Reply to reviewers: The vertical structure and spatial variability of lower tropospheric water vapor and clouds in the trades

Ann Kristin Naumann, Christoph Kiemle

We thank the reviewers for their helpful comments on the manuscript. In the following, reviewer's comments are in *italics*, authors' responses are in normal font.

Anonymous Referee 1

GENERAL COMMENTS

Despite its importance many questions on water vapor characteristics are still poorly understood. The paper uses unique water vapor profile measurements from two campaigns in the North Atlantic trades to investigates how well numerical simulations at different resolution capture the water vapor variability and its subsequent impact on cloudiness. This topic as well as the evaluation metrics used to investigate the problem are interesting and innovative making the paper well suited for ACP.

We thank the reviewer for the positive assessment.

Water vapor variability includes both spatial and temporal changes acting on different scales. The paper does not explicitly discuss the scale of variability addressed by its observations and the models and whether the same scales are captured. However, this is important as variability itself depends on the considered [scale (?)] *as shown by Steinkeet al. (2015, www.atmos-chem-phys.net/15/2675/2015/, their fig. 4) for a convective boundary layer with ICON. Due to the nature of the airborne measurements it is not possible to disentangle spatial and temporal variations from the observations but this could be done in the model world. In fact in the observations I suspect that there will be more correlation as in the model as spatially neighbouring profiles are correlated while model profiles are randomly distributed. Maybe this issue can be assessed by checking different approaches to select the model data, e.g. along straight lines resembling flightpaths?*

We agree that it is not possible to untangle spatial and temporal variability from the observations. For the simulations, the variability of each day is clearly dominated by the spatial variability. This can be seen in Fig. 1 in this document, where we show that the standard deviation of q_v does not differ considerably if we select one of the hourly output time steps (blue lines) or the full period (black solid line).

We also agree that we would expect different correlation scales for line-like observations than in

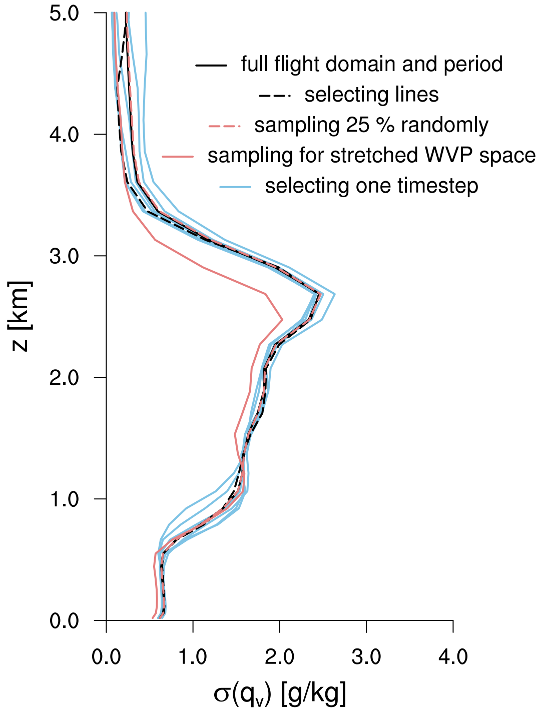

Figure 1: Standard deviation of q_v on 11 December 2013 from ICON-SRM 2.5 km for the full flight domain and period and for different sampling strategies.

the randomly selected profiles. However, for the variability the ordering does not play a role: if we select 10 latitudinal lines to mimic the flight path (dashed black line) or randomly sample 25 % of the profiles (red dashed line), the standard deviation of q_v does not differ considerably from the standard deviation of the full domain (solid black line). Only if we sample for the stretched WVP space, the standard deviation decreases because the sampling for the stretched WVP space is biased to where the lidar can measure, i.e., to outside of clouds where it is typically drier. We believe that the temporal variability plays a much smaller role in our analysis than for Steinkeel et al (2015) because of the steadiness of the trade wind regime compared to the midlatitudes. Also, the relatively large domain size might play a role.

We hesitate to add the figure to the manuscript because the direct comparison to the observations is not possible and because an in depth discussion of the temporal and spatial variability is beyond the scope of the manuscript. However, we add a clarification and short discussion of the topic in l. 159: "We analyse all model output in these domains instead of selecting profiles along the flight tracks because convection is not expected to trigger at the exact same location and time in simulations as it does in reality. Using the domain output is consistent with the statistical rather than spatio-temporal approach of this analysis and promotes the robustness of the results." and in line 197: "The analysis combines spatial and temporal variability but the contribution from spatial variability is dominating (not shown)."

The water variability assessed by the paper (with 2.5 km grid size for ICON-SRM) is not on the same scale as the shallow clouds which have typically much shorter dimensions. There should be some information on cloud dimensions available from lidar or other measurements. The issue *needs to be discussed and might be addressed in a follow-up analysis taking also cloud length into account which could be derived from backscatter lidar.*

Cloud size information from the lidar is provided in Gutleben et al. (2019). Using all flights from the two NARVAL campaigns they find that around two thirds of all trade wind clouds have a horizontal extent of less than 0.5 km. However, the contribution of these small clouds to overall cloud fraction scales with the size of the clouds. In ASTER satellite images of typical trade wind regimes (Mieslinger et al, 2019), clouds with a horizontal extent of less than 350 m contribute less than 10 % to the overall cloud fraction. We add this discussion in l. 282: "Typical cloud sizes obtained from the lidar are around 500 m (Gutleben et al 2019) and hence on the order of the grid spacing of the simulations. Because the contribution to overall cloud fraction scales with the size of the clouds, we do not expect the contribution of these small clouds to dominate the overall cloud fraction."

When trying to connect water vapor and clouds it is also interesting to look at the water budget. Water vapor mixing ratio qv might not differ much for the max and min scenarios but this difference is likely in the order of the liquid water vapor mixing ratio qc. Therefore checking how this difference translates in the end to cloud fraction might give new insights as also the microwave radiometer should be able to provide LWP simultaneously with WVP. This analysis could support the conclusion that water vapor variability not necessarily needs to relate to an adequate representation of clouds as these live at the tail of the water vapor distribution. In this respect it is interesting to know how strong temperature variability is? Would it be possible to look at relative humidity?

Here care is required in order not to confuse different things: the difference between the min and max estimate from the lidar quantifies the measurement uncertainty in the WVP due to the inability of the lidar to penetrate clouds. However, one might ask how the bias in the modeled WVP translates into cloud fraction from an order-of-magnitude water budget perspective. This is not an easy task either: Typical q_v values in the cloud layer (10 g kg^{-1}) are a factor of 30 larger than the in-cloud q_c (0.3 g kg^{-1}). From a water budget perspective, were the bias in WVP (Fig. 7 in the manuscript, say 1.2 kg m-2) not concentrated near the inversion but instead fully converted into liquid water and then distributed over the cloud layer depth (say 2000 m) with a typical in-cloud q_c, this would easily turn the whole column cloudy twice. However, we believe this estimate is flawed because the cloud layer on average is not saturated but typical relative humidities are below 90 %. As the other extreme, one could distribute all the WVP bias over the cloud layer in a way that it only increased the relative humidity without any additional saturation anywhere and hence no increase in cloud fraction. Due to the large range obtained by these two thought experiments, we believe such an order-of-magnitude argument is unfortunately not helpful at this point.

To look at relative humidity, one needs an estimate of the temperature profile, which is available from dropsondes but they are much more sparse in time and space than the lidar measurements. To avoid introducing this additional uncertainty, we concentrate on the specific humidity instead of relative humidity in the manuscript. In principle also the LWP is available from the microwave radiometer but because an order-of-magnitude budget analysis seem not to be useful in this context, we decided to focus on the in-depth analysis on the vapor phase and specific humidity for this study.

Being old fashion and looking at a printout several figures are very difficult to read and I make several suggestions for improvements in the technical section

Thanks for the comments. We mostly followed the suggestions and adjusted Fig. 1, 2, 4 and 7. Please see our answers in the technical section.

SPECIFIC COMMENTS

L7 please avoid the term "humidity inversion" as this would point at the classic polar phenomena of increasing moisture with altitude which is not what you mean. Also in line 94 this should be rephrased to make clear that you talk about the temperature inversion.

In line 7 we reformulate "... too weak humidity gradient at the inversion near the cloud top." and in line 95 "... the inversion that tops the cloud layer in the trades... ".

L8 "but is less pronounced" than what?

To clarify we added "... less pronounced than the moist model bias at the inversion".

L28 "with the decrease in subsaturation in the column" Is this true for all seasons. If you compare the different degrees of saturation in the free troposphere during the wet and dry season?

To our understanding this has been shown for all seasons in the deep convective regions. For shallow convection there are less studies but Nuijens et al (2009) show that it is true at least within the dry season in the trades, where they say that "... most of the variability in the humidity profiles, when conditioned on precipitation, is in the lower free troposphere and little is in the boundary layer ...". We clarify that in the manuscript l. 28: "... the amount of precipitation in deep convective regions correlates well with the decrease in subsaturation in the column (Bretherton et al., 2004; Holloway and Neelin, 2009). The same relation is found to hold within the dry season in the shallow convective regime (Nuijens et al., 2009)."

L27 "profiling moisture, aerosol, and clouds simultaneously with high accuracy and spatial resolution" is a bit overselling as there are limitations set by the strong lidar attenuation by clouds

We agree that the lidar is not able to profile clouds due to attenuation but rather able to detect cloud tops. We changed the text accordingly: "... profiling moisture and aerosols, and detecting cloud tops simultaneously with high accuracy and spatial resolution".

L116 For LCL it would be good to say how good the lidar approach is compared to dropsonde profiles?

We have tried to develop an algorithm to automatically retrieve the LCL from the lidar signals using cloud base heights detected from thin clouds the lidar can penetrate. These mostly correspond well with the LCL derived from the dropsonde profiles, however, due to the inability to detect the base of thick clouds with lidar the results from such an approach are slightly biased. Therefore we preferred a comprehensive case-to-case analysis of each situation, including the dropsonde results, together with auxiliary lidar information such as the aerosol and water vapor gradients at the top of the MBL. We added this indeed important information to the manuscript in line 120: "To find the LCL, we use the lidar signals from thin boundary layer clouds as well as dropsonde profiles and auxiliary lidar information such as aerosol and water vapor gradients at the top of the mixed layer. "

L133-135 I do not understand this sentence. How do you know that 1.5 % of the radiometer WVP data are affected by "saturation"?

Thanks for pointing this out. In this context, "saturation" is ambiguous and we replace it by "signal attenuation" in the manuscript.

L136: ICON has a complex grid such that resolution is not exactly the grid size, however, the true resolution of a model will always be coarser than the grid size. A discussion is needed.

The nominal grid spacing of the spherical icosahedron ICON grid is discussed by Wan et al. 2013 and Giorgetta et al. 2018. The effective resolution of ICON LEM is estimated to be a factor of six to ten larger than the nominal grid spacing by Hansen et al. 2020. We add this in the manuscript.

L153 SST fixed for each simulation day. However, SST shows spatio-temporal variation. Does SST have an influence on water vapor variability or cloud fraction?

Spatial variation of SST are included in the simulation setup. SST also differs from one simulation day to the other but not temporally within one simulation (which in our case is done for 36 h lead time). Ocean-atmosphere coupling and its influence on clouds in the trades is a topic of ongoing research (e.g., in the atmosphere-ocean coupling component of EUREC4A; Bony et al 2017). To our understanding in the tropics the effect of temporal variations of SST on the time scale of less than 2 days is most pronounced for tropical cyclones. For the trades, Vial et al (2019) show that the diurnal cycle of cloudiness is well represented, even if the SST is fixed for 36 h.

L184: Fig. 3 combines spatial and temporal variability. I would be good to split this up and check which limitation is imposed by the individual contributions. Just assuming a classical 10 m/s advection time scale gives an equivalent scale of 36 km for one hour time (ICON output frequency). That is of course a simplified view but could easily explain why models with scale 20 km and below are so similar.

Please see our answer to the first comment and Fig. 1 in this document.

L210-218: Are the numbers given here for the whole campaign or as in Fig. 4 only for 11 Dec 2013? In fact it might be good to explain before how the stretched WVP scale is generated for days and campaign?

The numbers given here are for 11 Dec 2013 and we added a clarification in the manuscript. Because all of Section 3 is about the case study, we fear that might be confusing to the reader to talk about the stretched WVP space of the seasonal composites at this point. Instead we clarified the peculiarities of the seasonal composite in Section 4 l. 310: "As for the case study in the previous section, we subsample all model results according to the percentages available from WALES in each 10 % bin of WVP for each flight individually. After the subsampling we concatenate the individual flights into the seasonal composite. The composite is thus weighted by the number of valid profiles per flight (which vary from flight to flight; Table 1).".

L329: Any idea why not? Are they to optically thick and thus as stratiform layers cover larger scales not in the data set?

Thanks for the comment. We checked the original unfiltered lidar data: Most of the stratiform cloud layers are indeed large and too optically thick, and hence not present in the analysed data set. We added this in the manuscript, l. 349: "... but are mostly removed from our analysis of the WALES data due to their opacity."

L348: The paper shows the better representation of cloud fraction for the "higher resolution" simulation. This does not necessarily need to be a resolution effect but might be due to the different cloud schemes employed by the SRM and LEM. One possibility might be autoconversion which might be too weak in SRM allowing further vertical development of the clouds.

