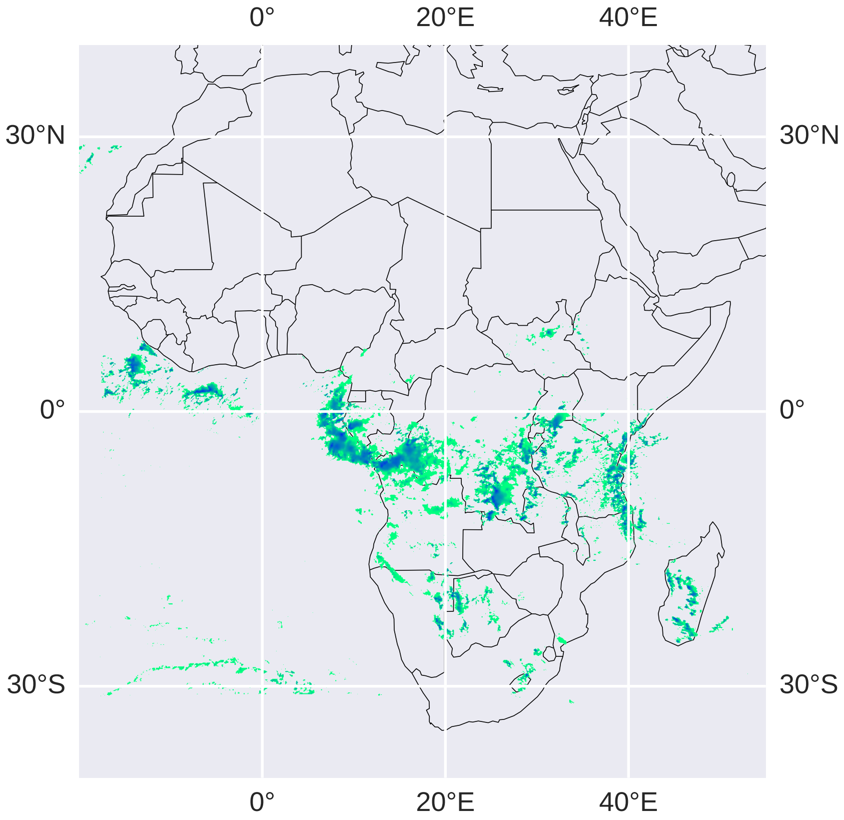

Rain over Africa

An Application of Quantile Regression Neural Networks to Retrieve Precipitation from Geostationary Satellites

Master's thesis in Complex Adaptive Systems

LILIAN HEE

DEPARTMENT OF SPACE, EARTH AND ENVIRONMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

Rain over Africa

An Application of Quantile Regression Neural Networks to Retrieve Precipitation from Geostationary Satellites

LILIAN HEE

Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden 2022

Rain over Africa
An Application of Quantile Regression Neural Networks to
Retrieve Precipitation from Geostationary Satellites
LILIAN HEE

Supervisors: Adrià Amell & Patrick Eriksson, Department of Space, Earth and Environment
Examiner: Patrick Eriksson, Department of Space, Earth and Environment

Master's Thesis 2022
Department of Space, Earth and Environment
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Precipitation over Africa as retrieved by quantile regression neural network applied in this work. Figure illustrates rain over Africa on December 1st 2021, 10:00 UTC.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Rain over Africa
An Application of Quantile Regression Neural Networks to
Retrieve Precipitation from Geostationary Satellites
LILIAN HEE
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

With an increasing change in Earth's climate, global precipitation assessments are essential to understand the hydrological cycle and prepare society for extreme weather events. As of today, many countries are equipped with networks of weather stations that individually measure precipitation over small areas. However, adequate spatial coverage is not feasible for large regions like the African continent. This problem can be overcome by utilising geostationary satellites. They scan large surfaces of the Earth with a short revisit period, making them suitable for global nowcasting. Unfortunately, direct measurements cannot be made as surface precipitation rates need to be derived from an indirect relation to measured infrared (IR) and/or visible radiation.

The problem is addressed by applying a Quantile Regression Neural Network (QRNN) to predict the a posteriori distribution of precipitation rates over Africa. The probabilistic retrievals yield not only estimated precipitation rate values but also their associated uncertainties. Surface precipitation rate estimates with a $3\,\mathrm{km} \times 3\,\mathrm{km}$ spatial resolution are achieved for Africa every 15 minutes. The QRNN is implemented with a convolutional neural network (CNN) trained on thermal IR data from the SEVIRI instrument onboard the geostationary Meteosat-11 satellite. Additionally, the GPM DPR and GMI combined precipitation product retrieved from the low Earth orbit GPM Core Observatory satellite is used as reference data.

Two set-ups are trained, first using all thermal IR channels of the SEVIRI instrument, and secondly using only two (6.2 µm and 10.8 µm). The second set-up indicates promising use of historical data, while the first set-up shows the overall best performance. Comparisons are made with IMERG, a leading global precipitation product. The posterior distributions retrieved by the QRNN are shown to be more revealing than the scalar-valued estimations from IMERG. Moreover, it is possible to increase the spatiotemporal resolution as well as reduce the latency of the retrievals. The relative simplicity of the methodology hints at potential improvements in global precipitation products available today.

Keywords: quantile regression, neural networks, CNN, precipitation, SEVIRI, GPM, nowcasting, Africa.

Acknowledgements

My deepest and foremost thanks go to Patrick Eriksson and Adrià Amell. Their supervision and expertise taught me much and helped me understand different aspects of the project. I thank Patrick for his valuable inputs, questions, and guidance, as well as acknowledge him for his admirable initiative on the topic. To Adrià, I express my sincere gratitude for his generous sharing of information, helpful discussions, and dedicated time. I also acknowledge his code for downloading SEVIRI data, data loading, training set-up, and much more that was reused in this project.

Additionally, I wish to extend my thanks to Ingrid Ingemarsson for her kind introduction at the start of the project.

Finally, I would like to mention my family and friends and thank them for their love and for the patience they have for me. A special thanks to Mimmi, who has always had my back and pushes me when it is needed. I give my warm thanks to Julia for being the great friend that she is and for the help with setting up a website where I could place the animations from the project. Also, as this work concludes my studies at Chalmers, I am grateful to Sofia for sharing the time spent in the library, labs, and various corners of the campus with me. I thank her and Isabella for making my years at Chalmers extra enjoyable. Lastly, thanks to my little Mito, the best dog in the world, who passed away during the writing of this thesis.

Lilian Hee, Gothenburg, August 2022

Acronyms

CDF	Cumulative distribution function
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station
CNN	Convolutional neural network
CRPS	Continous ranked probability score
DPR	Dual-frequency Precipitation Radar
GMI	GPM Microwave Imager
GPM	Global Precipitation Measurement
GPM-CO	GPM Core Observatory
IMERG	Integrated Multi-satellitE Retrievals for GPM
IR	Infrared
MAE	Mean absolute error
MLP	Multi-layer perceptron
MSE	Mean squared error
MSG	Meteosat second generation
NaN	Not a number
PDF	Probability density function
QRNN	Quantile Regression Neural Network
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
TAMSAT	Tropical Applications of Meteorology using SATellite data and ground-based observations
TIR	Thermal infrared
UTC	Coordinated Universal Time

Glossary

Convective precipitation	Precipitation from clouds that develop vertically.
Diurnal cycle	Here, precipitation pattern that recurs every 24 hours.
Geostationary orbit	Orbit in which a satellite is in fixed position relative to Earth.
Low Earth orbit	Orbit in lower altitudes where satellites are in motion relative to Earth.
Nadir	Point on Earth 90° below a satellite.
Precipitation	Any liquid or frozen water that forms in the atmosphere and falls to the Earth.
Product	The results of processed data.
Reanalysis	Estimation given by combining physical models with observations.
Revisit period	The time it takes for a satellite to observe the same location on Earth.

Satellite zenith angle	The angle between a straight line from a point on the Earth's surface to the satellite and a line from the same point on the Earth's surface that is perpendicular to the Earth's surface at that point.
Stratiform precipitation	Precipitation from clouds that develop horizontally.
Thermal IR	Infrared radiation that is emitted from Earth. Here we consider wavelengths between 4–15 µm without solar contribution as thermal infrared.

Contents

1

Introduction

This thesis concerns precipitation and, more specifically, the challenges of measuring precipitation satisfactorily over large regions. In this chapter, we motivate the need for better measurements and introduce approaches that can address challenges present in existing methods today.

1.1 The importance of rain

Human life and culture have been affected by rain for as long as they have existed. Historically, rain has acted both as a symbol of blessings and curses; conversely, the same can be said for the absence of it (Cauteruccio, Colli, et al., 2021). Rain provides us with fresh water, dictates how we grow our food and build our houses, and not least, affects our day-to-day life (Handmer et al., 2012). As the Earth's climate changes, we are expected to face the extreme ends, with wet events as well as dry spells (Held and Soden, 2006; Alexander et al., 2006; Allen and Ingram, 2002). These episodes can entail severe repercussions for life all over the globe, as the consequences of a disrupted hydrological cycle due to increasing temperatures can have a more significant impact on life than the warming of global temperatures by itself (Jiménez Cisneros et al., 2014).

Heavy and unexpected rain, in particular, is of extra importance in numerous applications. For example, when it comes to issuing weather warnings or navigation, early detection of hazardous rain events can save lives and minimise the damage caused. In such applications, short-term forecasts of up to two hours may be desired (WMO, 2017). These short-term forecasts are called nowcasting and are difficult to accurately assess using traditional numerical weather predictions (Browning and Collier, 1989). Instead, better estimates are often achieved by merely observing the current state of the weather. In order to quantify these observations, measurements must be made. However, since precipitation is a complex environmental variable that involves several processes, such measurements can prove to be a challenging task.

1.2 Measuring rain from space

Precipitation encompasses every form of water that accumulates in the atmosphere and falls to the Earth. Measuring precipitation is not only important for validating climate models but crucial for nowcasting. Available methods today for measuring precipitation all have their unique advantages and disadvantages. The most accurate way of measuring precipitation that reaches the ground is done with rain gauges (Michaelides et al., 2009). However, these gauges only provide point measurements, and the number of official rain gauge sites is highly scarce in relation to the surface of the Earth (Kidd, Becker, et al., 2017). Europe, in general, is well equipped with weather stations consisting of rain gauges or weather radars, which operate using remote sensing techniques. The issue is more severe for oceans or large land areas such as the continent of Africa, where networks of weather stations with sufficient spatial coverage are not feasible. Therefore, the use of in situ methods like rain gauges and weather radars is not enough for global measurements.

Luckily, satellite observations make it possible to achieve a global coverage. The Global Precipitation Measurement (GPM) mission is dedicated to measuring precipitation from space on a global scale. It is the most advanced precipitation mission in space we have today (Kidd, Takayabu, et al., 2020). The primary satellite in the GPM constellation carries a precipitation radar, allowing more direct interaction with precipitation particles. This satellite is in a low Earth orbit, and it observes only a swath of the Earth. Then, as the Earth rotates, it successively covers most of the Earth's surface. However, the time between observations for a single location is too long for the measurements to be useful for global nowcasting. To solve this problem, we aspire to make use of geostationary satellites. Illustrated in figure 1.1 together with a low Earth orbit, geostationary satellites are fixed in position relative to the Earth at a high altitude. With these, observations with an extensive spatial coverage and at a high temporal frequency can be obtained. The drawback is that they cannot carry radars due to physical constraints. Instead, they are equipped with instruments that passively sense radiation in the visible and infrared spectrum. Since no one-to-one mapping exists between such observations and precipitation, measuring rain with these instruments becomes much less direct.

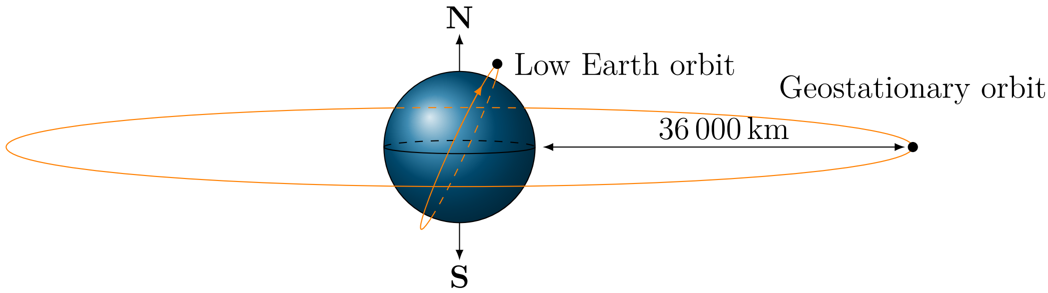

Figure 1.1: Illustration of geostationary and low Earth orbits. Black dots represent satellites. The satellite in the low Earth orbit moves in its orbit relative to the Earth, while the geostationary satellite is fixed in position relative to the Earth. Satellites in low Earth orbits have an altitude of less than 2000 km, while geostationary satellites orbit at an altitude of 36 000 km. Figure is not to scale.

In this project, we will address this problem by applying a machine learning method to retrieve precipitation rates over Africa from a geostationary satellite. With supervised machine learning, a network will be trained on input data consisting of geostationary infrared images from Meteosat-11. Precipitation measurements from

GPM will be used as ground truth. Since precipitation is highly variable, retrieval outputs should be probabilistic in order to be useful for interpretation (Palmer and Räisänen, 2002). With quantile regression, the posterior probability distribution is approximated, and a neural network can be adapted to perform quantile regression. We call such networks Quantile Regression Neural Networks (QRNNs), and with these, more informative retrievals in the form of probability distributions are obtained.

1.3 Previous and related works

Despite the difficulty of measuring precipitation using observations from geostationary satellites, these data have been available since the late 70s, incentivising their use. Different methods for deriving precipitation rates from infrared or visible radiation have been developed since then. Arkin and Meisner (1987) used a relatively simple threshold technique to discriminate between precipitating and non-precipitating pixels in the satellite image. Scofield (1987) and Adler and Negri (1988) both derived precipitation based on inferred cloud types. Scofield (1987) estimated convective rainfall by looking at the temporal changes in two consecutive observations, and Adler and Negri (1988) estimated both convective and stratiform precipitation by empirically eliminating non-precipitating clouds and assigning rain rates based on the sensed radiation. The screening of non-precipitating clouds in Adler and Negri (1988) was further developed in M. B. Ba and Gruber (2001), which improved upon Arkin and Meisner (1987). Because of the indirect relationship, these mentioned works rely on empirical correlations with reference data. As neural networks are efficient at modelling complex non-linear relationships, the data could perhaps be better correlated using a machine learning approach.

Applying machine learning methods to interpret satellite data is a well-motivated measure as satellite data are often plentiful, and data-driven machine learning methods improve with increased data. In the case of precipitation, machine learning methods have previously been applied to retrieve precipitation from geostationary satellites. An early and prominent example is the PERSIANN algorithm published in 1997 (Hsu et al., 1997). Its family has since evolved to improved products such as PERSIANN-CSS (Hong et al., 2004), which is now being replaced by PDIR-Now (Nguyen et al., 2020). Geostationary satellites have also been used in Lebedev et al. (2019), where a machine learning approach using a U-Net architecture was applied. Finally, a probabilistic approach to precipitation nowcasting using machine learning was employed in Ravuri et al. (2021), although here, a different task since weather radar data was used.

The intent of QRNNs is to retrieve informative probabilistic outputs with included uncertainties. Previous studies related to QRNNs applied to retrieve atmospheric variables include Ingemarsson (2021), Amell (2021), and Pfreundschuh, Ingemarsson, et al. (2022). Ingemarsson (2021) applied QRNNs to retrieve precipitation from a geostationary satellite over Brazil. It was investigated whether a convolutional neural network (CNN) or a multi-layer perception (MLP) would be more suitable for the task, and the CNN was found to outperform the MLP. The CNN was further developed in Pfreundschuh, Ingemarsson, et al. (2022), and slight changes were made to the network. Lastly, the work of Amell (2021) applied QRNNs to retrieve ice water path in clouds from a geostationary satellite. Here, the advantage of a CNN, as opposed to an MLP, was also suggested.

1.4 Aims and scope

This project aims to apply the methodologies of previous studies related to QRNNs to retrieve precipitation over Africa for the purpose of nowcasting. The project is conducted by training a machine learning model with input data from the Spinning Enhanced Visible and InfraRed Imager (SEVIRI, Schmid, 2000), the main instrument onboard Meteosat-11. For ground truth, GPM data is used. Further, we seek to explore the importance of input data. Two set-ups are therefore compared, one using fewer channels. This is interesting partly because of the reduced input size, but perhaps more motivating are the implications for the potential use of historical data. The considered channels are limited to wavelengths that do not have a solar contribution, and this is to avoid differences in day and night observations.

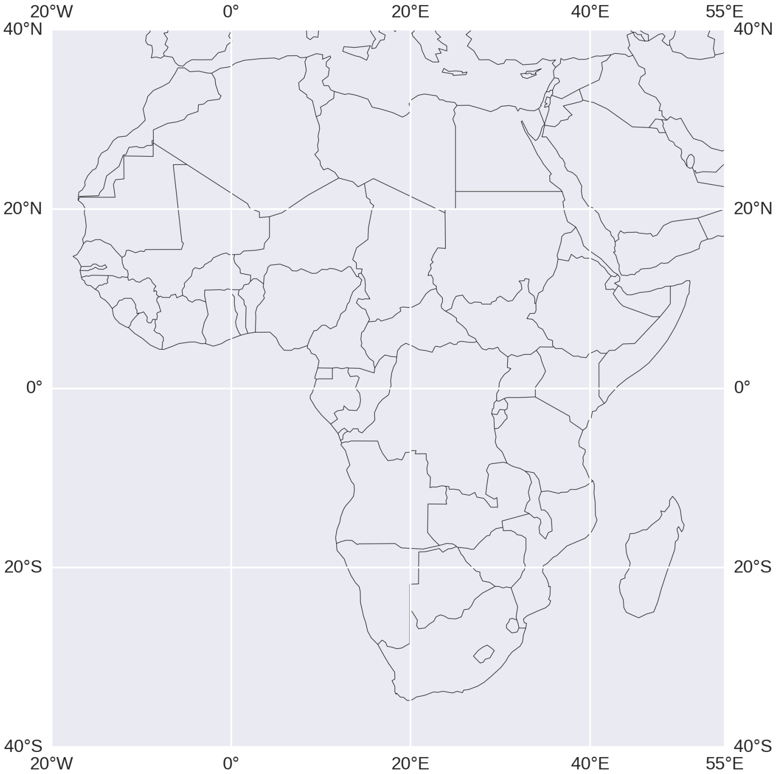

Figure 1.2: Defined region of interest for the project. The region coordinates range from 40°N–40°S, 20°W–55°E and encompass the continent of Africa including Madagascar.

The work is carried out by closely building on previous studies. Since the main objective is to apply suggested methods, not much focus is on exploring different network architectures. For the trainings, we define our region of interest as 40°N–40°S, 20°W–55°E. These coordinates encompass the continent of Africa, including Madagascar and are illustrated in figure 1.2. Rain gauge data were not possible to acquire for the region of interest in the scope of this work. Assessment of the model is then limited to evaluations on the test set and comparisons with a leading global precipitation product.

2

Data

As of today, an abundance of freely accessible satellite datasets and products exist. In this project, we wish to make use of this fact by training a neural network with data from geostationary and low Earth orbit satellites. The data sources, processing and properties of our produced dataset are described in this chapter.

2.1 Data sources

The input data to train our model consists of data from infrared sensors onboard the geostationary Meteosat-11 satellite. As ground truth, data from spaceborne passive and active microwave sensors from GPM are used. The retrievals are compared with IMERG, a leading global near real-time precipitation product.

2.1.1 Input data: Meteosat-11 SEVIRI

Meteosat-11 from the EUMETSAT Meteosat series is a geostationary satellite located above 0° latitude and longitude. The satellite is the fourth and last of the Meteosat Second Generation (MSG) satellites (Schmetz et al., 2002), also referred to as MSG-4. It started operating from its position on 20 February 2018 and has since provided images of the Earth in the visible and infrared spectrum. From Meteosat-11, we can obtain a full disc view of the Earth, encompassing both Europe and Africa, every 15 minutes.

