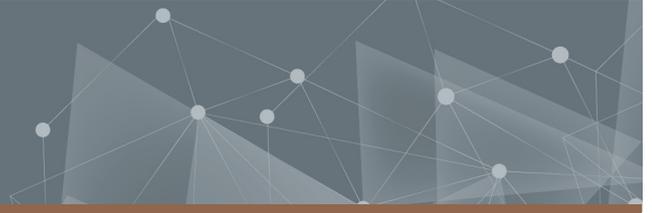




CHALMERS
UNIVERSITY OF TECHNOLOGY



Retrieving precipitation over Brazil

A Quantile Regression Neural Networks approach

Master's thesis in Engineering Mathematics and Computational Science

INGRID INGEMARSSON

DEPARTMENT OF SPACE, EARTH AND ENVIRONMENT

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2021

www.chalmers.se

MASTER'S THESIS 2021

Retrieving precipitation over Brazil

A Quantile Regression Neural Networks approach

INGRID INGEMARSSON



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Space, Earth and Environment
Division of Microwave and Optical Remote Sensing
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021

Retrieving precipitation over Brazil
A Quantile Regression Neural Networks approach
INGRID INGEMARSSON

© INGRID INGEMARSSON, 2021.

Supervisors: Simon Pfreundschuh & Patrick Eriksson,
Department of Space, Earth and Environment

Examiner: Patrick Eriksson,
Department of Space, Earth and Environment

Master's Thesis 2021
Department of Space, Earth and Environment
Division of Microwave and Optical Remote Sensing
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in L^AT_EX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2021

Retrieving precipitation over Brazil
A Quantile Regression Neural Networks approach
INGRID INGEMARSSON
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Close and accurate monitoring of precipitation on a global scale is key to understanding our future climate as well as our current weather. Geostationary weather satellites, as opposed to other measuring methods, provide high-resolution information covering large regions. The sensors carried can not, however, measure precipitation directly but are restricted to capturing cloud top temperatures (in IR radiation). Earlier work presents a range of models that aim to relate these geostationary observations to precipitation, including simple regression as well as more elaborate machine learning techniques. In this thesis we aim at predicting a posterior distribution instead of a single precipitation value for each set of cloud top temperatures. To achieve this, we make use of Quantile Regression Neural Networks (QRNNs), a supervised machine learning approach. The two main questions asked are as follows: Can this deep learning method be used to improve upon algorithms currently in operation? and Can spatial information be used to improve the retrieval? The models are trained on GOES-16 IR data over Brazil with a precipitation product from the GPM Core Observatory. Our results on held-out test data show that it is possible to model the precipitation distribution using a QRNN. Additionally, a 20% decrease in mean squared error and a 25% decrease in mean absolute error is observed on the test data when using the spatially aware model, which illustrates the general performance improvement by utilizing the spatial information. The QRNN models also show promising results on an independent rain gauge dataset where they are compared against the currently-in-operation Hydro-Estimator. Here our most promising QRNN shows a 30% decrease in mean squared error compared to the present model.

Keywords: quantile regression, neural networks, CNN, precipitation, GPM, GOES

Acknowledgements

First and foremost, I would like to thank my supervisor Patrick Eriksson for proposing this interesting project and for his valuable feedback as well as his patience throughout the process. Further, I would like to express my gratitude toward my co-supervisor Simon Pfreundschuh, for his supportiveness and for always taking on my extensive bullet-list questions very pedagogically. His solid work with the QRNNs should also be acknowledged, since it has been elemental for this thesis.

I would also like to express my gratitude toward Alan Calheiros and Daniel Alejandro Vila at the INPE for so generously providing us with the data needed for taking the analysis one step further.

My sincere thanks are also extended to my fellow student Adrià Amell Tosas for so kindly sharing his experiences when discussing common struggles, and to him as well as to Sara Nordin Hällgren for bringing some normality in these strange times with virtual coffee breaks.

Lastly, I want to thank Edvin, for his encouragement and unwavering belief in my ability.

Ingrid Ingemarsson, Gothenburg, November 2021

Contents

List of Figures	xi
List of Tables	xv
Glossary	xvi
Acronyms	xvi
Glossary	xvi
1 Introduction	1
2 Data	3
2.1 Sources	3
2.1.1 GOES-16: Input data	3
2.1.2 GPM: Ground truth data	3
2.1.3 Rain gauges: Reference data	5
2.1.4 HE: Baseline model	6
2.2 Processing	6
2.2.1 Region	6
2.2.2 Time period	7
2.2.3 Selection	7
2.2.4 Resampling	7
2.2.5 Cropping	7
2.2.6 Temporal matching	8
2.2.7 Split	10
2.2.8 Normalization	10
2.3 Dataset properties	10
2.3.1 Sample	10
2.3.2 Precipitation rate distribution	12
3 Modelling	13
3.1 The retrieval problem	13
3.1.1 Bayesian approach	13
3.2 Supervised machine learning	13
3.2.1 Quantile Regression	14
3.3 Artificial Neural Networks	14
3.3.1 Fully Connected Layer	15
3.3.2 Convolutional Layer	15

3.4	Architectures	17
3.4.1	Multilayer perceptron	17
3.4.2	XceptionFPN	17
3.5	Evaluation	20
3.5.1	Regression metrics	20
3.5.2	Categorical metrics	20
3.5.3	Probabilistic accuracy measures	21
3.6	Experimental setup	21
3.6.1	Data loading	21
3.6.2	Hyperparameters	22
4	Results	23
4.1	Evaluation on test set	23
4.2	Evaluation on rain gauge data	29
4.2.1	An extreme precipitation event	34
5	Discussion	37
6	Conclusion	39
	Bibliography	41
A	Appendix 1	I

List of Figures

2.1	Left: GOES-16 satellite geostationary view of Earth. Right: Example of geocolor image captured by the ABI instrument on-board of GOES-16 (image source CIRA/NOAA, <i>GOES Image Viewer</i> (2021)).	4
2.2	Earth in GOES-16 geostationary perspective (left) and in plate carrée projection (right), both overlaid with one 2BCMB data swath.	4
2.3	Map of rain gauge positions from the dataset over Brazil. Color legend corresponding to five regions.	5
2.4	Visualization of the region of interest in two different coordinate frames, the GOES-16 satellite view of Earth (left) and in the plate carrée projection (right).	7
2.5	Example of granulates intersecting a polygon that encloses Brazil in the Earth Data Search tool (<i>Earthdata Search</i> n.d.).	8
2.6	Example of a GPM granulate that is coordinate transformed and cut into three 256×256 pixel pieces. The shaded gray in the squares marks the area that lacks GPM data.	9
2.7	Example of an input tensor of standardized GOES channels.	11
2.8	Example of a GPM label, with precipitation rates (mm/h) and masked pixels marked in gray.	11
2.9	Precipitation distribution for the GPM <i>training</i> data overlaid by the rain gauge data, bin width 0.1 mm/h. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 0.4 mm/h with bin width 0.005 mm/h.	12
3.1	Fully Connected Layer with input $\mathbf{x} \in \mathbb{R}^n$ and output $\mathbf{y} \in \mathbb{R}^m$. Each arrow is associated with a weight w_{ij}	15
3.2	Example of a convolutional filter in the case of a single channel image. The 3×3 -kernel slides through the 7×7 input, performing the convolution at each position, which yields the output.	16
3.3	Illustration of Depth-wise separable convolution applied to a three channel image. First, an ordinary spatial convolution is applied to each channel separately. Second, a point-wise convolution (kernel size 1×1) is performed on the stack of output feature maps from the first step.	18

3.4	Schematic of the <i>XceptionFPN</i> model architecture. To the left: the general pyramid structure of the entire flow, and to the right: the Downsample/ <i>Xception</i> flow (where the dashed region only occurs in the downsampling), the Upsample flow and the Exit flow respectively. The n in the $nXception$ indicates that the module is repeated at this position.	19
4.1	Example predictions of selected QRNN models (right) on a sample from the test set (left). In the prediction images gray lines mark the edge of the GPM swath where the ground truth data exists.	23
4.2	Distribution of true and predicted precipitation rates for whole test data set. Dotted lines show the 95th quantile of the QRNN models predictions, while solid lines mark posterior means. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.	24
4.3	Difference between true and predicted precipitation rates for whole test data set. Right subplot shows a close up of the range -10.1 mm/h to 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.	25
4.4	2D-histogram showing frequency of predicted and true precipitation rates. Correct predictions fall on the dashed line. Bin sizes scales logarithmically along both axes. White bins are empty.	26
4.5	2D-histogram showing column-wise normalized frequency of predicted and true precipitation rates. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin sizes scales logarithmically along both axes. White bins are empty.	26
4.6	False Positives. The distribution of predictions above the threshold 10^{-1} mm/h that correspond to values below the threshold in the reference data. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.	27
4.7	False Negatives. The distribution of reference values above the threshold 10^{-1} mm/h that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.	28
4.8	Curve obtained by varying the precipitation threshold for classifying predictions as <i>precipitation/no precipitation</i> . Marks at the threshold 0.1 mm/h.	28
4.9	Distribution of true and predicted precipitation rates for the rain gauge dataset. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.	30
4.10	Difference between true and predicted precipitation rates for the corrected HE and the CNN on the rain gauge dataset. Right subplot shows a close up of the range -10.1 mm to 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.	30

4.11	2D-histogram showing frequency of predicted (by the corrected HE and the CNN) and true precipitation rates for the rain gauge dataset. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.	31
4.12	2D-histogram showing column-wise normalized frequency of predicted (by the corrected HE and the CNN) and true precipitation rates for the rain gauge dataset. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.	31
4.13	False Positives. The distribution of predictions above the threshold 10^{-1} mm that correspond to values below the threshold in the rain gauge data. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.	32
4.14	False Negatives. The distribution of rain gauge values above the threshold 10^{-1} mm that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.	33
4.15	Curve obtained by varying the precipitation threshold for classifying predictions as <i>precipitation/no precipitation</i> . Marks at the threshold 0.1 mm.	33
4.16	For each hour (local time) the mean precipitation over all available locations and dates in the rain gauge dataset and corresponding corrected HE and CNN predictions. Filled region represent plus minus one standard deviation of the locations mean over the dates.	34
4.17	Hourly observed and predicted precipitation in Xerém, Rio de Janeiro in the period Dec 22nd to 24th, 2020. Filled regions show hourly means of the CNN's estimated 95th and 99th quantile.	35
A.1	Missing values (represented in black) in the rain gauge dataset and the corresponding HE estimates.	I
A.2	Precipitation distribution for the GPM <i>validation</i> data, bin width 0.1 mm. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 10.1 mm/h.	II
A.3	Precipitation distribution for the GPM <i>test</i> data, bin width 0.1 mm. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 10.1 mm/h.	II
A.4	Difference between true and predicted precipitation rates for the HE and the QRNN models on the rain gauge dataset. Right subplot shows a close up of the range -10.1 mm to 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.	III
A.5	2D-histogram showing frequency of predicted (by the HE and the MLP) and true precipitation rates for the rain gauge dataset. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.	III

A.6 2D-histogram showing column-wise normalized frequency of predicted (by the HE and the MLP) and true precipitation rates for the rain gauge dataset. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty. IV

A.7 False Positives. The distribution of predictions above the threshold 10^{-1} mm that correspond to values below the threshold in the rain gauge data. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis. . . . IV

A.8 False Negatives. The distribution of rain gauge values above the threshold 10^{-1} mm that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis. V

A.9 Graph displays for each hour (local time) the mean precipitation over all available locations and dates in the rain gauge dataset and corresponding HE and MLP predictions. Filled region represent plus minus one standard deviation of the locations mean over the dates. . . . V

List of Tables

2.1	Time period, number of samples and number of available label pixels for each split data subset.	10
2.2	Mean and standard deviation for all GOES channels calculated on the pixels in the training set with corresponding GPM values.	10
3.1	Architecture of MLP.	17
3.2	Confusion matrix of our <i>precipitation/no precipitation</i> categories.	20
3.3	Hyperparameter setup for the MLP and the CNN models.	22
4.1	Bias, MAE and MSE calculated for the QRNN models posterior mean on either the whole test set, or on ranges corresponding to values in the reference data. The intervals are specified by the column "Ranges". The column "Fraction" displays the ratio of true values in the range and the total amount of true values. Bold font marks the best value in each row.	24
4.2	Loss and Continuous Ranked Probability Score (CRPS) mean and median over over all predicted quantiles calculated for the QRNN models on the test set. Bold font marks the best value in each row.	25
4.3	Bias, MAE and MSE calculated for the QRNN and HE models on the rain gauge dataset. Bold font marks the best value in each row.	29

Acronyms

2BCMB GPM DPR and GMI Combined Precipitation L2B 1.5 hours 5 km V06 product.

ABI Advanced Baseline Imager.

ANN Artificial Neural Network.

CNN Convolutional Neural Network.

CPTEC Centro de Previsão de Tempo e Estudos Climáticos (*eng.* Center for Weather Forecast and Climate Studies).

DPR Dual-frequency Precipitation Radar.

GMI Microwave Imager.

GOES Geostationary Operational Environmental Satellites.

GPM Global Precipitation Measurement.

HE Hydro-Estimator.

HM Histogram Matching.

INPE Instituto Nacional de Pesquisas Espaciais (*eng.* National Institute for Space Research).

IR infrared.

MLP Multilayer perceptron.

NOAA National Oceanic and Atmospheric Administration.

PERSIANN Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks.

QRNN Quantile Regression Neural Network.

TRMM Tropical Rainfall Measuring Mission.

Glossary

diurnal cycle Daily recurring pattern (here in precipitation).

nadir point on Earth directly beneath a satellite, the opposite of zenith (NASA 2021a).

nowcasting a description of current weather and a short-period forecast.

precipitation are particles formed by the condensation of water vapor in the atmosphere that falls under the force of gravity. Includes both liquid (rain) and solid (snow and ice) forms (Remote Sensing Systems 2020).

rain gauge any object that collects rain-water that can be measured (Kidd et al. 2017).

remote sensing is the science of obtaining information about objects or areas from a distance, typically from aircraft or satellites (NOAA 2021).

1

Introduction

Precipitation plays a crucial role in the global hydrological cycle, which is closely tied to Earth's climate system. With global warming, the hydrological cycle is expected to intensify, likely bringing increased extreme precipitation and risk of floods (Tabari 2020). It is of great importance for understanding our weather and climate that accurate and high resolution monitoring of precipitation is achieved on a global scale. In particular, a good temporal resolution is needed for nowcasting applications as well as studying the diurnal cycle.

The traditional way of measuring precipitation is by catching it in rain gauges, which provide point measurements at ground level locations. This is considered the most accurate method, but it suffers from low spatial coverage in a global perspective. Kidd et al. (2017) estimate that the broadest reaching available dataset of rain gauges represent less than 1% of the Earth's surface, and that the gauges themselves would cover less than half a soccer pitch if they were gathered up in the same place.

An alternative ground-based measuring technique is weather radars, that sample larger volumes than the rain gauges. But this method comes with other issues, such as blind spots caused by limitations in monitoring at higher altitudes and lower accuracy since the measurements are based on the back-scattered echo of the radar signal (Tapiador et al. 2012). The problem of sparsity in remote areas and over the ocean is also present for this method, with the addition of error sources such as beam blockage (Kidd et al. 2017).

The third option is satellite remote sensing, where instruments that can measure electromagnetic radiation are put in Earth orbit. The sensors can either be *passive*, detecting natural energy reflected by or emitted from the object observed, or *active*, sending out radiation and measuring the reflected or back-scattered signal returning from the target (NASA 2021b). Depending on its frequency, the radiation interacts with different particles in the atmosphere. Microwave radiation interacts only with the largest particles which are precipitation, while radiation in the visible and IR range interacts with the smaller cloud droplets, causing those signals to saturate further up in the clouds. Basically, this means microwave radiation can "look" into clouds, while the visible and IR radiation is constricted to capture cloud top brightness temperatures. Geostationary satellites fly in high orbit and provide good spatial and temporal coverage, but can only carry visible and IR sensors due to antenna size constraints. We can have active and microwave sensors with reasonable resolution on low-flying satellites, however this limits their spatial and temporal coverage.

The question is thus: Is it possible to obtain high quality precipitation retrievals from geostationary observations? Previous efforts have been made, building upon the general idea that cold cloud tops imply large vertical cloud development, which in turn would indicate more precipitation.

While all models require some ground-based or low orbit observations as a reference, the execution differs between them. To generalize, there is on one hand the more traditional methods that use regression and threshold rules to capture this relation. To mention a few well known, there is the GOES Precipitation Index (Arkin and Meisner 1987), the Convective-Stratiform Technique (Adler and Negri 1988), the Autoestimator (Vicente et al. 1998) and the GOES multispectral rainfall algorithm (Ba and Gruber 2001). On the other hand there is the machine learning approach, and especially Artificial Neural Networks (ANNs), that have been on the rise in recent years due to their ability to capture non-linear relations in large data. In this work we have studied the use of the latter, and this will be further introduced in section 3.3.

Commonly in previous works, the models consider the retrieval problem as the task of estimating a scalar quantity from the given observations, disregarding the underlying uncertainties in the prediction. In this thesis, we ask instead: Is it possible to learn a mapping from the observations to a *probability distribution* over possible precipitation rates to account for those uncertainties?

In order to do this, we study the use of Quantile Regression Neural Network, suggested for application to the general retrieval problem by Pfreundschuh et al. (2018). This thesis was conducted in collaboration with the National Institute for Space Research (INPE) in Brazil who have provided us with supplementary data and guidance for studying the geographic region of Brazil. To build a dataset that covers this region, we collect precipitation rate measurements from the GPM Core Observatory, and IR input data from the GOES-16 geostationary satellite.

2

Data

The data collection and processing played a large role in this thesis, and this chapter gives an introduction to the sources used (section 2.1) and a step-wise description of the dataset construction (section 2.2).

2.1 Sources

This section presents an overview of the raw data: The input in section 2.1.1, the ground truth in section 2.1.2 and an independent validation dataset consisting of rain gauge measurements in section 2.1.3. The Hydro-Estimator (HE) is also included in section 2.1.4 of this chapter since it was received as a precipitation product ready to use from INPE.

2.1.1 GOES-16: Input data

GOES-16 is a weather satellite in geostationary orbit, that delivers high spatial and temporal resolution measurements over the Western Hemisphere¹. Figure 2.1 shows the coverage of the GOES-16 satellite and an example image.

