THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Improving satellite measurements of clouds and precipitation using machine learning

Simon Pfreundschuh

Department of Earth, Space and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2022

Improving satellite measurements of clouds and precipitation using machine learning

Simon Pfreundschuh
ISBN 978-91-7905-657-5

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5123
ISSN 0346-718X

Department of Space, Earth and Environment
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden Telephone + 46 (0)31-772 1000

Cover:
Isosurfaces of rain (red) and snow (blue) concentrations in hurricane Ida shortly before making landfall in Lousiana on 2021-08-29 at 15:13 UTC calculated from observations of the Global Precipitation Measurement Microwave Imager using an algorithm developed as part of this thesis.

Chalmers digitaltryck
Gothenburg, Sweden 2022

Improving satellite measurements of clouds and precipitation using machine learning

SIMON PFREUNDSCHUH
Department of Space, Earth and Environment
Chalmers University of Technology

ABSTRACT

Observing and measuring clouds and precipitation is essential for climate science, meteorology, and an increasing range of societal and economic activities. This importance is due to the role of clouds and precipitation in the hydrological cycle and the weather and climate of the Earth. Furthermore, patterns of cloudiness and precipitation interact across continental scales and are highly variable in both space and time. Therefore their study and monitoring require observations with global coverage and high temporal resolution, which currently can only be provided by satellite observations.

Inferring properties of clouds or precipitation from satellite observations is a non-trivial task. Due to the limited information content of the observations and the complex physics of the atmosphere, such *retrievals* are endowed with significant uncertainties. Traditional methods to perform these retrievals trade-off processing speed against accuracy and the ability to characterize the uncertainties in their predictions.

This thesis develops and evaluates two neural-network-based methods for performing retrievals of hydrometeors, i.e., clouds and precipitation, that are capable of providing accurate predictions of the retrieval uncertainty. The practicality and benefits of the proposed methods are demonstrated using three real-world retrieval applications of cloud properties and precipitation. The demonstrated benefits of these methods over traditional retrieval methods led to the adoption of one of the algorithms for operational use at the European Organisation for the Exploitation of Meteorological Satellites. The two other algorithms are planned to be integrated into the operational processing at the Brazilian National Institute for Space Research, as well as the processing of observations from the Global Precipitation Measurement, a joint satellite mission by NASA and the Japanese Aerospace Exploration Agency.

The principal advantage of the proposed methods is their simplicity and computational efficiency. A minor modification of the architecture and training of conventional neural networks is sufficient to capture the dominant source of uncertainty for remote sensing retrievals. As shown in this thesis, deep neural networks can significantly improve the accuracy of satellite retrievals of hydrometeors. With the proposed methods, the benefits of modern neural network architectures can be combined with reliable uncertainty estimates, which are required to improve the characterization of the observed hydrometeors.

Keywords: Remote sensing, machine learning, clouds, precipitation

List of publications

Appended articles

This thesis is based on the work contained in the following papers:

Paper 1: S. Pfreundschuh, P. Eriksson, D. Duncan, B. Rydberg, N. Håkansson, and A. Thoss. A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems. *Atmos. Meas. Tech.*, 11(8):4627–4643, 2018. doi: 10.5194/amt-11-4627-2018

Paper 2: S. Pfreundschuh, P. J. Brown, C. D. Kummerow, P. Eriksson, and T. Norrestad. GPROF-NN: A neural network based implementation of the Goddard Profiling Algorithm. *Atmospheric Measurement Techniques Discussions*, 2022:1–37, 2022a. doi: 10.5194/amt-2022-15

Paper 3: S. Pfreundschuh, C. Guilloteau, P. J. Brown, C. D. Kummerow, and P. Eriksson. Evolution of the GPROF passive microwave precipitation retrievals evaluated against ground radar measurements over the continental US and the Pacific Ocean. *In preparation*, 2022c

Paper 4: S. Pfreundschuh, I. Ingemarsson, P. Eriksson, D. A. Vila, and A. J. P. Calheiros. An improved near real-time precipitation retrieval for Brazil. *EGUsphere*, 2022:1–30, 2022d. doi: 10.5194/egusphere-2022-78

Paper 5: I. Kaur, P. Eriksson, S. Pfreundschuh, and D. I. Duncan. Can machine learning correct microwave humidity radiances for the influence of clouds? *Atmospheric Measurement Techniques*, 14(4):2957–2979, 2021. doi: 10.5194/amt-14-2957-2021

Other relevant publications

1. I. Kaur, P. Eriksson, V. Barlakas, S. Pfreundschuh, and S. Fox. Fast Radiative Transfer Approximating Ice Hydrometeor Orientation and Its Implication on IWP Retrievals. *Remote Sensing*, 14(7), 2022. ISSN 2072-4292. doi: 10.3390/rs14071594
2. S. Pfreundschuh, S. Fox, P. Eriksson, D. Duncan, S. A. Buehler, M. Brath, R. Cotton, and F. Ewald. Synergistic radar and sub-millimeter radiometer retrievals of ice hydrometeors in mid-latitude frontal cloud systems. *Atmospheric Measurement Techniques*, 15(3):677–699, 2022b. doi: 10.5194/amt-15-677-2022

3. S. Pfreundschuh, P. Eriksson, S. A. Buehler, M. Brath, D. Duncan, R. Larsson, and R. Ekelund. Synergistic radar and radiometer retrievals of ice hydrometeors. *Atmospheric Measurement Techniques*, 13(8):4219–4245, 2020. doi: 10.5194/amt-13-4219-2020
4. R. Ekelund, P. Eriksson, and S. Pfreundschuh. Using passive and active observations at microwave and sub-millimetre wavelengths to constrain ice particle models. *Atmospheric Measurement Techniques*, 13(2):501–520, 2020. doi: 10.5194/amt-13-501-2020
5. J. Hagen, K. Hocke, G. Stober, S. Pfreundschuh, A. Murk, and N. Kämpfer. First measurements of tides in the stratosphere and lower mesosphere by ground-based Doppler microwave wind radiometry. *Atmospheric Chemistry and Physics*, 20(4):2367–2386, 2020. doi: 10.5194/acp-20-2367-2020
6. D. I. Duncan, P. Eriksson, and S. Pfreundschuh. An experimental 2D-Var retrieval using AMSR2. *Atmospheric Measurement Techniques*, 12(12):6341–6359, 2019a. doi: 10.5194/amt-12-6341-2019
7. D. I. Duncan, P. Eriksson, S. Pfreundschuh, C. Klepp, and D. C. Jones. On the distinctiveness of observed oceanic raindrop distributions. *Atmospheric Chemistry and Physics*, 19(10):6969–6984, 2019b. doi: 10.5194/acp-19-6969-2019

Contents

Acknowledgements

I am deeply thankful to Patrick Eriksson for his supervision and, more importantly, for the initial impetus to pursue the idea that ultimately led to the results presented in this thesis.

I would also like to thank all scientists and staff at the Department of Space, Earth and Environment at Chalmers University of Technology who went out of their way to make this journey as smooth and enjoyable as I would let them.

I would also like to thank everyone in the scientific community around me, especially the members of our small clouds and precipitation group; those who shared their knowledge and insights, gave feedback (on short notice) and took part in discussions. In particular, I would like to thank Oliver Lemke, Manfred Brath and Prof. Stefan A. Buehler for scientific the collaboration and a place to stay in Hamburg. Anke Thoss and Nina Håkansson for the collaboration on cloud retrievals. Prof. Christian D. Kummerow and Paula J. Brown for welcoming me in Fort Collins.

And, most important of all, Julia, for simply being who you are.

Finally, who must not be forgotten are the thousands of invisible helpers without whom this thesis would not have been possible. All those who develop and publish scientific software and make it free to use. The developers of Python, IPython (Pérez and Granger, 2007) and Jupyter (Kluyver et al., 2016), Matplotlib (Hunter, 2007), seaborn (Waskom, 2021), xarray (Hoyer and Hamman, 2017), pandas (Wes McKinney, 2010; pandas development team, 2020), NumPy (van der Walt et al., 2011) and SciPy (Virtanen et al., 2020), pyvista (Sullivan and Kaszynski, 2019), Paraview (Ahrens et al., 2005), Satpy (Raspaud et al., 2021), and PyTorch (Paszke et al., 2019).

Part I

Introductory chapters

Chapter 1

Introduction

Earth observation satellites play an essential role in many scientific and meteorological applications. Their unique ability to provide frequent observations of large parts of the globe allows meteorologists to predict the weather and scientists to study the Earth and its atmosphere. Weather predictions and the monitoring of the climate are of considerable societal and economic value, a value that is likely to increase as the Earth continues to warm.

Observations from these satellites consist of measurements of electromagnetic radiation, which is either reflected or emitted from the Earth and its atmosphere. An example of such measurements is given in Fig. 1.1. Besides demonstrating its delicate beauty, this image of infrared radiation emitted from a low-pressure system over the Mediterranean sea contains valuable information about the physical state of the atmosphere. The colors in the image represent the intensity of the measured radiation from strong to weak using bright to dark colors. At this specific wavelength, the measured radiation stems from water vapor and clouds in the atmosphere. Dry air is less opaque and allows the satellite to sense radiation from lower down in the atmosphere. Since the temperatures in the lower atmosphere increase with decreasing height, the radiation in these regions is more intense. Bright colors thus identify regions of relatively dry air. Moist air, which is more opaque, emits radiation from higher up in the atmosphere where temperatures are lower and thus produces the moderate intensities that are visible in the image. The lowest intensities stem from high clouds, which are fully opaque and thus emit radiation at high altitudes and low atmospheric temperatures.

This example illustrates that an understanding of the processes that generate the infrared radiation observed by a satellite allows an observer to infer moisture content and the presence of high clouds in the atmosphere. Formulated more generally: If a component of the atmosphere interacts sufficiently strongly with radiation, it generates an electromagnetic signal that can be measured using a suitable sensor. By inverting the component's interaction with the observed radiation, these measurements can be used to infer the properties defining the interaction. This inversion process is called *the retrieval* and is required to relate satellite observations to the physical state of the atmosphere.

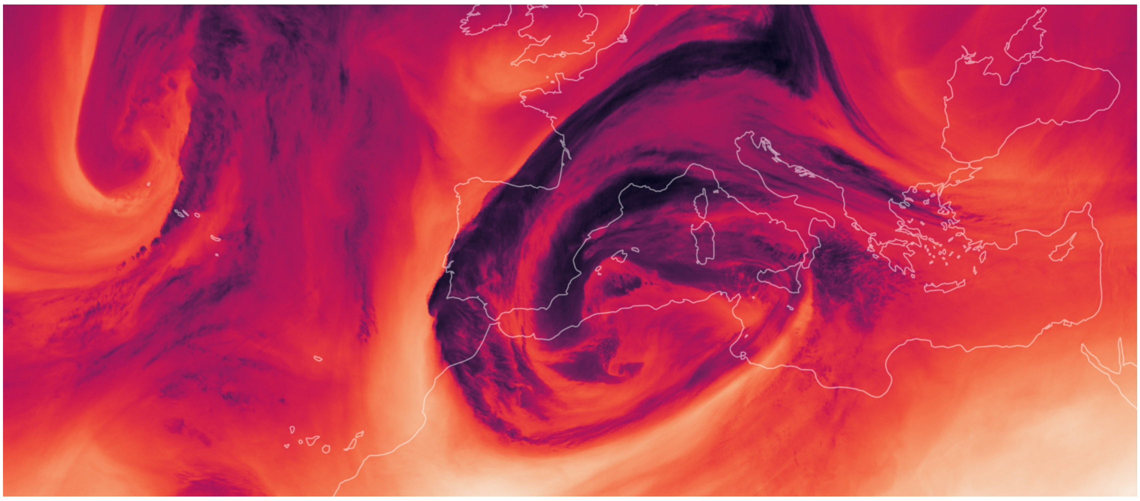


Figure 1.1: Water vapor (white to red) and high clouds (purple to black) over the Mediterranean as observed by the Spinning Enhanced Visible and Infrared Imager at a wavelength of 6.2 µm.

The subject of this thesis is the retrieval of properties of clouds and precipitation from satellite observations. These measurements are difficult due to the high variability in the appearance and composition of clouds and, secondly, the complexity of their interaction with radiation. At the same time, the important role that clouds and precipitation play in the weather and climate system makes these measurements highly relevant for science and society.

To set the scene for the discussion of satellite observations and retrieval techniques in the subsequent chapters, this introduction aims to provide an overview of the relevance of observing and measuring clouds and precipitation from space. Beginning with the societal significance of water, the discussion will move on to the scientific relevance of observing water as it moves through the atmosphere.

1.1 Water as resource

Water is essential to life on Earth. It is the bloodstream of the biosphere (Falkenmark et al., 2004) and a fundamental resource to all forms of human societies. The largest part of human freshwater consumption from lakes, rivers or ground water, so called *blue water* consumption, is used to irrigate crops, while domestic and industrial use play minor roles. Blue water consumption is distinguished from *green water* consumption, which refers to the part of precipitation which is involved in the photosynthesis process. Although rarely considered in consumption inventories, green water plays an important role in providing water for rain-fed agriculture, which accounts for 60 % to 70 % of the global food production, as well as in the sustaining of terrestrial ecosystems (Falkenmark et al., 2004).

Humans depend on water for food production both through irrigation from blue water flows and the provision of soil moisture by green water flows. Achieving food security has been recognized by the United Nations (UN) as a sustainable development goal (SDG, United Nations, 2018). The large contribution of food production to overall water consumption poses a challenge to the management of water resources. Diverting more blue or green water flows for the production of food reduces the amount of water that is available to sustain terrestrial and aquatic ecosystems. There is thus direct potential for conflict between SDG number two to end hunger and the SDGs 14 and 15 to sustain terrestrial and aquatic ecosystems.

At the same time, water requirements for the production of energy are expected to increase as fossil fuels are increasingly sourced from unconventional deposits, such as shale oil and gas, whose extraction consumes substantial amounts of water (Rosa et al., 2018). This is accompanied by a projected increase in hydropower (Zarfl et al., 2015) and the use of biofuels, both of which are reliant on and affect freshwater supplies. This emerging competition in water uses is recognized as the *food-energy-water* nexus (D'Odorico et al., 2018).

1.2 The hydrological cycle

Sustainable human water consumption depends on the replenishment of the sources from which water is obtained. Freshwater over land exists in the form of glaciers, ice sheets, lakes and reservoirs, snowpacks, wetlands, rivers, and a small part contained in the biomass. Water also exists within the land surface in the form of soil moisture, permafrost, and groundwater. The largest part of the water that is available on the surface of the Earth is stored in the oceans. Only a very small part is stored in the atmosphere. Most of that is in the form of water vapor, and only a minor fraction is contained in clouds in the form of liquid droplets or ice crystals (Abbott et al., 2019).

Despite containing only a tiny fraction of the global water reserves at any given moment, the atmosphere is responsible for essentially all of the water transport from oceans to land. An illustration of the atmospheric branch of the hydrological cycle is provided in Fig. 1.2. The largest flux in this part of the hydrological cylce is the influx of water vapor by evaporation from the ocean surface. Most of the water vapor that evaporates over the ocean never reaches the land but instead returns to it in the form of precipitation. Only about 10 % of the evaporation over oceans is transported to land, where it may form precipitation to replenish the water storages from which it is available for human consumption. A significant part of that precipitation evaporates again, allowing it to take part in the formation of precipitation further inland.

On the scale of river basins, the inflow of water through the atmosphere is the only sustainable source of freshwater (Falkenmark et al., 2004). Given the competing water requirements for the production of food and energy, as well as ensuring access to clean water and sustaining of ecosystems, it is clear that water management is required to avoid conflicts, minimize human suffering and reduce ecological damage. This management has to rely on the monitoring and modeling of the hydrological cycle. Since water fluxes within the hydrological cycle are highly variable in space and time and occur across continental scales, satellite observations are currently the only way through which continuous monitoring of the hydrological cycle can be realized.