Thank you for the suggestion. The role of the autoconversion rate has recently been highlighted by Jacob et al (2020). We include this in the discussion of Section 3, l. 289: "Another hypothesis, which has recently been developed by Jacob et al (2020), proposes that slight differences in the parameterization of autoconversion in the SRM and LEM might cause differences in the cloud's vertical extent."

L395: "in the vicinity of deep convection" is a significant part of the data from this region?

The majority of measurements are from the trade wind regime, hence we deleted the phrase here.

L409: "..the major features of the vertical distribution"

We added that.

TECHNICAL CORRECTIONS

Fig.1 is perfectly suited to add a fourth subplot with the difference between WVPmax-WVPmin which I find missing.

Thanks for the suggestion. We added a fourth plot with the WVP from $\mathrm{WALES}_{\mathrm{max}}$, $\mathrm{WVP}_{\mathrm{min}}$ und HAMP and ΔWVP to Fig. 1.

Fig. 2: I cant see anything in the water vapor plots. Maybe add a few contourlines. Wouldnt it make sense to show a microwave satellite field for WVP (from SSM/I,AMSR..)?

Unfortunately, contour lines don't work well for the ICON-LEM WVP because of the fine-scale structures at 300 m grid spacing. Instead we adjusted the scale from originally 20 - 60 kg m-2 to now 25 - 45 kg m-2, so that the contrast shows stronger now. For the WVP, we prefer to not add satellite data to keep the study focused on the comparison between lidar and simulations.

Fig. 4: I cant distinguish the different lines. Anyhow Fig. 4a already nicely shows both WVP scale so that I think that 4b could better show the difference of the models to values instead of repeating the full scale.

Fig. 4b differs from Fig. 4a by the applied subsampling. We think it is worthwhile to illustrate this transformation from "WVP space" to the "stretched WVP space" because it is a key method of the analysis and might not be very intuitive for the reader. However, we agree that the differences of the simulations to the measured values are also interesting but difficult to see in Fig. 4b. We therefore added an anomaly plot as Fig. 4c.

Fig. 7: similar to fig. 4 her b and c should be plotted as anomalies.

As for Fig. 4, we added the anomalies as Fig. 7 d and e.

Anonymous Referee 3

Synopsis: The present study investigates the variability of water vapor and clouds in the tropical Atlantic. The manuscript is mainly a model validation, i.e., it focuses on the comparison of observations with simulations that have different grid spacings. The authors found that the variability of water vapor is generally well represented by the simulations with little differences between the various model resolutions, whereas the simulated cloudfield shows a stark dependence on model resolution.

Overall comments: Overall, I think this study is appropriate for publication in ACP. I do not see any major flaws. The approach is sound, the results are sound, and, with a few exceptions detailed below, follow from the evidence. I do think, though, that the whole manuscript reads a little bit tedious; in other words, the clarity of the writing could be improved. Sometimes I feel like the authors make things more complicated than they really are. Examples are given in the specific comments below.

We followed the reviewers suggestions in the specific comments, please see our replies below.

Specific comments

I like that the authors make an effort to quantify uncertainty in the WVP measurements. This is really helpful in assessing potential model biases.

Thank you.

I think that the entire case study (Section 3) could be removed without diluting the main points of the manuscript. The results aggregated over Dec. 2013 show a similar story, and the case study is not necessary to understand the aggregated results.

We perform the case study in Section 3 for two main reasons: First, we think it is helpful for a reader to get an impression of the data, see its limitation and understand how it is connected to the synoptic situation. This would be repetitive to do for all research flights, so we decided to present a single day in more detail in the form of a case study. Second, a key point of the manuscript is how the sampled lidar data and the simulation output then translate into moisture space and how adequate subsampling translates moisture space into stretched moisture space. Since this is a new method, we strive to explain the method in detail in the relatively homogenous dataset of a single research flight. Untangling these aspects in an aggregated seasonal composite is much harder. The reviewer's next comment seems to support our notion that a careful presentation of the method is required.

We try to motivate the presentation of the case study more clearly in the introduction of Section 3, l. 173: "In this section, we use one day of the first NARVAL campaign, 11. December 2013, for a detailed case study. The aim of the case study is to introduce the central method of this study: the concept of a stretched moisture space. The stretched moisture space is obtained by selective subsampling of the model results and thereby allows for a fair comparison between lidar data and model results. The case study also illustrates some prominent features of covariation of clouds and moisture, before aggregated seasonal composite enable us to generalize the results to different regimes of water vapor structure in the trades in Section 4."

I am still unclear what "spanning the moisture space" really means. Is it just the ordering of all individual profiles with respect to their WVP? Also, it is not fully clear to me what the "stretched moisture space" accomplishes.

Correct, with moisture space we refer to an ordering of individual profiles with respect to their WVP value. This concept has been introduced by Bretherton et al (2005) and applied to observations by Schulz and Stevens (2018). Because the lidar quickly attenuates in clouds, its profiles are biased towards dry profiles (see next comment) and cannot be directly compared to simulation results. In stretched moisture space, we deliberately bias the sample of simulated profiles in the same way as the lidar profiles are biased. Therefore a fair comparison is accomplished only in stretched moisture space.

We explain the concept of moisture space and the motivation for comparing observations and simulations in stretched moisture space in detail in the first two paragraphs of Section 3.

It seems like the WALES instrument has some issues with sensing water vapor in cloudy/very moist areas that HAMPS does not have. Whats the reason for using WALES in conduction with HAMPS instead of HAMPS alone?

The HAMP radiometer measurements lack vertical profile information. Since most of our results are based on the analysis of vertical profiles, we use HAMP both as a cross check for WALES and also, and foremost, as an indicator of the WALES limits for high WVPs. We clarify this in line 135: "The nadir-viewing HAMP microwave radiometers lack vertical profile information but measure the WVP with 1 s (that is 210 m or 240 m) resolution along the HALO flight track also in the presence of shallow clouds (Jacob et al. 2019)."

How is the "cloud layer" defined?

We clarify that in the caption of Fig. 3: "The cloud layer ranges from cloud base at $z = 0.5$ km to the highest cloud tops at $z = 3.0$ km."

How are the model fields subsampled? Are model soundings drawn from under a virtual flight track?

Concerning the subsampling, please see our comment on the reviewer's question to the "stretched moisture space". The model soundings are drawn from the full domains given in Tab. 1. We clarify that in l. 159: "We analyse all model output in these domains instead of selecting profiles along the flight tracks because convection is not expected to trigger at the exact same location and time in simulations as it does in reality. Using the domain output is consistent with the statistical rather than spatio-temporal approach of this analysis and promotes the robustness of the results."

Text-figure mismatch hinders readability: For the multi-panel figures like Fig. 5, the authors first describe panel b), then c) and d), and lastly a). Making the text and figure panel order consistent would improve clarity.

Because many scientist first look at the figures of a paper before they read (or even might not read) a paper, we ordered the panels of all figures in the way that we think is most easy to understand. For Fig. 5 we think it is useful to first look at the cloud fraction in (a), which best gives an impression of the analysed regime. This sets the context for the mean, standard deviation and skewness of the humidity profiles in (b), (c) and (d). When reading the full manuscript, the reader already knows about the cloud regime and we find it reasonable to first discuss the humidity profiles and then the cloud profiles that live at the tail of the humidity distribution. To help the reader navigate, we always refer to the individual panel of each figure (e.g., Fig. 5 b).

paragraph beginning on line 183: I cant quite relate the text here to Fig. 3. Also, Im not sure what Fig. 3 is supposed to explain.

We clarified this in line 196 in the manuscript: "Averaging the results of the ICON-LEM 300 m simulation on squares of different side length, we analyse how the standard deviation of the water vapor mixing ratio, q_v, depends on the considered scales (Fig. 3)." Because the simulations and the lidar have different native scales, it is not obvious whether they need to be averaged to a common resolution for a fair comparison. The small differences between humidity variability on scales of 300 m and 20 km in Fig. 3, allow us to compare the different dataset at their native resolution without the need to artificially reduce information by averaging.

l. 252: "standard deviation of qv": How do you compute the standard deviation, in space or in time?

Please see our answer to the first reviewer's first comment. We add a short discussion in the manuscript in line 197: "The analysis combines spatial and temporal variability but the contribution from spatial variability is dominating (not shown)."

l. 275: "...but is still included in the range of uncertainty given by the observations." data from LEM 300m seem to fall outside the range estimated by WALES.

Given all the uncertainties that come with such a comparison we think that 5.8 % (ICON LEM 300 m) is still too close to 7.4 % ($\mathrm{WALES_{min}}$) to be considered substantially different. However, we agree that it is not strictly "in the range" and change the phrase to "... but is still close to the range of uncertainty given by the observations".

l. 325: "a moist model bias at the inversion" I think whats going on is that the simulated inversion is too high. This is equivalent to what you write but more intuitive when looking at the figure.

We replace the phrase in the manuscript by "a too high model inversion".

l. 352: I find the sentence construction "A too high low-cloud fraction" difficult to follow, because is has words that are the exact opposite of each other...I suggest rephrasing.

We rephrase: "If the low-cloud fraction is too large..."

l. 410-411: "In both seasons the model tends to smooth the moisture gradient at the inversion too much,..." Id say the more important bias is that the moist layer is too deep, or in other words, the inversion is too high in the models.

We rephrased the sentence before to stress the too high inversion in the model. This sentence, however, points to a second aspect.

Technicalities, typos, etc. l. 106/107: "and only in 34 % of all lidar profiles are more than half of the data points valid below this height"

We corrected that.

l. 123: What is the 12-s grid? I dont think this has been mentioned before.

The 12-s grid is introduced in line 102 but we reformulate the sentence in l. 132 to make it self-contained.

The vertical ~~Structure~~ structure and spatial ~~Variability~~ variability of lower tropospheric ~~Water Vapor~~ water vapor and ~~Clouds~~ clouds in the ~~Trades~~trades

Ann Kristin Naumann[1,2] and Christoph Kiemle[3]

[1]Max Planck Institute for Meteorology, Hamburg, Germany

[2]Meteorological Institute, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany

[3]Deutsches Zentrum für Luft- und Raumfahrt, Institute of Atmospheric Physics, Oberpfaffenhofen, Germany

Correspondence: Ann Kristin Naumann (ann-kristin.naumann@mpimet.mpg.de)

Abstract. Horizontal and vertical variability of water vapor is omnipresent in the tropics but its interaction with cloudiness poses challenges for weather and climate models. In this study we compare airborne lidar measurements from a summer and a winter field campaign in the tropical Atlantic with high-resolution simulations to analyse the water vapor distributions in the trade wind regime, its covariation with cloudiness and their representation in simulations. Across model grid spacing from 300 m to 2.5 km, the simulations show good skill in reproducing the water vapor distribution in the trades as measured by the lidar. An exception to this is a pronounced moist model bias at the top of the shallow cumulus layer in the dry winter season which is accompanied by a too weak humidity ~~inversion at the~~ gradient at the inversion near the cloud top. The model's underestimation of water vapor variability in the cloud and subcloud layer occurs in both seasons but is less pronounced than the moist model bias at the inversion. Despite the model's insensitivity to resolution from hecto- to kilometer scale for the distribution of water vapor, cloud fraction decreases strongly with increasing model resolution and is not converged at hectometer grid spacing. The observed cloud deepening with increasing water vapor path is captured well across model resolution but the concurrent transition from cloud-free to low cloud fraction is better represented at hectometer resolution. In particular, in the wet summer season the simulations with kilometer-scale resolution overestimate the observed cloud fraction near the inversion but lack condensate near the observed cloud base. This illustrates how a model's ability to properly capture the water vapor distribution does not need to translate into an adequate representation of shallow cumulus clouds that live at the tail of the water vapor distribution.

1 Introduction

Globally moisture fields, unlike temperature fields, are not smooth but they vary on the regional scale in particular in the lower troposphere where water vapor values can be large. The distribution of water vapor strongly interacts with the atmospheric circulation through the formation of clouds and convection and through radiation. This interplay has been studied in the tropics at the large scale (e.g., Pierrehumbert, 1995) but is less well understood in the lower tropical troposphere, where humidity is less well quantified from observations (Nehrir et al., 2017; Stevens et al., 2017). One way to fill this gap are airborne measurements taken during dedicated field campaigns. In this study, we use airborne lidar measurements from two field campaigns in the northern tropical Atlantic to analyse the vertical structure and the spatial variability of water vapor and clouds and their representation in simulations with resolution from hecto- to kilometer scale.

Water vapor has multiple roles in the atmosphere and is closely connected to cloudiness: The boundary layer humidity sets the potential for deep convection and determines cloud amount (e.g., Keil et al., 2008; Vial et al., 2017). As the vertically integrated amount of water vapor approaches its saturation value over the tropical oceans, precipitation sets in and the amount of precipitation in deep convective regions correlates well with the decrease in subsaturation in the column ~~(Bretherton et al., 2004; Holloway and Neelin, 2009; Nuijens et al., 2009).~~ (Bretherton et al., 2004; Holloway and Neelin, 2009). The same relation is found to hold within the dry season in the shallow convective regime (Nuijens et al., 2009). On a process level, the vertical distribution of moisture determines the amount and distribution of radiative cooling and can thereby drive large-scale and meso-scale circulations (e.g., Pierrehumbert, 1995; Muller and Bony, 2015; Naumann et al., 2019). Also, the humidity of cloud-free air in the vicinity of a cloud determines the strength of dilution of in-cloud water by entrainment. The strength of this dilution is a long-standing problem in convective parameterizations, a key ingredient of the thermostat and the iris hypothesis, and a popular tuning parameter (Ramanathan and Collins, 1991; Mauritsen et al., 2012; Mauritsen and Stevens, 2015).