The main instrument on-board of the GOES-16 is the Advanced Baseline Imager (ABI). It is a passive imaging radiometer with sixteen channels in the visible to IR range designed to observe Earth's surface, atmosphere and cloud cover by measuring emitted and reflected radiation. The ABI can not measure precipitation directly, but the high resolution and good coverage motivates making the effort of trying to derive a prediction from its measurements. In this work, we limit ourselves to only using IR bands to exclude the day and night differences in the visible radiation².

2.1.2 GPM: Ground truth data

As ground truth precipitation rate we will use the 2BCMB product (Olson 2017), which is derived by combining measurements from the Microwave Imager (GMI) and the Dual-frequency Precipitation Radar (DPR). These sensors are found on-board of the GPM Core Observatory, which is the main satellite in the Global Precipitation

¹Temporal resolution of 5-15 minutes and spatial resolutions of 0.5, 1.0 or 2.0 km at nadir (NOAA/NASA 2021).

²Also, channel 12 is discarded due to its main focus on ozone. Thus, the channels to be used for this thesis are 8-11 and 13-16.

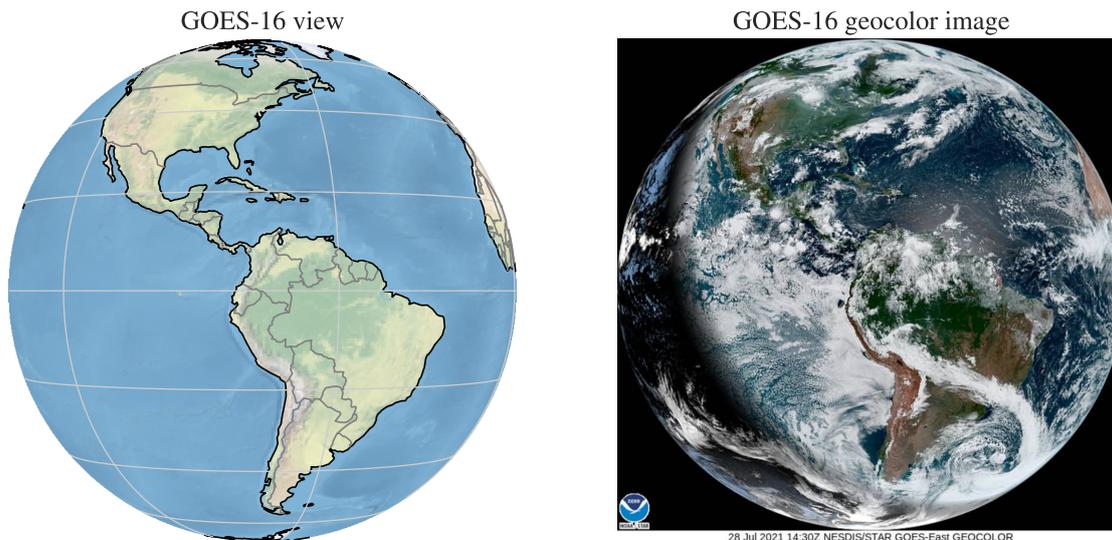


Figure 2.1: Left: GOES-16 satellite geostationary view of Earth. Right: Example of geocolor image captured by the ABI instrument on-board of GOES-16 (image source CIRA/NOAA, *GOES Image Viewer* (2021)).

Measurement (GPM) constellation. The combined precipitation product is developed to provide highly accurate measurements to be set as a reference standard for other constellation sensors and hence improve measurements globally. This quality makes it suitable as ground truth also for this task. Validation against the Global Precipitation Climatology Project (GPCP) shows good agreement in the tropics and midlatitudes (Grecu et al. 2016).

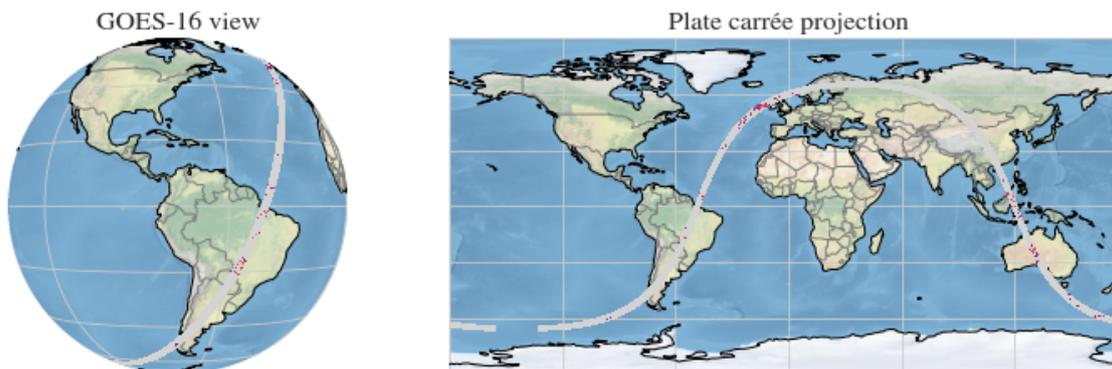


Figure 2.2: Earth in GOES-16 geostationary perspective (left) and in plate carrée projection (right), both overlaid with one 2BCMB data swath.

The Core Observatory flies at a 65 degrees inclination to the equator in a non-sun-synchronous circular orbit which takes about 1.5 hours to complete. As the Earth rotates, the GPM Core Observatory orbit is shifted and the surface is successively covered by its swath. Figure 2.2 shows an example of a swath corresponding to one 2BCMB orbit granulate.

2.1.3 Rain gauges: Reference data

A dataset of hourly rain gauge measurements from December 2020 provided by INPE is to be used as an independent reference for evaluation. As opposed to the training data, the rain gauges give a more direct measure in the sense that it captures rain that certainly has fallen rather than measuring it from a distance which entails retrieval uncertainties. Thus, comparing our estimates to gauge data is a way of checking how well the predictions correspond to 'real' rain.

The dataset consists of 609 rain gauges specified by state, station ID, altitude, longitude and latitude, each with a corresponding series of timestamps (date, hour) with precipitation measurements (mm) aggregated over the previous hour. This is in contrast to the GPM measurements which are instantaneous. Figure 2.3 shows the spatial distribution of the gauges.

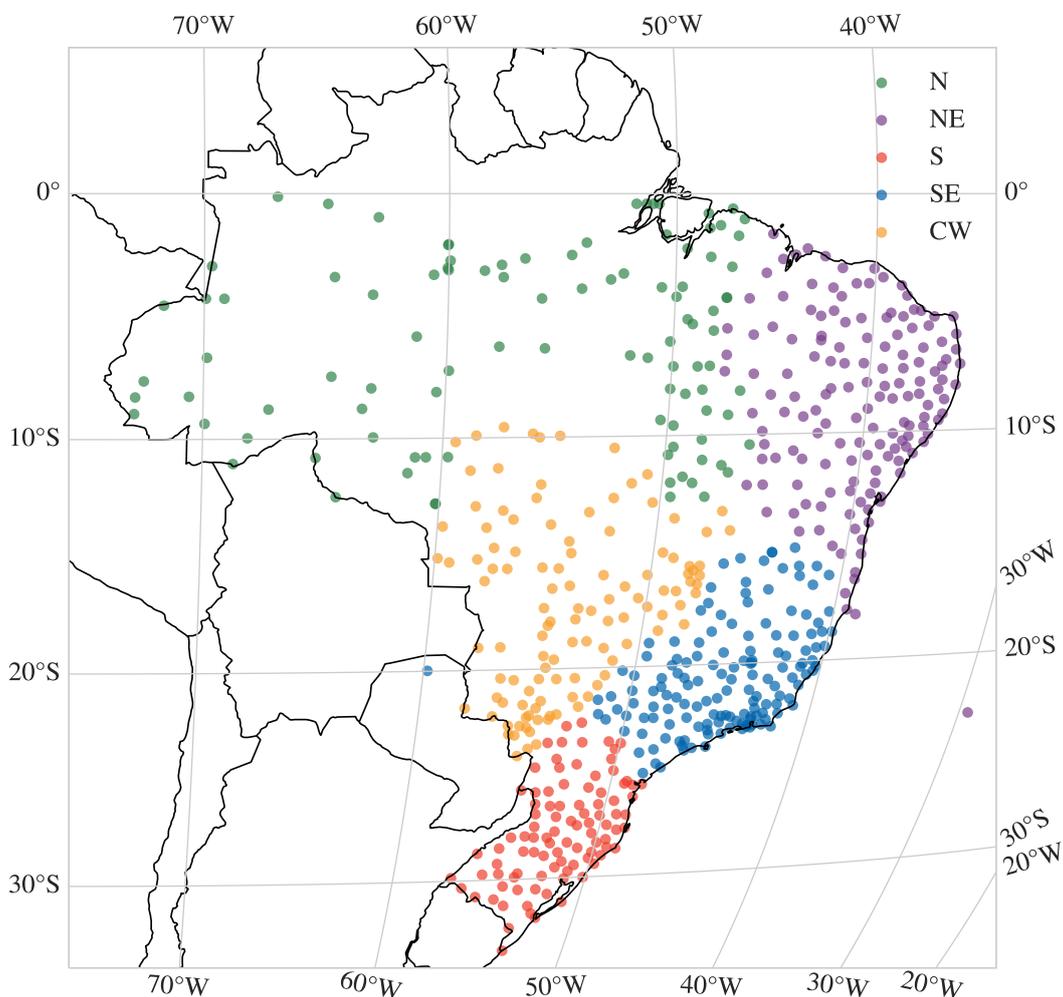


Figure 2.3: Map of rain gauge positions from the dataset over Brazil. Color legend corresponding to five regions.

An issue with comparing the gauge data with the models that are to be trained on the satellite data is the difference in resolution. Where the gauges measure pre-

precipitation on a very local scale of a decimetre (Kidd et al. 2017), the models give estimates for a few kilometre. This might have the effect that precipitation systems on a scale smaller than the prediction resolution appear smeared in the estimates that present a lower precipitation rate on the whole. In addition to this we have the difference in temporal resolution as mentioned above.

Note also that missing values occurs quite frequently in the rain gauge dataset, meaning that not all gauges have available data at all timestamps. A visualization of this can be found in figure A.1 in appendix.

2.1.4 HE: Baseline model

The Hydro-Estimator (HE) (Scofield and Kuligowski 2003) is an algorithm that produces precipitation estimates based on a single IR channel input from the GOES satellite. The HE is a later version of the Auto-Estimator developed by Vicente et al. (1998) at NOAA. The algorithm (NOAA 2012) was later adapted to South America conditions and used at the CPTEC/INPE in Brazil. The core of the model is a power-law expression that relates the measured brightness temperatures to precipitation rates, and the estimates are returned in a 4 km regular grid.

A technique called Histogram Matching (HM) was applied by Siqueira and Vila (2019) to adjust the HE estimates to better match the observed precipitation rates from the radar instruments onboard of GPM and the Tropical Rainfall Measuring Mission (TRMM), a precursor of the GPM. Different adjustment lookup tables are available based on season and region. As for the rain gauges the estimates will be taken for December 2020, and thus the adjustments for October to December. Also, the regional adjustments (north, northeast, south, southeast, centre-west) will be applied for the corresponding gauges.

2.2 Processing

This section deals with the accessing, pre-processing and matching of the raw data that is performed in order to create the main dataset. If C , H , and W represents number of channels, height and width, then the end result of this section is a set of input tensors of dimension $C \times H \times W$ containing the GOES channels data that corresponds to a tensor of dimension $H \times W$ with the GPM precipitation rate at the same time and location.

2.2.1 Region

The aim is to retrieve precipitation over Brazil, thus a region enclosing the country is defined that will act as an outer limit for data collection. For convenience, the region is chosen as a square in the view of the GOES-16 satellite, since this is the grid coordinate system of the input. Figure 2.4 shows the region in the GOES-16 satellite view (left) and on a plate carrée projection map (right).

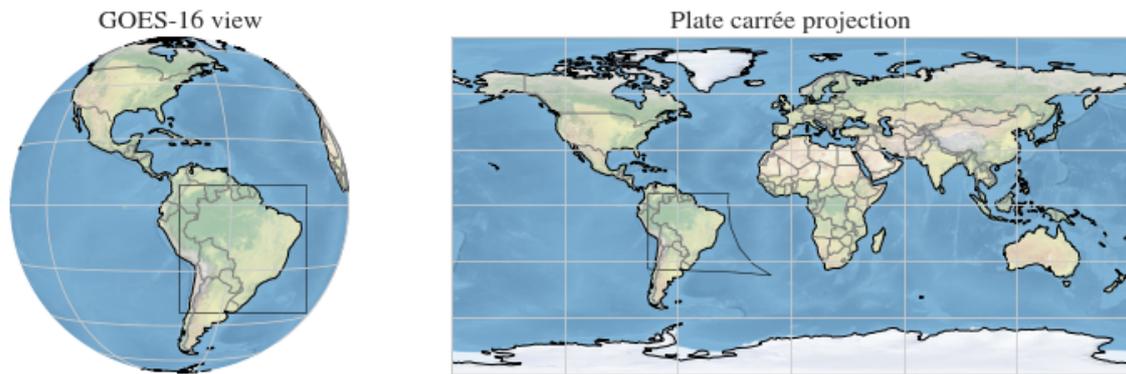


Figure 2.4: Visualization of the region of interest in two different coordinate frames, the GOES-16 satellite view of Earth (left) and in the plate carrée projection (right).

2.2.2 Time period

Since this dataset will be constructed based on data from two satellites, it is necessary for both of them to be in orbit for the whole time period considered. Because the GOES-16 was the last of the two to become operational, on December 18, 2017, this will be used as start time. Both satellites are still active, and are thus constantly capturing new data which makes it possible to extend the dataset later on, but for the purpose of this thesis the period end date was set to March 31, 2021.

2.2.3 Selection

The GPM data files are collected from an online archive. In order to only download the relevant data, i.e. the granulates that actually contain data on Brazil, an online tool called Earth Data Search (*Earthdata Search* n.d.) is used. With this tool one can filter out granulates that have no part of its swath inside a specified region. The remaining granulates extracted intersect the polygon shown in figure 2.5).

2.2.4 Resampling

Due to the different coordinate systems of the GPM granulates and the region in the GOES system, the former is to be resampled. This is done by the nearest neighbour method. To account for the discrepancy between the spatial resolution of the GOES data (2 km at nadir) and the GPM data (5 km × 5 km), the GOES data is aggregated to have a resolution of 4 km at nadir before the resampling of the GPM data. The region (see section 2.2.1) consists of 1024 × 1024 pixels in this grid.

2.2.5 Cropping

Once a subset of granulates is selected, it is to be further slimmed down by removing the granulate parts outside of the region presented in section 2.2.1. Now that the data is matched into the same coordinate system it is straightforward to crop to the region. But, in order to reduce file sizes and possibly improve temporal matching



Figure 2.5: Example of granulates intersecting a polygon that encloses Brazil in the Earth Data Search tool (*Earthdata Search* n.d.).

in some cases, each granulate is split into several smaller chunks. All operations are performed in the GOES-16 view coordinate system.

The splitting is done as follows:

- i) Start from the center of the region and add a uniformly sampled vertical offset, call this point y_{mid} .
- ii) Place out three squares of 256×256 pixels vertically centered about the point y_{mid} and with their horizontal centers on the swath.
- iii) Save all squares that lie entirely within the region.

Step i) is included to reduce the impact of edge effects on a certain subregion. The offset ensures that the squares are not always placed in the same vertical position. Figure 2.6 shows an example of a resampled and split up granulate.

2.2.6 Temporal matching

For each GPM granulate chunk a corresponding GOES chunk should be matched. Since the data is now on the same grid, it is the temporal matching that needs considering. Each GPM scan is annotated with a timestamp, hence the time period for each split of the overpasses are known. Given such a start and end time, a GOES-16 full disk image from this time span can be searched for in another online archive. To ensure that the overlap is sufficient, a tolerance is applied to the difference between the mid timestamps.

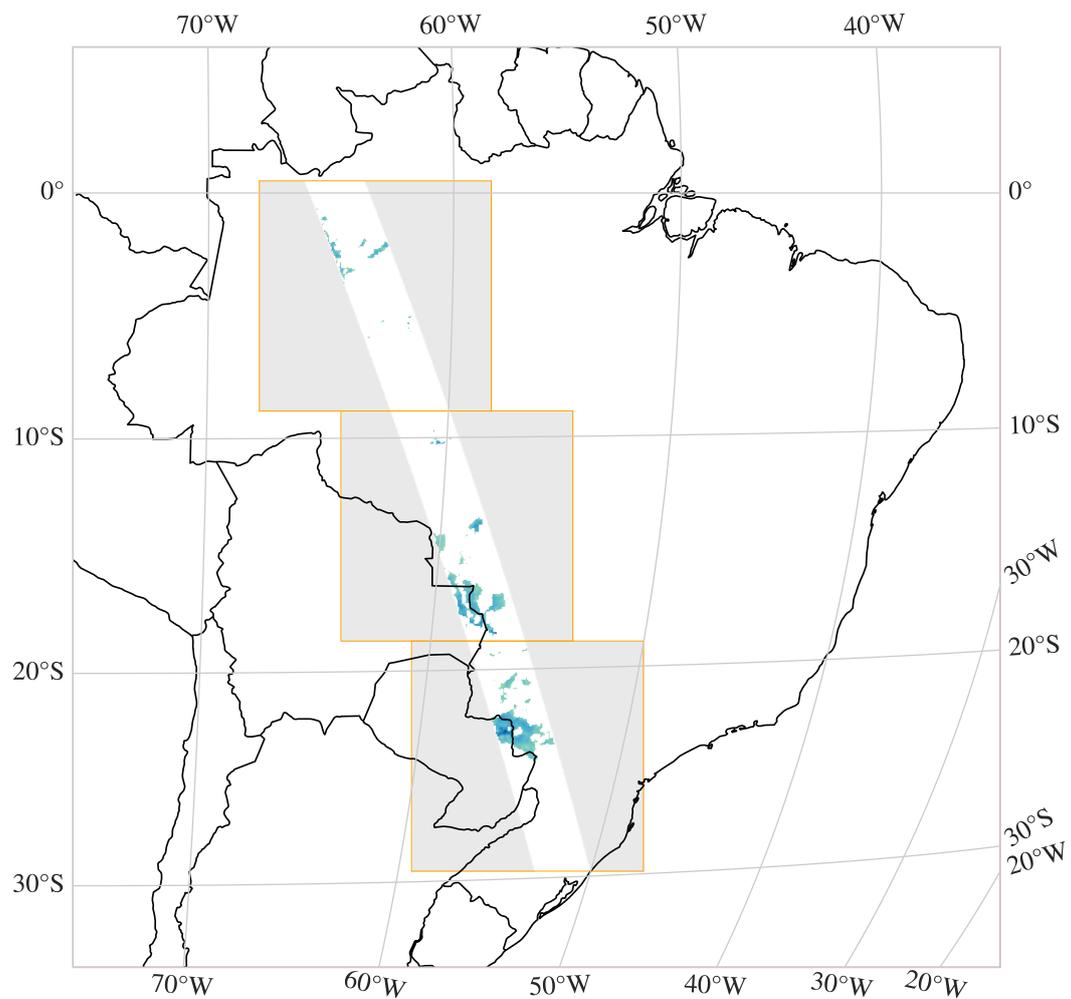


Figure 2.6: Example of a GPM granulate that is coordinate transformed and cut into three 256×256 pixel pieces. The shaded gray in the squares marks the area that lacks GPM data.