The primary instrument onboard is the Spinning Enhanced Visible and InfraRed Imager (SEVIRI), equipped with 12 spectral channels (Schmid, 2000). The channels of the SEVIRI instrument were designed to estimate atmospheric properties from which variables such as precipitation can be derived (Levizzani et al., 2001). In this project, we limit the channels of interest to 7 out of the 12 available ones as explained in section 1.4. The considered channels measure the intensity of thermal emission from Earth and its atmosphere, expressed in brightness temperature. Table 2.1 presents the relevant SEVIRI channels of this project. Here, the channels are assigned to a channel number which they will be referred to in this work. As seen in the table, 'window' in the absorption band column refers to an atmospheric window, meaning that these wavelengths can pass through the Earth's atmosphere.

Table 2.1: SEVIRI channels used in this project.

Channel number	SEVIRI channel name	Nominal centre wavelength (μm)	Absorption band
5	WV6.2	6.25	Water vapour
6	WV7.3	7.35	Water vapour
7	IR8.7	8.70	Window
8	IR9.7	9.66	Ozone
9	IR10.8	10.80	Window
10	IR12.0	12.00	Window
11	IR13.4	13.40	Carbon dioxide

While channels 5 and 6 in the presented table can be used to estimate water vapour in the atmosphere, the remaining channels range in wavelengths typically used to assess the temperature of the Earth's surface and atmospheric elements (Cermak, Trigo, and Fuchs, 2021). Because liquid water and ice absorb all of these emitted wavelengths, clouds in the satellite's view will block Earth's emitted radiation and emit their own radiation (Seman, 2020). Thus, the considered channels are left to sense cloud top temperatures. For future reference, the channels in table 2.1 will be referred to as all thermal infrared (TIR) channels[1].

The TIR channels have a spatial resolution of 3 km × 3 km, which progressively becomes lower further out from the nadir point and degrades to at most 6 km × 6 km for our region of interest (EUMETSAT, 2017). In order to account for the difference in resolution, the satellite zenith angle is computed for each retrieval. It is then fed as input to the network together with the channel data. The zenith angle is computed for each retrieval because there are minor oscillations in the position of the satellite so that it can vary slightly for each retrieval.

Level 1.5 image data (EUMETSAT, 2022b) of the scans can be downloaded through the EUTMETSAT Data Store (EUMETSAT, 2022a). These data are given in binary files and can be read using the `satpy` Python library (Raspaud et al., 2022). The data are gridded in their native geostationary projection. Pixels toward the grid's boundaries are pixels where the SEVIRI instrument has scanned space, and these, together with any pixels with missing information, are masked out (EUMETSAT, 2017). An example image retrieved from channel 5 of the SEVIRI instrument is shown in figure 2.1, together with the ground truth data presented in the upcoming section. We note to the reader that any future illustrations of brightness temperatures in this thesis are normalised to $[-1, 1]$ when a colour bar is not presented.

---

[1] Although the 3.9 µm channel of the SEVIRI instrument can be considered a TIR channel, this wavelength is also emitted by the Sun. The channel is then excluded because of the solar contribution.

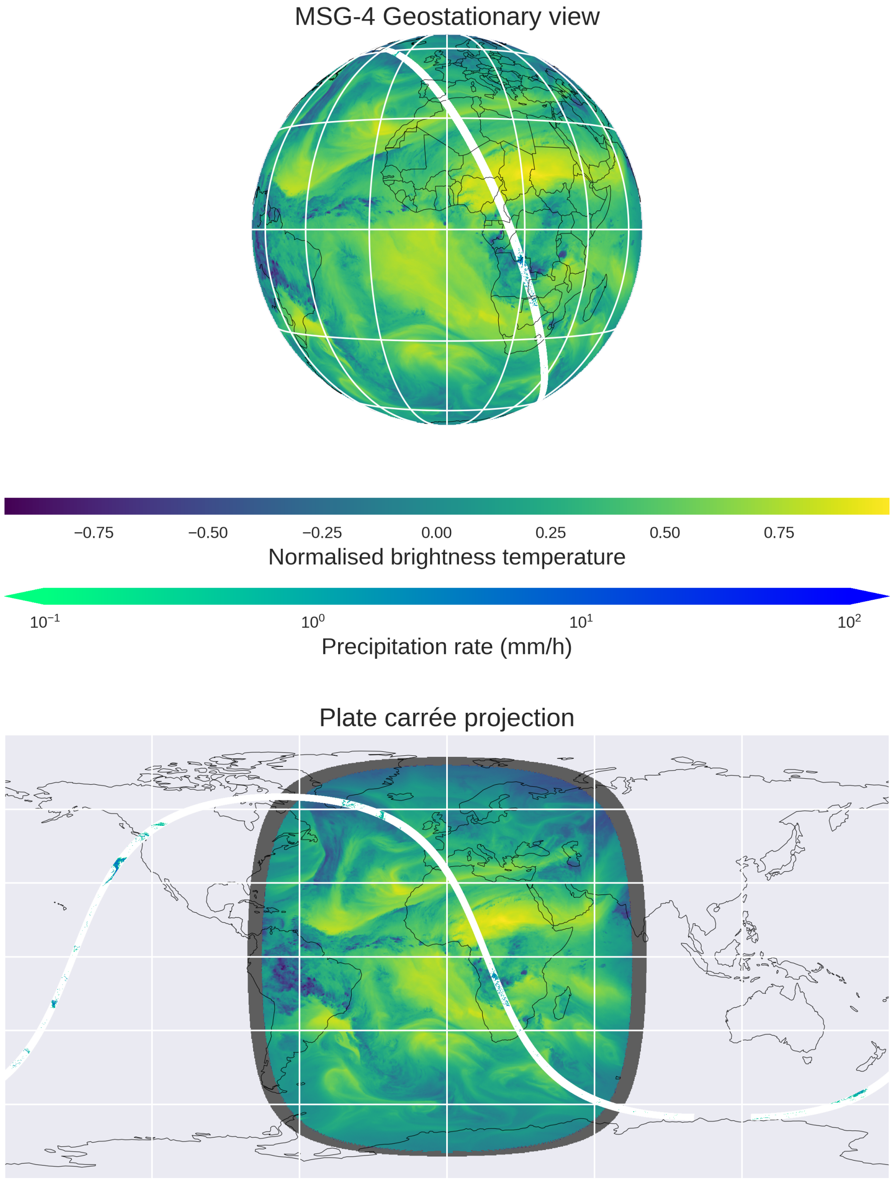

Figure 2.1: Brightness temperature as sensed from the water vapour channel 5 of the SEVIRI instrument overlaid by a swath from the GPM combined product. Darker values in the channel 5 data indicate radiation absorbed by water vapour in the atmosphere and thus the presence of water vapour. Invalid 'space pixels' are visible on the plate carrée projection, coloured in dark grey. GPM swath is visible in white where no precipitation is detected and graduates to blue as precipitation rates are measured. At the top are the observations projected on the geostationary view of MSG-4, and the same observations on the plate carrée projection are illustrated at the bottom.

2.1.2 Ground truth data: GPM combined product

The Global Precipitation Measurement (GPM) is a joint mission between international actors with the purpose of measuring precipitation globally using satellites (Hou et al., 2014). The mission comprises a constellation of satellites, and the primary satellite, the GPM Core Observatory (GPM-CO), was developed in collaboration between NASA and JAXA. This satellite was launched on 28 February 2014 and operates in low Earth orbit at an altitude of 407 km and with a 65° inclination from the equator. The orbit takes approximately 1.5 hours to complete, and its swath successively covers about 91% of the Earth's surface (Kidd, Takayabu, et al., 2020). Furthermore, due to its orbiting nature, it takes, on average, roughly 2.5 days for the satellite to revisit the same location on Earth (Kirschbaum, n.d).

It carries both a passive and an active microwave sensor designed to detect precipitation. These microwave sensors give more direct information about precipitation as these wavelengths can penetrate clouds. The passive sensor onboard is the GPM Microwave Imager (GMI), comprising 13 channels that sense emitted microwave radiation in the 10–183 GHz range (Draper et al., 2015). The second instrument is the Dual-frequency Precipitation Radar (DPR) which actively emits pulses to interact with matter (Iguchi, 2020). It does so in two frequencies, one in the K_u-band (13.6 GHz) and one in the K_a-band (35.5 GHz). With these frequencies, the two instruments were developed with the intent of measuring precipitation rates ranging in 0.2–111.0 mm/h (Kidd, Takayabu, et al., 2020).

Table 2.2: Summary of the GPM DPR and GMI combined product used in this work. Here we use version 6 and K_uPR refers to the K_u-band of the DPR.

GPM K_uPR and GMI combined product V06	
Swath width	245 km
Spatial coverage	−67° to 67° latitude[1], all longitudes
Temporal coverage	8 March 2014 to 2 December 2021
Spatial horizontal resolution	5 km × 5 km
Sensitivity threshold	0.2 mm/h[2]

[1] Iguchi et al. (2018)

[2] Tan et al. (2018)

The DPR interacts more directly with precipitation while the GMI senses a broader range of wavelengths. In order to benefit from both instruments, the retrievals are combined into one product (Grecu et al., 2016). This combined product holds retrievals of the currently most advanced instruments in space designed to measure precipitation and is thus suitable to use as ground truth data. The combined level 2B product (NASA, 2022b) can be downloaded via NASA's Earthdata Search website (NASA, 2022a), where it is possible to filter according to location and time. Version 6 and the swath matched with the K_u-band of the DPR is used in this project. A summary of the product is presented in table 2.2, where K_uPR refers to the K_u-band of the DPR. An example swath of the combined precipitation product is shown in figure 2.1, together with a SEVIRI retrieval. Here, it is also noted that figures illustrating precipitation in the remainder of the thesis are in the range 10^{-1}–10^{2} mm/h when no colour bar is provided.

2.1.3 Data comparison: IMERG

For comparing our retrievals, there are several potential datasets to choose from. The global dataset CHIRPS (Funk et al., 2015) has shown good performance over complex topography in the mountainous regions of Ethiopia (Geleta and Deressa, 2021). Additionally, there is TAMSAT (Maidment et al., 2017), which was specifically developed for Africa. Both CHIRPS and TAMSAT make use of geostationary thermal infrared data and rain gauge data, while CHIRPS also incorporates satellite microwave retrievals and reanalysis. These two datasets are however intended to use for long-term analyses and have a low temporal resolution and high latency compared to the retrievals of this work.

A precipitation product closer in temporal resolution is Integrated Multi-satellitE Retrievals for GPM (IMERG, Huffman, Bolvin, Braithwaite, et al., 2020). IMERG is a global near real-time precipitation product developed by the US GPM Science Team. The algorithm used to produce this dataset integrates data from numerous sources such as DPR and GMI retrievals, geostationary infrared sensors, various spaceborne passive microwave instruments and rain gauge analyses (Huffman, Bolvin, Nelkin, et al., 2020). The dataset has a spatial resolution of $10\,\text{km} \times 10\,\text{km}$ and temporal resolution of 30 minutes and is considered one of the most reliable global precipitation products derived from satellites (Beck et al., 2019).

IMERG comes in three different runs; early, late and final with a minimum latency of 4 hours, 12 hours and 3.5 months, respectively. For the time period considered for comparison, the final run had not been released. Furthermore, as near real-time products are of interest in this project, the run chosen for comparison is the late run, and version 6 of it is used. While the early and late run do not incorporate rain gauge data, the late run has the advantage over the early run that it has both back and forward propagation incorporated in the merging, while the early only has forward propagation. The final run includes any microwave overpasses that were missed when computing the other two runs and also adjusts the estimates according to a monthly gauge data analysis. The final run is considered the most accurate among the three, but because of the reasons stated above, this run was not used.

2.2 Dataset preparations

In order to create the dataset used for training, validating and testing the network model, the GPM swaths need to be spatially and temporally matched with SEVIRI data. The data fed to the network for training will consist of images with a height and width of 256×256 pixels, so data over our region will be cropped into smaller portions. Data are collected over nearly four years, from 20 February 2018 until 30 November 2021, with the former date bounded by the availability of SEVIRI data and the latter bounded by the time availability of the GPM combined product.

2.2.1 Spatial resampling

The considered channels of the SEVIRI instrument have a spatial resolution of 3 km × 3 km at nadir, while the GPM combined product has a spatial resolution of 5 km × 5 km. Furthermore, SEVIRI data are projected on a geostationary grid, where positions are expressed in the shortest distance (metres) away from 0° longitude and latitude, while swaths in the GPM combined product are given in longitude and latitude. In order to correspond input data with ground truth data, the pixels in the SEVIRI observations are matched with GPM precipitation retrievals. Resampling the GPM swath onto the SEVIRI grid can be done with the `satpy` library, and it was done using the nearest neighbour algorithm with a constant radius of influence of 3.6 km. A schematic illustration of the spatial resampling is shown in figure 2.2.

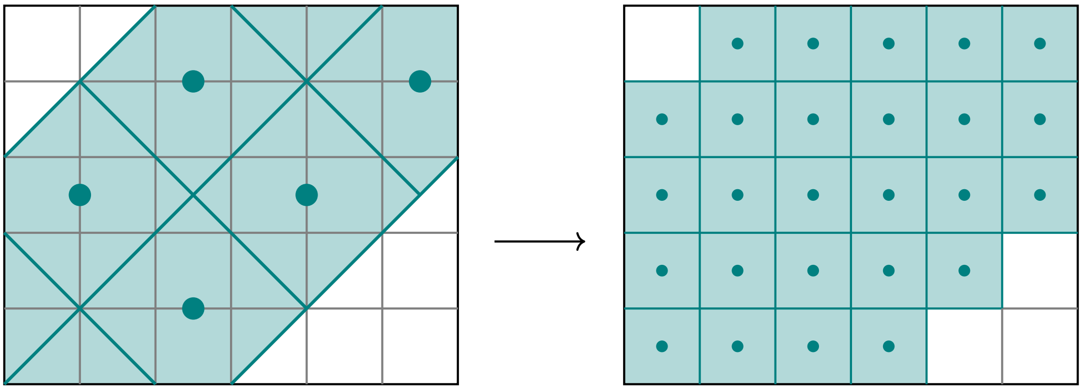

Figure 2.2: Schematic illustration of the nearest neighbour resampling of the GPM swath onto the SEVIRI grid. Figure is not to scale.

2.2.2 Cropping GPM boxes and matching SEVIRI observations

When the GPM data are resampled onto the SEVIRI grid, the input and reference data are provided in two equivalent grids. Each GPM swath that overpasses the region of interest is cropped into nine boxes, each with a 256 × 256 pixels resolution. The cropping of the boxes adheres to the methodology of Ingemarsson (2021) and is done in the following way:

i) Sample a random vertical offset y_{mid}.
ii) Stack nine squares of 256 × 256 pixels vertically, with the mid square centred about y_{mid}.
iii) Centre each square horizontally about the GPM swath if one exists in the same horizontal plane else disregard it.
iv) Save all squares that lie entirely within the region.

An example of cropped boxes is illustrated in figure 2.3. The SEVIRI observations and the GPM combined product are both given in UTC; thus, temporal matching is done more easily. The GPM timestamps in and out of the box are used to find the closest matching SEVIRI retrieval in time. The retrieved SEVIRI data are then cropped to the same box. Any box with SEVIRI data containing NaNs is discarded.

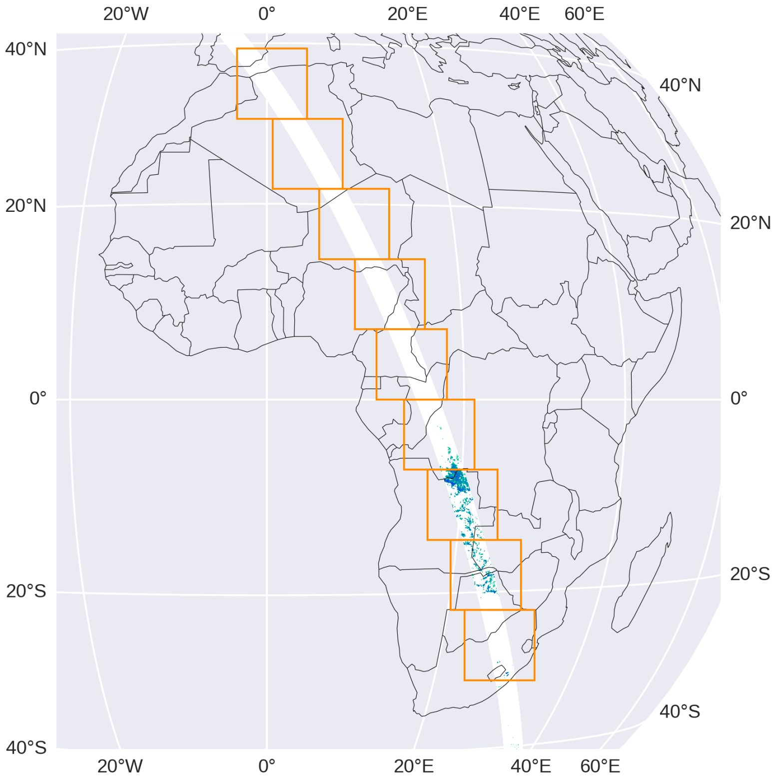

Figure 2.3: Shown is an example of how a GPM swath is cropped into boxes. For each box, a SEVIRI retrieval was matched temporally and cropped onto the same square. Note that the GPM swath is first resampled onto the geostationary SEVIRI grid.

2.3 Dataset properties

The data collocations over the selected time period and region resulted in nearly 50,000 samples. These were, in turn, split randomly into training, validation and test sets with a 60/20/20 ratio, respectively. Properties of the sets are summarised in table 2.3.

Table 2.3: Properties of data subsets.

Set	Train	Validation	Test
Number of samples	29944	9982	9981
Disk space (GB)	67	23	23

The distributions of precipitation rates for the training and validation sets are shown in figure 2.4. Noting the logarithmic scaling of the y-axis, most of the precipitation rates take place in the low range. There was no balancing being done to the datasets since a representation of the actual frequency distribution is relevant information to the network.

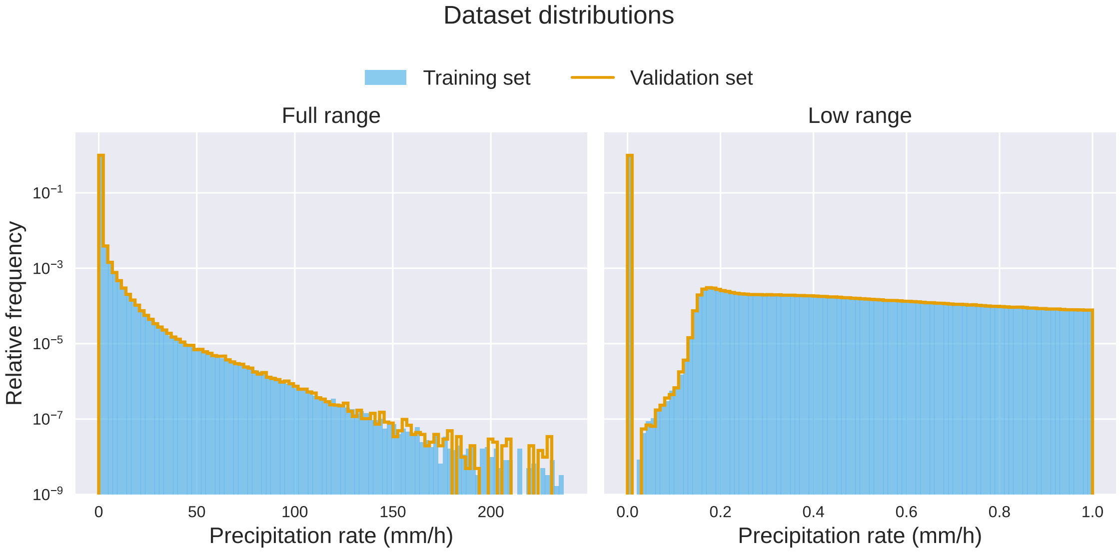

Figure 2.4: Precipitation rate distributions of the training (blue filled) and validation (orange line) sets. The entire range of distributions is shown to the left, while the right figure only displays low range rates. The GPM sensitivity threshold of 0.2 mm/h is visible in the right panel.

