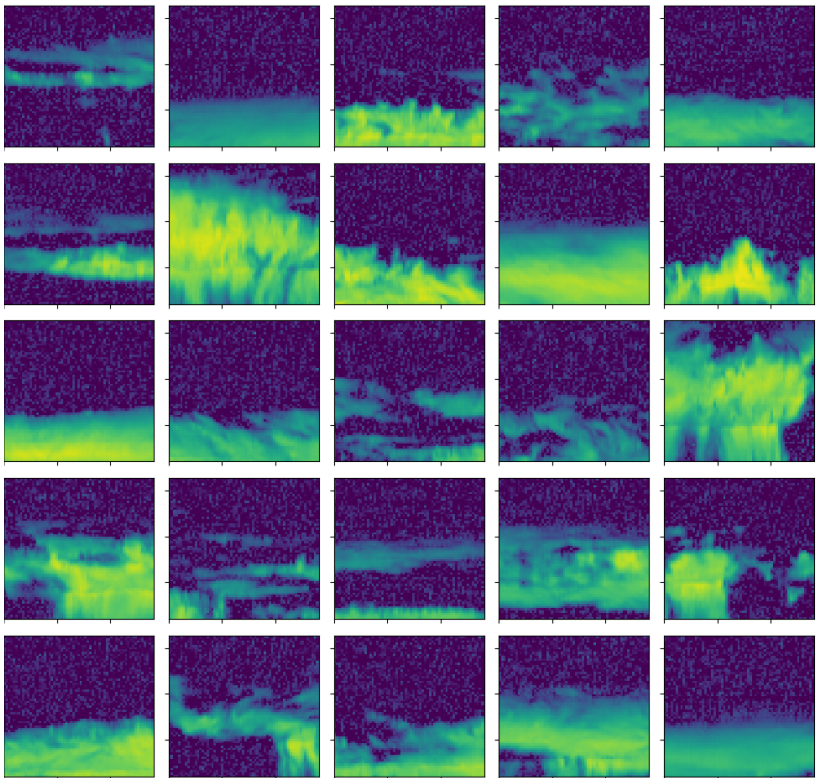
CHALMERS
UNIVERSITY OF TECHNOLOGY



Generation of atmospheric cloud fields using generative adversarial networks

Master's thesis in Complex Adaptive Systems & Physics and Astronomy

REBECKA RILEMARK
CARL SVENSSON

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020:SEEX30

Generation of atmospheric cloud fields using generative adversarial networks

REBECKA RILEMARK
CARL SVENSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Space, Earth and Environment
Microwave and optical remote sensing
Chalmers University of Technology
Gothenburg, Sweden 2020

Generation of atmospheric cloud fields using generative adversarial networks
REBECKA RILEMARK
CARL SVENSSON

Supervisor: Simon Pfreundschuh, Microwave and optical remote sensing
Examiner: Patrick Eriksson, Microwave and optical remote sensing

Master's Thesis 2020:SEEX30
Department of Space, Earth and Environment
Microwave and optical remote sensing
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Examples of synthetic vertical radar scenes of cloud fields generated with a conditional generative adversarial network (CGAN).

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Generation of atmospheric cloud fields using generative adversarial networks
REBECKA RILEMARK
CARL SVENSSON
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Cloud fields have a great impact on the weather and climate of the Earth. Modern climate models that include feedback from cloud fields present large uncertainties in their predictions. To improve models of climate systems large amounts of high-quality data of cloud fields are necessary. The Cloud Profiling Radar (CPR) on the CloudSat satellite provides high-quality vertical data of cloud fields but is limited in its geographical coverage. The Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite provides data that only discerns the top layer of clouds but has far greater spatial coverage.

This project implements two generative adversarial networks (GAN), with the intention of extending the current available CPR data set. One network generates vertical cloud data using random noise as input. The other network uses MODIS data as input to generate vertical cloud data tied to a specific geolocation, this network is accordingly a conditional generative adversarial network (CGAN). The two neural networks are compared to a common method used for generating synthetic cloud fields, the Iterative Amplitude Adjusted Fourier Transform (IAAFT). The methods are compared in regard to multiple different statistical and physical properties, including ice water path, cloud-top height, and spatial autocorrelation.

The results of both GAN are promising with regard to generating realistic cloud fields. Individual generated cloud scenes from GAN and CGAN show a stronger visual resemblance to real radar scenes than scenes generated with IAAFT. The three different methods vary in performance when analysed based on the statistics of their output. For example, the IAAFT method outperforms both the GAN and the CGAN when it comes to recreating the large scale vertical cloud distributions, but shows a clear weakness when it comes to capturing the internal structure of cloud fields, measured by the autocorrelation. A reoccurring problem with the GAN is the overfitting of the training data and mode collapse. This aspect needs further improvement through re-training of the networks with a larger data set and possibly uneven weighting between the generator and discriminator. After the inclusion of such improvements, the two GAN developed in this project are expected to show promising results for including generated radar data in applications such as verification of climate models and development of retrieval algorithms.

Keywords: Cloud fields, climate, remote sensing, MODIS, CloudSat, neural networks, generative adversarial networks, ice water path.

Acknowledgements

We would like to start by thanking both our thesis advisor Patrick Eriksson and our supervisor Simon Pfreundschuh at the Department of Space, Earth and Environment at Chalmers University of Technology. Both of which have provided us with all the support we could have asked for and have answered any and all questions we could think of. Thank you for sharing your expertise with us and pushing us to always perform better.

We would like to thank everyone who we have interacted with at the department. They have all shown great encouragement and we are very grateful to them for making us feel part of the research community also in these trying times.

Lastly, we would also like to thank our families and friends for putting up with us when we perhaps focused a bit too much on the project.

Rebecka Rilemark, Carl Svensson, Gothenburg, October 2020

Contents

List of Figures

List of Tables

List of Abbreviations

CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
CPR	Cloud Profiling Radar
CTH	Cloud-Top Height
ECMWF	European Centre for Medium-Range Weather Forecasts
GAN	Generative Adversarial Network
IAAFT	Iterative Amplitude Adjusted Fourier Transform
IWC	Ice Water Content
IWP	Ice Water Path
MODIS	Moderate Resolution Infrared Spectroradiometer
MSL	Mean Sea Level
TIR	Thermal Infrared

1
Introduction

1.1 Background

Clouds play an essential part in regulating the hydrological cycle and radiative balance of Earth, which in turn affects the climate and weather strongly. Cloud abundances and geographical distributions are suggested to play an even greater role in regulating the climate than greenhouse gases, human-produced aerosols and other factors (NASA, 2005). However, many of the physical processes governing clouds and their feedback on the climate are still poorly understood, especially the effect a warming climate will have on cloud fields (Bony et al., 2015).

Water vapour is one of the main absorbing agents of thermal infrared radiation in the atmosphere. The absorption of radiation leads to an increase in temperature, which can result in a larger amount of water vapour in a feedback loop. Water in solid or liquid form - as is the case in clouds of condensed water - exhibits an absorption more than 1000 times stronger than in its gaseous form. Clouds also play an important role in regulating the Earth's radiation budget - which can be described as the amount of radiation entering and exiting the Earth system - and temperature by reflecting incoming solar radiation. The scattering and absorption by clouds have a more complicated relation to other factors that also regulate the feedback from clouds than water vapour has (Stephens, 2005). Some parameters that affect the radiation budget of Earth include cloud height, vertical profile, liquid and ice water content and particle sizes. Clouds also act as both sources and sinks for water vapour in the atmosphere, thus also regulating the feedback mechanisms of water in its gaseous phase (Stephens, 2005). This means that clouds affect the climate response to changes in greenhouse gases, aerosols or other factors, leading to an increased complexity and uncertainty in many climate models.

The relationship between cloudiness and surface temperature is not entirely straightforward and cloud fields are considered to be governed by large scale motions of the atmosphere, at least to first order, meaning that it is connected to the three dimensional distribution of temperature in a complex way (Stephens, 2005). This presents itself as different levels of cloud cover at different geographical locations as well as altitudes. Since the radiative effects of clouds depend on their altitude, modelling the vertical distribution of cloud layers is important to improve cloud dependent climate models (Leinonen et al., 2019).

Observations of the atmosphere are important for both meteorology and climate science. On a global scale, such observations can only be obtained using remote sensing techniques. Remote sensing refers to the process of collecting information about a system or

part of a system from a distance, without physical contact (Rees, 2012). Remote sensing systems are generally divided into two main categories: passive systems, which detect naturally emitted radiation, and active systems, which emit radiation and analyse the back-propagated signals. All remote sensing techniques have in common that the sensors detect electromagnetic radiation that has either interacted with or been emitted by the target (Rees, 2012).

Many passive systems used for atmospheric observations, such as the Moderate-Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite, are able to collect data over large geographical areas, but are not able to make very detailed measurements. Active systems on the other hand, such as Cloud Profiling Radar (CPR) aboard the CloudSat satellite, are able to make very detailed measurements, but only on a limited geographical area. This means that the data set available from passive sensors has a much larger spatial coverage than for active sensors. Passive sensors are often limited to measurements of the cloud top only, but affected by radiative transfer through the entire cloud structure. The detailed vertical structure information needs to be provided by active sensors. The lack of spatial coverage by active sensors causes a shortage of global three-dimensional data from cloud fields (Leinonen et al., 2019). This in turn leads to a lack of accurate three-dimensional models of global cloud fields necessary for climate modelling and weather predictions.

Various algorithms have been proposed to simulate two and three-dimensional cloud fields from a combination of data from passive and active sensors. However, many of these algorithms are based on self-similarity in the cloud where cloud vertical columns are created from nearby columns with similar radiances (Leinonen et al., 2019). This approach fails to provide realistic two and three-dimensional cloud fields. Another approach involves using neural networks to create more complex spatial structures based on the available data.

The preliminary theoretical base of neural networks was proposed as early as the late nineteenth century independently by Alexander Bain and William James (Bain, 1873; James, 1890). With increased computational power neural networks have been able to produce impressive results. Neural networks can be implemented in various different ways, and to solve a plethora of different problems. A structure for generative networks called generative adversarial networks (GAN) was proposed by Goodfellow (2014). The GAN is structured into two parts, one part generates some distribution and the other part discriminates between the generated distribution and some distribution that is to be duplicated. The generative and the discriminatory part of the network are trained against each other in order to improve performance (Goodfellow et al., 2014). An extension of the GAN is the Conditional generative adversarial network (CGAN), which takes additional input in the generator and the discriminator.

Leinonen et al. (2019) proposed a solution to generating synthetic radar data from the data of the passive system, with the use of a CGAN. Their results are promising with regards to duplicating the nonlinearity that is present in the transformation between the data types (Leinonen et al., 2019). In order to prove the physical feasibility of this approach, more quantitative analysis should be performed on the generated data.

1.2 Aim

The overall aim of the project is to estimate and replicate vertical cloud structures using generative adversarial networks, with random noise or data from passive MODIS satellite imagery as input, using CloudSat radar data as reference. That is, the goal is to obtain vertical 2D cloud fields from the MODIS spectra - consisting of top-of-the atmosphere infrared radiances - to extend the data set available at geographical locations which are poorly covered by current active measurements. The derived 2D cloud fields can be used as input for various applications in weather and climate modelling and can also be used to improve the retrieval of ice cloud properties. The project further aims to perform quantitative analysis of the generated data, proving its physical feasibility or lack thereof.

1.3 Delimitations

It is assumed that the data from the CloudSat and MODIS observations are enough to create a realistic and reliable model of the cloud structures. That is, no further data is required. The method is limited to only include data obtained over ocean locations and land without high altitude topography. The altitudes closest to the surface of the Earth will not be recreated in the simulations due to extensive problems with surface clutter in space-borne radar observations. The dependence on daytime observations, based on reflected radiation from the Sun, is eliminated by only using MODIS channels that operate at wavelengths allowing for nighttime observations as well. The development and movement of clouds over time will not be analysed.

The training of the networks is limited to data from 2015. Scenes are chosen based on a criterion for the lowest amount of clouds present and scenes where the elevation of the Earth's surface is over a threshold are removed. Data from the two satellites needs to by synchronized and correspond to the same location within a small margin of error in position. Measurements over locations from one instrument that does not have a counterpart from the other instrument will not be considered.

1.4 Specification of problem

This project studies the potential of creating a neural network that generates realistic cloud fields from random noise and data from active sensors, and another network that takes input from active sensors along with passive measurements as a basis to cloud fields. The first tool is to be structured as a GAN and the second tool is to be structured as a CGAN. Both tools should be accurate from a physical standpoint, and also compared to other methods for generating artificial cloud fields, such as the surrogate clouds from the IAAFT algorithm created by Venema et al. (2006). The core problem is to generate stochastic two-dimensional cloud fields and to analyse their statistics.

The two data sets focused on in this project originate from MODIS on the Aqua satellite and the Cloud Profiling Radar (CPR) on CloudSat, where MODIS is a passive system that measures radiances at various wavelengths in the thermal infrared region, and the CPR is an active system that measures reflectivities at 94 GHz. Aqua and CloudSat are part of the A-train constellation, meaning they orbit the Earth in the same Sun-synchronous

trajectory, with a minor temporal displacement (Platnick et al., 2003; Oreopoulos et al., 2017).

In extension, the project also intends to obtain inversions from MODIS data and the generated fields to other properties of the cloud fields, such as the ice water path (IWP). In many cases the inversion from MODIS data works well for thin clouds, where the top reflectivities are representative of the properties of the entire vertical structure, but the inversions are not as accurate for thicker clouds. Thus, this project strives to generate accurate inversions also for thicker clouds from the higher dimensional structures generated by the CGAN. Ideally the network can generate cloud structures with high enough accuracy to be used in real life applications such as weather forecasting.

The architecture of the networks will be based on the structure of the networks used by Leinonen et al. (2019). Several pre-written packages in Python will also be used when building the networks.

2

Theory

This chapter introduces important concepts that are relevant to the thesis. Section 2.1 covers the basics of remote sensing, cloud detection as well as important parameters of the Earth's atmosphere. The underlying physics governing cloud fields and their distribution are introduced in Section 2.3. Retrieval theory and geophysical products are introduced in Section 2.4. Then in section 2.5 the structure and fundamentals of neural networks, necessary for understanding the development of the networks used in this thesis are presented. Another method for generating cloud fields is described in Section 2.6 and lastly, spatial autocorrelation is introduced in Section 2.7 as an important measurement of the statistics of cloud fields.

2.1 Remote sensing

Remote sensing refers to the process of collecting information about a system or part of a system from a distance, without physical contact (Rees, 2012). The technique is often applied when studying the Earth's surface and atmosphere using ground-based or satellite mounted instruments that analyse electromagnetic radiation in a wide range of frequencies. Remote sensing is essential for taking measurements of systems that are difficult, dangerous or expensive to reach with in situ techniques, requires no removal of samples and offers quick large scale surveys (Rees, 2012).

Often used in environmental research, remote sensing can measure properties in the atmosphere such as temperature, cloud field distributions, precipitation and wind speeds. It can also be used for monitoring concentrations of certain gases in the atmosphere, such as carbon dioxide, ozone and water vapour (Rees, 2012). Remote sensing techniques can also be applied to measure properties of the surface of the Earth. Over land these properties include topography, temperature, tectonic movement, moisture content in soil, vegetation states and storm damage etc. Over ocean surfaces properties such as temperature, ice fractions, colour, wave and ocean floor topography can be measured (Rees, 2012).

Remote sensing systems are generally divided into two main categories: passive systems, which detect naturally emitted radiation, and active systems, which emit radiation and analyse the back-propagated signals. All remote sensing techniques have in common that the sensors detect electromagnetic radiation that has either interacted with or been emitted by the target (Rees, 2012).

2.1.1 Passive techniques

The important variable in passive systems is how much radiation is detected. For thermal emission, this quantity of emitted radiation at each wavelength is determined by the temperature and emissivity of the target. For other types of radiation the received quantity depends on the incoming amount of radiation on the target and its reflectivity (Rees, 2012).

There exists two main types of passive systems in remote sensing depending on the source of the radiation they detect. The typical sources of radiation for passive systems are the Sun - which emits radiation mostly at ultraviolet, visible and infrared wavelengths - and objects that emit thermal radiation - this includes all bodies which are not at absolute zero (see Section 2.2.1) (Rees, 2012). At terrestrial temperatures the objects generally emit thermal infrared radiation with wavelengths around 10 µm, but emitted radiation can occur as far as in the microwave region (roughly wavelengths of 30 cm - 1 mm, corresponding to frequecies 1 GHz - 300 GHz). Passive techniques that rely on reflected radiation from the Sun can only operate during the day, whereas thermal radiation is emitted both during the day and the night, allowing for night time observations at the associated wavelengths.

The electromagnetic radiation from the Sun must propagate through the atmosphere before reaching the surface of the Earth. Along this propagation much of the energy is absorbed or scattered and only a portion reaches the Earth's surface.

2.1.2 Active techniques

Active systems in remote sensing use an inner source of electromagnetic radiation which is sent out and the response measured. Because of this, active systems can in theory use radiation at any wavelength, but in practice the technique is limited by the transparency and properties of the Earth's atmosphere (Rees, 2012). Furthermore, because active techniques provide their own source of electromagnetic radiation they are not limited to operation during a certain time of the day or year.

The radiation that is measured by remote sensing systems usually has to propagate at least one time through part of the Earth's atmosphere, twice if the detected radiation needs to be reflected, such as in active techniques. Each journey causes the radiation to be attenuated.

Active techniques depend on two important variables describing the radiation: the quantity of detected radiation and the arrival time of the signal. The second parameter is necessary for determining the distance from the emitting instrument to the target.

2.2 Earth's atmosphere

The Earth's atmosphere is composed of nitrogen (78%), oxygen (21%), argon (0.9%) and trace gases such as carbon dioxide, nitrous oxides, methane and ozone. In addition to this, water vapour also exists in the atmosphere with concentrations in the range $0 - 4\%$ depending on the location and time (NCCO, 2020). The concentrations of water vapour

are highest in the warm, humid, tropical regions (near the equator) and lowest in the cold, dry arctic regions (NCCO, 2020).

Gases and particles in the atmosphere absorb or reflect radiation at different wavelengths, affecting the transmissivity of the atmosphere, as can be seen in figure 2.1. Where the absorption is high, very little incoming radiation is transmitted and the path length through the absorbing media is very important. There exists certain transmission windows in the atmosphere where electromagnetic radiation with the specific wavelengths can pass through the atmosphere more easily. These windows thus play an important role in choosing the relevant wavelengths for observations of certain properties. For the purpose of observing targets other than the atmosphere, the instruments (mostly active systems) should preferably operate in one such window. Sounding instruments are typically used for measuring properties of the atmosphere, such as its temperature or composition, and operate at the edge of these windows or where the atmosphere is opaque (Rees, 2012). There are three main atmospheric windows, the first is in the visible and near-infrared (VNIR) region of the electromagnetic spectrum around wavelengths 0.3 µm - 3 µm. The second narrow window corresponds to the thermal infrared (TIR) part of the spectrum, around wavelengths 8 µm - 15 µm. The third window occurs in the microwave region with wavelengths from a few millimeters to a few meters (Rees, 2012).

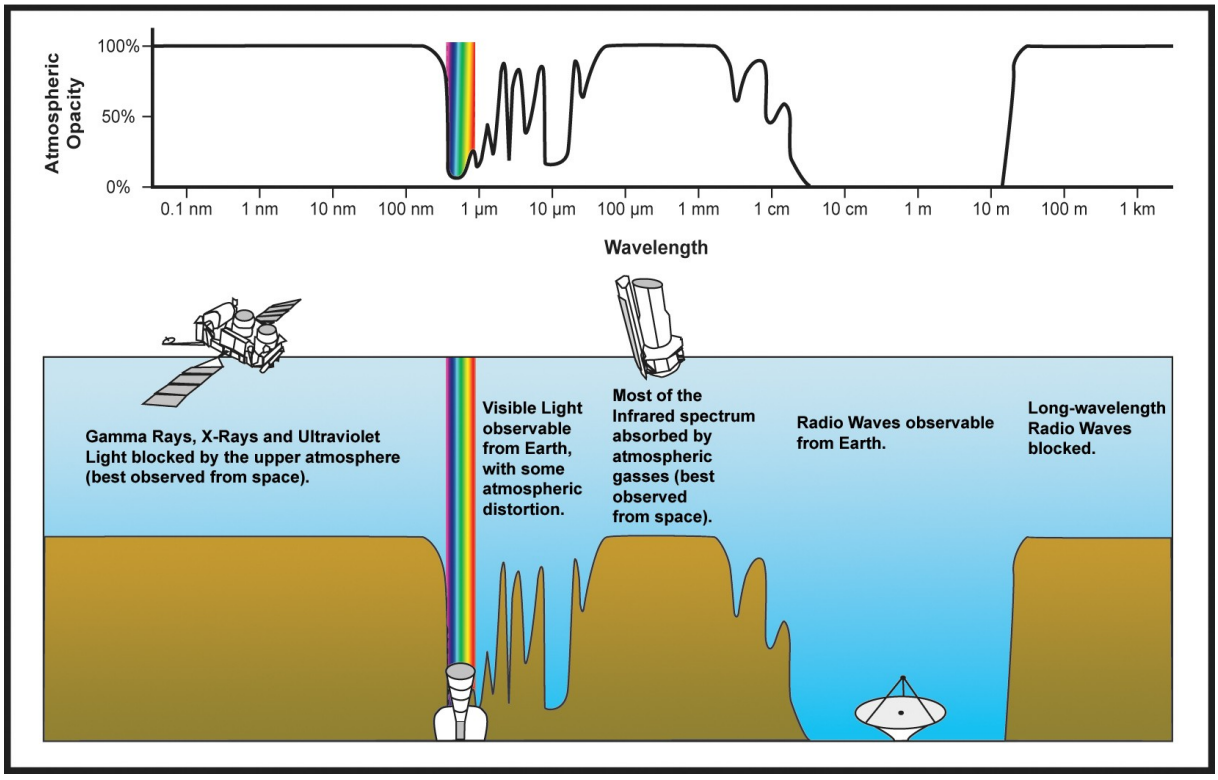


Figure 2.1: Atmospheric absorption in percentages at different wavelengths of the electromagnetic spectrum. Water vapour, carbon dioxide and ozone are responsible for the majority of the absorption. There are "windows" in the atmosphere where the absorption is very low. One such window occurs in connection with the visible region of the electromagnetic spectrum. Image credit: NASA

Both the temperature, pressure and the composition of the atmosphere vary with height. Pressure and density both decrease exponentially with height above the Earth's surface. The pressure at height h is given by

$$P(h) = P_0 \cdot \mathrm{e}^{-h/h_s}, \tag{2.1}$$

where P_0 is the pressure at mean sea level and h_s is the scale height ($h_s \approx 7.5$ km). The temperature dependence on height follows a more complex pattern. For this purpose the atmosphere is often divided into four layers: the troposphere, stratosphere, mesosphere and thermosphere. Each layer has a distinct border at the top where the properties of the atmosphere changes, as can be seen for the temperature in Figure 2.2. These transition borders between layers are known as the tropopause, stratopause, mesopause and thermopause, respectively (Liou, 2002).

The temperature in the troposphere decreases with altitude, from about 288 K at mean sea level to about 220 K at the tropopause with a lapse rate of $6.5\,\mathrm{K\,km^{-1}}$ (Liou, 2002). This means that the temperature-altitude relation can be modelled with the following equation:

$$T(h) = T_0 - 6.5h, \text{ [K]} \tag{2.2}$$

if the altitude h is given in kilometers and $T_0 = 288.15$ K. This temperature distribution is caused by the radiative balance and convective transport of energy from the Earth's surface to the atmosphere. This happens because most of the energy that heats the atmosphere comes from the surface and since the pressure decreases with increasing altitude, the air that rises into regions of lower pressure will tend to expand, causing the temperature to decrease. This is also known as adiabatic cooling. The absolute altitude of the tropopause depends on both time and location, as can be seen in figure 2.3. On average, the tropopause is located in the range of 8 to 18 km over mean sea level with higher altitudes over the warm, humid tropics and lower altitudes over cold, dry arctic regions.

Almost all water vapour, clouds and precipitation exist in the troposphere, with a very small amount of clouds also occuring in the lower stratosphere (*nacreous* clouds) or mesosphere (*noctilucent* clouds). The stratosphere consists of an isothermal layer from the tropopause up to around 35 km. After this layer the temperature increases with height. The stratosphere contains most of the ozone in the atmosphere, with the majority in a layer around 15 to 30 km. This is known as the ozone layer. Ozone reaches maximum concentrations at approximately 20 to 25 km. Apart from ozone, the stratosphere also contains thin layers of long-lived aerosols. (Liou, 2002)

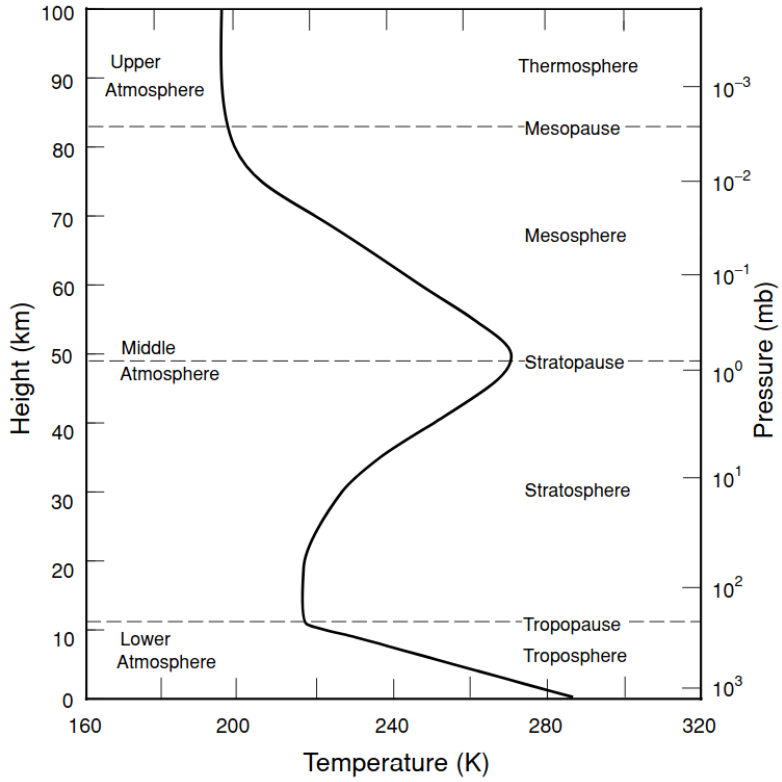


Figure 2.2: Vertical temperature profile for the standard atmosphere. This profile represents typical conditions in mid-latitude regions. Image credit: K. N. Liou (Liou, 2002)

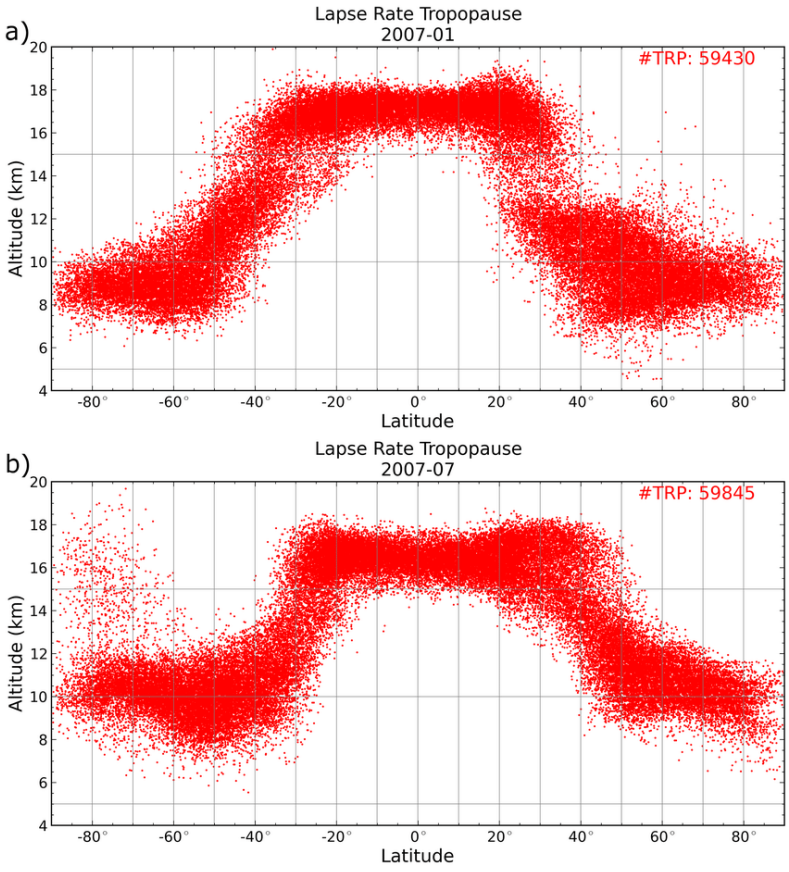


Figure 2.3: Altitude of tropopause for different latitudes during (a) winter and (b) summer in the Northern Hemisphere. The altitude of the tropopause varies during the different seasons, with slightly higher altitudes during the Northern Hemisphere winter. The altitude also varies with latitude, with the highest values near the tropics. Image credit: Biondi et al. (2015)

2.2.1 Thermal radiation

All objects with temperature above absolute zero emit thermal radiation in a continuous spectrum. A black body is an ideal body which absorbs all incident electromagnetic radiation at all frequencies and thus also acts as a perfect emitter (emissivity $\varepsilon = 1$). In thermal equilibrium the spectral radiance L_λ of a black body is given by **Planck's law**:

$$L_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}, \text{ [W sr}^{-1} \text{ m}^{-3}] \tag{2.3}$$

where λ is the wavelength of emitted radiation, h is the Planck constant, c is the speed of light, k_B is the Boltzmann constant and T is the absolute temperature of the object.

For real materials which are not perfect black bodies the emissivity, ε, is in general not equal to one, so the actual radiance needs to be adjusted by this value. The emissivity often depends on the wavelength, so the spectral radiance of a real material will be given by

$$L_\lambda = \varepsilon(\lambda) L_{\lambda,P}, \tag{2.4}$$

where $L_{\lambda,P}$ stands for the black body spectral radiance given by Equation (2.3). Following this, according to Kirchoff's law of radiation, materials acting as good emitters of radiation (high emissivity) will also act as good absorbers.

An important property that can be derived from the spectral radiance of an object is the *brightness temperature*, T_b. This value represents the temperature a black body would need to produce the same radiance at each wavelength. So for an object at temperature T with emissivity ε the brightness temperature T_b would be given by the following expression:

$$T_b = \frac{hc}{\lambda k_B \ln\left(1 + \frac{\exp(hc/\lambda k_B T) - 1}{\varepsilon}\right)}. \tag{2.5}$$

Substituting the spectral radiance in Equation (2.3) into Equation (2.5), the brightness temperature for a target with known spectral radiance L_λ is then given by

$$T_b = \frac{hc}{\lambda k_B \ln\left(1 + \frac{2hc^2}{\varepsilon \lambda^5 L_\lambda}\right)}. \tag{2.6}$$

In the thermal infrared region (wavelengths from 3 to 15 µm) and in the microwave region the emissivity, ε, of a body at any given wavelength is related to the reflectivity r by $\varepsilon = 1 - r$. Most materials have refractive indices close to 1 in this region, meaning that the reflectivity is low and the emissivity is high as a consequence (Rees, 2012).

In addition to the spectral radiance distribution given in Equation (2.3), the wavelength of the peak, λ_{peak}, of the black body radiation curve for a certain temperature T is given by **Wien's displacement law**:

$$\lambda_{\text{peak}} T = 2.898 \times 10^{-3}, \text{ [m K]} \tag{2.7}$$

This means that bodies at typical terrestrial temperatures have their peak emission around 10 µm.

The Sun can be modelled to high accuracy by a black body with effective temperature $T \sim 5800$ K, located at a distance of 1 AU (astronomical unit) from Earth. The spectral irradiance of the Sun, as seen in Figure 2.4, thus follows a typical black body curve, as given by Planck's law in Equation (2.3) modified by the emissivity (Rees, 2012). The spectral irradiance peaks in the visible region near 500 nm but has contributions from wavelengths around 300 nm up to about 3 µm. Multiple absorption bands (primarily from water vapour, carbon dioxide, oxygen and ozone, see also Figure 2.1) which drastically reduce the amount of radiation reaching the Earth's surface can bee seen at various wavelengths in Figure 2.4.

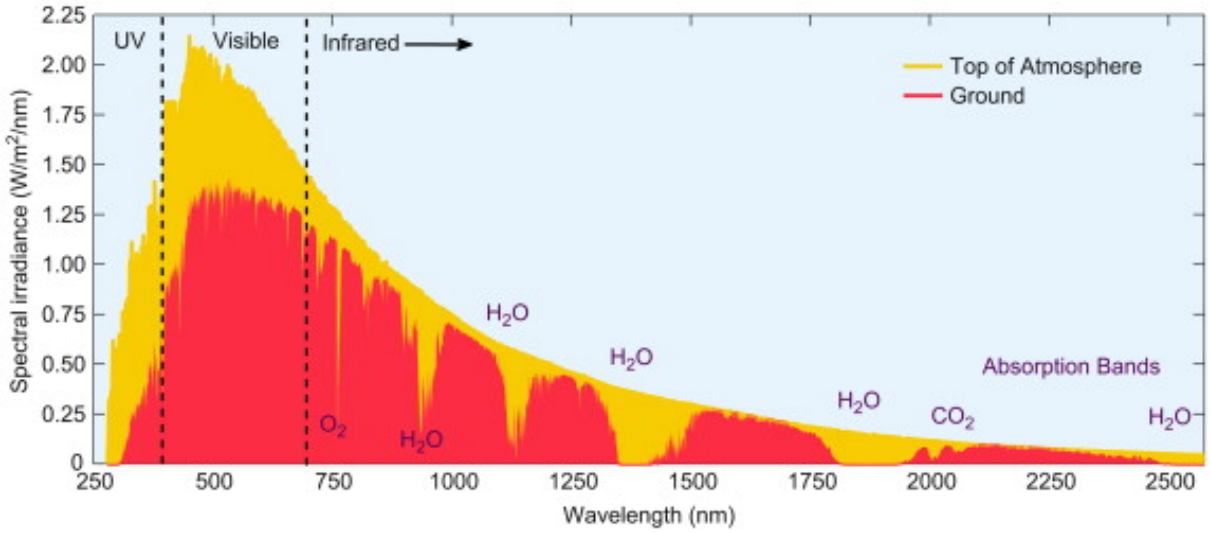


Figure 2.4: Solar spectrum as a function of wavelength. The spectrum follows a typical black body radiation curve with its peak near 500 nm. Multiple absorption bands from gases in the atmosphere lead to a decrease in the amount of radiation reaching the Earth's surface. Image credit: United States Department of Energy, National Renewable Energy Laboratory

2.3 Cloud fields

Clouds play an important part in regulating the amount of radiation that gets transmitted through the atmosphere and are thus essential for climate models. Clouds absorb and scatter thermal infrared radiation that has been emitted by the Earth's surface and underlying atmosphere. This means that a significant amount of thermal energy is trapped in the atmosphere resulting in a warming effect commonly known as the *greenhouse effect* (Jiang et al., 2017). However, clouds also reflect incoming solar radiation away from Earth, which has a cooling effect on the atmosphere. The net radiative effect is dependent on cloud parameters such as water content and ice or liquid particle size, which in turn is related to temperature, fall speeds and precipitation (Jiang et al., 2017).

Clouds are collections of water and ice particles which are large enough to be visible to the human eye. They form from water vapor that has entered the atmosphere through evaporation from the Earth's surface. The presence of aerosols is the second criterion necessary for clouds to form, since they act as cloud condensation nuclei in an unstable atmosphere (NASA, 2005).

Approximately 70% of the surface of the Earth is covered in clouds at any given time. About 33% is covered by high altitude clouds (over 6.5 km), 26% by middle altitude clouds (3-6.5 km) and 49% by low altitude clouds (below 3 km) (Wylie et al., 2005). Note that several types of clouds can occur in the same location, e.g. high altitude clouds over low altitude clouds. Clouds are not homogeneously distributed over the Earth's surface but rather tend to accumulate in distinct regions, as can be seen in Figure 2.5. Clouds are most common in the intertropical convergence zone (ITCZ) near the equator, at the ascending ends of the Hadley cells. Clouds are also frequently found in the mid-latitude storm belts over the North Atlantic, North Pacific and Antarctic Oceans (Wylie et al., 2005). In the subtropical zones (roughly 25° to 40° North and South latitudes) cloud coverage is much lower both over the oceans and over land (Wylie et al., 2005).