2.2.7 Split

The dataset is split up into subsets to be used for training, validation and testing respectively. The test set is separated from the train and validation sets by time period, i.e. the last year of samples is put into the test bin to ensure independence. A standard random 80/20 split is done to separate the training data from the validation data. Table 2.1 shows the compiled split information.

Table 2.1: Time period, number of samples and number of available label pixels for each split data subset.

Split	Period start	Period end	Samples	Label pixels
train	2017-12	2020-03	5 412	83 360 758
validation	2017-12	2020-03	1 354	20 805 499
test	2020-04	2021-03	2 928	44 869 385

2.2.8 Normalization

All channels are standardized based on the training data channel mean and standard deviation, shown in table 2.2.

Table 2.2: Mean and standard deviation for all GOES channels calculated on the pixels in the training set with corresponding GPM values.

Channel	8	9	10	11	13	14	15	16
mean	237.2	245.5	253.1	274.4	276.3	275.1	272.5	261.8
std	9.388	11.61	13.48	21.13	22.15	22.57	22.10	17.55

2.3 Dataset properties

The aim of this section is to give an overview of some data characteristics.

2.3.1 Sample

Figure 2.7 shows a sample from the (main) dataset, consisting of an input tensor of dimension $8 \times 256 \times 256$ which corresponds to a label of dimension 256×256 shown in figure 2.8, where the pixels outside the GPM swath is masked out (shown in gray).

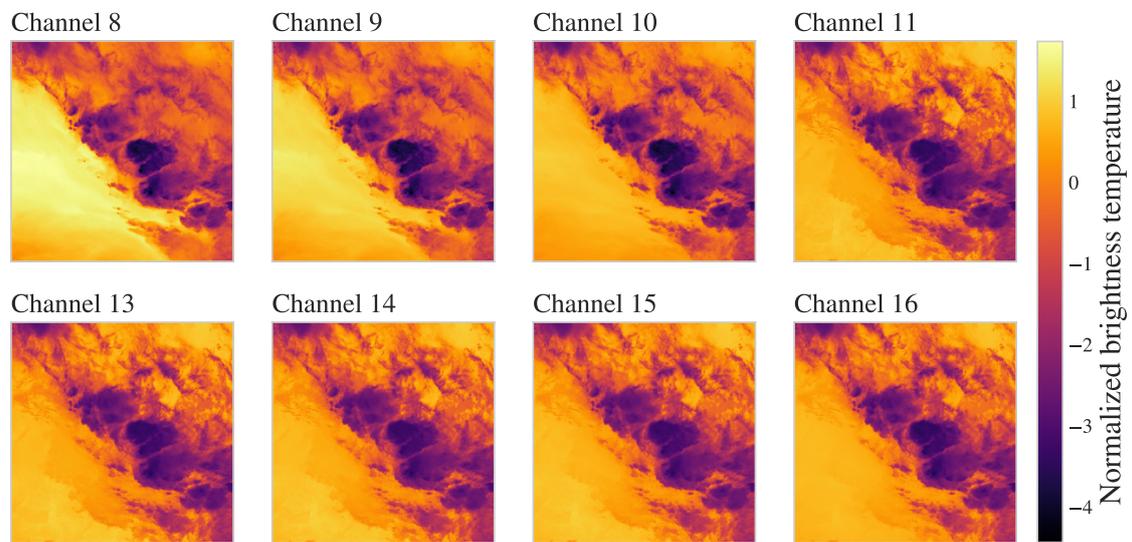


Figure 2.7: Example of an input tensor of standardized GOES channels.

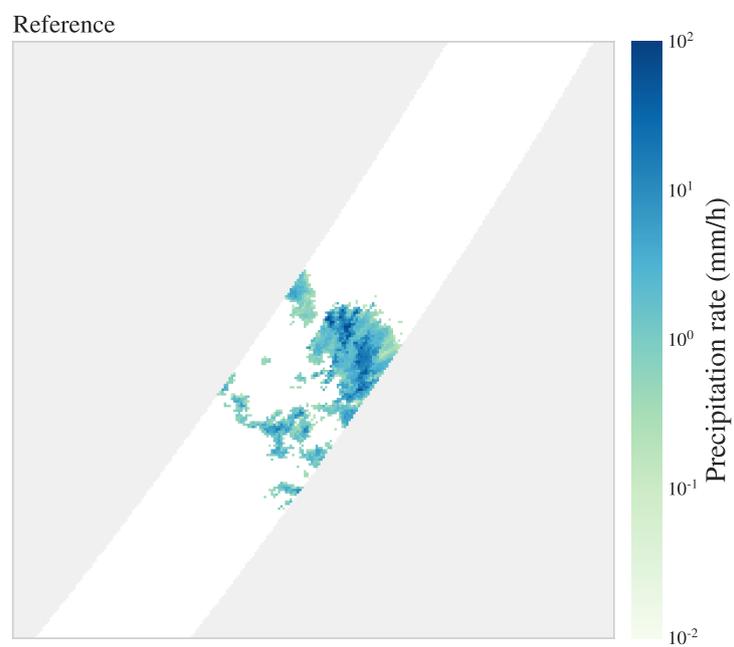


Figure 2.8: Example of a GPM label, with precipitation rates (mm/h) and masked pixels marked in gray.

2.3.2 Precipitation rate distribution

The intensity of precipitation is highly skewed towards light precipitation, as seen in figure 2.9 for the GPM *training* data and the rain gauge data. For the GPM *validation* and *test* distributions, see A.2 and A.3 in appendix A. The rain gauge data is discrete, in the sense that the precipitation intensities can take the values of 0.0, 0.2, 0.4, 0.6... and so on with multiples of 0.2. Compare this to the more continuous GPM distribution. Studying the left subplot in figure 2.9, we notice the apparent shortage of observations just above zero precipitation in the GPM data. This is a result of the nominal sensitivity threshold for the DPR and GMI sensors of around 0.2 mm/h (Skofronick-Jackson et al. 2017). The percentage of non-zero precipitation is 6% in the GPM datasets, and 8% in the rain gauge dataset.

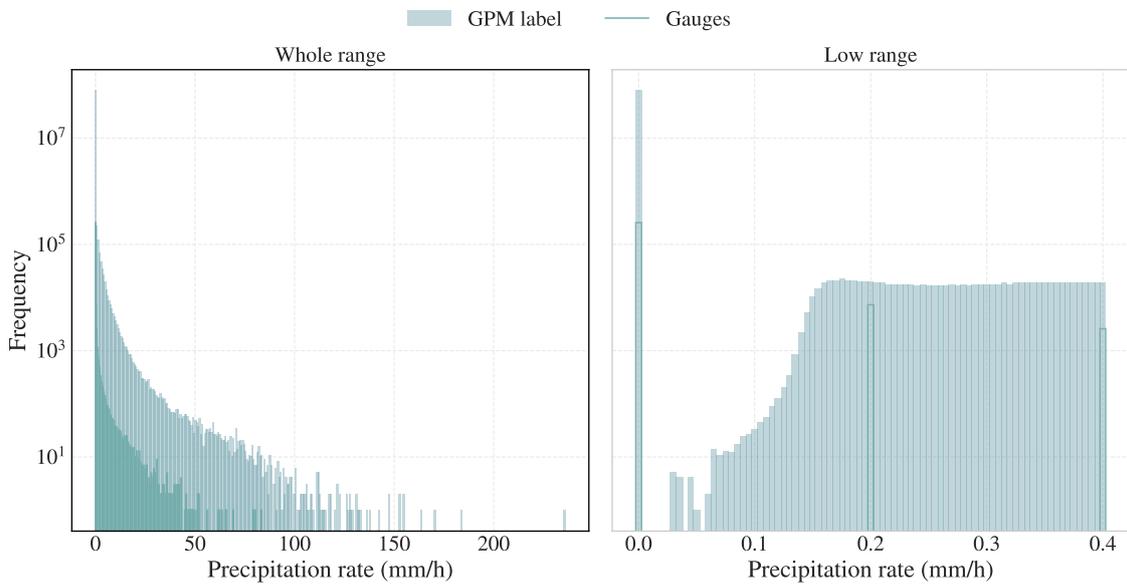


Figure 2.9: Precipitation distribution for the GPM *training* data overlaid by the rain gauge data, bin width 0.1 mm/h. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 0.4 mm/h with bin width 0.005 mm/h.

3

Modelling

With the dataset ready, it is time to consider the process of fitting a model to the data. This chapter includes a formulation of the problem, methods used and evaluation metrics to be applied.

3.1 The retrieval problem

The problem of retrieving precipitation or other atmospheric quantities by remote sensing can be described as the inverse problem of determining the retrieval quantity (e.g. the precipitation rate) x from the actual measurements (e.g. the brightness temperatures) \mathbf{y} . Due to measurement errors, approximations and limited sensitivity at the instruments, this problem is generally not uniquely solvable (Pfreunds Schuh et al. 2018).

3.1.1 Bayesian approach

Applying a Bayesian approach to account for the uncertainties in the problem has been suggested by Rodgers (2000) and Pfreunds Schuh et al. (2018). The solution to the inverse problem is then given by the a posteriori distribution $p(x|\mathbf{y})$ which can be obtained by Bayes' theorem

$$p(x|\mathbf{y}) = \frac{p(\mathbf{y}|x)p(x)}{\int p(x', \mathbf{y}) dx'}. \quad (3.1)$$

Methods that use expression (3.1) to compute a solution to the retrieval problem are referred to as Bayesian retrieval methods. Pfreunds Schuh et al. (2018) show that Quantile Regression Neural Network (QRNN) can be applied to estimate the posterior distribution $p(x|\mathbf{y})$ of Bayesian remote sensing retrievals, and that they are capable of providing probabilistic predictions on par with such Bayesian methods. The authors conclude that QRNN may even handle the curse of dimensionality better than traditional (non-analytical) methods which would make them more suitable for the application to high-dimensional retrieval problems. The QRNN will be formulated in the following section.

3.2 Supervised machine learning

The retrieval problem, disregarding the uncertainties, can be viewed as a simple multiple regression task that is well suited for supervised machine learning. With a

set $\{(\mathbf{y}, x)\}$ consisting of pairs of observations \mathbf{y} and retrieval quantities x a model f can be trained to map the input values \mathbf{y} to the expected output x . The model is trained through adjustment of its parameters to minimize the mean of a certain loss function $\mathcal{L}(f(\mathbf{y}), x)$ on the training dataset. Commonly, the squared error loss is used, but as will be seen in section 3.2.1 another loss function will be applied here.

3.2.1 Quantile Regression

Let $F(x)$ be the cumulative distribution function of a probability distribution $p(x)$. Then the τ th quantile x_τ of F is given by

$$x_\tau = \inf\{x : F(x) \geq \tau\}, \quad (3.2)$$

or equivalently: x_τ is the smallest value of x among all the values for which the cumulative distribution function exceeds τ . By Koenker (2005), the τ th quantile minimizes the expectation

$$\mathbb{E}(\mathcal{L}_\tau(x_\tau, x)) = \int_{-\infty}^{\infty} \mathcal{L}_\tau(x_\tau, x') p(x') dx' \quad (3.3)$$

of the function

$$\mathcal{L}_\tau(x_\tau, x) = \begin{cases} \tau|x - x_\tau|, & x_\tau < x \\ (1 - \tau)|x - x_\tau|, & x_\tau \geq x. \end{cases} \quad (3.4)$$

If the machine learning model f is trained to minimize the mean of the quantile loss function $\mathcal{L}_\tau(f(\mathbf{y}), x)$ over the training set $\{\mathbf{y}_i, x_i\}_{i=1}^n$, then it learns to predict the quantiles of $p(x|\mathbf{y})$. If the model is trained to estimate multiple quantiles, then an approximation of the cumulative distribution function $F_{x|\mathbf{y}}(x)$ can be obtained (Pfreundschuh et al. 2018).

3.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a class of flexible nonlinear models that are well-suited for representing complex relationships in data. They have been applied to precipitation retrieval problems for quite some time, with an early example being the PERSIANN algorithm that used IR imagery from geostationary satellites together with ground-based data to estimate rainfall rates over Japan and Florida in the late 1990s (Hsu et al. 1997). Numerous adaptations and later versions of the PERSIANN algorithm can be found, including PERSIANN-CNN that applies Convolutional Neural Networks (CNNs) to retrieve precipitation over the continental U.S. (Sadeghi et al. 2019).

In this section we will introduce some building blocks of ANNs, and thereafter the specific architectures of the networks trained for this thesis. As one objective of this thesis is to compare a model that makes use of spatial information to one that deals with each pixel individually, there will be two different types of models presented.

3.3.1 Fully Connected Layer

ANNs consists of layers of nodes, referred to as *neurons*, arranged in a network structure. Let $\mathbf{x} \in \mathbb{R}^n$ represent the input to a fully connected layer, and $\mathbf{y} \in \mathbb{R}^m$ the output. Then the j th element y_j of the output is connected to every input x_i by trainable weights w_{ij} , as illustrated in figure 3.1. Mathematically, this can be computed as in

$$y_j = \varphi \left(\sum_{i=1}^n w_{ij} x_i + b_j \right), \quad (3.5)$$

with the associated weights w_{ij} , bias b_j and non-linear activation function φ . A Multilayer perceptron (MLP) has at least three such layers, where the intermediate layers are called *hidden layers*.

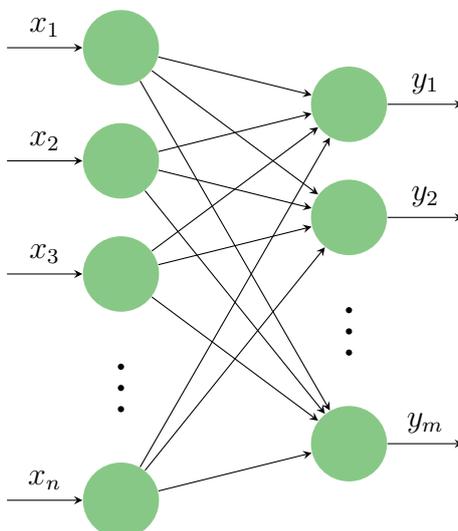


Figure 3.1: Fully Connected Layer with input $\mathbf{x} \in \mathbb{R}^n$ and output $\mathbf{y} \in \mathbb{R}^m$. Each arrow is associated with a weight w_{ij} .

3.3.2 Convolutional Layer

A layer type that is specialised in capturing the spatial structure of the input is the convolutional layer, which is based on the mathematical operation with the same name (LeCun et al. 1998). Given two functions $f, g : \mathbb{Z}^2 \rightarrow \mathbb{R}$, their convolution (in a 2-dimensional setting) at a point $(k_1, k_2) \in \mathbb{Z}^2$ is defined as

$$(f * g)[k_1, k_2] = \sum_{l_1 \in \mathbb{Z}} \sum_{l_2 \in \mathbb{Z}} f[k_1, k_2] g[k_1 - l_1, k_2 - l_2]. \quad (3.6)$$

This operation is illustrated in figure 3.2 where an input is convolved by a kernel which results in a output feature map. In the case of a multiple channel input a

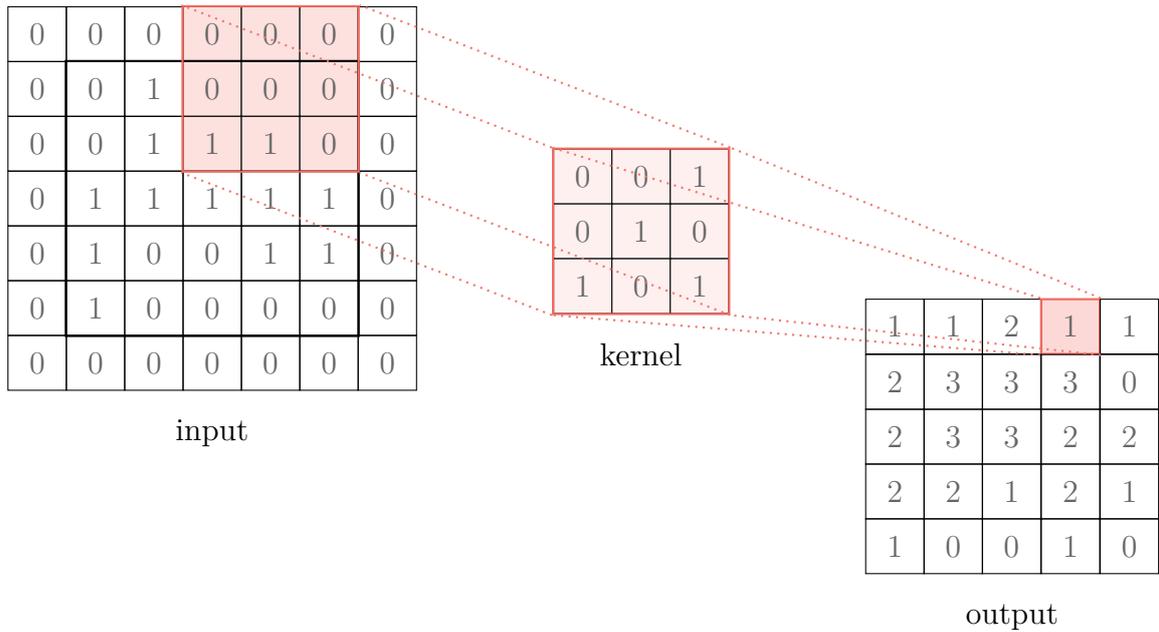


Figure 3.2: Example of a convolutional filter in the case of a single channel image. The 3×3 -kernel slides through the 7×7 input, performing the convolution at each position, which yields the output.

filter consisting of a stack of kernels is applied. Each channel is convolved with a kernel, and the outputs are summed-up element-wise.

