Characterizing maritime trade-wind convection using the HALO Microwave Package (HAMP)

INAUGURAL - DISSERTATION

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Marek Jacob

aus Henstedt-Ulzburg

Köln, 2020

Abstract

This thesis explores the marine trade-wind convection and the clouds forming within by using spatial-high-resolution airborne remote sensing observations taken from the German High Altitude and LOng range research aircraft (HALO). The nadir-pointing HALO Microwave Package (HAMP) is the central tool of this thesis. HAMP comprises a cloud radar and a 26-channel microwave radiometer (MWR, 22–183 GHz), for which the atmosphere and clouds are semitransparent. The shallow cumulus clouds, like they regularly occur in the trade-wind region, are of particular interest for better understanding the climate. Several studies (e.g., Bony and Dufresne, 2005; Schneider et al., 2017) identified such clouds as a main source of model spread in climate projections. The challenge of this kind of ubiquitous clouds in the models is partly due to large spread in global observations which can be related to the small scale of shallow cumuli and the coarse-scale observations from satellites. This thesis combines three studies around HAMP from the characterization of the HAMP MWR, over the development of MWR retrievals for liquid clouds to the application by evaluating two cloud-resolving simulations.

The HAMP MWR is characterized by investigating the random noise of each channel, the covariance within each of the five frequency bands, the brightness temperature (BT) offset, the offset stability, and by suggesting an offset correction. The offset and stability of the HAMP BT acquisitions are studied by comparing the measured BTs to synthetic measurements based on forward-simulated dropsondes. Offsets between -11 and $+6\,\mathrm{K}$ show a spectral dependency, which repeatedly appears but is shifted between flights. The offsets are most likely caused by uncertainties in the calibration method and changing environmental conditions of the MWR in the belly pod during take-off and ascending. However, an offset correction based on the dropsondes can be developed for each channel as a function of the flight.

To better interpret the HAMP BT observations, novel retrievals are developed based on a realistic database of synthetic measurements and corresponding atmospheric profiles. Retrievals of the liquid water path (LWP), rainwater path (RWP), and integrated water vapor (IWV) are developed to describe the clouds and their environment. The retrieved IWV using the offset-corrected BTs agrees with coincident dropsondes and water vapor lidar measurements by $1.4\,\mathrm{kg\,m^{-2}}$. The theoretical assessment of LWP shows that the LWP error is below $20\,\mathrm{g\,m^{-2}}$ for LWP below $100\,\mathrm{g\,m^{-2}}$. The absolute LWP error increases with increasing LWP, but the relative error decreases from $20\,\%$ at $100\,\mathrm{g\,m^{-2}}$ to $10\,\%$ at $500\,\mathrm{g\,m^{-2}}$. The RWP retrieval,

which uses the radar in addition to the MWR, can reliably detect RWP larger than $10\,\mathrm{g\,m^{-2}}$ with a Gilbert skill score > 0.75.

The retrieval results are summarized in a comparison of the clouds and their moisture environment in the two tropical seasons, which are represented by the field experiments in December 2013 (dry season) and in August 2016 (wet season). Clouds were more frequent, and their average LWP and RWP were higher in the dry season than in the wet season. However, deeper convection with the formation of large frozen particles was less frequent in the dry season. It is hypothesized, that the lower degree of cloud organization in the dry season led to smaller systems with more overall cloud cover. The higher degree of randomness in the dry season comes along with less extremes and is reflected by a narrower distribution of IWV. The variability between (especially the wet-season) flights shows, how statistics from airborne campaigns are affected by the choice of the individual flight pattern.

The more homogeneous and cloudy statistics of the dry season are used to assess the representation of shallow cumulus convection and the cloud formation over the ocean in two cloud-resolving simulations generated with the ICON model. The HAMP radar and a backscatter lidar are used for detecting cloud top height (CTH), base height, and precipitation, and the MWR stratifies the cases by LWP. Forward simulators are used to derive the same measurements synthetically from the model data while applying the same instrument-specific cloud-detection thresholds. The analysis reveals a bimodal structure of the CTH. The lower mode relates to boundary layer driven clouds, while the upper mode is driven by moist shallow convection, trapped under the trade inversion at about $2.3\,\mathrm{km}$ above sea level. The storm-resolving model (SRM) with $1.25\,\mathrm{km}$ horizontal grid spacing resolves the two cloud layers to a limited extend. Most CTHs in the SRM are above the observed lower CTH mode, and top height increases with LWP. The second model with a $300\,\mathrm{m}$ grid (large-eddy model, LEM) represents better the observed bimodal distribution of CTH. However, the microphysical schema of neither model can produce in-cloud drizzle-sized particles that were often observed by the radar. This application study shows, how HAMP on HALO provides insightful data to help closing the uncertainty in the models, if interpreted thoroughly.

Zusammenfassung

Diese Arbeit untersucht konvektive maritime Wolken im Passat. Dazu wurden Messflüge mit dem deutschen Forschungsflugzeug HALO (engl. *High Altitude and LOng range*) durchgeführt. Das in Nadirrichtung schauende HALO MikrowellenPaket (HAMP) ist dabei das zentrale Instrument dieser Arbeit. HAMP kombiniert ein Wolkenradar und ein Mikrowellenradiometer (MWR, 22–183 GHz) mit 26 Kanälen. Für beide sind Atmosphäre und Wolken halb-transparent. Die Untersuchung der niedrigen Cumulus Wolken, die sich regelmäßig in der Passatregion bilden, ist vor allem für ein besseres Klimaverständnis relevant. Verschiedene Studien (z.B. Bony and Dufresne, 2005; Schneider et al., 2017) identifizierten diese Wolken als einen Hauptgrund für die Vorhersagespannweite von Klimamodellen. Die Herausforderung diesen weit verbreiteten Wolkentyp zu modellieren, liegt zum Teil in der Variabilität, die globale Beobachtungsdaten geben. Diese Variabilität lässt sich auf die geringe räumliche Ausdehnung dieser Wolken im Verhältnis zur räumlichen Auflösung von Mikrowellensatelliten zurückführen. Die vorliegende Arbeit vereint drei Studien mit HAMP, beginnend mit der Charakterisierung der HAMP MWR, über die Entwicklung von MWR Retrievals für Flüssigwasserwolken bis hin zur Anwendung in der Bewertung von wolkenauflösenden Simulationen.

Das HAMP MWR wird durch die Untersuchung des Kanalrauschens, der Kovarianz in jedem der fünf Frequenzbänder, des Versatzes der Helligkeitstemperaturen (BT, *brightness temperature*) und der Versatzstabilität charakterisiert. Der HAMP-BT-Versatz und seine Stabilität werden untersucht, indem die gemessenen BT mit synthetischen BT aus Vorwärtsrechnungen basierend auf Fallsonden verglichen werden. Die Abweichungen zwischen -11 und $+6$ K zeigen eine spektrale Abhängigkeit, die sich, zwar mit Versatz, von Flug zu Flug wiederholt. Höchstwahrscheinlich wird dieser Versatz durch Unsicherheiten im Kalibrationsverfahren sowie Änderungen der Umgebungsparameter in der Instrumentenverkleidung des MWR hervorgerufen. Dennoch kann eine Korrektur für den BT-Versatz in jedem Kanal in Abhängigkeit vom Flugtag entwickelt werden.

Um die HAMP-BT-Beobachtungen einfacher zu interpretieren, werden neue Retrieval entwickelt, die auf einer Datenbank mit realistischen Fällen synthetischer Messungen und dazugehörigen Atmosphärenprofilen basieren. Um die Wolken und ihre Umgebung zu beschreiben, werden Retrievals für den Flüssigwasserpfad (LWP, *liquid water path*), den Regenwasserpfad (RWP, *rain water path*) und den integrierten Wasserdampf (IWV, *integrated water vapor*) entwickelt. Der gewonnene IWV, basierend auf um den BT-Versatz korrigierten BT, zeigt eine Übereinstimmung mit Fallsonden und Wasserdampflidarmessungen besser als $1.4\,\mathrm{kg\,m^{-2}}$. Die theoretische

Untersuchung der LWP-Genauigkeit ergibt, dass für einen LWP unter $100\,\mathrm{g\,m^{-2}}$ der Fehler kleiner als $20\,\mathrm{g\,m^{-2}}$ ist. Der absolute Fehler nimmt mit zunehmenden LWP zu, der relative Fehler verringert sich jedoch von $20\,\%$ bei $100\,\mathrm{g\,m^{-2}}$ zu $10\,\%$ bei $500\,\mathrm{g\,m^{-2}}$. Das RWP Retrieval nutzt zusätzlich zum MWR auch das Radar und erkennt verlässlich Regen mit RWP größer als $10\,\mathrm{g\,m^{-2}}$ mit einem Gilbert-Skill-Wert von über 0.75.

Die Retrievalergebnisse werden in einem Vergleich der Wolken und ihrer Wasserdampfumgebung in den zwei tropischen Jahreszeiten zusammengefasst. Die Jahreszeiten sind durch Feldexperimente im Dezember 2013 (Trockenzeit) und August 2016 (Regenzeit) vertreten. Der Bedeckungsgrad und der LWP und RWP in den Wolken waren höher in der Trockenzeit als in der Regenzeit. Tiefe Konvektion mit der Bildung von großen Eispartikeln war jedoch seltener in der Trockenzeit. Es wird vermutet, dass eine geringere Selbstorganisation der Wolken in der Trockenzeit generell zu kleineren, aber mehr Wolken führte. Die größere Grad an Zufälligkeit in der Trockenzeit geht mit weniger Extrema einher und stellt sich durch eine schmalere IWV-Verteilung dar. Die Variabilität zwischen den Flügen, vor allem jenen der 2016er Messungen, zeigen, wie die Statistik von Flugzeugmessungen durch die Wahl der Flugmuster beeinflusst wird.

Die homogenere und wolkenreichere Statistik der Trockenzeit wird genutzt, um die Wiedergabe von niedriger Konvektion und Wolken über dem Ozean in zwei wolkenauflösenden Simulationen des ICON Modells zu untersuchen. Das HAMP Radar und ein Rückstreulidar werden zur Erkennung von Wolkenoberkanten, -unterkanten und Niederschlag genutzt. Zusätzlich erlaubt das MWR die Klassifikation nach LWP. Vorwärtsrechnungen werden genutzt, um vergleichbare synthetische Messungen aus den Modelldaten zu erzeugen und um anschließend die gleichen instrumentenspezifischen Wolkenschwellwerte anzusetzen. Die Beobachtungen zeigen eine bimodale Verteilung der Wolkenoberkantenhöhen. Grenzschichtprozesse erzeugen die Wolken der niedrigeren Mode, während Wolken der oberen Mode mit feuchter Konvektion in Verbindung stehen und nach oben durch die Passatinversion bei etwa $2.3\,\mathrm{km}$ beschränkt sind. Das Modell mit einem horizontalen $1.25\,\mathrm{km}$-Gitter kann die zwei Moden nur bedingt auflösen. Die meisten Wolkenoberkanten in diesem Modell sind etwas über der beobachteten niedrigen Mode der Wolkenoberkanten. Das Modell mit einem $300\,\mathrm{m}$-Gitter gibt die beobachtete Bimodalität deutlich besser wieder. Jedoch ist das mikrophysikalische Modell in keiner der Simulationen in der Lage, Nieseltröpfchen, die häufig beobachtete Radarsignale erzeugen könnten, in den Wolken abzubilden. Diese Anwendungsstudie zeigt, wie HAMP Daten helfen können, Unsicherheiten in Modellen zu aufzuzeigen, wenn die Daten sorgfältig interpretiert werden.

Contents

Chapter 1.

Introduction

Clouds are fascinating atmospheric phenomena and play a remarkable role in the Earth climate system. Among them, maritime cumulus clouds, that are frequently triggered by shallow convection in the trade-wind regimes, are of particular importance. Reliable measurements are the key to foster our understanding expressed in models, but the enormous range of cloud sizes over several orders of magnitude defies all encompassing observations. From the wide range of in-situ to remote sensing observations in different spectral regimes, microwave remote sensing stands out, because clouds are semitransparent to microwaves such that the whole cloudy profile can be studied at once. This thesis contributes to the understanding of maritime convection using the airborne active and passive microwave package HAMP by characterizing the instrument, developing and assessing its retrievals, and by confronting cloud-resolving simulations with HAMP observations.

1.1. Motivation

Clouds cover on average about two thirds of our globe (Stubenrauch et al., 2013). Through their presence, they strongly regulate the incoming and outgoing radiation (e.g., Lohmann et al., 2016). Cloud-related processes redistribute energy in the atmosphere and control the water cycle (e.g., Bony et al., 2015). Therefore, clouds are an important component of the climate system. To better understand global warming, one must know the cloud feedbacks to global warming. Especially the shallow cumulus clouds over the oceans and in the trade-wind regimes were identified as a major source of uncertainty in climate simulations for cloud feedbacks (Bony and Dufresne, 2005; Schneider et al., 2017). Those clouds form over the vast tropical and subtropical oceans. Barbados is a particularly suitable, though not unique, location to study them, as it is a remote island in the undisturbed trades with well developed infrastructure (Medeiros and Nuijens, 2016). Shallow cumulus clouds form in various sizes and patterns, as shown in Fig. 1.1, within the lowest kilometers of the atmosphere. They are fed with moisture from the moist and warm surface layer at their base, and are capped by the dry free troposphere above. Many of these clouds are just a few hundred meters deep while others, especially those that precipitate, penetrate deeper and form outflow layers as depicted in Fig. 1.1d (see

also Nuijens et al., 2014). While condensation warms the layer the cloud forms in, evaporation of precipitation and droplets transported out of the cloud core cool the sub-cloud and detrainment layers. This sustains the trade-wind temperature inversion. The small size and low optical thickness of shallow cumuli make their impact on radiative fluxes particularly sensitive to changes in the microphysics, i.e., on the droplet scale (Turner et al., 2007). However, the radiative fluxes of a shallow cumulus field are not only controlled by the microphysics within an individual cloud, but also by the clouds' interdependency and organization on the mesoscale (Bony et al., 2020).

(a)

(b)

(c)

(d)
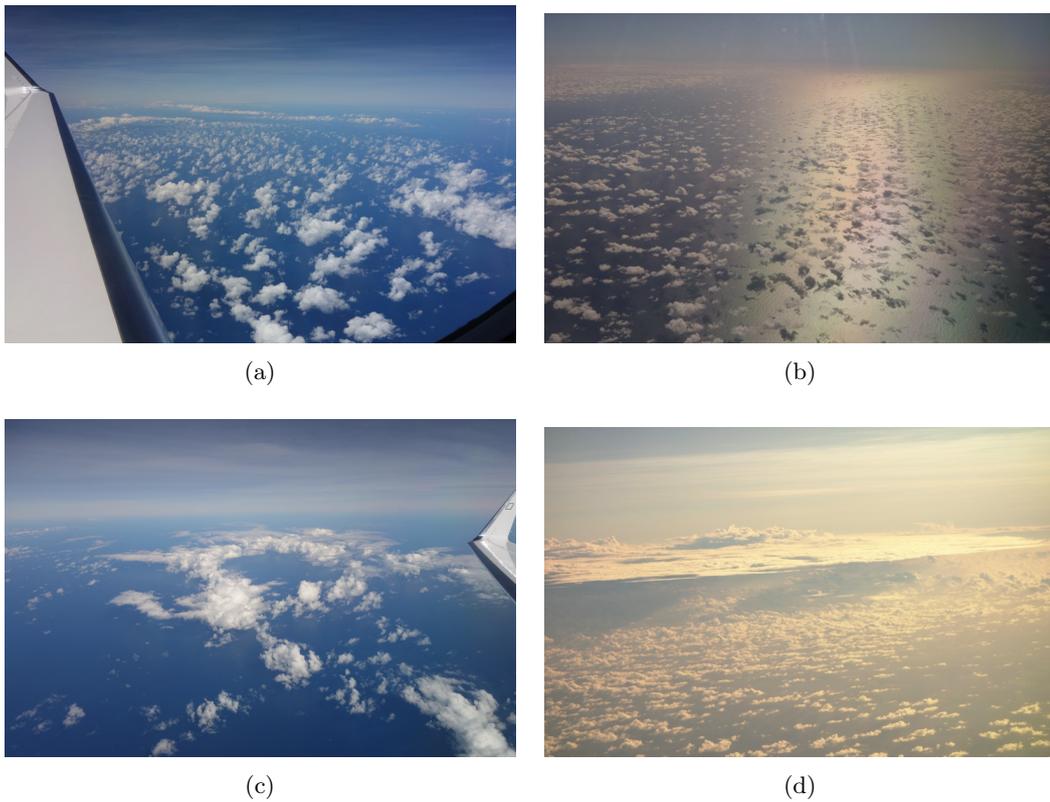
Figure 1.1.: Different sizes and patterns of shallow cumulus clouds. Scattered shallow cumulus (a, b). Shallow cumuli forming a ring around a cold pool (c). Shallow cumuli forming cloud streets in the front and a few convective updrafts flowing out into a shallow stratus cloud in the back (d).

The response of the cloud cover to a warming climate is predicted differently by current climate and process models. In general, warmer air can sustain higher amounts of water vapor especially in the boundary layer (Stevens and Bony, 2013). Also, evaporation in the marine boundary layer can increase with warming temperatures, and satellite data showed that the vertically integrated waver vapor path (IWV) already increased by $0.4\,\mathrm{kg\,m^{-2}}$ per decade on average over the ice-free oceans during

the last 30 years (Mears et al., 2018). The increase over the tropical oceans (20° S to 20° N) of $0.6\,\mathrm{kg\,m^{-2}}$ per decade is even higher. But whether increased water vapor acts as a negative feedback on global warming by also increasing the shallow cloud coverage and with it the Earth's albedo, or whether the positive feedback of moisture acting as the strongest greenhouse gas dominates, is still subject to research. While the different responses of the clouds to warming are found to explain most of the spread in climate sensitivity among different climate models, the shallow cloud fraction in process based large-eddy simulations (LES) tends to be less sensitive to environmental changes like boundary layer temperature (Rieck et al., 2012). However, other large-eddy simulation studies over large domains show that the shallow cumulus cloud fraction especially depends on how the clouds self-organize in patterns, i.e., the mesoscale organization (Jeevanjee and Romps, 2013; Vogel et al., 2016). These differences in model predictions are due to uncertainties related to assumptions and parametrizations of cloud processes on the micro scale that cannot be resolved directly, even in cloud-resolving large-eddy models. These assumptions are various (e.g., van Zanten et al., 2011) and observations are necessary to assess the models' abilities to reproduce reality.

Mesoscale organization is visible from satellite images (Stevens et al., 2020b), but also with the naked eye when flying with an aircraft. Such patterns can be evenly scattered like in Figs. 1.1a and b, form rings (Fig. 1.1c) around potential cold pools, or form cloud streets, and several convective updrafts can feed one veil-like cloud covering lots of seemingly independent clouds (Fig. 1.1d). A wide range of in-situ to remote sensing measurement principles are available to observe clouds more objectively on various size and time scales (Glackin, 2014). Most of the in-situ measurements are rather direct, but are in general limited to the availability of a measurement platform and thus have a limited spatial coverage. Remote sensing methods, in contrast, are generally more indirect as they measure the electromagnetic radiation or acoustic waves emitted, scattered, and/or absorbed by the phenomenon of interest. In atmospheric science, electromagnetic remote sensing uses a wide spectral range from ultraviolet over the visible and infrared to microwaves. Different frequencies or wavelengths are used for different purposes. Their main differences are the sensitivity to different components in the atmosphere, the opacity of the atmosphere related to sensing-depth or -range, and the technically feasible beamwidths, which correspond to spatial resolution. Visible and infrared sensors have the highest spatial resolution and can provide images with meter resolution even from $10\,\mathrm{km}$ or more distance. The atmospheric gases are mostly transparent or partially opaque to visible and infrared wavelengths, but cloud particles interact with electromagnetic radiation at these wavelengths, such that the most of the received information comes

from a thin sensor-facing layer of the cloud. In contrast, clouds are semitransparent to microwaves, such that the information content of microwave sensors comes from the whole cloud depth (e.g., Ulaby and Long, 2015). However, larger wavelengths require larger antennas limiting the spatial resolution.

Besides wavelengths, remote sensors are also categorized into active and passive sensors. Active instruments transmit radiation and receive the time-delayed scattering signal of atmospheric components. Thereby, ranging of the backscatterers is possible. Passive sensors measure the emitted radiation and radiation that is scattered from a natural source. Differential spectral absorption is exploited to separate different sources, i.e., components and ranges, with passive instruments. Remote sensing sensors can be deployed to multiple moving and stationary platforms on the ground like super sites, trucks, and ships, in the air on planes, kites, and balloons, and in space. The small size of shallow cumulus – 70 % having a length of less than 2 km in the Barbados region (Schnitt et al., 2017) – challenges spaceborne observations. The predominant existence of these clouds over the oceans complicates detailed ground-based observations. Island-based experiments like the Barbados Cloud Observatory (BCO; Stevens et al., 2015) provide useful data, although the special influence of the island has to be considered (Giangrande et al., 2019). However, airborne remote sensing can extend the spaceborne observations to finer scales, survey cloud fields and thus fill an observational gap.

The liquid water path (LWP), for example, is an important cloud characterization, as it describes the vertical integral of all liquid in an atmospheric column. However, for climate studies, only a few global LWP observational datasets exist and their mean zonal values differ by a factor of 2 with a maximum deviation in the tropical regions (Lohmann and Neubauer, 2018). The differences are partially due to different sensor designs with their advantages and disadvantages. One dataset is derived from visible to near-infrared spaceborne sensors, including the Moderate Imaging Spectroradiometer (MODIS) with a high spatial resolution, but these sensors mostly see information from cloud top (Zhou et al., 2016), while the considered microwave satellites see through the whole column, but have a coarse spatial resolution. The considered Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP; Elsaesser et al., 2017) also includes an estimation of the rain water path (RWP) to LWP ratio. However, the RWP in their study is derived with a rather simple parametrization based on the rain-shaft height estimation and microwave retrieved rain rate. Greenwald et al. (2018) reported on LWP biases of more than 50 % due to rain contributions in the MAC-LWP by using reference data from MODIS, the spaceborne CloudSat Profiling Radar (CPR; Stephens et al., 2002) and the spaceborne Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP; Winker

et al., 2007a), i.e., using additional passive and active measurements. Therefore one can ask, how a combination of active and passive microwave sensors available on a common platform could be used directly for a more elaborated RWP estimation.

Atmospheric models have to represent the effects of individual cloud and rain drops with variables that are resolved on the model grid. Vertically resolved models, for example, often consider the mass of all liquid cloud droplets within a unit volume of a cloud, which is the liquid water content (LWC) – a key parameter in many microphysical cloud parametrizations. Unfortunately, the instantaneous and direct observation of the LWC profile is difficult (Crewell et al., 2009). But, LWC can be estimated by distributing the LWP along the vertical using cloud boundary observations from a cloud radar (Frisch et al., 1998). This rather simple method resulting in a basically linearly increasing LWC can be improved by using Doppler radar spectra (Küchler et al., 2018). However, both methods were developed for ground-based radiometer measurements, but cannot be used to study precipitating clouds due to water on the radome. The radome of an airborne radiometer flying high above the raining clouds, however, is not affected by precipitation and thus its vertically integrated LWP estimates have the potential to be used during and after rain events. However, the method by Küchler et al. cannot be applied to airborne radar data, because the Doppler radar spectra are affected by Doppler broadening due to the platform movement (Mech et al., 2014). Therefore, the LWP has to be used directly instead of LWC, but cloud top and base height estimates with different cloud detection sensitivities from different instruments can be considered for a valuable synergistic model assessment. The transition from cloud to non-cloud is smooth in modeled clouds as well as in observed clouds. Modeled LWC is often mathematically different from zero due to numerical diffusion and sensors are limited by their sensitivity. Also with the naked eye it is hard to tell the exact boundary of a cloud. Therefore it is important to apply the same cloud detection thresholds on model data and observations for a fair comparison.

1.2. Historical Perspective on Tropical Campaigns

Already 65 years ago, Byers and Hall (1955) pioneered airborne tropical shallow-cumulus expeditions and found, that half of the 466 surveyed clouds with top height above 1.8 km precipitate with a strong correlation of cloud top height and precipitation probability. Byers and Hall studied clouds in the vicinity of Puerto Rico and reported temperatures at cloud top from 2 to 12.5 °C. Thus they showed that also warm over-water clouds with temperatures clearly above the freezing point precipitate at a time, when it was broadly assumed, that seeding with frozen particles is

required for precipitation to form. Subsequent initiatives like the Barbados Oceanographic and Meteorological Experiment (BOMEX; Holland, 1970) focused more on vertical transports of water vapor, heat, and momentum from the sea surface into the boundary trade-wind layer, on radiation, and on the general assessment of cloudiness. BOMEX did not include detailed cloud observations and also spaceborne observations were not ready at that time to support BOMEX with advanced cloud parameters. However, the deployment of satellites with active microwave radars like in the Tropical Rainfall Measuring Mission (TRMM), for example, allowed Short and Nakamura (2000) to study similar relations like Byers and Hall (1955) but considering the whole tropics. Short and Nakamura used TRMM data over several months and found two modes of precipitating cloud top heights with a shallow mode at about 2 km and a congestus mode near 5 km, where the rain top height of the shallow mode showed a linear relation to the rain rate. As sensitivity and spatial resolution like a 5 km footprint limit the use of spaceborne instruments for the study of individual shallow cumulus clouds, the Rain in Cumulus over the Ocean (RICO) field campaign (Rauber et al., 2007) was initiated. RICO took place during the winter trades in the western Atlantic near the Caribbean islands of Antigua and Barbuda to measure processes related to the rain formation in shallow cumuli and the subsequent modification of the cloud field by precipitation on scales from aerosol particles (µm) and cloud droplets to cloud organization (tens of kilometers). A LES reference case was created with RICO measurements to assess simulations (e.g., van Zanten et al., 2011). Van Zanten et al. showed that the 12 considered models with different microphysical parametrizations produce a variety of different cloud microphysical structures and emerging surface rainfall. The representation of the vertical structure of cloudiness and cloud water, and rain were plausible in all simulations. However, the simulated case is a composite constructed from several days during a less intense phase of the fieldwork. Further, airborne observations were biased to larger clouds by the desire to sample active clouds. Therefore, the sampled RICO data impeded further assessment of the model differences.

In order to better characterize and understand the distribution and structure of shallow convective clouds and precipitation in the trade-wind region and in the cold sector of developing winter storms, the Next-generation Aircraft Remote sensing for VALidation studies (NARVAL; Klepp et al., 2014) took place in December 2013 and January 2014. NARVAL-South and NARVAL-North were the first two campaigns to demonstrate the capability of the novel German research High Altitude and LOng range research aircraft (HALO; Krautstrunk and Giez, 2012) as a remote sensing platform for observing the state of the atmosphere with a focus on moist processes and water distribution. During NARVAL, HALO was equipped with a comprehen-

sive suite of active and passive remote sensing instruments ranging from the visible to microwave spectrum which supplement each other. The microwave spectrum was covered by HAMP. This NARVAL payload on HALO was also designed as a testbed for novel retrievals and synergetic measurement systems which could be later used on satellites. NARVAL-South targeted the trade-wind region east of Barbados and NARVAL-North targeted North Atlantic post frontal systems. In August 2016, the NARVAL series was extended by NARVAL2 (Stevens et al., 2019) using the similarly equipped HALO remote sensing platform to contrast the observations of the dry-season trades in December 2013 with observations of the warm-season trades. During NARVAL-South and NARVAL2, HALO was deployed to the eastern most Caribbean island of Barbados to probe the undisturbed trade-winds to the East and to make use of the long-term ground-based observations at BCO. NARVAL2 was also used to test methods to study the large-scale vertical motion with dropsondes released in circles. The area-averaged vertical motion is an important driver of the cloud environment and can be derived from the horizontal mass divergence estimated from the wind measurements of the sondes along the circle (Bony and Stevens, 2018). Following NARVAL2, the same instrumentation was used during the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX; Schäfler et al., 2018) to study diabatic processes in mid-latitude frontal systems over the North Atlantic in September and October 2016.

To follow up on the air-sea interaction studied during BOMEX, cloud and rain process investigated during RICO, and to facilitate the ultra-low-orbit satellite-like remote-sensing cloud-observatory established during NARVAL, the field campaign to Elucidate the Couplings between Clouds, Convection, and Circulation (EUREC4A; Bony et al., 2017) took place at the beginning of 2020. EUREC4A was the most recent deployment of HAMP. A timeline overview of the campaigns during which HAMP was deployed to HALO is shown in Fig. 1.2.
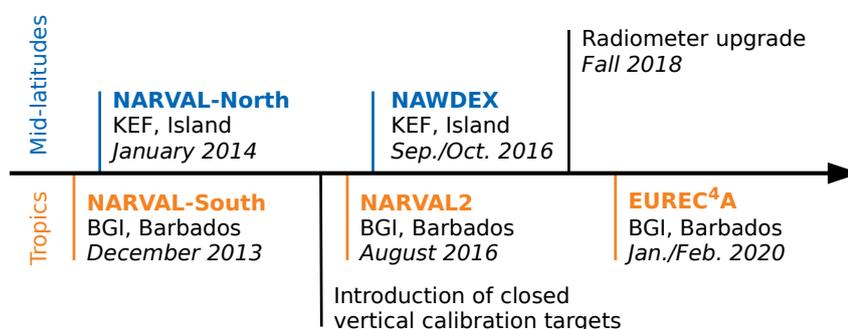
Figure 1.2.: Timeline of HAMP deployments to HALO. HALO was based in the Keflavík Airport (KEF) and the Grantley Adams International Airport (BGI) for the mid-latitude and tropical campaigns, respectively.

1.3. HALO as Flying Cloud Observatory

The NARVAL instrumentation of HALO includes a dropsonde system and a suite of active and passive remote-sensing instruments as depicted in Fig. 1.3. The dropsonde system allows for human-assisted launches of radiosounding-like probes attached to small parachutes to measure the wind and thermodynamic profiles. The remote sensing suite includes the water vapor and aerosol backscatter lidar WALES (WAter vapor Lidar Experiment in Space; Wirth et al., 2009) and spectral radiometer systems measuring up- and downwelling radiation at visible to near-infrared wavelengths. These instruments are complemented by the HALO Microwave Package (HAMP; Mech et al., 2014) comprising a cloud radar, and a suite of microwave radiometers (Fig. 1.4) at millimeter to centimeter wavelengths. The radar operates at $36\,\mathrm{GHz}$ and the radiometers have a total of 26 channels from 20 to $200\,\mathrm{GHz}$. This wide spectral range and high number of channels makes HAMP unique compared to other airborne and spaceborne microwave instruments (Wendisch et al., 2013; Ulaby and Long, 2015).
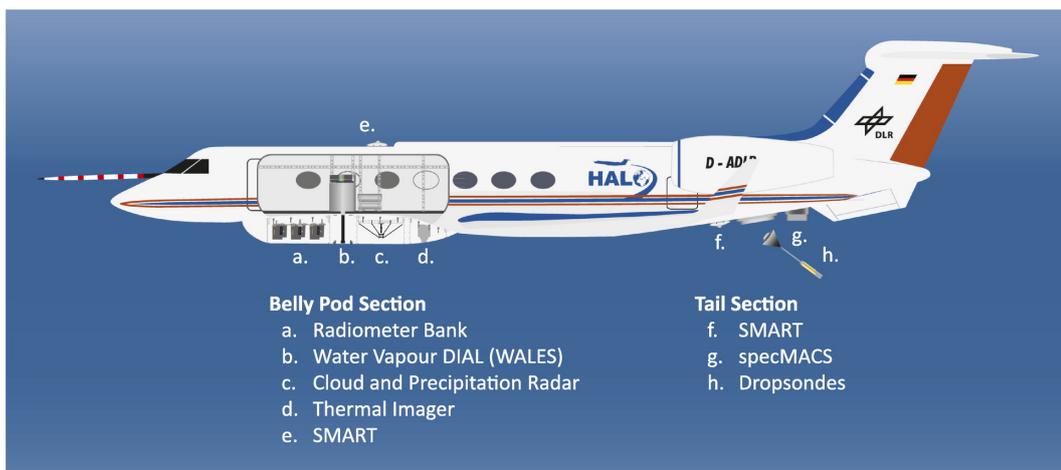
Figure 1.3.: Cross section of HALO with the NARVAL payload. Figure derived from Stevens et al. (2019). Courtesy of the American Meteorological Society.

The HAMP radar has been used, for example, by Oertel et al. (2019) to investigate embedded convection in warm conveyor belts associated with a mid-latitude cyclone and by Wolf et al. (2019b) to assess the ECMWF (European Centre for Medium-Range Weather Forecasts) radiation scheme analyzing another mid-latitude system during NAWDEX. Furthermore, the HAMP radar and radiometers have been used in a comparison study for different cloud masks during NARVAL-North by Albern (2014). Schnitt et al. (2017) developed the first HAMP radiometer retrievals of LWP and IWV using a simplified training dataset based on dropsonde data for NARVAL-South. Wolf et al. (2019a) used retrievals from HAMP, that were developed following

(a)

(b)

Figure 1.4.: HALO with closed belly pod (a) and remote sensing instruments inside the belly pot underneath the HALO fuselage (b). From left (front) to right in (b): HAMP-G, HAMP-KV, HAMP-WF, lidar window, and HAMP cloud radar antenna. Picture (b) by Mech et al. (2014) is distributed under Creative Commons Attribution 3.0 License.

the method by Schnitt et al. (2017) for NARVAL2, to improve the estimation of cloud droplet number concentration of trade wind cumuli based on spectral solar radiation measurements.

In general, the characterization and calibration of an instrument are unquestionable prerequisites for geophysical retrievals. The HAMP radar has been calibrated by Ewald et al. (2018). However, an in-depth characterization of the radiometer, a quality control of the calibration, and a refinement and assessment of the retrievals

have been missing. The calibration of microwave radiometers is especially sensitive as the radiance of the measured thermal emission is rather low compared to the emission of the radiometer itself. The accurate microwave radiometer calibration is a research subject on its own (e.g., Hardy, 1973; Han and Westwater, 2000; Küchler et al., 2016).

The signals detected by the HAMP radiometers are, like for every passive sensors, the sum of all sources and sinks of the radiation along the observation path. Different channels are affected by different features of the atmosphere. Forward simulators can simulate HAMP measurements given the state of the atmosphere and using radiative transfer equations. However, the challenge is to generalize and invert the forward simulator to retrieve geophysical quantities from the simultaneous signals seen by several sensors. Novel and specialized retrievals have to be developed to get the best of HAMP with its unique set of channels and supported by the special suite of further observations from HALO.

1.4. Research Questions and Outline

This thesis makes use of the unique capabilities of HAMP with a focus on the passive component and is structured around three major topics: instrument characterization, retrieval development, and application. Primarily, the data from the NARVAL-South and NARVAL2 campaigns are used, focusing on tropical clouds and convection. The latest observations from EUREC^{4}A are considered as an outlook. The thesis starts with an introduction into the theoretical background on microwave remote sensing, including the spectral features covered by HAMP as well as its measurement quantity *brightness temperature* (BT) in Chapter 2. That chapter also introduces the forward simulation of HAMP using the radiative transfer simulator PAMTRA. The subsequent three chapters address the three major topics, which are further refined by the following research questions.

1. How accurate are the HAMP brightness temperature BT measurements? This question involves a ground-based and airborne instrument characterization of the BT noise, its drift, and an assessment of the absolute calibration.

2. How can geophysical quantities be retrieved from HAMP observations? The focus is on the liquid hydrometeors, i.e. LWP and RWP, but also the cloud environment in terms of water vapor, i.e. IWV, matters. What is the uncertainty of LWP, RWP, and IWV retrievals from HAMP? Which differences between

the dry season in December 2013 (NARVAL-South) and the wet season in August 2016 (NARVAL2) can be identified from the retrieved products?

3. How can HAMP be used to assess cloud-resolving model simulations and how good are the considered models in representing the convective dry-season trade-wind clouds? Two different cloud-resolving simulations are confronted with HALO observations to tell their differences with respect to clouds and precipitation.

The instrument accuracy is studied at the processing level of BTs in Chapter 3 by analyzing measurements on reference targets as well as forward-simulated synthetic measurements based on dropsonde data. The analysis provides an uncertainty estimate and an offset correction with are subsequently considered in the retrieval development. Cloud-resolving models and forward simulations are used to developed novel retrievals of IWV, LWP, and RWP. Their description, accuracy assessment, and summary in a seasonal comparison are embraced in Chapter 4. The application of HAMP with its active and passive component is completed by using it to assess the clouds simulated by two cloud-resolving simulations in Chapter 5. Finally, the conclusion of using HAMP to characterize maritime trade-wind convection and an outlook on future possibilities are presented in Chapter 6.

Chapter 2.

Airborne Microwave Radiometry

This chapter gives a brief background on the principles that apply to airborne microwave remote sensing in general and on HAMP in particular.

2.1. Microwave Cloud Property Measurement

An advantage of microwave remote sensing compared to visible or infrared methods is the semitransparency of cloud droplets to microwaves such that microwave methods can also be used in cloudy scenes to retrieve information on the whole vertical column (Smith et al., 1994). The spectrum of microwave absorption coefficients (Fig. 2.1) shows how oxygen and water vapor affect the microwave radiation differently. Emission lines of water vapor and oxygen are broadened by the atmospheric pressure and thus influence certain frequency bands with decreasing intensity further away from the line centers. As scattering can be neglected due to the small molecule sizes compared to the wavelength, extinction is purely due to absorption and thus absorptivity equals emissivity by Kirchhoff's Law (Wendisch et al., 2013). Based on the absorption features or emissivity complexes, the sensing depth into the atmosphere of a passive sensor with a narrow channel bandpass depends on the position of the channel frequency relative to the absorption complex (e.g., Ulaby and Long, 2015). This means different microwave channels get most of their information from different altitudes and some are more sensitive to changes of water vapor while others are more sensitive to temperature changes due to changes in oxygen emission. Therefore, HAMP was designed with 26 channels related to 5 different spectral features as described in a following section.

The non-resonant emission from liquid droplets adds to the gaseous emission and depends on the amount of liquid and its temperature. This additional emission can be observed best in the so-called window channels around, e.g., 30 or 90 GHz, where gaseous emission is low. Frozen hydrometeors have a negligibly small emissivity, but scatter upwelling radiation and therefore reduce radiation observed from above (Kneifel et al., 2010). The radiation depression through scattering of ice generally increases with increasing frequency and particle size. The emission from the Earth's surface is also seen by downward looking air- and spaceborne sensors as the total
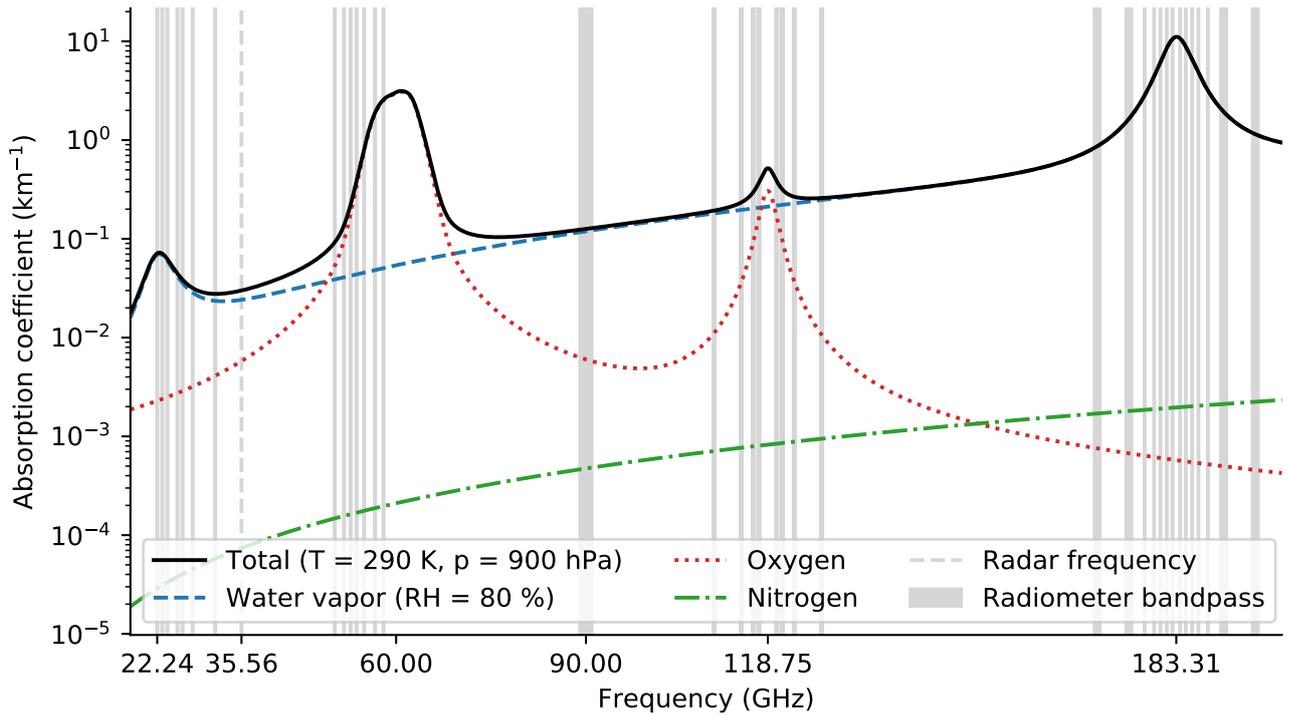
Figure 2.1.: Microwave absorption spectrum due to water vapor (H_2O), gaseous oxygen (O_2), and gaseous nitrogen (N_2) based on Rosenkranz (1998), Turner et al. (2009), and Liljegren et al. (2005).

atmosphere is especially transparent at microwave frequencies below $100\,GHz$ and in window channels. However, the microwave emissivity of the ocean is about 0.4 to 0.6 such that the ocean provides a relatively cold background signal compared to atmospheric emissions. The ocean emissivity depends mostly on the frequency, roughness, foam formation, salinity, temperature, and observation angle and is therefore spatially rather homogeneous. In the end, the signal seen by a passive sensor is the sum of all sources and sinks of the radiation along the observation path.

2.2. Radiative Transfer Equation

The quantitative amount of radiation received at a sensor, i.e., the radiative flux, can be derived from the radiative transfer equation (RTE). In general, the RTE describes the change of radiation as it is propagating through a medium like the atmosphere by interacting with its constituents like gas molecules or hydrometeors. The radiative transfer theory is well explained in textbooks (e.g., Liou, 2002; Petty, 2006) and is briefly described in the following.

Electromagnetic radiation propagating through the atmosphere is dampened by extinction. The extinction is quantified by the extinction coefficients β_e (in m^{-1}) and is the result of absorption (β_a) and scattering (β_s) as

$$\beta_e = \beta_a + \beta_s. \tag{2.1}$$

Absorption is the energy uptake by the medium and scattering the change of the propagation direction of an incident wave. β_a and β_s are intrinsic attributes of the medium and in principle functions of frequency and atmospheric temperature, pressure, and composition. For microwaves, the most important and variable part in the atmospheric gas composition is the water vapor. Also homogeneously distributed hydrometeors can be described with β_a and β_s coefficients. The following assumes the monochromatic case and that scatterers are small compared to the wavelength λ. This is the case for the gas molecules and cloud droplets at the considered microwave wavelengths and simplifies the RTE to one dimension as scattering can be neglected ($\beta_s = 0$).

The spectral radiance I_ν (in $W\,sr^{-1}\,m^{-2}\,Hz^{-1}$) of the incident radiation of frequency ν is reduced by absorption when propagating through an infinitesimal thin layer of air with thickness ds by the amount

$$dI_{\nu,\text{absorption}} = -\beta_{a,\nu}\, I_\nu\, ds. \tag{2.2}$$

The subscript ν is omitted in the following. At the same time, I increases through thermal emission of the layer. The emission is related to the air temperature T. According to Kirchhoff's Law, the emissivity of the air in thermodynamic equilibrium is equal to the absorptivity and the emitted radiance dI_{emission} is proportional to the black body radiance $B(T)$, which is explained later on.

$$dI_{\text{emission}} = \beta_a\, B(T)\, ds. \tag{2.3}$$

Combining the equations above, the monochromatic non-scattering form of the RTE describing the change of radiance is

$$dI = dI_{\text{absorption}} + dI_{\text{emission}} = -\beta_a\, I\, ds + \beta_a\, B(T)\, ds$$
$$= \beta_a\, (-I + B(T))\, ds. \tag{2.4}$$

This equation can be integrated and a solution can be found for the radiance $I(s)$ at a point s by knowing the radiance at a reference point s_0. The solution is

$$I(s) = I(s_0)\, e^{-\tau(s_0,s)} + \int_{s_0}^{s} \beta_a(s')\, B(T(s'))\, e^{-\tau(s',s)}\, ds', \tag{2.5}$$

with the optical depth $\tau(s_1, s_2)$ between s_1 and s_2 being defined as

$$\tau(s_1, s_2) = \int_{s_1}^{s_2} \beta_a(s)\, ds. \tag{2.6}$$

A black body is an idealized physical body that absorbs all incident radiation regardless of frequency or incident angle. In thermal equilibrium, i.e., at a constant temperature, the emitted radiation $B(T)$ is described by Planck's law as

$$B(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_{\mathrm{B}}T}} - 1}, \tag{2.7}$$

where h is the Planck constant ($h = 6.626\,070\,15 \times 10^{-34}\,\mathrm{J\,s}$), c is the speed of light in the medium ($c \approx 3 \times 10^8\,\mathrm{m\,s^{-1}}$ in air), and k_{B} is the Boltzmann constant ($k_{\mathrm{B}} = 1.380\,649 \times 10^{-23}\,\mathrm{J\,K^{-1}}$). The black body radiation is reduced by the surface-material-dependent emissivity factor $\epsilon < 1$ for real bodies like the sea surface. Such bodies are called gray. As a gray body does not absorb all incident radiation, the remaining part $(1 - \epsilon)$ is scattered or transmitted.

The theory above provides enough background to derive a simple form of the RTE for the airborne application. In this example a nadir looking radiometer with a narrow field of view is assumed in a non-scattering atmosphere. The radiometer is flying at the altitude A above the calm ocean surface at altitude $a = 0$, for which specular reflection is assumed. In this setup, the radiometer receives not only an atmospheric signal but also the emission of the sea surface with temperature T_{sea}. However, as the sea is a gray body ($\epsilon \approx 0.4 \ldots 0.6$), it also reflects radiation and cannot be used as background at s_0. Instead the downwelling radiance at the sea surface I_{sea} has to be considered. This can be directly derived from Eq. 2.5 knowing the cosmic microwave background radiation at $s_0 = \infty$ which emits with a temperature of about $T_\infty = 2.7\,\mathrm{K}$ (Noterdaeme et al., 2011). Therefore,

$$I_{\mathrm{sea}} = T_\infty\, e^{-\tau(a,\infty)} + \int_a^\infty \beta_{\mathrm{a}}(s')\, T(s')\, e^{-\tau(a,s')}\, ds' \tag{2.8}$$

where $I(\infty) = B(T_\infty)$.

