JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002209

Linking Large-Eddy Simulations to Local Cloud Observations

Vera Schemann[1], Kerstin Ebell[1], Bernhard Pospichal[1], Roel Neggers[1], Christopher Moseley[2], and Bjorn Stevens[3]

[1]Institute of Geophysics and Meteorology, University of Cologne, Cologne, Germany, [2]Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan, [3]Max Planck Institute for Meteorology, Hamburg, Germany

Key Points:
- Large-eddy simulations including external variability can bridge the gap between ground-based observations of clouds and large-scale models
- For comparison with local observations, it is important to take external variability (e.g., large-scale forcing and surface) into account
- ICON-LEM offers new possibilities to simulate small scales while considering external variability

Correspondence to:
V. Schemann,
vera.schemann@uni-koeln.de

Citation:
Schemann, V., Ebell, K., Pospichal, B., Neggers, R., Moseley, C., & Stevens, B. (2020). Linking large-eddy simulations to local cloud observations. *Journal of Advances in Modeling Earth Systems*, *12*, e2020MS002209. https://doi.org/10.1029/2020MS002209

Received 16 JUN 2020
Accepted 18 NOV 2020
Accepted article online 21 NOV 2020

Abstract In order to enhance our understanding of clouds and their microphysical processes, it is crucial to exploit both observations and models. Local observations from ground-based remote sensing sites provide detailed information on clouds, but as they are limited in dimension, there is no straightforward way to use them to guide large-scale model development. We show that large-eddy simulations (LES) performed on similar temporal and spatial scales as the local observations can bridge this gap. Recently, LES with realistic topography and lateral boundary conditions became feasible for domains spanning several 100 km. In this study, we show how these simulations can be linked to observations of the Jülich Observatory for Cloud Evolution (JOYCE) for a 9-day period in spring 2013. We discuss the advantages and disadvantages of very large versus small but more constrained domains as well as the differences compared to more idealized setups. The semi-idealized LES include time-varying forcing but are run with homogeneous surfaces and periodic boundary conditions. These assumptions seem to be the reason why they struggle to represent the observed varying conditions. The simulations using the "realistic" setup are able to represent the general cloud structure (timing, height, phase). It seems that the smaller and more constrained domain allows for a tighter control on the synoptic situation and is the preferred choice to ensure the comparability to the local observations. These simulations together with measures as the shown Hellinger distance will allow us to gain more insights into the representativeness of column measurements in the future.

Plain Language Summary Clouds are still a cause for uncertainty in our understanding of climate and climate feedbacks. Due to the large range of involved scales—from small droplets up to storm systems—their representation in weather and climate models is an ongoing challenge. While new and sophisticated measurements of the atmospheric column could provide new insights into important processes, their linking to models is not trivial and is ongoing research. In this study, we are presenting and exploring different approaches to combine local observations of clouds with state-of-the-art high-resolution simulations. And we are presenting a setup, which shows a promising representation of the observed clouds and is constrained enough to be applicable for long-term statistics—one of the key requirements for improvements and evaluation clouds in of weather and climate models.

1. Introduction

Clouds and cloud feedback mechanisms have, for quite some time, contributed substantial uncertainty to estimates of how the climate system responds to radiative forcing (Bony et al., 2006; Boucher et al., 2013; Cess et al., 1990; Stevens et al., 2016). Even as a new generation of climate models, with kilometer-scale horizontal meshes, is showing great promise for better representing precipitation processes (Satoh et al., 2019; Stevens et al., 2019), clouds remain challenging to represent, with expected, but largely unquantified sensitivity to cloud microphysical processes (Stevens et al., 2020). An ability to accurately represent clouds in meteorological models is important for all types of weather forecasts, but also new application sectors such as renewable energy. For these reasons, there has been a tremendous effort over the past decades to improve observations, simulations, and models of cloud processes, as well as interest in new methods for harmonizing these methodologies (Schneider et al., 2017).

In a new twist on an old approach, Schneider et al. (2017) propose to spawn multitudes of idealized large-eddy simulations (LES) for the large-scale conditions associated with important cloud regimes. The

simulations would then be constrained by satellite data, and their dynamics would be learned by machines. Advances in computing, and in machine learning, would thus allow the replication of the Global Energy and Water Experiment (GEWEX) Cloud Systems Studies (GCSS) approach outlined by Browning et al. (1993) on a massive scale. The GCSS approach, like its more modern incarnation, adopts the scale-separation hypothesis inherent to the parameterization problem, whereby it is assumed that small-scale processes respond to much larger-scale forcing, to set the properties of clouds and precipitation. This assumption gives relevance to the study of the dynamics of very high-resolution simulations over very small domains for large-scale conditions that are prescribed and stationary. Essentially, it allows for the study of quite idealized problems, whereby heterogeneity of any kind in the forcing is neglected. The simulation of stratocumulus idealized from observations taken during the First International Satellite Cloud Climatology Project (ISCCP) Regional Experiment (FIRE) is an early example of this approach (Moeng et al., 1996). But over the years, there have been great many studies of this kind, and the approach can under some conditions yield quite satisfactory results, remains quite popular (Holloway et al., 2014), and—forshadowing its almost industrial application as suggested by Schneider et al. (2017)—has begun to be applied also to routine observations from local sites (Neggers et al., 2012; Schalkwijk et al., 2015).

Parallel to these developments, some groups have been experimenting with approaches that relax the parameterization assumption, by embedding smaller domain very high-resolution simulations in a more dynamic large-scale environment. Notably, Chow et al. (2006) embedded LES in a mesoscale model to study boundary layer processes over complex terrain, an approach developed simultaneously and applied to idealized problems by Moeng et al. (2007). In doing so, Chow et al. (2006) noted not only the importance of an accurate representation of surface forcing but also sensitivities to how the nested simulations were set up, an issue also investigated by Moeng et al. (2007). In a later, related study, adopting a similar approach of nesting an LES within larger scale mesoscale model, Talbot et al. (2012) also highlight the importance of the mesoscale meteorological forcing for the LES. These approaches make it possible to use observations from regions, or for time periods, where there is not a strong separation between the large and small scales. As computational capacity has increased, it has also become possible to simply do away with the nesting and begin performing LES over very large domains, thereby coupling the mesoscale with the turbulence scale more organically, and allowing the representation of more realistic situations (Heinze, Dipankar, et al., 2017; Stevens et al., 2020).

In this study, we systematically explore the trade-offs associated with some of the different approaches outlined above. For instance, the benefits of a large domain which allows a realistic coupling between turbulent and mesoscale motions, versus a local domain which might allow a tighter prescription of the large-scale flow and a higher resolution representation of turbulent processes. In the latter case, one can further ask how much additional information is imprinted by heterogeneity in the lower boundary condition, or through the open boundary conditions. To perform the study, we take advantage of and expand upon the capabilities of the LES model configuration of ICON (ICON-LEM; Dipankar et al., 2015). ICON can be run with open lateral boundary conditions and a heterogeneous and complex surface over very large domains (Heinze, Dipankar, et al., 2017) as well as in semi-idealized mode (Heinze, Moseley, et al., 2017), or with the small and computationally more efficient setup as used in Marke et al. (2018) and Schemann and Ebell (2020). In this study, with applications such as the LES Symbiotic Simulation and Observation (LASSO) workflow (Gustafson et al., 2019, 2020) in mind, we also include in our comparison suite simulations with the Dutch Atmospheric Large-Eddy Model (DALES; Heus et al., 2010). Having these two different configurations of semi-idealized LES using two different forcing data sets provides some estimate for the space of possible realizations.

The focus of our study will be on the representation of clouds through the varying approaches to perform LES around local observations. As a reference site for comparison, we choose the Jülich Observatory for Cloud Evolution (JOYCE) (Löhnert et al., 2015), which provides several remote sensing observations and is surrounded by an area of modest heterogeneity. But in general, the setups should be applicable for different locations and conditions.

The manuscript is organized as follow: In section 2, the different model setups as well as the observational basis of evaluation are introduced. This will be followed by a basic comparison (section 3) and a discussion of the resolution dependency (section 4) as well as the dependency on different forcing data. Further details, such as the role of the modest topography in the study area, are explored through the analysis of a specific

Table 1
Overview of the Applied Model Setups Summarizing the Height of the Model top, the Domain Size (Linear Dimension), Horizontal Mesh Size (linear dimension), and Boundary Conditions

Setup	Top	Domain size	Horz. mesh	Boundary conditions
ICON-DE	21 km	1,000 km	156 m	O/Het (C)
"	"	"	312 m	"
"	"	"	624 m	"
ICON-LOC	"	20 km	78 m	O/Het (C,I)
"	"	30 km	156 m	"
"	"	60 km	312 m	"
"	"	110 km	624 m	"
ICON-SI	13 km	7 km	50 m	P/Hom (C)
DALES-SI	13 km	6.4 km	50 m	P/Hom (I)

Note. Boundary conditions are either open and heterogeneous (O/Het) and thus including different surface types as well as topography or periodic (lateral) and homogeneous (surface) as designated by P/Hom. Boundary conditions are provided by either COSMO-DE (C) or the ECMWF-IFS (I). The model configurations are explained in sections 2.2 and 2.3.

case study in section 5. We conclude, as is customary with a brief summary and a restatement of our major findings.

2. Model Setup and Data

Different models and model configurations (see Table 1) are applied to study the weaknesses and strengths in their ability to capture different synoptic conditions as well as the details provided by measurements. Simulations are compared to observations from the measurement site JOYCE (Löhnert et al., 2015). In this section, the different model setups as well as the observational site and its data are introduced. References are provided for information already in the published literature.

Often the word "resolution" is used as short hand for the grid spacing. As many studies have shown, they are not the same thing, but the former generally scales with the latter, and we use the terms synonymously. At least for the ICON model, there is also ambiguity in what is meant by grid spacing. Values given in Table 1, measure the edge length of a grid cell, which—due to the triangular grid—has to be scaled by a factor of 2/3 to provide an area-based resolution. Hence, an edge length of 78 m corresponds to a 50-m area-based resolution; however, each cell has less information than in a rectangular grid, i.e., because the velocities are defined on cell faces, triangles come with three velocities instead of four. This is expected to impact (reduce) the resolution for a given grid spacing as compared to a quadrilateral discretization.