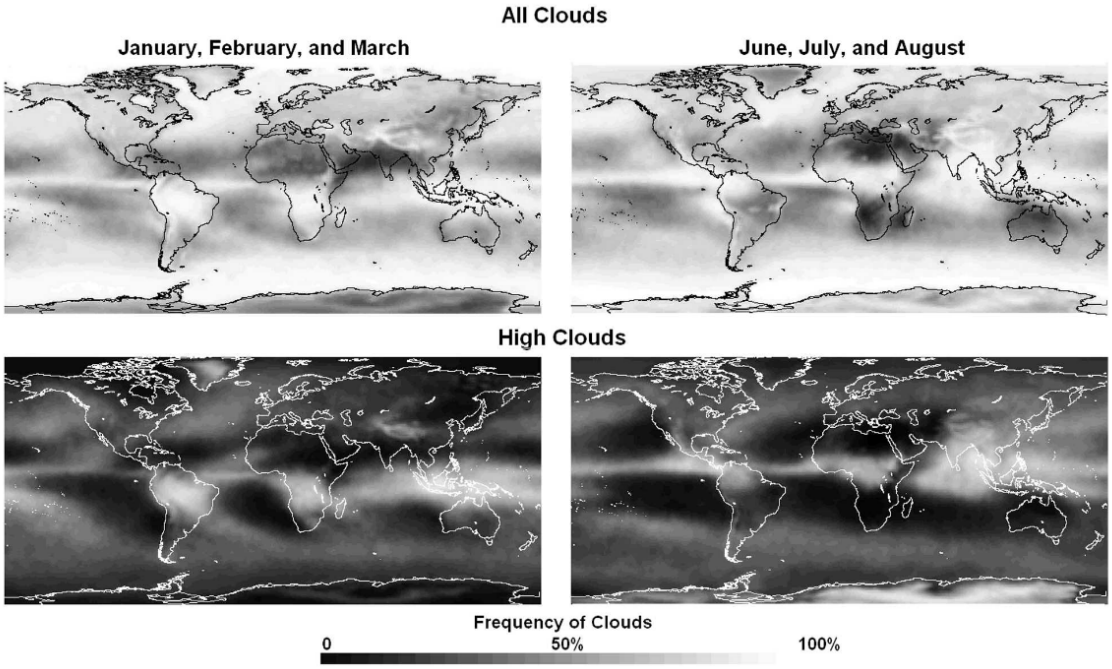


Figure 2.5: Global mean cloud cover. The top panels show frequencies of clouds for all types of clouds. The lower panel shows high clouds above 6.5 km. The data is collected with the High Resolution Infrared Radiometer Sounder (HIRS) through the years 1979 to 2001 and divided into winter (left) and summer (right). Image credit: Wylie et al. (2005)

Low altitude clouds consist of liquid water droplets with radii of 10 to 50 µm (Rees, 2012). Higher altitude clouds, on the other hand, typically consist of ice particles with mass mean effective diameters in the range of 50 to 1000 µm (Jiang et al., 2017). Because the refractive index, n, for liquid water is typically higher than for ice the absorption length (inversely proportional to the imaginary part of the complex refractive index) will be smaller in the case of liquid water, effectively meaning that liquid water clouds absorb more strongly than ice water clouds (Rees, 2012).

Defining the size parameter of a spherical particle as $\chi = 2\pi a/\lambda$, where a is the typical radius of the particle and λ is the wavelength of incident radiation, several types of scattering interactions with the particle can be identified. The first case occurs when the particle is much smaller than the wavelength ($\chi \ll 1$). Then the incoming radiation will

be in phase across the particle. This case is known as *Rayleigh scattering*, and is a result of the polarizability of the particle that acts as a dipole. In this limit the particle absorbs much less power than the power that would be contained in a cross-section of radiation that is equal to the particle's geometrical area. The absorption cross-section is proportional to χ. The scattering cross-section for Rayleigh scattering is highly dependent on the wavelength and scales as χ^4 ($\propto \lambda^{-4}$) meaning that shorter wavelengths experience much stronger scattering. The case where $\chi \approx 1$ is known as *Mie scattering*. The third case occurs when $\chi \gg 1$, commonly known as *geometric scattering* because in this limit the scattering and absorption cross-sections are proportional to the geometrical cross section of the particle. (Rees, 2012)

In the visible and thermal infrared region the attenuation of the atmosphere is mostly dominated by Rayleigh scattering from gas particles whereas atmospheric clouds are subject to Mie scattering or geometric scattering, as can be derived from the size parameter for the given particle sizes in clouds (Rees, 2012). In these regions scattering effects are dominant over absorption and the optical thickness is generally very high, up to 100, making the clouds opaque (Rees, 2012). In the visible region cloud optical depths are rather independent of wavelength, but increase with wavelength in the infrared. As a result of the high opacity of clouds in these regions of the electromagnetic spectrum, thermal radiation emitted from the Earth's surface is strongly attenuated by the clouds and does not interfere with satellite observations of clouds. On the other hand, satellite based remote sensing systems will mostly only be able to observe the top layer of clouds at these wavelengths. Because of this the measured brightness temperature of a cloud is approximately equal to the *cloud-top temperature* which in turn is similar to the temperature in the atmosphere at the altitude of the cloud-top (Rees, 2012). This means that higher altitude clouds will have lower brightness temperatures than low altitude clouds and the background surface emission. Channels operating at shorter wavelengths in the thermal infrared region (up to about 3 to 4 µm) will detect not only thermal emission from the cloud, but also reflected solar radiation during the daytime, which increases the detected brightness temperature.

In the microwave region the size parameter for particles in clouds approaches the Rayleigh limit and absorption is much stronger than scattering, and although strongly dependent on the wavelength, the attenuation of microwave radiation is often low enough even for thick layers of dense clouds to be fairly transparent (Rees, 2012).

2.4 Retrievals and geophysical products

The data interpretation in remote sensing is built on a set of models: a *forward model* that simulates observables based on a given set of model parameters that represent reality, and a *retrieval method* (also known as *inversion*) that aims to generate some geophysical product that estimates the state vector of the observed reality from the observations. The *forward model* is divided into a forward problem, which is supposed to produce a quantitative description of the physical processes that occur in the object being studied and an *instrument model* that evaluates the observables generated by the forward problem and provides a description for the measurement procedure. The forward model can always be expected to be subject to some measurement errors, leading to an uncertainty in the observation (Ustinov, 2014). The retrieval method takes measured observables as input

and outputs retrieved model parameters. Along with these the retrieval error for each parameter, related to the uncertainty in the measurement, should be provided (Ustinov, 2014).

Geophysical retrievals are indirect measurements, meaning that they are not directly related to a measured quantity but are dependent on the above mentioned models. In the case of remote sensing of the atmosphere the interesting parameters are measured in terms of radiances at the top of the atmospheric system (for passive systems) or reflectivities (active systems). The geophysical product of interest - such as temperature, pressure, optical thickness - can be determined by tracing back through the observation system via the different models, creating an inversion for the sought parameters. In atmospheric sounding the forward problem is comprised of equations for radiative transfer. The instrument model is dependent on the angular and spectral distributions of the instrument response. Finally, the inverse model is applied in order to get an estimate of the atmospheric parameters based on the observables and sensitivities. Sometimes a priori information about the state vector is necessary to solve the inversion problem. This is especially the case when the state vector contains some continuous parameters that are functions of space coordinates or time. (Ustinov, 2014)

2.4.1 Ice water path

The ice water path (IWP) of a cloud is defined as the integral of the ice water content (IWC) in a vertical column of the cloud layer. It is an important quantity for climate studies as it determines the interaction between the cloud and incoming and outgoing radiation. This is due to its close relationship with the cloud absorption and optical depth and importance for determining albedo and emissivity (Heymsfield et al., 2003). Current climate models show difficulties in determining precise and accurate IWPs for cloud fields, with values differing up to an order of magnitude (Protat et al., 2016). This means that the difference between derived IWPs for the same conditions is on par with the magnitude of the IWP. The retrieval of IWC in nonprecipitating clouds provides reasonably accurate results in climate models, but convective cloud systems (cumulus clouds) prove to be more challenging (Protat et al., 2016).

Several methods exist for determining IWP from remote sensing measurements. The choice of method depends on the type of sensor that is performing the observations, the measured parameters and the access to additional information about cloud microphysical properties. Common methods for passive techniques consist of measuring or estimating ice particle size distributions based on radiances at two different infrared wavelengths (at least for thin cirrus clouds) and relating this to the IWP (Heymsfield et al., 2003).

The IWP can also be obtained by empirical relationships between IWP (or IWC) and more easily observed quantities, such as radar reflectivities. Authors of several articles have found such relationships with slight variations, especially for different latitude regions and cloud types. Large measurement campaigns of the IWC in a variety of ice cloud types - including thin cirrus and stratiform ice clouds - in the midlatitudes and tropical regions using a 95 GHz radar resulted in the following relationship for radar reflectivities (Z_{95} [dBZ]) and IWC (Protat et al., 2016):

$$\mathrm{IWC} = 0.149 Z_{m95}^{0.681}, \tag{2.8}$$

where Z_{m95} is the radar reflectivity transformed from dBZ to linear units ($\mathrm{mm^6 m^{-3}}$):

$$Z_{m95} = 10^{Z_{95}/10}. \tag{2.9}$$

The relationship is based on an assumption of spherical particles with a given statistical relationship between ice crystal mass and diameter ($m(D)$ relationship), and the IWC is estimated from measured particle size distributions (PSDs) (Protat et al., 2016). This power-law relationship fails in accuracy and produces large errors for high radar reflectivities due to non-Rayleigh scattering effects in the cloud layer (Protat et al., 2016).

2.4.2 Cloud-top height

The cloud-top height (CTH) and cloud-top temperature (CTT) - which is the altitude of the top layer of the cloud field and the temperature at this altitude - are important parameters of a cloud field for determining its radiative effects. High-altitude ice clouds are especially important in modulating Earth's weather and climate since they regulate the hydrological cycle and energy balance in the atmosphere (Jiang et al., 2017). They have also been shown to be the most difficult cloud types for estimating parameters such as IWC (Jiang et al., 2017). This is due to a lack of understanding of the physical processes that govern ice clouds. The CTH can either be provided by meteorological satellites or estimated from measurements of other parameters.

The CTH can be estimated directly from active measurements or by using measurements of radiances by passive sensors, from which brightness temperatures in the infrared region can be calculated but the data needs to be supplemented by temperature vertical profiles. This model for passive techniques is based on radiative transfer methods, which relate the cloud temperature to the emitted radiance following Planck's law, and in extension relates this to the measured radiance. The vertical profiles of temperatures are necessary for finding the corresponding height. Some issues arise in determining the relationship between emitted and measured radiance since the atmosphere interacts with the radiation and emission from the Earth's surface may have an influence in the case of thin clouds. In order to get an estimate of the CTH using passive techniques and radiative transfer, knowledge about the atmospheric transmittance and the cloud emissivity at the relevant wavelength is thus necessary.

In general, clouds at lower altitudes emit a larger amount of radiation than clouds at higher altitudes. On average, low-altitude clouds also have a greater optical thickness than high-altitude clouds (NASA, 2005). Since the temperature of low-altitude clouds is similar to that of the underlying surface, this type of clouds has little impact on the amount of thermal infrared radiation escaping the Earth system. Low altitude clouds will absorb and re-emit radiation with similar properties to that which is emitted by the Earth's surface. Due to scattering of incoming solar radiation, low-altitude clouds have a net cooling effect on the atmosphere. High-altitude clouds, on the other hand, are optically thin and reflect a smaller part of the solar radiation. They do however absorb a significant amount of outgoing thermal radiation and have a net warming effect (NASA, 2005).

2.5 Neural Networks

Neural networks is a methodology that aims to recognize underlying relationships in a set of data. The structure of neural networks consists of a collection of units based on models of neurons. The neurons are sorted into layers, where information is passed and modified by different types of mathematical operations. The network is trained, meaning the parameters of the network are updated according to how the network performs with regards to an arbitrary loss-function.

2.5.1 Convolutional neural network

A special kind of neural network is the Convolutional Neural Network (CNN). The CNN processes data that has a known grid-like topology. There are several examples of data with such structures, such as time-series, and perhaps most well known image data, which can be thought of as a 2-D grid of pixels. Convolutional networks have been successful in several practical applications, such as in the field of image analysis. A convolutional network is defined as having at least one layer that uses the operation called Convolution. (Goodfellow et al., 2016)

The convolution operation is defined on a continuous interval by Equation 2.10.

$$s(t) = \int x(a)w(t-a)da, \tag{2.10}$$

where $w(a)$ is a weighting function, i.e. a valid probability density function.

When using computers the data is always discrete, and the convolution on a discrete interval is given by Equation 2.11.

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a). \tag{2.11}$$

In machine learning applications, the most common input is a multidimensional array of data, and the kernel is usually a multidimensional array of parameters adapted by the learning algorithm.

Convolutions are often used over more than one dimension at a time. For example, if a two-dimensional image I is used as input, a two-dimensional kernel K is often used. The convolution over two axis is given by Equation 2.12. (Goodfellow et al., 2016)

$$S(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \tag{2.12}$$

2.5.2 Gradient Descent

In order to optimize the neural network with regards to the cost function, which is an arbitrary function based on the desired output of the system, the weight and biases of the neural network need to be updated. An effective method for updating the neural network is Gradient Descent, which is based on the analytical fact that local minimas of a function is found where the value of the gradient is 0. The gradient descent updating method is defined by Equation 2.13.

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \epsilon \nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \tag{2.13}$$

Where $\boldsymbol{\theta}'$ is the updated parameter, $\boldsymbol{\theta}$ is the parameter before being updated, ϵ is the learning rate and f is the cost function. This expression will converge if $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \rightarrow 0$, which is a desired behaviour. There are multiple ways of setting the value for the learning rate, ϵ, with different benefits. Generally the learning rate is set to a small positive value, to make sure that the method does not update the parameter too much, and by doing so missing a minima. A small learning rate will make the network focus on local optimization (exploitation) and a larger learning rate will make the network focus on finding a more general area in the parameter space (exploration). (Goodfellow et al., 2016)

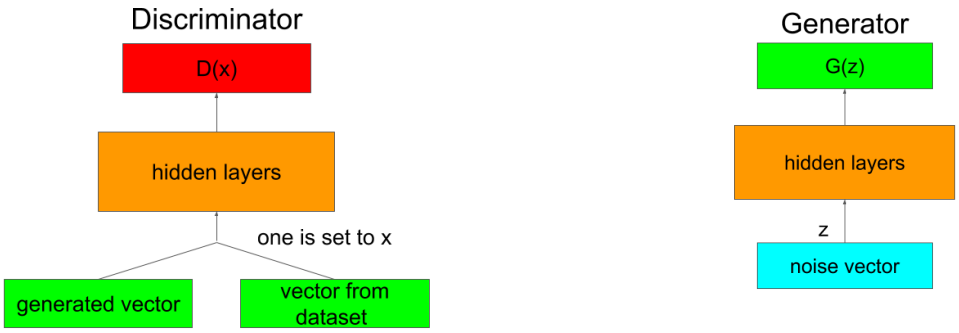
2.5.3 Generative adversarial network

Generative adversarial networks, or GAN for short, is a framework for estimating generative models via an adversarial process. The adversarial process means that two models are trained simultaneously: a generative model G designed to capture the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than the generator. The goal of the training for G is to maximize the probability for D to make a mistake. In the space of arbitrary functions G and D, a unique solution exists where G recovers the distribution of the training data and D is equal to 1/2 everywhere. This means that the discriminator considers the generated data to be true with the same probability as for the training data.(Goodfellow et al., 2014)

Assuming both the generator and discriminator are multilayer perceptrons, applying the adversarial modeling framework is straightforward. An input noise variable $p_z(z)$ needs to be defined in order to learn the distribution p_g of the generator over the data. This is mapped to data space as $G(z;\theta_g)$, where G is a differentiable function represented by a multilayer perceptron with parameters θ_g. Additionally, a second multilayer perceptron is defined, $D(x;\theta_d)$, which outputs a single scalar. $D(x)$ represents the probability that x came from the data rather than p_g. The discriminator, D, is trained to maximize the probability of assigning the correct label to both the data and samples from G. Simultaneously G is trained to minimize $\log(1 - D(G(z)))$. In other words, D and G play the following minimax game with value function $V(G, D)$ (Goodfellow et al., 2014):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_{data}(z)}[\log(1 - D(z))]. \tag{2.14}$$

The structure of GAN is visualized in Figure 2.6.



(a) Structure of the discriminator. **(b)** Structure of the generator.

Figure 2.6: Structures of the two parts of the generative adversarial network. The flow of information starts at the bottom of the structure.

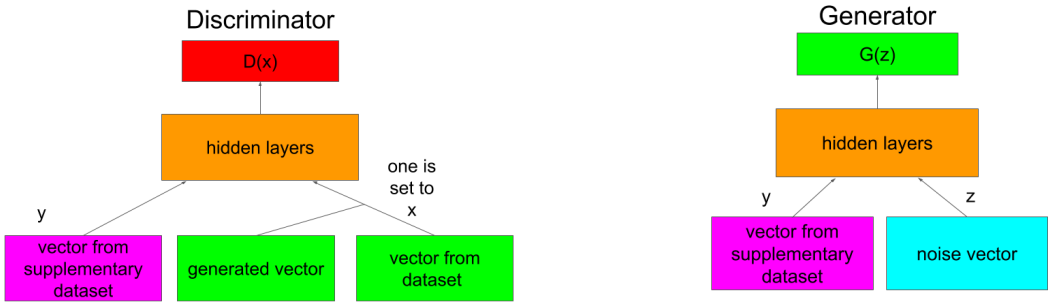
2.5.3.1 Conditional generative adversarial network

The generative adversarial network is extended by adding additional input y to both the generator and the discriminator. This extension is called conditional generative adversarial network (CGAN). The additional input, y, could be any form of auxiliary information.(Mirza and Osindero, 2014)

In the generator the input noise defined in previous section $p_z(z)$ and y are combined in joint hidden representation, with the rest of the structure being flexible in how its hidden representation is composed. In the discriminator, x and y are presented as input to a discriminative function.

The objective function of a two-player minimax game is a variation of the one proposed in Equation 2.14, where the new objective function is given in Equation 2.15 (Mirza and Osindero, 2014):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x|y)] + \mathbb{E}_{z\sim p_{data}(z)}[\log(1 - D(z|y))] \tag{2.15}$$



(a) Structure of the discriminator. **(b)** Structure of the generator.

Figure 2.7: Structures of the two parts of the conditional generative adversarial network. The flow of information starts at the bottom of the structure.

2.6 Iterative amplitude adjusted Fourier transform

The iterative amplitude adjusted Fourier transform algorithm (IAAFT) is an algorithm that can be used to generate randomized cloud fields. It is based on the assumption that time series are entirely determined by their power spectrum - or equivalently by their autocorrelation function. This holds in the case of stationary linear time series. Cloud fields, however, are known to be governed by nonstationary and nonlinear processes, so power spectrums and amplitude distributions are not enough to cover all statistics of cloud fields, although they are sufficient for describing cloud water structures responsible for radiative cloud properties. (Venema et al., 2006)

The IAAFT algorithm was developed to test nonlinearity in a system. The test aims to distinguish between a time series originating from linear and nonlinear dynamical systems. The nonlinearity of interest is the dynamical nonlinearity, therefore the null-hypothesis of this statistical test is that the time series should originate from a linear dynamical process

which is then altered by a nonlinear static filter (Venema et al., 2006).

The measured time series is denoted by $\{m_n\}$, where n is an index representing the time step. The algorithm is initiated by a random shuffle of the data points in $\{m_n\}$ in order to create a new time series that acts as a basis for the surrogate cloud field to be generated. In each iteration i, the Fourier spectrum is first adjusted and then the amplitudes. The power spectrum, M_k, can then be calculated with Equation 2.16 based on the original time series it is modelled from.

$$M_k^2 = |\sum_{n=0}^{N-1} m_n e^{i2\pi kn/N}|^2. \tag{2.16}$$

The Fourier transform, denoted S'_k, is calculated for the initial or iterated time series, $\{x_{2,i-1}\}$. The magnitudes of the Fourier coefficients are replaced by the coefficients of the original time series, the phases $\phi_k = S'_k/|S'_k|$ are left unchanged. This produces a time series with the desired power spectrum $\{x_{1,i}\}$. The complex Fourier coefficients of $\{x_i\}$ are given by Equation 2.17.

$$S_k = M_k \phi_k. \tag{2.17}$$

In a next step the altitudes of the iterate Fourier spectrum are adjusted based on their relative magnitudes. This means that the value of the highest altitude in the iterate time series is replaced by the highest value in the measured time series. This is done by sorting the values of the measured time series $\{m'_n\}$ and the iterate time series $\{x'_1\}$ into lists. The values of $\{m_n\}$ are then put at the index of the $\{x_1\}$ values with the same ranking in order to create a new time series. The function rank() returns the rank number in ascending order, starting at 1 for the highest amplitude. The new time series with adjusted amplitudes is then given by Equation 2.18:

$$x_2 = m'_{rank(x_1)}. \tag{2.18}$$

This amplitude adaption of the power spectrum and the spectral adaptation are repeated until a fixed convergence threshold is reached and the generated surrogate time series is returned. (Venema et al., 2006)

2.7 Spatial autocorrelation

Spatial autocorrelation is a measurement of how well a certain variable correlates to itself in space. It can be defined as the similarity of numerical values for neighbouring pixels in an image, where higher similarity suggests a higher spatial autocorrelation. If spatial autocorrelation is present this means that neighbouring locations share information and that the entire dataset contains an amount of information that to some degree could be considered redundant (Lee, 2017).

The autocorrelation function, $p_X(h)$ of $\{X_t\}$ at lag h, is defined as:

$$p_{i,j}(h) \equiv \frac{\gamma_X(h)}{\gamma_X(0)}, \tag{2.19}$$

where $\gamma_X(h) = Cov(X_{t+h}, X_t)$. The lag is often in time, but can also be defined as a spatial displacement. (Brockwell et al., 2002)

There exists several additional measurements of spatial autocorrelations, such as Pearson's r (mostly used for bivariate dependence) and Moran's I. The later uses an index I, which compares local and global coherence in an image or a data set according to the following equation (Schmal et al., 2017):

$$I = \frac{1}{\sum_{ij} w_{ij}} \frac{\sum_{ij} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{N^{-1} \sum_i (x_i - \bar{x})^2}, \tag{2.20}$$

where N is the number of spatial units that are indexed by i and j, x is the variable being measured and $\bar{x}$ is the average of x over all spatial units. The spatial weight matrix w_{ij} holds information about which pairs of locations are relevant for considering the correlation. This matrix is often chosen such that its diagonal elements are 0, and $w_{ij} = 1$ only for closest neighbours (and $w_{ij} = 0$ for all other combinations). It is also possible to expand the correlation search to include second closest or diagonal neighbours, setting $w_{ij} \neq 0$ for these combinations (Schmal et al., 2017). Moran's I takes values between -1 and 1, where -1 represents perfect dispersion (neighbouring pixels have maximal difference in value of the variable of interest, also called *negative spatial autocorrelation*), values of I close to 0 represent random distributions, and values close to 1 correspond to perfect (positive) spatial autocorrelation, which points to clustering of similar values.

3
Data and Methods

This chapter presents the methodology applied in order to meet the aim of the project. The project can be divided into three main parts. The first consists of selecting and processing the data from two principal instruments: The Cloud Profiling Radar (CPR) aboard CloudSat, and MODIS. Two data sets were created, the first intended for the GAN (containing only data from the CPR) and the second intended for the CGAN (including data from MODIS, CPR as well as external temperature data). Description of the data and creation of the data sets are presented in Sections 3.1 through 3.4. The second part involves developing and training the networks, the approach to which is described in Sections 3.5 and 3.6. The networks are also compared to another method of generating synthetic cloud fields, IAAFT. This method is presented briefly in Section 3.7. The final part amounts to performing a set of statistical analyses on the output from the networks as well as their performance and retrieving geophysical products from the output, all of which is described in Section 3.8. The Python code implemented in this project is available at `https://github.com/SEE-MOF/Generation_of_atmospheric_cloud_fields_using_GANs`.

3.1 CloudSat data

The CloudSat satellite carries a nadir-looking 94-GHz (millimeter-wavelength) CPR which measures the backscattered power from clouds as a function of the distance from the detector. CloudSat is a part of the *A-Train* satellite constellation at a 705 km Sun-synchronous near-polar orbit.

The CPR has a footprint of 2.5 km along track and 1.4 km across-track. Each CloudSat profile consists of 125 vertical bins, where each bin is roughly 240 m in depth, in a 1.1 km along track interval as seen in Figure 3.1 (CSU, 2017).

The primary data used is the radar reflectivity contained in the 2B-GEOPROF data product available at the ICARE Data and Services Center (ICARE, 2020). This product identifies levels in the vertical columns measured by CloudSat that show a significant radar echo - as opposed to noise or radar clutter - and estimates a radar reflectivity factor for each bin. The radar reflectivity is given in logarithmic dBZ units in the range -40dBZ to 50dBZ, although the reliable detection range is -30dBZ to 25dBZ. High sensitivity is needed since clouds present a much weaker scattering of microwave radiation than the underlying surface of the Earth. The product also includes several cloud masks and other geophysical products. Important data fields are presented in Table 3.1.

The satellite experienced a battery anomaly in April 2011, leaving the payload limited to "Daylight Only Operations" (DO-OP). Since 2011 the instrumentation on CloudSat has

not been able to operate in eclipse, when the satellite is in the shadow of the Earth and the corresponding geographical location experiences night (CSU, 2020). In extension, this means that the CPR only collected data during roughly half of each orbit. As a result, data is also missing from latitudes near the poles in many orbits. CloudSat remained in orbit (or in a lower, nearby orbit) with the A-train constellation from April 2006 to February 2018.

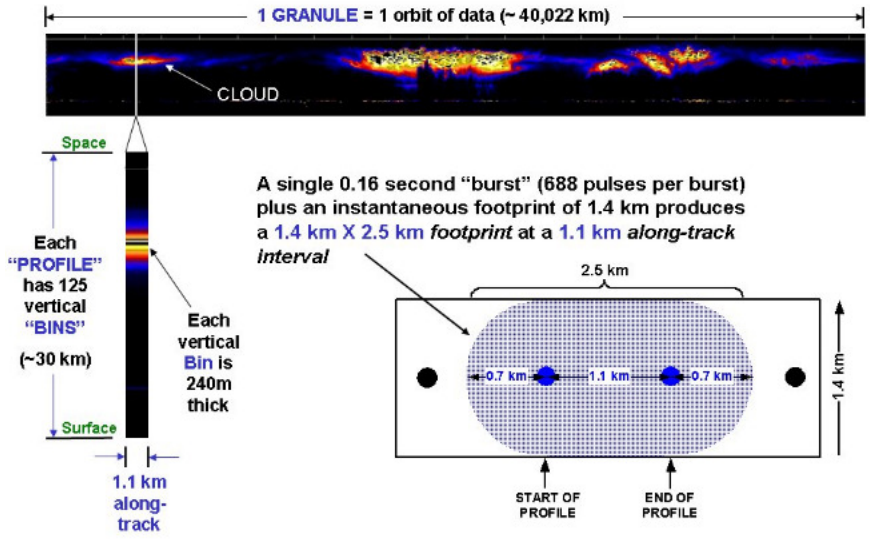


Figure 3.1: CloudSat profile from the CPR instrument. The CloudSat data is divided into granules, where each granule is defined as one orbit (beginning at the descending node), composing approximately 39400 profiles. Image credit: Colorado State University, Department of Atmospheric Science and NASA JPL.

Table 3.1: Information provided by the 2B-GEOPROF product, selected for relevance to the thesis (CloudSatDPC, 2020).

Field name	Information	Units	Range
CPR_Cloud_mask	Assigns a mask value to each resolution volume 0 = no cloud detected 1 = likely bad data 5 = likely ground clutter 5-10 = weak detection 20-40 = cloud detected	N/A	N/A
Radar_Reflectivity	Factor Ze calculated from the echo power	dBZe	-40 to 50
DEM_elevation	Elevation above Mean Sea Level (MSL) A value of -9999 indicates ocean	meters	-9999 to 8850
Height	Height of radar range bin above MSL	meters	-5000 to 30000
Latitude	Spacecraft geodetic latitude	degrees	-90 to 90
Longitude	Spacecraft geodetic longitude	degrees	-180 to 180
UTC_start	UTC seconds since 00:00 Z of first profile	seconds	0 to 86400

Several *auxiliary* data products from observation campaigns, mainly from instruments on other satellites in the A-train, have been produced. These auxiliary data products (systematically named with the suffix AUX) have been mapped horizontally and/or vertically to the center of each CloudSat profile bin (CSU, 2017).

3.2 MODIS data

The MODIS spectrometer aboard the Aqua satellite (launched in May 2002) is also part of the A-Train which makes it possible to allow close spatiotemporal collocation of the data from the two instruments. The Aqua satellite has an equatorial crossing time at about 1:30 PM local time, in ascending (northbound) direction. CloudSat lags Aqua by between 30 seconds and 2 minutes (NASA, 2020a). The orbit track of Aqua can be seen in Figure 3.2. The polar orbit has a slight inclination, resulting in a maximum and minimum geodetic latitude of approximately ± 82° for the spacecraft.

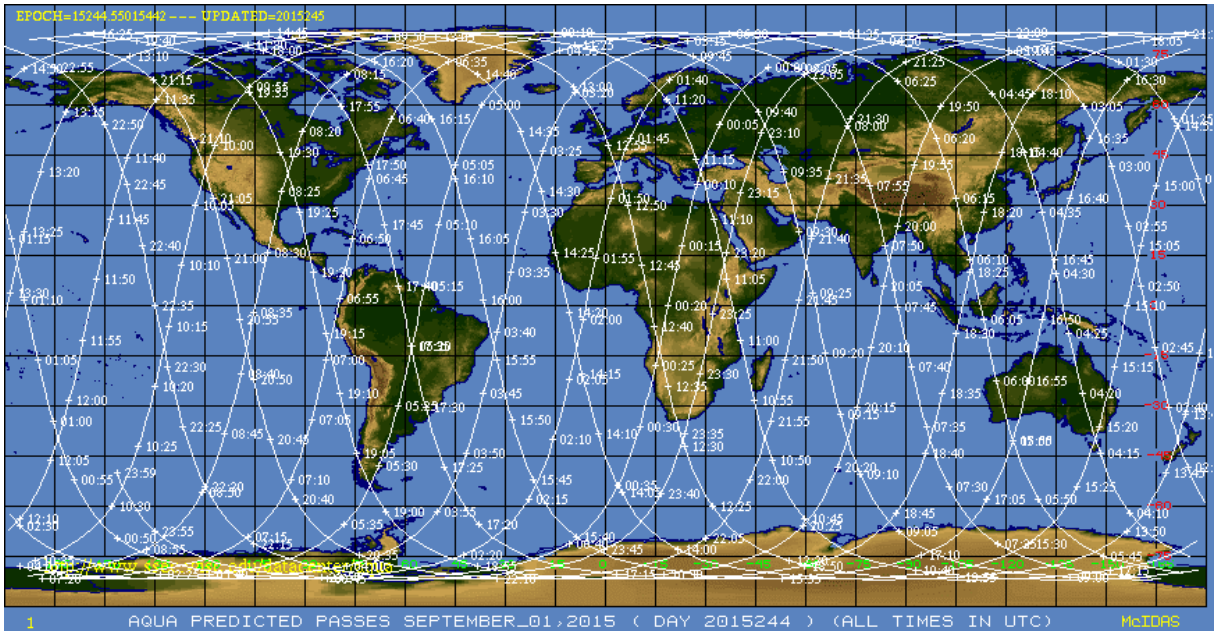


Figure 3.2: Aqua orbit track September 1, 2015. Image credits: SSEC, University of Wisconsin-Madison

MODIS has a swath of approximately 2300 km in the across track direction and 10 km along track at nadir (NASA, 2009). The CloudSat footprint is designed to always occur within the central 36 km of the MODIS swath (eoPortal, 2020). Because of its large swath, the MODIS instrument views the entire Earth's surface every 1 to 2 days. The instrument acquires data in 36 spectral bands with wavelengths in the range 0.4 µm to 14.4 µm. Several data products are made available from MODIS, with spatial resolutions of 250 m (bands 1-2), 500 m (bands 3-7) and 1000 m (bands 8-36) (NASA, 2020b). Bands 20 and 27-36 of MODIS are frequently referred to as *emissive bands* since they are not dependent on reflected solar radiation. The wavelengths of these bands as well as their primary use is presented in Table 3.2

Table 3.2: Wavelengths and primary fields of investigation for MODIS bands 20 and 27-36 (NASA, 2020b).

Band	Bandwidth (µm)	Primary use
20	3.660 - 3.840	Surface/Cloud Temperature
27	6.535 - 6.895	Cirrus Clouds Water Vapor
28	7.175 - 7.475	Cirrus Clouds Water Vapor
29	8.400 - 8.700	Cloud Properties
30	9.580 - 9.880	Ozone
31	10.780 - 11.280	Surface/Cloud Temperature
32	11.770 - 12-270	Surface/Cloud Temperature
33	13.185 - 13.485	Cloud Top Altitude
34	13.485 - 13.785	Cloud Top Altitude
35	13.785 - 14.085	Cloud Top Altitude
36	14.085 - 14.385	Cloud Top Altitude

The MODIS data contained in the CloudSat MOD-AUX product has been selected from the main data product MYD02_1KM_L1B (level 1B data product with a resolution of 1 km) to extract the 15 nearest measurement cells at each measured coordinate of the CloudSat data. This subset data is structured as a 3-pixel across-track by 5-pixel along-track grid centered on the CPR footprint (Cronk and Partain, 2017b). The MOD-AUX product contains bands 1-7, 17-20, and 26-36 from the original MODIS data product, however, only the emissive bands (Table 3.2) are of relevance for this project. The necessary information extracted from the MOD-AUX product is presented in Table 3.3. All relevant MODIS product are available at ICARE Data and Services Center (ICARE, 2020).

Table 3.3: Information provided by the MOD-AUX product, selected for relevance to the thesis (Cronk and Partain, 2017b).

Field name	Information	Units	Range
EV_1KM_Emissive	Radiances (unscaled) for bands 20 and 27-36	W/m^{2}srµm	0 to 32767
EV_1KM_Emissive_rad_scales	Radiance scales to convert unscaled radiance data	N/A	N/A
EV_1KM_Emissive_rad_offsets	Radiance offsets to convert unscaled radiance data	N/A	N/A
EV_1KM_Emissive_Uncert_Indexes	Uncertainty indexes for the radiance data	N/A	0 to 15
EV_1KM_Emissive_spec_uncert	Specified uncertainty	N/A	N/A
EV_1KM_Emissive_scaling_factor	Scaling factor for uncertainty	N/A	N/A
Latitude	Spacecraft geodetic latitude	degrees	-90 to 90
Longitude	Spacecraft geodetic longitude	degrees	-180 to 180
UTC_start	UTC seconds since 00:00 Z of first profile	seconds	0 to 86400

The scaled radiances for MODIS bands 20 and 27-36 are achieved by applying the radiance scales and radiance offsets to the unscaled radiances in the data fields presented in Table 3.3 for each measurement, according to the following equation:

$$L_{scaled} = (L_{unscaled} - L_{offset}) * s_L, \tag{3.1}$$

where L_{scaled} and $L_{unscaled}$ is the scaled and unscaled MODIS radiance, respectively. L_{offset} is the radiance offset and s_L is the scaling factor for the radiance. Radiance uncertainty for the MODIS data is calculated by:

$$u_m(\%) = u_{spec,m} * \exp(u_{index,m}/s_{u,m}), \tag{3.2}$$

where $u_m(\%)$ is the uncertainty of measurement m in percentage of the measured radiance, $u_{spec,m}$ is the specified uncertainty of the measurement, $u_{index,m}$ is the uncertainty index and s_m is the scaling factor for the uncertainty, all given in the corresponding data fields of Table 3.3.

3.3 ECMWF data

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides a large set of forecasts, climate re-analyses and multi-model data. ECMWF assembles data from around 90 satellite based instruments as well as ground based instruments. Much of this assembled data comes from the Copernicus Climate Change Service (C3S). Using data assimilation, ECMWF are able to make accurate forecasts by including current states of the atmosphere and the Earth's surface as initial conditions in the forecast models (ECMWF, 2020). Data assimilation also allows for improving models of the atmosphere historically, using data from previous years.

