RMetS
Royal Meteorological Society

Large-eddy simulations over Germany using ICON: a comprehensive evaluation

Rieke Heinze,[a,b] Anurag Dipankar,[a,c] Cintia Carbajal Henken,[d] Christopher Moseley,[a] Odran Sourdeval,[e] Silke Trömel,[f] Xinxin Xie,[f] Panos Adamidis,[g] Felix Ament,[h] Holger Baars,[i] Christian Barthlott,[j] Andreas Behrendt,[k] Ulrich Blahak,[l] Sebastian Bley,[i] Slavko Brdar,[l] Matthias Brueck,[e,a] Susanne Crewell,[m] Hartwig Deneke,[i] Paolo Di Girolamo,[n] Raquel Evaristo,[f] Jürgen Fischer,[d] Christopher Frank,[m,o] Petra Friederichs,[f] Tobias Göcke,[l] Ksenia Gorges,[g] Luke Hande,[j] Moritz Hanke,[g] Akio Hansen,[h] Hans-Christian Hege,[p] Corinna Hoose,[j] Thomas Jahns,[g] Norbert Kalthoff,[j] Daniel Klocke,[l,o] Stefan Kneifel,[m] Peter Knippertz,[j] Alexander Kuhn,[p] Thriza van Laar,[m] Andreas Macke,[i] Vera Maurer,[j,q,o] Bernhard Mayer,[r] Catrin I. Meyer,[s] Shravan K. Muppa,[k] Roeland A. J. Neggers,[m] Emiliano Orlandi,[m,t] Florian Pantillon,[j] Bernhard Pospichal,[e] Niklas Röber,[g] Leonhard Scheck,[r] Axel Seifert,[l] Patric Seifert,[i] Fabian Senf,[i] Pavan Siligam,[g] Clemens Simmer,[f] Sandra Steinke,[m] Bjorn Stevens,[a] Kathrin Wapler,[l] Michael Weniger,[f] Volker Wulfmeyer,[k] Günther Zängl,[l] Dan Zhang[l] and Johannes Quaas[e]*

[a]*Max-Planck-Institut für Meteorologie, Hamburg, Germany*
[b]*Institute of Meteorology and Climatology, Leibniz Universität Hannover, Germany*
[c]*Centre for Climate Research Singapore*
[d]*Institute for Space Sciences, Freie Universität Berlin, Germany*
[e]*Institute for Meteorology, Universität Leipzig, Germany*
[f]*Meteorological Institute, Universität Bonn, Germany*
[g]*Deutsches Klimarechenzentrum, Hamburg, Germany*
[h]*Meteorological Institute, Universität Hamburg, Germany*
[i]*Leibniz Institute for Tropospheric Research Leipzig, Germany*
[j]*Institute of Meteorology and Climate Research - Department Troposphere Research (IMK-TRO), Karlsruhe Institute of Technology, Germany*
[k]*Universität Hohenheim, Germany*
[l]*Deutscher Wetterdienst, Offenbach, Germany*
[m]*Institute for Geophysics and Meteorology, Universität zu Köln, Germany*
[n]*Scuola di Ingegneria, Università degli Studi della Basilicata, Potenza, Italy*
[o]*Hans-Ertel-Zentrum für Wetterforschung*
[p]*Zuse-Institut Berlin, Germany*
[q]*Institute for Atmospheric and Environmental Sciences, Goethe-Universität Frankfurt am Main, Germany*
[r]*Meteorological Institute, Ludwig-Maximilians-Universität München, Germany*
[s]*Jülich Supercomputing Centre, Forschungszentrum Jülich*
[t]*RPG Radiometer Physics GmbH, Meckenheim, Germany*

*Correspondence to: J. Quaas, Universität Leipzig, Institute for Meteorology, Stephanstr. 3, D-04103 Leipzig, Germany. E-mail: johannes.quaas@uni-leipzig.de

Large-eddy simulations (LES) with the new ICOsahedral Non-hydrostatic atmosphere model (ICON) covering Germany are evaluated for four days in spring 2013 using observational data from various sources. Reference simulations with the established Consortium for Small-scale Modelling (COSMO) numerical weather prediction model and further standard LES codes are performed and used as a reference. This comprehensive evaluation approach covers multiple parameters and scales, focusing on boundary-layer variables, clouds and precipitation. The evaluation points to the need to work on parametrizations influencing the surface energy balance, and possibly on ice cloud microphysics. The central purpose for the development and application of ICON in the LES configuration is the use of simulation results to improve the understanding of moist processes, as well as their parametrization in climate models. The evaluation thus aims

at building confidence in the model's ability to simulate small- to mesoscale variability in turbulence, clouds and precipitation. The results are encouraging: the high-resolution model matches the observed variability much better at small- to mesoscales than the coarser resolved reference model. In its highest grid resolution, the simulated turbulence profiles are realistic and column water vapour matches the observed temporal variability at short time-scales. Despite being somewhat too large and too frequent, small cumulus clouds are well represented in comparison with satellite data, as is the shape of the cloud size spectrum. Variability of cloud water matches the satellite observations much better in ICON than in the reference model. In this sense, it is concluded that the model is fit for the purpose of using its output for parametrization development, despite the potential to improve further some important aspects of processes that are also parametrized in the high-resolution model.

Key Words: large-eddy simulation; clouds and precipitation; evaluation with observations

Received 4 July 2016; Revised 30 September 2016; Accepted 17 October 2016; Published online in Wiley Online Library 11 January 2017

1. Introduction

The response of clouds to anthropogenic forcings is the main cause of diversity among climate models simulating global climate change (Boucher *et al.*, 2013). This is true for both the rapid response to perturbations in carbon dioxide concentrations and aerosols on time-scales of hours to weeks (Gregory and Webb, 2008; Sherwood *et al.*, 2015) and the feedbacks responding to sea-surface temperature changes (Bony *et al.*, 2006; Vial *et al.*, 2013).

A fundamental reason as to why clouds are not represented realistically in current general circulation models (GCMs) is their coarse grid spacing (horizontally, of $\mathcal{O}$ (100 km); in the remainder of this article, we will refer to 'grid spacing' as the 'resolution' of the model). Cloud processes occur across scales from the particle scale at $\mathcal{O}$ (1 μm) to the scale of individual clouds, $\mathcal{O}$ (1 km), and cloud systems, $\mathcal{O}$ (100 km). All scales are thus necessarily parametrized in GCMs and might become partially resolved with increasing resolution. Furthermore, the chain of parametrizations in land–atmosphere systems, such as land–atmosphere exchange and planetary boundary-layer turbulence, still suffers from significant systematic errors (Milovac *et al.*, 2016). Therefore, the representation of turbulence and land–atmosphere exchange must also be investigated and improved.

Large eddy simulations (LES), where a considerable part of the turbulence and thus of cloud-scale motions is resolved at resolutions of the order of 1–100 m, have been used for more than four decades (Lilly, 1967; Deardorff, 1970a, 1970b). Due to the substantial computational expense, such simulations usually covered small domains and involved many idealizations, especially with respect to boundary conditions. Thanks to increasing computational resources, increasingly larger domain cloud-resolving simulations are now possible (e.g. Hohenegger *et al.*, 2008; Khairoutdinov *et al.*, 2009; Love *et al.*, 2011; Schlemmer and Hohenegger, 2014; Schalkwijk *et al.*, 2015). Some groups have even started realistic, weather-forecast mode simulations and evaluated these with observational data (Caine *et al.*, 2013; Hanley *et al.*, 2015; Stein *et al.*, 2015). While high-resolution simulations for longer in time are also possible and beneficial (e.g. Neggers *et al.*, 2012), the parallel architecture of current and upcoming high-performance computers is particularly suited to handling large domains. This development is complemented by numerical weather prediction (NWP) and climate simulations that are performed at substantially better resolution than was possible in the past (Schalkwijk *et al.*, 2015), now regularly allowing users to switch off parametrizations for deep convection (Baldauf *et al.*, 2011; Holloway *et al.*, 2012, 2014; Prein *et al.*, 2015; Clark *et al.*, 2016).

In the High Definition Clouds and Precipitation for advancing Climate Prediction (HD(CP)^2) project, the ICOsahedral Non-hydrostatic (ICON) atmospheric model (Zängl *et al.*, 2015, hereafter Z15) was developed towards LES applications (Dipankar *et al.*, 2015, hereafter D15). The goal of the HD(CP)^2 project is to make use of the LES configuration of ICON for advancing climate prediction, through (i) an improved understanding of cloud and precipitation processes, thanks to a large, high-resolution, consistent dataset, (ii) developing, evaluating and improving parametrizations of moist diabatic processes for GCMs on the basis of the ICON output and (iii) assessing rapid responses of clouds to anthropogenic forcings in perturbed simulations, using ICON in the LES configuration. A prerequisite to all these approaches is to build confidence that the ICON simulations are sufficiently realistic.

In this study, ICON is used in LES configuration to simulate four days in spring 2013 covering Germany. The days were chosen so that data obtained from the HD(CP)^2 Observational Prototype Experiment (HOPE: Macke *et al.*, 2016) near Jülich could be used for evaluation. These first simulations with the new model are evaluated using multiple observational datasets created or compiled in the HD(CP)^2 project and also using further datasets from various sources. The ICON simulations are further compared with a reference simulation with the COnsortium for Small-scale MOdelling (COSMO) model (Baldauf *et al.*, 2011) at a horizontal resolution of 2.8 km in a corresponding set-up, which is similar to Barthlott and Hoose (2015). It is not to be expected that ICON should be superior to the established COSMO model with regard to synoptic or even mesoscale weather phenomena. However, the high resolution that allows us to resolve explicitly several processes that are parametrized even in the COSMO simulations should be beneficial to the representation of processes that are subgrid-scale for regular numerical weather forecast and climate models. We aim at a comprehensive evaluation approach for the ICON simulation, by taking further standard large-eddy models in semi-idealized configurations (Heinze *et al.*, 2016) into account. In particular, we focus on assessing the moist processes, including cloud, precipitation and convection dynamics and cloud and precipitation microphysical processes, as well as boundary-layer dynamics.

As the HD(CP)^2 project continues into its second phase, part of the intention of this study is also to document, besides the capabilities, the current shortcomings of ICON in order to monitor future progress. In the second phase, special attention will be paid to important cloud regimes including deep convection, which is not analyzed in detail in the current study.

The article is structured as follows: section 2 introduces ICON in the LES configuration, the set-up and the simulation output, while section 3 provides an overview of the observations used. The evaluation results are shown and discussed in section 4 for boundary-layer characteristics including turbulence, in section 5 for water-vapour distributions and variability, in section 6 for clouds and in section 7 for precipitation. Section 8 summarizes the main conclusions and outlines future work to improve the ICON model further.

2. Model description, set-up and simulation output

ICON as a unified modelling framework offers three basic physics packages, which are dedicated to NWP, climate modelling and large-eddy simulations, respectively. The LES configuration of ICON was validated for classical idealized LES configurations of a dry convective and a cumulus-topped boundary layer with doubly periodic horizontal boundary conditions and flat geometries (D15). In the present study, ICON is used in a real-case configuration with prescribed lateral boundary conditions and a one-way nesting approach (refined simulations embedded simultaneously in the model run) to perform high-resolution simulations over Germany. Section 2.1 provides an overview of the LES configuration and details about the set-up are presented in section 2.2. Simulations from additional models are used for validation purposes and are introduced in section 2.3. In section 2.4, a brief overview of the computational performance of ICON in the LES configuration is provided. Section 2.5 provides information on the simulation output. In order to take full advantage of remote sensing observations that allow us to investigate model skill at high resolution, different forward operators are used, which are described in section 2.6.

2.1. ICON configuration

ICON solves the prognostic variables, i.e. the ones suggested by Gassmann and Herzog (2008), on an unstructured triangular grid based on successive refinement of a spherical icosahedron (Wan *et al.*, 2013, Z15). The set of equations is described extensively in Z15 and D15 and is not repeated here. It includes the horizontal velocity component normal to the triangle edge, v_2, the vertical velocity component, v_3, the density of moist air, ρ (including condensate), the virtual potential temperature, θ_v, the mass fractions, q_x, and number densities, n_x, of a number of tracers, x, including the mass fraction of water vapour, q_v, and different hydrometeors as prognostic variables. The velocity component tangential to the triangle edge, v_1, is diagnosed using the radial basis reconstruction (Narcowich and Ward, 1994). The equations are implicitly filtered to reflect the scale separation inherent to LES (D15).

The horizontal discretization is formulated on triangular cells with C-type staggering. In other words, the normal velocity component v_2 is defined at triangle edge midpoints, while all other prognostic variables are defined at cell circumcentres.

The vertical discretization is formulated in a height-based terrain-following coordinate system. Here, the smooth level vertical (SLEVE) coordinate implementation (Leuenberger *et al.*, 2010) is used. Vertical staggering is of Lorenz type (Lorenz, 1960), with vertical velocity v_3 being defined at interface (half) levels and the remaining prognostic quantities (v_1, ρ, θ_v, q_x, n_x) defined at main (full) levels.

The LES physics package of ICON described by D15 was developed further to include metric correction terms stemming from the terrain-following coordinate system. These terms are added to the tendency terms due to the turbulence parametrization, which is a three-dimensional (3D), diagnostic Smagorinsky scheme with modifications from Lilly (1962) to account for thermal stratification. Details about the metric terms can be found in Baldauf and Brdar (2016) and in Section SI-1. While the details are given in D15, for the sake of completeness, it is to be noted that the Smagorinsky scheme in ICON is applied on prognostic winds, potential temperature, specific humidity and specific cloud liquid water.

Concerning microphysics, the two-moment mixed-phase bulk microphysical parametrization of Seifert and Beheng (2006, hereafter SB) is applied. The warm-rain part of the scheme is based on Seifert and Beheng (2001), with modifications following Seifert (2008). The SB scheme has a separate hail category, which includes wet growth processes and a spectral partitioning of freezing raindrops (Blahak, 2008; Noppel *et al.*, 2010). This version of the SB scheme predicts the mass fractions q_x and number densities n_x of cloud droplets, rain, cloud ice, snow, graupel and hail ($x \in \{c, r, i, s, g, h\}$). An additional prognostic tracer is used to track the number of activated ice nuclei to describe their depletion. Important parameters of the SB scheme and the chosen values are summarized in Table SI-1 of File S1.

Heterogeneous ice nucleation is parametrized based on mineral dust concentrations simulated with the COSMO Multi-Scale Chemistry Aerosol Transport (COSMO-MUSCAT) model (Wolke *et al.*, 2004, 2012), as described in Hande *et al.* (2015) using the data for spring as given in their table 1. In addition, ice nucleation via homogeneous freezing of aqueous solution droplets is included, following Kärcher and Lohmann (2002) and Kärcher *et al.* (2006). Activation of cloud condensation nuclei (CCN) is parametrized based on a simulation of the aerosol conditions for 17 April 2013 with COSMO-MUSCAT. This profile was chosen as it represents a typical condition over all simulation days, with approximately constant CCN concentrations in the lowest 1500 m and a decrease above. The CCN concentration is then parametrized as a function of pressure and vertical velocity (Hande *et al.*, 2016).

Further physics parametrizations include an all-or-nothing cloud fraction scheme that does not account for fractional cloud cover at subgrid scales. The multi-layer land-surface scheme Terra (Heise *et al.*, 2006) without subgrid land-cover variability is used. The surface transfer scheme is based on Louis (1979). For radiation, the Rapid Radiation Transfer Model (RRTM: Mlawer *et al.*, 1997) is applied.

Integration in time is performed using an explicit two-time-level predictor–corrector scheme, except for the contributions describing vertical sound-wave propagation, which are integrated implicitly. Different time steps (see next section for the set-up) are used for the dynamical core on the one side and tracer advection, numerical dissipation and physics parametrization on the other side (Z15). For efficiency reasons, the physics parametrizations are further split into fast-physics routines (turbulence, microphysics, saturation adjustment) and slow-physics routines (radiation). The fast-physics routines are called every physics time step, whereas the slow-physics routines are called after a user-defined multiple of the physics time step. To keep the integration numerically stable, the dynamical core is sub-stepped with respect to tracer advection, fast-physics parametrizations and horizontal diffusion. Usually, the time step used in the dynamical core is set to one fifth of the fast-physics time step.

The advective terms in the prognostic equations are discretized using second order for all variables except the vertical advection of tracers, which uses third order. For horizontal and vertical advection of momentum, centred differences are used. Horizontal advection of tracer variables (θ_v, ρ, q_x) is discretized by means of the upwind-biased scheme of Miura (2007). The discretization of vertical advection for θ_v and ρ largely follows the same ideas; however, it had to be adapted slightly to avoid the generation of spurious buoyancy waves (see Z15 for details). For the vertical advection of water species like q_v, as well as additional passive tracers, the third-order accurate Piecewise Parabolic Method (PPM: Colella and Woodward, 1984) is used.

The default artificial numerical dissipation is used to stabilize advection of velocity and temperature. It is a second-order Smagorinsky diffusion of velocity and temperature, which is combined with a fourth-order background diffusion of velocity. The Smagorinsky diffusion adjusts itself based on grid spacing. By calling a special divergence damping term in the v_2 equation every dynamical time step, we can afford to call the numerical dissipation only every physics time-step. Rayleigh damping on v_3 based on Klemp *et al.* (2008) is used to prevent reflection of gravity waves at the model top. For more detailed information on discretization and numerics, the reader is referred to Wan *et al.* (2013), Z15 and D15.

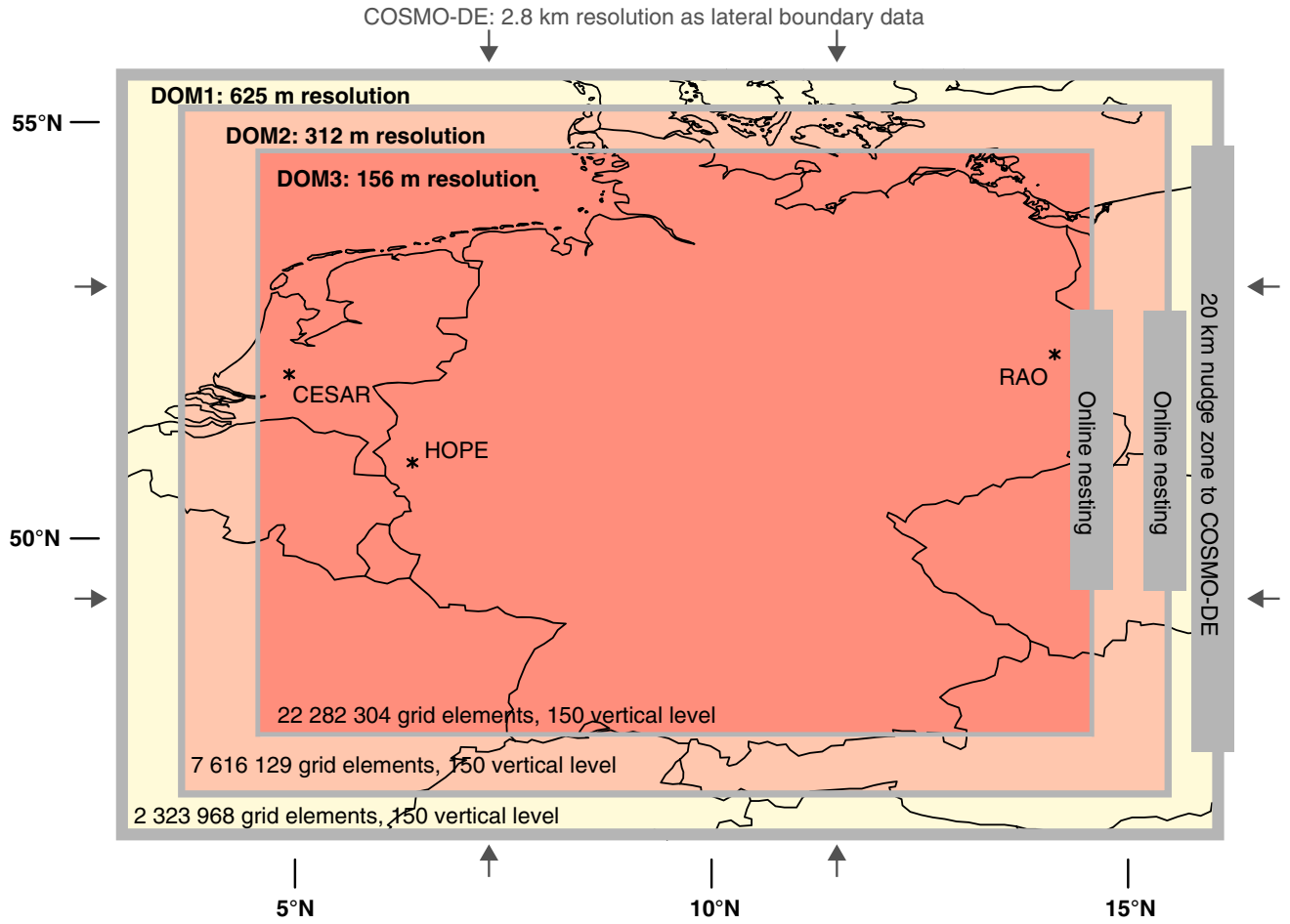

Figure 1. Simulation configuration: the simulation domain and its two nests with local grid refinement. The open lateral boundaries are relaxed towards the COSMO-DE analysis (see text for more details). The stars mark the approximate locations of the HD(CP)^2 Observational Prototype Experiment (HOPE), the Meteorological Observatory Lindenberg/Richard Aßman Observatory (RAO) and the Cabauw Experimental Site for Atmospheric Research (CESAR).

2.2. *Set-up*

Simulations of $\mathcal{O}(100\,\mathrm{m})$ over Germany are performed for four days in April and May 2013. These days in particular are chosen, as they occur within the period of the extensive measurement campaign HD(CP)^2 Observational Prototype Experiment (HOPE: e.g. Löhnert *et al.*, 2015; Steinke *et al.*, 2015; Macke *et al.*, 2016; Madhavan *et al.*, 2016; Maurer *et al.*, 2016) close to the city of Jülich (6.4 °E, 50.9 °N; see Figure 1) in the very west of Germany. Within the HOPE area of about $10 \times 10\,\mathrm{km}^2$, three supersites with advanced instrumentation for vertical profiling were located, namely the permanent Jülich Observatory for Cloud Evolution (JOYCE: Löhnert *et al.*, 2015), the Leipzig Aerosol and Cloud Remote Observations System (LACROS: Bühl *et al.*, 2013) and the Karlsruhe advanced mobile observation platform (KITcube: Kalthoff *et al.*, 2013). The specific days (24–26 April 2013 and 2 May 2013) were chosen to reflect a range of different spring-time central-European weather conditions (cloud-free, cumulus clouds, synoptic-scale driven; see section 3 for more details).

ICON is deployed in a limited-area set-up with local two-step grid refinement for the area of Germany, as shown in Figure 1. In each refinement step, the resolution is halved from 625 m, to 312 m and 156 m in the innermost domain. Here, the term *resolution* refers to the square root of the mean cell area in the icosahedral grid of ICON, which is equivalent to about 1.5 times the corresponding resolution in a regular grid (D15). 150 vertical levels are used, with grid stretching towards the model top at 21 km. The minimal layer thickness is 20 m near the surface and the lowest 1000 m encompass 20 layers. A fast-physics time-step of 3 s is used for the coarsest resolution. It is then halved with each refinement step. The slow-physics time-step for calling RRTM radiation is 900 s for each domain and is kept constant over the integration time.

Large eddy simulation over such a big domain is as much a computational challenge as a physical one. A total of more than 3.3 billion grid cells have to be advanced in time using a time step of 0.75 s in the innermost domain for a total of 4 days. Keeping this in mind, strong effort was made in optimizing ICON further for massive parallelization. The details are summarized in section 2.4.

Each simulated day is initialized at 0000 UTC from operational COSMO-DE (Baldauf *et al.*, 2011) analysis and runs for 24 h. The reason to initialize at midnight is that turbulence in the ICON model can develop in the morning, so that the model is spun up during the daytime, on which our analysis and evaluation is performed. COSMO-DE data are interpolated to the three domain grids by using a radial basis function (RBF) interpolation algorithm (Ruppert, 2007) and 3D variables are interpolated vertically during initialization. The soil moisture is converted to the soil moisture index for initialization and internally transferred back to soil moisture using the specifications of the soil types used. Vertical interpolation of the soil variables is not necessary, as COSMO-DE and ICON use the same soil model Terra in the same configuration. At the lateral boundaries of the outer domain, the numerical simulation of the model is relaxed towards hourly COSMO-DE analyses in a 20 km wide nudging zone (see outer grey margin in Figure 1).

Variables are linearly interpolated in time between the hourly boundary data analysis. The nudging is performed on the prognostic variables, namely the wind components, virtual potential temperature, density and, in subsaturated grid points, also specific humidity.

Table 1. Overview of the different models and simulations used. Note that the second column contains the specification for the inner domain (156 m resolution) only. The one-moment microphysics scheme in the semi-idealized ICON configuration (ICON-SI) is the one used operationally in COSMO-EU with two ice categories (Doms and Schättler, 2004).

Model	ICON	COSMO	ICON-SI	PALM	DALES
No. grid cells × levels	22 832 304 × 150	421 × 461 × 50	1 179 648 × 144	960 × 960 × 144	256 × 256 × 125
hor. res.	156 m	2.8 km	46 m	50 m	50 m
hor. domain	4.5°E – 14.5°E, 47.6°N – 54.6°N,	1.04°E – 19.84°N 44.72°N – 56.50°N,	38.4 × 38.4 km^2	48 × 48 km^2	12.8 × 12.8 km^2
Model top (km)	21	22	13	13	5
Equations	Compressible	Compressible	Compressible	Boussinesq	Boussinesq
hor. b. c.	Open lateral b. c. + 2 inner nests	Open lateral b. c.	Doubly-periodic	Doubly-periodic	Doubly-periodic
Land surface	Terra	Terra	homo. (MOST)	homo. (MOST)	ECMWF-IFS
Turbulence	diag. Smagorinsky	1D MY	diag. Smagorinsky	progn. Deardorff	progn. Deardorff
Microphysics	SB	SB	One-moment	SB (no ice)	SB (no ice)
Forcing	COSMO-DE (1-hrly.)	COSMO-EU (1-hrly.)	COSMO-DE (3-hrly.)	COSMO-DE (3-hrly.)	ECMWF (3-hrly.)

The abbreviations SI, MOST, ECMWF-IFS, MY, SB, hor., b. c., res., homo., hrly., diag. and progn. stand for semi-idealized, Monin–Obukhov similarity theory, integrated forecast system of the European Centre for Medium-Range Weather Forecasts, Mellor–Yamada, Seifert–Beheng, horizontal, boundary condition, resolution, homogeneous, hourly, diagnostic and prognostic, respectively.

The refined domains, so-called *nests*, are again nudged during run time (online) at every model time step to the next coarser domain. All prognostic variables are nudged in an eight grid-element wide nudging zone at the outer boundary of each nest (see inner grey margins in Figure 1). A one-way nesting approach is chosen, where information is passed only from a coarser to the next finest nest. The stepwise refinement of the resolution avoids large scale jumps from the driving boundary with 2.8 km resolution to the innermost nest and minimizes associated spin-up of small scales at the domain boundaries and associated artefacts.

The time-invariant data at the lower boundary taken from observational datasets are gridded on the three ICON grids separately with a pre-processor (Smiatek *et al.*, 2008) and read from file during model initialization. Topography (ASTER GDEM Validation Team, 2011), land-use data (GLOBCOVER: Bicheron *et al.*, 2008) and soil-type specifications (Harmonized World Soil Database, HWSD: Smiatek *et al.*, 2015) are used from dedicated high-resolution observational datasets appreciating the high resolution of the simulation with 30, 300 and 900 m resolution, respectively. A smoothing is applied to the topography, where a maximum height difference between adjacent grid points of 1000, 500 and 200 m is allowed for the individual domains, respectively.

2.3. Additional simulations

For evaluation of the ICON simulations, additional simulations using the COSMO model (Baldauf *et al.*, 2011) and several conventional LES models are used. COSMO ran in a nearly operational set-up in limited-area mode without data assimilation. Initial conditions and lateral boundary data are taken from hourly operational COSMO-EU analyses with horizontal grid spacing of 7 km. The topography, land-use data set and soil data sources are the same as for the ICON simulations. COSMO uses a rotated latitude/longitude C-grid with terrain-following hybrid height coordinates. The horizontal resolution is 2.8 km and 50 vertical levels are used, where 12 levels are in the lowest 1000 m. The domain corresponds to the operational COSMO-DE domain (see Table 1). The model (and physics) time step is 25 s. Concerning physics parametrizations, the convection scheme originates from Tiedtke (1989), but is restricted to shallow convection with a cloud depth not exceeding 250 hPa (see Baldauf *et al.*, 2011, for details); the Terra land surface model (Heise *et al.*, 2006), RRTM radiation (Mlawer *et al.*, 1997) with a calling frequency of 900 s, the two-moment SB microphysics parametrization and the 1D Mellor and Yamada (1974) turbulence parametrization are used.

