Bonsignori (2007)). In addition to this, a case study is presented that compares the retrieved surface precipitation from overpasses of the two sensors over Hurricane Harvey.

2   Data and methods

This section introduces the retrieval database that is used by the GPROF algorithm and serves as training data for the GPROF-
85   NN algorithms. This is followed by a description of the current implementation of GPROF as well as the implementation of the GPROF-NN retrievals.

Atmospheric
Measurement
Techniques
Discussions

Table 1. Retrieval quantities in the retrieval database

Retrieval variable	Unit	Type
Surface precipitation	mm h^{-1}	Scalar
Convective precipitation	mm h^{-1}	Scalar
Cloud water path	kg m^{-2}	Scalar
Rain water path	kg m^{-2}	Scalar
Ice water path	kg m^{-2}	Scalar
Cloud water content	g m^{-3}	Profile
Rain water content	g m^{-3}	Profile
Snow water content	g m^{-3}	Profile
Latent heating	K h^{-1}	Profile

2.1 The retrieval database

The GPROF retrieval database consists of pairs of retrieval inputs paired with corresponding retrieval outputs. The retrieval inputs of GPROF consist of PMW observations and ancillary data, while the outputs consist of the values of the retrieved variables. For simplicity, we refer to these pairs as 'database profiles' or simply 'profiles' due to them being derived from or corresponding to specific vertical profiles of atmospheric variables. GPROF's retrieval outputs include surface precipitation, profiles and path integrals of rain, snow and cloud water as well as latent heating profiles. A listing of all retrieval targets and corresponding units is provided in Tab. 1.

Values of the retrieval targets for the database of GPROF 2021 are derived from one year (October 2018 to September 2019) of retrieved hydrometeor profiles from the GPM combined product (Grecu et al., 2016). This data is complemented with surface precipitation from the currently operational Microwave Integrated Retrieval System (Boukabara et al., 2011) that adds light precipitation in areas where no echo is detected by the GPM Dual-Frequency Precipitation Radar. Observations over sea-ice and snow-covered surfaces are handled separately. For sea-ice, precipitation is derived from the ERA5 reanalysis (Hersbach et al., 2020). For snow-covered surfaces, precipitation is derived from five years of co-locations with gauge-corrected radar observations from the Multi-Radar Multi-Sensor (MRMS, Smith et al., 2016) product.

The retrieval targets are then combined with observations from each of the sensors of the GPM constellation: For GMI, this is done by remapping and averaging the retrieval outputs to the footprint of GMI's 18.7 GHz channel. Since co-locations with other sensors of the constellation are rare or not available at all, observations for these sensors are simulated and the footprint averaging is adapted to the expected sensor resolution.

The ancillary data that serve as additional retrieval inputs are derived from reanalysis datasets. They consist of two meter temperature (T_{2m}), total column water vapor (TCWV), a surface type as well as an air lifting index (ALI) that encodes information on atmospheric convergence in mountainous areas. The ancillary data for the databases used in this study were derived from the ERA5 reanalysis (Hersbach et al., 2020).

2.2 The GPROF algorithm

110 The current implementation of GPROF retrieves precipitation and hydrometeor profiles by resampling the profiles in the database based on the similarity of the observations to the input observations. In order to limit the amount of profiles to search through, the database is split up into bins with respect to the ancillary data. In addition to this, the profiles in each bin are clustered to further limit the number of profiles that need to be processed. An illustration of these steps and the Bayesian retrieval scheme, all of which will be explained in more detail below, is provided in Fig. 1.
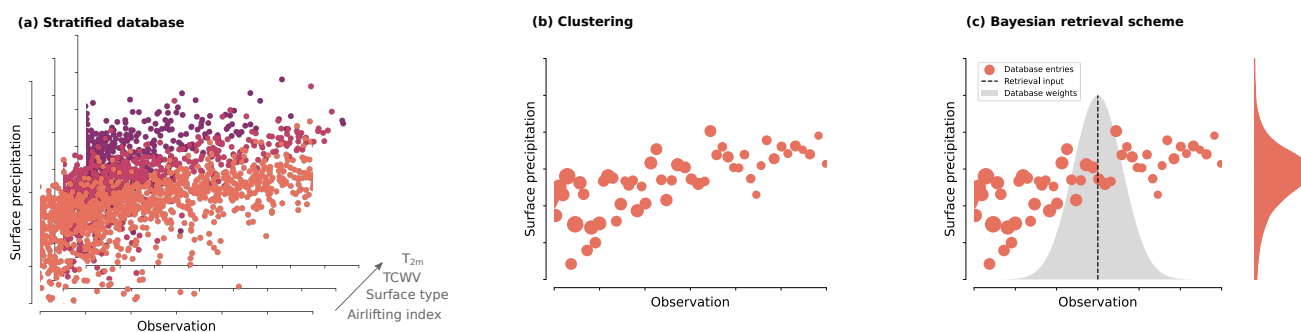

Figure 1. Components of the GPROF retrieval algorithm. Panel (a) illustrates the binning of the database with respect to the ancillary data consisting of two meter temperature (T_{2m}), total column water vapor (TCWV), surface type and airlifting index. Panel (b) illustrates the clustering of each database bin into self-similar clusters with the size of the markers representing the number of profiles in each cluster. Panel (c) illustrates the Bayesian scheme that is used to approximate the posterior distribution of the retrieval, which corresponds to the filled curve to the right, by weighting the samples in the database.

115 The binning of the profiles in the database is performed with respect to all ancillary variables, that is T_{2m}, TCWV, surface type and airlifting index (Fig. 1 (a)). The binning is hard in the sense that only profiles from the single bin corresponding to the ancillary data are considered in the retrieval. The exception to this rule are the edges of the database with respect to T_{2m} and TCWV for which the profiles are combined with those from neighboring bins until at least 2000 profiles are found. To further limit the amount of data to process during the retrieval, the profiles in each database bin are clustered into self-similar clusters 120 and only the mean of the observations and retrieval targets as well as the number of observations is retained (Fig. 1 (b)).

The binning and clustering of the database is performed offline, i.e. during the algorithm development. To retrieve precipitation from the binned database GPROF employs the following Bayesian retrieval scheme.

Given a vector of observations $\mathbf{y}$ and corresponding ancillary data, the retrieval first uses the ancillary data to select the corresponding database bin. Let $(\mathbf{y}_1, \mathbf{x}_1, n_1), \ldots, (\mathbf{y}_N, \mathbf{x}_N, n_N)$ denote the profile clusters in that database bin with $\mathbf{y}_i$ and $\mathbf{x}_i$ 125 the centroids of the observation and state vector in the ith cluster and n_i the number of profiles in the cluster. Assuming that the un-clustered profiles in the database bin are distributed according to the true a priori distribution $p(\mathbf{x})$, i.e. the climatology of the atmospheric state $\mathbf{x}$ corresponding to the given bin, the expected value of $\mathbf{x}$ with respect to the posterior distribution

$p(\mathbf{x}|\mathbf{y})$ can be approximated using

$$\int_{\mathbf{x}} \mathbf{x} p(\mathbf{x}|\mathbf{y})\,d\mathbf{x} = \int_{\mathbf{x}} \mathbf{x}\,\frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}\,d\mathbf{x} \approx \frac{\sum_i n_i p(\mathbf{y}|\mathbf{x}_i)\mathbf{x}_i}{\sum_i p(\mathbf{y}|\mathbf{x}_i)}. \tag{1}$$

130   The conditional probability $p(\mathbf{y}|\mathbf{x}_i)$ of the input observation $\mathbf{y}$ given atmospheric state $\mathbf{x}_i$ is taken as the probability of the deviations of $\mathbf{y}$ from the observations $\mathbf{y}_i$ corresponding to the state $\mathbf{x}_i$ to be caused by the random error in the observations, which is assumed to be unbiased and Gaussian:

$$p(\mathbf{y}|\mathbf{x}_i) = \frac{n_i}{\sqrt{2\pi \det(\mathbf{S}^{-1})}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{y}_i)^T \mathbf{S}^{-1}(\mathbf{y}-\mathbf{y}_i)\right\} \tag{2}$$

where $\mathbf{S}$ is a diagonal covariance matrix. The observation error represents sensor noise as well as other causes of deviations of
135   real observation from the observations in the database, such as calibration errors or modeling errors. It should be noted here, that the assumption of Gaussian errors with state independent, diagonal covariance matrix is made for simplicity but likely insufficient to accurately describe modeling errors that are state dependent and correlated between channels. As illustrated in Fig 1 (c), Eq. 1 corresponds to a resampling of the states in the database with case-specific weights calculated using Eq. 2. This approach can be extended to approximate the probability density function of the posterior distribution or to derive probabilities
140   of certain characteristics of the a posteriori state such as the presence of precipitation in a given observation.

2.3   The GPROF-NN algorithms

The principal objective guiding the design of the GPROF-NN algorithms was to develop a neural network retrieval that provides the exact same output as GPROF so that it can potentially replace the current implementation in a future update. GPROF produces several outputs that are probabilistic: A probability of precipitation as well as the mode and terciles of the posterior
145   distribution of precipitation. An implementation based on standard regression neural networks would not provide any principled way to produce these probabilistic outputs due to incompatibility of deterministic regression with the Bayesian retrieval formulation used in GPROF. To overcome this limitation, the implementation of the GPROF-NN retrievals is based on quantile regression neural networks (QRNNs). As has been shown in Pfreundschuh et al. (2018), when trained on a dataset that is distributed according to the a priori distribution of a Bayesian retrieval, QRNNs learn to predict quantiles of the posterior
150   distribution. They thus provide a simple and efficient way to reconcile neural network retrievals with the Bayesian framework that is applied in GPROF.

The QRNN approach can be generalized to predict a sequence of quantiles, which can be used to reconstruct the cumulative distribution function (CDF) of the a posteriori distribution of any scalar retrieval quantity. Since the distribution of a scalar variable is fully described by its CDF, the predicted CDF can be used to derive any relevant statistic of the a posteriori distribu-
155   tion. For the GPROF-NN retrievals, the predicted CDF is used to derive most likely and mean surface precipitation (the latter of which is identical to the solution that would have been obtained with common mean squared error regression), the terciles of the posterior distribution as well the probability of precipitation. An illustration the principle of the GPROF-NN retrievals is provided in Fig. 2: The retrieval employs a neural network to predict for each input pixel a vector of values. The elements

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

of this vector correspond to a sequence of quantiles of the a posteriori distribution. These quantiles are used to reconstruct a

160   piece-wise linear approximation of the CDF of the distribution from which the final retrieval outputs are derived.
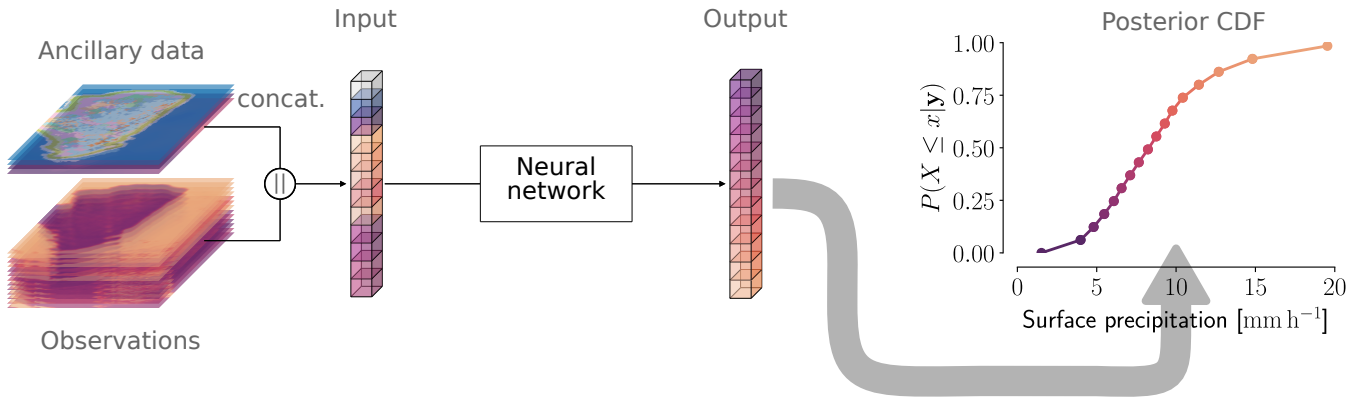

Figure 2. The basic principle of the implementation of the GPROF-NN retrievals. A Bayesian solution of the retrieval is obtained by predicting, for each input pixel, a sequence of quantiles of the a posteriori distribution that is used to reconstruct its CDF. The predicted CDF is then used to derive the scalar retrieval results.