2.1. Analyzed Time Period

To capture different synoptic situations and investigate the overall performance as well as looking into specific case studies, our study focuses on 9 days of the Observational Prototype Experiment (HOPE; Macke et al., 2017). HOPE comprised a 2-month field study in the vicinity of Jülich, Germany, during April and May 2013. The time period 24 April 2013 to 2 May 2013 was chosen to allow the use of previously performed simulations with the very large domain (ICON-DE). Within this period, two (relatively) clear-sky days were followed by a passage of a frontal system (26 and 27 April). The rest of the period consisted of more mixed conditions with the exception of 2 days with shallow cumulus clouds (1 and 2 May). Hence, we can investigate the performance of the different models and model configurations for different atmospheric situations.

2.2. Realistic Setup (ICON-LEM)

What we call the "realistic" set-up of the ICON Large-Eddy Model (LEM) is one where the simulations are subject to lateral boundary and surface conditions that attempt to mimic reality as closely as possible. For the surface conditions this includes both the specification of the topography and the land-surface properties. As a default, these simulations are initialized and forced every hour with output from the COSMO-DE, the operational numerical weather prediction model of the German Meteorological Service (Deutscher Wetterdienst, DWD) with a grid spacing of 2.8 km (e.g., Baldauf et al., 2011). As described below, both large and small

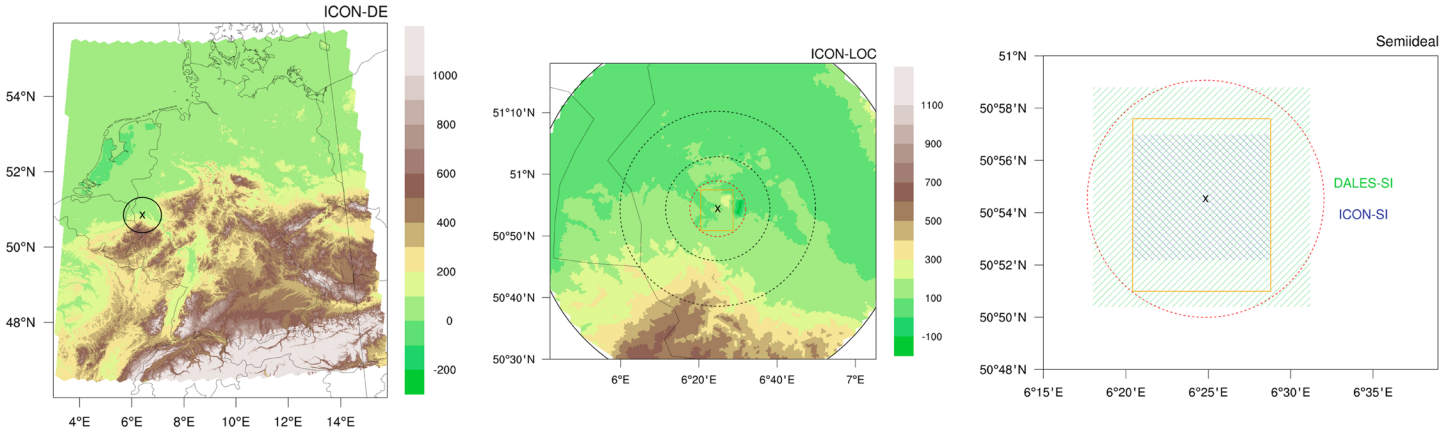

Figure 1. Model domains of the different setups. Left: ICON-DE. The circle indicates the domain of the ICON-LOC; middle: ICON-LOC with the four nested domains; right: sketch for domain size of DALES-SI (green shading), ICON-SI (blue shading) and ICON-LOC for smallest domain. The x indicates the location of JOYCE, and the orange rectangle encloses the subdomain used in the analysis. For ICON-DE and ICON-LOC, the colors show the topography (in m).

domain simulations are performed using the realistic setup. The small domain simulations (ICON-LOC) are performed twice, once with the COSMO-DE forcing and once with forcing data from the Integrated Forecasting System (IFS) of the European Centre for Medium-range Weather Forecasts (ECMWF). This allows us to assess the sensitivity to uncertainty in the boundary and initial data.

The large-domain simulations (ICON-DE) cover the whole of Germany (Figure 1, left) which allows mesoscale processes to develop and interact freely with smaller scale, turbulent, features as are the normal focus of LES. The simulations incorporate two nests. Each nest refines the grid spacing by a factor of 2 and slightly reduces the size of the domain to smooth the change in grid spacing at the lateral boundaries. Compared to the domain as a whole these transition regions are very small, and because the nesting is one way, this effectively provides three simulations with progressively refined meshes (from 624 to 156 m) over all of Germany. The ICON-DE simulations have been performed within the HD(CP)2 project (Heinze, Dipankar, et al., 2017; Stevens et al., 2020) and are used as a reference. They are very computationally expensive and thus have been performed only for selected days (24–26 April 2013, 2 May 2013).

The ICON-LOC simulations are a smaller version of the ICON-DE simulations. They start with a domain size of 110 km and a 624-m grid mesh. They are nested three times with the smallest domain having a size of ca. 20 km and a resolution of ca. 78 m (Figure 1, middle). Like the ICON-DE simulations, the one-way nesting effectively results in four simulations as described in Table 1. To reduce the computation expense and allow yet finer scale simulations, the domain size of each finer mesh is reduced more than is done for the ICON-DE simulations, combined with the smaller sizes of the domains to begin with this results in roughly a factor of two reduction in the domain size with each factor of two reduction in mesh size. An obvious advantage of the small domains is the limited computational demand, which allows the whole analysis period to be simulated. ICON's unstructured mesh and the use of open boundary conditions made it possible to define a roughly circular domain, centered on the JOYCE observational site. By choosing a circular domain, the quality of the simulation should not be effected by the direction of the flow. Experiments were performed with domains of different sizes, but systematic differences were difficult to identify and this aspect of the setup was not further explored.

All of the ICON setups share the same set of parameterizations including a Smagorinsky turbulence scheme (see Dipankar et al., 2015 for more details). For the cloud microphysics parameterization, the two-moment scheme by Seifert and Beheng (2006) is used, which is based on six hydrometeor classes (liquid, ice, rain, snow, graupel, and hail). The model nesting is for both setups one-way. This means that information is only provided from the coarser to the finer resolutions. For both realistic configurations (ICON-DE and ICON-LOC), 150 levels are used, reaching up to 21 km.

For the realistic simulations, we have different output possibilities. For most of our analysis, we will use the "meteogram" output. This consists of quantities taken from the model column closest to the location of the observational site. In case of the ICON-LOC setup, this is the center of the domain. As it is only the output

of one column, the output frequency is rather high with every 9 s. This output is designed to mimic how we observe the atmosphere with automated measurements, but it provides no (horizontal) spatial information, which is the main drawback of this type of output.

For comparison and to investigate the question of how valid the point-to-point evaluation is and how we can use models to put column observations into a three-dimensional context, we also use 2D information of vertically integrated quantities. The output frequency of the 2D data for the ICON-LOC is every 10 min.

2.3. Semi-Idealized Setup (DALES and ICON-LEM)

What we call the semi-idealized simulations follows the more traditional way of configuring and performing LES. These simulations are idealized in the sense that they adopt a simplified surface forcing (i.e., homogeneous land surface types) and periodic horizontal boundary conditions. For ICON-SI, time-dependent skin temperature and surface relative humidity, which are representative for the entire LES domain, are applied. In DALES, an interactive soil model is used, that is initialized in soil temperature and humidity and is then free to evolve during the simulation. The skin temperature and relative humidity follow from the surface energy budget and a simple representation of vegetation. In addition, the large-scale forcing, containing the geostrophic wind, and both horizontal advection and subsidence of the temperature and moisture field, is applied. Note that the contribution of horizontal advection is horizontally homogeneous, whereas the subsidence contribution is not. Newtonian relaxation (nudging) is applied in addition to the previously mentioned larger scale components. A nudging time scale of $\tau = 6$ h is used, which is long enough for the fast boundary layer physics to develop their own unique state and short enough so that larger scale disturbances, such as weather fronts, can be represented in the LES. We use the term "semi-idealized" in this study to point out that we still use time-varying large-scale forcing in order to introduce changes in the synoptic situation—the weather—to the LES instead of sticking to one special case (Neggers et al., 2012). The ICON-LEM model also offers the possibility to be run in a more fully idealized mode (Heinze, Moseley, et al., 2017). However, this setup is less well tested. For this reason, we decided to also include results from a more well-established model (DALES-SI). The DALES model (Heus et al., 2010) has been applied in semi-idealized mode over a wide variety of conditions (Neggers et al., 2012; Reilly et al., 2020), including the JOYCE site (Corbetta et al., 2015; van Laar et al., 2019). The DALES simulations discussed here were generated using the identical mixed-phase microphysics scheme as used in ICON-SI (Seifert & Beheng, 2006). Its recent implementation has been thoroughly tested against observations in the Arctic (Neggers et al., 2019).

Whereas the ICON-LEM semi-idealized version (ICON-SI) (Heinze, Moseley, et al., 2017) is forced with COSMO-DE data, the DALES-SI is forced with IFS data. The exact construction of the IFS forcing is described by van Laar et al. (2019). For semi-idealized simulations, different forcing data sets can result in different atmospheric conditions. For our study, this is an advantage as we would like to span a rather wide range of possible outcomes from semi-idealized models to investigate how close they can come to observations and how they compare to the more realistic setup. A sketch of the model domain of DALES-SI and ICON-SI as well as the domain size of the innermost domain of ICON-LOC can be seen in Figure 1 (right).

