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], and 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.

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

25 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. (2020) 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

30 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 asses 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

35 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., 2019). 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

40 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

45 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

50 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 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 bi-

55 modal 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

60 radiometer. The latter provides the vertically integrated LWP (Jacob et al., 2019b), which helps to approach 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

65 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; vanZanten

70 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).

75 To asses 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

80 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. 2 are followed by a brief description of

85 the model data in Sect. 3. Then, the forward simulations are presented in Sect. 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 including the analysis in LWP space. A summary and conclusions are given in Sect. 6.

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. 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 by Jacob et al. (2019b). 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.
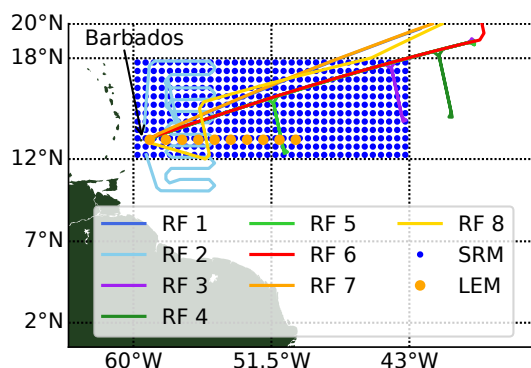
Figure 1. Map showing research flight (RF) tracks and the model columns, which are used in this study. The storm-resolving model (SRM, blue, original model grid spacing: 1.25 km) is thinned to a $0.5° \times 0.5°$ grid. From the large-eddy model (LEM, orange, original model grid spacing: 300 m), ten meteogram outputs are used.

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. (2019) and was operated at a vertical resolution of about 30 m with 1 Hz sampling. This

110   sampling frequency corresponds to a surface footprint of about 136 m × 376 m at a cruising speed of about 240 m s^{-1}.

The instruments minimal detectable signal (MDS) in dBZ decreases with range r and is estimated by Ewald et al. (2019) as

$$MDS(r) = -39.8 + 20 \log_{10} \left(\frac{r}{5\,\mathrm{km}} \right). \tag{1}$$

According to this equation, the MDS in the shallow cumulus layer is about −32 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

115   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., 2016). 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 asses the practical sensitivity limit of the HAMP radar.

120   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. 2. The BCO radar frequently measures reflectivity signals down to −70 dBZ at around 500 m with a clear maximum below 1 km for Z up to

125   −20 dBZ. Klingebiel et al. (2019) identify such weak signals at BCO below −50 dBZ as originating from sea salt aerosols and only signals above −50 dBZ are attributed to clouds. Clouds with reflectivity between the HAMP MDS (−32 dBZ) and −20 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"

130   in the observations and forward simulations.

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

135   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

140   of backscatter ratio (BSR) > 20 in the 532 nm channel to differentiate cloudy scenes from clear-sky or dusty scenes. As
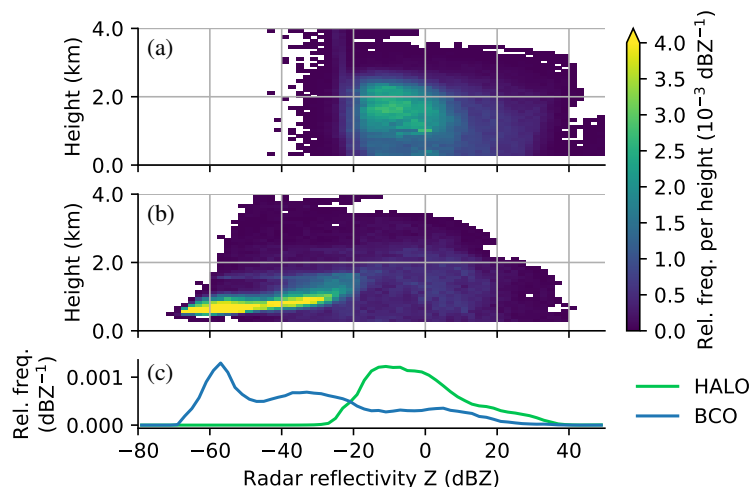
Figure 2. Height-resolved radar reflectivity distribution of shallow cumulus from (a) HAMP radar and (b) BCO radar during flight days of NARVAL-South. Marginal distributions (c) show the probability density of reflectivity from HAMP and BCO below 4 km. BCO data are limited to hours between 12:00 and 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.

