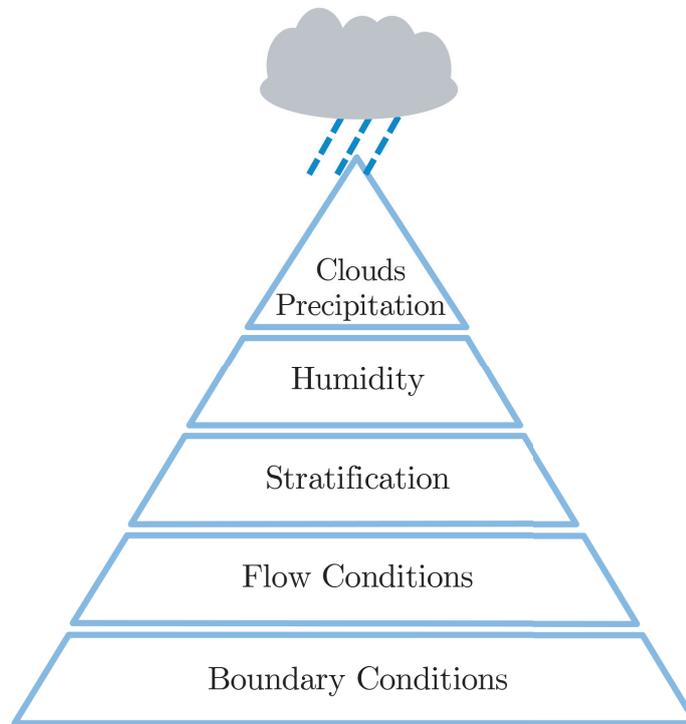# New Techniques for Ultra-High-Resolution Circulation Model Evaluation

Akio Hansen

Hamburg 2021

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49   (0)40 4 11 73 - 0
Fax: +49   (0)40 4 11 73 - 298

name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including Ph. D. theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Layout
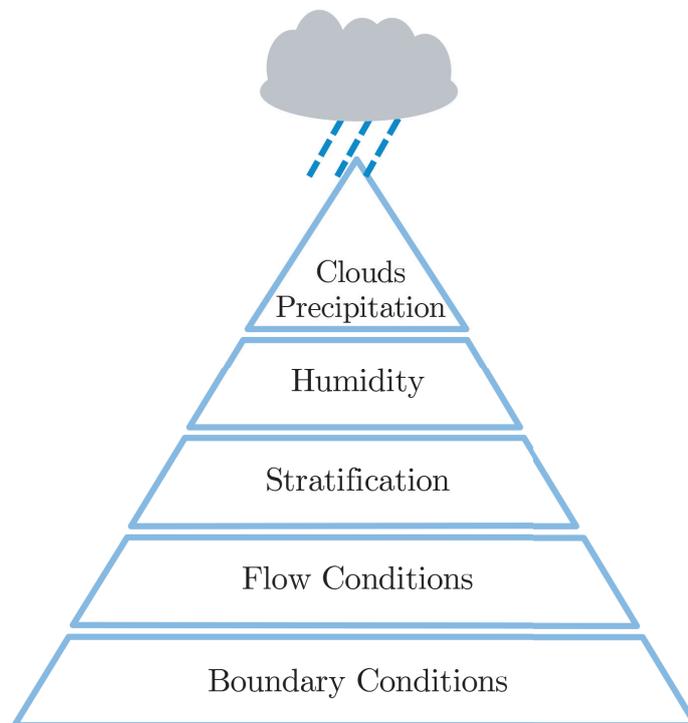
Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

New Techniques for Ultra-High-Resolution Circulation Model Evaluation

Clouds
Precipitation

Humidity

Stratification

Flow Conditions

Boundary Conditions

Akio Hansen

Hamburg 2021

Akio Hansen
aus Hamburg, Deutschland

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Universität Hamburg
Geowissenschaften
Meteorologisches Institut
Bundesstr. 55
20146 Hamburg

Abstract

A profound understanding of clouds and precipitation is crucial to reduce the largest uncertainties of current weather and climate predictions (IPCC, 2013). Relevant processes occur on scales of less than 1 km and cannot be explicitly simulated by today's climate simulations with horizontal resolutions on the order of 100 km and weather forecasts on the order of 1 km. Large Eddy Simulations (LES) for huge domains with realistic forcing are an emerging tool to bridge this gap. Nevertheless, physical consistency and realism as the prerequisite for model-based studies have to be ensured. For this, an overarching evaluation with new evaluation techniques is developed considering the demands of LES. This concept is applied to various simulations of the novel ICOsahedral Non-hydrostatic (ICON) LES model with realistic forcing data. The added value is explored through comparisons of the ICON LES with the cloud-resolving COSMO model in terms of basic atmospheric parameters and wind gusts.

Twelve days of Germany-wide ICON LES with different resolutions of down to 156 m and 2.8 km resolved COSMO simulations are used for the evaluation of the basic atmospheric state (e.g. wind, temperature, humidity). In situ observations from, for example, weather station networks and remote-sensing measurements are used as reference data. Cloud evaluation is conducted by two months of ICON LES with a resolution of down to 156 m and a circular domain of 220 km in diameter. The model output is compared with comprehensive cloud measurements and by means of the Cloudnet target classification, providing information about the cloud structure and phase. A novel cloud classification algorithm based on the direct model output is developed and applied. Additionally, physically consistent forward simulations of cloud radar, microwave radiometer, and lidar observations are performed to generate a forward-simulated cloud classification. The added value of LES regarding wind gusts compared to cloud-resolving models is explored by a one-day ICON LES case study around Hamburg with six nests down to 20 m and 20 Hz wind measurements of a boundary layer tower.

The basic atmospheric state is well represented by the ICON LES even though the well-tuned COSMO model is slightly better for most parameters and no added value is seen. In contrast to the expected higher accuracy due to the higher resolution, the errors are often largest for the finest resolved ICON LES. Overall, the simulated clouds by the ICON LES agree well with the observations at supersites. Nevertheless, frozen hydrometeors are overestimated by the ICON LES above 5 km with an ice water content of up to half an order of magnitude larger than the measurements. Additionally, liquid hydrometeors are overestimated below 5 km, detectable by an overestimated liquid water content of up to one order of magnitude. The cloud classification based on the direct model output is more practicable than the forward simulated approach with remaining technical issues. The diurnal cycle of the wind gust profiles of the 20 m resolved ICON LES show a clear added value by a good match with the observations even though not all wind gusts are explicitly resolved. A new wind gust parameterisation based on the turbulence spectrum for LES is developed and reduces the error of the simulated wind gusts by up to 60% compared to the non-parameterised model output.

Zusammenfassung

Ein umfassendes Verständnis von Wolken und Niederschlag ist essentiell zur Reduzierung der größten Unsicherheiten aktueller Wetter- und Klimavorhersagen (IPCC, 2013). Die relevanten Prozesse befinden sich auf Skalen von weniger als 1 km und können somit von aktuellen Klimasimulationen mit Auflösungen in der Größenordnung von 100 km und Wettervorhersagen in der Größenordnung von 1 km nicht explizit aufgelöst werden. Wirbelauflösende Simulationen (LES) für große Gebiete mit realistischem Antrieb sind ein vermehrt genutztes Werkzeug, um diese Lücke zu schließen. Nichtsdestotrotz müssen die physikalische Konsistenz und der Realismus als Grundlage modellbasierter Studien sichergestellt werden. Hierfür ist eine umfassende Evaluierung mit neuen Auswertunstechniken entwickelt worden, die die Anforderungen eines LES Modells berücksichtigen. Dieses Konzept wird auf verschiedene Simulationen mit dem neuen ICOsaeder Nicht-hydrostatischen (ICON) LES mit realistischen Antriebsdaten angewendet. Der Mehrwert wird anhand eines Vergleichs mit dem wolkenauflösenden COSMO Modell in Bezug auf atmosphärische Basisgrößen und Windböen untersucht.

Deutschlandweite ICON LES Simulationen mit unterschiedlichen Auflösungen von bis zu 156 m für zwölf Tage sowie 2,8 km aufgelöste COSMO Simulationen werden für die Evaluierung der atmosphärischen Basisgrößen wie z.B. Wind, Temperatur und Feuchte genutzt. In situ Beobachtungen von z.B. Wetterstationsnetzwerken sowie Fernerkundungsmessungen werden als Referenzdaten verwendet. Die Evaluierung der Wolken wird mit Hilfe einer zweimonatigen ICON LES Simulation mit einer Auflösung von bis zu 156 m und einem kreisförmigem Modellgebiet von 220 km im Durchmesser durchgeführt. Die Simulationsdaten werden mit umfassenden Wolkenmessungen und der Cloudnet Wolkenklassifikation, welche Informationen zur Wolkenstruktur und -phase liefert, verglichen. Ein neuartiger auf der direkten Modellausgabe basierender Algorithmus zur Wolkenklassifizierung wird entwickelt und getestet. Zusätzlich werden physikalisch konsistente Vorwärtssimulationen von Wolkenradar, Mikrowellenradiometer und Lidar Messungen zur Erzeugung einer vorwärtssimulierten Wolkenklassifikation durchgeführt. Der Mehrwert in Bezug auf Windböen wird mittels einer eintägigen ICON LES Fallstudie rundum Hamburg mit sechs genesteten Modellgebieten von bis zu 20 m sowie 20 Hz Windmessungen eines Grenzschichtmastes untersucht.

Die atmosphärischen Basisgrößen werden vom ICON LES gut dargestellt, auch wenn das gut abgestimmte COSMO Modell für die meisten Parameter leicht besser und bisher kein Mehrwert von ICON zu erkennen ist. Im Gegensatz zu den Erwartungen einer höheren Genauigkeit durch die höhere Auflösung, sind die Fehler für die am höchsten aufgelöste ICON LES Simulation zumeist am größten. Insgesamt stimmen die simulierten Wolken des ICON LES Modells gut mit den Beobachtungen an den Standorten der Supersites überein. Allerdings werden gefrorene Hydrometeore vom ICON LES oberhalb von 5 km mit einem um bis zu einer halben Größenordnung höheren Eiswassergehalt überschätzt. Ebenfalls werden flüssige Hydrometeore unter 5 km überschätzt, welches an

einem bis zu einer Größenordnung größerem Flüssigwassergehalt erkennbar ist. Die auf der direkten Modellausgabe basierende Wolkenklassifikation ist praktikabler als der Ansatz der Vorwärtssimulation mit verbleibenden technischen Problemen. Der Tagesgang des Grenzschichtprofiles der Windböen in der 20 m aufgelösten ICON LES zeigt einen eindeutigen Mehrwert, welches an einer guten Übereinstimmung mit den Messungen erkennbar ist, auch wenn weiterhin nicht alle Windböen explizit aufgelöst werden. Eine Windböenparametrisierung basierend auf dem Turbulenzspektrum ist für LES entwickelt worden und reduziert den Fehler der simulierten Windböen um bis zu 60% im Vergleich zu den nicht parameterisierten Modellergebnissen.

Contents

Chapter 1

Introduction

Weather and climate have a large impact on our daily lives, politics, and almost all economic sectors like agriculture, traffic, safety and many more. Therefore, accurate weather predictions are crucial among others for daily operations and an efficient integration of the varying availability of renewable energies. Precise climate predictions are the prerequisite for successful climate adaptation. Detailed information about local effects like wind gusts and changes of severe weather situations in a warming climate is especially essential considering the global trend of urbanisation with already more than half of all humans living in large cities (United Nations and Social Affairs, 2018).

However, atmospheric processes occur on a wide range of spatial and temporal scales, from long-lasting large planetary waves on scales of the order of $1000\,\mathrm{km}$ down to fast-changing, small-scale cloud microphysics on the order of $10^{-6}\,\mathrm{m}$. Therefore, current atmospheric models cannot resolve all those scales simultaneously because of limited computational resources. Today's climate models have grid resolutions on the order of $100\,\mathrm{km}$, like

Pohlmann et al. (2019), whereas numerical weather predictions (NWP) like van den Brink and Bosveld (2017); Nielsen and Gleeson (2018) and Reinert et al. (2019) have resolutions on the order of 1 km. However, relevant cloud processes act on scales of less than 1 km, and small-scale turbulence requires even finer scales. Therefore, those processes cannot be explicitly resolved because of a too-coarse grid resolution and are included by parametrisations, representing only the effects by several tuned parameters without simulating the underlying physics. Most parametrisations are based on theoretical concepts or observations of partially very idealised conditions, such as the Kansas field experiment in 1968 (Kaimal and Wyngaard, 1990), and the underlying physics are still not well understood. For these reasons, clouds and precipitation still induce the largest uncertainties of current weather and climate predictions (IPCC, 2013). Most current models do not provide information about small-scale turbulence like wind gusts because of their coarse resolutions. Nevertheless, such information is becoming increasingly relevant such as for air traffic, wind comfort in urban areas, and safety at construction sites.

Large Eddy Simulations (LES) with grid resolutions on the order of 100 m and less try to bridge this gap and resolve relevant cloud and turbulence processes explicitly. Therefore, commonly required parametrisations at climate and weather predictions such as for deep and shallow convection are no longer necessary and the induced uncertainties are avoided. Nevertheless, there remain smaller scales with unresolved processes like cloud microphysics, radiation, soil-moisture, sub-grid scale turbulence during for example a stable boundary layer (van Stratum and Stevens, 2018), and others for which LES models also need parametrisations. For a long time, the high computational demand of LES restricted them to only small-area and conceptual studies. Therefore, the first LES models were originally developed to study small-scale turbulence, like Smagorinsky (1963), Lilly (1966), and Deardorff (1970), all of whom neglected complex atmospheric processes like clouds and precipitation. The increasing computational power has enabled idealised case studies of short periods and small domains (Brown et al., 1994; Bechtold et al., 1996; Stevens et al., 2001; Brown et al., 2002; Siebesma et al., 2003; Randall et al., 2003; Ackerman et al., 2009; van Stratum and Stevens, 2018; Liu et al., 2019; van der Linden et al., 2019). Neggers et al. (2012) performed long-term LES with realistic boundary conditions for only one column. Today, the computational resources allow even for large-domain LES to study for example moist convection (Hohenegger et al., 2008; Love et al., 2011; Schlemmer and Hohenegger, 2014; Schalkwijk et al., 2015). Also, novel weather-hindcast LES runs with realistic forcing data like recent studies of Heinze et al. (2017a); Zhang et al. (2019) and Cui et al. (2019) are now feasible. However, most LES models still do not include all processes of current NWPs like a fully coupled soil-moisture model and non-periodic boundary conditions. The first global sub-kilometre simulations with the Non-hydrostatic Icosahedral Atmospheric Model (NICAM; Satoh et al. (2014)) were conducted by Miyamoto et al. (2013) to

study deep convection on a global scale. They deactivated the cumulus parameterisation to avoid any ambiguity even though they knew they might miss small cumulus clouds by this. Nevertheless, the added value of high-resolution LES, requiring large computational efforts, has to be investigated in more detail, as for instance in Stevens et al. (2020).

The ICOsahedral Non-hydrostatic (ICON) model (Zängl et al., 2015), jointly developed by the German Weather Service (DWD) and the Max-Planck Institute for Meteorology, can be used as a General Circulation Model (GCM) for climate predictions, as well as for global and limited-area weather forecasts, all sharing the same physics. The ICON model was extended by a dynamical LES core (Dipankar et al., 2015) within the project "High Definition Clouds and Precipitation for Advancing Climate Prediction" HD(CP)^2. In contrast to most simplified LES models, the novel ICON LES includes all components and parametrisations of a state-of-the-art weather prediction model like a soil-moisture model, a detailed one- and two-moment cloud microphysical scheme, and others. Furthermore, limited-area LES runs with realistic lateral boundary conditions of a weather prediction model are possible. Novel realistic weather-hindcast like LES with the new ICON LES were conducted for all of Germany for several days with a horizontal resolution of down to 156 m within HD(CP)^2. Usually, parameterised processes such as for shallow convection can therefore be explicitly resolved and uncertainties of these parameterisations avoided. The resulting high-resolution model output provides a consistent and three-dimensional dataset of the whole atmosphere, which is one of the major advantages over observations, which are limited in their spatial and temporal coverage. The goal of the project was to investigate clouds and precipitation in detail by this unique dataset to advance parametrisations of current atmospheric models. Furthermore, the ICON LES runs are used to assess the added value of high resolution climate simulations (Stevens et al., 2020).

The Germany-wide ICON LES simulations are already widely used to study different processes. For example, Brune et al. (2018) applies a wavelet analysis to the ICON LES simulations to assess convective organisation over Germany. Griewank et al. (2018) evaluates the probability density functions (PDFs) of clouds to improve them using the large domain ICON LES simulations. Different metrics to evaluate warm convective cloud fields are analysed by Bley et al. (2017). The first prototype realistic ICON LES runs are evaluated by Heinze et al. (2017a). Pscheidt et al. (2019) studies the organisation of deep convection over Germany using the simulation output. The various studies illustrate the large potential of these large-domain, high-resolution LES simulations.

Although emerging LES like the ICON simulations provide detailed insights into the atmosphere and are widely used for model-based studies, the fundamental prerequisite is the physical consistency of the model, which first has to be investigated in detail. Additionally, the realism for simulations with realistic initial and boundary conditions has

to be ensured. Most classical LES models, like the University of California, Los Angeles large eddy simulation model (UCLA-LES; Stevens et al., 2005), the Dutch Atmospheric Large Eddy Simulation (DALES; Heus et al., 2010), and the MicroHH LES model (van Heerwaarden et al., 2017) are designed for idealised cases for fundamental research on turbulence or the analysis of flow conditions within cities. Example studies considering the urban ventilation are Letzel et al. (2012) and Gronemeier et al. (2017) using the Parallelized Large-Eddy Simulation Model (PALM; Maronga et al., 2015). However, they do not include all components of a numerical weather prediction model such as a coupled soil-moisture model or different cloud phases and precipitation. On that account, LES models are often validated by idealised and well-defined setups for flows around cuboids, as for example by comparisons with well-prescribed wind tunnel reference simulations (Patnaik et al., 2009; Fischer et al., 2010; Schatzmann and Leitl, 2011; Hertwig et al., 2017). For example, the UCLA-LES is evaluated by Stevens et al. (2005) and the PALM LES model by Heinze et al. (2017b). Hanley et al. (2015) and Stein et al. (2015) performed realistic, weather-hindcast-like simulations down to 200 m and 100 m, respectively, but they only focused on cloud analysis and neglected other basic atmospheric quantities.

Concerning climate models, such as the Max-Planck Institute for Meteorology Earth System Model (MPI-ESM; Giorgetta et al., 2013), the Geophysical Fluid Dynamics Laboratory Climate Model (Held et al., 2019) and many others, a detailed and structured evaluation is conducted by routinely model intercomparisons to assess their uncertainties. One example is the Coupled Model Intercomparison Project (CMIP, Eyring et al., 2016) of the World Climate Research Programme (WCRP) to assess for model biases. Commonly used cloud-resolving models such as the Consortium for Small-scale Modeling (COSMO) model are extensively evaluated (Claussnitzer and Nevir, 2009; Pfeifer et al., 2010; Böhme et al., 2011; Akkermans et al., 2012). Also, the Weather Research and Forecasting (WRF) model is already assessed in detail (Jankov et al., 2011; Madala et al., 2014; Moya-Alvarez et al., 2018; Kalverla et al., 2019), but most of them focus only on single parameters like precipitation or single cloud properties, whereas a comprehensive evaluation is missing. Additionally, operational weather forecast models like the COSMO model covering Germany (COSMO-DE) are routinely evaluated by the national weather agencies, which is not the case for research-oriented models like LES. An extensive evaluation concept for realistic, large-domain, weather-hindcast LES considering all different quantities of the atmosphere from boundary conditions up to clouds and precipitation is still missing.

Clouds and precipitation are essential parts of an overarching evaluation to prove physical consistency but are at the same time one of the most complex quantities to evaluate due to their almost infinite number of parameters to analyse. There are several LES studies investigating clouds by parameters like cloud base, cloud top height, fraction or liquid

water path, and many more (Pressel et al., 2017; Heath et al., 2017; Angevine et al., 2018). Nevertheless, most of them neglect to check more basic atmospheric quantities like the vertical temperature and humidity profile to prove physical consistency. The Cloudnet project has developed a framework for comprehensive evaluation of cloud macrophysical properties by specially equipped Cloudnet supersites suitable for all kinds of atmospheric models. The Cloudnet target classification product is one of the most powerful integrated remote-sensing products, providing detailed information about the cloud structure and macrophysical properties, but there is so far no appropriate model output to compare it with. The valuable target classification has only been used to derive certain model quantities like the liquid or ice water path for model evaluation. The derived Cloudnet products are widely used by the community to assess, for example, the overall forecast quality of current NWP models (Hogan et al., 2009; Illingworth et al., 2015) or to look at mixed-phase clouds (Bühl et al., 2016). Boers et al. (2019) combine short- and long-wave radiation measurements with the Cloudnet vertical cloud distributions to analyse the surface radiative budget and cloud radiative forcing. Nomokonova et al. (2019) relates clouds based on the Cloudnet Target Classification to thermodynamic conditions and evaluates ICON simulations by the derived ice and liquid water path products of Cloudnet. Shallow boundary layer clouds are evaluated in studies such as Corbetta et al. (2015) using Cloudnet observations, as well as LES of the model DALES. The detailed cloud macrophysical properties of the Cloudnet target classification would be an ideally suited product for an overarching model evaluation but has so far not been directly used because of a missing model surrogate.

Along with the evaluation of cloud and precipitation processes, profound knowledge about local extremes of for example precipitation, wind, and even small-scale turbulence like wind gusts in a warming climate is of great importance for a successful climate adaptation. The information about local conditions, for example, is valuable to ensure wind and thermal comfort in densely populated areas or for engineering applications like wind turbines and safety regarding severe wind gusts. The interest in regional, high-resolution climate simulations has been increasing since its first attempts (Giorgi, 1990; Giorgi and Mearns, 1991; Jones et al., 1995; Almazroui, 2013; Jia et al., 2015; Gettelman et al., 2018). The new large-domain ICON-LES runs can resolve local processes of, for example, wind gusts explicitly, which are parameterised at weather and climate prediction models due to their coarse resolution. On that account, an overarching evaluation of realistic LES models also needs to consider the analysis of small-scale turbulence. However, most LES studies focusing on local circulations use idealised setups and only small domains. For example, the PALM model (Maronga et al., 2015) is widely used to study urban ventilation using typically idealised cases (Letzel et al., 2008; 2012; Gronemeier et al., 2017; Gronemeier and Sühring, 2019). Likewise, mostly idealised LES setups are applied to study

wind gusts (Storey et al., 2014; Knigge and Raasch, 2016; Ahmad et al., 2017; Ikegaya et al., 2017). Pantillon et al. (2020) use an ICON LES run of a winter storm to compare wind gust measurements of a light detection and ranging (lidar) instrument with the model and relate single measurements to the large-scale situation. Further studies use LES models to investigate wind gusts at cold pools (Moeng et al., 2009; Skyllingstad and de Szoeke, 2015; Hughes et al., 2015; Crosman and Horel, 2017). Additionally, common wind gust parametrisations are based on theoretical concepts, and the underlying physical processes are still not well understood, making them rising challenges to be solved (Sheridan, 2017). There is little knowledge about wind gusts within the entire boundary layer, because most studies, like Sallis et al. (2011); Valero et al. (2014); Seregina et al. (2014); Azorin-Molina et al. (2016); Brazdil et al. (2017) and Mashiko (2019), are only considering the near-ground 10 m wind gust observations of weather station networks. Suomi et al. (2015) analyse wind gusts within the first 100 m above ground, considering increasing hub heights of wind turbines. Nonetheless, small-scale turbulence processes within the boundary layer also affect, for example, the stability, fog, and boundary layer clouds, causing large uncertainties in current atmospheric models.

The novel, realistic, large-domain ICON LES runs provide detailed insights into cloud and precipitation processes, as well as into small-scale phenomena like wind gusts. However, an overarching evaluation of the model to prove its physical consistency and realism is still missing, as is the case for many other LES models. For that reason, a systematic and comprehensive evaluation concept is developed in this study for high-resolution output of LES covering the different levels of complexity of the involved atmospheric processes. The overarching evaluation adapts and enhances concepts of the mesoscale COSMO model evaluation of Hansen (2014) for the new demands of LES models. The large computational and disk space demands of realistic, high-resolution LES requires new workflows and ways to analyse the output. For example, the highest 156 m-resolved Germany-wide ICON LES domain contains almost three billion grid cells, and a single three-dimensional dump of a single variable for one time step consumes already about 13 GB disk space. Each simulated day uses roughly 91,000 node hours at the German Climate Computing Center (DKRZ).

Throughout the entire evaluation within this thesis of the ICON LES, the following overall scientific questions are addressed regarding the new opportunities and demands of large-domain, realistic LES:

1. **How can realistic high-resolution Large Eddy Simulations be evaluated?**

2. **Can an added value be identified for the realistic LES compared to state-of-the-art cloud-resolving models?**

The fundamental preconditions for physical consistency and for an accurate simulation of complex clouds and precipitation processes are well-represented basic atmospheric conditions. The precisely defined structured evaluation developed in this study ensures that all processes are right for the right reason and not because of some compensating errors. For example, if there are large errors in the temperature profile, but clouds are well represented, the clouds are right for the wrong reason. On that account, the newly developed evaluation concept starts with the basic atmospheric state consisting of boundary conditions, flow conditions, stratification, and humidity and ends up with the evaluation of most complicated clouds and precipitation, as well as small-scale turbulence. The realism of the realistic ICON LES is proven by using various quality-controlled, in situ and remote-sensing observations of instrument networks, supersites, and satellites providing an extensive basis for an in-depth evaluation. The added value of the computationally intensive realistic ICON LES is explored by comparisons with state-of-the-art cloud-resolving simulations of the well-established COSMO model for the same domain and in terms of wind gusts. The four initially simulated days with one of the first working ICON LES versions are comprehensively evaluated by Heinze et al. (2017a). The model has been further advanced and its issues solved, and the latest twelve simulated days will be extensively evaluated in this study by the new evaluation concept.

The detailed representation of clouds by realistic, high-resolution LES models raises new challenges for an adequate evaluation, as well as high demands on appropriate measurement data. Novel evaluation techniques using the powerful Cloudnet target classification product are analysed. Two different approaches to create a new consistent model surrogate, which has not been available before, are developed within this thesis. The first, the direct model output based classification algorithm, uses the vertical profiles of temperature, dew point, and specific hydrometeor masses to compute a Cloudnet-inspired cloud classification. The second approach forward simulates physically consistent all required Cloudnet remote-sensing instruments to generate synthetic measurements to which exactly the same Cloudnet algorithms are applied. For the first time, the new model cloud classification can be directly compared to the profound observed Cloudnet target classification. In addition, the evaluation by the physically consistent generated synthetic measurements avoids uncertainties of mostly statistically retrieved model quantities and thus enables detailed analysis of clouds and precipitation. These new cloud evaluation techniques are applied in the overarching evaluation of the ICON LES model for which a two-month run of a small domain of 220 km in diameter with a resolution of down to 156 m was conducted to derive profound cloud statistics.

New realistic LES simulate large parts of the turbulence explicitly and offers detailed insights into, for example, boundary layer processes, and small-scale turbulence like wind

gusts, demanding new evaluation methods to check for realism and physical consistency. Additionally, profound knowledge about the underlying physical processes, as well as wind gusts in general within the entire boundary layer, is still missing but crucial for the right representation of boundary layer turbulence by LES models. Therefore, wind gusts within the boundary layer are examined by the turbulence resolving $20\,\text{Hz}$ wind measurements of a $300\,\text{m}$ tall boundary layer tower in Hamburg. The required resolution for an explicit simulation of wind gusts by an LES is explored by a special one-day ICON LES case study with six nests down to $20\,\text{m}$ horizontal resolution. The capabilities of a correct representation of small-scale turbulence within the boundary layer is assessed as part of the overarching evaluation by the analysis of the turbulence spectra. High-resolution wind measurements of a boundary layer tower are used as reference data.

The detailed systematic evaluation concept is presented together with the performed ICON LES and COSMO simulations, and the various observational reference datasets in Chapter 2. The comprehensive evaluation of the HD(CP)^2 ICON LES runs is covered by Chapter 3, followed by the analysis of cloud macrophysical properties using the Cloud-net products in Chapter 4. In Chapter 5, the findings of the case study on the explicit modelling of wind gusts are shown. The results of the comprehensive ICON LES evaluation are summarised and discussed together with the overarching scientific questions of this study in Chapter 6. Furthermore, an outlook on future research is presented in Chapter 6.

Chapter 2

Evaluation Strategy, Simulations and Data

The new ICOsahedral Non-hydrostatic (ICON) model (Zängl et al., 2015) is a joint development of the Max-Planck Institute for Meteorology and the German Weather Service (DWD). The ICON model can be used for different atmospheric simulations, from long-term climate simulations, operational daily numerical weather predictions up to high-resolution large eddy simulations, sharing all the same code with different dynamical cores.

Independent of the scientific field, dynamical core, or resolution, the physical consistency of every model is inevitable for all model-based studies and developments. Therefore, a systematic and overarching evaluation strategy with appropriate reference data is of great value. The evaluation concept of an overarching evaluation for atmospheric models is presented in Section 2.1, and then details about the various involved simulations (Sect. 2.2, 2.3) and observational data (Sect. 2.4) are given.

2.1. Overarching Model Evaluation Strategy

Atmospheric models of different complexities try to simulate all the various atmospheric quantities like temperature, humidity, clouds, and precipitation, which all depend on and interact with each other. A systematic evaluation procedure is on that account very important, to analyse, if for instance precipitation is simulated correctly for the right reason or if compensating errors such as a wrong temperature and wrong humidity profile compensate for each other. Those compensating errors could even result in the right simulation of clouds and precipitation but of wrong physical reasons. In addition, realistic LES covering large domains deliver high-resolution model output for which often measurements of a comparable spatial and temporal resolution are lacking. This raises questions such as whether a classical baseline evaluation using ground-based weather stations is still appropriate and how such high-resolution simulations can be evaluated.
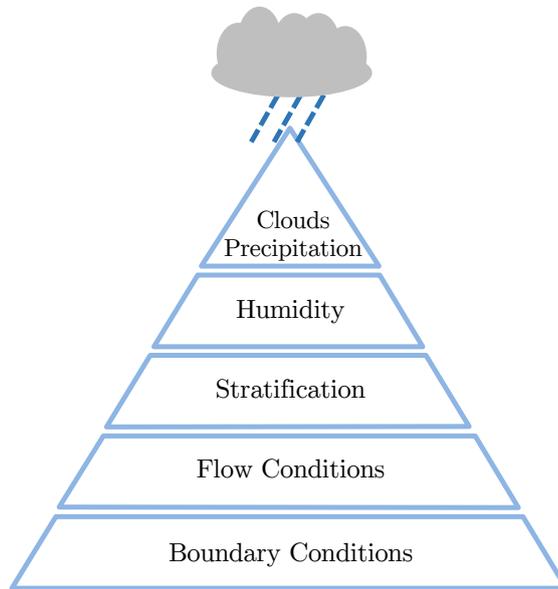
Figure 2.1.: Evaluation pyramid concept with increasing level of complexity of an atmospheric model from bottom to top, adapted from Hansen (2014).

The evaluation pyramid (Fig. 2.1) addresses these challenges and presents a generic evaluation concept for atmospheric models independent of their resolution. The proposed procedure starts with basic boundary conditions, evaluated by, for example, the mean sea level pressure field. Subsequently, the level of complexity is stepwise increased and the flow conditions may be analysed by the 10 m wind measurements of weather stations and the wind profile by soundings. The stratification and humidity can be, for example, inves-tigated similarly by near-ground temperature and humidity observations, boundary layer towers, and soundings, to name a few. The most complex processes of clouds and

precipi-tation are on top of the pyramid. This systematic approach allows analysis of such issues as whether clouds are simulated correctly because of well-represented atmospheric conditions or because of a wrong reason. This is why even for high-resolution LES, basic atmospheric quantities still need to be evaluated to ensure physical consistency. Fuzzy verification techniques, commonly used for precipitation evaluation (Ebert, 2008; Weusthoff et al., 2009), averaging model output over a certain area or period, should be applied for a phys- ical fair comparison of LES with local observations representing the same area and period, if possible. For example, a rain shower might only be shifted by a few kilometres within the model, which would be an error at a classical point-to-point comparison but still be considered correct by fuzzy methods.

Clouds can be evaluated by their various macro- and microphysical properties such as the cloud base height, liquid water content (LWC) and many more. The Cloudnet project (Illingworth et al., 2007) provides one of the most powerful and integrated multi-sensor product, the Cloudnet target classification. The Target Classification combines measurements of a cloud radar, a lidar, and a microwave radiometer to provide detailed information about the cloud structure and phase, which will be used for the comprehensive cloud evaluation in this study. Details about the performed ICON LES simulations and reference datasets of the COSMO simulations, as well as the observations, are given in the subsequent sections.

2.2. ICON Large Eddy Simulations

The novel ICON LES model with its new dynamical core was mainly developed as part of the German research project $HD(CP)^2$ (Dipankar et al., 2015). The used ICON LES configuration has terrain-following coordinates with 150 full vertical model layers. The lowest model level is at $10\,m$ above ground and has a thickness of $20\,m$. The layer thickness is stretched with increasing altitude, and the top level is at $21\,km$ in height. The triangular grid is an Arakawa-C type (Arakawa and Lamb, 1977), with the grid points at the centre of each grid cell. Advantages among others of the triangular grid type compared to a regular squared grid are the area-constant grid cells over the whole globe as well as further benefits at the nesting of model simulations (Zängl et al., 2015). The equations consider compressible air, and a three-dimensional, diagnostic Smagorinsky turbulence scheme with modifications from Lilly (1966) is integrated for the sub-grid scale turbulence to consider thermal stratification. A constant model time step of three seconds is used for all conducted ICON LES simulations in this study. Furthermore, the convection parameterisation is deactivated at all runs because of the high resolution of at least $624\,m$. The cloud microphysics is parametrised by the Seifert and Beheng (2006) two-moment bulk microphysics with six different hydrometeor classes (cloud droplets, ice, rain, snow, graupel, hail), cal-
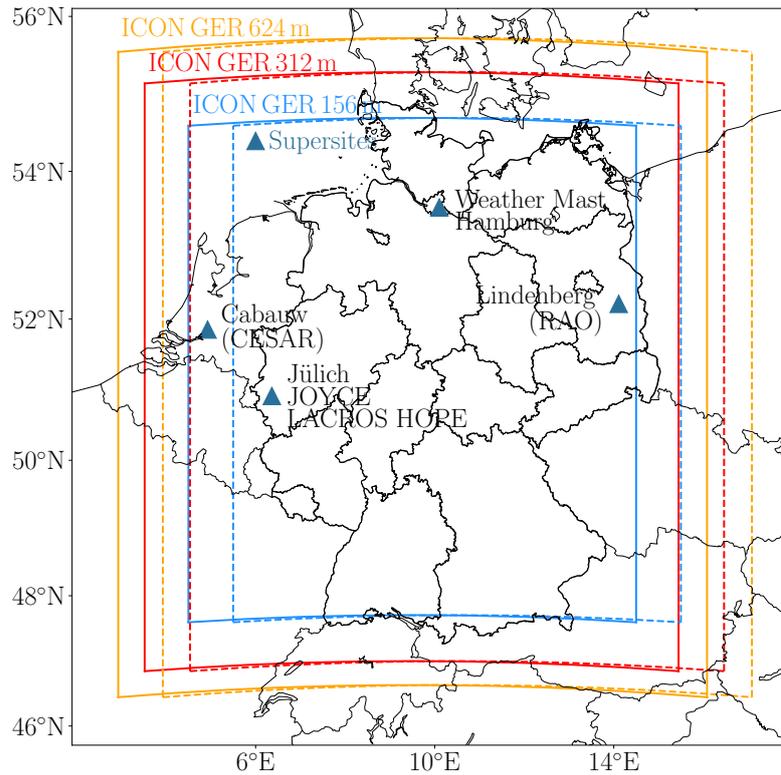
Figure 2.2.: Domains of the HD(CP)^2 Germany ICON GER (Germany) LES simulations, non-shifted domains (solid) and eastward shifted domains (dashed). Involved observational supersites at the evaluation (triangles), providing comprehensive atmospheric measurements.

culating the specific masses and number concentrations of each hydrometeor type. Further information about the ICON LES model can be found in Dipankar et al. (2015) and Zängl et al. (2015).

All conducted ICON LES simulations get their initial and boundary conditions of the operational COSMO-DE analyses of DWD covering whole Germany (Baldauf et al., 2011; 2016). The boundary conditions are updated hourly. For this study, three types of experiments with the ICON LES were performed:

1. Germany-wide ICON LES runs were conducted in the HD(CP)^2 project focusing on different research aspects (Heinze et al., 2017a). The setup consists of three one-way nested domains with 624 m, 312 m, and 156 m horizontal resolution (Fig. 2.2), which are named ICON GER (Germany) 624 m, ICON GER 312 m, and ICON GER 156 m hereafter. The outermost domain is nudged to the COSMO-DE analyses with a 20 km nudge zone. The two inner domains are nested by an eight-grid-cells-wide area at the lateral boundaries. Each simulation is initialised at 00 UTC and runs for 24 hours. Altogether, 14 bug-free single days and a continuous one-month simulation with only
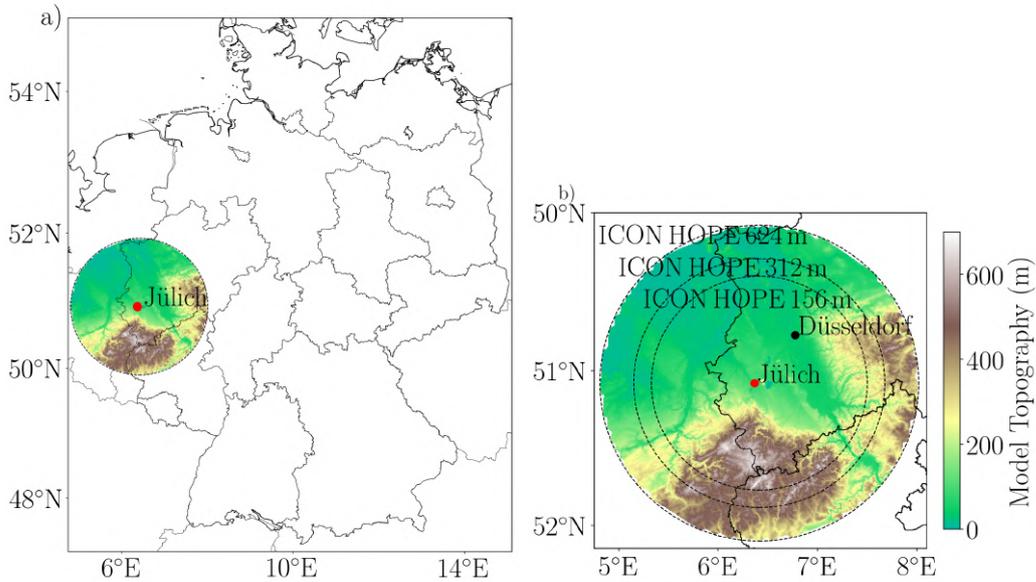
Figure 2.3.: Model topography of the ICON LES HOPE small-scale setup with three domains (ICON HOPE 624 m: 624 m horizontal resolution and 220 km diameter large domain; ICON HOPE 312 m: 312 m resolution and 180 km diameter; ICON HOPE 156 m: 156 m resolution and 160 km diameter). Overview of Germany with domains (a) and local zoom with all domains (b).

the 624 m domain for June 2016 were accomplished. Nevertheless, only the twelve days of Table 2.1 are included in this study because of some missing output of the ICON LES for some days of the overall 14 bug-free days, as well as missing COSMO reference simulations. The three simulated days of 2015 were conducted with a shifted domain towards east by approximately 65 km (Fig. 2.2) to better capture the severe convection occurred at Eastern Germany. Real and synthetic MODIS satellite images of the ICON LES 156 m can be found in Appendix A for each of the simulated days to provide an initial impression of the general weather situation.

2. A two-month ICON LES run was performed for a small domain, roughly 220 km in diameter, around Jülich (near Cologne), Germany, to generate "long-term" continuous cloud statistics. The simulated period is in April and May 2013 in accordance with the HD(CP)^2 Observational Prototype Experiment (HOPE) campaign (Macke et al., 2017), providing extensive cloud and precipitation measurements. The simulation setup is similar to the HD(CP)^2 setup (Heinze et al., 2017a; Marke et al., 2018) and consists of three one-way nested domains (624 m, 312 m, 156 m; Fig. 2.3) with initial and hourly boundary conditions of the COSMO-DE analyses. The simulations are named as ICON HOPE 624 m, 312 m, and 156 m hereafter. Each day was simulated separately, was initialised at 00 UTC, and lasted 24 hours.

Table 2.1.: Short synoptic overview of the included Germany-wide HD(CP)^2 ICON LES simulation dates of this study.

| Simulation Day	Date	Synoptic Description
1	20/04/2013	Almost clear-sky with few shallow cumulus clouds in North Germany, overcast conditions in South Germany with deep convection and light rain, Wind from north in North Germany and from east in South Germany.
2	24/04/2013	Almost clear-sky conditions in all of Germany, only few shallow cumulus clouds in North and East Germany, no rain, westerly winds all over Germany.
3	25/04/2013	Cumulus clouds in North and Central Germany with few rain showers, South Germany with only few shallow cumulus clouds, westerly winds all over Germany.
4	17/06/2014	Cumulus clouds all over Germany with rain and north to north-east wind direction all over Germany.
5	29/07/2014	Shallow cumulus clouds close to the North Sea, strong convection in Central and East Germany with intense precipitation, overcast in South Germany, wind from north-west.
6	17/06/2015	High cirrus clouds in North Germany heading to South Germany, clear-sky in Central Germany, day starting with shallow cumulus clouds close to the Alps, wind from north-west.
7	04/07/2015	High cirrus clouds over the North and Baltic Sea, remaining parts clear-sky conditions, developing thunderstorms during late afternoon with intense precipitation all over Germany, wind from south-west.
8	05/07/2015	Severe thunderstorms with strong convection in West, North and Central Germany with intense precipitation, almost clear-sky in South Germany.
9	29/05/2016	Deep convection almost all over Germany with rain, except for regions close to Poland with only few shallow cumulus clouds, south to south-easterly wind direction.
10	03/06/2016	Deep convection in Central and South Germany with single convective rain cells, shallow cumulus in North Germany, wind from east.
11	06/06/2016	Strong deep convection over Central Germany with intense precipitation, shallow cumulus in South Germany, clear-sky in North Germany, wind from east in South Germany and from north-west in North Germany.
12	01/08/2016	Convection in North and Central Germany with precipitation, shallow cumulus clouds in South Germany, wind from east.
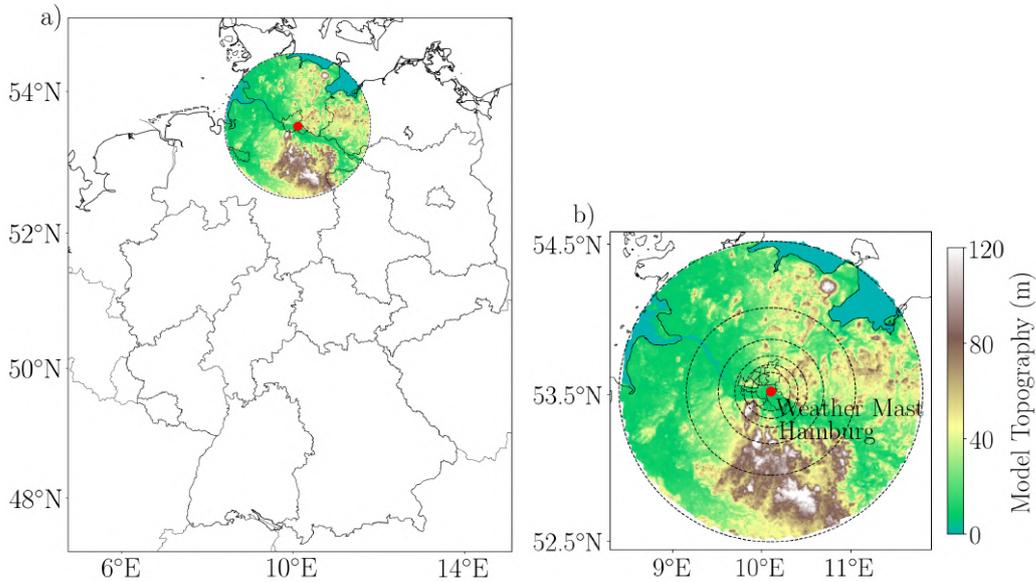
Figure 2.4.: Model topography of ICON LES small-scale setup with six domains (ICON HH 626 m: 626 m horizontal resolution and 220 km in diameter large domain; ICON HH 313 m: 313 m resolution and 110 km diameter; ICON HH 157 m: 157 m resolution and 66 km diameter; ICON HH 78 m: 78 m resolution and 44 km diameter; ICON HH 39 m: 39 m resolution and 26 km diameter; ICON HH 20 m: 20 m resolution and 18 km diameter). Overview of Germany with domains (a) and local zoom with all domains (b).

3. Another ICON LES experiment focuses on the representation of wind gusts in realistic, high-resolution LES for which a simulation with six nests with 624 m, 312 m, 156 m, 78 m, 32 m, and 20 m horizontal resolution is conducted. They are named according to their resolution and domain location hereafter: ICON HH (Hamburg) 624 m, ICON HH 312 m, ICON HH 156 m, ICON HH 78 m, ICON HH 32 m, and ICON HH 20 m. The domain is centred around the boundary layer tower of the weather mast Hamburg (Brümmer et al., 2012; Fig. 2.4) and the outer domain has roughly a diameter of 220 km, whereas the finest resolved domain has only a diameter of less than 20 km. Only one day, 24 April 2013, is simulated so far due to the large computational power and disk space required for such LES runs. The initial and hourly boundary conditions are again provided by the COSMO-DE analyses.

One big challenge of conducting such high-resolution large area LES runs is the high computational demand and disk space necessary for the output. One single three-dimensional dump of a single variable for all of Germany of the 156 m domain takes already 13 GB, which requires entirely new processing and analysing strategies like extracting only the data needed for a certain analysis, as well as new workflows. For that reason, only very selective output was written to disk for all ICON LES simulations, like the one-dimensional single column output for the locations of the sounding stations (Tab. 2.2).

Table 2.2.: Available output for the Germany-wide HD(CP)^2 ICON GER simulations with file sizes for one day; night files contain all hours between 00-06 UTC and day files all hours starting from 06 UTC.

Output Type	Number Variables	Time Resolution	Output Size All Domains
1D Columns at 33 stations	33	9 sec.	41 GB
2D Cloud variables (Day)	22	1 min.	2.8 TB
2D Cloud variables (Night)	20	5 min.	176 GB
2D Radiation variables (Day)	13	1 min.	1.7 TB
2D Radiation variables (Night)	13	5 min.	114 GB
2D Surface variables (Day)	19	1 min.	4.4 TB
2D Surface variables (Night)	19	5 min.	298 GB
3D Coarse 1.2 km (Day)	21	15 min.	965 GB
3D Coarse 1.2 km (Night)	21	1 hour	93 GB
3D Fine Snapshot (2 per day)	21	instant.	757 GB
3D Small Domain (50 x 50 km)	21	15 min.	58 GB
Total			**11.4 TB**

All two- and three-dimensional ICON LES output data are on the native triangular grid, which is directly used for the evaluation of the ICON LES simulations to avoid uncertainties due to the interpolation to a regular grid. However, for the assessment of precipitation and generation of satellite quicklooks, the ICON LES data are regridded to a regular lat-lon grid because of fewer required computational resources at the analysis. The interpolation is done by the Climate Data Operators (cdo; Schulzweida, 2019) using a first-order conservative remapping (CDO Reference manual, p. 155) conserving all properties of the regridded quantity.

2.3. COSMO Reference Simulations

The added value and general quality of the new high-resolution ICON LES simulations are compared to Consortium for Small-scale Modelling (COSMO) simulations. The non-hydrostatic COSMO model is a state-of-the-art numerical weather prediction model, used for example by the German Weather Service (DWD). The operational cloud-resolving COSMO-DE configuration (Baldauf et al., 2011; 2016) of the DWD covers all of Germany with a horizontal resolution of 2.8 km and 50 full levels up to 22 km in height (Fig. 2.5). The coarser 7 km COSMO-EU, covering all of Europe, provides the initial and hourly boundary conditions for the COSMO-DE model runs.

The COSMO reference simulations in this study are based on the COSMO-DE configuration (Baldauf et al., 2011; 2016) using the Germany-wide domain (Fig. 2.5). In contrast to the DWD COSMO-DE setup, the two-moment cloud microphysics of Seifert and Beheng
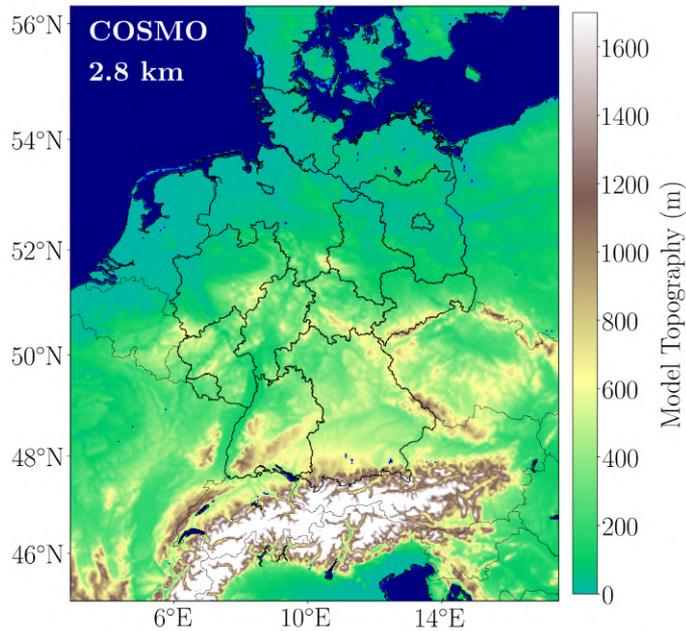
Figure 2.5.: Model topography of the Germany-wide COSMO reference simulation domain with 2.8 km horizontal resolution.

(2006) is applied instead of the operational one-moment scheme to be physically consistent with the ICON LES setup. The COSMO simulations are performed for each of the Germany-wide HD(CP)^2 ICON LES simulation days. The COSMO-EU analyses are used as initial and hourly boundary data instead of the COSMO-DE analyses applied to the ICON LES runs. The well-established COSMO reference simulations are used as reference data for a well-tuned and deep convection resolving atmospheric model on the scale of 1 km.

2.4. Observational Data

The comprehensive evaluation of the ICON LES requires a vast amount of in situ and remote-sensing observations to account for the various spatial and temporal scales of the high-resolution and large-area LES output. Therefore, different data sources from routine DWD observations to research networks and campaign data are included in the analysis. The Standardized Atmospheric Measurement Database - SAMD (Lammert et al., 2019), developed in HD(CP)^2, provides highly standardized measurement data and is used if the required data is available in the database.

The DWD ground-based weather station network consists of almost 400 stations distributed all over Germany. Depending on the instrumentation of the station, observations of the near-ground atmospheric conditions like the mean sea level pressure, 2 m temper-
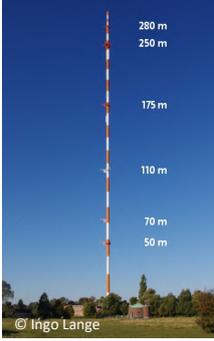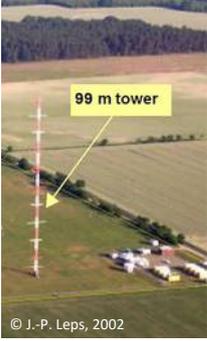
	Hamburg Weather Mast	Cabauw CESAR	Lindenberg DWD RAO
Max. Height:	280 m	200 m	98 m
Types:	Sonic	Cup anem./ Sonic	Cup anem./ Sonic
Time res.:	Up to 20 Hz.	Up to 10 Hz.	Up to 20 Hz.
Start Year:	1995	1986	1998
Area:	Urban	Rural	Rural

Figure 2.6.: Overview of main characteristics of the incorporated atmospheric boundary layer towers in this study.

ature and humidity, solar irradiance, and 10 m wind speed and direction are measured. The 14 DWD sounding stations provide detailed full troposphere profiles of the temperature, humidity, pressure, wind speed, and wind direction every twelve hours (00 UTC and 12 UTC). Six out of 14 sounding stations launch soundings every six hours at 00, 06, 12, and 18 UTC.

Several specially equipped boundary layer towers gather high-resolution atmospheric measurements across Europe. The data of the weather mast Hamburg, Cabauw, and Lindenberg (Fig. 2.6) are used in this study. The weather mast Hamburg is a 300 m tall tower at the eastern outskirts of Hamburg with six measurement platforms at $50, 70, 110, 175, 250,$ and 280 m and an additional 12 m lattice tower close by (Brümmer et al., 2012). Extensive meteorological measurements are gathered such as 20 Hz wind measurements on these platforms. The terrain is almost flat, but especially for the most common westerly winds, influenced by the city of Hamburg and climate of the North Sea. The 250 m tall tower of the Cabauw Experimental Site for Atmospheric Research (CESAR) supersite (van Ulden and Wieringa, 1996) is located in Cabauw, Netherlands and has five measurement heights at $20, 40, 60, 80,$ and 180 m (Fig. 2.6). The surroundings are flat and influenced by a maritime climate. The meteorological tower of the Richard Aßmann Observatory (RAO) of the DWD in Lindenberg consists of a 98 m lattice mast with sensors at $40, 60, 80,$ and 98 m (Fig. 2.6). The supersite is surrounded by small hills up to 100 m in height and woods.

Figure 2.7.: LACROS supersite container located at a sewage plant near Krauthausen with all Cloudnet instruments, the lidar was installed at the surface near the container, modified from Hansen (2014).

The Radar-Online-Calibration (RADOLAN) product (Weigl et al., 2004), using the 16 DWD C-Band weather radar stations in combination with online rain gauge data, provides precise quantitative precipitation observations for all of Germany. A comprehensive dataset of the spatial and temporal distribution of the integrated water vapour (IWV) is generated by the Geo Forschungs Zentrum (GFZ) Potsdam using the water vapour induced delay of the global positioning system (GPS) signal, observed at almost 300 global navigation systems stations (GNSS) across Germany (Gendt et al., 2004).

Specially equipped Cloudnet supersites (Illingworth et al., 2007) provide comprehensive and detailed macrophysical cloud measurements at several stations across the globe. Each Cloudnet supersite consists of a cloud radar, lidar, microwave radiometer, and rain gauge, which are combined to extensive cloud observations by the Cloudnet algorithms. The Cloudnet products of the Leipzig Aerosol and Cloud Remote Observations System (LACROS; Bühl et al., 2013) captured during the HOPE campaign (Macke et al., 2017) in April and May 2013 are included in this study (Fig. 2.7). The LACROS supersite consists of a 35 GHz METEK MIRA35 cloud radar (METEK, 2013), a 1064 nm Jenoptik CHMk15 lidar (Lufft, 2019), an RPG HATPRO multichannel microwave radiometer (RPG, 2011) and a rain gauge. The instruments were based at a sewage plant near Krauthausen (Jülich, Germany) during the campaign. Additionally, satellite data of the visible channels of the moderate image spectrometer (MODIS) instrument on board of the AQUA and TERRA satellite are used for a qualitative comparison of the clouds across Europe (NASA, 2019). The various measurements of the different in situ and remote-sensing instruments are of great value as reference data for the evaluation of the ICON LES model in the upcoming chapters.

Chapter 3

Baseline Evaluation

The first realistic Germany-wide large eddy simulations were conducted with the novel ICON LES model in the $HD(CP)^2$ project to investigate cloud and precipitation processes at a very high resolution. This new tool and its simulations should advance our knowledge of the underlying physics and improve current climate predictions, which suffer from their largest uncertainties due to clouds and precipitation (IPCC, 2013).

The physical consistency and realism constitute the foundation of every model-based study and have to be ensured by an overarching evaluation, especially for novel models like the ICON LES. Highly advanced atmospheric quantities like clouds and precipitation can agree well with the observations but for a wrong reason or compensating errors, which might lead to wrong conclusions of our understanding of the atmosphere. For that reason, the ICON LES is comprehensively evaluated and compared to state-of-the-art COSMO simulations. The following two specific scientific questions are addressed in this chapter:

1. **How well is the basic atmospheric state represented by the ICON GER at different resolutions?**

2. **Is there an added value of high-resolution LES detectable for basic atmospheric quantities compared to an operational weather forecast model?**

The results of the overarching evaluation in this study are compared to the assessment of the first four ICON GER prototype simulations by Heinze et al. (2017a). A short overview of the used data and error metrics is given in Section 3.1. Afterwards, the importance of the baseline evaluation is illustrated in Section 3.2 by an example of a physically unrealistic calculation of the surface momentum flux by a code error in an earlier ICON LES version, which caused a substantial overestimation of the near-ground wind speed. Based on the concept of the evaluation pyramid (Sect. 2.1), an extensive and systematic evaluation of the different basic atmospheric quantities is presented. Evaluating the high-resolution ICON GER simulations covering Germany at a resolution of down to 156 m is a big challenge in terms of computational requirements and of finding appropriate reference data, as it is shown in the following section.

3.1. Data and Methods

The used statistical metrics and observational reference datasets for the overarching model evaluation of the ICON GER simulations are described in this section. For a physically consistent model evaluation, the various temporal and spatial resolutions of the model and observational datasets have to be homogenised to a common resolution and format. Additionally, all missing data due to technical errors or missing simulation output have to be considered consistently. All datasets have to be quality checked for physically unrealistic values and measurement errors. Wherever needed, the high-resolution ICON GER output is averaged to the resolution of the observations for a physical fair comparison. In contrast, the temporal and spatial resolution of the COSMO simulations is often coarser than the measurements. The COSMO output is interpolated to the resolution of the observation to avoid degrading the ICON GER simulations to the coarse COSMO resolution and to enable still the analysis of high-resolution atmospheric processes.

The observational reference data used in this chapter consists of ground-based measurement networks, boundary layer towers, and soundings, which provide a comprehensive view of the full troposphere from the surface up to the top of the troposphere. The near-ground basic atmospheric parameters of pressure, wind, temperature, and specific humidity are evaluated using 10-minute resolved point measurements from the DWD weather stations. The observations of three boundary layer towers at Hamburg, Cabauw, and Lindenberg are

used to analyse the basic atmospheric parameters within the boundary layer. The DWD soundings, launched two or four times a day, depending on the specific site, have a high vertical resolution of only a few meters and extend the evaluation to the full troposphere up to 12 km in height. Additionally, the soundings provide a spatial representative wind direction and are accordingly used for this. The GNSS stations observe the integrated water vapour (IWV) operationally at about 300 locations across Germany with a temporal resolution of 10 minutes, which are incorporated in the evaluation of the IWV. The representation of precipitation is assessed by the RADOLAN rain radar measurements, which provide five-minute resolved precipitation rates with a one-kilometre spatial resolution all over Germany.