For the analysis of the semi-idealized simulations, we mostly focus on domain mean output. For the DALES model, we used output provided every 30 min, which is coarser than for the other models but still provides an impression on the general classification. For the ICON-LEM, we added 2D output for integrated values every 10 min similar to what is done for the realistic setup.

2.4. Observations (JOYCE)

The observations used in this study were performed at JOYCE, the Jülich Observatory for Cloud Evolution (Löhnert et al., 2015). JOYCE was founded in 2008 and became a comprehensive site for ground-based observations of the atmosphere with the main focus on profiling clouds, precipitation, wind, and the thermodynamic state of the atmospheric column using different remote sensing methods. The observations are performed by several cloud and precipitation radars, a microwave radiometer, Doppler lidar, ceilometer, and various other instruments. All these measurements are performed continuously with a temporal resolution of less than a minute. In 2013, JOYCE was part of the HOPE campaign (Macke et al., 2017) where additional ground-based remote sensing instruments were installed in the vicinity of JOYCE to observe local variability. Observational data from HOPE will be used in this study.

Since 2011, JOYCE is part of the European network Cloudnet (Illingworth et al., 2007) within the European Research Infrastructure for the observation of Aerosol, Clouds, and Trace Gases (ACTRIS). The Cloudnet

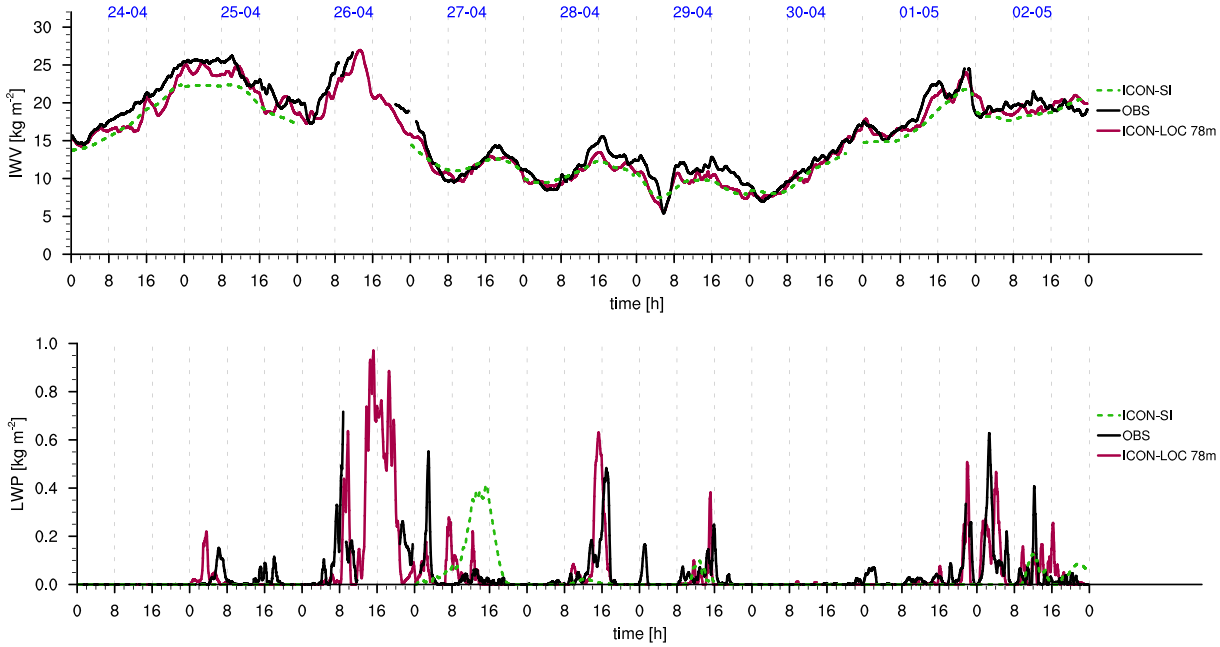

Figure 2. Time series of IWV (top) and LWP (bottom) for 29 April to 2 May 2013 for JOYCE. Values are 30-min running means with observations (black), ICON-LOC with 78-m resolution (red) and ICON-SI (green).

network consists of currently 15 stations around Europe which operate the combination of cloud radar, microwave radiometer and ceilometer. From these observations, Cloudnet provides many cloud properties, such as classification (phase, precipitation), extent and liquid water/ice water content on a constant temporal (30 s) and vertical grid (30 m).

3. Capturing the Weather

To enable the comparison of simulations around heavily instrumented observational sites, it is important to capture the general weather or synoptic conditions. These large-scale features should be provided by the forcing model, while the high-resolution model should resolve and focus on the small-scale features like turbulence and clouds within the given weather regime. To evaluate the representation of the general weather, we compare the integrated water vapor (IWV), which is describing large-scale changes in the atmospheric conditions. As the evolution of the IWV will be dominated by the large scale forcing models, it proves sufficient to compare this quantity from the ICON-LOC 78 m and the ICON-SI, as two examples covering the range of model configurations. Given that we are first interested in whether the general weather situation is well captured, we calculated a 30-min running mean of the IWV for the period 29 April to 2 May 2013. Figure 2 (top) shows a good agreement of the simulated IWV and the observed one. Even though the information is given at the boundaries in the ICON-LOC setup, while the output is taken in the center, it covers nicely the increase and decrease of the IWV over the 9 days. Nevertheless, both model setups tend to underestimate the IWV and show deviations up to 3 kg/m^2. Whereas the IWV is dominated by the large scales, the cloud liquid water path (LWP) provides an estimate of the model's ability to represent the small scales through the liquid cloud occurrence. With respect to the liquid cloud occurrence, the simulations differ more markedly. ICON-LOC shows a reasonable agreement with the observations (Figure 2, bottom), while ICON-SI often underestimates the observed LWP.

The LWP already gives a hint on the representation of cloudy versus noncloudy situations. To evaluate the representation of clouds in more detail, particularly their vertical distribution, we use the Cloudnet classification (Illingworth et al., 2007). The classification for the model data is done by simple thresholds. If the frozen hydrometeors are larger than 1×10^{-8} kg/kg the point is classified as "ice," if the liquid water is larger than the threshold, it is classified as "cloud droplets" and if both are larger as "ice & supercooled droplets." Similarly, we use the same threshold to define the "rain" and the "drizzle/rain and cloud droplets" category. For the semi-idealized simulations (DALES, ICON-SI), we calculated first the mean value and then applied

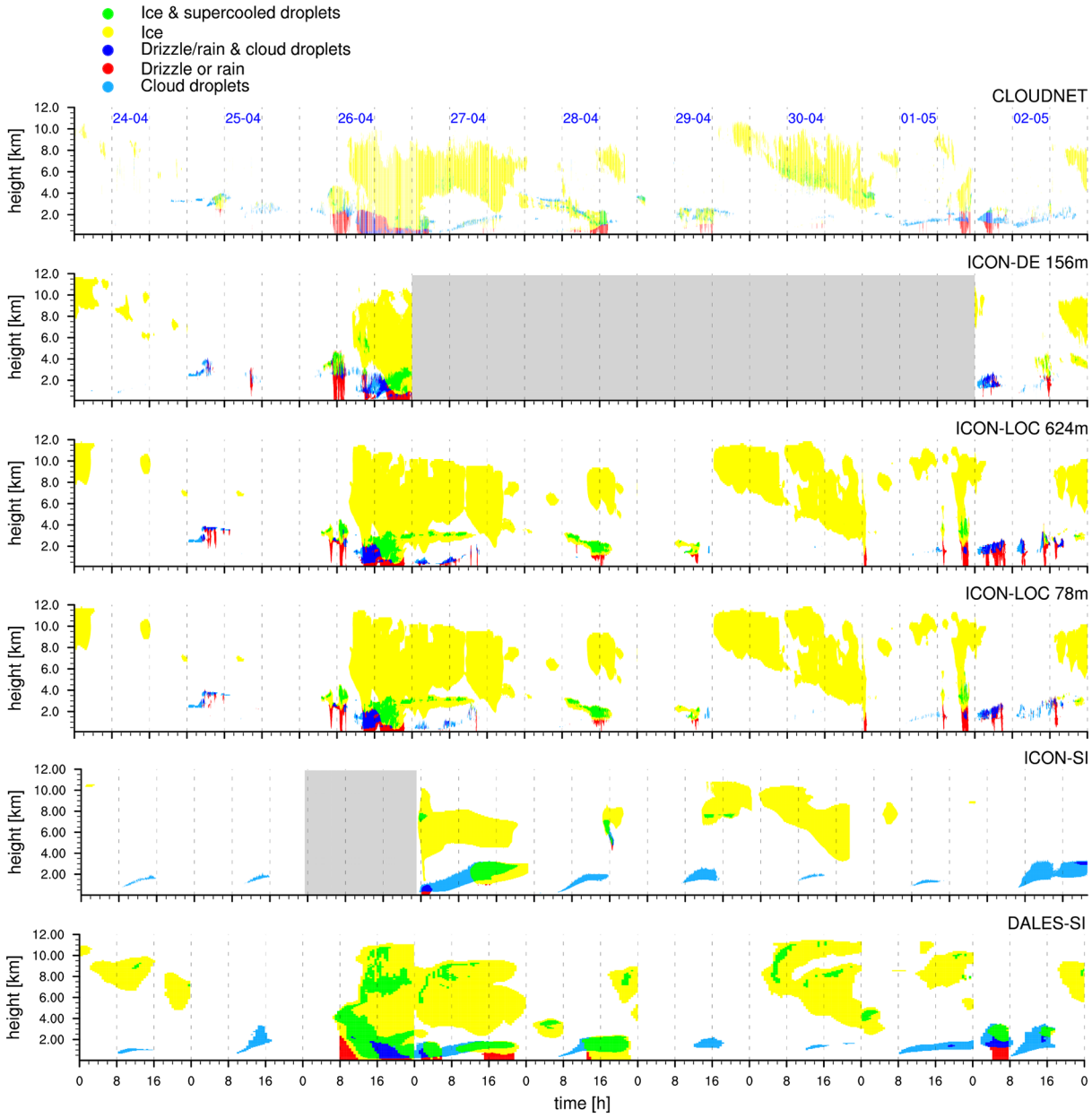

Figure 3. CLOUDNET classification at JOYCE for 26 April to 2 May 2013 and hydrometeor classification for varying model setups: ICON-DE, ICON-LOC with 624- and 78-m resolution, ICON-SI and DALES-SI (from top to bottom). Model classifications are calculated based on a threshold of 1.0×10^{-8} kg/kg for the different hydrometeors. Gray color indicates missing simulation days.

