https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

An improved near real-time precipitation retrieval for Brazil

Simon Pfreundschuh[1], Ingrid Ingemarsson[1], Patrick Eriksson[1], Daniel A. Vila[2], and Alan J. P. Calheiros[3]

[1]Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden
[2]Regional office for the Americas, World Meteorological Organization, Asunción, Paraguay
[3]Coordination of Applied Research and Technological Development, National Institute for Space Research (INPE), São José dos Campos, Brazil

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

Abstract. Observations from geostationary satellites offer the unique ability to provide spatially continuous coverage at continental scales with high spatial and temporal resolution. Because of this, they are commonly used to complement ground-based measurements of precipitation, whose coverage is often more limited.

We present a novel, neural-network-based, near real-time precipitation retrieval for Brazil based on visible and infrared (VIS/IR) observations from the Advanced Baseline Imager on the Geostationary Operational Environmental Satellite 16. The retrieval, which employs a convolutional neural network to perform Bayesian retrievals of precipitation, was developed with the aims of (1) leveraging the full potential of latest-generation geostationary observations and (2) providing probabilistic precipitation estimates with well-calibrated uncertainties. The retrieval is trained using co-locations with combined radar and radiometer retrievals from the Global Precipitation Measurement (GPM) Core Observatory. Its accuracy is assessed using one month of gauge measurements and compared to the precipitation retrieval that is currently in operational use at the Brazilian Institute for Space Research as well as two state-of-the-art global precipitation products: The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System and the Integrated Multi-Satellite Retrievals for GPM (IMERG). Even in its most basic configuration, the accuracy of the proposed retrieval is similar to that of IMERG, which merges retrievals from VIS/IR and microwave observations with gauge measurements. In its most advanced configuration, the retrieval reduces the mean absolute error for hourly accumulations by 22 % compared the currently operational retrieval, by 50 % for the MSE and increases the correlation by 400 %. Compared to IMERG, the improvements correspond to 15 %, 15 % and 39 %, respectively. Furthermore, we show that the probabilistic retrieval is well calibrated against gauge measurements when differences in a priori distributions are accounted for.

In addition to potential improvements in near real time precipitation estimation over Brazil, our findings highlight the potential of specialized data driven retrievals that are made possible through advances in geostationary sensor technology, the availability of high-quality reference measurements from the GPM mission and modern machine learning techniques. Furthermore, our results show the potential of probabilistic precipitation retrievals to better characterize the observed precipitation and provide more trustworthy retrieval results.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

1 Introduction

Timely and highly-resolved measurements of precipitation constitute an important source of information for weather forecasting, disaster response and hydrological modeling. These measurements can be provided by dense radar and gauge networks but their coverage is typically limited in less populated regions. However, even where these measurements are available, they are not necessarily without issues. The ability of rain gauges measurements to truthfully represent spatial precipitation statistics at larger scales is limited by their extreme localization (Smith et al., 1996). Ground-based precipitation radars are affected by beam blocking as well as measurement errors caused by the varying altitude of the radar beam along its range (Holleman, 2007).

Since satellite observations provide continuous spatial coverage, they are well suited to complement the measurements from gauges and ground radars. Microwave observations generally provide the most direct space-borne measurements of precipitation because of their sensitivity to emission and scattering from precipitating hydrometeors. Unfortunately, due to their comparably low spatial resolution, these sensors are currently employed only on polar orbiting platforms. Since this limits the width of the satellite swath, a large constellation of sensors on different platforms is required to achieve low revisit times. This approach is pursued by the Global Precipitation Measurement (GPM, Hou et al., 2014). Nonetheless, the revisit times for the passive microwave sensors of the GPM constellation still exceed 2 hours ins the tropics.

Visible and infrared (VIS/IR) observations from the latest generation of geostationary satellites (Schmit et al., 2005) provide spatial resolutions between 0.5 and 2 km at the sub-satellite point and a temporal resolution of up to 10 minutes for full disk observations. The disadvantage of these observations for measuring precipitation is that they are mostly sensitive to the top of clouds, which are only indirectly related to the precipitation near the surface. Their unrivaled spatial and temporal resolution nonetheless makes them a valuable source of information for satellite-based precipitation estimates.

The operational use of geostationary VIS/IR observations for precipitation retrievals dates back more than 40 years (Scofield and Oliver, 1977) and a large number of different algorithms have been developed over the years (Arkin and Meisner, 1987; Adler and Negri, 1988; Vicente et al., 1998; Sorooshian et al., 2000; Kuligowski, 2002; Scofield and Kuligowski, 2003a; Hong et al., 2004; Kuligowski et al., 2016). Due to the aforementioned indirect relationship between observations and precipitation, nearly all of these methods are based on empirical relationships derived from satellite observations co-located with reference data derived from more direct measurement techniques such as ground-based radar. Moreover, operational retrievals often require corrections to improve the accuracy of their estimates. The Self-Calibrating Real-Time GOES Rainfall Algorithm for Short-Term Rainfall Estimates (SCamPR, Kuligowski et al. 2016), for example, is dynamically calibrated using the latest available microwave precipitation estimates. Similarly, the PERSIANN CSS real time dataset is in the process of being superseded by the PERSIANN PDIR algorithm, which extends the PERSIANN CCS algorithm with a regional correction.

Another example is the HYDRO precipitation retrieval that is currently in operational use at the National Institute for Space Resarch (INPE) in Brazil, which is based on the Hydroestimator algorithm (Scofield and Kuligowski, 2003b). It employs an empirical relationship between the 10.7 $\mu\mathrm{m}$ IR channel and precipitation rates with additional corrections. To adapt it

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

for application over South America yet another correction was derived by de Siqueira and Vila (2019), which improved the accuracy of precipitation accumulations but not that of instantaneous precipitation rates.

A common shortcoming of all retrieval algorithms discussed above is that they neglect retrieval uncertainties. The retrieval of precipitation rates from VIS/IR observations constitutes an inverse problem that is strongly underconstrained. This is true even for microwave based retrievals and likely exacerbated by the less direct information content in the VIS/IR observations. The ill-posed character of the retrieval problem leads to significant retrieval uncertainties. Providing probabilistic estimates that quantify these uncertainties would help the characterization of precipitation estimates and thus increase their usefulness.

This study presents Hydronn, a novel real-time precipitation retrieval that uses VIS/IR observations from the GOES 16 Advanced Baseline Imager (ABI, Schmit et al., 2018) to retrieve precipitation over Brazil. It was designed with two aims: (1) To leverage the full potential of observations from the latest generation of geostationary sensors and (2) to develop a Bayesian precipitation retrieval algorithm that can provide well calibrated uncertainty estimates. The algorithm is trained using a large database of co-locations between observations from the ABI and the combined radar-radiometer retrievals from the GPM Core Observatory (Grecu et al., 2016) . The accuracy of the retrieval is evaluated using a month of gauge measurements. To assess its performance with respect to existing retrievals it is compared to the currently operational HYDRO algorithm as well as two other commonly used precipitation products: The hourly Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System (PERSIANN CCS, Hong et al., 2004) , which is based on IR observations only, and the Integrated Multi-Satellite Retrievals for GPM (IMERG, Huffman et al. 2020), which combines observations from microwave, geostationary sensors and rain gauges.

2 Data

This section presents the rain gauge data, which serve as reference measurements for this study, as well as the baseline algorithms to which Hydronn will be compared. This is followed by a description of the Hydronn retrieval algorithm.

2.1 Rain gauge data

The rain gauge measurements that are used in this study were compiled by the National Institute of Meteorology of Brazil and consists of hourly gauge measurement covering the time range May 2000 until May 2020. From this data December of 2020 will be used for the evaluation of Hydronn. Data from 2018 and 2019 is used to derive correction factors for the calibration of the hourly precipitation estimates produced by Hydronn, as will be described in Sec. 3.6 below.

From all available gauge stations only those with a data availability exceeding $90\ \%$ during December 2020 were selected. Their geographical distribution is displayed together with the mean precipitation in Fig. 1. While the gauge density is fairly high on the south-eastern coast of Brazil, it decreases markedly towards the Northwest. Climatologically, the highest precipitation occurs in the Amazonas and surrounding regions in the Northwest of the country, however this is not well represented in the gauge measurements due to their sparsity in this region. More evident are the high precipitation amounts on the south-western coast of the country extending towards the Northwest, which manifest the South Atlantic Convergence Zone (SACZ, Satya-

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

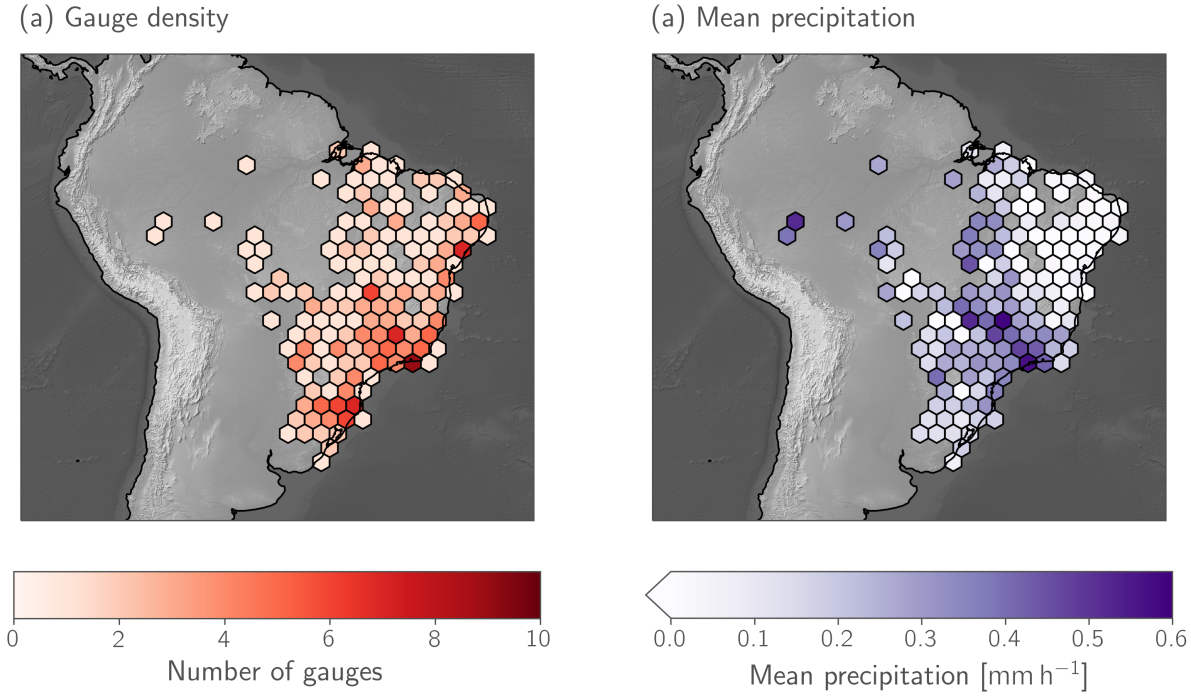

Figure 1. Overview of the rain gauge data from December 2020 used to validate the retrievals. Panel (a) displays their spatial distribution by means of the number of gauges falling into each hexagon. Hexagon-free areas are not covered by any gauges. Panel (b) shows the corresponding mean precipitation.

murty et al., 1998). Very low precipitation rates are observed in the Northeast, which is influenced by large scale subsistence patterns (de Siqueira and Vila, 2019).

2.2 HYDRO

HYDRO is the currently operational near real-time precipitation retrieval at the Center for Weather Forecast and Climate Studies/National Institute for Space Research (CPTEC/INPE). It is based on the Hydroestimator (Scofield and Kuligowski, 2003b) and thus uses a combination of empirical power-law relationships between 10.7 µm IR brightness temperatures and surface precipitation with correction factors taking into account model-derived moisture and wind parameters as well as cloud structure. The current version of the retrieval is described in de Siqueira and Vila (2019), which also introduces regional correction factors based on a climatology of surface precipitation rates derived from radar measurements of the Tropical Rainfall Measurement Mission (TRMM, Simpson et al., 1996) and GPM. For this study we use the corrected version of HYDRO proposed in de Siqueira and Vila (2019) with a regional correction for all of Brazil (referred to as HYDROBR in de Siqueira and Vila (2019)).