The radiance $I(A)$ received at the airborne radiometer is the sum of the upwelling radiation emitted by the atmosphere between A and the sea surface a, the sea surface emission and the radiation reflected at the sea surface. The latter two terms are dampened by the atmospheric absorption between a and A by $e^{-\tau(a,A)}$. Thus

$$\begin{aligned}
I(A) = \int_a^A \beta_{\mathrm{a}}(s')\, B(T(s'))\, e^{-\tau(s',A)}\, ds' \\
+ \epsilon\, B(T_{\mathrm{sea}})\, e^{-\tau(a,A)} \\
+ (1 - \epsilon)\, I_{\mathrm{sea}}\, e^{-\tau(a,A)}.
\end{aligned} \tag{2.9}$$

In the microwave regime, $\frac{h\nu}{k_{\mathrm{B}}T} \ll 1$. Thus, the Rayleigh-Jeans approximation can be used which simplifies Eq. 2.7 to

$$B(T) = \frac{2k_{\mathrm{B}}\nu^2}{c^2}\, T. \tag{2.10}$$

This equation shows the linear relation between the black body radiation and the physical temperature. In microwave remote sensing, it is convenient to use this linearity to translate the incident radiance into its radiatively-equivalent black body temperature called brightness temperature (BT) by inverting $B(T)$. In this thesis, BT is defined by inverting the exact Planck equation (Eq. 2.7), but also the inversion of the Rayleigh-Jeans (Eq. 2.10) approximation is common in the literature. Equation 2.9 can be simplified with the Rayleigh-Jeans approximation to an equation that is related to physical temperatures directly:

$$
\begin{aligned}
\mathrm{BT}(A) = \int_a^A & \beta_\mathrm{a}(s')\, T(s')\, e^{-\tau(s',A)}\, ds' \\
& + \epsilon\, T_\mathrm{sea}\, e^{-\tau(a,A)} \\
& + (1-\epsilon)\, e^{-\tau(a,A)} \left(T_\infty\, e^{-\tau(a,\infty)} + \int_a^\infty \beta_\mathrm{a}(s')\, T(s')\, e^{-\tau(a,s')}\, ds' \right).
\end{aligned}
\tag{2.11}
$$

2.3. Radiative Transfer Simulator

The radiative transfer equation (Eq. 2.11) presented in the section before is the idealized theory to calculate microwave BTs. Much complexity is hidden in parameters like β and ϵ which depend on atmospheric and surface attributes. Furthermore, the scattering by larger hydrometeors should be considered as well. Therefore, the comprehensive Passive and Active Microwave TRAnsfer package (PAMTRA; Mech et al., 2020) is used to forward simulate microwave radiometer measurements. PAMTRA components provide the necessary estimates of, for example, β and ϵ.

Within PAMTRA, gaseous absorption coefficients β_a of oxygen, water vapor, and nitrogen are calculated according to Rosenkranz (1998) with corrections of the water vapor continuum absorption by Turner et al. (2009) and the modification of the 22.235 GHz water vapor line width proposed by Liljegren et al. (2005) as depicted in Fig. 2.1. The ocean surface emissivity (ϵ) and reflectivity are calculated using the FAST microwave Emissivity Model version 5 (FASTEM5; Liu et al., 2011). FASTEM5 is a modification of the Fresnel coefficients and considers corrections for ocean surface roughness and foam building as a function of wind speed.

2.4. Microwave Radiometer Design

The objective of a microwave radiometer is to measure the radiative flux at a specified microwave frequency. The radiometer antenna defines the field of view, from which the radiation is received. Generally, a narrow field of view is desired for high spatial

resolution. The power of the radiative flux can be interpreted as thermal noise emitted by the observed scene, and is often expressed in terms of the flux-equivalent BT (cf. Eq. 2.10). Thus, the power output of an antenna can be related to the equivalent antenna temperature T_A. In case of an ideal, lossless antenna, T_A is the average BT of the field of view weighted by the antenna response pattern (Ulaby and Long, 2015). As the power of natural microwave radiation on Earth is low, the power is amplified with the amplification gain G before it can be measured by a detector. The most simple design of a microwave radiometer is the total power radiometer as depicted in Fig. 2.2.
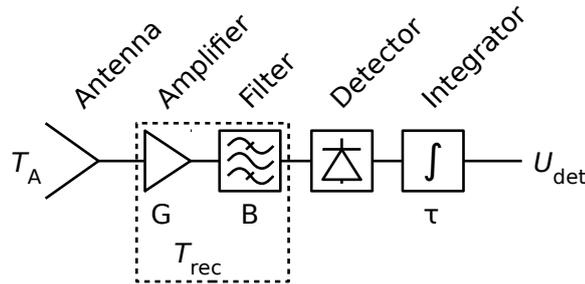
Figure 2.2.: Principal schematic of a one-channel total power radiometer. After Skou and Vine (2006) and Ulaby and Long (2015).

In a real radiometer, its components will generate noise, which can be expressed with the receiver noise temperature T_{rec} (Skou and Vine, 2006). Therefore, the detector output voltage U_{det} can be expressed as

$$U_{det} = G\left(T_A + T_{rec}\right). \tag{2.12}$$

The noise fluctuation is smoothed by a final integrator over the observation period τ. Two reference points with different T_A and corresponding U_{det} are necessary to determine G and T_{rec}. Subsequently, Eq. 2.12 can be inverted to obtain T_A from a measured U_{det}.

The simple total power radiometer assumes constant G and T_{rec} in Eq. 2.12. However, a stability of G of a thousandth is required when a signal of $1\,\mathrm{K}$ should be measured on top of $T_A + T_{rec}$ which is on the order of $1000\,\mathrm{K}$. Therefore two additional components can be added to the schematic as shown in Fig. 2.3 following Rose (2009) and Radiometer Physics GmbH (2015). First, a so called Dicke switch is installed as close as possible to the antenna. The Dicke switch can switch the radiometer input between the antenna and a resistor, with has a radiative-equivalent temperature T_{DS}. As waveguides guide the received radiation, the Dicke switch has to be realized as waveguide circuit. Such switch can potentially leak some of the

received power from the antenna into the detector when switched to the resistor. Therefore, the detector output of the closed Dicke switch is

$$U_{\mathrm{DS}} = G\left((1-\alpha)\,T_{\mathrm{DS}} + \alpha\,T_{\mathrm{A}} + T_{\mathrm{rec}}\right) \tag{2.13}$$

with the Dicke switch leakage parameter α, which is typically below 0.005. The radiometric Dicke switch temperature T_{DS} is close to its physical temperature T_{DSp} and can be described as

$$T_{\mathrm{DS}} = T_{\mathrm{DSp}} + \Delta T_{\mathrm{DS}}. \tag{2.14}$$

The second additional component is the noise injection. When the noise diode is connected via the noise injection, extra noise power is added onto the detected signal. The noise diode emits with an equivalent temperature of T_{noise} on the order of 1000 K or higher. The measured voltage on a scene with T_{A} is then

$$U_{\mathrm{A+noise}} = G\left(T_{\mathrm{noise}} + T_{\mathrm{A}} + T_{\mathrm{rec}}\right). \tag{2.15}$$
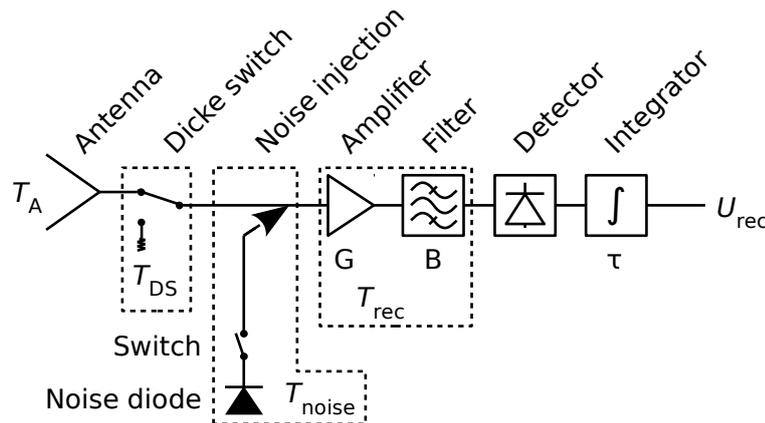
Figure 2.3.: Principal schematic of a one-channel radiometer with Dicke switch and noise injection. After Rose (2009).

Equations 2.13, 2.14, and 2.15 introduce the additional three parameters α, ΔT_{DS}, and T_{noise} which have to be calibrated. T_{DSp} can be monitored with a physical thermometer. ΔT_{DS} can be determined using Eqs. 2.13 and 2.14 by ignoring the influence of α. Such condition is established when T_{A} of a reference target is close to T_{DS}. When the radiometer is operated approximately at ambient temperature, it is convenient to use a hot target at ambient temperature. Subsequently, α can be determined from Eq. 2.13 with a target, that is significantly different from T_{DS}. Such a target can be a cold target, which is for example cooled with liquid nitrogen. T_{noise} can be derived from Eq. 2.15 after determining G and T_{rec} from Eq. 2.12 by using two known reference pairs of U_{det} and T_{A}.

T_{noise} is the most significant parameter of the three additional parameters. Noise diodes with much better stability compared to the amplifier gain can be built. Therefore, injecting known noise and closing the Dicke switch can be used to constantly calibrate the values of G and T_{rec} during operation. To do so, the voltage is measured at a scene with alternately closing the Dicke switch and injecting the noise, such that Eqs. 2.13 and 2.15 can be solved for G and T_{rec} by assuming T_{A} being constant for a short time.

2.5. The HALO Microwave Package (HAMP)

HAMP is a combined active and passive microwave instrument designed for the airborne operation in the belly pod of the HALO research aircraft. It was developed to study cloud and precipitation processes, test retrieval algorithms, motivate future satellite instrument proposals, and to validate and better understand already operating spaceborne instruments. HAMP comprises a 35.56 GHz Ka-band cloud radar and a suite of microwave radiometers (MWRs) measuring in five frequency bands between 22.24 and 183.31 ± 12.5 GHz. The radar antenna as well as the three radiometer modules are mounted in the belly pod under the front part of the HALO fuselage in nadir pointing orientation as depicted in Figs. 1.3 and 1.4. HAMP was operated at flight altitudes ranging between 6.4 and 15.0 km while HALO was flying with an average speed above ground of 237 and 207 m s^{-1} during NARVAL-South and NARVAL2, respectively. The radiometers were custom-manufactured by Radiometer Physics GmbH (RPG) and technical details are provided by Rose (2009). Mech et al. (2014) describes the design of HAMP and illustrated its potential. This section gives a brief overview on the passive part of HAMP with the features important for this thesis. The term "HAMP" will in general be used to refer to the three MWRs forming the passive part of HAMP, whereas the active part of HAMP will either be referred to as "active" or as "radar".

The passive part of HAMP is a microwave radiometer composed of three independent modules. Their main and name-giving distinction are the frequency bands observable with each module. The module in the front of the belly pod is called HAMP-WF and measures in a W-band window channel and in four double-sideband F-band channels along the 118.75 GHz oxygen line. The center module called HAMP-KV operates in seven K-band channels along the water vapor line between 22.24 and 31.40 GHz, and in seven V-band channels along another oxygen line from 50.3 to 58.0 GHz. The HAMP-G module in the rear has seven double-sideband channels around another water vapor line at 183.31 GHz in the G band. The center frequencies and bandpasses of each channel are presented in Table 2.1. The modules

are referred to as "HAMP-x" here instead of "HALO-x", which is the term Mech et al. (2014) use as a deviation from the manufacturer's nomenclature (Rose, 2009), which is HALO-KV, HALO-11990, and HALO-183 for HAMP-KV, HAMP-WF, and HAMP-G, respectively.

Table 2.1.: HAMP radiometer attributes and receiver noise temperatures (T_{rec}) (Rose, 2009; Mech et al., 2014).

Band	Channels (GHz)	Bandwidth (MHz)	Abs. feat.	T_{rec} (K)
K	22.24, 23.04, 23.84, 25.44, 26.24, 27.84, 31.40	230	H_2O	380–520
V	50.30, 51.76, 52.8, 53.75, 54.94, 56.66, 58.00	230	O_2	570–720
W	90.0	2000	window	≈ 1000
F	$118.75 \pm 1.4, 2.3, 4.2, 8.5$	400	O_2	1800–2300
G	$183.31 \pm 0.6, 1.5, 2.5, 3.5, 5.0$ $183.31 \pm 7.5, 12.5$	200 1000	H_2O	1800–2800

The double-sideband channels measure the averaged BT at two frequencies with symmetric offset around the center frequency, i.e., around 118.75 and 183.31 GHz in the F and G band, respectively. This is realized by a superheterodyne receiver that tunes the received frequency ν_{RF} down to an intermediate frequency ν_{IF} using a subharmonic mixer and a local oscillator at central frequency ν_{LO}. Subsequent bandpasses filters centered at ν_{IF} filter the mixed signal such that the signals with frequencies around $\nu_{LO} + \nu_{IF}$ and $\nu_{LO} - \nu_{IF}$ are combined. Further details can be found in the literature (e.g., Skou and Vine, 2006; Ulaby and Long, 2015).

The four antennas of HAMP-KV and -WF are realized by corrugated feed-horn aperture-lens combinations. HAMP-G uses an off-axis parabolic mirror in addition to the feed-horn antenna. This mirror can rotate to change the beam direction of the HAMP-G antenna on a temperature monitored internal target. This mirror and internal target replace the magnetically controlled Dicke switches (Sect. 2.4), which can close the receiver waveguides in the other modules. In addition, a noise diode signal can be injected into the receiver path in all modules. The noise diodes emit at a constant equivalent temperature between 1000 and 1500 K depending on the channel. The noise diode and Dicke switch or turning the mirror to the internal

target are used for a frequent and continuous update of the gain and receiver noise temperature.

After NARVAL2 and NAWDEX, HAMP was upgraded for improved performance during EUREC[4]A. This upgrade included a software update for a higher temporal resolution of 4 instead of 1 Hz. Also, the radiometers were upgraded to RGPs latest G5 technology which includes the change from a Dicke type radiometer to a noise-switching radiometer. The reader is referred to Skou and Vine (2006) for further details on the radiometer principles. Furthermore, the outermost G-band channel at 183.31 ± 12.5 GHz was removed with the upgrade.

2.6. HAMP's Sensitivity to the Atmosphere and Sea Surface

PAMTRA can be used so simulate HAMP measurements, given an atmospheric profile and sea surface attributes. This section demonstrates the sensitivity of the different HAMP channels to changes in the clear-sky atmosphere and input. The influence of liquid hydrometeor emissions will be discussed in Sect. 4.2.2.

The forward-simulated BT spectrum with unmodified input serves as reference and is depicted in Fig. 2.4. The related thermodynamical profile was measured with the sonde released from an altitude of 9633 m at 18:59:12 UTC on August 12, 2016. The BT spectrum reflects the features of the absorption coefficients (Fig. 2.1). This means, low BTs are simulated in regions of low absorption coefficients like in the K band. The emission of water vapor generates a local maximum at about 22 GHz of about 184 K. The BT increases with increasing absorption as the radiometer receives more atmospheric emission above the less-emitting sea surface. However, BT at frequencies close the center of the absorption features at 60, 119, and 183 GHz decreases again as the atmosphere becomes so opaque that all the received radiation originates in layers close to radiometer. These layers are colder than the lower layers. The higher emission of lower layers is almost completely absorbed before reaching the simulated radiometer.

To test the sensitivity of HAMP channels to the PAMTRA input, the input parameters are altered in individual simulations as summarized in Table 2.2. Modifications of gaseous emissions by modifying the air temperature and humidity inputs change the BT spectra (Fig. 2.5) as it can be expected from the gaseous absorption features (Fig. 2.1). This means, increased humidity increases the humidity absorption coefficient. Thus, increased humidity increases BTs at the water vapor absorption line in
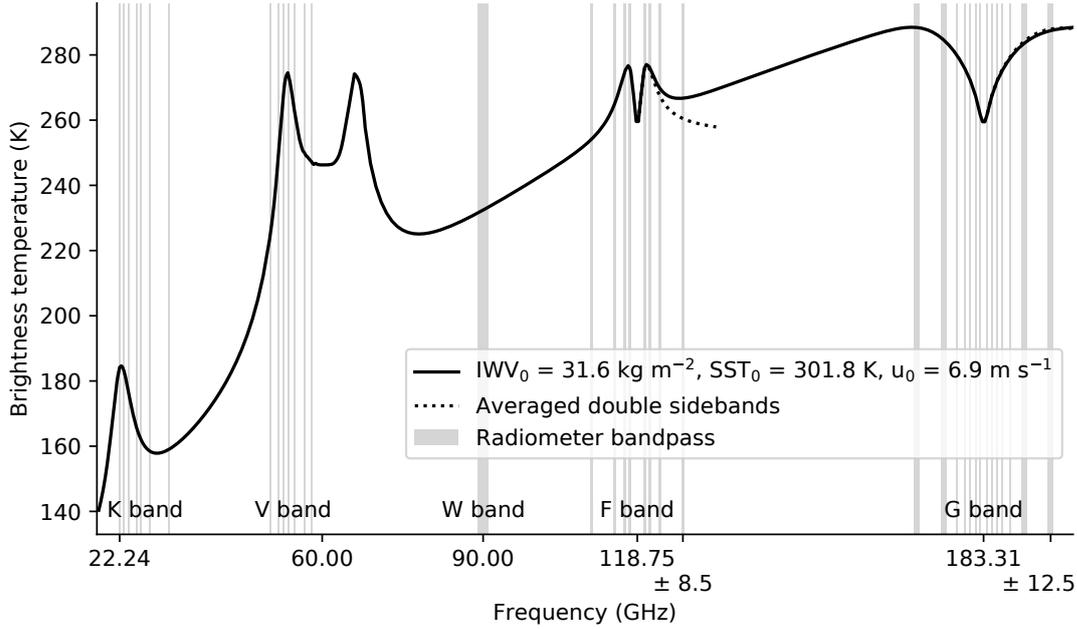
Figure 2.4.: Microwave brightness temperature spectrum simulated for a nadir pointing radiometer at $9633\,\mathrm{m}$ altitude. Input data is from the dropsonde profile of the sonde released during the third NARVAL2 flight at 18:59 UTC on August 12, 2016. The integrated water vapor (IWV_0) of the profile, the assumed sea surface temperature (SST_0), and near surface wind (u_0) are given in the legend.

the K band and in low-frequency V-band channels with little oxygen absorption, the W-band window, and the transparent outer F-band channels due to the water vapor continuum emission. An increase of the whole relative humidity profile by 0.05, increases the innermost K-band channel by $5.5\,\mathrm{K}$ and the $31.4\,\mathrm{GHz}$ channel by $1.8\,\mathrm{K}$. The lowest-frequency HAMP-V channel at $50.3\,\mathrm{GHz}$, the HAMP-W channel, and the outermost HAMP-F channel would increase by 1.5, 4.3, and $3.7\,\mathrm{K}$, respectively. The G-band BTs decrease with increased humidity as the sensing depth decreases with lower opacity due to more water vapor. An increase of the relative humidity by 0.05 would decrease channels near the G-band line center stronger ($-3.4\,\mathrm{K}$) than those on the wings ($-1.2\,\mathrm{K}$).

Decreasing temperature, decreases the BT at channels with high oxygen absorption like in the V-band and central F-band channels. But also the G-band BTs decrease as the temperature of water vapor molecules is decreasing as well. For example, decreasing the temperature by $2\,\mathrm{K}$ decreases the strongest absorption in V- and F-band channels by 2.1 and $1.8\,\mathrm{K}$, respectively, while leaving the least absorbing channels almost unaffected (0.2 and $0.0\,\mathrm{K}$). HAMP-G BTs decrease by 2.0 to $2.2\,\mathrm{K}$ and interestingly, some HAMP-K channels would decrease by up to $0.3\,\mathrm{K}$ while

Table 2.2.: Modifications of the input to test the dropsonde-forward-simulation sensitivity in Fig. 2.5.

Experiment	Explanation
Reference	Unmodified profiles of temperature, relative humidity, pressure, altitude, sea surface temperature (SST), and surface wind.
Humidity	Decrease and increase relative humidity (RH) in all layers by 0.10 (-10%) and 0.05 ($+5\%$), respectively. Minimum of relative humidity is 0.
Temperature	Decrease temperature (T) in all layers by 2 and 4 K. Keep Water vapor mixing ratio constant by modifying RH.
Wind	Set surface wind (u) to 0 and $10\,\mathrm{m\,s^{-1}}$.
SST	Decrease and increase SST by 4 K.

others would slightly increase by up to 0.3 K and the HAMP-W channel would increase by 0.4 K as well. The BT increase is probably related to a slight increase of the water vapor absorption coefficients. Changes of temperature or humidity with opposite sign would result in BT changes of similar magnitude with opposite sign.

The influence of the surface roughness controlled by the surface wind is most prominent in the transparent channels. Calm wind conditions would reduce the HAMP-K BTs by 2.4 K (at 22.24 GHz) to 3.4 K (at 31.4 GHz), HAMP-V BTs by up to 2.1 K in the most transparent channels, HAMP-W by 2.5 K, HAMP-F by 0.2 K in the innermost and 1.5 K in outermost channels, and HAMP-G by less than 0.1 K. The BT decrease is related to a decreased ocean emissivity in the surface model FASTEM5 as less formation of foam, which has in general a high emissivity (Kazumori et al., 2008), is assumed at low wind speed. There is also a weak influence of the sea surface temperature (SST) on BT of less than $0.25\,\mathrm{K\,K^{-1}}$ in the transparent channels.

The sensing depth into the atmosphere of each channel depends on the position of the channel relative to the spectral absorption features. The sensitivity of the HAMP channels for differential temperature and humidity variations is calculated by Mech et al. (2014) and is shown in Fig. 2.6. The sensitivities are also called weighting functions (WFs) and were calculated for the US 1976 Standard Atmosphere over a black surface. The WFs emphasize findings discussed before. For example, the sensitivity to water vapor of the HAMP-G channels is by a factor of up to 100 higher than the sensitivity of the HAMP-K. The innermost HAMP-G channels have their WF maximum close to the aircraft, while channels further away from the central

frequency also sense lower layers. Likewise the temperature WFs in the V band have their maximum in higher layers the higher the frequency is. This shows that the BT at 58.0 GHz originates from higher and colder layers than the BT at 53.75 GHz.
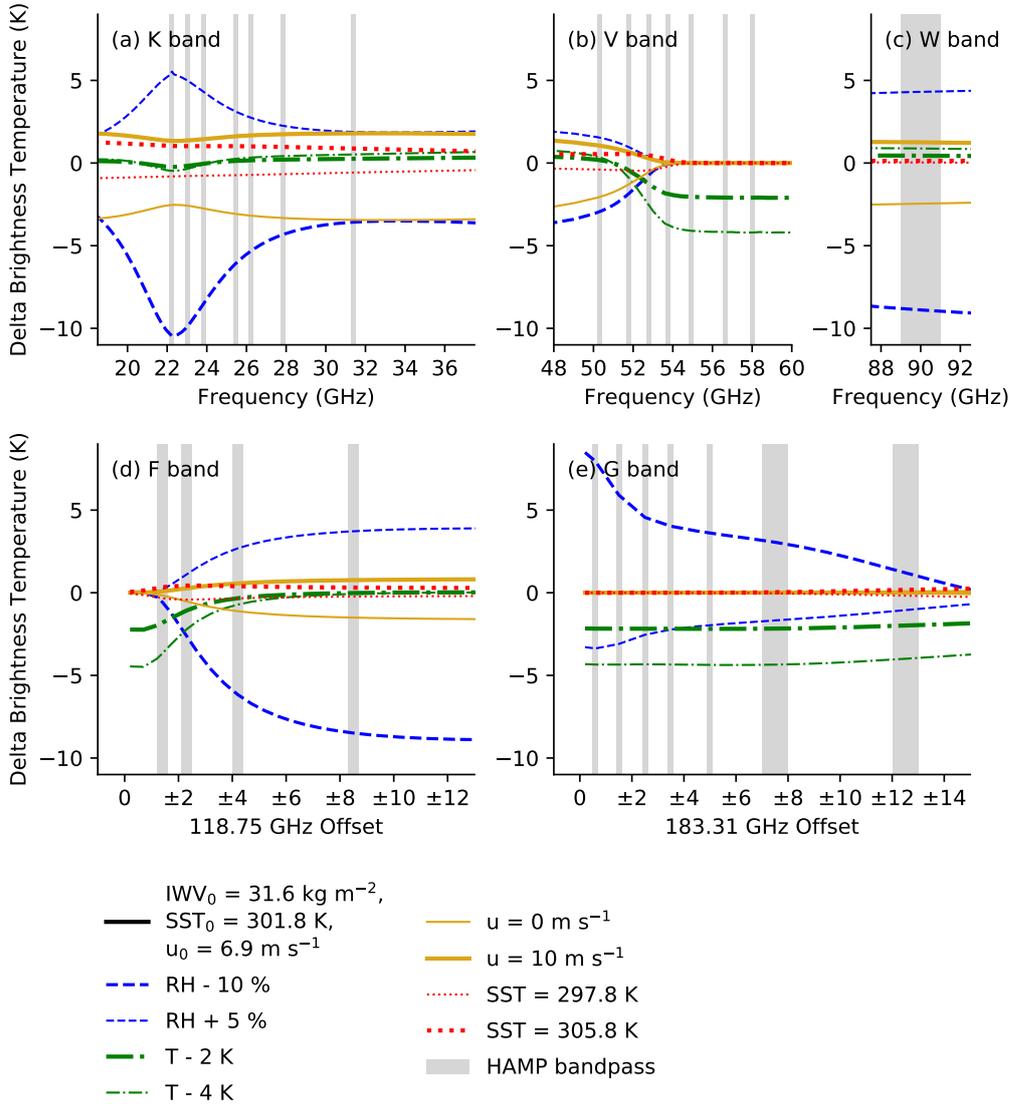
Figure 2.5.: Brightness temperature (BT) changes for slightly modified input to the forward simulation. The absolute BT of reference simulation is shown in Fig. 2.4. For the explanation of the sensitivity experiments see Table 2.2.

.
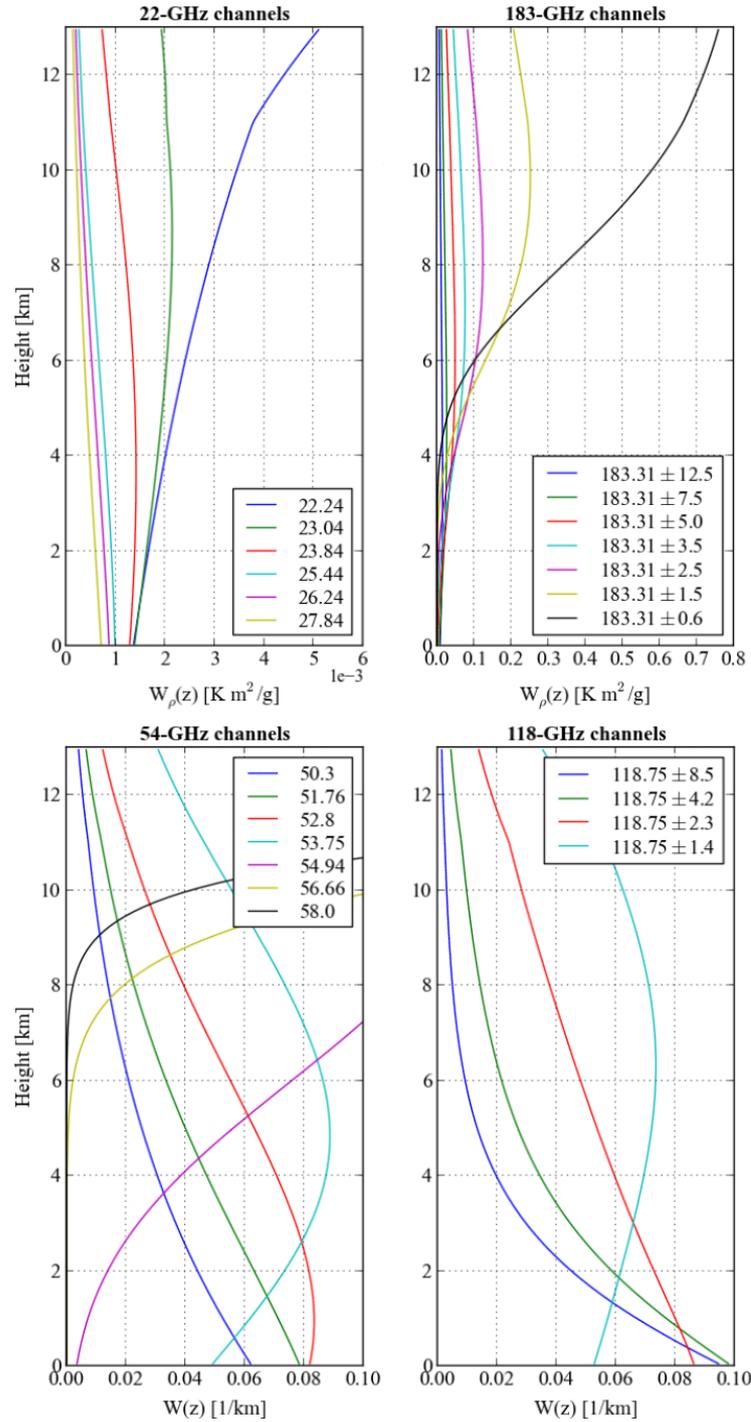
Figure 2.6.: "HAMP clear air weighting functions (nadir downward looking). Shown are the water vapor (top row) and the temperature (bottom row) weighting functions for a HALO ceiling height of about 13 km. The US 1976 Standard Atmosphere over a black surface was assumed for the calculations." The 22, 183, 54 and 118 GHz channels are from HAMP-K, -G, -V, and -F, respectively. Figure and quoted caption by Mech et al. (2014) are distributed under Creative Commons Attribution 3.0 License. The sub-figures have been rearranged.

Chapter 3.

Characterization of the HAMP
Microwave Radiometer

3.1. Introduction

The HALO Microwave Package, HAMP, comprises a cloud radar and a multi-band microwave radiometer as described in Sect. 2.5. While the radar has been calibrated by Ewald et al. (2018), the characterization and assessment of the calibration quality of the radiometer during airborne operations was still missing.

The proper characterization and calibration is an unquestionable prerequisite for geophysical retrievals from any instrument. The quality of MWR acquisitions depends on the instrument's stability, noise level and calibration accuracy (Fig. 3.1). The MWR calibration procedure estimates the parameters that are needed to transform measured receiver voltages and counts into brightness temperatures (BTs) as introduction in Sect. 2.4. Whether these parameters are reliable over time depends on the instrument's thermal stability. The continuous internal gain calibration corrects for gain drifts. It has to be noted, that the HAMP radiometers are thermally stressed during airborne operations, as the belly pod is unpressurized and has air ventilation slots. This means, that the rapid change of ambient conditions during climb and descent could impact the instruments and influence the measurements. However, instrument housings with thick walls, insulation, and build-in Peltier elements are used to counteract ambient influence such that receiver temperature is stabilized at a fixed temperature between 36 and 43 °C with a thermal stability better than 0.1 K according to the manufacturer (Rose, 2009).

MWRs are typically calibrated with a so-called hot-cold absolute calibration using two reference points that ideally span the full atmospheric measurement range of BTs (Küchler et al., 2016). These references are created by targets which behave like a black body at the respective radiometer frequency. This means, a target has to be a perfect absorber and emits radiation with a BT equal to its physical temperature following the Planck equation (Eq. 2.7) in the direction of the radiometer antenna. Often, the "hot" target is at ambient temperature, while the "cold" is cooled down with liquid nitrogen (LN$_2$), but also other references are used. For example, the airborne International SubMillimeter Airborne Radiometer (ISMAR) is permanently
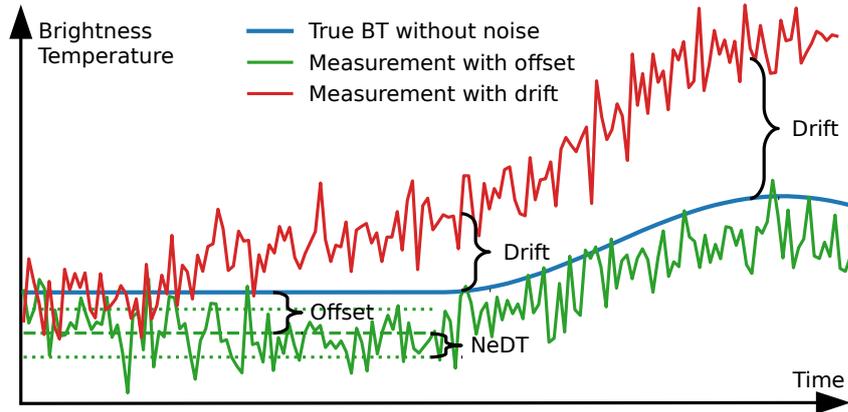
Figure 3.1.: Illustration of error types involved in brightness temperature (BT) measurements. The noise level of the signal is characterized by its standard deviation, also called noise-equivalent delta temperature (NeDT). The offset of the measurement is controlled by the radiometer calibration, while receiver instability results in a drift.

monitoring a cold target at aircraft ambient temperature and a hot target that is electrically heated (Fox et al., 2017). Similarly, spaceborne MWRs use the cosmic background radiation of about 2.7 K as cold reference. For ground-based MWRs, the cold reference can also be realized with the so-called tipping curve calibration involving measurements in clear sky with different elevation angles, including zenith pointing and almost horizontal measurements (Han and Westwater, 2000; Küchler et al., 2016).

3.1.1. HAMP Radiometer Calibration Method

External hot and cold targets are used for the absolute calibration of HAMP radiometers before flight as no cold targets are embedded and a tipping curve calibration is impossible with the radiometers mounted on the aircraft. The continuous gain calibration using Dicke switches and noise diodes does not provide enough reference points for an absolute calibration. The five radiometer bands are pointed subsequently at the hot and cold targets following the manufacturer's method for the absolute calibration. The targets are enclosed such that their containers can be filled with LN_2 and cooled down to the LN_2 boiling point, which is at 77.35 K under standard pressure. The boiling point increases with increasing atmospheric pressure following the Clausius–Clapeyron relation; but also impurities due to mixed-in oxygen can increase the temperature by a few Kelvin (Dodge and Dunbar, 1927; Lemmon et al., 2000). Targets with a horizontally oriented absorber and an open

air-LN$_2$ interface were used during NARVAL-South and -North as shown in Fig. 3.2. However, the handling of these rather large open targets filled with LN$_2$ was not ideal on a windy airport apron. Further, Küchler et al. (2016) identified a negative influence on the calibration by an additional signal due to standing waves reflecting on the air-LN$_2$ interface. The standing waves add an oscillating offset to the observed cold target temperature. This offset changes as a function of the LN$_2$ filling level, which changes through evaporation. To suppress the reflection, targets embedded in novel closed containers developed by RPG were used in later campaigns starting with NARVAL2. The containers are mainly made from foam that has a very low microwave emissivity and does not reflect such that it is considered transparent. The sides of the containers act as windows. The windows are milled at the Brewster's angle on the inside, such that no microwave emission of the antenna itself is reflected at the foam-LN$_2$ interface. As the absorber is oriented vertically in the new targets, a metal mirror at a 45° angle is used to direct the MWR field of view on the targets. The new calibration setup is depicted in Fig. 3.3.
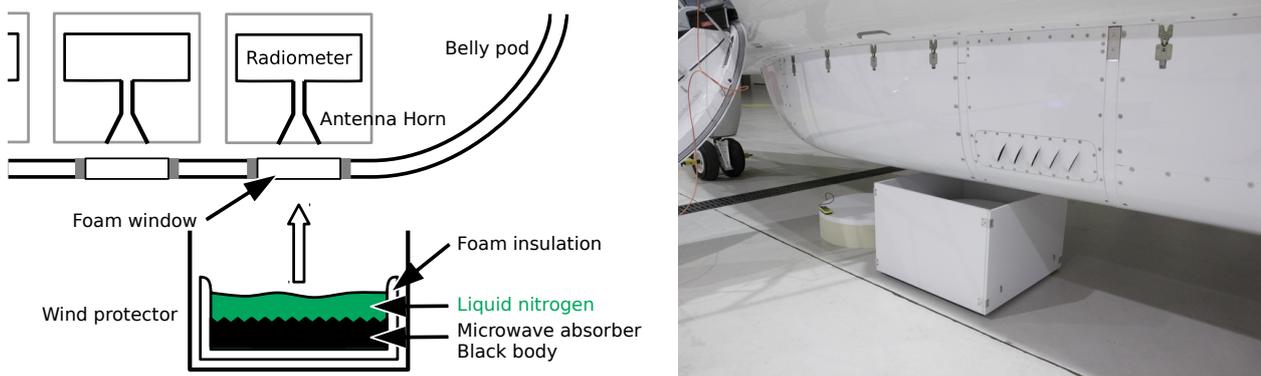
Figure 3.2.: Schematic and picture of calibration targets used during NARVAL-South and -North under the HAMP belly pod. In the picture, left target contains hot load, right target cold load. The picture is a courtesy of Heike Konow, 2014.

3.1.2. Goals and Structure

A preliminary study by Barrera Verdejo (2016) and the work by Schnitt et al. (2017) showed discrepancies between synthetic BT measurements form forward-simulated dropsonde profiles and the actual radiometer BT measured during the dropsonde releases during NARVAL-South. As new data was acquired during NARVAL2, a reassessment and deeper analysis of BT offsets are important steps before NARVAL2 and NAWDEX radiometer data can be used for further meteorological studies. Besides that, the instrument noise of the radiometers in practice has not been analyzed
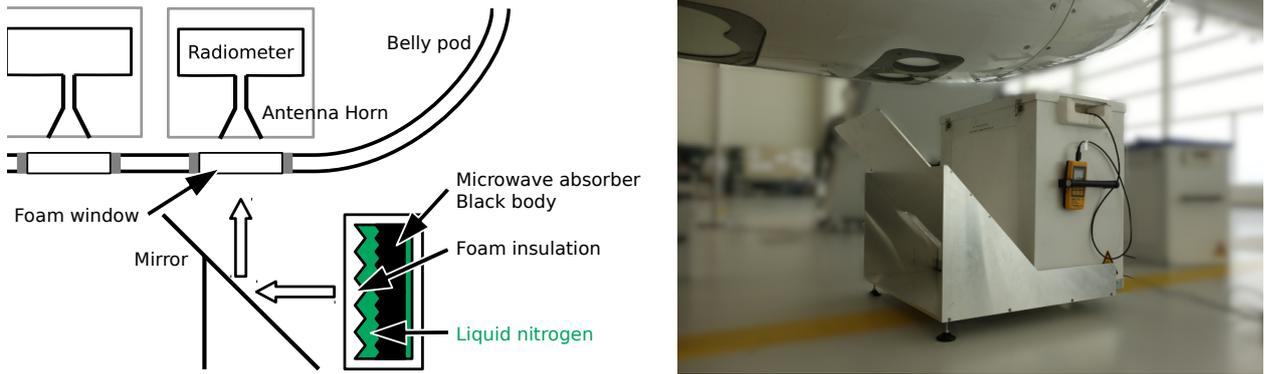
Figure 3.3.: Schematic and picture of calibration targets used during NARVAL2, NAWDEX, and EUREC[4]A under the HAMP belly pod. The absorber target in the white container emits through the foam window of the container. A metal mirror directs the emission into the radiometer antenna, which is located behind the white foam window in the gray frame in the belly pod.

yet. Therefore, the noise in terms of BT fluctuations (Sect. 3.2) as well as the absolute calibration and its stability (Sect. 3.3) are investigated in this preparatory study. A discussion and summary of both aspects will be presented in Sect. 3.4.

3.2. Noise Characterization

The noise-equivalent delta temperature (NeDT) is an important characterization of microwave radiometers. The NeDT is the standard deviation of a radiometric measurement due to the thermal noise of the observed scene with its temperature T_{sc} and the components of the radiometer like the antenna, amplifiers, and detector characterized by the receiver noise temperature T_{rec}. The ideal NeDT is defined as

$$\text{NeDT}_{ideal} = C \frac{T_{sc} + T_{rec}}{\sqrt{B\,\tau}} \tag{3.1}$$

with the width of the channel bandpass B, and the integration time for each acquisition τ (Ulaby et al., 1981). The ideal NeDT is modified for different kind of radiometers by the term C, which is about 2 to 3 in the case of HAMP (see Ulaby et al., 1981, Tab. 6.4). The bandpasses are different for different HAMP channels (Table 2.1). The NeDT is the lower limit of random errors that must be considered in the retrieval development to avoid overfitting and enable a reasonable retrieval error estimation when using synthetic measurements (Chapter 4).

The calibration setup provides two stable T_{sc} conditions which can be used to characterize the true NeDT. The NeDT can be estimated as the standard deviation over a several minutes long acquisition on the target under the assumption that the temperature of the scene, i.e., calibration target, and receiver are constant. At the same time, such a setup can be used to investigate covariances or correlations between the channels within each frequency band. The covariances are important for advanced retrieval methods like optimal estimation (Rodgers, 2000). Most of the amplifying and filtering waveguides circuits are independent components for each channel. Therefore, low correlation coefficients are expected, because the noise coming from most of the redundant circuits should be uncorrelated (Skou and Vine, 2006). A quantitative correlation expectation cannot be given beforehand, as the noise dependence of the specific components is unknown and their individual characterization would require a breakdown of the instrument which is beyond the scope of this study. However, high or irregular correlation patterns can either point to external variation of the reference target during the acquisition or to unreliable channels due to hardware malfunction. In addition to ground-based characterization using the calibration targets, also airborne measurements over very homogeneous scenes will be used in the following noise analysis.

3.2.1. Ground-based and Airborne Experiment Setup

To investigate the HAMP radiometer noise and channel correlations, the three modules were set up for a laboratory experiment in the HALO aircraft hanger in Oberpfaffenhofen, Germany. The modules were turned 90°, such that antennas were pointing sideways and targets could be placed into the beam without any external mirrors as shown in Fig. 3.4. Each antenna was subsequently pointed toward the cold and hot targets, and BTs were acquired using the standard measurement mode for at least five minutes. The data for the characterization of the dual-band modules HAMP-KV and -WF on the hot target were recorded at once, while the cold characterization was acquired for each band separately due to the polarization of the radiometers and targets. The corrugation on the inside of both sides of the target container is rotated by 90° to minimize microwave reflections at different polarization angles when filled with LN_2.

The NeDTs can also be estimated from atmospheric measurements in addition to the laboratory experiment. This requires as little change as possible of the atmospheric and surface emissivity and absorption during a scene as the NeDT estimation requires a stable signal source. It is expected to find such conditions in a cloud free scene. The NARVAL2 flights from August 15, 19, 21, and 27, 2016 have the lowest
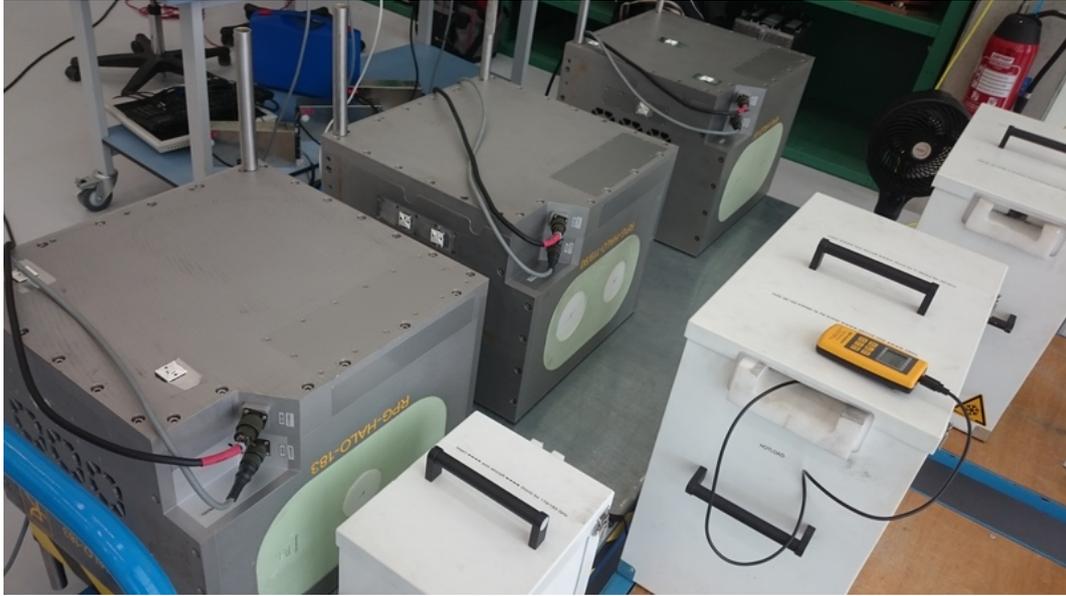
Figure 3.4.: Setup of HAMP radiometer modules for the ground-based noise characterization experiment. White containers in the front contain the calibration targets. Behind are the 3 HAMP modules from left to right: HAMP-G, HAMP-WF, and HAMP-KV. The fan in the back on the right hand side provides an air stream to avoid condensation on the LN$_2$-filled right-most big target.

overall cloud cover regarding the cloud radar data. Therefore, radiometer data from those flights is used to determine the in-flight NeDT as the smallest lowest standard deviation of BT measurements in each channel within any five-minute interval.