The ECMWF-AUX product (available at ICARE) is based on climate re-analyses from the years of interest. State variable data are interpolated for each CloudSat profile contained in the 2B-GEOPROF product from 3-hourly forecasts of the atmosphere. A single data value is given at the location for each CPR ray at each radar bin height (Cronk and Partain, 2017a). The necessary data from the ECMWF-AUX product is contained in the *Temperature* field, which provides the atmospheric temperature at each location and altitude. Other fields that are necessary for synchronizing the ECMWF data with the CloudSat and MODIS data include coordinates and timestamp for each profile. All fields which have been used from the ECMWF-AUX product are presented in Table 3.4

Table 3.4: Information provided by the ECMWF-AUX product, selected for relevance to the thesis (Cronk and Partain, 2017a).

Field name	Information	Units	Range
Temperature	Atmospheric temperature at each height above mean sea level	Kelvin	N/A
EC_height	Height of the ECMWF data bins Bin 105 is at 0 m above MSL Each profile uses the same height information, bin height ∼240 m	meters	-5000 to 30000
Latitude	Spacecraft geodetic latitude (CloudSat)	degrees	-90 to 90
Longitude	Spacecraft geodetic longitude (CloudSat)	degrees	-180 to 180
UTC_start	UTC seconds since 00:00 Z of first profile	seconds	0 to 86400

3.4 Data selection and processing

The data set used in this study was created from the entire year of 2015 of the 2B-GEOPROF, MOD-AUX and ECMWF-AUX products. The set was divided into non-overlapping rectangular patches of radar reflectivity, with a size of 64 × 64 radar bins. These patches are defined as "scenes". Each scene corresponds to a projected physical area of 15.4 km in height and 70.4 km in horizontal along-track distance, with a 1.1 km along-track resolution and 240 m vertical bin size. Only along-track scenes were studied in this project, but the cloud-chord-length distribution is considered to be horizontally isotropic, suggesting that similar cloud distributions should be found in the across-track direction

(Guillaume et al., 2018). This means that the two-dimensional "curtain" sampling of CloudSat can be viewed as a sufficiently accurate representation of the three-dimensional atmosphere. The division of the data set into discrete scenes in the horizontal plane rather than continuous orbit scenes or individual measurement columns is motivated by the fact that adjacent columns tend to be similar and so their probability distributions depend on each other (Leinonen et al., 2019). Cloud-cord-lengths are skewed towards shorter scales, meaning that the very large scale regions are not necessary in order to get a statistical representation of cloud fields (Guillaume et al., 2018). The altitudes closest to the surface of the Earth were not to be recreated in this project due to extensive problems with surface clutter, so the lowest vertical point of the scenes were set at 1 km. This causes the highest bin edge of the scenes to occur at 16.4 km. Since close to all clouds occur in the troposphere, the top of which reaches altitudes between 8 km and 18 km based on latitude, as presented in Figure 2.3 of the Theory chapter, the scene height is expected to cover the entire range where clouds are detected by CloudSat.

Scenes over both land and ocean were used, with the constraint that the elevation of the ground is less than 500 m. If any of the bins in the scene had an elevation over 500 m, the entire scene was removed from the data set. The CPR cloud mask from the 2B-GEOPROF product was used to evaluate the amount of clouds detected in each scene. Only scenes where at least 30 percent of the bins contained detected clouds were chosen. In this way the processing of a large number of almost empty scenes could be avoided. The data set for the GAN (see Section 3.5) was created from the remaining scenes after this selection.

The data from the MOD-AUX and ECMWF-AUX (referred to as MODIS and ECMWF data throughout the rest of this paper) was matched to the radar reflectivity bins at each coordinate in every scene. As mentioned, the AUX products are created from larger data sets that have been mapped to the CloudSat data using a generic interpolate-to-reference algorithm. However, both the MODIS and ECMWF data in the AUX products contain longer measurement cycles than the CloudSat data, meaning that the AUX products still include data at locations that lack CloudSat measurements. In order to synchronize the MODIS and ECMWF data with the CloudSat data, the data files needed to be sorted according to the start time of their first profile. The number of measured data points in the MODIS and ECMWF data was adjusted such that the first and last data point corresponded to the first and last profile, respectively, of the CloudSat data for each orbit. The pixel in the 15-pixel-grid of the MODIS data with the least difference in coordinates from the CloudSat profile at each position in a scene was chosen. If the difference in geographical position of these pixels and the corresponding CloudSat profiles was more than 0.7 km for more than 30 percent of the positions in a scene, the whole scene was removed. The final data set for the CGAN (described in Section 3.6) was created from the remaining scenes, after the removal of scenes that did not meet the criterion of ground elevation, cloud coverage and matching of the different data products.

After creating the two data sets, the CloudSat radar reflectivity Z_{dB} was linearly rescaled from the range [-35 dBZ, 20 dBZ] to [-1,1] as

$$Z'_{\mathrm{dB},i} = 2\frac{Z_{\mathrm{dB},i} - min(Z_{\mathrm{dB}})}{max(Z_{\mathrm{dB}}) - min(Z_{\mathrm{dB}})} - 1, \tag{3.3}$$

where $Z_{\mathrm{dB},i}$ denotes each radar reflectivity in the total vector Z_{dB} of reflectivities, and

$Z'_{\mathrm{dB},i}$ is the value of the rescaled reflectivity. The values of $min(Z_{\mathrm{dB}})$ and $max(Z_{\mathrm{dB}})$ are -35 dBZ and 20 dBZ, respectively, given by the initial reflectivity range. Missing data points and values below -35 dBZ were set to -1, and values above 20 dBZ set to 1. This mapping of missing values to the minimum value of the rescaled range is motivated by the fact that the observed radar reflectivity in clouds typically decreases at the cloud edge, so the mapping allows for a smooth transition between clouds and cloudless regions. This was done for both data sets.

The MODIS data contains several channels that correspond to different wavelengths, as seen in Table 3.2. The channels 27-29 and 31-35 were chosen because they have wavelengths that allow for nighttime observations. Channel 20 is subject to non-negligible influence from incoming solar radiation (see Figure 2.4). Channel 30 operates at a wavelength strongly affected by ozone and is thus mostly important for studying the ozone layer (primarily in the stratosphere, above the cloud scenes), therefore it is not included in this work. The radiance for each of the eight remaining channels were used for each location in the CloudSat scenes. This means that there were eight 64-bin time series of MODIS data for each scene. Similarly 64 $\times$ 64 bins of temperature data from the ECMWF data was stored for each scene.

The MODIS radiance, R, of each channel was linearly rescaled by using the maximum and the minimum value of the dataset for each respective channel according to Equation (3.4):

$$R'_{i,j} = 2\frac{R_{i,j} + 0.0001 + min(R_i)}{max(R_i) - min(R_i)} - 1, \tag{3.4}$$

where i is the channel number and j is a position index. This mapped the radiance values to the range (-1,1], and the missing values were set to -1. The term 0.0001 was added such that the minimum radiance values could be discerned from the missing values. By normalizing each channel individually, the relative intensity between channels remains for each position in a scene. The values of the maximum and minimum radiance for each channel is displayed in Table 3.5.

Table 3.5: Normalization parameters for the different MODIS bands. The radiance parameters are given in the unit $\mathrm{Wm^{-2}sr^{-1}\mu m^{-1}}$.

Band i	27	28	29	31	32	33	34	35
Max(R_i)	2.3465	3.6409	13.3189	13.7809	11.9740	5.3143	3.7898	3.2748
Min(R_i)	0.0555	0.1160	0.2200	0.4849	0.5801	0.7596	0.8438	0.9779

In total 147643 scenes were used for CGAN and 172849 for GAN, with 90% of these randomly chosen to be in the training set, and the remaining 10% were used in the test set.

3.5 Generative adversarial network

The structure of the generative adversarial network was based on the network implemented by Leinonen et al. (2019). The GAN in this project differs from the network of Leinonen et al. (2019), mainly in that it does not take any MODIS data as input. The input to the generator is a single noise vector, and the discriminator has an input that consists of a real or a generated scene. Since the discriminator does not have MODIS data as input, the upsampler block used by Leinonen et al. is obsolete. The generator takes an input vector that consists of normal distributed noise, with variance 1 and mean 0, and passes it to a dense layer. This is followed by a ReLU and a batch norm layer. Three layers follow that consist of one upsampling, one convolutional, one ReLU and lastly one batch norm layer. The final layer is a convolutional layer, which is followed by the tanh-function. The discriminator takes a scene, either real or generated, and passes it to four blocks that each consist of one convolutional layer and one leaky ReLU layer. The data is further sent to a dense layer and finally a sigmoid-function is applied. The output of the discriminator is thus a value between 0 and 1, representing the probability that the inserted scene is real. The structure of the generator and discriminator of the GAN is visualized in Figure 3.3 and 3.4.

The discriminator was trained by calculating the gradient in two steps. First, a batch of real samples, with *real* label (1), from the training set was passed through the discriminator and the loss calculated ($\log(D(x))$), then the gradient was calculated in a backward pass. Secondly, a batch of fake samples, with *fake* label (0), from the current generator was created. This batch was forwarded through the discriminator, and the loss was calculated ($\log(1 - D(G(z)))$). The gradient was calculated in a backward pass. The two gradients were combined and an update of the discriminator was performed with the optimizer of the discriminator.

The generator was trained by using the batch of fake samples used for training the discriminator, but switching the labels to *real* (1) and calculating the loss of the generator ($\log(D(G(z)))$). The gradient of the generator was then calculated in a backwards pass, and the parameters of the generator updated with an optimizer step. The generator and discriminator were both updated for each batch.

The GAN was implemented using the package PyTorch for Python, and optimized using the ADAM-algorithm, with the use of Binary Cross-Entropy (BCE) as the cost function separately for the generator and the discriminator. Both ADAM-optimizers used initial conditions $\beta_1 = 0.5$ and $\beta_2 = 0.999$, which control the decay rates of the moving average of the gradient. The initial learning rate was set at $\alpha = 0.0002$. Normal distributed noise, multiplied by a noise parameter was added to each reference image before passing through the discriminator. The noise parameter was initialized as 0.7 and decayed with factor 0.9 each epoch. The network was trained using minibatch-training, with a batch size of 64. The training was continued until convergence of both loss-functions, with the network parameters being saved every 100 epochs. The GAN was trained for 3000 epochs, which took approximately 12 days. The training was done on a Nvidia Tesla K40 GPU accelerator.

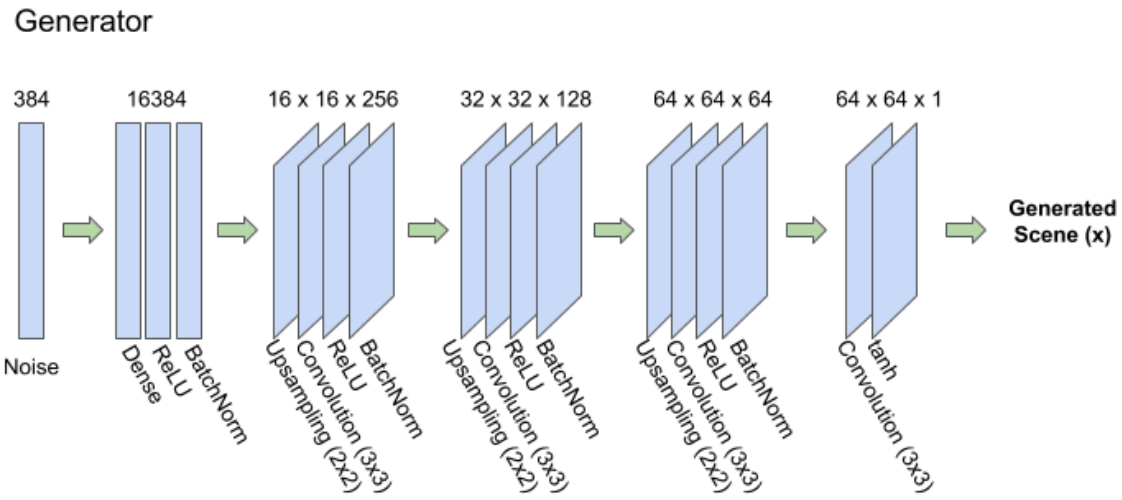


Figure 3.3: Schematic of the structure of the generator part of the GAN.

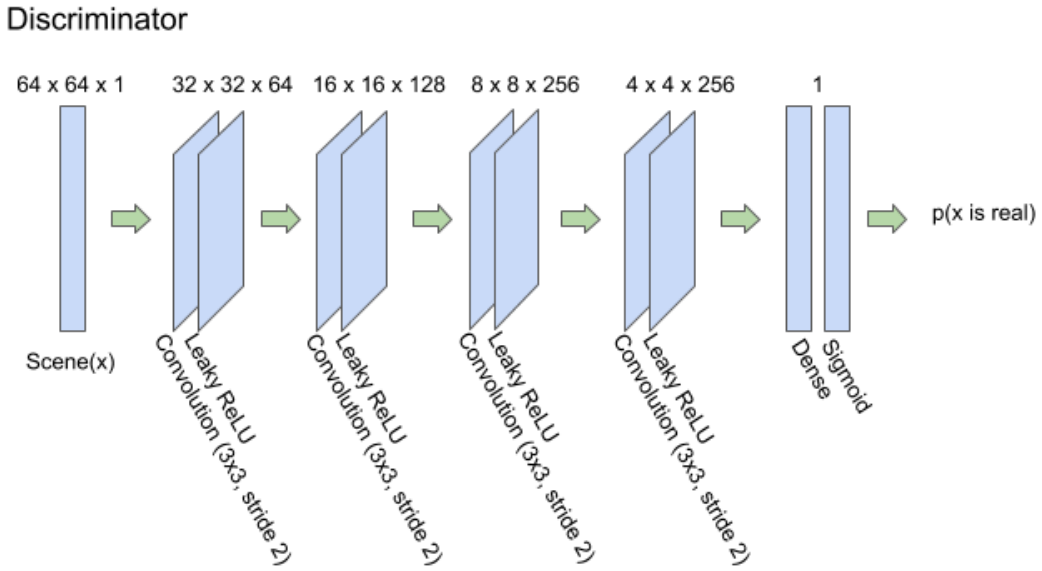


Figure 3.4: Schematic of the structure of the discriminator part of the GAN.

3.6 Conditional generative adversarial network

The structure of the conditional generative adversarial network was based on the network implemented by Leinonen et al. (2019), with a few changes. The main changes in this project was that instead of using derived physical quantities of the MODIS data as input, raw radiance data from selected MODIS channels was utilized. There were minor changes to the dimensions of inner layers, due to the change of the input dimensions. The change of input is due to the fact that derived quantities could lack important information for the network, but all the information for the physical quantities should be contained in the raw data. The generator takes an array of MODIS data and a noise vector as input. The inputs are flattened and concatenated, and sent to a block with a dense, a ReLu and a batch norm layer. After this follows three blocks consisting of an upsampling, a convolutional, a ReLu and a batch norm layer. The final layer is a convolutional layer followed by the tanh activation function. The discriminator starts by passing the MODIS input through an upsampler. The upsampler consists of four blocks, each one composed of an upsampling, a convolutional, a leaky ReLu, and a batch norm layer. The output of the upsampler is concatenated with a real or generated scene. Following this are four blocks consisting of a convolutional and a leaky ReLU layer. The final layer is a dense layer and the sigmoid-function is applied. The structure of the generator is visualised in Figure 3.5 and the structure of the discriminator is visualised in Figure 3.6.

The training of the discriminator and generator of the CGAN followed the same structure as the training of the GAN, which is described in Section 3.5.

The CGAN was implemented using the package PyTorch for Python. The network was optimized using the ADAM-algorithm, with the use of Binary Cross-Entropy (BCE) as the cost function, with a separate cost function for the generator and discriminator. The same initial conditions from GAN was used for the optimizer, i.e. $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\alpha = 0.0002$. Normal distributed noise, multiplied with a noise parameter was added to the reference images before the scenes and output from the upsampler was past through the final part of the discriminator. The noise parameter was initialized as 0.7 and decayed with a factor 0.8 each epoch. The network was trained using minibatch-training, with a batch size of 64. The training of the CGAN continued for 3500 epochs, with the parameters of the network being saved every 25 epochs, which took approximately 20 days. Convergence of the loss-functions was not fully reached due to time constraints. The training was done on a Nvidia Tesla K40 GPU accelerator.

3.7 Iterative amplitude adjusted Fourier transform

The IAAFT-algorithm proposed in Section 2.6 was implemented with CloudSat scenes as input. The algorithm was implemented in MATLAB. The data that was used was taken from the training and test set for the GAN.

The code developed by Venema et al. (2006) is freely available at the IAAFT main website (Venema, 2006). This project applied the 2D vertical version of the IAAFT algorithm in order to create surrogate cloud fields with similar vertical statistics as the template. The output of this method was to be compared to that of the neural networks.

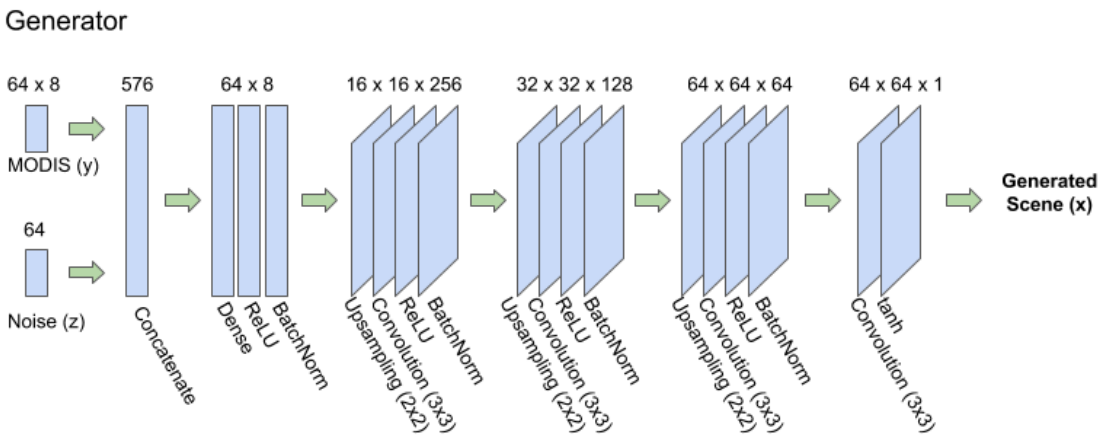


Figure 3.5: Schematic of the structure of the generator part of the CGAN. The structure is similar to that of Leinonen et al. (2019).

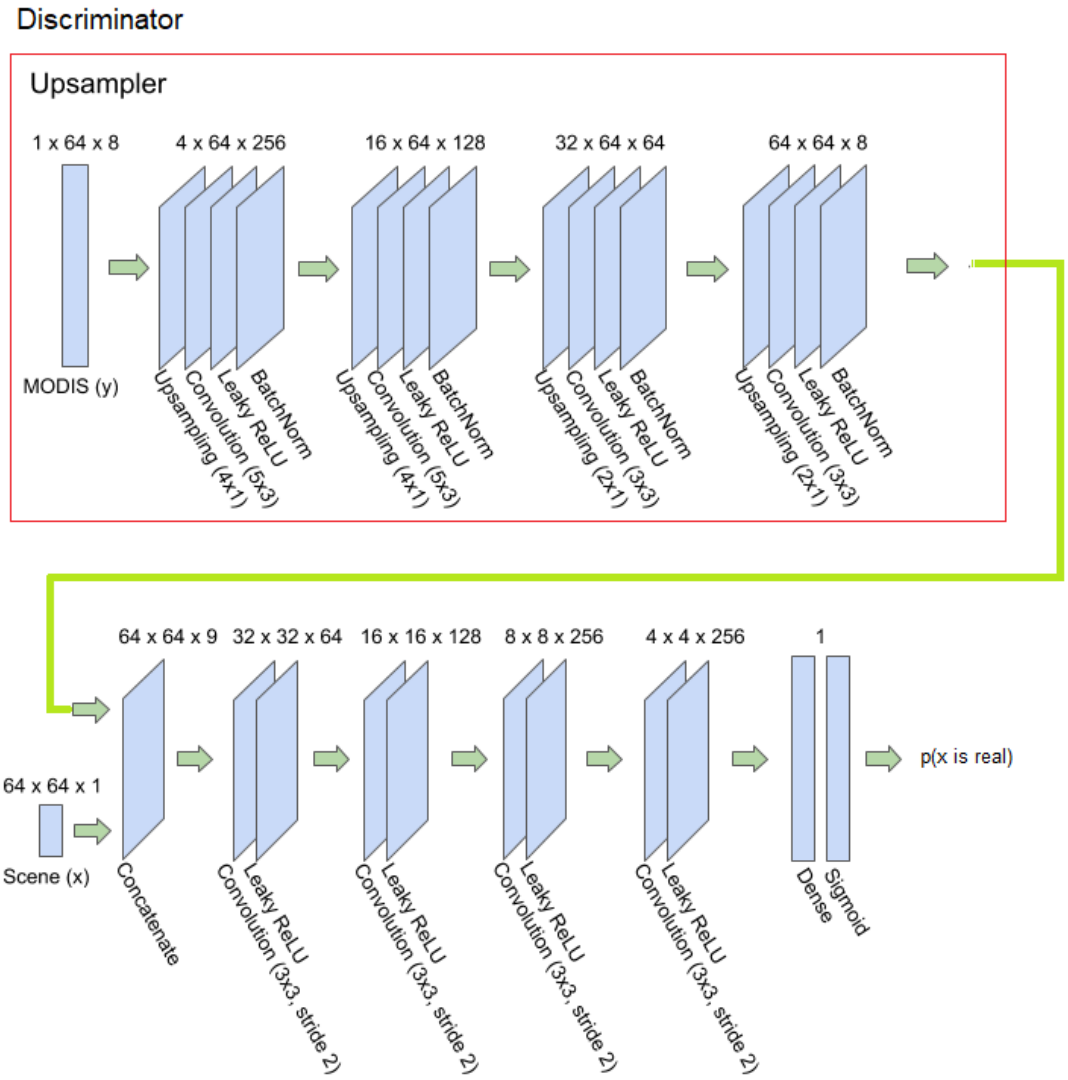


Figure 3.6: Schematic of the structure of the discriminator part of the CGAN. The upsampler component of the discriminator is marked by a red rectangle. The structure is similar to that of Leinonen et al. (2019).

3.8 Statistical analyses and retrieval of geophysical products

Several quantitative measures of the output of the GAN, CGAN and IAAFT were studied. The output of the generator part of the network for GAN and CGAN with scenes generated from each test set were scaled back to the initial range [-35 dBZ, 20 dBZ]. Reflectivities at -35 dBZ were considered as missing values throughout the statistical analyses, as a result of the rescaling of reflectivities. Since the CPR rarely detects clouds below -30 dBZ it is not a major limitation that the minimum values also get flagged as missing values. In order to analyse the ability of each network to replicate the overall vertical distribution of clouds, the global average of radar reflectivities at the different altitudes in a scene was mapped by histograms.

The downselection of data based on cloud coverage and ground elevation introduces some bias into the global distribution of cloud fields since not all cloud scenarios are represented. The use of a single year of data for training the networks may also be insufficient in regards to capturing the interannual variability of cloud properties. The extent of this bias needs to be analysed by testing the applicability of the trained network to other distributions. This was done by studying the performance of the two networks on data from the year 2016, with scenes selected based on the same criteria as above. In total 32615 scenes from 2016 were analysed for CGAN and 39804 scenes for GAN. The performance of the network on different geographical locations such as ocean or land or different latitude regions was also tested in order to identify possible bias in the performance.

3.8.1 Ice water path

The IWC of real and generated cloud scenes was calculated from Equations (2.8) and (2.9) in Section 2.4 of the Theory chapter. For real scenes from the test set for CGAN as well as its generated scenes, the altitude where the temperature in the atmosphere is 0 °C (273.15 K) was identified from the 64 vertical ECMWF bins in the data set. The IWC was then integrated upwards in a vertical column from this freezing level in order to estimate the IWP for each horizontal position.

A simpler model for the temperature in the atmosphere, based on Equation (2.2), was used for the IAAFT template and generated scenes. This model provides a global estimate of the altitude of the freezing level, from which the IWC was integrated upwards in a column to calculate an approximate IWP.

The IWP was not calculated for the scenes generated with GAN, since these generated scenes do not contain any reference to a specific location or time and can therefore not be compared to a real scene.

3.8.2 Cloud-top height

The CTH of each scene was found by scanning downwards from the top of the scene at each horizontal position until a radar reflectivity value above -20 dBZ was found. The corresponding physical altitude for the bin containing the first value above this threshold was denoted as the CTH of that position. If no value above -20 dBZ was found, the CTH

was set to 1 km, corresponding to the bottom of the scene. Note that this method only finds the first CTH. In the case of cloud fields with high fragmentation, several lower CTHs may also exist. These can be found by a similar approach, but were not studied in this project. The vertical resolution of the CTH is the same as that of the CloudSat data, i.e. 240 m.

3.8.3 Spatial autocorrelation

Two measures of the spatial autocorrelation were computed. The first was given by the autocorrelation function at different spatial lags in Equation (2.19). This can be said to be an indication of the local autocorrelation and was calculated for each altitude of both the real and generated scenes. The second measure was given by Moran's I from Equation (2.20). This was calculated for each scene as a whole, as it provides implications of the global coherence of an image. Only the nearest horizontal and vertical neighbors to each bin was used for calculating this index.

3.9 Network performance

3.9.1 Discriminator classification

The discriminator of the two networks was tested by using several different data sets as input and studying how they are classified. This is a test that shows an intermediate result of the network, which could explain why the network acts the way it does. In order to visualize the results histograms of how the discriminators classify the different sets are obtained. GANs discriminator is tested with five different data set: scenes generated by IAAFT, scenes generated by GAN, scenes generated by CGAN, test and training set for GAN. The discriminator of CGAN is tested with three different data sets: scenes generated by CGAN, test and training set for CGAN. The reason that CGAN is tested with fewer data sets is that its discriminator requires MODIS data, which the GAN discriminator does not.

3.9.2 Error sensitivity

The measurements of the radiances in the MODIS data are subject to some uncertainty as specified in Section 3.2. In order to get an indication of the network response to errors of different sizes (within the maximum uncertainty range) in the input data, scenes were also generated with additional uncertainty added to the MODIS input of the CGAN. Note that the original uncertainty present in the data cannot be removed, but its effect on the output of the network can be analysed. Three levels of uncertainty were added, with magnitudes corresponding to the minimum, average and maximum uncertainty of the data.

4

Results and Discussion

In this chapter, results from the experiments described in the previous chapter are presented and analysed. There are three main types of results. Firstly, example scenes generated by the three different methods are presented and analysed in Section 4.1. Next, several quantitative measures and comparisons between the generated and real data sets are laid out as well as comparisons of derived physical quantities. These are presented in Section 4.2. Finally, the network response to different variations in the input data is discussed in Sections 4.3 through 4.5. If nothing else is stated, the results are taken from GAN at epoch 400 and CGAN at epoch 3500 of the training.

4.1 Example scenes

An important part of generating the types of cloud distributions studied in this project is that individual samples are satisfactory. The individual generated scenes should have the same overall appearance as the real CloudSat scenes, with some distinct features following physical processes that occur in clouds. Most importantly, the intensity distribution in the generated scenes should resemble the one in the real scenes. In this section examples of real scenes as well as example scenes generated with the different methods (GAN, IAAFT, and CGAN) are visualized, along with some derived physical variables, to give an indication of how the methods perform. This type of visual analysis, however, is not enough to accept the results as satisfactory. The generated scenes need to be physically and statistically examined further. A quantitative analysis is presented in Section 4.2.

In Figure 4.1 some examples of real scenes from the CloudSat data in the test set of the GAN are shown. These scenes are to be used as a reference for the visual inspection of generated scenes, to make the reader more familiar with the structures that can be seen in cloud fields. The majority of clouds presented in 4.1 have reflectivities between approximately -20 dBZ to 15 dBZ, where higher reflectivities typically indicate a higher density of cloud water particles. An important feature that should be pointed out is the diversity of cloud structures, the example scenes in Figure 4.1 contain a wide range of cloud altitudes, vertical (and horizontal) thicknesses and densities. The presence of smaller scale structures also varies, where some scenes contain cloud fields that are broken up into smaller regions, and other scenes have larger continuous cloud masses. An important feature that occurs in some of the scenes is the horizontal line (often brighter than the surrounding cloud field, but sometimes also darker) occurring between 3 km and 5 km. This is a characteristic lines often seen in this type of radar data, and is caused by a transition from ice to liquid particles and precipitation.

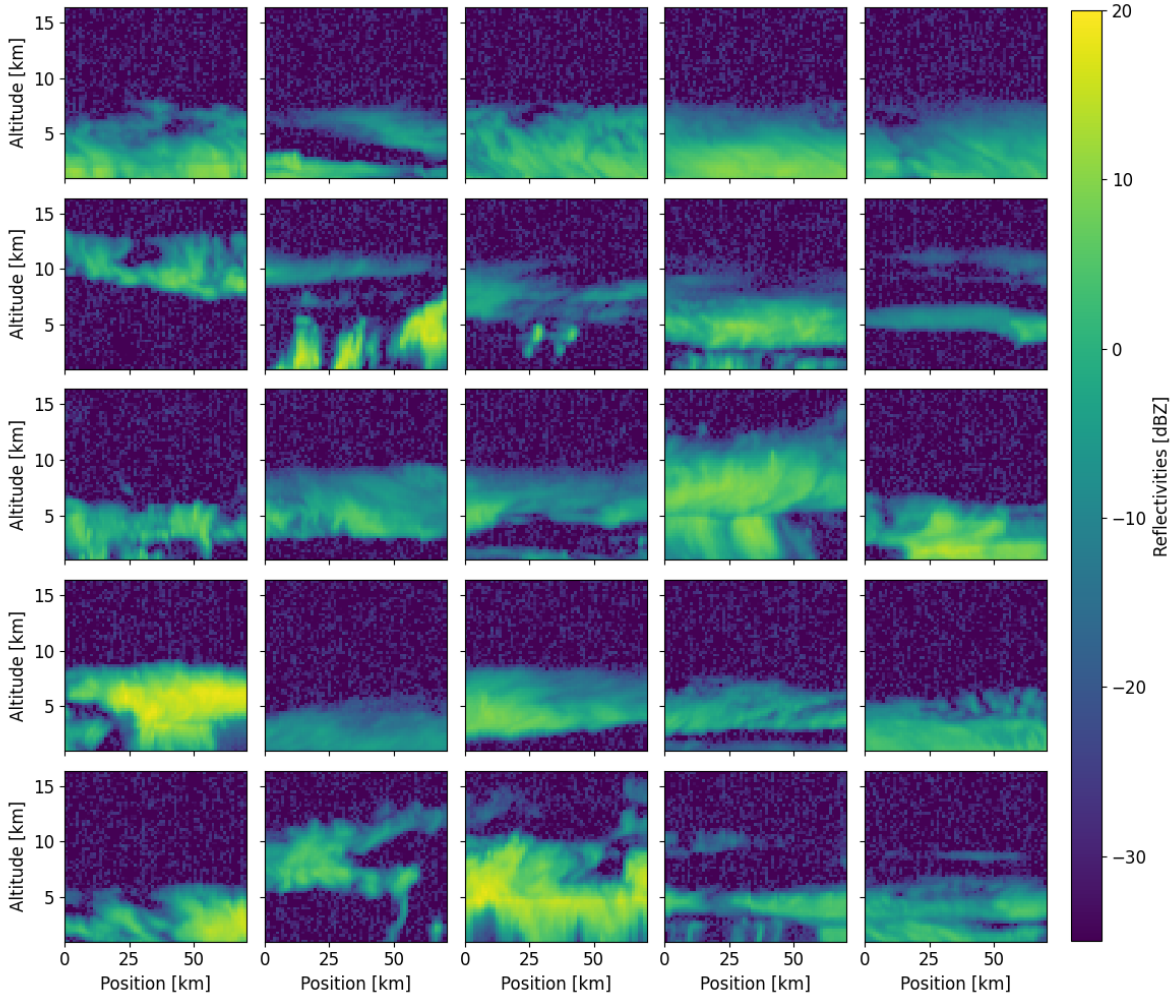


Figure 4.1: Example of 25 different real scenes consisting of vertical radar profiles from the CPR instrument on CloudSat. The reflectivity ranges from −35 dBZ to 20 dBZ.

4.1.1 Generative adversarial network

Examples of scenes generated with GAN using random noise as input are presented in Figure 4.2. An initial visual inspection of the cloud fields suggests that they may be representative of real cloud scenes. The transition of reflectivities between adjacent bins shows the same smoothness as in the real scenes of Figure 4.1. Structures that result from characteristic physical processes in clouds such as fall streaks, the bright line at the freezing level and precipitation below this level are visible in the scenes in the second row of Figure 4.2. The cloud fields seem to have a large variability in how they look, which is a desired feature. Both clouds with high and low degree of fragmentation can be seen in Figure 4.2. Clouds with large vertical thickness as well as cloud tops near the upper edge of the scene are present for examples in columns 3 and 5 of the first row. Thinner clouds are also noted, e.g. in row 3, as well as clouds with low cloud top heights. The density of the clouds also varies, which can be seen as a variability in the highest value of the reflectivity in each cloud. The GAN scenes do not have corresponding real scenes, so derived physical properties such as IWP and CTH are not shown for individual examples since a direct comparison to real examples cannot be made. The statistics of these quantities for the entire generated data set, however, should match those of the real data set and are shown in Section 4.2.6.

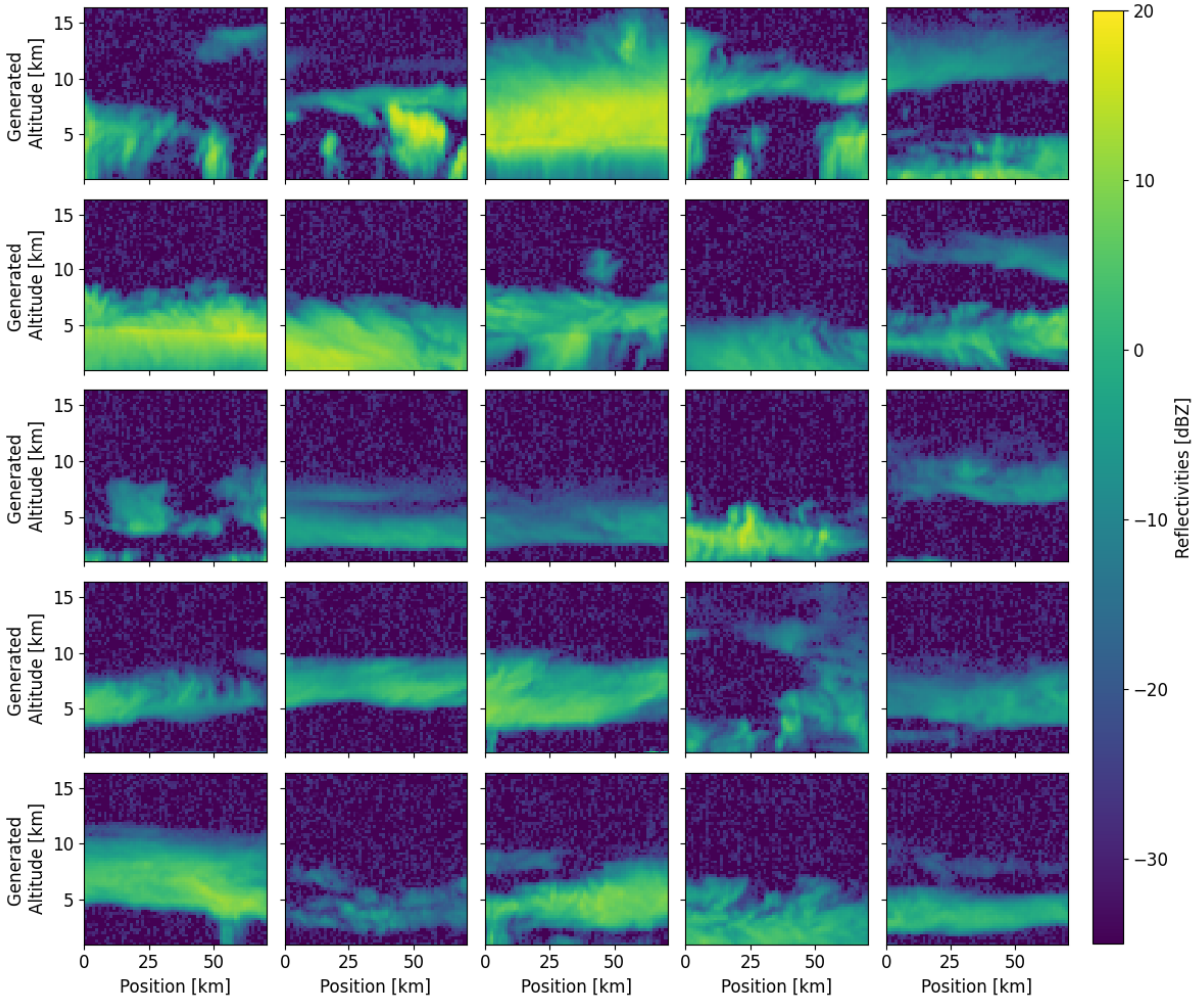


Figure 4.2: Example of 25 different scenes generated with GAN using random noise as input. The reflectivity ranges from −35 dBZ to 20 dBZ.