Each sample contains 8 input channels of size 256×256 pixels, consisting of the 7 TIR SEVIRI channels and the computed satellite zenith angle. All input channel values are standardised using statistics computed from the training set. The reference input comprised of the GPM swath is also a 256×256 pixel image, where invalid pixels outside the swath are masked. The loss function is only computed for valid pixels in the reference image, consisting of the GPM swath. An example of input channels is shown in figure 2.5 and the corresponding reference is displayed in figure 2.6. From observing the example inputs in figure 2.5, it can be noted how the channels offer rather similar information. The most distinguishable difference appears between the two water vapour channels (5 and 6) and the rest. This project additionally explores the necessity of using all TIR channels; as a set-up, only using two channels (5 and 9) is evaluated. This is further explained in chapter 3.

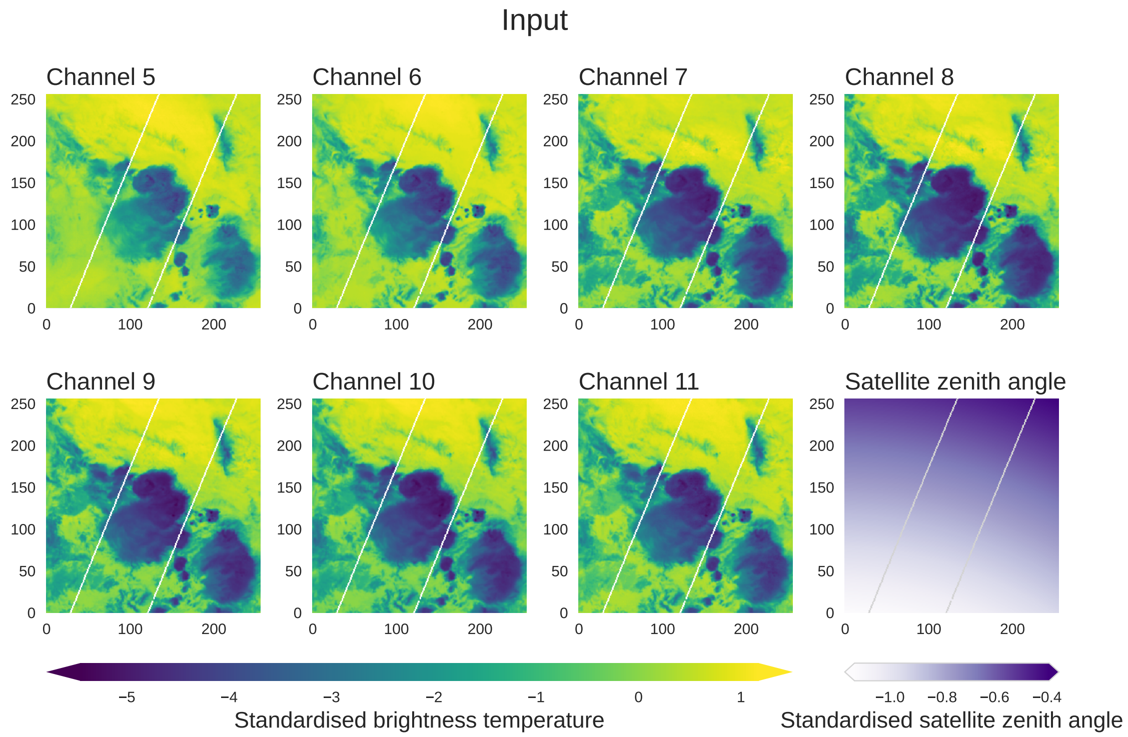

Figure 2.5: Example of input channels fed to the network model. All input values are standardised. Images here are overlaid with lines of the GPM swath boundary for illustration purposes.

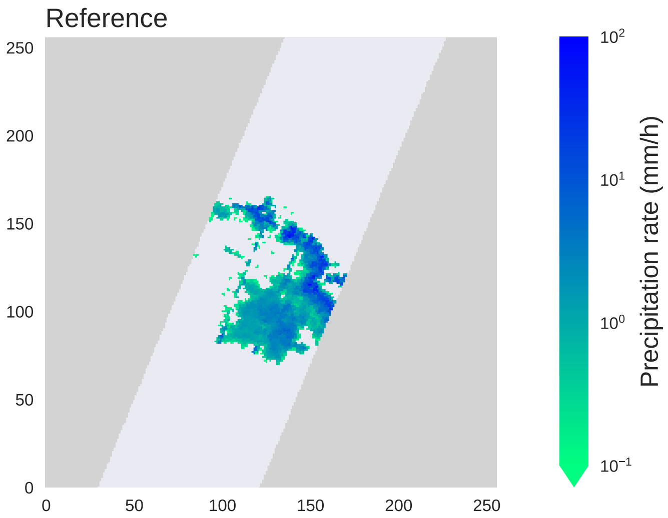

Figure 2.6: The corresponding reference image for inputs shown in figure 2.5. Masked out area is in dark grey, and parts of the swath in light grey are pixels where no precipitation was detected.

3

Machine learning

The challenge with our obtained data is that there is no direct relation between the geostationary SEVIRI data and the combined precipitation product from GPM. Therefore, we approach the problem using a machine learning method with Quantile Regression Neural Networks. In this chapter, we first motivate the choice of methodology by describing the underlying problem. This is followed by introducing the basic principles behind the method to illustrate how the problem is approached. Lastly, it is described how an assessment of our machine learning retrievals is made.

3.1 The retrieval problem

Deriving atmospheric parameters from remote sensing observations often involves indirect measurements of the atmospheric properties of interest. In this project, we sought to determine precipitation rates from measured thermal IR radiation, a retrieval problem that can be regarded as an inverse problem. If we define the forward function f as the function that characterises the physical relationship between surface precipitation rates y and our observed thermal IR radiation x, we have

$$x = f(y) + \epsilon,$$

where ϵ is a random error arising because of the fact that all real measurements include some error (Rodgers, 1997). Here, we are interested in the quantity y, which can be obtained by inverting the forward function f, and we are thus dealing with an inverse problem. A difficulty with inverse problems is that the forward function is generally not invertible. Furthermore, well-posed problems are characterised by the following properties;

i) a solution to the problem exists,

ii) the solution is unique,

iii) a small pertubation in the initial conditions results in a small pertubation in the solution's behaviour.

An ill-posed problem, on the other hand, is any problem for which at least one of the stated properties is absent. Our problem at hand cannot be uniquely determined because of the random noise ϵ present in measurements, and so it is an ill-posed inverse problem. In order to tackle ill-posed inverse problems, it is common to adopt a Bayesian approach. Instead of determining an exact solution, we treat the state y and the observations x as random variables and assess the conditional probability density function $p(y|x)$. In the Bayesian framework, prior knowledge is combined with observations to determine $p(y|x)$ through Bayes' theorem

$$p(y|x) = \frac{p(x|y)p(x)}{p(y)}. \tag{3.1}$$

Bayesian retrieval methods typically approximate $p(y|x)$ by dealing with the right-hand side of equation 3.1. A drawback of these methods is that they are often computationally expensive, and it is therefore of interest to explore alternatives. In Pfreundschuh, Eriksson, et al. (2018), a machine learning method with so-called Quantile Regression Neural Networks (QRNNs) was used to directly approximate $p(y|x)$. This was shown to yield consistently satisfactory results compared with traditional Bayesian methods, with the added benefit of computational efficiency and flexibility. Consequently, this is the approach taken for this project.

3.2 Quantile regression

The principle behind QRNNs is quantile regression, where the quantiles of a distribution function are found by minimising a quantile loss. We first recall that for any real-valued random variable X with probability density function (PDF) $p(x)$, we denote its cumulative distribution function (CDF) by $F(x)$ and

$$p(x) = \frac{dF(x)}{dx}. \tag{3.2}$$

Using $F(x)$, we define the τth quantile of X as the value x_τ such that

$$x_\tau = \inf\{x : F(x) \geq \tau\}, \qquad \forall \tau \in (0, 1).$$

In other words, the τth quantile is the smallest value x for which $F(x)$ is at least τ. Koenker (2005) shows that the τth quantile minimises the expectation with respect to x of the function

$$\mathcal{L}_\tau(x_\tau, x) = \begin{cases} \tau|x - x_\tau|, & x_\tau < x \\ (1-\tau)|x - x_\tau|, & x_\tau \geq x \end{cases} \tag{3.3}$$

and we refer to this function as the quantile loss. If $\mathcal{L}_\tau$ is minimised for many values of τ, we then obtain an approximation of the distribution function from which we can reconstruct the probability density function via equation (3.2).

3.3 Supervised machine learning

Artificial neural networks are sets of connected computational units. They have gained much attention due to their adeptness at modelling complex non-linear relationships. In supervised machine learning with artificial neural networks, the network is provided input and reference data. The input data propagates through the computational units, and the output is subsequently compared against the reference data. The network is trained to map input data to reference data by adjusting its internal parameters to minimise some defined loss between the learnt output and the reference data. Moreover, if the loss function is chosen to be the quantile loss function defined in equation (3.3), the network can solve our retrieval problem by learning multiple quantiles of the CDF of precipitation rates given the emitted IR radiation. Thus we do not only retrieve a scalar quantity but numerous quantiles from which we approximate the CDF.

Solving regression problems in satellite data with machine learning is a relatively recent practice, and it is not generally widespread. Because of the limited exploration, a standard network model has yet to be established. Entering this new field of satellite data, it is then sensible to draw inspiration from approaches taken to similar problems. For image data, it is common to consider multi-layer perceptrons (MLPs), convolutional neural networks (CNNs) or a combination of the two. In Ingemarsson (2021), it was shown that a CNN outperformed an MLP in retrieving precipitation rates from geostationary data over Brazil. The same was seen in Amell (2021) for a related task. Since we have intricate cloud structures in our input data, it is reasonable to consider CNNs as they are advantageous at capturing spatial features. While it is possible to further explore network architectures suitable for this task, it was settled for a CNN in this project.

3.3.1 Convolutional neural networks

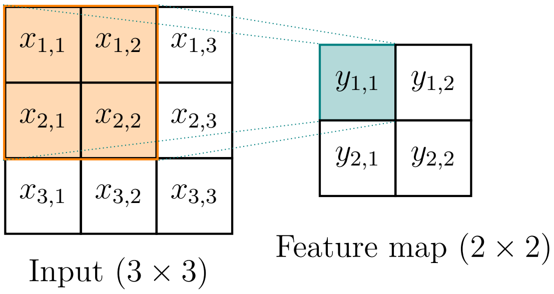

Figure 3.1: Input and feature map of a convolution. $x_{i,j}$ and $y_{i,j}$ are elements of the input and feature map respectively. The kernel is highlighted in orange, and the working convolution produces the element in the feature map highlighted in blue.

Convolutional neural networks use convolutional layers as building blocks and were designed for tasks such as image classification and object detection (Lecun et al., 1998). Convolutional layers take tensors as input, typically images with two spatial dimensions (height and width) and a depth reflecting the number of channels. The input elements are convolved with a kernel which produces a feature map. Each element in the feature map represent the convolution obtained from a specific region in the input image, and this region is also called the local receptive field. While the size of the feature map is partly determined by the size of the kernel, it is also influenced by two hyperparameters; padding and stride. The padding adds elements surrounding the border of the input, and the stride determines the step size used when sliding the kernel along the spatial dimensions of the input. An example of a simple convolution using a two-dimensional kernel is illustrated in figure 3.1. In the presented example, the input layer is a 3 × 3 matrix convolved with a 2 × 2 kernel, highlighted in orange. With a stride of 1 and no padding, the convolution produces a 2 × 2 feature map.

The neuron outputs $y_{i,j}$ of the example in figure 3.1 are computed as

$$y_{i,j} = g\left(\sum_{p=1}^{2}\sum_{q=1}^{2} w_{pq}x_{p+i-1,q+j-1} - b\right),$$

where g is some non-linear activation function and weights w_{pq} and bias b are learnable parameters of the kernel.

Whether it be the input data or between layers, the input to a convolutional layer often includes a depth dimension. Employing a kernel with a depth increases the number of learnable parameters, which can take longer to train and can additionally lead to overfitting. A depthwise convolution is done by convolving each channel separately, resulting in one feature map for each input channel. If this is done using one kernel for each input channel, the number of internal parameters is kept low. An illustration of a depthwise convolution with a stride of 1 and no padding is shown in figure 3.2. Furthermore, when producing many feature maps, it is sometimes desired to reduce their dimensionality. This can be done with a pointwise convolution, which is a convolution with a kernel of size $1 \times 1 \times C$, where C is the number of channels or depth of the input. In figure 3.3, a pointwise convolution with a stride of 1 and no padding is shown.

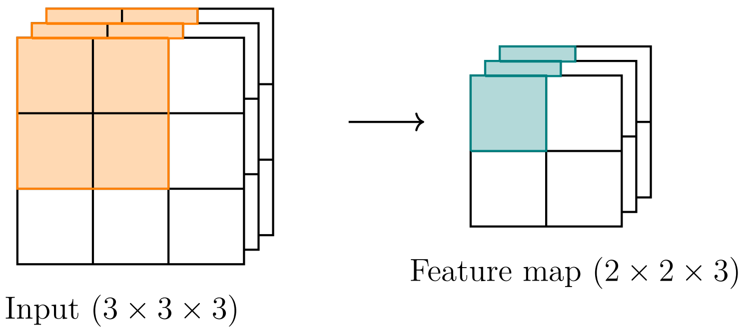

Figure 3.2: Depthwise convolution on a 3 × 3 × 3 input with one 2 × 2 kernel for each channel. The convolution results in three feature maps of size 2 × 2.

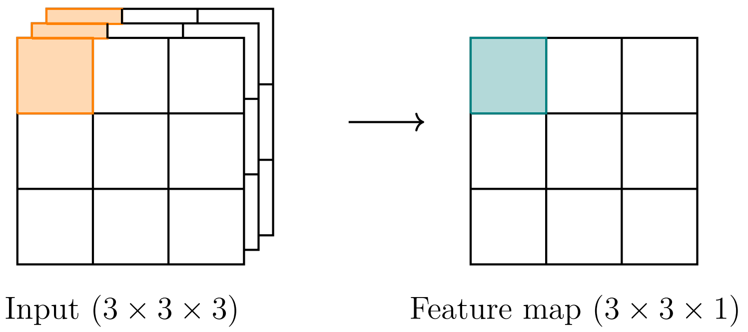

Figure 3.3: Pointwise convolution on a 3×3 input with one $1 \times 1 \times 3$ kernel. The resulting feature map is of size $3 \times 3 \times 1$.

3.4 Network model architecture

The decision of a network model was made from what was known to have previously worked for the implementation over Brazil in Ingemarsson (2021) and Pfreundschuh, Ingemarsson, et al. (2022), which resulted in a fully convolutional network. The architecture consists of a downsampling and upsampling path with long skip connections between stages, similar to the U-Net architecture described in Ronneberger, Fischer, and Brox (2015). With the downsampling path, the model learns to extract features at multiple scales and with the upsampling path, it localises the features in the input. The feature extraction and localisation are done with modules containing Xception blocks (Chollet, 2017). These blocks perform efficient convolutions by assuming that depth correlations are independent of spatial correlations and convolve these dimensions separately. Such convolutions are called depthwise separable convolutions and consist of a depthwise convolution followed by a pointwise convolution. By doing so, the number of learnable parameters is reduced together with the complexity of the model.

The architecture summarised in figure 3.4 was implemented using the `quantnn` library (Pfreundschuh, 2022b) and the network model is available from Pfreundschuh (2022a). The figure shows that the input goes through an asymmetric downsampling and upsampling path. The path is asymmetric in the sense that the downsampling blocks (DXception) are followed by n Xception blocks while the upsampling stages are not. Our model used $n = 2$ Xception blocks after each downsampling block. The data is normalised between layers with layer normalisation (LN, J. L. Ba, Kiros, and Hinton, 2016). Unlike the otherwise commonly used batch normalisation (Ioffe and Szegedy, 2015), layer normalisation is independent of batch size. This is suitable for our task as we have large amounts of data, and our batch size is restricted. The activation function used for each output is the Gaussian Error Linear Units (GELU) function (Hendrycks and Gimpel, 2016), which is similar to the basic ReLU function (Glorot, Bordes, and Bengio, 2011) but with the added benefit of some regularisation effects. For each convolutional layer, we use 128 kernels, except for the very last pointwise convolution where $M = 99$ is the number of quantiles to predict. MaxPool in the figure refers to a pooling layer which downsamples the input using the maximum value in the local receptive field.

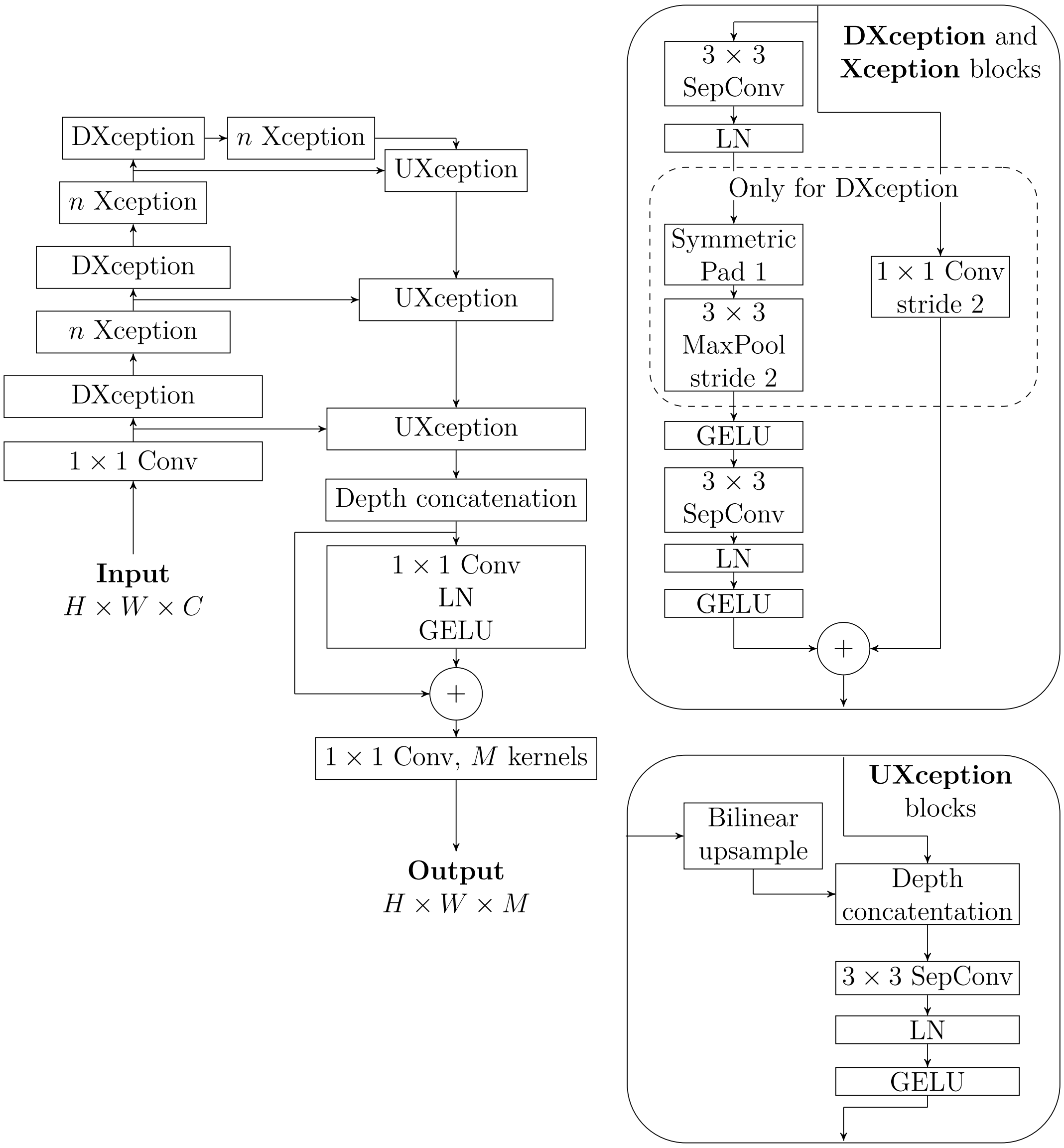

Figure 3.4: Network architecture consists of a downsampling path followed by an upsampling path, shown to the left. The blocks in the paths consist of Xception blocks, where DXception is used for downsampling and UXception for upsampling. Each DXception block is followed by $n = 2$ Xception blocks. All convolutional layers have 128 kernels, except for the output layer, which has $M = 99$ kernels.