3.2.2. Results

The ground-based experiment for the NARVAL HAMP configuration took place in March 2017 after all instruments that were used during NAWDEX were dismantled from HALO (see also Fig. 1.2). The integration time was set to 1 s during the considered flight campaigns. The experiment was repeated in May 2019 after the hardware upgrade for EUREC[4]A (last paragraph in Sect. 2.5). As this upgrade included a software update for a higher temporal resolution, data was recorded with 4 Hz during EUREC[4]A. The results for the individual radiometer bands are shown in Fig. 3.5 for NARVAL/NAWDEX and Fig. 3.6 for EUREC[4]A. The standard deviations shown on the main diagonal of each of the presented matrices are the NeDT for each channel.
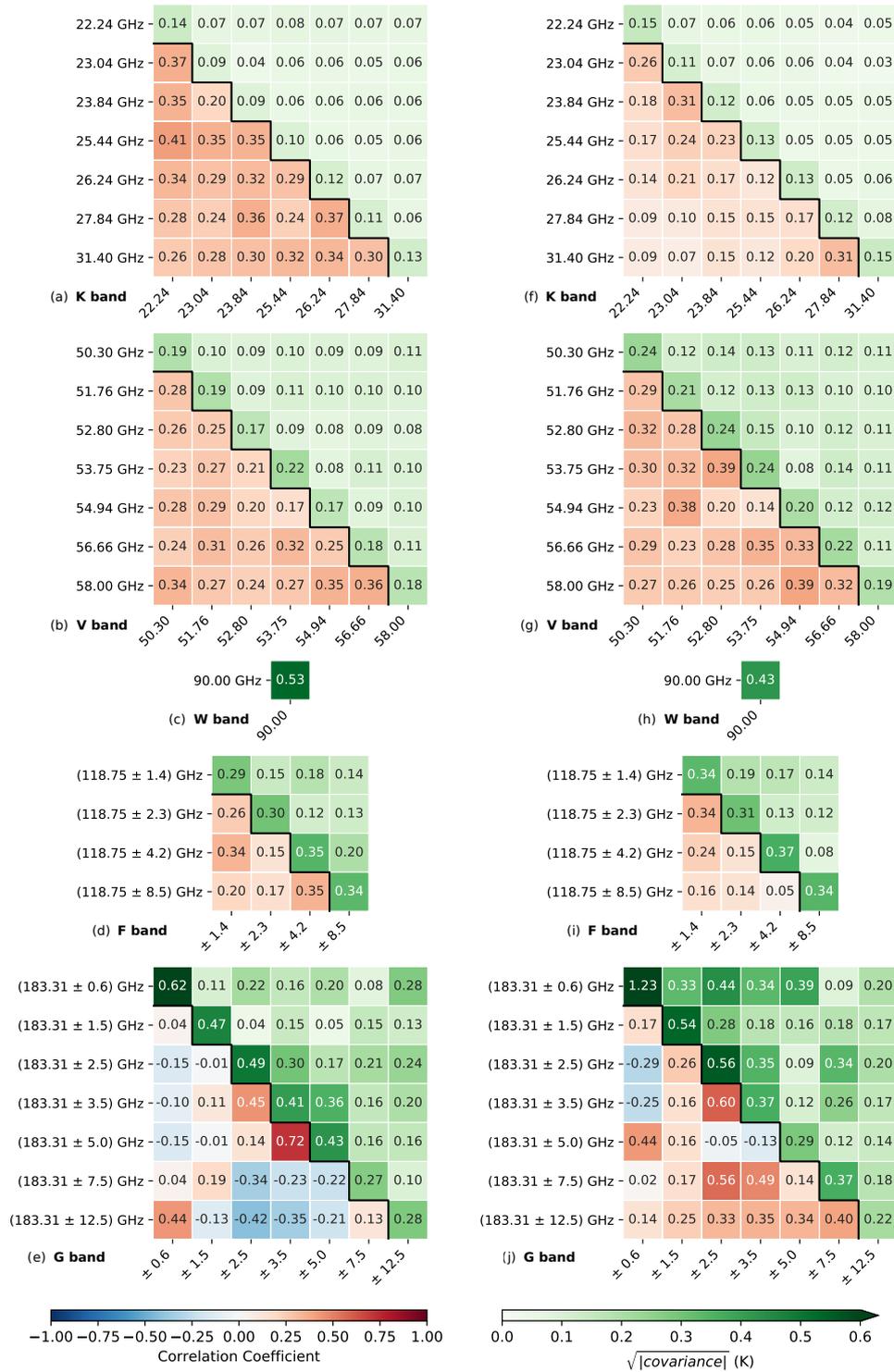
Figure 3.5.: Matrices of correlation and covariance of all HAMP bands. Entries on the diagonal show the standard deviation (i.e., NeDT) in Kelvin K. Entries above (green colors) are the square roots of the covariance in K and entries below (blue to red colors) the correlation coefficients. Left column (a-e) shows characteristics on cold load (77 K), right column (f-j) on hot load (293 K). Data shown from the experiment with HAMP in the NARVAL configuration recorded with 1 Hz sampling rate.

In the K band the NeDT is smaller than 0.16 K for each channel in the NARVAL configuration. The NeDT is slightly smaller when estimated using the cold target. This follows the relation of the ideal HAMP-K NeDT (Eq. 3.1), which is 0.08 and 0.11 K for a cold (77 K) and warm scene (300 K), respectively, when assuming an ideal Dicke radiometer with $C = 2$ and $T_{\text{rec}} = 520$ K. The difference between the ideal NeDT and the observed NeDT is due to non-ideal hardware components. The correlation coefficients and covariances between different channels are rather homogeneous with correlation coefficients on the order of 0.1 to 0.4. Higher correlations were registered on the cold target. Likewise, the V-band NeDT is below 0.25 K for all channels and also slightly smaller on the cold target. The correlation between V-band channels is also rather homogeneous with correlation coefficients between 0.2 and 0.4 for the hot and cold target experiment, respectively. In contrast to the K- and V-band channels, the 90 GHz channel NeDT is higher on the cold target (0.53 K) than on the hot target (0.43 K). The F-band NeDT is below 0.38 K for all channels, and is again slightly lower when estimated at the cold target. Correlation coefficients range between 0.05 and 0.35 without a clear hot and cold target difference nor spectral order. The G-band channels give a rather mixed impression in contrast to the bands discussed before. All channels showed standard deviations below or equal to 0.6 K as specified by the manufacturer (Rose, 2009) for the cold load measurement. But all but one channel also showed standard deviations smaller than 0.6 K on the hot target. However, some standard deviations were again lower than on the cold target and the 183.31 ± 0.6 GHz channel showed a rather high standard deviation of 1.23 K on the hot target. Judging from the respective 20 min-time series of the hot target data acquisition, most of the variance of this channel seems to be coming from variations and drifts on rather a one-minute time scale than from second-to-second noise. In principle, the NeDT of the 183.31 ± 7.5 GHz and ± 12.5 GHz channels should be smaller than the NeDT of the other G-band channels. However, this was not observed. Further, the covariance of the G-band channels is rather heterogeneous. While some channels have rather high correlation coefficients up to 0.72 others are negatively correlated with coefficients up to -0.42. As some channel pairs showed positively correlated measurements on the one and negatively correlated measurements on the other target, the behavior of the G-band channels is different to the other bands, where all measurements were slightly positively correlated.

With the EUREC[4]A upgrade the purity of the K- and V-band channels improved a lot (compare Figs. 3.5 and 3.6). The correlations between channel pairs basically vanished and individual correlation coefficients are now between -0.15 and $+0.07$ except for the two highest-frequency channels of the V band on the hot target. Their
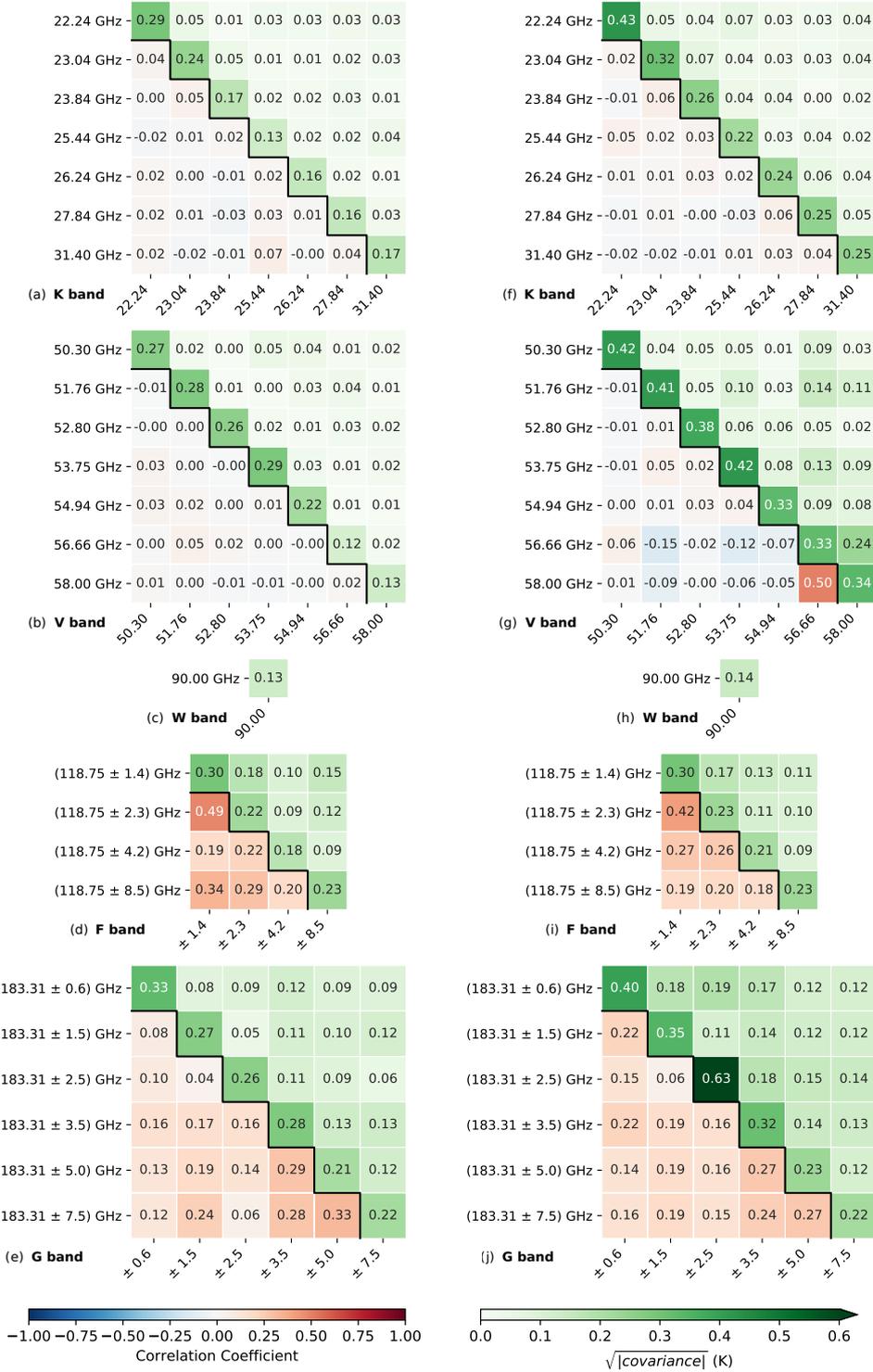
Figure 3.6.: Like Fig. 3.5 but with the EUREC⁴A upgraded radiometers recorded with 4 Hz sampling rate. Left column (a-e) shows characteristics on cold load (77 K), right column (f-j) on hot load (296 K).

correlation coefficient on the hot target was estimated 0.50, while it was 0.02 on the cold target. This prominent correlation cannot be explained by slow changes of the hot target temperature, as this should also increase the correlation with other channels. But also shorter subsets of the hot-target time series show correlation coefficients between 0.3 and 0.5. However, also systematic internal issues in the radiometer are unlikely as the cold-target measurements showed no correlation. The HAMP-K NeDT estimated on the cold and hot target is 0.13 to 0.29 and 0.22 to 0.34 K after the EUREC^{4}A upgrade, respectively. This is a strong improvement considering that the EUREC^{4}A NeDT is estimated at a four times higher resolution, which means, that these numbers have to be divided by 16 before comparing the numbers with the NARVAL setup. The upgraded V-band NeDT is smaller than 0.3 and about 0.4 K on the cold and hot targets, respectively. For the K and V band, the hot-target NeDTs are clearly higher which shows an increased influence of the scene temperature on NeDT relative to the receiver noise temperature, as the hot target temperature was about the same in the NARVAL and EUREC^{4}A characterization experiments. The NeDT of the 90 GHz channel decreased to about 0.14 K after the upgrade to the 4 times higher sampling rate. The F-band noise is also reduced such that the NeDT is now below 0.3 K for all channels at 4 Hz. However, the linear correlation of F-band channels stayed the same or increased even slightly with coefficients between 0.2 and 0.5. The wide variety of the G-band covariances and correlations is more uniform after the EUREC^{4}A upgrade. All correlations are now at least slightly positive with correlation coefficients up to 0.33. The NeDT of all but one channel is 0.4 K or smaller. The 183.31 $\pm$ 2.5 GHz shows an exceptional high NeDT of 0.63 K on the hot target while the cold-target NeDT of 0.26 K is within the range of the other channels.

The NeDT can also be estimated from in-flight observations as an alternative to the ground-based experiments. The comparison of in-flight and ground-based NeDT estimates is important to assess the applicability of the ground-based noise characterization to airborne measurements. To estimate the in-flight NeDT, scenes with very little atmospheric variation are needed as the atmosphere replace the reference target of the ground-based experiments. Thus, the in-flight NeDT is investigated by analyzing the scenes with the lowest variability in each channel. Five-minute intervals of four NARVAL2 flights are considered. The estimated in-flight NeDTs match the NeDTs determined in the NARVAL2 ground-based experiment quite well as shown in Fig. 3.7. The HAMP-KV and HAMP-WF NeDT are clearly below 0.3 and 0.5 K, respectively. Only the in-flight NeDT of the HAMP-G deviates significantly from the ground-based estimates. This might be related to the irregularities of the HAMP-G during the ground-test that were indicated by the varying and in-

homogeneous channel correlations. The ground-based estimates after the EUREC[4]A upgrade correspond better to the in-flight HAMP-G noise, considering that a certain increase of the NeDT for the EUREC[4]A is expected due to the higher sampling rate. Overall, this comparison demonstrates the applicability of the ground-based characterization to the airborne measurements.
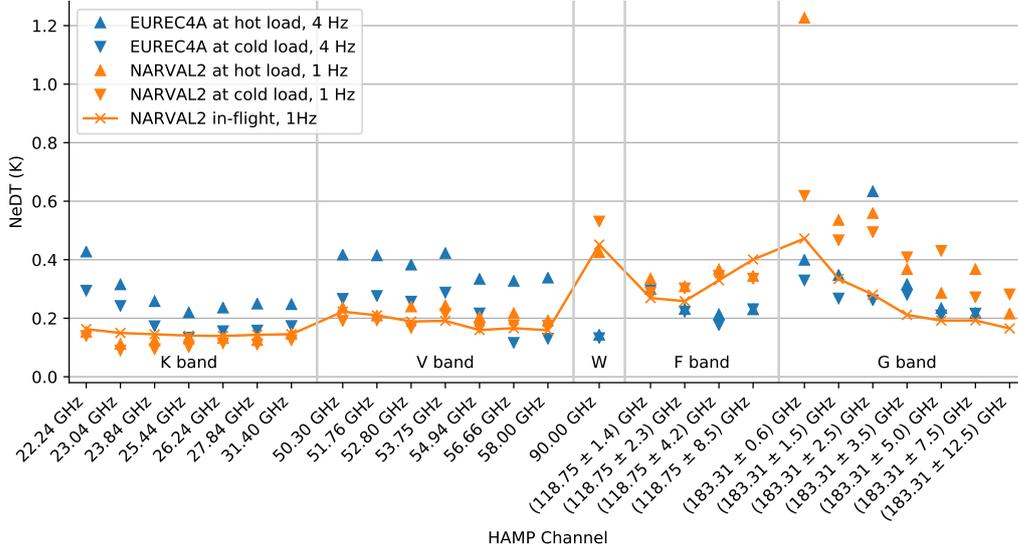
Figure 3.7.: Noise-equivalent delta temperature (NeDT) of HAMP. Derived from ground-based experiments using hot and cold targets (main diagonals from Figs. 3.5 and 3.6), and from the scenes during NARVAL2 flights with lowest variability.

3.3. Brightness Temperature Offset

Besides the sensor's noise (sensitivity) also absolute calibration (accuracy) is important. The HAMP radiometers were absolute calibrated with the manufacturer's calibration procedure before almost every flight. The temperature of the hot target was read manually using a handheld thermometer and the readings were inserted into the software. The boiling point temperature of LN_2 T_{LN_2} was assumed as cold target temperature and derived from an internal atmospheric pressure measurement following the Clausius–Clapeyron relation (Radiometer Physics GmbH, 2015):

$$\frac{\Delta H}{R}\left(\frac{1}{77.35\,\mathrm{K}} - \frac{1}{T_{LN_2}}\right) = \ln\left(\frac{p}{1013.25\,\mathrm{hPa}}\right) \tag{3.2}$$

R is the universal gas constant, ΔH is the enthalpy of LN_2 vaporization, and p is the atmospheric pressure. Equation 3.2 was the best available T_{LN_2} estimator

during NARVAL-South, -North, NARVAL2, and NAWDEX. However, a cryogenic thermometer was available during EUREC^{4}A to measure the physical temperature of the LN$_2$ and the cold target. The measured temperatures of the cold target and LN$_2$ available in Barbados were about 1.2 to 3.5 K warmer than calculated from pressure. These discrepancies indicate impure LN$_2$ due to low-grade production quality and oxygen mixing. It is likely, that LN$_2$ of the similar quality has been used for calibrations in Barbados before EUREC^{4}A as well, which could partly explain the NARVAL-South BT offsets indicated in the preliminary study by Barrera Verdejo (2016) and the work by Schnitt et al. (2017).

Please note, that BT offsets were observed during the noise characterization experiment (Sect. 3.2). Even though, the radiometers have been calibrated before that experiment using the manufacturer's method, deviations of the brightness temperatures from the expected cold and hot load temperatures of up to ± 5 K were observed (Fig. 3.8). As that experiment was conducted to investigate the noise and covariances of HAMP, absolute offsets were tolerated. However, these observations of BT offsets after proper hot-cold calibration motivate a closer examination of BT offsets during the airborne observations.

The reported differences between simulated BTs from clear sky dropsondes and simultaneous BT measurements (Schnitt et al., 2017) are further investigated with a unified method considering also the NARVAL2 observations. All NARVAL-South and NARVAL2 dropsondes are forward simulated and compared to HAMP measurements in clear sky. Comparisons of the midlatitude campaigns NARVAL-North and NAWDEX are provided in Appendix A (pp. 123) for completeness, but are excluded here as this thesis focuses on the tropical clouds. The latest EUREC^{4}A acquisitions are not considered here as their data processing is just starting and the detailed analysis of them is beyond the scope of this thesis.

The increased number of dropsonde released during NARVAL2 flights (9 to 50 sondes with an average of 18 per flight) compared to NARVAL-South (6 to 14, average 9.5) allows for a more robust comparison and offers the opportunity to also analyze the stability over time during one flight. The following research questions are investigated: What are the differences between synthetic and measured BTs in each channel? Does the offset drift over the course of a flight? Is there an offset difference between NARVAL-South and NARVAL2 which can be related to the new calibration targets? Can sources of the offset be identified?
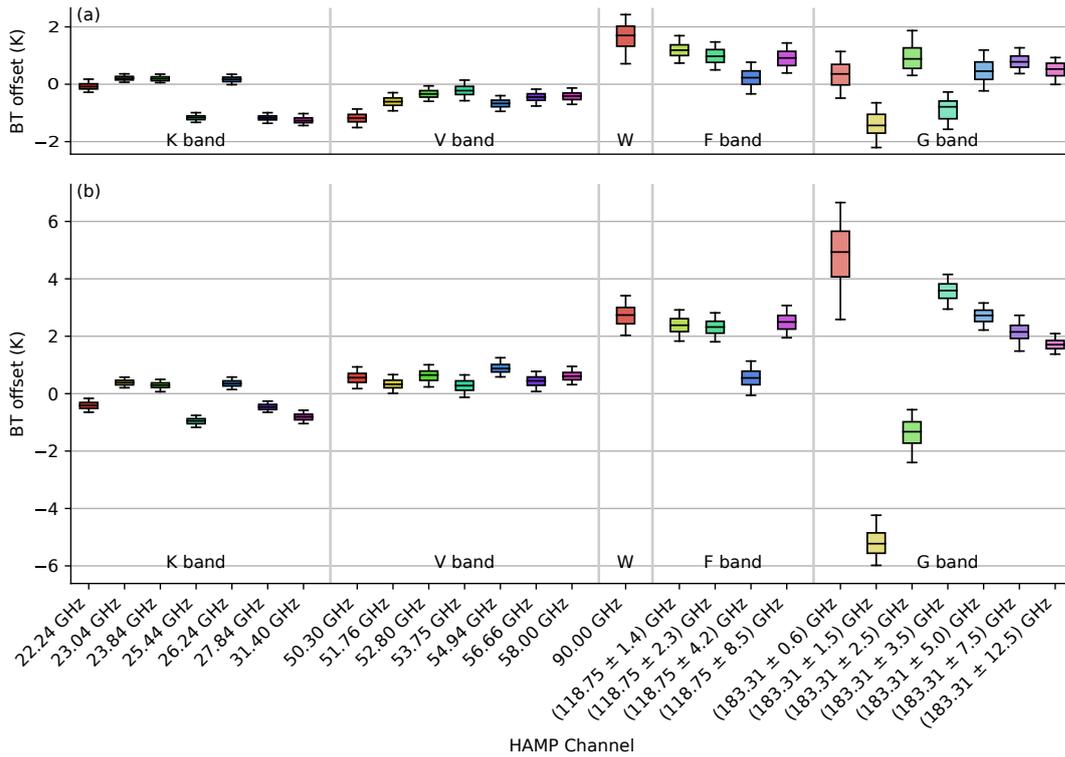
Figure 3.8.: Brightness temperature (BT) offset on cold (a) and hot (b) load after the calibration during the noise experiment using the HAMP in the NARVAL configuration in March 2017 (Sect. 3.2. Box-and-whiskers plots show the quantiles (box), and 5th and 95th percentiles (whiskers) of the data recorded with a 1 Hz sampling.

3.3.1. Methods

Radiative transfer models simulate microwave radiometer measurements based on thermodynamic profiles and further background information as presented in Sect. 2.3. Here, PAMTRA is used as a brightness temperature simulator, which simulates the nadir-pointing HAMP radiometer measurements. Most of the required input can be acquired by dropsondes as long as no hydrometeors emit nor scatter any radiation along the measurement path. Temperature, humidity, and pressure profiles, as well as surface parameters are the essential input parameters to a 1D clear-sky microwave radiative transfer simulation. The sea surface emissivity depends on the near surface wind speed and sea surface temperature. The wind speed is taken from the lowest available measurement of the dropsonde, which is typically 10 to 50 m above the sea surface. The sea surface temperature is taken from the Advanced Very High Resolution Radiometer (AVHRR) infrared satellite product (Reynolds et al., 2007; Reynolds, 2009). Temperature, humidity, and pressure profiles are taken from every dropsonde respectively. These profiles are extrapolated to the top of the available

profiles, as the sondes provide no data within the first few hundred meters of their descent.

The forward simulation of the sondes does not include any emission of rain or cloud droplets. Therefore, forward-simulated sondes can only be used to assess clear-sky measurements of HAMP. Thus a clear- and cloudy-sky discriminator is needed. Even though the WALES lidar or the spectrometer of the Munich Aerosol and Cloud Scanner (specMACS; Ewald et al., 2016) have proven to be the most sensitive cloud detector (Stevens et al., 2019, Fig. 5 therein), their measurements are not used as a discriminator to avoid dependencies on instrument co-availability. Instead, the variability of the HAMP radiometer channels themselves and the HAMP radar are used as a cloud detector. This radiometer approach is similar to the method by Albern (2014). The idea is, that the BT change due to hydrometeors emissions has a high spatial variability compared to other signals like a change in water vapor. The clear-sky variability filter is basically determined by a standard-deviation threshold and a temporal window relative to the dropsonde release. A scene is considered clear sky, if the following two statistics in the given time window are smaller than chosen thresholds: the maximum of the standard deviation in all channels but the G band, and the number of radar range gates with a signal above the noise level.

The G-band standard deviations are only considered when comparing the G-band acquisitions with dropsondes. This general exclusion of HAMP-G from the clear sky filter has two reasons. First, the usability of the HAMP-G was limited as it suffered from instabilities due to a broken receiver component during some flights which caused a sawtooth pattern in the BTs (Konow et al., 2019). The jumps of the BTs caused artificial variance in the data. Second, the other channels are more sensitive to liquid water droplets, such that they work better as a liquid-cloud detector, but they are less sensitive to scattering of large (precipitating) ice particles. This means, filtering for precipitating ice is only needed in G-band comparisons.

Tests with different parameter sets showed the best agreement of forward-simulated sondes and simultaneous radiometer measurements when filtering with 10 s and 1 K, which is about two to three times the radiometric noise NeDT. "Best" is considered here, when offsets between measured and simulated BTs are smallest and mostly constant within one flight. Sondes released during a turn of HALO cannot be use either, as the beam orientation of HAMP is changing quickly in off-nadir directions. The numbers of considered clear-sky sondes with stable aircraft orientation of each NARVAL-South and NARVAL2 flight is given are Table 3.1. During NARVAL-South 3 to 8 clear-sky sondes are available per flight, while there are 5 to 37 clear-sky

sondes for each NARVAL2 flight except for research flight 5, where a thick cirrus layer covered the whole study area.

Table 3.1.: Number of all released dropsondes and those released in clear sky with available measurements by the HAMP-KV, -FW, and -G module.

Campaign	Research flight	Date	Sondes released	Clear-sky sondes KV	WF	G
NARVAL-South	RF01	2013-12-10	15	4	4	4
	RF02	2013-12-11	6	2	2	0
	RF03	2013-12-12	9	8	8	0
	RF04	2013-12-14	11	6	6	3
	RF05	2013-12-15	9	5	5	3
	RF06	2013-12-16	10	6	6	4
	RF07	2013-12-19	9	4	4	4
	RF08	2013-12-20	8	3	3	0
NARVAL2	RF01	2016-08-08	9	7	7	7
	RF02	2016-08-10	30	19	19	11
	RF03	2016-08-12	50	37	37	18
	RF04	2016-08-15	10	7	7	1
	RF05	2016-08-17	12	0	0	0
	RF06	2016-08-19	50	33	33	19
	RF07	2016-08-22	13	5	5	1
	RF08	2016-08-24	12	8	8	2
	RF09	2016-08-26	12	11	11	3
	RF10	2016-08-30	17	14	14	0

3.3.2. Temporal Stability of the Radiometer Measurements

The NARVAL2 research flight on August 12, 2016, provides a good opportunity to study the radiometer stability over the course of a whole flight, as 50 sondes were released during an 8-hour flight. Most of the sondes were released during two intense phases of 24 sondes each lasting about 95 minutes. The flight track and locations of the sonde releases are depicted in Fig. 3.9. The 24th, 37th, and 49th sonde of the flight were launched while HALO was turning. This means, these sondes cannot be used in the following analysis. One more sonde was launched during the 135 minutes break between the intense phases and another one shortly before starting the final descent. In total, 47 sondes are available with simultaneous near-nadir

measurements of HAMP-K, -V, -W, and -F. The HAMP-G data is only available in the second half of the flight with 24 usable sondes as hardware instabilities imprinted an artificial sawtooth pattern with decreasing amplitude on the data during the first half of the flight. Ten of the 47 sondes are considered cloudy and are not considered in the following discussion (Fig. 3.11 e).
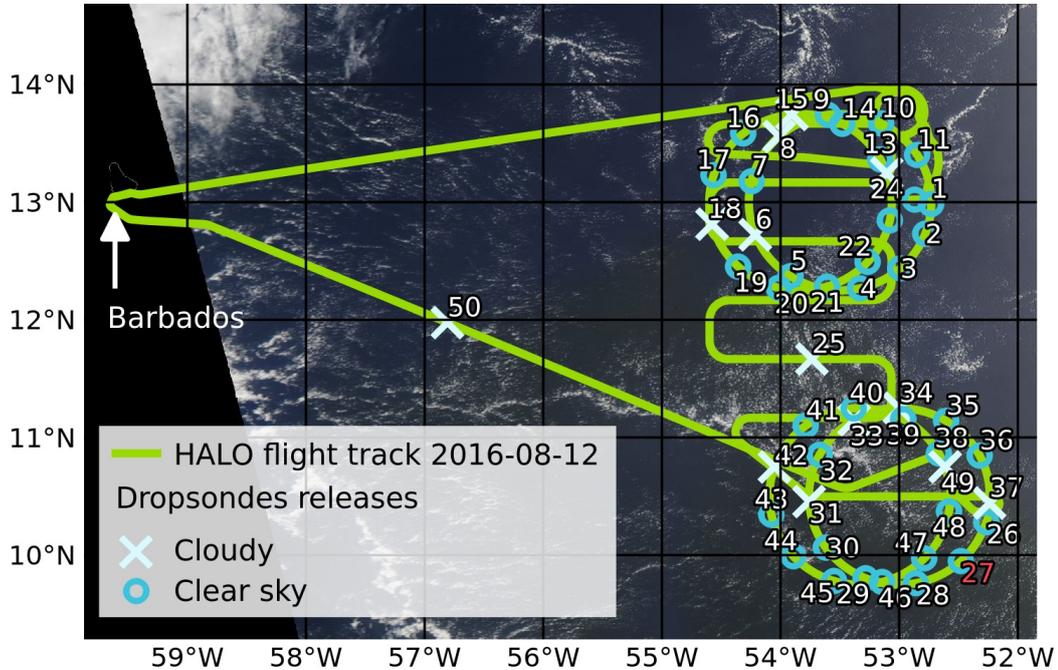
Figure 3.9.: HALO flight track and dropsonde locations from NARVAL2 research flight 3 on top of the MODIS corrected reflectance image from 16:40 UTC. Image retrieved from NASA Worldview.

The forward-simulated sondes should agree with the HAMP acquisition at the time of the release assuming perfectly calibrated and stable radiometers, a perfect forward simulation, and perfect dropsonde sensors with infinite falling velocity. A real sonde, however, needs about 10 minutes to sample one profile. HALO flies about 130 km in this time, while the sonde drifts typically about 4 km in a horizontal direction which likely differs from the flight track. An example of a forward-simulated dropsonde and the corresponding HAMP measurements are shown in Fig. 3.10. The HAMP measurements follow the simulated BTs according to the absorption features of the humid atmosphere. The water vapor absorption around 22 GHz increases the BTs of the lower-frequency channels in the K band. The HAMP-V has the highest BTs in the central channels, as the absorption of the 60 GHz oxygen line is intermediate at these frequency (Fig. 2.1) such that the received temperature mostly comes from warm layers of the lower troposphere but less from the ocean. Lower-frequency V-band channels would see more of the cold ocean, while the high-frequency V-band

channels are so opaque, that most of the signal comes from the atmosphere close to the aircraft. Likewise, high oxygen and water vapor absorption causes low BTs close to the 118.75 and 183.31 GHz lines, respectively. Therefore, the innermost F-band channel can be colder than the second F-band channel. Water vapor absorption around 183.31 GHz is so strong that the HAMP-G BTs even monotonously increase with distance to the line center. These simulated features are also present in the HAMP measurements shown in Fig. 3.10. However, some channels show offsets of up to 8 K. In addition to the offsets, the measured K- and G-band spectra (Fig. 3.10 a and e) are spiky compared to the simulated spectra. In both spectra, the middle channels seem a bit off.
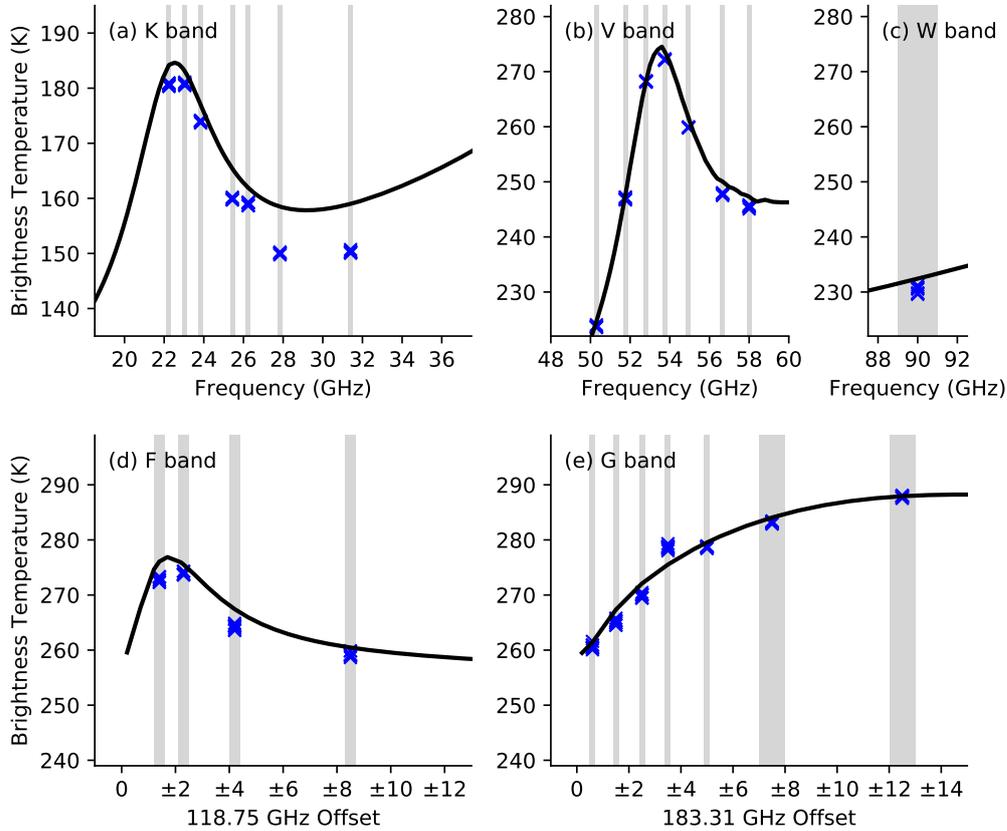
Figure 3.10.: Synthetic forward-simulated dropsonde and corresponding HAMP brightness temperature (BT) measurements measured during 10 seconds after the release of the sonde. All 10 HAMP measurements per channel are depicted as overprinted × markers. The sonde 27 was released at 18:59:12 UTC on August 12, 2016. The F- and G-band simulations are averaged around the central frequency to simulate the double-sideband receivers.

.

All remaining 46 dropsondes are forward simulated like the test case of sonde 27 in Fig. 3.10 and their differences to the observed BTs are analyzed in the following for more statistical evidence. Similar to the sonde discussed before, also the forward-simulated BTs from the other sondes deviate from the HAMP measurements and many channels are persistently offset as shown in Fig. 3.11. HAMP-K measurements in the off-line channels at 27.84 and 31.40 GHz are up to 11.1 K lower than the corresponding forward-simulated sondes, while channels on the water vapor line at 23.04 GHz are 3.7 K higher. Cloudy-sky sondes like the 42nd show even more positive deviations from the dropsonde, but these deviations are due to the additional emission of cloud droplets, which are not considered in the forward simulation. Therefore sondes that are identified as cloudy are excluded in the following. The offset of each HAMP-K channel variates with a mean standard deviation of 1.3 K. Parallel shifts of all channel offsets with time are the prominent pattern in the HAMP-K offsets and linear relations can explain most (minimum, median, maximum squared correlation coefficients r^2: 0.53, 0.88, 0.99) of the offset variation among the channels. However, no general offset drift over time is evident in any channel. Every channel's mean offset ranges between -1.9 and -8.6 K.

The offsets of other HAMP receivers are similar to HAMP-K, however, mean absolute offsets are typically smaller. The more transparent channels of the other HAMP receivers like 50.30, 51.76, 90.00, and 118.75 $\pm$ 4.2 and $\pm$ 8.5 GHz show offset variations with similar curves like the HAMP-K. The offsets of the 90 GHz channel can linearly explain 59 % to 88 % (median r^2: 80 %) of the K-band offset variation in any channel. The less transparent channels and those with higher sensitivity to the atmosphere closer to the aircraft show less variation between sondes. The HAMP-G channels are rather constant over time and show little variation among sondes. Their variations are rather uncorrelated with HAMP-K. This indicates that most of the offset variations in the transparent channels are not due to internal instabilities within one HAMP receiver or module. These variations rather depend on the lower-tropospheric input to the forward simulation and are probably mostly related to the humidity profiles. It has to be considered here, that the temporal matching between sonde and the remote sensing is better directly after the sonde release than when it is reaching the ground after approximately 10 minutes. Therefore we conclude, that discrepancies in the remote-sensed and in-situ-measured profiles are most likely related to the spatiotemporal mismatching of HAMP nadir measurements and the drifting sondes.

No general offset drift between any HAMP channel and the sondes can be noted. Thus, the mean and standard deviation can summarize the offset of the individual sonde comparisons for each channel. The mean and standard deviation are summa-
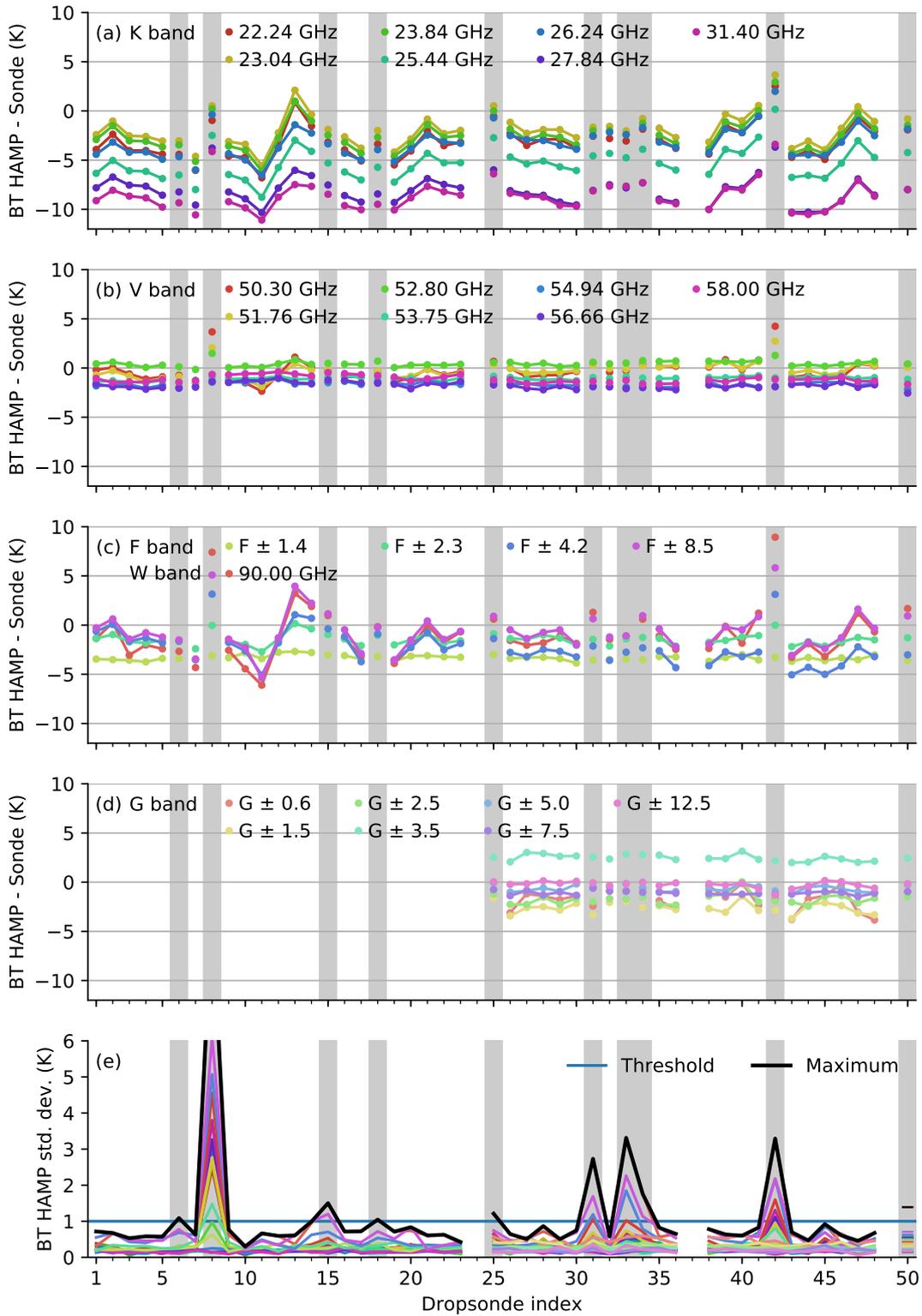
Figure 3.11.: Offset between HAMP brightness temperature (BT) and forward-simulated dropsondes during the NARVAL2 flight on August 12, 2016. HAMP-K in (a), -V in (b), -W, and -F (118.75 GHz) in (c), and -G (183.31 GHz) in (d). Gray shaded sondes are considered cloudy due to high BT standard deviation (e).

rized and discussed for all flights of NARVAL-South and NARVAL2 to set them in the broader picture, in the next section.

3.3.3. HAMP–Dropsonde Offset during NARVAL-South and NARVAL2

The analysis of offsets between forward-simulated clear-sky dropsondes and HAMP for all NARVAL-South and NARVAL2 flights is important in order to identify systematic and variable error patterns. The offset and its sign are defined by

$$\text{offset} = \text{BT}_{\text{HAMP}} - \text{BT}_{\text{sonde}}. \tag{3.3}$$

The mean offset and offset standard deviation for each flight and channel are depicted in Fig. 3.12. The number of considered dropsondes in given in Table 3.1. The offset in each band has a typical spectral shape, that is more or less evident in all flights. For example, the second, third and fifth K-band channels typically have the strongest positive offset, while the first and fourth channel offsets are a bit lower and the sixth and seventh channels have the most negative offset. The spikes in the offset spectra are similar to the spikes in Fig. 3.10. The range between the highest and lowest offset within one flight is about 6 K. The offset is spectrally speaking similar among different flights, but shifted in absolute terms. This means, that the spectral dependence of the offset is systematic, especially in the K band discussed for one flight in the section before.

The HAMP-V offsets are generally more positive in the lower-frequency channels and more negative in the high-frequency channels. The spectral offset amplitude is about 2 to 6 K, but the offset pattern is spectrally less persistent than in the K band. The mean 90 GHz offset ranges between -4 and 4 K in the NARVAL-South dataset and -4 and 0 K in the NARVAL2 dataset. The 118.75 ± 1.4 GHz has in general the most negative offset in the F band, while the second channel offset is higher and the third and fourth channel show the highest offsets, which are often about 2 to 6 K warmer. The spectral offset shape has a similar degree of consistence between flights in the F band as in the V band. HAMP-G BTs were almost always lower than the respective dropsonde forward simulation during NARVAL-South. The inner most channel has the most negative offset. The offset generally increases in absolute terms with distance to the line center. The spectral amplitude of the offset is about 7 to 10 K. The HAMP-G pattern is a bit different for NARVAL2. During NARVAL2, most channels have a mostly negative offset as well, but the 183.31 ± 3.5 GHz channel reported warmer BTs than simulated during most flights. The inner most channel has again the most negative offset and the HAMP-G offset amplitude during NARVAL2 is about 3 to 8 K. The NARVAL2 flight on
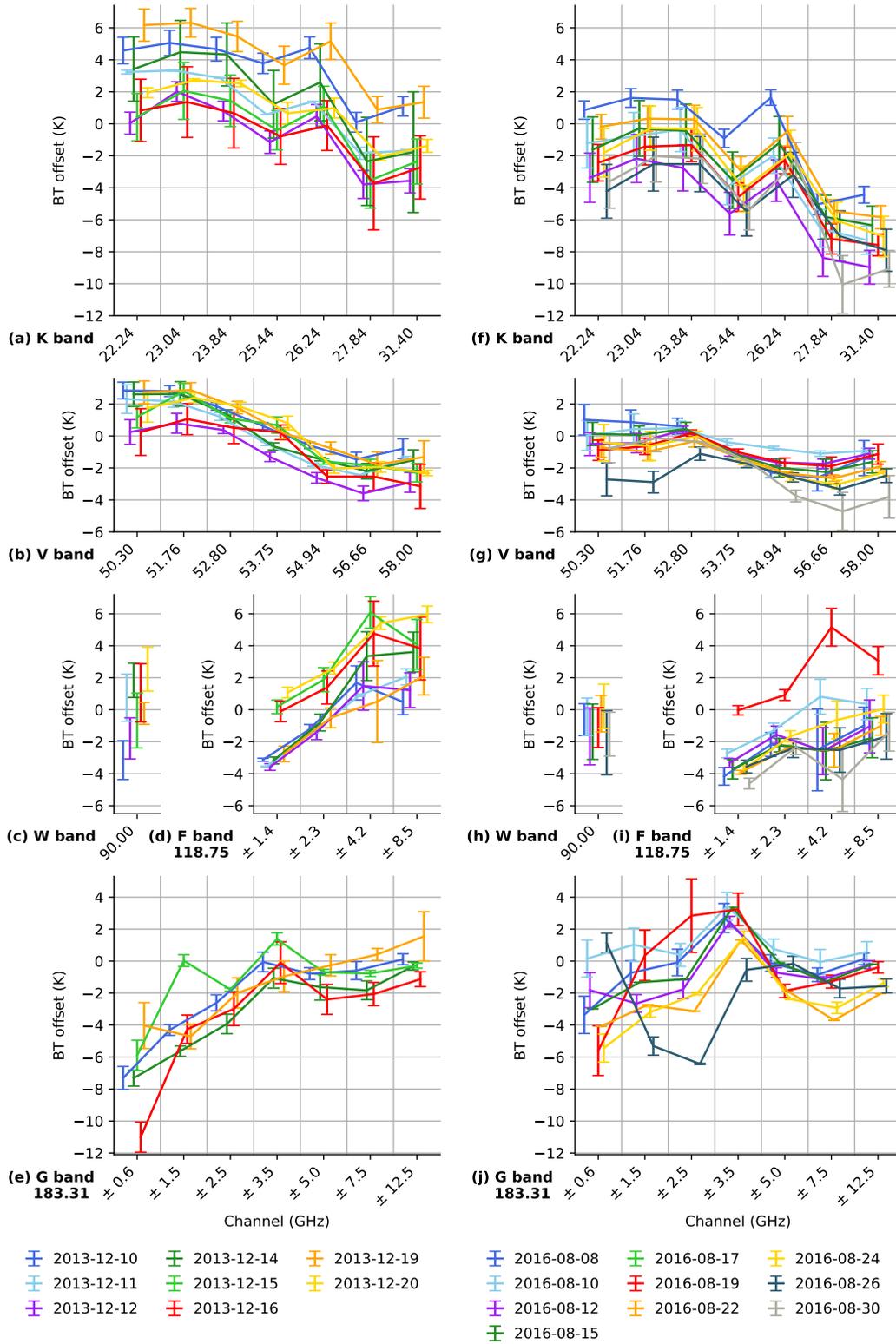
Figure 3.12.: Mean brightness temperature (BT) offset for each flight of NARVAL-South (left column, a-e) and NARVAL2 (right column, f-j). Offsets are HAMP BT minus forward-simulated dropsonde BT in clear-sky scenes. Error bars denote twice the standard deviation over all clear sky dropsondes within each flight. Number of considered sonde for each flight is given in Table 3.1.

the 26 August 2016 is an exception to the values reported before. Three clear-sky dropsondes with simultaneous HAMP-G acquisitions are available and these show the strongest positive offset of about $1\,\mathrm{K}$ in the innermost and the most negative offset of about $-6\,\mathrm{K}$ in the $183.31 \pm 2.5\,\mathrm{GHz}$ channel.

The standard deviation (shading in Fig. 3.12) of the offset in each channel of all sondes per flight is mostly below $2\,\mathrm{K}$. The standard deviations also show some spectral features. For example, the 53.75, 54.94, and 118.75 ± 1.4 and $2.3\,\mathrm{GHz}$ channels always have the smallest offset variance in their bands. However, significance of the variance is limited in many cases due to the small sample size during the flights.