Two conventional large-eddy models, the PArallelized Large eddy simulation Model (PALM: Maronga *et al.*, 2015) and the Dutch Atmospheric Large-Eddy Simulation model (DALES: Heus *et al.*, 2010), are used in a semi-idealized (SI) configuration to support the evaluation of ICON in the HOPE region. The SI configuration includes using doubly periodic boundary conditions and a homogeneous surface forcing combined with prescribed large-scale advective tendencies for temperature and humidity, as well as large-scale vertical motion for the HOPE region, together with Newtonian relaxation to prevent excessive model drift in time (Neggers *et al.*, 2012). This allows for semi-idealized simulations, which are representative for the larger HOPE area and can be compared with observational datasets. Further details, especially about the SI simulations with PALM, can be found in Heinze *et al.* (2016). Details about the model configurations can be found in Table 1. Major differences between the PALM and DALES simulations are the treatment of the surface (prescribed temperature and humidity versus interactive land-surface model) and the source for the forcing data (COSMO-DE versus European Centre for Medium-Range Weather Forecasts (ECMWF)).

To complement the analysis, ICON was also used in the same SI configuration with a flat mesh, homogeneously prescribed temperature and humidity at the surface in a doubly cyclic domain with exactly the same large-scale forcing data for the HOPE region as in the PALM simulation. Further details are listed in Table 1.

2.4. Performance of the ICON code

ICON in the LES configuration aims to resolve details of turbulence and moist processes for large domains. Such simulations are intensive, both computationally and concerning the amount of output data. Therefore, it is vital to be able to exploit the hardware resources of exascale high-performance computing (HPC) systems in an effective way. Current and upcoming HPC systems are massively parallel computers consisting of hundreds of thousands of cores. A good scaling behaviour of ICON is the key to using such architectures efficiently. Furthermore, the time needed for writing the simulation results out on the file system makes the scalability of the model a big challenge. In order to address this issue, a major refactoring of the code has been undertaken. Thereby, all the global fields were substituted with distributed data structures and the corresponding algorithms were parallelized. Successful strong scaling experiments (test runs with constant problem size and varying number of parallel processes) in the course of the refactoring of ICON were carried out by Jülich Supercomputing Centre and German Climate Computing Centre

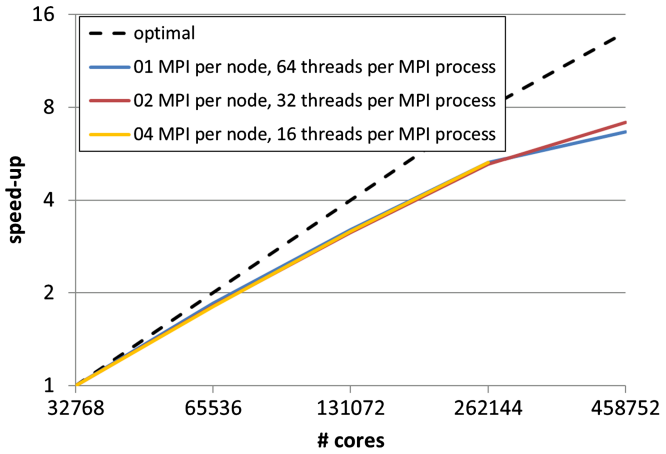

Figure 2. Results of strong scaling tests of ICON with a spatial resolution of 120 m performed on *JUQUEEN*. The dashed line denotes the optimal speed-up, solid coloured lines show test results with different number of Message Passing Interface (MPI) processes per node and threads per MPI process (hybrid parallelization). Note that one node consists of 16 cores.

(Deutsches Klimarechenzentrum, DKRZ) on the BlueGene/Q system *JUQUEEN* of Jülich Supercomputing Centre. Runtime measurements of ICON with a spatial resolution of 120 m and more than 6.7 billion grid cells (42 million horizontally with 160 vertical levels) show an excellent scaling up to 458 752 cores (Figure 2). This scaling result only assesses the time that the parallel algorithms need during the calculation, not the time for writing the results to files. The ICON simulations evaluated in the present article were carried out on the supercomputer *Mistral* at DKRZ, which has been available since July 2015.

2.5. *Simulation output and workflow*

Simulation output is generated on each of the three domains on the native triangular grid. Depending on the type of analysis performed, data sets used may be remapped on to a regular geographic (lat/lon) grid using distance-weighted average remapping of the four nearest neighbours. The output frequency for 1D profiles at different locations is 10 s. For 2D and 3D snapshots, day- and night-time frequency are distinguished: 10 s and 5 min for 2D output and 30 min and 1 h for 3D output, respectively. Day-time frequency starts at 0600 UTC and lasts until midnight. Furthermore, 3D snaphots are taken at overpass times of the *A-Train* satellite constellation (between about 1200 and 1300 UTC). In total about 50 TB model output is generated for one day of simulation. Additionally, 16 TB of restart files are stored on a tape archive. The simulations, including output on all three grids, took about 12 days on 7200 compute cores to simulate 1 day in the described set-up.

Advanced visualization of these large data sets poses a challenge to current visualization tools. The way taken here is described further in the Appendix.

As the long author list suggests, a large part of the HD(CP)^2 community took part in the evaluation of ICON. Taking into account the wide spread of users across Germany and their different scientific focuses, additional work on post-processing was required after a successful model run. The total amount of model output data for one simulated day and the data distribution workflow are shown in Figure 3. Since it is not possible to distribute such an amount of data efficiently across the community, post-processing on a personal level was developed and offered as a service to the community by a workflow team. Community members interested to use the simulation data specified their area of interest (e.g. campaign area or single station position), time resolution and a number of variables. The workflow team prepared the data sets and distributed them in the most convenient way for the community members (e.g. over cloud storage or Mistral working directory).

2.6. *Forward operators*

In order to exploit non-coventional observations for model evaluations, different forward operators listed in Table 2 are applied to ICON and COSMO output in the subsequent analysis. The visible satellite forward operator (VISOP) generates radiances for the solar channels of the MODerate Resolution Imaging

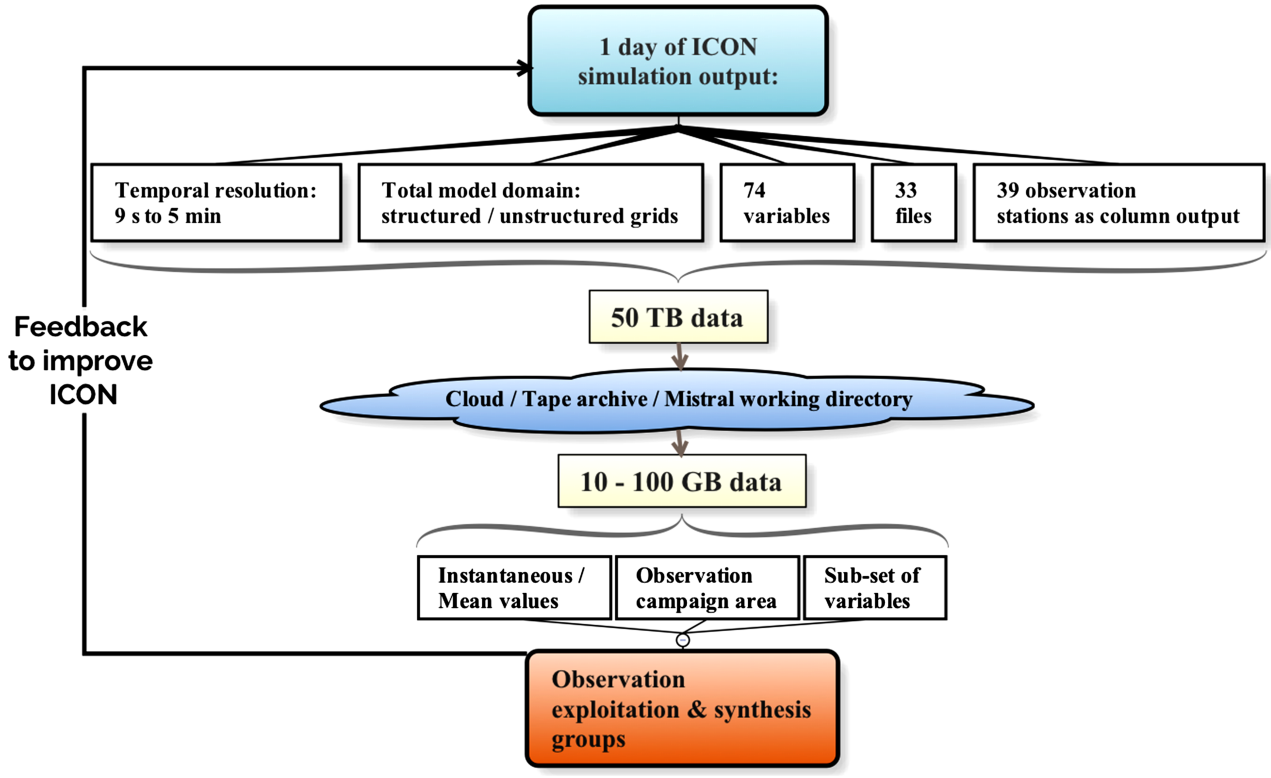

Figure 3. Data workflow in the HD(CP)^2 project.

Table 2. Overview of the forward operators used in the analysis.

Forward operator	Input from ICON/COSMO	Output from forward operator	References
VISOP	Temperature, pressure, humidity, cloud water and cloud ice content	0.6 and 0.8 µm MODIS and SEVIRI reflectances	Scheck *et al.* (2016)
SynSat	3D thermodynamic and hydrometeor fields, 2D fields of 2 m temperature, 2 m humidity, skin temperature	Infrared radiances for SEVIRI/MSG channels	Keil *et al.* (2006)
PAMTRA	3D thermodynamic and hydrometeor fields	Radar Doppler spectrum, effective radar, reflectivity factor, mean Doppler velocity, spectral width, brightness temperatures	Maahn *et al.* (2015)
EMVORADO (extended)	3D thermodynamic and hydrometeor fields	Horizontal and differential reflectivity, specific differential phase, cross-correlation coefficient, Doppler velocity	Zeng (2013)

Spectroradiometer (MODIS) satellite instrument (King *et al.*, 1992). It relies on the look-up table based Method for Fast Satellite Image Synthesis (MFASIS: Scheck *et al.*, 2016). The hybrid parallel VISOP code extracts the input parameters for the radiative transfer from the full-resolution model data along columns tilted towards the satellite position. Moreover, columns tilted towards the sun are extracted to compute cloud shadows. A similar strategy to that in Kostka *et al.* (2014) has been used to convert the model state into radiative transfer input variables, except for the parametrization of effective ice particle sizes, for which the method of McFarquhar *et al.* (2003) was adopted. For COSMO, the parametrized subgrid-scale clouds are also considered.

Infrared satellite data have been derived from the forward operator Synthetic Satellite imagery (SynSat: see e.g. Keil *et al.*, 2006; Eikenberg *et al.*, 2015; Senf and Deneke, 2017), which provides an interface to a radiative transfer model (RTTOV v11.2: Saunders *et al.*, 1999). The forward operator needs 3D fields of thermodynamic and hydrometeor variables, as well as surface fields, and simulates synthetic cloud-free and cloud-affected infrared radiances as observable by the *Meteosat Second Generation* Spinning Enhanced Visible and Infrared Imager (SEVIRI/MSG). SynSat has been applied in the operational setting, in which ice and snow content are combined into a frozen condensate content and the corresponding effective diameters are calculated following the bulk parametrization of McFarquhar *et al.* (2003).

The passive and active microwave radiative transfer radar simulator (PAMTRA: Kollias *et al.*, 2011; Maahn *et al.*, 2015) is applied to ICON and COSMO model output to allow for a comparison with cloud radar. For this comparison, the self-similar Rayleigh–Gans approximation has been used for the scattering of ice and snow particles (Hogan and Westbrook, 2014); for all other hydrometeors, Mie scattering theory has been used. Note that assumptions in the forward calculations, e.g. particle size distribution, mass–size relation, are identical to those in ICON and COSMO, respectively, and PAMTRA does not yet include a module for simulating the melting layer.

The Efficient Modular Volume RADar Operator (EMVORADO), described in Zeng (2013), Jerger (2014), Blahak (2016) and Zeng *et al.* (2016), is used to compare the 3D scan measurements from BoXPol and the C-band radar network of the German Weather Service (DWD) with the model simulations. The non-polarimetric operator EMVORADO (restricted to reflectivities and Doppler velocities) was extended in order also to simulate the pseudo-polarimetric radar observations using the T-matrix method (Mishchenko, 2000).

3. Observational data sources and selected days

This section provides an overview of the various observational data sources used in the ensuing sections and introduces the synoptic situations of the four selected days. Table 3 summarizes the various observations used to validate the ICON simulations.

Figures 4(a)–(d) give an overview of the synoptic-scale situations on the four selected days by showing the visible images of MODIS at the overpass times around noon of the polar-orbiting satellite *Aqua*, which hosts MODIS. These four days encompass different synoptic situations. For the first two days (24 and 25 April), high pressure dominated over Germany, which resulted in a rather calm, almost clear-sky day (24 April) and a day with mostly shallow cumulus (25 April). This situation changed noticeably on 26 April, as a frontal system passed from a northwesterly direction over Germany, accompanied by strong convection, deep clouds and precipitation. On 2 May, high pressure prevailed over Germany, with low- to mid-level convective clouds. Compared with 25 April, on this day stronger convection occurred, accompanied by thicker cloud layers in the eastern part of the domain (Figure 4(d)). Due to these different weather situations, in terms of cloud regimes in particular, ICON can be evaluated over a rather broad range of relevant conditions.

Figures 4(e)–(h) show the synthetic radiances for the solar MODIS channels generated from ICON model output. Note that ICON output on the highest resolved domain (156 m grid spacing) was used on all days, apart from 25 April (f), where the coarsest data (625 m) were used, as this data set is not available on the finer grids due to failure during output. Details about the method used to obtain synthetic radiances and a discussion of Figures 4(e)–(h) is presented in section 6.

4. Boundary-layer characteristics

Several output variables of ICON, as well as derived quantities within the planetary boundary layer (PBL) on a local scale, are evaluated using ground-based observations as well as radiosonde profiles. Further, the ICON simulations are compared with the coarser simulations by COSMO and with the semi-idealized LES simulations.

4.1. Near-surface temperature, humidity and winds

For a first assessment of the simulation quality, basic meteorological quantities, i.e. temperature and humidity at 2 m above ground as well as wind speed at 10 m of ICON and COSMO, are compared against the 196 DWD weather station measurements. All values during the daytime (0600–1800 UTC) over the four days 24–26 April and 2 May are included in the calculation of the box-whisker plots (Figure 5), the results for each individual day shown in Figure SI–3. The information of Figure 5 is also shown in Figure SI–4 as box-whisker plots of the bias of the two models versus observations. The stations are divided into three approximately equally sized areas, consisting of northern (north of 52.5 °N), central (between 52.5 °N and

Table 3. Overview of the different observations used in this study. Observations at JOYCE, LACROS and KITcube sites are in context to HOPE. The type denotes either surface (surf.), vertical (vert.), spatial (spat.) or volume scan (vol.) measurements. Parameter: actual temperature (T), specific humidity (q_v), absolute humidity (ρ_v), wind components (u, v, w), boundary-layer depth z_i, net radiation Q_0, surface sensible heat flux H_0, surface latent heat flux E_0, integrated water vapour (IWV), cloud water path (CWP), cloud water content (CWC), ice water content (IWC) and visual reflectance from satellite (reflectance). Daytime satellite overpass is within a few minutes across the full domain (sat overpass). Vertical and horizontal resolution are denoted by Δz and Δx, respectively. The RADOLAN products RY and SF are taken, with corrected shadowing effect and 24 h rain accumulation adjusted to rain-gauges, respectively (see also http://www.dwd.de/DE/leistungen/radolan/radolan.html). See text for details. MODIS resolutions are given for nadir, MSG for the approximate viewing angle.

Instrument/technique	Type	Parameter	Temporal characteristics	Spatial characteristics	References
DWD weather stations	surf.	T, q_v, u, v	10 min	Full domain, 196 stations	DWD
Tower	surf.	u, v	10 min	JOYCE, Lindenberg	Löhnert *et al.* (2015) and Beyrich and Adam (2007)
DWD radio soundings	vert.	T, q_v	Up to four times a day	Full domain, 11 stations, Δz: 10–30 m	DWD
Radio soundings	vert.	T, q_v	At least twice a day	KITcube site, Δz: 10–30 m	Kalthoff *et al.* (2013)
DWD ceilometer network, STRAT algorithm	vert.	z_i	1 h avg.	Full domain, 48 stations	Morille *et al.* (2007) and Haeffelin *et al.* (2012)
Ceilometer CHM15k	vert.	z_i	5 min avg.	Lindenberg, Δz: 15 m	Schween *et al.* (2014)
Doppler wind lidar HALO	vert.	z_i	5 min avg.	JOYCE, Δz: 30 m	Schween *et al.* (2014)
Raman lidar Polly^{XT}	vert.	z_i	30 s	LACROS, Δz: 30 m	Engelmann *et al.* (2016)
Raman lidar BASIL	vert.	z_i, $\overline{q'^2}$	1 min, 10 s	JOYCE, Δz: 30 m	Di Girolamo *et al.* (2009)
Doppler lidar WLS7-V2 ($z <$ 400 m), Doppler lidar Wind-Tracer WTX ($z \geq 400$ m)	vert.	$\overline{w'^2}$	2 h avg.	HOPE area, 3 locations, Δz: 25–60 m	Maurer *et al.* (2016)
Raman lidar RRL	vert.	$\overline{\theta'^2}$	10 s	KITcube, Δz: 109 m	Behrendt *et al.* (2015)
Differential absorption lidar (DIAL)	vert.	$\overline{\rho_v'^2}$	10 s	KITcube, Δz: 67.5 m	Muppa *et al.* (2016)
Energy balance stations	surf.	Q_0, H_0, E_0	30 min avg.	HOPE area, 5 stations	Maurer *et al.* (2016)
GPS network	surf.	IWV	15 min	124 stations	Gendt *et al.* (2004)[a]
Microwave radiometer	surf.	IWV	2 sec	JOYCE	Rose *et al.* (2005)
Sunphotometer	surf.	IWV	10 min	JOYCE	Alexandrov *et al.* (2009)
MODIS–FUB	spat.	IWV	Sat overpass	Full domain, Δx: 1×1 km^2	Diedrich *et al.* (2015b)
MODIS C6 (NIR)	spat.	IWV	Sat overpass	JOYCE, Δx: 1×1 km^2	Gao and Kaufman (2003)[b]
MODIS C6 (IR)	spat.	IWV	Sat overpass	JOYCE, Δx: 5×5 km^2	Seemann *et al.* (2003)
MODIS C6 (Vis+IR)	spat.	CWP	Sat overpass	Full domain, Δx: 1×1 km^2	Platnick *et al.* (2014)
SEVIRI/MSG (Vis+IR)	spat.	CWP	5 min	Full domain, Δx: 4×6 km^2	Roebeling *et al.* (2006)[c]
MODIS C6 (Vis)	vert.	Reflectance	Sat overpass	Full domain, Δx: 1×1 km^2, Δz: 250 m	Platnick *et al.* (2014)
Cloudnet	vert.	CWC	30 s	JOYCE, Δz: 30 m	Illingworth *et al.* (2007)[d]
CloudSat/DARDAR	vert.	IWC	sat overpass	Full domain, Δz: 60 m	Delanoë and Hogan (2010)[e]
SEVIRI/MSG (Vis)	vert.	Reflectance	15 min	Full domain, Δx:4×6 km^2	Platnick *et al.* (2014)[c]
Cloud radar MIRA	vol.	Precipitation	30 s	LACROS, Δz: 30 m	Görsdof *et al.* (2015)
X-band radar system BoXPol	vol.	Polarimetric moments	5 min	Bonn area, 100 m radial, 1° azimuthal, 10 elevations	Diederich *et al.* (2015a)[f]
DWD C-band radar network	vol.	Polarimetric moments	5 min	Full domain, 1 km radial, 1° azimuthal, 10 elevations	Helmert *et al.* (2014)
DWD radar network RADOLAN	surf.	Precipitation: RY and SF products	5 min, 60 min	Full domain, Δx: 1 × 1 km^2	DWD

Note: For data not obtained by the HD(CP)^2 project or by one of the institutions of the co-authors, the following data contributors are acknowledged: [a]GeoForschungsZentrum Potsdam, [b]MODIS data are obtained from the Level-1 & Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC), located in the NASA Goddard Space Flight Center, [c]EUMETSAT, [d]EU project Cloudnet (GA 654109), [e]ICARE data centre, Université de Lille, [f]Transregional Collaborative Research Centre 32 (German Research Foundation).

49.9 °N) and southern Germany (south of 49.9 °N), to analyze the spatial distribution of model performance. The original time resolution of the datasets is 10 min for the observations, 15 min for COSMO and 10 s for ICON. For comparison, all datasets are averaged to a common 15 min time resolution. The maps in Figure SI–2 show the spatial distribution of both the bias and the standard deviation (STD) for selected stations and Table SI-2 lists these quantities as averages within the PBL for some of these stations for temperature and humidity. Box and whisker plots of the bias are also shown in Figure SI–4. Here, the bias and STD are the mean and, respectively, the standard deviation of the difference between model and observational data taken for at each individual 15 min time slice. While the bias means a systematic deviation of the model results from the observations, the STD is a measure for random forecast errors. Inspection of the geographical distribution of the bias reveals a mainly latitudinal dependence, while the STD varies more strongly, but irregularly, throughout the domain.

Broadly, the distributions of the 2 m temperature of ICON and COSMO match the observations well for a large temperature range of almost 2–27 °C. The spread of the different box plots is of similar magnitude and the spatial variability is well represented by the models. However, a north–south gradient of the median and the distributions is visible for both models. In northern Germany, there is a positive bias by ICON of up to 1 K and even more by COSMO, with roughly 3 K. In contrast, ICON underestimates the temperature by 2 K and COSMO by approximately 1 K in southern Germany. The lowest differences are found for central Germany, where COSMO fits the observations almost perfectly, while ICON shows an underestimation by 1–2 K. The too-large sensible and latent heat flux at the HOPE area (see section 4.4 and Figure 10 later), which is within the central subregion, might be an explanation for the underestimation. For comparison, the heat fluxes of COSMO are better represented, which corresponds to a very good match of the 2 m temperature distribution. The outliers at the southern subregion of the observations belong to

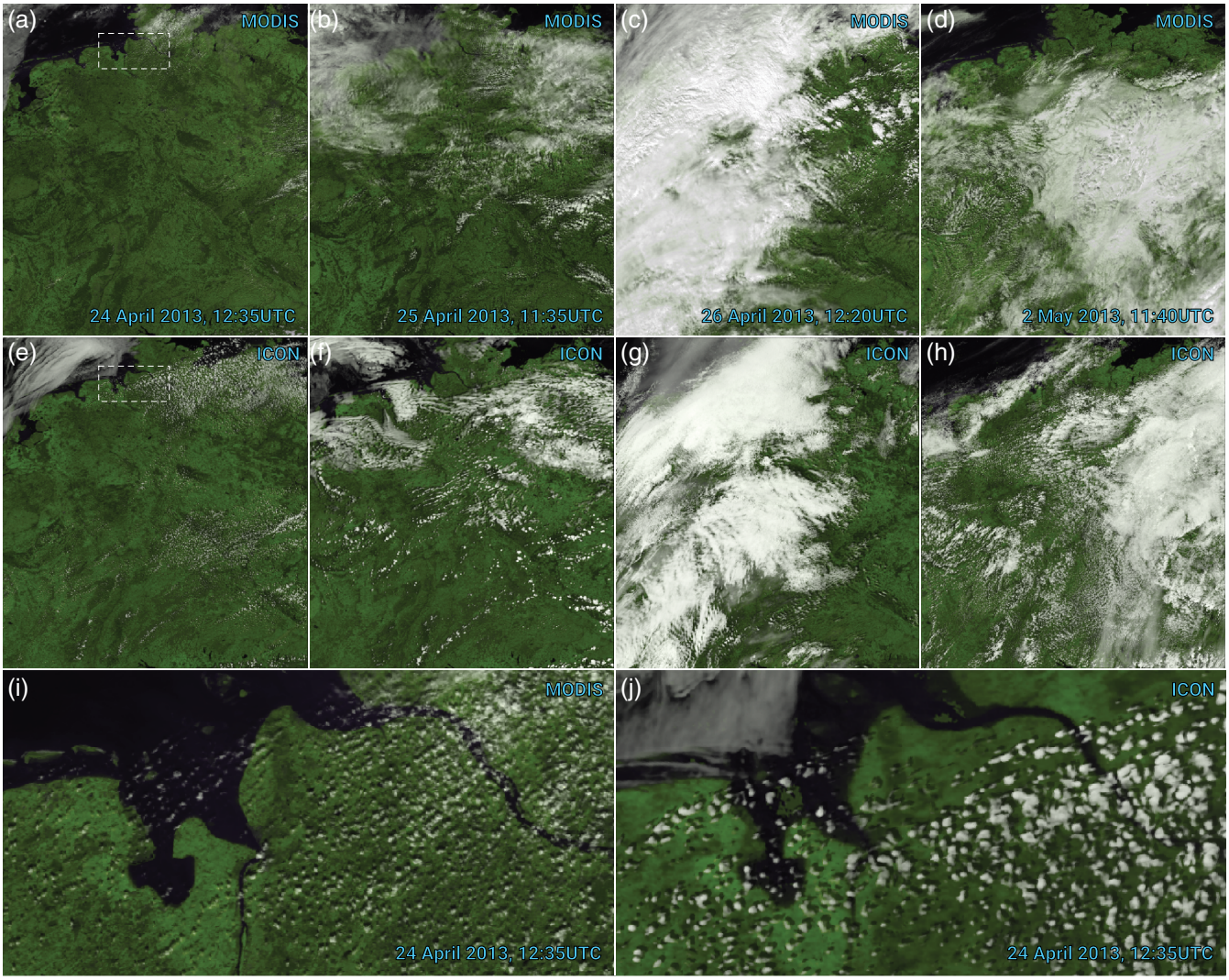

Figure 4. Snapshots of visible images of MODIS and synthetic radiances for the solar MODIS channels produced based on ICON output. The upper row (a–d) shows snapshots of visible images of MODIS over the model domain on the four selected days at overpass times around noon. The middle row (e–h) contains the respective synthetic images produced with ICON 3D model output on a 156 m grid for (e), (g) and (h) and on 625 m grid for (f). The lower row (i, j) contains a zoom into the North Sea coastal region on 24 April, depicted as the white dashed box in panels (a) and (e). In these colour images, the 0.6 µm reflectance, R_6, was used for the red channel, the 0.8 µm reflectance, R_8, for the green channel and $0.5(R_6 + R_8)$ for the blue channel. Invalid R_8 values, which can be encountered in bright clouds due to detector saturation, were replaced by the corresponding R_6 values.

stations in more complex terrain, which are not captured by both models.

The 2 m specific humidity of ICON and COSMO generally shows a high bias compared with the observations. The variability of the near-surface humidity is well represented by both models but shifted towards higher specific humidity values. As for the temperature, a north–south gradient is seen by the largest overestimation of up to 1.5 g kg^{-1} by ICON for the northern subregion and by up to 1 g kg^{-1} by COSMO. For the central and southern subregions, the simulated distributions match the observations much better and only a small overestimation of less than 0.8 g kg^{-1} is seen. The errors of ICON and COSMO are of similar magnitude, but ICON performs somewhat better (only approx. two-thirds of the bias) for central and southern Germany. The outliers of the observations again belong to more complex terrain.

For the 10 m wind speed, the distributions of ICON are too broad and shifted towards higher wind speeds compared with the observations. For COSMO, the width of the distribution fits the measurements much better, but a small underestimation of the wind speed by about 1 m s^{-1} is seen. The northern subregion shows the best match for ICON and COSMO, but, especially for ICON, many outliers are found at high wind speed. The largest differences are seen for the southern subregion, where e.g. the median of ICON is up to 3 m s^{-1} higher than the observations and also many outliers exist at high wind speeds. One reason for this might be problems with the right representation of strong orographic regions (Alps, Black Forest). In contrast to the overestimation of ICON, COSMO shows only a slight underestimation and the width of the distribution matches the measurements well. Overall, for the wind speed, COSMO shows a better performance, which could indicate the need of improving the turbulence closure and the land-surface coupling in ICON further.

On 26 April, when high wind speeds were observed, wind gusts are evaluated during the daytime in ICON 312 m and COSMO. Here, gusts are defined grid-box-wise as maxima of the 10 m wind speed over 15 min intervals. The models show similar spatial distributions of daily gust maxima; however, local differences are seen close to the western domain boundary (Figure SI–6). The gust intensities derived from the model output are compared with daily peak values from tower measurements, where gusts are defined as maxima of 10 m wind speed over 10 min intervals. The time evolution of the gust maxima over 15 min are shown for the supersites JOYCE and Lindenberg in Figure 6. When the front passes over the respective locations, the peak intensities of COSMO and ICON are almost identical and fit the observations well. However, intensities increase more slowly in ICON than in COSMO, which fits observations at JOYCE better, but those at Lindenberg worse.