3.5 Evaluation metrics

Our network minimises the quantile loss for multiple quantiles, and we thus obtain multiple outputs for each pixel over the region. In contrast, our reference precipitation rate is a single scalar value for each pixel of the GPM swath. Because the loss can be difficult to interpret, additional performance metrics are used to better understand the trends during training and evaluation on validation and test sets.

Three common regression metrics are bias, mean absolute error (MAE) and mean squared error (MSE) defined below

$$
\begin{aligned}
\mathrm{Bias}(\hat{x}, x) &= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i),\\
\mathrm{MAE}(\hat{x}, x) &= \frac{1}{N}\sum_{i=1}^{N}|\hat{x}_i - x_i|,\\
\mathrm{MSE}(\hat{x}, x) &= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2.
\end{aligned}
$$

These metrics require a scalar-valued estimate $\hat{x}$ for comparison with the reference value x. The predicted posterior mean will be used for computing these metrics. The bias gives information about the average under- and over-estimation of rain rates, while the MAE and MSE are the average absolute and squared differences, respectively.

Though the posterior mean can be used to represent the predicted distribution function, assessing how the model performs on each quantile level is also of interest. To do this, we use the continuous ranked probability score (CRPS), defined as

$$
\mathrm{CRPS}(\hat{F}, F) = \int_{-\infty}^{\infty}\left[\hat{F}(x) - F(x)\right]^2 dx,
$$

where $\hat{F}$ is the estimated cumulative distribution function and F is the true distribution. Since the actual distribution is not known and we only have point values for our reference, we will assume F to be the indicator function $\mathbb{1}_{x\geq y}(x) = \{1 \text{ if } x \geq y, 0 \text{ otherwise}\}$, where y is the reference value. While this is not the true distribution, it poses a good assumption for our metric. This is because we ideally would want the predicted probability distribution to have a small uncertainty (sharp) and for the probability to be concentrated around a value close to the reference (well-calibrated). These features are found in the indicator function. The CRPS measures the difference between our estimated distribution function and the ideal. All metrics used are interpreted as satisfactory for values close to zero.

3.6 Training set-up

Hyperparameters and training methods were decided through an iterative process of trial and error. The initial trial was based on the set-up in Ingemarsson (2021). It was then progressively adjusted and changed until an adequate result on the validation set was obtained using all TIR channels. The resulting set-up is a combination of settings and techniques from previous works relating to similar tasks (Ingemarsson, 2021; Amell, 2021; Pfreundschuh, Ingemarsson, et al., 2022). A summary of final hyperparameters used is found in table 3.1. In this table, the Adam optimiser refers to the algorithm described in Kingma and J. Ba (2014).

Table 3.1: Summary of chosen settings and hyperparameters used for training.

Training specifications	
Optimiser	Adam
Scheduler	None
Initial learning rate	10^{-3}
Batch size	64
Quantiles τ	$0.01, 0.02, ..., 0.99$

With the finalised parameters, the same training was done for one set-up using all TIR channels and one only using channels 5 and 9. The two channels were chosen based on instruments of earlier generation Meteosat satellites. These earlier satellites carried the MVIRI instrument, which had three channels, of which two in the infrared spectrum (Rüthrich et al., 2020). Channels 5 and 9 of the SEVIRI instrument are the closest match in wavelength to the two MVIRI infrared channels, which is why they were chosen for the second set-up. For ease of reference, we will refer to the training using all TIR channels as the main set-up and the second set-up only using channels 5 and 9 as the two-channel set-up.

Furthermore, because the precipitation rate values vary differently on different orders, a log-linear transformation was applied to the rain rates. Zero-valued rates were replaced with values drawn from a log-uniform distribution between 10^{-3}–10^{-2} mm/h, such that they fell between rates lower than the 0.2 mm/h sensitivity threshold of the GPM combined product. The log-linear transformation was subsequently applied as follows

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

At last, data augmentation methods were employed as they were seen to improve generalisation. The input channels were randomly rotated $0°, 90°, 180°$ or $270°$, and also randomly mirrored vertically and horizontally. In addition, the inputs were randomly cropped to 128×128 images, such that the network was training on inputs of size $64 \times 128 \times 128 \times (N + 1)$, where N is the number of SEVIRI channels used and the $+1$ term signifies the addition of the satellite zenith angle.

Both trainings were run on an NVIDIA Tesla V100 32 GB SMX2 GPU. They were each kept running for three weeks as the validation loss was not seen to increase, though both training and validation losses converged within the first few days. Some computed metrics for the trainings can be found in appendix A. The training infrastructure was implemented with the `quantnn` library (Pfreundschuh, 2022b) using a `pytorch` backend (Paszke et al., 2019).

4

Results

Instead of a scalar-valued output, our retrievals estimate a cumulative distribution function for every pixel in a retrieved image. In this chapter, the retrieval outputs are illustrated, and the performance is evaluated on the test set. Finally, the QRNN retrievals are compared to IMERG for three different sub-regions.

4.1 The retrieval output

A prediction with our QRNN on a single sample result in a tensor of size $H \times W \times 99$, where H, W is the height and width of the input, respectively, and we predict 99 quantiles for each pixel. With the 99 predicted quantiles, we can approximate the cumulative distribution function from which we derive the corresponding probability density function. An example of a single-pixel output is shown in figure 4.1. Here, the approximated CDF and the corresponding PDF are plotted. Also shown are the reference value and the posterior mean value.

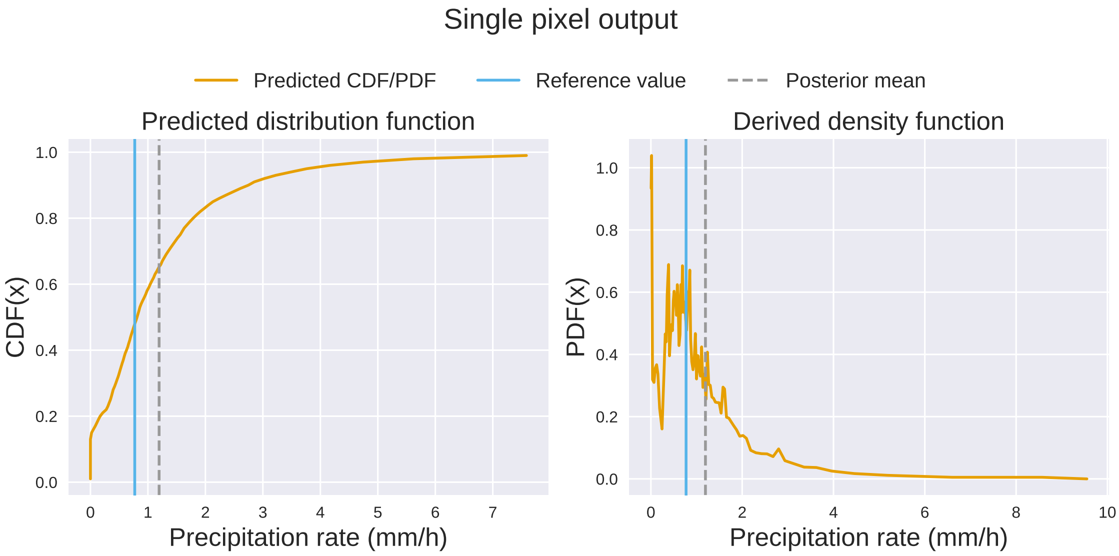

Figure 4.1: QRNN output for a single pixel in a sample. To the left is the CDF approximated by the predicted quantiles. The PDF can be derived from the CDF and is plotted to the right. Vertical lines represent the reference value in solid blue and the posterior mean in dashed grey.

Predictions can be visualised both using the posterior mean and the predicted quantiles. In figure 4.2, an example prediction is illustrated. The outputs are shown for the main set-up, with the reference image in the top panel of the figure. The corresponding inputs can be found in figure 2.5 from chapter 2. In the bottom left figure, the posterior mean for each pixel is plotted, and the bottom right figure shows a contour plot of probabilities for heavy precipitation (>10 mm/h).

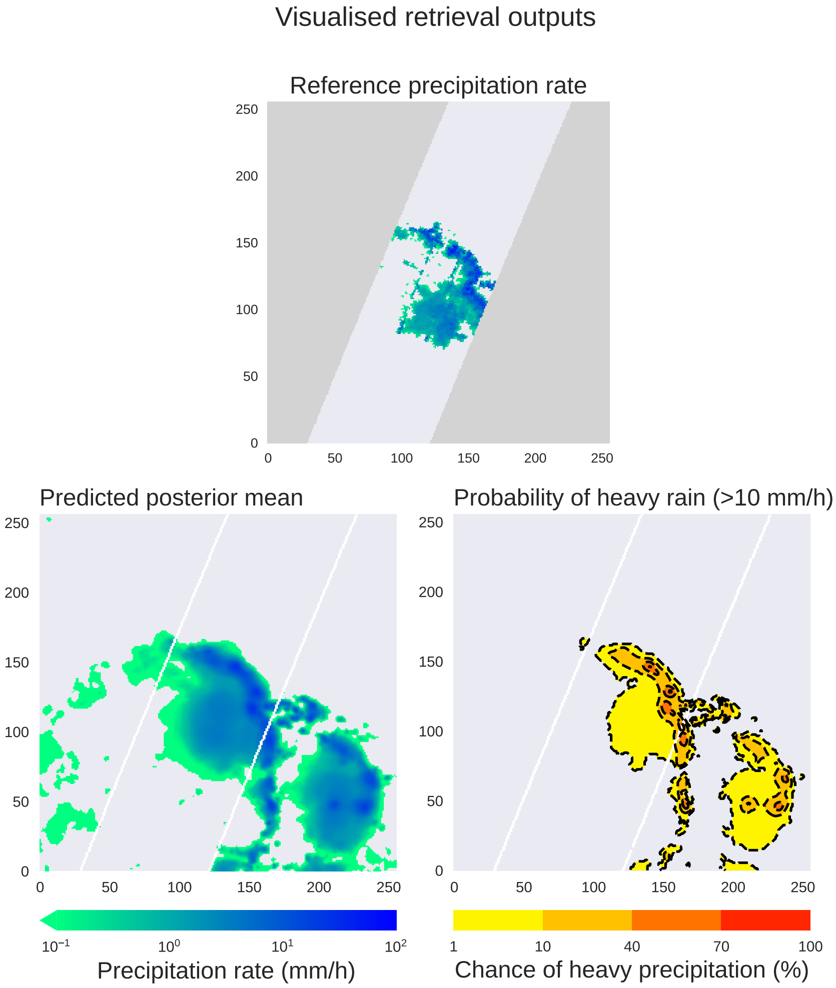

Figure 4.2: Retrieval outputs for a single sample. The top figure is the reference, and the bottom figures show outputs from the main set-up. The figure to the bottom left shows the posterior mean for each pixel. In the bottom right, the probabilities for heavy precipitation rates (>10 mm/h) are shown in a contour plot for different levels. The reference figure and posterior mean output have the same colour bar.

4.2 Evaluation on test set

First presented is a table of evaluation metrics in table 4.1. The test set images have heights and widths of 256×256, and the model was trained on cropped images with 128×128 height and width. To make a rough assessment of whether the inference spatial dimension size has any impact, we also include an evaluation of the main set-up where the test images are split into four quadrants. The table shows that the metrics are generally close to zero and close in value across different set-ups. The set-up with the best overall performance is the main set-up with inference on 256×256 image sizes. From the bias, it can be seen that the posterior mean slightly overestimates low precipitation rates between 0–10^0 mm/h while underestimating the remaining range and more severely for very high rates between 10^1–10^2 mm/h. As can be seen from the table, the CRPS median was reported as 0.0 mm/h for the full and lowest range. An explanation for this is the dominating non-precipitating pixels and the fact that the network detects them successfully. This information is not as apparently reflected in the other metrics as they are averaged quantities.

A more qualitative comparison of predicted posterior means of a single sample for the different set-ups is visualised in figure 4.3. Here, all three set-ups seem to capture the same overall patterns. For the main set-up on 128×128, edge effects where the image had been cropped are visible. This is expected as the CNN has less context around the borders. Apart from this, the predictions are very similar, and the 128×128 evaluation is not further presented in this chapter. More notable differences can be seen between the two-channel and main set-up, and the predictions from the main set-up seem to correspond better to the reference image. For example, the high precipitation rates in the top centre of the images are placed more on the left side of the cloud by the main set-up, while the two-channel set-up places them to the right. Also noted are differences in low precipitation rates in the swath, where the two-channel set-up seem to overestimate these.

We can also visualise the bias in terms of the distribution of errors, found in figure 4.4. The errors range in -200–50 mm/h for the main set-up, while the two-channel has a very slightly longer tail past -200 mm/h. The negative errors arise from an underestimation of precipitation rates. As was noted from the bias, it is also seen here that the posterior mean underestimates high precipitation rates more gravely than it overestimates precipitation rates. The two set-ups are very alike here, only noting minor differences.

The distributions of predicted precipitation rates on the test set are illustrated in figure 4.5. Since the posterior mean has its limitations in representing the full distribution, the sampled posterior is also included here. Instead of using the posterior mean for each pixel, we here sample the approximated CDF. This way, rates that are probable but to varying degrees are better represented. As seen in the figure, the sampled posterior is better at capturing the full distribution for both set-ups. This illustrates how condensing a distribution to a single value such as the posterior mean can be accompanied by a loss of information. The difference between the two set-ups is once more minimal, with their sampled posteriors being very close. Again, both posterior means can be seen to slightly overestimate low precipitation rates and underestimate higher rates more significantly.

Table 4.1: Evaluation metrics on the test set. Bias, MAE and MSE are computed using the posterior mean, while the CRPS is computed using the predicted quantiles. The metrics are computed for different intervals of the ground truth data. About 97% of the data points are in the 0–10^{-1} mm/h range, 1.3% in 10^{-1}–10^{0} mm/h, 1.2% in 10^{0}–10^{1} mm/h and only 0.13% of the data points are in the 10^{1}–10^{2} mm/h range. Bold font marks the value closest to zero in each row.

Metric	Ground truth interval (mm/h)	**Main**		**Two-channel**
		128×128	256×256	256×256
Bias (mm/h)	All values	**$-8.83 \cdot 10^{-4}$**	$-1.58 \cdot 10^{-3}$	$-4.54 \cdot 10^{-3}$
	$0 - 10^{-1}$	$1.84 \cdot 10^{-2}$	**$1.76 \cdot 10^{-2}$**	$2.16 \cdot 10^{-2}$
	$10^{-1} - 10^{0}$	$5.27 \cdot 10^{-1}$	$5.10 \cdot 10^{-1}$	**$4.31 \cdot 10^{-1}$**
	$10^{0} - 10^{1}$	$-5.13 \cdot 10^{-1}$	**$-5.08 \cdot 10^{-1}$**	$-8.80 \cdot 10^{-1}$
	$10^{1} - 10^{2}$	-13.3	**-13.2**	-14.1
MAE (mm/h)	All values	$7.34 \cdot 10^{-2}$	**$7.21 \cdot 10^{-2}$**	$7.66 \cdot 10^{-2}$
	$0 - 10^{-1}$	$1.84 \cdot 10^{-2}$	**$1.77 \cdot 10^{-2}$**	$2.16 \cdot 10^{-2}$
	$10^{-1} - 10^{0}$	$8.36 \cdot 10^{-1}$	$8.15 \cdot 10^{-1}$	**$7.63 \cdot 10^{-1}$**
	$10^{0} - 10^{1}$	2.08	**2.05**	2.07
	$10^{1} - 10^{2}$	13.8	**13.7**	14.6
MSE ($\mathrm{mm}^2/\mathrm{h}^2$)	All values	$7.21 \cdot 10^{-1}$	**$7.10 \cdot 10^{-1}$**	$7.46 \cdot 10^{-1}$
	$0 - 10^{-1}$	$3.63 \cdot 10^{-2}$	$3.35 \cdot 10^{-2}$	**$3.25 \cdot 10^{-2}$**
	$10^{-1} - 10^{0}$	2.55	2.43	**2.01**
	$10^{0} - 10^{1}$	8.93	**8.82**	8.85
	$10^{1} - 10^{2}$	333	**330**	359
CRPS (mm/h)				
mean	All values	$5.26 \cdot 10^{-2}$	**$4.57 \cdot 10^{-2}$**	$6.28 \cdot 10^{-2}$
	$0 - 10^{-1}$	$1.24 \cdot 10^{-2}$	**$1.22 \cdot 10^{-2}$**	$1.77 \cdot 10^{-2}$
	$10^{-1} - 10^{0}$	$4.40 \cdot 10^{-1}$	$4.33 \cdot 10^{-1}$	**$4.18 \cdot 10^{-1}$**
	$10^{0} - 10^{1}$	1.57	**1.56**	1.63
	$10^{1} - 10^{2}$	**11.2**	**11.2**	11.7
median	All values	**0.0**	**0.0**	**0.0**
	$0 - 10^{-1}$	**0.0**	**0.0**	**0.0**
	$10^{-1} - 10^{0}$	$3.07 \cdot 10^{-1}$	**$3.04 \cdot 10^{-1}$**	$3.13 \cdot 10^{-1}$
	$10^{0} - 10^{1}$	1.10	**1.09**	1.13
	$10^{1} - 10^{2}$	8.68	**8.63**	9.13

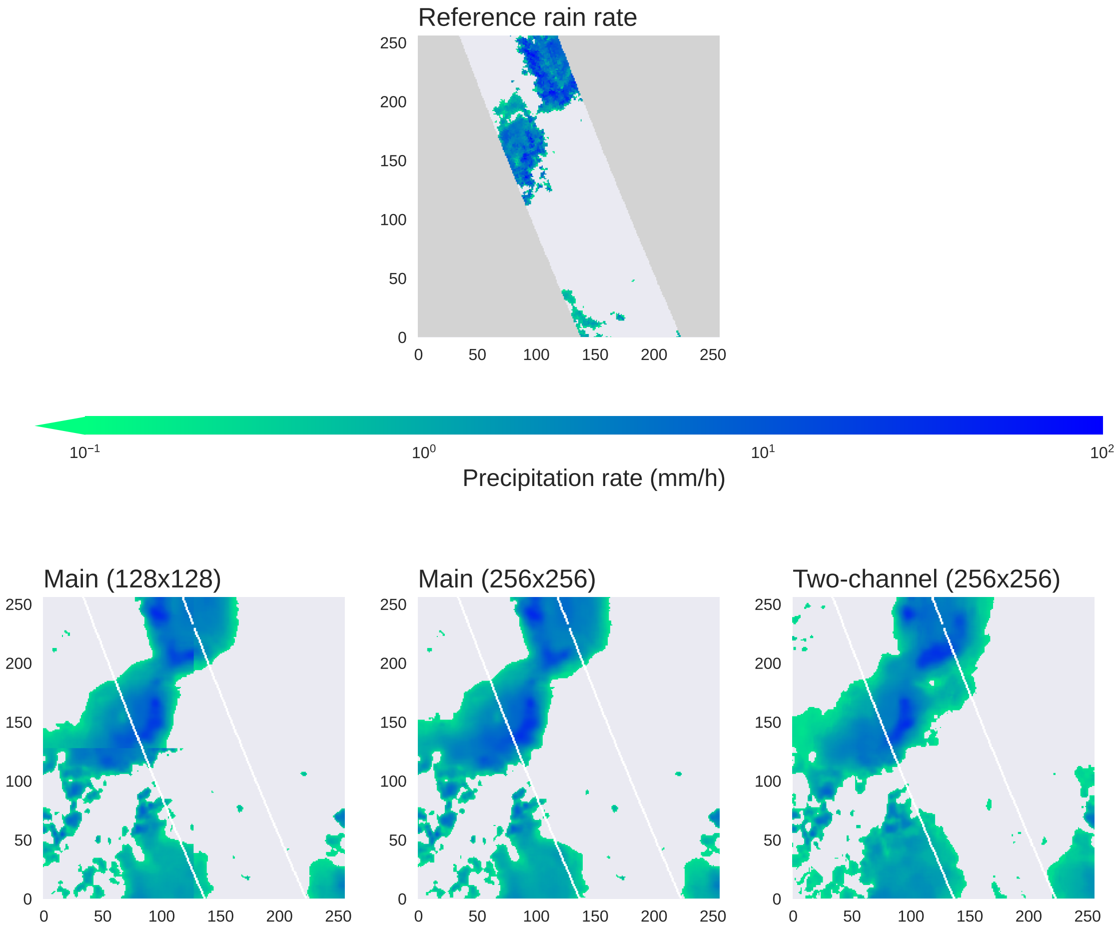

Figure 4.3: Posterior mean outputs on a single test sample for the two set-ups. The top figure shows the corresponding reference, and the bottom figures show predicted posterior mean retrievals. The bottom left figure shows the main set-up when inferring on images with 128×128 height and width, while the centre bottom figure is the same set-up but inferring on the entire 256×256 image. The two-channel set-up is shown in the bottom right figure. All figures share the same colour bar.