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

2.3 PERSIANN CSS

PERSIANN CCS (Hong et al., 2004) uses 10.7 µm IR observations from geostationary satellites to retrieve precipitation. Input images are first segmented using increasing temperature thresholds in order to identify pixels that correspond to convective activity. These pixels are consecutively assumed to be precipitating and classified using a neural network based algorithm. Quantitative precipitation estimates at pixel level are derived from this classification by applying a class-specific power law relationship that relates the 10.7 µm brightness temperatures to precipitation.

The dataset that is used for the evaluation against Hydronn are hourly precipitation rates that are distributed in near real time through the PERSIANN data portal (UCI CHRS Data Portal, 2022). Although the global CCS dataset is currently being replaced by the updated PERSIANN Dynamic Infrared-Rain rate (PDIR-Now, Nguyen et al., 2020) we were not able to use it for this study due to the first third of the evaluation period missing from the online archive.

2.4 GPM IMERG

GPM IMERG (Huffman et al., 2020) combines retrievals from passive microwave and IR observations as well as rain gauge measurements to produce global, half-hourly measurements of precipitation. Due to its reliance on a wealth of measurement sources as well as the sophistication of the retrieval pipeline, the product can be considered one of the most robust satellite-based precipitation products (Pradhan et al., 2022).

Three different configurations of IMERG products are available: IMERG-Early and IMERG-Late are based solely on satellite observations and available with latencies of 4 and 14 hours, respectively. IMERG-Final is adjusted using global gauge measurements but available first after 3.5 months. Although Hydronn has been designed to target near real-time applications and is thus more similar to IMERG-Early, we use IMERG-Final for our comparison as it constitutes the most elaborate precipitation estimates that are currently available and can thus be considered the state of the art of space-borne quantitative precipitation estimation.

3 Method

This section describes the implementation of Hydronn, the proposed near real-time precipitation retrieval algorithm for Brazil. It is based on a convolutional neural network (CNN), which is used to predict the a posteriori distribution of instantaneous precipitation. Following this, it is discussed how the probabilistic precipitation estimates can be combined to hourly accumulations and an a priori adjustment is proposed to account for differences between the training data and the gauge measurements that are used to evaluate the retrieval.

3.1 Training data

The training data for the Hydronn retrieval is generated from co-locations of input observations from the GOES-16 ABI (Schmit et al., 2018) and retrieved surface precipitation from the GPM combined product (Grecu et al., 2016). GOES ABI observations

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

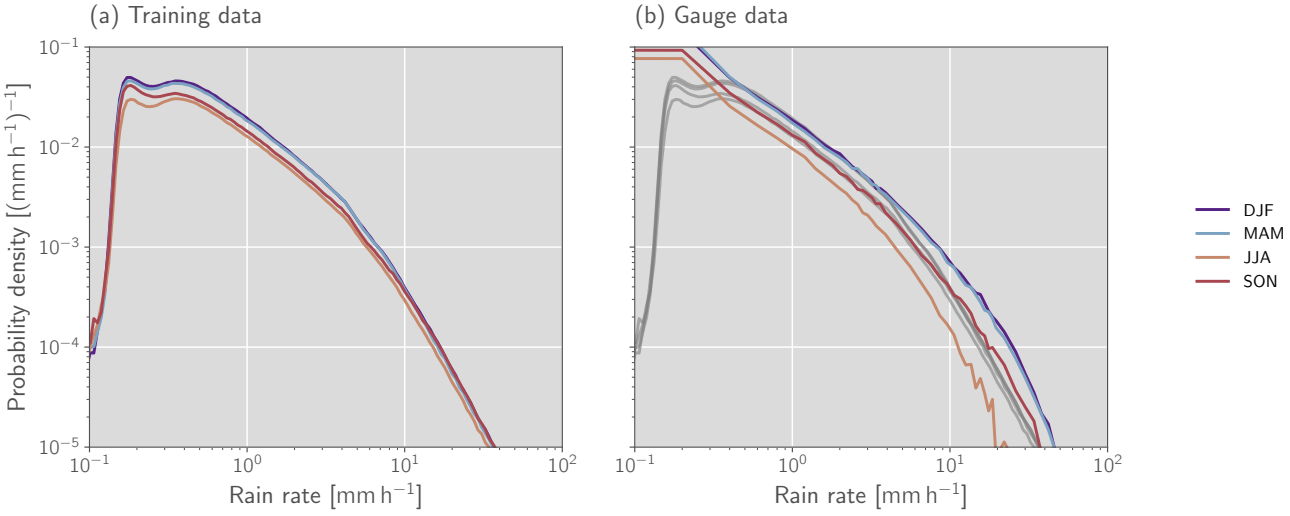

Figure 2. Distribution of reference precipitation rates. Panel (a) shows the seasonal PDFs of precipitation rates in the training data. Panel (b) shows the PDFs of precipitation rates measured by the gauges over the time period covered by the training data. Grey lines in the background trace the PDFs of the precipitation rates in the training data shown in Panel (a).

were extracted at their native resolutions and combined with the GPM combined surface precipitation by remapping them to the 2 km resolution of the ABI's IR channels using a nearest-neighbor criterion. Co-locations were extracted for the time range 2018-01-01 until 2020-01-01 and 2021-01-01 until 2021-09-01. Observations from the first, 11th and 21st day of every month of 2020 are used as test data set to establish the nominal performance of different retrieval configurations.

The results presented in Pfreundschuh et al. (2018) show the correspondence between the training data of neural network retrievals and the a priori distribution of explicitly Bayesian retrieval schemes. This perspective emphasizes the importance of training data distribution for the retrieval results and their interpretation. The distribution of precipitation rates in the training dataset is displayed in Fig. 2. The detection threshold for precipitation of the GPM radar between 0.2 and 0.4 mm h^{-1} is clearly visible in the distributions. In addition to this, a weak seasonal cycle is apparent, which mainly impacts the likelihood of moderate precipitation. The gauge measurements exhibit a stronger effect of the seasonal cycle especially for strong rain. It should be noted here, however, that the precipitation estimates in the training data correspond to instantaneous precipitation estimates while the gauge measurements are integrated over the time of an hour. Differences between the seasonal cycles of the datasets may therefore be caused by changes in the temporal evolution of precipitation events. An approach to reconcile the differences between the distributions of training and validation will be proposed in Sec. 3.6 below.

3.2 Retrieval configurations

The Hydronn retrieval has been implemented in three different configurations in order to assess how the choice of input observations and output resolution affects its performance. The most basic retrieval configuration is the Hydronn$_{4,\mathrm{IR}}$ retrieval,

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

which only uses brightness temperatures from the GOES 16 $10.3\ \mu\mathrm{m}$ channel as input. Due to their sensitivity to cloud top temperatures, longwave IR window channels are also used by HYDRO as well as the PERSIANN CCS retrieval. The availability of similar channels on a long range historical geostationary sensors makes them suitable for the generation of climate data records. The reliance on a single thermal IR channel has the additional advantage that the information content of the retrieval input is independent of the availability of sunlight and thus constant throughout the day. The second retrieval configuration, denoted as $\text{Hydronn}_{4,\text{All}}$, uses all available GOES channels at a resolution of $4\ \mathrm{km}$. It uses the same neural network model as the $\text{Hydronn}_{4,\text{IR}}$ configuration adapted to the larger number of input channels. The third configuration, $\text{Hydronn}_{2,\text{All}}$, aims to exploit the full potential of GOES observations for precipitation retrievals. The input includes observations from all GOES channels at their native resolution and precipitation is retrieved at a resolution of $2\ \mathrm{km}$ at nadir. The characteristics of the three configurations are summarized in Tab. 1.

Table 1. Hydronn retrieval configurations

Name	Input bands	Input resolution	Output resolution
$\text{Hydronn}_{4,\text{IR}}$	13	4 km	4 km
$\text{Hydronn}_{4,\text{All}}$	$1,\ldots,16$	4 km	4 km
$\text{Hydronn}_{2,\text{All}}$	$1,\ldots,16$	$0.5, 1$ and 2km	2 km

3.3 Neural network model

All Hydronn retrievals are based on a common convolutional neural network (CNN) architecture, which is illustrated in Fig. 3. A preliminary study found CNNs to yield significantly more accurate results than fully-connected neural networks that use only a single pixel as input (Ingemarsson, 2021). The fully-convolutional networks are constructed using what we refer to as Xception blocks, which are based on the Xception architecture proposed by Chollet (2017). These blocks are combined in an asymmetric encoder-decoder structure with 5 stages. Each downsampling stage consists of one downsampling Xception block followed by $N = 4$ standard Xception blocks.

Since the $\text{Hydronn}_{2,\text{All}}$ retrieval ingests observations at their native resolution, this architecture contains two additional downsampling blocks that are omitted for the $\text{Hydronn}_{4,\text{IR}}$ and $\text{Hydronn}_{4,\text{All}}$ retrievals. The number of internal features for all architectures was set to $n_f = 128$, which is probably low compared to other neural network architectures. This was mostly motivated by hardware limitations. Since it was found to be sufficient to achieve good retrieval performance, we did not investigate the impact of this decision further.

3.4 Probabilistic precipitation estimates

A defining characteristic of Hydronn is that precipitation is retrieved using a Bayesian framework. This means that, instead of predicting a single precipitation value, it provides an estimate of the full a posteriori distribution of the Bayesian retrieval problem. Although Pfreundschuh et al. (2018) proposed to use quantile regression neural networks (QRNNs) to perform

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

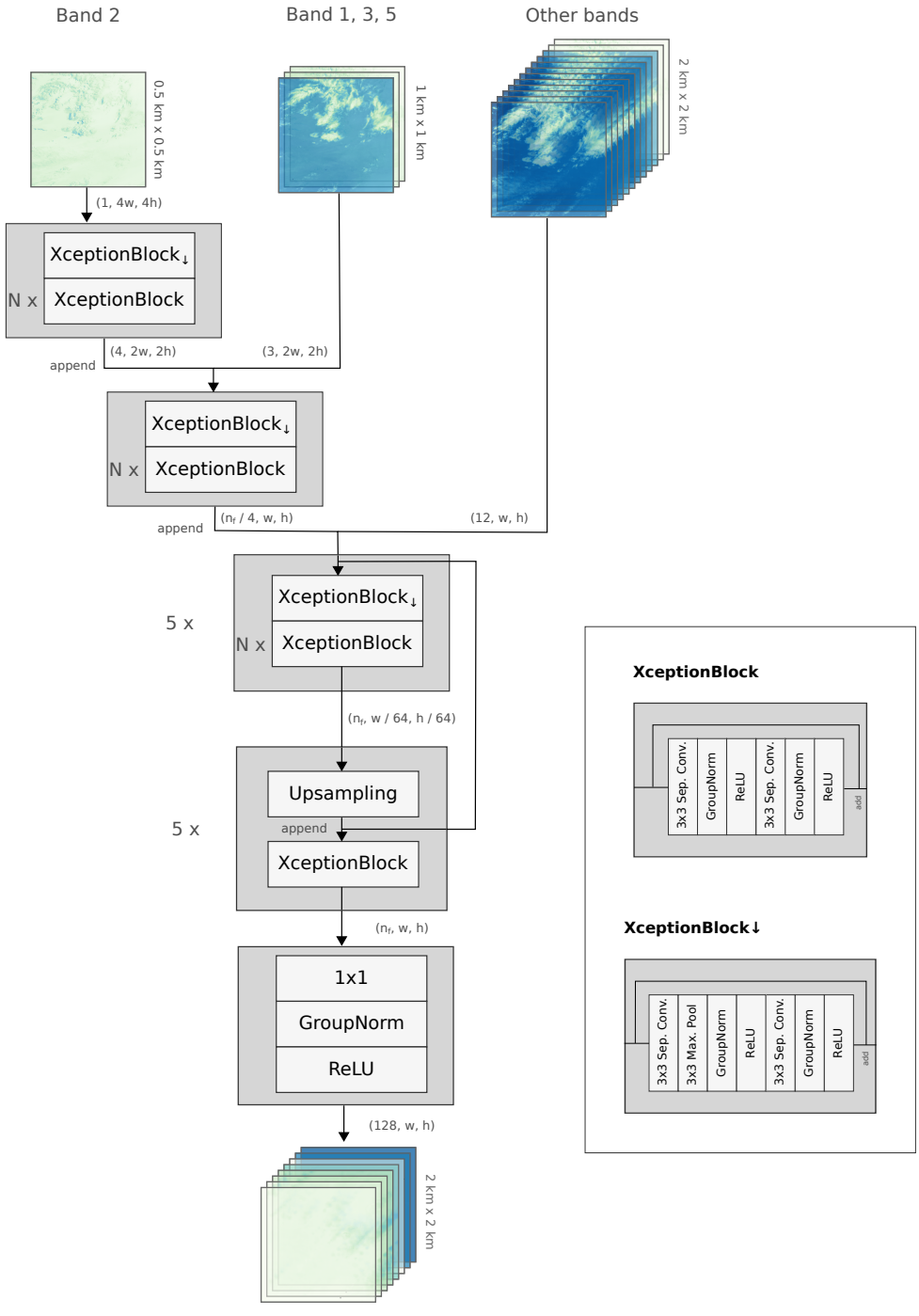

Figure 3. The neural network architecture used by the Hydronn$_{2,\mathrm{All}}$ retrieval.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Bayesian remote sensing retrievals with neural networks, a different approach is taken here. Following the work by Sønderby et al. (2020), the range of precipitation values is discretized and the probability of the observed precipitation falling into each bin is predicted. By normalizing the predicted probabilities, a binned approximation of the probability density function (PDF) of the Bayesian a posterior distribution can be obtained. We found this approach to be equivalent to QRNNs in retrieval accuracy. However, calculating the distribution of the sum of two temporally independent predictions is easier on the binned PDF than on the predicted quantiles, which is why the former approach was chosen for the implementation of Hydronn.