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.

3 ICON-NARVAL model data

145 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.

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
150 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

155 (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. 2.1. The data are spatially subsampled on a coarser $0.5° \times 0.5°$ grid to reduce the computational effort while still conserving the variety of atmospheric profiles. A compromise of
160 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. 1. The total number of analyzed SRM columns in this study is 97920.

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
165 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 q_{nc} and q_{nr} 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
170 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. 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 37030.

4 Forward Simulations

Forward simulators, also called forward operators, can simulate how the remote sensing instruments presented in Sect. 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
180 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
185 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 assume PSDs with modified Gamma distributions. The number concentration N of spherical drops with diameter D can be described as

190 $$N(D) = N_0 \, D^\mu \exp(-\Lambda \, D^\gamma) \qquad\qquad (2)$$

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 Tab. 1.

Table 1. Configuration of modified Gamma distribution (Eq. 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	$f(q_c)$	8	$f(q_c)$	3	$q_{nc} = 2 \times 10^8 \, \mathrm{kg^{-1}}$
one moment (SRM)	raindrops	$8 \times 10^6 \, \mathrm{m^{-4}}$	0	$f(q_r)$	1	
two moments (LEM)	cloud droplets	$f(q_c, q_{nc})$	8	$f(q_c, q_{nc})$	3	
two moments (LEM)	raindrops	$f(q_r, q_{nr})$	2	$f(q_r, q_{nr})$	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

195 microphysics following the relations in Tab. 1. Though CR-SIM can also simulate radar reflectivity, the Passive and Active Microwave TRANsfer package (PAMTRA, Mech et al., submitted) 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

200 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. 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

205 rainwater to generate a radar signal $Z > -20 \, \mathrm{dBZ}$ were simulated above or horizontally attached to a precipitating cloud (e.g., Fig. 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.

The approximated proportionality of the radar reflectivity to D^6 makes Z especially sensitive to larger raindrops. Therefore, q_r (and q_{nr}) has to be considered in addition to q_c (and q_{nc}) when simulating the radar signal. The size difference between

210 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. 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. 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 q_{nc} and q_{nr}. Therefore, the forward simulation broadens the relation between the water content and the forward simulated signals (compare Figs. 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   Model – observation comparison

Observations and forward simulations of the SRM and LEM runs are used to asses 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.

5.1   Case study

An example scene observed from HALO during research flight 5 is depicted in Fig. 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. 3).

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. 4a summarize the detected cloud layers in the scene. The integral over the histogram equals the shallow-cloud cloud fraction detected by the respective sensor. In the particular scene depicted in Fig. 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
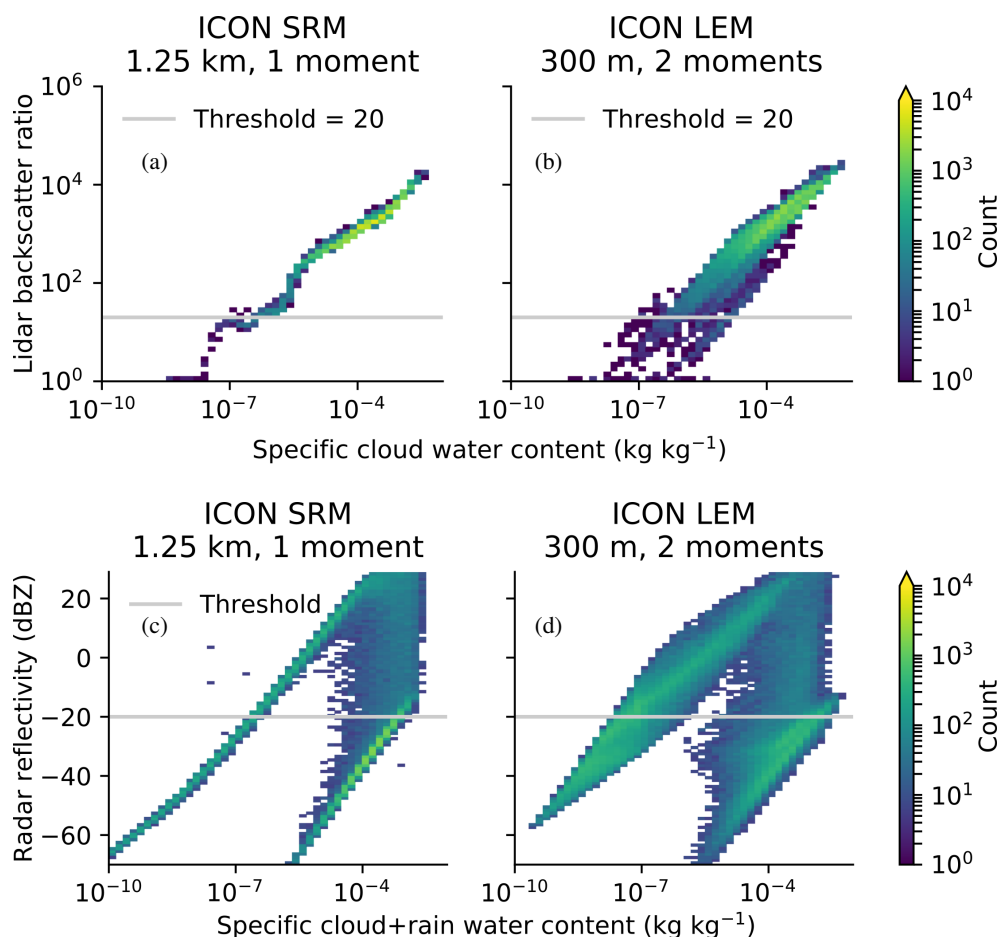
Figure 3. Simulated lidar and radar signals as function of hydrometeor contents. CR-SIM and PAMTRA simulate the observable lidar and radar signals from drop size distributions in the one-moment ICON SRM and two-moment ICON LEM microphysical models. Signals are simulated without attenuation as they would be sensed at cloud top.