In the comparison of NARVAL-South and NARVAL2 offsets it becomes clear, that the NARVAL-South offsets are generally higher in absolute terms than the respective NARVAL2 offsets in all but the G band. The mean absolute K-, V-, W-, and F-band offset of NARVAL-South is about 5, 1, 1, and $2\,\mathrm{K}$ higher than the respective NARVAL2 offset. This offset shifts could be related to the exchange of the calibration targets, which would provide a more reliable cold calibration reference, but it cannot be determined form the dropsonde comparison alone with absolute certainty. If the old cold target had a warmer BT due to reflection than it was assumed by the calibration routine, colder than real BTs would have been recorded by the software in cold scenes, in principle. Consequently, a cold offset could be expected for NARVAL-South. However, the software also included a correction function to account for the mean reflection, but not the oscillation caused by standing waves. Therefore, one could speculate, that strong wind drag could rough the LN_2 surface such that the reflection at the air-LN_2 interface would be effectively reduced turning the reflection correction in the software into a positive offset. This could at least partially explain, why NARVAL-South offsets were generally more positive than NARVAL2 offsets. An experiment with both target types under real condition could probably spread more light on the issue, but practical reasons have hindered further experiments and there are other factors which influence the radiometer stability similarly or even stronger. Finally from the practical side it has to be noted, that the new closed targets improved the handling of the LN_2 cooling during calibration a lot.

3.3.4. Influence of the Forward Simulation on the Offset

Besides the hardware and calibration method, offsets could also be caused by the forward simulation itself. The forward model is driven the by dropsonde measurements, but dropsonde, like radiosondes, are consumable devices. Therefore, their

sensors are rather simple. Studies have reported on measurement biases and especially on dry biases (e.g., Vömel et al., 2007, 2016). Such biases could cause offsets between HAMP and the forward-simulated dropsondes.

The sensitivity analysis in Sect. 2.6 uses the same dropsonde profile as shown with Fig. 3.10 in Sect. 3.3.2. In the sensitivity analysis in Sect. 2.6 it is shown, that an increase of the whole relative humidity profile by 0.05, would increases the innermost K-band channel by 5.5 K and the 31.4 GHz channel by 1.8 K, i.e., would imprint a spectral difference of 3.7 K on the K band. Increasing humidity would imprint a spectral difference of 1.5, 4.3, and 3.7 K on the HAMP-V, -W, and -F channels, respectively. The channel with the lowest frequency in the V band and the outermost F-band channel would be increased the most. The G-band BTs would decrease with increased humidity with the strongest decrease of -3.4 K near the line center compared to a decrease of $(-1.2$ K$)$ at the wings.

An assumed warm measurement bias of the sonde of 2 K, i.e., the simulation with decreased temperature by 2 K, would decrease the strongest absorption in V- and F-band channels by 2.1 and 1.8 K, respectively, while leaving the least absorbing channels almost unaffected (0.2 and 0.0 K). HAMP-G BTs would decrease by about 2.1 with no strong spectral dependency. HAMP-K and -W BTs would be less affected by a temperature bias. The spectral sensitivity to the surface parameters is comparably small in all channels.

Comparing Fig. 3.12 – the general spectral patterns of the offset – and Fig. 2.5 – possible variations of the forward simulations due to dropsonde sensors biases – is presumptuous on which dropsonde sensors biases could explain some of the spectral offsets. For example, a dry bias – compensated by the relative humidity $+5\,\%$ simulation – could explain the general shape of the spectral offset in the K and F band, partially in the V band and G band, excluding the central G band channels during NARVAL2. However, the spectral amplitude of the humidity increase is smaller than the spectral offset amplitude in all channels. Further, the 0.05 increase in relative humidity would also result in warmer measurements at 90 GHz which is opposite to most of the offsets that are negative in the W band. Also, the simulated BT increase in the outermost F-band channels objects the NARVAL2 offsets observed for those channels. Spectrally, a warm bias of about 4 K could explain the shape of the V-band and F-band offsets, but would also yield an almost constant negative offset in the G band. This contradicts the observed offsets of the outermost G-band channels, which are around 0 K. Apart from this, such a significant temperature bias of several Kelvin or a serious humidity bias larger than 0.05 would somehow be the minimum to partially explain the BT offsets. Such dropsonde biases, are larger

than reported issues and also the comparison of integrated water vapor from the differential absorption lidar WALES with dropsonde indicates no severe humidity error as it will be shown in Sect. 4.3.

Furthermore, the sensitivity study does not show any similarity to the spectral spikes in the offsets observed in the mid-K- and -G-band channels during all flights and in other channels during some flights. In general, this sensitivity study is not able to completely explain the apparent offset of the offset patterns between flights, especially since no clear systematic in the offset order of the flights among the different bands could be identified yet. This indicates a strong offset component that is rather band-dependent and independent of any possible dropsonde sensors bias. In addition the sensitivity of absorption models (e.g., Rosenkranz (1998) v.s. Liebe et al. (1993)) was tested but could not explain the offsets.

3.3.5. Results and Discussion

Brightness temperature offsets and their stability were analyzed using forward-simulated dropsonde measurements in clear-sky scenes. Positive and negative offsets of up to $+6$ and $-11\,\mathrm{K}$ in certain channels and flights were found. Typical spectral distributions of the offset are evident in most HAMP receivers. One flight with intense dropsonde deployment is used to investigate the temporal stability. No drift could be found. Although the offset during one flight seems to be stable, the mean offset spectrum is offset between different flights. This might be related to the regular execution of the absolute calibration procedure before most flights. However, the offsets of certain flights without pre-flight calibration, i.e., using the calibration parameters from the previous flights, are different from those of the previous flight. It has to be kept in mind, that the radiometer were turned off between those flights and cooled down, which might have caused changes. An underestimation of the cold target temperature due to impurities of the LN_2 could explain warm offsets in principle as sketched out in Fig. 3.13. Similarly, Maschwitz et al. (2013) identified a warm calibration offset related to a overestimation of the LN_2 temperature. The magnitude of such a calibration bias should be similar in all bands. However, there is no systematic offset order from different flights that applies to the offsets in all HAMP bands. This points to independent offset shifts in all bands and all flights as well. The sensitivity study shows that severe dropsonde sensor biases could explain some of the spectral offset shapes, but not in all channels at once. Further more, the crosscheck with other instruments make the required large dropsonde sensor biases rather unrealistic.

The spectral spikes in the offset patterns and the underlying observed spiky BT spectra (e.g., Fig. 3.10a) involve the same K-band channels that stood out from the offsets observed during the noise experiment (Fig. 3.8), where those channels had offsets in the same direction. This indicates issues related to the radiometer hardware and/or calibration software. It is speculated (Thomas Rose, RPG, personal communication, 2017), that different switching frequencies of the Dicke or noise diode switch in normal operation and the calibration mode could cause offsets directly after the calibration. The switching frequency could play a role in relation to the time constant of the low-pass filter which is used in the temporal integration of the radiometer. However, the software code is not accessible and further analysis is impeded. Maschwitz et al. (2013) investigated the ground-based calibration techniques of a HATPRO-G2 (RPG's Humidity And Temperature Profiler – Generation 2), which is similar to the HAMP-KV. In their study, they found discrepancies between the hot-cold and tipping-curve calibration methods with differences in certain channels of up to $1.5\,\mathrm{K}$. Furthermore, a slight variation of the amplifier gain between the calibration reference measurement on the cold target and on the hot target could play a role for the HAMP offsets. Nevertheless, the apparent repeatability of those BT offsets inspires the confidence to correct them with an additive offset correction.
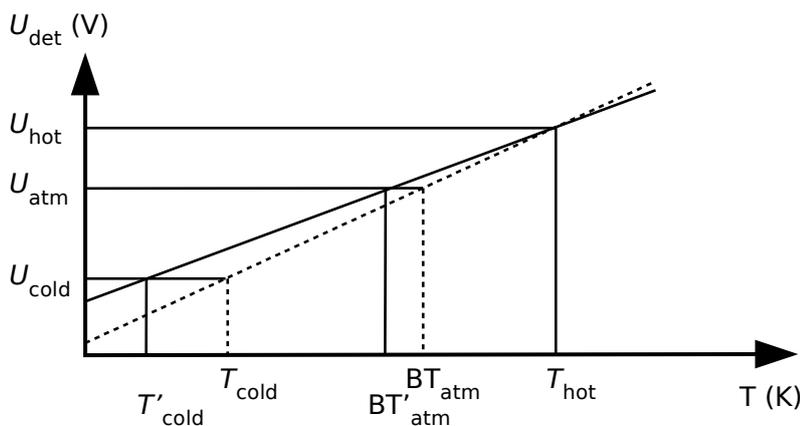
Figure 3.13.: Calibration diagram showing the influence of an underestimation T'_{cold} of the true cold-calibration-target temperature T_{cold}. (E.g., $T_{\mathrm{cold}} > T_{\mathrm{LN_2}}$ calculated from Eq. 3.2 due to impurities mixed in the $\mathrm{LN_2}$.) The receiver voltages U_{det} at the cold and hot targets are registered during the absolute calibration. The solid line is the assumed calibration curve based on the underestimated T'_{cold}, while the dashed line is the true calibration, that could be achieved with an unbiased cold-target temperature T_{cold}. The voltage measured at an atmospheric scene U_{atm} is related to an underestimated brightness temperature ($\mathrm{BT}'_{\mathrm{atm}}$) when using the assumed calibration curve compared to the true curve.

It is suggested to subtract the spectral mean offset (Fig. 3.12) from the corresponding BTs of each flight. The campaign averaged offset should be used as best guess for flights with three or less clear-sky sondes. It cannot be excluded, that the forward simulation and used absorption models themselves have errors, which could explain the spectral shape of the offsets. However, as the same forward simulator will be used in the retrieval development (Chapter 4) the same offsets would be assumed by the retrieval, i.e., the forward model offsets would cancel out.

3.4. Summary and Conclusions

The HAMP radiometer sensitivity, i.e., sensors noise, and accuracy, i.e., absolute calibration, are analyzed using a ground based experiment with calibration targets, forward-simulated dropsondes and in-flight HAMP acquisitions. The NeDT estimates from ground-based sampling on reference targets agree with noise estimate from radiatively-homogeneous atmospheric scenes. The NeDT of HAMP-KV and HAMP-WF are clearly below 0.3 and 0.6 K, respectively, in the NARVAL configuration, which was used during NARVAL-South and NARVAL2 before the HAMP was upgraded for EUREC[4]A. The ground-based NeDT estimates of the HAMP-G deviate from the in-flight estimated. The in-flight HAMP-G NeDT is lower than 0.5 K, but a variability up to 1.2 K was observed during a ground experiment. In addition, high positive and negative correlation coefficients were observed in the ground based noise experiment. These were probably related to failing hardware components, which were replaced with the upgrade for EUREC[4]A, such that the new covariance is smoother.

The absolute accuracy of HAMP, which is necessary for a bias-free retrieval application, is assessed by comparing synthetic HAMP measurements based on dropsonde data with actual HAMP measurements in clear-sky scenes. Systematic channel-dependent offsets are found between forward-simulated dropsondes and simultaneous HAMP measurements. The offsets cannot be explained by a realistic and systematic error in the dropsonde measurements. The offset of each channel does not drift during one flight, but varies similarly in channels with similar sensitivity to the atmosphere. This random covariance can be attributed to the spatiotemporal mismatching of HAMP nadir measurements and the drifting sondes.

More-positive warm offsets are found in the NARVAL-South data, while NARVAL2 offsets are more negative and have a slightly smaller absolute offset. This could be due to the upgraded calibration targets, but unfortunately no meaningful direct comparisons of the two targets were feasible under real conditions. However,

ground-based measurements indicate discrepancies in the calibration routine as some offsets, even though with significantly smaller amplitude, were already seen after the calibration. Nevertheless, the in-flight offsets are larger and most likely result from changes within the ventilated belly pod during takeoff. For example, an environmental influence on the foam windows in the belly pod through condensation cannot be excluded. However, the stability of the offset indicates, that the condensate wouldn't have changed much during one flight. In addition, impurities in the LN_2 used during the absolute calibration in Barbados probably caused a bias in the assumed temperature of the cold calibration reference. Hardware issues of HAMP-G resulted in gain instabilities which are visible as BT sawtooth patterns (Konow et al., 2019). These instabilities limit the availability of HAMP-G data, such that less dropsonde comparisons are possible to assess the absolute accuracy. However, those comparisons that are possible show no general drift of HAMP-G BTs. The stability of the offset during a flight and its systematic spectral shape allow for an offset correction of each channel based on the mean offsets of each flight. The mean standard deviation of the offsets within one flight is below 1.4 K in the K-, W-, and F-band channels and below 0.9 K in the V- and G-band channels. More than 80 % (derived from Fig. 3.11) of this variation in the K- and W-band channels is attributed to the spatiotemporal mismatching of HAMP nadir measurements and the drifting sondes. The results presented here allowed the author to contribute an offset correction to the HAMP datasets published by (Konow et al., 2019) in the Climate and Environmental Retrieval and Archive (CERA) (Konow et al., 2018a,b,c,d). The random uncertainty of the offset-corrected BT data in the K- and W-band, which are the essential channels used by the retrievals in Chapter 4, is estimated 0.5 K on average for all channels considering the low NeDTs and strong contribution of the sonde drifting error to the offsets.

Chapter 4.

Microwave Radiometer Retrievals of Hydrometeors and Water Vapor

The content of the following Chapter has been published by Jacob et al. (2019). It investigates the research question 2 posted in the introduction (Sect. 1.4):

2. How can geophysical quantities be retrieved from HAMP observations? The focus is on the liquid hydrometeors, i.e. LWP and RWP, but also the cloud environment in terms of water vapor, i.e. IWV, matters. What is the uncertainty of LWP, RWP, and IWV retrievals from HAMP? Which differences between the dry season in December 2013 (NARVAL-South) and the wet season in August 2016 (NARVAL2) can be identified from the retrieved products?

Novel retrievals are developed for liquid water path (LWP), which is the main characteristic of trade-wind clouds that is observable with HAMP. In addition, the separate contribution by raindrops, in terms of the rain water path (RWP), and cloud droplets to the LWP is studied, and the moisture environment of the clouds is characterized by newly developed retrievals of the integrated water vapor (IWV).

The present thesis focuses on trade-wind convection. Therefore we compare retrieval results from the first two field experiments NARVAL-South and NARVAL2 during which HAMP was used in the tropics. Thus the northern part of the first set of NARVAL campaigns in 2013–2014 is not considered and NARVAL-South is simply referred to as "NARVAL1". Furthermore, the HAMP-W channel is simply referred to by its frequency (90 GHz). On page 70, a footnote is added, which refers the BT offset correction published by Konow et al. (2019) to Chapter 3.

The author of this thesis conceptualized the study together with advice from Susanne Crewell. Furthermore, he developed the HAMP retrievals and a corresponding retrieval database, conducted retrieval assessments and the analysis of the two seasons, and wrote the paper with support and input from all co-authors. Felix Ament and Susanne Crewell contributed to the design of the field experiments by initiating the DFG HAMP project. Martin Wirth and Manuel Gutleben contributed the lidar water vapor profiles and cloud mask, respectively. The lidar water vapor profiles are used to asses the newly developed IWV retrieval, while the cloud mask is used to enhance the LWP retrieval by identifying clear-sky scenes. Heike Konow organized

the quality controlled HAMP data, considering the BT offset correction developed in Chapter 3, and dropsonde data in a unified file format. The latter is used to asses the IWV retrieval as well. Mario Mech designed the principle of the PAMTRA simulations, which are used to create the retrieval database.

Investigating the liquid water path over the tropical Atlantic with synergistic airborne measurements

Marek Jacob[1], Felix Ament[2], Manuel Gutleben[3], Heike Konow[2], Mario Mech[1], Martin Wirth[3], Susanne Crewell[1]

[1]Institute for Geophysics and Meteorology, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany

[2]Meteorological Institute, Universität Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

[3]German Aerospace Center, Münchener Str. 20, 82234 Oberpfaffenhofen-Wessling, Germany

Correspondence: Marek Jacob (marek.jacob@uni-koeln.de)

Abstract. Liquid water path (LWP) is an important quantity to characterize clouds. Passive microwave satellite sensors provide the most direct estimate on a global scale but suffer from high uncertainties due to large footprints and the superposition of cloud and precipitation signals. Here, we use high spatial resolution airborne microwave radiometer (MWR) measurements together with cloud radar and lidar observations to better understand the LWP of warm clouds over the tropical North Atlantic. The nadir measurements were taken by the German High Altitude and LOng range research aircraft (HALO) in December 2013 (dry season) and August 2016 (wet season) during two Next-generation Advanced Remote sensing for VALidation (NARVAL) campaigns.

Microwave retrievals of integrated water vapor (IWV), LWP, and rainwater path (RWP) are developed using artificial neural network techniques. A retrieval database is created using unique cloud-resolving simulations with $1.25\,\mathrm{km}$ grid spacing. The IWV and LWP retrievals share the same eight MWR frequency channels in the range from 22 to $31\,\mathrm{GHz}$ and at $90\,\mathrm{GHz}$ as their sole input. The RWP retrieval combines active and passive microwave observations and is able to detect drizzle and light precipitation. The comparison of retrieved IWV with coincident dropsondes and water vapor lidar measurements shows root-mean-square deviations below $1.4\,\mathrm{kg\,m^{-2}}$ over the range from 20 to $60\,\mathrm{kg\,m^{-2}}$. This comparison raises the confidence in LWP retrievals which can only be assessed theoretically. The theoretical analysis shows

that the LWP error is constant with $20\,\mathrm{g\,m^{-2}}$ for LWP below $100\,\mathrm{g\,m^{-2}}$. While the absolute LWP error increases with increasing LWP, the relative one decreases from $20\,\%$ at $100\,\mathrm{g\,m^{-2}}$ to $10\,\%$ at $500\,\mathrm{g\,m^{-2}}$. The identification of clear-sky scenes by ancillary measurements, here backscatter lidar, is crucial for thin clouds (LWP $<$ $12\,\mathrm{g\,m^{-2}}$) as the microwave retrieved LWP uncertainty is higher than $100\,\%$.

The analysis of both campaigns reveals that clouds were more frequent ($47\,\%$ vs. $30\,\%$ of the time) in the dry than in the wet season. Their average LWP (63 vs. $40\,\mathrm{g\,m^{-2}}$) and RWP (6.7 vs. $2.7\,\mathrm{g\,m^{-2}}$) were higher as well. Microwave scattering of ice, however, was observed less frequently in the dry season ($0.5\,\%$ vs. $1.6\,\%$ of the time). We hypothesize that a higher degree of cloud organization on larger scales in the wet season reduces the overall cloud cover and observed LWP. As to be expected, the observed IWV clearly shows that the dry season is on average less humid than the wet season (28 vs. $41\,\mathrm{kg\,m^{-2}}$). The results reveal that the observed frequency distributions of IWV are substantially affected by the choice of the flight pattern. This should be kept in mind when using the airborne observations to carefully mediate between long-term ground-based and spaceborne measurements to draw statistically sound conclusions.

4.1. Introduction

Clouds and precipitation are a fundamental part of the Earth's climate system and significantly contribute to the water and energy cycle. However, the great variability of clouds, the complex interaction of small-scale processes involved, and their coupling to atmospheric circulation make them a major source of uncertainty in numerical climate and weather models (e.g., Bony et al., 2015; Boucher et al., 2013). Sherwood et al. (2014) attribute especially shallow marine clouds to contribute largely to the intermodel spread of climate models. Such clouds are particularly difficult to assess from spaceborne sensors due to their small size, with about $70\,\%$ appearing in sizes of less than $2\,\mathrm{km}$ over the tropical North Atlantic (Schnitt et al., 2017). The accurate observation of thin liquid clouds is an ongoing and important challenge as they cover more than a quarter of the globe and are an important contribution to Earth's energy balance (Turner et al., 2007).

Liquid water content (LWC) is the key parameter to describe clouds in atmospheric models. Due to the even higher difficulty in observing LWC profiles (Crewell et al., 2009), we focus on the liquid water path (LWP). It describes the total mass of all liquid water droplets in an atmospheric column above a unit surface area. However, care has to be taken as to whether LWP only denotes the contribution by cloud

droplets, later on called CLWP, or whether it also includes the contribution by liquid precipitation, i.e., drizzle and raindrops (rainwater path, RWP). Thus, we define LWP as the sum of CLWP and RWP. Furthermore, the observed LWP per se is an average over the sensors' field of view, which is affected by cloud and rain inhomogeneity, and the clear-sky contribution. Therefore, the spatial resolution is key information to interpret LWP statistics.

Few global (C)LWP datasets exist, and differences in global mean (C)LWP of a factor of 2 are reported by Lohmann and Neubauer (2018). These findings reflect the different sensing principles, i.e., microwave radiometry and visible–near-infrared techniques. Satellite microwave imagers such as the Special Sensor Microwave Imager (SSM/I) provide CLWP estimates for several decades but are limited to the ice-free oceans where the background signal is low. (C)LWP is mainly derived from the thermal emission signal in window regions with low water vapor contribution. Microwave receivers also sense rainwater within the satellite footprint which can be as large as several tens of kilometers. Recently, the Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP; Elsaesser et al., 2017) covering the period 1988 to 2016 has been generated. Elsaesser et al. (2017) additionally estimate the contribution of RWP to the total LWP by a simple parametrization and recommend only using those values with a ratio RWP : LWP of less than 0.2. The average MAC RWP : LWP ratio in our area of interest is 0.23 and 0.30 in December 2013 and August 2016, respectively. Therefore, a more detailed assessment of the rain cloud partitioning is important to better interpret satellite measurements in our study area. Greenwald et al. (2018) evaluate MAC-LWP using measurements by the Moderate Imaging Spectroradiometer (MODIS), the CloudSat Profiling Radar (CPR; Stephens et al., 2002), and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP; Winker et al., 2007b). They found in some cases a net LWP bias of more than 50 percent of the mean CLWP due to the combined effects of the in-cloud and adjacent precipitation biases as well as the cloud–rain partition.

Visible–near-infrared techniques such as those applied to MODIS exploit the spectral response of reflected sunlight to derive CLWP from optical depth and effective radius retrievals and are therefore limited to daytime. As the signal mainly relates to the upper part of the cloud, assumptions of the cloud vertical structure introduce uncertainties (Zhou et al., 2016). The horizontal MODIS resolution of about 1 km is much better than that of microwave satellites. Therefore, MODIS data have also been used to assess the clear-sky bias of microwave retrievals (Greenwald et al., 2018), to combine them with microwaves for a better assessment of low clouds (Masunaga et al., 2002), and to detect the ratio of rainwater and cloud water in low-latitude shallow marine clouds via combination with CPR (Lebsock et al., 2011).

In summary, quantifying the accuracy of CLWP and RWP observations is a major challenge as no absolute reference exists. While shipborne microwave observations have potential for satellite CLWP evaluation (Painemal et al., 2016), they fail during precipitation events, due to a wet radome.

In this study, we use the Next-generation Advanced Remote sensing for VALidation studies (NARVAL; Stevens et al., 2019) expeditions for investigating LWP and its uncertainty over the tropical North Atlantic. The NARVAL missions aim at improving the understanding of clouds, their role in the distribution of water in the atmosphere, and their interaction with the environment (Bony et al., 2015). Within NARVAL, the German High Altitude and LOng range research aircraft (HALO; Krautstrunk and Giez, 2012) was configured as an airborne cloud observatory combining active and passive microwave instruments with water vapor lidar, solar reflectance measurements, and dropsondes. Measurements taken during two campaigns in December 2013 (dry season) and in August 2016 (wet seasons) allow the study of clouds with similar, but more sensitive and higher spatially resolving instrumentation than that available on satellites.

Schnitt et al. (2017) demonstrate the ability of the HALO NARVAL 2013 instrumentation to characterize shallow clouds over the tropical North Atlantic in terms of size, integrated water vapor (IWV), CLWP, and surface reaching precipitation using classical regression algorithms. Their study uses the 1 km resolution HAMP data to show the sub-footprint variability of spaceborne CLWP estimation of about 30 km resolution. Further they illustrate how MODIS products at 1 km resolution likely underestimate CLWP of thick clouds due to MODIS' sensitivity towards the upper part of the cloud. In this study, we refine the (C)LWP retrieval by making use of high-resolution simulations that start to resolve cloud-scale circulations and were performed over the full tropical North Atlantic with the ICON (ICOsahedral Non-hydrostatic) weather model to support the analysis (Klocke et al., 2017). We further assess the total LWP retrieval accuracy over a wide range of cases, extend the retrieval towards a separation of rain and clouds, and reanalyze the dry season measurements in relation to the wet season campaign.

First, we aim to provide an accurate LWP dataset including uncertainty estimates to support the NARVAL overall goals. For this purpose, we develop retrieval algorithms using multi-channel microwave radiometer measurements as input for LWP and – based on the similar principle – IWV. The novel cloud-resolving ICON simulations serve as a training dataset (Sect. 4.2). In contrast to LWP, IWV can be evaluated using simultaneous measurements by dropsondes and water vapor lidar. The evaluation is presented in Sect. 4.3. The assessment of LWP (Sect. 4.4) reveals the

importance of using ancillary measurements, e.g., lidar measurements, for low LWP values and cloud radar measurements for lightly precipitating cases. For the latter an RWP retrieval is developed and assessed (Sect. 4.5). Finally, the campaign data are analyzed to investigate differences between dry and wet seasons (Sect. 4.6).

4.2. Material and Methods

This section presents the data and methods used in this study. That includes an introduction to the two NARVAL campaigns and the relevant measurements that were conducted during both campaigns. Furthermore, the generation of the retrieval database and the subsequent retrieval development are presented.

4.2.1. Campaign Overview

During the NARVAL expeditions HALO was operated out of Grantley Adams International Airport in Barbados to observe trade wind cumuli and their environment over the tropical North Atlantic (Fig. 4.1). Different flight patterns were chosen to perform satellite underflights, survey the area, probe the environment of a tropical cyclone, and to determine the large-scale vertical motion by launching several dropsondes within circles of approx. 170 km diameter (Bony and Stevens, 2018). In total eight research flights were performed during NARVAL1-South in December 2013 and 10 flights during NARVAL2 in August 2016. NARVAL1 also included research flights in the northern sector of the Atlantic which are not considered here. For simplicity we refer to the southern part as NARVAL1 in the following.

Flights were scheduled during local daytime. Flight altitudes varied between 6.4 and 15.0 km with an average speed above ground of 237 and 207 m s^{-1} during NARVAL1 and NARVAL2, respectively. All further analyses refer to the area from 37 to 60° W and 7 to 20° N. A detailed description of the different research flights can be found in Klepp et al. (2014) for NARVAL1 and Stevens et al. (2019) for NARVAL2.

4.2.2. Measurements

The microwave radiometer (MWR) being part of the HALO Microwave Package (HAMP Mech et al., 2014) provides the key measurements for this study. HAMP was installed in a belly pod below the HALO fuselage in nadir-looking configuration. While the 26-channel MWR includes channels from 22 tp 195 GHz, we only use the seven K-band channels (22.24–31.40 GHz) and the 90 GHz channel to retrieve LWP
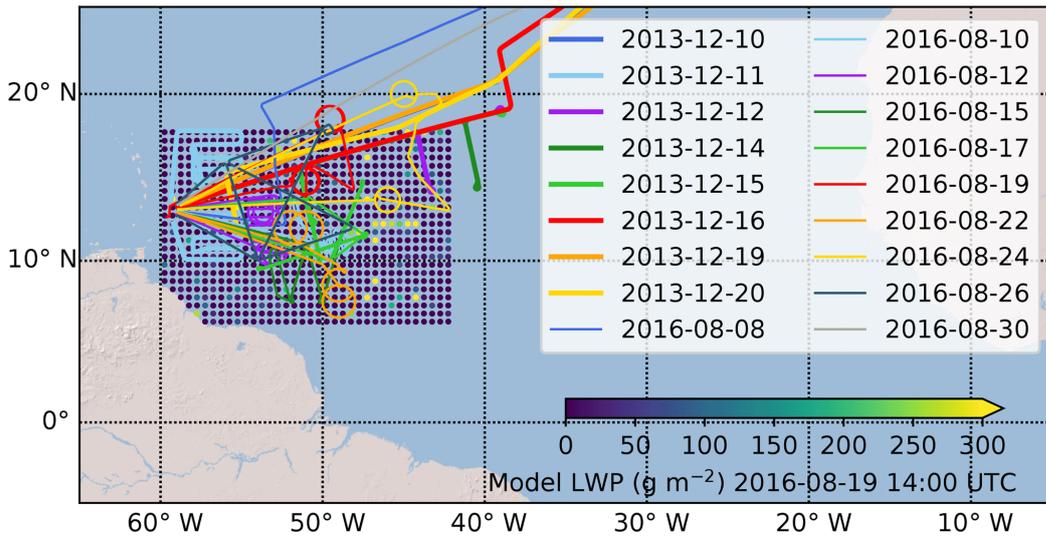
Figure 4.1.: NARVAL1 (thick lines) and NARVAL2 (thin lines) flight patterns. The study area of interest is depicted by subsampled ICON LWP from August 19, 2016 14:00 UTC. A grid point is shown every 0.5° as present in the retrieval database.

or IWV in the present work. At these frequencies ice particles do not influence the microwave signal substantially with the exception of precipitation sized particles.

As we focus on warm clouds only, cases of ice precipitation are filtered using the differential response of two frequencies along the 60 and 118 GHz oxygen lines. The channels at 53.75 and 118.75±1.4 GHz have similar temperature weighting functions but the higher frequency is more affected by ice scattering. The difference between a moving median of differential brightness temperature (BT) to the instantaneous differential BT is used to define the "ice flag". This procedure flags 1.2 % of the measurements of both campaigns.

Both liquid water and water vapor emit microwave radiation across the full microwave spectrum albeit with different spectral sensitivity (Fig. 4.2). BTs around the 22.235 GHz water vapor rotational line increase with increasing water vapor. The effect is strongest at the line center and decreases along the pressure-broadened wing of the absorption line. However, due to water vapor continuum absorption, BTs at window frequencies near 30 and 90 GHz are still affected. In contrast, the influence of liquid water is more dominant in the higher-frequency window channels than in absorption channels due to increasing emission with frequency. This can be best seen under low-humidity conditions by the increasing BT with increasing frequency. The near-surface wind speed slightly alters the BTs through modification of surface reflectivity and emissivity, as also shown in Fig. 4.2. This influence will

act as a random source of error on the LWP and IWV retrievals as no independent information to correct for wind influence is available.

Figure 4.2 illustrates the difficulty of retrieving LWP and IWV as in certain channels (e.g., 90 GHz) it is indistinguishable whether BT changes result from changes in IWV or LWP. Therefore, a combination of at least two channels is needed for retrieving IWV or LWP. Note that measurement errors in any of the channels affect both IWV and LWP retrievals (Crewell and Löhnert, 2003). This means that a good retrieval of either IWV or LWP indicates a good retrieval capability of the other. Thus, an accurate IWV retrieval is a prerequisite of a good LWP retrieval. Note that in most LWP retrievals (e.g., Wentz and Meissner, 2000) the liquid is assumed to consist of cloud droplets only, and therefore bulk approaches to calculate the liquid water absorption coefficients are used. However, for raindrops the Rayleigh approximation is not valid anymore, and Mie effects need to be considered, though the discrimination of the cloud and rain signal using MWRs is difficult.
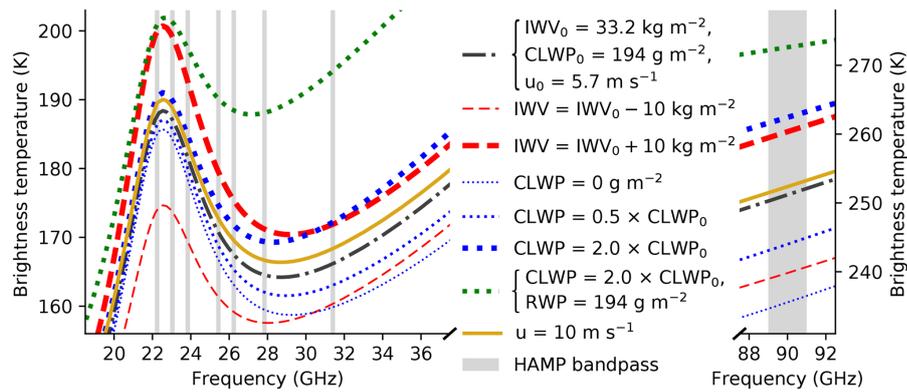
Figure 4.2.: Sensitivity of brightness temperatures in the K band and around 90 GHz to integrated water vapor (IWV), cloud liquid (CLWP), and rain water path (RWP), and 10 m surface wind (u). Dashed, dotted, and solid lines show variations in IWV and LWP (CLWP or RWP) and u, respectively. Bandpasses of the HALO Microwave Package (HAMP) channels are indicated by gray bars. Calculations are based on a thermodynamic dropsonde profile and a synthetic cloud in nadir geometry above the ocean.

The HAMP MWR measures BT with 1 s integration time and a noise level of less than 0.5 K in the considered channels. Despite ground calibration using hot/cold targets on the air field, BT offsets were identified by comparison with forward simulated dropsondes. Flight-dependent corrections were developed (Konow et al., 2019), and corrected BTs are available in the Climate and Environmental Retrieval and Archive (CERA) (Konow et al., 2018a,c).

HAMP also includes a 35.5 GHz cloud radar with a sensitivity of -30 dBZ at 13 km distance in the NARVAL setup. Profiles of the radar reflectivity factor (Z) and the linear depolarization ratio are recorded with 30 m vertical and 1 s temporal resolution. To supplement HAMP, Vaisala RD94 dropsondes were launched from HALO to provide the thermodynamic conditions of the environment. In total 76 and 215 sondes were released during NARVAL1 and NARVAL2, respectively.

To distinguish between clear-sky and cloudy conditions as well as possible, Schnitt et al. (2017) derive a cloud mask for NARVAL1 based on the nadir spectral solar radiance measurements by HALO-SR (HALO Solar Radiation; Fricke et al., 2014). Unfortunately, sun glint in August deteriorated the cloud mask retrieval during NARVAL2. Therefore, Gödde (2018) developed a cloud mask product using the imaging spectrometer specMACS (spectrometer of the Munich Aerosol and Cloud Scanner Ewald et al., 2016) which overcomes the sun glint problem. However, spec-MACS was not installed during NARVAL1. In order to have similar cloud mask performance during both campaigns the aerosol backscatter profile measured by the WAter vapor Lidar Experiment in Space (WALES) airborne demonstrator (Wirth et al., 2009) is used instead to provide an along-track cloud mask with 1 s resolution.

WALES also provides profiles of water vapor molecular number density based on the differential absorption lidar (DIAL) principle. These profiles are converted to volume mixing ratio profiles using temperature and pressure data from ECMWF analyses. A resolution of about 200 m vertical and 12 s temporal was chosen as a compromise between accuracy and resolution. The water vapor data are given on the vertical grid of the raw backscatter data which is 15 m but smoothed with an averaging kernel of 200 m width (full width at half maximum, FWHM). Water vapor profiles are provided down to about 250 m in cases with no or optically thin clouds, which can be penetrated by the lidar beam. Water vapor information is available below thin clouds, but the cloud itself is masked out in the profile.

The requirement of simultaneous measurements by all sensors reduces the dataset. While all research flights during NARVAL1 can be used, no data is available for some NARVAL2 flight days due to hardware issues as summarized in Table 4.1. The spatial sampling differs even with the same temporal sampling due to footprints differences. The HAMP MWR has the largest beamwidth in its lowest-frequency channel of 5.0° (FWHM). The corresponding surface footprint at 10 km altitude is about 870 m across and 1090 m along track. The HAMP radar beamwidth is 0.6° whereas WALES has a field of view of 1.6 mrad. The respective footprints are about 105 m × 335 m and 16 m × 216 m. We reduce the along-track sampling differences

by averaging temporally, but the cross-track sampling issues remain. This means a cloud covering a lateral part of the MWR footprint might be missed by the lidar or even the radar. Cross-track imagers such as specMACS could be used to assess these issues. However, specMACS was only installed on HALO for NARVAL2, and the detailed analysis of HAMP beam filling is beyond the scope of this study. The problem of different footprints and sensitivities of different NARVAL instruments for cloud masking is illustrated by Stevens et al. (2019).

Table 4.1.: Dataset availability. Days of research flights from which the datasets are used for the study of NARVAL1 and NARVAL2, respectively.

Dataset	NARVAL1 Day in December 2013	NARVAL2 Day in August 2016
HAMP radiometer	10, 11, 12, 14, 15, 16, 19, 20	8, 10, 12, 15, 17, 19, 22, 24, 26, 30
HAMP radar	10, 11, 12, 14, 15, 16, 19, 20	8, 10, 12, 15, 17, 19, 22
Dropsondes	10, 11, 12, 14, 15, 16, 19, 20	8, 10, 12, 15, 17, 19, 22, 24, 26, 30
WALES water vapor	10, 11, 12, 14, 15, 16, 19, 20	10, 12, 15, 17, 19, 22, 24, 26, 30
WALES cloud mask	10, 11, 12, 14, 15, 16, 19, 20	8, 12, 15, 19, 22, 24, 26, 30

4.2.3. Retrieval Database

Recently, high-spatial-resolution simulations with the storm-resolving ICON model were able to show how resolved convection and its associated circulation interact with and form the larger-scale circulation within the Atlantic intertropical convergence zone (Klocke et al., 2017). These simulations serve as training and testing data for the retrieval algorithms. The simulations were performed on a triangular grid, with a horizontal spacing of about 1.25 km and 75 vertical levels. The simulations cover the area of 4° S to 18° N and 64 to 42° W. The data were spatially subsampled to reduce the computational effort while still covering the variability of atmospheric profiles. To eliminate atmospheric columns with a high degree of correlation, columns are selected on a $0.5° \times 0.5°$ longitude–latitude grid, so that each time step includes 849 cases over the ocean as indicated in Fig. 4.1. Data from 24 days with hourly outputs each, spanning the period of each campaign, are alternately separated into test and training data. In general, the training and test data exclude cases with LWP greater than $1000\,\mathrm{g\,m^{-2}}$, and cases with ice. This means 86 % of all profiles over the ocean are used. This limitation is done as our focus is on liquid clouds and their transition to rain. Note that classical satellite algorithms (e.g., Wentz and Meissner, 2000) are trained with an upper LWP limitation of $300\,\mathrm{g\,m^{-2}}$.

Synthetic HAMP measurements, i.e., BTs and radar reflectivity profiles in nadir view, are simulated for each model column based on its thermodynamic profile and hydrometeors (cloud liquid water, rain, cloud ice, snow, and graupel). The Passive and Active Microwave TRAnsfer code (PAMTRA Maahn et al., 2015; Cadeddu et al., 2017) is used. It is configured with 27 output levels to mimic different flight altitudes (6–15 km). The ICON model was set up with a one-moment microphysics scheme (Baldauf et al., 2011). In PAMTRA, cloud and rain particles and their size distributions are described according to the microphysical scheme of ICON and the single scattering properties for each particle are approximated by the Lorentz–Mie theory. Cloud and rain particles are simulated with a 20 μm diameter monodisperse and exponential distribution of water spheres, respectively. The exponential distribution has its intersect N_0 classically fixed to $0.08 \, \text{cm}^{-4}$ (Marshall and Palmer, 1948). Absorption coefficients of atmospheric gases (i.e., oxygen, water vapor, nitrogen) are calculated according to Rosenkranz (1998) with corrections of the water vapor continuum absorption according to Turner et al. (2009) and the line width modification of the 22.235 GHz water vapor line as proposed by Liljegren et al. (2005). The emissivity and reflectivity of the sea ocean surface are calculated using the FAST microwave Emissivity Model version 5 (FASTEM5 Liu et al., 2011), which is a modification of the Fresnel coefficients including corrections for ocean surface roughness and foam building as a function of wind speed.

To test the realism of the retrieval database, histograms of BTs were compared with their observed counterparts. Joint histograms of an absorption (22 GHz) and a window channel (31 or 90 GHz) show that the relations between channel pairs are depicted in the model and observations in the same way (Fig. 4.3). In clear-sky conditions absorption and window channels are highly correlated with both increasing with increasing moisture albeit the increase is less in the window channels. Clear-sky cases with low BT_{31} and BT_{90} are visible as a line of high occurrence and reveal the linear relation between absorption and window channel BTs as a function of IWV. The simulations and measurements show the same relations but differ slightly in terms of the BT combination distribution within this line as the underlying IWV sampling is slightly different. If liquid water clouds occur, the window channel BTs increase compared to clear-sky cases (solid lines in Fig. 4.3). The window channel at 90 GHz has a higher sensitivity towards LWP compared to BT_{31} as it can be seen by the increased LWP line spread. Rainy cases show higher emissions in all channels (dotted lines in Fig. 4.3). For thick clouds and rain the most-liquid-sensitive channel (90 GHz) experiences saturation effects with BT_{90} approaching cloud temperatures. The joint histograms reveal the major signals by liquid and water vapor which are exploited within retrieval algorithms. However, multiple influence factors like the ex-

act vertical structure lead to the variability illustrated in Fig. 4.3. Minor deviations between observations and simulations are visible in the frequency of combinations of BT_{31} and BT_{90} with high BT_{22}. Those combinations are associated with heavy precipitation and were observed less frequently than present in the model as flight patterns avoided the heaviest precipitation.
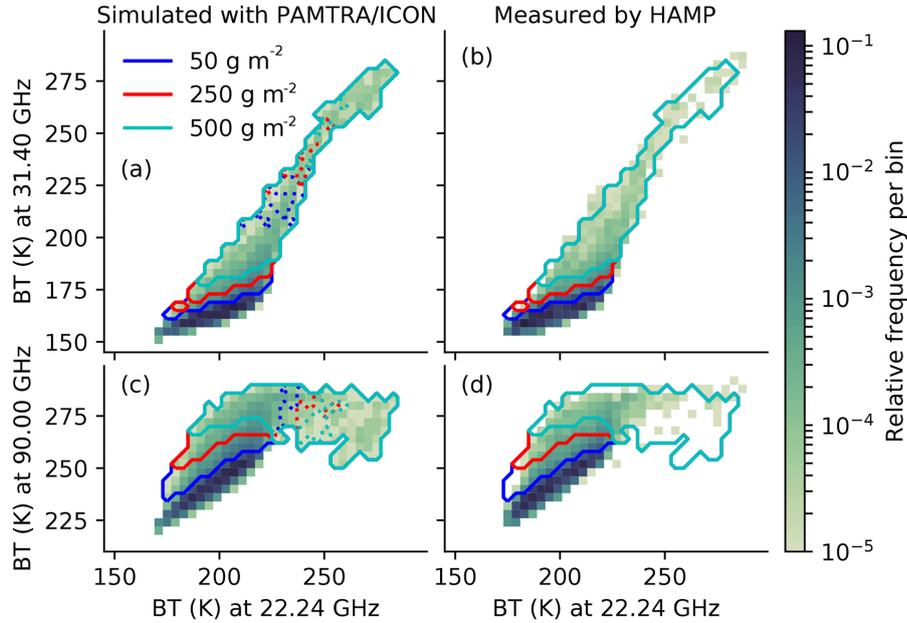
Figure 4.3.: Relation between brightness temperatures (BT) in an absorption channel (22.24 GHz) and two window channels, i.e., 31.40 GHz (a, b) and 90.00 GHz (c, d). Two-dimensional histograms of occurrences in simulations (a, c) and HAMP measurements (b, d). Solid contours highlight BT combinations in simulations that mostly occur with LWP higher than $50 \, g \, m^{-2}$, $250 \, g \, m^{-2}$, and $500 \, g \, m^{-2}$. Dotted lines highlight combinations of which RWP mostly exceeds the respective threshold. LWP contours in (b) and (d) are taken from (a) and (c) for guidance. HAMP data from all NARVAL2 flights and ICON–PAMTRA data of the corresponding dates are used. Profiles and measurements with ice are excluded.

4.2.4. Retrieval

The atmosphere emits radiation depending on the atmospheric state as illustrated in Fig. 4.2. In general, the retrieval of the atmospheric state from MWR measurements is under-determined as multiple atmospheric states can lead to the same set of BTs. Statistical relations have to be established to link the measurement to the most

common atmospheric state that can provide those measurements. To account for nonlinearity in this inverse modeling problem, we use an artificial neural network (NN) model similar to Cadeddu et al. (2009) to relate BTs to IWV and LWP. The NN is configured with eight input neurons and 15 hidden neurons in one hidden layer.

For testing and training the retrievals, Gaussian noise of $0.5\,\mathrm{K}$ is added to the simulated BTs to account for uncertainties of the HAMP MWR, the radiative transfer, and absorption model. The database is separated by alternating days into test and training data. Retrieval parameters are derived for each PAMTRA output level to account for the altitude dependence of the microwave signal. The output levels are chosen such, that a HALO flight level never deviates more than $90\,\mathrm{m}$ from the next output level. The parameters at output levels closest to HALO's altitude are interpolated to HALO's altitude in the retrieval application. Retrieval parameters are derived separately for both campaigns. For testing, each retrieval is applied to the test data of the campaign it is trained for.

In the retrievals, IWV and LWP, and later CLWP and RWP are the integrals of the water vapor and liquid water over the whole column as seen from space. Despite the fact that HALO flies lower, we chose the total integrals as they prevent artificial flight-level-depended biases in statistics and allow a comparison with satellite and model data. According to ICON model data, typically less than $0.1\,\mathrm{kg\,m^{-2}}$ water vapor is above a flight altitude of $10\,\mathrm{km}$. About $1\,\mathrm{kg\,m^{-2}}$ of IWV is not seen by the MWR at the lowest NARVAL2 flight altitude of $6.4\,\mathrm{km}$ but is included in the retrieval. The LWP retrieval is trained with the integral of all liquid water, that is given by the model either as cloud water or rainwater. The sum of both is used due to the difficulty of MWRs to distinguish clouds and rain (Fig. 4.2).

Neural network LWP retrievals are compared with linear regression (REG) models as used by Schnitt et al. (2017). The regression relates measured brightness temperatures BT_i to LWP including the quadratic terms of BT_i

$$\mathrm{LWP} = c + \sum_i \left(b_i \mathrm{BT}_i + a_i \mathrm{BT}_i^2 \right), \tag{4.1}$$

where a_i, b_i, and c are regression coefficients. Such REGs are less susceptible than NNs to extrapolation towards unforeseen input data, i.e., data values or combinations that are not covered by the training data. However, NNs are better in representing nonlinear effects that are apparent in microwave radiative transfer and thus can better adjust to the extremes of the LWP target space. The application of the retrievals to test data reveals overall uncertainties between $0.5\,\mathrm{kg\,m^{-2}}$ and $0.6\,\mathrm{kg\,m^{-2}}$ for IWV for both approaches, i.e. NN and REG, and $22\,\mathrm{g\,m^{-2}}$ and

$26\,\mathrm{g\,m^{-2}}$ for LWP using the NN and REG, respectively. For LWP the uncertainty strongly depends on atmospheric conditions as it will be investigated in Sect. 4.4.

