Received: 23 February 2025 | Revised: 27 July 2025 | Accepted: 30 August 2025

DOI: 10.1002/qj.70037

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Where numerical weather prediction meets large-eddy simulations in a convective boundary layer during the FESSTVaL field experiment

Mirjana Sakradzija[1,2] | Maike Ahlgrimm[2,3] | Frank Beyrich[4] | Chiel van Heerwaarden[5] | Eileen Päschke[4] | Ulrich Görsdorf[4] | Bastian Kirsch[6] | Juerg Schmidli[2,7] | Noviana Dewani[7]

[1]Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany

[2]Hans Ertel Centre for Weather Research (HErZ), Deutscher Wetterdienst, Offenbach am Main, Germany

[3]Deutscher Wetterdienst, Offenbach am Main, Germany

[4]Lindenberg Meteorological Observatory–Richard Aßmann Observatory, Deutscher Wetterdienst, Lindenberg, Germany

[5]Wageningen University & Research, Wageningen, the Netherlands

[6]Institute of Meteorology and Climate Research Troposphere Research (IMKTRO), Karlsruhe Institute of Technology, Karlsruhe, Germany

[7]Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany

Correspondence
Mirjana Sakradzija, Department of Geography, Ludwig-Maximilians-Universität, Luisenstraße 37, 80333 Munich, Germany.
Email: m.sakradzija@lmu.de

Funding information
Hans-Ertel Centre for Weather Research, Grant/Award Number: 4818DWDP4; Dutch Research Council, Grant/Award Number: VI.Vidi.192.068; Bundesministerium für Verkehr; BMV, Grant/Award Number: 4823DWDP3

Abstract

This study leverages the Field Experiment on Sub-mesoscale Spatio-Temporal Variability in Lindenberg (FESSTVaL), including comprehensive observations of the surface, atmospheric boundary layer (ABL), and clouds, to compare the performance of a numerical weather prediction (NWP) model at sub-km resolutions with traditional large-eddy simulations (LES). This comparison is both relevant and timely: as the typical grid spacings of both techniques are converging, so should their results if the LES is assumed to be a good virtual laboratory for the ABL and if NWP is resolving the turbulence well. The representative of NWP is the Icosahedral Nonhydrostatic (ICON) model run at horizontal resolutions ranging from 2.5 km to 78 m. The LES model MicroHH is run at resolutions from 75 m to 38 m. ICON is set up in a limited area with realistic boundary conditions and heterogeneous land surface, whereas the setup of MicroHH is doubly periodic above a homogeneous surface and flat terrain. We focus our comparison at a horizontal grid resolution of about 78 m, where the two models overlap. ICON can represent the ABL processes with high fidelity. It approaches the performance of the MicroHH-LES in representing surface turbulent and radiation fluxes due to better-resolved ABL dynamics, clouds, and land-surface properties at sub-km resolutions. The modeled turbulence shows good agreement with observations, highlighting the equal importance of resolved and subgrid turbulent mixing at sub-km resolutions. Our modeling setup also reproduces deep convective cold pools, although with only a qualitatively realistic onset and development. In addition to the FESSTVaL observations used here, extensive datasets are available for follow-up studies focusing on specific processes. The complete set of ICON simulations covers the intensive observation period between June 5 and July 5, 2021, and can be used in process studies as well as providing a benchmark for model development.

KEYWORDS
clouds, cold pools, FESSTVaL, field experiment, large-eddy simulations, surface radiation, turbulence

1 | INTRODUCTION

Numerical weather prediction (NWP) is performed today at grid spacings less than 1 km at several leading operational weather centers. The possibility of using the same model configuration across various applications and spatial scales, ranging from local large-eddy-resolving cases to global NWP and climate research, has become real. We explore one such possibility by using an operational NWP model at sub-km resolutions and examining its ability to represent relevant processes in the atmospheric boundary layer (ABL) in comparison with a standard large-eddy simulation (LES) and micro- and sub-mesoscale meteorological observations.

Applications of NWP models at sub-km spatial resolutions usually cover regions of meso-γ to meso-β scales, aimed at processes which are otherwise challenging to represent in traditional NWP. Such applications aim to resolve atmospheric flows over complex mountainous terrain better, such as valley winds (Schmidli & Quimbayo-Duarte, 2023), snow conditions during winter (Vionnet *et al.*, 2015), coastline terrain at high latitudes (Valkonen *et al.*, 2020), or urban coastal areas where improvements have been observed in the representation of mesoscale flows such as sea–land breezes (Leroyer *et al.*, 2014). Furthermore, sub-km resolutions are beneficial for processes that are difficult to parameterize, such as, for example, stable conditions and radiation fog (Smith *et al.*, 2021). Although some systematic model deficiencies remain at sub-km resolutions, such as considerable sensitivities to surface properties (Smith *et al.*, 2021), there is generally a better agreement with observations.

Although they are very promising, due to the possibility of resolving sub-mesoscale and mesoscale atmospheric flows, the sub-km applications of NWP are not without challenges (Dudhia, 2022). The processes that evolve at spatial scales close to the grid spacing of a model are entering a so-called "gray zone", which for sub-km resolutions includes turbulence, shallow convection, and shallow clouds (Field *et al.*, 2017; Honnert *et al.*, 2020; Wyngaard, 2004). At sub-km resolutions, large turbulent eddies (or shallow convective updrafts) emerge as an explicit numerical solution of the model dynamics. However, these eddies develop spatial characteristics and intensities driven by the grid spacing of a model and cannot interact with smaller-scale turbulence due to artificial scale separation (e.g., Ching *et al.*, 2014; Kealy *et al.*, 2019). At a subgrid scale, horizontal turbulent gradients must also be taken into account in otherwise one-dimensional (1D) turbulent closures in NWP. How to consolidate these insufficiently resolved but also inadequately parameterized turbulent scales remains an open research question.

At the other end of the gray zone in a daytime convective ABL are LES aimed at resolving the most energetic large turbulent eddies properly, while still parameterizing isotropic turbulent mixing at subgrid scales. Starting with the pioneering work of Smagorinsky (Smagorinsky, 1963) and Deardorff (Deardorff, 1970) with the first applications on turbulent flows at large Reynolds numbers, LES have been widely used in studies from atmospheric boundary layers (Mason, 1989) to tropical cyclones (Rotunno *et al.*, 2009), with the aim of understanding processes and the development of process parameterizations. LES solve the Navier–Stokes equations numerically, applying low-pass filtering for the smallest turbulent scales, which would otherwise require direct numerical simulation that is still too computationally expensive for most of the applications in atmospheric sciences. Traditionally, LES are performed with idealized representations of the initial conditions, over homogeneous flat surfaces, and with cyclic lateral boundary conditions (Moeng & Sullivan, 2015; van Heerwaarden *et al.*, 2017). Even though NWP is approaching LES, idealized LES remain vital for atmospheric modeling, since they are the main method from which parametrizations used in NWP are derived.

During the past decade, LES have evolved from idealized case studies representing average and uniform meteorological conditions to more realistic representations of the land surface and large-scale forcings with a fully interactive coupling of the land or ocean and atmospheric components. One of the main advantages of realistic configurations is the ability to compare simulation output with observations directly, but also to complement the observations by a complete three-dimensional (3D) and time-varying realistic representation of the atmosphere. For this purpose, NWP models are being extended to typical LES grid spacings, while switching to Smagorinsky or Deardorff-type turbulence schemes that represent 3D turbulent mixing instead of the simplified 1D mixing used in operational NWP. Thus, these realistic LES have become standard in accompanying observational field experiments (e.g., Bauer *et al.*, 2023; Fast *et al.*, 2019; Goger & Dipankar, 2024; Heinze *et al.*, 2017), due to the possibility of comparing the model output directly with observations of specific meteorological conditions.

Nevertheless, as the resolution of NWP models has increased to sub-km scales, 1D turbulence schemes have been extended to include 3D aspects of turbulence. An example of such a hybrid approach includes shear effects of horizontal winds as a relevant production term of turbulent kinetic energy (TKE) in an Alpine valley (Goger *et al.*, 2018). A more pragmatic approach to blending 1D and 3D turbulence schemes across the gray zone is applied in Boutle*et al.* (2014), Efstathiou and Beare (2015), Ito *et al.* (2015), and Shin and Hong (2015). Scale-adaptive 3D

turbulence schemes aimed at sub-km NWP applications unify horizontal and vertical mixing in an energetically consistent way by solving a 3D prognostic equation for unresolved TKE (Kosović *et al.*, 2020; Zhang *et al.*, 2018). The Icosahedral Nonhydrostatic (ICON) model that we use in this study includes a hybrid turbulence scheme where, in addition to the prognostic 1D TKE equation, 3D shear effects can be included, along with additional interaction terms originating from non-turbulent subgrid flows and a turbulent length-scale dependent on the model resolution (Raschendorfer, 2016). Therefore, in principle, the application of ICON-NWP can be extended to sub-km resolutions.

ICON currently runs operationally at 500-m resolution at the Deutscher Wetterdienst (DWD) using the hybrid 1D turbulence closure with additional terms representing horizontal shear effects (Reinert *et al.*, 2025). In this study, we explore to what degree the current limited-area setup of ICON is able to adapt to sub-km resolution and still represent selected boundary-layer processes adequately. As a reference for the assessment of ICON performance, we include a traditional LES model, MicroHH (van Heerwaarden *et al.*, 2017), which uses a 3D Smagorinsky type turbulence closure typical for LES. We focus our analysis and comparisons at a spatial resolution (approximately 78 m) at which the finest resolution used in our ICON setup overlaps with the coarsest resolution of our MicroHH setup.

The modeling output is evaluated using measurements from the Field Experiment on Sub-mesoscale Spatio-Temporal Variability in Lindenberg (FESSTVaL; Hohenegger *et al.*, 2023). The Lindenberg Meteorological Observatory is located southeast of Berlin in Brandenburg, Germany. FESSTVaL was designed to observe the spatiotemporal variability of near-surface atmospheric processes at scales from about 500 m to 5 km, with the aim of improving understanding of sub-mesoscale atmospheric variability, evaluating numerical weather prediction (NWP) models, and comparing different measurement strategies and instruments for future ground-based observational networks. As a testbed for using FESSTVaL in research and modeling, we propose three case studies representative of clear sky (June 14, 2021), cumuliform cloudy conditions (June 27, 2021), and a deep-convective precipitation event (June 29, 2021). In the first case study on June 14, we focus on the turbulent mixing in the ABL and surface turbulent fluxes of sensible and latent heat. The case of June 27 is ideal for studying shallow convective clouds and variability in surface radiation due to rich cloud growth and development during the day. The case study on June 29 provides the first combined observational, LES, and NWP representation of the spatial variability associated with cold pools resulting from a deep precipitating storm in the area. The storm was captured successfully by the observational ground network and is simulated within the domain of LES and NWP models, although at different locations and with resolution-dependent characteristics. By comparing different model configurations, forcing strategies, spatial resolutions, and domain sizes, we assess current limitations and requirements for successful modeling of the ABL processes, land-surface exchange fluxes, and interactions between ABL and cold-pool dynamics.

This article presents the FESSTVaL field experiment in Section 2 and the numerical models in Section 3. Selected case studies are described in Section 4. Section 5 starts with the first evaluation of the modeled diurnal and vertical changes in thermodynamic properties and winds in the ABL (Section 5.1). Section 5.2 presents the case study on June 14, focusing on the turbulent exchange and mixing. Section 5.3 focuses on surface radiation and clouds on June 27, while Section 5.4 presents June 29 as the benchmark case study of FESSTVaL for deep-convective cold-pool dynamics. Discussion of the resolution dependence of the ABL statistics is provided in Section 6 and a summary of the study is given in Section 7.