The vertical distribution of moisture and small-scale phenomena such as the dilution of clouds by entrainment are posing challenges to both modelling and observations. The WALES (WAter vapor Lidar Experiment in Space) lidar is capable of profiling moisture ~~, aerosol, and clouds~~ and aerosols, and detecting cloud tops simultaneously with high accuracy and spatial resolution (Wirth et al., 2009). High resolution in vertical profiles is of particular importance in the tropics since sharp moisture gradients at the trade inversion influence radiation locally (Stevens et al., 2017). Installed on an aircraft, measurements with WALES can be undertaken in regions of particular interest. In December 2013 and in August 2016 the NARVAL (Next-generation Aircraft Remote-sensing for VALidation) campaigns were the first tropical experiments in which an airborne water vapor lidar participated (Stevens et al., 2019b). For the two campaigns the German research aircraft HALO (High-Altitude Long-Range) sampled the western tropical Atlantic ~~East~~ east of Barbados to investigate the interactions between shallow moist convection, moisture distribution, and radiative effects with a state-of-the-art suite of remote sensing instruments and dropsondes.

The close coupling between clouds and water vapor and the capabilities of lidar measurements in the trade wind regime motivate the guiding questions of this study: What is the vertical structure and the spatial variability of water vapor in the trades? How does cloudiness covary with water vapor and are models able to represent the observed relationship correctly?

In numerical weather prediction, storm resolving model (SRM) simulations with kilometer-scale grid spacing are common and evaluated frequently (e.g., Bauer et al., 2015). Aiming to better resolve convection with higher resolution, traditional idealized large-eddy model (LEM) simulations lack the ability to represent the mesoscale and large-scale variability of observed cloud fields (Nuijens and Siebesma, 2019). LEM simulations with hectometer scale grid spacing are now becoming available on large domains with realistic boundary conditions (Heinze et al., 2017; Stevens et al., 2019b). These LEM simulations with realistic and varying large-scale states include the interaction with the large-scale circulation and at the same time the subgrid-scale flow is better constrained than in coarse resolution simulations. Although simulations with hectometer grid spacing still do not have a grid spacing fine enough to represent details of shallow convection, even kilometer-scale simulations are found to reproduce many features, such as the daily cycle in cloud amount and precipitation, better than climate models with convective parameterization (Stevens et al., 2019a; Vial et al., 2019). It is an open question whether hecto- and kilometer-scale simulations with realistic and varying large-scale states are able to represent water vapor variability and its co-variation with clouds in the trades and whether this ability depends on resolution.

In model simulations convection, due to its stochastic nature, is not expected to trigger in the exact same position and with the exact same timing as in reality. Therefore comparisons between observations from line-shaped research flights and models, where the comparison is based on co-location of the two in space and time, are often of limited use. To bypass the issue of co-location other means of comparison are needed. We propose to compare model and observations in moisture space, i.e., we sort water vapor profiles from the driest to the wettest profile, to identify differences in the vertical structure of water vapor and its change in moisture space. The depiction of humidity in moisture space is inspired by Bretherton et al. (2005), who compare model results as a function of column-relative-humidity to illustrate the mechanisms of convective self-aggregation in radiative convective equilibrium. In observations this technique has been first used by Schulz and Stevens (2018). With a comparison of observations and simulations in moisture space we avoid relying on co-location but retain the ability to quantify variability at high spatial resolution.

This paper is organized as follows: Section 2 describes the observations and model simulations used in this paper. In Sect. 3 we focus on the case study of a research flight on 11. December 2013, which is a case of typical shallow trade wind convection and is also used to explain our methodology in detail. In Sect. 4 we generalize the results of the case study by applying the same methodology to a set of research flights that allow us to analyse the seasonality of the water vapor structure in the trades. Conclusions are given in Sect. 5.

2 Observations and ~~Model Simulations~~model simulations

2.1 NARVAL winter and summer campaign

Two NARVAL field studies took place over the tropical Atlantic ocean east of Barbados (Stevens et al., 2019b). The first part of the field study counts eight research flights between 10 - 20 December 2013 and the second part ten flights between 8 - 30 August 2016. The details of the NARVAL field studies, such as the flight strategy and the ~~description~~ instrumentation of the HALO aircraft are described by Stevens et al. (2019b) and Konow et al. (2019). Not all data are to the same degree useful for this analysis, as some of the long flights (e.g., the transit flights between Germany and Barbados) are not contained in the modeling domain of the LES (see Sect. 2.3) and some other days have not been chosen to be modeled with LES. For the purpose of this study, we limit the available lidar and microwave radiometer data by the criterium of being included in our smallest modeling domain (see Section 2.3). The time and domain constraints are given in Table 1.

Basic differences between the winter and the summer trades appear in the cloud layer moisture and thickness (Table 1). While the winter situations are characterised by similar and undisturbed trade wind conditions, the summer flights encountered a significant layer of Saharan dust on August 12 and 19, the flight on August 22 was close to the intertropical convergence zone (ITCZ), and the flight on August 24 was close to the tropical storm Garcon (Gutleben et al., 2019).

Table 1. Specification of flight domains used in this study.

t in UTC	domain	N	p in %	q_c in g kg^{-1}	h_c in km
NARVAL 1					
11. Dec 2013 16 - 21	10.0 - 16.5 N, 58.0 - 55.0 W	531	34.2	4.0	3.0
12. Dec 2013 14-15, 19-20	14.0 - 16.5 N, 56.5 - 48.5 W	526	86.5	4.0	2.8
14. Dec 2013 14-15, 19-20	13.9 - 16.5 N, 57.2 - 48.5 W	296	48.9	4.0	2.5
15. Dec 2013 16 - 21	12.0 - 16.5 N, 57.5 - 48.5 W	668	72.8	4.0	2.7
20. Dec 2013 17 - 18	13.3 - 16.5 N, 56.0 - 51.6 W	168	70.3	4.0	3.0
NARVAL 2					
12. Aug 2016 13 - 19	9.5 - 14.0 N, 55.0 - 52.0 W	1317	69.0	6.0	1.9
19. Aug 2016 13 - 17, 20	13.5 - 16.0 N, 57.0 - 48.0 W	1115	85.4	8.0	2.6
22. Aug 2016 14-15, 20-21	10.0 - 12.8 N, 58.6 - 51.0 W	279	55.9	8.0	1.8
24. Aug 2016 13 - 16	13.0 - 14.5 N, 56.5 - 44.0 W	405	51.3	9.0	1.6

t: time period of analyzed flight, N: number of valid lidar profiles, p: percentage of valid profiles, q_c: water vapor mixing ratio threshold for detecting a cloud top with WALES, h_c: maximum shallow cloud top altitude

2.2 WALES ~~Lidar~~ lidar and HAMP ~~Radiometer~~radiometer

The differential absorption lidar WALES is installed pointing downwards on the HALO aircraft, measuring water vapor profiles throughout the tropical troposphere with three on-line laser wavelength positions in the near-infrared situated on three water vapor absorption lines of cascading strength (Wirth et al., 2009; Kiemle et al., 2017; Gutleben et al., 2019). The weakest line, specially selected for the tropics, permits accurate profiling of very moist layers below the ~~trade inversion~~ inversion that tops the cloud layer in the trades while the stronger two lines provide reliable data of the moisture jump across the inversion and the dry regions above. Backscatter from aerosol and clouds, corrected for extinction by aerosol, is simultaneously measured by a high spectral resolution lidar (HSRL) at 532 nm with a temporal resolution of 1 s, corresponding to a spatial horizontal resolution along the flight route of 210 m given the typical aircraft speed of 210 $\mathrm{m\,s^{-1}}$ during the summer campaign and a horizontal resolution of 240 m given an aircraft speed of 240 $\mathrm{m\,s^{-1}}$ during the winter campaign. Flight speed was higher in winter due to a higher average flight altitude. To achieve an acceptable measurement precision of typically 10 % in the cloud layer and above, the water vapor profiles are aggregated across 12 s or about 2.5 km in the summer campaign and 2.9 km in the winter campaign. The vertical resolution is about 250 m for water vapor and 15 m for backscatter. Water clouds quickly attenuate the lidar signal such that valid data are only obtained above cloud top~~clearly defined by~~ , which is visible in the backscatter signals (Fig. 1 a). Full profiles are obtained wherever the cloud gaps are larger than 2.5 km. Due to a methodical constraint, water vapor lidar data below 200 m is not available. The dropsonde profiles show that humidity is relatively constant with height within this layer, which agrees with the assumption of a well-mixed subcloud layer. To calculate the WVP, we therefore extend the measurements at 200 m down to the surface.

Since our focus is the cloud layer moisture variability we only use those lidar profiles where more than half of the data points below the maximum cloud top height, which is defined by q_c in Table 1, are valid. For example, on 11 December 2013, the cloud layer top height is 3.0 km, and only in 34 % of all lidar profiles are more than half of the data points ~~are~~ valid below this height (Fig. 1 a). The rest is unavailable due to clouds or laser adjustment phases. We consequently use only one third of all profiles of this flight (Fig. 1 b). This subset still contains small gaps mainly due to clouds which we fill with the saturation value by assuming saturation wherever the HSRL backscatter coefficient is $> 10\,(\mathrm{Mm\,sr})^{-1}$ which to sufficient approximation defines a water cloud (Kiemle et al., 2017). We deviate from this threshold only in two cases where the clouds are particularly small (on 12 August 2016 we use $5\,(\mathrm{Mm\,sr})^{-1}$ to compensate for the signal dilution) or large (on 24 August 2016 we use $15\,(\mathrm{Mm\,sr})^{-1}$). We fill the remaining gaps with the moisture of the nearest neighbor profile in the horizontal and call this gap-free result a minimum estimate (WALES$_{\mathrm{min}}$; Fig. 1 c). In a maximum estimate (WALES$_{\mathrm{max}}$) we additionally fill all original cloud shadows down to the lifting condensation level (LCL), i.e., missing data below lidar-detected clouds, with the saturation value. ~~We~~ To find the LCL, we use the lidar signals from thin boundary layer clouds ~~to find the LCL and temperature profiles from close dropsondes to determine the saturation humidity profile~~as well as dropsonde profiles and auxiliary lidar information such as aerosol and water vapor gradients at the top of the mixed layer. The saturation humidity profiles are calculated from the temperature profiles of nearby dropsondes. Since the thickness of the cloud cannot be determined by the lidar and also lower clouds may exist above the LCL, the maximum estimate gives an upper bound on cloudiness and water vapor path (WVP, defined as the vertically integrated specific humidity without contributions from liquid or ice). Likewise the minimum estimate provides a lower bound on cloudiness and WVP. Consequently, the difference between the minimum and the maximum estimates characterises to a satisfying extent the uncertainty of our attempt to quantify the lidar moisture distribution within and below the clouds while aiming to obtain a gap-free data curtain needed for the model comparisons. The difference in WVP between WALES$_{\mathrm{max}}$ and WALES$_{\mathrm{min}}$ is at maximum 5 % (Fig. 1 d). We will show later that the uncertainty in the measured humidity estimate is small compared to the difference between model and observation (see Section 3.2). To obtain the cloud fraction~~in the 12-s grid~~, we apply the abovementioned HSRL backscatter coefficient threshold for water clouds onto the 1-s lidar backscatter curtains, ~~using~~ use a similar min/max assumption to account for measurement and methodical uncertainties and aggregate it into a 12-s grid along the flight direction.