2.3.1   Training objectives

A neural network can be trained to predict a quantile $\hat{x}_\tau$ of a given conditional distribution by training it to minimize the quantile (or pinball) loss function $\mathcal{L}_\tau$ corresponding to the quantile fraction τ (Koenker and Hallock, 2001):

$$\mathcal{L}_\tau(\hat{x}_\tau, x) = (\tau - \mathbb{I}_{x \leq \hat{x}_\tau})(x - \hat{x}_\tau), \tag{3}$$

165   where $\hat{x}_\tau$ is the predicted quantile, x is the reference value from the training data and $\mathbb{I}_{x \leq \hat{x}}$ is the indicator function taking the value 1 when the condition $x \leq \hat{x}$ is true and 0 otherwise.

This principle can be extended to a sequence of quantiles corresponding to quantile fractions $\tau_1, \ldots, \tau_N$ by minimizing the mean of the loss functions corresponding to each quantile fraction:

$$\mathcal{L}_{\tau_1, \ldots, \tau_N}(\hat{\mathbf{x}}, x) = \frac{1}{N} \sum_{i=0}^{N} \mathcal{L}_{\tau_i}(\hat{x}_i, x) \tag{4}$$

170   where $\hat{x}_i$ is the ith component of the vector of predicted quantiles $\hat{\mathbf{x}}$. The GPROF-NN retrievals use this loss function with 128 equally spaced quantiles ranging from $\tau_1 = 0.001$ to $\tau_{128} = 0.999$ for all scalar retrieval variables.

A difficulty with predicting quantiles of precipitation is that that lower quantiles may become degenerate due to the high probability of no precipitation. For example, it is impossible to predict empirical quantiles with $0 < \tau < 0.5$ for a pixel with $50\,\%$ probability of precipitation. To allow monitoring of the ability of the network to correctly predict retrieval uncertainty up

175   to the degeneracy induced by non-raining pixels, we replace rain rates of non-raining pixels with random values from a log-uniform distribution that are smaller than the smallest rain rate in the training data. During the retrieval, predicted precipitation

rates that are smaller than this threshold are set to zero. The threshold is chosen as 10^{-4} mm h^{-1} and thus has negligible impact on mean or accumulated precipitation.

An additional advantage of the application of the quantile loss function is that the training can be performed on transformed
180   retrieval outputs without changing the statistical properties of the network predictions given that the transformation function is strictly monotonic. The training of all scalar, non-negative retrieval quantities uses a log-linear transformation function of the form

$$f(x) = \begin{cases} \log(x) & \text{if } x < 1 \\ x + 1 & \text{if otherwise.} \end{cases} \tag{5}$$

In addition to avoiding the prediction of negative values, we found this to slightly increase retrieval accuracy for quantities that
185   vary by multiple orders of magnitude, which precipitation rates and hydrometeor concentrations typically do.

For hydrometeor profiles, the retrieval is implemented in a slightly different manner. To reduce the number of network outputs, the posterior mean of hydrometeor profiles is predicted directly using mean squared error regression. Since the output of GPROF contains only the posterior mean of the hydrometeor concentrations, it was deemed unnecessary to predict their full posterior distribution at each level using quantile regression. To avoid the prediction of negative concentrations ReLU
190   activation functions are applied to the network outputs corresponding to hydrometeor concentrations.

2.3.2   GPROF-NN 1D

The GPROF-NN 1D algorithm was designed to be fully identical to the current GPROF algorithm. This means it not only produces the same outputs, but it also uses the same inputs, namely PMW observations from a single pixel together with the corresponding ancillary data.

195   The neural network architecture used for the GPROF-NN 1D retrieval is illustrated in Fig. 3. A single network is trained to predict all retrieval variables (c.f. Tab. 1) using the training objectives described in Sec. 2.3.1. The network consists of a shared body and a separate head for each retrieved variable. Body and heads are built-up of blocks consisting of a fully-connected layer followed by layer normalization (Ba et al., 2016) and GELU (Hendrycks and Gimpel, 2016) activation functions. During development we have experimented with different numbers of blocks in body (N_b) and each of the heads (N_h) but found only
200   marginal impact on the retrieval performance and settled for a configuration with $N_b = 6$ and $H_h = 4$.

The GPROF-NN 1D network is trained by simultaneously minimizing the sum of the losses of all retrieval variables. The training is performed over 70 epochs using the Adam optimizer (Kingma and Ba, 2014) with an initial learning rate of $5 \cdot 10^{-4}$ and a cosine annealing learning rate schedule (Loshchilov and Hutter, 2016). Warm restarts are performed after 10, 30 and 50 epochs.
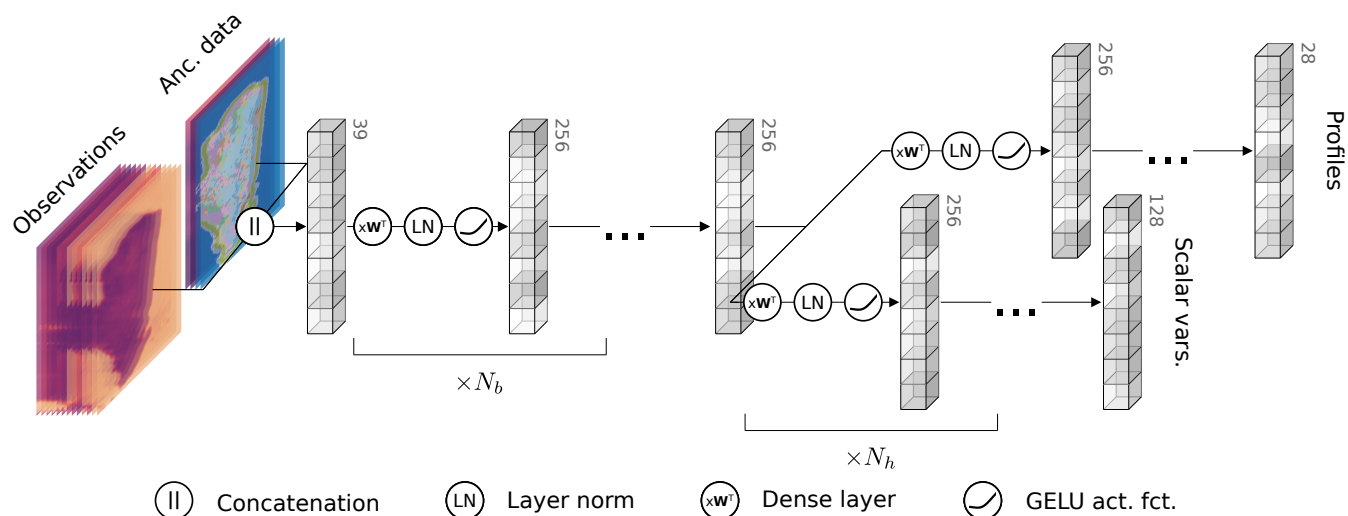
Figure 3. Illustration of the neural network architecture used in the GPROF-NN 1D algorithm. The network consists of a common body and one head for each retrieval variable. Each block in body and head consists of fully-connected layer, layer norm and GELU activation function.
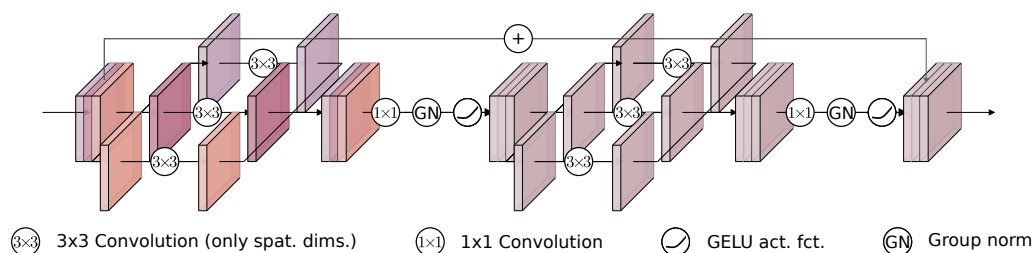
2.3.3 GPROF-NN 3D

The GPROF-NN 3D retrieval extends the GPROF and GPROF-NN 1D algorithms by incorporating structural information into the retrieval. To achieve this, the GPROF-NN 3D algorithm employs a convolutional neural network (CNN) that performs the retrieval for all pixels in the swath simultaneously.

The network architecture for the GPROF-NN 3D algorithm, illustrated in Fig. 4, consists of an asymmetric encoder-decoder structure followed by a separate head for each retrieved variable. The stages of the en- and decoder are built up of what we refer to here as Xception blocks (Fig. 4 (a)) because they are based on the Xception architecture introduced in Chollet (2017). Each block consists of two depthwise separable convolutions with a kernel size of 3 followed by group normalization layers with 32 groups and GELU activation functions. The first block in each stage of the encoder additionally contains a 3×3 max-pooling layer with a stride of 2 following the first 3×3 convolution layer. Each downsampling block in the encoder is followed by $N = 4$ standard Xception blocks. The stages of the decoder consist of a bi-linear upsampling layer followed by a single Xception block. The network architecture was chosen with the aim of maximizing the depth and width of the network while keeping the time required for processing an orbit low. Symmetric padding is performed before all convolution operations with a kernel size larger than one in order to conserve the input size.

The fully-convolutional network architecture employed by the GPROF-NN 3D retrieval enables the model to process inputs of arbitrary sizes. This allows processing of a whole orbit of observations as a single input, thus omitting the need for splitting up the observations into overlapping tiles to avoid edge effects.

Atmospheric
Measurement
Techniques
Discussions

Open Access
EGU

(a) Xception block

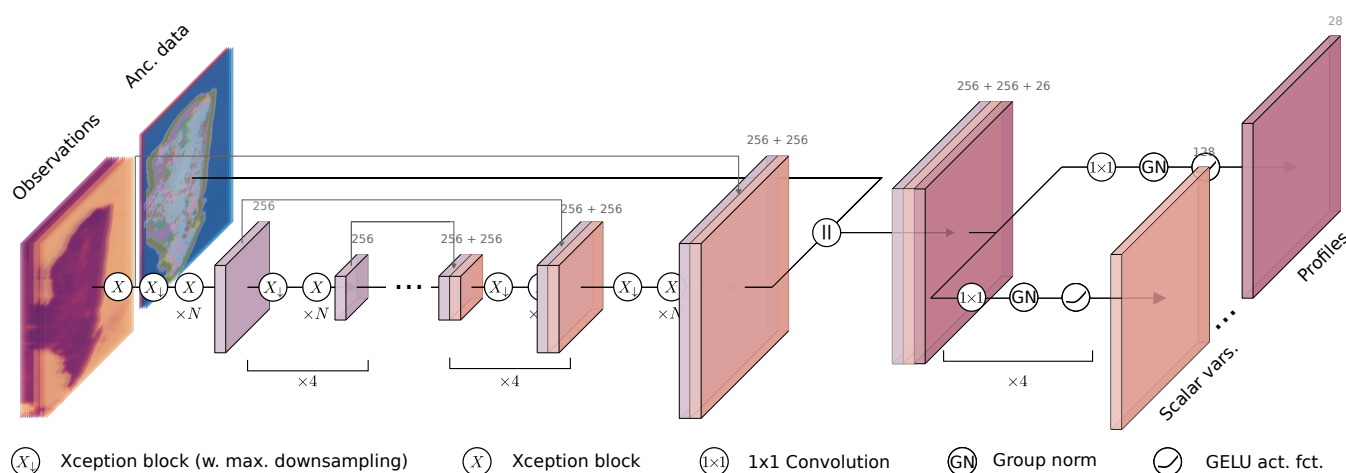**(b) GPROF-NN 3D architecture**

Figure 4. The neural network architecture of the GPROF-NN 3D retrieval. Illustration (a) displays the structure of the Xception blocks (Chollet, 2017) that form the building blocks of the GPROF-NN 3D model. An Xception block consists of two depthwise separable convolutions followed by a group normalization layer and a GELU activation function. Illustration (b) shows how the Xception blocks are used in an asymmetric encoder-decoder structure that forms the body of the network. Output from the body is combined with the ancillary data to form the inputs to the separate heads that predict the retrieval results for each of the retrieved variables.