the thresholds. The Cloudnet classification of the measurements, which is used as the reference data set, can be seen in the first panel of Figure 3. It provides an overview of the varying situations, comprising clear sky days with a frontal system and rather fair weather conditions. Already the coarse 624-m simulation of the ICON-LOC setup (Figure 3c) is able to reproduce this variability to a large extent. The higher resolution (78 m, Figure 3d) seems to be beneficial mainly for its improved representation of shallow cumulus clouds at the end of the time period. As those clouds are strongly influenced by the small scales, a higher resolution improves their representation. The higher resolution also shows less precipitation events on 1 and 2 May 2013, which is also closer to the observations. The ICON-DE simulation (Figure 3b) shows at the available days a similar representation of the daily variability, except for 2 May 2013, where the shallow cumulus clouds seem to be underestimated. We will investigate these differences further in section 5. Whereas the high ice clouds seem to be large-scale driven and are also reasonably well represented in the

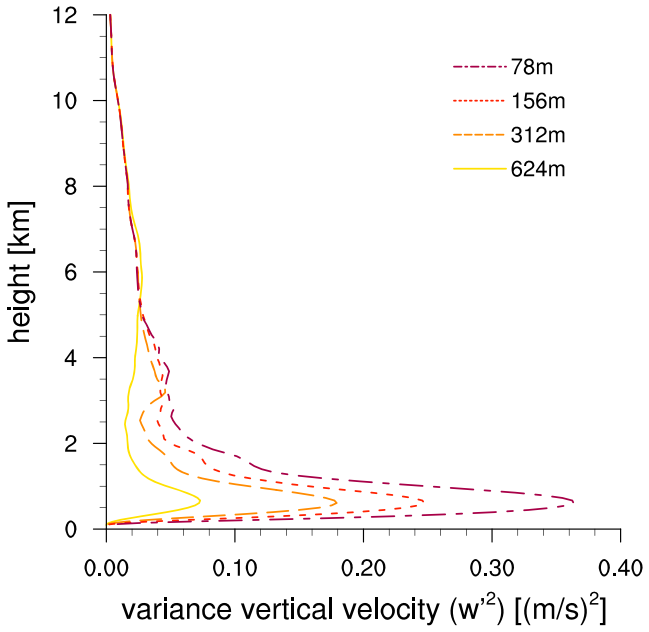

Figure 4. Mean variance of vertical wind from ICON-LOC meteogram output for different resolutions. The variance is calculated for each day and then averaged over all 9 days (29 April to 2 May 2013).

ICON-SI and DALES-SI, the representation of the boundary layer clouds in the semi-idealized simulations deviates noticably from the observed conditions in a consistent way. The rather smooth appearance and time evolution is due to the domain averaging that is applied, but additionally the semi-idealized setups emphasize the response of the small scales to the large-scale situation. The influence of mesoscales as well as a heterogeneous surface is neglected for a reduced complexity but proves detrimental for the comparison to the observations. Figure 3 suggests that these external drivers play an important role in setting the variability. For a day-to-day comparison between column observations and simulations, the realistic simulations (ICON-DE, ICON-LOC) seem to be more generally suitable than the semi-idealized simulations.

4. Methodological Biases

For the best representation of the turbulence and to facilitate comparisons with high-frequency measurements, it is helpful to simulate the atmosphere at the finest possible resolution. However, limited computational resources and a desire to simulate many different cases encourage the use of coarser simulation grids. The tension between these two demands motivates a study of the resolution dependency of our simulation output. A second question that arises is the trade-off between better resolution and the effects of variability associated with the local conditions of the measurement site. To the extent the latter is less important, it can be advantageous to use simpler and more computationally efficient semi-idealized setups, which by virtue of their reduced overhead would then allow simulations with higher resolution at the same cost. Finally, as a third question, we ask to what extent small differences in the forcing condition the response.

4.1. Resolution Effects

4.1.1. Vertical Wind

The vertical wind is fundamental for transport and is associated with both cloud and precipitation formation. Representing its variability should thus be a metric of model fitness. In the boundary layer, it mostly measures the structure of the turbulence, and above the boundary layer, it will be sensitive to the development of convection. For a quantitative idea about the effect of resolution on the vertical wind, we compare an average profile of the variance of the vertical wind from the meteogram output over all 9 days for the four different ICON-LOC simulations (Figure 4). All the simulations capture the basic structure of the vertical velocity field, but especially in the turbulent boundary layer (up to 2 km), the benefit of a higher resolution is clear. Between 2 and 4 km, only the coarsest resolution differs substantially from the finer resolution simulations, and even this difference vanishes above 5 km height. Below 2 km, differences between the two finest resolutions suggest that an even higher resolution than the 78 m will be required to fully resolve the fluctuations in the vertical velocity. On the other side, above 5 km, a 624-m model resolution might already be sufficient for most studies.

4.1.2. Liquid Water Path

As seen in Figure 2, LWP is more variable and probably more sensitive to resolution than IWV. For the LWP, two quantities are of interest—the mean amount of cloud water and its variance. In Figure 5, the difference between simulated mean (variance) and observed mean (variance) of cloud water is shown. The left panel of Figure 5 depicts the point-to-point comparison of the meteogram output and the column observations, which shows for many days an improvement with increasing resolution (e.g., for the 25 or 27 April). The shallow cumulus days (1 and 2 May) are also rather well represented, while the distribution of the almost clear sky or frontal system is more sensitive and difficult to capture. For this reason, days with more than 40% missing values or values smaller than 1 g/m^2 are highlighted. For the point measurements, we are still left with the question of how much of the differences between model and observations are due to potential mislocation in space or time of clouds (causing double penalty). To answer this question, we selected a subregion (see Figure 1), which is included in each domain of the ICON-LOC and ICON-DE and compared the

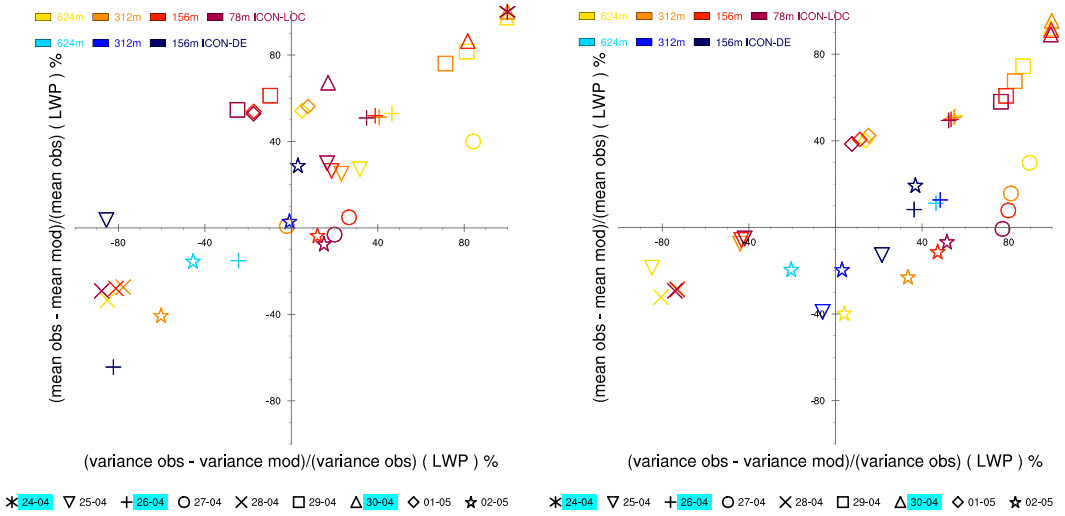

Figure 5. Percentage difference of mean LWP and variance of LWP between simulations and observations at JOYCE for each day (26 April to 2 May 2013). Model data are the meteogram output (9-s resolution, left) and the domain mean (10-min resolution, right). The colors indicate the different model setups, the symbols, and the different days. Days with more than 40% of missing values or values smaller than 1 g/m^2 are highlighted with blue.

domain mean of LWP for the different resolutions (Figure 5, right). For the domain means, the improvement by increasing resolution can be seen in the tendency for each setup to reduce the differences in mean LWP and in the variance of LWP; i.e., the symbols in Figure 5 (right) denoting higher resolution are shifted progressively toward the origin (esp. 29 April or 1 May). An interesting feature can be seen at the right panel for the ICON-LOC at the 2nd of May, where the difference in the mean LWP is decreasing, but the difference in the variance of LWP is increasing. In general, the symbols on the left plot are rather clustered around the y-axis, while the symbols on the right plot are closer to the x-axis. This supports the expected improved representation of the variability of LWP by applying the meteogram output vs. an improved representation of the general amount by taking the domain mean.

4.2. Representativeness of Column Observations

One important question for column observations is always how representative these observations are for the surrounding region. By including surface heterogeneity and mesoscale circulations, the model has the potential to tackle this question. As our main interest are clouds and their representation in the model, we continue analyzing the representativeness of LWP, as might be observed within a single column, for a larger domain, and vice versa. The question is how well the LWP distribution at one point compares to the LWP distributions of the neighboring points. To answer this, we need a measure to compare different density functions. For this, we use the Hellinger distance H, which is defined as

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2},$$

where $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_k)$ are two different discrete probability distributions. $H(P,Q) = 0$ implies that the distributions are identical, while $H(P,Q) = 1$ stands for completely disjunct distributions.