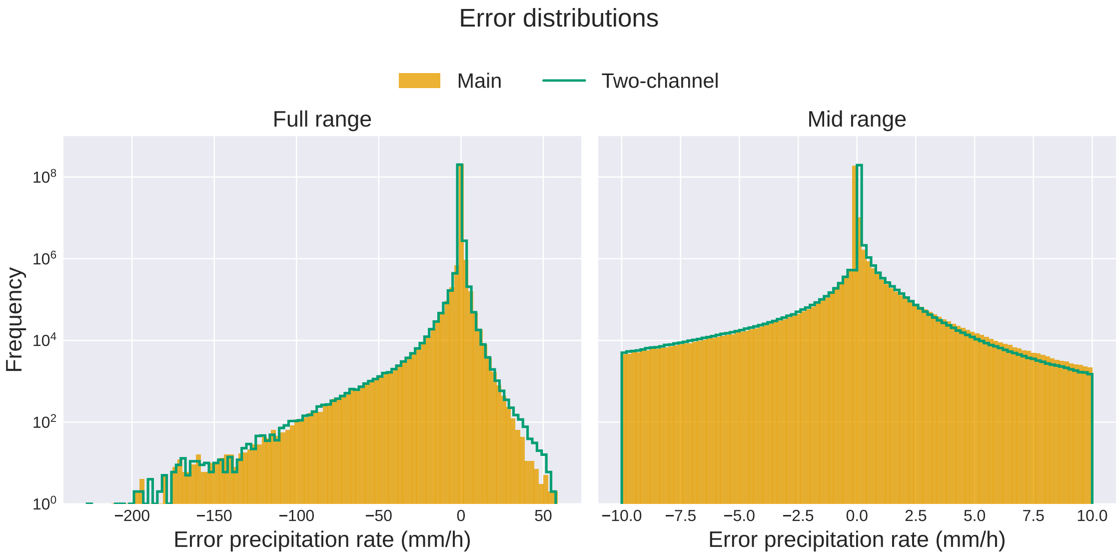

Figure 4.4: Error distribution for the main set-up in filled orange and two-channel in the green line. To the left is the full range, and the right figure is zoomed in on a mid range. Note the logarithmic scaling of the y-axis.

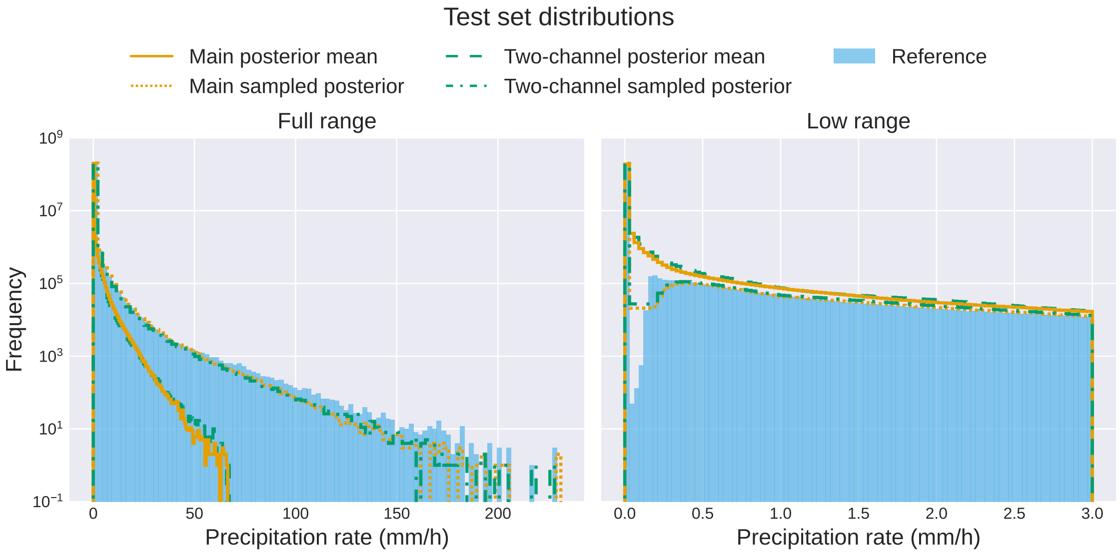

Figure 4.5: Predicted precipitation rates on the test set. The distributions are plotted for the full range to the left and for a lower range in the right figure. Y-axes are logarithmically scaled.

To visualise how the set-ups perform on a pixel-level, logarithmically binned 2D-histograms are found in figure 4.6. The figure reveals notable contrasts between the two set-ups, as the main set-up predictions follow a much more coherent diagonal line. Common for both set-ups is the high spread of low precipitation rate predictions and low frequency of high precipitation rate predictions. The low precipitation rates which are frequent can be seen to be difficult for the model to distinguish. The low frequency of high precipitation rate predictions is expected as these rates seldom occur.

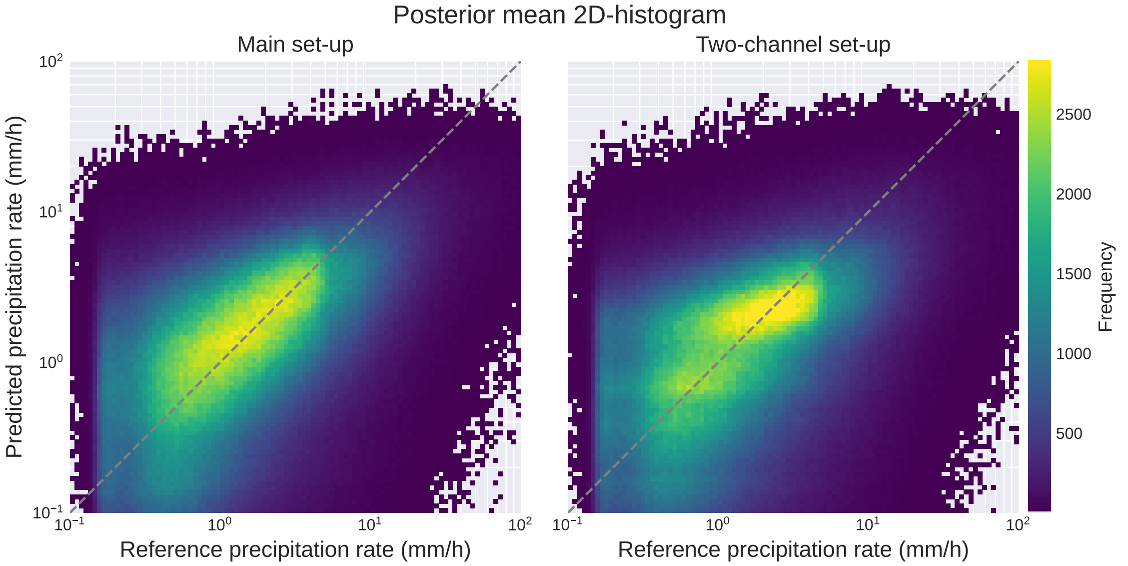

Figure 4.6: 2D-histogram plots of posterior mean predictions. Y-axes show the posterior mean predicted precipitation rate, and x-axes show the corresponding reference rate. The dashed grey diagonal line represents where correct predictions would lie. Note logarithmic scaling of both axes.

The predictions of the lower and higher rates are better illustrated when normalising the plots column-wise, and this is shown in figure 4.7. From the figure, we can observe that both set-ups seem to underestimate high precipitation rates and overestimate low precipitation rates, as was also previously noted. Further, it can be seen that the main set-up follows the dashed diagonal grey line rather well and has a relatively constant spread for most precipitation rates. In contrast, while the two-channel set-up does follow the line, the spread is less consistent. This reveals the model's lower confidence in predicting higher and lower precipitation rates.

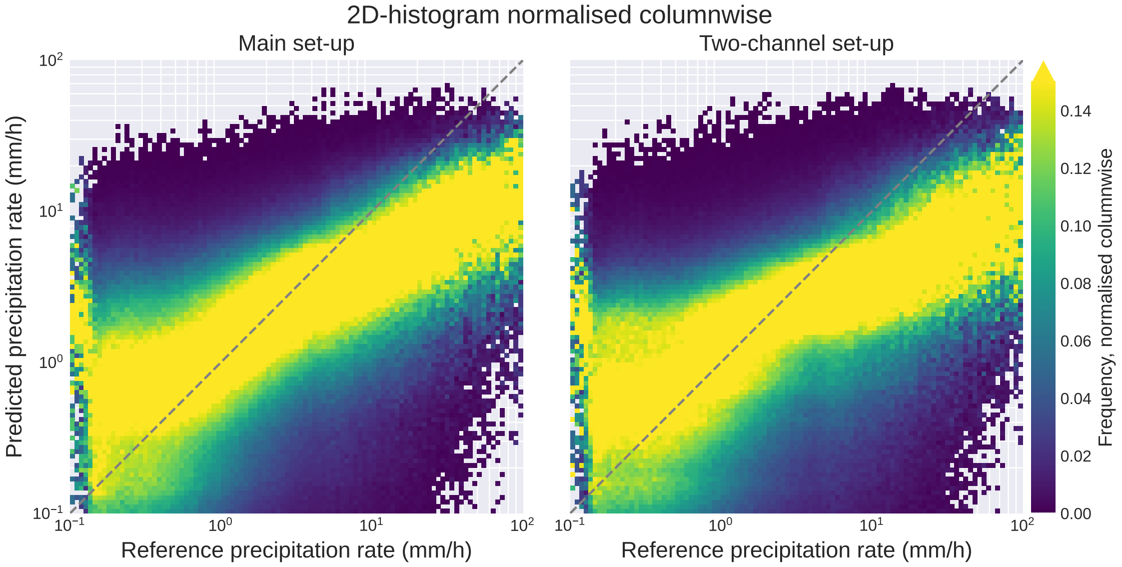

Figure 4.7: 2D-histogram plots of posterior mean predictions here normalised column-wise. Y-axes show the posterior mean predicted precipitation rate, and x-axes show the corresponding reference rate. The dashed grey diagonal line represents where correct predictions would lie. Logarithmic scaling of both axes. Note here the clipped colour bar.

4.3 Comparison with IMERG late run

The comparison with IMERG is made for the winter season, including December, January and February, during the shift from 2021 to 2022. These comparisons are made with the main set-up, using all TIR channels. Version 6 of the GPM combined product is not available for this period of time. Instead, GPM data for the winter season of the previous five years were included in the comparison as reference. It is important to emphasise here that there is no true answer for these comparisons as IMERG as well as our retrievals are only estimates. The GPM retrievals are additionally very sparse for specific regions such that even if they were available for the studied period, the observations would be very few.

The retrievals are compared over three regions, two of which are over the African continent and one over the ocean. The first continent region studied is an area over Algeria with borders to Libya and Tunisia, characterised by dry weather and is colder during the considered winter months. The second continent region is the border between Angola and the Democratic Republic of Congo in the tropics with a wet climate. Finally, a region over the ocean is considered here in the southern hemisphere quite far from the coast. The three subregions, which will be referred to as dry continent, wet continent and ocean, are shown in figure 4.8. Also shown in the figure is the brightness temperature from channel 5 cropped to images of 256×256 pixels for each subregion. Here, the difference in spatial resolution for the subregions can be seen as the geostationary channel 5 data is plotted on the plate carrée projection.

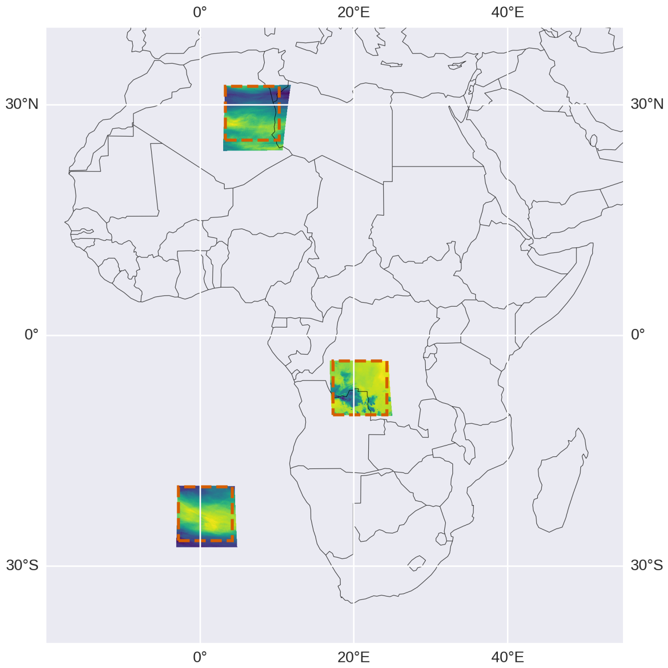

Figure 4.8: Regions for which we compare our retrievals with IMERG. The area inside of the red dashed box is considered for comparison. Also illustrated is brightness temperature from channel 5 cropped into 256 × 256 boxes surrounding the subregions. From top to bottom, we have regions characterised by dry continent, wet continent and ocean.

We first study the probability density functions of precipitation rates over the subregions as retrieved from our QRNN, IMERG and GPM. In figure 4.9 we note that the sampled posterior captures the density of historical data from GPM better than IMERG. Overall, the posterior mean and IMERG retrievals are relatively close for precipitation rates with high probabilities, though IMERG can be seen to be closer to the GPM density for rates with lower probabilities.

Observing all three regions, we note that both IMERG and the QRNN retrievals seem to be closer to the GPM density for the wet continent. This could be because there are more samples of precipitation events such that the distribution is more easily captured. For the dry continent, it can be seen that IMERG, the posterior mean and the sampled posterior all report lower precipitation rates than the GPM retrievals used as reference. This could potentially be explained by that the winter season 21/22 was an exceptionally dry period for the region or that the GPM data contain unusually wet years for the season. Finally, all three products report moderately low precipitation rates for the ocean area.

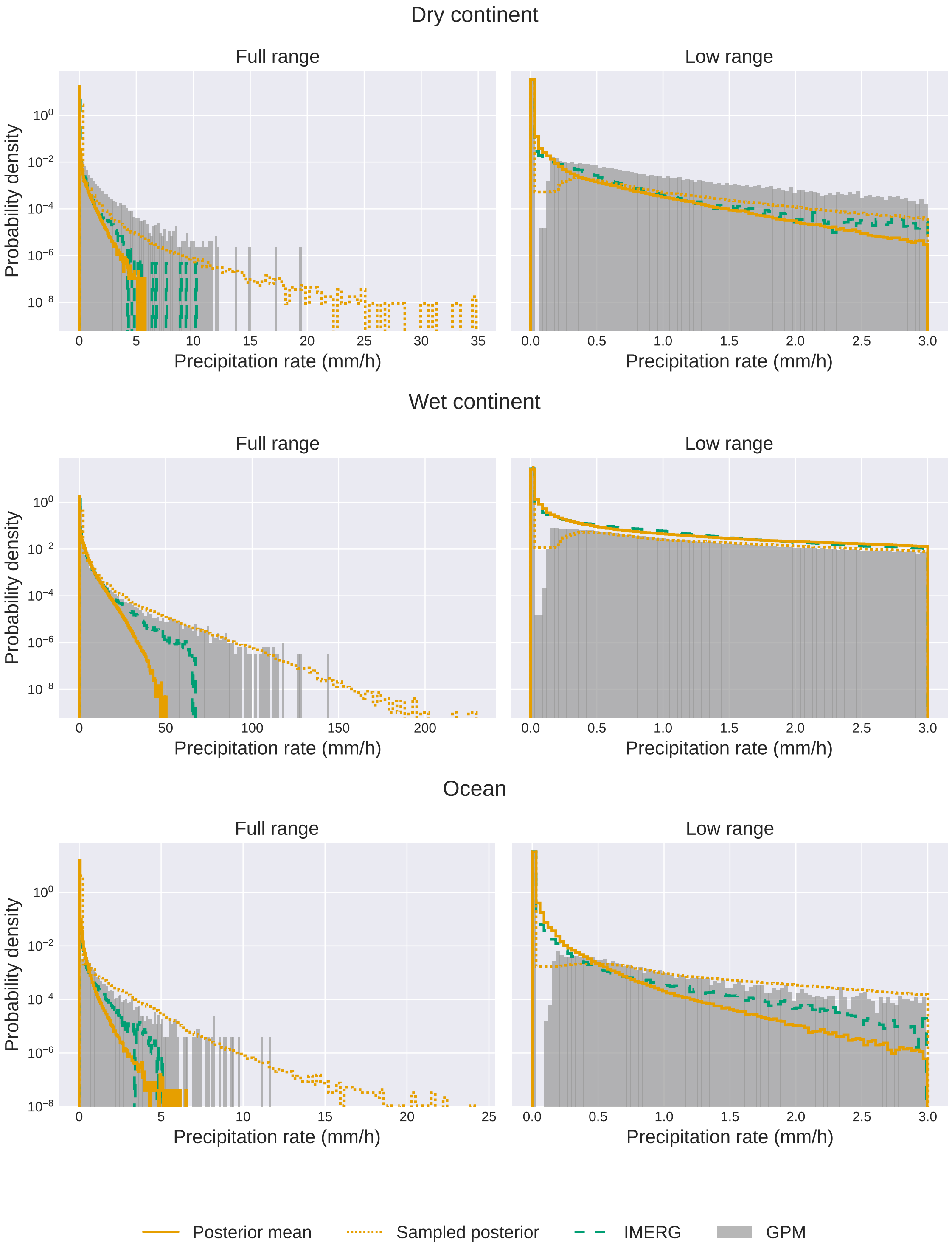

Figure 4.9: Probability densities for QRNN retrievals, IMERG and GPM over the three subregions. All figures share legends. GPM is plotted in filled grey boxes, posterior mean in solid orange line, sampled posterior in dashed orange line and IMERG in the solid green line. Logarithmic scaling of y-axes.

Turning to the diurnal cycles of precipitation, these are found in figure 4.10. We first make a general observation that the posterior mean generates smoother curves than both GPM and IMERG. This is because the QRNN retrievals have a higher temporal resolution of 15 minutes, while IMERG produces an output for every 30 minutes, and GPM is plotted for every hour. Because of the long revisit period of GPM-CO, only around 20 GPM samples were obtained per hour. This is compared to the 90 samples per timestamp that the QRNN and IMERG retrievals yield. Secondly, we note that IMERG and the posterior mean seem to follow similar overall patterns for all three regions.

The difference in probability densities previously noted for the dry continental region is also reflected in the diurnal cycle. The GPM retrievals are higher, while both IMERG and the posterior mean report lower rates. The precipitation rates here are meagre, and a distinct pattern cannot be deduced. For the wet continent, we see that all three products seem to capture some variation, with a peak in the late afternoon. A secondary peak during late-evening to early-morning hours is found, and it is more pronounced in the IMERG curve. For the ocean region, the variation is much lower, and a distinct maximum is found in the late-evening to early-morning hours in the GPM and QRNN retrievals. The IMERG curve, on the other hand, shows more severe disruptions.