Due the special structure of the GPROF retrieval database, it was necessary to adapt the training scheme for the GPROF-NN 3D retrieval. For the most part, the retrieval database contains values for the retrieved variables only at the 20 central pixels of the GMI swath, as illustrated by the light stripe in Fig. 5 (a). This means that, although the network predicts the retrieved

225 variables at all pixels in the swath, the loss can be calculated only where the values of the retrieval variables are known. In addition to this, the viewing geometry of PMW sensors breaks the translational symmetry that is common for digital images and constitutes one of the inductive biases of CNNs (Goodfellow et al., 2016). This is because the distance between pixels, as well as their orientation, varies across the swath of the sensor. An illustration of this is provided in Fig. 5 for the conical viewing geometry of GMI. When displayed on a uniform grid, similar shapes appear differently depending on their position across the

230 swath. To account for this in the training, training samples are transformed to simulate the effect of seeing a given scene at different locations across the GMI swath. These transformations are applied randomly when a training sample is loaded thus
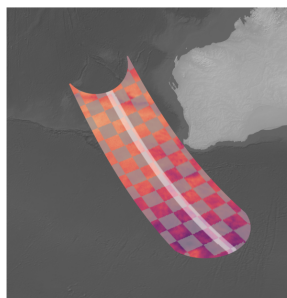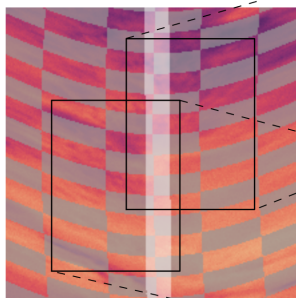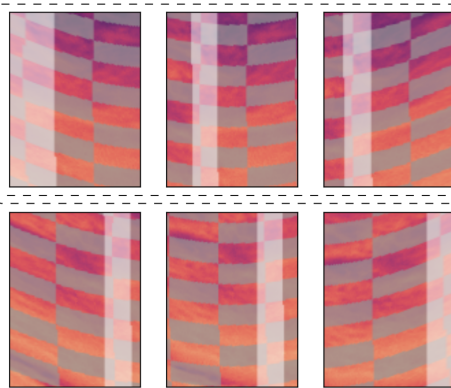
Atmospheric
Measurement
Techniques
Discussions

Figure 5. The effect of GMIs conical viewing geometry on observed features. Panel (a) displays geolocated observations of the 10.6 GHz channel. Grey squares mark equilaterals with a side length of 200km oriented along the swath. The highlighted stripe located at the swath center marks the region where the values of the retrieved variables are known. Panel (b) shows the same observations viewed as an image on a uniform grid. Panel (c) shows six synthetically generated training inputs based on two input regions marked in Panel (b). The first row shows three synthetic samples that simulate the effect of viewing the input in region A at a different position across the GMI swath. The second row shows the corresponding transformations for the input in region B.

ensuring that the network rarely or never sees a training scene from the same perspective. The transformations were furthermore designed to ensure that the relative location of the pixels at which values of retrieval variables are known varies across the full width of the swath instead of always being located at its center. Examples of transformed inputs are displayed in Fig. 5 (c).

235 **2.3.4 Extension to other sensors**

The GPROF retrieval for GMI is special because GMI is the only sensor for which the retrieval inputs used in the database correspond to real observations. For the other sensors of the GPM constellation, the observations used to construct the retrieval database are simulated. For conically scanning sensors a single simulation is performed for each profile in the database. For cross track scanning sensors several simulations are performed for each profile covering the range of viewing angles of the
240 sensor.

Since the databases for other sensors are based on simulations, it is not guaranteed that the distribution of brightness temperatures in the database matches those of real observations. The simulated observations are therefore corrected using a surface type and total column water vapor dependent correction that matches the quantiles of the conditional distributions of simulated and real observations. The GPROF algorithm uses a slightly different correction that distinguishes only 3 different surface
245 types, while the GPROF-NN algorithms use a correction with all 18 surface types. This is because the correction used by GPROF was found to be too crude over land surfaces leading to artifacts in the neural network results.

The training of the GPROF-NN 3D algorithm for other sensors than GMI is further complicated by the fact that the retrieval database for these sensors contains simulated observations only at the central pixels of the GMI swath (the highlighted pixels in Fig. 5 (a)). To obtain two-dimensional training scenes that are sufficiently wide to train a CNN, we make use of an intermediate 250 CNN based model to 'retrieve' simulated brightness temperatures across the full GMI swath. The extended simulated brightness temperatures are then remapped from the GMI viewing geometry to the viewing geometry of the target sensor. While this approach is certainly not ideal with respect to the realism of the generated scenes, it was the simplest and currently only feasible way to extend the GPROF-NN 3D retrieval to other sensors than GMI using only currently available data from the GPROF database.

3 Results

The first part of this section presents the evaluation GPROF and the GPROF-NN algorithms on a held-out test data set. The test data set consists of observations from the first three days of every month from the retrieval database. It should be noted that we have deliberately limited this evaluation to data from the retrieval database in order to isolate the effect of the retrieval algorithm from that of the database. We conclude this section with a case study of overpasses of Hurricane Harvey. These 260 results are based on real observations and thus provide an indication to what extent the performance on the retrieval database can be expected to generalize to real observations.

3.1 Precipitation and water columns

As described in Tab. 1, the scalar variables retrieved by GPROF are surface and convective precipitation as well as the column-integrated concentrations of cloud droplets, rain and snow, denoted as cloud water path (CWP), rain water path (RWP) and ice 265 water path (IWP), respectively. Scatter plots of the retrieval results for these five quantities are displayed in Fig. 6 for GMI and Fig. 7 for MHS. The frequencies in all plots have been normalized column-wise to ensure that results for high reference values remain visible.

For surface precipitation retrieved by GMI, consistent improvements in retrieval accuracy are observed between the GPROF and GPROF-NN 1D as well as GPROF-NN 1D and GPROF-NN 3D. The improvements are most pronounced for very small 270 rain rates between 10^{-2} and 10^{-1} mm h^{-1} but are consistent across the full range of displayed values. The comparably bad performance of GPROF for small precipitation values is likely due to the tuning of the assigned uncertainties to yield good results for larger rain rates that are more relevant for rainfall accumulation.

For convective precipitation, the results of GPROF deviate noticeably from the diagonal. The GPROF-NN 1D slightly improves upon the results of GPROF. Although the mode of the distribution is still displaced from the diagonal, the GPROF-275 NN 3D algorithm yields the best agreement with the reference data. For the path-integrated quantities, similar improvements between GPROF and GPROF-NN 1D as well as GPROF-NN 1D and GPROF-NN 3D are observed. Large cloud water path values are underestimated by all retrievals which is likely because these values are associated with precipitation but difficult
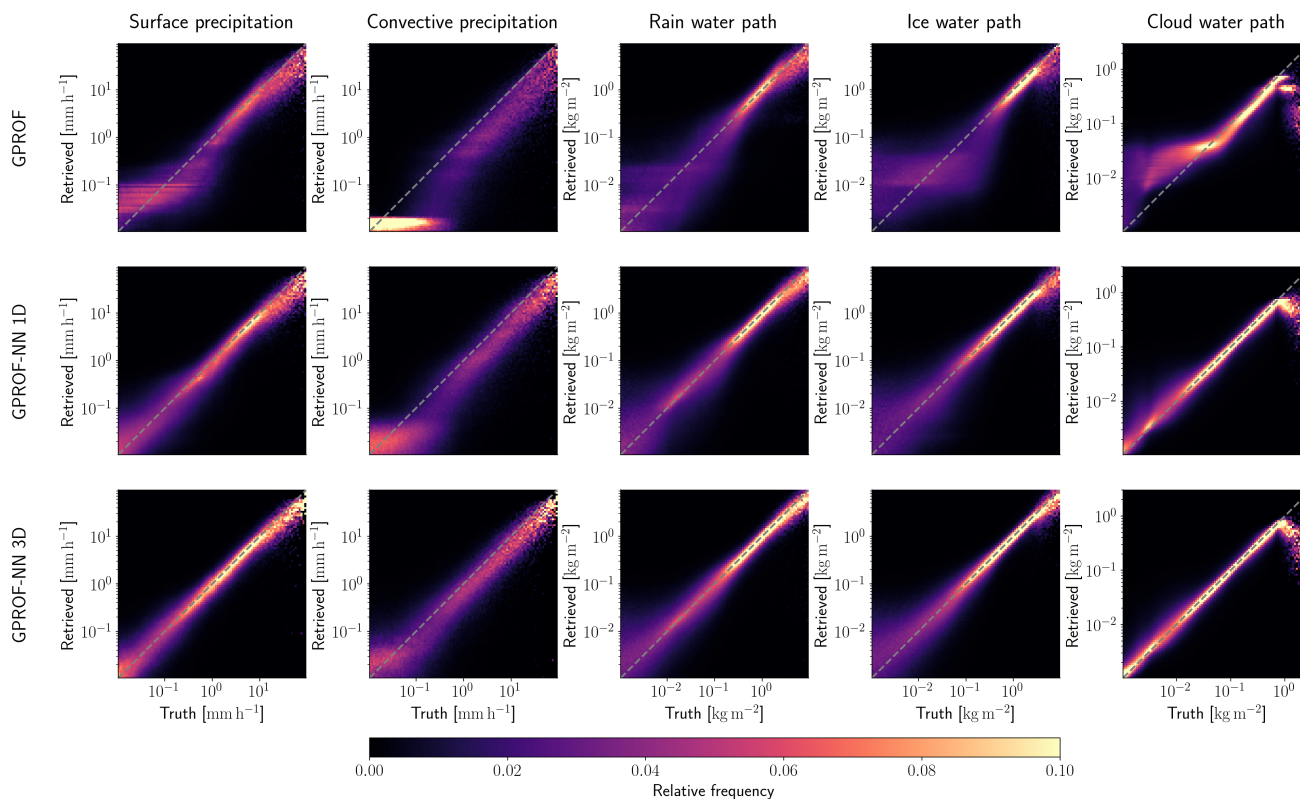
Figure 6. Scatter plots of scalar retrieval targets for the three retrieval algorithms for GMI. Rows display the results for the GPROF, GPROF-NN 1D and GPROF-NN 3D algorithms, respectively. Columns display the results for different retrieval targets. Frequencies in the plots have been normalized column-wise, i.e. per bin of the reference value.

to distinguish from it. Due to the lack of a cloud water path signal in raining profiles, all algorithms resort to predicting the climatology in the presence of significant rain.

280   The results for MHS, displayed in Fig 7, paint a similar picture. Although the overall accuracy of the retrieval is lower, the GPROF-NN 1D retrieval consistently yields more accurate results than GPROF. Also here the GPROF-NN 3D retrieval yields further, consistent improvements compared to the GPROF-NN 1D retrieval.

Quantitative measures of the retrieval accuracy for surface precipitation of the three retrieval algorithms are displayed in Tab. 2 for GMI and Tab. 3 for MHS. Similar tables for the other retrieval quantities are provided in Tab. A1-A8 in the appendix.