We calculated H for each day and for each grid cell in a sub-region, that is contained in all four nests, by comparing the LWP distribution of a given grid column to the reference grid column covering the observational site. For each day, the probability of the given grid column and the reference column is constructed from the temporal data, as if each measurement was an independent sample. Figure 6 shows H for each grid column averaged over all 9 days. By definition, $H = 0$ at the reference column. Even though the average is presented, all resolutions show a similarly distinct regional pattern. Higher values are apparent to the East, and there also appears wind-aligned (roughly east-west oriented) structures of small and large H. This points out the importance of taking the surface and also the meteorological conditions (e.g., wind direction) into account, as they are most likely dominating the pattern. While our statistic is still limited, the setup could be used to determine a region for which the column observations are still representative. This likely

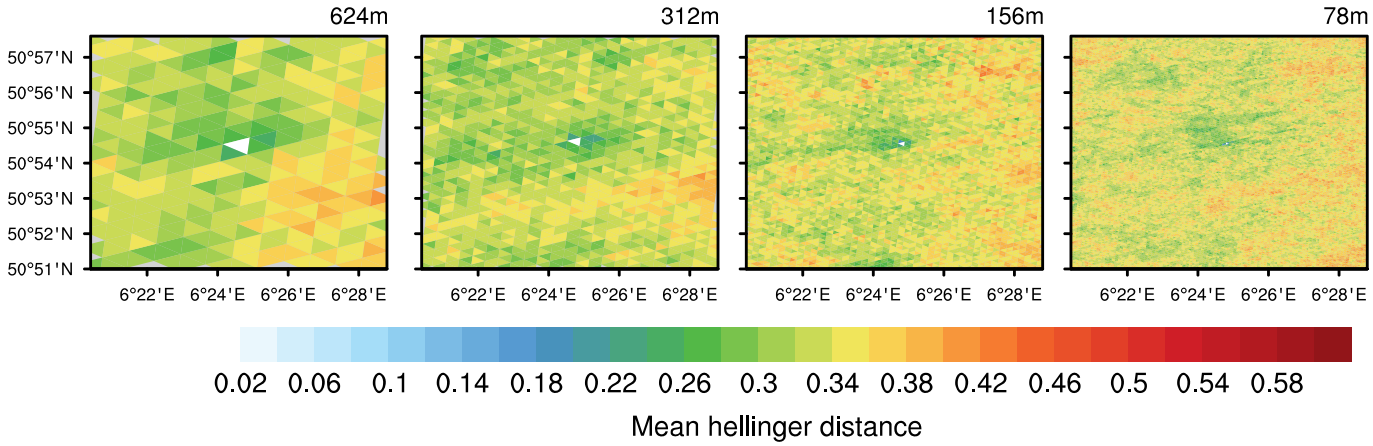

Figure 6. Mean Hellinger distance, H, for the frequency distributions of LWP with respect to the central cell closest to the location of JOYCE. The Hellinger distance has been calculated for each day (26 April to 2 May 2013) and then averaged over all 9 days.

depends on the meteorological regime, and for this, a longer time period of simulations would be beneficial. Especially in regions dominated by topography or surface features (like the open-pit mine a few kilometers East of the observation site) such a study could be informative for discussions of where to set up certain observing systems. Another application could be the investigation of changes in the pattern due to surface changes, created intentionally, e.g., the mining activities, or by climate change.

Figure 7 shows the relative bias of the domain average of LWP with respect to the observed mean and the corresponding domain mean of H. The domain mean of H is used as a measure for the similarity between the LWP density function at the grid cell of the observations and the surrounding. For H, we see two different clusters: one around $H \approx 0.25$ and one with $H \approx 0.45$. Additionally, there is a tendency toward larger H for larger mean LWP differences. So in cases where the bias is large, we also have a higher LWP variability within the domain, and the point-to-point comparison is less representative.

Figures 7 and 5 additionally show that the spread of the results for the different resolutions of ICON-DE is larger than for the different resolutions of ICON-LOC. This can most likely be explained by the more constrained forcing and setting for the small domains, compared to the large domains of the ICON-DE which allow for the representation of a wider range of scales. This can be beneficial as the mesoscale circulations have more time to evolve, but the more realistic setting, with rich mesoscale variability, also poses additional challenges to the comparison with point measurements.

4.3. Influence of the Forcing Data Set

An important question for limited area simulations (including regional climate models) is always the dependency on the large-scale forcing (e.g., Køltzow et al., 2011; Laprise et al., 2012; Warner et al., 1997). Especially the semi-idealized LES are known to depend strongly on the large-scale forcing (e.g., Gustafson et al., 2020). In this section, we will show that one advantage of the forcing at the open boundaries is a reduced dependency on the large-scale forcing. To do so, we compare the previous ICON-LOC simulations forced with COSMO-DE with an additional set of ICON-LOC simulations forced with IFS data.

Figure 8 shows the hydrometeor classification for the location of JOYCE from the COSMO-DE and the IFS, the two models used to create the local forcings. The two forecast systems produce a similar picture of the synoptic situation (cf., Figure 3a), something also shown by Barthlott and Hoose (2015), but differ substantially in their details. These differences are most pronounced in the lower atmosphere (below 4 km) where the IFS forcing supports the development of more liquid and mixed-phase clouds and precipitation in the lower boundary layer as compared to both COSMO-DE and the Cloudnet observations. The better representation of the lower atmosphere by the COSMO-DE simulations is by virtue of its much finer resolution to be expected. Our point here is not which system is better, but to then ask to what extent the LEM simulations inherit the differences apparent in the forcing data sets.

Despite differences in the host models used to produce the forcing data sets, the results of the ICON-LOC simulations forced with COSMO-DE and IFS, respectively (Figure 9), compare very well to each other. Thus,

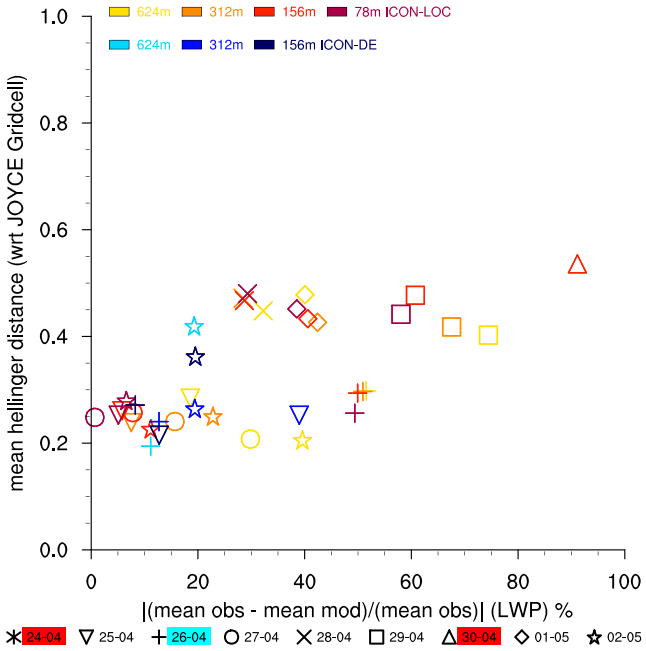

Figure 7. Absolute value of the difference (%) of domain mean LWP between simulations and observations at JOYCE and domain mean Hellinger distance for each day (26 April to 2 May 2013). Red indicates the 2 days without observed liquid clouds, and cyan the frontal day, which includes a large uncertainty in the measurements.

the differences in the forcing seem to be reduced due to the high-resolution setup. The simulations forced by the IFS seem to have an slightly enhanced precipitation frequency, suggesting that the higher amount of clouds and precipitation in the IFS itself may be partially forced. Past work has shown, in other context, that large differences can occur, as shown for an example in case of Arctic mixed-phased clouds (Schemann & Ebell, 2020). We speculate that this reflects a reduced role for surface driven turbulence and the complexity of mixed-phase clouds in those situations. In the present context of early summer convection over land, the results seem less sensitive to the forcing. The more realistic setups, which admit a larger role for the mesoscale, may also make the results less sensitive to the large-scale forcing.

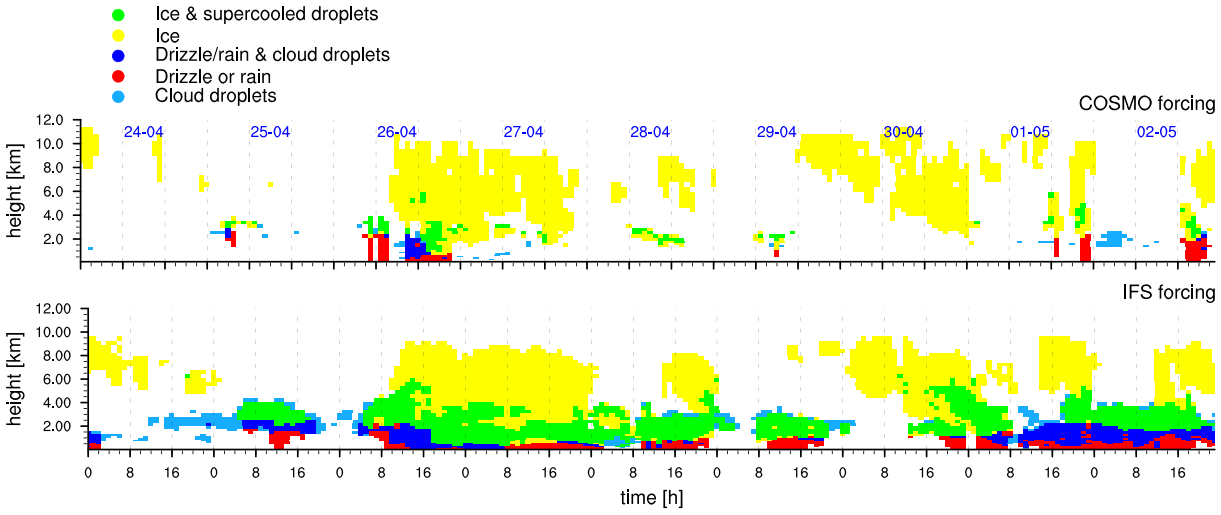

Figure 8. Hydrometeor classification at the location of JOYCE for 26 April to 2 May 2013 from the COSMO-DE forcing data (hourly data, top) and the IFS forcing data (hourly data, bottom). Model classifications are calculated based on a threshold of 1×10^{-8} kg/kg for the different hydrometeors.