4.1.2 Iterative amplitude adjusted Fourier transform

Examples of scenes generated by IAAFT can be seen in Figure 4.3. These specific generated scenes seem to replicate both the IWP and the CTH well compared to the original scenes. It should be noted that the altitude of the 0 °C level, used for calculation of the IWP for IAAFT is based on a temperature model of the atmosphere and not specific measurements at each scene. The real and generated scene use the same altitude information of this freezing level. In some cases CTH and IWP seem to be shifted along the horizontal axis in the generated scenes compared to the real scenes, this can be seen in column 2 and 4 of Figure 4.3. Visual inspection of the generated scenes suggests that much of the internal structure of the cloud fields is missing from the generated scenes. This is evident in column 1, where the generated scene fails to recreate the fall streaks seen in the real scene. It appears that the IAAFT method only reproduces the basic statistics of the cloud scenes and is unable to cover more advanced features. Moran's index I printed on each scene appears to be lower for the generated scenes than the real scenes on average, which could be explained by the clear discontinuities between adjacent pixels in the clouds. The smooth transition of reflectivities in nearby pixels in the real scenes is replaced by sharper changes in reflectivities in the generated scenes. This is discussed further in Section 4.2.7.

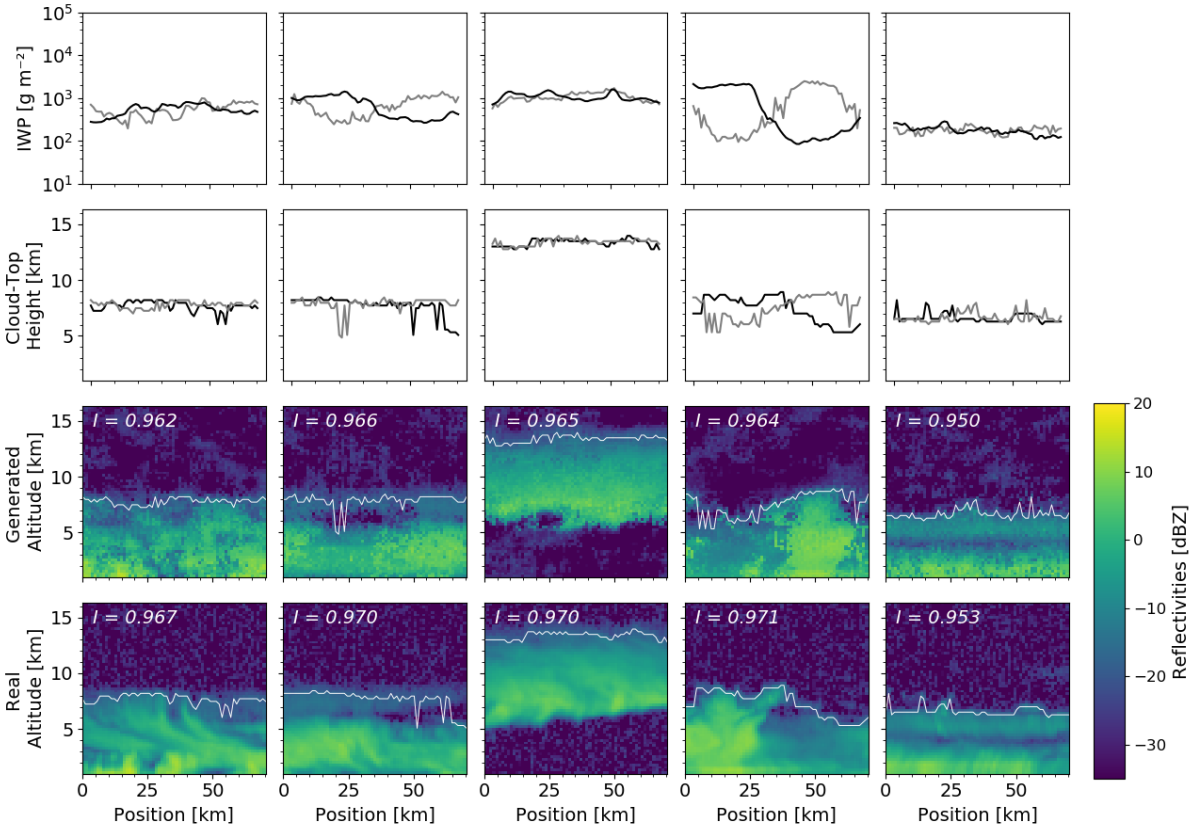


Figure 4.3: Examples of five different scenes generated with the IAAFT algorithm. Each column corresponds to a new scene. Row 1: IWP of the generated scene (grey) and the CloudSat template (black) shown in logarithmic scale with base 10. Row 2: Cloud-top height for the generated scene (grey) and the CloudSat template (black). The cloud-top height is also shown as a white outline in Row 3 and 4. Row 3: Generated scenes from the CloudSat template (Row 4) with reflectivities measured in dBZ. The spatial autocorrelation for the scene, given by Moran's index I is shown for each CloudSat template and generated scene.

4.1.3 Conditional generative adversarial network

Examples of scenes generated by CGAN are visualized in Figure 4.4, additional examples can be found in Section A.1 in the Appendix of this report. Real scenes are compared to the generated scenes with regards to IWP and CTH. These specific examples seem to capture the characteristics of the original scene rather well, both in terms of the visual inspection and the derived physical quantities. However, the generated scenes sometimes feature thin clouds at the top of the scene although these are not present in the original scene, visible in column 4. Apart from this, the generated scenes are able to recreate the top of the clouds very well, but have more difficulties reproducing the statistics of the lower parts of clouds, especially for thicker clouds. This is not surprising since the MODIS data is limited to the uppermost regions of clouds and does not perceive the underlying regions in most cases. It is, however, important that the network is able to recreate the reflectivity distribution in the lower part of the cloud field also when generating a large number of scenes, since the network had access to this information during the training. The performance of the network for scenes that consist mostly of a single uniform cloud is

significantly better than for scenes containing more complicated structures, both in terms of the visual appearance and the derived IWP and CTH (see also additional examples in Section A.1). These uniform clouds typically have a much flatter MODIS input and are easier for the network to replicate since all the important features of the cloud are available in the MODIS data, which is not the case when there are complex cloud structures below the top layer.

The normalized MODIS input for each generated scene is presented in the third row of Figure 4.4. The structures seen in this input are similar for all the MODIS bands except for one band (represented by the blue line) in each scene. The main difference between the MODIS input of different bands is the overall magnitude of the data. This suggests that all the bands register the same features in the top layer of the cloud, where an increase in the MODIS data typically corresponds to a decrease in the physical cloud-top height of the real scene.

There is a lack of variability in the generated scenes, especially for the lower parts of clouds, where precipitation should be present in some scenes. Since the MODIS instrument only provides data from the top layers of clouds, the underlying structure should present a variability that represents the distribution of possible scenarios, not just a single repeated structure in all generated scenes from the same original scene. This is apparent in the scenes of column 3, where the generated scenes capture the top of the clouds well but fail in the lower regions and also have a low variability with respect to each other. In column 2 (especially the first row of generated scenes) a dark line at an altitude around 4 km can be seen. This line, although not present in the original scene, is believed to represent the freezing level in clouds. It can be compared to the bright line seen at a similar altitude in the real scenes of column 1 and 3. This means that although the network does not always incorporate the line and the associated precipitation in generated scenes where it occurs in the original scene, the CGAN has learned that it is a feature that is present in the data set and thus incorporates it in some of the generated scenes. Below this line there is more variability present in the cloud distribution of the lower part of the scene than can be seen in generated scenes which lack this line.

4.2 Statistics and comparisons

An integral part of studying a system is to look at the statistics of relevant properties of said system. Which properties should be taken into account is a matter of discussion, and needs to be determined for each system and application. Several properties are considered significant to this project. The cloud vertical distribution is the most fundamental feature that needs to be well replicated by the generating methods, and is presented as histograms of the radar reflectivities of real and generated scenes at the different altitudes in Section 4.2.1. This analysis is also performed and presented in Section 4.2.2 and 4.2.3 for the network trained to different epochs to study the evolution of the network throughout the training. Comparisons of the cloud vertical structure are also made between data sets over ocean and land respectively, to check for variations in the performance of the methods over different types of surface backgrounds. These results are presented in Section 4.2.4. In Section 4.2.5 histograms of the radar reflectivities at different altitudes are compared for different intervals of latitudes. Since the atmospheric properties and thus also the structures of clouds present at different latitudes varies slightly, assessment of the output

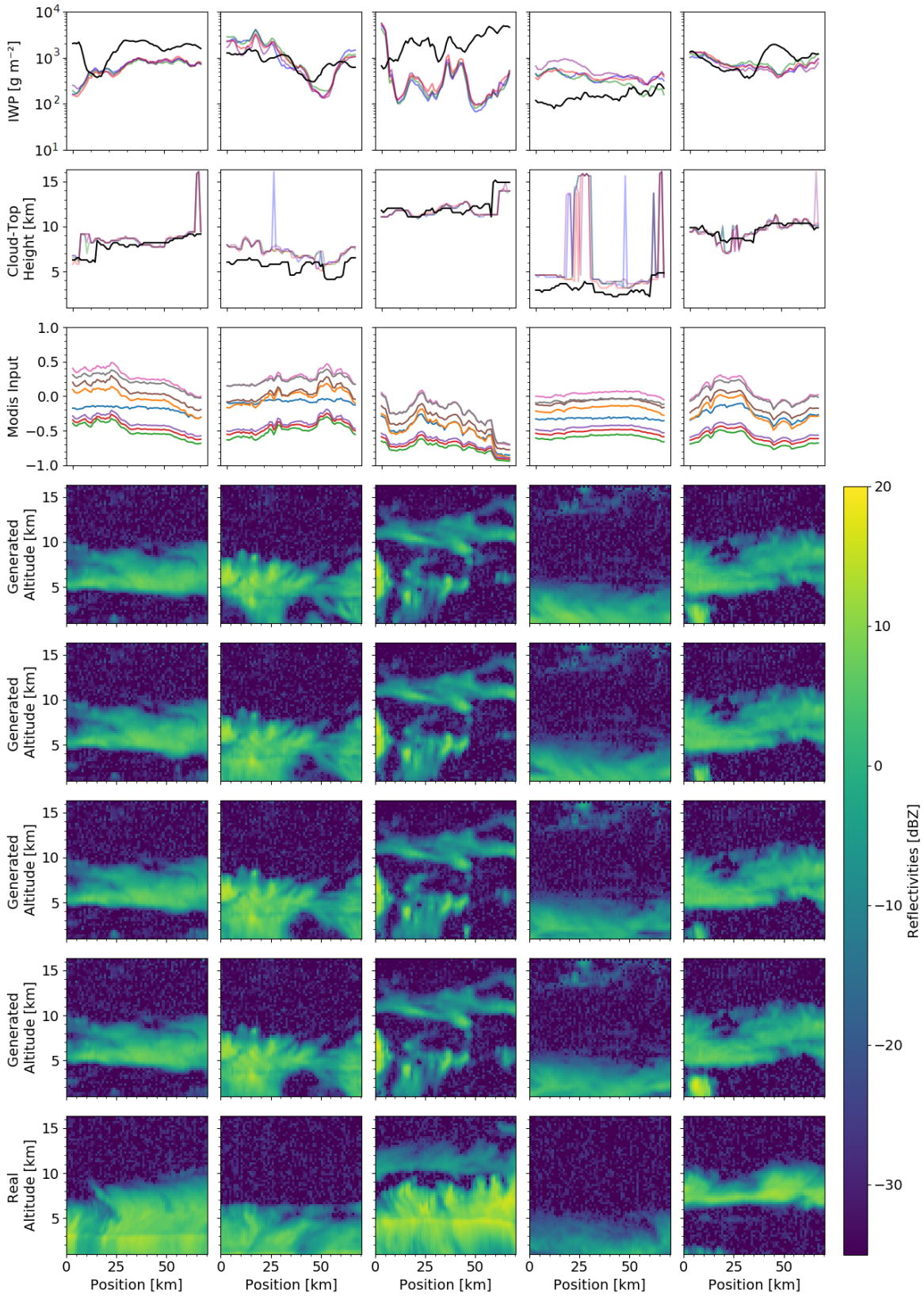


Figure 4.4: Examples of five different scenes for CGAN. Each column corresponds to one example. Row 1: IWP of the generated scene (colours) and the original scene (black) on a logarithmic y-axis. Row 2: CTH for generated (colours) and original (black) scenes. Row 3: Input of the eight MODIS channels. Row 4-7: Scenes generated with CGAN from the MODIS input, with different noise input. Row 8: Measured CloudSat scenes.

response to these variations is important. The physical implications of IWP and CTH and the distributions of these for different data sets are discussed in Section 4.2.6. Lastly, the two measures of spatial autocorrelation are analysed in Section 4.2.7.

4.2.1 Reflectivity distribution

The reflectivities of the different data sets are summarized in histograms in Figure 4.5 over all the scenes present in or generated by the test set relevant to each method. The histograms are normalized in units of dBZ for each altitude. This method of using frequency by altitude diagrams for analysing radar data climatologically is frequently used.

The first row of Figure 4.5 shows that the reflectivity distribution of the real data sets is very similar for the different models. Reflectivities above 15 dBZ are very uncommon for all altitudes and high reflectivity clouds are most common at low altitudes, possibly because the concentration of water vapour in the troposphere decreases with increasing altitude. The second row shows that there is a significantly higher variability between the different data sets of the generated scenes than there is for the real scenes. The diagonal line between 2 km to 5 km is present in the distribution of the generated scenes of both GAN and IAAFT, but not as clearly for CGAN. Furthermore CGAN has a high normalized occurrence of reflectivities at around 6 dBZ for very low altitudes (roughly 1 km - 4 km), which is not as significant in the other generated distributions, or the real distributions. Reflectivities below -25 dBZ are not visualized in the histograms since these reflectivities mostly represent noise and have a very large normalized occurrence, which makes the other more relevant reflectivities difficult to discern. In the distribution of generated scenes with CGAN there is some spillover of these reflectivities with very high occurrence, especially at higher altitudes, also for reflectivities up to -20 dBZ. Another interpretation is that these reflectivities are actually caused by generated low-density clouds.

The third row shows the differences between the generated distributions and the real distributions. For the GAN there is a large area where the generated scenes overestimate the amount clouds present at certain reflectivities at the altitudes 5 km to 15 km, which is seen as the red area. The difference histogram for IAAFT shows that the distribution of the generated scenes is very similar to that of the real data set. The difference histogram for CGAN shows very clearly the dissimilarities discussed earlier, with two region where the CGAN overestimates the occurrence of clouds with certain reflectivities. These areas are visualized as intensely red, between altitudes 10 km and 16 km with low reflectivities, and between 1 km and 6 km with higher reflectivities. The intense region at high altitudes with low reflectivity most likely corresponds to the addition of thin clouds at the top of the scene that could be seen in the examples of CGAN scenes in Figure 4.4.

4.2.2 Epoch evolution GAN

In order to assess which epoch of the training should be used for the final network, several properties of the training can be analysed. The loss is a measurement of how well the network performs. For a GAN it is important to study the loss of both the discriminator and the generator. The loss of the discriminator is defined as $\log D(x) + \log D(G(z))$, and the discriminator strives to minimize this value. The loss of the generator is defined as $\log D(G(z))$, which the generator instead works towards maximizing. The average loss for

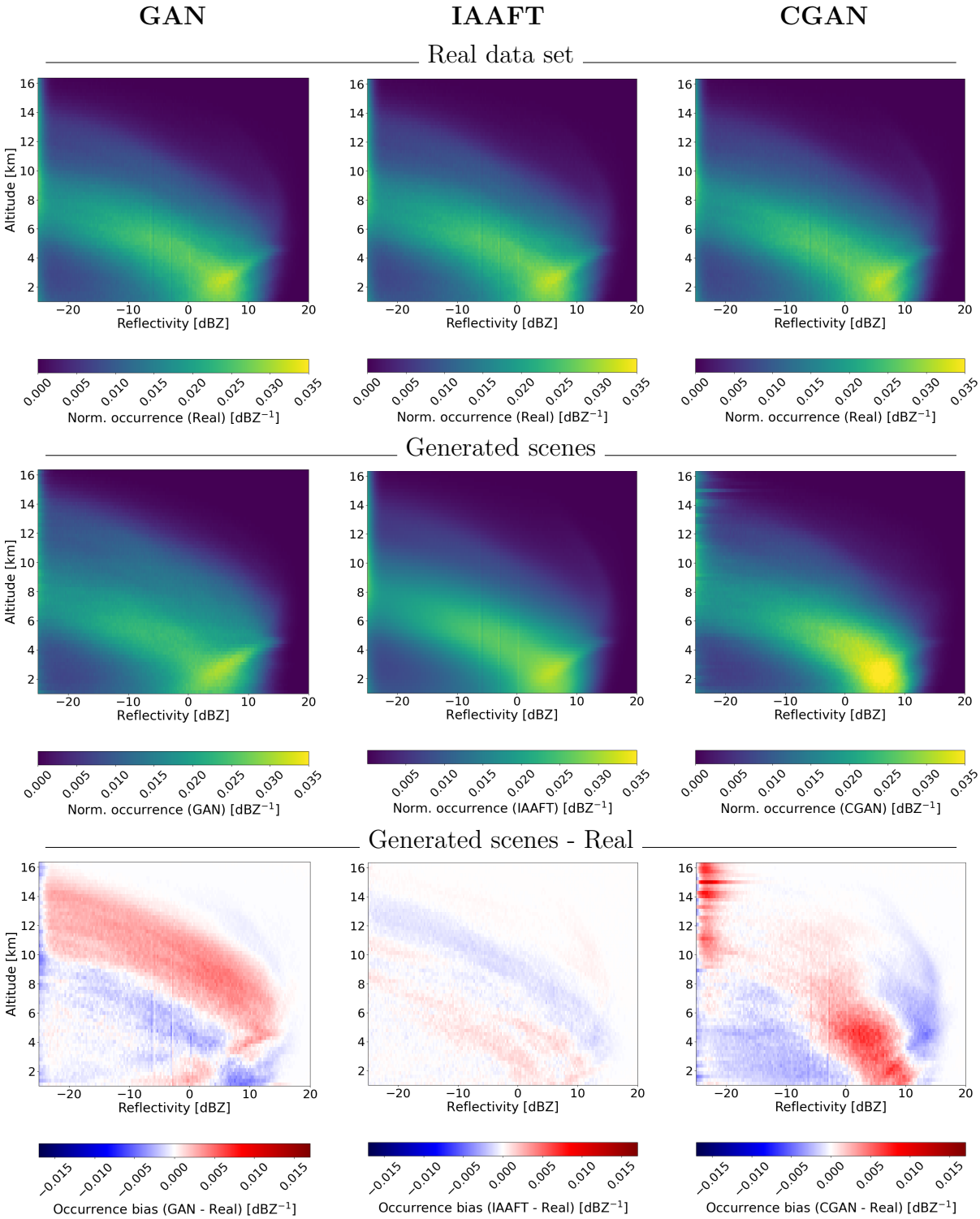
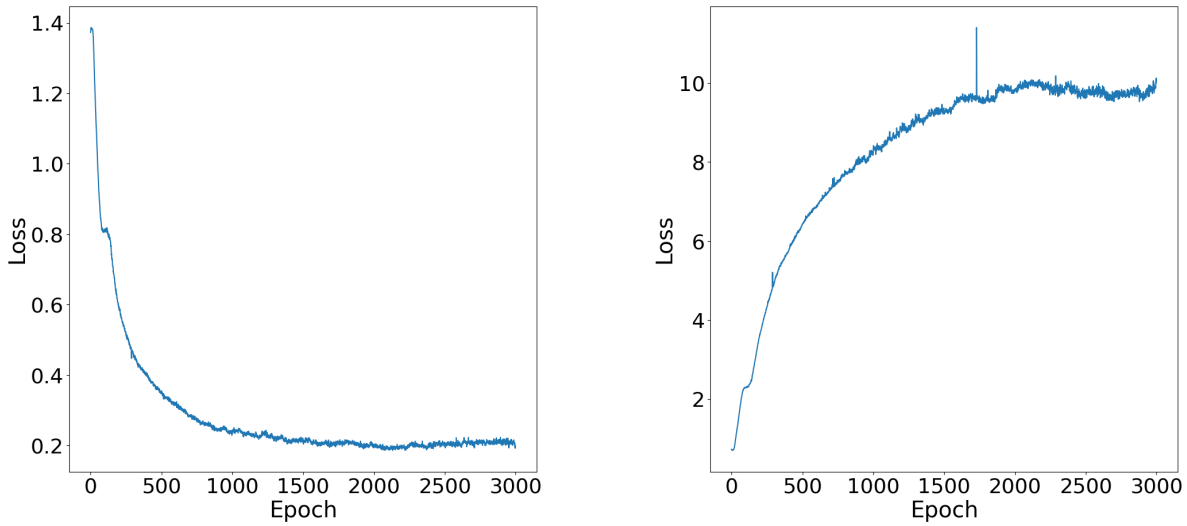


Figure 4.5: Normalized reflectivity-altitude histograms for GAN, IAAFT and CGAN. The first row contains data from the real data sets, note that the same data is used for GAN and IAAFT. The second row is based on data from the generated scenes. The third row shows the difference between the first and second row for each column. The columns contain, from left to right, GAN, IAAFT, and CGAN.

each epoch of the training of the GAN is presented in Figure 4.6. The average loss of the discriminator starts to converge already after epoch 1500, but the loss of the generator has not converged at this point. The average loss has converged for both the discriminator and the generator after 3000 epochs.



(a) The average loss-value for the discriminator plotted against epoch.

(b) The average loss-value for the generator plotted against epoch.

Figure 4.6: Graphs of the loss of the GAN.

Example scenes generated by GAN at epoch 3000 are visualized in Figure 4.7. The generated scenes do not show any discontinuities or artefacts. There is a degree of variability present and features such as fall streaks are clearly distinguishable (e.g. row 3 and 4 of column 3) and the bright freezing line (column 2). It is rather difficult to tell that the difference between the output of the network at this epoch and the network at epoch 400 from a visual inspection of example scenes only, because the scenes in Figure 4.7 are also very similar to real scenes. In fact they may even be too similar.

Figure 4.8 visualizes the difference histograms for GAN and real scenes at three different epochs of the training. All three difference histograms show larger occurrence bias than the difference histogram of epoch 400 (see Figure 4.5). This implies that there has been some event that reduces the accuracy of the results of the network in an epoch between 400 and 1400. For this reason the network from epoch 400 is chosen as the final version used for further analysis. It can also be noted that the network is significantly better at epoch 3000 than at epoch 1400 and 2400. A conclusion of examining the average loss over epochs in Figure 4.6 and the epoch evolution of Figure 4.8 is that the average loss is not a good standalone indicator of how the network is evolving with regards to the quality of the output. Visual inspection of individual scenes is also not an optimal method for discovering this type of changes in performance.

Additional statistical quantities (spatial autocorrelation and CTH) of the GAN at epoch 3000 are shown in Section A.5 in Appendix. The decrease in performance of the GAN at later epochs is apparent also for these properties.

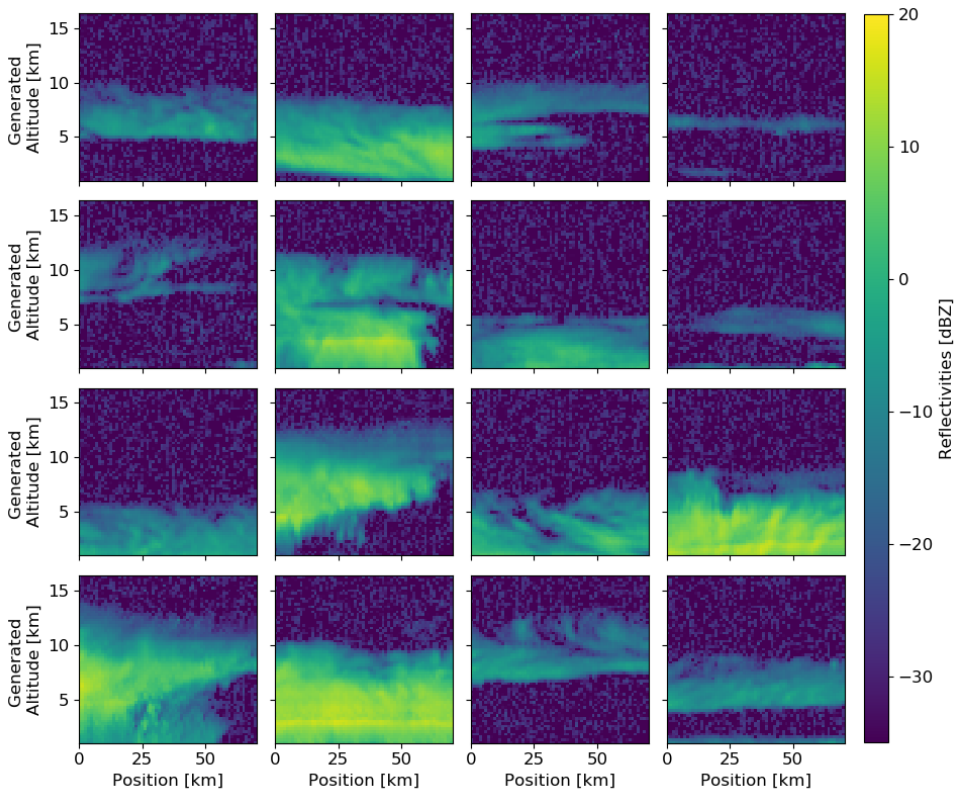


Figure 4.7: Example of 16 different scenes generated with GAN at epoch 3000 using random noise as input. The reflectivity ranges from −35 dBZ to 20 dBZ.

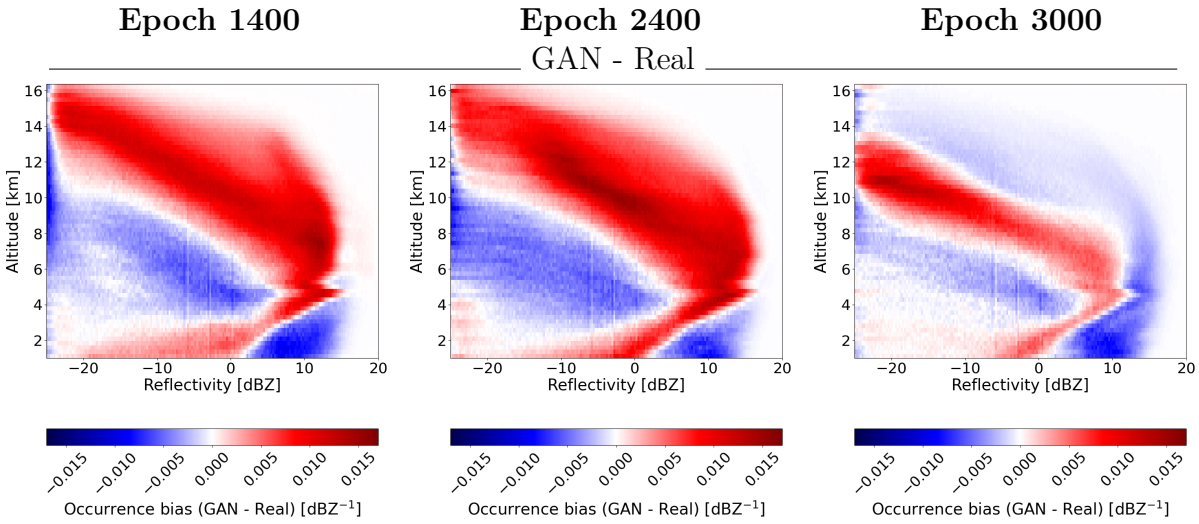
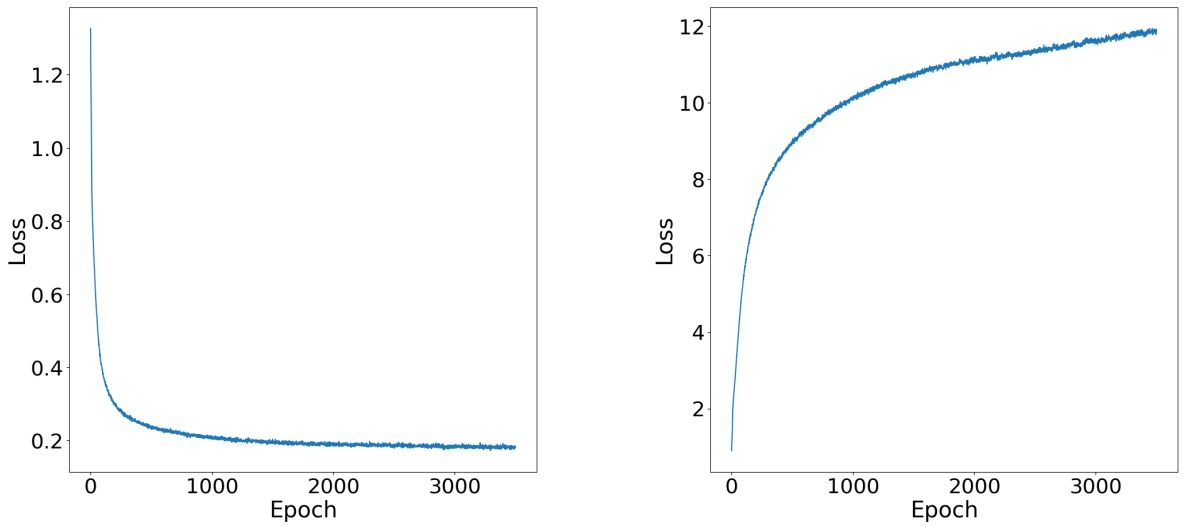


Figure 4.8: Difference histograms for GAN compared to the real data set. Evaluated at three different epochs of the training.

4.2.3 Epoch evolution CGAN

The average loss of the two parts of the CGAN are visualized in Figure 4.9. The average loss of the discriminator seems to have converged already after 1500 epochs, but the average loss of the generator does not seem to have converged even at 3500 epochs. The training was stopped due to time constraints.



(a) Average loss of the discriminator of the CGAN plotted against epoch number.

(b) Average loss of the generator of the CGAN plotted against epoch number.

Figure 4.9: Graphs of the loss of CGAN.

Figure 4.10 shows the difference histograms of scenes generated by CGAN and real scenes from three different epochs of the training. The performance of the output of the network seems to improve with training, and the output of the network at epoch 2500 is equivalent to that of the network at epoch 3500. This implies that the network has converged on a solution and further training is superfluous. The network of epoch 3500 is chosen as the final network. In this case the average loss and the performance of the network over the training seems to agree that the network is improved over the course of the training, which was not the case for GAN.

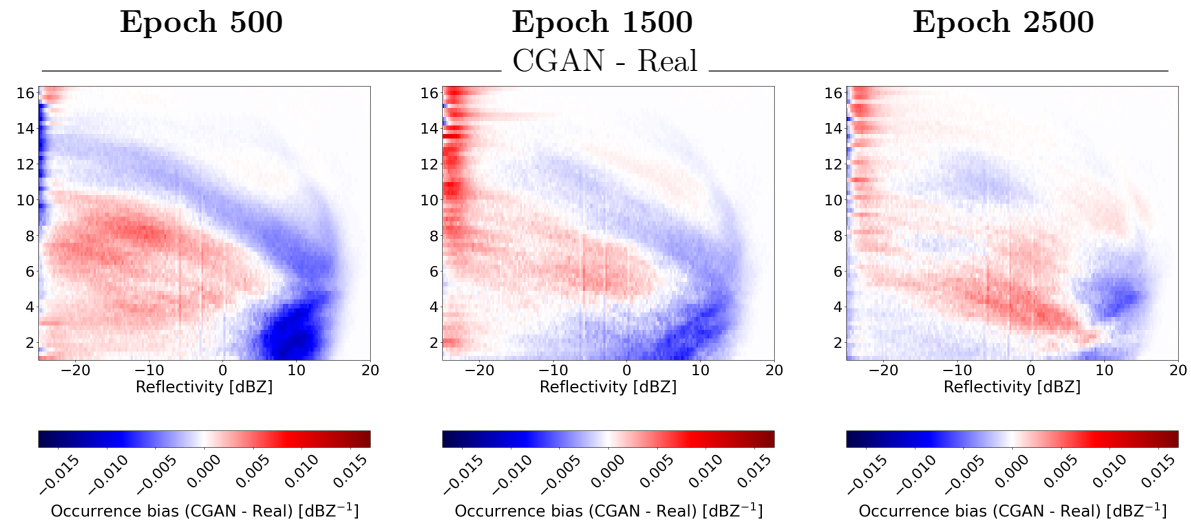


Figure 4.10: Difference histograms of scenes generated by CGAN and real scenes at three different epochs of training.

4.2.4 Ocean and land

The first row of Figure 4.11 shows the distribution for the CloudSat data set (for GAN) over geographic locations corresponding to ocean and land respectively. Setting the limit for *ocean scenes* to have at least two-thirds of positions in the scene over ocean tiles (scenes with less than two-thirds of positions over ocean are classified as *land scenes*) roughly 81% of scenes are classified as ocean scenes. Approximately 71% of the Earth's surface is covered in water. The deviation of the amount of ocean scenes from this value is likely caused by the inclination of the satellite orbits - causing no measurements near the poles - as well as the procedure for selecting the data set. Both the elevation and cloud coverage criteria are expected to be responsible for sorting away a higher fraction of the land scenes than ocean scenes since the cloud fraction is generally higher over ocean ($\sim$72%) than over land ($\sim$55%) (King et al., 2013).

The occurrence bias between cloud distributions over ocean and land, as seen in Figure 4.11 and Figure 4.12 (first row, last column), is rather small. This means that the overall cloud distributions over the two types of scenes are fairly similar, with clouds at slightly higher altitudes and thickness over oceans. The histograms also seem to suggest that clouds of higher density (higher reflectivities) are more common at altitudes around 3-5 km over ocean and around 1-2 km over land, as can be interpreted from the high intensity regions at these altitudes in the occurrence bias histograms of the real data set in Figure 4.11 and 4.12. The bias towards low-altitude, high-density clouds over oceans appear to be even stronger in the data set for the CGAN, visible as the high intensity region near 10 dBZ in occurrence bias histogram for the real data set of Figure 4.12.

The normalized radiance of the MODIS bands in Figure 4.13 is below 0 in general for all bands, over both ocean and land scenes. The dispersion of radiance, however, seems to be higher over ocean scenes than over land scenes which can be seen in the occurrence bias histogram of Figure 4.13. This histogram illustrates that the spread of the MODIS radiance in each band is larger over ocean scenes (broad red regions of radiances) than it is over land scenes (narrow dark blue regions of radiances). The interpretation of this is that a wider range of radiances is observed by MODIS over oceans than over land.

Overall, the difference in the data set over land and ocean is considered small enough not to affect the performance of the network for the combined distributions. In the last row of Figure 4.11 the occurrence bias between scenes generated with GAN and real scenes over ocean and land, respectively, can be seen to present similar distributions as the occurrence bias with respect to the entire data set seen in Figure 4.5 for GAN. A somewhat larger difference is seen between the total generated scenes and real CloudSat scenes over land than over ocean, suggesting that the network has trained to generate a distribution more like the one over ocean scenes. This means that the performance of the GAN is slightly better for cloud distributions over ocean, which is not surprising since the number of ocean scenes in the training data set is larger than the number of land scenes. Similarly, the last row of Figure 4.12 shows the occurrence bias between scenes generated with CGAN and real scenes over ocean and land respectively. This bias appears to be stronger for land scenes than for ocean scenes also for CGAN, suggesting that this network performs slightly better for cloud scenes over ocean. The performance of the CGAN for land scenes, however, is considered high enough for the scenes to be acceptable.