A convolutional layer is often followed by a *pooling layer* with the purpose of down-sampling. 2D-Max-pooling, which is applied in this thesis, is a filter that outputs the maximum of each input subregion. A convolutional network can be viewed as a fully connected network with a priori constraints encoded into its structure. This prior makes sure that the weights for each hidden unit is equal to the weights of its neighbor shifted in space. This makes convolutional neural networks learn local correlations and helps it in becoming invariant to translation. The prior also ensures that the weights are equal to zero everywhere except for a small spatial neighborhood to that hidden unit. Further, the pooling operation can be regarded as a prior that makes each unit invariant to small translations (Goodfellow et al. 2016). This makes convolutional networks suitable for many different computer vision tasks, where spatial information is important.

In addition, we apply batch normalization (Ioffe and Szegedy 2015) in order to stabilize and speed up the training. Each input element x_i in a mini-batch \mathcal{B} is re-scaled and re-centered according to equation

$$\text{BN}_{\gamma, \beta}(x_i) = \gamma \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} + \beta \quad (3.7)$$

where $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}$ is the mini-batch empirical mean and standard deviation respec-

tively and where γ and β are learnable parameters. ϵ is added for numerical stability.

3.4 Architectures

As stated above, this thesis is intended to compare an independent pixel based model to a spatially aware model. For this purpose, we will introduce two different architectures in this section. Common for both of them is the number of input channels C (or features in the case of independent pixels) which is eight due to the dimension of the sample described in section 2.3.1. Also, the quantiles to be predicted are set to $\tau = 0.01, 0.02, \dots, 0.99$ for both architectures.

3.4.1 Multilayer perceptron

The MLP to be used as our non-spatial QRNN model has the architecture given in table 3.1. Here the input $\mathbf{x} \in \mathbb{R}^C$ represent a single pixel with eight channel values, and $\mathbf{y} \in \mathbb{R}^{99}$ the estimated quantiles of the posterior distribution for this pixel is the output.

Table 3.1: Architecture of MLP.

Layer	Type	Activation	H_{in}	H_{out}
Input	Linear	ReLU	8	256
Hidden	Linear	ReLU	256	256
Hidden	Linear	ReLU	256	256
Output	Linear	None	256	99

As specified in the table 3.1, the activation function φ chosen is the rectified linear unit function (ReLU), given by

$$\text{ReLU}(x) = \max(0, x). \quad (3.8)$$

3.4.2 XceptionFPN

The *XceptionFPN* model architecture from the `quantnn` library (Pfreundschuh 2021) is based on combining concepts from the *Xception* network proposed by Chollet (2017), and Feature pyramid networks (FPN) proposed by Lin et al. (2017).

The first aforementioned network, *Xception*, is an 'extreme' version of its predecessor the *Inception* (Szegedy et al. 2016) and builds upon a hypothesis of decoupled cross-channel correlations and spatial correlations. The aim is to make the convolution operation more efficient. The *Xception* architecture is entirely based upon the Depth-wise separable convolution consisting of

- i) a spatial convolution performed independently over each channel in the input,
- ii) a 1×1 -convolution projecting the previous output onto a new channel space,

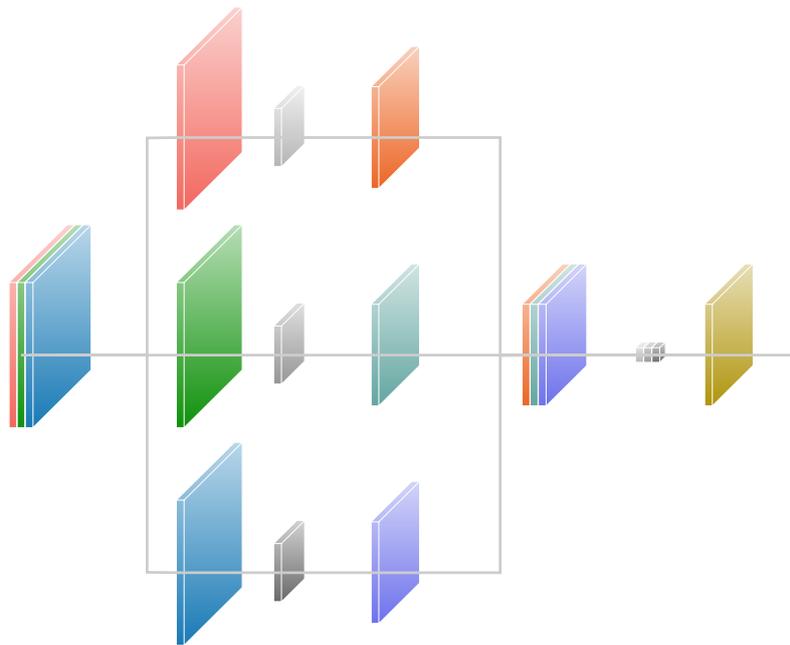


Figure 3.3: Illustration of Depth-wise separable convolution applied to a three channel image. First, an ordinary spatial convolution is applied to each channel separately. Second, a point-wise convolution (kernel size 1×1) is performed on the stack of output feature maps from the first step.

illustrated in figure 3.3.

The second network, the FPN, is a pyramidal structured feature extractor, as indicated by its name. The architecture consists of a bottom-up down-sampling pathway and a top-down up-sampling pathway, with lateral connections at every level. The goal is strong semantics on all scales.

Figure 3.4 shows a diagram of the *XceptionFPNs* general architecture. We recognise the pyramid structure of the FPN to the left, and the Depth-wise separable convolution as the 'SeparableConv' in the network blocks. The output of the *XceptionFPN* is a tensor with the quantiles of the posterior distribution along the channel dimension.

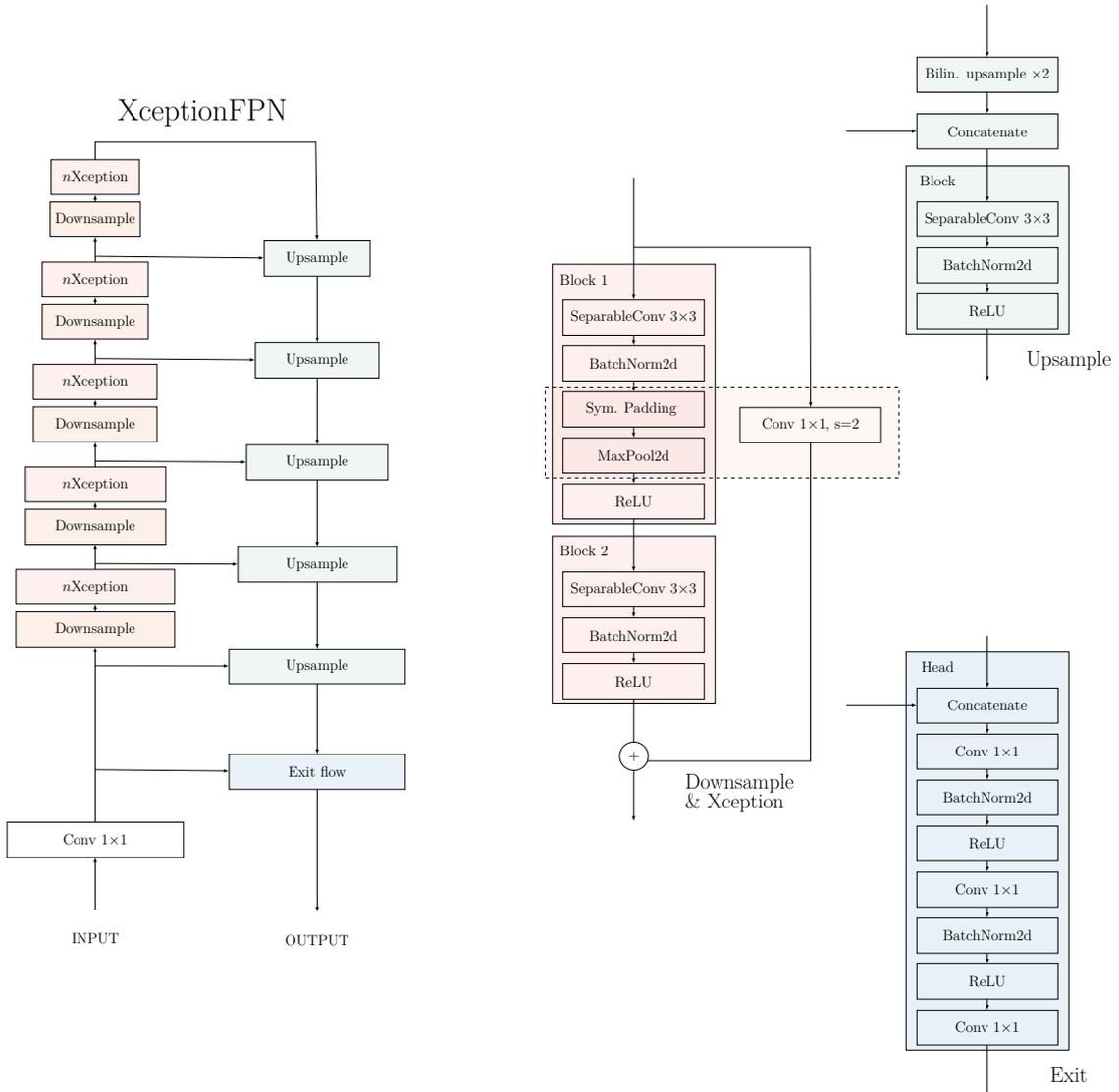


Figure 3.4: Schematic of the *XceptionFPN* model architecture. To the left: the general pyramid structure of the entire flow, and to the right: the Downsample/Xception flow (where the dashed region only occurs in the downsampling), the Upsample flow and the Exit flow respectively. The n in the $nXception$ indicates that the module is repeated at this position.

3.5 Evaluation

How to fairly and meaningfully evaluate the models is a question in itself, that is somewhat dependent upon the intended application. The metrics applied for this thesis are described in this section. Most metrics require a single estimate \hat{x} for each observed sample x . The estimate is then taken to be the posterior mean $E(x|\mathbf{y})$ whenever these metrics are applied.

3.5.1 Regression metrics

Common metrics for evaluating regression models are the bias (3.9), Mean Absolute Error (3.10) and Mean Squared Error (3.11). These will be applied to assess if the predicted posterior means are close to the ground truth.

$$\text{Bias}(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i) \quad (3.9)$$

$$\text{MAE}(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^N |\hat{x}_i - x_i| \quad (3.10)$$

$$\text{MSE}(\hat{x}, x) = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2 \quad (3.11)$$

3.5.2 Categorical metrics

We are interested in how well the models differentiate between no precipitation and precipitation. The regression metrics do not capture this and therefore we complement them with categorical metrics. Table 3.2 shows a confusion matrix

		<i>Predicted</i>	
		Precipitation	No precipitation
<i>Observed</i>	Precipitation	True Positives	False Negatives
	No precipitation	False Positives	True Negatives

Table 3.2: Confusion matrix of our *precipitation/no precipitation* categories.

with the classes *precipitation* and *no precipitation*. With n_k being the number of predictions in each category $k = \text{TP}, \text{TN}, \text{FP}, \text{FN}$, the categorical metrics True Positive Rate (3.12) False Positive Rate (3.13) can be computed.

$$r_{\text{TPR}} = \frac{n_{\text{TP}}}{n_{\text{TP}} + n_{\text{FN}}} \quad (3.12)$$

$$r_{\text{FPR}} = \frac{n_{\text{FP}}}{n_{\text{FP}} + n_{\text{TN}}} \quad (3.13)$$

To divide our continuous data into the two categories *precipitation/no precipitation* we apply a threshold as decision rule, i.e. all precipitation rates below the threshold will belong to the *no precipitation* class and vice versa. The threshold is set to 10^{-1} mm/h if not stated otherwise.

3.5.3 Probabilistic accuracy measures

We have now established the use of standard metrics for evaluating our predictions summarized as a point prediction \hat{x} against the observed sample x . But, since the goal is to estimate the posterior distribution, we would like a means of evaluating the accuracy of predicted conditional distributions. Such summary measures are called scoring rules. Ideally, a probabilistic prediction is sharp (i.e. have small uncertainty) and well calibrated (predicted probabilities match observed frequencies). To formalize this, there are so called *proper* scoring rules. A scoring rule is proper if the highest expected score for a given observation is achieved by reporting its true probability distribution (Gneiting and Raftery 2007).

The quantile loss function (3.4) used to optimize the QRNN is a proper scoring rule for quantile estimation and will be used also for evaluating and comparing the implemented models. Further, we will use the continuous ranked probability score (CRPS) defined by equation (3.14), which is a proper scoring rule for evaluating an estimated cumulative distribution function. The CRPS measures the quadric difference between the cumulative distribution function (F) and the empirical cumulative distribution function (Gneiting and Raftery 2007).

$$\text{CRPS}(F, x) = \int_{-\infty}^{\infty} (F(x') - \mathbb{1}(x \leq x'))^2 dx' \quad (3.14)$$

3.6 Experimental setup

The two models, further on referred to as "the CNN" and "the MLP", described in section 3.4 were trained on the training dataset, and the validation dataset was utilized for model selection (split described in section 2.2.7). As stated above, the loss to be minimized is the mean of equation (3.4), which is calculated on the pixels where the label data is present. This section contains further details of the experimental setup.

3.6.1 Data loading

The data is loaded differently for the CNN and the MLP models. Due to the fact that the MLP only considers single pixels, the inputs of shape $C \times H \times W = 8 \times 256 \times 256$ is converted to arrays of shape $HW \times C$ with corresponding label arrays of length HW . Further, the entries without corresponding GPM labels are masked out, and the remaining arrays are concatenated to a resulting array of shape $P \times C$, where P is the number of label pixels.

For the CNN the inputs are not reshaped, but since the 256×256 sample squares are all by construction centered around the GPM swath, a cropping step is included in the loading of the data to support the network in predicting on all parts of the sample image. Each time a sample is loaded, a random 128×128 pixel square is

returned instead of the whole sample of 256×256 pixel. In this way the network gets to see different parts of the data in different epochs, with the swath located in different parts of the image.

Additionally, when loading the data for both models, all zero-valued precipitation rates were reassigned random values in the range 10^{-4} mm/h to 10^{-3} mm/h.

3.6.2 Hyperparameters

A cosine annealing learning rate schedule was used for training (Loshchilov and Hutter 2016), together with the optimizer *Adam* (Kingma and Ba 2014). The hyperparameters were manually tuned and the most promising models were selected based on the validation dataset metrics computed. The final hyperparameter setup for the models are given in table 3.3. The batch size was set to the largest that could fit in memory.

Table 3.3: Hyperparameter setup for the MLP and the CNN models.

Model	Initial LR	Epochs	Batch size
MLP	0.001	100	32786 entries
CNN	0.01	100	64 images

4

Results

This chapter consists of two main parts, (i) the evaluation of the QRNN models on the unseen reference test set, and (ii) the comparison against the hydroestimator on the separate rain gauge dataset. Posterior mean estimates are considered as predictions, if not otherwise stated.

4.1 Evaluation on test set

The trained and selected MLP and CNN models are evaluated on the test set described in section 2.2.7. Figure 4.1 shows an example prediction for illustrational purposes.

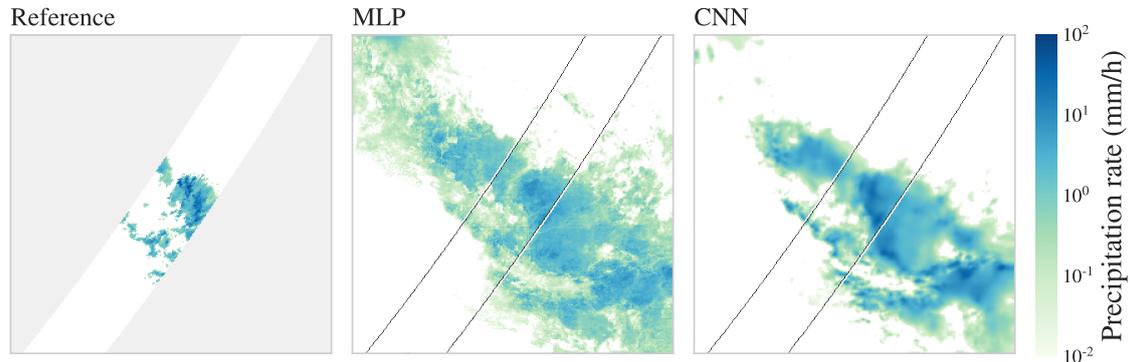


Figure 4.1: Example predictions of selected QRNN models (right) on a sample from the test set (left). In the prediction images gray lines mark the edge of the GPM swath where the ground truth data exists.

In figure 4.2 we see the observed precipitation rate distribution of the test set overlaid with the distributions of the QRNN posterior mean and 95th quantile estimates. It is clear that both posterior mean estimates fall short of capturing the heavy precipitation tail. However, the CNN appear to do a bit better than the MLP. The 95th percentile estimates are also included here. Those are shifted towards higher rates and appear to encase the tail.

Table 4.1 contains regression metric values for both models. The CNN outperforms the MLP on all but the 0.1 – 1 mm/h range which makes up 2.7% of the true rates. Comparing the bias and MAE suggest that both models have a tendency to overestimate low rates 0 – 0.1 mm/h and underestimate rates above 10 mm/h.

4. Results

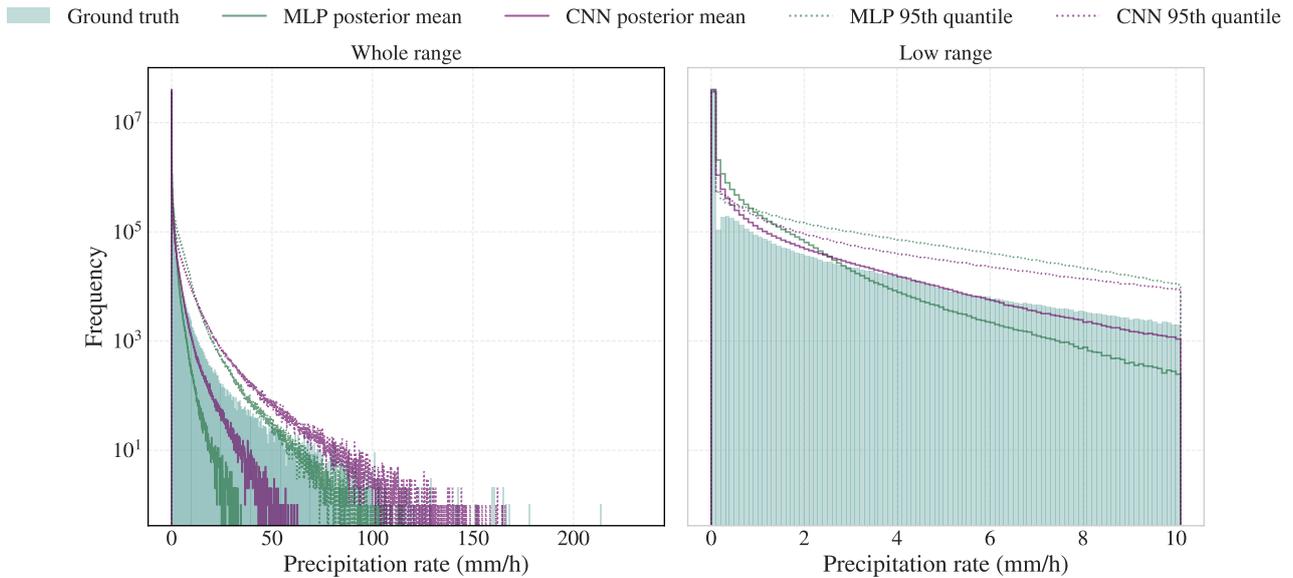


Figure 4.2: Distribution of true and predicted precipitation rates for whole test data set. Dotted lines show the 95th quantile of the QRNN models predictions, while solid lines mark posterior means. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.