column and are therefore different from profiles of vertical cloud fraction. In the case of multi-layer clouds, one layer is hidden

245   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. 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 4b displays an example time series from ICON LEM which also includes precipitating clouds (beginning) and a few

250   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.

255   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
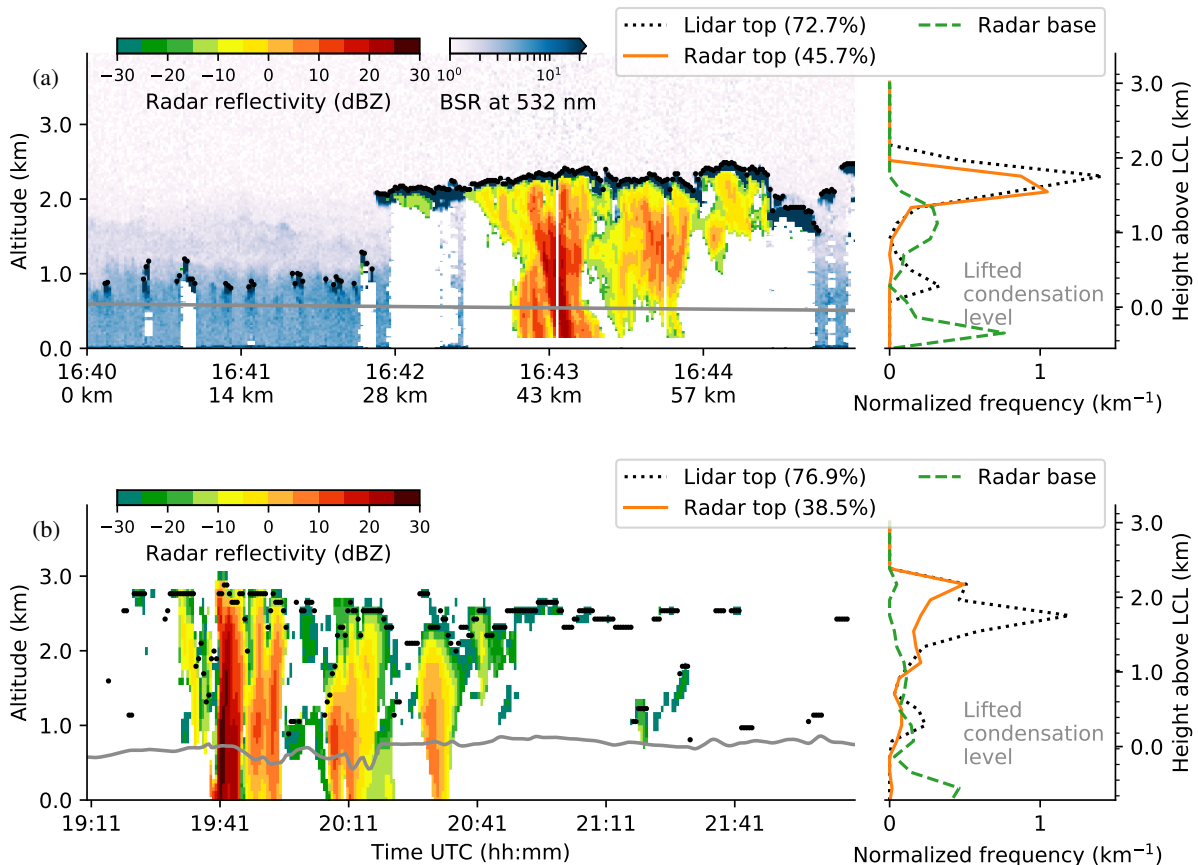
Figure 4. Case study time series of observed (a) and modeled (b) radar reflectivity, lidar-detectable cloud top height, lifted condensation level and their vertical distribution. Shallow cloud fraction detected by lidar and radar in each dataset is given in each legend. Observations (a) are from research 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.