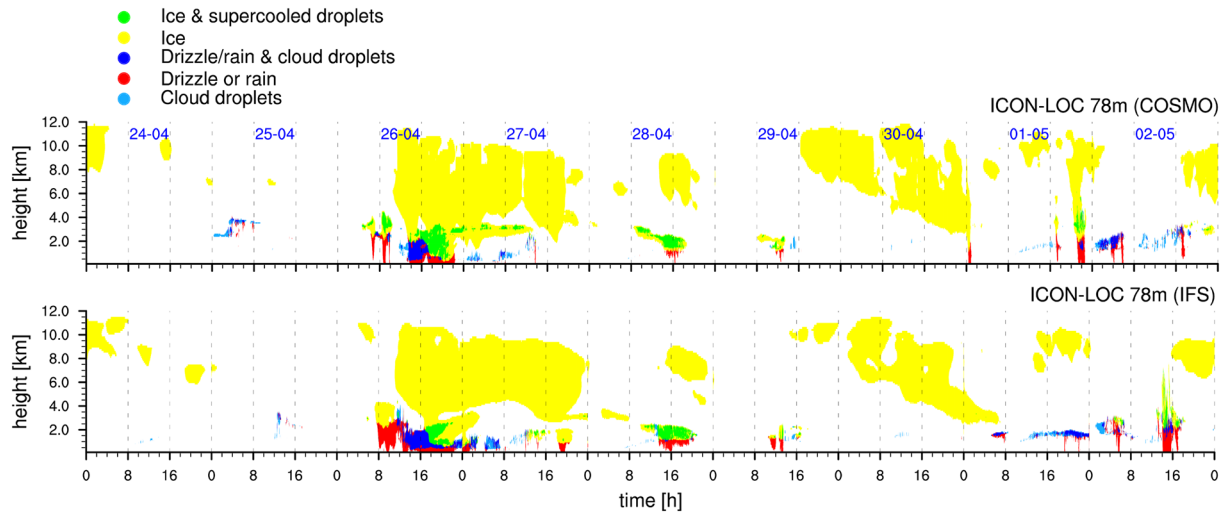

Figure 9. Hydrometeor classification at the location of JOYCE for 26 April to 2 May 2013 from ICON-LOC at 78-m resolution forced with COSMO-DE (top) and IFS (bottom). Model classifications are calculated based on a threshold of 1×10^{-8} kg/kg for the different hydrometeors.

5. Case Study: Zooming in on 2 May 2013

While large-scale forcing always plays a role, especially idealized LES are useful for highlighting particular features in a general way, e.g., shallow cumulus convection. Indeed, that is the purpose of the idealization. For this reason, we will focus in this section on 2 May 2013 where a convectively driven boundary-layer development topped with afternoon shallow cumulus was observed. This situation is typical of the type of situation often studied with LES, and the enhanced homogeneity is better suited for the application of ICON-SI and DALES-SI, allowing them to be compared to the more realistic setups in the most favorable manner possible. Our analysis focuses on the development of the cloud field and, at the end, explores to what extent differences between the ICON-LOC and ICON-SI/DALES-SI can be explained by the influence of topography alone.

5.1. Hydrometeor Classification

A more detailed assessment of the cloud classification of 2 May 2013 (Figure 10) shows that all model setups can capture the typical shallow cumulus clouds during midday. The cloud classification based on domain averages—for the semi-idealized (Figures 10e and 10f) as well as for the realistic setup (Figure 10d)—accentuates the cloud features. This is particularly pronounced for the case of the boundary layer cloud development; the semi-idealized cases emphasize the canonical development of the convective boundary layer with a growing cloud layer between approximately 12 noon and 4 p.m. (cf., Brown et al., 2002). In the meteogram output of the realistic setups (Figures 10a and 10b), the clouds are more scattered throughout the day and their representation seems to improve with resolution. The 78-m simulation with the ICON-LOC shows a cloud structure that is most similar to the observed clouds, suggesting that indeed as more detail is added to the turbulent flow and the surface representation, the simulations more closely approximate the observations.

Also, the clouds near the surface in the morning are apparent in the more realistic simulations, but either not apparent or distorted by the semi-idealized framework. Wind lidar measurements (not shown) suggest these to be decoupled from the surface. The absence of these clouds in the ICON-SI and their prevalence in DALES-SI suggests that these clouds are likely driven by differences in the large-scale flow, as DALES-SI is forced by the IFS and the ICON-SI by COSMO-DE.

For the ice clouds on 2 May 2013, more systematic differences occur. In the very early morning, all realistic simulations with COSMO-DE forcing (Figures 10b–10d) show some ice clouds between 7 and 11 km height which are not seen in the observations. These simulated ice clouds are related to ice clouds which have been observed late in the evening on the previous day and linger longer in the simulations than they did in reality. The high ice cloud seen by the observations between 7 p.m. and midnight on 2 May 2013 is well captured by the realistic setups and DALES-SI, but missed by ICON-SI. Both SI realizations use the same two-moment

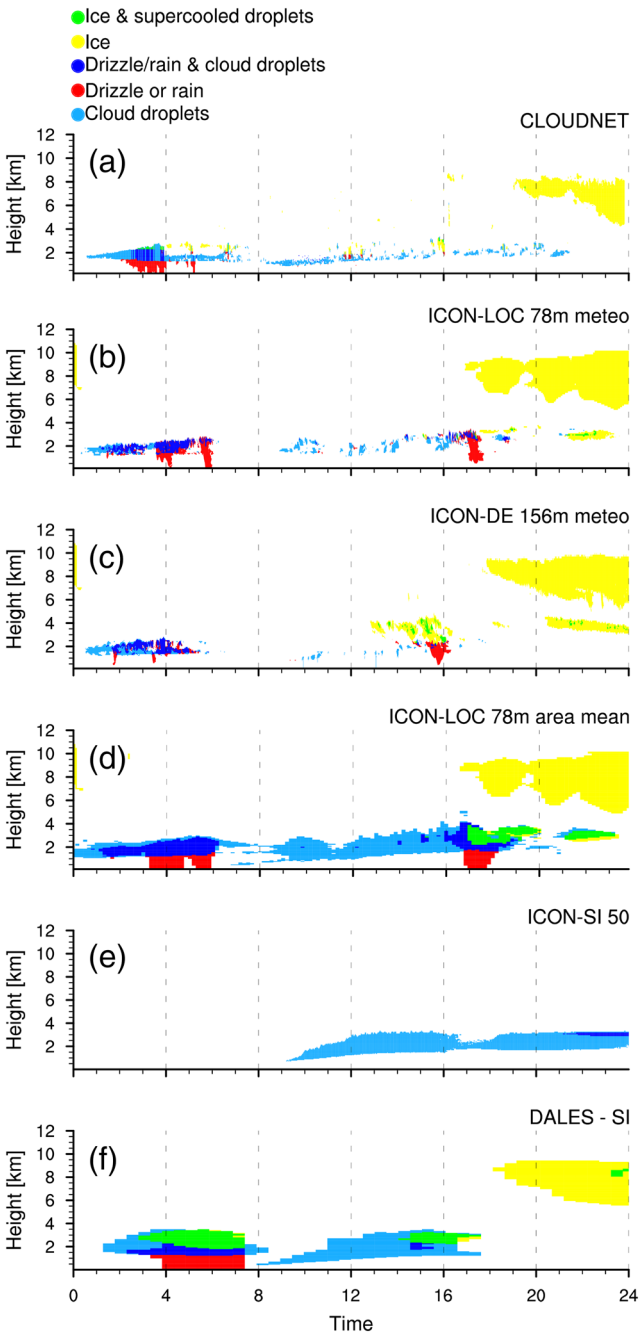

Figure 10. Hydrometeor classification at JOYCE for 2 May 2013. (a) CLOUDNET. Meteogram output of (b) ICON-LOC with 78-m resolution and (c) ICON-DE. Results based on domain mean profiles of (d) ICON-LOC with 78-m resolution, (e) ICON-SI, and (f) DALES-SI. Model classifications are calculated based on a threshold of 1×10^{-8} kg/kg for the different hydrometeors.

microphysics scheme as ICON-LOC, so this cannot explain the difference. Accordingly, it is probable that in the realistic ICON configurations, the ice cloud is due to inflow at the domain boundaries, while in ICON-SI, it is not captured by the mean nudging profile (in contrast to DALES-SI, which uses a different state [IFS] for nudging). Additionally, all realistic setups have ice/mixed-phase clouds at a height of around 4 km in the afternoon, which are less pronounced in the observations. These simulated ice clouds might trigger the precipitation development around 4 and 5 p.m. in ICON-LOC and ICON-DE which is not observed either. The ICON-DE setup produces even more ice clouds than the ICON-LOC, which leads especially for the coarse resolution to even more precipitation.