2 | THE FIELD EXPERIMENT

The data used in this study have been collected during the FESSTVaL field campaign. The field campaign took place in summer 2021 (from mid-May until the end of August) with an Intensive Observation Period (IOP) from June 5–July 5, 2021. FESSTVaL aimed at studying sub-mesoscale phenomena such as convection, wind gusts, and cold pools. To achieve this, a hierarchical measurement strategy was defined including regional-scale networks of around 100 simple automatic weather stations (basically measuring temperature, pressure, near-surface soil moisture) deployed in an area of about 15-km radius around the boundary-layer field site (in German: Grenzschichtmessfeld, GM) Falkenberg, vertical profile measurements of the ABL using ground-based remote sensing instruments (ceilometers, Doppler lidars, microwave radiometer profilers) at three supersites (GM Falkenberg, Lindenberg, and Birkholz), and spatially resolving/integrating techniques (X-band rain radar, uncrewed aerial vehicles, and an optical–microwave scintillometer). The side length of the triangle formed by the three profiling sites was 5-6 km, and the surface stations were deployed at distances between a few 100 m at and around GM Falkenberg up to about 5 km in the outer part of the measurement area. Moreover, FESSTVaL could rely on the routine measurement program of the Lindenberg Meteorological Observatory–Richard Aßmann Observatory (MOL-RAO). This includes operation of a broader suite of ground-based remote sensing systems (482-MHz

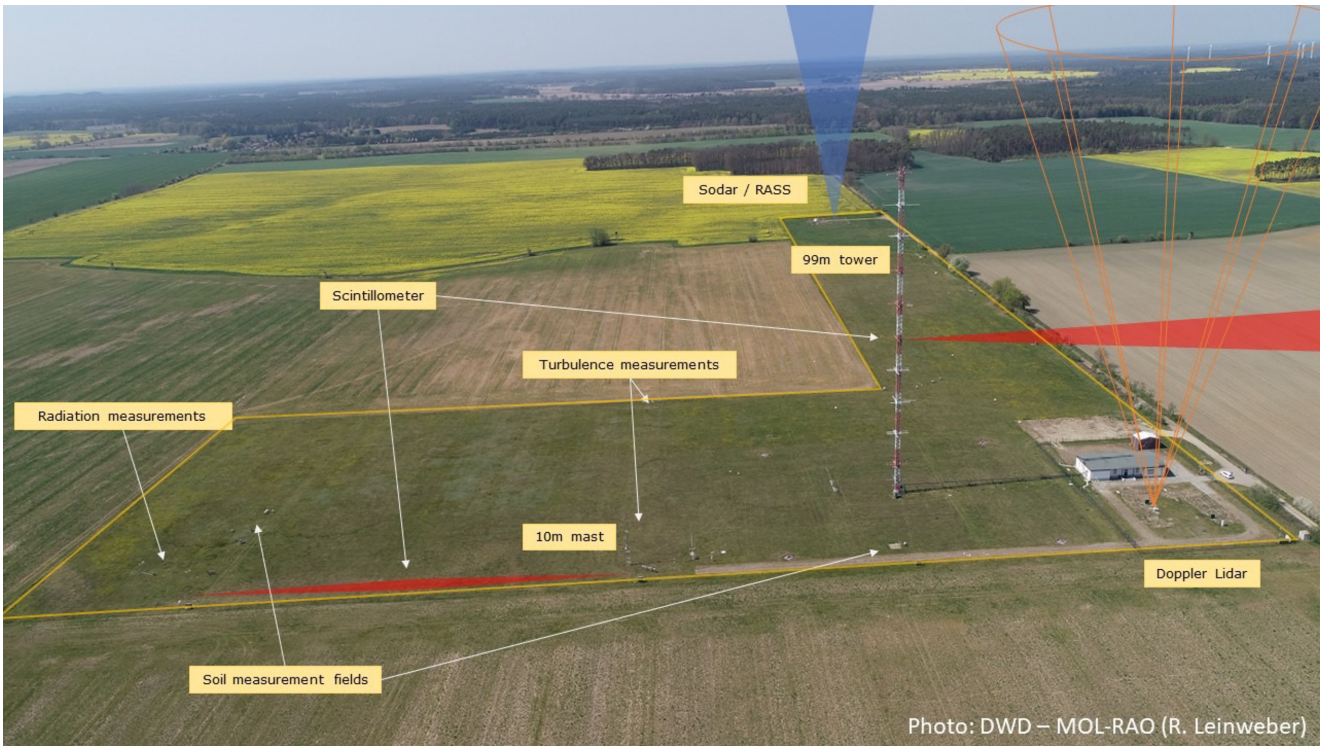

FIGURE 1 Aerial view of GM Falkenberg towards the west with the different measurement complexes and systems indicated. [Colour figure can be viewed at wileyonlinelibrary.com]

radar wind profiler, cloud radar, Raman lidar), four daily operational radiosoundings (launched regularly at 0445, 1045, 1645, and 2245 UTC), the Lindenberg Baseline Surface Radiation Network station, and micrometeorological measurements at GM Falkenberg and at a forest station. A comprehensive overview of FESSTVaL is given in Hohenegger *et al.* (2023). Here we confine ourselves to those measurements that were used for comparison with the model simulations in the following sections.

2.1 | Micrometeorological and in situ turbulence measurements

MOL-RAO performs routine micrometeorological measurements at its boundary-layer field site (GM Falkenberg) located south of Lindenberg and at a forest site (Kehrigk forest) about 10 km west of Falkenberg, in order to characterize the interaction between the atmosphere and the land surface for two vegetation classes typical for the region (grassland, representing low, agricultural vegetation, and a pine forest). These measurements include basic meteorological variables, soil status, and process variables, all relevant components of the surface energy balance, and profile measurements of wind, temperature, and humidity at towers up to a height of 98 m at Falkenberg and 30 m at Kehrigk forest, respectively. An aerial view of the GM Falkenberg site is given in Figure 1.

Data of the turbulent fluxes of momentum and sensible and latent heat are based on eddy-covariance measurements with a 20-Hz sampling rate. Fluxes were calculated as 30-minute averages using the EddyPro (V7.0.9) software. At GM Falkenberg, two eddy-covariance stations are operated at the eastern and western sides of grassland area (see Figure 1), such that the major part of the footprint area comes from the grassland for either of the stations, dependent on the actual wind direction. All other variables are sampled at 1 Hz, pre-averaged in the data loggers to 1-minute basic data, and finally provided as a quality-controlled data product for 10-minute averaging time. Basic principles of the quality-control scheme are described in Beyrich and Adam (2007).

2.2 | Network of surface weather stations

A dense network of surface weather stations distributed over the FESSTVaL domain was set up to quantify the sub-mesoscale variability of the near-surface temperature field. The network consisted of 99 autonomous weather stations (80 APOLLO stations and 19 WXT stations) that were separated by between 0.1 and 4.8 km and measured

TABLE 1 System parameters and scan configurations of the "Streamline/Streamline XR" Doppler lidar systems operated at GM Falkenberg during FESSTVaL.

	DL146	DL78
Instrument specification		
DL type	XR	–
Pulse length	413 ns	180 ns
Number of pulses per ray	30,000	2000
Pulse repetition frequency	10 kHz	10 kHz
Configuration features		
Mode	Vertical stare	Continous conical scan
Zenith angle	0°	54.7°
Azimuthal resolution	N/A	~ 1–2°
Vertical resolution	48 m	30 m
Focus setting	infinity	500 m
Duration of one circular scan	N/A	72 s
Application		
Processed variables	Vertical velocity variance	Turbulence kinetic energy (TKE)
Averaging times	1 min, 30 min	30 min

air temperature at 3-m height above ground. A detailed description of the instruments and the measurements during FESSTVaL can be found in Kirsch *et al.* (2022) and Kirsch *et al.* (2024), respectively.

2.3 | Doppler lidar wind and turbulence measurements

During FESSTVaL, eight "Streamline"/"Streamline XR" Doppler lidar (DL) systems manufactured by Halo Photonics Ltd were operated at the three supersites Falkenberg, Lindenberg, and Birkholz. For comparison purposes with the LES, measurements from two Doppler lidar systems (DL146, DL78) that were installed at the GM Falkenberg site are used here. The two systems differed in terms of their technical specifications and each was configured for different measurement applications (see Table 1).

DL146 was operated exclusively in vertical stare mode, thus providing profiles of the vertical wind vector component at high temporal resolution. Profiles of the vertical velocity variance (w'^2) were derived from a statistical analysis of the w time series at each measurement height over 30-minute averaging times (Dewani *et al.*, 2023). DL78 was used in a conically scanning mode to provide radial velocity measurements along different lines of sight (LOS) from ground to sky in different directions. Therewith the measurements met the required spatiotemporal resolution for a retrieval method introduced by Smalikho and Banakh (2017) to derive Doppler lidar-based turbulence variables such as the TKE. The peculiarity of the approach is that it includes additional correction terms to account for both the well-known underestimation of TKE due to the averaging effect over the pulse volume and an overestimation of TKE due to instrumental error.

Because of the conical scanning strategy, all retrieved turbulence variables represent spatial averages over differently sized circular areas for each measurement height, with a temporal mean value valid for 30 minutes. In FESSTVaL, the turbulence retrieval method by Smalikho and Banakh (2017) was implemented and tested for routine 24/7 measurements to enable a continuous profiling of the boundary layer up to 600-m height. Due to the system configuration (low number of pulses per ray), increased efforts were necessary in terms of a pre-processing of the measured radial velocities to minimize noise contamination (Päschke & Detring, 2024).

2.4 | Radiation measurements

Radiation fluxes at Lindenberg, GM Falkenberg, and the forest site are measured using CMP22/CM 24 pyranometers (Kipp & Zonen, shortwave hemispheric fluxes), CGR4 pyrgeometers (Kipp & Zonen, longwave fluxes at Lindenberg), DD-PIR pyrgeometers (Eppley, longwave fluxes at GM Falkenberg and forest site), and a CH1 pyrheliometer (Kipp & Zonen, direct solar radiation at Lindenberg). All

radiation flux sensors are operated in ventilated shields. The sensors for diffuse shortwave and longwave radiation at Lindenberg are shaded and mounted on a solar tracker, together with the pyrheliometer. In situ calibrations are performed regularly using reference sensors directly traceable to the World Radiometric Reference (WRR) and the World Infrared Standard Group (WISG) for shortwave and longwave radiation, respectively. Quality control follows the recommendations of the World Meteorological Organization (WMO) baseline surface radiation network (BSRN). It includes absolute value range tests and intercomparison with a second independent radiation flux measurement at the same site. The temperature of the emitting sensor surface of the pyrgeometer is checked for plausibility versus ambient air temperature.

During the FESSTVaL field campaign, a complementary grid of fast and low-cost radiometers (Heusinkveld *et al.*, 2023) was set up at the Falkenberg site to measure fluctuations in surface solar irradiance at 10 Hz and 18 wavelengths in a grid of 200 by 250 m using 20 devices. These fast observations provide insight into the properties of cloud shadows and related 3D effects of radiation scattering over the edges of clouds (Mol *et al.*, 2024).

2.5 | Scintillometer measurements

An optical–microwave scintillometer system (OMS) was operated during FESSTVaL over a path length of 4.8 km between the GM Falkenberg and Lindenberg supersites at an effective path height of 43 m. It consisted of a BLS900 optical large-aperture scintillometer (LAS, Scintec AG) and an MWSC-160 microwave scintillometer (MWS, RPG GmbH). While propagating through the turbulent atmosphere, the emitted electromagnetic radiation is scattered by turbulent eddies of different density. This results in high-frequency intensity fluctuations of the electromagnetic signal ("scintillations") recorded at the receiver. Using wave propagation theory and Monin–Obukhov similarity theory, the refractive-index structure parameter of the air at the two wavelengths and the turbulent fluxes of sensible and latent heat can be derived from these measurements. The signal at the receiver represents an integrated effect of the conditions along the path; scintillometers therefore provide area-averaged values of the turbulent fluxes (see Beyrich *et al.*, 2021). Data acquisition, data analysis, and flux calculations were performed with the mwsc.exe software package provided by the manufacturer. Structure parameters and the temperature–humidity correlation coefficient for each 10-minute time interval have been calculated twice based on different settings, that is, using the methods described in Hill (1997) and in Lüdi *et al.* (2005), respectively. The similarity model proposed by Kooijmans and Hartogensis (2016) was then used to derive the heat fluxes from the structure parameters. Using temperature and humidity-profile measurements at the Falkenberg tower and measurements of the radiation budget, the deduced fluxes have been checked for sign consistency with the mean gradients of temperature and humidity and for a violation of the energy budget. In the end, the "most plausible" fluxes from the two methods (Hill, 1997; Lüdi *et al.*, 2005, see above) have been merged to a composite to ensure a higher availability/quality of the fluxes, especially around sunrise and sunset, when the assumptions of the Hill (1997) approach typically fail.