Table 4.1: Bias, MAE and MSE calculated for the QRNN models posterior mean on either the whole test set, or on ranges corresponding to values in the reference data. The intervals are specified by the column "Ranges". The column "Fraction" displays the ratio of true values in the range and the total amount of true values. Bold font marks the best value in each row.

Metric	Range		Fraction	MLP	CNN
Bias	all		1	-0.00952	-0.00980
	0	10^{-1}	0.94	0.0788	0.0362
	10^{-1}	10^0	0.027	0.336	0.418
	10^0	10^1	0.030	-1.69	-0.717
	10^1	10^3	0.0025	-16.8	-13.4
MAE	all		1	0.197	0.148
	0	10^{-1}	0.94	0.0789	0.0364
	10^{-1}	10^0	0.027	0.639	0.730
	10^0	10^1	0.030	2.11	1.97
	10^1	10^3	0.0025	16.8	13.9
MSE	all		1	1.53	1.23
	0	10^{-1}	0.94	0.0951	0.0561
	10^{-1}	10^0	0.027	1.10	1.82
	10^0	10^1	0.030	8.21	7.91
	10^1	10^3	0.0025	458	351

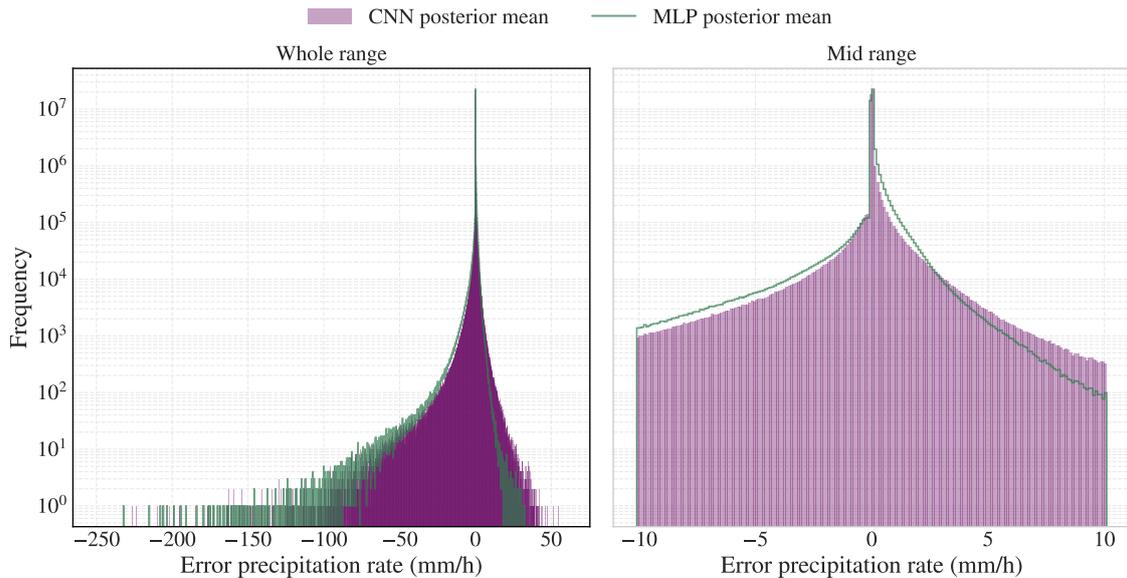


Figure 4.3: Difference between true and predicted precipitation rates for whole test data set. Right subplot shows a close up of the range -10.1 mm/h to 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.

Figure 4.3 shows the distribution of errors on the test set. We see that most errors are concentrated around zero, with the majority on the positive side i.e. predictions slightly above the observed values. Both models present a larger left tail. We can even see a handful of MLP errors larger than 200 mm/h that must originate from extreme precipitation rates being estimated as very small. On the whole, it seems that the MLP is more prone to large underestimation errors. When it comes to overestimation errors on the other hand, the CNN presents a longer right tail, however not as extensive as the left one. Note also the surplus of the MLP distribution for low positive error rates.

Table 4.2: Loss and Continuous Ranked Probability Score (CRPS) mean and median over over all predicted quantiles calculated for the QRNN models on the test set. Bold font marks the best value in each row.

Metric	MLP	CNN
Loss mean	0.0625	0.0504
CRPS mean	0.116	0.0945
CRPS median	0.000203	0.000159

Table 4.2 show probabilistic accuracy measures loss (3.4) and CRPS (3.14). Also here we can note that the CNN yields lower values than the MLP.

Figure 4.4 shows two-dimensional histograms or heatmaps of the predictions in relation to the true labels. Correct predictions fall on the $y = x$ line, drawn as the dashed diagonal. Due to the shape of the precipitation rate distribution (see figure 2.9) with a heavy predominance of zero and small rates, a linearly spaced heatmap

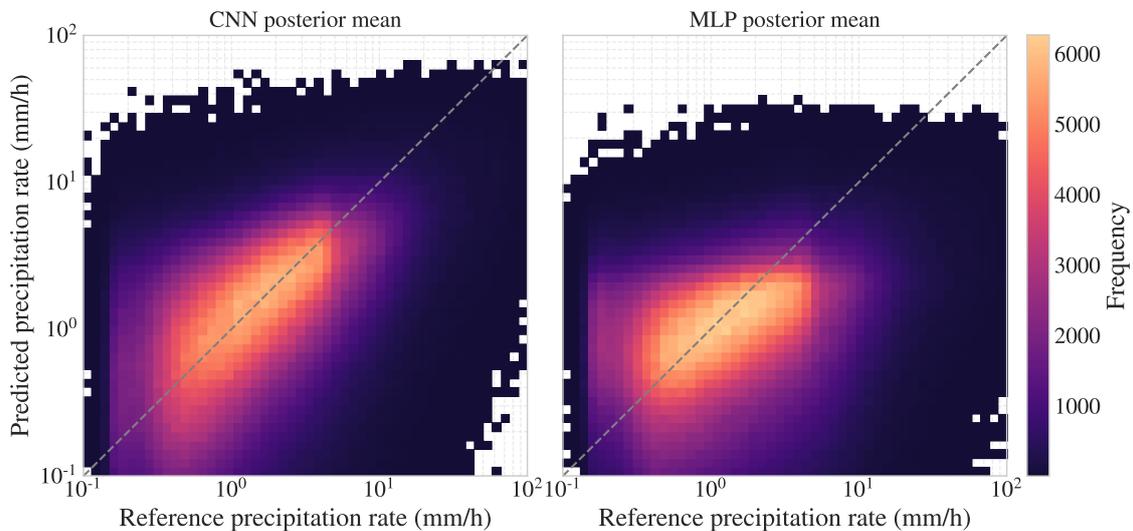


Figure 4.4: 2D-histogram showing frequency of predicted and true precipitation rates. Correct predictions fall on the dashed line. Bin sizes scales logarithmically along both axes. White bins are empty.

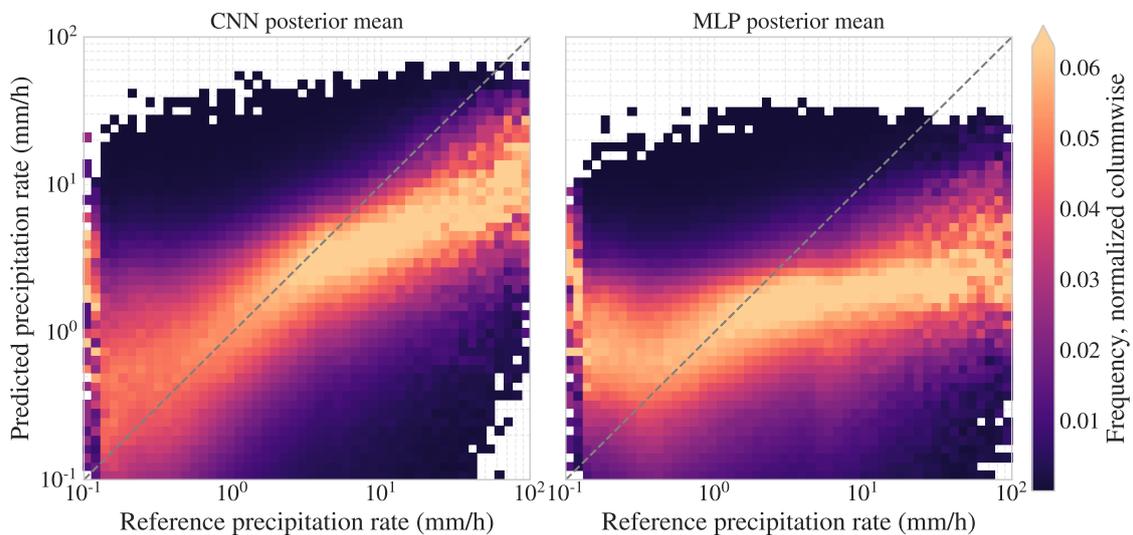


Figure 4.5: 2D-histogram showing column-wise normalized frequency of predicted and true precipitation rates. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin sizes scales logarithmically along both axes. White bins are empty.

proved not so useful here. Therefore the bins are not made equal in size, instead they are logarithmically spaced on both axes which counteracts the rate imbalance.

In this visualization, it seems the CNN predictions lie centered around the diagonal, while the MLP predictions seem to fall on a somewhat less steep line. For completeness, we also include a second visualization, figure 4.5, that shows a similar heatmap, with the difference being the color scaling. In this figure a column-wise normalization is applied. This enables us to study where the bulk of the predictions fall independently for each bin of reference values. Here we can distinguish the same behavior for the lower rates as in figure 4.4, but also that the predictions defer from the line at certain points for both models.

As stated in section 3.5.2 we categorize rates below 10^{-1} mm/h as no precipitation in the context of applying categorical metrics. Figure 4.6 and 4.7 show the distribution of values corresponding to false positives (no precipitation predicted as precipitation) and false negatives (precipitation predicted as no precipitation). First, looking at figure 4.6 we can note that the MLP gives rise to a larger amount of miss-classified observations, but that the CNN distribution has a longer tail with a few no precipitation observations being predicted as above 30 mm/h rates.

In figure 4.7 we see instead observed non-zero precipitation rates that have been miss-classified as no precipitation. Here, the MLP seems to be missing a good deal more than the CNN throughout all rates except for the lowest ones.

Additionally, we consider figure 4.8 which displays a Receiver Operating Charac-

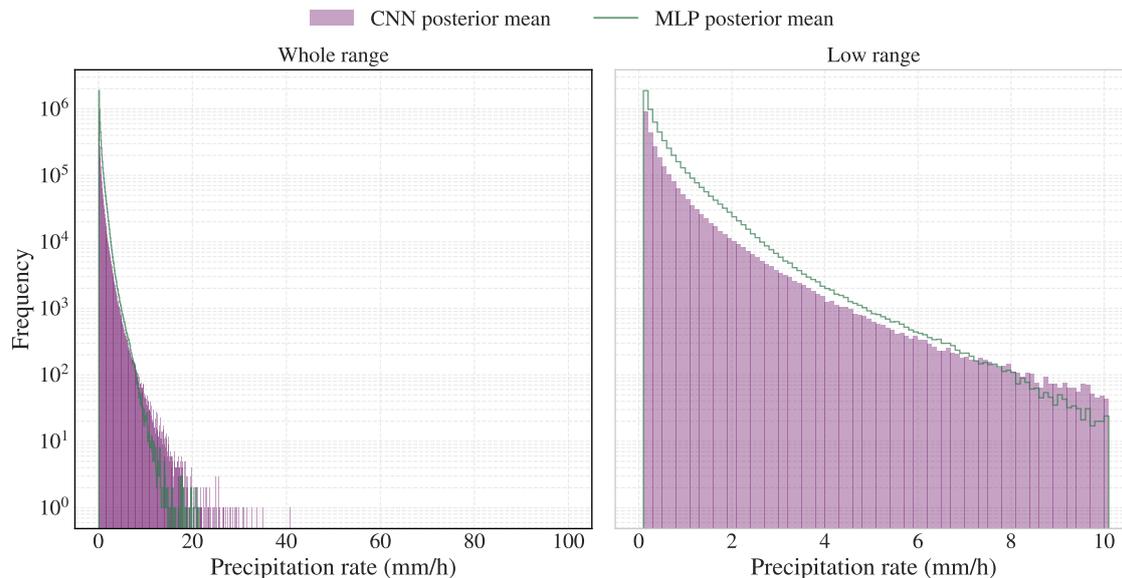


Figure 4.6: False Positives. The distribution of predictions above the threshold 10^{-1} mm/h that correspond to values below the threshold in the reference data. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.

4. Results

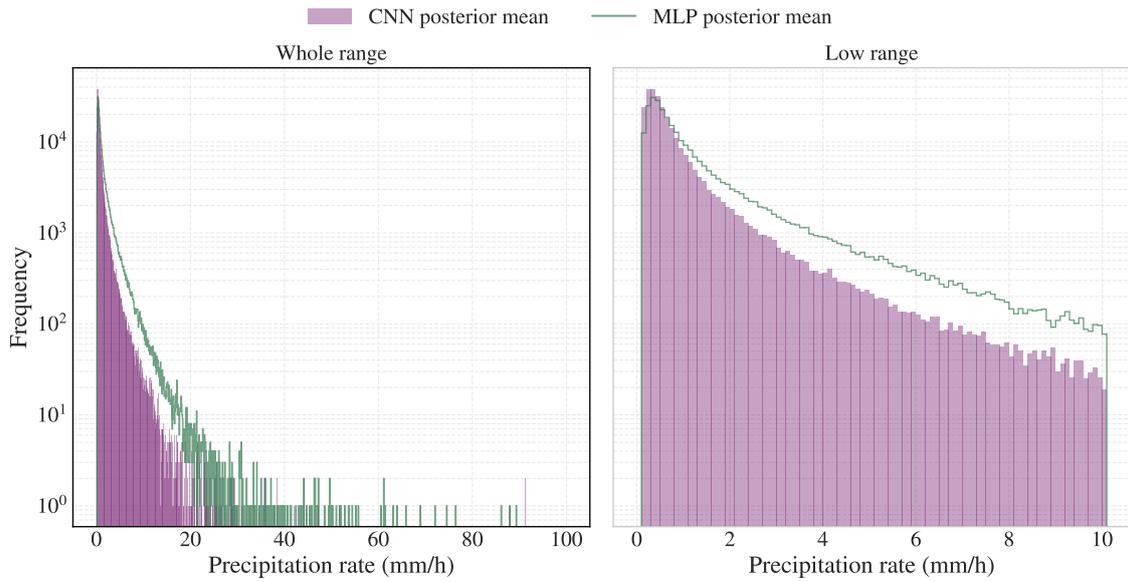


Figure 4.7: False Negatives. The distribution of reference values above the threshold 10^{-1} mm/h that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm/h. Bin width 0.1 mm/h. Logarithmic scaling of the vertical axis.

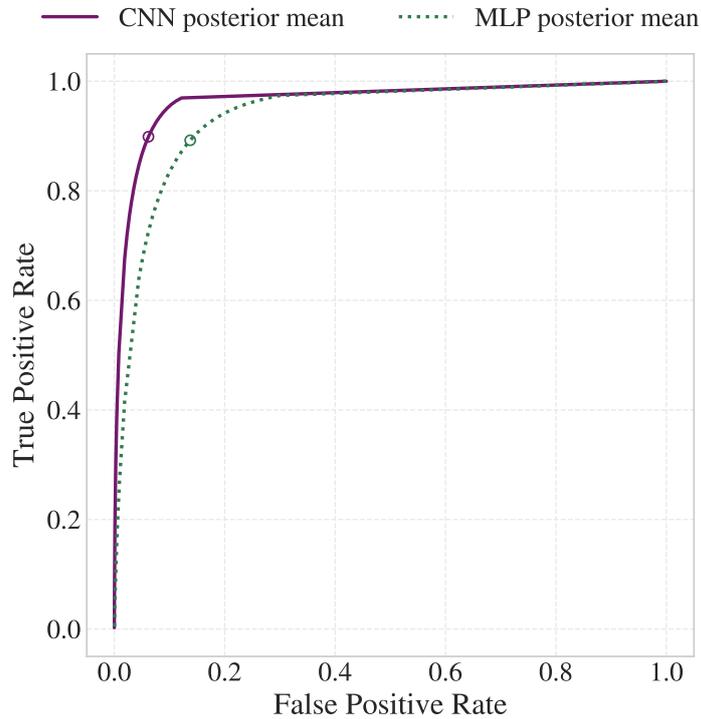


Figure 4.8: Curve obtained by varying the precipitation threshold for classifying predictions as *precipitation/no precipitation*. Marks at the threshold 0.1 mm/h.

teristic (ROC) curve that shows the performance in terms of True Positive Rate (3.12) and False Positive Rate (3.13) of a classifier at varied classifying thresholds. Here, we consider the classifiers to the QRNN models in combination with applying a precipitation rate threshold to their predictions. The curve starts out in $(0, 0)$ for thresholds exceeding all precipitation rates. As the threshold is successively lowered the curve grows monotonically, to finally end up in $(1, 1)$ for the threshold 0 mm/h . An ideal classifier that identifies all precipitation observations as precipitation without wrongly classifying no precipitation as precipitation would have $r_{\text{TPR}} = 1$, $r_{\text{FPR}} = 0$ in the upper left corner. A random guess would give a point on the diagonal between $(0, 0)$ and $(1, 1)$. In figure 4.8 we can see that both the MLP and the CNN curves come quite close to the corner.