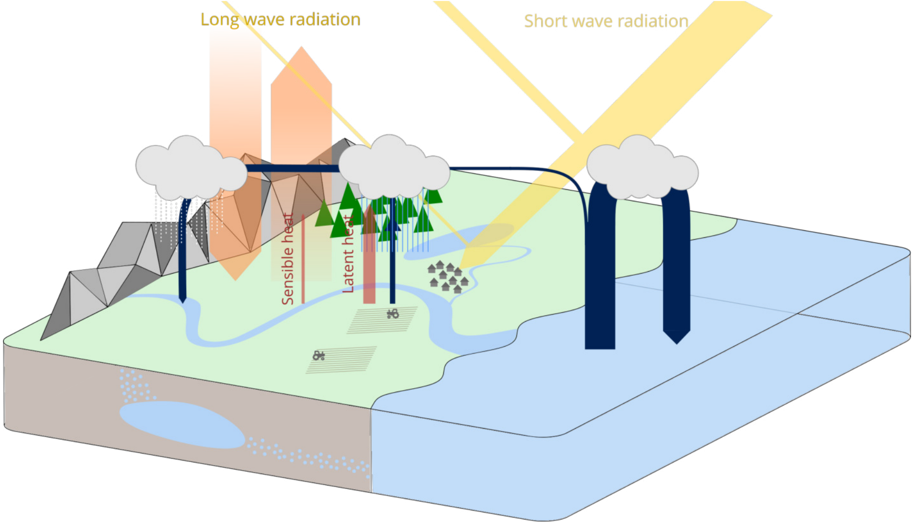


Figure 1.2: Illustration of the atmospheric branch of the hydrological cycle. Dark blue arrows show the flow of water into, within and out of the atmosphere. Their width approximately represents the extent of these fluxes. The semi-transparent arrows in orange, yellow and red show the energy fluxes in the Earth system, which are directly coupled to the hydrological cycle through the radiative effects of water vapor, clouds and the release of latent heat.

1.3 Weather and climate

Water in its various forms is a fundamental component of the Earth's weather (Stevens, 2013). When it evaporates, it takes up energy from the environment. This energy, the *latent heat*, is released when the air cools and the water vapor condensates. The release of latent heat acts as fuel for the development of vigorous convective storms. The water that is released by those storms in the form of precipitation can cause flooding and landslides.

The ability to predict the weather has considerable value for society, especially for high-impact events involving extreme wind speeds and precipitation. The availability of satellite observations has played an important role in the steady improvement of weather forecasts that occurred during the last three decades (Bauer et al., 2015). Weather forecasting systems use satellite data to determine the best estimate of the current state of the atmosphere, which is then evolved into the future to predict the weather. While advanced forecasting systems ingest satellite observations directly and are thus not dependent on satellite retrievals (Bauer et al., 2010), some systems use retrieved cloud properties for the initialization of short-range forecasts (de Haan and van der Veen, 2014; Benjamin et al., 2021). Besides that, retrievals of precipitation and cloud properties are used by weather forecasters to gain situational awareness and assess potential weather hazards.

At the same time, water also has a pivotal role in the climate system. As illustrated in Fig. 1.2, it is coupled to the Earth's energy budget through multiple processes. Water vapor is the most potent greenhouse gas and thus contributes strongly to the retention of radiative energy emitted from the Earth's surface. In the form of clouds, water both cools the Earth system through the reflection of solar radiation and warms it through the absorption of upwelling longwave radiation. Although the global effect of clouds is a pronounced cooling, the effect varies regionally and with the properties of the clouds. Evaporation, transpiration, and convection are important for the transfer of heat from the Earth's surface to the atmosphere. Since, on average, evaporation and transpiration have to be balanced by precipitation, the energy budget of the Earth is directly linked to precipitation (Trenberth et al., 2009).

Instead of a passive tracer of atmospheric dynamics, water must thus be understood as an active component of both the weather and the climate system linking these systems to the hydrological cycle. Because of the global extent and complex nature of these systems, satellite observations of the hydrological cycle are essential for climate science and meteorology.

1.4 A warmer planet

As a consequence of anthropogenic emissions of carbon dioxide, the Earth's climate is warming. As the Earth system adapts to the increased radiative forcing from greenhouse gases, its climate changes. These changes are not uniform but exhibit significant regional variability. Reliable regional predictions of the future climate are thus required to help societies adapt to them. However, the current understanding, as well as the ability to predict changes at regional scales, remains limited.

The hydrological cycle is changing together with the climate. Because of the ability of warmer air to hold more water, water vapor concentrations are increasing. An increase in evaporative demand is expected to lead to more frequent and severe droughts. At the same time, global precipitation is expected to increase. Extreme precipitation events are strengthening and becoming more frequent. At the scales of individual precipitation systems, precipitation increases with the capacity of warmer air to carry water vapor. Globally, however, evapotranspiration and thus also precipitation are constrained by the Earth's energy budget, which limits increases in global precipitation to a lower rate than that of individual storms (Collins et al., 2013). In addition to that, changes in the general circulation will modify the global distribution of precipitation.

As part of the hydrological cycle, clouds will change with the climate. Since clouds reflect incoming solar radiation and absorb outgoing long-wave radiation, these changes act as a feedback on the Earth's energy budget and will thus influence the response of the climate system to increased carbon dioxide concentrations. Whether an individual cloud exerts net warming or cooling on the climate system depends on its properties such as altitude, composition, and location. The representation of clouds remains a challenge for climate models and uncertainties in cloud feedbacks remain the primary source of uncertainty in predictions of the climate response to increased carbon dioxide concentrations in the atmosphere (Zelinka et al., 2020).

The understanding of climate change has progressed immensely during the last decade but remains insufficient to make confident predictions at regional scales. Such predictions are necessary for successful climate change adaptation. Satellite observations are essential for improving and validating climate models and monitoring changes in the climate system, making them a crucial guide for the passage into an uncertain, warmer future.

1.5 Why satellites?

Considering the importance of rain for a large range of human activities, it is not surprising that humans have been measuring precipitation and trying to improve their understanding of the weather system. The most direct and arguably intuitive way to measure precipitation is through the use of *rain gauges*. A rain gauge consists of a cavity that captures rain and from which the accumulated precipitation can be measured. First records of gauge measurements of rain date back to as long as 400 years BC (Strangeways, 2000). Today, gauge measurements are available from most populated regions across the globe. However, their density varies from country to country and is oftentimes tied to the local population density.

the main disadvantage of rain gauges is the highly localized character of their measurements, which limits their ability to accurately capture the spatial structure of precipitation on short time scales. moreover, because these measurements are performed by a large number of public and private entities across the globe, not all of them are easily accessible to scientists. effectively, only 1 of the surface of the earth is estimated to be within a distance of 5 from the nearest gauge measurement (Kidd et al., 2017).

Ground-based radars are another important technique for measuring precipitation over land. These radars send out beams of radiation, which are reflected by rain and snow. Precipitation within a radius of several hundreds of kilometers can be detected by measuring the amount of radiation that is reflected back to the radar. Ground-based radars are important for the real-time monitoring of precipitation over land. However, due to their high installation and maintenance costs, their spatial availability is irregular and typically follows the local population density. In addition to that, the availability and accuracy of radar measurements may be hampered by complex terrain as well as the distance from the radar location, which limits the utility of these observations for climatological applications.

The relevance of clouds and precipitation for the hydrological cycle has been known since before the times of Aristotle (Frisinger, 1972). Nonetheless, the systematic study of clouds has long remained a secondary focus of meteorology, and it took until the 1970s before the importance of interactions between clouds, weather, and climate put them at the forefront of atmospheric and climate research (Stephens, 2003). Traditionally, cloud observations were conducted as parts of routine meteorological observations performed on land and ships (Hughes, 1984). However, these observations were primarily qualitative, indicating only the presence of clouds, their type and an approximate height.

The principal drawback of the measurement techniques discussed above is their limited and irregular spatial coverage, which reduces their suitability for monitoring the global weather and climate. In contrast to that, satellites have the ability to provide observations across the globe at high temporal resolutions. Therefore, they ideally complement ground-based measurements and have become an essential source of information for a wide range of scientific, economic and societal activities.

Chapter 2

Satellite remote sensing of clouds and precipitation

The remote sensing of any atmospheric component is based on its interaction with electromagnetic radiation. By inverting this interaction, the measured radiation can be used to infer specific properties of this component. This chapter presents the physical processes underlying these interactions and relevant practical aspects of observing the atmosphere from space.

The first part of this chapter provides a brief introduction to radiative transfer theory, which can be used to describe the interaction of radiation with the atmosphere. Based on this, the principal characteristics of observations across the electromagnetic spectrum and their sensitivity to hydrometeors[1]are discussed. The focus is put on observations that measure radiation emitted from the sun or the Earth and its atmosphere, so called *passive observations*. The remaining part of this chapter then provides an overview of the types of different satellite observations that were used in the retrieval applications developed in this thesis.

2.1 Radiative transfer theory

Radiative transfer theory describes the propagation of a perfectly monochromatic beam of light through a medium. As the beam propagates, it interacts with the medium through electromagnetic and quantum mechanical processes. Radiative transfer theory provides a simplified model of these interactions, which is well suited for the conditions of passive remote sensing of the atmosphere. The following presentation is based on the more comprehensive texts by Mishchenko et al. (2002), Thomas and Stamnes (2002) and Wallace and Hobbs (2006).

Electromagnetic sensors measure the mean total energy that is transferred by electromagnetic radiation over a finite time interval. The way in which this energy is distributed across the components of the electric field is defined as its *polarization state*. Mathematically, the polarization state of the beam can be described using

---

[1] Hydrometeor is the collective term for the liquid and frozen particles that make up clouds and precipitation.

the four-dimensional Stokes vector

$$\boldsymbol{I} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} \tag{2.1}$$

The components I, Q, U, and V are called the Stokes parameters and have the unit of monochromatic energy flux per solid angle. The Stokes vector fully describes the state of a monochromatic beam of radiation to the extent that an electromagnetic sensor can measure it. Any electromagnetic measurement can thus be derived from knowledge of the Stokes vector at the position and orientation of the sensor. Radiative transfer theory describes how the Stokes vector changes as it propagates through a medium and thus allows the modeling of satellite observations of the atmosphere.

2.1.1 Interactions with matter

Radiative transfer theory distinguishes three fundamental processes through which matter interacts with radiation. These are the emission of electromagnetic radiation, its absorption and scattering.

Emission

At temperatures above absolute zero, all matter emits thermal radiation. Thermal radiation is produced when molecules transition from a quantum mechanical state of higher energy to one of lower energy causing the excess energy to be emitted in the form of radiation. The amount of radiation that a body emits varies with its material properties and temperature as well as the wavelength λ of the radiation.

In radiative transfer theory, emission from arbitrary matter is modeled using a material-dependent emissivity vector $\boldsymbol{e}$, which relates the emission of the material to that of a black body. A black body is an idealized body that absorbs all incoming radiation. The intensity of the radiation emitted by a black body at temperature T and wavelength λ is given by Planck's law

$$B(T, \lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}, \tag{2.2}$$

with c is the speed of light in vacuum, h the Planck constant and k the Boltzmann constant.

The emissivity vector $\boldsymbol{e}$ describes the amount of radiation that is emitted along an infinitesimal step ds along the propagation path of a beam:

$$\boldsymbol{I}_{\text{emitted}} = \boldsymbol{e} \cdot B(T, \nu)\ ds. \tag{2.3}$$

Eq. 2.3 provides a deceptively simple way to calculate the radiation emitted by arbitrary materials. It is clear that this formulation delegates much of the complexity of calculating the radiation emitted from a material to the calculation of the emissivity vector $\boldsymbol{e}$. In the atmosphere, $\boldsymbol{e}$ depends on the concentrations of gases, aerosols, and hydrometeors as well as the local temperature and pressure. A significant part of the difficulty of calculating thermal emission from the atmosphere arises from determining its emissivity.

Absorption

Absorption refers to the process of radiation being converted into internal energy of the matter it interacts with. Mathematically, this process is described by the absorption vector $\boldsymbol{\alpha}$, defined as the fraction of the beam's radiation that is absorbed along an infinitesimal step of length $\boldsymbol{ds}$ along its propagation path:

$$\boldsymbol{I}_{\text{absorbed}} = \boldsymbol{\alpha} \odot \boldsymbol{I}\ ds \tag{2.4}$$

Here $\odot$ denotes the element-wise product of the absorption vector and the Stokes vector $\boldsymbol{I}$ of the incoming radiation. Absorption may be understood as the inverse process of thermal emission. Formally, this is expressed by Kirchoff's law of radiation

$$\boldsymbol{\alpha} = \boldsymbol{\epsilon}, \tag{2.5}$$

which states that the absorption vector is identical to the emissivity vector defined in Eq. 2.3. This law is applicable to all matter in the atmosphere, given that it is in a state of local thermal equilibrium (LTE). LTE occurs when the density of matter is sufficiently high for the population rates of energy states above the ground state to be determined by thermal collisions rather than the absorption of radiation. This decouples the emission of radiation from the radiation field itself, allowing the simplified treatment of matter as thermal emitters whose emission rates are independent of the radiation field. LTE is a valid assumption for radiative transfer in the troposphere.

Scattering

Scattering describes the effect of inhomogeneities in the medium on the propagation of the beam. When a plane-parallel electromagnetic wave encounters such inhomogeneities, it is scattered in all directions. Mathematically, the scattering of a beam of light propagating in direction $\boldsymbol{n}$ into the direction $\hat{\boldsymbol{n}}$ is described by the phase matrix $\boldsymbol{Z}(\hat{\boldsymbol{n}}, \boldsymbol{n})$:

$$\boldsymbol{I}_{\text{scattered}}(\hat{\boldsymbol{n}}) = \boldsymbol{Z}(\hat{\boldsymbol{n}}, \boldsymbol{n})\boldsymbol{I}(\boldsymbol{n}) \tag{2.6}$$

Since parts of the energy transported by the beam are deviated from its propagation path, its intensity is reduced. As the beam propagates through the medium, its intensity is thus decreased by the effects of absorption and scattering. The combination of these two processes is referred to as *attenuation* or *extinction* and is described by the attenuation matrix $\boldsymbol{K}$, which is the sum of the absorption vector $\boldsymbol{\alpha}$ and the fraction of radiation scattered away from the propagation path:

$$\boldsymbol{K} = \begin{bmatrix} | & | & | & | \\ \boldsymbol{\alpha} & 0 & 0 & 0 \\ | & | & | & | \end{bmatrix} + \int_{\hat{\boldsymbol{n}}} d\hat{\boldsymbol{n}}\ \boldsymbol{Z}(\hat{\boldsymbol{n}}, \boldsymbol{n}) \tag{2.7}$$

The strength of the scattering interaction depends on the relation of the size of the inhomogeneities to the wavelength of the radiation. When the scale of the inhomogeneities is much smaller than the wavelength, the effects of scattering can often be neglected. As the relative size of the inhomogeneities increases, the strength of the scattering increases and its effects must be taken into account. The scattering interaction drastically complicates the modeling of the radiative transfer because it requires taking into account the radiation that is scattered into the line of sight from all directions. Further difficulty arises from the dependency of the scattering properties on the shape of the inhomogeneities, which makes it difficult to them for scattering materials with irregular shapes.

Scattering effects from hydrometeors need to be taken into account across most of the electromagnetic spectrum because of their comparably large sizes. Frozen hydrometeors also exhibit highly variable shapes, which makes the calculation of their scattering properties complex and uncertain. These factors make the modeling of satellite observations of hydrometeors a scientific and computational challenge.

2.1.2 The radiative transfer equation

The previous section introduced the fundamental interactions of radiation with matter and how they are described mathematically by radiative transfer theory. Combining the three processes of emission, absorption, and scattering, the change that a beam undergoes as it travels a distance $\boldsymbol{ds}$ along its propagation path through the atmosphere is described by the vector radiative transfer equation (VRTE):

$$\frac{d\boldsymbol{I}(\boldsymbol{n})}{ds} = -\boldsymbol{K}\boldsymbol{I}(\boldsymbol{n}) + \boldsymbol{\alpha} \cdot B_\nu(T) + \int_{\hat{\boldsymbol{n}}} d\hat{\boldsymbol{n}}\ \mathrm{Z}(\boldsymbol{n}, \hat{\boldsymbol{n}})\boldsymbol{I}(\hat{\boldsymbol{n}}). \tag{2.8}$$

The first term on the right hand side is the extinction term and represents the combined effects of absorption and scattering of radiation out of the propagation direction, both of which act to decrease the intensity of the radiation. The second term represents emission along the propagation path, with the emissivity vector $\boldsymbol{\epsilon}$ replaced by the absorption vector $\boldsymbol{a}$, according to Kirchoff's law of thermal radiation. The third term represents the radiation that is scattered into the line of sight. The two latter terms act to increase the intensity of the radiation in the beam.

2.2 Observations of hydrometeors

Radiative transfer theory describes the propagation of radiation through the atmosphere in terms of its radiative properties. These are represented by the absorption vector $\boldsymbol{\alpha}$, the extinction matrix $\boldsymbol{K}$ and the phase matrix $\boldsymbol{Z}$. In order to apply radiative transfer theory to gain an understanding of remote sensing observations of the atmosphere, its composition must be related to its radiative properties.