When retrieval algorithms are applied to HALO measurements, slight biases of LWP from 0 with slow changes over time are observed during clear-sky scenes. To reduce these biases and to improve the retrieval of low LWP values, we follow the synergistic approach by van Meijgaard and Crewell (2005). Herein, we use the WALES cloud mask for clear-sky identification. HAMP measurements are considered clear sky if no cloud is detected by WALES within $\pm\,2$ seconds flight time. The distance-weighted average clear-sky LWP within $\pm\,30$ minutes is then subtracted from each a priori retrieved LWP value.

In thick clouds we occasionally observed, that while the REG retrieval gave LWP $>1000\,\mathrm{g\,m^{-2}}$, the NN LWP time series showed a sudden decline. This is likely caused by the clipping of the NN retrieval at $1000\,\mathrm{g\,m^{-2}}$, which is expected as the retrieval database is limited to LWP $<1000\,\mathrm{g\,m^{-2}}$ and thus BTs associated with higher amounts of liquid are unknown to the retrieval. To avoid this behavior, we use a second NN retrieval trained with an extended database up to $4000\,\mathrm{g\,m^{-2}}$ to flag scenes that are potentially above $1000\,\mathrm{g\,m^{-2}}$. Overall, $0.76\,\%$ of the measurements were masked in this way. Note that these measurements often coincide with ice scattering depressions in channels at higher frequencies.

To retrieve the contribution of raindrops (RWP) to the total LWP, the vertically integrated radar reflectivity is used in addition to the MWR channels in another NN retrieval. The aim is separating the LWP into CLWP and RWP, i.e., splitting the contributions from small cloud droplets and larger raindrops by estimating the fraction

$$f = \frac{\mathrm{RWP}}{\mathrm{LWP}} = \frac{\mathrm{RWP}}{\mathrm{RWP} + \mathrm{CLWP}}. \tag{4.2}$$

This retrieval is based on the hydrometeor classes of rain and cloud liquid water in the ICON model. The RWP is calculated by multiplying f and the retrieved total LWP.

4.3. Assessment of Integrated Water Vapor

Three independent methods to derive IWV are available from HALO: the MWR retrieval, vertically integrated humidity from dropsondes, and vertically integrated humidity from WALES. Each of the three methods has its advantages and shortcomings. The microwave radiometry can not provide profile information but gives continuous IWV under nearly all-sky conditions. The dropsondes provide in situ

measurements, but no valid data up to about the first half kilometer below the aircraft because of the sensor's adjustment from the aircraft cabin conditions to the outside. Furthermore, wind drifts sondes out of the aircraft nadir with a typical horizontal drift during the decent of $4\,\mathrm{km}$. The dropsonde relative humidity sensor has a repeatability of $2\,\%$ according to the manufacturer (Vaisala, 2017). This relates to an IWV accuracy of about $1.4\,\mathrm{kg\,m^{-2}}$. WALES provides water vapor profiles, but they are only available when no cloud extinguishes the laser beam. This limits the application of WALES for the IWV retrieval to clear-sky scenes.

To compute the numerical derivative in the DIAL equation, the first data point is at about $250\,\mathrm{m}$ above the sea surface and centered at the retrieval interval. Therefore, in the vertical integration, the missing near-surface information is filled with the median mixing ratio in the lowest five range bins. The median is chosen to reduce any surface artifacts which can occur, when the first raw data signal point used in the retrieval contains the surface reflex. We estimate that the error of this assumption is about $0.3\,\mathrm{kg\,m^{-2}}$ by analyzing dropsonde humidity profiles. The IWV estimation is discarded if information of more than $400\,\mathrm{m}$ above sea level is missing or there is a gap due to a thin cloud. Also, stability of the estimated WALES IWV is required, which means that the differences to the preceding and succeeding IWV estimations have to be smaller than $2\,\mathrm{kg\,m^{-2}}$.

An example of water vapor retrievals on August 19, 2016 is shown in Fig. 4.4. An elevated moisture layer between 3 and $4\,\mathrm{km}$ altitude is visible in the first half of the scene. Around 14:53 a plume of moist air reaching up into even higher levels causes an IWV gradient of nearly $10\,\mathrm{kg\,m^{-2}}$ (26–$35\,\mathrm{kg\,m^{-2}}$) over a distance of about $110\,\mathrm{km}$. This gradient is captured well by WALES and HAMP. The two dropsondes that were released between 14:45 and 14:55 reconstruct this gradient, but both have a dry offset. This offset might be due to drifting of the sonde towards the drier air mass. After a short outage of WALES at around 15:00, shallow clouds below $2\,\mathrm{km}$ prevent the determination of lidar IWV frequently. Most of the IWV measurements from dropsondes agree with the coincident remote sensing estimates within the sondes' uncertainty.

A more quantitative comparison is achieved by considering all measurements from both campaigns which cover a wide variety of water vapor conditions (Fig. 4.5). Overall, the sondes agree well with HAMP over the whole observed range from very low ($20\,\mathrm{kg\,m^{-2}}$) to very high ($60\,\mathrm{kg\,m^{-2}}$) values of IWV (Fig. 4.5a). The root-mean-square deviation (RMSD) is $1.39\,\mathrm{kg\,m^{-2}}$ ($1.28\,\mathrm{kg\,m^{-2}}$) with a mean bias of $0.28\,\mathrm{kg\,m^{-2}}$ ($0.47\,\mathrm{kg\,m^{-2}}$) during NARVAL1 (NARVAL2) as summarized in Table 4.2. The positive biases of HAMP are most likely caused by the retrieval, which
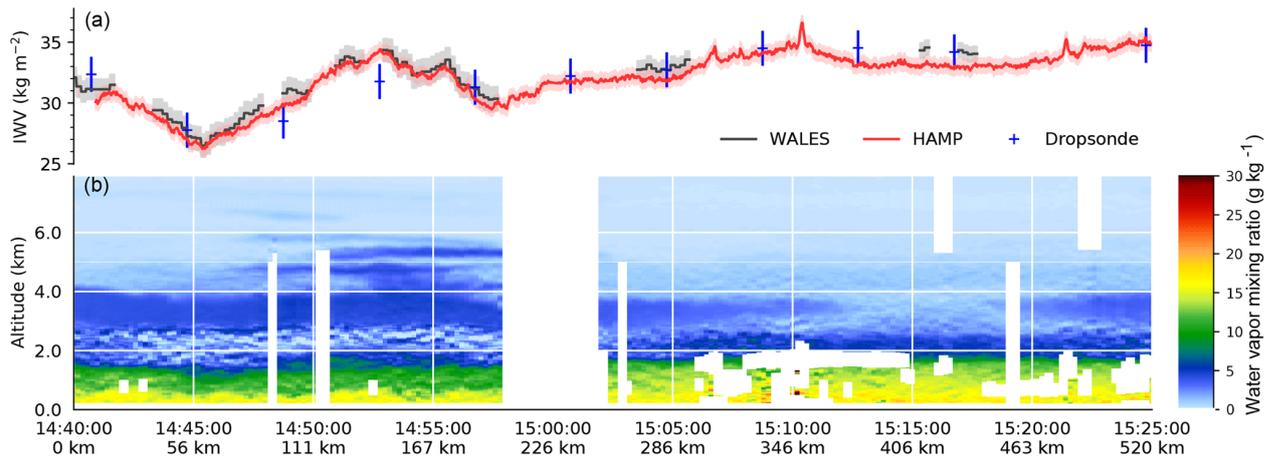
Figure 4.4.: Water vapor time series from NARVAL2 research flight 6 on August 19, 2016. Top: IWV time series by HAMP, WALES, and dropsondes with their uncertainties. Bottom: WALES water vapor profiles. White areas denote masked lidar data. The scene represents a circle around 14.8° N and 51.0° W over a distance of 520 km.

is trained with the whole column IWV, whereas the sonde IWV is only integrated along its measurement path. Most sondes were released above 9 km which would miss an IWV of about $0.2 \, \text{kg} \, \text{m}^{-2}$ according to ICON data. Note that dropsondes released from below 6.5 km are discarded in the comparison to avoid an artificial bias. The random error between HAMP and sondes ($1.2 \, \text{kg} \, \text{m}^{-2}$) is smaller than the estimated uncertainties of the dropsonde ($1.4 \, \text{kg} \, \text{m}^{-2}$) and the MWR retrieval ($0.6 \, \text{kg} \, \text{m}^{-2}$), which indicates the high quality of the measurements as uncertainties due to spatiotemporal mismatch are included in the RMSD as well. Note that uncertainties due to MWR calibration are largely compensated for as offsets between measured BT and those derived by radiative transfer calculations based on dropsondes have been corrected by Konow et al. (2019)[1].

WALES IWV can be used for continuous comparison to HAMP IWV along the flight track in clear-sky scenes. A comparison of all coincident measurements during NARVAL2 is depicted in Fig. 4.5b. The average bias between HAMP and WALES IWV is $-0.59 \, \text{kg} \, \text{m}^{-2}$. The bias is cut in half when considering only the 40 simultaneous measurements during which a dropsonde was launched (Table 4.2). The random error is smaller in contrast to the HAMP–dropsonde comparison. This is likely due to the better spatial match between the two nadir measurements compared to a drifting sonde. However, higher RMSD between HAMP and WALES IWV can be found during NARVAL1, which is mostly related to a higher bias. The bias increases

---

[1]This offset correction is described in Chapter 3.
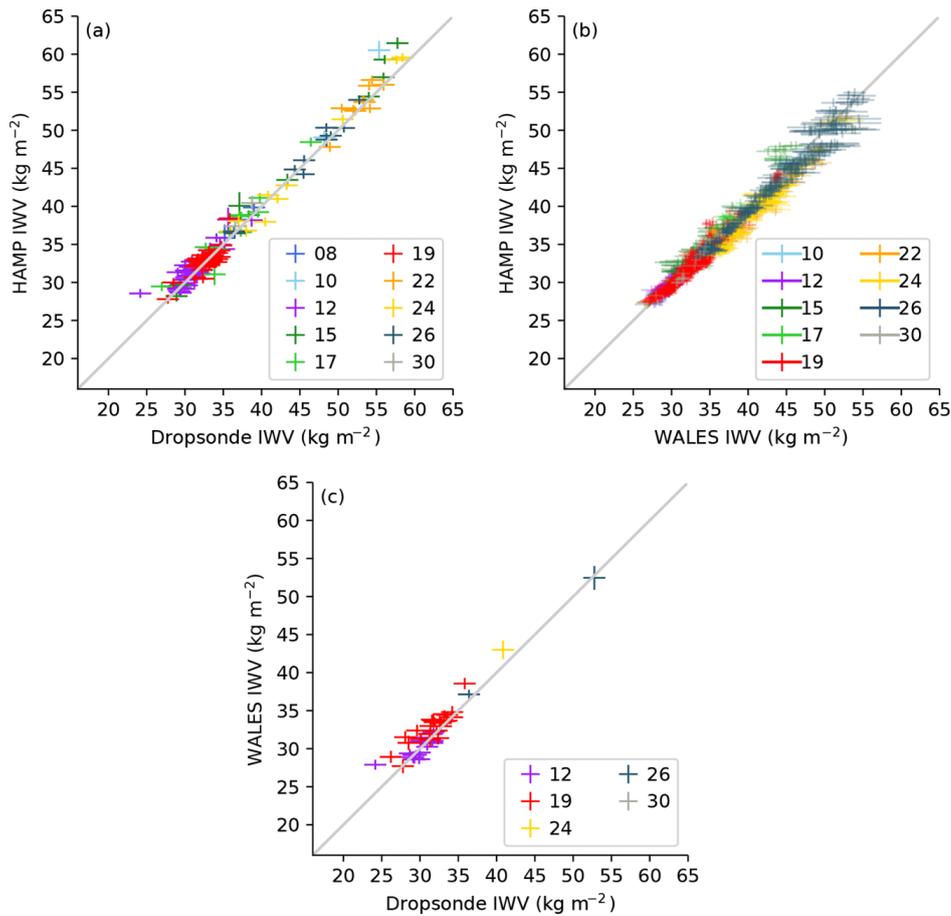
Figure 4.5.: IWV comparison of HAMP MWR and dropsondes (a), HAMP MWR and WALES (b), and WALES and dropsondes (c) during NARVAL2. The colors indicate the flight days in August 2016. Scores are given in Table 4.2.

to $1.70 \, \mathrm{kg \, m^{-2}}$ in the HAMP–WALES comparison when only considering measurements during which a sonde was released. A bias of similar magnitude is apparent between WALES and the dropsondes. Most likely the dry bias of WALES is due to the method of how the 12 s water vapor profiles are derived. The profiles only contain raw profiles (within the 12 s), that are not blocked by a cloud. For small-scale boundary layer convection, this means preferred sampling of downdraft regions. In these downdraft regions dry air is entrained from the rather dry free troposphere into the convection layer during NARVAL1 (Stevens et al., 2017). This results in biased sampling of rather dry profiles. During NARVAL2 humidity was reaching higher altitudes, which resulted in less entrainment of dry air in cloud gaps.

A small confounding effect from liquid water in cloudy scenes can be derived from the separation of the HAMP–dropsonde comparison into all ("observed pairwise")

Table 4.2.: Comparison of IWV retrieved from HAMP, WALES, and dropsondes. Pairwise observations of two instruments and the subsets of the observations for that all instruments were available. Bias, root-mean-square deviation (RMSD), and bias-corrected RMSD (standard deviation, SD) in kilograms per square meter $(\mathrm{kg\,m^{-2}})$. "Observed by all" implies only small or no clouds.

		HAMP - sondes	HAMP - WALES	WALES - sondes
NARVAL1 observed pairwise	bias	0.28	0.92	−1.21
	RMSD	1.39	1.36	1.60
	SD	1.38	1.01	1.07
	count	43	2482	24
NARVAL1 observed by all	bias	0.32	1.70	−1.37
	RMSD	1.21	2.20	1.70
	SD	1.20	1.41	1.03
	count	21	21	21
NARVAL2 observed pairwise (Fig. 4.5)	bias	0.47	−0.59	0.73
	RMSD	1.28	1.21	1.38
	SD	1.19	1.06	1.19
	count	146	1632	47
NARVAL2 observed by all	bias	0.32	−0.25	0.57
	RMSD	1.16	0.82	1.23
	SD	1.12	0.79	1.11
	count	40	40	40

and clear sky ("observed by all", i.e., when WALES is also available) in Table 4.2. In the NARVAL1 dataset, the bias for cloudy-sky sondes $(0.24\,\mathrm{kg\,m^{-2}})$ is somewhat smaller than that for clear sky $(0.32\,\mathrm{kg\,m^{-2}})$. However, RMSD and SD in cloudy scenes are about $0.3\,\mathrm{kg\,m^{-2}}$ larger than in clear sky. NARVAL2 also shows a larger bias in cloudy sky of about $0.53\,\mathrm{kg\,m^{-2}}$ in comparison to clear sky $(0.28\,\mathrm{kg\,m^{-2}})$. The cloudy-sky RMSD and SD of $1.32\,\mathrm{kg\,m^{-2}}$ and $1.21\,\mathrm{kg\,m^{-2}}$, respectively, are only slightly larger than their clear-sky counterparts. An increase of the random error for cloudy scenes is expected as also higher water vapor variations are expected in heterogeneous cloud fields.

With the exception of the HAMP–WALES comparison during NARVAL1, the RMSD between the different instrument pairs is found between $0.8\,\mathrm{kg\,m^{-2}}$ and $1.4\,\mathrm{kg\,m^{-2}}$

(Table 4.2). This corresponds to an error of 2% to 7% over the observed range of 20 to $60\,\mathrm{kg\,m^{-2}}$. For comparison, Mears et al. (2015) found random IWV deviations between different spaceborne MWR and ground-based GPS (Global Positioning System) instruments of 1.7 to $2.0\,\mathrm{kg\,m^{-2}}$ over a similar IWV range using 26 small island stations located mainly in the tropics.

The HAMP IWV retrieval has a theoretical uncertainty of about $0.6\,\mathrm{kg\,m^{-2}}$, which is derived by applying the IWV retrieval to simulated measurements from the test database (Sect. 4.2.4) and is constant over a wide IWV range (not shown). This is well in line with the RMSD derived in the pairwise comparisons taking into account the estimated uncertainties of WALES and dropsondes as well as uncertainties due to the spatiotemporal mismatch. In summary, the pairwise comparisons in relation to the individual uncertainties indicate high HAMP IWV performance and the suitability of our retrieval approach.

4.4. Assessment of Liquid Water Path

There are no independent measurements of sufficient quality to assess the quality of the LWP product. However, the large retrieval database (173 339 ice-free cases in the test dataset) allows a theoretical in-depth analysis of the retrieval performance. This approach is supported by the good consistency between the BTs in the database and the HAMP measurements in terms of relation resemblance (Fig. 4.3) and performance of IWV retrieval (Sect. 4.3). We analyze the retrieval error as a function of the true LWP as well as of the retrieved LWP using the database.

First, we analyze the difference of retrievals developed with all ice-free cases of the training database (all sky) and with cloudy cases only, which reduces the dataset size to about one-quarter. A model profile is regarded as cloudy if LWP $> 1\,\mathrm{g\,m^{-2}}$. REG and NN retrievals are trained with the all-sky and the cloudy-sky datasets separately. The errors of retrieved LWP from the test database are calculated for bins of the true LWP. Both REG and NN show similar behavior of the RMSD between the retrieved LWP and the model truth with increasing LWP (Fig. 4.6a). The RMSD is constant for LWP below about $30\,\mathrm{g\,m^{-2}}$ and increases with LWP, e.g. $50\,\mathrm{g\,m^{-2}}$ at $500\,\mathrm{g\,m^{-2}}$. For LWP values $> \sim 800\,\mathrm{g\,m^{-2}}$ the number of test cases reduces strongly, leading to less robust results. For LWP $< 30\,\mathrm{g\,m^{-2}}$, the errors are smaller for REG and NN retrieval types if the clear-sky cases are included in the training (compare Fig. 4.6a and 4.6b). Including clear sky in the training, the retrieval errors decrease slightly for a REG model and are almost cut to half for an NN. This shows the ability of an NN to nonlinearly relate a variety of BT combinations to zero LWP. However,

retrievals that are especially trained for all-sky scenes have a considerably larger RMSD for LWP $> 20\,\mathrm{g\,m^{-2}}$ than those trained with cloudy cases only as clear-sky cases make up $77\,\%$ of the data. Since we are targeting clouds and not clear sky, we chose a retrieval trained with data excluding the clear-sky model profiles. Instead of including clear sky directly in the retrieval, we make use of lidar measurements, which are better suited than MWR for cloud masking.

Regarding cloudy-sky retrievals, the RMSD for a given (true) LWP less than $40\,\mathrm{g\,m^{-2}}$ is smaller when using the NN retrieval instead of a REG model (Fig. 4.6a). This is related to a suppression of unphysical negative LWP values by the NN. Thus, in contrast to a REG which has a nearly Gaussian error characteristic, the NN tends to overestimate LWP. This results in a more negative mean LWP error (true minus retrieved) of clouds with less than $10\,\mathrm{g\,m^{-2}}$ but also in a smaller interquartile range of errors when using the NN instead of the REG. However, the retrieval error for true LWP $< 10\,\mathrm{g\,m^{-2}}$ remains on the order of 10 to $18\,\mathrm{g\,m^{-2}}$ even when using the NN.

The bias errors visible in Fig. 4.6a can not be used to adjust the retrieved LWP as the true LWP value is not known in practice. For the application of the error analysis on measurements, it is important to analyze the LWP error as a function of the retrieved LWP. The RMSDs of the NN and REG retrievals are larger than $100\,\%$ for a retrieved LWP below $12\,\mathrm{g\,m^{-2}}$, which can be regarded as a detection limit (Fig. 4.7). Therefore, ancillary measurements with higher sensitivity are needed to detect these thin liquid water clouds. The RMSD is below $20\,\mathrm{g\,m^{-2}}$ for REG LWP $< 50\,\mathrm{g\,m^{-2}}$ and NN LWP $< 100\,\mathrm{g\,m^{-2}}$ and moderately increases with increasing LWP. Therefore, the relative RMSD decreases from $50\,\%$ for a retrieved LWP of about $40\,\mathrm{g\,m^{-2}}$ to $20\,\%$ for LWP $> 100\,\mathrm{g\,m^{-2}}$ for both retrieval types. While the RMSD is rather similar for REG and NN, the NN succeeds in capturing the nonlinear retrieval, providing a nearly zero bias across the full LWP range, and is therefore preferred over REG.

Analyzing the retrieved LWP distribution for clear-sky scenes is a widely used method to assess an LWP retrieval (e.g., Liu et al., 2001; Greenwald et al., 2018) because this characterization can be made from measurements using ancillary observations that define clear-sky scenes. We use WALES measurements for the indication of cloud and clear sky. The distributions of LWP from HAMP MWR are depicted in Fig. 4.8 for observed clear-sky scenes (blue lines) along the track for both campaigns. The distributions are compared to the theoretical ones of retrieved LWP from all clear-sky (true LWP $< 1\,\mathrm{g\,m^{-2}}$) cases of the respective campaign in the ICON–PAMTRA database (orange lines in Fig. 4.8). The latter distributions
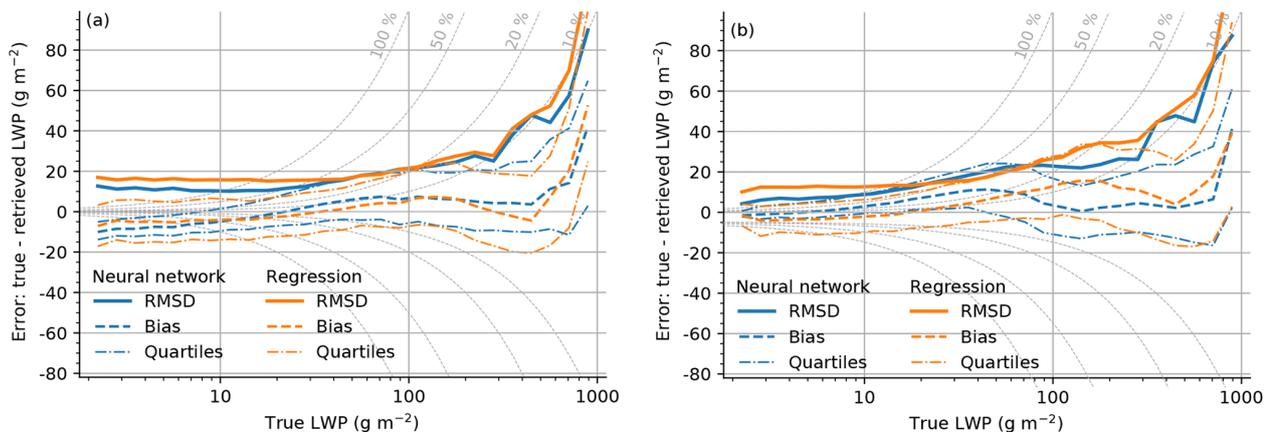
Figure 4.6.: Expected retrieval error as a function of true LWP for neural network and linear regression LWP retrievals. Retrievals (a) trained for $1\,\mathrm{g\,m^{-2}}$ < LWP < $1000\,\mathrm{g\,m^{-2}}$. Retrievals (b) trained including clear-sky cases (LWP < $1000\,\mathrm{g\,m^{-2}}$). Error measures (colored lines) for logarithmically distributed bins with 10 bins per LWP power of 10. Gray dashed lines denote the corresponding relative LWP error.
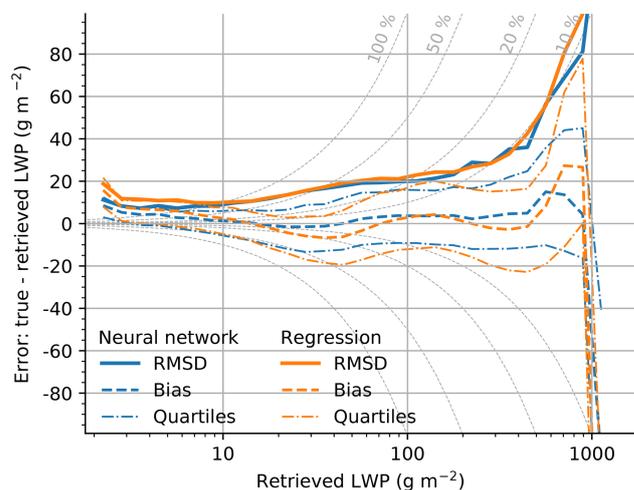
Figure 4.7.: As Fig. 4.6, but with errors shown as a function of retrieved LWP. Retrievals are trained and tested with $1\,\mathrm{g\,m^{-2}}$ < LWP < $1000\,\mathrm{g\,m^{-2}}$. The first bin contains all data with retrieved LWP < $2.5\,\mathrm{g\,m^{-2}}$ (including negative).

are closely related to the retrieval uncertainty of the lowest LWP bin in Fig. 4.6a as this represents the retrieval uncertainty for true LWP $< 2.5\,\mathrm{g\,m^{-2}}$. The distributions roughly resemble Gaussian behavior, with mean values of about $10\,\mathrm{g\,m^{-2}}$ and widths of about $9\,\mathrm{g\,m^{-2}}$. Some differences between NARVAL1 and NARVAL2 exist which are even stronger for the measured distributions. During NARVAL1, the measured distribution is skewed towards higher values. This might be caused by cloud patches that were only present in a lateral part of the MWR footprint such that the scene was falsely identified as clear sky by the lidar, which only slices though the center of the MWR footprint. As this effect is not visible for NARVAL2 measurements, it might be that clouds were generally smaller and more frequent during NARVAL1 (see Sect. 4.6).

For both campaigns the similar widths and standard deviations of the retrieved LWP indicate a good agreement between simulations and measurements for clear sky (Fig. 4.8). The apparent second mode at $20\,\mathrm{g\,m^{-2}}$ in the observed clear-sky LWP distribution during NARVAL2 is caused by different mean deviations during different flights, probably influenced by the calibration. Overall, the narrow Gaussian widths ($11.4\,\mathrm{g\,m^{-2}}$ and $8.3\,\mathrm{g\,m^{-2}}$ for NARVAL1 and NARVAL2) of the retrieved clear-sky LWP distributions demonstrate the good performance of HAMP compared to evaluation studies by Liu et al. (2001) ($28\,\mathrm{g\,m^{-2}}$, airborne) and Greenwald et al. (2018) ($\sim 30\,\mathrm{g\,m^{-2}}$, satellite). The better HAMP performance is likely due to its smaller footprint, additional frequency channels, and more recent technology. The sensor synergy of using the lidar cloud mask for clear-sky bias correction (Sect. 4.2.4) reduces the bias in clear-sky conditions to values barely above zero as small cloud patches can still be in the outer area of the MWR footprint which is not transected by the lidar beam. The bias correction further narrows the clear-sky LWP distributions. Note that a good agreement (small bias) is expected as the lidar cloud mask is also used to define clear sky. The deviations of the observed clear-sky LWP distributions from delta distributions are due to the moving window in the bias correction.

In summary, the ICON–PAMTRA database allows the expected uncertainty of the LWP retrieval to be estimated. This reveals a lower retrieval limit of about $12\,\mathrm{g\,m^{-2}}$ and an RMSD below $20\,\mathrm{g\,m^{-2}}$ for LWP below $100\,\mathrm{g\,m^{-2}}$ and below $20\,\%$ above $100\,\mathrm{g\,m^{-2}}$ for the NN retrieval. A narrow clear-sky distribution of HAMP measurements (SD $\sim 10\,\mathrm{g\,m^{-2}}$) is found that is in good agreement with the theoretical assessment, but a small bias on the order of $12\,\mathrm{g\,m^{-2}}$ remains which is eliminated by the clear-sky correction. The synergy of MWR and lidar removes the bias and reduces the clear-sky LWP noise to 5 to $7\,\mathrm{g\,m^{-2}}$.
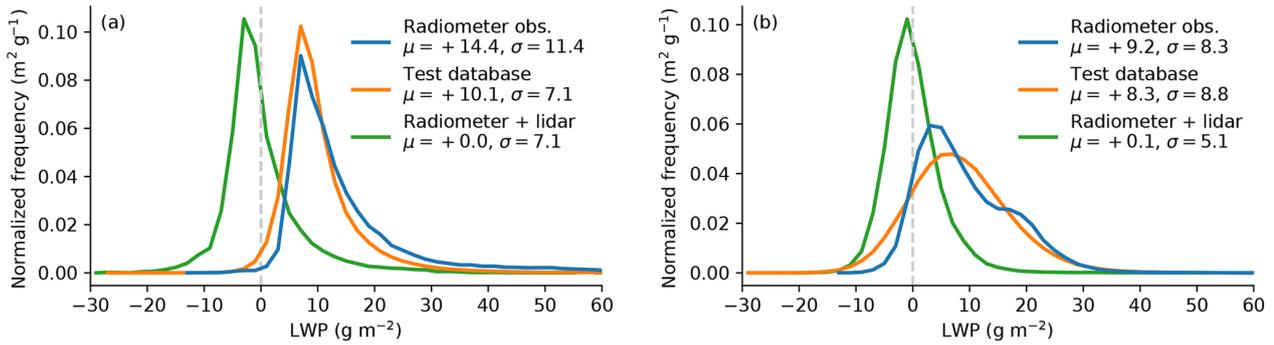
Figure 4.8.: LWP distributions retrieved during clear-sky scenes only. Shown are the LWP retrieved with the neural network based on radiometer observations (blue lines) during (a) NARVAL1 and (b) NARVAL2, the retrieved LWP from the test database (orange lines) for profiles with $LWP < 1\,g\,m^{-2}$, and the observed LWP after applying the lidar clear-sky correction (green lines). Mean (μ) and standard deviation (σ) are given for each distribution in grams per square meter ($g\,m^{-2}$).

4.5. Assessment of Rain

To investigate the formation of rain with HAMP measurements, this section extends the applicability of the LWP retrieval to drizzle and light precipitation by combining cloud radar with MWR. As described in Sect. 4.2.4, RWP is retrieved as the fraction $f = \frac{RWP}{LWP}$ by a NN using eight BTs and integrated radar reflectivity as input. Two physical effects are considered in the retrieval: hydrometeor scattering, which becomes more important with increasing droplet size and microwave frequency, and radar backscatter being sensitive to D^6, where D is the droplet diameter. The first effect is considered by including the $90\,GHz$ channel as proposed by Cadeddu et al. (2017). For the latter effect, the vertically integrated (linear) radar reflectivity is used as retrieval input in addition to the MWR channels also used in the LWP retrieval. This integrated reflectivity as a columnar quantity is more comparable to a BT and less noisy than the reflectivity of a single range gate and is thus used as retrieval input.

The Gilbert skill score (GSS) (Hogan et al., 2010), also known as equitable threat score (ETS), is used to rate how well retrieval "yes" events correspond to true yes events while accounting for hits due to chance. Yes events mark RWP above a given threshold. The GSS is defined as

$$GSS = \frac{hits - hits\_by\_chance}{hits + misses + false\_alarms - hits\_by\_chance} \tag{4.3}$$

using the common entries of the contingency table and the hits due to chance

$$\text{hits\_by\_chance} = \frac{(\text{hits} + \text{misses}) \times (\text{hits} + \text{false\_alarms})}{\text{hits} + \text{misses} + \text{false\_alarms} + \text{correct\_negatives}}. \quad (4.4)$$

GSS ranges from $-\frac{1}{3}$ to 1, with 1 being the perfect score.

The retrieval of RWP is evaluated for different RWP thresholds (Fig. 4.9). The GSS shows good performance, being higher than 0.75 for RWP thresholds from $10\,\mathrm{g\,m^{-2}}$ to about $50\,\mathrm{g\,m^{-2}}$ and higher than 0.5 for RWP up to $250\,\mathrm{g\,m^{-2}}$. Note that 762, 295, and 62 of the test cases have RWP greater than 10, 50, and $250\,\mathrm{g\,m^{-2}}$, respectively, and only few samples with higher RWP are available. The hit rate is higher than $80\,\%$ for thresholds between 10 and $250\,\mathrm{g\,m^{-2}}$, but the $250\,\mathrm{g\,m^{-2}}$ threshold also generates $37\,\%$ false alarms. Especially the high GSS for low RWP thresholds makes the f retrieval a useful tool combining cloud radar and MWR for detecting measurements, that contain warm precipitation.
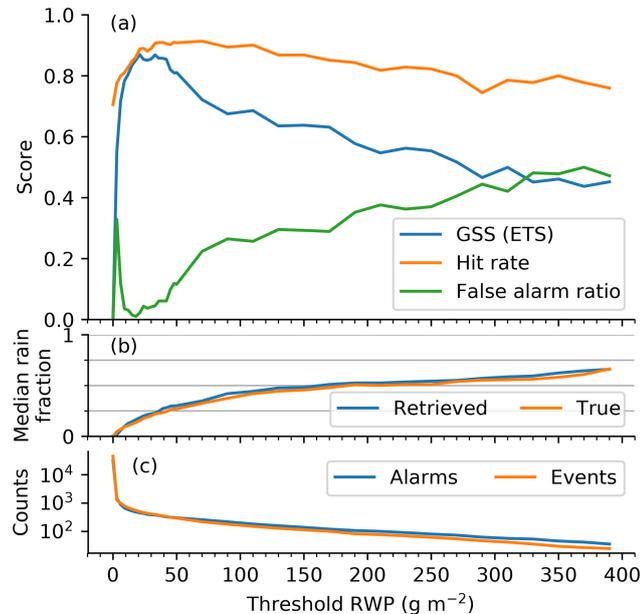
Figure 4.9.: Scores for detecting an atmospheric profile with RWP higher than the respective threshold. (a) Gilbert skill score (GSS), hit rate, and false alarm ratio. (b) Median fraction of rainwater as a function of RWP threshold. (c) Number of alarms and events for retrieved and true RWP above the threshold, respectively.

A case study of two showering shallow cumuli is shown in Fig. 4.10 to illustrate the capabilities of retrieving CLWP and RWP separately. The figure shows how HAMP is able to resolve spatial features of showering cells, which were observed with a cross section of several HAMP footprints. The precipitating core of both cells had maximum RWPs of probably more than $200\,\mathrm{g\,m^{-2}}$. The stronger relative gradients

of RWP compared to CLWP indicate the narrowness of the precipitating core. Note how the higher horizontally resolved information by radar (MWR footprints 3.3° to 5.0° vs. radar footprint 0.6°) contributes relatively stronger to the RWP retrieval than to the CLWP retrieval. The RWP retrieval consistently indicates no rain except for the time when the radar signal touches the surface or when there is a clearly visible fallstreak (17:42:30). The two showering clouds reveal maximum total LWP of more than $700\,\mathrm{g\,m^{-2}}$ and $1000\,\mathrm{g\,m^{-2}}$. The second shower core likely contains more water than indicated, as the retrieval sets the clipping flag. This case study also demonstrates the higher sensitivity of the lidar and the (C)LWP retrieval, which shows cloud signals between 17:38:30 and 17:39:10 of clouds which are too thin to be detected by the radar.
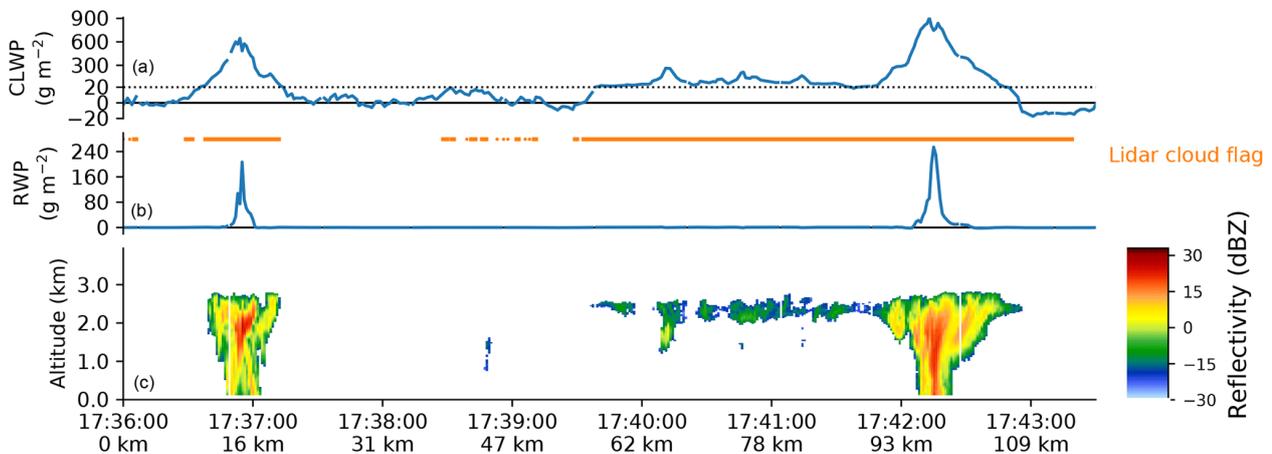
Figure 4.10.: Example scene of rain retrieval from NARVAL1 research flight 8 on December 20, 2013. Retrieved CLWP (a), retrieved RWP (b), and radar reflectivity profile (c). Note that the scale in (a) is piecewise linear, with a scale change at $20\,\mathrm{g\,m^{-2}}$. The IWV varies around $31.5 \pm 1.5\,\mathrm{kg\,m^{-2}}$ in this scene.

4.6. Comparison of Dry and Wet Season

The synergy of lidar, radar, and MWR is necessary to understand the difference of clouds in the dry and wet season as all instruments have their specific limitations. The lidar cloud mask indicates the more frequent occurrence of clear sky during the wet season (70.0 %) compared to the dry season (53.3 %, Table 4.3), even though the environment is characterized by less humid air in the dry season (Fig. 4.11a, b). The IWV distribution is clearly confined to moderate values with a mean of $28\,\mathrm{kg\,m^{-2}}$ in the dry season, which is mainly due to a rather dry middle troposphere seen in the

lidar water vapor profiles. During the wet season, IWV values up to $60\,\mathrm{kg\,m^{-2}}$ were observed, distributed into two modes around 35 and $52\,\mathrm{kg\,m^{-2}}$. These two modes are clearly distinct from the single mode observed in the dry season and reveal the expected humid characteristic of the wet season. The most-humid air during NARVAL1 was sampled in a deep convective system on the southernmost leg of research flight 2 on December 11, 2013. This was the NARVAL1 flight during which HALO was closest to the intertropical convergence zone (ITCZ). The NARVAL2 IWV distribution seems to be driven by the vicinity of the flight track to deeper convective systems and the ITCZ as it can be analyzed from satellite images and thus also by the selection of flight patterns.

Interestingly, clouds show a higher mean LWP of about $63\,\mathrm{g\,m^{-2}}$ in the dry season compared to a mean LWP of $40\,\mathrm{g\,m^{-2}}$ in the wet season. Likewise, thicker clouds (LWP $> 50\,\mathrm{g\,m^{-2}}$) were more frequent in the dry season (Fig. 4.11c, d); i.e., $27.1\,\%$ of the time when a cloud was seen in the dry season, it contained LWP $> 50\,\mathrm{g\,m^{-2}}$, whereas only $18.6\,\%$ of the time in the wet season, clouds exceeded this value. The dry-season clouds tend to produce light precipitation more frequently than the wet-season clouds as indicated by the more frequent exceedance of RWP thresholds (Table 4.3). The cumulative distributions of RWP occurrences of all cloudy measurements with LWP $> 50\,\mathrm{g\,m^{-2}}$ are depicted in Fig. 4.11e and f for each flight in the two seasons, when radar measurements are available. The vast majority (NARVAL1: $91\,\%$; NARVAL2: $96\,\%$) of all these measurements show RWP $< 10\,\mathrm{g\,m^{-2}}$. Higher amounts of light rain seem to be more frequent in the dry season dataset, although the small number of heavy RWP observations inhibits a statistical sound statement as RWP $> 100\,\mathrm{g\,m^{-2}}$ was only observed for 162 and $49\,\mathrm{s}$ in the radar–radiometer datasets of the dry and wet season, respectively. These time spans exclude measurements flagged as clipping (LWP $> 1000\,\mathrm{g\,m^{-2}}$) or frozen precipitation (ice scattering). While warm precipitation seems to occur less often, clouds associated with frozen precipitation were more often observed in the wet season ($1.6\,\%$ of the time) than in the dry season ($0.5\,\%$). Therefore, the lower LWP of the wet season clouds might be due to a higher precipitation efficiency compared to the dry season.

The higher LWP in the dry season might partly be explained by the choice of flight patterns. However, an analysis of ground-based LWP measurements at the Barbados Cloud Observatory (Stevens et al., 2015) over the years 2013-2018 confirms the generally higher LWP values during December than August (not shown). Thus, also changes in the organization of clouds could cause the differences in cloud fraction and LWP. The fact that the medium LWP range from 100 to $400\,\mathrm{g\,m^{-2}}$ is less frequent in the wet season could be due to the higher degree of organization causing more
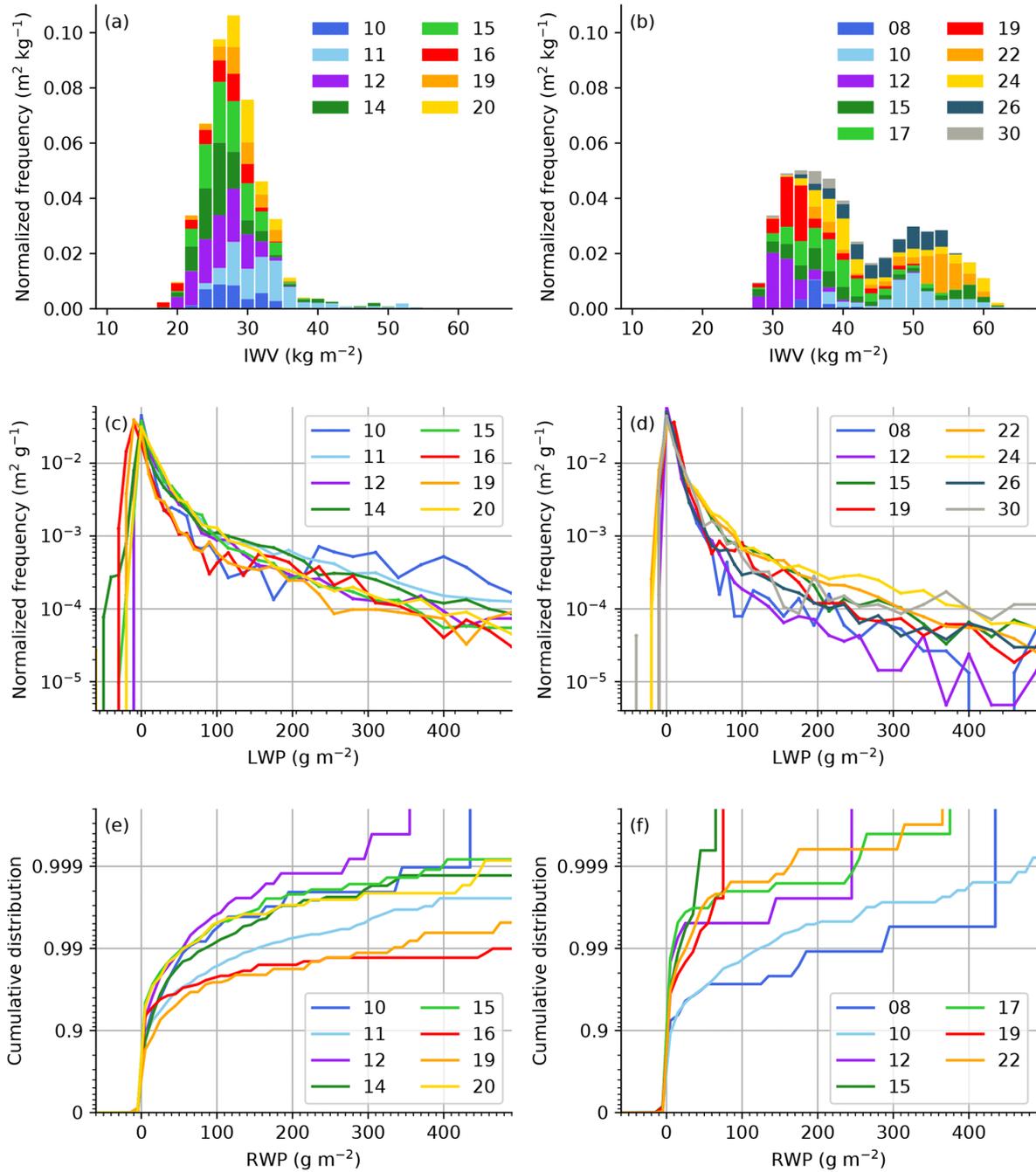
Figure 4.11.: Frequency distribution functions of IWV during the dry season (NARVAL1, a) and wet season (NARVAL2, b), LWP during NARVAL1 (c) and NARVAL2 (d), and cumulative distribution functions of RWP during NARVAL1 (e) and NARVAL2 (f). Colors denote the day of the month of the respective study. Colors in (a) and (b) denote the contribution of each flight to the total distribution. The bin edges are represented as minor ticks in (c) to (f). LWP distributions only include measurements for which the lidar cloud flag reports a cloud within $\pm$ 2 seconds. RWP distribution is based on the non-clear-sky-corrected LWP dataset (see note [a] in Table 4.3), where LWP $> 50\,\mathrm{g\,m^{-2}}$.

Table 4.3.: Comparing NARVAL1 and NARVAL2 cloud properties observed south of 20° N and with HALO altitude above 6 km. Percentages of flight time with available corresponding datasets during each study.

Dataset	NARVAL1 December 2013 radiometer, lidar and radar	NARVAL2 August 2016 radiometer and lidar	radiometer and radar
Clear sky	53.31 %	69.95 %	–
$LWP > 20\,\mathrm{g\,m^{-2}}$	21.62 %	10.60 %	–
$LWP > 50\,\mathrm{g\,m^{-2}}$	12.63 %	5.26 %	–
$LWP > 500\,\mathrm{g\,m^{-2}}$	1.18 %	0.33 %	–
$\overline{LWP}$ of clouds	$63\,\mathrm{g\,m^{-2}}$	$40\,\mathrm{g\,m^{-2}}$	–
RWP [a] $> 10\,\mathrm{g\,m^{-2}}$	1.85 %	–	0.30 %
RWP [a] $> 50\,\mathrm{g\,m^{-2}}$	0.43 %	–	0.07 %
Ice-flag	0.51 %	0.94 %	1.76 %
LWP clipping	0.97 %	0.45 %	0.53 %
Total hours	25:26:18	39:43:28	41:22:48

[a] Based on non-clear-sky-corrected LWP as radar and lidar cloud mask were only available during 5 of 10 flights during NARVAL2.

clear-sky areas and more intense clusters with higher amounts of precipitation. In that sense the latter would be missed by our flight patterns as we avoided strongly convective scenes with the formation of large ice particles.