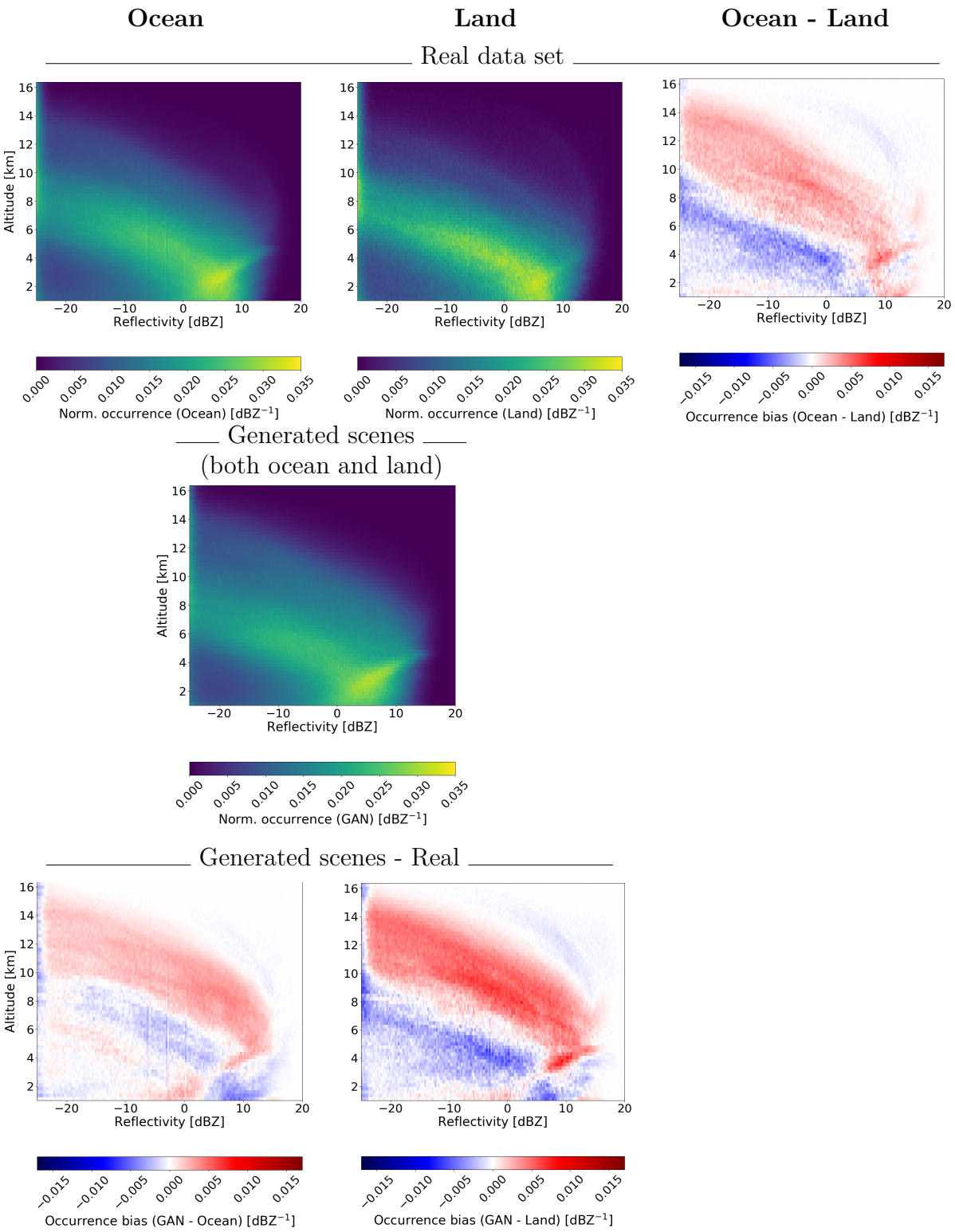


Figure 4.11: Normalized reflectivity-altitude histograms from the data set used for the GAN and generated scenes. Row 1: CloudSat data from ocean locations (left), land locations (middle) and the difference between ocean and land locations (right). Row 2: Scenes generated with GAN using the entire test set. Row 3: The occurrence difference between generated scenes and ocean scenes (left) or land scenes (right). Note that generated scenes for GAN do not have any indicator whether they are over land or ocean, so the occurrence bias in the last row is calculated for all generated scenes by GAN (both ocean and land) compared to the real data set over ocean or land.

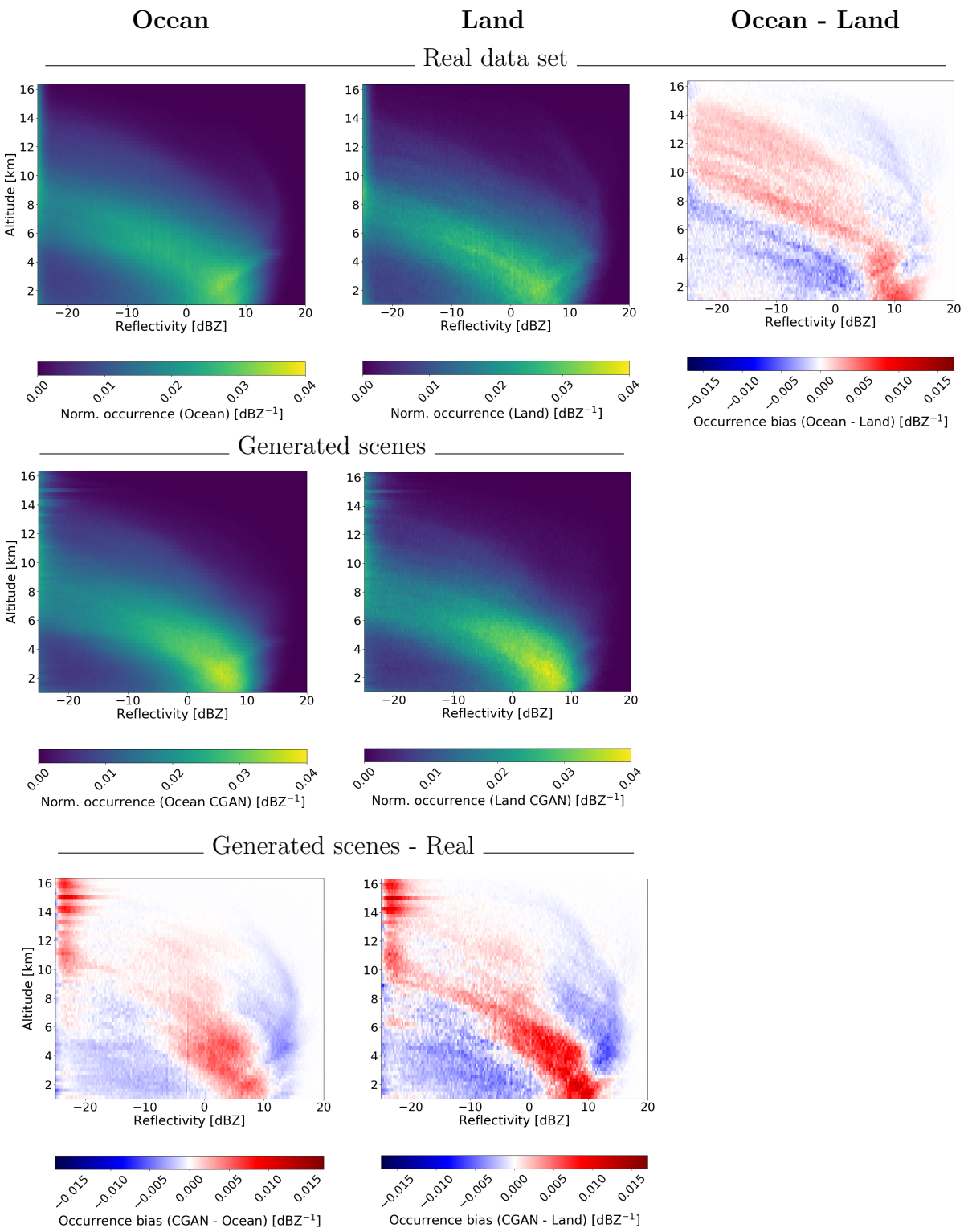


Figure 4.12: Normalized reflectivity-altitude histograms from the data set used for the CGAN and generated scenes over ocean and land, respectively. Row 1: CloudSat data from ocean locations (left), land locations (middle) and the difference between data over the two types of backgrounds (right). Row 2: Scenes generated with CGAN over ocean (left) and over land (right). Row 3: The occurrence difference between generated scenes and ocean scenes (left) or land scenes (right).

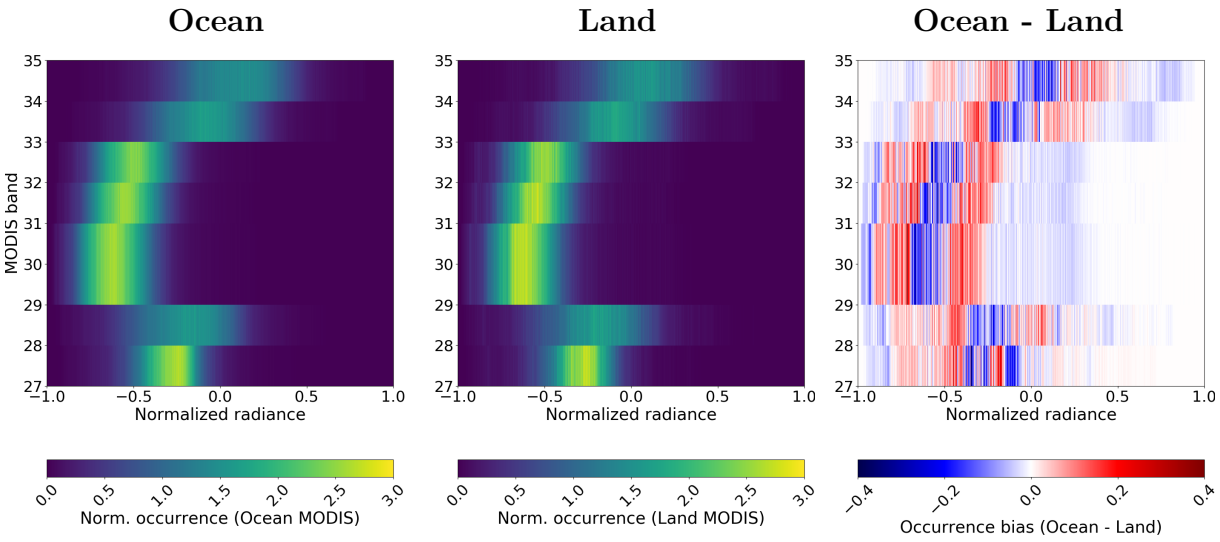


Figure 4.13: Normalized radiance-band histograms from the MODIS data set used for the CGAN. Left: Data from ocean locations. Middle: Data from land locations. Right: The difference between ocean and land locations.

4.2.5 Latitude variations

The test data set of CGAN was sorted into four different subsets, based on the latitude of the data points. The different intervals of latitudes are: 0° to 20°, 20° to 40°, 40° to 60° and 60° to 90°. The South latitudes are grouped with the corresponding North latitudes. There are 2615 scenes in the interval 0° to 20°, 2457 scenes in the interval 20° to 40°, 4871 scenes in the interval 40° to 60° and lastly, 4736 scenes in the interval 60° to 90° in the test set. The differences in the number of scenes in each latitude interval is most likely caused by the selection process of data for the final data set of CGAN. There is no direct correlation between latitude regions and regions of high topography of the Earth's surface on land. Regions such as the Tibetan plateau, which are located in a rather narrow latitude range near 30°'N, could be responsible for the decreased amount of scenes in the latitude regions closer to the equator.

The cloud vertical distributions at the different latitudes are shown in Figures 4.14 and 4.15. The distribution of the real data set in column 1 has clear variations for the different latitudes. For example, high normalized occurrences are present at higher altitudes in the first latitude region (0° to 20°), which is unsurprising since the altitude of the tropopause is higher near the equator (see Figure 2.3), meaning that clouds are also present up to about 16 km. The altitude of the tropopause, and therefore the top of the cloud fields, decreases with increasing latitude, which is also seen in the distributions of the first column.

An interesting feature that can be seen in the first row of Figure 4.14 is the distinct horizontal line in the distribution at approximately 5 km. This line is most likely a result from the high occurrence of precipitation in the tropics at these latitudes. It can also be discerned, although much less distinct, in the generated distribution. This points to the fact that the network identifies the presence of such features in the data and creates this also in some of the generated scenes.

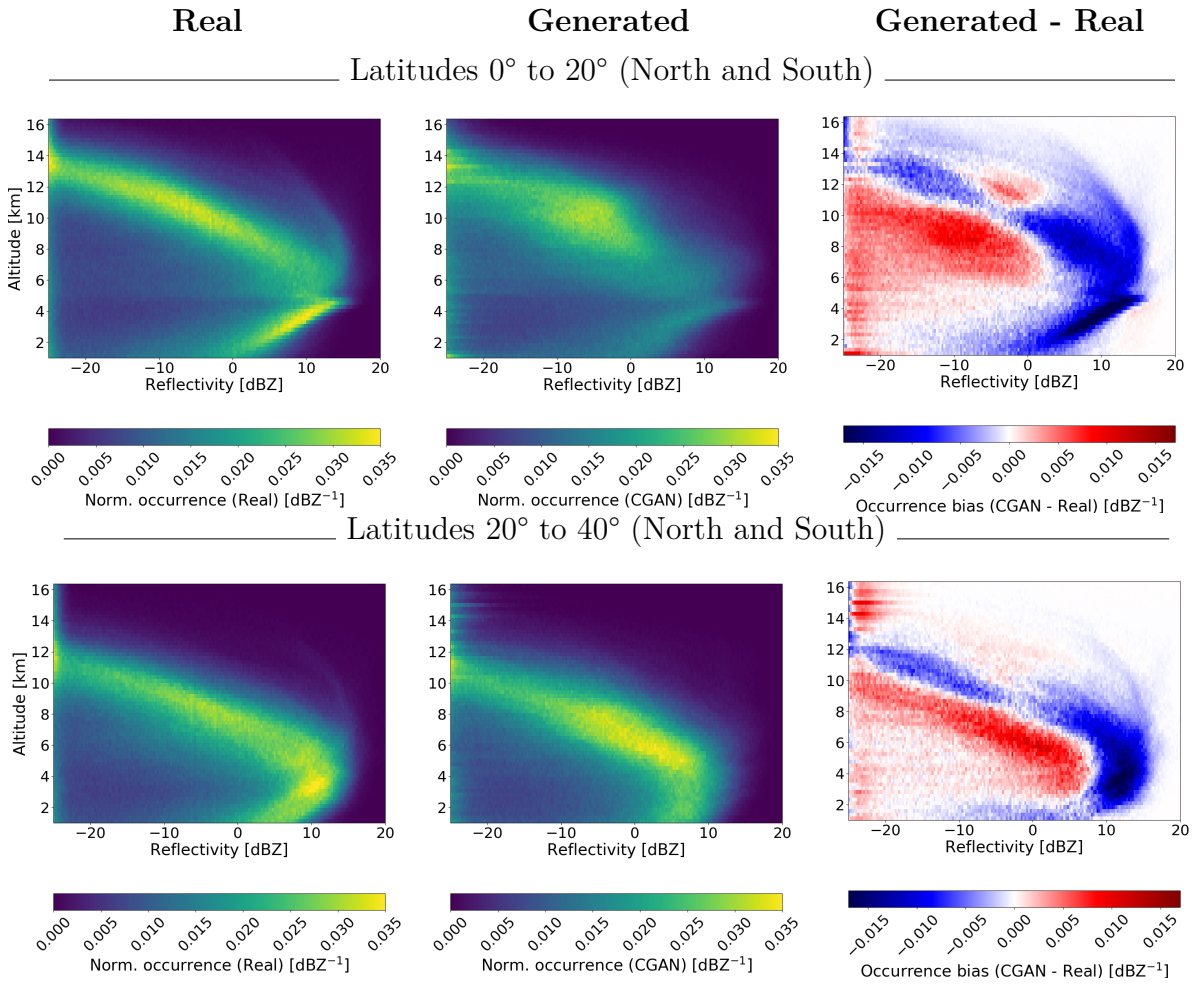


Figure 4.14: Normalized reflectivity-altitude histograms for CGAN in different latitude intervals. Column 1: CloudSat data of the test set for CGAN. Column 2: Generated scenes from CGAN. Column 3: Difference between the histograms from the real and generated data for each latitude interval. Row 1: Latitudes from 0° to 20°. Row 2: Latitudes from 20° to 40°. Note that each latitude interval includes both North and South latitudes.

The generated cloud fields at 0° to 20° appear to consist of clouds with reflectivities mainly in the region [-10 dBZ, 0 dBZ]. The average altitude of the clouds with reflectivity values in this region coincide with the altitude of the real scenes, but the spread of altitudes is larger for the generated scenes.

For all latitudes the CGAN manages to capture the overall outer shape of the occurrence histogram. This suggests that the vertical positions of clouds are accurately replicated. However, in the regions 0° to 20° and 20° to 40° the CGAN does not generate enough clouds with high reflectivities, which can be seen as the intense blue regions at high reflectivities in the rightmost column of Figure 4.14. On the other hand, in the other two latitude regions 40° to 60° and 60° to 90°, the CGAN generates too many clouds with high reflectivities, represented by the intense red regions at high reflectivities in the occurrence bias histogram of Figure 4.15. These two effects cancel each other out to some degree, which explains why the histogram of the generated scenes for all latitudes (Figure 4.5) shows a smaller difference from the real data set than what the individual latitude regions

do. However, since the number of scenes further away from the equator (very high and very low latitudes) is larger than the number of scenes in the latitude regions closer to the equator, the occurrence bias of the former mentioned regions has a much larger influence on the global distribution.

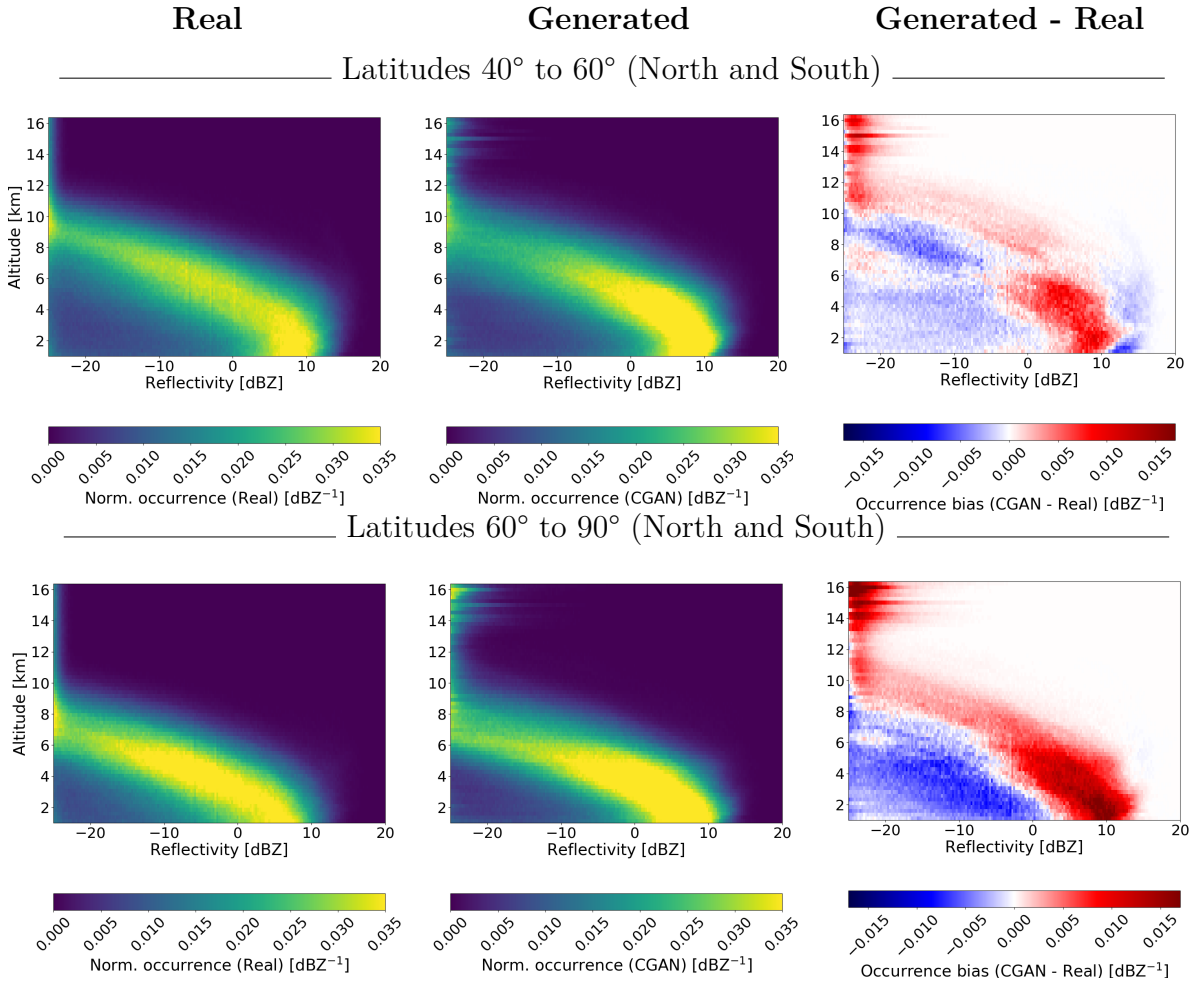


Figure 4.15: Normalized reflectivity-altitude histograms for CGAN in different latitude intervals. Column 1: CloudSat data of the test set for CGAN. Column 2: Generated scenes from CGAN. Column 3: Difference between the histograms from the real and generated data for each latitude interval. Row 1: Latitudes from 40° to 60°. Row 2: Latitudes from 60° to 90°. Note that each latitude interval includes both North and South latitudes.

4.2.6 Ice water path and cloud-top height

The IWP is considered to be one of the most important microphysical characteristics of clouds due to its key role in cloud radiative forcing, which is the effect clouds have on incoming solar radiation and outgoing thermal radiation from the Earth's surface. It is thus one of the important indirect features of the cloud structures that needs to be well preserved in the generated fields in order to be able to use generated scenes in climate models. Similarly, the CTH is important in determining the net warming or cooling effect of clouds, based on their altitude.

Histograms of IWP calculated from the real data set and generated scenes for CGAN are shown in Figure 4.16. The two distributions have similar appearances, which indicates that the CGAN captures the overall statistics of this derived physical quantity rather well. The peak of the distribution is located at almost the same IWP in both distributions, $75\,\mathrm{gm}^{-2}$ for the real scenes and approximately $50\,\mathrm{gm}^{-2}$ for the generated scenes. The real CloudSat data set has a maximum IWP of $13\,800\,\mathrm{gm}^{-2}$ whereas the generated scenes provide a maximum IWP of $25\,600\,\mathrm{gm}^{-2}$. Values above $4500\,\mathrm{gm}^{-2}$ are not presented in Figure 4.16 since the frequency of these tends towards zero in both distributions. The higher maximum value for the IWP of the generated scenes indicates a larger spread of IWP, which also presents itself as the lower maximum frequency in the histogram of generated scenes. The higher maximum IWP also suggests that the reflectivities in the generated scenes are sometimes higher, which we saw in the difference histogram for CGAN in Figure 4.5 for low altitudes (1 km - 5 km). However, this altitude region is expected to occur below the freezing level in many scenes and should thus not be part of the calculation of the IWP. The increased maximum IWP could also be because the generated clouds are thicker vertically than the real clouds. On the other hand, the peak of the distribution is shifted towards lower values of IWP for the generated scenes, which would mean that the average reflectivities and/or vertical thickness is smaller for the generated scenes. It is difficult to say exactly which of these two effects is the most important.

The calculation of the IWP is based on the temperature data from ECMWF for the test set. This provides an altitude for the 0 °C level. It is assumed that this level should occur at the same altitude in both the real and generated scene. It is, however, not unreasonable to believe that the generated scenes will have this level at slightly different altitudes from the real scenes, especially considering the fact that the network is not always able to generate features such as the bright line at the freezing level. If the 0 °C altitude is shifted in the generated scene, the vertically integrated IWC will contain additional bins below, or lack bins above the freezing level. This is expected to affect the calculated IWP quite a lot since the reflectivities are often high near this level. Above approximately 5 km the frequency of very high reflectivities (>5 dBZ) is clearly lower in the generated scenes than in the real scenes for CGAN (which is illustrated in the reflectivity-altitude occurrence bias in Figure 4.5). This might explain the lower average IWP of the generated scenes.

From Figure 4.16 it becomes clear that clouds with low IWP are much more common than clouds with higher IWP. The CGAN is able to reproduce this distribution rather well. In general, the IWP of the generated clouds is slightly higher than the IWP of the real data set as can be seen from the difference histogram in Figure 4.16 where the frequency of low IWP clouds is much higher for the real than the generated scenes. The generated scenes also contain more clouds with IWP between $500\,\mathrm{gm}^{-2}$ and $3000\,\mathrm{gm}^{-2}$ than the real scenes do. There is a discrete peak near $0\,\mathrm{gm}^{-2}$ in both the generated and real distribution, but with slightly higher occurrence frequency in the real data set. This peak is slightly lower than the peak at $75\,\mathrm{gm}^{-2}$ in the real distribution, but significantly higher than the other peaks of the generated distribution. One interpretation of this result is that the CGAN generates more scenes with completely cloud free regions above the freezing level than are present in the real scenes. In extension, this could mean that a higher frequency of CTH would be present at the typical altitudes of the freezing level and below this. This can, in fact, be seen in Figure 4.19 of the CTH for CGAN, especially at altitudes 4.5 km to 6 km.

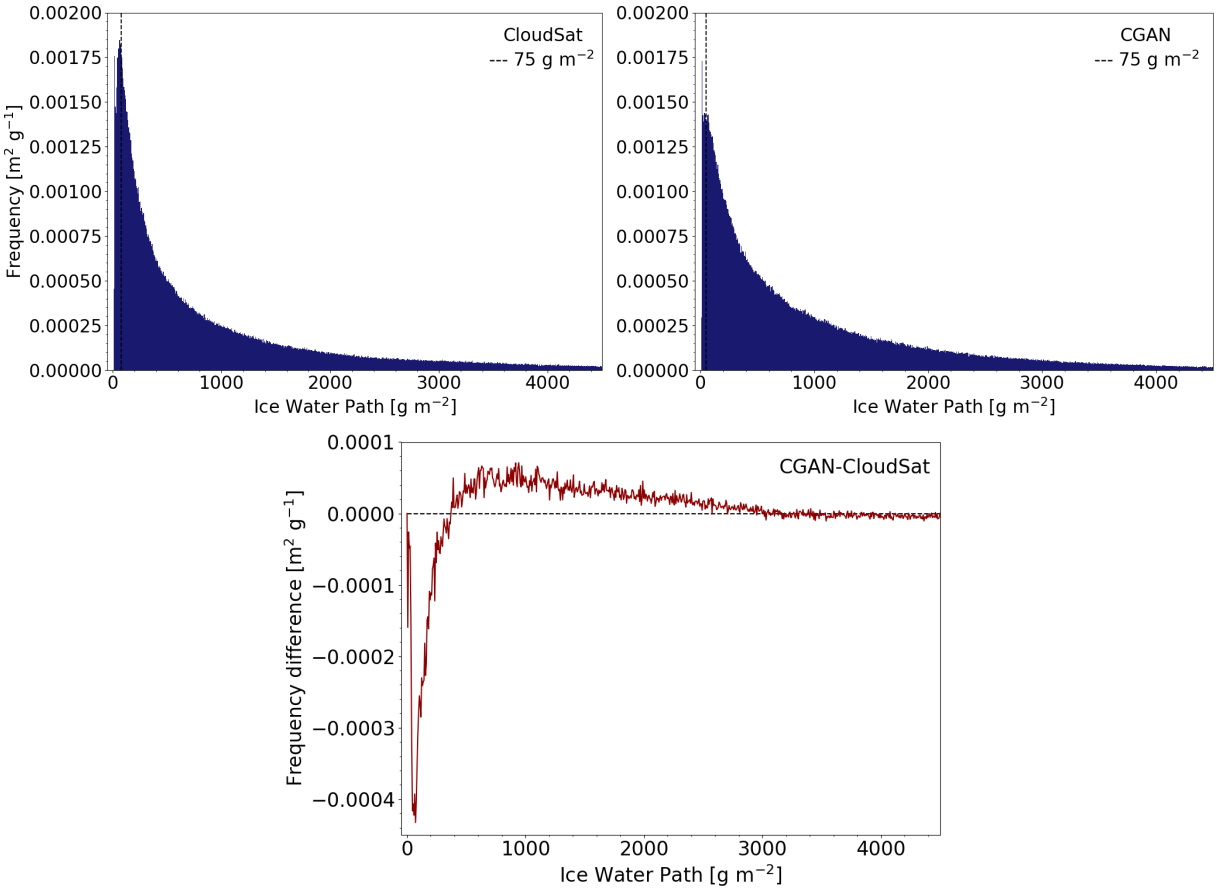


Figure 4.16: Ice water path for the real data set (left) and generated scenes with CGAN (right). The distribution of the real data set has an average of 1080 gm^{-2} and the generated data presents an average near 1050 gm^{-2}. The most frequent value of IWP is 75 gm^{-2} for the real scenes, marked by the dashed line in both distributions. The most frequent value in the distribution of the generated scenes is just below this. The difference between the two distributions is also shown in the lower image. The difference is largest for small values of IWP and tends towards zero for values above 3000 gm^{-2} approximately.

Histograms of the CTH calculated from both the real and generated scenes for GAN, IAAFT and CGAN are presented in Figures 4.17, 4.18 and 4.19 respectively. The overall distribution of generated CTH is similar to the CTH of the real data set for all models, which indicates that all three models are rather good at recreating this physical quantity. It should be noted, however, that it is only the highest cloud top that is considered in these distributions, so they are not fully representative of the case where several layers of clouds with individual CTH are present.

The IAAFT method is undoubtedly best at replicating the CTH since its distribution is almost identical to that of the real scenes. Both GAN and CGAN tend to generate too many cloud tops at high altitudes (above 10 km). An especially interesting feature of the CTH distribution from the CGAN is the discrete peaks near 14.5 km and 16.4 km, which are not present in the original distribution. These are most likely caused by the very thin clouds at the top of many generated scenes by CGAN. There is no clear reason why they always occur near the same distinct altitudes, but may be caused by a mode collapse where the network learns that this kind of cloud scene is acceptable to the discriminator.

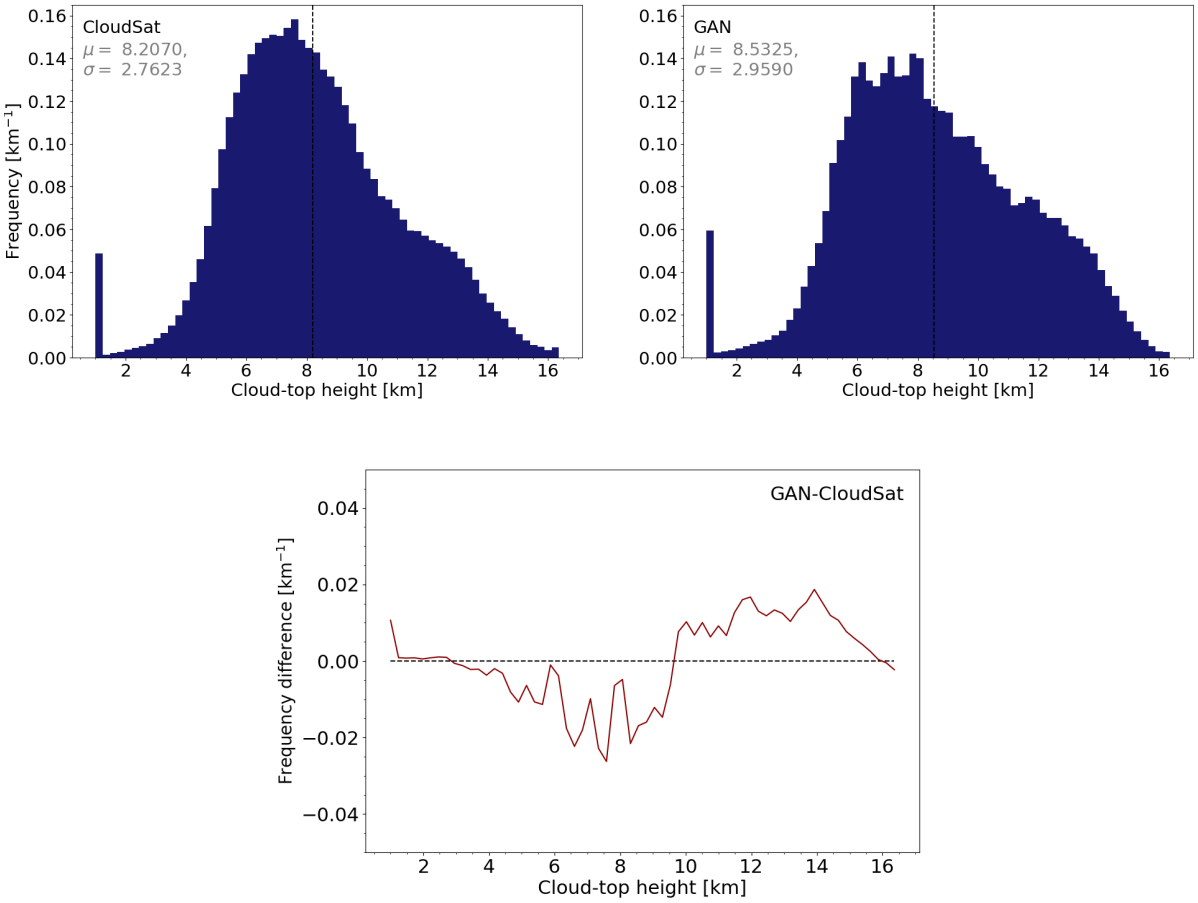


Figure 4.17: Cloud-top height (CTH) for the real data set (left) and generated scenes with GAN (right). The distribution of the real data set has a peak near 7.5 km. The generated data set shows two peaks around 7.5 km and 10.5 km. The distinct peak at 1 km in both histograms corresponds to locations without clouds. The dashed vertical line marks the average value of the CTH in each distribution. The difference between the two distributions is also shown in the lower image. The horizontal line marks zero difference.

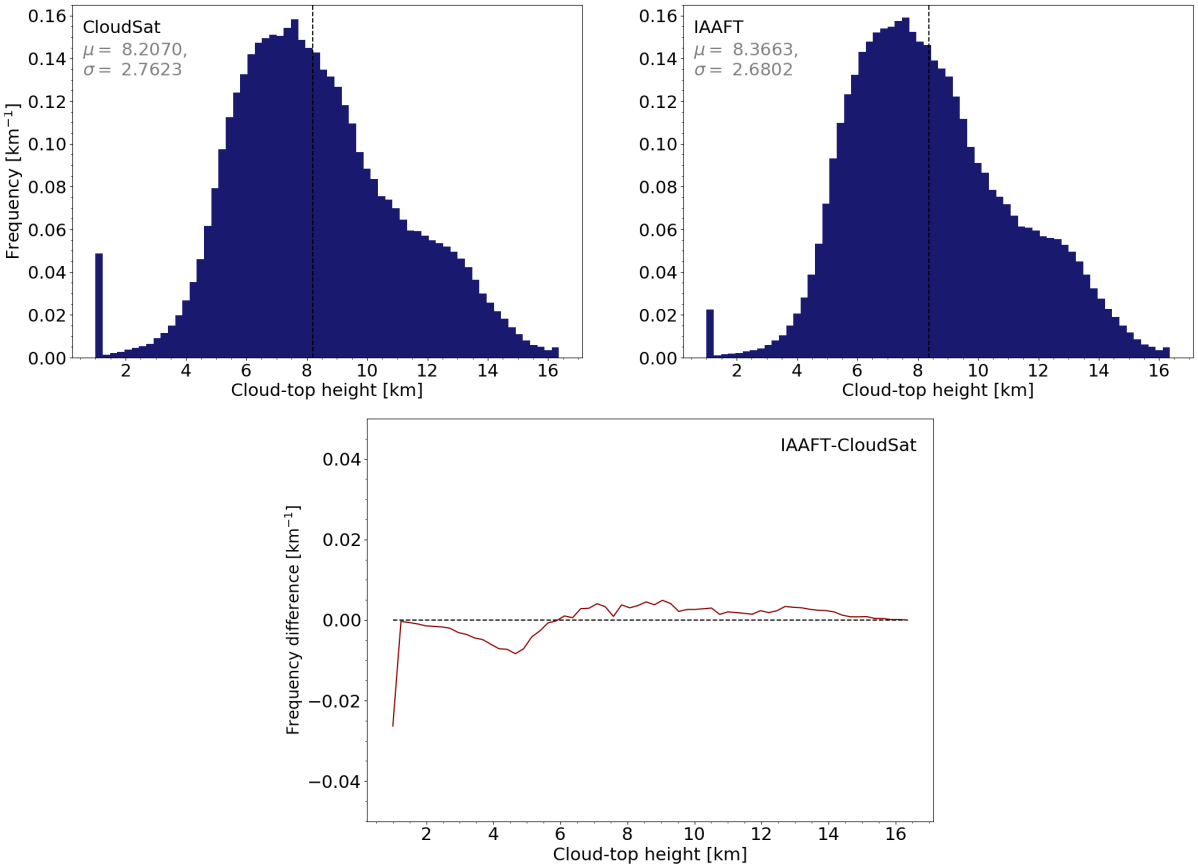


Figure 4.18: Cloud-top height for the real data set (left) and generated scenes with IAAFT (right). The distribution of both the real data set and the generated scenes has a peak near 7.5 km. The dashed vertical line marks the average value of the CTH in each distribution. The peak at 1 km in both histograms corresponds to locations without clouds. The difference between the two distributions is also shown.