The various ICON GER simulation outputs are included in the baseline evaluation, and cloud-resolving COSMO simulations (Sect. 2.3) are used as a benchmark. The ICON GER output is split into three different types: single three-dimensional dumps, the one-dimensional single column output, and two-dimensional surface fields (Sect. 2.2). The single column output is only available for the locations of the soundings, the boundary layer towers, and certain airports because of the large file sizes of the ICON GER output. At these specific locations, the output contains the vertical resolution of 150 levels and a temporal resolution of nine seconds. The two-dimensional surface output has the spatial resolution of the corresponding domain, such as 156 m in the ICON GER 156 m simulation, and the output interval is 10 minutes for the first six hours from 00 to 06 UTC and one minute during the remaining simulation. The COSMO simulation results are all saved on the native 2.8 km regular grid with a temporal resolution of 15 minutes with 50 height levels up to 22 km in height (Baldauf et al., 2011).

In preparation for the baseline evaluation, the datasets are temporally and spatially homogenised. Regarding the comparison with the DWD weather station data, the nearest grid cells within a 2 km radius of ICON GER and COSMO simulations are selected and averaged to consider natural variability by being fuzzy in space. The corresponding single-column output of the ICON GER for the sounding stations and boundary layer towers is used due to the missing three-dimensional output to include a spatial fuzzy verification. The ICON GER output is averaged over time to the 10-minute interval of the station observations. As opposed to the ICON GER output, the COSMO data are linearly interpolated to the common 10-minute resolution to avoid losing information of the ICON GER simulations and of the measurements. The observed tropospheric profiles are averaged to the 150 height levels of the ICON GER grid for a fair and physical consistent comparison. The 50 COSMO output levels are interpolated linearly in height coordinates to the 150 levels because the COSMO simulations are only used as reference data and should not reduce the resolution of the ICON GER, such as of the right representation of the

boundary layer properties. To overcome the problems of a matching point-to-point comparison, a fuzzy verification in time is applied for the analysis of the soundings. For the fuzzy verification, the model output is averaged over the 30 minutes before and after the official launch time of the sounding. The time averaging also helps to solve the problems of the real profiling time of up to one hour instead of the instantaneous model output. The model output is interpolated linearly to the height levels of the different boundary layer towers. For a consistent analysis, missing values in one dataset are mirrored to all other datasets.

The ICON GER and COSMO simulations are investigated mainly in terms of the mean error (bias) and the root mean squared error (RMSE). The bias is defined as the mean difference between two time series, calculated by equation 3.1:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (F_i - M_i) = \overline{F} - \overline{M} \tag{3.1}$$

with:

- n :   total number of compared time steps
- F_i :   the i-th simulated data point
- M_i :   the i-th measurement point
- $\overline{F}$:   is the average of the simulated values
- $\overline{M}$:   is the average of the measurements.

The root mean squared error (RMSE) depicts the expected overall error including the mean difference and random part of the error and is calculated according to equation 3.2:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F_i - M_i)^2} \tag{3.2}$$

The standard deviation (STD) quantifies the variability of a value set, as in a time-series, and is defined by equation 3.3. Furthermore, the standard deviation of the observed parameters depicts the natural variability, which is used to put the calculated errors of the model simulations into a larger context.

$$\mathrm{STD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \qquad (3.3)$$

with:

- n : total number of compared values
- x_i : the i-th data point
- $\overline{x}$: is the average of all values.

3.2. Importance of a Baseline Evaluation

The systematic evaluation of the basic atmospheric state is of great importance to guarantee physical consistency of the model simulations. The value of this analysis is exemplary illustrated by a physical unrealistic definition of the surface momentum flux in an earlier version of the ICON LES code, which was detected by the baseline evaluation and is shown in the following. For this, eleven Germany-wide ICON GER simulation days of an earlier ICON LES version are analysed together with 196 available DWD ground-based weather stations. The considered time of each day is restricted to 06-18 UTC to avoid any model spin-up effects or atmospheric phenomena such as very stable boundary layers during night-time. As an additional reference, state-of-the-art cloud-resolving COSMO simulations for Germany are performed.

The flow conditions of both models are evaluated by the 10 m wind speed observations. A strong overestimation of the 10 m wind speed by more than 3 m s^{-1} is visible for most stations in Germany in the erroneous ICON GER model (Fig. 3.1). A similar overestimation is found by Heinze et al. (2017a) using the erroneous ICON GER simulations. This overestimation is independent of station height and surface orography. The substantial overestimation is not detectable for the well-established COSMO model. Similar errors were found by Matthias Brueck (personal communication) for a tropical Atlantic domain with the same ICON LES version (not shown here). The COSMO simulations show also a much smaller standard deviation of about 1–2 m s^{-1} at the DWD weather station locations compared to a standard deviation of up to 5 m s^{-1} visible for the erroneous ICON GER simulations (Fig. 3.1).

Wind speeds are usually Weibull distributed (Conradsen et al., 1984), as seen by the one-minute observations of the 10 m wind speed at the weather mast Hamburg (Fig. 3.2). The corresponding ICON GER distribution is, however, almost equally distributed and shifted to higher wind speeds (Fig. 3.2). In contrast, the range of wind speeds of the COSMO simulations matches the observations much better than those of the erroneous
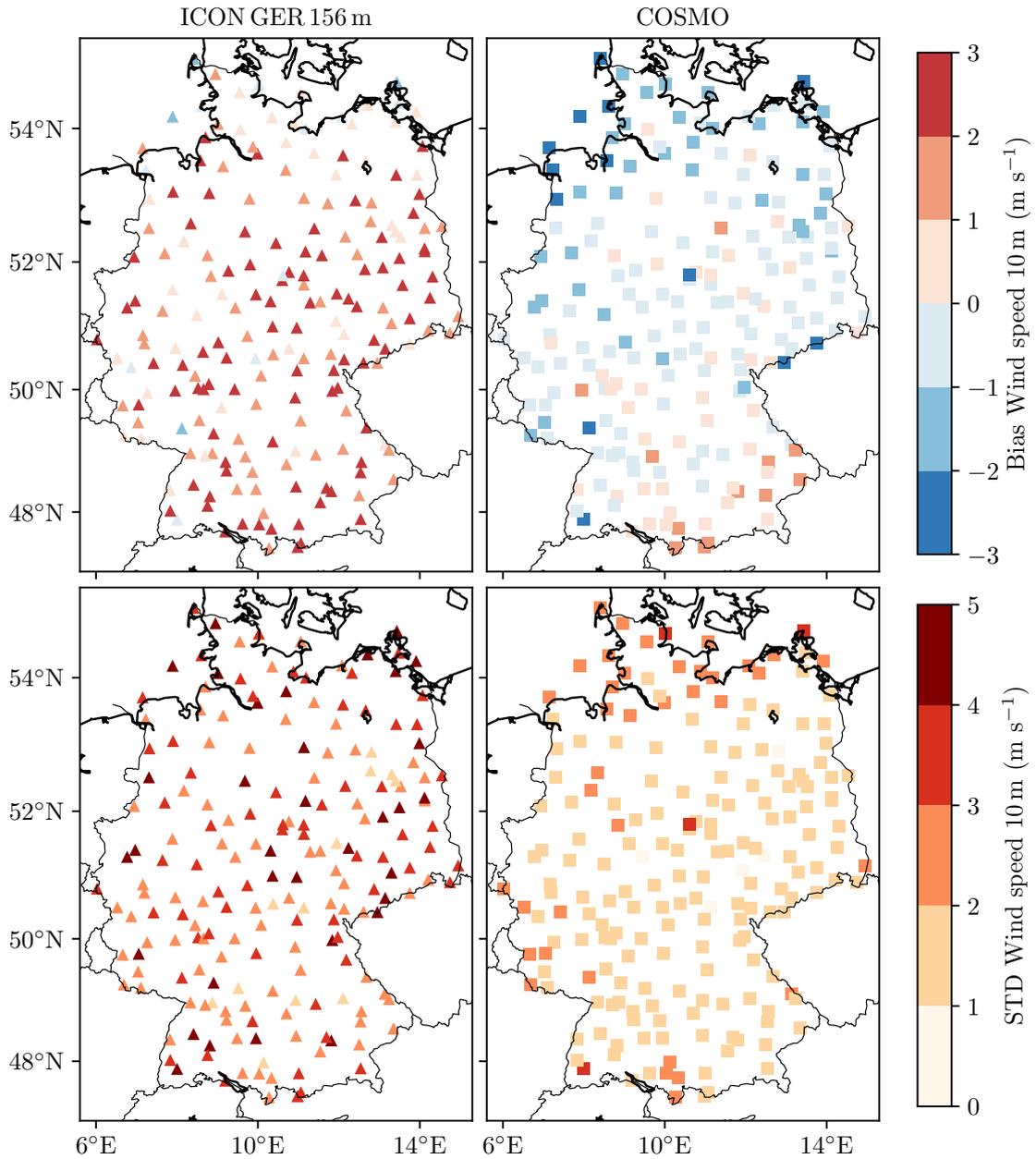
Figure 3.1.: Bias (upper) and standard deviation (bottom) of the 10 m wind speed of the erroneous ICON GER 156 m (left column) and COSMO (right column) using 196 DWD ground weather stations as reference. Only the 06-18 UTC output of eleven Germany-wide ICON GER 156 m simulations is considered. The corresponding COSMO simulation output is included for comparison.
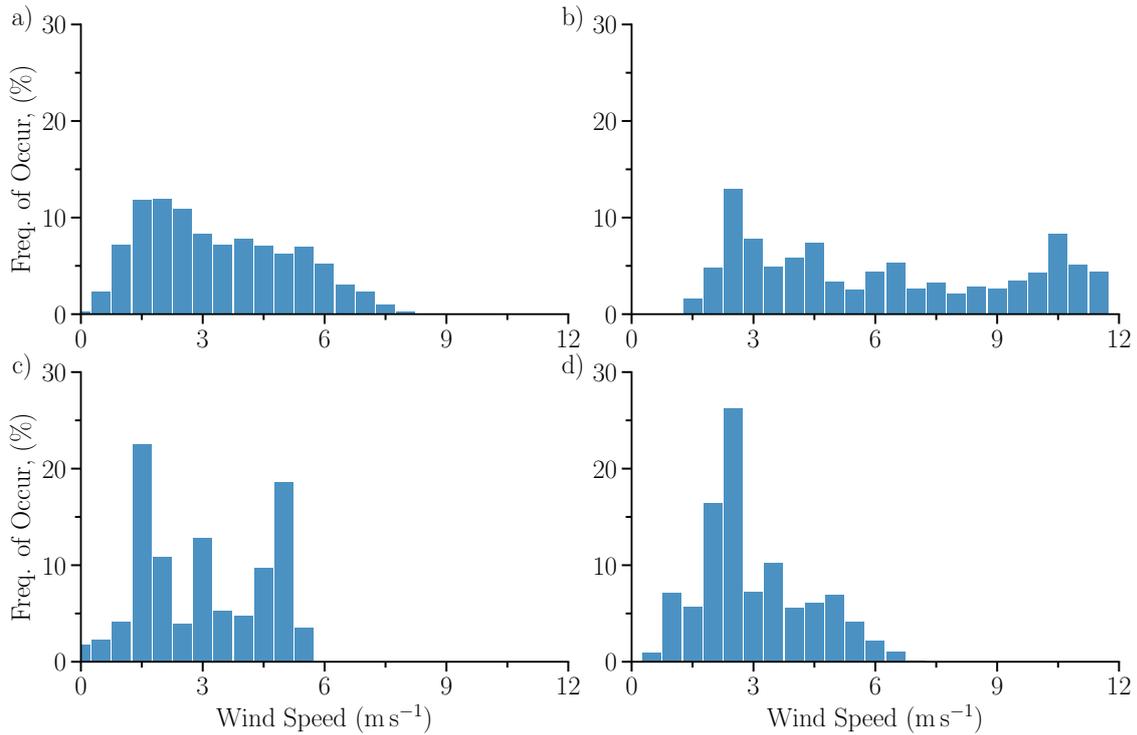
Figure 3.2.: Wind speed distributions in terms of the frequency of occurrence (Freq. of Occur.) for the observations from the weather mast Hamburg (a), the ICON GER 156 m version with the bug (b), the COSMO output (c) and the ICON GER 156 m fixed version (d), using all data of 02/05/2013. For the ICON GER fixed, only data of 00-14 UTC are included due to limited availability of the short test simulation.

ICON GER simulations. The wind speed distribution of the COSMO simulations is shifted to lower wind speeds compared to the measurements, which corresponds to the small negative bias of less than $1 \, \mathrm{m \, s^{-1}}$ at the DWD weather station locations (Fig. 3.1). Additionally, the distribution width of the COSMO simulations is substantially smaller than that of the erroneous ICON GER simulations and smaller than the measurements. Most likely, the wind speed distribution of the COSMO simulations still differs from a well-represented Weibull distribution due to the small sample size.

The stated findings indicate an error at the turbulence scheme or the transfer coefficients for which the surface momentum fluxes are analysed at the weather mast Hamburg. According to the flux-gradient parameterisation of the surface momentum fluxes within the Louis transfer scheme (Louis, 1979) of the ICON LES, the flux depends on the square of the wind speed. In theory, the flux gradient parameterisation is defined according to equation 3.4 (Stull, 1988):
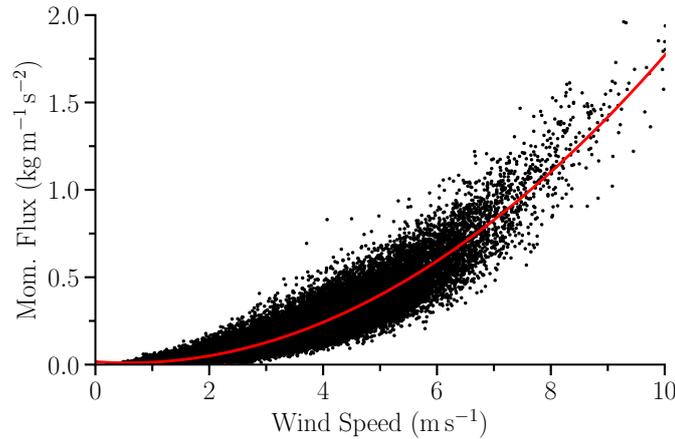
Figure 3.3.: Momentum flux (Mom. Flux) with increasing wind speed observed at the weather mast Hamburg using one-hourly observations of 2013 till 2019 including all stratification types. The red line shows a quadratic fit to the data.

Theory:

$$F = -C * \rho * |v| * v \tag{3.4}$$

Bug:

$$F = -C * \rho * \cancel{|v|} * v \tag{3.5}$$

with:

- F : Momentum flux at the surface ($\mathrm{kg\,m^{-1}\,s^{-2}}$)
- C : Transfer coefficient for momentum ($-$)
- ρ : Density of air ($\mathrm{kg\,m^{-3}}$)
- v : Wind speed ($\mathrm{m\,s^{-1}}$)

The theoretical concept of the flux gradient parameterisation is also supported by six years of momentum flux observations at the weather mast Hamburg (Fig. 3.3), which shows a quadratic relationship between flux and wind. In the erroneous ICON LES version, the surface momentum flux was only linearly dependent on the wind speed (Fig. 3.4a) instead of the squared dependence (Stull, 1988), which causes the substantial overestimation of the near-ground wind speed. The linear behaviour indicates a missing wind speed term in the flux gradient parameterisation. Indeed, the absolute value of the wind speed was missing in the earlier ICON LES version with the bug (Eq. 3.5), which was fixed afterwards.

The momentum fluxes of the corrected version show also the expected quadratic dependence on the wind speed (Fig. 3.4b) and Weibull distributed wind speeds (Fig. 3.2d). The linear dependence of the momentum fluxes underestimated the surface friction as the wind speed is mostly larger than $1\,\mathrm{m\,s^{-1}}$. Higher wind speeds were additionally less decelerated
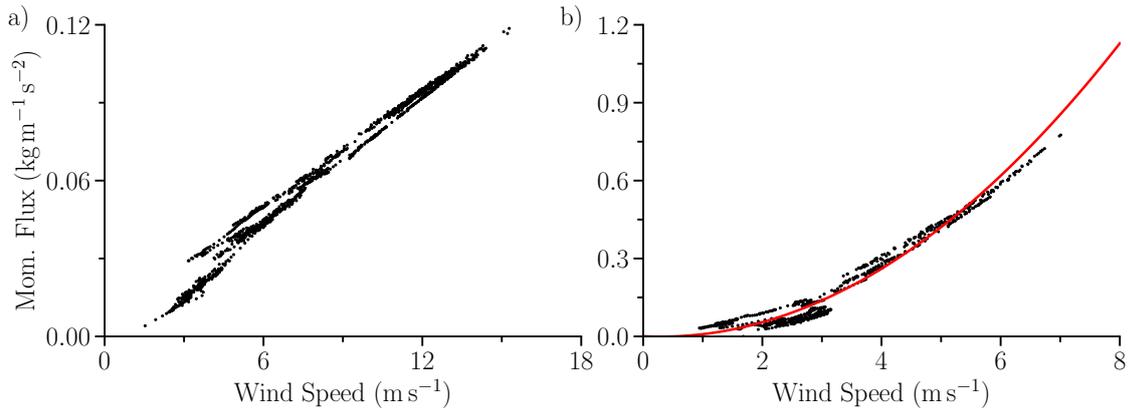
Figure 3.4.: As Fig. 3.3, but using the ICON GER single-column output of 02/05/2013 between 00-14 UTC of the version including the bug (a) and the corrected one (b). The red line shows a quadratic fit to the data. Note the different ranges of x- and y-axes.

than in reality due to the unphysical linear increase in friction, leading to the found wind speed overestimation, as seen in Figure 3.1.

Altogether, errors at the evaluation hint at a physical inconsistency in the model, that lead to inappropriate physical units in the shown case of momentum fluxes. Even though cloud fields and precipitation appeared reasonable in the erroneous ICON GER runs (not shown), such advanced quantities might be only right for wrong reasons, which highlights once more the great importance of a baseline evaluation, especially for high-resolution LES models.

3.3. Boundary Conditions

The simulation quality and all higher-level atmospheric processes of limited area models like the ICON LES decisively depend on the accuracy of the boundary conditions. One of the most important boundary condition is the pressure, as the overall geostrophic wind field is calculated based on the pressure field of the model. Accurate surface fluxes are another important boundary condition, relevant for the correct exchange of the atmosphere with the ground and accordingly for the near-ground temperature and humidity. For these reasons, the boundary conditions are the foundation of the evaluation pyramid (Sect. 2.1) and will be evaluated first. The boundary conditions of the ICON GER and COSMO simulations are evaluated in terms of the mean sea level pressure using the DWD weather station measurements within this section. The surface fluxes are not considered because of the large impact of local characteristics at the measurement location on the observations.
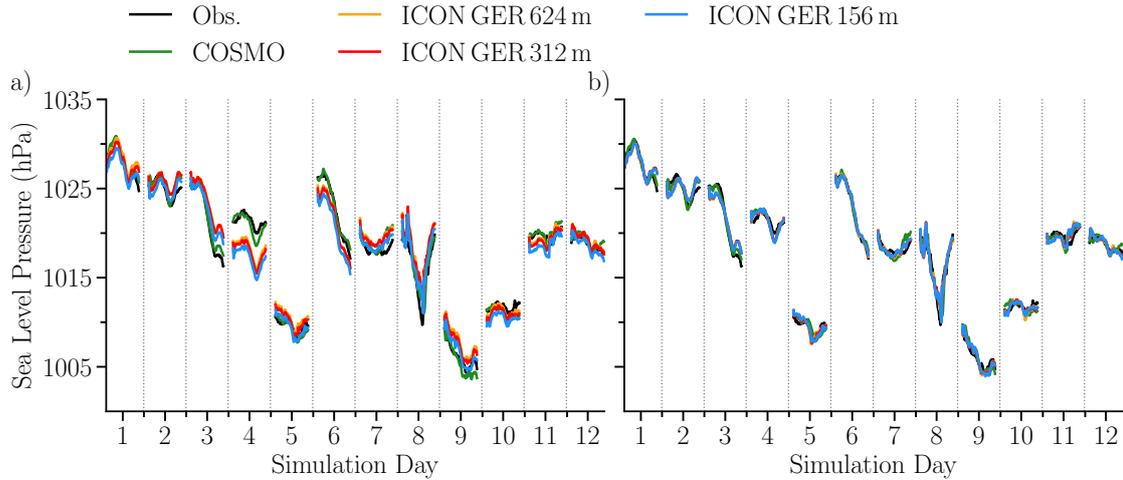
Figure 3.5.: Mean sea level pressure time series of all twelve considered ICON GER days (see Tab. 2.1 for exact dates) at the DWD Bonn weather station. Original output of all ICON GER simulations and COSMO (a) and with an applied daily bias correction (b).

The reduced mean sea level pressure is not included in the HD(CP)^2 ICON GER standard output and is for that reason calculated afterwards by the barometric formula 3.6:

$$p(h_1) = p(h_0) \cdot e^{\frac{-Mg}{RT} \cdot \Delta h} \tag{3.6}$$

with:

- h_1 : Height of target level, e.g. sea level (m)
- h_0 : Local height above sea level (m)
- $p(h_1)$: Corrected air pressure by height (Pa)
- $p(h_0)$: Local air pressure at altitude (Pa)
- M : Mean molar mass of air $= 0.0289644 \, \text{kg} \, \text{mol}^{-1}$
- g : Gravitational acceleration $= 9.806\,65 \, \text{m} \, \text{s}^{-2}$
- R : Universal gas constant $= 8.3144 \, \text{J} \, \text{mol}^{-1} \, \text{K}^{-1}$
- T : Temperature (K)
- Δh : Height difference between h_0 and h_1 (m).

The mean sea level pressure of the ICON GER and COSMO simulations is compared to all available observations from the DWD weather stations with 10-minute temporal resolution. An example time-series of the reduced mean sea level pressure for the DWD station in Bonn is illustrated for twelve days in Fig. 3.5. A distinct offset of the reduced mean sea level pressure by up to 2 hPa is clearly visible for some of the twelve simulated days at the DWD station in Bonn, such as on 17 June 2014, which is the fourth simulated day (Fig. 3.5a).

Table 3.1.: Standard deviation of daily bias corrected mean sea level pressure (MSLP) output for all ICON GER and COSMO simulations using all available DWD weather station observations and all twelve simulation days. The days with an almost factor of two higher STD of the ICON GER compared to COSMO are highlighted in gray.

MSLP (hPa) STD Corrected	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
20/04/2013	0.76	0.76	0.74	0.44
24/04/2013	0.97	0.95	0.93	0.59
25/04/2013	1.03	1.01	1.00	0.50
17/06/2014	0.55	0.55	0.57	0.51
29/07/2014	0.59	0.59	0.59	0.51
17/06/2015	0.48	0.47	0.46	0.39
04/07/2015	0.53	0.54	0.53	0.56
05/07/2015	1.23	1.26	1.26	0.96
29/05/2016	0.79	0.76	0.79	0.75
03/06/2016	0.44	0.45	0.44	0.39
06/06/2016	0.59	0.55	0.53	0.49
01/08/2016	0.57	0.57	0.57	0.41
Mean STD	0.71	0.71	0.70	0.54

The offset is independent of the station altitude or location and is consistent among all stations. In contrast, the COSMO simulations do not show any substantial difference. The offset in ICON GER output is most likely caused by issues in the initialisation or boundary data, a wrong remapping or preparation of the COSMO-DE analyses data as input for the ICON GER at those days with a substantial daily bias. Because of these issues, a daily bias correction is applied to each station individually of the ICON GER output to focus on the model error rather than on issues of the model input data. The resulting corrected time series for Bonn (Fig. 3.5b) is close to the observations and the offset corrected accordingly.

The higher resolution of the ICON GER 156 m simulation does not show any advantage over the coarser resolved simulations, and the standard deviations of the daily bias corrected mean sea level pressure are almost the same for all three resolutions with about 0.7 hPa on average (Tab. 3.1). In addition, the standard deviation is even 30% higher than the well-tuned COSMO model, indicating model issues in ICON GER. The 24 and 25 April 2013 exhibit nearly twice the standard deviation of COSMO. During both days, a high-pressure system located at the Alps is dominating the pressure field in Germany. Only small cumulus clouds and clear sky conditions were present in satellite observations (Appendix A). A wrong wind field, causing a too-slow or too-fast propagation of the pressure system, uncertainties of the COSMO-DE initialisation and boundary conditions, or issues in the model setup are possible reasons for the found differences.
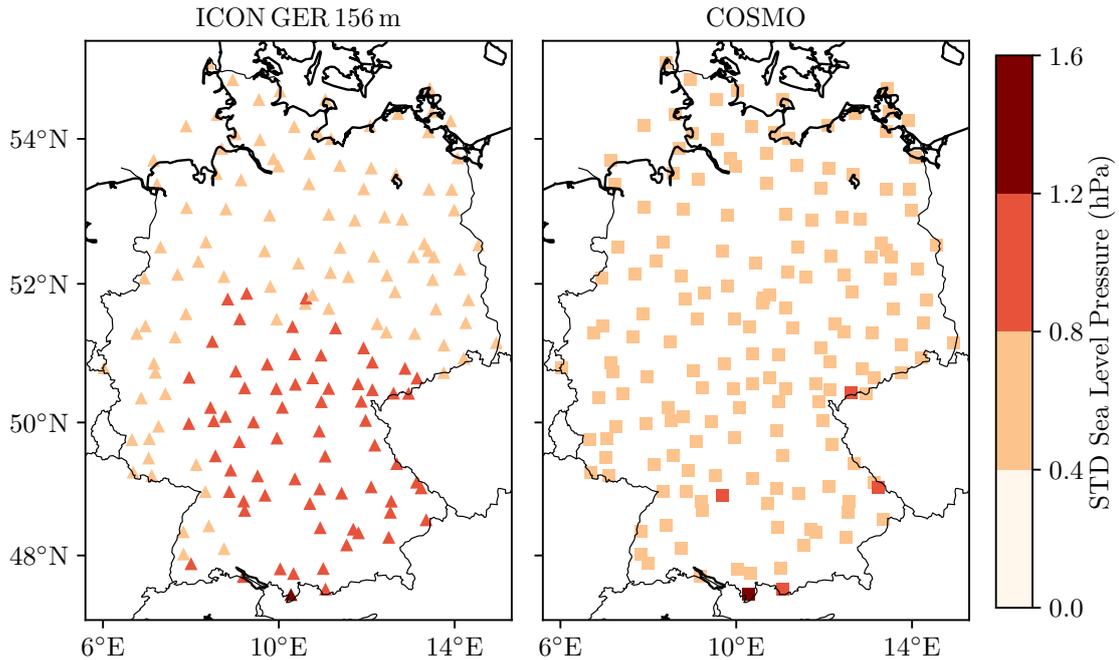
Figure 3.6.: Standard deviation of reduced mean sea level pressure of the daily bias corrected ICON GER 156 m (left) and COSMO (right) output using the DWD ground based weather stations and all twelve simulated days.

An unambiguous north-south gradient of the mean sea level pressure error is identifiable for the ICON GER 156 m, which is not the case for the COSMO simulations (Fig. 3.6). Almost all stations located at complex terrain in southern Germany have a standard deviation twice as large as for stations located at flat terrain at the ICON GER 156 m. Possibly, a not-well-adapted model physics for such high resolutions explains the large standard deviations of up to 1.2 hPa for the ICON GER 156 m simulations at regions with complex terrain.

In conclusion, a roughly 30% larger standard deviation of the ICON GER simulations compared to COSMO output highlights the importance of a further model evaluation and development, especially because the pressure field is the most fundamental quantity of every atmospheric model. Therefore, the ICON GER is currently worse than state-of-the-art cloud-resolving models, and no added value can be identified regarding the mean sea level pressure. Nevertheless, the average standard deviation of about 0.70–0.71 hPa for the ICON GER simulations is on the same order as of the well-established COSMO simulations with 0.54 hPa. Consequently, the impact of the larger errors of the ICON GER simulations on further model quantities should be rather low. Higher-order atmospheric parameters are evaluated within the upcoming sections to investigate whether this error propagates or is eventually counter-balanced by other effects.

Table 3.2.: Average standard deviation (STD) of $10\,\mathrm{m}$ wind speed across all locations of DWD weather stations for the observations, all ICON GER and COSMO simulations using all twelve days. Averaged bias and RMSE errors are calculated between the model output and the observations.

Wind Speed $(\mathrm{m\,s^{-1}})$	Obser-vations	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
STD	1.82	1.65	1.66	1.81	1.50
Bias		-0.01	0.06	0.26	-0.06
RMSE		1.47	1.47	1.51	1.40

3.4. Flow Conditions

Clouds and precipitation move mainly with the wind field at their specific height. Further, advection and exchange processes driven by momentum fluxes depend heavily on the three-dimensional wind field, showing the importance of its accurate representation by the model, which is analysed in this section. First, the near-ground characteristics will be analysed by the $10\,\mathrm{m}$ wind speed observations, followed by the investigation of the flow conditions in the boundary layer and the full troposphere.

High-resolution modelling allows for a more explicit representation of small-scale processes, such as local wind effects over complex terrain and sea-breeze effects. Therefore, the simulation accuracy should be higher due to the increased spatial resolution of the ICON GER. In contrast to these expectations, the highest resolved ICON GER $156\,\mathrm{m}$ shows the highest bias and RMSE values of all considered simulations regarding the analysis of the $10\,\mathrm{m}$ wind speed measurements of the DWD weather station network (Tab. 3.2). Differences of surface properties between the model and reality might lead to the larger bias of $0.26\,\mathrm{m\,s^{-1}}$ for ICON GER $156\,\mathrm{m}$ compared to the almost non-existing $-0.01\,\mathrm{m\,s^{-1}}$ for the coarser ICON GER $624\,\mathrm{m}$ simulation. The RMSE of the ICON GER $156\,\mathrm{m}$ is with $1.51\,\mathrm{m\,s^{-1}}$ about 7% higher than the coarser but well-tuned $2.8\,\mathrm{km}$-resolved COSMO model and shows a good agreement with the observations.

The errors could result from uncertainties in the mean sea level pressure field, where the ICON GER $156\,\mathrm{m}$ exhibits a 30% larger standard deviation compared to the COSMO simulations. An overestimated variability of the wind might explain the increasing RMSE for the higher resolved domains, as well as in comparison to COSMO (Tab. 3.2). However, in total, the $10\,\mathrm{m}$ wind speed is overall well-represented by all ICON GER simulations with a RMSE of $1.47\,\mathrm{m\,s^{-1}}$ (ICON GER $624\,\mathrm{m}$, ICON GER $312\,\mathrm{m}$) up to $1.51\,\mathrm{m\,s^{-1}}$ of ICON GER $156\,\mathrm{m}$, which is very similar to the established COSMO model with a RMSE of $1.40\,\mathrm{m\,s^{-1}}$. Additionally, the RMSE of the ICON GER simulations is about 18% lower than the natural variability of about $1.82\,\mathrm{m\,s^{-1}}$ of the observations depicted by the standard deviation (Tab. 3.2).

The spatial distribution of the bias and RMSE is analysed by means of the ground-based DWD weather stations. The errors are very similar for the ICON GER 156 m and COSMO model (Fig. 3.7) with a bias between -1.0 m s^{-1} and 1.5 m s^{-1} and an RMSE of about 1.0 m s^{-1} to 2.0 m s^{-1} for almost all considered station locations. The overestimated wind speed near the shore, as well as in central Germany, is slightly lower for the ICON GER model. Nevertheless, slightly larger errors are found for all models near the shore and in mountainous regions depicting problems in the representation of local wind effects, as for instance within valleys or on top of hills. The slightly larger errors at the complex terrain compared to flat terrain of the ICON GER 156 m are in accordance with the larger uncertainties of the mean sea level pressure. The slightly increased RMSE values near the shore are not found at the mean sea level pressure and are most probably related to shortcomings in the representation of local effects like sea-breeze circulations. Heinze et al. (2017a) assessed the first erroneous ICON GER prototype simulations with the momentum flux bug (Sect. 3.2) and found bias values of up to 3 m s^{-1} and RMSEs of up to 5 m s^{-1} at the locations of the ground-based weather stations. Accordingly, the fixed momentum flux bug ICON LES version, used in this study, depicts a substantially higher accuracy at the representation of the near-ground wind speed with bias values of less than 1.5 m s^{-1} and RMSE values in the range of 1–2 m s^{-1}.

The 10 m wind speed, averaged over all locations of ground-based weather stations, is used to evaluate the diurnal cycle of the wind speed at 10 m height of the ICON GER and COSMO simulations (Fig. 3.8). The simulated diurnal cycle depends strongly among others on an accurate representation of the stratification, turbulence, and coupling of the different model levels. Most likely a too-strong coupling of the different model levels during night of all ICON GER simulations leads to a too-large downward mixing of momentum from the free troposphere down to the bottom layers, which results in a remarkable overestimation of up to 0.5 m s^{-1} and 20%, during the night. Similar, the too-strong coupling of the model levels might cause the missing decrease of the 10 m wind speed after sunset around 18 UTC of the ICON GER simulations. The COSMO output is in good agreement with the observations during night time and captures well the decoupling of the bottom layers after sunset, leading to the correct decrease in wind speed.

In opposition to the too-strong coupling during the night, probably a too-weak coupling during the daytime of the ICON GER and COSMO simulations causes an underestimated downward mixing of momentum and might explain the underestimated 10 m wind speed by up to 0.5 m s^{-1} and 20%, of all simulations (Fig. 3.8). Presumably, a too-low surface roughness depicted by a too-low roughness length z_0 (Stull, 1988) of the 156 m resolved ICON GER simulations causes the constantly higher 10 m wind speeds than those of the coarser resolved ICON GER simulations. The higher wind speeds of the ICON GER 156 m
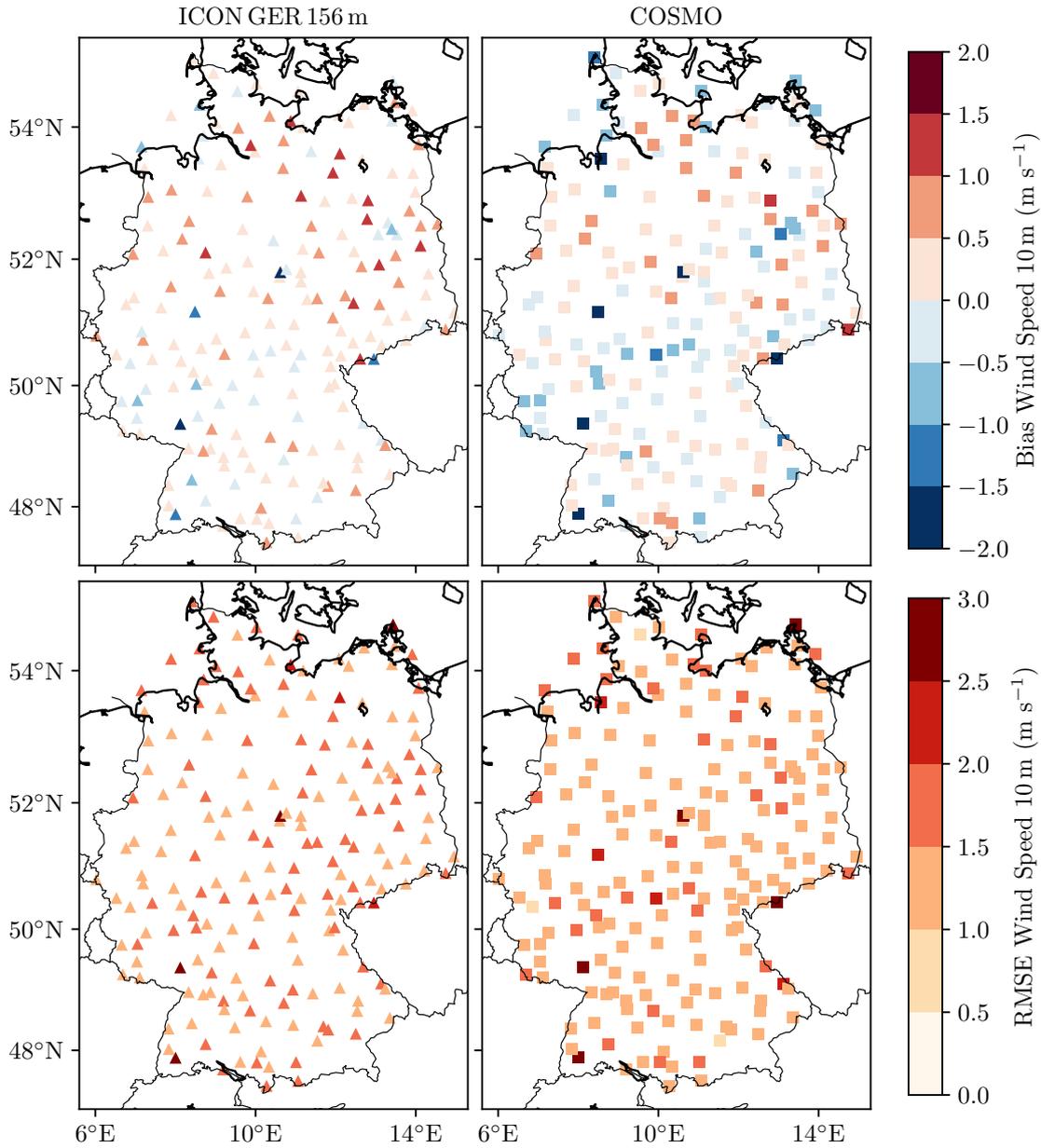
Figure 3.7.: Bias (upper row) and RMSE (bottom row) of 10-minute average 10 m wind speed of all available ground based DWD stations using all twelve days for the ICON GER 156 m (left column) and COSMO simulation output (right column).

are also consistent with the larger positive bias, which was found at the average bias across all stations and days (Tab. 3.2). Most likely, the larger positive bias of the ICON GER 156 m explains the missing underestimation of the 10 m wind speed at daytime compared to the observations despite the too-weak coupling of the model levels. In opposition, the coarser resolved ICON GER simulations show the expected underestimation of the

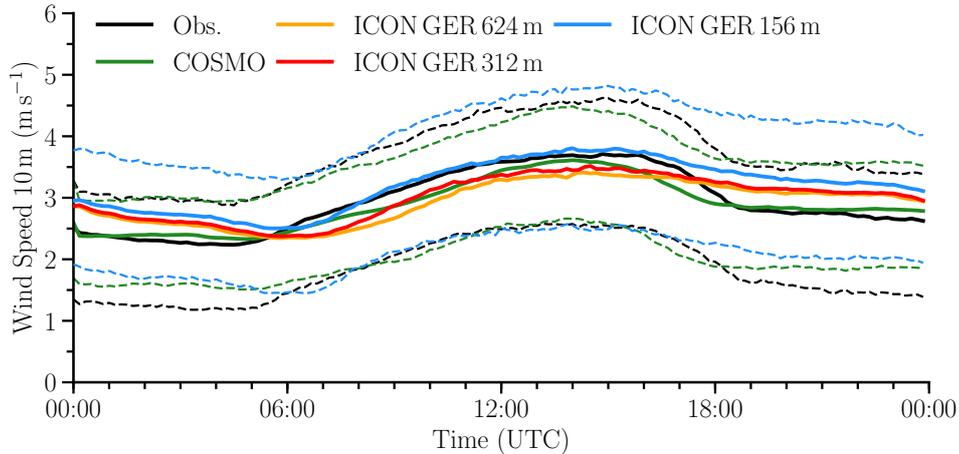**Figure 3.8.:** Mean diurnal cycle of the wind speed in 10 m height averaged over 10-minute intervals for the observations, all ICON GER and COSMO simulations using all locations of the DWD ground-based weather stations. The upper dashed lines shows the 95th percentile and the lower dashed lines the 5th percentile of all data, shown for the observations (black), COSMO (green), and ICON GER 156 m (blue).

10 m wind speed during the day as a consequence of the too-weak coupling. Nevertheless, the ICON GER 156 m still shows the largest errors at night due to the large positive bias. The too-weak coupling of the lower model levels by the COSMO simulations might also cause a higher influence of the near-ground levels leading to a too-early decoupling and decrease of the 10 m wind speed already around 15 UTC.

Inaccuracies of the COSMO-DE initialisation data applied to the ICON GER runs already induce an overestimation of the wind speed by up to $0.5\,\mathrm{m\,s^{-1}}$ at the beginning of all ICON GER simulations. In contrast, the COSMO-EU initialisation data used for the COSMO simulations seem to be more accurate, and no initial overestimation of the wind speed is identifiable for the COSMO output. A good representation, as also seen by the low bias values in Tab. 3.2, of local conditions might explain the general good agreement of the spread of the wind speeds across the weather stations for the ICON GER simulations apart from the abovementioned offset due to the general overestimation. In contrast, the COSMO model slightly underestimates the spread (Fig. 3.8).

The wind speed measurements of the weather mast Hamburg, Cabauw, and Lindenberg towers are used to evaluate the simulated wind profiles within the boundary layer of the ICON GER and COSMO simulations. The first 1,000 m of the troposphere are represented by 20 model levels by the ICON GER model, whereas the COSMO model contains only 13 model levels up to 1,000 m height. The additional model levels of the ICON GER compared to the COSMO simulations in combination with the explicit simulation of large eddies and the higher spatial resolution should lead to an overall better representation of the turbulence-dominated boundary layer compared to an Reynolds Averaged Navier
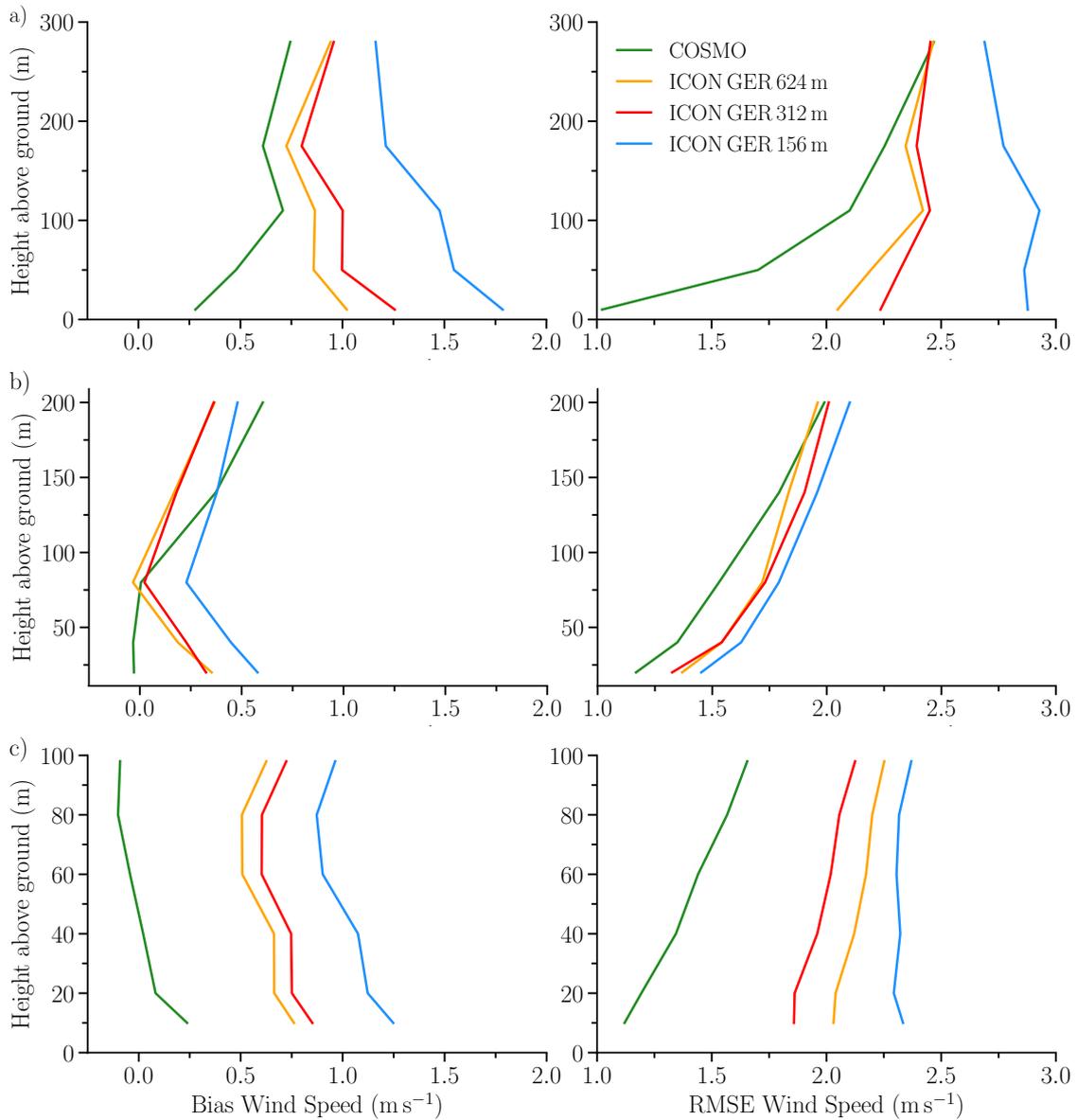
Figure 3.9.: Bias and RMSE of wind speed profiles at three boundary layer towers: weather mast Hamburg (a), Cabauw (b) and Lindenberg (c) using 10-minute averaged observed data and corresponding ICON GER and COSMO single-column output.

Stokes (RANS) model like COSMO. In contrast to this hypothesis, the highest resolved ICON GER 156 m simulation exhibits the largest bias and RMSE values at all three boundary layer towers (Fig. 3.9).

All ICON GER domains overestimate on average the wind speed within the boundary layer between $0.2\,\mathrm{m\,s^{-1}}$ and $1.7\,\mathrm{m\,s^{-1}}$, whereas COSMO has almost no bias at all. Except for the Cabauw tower, the RMSE of the ICON GER 156 m simulation is almost by a factor of two

larger than the RMSE of the COSMO simulations. The higher RMSE is distinctly illustrated in Figure 3.9a for the weather mast Hamburg with an average RMSE of $1.75\,\mathrm{m\,s^{-1}}$ for COSMO and $3\,\mathrm{m\,s^{-1}}$ for the ICON GER 156 m simulations. The other ICON GER simulations have slightly lower RMSE values but are still worse than for the established COSMO model. The lower errors at the location of Cabauw might be explained by the proximity to the model boundary and mostly westerly wind conditions and thus larger impact of the high quality COSMO-DE analyses used as boundary conditions (Fig. 3.9).

The tropospheric profiles from soundings at four different launch times at 06, 12, 18 and 00 UTC (Fig. 3.10) confirm the stated findings of the boundary layer towers and extend them to higher altitudes. Both bias and RMSE are increasing with model integration time (e.g., bias from $-0.1\,\mathrm{m\,s^{-1}}$ at 06 UTC to $0.8\,\mathrm{m\,s^{-1}}$ at 00 UTC, RMSE from $2.4\,\mathrm{m\,s^{-1}}$ to $4.8\,\mathrm{m\,s^{-1}}$ for ICON GER 156 m) for all ICON GER domains influenced by such issues as a wrong development of the pressure field or further issues of the model physics. However, the bias of the COSMO model is almost constant. The largest deviations can be seen at midnight at the 00 UTC sounding with a forecast time of 24 hours (Fig. 3.10d), where the bias is about $1\,\mathrm{m\,s^{-1}}$ for the ICON GER and only $-0.3\,\mathrm{m\,s^{-1}}$ for the COSMO. The RMSE of the ICON GER 156 m simulations is about $5\,\mathrm{m\,s^{-1}}$ above 1.5 km, which is almost a factor of two larger than COSMO.

Remarkable is also the large RMSE of up to $4\,\mathrm{m\,s^{-1}}$ for ICON GER 156 m above 6 km at 12 UTC, which is again twice as large as the RMSE of the COSMO simulations and not identifiable for the 06 and 18 UTC launch times. Only six out of the 14 considered sounding stations launch a radiosonde four times a day at 00, 06, 12, and 18 UTC, whereas the other stations only launch them at 00 and 12 UTC. Nevertheless, the previous stated findings, especially the notable differences found at 12 and 00 UTC, are independent of the selected sounding stations. Similar results are found if only the stations with all four launching times or those stations with only two launching times are selected.

Overall, large biases and standard deviations of the mean sea level pressure (Sect. 3.3), inaccuracies of the simulated pressure field throughout the troposphere, and probably an artificially increased variability by the model might explain the errors found in surface winds, as well as at higher altitudes. Issues of an incorrectly simulated stratification or uncertainties of the simulated turbulence have particularly large influences on the 10 m wind analysis. Most likely, a too-strong coupling of the different model levels during night time explains the substantially overestimated 10 m wind speed by all ICON GER simulations before and after sunset. In opposition, a too-weak coupling during daytime of ICON GER and COSMO might be the reason for the underestimated 10 m wind speed.

Figure 3.10.: Full tropospheric bias (left column) and RMSE (right column) profiles of the wind speed using the high-resolution DWD soundings and corresponding ICON GER and COSMO single-column model output for single launch times ((a,b) - 06 UTC, (c,d) - 12 UTC, (e,f) - 18 UTC, (g,h) - 00 UTC forecast 24 hr). Vertically averaged bias and RMSE for each dataset are provided with the labels. The ICON GER and COSMO model profiles are averaged 30 minutes before and after launch time of DWD sounding in terms of a fuzzy verification. At 06 UTC and 18 UTC only six of the 14 stations provide sounding data.
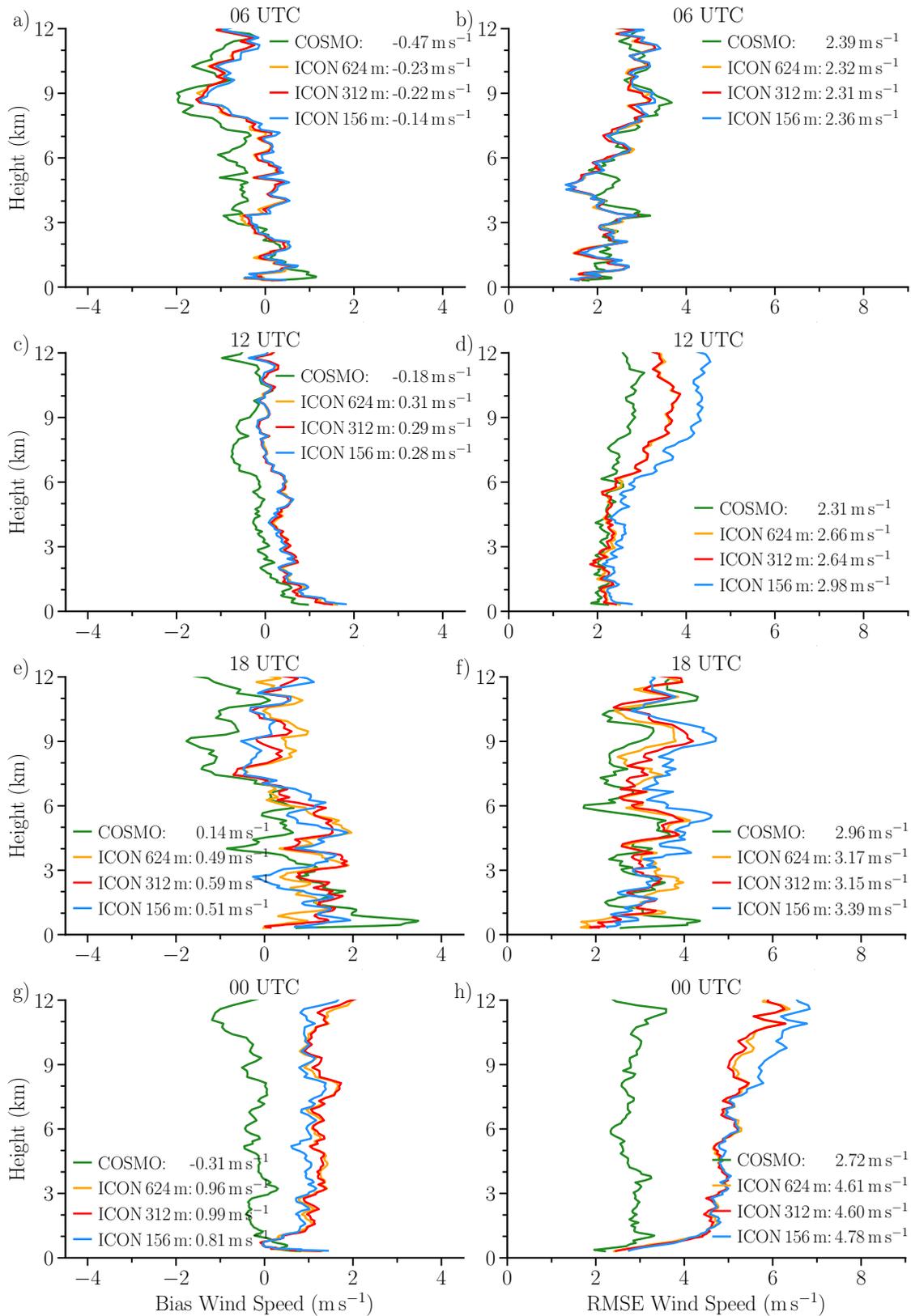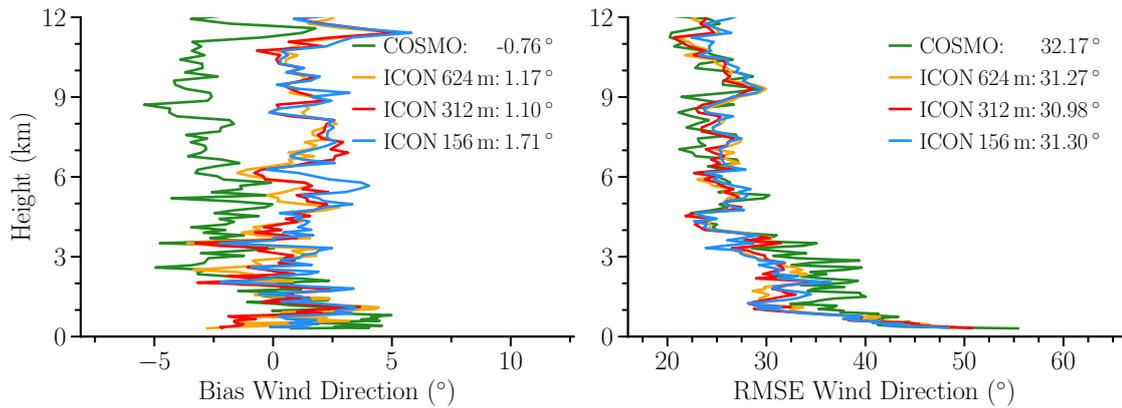
Figure 3.11.: Averaged full tropospheric bias (left) and RMSE (right) profiles of the wind direction using the high resolution DWD soundings and corresponding ICON GER and COSMO single-column model output using all four launch times of 06, 12, 18 and 00 UTC (with 24 hour forecast for the models). Vertically averaged bias and RMSE for each dataset are provided with the labels. The ICON GER and COSMO model profiles are averaged 30 minutes before and after the launch time of the DWD sounding in terms of a fuzzy verification. At 06 UTC and 18 UTC only six of the 14 stations provide sounding data.

Nevertheless, further in-depth evaluation of, for example, the pressure at high altitudes like at the middle of the troposphere, diurnal cycle of the mean sea level pressure, and other analyses will be valuable to explore the underlying issues resulting in these above mentioned deviations of the wind speed profiles. However, this is beyond the scope of the baseline evaluation of this study focusing on the general evaluation of the new ICON GER simulations to provide an overview on the model's quality.

In addition to the wind speed, the wind direction exhibits information about the right positioning of pressure systems by the model and is another essential parameter of the baseline evaluation. The wind direction of the ICON GER and COSMO simulations is only evaluated in terms of the full troposphere DWD soundings. The near-ground 10 m wind direction observations of the DWD weather stations are not considered for the evaluation because they are strongly influenced by local characteristics like obstacles, trees and fine-scale orographic features. Additionally, the ICON LES does not resolve obstacles like buildings or trees explicitly and consequently differences would be expected. All four launch times of 06, 12, 18, and 00 UTC of the DWD soundings are included for the wind direction analysis. The bias and RMSE are almost constant in time for all models and accordingly only the time averaged profiles are presented (Fig. 3.11).

The large-scale wind direction is mainly prescribed by the location of low- and high-pressure systems provided by the initialisation and boundary data, recognisable by similar errors of the COSMO and all ICON GER simulations. The errors are almost constant regarding height (Fig. 3.11). The synoptic situation and positions of the pressure systems

are well represented by all simulations, seeable by an average bias of less than 2°. As with the wind speed, the random error is dominating with a RMSE of about 31° for ICON GER and 32° for COSMO simulations, indicating an overestimated wind direction variability. The impact of the near-ground properties is visible by the larger RMSE of up to 50° within the first roughly 500 m above ground in all simulations.

Most probably, uncertainties of the pressure field affecting the wind field, lead to the stated errors in wind direction. Additionally, inaccuracies of the GPS-based wind direction measurement of the soundings induce a small uncertainty of about two degrees (Vaisala, 2013). Contrary to the wind speed evaluation, the representation of the wind direction by the ICON GER simulations is of a similar quality as of the well-established COSMO model albeit no added value is detected (Fig. 3.11).

Overall, the finer horizontal and vertical resolution results in larger errors in wind speed and similar differences in wind direction, respectively, and do not show the expected added value for the flow conditions due to the higher resolution. The near-ground wind speeds are slightly worse compared to the well-established COSMO model (RMSE of $1.5\,\mathrm{m\,s^{-1}}$ for ICON GER compared to $1.4\,\mathrm{m\,s^{-1}}$ for COSMO). The largest problems are seen at the diurnal cycle with a substantially overestimated wind speed during the night by about $0.5\,\mathrm{m\,s^{-1}}$ and underestimated wind speed during the day by also about $0.5\,\mathrm{m\,s^{-1}}$, most probably caused by a too-strong coupling at night and too-weak coupling during daytime by the ICON GER. Additionally, large errors of an RMSE of more than $5\,\mathrm{m\,s^{-1}}$ are identified above 1 km for the full troposphere after 24 hours of model integration by the ICON GER, which might be the result of a wrong propagation of pressure systems. The substantially increased errors of the ICON GER simulations after 24 hours of model integration are not visible for the COSMO simulations.

3.5. Stratification

The properties of the boundary layer and many exchange processes within the atmosphere are influenced by the stratification, which itself depends on the flow conditions and turbulence. The stratification is examined by the 2 m temperature, boundary layer, and full tropospheric temperature profiles, followed by an evaluation of the potential temperature within the boundary layer and the boundary layer height.

The height of the boundary layer can be determined by various methods and instruments like a lidar or radiosoundings. An overview about the different possibilities with their advantages and disadvantages is given by Hennemuth and Lammert (2006). The column

Table 3.3.: Average standard deviation (STD) of 2 m temperature across all locations of DWD weather stations for the observations, all ICON GER and COSMO simulations using all twelve days. Averaged bias and RMSE errors are calculated between the model output and the observations.

Temperature 2 m (K)	Obser- vations	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
STD	6.81	6.43	6.44	6.45	6.34
Bias		0.06	0.14	0.21	0.31
RMSE		1.91	1.90	1.93	1.91

output of the ICON GER simulations, available at the sounding stations, offers the possibility to evaluate the boundary layer height at the radiosounding locations. Different definitions of the boundary layer height using radiosoundings are described by Seibert et al. (2000). Within this study, the boundary layer height is determined by the first level above surface, where the bulk Richardson number (Richardson et al., 2013) is larger than the critical Richardson number, which is set to 0.28 as used within the ICON code and by the study of Heinze et al. (2017a). The bulk Richardson number provides information on the ratio between the turbulence and the shear production for a thin layer. Consequently, a small bulk Richardson number depicts the top of the boundary layer, where the turbulence is decreasing and the shear production increasing. The determination of the boundary layer height by the critical bulk Richardson number is applied because this method allows for a physical consistent and fair comparison between the DWD soundings and the ICON GER, as well as COSMO simulations with all data available.