285   Each table displays bias, mean absolute error (MAE), mean squared error (MSE) and the symmetric mean absolute percentage error (SMAPE$_t$) for all test samples with a reference value that exceeds a quantity-specific threshold t. The error metrics confirm the qualitative findings from Fig. 6 and 7: The neural network implementations outperform GPROF in terms of all considered metrics. Moreover, the GPROF-NN 3D algorithm further improves upon the performance of the GPROF-NN 1D algorithm. Although the errors are larger for MHS, the same tendency is observed.
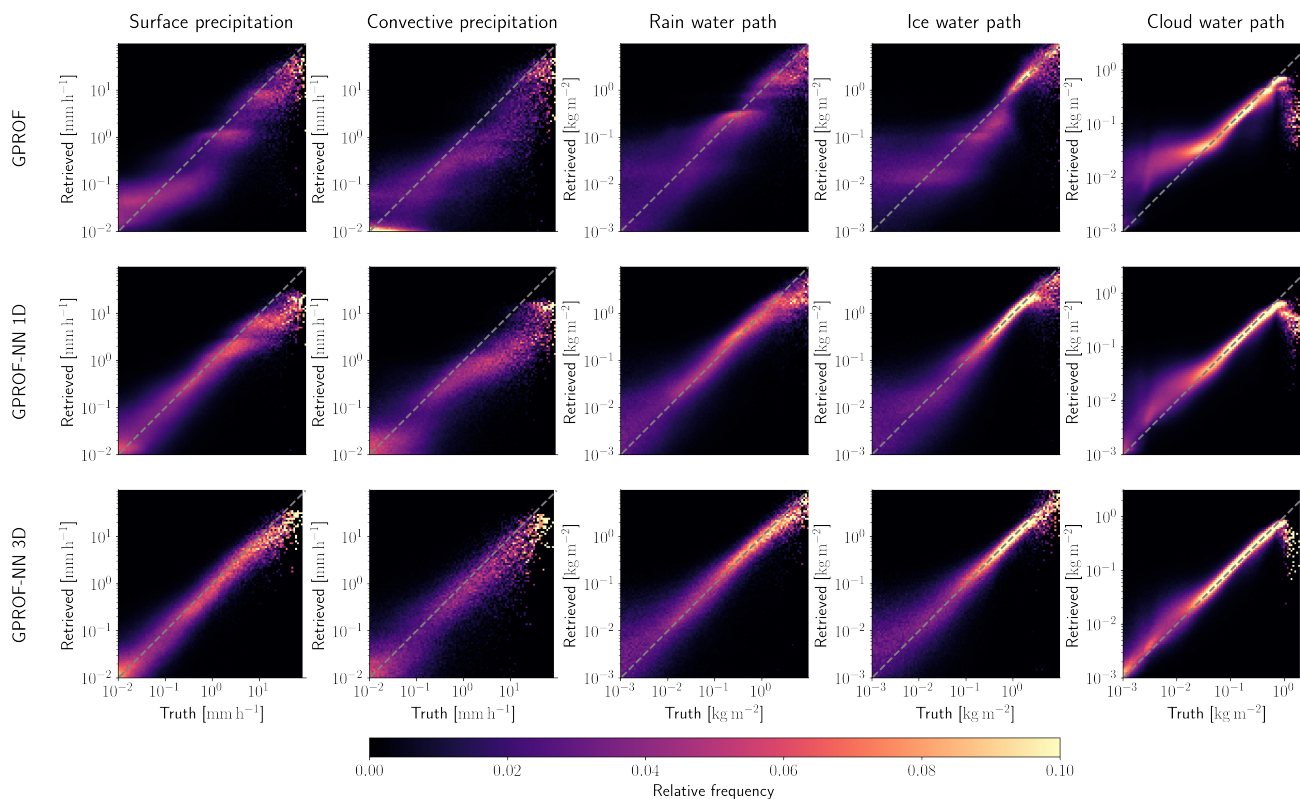
Figure 7. Like Fig. 6 but for MHS.

Table 2. Mean error metrics and estimated standard deviation for surface precipitation retrieved from GMI observations.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[mm\,h^{-1}]$	-0.0040 ± 0.0001	-0.0024 ± 0.0001	-0.0006 ± 0.0001
MAE $[mm\,h^{-1}]$	0.0785 ± 0.0001	0.0585 ± 0.0001	0.0444 ± 0.0001
MSE $[mm\,h^{-1}]$	0.1830 ± 0.0001	0.1379 ± 0.0001	0.0983 ± 0.0001
$SMAPE_{0.01}$ $[\%]$	76.9023 ± 0.0138	69.5382 ± 0.0127	56.0040 ± 0.0181

Table 3. Mean error metrics and estimated standard deviation for surface precipitation retrieved from MHS observations.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[mm\,h^{-1}]$	0.0070 ± 0.0001	-0.0053 ± 0.0001	0.0017 ± 0.0002
MAE $[mm\,h^{-1}]$	0.0948 ± 0.0001	0.0610 ± 0.0001	0.0524 ± 0.0002
MSE $[mm\,h^{-1}]$	0.3078 ± 0.0002	0.2088 ± 0.0002	0.1373 ± 0.0002
$SMAPE_{0.01}$ $[\%]$	80.9690 ± 0.0192	70.1140 ± 0.0189	63.4292 ± 0.0414

Atmospheric
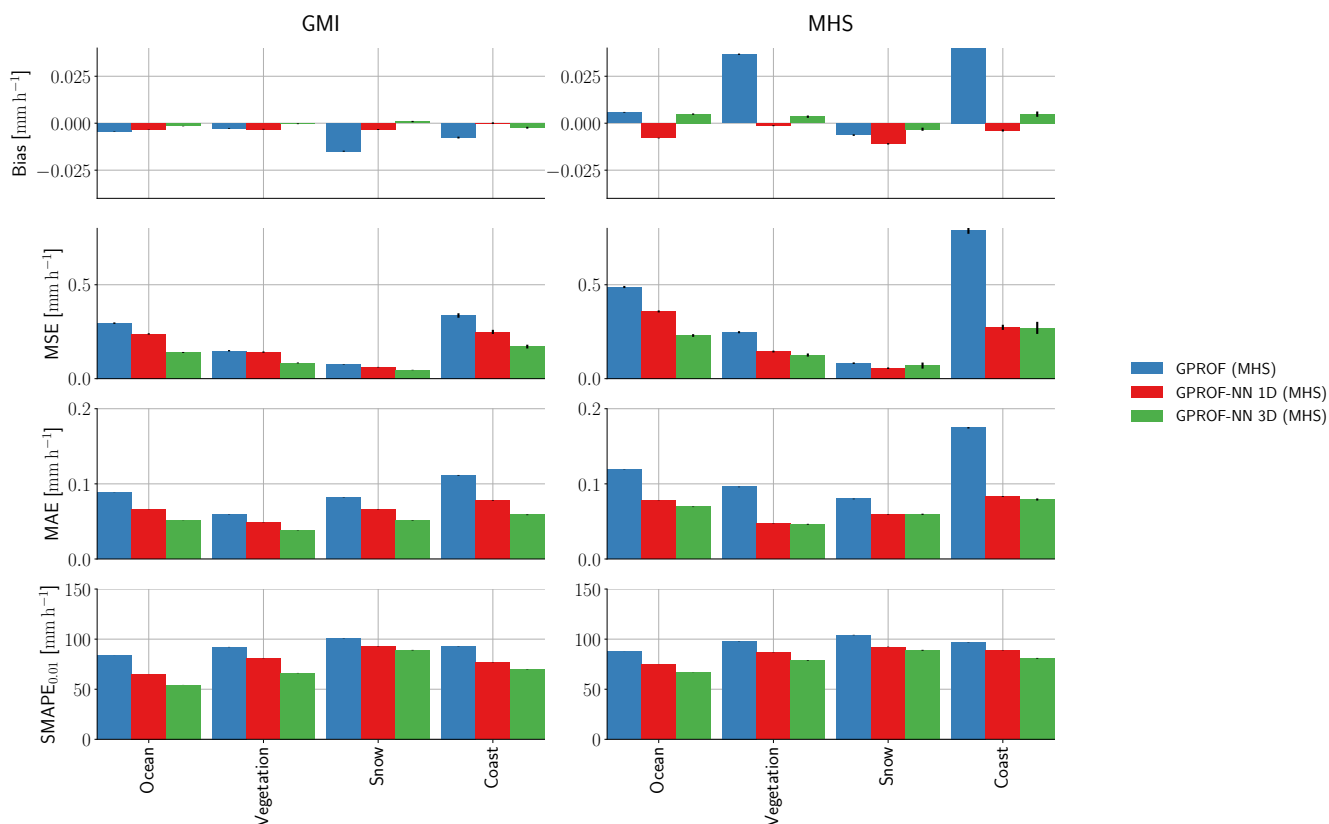Measurement
Techniques
Discussions

Open Access

EGU

Figure 8. Bias, MSE, MAE and SMAPE$_{0.01}$ of the retrieved surface precipitation w.r.t. surface type and retrieval algorithm. The results for the GMI sensor are displayed in the first column, results for the MHS sensor in the second column. Error bars mark one standard deviation around the mean calculated using bootstrap and thus quantify the uncertainty due to the random sampling of the test data.

290   Since the surface type has a considerable effect on the lower-frequency observations used in the retrieval, its impact on the retrieval of surface precipitation is assessed in Fig. 8. The figure displays bias, MSE, MAE and SMAPE for the most relevant classes of surface types. Also, when different surface types are considered separately, the results of the surface precipitation retrieval show the same pattern as the scatter plots in Fig. 6. The results of the GPROF-NN 1D retrieval are generally more accurate than those of GPROF and the results of the GPROF-NN 3D algorithm are in turn more accurate than those of the

295   GPROF-NN 1D algorithm. These findings are mostly consistent across the considered surface types and both sensors. Exceptions are the biases of the GPROF-NN 1D algorithm for GMI over vegetated surface and for MHS over snow, which are larger than those of GPROF as well as the MSE of the GPROF-NN 3D algorithm for MHS over snow, which is slightly larger than that of the GPROF-NN 1D retrieval. We suspect that this is caused by the relative scarcity of the observations in the retrieval database.

300   Figure 9 displays the geographical distribution of bias, MSE and MAPE for GMI in $5\,^\circ \times 5\,^\circ$ boxes. As could be expected from the previous results, the magnitudes of the biases of GPROF are considerably larger than for the other two algorithms.

Furthermore, GPROF exhibits consistent biases across geographical regions such as the Northwest Atlantic and Northwest Pacific, which is less the case for the neural network algorithms. Although spatial distribution of the MSE mostly reflects the global distribution of precipitation, a gradual decrease in MSE can be observed between the results of GPROF and GPROF-NN 1D as well as GPROF-NN 1D and GPROF-NN 3D. More consistent patterns are visible in the SMAPE: The largest errors for all three retrievals occur over land surfaces, which likely reflects the decrease in information content due to the reduced contrast in the lower frequency channels. Over Ocean, errors are generally higher in the sub-tropics and tropics compared to higher latitudes. Although these patterns are observed in the results of all algorithms, a clear and globally consistent decrease in SMAPE can be observed between the GPROF, GPROF-NN 1D and GPROF-NN 3D retrievals.

The corresponding results for MHS are provided in Fig. A1. Again, although the errors are slightly larger, the results are qualitatively similar.