Finally, we observe a snapshot of precipitation retrievals from QRNN and IMERG late run. Shown in the right panel of figure 4.11 is a QRNN retrieval over a cropped 2048 × 2048 SEVIRI observation. Inference on such a cropped input for the main set-up takes less than a minute to perform and can be done using a CPU. The left of the figure shows precipitation retrievals from IMERG for the same timestamp, 02:00 UTC on 2 December 2021. Here, we see that the general patterns are similar for both retrievals. There are differences in the details such as the dry regions of Libya in the north of Africa, and the concentration of rates over South Sudan in the tropics. These are low precipitation rates reported by IMERG late run, not represented in the posterior mean. For the ocean region in the very south of the map, a more coherent pattern of precipitation is retrieved by the posterior mean. The boundaries for the two retrievals differ because the SEVIRI data is given in a geostationary projection, and these are shown on a flat plate carrée projection. The cropped 2048 × 2048 observation is illustrated on both projections in appendix B. The retrievals are zoomed in for the wet continent region in figure 4.12. Here, the benefit of a higher spatial resolution is apparent, as the pixels on the IMERG retrieval are clearly noticable.

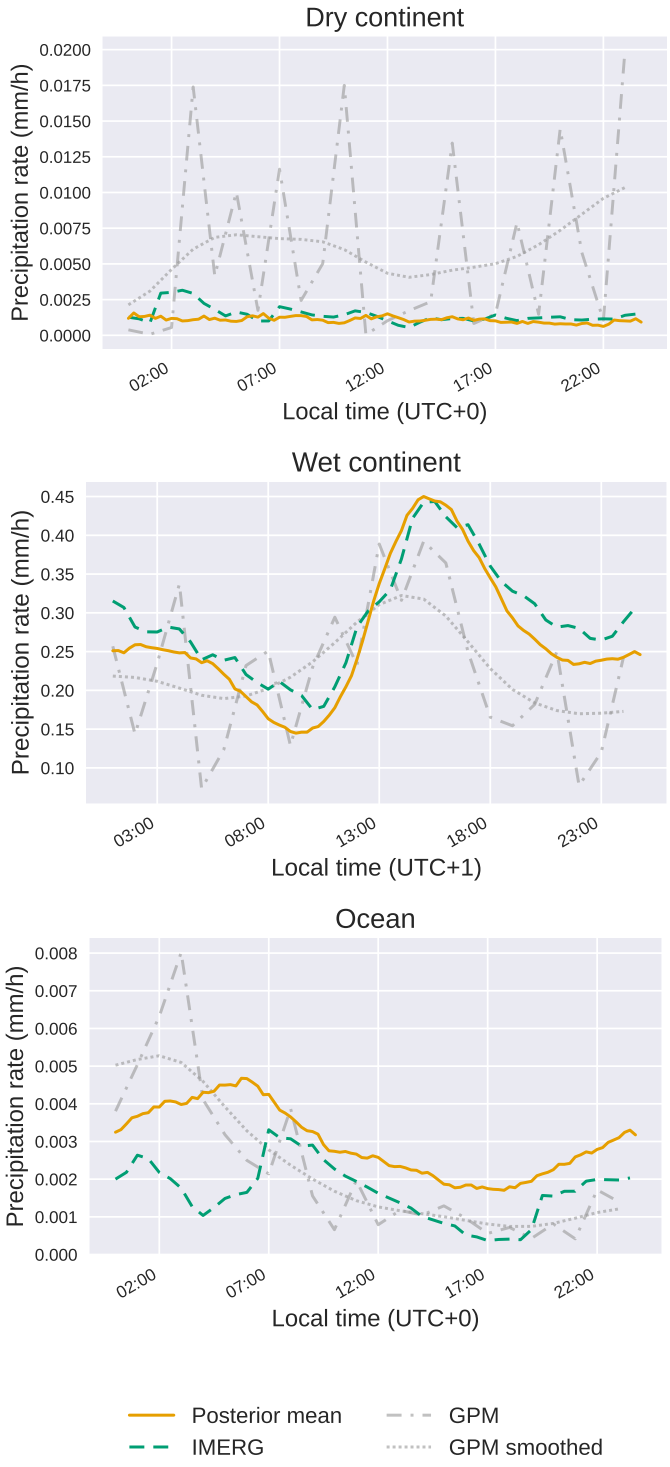

Figure 4.10: Diurnal variations retrieved by posterior mean, IMERG and GPM for subregions. All figures share legends. Because the GPM data contain fewer samples, the GPM curve is also shown smoothed using a one-dimensional Gaussian filter. Note the different scales in vertical and horizontal axes for the different regions.

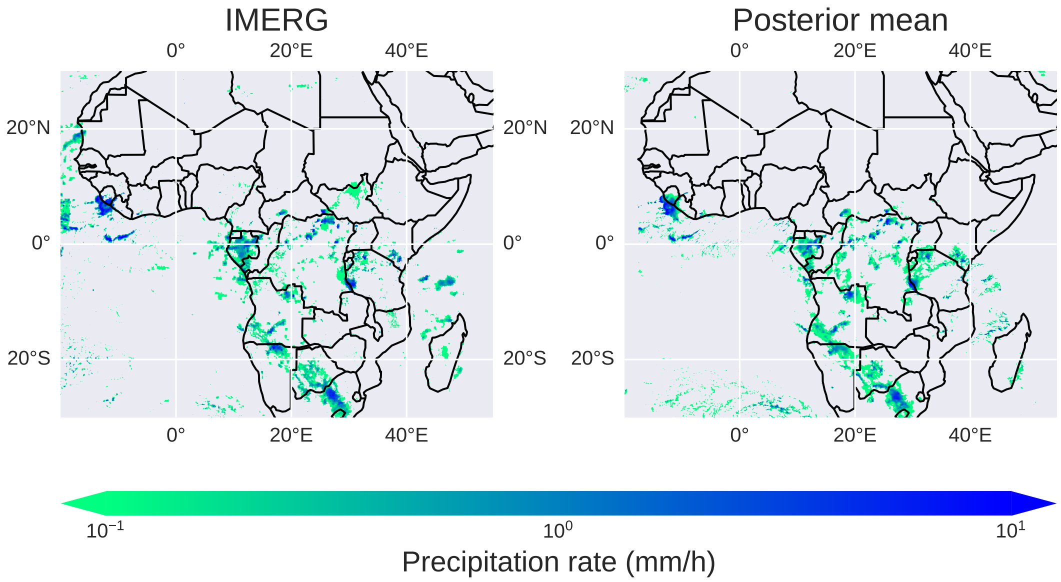

Figure 4.11: Precipitation retrievals over Africa at 02:00 UTC on 2 December 2021. The snapshot shows similarities in general precipitation patterns captured by IMERG and the posterior mean. Some minor differences are seen over the ocean, dry regions in the north and in the upper tropical regions. The boundaries of the snapshots differ as the QRNN retrievals are given in a geostationary grid.

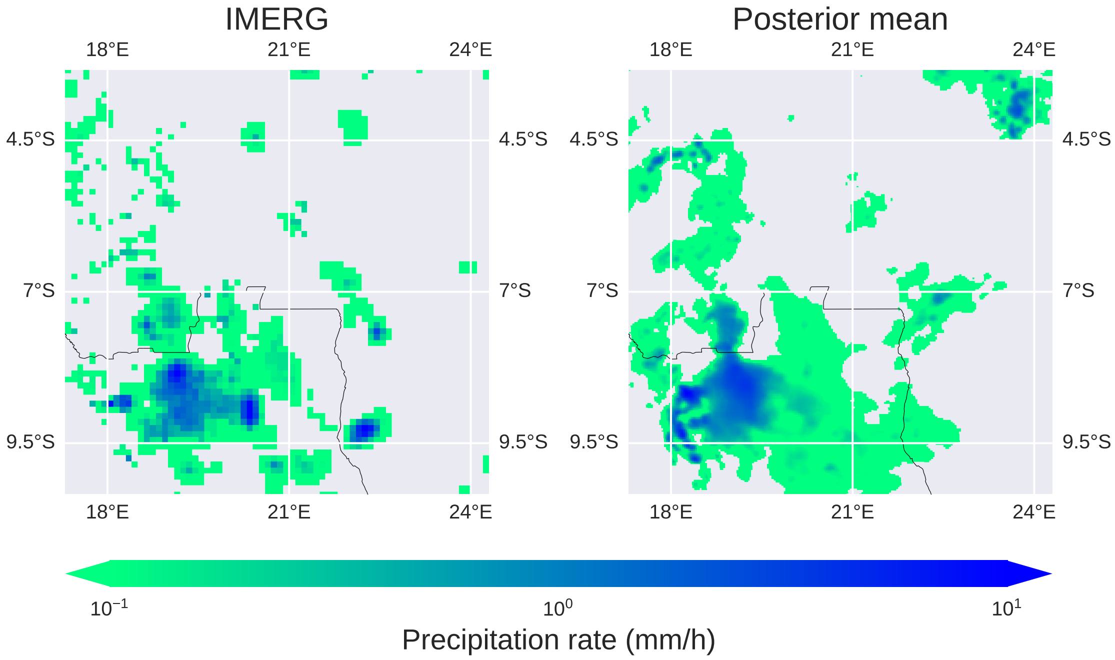

Figure 4.12: Precipitation retrievals over the wet continent region at 02:00 UTC on 2 December 2021. The 10 km × 10 km spatial resolution of IMERG is compared to the higher 3 km × 3 km resolution of the QRNN retrievals. As a result, the pixels in the IMERG snapshot are more perceptible.

5

Discussion

With the QRNN approach, high-resolution precipitation retrievals over Africa were obtained. Here we discuss the assessment of the retrievals, the methodology and possible or necessary extensions of the project.

5.1 Result findings

From evaluating the model on the test set, it was found that the posterior mean in all set-ups tends to overestimate low precipitation rates while underestimating high precipitation rates. This was also seen in previous works and was expected as the higher rates are less frequent while the lower rates are more frequent (Ingemarsson, 2021; Pfreundschuh, Ingemarsson, et al., 2022). The cause behind the difficulty of correctly capturing these rates is likely due to the vague relationship between cloud top temperatures and precipitation rates. Cold cloud top temperatures can generally be associated with heavier precipitation (Hanna, Schultz, and Irving, 2008). However, a non-precipitating cloud may very well rise in altitude due to environmental factors, causing its cloud top temperature to cool. Additionally, precipitation can form during both warm and cold processes, such that the relationship is certainly non-linear (McFarquhar, 2022). Evidently, different events can correspond to the same input features. Because of the low frequency of the higher precipitation rates, the posterior mean is more likely to underestimate these rates. In classification problems, it is usual to balance the datasets to ensure that the network encounters each class an equal number of times. Here, we have a regression problem, and more importantly, the input features cannot be well separated. If a balancing of the dataset would have been made (by for example discarding a portion of the non-precipitating pixels), the network would be more incentivised to predict the higher rates. Nonetheless, it would still have difficulty discriminating the feature space. This would then result in an introduced systematic bias. Instead, it is rather preferred that the bias arises from the natural distribution of precipitation rates, as seen by the model.

When comparing different inference sizes, the evaluation metrics were judged to be very similar. However, from the visualised predictions, it was noted that edge effects occur when arranging the smaller images to a bigger one. Since the CNN is given less context at the edges, this was reasonable. Though no evaluation on other image sizes was made, it can be suggested that even if the model was trained on 128 × 128 images, it can handle larger image sizes. Further, because of the edge effects, it is even preferable to use larger image sizes when the region of interest is larger. Inferring on an unnecessarily big image will however cost time without any added value. It is therefore advised to use images which are large enough such that they cover the entire region of interest while at the same time keeping it as small as possible.

With the two-channel set-up, we were able to train the network with comparable results. We reported only slightly lower accuracy in predicting higher and lower precipitation rates compared to the main set-up. Nevertheless, it seemed to capture the same overall structures as seen in the predictions from the main set-up. This indicates the potential use of older generation satellites that could cover 30 years of observations. However, how to undertake this task is not clear. The model could perhaps directly infer on the older generation data. Nevertheless, the accuracy of these predictions would be difficult to evaluate without reliable data records. If such records are unavailable, then at the very least, an interpretation of historical precipitation rates with high temporal and spatial resolution would be offered. Because of the lack of GPM data for years before 2014, it would be challenging to train a network using historical data. The Tropical Rainfall Measuring Mission (TRMM), which GPM was built upon, dates back to 1997 (Kummerow et al., 2000). Data from this mission could suggestively be used as ground truth. However, for earlier years before the launch of TRMM, not much reliable reference data is available. Employing a supervised machine learning method as in this work would then be less straightforward. Instead, other networks could be explored, or different ways of treating the available data could be considered.

Lastly, we discuss the comparison with IMERG late run. For the diurnal cycle in the wet continent region, we found a maximum in the late afternoon, which corresponds well with results from extensive studies of diurnal cycles such as in Yang and Smith (2008). While a number of complex mechanisms contribute to these variations, a general explanation of this peak is that it results from the Sun's heating of the surface during daytime and the hydrostatic instabilities that follow as the surface cools down (Yang and Smith, 2006). Additionally, the maximum during late-evening to early-morning hours found by the posterior mean and GPM over the ocean is also in accordance with what is expected. IMERG did not report a well-defined maximum over the ocean as the curve for this region was more disrupted. For the snapshot comparison, differences were also seen over the southern ocean, where the precipitation patterns in the posterior mean appeared more coherent. Animations of the snapshots can be found at *Rain over Africa* (2022), where the temporal evolution of the QRNN retrievals appear more natural in comparison to IMERG. Because there was no ground truth data available for these comparisons, the accuracy cannot be assessed. Instead, what can be said is that the QRNN retrieval is comparable to IMERG. It further has higher spatiotemporal resolution and added value of reporting distribution functions rather than scalar quantities. The relative simplicity of the method also implies that further improvements in available precipitation products today are likely possible.

5.2 Validating the retrievals

Validating our retrievals is challenging; we could only settle for a comparison with a superior global precipitation product in this scope. In Pfreundschuh, Ingemarsson, et al. (2022) and Ingemarsson (2021), similar machine learning retrievals over Brazil were validated against rain gauge data and showed good agreement. In Pfreundschuh, Ingemarsson, et al. (2022), comparisons were made with IMERG and PERSIANN-CCS. In general, it was found that while IMERG outperformed PERSIANN-CCS, the suggested machine learning retrievals showed better correlations with rain gauge data. Though the method in the aforementioned study differs slightly from the QRNN approach, they are conceptually parallel in that they both predict probabilities of precipitation rates and achieve similar accuracy. While these are interesting findings of closely related tasks, caution should be raised before inferring any implications on our achieved results. Instead, the retrievals of our model need to be assessed on their own, with suitable validation data. The natural choice for validation data would be rain gauges, as they are the most tangible way of measuring precipitation reaching the ground. Unfortunately, rain gauge data are very scarce for our region of interest. Moreover, they were not possible to acquire in the scope of this project. Even if the data were to be obtained, they only provide point measurements, and performance would, in turn, only be assessed for these points. Furthermore, they come with their own systematic biases, including those of environmental sources such as wind or placement (Cauteruccio, Brambilla, et al., 2021; WMO, 2018). How and if these errors are corrected can vary from site to site and are difficult to trace (Lanza et al., 2021). Needless to say, measuring precipitation accurately is a non-trivial task, and this makes validating precipitation products a demanding process. Nevertheless, to fully assess our model's accuracy, further work should include validation against independent, high-quality precipitation gauge data.

5.3 Decisions taken regarding the methodology

First, we make a note that though the data augmentation, layer input normalisation and GELU activation had some regularisation effects, not much effort was focused on this matter. The reason for that being that overfitting was not seen to be an issue for our task. This is mainly due to the large amount of data that was available for training. A machine learning approach was thus seen to be a suitable method for interpreting our data, and there are potentially numerous tasks in which satellite observations could benefit from machine learning. Furthermore, though slight alterations in network architecture were tested, this aspect was not extensively explored either. There are, however, numerous architectures that could be evaluated. As an example, Trebing, Staczyk, and Mehrkanoon (2021) approached precipitation nowcasting from radar data using a U-Net architecture with convolutional and attention modules. This was done using convolutional block attention modules (CBAM) described in Woo et al. (2018). The attention modules help the network direct focus in the input (both across channels and in the spatial dimensions), and a weighted feature map is obtained. Evidently, there are endless options to explore, and so there are possible improvements that could be made in future work.

When preparing the datasets for training, validating and testing, any sample containing missing channel information masked as NaN was discarded. This was done since such pixels would affect surrounding pixels during the convolution operation. Consequently, the model is not able to handle defective pixels. A quick inspection of inference on a corrupted image where NaNs were filled with a value of -999999 was made. The inspection revealed that unreasonable precipitation rates were being predicted. As missing information in channels can occasionally occur, it would be of interest to be able to handle such observations. One possible measure could be to fill these pixels with the averaged value of surrounding, valid pixels. If there are many faulty pixels, this becomes more difficult. In this case, it is incentivised to further explore alternative networks which can better handle corrupted images. One such network could be realised with partial convolutions described in Liu et al. (2018).

The quantile loss function in equation (3.3) is minimised without any constraints on the quantiles x_τ. A cumulative distribution function is monotonically increasing, meaning that the predicted quantiles must also be in increasing order for increasing quantile levels τ. Although the predicted quantiles were most often seen to be aligned in a sound order, the possibility for quantiles to cross each other exists. For one of the trainings, the predicted distributions were corrected to account for the crossings. The correction was made using an isotonic regression method. This minimises the displacement added to the quantiles necessary to make the distribution become monotonically increasing. The difference was not deemed to be significant when comparing the posterior mean of the corrected and uncorrected distributions. The correction was made for the entire test set and took roughly two days to complete. Based on this, we did not continue to correct for the crossing quantiles as they were neither frequent nor drastic enough to become a problem. Furthermore, it would have increased the latency of the retrievals. Still, if one is concerned with the correctness of the predicted distributions there are several possible ways to deal with this such as presented in Cannon (2018) or Moon et al. (2021). However, these methods can have potential weaknesses too, and it was not further investigated in this project. A rough assessment of corrected and uncorrected retrievals can be found in appendix C.

Regarding the separation of datasets using a random split, it is believed to be a fair choice of methodology. The test set is fully separated from both the training and validation sets. This choice was made to harness as much information as possible from the data available, with the training set containing samples from the entire time period. An alternative would be to split the sets by time. More careful consideration would then have been needed as there are daily, seasonal and even annual variations in precipitation (Wang and Picaut, 2004; Madden and Julian, 1971; Webster et al., 1998). If the test set, for instance, contained mostly dry events, an evaluation on this set might not yield a very informative assessment. Because the sets were split randomly, it is likely that samples in different sets belong to the same event, albeit reporting different observations of it. This should not affect the assessment of the model's capability to generalise as it is truly seeing new samples in the test set. From the comparison with IMERG over a completely separate time period, we saw its ability to capture expected trends. Furthermore, the model is learning the relationship between cloud top temperatures and precipitation rates. Although the future might see extreme events at different frequencies and durations, the physical relation between cloud top temperature and precipitation will not change. It is then trusted that if our model has learnt this relation well enough, it will be able to capture these events in a changing climate.

5.4 Limitations in the data

The input data was limited to the TIR channels of the SEVIRI instrument, and the network was trained to relate these with the GPM DPR and GMI combined precipitation product. The QRNN retrievals are then bounded by the limitations of the used data. For instance, a consequence of the 0.2 mm/h sensitivity threshold of the GPM product is that the QRNN will not accurately retrieve precipitation rates lower than this threshold. As previously discussed, precipitation from clouds with warm tops is difficult to detect using IR imagery. Remote sensing methods using microwave radiation also struggle with this type of precipitation because emission from the Earth's surface at the relevant wavelengths cause signal disturbances (M. B. Ba and Gruber, 2001). In M. B. Ba and Gruber (2001), channels in the visible spectrum were used to assess the microphysical structure of cloud tops better, and they were also used in Adler and Negri (1988) to discriminate thin, non-precipitating clouds with cold cloud tops. The inclusion of channels with reflected solar contribution could thus reasonably be considered, as was done in the related works of Amell (2021) and Pfreundschuh, Ingemarsson, et al. (2022). Such channels can, however, only be used during daytime.