The local near-ground conditions are examined by the 2 m temperature at the locations of the DWD weather stations. All simulations are on average slightly too warm with their positive bias values of 0.06 K to 0.31 K (Tab. 3.3). A better representation of small-scale characteristics and related processes, such as land-sea effects and other surface characteristics of the higher resolved ICON GER domains, compared to the COSMO model, might explain the lower bias values of 0.06-0.21 K of the ICON GER simulations in comparison to the 0.31 K bias of COSMO (Tab. 3.3). Nevertheless, the bias is almost linearly increasing with the ICON GER resolution from 0.06 K for the ICON GER 624 m up to 0.21 K for the ICON GER 156 m, which is still more than 30% lower than that of the COSMO simulations. A similar behaviour was seen for the 10 m wind speed, where also the ICON GER 156 m showed the largest bias compared to the other ICON GER resolutions and COSMO simulations.

The mean RMSE of the different resolved ICON GER simulations is almost identical to the COSMO simulations. An overestimated small-scale variability by the high-resolution ICON GER simulations could be responsible for the similar RMSE values of about 1.9 K
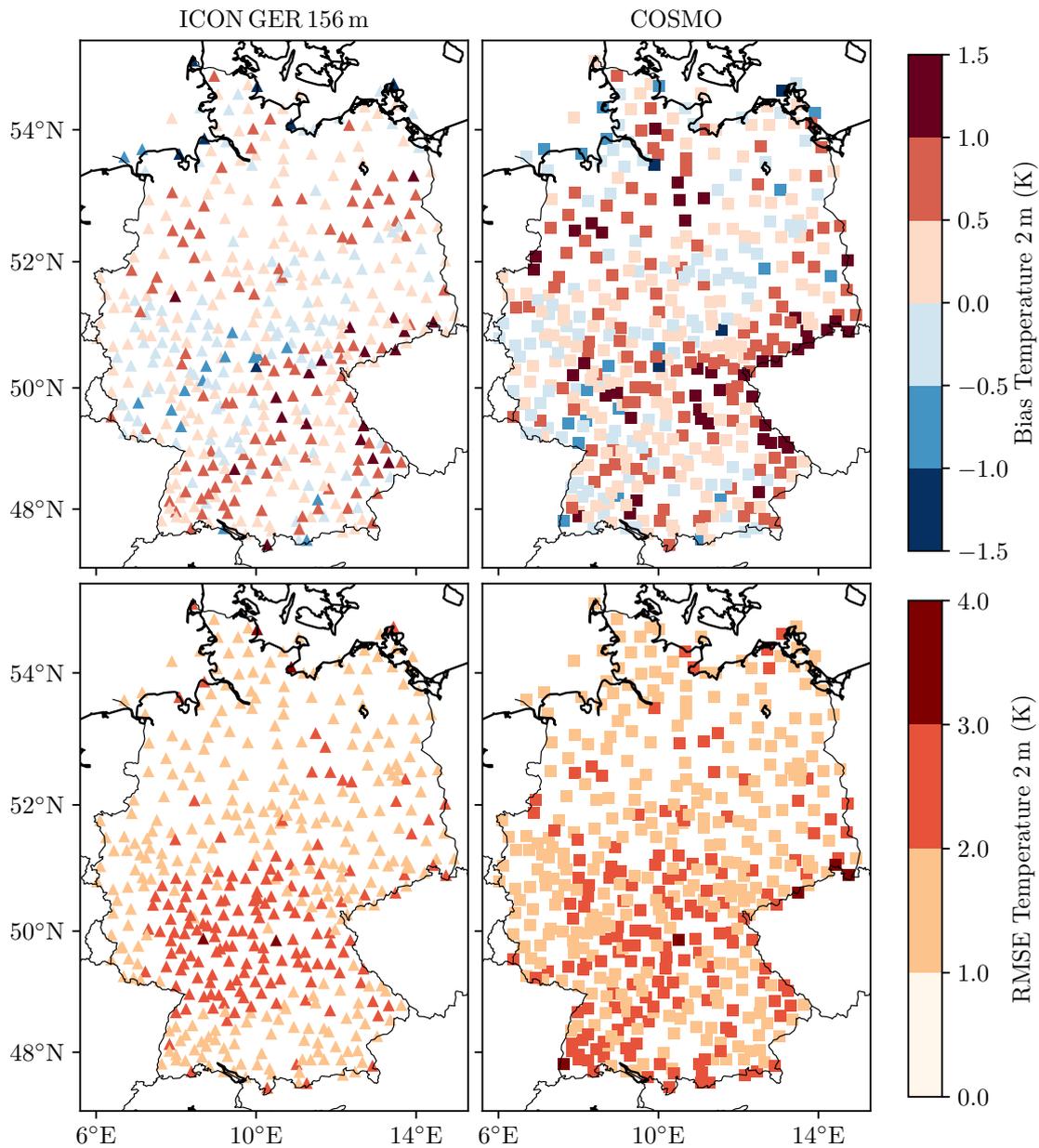
Figure 3.12.: Bias (upper row) and RMSE (bottom row) of 10-minute average 2 m temperature of all available ground based DWD stations using all twelve days for the ICON GER 156 m (left column) and COSMO simulation output (right column).

as of the COSMO simulations despite the lower bias. Therefore, the similar RMSE depicts the dominant random error for the 2 m temperature of the ICON GER simulations, which accounts for more than 99% of the RMSE. The 2 m temperature is overall well represented by all ICON GER simulations with very similar RMSE values as of the COSMO simulations. Furthermore, the RMSE of the ICON GER of roughly 1.9 K is much lower

than of the natural variability, illustrated by the standard deviation of the observations of 6.8 K (Tab. 3.3). Therefore, the RMSE is about 72% lower than the observed standard deviation exhibiting a good simulation quality.

The accuracy of the spatial representation of the 2 m temperature by the ICON GER and COSMO simulations is assessed in terms of the measurements of the Germany-wide ground-based weather stations. Local effects at stations with complex terrain, as well as the influence of the sea on near-shore stations, might be better resolved by the ICON GER simulations than by the COSMO simulations, as seen by a reduced bias in southern Germany and near coastlines (Fig. 3.12) compared to the COSMO output. On average, the ICON GER simulations are slightly too warm by 0.5–1.0 K at Northern and Southern Germany and slightly too cold at Central Germany with a bias of about -0.5 K. There is no clear spatial pattern visible for the bias of the COSMO simulations at the locations of the weather stations, which ranges from a slight underestimation of about -0.5 K to an overestimation of up to 1.5 K. The four erroneous ICON GER prototype simulations, assessed by Heinze et al. (2017a), exhibit a clear north-south gradient with a strong positive bias of up to 5 K for locations in North Germany and by up to -4 K for locations in South Germany. Consequently, the new twelve simulated ICON GER days show a substantially lower bias and improved results. The small underestimation by about -0.5 K for Central Germany, stated by Heinze et al. (2017a), is similar to the abovementioned results.

However, large RMSE values of up to 2–3 K remain in all simulations over complex terrain, possibly due to not well-represented processes of local heating effects within valleys or increased wind speeds at hills or other such factors. An increased RMSE at regions with complex terrain was also found for the near-ground wind speed. Uncertainties at the representation of the correct flow conditions and other such factors might change the ventilation within valleys and the mixing with the layers above or the surface fluxes and consequently influence also the accuracy of the simulated 2 m temperature. Apart from the larger RMSE values for locations at complex terrain, the 2 m temperature agrees well with the observations, which is depicted by a low RMSE of about 1–2 K for most locations. In Heinze et al. (2017a), larger RMSEs of about 2–3 K are found for almost all considered weather stations in Germany based on the four erroneous ICON GER prototype simulations. Accordingly, also an improvement at the representation of the 2 m temperature is identified for the twelve new ICON GER simulations. However, uncertainties at the simulation of the surface fluxes of the sensible and latent heat fluxes in general could be one possible reason for the found errors of the 2 m temperature by the ICON GER and COSMO simulations.
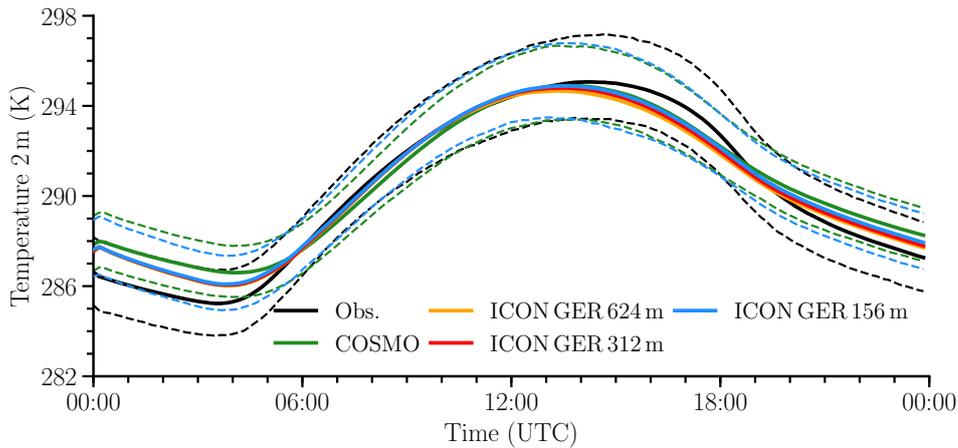
Figure 3.13.: Mean diurnal cycle of the 2 m temperature averaged over 10-minute intervals for the observations, all ICON GER and COSMO simulations using all locations of the DWD ground-based weather stations. The upper dashed lines shows the 95th percentile and the lower dashed lines the 5th percentile of all data, shown for the observations (black), COSMO (green), and ICON GER 156 m (blue).

The mean diurnal cycle of the 2 m temperature is investigated by the average over all ground-based weather station locations of the DWD observations and the ICON GER and COSMO simulation outputs. In accordance with the presumed too-strong coupling of the model levels during the night by the ICON GER simulations at the wind speed analysis, most likely the too-strong coupling of the levels also causes a too-strong downward mixing of warmer air from higher altitudes to the lower layers. Additionally, the too-strong downward mixing during the night counteracts the cooling from the ground. As a consequence, the 2 m temperature is substantially overestimated by up to 1.5 K during the night by all ICON GER simulations (Fig. 3.13).

Possibly, an underestimation of the radiation cooling rate during the night or uncertainties of the sensible heat fluxes by the COSMO simulations might cause the substantially overestimated 2 m temperature by up to 1.5 K during the night. In contrast to the ICON GER simulations, the coupling of the model levels should not be the reason for the temperature overestimation by the COSMO simulations, because the 10 m wind speed is well represented during the night. The overestimation of the 2 m temperature by the COSMO simulations is of a similar order as the ICON GER simulations.

Most probably, a too-weak coupling of the model levels by the ICON GER and COSMO models during daytime causes an underestimated mixing with the layers above and a higher influence of the ground on the 2 m temperature (Fig. 3.13). The larger impact of the ground might explain the faster response to the decreasing incoming solar radiation after noon and cooling by the ground, which causes a too-early decline of the 2 m temperature by about one hour for all ICON GER and COSMO simulations. The supposed too-weak
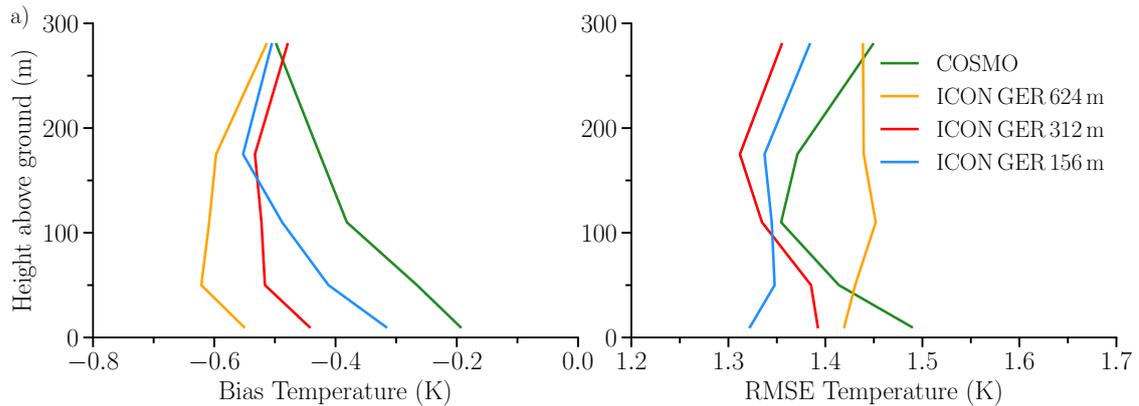
Figure 3.14.: Bias and RMSE of the temperature profiles at the boundary layer tower of the weather mast Hamburg (a) using 10-minute average observed data and corresponding ICON GER and COSMO single-column output. Cabauw and Lindenberg results are shown in Fig. Appendix B.1.

coupling is in agreement with the results of the diurnal cycle of the 10 m wind speed. Despite the too-early decrease, the overall daily maximum temperature is well captured by the ICON GER and COSMO output.

Issues of the COSMO-DE initialisation data used for the ICON GER are probably responsible for the positive 2 m temperature offset of about 1.5 K at the initialisation of the ICON GER simulations. The COSMO-EU analyses used as initialisation for the COSMO simulations point out similar inaccuracies causing an overestimated 2 m temperature of 1.5 K at the initialisation of the COSMO. The variability of the 2 m temperature of the ICON GER and COSMO simulations matches the observations well, except for the temperature offsets during the day and night. The good agreement of the variability between the simulations and the measurements indicate a good representation of the characteristics at the considered locations of the weather stations (Fig. 3.13).

An accurate simulation of the temperature profile of the boundary layer is among others important for the correct representation of the stratification. The temperature profile of the ICON GER and COSMO simulations is evaluated using the boundary layer tower observations of the weather mast Hamburg, Cabauw, and Lindenberg. Only the results of the weather mast Hamburg are presented, because the errors of Cabauw and Lindenberg are very similar (Appendix B.1). The bias and RMSE are almost constant with height within the boundary layer for all ICON GER and COSMO simulations, which can be seen at the weather mast Hamburg (Fig. 3.14). Both models are on average too cold by about -0.4 K to -0.6 K for ICON GER and -0.3 K for COSMO within the first 280 m above surface. This is in agreement with the findings of Heinze et al. (2017a), where the authors stated a too-cold boundary layer for the erroneous ICON GER prototype simulations.

A better representation of small-scale boundary layer processes, even though they are not seen for the wind speed, might lead to the lower RMSEs found for the ICON GER 312 m and 156 m with about 1.3 K compared to the COSMO model with an RMSE of roughly 1.4-1.5 K. Similar results, except for an almost zero bias at Cabauw, are found for the other two boundary layer towers of Lindenberg and Cabauw (Appendix Fig. B.1).

The previous evaluation of the temperature profile within the boundary layer is extended to the full troposphere by the DWD soundings and the corresponding output of the ICON GER and COSMO simulations. On average, the full tropospheric temperature profiles of the ICON GER and COSMO simulations are in very good agreement with the observations, as seen by a vertically averaged bias of less than 0.20 K and an RMSE of about 1 K above the boundary layer (Fig. 3.15). Both, the bias and RMSE are almost constant with height. The overestimated wind speed by almost $1\,\mathrm{m\,s^{-1}}$ (Fig. 3.10) within the full troposphere or an overestimated variability of for example clouds and further model uncertainties might cause the increasing RMSE by about 75% between 06 UTC and the 24 hour forecast time step of the ICON GER simulations. The large increase in errors during the model integration of up to 75% is especially seen for the ICON GER 156 m. The vertically averaged RMSE of the temperature of the COSMO simulations increase only by about 30% between 06 UTC and the 24 hour forecast time step. Additionally, at most heights of the 00 UTC sounding with 24 hour forecast time, the RMSE of the ICON GER 156 m simulation is about 0.5 K larger than that of the COSMO simulations.

The temperature stratification has a substantial impact on the boundary layer height. Therefore, the diurnal development of the mean profiles of the potential temperature at the three boundary layer towers is investigated in the following (Fig. 3.16). The potential temperature is the temperature of an unsaturated air parcel of dry air at a certain height adiabatically brought to a standard pressure level as for instance 1000 hPa and is a useful measure to investigate atmospheric stability (Holton, 2004).

All ICON GER simulations and the COSMO output are similar to the tower measurements. The stability is, as expected, near neutral to unstable during the daytime, as visible by a negative slope, and slightly stable for the 00 UTC profile with 24-hour forecast, as seen by a positive slope (Fig. 3.16). The temperature inversion at roughly 180 m for the weather mast Hamburg and at roughly 150 m for Cabauw at 06 UTC is well captured by all models, even though there is no benefit of the higher resolution of ICON GER visible. The convective boundary layer with unstable profiles during the daytime is well represented.

The stable conditions at 00 UTC are especially overestimated by the COSMO model at Cabauw with a potential temperature difference of 1.5 K between 10 m and 200 m,
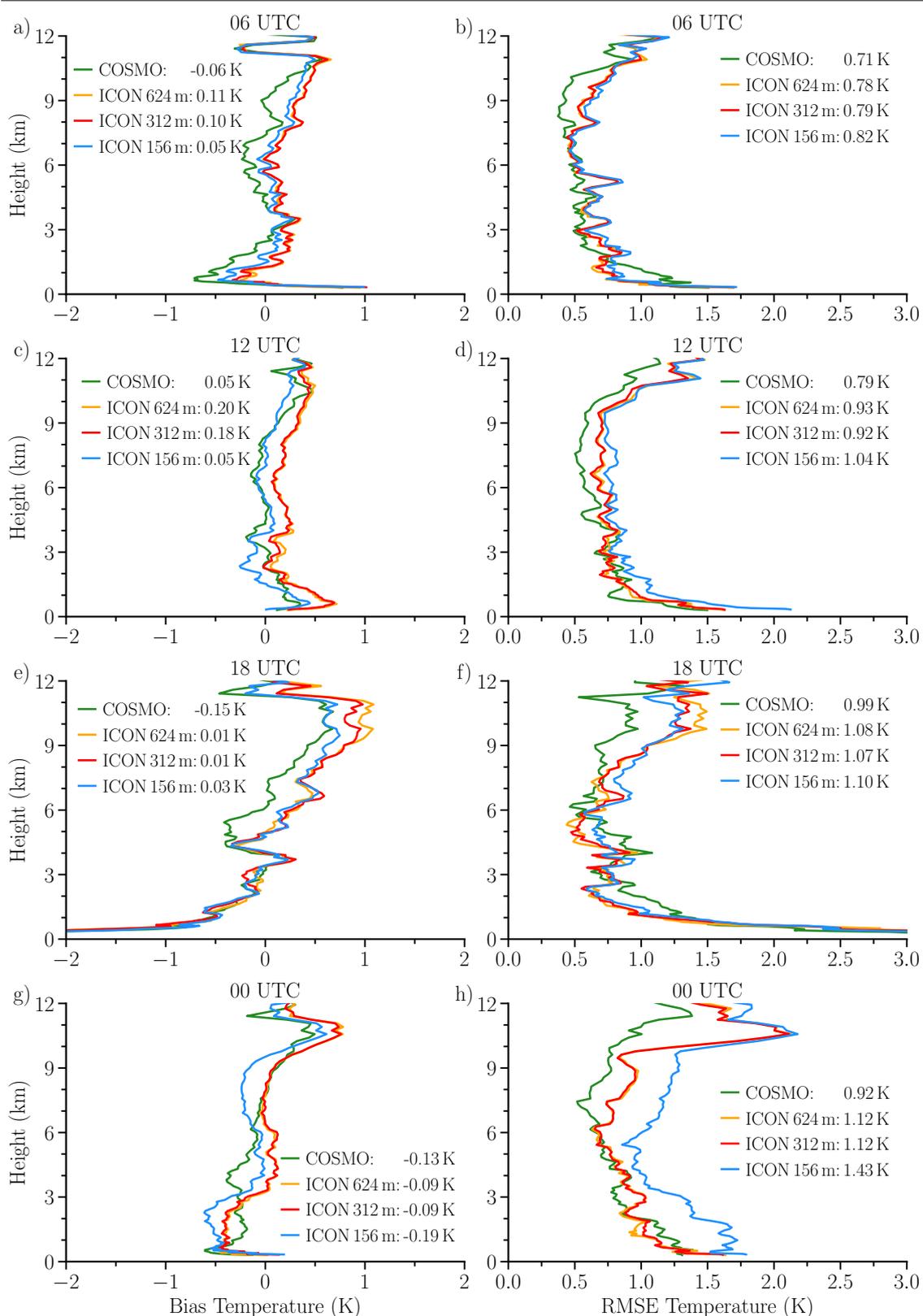
Figure 3.15.: Full tropospheric bias (left column) and RMSE (right column) profiles of the temperature using the high-resolution DWD soundings and corresponding ICON GER and COSMO single-column model output for single launch times ((a,b) - 06 UTC, (c,d) - 12 UTC, (e,f) - 18 UTC, (g,h) - 00 UTC forecast 24 hr). Vertically averaged bias and RMSE for each dataset are provided with the labels. The ICON GER and COSMO model profiles are averaged 30 minutes before and after launch time of DWD sounding in terms of a fuzzy verification. At 06 UTC and 18 UTC only six of the 14 stations provide sounding data.
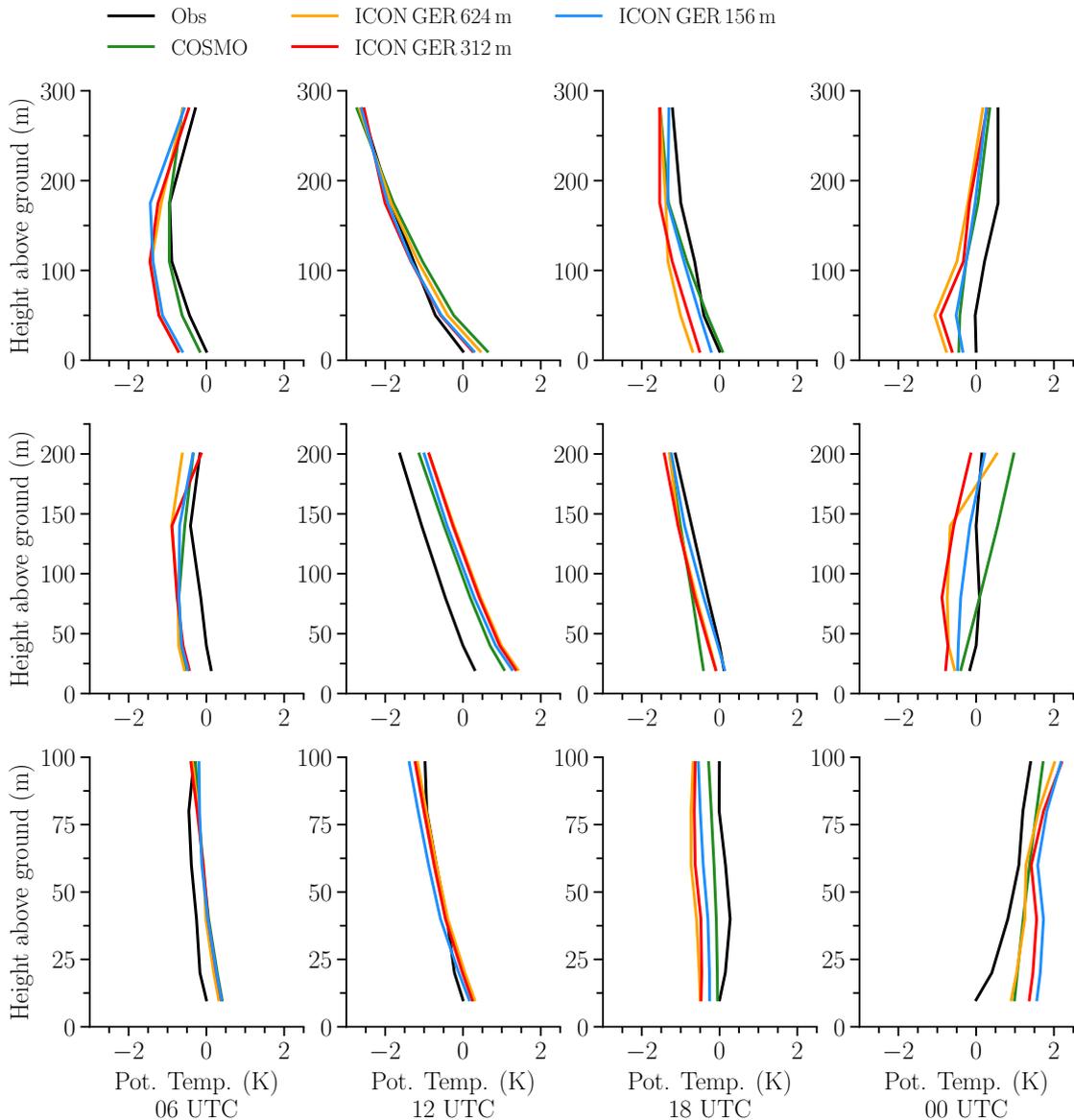
Figure 3.16.: Average diurnal cycle of vertical potential temperature within the boundary layer of the observations, all ICON GER simulations and COSMO model output for the weather mast Hamburg (top row), Cabauw (middle row), and Lindenberg (bottom row). All twelve days are included with 10-minute average values.

compared to the observed near-neutral conditions with an almost nonexsting potential temperature difference. The ICON GER profiles of the different resolutions fit well to the measurements, and only the ICON GER 624 m is slightly too stable with a potential temperature gradient of less than 0.5 K for Cabauw at 00 UTC.

Remarkable is the good agreement in atmospheric stability after sunset at 18 UTC and 00 UTC, when large errors of the near-ground wind speed are present due to a missing

decrease of the wind speed, as discussed in Section 3.4. Altogether, the potential temperature within the boundary layer is in a good agreement with the observations, and the diurnal cycle is well represented, except for small offsets at single locations and time steps, such as for Cabauw at 12 UTC. The ICON GER simulations do not show any added value but are of a similar quality as the COSMO simulations. The visible deviations might arise from such issues as those noted in previous sections of the wind speed and temperature profile.

The boundary layer heights are derived consistently by the bulk Richardson number (Richardson et al., 2013), which is defined according to equation 3.7:

$$R_{\mathrm{B}} = \frac{(g/T_{\mathrm{v}})\Delta\theta_{\mathrm{v}}\Delta z}{(\Delta U)^2 + (\Delta V)^2} \tag{3.7}$$

with:

- R_{B} : Bulk Richardson number (-)
- g : Gravitational acceleration $= 9.80665\,\mathrm{m\,s}^{-2}$
- T_{v} : Absolute virtual temperature (K)
- $\Delta\theta_{\mathrm{v}}$: Difference of virtual potential temperature across two considered layers (K)
- Δz : Height difference between two layers (m)
- ΔU^2 : Difference of horizontal wind speed of east-west component across two considered layers $(\mathrm{m\,s}^{-1})$
- ΔV^2 : Difference of horizontal wind speed of north-south component across two considered layers $(\mathrm{m\,s}^{-1})$.

The determination of the planetary boundary layer (PBL) height by the bulk Richardson number is applied to all sounding stations across Germany using the DWD soundings and the corresponding single column output of the ICON GER and COSMO simulations. The limited vertical resolution of the commonly used ICON GER grid for all datasets with 20 m resolution near surface and up to 80 m at about 1500 m above ground influences the overall accuracy at the PBL-height calculation. The DWD soundings have a vertical resolution of about 5 m and are averaged to the ICON GER grid. The even coarser COSMO grid with a vertical resolution of roughly 20 m near the surface and increasing to about 200 m at 1500 m height is interpolated to the ICON GER grid to treat all datasets consistently and not reduce the resolution of the ICON GER simulations.

The simulated boundary layer heights of the ICON GER and COSMO model match the observations well during the night, whereas larger deviations are clearly visible during the
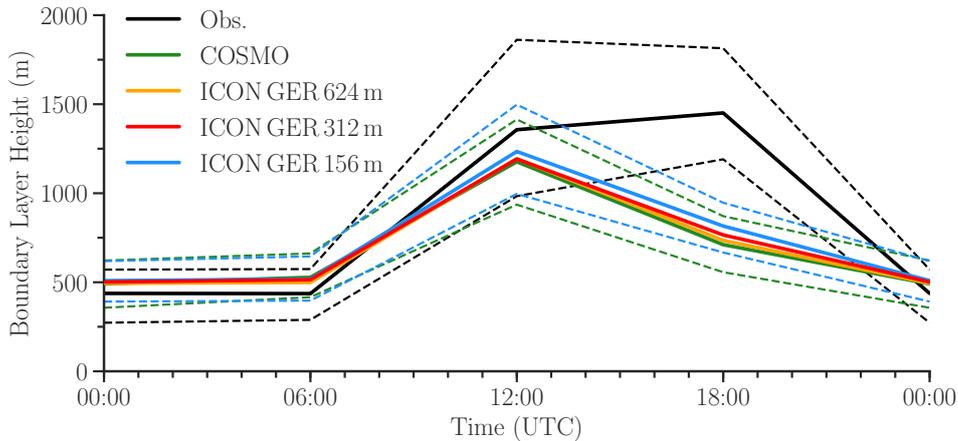
Figure 3.17.: Mean diurnal cycle of the planetary boundary layer (PBL) height derived using the critical Richardson bulk number for all ICON GER, COSMO simulations and the high-resolution DWD soundings of 06, 12, 18 and 00 UTC (with 24 hour forecast time) at all DWD sounding station locations. The upper dashed lines show the 95th percentile and the lower dashed lines the 5th percentile of all data, shown for the observations, COSMO and ICON GER 156 m. At 06 UTC and 18 UTC, only six of the 14 stations provide sounding data.

day and especially before sunset at 18 UTC (Fig. 3.17). A similar behaviour is seen for the variability of the boundary layer heights between the simulations and the measurements.

As mentioned in previous sections, the distinct overestimated 10 m wind speed (Fig. 3.8) due to the supposed too-strong coupling of the model levels during night would increase the downmixing of momentum. Consequently, a too-deep boundary layer would be expected. In contrast to this, the simulated boundary layer height of the ICON GER agrees well with the observations during night at 00 UTC and 06 UTC, with only a slight overestimation of a few meters (Fig. 3.17). Additionally, the boundary layer height is remarkably underestimated by about a factor of two between the observed 1400 m and the simulated 700 m at 18 UTC of the ICON GER and COSMO simulations. This underestimation is also in opposite to the supposed too-high boundary layer height. Furthermore, the PBL-height is similarly underestimated at 18 UTC by the COSMO simulations, even though the wind speed decreases correctly for the COSMO simulations. The large deviations of the boundary layer heights are also in opposition to the before-mentioned overall good representation of the potential temperature within the boundary layer by the ICON GER and COSMO simulations (Fig. 3.16). At 00 UTC and 06 UTC, the good representation of the 10 m wind speed by the COSMO simulations during the night leads to a good match of the simulated boundary layer heights with the observations.

Only six out of the 14 DWD sounding stations launch their soundings also at 06 UTC and 18 UTC, which has to be considered regarding the large deviation found at 18 UTC. Nevertheless, the six sounding stations are well distributed across Germany and provide

Table 3.4.: Average standard deviation (STD) of the planetary boundary layer (PBL) height derived by the critical bulk Richardson number at the locations of DWD soundings for the observations, all ICON GER and the COSMO simulations using all available soundings of 06,12,18 and 00 UTC (with 24 hour forecast) for all twelve days. Averaged bias and RMSE errors are calculated between the model output and the observations.

Boundary Layer Height (m)	Obser-vations	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
STD	656	481	479	487	450
Bias		-68	-60	-40	-80
RMSE		426	422	438	407

a good representation of the present atmospheric conditions. Furthermore, the results at 00 and 12 UTC are independent of selecting stations with four or only two launch times.

The suspected too-weak coupling of the model levels in the daytime by the ICON GER and COSMO simulations of previous sections lead most likely to a too-weak mixing and the observed underestimated 10 m wind speed (Fig. 3.8). The too-weak mixing might explain the underestimated boundary layer height at noon by up to 250 m by the ICON GER and COSMO simulations compared to the observations (Fig. 3.17). The reduced mixing could also be the reason for the missing simulated higher boundary layer heights of more than about 1,500 m and underestimated variability by the ICON GER and COSMO simulations at noon. The variability of the boundary layer heights across the various sounding stations is underestimated during the daytime (Fig. 3.17). High boundary layers measured at some stations are considerably underestimated by up to 500 m at 12 UTC and by up to 1,000 m at 18 UTC, which is seen by a substantially smaller spread of the mean diurnal cycle.

The large impact of the good agreement of the simulated and observed boundary layer heights at 00 UTC and 06 UTC leads to a small mean underestimation of about -40 m to -68 m for the ICON GER simulations and of -80 m for the COSMO simulations (Tab. 3.4). Nevertheless, the above stated large deviations of the simulated boundary layer heights during the daytime compared to the measurements result in a large RMSE of 426–438 m for the ICON GER and of 407 m for the COSMO simulations. The large RMSEs of the ICON GER and COSMO simulations are already on the order of the natural variability of boundary layer heights, depicted by the observed standard deviation of 656 m (Tab. 3.4). Accordingly, issues of the simulated boundary layer height of the ICON GER and COSMO simulations are observed during the daytime, which should be investigated in more detail.

However, the large errors of the boundary layer height of the ICON GER and COSMO simulations in this study are contrary to the overall good representation of the boundary layer heights stated by Heinze et al. (2017a) for the four analysed days of the ICON GER

prototype simulations, except for the 26 April 2013. The higher vertical and horizontal resolution of the ICON GER model does not show any added value compared to the coarser COSMO simulations regarding the boundary layer heights.

Issues of the right coupling of the model levels or the correct simulation of the turbulent mixing and momentum fluxes within the boundary layer might explain the found deviations of the boundary layer height during the daytime. The errors are also aligned with the too-early decrease of the before-mentioned near-ground temperature, which might influence the boundary layer height.

Altogether, the 2 m temperature and vertical temperature profiles, representing the stratification, are of a similar quality in the different ICON GER simulations compared to the much coarser but well-tuned COSMO simulations. All simulations agree well with the observations. The boundary layer height of the different ICON GER simulations are very similar to those of the COSMO output. The simulated boundary layer heights of the ICON GER and COSMO model agree well with the observations during the night but are substantially underestimated by up to 50% and 750 m, during the daytime, especially before sunset at 18 UTC, indicating substantial model issues. As for the flow conditions and in contrast to the expectations of a higher accuracy due to the higher resolution, there is no added value identifiable for the the high-resolution ICON GER simulations compared to the COSMO simulations regarding the stratification and the boundary layer height. The highest resolution of the ICON GER 156 m shows once more the largest uncertainties.

3.6. Humidity

In line with the concept of the evaluation pyramid (Sect. 2.1), humidity is evaluated next because it mainly depends on the air temperature's prescribing how much water vapour the air can contain. Furthermore, water vapour is the key prerequisite for the development of clouds and precipitation. For that reason, the near-ground, boundary layer, tropospheric humidity, and column integrated full tropospheric water vapour is investigated in detail in this section.

The near-ground humidity is investigated by means of the DWD ground-based weather stations and the corresponding ICON GER and COSMO output. All three resolutions of the ICON GER simulations are on average slightly too dry, as seen by a negative bias of -0.17 $\mathrm{g\,kg^{-1}}$ to -0.29 $\mathrm{g\,kg^{-1}}$ at the specific humidity, whereas the COSMO simulations are on average too wet with a positive bias of +0.31 $\mathrm{g\,kg^{-1}}$ (Tab. 3.5). As for the wind speed and temperature, the RMSE of the 2 m specific humidity is also increasing with resolution from 1.26 $\mathrm{g\,kg^{-1}}$ for ICON GER 624 m to 1.32 $\mathrm{g\,kg^{-1}}$ for ICON GER 156 m, which is in

Table 3.5.: Average standard deviation (STD) of 2 m specific humidity across all locations of DWD weather stations for the observations, all ICON GER and the COSMO simulations using all twelve days. Averaged bias and RMSE errors are calculated between the model output and the observations.

Specific Humidity 2 m (g kg^{-1})	Obser-vations	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
STD	2.79	2.81	2.81	2.82	2.91
Bias		-0.17	-0.25	-0.29	0.31
RMSE		1.26	1.29	1.32	1.14

contrast to the expectations of a higher accuracy due to the higher model resolution. The lowest RMSE is once more found for the well-tuned COSMO model with 1.14 g kg^{-1}, which is about 16% lower than the RMSE of the ICON GER 156 m.

However, the near-ground humidity is generally well represented by all simulations, and the RMSE is substantially lower than the observed natural variability represented by the standard deviation of the observations of 2.79 g kg^{-1}. This is in agreement with the similar low errors found for the 2 m temperature prescribing how much water vapour the air can contain. The variability of the specific humidity is well captured by all ICON GER simulations and the COSMO model with a similar standard deviation of about 2.8-2.9 g kg^{-1} compared to the observations with about 2.8 g kg^{-1} (Tab. 3.5).

A weak north-south gradient of the 2 m specific humidity bias is visible for the COSMO model with an underestimated humidity near the coastlines and an overestimated humidity south of central Germany using the DWD weather stations (Fig. 3.18). A similar structure is seen for the wind speed, where most probably issues at the simulation of the right flow field within complex terrain causes the north-south gradient of the error. The ICON GER simulations are on average too-dry, which is visible by a bias of up to -0.5 g kg^{-1}. Overall, the 2 m specific humidity agrees well with the observations, as illustrated by an RMSE of 1.0–1.5 g kg^{-1} for the ICON GER and COSMO simulations at almost all weather station locations. In contrast to the previously found north-south gradient of the RMSE of the wind speed and temperature of the ICON GER simulations, there is no distinct spatial pattern identifiable for the bias and RMSE of the 2 m specific humidity.

Regarding the erroneous ICON GER prototype simulations, Heinze et al. (2017a) found a distinct north-south gradient for the bias of the 2 m specific humidity of up to 2.0 g kg^{-1} at North Germany and of up to -1.5 g kg^{-1} for stations in South Germany. This is not identifiable for the considered twelve ICON GER simulations in this study. Furthermore, RMSE values between 1.0 g kg^{-1} and 3.0 g kg^{-1} are found by Heinze et al. (2017a) for the considered weather station locations in the ICON GER prototype simulations.
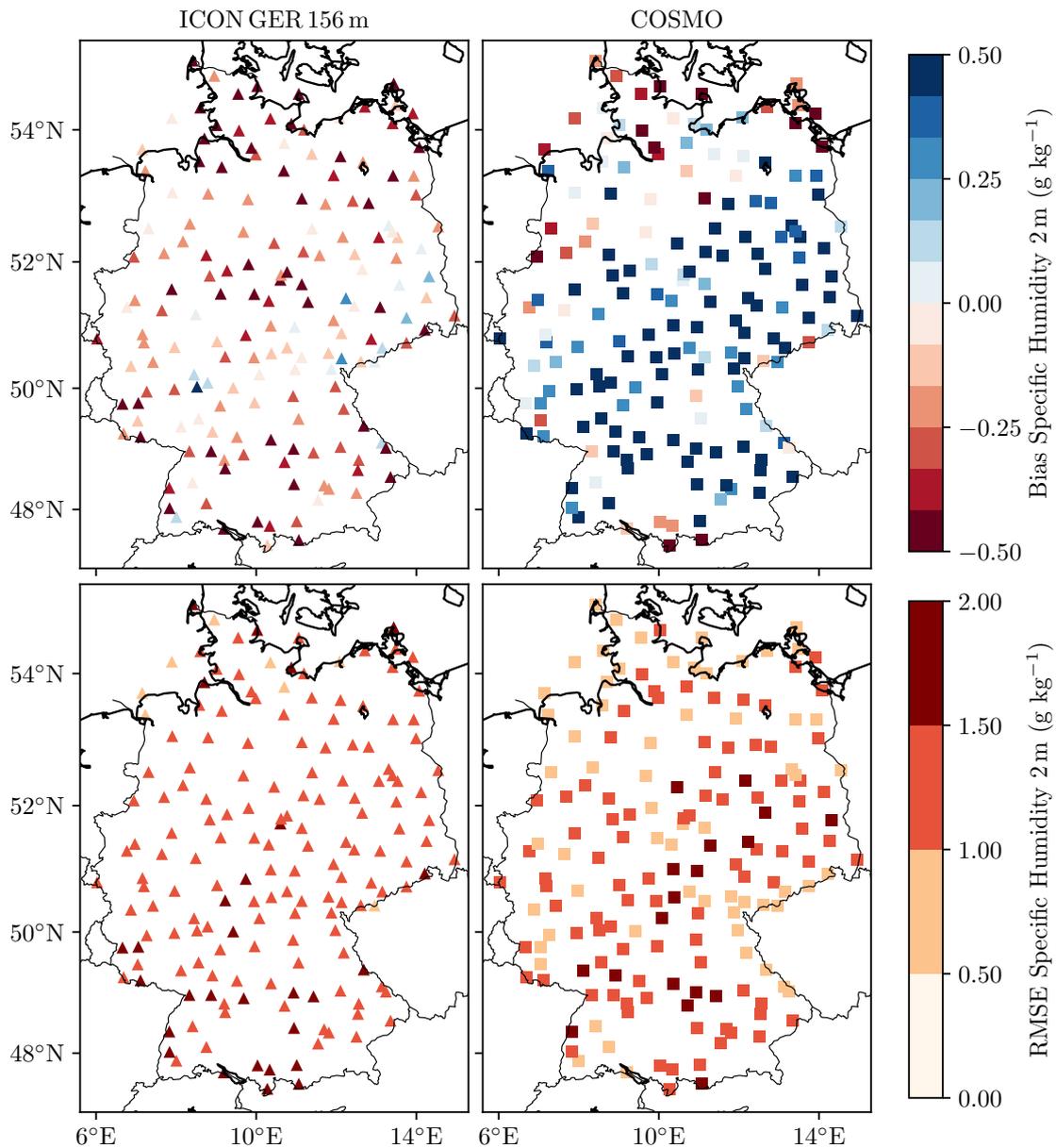
Figure 3.18.: Bias (upper row) and RMSE (bottom row) of 10-minute average 2 m specific humidity of all available ground based DWD stations using all twelve days for the ICON GER 156 m (left column) and COSMO simulation output (right column).

Therefore, a better representation of the near-ground humidity is detectable with an RMSE of about 1.0–1.5 g kg^{-1} at most stations for the twelve new ICON GER simulations using the improved ICON LES model. Among others, issues of the soil moisture model could be responsible for the found deviations between the ICON GER and the measurements.

Figure 3.19.: Mean diurnal cycle of the 2 m specific humidity averaged over 10-minute intervals for the observations, all ICON GER and COSMO simulations using all locations of the DWD ground-based weather stations. The upper dashed lines 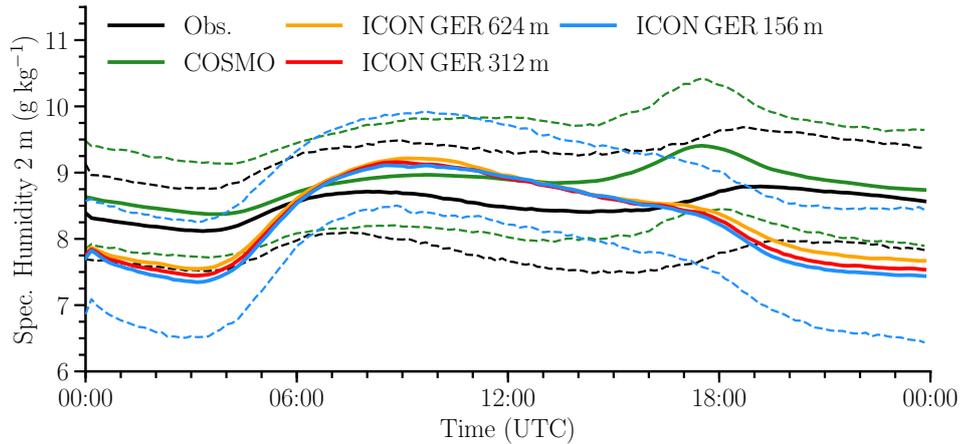shows the 95th percentile and the lower dashed lines the 5th percentile of all data, shown for the observations (black), COSMO (green), and ICON GER 156 m (blue).

The ground-based weather stations are also used to analyse the mean diurnal cycle of the 2 m specific humidity of the ICON GER and COSMO simulations for which the average over all stations is calculated (Fig. 3.19). A too-strong coupling of the lower model levels by the ICON GER model during night could cause a too-strong mixing of drier air from higher layers to the lower layers. The downward mixing of drier air might consequently explain the remarkable underestimation of the 2 m specific humidity by up to $0.75\,\mathrm{g\,kg^{-1}}$ before sunset, and by up to $1\,\mathrm{g\,kg^{-1}}$ and 13% after sunset for the ICON GER simulations (Fig. 3.19). The supposed too strong coupling is in agreement with the findings of the simulated diurnal cycle of the 10 m wind speed and 2 m temperature of the ICON GER. Additionally, the underestimation of the 2 m specific humidity after sunset might be even worse, if the boundary layer would also not be underestimated (Sect. 3.5) and consequently the available humidity distributed across a thicker layer. In contrast, the COSMO simulation matches the observations well during the night with only a small overestimation of less than $0.2\,\mathrm{g\,kg^{-1}}$.

Similar to the overestimated downmixing of dry air during the night by a too-strong coupling of the model levels, a too-weak coupling of the model levels by the ICON GER and COSMO during the day could explain the too-weak mixing of dry air from higher altitudes to lower layers. The too-weak mixing could be the reason for the notable overestimation of the 2 m specific humidity by up to $0.5\,\mathrm{g\,kg^{-1}}$ and about 6%, for the ICON GER and COSMO simulations (Fig. 3.19). The too-weak coupling during daytime would confirm the results of the 10 m wind speed and of the 2 m temperature.

Most likely, the underestimated boundary-layer height around 18 UTC (Fig. 3.17) by the COSMO simulations requires the model to distribute the available humidity within a thinner layer. Consequently, the simulated thinner boundary layer might lead to the substantial overestimation of the 2 m specific humidity by up to $0.75\,\mathrm{g\,kg^{-1}}$ between 16 and 19 UTC, visible for the COSMO simulations (Fig. 3.19). Apart from this overestimation, the COSMO simulations agree well with the observations concerning the diurnal cycle of the 2 m specific humidity with an almost constant small overestimation of about $0.2\,\mathrm{g\,kg^{-1}}$.

The uncertainties of the COSMO-DE analyses, used for the initialisation of the ICON GER simulations, lead to an underestimation of the 2 m specific humidity by about $0.5\,\mathrm{g\,kg^{-1}}$ already at the initialisation of the simulations (Fig. 3.19). Likewise, issues of the COSMO-EU data, used for the COSMO simulations, cause an overestimation of the 2 m specific humidity by about $0.2\,\mathrm{g\,kg^{-1}}$ at the beginning of the simulations. The spread of the specific humidity values of the stations is well captured by the ICON GER and COSMO simulations, except for the stated offset suggesting a good representation of the spatial variability and local conditions. In opposition to the expectations of a higher accuracy due to an increased horizontal resolution of the model, the highest resolved ICON GER 156 m exhibits the largest errors compared to the coarser resolved ICON GER simulations. Therefore, no added value of the higher resolution can be identified considering the 2 m specific humidity.

The specific humidity within the boundary layer is investigated at the three boundary layer towers of Hamburg, Cabauw, and Lindenberg (Fig. 3.20). On average, the specific humidity is well represented within the boundary layer with an overall low bias of $-0.5\,\mathrm{g\,kg^{-1}}$ to $+0.5\,\mathrm{g\,kg^{-1}}$, similar to the low biases found for the temperature profile of the boundary layer (Sect. 3.5). The humidity of the boundary layer seems to be underestimated by the ICON GER and COSMO simulations for locations with a maritime climate and overestimated for more continentally influenced locations. However, the sample size with three stations is very small. Accordingly, the specific humidity is on average underestimated at the maritime-influenced location of Cabauw by about $-0.32\,\mathrm{g\,kg^{-1}}$ for the ICON GER simulations and by about $-0.24\,\mathrm{g\,kg^{-1}}$ for the COSMO simulations. Contrary to this underestimation, the ICON GER and COSMO simulations overestimate the boundary layer humidity at the continental climate location of Lindenberg. At the weather mast Hamburg, which is located between Cabauw and Lindenberg, the vertically averaged bias of the specific humidity is almost zero.

The RMSE of the ICON GER and COSMO simulations is lowest with about $0.7\,\mathrm{g\,kg^{-1}}$ to $0.9\,\mathrm{g\,kg^{-1}}$ for the Cabauw site and largest at the weather mast Hamburg with up to $1.3\,\mathrm{g\,kg^{-1}}$ for the ICON GER and up to $1.1\,\mathrm{g\,kg^{-1}}$ for the COSMO simulations. Notable is the lowest RMSE of about $0.9\,\mathrm{g\,kg^{-1}}$ found for the highest-resolved ICON GER 156 m
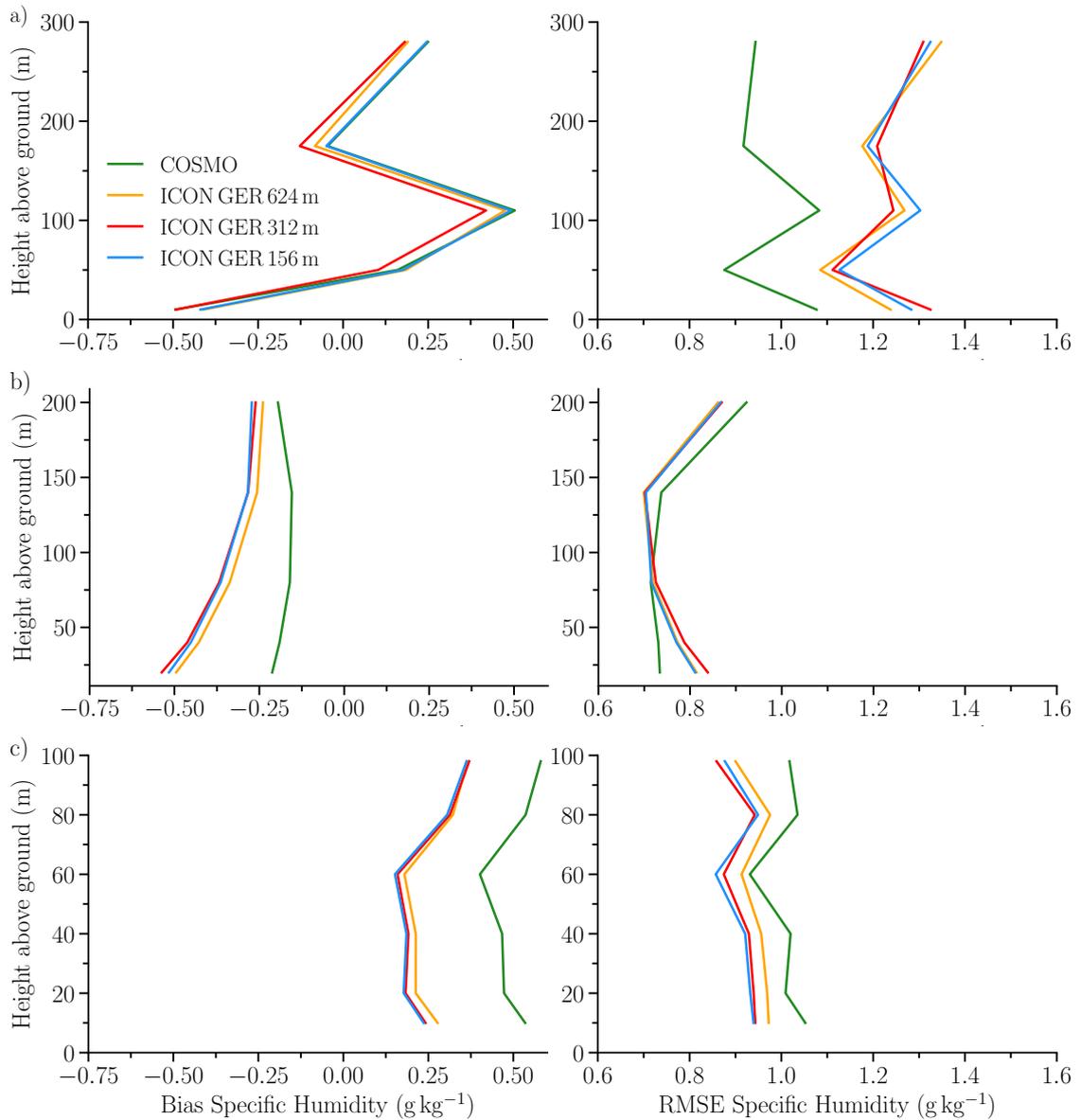
Figure 3.20.: Bias and RMSE of the specific humidity profiles at the boundary layer towers: weather mast Hamburg (a), Cabauw (b), and Lindenberg (c) using 10-minute average observed data and corresponding ICON GER and COSMO single-column output.

at Lindenberg compared to the coarser resolved ICON GER and COSMO simulations. The specific humidity within the boundary layer is in general well represented by all models, visible by an RMSE of less than $1.3\,\mathrm{g\,kg^{-1}}$ for all simulations compared to the observed standard deviation of the $2\,\mathrm{m}$ specific humidity of about $2.8\,\mathrm{g\,kg^{-1}}$, depicting the natural variability (Tab. 3.5). Nevertheless, there is once more no substantial advantage identifiable for the higher resolution of the ICON GER $156\,\mathrm{m}$.

The measurements of the DWD sounding stations provide detailed information about the vertical humidity distribution within the troposphere. Accordingly, the soundings are used to evaluate the full tropospheric profiles of the specific humidity of the ICON GER and COSMO simulations. The full tropospheric profiles of the specific humidity agree well with the measurements, seen by a low bias of -0.4 g kg^{-1} to +0.7 g kg^{-1} and a low RMSE of up to 1.6 g kg^{-1} for all ICON GER simulations and the COSMO model (Fig. 3.21). The overall decreasing available humidity with height in the troposphere leads to a continuously decreasing RMSE with altitude for all ICON GER and COSMO simulations. Additionally, this leads to an almost nonexisting RMSE at about 12 km height. Similarly, the larger amount of humidity and higher variability at lower tropospheric layers explain the larger RMSEs of up to 1.6 g kg^{-1} for the ICON GER and COSMO output at lower altitudes. In particular, the complex interactions of the available humidity with clouds might be responsible for large parts of the RMSE error.

The different ICON GER simulations agree well with the COSMO simulations, except for 00 UTC. After 24 hours of model integration, the ICON GER 156 m simulations exhibit a notably larger RMSE of about 0.2 g kg^{-1} and 35%, between 1 km and 4 km height than the COSMO simulations. Possibly, the in previous sections stated larger RMSEs of the flow conditions (Sect. 3.4) and stratification (Sect. 3.5), influence the found RMSE of the specific humidity at 00 UTC. Apart from the deviations at 00 UTC, the specific humidity profile matches the observations well. There is no substantial impact visible of errors found in the wind speed profiles (Sect. 3.4) on the specific humidity profiles up to 12 km in height.

In addition to the previous analysis of the specific humidity, the column-integrated humidity is evaluated by the IWV, which is a good indicator of the moist troposphere. More than 300 GNSS stations operationally observe the IWV across Germany. The measurement principle is based on the observed time delay of the GPS signal induced by the water vapour. This valuable dataset provides spatial information across Germany with 10-minute time resolution and is used as a reference to investigate the ICON GER and COSMO models.

On average, all models are too dry, which is identifiable by a negative bias of -0.51 kg m^{-2} to -0.76 kg m^{-2}. This is in line with the negative bias of the 2 m specific humidity for the ICON GER simulations mentioned in Table 3.5. Notable is also the on-average underestimated IWV by the COSMO simulations, which is in opposition to the previously found overestimation of the 2 m specific humidity analysis (Tab. 3.5).

An added value of the IWV is expected for the ICON GER simulations due to the higher resolution and explicit simulation of small-scale processes, such as shallow cumulus clouds, compared to the COSMO model. The analysis of the GNSS measurements shows the lowest
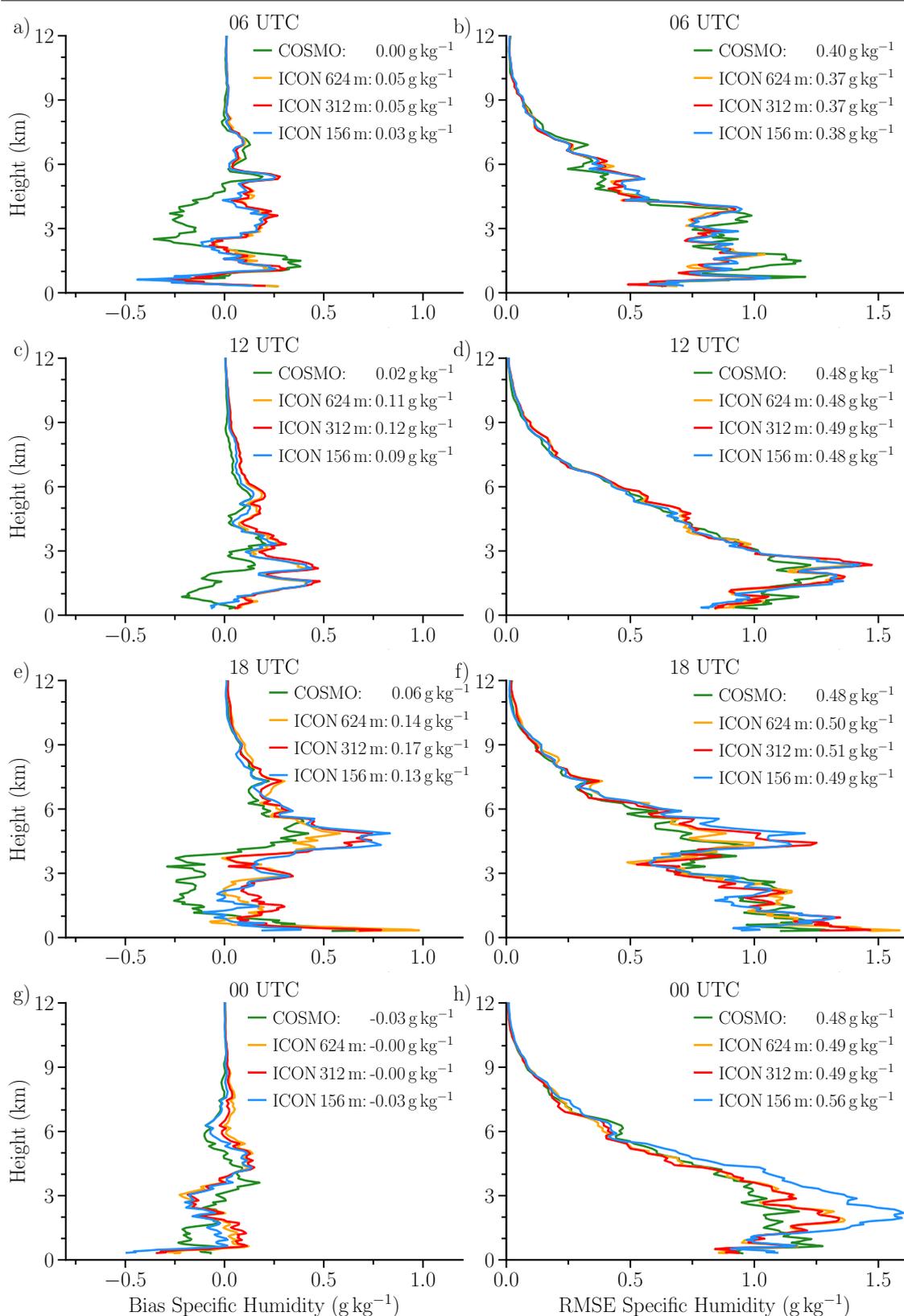
Figure 3.21.: Full tropospheric bias (left column) and RMSE (right column) profiles of the specific humidity using the high-resolution DWD soundings and corresponding ICON GER and COSMO single-column model output for single launching times ((a,b) - 06 UTC, (c,d) - 12 UTC, (e,f) - 18 UTC, (g,h) - 00 UTC forecast 24 hr). Vertically averaged bias and RMSE for each dataset are provided with the labels. The ICON GER and COSMO model profiles are averaged 30 minutes before and after launch time of DWD sounding in terms of a fuzzy verification. At 06 UTC and 18 UTC only six of the 14 stations provide sounding data.