This approach, which we refer to for simplicity as density regression neural network (DRNN), can be easily implemented by treating the retrieval as a classification problem over a discretized range of precipitation values and using a cross-entropy loss to train the network. During inference, the logits predicted by the network are transformed into a probability density by applying a softmax activation and normalizing the bin probabilities.

Hydronn predicts the a posteriori distribution over 128 logarithmically spaced bins covering the range from 10^{-3} to $10^{3}\ \mathrm{mm\ h^{-1}}$. The reference precipitation of pixels without rain was set to a log-uniform random value between 10^{-3} and $10^{-2}\ \mathrm{mm\ h^{-1}}$. While not strictly necessary for our approach, this has the advantage of breaking the degeneracy of low quantiles of the posterior distribution, which makes it easier to verify the calibration of the probabilistic predictions on the validation data.

Since storing the full posterior distribution for all pixels is of little use for operational processing, only a reduced number of relevant statistics are retained in the retrieval output. Those are the posterior mean as well as a sample and 14 quantiles of the posterior distribution. In addition to providing a direct measure of retrieval uncertainty, the quantiles can be used to reconstruct a piece-wise linear approximation of the posterior cumulative distribution function (CDF). The predicted posterior CDF can then be used, for example, to detect the exceedance of certain precipitation thresholds. Compared to training a separate classifier to perform this task, this approach has the advantage that the precipitation threshold can be chosen during inference.

3.5 Calculation of hourly accumulations

The precipitation estimates produced by Hydronn correspond to instantaneous precipitaiton rates. Since GOES 16 imagery is available every 10 minutes, a method is required to accumulate the posterior distributions of the instantaneous precipitation rates to hourly accumulations, which can then be compared to the gauge measurements. While this is not an issue when only the posterior mean is predicted, it is unclear how the retrieval uncertainties should be aggregated in time. In lack of a formal way to resolve this, we have implemented two heuristics for calculating probabilistic estimates of hourly accumulations from instantaneous measurements.

The first heuristic is to simply average the predicted posterior distributions. For the case of multiple identical observations, this preserves the retrieval uncertainties and thus corresponds to the assumption of strong dependence of the retrieval errors for consecutive observations. The second approach is to assume temporal independence of the retrieval uncertainty. For identical, consecutive observations this will generally cause the retrieval uncertainty to decay. Given the binned probability densities of two independent random variables, the PDF of their sum can be approximated by calculating weighted histograms of the outer sum of the bin centroids weighted by the product of the bin probabilities.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

For the evaluation of the Hydronn retrieval, we calculate PDFs of hourly accumulations using both approaches. Two types of accumulations are produced for each Hydronn configuration: One corresponding to the assumption of dependent retrieval errors, which will be identified with the qualifier '(dep.)', as well as one corresponding to the assumption of independent retrieval errors, which will be identified with the qualifier '(indep.)'. Since the assumptions only affect the predicted retrieval uncertainties and not the predicted mean values, such a distinction is not required when point estimates of precipitation are considered.

3.6 Correcting for a priori data

According to Bayes theorem, the posterior distribution of retrieved precipitation $p(x|\mathbf{y})$ for given input observations $\mathbf{y}$ is proportional to the product of the probability of observing $\mathbf{y}$ for a given precipitation rate x and the a priori probability of x:

$$p(x|\mathbf{y}) \propto p(\mathbf{y}|x)p(x) \tag{1}$$

One difficulty with machine learning based retrievals is that the a priori distribution cannot be chosen freely but is dictated by the training data distribution. Fig. 2 indicates that there are inconsistencies between the training data and gauge measurements. For example, the retrieval will learn from the training data that the probability of precipitation values between 10^{-1} and 10^{-2} mm h^{-1} is effectively zero.

This raises the question whether it is possible to correct for the effect of the a priori assumptions encoded in the training data of the retrieval. To explore this, we propose the following method to correct the probabilistic predictions. Let $p_{\text{Gauges}}(x)$ denote the PDF of precipitation as measured by the available gauges shown in Fig 2 (b). Moreover let

$$r(x) = \frac{p_{\text{Gauges}}(x)}{p(x)} \tag{2}$$

denote the ratio of the PDFs of the gauge a priori distribution and the a priori distribution of precipitation as defined by the training data. Assuming that $p_{\text{Gauges}}(x) = 0$ wherever $p(x) = 0$ and that the conditional distribution $p(\mathbf{y}|x)$ of the observations remains unchanged, a corrected posterior distribution can be obtained by point-wise multiplying the likelihood ratio r with the posterior distribution predicted by Hydronn:

$$p_{\text{Corrected}}(x|y) \propto p(y|x)r(x)p(x), \tag{3}$$

The a priori distribution of retrieved hourly accumulations is not necessarily the distribution of instantaneous rain rates displayed in Fig. 2, but depends on how these accumulations are calculated. This leads to a different a priori distributions for each of the two methods used to accumulate the precipitation (see Sec. 3.5). For the assumptions of temporally dependent retrieval uncertainties, the a priori is identical to that of the instantaneous rain rates. For the assumption of temporally independent retrieval uncertainties, however, the a priori corresponds to the distribution of the mean of 6 consecutive draws from this distribution.

An issue that we encountered here is the limited numerical resolution of the gauge measurements of 0.2 mm h^{-1}, which causes problems when the calibration of probabilistic predictions is evaluated against the gauge measurements. To counteract

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

this, we have added numerical noise to the gauge measurements: Uniform random values from the range $[-0.1, 0.1]$ are added to all non-zero measurements, while zero values are replaced by log-randomly distributed values from the range $[10^{-3}, 5{\cdot}10^{-3}]$. This slightly increases the mean of the precipitation by about 2 %, but this is likely negligible compared to other uncertainties that affect precipitation retrievals.

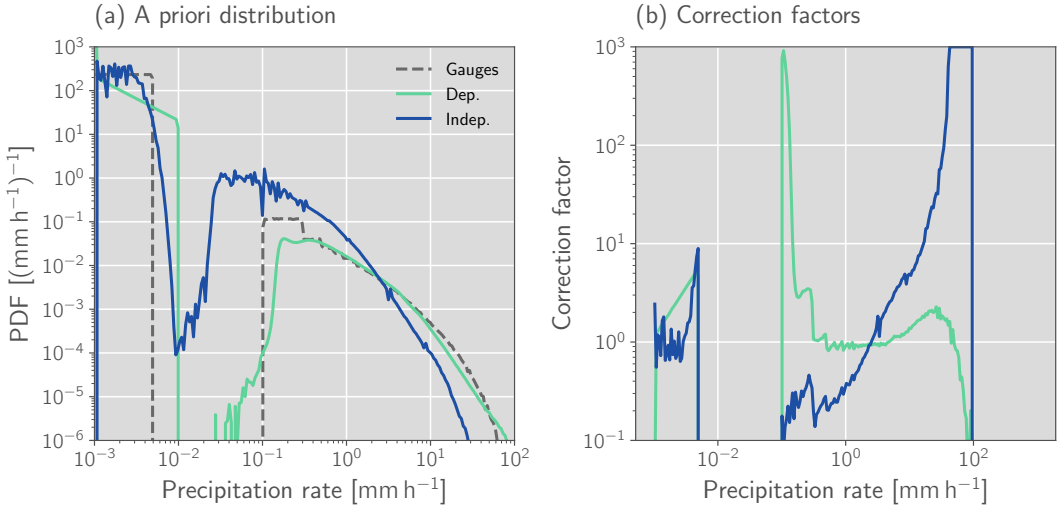

Figure 4. A priori distributions of hourly accumulations and derived correction factors r. Panel (a) displays the a priori distributions of hourly precipitation accumulations derived assuming strong temporal dependence of measurements (green) and complete independence (blue). The gray, dashed line shows the PDF of the gauge measurements. Panel (b) displays the corresponding correction factors for the two assumptions calculated as the ratio between the respective PDFs and the PDF of the gauge measurements.

The a priori distributions and corresponding derived correction factors are displayed in Fig. 4. It is apparent that the assumptions of temporally dependent uncertainties yields better agreement with the gauge data than the assumption of temporally independent uncertainties. The resulting correction factors are thus closer to the $y = 1$ line for the dependence assumption. We found that it was necessary to truncate the correction factors corresponding to the independence assumption at $r = 10^3$ because larger values would amplify numerical noise leading to the rare occurrence of unrealistically high precipitation values, which would distort the retrieval results.

4 Results

This section presents the evaluation of the Hydronn retrievals, which is split into three parts. The first part analyzes the nominal performance of the three Hydronn configurations on the held-out test data. The second part compares the retrieved hourly accumulations to the gauge measurements and the reference precipitation algorithms. Finally, the third part presents a case study of a heavy precipitation event that occurred during the validation period.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

4.1 Evaluation of Hydronn configurations

To obtain an unperturbed assessment of the relative performances of the three Hydronn configurations, we evaluate the performance on the test data, which was derived from the same source (albeit during a different time period) and is therefore guaranteed to have similar statistics as the training data. Fig. 5 displays PDFs of retrieved precipitation conditioned on the value of the reference precipitation. Due to the limited information content of the VIS/IR observations there are significant uncertainties in all results. These lead to significant wet biases for lightly raining pixels and dry biases for strong precipitation.

Nonetheless, slight improvements between the three Hydronn configurations are discernible. While the Hydronn$_{4,\mathrm{IR}}$ retrieval exhibits the weakest relationship between reference and retrieved precipitation, the Hydronn$_{4,\mathrm{All}}$ configuration yields slightly more accurate results. This can be seen in the sharpening of the conditional PDFs for precipitation rates occurring between 2 and 20 $\mathrm{mm\ h^{-1}}$ as well as an increase in the slope of the conditional mean retrieved precipitation for rain rates exceeding 2 $\mathrm{mm\ h^{-1}}$. Clearer improvements in retrieval accuracy are observed for the Hydronn$_{2,\mathrm{All}}$ configuration, which yields a slightly sharper distribution and an increased slope in the conditional mean of the retrieved precipitation for precipitation rates larger than 0.5 $\mathrm{mm\ h^{-1}}$.