4.7. Summary and Conclusions

Clouds play a critical role in the development of the future climate, and especially marine low-level clouds have been identified as source of uncertainty. An important cloud macrophysical quantity is LWP. Global observations are limited by satellite resolution or accuracy, and ground-based observations over the oceans are few. To fill this observational gap, the NARVAL studies were initiated to assess North Atlantic trade wind clouds using the HALO research aircraft. We use a multichannel microwave radiometer, a cloud radar, a lidar, and a dropsonde system deployed to HALO to provide insights into clouds on the kilometer scale. For NARVAL1 (December 2013) and NARVAL2 (August 2016) a unique retrieval training and test database was developed based on ICON simulations with 1.25 km grid spacing. The database contains more than 350 000 physically consistent profiles that characterize

the thermodynamic state of the atmosphere and the hydrometeor distributions during each of the two campaigns. Synthetic HAMP measurements in terms of BTs and radar reflectivity profiles in nadir view were simulated for each profile using PAM-TRA. The synthetic BT measurements show bivariate relations that are consistent with those observed and therefore show reliability that the database can be used to develop retrievals and assess LWP quality.

To estimate IWV, LWP, and RWP from HAMP measurements, artificial neural networks are trained with the retrieval database. BTs of seven K-band channels and the 90 GHz channel are used for IWV and LWP; vertically integrated radar reflectivity is used in addition for RWP.

Similar to LWP, an IWV retrieval is based on the spectral BT characteristics between the same water vapor absorption and window channels. A good retrieval of either IWV or LWP is a prerequisite for the other. The IWV comparison to dropsonde measurements and the continuous along-track comparison to the water vapor lidar WALES show good agreement with an RMSD smaller than $1.4\,\mathrm{kg\,m^{-2}}$ and no distinct error dependence of IWV itself. Overall, the IWV assessment shows the good practical performance of HAMP and the suitability of the ICON–PAMTRA database for developing microwave retrievals for NARVAL1 and NARVAL2.

LWP retrievals are theoretically assessed as a functions of retrieved LWP and true LWP. A slight advantage of the neural network compared to a linear regression retrieval is evident, especially at the limits of the LWP range (1 to $1000\,\mathrm{g\,m^{-2}}$). Both approaches show relative errors greater than $100\,\%$ for a retrieved LWP $< 12\,\mathrm{g\,m^{-2}}$, which can be regarded as detection limits. If more liquid water is contained in the column, the random error decreases to $20\,\%$ at LWP $\approx 100\,\mathrm{g\,m^{-2}}$ and $10\,\%$ at LWP $\approx 800\,\mathrm{g\,m^{-2}}$. Both retrievals show an offset error smaller than the random component for LWP $< 10\,\mathrm{g\,m^{-2}}$ with different signs depending on whether it is analyzed as a function of true or retrieved LWP. Because of the ambiguity of the error sign, we conclude that this bias can not be accounted for with the MWR retrieval alone, and we developed a synergistic clear-sky offset correction using the WALES lidar cloud mask. The cloud mask reduces the noise of clear-sky LWP to 7.1 and $5.0\,\mathrm{g\,m^{-2}}$ for NARVAL1 and NARVAL2, respectively.

To allow the onset of precipitation to be investigated, a neural network retrieval is trained to estimate the fraction between RWP and LWP from a combination of integrated radar reflectivity factors and BTs. Using the test database, a Gilbert skill score above 0.75 is found for RWP thresholds between $10\,\mathrm{g\,m^{-2}}$ and $50\,\mathrm{g\,m^{-2}}$, which shows good applicability for detection of rain or drizzle onset.

We used data from 36 flight hours in December 2013 (dry season, NARVAL1) and 64 flight hours in August 2016 (wet season, NARVAL2) to investigate differences between the seasons. The analysis shows that although clouds were more frequent and their LWP and RWP were higher during the flights in the dry season, more microwave scattering of ice was observed in the wet season indicating strong precipitation events. The difference between $\overline{\mathrm{LWP}}_{\text{dry season}} \approx 63\,\mathrm{g\,m^{-2}}$ and $\overline{\mathrm{LWP}}_{\text{wet season}} \approx 40\,\mathrm{g\,m^{-2}}$ is clearly larger than the LWP retrieval uncertainty. As expected, the IWV histograms reveal the dry season as being drier and more uniform and the wet season as more humid. However, the IWV distributions also reveal sampling biases due to flight track choices, especially for the wet season. Therefore, the airborne measurements need to be combined with long-term ground-based and spaceborne measurements to draw statistically sound conclusions. The fine-scale airborne microwave observations such as the measurements obtained with HAMP can be used to investigate the sub-satellite-footprint inhomogeneity of LWP and rain for a better error characterization of satellite measurements. Sound conclusions on the diurnal cycle can not be drawn from the data presented here, as the spatial variability of the clouds on the observed mesoscale was higher than an expected effect of the diurnal cycle.

The synergy of active and passive microwave observations could further benefit from using an optimal estimation approach including the full radar profile and all MWR channels to improve the partition of rain and cloud droplets and frozen particles (e.g Battaglia et al., 2016). With respect to trade wind cumuli, the products of the present study in combination with cloud boundary estimations from the radar and backscatter lidar will be used to evaluate the condensate loads of different shallow trade wind cumulus types in large eddy simulations. For example, radar and lidar both detect shallow convection or shallow outflow anvils as depicted in Fig. 4.10. But in addition, the lidar also allows boundary-layer-driven clouds, which have tops around 1 km and are below the radar sensitivity, to be detected.

An extension of the NARVAL observations is planned by the EUREC[4]A field study in early 2020 (elucidating the couplings between clouds, convection, and circulation; Bony et al., 2017), which among other objectives will investigate convective aggregation. The algorithms presented here will be applied, and together with additional measurements a better understanding of the governing processes that cause differences between the dry and wet season will be analyzed. For that, the campaign will provide additional observations of large-scale dynamics, horizontally resolved remote sensing, and in situ observations by additional aircraft in the cumulus layer. Also, more locally targeted flights, distributed over the daytime, are planned to study the

diurnal cycle. Together with ship, drone, and buoy measurements, a unique dataset for a better understanding of precipitation onset will be generated.

4.A. Remarks

Chapter 5.

Application: Confronting Models with Observations

The manuscript of the following Chapter has been submitted for publication in Geoscientific Model Development on 13 January, 2020 and is available as discussion paper since 6 March, 2020 (Jacob et al., 2020). This Chapter investigates the research question 3 posted in the introduction (Sect. 1.4):

3. How can HAMP be used to assess cloud-resolving model simulations and how good are the considered models in representing the convective dry-season trade-wind clouds? Two different cloud-resolving simulations are confronted with HALO observations to tell their differences with respect to clouds and precipitation.

A method of combined forward simulations and retrievals of active and passive HAMP observations in combination with WALES lidar backscatter is developed. The study demonstrates how the observations can spread light on the understanding of trade-wind convection established in the model formulations. The presented method is not limited to the two simulations assessed in the following, but can be applied to cloud resolving simulations on hectometer to kilometer grid scales in general.

The author of this thesis conceptualized the study together with advice from Pavlos Kollias. Furthermore, the author performed the analysis, prepared all plots, and wrote the manuscript with support and input from all co-authors. Felix Ament and Susanne Crewell designed the observational experiment setup and supported the interpretation of the measurements. Vera Schemann supported the interpretation of the model data and the implementation of the microphysical scheme in the models. The latter can cause differences in the cloud-life-cycle and will be identified as a further distinction between the models besides the obvious grid-size difference.

Modifications: References to Jacob et al. (2019) are replaced by respective references to Chapter 4 of this thesis and a comment on the dependence between the used retrieval and model datasets is added as Appendix 5.B.

Multi-layer Cloud Conditions in Trade Wind Shallow Cumulus – Confronting Models with Airborne Observations

Marek Jacob[1], Pavlos Kollias[1,2], Felix Ament[3], Vera Schemann[1], Susanne Crewell[1]

[1]Institute for Geophysics and Meteorology, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany

[2]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794-5000, USA

[3]Meteorological Institute, Universität Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany

Correspondence: Marek Jacob (marek.jacob@uni-koeln.de)

Abstract. Airborne remote sensing observations over the tropical Atlantic Ocean upstream of Barbados are used to characterize trade wind shallow cumulus clouds and to benchmark two cloud-resolving ICON (ICOsahedral Nonhydrostatic) model simulations at kilo- and hectometer scales. The clouds were observed by an airborne nadir pointing backscatter lidar, a cloud radar, and a microwave radiometer in the tropical dry winter season during daytime. For the model benchmark, forward operators convert the model data into the observational space for considering instrument specific cloud detection thresholds. The forward simulations reveal the different detection limits of the lidar and radar observations, i.e., most clouds with cloud liquid water content greater than $10^{-7}\,\mathrm{kg\,kg^{-1}}$ are detectable by the lidar, whereas the radar is primarily sensitive to the "rain"-category hydrometeors in the models and can detect even low amounts of rain.

The observations reveal two prominent modes of cumulus cloud top heights separating the clouds into two layers. The lower mode relates to boundary layer convection with tops closely above the lifted condensation level, which is at about 700 m above sea level. The upper mode is driven by shallow moist convection, also contains shallow outflow anvils, and is closely related to the trade inversion at about 2.3 km above sea level. The two cumulus modes are reflected differently by the lidar and the radar observations and under different liquid water path (LWP) conditions. The storm-resolving model (SRM) at kilometer scale reproduces the cloud modes barely and shows the most cloud tops slightly above the observed lower mode. The large-eddy model (LEM) at hectometer scale reproduces better the observed cloudiness distribution with a clear bimodal separation. We hypothesize that slight differences in the autoconversion parametrizations could have caused the different cloud development in the models. Neither model seems to account for in-cloud drizzle particles

that do not precipitate down to the surface but generate a stronger radar signal even in scenes with low LWP. Our findings suggest that even if the SRM is a step forward for better cloud representation in climate research, the LEM can better reproduce the observed shallow cumulus convection and should therefore in principle represent cloud radiative effects and water cycle better.

5.1. Introduction

The representation of low-level oceanic clouds contributes largely to differences between climate models in terms of equilibrium climate sensitivity (Schneider et al., 2017). Global atmospheric models with kilometer-scale resolution are considered as the way forward in forecasting future climate scenarios (Bony and Dufresne, 2005; Satoh et al., 2019). The increased model resolution and better matching scales with measurements allow for a more direct observational assessment by comparing the present day representation in the models with atmospheric measurements and thus anchoring models to reality. Recently, Stevens et al. (2020a) demonstrated the general advantage of high resolution simulations compared to typical climate models in terms of cloud representation using different versions of the ICOsahedral Non-hydrostatic model (ICON). The progress in such novel large-area high-resolution models and new capabilities of synergetic airborne measurements in the trades motivate the following guiding questions of this study. How do two cloud-resolving versions of the ICON model represent shallow cumuli in comparison to observations? What is an appropriate approach to assess the model clouds? How does the liquid water path (LWP) help to interpret differences between observed and simulated cloud structures?

Increased model resolution facilitates the model-observation comparisons. However, there are several other factors to be considered (Lamer et al., 2018). On the one hand, particle size distributions (PSDs) in models are typically represented by bulk and spectral microphysical schemes, or Lagrangian superparticles (e.g., Grabowski et al., 2018). Bulk microphysics schemes predict changes in condensate using one to three moments. These are usually the lower moments like particle number concentration and mixing ratio (Khain et al., 2015). On the other hand, radars and lidar, like those used in this study, observe different moments of the PSD. A backscatter lidar, for example, is primarily sensitive to the second moment, while a radar is sensitive to the sixth moment.

An objective definition of a cloud is required when comparing cloudiness in models with observations. If one asks different instrument operators to provide "cloud

fraction", one can get different answers, e.g., 19 to 46 %, for the very same scene as demonstrated by Stevens et al. (2019). This range is caused by different sensitivities due to different measurement principles and sampling methods by the remote sensing instruments involved. To find a common definition, it is favorable to compare clouds in models and observations in terms of the same quantities. Here, forward simulators can be used to simulate measurements as they would be recorded by a radar or lidar, based on the atmospheric state and assumptions in the model (Lamer et al., 2018).

The observations used in this study were recorded with the research aircraft HALO (High Altitude and LOng range; Krautstrunk and Giez, 2012) which was equipped as a flying remote sensing cloud observatory during the NARVAL-South experiment (Next generation Advanced Remote sensing for VALidation; Klepp et al., 2014) in December 2013. A reason to initiate the NARVAL expeditions was that satellites cannot provide sufficient resolution for multiple cloud variables. For example, the spaceborne Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) has frequently been used to investigate marine low clouds. Leahy et al. (2012) observed two modes of low clouds in the tropical Pacific trade wind, and reveal that CALIOP misses small clouds ($< 1\,\mathrm{km}$) and combines adjacent but separated clouds due to the CALIOP sampling rate. Genkova et al. (2007) compared trade wind cumuli cloud top heights from passive optical spaceborne instruments. They also observed bimodal distribution with data from three different satellites but found vertical biases of 250 to 500 m due to different retrieval approaches and spatial resolutions.

Since active instruments are advantageous for observing cloud heights, the HALO instrumentation included an aerosol backscatter lidar as part of the WALES (WAter vapor Lidar Experiment in Space; Wirth et al., 2009) airborne demonstrator, and a cloud radar. The radar is part of the HAMP (HALO microwave package; Mech et al., 2014) together with a microwave radiometer. The latter provides the vertically integrated LWP (Sect. 4.2.4), which helps approaching the liquid water content which is a key quantity to describe clouds in models like the ICON. The direct observation of the liquid water content profile is difficult (Crewell et al., 2009), but the LWP can be used to estimate the water content when combined with estimates of cloud vertical extend by lidar and radar either in a simple average approach or more sophisticated as a profile (Frisch et al., 1998; Küchler et al., 2018). In addition, dropsondes were released regularly during the flights to probe the temperature and humidity profile. Compared to ground-based observations, the airborne remote sensing instruments, especially the microwave radiometer, have the advantage of not being harmed by precipitation or sea spray deposition on the instrument (Rose et al., 2005).

The observations are used to confront the simulations of winter season trade wind cumuli in the tropical west Atlantic Ocean. Such clouds are regularly subject in idealized large-eddy simulation (LES) studies (e.g., Siebesma et al., 2003; van Zanten et al., 2011) due to their high relevance for the climate. As it is difficult for small domain LES models to generate realistic mesoscale cloud organization (Jeevanjee and Romps, 2013), we use simulations by Klocke et al. (2017) that were run on large domains (> 1500 x 900 km) with kilo- and hectometer horizontal grid spacings and were forced by numerical weather prediction output. Simulations with 1.25 km grid spacing were produced using the storm-resolving model (SRM) version of ICON, while simulations with 3 hectometer grid spacing were produced using ICON large-eddy model (LEM).

To assess vertically resolved cloudiness and shallow convection, we compare the vertical cloud boundaries. As the backscatter lidar is quickly attenuated completely by the presence of hydrometeors in a cloud, lidar measurements and their forward simulations are considered for a cloud top height estimate only. The radar, however, can penetrate through the cloud and precipitation layers and thus provides estimates of cloud or precipitation base height in addition to cloud top height. As shallow cumulus convection is not expected to trigger at the same time and place in a model and reality, a statistical approach is adopted here, in which the airborne observations are compared to their model counterpart for different LWP regimes. The analysis in LWP space is similar to the studies in moisture space that first have been published by Schulz and Stevens (2018) for ground-based observations and by Naumann and Kiemle (2019) for airborne observations. In the LWP space it is possible to study microphysical cloud processes like the transition from non-precipitating to precipitating clouds.

This paper is structured as follows: The observations and their sensitivities in Sect. 5.2 are followed by a brief description of the model data in Sect. 5.3. Then, the forward simulations are presented in Sect. 5.4 taking into account the instrument characteristics and specifications of model outputs. Finally, the model data of ICON SRM and LEM are confronted with the airborne observations in Sect. 5.5 including the analysis in LWP space. A summary and conclusions are given in Sect. 5.6.

5.2. Observations

The airborne measurements were taken during the NARVAL-South field experiment in the tropical Atlantic east of Barbados. The NARVAL remote sensing package (Stevens et al., 2019) recorded data during 8 research flights in the tropical domain

south of 20° N from 10 to 20 December 2013. The flight tracks are depicted in Fig. 5.1. A total of 22 000 km of HALO along track observations with about 91 thousand profiles were sampled at a frequency of 1 Hz from altitudes between 13 and 14.5 km. Further details of the experiment and flight planing are provided by Klepp et al. (2014) and Konow et al. (2019). In this study we use the backscatter lidar cloud top height time series, the radar reflectivity factor Z, liquid water path (LWP) retrieved from microwave radiometer, and the lifted condensation level (LCL) estimated from dropsondes. The remote sensing lidar, radar, and microwave radiometer were installed in a near-nadir pointing direction under the fuselage of the aircraft.

This section briefly describes the measurement principles of the radar and lidar and the respectively used thresholds for cloud detection. The LWP retrieval form the microwave radiometer has a high accuracy, which is better than $20\,\mathrm{g\,m^{-2}}$ for LWP $< 100\,\mathrm{g\,m^{-2}}$ and better than 20 and 10 % for LWP greater than $100\,\mathrm{g\,m^{-2}}$ and $500\,\mathrm{g\,m^{-2}}$, respectively, as described in Chapter 4. The LWP is defined as the integral of all liquid in the column comprising cloud liquid and rain water. The LCL is derived from the dropsonde temperature and relative humidity (RH) measurements closest to the surface using the code by Romps (2017). The LCL measurement uncertainty is mostly affected by the RH measurement, such that an overestimation on the order of the calibration repeatability of 2 % RH (Vaisala, 2017) would result in an about 60 m lower LCL. The LCL from dropsonde releases is temporally interpolated to generate a continuous time series along the flight track.
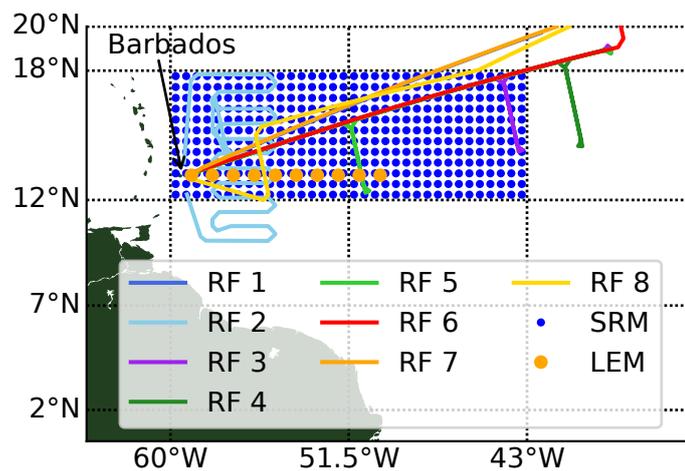
Figure 5.1.: Map showing research flight (RF) tracks and the model columns, which are used in this study. The storm-resolving model (SRM, blue, original model grid spacing: 1.25 km) is thinned to a $0.5° \times 0.5°$ grid. From the large-eddy model (LEM, orange, original model grid spacing: 300 m), ten meteogram outputs are used.

5.2.1. Radar Sensitivity

The radar reflectivity factor – short "reflectivity" – Z is measured by the HAMP radar at 35.5 GHz. In case of small spherical liquid droplets, the radar is approximately proportional to the sixth moment of the PSD at a given range. This means that larger raindrops show a higher reflectivity than smaller cloud droplets given the same mass mixing ratio. The HAMP radar is calibrated following Ewald et al. (2018) and was operated at a vertical resolution of about 30 m with 1 Hz sampling. This sampling frequency corresponds to a surface footprint of about $136\,\mathrm{m} \times 376\,\mathrm{m}$ at a cruising speed of about $240\,\mathrm{m\,s^{-1}}$.

The instruments minimal detectable signal (MDS) in dBZ decreases with range r and is estimated by Ewald et al. (2018) as

$$\mathrm{MDS}(r) = -39.8 + 20\log_{10}\left(\frac{r}{5\,\mathrm{km}}\right). \tag{5.1}$$

According to this equation, the MDS in the shallow cumulus layer is about $-32\,\mathrm{dBZ}$ when flying at 13 km. However, this does not include sensitivity reduction due to Doppler broadening caused by the aircraft motion (Mech et al., 2014). To estimate the practical sensitivity limit, HAMP radar statistics are compared to ground-based measurements. The ground-based measurements were taken at the Barbados cloud observatory (BCO) at the upstream eastern coast of Barbados at Ragged Point (Stevens et al., 2015). The BCO radar operates at the same Ka-band frequency as the airborne radar, but has a better sensitivity due to a larger antenna and longer integration time (Lamer et al., 2015). Therefore, the lower MDS of the BCO radar offers the opportunity to assess the practical sensitivity limit of the HAMP radar.

A comparison can only be made on a statistical basis as the BCO and HAMP radars do not sample the same volume. To avoid statistical effects of the diurnal cycle identified by Vial et al. (2019), BCO data are only considered roughly during the time when HALO was flying, i.e., between 12:00 and 21:00 UTC (8:00 and 17:00 local time) on the 8 flight days.

The higher BCO sensitivity compared to HAMP is notable in the height-resolved reflectivity histograms in Fig. 5.2. The BCO radar frequently measures reflectivity signals down to $-70\,\mathrm{dBZ}$ at around 500 m with a clear maximum below 1 km for Z up to $-20\,\mathrm{dBZ}$. Klingebiel et al. (2019) identify such weak signals at BCO below $-50\,\mathrm{dBZ}$ as originating from sea salt aerosols and only signals above $-50\,\mathrm{dBZ}$ are attributed to clouds. Clouds with reflectivity between the HAMP MDS ($-32\,\mathrm{dBZ}$) and $-20\,\mathrm{dBZ}$ and within 4 km above sea level are observed in 8.5 % of the time at BCO but only rarely ($< 1.2\,\%$) by HAMP. Only clouds with a reflectivity higher than

about -20 dBZ are similarly or more often observed by HAMP than at BCO. Thus, we use -20 dBZ as the practical cloud detection threshold of HAMP and use this value in the further analysis to define "radar-detectable clouds" in the observations and forward simulations.
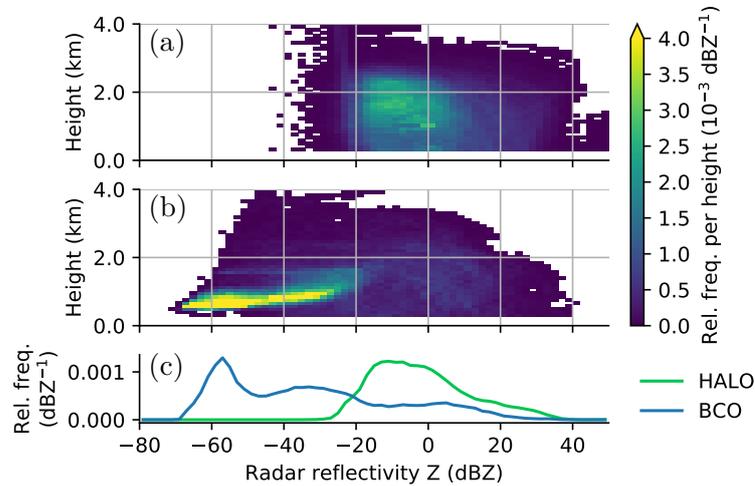
Figure 5.2.: Height-resolved radar reflectivity distribution of shallow cumulus from (a) HAMP radar and (b) BCO radar during flight days of NARVAL-South. Marginal distributions (c) show the probability density of reflectivity from HAMP and BCO below 4 km. BCO data are limited to hours between 12:00 and 21:00 UTC (8:00 and 17:00 local time) on every flight day to match aircraft operation time. The probability density function of each height is normalized to the maximal possible number of data points.

5.2.2. LIDAR

The lidar system WALES supplements the HAMP radar with optical active remote sensing on HALO. WALES comprises a water vapor differential absorption lidar system (DIAL) at different wavelengths and a high spectral resolution lidar (HSRL) which measures molecular and aerosol backscatter at 532 and 1064 nm. The scattering of an emitted laser pulse on a liquid hydrometeor mostly follows the principles of geometrical optics as the wavelength is much smaller than the particle. Therefore, the back-scattered energy is in first order approximation proportional to the hydrometeor diameter and thus to the second moment of the PSD (O'Connor et al., 2005). This means, that a backscatter lidar is more sensitive to the number of small droplets compared to a radar. Besides hydrometeors, also other aerosol particles like dust scatter the lidar pulse back. However, the aerosol signal is much smaller

than the hydrometeor signal. Therefore, we follow Gutleben et al. (2019) and use a threshold of backscatter ratio (BSR) > 20 in the 532 nm channel to differentiate cloudy scenes from clear-sky or dusty scenes. As hydrometeors attenuate the lidar signal strongly, the WALES lidar is used only to detect cloud top height using that threshold. The lidar top height is measured every second with a vertical accuracy of 15 m and the lidar footprint width at the surface is at about 22 m.

5.3. ICON-NARVAL Model Data

Two different versions of the ICON model were run to supplement the NARVAL experiment. The runs of the so called storm-resolving model (SRM) and the large-eddy model (LEM) are described by Klocke et al. (2017) and Vial et al. (2019). The most important aspects relevant for this study of the SRM and LEM are summarized in this section.

5.3.1. ICON SRM

The SRM (Zängl et al., 2015) was run at 1.25 km horizontal grid spacing with a stretched vertical grid of 75 levels up to 30 km which has 12 and 22 levels below 800 m and 3 km, respectively. The domain spans the western tropical North Atlantic from 4° S to 18° N and from 64° W to 42° W. The SRM is one-way nested into a coarser 2.5 km SRM which is initialized and nudged with lateral boundary data from the European Centre for Medium-Range Weather Forecasts (ECMWF). The SRM uses physical packages that are similar to those used in operational numerical weather prediction codes, but does not use a convection parametrization. The cloud and precipitation microphysics are represented by a one-moment microphysics scheme (Baldauf et al., 2011) that predicts the specific water contents of five different hydrometeor classes including liquid cloud water (q_c) and rain (q_r). 17 modeled days from 10 to 28 December 2013 are used and cover the whole NARVAL experiment.

The model output is archived hourly. This study only uses model output between 12:00 and 21:00 UTC to avoid influence of a diurnal cycle. This is analogous to the ground-based data described in Sect. 5.2.1. The data are spatially subsampled on a coarser 0.5° × 0.5° grid to reduce the computational effort while still conserving the variety of atmospheric profiles. A compromise of domain overlap between all available model data and observations is achieved by limiting the SRM data to the area of 12 to 18° N and 60 to 43° W as marked in Fig. 5.1. The total number of analyzed SRM columns in this study is 97 920.

5.3.2. ICON LEM

The LEM (Dipankar et al., 2015; Heinze et al., 2017) with 300 m grid spacing was run in a multi-step nested setup forced with the SRM. This means, that the LEM also has a realistic, non-idealized setup. The LEM vertical grid also reaches up to 30 km but has 150 levels with 14 and 37 of them below 800 m and 3 km, respectively. The LEM physics package differs from the SRM configuration. The most important difference for this study is that the microphysics are represented by the two-moment scheme of Seifert and Beheng (2001). This scheme predicts the hydrometeor number concentrations in addition to the specific water contents and thus provides N_c and N_r for liquid cloud water and rain, respectively.

In contrast to the SRM, the LEM was only run for the six days of research flights 2 to 6 and 8. However, the full hydrometeor state including rainwater and the number concentrations were only archived for four of the runs in the form of so called "meteogram output". This means that hydrometeor profiles are available with high temporal resolution (every 36 s) but only at 12 model columns. Such meteogram output was saved for the days of research flights 4, 5, 6, and 8. The ten model columns east of Barbados are used for this study and are also marked in Fig. 5.1. The LEM data are also limited to the time between 12:00 and 21:00 UTC. The total number of analyzed LEM columns in this study is 37 030.

5.4. Forward Simulations

Forward simulators, also called forward operators, can simulate how the remote sensing instruments presented in Sect. 5.2 would perceive a scene provided by an atmospheric model. A forward simulator requires input like model variables and the knowledge about the microphysical assumptions employed in the atmospheric model. The basic variables are temperature, pressure, layer height, and humidity for each model level in a column for a 1D vertical forward simulation. The variables describing the hydrometeors depend on the microphysical scheme. Typically, these include mass mixing ratios (e.g., q_c or q_r) of different hydrometeor classes. The forward simulator has to be configured such, that the PSD used to simulate hydrometeor characteristics matches the PSD assumed in the atmospheric model as accurately as possible. This means that for models with advanced microphysical schemes, also the variables describing those aspects of the PSD are important input parameters for the forward simulation and need to be saved during the model run. In the case of the ICON LEM, the two-moment scheme by Seifert and Beheng (2001) uses the particle number concentrations as additional variables.

As this study focuses on the tropical shallow cumulus below freezing level, we confine the following description and analysis to precipitating and non-precipitating liquid hydrometeors, which are the raindrops and cloud droplets in the ICON microphysical schemes. Both ICON models which are used in the subsequent Sect. 5.5 assume PSDs with modified Gamma distributions. The number concentration N of spherical drops with diameter D can be described as

$$N(D) = N_0 \, D^\mu \exp(-\Lambda \, D^\gamma) \tag{5.2}$$

with the scale parameters N_0 and Λ and the shape parameters μ and γ. These parameters are either fixed or derived from the input variables as described in Table 5.1.

Table 5.1.: Configuration of modified Gamma distribution (Eq. 5.2) for liquid hydrometeors in ICON one and two-moment microphysical schemes.

scheme	hydrometeor	N_0	μ	Λ	γ	additional constrain
one moment (SRM)	cloud droplets	$f(q_\mathrm{c})$	8	$f(q_\mathrm{c})$	3	$N_\mathrm{c} = 2 \times 10^8 \, \mathrm{kg}^{-1}$
one moment (SRM)	raindrops	$8 \times 10^6 \, \mathrm{m}^{-4}$	0	$f(q_\mathrm{r})$	1	
two moments (LEM)	cloud droplets	$f(q_\mathrm{c}, N_\mathrm{c})$	8	$f(q_\mathrm{c}, N_\mathrm{c})$	3	
two moments (LEM)	raindrops	$f(q_\mathrm{r}, N_\mathrm{r})$	2	$f(q_\mathrm{r}, N_\mathrm{r})$	1	

The lidar BSR is forward simulated using the Cloud Resolving Model Radar Simulator (CR-SIM; Oue et al., 2019). The code has been slightly modified such that the configuration for the two-moment ICON microphysics can be used for one-moment microphysics following the relations in Table 5.1. Though CR-SIM can also simulate radar reflectivity, the Passive and Active Microwave TRAnsfer package (PAMTRA; Mech et al., 2020) is used to forward simulate the radar as it offers a higher degree of flexibility.

The lidar forward simulations are used to detect the hydrometeors layer top and not for quantitative retrievals or estimates. Furthermore, as the airborne lidar is not affected by liquid collection on the telescope during raining conditions, there is no need to account for such effects. Thus, we decided to simplify the forward simulation of the backscatter lidar and assume that the raindrops are optically thin and thus ignore their contributions. Therefore, the BSR is primarily a function of q_c as shown in Figs. 5.3 a and b. One could further imagine that a raining cloud is always topped by small droplets contributing to q_c and that the lidar pulse hence would be scattered back by those cloud droplets, which would very likely have a BSR > 20, anyway and be thus identified as cloud top. However, this is not always true as some grid cells in ICON LEM with enough rainwater to generate a radar signal $Z > -20 \, \mathrm{dBZ}$ were simulated above or horizontally attached to a precipitating cloud

(e.g., Fig. 5.4b, at 20:11). Therefore, it has to be noted, that the cloud amount reported below for the simulated lidar clouds slightly underestimates the cloudiness and cloud top height seen by a real backscatter lidar.
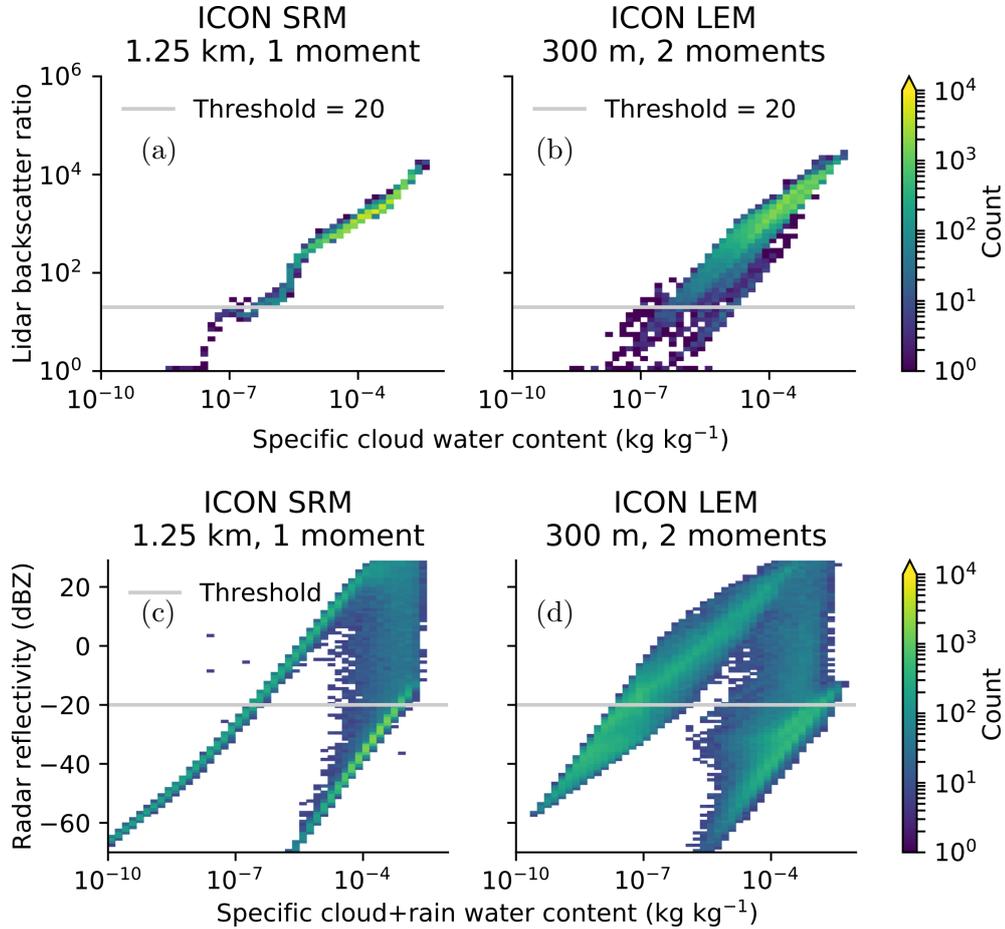
Figure 5.3.: Simulated lidar and radar signals as a function of hydrometeor contents. CR-SIM and PAMTRA simulate the observable lidar and radar signals from drop size distributions in the one-moment ICON SRM and two-moment ICON LEM microphysical models. Signals are simulated without attenuation as they would be sensed at cloud top.

The approximated proportionality of the radar reflectivity to D^6 makes Z especially sensitive to larger raindrops. Therefore, q_r (and N_r) has to be considered in addition to q_c (and N_c) when simulating the radar signal. The size difference between cloud droplets and raindrops produces a two-modal relation between the total liquid water concentration $q_t = q_c + q_r$ and Z as it can be deduced from Figs. 5.3 c and d. The mode along a line of low q_t corresponds to grid cells that predominantly feature rainwater. In this mode, even low amounts of liquid water in the rain category produce a reflectivity that can only by reached by cloud droplets with a three to four orders of magnitude higher cloud water content. Grid cells with such high q_c

and no q_r align in a second mode parallel to the rain mode. A mixture of cloud and rain water accordingly results in an intermediate Z which populates the space of $q_t >\approx 10^{-5}\,\mathrm{kg\,kg^{-1}}$ between the two main modes in Figs. 5.3 c and d. By setting the radar threshold to $-20\,\mathrm{dBZ}$, hardly any cloud-only grid cells in the lower right high-q_t mode can be detected by the simulated radar. This means, that in the forward-simulated dataset, all lidar-detectable hydrometeors are from the ICON cloud category while the radar-detectable hydrometeors have to contain at least a small amount of water from the ICON rain category.

The ICON LEM uses a two-moment scheme including N_c and N_r. Therefore, the forward simulation broadens the relation between the water content and the forward-simulated signals (compare Figs. 5.3 a to b and c to d). Especially the radar reflectivity of rain is amplified compared to the one-moment simulation, such that also some grid cells with lower q_r are above the radar detection threshold.

5.5. Model – Observation Comparison

Observations and forward simulations of the SRM and LEM runs are used to assess the vertical structures of the shallow clouds by focusing on the boundaries sensed by different instruments. In the following, shallow clouds are analyzed in terms cloud top heights estimated from lidar and radar measurements as well as the radar echo base height. All heights in the different scenes are set in relation to the theoretical cloud base of an adiabatic thermal-plume-driven boundary layer cloud by setting the height in relation to the lifted condensation level (LCL). First, a case study with example scenes from the observations and the LEM illustrates the approach. The case study is followed by the statistical analysis of the full datasets and the analysis stratified in the liquid water space to identify differences in microphysical processes.

5.5.1. Case Study

An example scene observed from HALO during research flight 5 is depicted in Fig. 5.4a. Here, several shallow clouds close to the LCL were observed first, followed by a precipitating cloud with stratiform shallow anvil outflow. The shallow clouds were only detected by the lidar, whereas the precipitating cloud was detected by both the lidar and the radar. However, the lidar detected cloud top heights about 50 to 100 m, i.e., up to three radar range gates, above the upper most recorded radar echo. Also, a larger part of the outflow layer was visible to the lidar. Thus,

we conclude, that the precipitating shallow cumulus has a thin layer of very small droplets on top which are only seen by the lidar due to its higher sensitivity (compare Fig. 5.3).
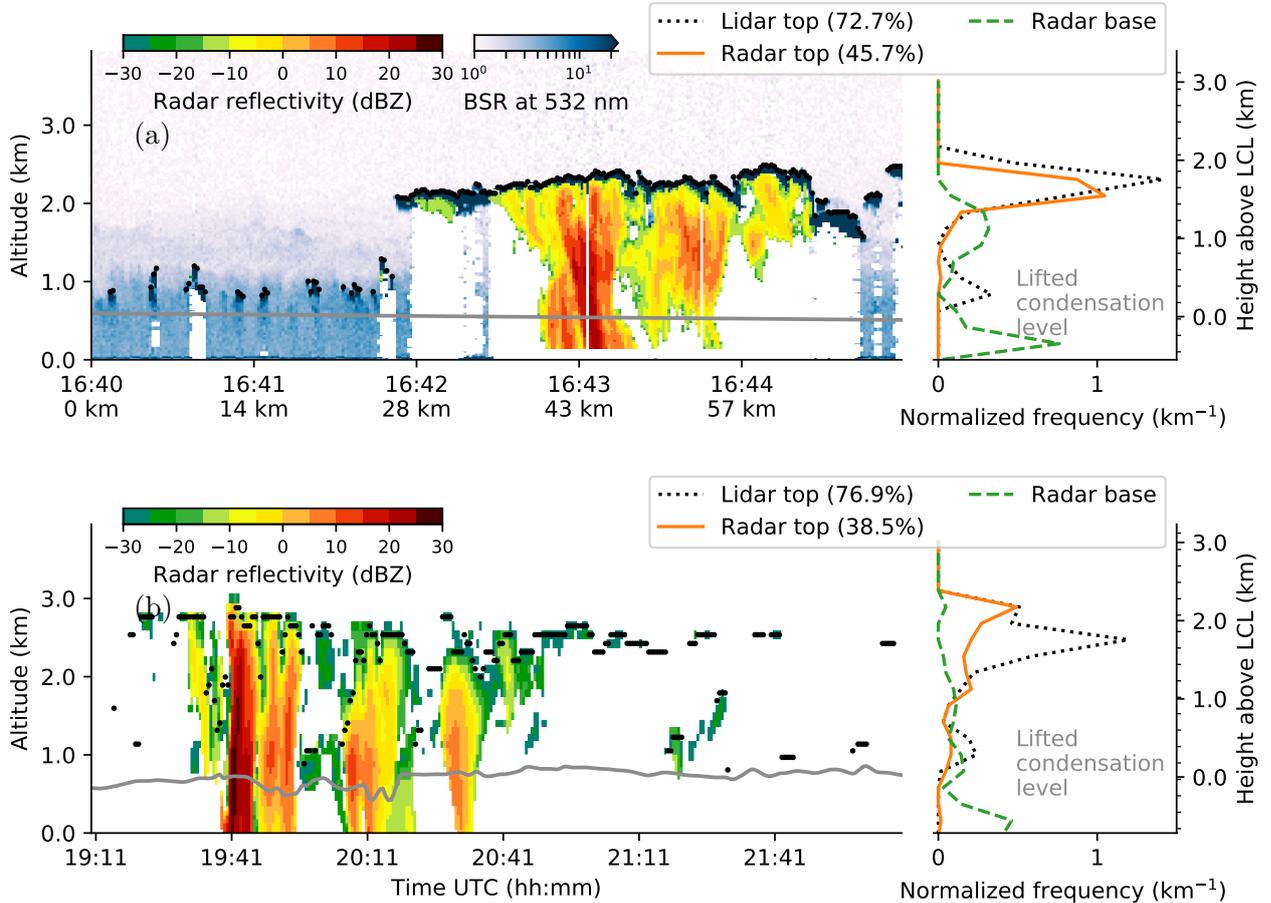
Figure 5.4.: Case study time series of observed (a) and modeled (b) radar reflectivity, lidar-detectable cloud top height, lifted condensation level and their vertical distribution. Shallow cloud fraction detected by lidar and radar in each dataset is given in each legend. Observations (a) are from research flight 5 on 2013-12-15 and also include lidar backscatter ratio (BSR) plotted below the reflectivity. Model simulation is from an ICON LEM meteogram station on 2013-12-16. The vertical distributions are normalized by the number of time steps in each scene.

A joint standard grid for the radar and lidar observations and forward simulations is used to facilitate additional analysis. A grid spacing of seven radar range gates is chosen, so that histograms are calculated as counts in 210 m high bins normalized by the bin width and the total number of cases in the total dataset. The histogram statistics in the right part of Fig. 5.4a summarize the detected cloud layers in the scene. The integral over the histogram equals the shallow-cloud cloud fraction de-

tected by the respective sensor. In the particular scene depicted in Fig. 5.4, the lidar sees a cloud in about 73 % of the time, while the radar cloud fraction is about 46 %. Note that the histograms depict the vertical distribution of detected cloud tops or base heights in a column and are therefore different from profiles of vertical cloud fraction. In the case of multi-layer clouds, one layer is hidden by the other. To limit the analysis to shallow clouds, an upper limit is set to 4 km above sea surface. The histogram in Fig. 5.4a reveals the separation of the radar echo base into non raining drizzle in the outflow layer and precipitation that falls out of the cloud base at LCL. Note, that the lowest usable radar range bin is at about 100 m above the sea surface to avoid any surface clutter artifacts.

Figure 5.4b displays an example time series from ICON LEM which also includes precipitating clouds (beginning) and a few thermal-driven clouds (in the end). The cloud tops seen by the lidar and radar are mostly in the upper mode about 2 km above the LCL. The peak of the radar cloud top heights is about 400 m above most of the lidar cloud tops. This order is contrary to the observed case study. The higher reaching radar signal originates from grid cells at cloud top containing only rainwater but no cloud water. This can be seen by the pixels with a radar reflectivity signal above the lidar cloud top height, e.g., at 20:11. As only a few thin lidar-only-visible clouds near LCL are present, the mode of lower clouds is not very pronounced.

Two short scenes illustrate the information content gained by analyzing the vertical distributions of lidar- and radar-detectable cloud top, and radar-detectable cloud base heights. More sound findings on the relative occurrence of upper- and lower-mode clouds and their typical heights can be gained by applying this method to the full dataset

5.5.2. Cloud Statistics

To investigate whether the findings of the case study apply generally, all observations and simulations are analyzed in this section. The histograms of the observed lidar cloud top heights (Fig. 5.5) reveal, similar to the case study, two modes of cloud top heights. While the lower one is about 300 m above LCL, the upper one is about 1.3 km above LCL. Frequency wise, the upper mode dominates over the lower mode by about 30 %. The lower mode of very shallow cumulus clouds on top of the well mixed boundary layer (Stevens et al., 2017) is very likely to be thermal driven and hardly produces precipitation. The radar, however, observes in principle just one mode of top heights with its maximum at about 1.3 km above LCL, consistent with the upper lidar mode. But, similarly to the example in Fig. 5.4, the distribution is

shifted slightly towards lower top heights than the lidar-visible cloud top distribution. Overall, the lidar sees clouds more than twice as often as the radar (43.2 vs. 18.2 %) due to its higher sensitivity that even responds to low cloud water contents of about $10^{-7}\,\mathrm{kg\,kg^{-1}}$ (compare Fig. 5.3).
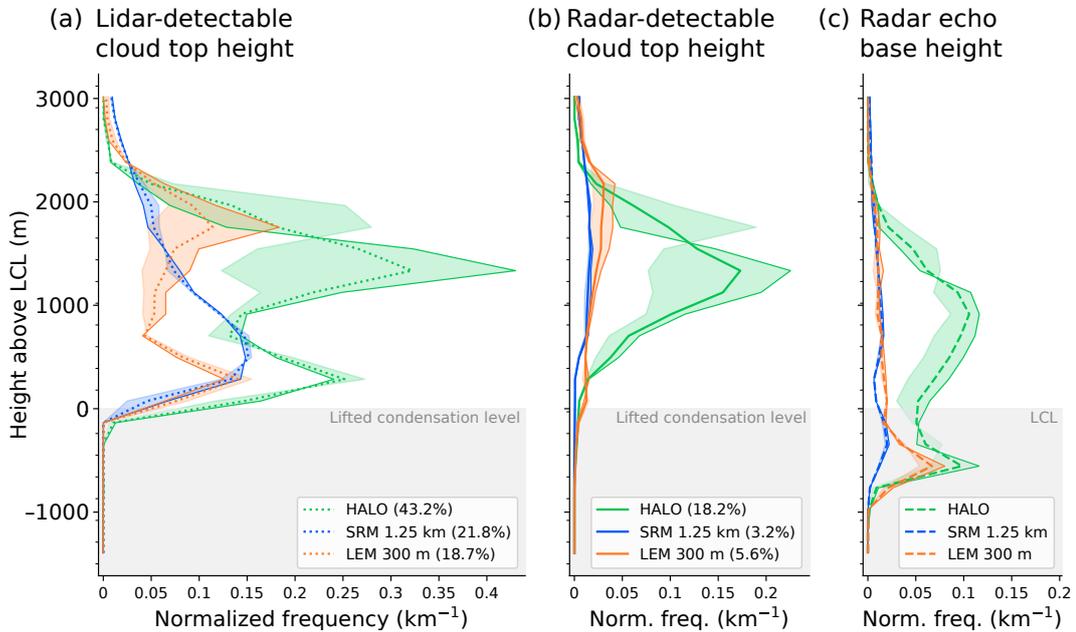
Figure 5.5.: Cloud boundary statistics on all observed and forward-simulated lidar and radar signals: (a) lidar cloud top, (b) radar cloud top, and (c) radar echo base. Same thresholds for cloud detection are used for the observed and simulated lidar and radar signals. Height is in relation to the lifted condensation level (LCL). Shadings depict western (bright edge) and eastern (dark edge) half of each dataset. The histogram bin edges are depicted as ticks on y-axis. Shallow cloud fraction detected by lidar and radar in each dataset is given in the legend.