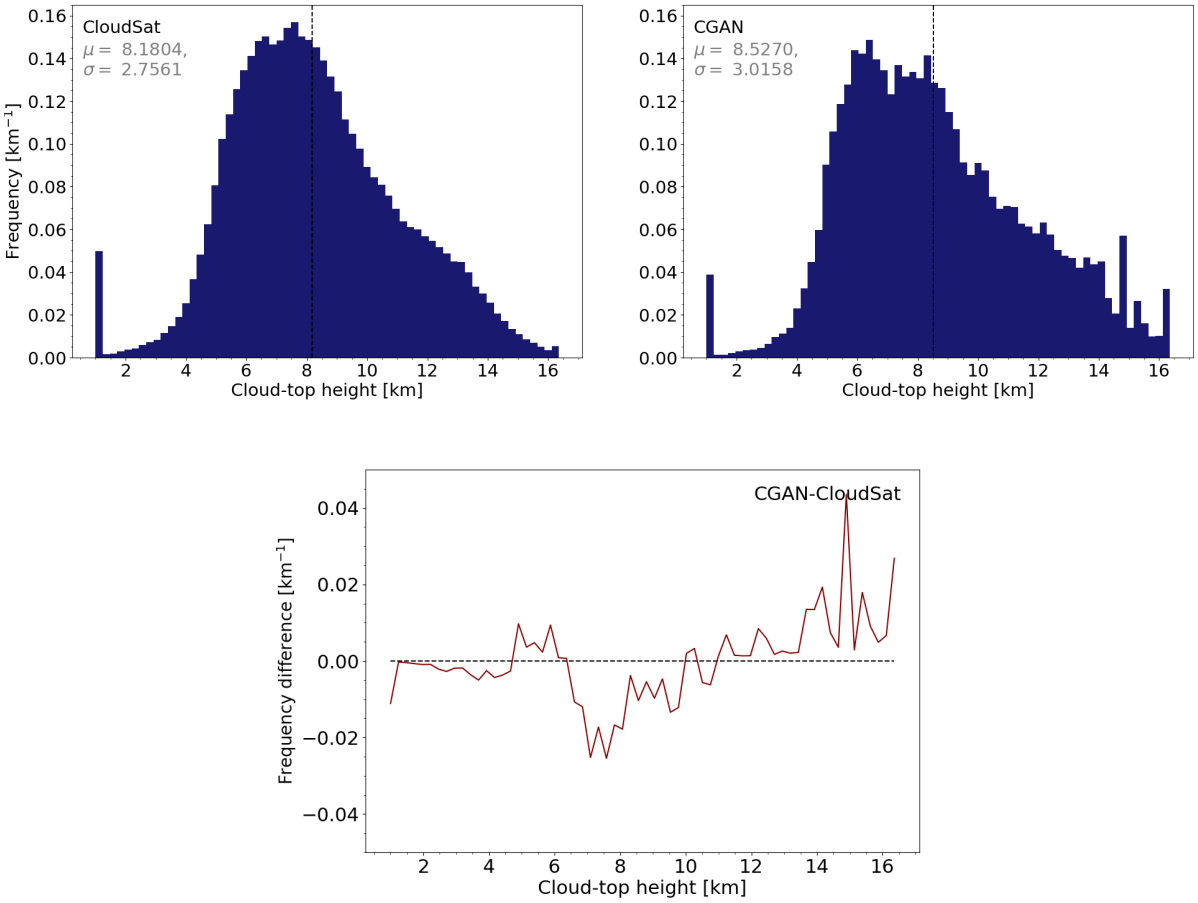


Figure 4.19: Cloud-top height for the real data set (left) and generated scenes with CGAN (right). The distribution of the real data set has a peak near 7.5 km. The generated data set shows peaks around 6 km and 8.5 km. The peak at 1 km in both histograms corresponds to locations without clouds. The difference between the two distributions is also shown.

4.2.7 Spatial autocorrelation

The average autocorrelation calculated at each altitude is visualized in Figure 4.20, where row 1 contains the average autocorrelation for the real data set, row 2 contains the average autocorrelation for the generated scenes, and finally, row 3 contains the difference between the average autocorrelation for the generated and real data set. Examples of the autocorrelation in individual scenes can be found in Section A.2 in the Appendix. The bottom left image of Figure 4.20 shows that GAN is able to capture the autocorrelation of the real distribution well at altitudes between 12 km and 16 km. In this region there is almost no difference between the real and generated scenes, most likely because there is typically only noise here in most scenes, which is easy to replicate. At lower altitudes GAN is a bit less successful, in the interval 7 km to 12 km GAN has lower autocorrelation than the real data set, and in the interval 0 km to 7 km GAN has higher autocorrelation than the real data set. In general, a lower spatial autocorrelation at small distances is an indication that nearby pixel values are more dissimilar, and on larger scales this could indicate a broader range of values (reflectivities). By similar reasoning, a higher autocorrelation at larger distances than the ones corresponding to nearest neighbouring pixels suggests a more narrow range of pixel values in the data set. For GAN this means that cloud fields between altitudes 0 km and 7 km typically include a more narrow range of reflectivities in each scene than the real cloud fields at the same altitudes, and the opposite between altitudes 7 km to 12 km.

The CGAN shows similar results as GAN, but it does not have as distinct regions as GAN, when it comes to the difference to the real data set. The network tends to underestimate the autocorrelation in almost the entire scene, as can be interpreted from the large blue region in the occurrence bias histogram for CGAN in Figure 4.20. At altitudes above approximately 12 km the occurrence bias of the autocorrelation for CGAN shows regions with positive bias. This is believed to be a result of the thin clouds that are sometimes generated by CGAN, where there is most often only noise in the real scenes.

With just a quick look it becomes apparent that difference between the autocorrelation in the real and generated scenes is much smaller for GAN and CGAN than it is for IAAFT. At higher altitudes, IAAFT performs well, but at altitudes below 12 km it displays dissimilarities to the real data set. IAAFT has a lower autocorrelation at short distances and a higher autocorrelation at long distance than is present in the real data set. It is not even able to capture the autocorrelation at very short distances, corresponding to adjacent pixels. This is speculated to be because of the discontinuities that could be seen in examples of scenes generated by IAAFT (see Figure 4.3). There are abrupt changes in radar reflectivity between adjacent pixels in the IAAFT generated scenes, compared to the scenes from the real data and GAN and CGAN, where there are smooth transitions in reflectivity between nearby pixels.

A major result is the fact that the intensity of the difference between the average autocorrelation of the generated and the real data set is much greater for IAAFT than GAN and CGAN. This points to the fact that IAAFT does not capture the autocorrelation of the real data set as well as GAN and CGAN. Since the IAAFT algorithm is based on preserving the autocorrelation of the template, the result that the local autocorrelation appears to be better preserved by the GAN and CGAN is very promising.

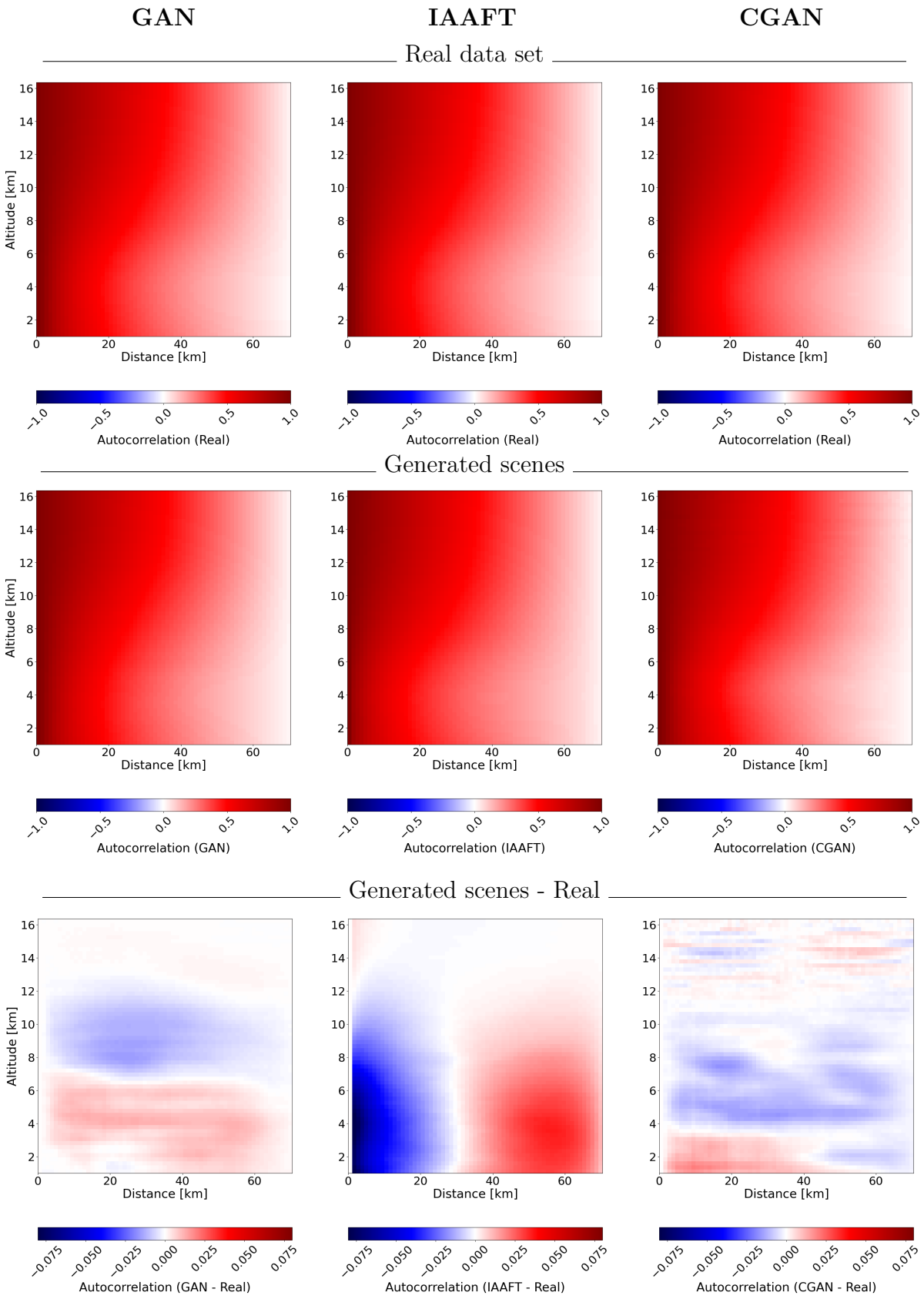


Figure 4.20: Average spatial autocorrelation calculated at each altitude. The first column contains histograms for GAN, the second column shows histograms for IAAFT and the last column shows histograms for CGAN. Note that the real data set is the same in the leftmost and middle histogram of the first row.

Figure 4.21 depicts the distribution of spatial autocorrelation of CloudSat scenes given by Moran's I from the real data set as well as scenes generated with the GAN. The distributions of Moran's I for IAAFT and CGAN can be found in Figures 4.22 and 4.23. Moran's I is a more global measurement of the spatial autocorrelation in entire scenes than the results shown in Figure 4.20 which are calculated more locally for each altitude in the data set.

The distributions of Moran's I for IAAFT and real scenes in Figure 4.22 are similar for both cases, with mean values at $\mu_I \approx 0.9408$ (CloudSat) and $\mu_I \approx 0.9278$ (IAAFT). This suggests that the spatial autocorrelation and thus the clustering of similar reflectivities is high for both data sets, albeit somewhat lower for IAAFT. The IAAFT generated scenes also present a larger standard deviation for I, which points to a more random distribution of reflectivities in the scenes. This is in line with what could be seen in Figure 4.20. The difference histogram of Figure 4.22 also reveals that the difference in the distributions for high values ($I > 0.80$) is large for IAAFT, with much less scenes in the IAAFT generated data set providing high global autocorrelation.

The scenes generated with the GAN shows a wider spread in spatial autocorrelation than both the real scenes and IAAFT scenes, measured by the larger standard deviation which is $\sigma_I \approx 0.0390$ for the real data set, $\sigma_I \approx 0.0446$ for IAAFT and $\sigma_I \approx 0.0538$ for GAN. This means that the distribution of reflectivities across a scene tends to be more random in a higher fraction of the scenes generated by GAN. However, if Gaussian curves were to be fitted to the distributions presented in Figures 4.22 and 4.21, the mean μ_{Gaussian} of such a normal distribution would appear to be lower for IAAFT compared to the CloudSat template, but closer to 1 for GAN. This points to the scenes generated by GAN having a higher spatial autocorrelation on average. The GAN also generates some scenes with a much lower value of Moran's I, which presents itself as a lower value of I_{min}. This means that some generated scenes have a distribution of reflectivities which is much more random than the distribution in any of the scenes in the real data set. These scenes are rather few however, since the frequency of scenes with index I below 0.70 is almost zero in both the real and generated distribution. Interestingly, the distribution of generated scenes from GAN also show a larger peak near $I = 0.99$ than in the real distribution, meaning that the clustering of similar reflectivity values is somewhat higher in the generated scenes.

The distribution of Moran's I for the scenes generated with CGAN is presented in Figure 4.23. The average value of I as well as the standard deviation are very close to the values of the real data set. In this aspect, the difference between real and generated scenes is smaller than for GAN and IAAFT. Both the maximum and minimum values of I are also close the values of the real data set, but not as close as IAAFT. The CGAN shows the least absolute difference between the real and generated distribution of I, as given by the last image of Figure 4.23 compared to the corresponding image in Figures 4.21 and 4.22, which suggests that the CGAN method is best equipped to model the global distribution of autocorrelation found in the real data set. This is in agreement with what was found also for the local autocorrelation in Figure 4.20.

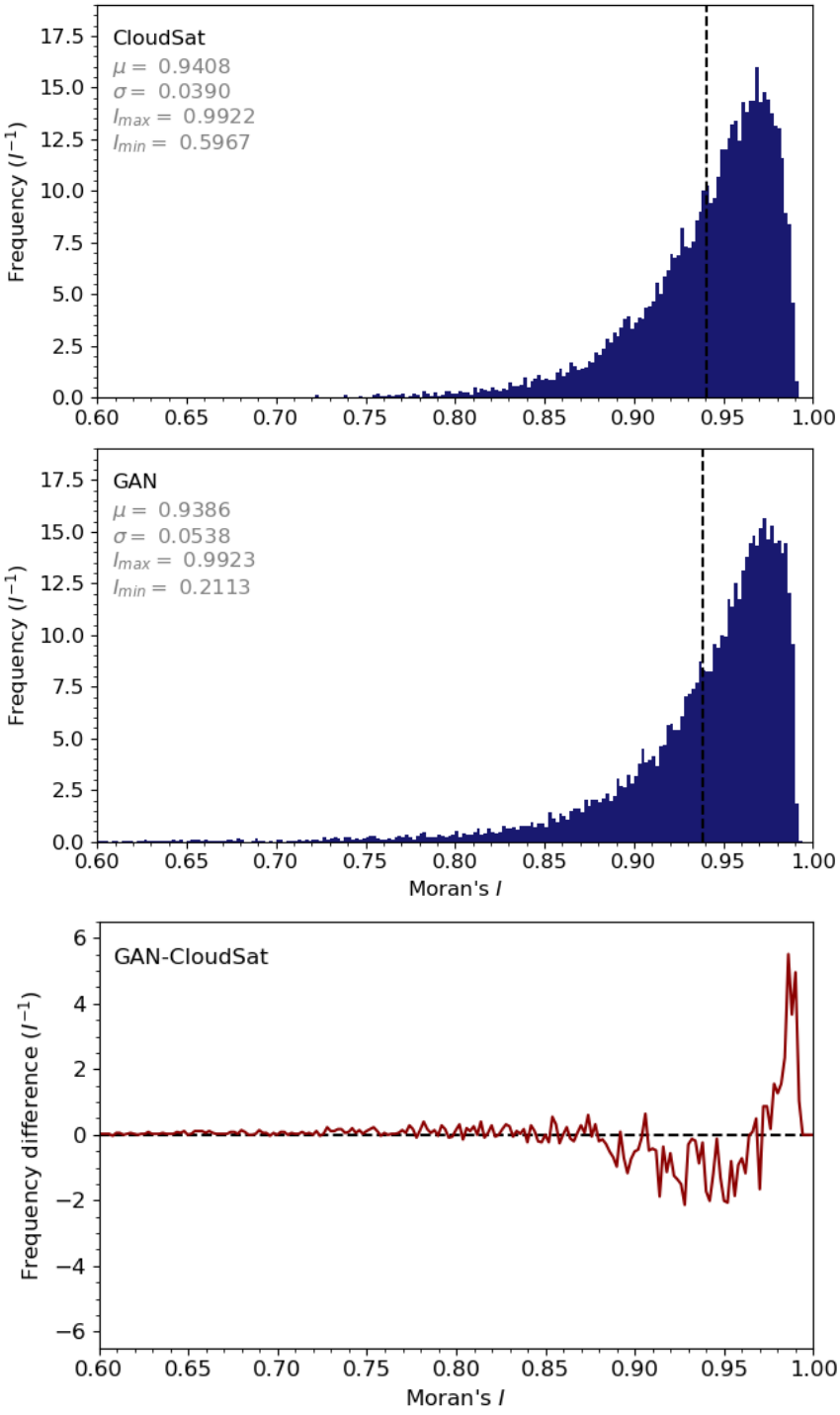


Figure 4.21: Spatial autocorrelation given by Moran's I. CloudSat data from the test set ($\sim$ 17400 scenes) for GAN and generated data from the GAN. The μ values represent the mean of each distribution and the standard deviation is given by σ. I_{max} and I_{min} are the maximum and minimum index I, respectively, for each distribution. The distributions tend towards normal distributions with the right flank cut off. Note that the mean values are not located at the peaks of the distributions. The mean values are marked with the vertical dashed line for each distribution. The peak of the distribution occurs around $I = 0.97$ for the real data set and near $I = 0.98$ for the generated scenes.

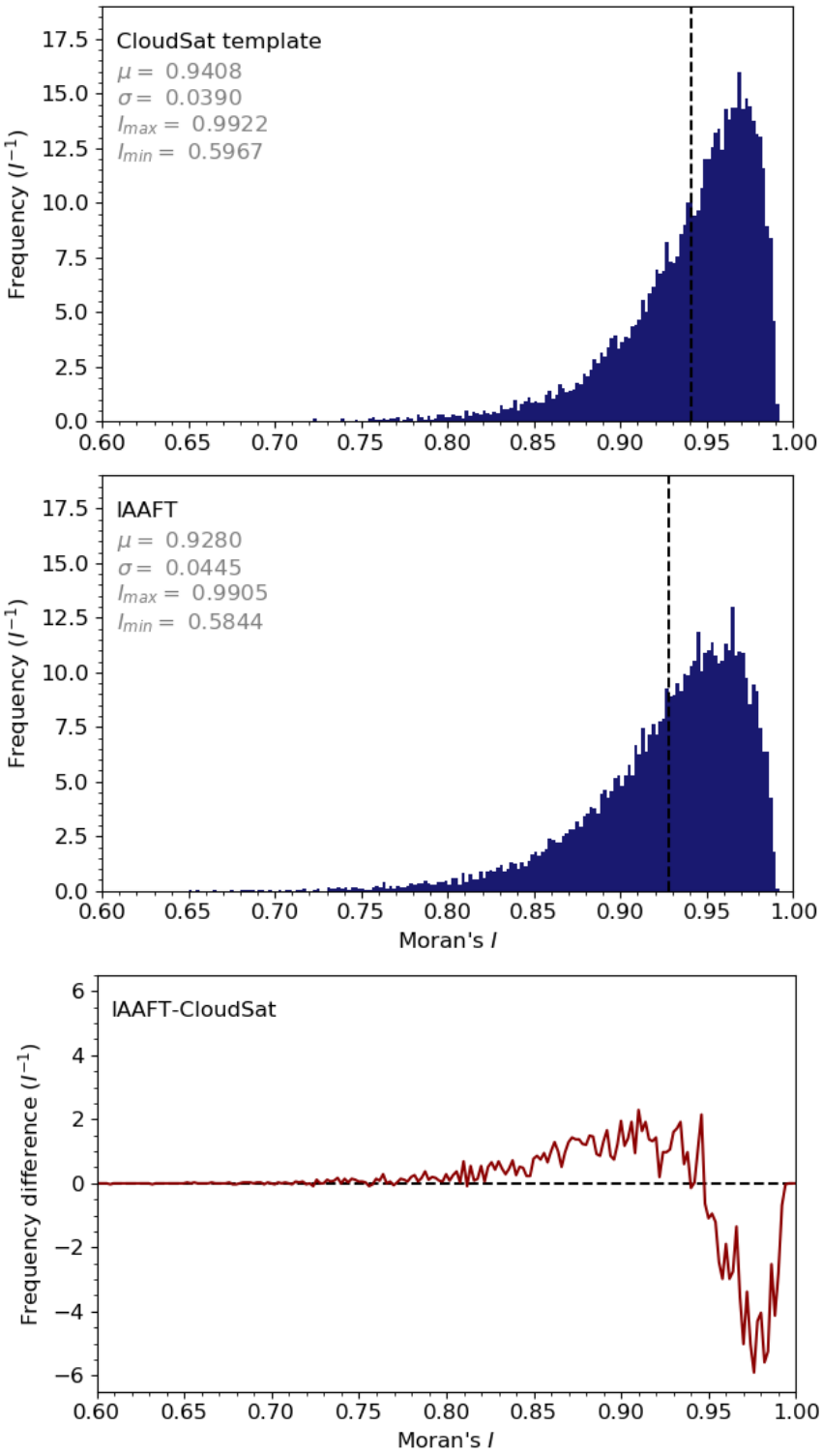


Figure 4.22: Spatial autocorrelation given by Moran's I. Data from the test set for GAN ($\sim$ 17400 scenes) and generated data from IAAFT. Values μ, σ, I_{max} and I_{min} as in Figure 4.21. Mean values μ marked with vertical dashed lines. The peak occurs around $I = 0.97$ for both distributions.

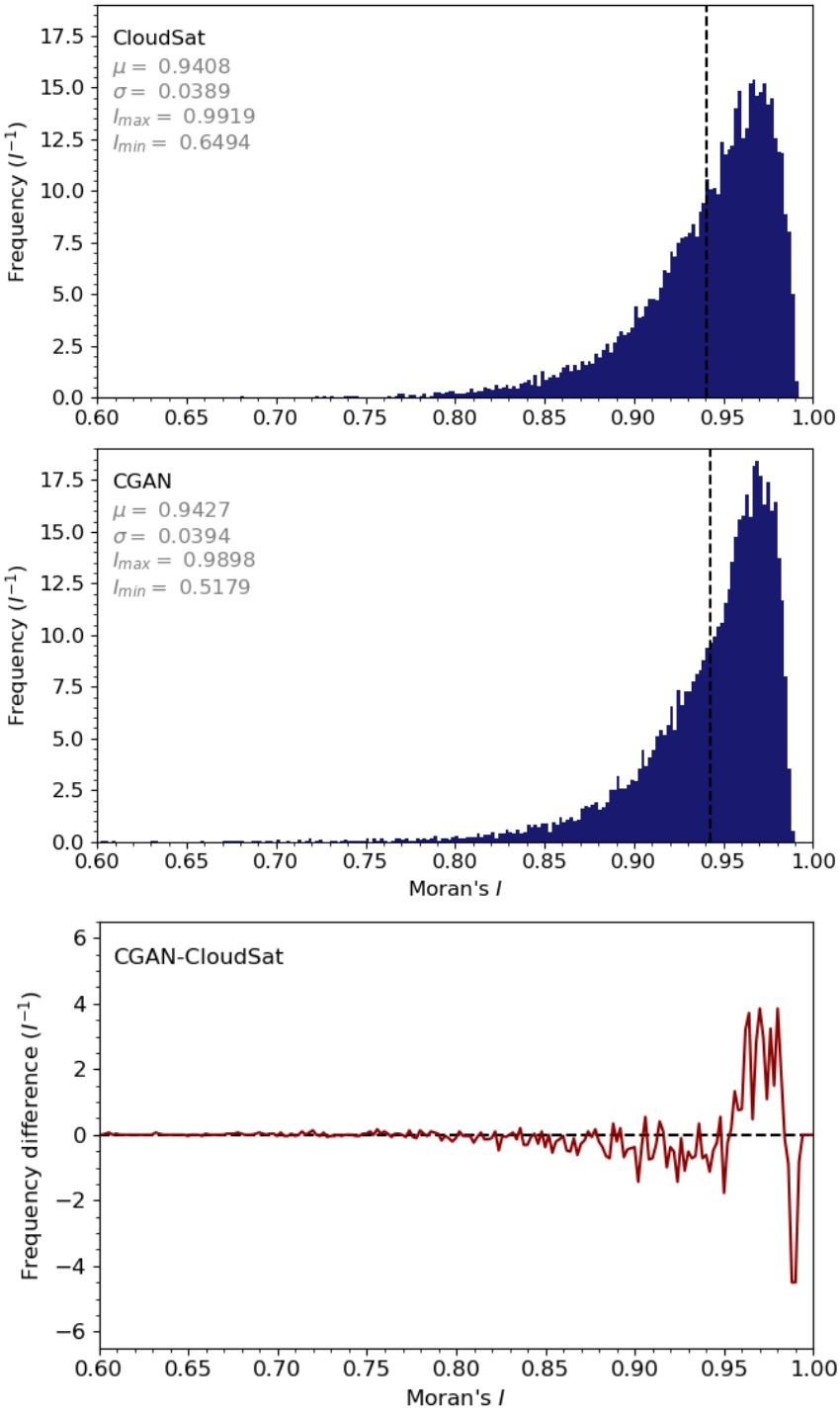


Figure 4.23: Spatial autocorrelation given by Moran's I. Data from the test set for CGAN ($\sim$ 14500 scenes) and generated scenes. Values μ, σ, I_{max} and I_{min} as in Figure 4.21. Mean values μ marked with vertical dashed lines. The peak occurs around $I = 0.97$ for both distributions.

4.3 Discriminator classification

Testing how the discriminator classifies different data sets is a way of testing an important part of the generative adversarial network. This section aims to test the discriminators of GAN and CGAN on relevant data sets, and discuss how they perform. The results in this section are to be considered as indirect results of the different parts of the networks, and should help with the understanding of why the network works as it does. It should not be used to draw conclusions on the results of the network as a whole, but instead it can indicate flaws or strengths of the network.

4.3.1 Generative adversarial network

In Figure 4.24 histograms of how the discriminator of the GAN classifies different data sets are visualized. This classification is performed by the discriminator of GAN trained for 400 epochs. The discriminator classification of later epochs in the training can be seen in Section A.3 in the Appendix.

From Figure 4.24 it becomes clear that the discriminator is able to classify generated scenes from both CGAN and IAAFT as fake with high certainty. One explanation to why the generated scenes from CGAN are classified to be fake to a larger extent than the generated scenes from GAN is that the CGAN generates scenes from a slightly different distribution than the one the discriminator is used to seeing for GAN, since the two have different data sets. This is not the case for IAAFT, since these scenes are generated from the test set for GAN, so the feasible explanation here is that the generated scenes simply do not resemble the real scenes well.

The training set and the test set from the real data set of GAN are classified differently by the discriminator. The discriminator classifies the training set as being true, but the test set is generally classified as false. This fact implies that there has been an overfitting to the training data, since the test and training set are sampled from the same distribution. The discriminator classifies the data set generated by GAN to be true to a higher extent than the test set. This could imply that the generator creates more realistic scenes than the test set. It is however more likely that the network is overfitted on the training set, i.e. the discriminator learns to recognize exact scenes in the training set, and that the GAN is able to recreate scenes from the training set more than actually capturing the distribution of the test and training set. This phenomenon is known as mode collapse. The overfitting of the network does not decrease with a longer training beyond this point, but it is more likely that it actually increases through the later epochs, as could be seen by the worse results of later epochs in Section 4.2.2.

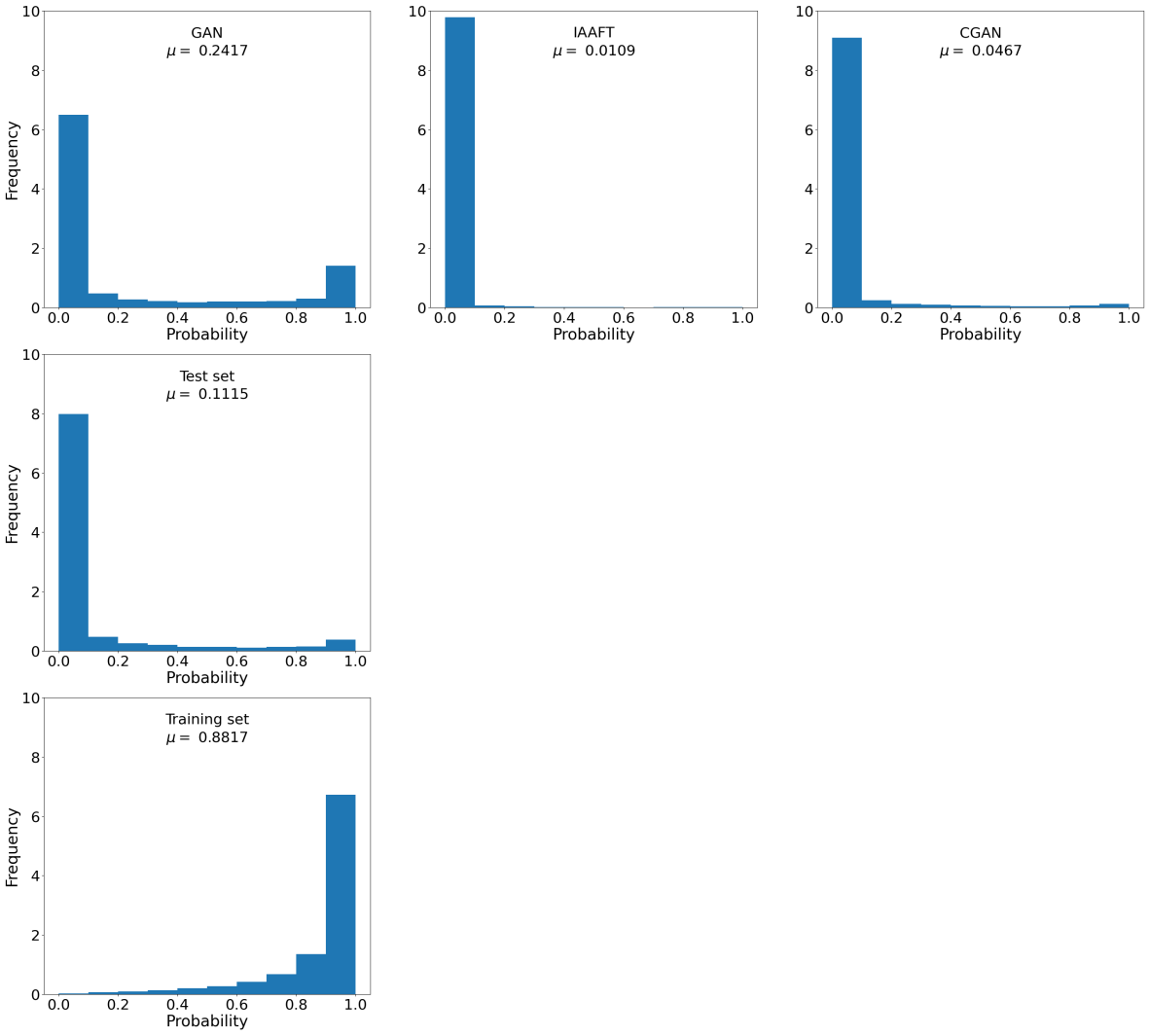


Figure 4.24: Histograms of different data sets classified by the discriminator of GAN at epoch 400. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

4.3.2 Conditional generative adversarial network

The resulting histograms of the discriminator of CGAN at epoch 3500 applied to different data sets are visualized in Figure 4.25. The discriminator classification of earlier epochs in the training can be seen in Section A.4 in the Appendix. The discriminator at epoch 3500 has a discrepancy between how it classifies the training set and the test set, however it is less severe than that of the GAN since it classifies the test set as true to a higher degree, which is shown in Figure 4.24. This means that there is probably some overfitting present, but perhaps not as severe as with the GAN. The generated scenes from CGAN are also classified as false in a higher extent than the test set, which is to be expected. The fact that the training set is almost entirely classified as true and the generated scenes are almost entirely classified as false implies that the discriminator performs much better than the generator, since the generator does not manage to fool the discriminator on many occasions. This is a sign that an uneven training of the discriminator and the generator could be beneficial, since the performance of the discriminator improves at a much higher rate than the performance of the generator, which could also be seen from the loss graphs

in Figure 4.9.

Both the GAN and the CGAN could benefit from a less complex network or a larger data set in order to avoid overfitting and mode collapse during the training. Another approach would be to include a validation set from the data set in order to test and stop the training before overfitting of the network occurs. The major problem with this approach is that it is not clear how to formulate a stop condition for the training.

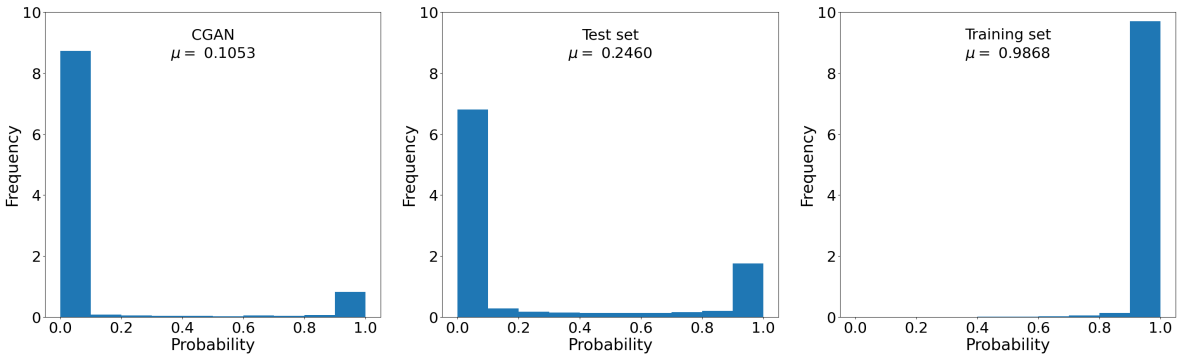


Figure 4.25: Histograms of different data sets classified by the discriminator of the CGAN. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

4.4 Error sensitivity

The minimum, maximum and average uncertainty of the MODIS bands are summarized in Table 4.1. It should be stated, once more, that the uncertainties are already present in the MODIS data in the form of measurement noise. This test is therefore only a indication on the effect of noise of different magnitudes on the output of the network.

Table 4.1: Calculated uncertainties for the MODIS bands.

MODIS Band	Min uncertainty (%)	Max uncertainty (%)	Average uncertainty (%)
27	0.5000	7.821	0.5895
28	0.5000	4.744	0.5184
29	0.5000	7.821	0.5203
31	0.3750	1.019	0.3750
32	0.3750	1.681	0.3750
33	0.5000	21.26	0.7570
34	0.5000	1.745	0.5003
35	0.5000	21.26	0.7081

In Figure 4.26 the results of the network with three levels of noise added to the MODIS input are visualized as reflectivity histograms. Row 2 and 3 show that the difference histograms with minimum uncertainty and average uncertainty are very similar, and also similar to the output of the network without additional added noise (see Figure 4.5). This implies that noise of this magnitude does not have a big effect of the output of the network, at least on average. In extension, this could imply that the uncertainty already

existing in the MODIS data is not of significant impact on the performance of the network. The middle column of Figure 4.26 shows histograms from the output of the network with maximum added uncertainty to the MODIS input. A fascinating discovery is that the performance of the network with this input appears to be better than the performance for the input without added uncertainty. The occurrence bias between the generated and real scenes is much smaller on average for this constellation than any other degree of uncertainty analysed.