To understand which part of the moisture space the WALES lidar misses in cloudy environments, we additionally make use of the HAMP (HALO Microwave Package) radiometers, whose data is available for NARVAL 1 (Jacob et al., 2019a) and NARVAL 2 (Jacob et al., 2019b). The nadir-viewing HAMP microwave radiometers lack vertical profile information but measure the WVP with 1 s (that is 210 m or 240 m) resolution along the HALO flight track ~~(Jacob et al., 2019c). Their~~ also in the presence of shallow clouds (Jacob et al., 2019c). The co-alignment of HAMP with the lidar field of view was checked by comparing the radiometer liquid water path with the lidar cloud backscatter signals, both available at 1 s resolution. The radiometer signals are interrupted by calibration events. Comparisons with the co-located lidar WVP reveal that those events are independent from the ambient humidity conditions. The radiometer WVP distributions are consequently not biased, except

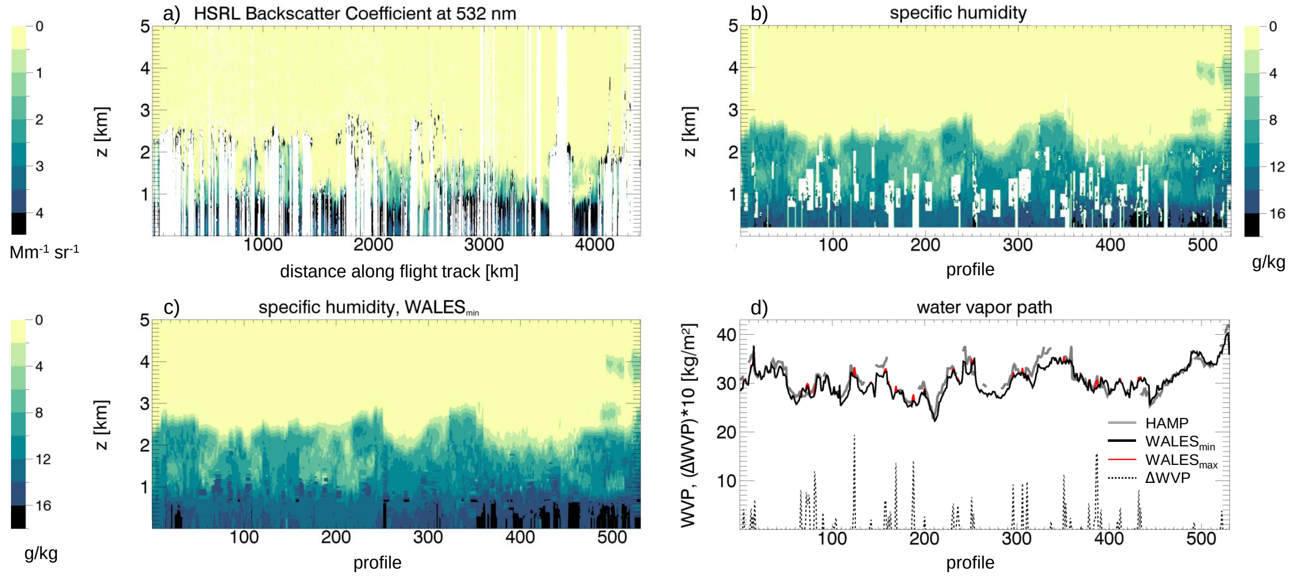

Figure 1. Lidar profiles of the flight on 11. December 2013: (a) atmospheric backscatter for the full flight, ~~and~~ (b) specific humidity with lidar gaps, (c) specific humidity ~~for~~ of WALES$_{max}$, and (d) WVP and the difference in WVP between WALES$_{max}$ and WALES$_{min}$, ΔWVP. (b-d) show those 531 profiles where more than 50 % of the lidar data within the cloud layer and below are available. The remaining gaps in the original data set in (b) are filled by assuming saturation in clouds, and by nearest neighbor values elsewhere, resulting in a gap-filled representation in (c) and (d). See text for details. Note that the aspect ratio is 1:500 in (a) and 1:150 in (b) and (c).

for a slight underrepresentation of the moistest scenes due to ~~saturation~~ signal attenuation which concerns less than 1.5 % of all WVP data.

2.3 ICON

Simulations are run with ICON (Icosahedral non-hydrostatic model; Zängl et al., 2015) with four different grid spacings between 2.5 km and 300 m and with two different model versions: ICON-SRM and ICON-LEM. The ICON-SRM was run with 75 vertical levels and with 2.5 km and 1.25 km nominal horizontal grid spacing. Details of the simulations are described by Klocke et al. (2017). The ICON-LEM (Dipankar et al., 2015; Heinze et al., 2017) was run with 150 vertical levels and with 600 m and 300 m nominal horizontal grid spacing. Details of the simulations are described by Stevens et al. (2019b). The effective resolution is estimated to be a factor of six to ten larger than the nominal grid spacing (Hansen, 2020). In all simulations the parameterizations for shallow and deep convection, gravity wave drag and subgrid-scale orography are switched off. The parameterizations for turbulence and microphysics differ between the SRM and the LEM. In addition, the SRM simulations apply a cloud cover parameterization while the LEM simulations use a binary approach. For this study, we set the LEM cloud fraction to 1 if the liquid water content in a grid box is non-zero, and 0 otherwise.

The SRM runs with the coarsest grid spacing of 2.5 km cover the largest domain including the entire tropical Atlantic (10.0 S - 20.0 N, 68.0 W - 15.0 E). The simulated domain size decreases with increasing resolution, so that the LEM run with the finest grid spacing of 300 m has the smallest domain, which still covers an area of 800 km × 1600 km in the western part of the Atlantic (8.0 - 16.5 N, 60.0 - 43.5 W). For the purpose of this study, we do not analyze model output from the full simulation domains of ICON at different resolutions but instead limit the domain analyzed to rectangles around those parts of the flight paths that took place within the smallest simulated domain. Because the flight paths and time periods differ from day to day, the analyzed domains and time periods also differ as given in Table 1. We analyse all model output in these domains instead of selecting profiles along the flight tracks because convection is not expected to trigger at the exact same location and time in simulations as it does in reality. Using the domain output is consistent with the statistical rather than spatial-temporal approach of this analysis and promotes the robustness of the results.

Initial and boundary conditions for the ICON-SRM 2.5 km simulations are taken from the European Centre for Medium-Range Weather Forecast (ECMWF) reanalysis and vary in time except for the SST, which is fixed for each simulation day. The simulations apply a one-way nesting of higher resolution simulations in low resolution simulations. The ICON-SRM simulations with 2.5 km horizontal grid spacing apply an online refinement to 1.25 km via nesting in the eastern part of the domain and start at 0 UTC for each day of December 2013 and August 2016. They are run forward in time for 36 hours. ICON-LEM simulations are initialized and nudged at the lateral boundaries from ICON-SRM and start at 9 UTC for selected days to match the flight operations of the NARVAL campaign. They are run forward in time for 27 hours. Simulations are analyzed from hourly model output starting earliest at 13 UTC (see Table 1) so that a sufficient spinup period is taken into account.

3 Case study: Covariation of clouds and moisture

In this section, we use one day of the first NARVAL campaign, 11. December 2013, ~~to explain our methodology of comparing lidar measurements with model results in moisture spaceand to illustrate~~ for a detailed case study. The aim of the case study is to introduce the central method of this study: the concept of a stretched moisture space. The stretched moisture space is obtained by selective subsampling of the model results and thereby allows for a fair comparison between lidar data and model results. The case study also illustrates some prominent features of covariation of clouds and moisture~~(Fig. 1 and Fig. 2).~~ , before aggregated seasonal composites enable us to generalize the results to different regimes of water vapor structure in the trades in Section 4.

3.1 Synoptic ~~Situation~~ situation and ~~Flight~~flight

We choose the 11. December 2013 for a detailed case study for two reasons: First, a regular meander flight pattern allows us to sample a well-defined region thoroughly, which aids a comparisons with simulations (Fig. 2). Second, the conditions seem preferential to sample the humidity space because the flight area includes typical shallow convection over most of the area but also approaches deeper convection with higher humidity towards the south.

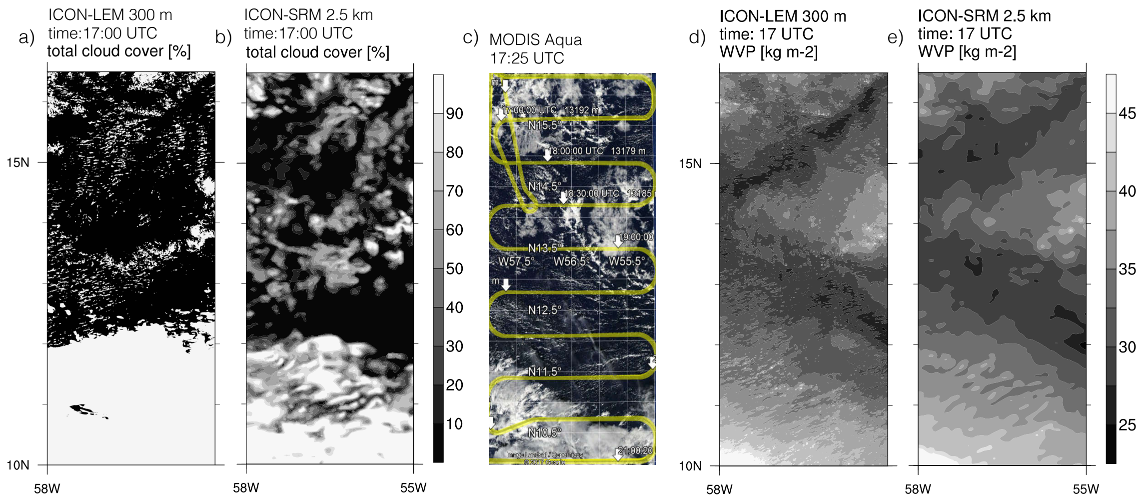

Figure 2. Cloud cover and water vapor path (WVP) in the flight domain on 11 December 2013. Cloud cover at 17:00 UTC from (a) ICON-LEM 300 m and (b) ICON-SRM 2.5 km; (c) MODIS Aqua corrected reflectance at 17:25 UTC overlaid with the flight path which was flown from north to south; WVP at 17:00 UTC from (d) ICON-LEM 300 m and (e) ICON-SRM 2.5 km.

The modeled cloud structures have similarities with the observed reflectance from MODIS showing organized structures of shallow clouds in the northern three quarters of the domain (Fig. 2). With a grid spacing of 2.5 km these shallow clouds have a too broad structure compared to observations. With higher resolution the cloud structures, not surprisingly, become finer but at 300 m grid spacing the model misses some stratiform outflow from shallow cumulus giving the shallow convective cloud field a less organized appearance than in satellite observations. In both simulations and in the satellite view the southern quarter of the domain is dominated by a cirrus shield originating from deep convection just south of the domain. This cirrus shield is reaching further north in the model than in the satellite observations. Because the deep convective system moves towards the south west with time and the flight itinerary is following the pattern from north to south, the lidar observations onboard the aircraft catch only a small amount of this regime (see Sect. 3.2).

The field of WVP shows more small-scale structure at 300 m grid spacing than with 2.5 km but changes less with resolution than the cloud cover does. All simulation show an increase of WVP from north to south and a c-shape of low WVP in the northern and central section of the domain. This c-shape in the modeled WVP can be surmised in a reduced presence of clouds in the satellite view but is less well reflected in the modeled cloud cover.

Averaging the results of the ICON-LEM 300 m simulation on squares of different side length, we analyse how the standard deviation of the water vapor mixing ratio, q_v, ~~changes with effective resolution~~ depends on the considered scales (Fig. 3). The analysis combines spatial and temporal variability but the contribution from spatial variability is dominating (not shown). Coarse graining the 300-m LEM results to 2.5 km does not change the standard deviation considerably. The relative contribution of small scales between 300 m and 2.5 km to the standard deviation of q_v is largest near cloud base and in the subcloud layer but generally well below 10 %. Even for a side length of 20 km the relative differences to the native grid spacing of

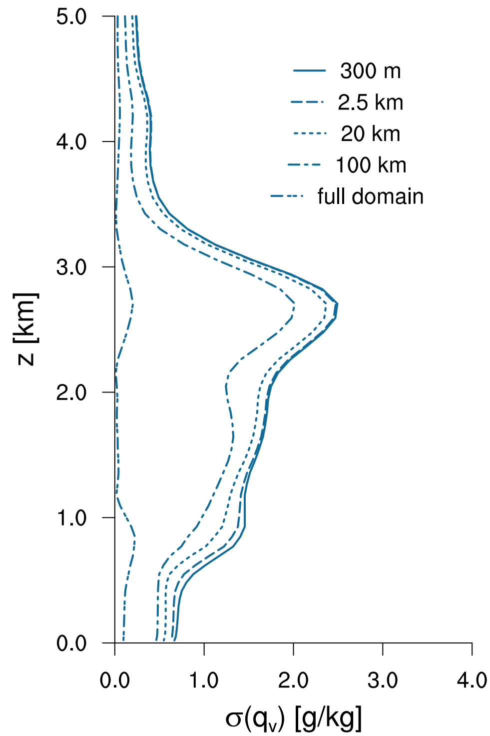

Figure 3. Contribution of different scales to the standard deviation of q_v on 11 December 2013 from ICON-LEM 300 m. Domain and temporal coverage are given in Table 1. Simulations with a grid spacing of 300 m have been coarsend to squares with side lengths of 2.5 km, 20 km, 100 km, and "full domain", which corresponds to a side length of about 400 km. Both spatial and temporal variability contribute to the standard deviation except for the "full domain", which only shows temporal variability. The cloud layer ranges from cloud base at $z = 0.5$ km to the highest cloud tops at $z = 3.0$ km.

300 m are maximum near cloud base (30 %) but are considerably smaller throughout the cloud layer and above (< 10 %). Because the differences are small, for the remainder of this analysis we show model results and observational data at their native scale (from 300 m to 2.5 km), which aids a direct evaluation of what a simulation is able to catch without artificially reducing information by averaging.

3.2 Spanning the ~~Moisture Space~~moisture space

Because of its stochastic nature convection is not expected to trigger at the exact same location and time in simulations as it does in reality. To bypass the issue of co-location, we sort water vapor profiles from the driest to the wettest profile and compare simulations and observations in moisture space (Bretherton et al., 2005; Schulz and Stevens, 2018). Comparing simulation results with data from HAMP, this procedure is straight forward because the HAMP dataset samples the whole domain well. WALES on the other hand is rapidly attenuated in clouds and saturated in the wettest profiles so that a fair comparison to simulations needs to take into account information on which situations WALES is not able to observe. We therefore use HAMP to span the moisture space, to quantify what WALES misses, in particular in the wet regions, and to construct a "stretched moisture space" that enables a fair comparison between WALES and ICON. This method works well during NARVAL because flight patterns were fixed before takeoff and hence measurements along the flight path represent a random sample of the encountered cloud regime. The validity of this method quickly reaches its limits if flight paths are adjusted to preferentially sample a feature of special interest – a trade-off to be aware of for future flight planning (e.g., in view of EUREC^{4}A, Bony et al., 2017).