Table 3.6.: Average standard deviation (STD) of column integrated water vapour (IWV) across all locations of GNSS stations for the observations, all ICON GER and the COSMO simulations using all twelve days. Averaged bias and RMSE errors are calculated between the model output and the observations.

IWV (kg m^{-2})	Obser-vations	ICON GER 624 m	ICON GER 312 m	ICON GER 156 m	COSMO
STD	6.34	5.78	5.86	5.76	5.96
Bias		-0.51	-0.76	-0.73	-0.73
RMSE		2.30	2.90	2.98	2.70

bias of -0.51 kg m^{-2} of all simulations for the ICON GER 624 m and a 15% lower RMSE for the ICON GER 624 m compared to the COSMO simulations (Tab. 3.6), which might result from the explicit simulation of small-scale processes. Similar to previous sections, where largest bias and RMSE values of the near-ground wind speed, temperature, and humidity are found for the highest-resolved ICON GER 156 m simulations, the bias and RMSE of the IWV is again increasing with the ICON GER resolution (Tab. 3.6). The increased RMSE is visible by a 30% higher RMSE for the ICON GER 156 m than of the ICON GER 624 m. The larger uncertainties of the higher-resolved ICON GER simulations is again in contrast to the above stated hypothesis of an increased accuracy due to the better representation of small-scale processes. Possibly, an overestimated variability of the humidity and cloud field explains the increasing RMSE for the higher-resolved ICON GER simulations. Except for the ICON GER 624 m, the RMSE values of the other two ICON GER simulations are higher than the RMSE of the COSMO simulations.

The IWV of the different ICON GER and COSMO simulations matches the observations well with an RMSE of 2.30-2.98 kg m^{-2} for the different ICON GER simulations (Tab. 3.6). Accordingly, the RMSE is by a factor of two-to-three smaller than the observed natural variability depicted by an observed standard deviation of 6.34 kg m^{-2} (Tab. 3.6), which is in accordance with the good representation of the full troposphere humidity (Sect. 3.6). At the evaluation of the four ICON GER prototype simulations, Heinze et al. (2017a) state an average RMSE of 4.49 kg m^{-2}. Therefore, a distinct lower RMSE is seen for the twelve considered ICON GER simulations in this study.

The spatial distributions of the IWV bias and RMSE of the ICON GER and COSMO simulations are very similar, even though there is no clear pattern like a north-south gradient identifiable (Fig. 3.22). Also, the magnitudes of the bias and RMSE are similar between the ICON GER and COSMO simulations. The large occurred thunderstorms and deep convection during the simulated summer days (Appendix A) might have an influence on the humidity field. Most likely, the complex interactions and uncertainties of the
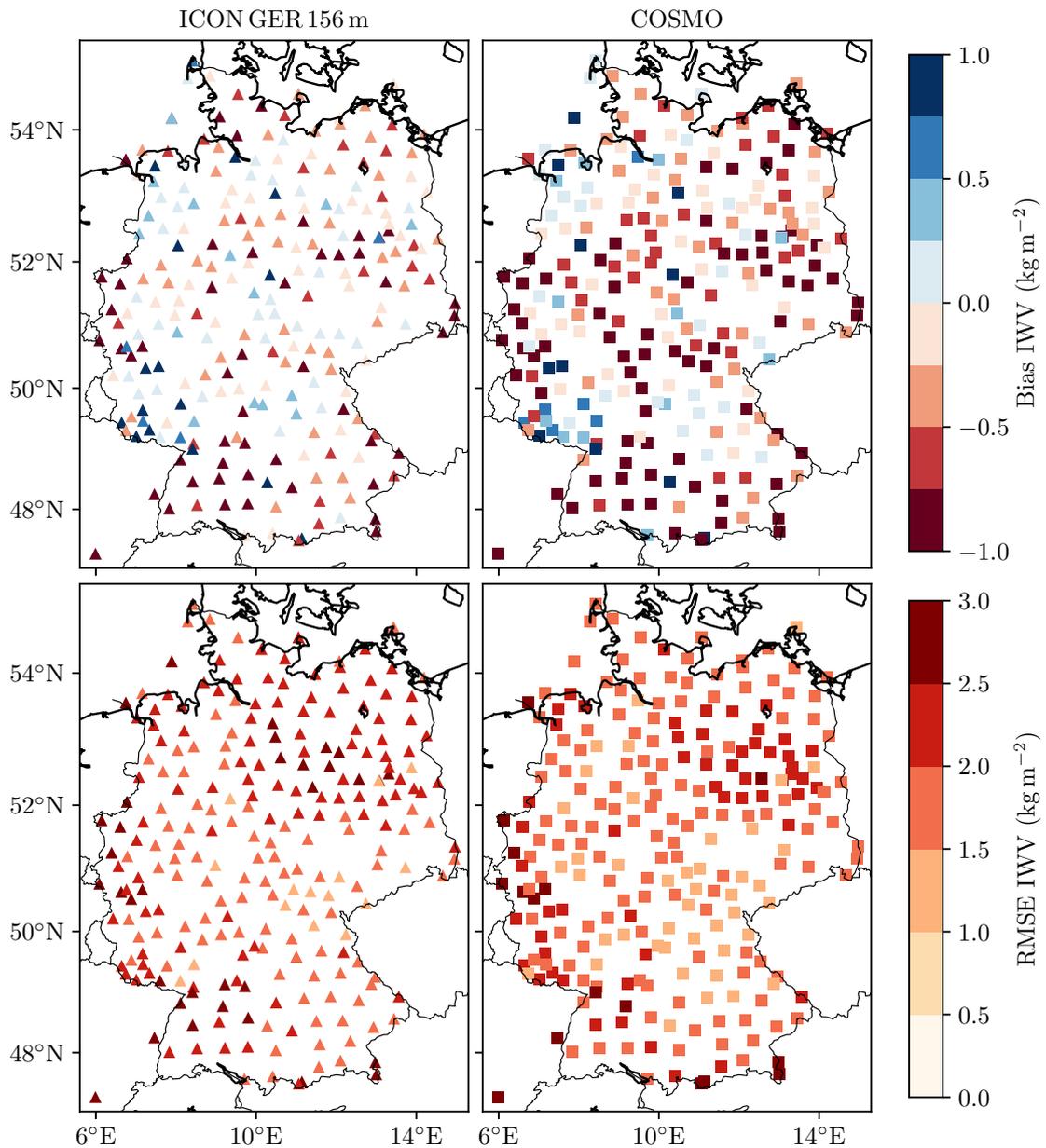
Figure 3.22.: Bias (upper row) and RMSE (bottom row) of the 10-minute average integrated water vapour (IWV) for ICON GER 156 m (left column) and COSMO (right column) using the available GNSS observations of all twelve simulated days.

representation of these thunderstorms might explain the two regions of slightly higher RMSE values around Brandenburg and close to the Black Forest. Apart from these two patterns, most RMSE values are in the range of 1.5–2.5 kg m^{-2} (Fig. 3.22). Presumably, the large variability of the tropospheric humidity and the complex interactions of the humidity with the clouds explain the previous stated RMSE errors of the ICON GER and

COSMO simulations. There is again no added value in the integrated water vapour of the ICON GER 156 m simulations detectable compared to the coarser-resolved ICON GER 624 m, 312 m and COSMO simulations.

Presumably, issues of the soil moisture model or uncertainties of the ground-to-atmosphere exchange processes or of the latent heat flux might be reasons for the mentioned deviations in the specific humidity and should be further analysed. Most notable are the large errors found at the diurnal cycle of the 2 m specific humidity. Most probably, the too-strong coupling of the model levels during the night cause a down mixing of drier air from higher altitudes leading to a substantial underestimation of the 2 m specific humidity by the ICON GER simulations. In the daytime, most likely a too-weak coupling of the model levels reduces the mixing with higher layers and increases the influence of the soil moisture, which leads to a remarkable overestimation of the 2 m specific humidity visible for all ICON GER simulations.

In summary, the representation of the specific humidity of the different ICON GER simulations is of a similar quality to that of the COSMO simulations, except for substantial differences found at the diurnal cycle of the 2 m specific humidity. Overall, the different simulations agree well with the observations. In contrast to the expectations of a higher accuracy due to the increased resolution and explicit simulation of certain processes, like the deactivated cumulus parameterisation of the ICON GER compared to the COSMO model, no substantial improvements are identified regarding the specific humidity. The remarkable deviations at the diurnal cycle of ICON GER compared to the observations indicate for relevant model issues that should be analysed and fixed in future ICON LES versions.

Impact of Atmospheric State Uncertainties

The objective of the evaluation pyramid is first to check basic atmospheric quantities like boundary conditions or flow conditions to investigate whether more complex quantities like clouds and precipitation are simulated correctly for the right reason or only because of compensating errors (Sect. 2.1). This concept is tested in an exemplary manner for all simulations by the 2 m specific humidity. For this, a selection of low- and high-quality data for the basic atmospheric quantities of mean sea level pressure (boundary conditions), wind speed (flow conditions) and 2 m temperature (stratification) is used. The absolute difference between the 10-minute interval DWD weather station observations and the corresponding model dataset is calculated, and a threshold for the lowest/highest 25% deviations is determined for each dataset separately to define values of good and bad quality. The time series of the 2 m specific humidity measured at the DWD weather stations

Table 3.7.: Error dependence of 2 m specific humidity in an increasingly quality-controlled (QC) basic atmospheric state following the evaluation pyramid (Fig. 2.1). The upper table displays the RMSE of the best 25% of all data, and the lower table is based on the worst 25% values in each QC step. The data are filtered first for the boundary conditions (Bound.), then for the flow conditions (Flow.), and then for the stratification (Strat.).

Specific Humidity RMSE - Good $(\mathrm{g\,kg^{-1}})$	All Data	Bound. QC	Bound./Flow. QC	Bound./Flow./ Strat. QC
ICON GER 624 m	1.26	1.18	1.13	0.97
ICON GER 312 m	1.29	1.20	1.16	0.95
ICON GER 156 m	1.32	1.25	1.21	0.96
COSMO	1.14	0.95	0.88	0.75

Specific Humidity RMSE - Worst $(\mathrm{g\,kg^{-1}})$	All Data	Bound. QC	Bound./Flow. QC	Bound./Flow./ Strat. QC
ICON GER 624 m	1.26	1.32	1.45	1.72
ICON GER 312 m	1.29	1.36	1.50	1.77
ICON GER 156 m	1.32	1.39	1.54	1.80
COSMO	1.14	1.44	1.54	1.79

is first filtered by the 25% best/worst mean sea level pressure values, followed by the same analysis for wind speed and stratification. The resulting error propagation of all models is compiled in Table 3.7.

As expected, well-represented boundary, flow, and stratification conditions reduce the RMSE of the 2 m specific humidity by 23% for the ICON GER 624 m (1.26 to 0.97 $\mathrm{g\,kg^{-1}}$) and by up to 34% for the COSMO model (1.14 to 0.75 $\mathrm{g\,kg^{-1}}$), which can be seen in the upper part of Table 3.7. Similarly, the error increases by 36% for ICON GER 156 m (1.32 to 1.80 $\mathrm{g\,kg^{-1}}$) and by up to 57% for COSMO (1.14 to 1.79 $\mathrm{g\,kg^{-1}}$), if the basic atmospheric state is not well represented (Tab. 3.7, bottom). The stratification has the largest impact on the quality of the specific humidity, which increases the quality by up to 20% for the ICON GER 156 m instead of about 5% for the boundary conditions and only 3% for the flow conditions. The temperature determines how much humidity the air can contain, and therefore the stratification might be the most relevant quantity. Nevertheless, the air pressure has also an influence on the density and the flow conditions, such as on the advection and distribution of humidity, which are subsidiary processes.

This example analysis proves the hypothesis that a good simulation of the basic atmospheric quantities is crucial for an accurate representation of more complex atmospheric quantities like humidity and clouds. Additionally, the humidity example proves the concept of the evaluation pyramid approach.

3.7. Precipitation

Clouds and precipitation are the most complex atmospheric processes influenced by all other in previous sections evaluated quantities like wind, temperature, and humidity. A detailed cloud evaluation is presented in Chapter 4, focusing on the cloud macrophysical properties. Concerning the precipitation, only a brief and general evaluation is conducted for the available ICON GER simulations, as the focus of this chapter is on the basic atmospheric state.

Detailed investigations of the modelled precipitation are done by Nürenberg (2018) and Heinze et al. (2017a) using the first available Germany-wide ICON GER prototype runs. Heinze et al. (2017a) analysed the 312 m-resolved ICON GER simulation of 26 April 2013 with a frontal passage across Germany. The authors found that precipitation rates below about $16\,\mathrm{mm\,hr^{-1}}$ are underestimated and larger precipitation rates are overestimated by the ICON GER simulations compared to the RADOLAN measurements. In contrast, the COSMO simulations generally underestimate the precipitation rates compared to the observations.

Nürenberg (2018) evaluated three simulated days (26 April 2013, 11 May 2013, 28 May 2013) of the coarser 624 m-resolved ICON GER in respect to precipitation rates and compared the ICON GER output to the RADOLAN observations and COSMO reference simulations. According Nürenberg (2018), the frequency of occurrence of small precipitation rates of less than $2\,\mathrm{mm\,hr^{-1}}$ is overestimated by the ICON GER and COSMO simulations, whereas precipitation rates between $2\,\mathrm{mm\,hr^{-1}}$ and $8\,\mathrm{mm\,hr^{-1}}$ occur too rarely within the ICON GER and COSMO simulations compared to the RADOLAN observations. Intense precipitation rates of more than $8\,\mathrm{mm\,hr^{-1}}$ are overestimated in their frequency by the ICON GER simulations, as stated by Nürenberg (2018). The COSMO simulations also underestimate those intense precipitation rates. Consequently, an in-depth evaluation of the simulated precipitation of the twelve new simulated days with the ICON GER would be valuable to assess its accuracy and gather statistically robust results. Nevertheless, such a detailed assessment is beyond the scope of the baseline evaluation and therefore only a general evaluation of the simulated precipitation is conducted.

The precipitation representation within the ICON GER simulations is evaluated in terms of the precipitation rate distributions using the RADOLAN radar observations (Weigl et al., 2004) as reference. Additionally, the COSMO simulations are included as a benchmark. The quality-controlled RADOLAN data are obtained from the SAMD database (Lammert et al., 2019) to avoid measurement errors like radar spikes at the evaluation. All model datasets are regridded to the common RADOLAN grid with $1\times1\,\mathrm{km}$ spatial resolution by a nearest-neighbour approach. The five-minute instantaneous ICON GER data are
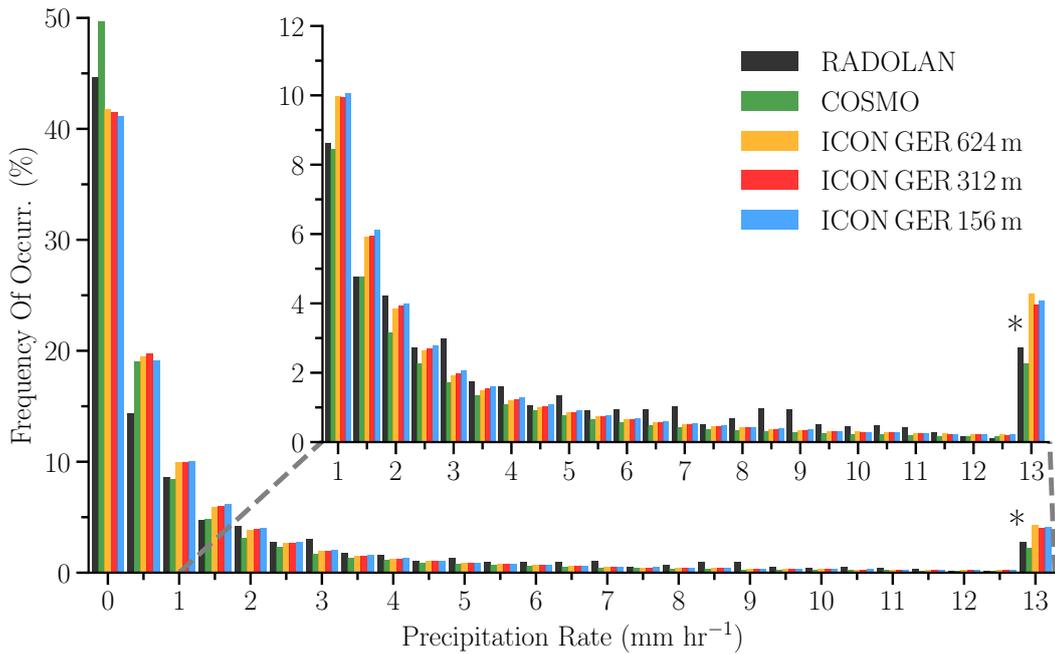
Figure 3.23.: Frequency of occurrence (Freq. Of Occurr.) of precipitation rates using all twelve simulated days based on five-minute data for DWD RADOLAN observations and COSMO, ICON GER LES 624 m, 312 m, 156 m model data. All datasets are regridded to the common RADOLAN resolution of 1 x 1 km. Histograms are normalised for each dataset separately. Bin size is $0.5 \, \text{mm} \, \text{hr}^{-1}$ and last bin with "*" contains all events between $13\text{-}100 \, \text{mm} \, \text{hr}^{-1}$.

matched with the five-minute RADOLAN observations. The coarser 2.8 km-resolved COSMO data are interpolated by a distance-weighted interpolation to the RADOLAN resolution and the original 15-minute data are used for the statistical analysis. All precipitation rates below $0.1 \, \text{mm} \, \text{hr}^{-1}$ and above $100 \, \text{mm} \, \text{hr}^{-1}$ are filtered to exclude most probably measurement errors and observation uncertainties regarding very low precipitation rates. Missing data of the radar or model are homogenised across all datasets.

Overall, the different ICON GER and COSMO simulations fit well to the radar observations regarding the frequency of occurrence of precipitations rates, which is visible by deviations of less than 5 percentage points (p.p.) (Fig. 3.23). The precipitation rates of the observations and of the different ICON GER and COSMO simulations are, as expected, log-normally distributed (Cho et al., 2004). The largest deviations are seen for low rain rates below $2 \, \text{mm} \, \text{hr}^{-1}$, where all models overestimate the frequency of occurrence, which is in agreement with the results of Nürenberg (2018). In particular, the range of $0.5\text{–}1.0 \, \text{mm} \, \text{hr}^{-1}$, representing drizzle, is overestimated by up to 5 p.p. by the ICON GER and COSMO output. Likewise, the COSMO simulations overestimate rain rates below $0.5 \, \text{mm} \, \text{hr}^{-1}$, depicting very light drizzle, by about 5 p.p. which are well represented by
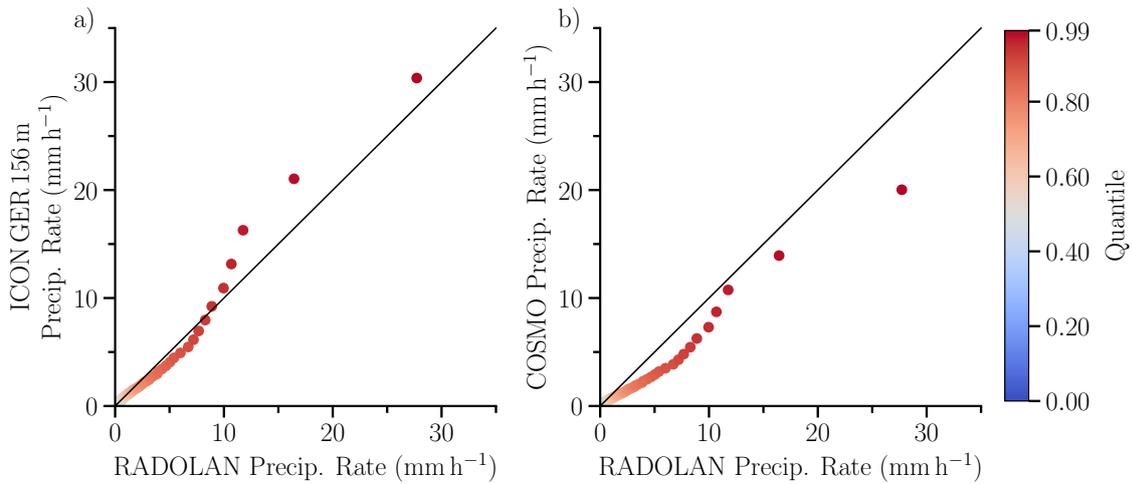
Figure 3.24.: Quantile-Quantile comparison of precipitation rates between reference observations of the $1 \times 1\,\mathrm{km}$ resolved DWD RADOLAN observations and corresponding ICON GER 156 m (a) and COSMO (b) simulation output using all 12 simulation days. All data is regridded to the common RADOLAN resolution of 1 km.

the ICON GER simulations. Precipitation rates between $2\,\mathrm{mm\,hr^{-1}}$ and $11.5\,\mathrm{mm\,hr^{-1}}$ are slightly underestimated in their frequency by less then 1 p.p. by the different ICON GER and COSMO simulations.

Intense precipitation with precipitation rates of more than $12\,\mathrm{mm\,hr^{-1}}$ is simulated too frequently by the ICON GER simulations compared to the observations. Considering all precipitation rates above $13\,\mathrm{mm\,hr^{-1}}$, the frequency of occurrence is overestimated by almost 5 p.p. in comparison with the RADOLAN measurements. In opposite, these intense precipitation events occur slightly too rarely within the COSMO simulations, which is detectable by an underestimation of less than 1 p.p.. These findings are similar to the results of Nürenberg (2018), who stated an underestimation of precipitation rates between $2\,\mathrm{mm\,hr^{-1}}$ and $8\,\mathrm{mm\,hr^{-1}}$ and an overestimation of precipitation rates above $8\,\mathrm{mm\,hr^{-1}}$ for the analysed three ICON GER 624 m simulations. The different resolved ICON GER simulations are very similar, and no added value is visible for the highest-resolved ICON GER 156 m in terms of the precipitation rates.

Furthermore, the distributions of the simulated and observed precipitation rates are investigated in more detail by comparing the quantiles of the different datasets (Upton and Cook, 1996). First, the datasets are sorted with increasing precipitation rate and then the quantiles are calculated. For example, the 0.10 quantile illustrates the precipitation rate, for which 10% of all precipitation rates are below this threshold. The 100 equally sized quantiles between 0.00 and 0.99 are calculated for the ICON GER and COSMO simulations, as well as the RADOLAN observations for the evaluation of the precipitation

rates. The computed quantiles of the ICON GER 156 m and COSMO simulations are compared with the quantiles of the observations (Fig. 3.24). Specific quantiles of the different datasets are listed in Appendix Table B.2.

The simulated distribution of the ICON GER 156 m precipitation rates agrees well up to the median with the RADOLAN observations with deviations of less than $0.03\,\mathrm{mm\,hr^{-1}}$. The previously mentioned underestimation of precipitation rates of about $3\text{-}12\,\mathrm{mm\,hr^{-1}}$ is again visible by lower quantile values in this range of the ICON GER 156 m simulations compared to the observations. Intense precipitation of more than about $13\,\mathrm{mm\,hr^{-1}}$ is overestimated by the ICON GER 156 m simulations in comparison with the RADOLAN data. This is identifiable by larger quantile values above the 0.94 quantile of the ICON GER 156 m simulations than the measurements. The quantiles of the precipitation rates of the COSMO simulations are overall underestimated compared to the observations. The median of the COSMO simulations is for instance lower by $0.21\,\mathrm{mm\,hr^{-1}}$ than RADOLAN. The stated results are similar to the findings of Nürenberg (2018).

Possibly, an overestimation of small and short-living convective rain cells by the ICON GER simulations, as stated by Heinze et al. (2017a) and Nürenberg (2018), might explain the found too intense precipitation rates compared to the observations. However, a more in-depth evaluation is required to investigate the detailed quality of the simulated precipitation within the ICON GER and COSMO simulations. Further research would be valuable to investigate the underlying reasons causing for instance the found differences at the precipitation rate distributions between the ICON GER and the observations.

In conclusion, the general evaluation of the precipitation rates of the ICON GER simulations exhibit a reasonable agreement with the observations. The uncertainties are similar to those of the well-established COSMO model with deviations regarding the frequency of occurrence of less than 5 p.p. compared to the observations, except for drizzle. The frequency of occurrence of drizzle is overestimated by up to 5 p.p. by the ICON GER simulations compared to the observations. Low rain rates of about $3\text{-}12\,\mathrm{mm\,hr^{-1}}$ are underestimated by the ICON GER and COSMO simulations. In contrast, intense precipitation of more than $13\,\mathrm{mm\,hr^{-1}}$ is overestimated by the ICON GER simulations, which are also underestimated by the COSMO simulations. However, even for these more complex quantities, no added value of the high-resolution ICON GER is found compared to the coarser COSMO simulations. Therefore, a comprehensive precipitation evaluation for further parameters as for instance of the size distribution of single rain cells and their lifetime or features of severe convection would be interesting to investigate the different ICON GER simulations in respect to an added value. For example, the explicit simulation of shallow convection due to the high-resolution of the ICON GER simulations might lead to an added value of higher-order parameters compared to the coarser 2.8km resolved COSMO simulations.

Summary

A comprehensive evaluation of the basic atmospheric state according to the evaluation pyramid concept is conducted for the Germany-wide ICON GER simulations for spatial resolutions of 624 m, 312 m, and 156 m. Furthermore, various observational datasets and 2.8 km cloud-resolving COSMO simulations are included as reference. Conditions near the ground, within the boundary layer, as well as in the full troposphere, were analysed. Based on the results of this chapter, the scientific questions posed in the beginning are answered as follows:

1. **How well is the basic atmospheric state represented by the ICON GER at different resolutions?**

   The basic atmospheric state including the pressure, wind, temperature, and humidity of all ICON GER simulations is overall well-represented for the near-ground conditions, boundary, and full-troposphere profiles compared to the observations. On average, the RMSE of the near-ground parameters of the ICON GER simulations are all substantially lower by about 48% than the observed natural variability, depicted by the standard deviation of the observations, for the various parameters. The largest errors of the 10 m wind speed and 2 m temperature are identifiable at regions with complex terrain for all ICON GER simulations. The random error dominates the RMSE for most considered parameters, and only small bias values are found, which usually account for less than 13%. However, a large bias is seen for the mean sea level pressure. Two out of the twelve considered days also show a considerably larger standard deviation for the mean sea level pressure.

   Most remarkable are the deviations at the diurnal cycle of the near-ground wind speed, temperature, and specific humidity of all ICON GER simulations compared to the observations. Possibly a too-strong coupling of the model levels by the ICON GER simulations during night induces an overestimation of the downward mixing of momentum from the layers above, which results in a substantial overestimation of the 10 m wind speed by up to $0.5\,\mathrm{m\,s^{-1}}$ and 20%. The decrease of the 10 m wind speed after sunset because of the layer decoupling is completely missed by all ICON GER simulations. Furthermore, the too-strong coupling might also cause an increased downward mixing of warmer and drier air from higher altitudes, which leads to an overestimation of the 2 m temperature by up to 1 K and a remarkable underestimation of the 2 m specific humidity by up to $1.25\,\mathrm{g\,kg^{-1}}$ and 14%.

   In contrast, presumably a too-weak coupling of the model levels by the ICON GER model in the daytime reduces the downward mixing of momentum and causes ac-

cordingly a distinct underestimation of the $10\,\mathrm{m}$ wind speed by up to $0.5\,\mathrm{m\,s^{-1}}$ and 14%. Similarly, the influence of the ground on the near-ground conditions could be higher due to the reduced mixing with the air from higher layers. Most likely, the higher influence of the ground leads to a too-early decrease of the $2\,\mathrm{m}$ temperature by one hour after the maximum at noon. Additionally, the reduced downward mixing of dry air during the day could explain the notable underestimation of the $2\,\mathrm{m}$ specific humidity of all ICON GER simulations by up to $0.5\,\mathrm{g\,kg^{-1}}$ and 5%.

On average, the ICON GER simulations are too dry, and the $2\,\mathrm{m}$ specific humidity is underestimated by about -$0.2\,\mathrm{g\,kg^{-1}}$ and 2-3%. The simulated IWV of the different ICON GER simulations agree well with the observations. The boundary layer height is well represented during the night by the ICON GER simulations, whereas a substantial underestimation by up to a factor of two and of more than $700\,\mathrm{m}$ is seen in the early evening. The brief evaluation of the complex precipitation shows an overall reasonable agreement with the observations. The frequency of occurrence of drizzle of about $0.5\,\mathrm{mm\,hr^{-1}}$ is overestimated by about 5 p.p.. Precipitation rates between 3–$12\,\mathrm{mm\,hr^{-1}}$ are underestimated, whereas intense precipitation of more than $12\,\mathrm{mm\,hr^{-1}}$ is overestimated by the ICON GER simulations.

2. **Is there an added value of high-resolution LES detectable for basic atmospheric quantities compared to an operational weather forecast model?**

The three high-resolution ICON GER domains with a horizontal resolution of down to $156\,\mathrm{m}$ are compared to $2.8\,\mathrm{km}$-resolved COSMO simulations. In contrast to the expectations of a higher simulation accuracy by the increased resolution, there is no added value visible for the higher-resolved ICON GER regarding the basic atmospheric state quantities. The uncertainties of most considered parameters of the ICON GER are slightly larger by up to 20% than that of the well-tuned COSMO model. Even more notable is that for most quantities, the errors are largest for the highest-resolved ICON GER $156\,\mathrm{m}$ simulation, and the best results are seen for the coarser ICON GER $624\,\mathrm{m}$ simulation. Regarding the near-ground, boundary-layer, and full troposphere profiles of the wind speed, temperature, and specific humidity, the RMSE of the ICON GER $156\,\mathrm{m}$ is on average about 7% larger than that of the ICON GER $624\,\mathrm{m}$. The RMSE of the ICON GER $312\,\mathrm{m}$ is on average about 3% larger than the coarser resolved $624\,\mathrm{m}$ ICON GER simulations. Nevertheless, the found uncertainties of the ICON GER and COSMO model are already in a similar range. This is remarkable for a completely novel and untuned model in comparison with an operational, well-tuned weather forecast model like the COSMO. Additionally, substantial improvements in the accuracy are detected for the twelve new ICON GER

simulations with the advanced ICON LES model compared to the ICON GER prototype runs with the erroneous ICON LES version, which were analysed by Heinze et al. (2017a). Nevertheless, further investigations of the ICON GER are crucial to improve the found issues of the various evaluation analyses.

Chapter 4

Cloud Evaluation by Classification

Clouds are one of the most beautiful but at the same time most complex quantities of the atmosphere with almost an infinite number of characteristics to analyse. These characteristics like cloud distribution, structure, phase, development, and micro- and macrophysical properties with their high temporal and spatial variability make clouds very challenging to evaluate. Furthermore, many cloud properties like the ice or liquid water path can be directly determined by model quantities, yet they cannot be directly observed. These quantities need to be calculated from measurements with statistical retrievals using for example brightness temperatures at specific wavelengths.

New realistic large eddy simulations provide model output at a temporal and spatial resolution similar to those of observations. Forward models, running well-known radiative transfer simulations for given tropospheric profiles, allow the generation of physically consistent virtual measurements. Therefore, atmospheric models can be directly compared to the observed quantities, avoiding uncertainties due to retrievals at the evaluation. This chapter will focus on the following two scientific questions:

1. **Which advantages and disadvantages provide a forward-simulated cloud classification compared to a direct model output-based one?**

2. **How well are clouds represented by the ICON HOPE LES from a forward simulation perspective and in terms of a cloud classification?**

4.1. The Cloudnet Project and Target Classification

The Cloudnet project, started in 2002, developed a framework for an almost near-real-time cloud evaluation of operational numerical weather prediction models based on specially equipped Cloudnet supersites (Illingworth et al., 2007; Hogan et al., 2009). The various automatically generated analyses and forecast scores provide model developers fast feedback from current model simulations and improvements of new model developments. A worldwide community of scientists steadily advances the algorithms and products, and integrates new supersites. The implemented numerical weather prediction (NWP) models range from global to small, limited-area models like the COSMO-DE with 2.8 km of the DWD. The first three Cloudnet supersites were installed in Chilbolton (United Kingdom), Jülich (Germany), and Copenhagen (Denmark) (Illingworth and CloudNet-Team, 2004). The network has been continuously growing to currently about 30 supersites worldwide participating in the project. Every Cloudnet supersite is specially equipped with a cloud radar, a lidar or ceilometer, a dual- or multi-wavelength microwave radiometer, and a rain gauge. All these measurements are combined with the temperature and humidity profiles of an atmospheric model by the advanced Cloudnet algorithms (Illingworth et al., 2007).

The measured radar reflectivity, Doppler velocity, spectral width, and linear depolarisation ratio of the cloud radar is used to infer the overall cloud shape, height of the melting layer, and much more. The lidar observations help with the cloud base height detection of thin liquid clouds, not accurately determined by the cloud radar. The liquid water path (LWP) is calculated according to Karstens et al. (1994), based on the microwave radiometer measurements and used among other for the computation of the liquid water content within the troposphere. All observations are automatically quality-flagged, attenuation corrected, and afterwards combined to an advanced and intuitive multi-sensor product, the Cloudnet target classification (Hogan and O'Connor, 2006). The Cloudnet target classification for 26 April 2013 is illustrated as an example in Figure 4.1.

This product provides detailed information about the cloud structure and phase by a time versus height slice up to 12 km. The classification differentiates between eleven different targets, namely: "Clear Sky", "Cloud droplets only", "Drizzle or rain", "Drizzle/rain & cloud droplets", "Ice", "Ice & supercooled droplets", "Melting ice", "Melting ice & cloud
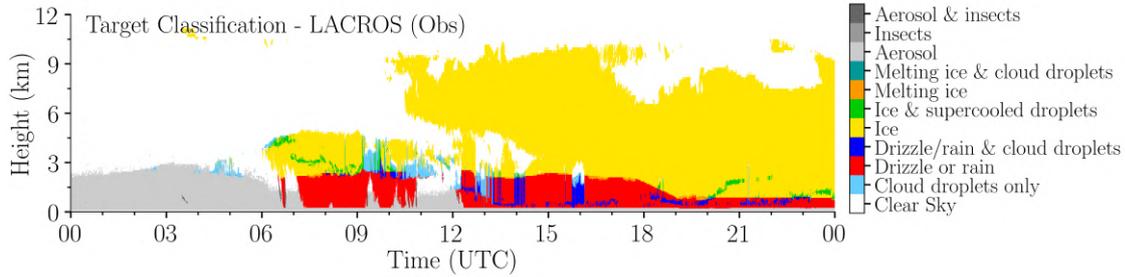
Figure 4.1.: Example Cloudnet target classification product with all categories for 26/04/2013 using LACROS HOPE measurement data.

droplets", "Aerosol", "Insects" and "Aerosol & Insects". Complex and diverse remote-sensing measurements are therefore combined to a quick and easy-to-understand product. Nevertheless, the target classification has currently only been used to derive advanced model parameters like cloud fraction, ice-, and liquid water path, but is not directly compared to the model. So far, no suitable model output was available for the evaluation of the target classification. For that reason, two different approaches are presented in the upcoming two sections to generate a comparable cloud classification for the output of atmospheric models.

4.2. Synthetic Cloud Classification

The Cloudnet target classification is a detailed and comprehensive dataset for the evaluation of cloud macrophysics, but so far it has only been available for observations and not for atmospheric models. Two different concepts to derive a model surrogate are developed and explained in this study (Fig. 4.2). The first one is based on the direct model output using the temperature, dew point, and hydrometeor concentration profiles (Sect. 4.2.1). The second approach uses forward operators to create synthetic measurements for all necessary Cloudnet instruments on which the original Cloudnet algorithms are applied afterwards (Sect. 4.2.2) to create a physical consistent cloud classification.

4.2.1. Model to Cloud Classification Algorithm

The direct model output based "model to cloud classification" (MC) algorithm requires the single-column output of all available specific hydrometeor masses and the temperature and dew point profiles of an atmospheric model. The hydrometeor categories of cloud liquid water (QC), cloud ice (QI), rain (QR), graupel (QG), hail (QH), and snow (QS) are currently included in the classification algorithm, which can be easily extended. Based on physical principles, the algorithm (Fig. 4.3) calculates a cloud classification inspired by the Cloudnet target classification with the same categories. All aerosol and
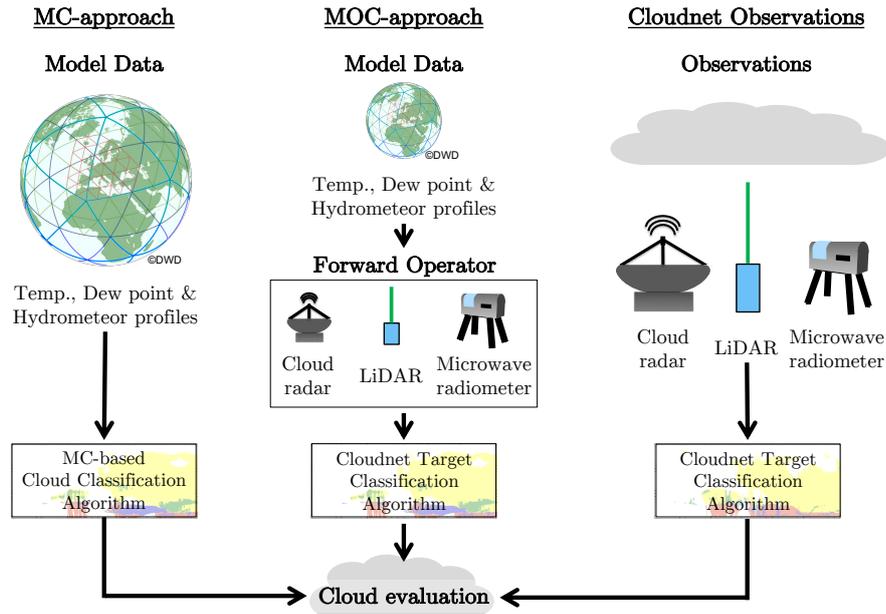
75

Figure 4.2.: Overview of the three different approaches to generate a cloud classification: the model to cloud classification (MC) approach (left), the model to observation to classification (MOC) approach using forward operators to generate synthetic measurement data to which the Cloudnet algorithms are applied (middle) and the original Cloudnet target classification using observations (right), advanced from Hansen (2014).

insect categories are not considered because commonly used atmospheric models do not contain those target types. The MC algorithm determines for every single grid box at a certain height and time the cloud phase or clear sky conditions independently, which is one of the assumptions to be considered. For example, the category of "Ice" is determined based on a specific ice hydrometeor mass concentration exceeding a certain threshold, a specific cloud liquid water hydrometeor concentration not exceeding a certain threshold, and a dew point below zero degrees Celsius. The order of the case selections is crucial for physically consistent results. For example, if the category of "Melting ice & cloud droplets" were analysed before and afterwards the category of "Melting ice", all grid boxes would be set to "Melting ice" because the additional condition of $QC > m_{\text{thresh,QC}}$ for "Melting ice & cloud droplets" would not be considered at the classification of "Melting ice" and therefore overwritten by this category. The freezing of hydrometeors is calculated based on the dew point temperature, consistent with the Cloudnet algorithms (Illingworth et al., 2007). For every hydrometeor category, a certain threshold of a significant amount of each hydrometeor mass has to be chosen according to the model's cloud microphysics and observation characteristics. The same threshold of $10^{-6}\,\text{g}\,\text{kg}^{-1}$ is applied to all ICON hydrometeors within this study, which is a first estimate for a significant hydrometeor mass concentration neglecting numerical artefacts (Axel Seifert, personal communication) of the ICON model.
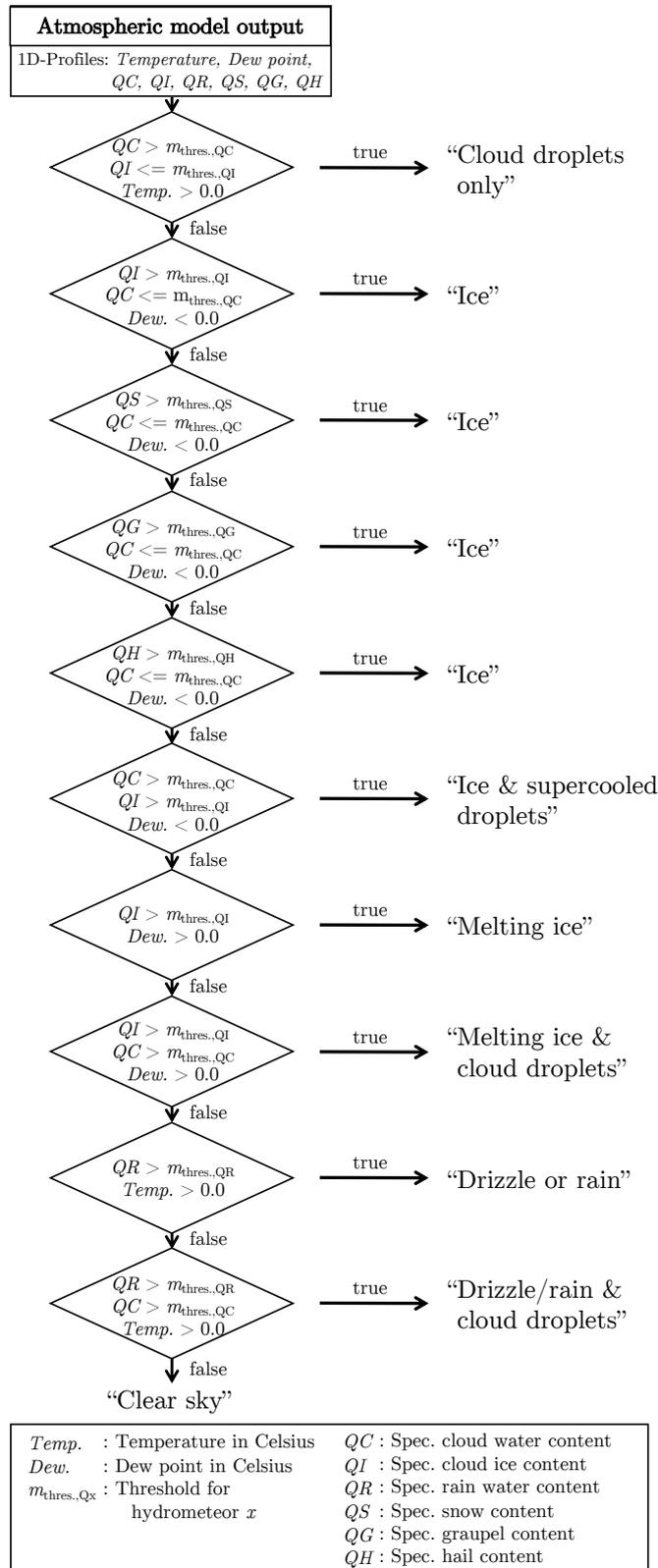
Figure 4.3.: Flowchart of the MC cloud classification algorithm using the output of an atmospheric model to generate a Cloudnet inspired target classification, differentiating clouds by eight different categories, enhanced from Hansen (2014).

Nevertheless, this rough assumption has to be kept in mind for the interpretation of the results. Different thresholds and further assumptions would, however, induce more complexity and thus lower the comprehensibility of the algorithm. Additionally, the algorithm does not take into account the characteristics of the instruments like the limited sensitivity or attenuation of remote-sensing devices, which will be very complex to determine in a physically correct way and probably cause larger uncertainties rather than neglecting them. For these reasons, the MC-algorithm is designed in the simplest way to ensure a quick applicability to new model data while at the same time trying to be as physical consistent as possible with the Cloudnet output. The structure also increases comprehensibility and easy traceability of the classification generated. The newly developed algorithm is applied to the two months of ICON HOPE data (Sect. 2.2) and results are analysed in the upcoming sections.

4.2.2. Model to Obs to Classification Approach

In the forward-simulated "Model to Observation (Obs) to Classification" (MOC) approach, different forward models are applied to the atmospheric model output to create synthetic observations (Model to Obs), which are physically consistent with the model's assumptions of the cloud microphysics. Additionally, they incorporate the properties of real instruments like attenuation. Afterwards, exactly the same Cloudnet algorithm is applied to the virtual Cloudnet supersite to create a fully synthetic Cloudnet target classification (Model to Obs to Classification), (Fig. 4.2).

Forward operators for all three basic Cloudnet supersite remote-sensing instruments of a cloud radar, lidar, and microwave radiometer are necessary to generate all synthetic measurements for the Cloudnet algorithms. Additionally, the in situ rain gauge is derived directly from the model output. Most forward operators use the single-column output of temperature, humidity, pressure, specific hydrometeor masses together with the number concentrations if available, and a description of the cloud microphysical scheme as input (Fig. 4.4). Accordingly, the forward operator has detailed information about the atmospheric conditions and can run the well-known radiative transfer for a certain frequency to simulate physically consistent measurements including instrument characteristics (Fig. 4.4). To derive the temperature profile from a commonly used microwave radiometer, for example, only 14 brightness temperature measurements at different frequencies can be used. These observed brightness temperatures are then combined with a typically statistically-based retrieval to derive a full tropospheric temperature profile with values at various heights from a few measurements. In contrast, forward operators get all required information for the radiative transfer calculation about the atmosphere from the atmospheric model to compute for instance only 14 brightness temperatures from 150 dif-
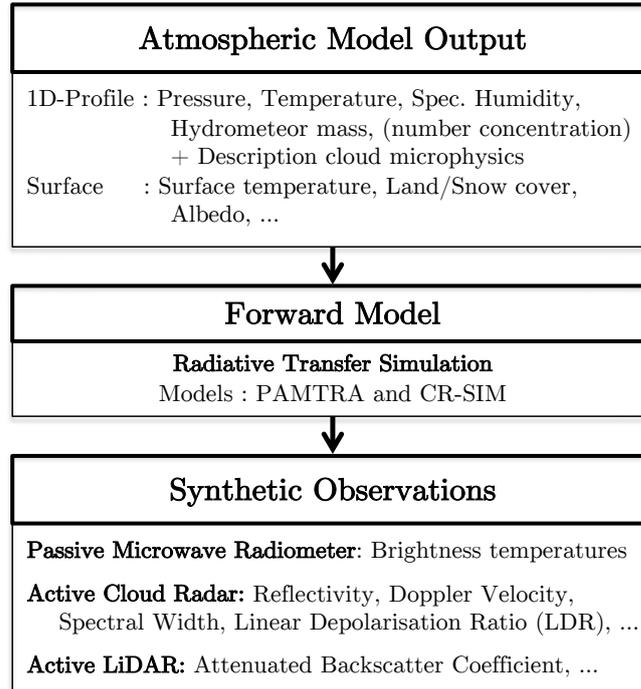
```
┌─────────────────────────────────────────────────┐
│           Atmospheric Model Output              │
├─────────────────────────────────────────────────┤
│ 1D-Profile : Pressure, Temperature, Spec. Humidity,│
│              Hydrometeor mass, (number concentration)│
│              + Description cloud microphysics    │
│ Surface    : Surface temperature, Land/Snow cover,│
│              Albedo, ...                         │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────────────────────────┐
│                 Forward Model                   │
├─────────────────────────────────────────────────┤
│           Radiative Transfer Simulation         │
│           Models : PAMTRA and CR-SIM            │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────────────────────────┐
│              Synthetic Observations             │
├─────────────────────────────────────────────────┤
│ Passive Microwave Radiometer: Brightness temperatures│
│                                                 │
│ Active Cloud Radar: Reflectivity, Doppler Velocity,│
│     Spectral Width, Linear Depolarisation Ratio (LDR), ...│
│                                                 │
│ Active LiDAR: Attenuated Backscatter Coefficient, ...│
└─────────────────────────────────────────────────┘
```

Figure 4.4.: Flow chart of the PAMTRA and CR-SIM forward operators to simulate synthetic measurements based on the output of an atmospheric model, adapted from Hansen (2014).

ferent model levels of the ICON HOPE. Additionally, no retrieval is necessary, which might induce further uncertainties. In combination with the physical consistency regarding the model output, these are the major advantages of this approach over derived model quantities by typical statistical retrievals. Nevertheless, the model's cloud microphysical scheme has to be accurately implemented by the forward model and small differences can induce also uncertainties in the forward simulation.

An additional advantage of the forward operator approach over retrievals of observations are the generated synthetic observations themselves, which can be directly compared to the measurements in terms of the observed quantities. These measurements often exhibit detailed information about underlying physics. Furthermore, the virtual observations help to create a link between the modelling and observational experts because the model output can be handled like real observations.

Two different forward models are used to simulate all remote-sensing instruments required by the Cloudnet algorithms. The active cloud radar and passive microwave radiometer instruments are generated by the Passive and Active Microwave TRAnsfer model (PAMTRA) of the University of Cologne. The PAMTRA forward operator has already been widely tested by studies like Maahn et al. (2015); Marke et al. (2018) and Cadeddu et al. (2017). Additionally, a physically consistent configuration for the ICON LES model

is already available, which is for instance used by Heinze et al. (2017a). PAMTRA does not contain a forward operator for a lidar. Therefore, the lidar is generated by the Cloud Resolving Model Radar Simulator (CR-SIM) of Stony Brook University. CR-SIM can also simulate a cloud radar and microwave radiometer, but the ICON LES cloud microphysics of Seifert and Beheng (2006) were just recently integrated as part of this study and are yet not well evaluated. For that reason, the PAMTRA and CR-SIM are used as forward operators in this study.

PAMTRA Forward Operator

The PAMTRA model can simulate active and passive instruments with frequencies in the microwave band (Kollias et al., 2011; Maahn et al., 2015; Cadeddu et al., 2017). Both ground-based and space-borne/airborne instruments, either on satellites or on research aircrafts, can be simulated. The radiative transfer model is based on the RT4 model of Evans and Stephens (1995) and uses a one-dimensional plane-parallel assumption for the radiative transfer calculations. The radar reflectivities are computed following Smith (1984). The cloud microphysics and hydrometeor size distribution are configured consistently with the two-moment scheme of Seifert and Beheng (2006) of the ICON LES model. The single-scattering characteristics are adopted according to the two-moment cloud microphyiscs of the ICON model, as well as the soft-sphere approximation with fixed densities and the discrete dipole approximation. The Mie theory for single-scattering of frozen hydrometeors like ice, snow, hail, and graupel is assumed with constant densities.

As input for the PAMTRA simulation, the ICON HOPE 156 m surface temperature, albedo, land fraction, and single-column output of temperature, specific humidity, pressure, height, specific hydrometeor masses, and their number concentrations are necessary for the two-moment microphysical schemes. Additionally, the vertical profiles of the horizontal and vertical wind, as well as of the turbulent kinetic energy (TKE), are required for the computation of higher-order cloud radar moments like the spectral width. The high-resolution single-column output of the two-month ICON HOPE 156 m run (Sect. 2.2) is used as input for PAMTRA. An active polarimetric 35 GHz cloud radar is simulated providing synthetic measurements of the cloud radar reflectivity, mean Doppler velocity, spectral width, and linear depolarisation ratio (LDR). The instrument's characteristics are configured physical consistently to the real METEK 35 GHz cloud radar of LACROS. In addition, a synthetic passive 14-channel microwave radiometer with seven channels of the water vapour window around 22 GHz and seven channels of the oxygen line around 58 GHz is generated by PAMTRA. This instrument is also set up according to the real Radiometer Physics GmbH Humidity And Temperature PROfilers (HATPRO) instrument of LACROS. The liquid water path required by the Cloudnet algorithms is computed afterwards, as to be consistent with the observations, based on the synthetic brightness temperatures.

CR-SIM Forward Operator

The CR-SIM is developed by the Applied Radar Science Group of the Stony Brook University (Tatarevic et al., 2019) to simulate physical consistent cloud radar measurements for the Weather Research and Forecasting (WRF) model. The synthetic observations can be directly compared to real measurements without any retrievals. Additionally, a virtual profiling ceilometer at 905 nm and a micropulsed lidar with wavelengths of 353 nm and 532 nm have been integrated into recent versions. Ground-based and space-borne forward simulations are possible. CR-SIM is continuously extended by new atmospheric models with their cloud microphysical schemes. In this study, the two-moment cloud microphysics (Seifert and Beheng, 2006) of ICON were implemented together with Mariko Oue, Axel Seifert, and the CR-SIM team.

The scattering properties of spherical hydrometeors at the radiative transfer are calculated by a T-matrix method. Pre-computed look-up tables (LUTs) of the Mishchenko's T-matrix algorithm (Mishchenko, 2000) are used for non-spherical, complex scattering characteristics. The forward operator includes all available hydrometeor types of the various cloud microphysics with their specific masses and, if possible, their number concentrations, for which different scattering models are incorporated. The general input parameters of CR-SIM are the single column output of temperature, specific humidity, pressure, height, horizontal and vertical wind velocities, turbulent kinetic energy (TKE), and specific hydrometeor masses together with their number concentrations for two-moment schemes. All atmospheric model data have to be converted first to the WRF-model format structure to be used by CR-SIM, which is done for the ICON HOPE simulations.

The CR-SIM forward operator is also applied to the two-month ICON HOPE 156 m run to generate synthetic micropulsed lidar measurements as input for the Cloudnet algorithm (Sect. 4.1). The required observed attenuated backscatter of the 532 nm wavelength micropulsed lidar is calculated according to the Bohren and Huffman Mie (BHMIE) code (Craig F. Bohren, 1998) assuming only spherical hydrometeors. In addition, all rain hydrometeors are neglected by CR-SIM during the simulation of the lidar backscatter (Mariko Oue, personal communication). The output parameters of the simulated micropulsed lidar are the observed attenuated lidar backscatter at 532 nm, the lidar extinction coefficient, the lidar ratio, and the height of the first cloud base detected. Many more output parameters, as well as instruments like a cloud radar and ceilometer, can be included. The synthetic remote-sensing instruments of PAMTRA and CR-SIM are investigated in detail in the following sections.

4.3. The Virtual Cloudnet Supersite

The minimal instrumentation of a Cloudnet supersite consists of a cloud radar, microwave radiometer, lidar or ceilometer, and rain gauge. The cloud radar and microwave radiometer are forward simulated by PAMTRA, the lidar by CR-SIM, and the rain measurements are directly derived from the model output to create a virtual Cloudnet supersite for the ICON HOPE 156 m simulations. All synthetic instruments are configured consistently to the real instruments of the LACROS supersite and can be consequently compared directly to the measurement data in their native units without using any retrieval. The various observational parameters are analysed in detail to evaluate the cloud's representation of the ICON HOPE simulations. Furthermore, the synthetic instruments, used as input for the Cloudnet algorithms, are checked for issues.

The real non-virtual Cloudnet observations are provided by the mobile LACROS supersite of the HOPE campaign (Sect. 2.4). The instrument data are already pre-processed and homogenised by the Cloudnet algorithms to a common 30-second time resolution and 491 height levels up to 12 km with a vertical resolution of roughly 100 m decreasing with height. The first measurements start at about 200 m above ground due to the technical limitations of a real radar switching between sending and receiving (Rinehart, 2010) and the configured settings. The forward simulations of the synthetic instruments are based on the native nine-second ICON HOPE 156 m column output with 150 vertical model levels up to 21 km in height. The vertical resolution depends on the model layer thickness, which is 20 m for the lowest layer and increases with altitude up to about 170 m at 12 km height. The different temporal and spatial resolutions of the two datasets require data preprocessing for a consistent and fair comparison. For this, the real observations are linearly averaged to the ICON HOPE heights, and the forward simulated data is averaged over time to the 30-second resolution of Cloudnet, always using the coarser resolution to avoid interpolating unknown data. If for a certain time step one of the data sources has no data available, all datasets are commonly set to a missing value to capture the same atmospheric conditions.

In the following, the cloud's representation of the ICON HOPE 156 m is evaluated first from the observational perspective using newly created forward simulated measurements in combination with real observations. Afterwards, the cloud macrophysical properties are assessed by the two newly created synthetic cloud classifications for the ICON HOPE simulations. The forward simulated observations are analysed in terms of time series and two-dimensional histograms. An example two-day period of 26-27 April 2013 with a frontal passage in the afternoon of 26 April 2013 is selected for the time series including various cloud features like rain showers, liquid clouds, ice clouds, and intense precipitation. Therefore, these two days provide a good representation of the overall two-month time frame.

4.3.1. Cloud Radar Simulation

The cloud's representation of the ICON HOPE 156 m simulations is evaluated in terms of the simulated cloud radar of the PAMTRA forward operator and compared to the LACROS observations. The two selected days are used for a qualitative analysis and the two months for a statistical investigation.

At a first glance, the forward-simulated cloud radar quantities agree remarkably well with the real observations of the LACROS supersite for the selected two-day time-series (Fig. 4.5). This is especially noteworthy taking into account the matching point-to-point comparison with the deterministic ICON HOPE run. The large-scale pattern of the cold front passage, with deep convection starting in the afternoon of 26 April 2013, long-lasting rain, and high clouds afterwards, is well represented by the ICON HOPE. However, small liquid clouds such as at 05 UTC of 26 April 2013 are not captured by ICON HOPE, most probably because of the point-to-point comparison. The new initialisation of the ICON HOPE after 24 hours of integration causes the visible leap of all simulated parameters at midnight.

The cloud radar reflectivity (Fig. 4.5a,b) exhibits information about the general cloud shape and to a certain extent also about the hydrometeor types, as for instance ice has a much higher reflectivity than water droplets. The simulation of a melting layer is not yet implemented by PAMTRA and for that reason missed by the virtual cloud radar, clearly visible for instance at about 1.5 km height around 00 UTC for the observed radar reflectivity but not for the forward simulation.

Presumably frozen hydrometeors are substantially overestimated by the ICON HOPE above roughly 4-5 km in height, as is illustrated by higher reflectivities of up to 10 dBZ, as well as by a larger extent of areas with reflectivity values to the top compared to the observations (Fig. 4.5a,b). This is consistent with the peak at the two-dimensional histogram of the simulated radar reflectivities (Fig. 4.6b) between 8 km and 10 km at around -30 dBZ, which is not the case for the observations (Fig. 4.6a). Additionally, the marginal distributions of the virtual cloud radar are shifted to higher altitudes by roughly 2 km and to higher reflectivities by about 5 dBZ due to the overestimation of the frozen hydrometeors compared to the real instrument. Similar results are found by Heinze et al. (2017a) for the ICON LES when the authors analysed the first Germany-wide ICON LES prototype simulations. A too-high ice nucleation rate or a too-low threshold for the nucleation of frozen hydrometeors of the cloud microphysical scheme might be possible hypotheses explaining the overestimation of the frozen hydrometeors by the ICON HOPE simulations.