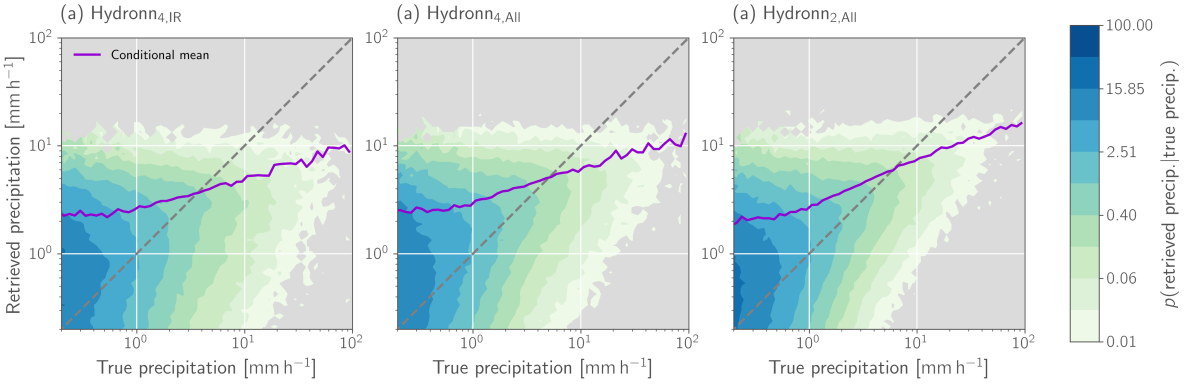

Figure 5. PDFs of retrieved precipitation conditioned on the reference precipitation. The purple line shows the mean retrieved precipitation conditioned on the value of the reference precipitation.

For a more quantitative analysis, Tab. 2 summarizes the retrieval performance using a range of accuracy metrics. The metrics considered here are the bias, mean absolute error (MAE), mean squared error (MSE), the mean of the continuous ranked probability score (MCRPS) and the correlation coefficient. Given a predicted cumulative distribution function F and a reference value x, the continuous ranked probability score (CRPS) is defined as

$$\mathrm{CRPS}(F, x) = \int_{-\infty}^{\infty} \left(F(x') - \mathrm{I}_{x' > x}\right)^2 \, dx', \tag{4}$$

where $\mathrm{I}_{x' > x}$ is the indicator function taking the value 1 where $x' > x$ and 0 otherwise.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Table 2. Accuracy metrics for the three Hydronn configurations evaluated on test data. The value corresponding to the highest accuracy in each column is marked using bold font.

Algorithm	Bias [mm h^{-1}]	MAE [mm h^{-1}]	MSE [(mm h^{-1})^2]	CRPS	Correlation
Hydronn$_{4,IR}$	0.0009	0.1988	1.3356	0.1807	0.4343
Hydronn$_{4,All}$	0.0006	0.1626	1.1848	0.0865	0.5296
Hydronn$_{2,All}$	**-0.0001**	**0.1508**	**1.0827**	**0.0757**	**0.5844**

The 14 quantiles that are produced as retrieval output are used to calculate the CRPS instead of the full predicted PDF, which ensures that the evaluation is representative of the actual retrieval output. In contrast to the other metrics considered in Tab. 2, the CRPS takes into account not only the accuracy of the predicted posterior mean but both sharpness and calibration of the probabilistic precipitation estimates (Gneiting and Raftery, 2007).

Overall, the results from Tab. 2 confirm the tendencies observed in Fig 5. The retrieval accuracy increases as the information content in the input observations is increased. In absolute terms, the largest improvements are achieved when the inputs are extended from a single channel to all channels of the ABI. However, further improvements can be achieved by ingesting all observations at their native resolutions and retrieving precipitation at 2 km resolution. It should be noted that the test dataset contains observations from all times of the day, so these improvements are not constrained by the availability of daylight.

4.2 Validation against rain gauge data

In contrast to Hydronn, all of the reference algorithms considered here neglect the probabilistic nature of the retrieval and provide only a single precipitation estimate. The evaluation of Hydronn against the reference algorithms is therefore performed in two steps. Firstly, the accuracy of the deterministic quantitative precipitation estimates is evaluated against the gauge measurements. Secondly, the probabilistic estimates produced by Hydronn and their potential to improve the characterization of the observed precipitation are assessed.

4.2.1 Quantitative precipitation estimates

Accuracy metrics of all retrievals evaluated against the gauge measurements for hourly, daily and monthly precipitation means are provided in Tab. 3. In terms of correlations for hourly means, HYDRO yields the worst performance with a correlation of 0.134, followed by PERSIANN CCS with a correlation of 0.26. IMERG and Hydronn$_{4,IR}$ achieve similar accuracy for hourly estimates with a correlation around 0.4. The Hydronn$_{4,All}$ and Hydronn$_{2,All}$ retrievals further improve the accuracy with correlations of 0.5 and 0.545, respectively. As the integration time increases, the accuracy of all retrievals improves. For daily means, the ranking of the retrieval algorithms remains the same as for hourly means. This is also the case for monthly means with the exception that the accuracy of IMERG increases to the level of the best Hydronn configuration. A likely explanation for this is the calibration that is applied to the IMERG Final product, which matches it to monthly gauge measurements.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Table 3. Accuracy metrics for the retrieved mean precipitation compared to gauge measurements at different time scales. The best values in each column are marked using bold font.

		MAE [mm h^{-1}]			MSE [(mm h^{-1})^2]			Correlation		
Retrieval	Bias [mm h^{-1}]	Hourly	Daily	Monthly	Hourly	Daily	Monthly	Hourly	Daily	Monthly
HYDRO	-0.030	0.308	0.208	0.104	2.964	0.212	0.019	0.134	0.421	0.629
PERSIANN CCS	0.088	0.382	0.274	0.144	3.417	0.293	0.04	0.26	0.415	0.55
IMERG	0.015	0.282	0.191	0.077	2.262	0.178	**0.013**	0.389	0.574	0.741
$\text{Hydronn}_{4,\text{IR}}$	-0.023	0.277	0.191	0.096	2.091	0.163	0.017	0.412	0.573	0.650
$\text{Hydronn}_{4,\text{All}}$	**0.002**	0.247	0.168	0.081	1.874	0.135	0.014	0.502	0.662	0.731
$\text{Hydronn}_{2,\text{All}}$	0.011	**0.239**	**0.162**	**0.075**	**1.760**	**0.128**	**0.013**	**0.545**	**0.685**	**0.756**

A graphical analysis of the accuracy of the retrieved daily accumulations is provided in Fig 6. In this representation the large uncertainties that are present in all retrievals are evident. Nonetheless, the results confirm the general findings from the analysis above. The two conventional VIS/IR retrievals, HYDRO and PERSIAN CCS, yield the least accurate results. In particular, both retrievals show a tendency to miss or strongly underestimate accumulations below 50 mm d^{-1}. This tendency is decreased in the IMERG results for accumulations > 10 mm h^{-1} but still evident for weaker precipitation. Overall, the Hydronn retrievals achieve higher accuracy for both low and high precipitations, which increases with the information content of the input. Nonetheless, systematic underestimation of strong rain rates affects all Hydronn retrievals.

The spatial distribution of the biases of the monthly mean precipitation is displayed in Fig. 7. The strongest biases are observed in the PERSIANN CCS results, which strongly overestimate precipitation in central and northern Brazil. HYDRO and $\text{Hydronn}_{4,\text{IR}}$, as well as to a lesser extent PERSIANN CCS, $\text{Hydronn}_{4,\text{All}}$ and $\text{Hydronn}_{2,\text{All}}$, exhibit a systematic dry bias in southern Brazil. Overall, the biases of IMERG are smallest in magnitude and exhibit the least extent of spatial correlation. However, the differences between IMERG and the best Hydronn configuration, $\text{Hydronn}_{2,\text{All}}$, are small.

Finally, we consider the derived daily cycles of precipitation, which are displayed in Fig. 8. From the reference retrievals, both IMERG and HYDRO yield relatively good agreement with the gauge measurements. IMERG is slightly closer to the gauge measurements during morning and early afternoon but overestimates precipitation in the afternoon and evening. HYDRO slightly underestimates precipitation during the first half of the day but its afternoon peak, despite being close in magnitude to that of the gauge measurements, is delayed by about three hours. PERSIANN CCS shows good agreement with the gauge measurements in the first half of the day but strongly overestimates the afternoon peak. All Hydronn configurations yield good agreement with gauge measurements. The $\text{Hydronn}_{2,\text{All}}$ configurations slightly overestimates precipitation before 10 am, while $\text{Hydronn}_{4,\text{IR}}$ underestimates the afternoon peak.

4.2.2 Probabilistic estimates

We now proceed to evaluate the probabilistic precipitation estimates that are produced by Hydronn. As explained in Sec. 3.5, two probabilistic estimates of the hourly precipitation rates were produced. The first one, (dep.), assumes strong dependence

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

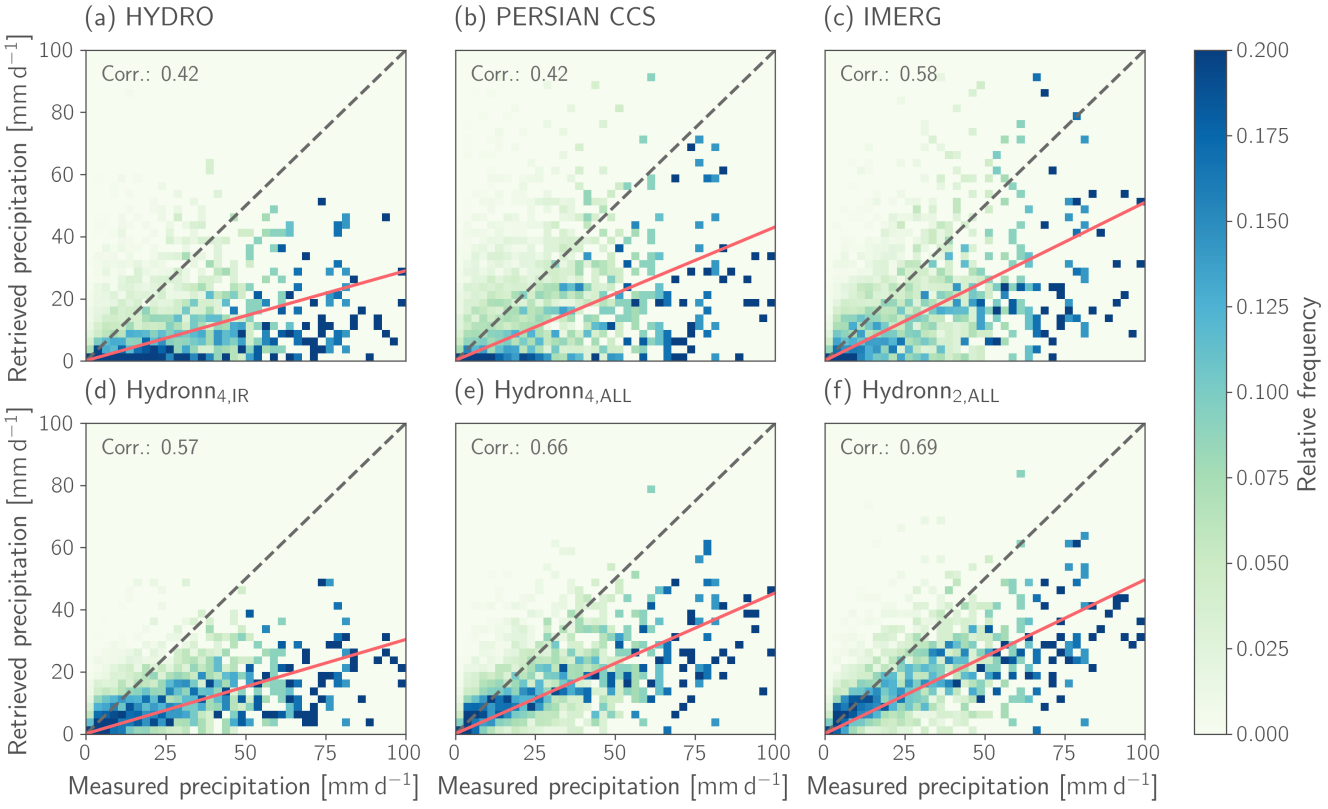

Figure 6. Scatter plots of retrieved daily accumulations against gauge measurements for the reference retrievals and the three Hydronn configurations. Frequencies have been normalized column-wise to improve the visibility of high reference precipitation.

of retrieval errors for consecutive observations, while the second one, (indep.), assumes independent errors. In addition to that, for each of these predictions a corrected distribution has been calculated using the correction factors described in Sec. 3.6. This yields four probabilistic predictions for each Hydronn configuration. These predictions will be referred to with the configuration name and the qualifiers (dep.), (dep., corr.) for the predictions derived assuming dependent uncertaintes with and without correction, respectively, and (indep.), (indep. corr) for the corresponding predictions derived using the independence assumption.