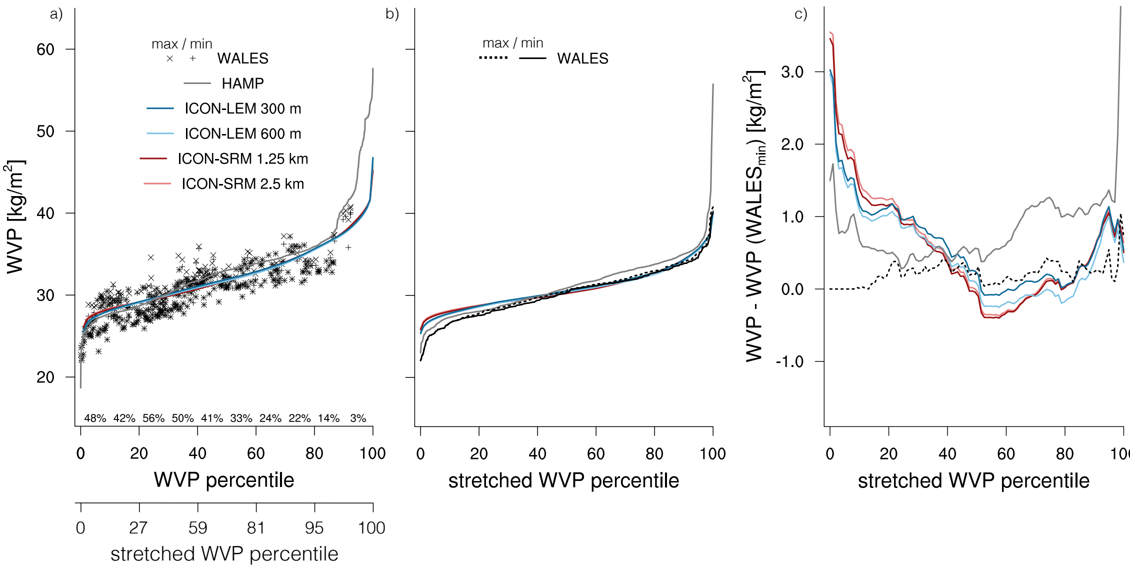

Figure 4. Water vapor path (WVP) on 11th of December 2013 in WVP space. (a) ICON simulations and HAMP observations of WVP are sorted by each one's WVP values. WALES data is plotted as co-located with HAMP. Percentages above the x axis tell how many valid WALES measurements have been obtained in each 10 % interval of HAMP's WVP space. (b) ICON results and HAMP data is randomly selected according to those percentages in each 10 % interval resulting in a stretched WVP space, which is also shown as an additional x axis in (a). In (b) WALES data is sorted by its own WVP instead of being co-located with HAMP. (c) as in (b) but for the difference to WALES$_{min}$. Further details are discussed in the text.

We proceed as follows: All available WVP values from HAMP and the ICON simulations at different resolution are sorted from the lowest to the highest value (Fig. 4 a). This representation corresponds to the cumulative distribution function of ~~WVP, which is rotated by 90°~~ the WVP but with swapped x- and y-axis compared to the common depiction. Since WALES and HAMP measure the same location at the same time, a co-location between those two instruments is eligible. For 11th of December 2013 WALES$_{min}$ values scatter around HAMP values with a standard deviation of 1.48 kg/m^2 (1.62 kg/m^2 for WALES$_{max}$), which is consistent with Jacob et al. (2019c). Because WALES measurements attenuate quickly in clouds and for high WVP (Sect. 2), data gaps are not randomly distributed in moisture space but instead preferentially occur where WVP is high: of the driest 10 % of HAMP measurements 48 % have a corresponding measurement from WALES, while for the moistest 10 % of HAMP measurements only 3 % have a corresponding measurement in WALES. To account for this biased sampling of WALES, we randomly select model results and HAMP according to these percentages of WALES counterparts in each 10 % interval. Then we sort all WALES WVP by its increasing value. The resulting new moisture space of all valid WALES data points and those subsampled from ICON and HAMP is effectively stretched in its drier part and compressed in the moister part (Fig. 4 b and lower x axis in Fig. 4 a). We call this new moisture space the stretched WVP space according to WALES or, in short, the "stretched moisture space". This stretched moisture space enables a fair comparison between WALES and ICON.

In stretched moisture space, the distribution of WVP from ICON simulation results, and WALES and HAMP measurements overall agree well (Fig. 4 b, c). The differences between the three observational estimates, HAMP, $\text{WALES}_{\text{min}}$, and $\text{WALES}_{\text{max}}$ are small with a median of absolute difference around 0.6 kg/m^2 ($\text{WALES}_{\text{min}}$ vs. HAMP: 0.60 kg/m^2, $\text{WALES}_{\text{max}}$ vs. HAMP: 0.56 kg/m^2). The differences in the distributions of WVP between simulations at different grid spacing are much smaller. This possibly reflects the nested modeling approach, which ensures consistent initial and boundary conditions and where domains are nudged with a time scale of 3 h, to ensure that they do not deviate too much in the two-way setup. However, the differences in cloud fraction are considerably larger (see Sect. 3.3), which indicates that the effect of grid spacing in the range of hecto- to kilometer scale is small for the distribution of the WVP.

The small intra-observational and intra-model differences enable a meaningful interpretation of the difference between model and observation. Compared to observations the modeled variability of WVP is too small. The driest model areas are too wet, while the wettest model areas agree well with WALES (Fig. 4 ~~b~~c). This underestimation of the variability in WVP can be attributed to too low variability of moisture in the cloud layer (see Sect. 3.3). If WVP is not subsampled for valid WALES profiles, there is also a dry model bias for very wet profiles as compared to HAMP (Fig. 4 a). Here, the wettest 15 % of HAMP's moisture space seem to be not well represented in the model. Two factors are expected to contribute to this deviation: On 11 December 2013 there is a little change in the flight track near 11 N 56 W. This was made to try to fly over the deepest turret of the towering convection and try to drop a sonde through this (~~personal communication,~~ Bjorn Stevens, personal communication, 2019). Hence this flight segment is purposely biased to the moistest cell and may contribute to differences in the moist part of the space of Fig. 4 a. Also, extending the analysed model domain to south of 10 N, decreases this bias which suggests that the deep convective system on 11 December 2013 is consistently placed too far south in all four simulations (not shown). Because both the deepest turret of the towering convection and in general the moistest profiles towards the south of the domain contain less valid WALES samples than the drier profiles, this feature is much less visible in stretched moisture space and is therefore less important for the remainder of this analysis.

3.3 Vertical ~~Distribution~~ distribution of ~~Water Vapor~~ water vapor and ~~Cloud Fraction~~cloud fraction

With the framework of the stretched moisture space, we can now also analyse the vertical structure of water vapor and cloud fraction by comparing valid WALES profiles with ICON profiles that are subsampled according to percentages of the WALES counterpart. The analysis therefore does not represent the real space as an omniscient observer would see it but only that part that WALES is equipped to measure.

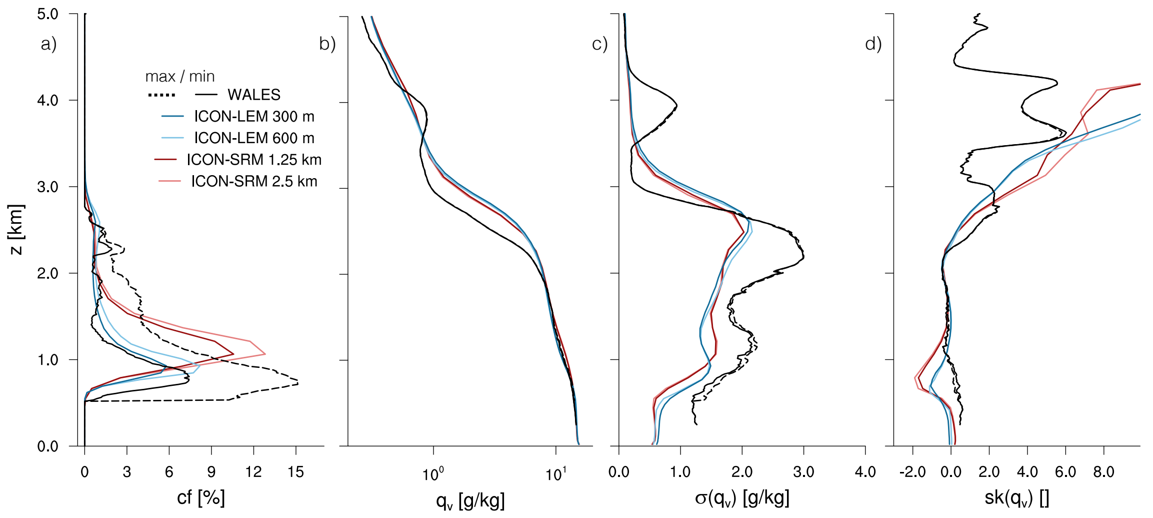

Figure 5. Profiles of (a) cloud fraction, (b) mean water vapor, q_v, and its (c) standard deviation and (d) skewness for the 11. December 2013 in stretched moisture space as defined in Fig. 4.

The mean water vapor mixing ratio compares well between WALES and ICON (Fig. 5 b). As for the integrated quantity WVP, also in the vertical structure of q_v there is no dependence on grid spacing. Compared to WALES the inversion is too high in the model, a feature that is common to all analysed days in December 2013. Both the observed and the modeled heights of the inversion increase with increasing WVP but this increase is less pronounced in the simulations (Fig. 6 b). For the dry profiles the modeled inversion is also less steep, which implies a less concentrated radiative cooling in the simulations at the cloud layer top with possible implications for mesoscale circulations (Naumann et al., 2019).

The higher moments of the water vapor distribution do not agree as well as the mean but still capture the main features and the right magnitude. The two maxima of the standard deviation of q_v in the cloud layer are well captured but are underestimated by the model compared to the observations (Fig. 5 c). This is also evident from the change in bias with increasing WVP: in the cloud layer the driest profiles tend to be too moist in the model (Fig. 6 c).

The skewness, which is defined as the ratio of the third central moment of the distribution to the $3/2$ power of the variance, is reasonably well represented from the middle of the cloud layer up to the cloud layer top (Fig. 5 d). Near cloud base the model simulates a negative skewness, that is, few very dry locations associated with cloud free regions, while the observations indicate slightly positive values, that is, few very moist locations. This difference in sign between model and observations is also found on the 14. and 15. December 2013 but not on the other days (not shown). Above cloud top between 4 km and 7 km the modeled skewness is very large, which is caused by a single deep convective cell near the south-western corner of the domain that dominates the skewness but has not been sampled ~~during the campaign~~ by the lidar and is therefore not represented in the observations.

While these properties are characteristic also for other flight days of the NARVAL campaign, a feature that is special to the observations on 11. December 2013 is a secondary maximum at 4 km height (Fig. 5 b). This secondary maximum is evident

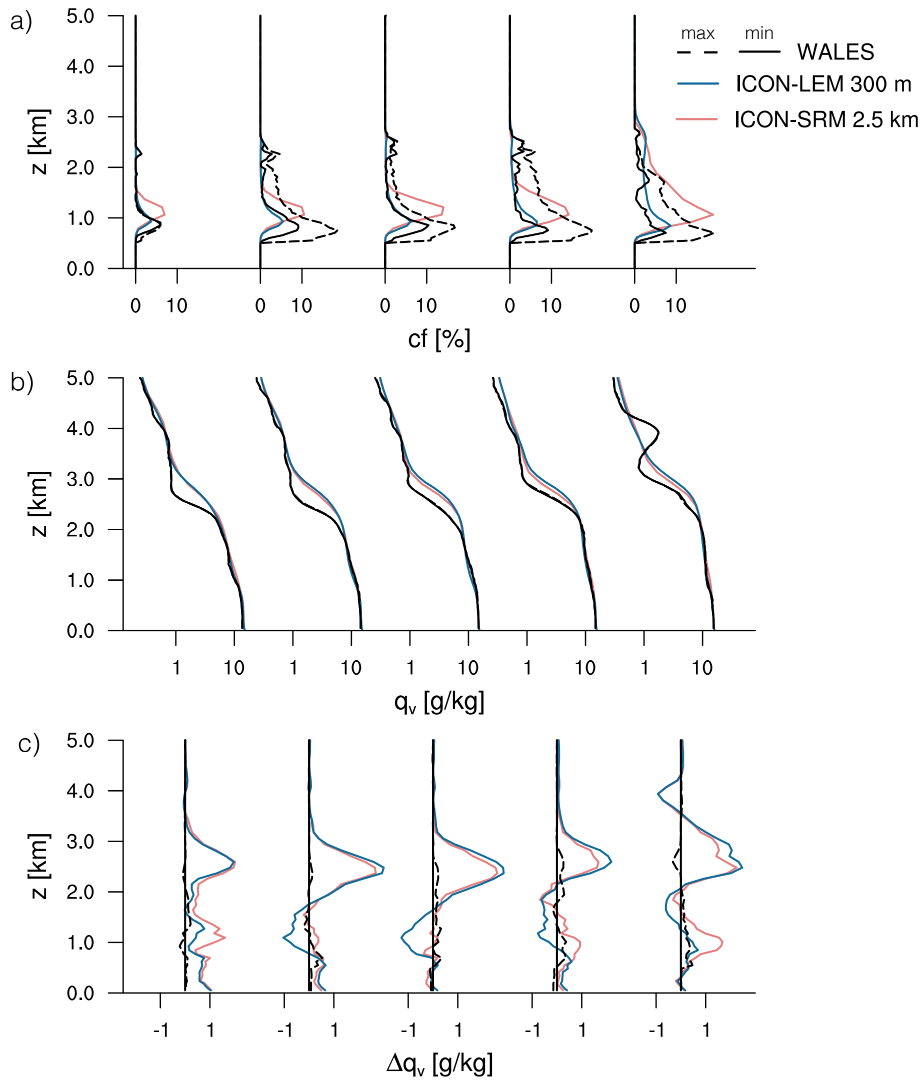

Figure 6. Profiles of (a) cloud fraction, (b) water vapor and (c) the difference of water vapor to the $WALES_{min}$ estimate for 11 December 2013. Each profile shows the mean for a 20-percentile range of WVP in stretched moisture space from driest profiles on the left to moistest profiles on the right (see Fig. 4). To retain fluctuations due to a limited number of samples, as many profiles as are available from WALES have been randomly subsampled from ICON results (here 531 samples, see Table 1). At a given height level $WALES_{max}$ can be lower than $WALES_{min}$ because the sorting of profiles is done according to the column integrated WVP separately for the minimum and the maximum estimate.