We attribute the upper mode to shallow convection, precipitating clouds and their shallow anvil outflow. This interpretation is supported by the distribution of reflectivity bases detected by the radar. These bases are also bimodal with the upper mode about 400 m below the mode of radar top heights. This upper mode of radar base heights is related to the outflow anvils and not-yet precipitating clouds in which the layer of radar-detectable hydrometeors is only a few hundred meters thick. The lower mode of radar base heights is below the LCL, i.e., comprises clearly precipitating cases even if the precipitation occasionally evaporates before reaching the surface.

A deepening of the shallow cumulus cloud layer in accordance with a sea surface temperature increase is expected from the stratocumulus decks in the east tropical

Atlantic to the cumulus regime in the west (e.g., Wyant et al., 1997). A temperature increase of about 2 K from east to west in the flight area motives a separation of our data by longitude. The deepening of the cumulus cloud layer can be seen in the HALO observations as the lidar and radar detect the upper mode about 400 m higher in the observations west of 51.5° W than east of it. However, the frequency and height of the lower mode of the lidar-visible clouds is almost the same in the western and eastern parts.

Better pronounced than in the case study, a bimodal distribution of cloud top heights is also present in all available ICON LEM data (Fig. 5.5). The lower mode behaves very similar to the observations. It has its maximum frequency at the same height and is also detectable with the sensitivity of the lidar only. However, the frequency of this mode and the overall shallow cloud fraction is only half the observed cloud fraction (18.7 vs. 43.2 %). The height of upper mode is about 400 m above the observed upper mode of the whole dataset but matches the upper mode of the western part of observations very well. This is in line with the fact, that the LEM is only represented by meteograms in the western flight area. The shallow clouds are detected by the forward-simulated radar in only 5.6 % of the LEM scenes compared to 18.2 % in the observations. In agreement with the observations, the radar cloud tops are mostly modeled in the upper layer of the LEM, but with the maximum higher than the lidar cloud tops – similar to the example discussed before. The distribution of the modeled radar signal base heights indicates, that most clouds in the LEM are precipitating if they are visible to the radar.

The ICON SRM represents the clouds rather differently than the LEM. The clouds visible to the lidar generally form one broad mode with the most frequent lidar cloud top heights around 500 to 700 m above LCL. The frequency of shallow cloud tops decreases with altitude until they disappear at 2.6 km above LCL, which is similar to the other two datasets. The clear separation of cloud tops into two layers, however, is not evident in contrast to the observations and LEM. While a double layer structure could be seen on individual days in the SRM data (not shown), this is likely caused by the significantly varied altitude of the upper layer. Radar-detectable clouds and precipitation are also modeled but only in about 3 % of the SRM scenes which is much less than observed (18 %) and in the LEM (6 %). The radar top height distribution, however, has a similar shape as the observed radar clouds. Even if less frequent, the relative distribution of radar signal base heights in the SRM is similar to the observations with one peak between LCL and LCL + 1 km and the second peak few hundred meters below LCL. The distribution of the upper edge of the upper mode is relatively more gentle than in the observations. The clear difference of observed outflow and precipitating cloud layer between the eastern and western

part of the data is not pronounced in the SRM data, even though the coverage of the model fits better to the observations than the LEM. This indicates, that the shallow convection and outflow process is not modeled as seen during the field experiment.

Bimodal distributions of cloud top heights were also observed from space by Genkova et al. (2007) and Leahy et al. (2012). The former identified cloud top height maxima at 650 and 1500 m above sea level in an area similar to this study from about 150 scenes between September 2004 and March 2005. Both modes seem to be lower than observed in the present study, considering that the heights of the LCL is in the dropsonde, SRM, and LEM datasets in this study have means and standard deviations of 720 ± 135, 763 ± 144, and (777 ± 121) m respectively. However, Genkova et al. (2007) denote vertical uncertainty of 250 to 500 m. Leahy et al. (2012) observed the upper layer around 2 km, the lower at about 0.8 km above the sea surface in tropical Pacific trade wind cumulus (15° S, 155° W). These values are closer to the values in the present study even though the similar cloud regimes are investigated in different areas.

To conclude: Bimodal lidar cloud top height distributions were observed and their clear separation is well reproduced by the LEM but not by the SRM. The lower mode of thermal driven clouds is closely above the LCL, while the upper is closely below the trade inversion (Stevens et al., 2017), i.e., about 1.5 km higher up. The SRM, however, shows one prominent mode of cloud top heights with its maximum at rather lower heights. However, the SRM also produces deeper clouds with their frequency decreasing with height. Neither model reproduces the often observed radar echoes embedded in the non-precipitating upper stratiform outflow mode. To shed light on the conditions under which these clouds are simulated infrequently compared to observations, comprehensive LWP observations refine the statistics in the next section.

5.5.3. LWP Classes

The stratification of the observations and model data into different LWP classes can give more detailed insight into the regimes under which the models perform better or worse. LWP classes are chosen to represent barely detectable clouds ($<$ $10 \, \mathrm{g \, m^{-2}}$), clouds which are not completely optical thick ($< 50 \, \mathrm{g \, m^{-2}}$), classical cumulus clouds ($< 100 \, \mathrm{g \, m^{-2}}$), thicker clouds which are still considered in satellite retrievals ($< 300 \, \mathrm{g \, m^{-2}}$) (Wentz and Meissner, 2000), and even more water bearing clouds. An overview of cloud top heights and radar base in the different datasets and LWP ranges is presented in Fig. 5.6 and discussed in the following. To ease this discussion, we define three layers in which the lidar and radar signals occur. Every

signal below LCL is in the "precipitation" layer. Typically, only the radar base is in this layer. Signals within $1\,\mathrm{km}$ above LCL are in the so called "thermal" cloud layer. Signals above $\mathrm{LCL} + 1\,\mathrm{km}$ are called "outflow" which also includes the tops of raining clouds as depicted in Fig. 5.4.

It is remarkable, that high cloud top heights in the outflow layer were often observed by the lidar under low LWP conditions (below $10\,\mathrm{g\,m^{-2}}$). However, the lidar top heights in the outflow layer are relatively more frequent, when extending the class from $\mathrm{LWP} < 10\,\mathrm{g\,m^{-2}}$ to $\mathrm{LWP} < 50\,\mathrm{g\,m^{-2}}$. In all cases with $\mathrm{LWP} < 50\,\mathrm{g\,m^{-2}}$, the outflow layer was observed more often by the lidar than the thermal layer clouds. In general, it is no surprise that the distribution of lidar cloud tops in low LWP conditions (Fig. 5.6a and d) is similar to those of the whole dataset (Fig. 5.5a), as most of the scenes have a low LWP. For example, the statistics of the lidar-detectable top-height of scenes with $\mathrm{LWP} < 50\,\mathrm{g\,m^{-2}}$ in the SRM with only one mode and in the LEM with two modes is in general the same as discussed in the previous Sect. 5.5.2. However, the classification by LWP shows the trend in both the observations and the LEM data that outflow or precipitating clouds are more likely with higher LWP. Likewise, the thermal mode disappears in the observation and LEM datasets for higher LWP ($> 100\,\mathrm{g\,m^{-2}}$).

The statistics of radar-detectable cloud top and base heights in scenes with LWP < 10 and $50\,\mathrm{g\,m^{-2}}$ in Figs. 5.6b, c, e and f are different to the overall statistics (Fig. 5.4b), as the radar is often not sensitive enough to detect clouds with such little LWP. The lidar-detected clouds are about seven (three) times more frequent than those detected by the radar on HALO in scenes with $\mathrm{LWP} < 10\,\mathrm{g\,m^{-2}}$ ($< 50\,\mathrm{g\,m^{-2}}$). In the LEM simulations, this ratio is about five for both LWP limits. The relative smaller increase of radar-detectable clouds means that clouds in the LEM with $10\,\mathrm{g\,m^{-2}} < \mathrm{LWP} < 50\,\mathrm{g\,m^{-2}}$ probably consist out of too small droplets and thus miss a radar-detectable drizzle component. About a twelfth of the observed radar clouds with $\mathrm{LWP} < 50\,\mathrm{g\,m^{-2}}$ are categorized as precipitating, while the LEM depicts half of them as precipitating. No statement on the SRM precipitation fraction can be made as only $0.2\,\%$ (i.e., less than 200 profiles) of the SRM scenes with LWP $< 50\,\mathrm{g\,m^{-2}}$ show radar-visible cloud tops below $4\,\mathrm{km}$ at all.

The lidar detected a cloud in $96\,\%$ of the observed scenes with $\mathrm{LWP} > 50\,\mathrm{g\,m^{-2}}$. In the remaining cases, the lidar probably missed clouds with only partially coverage in the microwave radiometer footprint ($\approx 1\,\mathrm{km}$). Further, not all clouds in scenes with $\mathrm{LWP} > 50\,\mathrm{g\,m^{-2}}$ contained radar-detectable hydrometeors. This difference between lidar- and radar-detectable clouds with $\mathrm{LWP} > 50\,\mathrm{g\,m^{-2}}$ is in principle also reproduced by both models. In the observations, about four of five clouds detected by
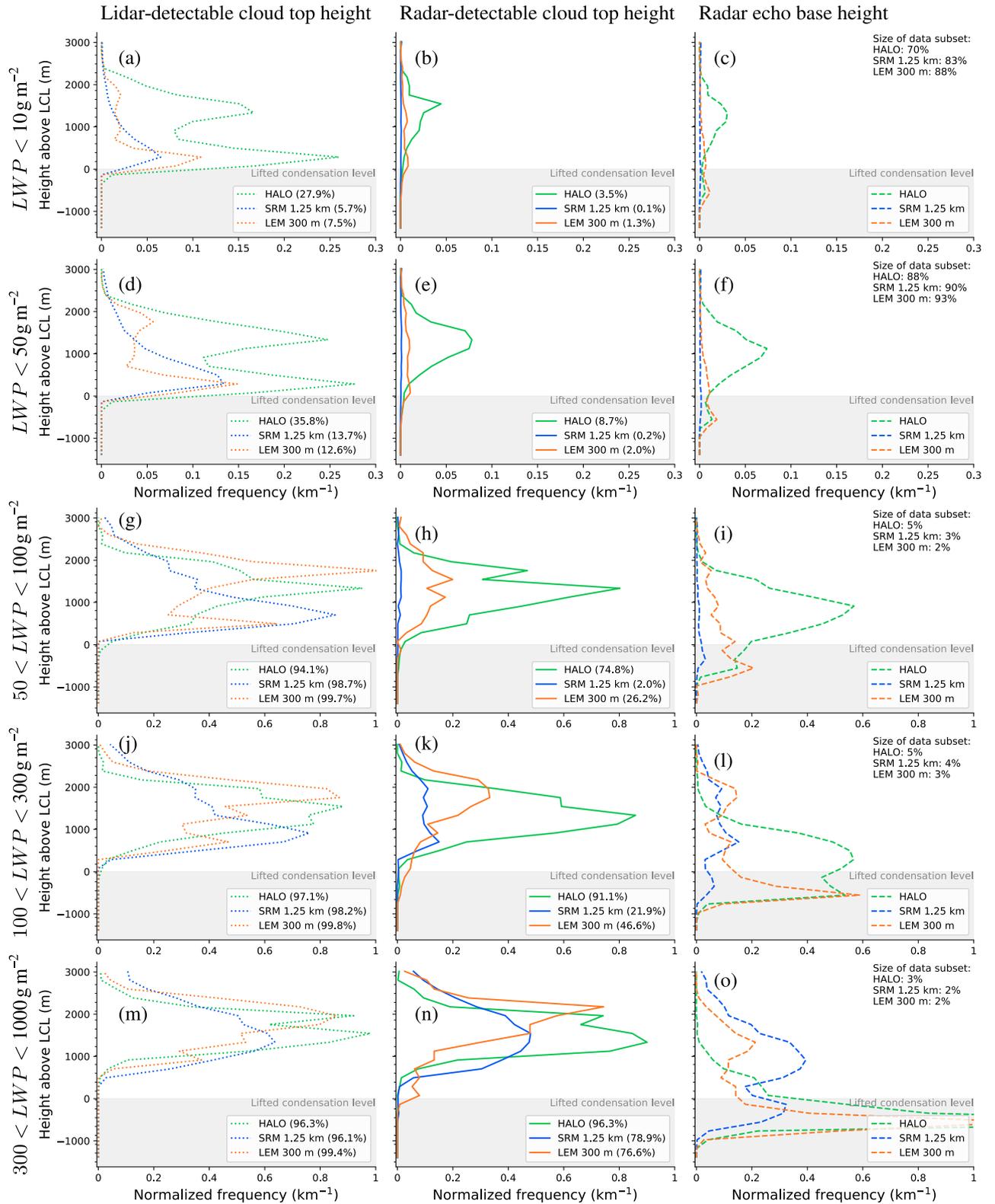
Figure 5.6.: Like Fig. 5.5 but classified by liquid water path (LWP). Columns represent lidar cloud top, radar cloud top, and radar base of observed and forward-simulated lidar and radar signals. Rows represent different LWP ranges. Note the different x-scale used in the upper two rows.

the lidar were also seen by the radar in the 50 to $100 \, \text{g m}^{-2}$ LWP class. However, only a quarter of the lidar-detectable LEM clouds are also detectable by the simulated radar. The ratio in the SRM simulations is even smaller. The radar base on the other hand shows, that the LEM models about half of the radar-detectable clouds as precipitating, while precipitation was only observed for a quarter of the observed radar clouds with $50 < \text{LWP} < 100 \, \text{g m}^{-2}$.

In scenes with LWP between 100 and $1000 \, \text{g m}^{-2}$, the radar-detectable clouds in both models form two groups. They either precipitate or form an outflow like structure with a base clearly above the LCL. Such a separation was not observed from HALO. In the observations, about a third of the 100 to $300 \, \text{g m}^{-2}$ clouds precipitate, while most others have base heights within $1 \, \text{km}$ above LCL. In the observed dataset with $\text{LWP} > 300 \, \text{g m}^{-2}$, about four fifths precipitate. The single mode of lidar-detectable cloud top height in the SRM increases with LWP. Finally, radar-detectable clouds appear more frequently in the SRM when $\text{LWP} > 300 \, \text{g m}^{-2}$. However, in these cases only a quarter of the radar-visible SRM clouds actually show a precipitating signal below the LCL.

The mode of non-precipitating radar-visible clouds under high LWP conditions in both models can also be explained by heavy clouds in the model consisting of cloud droplets only. A model cloud with $\text{LWP} > 300 \, \text{g m}^{-2}$, for example, which is $300 \, \text{m}$ deep must on average contain at least about $10^{-4} \, \text{kg kg}^{-1}$ liquid. This means that such cloud doesn't need any contribution from raindrops to be radar-detectable (compare Fig. 5.3). However, such heavy non-precipitating clouds are observed rather infrequently.

The stratification of the data by LWP shows that both models cannot represent non-precipitating but radar-visible drops that were observed under all LWP conditions. These drops are probably larger than those represented by the Gamma distributions of the cloud hydrometeor class in both models. Radar-visible model clouds precipitate more often than observed, which means they consist of already very large droplets, but the fraction of radar-visible clouds is in general too small. Non-precipitating clouds, consisting presumably of cloud-type hydrometeors only, were produced by both models under high LWP conditions ($> 300 \, \text{g m}^{-2}$), but such cases were not observed.

5.6. Summary and Conclusions

Observed statistics of hydrometeor profiles and liquid water path (LWP) of oceanic shallow cumulus clouds are compared against those produced by two high resolution

models. The observations and model runs were part of the NARVAL experiment over the tropical Atlantic east of Barbados in the dry winter season 2013. The instruments were operated from the research aircraft HALO at an altitude between 13 and 14.4 km in a nadir-pointing orientation. The two models from the ICON family are the so called storm resolving model (SRM) and the large-eddy model (LEM) with 1.25 km and 300 m grid-spacing, respectively. The SRM resolves the shallow cumulus layer with 10 to 13 layers, while the LEM has 23 to 28 levels in that layer.

The upper part of the hydrometeor profile is characterized by radar and lidar observations, while the lower part of the hydrometeor profile is characterized by the radar only due to lidar extinction. The LWP is retrieved from microwave radiometer measurements. When looking at the high occurrence of low-LWP scenes in the models (83 and 88 % below $10 \, \mathrm{g \, m^{-2}}$, Fig. 5.6), it becomes evident that common sensitivity thresholds for the instruments and models are urgently needed to assess clouds in this regime. Thus, forward simulations of the radar and lidar observations using instrument specific sensitivity thresholds and relationships between the observables and the model output are used to allow an apples-to-apples comparison between the HALO observations and the ICON model output (Lamer et al., 2018). A lidar backscatter ratio threshold of 20 suggested by Gutleben et al. (2019) is applied to clearly distinguish between backscatter from dust aerosols and cloud droplets. A comparison of the airborne measurements to ground-based radar records reveals a reliable radar reflectivity detection threshold of $-20 \, \mathrm{dBZ}$ for the airborne radar over the full column. The forward simulations show that most clouds with $q_\mathrm{c} > 10^{-7} \, \mathrm{kg \, kg^{-1}}$ in the model are detectable with the respective backscatter lidar threshold. The radar, in contrast, is primarily sensitive to the "rain"-category hydrometeor in ICON. Only the highest amounts of liquid q_c in a cloud-water-only cloud in the model are detectable by the radar.

The observations reveal two prominent modes of cumulus cloud top heights separating the clouds into two layers. The lower mode of cloud tops relates to shallow, non-precipitating boundary layer clouds reaching up to a few hundred meters above the lifted condensation level (LCL). The upper mode is mostly driven by shallow moist convection and also contains stratiform shallow outflow anvils around 1.3 km above LCL. The lower mode consists of mostly thin water clouds that are best seen by the backscatter lidar and are frequently missed by the radar. In contrast, the upper mode clouds contain more and larger droplets that scatter sufficient microwave radiation to be detected by the radar in addition to the lidar. Overall, the upper mode was observed more frequently, but both modes are similarly frequent in scenes with little condensate (LWP $< 50 \, \mathrm{g \, m^{-2}}$). In the outflow layer, the lidar detected

the cloud tops slightly higher than the radar. This indicates that small particles with low radar reflectivity are present at the upper part of the outflow layer. Higher LWP values are associated with more precipitation echoes below the LCL and with deeper outflow layers. Also, a clear trend with higher cloud tops in the upper mode in the western part of flight tracks is observed that is probably related to higher sea surface temperatures in that area enforcing convection.

The bimodal cloud top height distribution is reproduced by the LEM, although the total cloud fraction is lower than observed. The radar forward simulations suggest that the LEM produces less large particles in the outflow regimes. The observed increase of radar-detectable clouds between LWP of 10 and $50\,\mathrm{g\,m^{-2}}$ is not reproduced by the LEM. This is consistent with the overall trend of the models that produce smaller than observed particle sizes. However, the LEM describes more of the radar-detectable clouds as precipitating. This indicates that large radar-visible drops probably cannot be kept long enough in the model cloud layer before falling out. An observed cloud layer deepening with LWP can be also found in the LEM.

Different than the LEM, the SRM produces no clear separation between the two cloud layers. Cloud tops are typically at 500 to 700 m above LCL. Small differences in the warm autoconversion (AU) parametrizations might be a reason for the reduced frequency of deeper shallow clouds. The AU formulation is similar in the LEM and the SRM, but as the SRM cloud droplet number concentration N_c is constant (Table 5.1) but smaller than the average in-cloud N_c in the LEM (not shown), and as the AU rate increases with decreasing N_c (Seifert and Beheng, 2001, eq. 16), the AU in the SRM is expected to be stronger on average. Therefore rain could form quicker in the SRM and thereby reduce the average cloud life time, cloudiness, and also cloud top height. Indeed, especially the radar-visible cloud top heights of the LWP heavy clouds in the SRM are in general lower than in the LEM (Figs. 5.6 k and n). One could hypothesize further that a faster warm precipitation cycle reduces the strength of the shallow convection, so than in consequence, less clouds would reach the tropical inversion layer, which could create the shallow outflow, that is produced by the SRM too seldom. However, there are other differences between the LEM and SRM that could contribute to differences in cloudiness and rain production. For example, the lack of a clear gap might be also due to the lower vertical resolution of the SRM with 10 to 13 layers in the shallow cumulus layer (compared to 23 to 28 in the LEM) as the gap would require that always the same few model layers contain no cloud top. The clearly observed east-west difference in height of the upper cloud layer is only weak in the SRM. This indicates that processes of the precipitating shallow-convection cumulus clouds are not fully represented in the SRM. The SRM cloud distribution is rather insensitive for different LWP classes except for a cloud

deepening and precipitation increase with increasing LWP. This study primarily considers the grid-resolved clouds in the SRM. This might an unfair comparison as the SRM also contains a diagnostic scheme for sub-grid-scale cloudiness used in the radiation calculations. Thus, the additional sub-grid-scale cloudiness is briefly assessed in the Appendix 5.A. In summary, clouds modeled from diagnostic equations would moderately increase the SRM cloudiness, but would not alter the vertical structure significantly, i.e., the diagnosis does not solve the missing cloudiness in the outflow layer.

Both models show clearly non-precipitating radar-visible clouds with LWP $> 300\,\mathrm{g\,m^{-2}}$ which were not observed in that way and probably come from very high amounts of pure cloud water. In other cases, both models tend to produce precipitation that is also detectable below LCL once the cloud is visible to the radar and it seems that large radar-visible but just slowly sedimenting non-precipitating drops like in drizzle are missing. This is probably due to the size constraint in the ICON microphysics (Seifert and Beheng, 2001), that implies a threshold between cloud PSD and rain DSP at $40\,\mu\mathrm{m}$, i.e., cloud PSD is assumed to not contain a significant number of droplets with diameter larger than this threshold. Our observation of larger but non-precipitating particles is in line with findings by Siebert et al. (2013) and Wolf et al. (2019a) who observed cloud droplet effective radii on the order of this threshold in the same region but in generally moister months, i.e., they also note the principle presence of large cloud droplets.

Finally, it has to be noted that the available datasets have a great spatiotemporal overlap but do not match perfectly. The consequences of this are probably less severe than they would be for example in the mid-latitudes, a region that is heavily influenced by synoptic systems, because the study area and period is characterized as mostly undisturbed (Vial et al., 2019) and the variation from flight to flight in the winter season is limited (Fig. 4.11). Nevertheless, the methods presented in this study show high potential to benchmark realistically driven large-eddy simulations. Even with slightly different underlying meteorological statistics the analysis provides insight into processes that are well represented by the models and which phenomena are difficult to model with the respective setup. However, absolute numbers of cloud frequencies should be interpreted carefully.

Enhanced observations with several research aircraft, vessels, and autonomous platforms and coordinated model applications during the upcoming EUREC[4]A field study in early 2020 (ElUcidating the Role of Cloud-Circulation Coupling in Climate Bony et al., 2017) will provide an even more comprehensive view on the trade wind shallow cumulus clouds. For that, the methods presented here are ready to by ap-

plied to future EUREC[4]A studies. Also, cloud-chasing ship-based observations can observe individual cloud cycles including the transition from pure cloud to drizzle onset and probably rain production, while airborne observations survey the cloud field to report on the representativeness of the in-detail studied cloud. As shallow cumulus clouds also will be probed in-situ in addition to the remote sensing setup used in this study, a closer look into the drop size distributions in the outflow layers will be enabled.

5.A. Appendix: Sub-grid Clouds in the SRM

The SRM vertical cloud structure deviates stronger from the observations than the LEM, as discussed in Sect. 5.5. This might be because the forward simulations of the SRM clouds and precipitation are analyzed based on the prognostic model equations under the assumption that these clouds are resolved by the model grid. However in addition to the prognostic cloud scheme, the SRM uses a diagnostic cloud scheme to model the sub-grid-scale cloud distribution used in the SRM radiation scheme. This appendix presents a rough estimation, whether the diagnostic cloud scheme provides the missed outflow clouds.

The diagnostic cloud scheme uses a simple box probability density function of total water content and provides the diagnostic cloud cover (CLC) and liquid cloud water content ($q_{c,dia}$) (Martin Köhler, personal communication). In that scheme, the total amount of water is conserved but redistributed between the vapor, and liquid and solid cloud phases. In principle, the diagnostic clouds should be analyzed as filling only their specific cloud fraction of each grid box. This means, that the diagnostic in-cloud cloud water $\frac{q_{c,dia}}{CLC}$ covers the CLC fraction of a grid box. The lidar-detectable cloud fraction *cf* in each height can then be calculated as

$$cf = \frac{1}{N} \sum_{i=1}^{N} c_i \tag{5.A.1}$$

$$c_i = \begin{cases} CLC, & \text{if } \frac{q_{c,dia}}{CLC} > t \\ 0, & \text{else,} \end{cases} \tag{5.A.2}$$

with N being the number of model columns, i the column index, and t the detection threshold. *cf* describes the spatial cover in each height that contains enough cloud water to be detectable by the lidar. Analogous to the analysis in the previous section, the prognostic cloud fraction is calculated as fraction of cells in one height level, where $q_c > t$. This is a binary assumption that implies full cloud cover, if the cloud simulated from the prognostic equations is lidar detectable.

The additional cloud fraction due to the diagnostic scheme is largest (about 3.5%) near the LCL (Fig. 5.A.1) using the sensitivity threshold $t = 10^{-7}\,\text{kg}\,\text{kg}^{-1}$ estimated from Fig. 5.3a. However, sensitivity tests (not shown) indicated, that the diagnostic and prognostic cloud fraction profiles derived from sensitivity thresholds between 10^{-5} and $10^{-8}\,\text{kg}\,\text{kg}^{-1}$ are not significantly different. The highest diagnostic cloud fraction is at the same height as the prognostic cloud fraction at about $500\,\text{m}$ above LCL but about a third higher. Above its maximum, the additional cloud fraction decreases until it approaches the prognostic cloud fraction. The diagnostic lidar-detectable cloud cover profile follows the profile of diagnostic cloud cover (clc) from the model very closely. This means, the lidar is so sensitive, that it detects all (diagnostic) model clouds with meaningful spatial extent.
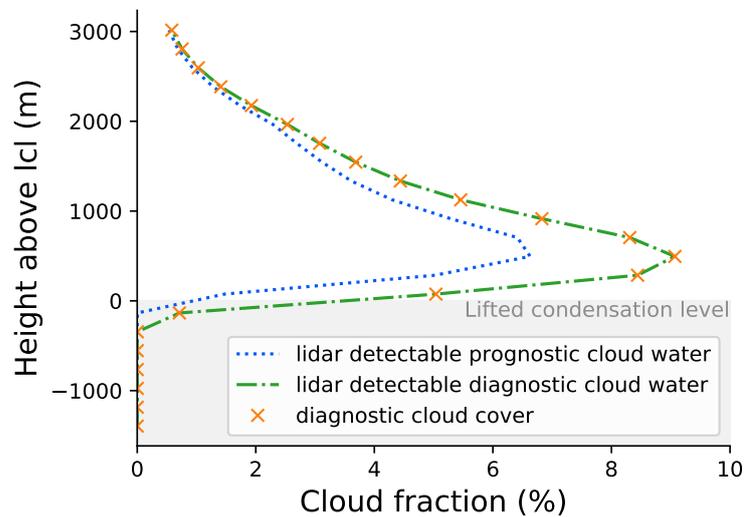
Figure 5.A.1.: Mean cloud fraction profile with for resolved and diagnostic lidar-detectable clouds in the SRM over all cases. Additionally the cloud cover profile given by the diagnostic equations is shown.

As the profile shape of diagnostic clouds is very similar to the profile of prognostic clouds, we do not expect the statistics of forward-simulated diagnostic clouds to differ much from what is discussed in Sections 5.5.2 and 5.5.3 except for a somewhat higher frequency of lidar-detectable cloud tops. However, a proper forward simulation would have to take the sub-grid cloud overlap problem into account. The radar cloud top and base statistics are almost unaffected by the diagnostic cloud water content, as the maximum additionally diagnosed cloud water content in the SRM is only $2.2 \times 10^{-4}\,\text{kg}\,\text{kg}^{-1}$. Such contribution is insignificant for the radar-detectable cloudiness in relation to the radar detection threshold (compare Fig. 5.3c).

5.B. Appendix: Comment on Data Interdependence

The ICON SRM dataset is the day-time subset of the dataset which was used for the LWP retrieval development and assessment. In consequence, the observational LWP data, that is used to classify the radar and lidar observations by LWP, is not fully independent from the SRM dataset. We are aware, that this may by a limitation of this study. This independence could potentially increase the agreement between the observations and the SRM, compared to the LEM. However, no such evidence is found as the general agreement of the LEM is better. Thus, the data independence should be kept in mind when developing retrievals from model data and assessing model data, but the independence seems acceptable in this study.

5.C. Remarks

height data available. Furthermore, we would like to acknowledge the discussion on the ICON microphysics with Axel Seifert and diagnostics with Martin Köhler and Harald Rybka from the German Weather Service (DWD).

Chapter 6.

Conclusions and Perspectives

The thesis at hand presents a comprehensive study to better characterize tropical marine shallow-convection clouds using the HALO Microwave Package *HAMP*. Shallow cumulus clouds are a globally ubiquitous and important component of the climate system. Their characteristics are not yet fully understood in current climate, weather, and large-eddy models – partly due to lacking observations. Here, the usefulness of airborne microwave remote sensing for studying these clouds is shown. The general advantage of microwave techniques like those used by HAMP is the semitransparency of the cloudy atmosphere to microwaves. In this thesis, the passive component of the HAMP instrument is characterized and novel retrievals based on artificial neural networks are developed. It is shown, that they can retrieve characteristics over a wide range from clear sky over thin and at most slightly precipitating clouds with a LWP of less than $100\,\mathrm{g\,m^{-2}}$ to mostly precipitating scenes with more than $1000\,\mathrm{g\,m^{-2}}$ of liquid. Subsequently, HAMP is used to demonstrate an assessment method for cloud-resolving simulations.

Before HAMP data can be used for physical retrievals, the HAMP radiometers are characterized in terms of sensitivity and accuracy. NeDTs, i.e., noisy BT variations, are estimated for all channels and their results are considered in the retrieval development. The estimated NeDT is well below $0.3\,\mathrm{K}$ for all HAMP-KV channels and below $0.6\,\mathrm{K}$ for the HAMP-FW and -G channels. Significant offsets between synthetic measurements from forward-simulated dropsondes and simultaneous HAMP measurements of more than ten times the NeDT are found. The offsets are positive as well as negative and systematic channel-dependent spectral patterns are observed in different HAMP bands. No systematic errors of realistic magnitude in the dropsonde data and forward simulations can explain the offsets consistently. A bias in the gaseous absorption model could also contribute to the spectral shape of the offset, however, such bias is assumed to cancel out, when the same model is used for correcting the offset as well as for developing retrievals. It is concluded, that the offsets are more likely caused by uncertainties in the absolute-calibration procedure regarding the cold-load temperature and changes in the belly pod during take-off and subsequent climbing to the final flight level. The offsets are stable during each flight and are only randomly influenced by spatiotemporal mismatches between the drifting sonde and the nadir observations. Because of this stability, a simple off-

set correction is developed based on the mean offset in each channel during each flight. The resulting offset-corrected HAMP BT data contributed to the dataset published by Konow et al. (2019) in the Climate and Environmental Retrieval and Archive (CERA) database for the tropical campaigns NARVAL-South, NARVAL2 and the two mid-latitude campaigns NARVAL-North and NAWDEX (Konow et al., 2018a,b,c,d). The characterization study also revealed instabilities in HAMP-G, one of the five HAMP bands, related to failing hardware. This failing is related to the sawtooth pattern in the acquired HAMP-G data which was removed from the data publication by Konow et al. (2019). Over all, the assumed accuracy of the offset-corrected BT data in all but the G band is 0.5 to 1 K.

Novel retrievals of IWV, LWP and RWP are developed and applied to compare HAMP observations from the tropical dry and wet seasons. The retrievals are based on a database of atmospheric profiles from cloud-resolving ICON simulations at 1.25 km grid spacing and related synthetic HAMP measurements. The developed IWV retrieval has a theoretical uncertainty of about $0.6\,\mathrm{kg\,m^{-2}}$ and root-mean-square deviations to water vapor lidar and dropsonde data are below $1.4\,\mathrm{kg\,m^{-2}}$ over the range from $20\,\mathrm{kg\,m^{-2}}$ to $60\,\mathrm{kg\,m^{-2}}$. For LWP and RWP, synthetic assessments are necessary as no reference measurements of column integrated hydrometeor contents are available. The LWP retrieval can be assessed as a function of LWP owing to the comprehensive size of the ICON datasets. LWP uncertainty between $12\,\mathrm{g\,m^{-2}}$ and $20\,\mathrm{g\,m^{-2}}$ is found for LWP below $100\,\mathrm{g\,m^{-2}}$ and a relative uncertainty better than 20 % is found for higher values. The RWP retrieval bases on the vertically integrated radar reflectivity in addition to BTs of eight K- and W-band channels, which are the sole input to the IWV and LWP retrievals. The RWP retrieval shows good performance with Gilbert skill scores (GSS) being higher than 0.75 for RWP thresholds from $10\,\mathrm{g\,m^{-2}}$ to about $50\,\mathrm{g\,m^{-2}}$ and higher than 0.5 for RWP up to $250\,\mathrm{g\,m^{-2}}$. This is due to the high sensitivity of the radar reflectivity to the presence of any rain.

The novel retrievals are first applied to contrast the tropical dry season in December 2013 (NARVAL-South) and the wet season in August 2016 (NARVAL2). The contrast of IWV (NARVAL-South: 28 vs. NARVAL2: $41\,\mathrm{kg\,m^{-2}}$) clearly justifies the differentiation of the two seasons. Regardless of more water vapor in the atmosphere, clouds were seen less often (30 % vs. 47 % of the time) in the wet than in the dry season. Their average LWP (40 vs. $63\,\mathrm{g\,m^{-2}}$) and RWP (2.7 vs. $6.7\,\mathrm{g\,m^{-2}}$) is lower, too. However, deeper convection into layers above the freezing level was more likely in the wet season, as microwave scattering of frozen precipitation was observed more frequently in the wet season (1.6 % vs. 0.5 % of the time). In the respective study, it is speculated, that a higher degree of cloud organization on larger scales in

the wet season reduced the overall cloud cover and observed LWP. The organization into larger cloud systems during wet season also caused an overall broader IWV distribution with higher inter-flight variability during NARVAL2. However it needs to be noted that due to flight safety regions of deep convection were avoided especially in the wet season.

The dry season with smaller clouds scales but more cloudiness seemed more typical for shallow cumuli. Therefore the NARVAL-South dataset is used to assess the ability of two versions of the ICON model with different horizontal resolutions (0.3 and 1.25 km) to simulate theses clouds. Products from the active and passive HAMP measurements are used together with the backscatter lidar in WALES to characterize cloud vertical extend statistics at different hydrometeor loads. For this model benchmark, forward operators converted the model data into the observational space for considering cloud detection thresholds. The forward simulations reveal the different cloud detection limits of the lidar and radar observations. It is shown, that most clouds are detectable by the lidar, whereas the radar is primarily sensitive to the "rain"-category hydrometeors in the models, and can detect rain of three to four orders of magnitude lower water content than it could detect pure clouds.

Two prominent modes of the shallow cumulus cloud top heights were observed. The lower mode is related to boundary layer convection with cloud tops closely above the lifted condensation level, which is at an altitude of about 700 m. The upper mode, however, is driven by deeper shallow convection and includes clouds with forming precipitation and shallow outflowing anvils that cover a larger area like a veil. This upper layer is close to the trade inversion at an altitude of about 2.3 km. The different sensitivity of the lidar and radar results in different cloud statistics which depend differently on the LWP. The kilometer-scale model reproduces the lower cloud mode of lidar-visible clouds. The upper mode of radar- and lidar-visible clouds is also reproduced in principle, but the observed gap between the layers and the relative frequency of occurrence of both modes is simulated differently than observed. Neither model accounts for drizzle sized drops, that do not precipitate but generate a detectable radar signal which was observed even in scenes with low LWP. The higher resolving model reproduces the bimodal distribution of clouds better. The different cloud representations in the two models are thought to be partially related to the different model resolutions. However, indirectly dependent on the resolution, different parametrizations were used in the two model runs. These parametrizations and especially the two-moment microphysics scheme help the higher resolved model to better represent the two-layer cloud structure.

In general, the presented studies demonstrate HAMPs suitability as a tool for researching the clouds forming in maritime shallow convection. However, the full potential of HAMP could only be utilized by combining HAMP with other instruments on HALO. The dropsondes are an important tool to correct BT offsets of the HAMP radiometers. The water vapor profiling capabilities of WALES are good to assess the water vapor measurements of the sondes and the IWV retrieval from HAMP. The backscatter channel in WALES helps to remove a slowly varying bias component in the LWP retrieval. Further, the backscatter lidar supports the HAMP radar capabilities to detect thin water clouds. The synergistic combination of instruments like in the HALO NARVAL payload is the key to interpreting the observations and providing insights into shallow cumulus convection.

Perspectives

The space of using HAMP for novel studies is still wide open. Further retrievals can be developed and refined by finding new synergies between HAMP and other instruments on HALO or even on separate platforms. Quantities that influence HAMP measurements but that are too ambiguous for retrievals could be assessed by further usage of forward simulators. But also the thorough documentation of the clouds and their interaction in the current climate state should be continued by adjusting and applying the techniques developed in this thesis or for example by Wolf et al. (2019a) to new observations like those collected during EUREC[4]A. These datasets could then be used for evaluating further datasets like those from cloud-resolving models or microwave satellites which use the same principles as HAMP but with larger footprints.

EUREC[4]A has gathered the largest continuous record of measurements of the atmospheric mass divergence profile together with comprehensive observations aiming at a four-dimensional assessment of clouds. Besides the presented HAMP capabilities, cloud observations include a horizontal and vertical cloud survey by multiple radars and lidars, combined measurements of turbulence and cloud microphysics throughout the whole boundary layer, water isotope characterization, study of ocean-atmosphere interactions across mesoscale ocean eddies, and much more. To sample all this, the multi-platform field experiment involved four research aircraft, four research vessels, an unprecedented number of autonomous air- and seaborne vehicles as well as satellites and ground-based observations, and the passion of hundreds of scientists. Within this context, HAMP acquisitions aboard HALO can provide the statistics on vertically integrated hydrometeor content. EUREC[4]A has a well

defined core area of about $200\,\mathrm{km} \times 200\,\mathrm{km}$ for atmospheric observations. HALO was mostly flying large repetitive circles in this area to obtain an ideal sampling of the clouds with the remote sensing instruments while releasing a total of 840 of dropsondes in 13 research flights to measure the large-scale vertical motion. HALO was sampling along the circles for about 6 to 8 hours during each flight. In this time, an air parcel with the speed of the trade winds (about $7\,\mathrm{m\,s}^{-1}$) moves once through the study area. This means, that HALO was sampling approximatively the whole lower troposphere in the study area once within one flight. This repetitive flight pattern could diminish statistical sampling issues raised in the comparison of NARVAL-South and NARVAL2 (Sect. 4.6).

Novel EUREC[4]A datasets as well as the now available NARVAL-South and NAR-VAL2 datasets cannot be used for assessing only cloud-resolving simulation datasets, but also satellite records of clouds. Schnitt et al. (2017) made the first step by comparing single cases of HAMP LWP to LWP from the MODIS and the Special Sensor Microwave Imager and Sounder (SSMIS). They conclude that the space-borne microwave radiometer misses quite some clouds due to it coarse resolution. With now three successfully completed campaigns in tropical shallow cumulus, a comprehensive dataset is gathered, firstly to assess more satellite overpasses in one-to-one comparisons, and secondly also to study the spatial cloud variability more in depth. Important questions to ask are whether there is a constant bias of the mean LWP due to the cloud variability within the satellite footprint or to which extent the satellite statistics are able to document light precipitation events. Again, the repetitive EUREC[4]A flight patterns promise good sampling to assess the spatial satellite-footprint inhomogeneity of the clouds. Seaborne observations from the ships could be used as well to study the temporal cloud evolution at one point. Furthermore, observations from the ships as well as from Barbados could be used to assess the assumptions made in spaceborne retrievals. For example, the easily accessible and widely used retrievals by Remote Sensing Systems (Wentz and Meissner, 2000) assume, that the rain water path depends on the rain column height, which is related to the freezing level, which is parameterized as a function of sea surface temperature (Wentz and Spencer, 1998). Related observations are now available from micro rain radar (MRR; Peters et al., 2002) and atmospheric and oceanic in-situ instruments.

The MRR might become handy also for verifying HAMP products of rain rate estimates which could extend the IWV, LWP, and RWP products in the future. One could also try to tackle the difficulties in retrieving the vertical liquid water content profile by adjusting the optimal estimation based retrieval developed by Battaglia et al. (2016) to HAMP acquisitions. The original algorithm has been developed for

multi-frequency radars and a four-channel microwave radiometer aboard the ER-2 aircraft and adjustments would have to made to a single radar frequency, but more radiometer channels and probably also the backscatter lidar.

Improvements cannot only be achieved in vertical resolution but also in horizontal resolution. The spectrometer of the Munich Aerosol and Cloud Scanner (spec-MACS; Ewald et al., 2016) is a downward looking hyper-spectral imager aboard HALO. Combining HAMP microwave observations (footprint $0.3 - 1\,\mathrm{km}$) with the visible–near-infrared passive sensors like the specMACS (ground resolution of about $10\,\mathrm{m}$) provides the opportunity of fusing the cloud penetrating information with high spatial resolution. For example, Höppler et al. (2020) started investigating these possibilities by extending the method by Barker et al. (2011) and finding similarities in the specMACS nadir and off-nadir pixels to map measurements from nadir to the off-nadir swath. LWP retrieval results from Chapter 4 allowed the construction of a 2D LWP field through this approach. However, further refinements are possible. The ability of specMACS to resolve the cloud filed within one footprint of HAMP could be used to effectively increase the spatial HAMP LWP resolution by distributing the LWP only among the area identified as cloudy by the imager.

In addition to the tropical campaigns NARVAL-South, NARVAL2, and EUREC[4]A, which provide many possibilities for subsequent studies, also the mid-latitude campaigns NARVAL-North and NAWDEX have great potential to foster knowledge by analyzing HAMP acquisitions. However, those datasets are more challenging for absolute retrievals, as the absolute BT offset correction is more uncertain due to the small number of clear-sky dropsondes as there are only one or less usable sondes for several flights (Table A.1). Therefore, data exploitation should rely stronger on the temporal variation of the data and spectral BT difference changes. HAMP capabilities for atmospheric soundings could facilitate humidity and temperature profile retrievals from an ideal combination of passive microwave (HAMP), active (WALES lidar), and in-situ (dropsondes) observations. Temperature and humidity profiles are also of special interest for the upcoming deployment of HAMP and HALO to the Arctic to study meridional warm air transport into the Arctic and marine cold-air outbreaks within the Transregional Collaborative Research Centre TR 172 *(AC)³* (ArctiC Amplification: Climate relevant Atmospheric surfaCe processes and feedback mechanisms; Wendisch et al., 2017; Brückner, 2020).

In general, retrieved HAMP products should be made available like the results of Chapter 4 that were published in the CERA database (Jacob et al., 2019a,b). In that way, the products support the broader community, for example, in cloud-resolving model studies. The IWV retrievals already allowed Reilly et al. (2020) to

study the representativeness of the local conditions in large-eddy simulations that were partially forced with HALO dropsonde data. The published HAMP data also strengthened the water vapor based assessment of cloud-resolving model simulations by Naumann and Kiemle (2019). They combined HAMP and WALES water vapor products to study the all-sky moisture range and assess the cloud representation in covariation with humidity profiles in cloud-resolving simulations with different grid spacings, including the two simulations also assessed here in Chapter 5. They found a general good agreement of the observed and simulated water vapor profiles, but also stress, that an agreement in the water vapor distribution does not need to translate into an adequate cloud representation. These are just two examples of how HAMP products support model assessments. Several model studies are planed around the EUREC[4]A field experiment and methods like demonstrated in Chapter 5 or by Naumann and Kiemle (2019) are ready to test their cloud representations.

Appendix A.

Brightness Temperature Offset During NARVAL-North and NAWDEX

The mean offset of synthetic forward-simulated dropsondes and HAMP measurements during NARVAL-North and NAWDEX are displayed in Fig. A.1. Significantly less sondes per flight could be considered here than during NARVAL-South and NARVAL2 (Sect. 3.3) due to less sondes released during clear-sky conditions (Table A.1).

Table A.1.: Number of all released dropsondes and those released in clear sky with available measurements by the HAMP-KV, -FW, and -G module during NARVAL-North and NAWDEX.

Campaign	Research flight	Date	Sondes released	Clear-sky sondes		
				KV	WF	G
NARVAL-North	RF01	2014-01-09	15	1	1	1
	RF02	2014-01-12	6	1	1	0
	RF03	2014-01-18	9	1	1	1
	RF04	2014-01-20	11	1	1	1
	RF05	2014-01-21	8	2	2	2
NAWDEX	RF01	2016-09-17	10	0	0	0
	RF02	2016-09-21	14	2	2	2
	RF03	2016-09-23	21	7	7	7
	RF04	2016-09-26	25	3	3	3
	RF05	2016-09-27	20	7	7	6
	RF06	2016-10-06	20	2	2	2
	RF07	2016-10-09	1	0	0	0
	RF08	2016-10-10	19	2	2	2
	RF09	2016-10-13	24	1	1	0
	RF10	2016-10-14	7	1	1	1
	RF11	2016-10-15	12	1	1	0
	RF12	2016-10-18	15	2	2	0

Figure A.1.: As Fig. 3.12, but for NARVAL-North (a-e) and NAWDEX (f-j) flights.

Bibliography

Albern, N., 2014: Wolken über dem Nordatlantik: Vergleich von flugzeuggetragenen Messungen eines Lidars, Wolkenradars und Mikrowellenradiometers während NARVAL-Nord. B.Sc. thesis, Meteorologisches Institut, Universität Hamburg.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review*, **139 (12)**, 3887–3905, doi:10.1175/MWR-D-10-05013.1.

Barker, H. W., M. P. Jerg, T. Wehr, S. Kato, D. P. Donovan, and R. J. Hogan, 2011: A 3D cloud-construction algorithm for the EarthCARE satellite mission. *Quarterly Journal of the Royal Meteorological Society*, **137 (657)**, 1042–1058, doi:10.1002/qj.824.

Barrera Verdejo, M., 2016: Lidar and Microwave Radiometer Synergy for High Vertical Resolution Thermodynamic Profiling. Ph.D. thesis, Universität zu Köln.

Battaglia, A., K. Mroz, T. Lang, F. Tridon, S. Tanelli, L. Tian, and G. M. Heymsfield, 2016: Using a multiwavelength suite of microwave instruments to investigate the microphysical structure of deep convective cores. *Journal of Geophysical Research: Atmospheres*, **121 (16)**, 9356–9381, doi:10.1002/2016JD025269.