We first consider the distribution of precipitation rates as measured by gauges and the retrieval algorithms, which is shown in Fig. 9. All reference retrievals overestimate the frequency of low and moderate rain rates while underestimating the frequency of high rain rates. For each Hydronn configuration, the distributions of the mean as well as a random sample from the retrieval posterior are displayed. The results for the three Hydronn configurations are almost identical to each other. The distribution of the posterior mean exhibits similar characteristics as the retrieved values of the reference retrievals. Although slightly closer to the distribution of gauge measurements, the uncorrected distribution of samples of the posterior distribution obtained assuming temporally independent errors still does not match the gauge distribution very well. This is improved by the a priori correc-

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

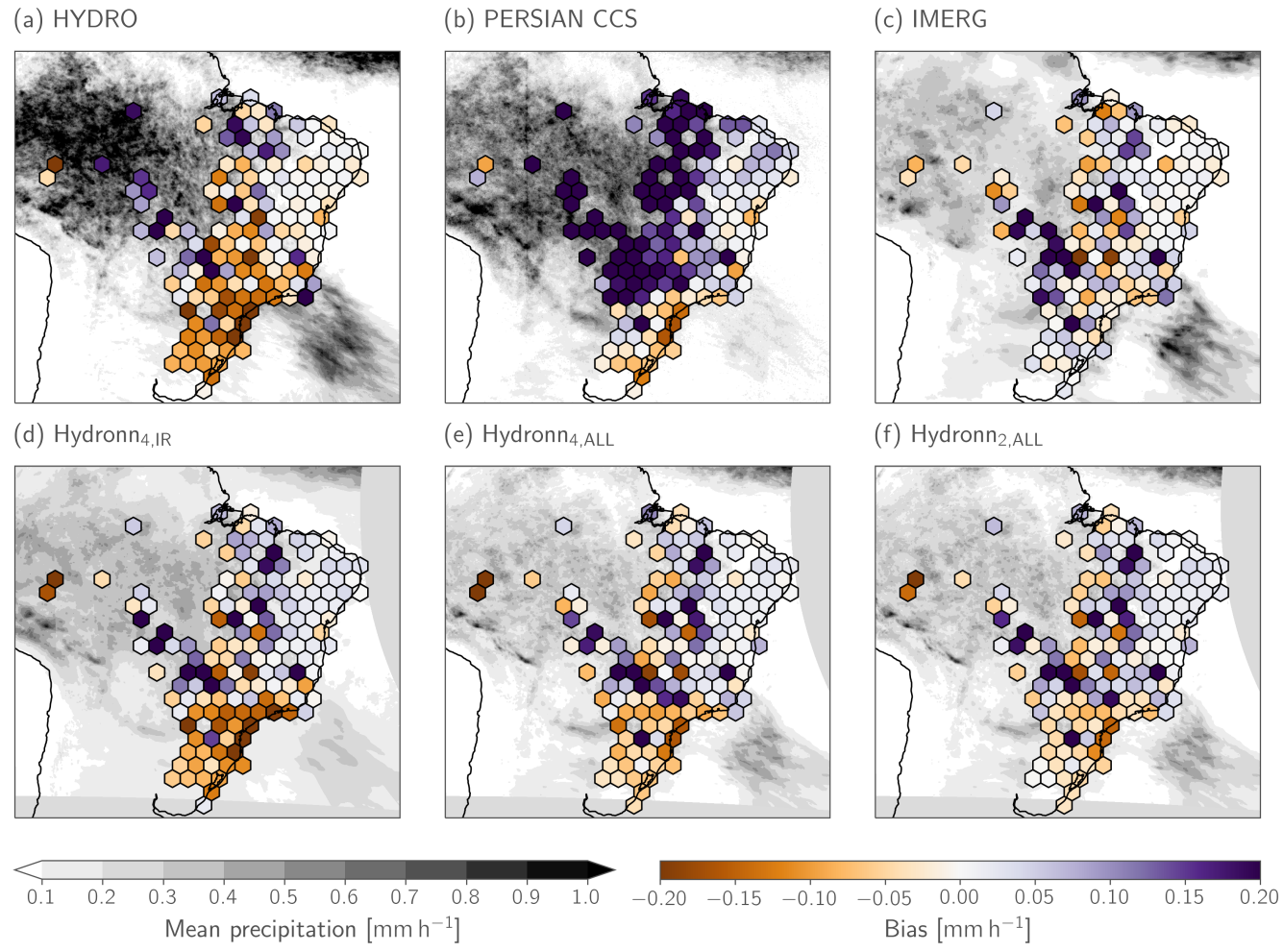

Figure 7. Mean precipitation during December 2020. Shading in the background of each panel shows the spatial distribution of the mean precipitation of the corresponding retrieval. Colored hexagons show the spatial distributions of the retrieval biases with respect to the gauge measurements.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

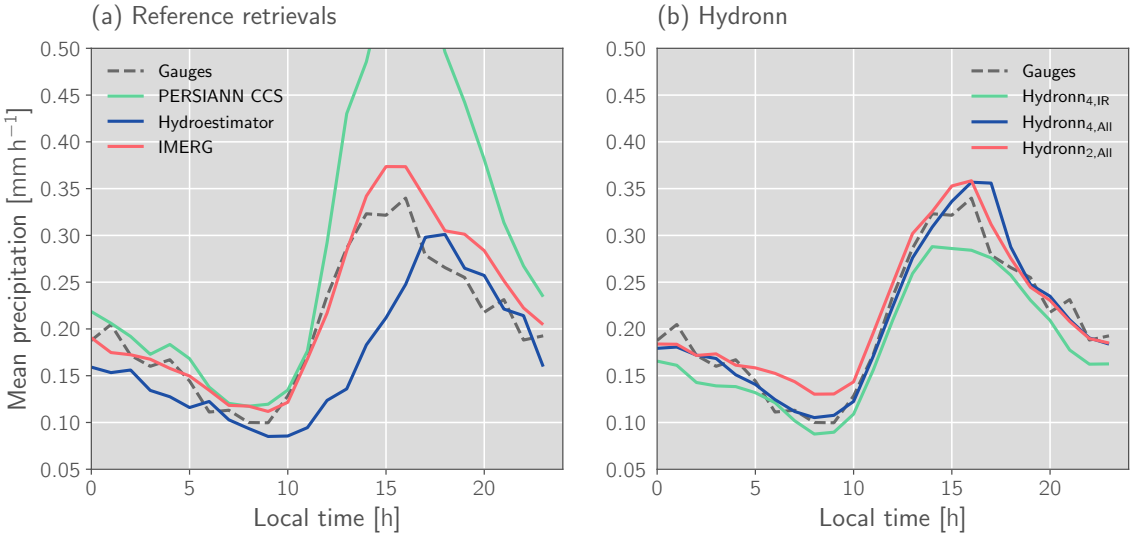

Figure 8. Measured and retrieved daily cycles of precipitation. Panel (a) displays the daily cycles retrieved by the three reference retrievals (solid lines) and the gauge measurements (dashed line) for reference. Panel (b) displays the corresponding diurnal cycles for the three Hydronn configurations (solid lines).

tion but an overestimation of moderate precipitation remains. Sampling from the posterior distributions corresponding to the assumption of temporally dependent uncertainties yields fairly good agreement with the distribution of gauge measurements. It is improved further by the a priori correction, which mostly improves the agreement at low and moderate rain rates.

The deviations between the distribution of retrieved posterior mean values and the reference distribution can thus be understood as an effect of the uncertainties in the retrieval results. These uncertainties lead to a lack of very high precipitation values in the distribution of retrieved means of the retrieval posterior. When instead of the mean samples from the posterior distribution are considered, and the differences between the a priori and gauge measurement distributions are taken into account, the extreme values of the distribution are correctly reproduced.

In addition to sampling from the posterior distribution, the retrieved quantiles can be used to derive confidence intervals for the predicted precipitation. Their reliability is assessed in Fig. 10, which displays the calibration curves of the confidence intervals. While the assumption of dependent retrieval errors leads to an overestimation of the uncertainties, assuming independent errors leads to underestimation. However, for both assumptions this is corrected by the a priori correction, albeit some underestimation remains for the (indep.) results. The results presented in Fig. 10 use the modified gauge measurements described in Sec. 3.6 to which small random noise has been added to non-zero measurements and zeros were replaced with small random values. This was required because quantiles are ill-defined when the CDF of a quantity is discontinuous. So while the results above show that the predicted uncertainties from the Hydronn retrieval are well calibrated, they would not be when compared against the raw gauge measurements. However, since the modifications to the measured precipitation are well within their uncertainty this still demonstrates the meaningfulness of the predicted confidence intervals.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

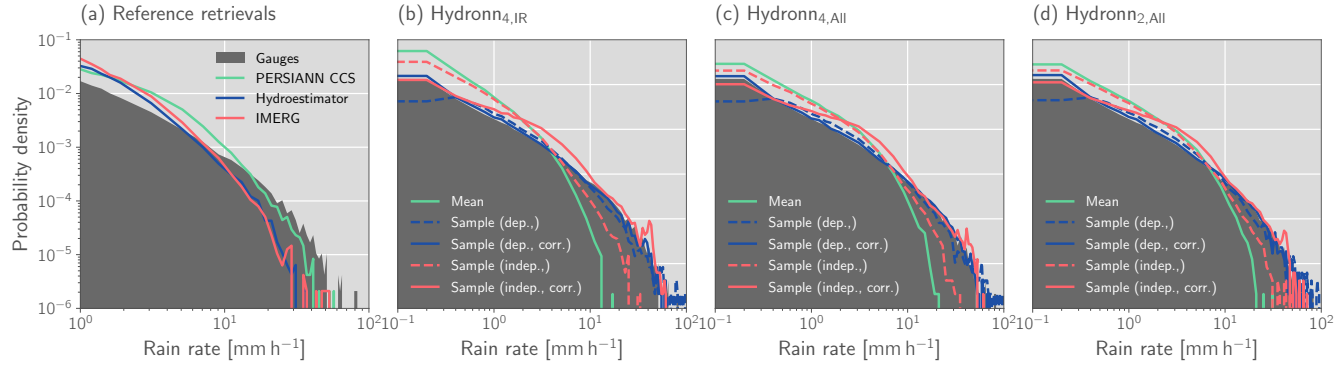

Figure 9. Distributions of measured and retrieved rain rates. The dark grey, filled curve in the background of each panel shows the PDF of the gauge measurements. Colored lines drawn on top show the corresponding PDFs of the retrieved precipitation. The distributions for the reference algorithms are shown in Panel (a). Panel (b), (c) and (d) show the distribution of mean values and random samples from the retrieval posterior.