The radiative properties of the atmosphere are strongly dependent on the wavelength λ at which it is observed. The applications presented in this thesis make use of observations from the visible to the microwave region of the electromagnetic spectrum. To gain an understanding of the principal properties of these observations, it is practical to distinguish the following four different wavelength domains:

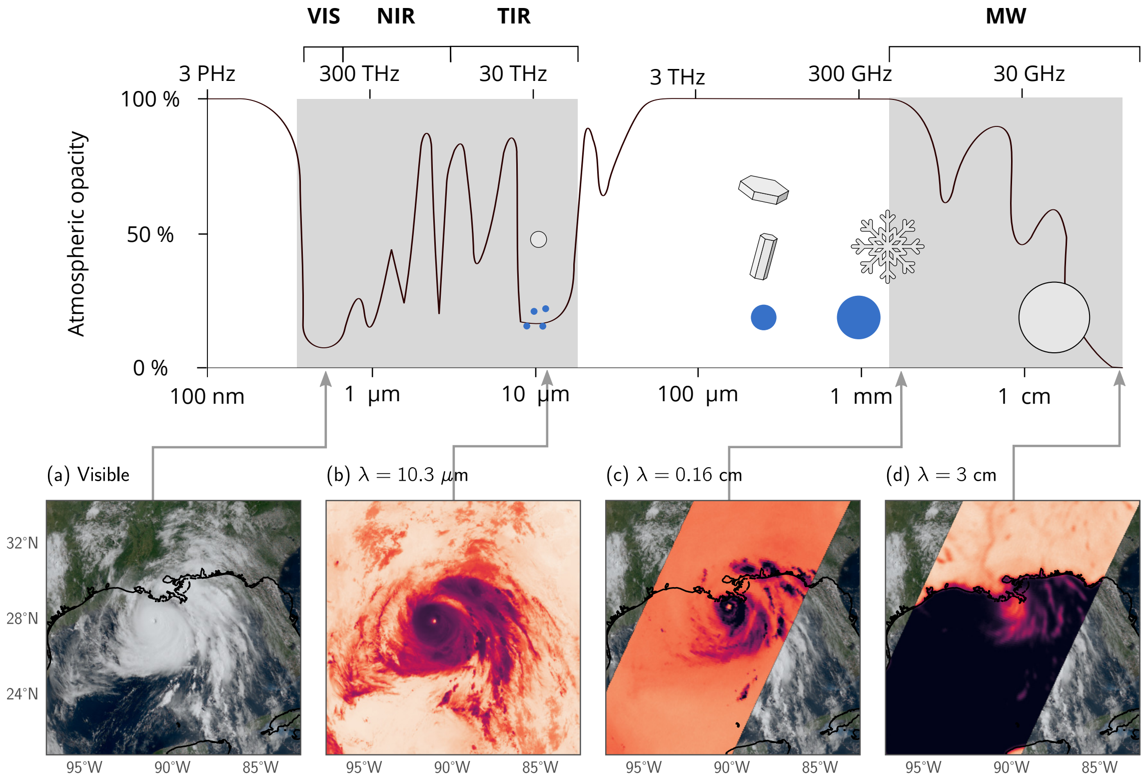


Figure 2.1: Overview of the electromagnetic spectrum used for hydrometeor retrievals. The black curve in the top panel shows the variation of the atmospheric opacity across the spectrum. Principal classes of hydrometeors are drawn at the wavelength corresponding to their approximate size. Liquid hydrometeors are drawn in blue, frozen hydrometeors in gray. The four lower panels show example observations of hurricane Ida at different wavelengths.

Visible (VIS): $350\,\text{nm} \leq \lambda < 720\,\text{nm}$

Near infrared (NIR): $720\,\text{nm} \leq \lambda < 2.5\,\mu\text{m}$

Thermal infrared (TIR): $2.5\,\text{nm} \leq \lambda < 20.0\,\mu\text{m}$

Microwave (MW): $1\,\text{mm} \leq \lambda < 1\,\text{m}$

An overview of these regions within the spectrum of electromagnetic radiation is provided in Fig. 2.1. The top panel shows a simplified graph of the variation of the opacity of a cloud-free atmosphere. The opacity is defined as the fraction by which the intensity of a beam is reduced as it propagates from the bottom to the top of the atmosphere. When viewed from a satellite, the opacity gives an indication of how far down into the atmosphere a sensor can 'see'. Generally, the satellite can see down to the surface where the opacity is low, while it is sensitive only to the the atmosphere where it is high. Fig. 2.1 also contains four examples of observations of hurricane Ida covering the electromagnetic spectrum, which will be used in the following to illustrate their principal characteristics.

Observations of the atmosphere are the result of the combined effects of gases, aerosols, and hydrometeors. It is helpful to decompose the contributions into a clear-sky component, which considers only the effects of molecules and aerosols, and a cloudy-sky component, which also considers the effects of hydrometeors. As a first approximation, the clear-sky component may be viewed as the background upon which the hydrometeors are observed.

2.2.1 The atmospheric background

Observations at short wavelengths differ from those at long wavelengths with respect to the origin of the observed radiation. In the visible and near-infrared regions, the largest part of the observed radiation stems from the sun. This is because, at these short wavelengths, black body radiation at typical atmospheric temperatures is small compared to the reflection of solar radiation. As the wavelength increases, the intensity of solar radiation decreases while that of a black body at atmospheric temperatures increases. At wavelengths exceeding approximately 2.5 µm, emission from atmospheric constituents dominates the intensity of reflected solar radiation. This threshold separates the thermal from the near-infrared. Solar radiation at these longer wavelengths can generally be neglected, and all observed radiation originates from the atmosphere itself or the Earth's surface.

The atmospheric opacity displayed in the top panel of Fig. 2.1 corresponds to the combined effects of absorption and extinction by gases in the atmosphere. Gases affect radiation mostly through absorption and emission, while scattering plays a role only in the visible range. The reason for this is the small size of the molecules compared to the wavelengths of the radiation.

There is little absorption and molecular scattering in the visible range, which makes the atmosphere transparent to the human eye. Therefore, Satellite observations see down to the surface of the Earth where it is not covered by clouds (see panel (a) of Fig. 2.1). In the near and thermal infrared, absorption increases but is concentrated in discrete absorption bands, which makes the opacity of the atmosphere highly variable across wavelengths. The principal gaseous absorbers in the infrared region are water vapor, carbon dioxide, and ozone. The banded structure of the atmospheric opacity gives rise to so-called *window channels*, which are channels in which the atmospheric opacity is low. An example of such a window channel is given in panel (b) of Fig. 2.1. Because of the low opacity in this channel, the radiation observed in the cloud-free regions of the image corresponds to thermal radiation from the Earth's surface. This radiation is intense because of the relatively high temperatures and the high emissivity of the land and ocean surfaces.

Molecular absorption also plays an important role in the microwave region[2] with significant contributions from water vapor, oxygen, nitrogen, and ozone. At the short-wavelength end of the region, the atmosphere is relatively opaque because of absorption from water vapor. The observations shown in panel (c) of Fig. 2.1 are located close to a strong water vapor absorption line at 183 GHz ($\lambda \approx 1.6\,\text{mm}$). Because the atmosphere is fully opaque at this wavelength, the observed thermal radiation from water vapor stems from higher up in the atmosphere than at channels with lower opacity. This causes the low intensities observed in the cloud-free regions

---

[2] For microwave applications it is common practice to refer to wavelengths in frequency units. We will follow this convention here but give the corresponding wavelength in parentheses.

of the observations compared to those at $\lambda = 10.3\,\mu\text{m}$. At the long-wavelength end of the microwave region, the opacity of the atmosphere is low, and the observations are again sensitive to the surface. As can be seen in panel (d) of Fig. 2.1, there is a very strong contrast between land and ocean surfaces. This is because of the different radiative properties of land and water surfaces at these wavelengths and the angle from which the observations are performed.

2.2.2 Physical and radiative properties of hydrometeors

The radiative properties of hydrometeors depend on their size and whether they are in the liquid or frozen phase. Because the dielectric properties of ice and water differ significantly across the electromagnetic spectrum, hydrometeors of similar sizes may have drastically different radiative properties at certain wavelengths. To provide an overview of the principal classes of hydrometeors with respect to the observation wavelength, they are displayed in Fig. 2.1 at the wavelengths corresponding to their approximate sizes. Liquid hydrometeors are identified using blue coloring, while white corresponds to frozen hydrometeors.

Among the smallest hydrometeors are liquid cloud droplets with typical sizes around 5 µm. At sizes of around 25 µm, water drops become large enough to fall out of the clouds. Small, precipitating liquid droplets that range in size from 25 µm to 250 µm are referred to as drizzle. Larger drops are classified as rain, whose typical sizes are around 1 mm but can become as large as 5 mm.

At temperatures below 0 °C ice crystals can form in the atmosphere. Their sizes range from 1 µm to 1.5 mm with typical sizes around 100 µm. Snowflakes are aggregates that form through the collision of ice crystals. These range in size from hundreds of micrometers to several centimeters. When snowflakes collide with liquid drops at temperatures below 0 °C, the liquid drop freezes upon the snowflake, causing it to grow. This process is called riming. Heavily rimed snowflakes are referred to as graupel and have typical sizes of around 1 cm. Finally, the largest hydrometeors are hailstones, which form only in strong thunderstorms and can reach sizes of up to 10 cm.

2.2.3 Cloudy sky

Due to their comparably large sizes, the scattering effects of hydrometeors need to be taken into account across most wavelengths from the visible to the microwave regions. In the visible and near-infrared there is little absorption from either water or ice. Hydrometeors thus mostly deviate radiation from its propagation path without significantly decreasing its intensity. Clouds observed at these wavelengths typically appear bright because the solar radiation scattered back towards the sensor is much more intense than the reflection from the surface.

Both water and ice are strongly absorbing in the thermal infrared region. Because the ambient temperature decreases with altitude in the troposphere, the intensity of the radiation that is emitted by a cloud is related to its altitude. The effect of this is clearly visible in panel (b) of Fig. 2.1. The clouds which appear uniformly white in the natural color composite exhibit more structure in the observations at

$\lambda = 10.3\,\mu\text{m}$. The lowest intensities are observed close to the center of the hurricane where the clouds extend highest into the atmosphere.

In the microwave region, absorption from ice is generally much weaker than that from liquid water. Frozen hydrometeors therefore interact with radiation primarily through scattering. However, due to the relatively large wavelengths of microwave observations, significant scattering effects are limited to snow, hail and graupel and frequencies above 50 GHz ($\lambda \approx 6\,\text{mm}$). Because they are only sensitive to large hydrometeors, only clouds that produce snow are visible in the observations at 183 GHz ($\lambda \approx 1.6\,\text{mm}$).

For passive observations the scattering from hydrometeors at frequencies below 50 GHz can be neglected. The principal interaction with hydrometeors is therefore thermal emission and absorption from liquid cloud droplets and precipitation. Since water surfaces have low emissivities at these wavelengths, emission from liquid hydrometeors causes a distinct signal in the observations. Over ocean, the thermal emission from the rain in the hurricane is clearly visible in the observations at 10 GHz shown in panel (d) of Fig. 2.1.

2.3 Satellite observations

The discussion in the previous section demonstrated that observations at different wavelengths contain different information on hydrometeors. Besides that, the potential of satellite observations to characterize hydrometeors from space depends on the type of sensor and its temporal and spatial resolution. Due to the technical complexity of operating specialized sensors in space, the availability of observations and their characteristics are ultimately determined by a trade-off between technical and financial constraints on one side and scientific value on the other. The following section briefly reviews the principal sensors types and satellite platforms that are commonly used for the remote sensing of hydrometeors together with the principal characteristics of the observations that they provide.

2.3.1 Sensor types

The above discussion of radiative transfer in the atmosphere focused on observations of radiation that originates either from the sun or the Earth and its atmosphere. Sensors that measure these types of radiation are called passive sensors because they do not emit any radiation themselves. *Active sensors*, on the other hand, emit radiation and measure how much of that radiation is scattered back to the sensor. The active sensors that are used for the remote sensing of hydrometeors are radars and lidars. These sensors measure the time of travel of the radiation, which can be used to infer the distance of the observed hydrometeors from the sensor. This allows active sensors to profile the atmosphere along the line of sight, leading to much higher vertical resolution than what can be achieved with passive sensors.

From an observational perspective, the principal difference between radars and lidars is the wavelength at which they operate. Lidars make use of radiation with relatively short wavelengths making them sensitive to scattering from molecules, aerosol, and

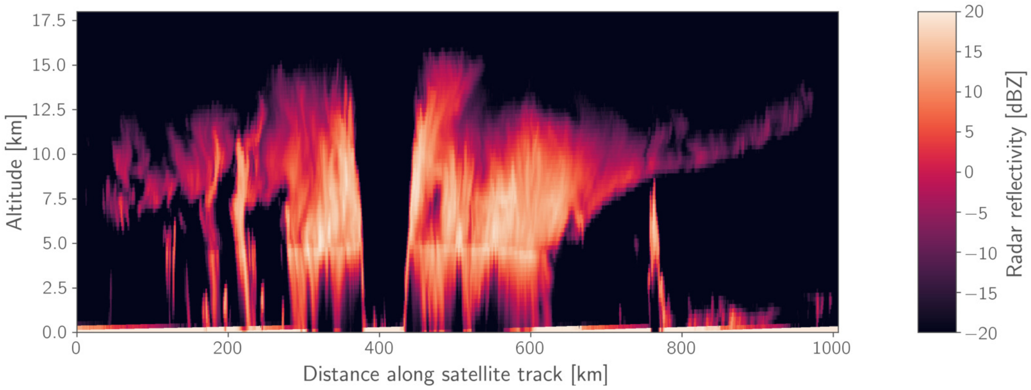


Figure 2.2: Radar observations of hurricane Nicole from 2016-10-12 at 17:55 UTC recorded by the CPR on the CloudSat satellite.

hydrometeors. In contrast to that, radars operate at microwave wavelengths, which makes them sensitive to scattering only from hydrometeors. Because they emit a powerful beam of radiation, active sensors are much more sensitive to the scattering from small particles than their passive counterparts.

An example of observations from the Cloud Profiling Radar (CPR, Tanelli et al., 2008) on the CloudSat satellite is shown in Fig. 2.2. The CPR operates at a relatively high frequency of 94 GHz, which makes it sensitive not only to precipitation but also clouds. For these specific observations, the satellite passed directly over the eye of the hurricane, which is clearly visible in the vertically resolved observations.

The disadvantage of active sensors is that the horizontal extent of their observations across the direction of movement, i.e., the width of their *swath*, is typically very low. This severely reduces the spatial coverage of the observations and leads to very long times between consecutive overpasses for a fixed position on Earth.

2.3.2 Resolution and coverage

In addition to the information content that observations can provide on hydrometeors, their ability to resolve spatial and temporal variability as well as how much of the globe they cover determine their utility for different applications.

The orbit in which a satellite is placed plays an important role in determining these characteristics. For earth observations, two principal classes of orbits can be distinguished: Geostationary and low-Earth orbits. In geostationary orbit, satellites orbit the Earth at the same angular velocity as the Earth rotates around its axis. This allows them to hover over the same position on the Earth's surface. Since the field of view of the satellite is constant, it can provide observations with high temporal resolution at all locations below the satellite. This makes these observations from these satellites ideal for near-real-time applications.

The disadvantage of geostationary platforms is their long distance from the Earth, which is around 36 000 km. Since the spatial resolution of microwave sensors is limited by the long wavelengths of the radiation, they are currently not deployed on these platforms. Therefore, geostationary satellites typically carry visible and infrared sensors, which can produce observations at very high spatial resolutions. Besides that, their orbit restricts them to locations over the equator. The resulting viewing geometry causes the resolution to degrade with increasing latitudes and makes the observations unsuitable for observations at very high latitudes.

Polar orbits are a common type of low-Earth orbit used for Earth-observing satellites. In a polar orbit, the satellite passes above or nearly above the poles. The rotation of the Earth between consecutive orbits allows these satellites to achieve global coverage. The low altitude of these orbits ($300 - 1000\,\text{km}$) makes them suitable also for microwave sensors but decreases the spatial coverage of the sensor. The swaths of sensors in low-Earth orbits are typically limited to a few thousand kilometers in width. Depending on the exact size of the field of view and the location on Earth, the time between consecutive overpasses for a fixed position can be as high as 12 h.

The fields of view of the Advanced Baseline Imager (ABI, Schmit et al., 2018) on the Geostationary Operational Environmental Satellite (GOES) 16 and the Global Precipitation Measurement (GPM) Microwave Imager (GMI) on the polar-orbiting GPM core observatory are shown in Fig. 2.3. The observations from both satellites are projected onto the corresponding location on the Earth. While the geostationary satellite can permanently observe a large part of the hemisphere below it, the polar-orbiting satellite sees only a comparably thin stripe of the globe.

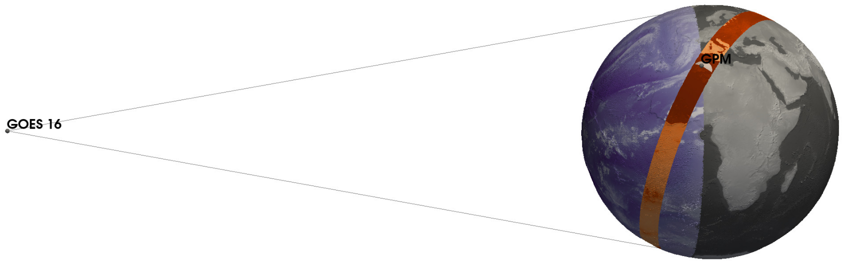


Figure 2.3: Viewing geometries of geostationary and polar-orbiting satellites drawn to scale. The illustration shows geo-located satellite observations from the GOES 16 geostationary satellite in purple and from the polar-orbiting GPM core observatory in orange. The spheres mark the instantaneous satellite positions.