2.6 | Cloud measurements

Two ceilometers CHM15k (Jenoptik/Lufft) were operated at the MOL-RAO and GM Falkenberg sites. The ceilometers send short laser pulses at a wavelength of 1064 nm into the atmosphere and measure the backscattering from molecules, hydrometeors, and aerosols. From the backscatter profiles, cloud bases and aerosol-layer heights are derived as 15-s averages from near the ground up to a maximum height of 15 km with a vertical resolution of 10 or 15 m, respectively. Depending on the cloud optical thickness, up to three cloud-layer heights can be determined. Additional parameters are the vertical visibility, cloud penetration depth, and cloud cover; the latter is derived from the percentage of clouds that have been detected in a given layer during a prescribed time interval. The backscatter profiles are also useful for estimating the convective boundary-layer height.

The Cloudnet dataset is a synergistic product of 35-GHz cloud radar, ceilometer, and multi-frequency microwave radiometer (MWR) measurements. The corresponding retrieval package has been developed in the Cloudnet project (Illingworth *et al.*, 2007). Data are available (24/7) with a temporal and vertical resolution of 30 s and 30 m, respectively. Due to the low attenuation of the radar signals at this wavelength in the cloudy atmosphere, clouds are detected in almost their entire vertical extent depending on the radar sensitivity. Only in situations with strong precipitation is the attenuation higher and thus the cloud detection capability reduced.

In the first step, a target classication including the determination of cloud base and top is performed from the radar profiles of reflectivity, Doppler velocity, and ceilometer backscatter profiles, as well as temperature and humidity profiles provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) model. The liquid-water content (LWC) profile is estimated by calculating the theoretical adiabatic LWC gradient for each liquid-water cloud layer.

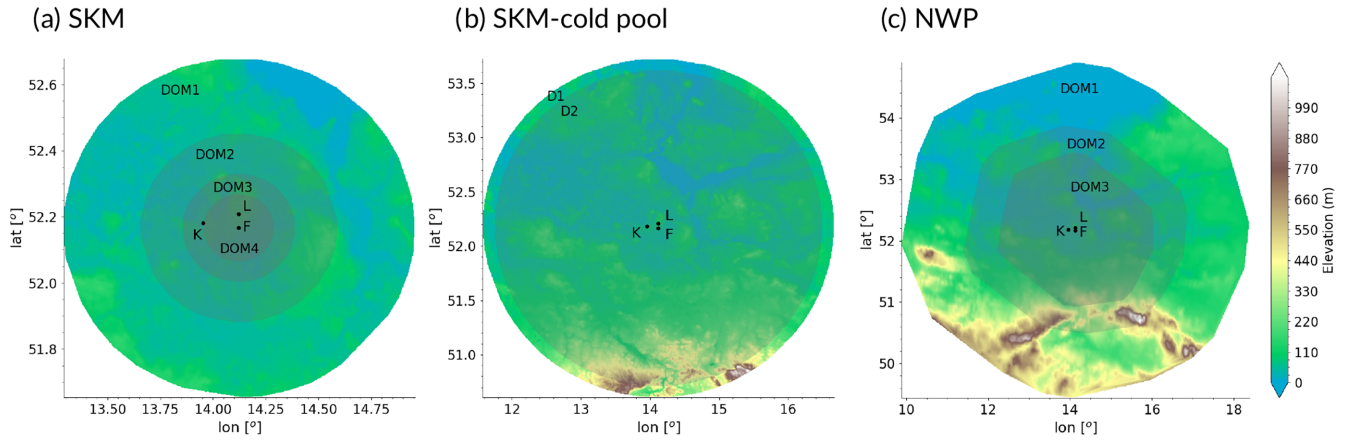

FIGURE 2 (a) ICON-SKM, (b) larger ICON-SKM-cold pool, and (c) ICON-NWP modeling domains overlaid over the orography map representing terrain height in meters, source ASTER Global Digital Elevation Model V003 (2019). Grid spacings are given in Table 2. The three supersite locations are denoted by the first letters of their names: L—Lindenberg, F—Falkenberg, and K—Kehrigk. [Colour figure can be viewed at wileyonlinelibrary.com]

The integrated LWC values are finally scaled to the MWR measured liquid water path. It should also be noted that, due to the limited sensitivity of the cloud radar, very thin clouds (e.g., with very small cloud droplets or ice crystals) may not be detected completely.

3 | DESCRIPTION OF THE MODELS

3.1 | ICON

The ICON model solves the fully compressible non-hydrostatic atmospheric equations of motion following Gassmann and Herzog (2008). For an extensive description of the model equations and the model configurations used in NWP (ICON-NWP) and current operational configurations, we refer to Zängl *et al.* (2015) and Reinert *et al.* (2025), respectively. The equations are discretized on an icosahedral-triangular Arakawa-C grid with the mass points located in the cell centers, and the horizontal velocity component normal to the triangle edges defined at the edge midpoints (Zängl *et al.*, 2015). The vertical discretization is formulated using a height-based terrain-following coordinate, the smooth level vertical (SLEVE) coordinate (Leuenberger *et al.*, 2010; Schär *et al.*, 2002). The model solves for the prognostic variables, including the horizontal velocity component normal to the triangle edges v_n, the vertical wind component w, the density ρ, the virtual potential temperature θ_v, and the specific masses and number densities of tracers $q_i, i = 1, 2, 3, \ldots, N$. Depending on the choice of microphysics scheme, the tracers can include water vapor q_v, liquid water q_l, snow q_s, ice q_i, and other hydrometeors or trace substances. The vertical velocity w is defined at half levels and other prognostic variables at full levels, following Lorenz (1960). Localized grid refinement is possible by successive division of the cells of a spherical icosahedron and a nesting approach (Wan *et al.*, 2013). ICON can also be run on a predefined set of grids in a limited-area mode (LAM) by specifying the lateral boundary conditions from coarser-resolution runs or reanalysis datasets. In addition to its NWP and climate applications (Giorgetta *et al.*, 2018), ICON can be run in an LES mode (Dipankar *et al.*, 2015) and a single-column mode (Bašták Ďurán *et al.*, 2021).

The ICON modeling setup for FESSTVaL covers grid resolutions from about 5 km to 78 m, presented in Figure 2 and Table 2. We distinguish between two different setups: ICON sub-kilometer (ICON-SKM) and ICON-NWP. ICON-SKM is used in a limited-area mode with a two-way nesting configuration, including four nested domains with a sub-km grid spacing, as shown in Figure 2a. The domains used in NWP are set in a similar way, centered at Falkenberg, encompassing the finer-resolution grids and expanding the domain size as the resolution coarsens (Figure 2c). The simulations at the coarsest-resolution grid of the sub-kilometer (SKM) setup (DOM1, 626 m) are forced by the ICON operational forecasts at a horizontal grid resolution of about 2.2 km, while three SKM domains are nested with two-way communication between the parent and child grids, with sequentially decreasing grid spacings of 313, 156, and 78 m, respectively. The innermost domain size (DOM4) is about 22 km in the north–south direction and 24 km in the east–west direction. The number of vertical levels in all nested grids is 97, with the smallest vertical distance near the surface of about 20 m. The ICON-SKM grids used in the cold-pool case are presented in Figure 2b and include two nests of

TABLE 2 Setup of the modeling grids for ICON-SKM, MicroHH, and ICON-NWP in FESSTVaL.

Domain abbreviation	Grid spacing (m)	Domain size/diameter (approx. km)
ICON-SKM		
DOM1	626	114
DOM2	313	63
DOM3	156	36
DOM4	78	23
ICON-SKM—cold pool		
D1	525	348
D2	260	324
MicroHH-LES		
D1	75	19.2
D2	37.5	19.2
ICON-NWP		
DOM1	5000	610
DOM2	2500	382
DOM3	1200	272

525- and 260-m resolution and relatively large domains that are comparable in size to the ICON-NWP domain DOM2. This study focuses primarily on the SKM analysis of DOM3 and DOM4 with grid spacings of 156 and 78 m, respectively, and the NWP domain DOM2 with a grid spacing of about 2.5 km. The latter setup is similar to the operational ICON-D2 configuration (Reinert *et al.*, 2025), but instead of the D2 domain covering Germany, we focus on a smaller region in Eastern Germany centered at Falkenberg. For reference, the innermost ICON-SKM domain, DOM4, is covered mostly by croplands (43.5%) and forest (33.7%), followed by shrubland and herbaceous vegetation (11.5%) and water bodies (4.05%). In DOM3, the coverage by croplands is slightly lower (32.5%) compared with that by forests (42.%), while shrubland and herbaceous vegetation take about 13.1% and water bodies about 4.7%.

The ICON model is configured using the operational NWP physics schemes. The ABL turbulence is parameterized by the 1D TKE turbulence scheme of Raschendorfer (2001, 2011) and Raschendorfer *et al.* (2003), including additional horizontal shear terms contributing to TKE production, additional scale-interaction terms originating from non-turbulent subgrid flows, a turbulent length-scale dependent on the model resolution, and a subgrid-scale cloud cover scheme (Raschendorfer, 2016). ICON can also be configured in a LES mode (Dipankar *et al.*, 2015), which employs the 3D turbulence scheme of Smagorinsky–Lilly (Lilly, 1962), the surface-layer scheme based on Louis (1979) and no subgrid cloud scheme. We have performed our simulations using both turbulence schemes in ICON-SKM and investigated the benefits of using the Smagorinsky-type closure for the representation of turbulent quantities across the SKM resolutions. The difference in the performance of the two configurations is minimal, with a slight benefit from the NWP setup due to a better representation of the contrast in the surface turbulent heat fluxes between the areas covered by forests, grasslands, or crops (not shown here; see also Goger & Dipankar, 2024). As one of the main goals of our present study is to assess the gains of increasing the resolution of the operational ICON to sub-km scales, we focus the analysis on the ICON-NWP configuration across sub-km resolutions. For either configuration option at sub-km resolutions, the parameterizations of convection, subgrid-scale orographic effects, and gravity-wave drag are switched off (for a detailed description see Dipankar *et al.*, 2015). For the land-surface component, we use the TERRA land-surface model (Schulz & Vogel, 2020) with the option for subgrid land-surface variability turned off, due to the high horizontal resolution. The SKM setup includes the static land-surface data as used in the operational ICON-NWP, with addition of the ASTER orography data with an original resolution of 1 arcsecond (approximately 30 m) (Figure 2, NASA/METI/AIST/-Japan Spacesystems ans U.S./Japan ASTER Science Team, 2019) and the Corine land-cover data at 100-m resolution (CORINE Land Cover 2018, raster 100 m, 2020).

3.2 | MicroHH

MicroHH solves the anelastic governing equations on an Arakawa-C grid. The variables being solved with prognostic equations are the three velocity components u, v, w, the liquid water potential temperature θ_l, and the specific humidity q_t, which is the sum of water vapor q_v, liquid water q_l, and ice q_i. Division of total water between these three components is done via saturation adjustment. A full description of the code is found in van Heerwaarden *et al.* (2017); hereafter, we describe the settings relevant to the FESSTVaL simulations.