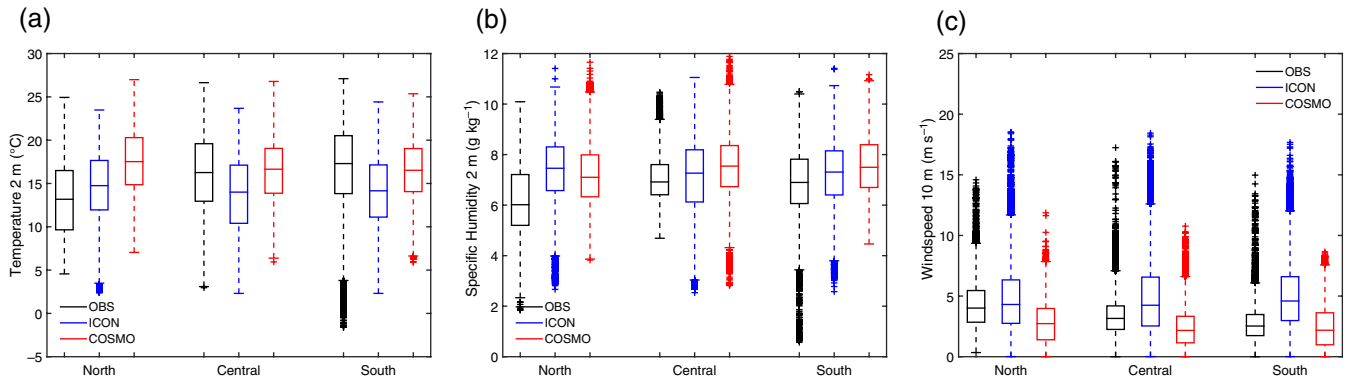

Figure 5. Box-whisker plots of (a) 2 m temperature, (b) 2 m specific humidity and (c) 10 m wind speed at the DWD weather stations. Observations are shown in black, ICON results in blue and COSMO results in red. The analysis covers all observation times between 0600 and 1800 UTC for 24–26 April and 2 May. The locations are divided into three subregions of northern (left three boxes), central (middle three boxes) and southern (right three boxes) Germany. The central line shows the median, the box the middle 50% of the distribution and individual values ('outliers') that fall above or below ±2.7 standard deviations (limits shown by horizontal bars) of the data are marked by crosses.

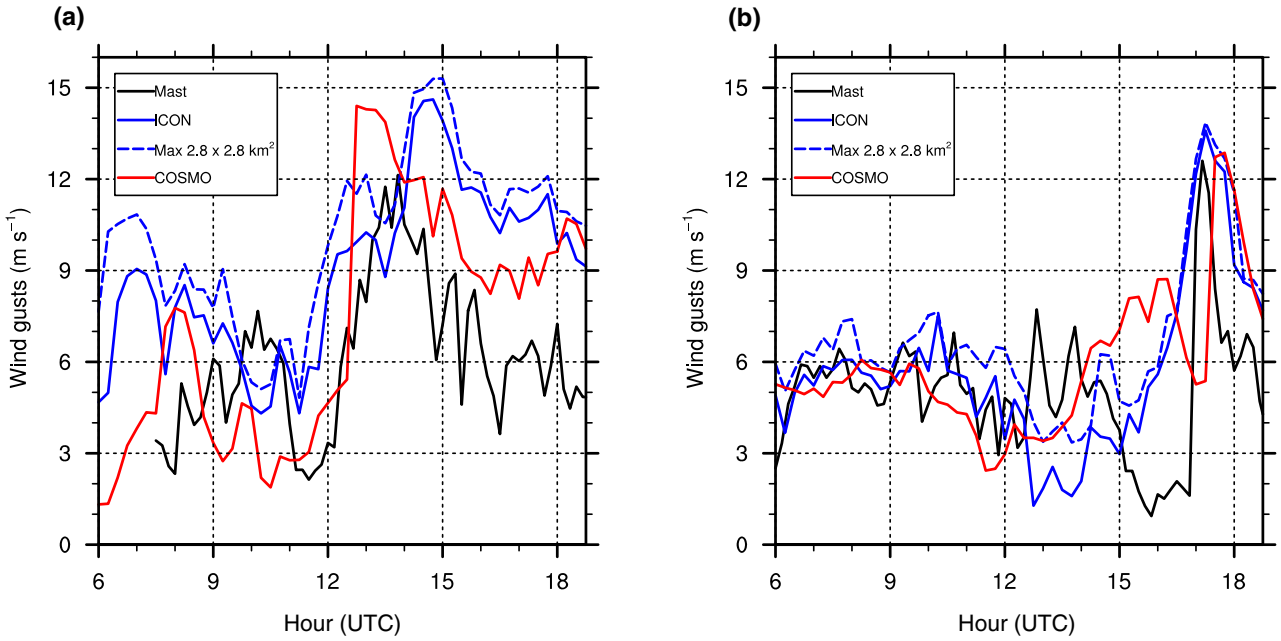

Figure 6. Time evolution of wind gust intensity (maximum 10 m wind speed over 15 min time intervals) for ICON (blue) and COSMO (red) on 26 April and tower (mast) observations (black) at (a) JOYCE and (b) Lindenberg. For ICON, the maximum over a 2.8 × 2.8 km^2 area around the measurement sites is also shown (dashed blue).

4.2. *Thermodynamic profiles*

Figure 7 shows exemplary profiles of temperature and moisture of ICON and COSMO simulation results compared with radio soundings for the two days with cumulus clouds, 25 April and 2 May, both for KITcube at 1700 UTC. On 25 April, both models show similar problems in simulating the PBL, with too low temperatures and – in consistency with the near-surface observations shown in Figure 5 – a too moist PBL. However, the bias is stronger in ICON (temperature about 2–4 K too cold and specific humidity 2 g kg^{-1} too high). Both models simulate a too low PBL height (about 1200 m above ground). The ICON profile is too stable, with potential temperature continuously increasing with height. In contrast, 2 May is simulated far better in both models. Only in the lowest levels is COSMO too cold and too moist, while ICON performs well even there.

The averaged biases and standard deviations (STD) within the PBL for a variety of stations (see Table SI-2) confirm the generality of the features shown by the example soundings of Figure 7, with a too cold and too moist PBL of ICON and COSMO. Biases are of similar magnitude for both models, although COSMO shows slightly more skill (lower errors for COSMO in 30 cases, compared with 18 cases for ICON).

4.3. *PBL height*

The planetary boundary-layer height can be regarded as an integral quantity to evaluate the quality of the representation of surface processes on a local scale. There are several definitions for the height of the PBL. An overview of observing-system methodologies for thermodynamic profiling and their scientific applications is given in Wulfmeyer *et al.* (2015). For our analysis of the model data, we define the PBL height as the lowest level where the bulk Richardson number exceeds the value 0.28 (e.g. Richardson *et al.*, 2013). Time series of model results from ICON and COSMO for the four days 24–26 April and 2 May are compared against observations from different instruments at the supersites JOYCE and Lindenberg (Figure 8). Note that the derivation of PBL height in observations relies on different principles, leading to differences between dynamically and tracer-based techniques (Schween *et al.*, 2014): for the wind lidar HALO, the vertical velocity variance is used as a criterion, while the maximum variance of water vapour is used for the Raman lidar BASIL and criteria based on aerosol layering are used for the aerosol RAMAN lidar Polly^{XT} (Baars *et al.*, 2008) and the ceilometer CHM15k. Only for the radio soundings can the same method as for the models be applied, except that erroneous surface

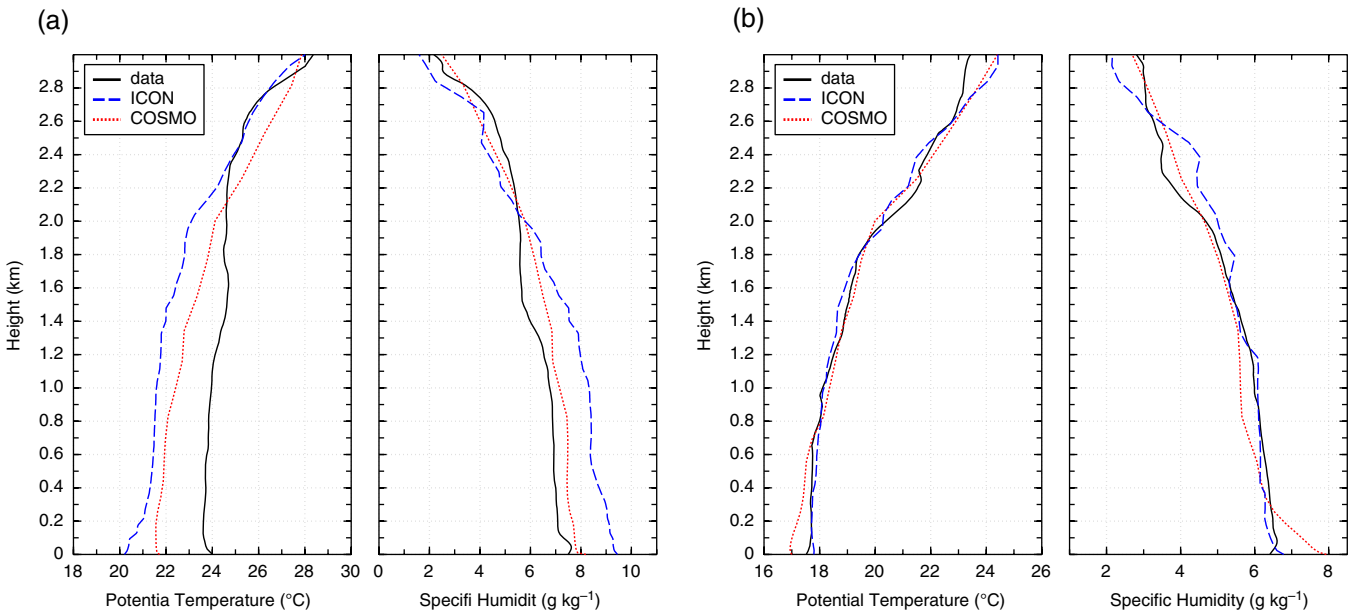

Figure 7. Simulated and observed vertical profiles of potential temperature and specific humidity at the KITcube supersite for (a) 25 April (1700 UTC) and (b) 2 May (1700 UTC).

values of the sounding data had to be excluded by detecting the critical bulk Richardson number above 100 m only.

For the HOPE region, the various measurements show a considerable spread, giving an approximate corridor for the most likely boundary-layer depth, which could be explained by the different locations and methods of the instruments. The different locations are characterized by different land-use types in this rural region, which is characterized by patchy fields of about 200 m size dominated by bare soil (sugar beet to be planted) and winter wheat with already well-developed plant cover, as well as small settlements and forest areas. For 24––26 April and 2 May, ICON shows higher PBL height than COSMO and is closer to observations. On 26 April, ICON and COSMO are similar, but do not show a clear boundary-layer signal, due to the frontal passage. The semi-idealized ICON shows higher peak heights compared with PALM, although the applied large-scale forcing is exactly the same. However, the implementation of the Monin–Obukhov surface flux calculation is different in these models, which possibly explains the difference. Comparing the peak sensible and latent heat fluxes of PALM and ICON for the four days, one can see that they are considerably higher in ICON than in PALM (about 100–200 W m^{-2}; not shown), thus providing more energy input at the surface. This may result in larger thermals, stronger turbulence and a deeper PBL.

Overall, the models are inside the spread shown by the measurements. In contrast to the HOPE site, at the Lindenberg location, COSMO shows a higher PBL than ICON and is closer to ceilometer observations on 24–26 April. However, especially on 26 April, the performance of ICON is worse than COSMO, in the sense that the drop in PBL depth due to the frontal passage seems to be too early in comparison with ceilometer and radiosonde measurements; on the other hand, ICON is closer to the radio sounding at 1800 UTC. On 2 May, ICON and COSMO perform equally well, but the observed increase in the afternoon is not caught by either model.

Besides looking at time series of PBL heights at fixed locations, we are also interested in spatial variations at a fixed point in time, here at 1100 UTC (local solar noon) on the four days 24–26 April and 2 May (Figure 9). The same definition as for the time series is applied to the ICON, COSMO and sounding observations to determine the PBL height. For the observations, the 11 available DWD soundings plus the HOPE KITcube soundings are used. Additionally, 48 DWD ceilometer stations are utilized for the comparison, where the PBL height was derived using the STRAT method (Morille *et al.*, 2007; Haeffelin *et al.*, 2012). The model values of ICON and COSMO are averaged over a circle of 5 km radius, as in Figure 8. The model domain is divided into three approximately equally sized areas, as in Figure 5. The distributions of PBL heights of the models and observations are shown by individual box-whisker plots for each of these three regions (Figure 9). Maps of selected stations showing the spatial distribution are shown in Figure SI–5.

Overall, a good match of models and observations is found for 24 and 25 April and 2 May, with deviations around ±200 m throughout the model domain (except for the northern region on 2 May, see below). Mostly, the errors of ICON and COSMO are similar in sign and magnitude. The spatial variability is well represented by both models. There does not seem to be any systematic over- or underestimation of the PBL height and there is little geographical variation in errors. Nevertheless, on 2 May, ICON and COSMO underestimate the PBL heights for northern Germany by up to 1000 m compared with the observations. This could be the result of the simulated clouds (see Figure 4(h)), which are not seen by the observations (see Figure 4(d)). Moreover, on 26 April, where a frontal passage took place, a large underestimation for all three subregions by ICON of more than 1500 m in the median is also visible, consistent with Figure 8. Also, COSMO largely underestimates the PBL height for central and southern Germany. In addition, for the northern subregion, ICON underestimates and shows almost no variability at the PBL height, whereas COSMO significantly overestimates the PBL height and shows a large variability of the values.

These results show the importance of including a variety of weather situations, as systematic uncertainties in the COSMO model, such as the 20% underestimation in PBL height found by Baars *et al.* (2008) and the underestimation of PBL moisture found by Crewell *et al.* (2008), refer to much longer time-scales.

4.4. *Surface energy balance*

The surface energy balance largely impacts the properties and time evolution of the PBL. It is mainly a result of land–atmosphere (PBL) feedback processes. Therefore, errors in the surface fluxes can also be due to errors in the simulation of PBL moisture, temperature and dynamics. We compare the different components of the surface energy balance with observations from energy-balance stations over the HOPE area around Jülich, i.e.

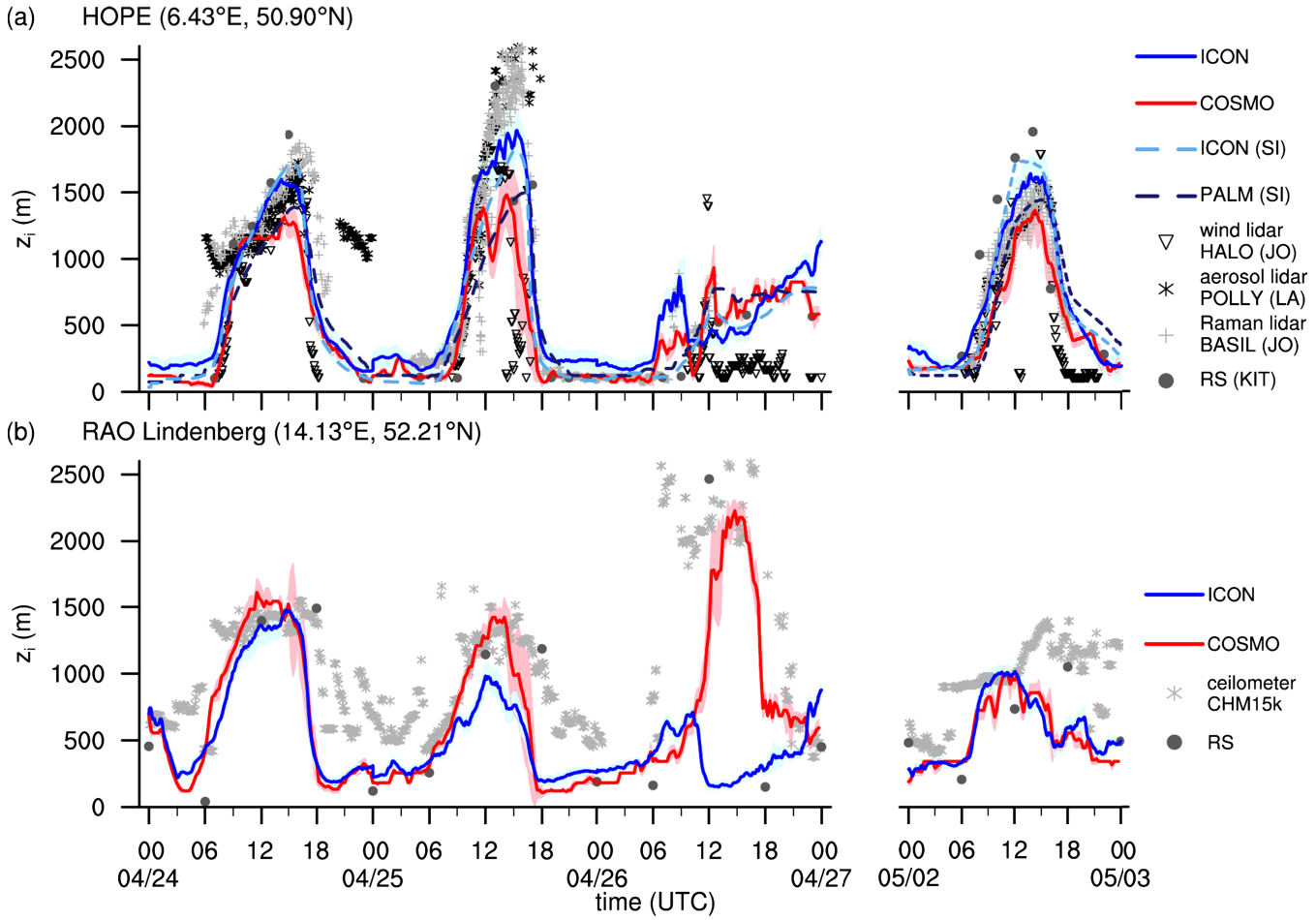

Figure 8. Temporal evolution of the boundary-layer depth z_i at (a) the HOPE site and (b) RAO Lindenberg, for all four simulated days. The boundary-layer depth is determined based on a bulk Richardson number criterion in all three models (ICON, COSMO, PALM) and in the radiosonde data. In ICON and COSMO, a spatial average with a radius of 5 km centred around (6.43°N, 50.90°E) and (14.13°N, 52.21°E) for HOPE and RAO, respectively, was used. Wind, aerosol and Raman lidar took measurements at the JOYCE (JO) or LACROS (LA) sites in the HOPE domain. Radiosondes were launched at the KITcube site (KIT) during the HOPE campaign. For the HOPE site, semi-idealized (SI) LES were performed with ICON and PALM, where boundary-layer depth is provided as a spatial average over the total horizontal model domain. Blue and red shading denote twice the standard deviation of z_i in ICON and COSMO, respectively. Note that Polly^{XT} aerosol lidar data is missing on 2 May.

the surface net radiation, as well as latent and sensible heat fluxes. For this area, at least five energy balance stations were available. They reflect the observed spatial variability of the surface forcing connected to the different land uses (Maurer *et al.*, 2016) and are thus suited for a well-founded model validation. Diurnal cycles of the three days 24 and 25 April and 2 May are shown in Figure 10. 26 April has been excluded, due to incomplete simulation output, which does not impact our conclusions as the surface forcing is presumably less important on that day because of the frontal passage.

Simulated net radiation of ICON fits well with observations concerning both absolute values and diurnal cycle for 24 and 25 April. On 2 May, which is a day with boundary-layer clouds, net radiation is overestimated, as most points are cloud-free in the model. The 5% percentile, however, indicates that at least some grid boxes fit the observations well. COSMO also shows good agreement with observations on 24 April, but an underestimation of net radiation due to cloud occurrence on the afternoon of 25 April. Like ICON, there is an overestimation of net radiation on 2 May, but of a smaller magnitude. Both latent and sensible heat fluxes seem to be overestimated by ICON by about 50 W m^{-2} on 24 April, which was a cloud-free day. They are even more overestimated on 2 May, which is consistent with too high net radiation. In contrast, COSMO mainly produces only too high latent heat fluxes. This could be due to incorrect parametrizations of surface fluxes, but also erroneous representation of PBL variances in the models. Both models reproduce the observed spatial variability.

4.5. *Turbulence profiles*

For the evaluation of turbulence profiles simulated by ICON, we compare the variances of temperature, specific humidity and vertical velocity with lidar observations at the KITcube site (Figure 11). We chose two hours around local noon time of a clear-sky day (24 April, 1200–1400 UTC) for this purpose. To improve comparability of the results, we scaled the z-axis with the boundary-layer height z_i and the x-axis with the convective velocity/temperature/humidity scale w_*, θ_*, q_*, respectively. The values of these quantities have been calculated separately for the three ICON domains, the two semi-idealized simulations and the observations (Table 4). The length of the time period of 2 h is a compromise between two competing demands. On the one hand, the time period should be long enough to ensure an adequate sample size to prevent large systematic errors, as described by Lenschow *et al.* (1994). On the other hand, the time period should be short enough that the variations of PBL height are small and the scaling of the z-axis is consistent. For the 2 h period, the temporal variation of PBL height is less than 10% for both model output and observations, which is considered to be small enough. Behrendt *et al.* (2015) showed that temperature higher-order moments can be derived from the high-resolution temperature measurements of the University of Hohenheim (UHOH) Rotational Raman lidar (Hammann *et al.*, 2015). The specific humidity fluctuations are obtained from high-resolution measurements of the UHOH water-vapour Differential Absorption Lidar (DIAL) (Muppa *et al.*, 2016; Späth *et al.*, 2016). As for

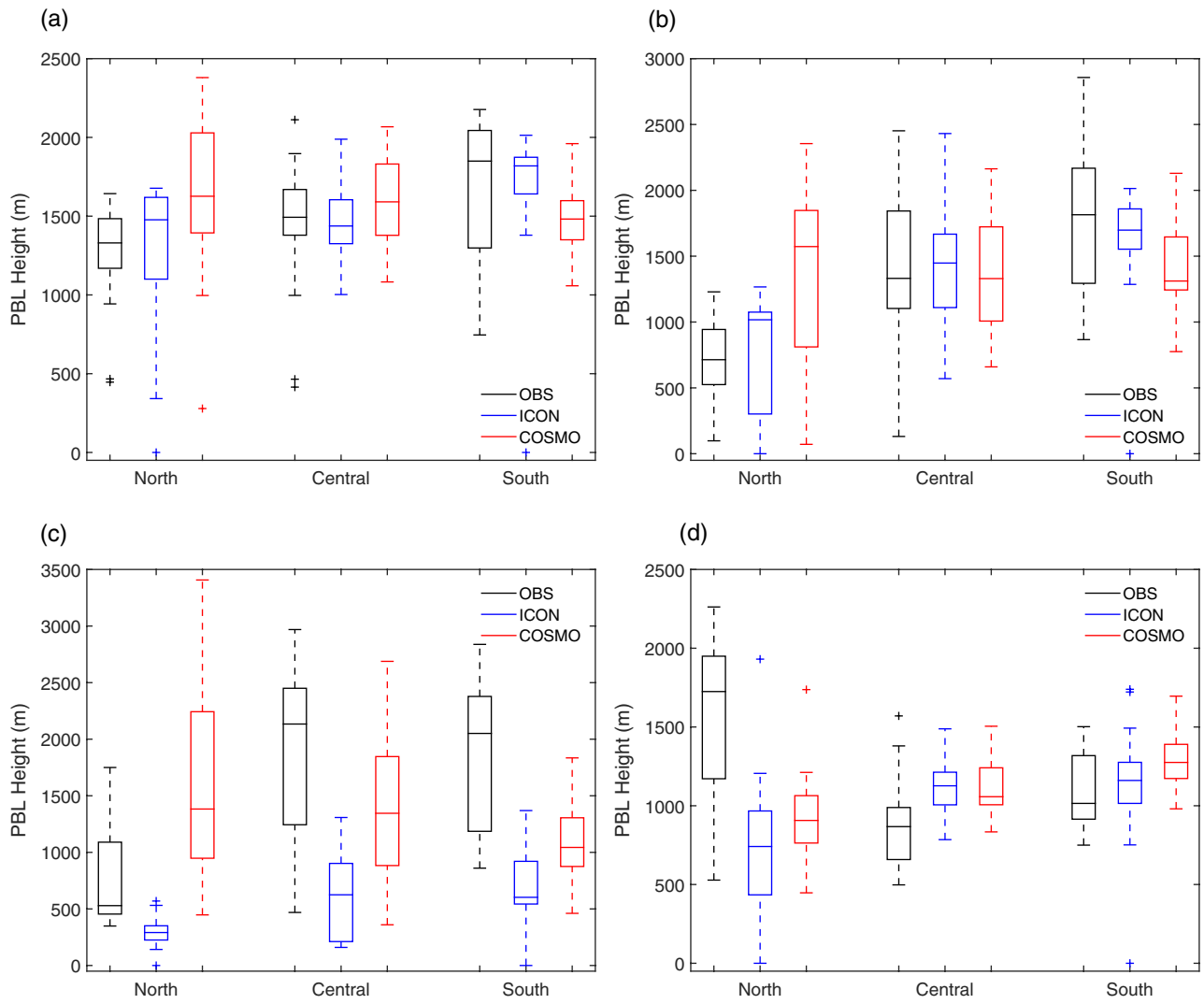

Figure 9. Box-whisker plots of the PBL heights for observations by radiosondes and the ceilometer network, ICON and COSMO for 1100 UTC on (a) 24 April, (b) 25 April, (c) 26 April and (d) 2 May. The data is divided into the three regions of northern (three left boxes), central (three middle boxes) and southern Germany (three right boxes). Box-whisker plots are defined as in Figure 5.

previous studies, we used the procedure of Lenschow *et al.* (2000) for analyzing the higher order moments of the lidar data as well as their errors. Here, we scaled the variance profiles with the surface scaling traditionally used in the literature (Lenschow *et al.*, 1980).

Especially for the vertical wind variance, a clear convergence towards the Doppler lidar derived profiles with increasing model resolution is seen. Deviations between the lidar-derived variance profiles and the simulations cannot be explained by a loss of variance in the observations. Due to their high vertical and temporal resolution, it can be demonstrated by auto-covariance and spectral analyses that the variance was almost completely resolved. The 156 m resolution result almost agrees with the lidar observation above approximately $0.4z_i$, although D15 examined variances of vertical wind as simulated with ICON at various resolutions. Only for a very high resolution of 25 m did the results agree well with Direct Numerical Simulations (DNS). Nevertheless, the best agreement with the observations is shown by the semi-idealized simulations with 50 m resolution. For completeness, the idealized profile of Lenschow *et al.* (1980) is shown in Figure 11(e). Its deviation from the measured variance illustrates the remaining uncertainty of single-point variance profiles due to the limited sample size as well as the accurateness of measured scaling variables.

For temperature, no large differences can be seen for the three resolutions, but the model data show much lower temperature variances than the Raman lidar data within the PBL. This possibly indicates remaining issues of the models with respect to the simulation of temperature fluctuations in the entrainment layer near the PBL top. However, ICON at its highest resolution agrees well with the semi-idealized simulations, which also show lower temperature variance compared with the lidar.

For the specific humidity variance, there is a resolution dependence, but a consistent convergence cannot be seen. The specific humidity variance profile resolved from the ICON 625 m domain output is larger than the ICON 312 m domain profile. The peak values of ICON 156 m and ICON 625 m are similar and agree well with the lidar observations, but for ICON 312 m the peak value is too low. Furthermore, the locations of the peaks do not agree between model and lidar retrievals. This points to a large uncertainty of the specific humidity variance at that resolution. Moreover, it is well known that surface scaling is not sufficient to explain the turbulence statistics from the mixed layer up to the entrainment layer (EL). Therefore, deviations between the variance maxima in the EL may not be due to differences in surface forcing but rather different gradients of temperature and moisture in the EL, as well as different wind shear and gravity-wave activities. Suitable scaling variables in the EL were proposed in Wulfmeyer *et al.* (2016).