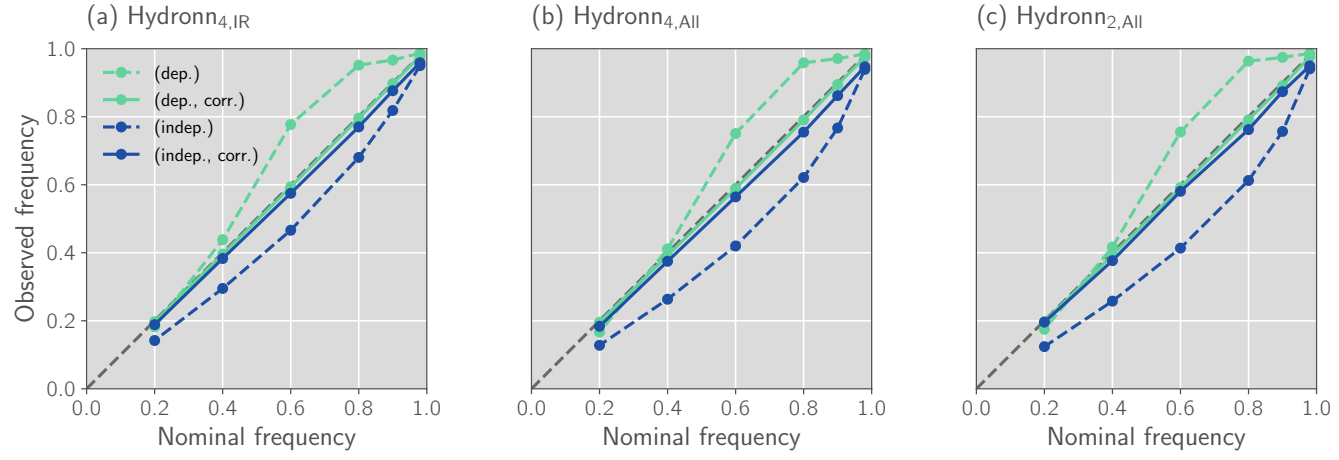

Figure 10. Calibration of the confidence intervals derived using the quantiles of the retrieval posterior distribution predicted by each Hydronn configuration. Grey, dashed line in the background shows the expected results for perfectly calibrated results.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

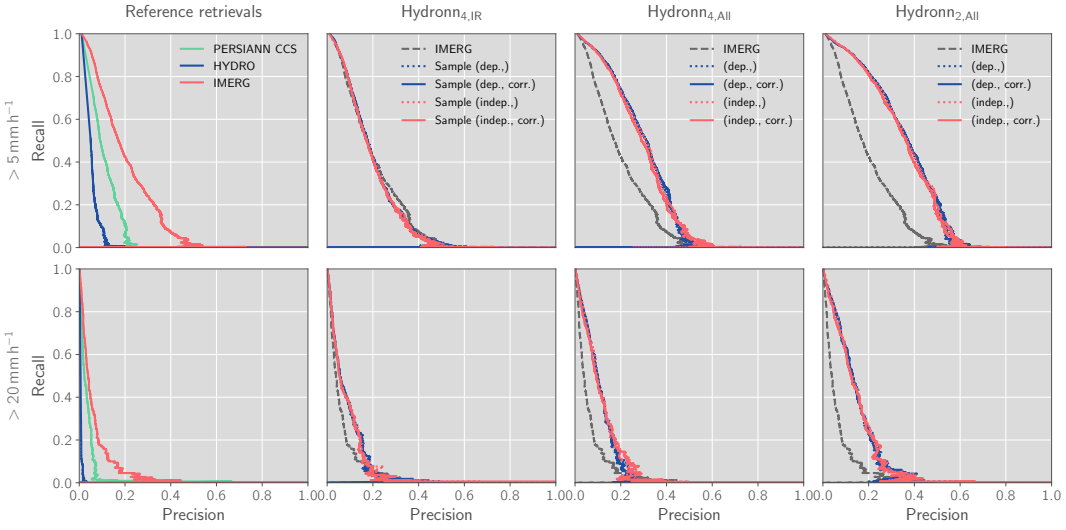

Figure 11. Precision-recall curves for the detection of precipitation events with precipitation rates larger than $5\ \mathrm{mm\ h^{-1}}$ (first row) and $20\ \mathrm{mm\ h^{-1}}$ (second row). Columns show the results for the reference retrievals as well as for each of the Hydronn configurations.

The retrieved quantiles can also be used derive probabilities of an observed pixel exceeding certain precipitation thresholds. This has been used to derive probabilities of the hourly precipitation exceeding 5 and $20\ \mathrm{mm\ h^{-1}}$, which correspond roughly to the 99 and $99.9\ \%$ of the CDF of precipitation rates. The ability of the retrievals to detect high-impact precipitation events is assessed using precision-recall (PR) curves in Fig. 11. For the non-probabilistic retrievals the curves were generated using the predicted precipitation and classifying all pixel above a varying threshold as exceeding the sought-after precipitation rate. The corresponding curves for the Hydronn retrievals were obtained by varying the probability threshold above which a pixel is classified as an high-impact event.

For the detection of events exceeding $5\ \mathrm{mm\ h^{-1}}$, HYDRO again exhibits the least skill, followed by PERSIANN CCS. Compared to these two, IMERG performs clearly better. The detection skill of $\mathrm{Hydronn_{4,IR}}$ is close to that of IMERG, while the two other configurations yield significantly better detection performance than IMERG. Interestingly, the assumption used to accumulate the uncertainties as well as the a priori correction do not have a significant effect on the detection skill. The reason for this is likely that the variation of the probability threshold for the generation of the PR curves has a calibrating effect on the retrieval results. For events exceeding $20\ \mathrm{mm\ h^{-1}}$, all retrievals yield worse detection accuracy. Also here HYDRO exhibits the least skill, followed by PERSIANN CCS and IMERG, which yield very similar results. All Hydronn configurations outperform the reference retrievals.

Since the Hydronn retrievals can be used to derive a probability of an observation exceeding a given precipitation threshold, a relevant question is how accurate these probabilities are. The calibration of the detection probabilities of events exceeding $5\ \mathrm{mm\ h^{-1}}$ is displayed in Fig. 12. While the predictions derived assuming temporally dependent uncertainties are fairly well

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

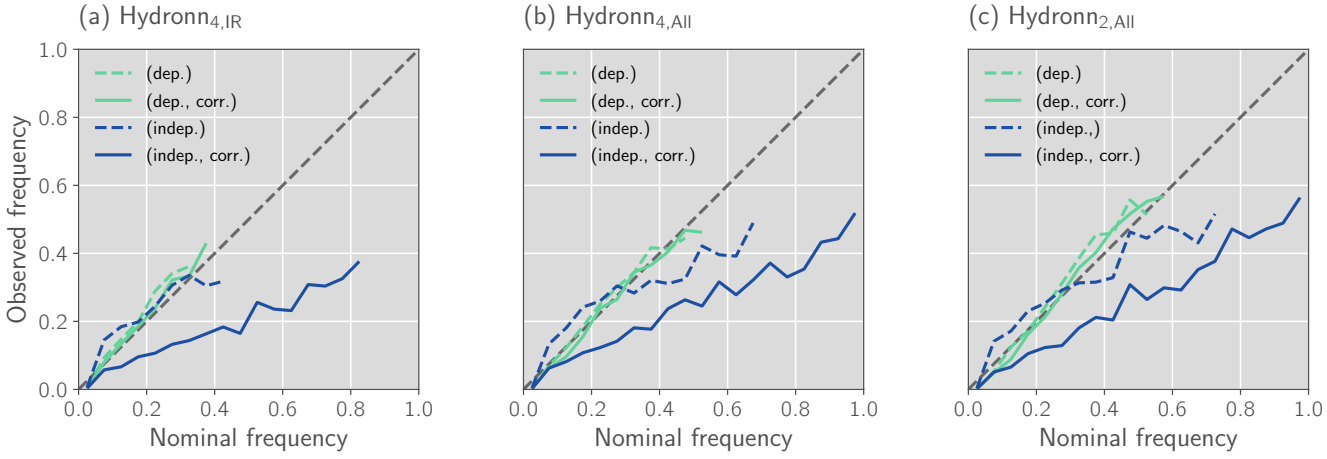

Figure 12. Calibration of the probabilistic precipitation event detection.

calibrated, larger deviations are observed for the probabilities derived assuming independent uncertainties. Moreover, while the correction marginally improves the calibration of the results for the dependent uncertainties, it degrades that of the probabilities derived assuming independent uncertainties. While the results are qualitatively similar across all Hydronn configurations, the detection probabilities increase in magnitude with the information content of the retrieval inputs, indicating that the sharpness of the probabilistic predictions increases. Nonetheless, even for an event with a threshold of only $5\ \mathrm{mm\ h^{-1}}$ probabilities do not exceed $60\ \%$, which highlights the limitations of the VIS/IR based retrievals for the detection of extreme precipitation.

In addition to the PR curves, we have calculated commonly used detection metrics for all retrievals in order to simplify the comparison to other studies. These metrics include the probability of detection (POD, the fraction of events that were correctly detected), the false alarm rate (FAR, the fraction of all detections that are wrong) and the critical success index (CSI) for the detection of events with more than $20\ \mathrm{mm\ h^{-1}}$. The results are displayed in Tab. 4. For the calculation of POD and FAR for the reference retrievals, a retrieved precipitation rate exceeding $20\ \mathrm{mm\ h^{-1}}$ is counted as a detection. For the Hydronn retrievals the probability threshold has been tuned to yield a FAR close to that of IMERG. As was to be expected from the PR curves in Fig. 11, HYDRO exhibits the least detection skill and does not detect any of the actual occurrences of strong precipitation but still produces false positives. Compared to this, IMERG does significantly better but suffers from a very low POD. PERSIANN performs better than IMERG in terms of POD but also has a higher FAR. The Hydronn retrievals achieve higher POD at lower FAR for all configurations. The POD of extreme precipitation increases markedly as more channels and higher resolution are incorporated. Similar results are observed for the CSI.

4.3 Case study

As final part of this evaluation, a case of heavy precipitation in the city of Duque de Caxias in the State of Rio de Janeiro is considered, which occurred between the 22nd and 24th of December 2020 and lead to floodings (Fohla De S. Paulo, 2020). About

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Table 4. Error metrics for the detection of precipitation events with rain rates exceeding $20\ \mathrm{mm\ h^{-1}}$.

Retrieval		POD	FAR	CSI
PERSIANN CCS		0.006	0.987	0.004
PERSIANN CCS		0.039	0.926	0.026
HYDRO		0.000	1.000	0.000
IMERG		0.018	0.867	0.016
$\mathrm{Hydronn_{4,IR}}$	(dep.)	0.085	0.803	0.063
	(dep., corr.)	0.121	0.811	0.080
	(indep.)	0.079	0.806	0.059
	(indep., corr.)	0.076	0.798	0.058
$\mathrm{Hydronn_{4,All}}$	(dep.)	0.109	0.800	0.076
	(dep., corr.)	0.133	0.808	0.085
	(indep.)	0.167	0.801	0.100
	(indep., corr.)	0.145	0.800	0.092
$\mathrm{Hydronn_{2,All}}$	(dep.)	0.242	0.799	0.123
	(dep., corr.)	0.270	0.804	0.128
	(indep.)	0.267	0.799	0.129
	(indep., corr.)	0.242	0.801	0.123

$250\ \mathrm{mm}$ of accumulated precipitation were measured by the rain gauge in Xerém over the period of two days. An overview of the spatial distribution of retrieved accumulated precipitation over the two days is provided in Fig. 13. Fairly good agreement is observed between IMERG and the $\mathrm{Hydronn_{4,All}}$ and $\mathrm{Hydronn_{2,All}}$ configurations. $\mathrm{Hydronn_{4,IR}}$ yields accumulations of slightly lower magnitude and is spatially less well defined than the two other Hydronn retrievals. The accumulations from PERSIANN CCS are lowest in terms of magnitude, whereas those from HYDRO are highest. However, neither of the two agrees well with the spatial distribution retrieved by IMERG or the Hydronn retrievals. Both PERSIANN CCS and HYDRO exhibit systematic dry biases in the region around Duque de Caxias compared to the gauge measurements, which is present to a lesser degree also in the results of $\mathrm{Hydronn_{4,IR}}$. This dry bias is less pronounced in the results of IMERG, $\mathrm{Hydronn_{4,All}}$ and $\mathrm{Hydronn_{2,All}}$.