only in the moistest profiles (Fig. 6 b), manifests in the southern part of the domain towards the end of the flight (Fig. 1) and is caused by a moist outflow from convectively more active regions. This feature is also reflected in higher values of standard deviation and skewness but is absent in all three moments in the model, which misses the moist outflow (Fig. 5 c, d).

For the mean cloud fraction, both uncertainties from observations and sensitivity to model resolution are larger than for q_v (Fig. 5 a). Typical cloud sizes obtained from the lidar are around 500 m (Gutleben et al., 2019) and hence on the order of the grid spacing of the simulations. Because the contribution to overall cloud fraction scales with the size of the clouds, we do not expect the contribution of these small clouds to dominate the overall cloud fraction. From WALES the uncertainty in maximum cloud fraction is a factor of two (between 7.4 % for $\text{WALES}_{\text{min}}$ and 15.2 % for $\text{WALES}_{\text{max}}$) but the vertical structure is similar with a clear maximum in cloud fraction near cloud base and few shallow clouds deepening up to 3 km. This structure is also represented well by the simulations, except that the cloud fraction maximum near cloud base is placed too high. We suspect that this upward shift in cloud fraction maximum is linked to the resolution because the shift is stronger for the SRM than the LEM simulations. Another hypothesis, which has recently been developed by Jacob et al. (2020), proposes that slight differences in the autoconversion parameterization in the SRM and LEM might cause differences in the cloud's vertical extent. A further resolution dependent feature is the value of the maximum cloud fraction, which decreases substantially by a factor of two between 12.8 % (ICON-SRM 2.5 km) and 5.8 % (ICON-LEM 300 m) but is still ~~included in~~ close to the range of uncertainty given by the observations. Hohenegger et al. (2019) find similar dependencies of cloud fraction on grid spacing between 2.5 km and 80 km and hypothesize that if horizontal resolution is not sufficient for proper mixing, the boundary layer grows and clouds form higher at colder temperatures leading also to more cloudiness. The decrease in cloud fraction between the simulations with 600 m and 300 m grid spacing is still substantial and not converged, which is in agreement with idealized modelling studies showing that LEM underestimates cloud fraction when the grid spacing becomes as fine as 50 m (Vogel et al., 2019).

With increasing WVP the clouds deepen from very shallow cloud tops around 1 km up to cloud tops around 3 km both in the simulations and in observations (Fig. 6 a). Whether the maximum cloud fraction also increases with increasing WVP is not clear: for $\text{WALES}_{\text{min}}$ the maximum cloud fraction stays about constant while for $\text{WALES}_{\text{max}}$ the maximum cloud fraction increases with increasing WVP. Cloud fraction from the LEM simulations agrees well with the $\text{WALES}_{\text{min}}$ estimate but in the SRM simulations the maximum cloud fraction increases similar to the $\text{WALES}_{\text{max}}$ estimate. For features other than the height of the maximum cloud fraction, which is shifted upward in particular in the SRM simulation, it therefore remains unclear for this case study whether the modeled cloud fraction improves with resolution or not. For the season of August 2016 a better representation of cloud fraction with higher resolution becomes apparent and will be discussed in the next section.

4 Seasonal ~~Composites~~composites

In this section we generalize the results of the case study by applying the same methodology to composites of several research flights that allow us to analyse different regimes of the water vapor structure in the trades. We combine five research flights in December 2013 to one composite case and four research flights from August 2016 for another composite case (Table 1), both of which represent different seasons in the trades. As for the case study in the previous section, we subsample all model results according to the percentages available from WALES in each 10 % bin of WVP ~~to enable a fair comparison between model results and observational data. The~~ for each flight individually. After the subsampling we concatenate the individual flights to obtain the seasonal composite. The composite is thus weighted by the number of valid profiles per flight (which vary from flight to flight; Table 1). The analysis in this section is ~~therefore~~ discussed in the resulting stretched moisture space.

4.1 Stretched ~~Moisture Space~~moisture space

Boreal winter in the northern trades near Barbados is generally characterized by a drier free troposphere compared to boreal summer, which is characterized by more frequent disturbances, a closer proximity of deep convection associated with the ITCZ, and a moister free troposphere (e.g., Stevens et al., 2017). All research flights in December 2013 took place in a period of undisturbed shallow convection (Vial et al., 2019). To analyse whether the chosen research flights characterize a meaningful regime of water vapor structure, we test their representativeness by extending the analyzed period to the ambient days (10. to 21. December 2013) and choosing the mean borders of their domains (12.7 - 16.5 N, 57.0 - 50.4 W). For December 2013 the research flights represent the extended period very well (Fig. 7 a). For August 2016, we extend the period and domain in the same way except for the southern border (11. to 25. August 2016; 13.0 - 14.3 N, 56.8 - 48.8 W). Compared to the mean border, the southern border is shifted 1.5° north to avoid inclusion of deep convection from the ITCZ on a few days, where it reaches further north. In August 2016, the extended period is several kg/m^2 moister than the flight period and domain. This difference can be explained by two factors: On several of the flights in August 2016 dry sectors were sought out purposely biasing the flight periods compared to the extended period (Bjorn Stevens, personal communication, 2019). This illustrates the problem of flying toward specific features, rather than fixing a flight pattern to sample a region evenly (see also Sect. 3.2; Jacob et al., 2019c). In addition, on 20 - 22 August 2016 the tropical cyclone Fiona runs by north of the domain and brings some very moist air into the domain behind it on 23 August 2016 contributing to a moister extended period. Because the difference between the moist August flights and the dry December flights is considerably larger than the difference between the flight periods and their extended periods, both composite cases can be seen as representative for different regimes. A good representation of the NARVAL flights for their respective season is also found by a comparison with a 8 year long time series at the Barbados cloud observatory in terms of cloud depth and base (Heike Konow, personal communication, 2019).

As for the case study also in the seasonal composites of the flight domains the stretched distribution of WVP agrees well between model and observation (Fig. 7 ~~b, c~~b-e). The uncertainty in the observational estimate as well as the sensitivity to model resolution is small for both seasons. In December 2013 the model tends to be too moist with the largest bias up to 2 kg/m^2 between the ~~10th~~ 20th and the 60th percentile and a smaller moist bias for the very low and the high WVPs. In August 2016, the agreement is excellent. The LEM results fall almost exactly on the WALES estimate for the lower half of the stretched moisture space and the SRM results coincide with the WALES estimate in the upper half of the stretched moisture space.

4.2 Vertical ~~Distribution~~ distribution of ~~Water Vapor~~ water vapor and ~~Cloud Fraction~~cloud fraction

For the December composite the vertical distribution of mean water vapor, its first moments, and the cloud fraction is very similar to the case study on 11. December 2013 (Sect. 3). We find good agreement between model and observation both in value and shape of the vertical profiles with a few exceptions (Fig. 8 a-d): a ~~moist model bias at the~~ too high model inversion, an underestimation of the standard deviation of q_v in the cloud layer by the model, the model's negative skewness of q_v at

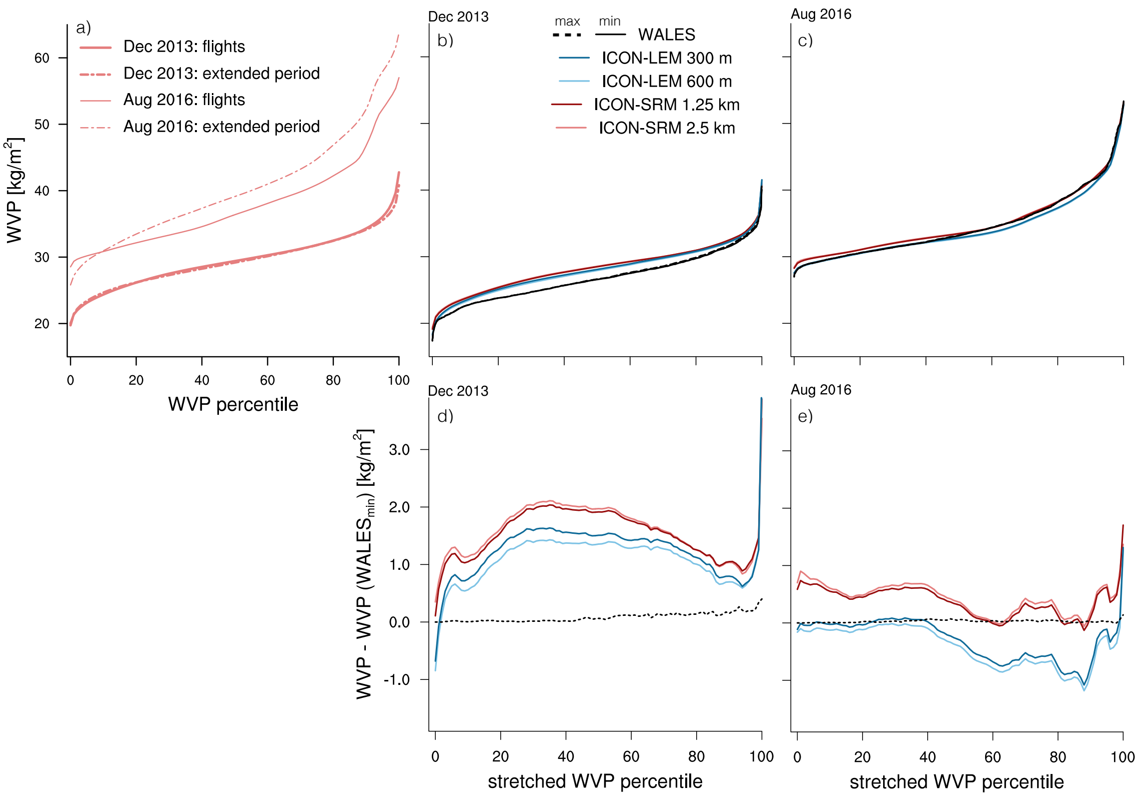

Figure 7. WVP as a function of WVP percentiles (a) for ICON-SRM 2.5 km and the flight period and domain in December 2013 and August 2016 (see Table 1) as well as for an extended period that includes a longer time period for a domain with mean borders (see text for details); (b) for the flight composite in December 2013; and (c) for the flight composite in August ~~2016.~~2016; (d,e) as in (b,c) but for the difference to WALES$_{min}$.

cloud base as compared to a positive value in observations, and an upward shift of the modeled height of the maximum cloud fraction. One difference to the case studies of 11. December 2013 is a stronger secondary maximum of cloud fraction near 2 km height in the simulations with 600 m to 2.5 km grid spacing. These small stratiform cloud shields below the inversion are often present in both model and observations (Lamer et al., 2015; Vogel et al., 2019) but ~~cannot be found in~~ are mostly removed from our analysis of the WALES data ~~in our time period~~due to their opacity. The LEM simulations with finest grid spacing (300 m) are closer to the observations in this case.

Compared to the December composite the August composite is characterized by a moister free troposphere and a shallower cloud layer (< 2 km, Fig. 8 e-h). This supports the understanding that a moister free troposphere promotes shallower cumuli because both the entrainment of moister air into the boundary layer, which decreases surface fluxes, and a weaker radiative cooling at the cloud layer top lead to a weaker buoyancy excess in clouds compared to their environment and therefore convection remains shallower (e.g. Nuijens and Siebesma, 2019).