The Doppler velocity (Fig. 4.5c,d) is the second moment of a cloud radar measurement and gives information about the mean fall velocity of particles. For example, large, heavy
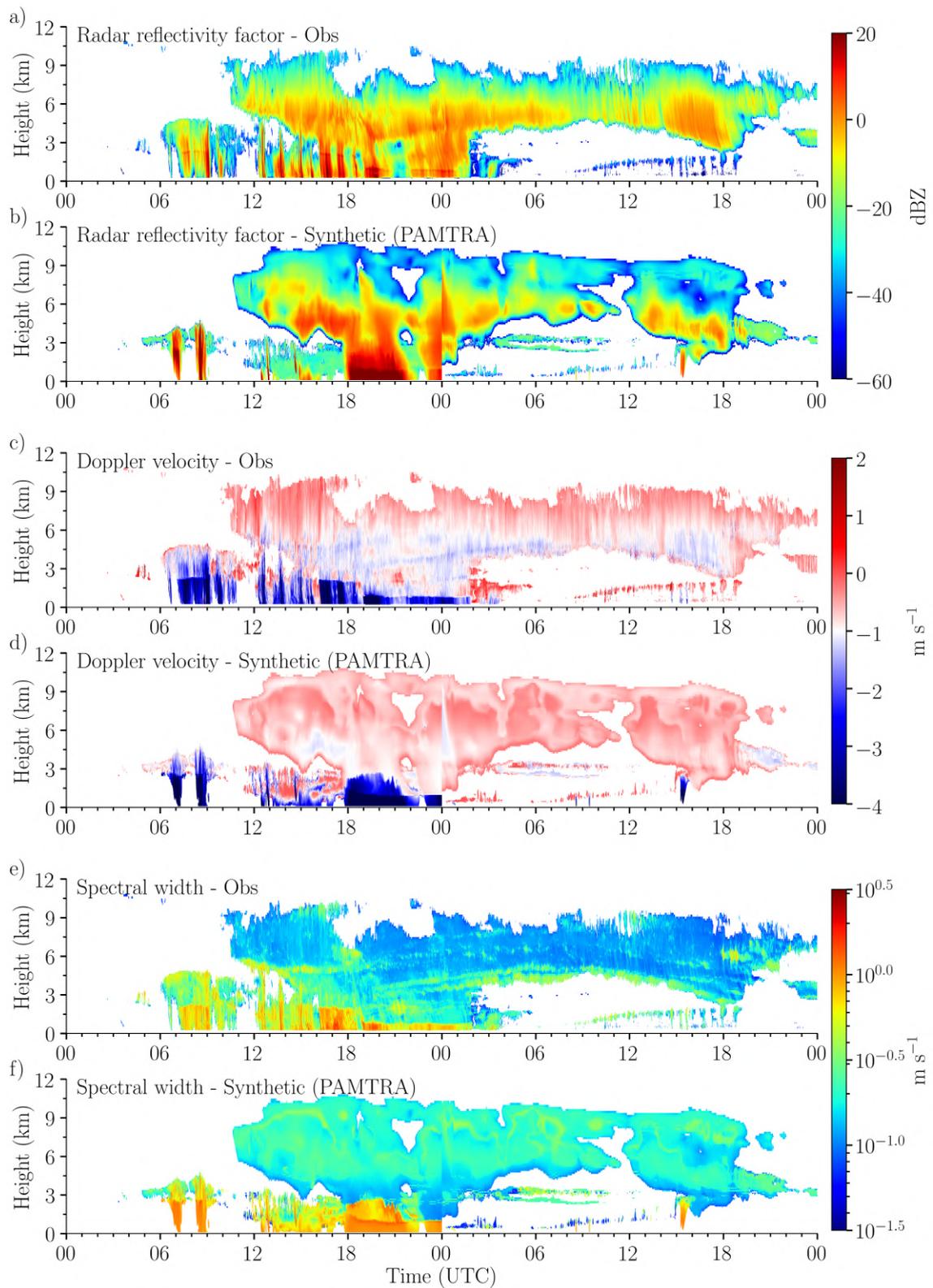
Figure 4.5.: Cloud radar reflectivity (a,b), Doppler velocity (c,d) and spectral width (e,f) of LACROS HOPE supersite observations (a,c,e) and of PAMTRA simulation (b,d,f) using the ICON HOPE 156 m simulation data as input for 26-27 April 2013.

raindrops fall faster than lightweight cloud droplets. A negative fall velocity is defined as downward motion within this study. The spectral width of a cloud radar (Fig. 4.5e,f) measures the turbulence intensity, depicting the differences of velocities of the various atmospheric particles inside the observed volume. The spectral width is higher for small cloud droplets moving with the surrounding turbulence than for larger rain droplets that fall because of their weight (Rinehart, 2010).

The geometric and mass properties of the frozen hydrometeors seem to be well captured by the ICON HOPE, which is visible by a good agreement of the Doppler velocity distribution with the measurements above 4 km (Fig. 4.5c,d). Cloud ice hydrometeors are mainly present above 5 km with fall velocities of less than $0.5 \, \text{m s}^{-1}$. The ICON HOPE overestimates the local turbulence within clouds above 4 km, as seen by an overestimated spectral width of about $0.3 \, \text{m s}^{-1}$ (Fig. 4.5e,f), indicating an overestimated TKE of the ICON HOPE. The reason cannot be wrong hydrometeor properties because otherwise the Doppler velocities would also not be correct.

Furthermore, rain events are often too intense and short in the ICON HOPE, as seen for instance by the two very intense rain showers between 06-09 UTC of the first day with an overestimated reflectivity of about 10 dBZ (Fig. 4.5a,b). The rain showers are only about 30 minutes long compared to the less intense and longer rain shower of the measurements between 09 UTC and 11 UTC. Similarly, the rain intensity of the main cold frontal passage after 18 UTC is overestimated, illustrated by roughly 5 dBZ higher reflectivities than the measurements. Only 0.7 mm of rain was observed between 18 UTC and 22 UTC at the JOYCE supersite, whereas more than a factor of two higher accumulated rain sum of 1.8 mm is seen for the ICON HOPE 156 m simulations within the same timeframe (not shown). The overestimated rain intensities cause also a shifted peak of the reflectivities at the frequency of occurrence between surface and 2.5 km to higher values by about 5 dBZ (Fig. 4.6a,b). These findings are in accordance with results of the precipitation evaluation in Chapter 3, as well as with the findings of Heinze et al. (2017a) in stating also overestimated heavy rain intensities. The too-heavy rain intensities also lead to Doppler velocities of more than $-2 \, \text{m s}^{-1}$, seen by the marginal distribution in Fig. 4.5d, which are almost not present in the measurements (Fig. 4.5c). The quickly falling rain drops are less affected by the local turbulence due to their speed, visible by a very low spectral width of $0.05 \, \text{m s}^{-1}$ which does not exist in the observations. Most probably, the rain events are too short as a consequence of the overestimated precipitation intensity.

Within the lower troposphere, below 3 km, liquid hydrometeors are most likely overestimated by the ICON HOPE, whereas the amount of rain hydrometeors might be underestimated. This is seen by two discrete peaks at the histograms of the Doppler velocity (Fig. 4.6d) and spectral width (Fig. 4.6f) below 3 km compared to the continuous distribu-
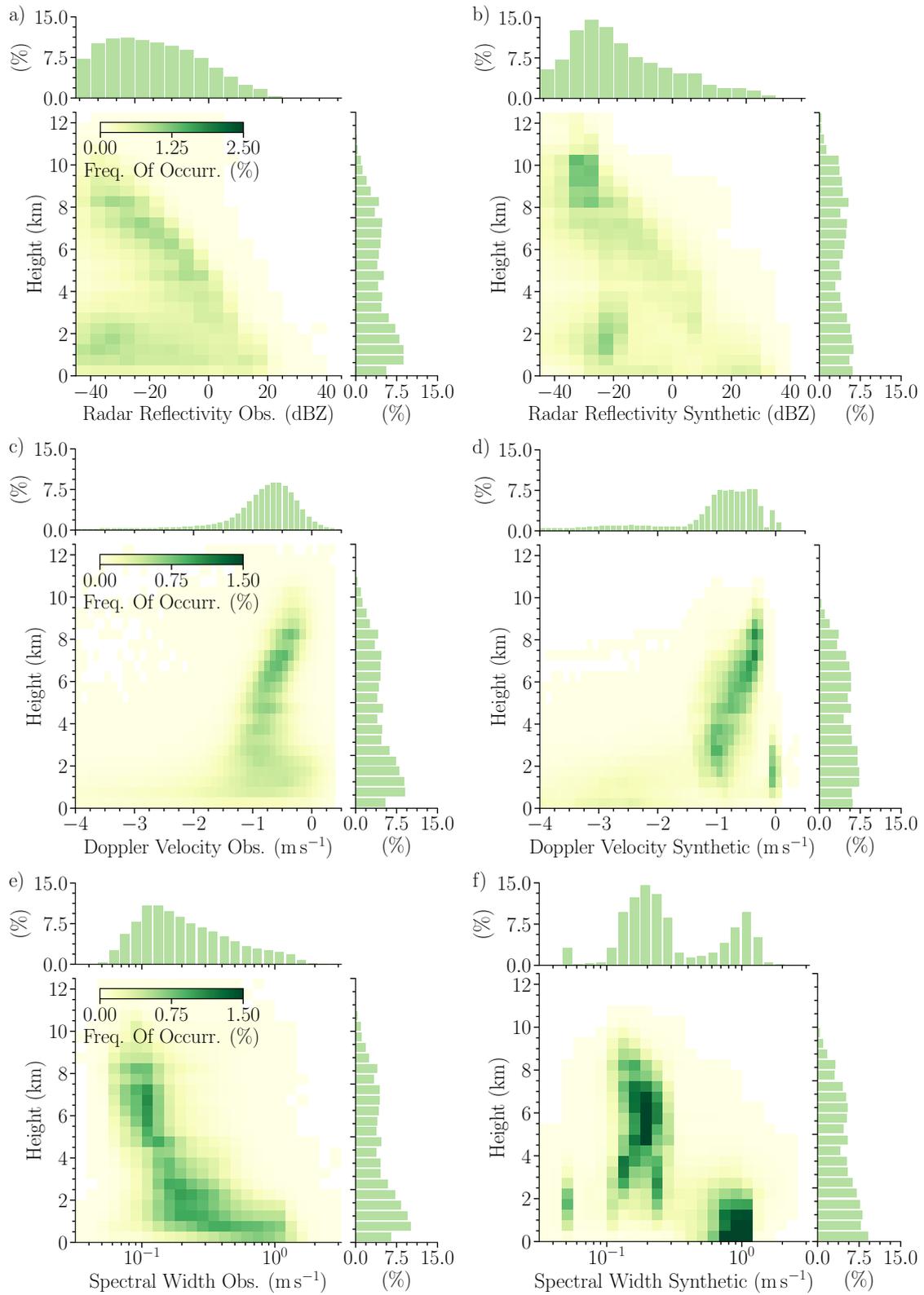
Figure 4.6.: Two-dimensional histograms of cloud radar reflectivity (a,b), Doppler velocity (c,d) and spectral width (e,f) as function of height for LACROS observations (a,c,e) and simulated cloud radar by the ICON HOPE 156 m PAMTRA simulation (b,d,f) using data of April and May 2013. The frequency shown by the colour is normalised for each dataset. Marginal distributions show integrated frequency of occurrence for each interval, normalised by the each dataset.

tions of the measurements (Fig. 4.6c,e). Uncertainties of the simulated TKE cannot explain the two peaks because otherwise the Doppler velocity would be correctly simulated and only the spectral width would be influenced. Small liquid cloud droplets have almost no fall velocity and move mainly with the surrounding turbulence due to their small size and weight. The overestimation is seen by a substantial peak around -0.25 m s^{-1} to 0.00 m s^{-1} in the histogram of the Doppler velocity and by the peak of a large spectral width of about 1 m s^{-1} below 3 km. Rain droplets have a larger fall velocity and thus are less affected by the turbulence leading to a smaller spectral width in the cloud radar measurements. A too-infrequent occurrence of Doppler velocities of more than -1 m s^{-1} as well as of small spectral width values around 0.1 m s^{-1}, representing rain, illustrate the underestimation of rain hydrometeors by the ICON HOPE compared to the observations. The dominating liquid hydrometeors also cause the bi-modal marginal distribution of the Doppler velocity and spectral width, with an extra peak at low spectral width values for heavy rain. Contrary to this, the observed Doppler velocities are continuously Gaussian distributed and the spectral width values similar to a Weibull distribution. Possibly a too-early melting of the before mentioned overestimated frozen hydrometeors could be the reason for the overestimation of liquid hydrometeors by the ICON HOPE simulations.

Updrafts at cloud edges and within deep convection are substantially underestimated by the ICON HOPE simulations. For example, this is visible in the forward-simulated time-series of the Doppler velocity (Fig. 4.5c,d) of the first day around 12 UTC at the cloud edges and near the top of the deep convection compared to the observations. The underestimated updrafts are also confirmed by the histogram of the ICON HOPE with values of less than +0.125 m s^{-1} and a frequency of occurrence of less than 1% (Fig. 4.6c,d). In contrast, more frequent updrafts of up to +0.5 m s^{-1} are seen for the observations. Even the 156 m horizontal resolution with 150 vertical model levels of the ICON HOPE simulations are presumably too coarse to accurately resolve the small-scale updrafts within deep convection, as well as at the cloud edges, explaining the missing upward motions.

The linear depolarisation ratio (LDR), defined as the quotient of the reflectivity of vertically polarised sent and horizontally received signals divided by the reflectivity of sent and received horizontally polarised signals, provides information about the geometry of hydrometeors. For perfectly spherical hydrometeors, the LDR is minus infinite (Rinehart, 2010), which is measured as -40 dB by real cloud radars due to their technical limitations. Non-spherical hydrometeors like ice crystals or melting ice have larger linear depolarisation ratios due to their shape and reflectivity properties. The LDR can be used for example to detect the melting layer as it is also done by the Cloudnet algorithms (Illingworth et al., 2007). A technical error of the used PAMTRA version (git #ca96ff6) during the time of this study neglects the different polarisations, so the LDR is always minus infinite, which is

set to -40 dB as for a real cloud radar. The missing LDR information has to be considered later at the computed MOC-based Cloudnet target classification.

The forward-simulated cloud radar provides a comprehensive view on the cloud's representation of the ICON HOPE simulations, which are further investigated in the following sections using the two additional instruments of the microwave radiometer and the lidar.

4.3.2. Passive Microwave Radiometer Simulation

The 14 channels of the commonly at Cloudnet supersites-installed RPG HATPRO passive microwave radiometer are simulated by PAMTRA for the virtual Cloudnet supersite to evaluate the tropospheric water vapour and temperature distribution of the ICON HOPE 156 m simulations. The lower seven channels between 22.24 GHz and 31.40 GHz of the water vapour window provide information about the tropospheric humidity at the different height levels. The centre of the H_2O absorption line is at 22.235 GHz (Crewell et al., 2010). The signal of channels close to the line, such as the 22.24 GHz, are affected more by the pressure broadening and so the information is dominated by layers of high altitude (Crewell et al., 2010). Channels further away from the absorption line like 31.40 GHz are less affected by the pressure broadening and can thus provide information about the humidity at lower levels of a ground-based instrument. The V-band channels between 51.26 GHz and 58.00 GHz of the oxygen line with its centre at 60 GHz (Crewell et al., 2010) can be used to determine the temperature at different heights. The signals of channels like 58.00 GHz, which is close to the absorption line, are attenuated very quickly and provide information about the temperature near the instrument. The microwave radiometer itself measures only the brightness temperatures at the different frequencies, which can be used typically in combination with statistical-based retrievals to calculate for instance a temperature- and humidity profile, liquid water path (LWP), and integrated water vapour (IWV). However, the retrievals induce uncertainties because of the assumed average climatology of the parameters at the calculation of the values, as well as by having only few measurements available (Sect. 4.2.2). The Cloudnet algorithm requires only the LWP of the microwave radiometer, which is calculated in this study based on a well-established retrieval using DeBilt radiosondes, minimising the induced inaccuracies.

The microwave radiometer measurements of the LACROS supersite are used as reference for the comparison with the PAMTRA forward simulated microwave channels using the ICON HOPE 156 m run as input. All values during rain events of the observations, where the microwave radiometer window might be wet and unusable, are filtered. The existing quality flags of the instruments are also incorporated to consider technical issues of the instrument. The two datasets are homogenised, and missing values neglected consistently.

The two-month time-series of two microwave radiometer channels, one for the water vapour (23.84 GHz) and one for the temperature (58.00 GHz), are illustrated in Figure 4.7 to give a first impression of the results to be expected. The near-ground temperature and diurnal cycle are well represented by the ICON HOPE simulations, seen by a good match of the example two-month time-series of the forward simulated 58.00 GHz brightness temperatures with the observations (Fig. 4.7a). As explained above, the 58.00 GHz channel is attenuated very quickly and provides information about the temperature close to the instrument. An extremely small bias of only -0.02 K and low RMSE of 0.77 K (Tab. 4.1) for the 58.00 GHz channel confirms the good representation of the near-ground temperature. This is also in accordance with the low errors found for the ICON GER simulations in the baseline evaluation of the 2 m temperature (Sect. 3.5). Similarly, the temperature of the lower troposphere is well captured by the ICON HOPE 156 m, visible by bias values of less than -0.12 K and RMSE values of less than 0.77 K for the microwave channels between 58.00 GHz and 54.94 GHz. These channels are dominated by the temperature distribution of the lower troposphere. All RMSE values are substantially lower than the natural variability of these channels, depicted by the standard deviation of about 5 K. These findings are again consistent with the baseline evaluation (Sect. 3.5). Contrary to the good representation of the full tropospheric temperature profile by the ICON GER simulations at the analysis of the soundings (Sect. 3.5), the forward-simulated microwave radiometer exhibits large temperature uncertainties at high altitudes. This can be seen by large RMSE values of 1.51 K for the 53.86 GHz channel and of up to 7.07 K for the 51.26 GHz channel (Tab. 4.1), which are mainly influenced by higher altitudes. The stated uncertainties of the hydrometeor distributions in the cloud radar analysis affect the radiative transfer calculations and might explain therefore the large errors for these microwave radiometer channels, as well as the contradictory results compared to the soundings.

The exemplary forward-simulated two-month time-series of the 23.84 GHz water vapour channel of the ICON HOPE exhibits substantial deviations of the middle tropospheric humidity, as seen at 10 May 2013, indicating large uncertainties of the ICON HOPE 156 m (Fig. 4.7b). The 23.84 GHz channel is more influenced by layers away from the instrument due to the pressure broadening. Accordingly, the 23.84 GHz channel is dominated by the signals of the middle tropospheric humidity. On average, the ICON HOPE 156 m is too dry, as seen by a negative bias for all water vapour channels, which is in agreement with the baseline evaluation of the ICON GER simulations in Section 3.6. The upper tropospheric humidity is better represented than the near-ground humidity by the ICON HOPE due to the overall lower humidity at high altitudes. This is visible by an increasing bias of -0.16 K and RMSE of 4.09 K for the upper troposphere 22.24 GHz channel up to a bias of -1.57 K and RMSE of 5.55 K for the near-ground channel of 31.40 GHz. The RMSE of the higher frequencies, like for the 31.40 GHz with 5.55 K, is already slightly higher than that
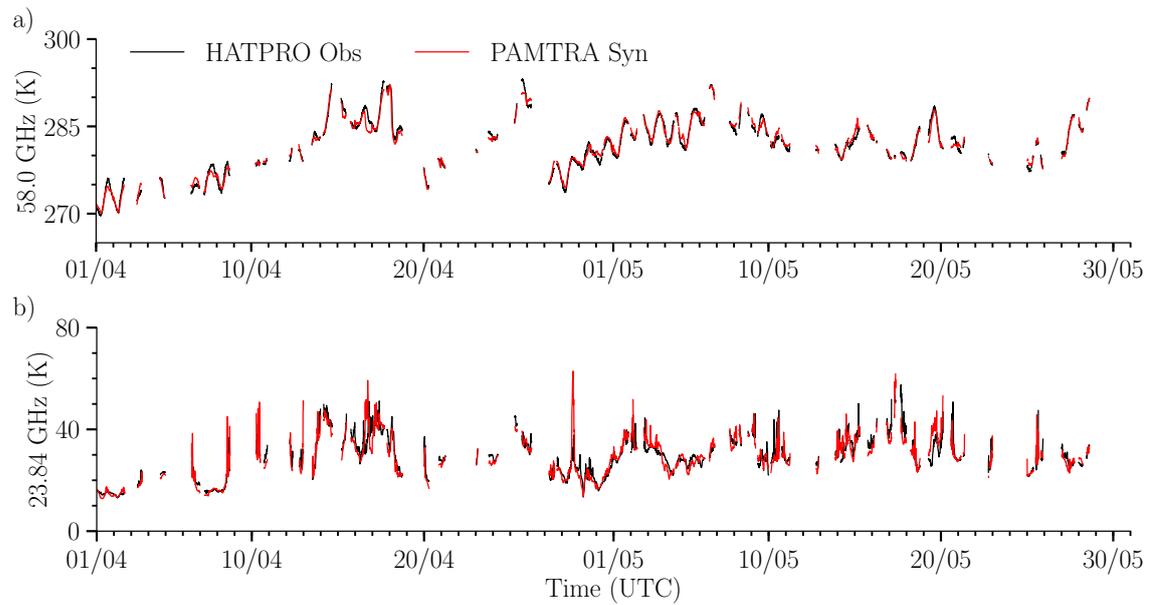
Figure 4.7.: Brightness temperature time series for HATPRO microwave radiometer observations from LACROS HOPE supersite and corresponding PAMTRA simulation using the ICON HOPE 156 m simulation data for April and May 2013 for 58.00 GHz (a) and 23.84 GHz (b) channels.

of the natural variability illustrated by the observed standard deviation of 5.41 K. Also, the RMSE of the 22.24 GHz with 4.09 K is already about half of the observed natural variability of 9.53 K. The large errors of the forward simulated water vapour channels suggest large uncertainties of the humidity representation by the ICON HOPE. However, this is again in contrast to the well-represented near-ground and full tropospheric humidity found at the baseline evaluation of the ICON GER simulations using in situ measurements of the weather stations and soundings. Most probably, problems of the matching point-to-point comparison of for instance shifted or delayed clouds explain the large errors seen by the moderate-to-low bias values and considerably large RMSE values. Nevertheless, also the issues of the hydrometeors, especially the overestimation of liquid hydrometeors, affect the radiative transfer calculation and could also be the reason for the large errors found.

The liquid water path (LWP) is the column-integrated liquid water above the station, which is required by the Cloudnet algorithms for the classification of liquid clouds and calculation of the liquid water content. The LWP is calculated consistently based on the water vapour channels for the simulated and real measurements according to Karstens et al. (1994). The applied LWP retrieval, provided by the University of Cologne, is created based on the radiosonde climatology of DeBilt (Netherlands). The DeBilt soundings represent the western Europe climate, which is similar to the considered location of Jülich. The cloud analysis algorithm uses a 95.0% threshold with an adiabatic profile. All seven water vapour channels at zenith elevation of the microwave radiometer are included in the

Table 4.1.: Average standard deviation of observed (STD_OBS) oxygen (V-band, upper table) and water vapour (K-band, lower table) microwave channels using April and May 2013 data of the LACROS observations. Averaged bias and RMSE errors are calculated between the forward-simulated PAMTRA output of the corresponding ICON HOPE 156 m column output and the LACROS observations for the different microwave channels.

V-Band (GHz)	51.26	52.28	53.86	54.94	56.66	57.30	58.00
STD_OBS (K)	6.68	5.76	4.50	4.48	4.69	4.73	4.72
Bias (K)	-0.66	1.28	-0.74	0.00	-0.12	-0.11	-0.02
RMSE (K)	7.07	5.62	1.51	0.49	0.72	0.76	0.77
K-Band (GHz)	**22.24**	**23.04**	**23.84**	**25.44**	**26.24**	**27.84**	**31.40**
STD_OBS (K)	9.53	9.21	7.94	5.93	5.41	4.97	5.41
Bias (K)	-0.16	-0.13	-0.29	-0.19	-0.64	-0.75	-1.57
RMSE (K)	4.09	4.12	4.01	4.04	4.17	4.50	5.55

LWP calculation. The forward-simulated LWP is evaluated by the LACROS reference observations. Furthermore, the LWP of the direct model output is compared to the forward simulated and observed LWP to examine differences regarding the determination of the LWP in the ICON HOPE 156 m simulations.

The stated overestimation of liquid hydrometeors by the ICON HOPE 156 m simulations in the cloud radar analysis is confirmed by a too-high LWP. This is seen by substantially larger LWP values of the selected one-day time-series (Fig. 4.8), and by a positive bias of $1.7 \, \mathrm{g \, m^{-2}}$ for the forward simulation and of $2.6 \, \mathrm{g \, m^{-2}}$ for the direct model output of the ICON HOPE 156 m. The large uncertainties of the forward-simulated water vapour channels (Tab. 4.2) used for the LWP calculation are again affected by the point-to-point comparison and found issues of the hydrometeors. This causes also the remarkably large RMSE of $118.8 \, \mathrm{g \, m^{-2}}$ for the forward-simulated LWP. The large errors of the liquid hydrometeors in the ICON HOPE 156 m are approved by a similarly large RMSE of $115.8 \, \mathrm{g \, m^{-2}}$ using the direct model output of the ICON HOPE 156 m simulations. The similar bias and similar RMSE of the forward-simulated values and of the direct model output clearly show the robustness of these findings and independence of the method to derive the LWP.

The large deviations of the simulated LWP compared to the observations are also recognisable by the larger RMSE of about $116\text{-}118 \, \mathrm{g \, m^{-2}}$ than that of the natural variability, depicted by the observed standard deviation of $93.7 \, \mathrm{g \, m^{-2}}$ (Tab. 4.2). The cloud base height, necessary for the Cloudnet algorithm, is examined in the following section using a forward-simulated micropulsed lidar.
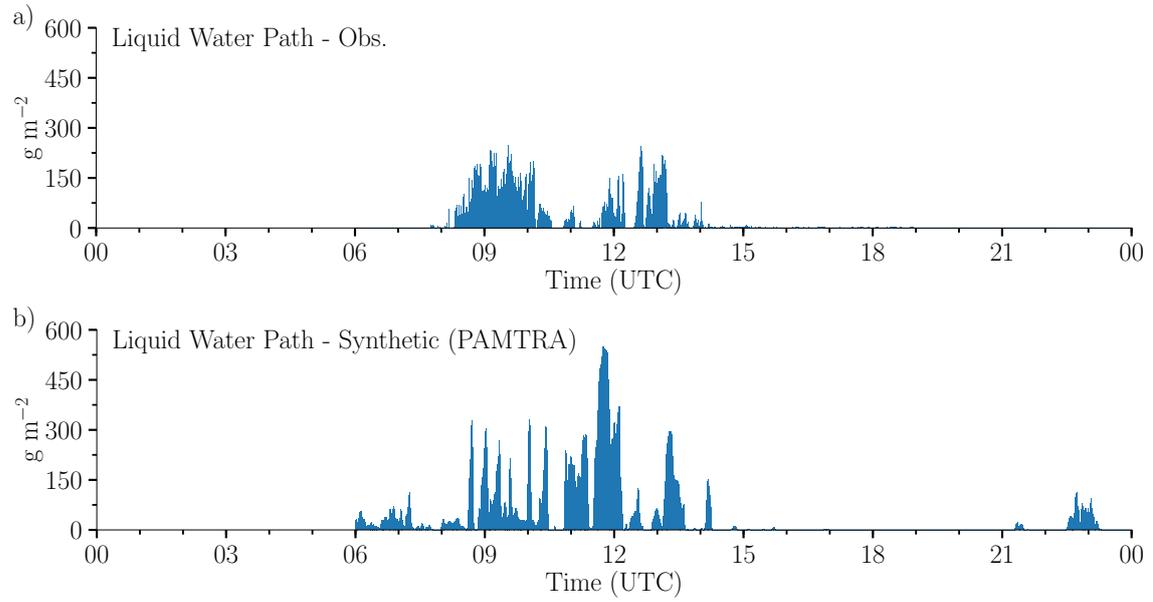
Figure 4.8.: Retrieved integrated liquid water path (LWP) of the microwave radiometer of the LACROS supersite (a) and PAMTRA simulation (b) using the ICON HOPE 156 m model data as input for 26-27 April 2013.

Table 4.2.: Average standard deviation (STD) of liquid water path (LWP) at the LACROS HOPE supersite using the LACROS HATPRO observations, the PAMTRA forward-simulated Microwave Radiometer (MWR) simulations and direct model output of the ICON HOPE 156 m simulations for April and May 2013. Averaged bias and RMSE errors are calculated between model and observations.

LWP (g m^{-2})	Obser-vations	Forward-simulated MWR (PAMTRA)	ICON HOPE 156 m Direct Model Output
STD	93.7	109.3	103.6
Bias		2.6	1.7
RMSE		118.8	115.8
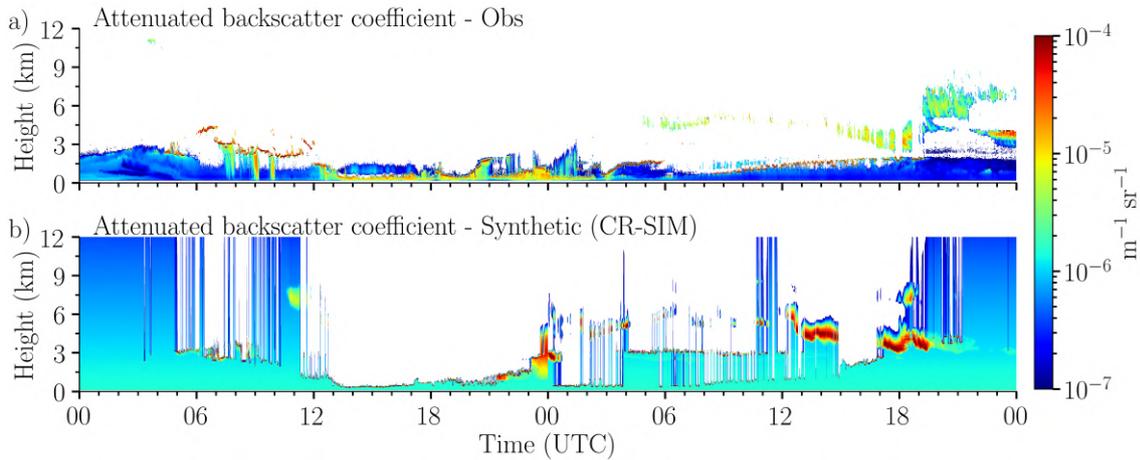
Figure 4.9.: Attenuated backscatter coefficient at 1064 nm of the Jenoptik CHM15kx LACROS lidar (a) and of the corresponding 532 nm CR-SIM simulated micropulsed lidar (b) using the ICON HOPE 156 m model data as input for 26-27 April 2013.

4.3.3. Lidar Simulation

The attenuated backscatter coefficient of a lidar, respectively ceilometer, is used by Cloudnet to detect the cloud base height of liquid clouds. Those clouds are often not accurately captured by the cloud radar measurements because of their small radar reflectivity, which is the reason to use a lidar for this. A 532 nm micropulsed lidar is simulated by CR-SIM based on the ICON HOPE 156 m simulation output (Sect. 2.2) and compared to the available LACROS Jenoptik CHM15kx reference observations at 1064 nm. The shorter wavelength of 532 nm of the forward-simulated lidar is more sensitive to smaller particles and the attenuated backscatter coefficient signal is usually larger than that of the observed 1064 nm lidar due to the higher molecular scattering (Weitkamp, 2005). Nevertheless, the derived cloud base heights relevant for Cloudnet are very similar. CR-SIM can only simulate a micropulsed lidar with a wavelength of 353 nm and 532 nm (Tatarevic et al., 2019), which is the reason for the different wavelengths used in this study. The cloud base height is calculated consistent with Cloudnet by the first height above ground where the attenuated backscatter coefficient is larger than the threshold of 10^{-6} m^{-1} sr^{-1} (Illingworth et al., 2007). Additionally, the direct model output of the cloud base height from the ICON HOPE 156 m simulations is included in this analysis to investigate the influence of the cloud base height determination method on the results.

As expected, overall larger attenuated backscatter coefficients of about $0.5*10^{-5}$ m^{-1} sr^{-1} are identified for the shorter wavelength of 532 nm of the forward-simulated CR-SIM lidar than the observations at 1064 nm (Fig. 4.9). The CR-SIM lidar simulates a homogeneous attenuated backscatter coefficient value for cloud-free regions decreasing with height because of the missing information about aerosols and insects, which are usually not provided
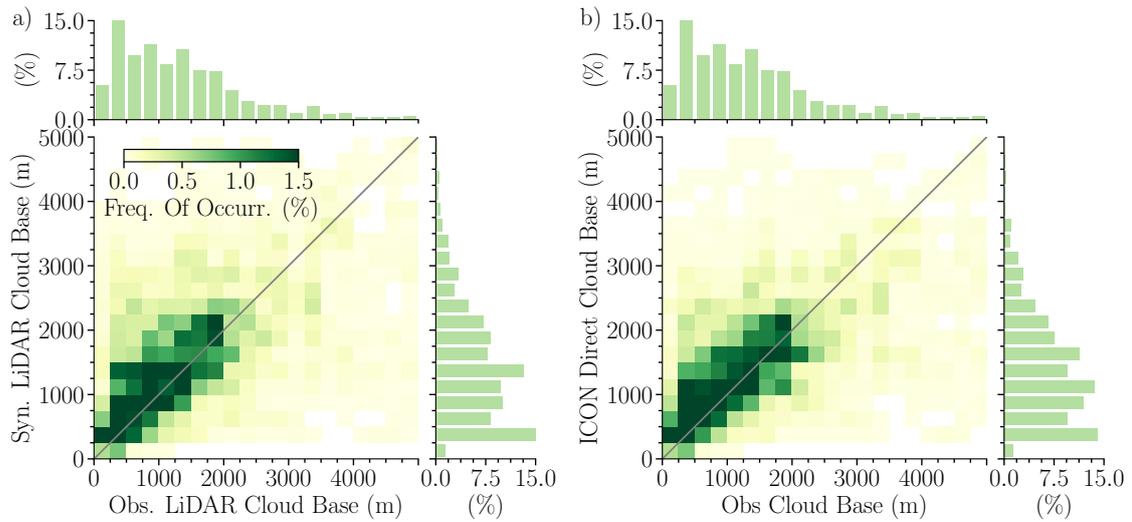
Figure 4.10.: Two-dimensional histograms of the cloud base height derived from the observed Cloudnet lidar versus the synthetic forward simulated lidar (a) and direct model output (b) based on the ICON HOPE 156 m simulation using all data of April and May 2013. The colour of the 2D histogram shows the frequency of occurrence and is normalised by the dataset. Likewise, the marginal distributions are normalised.

by atmospheric models like the ICON LES. In contrast, the aerosols and insects are clearly seen in the observed lidar time-series within the boundary layer as regions with larger and smaller attenuated backscatter coefficients, as, for example, between 00 UTC and 06 UTC of the first day. In addition, all rain hydrometeors are neglected by CR-SIM, and the background signal is used for these regions, as is visible during the rain period between 12 UTC and 00 UTC of the first day. However, the rain is clearly measured by larger attenuated backscatter coefficient values identifiable during the same period for the observations. The rain hydrometeors are not considered by the forward-simulated lidar of CR-SIM because the application focus of this tool is on clouds rather than on rain events (Mariko Oue, personal communication). Nevertheless, the cloud base heights of the ICON HOPE 156 m simulations are well detected by the CR-SIM forward simulation, as seen by distinct regions of larger attenuated backscatter coefficient values, similar to the observations. Also, high cirrus clouds such as those seen around 11 UTC of the first day or multi-layer cloud bases during the second day are captured by the forward-simulated lidar. As expected, distinct deviations between the forward-simulated cloud base height of the ICON HOPE simulations and the observations can be identified because of the matching point-to-point comparison. For example, around 12 UTC of the first day, clouds are present in the observations, whereas ICON HOPE 156 m does not simulate a cloud for this exact grid cell and time.

The overall distribution of different cloud base heights is well captured by the ICON HOPE 156 m simulations, seen by similar marginal distributions of the forward simulated cloud

Table 4.3.: Time averaged standard deviation (STD) for the cloud base height determined from the observations, the forward simulated lidar of ICON HOPE 156 m, and the direct ICON HOPE 156 m model output for April and May 2013. The averaged bias and RMSE are computed between the simulated datasets and the observations.

Cloud Base Height (m)	Observations	Forward-simulated Synthetic Lidar	ICON HOPE 156 m Direct Model Output
STD	1031	961	918
Bias		180	93
RMSE		901	824

base heights and of the observations (Fig. 4.10a). This is consistent with the marginal distribution of the direct model output of the ICON HOPE simulations (Fig. 4.10b). On average, a small overestimation of the cloud base height is found for the ICON HOPE 156 m simulations, as seen in the two-dimensional histograms in Fig. 4.10a,b and by a positive bias of 180 m for the forward simulated lidar and of 93 m for the direct model output (Tab. 4.3).

However, the problems of the matching point-to-point comparison of the cloud base height between the ICON HOPE 156 m simulations and the observations, as well as the high variability of cloud base heights, depicted by an observed natural variability of 1031 m, leads also to the large spread seen in the two-dimensional histograms (Fig. 4.10a,b). This is consistent with the derived cloud base heights from the forward-simulated lidar and the direct model output of the ICON HOPE 156 m. For that reason, a remarkably large RMSE of 901 m for the forward-simulated lidar and of 824 m for the direct model output is found, which is of a similar magnitude as of the observed natural variability. However, the variability of cloud base heights is well represented by the ICON HOPE, depicted by similar standard deviation values of 961 m for the forward simulation and 918 m for the direct model output compared to the observed 1031 m. The similar histograms and error statistics between the forward simulated cloud base heights and the ones of the direct model output increases again the robustness of the results and proves the independence of the results from the derivation method of the cloud base height.

To sum up, the ICON HOPE simulations overestimate frozen hydrometeors above 4 km and the intensity of rain showers, as seen by an overestimation of the forward simulated cloud radar reflectivity by up to 10 dBZ. Updrafts at deep-convection and cloud edges are almost missed by the ICON HOPE, depicted by only few simulated positive Doppler velocity values. Liquid hydrometeors are overestimated and rain hydrometeors underestimated, especially between 500 m and 3 km height, found by a bimodal distribution of Doppler velocities and spectral width. This was confirmed by a positive bias of the LWP. The matching point-to-point evaluation causes large uncertainties of the forward-simulated

brightness temperatures, LWP and of the cloud base height with a RMSE of about 900 m, seen in the forward-simulated lidar data, as well as in the ICON HOPE 156 m direct model output.

4.4. Cloud Classification Analysis

The Cloudnet target classification is a multi-sensor product providing detailed information about cloud macrophysics, phase, and structure, which is a good starting point for an in-depth cloud evaluation. The observed Cloudnet products (Sect. 4.1) of the LACROS HOPE supersite are used as reference data. The model-to-classification (MC) based cloud classification (Sect. 4.2.1), inspired by Cloudnet, and the model to obs to classification (MOC) approach (Sect. 4.2.2) using forward operators to generate a virtual Cloudnet supersite are applied to the two-month ICON HOPE 156 m run for the time of the HOPE campaign (Sect. 2.4). All aerosol and insect targets of the observations are set to "clear sky" to be consistent with the ICON HOPE, which does not provide information about those. If the model or the observations contain missing data due to technical issues or other, all other datasets are homogenised and set to missing data as well. The 491 observed vertical levels of Cloudnet are averaged to the 150 ICON LES levels by using the most frequent category of each interval. Likewise, the nine-second ICON HOPE 156 m output is averaged by the most frequent category to the common 30 seconds of the Cloudnet products to create consistent datasets for a meaningful and fair comparison of the cloud classifications. In this section, the cloud's representation of the ICON HOPE 156 m simulations is evaluated for different aspects like cloud structure, mean cloud fraction profiles considering their phase, cloud geometry, and much more using the new synthetic cloud classifications as the basis.

The overall structure of the exemplary chosen frontal-passage, with small clouds in the beginning and rising deep convection with long-lasting rain, is well captured by the ICON HOPE 156 m simulations (Fig. 4.11). This is visible by a good agreement of the MC-based and MOC-based cloud classifications with the observed Cloudnet classification, especially considering the matching point-to-point comparison of the deterministic ICON HOPE 156 m run. However, the missing linear depolarisation ratio and therefore missing information about the melting layer of the forward simulated cloud radar, as well as the uncertainties of the forward simulated lidar, such as neglecting all rain hydrometeors, lead to the issues that only "Ice" and "Drizzle or rain" are detected by the MOC-based classification (Fig. 4.11c). Consequently, all mixed-phase categories, as well as "Melting ice" and "Cloud droplets only", are not captured by the MOC-approach. The inaccuracies of the forward-simulated lidar lead to such unrealistic layers as "Drizzle or rain" with no clouds above, visible between 00 UTC and 12 UTC of the second day. These issues clearly
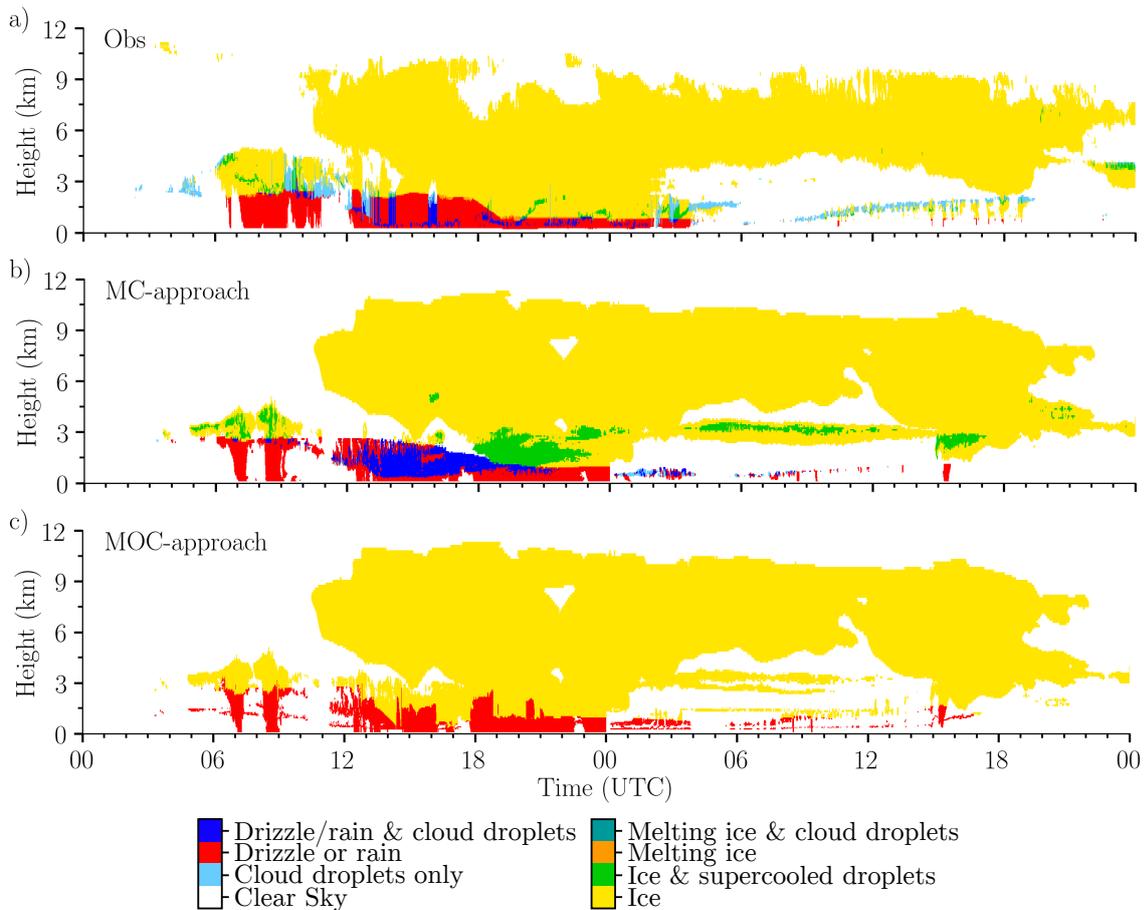
Figure 4.11.: Example Cloudnet target classification (a), MC-based cloud classification (b) and forward-simulated MOC-based cloud classifcation (c) for 26-27 April 2013 using the ICON HOPE 156 m simulation output.

point out the challenges of the MOC approach, requiring accurate and physically consistent forward simulations of all necessary instruments including higher moments of a cloud radar such as the linear depolarisation ratio. In consequence, detailed model output such as of the TKE, vertical wind, and the cloud microphysics is necessary as input for the forward simulations. However, the MOC approach provides a comprehensive view of the simulated clouds, and would enable a physically consistent cloud classification, but further research on the forward simulations is necessary to accomplish an accurate MOC cloud classification.

In contrast, the MC-based cloud classification inspired by Cloudnet can be applied very quickly to most atmospheric models like the ICON LES because only information about the profiles of temperature, dew point, and the different hydrometeor masses are necessary. The MC-based cloud classification of the ICON HOPE 156 m simulations shows very

promising results containing all categories of the observations including mixed phases (Fig. 4.11b). This synthetic MC-based cloud classification is even difficult to differentiate from real observations because of the similar spatial and temporal resolved model output as of the measurements. The selected two-day time-series of the ICON HOPE MC-based classification confirms the findings of the forward simulation analysis and illustrates the found issues comprehensively.

The overestimation of frozen hydrometeors above roughly 4 km by the ICON HOPE is again visible by a larger ice cloud fraction and higher ice cloud top heights of the MC-based cloud classification than seen in the observations. This is also in accordance with the MOC-based classification and the findings of Heinze et al. (2017a). Also, the too-high amount of liquid hydrometeors and underestimation of rain hydrometeors by the ICON HOPE simulations between 500 m and 3 km is proven by a substantially larger fraction of the mixed-phase categories of "Drizzle/rain & cloud droplets" and of "Ice & supercooled droplets" than for the observations (Fig. 4.11b). For example, the overestimation is seen between 12 UTC and 00 UTC of the first day, where the MC-based classification of the ICON HOPE 156 m contains mostly mixed-phase categories with cloud droplets, whereas only pure "Ice" or "Drizzle or rain" is present in the observations. This example illustrates the added value of the cloud classification by making such issues clearly apparent. Contrary to this, the found overestimation of the rain intensity by the ICON HOPE in the analysis of the forward-simulated cloud radar is not recognisable any longer in the cloud classification due to reduction to only discrete categories. The MC-based cloud classification can only prove the too-short rain showers such as between 06-12 UTC, as is consistent with the classification of the MOC approach. Therefore, these two examples show distinctively the synergy of the combined analysis of the single forward-simulated instruments with the cloud classification for a comprehensive cloud evaluation.

Mean Cloud Cover Profile

In addition to the previous qualitative analysis of the cloud classification, further quantitative parameters like the cloud fraction profile can be derived consistently from the observed and the new synthetic cloud classifications. This comparison avoids uncertainties due to for example empirically tuned model output variables of the cloud fraction and enables a fair and physically consistent model evaluation. The vertical mean cloud fraction profile is calculated on the basis of the observed, the MC-based, and MOC-based cloud classifications considering all categories except for "Drizzle or rain" as a cloud and is averaged over the two-month ICON HOPE 156 m simulation.

The missing liquid categories of the MOC-based cloud classification below roughly 4 km cause the deviations of up to 10 percentage points (p.p.) in the mean cloud fraction profile
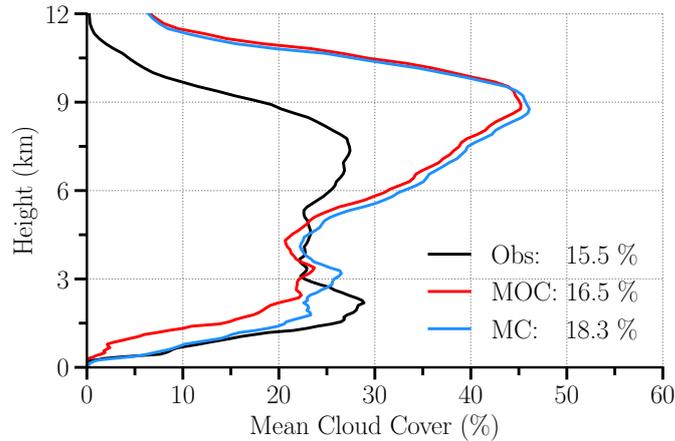
Figure 4.12.: Mean cloud cover profile for the Cloudnet observations (black), the MOC-based cloud classification (red) and the MC-based cloud classification (blue) using the ICON HOPE 156 m simulation output for Jülich for April and May 2013.

in relation to the MC-based cloud classification (Fig. 4.12). Above 4 km, both approaches agree very well, strengthening the robustness of the ICON HOPE simulation results and their independence from the method used to derive the cloud fraction. The mean cloud fraction profile of the ICON HOPE, based on the MC approach, is well represented below 5 km compared to the observations, except for an underestimation of roughly 5 p.p. between 1 km and 3 km height. The already stated too-high amount of frozen hydrometeors above 4 km of the ICON HOPE output leads to a notable overestimation of the mean cloud fraction by up to 25 p.p. above roughly 6 km. For that reason, the vertically averaged cloud cover is also too high by 2.8 p.p. for the MC-based classification and by 1.0 p.p. for the MOC approach.

Mean Cloud Cover Profile by Cloud Type

One big advantage of using the cloud classification compared to others such as single cloud radar measurements for the evaluation of clouds is the possibility to include different cloud macrophysical properties like their phase. The eight different categories, neglecting aerosols and insects, are combined to four cloud types of "Clear sky", "Drizzle or rain", "Liquid clouds", and "Ice clouds" with the focus on the model's hydrometeors. The categories of "Clear sky" and "Drizzle or rain" are not modified. The new cloud type of "Liquid clouds", representing liquid hydrometeors, includes the categories of "cloud droplets only", "Drizzle/rain & cloud droplets", "Ice & supercooled droplets", and "Melting ice & cloud droplets". The category of "Ice & supercooled droplets" is considered as "Liquid clouds" because supercooled droplets are still water droplets in liquid form below the freezing point due to factors such as the absence of a freezing nuclei (Galvin, 2016), and
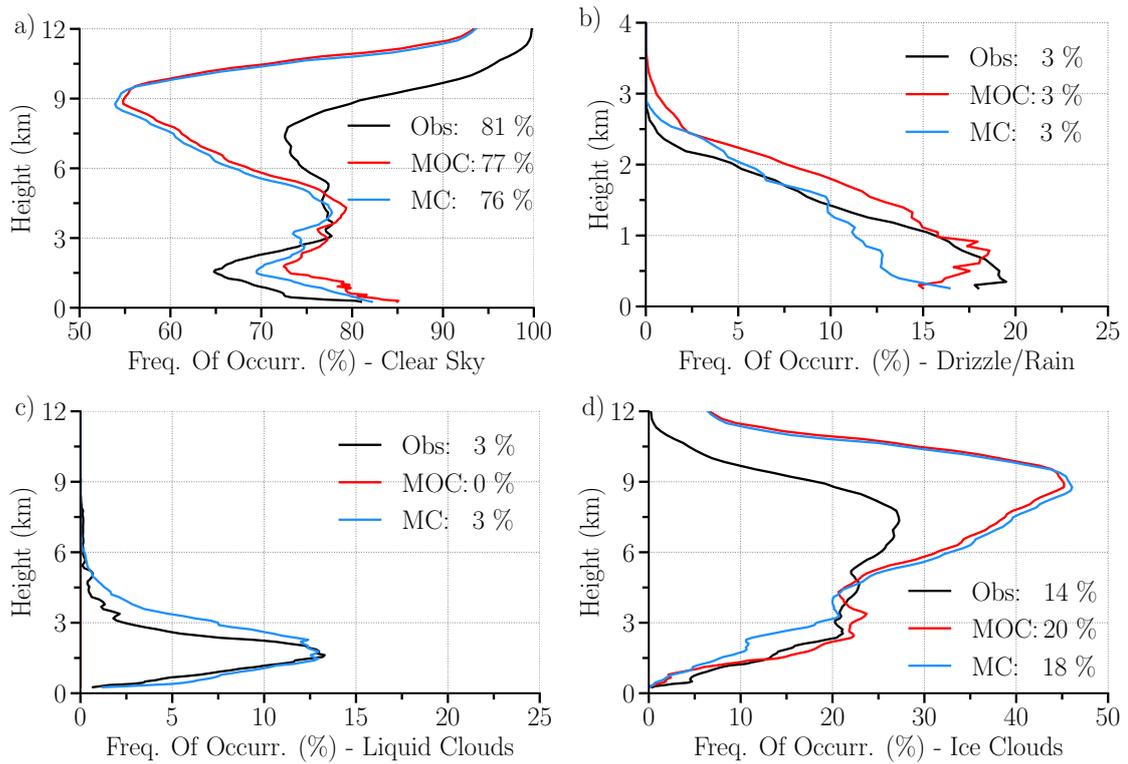
Figure 4.13.: As Figure 4.12, but only for the newly combined "Clear sky" (a), "Drizzle or rain" (b), "Liquid clouds" (c), and "Ice clouds" (d) types.

the category of "Ice clouds" should only contain pure ice. Likewise, "Melting ice & cloud droplets" is regarded as "Liquid clouds" due to the presence of the liquid cloud droplets. Therefore, the new "Ice clouds" type consists only of the pure ice categories of "Ice" and "Melting ice" without any further mixed phases. The merging of the eight categories is conducted because the "Liquid clouds" mainly consists of "Cloud droplets only" and "Ice & supercooled droplets", as well as the "Ice clouds" mainly of the category of "Ice". Additionally, clarity and comprehensibility are increased due to the combination. The time averaged mean cloud fractions are calculated for all new four cloud types using the observed, the MOC-based, and the MC-based cloud classification (Fig. 4.13).

The analysis of the mean cloud fraction profiles considering the newly defined cloud types quantifies the findings of the qualitative comparison of the cloud classification at the beginning and supports the overall results of this section. The previously found too-short and rare rain showers of the ICON HOPE simulations (e.g., Fig. 4.11), result in an underestimation of the frequency of occurrence of "Drizzle or rain" by about 7 p.p. below 3 km for the MC-based classification compared to the observations (Fig. 4.13b). Consequently, the "Clear sky" conditions are overestimated at that height range by roughly

5 p.p. (Fig. 4.13a). The missing detection of "Liquid cloud" categories, often misclassified as "Drizzle or rain" by the MOC approach, causes a too high frequency of occurrence of "Drizzle or rain" by up to 8 p.p. above 1 km in comparison with the observations. Accordingly, the frequency of occurrence of "Liquid clouds" is completely missed by the MOC-classification (Fig. 4.13c).

The stated too-high amount of liquid hydrometeors by the ICON HOPE 156 m simulations is confirmed by the mean cloud profile of "Liquid clouds" of the MC-based classification. This is seen by a slight overestimation of up to 2 p.p. below 1.5 km and a too-high frequency of occurrence by up to 8 p.p. between 1.5 km and 5 km compared to the observations. As a consequence, the frequency of occurrence of "Ice clouds" of the MC-based classification is underestimated by a similar magnitude between 1.5 km and 3 km compared to the measurements (Fig. 4.13d). This underestimation is not visible for the MOC-approach because of the absence of the "Liquid clouds" category, resulting in a too-high frequency of occurrence of "Clear sky" conditions of up to 10 p.p. below 3 km. Similarly, the overestimated "Liquid clouds" of up to 5 p.p. between 3 km and 5 km of the MC based classification explain the underestimation of "Clear sky" conditions by a similar order of magnitude.

The too-high amount of frozen hydrometeors of the ICON HOPE simulations is clearly visible as an overestimated frequency of occurrence of up to 25 p.p. of "Ice clouds" above roughly 5 km for the MC-based classification compared to the observations (Fig. 4.13d). Consequently, the frequency of occurrence of "Clear sky" conditions are too low by a similar magnitude. The MOC approach agrees well with the MC-based approach concerning the overestimation of "Ice clouds" above 5 km and therefore strengthens the robustness of these findings, which are also seen in previous analyses.

In conclusion, the missing linear depolarisation ratio and detection of the melting layer by the forward-simulated cloud radar together with the uncertainties of the forward-simulated lidar have the consequence that currently the MOC approach is only able to detect the categories of "Ice" and "Rain". Consequently, all liquid categories and mixed-phase targets are not detected, which requires further research on the optimisation of the forward simulated instruments. Nevertheless, the physical consistency with the atmospheric model of the MOC approach is the major advantage compared to the MC-based cloud classification. The MC approach applied to the ICON HOPE 156 m simulations captures all different categories of ice, liquid, rain and mixed phases very well. The ICON HOPE cloud classification based on the MC method agrees overall remarkably well with the observations, as seen in the two-day time-series. The analysis of the new synthetic cloud classifications confirms the findings of the forward simulated instruments and quantifies them, such as in relation to the frequency of occurrence of the different categories. The previously at the

cloud radar analysis found too-short and rare rain showers of the ICON HOPE 156 m simulations, are identifiable by a roughly 7 p.p. lower frequency of occurrence of "Rain" at the MC-based cloud classification than the observations. The overestimated frozen hydrometeors by the ICON HOPE 156 m result in an overestimation of up to 25 p.p. of "Ice clouds" above 5 km compared to the observations. The too-high amount of liquid hydrometeors is seen by an overestimation of "Liquid clouds" by up to 10 p.p. below about 5 km.

4.5. Cloud Object Analysis

The synthetic cloud classification product of the ICON HOPE 156 m simulations enables new evaluation techniques, which can be applied to analyse for such parameters as cloud geometric properties in relation to the cloud macrophysics as their phase. The cloud classification facilitates to easily differentiate a cloud from rain, which is still challenging, such as when only cloud radar measurements are available. This illustrates one of the advantages of the new methods for a comprehensive cloud evaluation of the ICON HOPE simulations. Single cloud objects considering their type like "Ice clouds" or "Liquid clouds", based on the definition of previous Section 4.4, are identified by a connected component analysis (Fiorio and Gustedt, 1996; Wu et al., 2005; van der Walt et al., 2014) using the cloud classification data as input.