In the maximum uncertainty case, the magnitude of the added uncertainty is different for the various MODIS bands. Two of the bands have rather large magnitudes of noise ($\sim 20\%$), making the original shape of the input difficult to discern, as can be seen in the third row of Figure 4.27 which shows example scenes from the CGAN with added uncertainty (maximum uncertainty) of the MODIS input. The magnitude of the noise added to the other six bands is significantly smaller. This suggests that all eight bands may not be necessary as input to the network, at least not in the form of pure measurements. Since there is no unambiguous reason why this input gives the result with the closest similarity to the real data set, it would be beneficial to perform more tests with noise added to the input data in the future. The number of degrees of freedom in the input data may be smaller than what is used throughout this work, so this should also be analysed further. The example scenes of the CGAN with maximum added uncertainty seen in Figure 4.27 present similar features as for the example scenes without additional uncertainty (Figure 4.4). The network appears to be equally good at recreating the CTH and the IWP for individual scenes, with or without added noise. If the same uncertainty is added to the MODIS input for all instances of the generated version the same degree of variability is found as for the examples in Figure 4.4. When different uncertainties (with the same magnitude) are added to the MODIS input for each version of the generated scenes, the variability of the scenes is slightly increased, as can be seen in rows 4 through 7 of Figure 4.27, especially for thicker clouds. The variance is also clearly distinguishable in the calculated IWP for the different generated scenes of each real scene. Based on this analysis it could be beneficial to further investigate adding noise of different magnitudes to the input data of the networks, perhaps also during training.

4.5 Applicability on other data sets

The performance of the two networks for other data sets than the one used for training is presented in Figure 4.28 and A.17 for data from the year 2016.

The histograms presented in Figure 4.28 are to be compared to Figure 4.5. For GAN the difference histogram has a slightly higher intensity for the 2016 data set, compared to the difference histogram from 2015. This could imply that the GAN is overtrained, and that it has learned attributes of the 2015 set, that are not present in the 2016 set. The CGAN difference histograms look similar for 2016 and 2015, which is an indication that the CGAN is able to be used on data sets consisting of other years than 2015. However, the real data set of 2016 shows a very similar distribution to the real data set from 2015. It is possible that years further apart have larger differences in distribution from 2015, and the performance of the network for these years is difficult to predict.

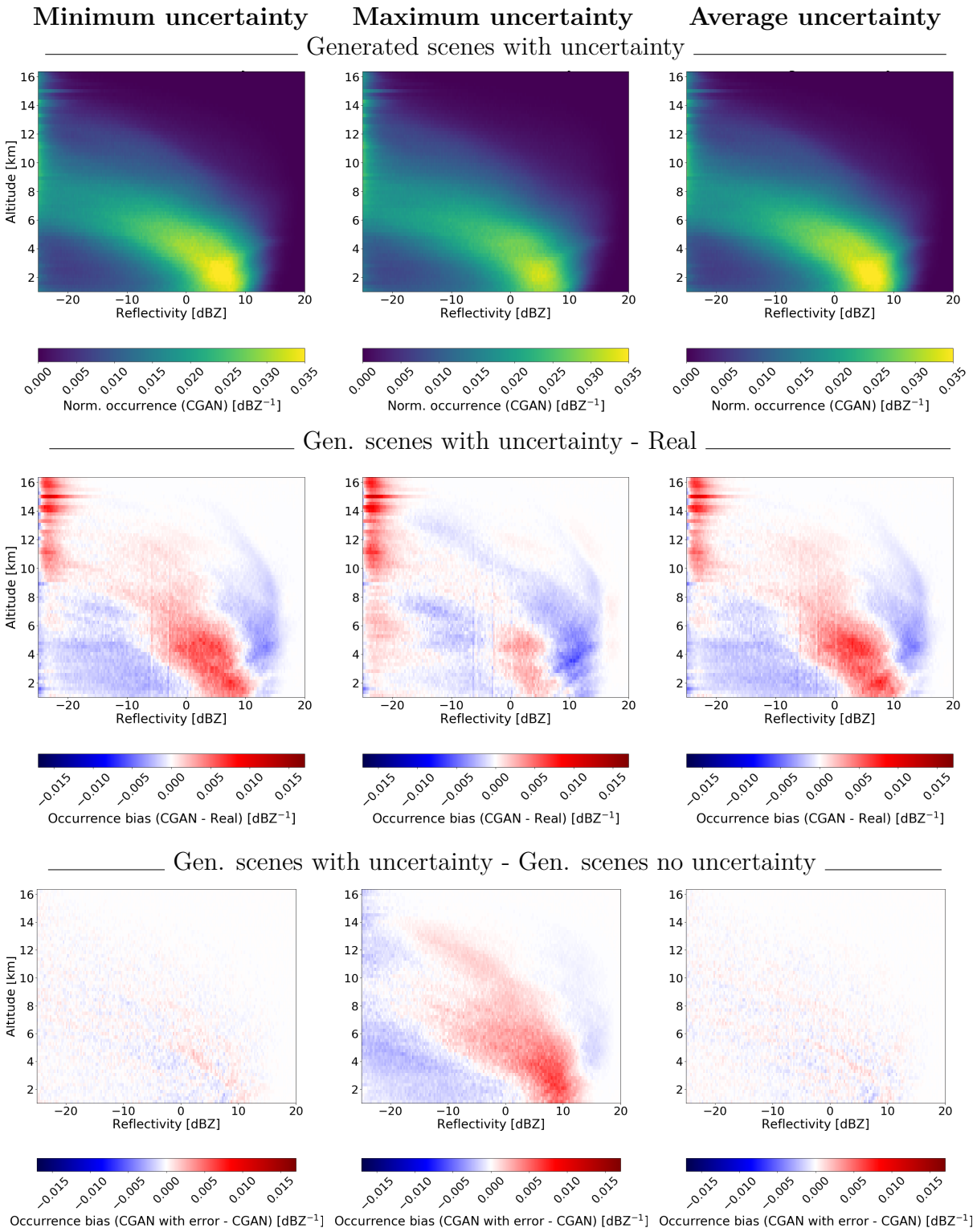


Figure 4.26: Normalized reflectivity-altitude histograms for added error to the MODIS data. Row 1: Generated scenes from CGAN with varying amounts of error added to the MODIS input. Row 2: Occurrence difference between generated scenes with added error and the real data set. Row 3: Occurrence difference between the generated scenes with added error and scenes generated without error in the MODIS data. The left column uses the minimum uncertainties for MODIS presented in Table 4.1, the middle column uses the maximum uncertainties and the right column uses the average uncertainties from the same table.

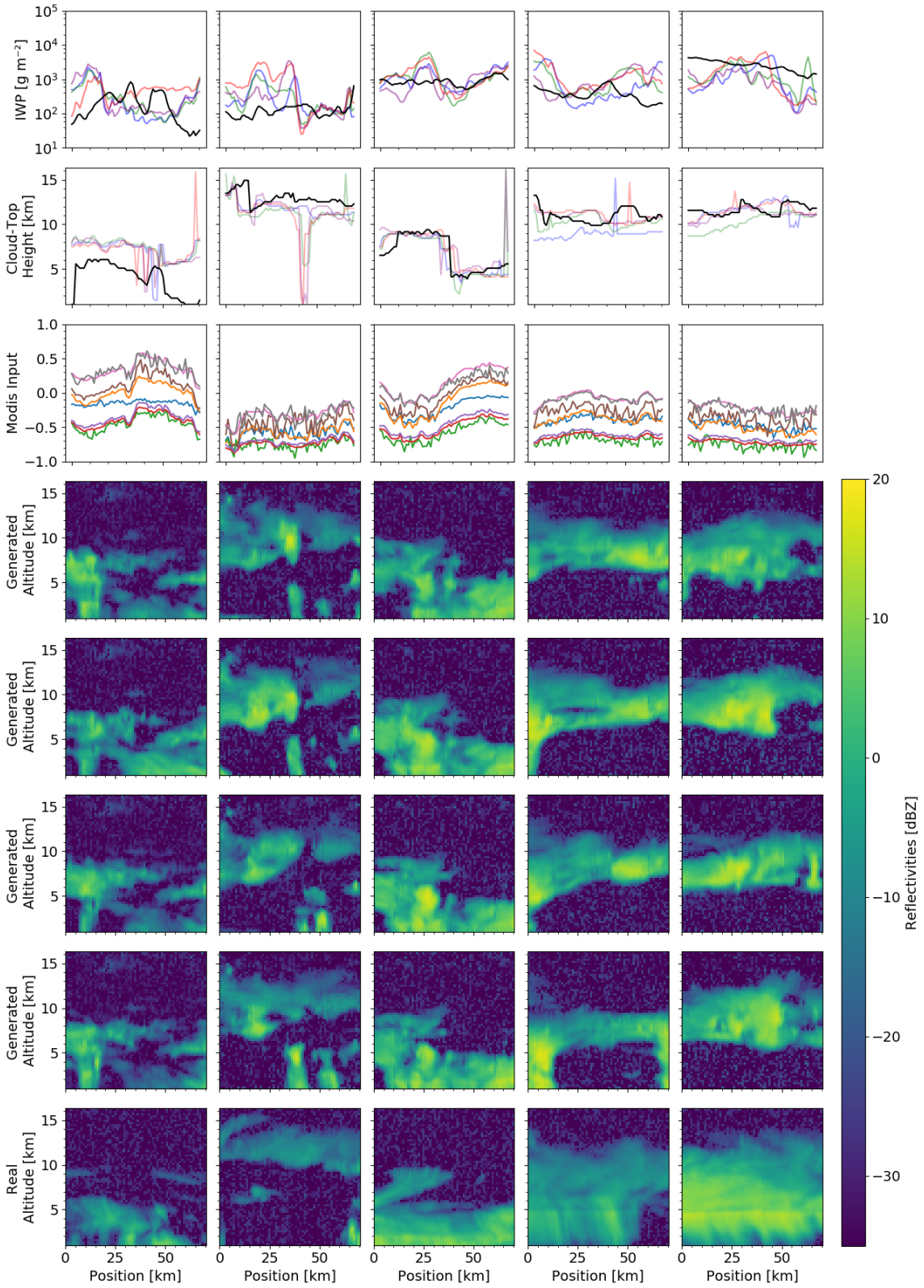


Figure 4.27: Examples of five different scenes from CGAN. Row 1: IWP of the generated (colours) and the original (black) scenes. Row 2: CTH for generated (colours) and original (black) scenes. Row 3: Input of the eight MODIS channels with added error corresponding to the maximum uncertainty. Row 5-7: Scenes generated with CGAN from the MODIS input, with different noise input and new instance of uncertainty added for each generated version. Row 8: Measured CloudSat scenes.

Example scenes generated with CGAN using MODIS data from 2016 as input are available in Section A.6 of the Appendix. The examples show the same amount of resemblance to the real scenes as the generated scenes from 2015 did, which further verifies that CGAN performs well also for other data sets.

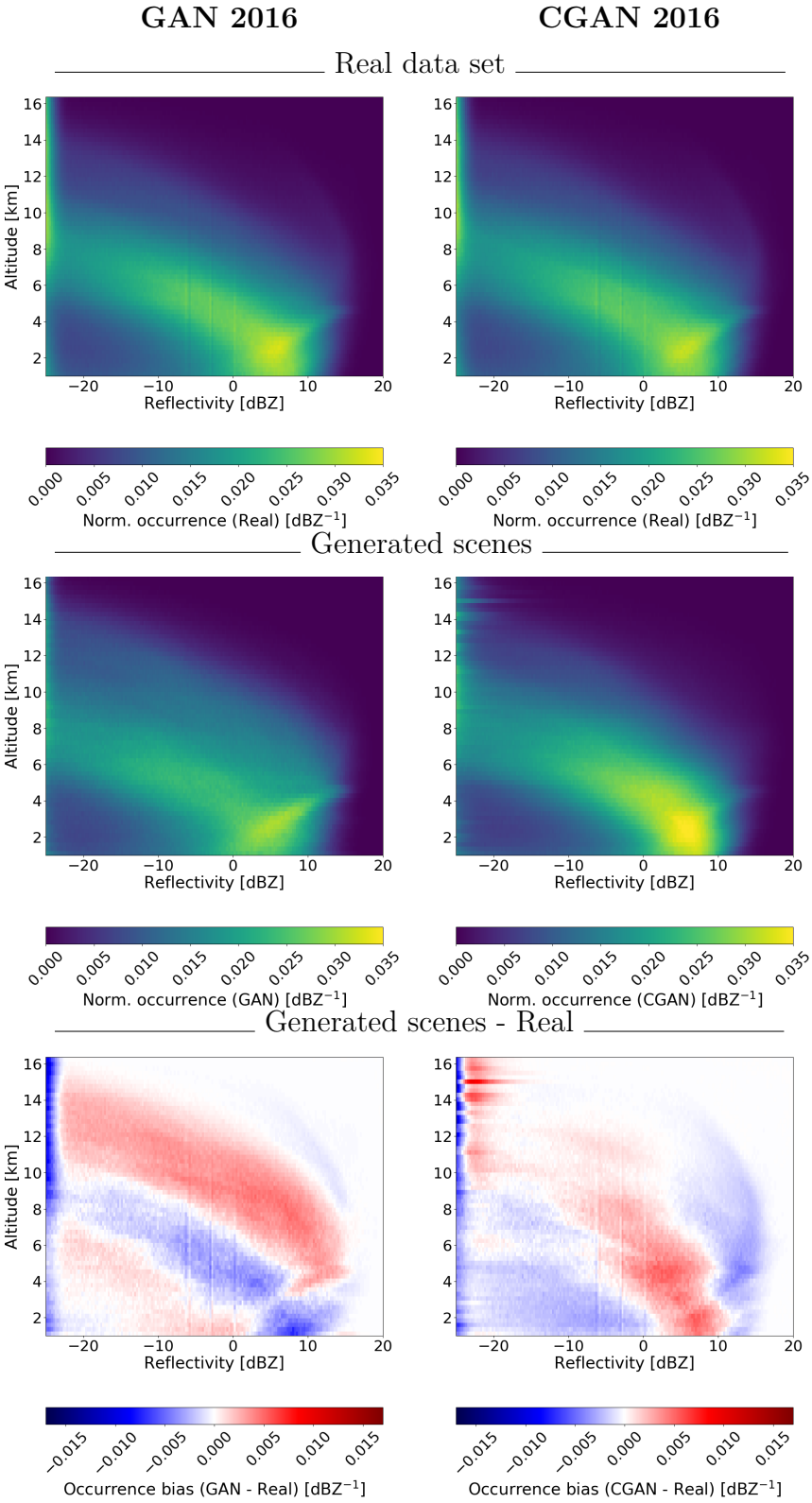


Figure 4.28: Normalized reflectivity-altitude histograms for the data set for GAN and CGAN from 2016.

5
Conclusions

The GAN and CGAN developed in this project are able to generate synthetic scenes with radar reflectivities that resemble the ones in scenes with real measurements. Thus GAN and CGAN can both function as methods for acquiring a larger synthetic radar data set. CGAN also presents promising results for acting as an inversion method for MODIS data. However, in order for it to be used as such it would need to be able to replicate the reflectivity-altitude distribution better.

It has been shown that only visual inspection of example scenes or the analysis of a single measurement of some statistic of the real and generated scenes is not enough to evaluate the validity of any of the methods. Both the GAN and CGAN method perform better than the IAAFT method when it comes to a visual inspection of the scenes as well as recreating for example the spatial autocorrelation present in scenes, whereas IAAFT is better at replicating the global cloud vertical distribution given by reflectivity-altitude histograms. Thus, in order to make a scientifically supported statement about which method is to be preferred for generating synthetic radar scenes, many features of the methods need to be considered and compared. The intended application of the generated data set is also important to consider when choosing the best method. In this regard, this project acts as an important contribution to understanding which statistical properties are important for determining the quality of a method by studying a large number of variables of cloud fields, something that has not been done before. The analysis of the spatial autocorrelation in real and generated cloud scenes is one example of a new and valuable approach tested in this project.

Both GAN and CGAN perform well enough already to be considered as a useful alternative to IAAFT, but the user needs to be aware of some of the issues associated with the generative adversarial networks. These issues include the lack of variability in generated scenes, overtraining and mode collapse. All of these problems present themselves in different ways throughout the visual inspection and statistical analysis. In order to reduce the occurrence of these types of problems we suggest some means of improvement.

A larger training set may be useful in order to avoid the issue of early overtraining of the networks by making it more difficult for the network to learn the exact scenes that are included. Another attempt to solve this problem could be not weighting the training equally between the generator and the discriminator, since the discriminator seems to improve at a much higher rate. An alternative could also be to use the discriminator and generator from different epochs of the training, since the discriminator improves faster than the generator. The most typical improvement is to include a validation set for evaluating the network during training and developing a stopping criterion. Such an early stop could be implemented by evaluating the sum of absolute values of all bins in the

reflectivity-altitude difference histograms. This measure should be minimized, so if this measure does not improve in subsequent epochs the training should be stopped.

The number of input channels needed for the CGAN should be tested further in order to find the optimal conditions of the input data to the network while also minimizing the computation power needed. The amount of noise added to the MODIS input data should also be investigated further. The results in Section 4.4 imply that this could improve the performance, especially when it comes to variability in the example scenes and the derived IWP, but also for the cloud vertical distribution of the entire generated data set.

In order to further assess the seasonal or yearly bias of the network, data from several years could be used. This would also increase the size of the data set, which could be desired as described above.

An interesting further development of these networks would be to implement CGAN to generate three-dimensional cloud fields, i.e. to create a tool for modelling cloud distributions also in the across-track direction.

Bibliography

Bain, A. (1873). *Mind and body: The theories of their relation*, volume 4. Henry S. King.

Biondi, R., Kirchengast, G., and Rieckh, T. (2015). Characterization of thermal structure and conditions for overshooting of tropical and extratropical cyclones with gps radio occultation. *Atmospheric Chemistry and Physics*, 15.

Bony, S., Stevens, B., Frierson, D., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T., Sherwood, S., Siebesma, A., Sobel, A., Watanabe, M., and Webb, M. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8:261–268.

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to time series and forecasting*, volume 2. Springer.

CloudSatDPC (2020). 2b-geoprof. `http://www.cloudsat.cira.colostate.edu/data-products/level-2b/2b-geoprof?term=42`. Accessed: 2020-09-22.

Cronk, H. and Partain, P. (2017a). Cloudsat ecmwf-aux auxillary data product - process description and interface control document. `http://www.cloudsat.cira.colostate.edu/sites/default/files/products/files/ECMWF-AUX_PDICD.P_R05.rev0_.pdf`. Accessed: 2020-09-22.

Cronk, H. and Partain, P. (2017b). Cloudsat modis-aux auxiliary data - process description and interface control document. `http://www.cloudsat.cira.colostate.edu/sites/default/files/products/files/MODIS-AUX_PDICD.P_R05.rev1_.pdf`. Accessed: 2020-09-22.

CSU (2017). Cloudsat overview. `https://cloudsat.atmos.colostate.edu/CloudSat_overview.pdf`. Accessed: 2020-09-22.

CSU (2020). Cloudsat radar status. `https://cloudsat.atmos.colostate.edu/news/CloudSat_status`. Accessed: 2020-08-22.

ECMWF (2020). Data assimilation. `https://www.ecmwf.int/en/research/data-assimilation`. Accessed: 2020-09-22.

eoPortal (2020). Cloudsat. `https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/cloudsat`. Accessed: 2020-04-14.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *arXiv preprint arXiv:1406.2661*.

Guillaume, A., Kahn, B. H., Yue, Q., Fetzer, E. J., Wong, S., Manipon, G. J., Hua, H., and Wilson, B. D. (2018). Horizontal and vertical scaling of cloud geometry inferred from cloudsat data. *Journal of the Atmospheric Sciences*, 75:2187 – 2197.

Heymsfield, A. J., Matrosov, S., and Baum, B. (2003). Ice water path–optical depth relationships for cirrus and deep stratiform ice cloud layers. *Journal of Applied Meteorology and Climatology*, 42.

ICARE (2020). Icare data and services center. `https://www.icare.univ-lille.fr/`.

James, W. (1890). The principles of psychology. new york: H. holt and company.

Jiang, J. H., Yue, Q., Reising, S. C., Kangashlahti, P. P., Deal, W. R., Schlecht, E. T., Wu, L., and Evans, K. F. (2017). A simulation of ice cloud particle size, humidity,and temperature measurementsfrom the twice cubesat. *Earth and Space Science*, 4.

King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., and Hubanks, P. A. (2013). Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7).

Lee, S. I. (2017). *Correlation and Spatial Autocorrelation. In: Shekhar S., Xiong H., Zhou X. (eds) Encyclopedia of GIS*. Springer.

Leinonen, J., Guillaume, A., and Yuan, T. (2019). Reconstruction of cloud vertical structure with a generative adversarial network. *Geophysical Research Letters*, 46(12):7035–7044.

Liou, K. N. (2002). *An Introduction to Atmospheric Radiation*. Academic Press, San Diego, USA, 2 edition.

Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *CoRR*, abs/1411.1784.

NASA (2005). The importance of understanding clouds. `TheImportanceofUnderstandingClouds`. Accessed: 2020-09-23.

NASA (2009). Aqua satellite and modis swath. `https://svs.gsfc.nasa.gov/3348`. Accessed: 2020-04-14.

NASA (2020a). A-train constellation. `https://www-calipso.larc.nasa.gov/about/atrain.php`. Accessed: 2020-09-21.

NASA (2020b). Terra & aqua moderate resolution imaging spectroradiometer (modis). `https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/`. Accessed: 2020-09-22.

NCCO (2020). Composition of the atmosphere. `https://climate.ncsu.edu/edu/Composition`. Accessed: 2020-07-26.

Oreopoulos, L., Cho, N., and Lee, D. (2017). New insights about cloud vertical structure from cloudsat and calipso observations. *Journal of Geophysical Research: Atmospheres*, 122(17):9280–9300.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., and Frey, R. A. (2003). The modis cloud products: algorithms and examples from terra. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):459–473.

Protat, A., Delanoe, J., Strapp, J. W., Fontaine, E., Leroy, D., Schwarzenboeck, A., Lilie, L., Davison, C., Dezitter, F., Grandin, A., and Weber, M. (2016). The measured relationship between ice water content and cloud radar reflectivity in tropical convective clouds. *American Meterological Society*, 55:1707 – 1729.

Rees, W. G. (2012). *Physical Principles of Remote Sensing*. Cambridge University Press, New York, USA, 3 edition.

Schmal, C., Myung, J., Herzel, H., and Bordyugov, G. (2017). Moran's I quantifies spatio-temporal pattern formation in neural imaging data. *Bioinformatics*, 33(19).

Stephens, G. L. (2005). Cloud feedbacks in the climate system: A critical review. *Journal of climate*, 18:237 – 273.

Ustinov, E. (2014). *Geophysical Retrieval, Forward Models in Remote Sensing. In: Njoku E.G. (eds) Encyclopedia of Remote Sensing. Encyclopedia of Earth Sciences Series*. Springer.

Venema, V. (2006). Iaaft surrogate clouds. `https://www2.meteo.uni-bonn.de/mitarbeiter/venema/themes/surrogates/iaaft/`.

Venema, V., Meyer, S., Garcìa, S. G., Kniffka, A., Simmer, C., Crewell, S., Löhnert, U., Trautmann, T., and Macke, A. (2006). Surrogate cloud fields generated with the iterative amplitude adapted fourier transform algorithm. *Tellus A: Dynamic Meteorology and Oceanography*, 58(1):104–120.

Wylie, D., Jackson, D. L., Menzel, W. P., and Bates, J. J. (2005). Trends in global cloud cover in two decades of hirs observations. *Journal of Climate*, 18.

A

Appendix 1

In this appendix additional figures of example scenes and histograms that are not essential for understanding and/or motivating the results in the main paper are presented.

A.1 CGAN examples

Below are some additional randomly selected examples of generated scenes with CGAN and derived properties such as IWP and CTH. These figures are similar to Figure 4.4 of the main report.

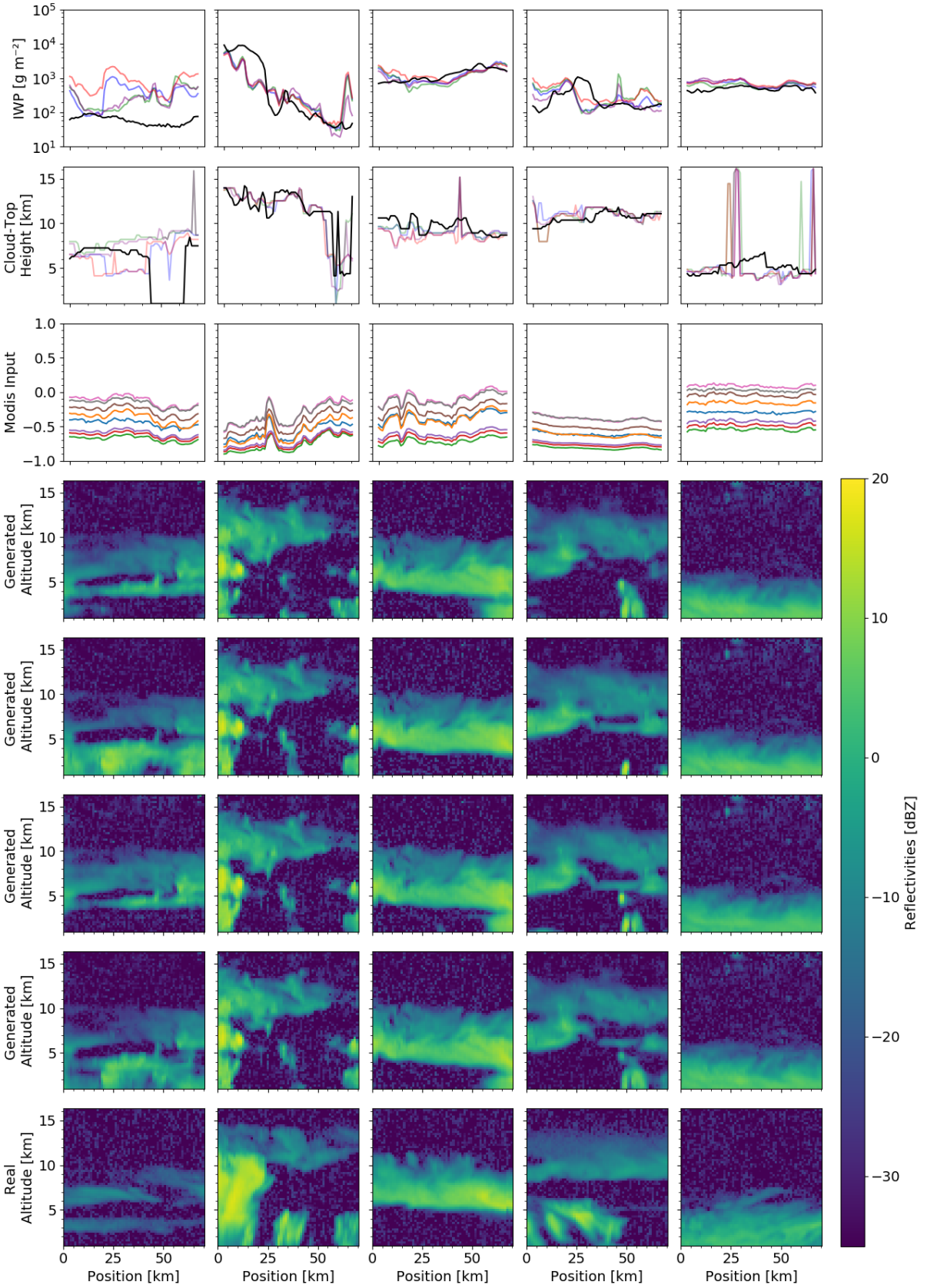


Figure A.1: Examples of five different scenes from CGAN. Each column corresponds to one example. Row 1: IWP of the generated scene (colours) and the original scene (black) on a logarithmic y-axis. Row 2: CTH for generated (colours) and original (black) scenes. Row 3: Input of the eight MODIS channels. Row 5-7: Scenes generated with CGAN from the MODIS input, with different noise input. Row 8: Measured CloudSat scenes.

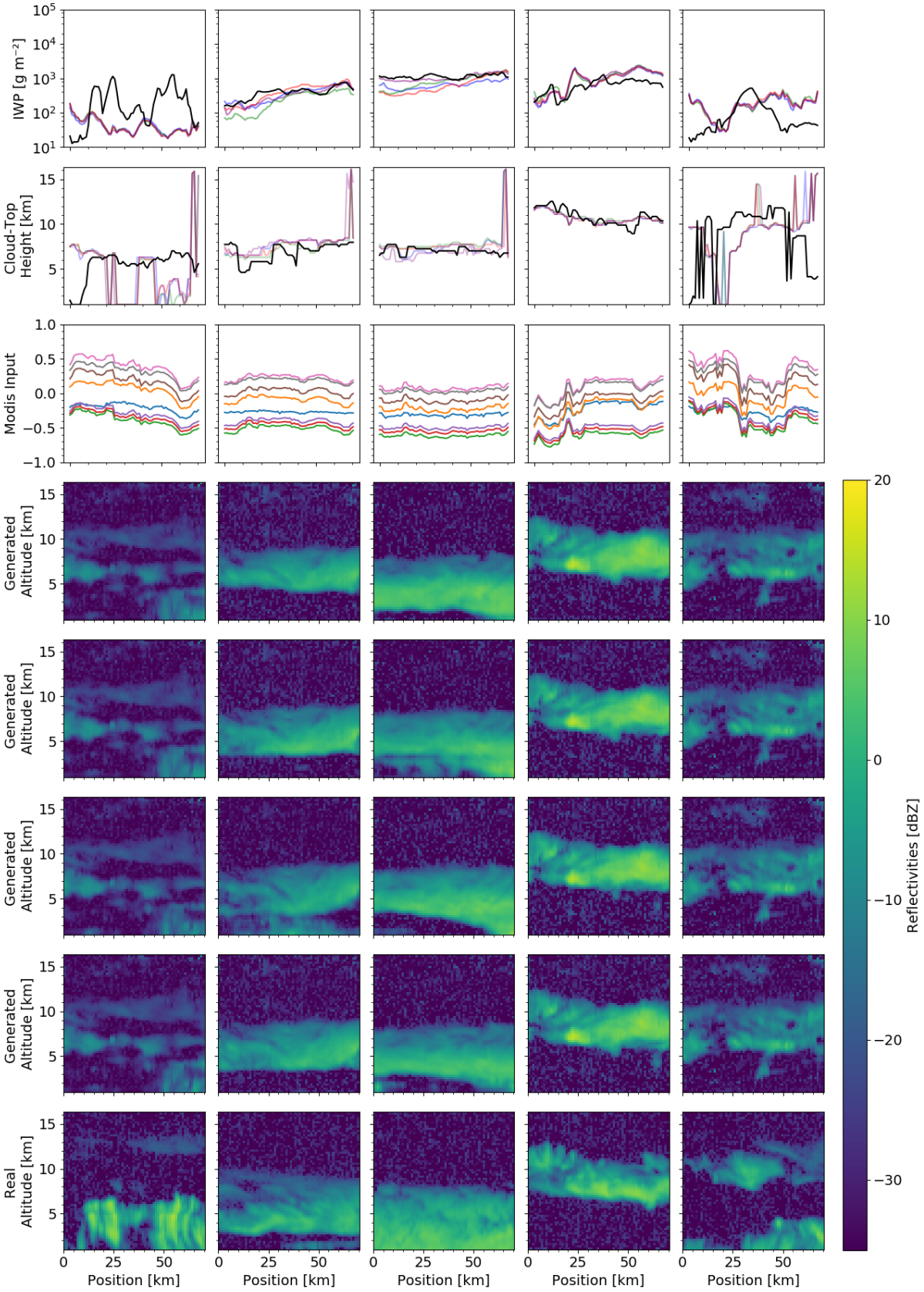


Figure A.2: Examples of five different scenes from CGAN. Like Figure 4.4

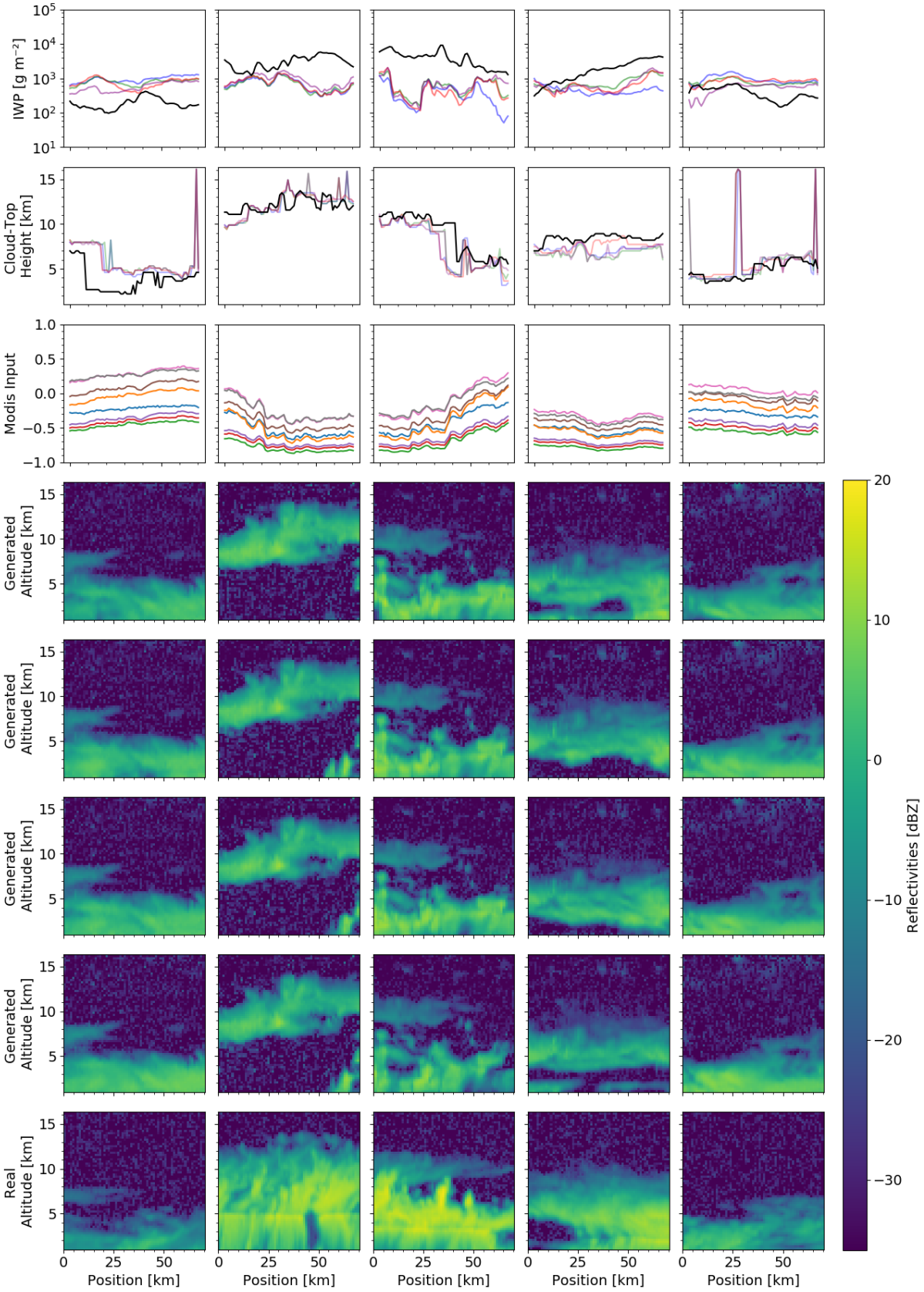


Figure A.3: Examples of five different scenes from CGAN. Like Figure 4.4

A.2 Spatial autocorrelation examples

Below are some examples of calculated spatial autocorrelation for scenes generated by GAN, CGAN or IAAFT. The spatial autocorrelation histograms in Figure 4.20 are calculated from the average of histograms of the type in Figure A.4 to Figure A.7 for the entire data set.