Unlike for the vertical wind variance, the semi-idealized simulations disagree in the variance profiles for temperature and specific humidity, with PALM showing higher variance than ICON. This could be due to the different turbulence schemes in the two models. Especially for specific humidity, variance of both semi-idealized simulations is lower than for the lower resolved

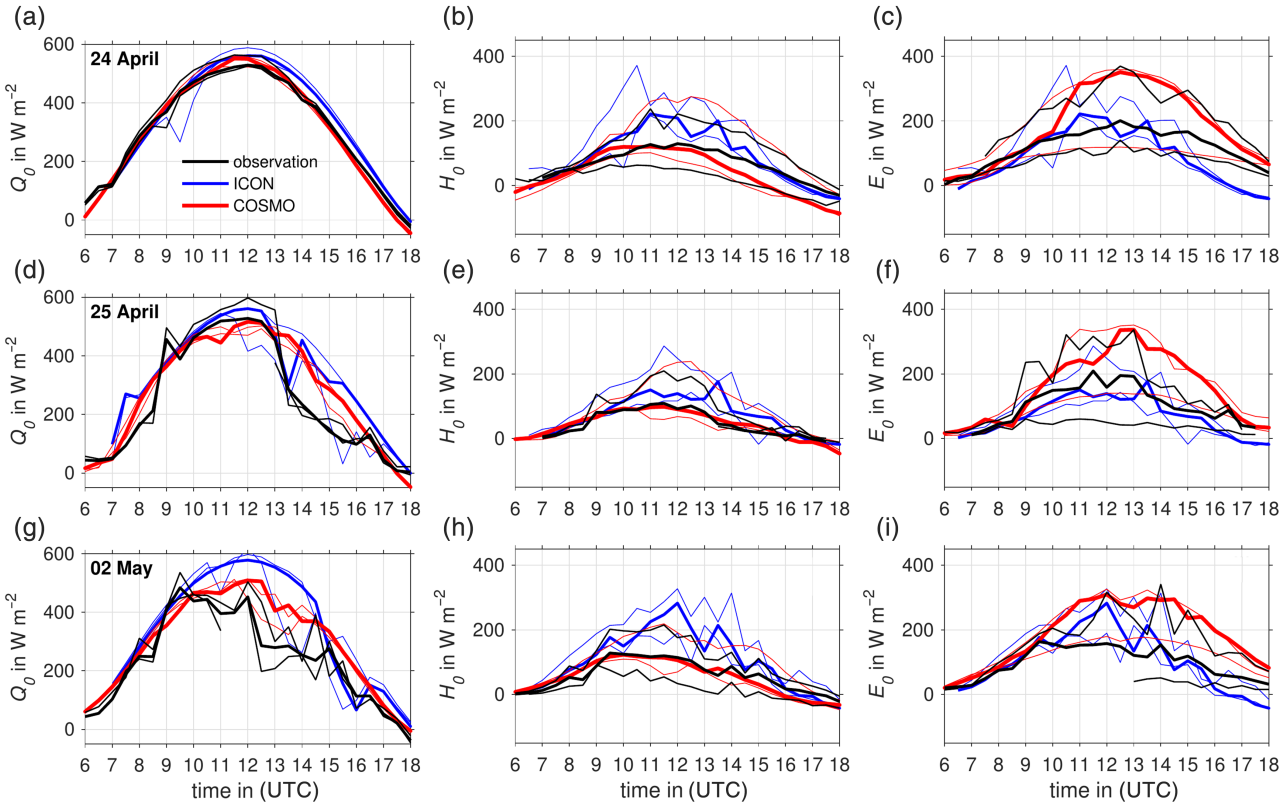

Figure 10. Comparison of energy-balance components: (a,d,g) net radiation Q_0; (b,e,h) surface sensible heat flux H_0; and (c,f,i) surface latent heat flux E_0 of ICON and COSMO simulations, including observations from energy-balance stations in the HOPE area (Tereno and KITcube) for three days. All values are averages of 30 min. For the observations, the thick line corresponds to the weighted-averaged values and the thin lines to the energy-balance stations with daily minimum and maximum values, respectively; for ICON and COSMO, the thick lines correspond to the median in an area of 30 × 30 km^2 around the HOPE stations and the thin lines to the 5 and 95% percentiles.

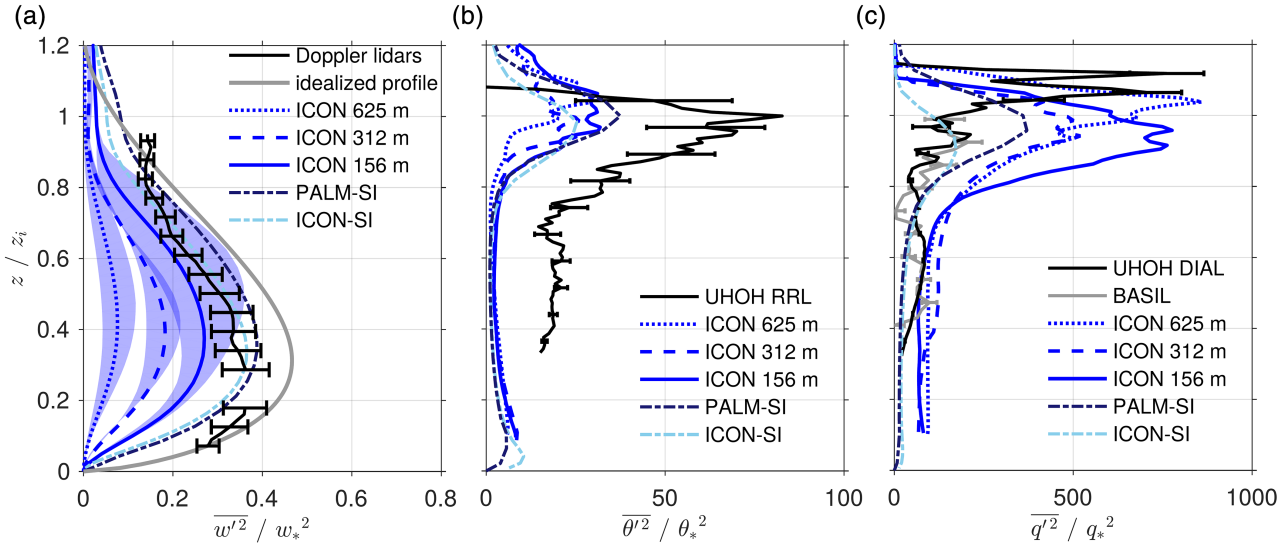

Figure 11. Normalized variance profiles of (a) vertical velocity, (b) potential temperature and (c) specific humidity for ICON simulations (three domains) and the semi-idealized simulations with PALM and ICON, for 24 April, 1200–1400 UTC (1 grid point for each domain), in comparison with variance profiles derived from lidar observations during HOPE, with error bars showing the statistical uncertainty according to Lenschow and Wyngaard (2010); shaded areas in (a) denote the spatial variation of simulated variances within a range of about 10 km.

ICON simulations, probably due to the absence of soil-moisture variations and the missing topography in the semi-idealized simulations.

4.6. *Energy spectra*

Energy spectra provide additional information on turbulence within the simulations. We calculated the spectra for the three ICON simulations at an instant in time on 26 April, 1200 UTC, as shown in Figure 12. Model output for the three resolutions is interpolated to a latitude–longitude grid of approximately 156 m resolution (corresponding to the highest model resolution) and detrended. The spectra are integrated radially over identical areas and averaged over five levels in the free troposphere.

All simulations reproduce the theoretical Kolmogorov −5/3 spectrum for a scale of $\mathcal{O}$(10 km). At the synoptic scale (> 100 km), which is prescribed by the forcing, spectra follow a power law with the exponent −3. As a rough definition, we identify the effective resolution at the point when the spectra fall below 50% of the theoretical Kolmogorov spectrum, which is about eight times the grid spacing for the ICON simulations and five times for COSMO. This is typical for grid-point models

Table 4. Values derived for the boundary-layer height and the convective velocity, temperature and humidity scales, for the three ICON domains, the semi-idealized (SI) simulations with PALM and ICON and the observations.

	z_i (m)	w_* (m s^{-1})	θ_* (K)	q_* (mg kg^{-1})
ICON 625 m	1315	2.03	0.11	42.2
ICON 312 m	1385	1.97	0.10	46.6
ICON 156 m	1424	1.99	0.10	46.0
PALM-SI	1175	2.03	0.10	33.4
ICON-SI	1528	2.39	0.11	48.2
OBS	1395	1.66	0.07	37.6

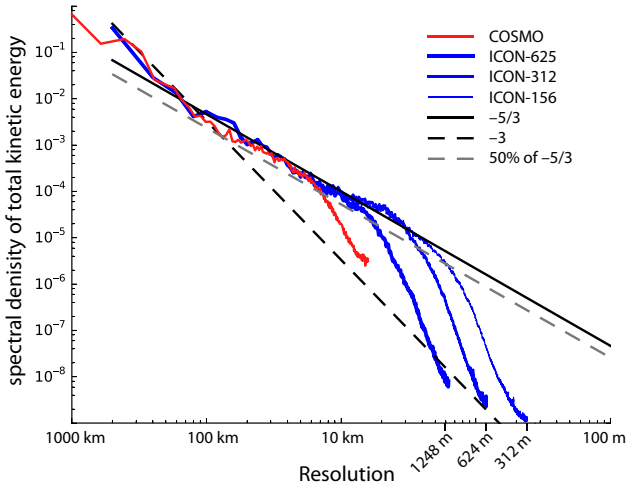

Figure 12. Spectra of total kinetic energy of the three nested domains and COSMO on 26 April at noon. The black solid and dashed lines indicate the power laws with the exponents −5/3 (Kolmogorov spectrum) and −3 (synoptic spectrum), respectively. The grey dashed line indicates a value 50% below the Kolmogorov spectrum.

with a C-grid and has been found for global weather prediction models (Z15), as well as for idealized LES (D15).

5. Water vapour

For the evaluation of temporal variations of integrated water vapour (IWV), a time series for the days 24–26 April and 2 May 2013 was used at the supersite JOYCE, where various ground-based observations of IWV were also available. To evaluate IWV on the entire simulation domain, it is compared with IWV measurements of the global positioning system (GPS). Furthermore, simulations from COSMO and IWV retrievals from MODIS on board the polar-orbiting satellite *Aqua* are also used as references.

5.1. *Temporal evolution and variability*

Figure 13 shows the temporal evolution, as well as small-scale temporal variability of IWV at JOYCE. The temporal resolution of the various instruments and models is as follows: GPS 15 min, microwave radiometer (MWR) 2 s (Rose *et al.*, 2005), sunphotometer 10 min (Alexandrov *et al.*, 2009), radiosonde several times daily, MODIS near-infrared (NIR, Gao and Kaufman, 2003) and infrared (IR, Seemann *et al.*, 2003) at the *Aqua* and *Terra* overpass times and COSMO 15 min. The temporal resolution of the time series at the ICON grid points is 5 min during the spin-up phase of the model, 0000–0600 UTC for 24 and 25 April and, due to output problems, 0000–1800 UTC on 26 April. During the remaining time, the resolution is 10 s. For ICON and COSMO, the IWV value of the nearest model grid point to JOYCE was taken.

On 24 April 2013, IWV increases constantly by about 10 kg m^{-2} from approximately 15 kg m^{-2}, with strong small-scale variations after noon (1200–1500 UTC). Afterwards, IWV is more or less constant until noon on 25 April and starts to decrease subsequently. On the third day, there is a strong increase of IWV from 0100 UTC to 1300 UTC from 17 to 28 kg m^{-2}. The IWV decreases to 17 kg m^{-2} until the end of 26 April. On 2 May there is no strong tendency in IWV. It varies between 17 and 22 kg m^{-2} during the day. In general, both models, ICON and COSMO, match the measurements well in terms of the temporal evolution. Most of the time, the models lie within the range of the various IWV measurements. However, the small-scale temporal variability of IWV, which can be seen in the measurements of the MWR, is only captured by ICON due to the higher temporal resolution. This is demonstrated more clearly in the zoomed-in time period, shown in the upper panel of Figure 13, during which strong small-scale temporal variability was observed. While the high variations on 24 April occur at the same times on 25 April, ICON shows high variability a few hours earlier than the MWR measurements.

5.2. *Spatial distribution*

Figure 14 shows the MODIS–Freie Universität Berlin (FUB) IWV distribution (Diedrich *et al.*, 2015b), with a horizontal resolution of 1 km at nadir, and ICON and COSMO simulations at the MODIS/*Aqua* overpass time at about 1230 UTC on 24 April 2013. The IWV fields are overlaid with IWV derived from GPS observations (Gendt *et al.*, 2004) provided by the GeoForschungsZentrum (GFZ) Potsdam. Note that the observed IWVs are not height-corrected; however, only GPS stations with a height difference lower than 20 m and located at a height below 400 m above sea level are used. There is no dependence of bias due to any height difference. Further quality checks were applied to the GPS data to exclude erroneous stations due to e.g. wrong meteorological data or receiver problems. This resulted in 124 GPS stations out of the available 400 in the dataset used for comparison.

The MODIS–FUB IWV retrieval is only applied to pixels identified as clear-sky and over land surfaces. Retrieved low values of IWV around cloudy areas may be due to optically thin clouds or sub-pixel clouds not correctly masked out, e.g. in the northeast corner of the domain. IWV retrievals near water surfaces can also be problematic and may lead to too high IWV values.

The observations show that IWV values within the domain vary from about 5 to 25 kg m^{-2}, but also that geographical variability of the IWV field occurs at large scales. For most stations, GPS shows slightly higher IWV values than MODIS. In comparison with the observations from MODIS and GPS, the simulated horizontal distributions of IWV look very similar, with the highest values in the northwest and lowest values in the south of the model domain. Some regional differences can be identified between the observations and the models, e.g. in the eastern part of Germany both models underestimate the IWV, while further north towards the coast the IWV is overestimated. Overall, both ICON and COSMO appear to have similar biases when compared with both MODIS and GPS. The high spatial resolution of ICON is, for example, clearly visible in the northwest, where small-scale variability in the IWV field with high values up to about 25 kg m^{-2} are simulated, which is not visible in the COSMO IWV field.

Table 5 shows the bias and bias-corrected root-mean-square error (RMSE) between IWV from all GPS stations shown in Figure 14 and IWV from ICON and COSMO simulations, allowing for a more quantitative evaluation. Results are computed for the MODIS/*Aqua* overpass time on 24 April 2013, as well as for all matching time steps, with a temporal resolution of 15 min, for the days 24–25 April and 2 May 2013. 26 April is not included in this comparison, since for this day the data obtained from GFZ were incomplete and therefore were not yet included in

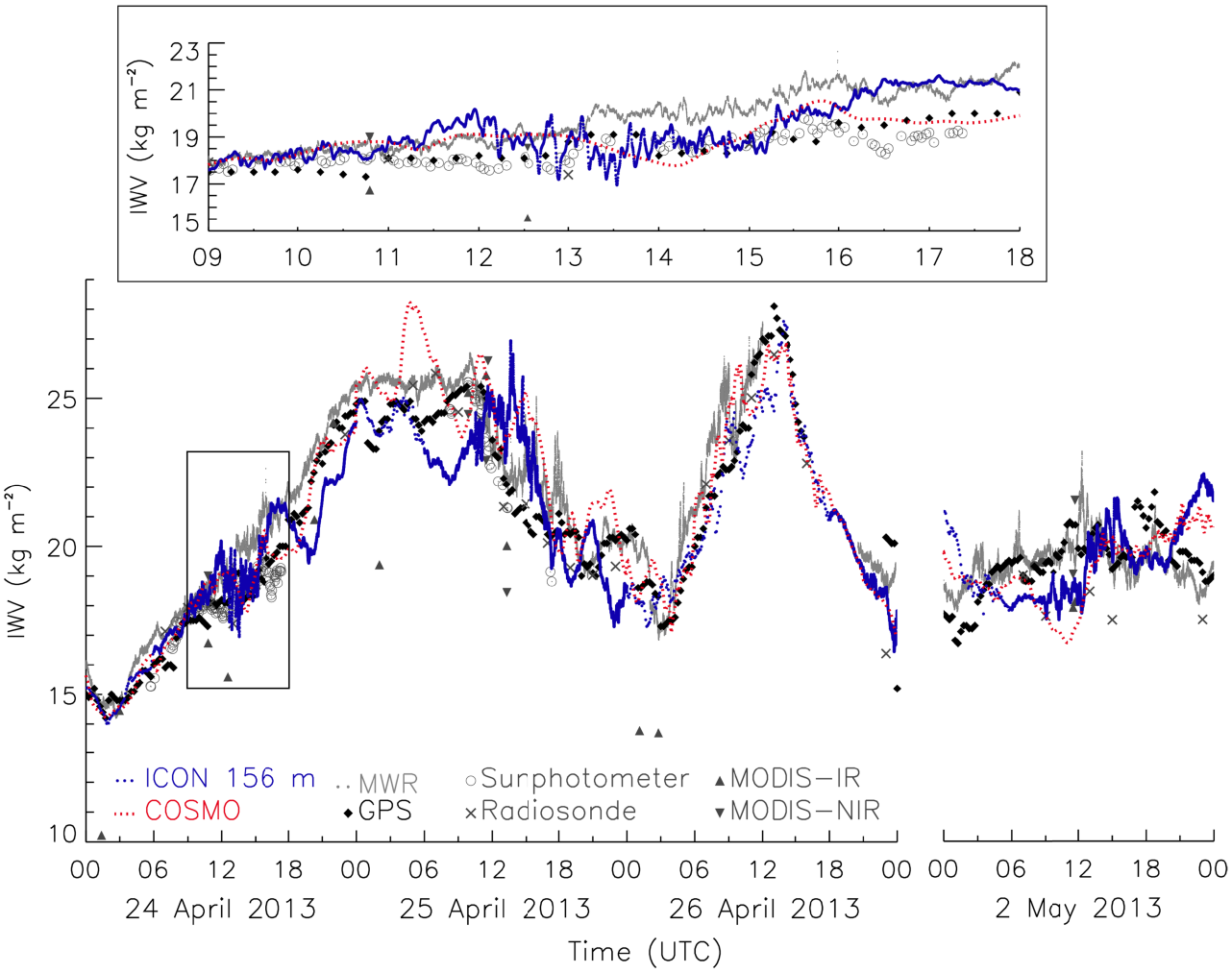

Figure 13. IWV time series from ICON, COSMO and several measurements for 24–26 April and 2 May 2013 at JOYCE supersite. The upper panel (a) shows the zoomed-in IWV daytime time series on 24 April, as indicated by the box in the lower panel (b).

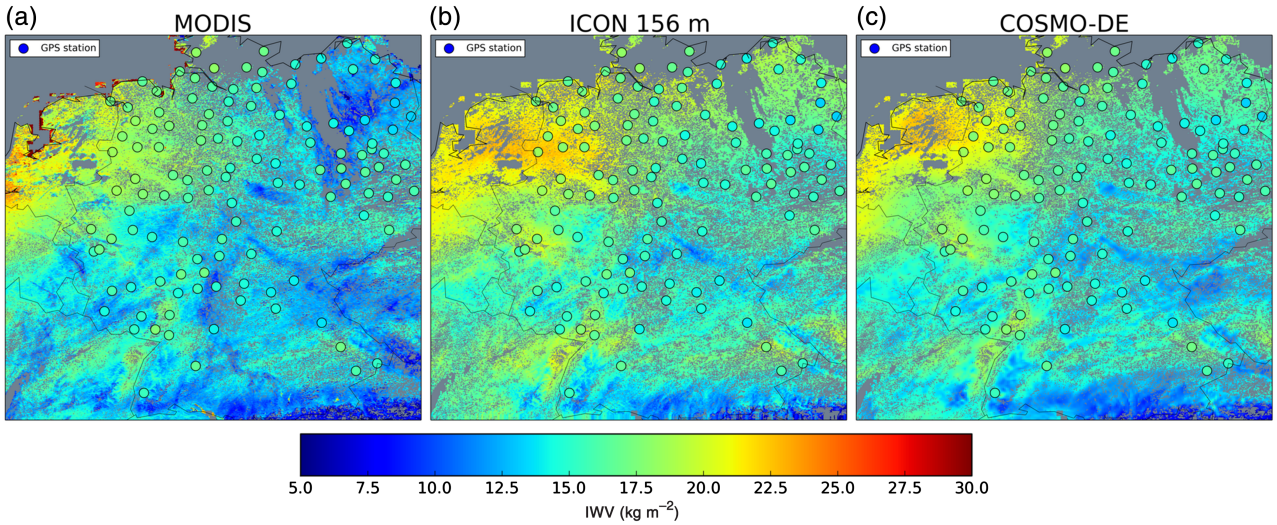

Figure 14. IWV distribution for (a) MODIS–FUB, (b) ICON 156 m and (c) COSMO, overlaid with IWV obtained from GPS measurements from the German GPS network for 24 April 2013 at the MODIS/*Aqua* overpass time of about 1230 UTC. For a better model–data comparison, areas for which no observations are available are marked by grey shading in all panels.

the HD(CP)^2 data portal. The IWV values of the models and MODIS–FUB were averaged over an area of about $10 \times 10\,\mathrm{km}^2$ around each GPS station. At the MODIS/*Aqua* overpass time, both ICON and COSMO show a positive bias, 1.73 and 0.61 kg m^{-2}, respectively. In this case, the bias and RMSE for ICON are larger than for COSMO. When looking at the bias and RMSE for all GPS time steps for 3 days, ICON shows a slight positive bias of 0.64 kg m^{-2}, while COSMO does not show any bias. The RMSEs for ICON and COSMO, 4.49 and 4.44 kg m^{-2}, respectively, are very similar.

6. Clouds

This section compares cloud properties simulated by ICON with numerous products derived from active and passive remote sensing instruments. The COSMO model is also included in

Table 5. Bias and bias-corrected root-mean-square error (RMSE) between GPS and ICON 156 m, COSMO and MODIS. Results are shown for the time period 24–25 April and 2 May 2013 and for the MODIS/*Aqua* overpass time of about 1330 UTC on 24 April, as shown in Figure 14.

	Bias ($kg\,m^{-2}$)	RMSE ($kg\,m^{-2}$)
ICON 156 m	0.64	4.49
COSMO	0.00	4.44
MODIS	−0.96	2.35
ICON 156 m (MODIS overpass time)	1.73	2.62
COSMO (MODIS overpass time)	0.61	1.85

these comparisons as a reference. Section 6.1 focuses on the representation of the horizontal distribution of cloud properties, while the vertical distribution is investigated in section 6.2. Finally, section 6.3 analyzes the subsequent capabilities of ICON to reproduce radiative properties that depend strongly on cloud parameters.

6.1. *Horizontal distribution of cloud parameters*

Figures 4(a)–(d) have illustrated that the large variability of meteorological situations found during the four simulated days should offer a deep qualitative overview for evaluating the representation of small- and mesoscale liquid, ice and mixed-phase clouds in ICON. Complementing these figures, synthetic MODIS images obtained from these four ICON simulations are presented in Figures 4(e)–(h). In these figures, radiances for the solar MODIS channels are generated from ICON model output using the radiative transfer code MFASIS (described in section 2.5). Figures 4(e)–(h) indicate that synoptic systems are simulated well by ICON. The representation of cumulus cloud fields looks particularly accurate in ICON 156 m simulations (e,g,h), while these clouds appear slightly larger for the 625 m simulation (f). However, their amount may be overestimated in comparison with MODIS images, notably on 24 April over central Germany. It can also be noticed that cirrus clouds, for instance in the northwest of Germany, are largely missed or are too thin in ICON. These aspects are analyzed further throughout this section.

Figure 15 compares spatial distributions of cloud water path (CWP, i.e. the sum of the vertically integrated ice and liquid cloud water contents) that were observed and simulated for the four scenes presented in Figure 4. The MODIS CWP retrievals Figures 15(a,a–d) are extracted from the collection 6 (C6) of the operational atmosphere products (Platnick *et al.*, 2014) and are provided with a horizontal resolution of 1 km. The SEVIRI/MSG CWP (b,a–d) is based on the Cloud Physical Properties (CPP) algorithm (Roebeling *et al.*, 2006) and is provided at SEVIRI's $4 \times 6\,km^2$ horizontal resolution (depending on viewing angle) over the simulation domain. Both products retrieve the CWP from a combination of visible and near-infrared radiometric measurements (Nakajima and King, 1990). ICON and COSMO simulations (c,a–d and d,a–d, respectively) are both shown at their highest horizontal resolution, i.e. 156 m and 2.8 km, respectively. It should be noted that the COSMO CWP includes both the resolved, grid-scale and the parametrized, subgrid-scale cloud water and ice. A cut-off threshold of $1\,g\,m^{-2}$ is used on the CWP simulated by both models. This choice is made in order to fit the lower sensitivity threshold of both instruments, as later shown in Figure 16. Very good overall agreements are found between the CWP spatial distributions simulated by ICON and the retrievals by MODIS and MSG. Note that, here, MODIS constitutes a better absolute reference, due to its finer horizontal resolution, while MSG as a geostationary instrument with 5 min temporal resolution is able to resolve the temporal development of clouds (not exploited in the current analysis). It can be expected that the high-resolution and explicit resolving of convective processes in ICON should, in particular, allow for an accurate representation of small-scale convective clouds observable by MODIS. This can particularly be noticed on 24–25 April and 2 May, when many cumulus cloud fields are found. Nevertheless, in agreement with observations from Figure 4, it can be noted that some of these cloud fields do not appear clearly in MODIS CWP retrievals. This is especially noticeable during the 24 April overpass, which could imply an overestimation of small-scale structures in ICON during that day. However, the MODIS CWP presented in Figure 15 contains many gaps, due to the strict quality filtering of the data for fractional cloudiness, which basically excludes all pixels below a CWP of $10\,g\,m^{-2}$ (see Figure 16). For this reason, a more thorough evaluation of small-scale clouds by ICON versus MODIS 250 m observations is presented later in this section (e.g. see Figure 17). By comparison, COSMO tends to overestimate the cloud coverage and subsequently underestimate the CWP in cumulus cloud fields and around cloud edges. This can be explained by its coarser resolution. For these highly variable clouds, the MSG data also underestimate MODIS values, due to the well-known plane-parallel albedo bias caused by the coarse resolution of the SEVIRI instrument (Marshak *et al.*, 2006; Wolters *et al.*, 2010). The CWP simulated by COSMO is underestimated by one order of magnitude compared with MODIS retrievals for the thick liquid and ice cloud layers that appear on 26 April and 2 May. This can be attributed to an averaging effect due to its lower grid resolution. ICON shows a slight underestimation of the cloud fraction during these days, but a better agreement is found with regard to the magnitude of CWP predictions.

In order to evaluate the statistical representation of the CWP in ICON further, probability density functions (PDFs) corresponding to these four scenes are presented in Figure 16. To account for the different resolution of each instrument and model, the CWP has first been averaged to the MSG horizontal resolution. This figure again indicates that COSMO (dotted red line) tends to underestimate the amount of clouds with a CWP greater than about $100\,g\,m^{-2}$ by comparison with MSG (black) and MODIS (grey) retrievals. ICON seems to predict the CWP of these clouds in comparison with the satellite data more accurately than COSMO. On the other hand, the amount of thin clouds with CWP less than about $2\,g\,m^{-2}$ is perfectly consistent between both models. Such values could be attributed to sub-visible cirrus clouds, which for instance appear on 24 April over northern Germany. The MSG and MODIS products contain large uncertainties for these thin clouds, which are therefore further investigated in section 6.2 with the use of lidar products. The observed difference in sensitivity between both retrieval methods for the range $1–10\,g\,m^{-2}$ can be explained by the very strict quality filtering of the MODIS cloud products and by the above-mentioned plane-parallel albedo bias, which affects MSG more strongly. The two distinct peaks in the MSG PDF around $10–20\,g\,m^{-2}$ should also be interpreted with caution, because of high retrieval uncertainties for these optically thin clouds, especially for broken cloud fields, where retrievals can be influenced by a priori information about the surface as well as jumps between the liquid and ice phase. It can nevertheless be concluded from Figure 16 that the distribution of CWP in ICON is in very good agreement with that of MODIS within their common sensitivity area, with, however, a small underestimation of clouds with very high CWPs. The agreement with MSG is also very good from about $30\,g\,m^{-2}$.