5.2 Cloud statistics

To investigate whether the findings of the case study apply generally, all observations and simulations are analyzed in this

260   section. The histograms of the observed lidar cloud top heights (Fig. 5) reveal, similar to the case study, two modes of cloud top heights. While the lower one is about $300\,\mathrm{m}$ above LCL, the upper one is about $1.3\,\mathrm{km}$ 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

265   upper lidar mode. But, similarly to the example in Fig. 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} kg kg^{-1} (compare Fig. 3).
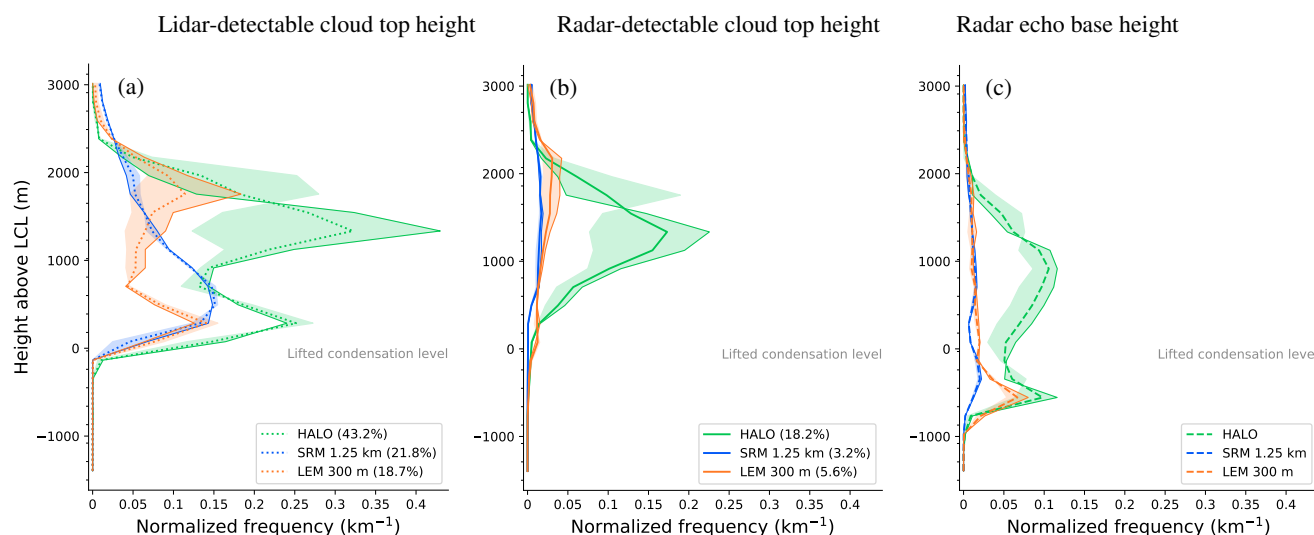
Figure 5. Cloud boundary statistics on all observed and forward simulated lidar and radar signals: (a) lidar cloud top, (b) radar cloud top, and (c) radar echo base. Same thresholds for cloud detection are used for the observed and simulated lidar and radar signals. Height is in relation to the lifted 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

270   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

275   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.