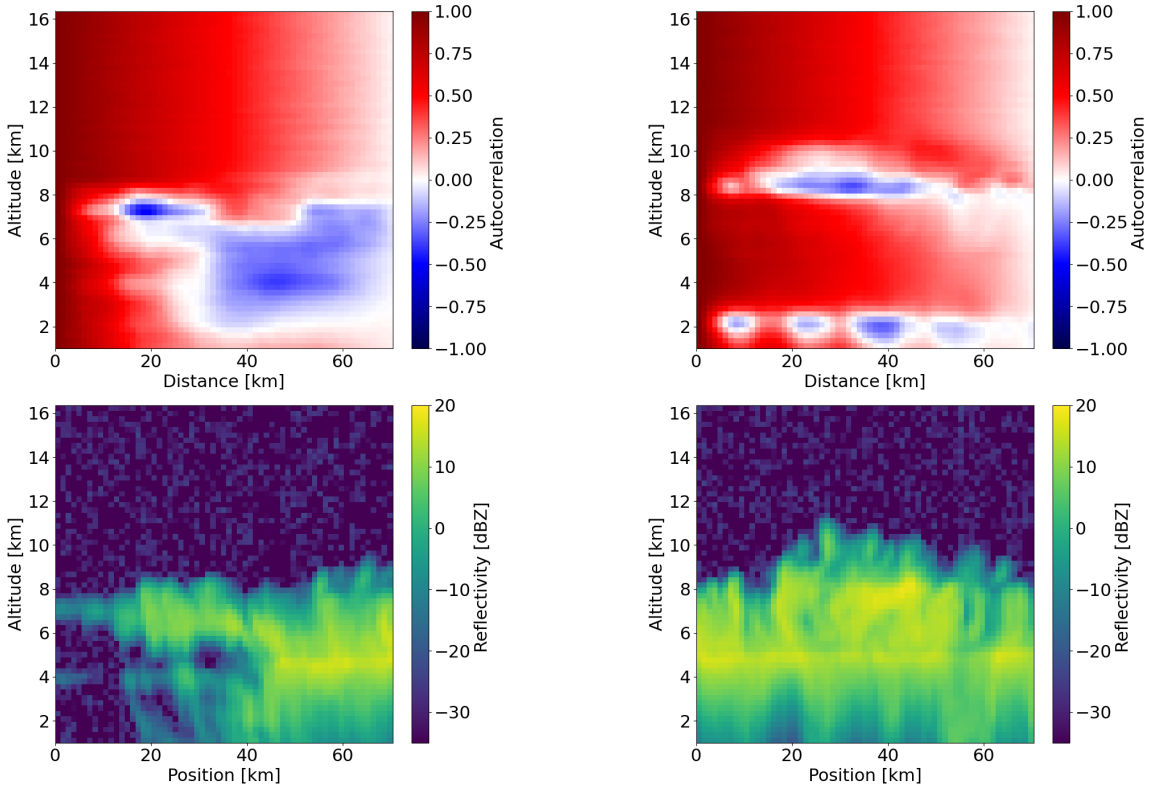


Figure A.4: Examples of spatial autocorrelation used to calculate the averages in column 1 and 2 of row 1 in Figure 4.20. Here the top image shows the spatial autocorrelation calculated from the real CloudSat scene in the lower image.

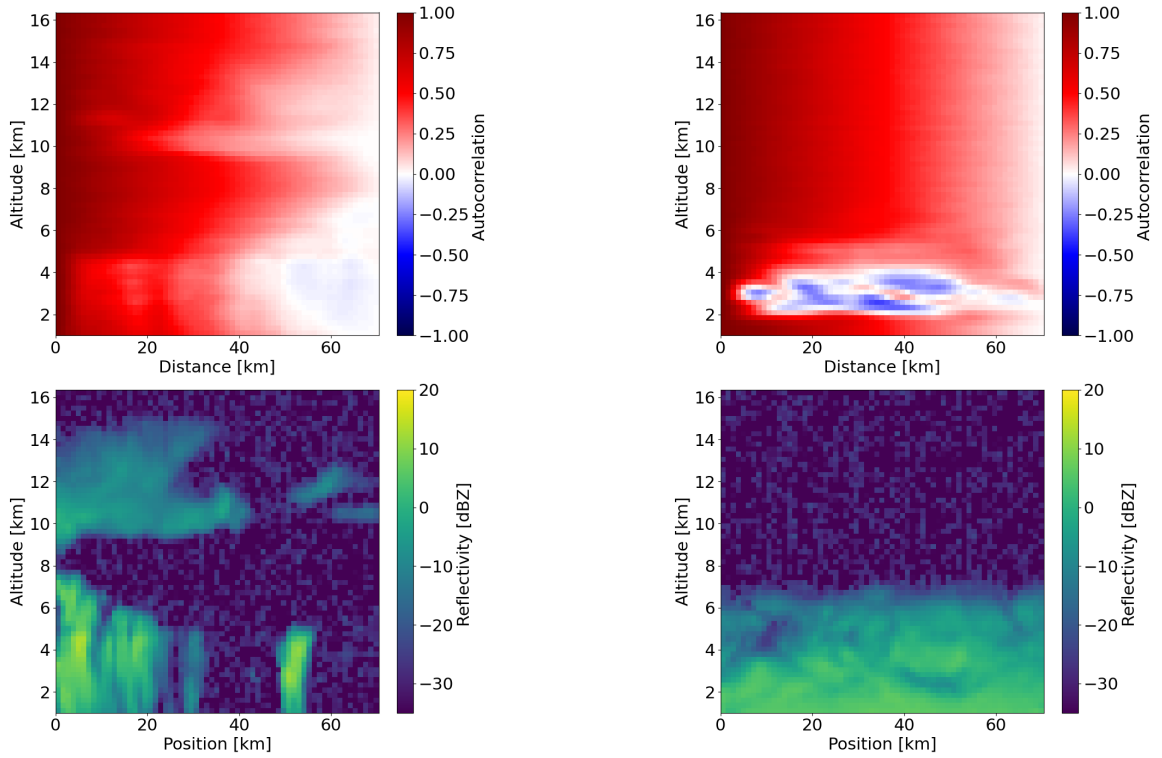


Figure A.5: Examples of spatial autocorrelation used to calculate the averages in row 2, column 1 of Figure 4.20. Here the top image shows the spatial autocorrelation calculated from the GAN generated scene in the lower image.

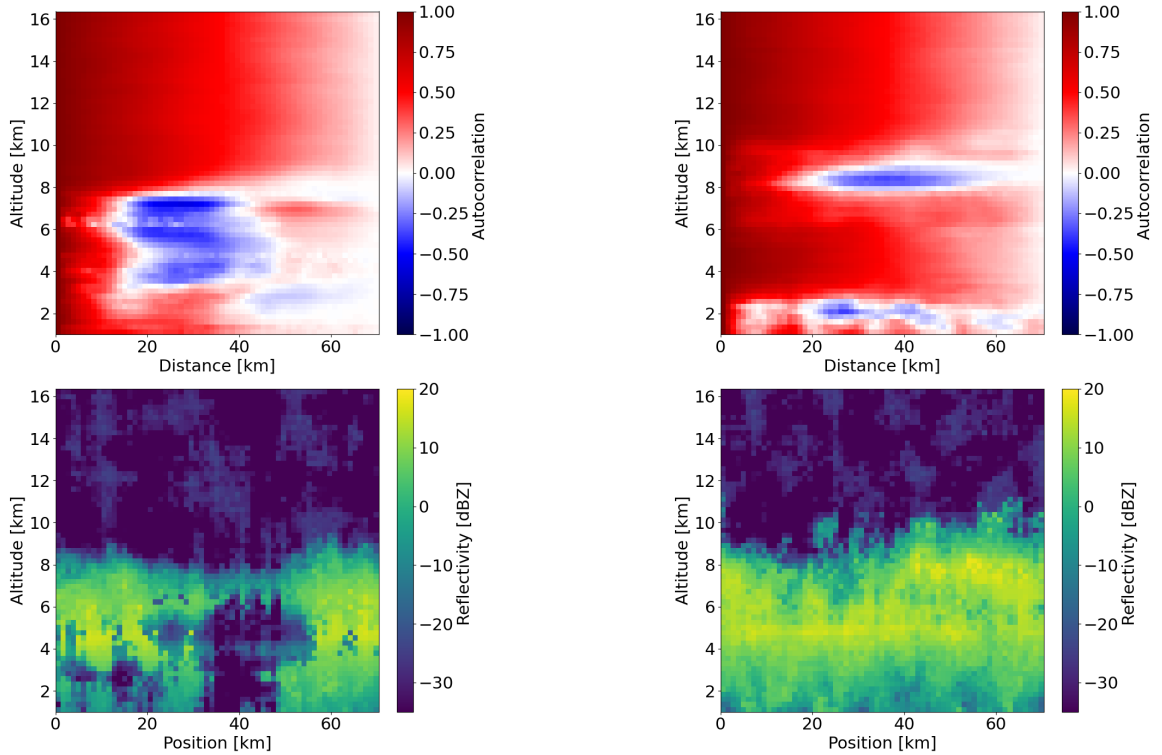


Figure A.6: Examples of spatial autocorrelation used to calculate the averages in row 2, column 2 of Figure 4.20. Here the top image shows the spatial autocorrelation calculated from the IAAFT generated scene in the lower image.

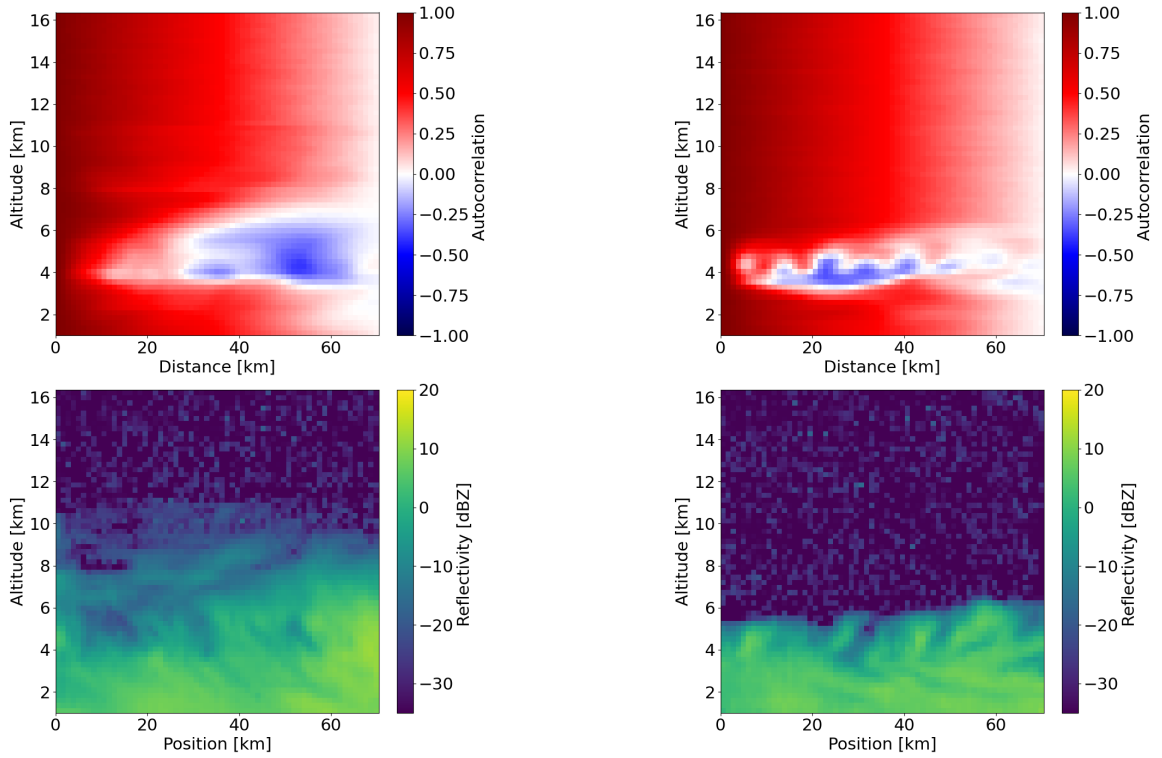


Figure A.7: Examples of spatial autocorrelation used to calculate the averages in row 2, column 3 of Figure 4.20. Here the top image shows the spatial autocorrelation calculated from the CGAN generated scene in the lower image.

A.3 Later epochs - GAN

In this section discriminator tests for the network at later epochs than 400 are shown. The increase in overtraining is apparent.

A.3.1 Epoch 1400

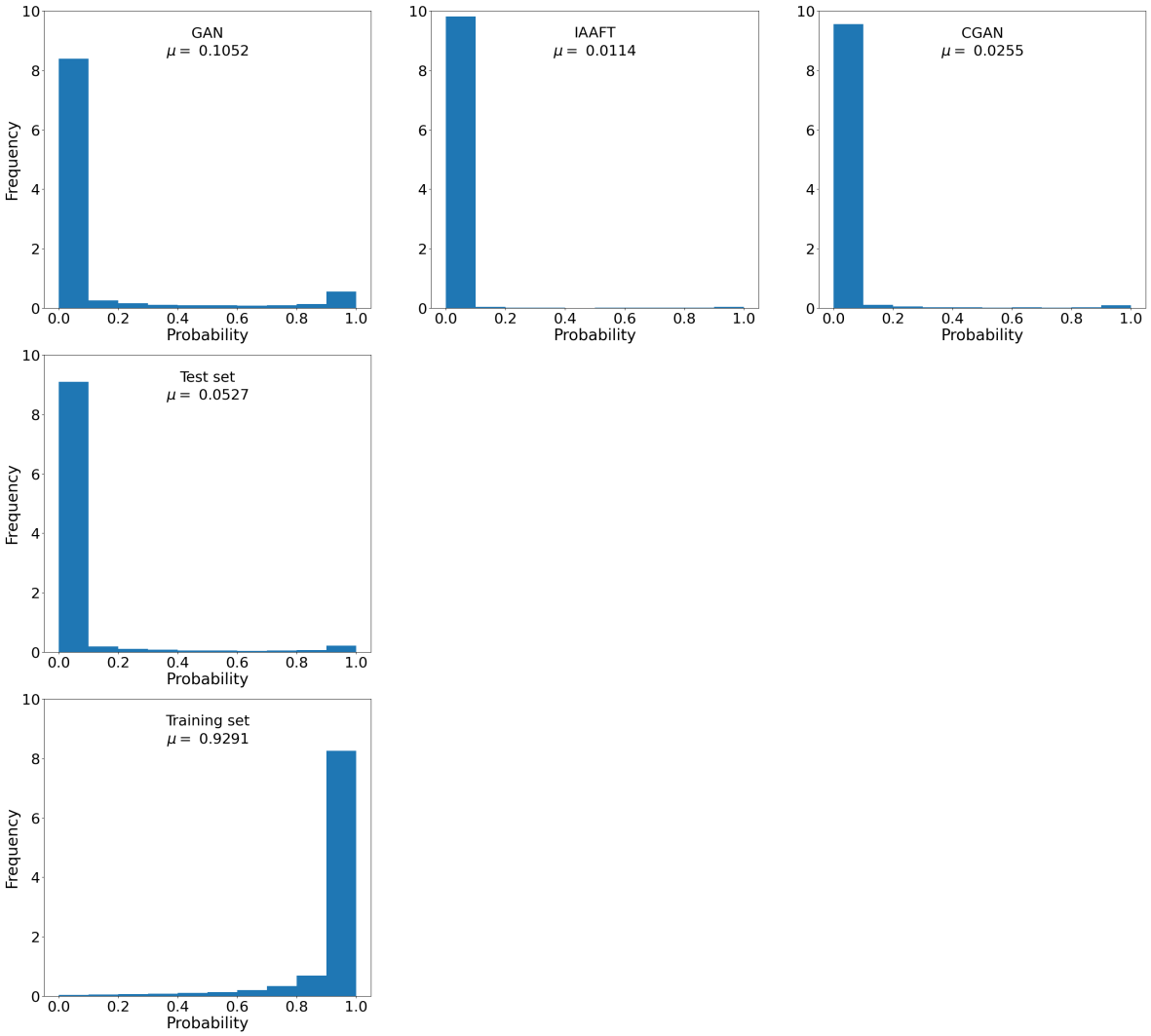


Figure A.8: Histograms of different data sets classified by the discriminator of GAN at epoch 1400. Scenes from GAN are generated from the generator at epoch 1400 and scenes from CGAN are generated from epoch 3500. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.3.2 Epoch 2400

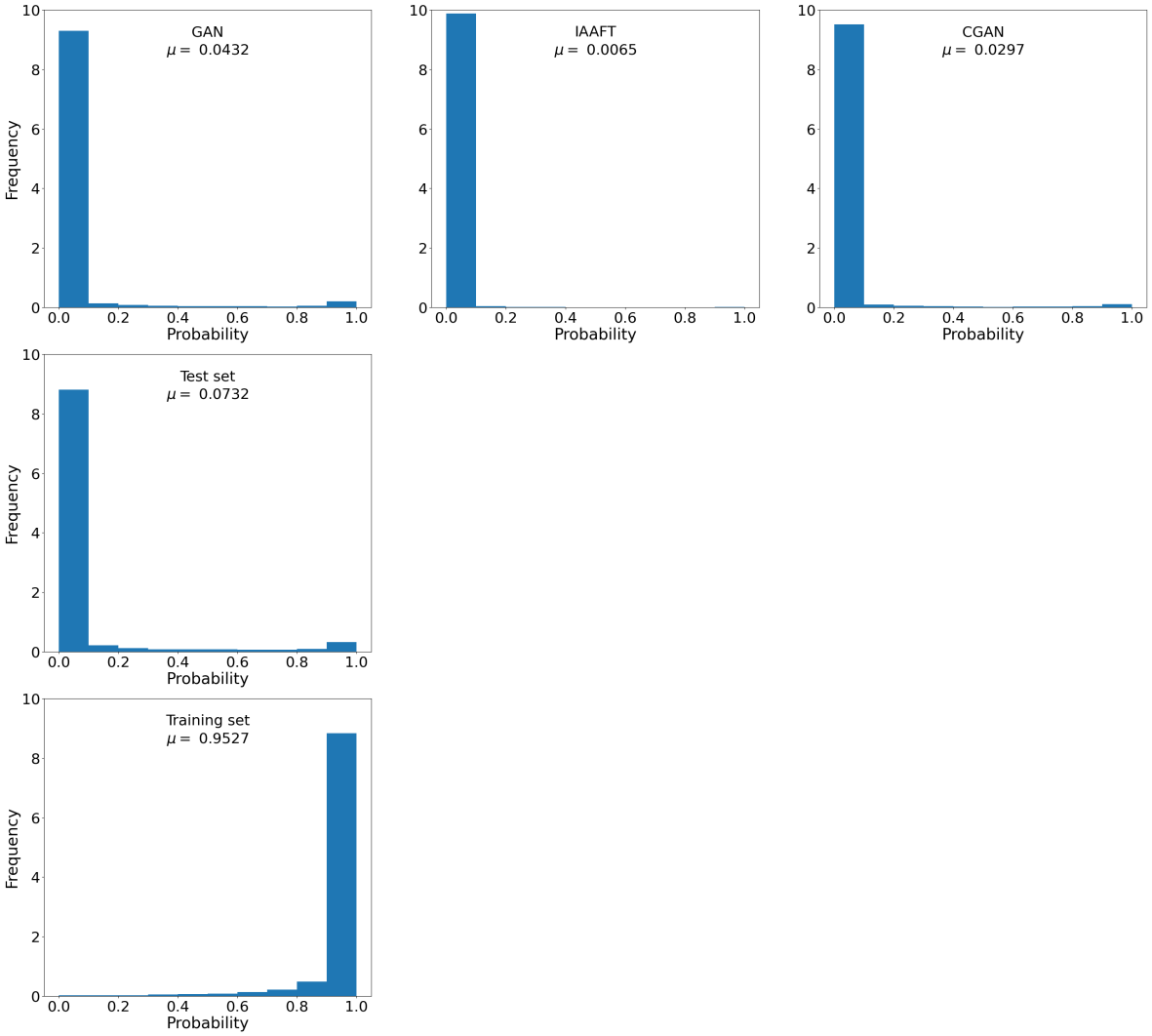


Figure A.9: Histograms of different data sets classified by the discriminator of GAN at epoch 2400. Scenes from GAN are generated from the generator at epoch 2400 and scenes from CGAN are generated from epoch 3500. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.3.3 Epoch 3000

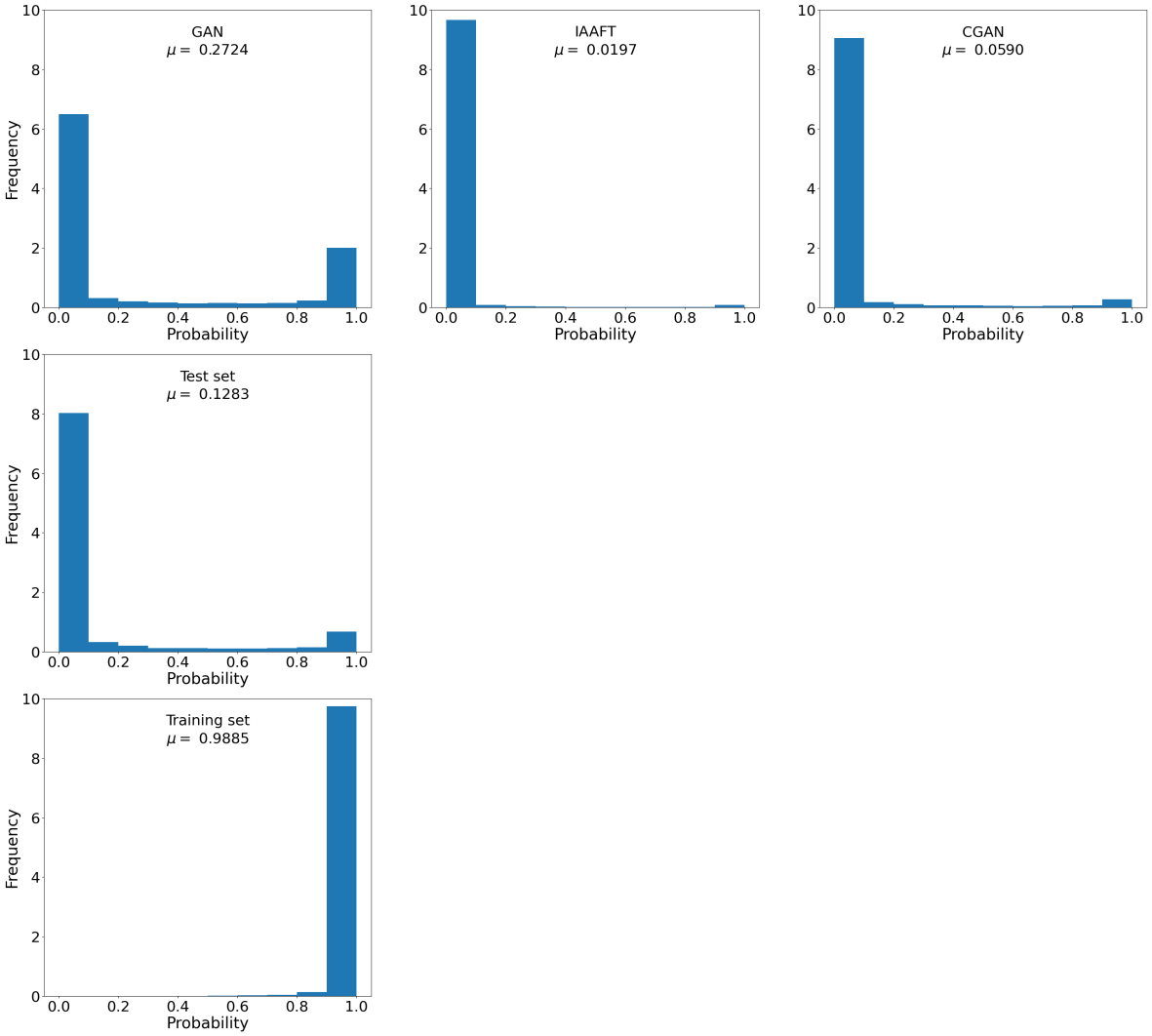


Figure A.10: Histograms of different data sets classified by the discriminator of GAN at epoch 3000. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.4 Early epochs - CGAN

In this section discriminator tests for the CGAN at earlier epochs than 3500 are shown.

A.4.1 500

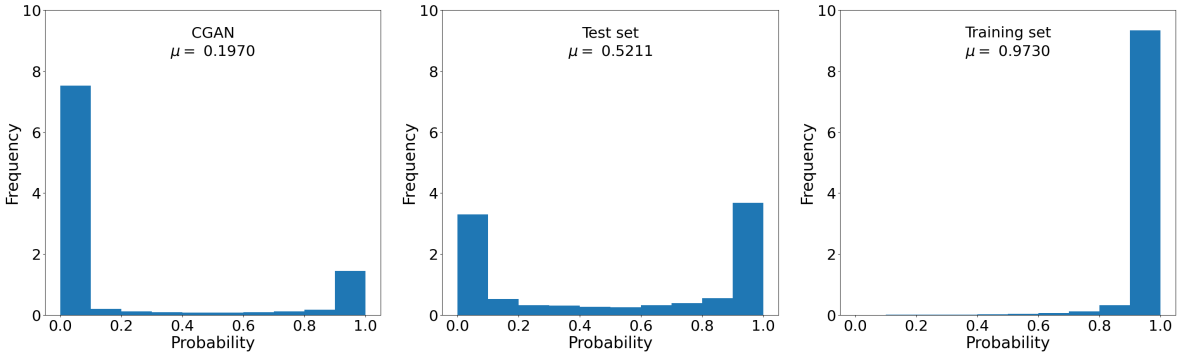


Figure A.11: Histograms of different data sets classified by the discriminator of the CGAN at epoch 500. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.4.2 1500

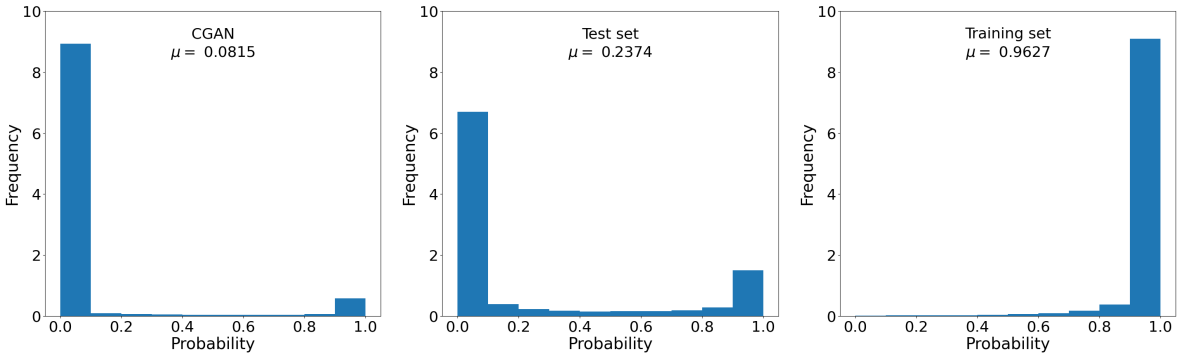


Figure A.12: Histograms of different data sets classified by the discriminator of the CGAN at epoch 1500. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.4.3 2500

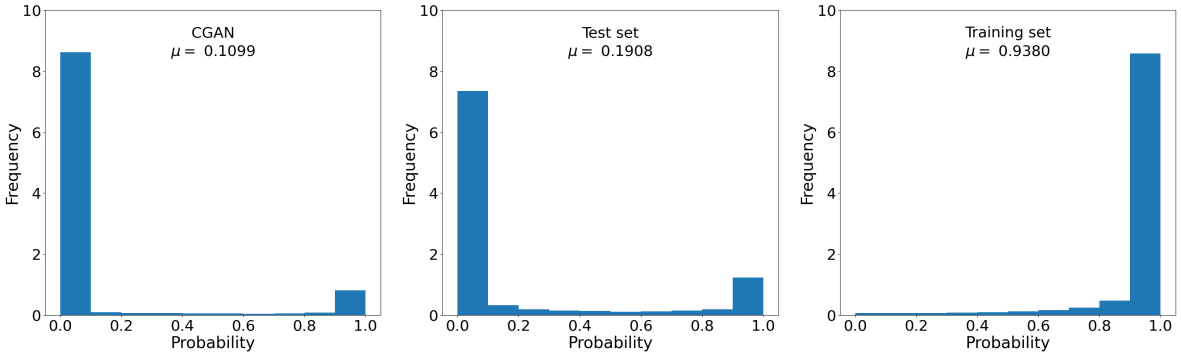


Figure A.13: Histograms of different data sets classified by the discriminator of the CGAN at epoch 2500. A probability of 1 means that the discriminator classifies the scene as true, and conversely a probability of 0 means the discriminator classifies the scene as false.

A.5 Results of an overtrained network

Results of the GAN at epoch 3000 of the training are presented in this section. The figures should be compared to corresponding figures from epoch 400 of GAN used in the report.

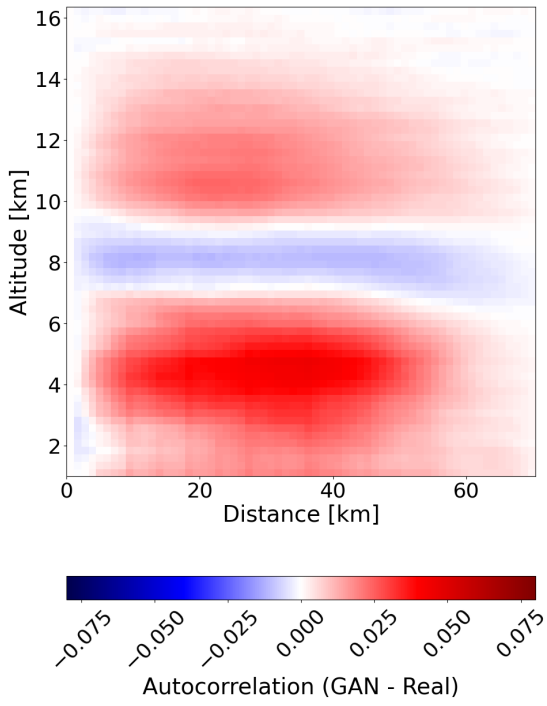


Figure A.14: Difference in average autocorrelation between scenes generated by GAN at epoch 3000 and real scenes.

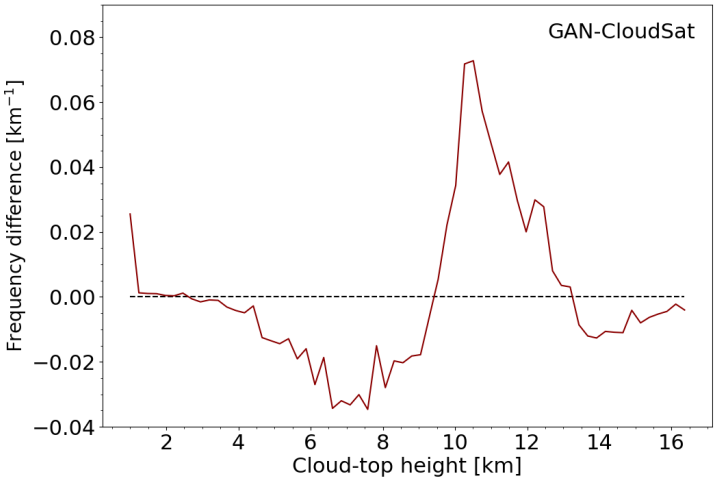


Figure A.15: Difference in cloud-top height between scenes generated by GAN at epoch 3000 and real scenes.

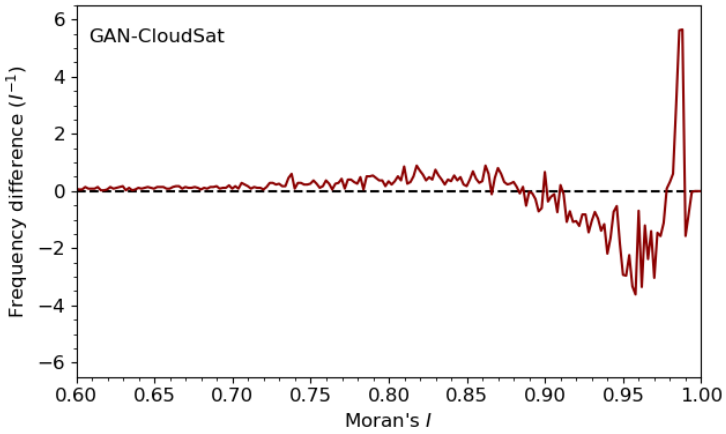


Figure A.16: Difference in spatial autocorrelation measured by Moran's I between scenes generated by GAN at epoch 3000 and real scenes.

A.6 Scenes generated with 2016 data

Examples of scenes generated with data from 2016 are visualized in Figure A.17.

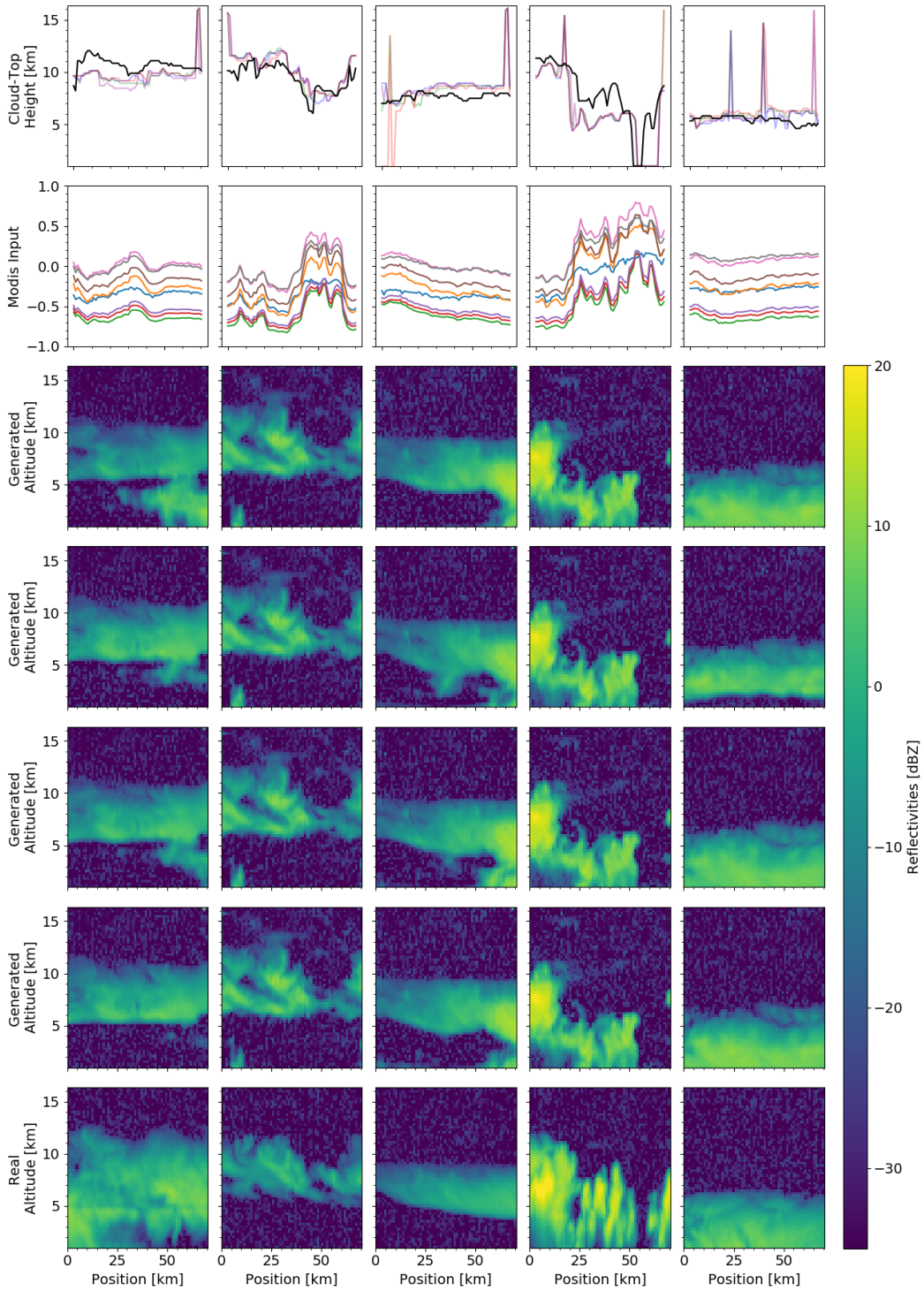


Figure A.17: Examples of five different scenes for CGAN using the data set from 2016. Each column corresponds to one example. Row 1: CTH for generated (colours) and original (black) scenes. Row 2: Input of the eight MODIS channels from 2016. Row 3-6: Scenes generated with CGAN from the MODIS input, with different noise input. Row 7: Measured CloudSat scenes from 2016.