4.2 Evaluation on rain gauge data

Moving forward in the evaluation, this section contains the comparison of the Hydroestimator (HE) and the QRNN models on the rain gauge dataset introduced in section 2.1.3. Note that the quantity considered here is hourly precipitation (as opposed to precipitation rate), and thus the retrieval predictions are aggregated to match these rain gauge measurements. To promote readability, we include only the CNN and the corrected HE for some of the figures in this chapter. For complementary figures, see appendix A.

Table 4.3: Bias, MAE and MSE calculated for the QRNN and HE models on the rain gauge dataset. Bold font marks the best value in each row.

Metric	HE	HE corr.	MLP	CNN
Bias	0.104	-0.00655	-0.0213	0.00423
MAE	0.394	0.300	0.265	0.232
MSE	3.94	2.44	2.04	1.70

Table 4.3 displays regression metrics for the models on the rain gauge dataset. According to these metrics, the CNN produces the best results. The MLP does second best when looking at the MAE and MSE scores, however the corrected HE shows a better bias.

In figure 4.9 we see the rain gauge distribution overlaid by the distributions of the HE and QRNN predictions. The standard HE is shifted towards higher values. It predicts too few non-zero rates below 0.7 mm and too many above, except for in the tail where it covers the higher precipitation values quite well. The MLP and the CNN seem to follow the same pattern as in figure 4.2; the MLP distribution starts off on top, to switch places with the CNN distribution somewhere just above 2 mm . Both distributions are shifted toward too low. The corrected HE distribution resembles the CNN distribution for most values, but diverges from it in the low range, where it displays a sharp drop.

4. Results

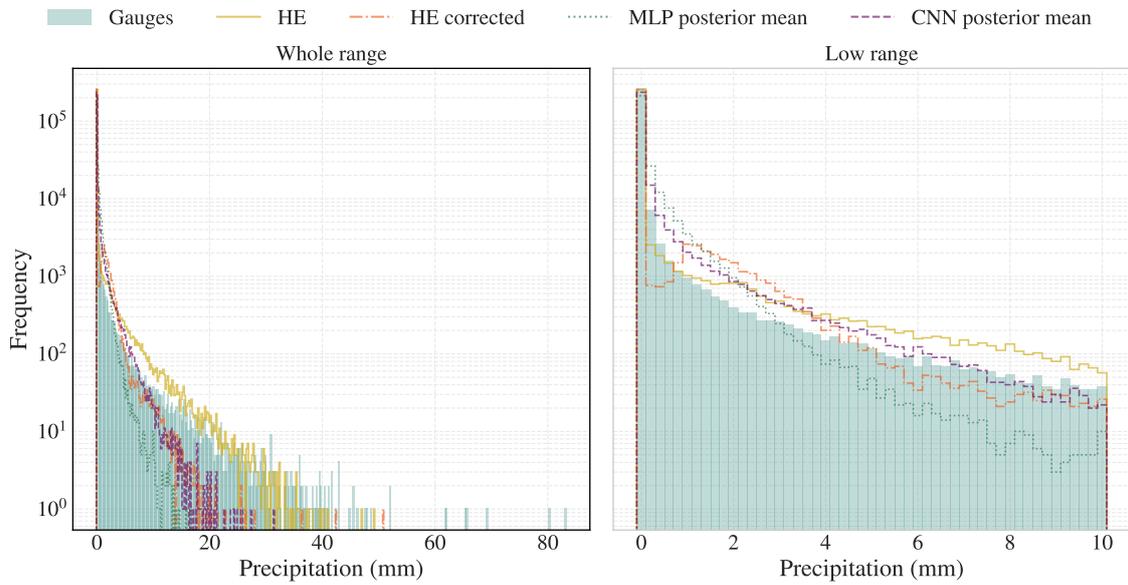


Figure 4.9: Distribution of true and predicted precipitation rates for the rain gauge dataset. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

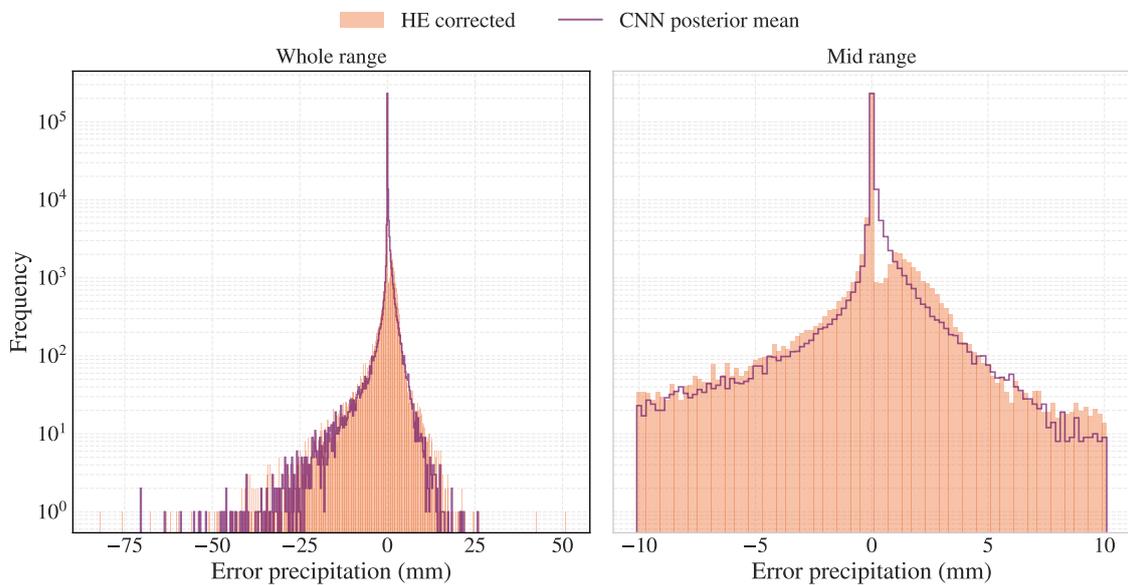


Figure 4.10: Difference between true and predicted precipitation rates for the corrected HE and the CNN on the rain gauge dataset. Right subplot shows a close up of the range -10.1 mm to 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

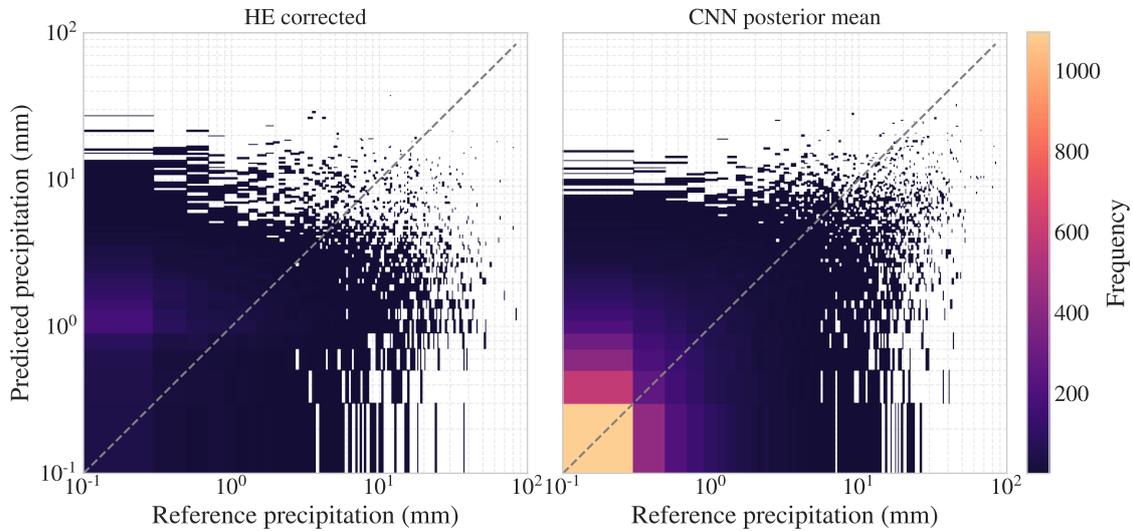


Figure 4.11: 2D-histogram showing frequency of predicted (by the corrected HE and the CNN) and true precipitation rates for the rain gauge dataset. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.

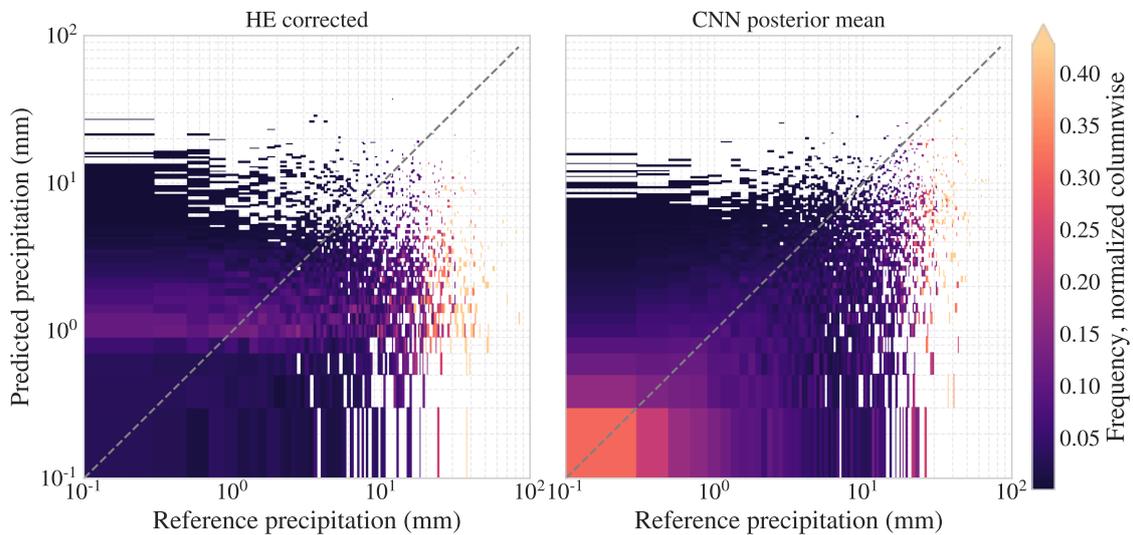


Figure 4.12: 2D-histogram showing column-wise normalized frequency of predicted (by the corrected HE and the CNN) and true precipitation rates for the rain gauge dataset. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.

4. Results

Figure 4.10 shows error distributions for the corrected HE and the CNN on the rain gauge dataset. Both distributions have a longer left tail, and seem to be quite similar except for in the 0 – 4 mm range.

Figure 4.11 and 4.12 show two-dimensional histograms of predictions from the corrected HE and the CNN in relation to the true rain gauge observations. Accurate predictions fall on the $y = x$ line. It appears as if the predictions from the corrected HE follow a horizontal line while the CNN predictions lie closer to the diagonal.

Figure 4.13 and 4.14 show the distribution of values corresponding to false positives and false negatives for the corrected HE and the CNN predictions on the rain gauge dataset. The false positives distribution (no precipitation predicted as precipitation) for the two models look to be similar overall, except for in the low range. The false negative distributions (precipitation predicted as no precipitation) on the other hand, displays differences. Here, the CNN distribution seems to have a shorter tail and to lie below the corrected HE for all values.

Figure 4.15 shows ROC curves for *precipitation/no precipitation* classifiers based on the HE and QRNN models. Compared to figure 4.8 in the previous section, the QRNN models do somewhat worse on the gauge dataset. The HE classifier curves defer early on.

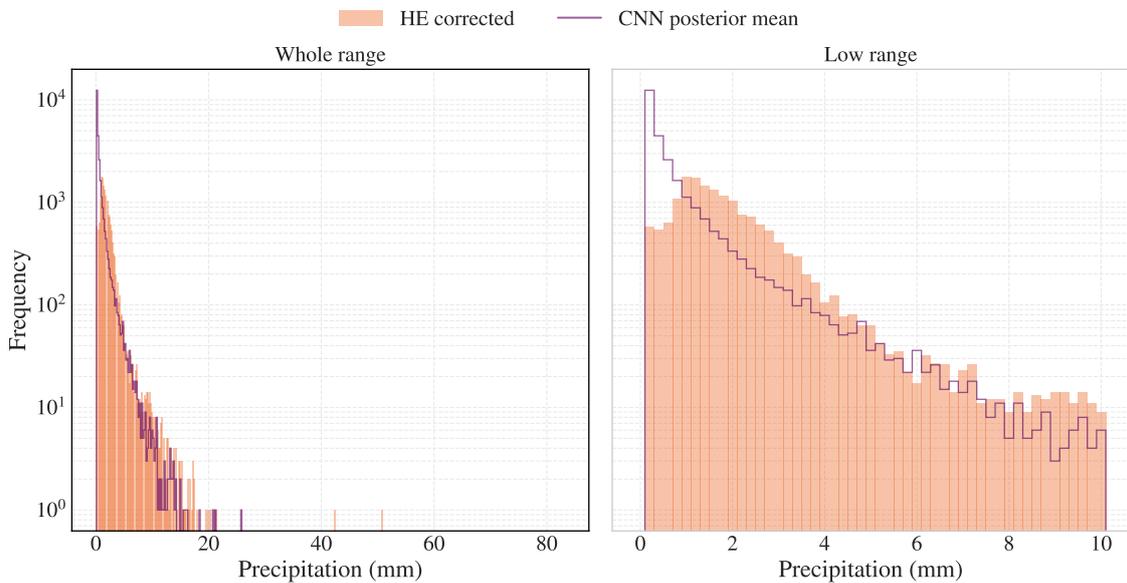


Figure 4.13: False Positives. The distribution of predictions above the threshold 10^{-1} mm that correspond to values below the threshold in the rain gauge data. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

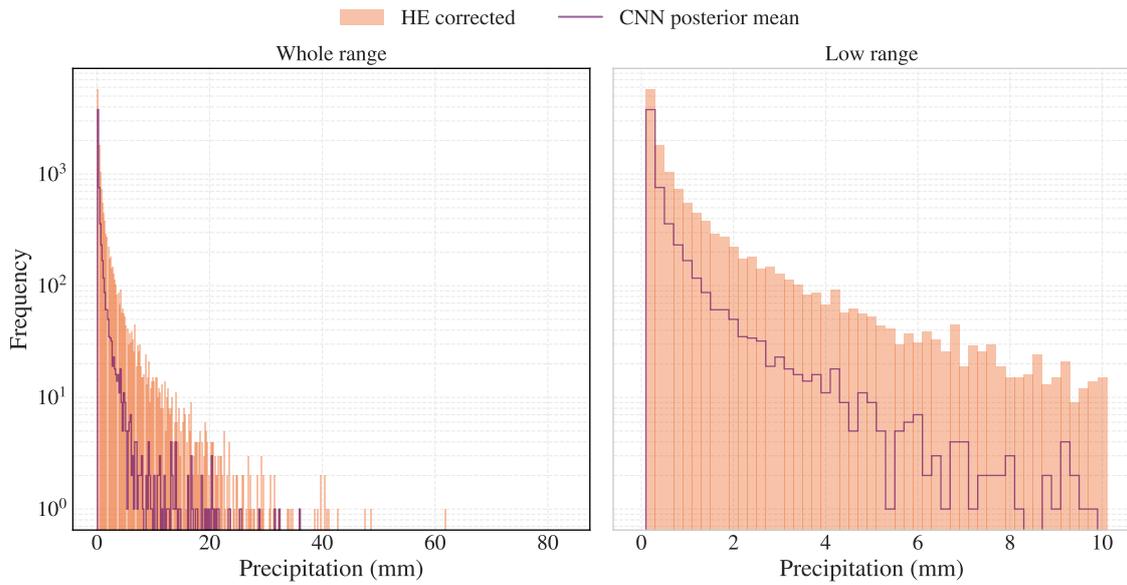


Figure 4.14: False Negatives. The distribution of rain gauge values above the threshold 10^{-1} mm that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

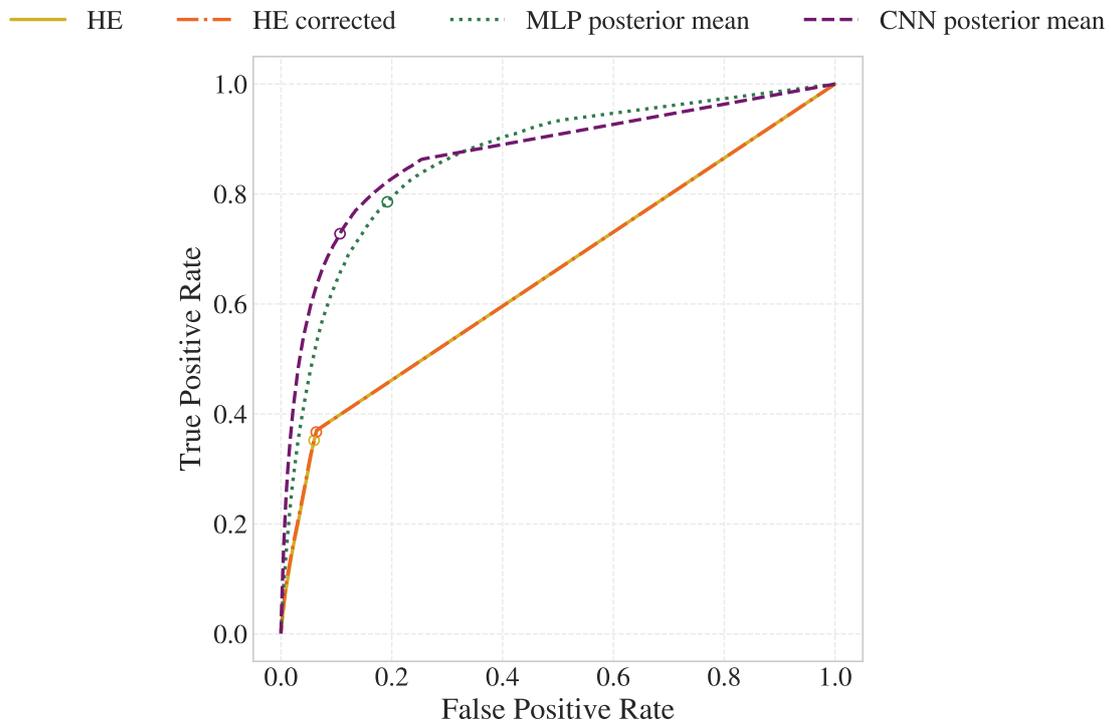


Figure 4.15: Curve obtained by varying the precipitation threshold for classifying predictions as *precipitation/no precipitation*. Marks at the threshold 0.1 mm.