The rain rates at the gauge station in Xerém (location marked by green star in Fig. 13) are displayed in Fig. 14. The plots show the hourly precipitation rates retrieved by the reference retrievals as well as the mean and posterior distribution for all Hydronn retrievals. Only results obtained with the assumption of dependent retrieval uncertainties and the a priori correction applied are shown. The precipitation measured at the rain gauge far exceeds the precipitation measured by any of the reference retrievals or the retrieved mean of the Hydronn retrievals. The Hydronn retrievals predict elevated uncertainties for the period during which the strongest precipitation is observed. However, the precipitation peaks still exceed the 99th percentile. Two factors may explain that more than the expected $1\ \%$ of gauge measurements lie outside the predicted uncertainty range. Firstly, the observations considered here are not randomly sampled but correspond to an event that is known to be extreme.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

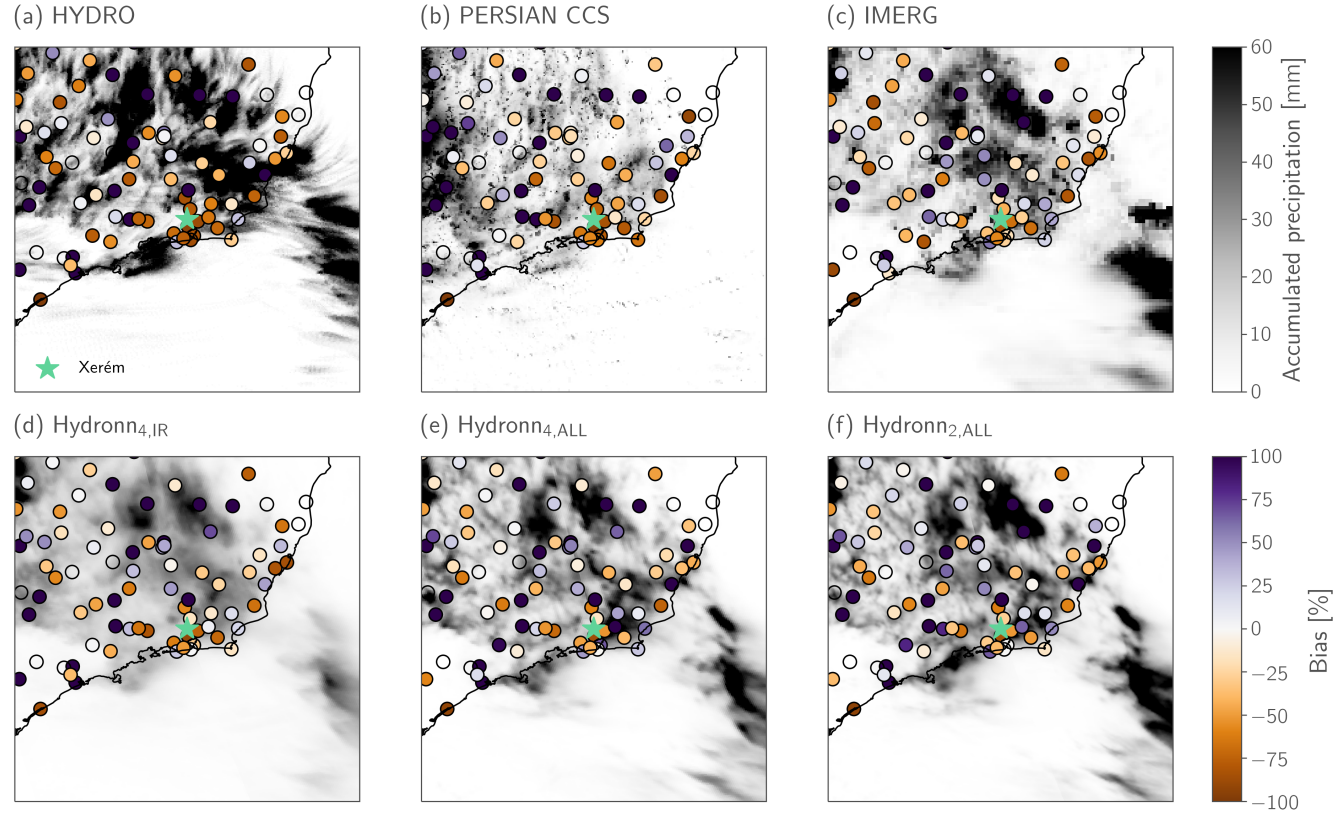

Figure 13. Retrieved precipitation for an extreme precipitation event in the city of Duque de Caxias in the state of Rio de Janeiro. Shading in background shows the accumulated precipitation between 2020-12-22 00:00 to 2020-12-24 00:00. Colored markers show the relative bias of the retrieved accumulated precipitation compared to gauges. The green star marks the location of Xerém, where flooding occurred.

Secondly, as stated in the article in Fohla De S. Paulo (2020), heavy precipitation events are common in this region. This may indicate that regional factors may act to intensify the precipitation, which is unlikely to be captured in the training data of the retrieval.

Nonetheless, an encouraging results is that the predicted value of the 99th percentile increases with the information content in the retrieval input. This indicates that the neural network can leverages the additional information to produce sharper uncertainty estimates.

5 Discussion

The study presented Hydronn, a neural-network-based precipitation retrieval for Brazil, which has been trained using combined radar and radiometer measurements from the GPM Core Observatory satellite. The retrieval compares favorably against the

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

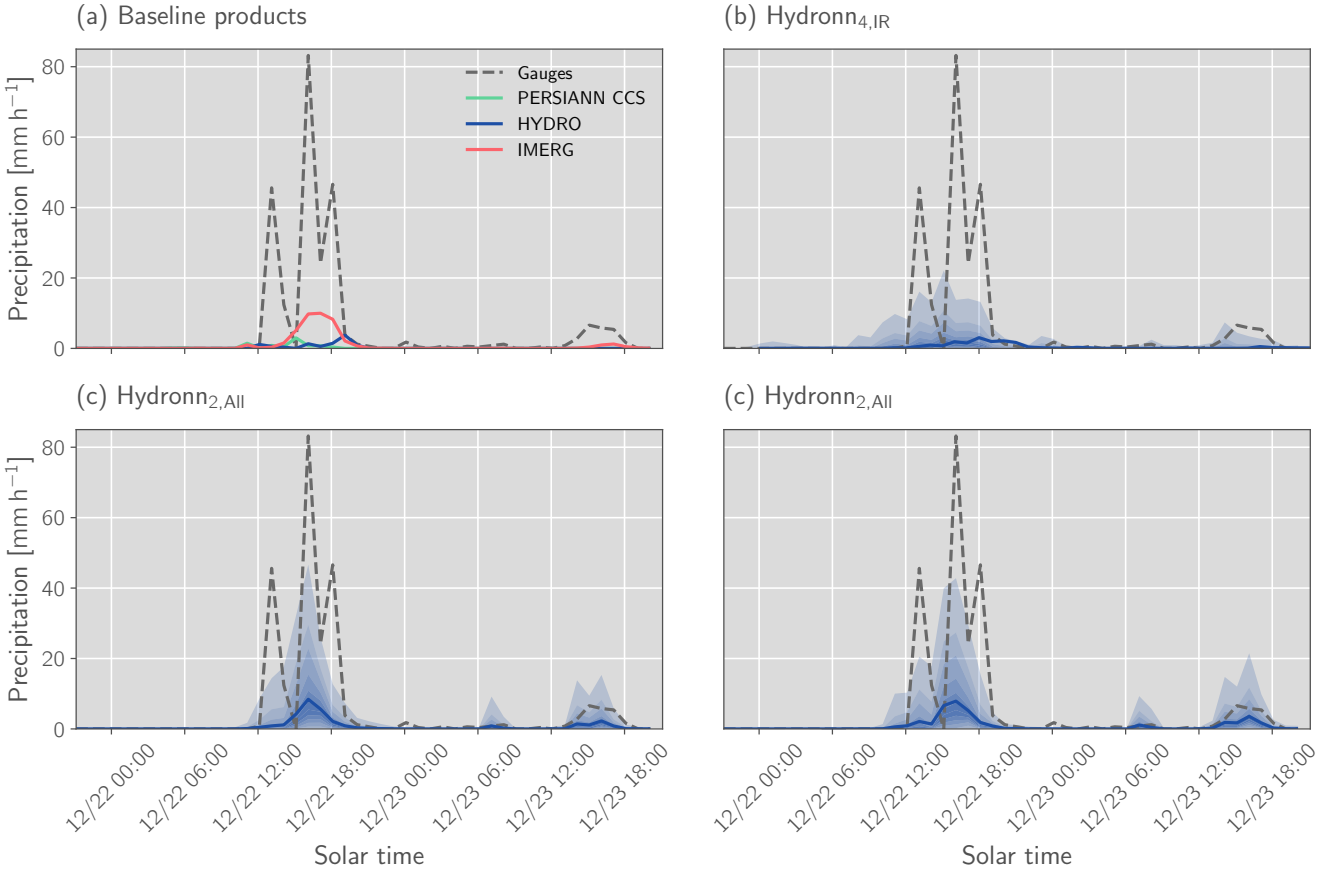

Figure 14. Retrieved precipitation for an extreme precipitation even that occurred between 2020-12-22 and 2020-12-24 in the cite of Duque de Caxias in the state of Rio de Janeiro. Grey, dashed lines show the precipitation by the gauge station in Xerém. Solid lines show the retrieved mean precipitation for each retrieval algorithm. The shading shows filled contours of the posterior CDF at values $[0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99]$.

currently operational precipitation retrieval, HYDRO, as well as the PERSIANN CCS product. In its best configuration the Hydronn retrieval yields retrieval accuracy superior to that of the IMERG Final product across most considered metrics.

5.1 Information content of VIS/IR observations

The three tested retrieval configurations use input observation of increasing information content. The $\text{Hydronn}_{4,\text{IR}}$ configuration only uses a single IR channel at a resolution of $4\ \text{km}$ while $\text{Hydronn}_{4,\text{All}}$ uses all available bands. The best performing retrieval, $\text{Hydronn}_{2,\text{All}}$, combines the observations from all channels of the GOES ABI at their native resolutions. Clear increases in retrieval performance are observed when all ABI bands are incorporated into the retrieval as well as when all channels are ingested at their native resolutions. This demonstrates the ability of the neural-network-based retrieval to learn complex re-

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

lationships even from input observations with low information content. The fact that HYDRO and Hydronn$_{4,\mathrm{IR}}$ use the same observations as retrieval input further highlights the need for advanced statistical retrieval techniques to fully exploit the potential of geostationary VIS/IR observations. This is in good agreement with the findings from Sadeghi et al. (2019), who also report improvements when comparing PERSIANN CCS to a CNN-based retrieval.

5.2 Probabilistic precipitation retrievals

A novel aspect of the proposed precipitation retrievals is their ability to provide probabilistic precipitation estimates. In this study we have demonstrated multiple ways in how this may improve the utility of the retrieval results:

1. The results in Fig. 9 show that samples from the retrieval posterior reproduce the gauge-measured distribution of rain rates fairly accurately. The deviations of the distribution of the posterior mean from the gauge measurements can thus be understood as a consequence by the statistical properties of this estimator instead of a retrieval deficiency. The random samples may be useful for applications that are sensitive to heavy precipitation rates, such as run off modeling or climatological studies. As an example, Fig. 15 shows scatter plots of the 99th percentile of the distribution of gauge-measured and retrieved precipitation during December 2020 for all gauge station. Also here the Hydronn retrievals yield the best estimates. For this evaluation, HYDRO and PERSIANN CCS yield similar accuracy as IMERG despite IMERG having higher accuracy for all other metrics considered in this study. This is likely because both HYDRO and PERSIANN CCS were both developed to correctly represent heavy precipitation, which harms their accuracy in terms of other statistics. By explicitly resolving the probabilistic nature of the precipitation retrieval, HYDRONN can provide both climatologically accurate accumulations (see Tab. 3) and correct representation of heavy precipitation.

   It should be noted, however, that these random samples do not take into account spatial correlations. To what extent this may negatively impact applications of the retrieval results remains to be investigated in a follow-up study.

2. The retrieved quantiles allow the derivation of confidence intervals to quantify retrieval uncertainty. By correcting for the difference in a priori distributions as well as the degeneracy of quantiles due to discontinuities in the CDF of gauge measurements, we were able to show that the retrieval uncertainties are well calibrated even against gauge measurements (Fig. 10). Due to the large uncertainties that are inherent to precipitation retrieval from VIS/IR observations (Fig. 5, 6), quantifying them increases the trustworthiness of the predictions.

3. We have shown that the retrieved quantiles can be used to detect heavy precipitation events (Fig. 11, Tab. 3). Here all Hydronn retrievals perform better than IMERG although they are based on observations with a significantly lower information content. This clearly shows the benefits of quantifying the retrieval uncertainties. Moreover, we were able to show that the probabilistic detection of these events is fairly well calibrated (Fig.12).