A new cloud object is generated and numbered consecutively by the connected component analysis, if one of the eight surrounding pixels of a certain selected starting point is of the same type. The object is extended as long as the surrounding pixels fulfil this condition, illustrated schematically by Figure 4.14. Only cloud objects of a certain type like "Ice clouds" or "Liquid clouds" are considered, and all others are neglected.
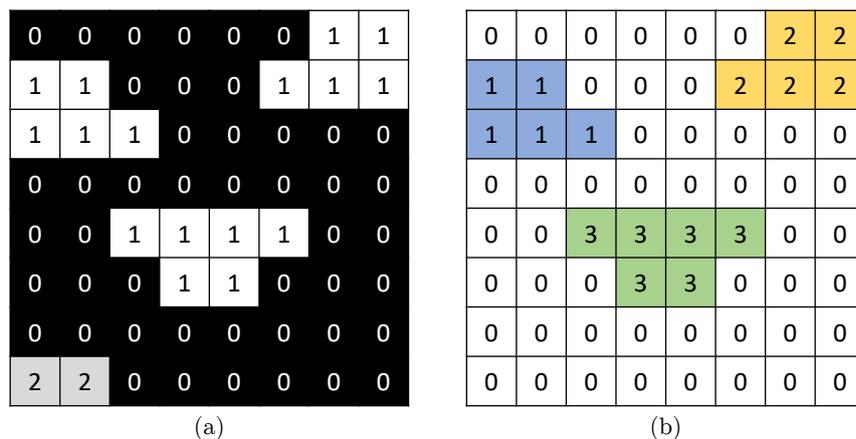
Figure 4.14.: Principle of a connected component analysis converting a cloud mask input (a) with Ice clouds (1) and Liquid clouds (2) to individual selectable Ice cloud objects (b).
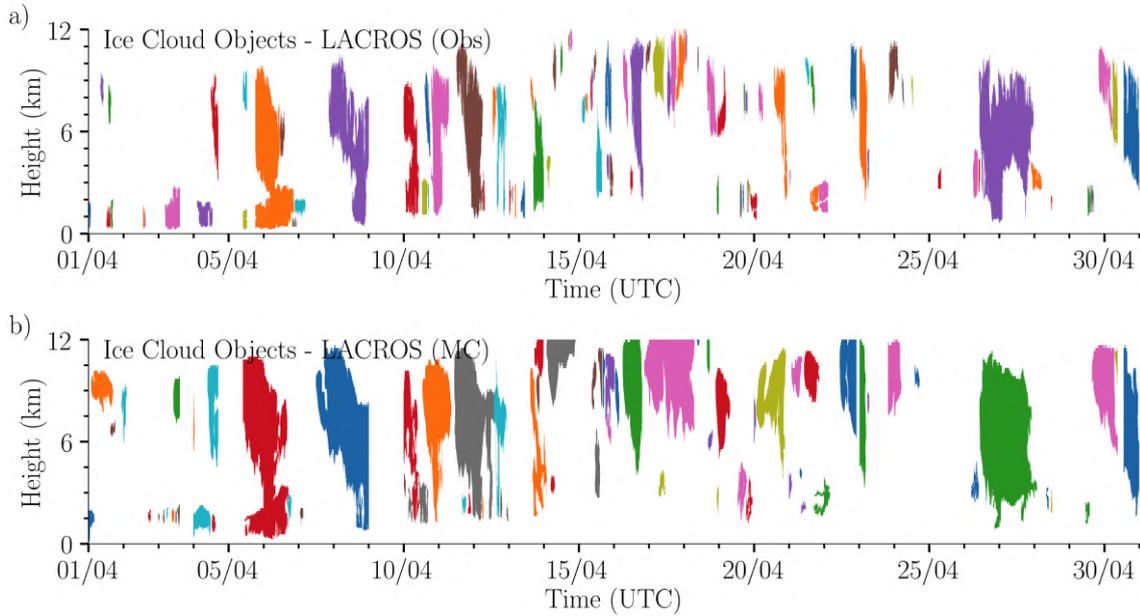
Figure 4.15.: Connected component analysis considering only "Ice clouds" applied to the observed (a) and ICON HOPE 156 m MC-based (b) cloud classification for April 2013. Each recognised cloud object is marked by a different colour.

The single cloud object analysis makes direct use of the newly developed synthetic cloud classification product, which enables the evaluation of parameters like the cloud top, base height, thickness, size distribution, structure, and further geometric cloud properties. The connected component algorithm is applied to the MC-based and observed cloud classification for an in-depth cloud evaluation of the two-month ICON HOPE 156 m simulations. The MOC-based classification is not further considered because of the missing detection of "Liquid clouds" and the observed shortcomings in previous sections. Each pixel of the homogenised cloud classification has a resolution of 30 seconds in time and a vertical resolution of the 150 levels of the ICON HOPE simulations with 20 m layer thickness near-surface. The resulting cloud objects are filtered for objects with a minimum of 100 connected pixels to eliminate unrealistically small objects of only a few pixels and measurement errors. The geometric cloud properties of the two cloud types "Liquid clouds" and "Ice clouds" are computed based on the new cloud object product.

The resulting single "Ice cloud" objects of the connected component analysis are exemplary illustrated in Figure 4.15 for April 2013 to give a first impression on the newly generated cloud objects. The single "Ice cloud" objects of the MC-based cloud classification using the ICON HOPE 156 m simulations and of the observed classification are well detected by the connected component algorithm. This is seen by the different colours of the single objects of the selected one-month time-series (Fig. 4.15). Again, the too-high amount of frozen hydrometeors is identifiable by a larger fraction of "Ice clouds" visible for the

MC-based classification of the ICON HOPE 156 m simulations compared to the observations. The geometric characteristics of for example cloud base height, cloud top, thickness and cloud object length, are analysed for the "Liquid clouds" and "Ice clouds" based on the new single cloud objects (Fig. 4.16). The cloud length is defined as the time in which the cloud passes the instrument with the assumption of a similar wind speed profile of the model and observations, which is reasonable seen by the overall well-represented wind speed profile at the baseline evaluation of the ICON GER simulations (Sect. 3.4).

The cloud base height distribution of "Liquid clouds" of the ICON HOPE 156 m simulations based on the MC-approach classification matches the observations well except for an overestimation by about 5 p.p. of the ICON HOPE for cloud base heights between 2750 m and 3250 m (Fig. 4.16a). The good match of the distributions confirms the findings of the analysis of the overall cloud base heights at the investigation of the forward-simulated lidar and direct model output (Sect. 4.3.3). Contrary to this, the already-stated too-high amount of liquid hydrometeors of previous analyses leads to a shifted distribution of the simulated "Liquid clouds" by the ICON HOPE 156 m to higher cloud top heights compared to the observations. The ICON HOPE 156 m simulations underestimate cloud top heights below 2750 m by up to 5 p.p. and overestimate them above by up to 7.5 p.p. in relation to the measurements. Consequently, the "Liquid clouds" of the ICON HOPE 156 m simulations are too thick, visible by a distinct overestimation of cloud depths of more than 300 m by up to 5 p.p. and underestimation of thin clouds of less than 300 m by up to 12 p.p. (Fig. 4.16e). A similar behaviour is identifiable for the distributions of the cloud lengths. "Liquid clouds" shorter than 30 minutes are too rarely simulated by the ICON HOPE by more than 20 p.p., and clouds of less than ten minutes are completely missed. In contrast, long-lasting "Liquid clouds" of more than 30 minutes are consistently overestimated by the ICON HOPE 156 m simulations by up to 7.5 p.p.. In combination with the overestimation of the cloud thickness, the "Liquid clouds" of the ICON HOPE are too large, which is in accordance with the too-high amount of the liquid hydrometeors, especially between 1.5 km and 5 km, of previous investigations.

The simulated "Ice clouds" of the ICON HOPE 156 m are in general shifted to higher altitudes, seen by shifted distributions of the cloud base height and cloud top height to larger heights in comparison with the observations (Fig. 4.16b,d). Low "Ice clouds" with cloud base heights of less than 500 m with an observed frequency of occurrence of 4% are almost not captured by the ICON HOPE 156 m simulations. The too-high amount of liquid hydrometeors and definition of "Liquid clouds" result in an underestimation of the "Ice cloud" base heights by up to 5 p.p. for heights between 500 m and 2500 m compared to the observations.
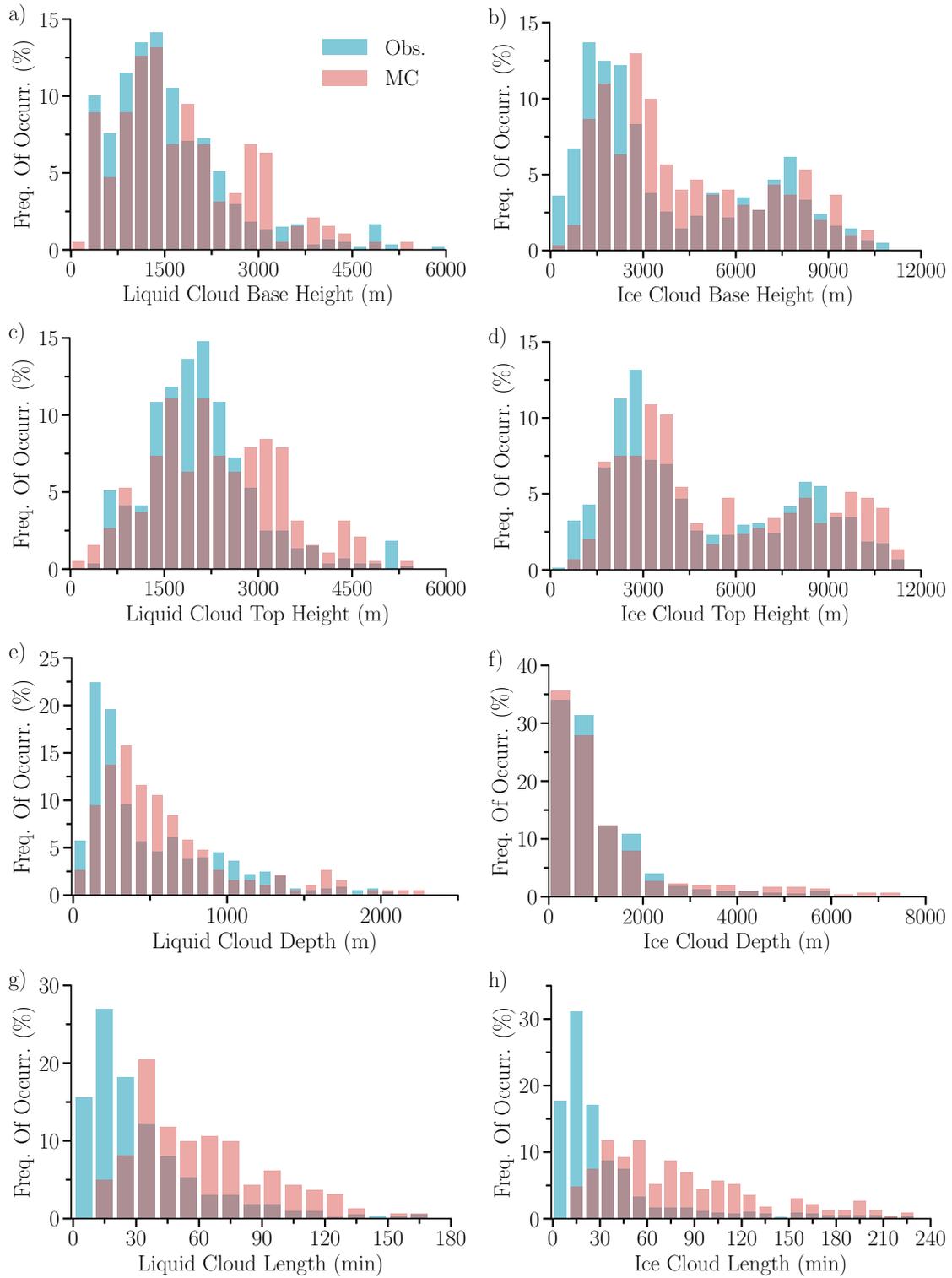
Figure 4.16.: Histograms of connected component analysis for liquid clouds (left column) and ice clouds (right column) using Cloudnet observations (blue) and the MC-based classification (red). The histograms are normalised for each dataset and parameter separately. Mind the different x- and y-axes.

Similarly the cloud top heights of "Liquid clouds" of less than 3 km are too rarely simulated by the ICON HOPE 156 m by a similar magnitude. Notable is the distinct overestimation of "Ice cloud" top heights of more than 9 km by the ICON HOPE 156 m simulations by about 2–3 p.p., which is in agreement as for example with the mean cloud fraction profiles of the "Ice clouds" (Sect. 4.4). Again, the too-high amount of frozen hydrometeors found in previous analyses leads to a consistent overestimation of cloud depths of more than 3 km in the ICON HOPE 156 m simulations by about 1–2 p.p. compared to the measurements. Likewise, the cloud length of the simulated "Ice clouds" of the ICON HOPE 156 m is shifted to longer time periods. Short "Ice clouds" of less than ten minutes are not seen at all for the ICON HOPE 156 m simulations, similar to the "Liquid clouds", and clouds of less than 30 minutes are substantially underestimated by more than 25 p.p. compared to the observations. In contrast, "Ice clouds" of more than 30 minutes are consistently too frequently simulated by up to 10 p.p.. Combined with the stated too-large cloud depth, the "Ice clouds" of the ICON HOPE 156 m simulations are too large overall, which confirms once more the overestimation of frozen hydrometeors by the ICON HOPE 156 m compared to the measurements.

4.6. Liquid and Ice Water Content Analysis

The accuracy of the correctly simulated vertical distribution of liquid and frozen hydrometeor masses can be assessed by the vertically resolved liquid water content (LWC) and ice water content (IWC). Both parameters can be computed based on the direct model output of the ICON HOPE 156 m simulations, but they cannot be observed by a single instrument because among others of the missing knowledge about the hydrometeor locations of the liquid or ice phase. For that reason, this information is derived from the observed Cloudnet target classification. In combination with the remote-sensing measurements and the temperature and pressure profiles of an atmospheric model, the Cloudnet algorithms (Illingworth et al., 2007) can compute a precise LWC and IWC for the observations.

The observed LWC is calculated based on the temperature and pressure profiles of an atmospheric model to determine the theoretical adiabatic liquid water content gradient for the different liquid cloud base heights of the classification. The estimated LWC is then scaled to the measured integral of the microwave radiometer. The IWC is computed by an empirical formula (Illingworth et al., 2007) using the 35 GHz cloud radar measurements together with the temperature profile of an atmospheric model for all ice targets of the classification. The MOC approach would enable a physically consistent determination of the LWC and IWC for the model simulations applying the same Cloudnet algorithms. Nevertheless, the ICON HOPE 156 m direct model output has to be used because of the
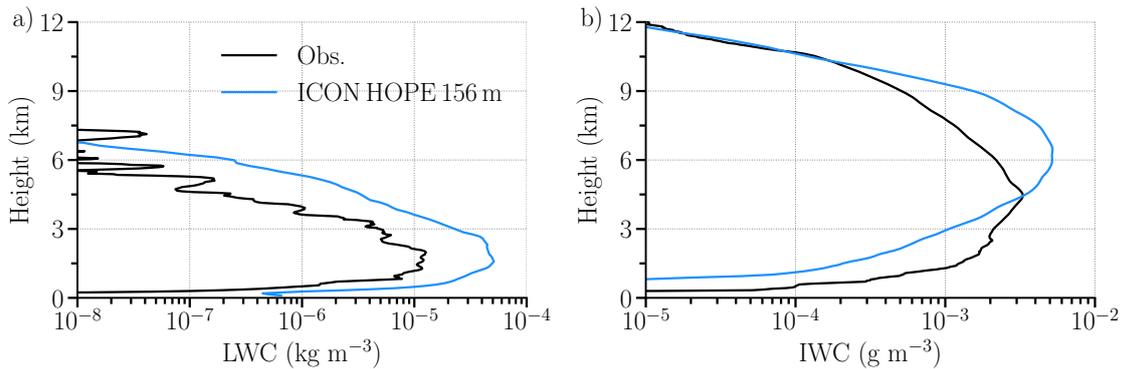
Figure 4.17.: Mean vertical profiles of liquid (LWC, a) and ice water content (IWC, b) for the Cloudnet LACROS HOPE observations and corresponding direct ICON HOPE 156 m simulation output for April and May 2013.

stated issues of the MOC method mentioned in previous sections, which prevents a precise computation of the LWC and IWC for the ICON HOPE 156 m simulations.

The time-averaged vertical profiles of the LWC and IWC are calculated on the basis of the direct model output of the ICON HOPE 156 m simulations and the corresponding LACROS HOPE observations for the two-month period of April and May 2013 (Fig. 4.17). The operational COSMO-DE analyses provide the required temperature and pressure profile information for the Cloudnet algorithms for the calculation of the LWC and IWC of the observations.

A remarkable overestimation of the LWC by almost an order of magnitude below 6 km above ground is clearly recognisable for the simulated LWC profile of the ICON HOPE 156 m simulations in comparison with the observations (Fig. 4.17a). This quantifies the overestimation of the liquid hydrometeors masses and distinctly proves the stated too-high amount of liquid phased hydrometeors of previous analyses. Most probably, the overestimated liquid hydrometeors cause the substantially too-low IWC by more than half an order of magnitude below 4 km height of the ICON HOPE 156 m simulations compared to the measurements (Fig. 4.17b). Issues in the cloud microphysics might explain the completely missing IWC below roughly 1 km above the surface, seen for the ICON HOPE output. These findings are also in accordance with the evaluation of the mean profiles of the "Ice clouds" of Section 4.4.

The too-high amount of frozen hydrometeors in the ICON HOPE 156 m simulations, found also in previous sections, is clearly identifiable by a distinct overestimation of the IWC between 4 km and 10 km by up to half an order of magnitude in comparison with the observations (Fig. 4.17b). Above 10 km, the ICON HOPE 156 m simulations overestimate the frequency of too-small IWC amounts. This explains the good agreement of the mean

IWC profile between 10 km and 12 km of the model with the observations in combination with the found overestimation of the frequency of occurrence of frozen hydrometeors at the analysis of the mean "Ice clouds" profiles by up to 25 p.p. (Sect. 4.4). As already mentioned in the cloud radar analysis, an overestimated ice nucleation rate or too-early formation of frozen hydrometeors by the cloud microphysics might be one possible reason for the substantial overestimation of the ice water content by the ICON HOPE 156 m simulations. The remarkably overestimated liquid water content below 6 km might arise from a too-early melting of the too-high amount of frozen hydrometeors. Both hypothesis should be investigated in more detail to further improve the cloud's representation by the ICON HOPE 156 m simulations. Overall, the LWC and IWC evaluation confirms the findings of the analyses of the forward simulations and of the cloud classifications.

Summary

In conclusion, the quality of cloud representation in the ICON LES model is comprehensively evaluated by two months of daily ICON HOPE runs and corresponding Cloudnet observations of the LACROS HOPE supersite. The clouds are evaluated by forward-simulated remote-sensing instruments and by new synthetic cloud classifications, one based on forward simulations (MOC-approach) and the other on the direct model output (MC-approach) to create a surrogate for the Cloudnet target classification. The cloud classifications are used for new evaluation techniques of clouds. Based on the various analyses, the scientific questions formulated in the beginning can be answered by following:

1. **Which advantages and disadvantages provide a forward-simulated cloud classification compared to a direct model output-based one?**

   The forward-simulated MOC cloud classification computes first the synthetic measurements of all required Cloudnet instruments, to which the same Cloudnet algorithms can be applied afterwards to create a consistent cloud classification for an atmospheric model. The forward-simulated instruments are fully consistent with the atmospheric model, its cloud microphysics, and the real instrument characteristics including for instance attenuation. The physical consistency is one of the major advantages of this approach. Additionally, the synthetic measurements provide detailed insights into the model's cloud properties in the observational space. For example, the right representation of hydrometeors can be assessed by the various parameters of a cloud radar. Furthermore, uncertainties induced by commonly used statistically-based retrievals, which assume mostly a mean climatology can be avoided. Despite these advantages, each forward operator has to be precisely adapted to the assump-

tions of the atmospheric model and the characteristics of the real instruments. Very detailed output of the atmospheric model like of the TKE is necessary for the computation of higher moments of a cloud radar. Likewise, the forward simulations are not trivial and time-consuming in respect to computational- and manpower. The Cloudnet algorithms are not able to capture all categories, unless all synthetic measurements are accurate enough. Currently, the applied PAMTRA forward operator contains technical issues with the simulation of the linear depolarisation ratio and does not support the detection of the melting layer. In combination with the uncertainties of the synthetic lidar, the current forward simulated cloud classification can only detect "Drizzle or rain" and "Ice" targets. For that reason, further research on the forward simulations is necessary regarding the forward-simulated cloud classification approach, which is beyond the scope of this thesis to investigate the general feasibility of this method.

The MC-based cloud classification, inspired by Cloudnet, is using the direct model output of the profiles of temperature, dew point, and hydrometeor mass concentrations to compute a cloud classification. The new algorithm uses simple case selections based on physical principles to determine the different categories. Thus, the algorithm and the assumptions are easy to understand. The MC-based classification can be applied quickly and easily to any atmospheric model, which is one of the major benefits of this approach. The generated MC-based cloud classification of the ICON HOPE within this study contains all different categories as of the observations including mixed phases and shows promising results. However, the integration of real instrument characteristics, such as the attenuation, is very challenging. Additionally, possible deviations at the exact definition of the different categories can occur because not the same Cloudnet algorithm is used to create the classification product. All decisions rely on empirically derived thresholds for the different hydrometeor mass concentrations. Furthermore, this approach does not provide synthetic measurement data such as in the forward-simulated method, and therefore further in-depth analysis like of the rain intensity of rain showers, as seen in the classification, is not possible.

In general, the new synthetic cloud classifications, especially based on the MC approach of the ICON HOPE simulations, enable a comprehensive cloud evaluation with a focus on cloud macrophysical properties. For example, different cloud types and their individual geometric properties can be assessed. Furthermore, the cloud classification provides a very intuitive possibility to get a first impression of the cloud's representation.

2. **How well are clouds represented by the ICON HOPE LES from a forward-simulated perspective and in terms of a cloud classification?**

The synergy of the forward-simulated instruments with the new synthetic cloud classifications allows for a comprehensive evaluation of the simulated clouds of the ICON HOPE 156 m simulations. Liquid hydrometeors are substantially overestimated, and rain hydrometeors are underestimated below about 5 km by the ICON HOPE 156 m simulations compared to the measurements. For example, this is visible by bi-modal distributions of the Doppler velocity and spectral width of the forward-simulated cloud radar. The cloud classification analysis confirms these findings. For instance, "Liquid clouds" are overestimated by up to 10 p.p. below 5 km compared to the observations. The remarkable overestimation of the LWC below 6 km by almost an order of magnitude by the ICON HOPE simulations in comparison with the measurements confirm these results of the forward simulations and of the MC-based cloud classification.

Furthermore, the ICON HOPE 156 m simulations substantially overestimate frozen hydrometeors above roughly 5 km, visible for example by a too-high reflectivity of up to 10 dBZ and reflectivity values above 10 km of the forward-simulated cloud radar compared to the measurements. The ICON HOPE cloud classification based on the MC-approach exhibit a notable overestimation of "Ice clouds" above 5 km by up to 25 p.p. in comparison with the observations. This is also found by the MOC-based cloud classification of the ICON HOPE and is in accordance with the forward simulations. The too-high amount of frozen hydrometeors is also confirmed by the substantial overestimation of the IWC above 4 km of the ICON HOPE output by up to half an order of magnitude compared to the measurements. On contrary, below 4 km, the IWC is underestimated by a similar order by the ICON HOPE 156 m simulations, which is not seen from a forward-simulation or cloud-classification perspective.

The simulated rain events are too intense and rare in the ICON HOPE 156 m simulations in comparison with the measurements. The overestimated intensity is visible by up to 5 dBZ higher reflectivities of the forward-simulated cloud radar than of the real instrument, which was confirmed by additional in situ rain gauge observations. The ICON HOPE simulates especially rain showers that are often too short compared to the measurements. The underestimated frequency of occurrence by up to 5 p.p. of rain below 3 km by the ICON HOPE cloud classification based on the MC approach confirms the too-short rain events compared to the measurements.

Updrafts within deep convection, and at the cloud edges are underestimated by the ICON HOPE 156 m simulations. This is recognisable by an underestimation of positive Doppler velocities and no values larger than $0.25\,\mathrm{m\,s^{-1}}$ of the forward-simulated cloud radar in comparison with the observations. The distribution of cloud base heights of the ICON HOPE 156 m simulations agrees well with the measurements. The cloud base heights are consistently derived from the forward-simulated lidar of the ICON HOPE 156 m output, the direct model output and of the real lidar measurements. Most probably, the matching point-to-point comparison of the cloud base heights leads to large errors of a RMSE of about 900 m.

Chapter 5

Wind Gusts at High-Resolution Large Eddy Simulations

High-resolution LES models can resolve small-scale turbulence and large parts of the energy spectra explicitly and consequently avoid uncertainties due to usually necessary parameterisations. For that reason, LES simulations provide new insights into the underlying processes of turbulence such as wind gusts, which helps to increase our knowledge about the relevant physics. Nevertheless, most LES studies have been limited to idealised setups and only small domains for short periods due to limited computational resources. A realistic weather hindcast like ICON LES case study simulation around Hamburg with a horizontal resolution of down to $20\,\text{m}$ is performed for 24 April 2013 (Sect. 2.2). The assessment of small-scale wind gusts at high-resolution LES requires new evaluation techniques, as well as appropriate observational datasets, which are investigated together with the following scientific questions of this chapter:

1. **Which resolution is required to simulate wind gusts explicitly by large eddy simulations and is there an added value of the conducted ICON HH LES compared to current weather forecast simulations?**

2. **How can wind gusts be parameterised at LES models with insufficient resolution to resolve them explicitly?**

5.1. Wind Gusts and Turbulence Spectrum Theory

So far, there is no commonly agreed-upon definition of a wind gust. For example, according to the US Weather Observing Practice (NOAA, 1998), a wind gust is defined as a peak wind speed with a minimum variation between peaks and lulls of 10 knots and a duration of less than 20 seconds. Knoop et al. (2019) propose a wind gust definition based on a wavelet-analysis. The World Meteorological Organisation (WMO) defines a wind gust as the highest wind speed of a 10-minute interval with a measurement sampling rate of 3 seconds (WMO, 1987) because a 3-second duration has the largest influence on structures like buildings. Beljaars (1987) investigated the influence of different sampling rates and averaging intervals of the measurements on wind gusts. Further studies like Kristensen et al. (1991) investigated possible definitions of wind gusts and their effects on extreme events. The German Weather Service, and many others, use the common WMO definition, which is also applied in this study. Additionally, only very limited high-resolution wind measurements with a resolution of finer than 3 seconds are available from research campaigns or boundary layer towers for the analysis of wind gusts. Most weather stations, like those of the DWD station network, only save the 10-minute average and wind gust speed of each interval and delete the raw measurements afterwards, which would be of great value for in-depth research on wind gusts. The unique dataset of the weather mast Hamburg (Brümmer et al., 2012) provides 20 Hz ultrasonic wind measurements at seven heights up to 280 m since 2004, as used for the analysis in this study (Sect. 2.4). Wind gusts are usually examined by the wind gust factor, as defined by equation 5.1:

$$F_{\text{gust}} = \frac{v_{\text{gust}}}{v_{\text{mean}}} \tag{5.1}$$

- F_{gust} : Gust factor
- v_{gust} : Highest wind speed within a 10-minute interval
- v_{mean} : Average wind speed of 10-minute interval.

Table 5.1.: Dependence of maximum wind gust speed and gust factor on measurement sampling rate using 20 Hz weather mast Hamburg observations of April and May 2013.

Averaging Interval	20 Hz	1 sec.	3 sec.	9 sec.	1 min.	10 min.
Max. Wind Speed ($\mathrm{m\,s^{-1}}$)	21.5	18.8	17.6	15.4	11.7	9.6
Gust Factor	2.03	1.83	1.73	1.60	1.31	1.00

The wind gust factor, as well as observed wind gusts in general, strongly depend on the sampling frequency of the measurement device (Beljaars, 1987), as seen by the analysis of different virtually calculated frequencies using the real 20 Hz weather mast Hamburg data of April and May 2013 (Tab. 5.1).

The widely used 3-second measurement frequency underestimates, for example, the overall maximum of the two-month period by $3.9\,\mathrm{m\,s^{-1}}$ compared to the 20 Hz measurements, and the gust factor is underestimated by 0.3 and 18%. The underestimation even increases rapidly by using less resolved sampling rates. A one-minute sampling rate would already underestimate the overall maximum by $9.8\,\mathrm{m\,s^{-1}}$ and almost 46%, for example. The substantial underestimation exhibits the importance of high-resolution measurement data for wind gust research and for upcoming atmospheric models with similar resolutions.

Turbulence Spectrum

A comprehensive evaluation of the turbulence is possible by computing the turbulence spectrum based on the model output and measurement data. The turbulence spectrum illustrates the energy of the different time periods and spatial scales, linked to the various atmospheric processes (Fig. 5.1; Stull, 1988). Large synoptic systems, like low-pressure systems, generate the atmospheric energy, which is then converted by an energy cascade to smaller eddies ending at the molecular dissipation to heat. The "van der Hoven" gap is defined as the low-energy range between the large synoptic systems and the turbulent scale. Wind gusts occur at the inertial range of the turbulent scale, where mid-size eddies do not generate TKE but are also not affected by viscosity (Stull, 1988). The energy decreases according to the empirical Kolmogorov law (Stull, 1988) by -5/3 with increasing frequency or wave number, in the double logarithmic scale at the intertial range.

The turbulence spectrum is derived by a Fourier transformation of the high-resolution wind speed time-series of the ICON HH model output and of the observations to calculate the energy of the different frequencies, followed by the computation of the spectral density. A Hanning window filter (Blackman and Tukey, 1958) is applied to the resulting spectra to eliminate noisy frequencies and smooth the overall spectrum to highlight relevant features. The turbulence spectrum allows for a detailed evaluation of the simulated turbulence.
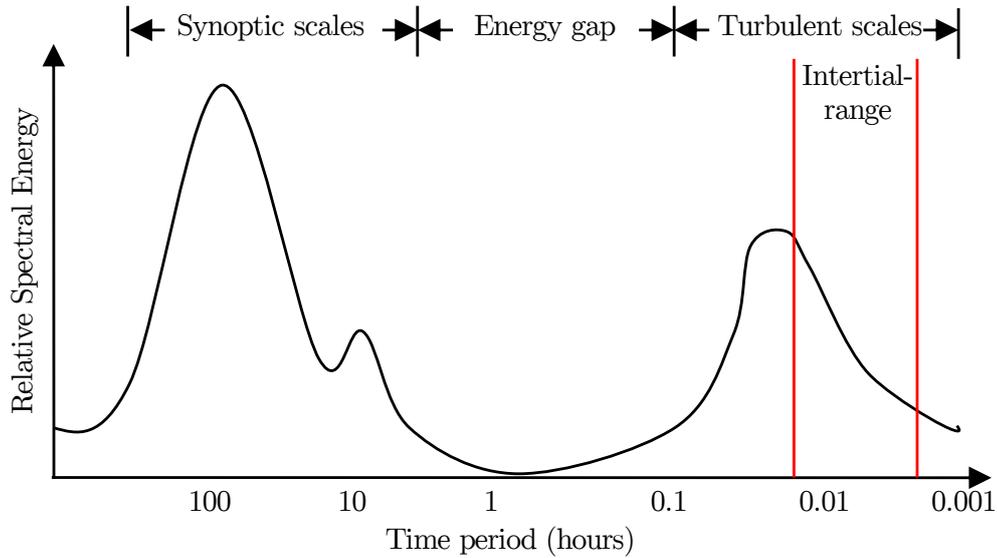
Figure 5.1.: Theoretical energy spectrum of near-ground wind speed after Stull (1988).

The effective model resolution of an atmospheric model can be determined according to Skamarock (2004) by the point where the turbulence spectrum starts to drop from the theoretical -5/3 Kolmogorov law. In this study, the breakdown point is defined as the point where the spectral energy of the computed turbulence spectrum is less than 50% of the theoretical Kolmogorov law to have an objective criteria, which is as well used in Heinze et al. (2017a) and similar to the concept of Skamarock (2004). The 50% criterion is applied on all six nests of the conducted ICON HH simulations. The effective model resolution is, according to Skamarock (2004, 2011), about six to ten times the nominal resolution and depicts the size of the smallest directly resolved atmospheric processes. The mean long-term wind gust characteristics within the atmospheric boundary layer are presented in the upcoming section, followed by a detailed analysis of the turbulence spectra.

5.2. Wind Gust Observations within the Entire Planetary Boundary Layer

The magnitude of the wind gust factor depends upon such factors as the height above ground, diurnal cycle, and season. Further relevant parameters are the surface properties, topography, buildings, and many more. The long-term wind gust statistics within the entire planetary boundary layer are analysed based on the ultrasonic 20 Hz wind measurements of the weather mast Hamburg for the twelve-year period from 2004–2016. According to the proposed 3-second sampling rate of the WMO, the 20 Hz raw wind measurements
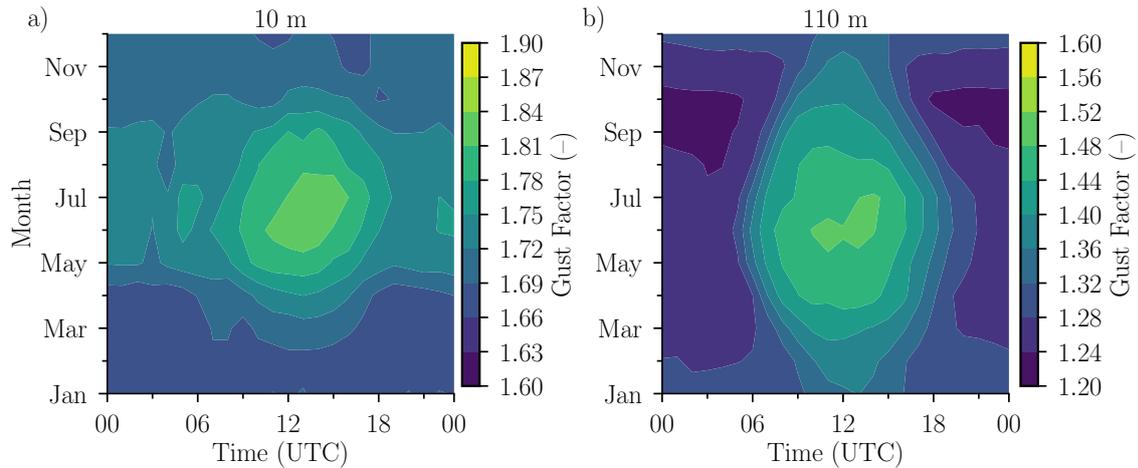
Figure 5.2.: Annual course of the diurnal cycle of the wind gust factor at $10\,\mathrm{m}$ (a) and $110\,\mathrm{m}$ (b) height based on the 10-minute weather mast Hamburg data of 2004–2016. Mind the different colorbars.

are initially averaged over time to 3-second values. The 10-minute average wind speed, the maximum wind gust, and the wind gust factor are calculated afterwards using the 3-seconds average wind speeds in correspondence to the wind gust definition of the WMO.

The annual variations of the diurnal cycle are computed for the near-ground $10\,\mathrm{m}$ and $110\,\mathrm{m}$ height (Fig. 5.2). The larger global radiation during summer triggers the turbulence near surface and unstable stratification. For that reason, the largest wind gusts with speeds of up to 1.85 times higher than the average wind speed are seen in summer during the daytime at $10\,\mathrm{m}$ height. The lower solar irradiation and turbulence during winter lead to smaller wind gust factors of up to 1.67 near-ground.

The stratification of the boundary layer is more frequently near-neutral at night during winter than at night in summer. A strong radiative cooling in summer during the night causes more stable stratified conditions. Most likely, the more near-neutral stratification during winter nights leads to the larger wind gust factors of up to 1.4 compared to wind gust factors of up to 1.2 present at night during summer at $110\,\mathrm{m}$ height. The turbulence, mainly generated by the surface and rising warm air, decreases with height. Accordingly, the occurring wind gusts are also lower at higher altitudes, visible for instance by a gust factor of about 1.8 during July at $10\,\mathrm{m}$ decreasing to 1.5 at $110\,\mathrm{m}$ height (Fig. 5.2).

The mean diurnal cycle of the wind gust vertical profile up to $280\,\mathrm{m}$ height during summer (June, July, August) and winter (December, January, February) is generated using the twelve years of measurements (Fig. 5.3). The higher planetary boundary layer during summer caused by the higher solar irradiation of the sun and increased mixing induces stronger wind gusts at higher altitudes during daytime than at winter times because above
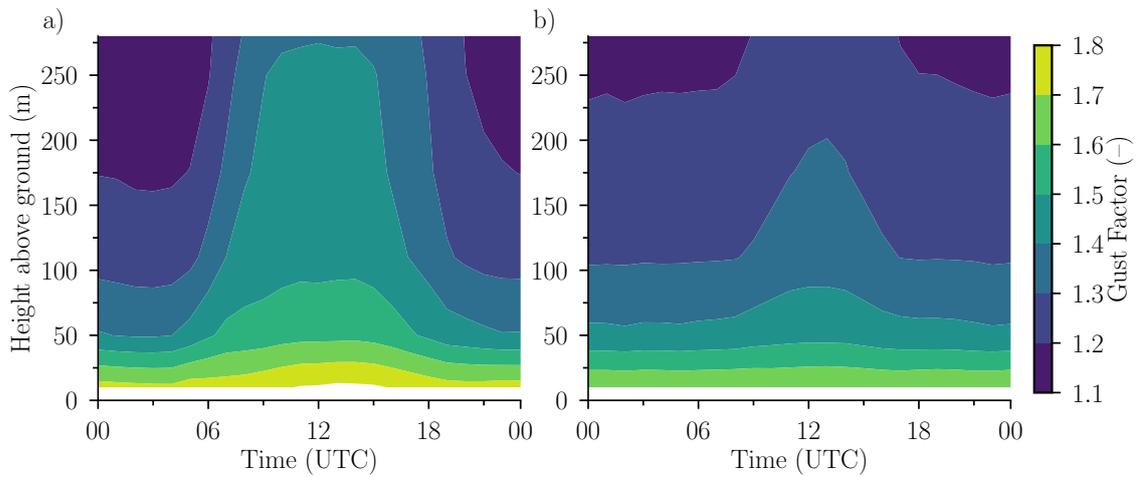
Figure 5.3.: Average diurnal cycle of wind gust factors within the planetary boundary layer of summer (June, July, August; a) and winter (December, January, February ;b) using the weather mast Hamburg data of 2004–2016.

the boundary layer the turbulence is essentially decreased. For example, the wind gusts are on average about 40% stronger than the average wind during summer at 250 m around 12 UTC, whereas during winter the wind gusts are just only roughly 20% stronger than the average wind. Similar results of the wind gusts were found by Förster (2014), who analysed the wind gust observations of the 98 m-tall Lindenberg tower of the DWD. Substantial wind gusts can be observed within the entire boundary layer. Additional turbulence resolving wind measurements at other locations would be very beneficial for further wind gust research instead of the typically available measurements at only 10 m height.

5.3. Case Study on Explicit Wind Gust Simulation by an LES

A single-day case study regarding wind gusts is performed for the 24 April 2013 around the weather mast Hamburg to investigate the following two approaches to resolve wind gusts at the LES scale:

1. Dynamically down-scaled ICON LES simulation,

2. Wind gust parameterisation based on turbulence spectrum.

The first approach uses a realistic, dynamically down-scaled ICON LES simulation with six nests and a horizontal resolution of down to 20 m (Sect. 5.4). The outermost domain covers an area of about 220 km in diameter around Hamburg with a resolution of 624 m. The innermost nest has a diameter of roughly 18 km with a horizontal resolution of 20 m to study wind gusts. Further details about the model setup are described in Section 2.2.
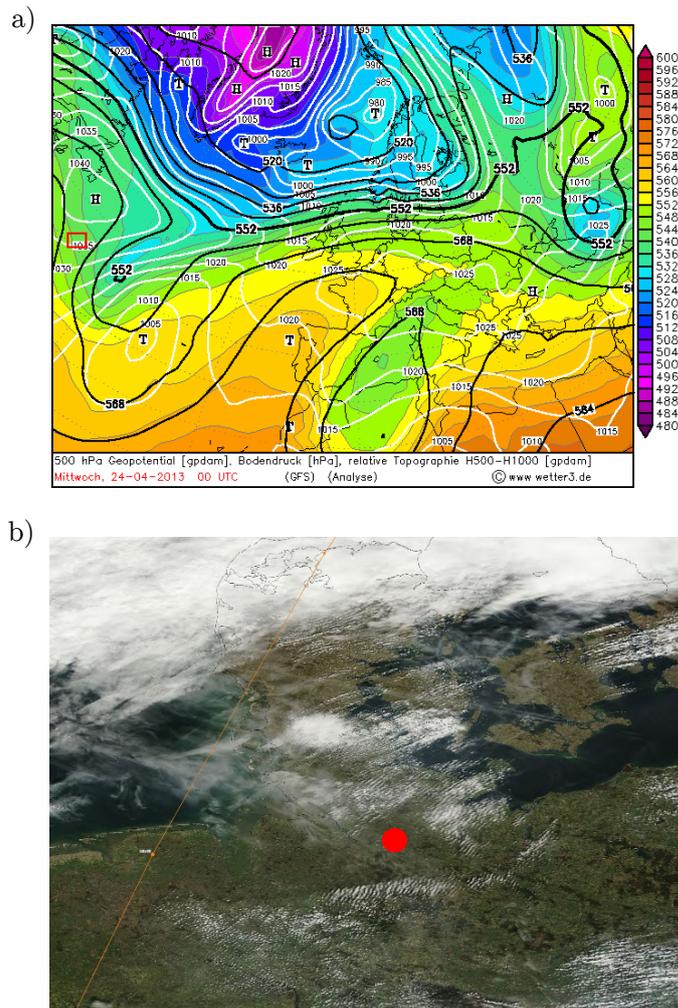
Figure 5.4.: Synoptic situation at central Europe on 24 April 2013. 500 hPa geopotential map (a) with isobars of mean sea level pressure (white lines) and relative topography H500-H1000 (black lines), source: wetter3.de. MODIS visible satellite image (b) of AQUA satellite, source: NASA WorldView.

The large city area of Hamburg, located to the west of the weather mast Hamburg, is only represented in the ICON LES by a higher surface roughness parameter and urban land-use class properties. However, single buildings are not explicitly resolved by the ICON LES even at the 20 m nest of the ICON HH simulations. This simplification has to be considered at the interpretation of the results. The second approach to resolve wind gusts consists of a new parameterisation based on the turbulence spectrum applied to high-resolution LES output (Sect. 5.5). The reference data are the full-resolution 20 Hz wind measurements of the weather mast Hamburg.

The weather on 24 April 2013 was dominated by local effects rather than by large synoptic systems, seen among others by low wind speeds, which was the reason to select this

day. Consequently, the small inner domains should be large enough to capture all local atmospheric processes. Hamburg was on the backside of a trough (Fig. 5.4) with a low wind speed of about $2\,\mathrm{m\,s^{-1}}$ from south-west at $10\,\mathrm{m}$ height. Almost clear-sky conditions were present throughout the entire day with only few shallow cumulus clouds during the daytime. Therefore, this day is well suited to analyse the mainly locally induced wind gusts within the atmospheric boundary layer.

5.4. Wind Gusts Resolved by Dynamical Downscaling

High spatial and temporally resolved simulations and measurements are necessary to capture small-scale wind gusts. A sampling rate of 3 seconds is suggested by the WMO for the observations. For that reason, the ICON HH case study simulation is dynamically downscaled to a horizontal resolution of $20\,\mathrm{m}$ and a 3-second output interval. The ICON HH wind gusts are calculated according to the measurements by the highest simulated 3-second output wind speed for each 10-minute average wind speed interval. Nevertheless, the effective resolution is usually six to ten times lower than the nominal resolution (Skamarock, 2011), which is determined first by the temporal turbulence spectrum using the 3-second and $10\,\mathrm{m}$ wind speed output (Fig. 5.5). As a reference, the $20\,\mathrm{Hz}$ weather mast Hamburg spectrum is added.

The effective model resolution is determined in this study by the point where the energy of the computed turbulence spectrum is for the first time less than 50% of the theoretical -5/3 Kolmogorov law. This criterion allows for an objective and replicable computation of the point. The method is based on the concept of Skamarock (2004), where the effective model resolution is defined as the point where the spectrum starts to break down. An average wind speed at $10\,\mathrm{m}$ height of $3\,\mathrm{m\,s^{-1}}$, comparable to the multi-annual mean of the weather mast Hamburg (Brümmer et al., 2012), is assumed to convert the temporal scale into a spatial scale as a first approximation.

The resolved turbulence of the parent $624\,\mathrm{m}$ ICON HH simulation already breaks down at a time period of 1763 seconds and about $5300\,\mathrm{m}$ (Fig. 5.5). This is eight times larger than the nominal resolution and therefore far too coarse to simulate 3-second wind gusts. The $20\,\mathrm{m}$ ICON HH nest resolves many more scales of the turbulence spectrum explicitly but also starts to break down already at a period of 72 seconds, corresponding to a spatial scale of about $220\,\mathrm{m}$. Accordingly, the effective resolution is by a factor of about eleven lower than the nominal resolution. The factors between eight and eleven for the effective model resolution are similar to those found by Skamarock (2004) for atmospheric models. Therefore, even the high nominal $20\,\mathrm{m}$ ICON HH resolution is still too coarse to resolve all wind gusts explicitly. As a consequence, a nominal LES resolution of $0.9\,\mathrm{m}$ would be
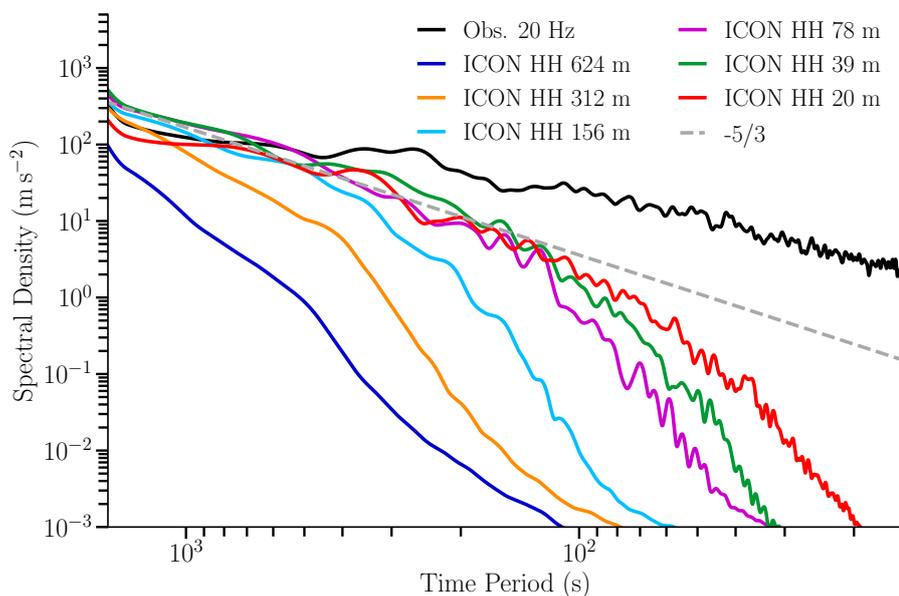
Figure 5.5.: Energy spectrum of inertial range for 24 April 2013 using the 10 m wind speed of the 20 Hz weather mast observations and the model output of all six ICON HH resolutions down to 20 m. The -5/3 Kolmogorov law is illustrated as gray dashed line.

required to simulate all 3-second wind gusts of the WMO-definition explicitly. This would be valid under the assumption of an effective resolution factor of ten and an average wind speed of $3\,\mathrm{m\,s^{-1}}$ to convert the temporal to a spatial scale. Similar to this theoretical estimate, Knigge and Raasch (2016) and Ahmad et al. (2017) use LES with horizontal resolutions of 2 m to study wind gusts.

The diurnal cycle of the vertical wind gust factor profiles is generated for the ICON HH 624 m and ICON HH 20 m nest, as well as for the 20 Hz wind speed observations averaged to 3 seconds. All datasets are processed according to the WMO wind gust definition (Fig. 5.6). As expected from previous analyses, the observations show wind gust factors of up to 1.8 near-ground at roughly 10 m in height. In contrast, the 624 m resolved ICON HH simulation contains almost no wind gust at all, and only three small areas with wind gust factors of less than 1.1 are identifiable. A distinct and reasonable turbulence structure with wind gusts of up to 50% faster than the mean wind speed is seen for the high-resolution 20 m ICON HH output. The simulated diurnal cycle of the wind gust factor contains similar patterns as of the observations. Consequently, a clear added value of LES simulations at resolutions on the order of 10 m is visible by resolving even small-scale turbulent structures compared to coarser resolutions of for instance on the order of 100 m. The explicit simulation is especially valuable for wind gust studies.
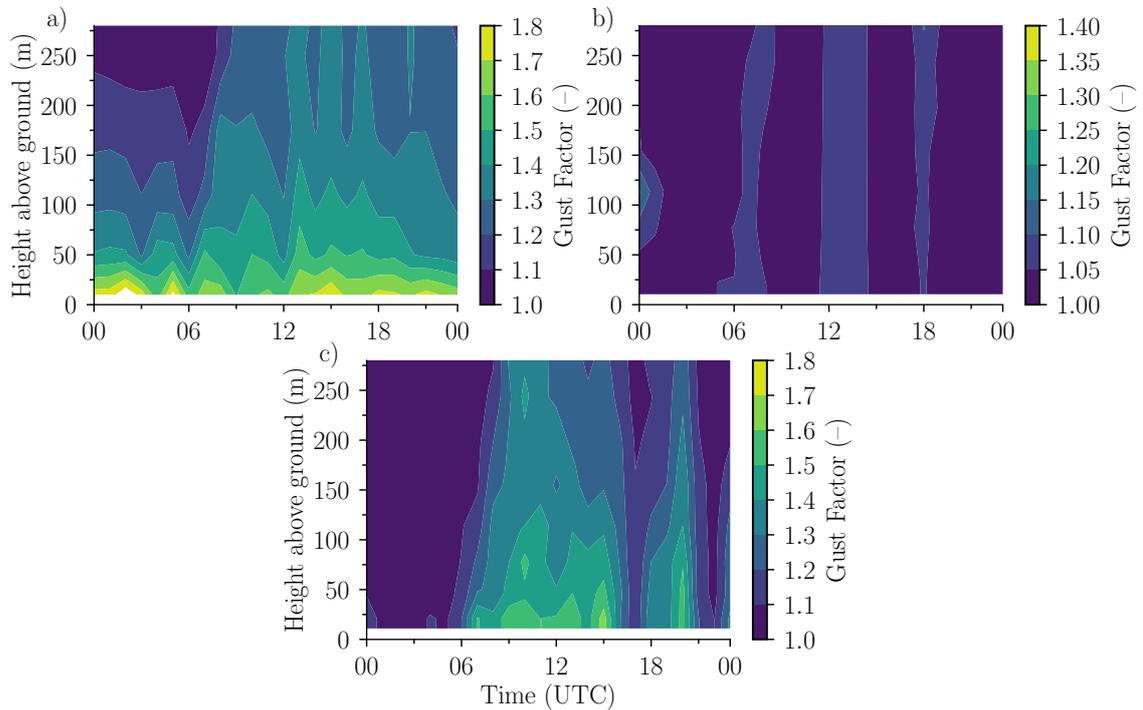
Figure 5.6.: Diurnal cycle of wind gust factor profile on 24 April 2013 of the weather mast Hamburg (a), the ICON HH 624 m (b) and ICON HH 20 m (c) nest. Mind the different colorbars.

Nevertheless, the maximum wind gust factors near-ground are still underestimated by about 0.3 and 17% due to the still-insufficient resolution to resolve the full turbulence spectrum (Fig. 5.6). Consequently, these findings indicate once more a necessary higher LES resolution to resolve all relevant wind gusts within the entire boundary layer explicitly. Similar results are found at the analysis of the energy spectrum before. Most probably, the matching point-to-point comparison and chaotic properties of turbulence explain the stated deviations of the overall structure between the ICON HH simulation results and the observations. Uncertainties at the initial and boundary conditions of the parent domain, as well as inaccuracies of the land surface properties could also explain parts of the differences. Deviations between the simulated cloud and real cloud field might affect the solar irradiation and can thus influence the turbulence structure, leading to a different diurnal cycle.

The 20 m nominal resolution of the ICON HH simulations is according to the stated results too coarse to directly resolve all wind gusts explicitly, as discussed above, and an appropriate parameterisation is still required. A new turbulence spectrum-based parameterisation for wind gusts is proposed in the following section.
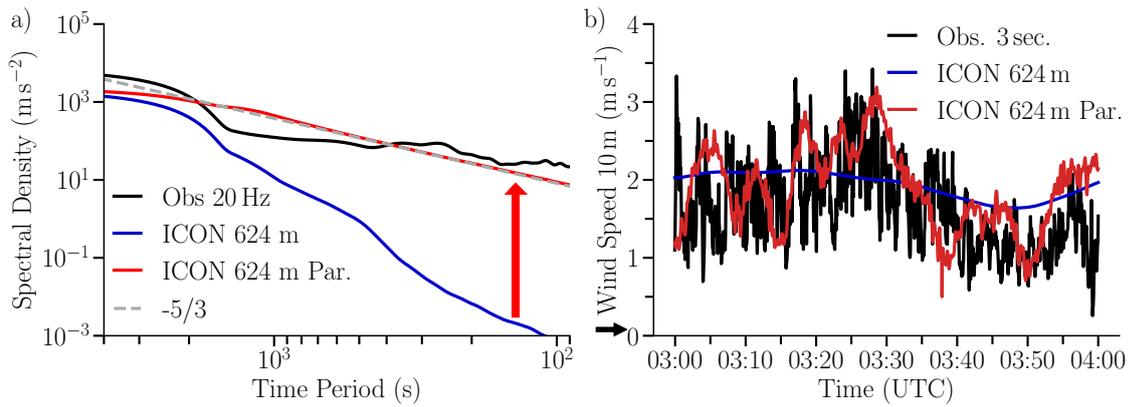
Figure 5.7.: Energy spectra of 10 m wind speed of 24 April 2013 (a) for the weather mast Hamburg (black), the ICON HH 624 m original output (blue), the parameterised (Par.) ICON HH 624 m output (red), and the Kolmogorov law (dashed gray). Wind speed time series between 03 UTC and 04 UTC (b) of the observations (black), the original ICON HH 624 m (blue), and the ICON HH 624 m parameterised (Par.) output (red).

5.5. Wind Gust Parameterisation based on Turbulence Spectrum

Even high-resolution LES models with resolutions on the order of 10 m are not able to directly resolve the full turbulence spectrum and all wind gusts explicitly, as seen by the results in the previous section. For that reason, a new wind gust parameterisation is developed for LES models based on the turbulence spectrum. The energy spectra of LES models like the ICON HH 624 m breaks down too early from the theoretical -5/3 Kolmogorov law at the inertial range (Fig. 5.5). Consequently, the energy of higher frequent variations is too low compared to the observations.

The proposed wind gust parameterisation assumes that the simulated turbulence spectrum contains already even high frequent variations, but their amplitude is only damped. Therefore, the parameterisation adds artificial turbulent kinetic energy to the smaller eddies by adjusting the turbulence spectrum after the breaking point to the theoretical -5/3 Kolmogorov law, as illustrated by Figure 5.7a. The additional artificial kinetic energy is physically inconsistent, which has to be kept in mind. Nevertheless, the energy is only added at the post-processing of wind gusts and not during the model integration to avoid physical errors. The wind speed time-series is then reconstructed by an inverse Fourier transformation using the modified turbulence spectrum, depicting the additional induced variability of higher frequencies (Fig. 5.7b).

The parameterisation is exemplary applied to the 24-hour long ICON HH simulation with 624 m and 20 m horizontal resolution. The turbulence spectra are adjusted for each height
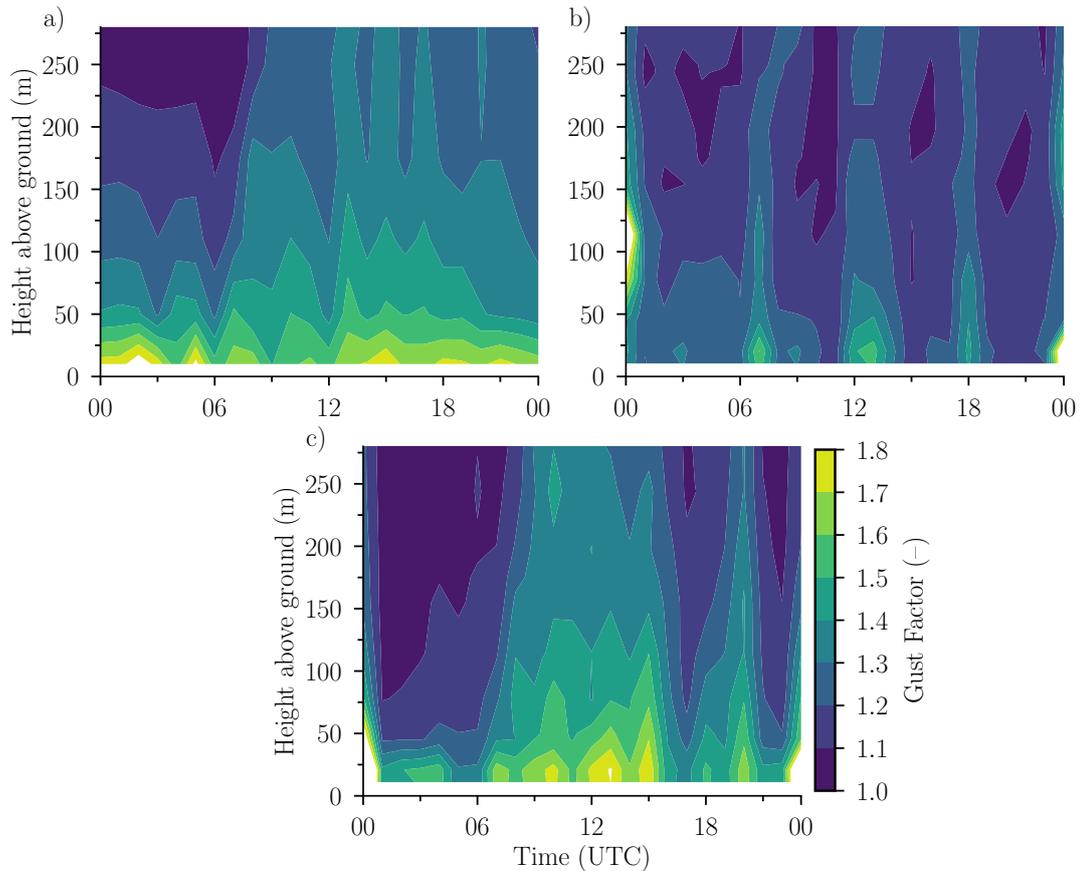
Figure 5.8.: As Figure 5.6, but for parameterised ICON HH 624 m (b) and parameterised ICON HH 20 m (c) output.

level separately. The diurnal cycles of the vertical wind gust factor profiles (Fig. 5.8) are computed based on the reconstructed wind speed time-series of the modified turbulence spectra. The original output, as well as the weather mast Hamburg measurements, are used as references (Fig. 5.6).

As expected, the parameterisation amplifies the small signals of the original ICON HH 624 m output (Fig. 5.8b), but the highest wind gust factors found at low levels are still underestimated by about 17 % (Fig. 5.6a,b). Nevertheless, this is a major improvement to the previous underestimation of more than 40 % found for the non-parameterised output. The largest simulated gust factors improved to roughly 1.5 instead of the almost nonexistent gust factors of 1.05, although the observed wind gust factor is still about 0.3 higher at about 1.8. Consequently, the error of the wind gust factor is remarkably reduced by up to 60 %.