In order to evaluate the spatial distribution of small-scale cloud fields in ICON, which could not be done from the above CWP distributions, the MODIS reflectances at 250 m resolution used in Figure 4 are utilized directly to infer cloud size distributions down to the model effective resolution. Figure 17 presents the distributions inferred from MODIS real (black line) and synthetic observations from ICON 156 m (blue), 312 m (light blue) and 625 m (dark blue) on 2 May. This overpass is of particular interest, due to the high occurrence of small- and mesoscale liquid clouds,

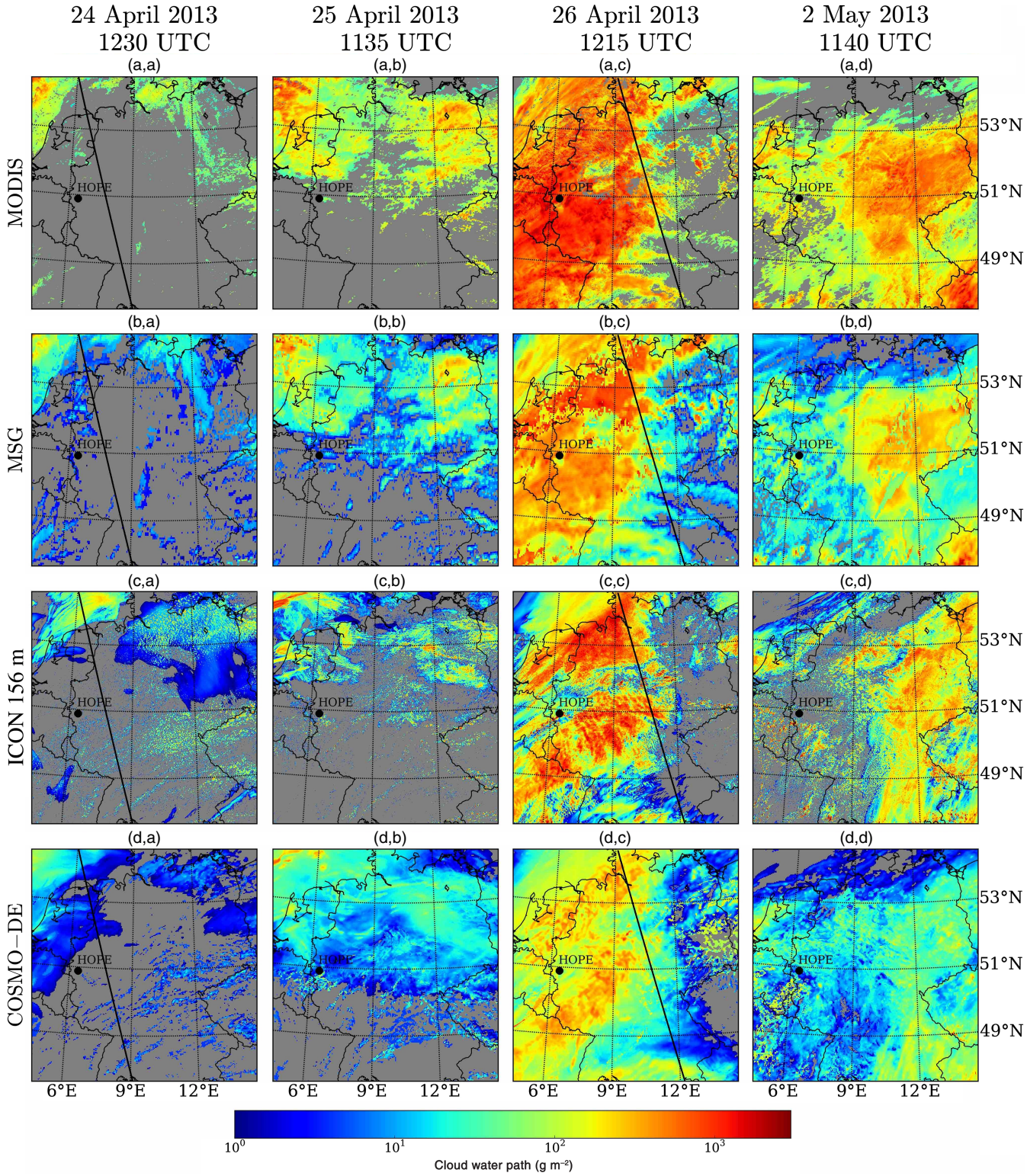

Figure 15. Spatial distribution of the CWP retrieved by (a,a–d) MODIS/*Aqua* and (b,a–d) SEVIRI/MSG and simulated by (c,a–d) ICON and (d,a–d) COSMO over the HD(CP)^2 domain at the MODIS overpass times on 24–26 April and 2 May 2013. The spatial resolution of each product is left unchanged. The solid black lines on 24 and 26 April show the overpass track of *CloudSat*. The HOPE label indicates the location where LACROS observations were taken during the HOPE campaign.

as indicated by Figure 4(d). Because of the push-broom approach of MODIS, its pixels can be associated with ground areas of strongly differing sizes. Hence, the synthetic and observed MODIS images are mapped on to a regular latitude–longitude grid with a resolution of $0.002^\circ \times 0.002^\circ \cos(\phi_C)$ (where $\phi_C = 51^\circ$ is a latitude near the centre of the simulation domain), in which all pixels correspond to areas on the ground that have similar sizes. The 0.6 µm MODIS images are mapped on to the regular grid and clouds are detected where the reflectance exceeds a threshold value of 0.25. Figure 17 shows that, for cloud sizes between 1.5 and 10 km, the distribution functions for MODIS and the ICON 156 m simulation are nearly straight lines in the double-logarithmic plot, i.e. they follow power laws. Therefore, assuming the distribution function is given as $N(r) = ar^b$, the corresponding slopes (i.e. the power law exponent b) for these sizes are −3.01 and −3.09 for ICON and MODIS, respectively. These slopes appear steeper than those reported in previous studies (e.g. between −1.7 and 2.0 from Neggers *et al.* (2003) and Dawe and Austin (2012), based on domain sizes of 6.4 km) but remain consistent with those obtained from larger LES simulations (e.g. around −2.7 from Heus and Seifert (2013) for a domain size of 50 km) and from observations for large cloud systems (Benner and Curry,

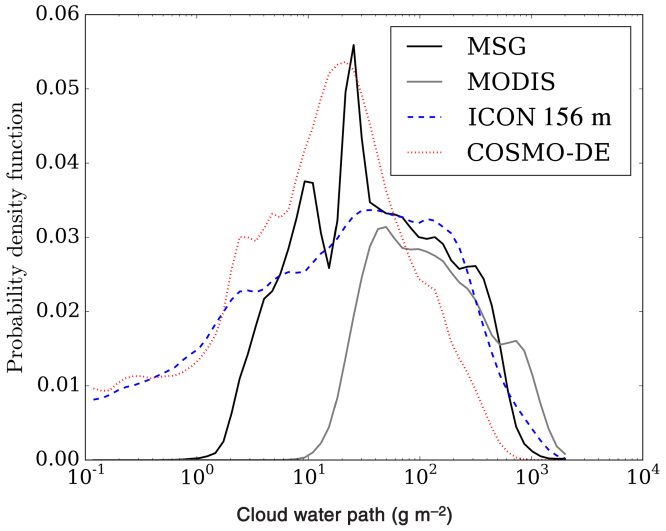

Figure 16. Probability density functions of the CWP retrieved by SEVIRI/MSG (black line) and MODIS/*Aqua* (grey) and simulated by ICON 156 m (dashed blue) and COSMO (dotted red) over the same domain and for the same days as in Figure 15. ICON, COSMO and MODIS are downscaled to the MSG resolution (4×6 km^2).

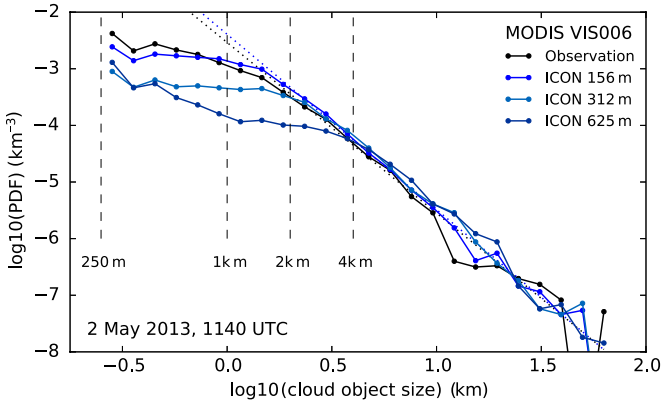

Figure 17. Normalized probability distribution function for cloud sizes observed over the HD(CP)^2 domain by MODIS (black line) and simulated by ICON at 156 m (blue), 300 m (light blue) and 600 m (dark blue) resolution on 2 May at 1140 UTC. The ICON curves have been multiplied by a factor n_{obs}/n_{156m}=1.80, where n_{obs} and n_{156m} are the total numbers of clouds detected in the MODIS and ICON 156 m images, respectively. Linear regressions of the MODIS and ICON 156 m size distributions between 1.5 and 10 km are indicated by dotted black and blue lines, respectively.

1998). The size distribution consistently deviates from the power law from about 10 km for ICON 156 m simulations and the observations. This value is higher than reported in the above-mentioned studies, but could be explained by the fact that the simulation domain allows clouds to grow to larger sizes. These previous studies also focused largely on shallow cumulus cloud fields, while our selected simulation represents a much more inhomogeneous situation. It can also be noted that the size distributions only start following a power law for cloud sizes from about 1.2, 2.3 and 5.0 km for ICON simulations obtained at 156, 312 and 625 m resolution, respectively. This behaviour is expected, assuming that the effective model resolution of ICON, from which cloud processes are actually resolved, is about seven to eight times the grid resolution (cf. Figure 12 and Z15). This indicates that comparisons with COSMO are not relevant here since, assuming an effective resolution that is a factor of 5 larger than the nominal resolution of 2.8 km (Bierdel *et al.*, 2012), it misses many of the small clouds that are reproduced well in ICON. Finally, the shallower slope of the observed size distribution for sizes below 1 km is consistent with Benner and Curry (1998), but could also be related to the fact that the effective resolution of MODIS 250 m observations is only 344–835 m, due to viewing geometry effects (Campagnolo and Montano, 2014).

Another parameter that is highly relevant for the planned application of ICON results for climate model development is the variability of the CWP over a large spatial scale. One can note from Figures 15 and 16 that the spatial and density distributions of CWP at model resolution display strong similarities between simulations and retrievals, but also differences that are interconnected with their respective horizontal resolution. Nevertheless, when averaged over a coarser grid, the subgrid CWP PDF carries additional information for distinguishing between cloud regimes that should also be properly captured, as it is deeply connected with underlying cloud-microphysical processes and cloud radiative properties (Kawai and Teixeira, 2012). Due to the strong nonlinearity of these relationships, the pixel mean CWP is indeed not sufficient for characterizing cloud regimes and the subgrid variability of this parameter is necessary. In climate models with a grid resolution much coarser than the ICON resolution, the subgrid cloud variability is parametrized. This variability is poorly constrained (e.g. Quaas, 2012; Rosch *et al.*, 2015) and one of the goals of HD(CP)^2 is to improve its representation in climate models. A prerequisite is therefore that ICON represents the CWP PDF adequately at scales that are unresolved by climate models. Figure 18 presents the spatial distribution of the CWP standard deviation obtained after averaging MODIS (a,a–d), ICON 156 m (b,a–d) and COSMO (c,a–d) CWP on a 25×25 km^2 grid. Unlike Figure 15, MODIS retrievals obtained in partial cloudiness have been included. A threshold of 5 g m^{-2} is used on CWP for the simulation output, consistent with the sensitivity limitations of MODIS (see Figure 16). It is observed that the signature carried in the 25×25 km^2 subgrid-scale variability of CWP is well represented by ICON compared with MODIS. The simulations display a rich spatial pattern that can be attributed to the different cloud regimes that appear during each scene studied. By comparison, COSMO simulations have a smaller variability and exhibit smaller spatial gradients, but remain consistent with MODIS retrievals, too.

6.2. *Vertical distribution of cloud parameters*

Besides the above analysis of the horizontal spatial distributions of cloud parameters, vertical profiles are investigated in this section. Figure 19 compares ground-based retrievals from the LACROS station operating during HOPE with ICON and COSMO simulations. The position of LACROS during HOPE is indicated by a thick black dot in Figure 15. This station is part of the Cloudnet project and therefore provides retrievals of cloud properties by following their standard procedure (Illingworth *et al.*, 2007). Unlike previous retrievals from MSG and MODIS, Cloudnet has the advantage of discriminating cloud water content (CWC) vertically between ice and liquid clouds within the same atmospheric column. Ice water content (IWC) profiles are obtained from cloud radar reflectivity and in-cloud temperature (here provided by COSMO-EU), using the parametrization by Hogan *et al.* (2006). Liquid water content (LWC) profiles are inferred based on the cloud-top and cloud-base altitudes provided by combined lidar/radar information, assuming adiabaticity. When possible, LWC profiles are then scaled in order to fit the LWP obtained from a microwave radiometer. Cloudnet provides CWC profiles every 30 s, while the temporal resolution is 9 s and 5 min for ICON and COSMO simulations, respectively. In order to account for these differences, each CWC profile is averaged to the vertical and temporal resolution of COSMO for the subsequent analysis. Further, a threshold of 10^{-4} g m^{-3} is set as a lower bound for CWC for this evaluation. Finally, it should be noted that data of the LACROS cloud radar and microwave radiometer are not available between 0900 and 1500 UTC on 25 April and from 0700–1900 UTC on

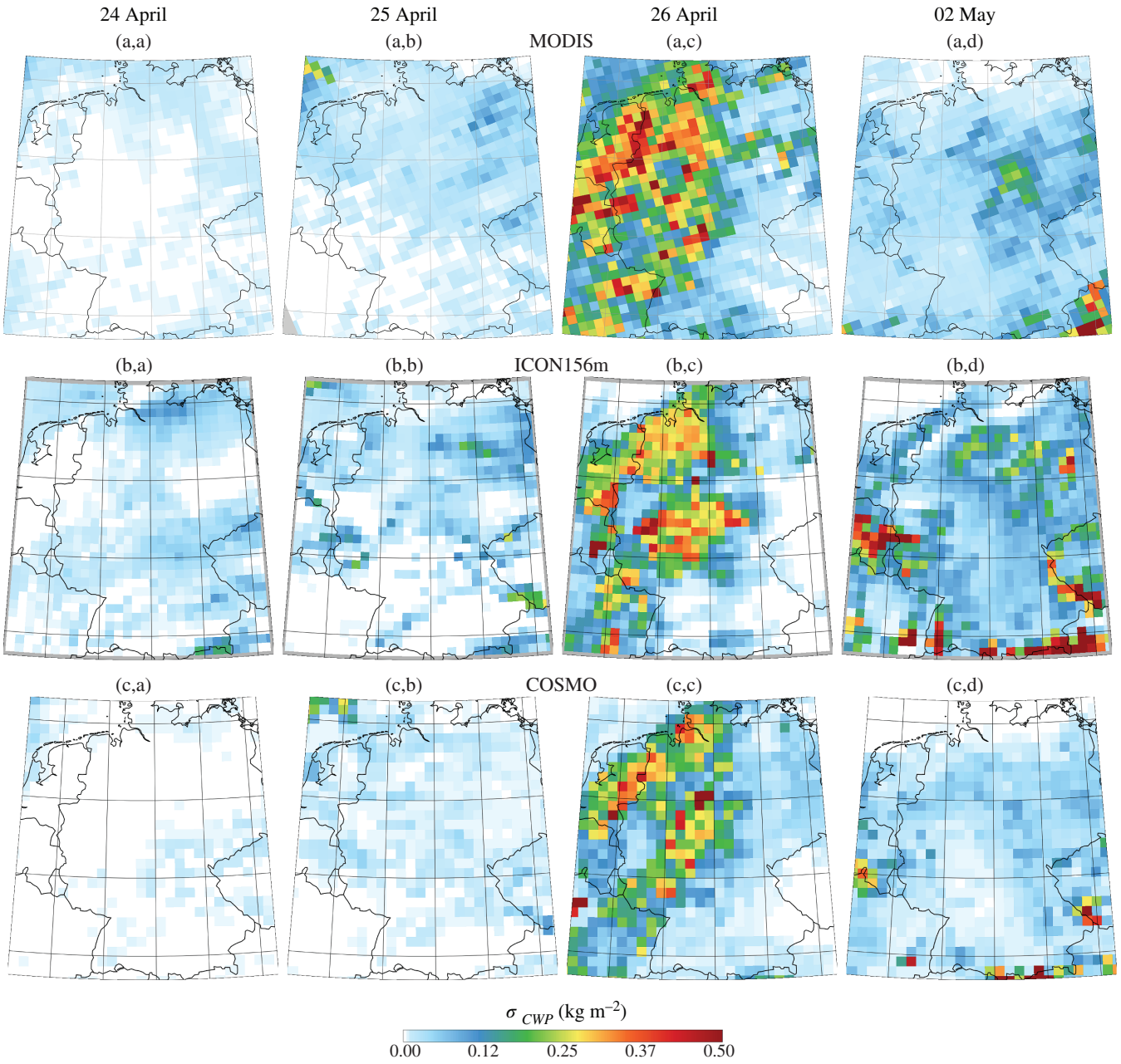

Figure 18. Spatial distribution of CWP standard deviation within 25 × 25 km^2 grid cells retrieved by (a,a–d) MODIS and simulated by (b,a–d) ICON and (c,a–d) COSMO.

2 May. In the absence of CWC retrievals, the position of cloud layers indicated by the Polly^{XT} lidar of LACROS are indicated in black. Cloud layers are defined from altitudes where the attenuated backscatter coefficient $\beta_{att} < 2.10$ until 300 m above this layer. Figure 19 shows good overall agreement between ICON simulations and Cloudnet retrievals. More specifically, the position and vertical extent of a low cloud deck observed by Cloudnet before 0600 UTC on 2 May are well reproduced by ICON, whereas this cloud field is virtually missing in the COSMO output. The LWC values are also in good agreement, despite differences in the vertical profiles and higher values towards the cloud top for Cloudnet, possibly due to its assumption of adiabaticity. ICON does not precisely predict the daily cycle of the development of planetary boundary-layer clouds that occur in the course of that day, but the double-layered cumuli cloud structure observed by Polly^{XT} during the afternoon is captured well. Similar disagreements between observations and both models are found for 25 April, for which the observed boundary-layer clouds are not present in the output of ICON and COSMO. However, the evolution of low liquid cloud layers occurring on 25 April before 0600 UTC is well simulated by both models. On 24 April, the boundary-layer clouds simulated for around 1200 UTC by ICON were not observed by Cloudnet. It is nevertheless possible that such small-scale clouds are missed by the Cloudnet station. It should also be kept in mind that discrepancies in the Cloudnet CWC can occur due to a lack of LWC retrievals in the presence of precipitation. This is, for example, clearly the case on 2 May. With regard to ice and mixed-phase clouds, the synoptic cloud layers observed on 24 April and 2 May are represented well in both models, with, however, a slight overestimation of the IWC by ICON during the latter day. The thick ice/mixed-phase cloud layer resulting from the cold front that passed the HOPE area during 26 April is also well simulated by both models, however with more homogeneous vertical distributions of the IWC in comparison with Cloudnet.

In order to evaluate further the representation of ice clouds in ICON, its simulations are compared with retrievals from the liDAR-raDAR (DARDAR) algorithm. DARDAR provides profiles of ice-cloud properties with 60 m vertical resolution, based on the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) extinction and the Cloud Profiling Radar (CPR) reflectivity (Delanoë and Hogan, 2010). Figure 20 compares these retrievals

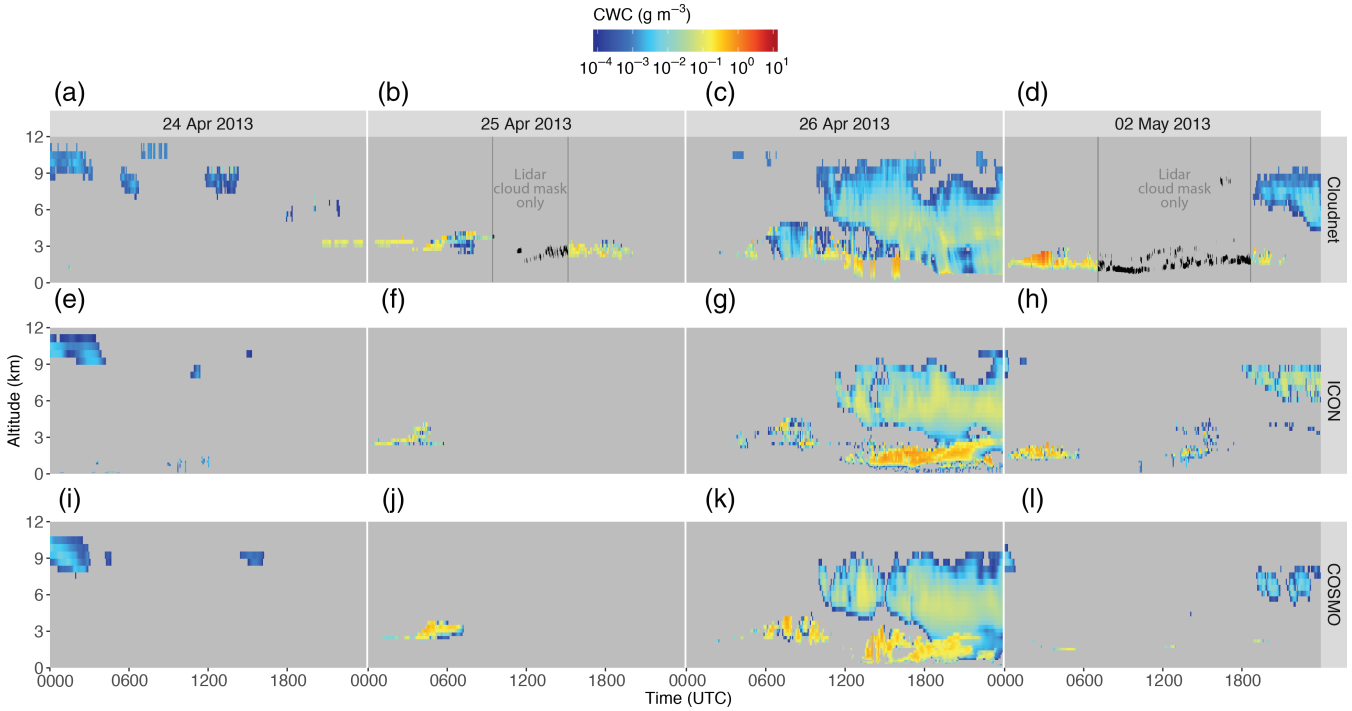

Figure 19. Vertical profiles of CWC (a–d) retrieved by Cloudnet at the LACROS station and simulated by (e–h) ICON and (i–l) COSMO from 24–26 April 2013 and on 2 May 2013. Gap areas in Cloudnet CWC retrievals are delimited by vertical grey lines and positions of cloud layers provided the Polly^{XT} lidar are indicated in black for these two time periods.

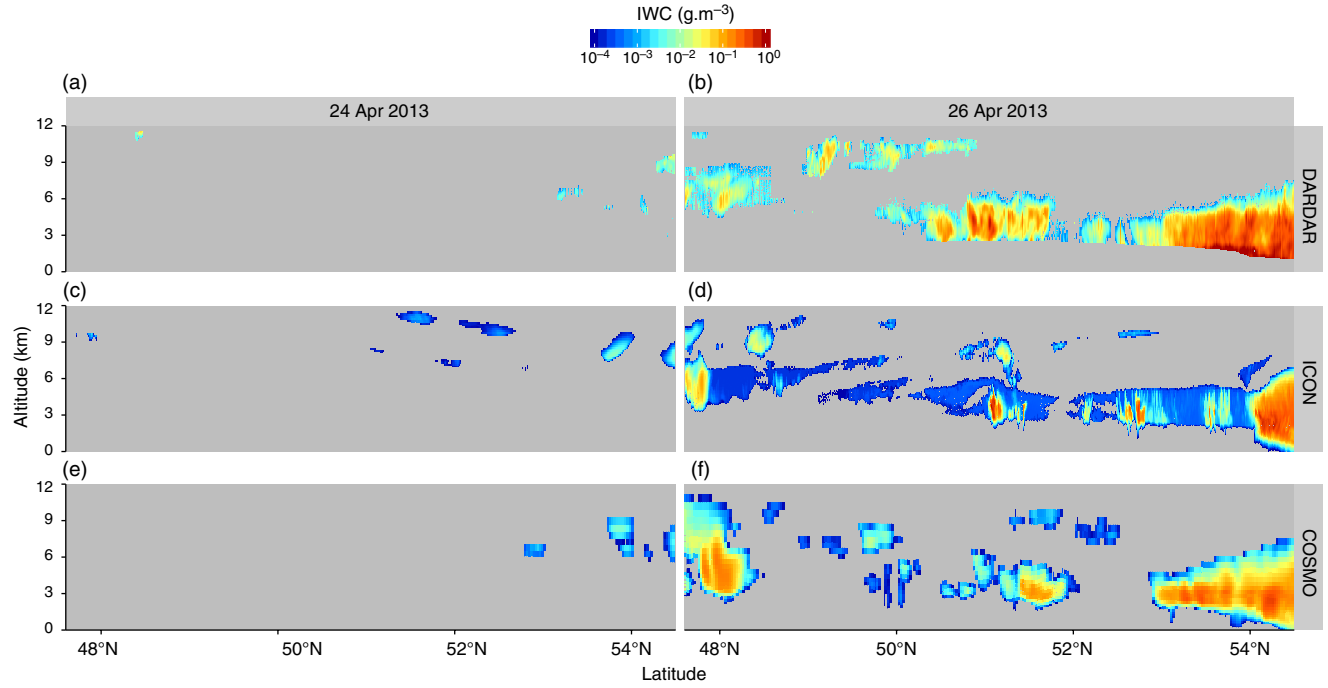

Figure 20. Vertical profiles of the IWC (a,b) retrieved by DARDAR and simulated by (c,d) ICON 156 m and (e,f) COSMO along the CloudSat overpass on 24 April 2013 (1234–1236 UTC) and 26 April 2013 (1221–1224 UTC).

with ICON and COSMO simulations. The IWC profiles are provided across the *A-Train* track at the CloudSat overpass between 1234 and 1236 UTC on 24 April (Figures 20(a),(c),(e)) and from 1222–1224 UTC on 26 April (Figures 20(b),(d),(f)). These overpasses are indicated by a black line in Figure 15. No significant *A-Train* overpass with nadir observations occurred during other simulation days. Note that the IWC profiles from ICON and COSMO include concentrations of ice as well as snow or graupel, as DARDAR does not separate these hydrometeors from ice crystals in its IWC retrievals. Not many ice clouds are found along the lidar–radar track during 24 April, but an overall coherence is still found between DARDAR and both models. In contrast, many more ice cloud layers occur on 26 April, including the presence of convective and mixed-phase clouds. The position of these cloud layers in ICON is consistent with DARDAR observations, but the IWC in the former is underestimated by one to two orders of magnitudes outside a few convective cores and away from the domain edges. By comparison, the temporal and vertical distributions of the IWC provided by COSMO are in much better agreement. An overestimation of the IWC for very thin ice clouds is known to occur in DARDAR retrievals (Deng *et al.*, 2012), but cannot explain these differences fully. Figure 20 therefore suggests that ICON can strongly underestimate the IWC of non-precipitating

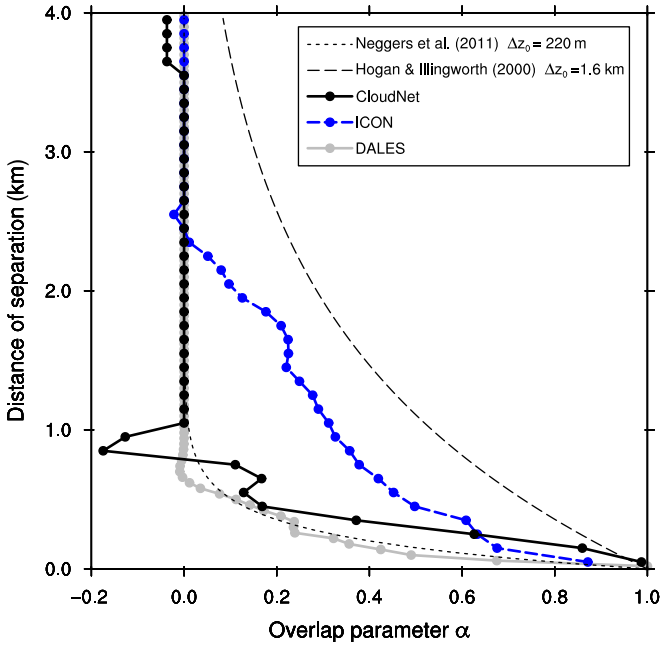

Figure 21. Vertical overlap in low-level boundary-layer clouds observed at JOYCE (black line) on 25 April 2013 between 1400 and 1600 UTC. Results from the corresponding simulations by ICON and DALES appear as blue and grey lines, respectively, and functional fits from Neggers *et al.* (2011) and Hogan and Illingworth (2000) are presented as thin dotted and dashed lines, respectively).

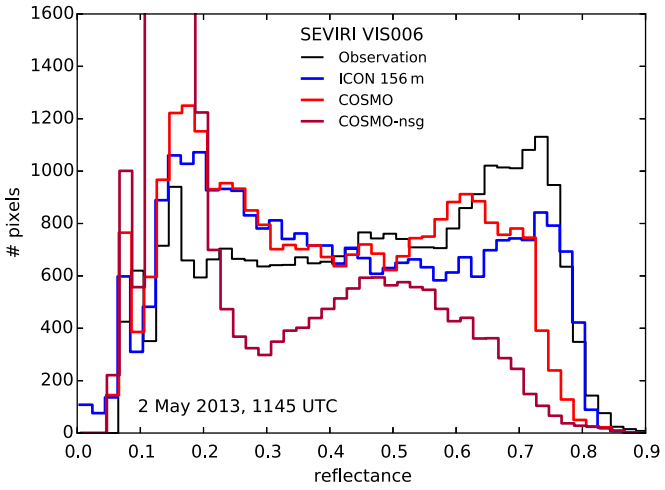

Figure 22. Histogram of the visible reflectance measured by SEVIRI/MSG (black line) and simulated by ICON 156 m (blue) and COSMO with (light red) and without (dark red) accounting for subgrid-scale variability.

ice clouds. It is worth noting that this underestimation is not clearly reflected in Figure 19 (where hydrometeor concentrations were, however, included), which illustrates the difficulty of comparing ice cloud retrievals due to discrepancies in their microphysical assumptions (e.g. shape and distribution of ice crystal habits, mass–dimensional relationships). Therefore, in complementarity, the representation of ice clouds in ICON is later analyzed through forward model simulations in section 6.3.