In this study, MicroHH is run at 75 and 38 m horizontal grid spacing (Table 2) and using a stretched grid in the vertical direction with doubly periodic boundary conditions in the lateral directions, no-slip and no-penetration at the bottom, and free-slip and no-penetration at the top. To force the model with the appropriate weather conditions, the Large-eddy simulation and Single-column model-Large-Scale Dynamics ((LS)^{2}D) package (van Stratum *et al.*, 2023) has been used, in order to have initial

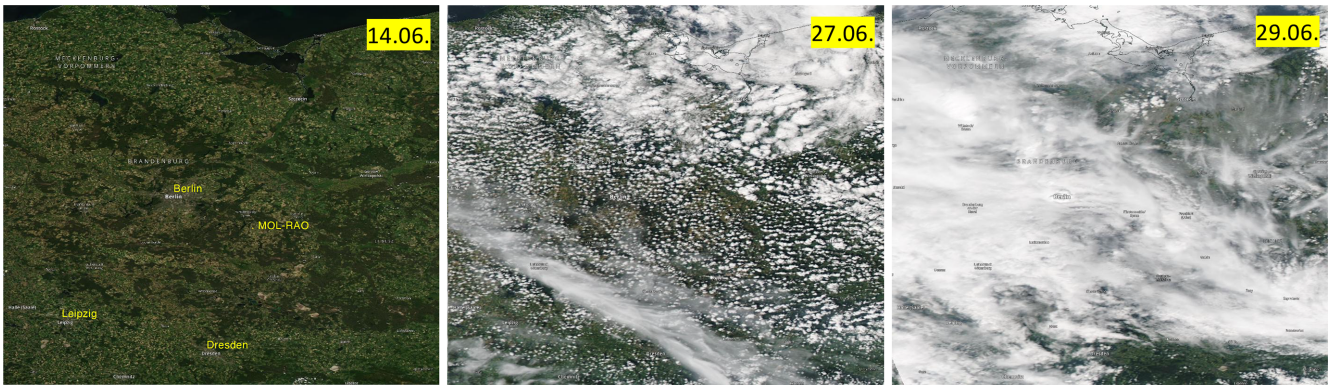

FIGURE 3 Cloud conditions on selected days during the IOP of FESSTVaL in June 2021. The images are captured by the *Aqua* MODIS instrument with overpass time at about 1200, 1130, and 1120 UTC, respectively [Satellite Image from https://worldview.earthdata.nasa.gov]. [Colour figure can be viewed at wileyonlinelibrary.com]

conditions and large-scale advection terms consistent with ERA5 reanalysis (Hersbach *et al.*, 2020). ERA5 is used as the most optimal forcing data for MicroHH previously validated in multiple experiments (van Stratum *et al.*, 2023) and to keep the LES independent of the ICON modeling framework.

The simulations are run with a flat and spatially homogeneous land surface representative of the Falkenberg site. MicroHH has an interactive land-surface model based on Balsamo *et al.* (2009), solar and thermal radiation fluxes computed using RTE+RRTMGP (Pincus *et al.*, 2019), and microphysics computed using a two-moment scheme (Seifert & Beheng, 2001; Stevens & Seifert, 2008). One additional 38-m grid spacing simulation has been performed for June 27, 2021 using a ray tracer (Veerman *et al.*, 2022) for shortwave radiation, for comparison with grid observations taken during the campaign (Mol *et al.*, 2024).

4 | METEOROLOGICAL OVERVIEW OF THREE CASE STUDIES

Based on the FESSTVaL field campaign, we simulate three example case studies representing typical ABL conditions: a predominantly clear-sky day (June 14, 2021), a shallow convective day (June 27, 2021), and a precipitating deep convective day representative of cold-pool dynamics (June 29, 2021, Figure 3). These three days are selected because they are covered by extensive boundary-layer and surface radiation measurements during the FESSTVaL campaign. In this study we use a subset of these extensive observations (see next section).

In brief, the three days can be characterized as follows.

On June 14, the study area was under the influence of a high-pressure system over Central Europe, moving slowly eastward. Boundary-layer winds were weak (2–4 $m \cdot s^{-1}$) from south to southwest. A weak low-level jet (axis wind speed 7–8 $m \cdot s^{-1}$) formed at the end of the day. A few isolated shallow cumulus clouds developed after 0630 UTC in the morning, but disappeared until 0930 UTC, such that the maximum incoming shortwave radiation reached values around 900 $W \cdot m^{-2}$. Near-surface temperature showed a pronounced diurnal cycle, with minimum values of 7 °C in the morning and maximum values of 25°C in the afternoon. The temperature difference between 2 and 98 m exceeded 7 °C at around 0500 UTC, indicating strongly stable stratification; however, this surface inversion was broken by strong surface heating within about one hour. The convectively mixed ABL reached a height of about 1700 m in the afternoon. This day is thus representative of midlatitude anticyclonic clear-sky conditions over relatively dry soils.

On June 27, the FESSTVaL area was under the influence of an extended high-pressure zone over Central Europe again. Pressure gradients were very weak, resulting in weak winds (1–3 $m \cdot s^{-1}$) from varying directions during daytime. At night, winds from northeast to east reached a maximum of 4–6 $m \cdot s^{-1}$ at heights between 100 and 250 m. The formation of shallow fair-weather cumulus clouds started at around 0830 UTC. These clouds moved very slowly, due to the weak winds, and disappeared in the evening at around 1800 UTC. Overall, more than 12 hours of sunshine were recorded, but the cumulus clouds lead to frequent abrupt changes of incoming shortwave radiation between values around 1000 and 250 $W \cdot m^{-2}$, respectively. Near-surface temperature increased from 13 °C in the morning to 27°C in the afternoon. The convective ABL grew to maximum of heights around 2100 m in the afternoon. This day is thus representative of undisturbed midlatitude anticyclonic conditions over relatively dry soils that favored shallow convective cloud development.

On June 29, the study region was situated at the eastern flank of an upper-air trough. At the surface, a

low-pressure system was situated over Western Europe, but the overall pressure gradients were weak. A convergence line over East Germany moved slowly northward in the course of the day; thunderstorms (first isolated, later clustering) developed in the warm and humid air (maximum temperatures around 30°C, specific humidity 10 g · kg^{-1}) and moved northwestward along that line. Thunderstorm activity reached the FESSTVaL region at around 1330 UTC and led to heavy rain in the northeast part of the area. At GM Falkenberg, no precipitation was measured at all, but a significant cold pool hit the site at 1410 UTC (temperature drop of about 6 K, wind gusts up to 18 m · s^{-1}). At the other supersites (Birkholz, Lindenberg) that were hit by the thunderstorm, the cold pool was associated with precipitation and the temperature drop (up to 10 K) and gust wind speeds (up to 23 m · s^{-1}) were even stronger. Before the arrival of the thunderstorms, winds were weak (2–4 m · s^{-1}) from the southeast; they turned towards east–northeast with speeds of 3–6 m · s^{-1} after the passage of the cold pool. A nocturnal jet with axis wind speeds of about 11 m · s^{-1} at 200–300-m height was present in the morning before the convective ABL developed.

5 | MODEL EVALUATION BASED ON CASE STUDIES

5.1 | Diurnal changes in the convective ABL across three case-study days

As the first evaluation of the ICON-SKM performance, we examine the diurnal cycle of thermodynamic properties and wind speed in the ABL. We focus on the diurnal changes of the domain average temperature, relative humidity, and wind speed, the vertical profiles of potential temperature and absolute humidity, and the spatial standard deviation of near-surface temperature.

5.1.1 | Domain average

Figure 4 shows the diurnal changes in the temperature and relative humidity at 2 m above the surface and the wind speed at 10-m height. To give a better impression of the representativeness of the comparison, observations from the three supersites are shown as black curves, while for ICON we consider grid points falling within a rectangular area

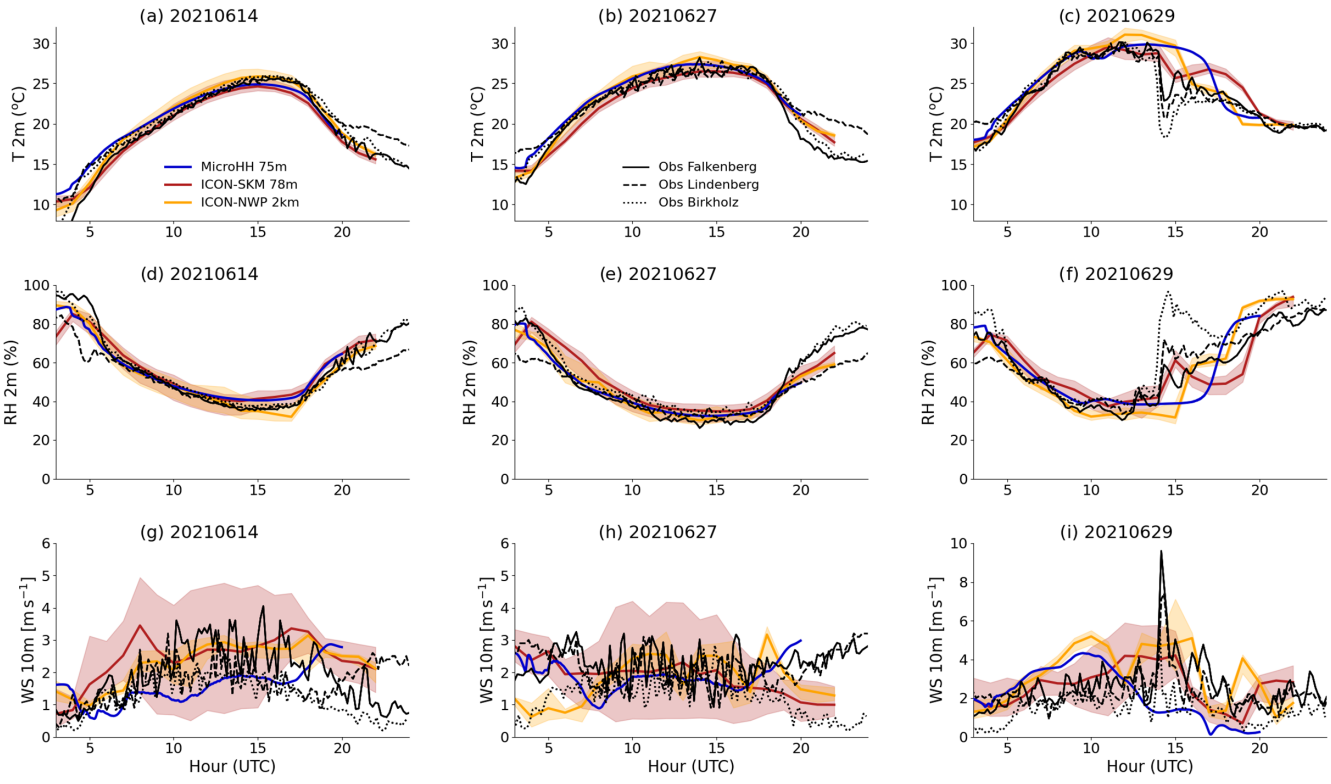

FIGURE 4 Diurnal changes of the 2-m temperature (T 2 m), relative humidity (RH 2 m), and the wind speed measured at the height of 10 m (WS 10 m). The modeled fields are averaged over the area surrounding the three observational sites (Falkenberg, Lindenberg, and Birkholz) from 52.16°–52.22°N and from 14.11°–14.2°E in ICON and over the LES domain in MicroHH. The median is shown as a solid curve, while the range from the fifth to the 95th percentile is shaded and provided for ICON runs. The values provided by Micro-HH are treated as a single-column output due to its homogeneous forcing. [Colour figure can be viewed at wileyonlinelibrary.com]

containing all three supersites (latitude 52.16°–52.22°, longitude 14.11°–14.2°). The median of each model's distribution is shown as a solid curve, while the range from the fifth to the 95th percentile is shaded. The region contains 6675 points for the ICON-SKM at 78-m resolution, but only eight grid points for the NWP configuration. Observations from the three sites are very similar during the daytime, but diverge during the night, with warmer and drier conditions observed at the Lindenberg site. Overall, all models simulate the diurnal cycle of temperature and humidity very well in all three cases, with the observed range of values falling within the model spread for the region considered during the daytime. However, the model spread at night remains low, no longer matching the range of values observed at the three sites. This indicates that land-surface and boundary-layer processes are represented well during daytime, while the model faces difficulties in representing non-convective (stable) ABLs and near-surface atmospheric flows. ICON-NWP produces a slightly higher average temperature and lower relative humidity compared with ICON-SKM, while ICON-SKM approaches the reference LES closely, especially on the clear-sky day. The abrupt changes in temperature and relative humidity with the passage of a cold-pool front on the afternoon of June 29 are imposed by the boundary conditions and are captured well at sub-km resolutions. ICON-NWP delays the temperature drop and relative humidity increase by approximately one hour compared with ICON-SKM and the measurements. Note that the forcing for ICON-NWP comes from ICON-EU, while the ICON-SKM forcing is derived from ICON-D2, that is, differences in forcing likely contribute to the time shift. The observations show strong temperature oscillations after the cold-pool passage, which none of the simulations reproduces to this degree. The NWP simulation with the strongest simulated cold pools does show a weak gravity-wave signature in the surface pressure field (not shown), but this does not translate into discernible temperature oscillations. In MicroHH, the development of the cold pool occurs later in the afternoon, which is visible in the abrupt change of the 2-m temperature and relative humidity at about 1700 UTC, closely following its ERA5-based boundary conditions. The idealized LES in the latter case is not an adequate reference for ICON-SKM runs. We discuss the reasons for this further and provide more details about the cold-pool development on June 29 in different models and model configurations in Section 5.4.