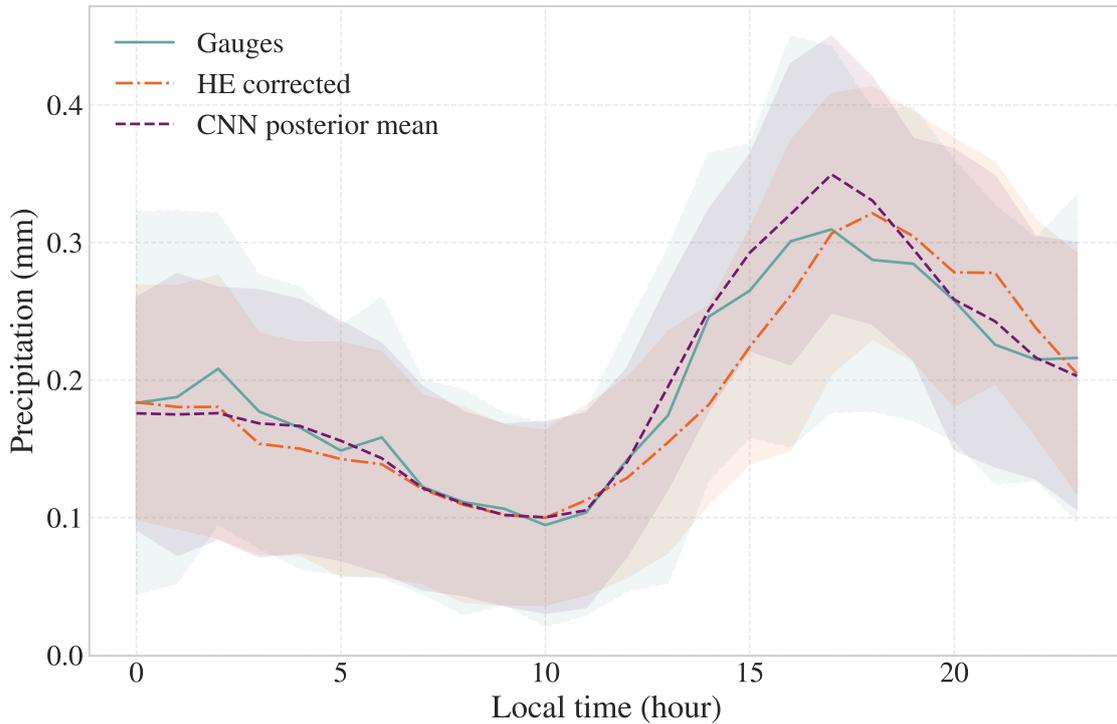


Figure 4.16: For each hour (local time) the mean precipitation over all available locations and dates in the rain gauge dataset and corresponding corrected HE and CNN predictions. Filled region represent plus minus one standard deviation of the locations mean over the dates.

The diurnal cycle is an important aspect of climate variability that should be reflected by a good model. In figure 4.16, the diurnal variation in the rain gauge dataset is shown, together with corresponding predictions from the corrected HE and the CNN. Both follow the fluctuations of the gauges quite closely, except for in the afternoon. Here, the corrected HE present a shifted peak, and the CNN an overshoot.

4.2.1 An extreme precipitation event

On December 22nd and 23rd, 2020, an extreme precipitation event was recorded in Xerém in the municipality of Duque de Caxias, Rio de Janeiro State. 224.2 mm was measured in less than 24 hours, causing severe flooding and landslides (Davies 2021). Figure 4.17 show the measurements from the Xerém rain gauge together with predictions from the corrected HE and the CNN. For the latter, the 95th and 99th quantiles are included – even though the reader is urged to approach these with some caution, since what is displayed should be viewed as a preliminary approximation. These hourly uncertainty estimates were computed in the same manner as the posterior means, disregarding correlation between subsequent retrievals, which may be a strong assumption to make. However, it is promising that an extreme precipitation event such as this one might be included in the QRNN distribution tail.

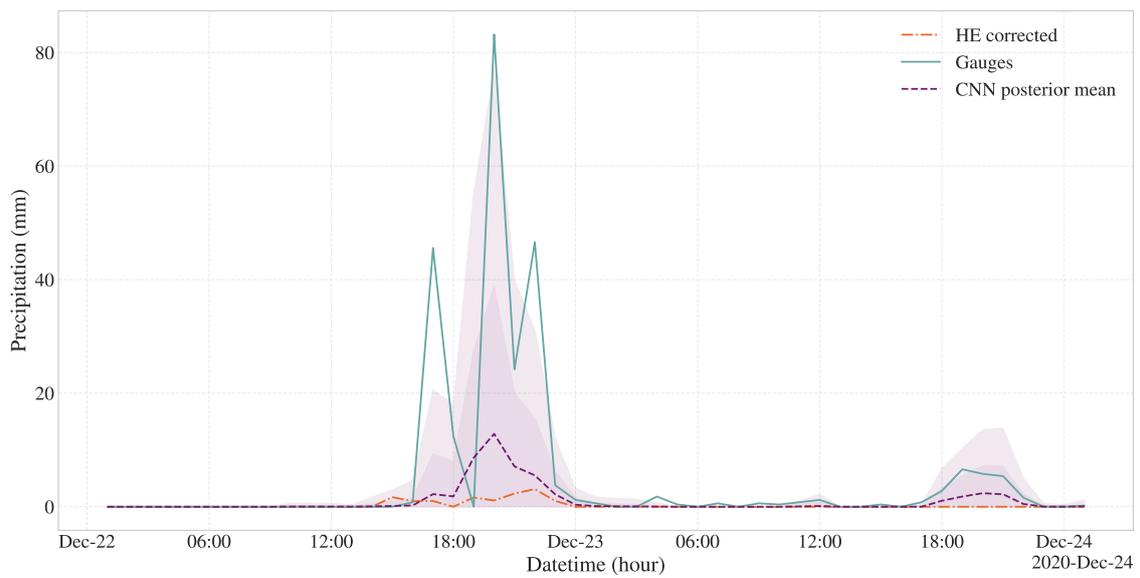


Figure 4.17: Hourly observed and predicted precipitation in Xerém, Rio de Janeiro in the period Dec 22nd to 24th, 2020. Filled regions show hourly means of the CNN’s estimated 95th and 99th quantile.

5

Discussion

We restate the main question of this thesis: is it possible to learn a mapping from the GOES-16 IR observations to a probability distribution over precipitation rates by use of QRNNs? The results indicate that this is indeed doable, and that the QRNN model performances go beyond the currently-in-operation HE. Moreover, the comparison of the two QRNN models overall suggests that the CNN is the preferred model for the task. In the following chapter we will shine some light upon the main findings.

In figure 4.2 we saw that neither the CNN nor the MLP posterior mean estimates captures the distribution of rates in the test dataset properly. This result is expected statistically because of the retrieval uncertainties which motivated the probabilistic approach in the first place. The far right of the distribution tail stems from rare precipitation occurrences that might rather be explained by the outskirts of the probability distributions than their posterior means. Another contributing factor to the underestimation can be attributed to the physics. In order to produce high precipitation, the clouds need to be very thick and deep, causing saturation of the IR signal. Nevertheless, what can be read from the plot is that the 95th quantile estimates encloses the distribution – which means that the observed extreme intensities fall inside the estimated probability distributions.

The CNN comes closer to the observed distribution than the MLP in figure 4.2. This resonates well with the calculated scoring rules in table 4.2, and most metrics in table 4.1. The 2D-histograms in figure 4.4 and 4.5 indicates stronger correlation between predicted and true rates and the ROC curve in figure 4.8 improved classification properties for the CNN. In short, the results show that the CNN outperforms the MLP. This suggests that including spatial information improves the retrieval.

Figure 4.17 shows predictions for an extreme precipitation event where the right tail quantities display way higher precipitation intensities than the posterior mean, matching the observed intensities. While this is an interesting result that certainly highlights the importance of considering uncertainties in the retrieval, the reader should be aware that the current implementation of the retrieval calculated the percentiles of the hourly precipitation rates only approximately. A thorough evaluation of the predicted uncertainties against independent validation data should therefore revisit the propagation of uncertainties during the accumulation of the instantaneous retrieval results to hourly averages.

Naturally, retrievals derived by supervised machine learning algorithms will always be limited by the reference data used in training, in our case the GPM combined product. This reference is itself a result of satellite retrievals, and therefore arguably suffers from retrieval uncertainties as well, to some extent. However, the assessment of the GPM products accuracy is clearly out of scope for this work.

6

Conclusion

In this work we have found that using a machine-learning based approach to the precipitation retrieval problem works better than the algorithm in operation for the region studied. In addition to this we have also seen that including spatial information in the retrieval improves the performance. This points to the potential in applying neural networks for this task, and the flexibility in the approach opens up for the opportunity of further boosting the retrieval by switching algorithms, which is a far lower cost than launching a new satellite.

Our results constitutes a promising first step in developing a new operational retrieval algorithm. With this as base, future research is encouraged to extend to all available GOES-16 channels. Other additional input might also be included, like *land/ocean* categories. The need for further validation on and comparison to other available independent reference datasets and retrieval algorithms are also emphasised as well as considering seasonal aspects. The preliminary analysis of the Xerém case presented in this work illustrates the importance of considering the retrieval uncertainties, but the retrieval performance for a larger set of extreme precipitation events should be considered in future studies.

6. Conclusion

Bibliography

- Adler, R. F. and A. J. Negri (1988). “A satellite infrared technique to estimate tropical convective and stratiform rainfall”. In: *Journal of Applied Meteorology and Climatology* 27.1, pp. 30–51.
- Arkin, P. A. and B. N. Meisner (1987). “The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84”. In: *Monthly Weather Review* 115.1, pp. 51–74.
- Ba, M. B. and A. Gruber (2001). “GOES Multispectral Rainfall Algorithm (GM-SRA)”. In: *Journal of Applied Meteorology* 40.8, pp. 1500–1514. DOI: 10.1175/1520-0450(2001)040<1500:GM-RAG>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/apme/40/8/1520-0450_2001_040_1500_gmrags_2.0.co_2.xml.
- Chollet, F. (2017). “Xception: Deep Learning with Depthwise Separable Convolutions”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.
- Davies, R. (2021). *Brazil – Deadly Flash Floods in Rio De Janeiro State*. URL: <https://floodlist.com/america/brazil-floods-baixada-fluminense-december-2020> (visited on Sept. 8, 2021).
- Earthdata Search* (n.d.). URL: <https://search.earthdata.nasa.gov/search/>.
- Gneiting, T. and A. E. Raftery (2007). “Strictly proper scoring rules, prediction, and estimation”. In: *Journal of the American statistical Association* 102.477, pp. 359–378.
- GOES Image Viewer* (July 28, 2021). *GOES-East Full Disk - GeoColor*. URL: https://www.star.nesdis.noaa.gov/GOES/fulldisk_band.php?sat=G16&band=GEOCOLOR&length=24 (visited on July 28, 2021).
- Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Greco, M., W. S. Olson, S. J. Munchak, S. Ringerud, L. Liao, Z. Haddad, B. L. Kelley, and S. F. McLaughlin (2016). “The GPM combined algorithm”. In: *Journal of Atmospheric and Oceanic Technology* 33.10, pp. 2225–2245.
- Hsu, K., X. Gao, S. Sorooshian, and H. V. Gupta (Sept. 1, 1997). “Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks”. In: *Journal of Applied Meteorology* 36.9, pp. 1176–1190. DOI: 10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

- Ioffe, S. and C. Szegedy (2015). “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. PMLR, pp. 448–456.
- Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum (2017). “So, How much of the Earth’s surface is covered by rain gauges?” In: *Bulletin of the American Meteorological Society* 98.1, pp. 69–78. DOI: <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- Kingma, D. P. and J. Ba (2014). “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980*.
- Koenker, R. (2005). “Cambridge University Press; Cambridge and New York: 2005”. In: *Quantile regression: Econometric society monographs* 38.
- LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.
- Lin, T., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie (2017). *Feature Pyramid Networks for Object Detection*. arXiv: 1612.03144 [cs.CV].
- Loshchilov, I. and F. Hutter (2016). “Sgdr: Stochastic gradient descent with warm restarts”. In: *arXiv preprint arXiv:1608.03983*.
- NASA (2021a). *Glossary. nadir*. URL: <https://earthobservatory.nasa.gov/glossary/nadir> (visited on July 29, 2021).
- NASA (2021b). *Remote Sensors*. URL: <https://earthdata.nasa.gov/learn/remote-sensors> (visited on July 25, 2021).
- NOAA (Apr. 27, 2012). *Global Hydro-Estimator - Algorithm Description*. URL: <https://www.ospo.noaa.gov/Products/atmosphere/ghe/algo.html>.
- NOAA (2021). *What is remote sensing?* URL: <https://oceanservice.noaa.gov/facts/remotesensing.html> (visited on July 29, 2021).
- NOAA/NASA (2021). *Instruments: Advanced Baseline Imager (ABI)*. URL: <https://www.goes-r.gov/spacesegment/abi.html> (visited on July 26, 2021).
- Olson, W. (2017). *GPM DPR and GMI Combined Precipitation L2B 1.5 hours 5 km V06*. DOI: 10.5067/GPM/DPRGMI/CMB/2B/06. (Visited on May 27, 2021).
- Pfreundschuh, S. (2021). *Quantile regression neural networks on top of Keras and Pytorch*. URL: <https://github.com/simonpf/quantnn> (visited on Apr. 30, 2021).
- Pfreundschuh, S., P. Eriksson, D. Duncan, B. Rydberg, N. Håkansson, and A. Thoss (2018). “A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems”. In: *Atmospheric Measurement Techniques* 11.8, pp. 4627–4643. DOI: 10.5194/amt-11-4627-2018.
- Remote Sensing Systems (2020). *Rain Rate*. URL: <https://www.remss.com/measurements/rain-rate/> (visited on July 22, 2020).
- Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: theory and practice*. Vol. 2. World scientific.

- Sadeghi, M., A. A. Asanjan, M. Faridzad, P. Nguyen, K. Hsu, S. Sorooshian, and D. Braithwaite (2019). “PERSIANN-CNN: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks?Convolutional Neural Networks”. In: *Journal of Hydrometeorology* 20.12, pp. 2273–2289. DOI: 10.1175/JHM-D-19-0110.1. URL: https://journals.ametsoc.org/view/journals/hydr/20/12/jhm-d-19-0110_1.xml.
- Scofield, R. A. and R. J. Kuligowski (2003). “Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events”. In: *Weather and Forecasting* 18.6, pp. 1037–1051.
- Siqueira, R. and D. Vila (Jan. 2019). “Hybrid methodology for precipitation estimation using Hydro-Estimator over Brazil”. In: *International Journal of Remote Sensing* 40, pp. 1–20. DOI: 10.1080/01431161.2018.1562262.
- Skofronick-Jackson, G., W. A. Petersen, W. Berg, C. Kidd, E. F. Stocker, D. B. Kirschbaum, R. Kakar, S. A. Braun, G. J. Huffman, T. Iguchi, P. E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W. S. Olson, Y. N. Takayabu, K. Furukawa, and T. Wilheit (2017). “The Global Precipitation Measurement (GPM) Mission for Science and Society”. In: *Bulletin of the American Meteorological Society* 98.8, pp. 1679–1695. DOI: 10.1175/BAMS-D-15-00306.1. URL: <https://journals.ametsoc.org/view/journals/bams/98/8/bams-d-15-00306.1.xml>.
- Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna (June 2016). “Rethinking the Inception Architecture for Computer Vision”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Tabari, H. (2020). “Climate change impact on flood and extreme precipitation increases with water availability”. In: *Scientific reports* 10.1, pp. 1–10.
- Tapiador, F. J., F. J. Turk, W. Petersen, A. Y. Hou, E. Garcia-Ortega, L. A. T. Machado, C. F. Angelis, P. Salio, C. Kidd, G. J. Huffman, et al. (2012). “Global precipitation measurement: Methods, datasets and applications”. In: *Atmospheric Research* 104, pp. 70–97.
- Vicente, G. A., R. A. Scofield, and W. P. Menzel (1998). “The Operational GOES Infrared Rainfall Estimation Technique”. In: *Bulletin of the American Meteorological Society* 79.9.

A

Appendix 1

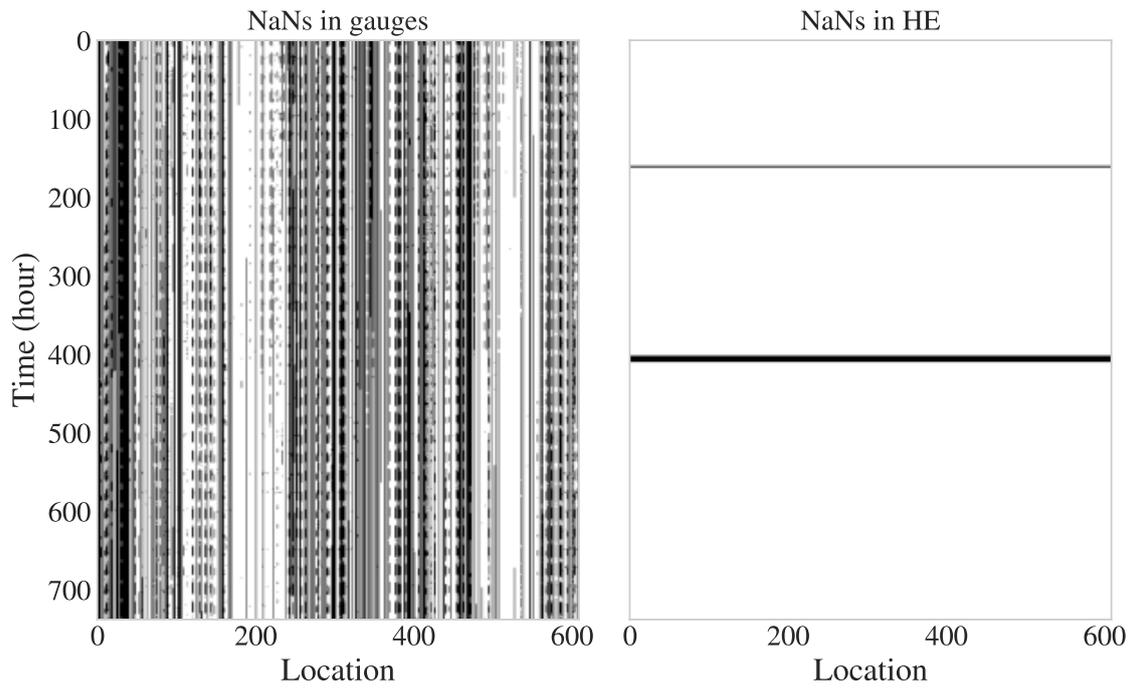


Figure A.1: Missing values (represented in black) in the rain gauge dataset and the corresponding HE estimates.