The representation of the vertical structure of clouds in ICON can also be evaluated with the help of the overlap parameter α (Hogan and Illingworth, 2000). The latter provides a direct measurement of the degree of vertical overlap, or 'overlap efficiency', of cloud layers separated by a given distance: a value of 1 indicates a maximum overlap, whereas a value of 0 implies a random overlap. Figure 21 shows the vertical overlap in low level boundary-layer clouds simulated by ICON at its 156 m horizontal resolution between 1400 and 1600 UTC on 25 April 2013, within a $10 \times 10\,\mathrm{km}^2$ subdomain around the JOYCE supersite. Comparisons with Cloudnet observations show that ICON can reproduce the random overlap typical of cumuliform boundary-layer cloud fields when the distance of separation is greater than about 2.5 km. However, it appears that the overlap is not yet random enough when the distance of separation decreases. For comparison, results from DALES simulations (cf. section 2.3) over the $12.8 \times 12.8\,\mathrm{km}^2$ subdomain are included. Parametrizations by Hogan and Illingworth (2000) and Neggers *et al.* (2011) are also added to Figure 21 for reference. These comparisons indicate that, while ICON shows signs of reproducing boundary-layer turbulence and associated clouds at its nominal 156 m resolution, a finer discretization is still required to resolve the overlap completely.

6.3. *Radiative representation of clouds*

Sections 6.1 and 6.2 have shown that ICON in general reproduces the vertical and horizontal structure of cloud properties with respect to a wide range of space-borne and ground-based retrievals well. One major improvement in comparison with the COSMO model lies in the better representation of small-scale convective cloud fields, but possible problems also appear, such as an observed underestimation of IWC in ICON simulations. This section summarizes these results in terms of histograms of observed and synthesized top-of-atmosphere radiation measurements in solar and terrestrial spectra.

Figure 22 compares the visible reflectances measured by SEVIRI (as shown in their geographical distribution in Figure 4) with corresponding synthetic reflectances obtained from ICON and COSMO with and without including subgrid clouds. The *A-Train* overpass scene on 2 May is chosen for this analysis, due to the high occurrence of low- and mid-level clouds, as shown in Figure 15(a–d,d). It can first be observed that the shape of the reflectance histograms from ICON agrees qualitatively with that from SEVIRI. Both histograms display a bimodal behaviour, with a surface peak near 0.15 and a cloud peak near 0.75. The range of reflectances is also very similar. However, ICON displays a higher surface peak and a lower cloud peak than indicated by the observations. This difference is caused by the fact that the large homogeneous cloud field on that day is located too far to the east, as already concluded from Figure 4(h). It can also be noted from the histogram of COSMO reflectances without subgrid clouds that the latter are essential to achieve a reasonable total cloud cover. Indeed, ignoring them results in too few pixels with reflectances greater than 0.6 and a clear overestimation of

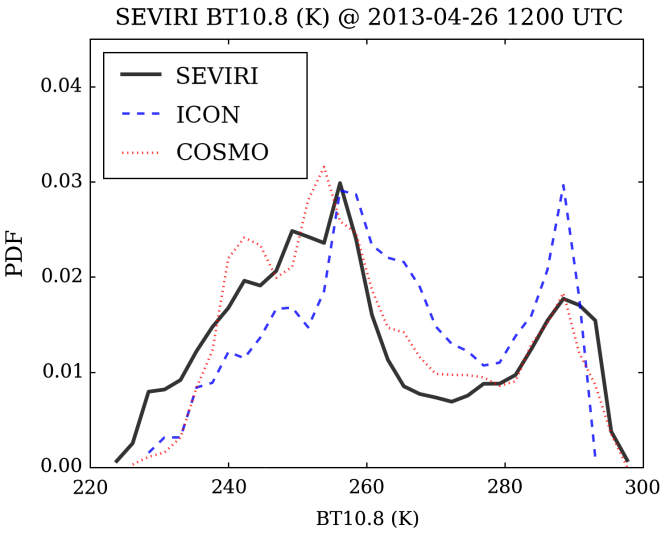

Figure 23. Normalized frequency of 10.8 µm brightness temperature occurrence rates for SEVIRI/MSG (black solid line), ICON 156 m (blue dashed) and COSMO (red dotted) synthetic imagery. Synthetic brightness temperatures have been derived from the ICON and COSMO simulations via the fast radiative transfer model RTTOV.

the surface peak in comparison with the observations. COSMO reflectances with subgrid clouds are in much better agreement with SEVIRI observations, but still with too few pixels with high reflectances. In contrast, the maximum reflectances related to the grid-scale clouds in ICON agree well with the observed distribution, which indicates a better representation of cloud cover due to its higher resolution.

Thermal infrared brightness temperatures (BT) are highly sensitive to absorption processes within cloud layers and therefore carry additional information on the altitude of the cloud layer in order to discriminate between liquid, mixed-phase and thick ice clouds. Figure 23 presents the probability density function (PDF) of BTs measured at 10.8 µm by SEVIRI on 26 April at 1200 UTC, compared with PDFs of synthetic measurements obtained from the corresponding ICON and COSMO simulations. This scene on 26 April is particularly suitable for this analysis, due to the high occurrence of ice and mixed-phase clouds, as previously observed in Figures 15(a–d,c), 19 and 20. BTs corresponding to ICON and COSMO simulations have been calculated using the radiative transfer code RTTOV (see section 2.6). The BT PDFs feature a bimodal structure, with a warm peak (around 290 K) representative of clear-sky radiances and a cold peak (around 255 K) connected to a passing frontal cloud system on that day. Firstly, a slight shift of the former peak towards cold temperatures can be noticed in ICON by comparison with the observations. These result are in agreement with the underestimation of the near-surface temperature discussed in section 4.1. The cold peak is representative of the high occurrence of liquid water in the mid-troposphere during that scene (e.g. through liquid/mixed-phase clouds or precipitation). Its position in ICON is in good agreement with that of SEVIRI, while a slight shift appears towards colder temperatures for COSMO. The strong underestimation of BTs below 250 K in ICON by comparison with SEVIRI indicates a lack of thick and high ice clouds for that scene, as previously observed in Figure 20 along the *CloudSat* overpass. In contrast, the overestimation of the occurrence of BTs above 260 K could suggest an overestimation of either mid-level or thin ice clouds. A similar behaviour is observed for COSMO, however with a better overall agreement with SEVIRI. It can be noted that biases related to the ice scheme have previously been investigated for COSMO (e.g. Böhme *et al.*, 2011; Eikenberg *et al.*, 2015), but remain difficult to attribute due to strong intercorrelations between emissions from the surface, ice clouds, liquid clouds, precipitations and, to a lesser extent, water vapour at 10.8 µm. Therefore, despite the fact that Figure 23 strongly suggests an issue with the representation of ice and mixed-phase clouds in ICON, further dedicated analyses remain necessary to validate this conclusion.

7. Precipitation

In this section, we examine the performance of ICON on precipitation prediction and the potential improvements through comparisons with COSMO simulations and radar observations for a case study on 26 April 2013. Demanding a relatively high temporal resolution of the model output (30 min or higher), the comparisons presented are based on ICON 312 m, aggregated to 1.2 km horizontal resolution.

The vertically pointing *Ka*-band cloud radar (MIRA) at the LACROS site (50.88°N, 6.41°E, Figure 15) provides a detailed insight into the temporal evolution of precipitating systems with a high spatial resolution of 30 m and a temporal resolution of 10 s. The Doppler polarimetric *X*-band radar in Bonn (BoXPol, 50.73°N, 7.07°E) provides 3D-volume scan data, which contain ten Plan Position Indicators (PPI) with elevations ranging from 0.5–28°. The resolution of each PPI is 1° × 100 m and the maximum detection range is 100 km. The radar data offer additional insight into hydrometeor microphysics, due to their dual polarization capabilities. Compared with precipitation radars with longer wavelengths (e.g. BoXPol), MIRA provides a higher sensitivity to non-precipitating hydrometeors such as cloud droplets and ice particles, but is affected by strong attenuation in precipitation cores. The German precipitation radar network, which is composed of 17 C-band radars, covers the entire area of Germany. Each radar site provides 3D-volume scan data, which contains 10 PPI with elevations ranging from 0.5° to 25°. The resolution of each PPI is 1° × 1 km and the maximum detection range is 180 km. The radar network thus provides surface precipitation with a 5 min temporal resolution and 1 km spatial resolution and also gives insights into the 3D structures of precipitation systems.

Since a direct comparison between the prognostic model variables and radar observations is rather difficult, we use a radar forward operator, which calculates pseudo-observations based on the modelled atmospheric states. The radar forward operator EMVORADO (see section 2.6) is used as a basis to compare 3D volume scan measurements from the DWD C-band radar network with model simulations. The zenith-pointing MIRA observations at the LACROS site were simulated with the radar forward operator implemented in PAMTRA (Maahn *et al.*, 2015, see also section 2.6).

This section is structured as follows. First, the representation of the surface precipitation fields in the ICON and COSMO models is investigated, including their rain-rate distribution and the organization in precipitation objects. Second, the vertical representation of the cloud and precipitation systems, including their temporal evolution, is evaluated at the LACROS and BoXPol sites, respectively. An investigation of the representation of

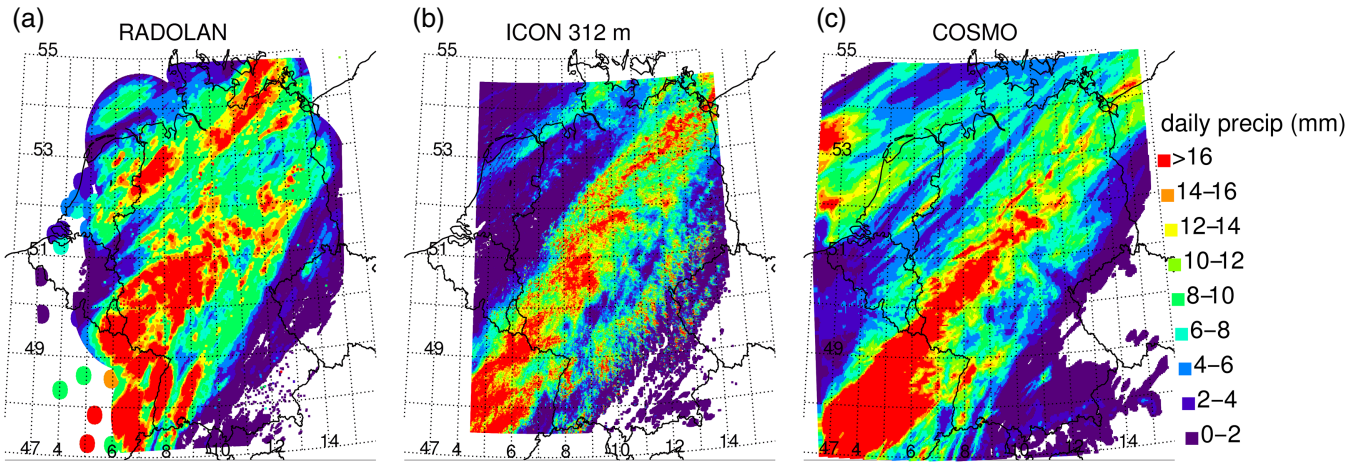

Figure 24. Surface precipitation accumulation on 26 April 2013 from (a) the radar network observation, (b) ICON 312 m and (c) COSMO, respectively. Note that in (a), precipitation accumulation indicated by isolated solid circles represents in situ rain-gauge observations.

the drop-size distributions in the model is presented through comparisons with the precipitation radar BoXPol. Third, the 3D spatial distribution of reflectivities is compared with the modelled pseudo-reflectivities.

7.1. Geographical precipitation distribution

Based on the German radar network, Figure 24 compares the accumulated surface precipitation on 26 April 2013 as provided by the RADOLAN composite product SF (24 h rain accumulation, adjusted to rain-gauges; see also Table 3) with the ICON 312 m and COSMO simulations. More than 90% of the grid boxes over Germany have a daily precipitation accumulation less than 25 mm on 26 April 2013. Two separate rain bands are identified from the radar observations (Figure 24(a)), one extending from southwestern to eastern Germany and the other one from western to northeastern Germany. The first is well represented in both models, whereas the latter is underestimated in both models (Figures 24(b) and (d)). Note that the solid circles outside the radar observation range are the precipitation measured by in situ rain gauges.

The Intensity Scale Skill score (ISS) technique (Casati *et al.*, 2004; Casati and Wilson, 2007; Casati, 2010) is used to evaluate the location of daily accumulated precipitation in the models on this day. For this analysis, all data were coarse-grained on to a regular grid with 637 × 589 grid points and a resolution of 1.2 km. For each threshold and each spatial scale, which can be understood as the size of an averaging window smoothing the data, the mean square error (MSE) between the model and observation is calculated and compared with the expected MSE of a random forecast, where the position of each rain pixel is chosen independently following a field-wide uniform distribution neglecting any spatial correlations. Positive skill (ISS > 0) indicates that the model hits the locations of the observed precipitation better than the random forecast. For the results presented in Figure 25 and the ensuing figures in this section, the observed precipitation product RADOLAN refers to the composite product RY (quality-controlled rain rates at 5 min temporal resolution). Figure 25 shows that both models exhibit good ISS results for all thresholds, with the exception of intense precipitation above 32 mm day^{-1} (indicated by the dark red colours in the upper left of panels (a) and (b)), where the first positive skill can be found at 19.2 km spatial scale. This length is denoted as the 'skilful scale' and characterizes the model accuracy regarding the location of precipitation. The amplitudes of the daily accumulated precipitation are captured by the frequency

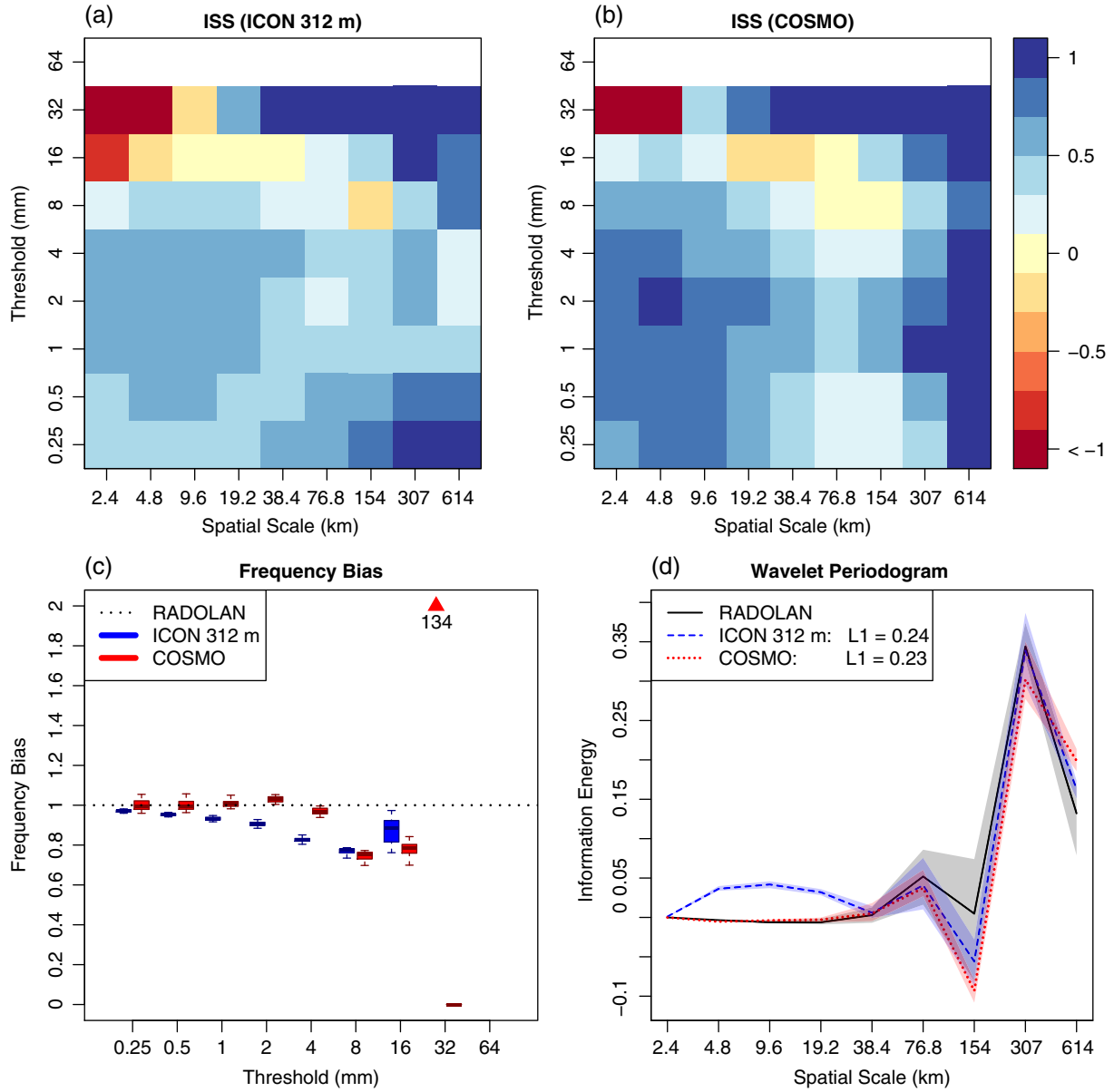

Figure 25. The Intensity-Scale Skill Score (ISS) evaluates the location information of daily accumulated precipitation data for (a) ICON 312 m and (b) COSMO, compared with RADOLAN observations at 2400 UTC on 26 April 2013. (c) The frequency bias captures intensity errors, while (d) the wavelet periodogram shows the distribution of information energy over different spatial scales.

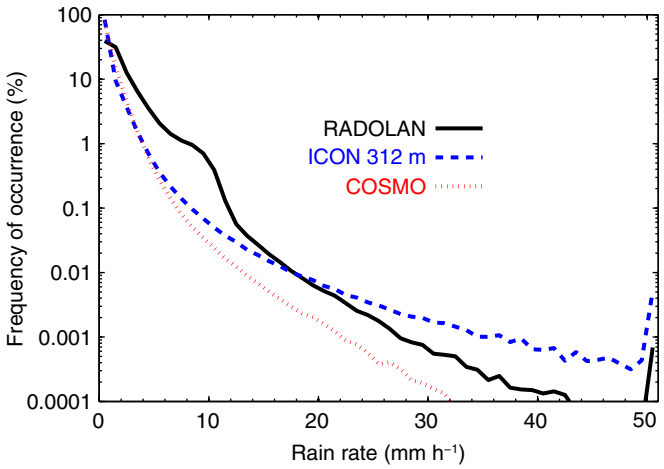

Figure 26. Surface rain-rate frequency of occurrences from the RADOLAN radar network, ICON 312 m and COSMO on 26 April 2013 over the full simulation domain.

bias (FB) at 2400 UTC, which measures the ratio of threshold exceedance in the models compared with observations. The FB shows that ICON 312 m underestimates low and intermediate precipitation sums with a threshold of up to 16 mm day^{-1}, but severely overestimates intense rain events above 32 mm day^{-1} (Figure 25(c)), while COSMO underestimates the occurrence of intermediate and strong precipitation sums above 8 mm day^{-1}.

The texture of the accumulated precipitation fields is evaluated by the (normalized) wavelet periodograms (WPG). This approach combines the advantages of spectral analysis (similar to Fourier power spectra) with the wavelets' inherent ability to decompose information into distinct spatial scales. The WPG is an estimator for the distribution of information energy, which can be understood as a measure for spatial variability, across separated spatial scales. We refer to Eckley *et al.* (2010) and Weniger *et al.* (2015) and the references therein for the mathematical details, which are outside the scope of this study. Since the periodogram is normalized for each data set (RADOLAN, ICON 312 m and COSMO), it is independent of FB and ISS. Therefore, the texture analysis complements the results concerning intensity (FB) and location (ISS). The daily precipitation texture from COSMO agrees very well with the observations. ICON 312 m has significantly more information energy on small scales compared with the observations, i.e. the features in ICON 312 m are too small and/or too scattered (Figure 25(d)). However, this could be a consequence of the much higher native spatial resolution of ICON 312 m and thus higher-resolved remote sensing products may be required to evaluate the small-scale structure of the accumulated precipitation properly.

To complement the information from FB at one single time step shown in Figure 25, Figure 26 compares directly the distribution of surface rain rates from RADOLAN, ICON 312 m and COSMO for the entire day. An apparent power law in the distribution of observed and modelled rain rates is visible. The shape of the distribution as simulated by COSMO shows a better agreement with the observations, while ICON 312 m generates a flatter distribution. Note that ICON 312 m and COSMO generate nearly indistinguishable surface rain-rate distributions for a day with dominant cumulus clouds (2 May 2013; not shown). During the day analyzed here (26 April 2013), however, ICON 312 m underestimates the frequency of occurrence of low rain rates but overestimates the occurrence of higher rain rates. COSMO, however, generally underestimates the frequency of occurrences compared with the RADOLAN observations over Germany (Figure 26). Both ICON 312 m and RADOLAN reveal occurrences of rain rate >50 mm h^{-1}, which are accumulated at the rain rate of 51 mm h^{-1} in Figure 26. Also, additional analysis of FB at each time step of ICON 312 m indicates that both models underestimate low rain rates (not shown), which is consistent with the conclusions from Figure 26. Note that both precipitation fields provided by RADOLAN and COSMO have been inter-/extrapolated to the grid with 1.2 km spatial resolution also used for the ICON 312 m data. The local maximum in the frequency of occurrence for rain rates around 10 mm h^{-1} as indicated by RADOLAN is also observed on other days and may be an artefact produced by the use of different $z - R$ relationships (between the linear radar reflectivity z and the rain rate R) for different rain intensities.

In order to compare the statistics of precipitation objects in models and observations, a contour-segmentation tracking (Bremer *et al.*, 2010; Kuhn *et al.*, 2015) is applied to all three data sets (RADOLAN, ICON 312 m and COSMO). The extraction corresponds to a (maximum) watershed algorithm with a persistence-based hierarchical merging step (Edelsbrunner *et al.*, 2002; Kuhn *et al.*, 2015). For object definition, a minimum threshold of $h = 0.5$ mm h^{-1} and a persistence-based merging threshold $p = 0.5$ mm h^{-1} are used. For tracking, we compute a bidirectional overlap function between two subsequent time steps to construct sets of object track graphs (similar to Bremer *et al.*, 2010). The resulting tracking statistics shown in Figure 27 include (a) the empirical life-time and (b) object-size distributions. Additionally, the location of the 40% largest object centres identified in RADOLAN, ICON 312 m and COSMO over the entire day (Figure 27(c)) and the object outlines for all objects identified at 1700 UTC (Figure 27(d)) are shown. For 26 April 2013, both models capture the main characteristics of rain

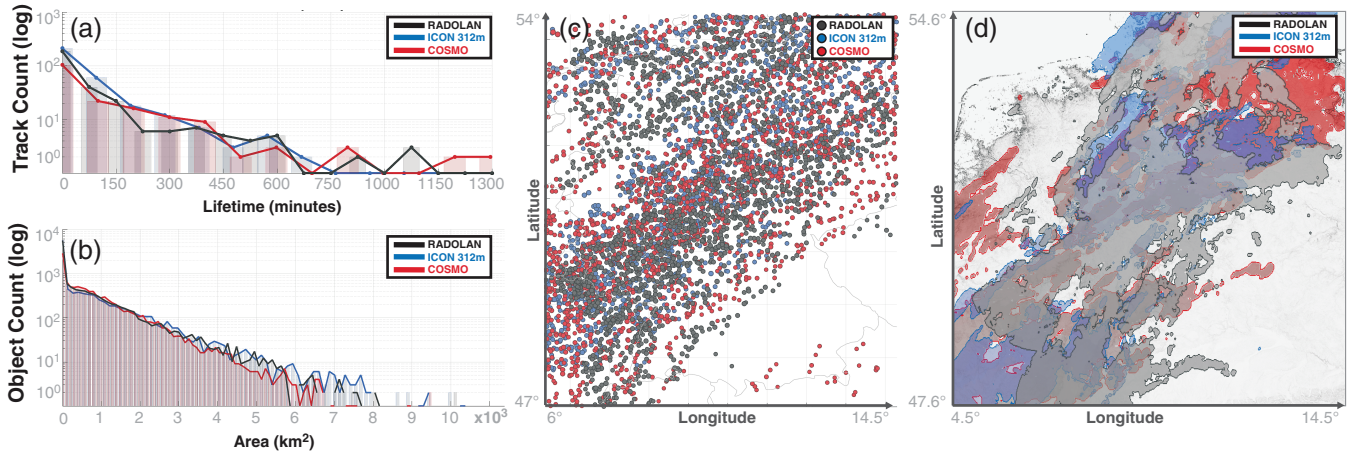

Figure 27. (a) The empirical life-time and (b) object-size distributions for tracked precipitation objects on 26 April 2013 in RADOLAN, ICON 312 m and COSMO. Additionally, (c) the location of the 40% largest object centres identified in all three data sets over the entire day and (d) the object outlines for all objects identified at 1700 UTC for each modality are shown. Precipitation objects are calculated by the contour-segmentation tracking algorithm using a minimum threshold of $h_{\min} = 0.5$ mm h^{-1} and persistence value $p = 0.5$ mm h^{-1}.

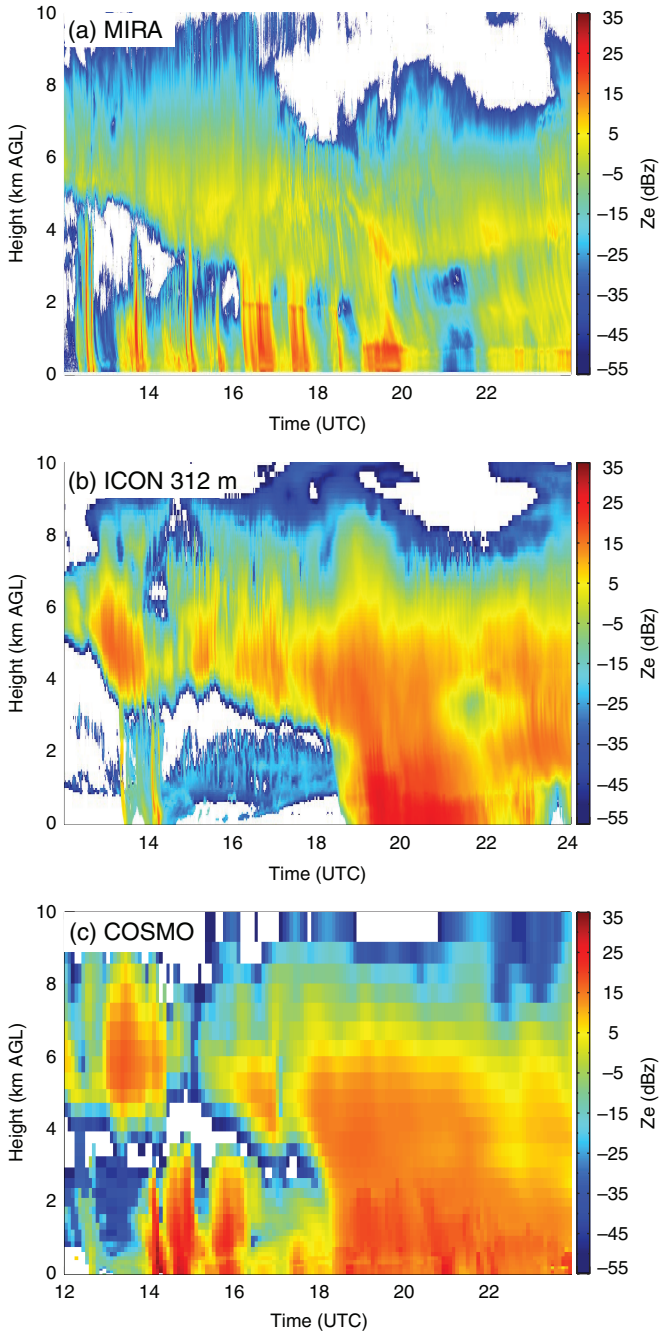

Figure 28. (a) Radar reflectivity factor as a function of time and height measured by the zenith-pointing 35.5 GHz cloud radar at the LACROS site on 26 April 2013 compared with the simulated radar reflectivity factor based on (b) the ICON 312 m output fields and (c) COSMO. The model output fields have been forward-simulated with the radar forward operator implemented in PAMTRA (Maahn *et al.*, 2015).

objects at the 312 m scale well compared with RADOLAN. More specifically, ICON 312 m produces more smaller ($<100\,km^2$) and larger ($>4000\,km^2$) short-lived (<150 min) objects, while COSMO creates more medium-sized ($200–1500\,km^2$) and few very large ($>12\,000\,km^2$) long-lived objects. The overall moving directions and passage of the precipitation systems are reproduced by both models, which is also corroborated by investigation of the locations of the precipitation bands shown in Figures 24 and 25.

7.2. *Vertical–temporal hydrometeor distribution*

The temporal evolution of the vertical reflectivity profile is monitored by MIRA at the LACROS site (Figure 28). The MIRA-observed ice clouds above the melting layer descend gradually from 1200 UTC to 1600 UTC (Figure 28(a)), which is well captured in both models (Figures 28(b) and (c)). Above 3 km, where the frozen hydrometeors reside, ICON 312 m and COSMO overestimate the amount of frozen hydrometeors, since simulated reflectivities are significantly higher for ICON and COSMO (up to 10 dBZ) compared with observations. At this single site, COSMO tends to locate even more ice at higher altitudes (see for example the ice cloud feature at 1300–1400 UTC between 4 and 7 km). These findings are consistent with earlier studies (Böhme *et al.*, 2011; Bollmeyer *et al.*, 2015). More detailed analysis with PAMTRA revealed that the large reflectivity overestimations in the ice part by ICON 312 m and COSMO are caused mainly by an overestimation of snow particles (not shown).