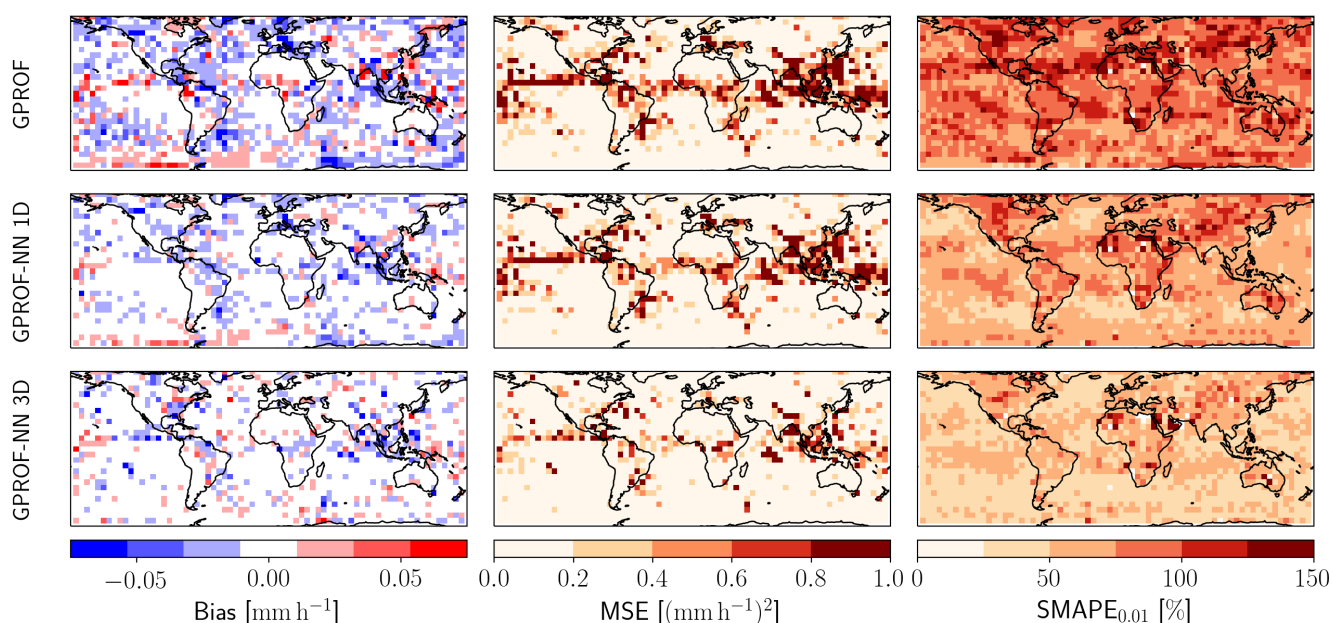
Figure 9. Spatial distributions of bias (Column 1), MSE (Column 2) and SMAPE$_{0.01}$ (Column 3) for the GPROF (Row 1), GPROF-NN 1D (Row 2), GPROF-NN 3D (Row 3) algorithms for GMI in $5\,° \times 5\,°$ grid boxes.

3.1.1 Predicted retrieval uncertainties and probabilistic rain detection

In addition to quantitative precipitation estimates, GPROF produces estimates of the first and second tercile of the posterior distribution of surface precipitation, which provide an uncertainty estimate for the retrieved mean surface precipitation, as well as a probabilistic classification of pixels into raining and non-raining pixels based on an estimated probability of precipitation. Due to their probabilistic nature, similar results can be produced using GPROF-NN algorithms. From the predicted quantiles, the terciles can be predicted by interpolating them to the fractions $\frac{1}{3}$ and $\frac{2}{3}$, respectively. The probability of precipitation is

Table 4. Calibration of the predicted terciles of the posterior distribution of surface precipitation for the three retrieval algorithms and the GMI and MHS sensors.

Tercile	Nominal	GMI			MHS		
		GPROF	GPROF-NN 1D	GPROF-NN 3D	GPROF	GPROF-NN 1D	GPROF-NN 3D
First	0.333	0.461	0.351	0.349	0.274	0.332	0.335
Second	0.667	0.514	0.652	0.654	0.480	0.663	0.659

calculated by using the predicted posterior distribution to calculate the probability of the retrieved surface precipitation to be larger than the smallest non-zero rain rate in the training data.

320   To assess the accuracy of the uncertainty estimates from GPROF and the GPROF-NN algorithms, Tab. 4 lists the calibration, i.e. the frequency with which each predicted tercile was larger than the true surface precipitation. The evaluation of the results for the GPROF-NN algorithms was performed with the replacement of non-raining values described in Sec. 2.3.1, which allows to account for degenerate quantiles. Since no such mechanism is available for GPROF it is not possible to evaluate the calibration of the predicted terciles without their effect. At least partially because of this, the calibration of GPROF deviates

325   from the nominal frequencies. For the GPROF-NN algorithms, however, both algorithms yield frequencies that are close to the expected frequencies of $\frac{1}{3}$ and $\frac{2}{3}$, respectively.
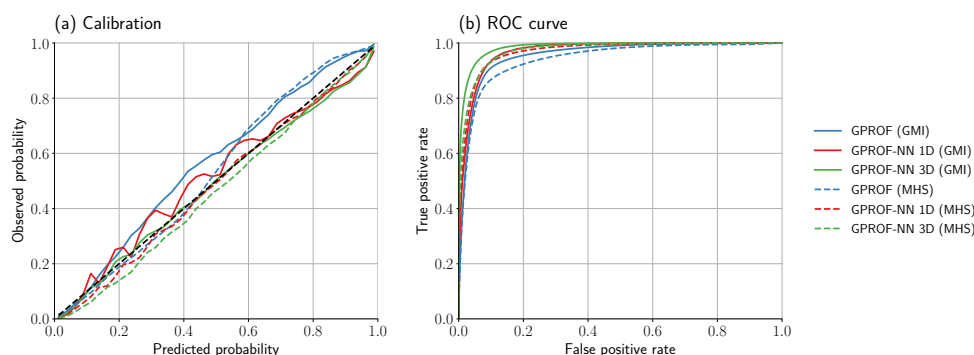

Figure 10. Calibration (Panel (a)) and receiver-operating characteristic (ROC, Panel(b)) for the predicted probability of precipitation.

The quality of the raining/non-raining classification is assessed in Fig. 10, which displays the calibration of the predicted probability and the receiver-operating characteristic (ROC) curve. The predicted probabilities are fairly well calibrated for all algorithms and sensors. Nonetheless, the GPROF-NN algorithms yield results that are slightly closer to the diagonal. The

330   results for the ROC curves are analogous: The GPROF-NN 1D algorithm yields better precipitation detection than GPROF and the GPROF-NN 3D retrieval in turn yields slightly better performance than the 1D version. In terms of classification skill, worse performance is achieved for GMI than for MHS by all algorithms, but again the relative performance of the retrievals is the same.

3.1.2 Effective resolution

Next, we aim to assess the impact of the retrieval method on the effective resolution of the retrieved precipitation fields that is important in hydrologic applications. For this, we adopt the approach from Guilloteau et al. (2017), who have studied the effective resolution of the previous version of GPROF for the GMI and the TRMM Microwave Imager sensors. A 1-dimensional Haar wavelet decomposition in along-track direction over all 128 pixel long sequences in the test data is performed to calculate the effective resolution. We do not consider observations for different surface types separately. Following, Guilloteau et al. (2017) we examine energy spectra as well as correlation coefficients and Nash-Sutcliffe (NS) efficiency of the coefficients of the wavelet decomposition for the reference and retrieved surface precipitation. The results are displayed in Fig. 11.
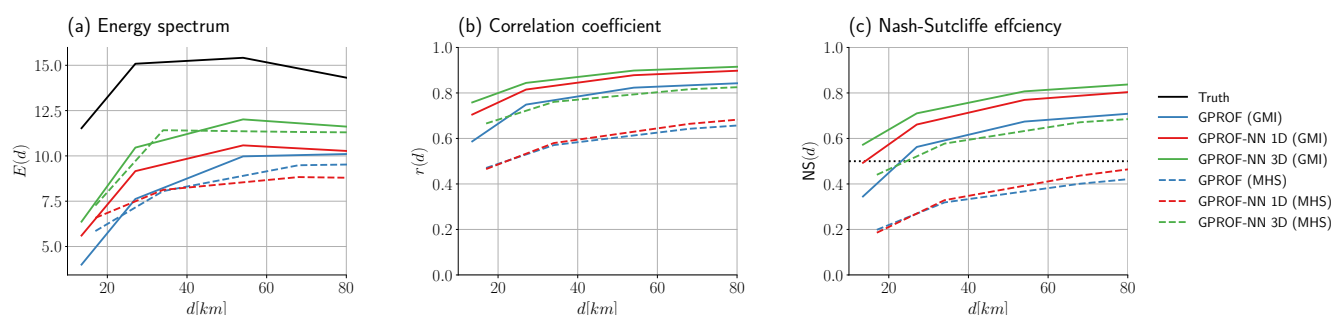

Figure 11. Spatial variability of retrieved and reference fields. Panel (a) shows the average of the total energy defined as the sum of the squared wavelet coefficient at different length scales for the reference and retrieved surface precipitation fields. Panel (b) shows the correlation coefficient between the coefficients of the reference and the retrieved precipitation field. Panel (c) shows the corresponding Nash-Sutcliffe efficiency.

An obvious difference to the results from Guilloteau et al. (2017) is that the energy spectra of the reference precipitation field is not monotonically decreasing. The reason for this is that the reference precipitation field in the retrieval database is smoothed using an averaging filter adapted to the footprint size of the respective sensor. For GMI, the GPROF-NN 3D algorithm has the highest variability in the retrieved precipitation field, followed by the GPROF-NN 1D algorithm and GPROF. However, for all retrievals the variability remains lower than that of the reference field. In terms of correlation of the wavelet coefficients at different scales (Panel (b)), the GPROF-NN 3D algorithm exhibits the highest correlation with the coefficients of the reference field, again followed by the GPROF-NN 1D algorithm and GPROF. The same pattern is observed for the NS efficiency. In terms of effective resolution, defined following Guilloteau et al. (2017) as the smallest scale at which the NS efficiency exceeds 0.5, the GPROF-NN 3D algorithm for the GMI sensor achieves a resolution solution of 13.5 km, which is the distance between consecutive pixels in along-track direction and thus the smallest spatial scale that can be resolved in this analysis. The effective resolutions for the GPROF-NN 1D algorithm is 14.1 km and for GPROF 23.1 km.

For MHS, the effective resolutions of GPROF and GPROF-NN 1D with 128 km and 95 km, respectively, is significantly higher than for GMI. Since the resolution is averaged over the viewing angles of the cross track scanner and because of its

355 generally lower sensitivity to precipitation, a certain degradation of the resolution was expected. Despite this, the GPROF-NN 3D algorithm achieves a resolution of 24.4 km, which is close to that of GPROF for GMI.

3.1.3 Profile retrieval variables

In addition to precipitation fields and path-integrated hydrometeor concentrations, GPROF retrieves concentration profiles of rain, snow and cloud water. The retrieval database also contains latent heating rates as target variable but there is currently no
360 plan to include them in the operational output of GPROF 2021. For this study, latent heating rates were nonetheless included in the output of the GPROF-NN retrievals to investigate the feasibility of the retrieval.

The error statistics for the profile retrievals are displayed in Fig. 12. For GMI, the results are qualitatively similar to those observed for the scalar retrieval variables: The GPROF-NN 1D retrieval has slightly lower biases than than GPROF with the GPROF-NN 3D algorithm yielding the lowest biases. Similar patterns are observed for MSE and for SMAPE throughout most
365 of the atmosphere. For rain and snow water content the SMAPE of the GPROF-NN 3D retrieval increases and even exceeds that of GPROF-NN 1D at the topmost levels where the hydrometeors are present. This is presumably due the scarcity of profiles with hydrometeors at these altitudes, leading to decreased accuracy for the more complex neural network model employed by the GPROF-NN 3D algorithm.

The results for MHS are again mostly similar to those of GMI. A noticeable difference, however, is that the biases of
370 the GPROF-NN retrievals are slightly larger than those of GPROF. This is somewhat surprising considering that there is a noticeable decrease for GMI. This indicates that those biases are due to the absolute sensitivity of the sensor rather than the retrieval method.