2.3.3 Synergies

Because of the specific strengths of different sensor types and satellite platforms, leveraging the full potential of available satellite data often involves exploiting synergies between different observation types. An example of this is the Global Precipitation Measurement (GPM, Hou et al., 2014). This joint mission led by NASA and the Japanese Space Agency aims to provide accurate, global measurements of precipitations with a temporal resolution of a few hours. The mission comprises a dedicated precipitation satellite, the core observatory, which carries a passive microwave sensor, the GPM Microwave Imager (GMI), and a precipitation radar. The combination of the observation from these two dedicated precipitation sensors is used to produce the most accurate space-borne measurements of precipitation that are currently available.

However, because of the narrow swath of the radar, the measurements alone are too sparse to be useful for many applications. The GPM mission overcomes this by employing an opportunistic constellation of passive microwave sensors. This constellation consists mostly of operational, meteorological sensors, whose primary purpose is the provision of observations for weather forecasting. However, due to their reliance on microwave observations, they are also sensitive to precipitation. The high-accuracy retrievals from the combined observations afforded by core observatory are used as a reference standard for the development of the retrievals for the passive sensors of the GPM constellation. These retrievals are the subject of two of the studies included in this thesis.

The GPM constellation currently comprises nine sensors, in addition to GMI. A snapshot of these sensors in orbit is shown in Fig. 2.4. From the satellite swaths drawn on the surface of the Earth, it becomes clear that the constellation greatly increases the spatial coverage of the measurements. The crossing of the swaths of GMI and the Advanced Technology Microwave Sounder (ATMS) on Suomi NPP shows the complementary properties of the different sensors of the constellation. GMI has a relatively narrow swath and high resolution whereas the swath of ATMS is almost three times as wide but also has a significantly lower spatial resolution.

The retrievals from the passive microwave sensors are finally combined with observations from geostationary satellites and rain gauges to provide global precipitation measurements with half-hourly resolution. The resulting Integrated Multi-Satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG, Huffman et al., 2020) product is widely considered the state of the art of instantaneous, global precipitation measurement.

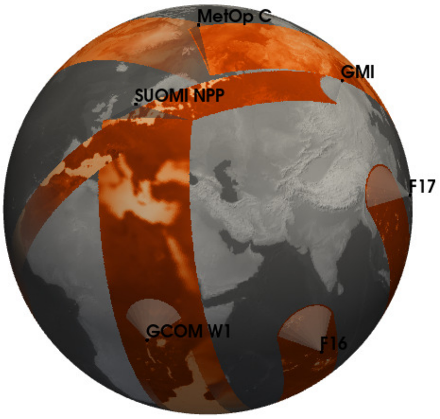


Figure 2.4: Currently operational passive microwave sensors of the GPM constellation in orbit around the Earth on 2022-04-19 11:26:00 UTC. Black spheres show the location of the satellites and their names. Orange coloring on the Earth's surface show the observations closest to 89 GHz for each sensor. Note that only six of the currently ten operational sensors are visible.

Chapter 3

Machine learning for remote sensing retrievals

Radiative transfer theory describes how radiation that interacts with hydrometeors produces an electromagnetic signal in satellite observations. Since the purpose of the observations is to provide information on the physical properties of hydrometeors, a way must be found to relate the observations to those properties. This process is referred to as retrieval. This chapter presents the mathematical formulation of the retrieval problem and provides a brief introduction to the machine-learning-based methods to solve it.

3.1 The retrieval problem

Mathematically, the problem of determining properties of hydrometeors from satellite observations can be formulated as finding a state vector[1] $\boldsymbol{y} \in \mathrm{R}^n$ that describes the physical properties of the atmosphere from a vector of observations $\boldsymbol{x} \in \mathrm{R}^m$. This so-called *retrieval problem* can be solved using the mathematical framework of inverse problem theory. In inverse problem theory, the underlying assumption is that the observations $\boldsymbol{x}$ are generated by a physical processes, which can be described using a forward model function $f : \mathrm{R}^n \rightarrow \mathrm{R}^m$, but may be affected by a random error $\boldsymbol{\epsilon} \in \mathrm{R}^M$:

$$\boldsymbol{x} = f(\boldsymbol{y}) + \boldsymbol{\epsilon} \tag{3.1}$$

An exact solution to this problem does not exist. Because of the random noise $\boldsymbol{\epsilon}$, the equation can never be strictly satisfied. Furthermore, there usually exist many states $\boldsymbol{y}$ that produce similar measurements $\boldsymbol{x}$, thus rendering the problem underconstrained. Since it does not admit a unique solution, the retrieval problem is *ill-posed*.

---

[1] The notation for the state and observation vectors was deliberately chosen to be the opposite of the conventions of inverse problem theory to make it consistent with the conventions used in machine learning.

Although an exact solution to the retrieval problem is not possible, Bayesian statistics provides a framework to extract the available information from the observations. In the Bayesian approach, both the observations $\boldsymbol{x}$ and the atmospheric state vector $\boldsymbol{y}$ are assumed to be random variables. Furthermore, it is assumed that $\boldsymbol{y}$ is distributed according to a known *a priori* distribution $p(\boldsymbol{y})$. Instead of a single, unique state $\boldsymbol{y}$, the solution of the problem then becomes the *a posteriori* (or *posterior*) distribution $p(\boldsymbol{y}|\boldsymbol{x})$ of the atmospheric state. According to Bayes theorem, the a posteriori distribution for given observations $\boldsymbol{x}$ is given by

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})}{p(\boldsymbol{x})}. \tag{3.2}$$

Traditionally, Bayesian methods for solving the retrieval problem make use of the right-hand side of Eq. 3.2 to approximate the posterior distribution. The approach pursued in this thesis is to learn the posterior distributions $p(\boldsymbol{y}|\boldsymbol{x})$ directly from data. How this can be done is the topic of the remainder of this chapter.

3.2 Machine learning

The field of machine learning is concerned with the development of algorithms that can learn to solve computational tasks from data. More specifically, the methods considered here solve the problem of *supervised learning.* In supervised learning, the task is to learn how to produce a certain output $\boldsymbol{y}$ from inputs $\boldsymbol{x}$ given a set of pairs $\{(x_i, y_i)\}$ for $i = 1, \ldots, N$ of specific input values $\boldsymbol{x}_i$ and corresponding output values $\boldsymbol{y}_i$. In general, the inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$ may represent anything from simple numbers to images or texts. For the applications considered here, the inputs are typically satellite images, and the outputs are the corresponding physical quantities to be retrieved.

Machine learning has gone through a wave of immense popularity during the 2010s triggered by the success of *deep learning* methods in addressing a range of computational problems in computer vision, natural language processing, and artificial intelligence. The idea behind deep learning is to construct increasingly powerful models from a hierarchy of simple models. Given sufficient data, these models can learn highly complex relationships and require less guidance from human experts to achieve good performance.

3.2.1 Shallow machine learning

Although the focus of this thesis is deep learning, this section will use a very shallow model to illustrate the basic principles of machine-learning-based remote sensing retrievals. The considered example tries to estimate rain at the surface from infrared (IR) observations at a wavelength of $\lambda = 10.3\,\mu\text{m}$. The training data, which contains of the in- and outputs from which a solution to the retrieval problem should be learned, is derived from co-locations of observations from the GOES 16 geostationary satellite and the combined radar and passive microwave retrievals from the GPM core observatory (see Sec. 2.3.3).

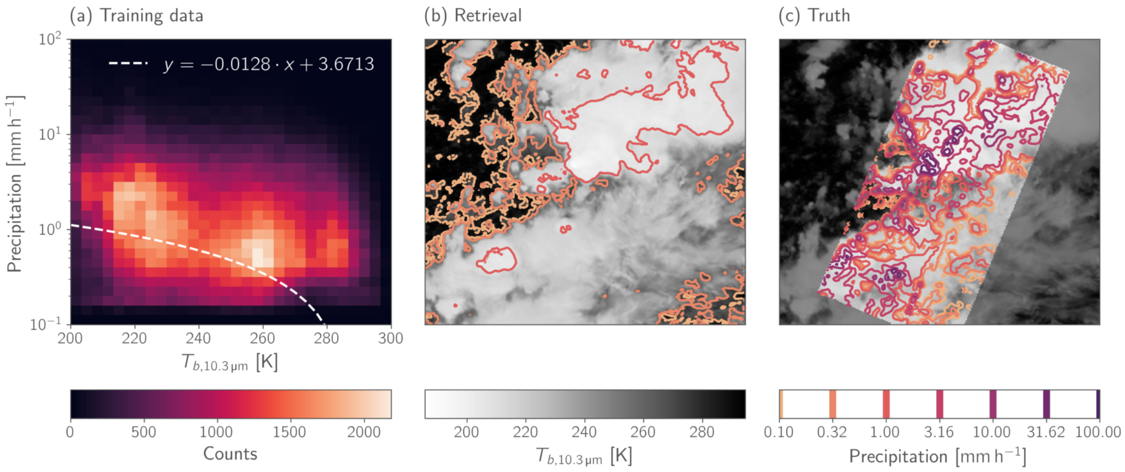


Figure 3.1: Example of a very simple precipitation retrieval. Panel (a) shows a density plot of the distribution of the input-output pairs, which consist of IR brightness temperatures and surface precipitation rates. The white line shows the linear relationship between observations and rain rate obtained by linear regression. Panel (b) shows retrieved precipitation rates when the learned relationship is applied to real observations. Panel (c) shows the true precipitation rates as present in the training data. Because of the narrow swath of the GPM precipitation radar the precipitation is not known across the full scene.

The input x for this particular example corresponds to the measured radiation at a single pixel of the geostationary observations while the output y is given by the rain rate measured by the GPM core observatory. The retrieval assumes a linear relationship between the observations x and the rain rate y in order to determine the precipitation corresponding to the geostationary observations:

$$y = ax + b \tag{3.3}$$

The two parameters of the linear regression model are the slope a and the intercept b. They can be determined by finding the values that minimize the squared error between the predictions and the true rain rates.

Panel (a) in Fig. 3.1 shows the distribution of the training data together with the learned relationship between input and output data. This view reveals that the relationship is not very robust, which becomes even clearer when the retrieval results (panel (b)) are compared to the reference precipitation (panel (c)). The linear regression model simply predicts rain almost everywhere where there is a cloud but completely fails to reproduce the structure in the reference precipitation.

Given the bad performance of the simple linear regression model, it is natural to ask how a better-performing model can be found. In general, there are three dimensions along which a machine learning model can be improved:

Model expressivity: The model used in this example can only learn linear relationships between input and output. The range of mathematical functions that a model can represent is referred to as its *expressivity*. Increasing the model expressivity enables it to learn more complex relationships between the input and output.

Amount of input information: The scatter plot in panel (a) of Fig. 3.1 shows that the problem of retrieving precipitation from infrared (IR) observations is highly degenerate. For any value of the observations, the training data contains a wide range of different precipitation values. This makes it impossible to assign a unique 'right' precipitation rate to an observed brightness temperature. The degeneracy can be decreased by adding more information to the input. One way to achieve this, for example, is to add observations from other channels or information from neighboring pixels. Depending on the problem, this may provide more flexibility to distinguish different precipitation rates to the model.

Amount of data: A machine learning model can only learn relationships that are sufficiently well represented in the training data. For more complex retrievals than this one, larger amounts of training data are crucial for good retrieval performance.

Traditionally, the difficulty in improving a machine learning model was that it is not sufficient to just increase the model expressivity or the amount of input information. Both measures increase the flexibility of the model and may thus cause it to *overfit*. Overfitting occurs when the model learns spurious relations from the training data, which are not actually representative of the true relationship between input and output. This typically causes predictions on unseen data to deteriorate drastically. Overfitting can be counteracted by artificially reducing the expressivity of the model, a process that is called *regularization*, as well as by increasing the amount of data that is used to train the model. However, the time and computational resources that are required for training typically increase with the amount of data. If the increase is too rapid, the amount of data that a machine learning model can be trained on may be limited.

Because of this, machine learning models traditionally depend on careful tuning to the task at hand. Especially the trade-off between model expressivity and the amount of data that can be used to train it required a multi-disciplinary approach that combined domain-specific knowledge with machine-learning experience to develop a good-performing model. A particular difficulty of shallow models was their inability to handle large inputs with low information density, such as the pixels from an image. Providing a shallow model with too many inputs typically makes the models prone to overfitting and more difficult to train. This gave rise to the practice of *feature engineering*, which consists of manually designing input variables that encode high-level information in task-specific input features.

It should be noted that the linear regression model used here is extremely simple and that there exist shallow machine learning methods that would likely work much better for this example. Nonetheless, the application is not totally unrealistic, given that it is similar to early precipitation retrievals from geostationary observations (Vicente et al., 1998). To overcome the limitations of the linear regression, the model from Vicente et al. (1998) predicts precipitation in log-space and uses weather model data to post-process the retrieval results. Modified versions of this retrieval are still in operational use today (de Siqueira and Vila, 2019).

3.2.2 Deep learning

The example discussed above illustrated the limitations of very shallow machine learning methods that deep learning aims to overcome. The essential promise of deep learning is that it decouples the dimensions along which a machine learning model can be improved. Deep learning models achieve high expressivity by stacking a large number of simple models and maximizing the amount of input data. Since the training time of deep learning methods scales linearly with the amount of available data while the required memory remains constant, the increased model expressivity can be balanced with large amounts of training data. This enables deep learning models to learn highly complex relationships from data without the risk of overfitting.

3.2.3 Neural networks

Seen from a more general perspective, the linear regression model considered above is just a transformation from the input $\boldsymbol{x}$ into the output $\boldsymbol{y}$ with a set of learnable parameters, in this case, the slope and intercept of the regression. A neural network is just a sequence of such parametrized transformations, which are referred to as *layers*. Each layer transforms an input vector into an output vector, which is passed on to the next layer. The exact form of each transformation can be adapted through its parameters, which is done by minimizing an error metric over the training data.

Defined in this way, every neural network also satisfies the definition of a layer. This allows any neural network to be used as a component within a larger network.

The multi-layer perceptron

Extending the simple linear regression model to accommodate for multiple in- and outputs yields one of the most fundamental transformations used in neural networks. It is typically referred to as *linear* or *fully-* or *densely-connected* layer. Mathematically, this layer implements an affine transformation of the input vector $\boldsymbol{x}$ of the form

$$\boldsymbol{y} = \boldsymbol{W}x + \boldsymbol{b} \tag{3.4}$$

where $\boldsymbol{W} \in \mathrm{R}^{m \times n}$ is a matrix of weights $W_{i,j}$, and $\boldsymbol{b} \in \mathrm{R}^m$ a vector of biases.

Fully-connected layers are typically used in combination with a non-linear function that is applied element-wise to its inputs. These functions are called *activation functions*. Without them, stacking fully-connected layers does not increase the expressivity of the neural network, which in this case depends only on the number of input and output variables.

There exist a wide range of activation functions. One of the currently most commonly used activation functions in deep neural networks is the Rectified Linear Unit (ReLU).

$$\mathrm{ReLU(x)} = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{3.5}$$

It should be noted, however, that neural network architectures are an area of active research and several alternatives and modifications of the ReLU activation have been proposed.

A neural network consisting of one or several layers of fully-connected layers followed by activation functions is typically referred to as multi-layer perceptron (MLP). Since all numeric inputs can be brought into vector form and thus fed into an MLP, the MLP is one of the most fundamental forms of neural networks.

Convolutional neural networks

Although MLPs can be applied to image data directly by flattening the input into a single vector, this approach was found to not work well in practice. The principal reason for this is that the number of parameters in the model becomes very large, making it prone to overfitting and the training more difficult.

An important technique in the application of neural networks to image data was the introduction of the convolutional layer, which replaces the matrix multiplication of the fully-connected layer by a convolution with a learnable filter kernel. Let $\boldsymbol{X} = [X_{i,j}]$ for $i = 1, \ldots, M, j = 1, \ldots, N$ be an input image of size $M \times N$ and $\boldsymbol{K} = [K_{i,j}]$ for $i = -k \ldots k, j = -k, \ldots, k$ a given convolution kernel of size $(2k+1)(2k+1)$. In two dimensions, the convolution operation $\boldsymbol{Y} = \boldsymbol{X} * \boldsymbol{K}$ of $\boldsymbol{X}$ and K used in machine learning is defined as

$$Y_{i,j} = \sum_{l=-k}^{k} \sum_{m=-k}^{k} X_{i+l,j+m} K_{i+l,j+m} \tag{3.6}$$

This operation corresponds to a filter mask whose response is calculated for all possible locations in the input image by sliding the filter window across the width and height of the image. Since digital images, as well as satellite observations, contain information also along a spectral dimension, the images actually correspond to three-dimensional tensors with an additional channel dimension in addition to the two spatial dimensions. To account for this, the kernel and sum in Eq. 3.6 are extended across all channels of the input image.

An illustration of the convolution operation for an input image with multiple channels is shown Fig. 3.2. The input corresponds to a rank-three tensor with one channel and two spatial dimensions. The kernel is limited in its extent in the spatial dimensions but extends over all channels of the input. A two-dimensional output tensor is obtained by calculating the kernel activation for all possible locations in the input tensor. A convolutional layer consists of a set of convolution kernels. The two-dimensional filter responses of each kernel are stacked along the channel dimension to produce a rank-three tensor as output of the layer.

In a convolution layer, the convolution operation is typically combined with the application of a bias term to each element in the output tensor. The learnable parameters of a convolution layer thus correspond to the elements of each of its kernels as well as the bias terms for each output.

Since the convolution transformation is linear, it is a special case of the fully-connected transformation of the corresponding flattened input image. However,

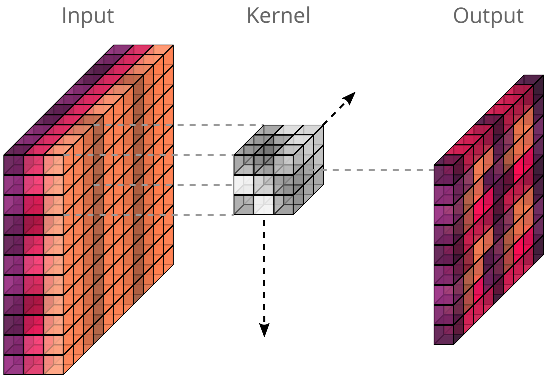


Figure 3.2: Illustration of the convolution operation. The output of the convolution operation is obtained by applying the filter kernel sequentially to every possible position in the input image.

in the fully-connected case, each pixel in the output image would be connected by independent parameters to all pixels in the input image, whereas for the convolutional layers, the parameters are shared between all pixels in a given channel of the output image. This greatly reduces the number of learnable parameters of the convolutional layer compared to a fully-connected one.

The convolution encodes two important properties of images into the neural network model, which are translation invariance and locality. Translation invariance refers to the property of the convolution operation that its output is independent of the relative location within the image. It is clear that this is a suitable assumption for many image processing tasks since features that identify an object should not be affected by their position in an image. The property of locality refers to the fact that each output of a convolution layer is determined only by a limited, connected subset of pixels of the input image, the number of which depends on the kernel size k. This encodes the assumption that for understanding the content of a given image region only neighboring pixels should be relevant.

Convolutional neural networks are neural networks built-up of convolution layers. These networks typically combine convolutional layers with downsampling layers that successively decrease the size of the input image, which allows these networks to combine information across the spatial dimensions of the image.

The motivation for this is to allow the model to learn a hierarchy of transformations from low-level information on the pixel scale to a reduced amount of high-level information that is relevant to the task that the network is being trained to solve.

Training

The basic training approach for neural networks is similar to that of the fitting of the parameters of a linear regression model, which are found by minimizing a certain loss criterion over the training data. For linear regression, this is typically the mean squared error, which can also be used to train neural networks. In contrast to linear regression, however, the parameter optimization for the neural network can not be solved explicitly. Instead, neural networks rely on iterative minimization of the loss function using gradient descent optimization.

While naive gradient descent would require traversing the whole training dataset to compute the gradients required to perform a single descent step, it is in practice sufficient and even beneficial to only use a small subset of samples from the training data. This modification of traditional gradient descent, which is known as *stochastic (mini-batch) gradient descent*, allows neural networks to scale to very large datasets. In addition to that, the stochasticity of the calculated gradients prevents the optimization process from getting stuck in local minima of the cost function.

Uncertainty in machine learning

The ill-posed character of the retrieval problem discussed in Sec. 3.1 leads to significant uncertainties in most remote sensing retrievals. In addition to that, the machine learning approach gives rise to two additional sources of uncertainties in the retrieval.

The total error that a machine learning model makes in its prediction is its *predictive uncertainty.* The predictive uncertainty can be decomposed into three independent sources. The first one, referred to as *epistemic uncertainty*, is the uncertainty caused by a lack of training data. Its defining property is that it can be reduced by collecting more data to train the model on. The second type is called *aleatoric uncertainty.* This uncertainty refers to the uncertainty that cannot be reduced by increasing the amount of data that the model is trained on. In the context of satellite observations, this uncertainty is caused by the inability of the observations to fully determine the observed atmospheric state. It is thus a consequence of the ill-posed nature of the retrieval problem discussed in Sec. 3.1.

An illustration of the concepts of aleatoric and epistemic uncertainty is provided in Fig. 3.3. The plot shows the joint distribution of observations and the retrieval target for a hypothetical retrieval scenario. The filled contours in the background show the actual joint distribution that relates the observations to the targets. The grey markers show the available training samples. If these samples were to be used to train a neural network model, the uncertainty in the left part of the graph would be dominated by the aleatoric uncertainty, because the spread of the underlying, true relationship is well represented in the training data. In the right part of the graph, however, the training samples are too few to capture the relevant features of the relationship and uncertainties would be dominated by epistemic uncertainty.

The final source of uncertainty is caused by differences between the training data and the data that the model is actually applied on. It is typically referred to as *covariate shift.* It is obvious that when the wrong data is used to train the model, it is likely to produce wrong results. However, subtle changes between training data and the actual data that the model is applied on are not uncommon and can increase the predictive uncertainty of the model.

3.2.4 Revisiting the precipitation retrieval

To provide a demonstration of the capabilities of the deep learning methods, Fig. 3.4 revisits the retrieval example introduced above and displays the results from two

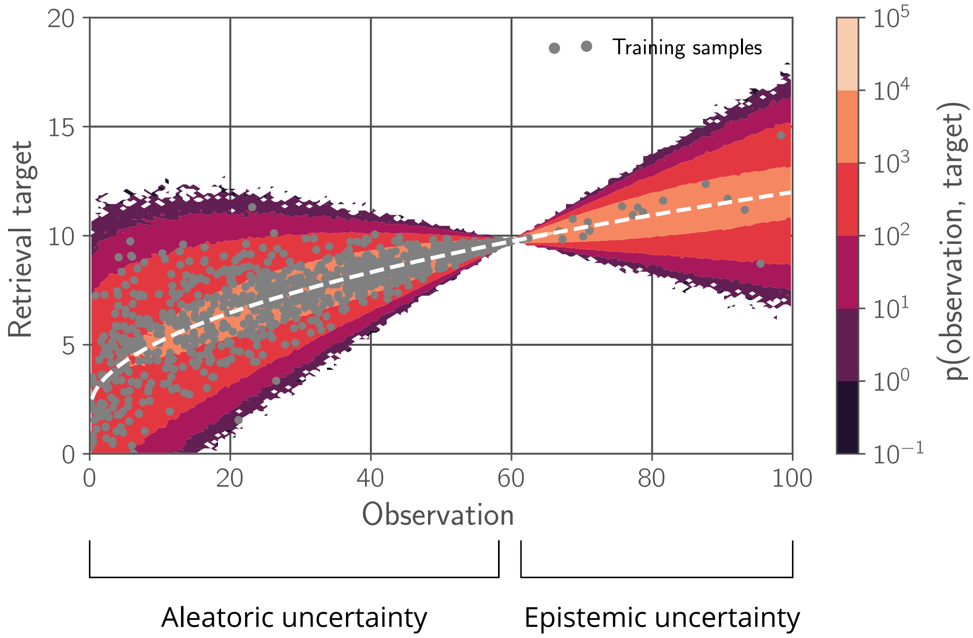


Figure 3.3: Illustration of aleatoric and epistemic uncertainty. The graph displays the relationship between observations and retrieval target for a hypothetical retrieval scenario. In the left part of graph the uncertainties are dominated by aleatoric uncertainty, whereas in the right part they are dominated by epistemic uncertainty.

additional retrievals, which have been trained using the same data.

The first one is based on an MLP, which uses the same single-pixel input as the linear regression model. The results from this retrieval are shown in panel (b). Compared to the reference precipitation shown in panel (d), the results show modest improvements in the structure of the precipitation, but its spatial extent is still strongly overestimated. Since the MLP uses only information from a single input channel, it is severely limited in the relations that it can learn. It is therefore not surprising that it does not work much better than the linear regression model.

To explore the full potential of deep learning methods, the second model uses a CNN to retrieve precipitation using the full input image and all 16 channels of the geostationary observations simultaneously. These results show clear improvements both in the structure of the retrieved precipitation as well as its spatial extent.

This example clearly demonstrates the ability of deep learning techniques to improve remote sensing retrievals. However, it should also be noted that the computational effort required to train these models increases by orders of magnitude. The simple linear regression model used here takes at most a few seconds to train. The training of the MLP and CNN models, on the other hand, takes hours respectively days and uses dedicated computing hardware. This is in general not an issue because the training is a one-time effort and the evaluation of the model is much faster, but nonetheless a fact that deserves consideration.

The results in Fig. 3.4 show that even the more advanced retrieval model fails

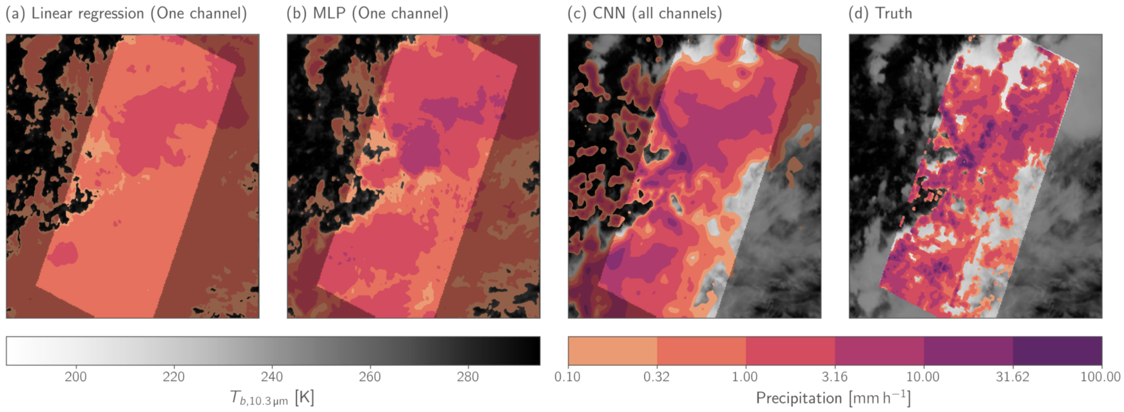


Figure 3.4: Retrieved precipitation from a linear regression model (panel (a)), an MLP applied separately to each input pixel (panel (b)) and a CNN that predicts precipitation for the full input image using all available channels (panel (c)). The reference precipitation measured by the combined radar and passive microwave retrievals of the GPM core observatory are shown in panel (d).

to truthfully reproduce the precipitation in the reference data. Since VIS and IR observations from geostationary satellites are only sensitive to the upper parts of the clouds, the inputs provided to the retrieval are only indirectly related to precipitation at the surface. The ill-posed character of the retrieval problem dictates that any specific prediction from a neural network will be wrong. Given this fundamental nature of remote sensing retrievals, the next section explores how neural networks can learn how wrong they expect to be.

3.3 Handling uncertainty in remote sensing retrievals

One of the issues that this thesis aims to address is the quantification of retrieval uncertainties in machine-learning-based remote sensing retrievals. The two methods that were explored in the appended studies fall into the category of probabilistic regression methods. This means that they account for the aleatoric uncertainty in the prediction but neglect epistemic uncertainty and covariate shift. Below, these two approaches are presented followed by a discussion of other approaches for quantifying uncertainties in neural network predictions.

3.3.1 Probabilistic regression

Aleatoric uncertainty arises due to ambiguous samples in the training data. The fact that these ambiguities are represented in the training data means that they can be predicted by a suitably trained neural network model. To allow for this, the deterministic formulation of regression, i.e. to predict a single value $y = f(\mathbf{x})$, is replaced by a probabilistic formulation, which aims predict the probability distribution $p(y|\mathbf{x})$ of y given the input vector $\mathbf{x}$.

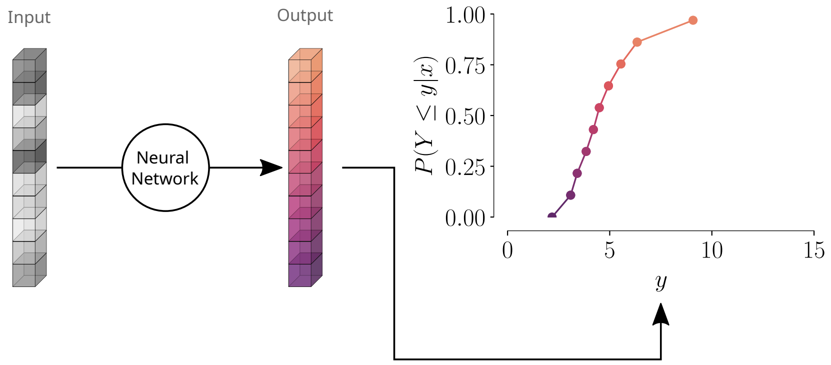


Figure 3.5: Basic principle of QRNNS. QRNNs predict the distribution of a scalar output variable y using a vector of quantiles from which the an approximation of the CDF of the distribution can be constructed.

Quantile regression neural networks

Quantile regression neural networks (QRNNs) predict the distribution $p(y|\boldsymbol{x})$ for a scalar output y using a sequence of its quantiles. Given a fraction $\tau \in [0, 1]$, the corresponding quantile y_τ is defined as

$$y_\tau = F^{-1}(\tau) \tag{3.7}$$

with $F(y) = \int_{-\infty}^{y} p(y|\boldsymbol{x})\ dy$ the cumulative distribution function (CDF) of the distribution $p(y|\boldsymbol{x})$. A neural network can learn to predict a quantile y_τ by training it to minimize the quantile loss

$$\mathcal{L}(y_\tau, y) = \begin{cases} \tau|y_\tau - y| & \text{if } y > y_\tau \\ (1-\tau)|y_\tau - y| & \text{otherwise} \end{cases} \tag{3.8}$$

where y_τ is the predicted quantile and y is the reference output value from the training data. It is important to realize here that y is only required to be an ordinary sample of the distribution $p(y|\boldsymbol{x})$ and does not require the distribution $p(y|\boldsymbol{x})$ to be explicitly represented in the training data.

The approach can be extended to the prediction of multiple quantiles corresponding to a sequence of quantile fractions $\mathrm{T} = \{\tau_1, \ldots, \tau_N\}$ by training the network to minimize the sum of the quantile losses

$$\mathcal{L}_\mathrm{T}(\boldsymbol{y}_T, y) = \sum_{\tau \in \mathrm{T}} \mathcal{L}_\tau(y_\tau, y) \tag{3.9}$$

where now $\boldsymbol{y}_T = [y_{\tau_1}, \ldots, y_{\tau_N}]$ is a vector of predicted quantiles.

The basic principle of QRNNs is illustrated Fig. 3.5. To predict the distribution of a scalar variable y conditioned on a vector of inputs $\boldsymbol{x}$, QRNNs use a neural network to transform the vector of inputs into a vector of outputs. The outputs correspond to a sequence of quantiles, which can be used to derive a piece-wise linear approximation of the CDF of $p(y|\boldsymbol{x})$.

Using quantiles to represent the distribution $p(y|\mathbf{x})$ has the advantage that the neural network can learn the shape of the posterior distribution of the retrieval. This is in contrast to other probabilistic regression methods, which often impose a more constrained parametrized form of $p(y|x)$, such as a Gaussian whose mean and standard deviation are predicted.

An unresolved issue with QRNNs is quantile crossing. There is nothing that ensures that the quantiles predicted by a QRNN as they are employed here are in increasing order. QRNNs can thus produce mathematically inconsistent predictions, where quantiles of lower fractions exceed those of higher fractions. However, at least for the applications considered in this thesis, this was not found to be a critical issue.

Density regression neural networks

The second approach for predicting the distribution $p(y|\mathbf{x})$ is based on the work by Oord et al. (2016) and Sønderby et al. (2020). Since it has not been given an explicit name by the authors, it is referred to here as density regression neural networks (DRNNs). In this approach, the regression problem is cast as a classification problem by discretizing the domain of output values y into bins $y_0, \ldots, y_n$ and then predicting for each bin the probability $p_i(y >= y_{i-1}, y < y_i)$ of the observed y falling into the ith bin.

Mathematically this is implemented by minimizing the categorical cross-entropy loss

$$\mathcal{L}(\hat{\mathbf{p}}, y) = -\log(\hat{p}_i) \text{ with i such that } y_{i-1} \leq y < y_i. \tag{3.10}$$

where $\hat{\mathbf{p}} = [\hat{p}_1, \ldots, \hat{p}_n]$ is a vector of predicted probabilities.

The predicted vector of bin probabilities can then be used as a piece-wise constant approximation of the PDF of $p(y|\mathbf{x})$, as illustrated in Fig. 3.6. Since the CDF of $p(y|\mathbf{x})$ can be calculated from the PDF and vise versa, QRNNs and DRNNs are functionally very similar. Compared to QRNNs, DRNNs have the advantage of avoiding quantile crossing by construction. What may be a disadvantage of DRNNs is that they require a relatively high number of outputs when the values of y vary across a wide range.

The aleatoric approximation

Both QRNNs and DRNNs only learn to quantify the aleatoric uncertainty and neglect epistemic uncertainty and covariate shift. It is thus necessary that the aleatoric component of the predictive uncertainty dominates the two other types of uncertainties for these models to produce reliable uncertainty estimates. Given that remote sensing observations are produced in a controlled environment and the range of possible observations is generally well understood, it is usually possible to obtain large amounts of reliable training data, which minimizes epistemic uncertainty and the effects of covariate shift. Given the cost of designing and operating these observation systems, one should expect that an effort is made to minimize both epistemic uncertainty and covariate shift, thus leaving the aleatoric uncertainty as the only irreducible source of uncertainty in the measurements. Additional, empirical evidence for the usefulness of the aleatoric approximation comes from the results presented

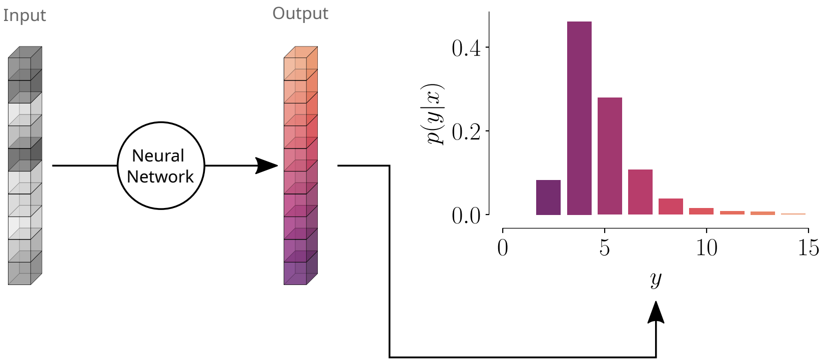


Figure 3.6: The basic functioning of QRNNS. QRNNs predict the distribution of a scalar output variable y using a vector of quantiles from which the an approximation of the CDF of the distribution can be constructed.

in this thesis, where the probabilistic predictions from QRNNs and DRNNs were found to provide reliable estimates of the retrieval uncertainty.

Further limitations

A further limitation of QRNNs and DRNNs is their handling of multiple retrieval outputs. The formulations presented above were based on the assumption of a scalar retrieval output y. These approaches can easily be extended to multiple output variables by simply minimizing the mean of their individual losses. This was found to work well in practice but provides no way of representing the correlations between the multiple outputs in the results. QRNNs and DRNNs are therefore not able to produce random samples that accurately the correlations between the output variables.

3.3.2 Other approaches for quantifying uncertainties

Because of their limitations, QRNNs and DRNNs may not be the best choice for all applications. Therefore, the following section briefly reviews other approaches for quantifying retrieval uncertainty and compares them to the probabilistic regression approach.

Bayesian neural networks

Bayesian neural networks (BNNs) can handle both epistemic and aleatoric uncertainty. For BNNs, not only the target y is assumed to be a random variable, but also the parameters $\boldsymbol{\theta}$ of the neural network model. The random distribution of these parameters represents the uncertainty caused by the limited amount of training data that the model is trained on. Instead of learning specific parameters, a BNN learns a distribution of each of its parameters.

A single prediction from a BNN $p(y|\boldsymbol{x}, \boldsymbol{\theta})$ for a specific input $\boldsymbol{x}$ is obtained by sampling values of the model parameters from the learned distributions and using them to evaluate the model. The epistemic uncertainty in the model predictions is typically quantified by sampling multiple instances of the model parameters $\boldsymbol{\theta}$ and evaluating the model for each of them. For inputs $\boldsymbol{x}$ that the model has encountered often during training, the distribution of $\boldsymbol{\theta}$ values will be sharp so that the distribution $p(y|\boldsymbol{x}, \boldsymbol{\theta})$ does not change much for different realizations of the model parameters $\boldsymbol{\theta}$. For samples where this is not the case, there will be a larger spread and thus higher epistemic uncertainty in the predictions.

Despite their ability to quantify both aleatoric and epistemic uncertainty, BNNs have not yet found widespread adoption in practical applications. One reason for this is likely that their training takes significantly more time. Another disadvantage is that the quantification of epistemic uncertainty requires evaluating the network multiple times. This typically increases the time required to evaluate the model by an order of magnitude, which is a disadvantage for applications that require high throughput.

Despite these potential disadvantages, Orescanin et al. (2022) have recently demonstrated the application of Bayesian neural networks to the classification of precipitation types and shown that the predicted uncertainties are well-calibrated. They also found that the predictive uncertainty is dominated by aleatoric uncertainty, which provides further evidence for the validity of the aleatoric approximation.

Generative models

Generative models are another approach for quantifying uncertainties in neural network predictions. Instead of predicting the conditional probability distribution $p(\boldsymbol{y}|\boldsymbol{x})$ explicitly, these methods are trained to generate samples from the distribution directly. These methods have gained popularity in computer vision tasks due to their ability to generate samples from highly complex probability distributions such as images of faces or generic objects.

The advantage of this approach is that it can handle spatial correlations in the uncertainties, which is not the case for the other methods discussed so far. This means that they can produce spatially coherent samples of the retrieved variables that look very realistic. Recent work has explored the application of generative models for short-term weather forecasting (Ravuri et al., 2021) and probabilistic downscaling of precipitation (Harris et al., 2022; Leinonen et al., 2020). However, these models are also known to be more difficult and take longer to train. In addition to that, they also need to be evaluated multiple times to calculate statistics of the posterior distribution, which increases the computational requirements of the retrieval.

Chapter 4

Contributions

The previous chapter introduced QRNNs and DRNNs as machine-learning-based methods to perform remote sensing retrievals. The motivation for this was to find a practical approach to combine the capabilities of modern deep neural networks with the theoretically sound handling of uncertainties of conventional, inverse-theory-based retrieval methods. To explore the validity and practicality of the proposed methods, they have been applied to idealized scenarios and multiple real-world retrieval applications. This section presents this work and summarizes the main results.

4.1 Handling retrieval uncertainty with neural networks

The first of the appended papers, entitled *A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems* and published as Pfreundschuh et al. (2018), proposes using QRNNs for remote sensing retrievals. The study was motivated by the shortcomings of the retrieval methods that were available at the time of its writing. The Bayesian framework provides a principled way of handling uncertainties in remote sensing retrievals (see Sec. 3.1), but the commonly used methods are computationally complex. While retrievals based on neural networks were already common and often offered superior performance in terms of computational cost as well as accuracy, the retrieval uncertainties were typically neglected.

The aims of the study are two-fold: (1) to demonstrate the compatibility of QRNNs with conventional Bayesian retrieval methods and (2) to demonstrate the practicality of QRNNs by applying them to a real-world retrieval application.

An idealized but realistic retrieval scenario in which the full Bayesian solution could be calculated to arbitrary precision was designed to demonstrate the compatibility of QRNNs with conventional Bayesian retrieval methods. The study applies QRNNs and Bayesian Monte Carlo Integration (BMCI), a commonly used Bayesian retrieval method, to solve the retrieval and compares their solutions to the true solution. BMCI calculates an approximate solution of the retrieval problem using a database of observations and corresponding reference values. This makes it similar in this regard to machine learning methods. The results show that QRNNs work at least as well as BMCI in solving the retrieval problem, given that both are based on a sufficiently large dataset. There is a clearer advantage for QRNNs for smaller datasets, indicating that they cope better with the curse of dimensionality.

This first result has two important implications. Firstly, it showed that QRNNs can be used instead of BMCI and expected to work at least as good given the same training data. Even if the advantage of neural networks over BMCI on the considered dataset was marginal when sufficient training data is available, the use of neural networks offers distinct advantages compared to BMCI. For one, the time required to produce a neural network prediction is independent of the training data size. At least for a naive implementation of BMCI this time scales linearly with the size of the training database. This means that the amount data that can be used by the method may be limited by operational processing requirements. Besides that, because of the flexibility of QRNNs, they can be easily extended to image or time-series data, which is not as simple for BMCI or other Bayesian retrieval methods.

The second important implication of these results is that they show the compatibility of probabilistic machine learning methods with Bayesian retrieval methods. These results link the extensive theory on Bayesian retrievals (Rodgers, 2000; Tarantola, 2005) with machine-learning-based methods. In particular, the correspondence between the training data of the neural network and the a priori distribution of a Bayesian retrieval highlights their importance for accurate quantification of retrieval uncertainties.

The second experiment from this study applied QRNNs to the retrieval of cloud top pressure (CTP) from infrared observations from the Moderate Resolution Imaging Spectroradiometer (MODIS). The comparison of the QRNN retrieval to an existing algorithm based on a deterministic neural network shows that QRNNs yield comparable or better accuracy for point predictions with the added benefit of providing reliable uncertainty estimates.

A principal result from this experiment is shown in Fig. 4.1, which shows the predicted and observed distributions of retrieval errors for QRNNs as well as for the deterministic retrieval combined with a Gaussian error model fitted on the test data. As the comparison of the distribution shows, the Gaussian error model does not provide a good description of the retrieval error, whereas the QRNN successfully predicts its distribution. This results shows the ability of the QRNN to represent non-Gaussian retrieval errors and the superiority over commonly assumed Gaussian error models. A consequence of these findings was the adoption of QRNNs for the operational production of near-real-time retrievals of cloud top pressure at the European Organisation for the Exploitation of Meteorological Satellites.

4.2 Passive microwave precipitation retrievals

The second study included in this thesis, entitled *GPROF-NN: A neural network based implementation of the Goddard Profiling Algorithm* and published as Pfreundschuh et al. (2022a), applied the QRNN methodology to the retrieval of precipitation

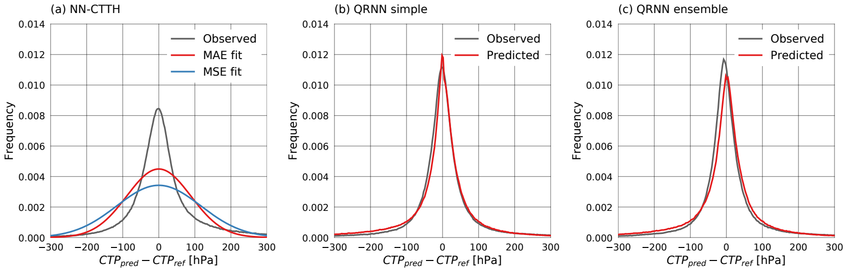


Figure 4.1: Distributions of predicted and observed cloud-top pressure (CTP) retrieval errors. Panel (a) shows the results for a deterministic neural network combined with a Gaussian error models fitted used the mean squared error (blue) and the mean absolute error. Panel (b) show the corresponding results for the QRNN retrieval.

from passive microwave (PMW) observations of the GPM mission (see Sec.2.3.3). GPM is an international satellite mission lead by NASA and JAXA, which aims to provide global measurements of precipitation at high spatial and temporal resolution.

The algorithm that is used to retrieve precipitation from GMI, the PMW sensor on the GPM core observatory, and the other PMW sensors of the GPM constellation is the Goddard Profiling Algorithm (GPROF, Kummerow et al., 2015). GPROF is based on the BMCI method and retrieves surface precipitation as well as profiles of hydrometeor concentrations and latent heating rates. The aim of the study was to assess the potential benefits of replacing BMCI with a neural-network-based retrieval. Besides that, it aims to explore to what extent the accuracy can be improved if spatial information is incorporated into the retrieval instead of processing each pixel separately as the current implementation does. To this end, two neural-network-based implementations were developed. The first one, named GPROF-NN 1D, uses an MLP to retrieve precipitation from a single pixel. The second one, GPROF-NN 3D, uses a CNN to retrieve precipitation using the full swath of observations. This allows the GPROF-NN 3D retrieval to leverage structural information in the observations, which is not available to the two other implementations.

Both the GPROF-NN 1D and GPROF-NN 3D retrieval were developed to be functionally equivalent to the currently operational GPROF algorithm so that they can potentially replace it in an upcoming update. Moreover, the implementations were restricted to use exactly the same data for their training as the current method, which was done to ensure that the comparison of the three methods isn't obscured by differences in the training data.

While the application of QRNNs to the GPROF algorithm is in principle straightforward, the requirement to develop a retrieval that uses and produces identical in- and output data and is based on the same data as the current implementation of GPROF proved challenging. The training data for the sensors of the GPM constellation makes use of radiative transfer simulations. However, these are not generated for full swaths of observations but only for single pixels. This makes the simulation data unsuitable for the training of CNNs. Since extending the simulations to the full

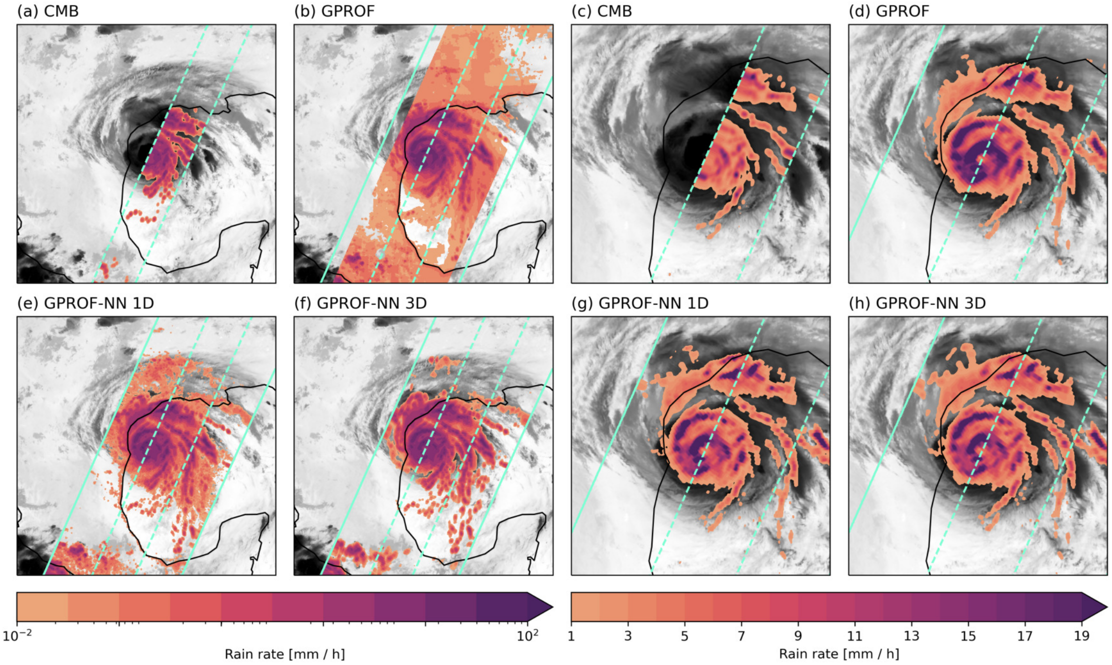


Figure 4.2: Retrieved surface precipitation from the GPM combined product (Panels (a), (c)), the GPROF algorithm (Panels (b), (d)), and the neural network versions GPROF-NN 1D (Panels (e), (g)) and GPROF-NN 3D (Panels (f), (h)) in and around Hurricane Harvey on 2017-08-25 at 11:50:00 UTC. Panels (a), (b), (e), (f) display the precipitation on a logarithmic scale over an area spanning 2 000 km in zonal and meridional directions. Panels (c), (d), (g), (f) display the surface precipitation on a linear color scale in an enlarged area spanning 500 km around the Hurricane. The background shows 10.3 μm brightness temperatures from the GOES 16 observations closest to the overpass.

swath is currently not possible, an intermediate CNN-based 'retrieval' was trained to emulate the simulations of PMW observations.

The main results from this study are estimates of potential improvements in retrieval accuracy that are achievable by upgrading GPROF to either the GPROF-NN 1D or GPROF-NN 3D retrieval. Consistent improvements for the GPROF-NN 1D algorithms are found across a range of accuracy metrics for the retrieved surface precipitation as well as the hydrometeor profiles. The accuracy is improved further by the GPROF-NN 3D retrieval, which yields improvements similar in magnitude to those provided by the GPROF-NN 1D algorithm over the GPROF baseline retrieval. Furthremore, we found that the effective resolution of the retrieval improves by at least 40 % for the neural-network-based retrievals.

An example of the improvements afforded by the neural-network-based retrievals is shown in Fig. 4.2. This plot shows the smoothed surface precipitation from the combined retrieval from the GPM core observatory as a reference and the retrievals from the GPROF and GPROF-NN algorithms. The comparison shows that the GPROF-NN retrievals are spatially better resolved and improve the agreement with the combined retrievals.

Although the results from this study were promising, their significance was limited because the evaluation of the retrieval was restricted to the same source of data that was used for the generation of the training data. Since the objective of the study was to compare the BMCI retrieval method with QRNNs, this simplification was justified. In a validation against independent data, potential benefits may be obscured by deviations between training data and the data used for validation. However, the practically more relevant question is, of course, to what extent the neural network retrievals improve the precipitation estimates when compared to independent validation data.

A study that addresses this question is currently under preparation and included in this thesis as the third paper, entitled *Evolution of the GPROF passive microwave precipitation retrievals evaluated against ground radar measurements over the continental US and the Pacific Ocean.* Since the development of the GPROF-NN algorithms was performed in parallel with a new version of the GPROF retrieval, this study aimed to assess improvements in this new version of GPROF (GPROF V7) against the previous version (GPROF V5) , as well as the potential benefits of replacing GPROF with GPROF-NN 1D or 3D in a future update. Furthermore, the study aims to identify potential outstanding issues impeding the adaptation of the GPROF-NN retrievals for operational processing.

The validation is based on ground-radar measurements of precipitation specifically created for the validation of GPM measurements. It is based on measurements from the continental United States (CONUS) and a ground radar station on the Kwajalein atoll in the Pacific Ocean.

The main focus of this study is put on retrievals from the GMI radiometer, which plays a special role in the GPM constellation due to it being designed specifically for the remote sensing of precipitation. The accuracy of the precipitation retrieved from the GMI sensor using the two versions of the conventional GPROF algorithm as well as the two GPROF-NN retrievals is evaluated for two years of co-locations. The results clearly show that the benefits of the neural-network-based implementation of GPROF carry over to validation against independent precipitation measurements. In particular, the effective resolution of the retrieved precipitation over land is improved by more than a factor of two.

The principal results from this study are shown in Fig. 4.3. The graphs show error metrics of the retrieval when compared against the training data (labeled Database) as well as compared against ground-based radar measurements over the period of two years. These results clearly show that the GPROF-NN retrievals consistently improve the instantaneous errors in the retrieval. For the biases, the results are less obvious, but the fact that the combined retrieval (labeled GPM-CMB) exhibits similar biases suggests that the origin for these biases is at least in part in the training data itself.

In addition to GMI, also the retrieval accuracy of a limited number of other sensors of the GPM constellation is evaluated. For the two sensors that have been evaluated so far, smaller improvements in retrieval accuracy are found. This is nonetheless an encouraging result because these sensors is affected by errors in the radiative transfer simulation required to generate the training data, which is not the case for

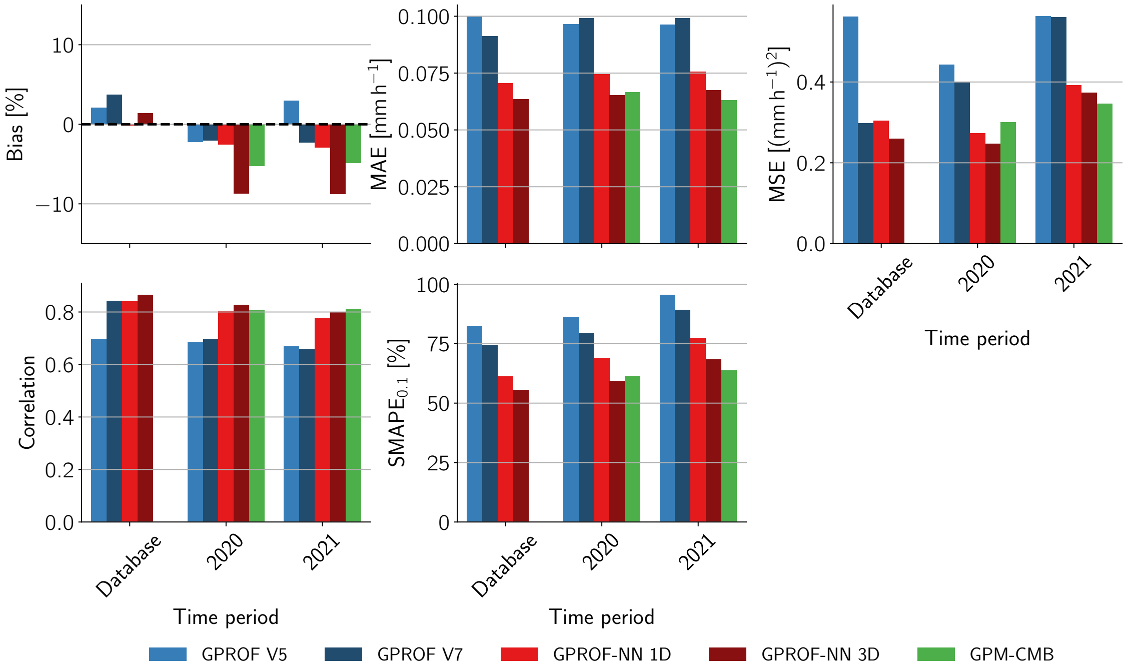


Figure 4.3: Accuracy metrics of the GPROF retrievals compared to the training data (Database) and gauge-corrected ground-radar measurements for the water years 2020 and 2021.

the retrieval for GMI.

4.3 Near real-time rain retrievals over Brazil

While the neural-network-based implementation of GPROF provided evidence for the potential of neural-network-based precipitation retrievals, the constraint of a retrieval that provides the same output as the current operational GPROF algorithm did not leave much room to explore retrieval improvements afforded by the probabilistic retrieval results. The two studies on the development and assessment of the GPROF-NN retrievals presented above therefore focused mostly on deterministic precipitation estimates and did not explore the full potential of the probabilistic predictions afforded by the neural networks.

The aim of the third study presented in this thesis, titled *An improved near real-time precipitation retrieval for Brazil* and published as Pfreundschuh et al. (2022d), was to explore the full potential of probabilistic neural-network based precipitation retrievals in the context of near real-time retrievals from VIS and IR observations over Brazil. The input data for the retrieval comes from the advanced baseline imager (ABI, Schmit et al., 2018) on the geostationary operational environmental satellite (GOES) 16. The input observations were co-located with retrievals from the combined radar and passive microwave observations from the GPM Core Observatory to generate the training data.

To validate the neural-network-based retrievals, their results were compared to one month of gauge measurements. The retrieval accuracy was assessed by comparing it to the retrieval algorithm that is currently used operationally at the Brazilian National Institute of Space Research as well as two other commonly used, global precipitation retrievals. This comparison shows that, despite the limited information content of VIS/IR observations, deep-learning-based retrievals outperform currently available methods, even those that merge IR observations from geostationary satellites with passive microwave observations.

Assessment of the predicted uncertainties against the gauge measurements shows that they are reliable, given that differences between training and reference data are taken into account. In addition to that, the study presented ways of leveraging probabilistic retrieval results to improve the characterization of precipitation. The predicted uncertainties can be used, for example, to derive the probability of the observed rain exceeding a given threshold. When the probabilistic prediction are used in this way, the retrieval outperforms all reference retrievals for the detection of heavy precipitation events. Moreover, when the predicted distributions are used to generate samples from the a posteriori distribution of the retrieval, the agreement between the distributions of retrieved and gauge-measured precipitation is improved. Including these samples in the retrieval output thus provides a way of improving the representation of precipitation extremes in the retrieval results.

Fig. 4.4 shows the results of the reference retrievals as well as the Hydronn neural-network-based precipitation retrievals that were developed in this study. The results correspond to a flood-producing precipitation event that occurred during the evaluation period. Each panel shows the results for one or more retrievals together with the reference gauge measurement. Panel (a) shows the results for the assessed reference retrievals. Only one of them predicts significant precipitation during the event. Panels (b), (c), and (d) show the corresponding results for the three configurations of the Hydronn retrieval with the predicted mean precipitation shown by the solid line and the shading showing the posterior CDF at quantile fractions $[0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99]$. Although the predicted mean precipitation is as low as that of the best reference retrieval, the spread in the uncertainties clearly shows the risk of much higher precipitation.

Besides a significantly improved near real-time retrieval for application over Brazil, the principal results from this study are further proof of the ability of neural-network-based retrievals to leverage the full potential of latest-generation satellite imagery. The configuration that uses all available channels at the highest possible resolution outperforms the GPM IMERG product, which is one of the most advanced precipitation retrievals currently available and integrates observations from a wide range of sensors and gauge measurements.

However, as can be seen in the results in Fig. 4.4, the retrievals from VIS and IR observations still remain limited in their capabilities of detecting extreme precipitation. While deep neural networks yield significant improvements in retrieval performance, their instantaneous accuracy will always be limited by the low information content of the observations. This suggest that future retrieval development should try to further increase the amount of information available to the retrieval, for example by incorporating the temporal evolution of the observations or by merging observations from multiple sensors types.

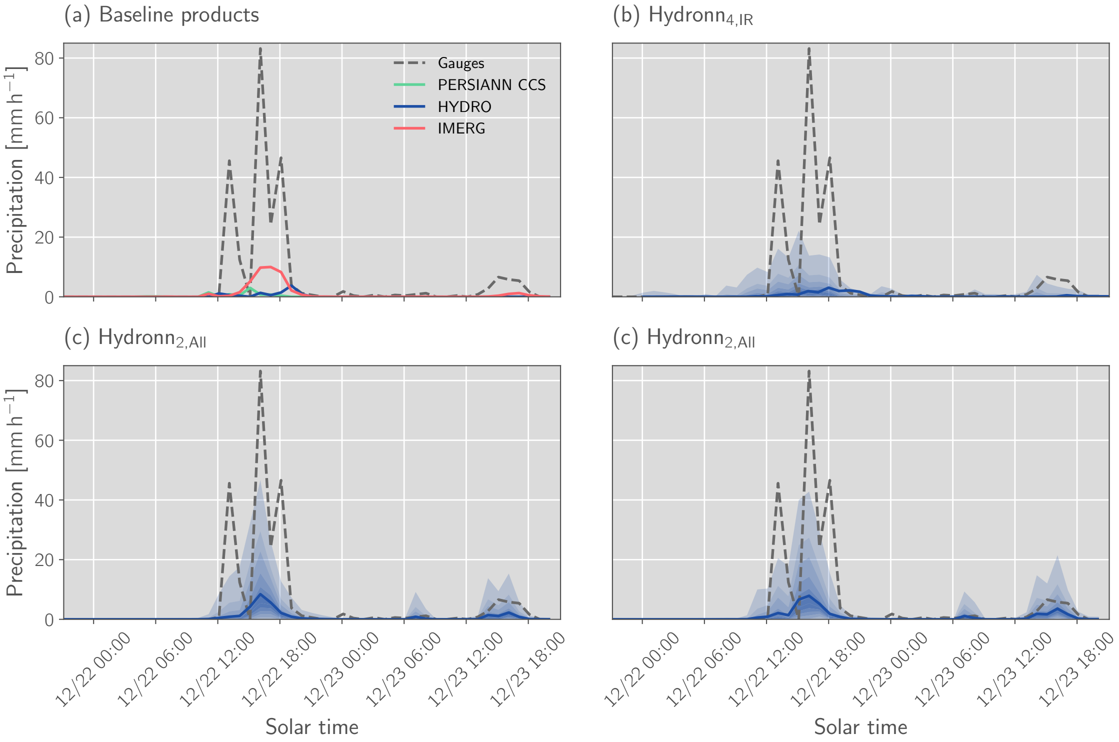


Figure 4.4: Retrieved precipitation for a flood-producing precipitation even that occurred between 2020-12-22 and 2020-12-24 in the cite of Duque de Caxias in the state of Rio de Janeiro. Grey, dashed lines show the precipitation by the gauge station in Xerém. Solid lines show the retrieved mean precipitation for each retrieval algorithm. The shading shows filled contours of the posterior CDF at values $[0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99]$.

4.4 Cloud correction for data assimilation

The final study included in this thesis is entitled *Can machine learning correct microwave humidity radiances for the influence of clouds?* and published as Kaur et al. (2021) explores the application of QRNNs to remove the effect of clouds from microwave observations. These observations contain information on the temperature and distribution of water vapor in the atmosphere and are used in data assimilation systems to find good initial conditions for numerical weather forecasts. Since the assimilation usually involves simulating the observations, the presence of hydrometeors in the atmosphere makes the handling of observations that are affected by clouds much more complex. The assimilation of both cloud-free and cloudy observations is referred to as *all-sky* assimilation. However, due to the complexity of handling the effects of hydrometeors, less advanced assimilation systems discard cloudy observations and only assimilate clear-sky observations.

This study investigates whether QRNNs can be used to detect and remove the effect of clouds on microwave observations. If a model existed that could accurately predict how cloudy-sky observations would look when the effect of hydrometeors is removed, then this could, in principle, be used to handle these observations in clear-sky assimilation systems. Since the ability to handle cloud-affected observations significantly increases the number of observations that can be ingested by the system, such a model could thus help to improve the initialization of weather forecasts.

Because of their impact on the radiative transfer, operational data assimilation systems have to detect observations affected by hydrometeors to either inflate the uncertainties assigned to the observations (for all-sky assimilation) or discard them altogether. Therefore, the first experiment in this study assesses the potential of applying QRNNs to predict a correction for cloud-contaminated observations using an existing satellite sensor and compares the performance to existing methods. The experiment shows that the QRNN-derived correction leads to a lower bias in the observations than the residual biases in currently used filtering methods, which reject up to 30 % of the observations.

The second experiment extended the approach to an upcoming satellite that will be launched soon and a concept for a future satellite. Both of these satellites will carry microwave sensors with channels above 300 GHz. These high frequencies make the observations more sensitive to scattering from ice particles, which further complicates their use in data assimilation. In these cases, using a QRNN-based cloud correction would provide a way to use the additional information from the high-frequency channels to provide a more accurate correction of lower frequency observations, which would improve the impact of the lower frequency observations in the assimilation system. The results show that observations at frequencies exceeding 300 GHz are well suited for cloud correction and thus demonstrate the potential of this approach.

The principal result from this study is a proposed novel application of QRNNs for the correction of microwave observations for use in data assimilation. QRNN-based cloud correction provides superior performance to existing methods, and the predicted uncertainties are more accurate than error estimates from currently used error models. Moreover, the simulation-based assessment of the potential of microwave observations at sub-millimeter wavelengths led to the selection of these channels for an upcoming satellite mission (ESA, 2021).

4.5 Future work

The findings of this thesis suggest several directions of future research. These will be discussed below.

4.5.1 Handling retrieval uncertainty with neural networks

This thesis has proposed and assessed neural-network-based methods for handling uncertainties in remote sensing retrievals and shown their practicality across multiple retrieval applications. The principal advantage of these methods is certainly their simplicity. A minor modification in the training process is required to migrate a deterministic neural network retrieval to a probabilistic one. There are, however, two important limitations of the approaches: Their incapability to handle correlations in the retrieval outputs and their reliance on the aleatoric assumption, which postulates that the predictive uncertainty is dominated by the aleatoric uncertainty.

It would therefore be valuable to systematically evaluate the approaches against Bayesian neural networks and generative models on a set of atmospheric retrieval problems. Such an evaluation should assess not only the retrieval accuracy but also to take into account the effect on downstream applications of the data. This would provide guidance in what scenarios the computationally more complex alternative approaches should be applied.

4.5.2 GPM precipitation retrievals

Due to the good performance of the developed GPROF-NN retrievals, they are being considered for the operational processing of the passive microwave retrievals of the GPM mission. This will require additional investigations regarding the climatological stability of the GPROF-NN retrievals across different satellites.

In addition to that, the migration to a neural-network-based implementation opens up a number of opportunities for further improvements of the GPM passive microwave retrievals. One of them is the extension of the training data to cover multiple years of observations. It was found in (Pfreundschuh et al., 2022c) that the limited training period may be one reason for the retrieval errors on climatological time scales. Besides that, including samples from the posterior distribution may improve the representation of extreme precipitation in the retrieval results, which is highly relevant for climate change studies.

Novel retrieval applications

As its name suggests, GPROF, the Goddard Profiling Algorithm, also retrieves profiles of hydrometeors. These profiles consist comprise 28 vertical levels and are retrieved at each pixel of the observations. However, because of storage limitations, the profiles are compressed in the retrieval results, which causes their accuracy to degrade.

One operational aspect of the GPROF-NN 3D retrieval that was discovered during its development is that it does not require the ancillary data used by GPROF to yield accurate. The ancillary provides information on the water content and temperature of the atmosphere as well as the surface type to the retrieval. Since these inputs need to be prepared, running the conventional GPROF retrieval requires access to very specific datasets in addition to the satellite observations.

Since the GPROF-NN 3D retrieval can be trained to work without the ancillary data, it becomes much simpler to run the retrieval. It only requires access the satellite observations, which can be readily downloaded from the internet, and the neural network model, which has a non-compressed size of about 250 MB and can thus also easily be distributed over the internet. This opens up the possibility of providing users with uncompressed retrieval results by giving them the opportunity to run the retrieval themselves.

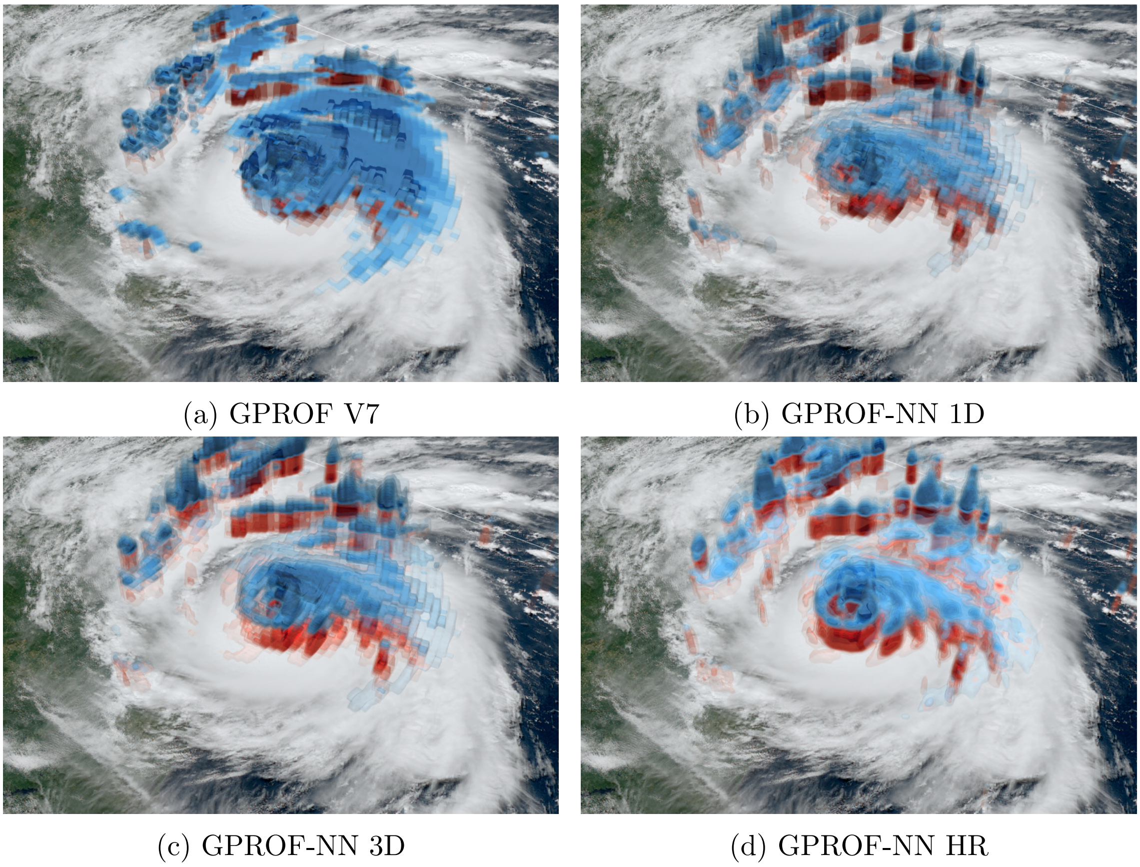
Since this lifts size constraints on the retrieval output, it provides a practical way of developing and distributing a super-resolution retrieval for the observations of the GMI microwave sensor. The spatial sampling of GMI is 13.5 km in the along-track direction and 5 km in the across-track direction. The training data, which is derived from the combined retrievals of the GPM core observatory, has a spatial sampling of 5 km × 5 km. Since the conventional GPROF retrieval operates on single pixels, the training data is sub-sampled and averaged to the footprint of the GMI pixels causing the high resolution of the reference data to be lost. This, however, is not necessary for the CNN based GPROF-NN retrieval. To explore the potential of performing super-resolution retrievals, an experimental retrieval, named GPROF-NN HR, has been developed, which retrieves precipitation at a three-times higher resolution in the along-track direction of the GMI swath.

Very early results of such a retrieval are shown in Fig. 4.5, which shows isosurfaces of retrieved rain and snow water content as retrieved by GPROF and the three GPROF-NN retrievals. The results for GPROF include the effects of the profile compression that is applied to store them. The GRPOF-NN 1D and 3D retrievals yield clearly better results, but the effects of the coarse along-track resolution are still visible. The results of the GPROF-NN HR retrieval demonstrate its ability to better resolve the rain bands in the hurricane. This suggests that the effective resolution of the retrievals from GMI can be improved even further than what is already achieved by the GPROF-NN retrievals.

4.5.3 Near real-time precipitation retrievals

The results presented in Pfreundschuh et al. (2022d) clearly show the potential of deep neural networks for precipitation retrievals from geostationary satellites. According to personal communication, the developed retrieval is also being considered for operational application.

An interesting result that emerged from this study was that, despite the low information content of the geostationary observations on precipitation, deep-learning-based



(a) GPROF V7 (b) GPROF-NN 1D

(c) GPROF-NN 3D (d) GPROF-NN HR

Figure 4.5: Iso surfaces of rain water content (red) and snow water content (blue) in hurricane Ida shortly before landfall on 2021-08-29 15:14 UTC. For visualization purposes the vertical dimension is exaggerated by a factor of 10.

retrievals can outperform highly complex retrieval pipelines which integrate observations from multiple sensors. This indicates that the use of information in current multi-sensor retrievals is sub-optimal. An upcoming project will therefore explore the potential of directly fusing observations from different sensors using a single neural network.

Bibliography

B. W. Abbott, K. Bishop, J. P. Zarnetske, C. Minaudo, F. Chapin, S. Krause, D. M. Hannah, L. Conner, D. Ellison, S. E. Godsey, et al. Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, 12(7):533–540, 2019.

J. Ahrens, B. Geveci, and C. Law. Paraview: An end-user tool for large data visualization. *The visualization handbook*, 717(8), 2005.

P. Bauer, A. J. Geer, P. Lopez, and D. Salmond. Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quarterly Journal of the Royal Meteorological Society*, 136(652):1868–1885, 2010. doi: 10.1002/qj.659.

P. Bauer, A. Thorpe, and G. Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

S. G. Benjamin, E. P. James, M. Hu, C. R. Alexander, T. T. Ladwig, J. M. Brown, S. S. Weygandt, D. D. Turner, P. Minnis, W. L. Smith, and A. K. Heidinger. Stratiform Cloud-Hydrometeor Assimilation for HRRR and RAP Model Short-Range Weather Prediction. *Monthly Weather Review*, 149(8):2673 – 2694, 2021. doi: 10.1175/MWR-D-20-0319.1.

M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, et al. Long-term climate change: projections, commitments and irreversibility. In *Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1029–1136. Cambridge University Press, 2013.

S. de Haan and S. H. van der Veen. Cloud Initialization in the Rapid Update Cycle of HIRLAM. *Weather and Forecasting*, 29(5):1120 – 1133, 2014. doi: 10.1175/WAF-D-13-00071.1.

R. A. de Siqueira and D. Vila. Hybrid methodology for precipitation estimation using Hydro-Estimator over Brazil. *International Journal of Remote Sensing*, 40 (11):4244–4263, 2019. doi: 10.1080/01431161.2018.1562262.

P. D'Odorico, K. F. Davis, L. Rosa, J. A. Carr, D. Chiarelli, J. Dell'Angelo, J. Gephart, G. K. MacDonald, D. A. Seekell, S. Suweis, and M. C. Rulli. The Global Food-Energy-Water Nexus. *Reviews of Geophysics*, 56(3):456–531, 2018. doi: 10.1029/2017RG000591.

D. I. Duncan, P. Eriksson, and S. Pfreundschuh. An experimental 2D-Var retrieval using AMSR2. *Atmospheric Measurement Techniques*, 12(12):6341–6359, 2019a. doi: 10.5194/amt-12-6341-2019.

D. I. Duncan, P. Eriksson, S. Pfreundschuh, C. Klepp, and D. C. Jones. On the distinctiveness of observed oceanic raindrop distributions. *Atmospheric Chemistry and Physics*, 19(10):6969–6984, 2019b. doi: 10.5194/acp-19-6969-2019.

R. Ekelund, P. Eriksson, and S. Pfreundschuh. Using passive and active observations at microwave and sub-millimetre wavelengths to constrain ice particle models. *Atmospheric Measurement Techniques*, 13(2):501–520, 2020. doi: 10.5194/amt-13-501-2020.

ESA. ESA - Arctic Weather Satellite. `https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/Arctic_Weather_Satellite`, 2021.

M. Falkenmark, J. Rockstrom, and J. Rockström. *Balancing water for humans and nature: the new approach in ecohydrology*. Earthscan, 2004.

H. H. Frisinger. Aristotle and his "Meteorologica". *Bulletin of the American Meteorological Society*, 53(7):634–638, 1972.

J. Hagen, K. Hocke, G. Stober, S. Pfreundschuh, A. Murk, and N. Kämpfer. First measurements of tides in the stratosphere and lower mesosphere by ground-based Doppler microwave wind radiometry. *Atmospheric Chemistry and Physics*, 20(4):2367–2386, 2020. doi: 10.5194/acp-20-2367-2020.

L. Harris, A. T. McRae, M. Chantry, P. D. Dueben, and T. N. Palmer. A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts. *arXiv preprint arXiv:2204.02028*, 2022.

A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bull. Amer. Met. Soc.*, 95(5):701–722, 2014. doi: 10.1175/BAMS-D-13-00164.1.

S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *In revision, J. Open Res. Software*, 2017.

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

N. A. Hughes. Global Cloud Climatologies: A Historical Review. *Journal of Applied Meteorology and Climatology*, 23(5):724 – 751, 1984. doi: 10.1175/1520-0450(1984)023⟨0724:GCCAHR⟩2.0.CO;2.

J. D. Hunter. Matplotlib: A 2D graphics environment. *Comput. Sce. Eng.*, 9(3): 90–95, 2007. doi: 10.1109/MCSE.2007.55.

I. Kaur, P. Eriksson, S. Pfreundschuh, and D. I. Duncan. Can machine learning correct microwave humidity radiances for the influence of clouds? *Atmospheric Measurement Techniques*, 14(4):2957–2979, 2021. doi: 10.5194/amt-14-2957-2021.

I. Kaur, P. Eriksson, V. Barlakas, S. Pfreundschuh, and S. Fox. Fast Radiative Transfer Approximating Ice Hydrometeor Orientation and Its Implication on IWP Retrievals. *Remote Sensing*, 14(7), 2022. ISSN 2072-4292. doi: 10.3390/rs14071594.

C. Kidd, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum. So, How Much of the Earth's Surface Is Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, 98(1):69 – 78, 2017. doi: 10.1175/BAMS-D-14-00283.1.

T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and J. development team. Jupyter Notebooks - a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90, Netherlands, 2016. IOS Press. URL `https://eprints.soton.ac.uk/403913/`.

C. D. Kummerow, D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic. The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *J. Atmos. Oceanic Technol.*, 32(12): 2265–2280, 2015. doi: 10.1175/JTECH-D-15-0039.1.

J. Leinonen, D. Nerini, and A. Berne. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7211–7223, 2020.

M. I. Mishchenko, L. D. Travis, and A. A. Lacis. *Scattering, absorption, and emission of light by small particles.* Cambridge university press, 2002.

A. V. Oord, N. Kalchbrenner, and K. Kavukcuoglu. Pixel Recurrent Neural Networks. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1747–1756, New York, New York, USA, 20–22 Jun 2016. PMLR.

M. Orescanin, V. Petković, S. W. Powell, B. R. Marsh, and S. C. Heslin. Bayesian Deep Learning for Passive Microwave Precipitation Type Detection. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2021.3090743.

T. pandas development team. pandas-dev/pandas: Pandas, February 2020.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

F. Pérez and B. E. Granger. IPython: a System for Interactive Scientific Computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.53. URL `https://ipython.org`.

S. Pfreundschuh, P. Eriksson, D. Duncan, B. Rydberg, N. Håkansson, and A. Thoss. A neural network approach to estimating a posteriori distributions of Bayesian retrieval problems. *Atmos. Meas. Tech.*, 11(8):4627–4643, 2018. doi: 10.5194/amt-11-4627-2018.

S. Pfreundschuh, P. Eriksson, S. A. Buehler, M. Brath, D. Duncan, R. Larsson, and R. Ekelund. Synergistic radar and radiometer retrievals of ice hydrometeors. *Atmospheric Measurement Techniques*, 13(8):4219–4245, 2020. doi: 10.5194/amt-13-4219-2020.

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

S. Pfreundschuh, S. Fox, P. Eriksson, D. Duncan, S. A. Buehler, M. Brath, R. Cotton, and F. Ewald. Synergistic radar and sub-millimeter radiometer retrievals of ice hydrometeors in mid-latitude frontal cloud systems. *Atmospheric Measurement Techniques*, 15(3):677–699, 2022b. doi: 10.5194/amt-15-677-2022.

S. Pfreundschuh, C. Guilloteau, P. J. Brown, C. D. Kummerow, and P. Eriksson. Evolution of the GPROF passive microwave precipitation retrievals evaluated against ground radar measurements over the continental US and the Pacific Ocean. *In preparation*, 2022c.

S. Pfreundschuh, I. Ingemarsson, P. Eriksson, D. A. Vila, and A. J. P. Calheiros. An improved near real-time precipitation retrieval for Brazil. *EGUsphere*, 2022: 1–30, 2022d. doi: 10.5194/egusphere-2022-78.

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

S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021.

C. D. Rodgers. *Inverse methods for atmospheric sounding: theory and practice*, volume 2. World scientific, 2000.

L. Rosa, M. C. Rulli, K. F. Davis, and P. D'Odorico. The Water-Energy Nexus of Hydraulic Fracturing: A Global Hydrologic Analysis for Shale Oil and Gas Extraction. *Earth's Future*, 6(5):745–756, 2018. doi: 10.1002/2018EF000809.

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

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

G. L. Stephens. The useful pursuit of shadows: the study of clouds has profoundly influenced science and human culture and stands poised to lead climate science forward again. *American Scientist*, 91(5):442–449, 2003.

B. Stevens. Waterin the atmosphere. *Phys. Today*, 66(6):29, 2013.

I. Strangeways. *Measuring the Natural Environment.* Cambridge University Press, 2000. ISBN 9780511151279.

C. B. Sullivan and A. Kaszynski. PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *Journal of Open Source Software*, 4(37):1450, may 2019. doi: 10.21105/joss.01450.

S. Tanelli, S. L. Durden, E. Im, K. S. Pak, D. G. Reinke, P. Partain, J. M. Haynes, and R. T. Marchand. CloudSat's Cloud Profiling Radar After Two Years in Orbit: Performance, Calibration, and Processing. *IEEE T. Geosci. Remote*, 46 (11):3560–3573, Nov 2008. ISSN 0196-2892. doi: 10.1109/TGRS.2008.2002030.

A. Tarantola. *Inverse problem theory and methods for model parameter estimation.* SIAM, 2005.

G. E. Thomas and K. Stamnes. *Radiative transfer in the atmosphere and ocean.* Cambridge University Press, 2002.

K. E. Trenberth, J. T. Fasullo, and J. Kiehl. Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, 90(3):311 – 324, 2009. doi: 10.1175/2008BAMS2634.1.

United Nations. Sustainable development goals. `https://sdgs.un.org`, 2018. Accessed: 2022-04-20.

S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2):22–30, March 2011. doi: 10.1109/MCSE.2011.37.

G. A. Vicente, R. A. Scofield, and W. P. Menzel. The Operational GOES Infrared Rainfall Estimation Technique. *Bulletin of the American Meteorological Society*, 79(9):1883 – 1898, 1998. doi: 10.1175/1520-0477(1998)079⟨1883:TOGIRE⟩2.0.CO;2.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

J. M. Wallace and P. V. Hobbs. *Atmospheric science: an introductory survey*, volume 92. Elsevier, 2006.

M. L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021.

Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.

C. Zarfl, A. E. Lumsdon, J. Berlekamp, L. Tydecks, and K. Tockner. A global boom in hydropower dam construction. *Aquatic Sciences*, 77(1):161–170, 2015.

M. D. Zelinka, T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor. Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1):e2019GL085782, 2020. doi: 10.1029/2019GL085782. e2019GL085782 10.1029/2019GL085782.