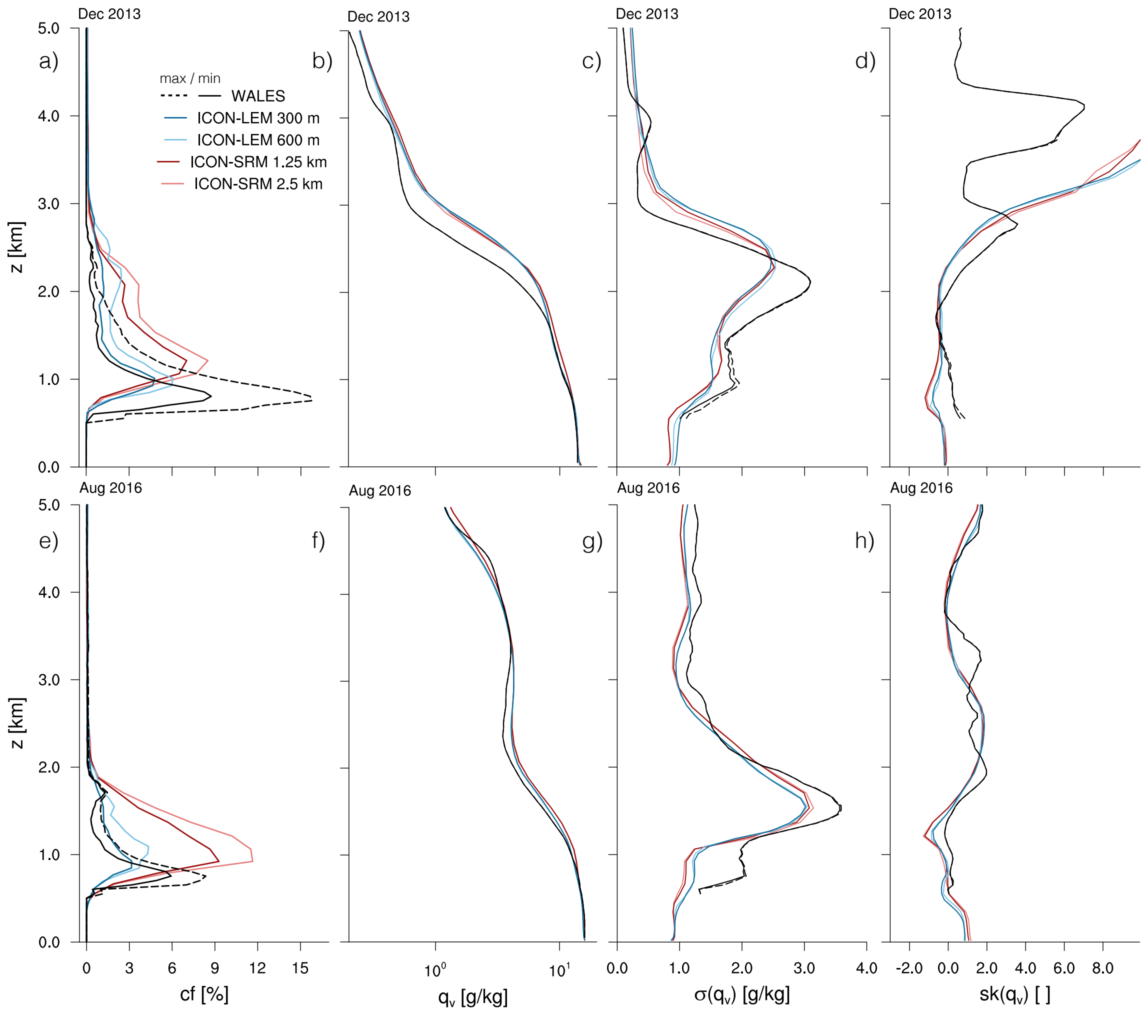

Figure 8. Profiles of (a,e) cloud fraction, (b,f) mean water vapor. q_v, and its (c,g) standard deviation and (d,h) skewness for the flight composites of (a-d) December 2013 and (e-h) August 2016 in stretched WVP space as defined in Fig. 7 b and c.

For two features there is better agreement between model and observations in the August composite than in the December composite: the moist model bias at the inversion is strongly reduced in August, and model and observations agree on a near-zero skewness of q_v near cloud base. However, the upward shift in the modeled height of the maximum cloud fraction and the underestimation of the standard deviation of q_v in the cloud layer by the model remain. Compared to the SRM simulations at coarser resolution, the LEM simulations are better able to capture the height of the cloud maximum and the amount of cloud fraction except for the cloud base cloud fraction. The SRM simulations clearly overestimate the cloud fraction throughout the cloud layer above cloud base. Because cloud fraction is not converged in the LEM simulations, we expect an underestimation of cloud fraction as grid spacing approaches decameter scale.

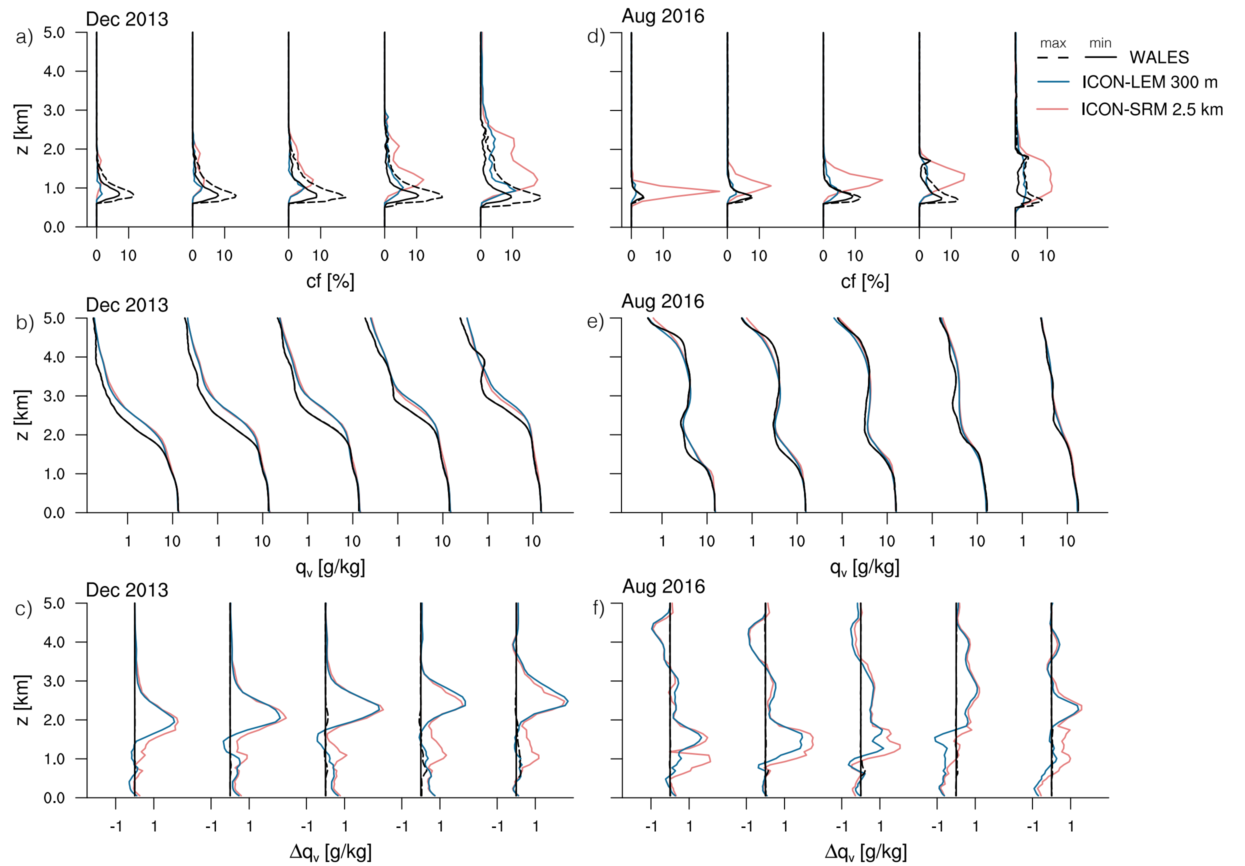

Figure 9. Profiles of (a,d) cloud fraction, (b,e) water vapor and (c,f) the difference to water vapor WALES$_{min}$ for flight composites (a-c) December 2013 and (d-f) August 2016. Each profile shows the mean for a 20-percentile range of WVP in stretched moisture space from driest profiles on the left to moistest profiles on the right (see Fig. 7).

A robust feature of the December and the August composite is the observed deepening of the cloud layer with increasing WVP from a few hundred meters for low WVP to the top of the inversion for high WVP (at 3 km in December 2013 and at 2 km in August 2016, Fig. 9 a, d). This deepening is well captured by the simulations across resolution.

A better representation of cloud fraction with higher resolution becomes apparent for the covariation of cloud fraction with WVP. In the August composite the LEM simulations capture the observed increase in cloud fraction from cloud-free to about 10 % (Fig. 9 d). However, the transition from cloud-free to low cloud fractions occurs too late in moisture space in the LEM. In contrast the coarse resolution SRM simulates clearly too much cloud fraction in the driest part of the moisture space where none is observed and overestimates cloud fraction at high WVP throughout the cloud layer above cloud base. ~~A too high~~ If the low-cloud fraction is too large, this increases the radiative cooling of the subcloud layer and can perhaps artificially promote convective self-aggregation too strongly when it is driven by low-level radiative cooling outside deep convective regions (e.g., Muller and Held, 2012; Hohenegger and Stevens, 2016; Wing et al., 2017).

Different from the August composite, in the December composite even for the driest part of the moisture space a distinct cloud fraction is observed (Fig. 9 a). Neither the SRM nor the LEM are able to capture this cloud regime but instead simulate cloud-free conditions. While both observational estimates of cloud fraction agree well for the dry part of the moisture space, the picture is less clear for the moist part of the moisture space. For $\mathrm{WALES_{max}}$ the maximum cloud fraction increases with increasing WVP but for $\mathrm{WALES_{min}}$ it is close to constant. The SRM and LES results both show increasing cloud fraction with increasing WVP but due to the uncertainty from the observational estimate, we cannot confirm this behaviour with WALES. Using ground based observations that are better able to estimate cloud fraction near cloud base, Nuijens et al. (2013) find that most of the variability in cloud fraction comes from clouds aloft and that clouds near the LCL are rather invariant with time. Although the variability depends on the time scale considered, this and the theory of the cumulus valve mechanism (Neggers et al., 2006; Bellon and Stevens, 2013) seem to be supported by the $\mathrm{WALES_{min}}$ estimate of a constant cloud fraction near cloud base in moisture space, but not by $\mathrm{WALES_{max}}$.

Differences in the vertical distribution of water vapor between model and observations are more subtle than those in cloud fraction. The observed rate of increase in inversion height in moisture space is well captured by the simulations (Fig. 9 b,e). In both the model and the observations the increase in WVP is mostly accomplished by a deepening of the moist layer and to a lesser extent by increasing moisture in the subcloud layer or above. If the increase in WVP was solely due to a deepening of the moist layer, then the agreement in the deepening rate between observations and simulations would directly follow from their agreement in percentile distribution of WVP (Fig. 7). It can therefore not be seen as a fully independent feature.

In the December composite the simulated inversion is shifted upward independent of WVP, which causes a strong bias around 2 km height (Fig. 9 c). For the December and the August composite the simulated gradients at the inversion are smoother than those observed, a well-known difficulty of simulating inversions in particular if vertical resolution is moderate. (In ICON-LEM the vertical grid spacing is about 100 m at 2 km height, for ICON-SRM 200 m.) Because the gradient of moisture at the inversion plays an important role for the local radiative fluxes, the weaker gradient implies a less concentrated radiative cooling in the simulations at the cloud layer top. Besides the too high cloud fraction at kilometer-scale resolution discussed above, the too smooth moisture gradient at the inversion is another model feature that distorts the interaction between radiation, subsidence and cloud development.

Model biases in q_v also lead to misrepresentations in modeled cloud fraction. In the August composite in the driest 20 percentiles of moisture space, the SRM is too moist between 500 m and 1000 m that is where there is too high cloud fraction. For the mid-range percentiles of moisture space (between the 20 percentile and 60 percentile) the bias in modeled q_v shows a bipolar structure for both SRM and LEM: On the one hand, at the height of the observed cloud maximum the modeled q_v is slightly too low, coinciding with modeled spurious too low cloud fraction at the observed cloud base. On the other hand, around the inversion the modeled q_v is too high, coinciding with spurious cloud fraction in the SRM at a height where there are much less clouds observed. We suspect that the latter feature only appears in the SRM simulation and not in the LEM simulation because the SRM applies a cloud fraction parameterization which can produce cloud cover at subsaturation. Taken together, the model smooths q_v in the inversion and thereby puts moisture too high into the inversion region where it produces clouds in the SRM and lacks moisture below the inversion where clouds are observed but not represented in the model.

5 Conclusions

In this study, we analyse the distribution of water vapor and clouds in the trades and how their covariation differs in observations and high-resolution models. The NARVAL campaigns, which took place in the northern tropical Atlantic east of Barbados, provide the opportunity to analyse the distribution of water vapor in the trade wind regime of shallow cumulus cloud fields ~~and in the vicinity of deep convection~~ (Stevens et al., 2019b). In this study, we analyse five research flights from December 2013 probing the region's dry season and four research flights from August 2016 probing the region's moist season. With a horizontal resolution of 2.5 km, the WALES lidar during the NARVAL campaigns provides accurate measurements of the water vapor distributions primarily in the cloud-free gaps of the shallow cumulus regime. The lidar data are compared with results from nested ICON model runs that are available at four grid spacings from 2.5 km to 300 m and that include the area and period of the flight domains.