The already good results of the 20 m ICON HH simulations (Fig. 5.6c) are further improved by the new parameterisation and are also not amplified too strongly (Fig. 5.8c). The near-

ground wind gust factors are very similar to the observations with values of up to 1.8, which are underestimated by 0.2 for the non-parameterised model output. The parametrised output shows, in general, a reasonable diurnal cycle of the vertical wind gust profile. The differences of the structure compared to the observations are caused by the original ICON HH 20 m input (Sect. 5.4). The overall results of the proposed parameterisation look very encouraging and should be investigated in more detail.

Summary

The twelve years (2004–2016) of high-resolution 20 Hz wind measurements of the 300 m tall weather mast Hamburg are used to study wind gusts within the entire boundary layer. The highest wind gusts of up to a factor of 1.85 larger than the average wind speed are found near-ground during the daytime in summer. The strength of wind gusts is decreasing with height and overall lower during the night. Likewise, the strength of wind gusts is lower at daytime during winter than during summer. For example, the wind gust factors decrease in winter during noon down to about 1.67 compared to 1.85 in summer at 10 m height.

Large eddy simulations resolve small-scale turbulence and large parts of the energy spectrum explicitly due to their high resolution. Therefore, they allow for detailed investigations of the underlying processes of for instance wind gusts. A one-day ICON LES case study simulation with a horizontal resolution of down to 20 m is performed and used together with high-resolution 20 Hz wind measurements of the weather mast Hamburg to analyse wind gusts. The turbulence spectrum and the boundary layer profiles of the wind gust factor are evaluated for the ICON HH LES simulations. Based on the results, the initially stated scientific questions can be answered as follows:

1. **Which resolution is required to simulate wind gusts explicitly by large eddy simulations and is there an added value of the conducted ICON HH LES compared to current weather forecast simulations?**

   The stated theoretical estimation notes for a necessary nominal LES resolution of roughly 1 m to resolve all 3-second wind gusts of the WMO definition explicitly. This is based on the findings of the energy spectrum analysis and an assumed effective resolution, which is ten times larger than the nominal resolution. The performed 624 m-resolved ICON HH simulation is too coarse to resolve wind gusts explicitly, which is recognisable by the largest wind gust factors seen of only 1.05. In contrast, a reasonable diurnal cycle of the wind gust factor profile within the boundary layer is visible for the ICON HH simulation with a 20 m horizontal resolution. The results indicate a correct simulation of the underlying physical processes relevant for

wind gusts. Consequently, a clear added value can be identified for the 20 m-resolved ICON HH simulation compared to current weather prediction models with resolutions on the order of 1 km, which is not the case for the coarser 624 m ICON HH simulation. Nevertheless, the ICON HH 20 m still underestimates wind gusts by about 17% and requires an appropriate parameterisation to simulate the wind gust strength correctly. Also, the effective resolution of about 72 seconds and 220 m is still too coarse to resolve all wind gusts explicitly, as seen by the analysis of the energy spectrum.

2. **How can wind gusts be parameterised at LES models with insufficient resolution to resolve them explicitly?**

A new wind gust parameterisation is proposed based on the turbulence spectrum and the assumption that even coarse LES models already contain all signals of high frequent turbulence, which are only damped in their amplitude. For this, the turbulence spectrum of the simulation is aligned to the empirical -5/3 Kolmogorov law by adding artificial turbulent kinetic energy to the higher frequencies, representing wind gusts and small-scale turbulence. The simulated spectrum is modified after the drop from the theoretical spectrum. The parameterisation increases the underestimated amplitudes of the wind speed variability. The exemplary parameterised wind gusts of the 624 m and 20 m resolved ICON HH simulations exhibit promising results for the diurnal cycle of the boundary layer profiles. A remarkable improvement of the wind gust representation is achieved for the 624 m ICON HH output, which is visible by reducing the underestimation of the wind gust factors from 42% to 17% compared to the measurements.

Chapter 6

Conclusions and Outlook

An overarching evaluation for atmospheric models based on a bottom-up approach, illustrated by the newly introduced evaluation pyramid, is presented in this thesis, which considers the different levels of complexity of an atmospheric model. The new concept is appropriate for Large Eddy Simulations (LES) and applied to the new ICON LES model. The evaluation strategy advances and applies techniques from the comprehensive COSMO evaluation by Hansen (2014) for the new challenges of LES. The new evaluation approach includes classical verification methods for basic atmospheric state parameters like of the wind, temperature and humidity profiles, as well as novel evaluation techniques and products regarding clouds. Additionally, new LES capabilities such as of the explicit simulation of small-scale turbulence like wind gusts are assessed.

The comprehensive evaluation ensures physical consistency and realism for simulations with realistic initial and boundary conditions, which is the basic prerequisite of all model-based studies. The example of a physically wrong implementation of the surface momentum flux in an earlier ICON LES version resulted in a substantial overestimation of the near-ground wind speed, which was found at the evaluation of the 10 m wind speed.

This example distinctly illustrates the importance of the evaluation of even basic atmospheric parameters. The incremental assessment of the different stages of an atmospheric model according to the evaluation pyramid also examines whether higher-order atmospheric quantities such as clouds are simulated correctly because of well-represented basic atmospheric parameters, or if this is the case because of compensating errors.

The novel ICON LES model is extensively evaluated in this study according to the evaluation pyramid by several conducted simulations considering different aspects. Germany-wide ICON LES runs, named ICON GER within this study, with a horizontal resolution of down to 156 m of the HD(CP)^2 project (Heinze et al., 2017a) are used for the evaluation of basic atmospheric parameters (Chapter 3). Clouds are assessed by a small-scale ICON LES setup around Jülich, Germany, with a horizontal resolution of down to 156 m for April and May 2013 (Chapter 4). This is the same time period and region as of the HOPE measurement campaign. Accordingly, those simulations are referred as ICON HOPE. The small-scale turbulence and wind gusts are investigated by a one-day ICON LES case study simulation with six nests and a resolution of down to 20 m around Hamburg for 24 April 2013 (Chapter 5), called ICON HH in this study. A wide variety of observations ranging from ground-based measurements, soundings, boundary layer towers, and remote-sensing instruments up to satellite data are used as reference data at the evaluation.

Baseline evaluation

The physical consistency of the ICON LES model as the fundamental prerequisite of all model-based studies is first evaluated by a comprehensive baseline evaluation. The basic atmospheric state consists of the boundary and flow conditions, the stratification, and humidity. For this purpose, the near-ground conditions, boundary, and full troposphere profiles of the wind, temperature, and humidity, and the mean sea level pressure are considered. In total, twelve days of realistic Germany-wide ICON GER simulations with three domains of 624 m, 312 m, and 156 m resolution are included in this overarching evaluation. All simulations get their realistic initial and hourly boundary conditions from the operational COSMO-DE analyses. The days are from spring and summer between 2013 and 2016. Additionally, state-of-the-art, cloud-resolving COSMO simulations with 2.8 km horizontal resolution are conducted for the same days and domain as a benchmark to investigate the accuracy and to identify an added value of the high-resolution LES. The same cloud microphysics are used for the COSMO simulations as for the ICON GER simulations.

Overall, the basic atmospheric state is well-represented by all ICON GER simulations, which is among others visible by an on average 48% lower RMSE of all considered near-

ground quantities than the observed natural variability, represented by the standard deviation of the measurements. The uncertainties of most basic atmospheric quantities of the novel ICON GER simulations are already of a similar magnitude or slightly above compared to the well-established COSMO model with general low errors. This is indicated by an RMSE of the wind, temperature, and specific humidity of the ICON GER simulations, which is less than 20% higher than the RMSE of the COSMO simulations for most locations and times, including near-ground conditions, and boundary layer, and full troposphere profiles. Only the mean sea level pressure of the ICON GER simulations exhibit a substantially higher uncertainty of about 31% for the standard deviation than the COSMO output.

The model uncertainties of the ICON GER and COSMO simulations are dominated by the random error, and the bias is very low for almost all considered parameters. The bias of the basic atmospheric quantities accounts for less than 13% of the RMSE for all near-ground parameters and for most boundary layer and full tropospheric profiles. Despite the common expectations of a higher accuracy due to the higher resolution, the errors increase with the horizontal resolution of the different ICON GER simulations for almost all parameters. On average, across all basic atmospheric parameters, the RMSE of the ICON GER 156 m is about 7% larger than the 624 m-resolved ICON GER simulations. In contrast, the RMSE of the 312 m ICON GER is only about 3% larger than the 624 m-resolved ICON GER output. Similarly, the simulation of the boundary layer height and boundary layer profiles do not improve by the additional vertical levels of the ICON GER with 150 layers compared to the 50 layers of COSMO in combination with the finer horizontal resolution. All simulations consistently underestimate the boundary layer height by up to a factor of two in the evening at 18 UTC, which is similar for all ICON GER simulations, as well as for the COSMO simulations.

The largest deviations to the measurements are especially seen at mountainous regions in South Germany. Also, a horizontal resolution of 156 m of the ICON GER seems still to be too coarse to resolve all small-scale effects at regions with complex terrains. This is for example recognisable by an RMSE of the 2 m temperature, which is twice as large for complex terrain regions than for locations with flat terrain, as well as by a north-south gradient across Germany. The larger errors found at complex terrain is consistent with the COSMO simulations. Therefore, an added value of the higher resolution of the ICON GER simulations could not be identified in terms of the basic atmospheric quantities.

The most remarkable model issues of the ICON GER simulations are found at the diurnal cycles of the 10 m wind speed, 2 m temperature, and 2 m specific humidity. Most likely a too-strong coupling of the model levels by the ICON GER during the night leads to a too-strong downmixing of momentum resulting in an overestimated 10 m wind speed by

up to $0.5\,\mathrm{m\,s^{-1}}$ and 20% compared to the measurements. Furthermore, the decoupling of the lower layers after sunset, causing a decrease of the near surface wind, is completely missed by all ICON GER simulations. Similarly, the supposed increased downmixing of warmer and drier air from higher layers during the night causes an overestimation of the $2\,\mathrm{m}$ temperature by about $1.5\,\mathrm{K}$ and an underestimated $2\,\mathrm{m}$ specific humidity of roughly 14% in comparison with the observations. In contrast, the well-tuned COSMO simulations agree overall well with the measurements during the night.

During the day, a possibly too-weak coupling of the model levels by the ICON GER and COSMO simulations might cause a too-low transport of momentum from higher levels to lower levels. The too-weak momentum transport results in an underestimation of the $10\,\mathrm{m}$ wind speed by up to $0.5\,\mathrm{m\,s^{-1}}$ and 14% for the ICON GER and COSMO simulations. The reduced mixing and therefore higher influence of the ground could explain the roughly one-hour too-early decline of the $2\,\mathrm{m}$ temperature after the maximum at noon for the ICON GER and COSMO simulations. Likewise, less dry air from higher levels is downmixed and might lead to the remarkable overestimated $2\,\mathrm{m}$ specific humidity by up to $0.5\,\mathrm{g\,kg^{-1}}$ and 6% for the ICON GER and COSMO simulations.

The ICON GER simulations are on average too dry and underestimate the near-ground specific humidity by -0.17 $\mathrm{g\,kg^{-1}}$ to -0.29 $\mathrm{g\,kg^{-1}}$ and 2-3% for the different domains, whereas the COSMO simulations are too wet and overestimate the specific humidity by roughly $0.31\,\mathrm{g\,kg^{-1}}$ and 4%. The complex precipitation is only briefly investigated in terms of precipitation rates using Germany-wide rain radar measurements. The different ICON GER and COSMO simulations agree reasonably well with the observations, which is visible by deviations of the frequency of occurrence of less than 2 percentage points except for light rain of less than $1\,\mathrm{mm\,hr^{-1}}$. Those rain rates are overestimated by up to 5 percentage points by all ICON GER simulations but not by the COSMO simulations. Light rain of less than $3\,\mathrm{mm\,hr^{-1}}$ is overestimated by all ICON GER simulations, whereas precipitation rates between 3-12 $\mathrm{mm\,hr^{-1}}$ are underestimated. Intense rain events with precipitation rates of more than $12\,\mathrm{mm\,hr^{-1}}$ are simulated too frequently by the ICON GER simulations.

The concept of the evaluation pyramid is tested by the influence of a well and poorly represented basic atmospheric state on the $2\,\mathrm{m}$ specific humidity. For this purpose, the basic parameters of the mean sea level pressure, the near-ground wind speed, and $2\,\mathrm{m}$ temperature are considered. All simulated time steps of the ICON GER simulations with the lowest and highest 25% deviations compared to the observations are selected to filter the $2\,\mathrm{m}$ specific humidity time series for a well and for a poorly represented basic atmospheric state. A remarkable reduction by up to 23% for the ICON GER simulations and by up to 34% for the COSMO simulations is achieved for the RMSE of the $2\,\mathrm{m}$ specific humidity considering only time steps with a well-represented basic atmospheric state. The

substantial error reduction confirms the evaluation pyramid concept, which assumes a well-represented basic atmospheric state as a basic prerequisite for an accurate simulation of clouds and precipitation.

Altogether, the basic atmospheric state is well represented by the ICON GER simulations. The errors are only slightly higher than the long-established COSMO model except for the remarkable deviations found at the mean diurnal cycle of the near-ground wind speed, temperature, and specific humidity. The overall good results of the ICON GER simulations are quite remarkable for a novel and untuned model. The high resolution of the ICON GER allows the simulation of many processes like shallow convection explicitly and can therefore avoid uncertainties due to usually necessary parameterisations such as those used by the COSMO simulations. This might explain the general low uncertainties of the untuned model found at the various analyses.

Cloud Evaluation

The overall low uncertainties found at the baseline evaluation ensures the physical consistency of the new ICON LES model, which is the basic prerequisite for the correct simulation of clouds. Therefore, a comprehensive cloud evaluation is conducted for the ICON LES using the ICON HOPE simulations according to the bottom-up evaluation concept of this study. However, clouds are one of the most difficult quantities to evaluate in atmospheric models due to their various properties and high spatial and temporal variability. For this purpose, the algorithms of the Cloudnet project combine measurements of a cloud radar, lidar, microwave radiometer, and simple rain gauge to the comprehensive Cloudnet target classification. The target classification provides detailed information about the cloud structure, phase, and macrophysical properties. This product allows for an intuitive view on clouds and is well-suited for a comprehensive cloud evaluation, but so far there was no appropriate model surrogate available. Two different approaches are developed in this study and applied to the realistic two-month ICON HOPE simulations for a supersite location of a 220 km-in-diameter large domain for April and May 2013. The forward-simulated cloud remote-sensing measurements are used in combination with the new cloud classifications for an overarching cloud evaluation of the ICON HOPE simulations. Additionally, a connected component analysis is applied to the cloud classification to investigate single cloud objects and their geometric properties considering the cloud phase.

The first method to generate a synthetic cloud classification is the Model to Observations to Classification (MOC) approach, which uses forward operators to compute virtual observations of all by the Cloudnet algorithms required remote-sensing instruments for an atmospheric model. The forward operators are based on the well-known radiative transfer calculations. The same Cloudnet algorithms are applied afterwards to the synthetic

instruments to create a consistent cloud classification. Accordingly, the MOC approach is physically consistent with the atmospheric model, its cloud microphysical scheme, and the Cloudnet algorithms. Furthermore, it can incorporate real instrument's characteristics like attenuation. The physical consistency is one of the major advantages of this method. Additionally, the synthetic measurements of the atmospheric model itself enable an in-depth cloud evaluation. However, each forward operator has to be adapted carefully to the model's physics. Furthermore, the forward simulations of all instruments are not trivial and are time consuming in terms of manpower and of computational efforts. Detailed model output of for instance the TKE is required for the simulation of higher moments of the cloud radar. The Cloudnet algorithms require accurate forward simulations to detect all different cloud categories. So far, the MOC-based cloud classification can only detect "rain" and "ice" targets, whereas all liquid and mixed-phase clouds are not captured. These problems are caused by technical issues at the simulation of the linear depolarisation ratio and detection of the melting layer by the cloud radar of the PAMTRA forward operator, as well as by uncertainties of the simulated lidar of the CR-SIM forward operator. Nevertheless, the general feasibility of this approach is confirmed, and a further development of the forward simulations would be very worthwhile.

The second approach, the Model to Classification (MC) algorithm, inspired by Cloudnet, uses the direct atmospheric model output of the temperature, dew point, and specific hydrometeor mass profiles. These profiles are used in combination with several physically based case selections to compute a comparable cloud classification for the atmospheric model output. The MC-based approach can be applied quickly to new atmospheric model simulations, and the classification decisions are well comprehensible, which are the major advantages of this method. Nevertheless, the thresholds for the hydrometeor mass concentrations to determine the different categories are only empirically derived and not physically based. Also, real instruments' characteristics are very challenging to be considered. Nevertheless, the example two-month MC-based cloud classification of the ICON HOPE 156 m simulation contains all different categories as of the observations and shows overall promising results.

The forward-simulated instruments are used in combination with the MC- and MOC-based cloud classifications of the ICON HOPE 156 m simulations for an in-depth cloud evaluation. An example two-day time period of the simulated MC-based cloud classification agrees well with the observations, as shown by similar cloud structures and phases. This is especially remarkable considering the matching point-to-point comparison of the deterministic ICON HOPE 156 m simulations.

Frozen hydrometeors are substantially overestimated by the ICON HOPE 156 m simulations above roughly 5 km in terms of their frequency of occurrence and of their mass

concentrations. This is seen by higher reflectivities of up to 10 dBZ and more frequently occurring reflectivities above 10 km by the forward-simulated cloud radar compared to the observations, among other aspects. These findings are in accordance with the results of Heinze et al. (2017a). A remarkable too-high frequency of occurrence of up to 25 percentage points is identified for "Ice clouds" above 5 km by the ICON HOPE 156 m simulations in comparison with the observations. Also, the cloud depth and length of the simulated "Ice clouds" are overestimated. Remarkable is the overestimation of the IWC between 4 km and 10 km by almost half an order of magnitude by the ICON HOPE 156 m simulations in comparison to the measurements. Below about 4 km, the IWC of the ICON HOPE 156 m simulations is underestimated by half an order of magnitude.

The liquid hydrometeors are overestimated and rain hydrometeors underestimated below about 5 km by the ICON HOPE 156 m simulations. This is seen by such issues as the bimodal distributions of the forward-simulated Doppler velocity, spectral width of the cloud radar and a too-high LWP compared to the observations. The frequency of occurrence of "Liquid clouds" is overestimated by up to 10 percentage points below 5 km. Notable is the overestimation of the LWC by up to one order of magnitude below 5 km of the ICON HOPE 156 m simulations compared to the measurements.

The simulated rain showers of the ICON HOPE 156 m are often too rare and intense, visible for instance by higher reflectivities of up to 5 dBZ of the forward simulated cloud radar and by the accumulated surface precipitation. This is again in agreement with the findings of Heinze et al. (2017a). The distribution of cloud base heights of the ICON HOPE 156 m simulations matches the observations well. Despite the issues of the frozen and liquid hydrometeors, the overall cloud structure and distributions are well captured by the ICON HOPE 156 m simulations.

In conclusion, a comprehensive evaluation of the cloud macrophysical properties is possible by the newly developed synthetic cloud classification using for example a connected component analysis to derive further higher-order cloud properties. Nevertheless, the forward simulations still need to be improved to accurately capture, among other aspects, the different target categories and mixed-phase clouds of the ICON HOPE 156 m simulations.

Small-Scale Turbulence Evaluation

In addition to the direct simulation of even small-scale clouds such as shallow convection due to the high resolution, LES models also resolve large parts of the energy spectrum, as well as turbulent fluctuations explicitly. Accordingly, the new ICON LES simulations are well suited to study even small-scale turbulence structures like wind gusts, which are extensively evaluated using the ICON HH simulations. A one-day ICON HH case study

simulation is conducted for a 220 km in diameter large domain around Hamburg by a dynamical downscaling of six nests down to a horizontal resolution of 20 m for 24 April 2013. The 20 Hz wind measurements of the 300 m-tall weather mast Hamburg tower are used as observational reference to investigate wind gusts in general and to assess the ICON HH model simulations.

The underlying physical processes of wind gusts are still not well understood, and there is so far no common wind gust definition agreed upon in the community. For example, the measured strength of wind gusts strongly depends on the sampling rate. This is visible by an overall captured maximum wind gust of $21.5 \, \mathrm{m \, s^{-1}}$ for April and May 2013 using the 20 Hz raw measurements, which is already lower by $3.9 \, \mathrm{m \, s^{-1}}$ and 18% using the 3-second average measurements. The wind gust factor, which is the ratio between the highest wind speed within a 10-minute interval and the 10-minute average wind speed, is therefore underestimated by 0.3 compared to the 20 Hz sampling rate. The seasonal and diurnal cycle of the wind gust profiles within the boundary layer are investigated using the 3-second average measurements of the twelve years (2004–2016) of the weather mast Hamburg. A strong dependence of wind gusts on the incoming solar radiation, causing the generation of turbulence, and on the present stratification is visible. For that reason, the highest wind gust factors of up to 1.8 are found near-ground during the daytime in summer with unstable conditions. The wind gusts are decreasing with height. During winter, the wind gust factors are lower, as well as during the night, where wind gust factors of only up to 1.6 are present.

The effective resolution of an atmospheric model is usually six to ten times larger than the nominal resolution (Skamarock, 2004), which is relevant among others to explicitly resolve wind gusts. The turbulence spectra of all six ICON HH simulations are computed to determine their effective resolutions. The outermost 624 m domain has an effective resolution of only 1763 seconds and 5300 m (factor of eight), but even the 20 m nominally resolved domain has only an effective resolution of about 72 seconds and 220 m (factor eleven). Therefore, the effective resolution is still too coarse to capture all wind gusts of for example the 3-second wind gust definition of the WMO. A nominal horizontal LES resolution of about 1 m would be required to simulate all 3-second wind gusts explicitly, based on a theoretical estimation using the findings of the turbulence spectrum analysis and an assumption of a ten-times-larger effective resolution as of the nominal resolution.

Nevertheless, the 20 m-resolved ICON HH simulations exhibit already a reasonable and similar diurnal cycle of the wind gust profile within the entire boundary layer as of the observations. This is not the case for the 624 m-resolved ICON HH simulations, where almost no wind gusts are directly simulated. Consequently, an added value of the very high-resolution 20 m resolved ICON HH simulations can be identified. However, the wind

gust factors are still underestimated by up to 0.3 near-ground for the 20 m ICON HH simulations. For that reason, a wind gust parameterisation is still required, even for a LES with a horizontal resolution of 20 m to represent the strength of all wind gusts correctly.

Therefore, a new wind gust parameterisation for LES models is developed, that adjusts the turbulence spectrum at the inertial range to the empirical $k^{-5/3}$ Kolmogorov law after the drop from the theoretical -5/3 spectrum. The parameterisation is based on the assumption, that the LES model output already contains all variations of the high-frequent turbulence, which are only damped in their amplitude. Accordingly, the artificially added turbulent kinetic energy to higher frequencies of the turbulence spectrum induces additional variability at the wind speed time series, which represent the underestimated wind gusts. The wind speed time series is recovered by an inverse Fourier transformation.

The parameterisation is applied to the 624 m and 20 m resolved ICON HH model output. As expected, the wind gust factors of the 624 m domain are amplified by the parameterisation, but there is still no realistic diurnal cycle of the wind gusts visible within the boundary layer. Nevertheless the underestimation of the wind gust factors of up to 40% of the non-parameterised ICON HH 624 m output is reduced to only about 17% by the new parameterisation, which is a substantial improvement. The original 20 m ICON HH simulations already seem to resolve the underlying physical processes correctly and the parameterisation further improves the results. Consequently, the maximum wind gust factors of the parametrised output are of about 1.8 like of the observations.

The stated findings indicate a necessary resolution on the order of 1 m to explicitly simulate all wind gusts by an LES. The newly introduced wind gust parameterisation presents promising results, which need to be investigated in more detail.

Overarching Scientific Questions

In summary, the initially stated overall scientific questions in the introduction can be answered as following, based on the novel findings of this study:

1) **How can realistic high-resolution Large Eddy Simulations be evaluated?**

   The proposed evaluation pyramid is well-suited for an overarching assessment of realistic LES to ensure physical consistency, as well as realism for cases with realistic initial and boundary conditions, which are the fundamental prerequisites for every model-based study. Ground-based measurement networks and a classical upper-air verification are appropriate and valuable to analyse the basic atmospheric state of the model. The Cloudnet products and especially the newly developed synthetic cloud classification for atmospheric models provide an ideal basis for an in-depth cloud evaluation. The resolved small-scale turbulence by LES models can be assessed by the turbulence spectrum and wind gusts by the wind gust factor profiles, both requiring appropriate high-resolution observations.

2) **Can an added value be identified for the realistic LES compared to state-of-the-art cloud-resolving models?**

   In terms of the basic atmospheric state, there is currently no added value detected for the high-resolution ICON GER simulations. The errors of the novel and untuned ICON GER model are of a similar magnitude as of the well-established COSMO model. However, the results of the different ICON GER simulations are slightly worse than the COSMO simulations for most parameters. In contrast to the common expectations of a higher accuracy due to the higher resolution, the errors are increasing with resolution of the ICON GER simulations for almost all considered quantities. Nevertheless, regarding clouds and especially small-scale turbulence like wind gusts, a clear added value is identifiable. The forward-simulated observations, as well as the synthetic MC-based cloud classification, are for instance difficult to differentiate from real observations due to the high resolution. Furthermore, at LES resolutions on the order of $10\,\mathrm{m}$, even very small-scale wind gust structures start to become explicitly resolved by the ICON HH simulations, which is not possible at all for a state-of-the-art RANS model.

Outlook

The various analyses and findings of this study raise several new open scientific questions worthy of being investigated in more detail. The found uncertainties of the ICON LES model at the overarching evaluation should be further investigated. These include the possible issues of a presumably too-strong coupling of the model levels during night and too-weak coupling at day, causing among others a substantial overestimation of the 10 m wind speed at night. The possibly too-strong coupling of the lower model levels and resulting overestimated downward momentum flux can be assessed for example by sensitivity studies of the ICON LES. For this, parameters like the turbulence length scale and the diffusivity parameters can be varied. The momentum fluxes can be evaluated for instance by in-situ measurements of boundary layer towers.

Furthermore, the increasing bias and RMSE with higher resolution of the ICON GER simulations, which is in contrast to the common expectations of an increased accuracy by a higher resolution, should be further analysed. For example, a coarse-grained output to 624 m of the ICON GER 156 m could be compared to the 624 m-resolved ICON GER simulations to investigate the effects of the explicitly resolved processes at the different resolutions. The physical consistency and simulation quality of more complex processes like those of the turbulence or interaction of clouds with the radiation would be interesting to analyse. For this, observations of the momentum fluxes, short- and long-wave radiation, and other parameters would be valuable to evaluate. The representation of the hydrological cycle and of individual rain cells should be studied by an extensive precipitation assessment. Further model issues and possibilities to advance the ICON LES could be identified by model intercomparisons of well-defined cases with other realistic LES models like the WRF LES or NICAM in the LES mode.

So far, the comprehensive ICON GER evaluation neglected for instance the winter with snow or fog conditions and other synoptic situations such as severe wind storms. Therefore, long-term simulations of at least one full year and other such simulations would enhance the results and strengthen the robustness. Similarly, the findings of the ICON GER evaluation of this study should be verified and extended to other climate regions as for instance by the already-conducted tropical Atlantic ICON LES runs and by the North Atlantic simulations of the HD(CP)^2 project. The spatial assessment and cloud evaluation in general could be advanced by the incorporation of further satellite products like of the cloud top heights, depth, and additional geometric properties. There are almost infinite possibilities to extend the evaluation, to further improve the ICON LES model, find issues, and ensure physical consistency for future model-based studies. The issues found at the assessment of the basic atmospheric state illustrate the value of an overarching evaluation and should therefore be followed up on.

The correct forward simulation of the linear depolarisation ratio of the virtual cloud radar and detection of the melting layer by the PAMTRA forward operator would be the most important issues to resolve to improve the quality of the MOC-based cloud classification. Simple case studies with idealised conditions of for example only perfectly spherical hydrometeors and another one with non-spherical hydrometeors like ice crystals would help to find the issues at the polarisation of the simulated cloud radar. The consideration of rain hydrometeors at the forward simulation of the lidar by CR-SIM would be of high relevance for an accurate simulation of a physically consistent instrument and should be implemented together with the CR-SIM team. A higher quality of the forward-simulated instruments will help to improve the overall synthetic MOC cloud classification and to detect mixed-phase categories. The abovementioned accurate LDR simulation and detection of the melting layer are the basic prerequisites for capturing mixed-phase categories.

The cloud radar and microwave radiometer could also be simulated by CR-SIM. This would help to find forward operator-related uncertainties in the MOC approach. Nevertheless, the other forward simulated parameters should also be investigated in more detail to optimise them and improve the results for the cloud classification. For instance, idealised and well-defined test cases can be used to find issues at the forward simulations and thus increase their quality. The forward-simulated microwave radiometer could be assessed by clear sky conditions to avoid uncertainties induced by clouds. Additionally, an uncertainty estimate of all forward-simulated quantities would help strengthen the statistical robustness of the findings and to know about the uncertainties at the comparison with real instruments. Again, idealised test case simulations of for instance a prescribed cloud with known reflectivities for the cloud radar, and intercomparisons with other forward operators like CR-SIM or the Atmospheric Radiative Transfer Simulator (ARTS; Buehler et al., 2018), among many others, are valuable to assess the uncertainties of the forward simulations.

The MC-based cloud classification exhibits very promising results, that should be further investigated to use the full potential of this new product. The chosen thresholds for the different hydrometeor mass concentrations at the MC algorithm should be validated for example by comparisons with the MOC approach as soon as this works. Idealised case studies might help to include real instrument's characteristics and improve by this the physical consistency of this approach. The evaluation of parameters like the global radiation or of precipitation can be related to macrophysical cloud properties by the new cloud classification, which helps to put the results into a larger context and find possible issues of the model.

The connected component analysis makes direct use of the new synthetic cloud classification. The derived single cloud objects of the connected component algorithm are well

suited for in-depth cloud studies considering single cloud properties, and their phase. The quality of representation of clouds and precipitation can be assessed by further higher-order properties like their size, lifetime, starting point of rain, and other parameters using the individual cloud objects. The influence of a changing climate on the cloud macrophysical properties such as their phase and size can be examined by the cloud classification using different sensitivity experiments of an atmospheric model. For example, the conducted simulations of Costa-Surós et al. (2019) with different aerosol concentrations of 1985 and of 2013 could be used to investigate the interaction of cloud macrophysics with aerosols. Also, the performed simulations of two times the current cloud condensation nuclei (CCN) concentration and of four times the current CO_2 concentration of the HD(CP)^2 project are valuable datasets to study cloud macrophysical properties under different atmospheric conditions.

The quality and especially physical consistency of the MC-based algorithm with the atmospheric model should be advanced by the consideration of the hydrometeor number concentrations for two-moment cloud microphysical schemes. Intercomparisons of different atmospheric models regarding the representation of cloud macrophysical properties, such as the structure and phase, would be valuable to learn about certain model biases and differences at the simulated clouds. For this, the MC-based cloud classification might be applied and tested for atmospheric models like the WRF, NICAM, and MPI-ESM model. The MC-based approach has already been successfully tested for the COSMO model (Hansen, 2014). The detailed development and shape of the simulated clouds including their phase, such as of large thunderstorms, could be investigated by a possible extension of the MC-based cloud classification to the full four dimensions of an atmospheric model. This would allow for in-depth analyses to advance our understanding of involved processes, such as of cold pools. The observational reference data might be generated by a network of scanning cloud radars to gather an extensive cloud tomography. More advanced products like the single cloud object analysis can be derived from the novel cloud classification to evaluate weather and climate models, and to investigate various cloud processes and properties.

The explicit simulation of small-scale turbulent processes such as of wind gusts in a fully coupled atmospheric model like the ICON LES is one of the major advantages of new realistic LES models, which are worthwhile to investigate in more detail. Nevertheless, a horizontal resolution of 20 m is still not sufficient to resolve all 3-second wind gusts explicitly as previously found at the analysis of the turbulence spectrum of the ICON HH 20 m. Therefore, a further downscaling to a spatial resolution of about 1 m, as proposed by the stated theoretical estimate, would be interesting to investigate, if this resolution is sufficient to resolve all wind gusts of the 3-second definition explicitly. Furthermore,

1 m-resolved ICON LES simulations would explicitly resolve the underlying physical processes, which are worthwhile to analyse to advance our understanding and improve currently applied wind gust parameterisations of today's weather forecast models. The new proposed wind gust parameterisation for LES models based on the turbulence spectrum should be assessed by additional simulations and for other boundary layer towers such as Cabauw and Lindenberg to investigate if this approach is worthwhile being further developed and integrated into a LES model.

Currently, buildings, street canyons and the like are not explicitly resolved by the high-resolution ICON LES model and are only parametrised for instance by a higher surface roughness, even at the conducted 20 m-resolved ICON HH simulations. However, such structures have a large influence on the flow conditions at these high-resolution LES simulations. Furthermore, the wind comfort and thermal ventilation in cities considering the influence of the atmosphere would be very interesting to study under changing climate conditions. Accordingly, an explicit implementation of buildings and of other obstacles would be worthwhile to be integrated in the ICON LES model to enable studies of urban environments.

High-resolution LES models are well suited and already widely used to study wake effects of large wind parks, but most of the time using only small domains and idealised cases. An implementation of an appropriate wind farm scheme into the ICON LES model would be interesting to study the interaction of the turbulence induced by wind farms with the atmosphere and their influence on the weather conditions downstream considering realistic scenarios. For example, Boettcher et al. (2015) investigated the impact of wind farms on the weather using the mesoscale model METRAS (Schlünzen et al., 2018) with a wind farm parameterisation and found changing weather conditions downstream. The turbulence induced by the wind turbines would be explicitly resolved by the ICON LES simulations, which would enable a detailed analysis of the underlying physics causing the stated changes by Boettcher et al. (2015). Furthermore, the ICON LES simulations would extend the results to the microscale. A comparison of the ICON LES model in respect to wind gusts with other LES models commonly used to study wind gusts such as the PALM model or the engineering-oriented OpenFOAM (OpenFOAM, 2019) model would be valuable to identify certain model biases or issues at the simulated turbulence.

Likewise, also more fundamental research on wind gusts such as on the definition of a wind gust for different applications like weather forecasts, wind comfort, and security purposes at construction sites, is of high relevance. The shown example of different sampling rates illustrates the large impact of different definitions and technical instrumentation on wind gust statistics. There are plenty of open research questions in the new field of wind gusts, which are very worthwhile to be investigated.

This study analysed various aspects of the new ICON LES model to show the potential of novel large-domain, realistic large eddy simulations. The ICON LES model explicitly resolve commonly at RANS-models parameterised processes due to the high resolution. The explicit simulation of for instance clouds and small-scale processes avoids uncertainties due to the parameterisations and provides new insights into the atmospheric processes. The new high-resolution model output helps to advance our understanding of the underlying physics and to improve the quality of today's weather and climate predictions.

Appendix

A. Real-color Satellite Images

The real satellite images (left column) are captured by the MODIS instrument on board of the AQUA/TERRA satellites, which are located in the polar orbit. The synthetic satellite images of the ICON GER 156 m simulations (right column) show the total column integrated cloud water for a range of 0.0 to 0.5 $kg\,m^{-2}$. This output is overlayed on top of a cloud-free MODIS scene of 6 May 2018. For each simulated ICON GER day, a real-color satellite image is shown below:

20 April 2013 - 12:55 UTC
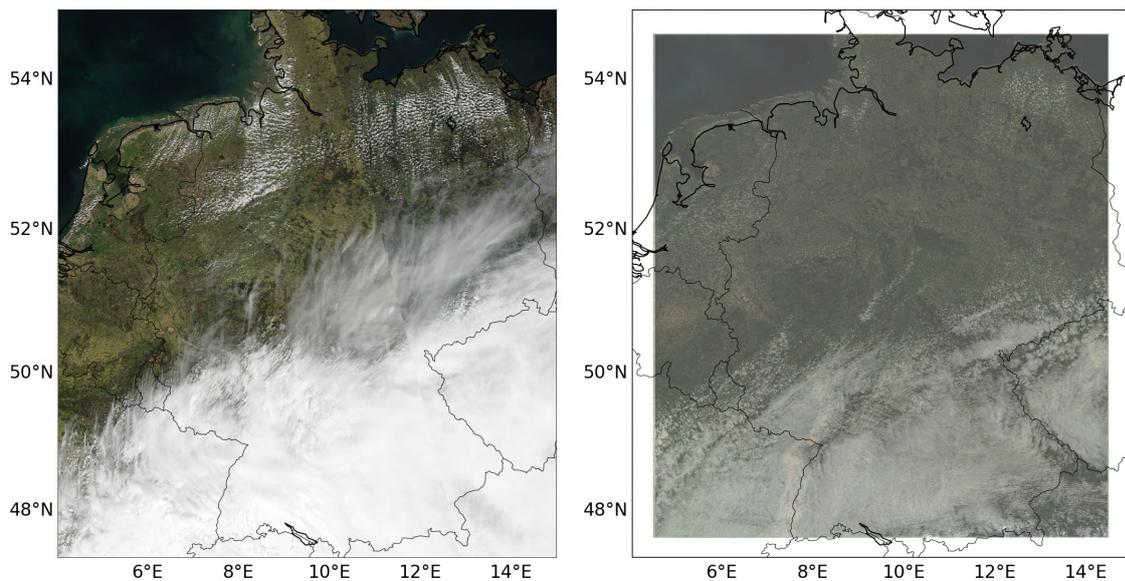
Figure A.1.: Real (left) and synthetic ICON GER 156 m (right) MODIS Satellite image for 20 April 2013 - 12:55 UTC.

24 April 2013 - 12:30 UTC
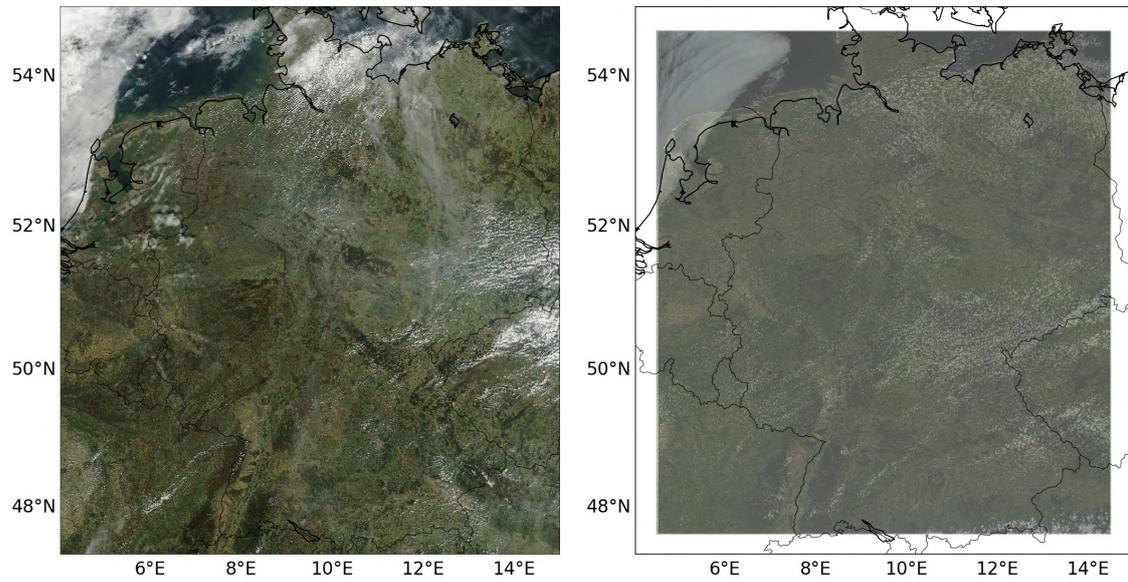
Figure A.2.: As Figure A.1, but for 24 April 2013 - 12:30 UTC.

25 April 2013 - 11:35 UTC
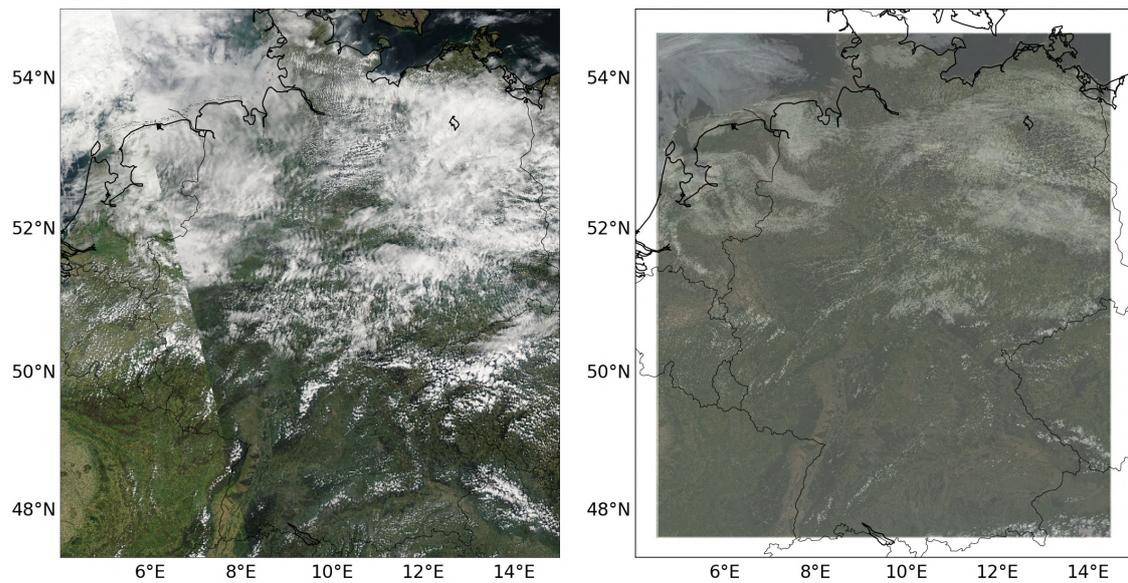
Figure A.3.: As Figure A.1, but for 25 April 2013 - 11:35 UTC.

17 June 2014 - 11:25 UTC
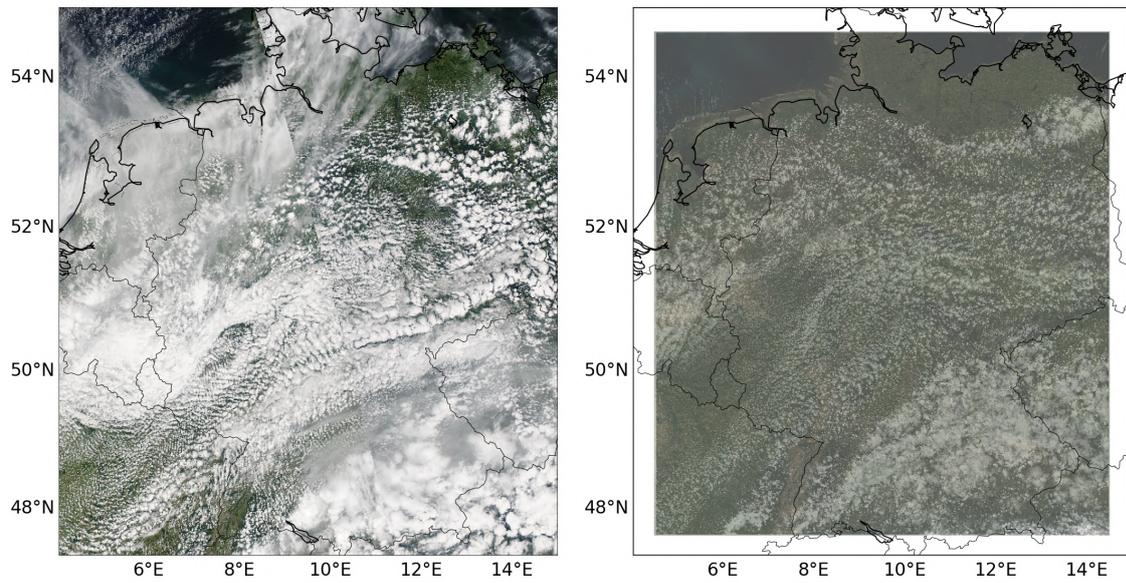
Figure A.4.: As Figure A.1, but for 17 June 2014 - 11:25 UTC.

29 July 2014 - 12:00 UTC
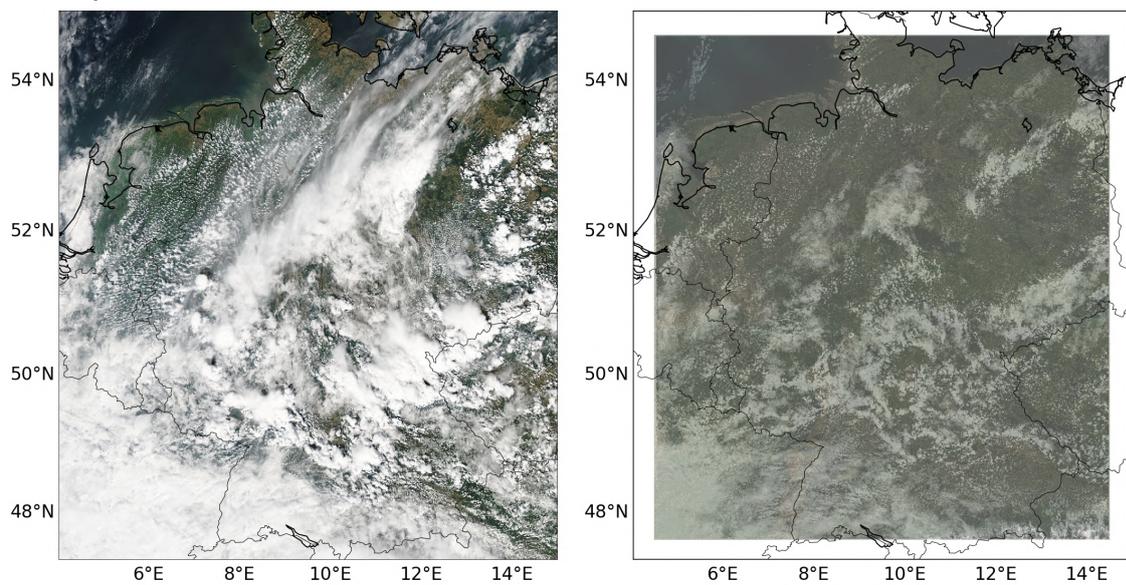
Figure A.5.: As Figure A.1, but for 29 July 2014 - 12:00 UTC.

17 June 2015 - 12:30 UTC (Shifted Domain)
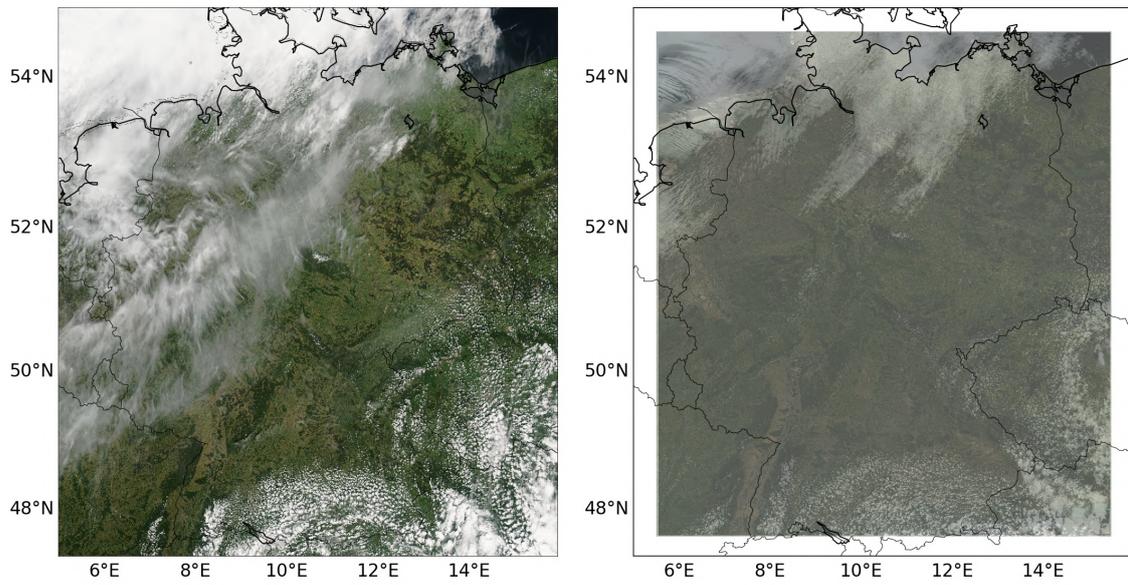
Figure A.6.: As Figure A.1, but for 17 June 2015 - 12:30 UTC.

04 July 2015 - 13:15 UTC (Shifted Domain)
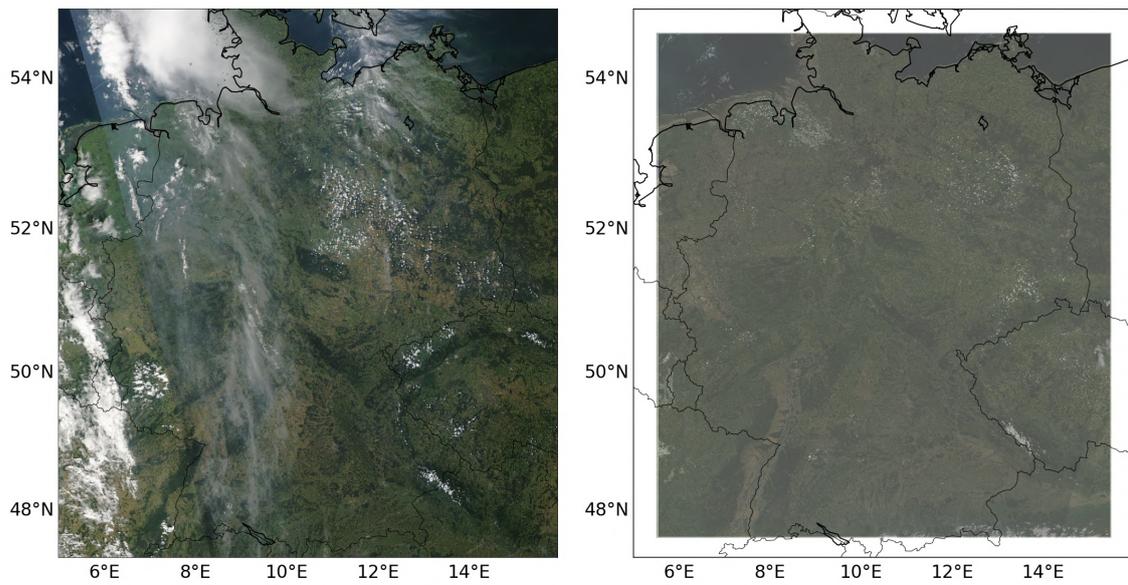
Figure A.7.: As Figure A.1, but for 04 July 2015 - 13:15 UTC.

05 July 2015 - 12:20 UTC (Shifted Domain)
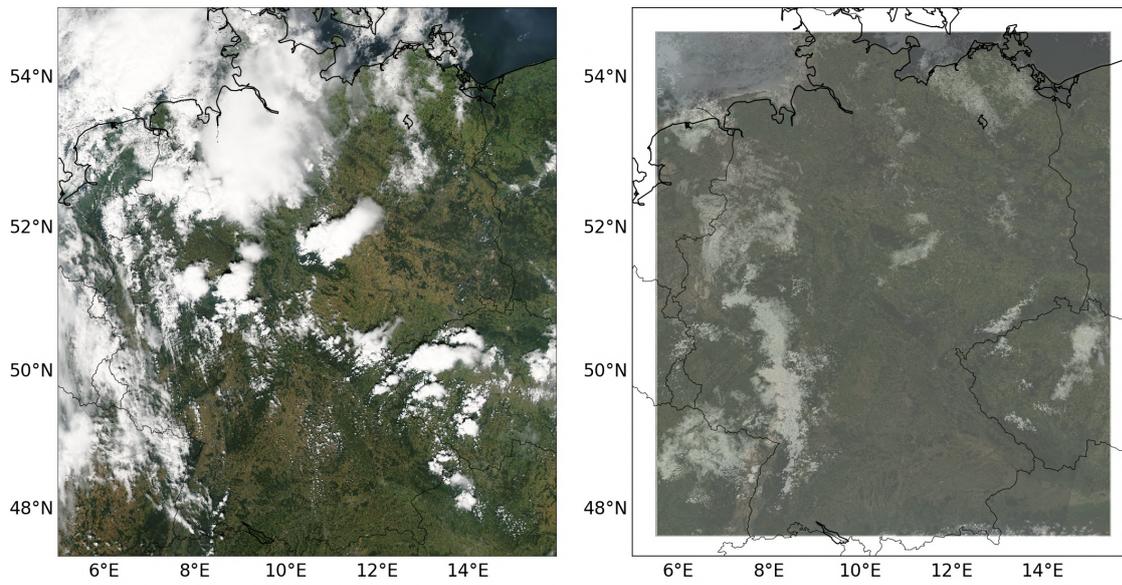
Figure A.8.: As Figure A.1, but for 05 July 2015 - 12:20 UTC.

29 May 2016 - 12:15 UTC
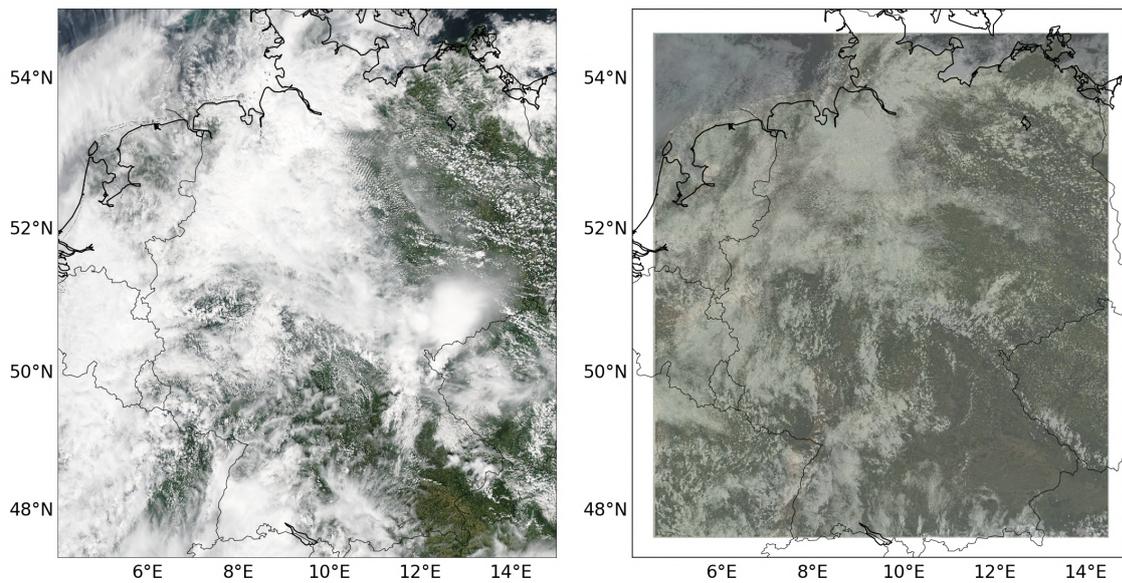
Figure A.9.: As Figure A.1, but for 29 May 2016 - 12:15 UTC.

03 June 2016 - 12:30 UTC
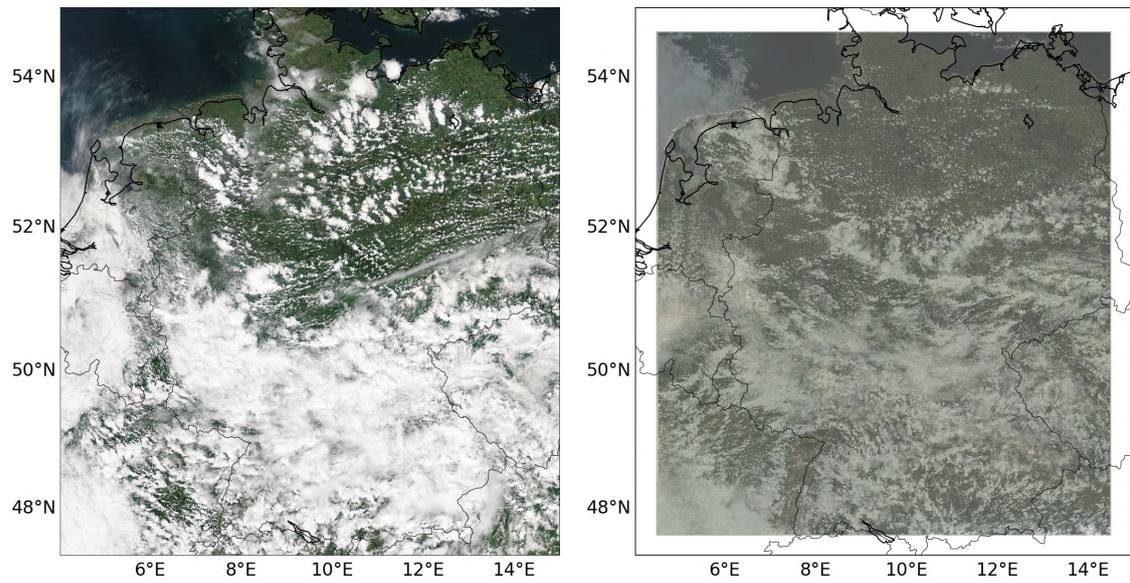
Figure A.10.: As Figure A.1, but for 03 June 2016 - 12:30 UTC.

06 June 2016 - 13:00 UTC
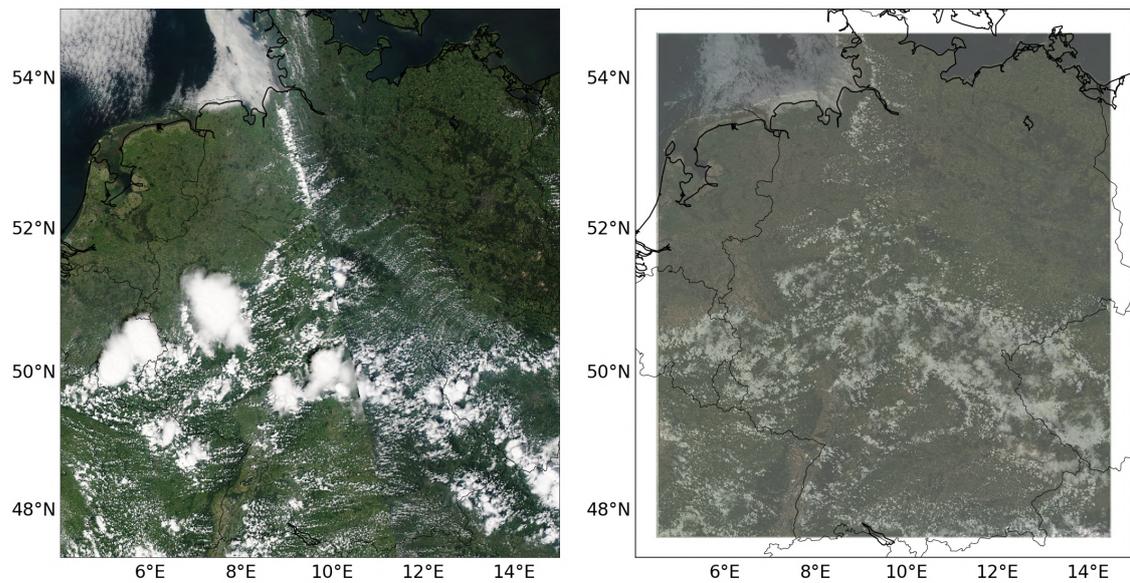
Figure A.11.: As Figure A.1, but for 06 June 2016 - 13:00 UTC.