Wind speed is more challenging to model and compare with the observations, due to its strong sensitivity to landscape characteristics and quick response to changes in the forcing. In the lowest panel of Figure 4, we compare the model output with the measurements at the three supersites. The average wind speed generally falls into the range of the observed values. Generally speaking, the 78-m SKM simulation tends to overestimate the wind variability, while the NWP model underestimates it. On June 27, ICON's modeled wind speed drops below the observed values during morning and evening hours. The observations show evidence of a low-level jet in the early morning hours. ICON-NWP underestimates the jet's wind speed and overestimates the height of the jet's nose (not shown here), which likely also explains the wind underestimate at 10 m. This is a common problem for NWP, where enhanced mixing in stable ABLs compensates for near-surface biases in temperature and humidity but degrades the evolution of low-level jets (e.g., Sandu *et al.*, 2013). The situation in the evening is less clear. Observations show higher wind speeds (up to 6 $m \cdot s^{-1}$) throughout the lowest 600 m of the ABL compared with ICON. In the reference LES, there is an increase in low-level wind speed at the end of the day. The model has a small nudging term that moves the model to slightly higher wind speeds, as local turbulence has stopped being a dominant force after the day has ended. On June 29, the wind gusts related to the cold-pool fronts are represented slightly better in ICON-NWP. However, the peak wind speed is delayed by approximately one hour following the delayed onset of deep convection and cold pools, consistent with the plots of temperature at 2-m height.

In Figure 5, the vertical profiles of potential temperature and absolute humidity are plotted for the times of routine soundings at the Lindenberg site. Here, we compare instantaneous values taken from the model output and averaged over an area around the observational site with the sounding values collected during the time of vertical ascent of the balloons. The vertical structure of the ABL and its growth during the day are, overall, represented well by the models for all three case studies (Figure 5). The midday boundary layer is well mixed and the changes at the top of the mixed layer are captured. ICON-SKM shows potential temperature values lower by about 1 K in the afternoon on shallow convective days, while in ICON-NWP the vertical structure of potential temperature follows the observed values closely. Diurnal changes in absolute humidity are less pronounced compared with those in temperature. This is especially visible on June 27, where the vertical profiles fall together, indicating a well-mixed and almost invariant absolute humidity in the boundary layer.

5.1.2 | Variability across the domain

The dense network of surface weather stations distributed over the FESSTVaL domain allows one to quantify the sub-mesoscale variability of the near-surface temperature field. Figure 6 shows the spatial standard deviation

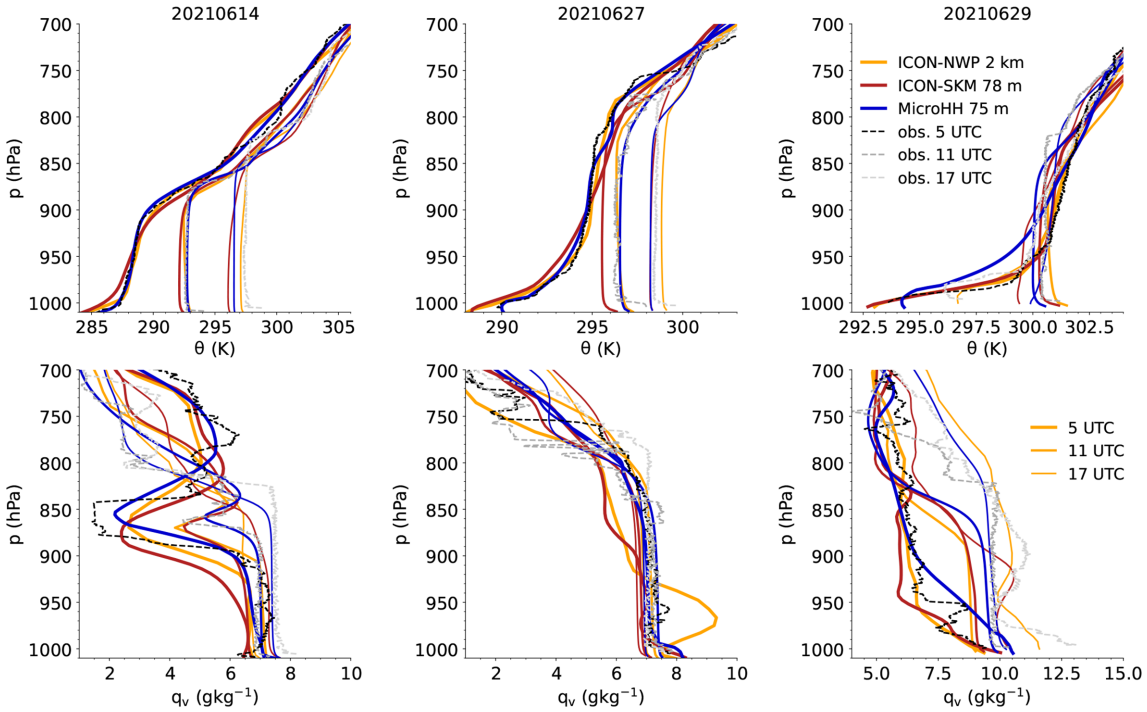

FIGURE 5 Vertical profiles of temperature and absolute humidity in the boundary layer are taken at the three time instances when operational soundings are conducted, at 0500, 1100, and 1700 UTC. The averaging area of the modeled values corresponds to the area used in Figure 4. [Colour figure can be viewed at wileyonlinelibrary.com]

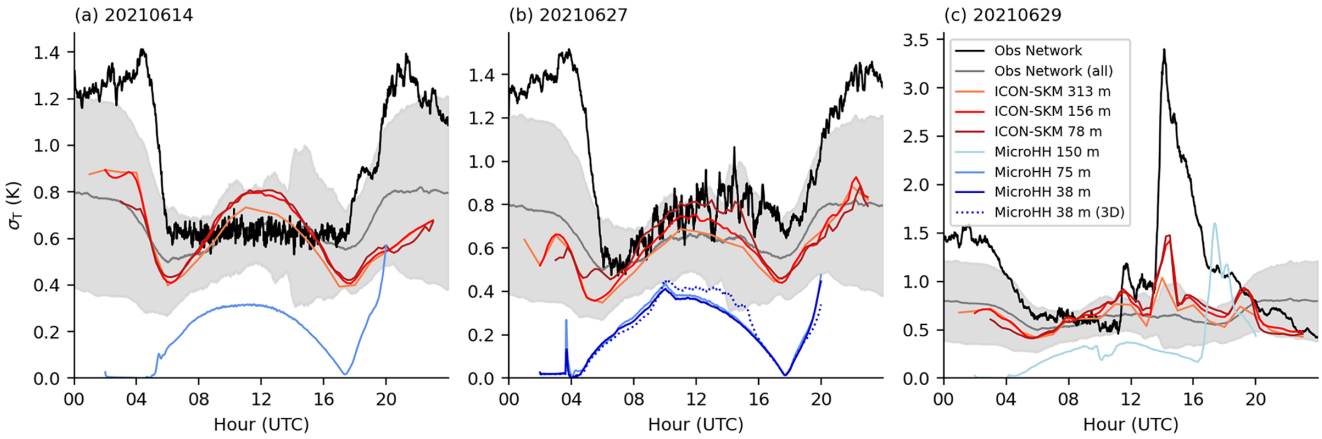

FIGURE 6 Diurnal cycles of spatial temperature standard deviation σ_T on (a) June 14, (b) June 27, and (c) June 29, 2021 observed by the network of automated weather stations, as well as for ICON-SKM and MicroHH simulations at different resolutions. The gray line and shading in each panel show the observed mean and standard deviation of σ_T for the full FESSTVaL period (May 17–August 27, 2021). [Colour figure can be viewed at wileyonlinelibrary.com]

of temperature σ_T calculated for the 99 measurement stations, as well as for the ICON-SKM and MicroHH simulations. The observational data with temporal resolutions of 1 s (APOLLO stations) and 10 s (WXT) are averaged at 1-min intervals to reduce the impact of instrumental noise. The simulation datasets are analyzed at the temporal resolution of the model outputs of 15 min (ICON-SKM) and 1 min (MicroHH), while all model grid points within 15 km around the network center (GM Falkenberg) are used.

On average over the full campaign period, the observed σ_T exhibits a w-shaped diurnal cycle with a broad nocturnal maximum between 2000 and 0400 UTC and a secondary afternoon maximum peaking around 1500 UTC, while minima are present at around 0600 and 1700 UTC. We hypothesize that the temporal evolution of σ_T is indicative of distinct processes over the course of the day. While convectively driven circulations and differential heating by variable cloud cover increase the heterogeneity of the near-surface temperature during the day, local radiative

and topographic effects in the stable boundary layer dominate the spatial differences at night. The same qualitative signal is observed for the individual days of June 14 and 27 (Figure 6a,b), although on June 14 σ_T stays nearly constant after the morning transition at 0600 UTC, likely the result of a well-mixed and shallower boundary layer with weaker convective updrafts compared with June 27 and the absence of clouds during the day.

The ICON simulations are capable of reproducing the w-shaped diurnal cycle of σ_T, although they underestimate its magnitude during non-convective hours. This indicates that a grid spacing of 78 m is still too coarse to represent the microscale temperature variability that impacts point measurements during the non-convective hours and points to the missing subgrid-scale variability not represented by subgrid-scale parameterizations. On June 14, ICON simulates a pronounced diurnal cycle of σ_T during daytime, which is not present in the observations. A possible reason is an effect of resolved large convective eddies on the thermodynamics in the lowest model layers on this day. The temperature variance increases for finer grid spacing. This results in a better agreement of the simulation with the observational data in the example of June 27; however, only for the day-time regime, while the impact of model resolution is negligible or even opposite at night. The impact of model resolution on the daytime σ_T can be explained by the fact that the size of explicitly resolved large eddies depends on model resolution: a finer grid spacing allows for the explicit modeling of smaller-scale eddies, which introduces more and finer spatial variability. During the night, the LES models fail to resolve the spatial variability of residual and stable boundary layers. Another reason for the underestimation is the unresolved spatial variability in land-surface properties, such as fine-scale patterns in topography and in the soil and vegetation characteristics, which only becomes dominant in the absence of daytime convection. On June 29, the impact of a cold pool causes an exceptional peak in σ_T (eight standard deviations above average) and a considerable deviation from the mean diurnal cycle, which is also underestimated significantly in the ICON simulations. A detailed discussion of this case is presented in Grant *et al.* (2024).

The idealized MicroHH simulations exhibit a diurnal cycle of σ_T similar to ICON-SKM, although at a significantly lower magnitude, as expected due to its idealization of the land-surface properties and cyclic boundary conditions. Most strikingly, MicroHH reproduces the timing and sharp increase in σ_T of the evening transition on June 14 and 27 correctly. However, the enhanced variability in the morning is not modeled, most likely because it falls into the model spin-up period. For the cold-pool case on June 29, σ_T reaches a peak similar to that in the ICON-SKM simulations, despite the considerably lower values before. In contrast to ICON-SKM, the change in the grid spacing from 75 to 38 m barely impacts the magnitude of σ_T, as demonstrated for June 27. This signifies that most of the spatial variability is already resolved at the grid spacing of 75 m, while the contribution from subgrid scales is missing, as in ICON-SKM. The fast convergence with resolution of MicroHH can be attributed to low interpolation errors of the sixth-order advection schemes.

5.2 | Case study June 14: turbulent exchange and mixing

5.2.1 | Surface energy balance

In situ measurements provide the turbulent surface fluxes of sensible and latent heat representative of the land-surface conditions within the footprints of the energy-balance stations, of which one is located above a grassland surface surrounded by agricultural crops (Falkenberg) and another in a forest (Kehrigk). Area-averaged turbulent heat fluxes are estimated using an OMS that operates along a path of about 5 km length between the two sites at Lindenberg and Falkenberg (Beyrich & Mengelkamp, 2006). The modeled heat flux is from the same region containing the three supersites as described in Section 5.1.1, and the colored curves and shading show the median and fifth–95th percentile ranges of surface fluxes within this region. Therefore, the three different methods that estimate the surface heat flux are not directly comparable; nevertheless, showing different observations and model output together provides a rich basis to improve our understanding of the advantages and limitations of each method.