   These results, however, also show the limitation of VIS/IR retrievals since even for events exceeding $5\ \mathrm{mm\ h^{-1}}$ the maximum detection confidence is $60\ \%$. This certainly puts the suitability of any of the considered retrievals to detect heavy

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

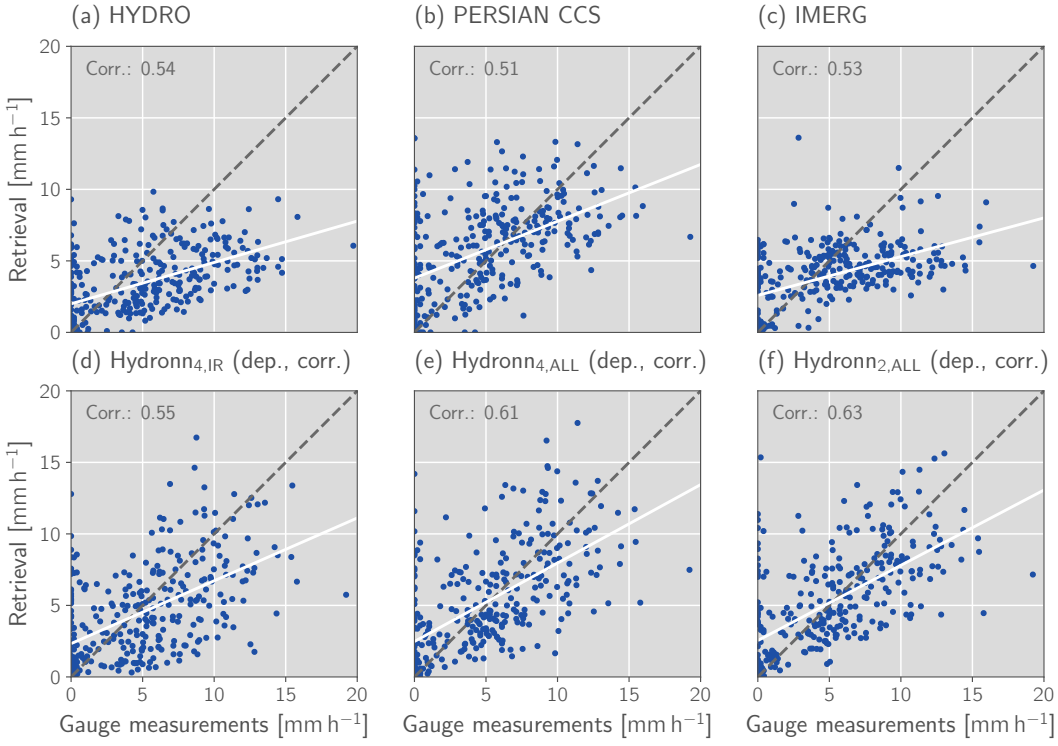

Figure 15. The 99th quantile of the distribution of gauge measurements of each gauge station plotted against the 99th quantile of the corresponding distribution of retrieved precipitation.

precipitation events into question and calls for new approaches that combine observations from different observational sources.

Finally, we have also investigated how uncertainties from instantaneous precipitation retrievals can be propagated to the full hour. The two approaches that we have tested correspond to assuming temporally independent and temporally dependent retrieval uncertainties. Our results indicate (Fig. 9, Fig. 10) that the assumption of dependent uncertainties overestimates the actual retrieval uncertainty, whereas assuming independent uncertainties underestimates actual uncertainties. It is interesting to note that the way the uncertainties are accumulated does not affect the detection of extreme events (Fig. 11), which indicates that the probabilities could also be re-calibrated a posteriori.

5.3 Utility of a priori corrections

We have proposed a method to correct for the distribution of precipitation rates in the training data. The corrections have improved the agreement between the distribution of retrieved precipitation rates as well as the calibration of the uncertainty intervals (Fig. 9, Fig. 10). Although for the assumed independent uncertainties the calibration was improved, the distribution of precipitation rates did exhibit slight deviations from the distribution of the gauge measurements. We suspect that the reason

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

for the correction working worse in the latter case is that the a corresponding a priori assumption deviates stronger from the distribution of the gauge measurements (Fig. 4). This led to much higher correction factors, which were truncated to avoid numerical issues.

The clearest effect of the a priori correction was observed when the predicted confidence intervals were evaluated against gauge data (Fig. 10). This allowed us to show that the Hydronn retrievals can provide well-calibrated uncertainty estimates for their predictions when the differences between the a priori distributions of the training data and the gauge measurements are taken into account. Nonetheless, the correction did not affect the detection of strong precipitation. We suspect the reason for this to be that the correction mostly affects small precipitation rates due to their high occurrence in the training and validation data as well as the re-calibrating effect of the varying probability threshold in the generation of the PR curves.

The correction relies on the assumptions that the conditional distribution of the observations vector $p(\mathbf{y}|x)$ given the rain rate x remains constant. It can thus only correct for differences in the measurement characteristics between the rain gauge data, which is used to evaluate the retrieval, and the GPM data, which was used to derive the training data. It can not, however, correct for differences between the training and evaluation data that involves changes in the observed processes, which would change $p(\mathbf{y}|x)$. We argue that this is not an issue for this study since the evaluation and training data are overlapping geographically.

6 Conclusions

Hydronn, the presented neural-network-based precipitation retrieval, improves real time precipitation estimates over Brazil. Its performance is superior to both the currently operational algorithm as well as the much more complex, global IMERG Final product, which combines observations from both VIS/IR and passive microwave sensors as well as global gauge measurements.

Our results demonstrate the potential of region-specific retrieval algorithms, which exploit the full potential of locally available satellite observations. This is made possible by the availability of accurate surface precipitation retrievals from the GPM CO satellite, which were used to derived the training data for the retrieval. Since this data is available globally between -60 and $60\ ^{\circ}\mathrm{N}$, the approach can potentially be applied to most other regions around the world.

Although our evaluation focused on Brazil, many of the results presented here should be of interest for precipitation retrievals from geostationary satellites in general. In addition to providing further evidence of the potential of deep neural networks to improve quantitative precipitation estimates, we show how a probabilistic regression approach can be used to perform VIS/IR precipitation retrievals using a Bayesian framework and that the probabilistic predictions improve the representation of the characteristics of the observed precipitation.

Finally, the fact that our relatively simple retrieval outperforms state-of-the-art precipitation products despite being solely based on VIS/IR observations, shows the potential of algorithmic innovation for quantitative precipitation estimation. The ability of the neural network retrieval to leverage information from all channels of the ABI at their native resolutions, shows the strength of the end-to-end approach to retrieval design. This suggests that further improvements for precipitation retrievals should be achievable by expanding the retrieval input to incorporate additional spectral as well as temporal information.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Code availability. The code to generate the training data, train the machine learning models, run the retrievals and analyze the results is available from a public repository (Pfreundschuh, 2022).

Author contributions. II, PE and SP designed the study. SP and II developed the retrieval and analyzed the retrieval results. SP prepared the manuscript. AC and DV provided the gauge measurements, HYDRO retrieval results and valuable feedback.

Competing interests. No competing interests are present.

Acknowledgements. We would like to acknowledge the Brazilian National Institute of Meteorology for the provision of the gauge measurements.

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) were used for the processing of satellite data, PyTorch (Paszke et al., 2019) for implementing the machine learning models as well as matplotlib (Hunter, 2007) and cartopy (Met Office, 2010 - 2015) for generating figures.

The training of the machine learning models used in the study were performed on resources at Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC).

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

References

Adler, R. F. and Negri, A. J.: A satellite infrared technique to estimate tropical convective and stratiform rainfall, Journal of Applied Meteorology and Climatology, 27, 30–51, 1988.

Arkin, P. A. and Meisner, B. N.: The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84, Monthly Weather Review, 115, 51–74, 1987.

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.

de Siqueira, R. A. and Vila, D.: Hybrid methodology for precipitation estimation using Hydro-Estimator over Brazil, International Journal of Remote Sensing, 40, 4244–4263, https://doi.org/10.1080/01431161.2018.1562262, 2019.

Fohla De S. Paulo: https://www1.folha.uol.com.br/cotidiano/2020/12/chuva-causa-morte-derruba-casas-e-deixa-familias-desalojadas-em-duque-de-caxi shtml, accessed: 2022-01-31, 2020.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, J. Atmos. Sci., 102, 359–378, https://doi.org/10.1198/016214506000001437, 2007.

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.

Holleman, I.: Bias adjustment and long-term verification of radar-based precipitation estimates, Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 14, 195–203, 2007.

Hong, Y., Hsu, K. L., Sorooshian, S., and Gao, X. G.: Precipitation estimation from remotely sensed imagery using an artificial neural 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.

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. 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.

Ingemarsson, I.: Retrieving precipitation over Brazil. A quantile regression neural networks approach, https://hdl.handle.net/20.500.12380/304390, 2021.

Kuligowski, R. J.: A Self-Calibrating Real-Time GOES Rainfall Algorithm for Short-Term Rainfall Estimates, Journal of Hydrometeorology, 3, 112 – 130, https://doi.org/10.1175/1525-7541(2002)003<0112:ASCRTG>2.0.CO;2, 2002.

Kuligowski, R. J., Li, Y., Hao, Y., and Zhang, Y.: Improvements to the GOES-R Rainfall Rate Algorithm, Journal of Hydrometeorology, 17, 1693 – 1704, https://doi.org/10.1175/JHM-D-15-0186.1, 2016.

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

Nguyen, P., Ombadi, M., Gorooh, V. A., Shearer, E. J., Sadeghi, M., Sorooshian, S., Hsu, K., Bolvin, D., and Ralph, M. F.: PERSIANN Dynamic InfraredâRain Rate (PDIR-Now): A Near-Real-Time, Quasi-Global Satellite Precipitation Dataset, Journal of Hydrometeorology, 21, 2893 – 2906, https://doi.org/10.1175/JHM-D-20-0177.1, 2020.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

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.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.: Hydronn, https://doi.org/10.5281/zenodo.6371712, 2022.

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.

Pradhan, R. K., Markonis, Y., Godoy, M. R. V., Villalba-Pradas, A., Andreadis, K. M., Nikolopoulos, E. I., Papalexiou, S. M., Rahim, A., Tapiador, F. J., and Hanel, M.: Review of GPM IMERG performance: A global perspective, Remote Sensing of Environment, 268, 112 754, 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.

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

Satyamurty, P., Nobre, C. A., and Silva Dias, P. L.: South America, pp. 119–139, American Meteorological Society, Boston, MA, https://doi.org/10.1007/978-1-935704-10-2_5, 1998.

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., and Bachmeier, A. S.: Introducing the next-generation Advanced Baseline Imager on GOES-R, Bulletin of the American Meteorological Society, 86, 1079–1096, 2005.

Schmit, T. J., Lindstrom, S. S., Gerth, J. J., and Gunshor, M. M.: Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI), 2018.

Scofield, R. A. and Kuligowski, R. J.: Status and Outlook of Operational Satellite Precipitation Algorithms for Extreme-Precipitation Events, Weather and Forecasting, 18, 1037 – 1051, https://doi.org/10.1175/1520-0434(2003)018<1037:SAOOOS>2.0.CO;2, 2003a.

Scofield, R. A. and Kuligowski, R. J.: Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events, Weather and Forecasting, 18, 1037–1051, 2003b.

Scofield, R. A. and Oliver, V. J.: A scheme for estimating convective rainfall from satellite imagery, 1977.

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

Smith, J. A., Seo, D. J., Baeck, M. L., and Hudlow, M. D.: An Intercomparison Study of NEXRAD Precipitation Estimates, Water Resources Research, 32, 2035–2045, https://doi.org/https://doi.org/10.1029/96WR00270, 1996.

Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J., and Kalchbrenner, N.: Metnet: A neural weather model for precipitation forecasting, arXiv preprint arXiv:2003.12140, 2020.

https://doi.org/10.5194/egusphere-2022-78
Preprint. Discussion started: 6 April 2022

Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D.: Evaluation of PERSIANN system satellite based estimates 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.

UCI CHRS Data Portal: UCI CHRS data portal, http://persiann.eng.uci.edu/CHRSdata, accessed: 2022-01-27, 2022.

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.

Vicente, G. A., Scofield, R. A., and Menzel, W. P.: The Operational GOES Infrared Rainfall Estimation Technique, Bulletin of the American Meteorological Society, 79, 1883 – 1898, https://doi.org/10.1175/1520-0477(1998)079<1883:TOGIRE>2.0.CO;2, 1998.