3.2 Case study: Hurricane Harvey

All of the results presented above were based on a test dataset with the same statistics as the retrieval database. While for GMI
375 the observations can be expected to be consistent with those in the database, this is not necessarily the case for sensors for which the retrieval database contains mostly simulated observations. While a comprehensive analysis of the retrieval performance on real observations is outside the scope of this study, this section presents retrieval results from two overpasses over Hurricane Harvey to provide an indication as to whether the performance characteristics of the retrieval algorithm can be expected to carry over to real observations.

380 The first considered overpass is of the GPM Core Observatory over Hurricane Harvey and occurred on August 25, 2017 at 11:50 UTC. Retrieved surface precipitation is displayed in Fig. 13. In addition to the retrieved surface precipitation from GPROF and the GPROF-NN algorithms, the figure also displays the surface precipitation from the GPM combined product averaged to the resolution of GMI's 18.7 GHz channel. When considered at log-scale, obvious structural differences in the retrieved precipitation are visible: The GPROF retrieval produces large areas of low precipitation that cover most of the scene,
385 which are not present in the combined retrieval. This is consistent with the overestimation of very small precipitation observed in Fig. 6. These artifacts are significantly reduced in the results of GPROF-NN 1D algorithm and even more so in the results of the GPROF-NN 3D retrieval. When considered at linear scale, all three retrievals capture the global structure of the precipitation
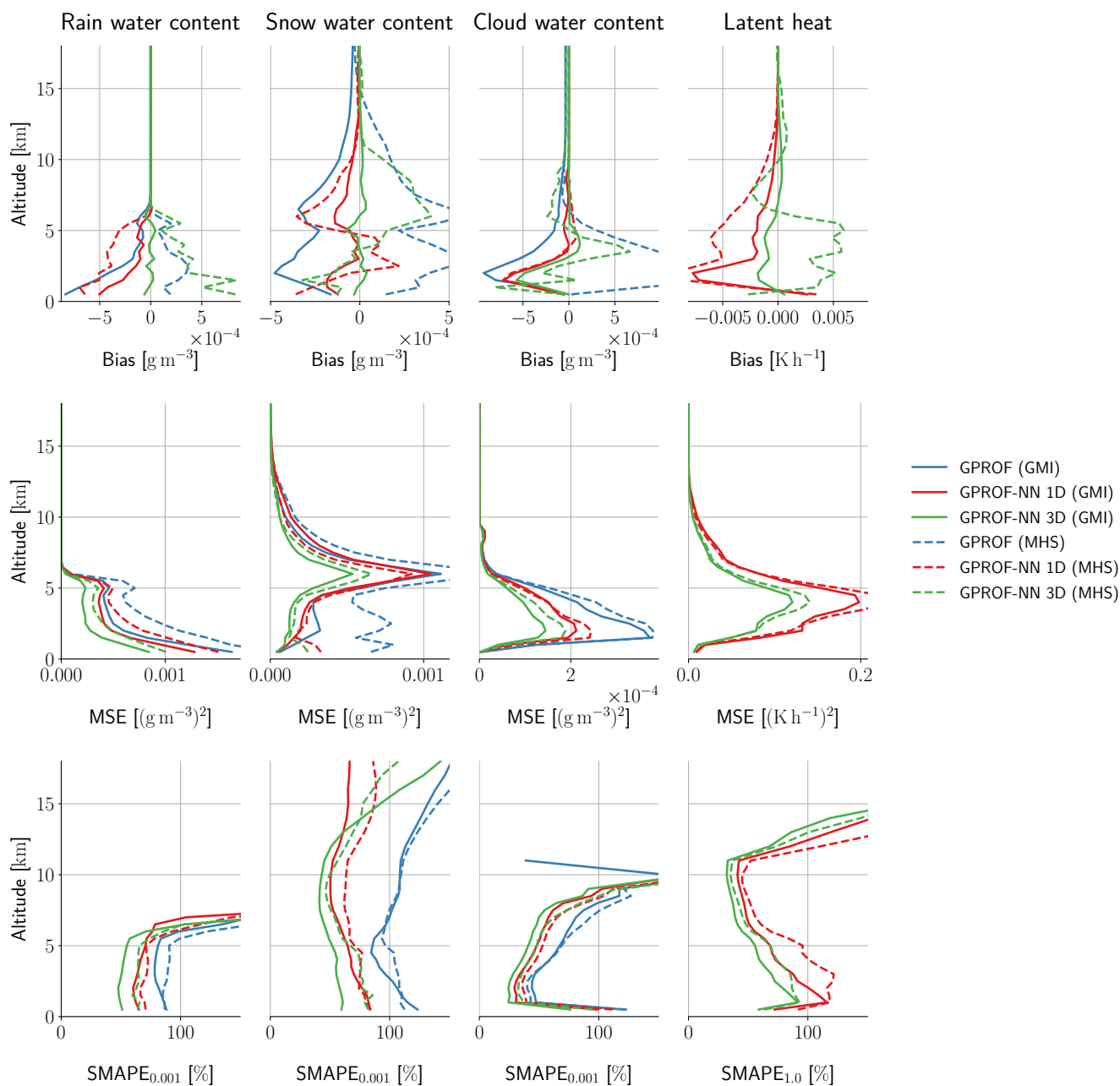
Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

Figure 12. Error statistics of the retrieved profile variables. Columns show the errors for the different retrieved variables, whereas rows show altitude averaged bias, RMSE and MAPE, respectively.
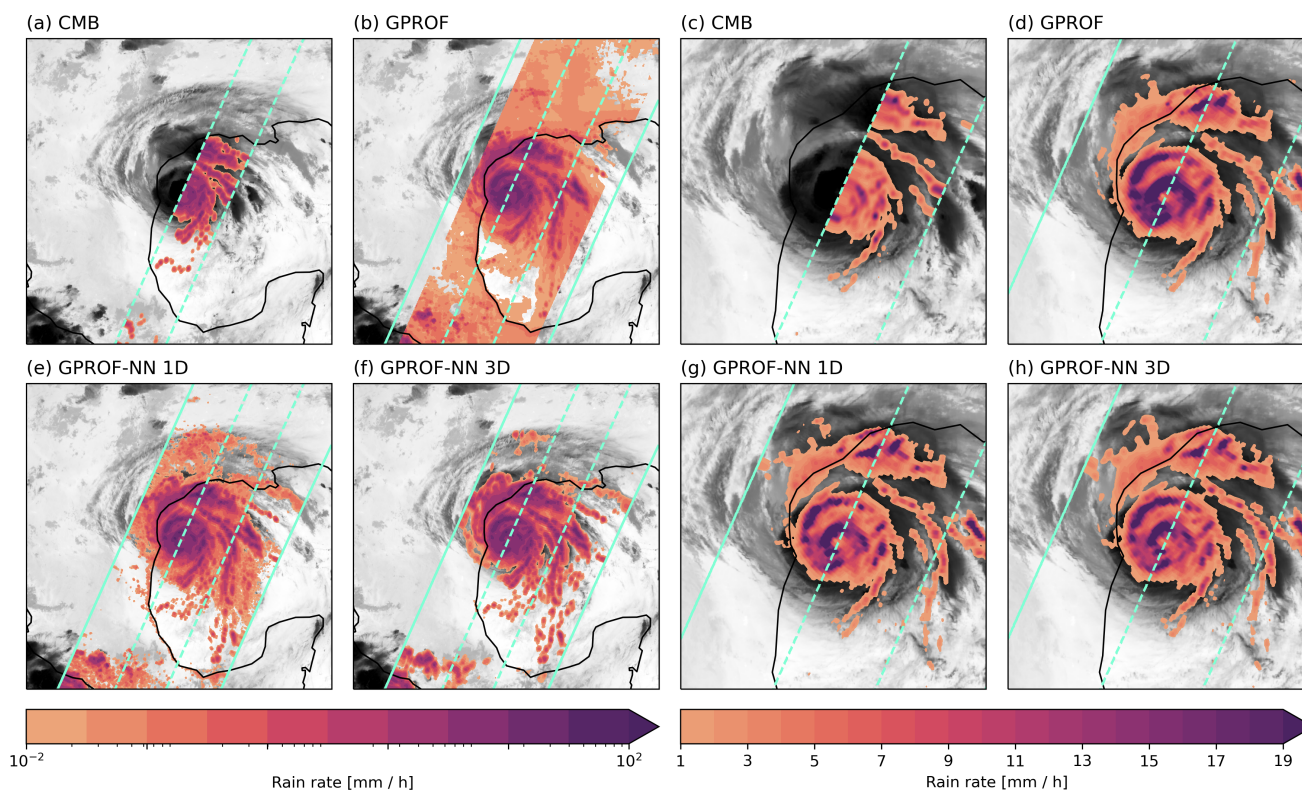
Atmospheric
Measurement
Techniques
Discussions

Figure 13. Retrieved surface precipitation from the GPM combined product (Panels (a), (c)), the GPROF algorithm (Panels (b), (d)), and the neural network versions GPROF-NN 1D (Panels (e), (g)) and GPROF-NN 3D (Panels (f), (h)) in and around Hurricane Harvey on 2017-08-25 at 11:50:00 UTC. Panels (a), (b), (e), (f) display the precipitation on a logarithmic scale over an area spanning 2 000 km in zonal and meridional directions. Panels (c), (d), (g), (f) display the surface precipitation on a linear color scale in an enlarged area spanning 500 km around the Hurricane. The background shows 10.3 μm brightness temperatures from the GOES 16 observations closest to the overpass.

but tend to overestimate its magnitude compared to the combined retrieval. This effect is strongest for GPROF but is also apparent for the neural network based retrievals. Nonetheless, both neural network based retrievals better reproduce the fine

390　scale structure that is visible in the results of the combined retrieval, which is inline with the analysis of the effective resolution of the retrievals.

　　Fig. 14 presents retrieved surface precipitation from an overpass of the MHS sensor on board the NOAA-18 satellite over the same storm at 13:58 UTC. The figure also presents radar derived precipitation rates from NOAA's MRMS product suite as reference data. Similar as for GMI, GPROF predicts low precipitation rates across large parts of the scene and even in areas that are cloud free. Albeit to a lesser degree, this seems to be an issue also for the GPROF-NN 1D algorithm. The GPROF-NN