Besides considering additional SEVIRI channels, including different data sources could also be discussed. When comparing the QRNN retrieval to IMERG, a secondary maximum was observed in the diurnal cycle for the wet content region in figure 4.10. The study of Yang and Smith (2008) suggests that this secondary peak is a contribution from stratiform precipitation. Stratiform precipitation is precipitation from clouds with mainly horizontal development. Since these clouds do not necessarily rise in altitude, their cloud tops are not necessarily cold. Although a secondary maximum was observed in the posterior mean, the QRNN retrievals possibly struggle slightly at capturing these types of precipitation. The amplitude of the secondary peak was more pronounced in IMERG, while the phase was relatively similar for all three products. IMERG, which is a highly sophisticated precipitation product with numerous data sources, perhaps benefits from the abundance of information. However, we also see compromises in terms of resolution and availability. The limited number of data sources in the QRNN method has its benefits, as complexity is reduced because no major merges are needed. With this, the errors from such merges are not present. Additionally, the availability of the retrievals only depends on the input source for inference. In contrast, some disadvantages arise due to the fact that not all precipitation phenomena can be observed with a single remote sensing method. Incorporating multiple data sources is then a decision most appropriately dictated by the product's intended use.

Finally, we restate that the SEVIRI observations have a revisit period of 15 minutes. A natural expansion of the project could include a temporal dimension to predict retrievals minutes or hours into the future. This could suggestively be done by matching a cropped GPM box as described in section 2.2, not with the SEVIRI observation for the same timestamp but for some consecutive observations earlier in time. Here, we are limited to observations 15 minutes apart. For precipitation, this can sometimes mean much change. Geostationary instruments usually offer a higher temporal sampling of 5 minutes or less for certain regions, and radar data are also available at a higher frequency of 5–10 minutes. None of these observations are currently available for our entire region of interest. However, with the upcoming Meteosat third generation satellites, the temporal resolution will be increased to 10 minutes and the spatial resolution to $2\,\mathrm{km} \times 2\,\mathrm{km}$ at nadir (Holmlund et al., 2021). The first satellite of the series is scheduled for launch in November 2022, and with it, the prospect of geostationary precipitation retrievals is bright.

6

Conclusion

The aim of this thesis was to apply a Quantile Regression Neural Network (QRNN) to retrieve precipitation over Africa. This was done with TIR channels of the SEVIRI instrument onboard geostationary Meteosat-11, using GPM retrievals as reference. With the QRNN, probabilities of precipitation were retrieved. These approximated probability functions contain information richer than that obtained from a single scalar quantity. The retrievals can be obtained at a high frequency as the SEVIRI instrument scans the continent every 15 minutes, and inference on a 2048×2048 image covering most of Africa takes under a minute to perform. Additionally, the retrievals have a high spatial resolution, $3\,\mathrm{km} \times 3\,\mathrm{km}$ at nadir.

The GPM combined product retrievals could be replicated using geostationary TIR channels in two settings. First, using all TIR channels without a solar contribution, and secondly, only using channel 5 ($6.2\,\mu\mathrm{m}$) and channel 9 ($10.8\,\mu\mathrm{m}$). The first set-up proved the overall best performance, though the set-up only using two channels also showed adequate results. The obtained retrievals were shown to be comparable with IMERG late run in distributions and diurnal cycles over continent and ocean subregions. As opposed to IMERG, a highly sophisticated global precipitation product, the QRNN approach is a relatively simple method. It uses only one data source for input, and all software and data used are public and freely available. It also has the benefit of a higher spatiotemporal resolution and lower latency of retrievals. From this, it can be implied that there is potential for further improvements in the global precipitation products available today.

Recommendations for future work include validating the retrievals against high-quality, independent gauge data and better handling invalid data points in the SEVIRI observations. Further, there are several interesting aspects to build on. A few suggestions could be exploring different network architectures, using historical data, considering other data sources, or predicting retrievals some time into the future.

Bibliography

Adler, R. F. and Negri, A. J. (1988). "A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall". In: *Journal of Applied Meteorology and Climatology* 27.1, pp. 30–51. DOI: `10.1175/1520-0450(1988)027<0030:ASITTE>2.0.CO;2`.

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L. (2006). "Global observed changes in daily climate extremes of temperature and precipitation". In: *Journal of Geophysical Research: Atmospheres* 111.D5. DOI: `10.1029/2005JD006290`.

Allen, M. R. and Ingram, W. J. (2002). "Constraints on future changes in climate and the hydrologic cycle". In: *Nature* 419.6903, pp. 228–232. ISSN: 1476-4687. DOI: `10.1038/nature01092`.

Amell, A. (2021). "Geostationary passive retrieval of ice water path with quantile regression neural networks". MA thesis. Chalmers University of Technology. URL: `https://hdl.handle.net/20.500.12380/302612` (visited on 07/22/2022).

Arkin, P. A. and Meisner, B. N. (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. DOI: `10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2`.

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). *Layer Normalization.* DOI: `10.48550/ARXIV.1607.06450`.

Ba, M. B. and Gruber, A. (2001). "GOES Multispectral Rainfall Algorithm (GMSRA)". In: *Journal of Applied Meteorology* 40.8, pp. 1500–1514. DOI: `10.1175/1520-0450(2001)040<1500:GMRAG>2.0.CO;2`.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Dijk, A. I. J. M. van, Huffman, G. J., Adler, R. F., and Wood, E. F. (2019). "Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS". In: *Hydrology and Earth System Sciences* 23.1, pp. 207–224. DOI: `10.5194/hess-23-207-2019`.

Browning, K. A. and Collier, C. G. (1989). "Nowcasting of precipitation systems". In: *Reviews of Geophysics* 27.3, pp. 345–370. DOI: `10.1029/RG027i003p00345`.

Cannon, A. J. (2018). "Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes". In: *Stochastic Environmental Research and Risk Assessment* 32.11, pp. 3207–3225. ISSN: 1436-3259. DOI: `10.1007/s00477-018-1573-6`.

Cauteruccio, A., Brambilla, E., Stagnaro, M., Lanza, L. G., and Rocchi, D. (2021). "Experimental evidence of the wind-induced bias of precipitation gauges using particle image velocimetry and particle tracking in the wind tunnel". In: *Journal of Hydrology* 600, p. 126690. ISSN: 0022-1694. DOI: `10.1016/j.jhydrol.2021.126690`.

Cauteruccio, A., Colli, M., Stagnaro, M., Lanza, L. G., and Vuerich, E. (2021). "Insitu Precipitation Measurements". In: *Springer Handbook of Atmospheric Measurements*. Ed. by T. Foken. Cham: Springer International Publishing, pp. 359–400. ISBN: 978-3-030-52171-4. DOI: `10.1007/978-3-030-52171-4_12`.

Cermak, J., Trigo, I. F., and Fuchs, J. (2021). "Imaging Techniques". In: *Springer Handbook of Atmospheric Measurements*. Ed. by T. Foken. Cham: Springer International Publishing, pp. 1171–1184. ISBN: 978-3-030-52171-4. DOI: `10.1007/978-3-030-52171-4_42`.

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`.

Draper, D. W., Newell, D. A., Wentz, F. J., Krimchansky, S., and Skofronick-Jackson, G. M. (2015). "The Global Precipitation Measurement (GPM) Microwave Imager (GMI): Instrument Overview and Early On-Orbit Performance". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.7, pp. 3452–3462. DOI: `10.1109/JSTARS.2015.2403303`.

EUMETSAT (2017). *MSG Level 1.5 Image Data Format Description*. EUMETSAT. URL: `https://www.eumetsat.int/media/45126` (visited on 07/22/2022).

EUMETSAT (2022a). *Data Store*. URL: `https://data.eumetsat.int/` (visited on 07/22/2022).

EUMETSAT (2022b). *High Rate SEVIRI Level 1.5 Image Data - MSG - 0 degree*. URL: `https://navigator.eumetsat.int/product/EO:EUM:DAT:MSG:HRSEVIRI` (visited on 07/22/2022).

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J. (2015). "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes". In: *Scientific Data* 2.1, p. 150066. ISSN: 2052-4463. DOI: `10.1038/sdata.2015.66`.

Geleta, C. D. and Deressa, T. A. (2021). "Evaluation of Climate Hazards Group InfraRed Precipitation Station (CHIRPS) satellite-based rainfall estimates over Finchaa and Neshe Watersheds, Ethiopia". In: *Engineering Reports* 3.6, e12338. DOI: `10.1002/eng2.12338`.

Glorot, X., Bordes, A., and Bengio, Y. (2011). "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, pp. 315–323. URL: `http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf` (visited on 07/22/2022).

Grecu, M., Olson, W. S., Munchak, S. J., Ringerud, S., Liao, L., Haddad, Z., Kelley, B. L., and McLaughlin, S. F. (2016). "The GPM Combined Algorithm".

In: *Journal of Atmospheric and Oceanic Technology* 33.10, pp. 2225–2245. DOI: `10.1175/JTECH-D-16-0019.1`.

Handmer, J., Honda, Y., Zbigniew, K., Arnell, N., Benito, G., Hatfield, J., Mohamed, I., Peduzzi, P., Wu, S., Sherstyukov, B., Takahashi, K., Yan, Z., Abdulla, A., Bouwer, L., Campbell, J., Hashizume, M., Hattermann, F., Heilmayr, R., Keating, A., and Yamano, H. (2012). "Changes in impacts of climate extremes: human systems and ecosystems". In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* Ed. by C. Field, V. Barros, T. Stocker, D. Qin, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G.-K. Plattner, S. Allen, M. Tignor, and P. Midgley. Cambridge University Press, pp. 231–290. URL: `https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap4_FINAL-1.pdf` (visited on 07/22/2022).

Hanna, J., Schultz, D., and Irving, A. (2008). "Cloud-Top Temperatures for Precipitating Winter Clouds". In: *Journal of Applied Meteorology and Climatology* 47, pp. 351–359. DOI: `10.1175/2007JAMC1549.1`.

Held, I. M. and Soden, B. J. (2006). "Robust Responses of the Hydrological Cycle to Global Warming". In: *Journal of Climate* 19.21, pp. 5686–5699. DOI: `10.1175/JCLI3990.1`.

Hendrycks, D. and Gimpel, K. (2016). *Gaussian Error Linear Units (GELUs).* DOI: `10.48550/ARXIV.1606.08415`.

Holmlund, K., Grandell, J., Schmetz, J., Stuhlmann, R., Bojkov, B., Munro, R., Lekouara, M., Coppens, D., Viticchie, B., August, T., Theodore, B., Watts, P., Dobber, M., Fowler, G., Bojinski, S., Schmid, A., Salonen, K., Tjemkes, S., Aminou, D., and Blythe, P. (2021). "Meteosat Third Generation (MTG): Continuation and Innovation of Observations from Geostationary Orbit". In: *Bulletin of the American Meteorological Society* 102.5, E990–E1015. DOI: `10.1175/BAMS-D-19-0304.1`.

Hong, Y., Hsu, K.-L., Sorooshian, S., and Gao, X. (2004). "Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System". In: *Journal of Applied Meteorology* 43.12, pp. 1834–1853. DOI: `10.1175/JAM2173.1`.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T. (2014). "The Global Precipitation Measurement Mission". In: *Bulletin of the American Meteorological Society* 95.5, pp. 701–722. DOI: `10.1175/BAMS-D-13-00164.1`.

Hsu, K.-l., Gao, X., Sorooshian, S., and Gupta, H. V. (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`.

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. (2020). "Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG)". In: *Satellite Precipitation Measurement: Volume 1.* Ed. by V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, and F. J. Turk. Cham: Springer International Publishing, pp. 343–353. ISBN: 978-3-030-24568-9. DOI: `10.1007/978-3-030-24568-9_19`.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., and Tan, J. (2020). “Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation”. In: URL: `https://gpm.nasa.gov/sites/default/files/2020-10/IMERG_doc_201006.pdf` (visited on 07/22/2022).

Iguchi, T. (2020). “Dual-Frequency Precipitation Radar (DPR) on the Global Precipitation Measurement (GPM) Mission’s Core Observatory”. In: *Satellite Precipitation Measurement: Volume 1*. Ed. by V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, and F. J. Turk. Cham: Springer International Publishing, pp. 183–192. ISBN: 978-3-030-24568-9. DOI: `10.1007/978-3-030-24568-9_11`.

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., and Kubota, T. (2018). *GPM/DPR Level-2 Algorithm Theoretical Basis Document*. URL: `https://www.eorc.jaxa.jp/GPM/doc/algorithm/ATBD_DPR_201811_with_Appendix3b.pdf` (visited on 07/22/2022).

Ingemarsson, I. (2021). “Retrieving precipitation over Brazil. A quantile regression neural networks approach”. MA thesis. Chalmers University of Technology. URL: `https://hdl.handle.net/20.500.12380/304390` (visited on 07/22/2022).

Ioffe, S. and Szegedy, C. (2015). “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *International conference on machine learning*. PMLR, pp. 448–456. URL: `http://proceedings.mlr.press/v37/ioffe15.pdf` (visited on 07/22/2022).

Jiménez Cisneros, B. E., Oki, T., Arnell, N. W., Benito, G., Cogley, J. G., Döll, P., Jiang, T., and Mwakalila, S. S. (2014). “Freshwater resources”. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*. Ed. by C. Field, V. Barros, D. Dokken, K. Mach, M. Mastrandrea, T. Bilir, M. Chatterjee, K. Ebi, Y. Estrada, R. Genova, B. Girma, E. Kissel, A. Levy, S. MacCracken, P. Mastrandrea, and L.L.Whit. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 229–269. URL: `https://archive.ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-Chap3_FINAL.pdf` (visited on 07/22/2022).

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., and Kirschbaum, D. B. (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: `10.1175/BAMS-D-14-00283.1`.

Kidd, C., Takayabu, Y. N., Skofronick-Jackson, G. M., Huffman, G. J., Braun, S. A., Kubota, T., and Turk, F. J. (2020). “The Global Precipitation Measurement (GPM) Mission”. In: *Satellite Precipitation Measurement: Volume 1*. Ed. by V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, and F. J. Turk. Cham: Springer International Publishing, pp. 3–23. ISBN: 978-3-030-24568-9. DOI: `10.1007/978-3-030-24568-9_1`.

Kingma, D. P. and Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. DOI: `10.48550/ARXIV.1412.6980`.

Kirschbaum, D. (n.d). *The Global Precipitation Measurement (GPM) Mission*. URL: `https://commons.esipfed.org/sites/default/files/Dalia_Kirschbaum_Global_Precipitation_Measurement_Mission.pdf` (visited on 07/22/2022).

Koenker, R. (2005). “Introduction”. In: *Quantile Regression.* Econometric Society Monographs. Cambridge University Press, pp. 5–6. DOI: `10.1017/CBO9780511754098.002`.

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A. T. C., Stocker, E., Adler, R. F., Hou, A., Kakar, R., Wentz, F., Ashcroft, P., Kozu, T., Hong, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z., Huffman, G., Ferrier, B., Olson, W. S., Zipser, E., Smith, E. A., Wilheit, T. T., North, G., Krishnamurti, T., and Nakamura, K. (2000). “The Status of the Tropical Rainfall Measuring Mission (TRMM) after Two Years in Orbit”. In: *Journal of Applied Meteorology* 39.12, pp. 1965–1982. DOI: `10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2`.

Lanza, L. G., Merlone, A., Cauteruccio, A., Chinchella, E., Stagnaro, M., Dobre, M., Garcia Izquierdo, M. C., Nielsen, J., Kjeldsen, H., Roulet, Y. A., Coppa, G., Musacchio, C., Bordianu, C., and Parrondo, M. (2021). “Calibration of non-catching precipitation measurement instruments: A review”. In: *Meteorological Applications* 28.3, e2002. DOI: `10.1002/met.2002`.

Lebedev, V., Ivashkin, V., Rudenko, I., Ganshin, A., Molchanov, A., Ovcharenko, S., Grokhovetskiy, R., Bushmarinov, I., and Solomentsev, D. (2019). “Precipitation Nowcasting with Satellite Imagery”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD ’19).* ACM. DOI: `10.1145/3292500.3330762`.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324. DOI: `10.1109/5.726791`.

Levizzani, V., Schmetz, J., Lutz, H. J., Kerkmann, J., Alberoni, P. P., and Cervino, M. (2001). “Precipitation estimations from geostationary orbit and prospects for METEOSAT Second Generation”. In: *Meteorological Applications* 8.1, pp. 23–41. DOI: `10.1017/S1350482701001037`.

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. (2018). *Image Inpainting for Irregular Holes Using Partial Convolutions.* DOI: `10.48550/ARXIV.1804.07723`.

Madden, R. A. and Julian, P. R. (1971). “Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific”. In: *Journal of Atmospheric Sciences* 28.5, pp. 702–708. DOI: `10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2`.

Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., Allan, R. P., Stein, T., Nkonde, E., Senkunda, S., and Alcántara, E. M. U. (2017). “A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa”. In: *Scientific Data* 4.1, p. 170063. ISSN: 2052-4463. DOI: `10.1038/sdata.2017.63`.

McFarquhar, G. M. (2022). “Rainfall microphysics”. In: *Rainfall: Modeling, Measurement and Applications.* Ed. by R. Morbidelli. Elsevier Science, pp. 1–26. ISBN: 9780128225455.

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., and Lane, J. (2009). “Precipitation: Measurement, remote sensing, climatology and modeling”. In: *Atmospheric Research* 94.4, pp. 512–533. ISSN: 0169-8095. DOI: `https://doi.org/10.1016/j.atmosres.2009.08.017`.

Moon, S. J., Jeon, J.-J., Lee, J. S. H., and Kim, Y. (2021). “Learning Multiple Quantiles With Neural Networks”. In: *Journal of Computational and Graphical Statistics* 30.4, pp. 1238–1248. DOI: `10.1080/10618600.2021.1909601`.

NASA (2022a). *Earthdata search.* URL: `https://search.earthdata.nasa.gov/search` (visited on 07/22/2022).

NASA (2022b). *GPM DPR and GMI Combined Precipitation L2B 1.5 hours 5 km V06 (GPM_2BCMB).* URL: `https://disc.gsfc.nasa.gov/datasets/GPM_2BCMB_06/summary` (visited on 07/22/2022).

Nguyen, P., Ombadi, M., Gorooh, V. A., Shearer, E. J., Sadeghi, M., Sorooshian, S., Hsu, K., Bolvin, D., and Ralph, M. F. (2020). “PERSIANN Dynamic Infrared–Rain Rate (PDIR-Now): A Near-Real-Time, Quasi-Global Satellite Precipitation Dataset”. In: *Journal of Hydrometeorology* 21.12, pp. 2893–2906. DOI: `10.1175/JHM-D-20-0177.1`.

Palmer, T. N. and Räisänen, J. (Jan. 2002). “Quantifying the risk of extreme seasonal precipitation events in a changing climate”. In: *Nature* 415.6871, pp. 512–514. ISSN: 1476-4687. DOI: `10.1038/415512a`.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* DOI: `10.48550/ARXIV.1912.01703`.

Pfreundschuh, S. (2022a). *hydronn.* Version 0.1. DOI: `10.5281/zenodo.6371712`.

Pfreundschuh, S. (2022b). *quantnn.* Version v0.0.4dev. DOI: `10.5281/zenodo.6660008`.

Pfreundschuh, S., Eriksson, P., Duncan, D., Rydberg, B., Håkansson, N., and Thoss, A. (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`.

Pfreundschuh, S., Ingemarsson, I., Eriksson, P., Vila, D. A., and Calheiros, A. J. P. (2022). “An improved near real-time precipitation retrieval for Brazil”. In: *EGUsphere* 2022. Preprint, pp. 1–30. DOI: `10.5194/egusphere-2022-78`.

Rain over Africa (2022). URL: `https://rain-over-africa.surge.sh` (visited on 08/17/2022).

Raspaud, M., Hoese, D., Lahtinen, P., Finkensieper, S., Holl, G., Proud, S., Dybbroe, A., Meraner, A., Feltz, J., Zhang, X., strandgren, Joro, S., Roberts, W., Rasmussen, L. Ø., BENR0, Méndez, J. H. B., Zhu, Y., mherbertson, rdaruwala, Buyl, P. de, Jasmin, T., Kliche, C., Barnie, T., Sigurðsson, E., Brodehl, S., R.K.Garcia, Leppelt, T., Tsukada, T., and ColinDuff (2022). *pytroll/satpy: Version 0.36.0 (2022/04/14).* Version v0.36.0. DOI: `10.5281/zenodo.6460193`.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., and Mohamed, S. (2021). “Skilful precipitation nowcasting using deep generative models of radar”. In: *Nature* 597.7878, pp. 672–677. ISSN: 1476-4687. DOI: `10.1038/s41586-021-03854-z`.