280 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). 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

285 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.

290 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

295 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

300 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

305 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

310 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

315 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.

5.3 LWP classes

The stratification of the observations and model data into different LWP classes can give more detailed insight into the regimes

320 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. 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

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

330 from LWP $< 10\,\mathrm{g\,m^{-2}}$ to LWP $< 50\,\mathrm{g\,m^{-2}}$. In all cases with 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. 6a and d) is similar to those of the whole dataset (Fig. 5a), as most of the scenes have a low LWP. For example, the statistics of the lidar-detectable top-height of scenes with 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. 5.2. However, the classification by LWP shows the

335 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. 6b, c, e and f are different to the overall statistics (Fig. 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

340 with 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 $10\,\mathrm{g\,m^{-2} < LWP < 50\,g\,m^{-2}}$ probably consist out of too small droplets and thus miss a radar-detectable drizzle component. About a twelfth of the observed radar clouds with 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

345 radar-visible cloud tops below 4 km at all.
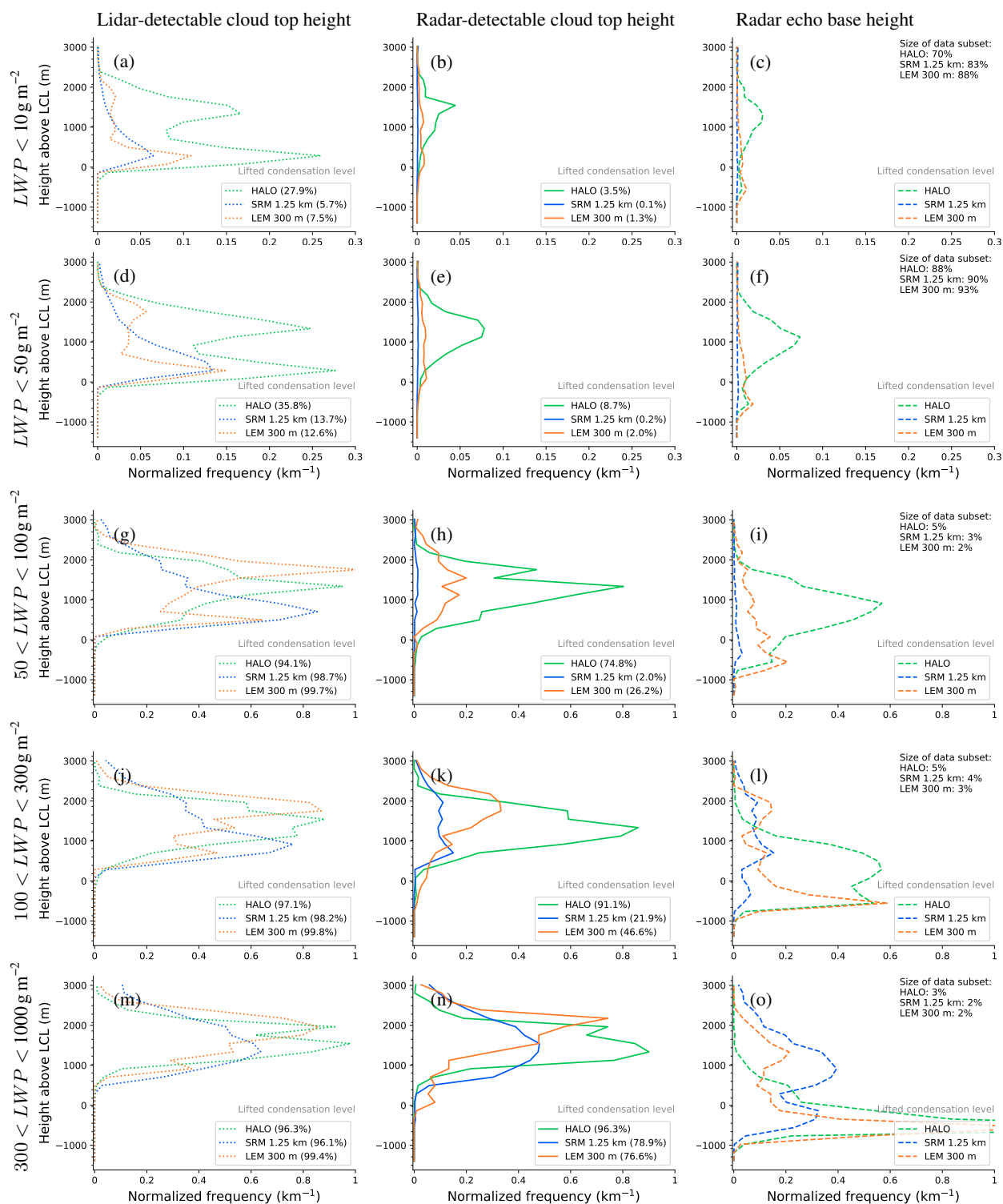
Figure 6. Like Fig. 5 but classified by liquid water path (LWP). Columns represent lidar cloud top, radar cloud top, and radar base of observed and forward simulated lidar and radar signals. Rows represent different LWP ranges. Note the different x-scale used in the upper two rows.