395　3D algorithm is even less affected by this issue and also yields better agreement with the MRMS precipitation rates over land. For the higher rain rates in and around the Hurricane, the overall structure agrees well with the radar measurements. High
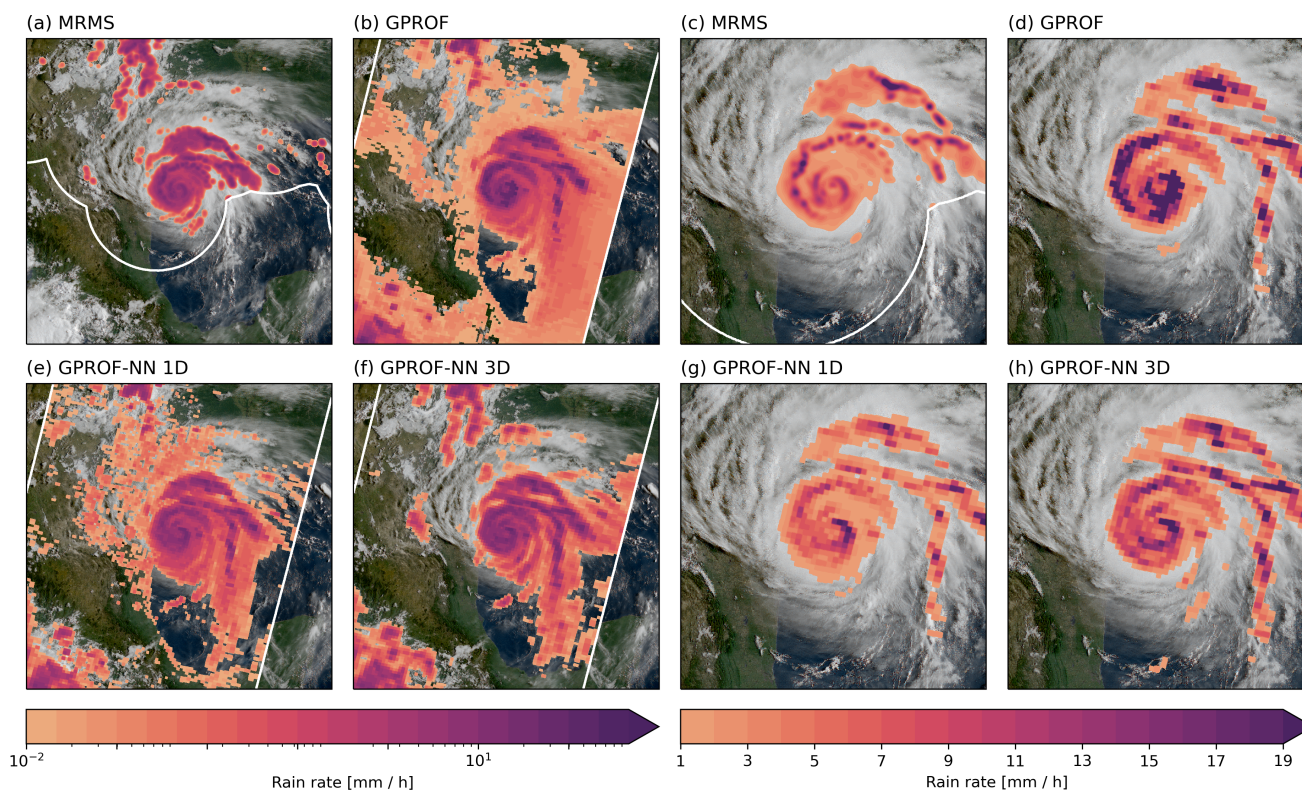
Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Figure 14. Surface precipitation retrieved from ground radar, (Panels (a), (c)), and MHS PMW observations using the GPROF algorithm (Panels (b), (d)), and the neural network versions GPROF-NN 1D (Panels (e), (g)) and GPROF-NN 3D (Panels (f), (h)) in Hurricane Harvey on 2017-08-25 at 13:58:00 UTC. Panels (a), (b), (e), (f) display the precipitation on a logarithmic scale over an area spanning 2 000 km in zonal and meridional directions. Panels (c), (d), (g), (f) display the surface precipitation on a linear color scale in an enlarged area spanning 500 km around the Hurricane. In the background the closest true-color composite from GOES 16 is shown.

precipitation rates are strongly overestimated by GPROF. Results of the GPROF-NN algorithms are closer to the MRMS rain rates, however, due to the distance of the observations from the coast it is not clear how reliable the ground radar measurements

400 are.

3.3 Processing time

Another relevant question regarding the retrieval algorithm is the computational cost of processing the satellite observations. Since GPROF is used operationally on PMW observations from a constellation of sensors spanning several decades of observations, the processing time must not be excessively high. Although neural networks are know to be efficient to evaluate, this

405 often assumes dedicated hardware, which is not yet available at the processing centers. We therefore evaluated the processing

time that is required for a retrieval of an orbit of observations using a single CPU core. The results are displayed in Fig. 15. Despite their superior retrieval accuracy, the GPROF-NN algorithms are about twice as fast as GPROF for GMI and MHS.
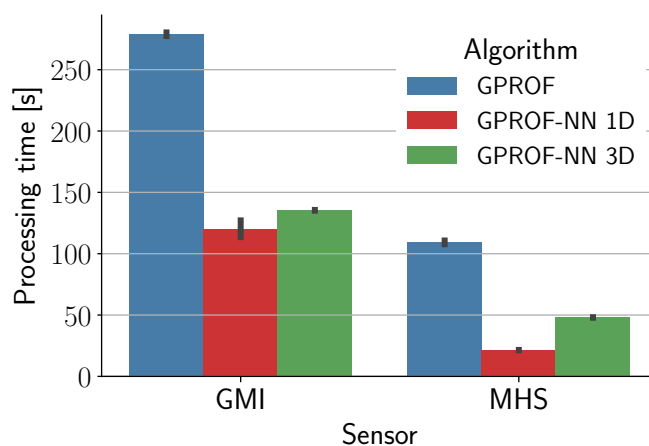

Figure 15. Single CPU core processing time for an orbit of observations for the three retrieval algorithms for GMI and MHS. Error bars show the range of one standard deviation around the mean.

4 Discussion

This study presented two novel, neural network based implementations of the GPROF retrieval algorithm and evaluated their
410 performance for the GMI and MHS sensors against the current implementation.

4.1 Retrieval performance

The evaluation of the GPROF-NN 1D algorithm against GPROF, showed that retrieval accuracy as well as effective resolution can be improved simply by replacing the current retrieval method with a fully-connected neural network. Although both GPROF and the GPROF-NN 1D algorithm are based on the same retrieval database and use the same information as retrieval
415 input, the neural network provides more accurate results. A potential explanation for this may lie in the way the two algorithms handle observation uncertainties, which were shown by Elsaesser and Kummerow (2015) to have a significant effect on the retrieval accuracy. For GPROF, these uncertainties must be specified manually. Apart from sensor noise, observations from other sensors than GMI are affected by modeling errors in the simulated observations, which are difficult to estimate and unlikely to be well described by the assumed Gaussian error model. In addition to this, uncertainties are inflated to account
420 for the sparsity of the retrieval database and the effects of clustering. Since samples with low precipitation rates are generally better represented in the database, this likely makes the uncertainties too large for the retrieval of low precipitation rates, which may explain the inferior performance of GPROF for low precipitation rates observed in Fig. 6 and Fig. 7. The neural network

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

based algorithms infer observation errors directly from the data, and can thus handle arbitrary observation errors, given that they are accurately represented in the training data.

425   Another advantage of the neural network retrievals that may explain the improved accuracy is that they scale more easily to large retrieval databases. While the GPROF algorithm requires compressing the retrieval database in a way that causes information loss, the training of the neural networks uses the full database. However, even in the absence of clustering, Pfreundschuh et al. (2018) provided empirical evidence that neural network based retrievals are less affected by the curse of dimensionality, which means they yield more accurate results when limited data is available. Although the GPROF database is fairly large,

430 heavy precipitation events are likely still underrepresented, which may be exacerbated by the clustering performed by GPROF. In this context, it may also be worth pointing out that, since the database size influences only the training time of the GPROF-NN algorithms, they can potentially be applied with even larger retrieval databases than the one currently used, which may help to further improve the retrieval accuracy in the future.

    The second important finding from this study is that by extending the retrieval to incorporate structural information, its

435 accuracy can be further improved by another 10 to 25 % compared to the GPROF-NN 1D retrieval at the same time as the effective resolution in along track direction is decreased to its lower limit of 13.5 km. The use of structural information for precipitation retrievals is common practice in algorithms based on infrared observations (Sorooshian et al., 2000; Hong et al., 2004) and the potential benefits of CNN based retrievals have been shown in Sadeghi et al. (2019). While basic structural information has been used in earlier PMW precipitation retrieval algorithms, as e.g. by Kummerow and Giglio (1994a), we are

440 not aware of any other operational PMW algorithms that incorporate structural information using CNNs.

    As a concluding remark regarding the retrieval performance, it should also be noted that this study focused on the development of a generic retrieval algorithm applicable to all sensors of the GPM constellation within the operational constraints of the current GPROF retrieval. This means that the used neural network models were not optimized exhaustively and that the performance of neural network based PMW precipitation retrievals can likely be improved further by dedicated tuning of the

445 architecture.

4.2 Limitations

It is important to consider the limitations of the results presented in this study. We have deliberately limited the evaluation of the retrieval accuracy to test data with the same statistical properties as the retrieval database. This was done to isolate the effect of the retrieval method from potential aliasing effects that would be introduced by the use of external validation data. The

450 presented retrieval accuracy should therefore be interpreted as an upper bound on the accuracy that can be achieved with respect to external validation data. Since the GMI retrieval is trained using real observations, the performance on real observations can be expected to be close to the results presented, which was confirmed by the results from the GMI overpass over Hurricane Harvey (Fig. 13). For other sensors, however, the observations in the database can only be simulated and may significantly from true observations. For the GPROF-NN 3D algorithm an additional neural network model was required to transform the

455 data from the retrieval database into a form that is amenable for training a CNN. Despite this, the retrieval results for the MHS

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

overpass of Hurricane Harvey actually indicated that the retrieval for MHS is improved by the GPROF-NN retrieval even on real observation.

5 Conclusions

The results presented in this study clearly demonstrate the potential of a neural network based implementation of GPROF to
460 improve accuracy as well as the effective resolution of retrievals of precipitation and hydrometeor profiles. In addition to the benefits of simply replacing the current retrieval scheme with an identical neural network implementation, we found similar, additional improvements by incorporating spatial information into the retrieval through the use of a CNN based implementation.

Although the results presented here cannot fully answer the question to what extent the improvements observed for the
465 GPROF-NN algorithms carry over to operational application of the retrievals, they show the extent to which the current implementation of GPROF would limit the retrieval accuracy given an ideal retrieval database. Since the retrieval method has an impact comparable to the differences in retrieval accuracy between MHS and GMI, upgrading the retrieval to a neural network based implementation has the potential of being a very cost efficient way to improve global measurements of precipitation with the added advantage of being applicable even to historical observations.

470 The GPROF implementations presented in this study constitute a first step towards such an upgrade of GPROF. The next step will be to run the GPROF-NN retrievals along side GPROF 2021 for all sensors of the GPM constellation and to validate the retrieval results against independent validation data. The Python based software package that implements the retrieval and training framework is made available together with all trained models as free software (Pfreundschuh, 2022).

Although the effective improvements that will be achieved in operational use still remain to be investigated, we take the
475 results presented here as a promising indication of the potential of the GPROF-NN retrievals to improve PMW retrievals from the sensors of the GPM constellation. With this these algorithms may constitute a step towards improving our ability to measure the global hydrological cycle and its changes in a warming climate.

Code availability. The implementation of the GPROF-NN retrievals is published is as free software online (Pfreundschuh, 2022).
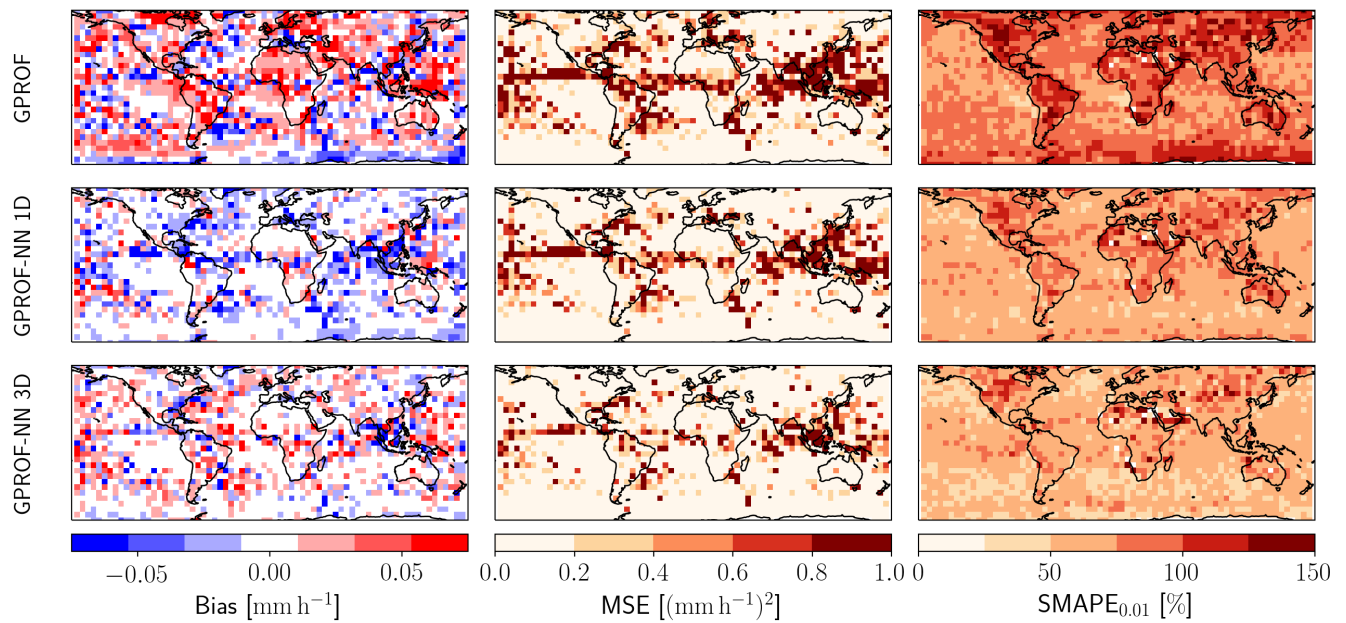
Appendix A