Because of its stochastic nature, shallow convection is not expected to trigger at the exact same location and time in simulations as it does in reality. To bypass the issue of co-location but retain information on variability, we sort water vapor profiles from the driest to the wettest profile and compare simulations and observations in moisture space (Bretherton et al., 2005; Schulz and Stevens, 2018). Because the signal of the WALES lidar is attenuated rapidly when encountering a cloud and therefore preferentially misses cloudy, high moisture profiles, information from the HAMP radiometers co-located with the lidar is used to construct a "stretched moisture space" that enables a fair comparison between WALES and ICON.

Across model grid spacing from hecto- to kilometer scale, ICON is able to represent the observed features of the water vapor distribution well. In stretched moisture space it correctly captures the full range of WVP from 20 kg/m^2 to 55 kg/m^2, the main features of the vertical distribution of the first three moments of water vapor, and the variability of water vapor profiles across moisture space. An exception in the vertical distribution is a persistent moist model bias at the trade wind inversion in the dry season, where the model simulates the inversion too high. In both seasons the model tends to smooth the moisture gradient at the inversion too much, which is a known feature of excessive model diffusion and might also be a result of underresolving shallow convection with low horizontal resolution. In addition, the simulations slightly underestimate the variability of water vapor in the cloud and subcloud layer in both seasons. Both the too smooth inversion gradient and the too weak cloud layer variability are expected to distort the interaction between radiation, subsidence and cloud development. That there is little dependence of these features on grid spacing and the general good agreement with observations implies no advantage of hectometer grid spacing over kilometer grid spacing in representing the water vapor distribution in the trade wind regime.

In contrast to water vapor, the modelled cloud fraction strongly depends on grid spacing. While the observed cloud deepening with increasing moisture is captured well across model resolutions, the modeled cloud fraction strongly decreases with increasing grid resolution. In the dry season the observational uncertainty in cloud fraction is too large to make a firm statement. In the wet season simulations with hectometer grid spacing agree better with observations than simulations with kilometer grid spacing. In particular, the transition from cloud-free to low cloud fraction with increasing moisture, which reflects the close connection between the distribution of water vapor and clouds, is better represented at hectometer resolution. Also, the height of maximum cloud fraction, which is observed just above cloud base, is shifted upward in the model in both seasons but decreases with higher resolution towards the observed values. Although cloud amount and its vertical distribution is compelling at 300 m grid spacing, it is not converged yet, which is in line with idealized modelling studies showing that LEM underestimates cloud fraction for decameter grid spacing (Vogel et al., 2019).

In conclusion, we show that high-resolution simulations of the shallow cumulus trade wind regime with kilometer scale grid spacing and realistic boundary conditions are able to capture the characteristics of the lower tropospheric water vapor distribution well (Heinze et al., 2017; Stevens et al., 2019a). They however have difficulties to reproduce the observed covariation of water vapor and cloud statistics, which is improved at hectometer resolution. As has been shown for conventional climate models, which apply a convective parameterization at much coarser resolution (e.g., Jiang et al., 2012), this means that capturing the water vapor distribution correctly does not imply that shallow clouds that live at the tail of the water vapor distribution are also well represented. It remains an open question which role such shallow cloud biases in kilometer-scale simulations play for the heat budget of the cloud layer and how they interact with the large-scale environment, e.g., in global storm resolving models (Satoh et al., 2019). The latter question of whether and how shallow cloud biases depend on the large-scale environment also prompts itself to be pursued further in the light of EUREC^{4}A, which sets out for measuring the distribution of water vapor and clouds in conjunction with the large-scale environment (Bony et al., 2017).

Code and data availability. Model results and observational data used in this study are published in different peer reviewed papers, that is ICON-SRM NARVAL 1+2: Klocke et al. (2017); ICON-LEM NARVAL 1+2: Stevens et al. (2019b); WALES NARVAL 1: Kiemle et al. (2017); WALES NARVAL 2: Gutleben et al. (2019); HAMP NARVAL 1+2: Jacob et al. (2019c, a, b).

Author contributions. AKN and CK developed the idea of the study and carried out the analysis for this manuscript. AKN took the leading role in writing the manuscript with input from CK.

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

Acknowledgements. We thank Bjorn Stevens for inspiring discussions and in particular for his idea to include radiometer data in this study. We also thank Marek Jacob for kindly providing the HAMP data, Silke Gross and Martin Wirth for the WALES data, and Daniel Klocke and Matthias Brück for the ICON results used in this study. We thank the editor Andreas Richter and two anonymous reviewers for their helpful comments. The data used in this publication was gathered in the NARVAL 1/NARVAL 2 campaigns and WALES data is made available through the German Aerospace Center (DLR). NARVAL was funded with support of the Max Planck Society, the German Research Foundation (DFG, project HALO-SPP 1294), the European Research Council (ERC), the German Meteorological Weather Service (DWD) and DLR. Primary data and scripts used in the analysis and other supplementary information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de. During part of this research AKN was funded as part of the Hans-Ertel Centre for Weather Research. This research network of Universities, Research Institutes and the Deutscher Wetterdienst is funded by the BMVI (Federal Ministry of Transport and Digital Infrastructure). AKN also acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2037 'Climate, Climatic Change, and Society' – Project Number: 390683824.

References

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47, 2015.

Bellon, G. and Stevens, B.: Time scales of the trade wind boundary layer adjustment, J. Atmos. Sci., 70, 1071–1083, 2013.

Bony, S. et al.: EUREC^{4}A: a field campaign to elucidate the couplings between clouds, convection and circulation, Surv. Geophys., 38, 1529–1568, 2017.

Bretherton, C. S., Peters, M. E., and Back, L. E.: Relationships between water vapor path and precipitation over the tropical oceans, J. Climate, 17, 1517–1528, 2004.

Bretherton, C. S., Blossey, P. N., and Khairoutdinov, M.: An energy-balance analysis of deep convective self-aggregation above uniform SST, J. Atmos. Sci., 62, 4273–4292, 2005.

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., and Brdar, S.: Large eddy simulation using the general circulation model ICON, J. Adv. Model. Earth Syst., 7, 963–986, 2015.

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 study, Atmos. Chem. Phys., 19, 10 659–10 673, https://doi.org/10.5194/acp-19-10659-2019, 2019.

Hansen, A.: New techniques for ultra-high-resolution circulation model evaluation, PhD thesis, Universität Hamburg, 2020.

Heinze, R. et al.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Quart. J. Roy. Met. Soc., 143, 69–100, 2017.

Hohenegger, C. and Stevens, B.: Coupled radiative convective equilibrium simulations with explicit and parameterized convection, J. Adv. Model. Earth Syst., 8, 1468–1482, https://doi.org/10.1002/2016MS000666, 2016.

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., Schulzweida, U., and Stevens, B.: Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing, p. in review, 2019.

Holloway, C. E. and Neelin, J. D.: Moisture vertical structure, column water vapor, and tropical deep convection, J. Atmos. Sci., 66, 1665–1683, 2009.

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.: Liquid water path and integrated water vapor over the tropical Atlantic during NARVAL2, https://doi.org/10.26050/WDCC/HALO_measurements_6, 2019b.

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, Atmos. Meas. Tech., 12, 3237–3254, https://doi.org/10.5194/amt-2019-18, 2019c.

Jacob, M., Kollias, P., Ament, F., Schemann, V., and Crewell, S.: Multi-layer cloud conditions in trade wind shallow cumulus – Confronting models with airborne observations, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-14, in review, 2020.

Jiang, J. H., Su, H., Zhai, C., Perun, V. S., Del Genio, A., Nazarenko, L. S., Donner, L. J., Horowitz, L., Seman, C., Cole, J., et al.: Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations, J. Geophys. Res. Atmos., 117, https://doi.org/10.1029/2011JD017237, 2012.

Keil, C., Röpnack, A., Craig, G. C., and Schumann, U.: Sensitivity of quantitative precipitation forecast to height dependent changes in humidity, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008GL033657, 2008.

Kiemle, C., Groß, S., Wirth, M., and Bugliaro, L.: Airborne lidar observations of water vapor variability in tropical shallow convective environment, Surv. Geophys., 38, 1425–1443, https://doi.org/10.1007/s10712-017-9431-5, 2017.

Klocke, D., Brueck, M., Hohenegger, C., and Stevens, B.: Rediscovering the Doldrums in Cloud Resolving Simulations of the Tropical Atlantic, Nat. Geosci., 10, 891–896, https://doi.org/10.1038/s41561-017-0005-4, 2017.

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 Syst. Sci. Data, 11, 921–934, https://doi.org/10.5194/essd-11-921-2019, 2019.

Lamer, K., Kollias, P., and Nuijens, L.: Observations of the variability of shallow trade wind cumulus cloudiness and mass flux, J. Geophys. Res. Atmos., 120, 6161–6178, 2015.

Mauritsen, T. and Stevens, B.: Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models, Nat. Geosci., 8, 346–351, 2015.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., et al.: Tuning the climate of a global model, J. Adv. Model. Earth Syst., 4, 2012.

Muller, C. and Bony, S.: What favors convective aggregation and why?, Geophys. Res. Lett., 42, 5626–5634, 2015.

Muller, C. J. and Held, I. M.: Detailed investigation of the self-aggregation of convection in cloud-resolving simulations, J. Atmos. Sci., 69, 2551–2565, 2012.

Naumann, A. K., Stevens, B., and Hohenegger, C.: A moist conceptual model for the boundary layer structure and radiatively driven shallow circulations in the trades, J. Atmos. Sci., 76, 1289–1306, https://doi.org/10.1175/JAS-D-18-0226.1, 2019.

Neggers, R., Stevens, B., and Neelin, J. D.: A simple equilibrium model for shallow-cumulus-topped mixed layers, Theoretical and Computational Fluid Dynamics, 20, 305–322, 2006.

Nehrir, A. R., Kiemle, C., Lebsock, M. D., Kirchengast, G., Buehler, S. A., Löhnert, U., Liu, C.-L., Hargrave, P. C., Barrera-Verdejo, M., and Winker, D. M.: Emerging technologies and synergies for airborne and space-based measurements of water vapor profiles, Surv. Geophys., 38, 1445–1482, 2017.

Nuijens, L. and Siebesma, A. P.: Boundary Layer Clouds and Convection over Subtropical Oceans in Our Current and in a Warmer Climate, Current Climate Change Reports, https://doi.org/10.1007/s40641-019-00126-x, 2019.

Nuijens, L., Stevens, B., and Siebesma, A. P.: The environment of precipitating shallow cumulus convection, J. Atmos. Sci., 66, 1962–1979, 2009.

Nuijens, L., Serikov, I., Hirsch, L., Lonitz, K., and Stevens, B.: The distribution and variability of low-level cloud in the North Atlantic trades, Quart. J. Roy. Met. Soc., 140, 2364–2374, 2013.

Pierrehumbert, R. T.: Thermostats, radiator fins, and the local runaway greenhouse, J. Atmos. Sci., 52, 1784–1806, 1995.

Ramanathan, V. and Collins, W.: Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Nino, Nature, 351, 27, 1991.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., and Düben, P.: Global Cloud-Resolving Models, Curr. Clim. Change Rep., https://doi.org/10.1007/s40641-019-00131-0, 2019.

Schulz, H. and Stevens, B.: Observing the tropical atmosphere in moisture space, J. Atmos. Sci., 75, 3313–3330, https://doi.org/10.1175/JAS-D-17-0375.1, 2018.

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, Surv. Geophys., 38, 1371–1397, 2017.

Stevens, B., Acquistapace, C., Hansen, A., Klinger, C., Klocke, D., Schubotz, W., Windmiller, J., and coauthors: Large-eddy and Storm Resolving Models for Climate Prediction – The Added Value for Clouds and Precipitation, p. in review, 2019a.

Stevens, B., Ament, F., Bony, S., Crewell, S., Ewald, F., Gross, S., Hansen, A., Hirsch, L., Jacob, M., Kölling, T., et al.: A high-altitude long-range aircraft configured as a cloud observatory – the NARVAL expeditions, Bull. Am. Met. Soc., https://doi.org/10.1175/BAMS-D-18-0198.1, 2019b.

Vial, J., Bony, S., Stevens, B., and Vogel, R.: Mechanisms and model diversity of trade-wind shallow cumulus cloud feedbacks: a review, Surv. Geophys., 38, 159–181, 2017.

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, J. Adv. Model. Earth Syst., p. early online, https://doi.org/10.1029/2019MS001746, 2019.

Vogel, R., Nuijens, L., and Stevens, B.: Influence of deepening and mesoscale organization of shallow convection on stratiform cloudiness in the downstream trades, Quart. J. Roy. Met. Soc., p. early online view, 2019.

Wing, A. A., Emanuel, K., Holloway, C. E., and Muller, C.: Convective self-aggregation in numerical simulations: A review, in: Shallow Clouds, Water Vapor, Circulation, and Climate Sensitivity, pp. 1–25, Springer, 2017.

Wirth, M., Fix, A., Mahnke, P., Schwarzer, H., Schrandt, F., and Ehret, G.: The airborne multi-wavelength water vapor differential absorption lidar WALES: system design and performance, Applied Physics B, 96, 201, 2009.

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, Quart. J. Roy. Met. Soc., 141, 563–579, 2015.