The lidar detected a cloud in 96 % of the observed scenes with LWP > 50 g m^{-2}. In the remaining cases, the lidar probably missed clouds with only partially coverage in the microwave radiometer footprint ($\approx$ 1 km). Further, not all clouds in scenes with LWP > 50 g m^{-2} contained radar-detectable hydrometeors. This difference between lidar- and radar-detectable clouds with LWP > 50 g m^{-2} is in principle also reproduced by both models. In the observations, about four of five clouds detected by the lidar were also seen by the radar in the 50 to 100 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 < LWP < 100 g m^{-2}.

In scenes with LWP between 100 and 1000 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 g m^{-2} clouds precipitate, while most others have base heights within 1 km above LCL. In the observed dataset with LWP > 300 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 LWP > 300 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 LWP > 300 g m^{-2}, for example, which is 300 m deep must on average contain at least about 10^{-4} kg kg^{-1} liquid. This means that such cloud doesn't need any contribution from raindrops to be radar-detectable (compare Fig. 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 g m^{-2}), but such cases were not observed.

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

380 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. 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 liar 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

385 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 $q_c > 10^{-7}\,\mathrm{kg\,kg^{-1}}$ in the model are detectable with the respective backscatter lidar threshold. The radar, in contrast, is primarily sensitive to the "rain"-category

390 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

395 shallow outflow anvils around $1.3\,\mathrm{km}$ 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

400 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

405 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 can not 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

410 to 700 m above LCL. Small differences in the warm autoconversion (AC) parametrizations might be a reason for the reduced frequency of deeper shallow clouds. The AC formulation is similar in the LEM and the SRM, but as the SRM cloud droplet number concentration q_{nc} is constant (Table. 1) but smaller than the average in-cloud q_{nc} in the LEM (not shown) and as the AC rate increases with decreasing q_{nc} (Seifert and Beheng, 2001, eq. 16), the AC 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

415   height. Indeed, especially the radar-visible cloud top heights of the LWP heavy clouds in the SRM are in general lower then in the LEM (Figs. 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

420   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

425   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 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 g m^{-2} which were not observed in that way

430   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 μm, 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

435   particles is in line with findings by Siebert et al. (2013) and Wolf et al. (2019) 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

440   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 (Jacob et al., 2019b). 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

445   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 applied 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.

Code and data availability. The source code of CR-SIM was made available by Oue et al. (2019) at https://www.bnl.gov/CMAS/cr-sim.php
(last accessed online: Nov. 6, 2019). The PAMTRA source code was made available by Mech et al. (submitted) at https://github.com/igmk/
pamtra/ (last accessed online: Nov. 6, 2019). The airborne radar and dropsonde data were published by Konow et al. (2019) as Konow et al.
(2018). The LWP retrieval data from the HAMP microwave radiometer were published by Jacob et al. (2019b) as Jacob et al. (2019a). The
BCO data are accessible to the broader community through Stevens et al. (2016). The ICON SRM and LEM data were produced by Klocke
et al. (2017) and made further public by Vial et al. (2019).

Appendix A: Sub-grid clouds in the SRM

The SRM vertical cloud structure deviates stronger from the observations than the LEM, as discussed in section 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

$$cf = \frac{1}{N} \sum_{i=1}^{N} c_i \tag{A1}$$

$$c_i = \begin{cases} clc, & \text{if } \frac{q_{c,dia}}{clc} > t \\ 0, & \text{else,} \end{cases} \tag{A2}$$

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. A1) using the sensitivity threshold $t = 10^{-7}\,\mathrm{kg\,kg^{-1}}$ estimated from Fig. 3a. However, sensitivity tests (not shown) indicated, that the diagnostic

480 and prognostic cloud fraction profiles derived from sensitivity thresholds between 10^{-5} and $10^{-8}\,\mathrm{kg\,kg^{-1}}$ are not significantly different. The highest diagnostic cloud fraction is at the same height as the prognostic cloud fraction at about 500 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.