The measured sensible heat flux is lower at the Falkenberg site above short grass, while the latent heat flux is higher than the flux measured at the Kehrigk site in the middle of a forest (Figure 7a). The sensible heat flux estimated using the optical LAS is significantly higher compared with the sensible heat flux measured using the eddy covariance (EC) method at the Falkenberg site, although an underestimation of the midday flux values in the LAS estimates is still possible due to signal saturation (Kohsiek *et al.*, 2006). The latent heat flux is lower over the forest than over the grassland, while the two measurement methods (EC and LAS/MWR) at Falkenberg show closer agreement compared with the sensible heat flux. The sensitivity of the surface flux to the variability in the surface properties, such as soil moisture, land cover, vegetation type, and convective circulations across different footprint areas of the methods used, is largely responsible for the discrepancy between the modeled and observed heat fluxes.

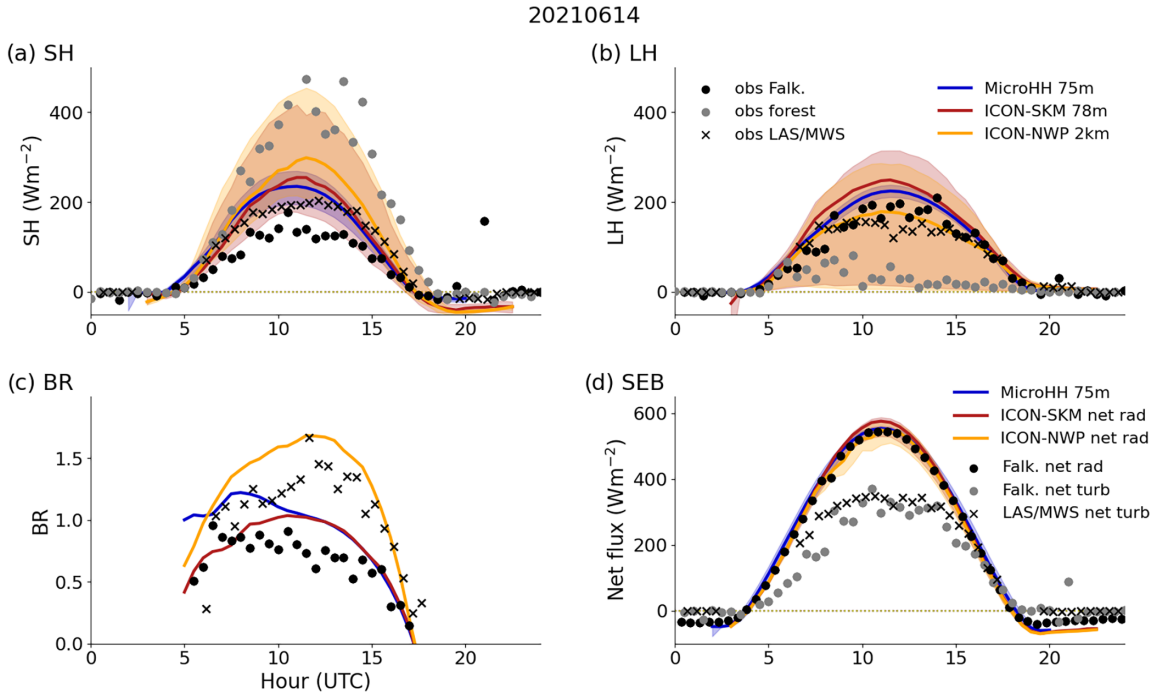

FIGURE 7 Surface turbulent flux of sensible (SH) and latent heat (LH), surface Bowen ratio (BR) and surface energy balance (SEB) on June 14: comparison of the surface turbulent fluxes of (a) SH and (b) LH measured by energy balance stations at Falkenberg and Kehrigk, estimated using the LAS/MWR operating between Lindenberg and Falkenberg sites, and modeled by ICON and MicroHH; (c) modeled surface BR compared with the BR estimated from observations at Falkenberg; and (d) comparison of the net surface radiation flux and the sum of the surface turbulent heat fluxes, indicating the gap in the SEB closure. [Colour figure can be viewed at wileyonlinelibrary.com]

Nevertheless, the range of modeled values is similar to the range of the three flux measurements, even for the NWP configuration represented by only eight grid points in this figure. This appears reasonable, as a mixture of farmland and forest are the dominant land-use classes in the model area. Unsurprisingly, the classical LES setup used for the MicroHH simulation, lacking heterogeneous surface conditions, produces a much reduced spread of the surface fluxes. ICON-NWP shows slightly higher values of sensible heat flux than ICON-SKM and MicroHH. The SKM and LES models agree better on the magnitude of the latent heat flux, while ICON-NWP shows lower values closer to the observed ones at a grassland site (Figure 7b). In general, these modeled latent heat fluxes appear too high, taking into account that the area-representative flux values also represent the lower latent heat fluxes over the forest. However, one has to remark that the measured turbulent energy fluxes do not close the energy balance (see below).

The difference between the flux measurements obtained from the EC station and the area-averaged heat fluxes estimated from the LAS/MWR system is demonstrated further by the changes in the surface Bowen Ratio (BR) during daytime (Figure 7c). The BR estimated by the LAS/MWR system has higher values that peak around noon, while the BR estimated from the EC measurements peaks in the morning and decreases during the day. As expected, the area-average BR modeled by ICON follows the LAS/MWR estimates closely. Interestingly, the BR estimated from the area-averaged fluxes modeled by the MicroHH model follows the diurnal evolution of the EC measurements, most likely because of its spatial homogeneity. The match in BR between SKM and LES after the morning hours is remarkable.

A comparison between the modeled and observed components of the surface energy budget is plotted in Figure 7d. The net radiation flux at the surface is calculated as a balance between incoming and outgoing shortwave and longwave radiation fluxes reduced by the ground heat flux. The net radiation flux modeled by ICON-SKM overestimates the observed net radiation flux at Falkenberg on June 14 slightly. The simulations carried out using the MicroHH and ICON-NWP models match the observed total radiation flux closely. The difference between the SKM and NWP simulations can be explained by a small amount of shallow cloud present in the ICON-NWP simulations on this day. The sum of the observed surface turbulent heat fluxes, compared with the net radiation flux, shows a closure gap in the observed energy balance of more than 100 W·m^{-2} during the midday hours. This gap is reduced slightly in the flux estimated using the scintillometry approach, which can represent the area-averaged turbulent heat fluxes. The review

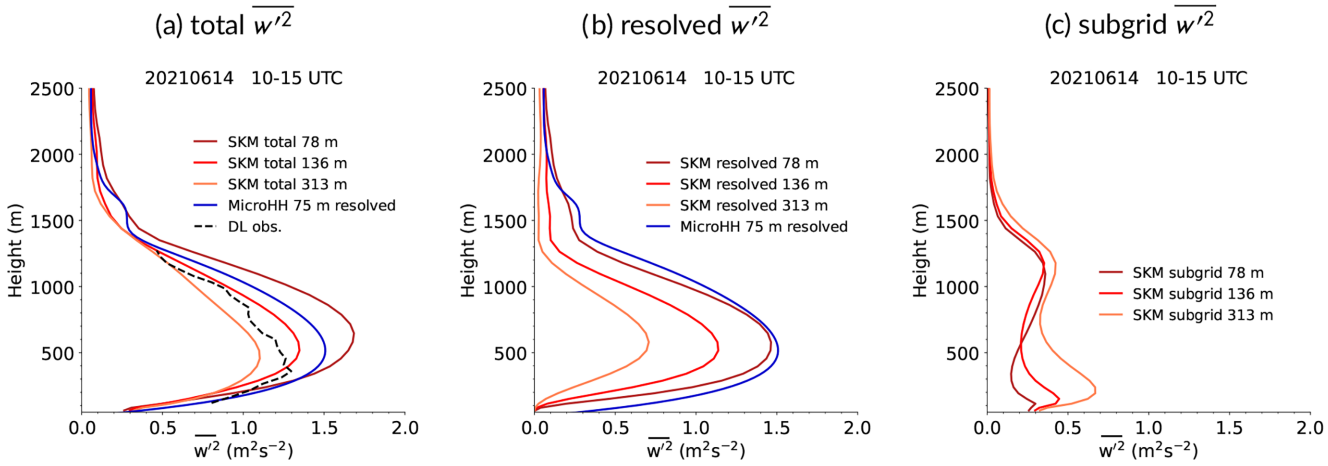

FIGURE 8 Comparison of the variance of vertical velocity averaged over the time window from 1000 to 1500 UTC modeled by ICON-SKM and MicroHH with the variance retrieved from Doppler lidar. The profiles retrieved from Doppler lidar are half-hourly estimates. The SKM variance is the estimate of the spatial variance taken every 15 minutes, averaged between 1000 and 1500 UTC. The area of consideration is equal for the three SKM domains, which results in a different number of points as the resolution increases. The area is the same as the one taken for the analysis in Figure 4. [Colour figure can be viewed at wileyonlinelibrary.com]

article of Mauder *et al.* (2020) narrows the possible reasons for the lack of closure in the surface energy balance to two major contributions in addition to heterogeneity in land-surface properties: secondary convective circulations that are not captured by conventional measurement methods and biomass heat storage. In ICON, the closure of the surface energy balance is ensured by the model formulation, where the ground heat flux is calculated as a residual of the surface energy balance (Kracher *et al.*, 2009). The total surface heat flux, as modeled by ICON, is thus not plotted.

5.2.2 | Turbulence measures

As a measure of variability and intensity of turbulence in the ABL, the variance of vertical velocity, $\overline{w'^2}$, as modeled in SKM, is compared with the variance resulting from the MicroHH-LES and the FESSTVaL measurements derived from a Doppler lidar in vertical stare mode (Table 1, Figure 8). We estimate the spatial variance from the instantaneous model output at 15-minute intervals over the five-hour period, while the observed variance is sampled half-hourly from one-minute DL measurements at a single measurement location. To avoid the morning and evening transitions of the ABL regimes, $\overline{w'^2}$ is averaged over the period with a well-developed convective ABL from 1000 to 1500 UTC (as in Dewani *et al.*, 2023). The modeled vertical profiles of $\overline{w'^2}$ are sampled in a rectangular area around the supersites, as in the analysis in previous sections.

The degree to which the models will reproduce the observed $\overline{w'^2}$ depends mainly on the correct estimation of the surface turbulent heat fluxes and the depth of the boundary layer, but also on the resolution of the model. For the day representative of clear-sky conditions, the total $\overline{w'^2}$, consisting of the resolved and subgrid parts, increases with the resolution of ICON-SKM (Figure 8a), similarly to the LES studies of Heinze *et al.* (2017) or Sullivan and Patton (2011). The change in the magnitude of the total variance with model resolution results from the resolved part of the spectrum of turbulent eddy sizes (Figure 8b), which is highly dependent on the model resolution. The effective resolution of ICON is about 7–8 times the horizontal grid spacing (e.g., Heinze *et al.*, 2017), which means that the size of the finest resolved turbulent eddies in our innermost domain starts at about 600 m, leaving eddies of finer scales unresolved. For the next coarse resolution nest, the size of effectively resolved eddies starts at about 900 m, followed by the coarsest SKM nest with eddies of about 2 km size.

Both models show a higher total variance compared with the observed variance sampled half hourly at a single point of observation. However, the general shape of the profiles agrees quite well. In the previous work of Heinze *et al.* (2017, their figure 11a), it was possible to estimate the spread of the observed variance profiles (in their case, the variance normalized by the vertical velocity scale) over multiple spatial points, which resulted in a spread of more than one third of the peak variance in their ICON-LES at a resolution of 156 m. Our ICON-SKM and MicroHH-LES produce variance in the observed range, taking into account that our single-point observations are most likely not entirely representative of the wider spatial area around the observational site (see also Maurer *et al.*, 2016). We

note here that a one-to-one comparison between ICON and DL observations is not representative at a single location, due to the non-exactness of land-surface spatial heterogeneities and model discretization.