In the following, polarimetric measurements from BoXPol are exploited to examine the representation of drop numbers and water contents in ICON 312 m. Figure 29 compares the pseudo-polarimetric radar observations using the ICON 312 m simulations with the so-called Quasi-Vertical Profiles based on BoXPol measurements at 18° elevation angle between 1200 UTC and 1730 UTC. The pseudo-polarimetric radar moments have been calculated using EMVORADO extended to polarimetry at the University of Bonn (configuration: Mie scattering taking into account attenuation and partially melted snow, graupel and hail, treated as described in Zeng *et al.*, 2016). Additional information about the forward operator is provided in section 2.6. In order to construct Quasi-Vertical Profiles, data from a given elevation angle scan ($>10°$) are azimuthally averaged and the range coordinate is converted to height. The methodology of Quasi-Vertical Profiles was demonstrated in Trömel *et al.* (2014) and expanded further in Ryzhkov *et al.* (2016). Quasi-Vertical Profiles are especially beneficial for monitoring the temporal evolution of precipitating systems at a larger scale. The Quasi-Vertical Profiles of the polarimetric radar variables are computed by azimuthal averaging of the data measured during standard conical scans, i.e. PPIs at higher antenna elevation angles (18° in this case) and display the results in a time-versus-height format. In the Quasi-Vertical Profiles, the melting layer between 1200 and 1730 UTC can be easily identified by the enhanced radar reflectivity (Z) and differential reflectivity (Z_{DR}) between 2 and 3 km (Figure 29(a) and (b)). The environmental 0 °C level predicted by ICON follows the descent of the BoXPol observed melting layer (Figures 29(a) and (b)). The pseudo-polarimetric radar observations from the ICON simulations show a clear melting signature, with Z and Z_{DR} enhancement between 2 and 3 km (Figures 29(c) and (d)), which is consistent with the BoXPol observations. However, an overestimation of Z_{DR} is found in the ICON 312 m simulations, with values up to 3 dBZ, while in the Quasi-Vertical Profiles Z_{DR} is lower than 1 dBZ below the melting layer. Z_{DR} is a measure of the mean particle size and does not depend on the concentration of the drops. While magnitudes of Z_{DR} around 3 dBZ suggest raindrops of around 5 mm in diameter, values below 1 dBZ corresponds to a diameter around 2 mm. Thus, the high Z_{DR} below the melting layer is caused by the relatively low drop number and relatively high rain-water content in the ICON 312 m simulations, which introduces a relatively high number concentration of large raindrops and finally results in an overestimation of the mean drop diameter. The tendency towards an overestimation in differential reflectivity Z_{DR} and radar reflectivity Z below the melting layer has already been recognized in COSMO simulations and needs further investigation.

7.3. *Three-dimensional precipitation distribution*

To verify the 3D distribution of precipitation fields, the object-based verification method 3D-SALH, following Wernli *et al.* (2008) and Zhang *et al.* (2015), is employed. 3D-SALH is applied to simulated and observed radar reflectivity data from hourly volume scans of 14 DWD radar stations (three radar stations are not available for the day investigated) for the entire day in native polar radar coordinates. The simulated volume scans from

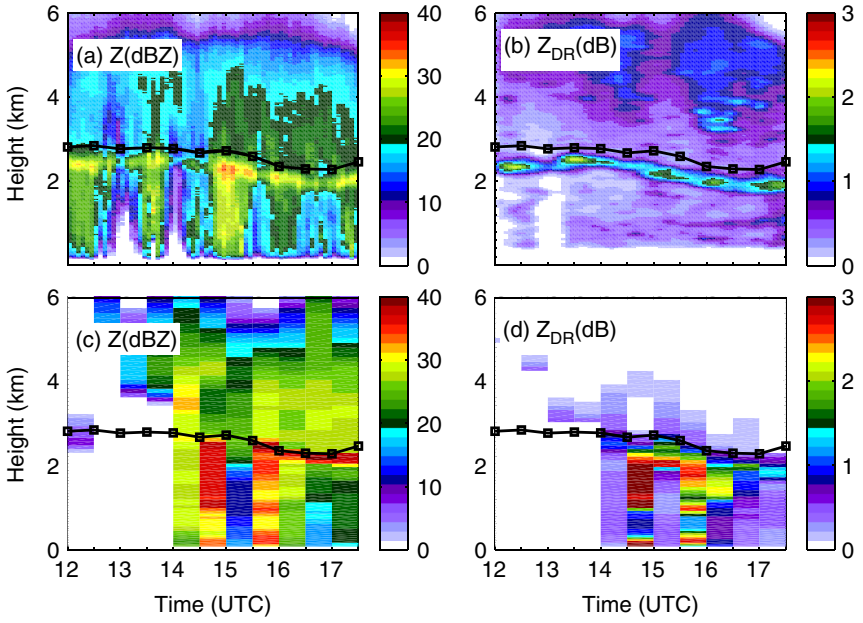

Figure 29. Quasi-vertical profiles of (a) reflectivity Z and (b) differential reflectivity Z_{DR} observed by BoXPol at 18° elevation between 1200 UTC and 1730 UTC on 26 April 2013 are compared with (c) forward simulated reflectivity Z and (d) differential reflectivity Z_{DR} using ICON 312 m results at the BoXPol site as input. The black lines with square markers indicate the height of the melting layer predicted by the ICON 312 m simulations.

ICON and COSMO have been obtained by non-polarimetric EMVORADO.

Figure 30 shows the 3D-SALH diagrams using the entire area at a radar site with reflectivities exceeding the 28 dBZ threshold for the definition of a precipitation object. Thus, objects are not necessarily continuous and could consist of many smaller precipitation cells. The verification data set includes 270/252/245 samples (ICON 312 m/ ICON 156 m/ COSMO) with reflectivities larger than 28 dBZ in both simulations and observations. Each dot in Figure 30 represents one set of structure/amplitude/location/height ($S/A/L/H$) values, which describe the averaged characteristics of the simulated precipitation objects compared with the observation at one radar site at 1 h. Thus, Figure 30 describes the overall information for the precipitation objects in Germany (14 radar sites) in the simulations compared with the radar data on the given day (24 h). In the ICON 312 m (Figure 30(a)) and ICON 156 m (Figure 30(b)) simulations, most dots are found in the first quadrant (top right) and third quadrant (bottom left) of the S–A plane, where the S and A components are either both positive or both negative. Compared with the ICON simulations, most dots of the COSMO simulations are found in the first quadrant of the S–A plane. Positive values of A indicate an overestimation of object amplitudes and vice versa, while positive S indicates that objects are too flat (or too large horizontally) compared with observations.

Considering first the A component, ICON 312 m and ICON 156 m simulate both precipitation objects with overestimated and underestimated amplitudes (positive and negative values of A in Figures 30(a) and (b)), which is in agreement with the findings presented in Figure 26. Note that the threshold of 28 dBZ corresponds to rain rates of around 2 mm h^{-1}. COSMO (Figure 30(c)) shows an overall strong model overestimation of the precipitation amount (about +83%). This could be caused by an overestimation of the frozen hydrometeors of COSMO as shown in Figure 28. With respect to the S component, the positive median values of the S component from the ICON simulations indicate that ICON simulates too large/flat precipitation objects compared with the radar data, of which ICON 156 m (median of $S = 0.53$) has smaller structure errors than ICON 312 m (median of $S = 0.77$). The large positive median value of the S component in COSMO (median of $S = 1.69$) indicate that the COSMO model simulates even larger/flatter precipitation objects than ICON, which is in agreement with the findings presented in Figure 27. Both models tend to simulate precipitation similarly well regarding the locations of the precipitation objects, as shown by the bluish colours of the L component. In terms of H errors, both models simulate higher centres than observed. ICON 156 m has the smallest median height errors (median of $H = +0.24$ km, vs +0.45 km and +0.38 km in ICON 312 m and COSMO, respectively) with the smallest interquartile range (IQR of $H = 0.37$ km, 0.55 and 0.55 km in ICON 312 m and COSMO, respectively).

It was shown in this section that ICON underestimates low rain rates. When compared with COSMO, ICON still shows an improvement, particularly for rain rates between 8 and 20 mm h^{-1}, where it is much closer to the radar observations. For higher precipitation rates, ICON overestimates rain rates. ICON also predicts precipitation rates higher than 50 mm h^{-1} in agreement with the observations, which are not predicted by COSMO-DE. The wavelet analysis also confirms the underestimation (overestimation) of low (high) daily accumulated precipitation in ICON (Figure 25). In general, ICON captures the evolution of clouds and the moving direction of precipitation systems well. With the polarimetric radar observations, investigation of the representation of drop number concentration indicates that ICON tends to produce unrealistically large mean drop diameters below the melting layer. The object-based statistics of the surface rain fields (Figure 27) as well as the differences of the 3D characteristics between the simulated and observed precipitation objects (Figure 30) show that both ICON and COSMO tend to simulate too large precipitation objects and ICON also simulates more smaller and short-lived cells compared with observations. Both ICON and COSMO perform similarly well in simulating the locations of the centres of 3D precipitation objects, as shown by the small values of IQR and median of the L and H components. Both models simulate higher centres of precipitation objects than observations, whereas ICON 156 m has smaller height errors than ICON 312 m and COSMO.

8. Conclusions

The aim of this study was to evaluate comprehensively simulations using ICON in large-eddy configuration at a horizontal resolution of 156 m over Germany. The model was developed on the basis of the ICON climate and weather prediction model in the context of the High Definition Clouds and Precipitation for advancing Climate Prediction (HD(CP)^2) project, where

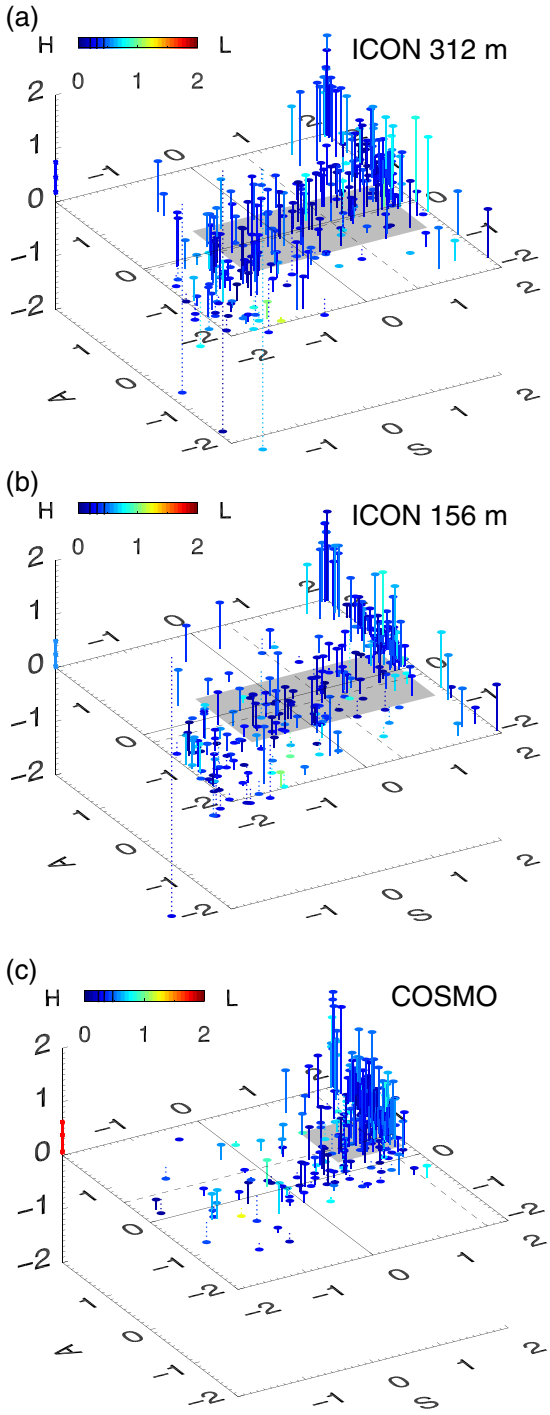

Figure 30. 3D-SALH diagrams for 24 h forecasts from (a) ICON 312 m, (b) ICON 156 m and (c) COSMO, using radar observations from 14 stations in Germany. Precipitation objects are defined using a reflectivity threshold of 28 dBZ. Every dot shows the values of the four components *S*, *A*, *L* and *H*, which describe the averaged characteristics of the simulated precipitation objects compared with the objects observed at one hour. The *L* component is indicated by the colour of the dots (the 25th, 50th and 75th percentiles are marked by black lines). Median values of the *S* and *A* components are shown as dashed lines and the grey box extends from the 25th to the 75th percentile of *S* and *A*, respectively. The *H* component is indicated by the line that connects the dots and the *S*–*A* plane ($H = 0$) with the 25th, 50th and 75th percentiles marked by the blue/light blue/red lines on the *Z*-axis. Good forecasts (small values of *S*, *A*, *L* and *H*) are thus in the centre of the diagrams.

a comprehensive observational dataset was also compiled and created. This evaluation study thus made use of a broad range of observations, as well as reference models and theoretical concepts, e.g. for turbulence scaling. A caveat on the present analysis is that it relies only on four simulated days and a detailed investigation of e.g. deep convection is left for future research.

Since the HD(CP)^2 project aims at a better understanding of the role of clouds and precipitation in the climate context, our main focus was on moist processes. In order to put the evaluation of ICON in LES configuration into a sensible context, a consistent simulation at a horizontal resolution of 2.8 km with the well-established numerical weather prediction model COSMO was conducted and evaluated along with ICON.

The aim of the present evaluation is twofold.

(1) Shortcomings in the current configuration and parametrizations of ICON in LES configuration should be identified. As the HD(CP)^2 project moves into its second phase, this identification will help to improve the model and document improvements in the foreseeable future.
(2) We intended to assess to what extent the model is fit for purpose, in the sense that it fulfils the criteria necessary to make use of the simulation results to improve our understanding of cloud and precipitation processes and to improve their representation via parametrizations in global climate models. The present analysis focuses on the representation of variability at small- to mesoscales that are subgrid-scale for global models and investigates turbulence and moist processes in particular.

An important result inferred from the analysis of the kinetic energy spectra is that the effective resolution of the 156 m resolution model is about a factor of 7–8 coarser than the nominal resolution, which is not unusual for atmospheric models. In consequence, small shallow cumulus clouds are too big in the model.

The weather conditions, as defined by water-vapour distributions, clouds seen from satellite or surface precipitation fluxes, are broadly simulated similarly well by ICON and COSMO. This behaviour is expected, as both models are forced with nearly identical initial and boundary conditions.

The results further show that ICON in LES configuration is already, at this early stage of development, almost as good in most aspects of weather prediction as the established COSMO model. It still shows some biases in the thermodynamic profiles, especially in the boundary layer, which are mostly comparable in magnitude with the COSMO model. In particular, the surface temperatures tend to be too low over extended areas. However, the turbulent sensible heat flux is substantially larger than observed, in combination with a too large net radiation flux. These results point to the need to work on the parametrization of the surface energy budget in the model. Possibly as a consequence, ICON is not superior to COSMO in terms of boundary-layer height and its variability. Also, the wind gusts that are resolved by ICON do not show an improvement compared with the gusts as empirically parametrized in COSMO.

ICON in its highest resolution matches the turbulence profiles as observed by lidar in the mixed layer quite well. The humidity variance profile is well resolved and close to the DIAL observations, except for a displacement of the maxima at the PBL top. However, the temperature variance in the entrainment layer is significantly underestimated in comparison with the lidar retrievals. This suggests that more work would be useful to represent turbulent transport better, particularly in the entrainment layer of the PBL.

The temporal variability in water vapour as observed by high temporal resolution microwave remote sensing, which a model such as COSMO with a longer time step cannot simulate, is captured well by ICON.

Despite the fact that cumulus tends to be too large in comparison with satellite data, the high-resolution model is much better at simulating these small clouds than the coarser-resolved versions. ICON generates clouds with a size distribution that is very similar to the observed one for cloud areas between 1 km^2 and 100 km^2. For smaller clouds, the power-law exponent in the model deviates somewhat from the observed one, as is to be expected on scales smaller than the effective model resolution, but it is still substantially better than at the coarser (312 m) resolution. In contrast, COSMO has to rely on simplistic subgrid cloud

parametrizations for this range of scales, which is quite important from a radiative perspective. ICON simulates the variance of cloud liquid water path within grid boxes of $25 \times 25\,km^2$ (representative for a next-generation global climate model) quite well. The high-resolution model shows a substantial improvement in this variability compared with the COSMO model.

Cloud and precipitation microphysics are also parametrized in ICON, although one might expect that some processes benefit from the better-resolved cloud variability. ICON in general is substantially better at simulating the occurrence of thick clouds than COSMO. This is, however, mostly true for liquid clouds, as ice clouds are often too thin compared with the data. In comparison with polarimetric radar data, it was found that precipitation particles in ICON are too large. These two results point to a potential to revise aspects of ice-phase microphysical processes.

In terms of precipitation, ICON underestimates the frequency of occurrence of low-intensity rain and simulates heavy-rain events too frequently. These strong events are only captured skilfully if occurring at a larger scale. The precipitation objects are too large in comparison with radar data, but ICON is substantially better at simulating the structure of precipitation than COSMO. Also, the lifetime for showers is better simulated by ICON than COSMO.

These results consistently show that the high-resolution model significantly improves the representation of small- to mesoscale variability. This generates confidence in the ability to simulate moist processes with fidelity. When using the model output to assess turbulent and moist processes and to evaluate and develop climate model parametrizations, it seems relevant to make use of the highest resolution, since the coarser-resolved model variants fail to reproduce aspects of the variability. The high-resolution ICON model, in terms of variability, is superior to the coarser-resolved COSMO model in virtually all aspects. This now allows us to make use of the large statistics for actual, realistic weather conditions with a comprehensive and consistent parameter dataset from the model simulation for analysis towards a better understanding of climate processes.

Acknowledgements

This work was funded by the Federal Ministry of Education and Research in Germany (Bundesministerium für Bildung und Forschung; BMBF) through the research programme *High Definition Clouds and Precipitation for Advancing Climate Prediction – HD(CP)*^{2}. The authors gratefully acknowledge the computing time granted by the German Climate Computing Centre (Deutsches Klimarechenzentrum, DKRZ) and by the Forschungszentrum Jülich at the supercomputer JUQUEEN (JARA-HPC).

Providers of observational data are acknowledged; for details about the datasets used, refer to Table 3.

We thank Armin Mathes, Paul Dostal, Jochen Elberskirch, Martin Rieland and their team at DLR Projektträger for administrating the project very constructively; Florian Rauser for coordinating the research; and Joachim Biercamp, Ulrike Burkhardt, Sarah Jones and Pier Siebesma for valuable advice on the project. We further thank Leonidas Lindardakis for his support in generating the ICON grid and D. Reinert for giving valuable comments on the draft version. We thank two anonymous reviewers for their supportive and constructive comments.

Appendix

Visualization

Data visualization and analysis are imperative tasks, requiring an adequate workflow matching the visualization goal. Visualization of HD(CP)^2 data, either for an explorative data analysis or for communicating the final results to a broader scientific community, requires a 3D interactive environment, yet the sheer size of HD(CP)^2 data poses several challenges. Currently we use ParaView (Ayachit, 2015) and Vapor (Clyne *et al.*, 2007) for both tasks and have written extensions that allow us to read, process and visualize large HD(CP)^2 data sets. Figure A1 shows a snapshot from an animation created with ParaView. The entire animation is uploaded as File S2. More in depth data analysis techniques, such as scatter plots, linked views and brushing, are also available within ParaView and are employed in our research. In order to handle future data output for even higher resolutions, techniques for in situ data visualization and compression are currently being examined and implemented (Jubair *et al.*, 2015).

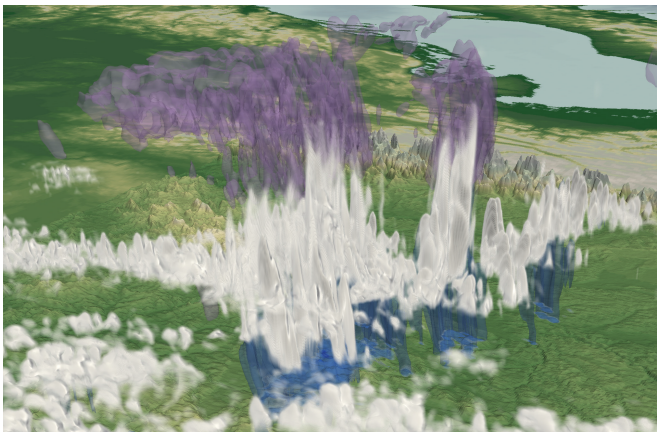

Figure A1. Example of data visualization of one time step of 26 April 2013 using ParaView (Ayachit, 2015). Liquid clouds are shown in white, ice clouds in purple and precipitation in blue colours. The snapshot shows a view from the north towards the Alps, exhibiting a large cloud cluster over south/east Germany.

Supporting information

The following supporting information is available as part of the online article.

File S1. Metric terms in the turbulence closure.

File S2. Visualization of HD(CP)^2 data.

References

Alexandrov MD, Schmid B, Turner DD, Cairns B, Oinas V, Lacis AA, Gutman SI, Westwater ER, Smirnov A, Eilers J. 2009. Columnar water vapor retrievals from multifilter rotating shadowband radiometer data. *J. Geophys. Res.* **114**: D02306, doi: 10.1029/2008JD010543.

ASTER GDEM Validation Team. 2011. 'ASTER global digital elevation model version 2 – summary of validation results', Technical report. NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center: Sioux Falls, South Dakota. http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary_GDEM2_validation_report_final.pdf (accessed 6 December 2016).

Ayachit U. 2015. *The ParaView Guide: A Parallel Visualization Application*. Kitware: New York, NY.

Baars H, Ansmann A, Engelmann R, Althausen D. 2008. Continuous monitoring of the boundary-layer top with lidar. *Atmos. Chem. Phys.* **8**: 7281–7296, doi: 10.5194/acp-8-7281-2008.

Baldauf M, Brdar S. 2016. 3D diffusion in terrain-following coordinates: Testing and stability of horizontally explicit, vertically implicit discretizations. *Q. J. R. Meteorol. Soc.* **142**: 2087–2101, doi: 10.1002/qj.2805.

Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Weather Rev.* **139**: 3887–3905, doi: 10.1175/MWR-D-10-05013.1.

Barthlott C, Hoose C. 2015. Spatial and temporal variability of clouds and precipitation over Germany: Multiscale simulations across the 'gray zone'. *Atmos. Chem. Phys.* **15**: 12361–12384, doi: 10.5194/acp-15-12361-2015.

Behrendt A, Wulfmeyer V, Hammann E, Muppa SK, Pal S. 2015. Profiles of second to fourth-order moments of turbulent temperature fluctuations in the convective boundary layer: First measurements with rotational Raman lidar. *Atmos. Chem. Phys.* **15**: 5485–5500, doi: 10.5194/acp-15-5485-2015.

Benner T, Curry J. 1998. Characteristics of small tropical cumulus clouds and their impact on the environment. *J. Geophys. Res.* **103**: 28753–28768, doi: 10.1029/98JD02579.

Beyrich F, Adam WK. 2007. 'Site and data report for the Lindenberg reference site in CEOP – phase I', Technical report. Deutscher Wetterdienst: Offenbach am Main, Germany. http://icdc.zmaw.de/fileadmin/user_upload/icdc_Dokumente/DWD230_CEOP_report.pdf (accessed 8 December 2016).

Bicheron P, Defourny P, Brockmann C, Schouten L, Vancutsem C, Huc M, Bontemps S, Leroy M, Achard F, Herold M, Ranera F, Arino O. 2008. 'GLOBCOVER products description and validation report', Technical report. European Space Agency: Toulouse, France. http://due.esrin.esa.int/files/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf (accessed 8 December 2016).

Bierdel L, Friederichs P, Bentzien S. 2012. Spatial kinetic energy spectra in the convection-permitting limited-area NWP model COSMO-DE. *Meteorol. Z.* **21**: 245–258, doi: 10.1127/0941-2948/2012/0319.

Blahak U. 2008. 'Towards a better representation of high density ice particles in a state-of-the-art two-moment bulk microphysical scheme'. In *Proceedings of the 15th International Conference Clouds and Precipitation*, Cancun, Mexico, p. 9.

Blahak U. 2016. 'Calculation of radar reflectivity from model output', Technical report. Consortium for Small Scale Modeling (COSMO): Offenbach, Germany. http://www.cosmo-model.org/content/model/documentation/core/cosmoPhysParamtr.pdf (accessed 8 December 2016).

Böhme T, Stapelberg S, Akkermans T, Crewell S, Fischer J, Reinhardt T, Seifert A, Selbach C, van Lipzig N. 2011. Long-term evaluation of COSMO forecasting using combined observational data of the GOP period. *Meteorol. Z.* **20**: 119–132, doi: 10.1127/0941-2948/2011/0225.

Bollmeyer C, Keller J, Ohlwein C, Bentzien S, Crewell S, Friedrichs P, Hense A, Keune J, Kneifel S, Pscheidt I, Redl S, Steinke S. 2015. Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. R. Meteorol. Soc.* **141**: 1–15, doi: 10.1002/qj.2486.

Bony S, Colman R, Kattsov VM, Allan RP, Bretherton CS, Dufresne JL, Hall A, Hallegatte S, Holland MM, Ingram W, Randall DA, Soden BJ, Tselioudis G, Webb MJ. 2006. How well do we understand and evaluate climate change feedback processes? *Clim. Dyn.* **19**: 3445–3482, doi: 10.1175/JCLI3819.1.

Boucher O, Randall D, Artaxo P, Bretherton C, Feingold G, Forster P, Kerminen VM, Kondo Y, Liao H, Lohmann U, Rasch P, Satheesh S, Sherwood S, Stevens B, Zhang X. 2013. Clouds and aerosols. In *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, Qin D, Plattner GK, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P. (eds.): 571–658. Cambridge University Press: Cambridge, UK and New York, NY, doi: 10.1017/CBO9781107415324.

Bremer PT, Weber GH, Pascucci V, Day M, Bell JB. 2010. Analyzing and tracking burning structures in lean premixed hydrogen flames. *IEEE Trans. Vis. Comput. Graph.* **16**: 248–260, doi: 10.1109/TVCG.2009.69.

Bühl J, Seifert P, Wandinger U, Baars H, Kanitz T, Schmidt J, Myagkov A, Engelmann R, Skupin A, Heese B, Klepel A, Althausen D, Ansmann A. 2013. LACROS: The Leipzig aerosol and cloud remote observations system. *Proc. SPIE* **8890**: 889002,doi: 10.1117/12.2030911.

Caine S, Lane TP, May PT, Jakob C, Siems ST, Manton MJ, Pinto J. 2013. Statistical assessment of tropical convection-permitting model simulations using a cell-tracking algorithm. *Mon. Weather Rev.* **141**: 557–581, doi: 10.1175/MWR-D-11-00274.1.

Campagnolo ML, Montano EL. 2014. Estimation of effective resolution for daily MODIS gridded surface reflectance products. *IEEE Trans. Geosci. Remote Sens.* **52**: 5622–5632, doi: 10.1109/TGRS.2013.2291496.

Casati B. 2010. New developments of the intensity-scale technique within the spatial verification methods intercomparison project. *Weather and Forecasting* **25**: 113–143.

Casati B, Wilson LJ. 2007. A new spatial-scale decomposition of the Brier score: Application to the verification of lightning probability forecasts. *Mon. Weather Rev.* **135**: 3052–3069.

Casati B, Ross G, Stephenson DB. 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorol. Appl.* **11**: 141–154.

Clark P, Roberts N, Lean H, Ballard SP, Charlton-Perez C. 2016. Convection-permitting models: A step-change in rainfall forecasting. *Meteorol. Appl.* **23**: 165–181, doi: 10.1002/met.1538.

Clyne J, Mininni P, Norton A, Rast M. 2007. Interactive desktop analysis of high resolution simulations: Application to turbulent plume dynamics and current sheet formation. *New J. Phys.* **9**: 301.

Colella P, Woodward PR. 1984. The Piecewise Parabolic Method (PPM) for gas-dynamical simulations. *J. Comput. Phys.* **54**: 174–201, doi: 10.1016/0021-9991(84)90143-8.

Crewell S, Mech M, Reinhardt T, Selbach C, Betz H, Brocard E, Dick G, O'Connor E, Fischer J, Hanisch T, Hauf T, Hnerbein A, Delobbe L, Mathes A, Peters G, Wernli H, Wiegner M, Wulfmeyer V. 2008. The general observation period 2007 within the priority program on quantitative precipitation forecasting: Concept and first results. *Meteorol. Z.* **17**: 849–866, doi: 10.1127/0941-2948/2008/0336.

Dawe JT, Austin PH. 2012. Statistical analysis of an LES shallow cumulus cloud ensemble using a cloud tracking algorithm. *Atmos. Chem. Phys.* **12**: 1101–1119.

Deardorff JW. 1970a. A numerical study of three-dimensional turbulent channel flow at large Reynolds number. *J. Fluid Mech.* **41**: 453–480.