485 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 5.2 and 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 $2.2 \times 10^{-4}\,\mathrm{kg\,kg^{-1}}$. Such contribution is insignificant

490 for the radar-detectable cloudiness in relation to the radar detection threshold (compare Fig. 3c).
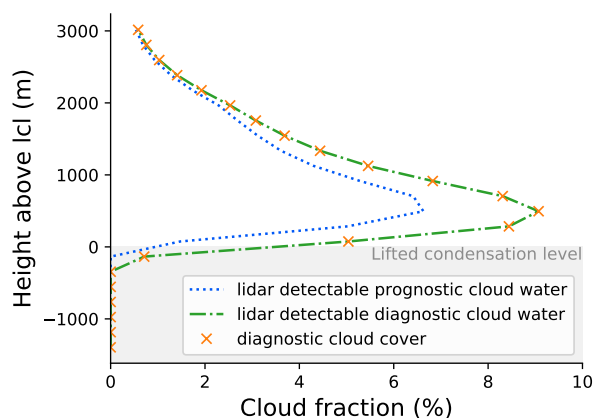
Figure A1. Mean cloud fraction profile with for resolved and diagnostic lidar-detectable clouds in the SRM over all cases. Additionally the cloud cover profile given by the diagnostic equations is shown.

Author contributions. PK and MJ conceptualized this study. FA and SC designed the observational experiment setup and supported the interpretation of the measurements. VS supported the model interpretation. MJ performed the analysis, prepared all plots, and wrote the paper with support and input from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

495 *Acknowledgements.* The work has been supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the DFG Priority Program (SPP 1294) "Atmospheric and Earth System Research with the Research Aircraft HALO (High Altitude and Long Range Research Aircraft)" under grant CR 111/12-1. We would like to thank Daniel Klocke and Matthias Brück for running the ICON simulations and the German Climate Computing Centre (DKRZ) for storing and supplying the data. Manuel Gutleben is thanked for making the WALES cloud top height data available. Furthermore, we would like to acknowledge the discussion on the ICON microphysics with Axel
500 Seifert and diagnostics with Martin Köhler and Harald Rybka from the German Weather Service (DWD).

References

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

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

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

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

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

Ewald, F., Groß, S., Hagen, M., Hirsch, L., Delanoë, J., and Bauer-Pfundstein, M.: Calibration of a 35 GHz airborne cloud radar: lessons learned and intercomparisons with 94 GHz cloud radars, Atmospheric Measurement Techniques, 12, 1815–1839, https://doi.org/10.5194/amt-12-1815-2019, 2019.

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

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

Grabowski, W. W., Morrison, H., Shima, S.-I., Abade, G. C., Dziekan, P., and Pawlowska, H.: Modeling of Cloud Microphysics: Can We Do Better?, Bulletin of the American Meteorological Society, 100, 655–672, https://doi.org/10.1175/BAMS-D-18-0005.1, 2019.

Gutleben, M., Groß, S., and Wirth, M.: 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, 10 659–10 673, https://doi.org/10.5194/acp-19-10659-2019, 2019.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Di Girolamo, P., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke, D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D., and Quaas, J.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Quarterly Journal of the Royal Meteorological Society, 143, 69–100, https://doi.org/10.1002/qj.2947, 2017.

Jacob, M., Ament, F., Gutleben, M., Konow, H., Mech, M., Wirth, M., and Crewell, S.: Liquid water path and integrated water vapor over the tropical Atlantic during NARVAL-South, https://doi.org/10.26050/WDCC/HALO_measurements_5, 2019a.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

570   Mech, M., Maahn, M., Kneifel, S., Ori, D., Orlandi, E., Kollias, P., Schemann, V., and Crewell, S.: PAMTRA 1.0: A Passive and Active Microwave radiative TRAnsfer tool for simulating radars and radiometers in cloudy and precipitating atmospheres, Geoscientific Model Development, submitted.