Rodgers, C. (1997). “Satellite Remote Sensing: The Retrieval Problem”. In: *The Stratosphere and Its Role in the Climate System.* Ed. by G. P. Brasseur. Vol. 54. Nato ASI Series. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 133–171. ISBN: 978-3-662-03327-2. DOI: `10.1007/978-3-662-03327-2_9`.

Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4. DOI: `10.1007/978-3-319-24574-4_28`.

Rüthrich, F., John, V. O., Roebeling, R. A., Quast, R., Govaerts, Y., Woolliams, E. R., and Schulz, J. (2020). *MVIRI Level 1.5 Climate Data Record Release 1 - MFG - 0 degree.* European Organisation for the Exploitation of Meteorological Satellites. DOI: `10.15770/EUM_SEC_CLM_0013`.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A. (2002). “AN INTRODUCTION TO METEOSAT SECOND GENERATION (MSG)”. In: *Bulletin of the American Meteorological Society* 83.7, pp. 977–992. DOI: `10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2`.

Schmid, J. (2000). “The SEVIRI instrument”. In: *Proceedings of the 2000 EUMETSAT meteorological satellite data user’s conference, Bologna, Italy.* Vol. 29, pp. 13–32. URL: `https://www.eumetsat.int/media/8161` (visited on 07/22/2022).

Scofield, R. A. (1987). “The NESDIS Operational Convective Precipitation- Estimation Technique”. In: *Monthly Weather Review* 115.8, pp. 1773–1793. DOI: `10.1175/1520-0493(1987)115<1773:TNOCPE>2.0.CO;2`.

Seman, S. (2020). “Lesson 5: Remote Sensing of the Atmosphere”. In: *Meteo 3 Introductory Meteorology.* Ed. by S. Seman. Lecture material. Department of Meteorology, Atmospheric Science, College of Earth, and Mineral Sciences, The Pennsylvania State University. URL: `https://www.e-education.psu.edu/meteo3/l5_p5.html` (visited on 07/22/2022).

Tan, J., Petersen, W. A., Kirchengast, G., Goodrich, D. C., and Wolff, D. B. (2018). “Evaluation of Global Precipitation Measurement Rainfall Estimates against Three Dense Gauge Networks”. In: *Journal of Hydrometeorology* 19.3, pp. 517–532. DOI: `10.1175/JHM-D-17-0174.1`.

Trebing, K., Staczyk, T., and Mehrkanoon, S. (2021). “SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture”. In: *Pattern Recognition Letters* 145, pp. 178–186. ISSN: 0167-8655. DOI: `https://doi.org/10.1016/j.patrec.2021.01.036`.

Wang, C. and Picaut, J. (2004). “Understanding ENSO physics - A review”. In: *Earth’s Climate: The Ocean-Atmosphere Interaction.* Ed. by C. Wang, S. Xie, and J. Carton. Vol. 147, pp. 21–48. ISBN: 9780875904122. DOI: `10.1029/147GM02`.

Webster, P. J., Magaña, V. O., Palmer, T. N., Shukla, J., Tomas, R. A., Yanai, M., and Yasunari, T. (1998). “Monsoons: Processes, predictability, and the prospects for prediction”. In: *Journal of Geophysical Research: Oceans* 103.C7, pp. 14451–14510. DOI: `https://doi.org/10.1029/97JC02719`.

WMO (2017). “Applications”. In: *Guidelines for Nowcasting Techniques.* WMO- No. 1198, pp. 32–51. ISBN: 978-92-63-11198-2. URL: `https://library.wmo.int/doc_num.php?explnum_id=3795` (visited on 07/22/2022).

WMO (2018). “MEASUREMENT OF PRECIPITATION”. In: *Guide to instruments and methods observation. Volume I. Measurement of meteorological variables*. WMO-N0.8, pp. 214–246. ISBN: 978-92-63-10008-5. URL: `https://library.wmo.int/doc_num.php?explnum_id=10616` (visited on 07/22/2022).

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). “CBAM: Convolutional Block Attention Module”. In: *Computer Vision – ECCV 2018*. Ed. by V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss. Cham: Springer International Publishing, pp. 3–19. ISBN: 978-3-030-01234-2. DOI: `https://doi.org/10.1007/978-3-030-01234-2_1`.

Yang, S. and Smith, E. A. (2006). “Mechanisms for Diurnal Variability of Global Tropical Rainfall Observed from TRMM”. In: *Journal of Climate* 19.20, pp. 5190–5226. DOI: `10.1175/JCLI3883.1`.

Yang, S. and Smith, E. A. (2008). “Convective–Stratiform Precipitation Variability at Seasonal Scale from 8 Yr of TRMM Observations: Implications for Multiple Modes of Diurnal Variability”. In: *Journal of Climate* 21.16, pp. 4087–4114. DOI: `10.1175/2008JCLI2096.1`.

A

Training metrics

Some metrics computed during training is shown for the main set-up in figure A.1 and for the two-channel set-up in figure A.2. Both set-ups ran for three weeks, with the main set-up reaching 550 epochs while the two-channel set-up reached 649 epochs in the same time.

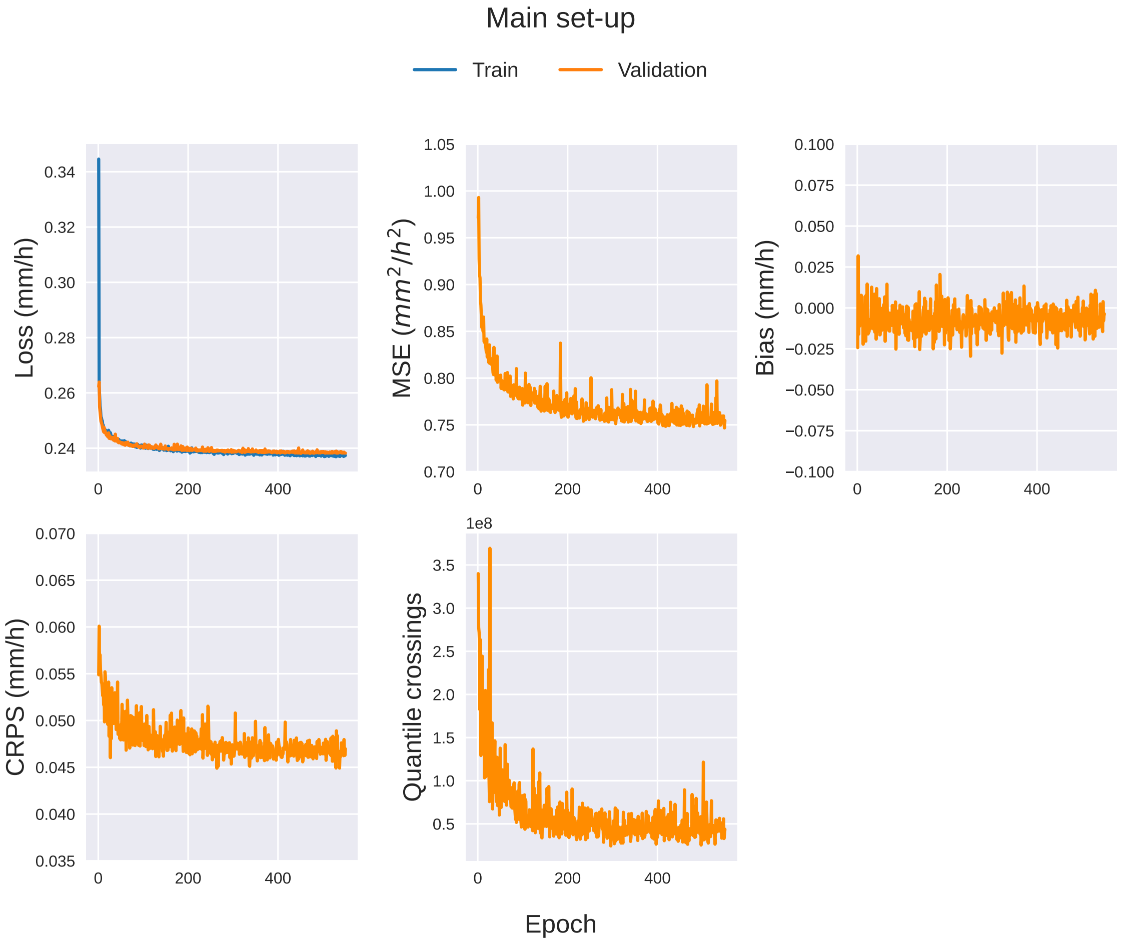

Figure A.1: Training metrics for the main set-up.

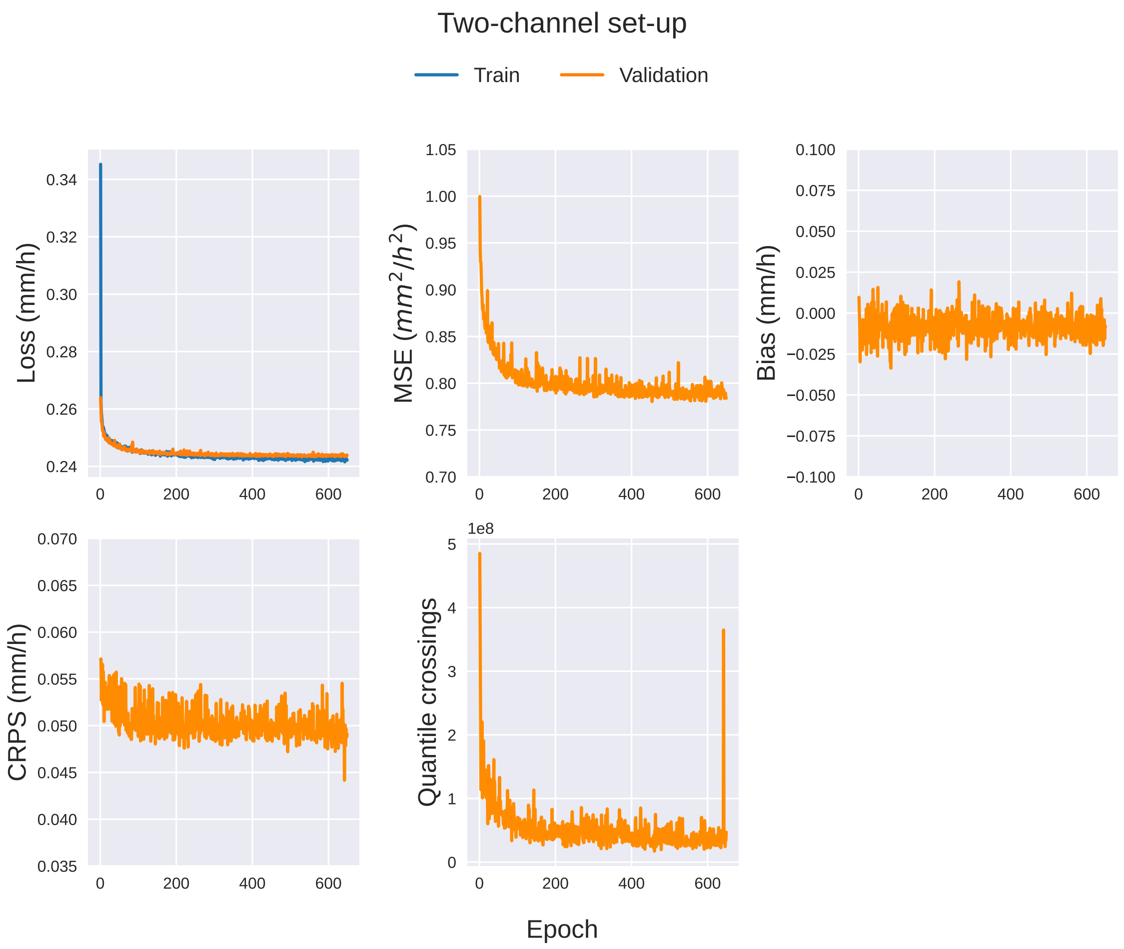

Figure A.2: Training metrics for the two-channel set-up.

B

Projections

Since the input SEVIRI data is projected on a geostationary grid, so are the retrievals. For inference over the African continent, a 2048 × 2048 square in the geostationary grid was cropped. It was chosen as to not contain any space pixels with NaNs and resulted in a coverage slightly smaller than the region of interest. Nontheless, it is possible to perform inference for the full region of interest avoiding NaNs if the region is split into smaller pieces. The cropped square is shown the left in figure B.1. To the right in the same figure is the image on the plate carrée projection A snapshot of retrievals over this region is compared with IMERG in section 4.3.

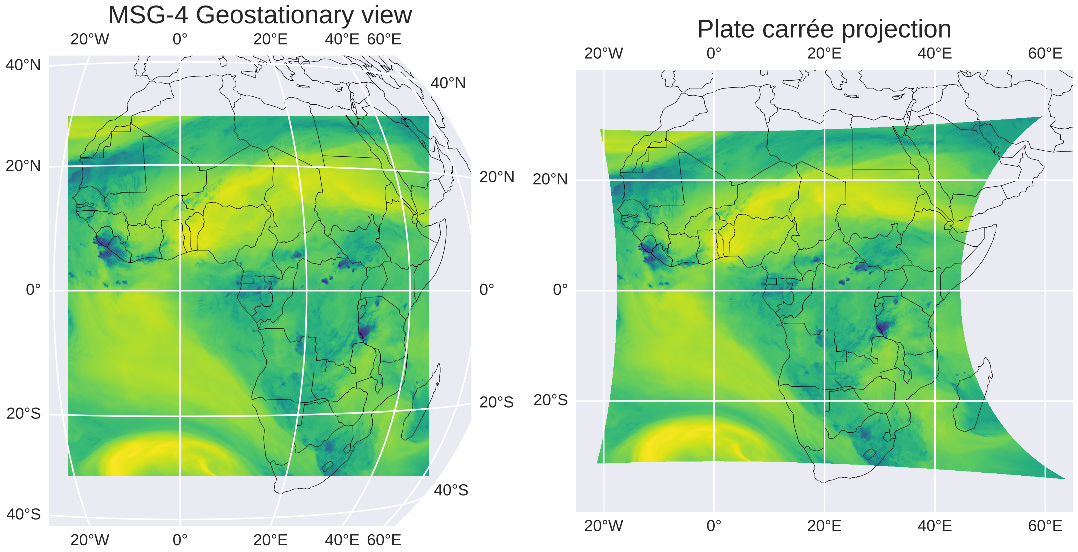

Figure B.1: Channel 5 input image cropped on a 2048 × 2048 square. The same square is then shown on the plate carrée projection to the right.

C

Quantile crossing

Rough assessment on the correction of quantile crossings is found in this section. Figure C.1 shows an example of a predicted CDF with quantile crossing and the same CDF after correction. As can be seen, the posterior mean is not much affected for this example. Metrics for the main set-up before and after corrections is presented in table C.1, and 2D histogram plots in figure C.2.

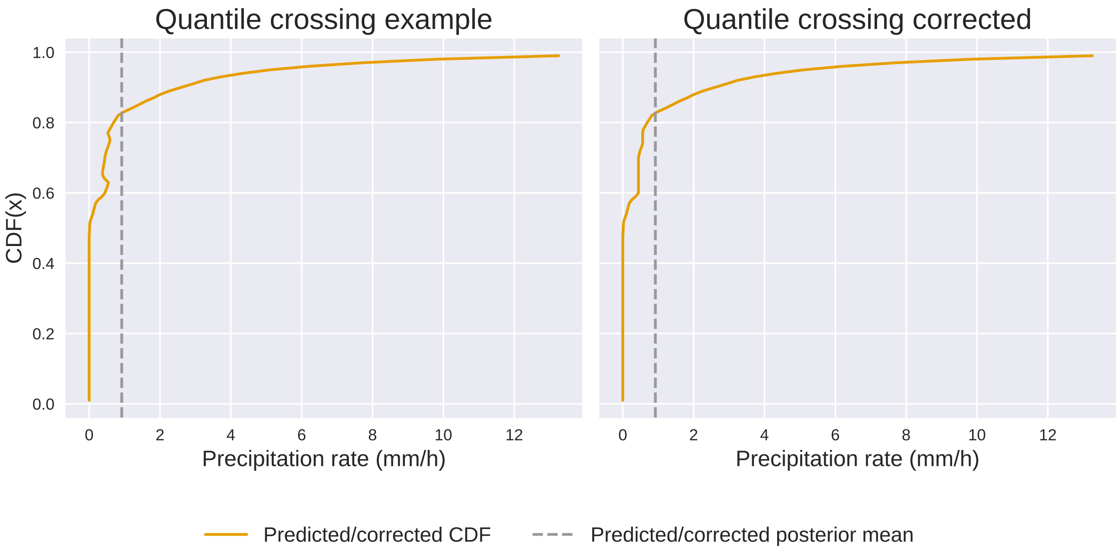

Figure C.1: Example of a predicted CDF with quantile crossing and the same CDF after correction.

Table C.1: Metrics for the main set-up and for the main set-up with quantile crossings corrected. Metrics computed on the test set and for all values of the ground truth. The value closest to zero in each row is marked in bold font.

Metric	Main set-up	Main set-up QC corrected
Bias (mm/h)	$-1.58 \cdot 10^{-3}$	$\mathbf{-5.70 \cdot 10^{-4}}$
MAE (mm/h)	$\mathbf{7.21 \cdot 10^{-2}}$	$7.36 \cdot 10^{-2}$
MSE (mm^2/h^2)	$\mathbf{7.10 \cdot 10^{-1}}$	$7.21 \cdot 10^{-1}$
CRPS (mm/h)		
mean	$\mathbf{4.57 \cdot 10^{-2}}$	$\mathbf{4.57 \cdot 10^{-2}}$
median	**0.0**	**0.0**

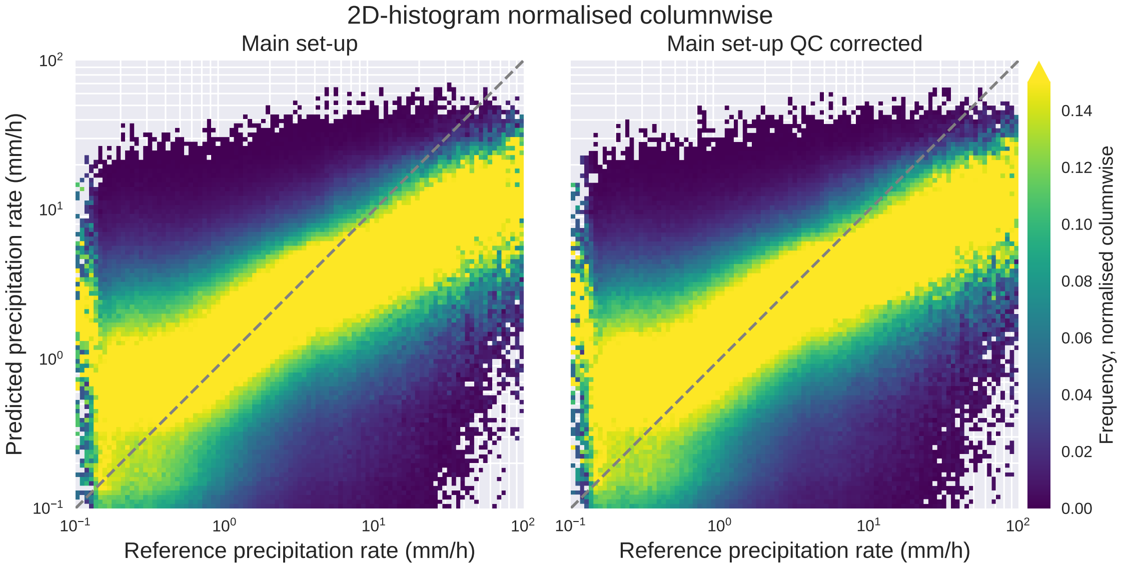

Figure C.2: 2D-histograms for the main set-up before and after quantile crossings correction.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se