The unresolved part of the eddy spectrum is parameterized using the TKE turbulence scheme of Raschendorfer (2001, 2016). The subgrid part of $\overline{w'^2}$ can be estimated as two-thirds of the subgrid-scale TKE, assuming that the turbulence is isotropic here, and is plotted in Figure 8c. The contribution of the subgrid-scale variance is considerable at all model resolutions and can reach from about a third to a fifth of the total variance depending on the model resolution. In the lower part of the mixed layer close to the surface, the contribution of the subgrid-scale variance increases as the model resolution coarsens. This signifies that ICON is capable of resolving more of the small-scale eddies near the land surface as its resolution is refined. The subgrid turbulence scheme only reflects changes in the profiles in the lowest part of the mixed layer, whereas in the mid and upper parts of the mixed layer the profiles of the subgrid variance do not change significantly with model resolution. The dependence of the total variance on the model resolution points to the need to develop further turbulence schemes that are able to adapt to changes in grid spacing better.

In addition to the Doppler lidar in vertical stare mode, a conically scanning Doppler lidar was also deployed during FESSTVaL, with the goal of retrieving TKE (Section 2.2). For a meaningful comparison of simulated TKE with the observations, both subgrid-scale (SGS) and resolved or grid-scale (GS) TKE from the model should be considered, representative of the temporal and spatial scales corresponding to those of the retrieval. Given the Doppler lidar's elevation angle of 35.3° (or zenith angle of 54.7°; see Table 1), the diameter of the scan circle changes with height, between about 280 m at 100-m height and 1700 m at 600 m above ground. The TKE retrieval has a temporal resolution of 30 min, meaning the observed TKE convolves both temporal and spatial wind variances.

Grid-scale TKE is calculated first for ICON-SKM at the highest horizontal resolution (78 m) saved at 5-min intervals. In a first step, the winds are interpolated from the ICON native grid to a regular latitude–longitude grid with 0.001° resolution (i.e., similar mesh size) using the Climate Data Operators (CDO) conservative remapping tool (Schulzweida, 2023). Then winds are averaged horizontally over a neighborhood $(\Delta x)^2$ for each grid point, and averaged temporally over a 30-min window. (Δx) varies vertically according to the Doppler lidar scan diameter up to a height of 678 m (and remains at 1875 m for layers above). Wind perturbations u', v', and w' are then derived relative to these temporally–spatially averaged winds in order to calculate TKE. The sum of this GS TKE and the SGS TKE predicted by the turbulence scheme are then considered for comparison with the observations. Figure 9 (top row) shows the GS, SGS, and total TKE for the clear-sky case alongside the observed TKE. It is evident that both GS and SGS contributions to the TKE are of equal importance at this resolution, and the magnitude of the total TKE is in good agreement with observations. The sonic anemometer mounted on the tower in Falkenberg provides independent observations of TKE (Figure 10, at 90-m height) to validate the DL retrieval (at 95-m height), and is shown alongside the simulated total TKE (solid line, at 78-m height). We see that the two observations agree well, and the temporal evolution and magnitude of the TKE at this height are captured well by the model.

In Figure 9 (bottom row) we present the simulated GS TKE if one considers (a) only the spatial variance of the winds, using vertically varying spatial scales according to the size of the Doppler lidar scan cone, (b) only temporal variance over a 30-min period, and (c) an assumption of a vertically uniform spatial scale of 21 grid points (approximately 1575 m), but ignoring any temporal variance. It is evident that most of the TKE at lower levels comes from the temporal variance, while spatial variance is equally important higher up. Results for the GS TKE are approximated well during the daytime if one chooses an appropriate vertically uniform spatial scale and neglects the temporal variance (see also the dashed curve in Figure 10). This is equivalent to assuming that the "frozen turbulence" hypothesis is valid, which appears to be the case here. It is also convenient to use, as it allows for an approximation of total TKE for individual time steps without needing to archive winds at high temporal resolution.

5.3 | Case study June 27: surface radiation and clouds

On June 27, 2021, shallow cumulus clouds evolved during the day. Observations from cloud radar, microwave radiometer (MWR), and ceilometer provide an opportunity to characterize the observed clouds and their impact on surface radiation. For better presentability of the inter-model comparison, we compare area averages, even though single gridpoint output might match the characteristics of the profiling observations a bit better. Due to the heterogeneity of the ICON surface and the widely varying domain sizes for different configurations, we average over the same rectangular area (latitude 52.16°–52.22°, longitude 14.11°–14.2°) containing the three supersites for

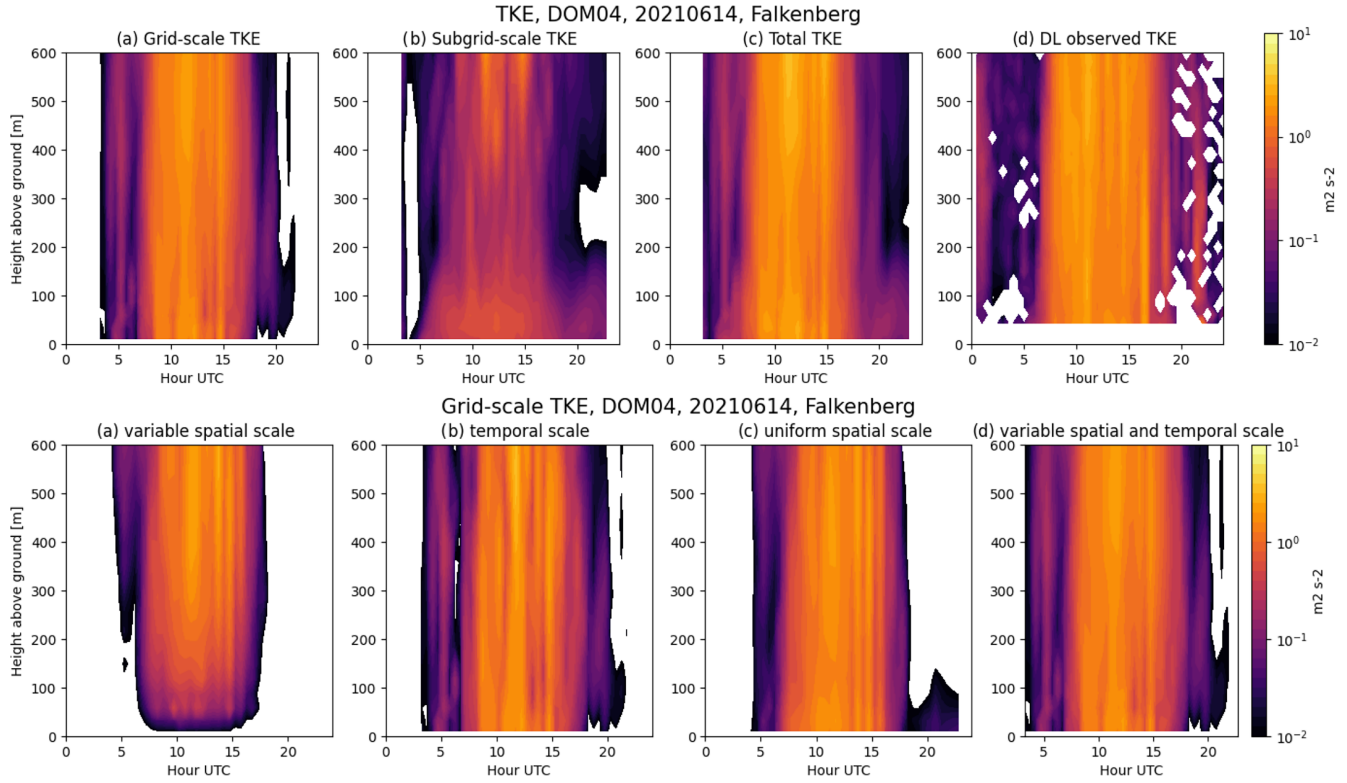

FIGURE 9 Top row: simulated (a) grid-scale, (b) subgrid-scale, and (c) total TKE, alongside (d) Doppler lidar observed TKE for the clear-sky day June 14, 2021. Bottom row: simulated grid-scale TKE for June 14, 2021 calculated using various assumptions for spatial and temporal averaging: (a) vertically varying spatial scale equivalent to Doppler lidar scan diameter, instantaneous wind fields, (b) temporal variance across 30 min only, no spatial variance, (c) vertically invariant spatial scale, instantaneous wind fields, and (d) vertically varying spatial scale and temporal variance across 30 min. [Colour figure can be viewed at wileyonlinelibrary.com]

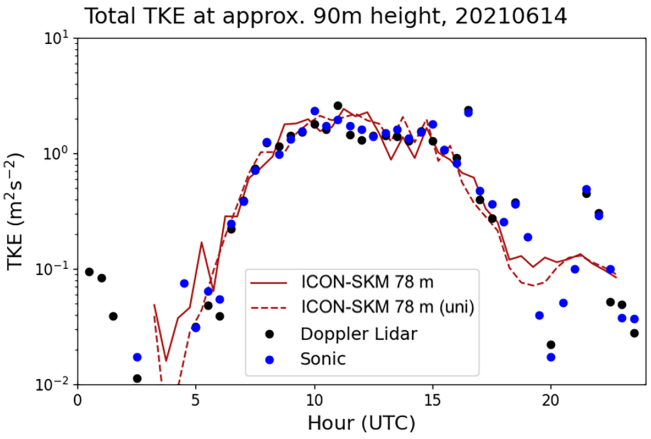

FIGURE 10 Diurnal cycle of TKE on June 14, 2021 at approximately 90-m height from observations and ICON-SKM. [Colour figure can be viewed at wileyonlinelibrary.com]

all ICON simulations (marked in Figure 11e). This comprises 6675 grid points at 78-m resolution, but only nine points for the NWP simulation at 2 km. Since MicroHH has periodic lateral and uniform lower boundary conditions, the full domain average is used for MicroHH.

To give an impression of the cloud scenes on this day, Figure 11 shows a photograph from the Falkenberg Cloudcam at 1100 UTC, together with maps of the simulated liquid water path (LWP) from various model configurations. The rectangular area in Figure 11e shows the averaging region with the three supersite locations mentioned above.

Figure 12 shows time–height cross-sections of liquid condensate retrieved with the Cloudnet algorithm, alongside estimates from five of the model simulations. The onset and decay of Cu clouds are in good agreement between the models and the observations. The clouds tend to grow deeper in the MicroHH simulation compared with the ICON-SKM simulation at similar grid spacing, and deeper also when compared with the observations. For the MicroHH setup, the increase in resolution changes little, but 3D radiation leads to greater condensate amount and deeper clouds. The ICON-NWP configuration produces clouds earlier and later in the day compared with ICON-SKM. This is unlikely to be related to the shallow-convection parameterization used in this configuration, but rather the early morning clouds are non-convective and most likely occur due to the ICON-EU