Bony, S. and J.-L. Dufresne, 2005: Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, **32 (20)**, doi:10.1029/2005GL023851.

Bony, S., H. Schulz, J. Vial, and B. Stevens, 2020: Sugar, Gravel, Fish and Flowers: Dependence of Mesoscale Patterns of Trade-wind Clouds on Environmental Conditions. *Geophysical Research Letters*, **47 (7)**, e2019GL085 988, doi:10.1029/2019GL085988.

Bony, S. and B. Stevens, 2018: Measuring area-averaged vertical motions with dropsondes. *Journal of the Atmospheric Sciences*, doi:10.1175/JAS-D-18-0141.1.

Bony, S., B. Stevens, D. M. W. Frierson, C. Jakob, M. Kageyama, R. Pincus, T. G. Shepherd, S. C. Sherwood, A. P. Siebesma, A. H. Sobel, M. Watanabe, and M. J. Webb, 2015: Clouds, circulation and climate sensitivity. *Nature Geoscience*, **8 (4)**, 261–268, doi:10.1038/ngeo2398.

Bony, S., B. Stevens, F. Ament, S. Bigorre, P. Chazette, S. Crewell, J. Delanoë, K. Emanuel, D. Farrell, C. Flamant, S. Gross, L. Hirsch, J. Karstensen, B. Mayer, L. Nuijens, J. H. Ruppert, I. Sandu, P. Siebesma, S. Speich, F. Szczap, J. Totems, R. Vogel, M. Wendisch, and M. Wirth, 2017: EUREC4A: A Field Campaign to Elucidate the Couplings Between Clouds, Convection and Circulation. *Surveys in Geophysics*, **38 (6)**, 1529–1568, doi:10.1007/s10712-017-9428-0.

Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S. Satheesh, S. Sherwood, B. Stevens, and X. Zhang, 2013: *Clouds and Aerosols*, book section 7, 571–658. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.016.

Brückner, M., 2020: HALO-(AC)^3 2021. News post, (AC)^3. URL: `http://www.ac3-tr.de/overview/observations/halo-ac3/`, accessed online 2020-04-14.

Byers, H. R. and R. K. Hall, 1955: A census of cumulus-cloud height versus precipitation in the vicinity of Puerto Rico during the winter and spring of 1953-1954. *Journal of Meteorology*, **12 (2)**, 176–178, doi:10.1175/1520-0469(1955)012<0176:ACOCCH>2.0.CO;2.

Cadeddu, M., D. Turner, and J. Liljegren, 2009: A Neural Network for Real-Time Retrievals of PWV and LWP From Arctic Millimeter-Wave Ground-Based Observations. *IEEE Transactions on Geoscience and Remote Sensing*, **47 (7)**, 1887–1900, doi:10.1109/tgrs.2009.2013205.

Cadeddu, M. P., R. Marchand, E. Orlandi, D. D. Turner, and M. Mech, 2017: Microwave Passive Ground-Based Retrievals of Cloud and Rain Liquid Water Path in Drizzling Clouds: Challenges and Possibilities. *IEEE Transactions on Geoscience and Remote Sensing*, **55 (11)**, 6468–6481, doi:10.1109/tgrs.2017.2728699.

Crewell, S., K. Ebell, U. Löhnert, and D. D. Turner, 2009: Can liquid water profiles be retrieved from passive microwave zenith observations? *Geophysical Research Letters*, **36 (6)**, doi:10.1029/2008GL036934.

Crewell, S. and U. Löhnert, 2003: Accuracy of cloud liquid water path from ground-based microwave radiometry 2. Sensor accuracy and synergy. *Radio Science*, **38 (3)**, 7–1–7–3.

Dask Development Team, 2018: *Dask: Library for dynamic task scheduling*. URL: `https://dask.org`.

Dipankar, A., B. Stevens, R. Heinze, C. Moseley, G. Zängl, M. Giorgetta, and S. Brdar, 2015: Large eddy simulation using the general circulation model ICON. *Journal of Advances in Modeling Earth Systems*, **7 (3)**, 963–986, doi:10.1002/2015MS000431.

Dodge, B. F. and A. K. Dunbar, 1927: An Investigation of the Coexisting Liquid and Vapor Phases of Solutions of Oxygen and Nitrogen. *Journal of the American Chemical Society*, **49 (3)**, 591–610.

Elsaesser, G. S., C. W. O'Dell, M. D. Lebsock, R. Bennartz, T. J. Greenwald, and F. J. Wentz, 2017: The Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP). *Journal of Climate*, **30 (24)**, 10 193–10 210, doi:10.1175/JCLI-D-16-0902.1.

Ewald, F., S. Groß, M. Hagen, L. Hirsch, J. Delanoë, and M. Bauer-Pfundstein, 2018: Calibration of a 35-GHz Airborne Cloud Radar: Lessons Learned and Intercomparisons with 94-GHz Cloud Radars. *Atmospheric Measurement Techniques Discussions*, 1–32, doi:10.5194/amt-2018-269.

Ewald, F., T. Kölling, A. Baumgartner, T. Zinner, and B. Mayer, 2016: Design and characterization of specMACS, a multipurpose hyperspectral cloud and sky imager. *Atmospheric Measurement Techniques*, **9 (5)**, 2015–2042, doi:10.5194/amt-9-2015-2016.

Fox, S., C. Lee, B. Moyna, M. Philipp, I. Rule, S. Rogers, R. King, M. Oldfield, S. Rea, M. Henry, H. Wang, and R. C. Harlow, 2017: ISMAR: an airborne sub-millimetre radiometer. *Atmospheric Measurement Techniques*, **10 (2)**, 477–490, doi:10.5194/amt-10-477-2017.

Fricke, C., A. Ehrlich, E. Jäkel, B. Bohn, M. Wirth, and M. Wendisch, 2014: Influence of local surface albedo variability and ice crystal shape on passive remote sensing of thin cirrus. *Atmospheric Chemistry and Physics*, **14 (4)**, 1943–1958, doi:10.5194/acp-14-1943-2014.

Frisch, A. S., G. Feingold, C. W. Fairall, T. Uttal, and J. B. Snider, 1998: On cloud radar and microwave radiometer measurements of stratus cloud liquid water profiles. *Journal of Geophysical Research: Atmospheres*, **103 (D18)**, 23 195–23 197, doi:10.1029/98JD01827.

Genkova, I., G. Seiz, P. Zuidema, G. Zhao, and L. Di Girolamo, 2007: Cloud top height comparisons from ASTER, MISR, and MODIS for trade wind cumuli. *Remote Sensing of Environment*, **107 (1)**, 211–222, doi:10.1016/j.rse.2006.07.021.

German Aerospace Center, 2016: HALO database. re3data.org - Registry of Research Data Repositories, doi:10.17616/r39q0t.

Giangrande, S. E., D. Wang, M. J. Bartholomew, M. P. Jensen, D. B. Mechem, J. C. Hardin, and R. Wood, 2019: Midlatitude Oceanic Cloud and Precipitation Properties as Sampled by the ARM Eastern North Atlantic Observatory. *Journal of Geophysical Research: Atmospheres*, **124 (8)**, 4741–4760, doi:10.1029/2018JD029667.

Glackin, D. L., 2014: *Remote Sensing, Physics and Techniques*, 691–701. Springer New York, New York, NY, doi:10.1007/978-0-387-36699-9_160.

Gödde, F., 2018: Detecting Clouds in the Presence of Sunglint: An Approach Using Spectral Water Vapor Absorption. M.Sc. thesis, Meteorological Institute Munich, Ludwig-Maximilians-University Munich, URL: `https://macsserver.physik.uni-muenchen.de/campaigns/documents/theses/master_thesis_felix_goedde_2018.pdf`, accessed online 2018-11-27.

Grabowski, W. W., H. Morrison, S.-I. Shima, G. C. Abade, P. Dziekan, and H. Pawlowska, 2018: Modeling of Cloud Microphysics: Can We Do Better? *Bulletin of the American Meteorological Society*, **100 (4)**, 655–672, doi:10.1175/BAMS-D-18-0005.1.

Greenwald, T. J., R. Bennartz, M. Lebsock, and J. Teixeira, 2018: An Uncertainty Data Set for Passive Microwave Satellite Observations of Warm Cloud Liquid Water Path. *Journal of Geophysical Research: Atmospheres*, **123 (7)**, 3668–3687, doi:10.1002/2017JD027638.

Gutleben, M., S. Groß, and M. Wirth, 2019: Cloud macro-physical properties in Saharan-dust-laden and dust-free North Atlantic trade wind regimes: a lidar case study. *Atmospheric Chemistry and Physics*, **19 (16)**, 10 659–10 673, doi:10.5194/acp-19-10659-2019.

Han, Y. and E. Westwater, 2000: Analysis and improvement of tipping calibration for ground-based microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, **38 (3)**, 1260–1276, doi:10.1109/36.843018.

Höppler, L., F. Gödde, M. Gutleben, T. Kölling, B. Mayer, and T. Zinner, 2020: Synergy of Active- and Passive Remote Sensing: An Approach to Reconstruct Three-Dimensional Cloud Macro- and Microphysics. *Atmospheric Measurement Techniques Discussions*, 1–37, doi:10.5194/amt-2020-49.

Hardy, W. N., 1973: Precision temperature reference for microwave radiometry (Short Papers). *IEEE Transactions on Microwave Theory and Techniques*, **21 (3)**, 149–150, doi:10.1109/TMTT.1973.1127954.

Heinze, R., et al., 2017: Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, **143**, 69–100, doi:10.1002/qj.2947.

Hogan, R. J., C. A. T. Ferro, I. T. Jolliffe, and D. B. Stephenson, 2010: Equitability Revisited: Why the "Equitable Threat Score" Is Not Equitable. *Weather and Forecasting*, **25 (2)**, 710–726, doi:10.1175/2009WAF2222350.1.

Holland, J. Z., 1970: preliminary report on the BOMEX sea-air interaction program. *Bulletin of the American Meteorological Society*, **51 (9)**, 809–821, doi:10.1175/1520-0477(1970)051<0809:PROTBS>2.0.CO;2.

Hoyer, S. and J. Hamman, 2017: xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, **5 (1)**, doi:10.5334/jors.148.

Hunter, J. D., 2007: Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, **9 (3)**, 90–95, doi:10.1109/MCSE.2007.55.

Jacob, M., F. Ament, M. Gutleben, H. Konow, M. Mech, M. Wirth, and S. Crewell, 2019: Investigating the liquid water path over the tropical Atlantic with synergistic airborne measurements. *Atmospheric Measurement Techniques*, **12 (6)**, 3237–3254, doi:10.5194/amt-12-3237-2019.

Jacob, M., F. Ament, M. Gutleben, H. Konow, M. Mech, M. Wirth, and S. Crewell, 2019a: Liquid water path and integrated water vapor over the tropical Atlantic during NARVAL-South. World Data Center for Climate (WDCC) at DKRZ, doi:10.26050/WDCC/HALO_measurements_5.

Jacob, M., F. Ament, M. Gutleben, H. Konow, M. Mech, M. Wirth, and S. Crewell, 2019b: Liquid water path and integrated water vapor over the tropical Atlantic during NARVAL2. World Data Center for Climate (WDCC) at DKRZ, doi:10.26050/WDCC/HALO_measurements_6.

Jacob, M., P. Kollias, F. Ament, V. Schemann, and S. Crewell, 2020: Multi-layer Cloud Conditions in Trade Wind Shallow Cumulus – Confronting Models with Airborne Observations. *Geoscientific Model Development Discussions*, **2020**, 1–25, doi:10.5194/gmd-2020-14.

Jeevanjee, N. and D. M. Romps, 2013: Convective self-aggregation, cold pools, and domain size. *Geophysical Research Letters*, **40 (5)**, 994–998, doi:10.1002/grl.50204.

Kazumori, M., Q. Liu, R. Treadon, and J. C. Derber, 2008: Impact study of AMSR-E radiances in the NCEP global data assimilation system. *Monthly weather review*, **136 (2)**, 541–559.

Khain, A. P., K. D. Beheng, A. Heymsfield, A. Korolev, S. O. Krichak, Z. Levin, M. Pinsky, V. Phillips, T. Prabhakaran, A. Teller, S. C. v. d. Heever, and J.-I. Yano, 2015: Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Reviews of Geophysics*, **53 (2)**, 247–322, doi:10.1002/2014RG000468.

Klepp, C., F. Ament, S. Bakan, L. Hirsch, and B. Stevens, 2014: NARVAL Campaign Report. Reports on Earth System Science 164, Max-Planck-Institut für Meteorologie. URL: `http://www.mpimet.mpg.de/fileadmin/publikationen/Reports/WEB_BzE_164_last.pdf`, accessed online 2020-04-20.

Klingebiel, M., V. P. Ghate, A. K. Naumann, F. Ditas, M. L. Pöhlker, C. Pöhlker, K. Kandler, H. Konow, and B. Stevens, 2019: Remote Sensing of Sea Salt Aerosol below Trade Wind Clouds. *Journal of the Atmospheric Sciences*, **76 (5)**, 1189–1202, doi:10.1175/JAS-D-18-0139.1.

Klocke, D., M. Brück, C. Hohenegger, and B. Stevens, 2017: Rediscovery of the doldrums in storm-resolving simulations over the tropical Atlantic. *Nature Geoscience*, **10 (12)**, 891–896, doi:10.1038/s41561-017-0005-4.

Kneifel, S., U. Löhnert, A. Battaglia, S. Crewell, and D. Siebler, 2010: Snow scattering signals in ground-based passive microwave radiometer measurements. *Journal of Geophysical Research: Atmospheres*, **115 (D16)**, doi:10.1029/2010JD013856.

Konow, H., M. Jacob, F. Ament, S. Crewell, F. Ewald, M. Hagen, L. Hirsch, F. Jansen, M. Mech, and B. Stevens, 2018a: HALO Microwave Package measurements during Next-generation Remote sensing for VALidation Studies - South (NARVAL-South). World Data Center for Climate (WDCC) at DKRZ, type: dataset, doi:10.1594/WDCC/HALO_measurements_2.

Konow, H., M. Jacob, F. Ament, S. Crewell, F. Ewald, M. Hagen, L. Hirsch, F. Jansen, M. Mech, and B. Stevens, 2018b: HALO Microwave Package measurements during Next-generation Remote sensing for VALidation Studies - North (NARVAL-North). World Data Center for Climate (WDCC) at DKRZ, type: dataset, doi:10.1594/WDCC/HALO_measurements_1.

Konow, H., M. Jacob, F. Ament, S. Crewell, F. Ewald, M. Hagen, L. Hirsch, F. Jansen, M. Mech, and B. Stevens, 2018c: HALO Microwave Package measurements during Next-generation Remote sensing for VALidation Studies 2 (NAR-VAL2). World Data Center for Climate (WDCC) at DKRZ, type: dataset, doi:10.1594/WDCC/HALO_measurements_3.

Konow, H., M. Jacob, F. Ament, S. Crewell, F. Ewald, M. Hagen, L. Hirsch, F. Jansen, M. Mech, and B. Stevens, 2018d: HALO Microwave Package measurements during North Atlantic Waveguide and Downstream impact EXperiment (NAWDEX). World Data Center for Climate (WDCC) at DKRZ, type: dataset, doi:10.1594/WDCC/HALO_measurements_4.

Konow, H., M. Jacob, F. Ament, S. Crewell, F. Ewald, M. Hagen, L. Hirsch, F. Jansen, M. Mech, and B. Stevens, 2019: A unified data set of airborne cloud remote sensing using the HALO Microwave Package (HAMP). *Earth System Science Data*, **11 (2)**, 921–934, doi:10.5194/essd-11-921-2019.

Krautstrunk, M. and A. Giez, 2012: The Transition From FALCON to HALO Era Airborne Atmospheric Research. *Atmospheric Physics*, Springer Berlin Heidelberg, 609–624, doi:10.1007/978-3-642-30183-4_37.

Küchler, N., S. Kneifel, P. Kollias, and U. Löhnert, 2018: Revisiting Liquid Water Content Retrievals in Warm Stratified Clouds: The Modified Frisch. *Geophysical Research Letters*, **45 (17)**, 9323–9330, doi:10.1029/2018GL079845.

Küchler, N., D. D. Turner, U. Löhnert, and S. Crewell, 2016: Calibrating ground-based microwave radiometers: Uncertainty and drifts. *Radio Science*, **51 (4)**, 311–327, doi:10.1002/2015RS005826.

Lamer, K., A. M. Fridlind, A. S. Ackerman, P. Kollias, E. E. Clothiaux, and M. Kelley, 2018: (GO)²-SIM: a GCM-oriented ground-observation forward-simulator framework for objective evaluation of cloud and precipitation phase. *Geoscientific Model Development*, **11 (10)**, 4195–4214, doi:10.5194/gmd-11-4195-2018.

Lamer, K., P. Kollias, and L. Nuijens, 2015: Observations of the variability of shallow trade wind cumulus cloudiness and mass flux. *Journal of Geophysical Research: Atmospheres*, **120 (12)**, 6161–6178, doi:10.1002/2014JD022950.

Leahy, L. V., R. Wood, R. J. Charlson, C. A. Hostetler, R. R. Rogers, M. A. Vaughan, and D. M. Winker, 2012: On the nature and extent of optically thin marine low clouds. *Journal of Geophysical Research: Atmospheres*, **117 (D22)**, doi:10.1029/2012JD017929.

Lebsock, M. D., T. S. L'Ecuyer, and G. L. Stephens, 2011: Detecting the ratio of rain and cloud water in low-latitude shallow marine clouds. *Journal of Applied Meteorology and Climatology*, **50 (2)**, 419–432, doi:10.1175/2010JAMC2494.1.

Lemmon, E. W., R. T. Jacobsen, S. G. Penoncello, and D. G. Friend, 2000: Thermodynamic Properties of Air and Mixtures of Nitrogen, Argon, and Oxygen From 60 to 2000 K at Pressures to 2000 MPa. *Journal of Physical and Chemical Reference Data*, **29 (3)**, 331–385, doi:10.1063/1.1285884.

Liebe, H. J., G. A. Hufford, and M. G. Cotton, 1993: Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz. *Adv. Group for Aerosp. Res. and Dev*, Specialist Meeting of the Electromagnetic Wave Propagation Panel, Palma de Mallorca, Spain.

Liljegren, J. C., S.-A. Boukabara, K. Cady-Pereira, and S. A. Clough, 2005: The effect of the half-width of the 22-GHz water vapor line on retrievals of temperature and water vapor profiles with a 12-channel microwave radiometer. *Geoscience and Remote Sensing, IEEE Transactions on*, **43 (5)**, 1102–1108.

Liou, K.-N., 2002: *An Introduction to Atmospheric Radiation.*, International Geophysics Series, Vol. 2nd ed. Academic Press.

Liu, G., J. A. Curry, J. A. Haggerty, and Y. Fu, 2001: Retrieval and characterization of cloud liquid water path using airborne passive microwave data during INDOEX. *Journal of Geophysical Research: Atmospheres*, **106 (D22)**, 28 719–28 730, doi:10.1029/2000JD900782.

Liu, Q., F. Weng, and S. J. English, 2011: An Improved Fast Microwave Water Emissivity Model. *IEEE T. Geosci. Remote*, **49 (4)**, 1238–1250, doi:10.1109/TGRS.2010.2064779.

Lohmann, U., F. Lüönd, and F. Mahrt, 2016: *An Introduction to Clouds: From the Microscale to Climate.* Cambridge University Press, doi:10.1017/CBO9781139087513.

Lohmann, U. and D. Neubauer, 2018: The importance of mixed-phase and ice clouds for climate sensitivity in the global aerosol–climate model ECHAM6-HAM2. *Atmospheric Chemistry and Physics*, **18 (12)**, 8807–8828, doi:10.5194/acp-18-8807-2018.

Maahn, M., U. Löhnert, P. Kollias, R. C. Jackson, and G. M. McFarquhar, 2015: Developing and Evaluating Ice Cloud Parameterizations for Forward Modeling of

Radar Moments Using In Situ Aircraft Observations. *Journal of Atmospheric & Oceanic Technology*, **32 (5)**, 880–903, doi:10.1175/JTECH-D-14-00112.1.

Marshall, J. S. and W. M. K. Palmer, 1948: The Distribution of Raindrops with Size. *Journal of Meteorology*, **5 (4)**, 165–166, doi:10/fdzhp6.

Maschwitz, G., U. Löhnert, S. Crewell, T. Rose, and D. D. Turner, 2013: Investigation of ground-based microwave radiometer calibration techniques at 530 hPa. *Atmospheric Measurement Techniques*, **6 (10)**, 2641–2658, doi:10.5194/amt-6-2641-2013.

Masunaga, H., T. Y. Nakajima, T. Nakajima, M. Kachi, and K. Suzuki, 2002: Physical properties of maritime low clouds as retrieved by combined use of Tropical Rainfall Measuring Mission (TRMM) Microwave Imager and Visible/Infrared Scanner 2. Climatology of warm clouds and rain. *Journal of Geophysical Research: Atmospheres*, **107 (D19)**.

Mears, C. A., D. K. Smith, L. Ricciardulli, J. Wang, H. Huelsing, and F. J. Wentz, 2018: Construction and Uncertainty Estimation of a Satellite-Derived Total Precipitable Water Data Record Over the World's Oceans. *Earth and Space Science*, **5 (5)**, 197–210, doi:10.1002/2018EA000363.

Mears, C. A., J. Wang, D. Smith, and F. J. Wentz, 2015: Intercomparison of total precipitable water measurements made by satellite-borne microwave radiometers and ground-based GPS instruments. *Journal of Geophysical Research: Atmospheres*, **120 (6)**, 2492–2504, doi:10.1002/2014JD022694.

Mech, M., M. Maahn, S. Kneifel, D. Ori, E. Orlandi, P. Kollias, V. Schemann, and S. Crewell, 2020: PAMTRA 1.0: A Passive and Active Microwave radiative TRAnsfer tool for simulating radiometer and radar measurements of the cloudy atmosphere. *Geoscientific Model Development Discussions*, **2020**, 1–34, doi:10.5194/gmd-2019-356.

Mech, M., E. Orlandi, S. Crewell, F. Ament, L. Hirsch, M. Hagen, G. Peters, and B. Stevens, 2014: HAMP – the microwave package on the High Altitude and LOng range research aircraft (HALO). *Atmospheric Measurement Techniques*, **7 (12)**, 4539–4553, doi:10.5194/amt-7-4539-2014.

Medeiros, B. and L. Nuijens, 2016: Clouds at Barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences*, **113 (22)**, E3062–E3070, doi:10.1073/pnas.1521494113.

Naumann, A. K. and C. Kiemle, 2019: The vertical Structure and spatial Variability of lower tropospheric Water Vapor and Clouds in the Trades. *Atmospheric Chemistry and Physics Discussions*, 1–24, doi:10.5194/acp-2019-1015.

Noterdaeme, P., P. Petitjean, R. Srianand, C. Ledoux, and S. López, 2011: The evolution of the cosmic microwave background temperature-measurements of TCMB at high redshift from carbon monoxide excitation. *Astronomy & Astrophysics*, **526**, L7.

Nuijens, L., I. Serikov, L. Hirsch, K. Lonitz, and B. Stevens, 2014: The distribution and variability of low-level cloud in the North Atlantic trades. *Quarterly Journal of the Royal Meteorological Society*, **140 (684)**, 2364–2374, doi:10.1002/qj.2307.

O'Connor, E. J., R. J. Hogan, and A. J. Illingworth, 2005: Retrieving Stratocumulus Drizzle Parameters Using Doppler Radar and Lidar. *Journal of Applied Meteorology*, **44 (1)**, 14–27, doi:10.1175/JAM-2181.1.

Oertel, A., M. Boettcher, H. Joos, M. Sprenger, H. Konow, M. Hagen, and H. Wernli, 2019: Convective activity in an extratropical cyclone and its warm conveyor belt – a case-study combining observations and a convection-permitting model simulation. *Quarterly Journal of the Royal Meteorological Society*, **145 (721)**, 1406–1426, doi:10.1002/qj.3500.

Oue, M., A. Tatarevic, P. Kollias, D. Wang, K. Yu, and A. M. Vogelmann, 2019: The Cloud Resolving Model Radar Simulator (CR-SIM) Version 3.2: Description and Applications of a Virtual Observatory. *Geoscientific Model Development Discussions*, 1–31, doi:10.5194/gmd-2019-207.

Painemal, D., T. Greenwald, M. Cadeddu, and P. Minnis, 2016: First extended validation of satellite microwave liquid water path with ship-based observations of marine low clouds. *Geophysical Research Letters*, **43 (12)**, 6563–6570, doi:10.1002/2016GL069061.

Pérez, F. and B. E. Granger, 2007: IPython: a System for Interactive Scientific Computing. *Computing in Science and Engineering*, **9 (3)**, 21–29, doi:10.1109/MCSE.2007.53.

Peters, G., B. Fischer, and T. Andersson, 2002: Rain observations with a vertically looking Micro Rain Radar (MRR). *Boreal environment research*, **7 (4)**, 353–362.

Petty, G., 2006: *A first course in atmospheric radiation.* Sundog Publishing.

Radiometer Physics GmbH, 2015: Instrument Operation and Software Guide - Operation Principles and Software Description for RPG standard single polarization radiometers (G5 series). Tech. Rep. RPG-MWR-STD-SW, Radiometer Physics GmbH.

Rauber, R. M., et al., 2007: Rain in Shallow Cumulus Over the Ocean: The RICO Campaign. *Bulletin of the American Meteorological Society*, **88 (12)**, 1912–1928, doi:10.1175/BAMS-88-12-1912.

Reilly, S., S. D. Gesso, and R. Neggers, 2020: Configuring LES Based on Dropsonde Data in Sparsely Sampled Areas in the Subtropical Atlantic. *Journal of Applied Meteorology and Climatology*, **59 (2)**, 297–315, doi:10.1175/JAMC-D-19-0013.1.

Reynolds, R. W., 2009: What's New in Version 2. Tech. rep., NOAA.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Journal of Climate*, **20 (22)**, 5473–5496, doi:10.1175/2007JCLI1824.1.

Rieck, M., L. Nuijens, and B. Stevens, 2012: Marine Boundary Layer Cloud Feedbacks in a Constant Relative Humidity Atmosphere. *Journal of the Atmospheric Sciences*, **69 (8)**, 2538–2550, doi:10.1175/JAS-D-11-0203.1.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding.* WORLD SCIENTIFIC, doi:10.1142/3171.

Romps, D. M., 2017: Exact Expression for the Lifting Condensation Level. *Journal of the Atmospheric Sciences*, **74 (12)**, 3891–3900, doi:10.1175/JAS-D-17-0102.1.

Rose, T., 2009: HAMP (HALO Microwave Package) Radiometer Systems, (HALO-KV, HALO-119-90, HALO-183), Version 1.00 (30. May 2009), System Specific Information. Tech. rep., Radiometer Physics GmbH.

Rose, T., S. Crewell, U. Löhnert, and C. Simmer, 2005: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmospheric Research*, **75 (3)**, 183–200, doi:10.1016/j.atmosres.2004.12.005.

Rosenkranz, P. W., 1998: Water vapor microwave continuum absorption: A comparison of measurements and models. *Radio Science*, **33 (4)**, 919–928, doi:10.1029/98RS01182.

Satoh, M., B. Stevens, F. Judt, M. Khairoutdinov, S.-J. Lin, W. M. Putman, and P. Düben, 2019: Global Cloud-Resolving Models. *Current Climate Change Reports*, **5 (3)**, 172–184, doi:10.1007/s40641-019-00131-0.

Schäfler, A., et al., 2018: The North Atlantic Waveguide and Downstream Impact Experiment. *Bulletin of the American Meteorological Society*, **99 (8)**, 1607–1637, doi:10.1175/BAMS-D-17-0003.1.

Schneider, T., J. Teixeira, C. S. Bretherton, F. Brient, K. G. Pressel, C. Schär, and A. P. Siebesma, 2017: Climate goals and computing the future of clouds. *Nature Climate Change*, **7**, 3–5, doi:10.1038/nclimate3190.

Schnitt, S., E. Orlandi, M. Mech, A. Ehrlich, and S. Crewell, 2017: Characterization of Water Vapor and Clouds During the Next-Generation Aircraft Remote Sensing for Validation (NARVAL)-South Studies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10 (7)**, 3114–3124, doi:10.1109/JSTARS.2017.2687943.

Schulz, H. and B. Stevens, 2018: Observing the Tropical Atmosphere in Moisture Space. *Journal of the Atmospheric Sciences*, **75 (10)**, 3313–3330, doi:10.1175/JAS-D-17-0375.1.

Seifert, A. and K. D. Beheng, 2001: A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, **59-60**, 265–281, doi:10.1016/S0169-8095(01)00126-0.

Sherwood, S. C., S. Bony, and J.-L. Dufresne, 2014: Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, **505 (7481)**, 37–42, doi:10.1038/nature12829.

Short, D. A. and K. Nakamura, 2000: TRMM Radar Observations of Shallow Precipitation over the Tropical Oceans. *Journal of Climate*, **13 (23)**, 4107–4124, doi:10.1175/1520-0442(2000)013<4107:TROOSP>2.0.CO;2.

Siebert, H., M. Beals, J. Bethke, E. Bierwirth, T. Conrath, K. Dieckmann, F. Ditas, A. Ehrlich, D. Farrell, S. Hartmann, M. A. Izaguirre, J. Katzwinkel, L. Nuijens, G. Roberts, M. Schäfer, R. A. Shaw, T. Schmeissner, I. Serikov, B. Stevens, F. Stratmann, B. Wehner, M. Wendisch, F. Werner, and H. Wex, 2013: The finescale structure of the trade wind cumuli over Barbados – an introduction to the CARRIBA project. *Atmospheric Chemistry and Physics*, **13 (19)**, 10 061–10 077, doi:10.5194/acp-13-10061-2013.

Siebesma, A. P., C. S. Bretherton, A. Brown, A. Chlond, J. Cuxart, P. G. Duynkerke, H. Jiang, M. Khairoutdinov, D. Lewellen, C.-H. Moeng, E. Sanchez, B. Stevens, and D. E. Stevens, 2003: A Large Eddy Simulation Intercomparison Study of Shallow Cumulus Convection. *Journal Of The Atmospheric Sciences*, **60**, 19.

Skou, N. and D. Vine, 2006: *Microwave Radiometer Systems: Design and Analysis.* Artech House remote sensing library, Artech House.

Smith, E. A., C. Kummerow, and A. Mugnai, 1994: The emergence of inversion-type profile algorithms for estimation of precipitation from satellite passive microwave measurements. *Remote Sensing Reviews*, **11 (1-4)**, 211–242, doi:10.1080/02757259409532266.

Stephens, G. L., D. G. Vane, R. J. Boain, G. G. Mace, K. Sassen, Z. Wang, A. J. Illingworth, E. J. O'connor, W. B. Rossow, S. L. Durden, S. D. Miller, R. T. Austin, A. Benedetti, C. Mitrescu, and the CloudSat Science Team, 2002: The Cloudsat Mission and the A-train. *Bulletin of the American Meteorological Society*, **83 (12)**, 1771–1790, doi:10.1175/BAMS-83-12-1771.

Stevens, B. and S. Bony, 2013: Water in the atmosphere. *Physics Today*, **66 (6)**, 29–34, doi:10.1063/PT.3.2009.

Stevens, B., H. Brogniez, C. Kiemle, J.-L. Lacour, C. Crevoisier, and J. Kiliani, 2017: Structure and Dynamical Influence of Water Vapor in the Lower Tropical Troposphere. *Surveys in Geophysics*, **38 (6)**, 1371–1397, doi:10.1007/s10712-017-9420-8.

Stevens, B., D. Farrell, L. Hirsch, F. Jansen, L. Nuijens, I. Serikov, B. Brügmann, M. Forde, H. Linne, K. Lonitz, and J. M. Prospero, 2015: The Barbados Cloud Observatory: Anchoring Investigations of Clouds and Circulation on the Edge of the ITCZ. *Bulletin of the American Meteorological Society*, **97 (5)**, 787–801, doi:10.1175/BAMS-D-14-00247.1.

Stevens, B., et al., 2019: A High-Altitude Long-Range Aircraft Configured as a Cloud Observatory: The NARVAL Expeditions. *Bulletin of the American Meteorological Society*, **100 (6)**, 1061–1077, doi:10.1175/BAMS-D-18-0198.1.

Stevens, B., et al., 2020a: Large-eddy and Storm Resolving Models for Climate Prediction - The Added Value for Clouds and Precipitation. *Journal of the Meteorological Society of Japan. Ser. II*, **98 (2)**, 395–435, doi:10.2151/jmsj.2020-021.

Stevens, B., S. Bony, H. Brogniez, L. Hentgen, C. Hohenegger, C. Kiemle, T. S. L'Ecuyer, A. K. Naumann, H. Schulz, P. A. Siebesma, J. Vial, D. M. Winker, and P. Zuidema, 2020b: Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, **146 (726)**, 141–152, doi:10.1002/qj.3662.

Stubenrauch, C. J., W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. Di Girolamo, B. Getzewich, A. Guignard, A. Heidinger, B. C. Maddux, W. P. Menzel, P. Minnis, C. Pearl, S. Platnick, C. Poulsen, J. Riedi, S. Sun-Mack, A. Walther, D. Winker, S. Zeng, and G. Zhao, 2013: Assessment of Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel. *Bulletin of the American Meteorological Society*, **94 (7)**, 1031–1049, doi:10.1175/BAMS-D-12-00117.1.

Turner, D. D., M. P. Cadeddu, U. Löhnert, S. Crewell, and A. M. Vogelmann, 2009: Modifications to the water vapor continuum in the microwave suggested by ground-based 150-GHz observations. *Geoscience and Remote Sensing, IEEE Transactions on*, **47 (10)**, 3326–3337, doi:10.1109/TGRS.2009.2022262.

Turner, D. D., A. M. Vogelmann, R. T. Austin, J. C. Barnard, K. Cady-Pereira, J. C. Chiu, S. A. Clough, C. Flynn, M. M. Khaiyer, J. Liljegren, K. Johnson, B. Lin, C. Long, A. Marshak, S. Y. Matrosov, S. A. McFarlane, M. Miller, Q. Min, P. Minimis, W. O'Hirok, Z. Wang, and W. Wiscombe, 2007: Thin Liquid Water Clouds: Their Importance and Our Challenge. *Bulletin of the American Meteorological Society*, **88 (2)**, 177–190, doi:10.1175/BAMS-88-2-177.

Ulaby, F. and D. Long, 2015: *Microwave Radar and Radiometric Remote Sensing*. Artech House.

Ulaby, F. T., R. K. Moore, and A. K. Fung, 1981: *Microwave Remote Sensing: Active and Passive*, Vol. I. Artech House.

Vaisala, 2017: Vaisala RD94 Dropsonde Datasheet. Technical data, Vaisala. URL: `https://www.vaisala.com/sites/default/files/documents/ RD94-Datasheet-B210936EN-B.pdf`, accessed online 2018-10-10.

Vömel, Hock, and Young, 2016: NCAR/EOL Technical Note –Dropsonde Dry Bias. Tech. rep., NCAR/EOL. URL: `https://www.eol.ucar.edu/system/ files/software/Aspen/Windows/W7/documents/Tech%20Note%20Dropsonde_ Dry_Bias_20160527_v1.3.pdf`, accessed online on 25 Mar 2020.

Vömel, H., H. Selkirk, M. LM, J. Valverde-Canossa, J. Valdés, E. Kyrö, R. Kivi, W. Stolz, G. Peng, and J. Diaz, 2007: Radiation Dry Bias of the Vaisala RS92 Humidity Sensor. *Journal of Atmospheric and Oceanic Technology*, **24**, doi:10.1175/JTECH2019.1.

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

van Meijgaard, E. and S. Crewell, 2005: Comparison of model predicted liquid water path with ground-based measurements during CLIWA-NET. *Atmospheric Research*, **75 (3)**, 201 – 226, doi:10.1016/j.atmosres.2004.12.006.

van Zanten, M. C., B. Stevens, L. Nuijens, A. P. Siebesma, A. S. Ackerman, F. Burnet, A. Cheng, F. Couvreux, H. Jiang, M. Khairoutdinov, Y. Kogan, D. C. Lewellen, D. Mechem, K. Nakamura, A. Noda, B. J. Shipway, J. Slawinska, S. Wang, and A. Wyszogrodzki, 2011: Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *Journal of Advances in Modeling Earth Systems*, **3 (2)**, doi:10.1029/2011MS000056.

Vial, J., R. Vogel, S. Bony, B. Stevens, D. M. Winker, X. Cai, C. Hohenegger, A. K. Naumann, and H. Brogniez, 2019: A new look at the daily cycle of tradewind cumuli. *Journal of Advances in Modeling Earth Systems*, **11 (10)**, 3148–3166, doi:10.1029/2019MS001746.

Virtanen, P., et al., 2020: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272, doi:10.1038/s41592-019-0686-2.

Vogel, R., L. Nuijens, and B. Stevens, 2016: The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *Journal of Advances in Modeling Earth Systems*, **8 (2)**, 843–862, doi:10.1002/2015MS000568.

Wendisch, M., P. Pilewskie, B. Bohn, A. Bucholtz, S. Crewell, C. Harlow, E. Jäkel, K. S. Schmidt, R. Shetter, J. Taylor, D. D. Turner, and M. Zöger, 2013: *Atmospheric Radiation Measurements*, chap. 7, 343–411. John Wiley & Sons, Ltd, doi:10.1002/9783527653218.ch7.

Wendisch, M., M. Brückner, J. P. Burrows, S. Crewell, K. Dethloff, K. Ebell, C. Lüpkes, A. Macke, J. Notholt, J. Quaas, A. Rinke, and I. Tegen, 2017: Understanding causes and effects of rapid warming in the Arctic. *Eos*, **98**, doi:10.1029/2017EO064803.

Wentz, F. J. and T. Meissner, 2000: Algorithm Theoretical Basis Document (ATBD), Version 2, AMSR Ocean Algorithm. Tech. rep., Remote Sensing Systems. URL: `http://images.remss.com/papers/rsstech/2000_121599A-1_Wentz_AMSR_Ocean_Algorithm_ATBD_Version2.pdf`, accessed online 2020-04-20.

Wentz, F. J. and R. W. Spencer, 1998: SSM/I Rain Retrievals within a Unified All-Weather Ocean Algorithm. *Journal of the Atmospheric Sciences*, **55 (9)**, 1613–1627, doi:10.1175/1520-0469(1998)055<1613:SIRRWA>2.0.CO;2.

Winker, D. M., W. H. Hunt, and M. J. McGill, 2007a: Initial performance assessment of CALIOP. *Geophysical Research Letters*, **34 (19)**, doi:10.1029/2007GL030135.

Winker, D. M., W. H. Hunt, and M. J. McGill, 2007b: Initial performance assessment of CALIOP. *Geophysical Research Letters*, **34 (19)**, doi:10.1029/2007GL030135.

Wirth, M., A. Fix, P. Mahnke, H. Schwarzer, F. Schrandt, and G. Ehret, 2009: The airborne multi-wavelength water vapor differential absorption lidar WALES: system design and performance. *Applied Physics B*, **96 (1)**, 201–213, doi:10.1007/s00340-009-3365-7.

Wolf, K., A. Ehrlich, M. Jacob, S. Crewell, M. Wirth, and M. Wendisch, 2019a: Improvement of airborne retrievals of cloud droplet number concentration of trade wind cumulus using a synergetic approach. *Atmospheric Measurement Techniques*, **12 (3)**, 1635–1658, doi:10.5194/amt-12-1635-2019.

Wolf, K., A. Ehrlich, M. Mech, R. J. Hogan, and M. Wendisch, 2019b: Evaluation of ECMWF Radiation Scheme Using Aircraft Observations of Spectral Irradiance above Clouds. *Journal Of The Atmospheric Sciences.*

Wyant, M. C., C. S. Bretherton, H. A. Rand, and D. E. Stevens, 1997: Numerical Simulations and a Conceptual Model of the Stratocumulus to Trade Cumulus Transition. *Journal of the Atmospheric Sciences*, **54 (1)**, 168–192, doi:10.1175/1520-0469(1997)054<0168:NSAACM>2.0.CO;2.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, **141 (687)**, 563–579, doi:10.1002/qj.2378.

Zhou, L., Q. Liu, D. Liu, L. Xie, L. Qi, and X. Liu, 2016: Validation of MODIS liquid water path for oceanic nonraining warm clouds: Implications on the vertical profile of cloud water content. *Journal of Geophysical Research: Atmospheres*, **121 (9)**, 4855–4876, doi:10.1002/2015JD024499.

Acknowledgment

I would like to use the opportunity to thank everyone who made this work possible. First of all, I thank my supervisors Susanne, Felix, and Mario for guiding me through my doctoral studies, giving advices, and pointing to opportunities. Then, I would also like to thank Pavlos for all his inspiration. Thank you Ana, Bernhard, Birte, Christoph, Claudia, Davide, Kerstin, Lisa, Lukas, María, Nils, Pavel, Philipp, Rainer, Rosa, Sabrina, Stephanie, Uli, Vera, and everybody else in the institute of Geophysics and Meteorology for exchanging a word or more and making every day in the institute a good day. Furthermore, I would like to thank the kind proxies to the broader university administration Dagmar and Nadine for organizing the everyday life in the background. In particular, I would like to thank the co-authors of my two publications (Chapters 4 and 5) for joining the studies and would also like to refer to the acknowledgments at the end of the respective Chapters. I would like to thank everybody involved in the NARVAL, NAWDEX, and EUREC^{4}A field experiments for their productive efforts, joyful mood, and inspiring enthusiasm. The NAWDEX, *(AC)^3* ACLOUD, and EUREC^{4}A field experiments, which I could join, impressed me in their own, unforgettable, and unique spirits. Thank you Kevin, Tobi, Manu et al. for having such a great time. Especially, I would like to thank Heike for collaborating in our HALO HAMP subproject, and for finding and providing Fig. 3.2.

Finally, I would like to thank my parents for all the support they gave me over the last decades, and give special thanks to Mareike.

Software Packages

In addition to thanking all the persons that I met face to face, I must express my very profound gratitude to all contributors for free open-source software, datasets, and publications, which I could use while working on this dissertation. The major software packages that made this work possible are: The programming language Python[1], the Python packages scipy[2] as the ecosystem of open-source science including IPython[3] as interactive Python command line interface, NumPy[4] for numerical computations, Matplotlib[5] for plotting, and pandas[6] for convenient statistics, xarray[7] for seamlessly working with netCDF files and multidimensional dataset with dask[8] for advanced parallelism for analytics and enabling performance at scale and netCDF4[9] as interface for netCDF4 files; seaborn[10] for statistical data visualization and color schemes, basemap[11] for plotting 2D data on maps, and pip[12] for organizing all Python packages. The netCDF Operators[13] and Climate Data Operators (CDO)[14] were used to prepare netCDF files. I would like to acknowledge the use of imagery from the NASA Worldview application[15], part of the NASA Earth Observing System Data and Information System (EOSDIS). Presenting the results involved the TeX Live[16] LaTeX distribution to render this thesis, the version control system git[17], Inkscape[18], LibreOffice Draw[19], the GNU Image Manipulation Program(GIMP)[20], and QGIS[21]. Most software was run on the Ubuntu[22] operating system including many pre-installed tools.

---

[1] Python Software Foundation, `https://www.python.org`, versions 2.7.12, 3.5.2, 3.6.7.

[2] Virtanen et al. (2020), `https://www.scipy.org/`, versions 0.19.0, 1.0.0, 1.2.0.

[3] Pérez and Granger (2007), `https://ipython.org/`, versions, 5.1.0, 6.2.1, 7.2.0.

[4] van der Walt et al. (2011), `https://www.numpy.org`, versions 1.12.0, 1.14.2, 1.18.1.

[5] Hunter (2007), `https://matplotlib.org`, versions 1.5.2, 2.2.0, 3.1.2.

[6] `https://pandas.pydata.org/`, versions 0.22.0, 0.25.3

[7] Hoyer and Hamman (2017), `https://xarray.pydata.org`, versions 0.10.0, 0.10.7, 0.14.1.

[8] Dask Development Team (2018), `https://dask.org`, versions 1.1.3, 2.9.2.

[9] `https://unidata.github.io/netcdf4-python/`, versions 1.3.1, 1.5.3, 1.4.2.

[10] `https://seaborn.pydata.org/`, versions 0.8.1, 0.9.0.

[11] `https://matplotlib.org/basemap/`, versions 1.0.7, 1.2.0.

[12] `https://pip.pypa.io`, versions 19.1.1, 19.3.1, 20.0.1.

[13] `http://nco.sourceforge.net/`, version 4.5.4, 4.7.5

[14] `https://mpimet.mpg.de/cdo/`, versions 1.7.0, 1.9.8.

[15] `https://worldview.earthdata.nasa.gov`

[16] TeX user groups, `https://www.tug.org/texlive`, version 2017.

[17] Junio Hamano, Linus Torvalds, and many others, `https://git-scm.com`, versions 2.7.4, 2.17.1.

[18] The Inkscape Community, `https://www.inkscape.org`, versions 0.91 r13725, 0.92.3 (2405546, 2018-03-11)

[19] The Document Foundation, `https://www.libreoffice.org/`, version 6.0.7.3.

[20] S. Kimball, P. Mattis and the GIMP Development Team, `https://www.gimp.org/`, version 2.10.14.

[21] `https://qgis.org/`, version 2.18.17

[22] Canonical, `https://ubuntu.com/`, versions 16.04 and 18.04 including recent updates.

– Last access to all URLs on 22 April 2020.

Eigenständigkeitserklärung

Erklärung nach dem Wortlaut der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln vom 12. März 2020:

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten – noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen

Die in Kapitel 3 entwickelte Helligkeitstemperaturkorrektur wurde in Konow et al. (2019) veröffentlicht. Darauf basierende Datenprodukte wurden als Konow et al. (2018a,b,c,d) veröffentlicht.

Kapitel 4 wurde als Jacob et al. (2019) veröffentlicht. Darauf basierende Datenprodukte wurden als Jacob et al. (2019a,b) veröffentlicht.

Kapitel 5 wurde als Manuskript eingereicht und als Diskussionspapier (Jacob et al., 2020) veröffentlicht.

Datum: ＿＿＿＿＿＿＿   ＿＿＿＿＿＿＿＿＿＿

Marek Jacob

Datenverfügbarkeit

Die für diese Dissertation gewonnenen Primärdaten und die zu Reproduzierbarkeit der in der Dissertation dargestellten Ergebnisse genutzten Datenanalysen und Versuchsprotokolle sind im digitalen Archivsystem des Instituts für Geophysik und Meteorologie der Universität zu Köln gesichert und dort auf Anfrage verfügbar. Datenprodukte der Teilpublikationen sind im Langzeitarchiv des Deutschen Klimarechenzentrums (DKRZ) *Climate and Environmental Retrieval and Archive (CERA)* abgelegt (Konow et al., 2018a,b,c,d; Jacob et al., 2019a,b). Es sei hier ebenfalls auf die Hinweise der Daten- und Codeverfügbarkeit der Teilpublikationen am Ende der Kapitel 4 und 5 verwiesen.