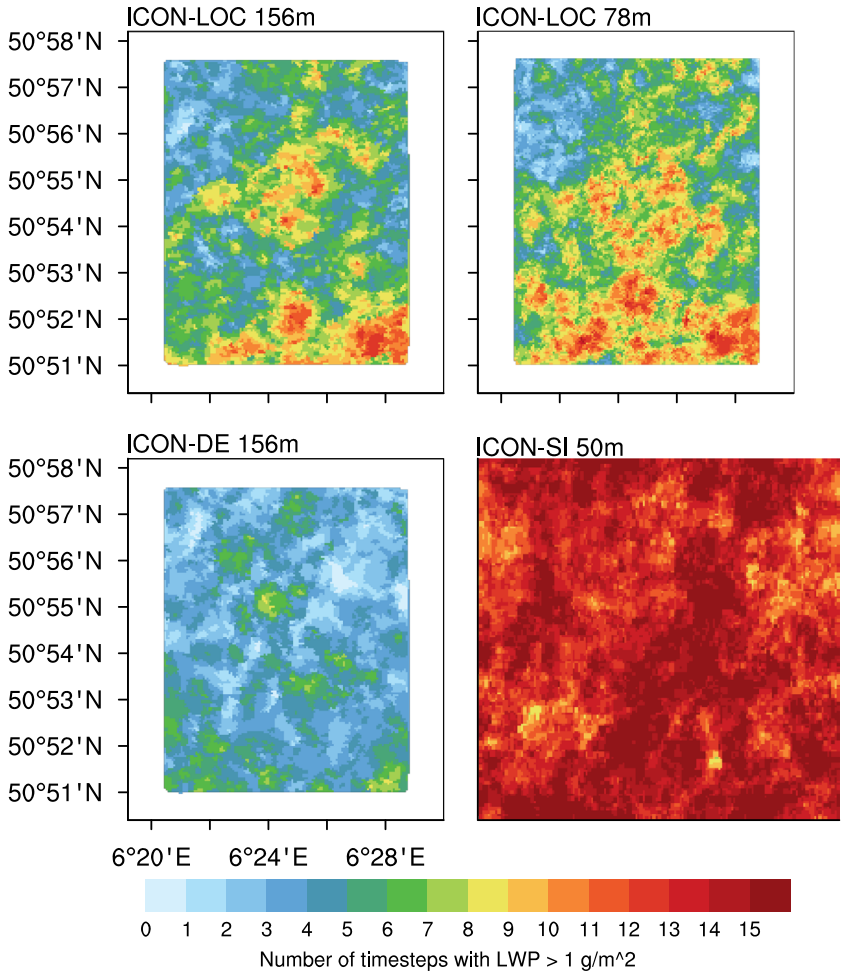

Figure 11. Number of timesteps per gridcell with LWP >1 g/m^2 between 11 a.m. and 2 p.m. on 2 May 2013 (18 time steps) for selected model setups.

Based on these analyses, the early boundary layer clouds are probably due to inflow into the boundary, the midday clouds due to typical boundary layer development and the afternoon clouds due to the influence by the topography which will be analyzed in more detail in section 5.3.

5.2. Horizontal LWP Variability

As seen in the previous section, it is difficult to establish if a disagreement between observations and simulations is due to physical reasons or due to a displacement in space or time. For liquid clouds, the assessment of the two-dimensional output of LWP can provide some insights here. We thus selected a subdomain that is included in all domains of the ICON-LOC and ICON-DE setups and counted all time steps with LWP greater than 1 g/m^2 between 11 a.m. and 1 p.m. on 2 May 2013. Figure 11 shows the occurrence of liquid clouds in the selected domain for ICON-LOC (156 and 78 m), ICON-DE, and ICON-SI. Indeed, the ICON-LOC simulations show clearly more liquid cloud cases than the ICON-DE 156-m simulations. However, by far, the most liquid clouds are counted for the ICON-SI simulation. For the ICON-LOC simulations, the amount of clouds around JOYCE (central point) is increasing with increasing resolution. The two-dimensional picture shows that the underestimation of the midday boundary layer clouds on 2 May 2013 for ICON-DE at 156-m resolution (Figure 10) is not simply due to a misplacement of the clouds. Overall, the comparison gives the impression that at least for this case, enhanced spatial variability reduces cloudiness.

5.3. Topography Experiment

To test the hypothesis that the afternoon clouds are less synoptically, and more topographically driven, we performed a sensitivity experiment with ICON-LOC at 624-m resolution where the topography

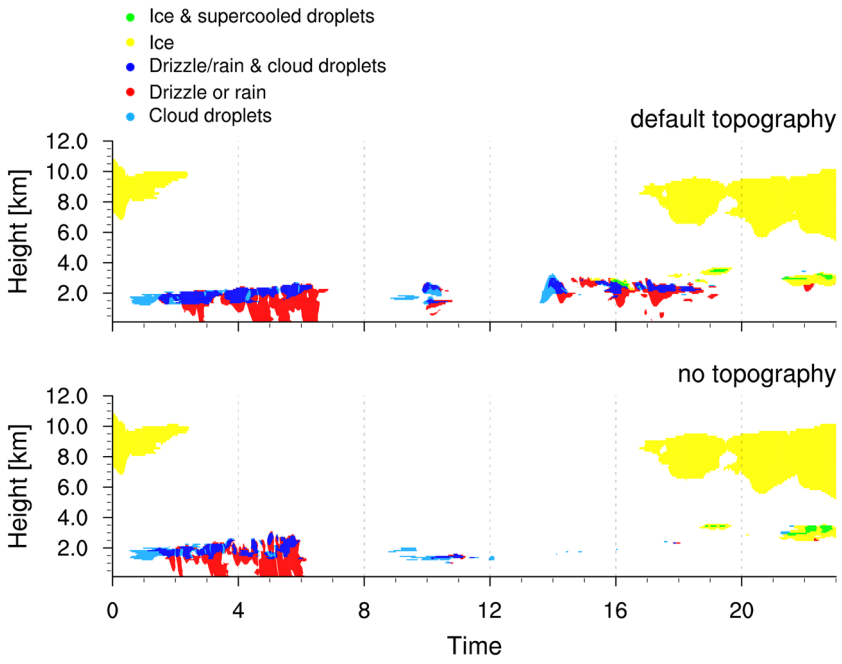

Figure 12. Hydrometeor classification at JOYCE for 2 May 2013 for ICON-LOC with 624-m resolution and (a) the default topography and (b) no explicit topography. Model classifications are calculated based on a threshold of 1×10^{-8} kg/kg for the different hydrometeors.

(Figure 1, middle) has been removed. For this, the surface height was set to 110 m in all grid cells which is approximately the surface height at JOYCE. This reduces the influence of the topography, even though some trace of it will still be present in the forcing, e.g., pressure profiles or humidity gradients. The comparison of the hydrometeor classification between the runs with and without topography (Figure 12) supports our hypothesis that the topography mainly influences the afternoon boundary layer clouds. While the morning and midday clouds are almost not at all influenced by the change in the topography, the afternoon clouds disappear in the model run without topography. The result is a litte surprsing, because the semi-idealized frameworks also lack topography but have a very strong development of fair-weather cumulus in the afternoon. We suspect that the presence of topography either contributes to the moistening or deepening of the boundary layer in ways that support cloud development. Further experiments, not shown, but with less extreme changes in topography support this finding. In the realistic configuration of the model, cloudiness increases with the strength of the topographic forcing. In some ways, this finding is counter to what we found previously, whereby the inclusion of mesoscale variability as we progressively transition from the semi-idealized to the large-domain ICON-DE simulations (e.g., Figure 11) led to a reduction in cloudiness. It suggests that the enhanced cloudiness of the semi-idealized simulations is if anything understated by virtue of their missing topographic forcing.

6. Summary and Conclusion

With the ongoing evolution of observational and computational capabilities, the interest to compare high-resolution simulations and observations on a day-to-day basis has grown (e.g., Gustafson et al., 2020; van Laar et al., 2019). Such comparisons are difficult if the models exhibit large biases in the representation of the synoptic setting. In this study, we compared three different approaches for bringing models together with observations from a fixed ground location: the traditional semi-idealized LES (ICON-SI, DALES-SI), defined as simulations without externally imposed heterogeneity, neither at the surface nor in the forcing; the more realistic setup on a very large domain (ICON-DE); and the realistic setup on a small and constrained domain (ICON-LOC). By analyzing a 9-day period in spring 2013 (26 April to 2 May 2013) in Germany, we could point out advantages and disadvantages of the various setups.

The semi-idealized LES are designed to emphasize particular flow features; this leads to a distortion—usually by overemphasis—of those features as compared to what is observed. Especially for the shallow

cumulus days, they produce, as expected, cumulus clouds on top of a well-mixed boundary layer. These setups may be suitable to analyze processes but are less well adapted to assessing their compatibility with observations, particularly over land sites with even modest heterogeneity.

The more realistic setups that take these effects into account by incorporating lateral boundary conditions from NWP models and a heterogeneous surface capture the different atmospheric conditions of the 9-day period: they show a reasonable representation of the general cloud structure, including height, time, and phase. Especially for the analyzed days when small-scale processes are more important—as the mentioned shallow-cumulus days (1 and 2 May)—higher resolution and smaller domains are beneficial for a better cloud representation. In initiating this study, we expected that the very large domain of the ICON-DE would lead to the best results, due to the possibility of freely evolving mesoscale processes. As we learned, this free evolution causes some drawbacks. It seems that a more constrained and smaller domain allows for a tighter control on the synoptic situation and may be the preferred choice if the aim is a better comparison to observations with point measurements from the surface.

Another advantage of the small domain is the relatively low computational demand, which makes it possible to run enough simulations for a statistical analysis and to investigate sensitivities by additional experiments. We shortly touched the issue of representativeness, which is a longstanding question for column observations and also gains importance due to specific output strategies, such as the meteogram output used in much of our analysis. A small domain setup as the ICON-LOC provides a reasonable representation of the cloud structure and can be used to tackle the question of representativeness in the future by using long-term simulations and, e.g., analyzing measures as the Hellinger distance to compare distributions of atmospheric variables at different points in space and time.

We highlighted the importance of including a realistic topography in the high-resolution simulations by means of a sensitivity study. Such model experiments are not only limited to changes in topography but also can be applied to changes of other surface properties, e.g., land cover, which can either be natural or man-made. The potential of the model to characterize the impact of such changes will play a large role in future research.

By comparing three different model setups with column observations, we showed the advantages and disadvantages of the different setups. An encouraging aspect of the exercise was that as more "realism" was added, either by the inclusion of finer scales of turbulence or through more realistic boundary conditions, the simulations more closely approximated the observations. Simulations over a realistic domain localized around the observational site appear to be a computationally expedient and effective way to bring modeling and observations together to develop understanding the physics underpinning how condensate forms and is distributed within atmospheric circulations.