480 **A1**

Figure A1. Like Fig. 9 but for MHS

Table A1. Like Tab. 2 but for convective precipitation.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\mathrm{mm\,h^{-1}}]$	-0.0010 ± 0.0001	-0.0015 ± 0.0001	-0.0011 ± 0.0001
MAE $[\mathrm{mm\,h^{-1}}]$	0.0310 ± 0.0001	0.0239 ± 0.0001	0.0204 ± 0.0001
MSE $[\mathrm{mm\,h^{-1}}]$	0.1766 ± 0.0001	0.1298 ± 0.0001	0.0854 ± 0.0001
$\mathrm{MAPE_{0.01}}$ $[\%]$	108.7784 ± 0.0391	107.1976 ± 0.0378	92.8343 ± 0.0542

Table A2. Like Tab. 2 but for RWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\mathrm{mm\ h^{-1}}]$	0.0017 ± 0.0000	-0.0005 ± 0.0000	-0.0003 ± 0.0000
MAE $[\mathrm{mm\ h^{-1}}]$	0.0183 ± 0.0000	0.0127 ± 0.0000	0.0094 ± 0.0000
MSE $[\mathrm{mm\ h^{-1}}]$	0.0112 ± 0.0000	0.0086 ± 0.0000	0.0047 ± 0.0000
$\mathrm{MAPE}_{0.001}$ $[\%]$	83.8520 ± 0.0287	69.6918 ± 0.0284	61.8979 ± 0.0315

Atmospheric
Measurement
Techniques
Discussions

Table A3. Like Tab. 2 but for IWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\text{mm h}^{-1}]$	-0.0022 ± 0.0000	-0.0006 ± 0.0000	-0.0002 ± 0.0000
MAE $[\text{mm h}^{-1}]$	0.0199 ± 0.0000	0.0123 ± 0.0000	0.0085 ± 0.0000
MSE $[\text{mm h}^{-1}]$	0.0129 ± 0.0000	0.0123 ± 0.0000	0.0053 ± 0.0000
$\text{MAPE}_{0.001}$ $[\%]$	88.0565 ± 0.0312	67.3705 ± 0.0305	58.5831 ± 0.0334

Table A4. Like Tab. 2 but for CWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\mathrm{mm\ h^{-1}}]$	-0.0019 ± 0.0000	-0.0005 ± 0.0000	-0.0005 ± 0.0000
MAE $[\mathrm{mm\ h^{-1}}]$	0.0268 ± 0.0000	0.0157 ± 0.0000	0.0115 ± 0.0000
MSE $[\mathrm{mm\ h^{-1}}]$	0.0027 ± 0.0000	0.0015 ± 0.0000	0.0009 ± 0.0000
$\mathrm{MAPE}_{0.001}$ $[\%]$	62.2267 ± 0.0100	36.6584 ± 0.0078	27.9016 ± 0.0087

Atmospheric
Measurement
Techniques
Open Access
Discussions
EGU

Table A5. Like Tab. 3 but for convective precipitation.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\mathrm{mm\ h^{-1}}]$	0.0072 ± 0.0001	-0.0033 ± 0.0001	0.0019 ± 0.0001
MAE $[\mathrm{mm\ h^{-1}}]$	0.0404 ± 0.0001	0.0280 ± 0.0001	0.0240 ± 0.0001
MSE $[\mathrm{mm\ h^{-1}}]$	0.2119 ± 0.0001	0.1417 ± 0.0001	0.0908 ± 0.0001
$\mathrm{SMAPE_{0.01}}\ [\%]$	109.8088 ± 0.0483	110.0042 ± 0.0506	95.5691 ± 0.1124

Table A6. Like Tab. 3 but for RWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias [mm h^{-1}]	0.0038 ± 0.0000	-0.0012 ± 0.0000	0.0004 ± 0.0000
MAE [mm h^{-1}]	0.0246 ± 0.0000	0.0145 ± 0.0000	0.0120 ± 0.0000
MSE [mm h^{-1}]	0.0196 ± 0.0000	0.0126 ± 0.0000	0.0071 ± 0.0000
SMAPE$_{0.001}$ [%]	91.8829 ± 0.0350	76.0732 ± 0.0349	73.0157 ± 0.0750

Table A7. Like Tab. 3 but for IWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\text{mm h}^{-1}]$	0.0051 ± 0.0000	-0.0011 ± 0.0000	0.0015 ± 0.0001
MAE $[\text{mm h}^{-1}]$	0.0290 ± 0.0000	0.0120 ± 0.0000	0.0114 ± 0.0001
MSE $[\text{mm h}^{-1}]$	0.0270 ± 0.0001	0.0119 ± 0.0001	0.0078 ± 0.0001
$\text{SMAPE}_{0.001}$ [%]	92.5347 ± 0.0364	74.4043 ± 0.0371	68.5830 ± 0.0760

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Table A8. Like Tab. 3 but for CWP.

Metric	GPROF	GPROF-NN 1D	GPROF-NN 3D
Bias $[\mathrm{mm\,h^{-1}}]$	0.0051 ± 0.0000	-0.0003 ± 0.0000	-0.0003 ± 0.0000
MAE $[\mathrm{mm\,h^{-1}}]$	0.0299 ± 0.0000	0.0193 ± 0.0000	0.0156 ± 0.0000
MSE $[\mathrm{mm\,h^{-1}}]$	0.0033 ± 0.0000	0.0016 ± 0.0000	0.0012 ± 0.0000
SMAPE$_{0.001}$ [%]	64.4350 ± 0.0137	46.5956 ± 0.0112	39.0331 ± 0.0238

Atmospheric
Measurement
Techniques

Open Access

EGU

Discussions

Atmospheric
Measurement
Techniques
Discussions

References

495 Ba, J. L., Kiros, J. R., and Hinton, G. E.: Layer normalization, arXiv preprint arXiv:1607.06450, 2016.

Bonsignori, R.: The Microwave Humidity Sounder (MHS): in-orbit performance assessment, in: Sensors, systems, and next-generation satellites XI, vol. 6744, p. 67440A, International Society for Optics and Photonics, 2007.

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
500 Geoscience and Remote Sensing, 49, 3249–3272, https://doi.org/10.1109/TGRS.2011.2158438, 2011.

Chollet, F.: Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258, 2017.

Elsaesser, G. S. and Kummerow, C. D.: The Sensitivity of Rainfall Estimation to Error Assumptions in a Bayesian Passive Microwave Retrieval Algorithm, J. Appl. Meteorol. Clim., 54, 408 – 422, https://doi.org/10.1175/JAMC-D-14-0105.1, 2015.

505 Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Grecu, M., Olson, W. S., Munchak, S. J., Ringerud, S., Liao, L., Haddad, Z., Kelley, B. L., and McLaughlin, S. F.: The GPM Combined 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
510 Hydrometeorology, 18, 3051–3070, 2017.

Hendrycks, D. and Gimpel, K.: Gaussian error linear units (gelus), arXiv preprint arXiv:1606.08415, 2016.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hong, Y., Hsu, K. L., Sorooshian, S., and Gao, X. G.: Precipitation estimation from remotely sensed imagery using an artificial neural
515 network cloud classification system, J. Appl. Meteor., 43, 1834–1852, 2004.

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 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.

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

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Koenker, R. and Hallock, K. F.: Quantile regression, Journal of economic perspectives, 15, 143–156, 2001.

Kummerow, C. and Giglio, L.: A Passive Microwave Technique for Estimating Rainfall and Vertical Structure Information from Space. Part I: Algorithm Description, Journal of Applied Meteorology and Climatology, 33, 3 – 18, https://doi.org/10.1175/1520-
525 0450(1994)033<0003:APMTFE>2.0.CO;2, 1994a.

Kummerow, C. and Giglio, L.: A passive microwave technique for estimating rainfall and vertical structure information from space. Part I: Algorithm description, J. Appl. Meteorol. Clim., 33, 3–18, 1994b.

Kummerow, C. and Giglio, L.: A passive microwave technique for estimating rainfall and vertical structure information from space. Part II: Applications to SSM/I data, J. Appl. Meteorol. Clim., pp. 19–34, 1994c.

Atmospheric
Measurement
Techniques
Discussions

Open Access
EGU

530 Kummerow, C., Olson, W., and Giglio, L.: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive
     microwave sensors, IEEE T. Geosci. Remote, 34, 1213–1232, https://doi.org/10.1109/36.536538, 1996.

Kummerow, C. D., Ringerud, S., Crook, J., Randel, D., and Berg, W.: An Observationally Generated A Priori Database for Microwave
     Rainfall Retrievals, Journal of Atmospheric and Oceanic Technology, 28, 113 – 130, https://doi.org/10.1175/2010JTECHA1468.1, 2011.

Kummerow, C. D., Randel, D. L., Kulie, M., Wang, N.-Y., Ferraro, R., Joseph Munchak, S., and Petkovic, V.: The Evolution of the Goddard
535 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.

Loshchilov, I. and Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts, arXiv preprint arXiv:1608.03983, 2016.

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

NASA: Precipitation Processing System (PPS), GPM Version 07 Release Notes and Documentation, https://arthurhou.pps.eosdis.nasa.gov/
540 GPMprelimdocs.html, accessed: 2021-12-19, 2021.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf,
     A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative
     Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems 32, edited by Wallach, H.,
     Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., pp. 8024–8035, Curran Associates, Inc., http://papers.neurips.
545 cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf, 2019.

Perez, F. and Granger, B. E.: IPython: A System for Interactive Scientific Computing, Computing in Science Engineering, 9, 21–29,
     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.

550 Pfreundschuh, S., Eriksson, P., Duncan, D., Rydberg, B., Håkansson, N., and Thoss, A.: A neural network approach to estimating a posteriori
     distributions of Bayesian retrieval problems, Atmos. Meas. Tech., 11, 4627–4643, https://doi.org/10.5194/amt-11-4627-2018, 2018.

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,
555 https://doi.org/10.5281/zenodo.5789830, 2021.

Sadeghi, M., Asanjan, A. A., Faridzad, M., Nguyen, P., Hsu, K., Sorooshian, S., and Braithwaite, D.: PERSIANN-CNN: Precipitation esti-
     mation from remotely sensed information using artificial neural networks–convolutional neural networks, Journal of Hydrometeorology,
     20, 2273–2289, 2019.

Simpson, J., Kummerow, C., Tao, W.-K., and Adler, R. F.: On the tropical rainfall measuring mission (TRMM), Meteorology and Atmo-
560 spheric physics, 60, 19–36, 1996.

Smith, T. M., Lakshmanan, V., Stumpf, G. J., Ortega, K. L., Hondl, K., Cooper, K., Calhoun, K. M., Kingfield, D. M., Manross, K. L.,
     Toomey, R., and Brogden, J.: Multi-Radar Multi-Sensor (MRMS) Severe Weather and Aviation Products: Initial Operating Capabilities,
     BAMS, 97, 1617 – 1630, https://doi.org/10.1175/BAMS-D-14-00173.1, 2016.

Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN system satellite based estimates
565 of tropical rainfall, Bull. Amer. Meteor. Soc., 81, 2035–2046, 2000.

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

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.