01 August 2016 - 12:15 UTC
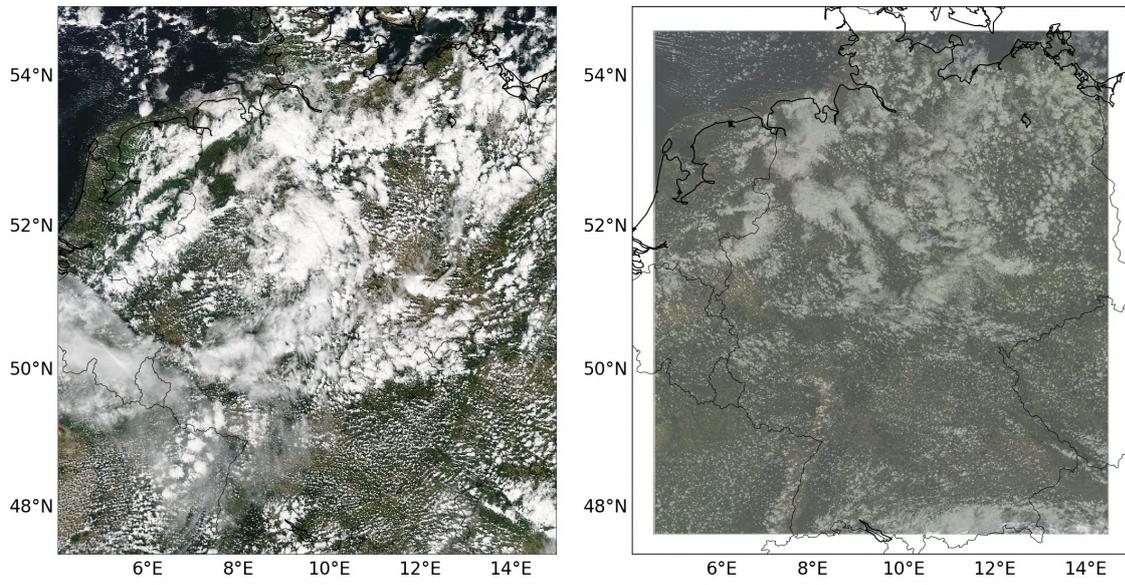
Figure A.12.: As Figure A.1, but for 01 August 2016 - 12:15 UTC.

B. Supplementary Baseline Evaluation Analyses
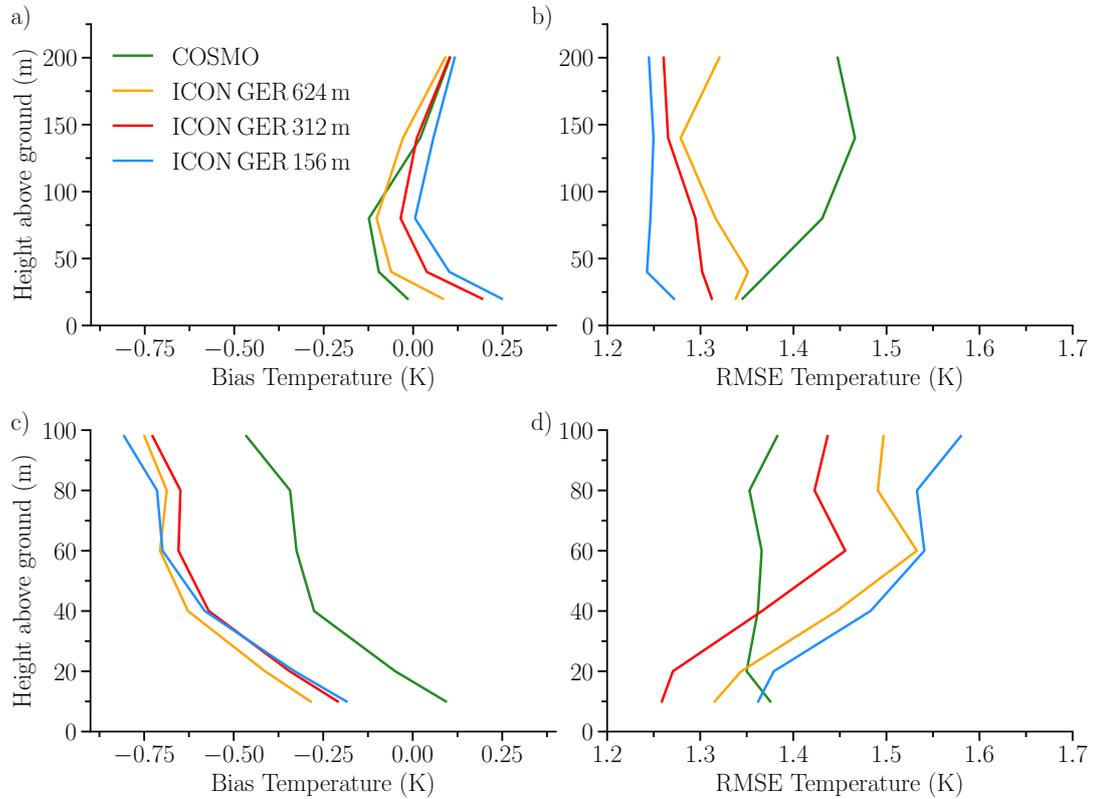
Figure B.1.: Bias and RMSE of the temperature profiles at the boundary layer towers of Cabauw (a,b) and Lindenberg (c,d) using 10-minute average observed data and corresponding ICON GER and COSMO single-column output.

Table B.1.: Selected quantiles of precipitation rates for the DWD RADOLAN observations, ICON GER 624 m, ICON GER 312 m, ICON GER 156 m and COSMO, regridded to the RADOLAN 1 x 1 km grid output using all 12 simulation days.

Quantile	RADOLAN Observation (mm hr^{-1})	ICON GER 624 m (mm hr^{-1})	ICON GER 312 m (mm hr^{-1})	ICON GER 156 m (mm hr^{-1})	COSMO (mm hr^{-1})
0.10	0.12	0.15	0.15	0.15	0.14
0.20	0.24	0.22	0.23	0.23	0.19
0.30	0.27	0.33	0.33	0.33	0.26
0.40	0.48	0.47	0.47	0.48	0.36
0.50	0.72	0.67	0.67	0.69	0.51
0.60	1.08	0.96	0.96	0.99	0.72
0.70	1.68	1.43	1.43	1.48	1.06
0.80	3.12	2.37	2.35	2.44	1.77
0.90	6.72	5.52	5.32	5.47	3.87
0.99	27.72	32.45	30.37	30.37	20.03

Acronyms

cdo	Climate Data Operators
COSMO	Consortium Consortium for Small-scale Modeling
CR-SIM	Cloud Resolving Model Radar Simulator
DWD	German Weather Service
GCM	General Circulation Model
GNSS	Global Navigation Systems Station
HATPRO	Humidity And Temperature PROfilers
HD(CP)^2	High Definition Clouds and Precipitation for Advancing Climate Prediction
HOPE	HD(CP)^2 Observational Prototype Experiment
ICON	ICOsahedral Non-hydrostatic
IWC	Ice Water Content
IWV	Integrated Water Vapour
LACROS	Leipzig Aerosol and Cloud Remote Observations System
LDR	Linear Depolarisation Ratio
LES	Large Eddy Simulation
Lidar	Light detection and ranging
LWC	Liquid Water Content
LWP	Liquid Water Path
MC	Model to Cloud Classification
MOC	Model to Observation to Classification

MODIS	Moderate Image Spectrometer
MSLP	Mean Sea Level Pressure
NWP	Numerical Weather Prediction
p.p.	percentage points
PAMTRA	Passive and Active Microwave TRAnsfer
PBL	Planetary Boundary Layer
PDF	Probability Density Function
Radar	Radio detection and ranging
RADOLAN	Radar Online Calibration
RANS	Reynold Averaged Navier Stokes
RMSE	Root Mean Squared Error
SAMD	Standardized Atmospheric Measurement Database
STD	Standard Deviation
TKE	Turbulent Kinetic Energy
WMO	World Meteorological Organisation
WRF	Weather Research and Forecasting

Bibliography

Ackerman, A. S., M. C. van Zanten, B. Stevens, V. Savic-Jovcic, C. S. Bretherton, A. Chlond, J.-C. Golaz, H. Jiang, M. Khairoutdinov, S. K. Krueger, D. C. Lewellen, A. Lock, C.-H. Moeng, K. Nakamura, M. D. Petters, J. R. Snider, S. Weinbrecht, and M. Zulauf, 2009: Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer. *Monthly Weather Review*, **137**, 1083–1110, doi:10.1175/2008MWR2582.1.

Ahmad, N. H., A. Inagaki, M. Kanda, N. Onodera, and T. Aoki, 2017: Large-Eddy Simulation of the Gust Index in an Urban Area Using the Lattice Boltzmann Method. *Boundary-Layer Meteorology*, **163**, 447–467, doi:10.1007/s10546-017-0233-6.

Akkermans, T., T. Boehme, M. Demuzere, S. Crewell, C. Selbach, T. Reinhardt, A. Seifert, F. Ament, and N. P. M. van Lipzig, 2012: Regime-dependent evaluation of accumulated precipitation in COSMO. *Theoretical and Applied Climatology*, **108**, 39–52, doi:10.1007/s00704-011-0502-0.

Almazroui, M., 2013: Simulation of present and future climate of Saudi Arabia using a regional climate model (PRECIS). *International Journal of Climatology*, **33**, 2247–2259, doi:10.1002/joc.3721.

Angevine, W. M., J. Olson, J. Kenyon, W. I. Gustafson, S. Endo, K. Suselj, and D. D. Turner, 2018: Shallow Cumulus in WRF Parameterizations Evaluated against LASSO Large-Eddy Simulations. *Monthly Weather Review*, **146**, 4303–4322, doi:10.1175/MWR-D-18-0115.1.

Bibliography

Arakawa, A. and V. R. Lamb: 1977, Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model. *General Circulation Models of the Atmosphere*, Julius Chang, ed., Elsevier, volume 17 of *Methods in Computational Physics: Advances in Research and Applications*, 173 – 265.

Azorin-Molina, C., J.-A. Guijarro, T. R. McVicar, S. M. Vicente-Serrano, D. Chen, S. Jerez, and F. Espírito-Santo, 2016: Trends of daily peak wind gusts in Spain and Portugal, 1961–2014. *Journal of Geophysical Research: Atmospheres*, **121**, 1059–1078, doi:10.1002/2015JD024485.

Baldauf, M., J. Förstner, S. Klink, T. Reinhardt, C. Schraff, A. Seifert, and K. Stephan, 2016: Kurze Beschreibung des Lokal-Modells Kürzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD.

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

Bechtold, P., S. K. Krueger, W. S. Lewellen, E. van Meijgaard, C.-H. Moeng, D. A. Randall, A. van Ulden, and S. Wang, 1996: Modeling a Stratocumulus-Topped PBL: Intercomparison among Different One-Dimensional Codes and with Large Eddy Simulation. *Bulletin of the American Meteorological Society*, **77**, 2033–2042, doi:10.1175/1520-0477-77.9.2033.

Beljaars, A. C. M., 1987: The Influence of Sampling and Filtering on Measured Wind Gusts. *Journal of Atmospheric and Oceanic Technology*, **4**, 613–626, doi:10.1175/1520-0426(1987)004¡0613:TIOSAF¿2.0.CO;2.

Blackman, R. B. and J. W. Tukey, 1958: The Measurement of Power Spectra from the Point of View of Communications Engineering — Part I. *Bell System Technical Journal*, **37**, 185–282, doi:10.1002/j.1538-7305.1958.tb03874.x.

Bley, S., H. Deneke, F. Senf, and L. Scheck, 2017: Metrics for the evaluation of warm convective cloud fields in a large-eddy simulation with Meteosat images. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2050–2060, doi:10.1002/qj.3067.

Boers, R., F. Bosveld, H. K. Baltink, W. Knap, E. Van Meijgaard, and W. Wauben, 2019: Observing and Modelling the Surface Radiative Budget and Cloud Radiative Forcing at the Cabauw Experimental Site for Atmospheric Research (CESAR), the Netherlands, 2009-17. *Journal of Climate*, **32**, 7209–7225, doi:10.1175/JCLI-D-18-0828.1.

Boettcher, M., P. Hoffmann, H.-J. Lenhart, K. H. Schlünzen, and R. Schoetter, 2015: Influence of large offshore wind farms on North German climate. *Meteorologische Zeitschrift*, **24**, 465–480, doi:10.1127/metz/2015/0652.

Brazdil, R., J. Hostynek, L. Reznickova, P. Zahradnicek, R. Tolasz, P. Dobrovolny, and P. Stepanek, 2017: The variability of maximum wind gusts in the Czech Republic between 1961 and 2014. *International Journal of Climatology*, **37**, 1961–1978, doi:10.1002/joc.4827.

Brown, A. R., R. T. Cederwall, A. Chlond, P. G. Duynkerke, J.-C. Golaz, M. Khairoutdinov, D. C. Lewellen, A. P. Lock, M. K. MacVean, C.-H. Moeng, R. A. J. Neggers, A. P. Siebesma, and B. Stevens, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, **128**, 1075–1093, doi:10.1256/003590002320373210.

Brown, A. R., S. H. Derbyshire, and P. J. Mason, 1994: Large-eddy simulation of stable atmospheric boundary layers with a revised stochastic subgrid model. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1485–1512, doi:10.1002/qj.49712052004.

Brümmer, B., I. Lange, and H. Konow, 2012: Atmospheric boundary layer measurements at the 280 m high Hamburg weather mast 1995-2011: mean annual and diurnal cycles. *Meteorologische Zeitschrift*, **21**, 319–335, doi:10.1127/0941-2948/2012/0338.

Brune, S., F. Kapp, and P. Friederichs, 2018: A wavelet-based analysis of convective organization in ICON large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, **144**, 2812–2829, doi:10.1002/qj.3409.

Buehler, S. A., J. Mendrok, P. Eriksson, A. Perrin, R. Larsson, and O. Lemke, 2018: ARTS, the Atmospheric Radiative Transfer Simulator – version 2.2, the planetary toolbox edition. *Geoscientific Model Development*, **11**, 1537–1556, doi:10.5194/gmd-11-1537-2018.

Bühl, J., P. Seifert, A. Myagkov, and A. Ansmann, 2016: Measuring ice- and liquid-water properties in mixed-phase cloud layers at the Leipzig Cloudnet station. *Atmospheric Chemistry and Physics*, **16**, 10609–10620, doi:10.5194/acp-16-10609-2016.

Bühl, J., P. Seifert, U. Wandinger, H. Baars, T. Kanitz, J. Schmidt, A. Myagkov, R. Engelmann, A. Fahrwald, B. Heese, A. Klepel, D. Althausen, and A. Ansmann, 2013: LACROS: the Leipzig Aerosol and Cloud Remote Observations System. *Proc SPIE*, **8890**, 02–, doi:10.1117/12.2030911.

Böhme, T., S. Stapelberg, T. Akkermans, S. Crewell, J. Fischer, T. Reinhardt, A. Seifert, C. Selbach, and N. van Lipzig, 2011: Long-term evaluation of COSMO forecasting using combined observational data of the GOP period. *Meteorologische Zeitschrift*, **20, No. 2**, 119–132.

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

Cho, H.-K., K. P. Bowman, and G. R. North, 2004: A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission. *Journal of Applied Meteorology*, **43**, 1586–1597, doi:10.1175/JAM2165.1.

Claussnitzer, A. and P. Nevir, 2009: Analysis of quantitative precipitation forecasts using the Dynamic State Index. *Atmospheric Research*, **94**, 694–703, doi:10.1016/j.atmosres.2009.08.013, General Assembly of the European-Geosciences-Union, Vienna, Austria, Apr 02-07, 2006.

Conradsen, K., L. B. Nielsen, and L. P. Prahm, 1984: Review of Weibull Statistics for Estimation of Wind Speed Distributions. *Journal of Climate and Applied Meteorology*, **23**, 1173–1183, doi:10.1175/1520-0450(1984)023¡1173:ROWSFE¿2.0.CO;2.

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

Costa-Surós, M., O. Sourdeval, C. Acquistapace, H. Baars, C. Carbajal Henken, C. Genz, J. Hesemann, C. Jimenez, M. König, J. Kretzschmar, N. Madenach, C. I. Meyer, R. Schrödner, P. Seifert, F. Senf, M. Brueck, G. Cioni, J. F. Engels, K. Fieg, K. Gorges, R. Heinze, P. K. Siligam, U. Burkhardt, S. Crewell, C. Hoose, A. Seifert, I. Tegen, and J. Quaas, 2019: Detection and attribution of aerosol-cloud interactions in large-domain large-eddy simulations with ICON. *Atmospheric Chemistry and Physics Discussions*, **2019**, 1–29, doi:10.5194/acp-2019-850.

Craig F. Bohren, D. R. H., 1998: *Absorption and Scattering of Light by Small Particles*. WILEY-VCH Verlag GmbH & Co. KGaA.

Crewell, S., U. Löhnert, and C. Simmer, 2010: Mikrowellensradiometrie für Wasserdampf- und Wolkenbeobachtungen. *promet*, **36. Jahrgang, Heft 3/4**, 109–118.

Crosman, E. T. and J. D. Horel, 2017: Large-eddy simulations of a Salt Lake Valley cold-air pool. *Atmospheric Research*, **193**, 10–25, doi:10.1016/j.atmosres.2017.04.010.

Cui, C., Y. Bao, C. Yuan, Z. Li, and C. Zong, 2019: Comparison of the performances between the WRF and WRF-LES models in radiation fog - A case study. *Atmospheric Research*, **226**, 76–86, doi:10.1016/j.atmosres.2019.04.003.

Deardorff, J. W., 1970: A numerical study of three-dimensional turbulent channel flow at large Reynolds numbers. *Journal of Fluid Mechanics*, **41**, 453–480, doi:10.1017/S0022112070000691.

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

Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, **15**, 51–64, doi:10.1002/met.25.

Evans, K. F. and G. L. Stephens, 1995: Microwave Radiative Transfer through Clouds Composed of Realistically Shaped Ice Crystals. Part I. Single Scattering Properties. *Journal of the Atmospheric Sciences*, **52**, 2041–2057, doi:10.1175/1520-0469(1995)052¡2041:MRTTCC¿2.0.CO;2.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9**, 1937–1958, doi:10.5194/gmd-9-1937-2016.

Fiorio, C. and J. Gustedt, 1996: Two linear time Union-Find strategies for image processing. *Theoretical Computer Science*, **154**, 165 – 181, doi:https://doi.org/10.1016/0304-3975(94)00262-2.

Fischer, R., I. Bastigkeit, B. Leitl, and M. Schatzmann, 2010: Generation of spatio-temporally high resolved datasets for the validation of LES-models simulating flow and dispersion phenomena within the lower atmospheric boundary layer. *The Fifth International Symposium on Computational Wind Engineering (CWE2010), Chapel Hill, North Carolina, USA May 23-27*.

Förster, E., 2014: Empirische Untersuchung der Änderungen von Mittelwind und Böen mit der Höhe in der Prandtl-Schicht. Bachelor's thesis.

Galvin, J., 2016: *An introduction to the meteorology and climate of the tropics*. 234 pp.

Gendt, G., G. Dick, C. Reigber, M. Tomassini, Y. Liu, and M. Ramatschi, 2004: Near Real Time GPS Water Vapor Monitoring for Numerical Weather Prediction in Germany. *Journal of the Meteorological Society of Japan. Ser. II*, **82**, 361–370, doi:10.2151/jmsj.2004.361.

Gettelman, A., P. Callaghan, V. E. Larson, C. M. Zarzycki, J. T. Bacmeister, P. H. Lauritzen, P. A. Bogenschutz, and R. B. Neale, 2018: Regional Climate Simulations With

the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, **10**, 1245–1265, doi:10.1002/2017MS001227.

Giorgetta, M. A., J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, **5**, 572–597, doi:10.1002/jame.20038.

Giorgi, F., 1990: Simulation of Regional Climate Using a Limited Area Model Nested in a General Circulation Model. *Journal of Climate*, **3**, 941–963, doi:10.1175/1520-0442(1990)003¡0941:SORCUA¿2.0.CO;2.

Giorgi, F. and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: A review. *Reviews of Geophysics*, **29**, 191–216, doi:10.1029/90RG02636.

Griewank, P. J., V. Schemann, and R. A. J. Neggers, 2018: Evaluating and Improving a PDF Cloud Scheme Using High-Resolution Super Large Domain Simulations. *Journal of Advances in Modeling Earth Systems*, **10**, 2245–2268, doi:10.1029/2018MS001421.

Gronemeier, T., S. Raasch, and E. Ng, 2017: Effects of Unstable Stratification on Ventilation in Hong Kong. *Atmosphere*, **8**, doi:10.3390/atmos8090168.

Gronemeier, T. and M. Sühring, 2019: On the Effects of Lateral Openings on Courtyard Ventilation and Pollution—A Large-Eddy Simulation Study. *Atmosphere*, **10**, doi:10.3390/atmos10020063.

Hanley, K. E., R. S. Plant, T. H. M. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, **141**, 272–284, doi:10.1002/qj.2356.

Hansen, A., 2014: *Evaluation of the COSMO-DE Ensemble Prediction System using the HOPE measurements of HD(CP)^2*. Master's thesis, Meteorological Institute, University Hamburg.

Heath, N. K., H. E. Fuelberg, S. Tanelli, F. J. Turk, R. P. Lawson, S. Woods, and S. Freeman, 2017: WRF nested large-eddy simulations of deep convection during SEAC4RS. *Journal of Geophysical Research: Atmospheres*, **122**, 3953–3974, doi:10.1002/2016JD025465.

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

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

Held, I., H. Guo, A. Adcroft, J. Dunne, L. Horowitz, J. Krasting, E. Shevliakova, M. Winton, M. Zhao, M. Bushuk, A. Wittenberg, B. Wyman, B. Xiang, R. Zhang, W. Anderson, V. Balaji, L. Donner, K. Dunne, J. Durachta, P. Gauthier, P. Ginoux, J.-C. Golaz, S. Griffies, R. Hallberg, L. Harris, M. Harrison, W. Hurlin, J. John, P. Lin, S.-J. Lin, S. Malyshev, R. Menzel, P. Milly, Y. Ming, V. Naik, D. Paynter, F. Paulot, V. Ramaswamy, B. Reichl, T. Robinson, A. Rosati, C. Seman, L. Silvers, S. Underwood, and N. Zadeh, 2019: Structure and Performance of GFDL's CM4.0 Climate Model. *Journal of Advances in Modeling Earth Systems*, doi:10.1029/2019MS001829.

Hennemuth, B. and A. Lammert, 2006: Determination of the Atmospheric Boundary Layer Height from Radiosonde and Lidar Backscatter. *Boundary-Layer Meteorology*, **120**, 181–200, doi:10.1007/s10546-005-9035-3.

Hertwig, D., G. Patnaik, and B. Leitl, 2017: LES validation of urban flow, part I: flow statistics and frequency distributions. *Environmental Fluid Mechanics*, **17**, 521–550, doi:10.1007/s10652-016-9507-7.

Heus, T., C. C. van Heerwaarden, H. J. J. Jonker, A. Pier Siebesma, S. Axelsen, K. van den Dries, O. Geoffroy, A. F. Moene, D. Pino, S. R. de Roode, and J. Vilà-Guerau de Arellano, 2010: Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geoscientific Model Development*, **3**, 415–444, doi:10.5194/gmd-3-415-2010.

Hogan, R. J. and E. J. O'Connor, 2006: Facilitating cloud radar and lidar algorithms: The

Cloudnet Instrument Synergy/Target Categorization product. Cloudnet documentation. *Available online at www.cloud-net.org/data/products/categorize.html*.

Hogan, R. J., E. J. O'Connor, and A. J. Illingworth, 2009: Verification of cloud-fraction forecasts. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1494–1511, doi:10.1002/qj.481.

Hohenegger, C., P. Brockhaus, and C. Schär, 2008: Towards climate simulations at cloud-resolving scales. *Meteorologische Zeitschrift*, **17**, 383–394, doi:10.1127/0941-2948/2008/0303.

Holton, J. R., 2004: *An introduction to dynamic meteorology*. International Geophysics Series, Elsevier Academic Press,, 4 edition, 535 pp.

Hughes, J. K., A. N. Ross, S. B. Vosper, A. P. Lock, and B. C. Jemmett-Smith, 2015: Assessment of valley cold pools and clouds in a very high-resolution numerical weather prediction model. *Geoscientific Model Development*, **8**, 3105–3117, doi:10.5194/gmd-8-3105-2015.

Ikegaya, N., Y. Ikeda, A. Hagishima, and J. Tanimoto, 2017: Evaluation of rare velocity at a pedestrian level due to turbulence in a neutrally stable shear flow over simplified urban arrays. *Journal of Wind Engineering and Industrial Aerodynamics*, **171**, 137–147, doi:10.1016/j.jweia.2017.10.002.

Illingworth, A. J., D. Cimini, C. Gaffard, M. Haeffelin, V. Lehmann, U. Löhnert, E. J. O'Connor, and D. Ruffieux, 2015: Exploiting Existing Ground-Based Remote Sensing Networks to Improve High-Resolution Weather Forecasts. *Bulletin of the American Meteorological Society*, **96**, 2107–2125, doi:10.1175/BAMS-D-13-00283.1.

Illingworth, A. J. and CloudNet-Team, 2004: Comparison of observed cloud properties at three ground based sites with their representation in operational models: the EU Cloudnet project. *Proc. 14th Int. Conf. on Clouds and Precipitation, Bologna, Italy*.

Illingworth, A. J., R. J. Hogan, E. J. O'Connor, D. Bouniol, M. E. Brooks, J. Delanoe, J. D. E. D. P. Donovan, N. Gaussiat, J. W. F. Goddard, M. Haeffelin, H. K. Baltink, O. A. Krasnov, J. Pelon, J.-M. Piriou, A. Protat, H. W. J. Russchenberg, A. Seifert, A. M. Tompkins, G.-J. V. Zadelhoff, F. Vinit, U. Willén, D. R. Wilson, and C. L. Wrench, 2007: CLOUDNET - Continuous Evaluation of Cloud Profiles in Seven Operational Models Using Ground-Based Observations. *Bulletin of the American Meteorological Society*, **June 2007**, 1–16, doi:10.1175/BAMS-88-6-xxx.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate

Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*, 1535 pp.

Jankov, I., L. D. Grasso, M. Sengupta, P. J. Neiman, D. Zupanski, M. Zupanski, D. Lindsey, D. W. Hillger, D. L. Birkenheuer, R. Brummer, and H. Yuan, 2011: An Evaluation of Five ARW-WRF Microphysics Schemes Using Synthetic GOES Imagery for an Atmospheric River Event Affecting the California Coast. *Journal of Hydrometeorology*, **12**, 618–633, doi:10.1175/2010JHM1282.1.

Jia, W., G. Xue-Jie, X. Yin-Long, and P. Jie, 2015: Regional Climate Change and Uncertainty Analysis based on Four Regional Climate Model Simulations over China. *Atmospheric and Oceanic Science Letters*, **8**, 147–152, doi:10.3878/AOSL20150013.

Jones, R. G., J. M. Murphy, and M. Noguer, 1995: Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Quarterly Journal of the Royal Meteorological Society*, **121**, 1413–1449, doi:10.1002/qj.49712152610.

Kaimal, J. C. and J. C. Wyngaard, 1990: The Kansas and Minnesota experiments. *Boundary-Layer Meteorology*, **50**, 31–47, doi:10.1007/BF00120517.

Kalverla, P., G.-J. Steeneveld, R. Ronda, and A. A. M. Holtslag, 2019: Evaluation of three mainstream numerical weather prediction models with observations from meteorological mast IJmuiden at the North Sea. *Wind Energy*, **22**, 34–48, doi:10.1002/we.2267.

Karstens, U., C. Simmer, and E. Ruprecht, 1994: Remote sensing of cloud liquid water. *Meteorology and Atmospheric Physics*, **54**, doi:10.1007/BF01030057.

Knigge, C. and S. Raasch, 2016: Improvement and development of one- and two-dimensional discrete gust models using a large-eddy simulation model. *Journal of Wind Engineering and Industrial Aerodynamics*, **153**, 46 – 59, doi:https://doi.org/10.1016/j.jweia.2016.03.004.

Knoop, H., F. Ament, and B. Maronga, 2019: A generic gust definition and detection method based on wavelet-analysis. *Advances in Science and Research*, **16**, 143–148, doi:10.5194/asr-16-143-2019.

Kollias, P., J. Rémillard, E. Luke, and W. Szyrmer, 2011: Cloud radar Doppler spectra in drizzling stratiform clouds: 1. Forward modeling and remote sensing applications. *Journal of Geophysical Research: Atmospheres*, **116**, doi:10.1029/2010JD015237.

Kristensen, L., M. Casanova, M. S. Courtney, and I. Troen, 1991: In search of a gust definition. *Boundary-Layer Meteorology*, **55**, 91–107, doi:10.1007/BF00119328.

Lammert, A., A. Hansen, F. Ament, S. Crewell, G. Dick, V. Grützun, H. Klein-Baltink, V. Lehmann, A. Macke, B. Pospichal, W. Schubotz, P. Seifert, E. Stamnas, and B. Stevens, 2019: A Standardized Atmospheric Measurement Data Archive for Distributed Cloud and Precipitation Process-Oriented Observations in Central Europe. *Bulletin of the American Meteorological Society*, **100**, 1299–1314, doi:10.1175/BAMS-D-18-0174.1.

Letzel, M. O., C. Helmke, E. Ng, X. An, A. Lai, and S. Raasch, 2012: LES case study on pedestrian level ventilation in two neighbourhoods in Hong Kong. *Meteorologische Zeitschrift*, **21**, 575–589, doi:10.1127/0941-2948/2012/0356.

Letzel, M. O., M. Krane, and S. Raasch, 2008: High resolution urban large-eddy simulation studies from street canyon to neighbourhood scale. *Atmospheric Environment*, **42**, 8770 – 8784, doi:https://doi.org/10.1016/j.atmosenv.2008.08.001.

Lilly, D. K., 1966: On the application of the eddy viscosity concept in the inertial sub-range of turbulence. *NCAR Tech. Rep. 123*, 22 pp., doi:10.5065/D67H1GGQ.

Liu, Z., S. Cao, H. Liu, and T. Ishihara, 2019: Large-Eddy Simulations of the Flow Over an Isolated Three-Dimensional Hill. *Boundary-Layer Meteorology*, **170**, 415–441, doi:10.1007/s10546-018-0410-2.

Louis, J.-F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, **17**, 187–202, doi:10.1007/BF00117978.

Love, B. S., A. J. Matthews, and G. M. S. Lister, 2011: The diurnal cycle of precipitation over the Maritime Continent in a high-resolution atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, **137**, 934–947, doi:10.1002/qj.809.

Lufft, 2019: *Betriebsanleitung - Lufft CHM 15k Ceilometer*. G. Lufft Mess- und Regeltechnik GmbH.

Maahn, M., U. Löhnert, P. Kollias, R. C. Jackson, and G. M. McFarquhar, 2015: Developing and Evaluating Ice Cloud Parameterizations for Forward Modeling of Radar Moments Using in situ Aircraft Observations. *Journal of Atmospheric and Oceanic Technology*, **32**, 880–903, doi:10.1175/JTECH-D-14-00112.1.

Macke, A., P. Seifert, H. Baars, C. Barthlott, C. Beekmans, A. Behrendt, B. Bohn, M. Brueck, J. Bühl, S. Crewell, T. Damian, H. Deneke, S. Düsing, A. Foth, P. Di Girolamo, E. Hammann, R. Heinze, A. Hirsikko, J. Kalisch, N. Kalthoff, S. Kinne, M. Kohler, U. Löhnert, B. L. Madhavan, V. Maurer, S. K. Muppa, J. Schween, I. Serikov, H. Siebert, C. Simmer, F. Späth, S. Steinke, K. Träumner, S. Trömel, B. Wehner, A. Wieser, V. Wulfmeyer, and X. Xie, 2017: The HD(CP)^2 Observational Prototype

Experiment (HOPE) – an overview. *Atmospheric Chemistry and Physics*, **17**, 4887–4914, doi:10.5194/acp-17-4887-2017.

Madala, S., A. Satyanarayana, and T. N. Rao, 2014: Performance evaluation of PBL and cumulus parameterization schemes of WRF ARW model in simulating severe thunderstorm events over Gadanki MST radar facility — Case study. *Atmospheric Research*, **139**, 1 – 17, doi:https://doi.org/10.1016/j.atmosres.2013.12.017.

Marke, T., S. Crewell, V. Schemann, J. H. Schween, and M. Tuononen, 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**, 1155–1170, doi:10.1175/JAMC-D-17-0341.1.

Maronga, B., M. Gryschka, R. Heinze, F. Hoffmann, F. Kanani-Sühring, M. Keck, K. Ketelsen, M. O. Letzel, M. Sühring, and S. Raasch, 2015: The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives. *Geoscientific Model Development*, **8**, 2515–2551, doi:10.5194/gmd-8-2515-2015.

Mashiko, W., 2019: A Statistical Study of Wind Gusts in Japan Using Surface Observations. *Journal of the Meteorological Society of Japan. Ser. II*, **97**, 39–54, doi:10.2151/jmsj.2019-001.

METEK, 2013: *Cloud Radar MIRA-35*. METEK Meteorologische Messtechnik GmbH, Fritz-Strassmann-Str. 4, 25337 Elmshorn, Germany.

Mishchenko, M. I., 2000: Calculation of the amplitude matrix for a nonspherical particle in a fixed orientation. *Applied Optics*, **39**, 1026–1031, doi:10.1364/AO.39.001026.

Miyamoto, Y., Y. Kajikawa, R. Yoshida, T. Yamaura, H. Yashiro, and H. Tomita, 2013: Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, **40**, 4922–4926, doi:10.1002/grl.50944.

Moeng, C.-H., M. A. LeMone, M. F. Khairoutdinov, S. K. Krueger, P. A. Bogenschutz, and D. A. Randall, 2009: The Tropical Marine Boundary Layer Under a Deep Convection System: a Large-Eddy Simulation Study. *Journal of Advances in Modeling Earth Systems*, **1**, doi:10.3894/JAMES.2009.1.16.

Moya-Alvarez, A. S., J. Galvez, A. Holguin, R. Estevan, S. Kumar, E. Villalobos, D. Martinez-Castro, and Y. Silva, 2018: Extreme Rainfall Forecast with the WRF-ARW Model in the Central Andes of Peru. *Atmosphere*, **9**, doi:10.3390/atmos9090362.

NASA, 2019: NASA EOSDIS WorldView. Web: https://worldview.earthdata.nasa.gov, accessed: 24/08/2019.

Bibliography

Neggers, R. A. J., A. P. Siebesma, and T. Heus, 2012: Continuous Single-Column Model Evaluation at a Permanent Meteorological Supersite. *Bulletin of the American Meteorological Society*, **93**, 1389–1400, doi:10.1175/BAMS-D-11-00162.1.

Nielsen, K. P. and E. Gleeson, 2018: Using Shortwave Radiation to Evaluate the HARMONIE-AROME Weather Model. *Atmosphere*, **9**, doi:10.3390/atmos9050163.

NOAA, 1998: National Oceanic and Atmospheric Administration - Training Guide in Surface Weather Observations. **May**, 2–1.

Nomokonova, T., K. Ebell, U. Loehnert, M. Maturilli, C. Ritter, and E. O'Connor, 2019: Statistics on clouds and their relation to thermodynamic conditions at Ny-Alesund using ground-based sensor synergy. *Atmospheric Chemistry and Physics*, **19**, 4105–4126, doi:10.5194/acp-19-4105-2019.

Nürenberg, H., 2018: *Evaluierung des Niederschlags aus ICON-LES mit Radarmessungen*. Master's thesis, Meteorological Institute, University Hamburg.

OpenFOAM, 2019: *OpenFOAM - The Open Source CFD Toolbox*. Version v1906 edition.

Pantillon, F., B. Adler, U. Corsmeier, P. Knippertz, A. Wieser, and A. Hansen, 2020: Formation of Wind Gusts in an Extratropical Cyclone in Light of Doppler Lidar Observations and Large-Eddy Simulations. *Monthly Weather Review*, **148**, 353–375, doi:10.1175/MWR-D-19-0241.1.

Patnaik, G., J. Boris, M.-Y. Lee, T. Young, B. Leitl, F. Harms, and M. Schatzmann, 2009: Validation of an LES urban aerodynamics model with model and application specific wind tunnel data. *7th Asia-Pacific Conference on Wind Engineering, APCWE-VII*.

Pfeifer, M., W. Yen, M. Baldauf, G. Craig, S. Crewell, J. Fischer, M. Hagen, A. Hühnerbein, M. Mech, T. Reinhardt, M. Schröder, and A. Seifert, 2010: Validating precipitation forecasts using remote sensor synergy: A case study approach. *Meteorologische Zeitschrift*, **19, No. 6**, 601–617, doi:10.1127/0941-2948/2010/0487.

Pohlmann, H., W. A. Müller, M. Bittner, S. Hettrich, K. Modali, K. Pankatz, and J. Marotzke, 2019: Realistic Quasi-Biennial Oscillation Variability in Historical and Decadal Hindcast Simulations Using CMIP6 Forcing. *Geophysical Research Letters*, **46**, 14118–14125, doi:10.1029/2019GL084878.

Pressel, K. G., S. Mishra, T. Schneider, C. M. Kaul, and Z. Tan, 2017: Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *Journal of Advances in Modeling Earth Systems*, **9**, 1342–1365, doi:10.1002/2016MS000778.

Pscheidt, I., F. Senf, R. Heinze, H. Deneke, S. Troemel, and C. Hohenegger, 2019: How organized is deep convection over Germany? *Quarterly Journal of the Royal Meteorological Society*, **145**, 2366–2384, doi:10.1002/qj.3552.

Randall, D., S. Krueger, C. Bretherton, J. Curry, P. Duynkerke, M. Moncrieff, B. Ryan, D. Starr, M. Miller, W. Rossow, G. Tselioudis, and B. Wielicki, 2003: Confronting Models with Data: The GEWEX Cloud Systems Study. *Bulletin of the American Meteorological Society*, **84**, 455–470.

Reinert, D., F. Prill, H. Frank, M. Denhard, and G. Zängl, 2019: Database Reference Manual for ICON and ICON-EPS. *Research and Development at DWD*, **1.2.11**.

Richardson, H., S. Basu, and A. A. M. Holtslag, 2013: Improving Stable Boundary-Layer Height Estimation Using a Stability-Dependent Critical Bulk Richardson Number. *Boundary-Layer Meteorology*, **148**, 93–109, doi:10.1007/s10546-013-9812-3.

Rinehart, R. E., 2010: *Radar for Meteorologists*. Rinehart Publishing, 5 edition.

RPG, 2011: *Technical Instrument Manual of HATPRO, LHATPRO, TEMPRO, HUMPRO, LHUMPRO, LWP, LWP-U90, LWP-U72-82, LWP-90-150, Tau-225, Tau-225-350*. Radiometer Physics GmbH.

Sallis, P., W. Claster, and S. Hernández, 2011: A machine-learning algorithm for wind gust prediction. *Computers & Geosciences*, **37**, 1337 – 1344, doi:https://doi.org/10.1016/j.cageo.2011.03.004.

Satoh, M., H. Tomita, H. Yashiro, H. Miura, C. Kodama, T. Seiki, A. T. Noda, Y. Yamada, D. Goto, M. Sawada, T. Miyoshi, Y. Niwa, M. Hara, T. Ohno, S.-i. Iga, T. Arakawa, T. Inoue, and H. Kubokawa, 2014: The Non-hydrostatic Icosahedral Atmospheric Model: description and development. *Progress in Earth and Planetary Science*, **1**, 18, doi:10.1186/s40645-014-0018-1.

Schalkwijk, J., H. Jonker, A. Siebesma, and E. Meijgaard, 2015: Weather Forecasting Using GPU-Based Large-Eddy Simulations. *Bulletin of the American Meteorological Society*, **96**, 715–723, doi:10.1175/BAMS-D-14-00114.1.

Schatzmann, M. and B. Leitl, 2011: Issues with validation of urban flow and dispersion CFD models. *Journal of Wind Engineering and Industrial Aerodynamics*, **99**, 169 – 186, doi:https://doi.org/10.1016/j.jweia.2011.01.005, The Fifth International Symposium on Computational Wind Engineering.

Schlemmer, L. and C. Hohenegger, 2014: The Formation of Wider and Deeper Clouds as a Result of Cold-Pool Dynamics. *Journal of the Atmospheric Sciences*, **71**, 2842–2858, doi:10.1175/JAS-D-13-0170.1.

Schlünzen, K. H., M. Boettcher, B. H. Fock, A. Gierisch, D. Grawe, and M. Salim, 2018: Scientific Documentation of the Multiscale Model System M-SYS (METRAS, MITRAS, MECTM, MICTM, MESIM). techreport, Meteorologisches Institut, Centrum für Erdsystemforschung und Nachhaltigkeit, CEN, Universität Hamburg.

Schulzweida, U., 2019: *CDO User Guide*. Max Planck Institute for Meteorology, Version 1.9.6 edition.

Seibert, P., F. Beyrich, S.-E. Gryning, S. Joffre, A. Rasmussen, and P. Tercier, 2000: Review and intercomparison of operational methods for the determination of the mixing height. *Atmospheric Environment*, **34**, 1001 – 1027, doi:https://doi.org/10.1016/S1352-2310(99)00349-0.

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

Seregina, L. S., R. Haas, K. Born, and J. G. Pinto, 2014: Development of a wind gust model to estimate gust speeds and their return periods. *Tellus A: Dynamic Meteorology and Oceanography*, **66**, 22905, doi:10.3402/tellusa.v66.22905.

Sheridan, P.: 2017, Current gust forecasting techniques, developments and challenges. *European Meteorological Society Conference UP1.4*.

Siebesma, A. P., C. S. Bretherton, A. Brown, A. Chlond, J. Cuxart, P. G. Duynkerke, H. Jiang, M. Khairoutdinov, D. Lewellen, C.-H. Moeng, E. Sanchez, B. Stevens, and D. E. Stevens, 2003: A Large Eddy Simulation Intercomparison Study of Shallow Cumulus Convection. *Journal of the Atmospheric Sciences*, **60**, 1201–1219, doi:10.1175/1520-0469(2003)60¡1201:ALESIS¿2.0.CO;2.

Skamarock, W. C., 2004: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra. *Monthly Weather Review*, **132**, 3019–3032, doi:10.1175/MWR2830.1.

Skamarock, W. C.: 2011, *Kinetic Energy Spectra and Model Filters*, Springer Berlin Heidelberg. 495–512.

Skyllingstad, E. D. and S. P. de Szoeke, 2015: Cloud-Resolving Large-Eddy Simulation of Tropical Convective Development and Surface Fluxes. *Monthly Weather Review*, **143**, 2441–2458, doi:10.1175/MWR-D-14-00247.1.

Smagorinsky, J., 1963: General Circulation Experiments with the Primitive Equations. *Monthly Weather Review*, **91**, 99–164, doi:10.1175/1520-0493(1963)091¡0099:GCEWTP¿2.3.CO;2.

Smith, P. L., 1984: Equivalent Radar Reflectivity Factors for Snow and Ice Particles. *Journal of Climate and Applied Meteorology*, **23**, 1258–1260, doi:10.1175/1520-0450(1984)023¡1258:ERRFFS¿2.0.CO;2.

Stein, T. H. M., R. J. Hogan, P. A. Clark, C. E. Halliwell, K. E. Hanley, H. W. Lean, J. C. Nicol, and R. S. Plant, 2015: The DYMECS Project: A Statistical Approach for the Evaluation of Convective Storms in High-Resolution NWP Models. *Bulletin of the American Meteorological Society*, **96**, 939–951, doi:10.1175/BAMS-D-13-00279.1.

Stevens, B., A. S. Ackerman, B. A. Albrecht, A. R. Brown, A. Chlond, J. Cuxart, P. G. Duynkerke, D. C. Lewellen, M. K. Macvean, R. A. J. Neggers, E. Sánchez, A. P. Siebesma, and D. E. Stevens, 2001: Simulations of Trade Wind Cumuli under a Strong Inversion. *Journal of the Atmospheric Sciences*, **58**, 1870–1891, doi:10.1175/1520-0469(2001)058¡1870:SOTWCU¿2.0.CO;2.

Stevens, B., C. Acquistapace, A. Hansen, R. Heinze, C. Klinger, D. Klocke, H. Rybka, W. Schubotz, J. Windmiller, P. Adamidis, I. Arka, V. Barlakas, J. Biercamp, M. Brueck, S. Brune, S. Buehler, U. Burkhardt, G. Cioni, M. Costa-Surros, S. Crewell, T. Crueger, H. Deneke, P. Friederichs, C. Carbajal Henken, C. Hohenegger, M. Jacob, F. Jakub, N. Kalthoff, M. Köhler, P. Li, U. Löhnert, A. Macke, N. Madenach, B. Mayer, C. Nam, A. K. Naumann, K. Peters, S. Poll, J. Quaas, N. Röber, N. Rochetin, L. Scheck, V. Schemann, S. Schnitt, A. Seifert, F. Senf, M. Shapkalijevski, C. Simmer, S. Singh, O. Sourdeval, D. Spickermann, J. Strandgren, O. Tessiot, T. van Laar, N. Vercauteren, J. Vial, A. Voigt, and G. Zängl, 2020: The Added Value of Large-Eddy and Storm-Eesolving Models for Simulating Clouds and Precipitation. *Journal of the Meteorological Society of Japan*, **98**, doi:10.2151/jmsj.2020-021.

Stevens, B., C.-H. Moeng, A. S. Ackerman, C. S. Bretherton, A. Chlond, S. de Roode, J. Edwards, J.-C. Golaz, H. Jiang, M. Khairoutdinov, M. P. Kirkpatrick, D. C. Lewellen, A. Lock, F. Müller, D. E. Stevens, E. Whelan, and P. Zhu, 2005: Evaluation of Large-Eddy Simulations via Observations of Nocturnal Marine Stratocumulus. *Monthly Weather Review*, **133**, 1443–1462, doi:10.1175/MWR2930.1.

Storey, R. C., S. E. Norris, and J. E. Cater: 2014, Modelling Turbine Loads during an Extreme Coherent Gust using Large Eddy Simulation. *Science Making Torque from Wind 2014 (Torque 2014)*, C. Bak, A. Bechmann, F. Bingol, E. Dellwik, N. Dimitrov, G. Giebel, M. Hansen, D. Jensen, G. Larsen, H. Madsen, J. Mann, A. Natarajan, O. Rathmann, A. Sathe, J. Sorensen, and N. Sorensen, eds., European Acad Wing Energy, volume 524 of *Journal of Physics Conference Series*, 5th Science of Making Torque from Wind Conference, Tech Univ Denmark, Copenhagen, Denmark, Jun 18-20, 2014.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers.

Suomi, I., S.-E. Gryning, R. Floors, T. Vihma, and C. Fortelius, 2015: On the vertical structure of wind gusts. *Quarterly Journal of the Royal Meteorological Society*, **141**, 1658–1670, doi:10.1002/qj.2468.

Tatarevic, A., P. Kollias, M. Oue, D. Wang, and Y. Kwangmin, 2019: *User's Guide CR-SIM SOFTWARE v 3.2*. McGill University Applied Radar Science Group.

United Nations, D. O. E. and P. D. Social Affairs, 2018: The World's Cities in 2018.

Upton, G. and I. Cook, 1996: *Understanding Statistics*. Oxford University Press.

Vaisala, 2013: Vaisala Radiosonde RS92-SGP Datasheet. Web: https://www.vaisala.com/sites/default/files/documents/RS92SGP-Datasheet-B210358EN-F-LOW.pdf, accessed: 29/11/2019.

Valero, F., A. Pascual, and M. Martin, 2014: An approach for the forecasting of wind strength tailored to routine observational daily wind gust data. *Atmospheric Research*, **137**, 58 – 65, doi:https://doi.org/10.1016/j.atmosres.2013.09.019.

van den Brink, H. W. and F. C. Bosveld, 2017: Bringing modelled and observed surface winds together. *International Journal of Climatology*, **37**, 4783–4790, doi:10.1002/joc.5121.

van der Linden, S. J. A., J. M. Edwards, C. C. van Heerwaarden, E. Vignon, C. Genthon, I. Petenko, P. Baas, H. J. J. Jonker, and B. J. H. van de Wiel, 2019: Large-Eddy Simulations of the Steady Wintertime Antarctic Boundary Layer. *Boundary-Layer Meteorology*, **173**, 165–192, doi:10.1007/s10546-019-00461-4.

van der Walt, S., J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, 2014: scikit-image: image processing in Python. *PeerJ*, **2**, e453, doi:10.7717/peerj.453.

van Heerwaarden, C. C., B. J. H. van Stratum, T. Heus, J. A. Gibbs, E. Fedorovich, and J. P. Mellado, 2017: MicroHH 1.0: a computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows. *Geoscientific Model Development*, **10**, 3145–3165, doi:10.5194/gmd-10-3145-2017.

van Stratum, B. J. H. and B. Stevens, 2018: The Impact of Vertical Mixing Biases in Large-Eddy Simulation on Nocturnal Low Clouds. *Journal of Advances in Modeling Earth Systems*, **10**, 1290–1303, doi:10.1029/2017MS001239.

van Ulden, A. and J. Wieringa, 1996: Atmospheric boundary layer research at Cabauw. *Boundary-Layer Meteorology*, **78**, 39–69, doi:10.1007/BF00122486.

Weigl, E., T. Reich, P. Lang, A. Wagner, O. Kohler, and N. Gerlach, 2004: Projekt RAD-OLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer). Technical report.

Weitkamp, C., 2005: *Lidar - Range-Resolved Optical Remote Sensing of the Atmosphere*. Springer.

Weusthoff, T., F. Schubiger, and F. Ament, 2009: Verification of precipitation forecasts of the MAP D-PHASE data set with fuzzy methods. *4th International Verification Methods Workshop, Helsinki*.

WMO, 1987: The Measurement of Gustiness at Routine Wind Stations - A Review. *IOM Report- No. 31*.

Wu, K., E. Otoo, and A. Shoshani: 2005, Optimizing connected component labeling algorithms. *Medical Imaging 2005: Image Processing*, J. Michael Fitzpatrick and Joseph M. Reinhardt, ed., International Society for Optics and Photonics, SPIE, volume 5747, 1965 – 1976.

Zhang, X., J. Huang, G. Li, Y. Wang, C. Liu, K. Zhao, X. Tao, X.-M. Hu, and X. Lee, 2019: Improving Lake-Breeze Simulation with WRF Nested LES and Lake Model over a Large Shallow Lake. *Journal of Applied Meteorology and Climatology*, **58**, 1689–1708, doi:10.1175/JAMC-D-18-0282.1.

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

Publications and Supervised Theses

Peer Reviewed Publications

Heinze, R., A. Dipankar, C. C. Henken, C. Moseley, O. Sourdeval, S. Trömel, X. Xie, P. Adamidis, F. Ament, H. Baars, C. Barthlott, A. Behrendt, U. Blahak, S. Bley, S. Brdar, M. Brueck, S. Crewell, H. Deneke, P. Di Girolamo, R. Evaristo, J. Fischer, C. Frank, P. Friederichs, T. Göcke, K. Gorges, L. Hande, M. Hanke, **A. Hansen**, H.-C. Hege, C. Hoose, T. Jahns, N. Kalthoff, D. Klocke, S. Kneifel, P. Knippertz, A. Kuhn, T. van Laar, A. Macke, V. Maurer, B. Mayer, C. I. Meyer, S. K. Muppa, R. A. J. Neggers, E. Orlandi, F. Pantillon, B. Pospichal, N. Röber, L. Scheck, A. Seifert, P. Seifert, F. Senf, P. Siligam, C. Simmer, S. Steinke, B. Stevens, K. Wapler, M. Weniger, V. Wulfmeyer, G. Zängl, D. Zhang, and J. Quaas, 2017a: Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, **43**, 69–100, doi:10.1002/qj.2947.

Lammert, A., **A. Hansen**, F. Ament, S. Crewell, G. Dick, V. Grutzun, H. Klein-Baltink, V. Lehmann, A. Macke, B. Pospichal, W. Schubotz, P. Seifert, E. Stamnas, and B. Stevens, 2019: A Standardized Atmospheric Measurement Data (SAMD) Archive for distributed cloud and precipitation process-oriented observations in Central Europe. *Bulletin of the American Meteorological Society*, **100**, 1299–1314, doi:10.1175/BAMS-D-18-0174.1.

Stevens, B., F. Ament, S. Bony, S. Crewell, F. Ewald, S. Gross, **A. Hansen**, L. Hirsch, M. Jacob, T. Klling, H. Konow, B. Mayer, M. Wendisch, M. Wirth, K. Wolf, S. Bakan, M. Bauer-Pfundstein, M. Brueck, J. Delano, A. Ehrlich, D. Farrell, M. Forde, F. Gdde, H. Grob, M. Hagen, E. Jkel, F. Jansen, C. Klepp, M. Klingebiel, M. Mech, G. Pe-ters, M. Rapp, A. A. Wing, and T. Zinner, 2019: A High-Altitude Long-Range Aircraft Configured as a Cloud Observatory: The NARVAL Expeditions. *Bulletin of the American Meteorological Society*, **100**, 1061–1077, doi:10.1175/BAMS-D-18-0198.1.

Pantillon, F., B. Adler, U. Corsmeier, P. Knippertz, A. Wieser, and **A. Hansen**, 2020: Formation of Wind Gusts in an Extratropical Cyclone in Light of Doppler Lidar Observations and Large-Eddy Simulations. *Monthly Weather Review*, **148**, 353–375, doi:10.1175/ MWR-D19-0241.1.

Stevens, B., C. Acquistapace, **A. Hansen**, R. Heinze, C. Klinger, D. Klocke, H. Rybka, W. Schubotz, J. Windmiller, P. Adamidis, I. Arka, V. Barlakas, J. Biercamp, M. Brueck, S. Brune, S. Buehler, U. Burkhardt, G. Cioni, M. Costa-Surros, S. Crewell, T. Crueger, H. Deneke, P. Friederichs, C. Carbajal Henken, C. Hohenegger, M. Jacob, F. Jakub, N. Kalthoff, M. Köhler, P. Li, U. Löhnert, A. Macke, N. Madenach, B. Mayer, C. Nam, A. K. Naumann, K. Peters, S. Poll, J. Quaas, N. Röber, N. Rochetin, L. Scheck, V. Schemann, S. Schnitt, A. Seifert, F. Senf, M. Shapkalijevski, C. Simmer, S. Singh, O. Sourdeval, D. Spickermann, J. Strandgren, O. Tessiot, T. van Laar, N. Vercauteren, J. Vial, A. Voigt, and G. Zängl, 2020: The Added Value of Large-Eddy and Storm-Eesolving Models for Simulating Clouds and Precipitation. *Journal of the Meteorological Society of Japan*, **98**, doi:10.2151/jmsj.2020-021.

Stevens, B., C. Acquistapace, **A. Hansen**, R. Heinze, C. Klinger, D. Klocke, H. Rybka, W. Schubotz, J. Windmiller, P. Adamidis, I. Arka, V. Barlakas, J. Biercamp, M. Brueck, S. Brune, S. Buehler, U. Burkhardt, G. Cioni, M. Costa-Surros, S. Crewell, T. Crueger, H. Deneke, P. Friederichs, C. Carbajal Henken, C. Hohenegger, M. Jacob, F. Jakub, N. Kalthoff, M. Köhler, P. Li, U. Löhnert, A. Macke, N. Madenach, B. Mayer, C. Nam, A. K. Naumann, K. Peters, S. Poll, J. Quaas, N. Röber, N. Rochetin, L. Scheck, V. Schemann, S. Schnitt, A. Seifert, F. Senf, M. Shapkalijevski, C. Simmer, S. Singh, O. Sourdeval, D. Spickermann, J. Strandgren, O. Tessiot, T. van Laar, N. Vercauteren, J. Vial, A. Voigt, and G. Zängl: The Atlantic ITCZ, *In Preparation*.

Supervised Theses

Radtke, J., 2018: Evaluierung der Variabilität solarer Strahlung in Large Eddy Simulationen mit ICON. Bachelor thesis, Universität Hamburg, Hamburg.

Latsch, M., 2019: Wind fields of the Atmospheric Boundary Layer dynamically downscaled by high-resolution mesoscale Simulations. Master thesis, Universität Hamburg, Hamburg.

Thielke, L., 2019: Vertical momentum flux measurements at a Boundary Layer Tower and its parameterization in numerical atmospheric models. Master thesis, Universität Hamburg, Hamburg.

Marshall, J., 2019: Windtrends in Nabenhöhe aus Reanalysedaten. Master thesis, Universität Hamburg, Hamburg.

Borowski, J., 2020: Johanna Borowski: Charakterisierung von Windvorhersagefehlern des ICON-EU Modells auf Nabenhöhe. Master thesis, Universität Hamburg, Hamburg.

Aue, L., 2020: Klassifizierung von Fehlern in der Strahlungsvorhersage des ICON-EU Modells und deren Einfluss auf Solarleistungsprognosen. Master thesis, Universität Hamburg, Hamburg.

Acknowledgements

First of all, I would like to thank my supervisors Felix Ament and Andrea Lammert, for their excellent support, all the fruitful and valuable discussions and the opportunity to write my PhD thesis within the motivating $HD(CP)^2$ project. Especially, I want to thank them for the possibility to ask them for advice at any time, as well as for the chance to participate in various workshops and conferences. Furthermore, I cordially want to thank Felix and Andrea for all their confidence in me and the opportunities to develop my own lectures and to supervise bachelor and master thesis, which was very encouraging for me and I've learned a lot from these experiences. I could not have imagined of a better supervision and working environment for my PhD. Thank you very much!

Additionally, I would like to thank Stefan Bühler for taking the time to be my panel chair and for all his valuable suggestions and inputs. Likewise, I wish to express my thank to Thomas Ludwig, Ann Kristin Naumann and Jürgen Scheffran for their willingness and interest of being members of my PhD committee. I also want to thank the IMPRS-ESM and especially Antje Weitz, Connie Kampmann and Wiebke Böhm for all their advice and especially for the opportunity to write my PhD within the encouraging Max Planck society. This was a fantastic experience for me!

I cordially want to thank the whole $HD(CP)^2$ community for the very encouraging discussions, advice and the motivating project environment. Thank you all! My special thanks go to Matthias Brueck, Stefan Poll, Katrin Gehrke, Anurag Dipanker, Rieke Heinze and Martin Köhler for the excellent collaboration and the fruitful discussions.

Versicherung an Eides statt
Declaration on oath

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: "New Techniques for Ultra-High-Resolution Circulation Model Evaluation" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen - benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, den 05.02.2020   Akio Hansen

Eidesstattliche Versicherung / *Declaration on oath*

Hiermit erkläre ich an Eides Statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg _____   Unterschrift / Signature _____
           (Datum / Date)