Deardorff JW. 1970b. Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection. *J. Atmos. Sci.* **27**: 1211–1213, doi: 10.1175/1520-0469(1970)027¡1211:CVATSF¿2.0.CO;2.

Delanoë J, Hogan RJ. 2010. Combined CloudSat–CALIPSO–MODIS retrievals of the properties of ice clouds. *J. Geophys. Res.* **115**: D00H29, doi: 10.1029/2009JD012346.

Deng M, Mace GG, Wang Z, Lawson RP. 2012. Evaluation of several *A-Train* ice cloud retrieval products with in situ measurements collected during the Sparticus campaign. *J. Appl. Meteorol.* **52**: 1014–1030.

Diederich M, Ryzhkov A, Simmer C, Zhang P, Trömel S. 2015a. Use of specific attenuation for rainfall measurement at X-band radar wavelengths – Part 1: Radar calibration and partial beam blockage estimation. *J. Hydrometeorol.* **16**: 487–502, doi: 10.1175/JHM-D-14-0066.1.

Diedrich H, Preusker R, Lindstrot R, Fischer J. 2015b. Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces. *Atmos. Meas. Tech.* **8**: 823–836.

Di Girolamo P, Summa D, Ferretti R. 2009. Multiparameter Raman lidar measurements for the characterization of a dry stratospheric intrusion event. *J. Atmos. Oceanic Technol.* **26**: 1742–1762, doi: 10.1175/2009JTECHA1253.1.

Dipankar A, Stevens B, Heinze R, Moseley C, Zängl G, Giorgetta M, Brdar S. 2015. Large eddy simulations using the general circulation model ICON. *J. Adv. Model. Earth Syst.* **7**: 963–986, doi: 10.1002/2015MS000431.

Doms G, Schättler U. 2004. 'A description of the nonhydrostatic regional model LM. Part II: Physical parameterization', Technical report. Deutscher Wetterdienst: Offenbach am Main, Germany. http://www.cosmo-model.org/public/documentation.htm (accessed 8 December 2016).

Eckley IA, Nason GP, Treloar RL. 2010. Locally stationary wavelet fields with application to the modelling and analysis of image texture. *J. R. Stat. Soc.* **59**: 595–616.

Edelsbrunner H, Letscher D, Zomorodian A. 2002. Topological persistence and simplification. *Discrete Comput. Geom.* **28**: 511–533.

Eikenberg S, Köhler C, Seifert A, Crewell S. 2015. How microphysical choices affect simulated infrared brightness temperatures. *Atmos. Res.* **156**: 67–79, doi: 10.1016/j.atmosres.2014.12.010.

Engelmann R, Kanitz T, Baars H, Heese B, Althausen D, Skupin A, Wandinger U, Komppula M, Stachlewska IS, Marinou VAE, Mattis I, Linn H, Ansmann A. 2016. The automated multiwavelength Raman polarization and water-vapor lidar PollyXT: The neXT generation. *Atmos. Meas. Tech.* **9**: 1767–1784, doi: 10.5194/amt-9-1767-2016.

Gao BC, Kaufman YJ. 2003. Water vapor retrievals using moderate resolution imaging spectroradiometer (MODIS) near-infrared channels. *J. Geophys. Res.* **108**: 4389, doi: 10.1029/2002JD003023.

Gassmann A, Herzog HJ. 2008. Towards a consistent numerical compressible non-hydrostatic model using generalized Hamiltonian tools. *Q. J. R. Meteorol. Soc.* **134**: 1597–1613, doi: 10.1002/qj.297.

Gendt G, Dick G, Reigber C, Tomassini M, Liu Y, Ramatschi M. 2004. Near real time GPS water vapor monitoring for numerical weather prediction in Germany. *J. Meteorol. Soc.* **82**: 361–370.

Görsdof U, Bauer-Pfundstein M, Peters G, Vavriv D, Vinogradov V, Volkov V. 2015. A 35-GHz polarimetric Doppler radar for long-term observations of cloud parameters – Description of system and data processing. *J. Atmos. Oceanic Technol.* **32**: 675–690, doi: 10.1175/JTECH-D-14-00066.1.

Gregory JM, Webb MJ. 2008. Tropospheric adjustment induces a cloud component in CO2 forcing. *J. Clim.* **21**: 58–71, doi: 10.1175/2007JCLI1834.1.

Haeffelin M, Angelini F, Morille Y, Martucci G, Frey S, Gobbi GP, Lolli S, O'Dowd CD, Sauvage L, Xueref-Rémy I, Wastine B, Feist DG. 2012. Evaluation of mixing-height retrievals from automatic profiling lidars and ceilometers in view of future integrated networks in Europe. *Boundary Layer Meteorol.* **143**: 49–75, doi: 10.1007/s10546-011-9643-z.

Hammann E, Behrendt A, Mounier FL, Wulfmeyer V. 2015. Temperature profiling of the atmospheric boundary layer with rotational Raman lidar during the HD(CP)2 observational prototype experiment. *Atmos. Chem. Phys.* **15**: 2867–2881.

Hande LB, Engler C, Hoose C, Tegen I. 2015. Seasonal variability of Saharan desert dust and ice nucleating particles over Europe. *Atmos. Chem. Phys.* **15**: 4389–4397, doi: 10.5194/acp-15-4389-2015.

Hande LB, Engler C, Hoose C, Tegen I. 2016. Parameterizing cloud condensation nuclei concentrations during HOPE. *Atmos. Chem. Phys.* **16**: 12059–12079, doi: 10.5194/acp-16-12059-2016.

Hanley KE, Plant RS, Stein THM, Hogan RJ, Nicol JC, Lean HW, Halliwell C, Clark PA. 2015. Mixing-length controls on high-resolution simulations of convective storms. *Q. J. R. Meteorol. Soc.* **141**: 272–284, doi: 10.1002/qj.2356.

Heinze R, Moseley C, Böske LN, Muppa SK, Maurer V, Raasch S, Stevens B. 2016. Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign. *Atmos. Chem. Phys. Discuss.*, doi: 10.5194/acp-2016-498.

Heise E, Ritter B, Schrodin E. 2006. 'Operational implementation of the multilayer soil model TERRA', Technical report. Deutscher Wetterdienst: Offenbach am Main, Germany. http://www.cosmo-model.org (accessed 8 December 2016).

Helmert K, Tracksdorf P, Steinert J, Werner M, Frech M, Rathmann N, Hengstebeck T, Mott M, Schumann S, Mammen T. 2014. DWD's new radar

network and post-processing algorithm chain. In *8th European Conference on Radar in Meteorology and Hydrology. Garmisch-Partenkirchen, Germany, p.6. http://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/237_Helmert.pdf (accessed 8 December 2016)*.

Heus T, Seifert A. 2013. Automated tracking of shallow cumulus clouds in large domain, long duration large eddy simulations. *Geosci. Model Dev.* **6**: 1261–1273.

Heus T, van Heerwaarden CC, Jonker HJJ, Siebesma AP, Axelsen S, van den Dries K, Geoffroy O, Moene AF, Pino D, de Roode SR, de Arellano JVG. 2010. Formulation of the Dutch atmospheric large-eddy simulation (DALES) and overview of its applications. *Geosci. Model Dev.* **3**: 415–444, doi: 10.5194/gmd-3-415-2010.

Hogan RJ, Illingworth AJ. 2000. Deriving cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.* **126**: 2903–2909.

Hogan RJ, Westbrook CD. 2014. Equation for the microwave backscatter cross section of aggregate snowflakes using the self-similar Rayleigh–Gans approximation. *J. Atmos. Sci.* **71**: 3292–3301, doi: 10.1175/JAS-D-13-0347.1.

Hogan RJ, Mittermaier MP, Illingworth AJ. 2006. The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model. *J. Appl. Meteorol.* **45**: 301–317, doi: 10.1175/JAM2340.1.

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

Holloway C, Woolnough S, Lister G. 2012. Precipitation distributions for explicit versus parameterized convection in a large-domain high-resolution tropical case study. *Q. J. R. Meteorol. Soc.* **138**: 1692–1708, doi: 10.1002/qj.1903.

Holloway CE, Petch JC, Beare RJ, Bechtold P, Craig GC, Derbyshire SH, Donner LJ, Field PR, Gray SL, Marsham JH, Parker DJ, Plant RS, Roberts NM, Schultz DM, Stirling AJ, Woolnough SJ. 2014. Understanding and representing atmospheric convection across scales: Recommendations from the meeting held at Dartington Hall, Devon, UK, 28–30 January 2013. *Atmos. Sci. Lett.* **15**: 348–353, doi: 10.1002/asl2.508.

Illingworth AJ, Hogan RJ, O'Connor EJ, Bouniol D, Delano J, Pelon J, Protat A, Brooks ME, Gaussiat N, Wilson DR, Donovan DP, Baltink HK, van Zadelhoff GJ, Eastment JD, Goddard JWF, Wrench CL, Haeffelin M, Krasnov OA, Russchenberg HWJ, Piriou JM, Vinit F, Seifert A, Tompkins AM, Willn U. 2007. Cloudnet – Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Am. Meteorol. Soc.* **88**: 883–898, doi: 10.1175/BAMS-88-6-883.

Jerger D. 2014. 'Radar forward operator for verification of cloud resolving simulations within the COSMO-model', PhD thesis. Karlsruhe Institute of Technology: Karlsruhe, Germany.

Jubair MI, Alim U, Röber N, Clyne J, Mahdavi-Amiri A, Samavati F. 2015. 'Multiresolution visualization of digital earth data via hexagonal box-spline wavelets'. In *IEEE Scientific Visualization Conference (SciVis)*, Chicago, IL.

Kalthoff N, Adler B, Wieser A, Kohler M, Träumner K, Handwerker J, Corsmeier U, Khodayar S, Lambert D, Kopmann A, Kunka N, Dick G, Ramatschi M, Wickert J, Kottmeier C. 2013. KITcube – a mobile observation platform for convection studies depolyed during HyMeX. *Meteorol. Z.* **22**: 633–647, doi: 10.1127/0941-2948/2013/0542.

Kärcher B, Lohmann U. 2002. A parameterization of cirrus cloud formation: Homogeneous freezing of supercooled aerosols. *J. Geophys. Res.* **107**: AAC 4-1-AAC 4-10, doi: 10.1029/2001JD000470.

Kärcher B, Hendricks J, Lohmann U. 2006. Physically based parameterization of cirrus cloud formation for use in global atmospheric models. *J. Geophys. Res.* **111**: D01205, doi: 10.1029/2005JD006219.

Kawai H, Teixeira J. 2012. Probability density functions of liquid water path and total water content of marine boundary-layer clouds: Implications for cloud parameterization. *J. Clim.* **25**: 2162–2177, doi: 10.1175/JCLI-D-11-00117.1.

Keil C, Tafferner A, Reinhardt T. 2006. Synthetic satellite imagery in the Lokal-Modell. *Atmos. Res.* **82**: 19–25.

Khairoutdinov MF, Krueger SK, Moeng CH, Bogenschutz PA, Randall DA. 2009. Large-eddy simulation of maritime deep tropical convection. *J. Adv. Model. Earth Syst.* **1**: 15, doi: 10.3894/JAMES.2009.1.15.

King MD, Kaufman YJ, Menzel WP, Tanré D. 1992. Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS). *IEEE Trans. Geosci. Remote Sens.* **30**: 1–27.

Klemp JB, Dudhia J, Hassiotis AD. 2008. An upper gravity-wave absorbing layer for NWP applications. *Mon. Weather Rev.* **136**: 3987–4004, doi: 10.1175/2008MWR2596.1.

Kollias P, Rémillard J, Luke E, Szyrmer W. 2011. Cloud radar Doppler spectra in drizzling stratiform clouds: 1. Forward modeling and remote sensing applications. *J. Geophys. Res.* **116**: D13201, doi: 10.1029/2010JD015237.

Kostka PM, Weissmann M, Buras R, Mayer B, Stiller O. 2014. Observation operator for visible and near-infrared satellite reflectances. *J. Atmos. Oceanic Technol.* **31**: 1216–1233, doi: 10.1175/JTECH-D-13-00116.1.

Kuhn A, Engelke W, Flatken M, Hege HC, Gerndt A, Hotz I. 2015. *Topology-based Methods in Visualization*. Kurhaus Trifels: Annweiler, Germany.

Lenschow DH, Wyngaard JC. 2010. Mean-field and second-moment budgets in a baroclinic, convective boundary layer. *J. Atmos. Sci.* **7**: 1313–1326.

Lenschow D, Wyngaard J, Pennell W. 1980. Mean-field and second-moment budgets in a baroclinic, convective boundary layer. *J. Atmos. Sci.* **37**: 1313–1326.

Lenschow DH, Mann J, Kristensen L. 1994. How long is long enough when measuring fluxes and other turbulence statistics? *J. Atmos. Oceanic Technol.* **11**: 661–673, doi: 10.1175/1520-0426(1994)011¡0661:HLILEW¿2.0.CO;2.

Lenschow DH, Wulfmeyer V, Senff C. 2000. Measuring second-through fourth-order moments in noisy data. *J. Atmos. Oceanic Technol.* **17**: 1330–1347.

Leuenberger D, Koller M, Fuhrer O, Schär C. 2010. A generalization of the SLEVE vertical coordinate. *Mon. Weather Rev.* **138**: 3683–3689, doi: 10.1175/2010MWR3307.1.

Lilly DK. 1962. On the numerical simulation of buoyant convection. *Tellus* **14**: 148–172, doi: 10.1111/j.1962.tb00128.x.

Lilly DK. 1967. The representation of small-scale turbulence in numerical simulation experiments In *IBM Scientific Computing Symposium on Environmental Sciences*: 195–210. Thomas J. Watson Research Center: Yorktown Heights, NY.

Löhnert U, Schween JH, Acquistapace C, Ebell K, Maahn M, Barreraverdejo M, Hirsikko A, Bohn B, Knaps A, O'Connor E, Simmer C, Wahner A, Crewell S. 2015. JOYCE: Jülich observatory for cloud evolution. *Bull. Am. Meteorol. Soc.* **96**: 1157–1174, doi: 10.1175/BAMS-D-14-00105.1.

Lorenz EN. 1960. Energy and numerical weather prediction. *Tellus* **12**: 364–373, doi: 10.1111/j.2153-3490.1960.tb01323.x.

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

Love BS, Matthews AJ, Lister GMS. 2011. The diurnal cycle of precipitation over the maritime continent in a high-resolution atmospheric model. *Q. J. R. Meteorol. Soc.* **137**: 934–947, doi: 10.1002/qj.809.

Maahn M, Löhnert U, Kollias P, Jackson R, McFarquhar G. 2015. Developing and evaluating ice cloud parameterizations for forward modeling of radar moments using in situ aircraft observations. *J. Atmos. Oceanic Technol.* **32**: 880–903, doi: 10.1175/JTECH-D-14-00112.1.

Macke A, Seifert P, Baars H, Beekmans C, Behrendt A, Bohn B, Madhavan BL, Brügmann B, Hammann E, Hirsikko A, Bühl J, Cacciani M, Corsmeier U, Crewell S, Damian T, Di Girolamo P, Graf A, Handwerker J, Kalisch J, Kalthoff N, Kinne S, Kohler M, Linné H, Löhnert U, Metzendorf S, Muppa SK, Pfeiffer M, Pospichal B, Simmer C, Späth F, Riede A, Ruhtz T, Stelitano D, Scaccione A, Serikov I, Summa D, Träumner K, Wieser A, Wulfmeyer V. 2016. The HD(CP)2 Observational Prototype Experiment HOPE – an overview. *Atmos. Chem. Phys.*, doi: 10.5194/acp-2016-990.

Madhavan BL, Kalisch J, Macke A. 2016. Shortwave surface radiation network for observing small-scale cloud inhomogeneity fields. *Atmos. Meas. Tech.* **9**: 1153–1166, doi: 10.5194/amt-9-1153-2016.

Maronga B, Gryschka M, Heinze R, Hoffmann F, Kanani-Sühring F, Keck M, Ketelsen K, Letzel MO, Sühring M, Raasch S. 2015. The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: Model formulation, recent developments, and future perspectives. *Geosci. Model Dev. Discuss.* **8**: 1539–1637, doi: 10.5194/gmdd-8-1539-2015.

Marshak A, Platnick S, Vrnai T, Wen G, Cahalan RF. 2006. Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes. *J. Geophys. Res.* **111**: D09207, doi: 10.1029/2005JD006686.

Maurer V, Kalthoff N, Wieser A, Kohler M, Mauder M, Gantner L. 2016. Observed spatiotemporal variability of boundary-layer turbulence over flat, heterogeneous terrain. *Atmos. Chem. Phys.* **16**: 1377–1400, doi: 10.5194/acp-16-1377-2016.

McFarquhar GM, Iacobellis S, Somerville RCJ. 2003. SCM simulations of tropical ice clouds using observationally based parameterizations of microphysics. *J. Clim.* **16**: 1643–1664, doi: 10.1175/1520-0442(2003)016¡1643:SSOTIC¿2.0.CO;2.

Mellor GL, Yamada T. 1974. A hierarchy of turbulence closure models for planetary boundary layers. *J. Atmos. Sci.* **31**: 1791–1806, doi: 10.1175/1520-0469(1974)031¡1791:AHOTCM¿2.0.CO;2.

Milovac J, Warrach-Sagi K, Behrendt A, Späth F, Ingwersen J, Wulfmeyer V. 2016. Investigation of PBL schemes combining the WRF model simulations with scanning water vapor differential absorption lidar measurements. *J. Geophys. Res.* **121**: 624–649, doi: 10.1002/2015JD023927.

Mishchenko MI. 2000. Calculation of the amplitude matrix for a nonspherical particle in a fixed orientation. *Appl. Opt.* **39**: 1026–1031.

Miura H. 2007. An upwind-biased conservative advection scheme for spherical hexagonal-pentagonal grids. *Mon. Weather Rev.* **135**: 4038–4044, doi: 10.1175/2007MWR2101.1.

Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.* **102**: 16663–16682, doi: 10.1029/97JD00237.

Morille Y, Haeffelin M, Dobrinski P, Pelon J. 2007. STRAT: An automated algorithm to retrieve the vertical structure of the atmosphere from single-channel lidar data. *J. Atmos. Oceanic Technol.* **24**: 761–775, doi: 10.1175/JTECH2008.1.

Muppa SK, Behrendt A, Späth F, Wulfmeyer V, Metzendorf S, Riede A. 2016. Turbulent humidity fluctuations in the convective boundary layer: Case studies using water vapour differential absorption lidar measurements. *Boundary Layer Meteorol.* **158**: 43–66, doi: 10.1007/s10546-015-0078-9.

Nakajima T, King MD. 1990. Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory. *J. Atmos. Sci.* **47**: 1878–1893.

Narcowich FJ, Ward JD. 1994. Generalized Hermite interpolation via matrix-valued conditionally positive definite functions. *Math. Comput.* **63**: 661–687.

Neggers RAJ, Jonker HJJ, Siebesma AP. 2003. Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.* **60**: 1060–1074.

Neggers RAJ, Heus T, Siebesma AP. 2011. Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations. *J. Geophys. Res.* **116**: D21202, doi: 10.1029/2011JD015650.

Neggers RAJ, Siebesma AP, Heus T. 2012. Continuous single-column model evaluation at a permanent meteorological supersite. *Bull. Am. Meteorol. Soc.* **93**: 1389–1400, doi: 10.1175/BAMS-D-11-00162.1.

Noppel H, Blahak U, Seifert A, Beheng KD. 2010. Simulations of a hailstorm and the impact of CCN using an advanced two-moment cloud microphysical scheme. *Atmos. Res.* **96**: 286–301.

Platnick S, King MD, Meyer KG, Wind G, Amarasinghe N, Marchant B, Arnold GT, Zhang Z, Hubanks PA, Ridgway B, Riedi J. 2014. 'MODIS cloud optical properties: User guide for the collection 6 Level-2 MOD06/MYD06 product and associated Level-3 Datasets'. Version 0.9 (beta) 17.

Prein AF, Langhans W, Fosser G, Ferrone A, Ban N, Goergen K, Keller M, Tölle M, Gutjahr O, Feser F, Brisson E, Kollet S, Schmidli J, van Lipzig NPM, Leung R. 2015. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Rev. Geophys.* **53**: 323–361, doi: 10.1002/2014RG000475.

Quaas J. 2012. Evaluating the critical relative humidity as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *J. Geophys. Res.* **117**: D09208, doi: 10.1029/2012JD017495.

Richardson H, Basu S, Holtslag AAM. 2013. Improving stable boundary-layer height estimation using a stability-dependent critical bulk Richardson number. *Boundary Layer Meteorol.* **148**: 93–109.

Roebeling RA, Feijt AJ, Stammes P. 2006. Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *J. Geophys. Res.* **111**: D20210, doi: 10.1029/2005JD006990.

Rosch J, Heus T, Brueck M, Salzmann M, Mülmenstädt J, Schlemmer L, Quaas J. 2015. Analysis of diagnostic climate model cloud parametrizations using large-eddy simulations. *Q. J. R. Meteorol. Soc.* **141**: 2199–2205.

Rose T, Crewell S, Löhnert U, Simmer C. 2005. A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere. *Atmos. Res.* **75**: 183–200.

Ruppert T. 2007. 'Vector field reconstruction by radial basis functions', Master's thesis. Department of Mathematics, Technical University Darmstadt: Darmstadt.

Ryzhkov A, Zhang P, Reeves H, Kumjian M, Tschallener T, Simmer C, Trömel S. 2016. Quasi-vertical profiles – a new way to look at polarimetric radar data. *J. Atmos. Oceanic Technol.* **33**: 551–562, doi: 10.1175/JTECH-D-15-0020.1.

Saunders R, Matricardi M, Brunel P. 1999. An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. R. Meteorol. Soc.* **125**: 1407–1425.

Schalkwijk J, Jonker HJJ, Siebesma AP, Van Meijgaard E. 2015. Weather forecasting using GPU-based large-eddy simulations. *Bull. Am. Meteorol. Soc.* **96**: 715–723, doi: 10.1175/BAMS-D-14-00114.1.

Scheck L, Frèrebeau P, Buras-Schnell R, Mayer B. 2016. A fast radiative transfer method for the simulation of visible satellite imagery. *J. Quant. Spectrosc. Radiat. Transfer* **175**: 54–67.

Schlemmer L, Hohenegger C. 2014. The formation of wider and deeper clouds as a result of cold-pool dynamics. *J. Atmos. Sci.* **71**: 2842–2858, doi: 10.1175/JAS-D-13-0170.1.

Schween JH, Hirsikko A, Löhnert U, Crewell S. 2014. Mixing-layer height retrieval with ceilometer and Doppler lidar: From case studies to long-term assessment. *Atmos. Meas. Tech.* **7**: 3685–3704, doi: 10.5194/amt-7-3685-2014.

Seemann SW, Li J, Gumley LE, Strabala KI, Menzel WP. 2003. Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances. In *Third International Asia-Pacific Environmental Remote Sensing of the Atmosphere, Ocean, Environment, and Space*: 168–176. International Society for Optics and Photonics: Hangzhou, China.

Seifert A. 2008. On the parameterization of evaporation of raindrops as simulated by a one-dimensional rainshaft model. *J. Atmos. Sci.* **65**: 3608–3619.

Seifert A, Beheng KD. 2001. A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmos. Res.* **59–60**: 265–281.

Seifert A, Beheng K. 2006. A two-moment cloud microphysics parameterization for mixed-phase clouds. Part I: Model description. *Meteorol. Atmos. Phys.* **92**: 45–66.

Senf F, Deneke H. 2017. Uncertainties in synthetic Meteosat SEVIRI infrared brightness temperatures in the presence of cirrus clouds and implications for evaluation of cloud microphysics. *Atmos. Res.* **183**: 113–129, doi: 10.1016/j.atmosres.2016.08.012.

Sherwood SC, Bony S, Boucher O, Bretherton C, Forster PM, Gregory JM, Stevens B. 2015. Adjustments in the forcing-feedback framework for understanding climate change. *Bull. Am. Meteorol. Soc.* **96**: 217–228, doi: 10.1175/BAMS-D-13-00167.1.

Smiatek G, Rockel B, Schättler U. 2008. Time invariant data preprocessor for the climate version of the COSMO model (COSMO-CLM). *Meteorol. Z.* **17**: 395–405.

Smiatek G, Helmert J, Gerstner EM. 2015. Impact of land use and soil data specifications on COSMO-CLM simulations in the CORDEX-MED area. *Meteorol. Z.* **25**: 1–16.

Späth F, Behrendt A, Muppa SK, Metzendorf S, Riede A, Wulfmeyer V. 2016. 3-D water vapor field in the atmospheric boundary layer observed with scanning differential absorption lidar. *Atmos. Meas. Tech.* **9**: 1701–1720.

Stein THM, Hogan RJ, Clark PA, Halliwell CE, Hanley KE, Lean HW, Nicol JC, Plant RS. 2015. The DYMECS project – a statistical approach for the evaluation of convective storms in high-resolution NWP models. *Bull. Am. Meteorol. Soc.* **96**: 939–951, doi: 10.1175/BAMS-D-13-00279.1.

Steinke S, Eikenberg S, Löhnert U, Dick G, Klocke D, Di Girolamo P, Crewell S. 2015. Assessment of small-scale integrated water vapour variability during HOPE. *Atmos. Chem. Phys.* **15**: 2675–2692, doi: 10.5194/acp-15-2675-2015.

Tiedtke M. 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Weather Rev.* **117**: 1779–1799, doi: 10.1175/1520-0493(1989)117¡1779:ACMFSF>2.0.CO;2.

Trömel S, Ryzhkov AV, Zhang P, Simmer C. 2014. Investigations of backscatter differential phase in the melting layer. *J. Appl. Meteorol. Climatol.* **53**: 2344–2359, doi: 10.1175/jamc-d-14-0050.1.

Vial J, Dufresne JL, Bony S. 2013. On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.* **41**: 3339–3362, doi: 10.1007/s00382-013-1725-9.

Wan H, Giorgetta MA, Zängl G, Restelli M, Majewski D, Bonaventura L, Fröhlich K, Reinert D, Rpodas P, Kornblueh L, Förstner J. 2013. The ICON-1.2 hydrostatic atmospheric dynamical core on triangular grids – Part 1: Formulation and performance of the baseline version. *Geosci. Model Dev.* **6**: 735–763, doi: 10.5194/gmd-6-735-2013.

Weniger M, Kapp F, Friederichs P. 2015. Spatial verification using wavelet transforms: A review. *Q. J. R. Meteorol. Soc.*, doi: 10.1002/qj.2881.

Wernli H, Paulat M, Hagen M, Frei C. 2008. SAL-A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Weather Rev.* **136**: 4470–4487.

Wolke R, Knoth O, Hellmuth OK, Schröder W, Renner E. 2004. The parallel model system LM-MUSCAT for chemistry-transport simulations: Coupling scheme, parallelization and application. In *Parallel Computing: Software Technology, Algorithms, Architectures, and Applications*, Joubert G, Nagel W, Peters F, Walter W. (eds.): 363–370. Elsevier: Amsterdam.

Wolke R, Schroeder W, Schroedner R, Renner E. 2012. Influence of grid resolution and meteorological forcing on simulated European air quality: A sensitivity study with the modeling system COSMO-MUSCAT. *Atmos. Environ.* **53**: 110–130, doi: 10.1016/j.atmosenv.2012.02.085.

Wolters ELA, Deneke HM, van den Hurk BJJM, Meirink JF, Roebeling RA. 2010. Broken and inhomogeneous cloud impact on satellite cloud particle effective radius and cloud-phase retrievals. *J. Geophys. Res.* **115**: D10214, doi: 10.1029/2009JD012205.

Wulfmeyer V, Hardesty R, Turner DD, Behrendt A, Cadeddu M, Di Girolamo P, Schlüssel P, Van Baelen J, Zus F. 2015. A review of the remote sensing of lower-tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles. *Rev. Geophys.* **53**: 819–895, doi: 10.1002/2014RG000476.

Wulfmeyer V, Muppa SK, Behrendt A, Hammann E, Späth F, Sorbjan Z, Turner D, Hardesty R. 2016. Determination of the convective boundary layer entrianment fluxes, dissipation rates and the molecular destruction of variances: Theoretical description and a strategy for its confirmation with a novel lidar system synergy. *J. Atmos. Sci.* **73**: 667–692.

Zängl G, Reinert D, Ripodas MP, Baldauf M. 2015. The ICON (ICOsahedral Nonhydrostatic) modelling framework of DWD and MPI-M: Description of the nonhydrostatic dynamical core. *Q. J. R. Meteorol. Soc.* **141**: 563–579, doi: 10.1002/qj.2378.

Zeng Y. 2013. 'Efficient radar forward operator for operational data assimilation within the COSMO-model', PhD thesis. Karlsruhe Institute of Technology: Karlsruhe.

Zeng Y, Blahak U, Jerger D. 2016. An efficient modular volume scanning radar forward operator for NWP-models: Description and coupling to the COSMO-model. *Q. J. R. Meteorol. Soc.* **142**: 3234–3256, doi: 10.1002/qj.2904.

Zhang D, Wapler K, Blahak U. 2015. Verification and uncertainty analysis of clouds and precipitation in high resolution models using 3D radar data. *12th European Conference on Applications of Meteorology (ECAM)*, 7-11 September 2015, Sofia, Bulgaria.