Data Availability Statement

Data were provided by Jülich Observatory for Cloud Evolution (JOYCE-CF), a core facility funded by Deutsche Forschungsgemeinschaft via grant DFG LO 901/7-1. JOYCE-CF is an integral part of CPEX-Lab, a competence center of the Geoverbund ABC/J. The ICON-DE simulations are archived within the HD(CP)2 project. The ICON-LOC, ICON-SI and DALES simulations are stored at the long-term archive of the German Climate Computing Center (DKRZ, https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=DKRZ_LTA_1086_ds00002).

Acknowledgments
The project HD(CP)2 (High Definition Clouds and Precipitation for advancing Climate Prediction) was funded by the German Federal Ministry of Education and Research within the framework programme "Research for Sustainable Development (FONA)" under the numbers 01LK1501 to 01LK1508A and 01LK1509A. The authors gratefully acknowledge the computing time granted on the supercomputer MISTRAL at Deutsches Klimarechenzentrum GmbH (DKRZ) through its Scientific Steering Committee (WLA). The Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) is acknowledged for providing computing time on the GCS Supercomputer JUWELS at the Jülich Supercomputing Centre (JSC) under project CHKU28 for the DALES experiments used in this study.

References

Baldauf, M., Seifert, A., Foerstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, *139*, 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1

Barthlott, C., & Hoose, C. (2015). Spatial and temporal variability of clouds and precipitation over germany: Multiscale simulations across the "gray zone". *Atmospheric Chemistry and Physics*, *15*(21), 12,361–12,384. https://doi.org/10.5194/acp-15-12361-2015

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., et al. (2006). How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, *19*(15), 3445–3482. https://doi.org/10.1175/JCLI3819.1

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols. In T. F. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 571–657). Cambridge, UK: Cambridge University Press.

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C., Khairoutdinov, M., et al. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, *128*(582), 1075–1093. https://doi.org/10.1256/003590002320373210

Browning, K. A., Jonas, P. R., Kershaw, R., Manton, M., Mason, P. J., Miller, M., et al. (1993). The GEWEX cloud system study (GCSS). *Bulletin of the American Meteorological Society*, *74*(3), 387–399.

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Del Genio, A. D., Déqué, M., et al. (1990). Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *Journal of Geophysical Research*, *95*(D10), 16,601–16,615.

Chow, F. K., Weigel, A. P., Street, R. L., Rotach, M. W., & Xue, M. (2006). High-resolution large-eddy simulations of flow in a steep alpine valley. Part I: Methodology, verification, and sensitivity experiments. *Journal of Applied Meteorology and Climatology*, *45*(1), 63–86.

Corbetta, G., Orlandi, E., Heus, T., Neggers, R., & Crewell, S. (2015). Overlap statistics of shallow boundary layer clouds: Comparing ground-based observations with large-eddy simulations. *Geophysical Research Letters*, *42*, 8185–8191. https://doi.org/10.1002/2015GL065140

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

Gustafson, W. I., Vogelmann, A. M., Cheng, X., Dumas, K. K., Endo, S., Johnson, K. L., et al. (2019). Description of the LASSO data bundles product. DOE Atmospheric Radiation Measurement (ARM) user facility (*Tech. Rep.*): DOE/SC-ARM-TR-216.

Gustafson, W. I., Vogelmann, A. M., Li, Z., Cheng, X., Dumas, K. K., Endo, S., et al. (2020). The large-eddy simulation (LES) Atmospheric Radiation Measurement (ARM) Symbiotic Simulation and Observation (LASSO) activity for continental shallow convection. *Bulletin of the American Meteorological Society*, *101*(4), E462–E479. https://doi.org/10.1175/BAMS-D-19-0065.1

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., et al. (2017). Large-eddy simulations over Germany using ICON: A comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 69–100. https://doi.org/10.1002/qj.2947

Heinze, R., Moseley, C., Böske, L. N., Muppa, S. K., Maurer, V., Raasch, S., & Stevens, B. (2017). Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign. *Atmospheric Chemistry and Physics*, *17*(11), 7083–7109. https://doi.org/10.5194/acp-17-7083-2017

Heus, T., van Heerwaarden, C. C., Jonker, H. J. J., Pier Siebesma, A., Axelsen, S., van den Dries, K., et al. (2010). Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geoscientific Model Development*, *3*(2), 415–444. https://doi.org/10.5194/gmd-3-415-2010

Holloway, C. E., Petch, J. C., Beare, R. J., Bechtold, P., Craig, G. C., Derbyshire, S. H., et al. (2014). Understanding and representing atmospheric convection across scales: Recommendations from the meeting held at dartington hall, devon, uk, 28–30 january 2013. *Atmospheric Science Letters*, *15*(4), 348–353. https://doi.org/10.1002/asl2.508

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoé, J., et al. (2007). Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, *88*(6), 883–898.

Køltzow, M. A., Iversen, T., & Haugen, J. E. (2011). The importance of lateral boundaries, surface forcing and choice of domain size for dynamical downscaling of global climate simulations. *Atmosphere*, *2*, 67–95.

Laprise, R., Kornic, D., Rapaic, M., Separovic, L., Leduc, M., Nikiema, O., et al. (2012). *Considerations of domain size and large-scale driving for nested regional climate models: Impact on internal variability and ability at developing small-scale details. Climate change*. Wien: Springer-Verlag.

Löhnert, U., Schween, J. H., Acquistapace, C., Ebell, K., Maahn, M., Barrera-Verdejo, M., et al. (2015). JOYCE: Jülich Observatory for Cloud Evolution. *Bulletin of the American Meteorological Society*, *96*(7), 1157–1174. https://doi.org/10.1175/BAMS-D-14-00105.1

Macke, A., Seifert, P., Baars, H., Barthlott, C., Beekmans, C., Behrendt, A., et al. (2017). The HD(CP)^2 Observational Prototype Experiment (HOPE)—An overview. *Atmospheric Chemistry and Physics*, *17*(7), 4887–4914. https://doi.org/10.5194/acp-17-4887-2017

Marke, T., Crewell, S., Schemann, V., Schween, J. H., & Tuononen, M. (2018). Long-term observations and high-resolution modeling of midlatitude nocturnal boundary layer processes connected to low-level jets. *Journal of Applied Meteorology and Climatology*, *57*(5), 1155–1170. https://doi.org/10.1175/JAMC-D-17-0341.1

Moeng, C.-H., Cotton, W. R., Bretherton, C., Chlond, A., Khairoutdinov, M., Krueger, S., et al. (1996). Simulation of a stratocumulus-topped planetary boundary layer: Intercomparison among different numerical codes. *Bulletin of the American Meteorological Society*, *77*(2), 261–278. https://doi.org/10.1175/1520-0477(1996)077<0261:SOASTP>2.0.CO;2

Moeng, C. H., Dudhia, J., Klemp, J., & Sullivan, P. (2007). Examining two-way grid nesting for large eddy simulation of the PBL using the WRF model. *Monthly Weather Review*, *135*(6), 2295–2311.

Neggers, R. A. J., Chylik, J., Egerer, U., Griesche, H., Schemann, V., Seifert, P., et al. (2019). Local and remote controls on arctic mixed-layer evolution. *Journal of Advances in Modeling Earth Systems*, *11*, 2214–2237. https://doi.org/10.1029/2019MS001671

Neggers, R. A. J., Siebesma, A. P., & Heus, T. (2012). Continuous single-column model evaluation at a permanent meteorological supersite. *Bulletin of the American Meteorological Society*, *93*(9), 1389–1400. https://doi.org/10.1175/BAMS-D-11-00162.1

Reilly, S., Gesso, S. D., & Neggers, R. (2020). Configuring les based on dropsonde data in sparsely sampled areas in the subtropical atlantic. *Journal of Applied Meteorology and Climatology*, *59*(2), 297–315. https://doi.org/10.1175/JAMC-D-19-0013.1

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184.

Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., & Bosveld, F. C. (2015). A year-long large-eddy simulation of the weather over Cabauw: An overview. *Monthly Weather Review*, *143*(3), 828–844. https://doi.org/10.1175/MWR-D-14-00293.1

Schemann, V., & Ebell, K. (2020). Simulation of mixed-phase clouds with the ICON large-eddy model in the complex Arctic environment around Ny-Ålesund. *Atmospheric Chemistry and Physics*, *20*(1), 475–485. https://doi.org/10.5194/acp-20-475-2020

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth System Modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*, 12,396–12,417. https://doi.org/10.1002/2017GL076101

Seifert, A., & Beheng, K. D. (2006). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorology and Atmospheric Physics*, *92*(1), 45–66. https://doi.org/10.1007/s00703-005-0112-4

Stevens, B., Acquistapace, C., Hansen, A., Klinger, C., Klocke, D., Schubotz, W., et al. (2020). Large-eddy and storm resolving models for climate prediction – the added value for clouds and precipitation. *Journal of the Meteorological Society of Japan Series II*, *98*, 1–115.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Sciences*, *6*(1), 1–17.

Stevens, B., Sherwood, S. C., Bony, S., & Webb, M. J. (2016). Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, *4*(11), 512–522.

Talbot, C., Bou-Zeid, E., & Smith, J. (2012). Nested mesoscale large-eddy simulations with WRF: Performance in real test cases. *Journal of Hydrometeor*, *13*(5), 1421–1441.

van Laar, T. W., Schemann, V., & Neggers, R. A. J. (2019). Investigating the diurnal evolution of the cloud size distribution of continental cumulus convection using multiday les. *Journal of the Atmospheric Sciences*, *76*(3), 729–747. https://doi.org/10.1175/JAS-D-18-0084.1

Warner, T. T., Peterson, R. A., & Treadon, R. E. (1997). A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. *Bulletin of the American Meteorological Society*, *78*(11), 2599–2618. https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2