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

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

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

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

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

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global Cloud-Resolving Models, Current

585 Climate Change Reports, 5, 172–184, https://doi.org/10.1007/s40641-019-00131-0, 2019.

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

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

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

Siebert, H., Beals, M., Bethke, J., Bierwirth, E., Conrath, T., Dieckmann, K., Ditas, F., Ehrlich, A., Farrell, D., Hartmann, S., Izaguirre, M. A., Katzwinkel, J., Nuijens, L., Roberts, G., Schäfer, M., Shaw, R. A., Schmeissner, T., Serikov, I., Stevens, B., Stratmann, F., Wehner, B., Wendisch, M., Werner, F., and Wex, H.: The fine-scale structure of the trade wind cumuli over Barbados and an introduction to the

595 CARRIBA project, Atmospheric Chemistry and Physics, 13, 10 061–10 077, https://doi.org/10.5194/acp-13-10061-2013, 2013.

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

Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., Brügmann, B., Forde, M., Linne, H., Lonitz, K., and Prospero, J. M.:

600 The Barbados Cloud Observatory: Anchoring Investigations of Clouds and Circulation on the Edge of the ITCZ, Bulletin of the American Meteorological Society, 97, 787–801, https://doi.org/10.1175/BAMS-D-14-00247.1, 2016.

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

Stevens, B., Ament, F., Bony, S., Crewell, S., Ewald, F., Gross, S., Hansen, A., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Mayer, B.,

605 Wendisch, M., Wirth, M., Wolf, K., Bakan, S., Bauer-Pfundstein, M., Brueck, M., Delanoë, J., Ehrlich, A., Farrell, D., Forde, M., Gödde, F., Grob, H., Hagen, M., Jäkel, E., Jansen, F., Klepp, C., Klingebiel, M., Mech, M., Peters, G., Rapp, M., Wing, A. A., and Zinner, T.: A high-altitude long-range aircraft configured as a cloud observatory – the NARVAL expeditions, Bulletin of the American Meteorological Society, 100, 1061–1077, https://doi.org/10.1175/BAMS-D-18-0198.1, 2019.

Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., Rybka, H., Schubotz, W., Windmiller, J., Adamidis, P.,

610 Arka, I., Barlakas, V., Biercamp, J., Brueck, M., Brune, S., Buehler, S., Burkhardt, U., Cioni, G., Costa-sururos, M., Crewell, S., Crueger, T., Deneke, H., Friederichs, P., Carbajal Henken, C., Hohenegger, C., Jacob, M., Jakub, F., Kalthoff, N., Kohler, M., Li, P., Lohnert, U., Macke, A., Madenach, N., Mayer, B., Nam, C., Naumann, A. K., Peters, K., Poll, S., Quaas, J., Rober, N., Rochetin, N., Scheck,

L., Schemann, V., Schnitt, S., Seifert, A., Senf, F., Shapkalijevski, M., Simmer, C., Singh, S., Sourdeval, O., Spickermann, D., Strand-gren, J., Tessiot, O., Laar, T. V., Vercauteren, N., Vial, J., Voigt, A., and Zangl, G.: Large-eddy and Storm Resolving Models for Cli-
615   mate Prediction - The Added Value for Clouds and Precipitation, Journal of the Meteorological Society of Japan, accepted 31.12.2019, https://doi.org/10.2151/jmsj.2020-021, 2020.

Vaisala: Vaisala RD94 Dropsonde Datasheet, Technical data, Vaisala, https://www.vaisala.com/sites/default/files/documents/ RD94-Datasheet-B210936EN-B.pdf, 2017.

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

Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., Hohenegger, C., Naumann, A. K., and Brogniez, H.: A new look at the daily cycle of tradewind cumuli, Journal of Advances in Modeling Earth Systems, 0, ahead of print, https://doi.org/10.1029/2019MS001746,
625   2019.

Wentz, F. J. and Meissner, T.: Algorithm Theoretical Basis Document (ATBD), Version 2, AMSR Ocean Algorithm, Tech. Rep. 2, Re-mote Sensing Systems, http://images.remss.com/papers/rsstech/2000_121599A-1_Wentz_AMSR_Ocean_Algorithm_ATBD_Version2. pdf, 2000.

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

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

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

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: 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, 563–579, https://doi.org/10.1002/qj.2378, 2015.