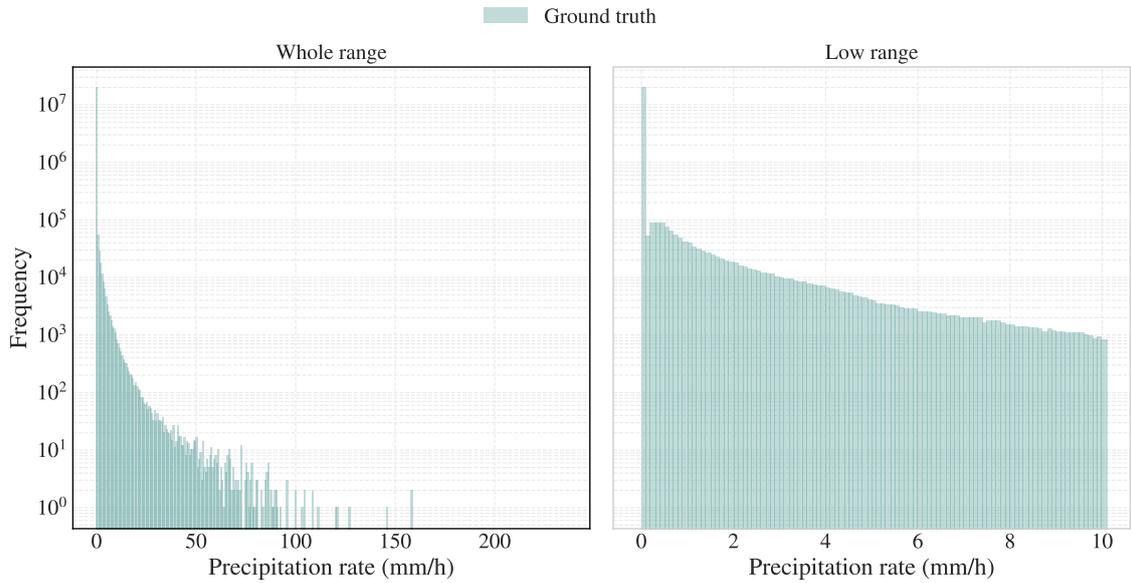


Figure A.2: Precipitation distribution for the GPM *validation* data, bin width 0.1 mm. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 10.1 mm/h.

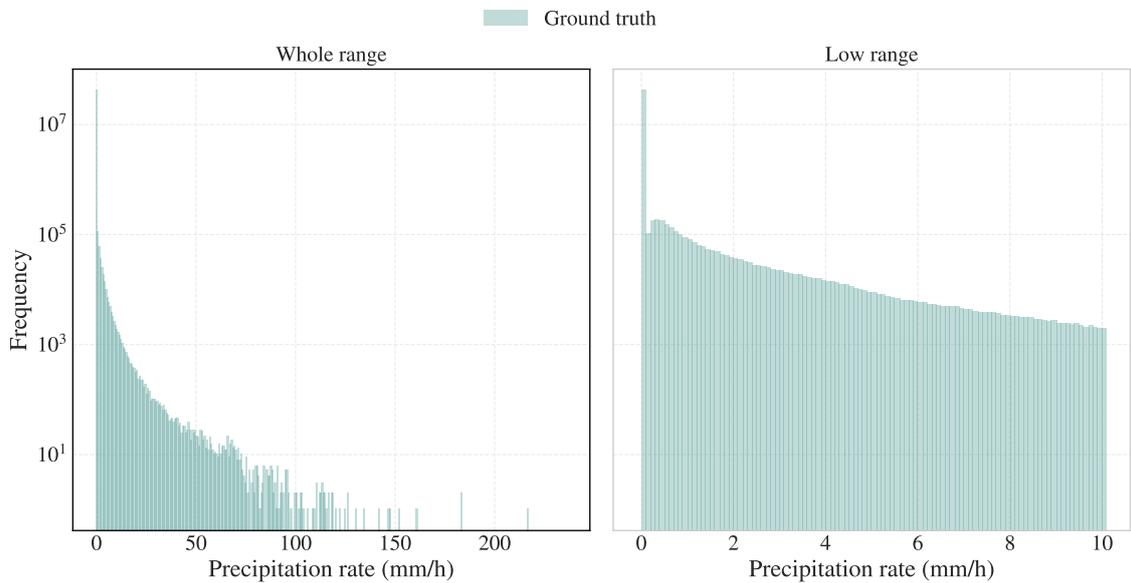


Figure A.3: Precipitation distribution for the GPM *test* data, bin width 0.1 mm. Note the logarithmic scaling on the vertical axis. Right image is a close up of the range below 10.1 mm/h.

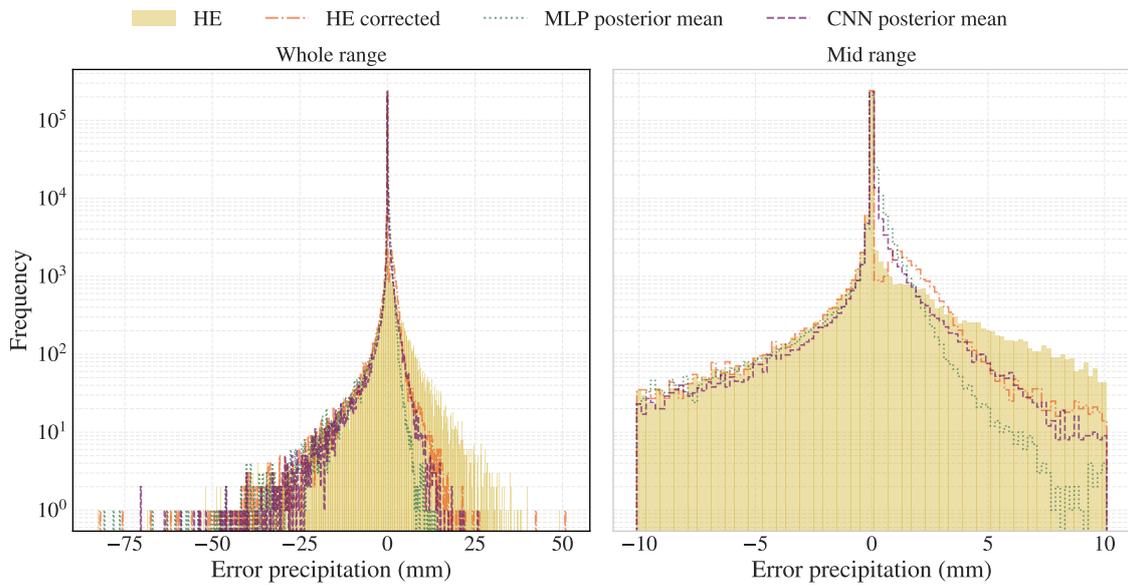


Figure A.4: Difference between true and predicted precipitation rates for the HE and the QRNN models on the rain gauge dataset. Right subplot shows a close up of the range -10.1 mm to 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

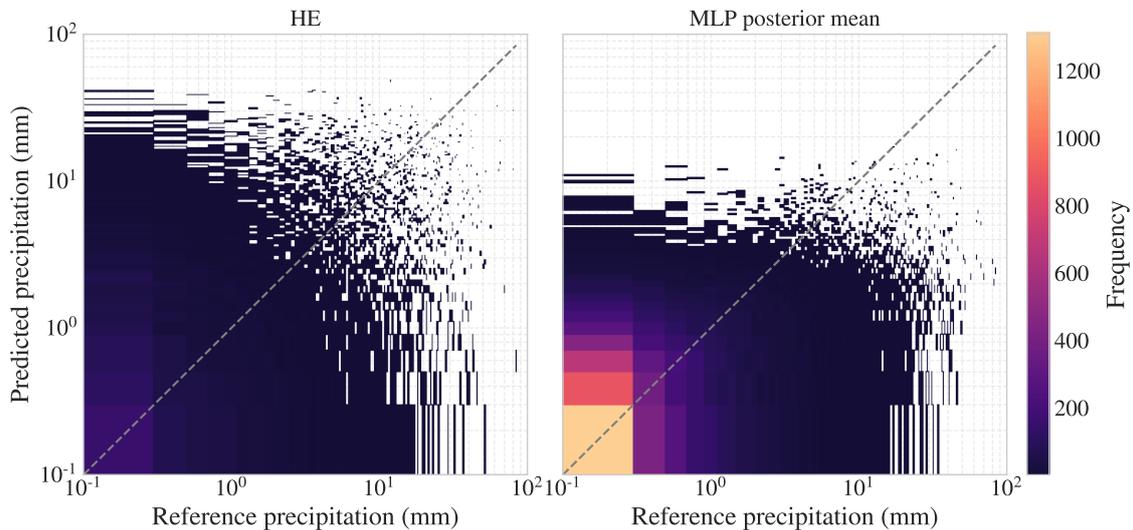


Figure A.5: 2D-histogram showing frequency of predicted (by the HE and the MLP) and true precipitation rates for the rain gauge dataset. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.

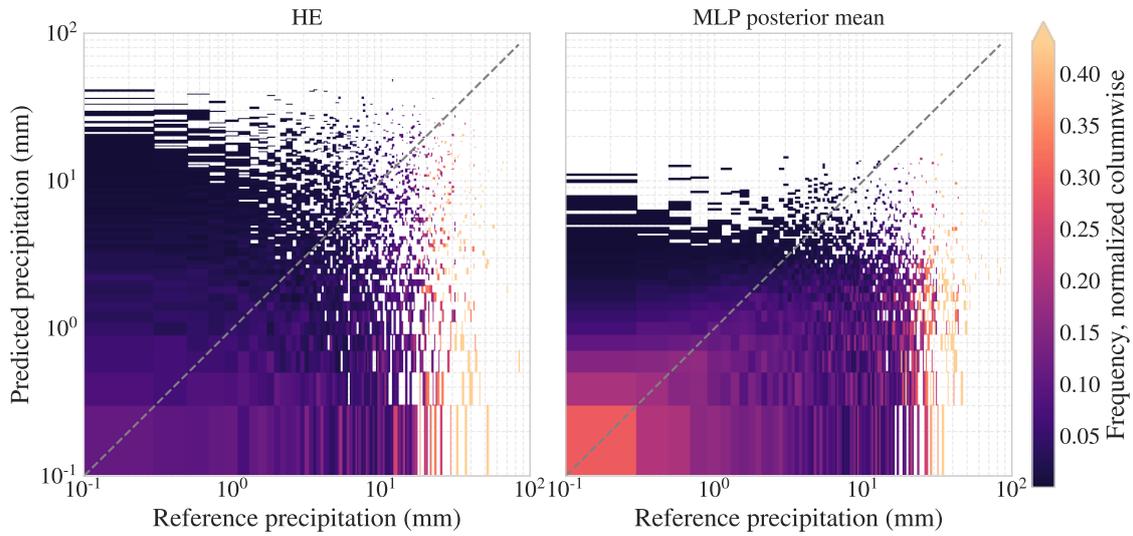


Figure A.6: 2D-histogram showing column-wise normalized frequency of predicted (by the HE and the MLP) and true precipitation rates for the rain gauge dataset. Note the clipped colorbar. Correct predictions fall on the dashed line. Bin width 0.2 mm, logarithmic scaling of both axes. White bins are empty.

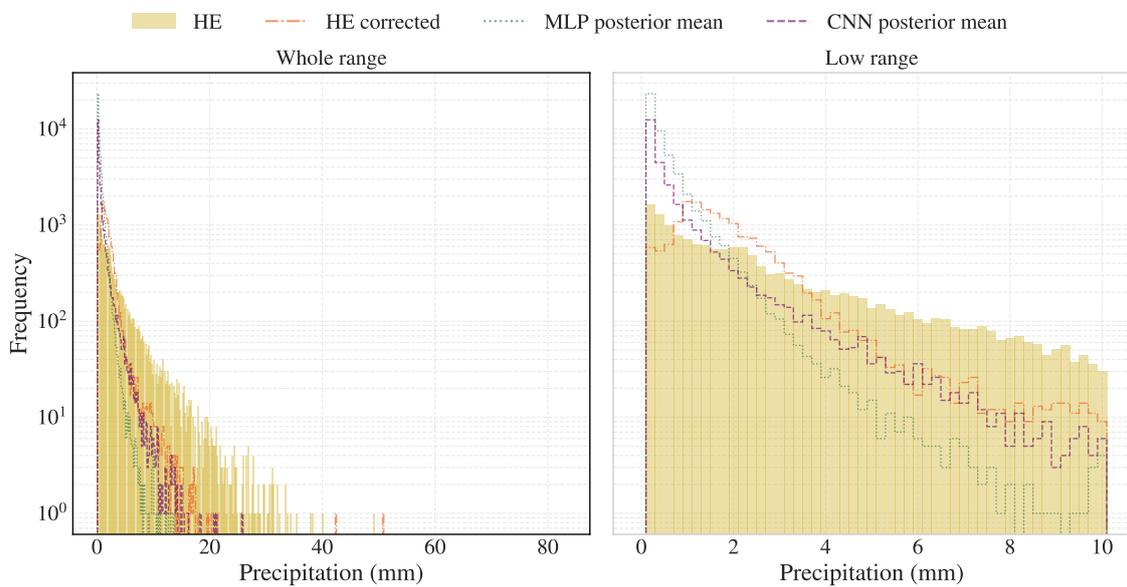


Figure A.7: False Positives. The distribution of predictions above the threshold 10^{-1} mm that correspond to values below the threshold in the rain gauge data. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

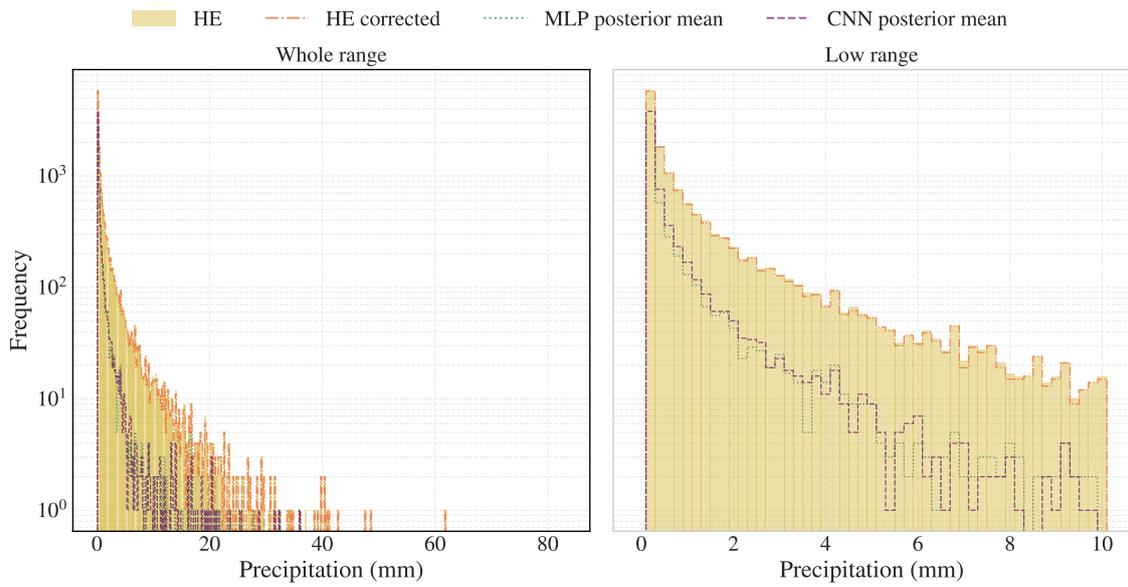


Figure A.8: False Negatives. The distribution of rain gauge values above the threshold 10^{-1} mm that correspond to predictions below the threshold. Right subplot shows a close up of the range below 10.1 mm. Bin width 0.2 mm. Logarithmic scaling of the vertical axis.

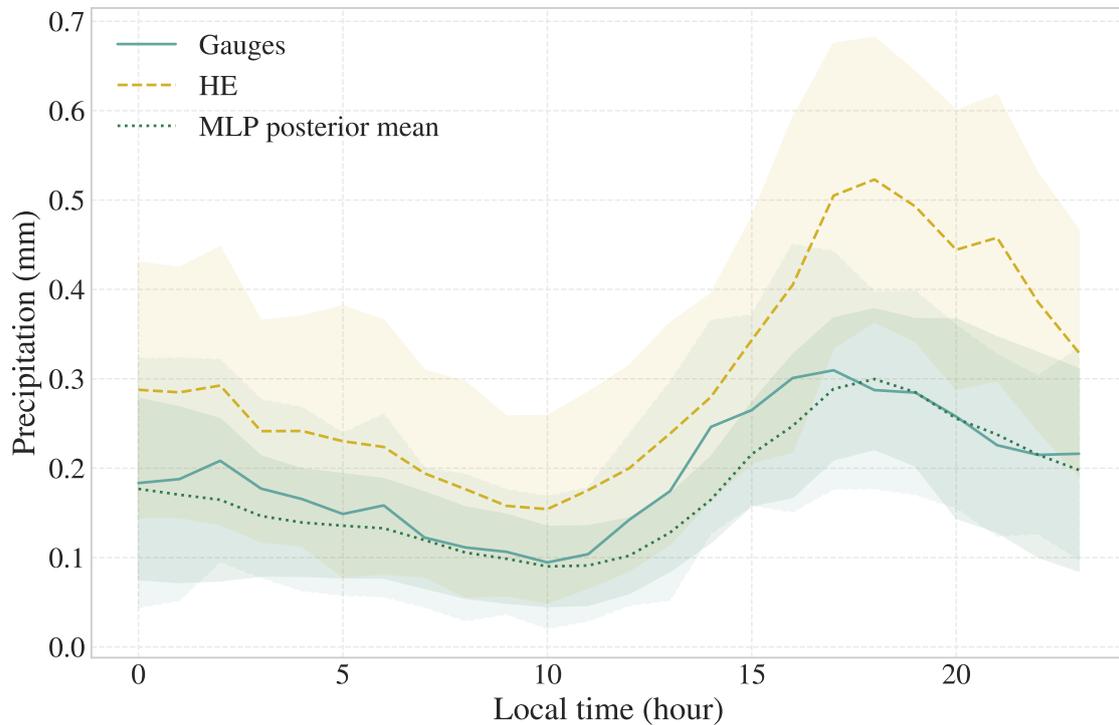


Figure A.9: Graph displays for each hour (local time) the mean precipitation over all available locations and dates in the rain gauge dataset and corresponding HE and MLP predictions. Filled region represent plus minus one standard deviation of the locations mean over the dates.

DEPARTMENT OF SPACE, EARTH AND ENVIRONMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY