Thesis for The Degree of Licentiate of Engineering

Retrieving cloud ice masses from geostationary images with neural networks

Adrià Amell

Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden, 2023

Retrieving cloud ice masses from geostationary images with neural networks

Adrià Amell

Department of Space, Earth and Environment
Division of Geoscience and Remote Sensing
Chalmers University of Technology
SE-412 96 Gothenburg,
Sweden
Phone: +46(0)31 772 1000

Printed by Chalmers Digitaltryck,
Gothenburg, Sweden, 2023

Retrieving cloud ice masses from geostationary images with neural networks

Adrià Amell

Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Clouds are essential to the Earth's energy budget and atmospheric circulation. Despite this, many cloud parameters are poorly known, including the mass of frozen hydrometeors. On the one hand, there will be specialized satellite missions targeting such hydrometeors. On the other hand, existing satellite data can be leveraged. There should be a particular interest in using geostationary satellite observations since they provide continuous coverage. Traditionally, retrievals of cloud ice masses from geostationary measurements require solar reflectances, ignore any spatial correlations, and solely retrieve the vertically-integrated ice mass density, known as the ice water path.

This thesis challenges the traditional approach by applying supervised learning against CloudSat collocations, the only existing satellite mission targeting ice clouds. A set of neural networks is assembled to compare the performance of using different visible or infrared channels as retrieval input as well as the added value of using spatial context. The retrievals are probabilistic, in the sense that all neural networks predict quantiles to estimate the retrieval irreducible uncertainty, and thus represent the state of the art for atmospheric retrievals.

With several spectral channels, infrared retrievals are found to have a similar performance compared to the peak accuracy offered by the combination of visible and infrared channels. However, the infrared-only retrievals enable a consistent diurnal performance. The use of spatial information reinforces the retrievals, which is demonstrated by the ability to provide skilful three-dimensional estimates of ice masses, known as ice water content, from only one infrared channel. The latter retrieval scheme is supported by an extensive validation with independent measurements.

These neural network-based retrievals offer the possibility to derive new insights into cloud physics, reduce present ice cloud uncertainties, and validate climate models. Ideally, such retrieval schemes will complement the sparse measurements from specialized instruments. Finally, this thesis contains the groundwork for executing retrievals on multidecadal geostationary observations, offering unprecedented spatially and temporally continuous three-dimensional data for the tropics and mid-latitudes. The implementation of these ongoing retrievals is publicly released as part of the Chalmers Cloud Ice Climatology.

Keywords
Machine learning, geostationary satellites, ice clouds

Acknowledgments

I am profoundly grateful to my supervisor Patrick Eriksson for his exceptional support throughout my doctoral journey. His availability and willingness to provide detailed and constructive feedback have not only enhanced my research but also inspired me. The freedom he grants me as well as his finesse are invaluable to me on a personal level.

I extend my heartfelt thanks to Simon Pfreundschuh for the continuous discussions that enriched my research. His willingness to have conversations at his 7 a.m. or even earlier has been pivotal in facilitating a transatlantic collaboration and stands as a testament to his commitment to exchange knowledge.

My gratitude also extends to my fellow colleagues and doctoral students, both within and beyond my university. Enlightening interactions with them have refined my questions and helped me recognize the limits of my knowledge.

I would like to express my gratitude to my examiner Lars Ulander, co-supervisors, and the administration staff for their instrumental role. I also wish to express my appreciation to the division of geoscience and remote sensing for funding my position.

Adrià Amell
Gothenburg, Sweden
September 2023

List of Works

Appended papers

This thesis is based on the following papers:

[**Paper I**] **A. Amell**, P. Eriksson, and S. Pfreundschuh, *Ice water path retrievals from Meteosat-9 using quantile regression neural networks Atmos. Meas. Tech., 15, 5701–5717, 2022,* `https://doi.org/10.5194/amt-15-5701-2022`.

[**Paper II**] **A. Amell**, and S. Pfreundschuh, and P. Eriksson, *The Chalmers Cloud Ice Climatology: retrieval implementation and validation Submitted to Atmos. Meas. Tech., 2023.*

Other works

The following work has been done during my PhD studies, but a manuscript has not been prepared. Furthermore, the objectives of the work are not directly related to the content of the thesis.

A. Amell, P. Eriksson, L. Hee, and S. Pfreundschuh, *Nearly instantaneous probabilistic retrievals of Rain over Africa EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-9544, 2023,* `https://doi.org/10.5194/egusphere-egu23-9544`.

Contents

List of Acronyms

CCIC	Chalmers Cloud Ice Climatology
CNN	convolutional neural network
GPU	graphics processing unit
IWC	ice water content
IWP	ice water path
MLP	multilayer perceptron
NN	artificial neural network
QPD	quantile-parameterized distribution
QRNN	quantile regression neural network
T_B	brightness temperature

Part I

Summary

1 Introduction

Presently, one of the key elements that contributes to progress in remote sensing and geoscience are advances in machine learning. The wealth of observational datasets and developments in computational systems have catalysed the exploration and implementation of data-driven algorithms. It suffices to attend any conference, engage in scientific discussions or follow remote sensing journals to notice that machine learning serves not only as a powerful tool for statistical data processing but increasingly also for learning, understanding, and extracting knowledge. This paradigm shift in modern remote sensing is also the basis for this licentiate thesis.

Acquiring atmospheric information is one of the applications of remote sensing. Clouds are an integral and dynamic part of the atmosphere. These perceivable masses of particulate matter play a crucial role in the Earth's weather and climate systems. Clouds are critical for the radiation budget by reflecting incoming sunlight, resulting in a cooling effect, and by retaining outgoing heat radiated from the Earth's surface, leading to a warming effect. Therefore, changes in surface temperature are sensitive to the distribution of clouds (Romano, 2020), including their shape, size, altitude, and constituents. According to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (2023), 'the cloud feedback remains the largest contribution to overall uncertainty' in the global warming.

Meteorological satellites, in particular geostationary satellites, enable a continuous monitoring of the atmosphere. They offer a larger spatial coverage than ground-based instruments and, when several satellites observations are combined, they can deliver near real-time measurements at a global scale. Consequently, observations from geostationary satellites present themselves as candidates to study cloud distributions, though their instruments cannot avoid the common ill-posedness of atmospheric retrieval problems.

Particles of ice or water in the atmosphere are known as hydrometeors. They are a predominant constituent of cloud distributions and can even be the only component of clouds, depending on the definition of *cloud* (Spänkuch et al., 2022). Multidecadal climate data records show that the global average cloudiness exceeds 60% (Karlsson and Devasthale, 2018), while specialized observations indicate that clouds composed solely of frozen hydrometeors can surpass the global average cloudiness in tropical regions (Nazaryan et al., 2008). Frozen hydrometeors are complex and challenging when compared to their liquid counterpart (Kneifel et al., 2020).

This thesis focuses on quantifying concentrations of frozen hydrometeors, that is, atmospheric ice masses, from geostationary imagery. Unsurprisingly, the scientific community has already addressed the challenge of retrieving the vertically-integrated atmospheric ice mass, technically the ice water path (IWP), as further explained in section 2.4. However, the standard technique for quantifying IWP using geostationary images suffers from a major drawback: it is only applicable during daytime. This implies that about half of the observations are not used and downstream analyses, such as diurnal cycles, are incomplete.

The primary goal of this thesis is to overcome this limitation when retrieving IWP from geostationary imagery. This is achieved by identifying and employing an appropriate machine learning approach capable of providing probabilistic estimates for sophisticated downstream applications. Secondly, this thesis defies the challenge of offering vertical atmospheric ice density estimates, technically known as the ice water content (IWC), as well as complementary information, also from geostationary images.

To address the goals, the works in this thesis make virtually no physical assumptions in its foundation, although the different data sources contain assumptions. Therefore, these assumptions are inherited. Artificial neural networks (NNs) are trained with geostationary imagery to best reproduce retrievals from CloudSat, a sun-synchronous satellite that has been the standard for profiling ice clouds globally and is a usual reference for IWP and IWC. The information supplied to the NNs is chosen through exploration of the available geostationary data and the targets at stake, for example, consistent diurnal retrievals by considering only infrared measurements. The choices are guided by black-box experimental results rather than physical models. The NNs used also vary in size and predefined skills, for example, the ability to leverage spatial information, but all follow the same machine learning method capable of offering probabilistic estimates: non-linear quantile regression.

This framework also sets the limitations of this thesis. It is not considered to explore the explainability of the NNs and, therefore, the retrievals are simply treated as the output of a black-box. Although different NNs have been used, no exhaustive survey is conducted to find an optimal architecture, albeit it can be argued that such an optimization would consume plenty of resources for slight improvements. The lack of suitable probabilistic reference data hinders the validation of retrieval uncertainties. Therefore, the probabilistic nature of the retrievals is limited to the evaluation of the internal consistency with the training data, in the sense of comparing the distribution of random samples with the training data distribution.

Overall, the work presented in this thesis offers the possibility of getting a more complete picture of atmospheric ice masses from geostationary observations than the corresponding standard methods. A better understanding of the changing climate and repercussions of clouds in human activities can be facilitated by retrievals of this kind, which provide new information on the distribution and evolution of clouds.

2 Background

2.1 Physical principles

Figure 2.1 shows the ranges in the electromagnetic spectrum in which the remote sensing instruments used here measure. Naturally, measurements in different regions can capture distinct and complementary information. As an example, ice clouds strongly absorb radiation at $\lambda = 1.64\,\mu\text{m}$, but measurements at 94 GHz can interact much better with ice clouds, in the sense that they can provide better estimates, for instance, of cloud thickness. These characteristics are explained by the emission, absorption, and scattering properties of clouds, which are determined, among others, by their shape, density, and composition.

An important concept used throughout this thesis is the brightness temperature T_B. This parameter indicates the intensity of the thermal radiation of a physical body at a given wavelength. It is obtained by solving the Planck radiation law. That is, the brightness temperature is not necessarily equivalent to the actual temperature of the elements measured. Nevertheless, the use of brightness temperatures provides an intuitive interpretation when handling measurements across various wavelengths, in particular at the infrared and microwave regions; ice clouds exhibit a lower thermal emission than its background when seen from space, the Earth's surface, which results in lower brightness temperatures.

As exemplified above, the various regions of the electromagnetic spectrum,

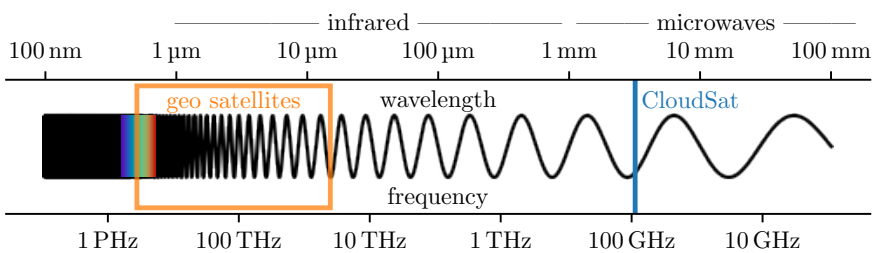


Figure 2.1: Part of the electromagnetic spectrum with the region measured by geostationary satellites highlighted in orange and the frequency of CloudSat in blue. The visible part of the spectrum is also illustrated.

and more precisely different wavelengths, may not offer the same performance for retrieving a geophysical parameter. The atmosphere is a complex medium consisting of many species of gas, liquid, and solid particles. In general, longer wavelengths penetrate better the atmosphere than shorter wavelengths, specifically in the presence of liquid or solid particles. This phenomenon is a consequence of the scattering and absorption properties of the various atmospheric constituents. Therefore, longer wavelengths, such as microwaves, are commonly preferred over infrared to retrieve the vertical distributions of frozen hydrometeors. This is relevant, for example, for thick ice clouds or in the presence of multilayer clouds.

2.2 Geostationary satellites

The geostationary orbit is located about 35 786 km above the equator. Satellites in gravitational equilibrium in this orbit are geostationary satellites. Their main characteristic is that they remain stationary relative to a fixed point on the Earth's surface. Meteorological geostationary satellites accommodate instruments that measure in the visible and infrared regions of the spectrum. One motivation for measuring in these regions is purely practical: the size required for the antenna. Although microwave geostationary imagers are possible (Lambrigtsen et al., 2022), the antenna size required to capture microwave signals is theoretically much larger than that of an antenna for visible or infrared signals. As an example, a rough calculation[1] indicates that achieving a resolution of 3 km at the sub-satellite point necessitates an antenna with a 0.13 m diameter for a wavelength of 11 µm, while for 94 GHz it is required a diameter of 38 m, almost 300 times larger. Therefore, placing and maintaining an instrument to capture microwaves at an altitude of 35 786 km is a cumbersome challenge.

Combining different geostationary satellites in a constellation, hereinafter referred to as a geo-ring, offers the possibility to have quasi-global continuous observations in the infrared region, only limited in latitude by the field of view of these satellites. The imagers on-board each satellite can differ in spatial resolution, revisit time, or measured wavelengths, among others. On the one hand, and as an example, the imager on-board Meteosat-9, the data source for paper I, located at the prime meridian during the time of the analysis, provided a 3 km resolution with a full-disc revisit time of 15 min for 11 spectral channels. On the other hand, the imager on-board the coexisting operational satellite GOES-11 (Menzel and Purdom, 1994), located at −135°, offered only five spectral channels, with resolutions between 1 and 8 km and a worse full-disc revisit time in the routine mode.

The characteristics of each instrument will depend on the agencies responsible for the satellite and, inevitably, the technological knowledge during their design phase. It should come as no surprise that data products tailored for one family of imagers cannot be applied directly to others. Nevertheless, ob-

[1] Following the formula diameter = altitude · wavelength/resolution. See, for example, *Physical Principles of Remote Sensing* (Rees, 2012) for details.

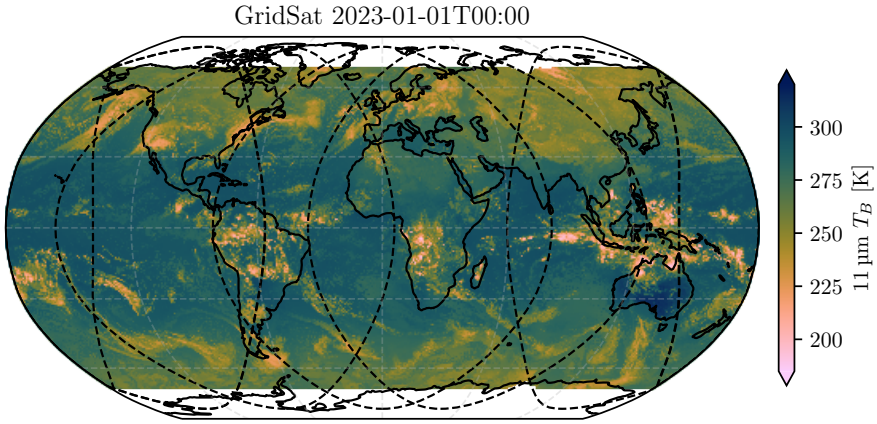


Figure 2.2: A 11 µm GridSat observation. According to the GridSat file metadata, the geo-ring consists of GOES-17 (−137.2°), GOES-18 (−137.0°), GOES-16 (−75.2°), Meteosat-11 (0.0°), Meteosat-9 (45.5°), and Himawari 9 (140.7°); approximate field of view of each satellite indicated in dashed black lines.

servations from a geo-ring can be combined onto a predefined grid through selecting shared spectral channels and applying intersatellite normalization. Figure 2.2 illustrates this with a sample from GridSat (Knapp et al., 2011), which is a long record of globally gridded geostationary satellite observations. Consequently, these observations can be used to develop data products for the global scale.

2.3 CloudSat

In 2006 a new satellite joined the A-Train constellation (L'Ecuyer and Jiang, 2010) of Earth observing satellites: CloudSat (Stephens et al., 2002). Its single instrument is a radar operating at 94 GHz, a frequency determined to be an optimal compromise between maximizing sensitivity to cloud reflectivities and atmospheric attenuation. Radiation at this microwave frequency interacts well with clouds and, consequently, the main application is to provide reflectivities of cloud profiles and derive products from them. Figure 2.3 illustrates one application: retrievals of IWP and IWC. According to Stephens et al. (2018), the new dimension offered by CloudSat has aided in updating the interpretation of radiance observations as well as provided new insights and understanding of cloud processes.

CloudSat has been the gold standard for global cloud profiling. However, it has four characteristics that can be considered limitations for the study of cloud distributions on a global scale. They are listed and discussed in the following paragraphs.

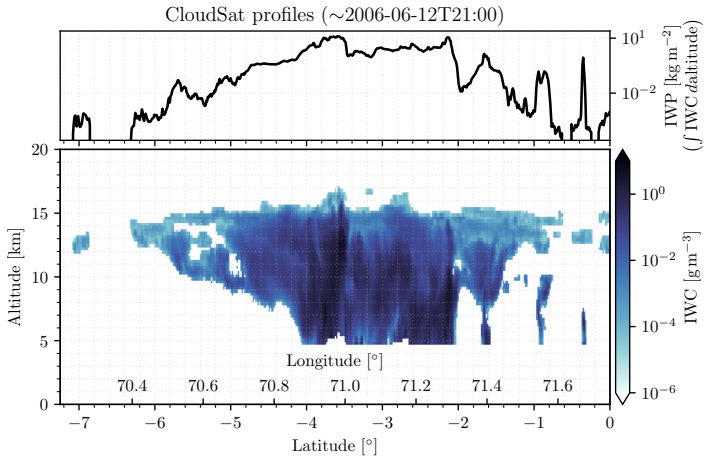


Figure 2.3: An example of atmospheric ice mass retrieval by the CloudSat product 2C-ICE (Deng et al., 2015).

Spatial sampling. The CloudSat radar provides vertical information at a 500 m resolution from the ground level up to 30 km. This should not be interpreted as any drawback. However, the limitation is in the planar dimensions: the cross-track effective field of view is approximately 1.4 km, which is also its cross-track coverage. This narrow coverage is akin to taking a slice of the atmosphere along the CloudSat orbit. Consequently, it is difficult to completely assess any spatial variability with the CloudSat reflectivities alone.

Sun-synchronous orbit. CloudSat is in a sun-synchronous orbit, that is, it always crosses the equator at the same time, specifically after 13:30 local solar time in ascending orbit. This implies three limitations. Firstly, CloudSat observations will be discontinuous for any location on Earth, although the observations will always be at the same local time. Secondly, its repeat cycle is 16 days. These two limitations, jointly with its narrow field of view, cause a third limitation, which is its spatial coverage and revisit time. Figure 2.4 illustrates the issue: a considerable portion of the atmosphere is unobserved by CloudSat. For most locations observed, the revisit time is 16 days. However, for fortunate locations it is 8 days, consisting of a day and a night observation, always at the same local mean solar time.

Temporal coverage. The concluding sentence in the previous paragraph applies only from its launch in mid 2006 through the first quarter of 2011. Thereafter, CloudSat has only acquired data in sunlit portions of its orbit due to a battery anomaly (Nayak, 2012). CloudSat has been operational until the time of writing, exceeding its three-year design life (Parkinson et al., 2006, p. 128), and has bypassed a series of unfortunate events, leading to its repositioning to the graveyard orbit of its partner satellite, which established

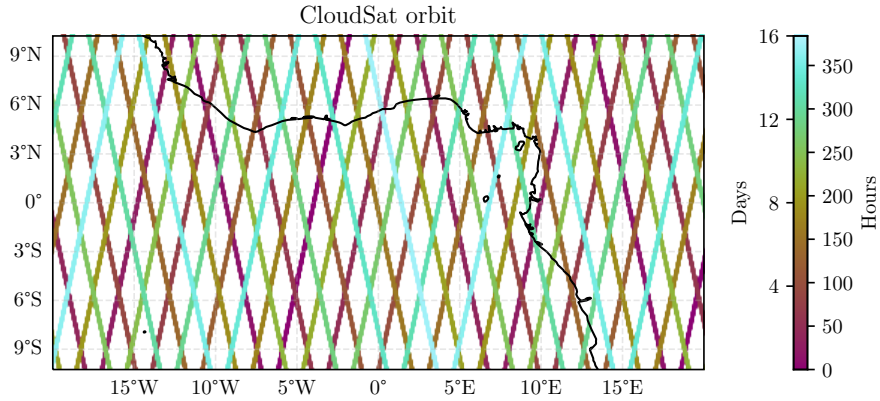


Figure 2.4: One full CloudSat repeat cycle for a region centred at 0° latitude and longitude. The cross-track field of view is not to scale, therefore even a larger part of the region is unobserved by CloudSat. The colours indicate the elapsed time since a hypothetical start time of observation.

the C-Train (Braun et al., 2019). Nonetheless, its expected end of life in late 2024 (World Meteorological Organization, 2023) will bring the time period containing CloudSat observations to an end.

The EarthCARE satellite (Illingworth et al., 2015) is the successor to CloudSat, containing also a 94 GHz cloud profiling radar. The Metop-SG B satellite will carry the Ice Cloud Imager (Bergadà et al., 2016; Kangas et al., 2014), a radiometer that will be used to derive data on the bulk ice mass from a new spectral range in space observations (Eriksson et al., 2020). However, both of these satellites will orbit in a sun-synchronous orbit, therefore presenting challenges similar to those of CloudSat, and they remain yet to be launched. Later upcoming missions will suffer from similar limitations. Consequently, observations from a geo-ring will persist as the unique source for a spatiotemporal, quasi-global continuous monitoring of cloud distributions.

2.4 Algorithms for visible and infrared imagery

The findings of Nakajima and King (1990) can be regarded as foundational for the retrieval of IWP from visible and infrared satellite imagers. Nakajima and King identified relationships between two spectral channels and two parameters, the cloud optical thickness τ and the effective droplet radius r_e. The first parameter quantifies the efficiency in preventing light from passing through a medium, while r_e is an average size parameter for the particles in the medium. It is worth noting that there are several different formulations of r_e for ice particles (McFarquhar and Heymsfield, 1998). Nakajima and King established that reflectances at 0.75 µm are primarily sensitive to τ, while reflectances at 2.16 µm are more sensitive to r_e.

Several algorithms, listed in paper I, retrieve IWP from visible and infrared radiances by adapting the Nakajima and King method. These algorithms estimate IWP as a function of τ and r_e, given by

$$\text{IWP} \propto \tau r_e \tag{2.1}$$

following Stephens (1978), and where the missing constant depends on each algorithm. However, this method presents three main shortcomings for the retrieval of IWP. Firstly, simplifying the particle size distribution to only r_e can render the relationship in eq. (2.1) incomplete for ice clouds (Mitchell et al., 2011). Secondly, it requires discriminating the ice cloud phase from the water phase, discouraging retrievals for mixed-phase clouds. Thirdly, the dependence on a visible channel to retrieve τ makes this method only applicable during daytime observations. There exists at least one technique that aims to address this last shortcoming by using infrared-only radiances (Wielicki et al., 1995). However, its application in eq. (2.1) should be considered experimental according to Minnis et al. (2011). The shortcomings of this physics-based approach may be overcome with machine learning. Few works, also listed in paper I, use machine-learning techniques for the retrieval of IWP. However, they face either a suboptimal choice of reference data when training the model for IWP retrieval or combine microwave and infrared observations, which makes the method inapplicable for geostationary imagers alone.

De Laat et al. (2017) and Yost et al. (2018) proposed physics-based algorithms to retrieve high IWC events from visible and infrared radiances. However, the method from de Laat et al. relies on eq. (2.1) and solar radiation. Arguably, it suffers from the shortcomings listed above. On the other hand, Yost et al. use a different approach by switching the method applied as a function of the solar radiance. In any case, both methods limit themselves to indicate the presence of IWC above a fixed threshold in the pixel of the geostationary image. In other words, they do not retrieve the vertical distribution of IWC.

2.5 Reshaping of the retrieval landscape with machine learning

The word *retrieval* has not been introduced in this thesis. In this context, it can be defined as the acquisition of information about a geophysical parameter through indirect measurements. Therefore, and mathematically, the measurements are functional results of the geophysical parameter. The inversion of such functions constitutes a retrieval.

The classical approach to solving remote sensing inversion problems starts with modelling the forward function. This function defines the perfect relationship between the measurements, the geophysical parameter, other parameters, and, depending on the interpretation used, experimental errors. That is, the forward function contains any physics necessary to relate the geophysical parameter to the measurement. Modelling the forward function correctly is fundamental to a successful retrieval method. Unsurprisingly, designing a forward model

that accurately matches the forward function can be difficult: the real physics may be too complex, or the level of detail required can still be uncertain. In addition, a forward model needs to be numerically efficient, particularly when large amounts of data are processed, as in the case of geostationary imagery.

Provided a forward model, one can design a retrieval method by mathematically inverting it. However, it is likely that this inversion is algebraically difficult or that it returns a non-unique solution, yielding an ill-posed problem. Variational methods can solve the inversion problem by finding the minimizer of a cost functional based on the forward model, but in their naive formulation they lack important information: the uncertainty of the retrieval. It is of little value to retrieve a quantity without a corresponding measure of error, particularly for sophisticated applications. Furthermore, in this context, external parameters or measurement noise can influence the retrieval and are a source of uncertainty.

One solution for this need is to consider the retrieved parameter as a random variable instead of a point estimate, and the measurement as a realization of another random variable. The Bayesian framework is well suited for this task, where the idea is that the measurement refines previous knowledge or assumptions of the parameter, yielding a posterior knowledge distribution for the parameter. The optimal estimation method (OEM, Rodgers, 2000) is widely used for solving atmospheric retrieval problems using the Bayesian framework. One characteristic of OEM is that it is an iterative method, since it repeatedly evaluates the Jacobian of the forward model. Therefore, one has to define the convergence criteria, which should include the possibility that there is no convergence. Furthermore, OEM is inherently limited to Gaussian statistics. In addition, this method also requires initial conditions, the formal specification of prior knowledge, and that the linearization of the forward model is reasonable, which all can have a substantial impact in the retrieval.

Although OEM has a solid statistical foundation, its reliance on a forward model, iterations that can be expensive, and predefined convergence criteria can render the retrieval sub-optimal. More general methods, such as Markov chain Monte Carlo, are computationally heavy. On the other hand, machine learning can be employed to execute efficient retrievals that overcome such inconveniences at the expense of a costly training step. The straightforward approach in this other paradigm is to program a machine to directly retrieve the geophysical parameter from the measurement. Therefore, solving the retrieval problem diverts from physical modelling and becomes instead a data-driven task, though it can be argued that the machine could mimic the inversion of a physical forward model.

This thesis followed the straightforward machine learning approach, but which is in principle consistent with general Bayesian methods. It is important to note that this approach relies on the quality of the data used to train the machine. Intuitively, the relationships learnt between inputs to the model and the corresponding expected outputs will be affected by the variability in the data representing these relationships. In the context of this thesis, the natural variability of the data is large and, therefore, strongly drives the uncertainty of the retrievals. Mathematical arguments are given in section 4.2.

In any case, the use of machine learning facilitates incorporating valuable features into the measurement, for example, spatiotemporal information. Furthermore, it is relatively easy to increase the capacity and expressivity of machine learning models. This is not only because there exist flexible models, but also because of the software and learning methods available as well as its vibrant scientific community. Machine learning models are virtually unlimited, including methods for obtaining uncertainties as in OEM. One of these methods is non-linear quantile regression, used throughout this thesis and detailed in section 4.2, which yields an equivalent of the posterior distribution obtained with Bayesian methods for scalar retrieval targets.

3 Summary of appended papers

Figure 3.1 illustrates the scope and main targets of each paper. Paper II builds upon the findings of paper I, as will be argued. The common methodology followed in both papers is outlined later in chapter 4. The following sections summarize each paper as well as selected results.

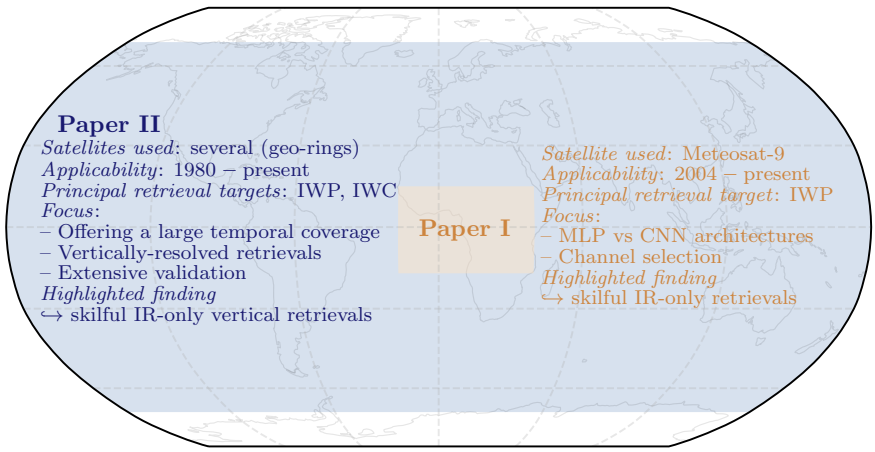


Figure 3.1: Summary of papers I and II. The shading indicates the areas covered by the papers.

3.1 Paper I – Ice water path retrievals from Meteosat-9 using quantile regression neural networks

The Meteosat satellites are the operational satellites at the prime meridian. For the last two decades Meteosats have carried the same imager: SEVIRI (Aminou et al., 1997; Schmid, 2000). This paper proposes retrieving IWP from SEVIRI images using quantile regression NNs trained against CloudSat IWP retrievals. In particular, it focusses on Meteosat-9 observations that cover a major part of the African continent, including a part of the Atlantic Ocean.

In total, five NN candidates for the retrieval of IWP are considered. It is first evaluated three candidates, where each consists of a pixel-wise multilayer perceptron (MLP) with the same number of hidden layers and hidden neurons, which were determined after an exploration with a development set. The difference between the three candidates is the number of input features required: one candidate required all SEVIRI spectral channels, that is, the 11 visible and infrared channels, while the other two candidates only supported infrared information that is known to not have any solar contamination, that is, thermal infrared. The difference between the infrared-only NNs also resided in the spectral channels used: one network required all seven SEVIRI thermal infrared channels, while the other aimed to approximate the SEVIRI predecessor using only two channels. These three candidates were trained and evaluated with daytime-only observations to fully explore the potential of the NN using visible and infrared radiances. As expected, this network best reproduced the retrievals from a held-out test dataset. However, the infrared-only NN retrievals followed closely, with better performance when more channels are used.

The satisfactory infrared-only retrievals encouraged to re-train the seven-infrared-channels MLP with a larger dataset containing both daytime and nighttime observations. The corresponding evaluation showed similar performance as the daytime-only trained NN. The fifth candidate consisted of a fully convolutional neural network (CNN) using the same spectral information as the seven-infrared-channels MLP. This CNN, grounded in similar retrieval problems and tuned over a small hyperparameter space with a development set, enabled incorporating spatial information into the retrieval. The lack of suitable ground-truth data hinders any exhaustive evaluation, including the probabilistic information of the retrieval method, but, in general, the CNN offered more satisfactory retrievals.

Retrievals given by the CNN were compared with physics-based retrievals from a readily available dataset, also derived from Meteosat-9. The comparison was two-fold: agreement with CloudSat retrievals and an analysis of diurnal cycles. Unsurprisingly, the machine-learning technique trained against CloudSat retrievals better reproduced the CloudSat retrievals. However, since CloudSat retrievals are considered an absolute ground truth, it was concluded that the CNN retrievals compared favourably. Concerning the diurnal cycles, a continental area and a maritime area, both expected to have a high IWP on average, were analysed. In general, the machine-learning and physics-based patterns of diurnal cycles correlate well, albeit they disagree in the IWP magnitude. However, given its dependence only on thermal infrared radiances, the NN has the advantage of providing the missing nighttime period in the physics-based approach.

3.2 Paper II – The Chalmers Cloud Ice Climatology: retrieval implementation and validation

Figure 3.1 highlights a limitation of paper I: its geographically-limited coverage. Paper II, also referred to as CCIC, extends the scope of these geostationary retrievals: not only for full-disc retrievals, as suggested in paper I, but also for a global coverage, only limited in latitude by the field of view of geostationary satellites. Consequently, more observations than those from SEVIRI are required; two datasets of globally gridded geostationary satellite observations are used as input data to a neural network. These datasets, which can be considered to complement each other, also cover a larger time period: one extends back to 1980, spanning more than forty years.

The larger coverage of the global datasets, both in the geographical and temporal domains, poses a principal challenge: harmonizing spectral information from different satellites. Consequently, paper II only uses the 11 µm radiances provided by these datasets. Given this limited information and the favourable retrievals by the CNN in paper I, another quantile regression CNN was adapted for this goal.

Ice water path is not the only retrieval target in CCIC. The same CNN targets a vertical resolution of the retrieval by providing IWC estimates, uncommon for retrievals from geostationary imagery. Additionally, CNN also provides estimates of cloud probabilities as a function of altitude, including the cloud class. The reference data to train the CCIC model was also derived from CloudSat products.

The evaluation of the retrievals against a held-out test dataset showed reasonable agreement. CCIC-retrieved IWP correlates well with CloudSat-retrieved IWP, as well as the corresponding zonal means. The performance of the CCIC-retrieved IWC is observed to be relatively worse than the performance of IWP retrieval. However, this should come as no surprise as determining the vertical distribution of the ice mass is, intuitively, a more complex problem than solving for IWP. This argument remains applicable to the retrieval of cloud classes. In any case, the vertically-resolved retrievals exceed expectations, as there is no suitable published reference for comparison, taking into account the limited penetration capability of 11 µm radiances.

The CCIC work goes beyond an evaluation against a held-out dataset. The IWP and IWC CCIC retrievals were compared with completely independent measurements outside the time period used for training the NN. These measurements, which cover different and distant regions, are taken from two series of flight campaigns as well as one ground-based cloud radar. All flight campaigns incorporated in-situ measurements of total hydrometeor water content, which consists of the mass of both frozen and liquid hydrometeors. However, most of the measurements were sampled in glaciated environments, and thus the in-situ measurements were considered here a proxy for IWC. It is observed that the CCIC IWC displays reasonable alignment with the in-situ measurements. Furthermore, observations from cloud radars on-board two of the flight cam-

paigns were processed to retrieve IWP and IWC. CCIC retrievals demonstrated similar agreement with the radar retrievals as with the in-situ measurements. Finally, CCIC retrievals were found to also exhibit reasonable performance when compared with ground-based radar IWC and IWP retrievals.

4 Methodology

Neural networks are at the core of the papers included in this thesis. Using statistical notation, this section outlines the common methodology followed to train NNs, an overview of their principles of operation, and the leading factors for some decisions. This section assumes a certain degree of familiarity with machine learning concepts, covered, for example, by Goodfellow et al. (2016).

4.1 Labelling the geostationary observations

The works in this thesis aim to reproduce CloudSat retrievals, here considered an absolute ground truth. Therefore, it is natural to frame the tasks in a supervised learning process, mainly as a regression problem, even though a classifier is required for the cloud class estimation in paper II. Supervised learning consists of maximizing the performance of an algorithm when mapping input data points $\mathbf{x}$ to known associated outcomes $\mathbf{y}$. Consequently, any supervised learning task will require a training dataset $\mathcal{D}_{\text{training}} = \{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{m}$ generated from a distribution $P_{\mathbf{X},\mathbf{Y}}$ on $\mathcal{X} \times \mathcal{Y}$. Note that the dimensionality of $\mathbf{x}$ and $\mathbf{y}$ will depend on the problem addressed. In this context, $\mathcal{X}$ are geostationary observations and $\mathcal{Y}$ CloudSat retrievals. The ideal situation is that a database $\mathcal{D}$ generated from $P_{X,Y}$ is readily available. However, $\mathcal{D}$ had to be constructed for each paper. In other words, the CloudSat retrievals had to be collocated with geostationary observations.

There is no unique way to construct $\mathcal{D}$: differences in projections of the data products and spatiotemporal resolutions make the collocation process challenging. The works in this thesis used more advanced approaches than a nearest neighbour search for the spatial collocation. This is hypothesized to improve the quality of $\mathcal{D}$, and consequently the quality of $\mathcal{D}_{\text{training}} \subset \mathcal{D}$ used to guide the algorithms. Temporal collocation was determined by the temporal resolution of the geostationary observations. Finally, the existence of two distinct CloudSat products that target the same variables, DARDAR-cloud (Delanoë and Hogan, 2010) and 2C-ICE (Deng et al., 2015), raised the question of which one to employ. The dataset of physics-based retrievals used for comparison in paper I was validated against DARDAR by the dataset authors. This data product was also the reference data in paper I. However, paper II used 2C-ICE, mainly motivated by a data policy.

4.2 Quantile regression

The goal of a neural network $f_{\boldsymbol{\theta}}$ is to learn optimal values for the learnable parameters $\boldsymbol{\theta}$ such that $f_{\boldsymbol{\theta}}$ can accurately map inputs from $\mathcal{X}$ to outputs in $\mathcal{Y}$, according to the underlying distribution. In the context of this thesis, the optimal $f_{\boldsymbol{\theta}}$ will not have a one-to-one correspondence: the limited sensitivity in the satellite instruments already introduces a degree of irreducible uncertainty. Therefore, it is appropriate to address this ill-posed problem by formulating $f_{\boldsymbol{\theta}}$ as the estimation of a conditional probability distribution $P_{Y|\mathbf{X}}$.

Koenker and Bassett (1978) introduced linear quantile regression. This method provides a more comprehensive relationship between two variables, an endogenous variable y and an exogenous $\mathbf{x}$, than the common least squares regression. It achieves a better description by estimating conditional quantiles. Given the cumulative distribution function F_Z for a distribution P_Z, the quantile z_τ at level $\tau \in [0, 1]$ is defined to be the value such that $z_\tau = \inf\{z : \tau \leq F_Z(z)\}$. The quantile loss function is defined as

$$\mathcal{L}_\tau(z_\tau, z) = \begin{cases} \tau|z - \hat{z}_\tau| & \text{if } \hat{z}_\tau < z \\ (1-\tau)|z - \hat{z}_\tau| & \text{otherwise} \end{cases} . \tag{4.1}$$

It can be proven that the expectation $\mathbb{E}_{Z\sim P_Z}[\mathcal{L}_\tau(\hat{z}_\tau, Z)]$ is minimized when $\hat{z}_\tau = z_\tau$ (Koenker, 2005, pp. 5–9).

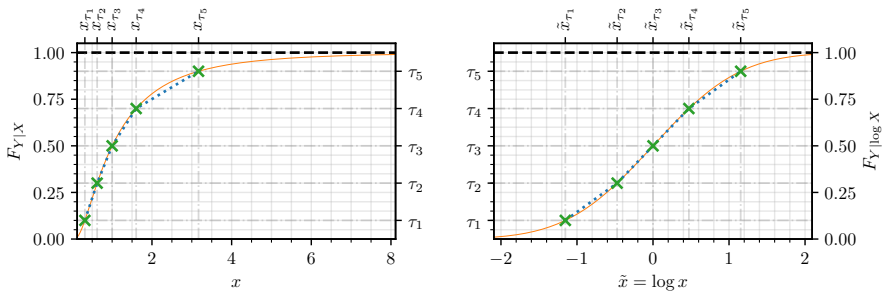
A quantile regression neural network (QRNN) is a non-linear extension of quantile regression, suitable for the non-linear nature of remote sensing retrievals. Therefore, using the notation in this thesis, a QRNN seeks $f_{\boldsymbol{\theta}}$ able to minimize $\mathbb{E}_{Y\sim P_{Y|\mathbf{X}}}[\mathcal{L}_\tau(f_{\boldsymbol{\theta}}(\mathbf{x}), Y)]$, where $\mathbf{x}$ are realizations of the random variable $\mathbf{X}$.

Pfreundschuh et al. (2018) analysed the performance of QRNNs as an alternative to estimating the posterior distribution of Bayesian remote sensing retrievals. Their work can be considered foundational for this thesis, as QRNNs are the unique method used here for the retrieval of ice masses. The actual implementation of QRNNs also follows the approach of Pfreundschuh et al.: $f_{\boldsymbol{\theta}}$ is designed to estimate a set of quantiles at fixed levels $\mathcal{T}$. That is, the loss function of QRNNs for a sample $(\mathbf{x}, y)$ is given by

$$\mathcal{L}\left((f_{\boldsymbol{\theta}}(\mathbf{x}), y\right) = \frac{1}{|\mathcal{T}|} \sum_{\tau\in\mathcal{T}} \mathcal{L}_\tau(f_{\boldsymbol{\theta}}(\mathbf{x}; \tau), y) \tag{4.2}$$

where $f_{\boldsymbol{\theta}}(\mathbf{x}; \tau)$ indicates the estimated quantile at level τ. Given the finite set $\mathcal{T}$ and the fact that the exact $P_{Y|\mathbf{X}}$ is not known, the output of the network results in a quantile-parameterized distribution (QPD), as illustrated in fig. 4.1. The distribution $P_{Y|\mathbf{X}}$ can then be approximated with, for example, a linear extension of the predicted QPD.

It should be emphasized that the regular and bold font faces in this section are not arbitrary. This formulation of quantile regression only allows for scalar target variables. Consequently, it is only possible to obtain marginal distributions for each variable constituting vector retrievals, as in the case of IWC retrievals or spatial distributions for IWP. How to perform the so-called



(a) QPD from quantile regression.

(b) Monotonic transform applied to (a).

Figure 4.1: Optimal quantile regression examples for arbitrary random variables and quantile levels $\mathcal{T} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The orange curve is the true cumulative distribution function $F_{Y|\cdot}$, the markers indicate quantiles estimated with quantile regression, therefore a QPD, and the blue curve the corresponding linear interpolation of the QPD.

vector quantile regression to estimate joint densities is an active area of research and is out of the scope of this thesis.

In any case, the essence of QRNNs consists of training NN to estimate quantiles of the irreducible uncertainty, also known as aleatoric uncertainty. These quantiles can be processed to obtain, for example, an estimate for the expected value of a retrieval. The QRNN approach does not require assumptions on the target variable and can be completely data-driven. In particular, it is not required to determine a plausible family of distributions for the target variable. This enables to obtain non-Gaussian uncertainties, which are expected in atmospheric ice retrievals; in the hypothetical case that the retrieval returns the distribution of the reference data, such distribution is far from being Gaussian, as illustrated in fig. 4.2. It is also reasonable to assume a context of heteroskedasticity in the works of this thesis, that is, that the uncertainties are not constant. Intuitively, the uncertainty of an IWP retrieval for a cloud-free pixel should be minimal when compared against a cloudy pixel. The non-linear nature of QRNNs is well-suited for this challenge. In addition, quantile regression has another property: invariability to monotonic transformations. Figure 4.1(b) provides a graphical proof of this property, but mathematically: if h is a non-decreasing function, then $\tau = \Pr(Z \leq z_\tau) = \Pr(h(Z) \leq h(z_\tau)) = \tau$. Finally, advanced network architecture components, for instance, convolutional layers, are easily integrated into a QRNN.

4.3 Network architectures

The use of neural networks for the retrieval of ice masses from geostationary observations is relatively novel. Therefore, there is a shortage of references to help determine an appropriate neural network architecture. As a consequence, it can be considered necessary to design the architecture from scratch. Given the

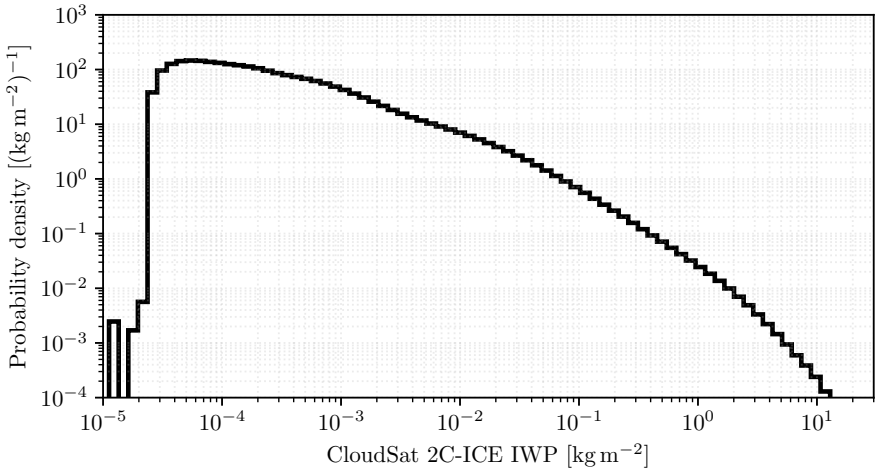


Figure 4.2: Probability density function obtained binning one year of CloudSat 2C-ICE IWP retrievals. Note the logarithmic axes. The plot shows only half of the distribution, since $\Pr(\text{IWP} \leq 10^{-5}\,\text{kg}\,\text{m}^{-2}) = 49.7\%$.

black-box nature of NNs, the design of an architecture will inevitably depend on the intuition, creativity, and expertise of the designer. Effective training algorithms can compensate for the use of excessively complex architectures, but the use of superfluous operations will result in longer inference runtimes or larger computational footprints. Therefore, it may be deemed necessary to optimize the hyperparameters.

Paper I started with few assumptions about a suitable architecture. First, several principles were gathered to propose a pixel-wise MLP as the QRNN backbone. The hyperparameters of this MLP were tuned over a regular grid to obtain a satisfactory performance. Although it can be argued that a grid search is not as efficient as, for example, a Bayesian optimization, the search space was small and manageable. Therefore, hyperparameter tuning for the MLP was simple and straightforward.

The CNNs used for paper I and paper II consisted of many more parameters and required more input data, that is, an image instead of a single pixel. This posed a challenge: both hyperparameter tuning and designing architectures from scratch can result prohibitive as well as distracting. Consequently, CNN backbones hypothesized to work reasonably well for the retrievals of atmospheric ice were used. Only paper I focused on the hyperparameter optimization for the CNN, even though being quite limited when compared with the optimization for the MLP.

The CNN backbones are fully CNNs. This is a type of network that preserves the spatial dimensions of the input data. Therefore, it enables one retrieval per input pixel. Furthermore, fully CNNs support different input sizes. The two papers in this thesis use slightly different backbones, mainly motivated by new machine learning papers and discussions. However, they share

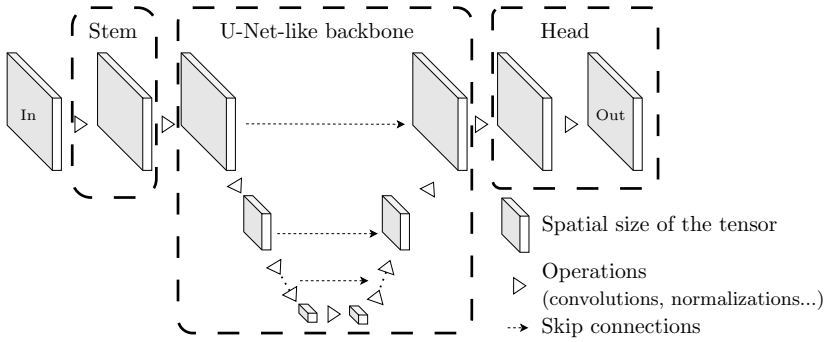


Figure 4.3: Schema of a U-Net-like architecture.

a common idea: the use of a U-Net-like architecture (Ronneberger et al., 2015). Figure 4.3 depicts a U-Net-like backbone in a QRNN. The U-Net architecture was specifically designed for semantic segmentation tasks, that is, pixel-wise classification; semantic segmentation with an infinite number of classes can be seen as a proxy for quantile regression. In a U-Net-like architecture, the encoder path is hypothesized to be able to extract high-level features, while the decoder path recovers the original spatial size and merges low- and high-level features, aided by skip connections.

Figure 4.3 is a simplistic representation of the network in paper I, in which the network head also follows a fully CNN. The network in paper II also follows the schema in Figure 4.3. However, it contains multiple heads, one per target variable; in this case, not all network heads lead to solve a quantile regression problem, since there are categorical target variables.

4.4 Training

The training of the networks followed standard machine learning practices: using a training, validation, and test split of $\mathcal{D}$, normalization of the inputs as well as throughout the network, mini-batch stochastic gradient descent, non-constant learning rates for the parameter updates, and data augmentation techniques, among others. In addition, log-based transforms were applied to the reference data, empirically observed to improve retrieval performance. These transforms leverage the property of invariability to monotonic transformations of quantile regression, given that both the transform and its inverse are monotonic.

The network in paper II outputs several variables using the same backbone. In this case, the loss function minimized was the sum of the losses for each variable. Using this definition, the errors for one variable might lead the parameter optimization of the backbone, resulting in better performance for this variable. Nevertheless, it was not considered to use more complex approaches, for example, introducing a scaling factor for each term; the straightforward approach already provided skilful retrievals.

Regardless of the number of output variables, the training was monitored not

only by the loss curve, but also by summary statistics, such as the mean squared error of the retrieval expected value. Monitoring these auxiliary parameters offered a physical interpretation for the training curve. However, the success of the training was only determined after visual examination of more descriptive tools, such as scatter plots. This resulted in a tedious process. Consequently, the final network models were chosen as a trade-off between the time needed for their implementation, training, and evaluation, and the retrieval performance.

One important requirement throughout the trainings was access to specialized hardware, which makes the learning process efficient, in particular powerful general purpose graphics processing units (GPUs) as well as fast storage systems. GPUs excel at matrix multiplication in parallel. Therefore, neural networks, which contain substantial linear algebra operations, make GPUs a natural fit for their optimization. Goodfellow et al. (2016, section 12.1.2) elaborate on this compelling argument. Fortunately, high-level machine learning libraries simplify working with such systems. This was particularly beneficial for paper II, were the PyTorch (Paszke et al., 2019) implementation of distributed data parallel training (Li et al., 2020) was used to train using four GPUs. In this technique, each GPU has a copy of the model and optimizer. Different inputs are sent to each GPU to compute the gradients for the parameter updates. However, the gradients will be different in each GPU as the inputs are different. Distributed data parallel training aggregates and synchronizes all gradients before applying the optimizer, resulting in the same updated copy across GPUs.

5 Contributions to the field and conclusion

This thesis could be categorized as the application of machine learning for remote sensing retrievals. The principal focus is on retrievals of atmospheric ice masses from spaceborne instruments. In particular, the use of geostationary satellites for this goal is motivated, since they offer an extensive record of spatially and temporally continuous observations. However, geostationary satellites measure a region of the electromagnetic spectrum that hinders the retrieval of atmospheric ice masses. One main contribution of this thesis, based on paper I, shows that through machine learning methods one can obtain retrievals that compare favourably to existing, physics-based methods. In particular, infrared-only retrievals demonstrate an unprecedented level of performance, and this retrieval scheme is applicable regardless of the time of day.

The continuous record of geostationary satellite observations can be leveraged to analyse local, regional, or global patterns in cloud distributions. This thesis makes two more significant contributions with paper II. Firstly, it is possible to provide skilful three-dimensional retrievals of cloud ice masses from one single infrared channel, partly owing to the use of spatial information in the retrieval. Secondly, the retrieval scheme developed for CCIC is applicable to most geostationary satellites and, therefore, addresses a global scale, limited in latitude by their field of view. The CCIC work can target a broad range of applications, such as long-term climate analyses, studies of cloud processes, or even applications for the aviation industry. Ongoing work involves comparing CCIC with other retrievals and formulating a study based on CCIC.

An intentional by-product of CCIC was to facilitate executing CCIC retrievals by the scientific community. Consequently, a software package has been developed. It has been designed to require no more computational power than any modern workstation. The software package also includes, among others, options to tailor the retrievals to regions of interest and an efficient output file size through a custom compression algorithm.

The machine learning models presented in this thesis can probably benefit from more advanced architectures, training schemes, or additional features, for example, a time dimension to incorporate cloud dynamics into the retrieval. However, it is unclear where the performance limit is: there is an ultimate constraint set by physics. Hence, other aspects could be improved, such as more

efficient neural network architectures, the introduction of vector quantile regression, or a validation of the retrieval uncertainties, though adequate reference data is lacking for an exhaustive validation. In any case, the contributions in this thesis can, on the one hand, influence established algorithms and datasets, and, on the other hand, complement retrievals from upcoming mission-specific satellite instruments.

References

Aminou, D. M. A., B. Jacquet and F. Pasternak (1997). 'Characteristics of the Meteosat Second Generation (MSG) Radiometer/Imager: SEVIRI'. In: Aerospace Remote Sensing '97. Ed. by H. Fujisada. London, United Kingdom, pp. 19–31. DOI: 10.1117/12.298084. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=932561 (visited on 2023-08-07) (cit. on p. 13).

Bergadà, M., M. Labriola, R. González et al. (2016). 'The Ice Cloud Imager (ICI) Preliminary Design and Performance'. In: *2016 14th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*. 2016 14th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad). Espoo: IEEE, pp. 27–31. ISBN: 978-1-5090-2951-8. DOI: 10.1109/MICRORAD.2016.7530498. URL: https://ieeexplore.ieee.org/document/7530498/ (visited on 2023-08-07) (cit. on p. 9).

Braun, B. M., T. H. Sweetser, C. Graham and J. Bartsch (2019). 'CloudSat's A-Train Exit and the Formation of the C-Train: An Orbital Dynamics Perspective'. In: *2019 IEEE Aerospace Conference*. 2019 IEEE Aerospace Conference. Big Sky, MT, USA: IEEE, pp. 1–10. ISBN: 978-1-5386-6854-2. DOI: 10.1109/AERO.2019.8741958. URL: https://ieeexplore.ieee.org/document/8741958/ (visited on 2023-08-07) (cit. on p. 9).

De Laat, A., E. Defer, J. Delanoë, F. Dezitter, A. Gounou, A. Grandin, A. Guignard, J. F. Meirink, J.-M. Moisselin and F. Parol (2017). 'Analysis of Geostationary Satellite-Derived Cloud Parameters Associated with Environments with High Ice Water Content'. In: *Atmospheric Measurement Techniques* 10.4, pp. 1359–1371. ISSN: 1867-8548. DOI: 10.5194/amt-10-1359-2017. URL: https://amt.copernicus.org/articles/10/1359/2017/ (visited on 2022-10-27) (cit. on p. 10).

Delanoë, J. and R. J. Hogan (2010). 'Combined CloudSat-CALIPSO-MODIS retrievals of the properties of ice clouds'. In: *Journal of Geophysical Research: Atmospheres* 115.D4. DOI: 10.1029/2009JD012346 (cit. on p. 17).

Deng, M., G. G. Mace, Z. Wang and E. Berry (2015). 'CloudSat 2C-ICE Product Update with a New Ze Parameterization in Lidar-Only Region'. In: *Journal of Geophysical Research: Atmospheres* 120.23, pp. 12, 198–12, 208. DOI: 10.1002/2015JD023600 (cit. on pp. 8, 17).

Eriksson, P., B. Rydberg, V. Mattioli, A. Thoss, C. Accadia, U. Klein and S. A. Buehler (2020). 'Towards an Operational Ice Cloud Imager (ICI)

Retrieval Product'. In: *Atmospheric Measurement Techniques* 13.1, pp. 53–71. ISSN: 1867-8548. DOI: 10.5194/amt-13-53-2020. URL: https://amt.copernicus.org/articles/13/53/2020/ (visited on 2023-08-07) (cit. on p. 9).

Goodfellow, I., Y. Bengio and A. Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press (cit. on pp. 17, 22).

Illingworth, A. J., H. W. Barker, A. Beljaars et al. (2015). 'The EarthCARE Satellite: The Next Step Forward in Global Measurements of Clouds, Aerosols, Precipitation, and Radiation'. In: *Bulletin of the American Meteorological Society* 96.8, pp. 1311–1332. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-12-00227.1. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-D-12-00227.1 (visited on 2023-08-07) (cit. on p. 9).

Intergovernmental Panel on Climate Change (2023). 'The Earth's Energy Budget, Climate Feedbacks and Climate Sensitivity'. In: *Climate Change 2021 – the Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 923–1054. DOI: 10.1017/9781009157896.009 (cit. on p. 3).

Kangas, V., S. D'Addio, U. Klein, M. Loiselet, G. Mason, J.-C. Orlhac, R. Gonzalez, M. Bergada, M. Brandt and B. Thomas (2014). 'Ice Cloud Imager Instrument for MetOp Second Generation'. In: *2014 13th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*. 2014 Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad). Pasadena, CA, USA: IEEE, pp. 228–231. ISBN: 978-1-4799-4645-7. DOI: 10.1109/MicroRad.2014.6878946. URL: http://ieeexplore.ieee.org/document/6878946/ (visited on 2023-08-07) (cit. on p. 9).

Karlsson, K.-G. and A. Devasthale (2018). 'Inter-Comparison and Evaluation of the Four Longest Satellite-Derived Cloud Climate Data Records: CLARA-A2, ESA Cloud CCI V3, ISCCP-HGM, and PATMOS-x'. In: *Remote Sensing* 10.10, p. 1567. ISSN: 2072-4292. DOI: 10.3390/rs10101567 (cit. on p. 3).

Knapp, K. R., S. Ansari, C. L. Bain et al. (2011). 'Globally Gridded Satellite Observations for Climate Studies'. In: *Bulletin of the American Meteorological Society* 92.7, pp. 893–907. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/2011BAMS3039.1. URL: https://journals.ametsoc.org/doi/10.1175/2011BAMS3039.1 (visited on 2023-08-07) (cit. on p. 7).

Kneifel, S., J. Leinonen, J. Tyynelä, D. Ori and A. Battaglia (2020). 'Scattering of Hydrometeors'. In: *Satellite Precipitation Measurement: Volume 1*. Ed. by V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura and F. J. Turk. Cham: Springer International Publishing, pp. 249–276. ISBN: 978-3-030-24568-9. DOI: 10.1007/978-3-030-24568-9_15 (cit. on p. 3).

Koenker, R. (2005). *Quantile regression*. Cambridge university press (cit. on p. 18).

Koenker, R. and G. Bassett (1978). 'Regression Quantiles'. In: *Econometrica* 46.1, p. 33. ISSN: 00129682. DOI: 10.2307/1913643. JSTOR: 1913643. URL: https://www.jstor.org/stable/1913643?origin=crossref (visited on 2023-08-07) (cit. on p. 18).

L'Ecuyer, T. S. and J. H. Jiang (2010). 'Touring the Atmosphere Aboard the A-Train'. In: *Physics Today* 63.7, pp. 36–41. ISSN: 0031-9228, 1945-0699. DOI: 10.1063/1.3463626. URL: http://physicstoday.scitation.org/doi/10.1063/1.3463626 (visited on 2023-08-07) (cit. on p. 7).

Lambrigtsen, B., P. Kangaslahti, O. Montes, N. Niamsuwan, D. Posselt, J. Roman, M. Schreier, A. Tanner, L. Wu and I. Yanovsky (2022). 'A Geostationary Microwave Sounder: Design, Implementation and Performance'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15, pp. 623–640. ISSN: 1939-1404, 2151-1535. DOI: 10.1109/JSTARS.2021.3132238. URL: https://ieeexplore.ieee.org/document/9634842/ (visited on 2023-08-07) (cit. on p. 6).

Li, S., Y. Zhao, R. Varma et al. (2020). 'PyTorch Distributed: Experiences on Accelerating Data Parallel Training'. In: *Proceedings of the VLDB Endowment* 13.12, pp. 3005–3018. ISSN: 2150-8097. DOI: 10.14778/3415478.3415530. URL: https://dl.acm.org/doi/10.14778/3415478.3415530 (visited on 2023-08-07) (cit. on p. 22).

McFarquhar, G. M. and A. J. Heymsfield (1998). 'The Definition and Significance of an Effective Radius for Ice Clouds'. In: *Journal of the Atmospheric Sciences* 55.11, pp. 2039–2052. ISSN: 0022-4928, 1520-0469. DOI: 10.1175/1520-0469(1998)055<2039:TDASOA>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0469(1998)055%3C2039:TDASOA%3E2.0.CO;2 (visited on 2023-08-07) (cit. on p. 9).

Menzel, W. P. and J. F. W. Purdom (1994). 'Introducing GOES-I: The First of a New Generation of Geostationary Operational Environmental Satellites'. In: *Bulletin of the American Meteorological Society* 75.5, pp. 757–781. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/1520-0477(1994)075<0757:IGITFO>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0477(1994)075%3C0757:IGITFO%3E2.0.CO;2 (visited on 2023-08-07) (cit. on p. 6).

Minnis, P., S. Sun-Mack, D. F. Young et al. (2011). 'CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms'. In: *IEEE Transactions on Geoscience and Remote Sensing* 49.11, pp. 4374–4400. ISSN: 0196-2892, 1558-0644. DOI: 10.1109/TGRS.2011.2144601. URL: http://ieeexplore.ieee.org/document/5783916/ (visited on 2023-08-07) (cit. on p. 10).

Mitchell, D. L., R. P. Lawson and B. Baker (2011). 'Understanding Effective Diameter and Its Application to Terrestrial Radiation in Ice Clouds'. In: *Atmospheric Chemistry and Physics* 11.7, pp. 3417–3429. ISSN: 1680-7324. DOI: 10.5194/acp-11-3417-2011. URL: https://acp.copernicus.org/articles/11/3417/2011/ (visited on 2023-08-02) (cit. on p. 10).

Nakajima, T. and M. D. King (1990). 'Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory'. In: *Journal of the Atmospheric Sciences* 47.15, pp. 1878–1893. ISSN: 0022-4928, 1520-0469. DOI: 10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0469(1990)047%3C1878:DOTOTA%3E2.0.CO;2 (visited on 2023-08-07) (cit. on pp. 9, 10).

Nayak, M. (2012). 'CloudSat Anomaly Recovery and Operational Lessons Learned'. In: *SpaceOps 2012 Conference*. SpaceOps 2012. Stockholm, Sweden: American Institute of Aeronautics and Astronautics. DOI: 10.2514/6.2012-1295798. URL: https://arc.aiaa.org/doi/10.2514/6.2012-1295798 (visited on 2023-08-07) (cit. on p. 8).

Nazaryan, H., M. P. McCormick and W. P. Menzel (2008). 'Global characterization of cirrus clouds using CALIPSO data'. In: *Journal of Geophysical Research: Atmospheres* 113.D16. DOI: https://doi.org/10.1029/2007JD009481. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD009481. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009481 (cit. on p. 3).

Parkinson, C., A. Ward and M. King (2006). 'Earth science reference handbook'. In: *National Aeronautics and Space Administration: Washington, DC, USA* (cit. on p. 8).

Paszke, A., S. Gross, F. Massa et al. (2019). 'PyTorch: An Imperative Style, High-Performance Deep Learning Library'. In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett. Curran Associates, Inc., pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf (visited on 2023-08-01) (cit. on p. 22).

Pfreundschuh, S., P. Eriksson, D. Duncan, B. Rydberg, N. Håkansson and A. Thoss (2018). 'A Neural Network Approach to Estimating a Posteriori Distributions of Bayesian Retrieval Problems'. In: *Atmospheric Measurement Techniques* 11.8, pp. 4627–4643. ISSN: 1867-8548. DOI: 10.5194/amt-11-4627-2018. URL: https://amt.copernicus.org/articles/11/4627/2018/ (visited on 2023-08-07) (cit. on p. 18).

Rees, W. G. (2012). *Physical Principles of Remote Sensing*. 3rd ed. Cambridge University Press. DOI: 10.1017/CBO9781139017411 (cit. on p. 6).

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. WORLD SCIENTIFIC. DOI: 10.1142/3171 (cit. on p. 11).

Romano, F. (2020). 'Editorial for the Special Issue "Remote Sensing of Clouds"'. In: *Remote Sensing* 12.24, p. 4085. ISSN: 2072-4292. DOI: 10.3390/rs12244085. URL: https://www.mdpi.com/2072-4292/12/24/4085 (visited on 2023-08-07) (cit. on p. 3).

Ronneberger, O., P. Fischer and T. Brox (2015). 'U-Net: Convolutional Networks for Biomedical Image Segmentation'. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi. Vol. 9351. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28. URL: http://link.springer.com/10.1007/978-3-319-24574-4_28 (visited on 2023-08-07) (cit. on p. 21).

Schmid, J. (2000). 'The SEVIRI instrument'. In: *Proceedings of the 2000 EUMETSAT meteorological satellite data user's conference, Bologna, Italy*. Vol. 29, pp. 13–32 (cit. on p. 13).

Spänkuch, D., O. Hellmuth and U. Görsdorf (2022). 'What Is a Cloud? Toward a More Precise Definition'. In: *Bulletin of the American Meteorological Society*

103.8, E1894–E1929. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-21-0032.1. URL: https://journals.ametsoc.org/view/journals/bams/103/8/BAMS-D-21-0032.1.xml (visited on 2023-07-28) (cit. on p. 3).

Stephens, G. L. (1978). 'Radiation Profiles in Extended Water Clouds. II: Parameterization Schemes'. In: *Journal of the Atmospheric Sciences* 35.11, pp. 2123–2132. ISSN: 0022-4928, 1520-0469. DOI: 10.1175/1520-0469(1978)035<2123:RPIEWC>2.0.CO;2. URL: http://journals.ametsoc.org/doi/10.1175/1520-0469(1978)035%3C2123:RPIEWC%3E2.0.CO;2 (visited on 2023-08-07) (cit. on p. 10).

Stephens, G., D. Winker, J. Pelon, C. Trepte, D. Vane, C. Yuhas, T. L'Ecuyer and M. Lebsock (2018). 'CloudSat and CALIPSO within the A-Train: Ten Years of Actively Observing the Earth System'. In: *Bulletin of the American Meteorological Society* 99.3, pp. 569–581. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-16-0324.1. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-D-16-0324.1 (visited on 2023-08-07) (cit. on p. 7).

Stephens, G. L., D. G. Vane, R. J. Boain et al. (2002). 'THE CLOUDSAT MISSION AND THE A-TRAIN: A New Dimension of Space-Based Observations of Clouds and Precipitation'. In: *Bulletin of the American Meteorological Society* 83.12, pp. 1771–1790. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-83-12-1771. URL: https://journals.ametsoc.org/doi/10.1175/BAMS-83-12-1771 (visited on 2023-08-07) (cit. on p. 7).

Wielicki, B. A., B. R. Barkstrom, B. A. Baum et al. (1995). *Clouds and the Earth's Radiant Energy System (CERES) Algorithm Theoretical Basis Document Volume III—Cloud Analyses and Determination of Improved Top of Atmosphere Fluxes (Subsystem 4)*. URL: https://www-pm.larc.nasa.gov/ceres/pub/journals/Minnis.etal.95.III.pdf (cit. on p. 10).

World Meteorological Organization (2023). *List of all Satellites*. Observing Systems Capability Analysis and Review Tool. URL: https://space.oscar.wmo.int/satellites (visited on 2023-08-01) (cit. on p. 9).

Yost, C. R., K. M. Bedka, P. Minnis et al. (2018). 'A Prototype Method for Diagnosing High Ice Water Content Probability Using Satellite Imager Data'. In: *Atmospheric Measurement Techniques* 11.3, pp. 1615–1637. ISSN: 1867-8548. DOI: 10.5194/amt-11-1615-2018. URL: https://amt.copernicus.org/articles/11/1615/2018/ (visited on 2023-01-31) (cit. on p. 10).