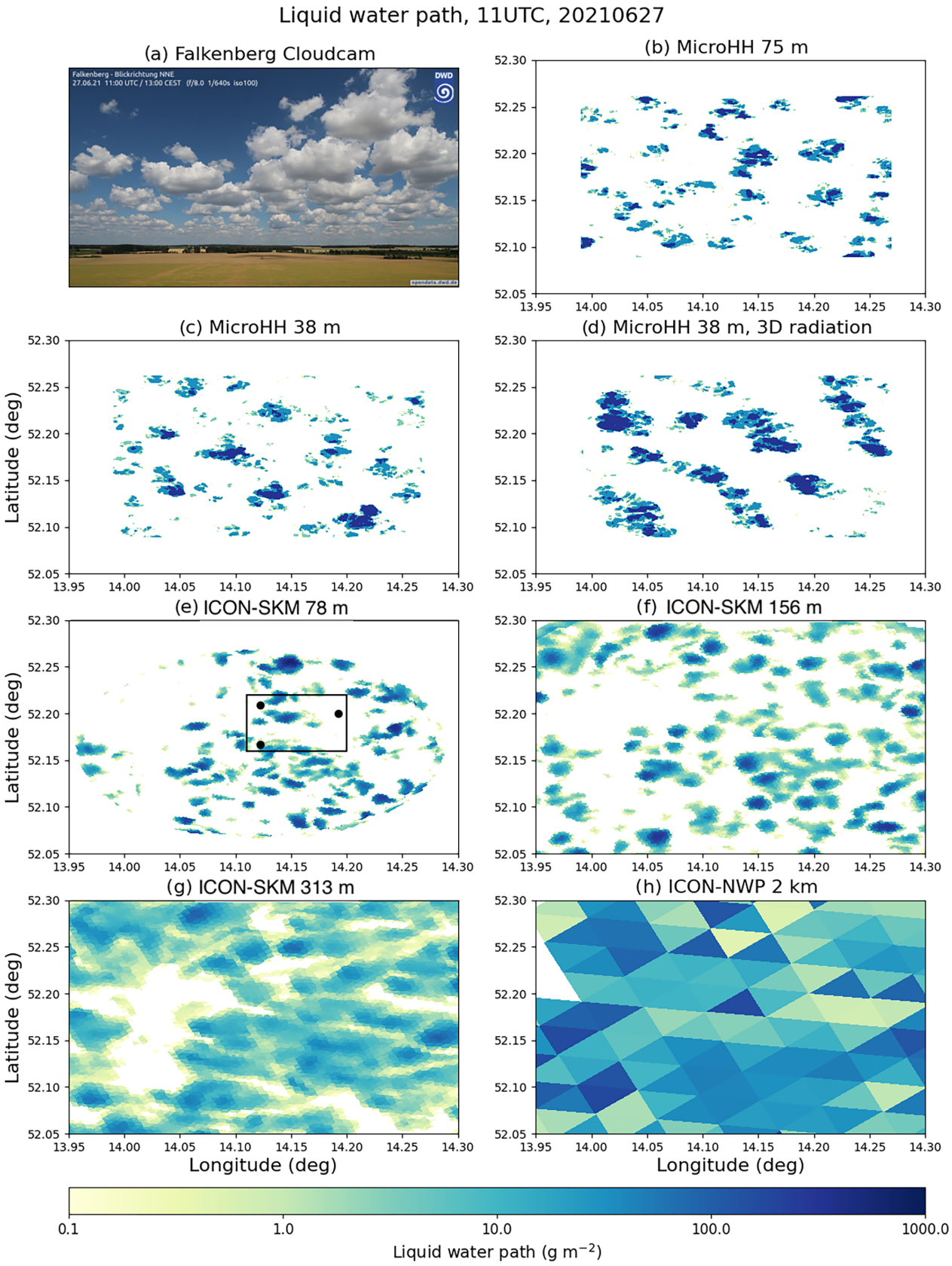

FIGURE 11 Cloud scenes at 1100 UTC on June 27, 2021. (a) Camera view at the Falkenberg measurement field and maps of simulated liquid water path from (b) MicroHH 75 m, (c) MicroHH 38 m, (d) MicroHH 38 m with 3D radiation, (e) ICON-SKM 78 m, (f) ICON-SKM 156 m, (g) ICON-SKM 313 m, and (h) ICON-NWP 2 km. The same spatial region is plotted for all simulations, that is, at coarser resolutions only a fraction of the ICON model domain is shown. [Colour figure can be viewed at wileyonlinelibrary.com]

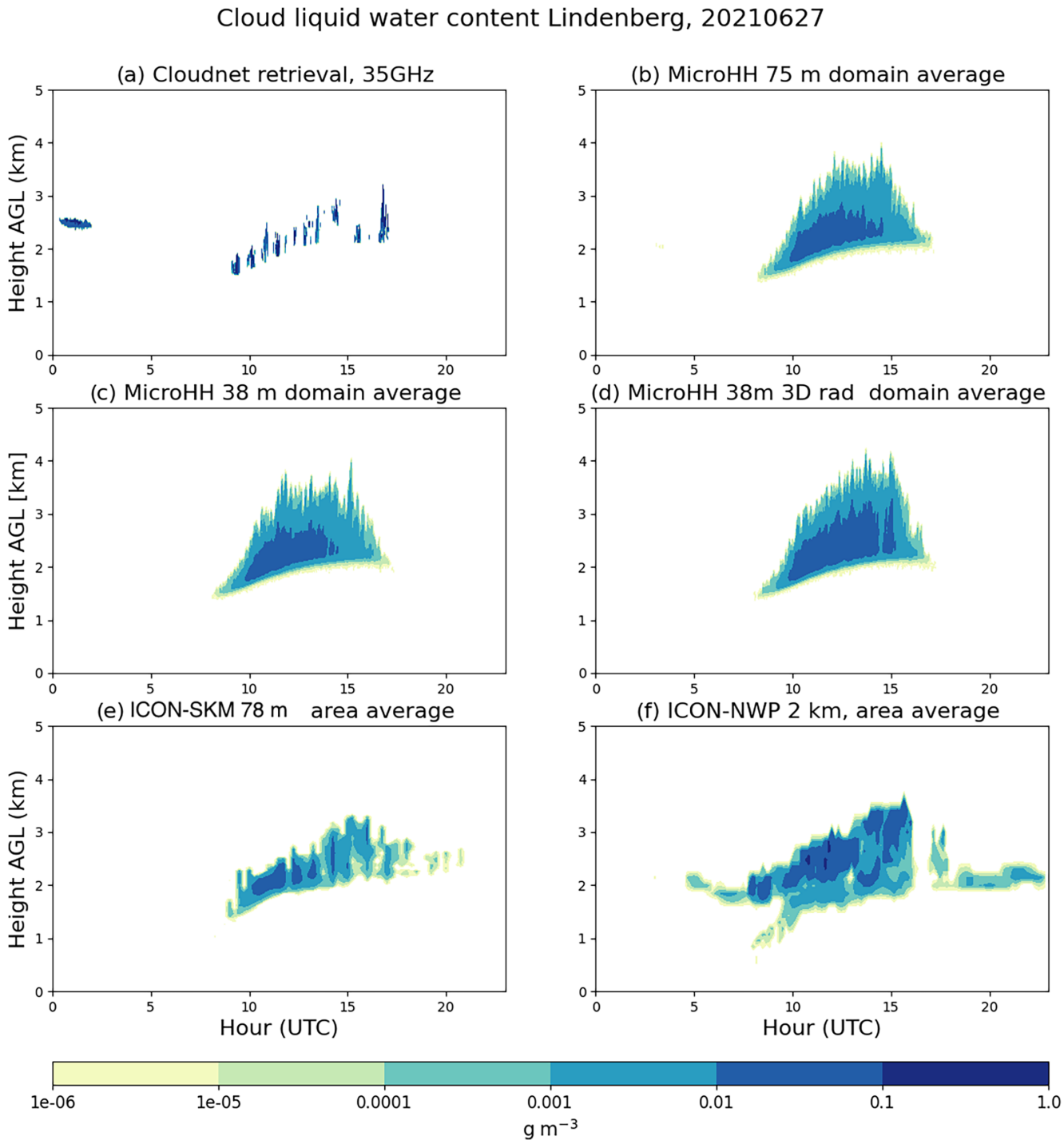

FIGURE 12 Time–height cross-section of (a) liquid cloud condensate from the Cloudnet retrieval at Lindenberg, (b) domain-average liquid condensate from MicroHH 75 m, (c) MicroHH 38 m, (d) MicroHH 38 m with 3D radiation, (e) area-averaged liquid condensate from ICON-SKM simulation 78 m, and (f) area-averaged liquid condensate from the ICON-NWP 2-km simulation. [Colour figure can be viewed at wileyonlinelibrary.com]

forcing used for this configuration leading to a moister residual layer. Note also a cloud layer with lower condensate concentration near cloud base, unlike MicroHH with the highest concentrations near cloud base. A quantitative comparison with the observations is difficult, since the vertically pointing ground-based remote sensing instruments can only observe individual clouds advected overhead. Clearly, in-cloud condensate concentrations in an individual cloud will be expected to be higher than the all-sky average condensate amounts from the models, which dilute in-cloud values by including cloud-free areas in the average. While the cloud base is determined quite reliably

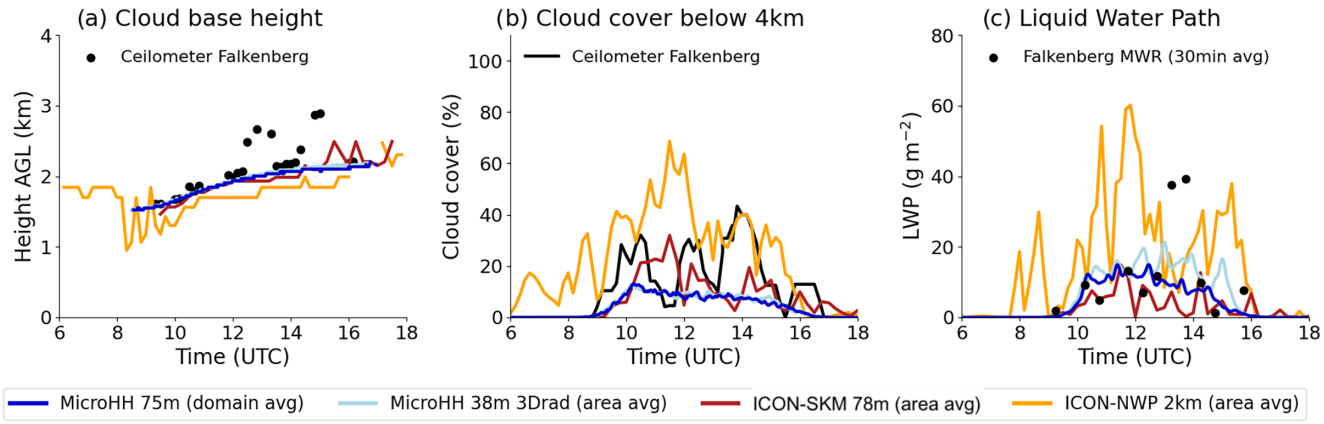

FIGURE 13 Observed (dots) and simulated (colored lines) (a) cloud-base height, (b) total cloud cover, and (c) liquid water path. The solid colored curve marks the average across a lat/lon region encompassing all three FESSTVaL supersites. Ceilometer estimates of cloud fraction are hourly running mean values. [Colour figure can be viewed at wileyonlinelibrary.com]

by the ceilometer, the cloud depth derived by Cloudnet has greater uncertainties due to the limited sensitivity of the radar. In this case, however, the derived cloud depths of 500–800 m agree with visual observations and are only slightly below those of the SKM simulations. Thus, this cross-section also serves as a more qualitative description of the cloud structure on the day. The following figure will consider more quantitative measures.

Figure 13 shows the observed cloud-base height, total cloud cover, and liquid water path for this case study at Falkenberg. Shown alongside are the same parameters derived from the ICON simulations in NWP mode, in SKM mode (78-m resolution), and from MicroHH (75- and 38-m resolution).

The model cloud-base height shown in Figure 13a is calculated in all three simulations from the domain- or area-averaged liquid water profile, searching for the first model layer exceeding a threshold of 1×10^{-4} g · m^{-3}. The ICON-SKM and the LES model agree quite well with the ceilometer-observed cloud-base height, while the NWP model (with a shallow convection parameterization active) produces clouds earlier in the day and with lower cloud base. However, the convection parameterization offers an alternative cloud-base height definition, based on a test parcel ascent identifying the first model layer where condensation occurs. This convection base height is 100–200 m higher (not shown), located just below the highest condensate values in Figure 12f. It appears that different processes produce different estimates of the cloud base in the NWP configuration. One should keep in mind that only eight grid points contribute to the cloud-base height estimate in the specified region.

In Figure 13b, the corresponding ceilometer-derived cloud cover is shown. Here, only cloud cover associated with clouds below 4 km is considered. For the model simulations, a liquid-only cloud cover below 4 km is shown. Given that the wind speed was low on this day and clouds are advected relatively slowly past the ceilometer, a 60-min running mean of the observed cloud cover is shown, to be representative of an area that is more similar in size to the model averaging area. The NWP simulation produces about twice as much cloud cover as the ICON-SKM simulation, and MicroHH has the lowest cloud cover, consistent with the qualitative views in Figure 11.

Figure 13c shows the liquid water path. The MWR retrieval is averaged for 30-min intervals to approximate the all-sky LWP values from the models better. The HATPRO MWRs estimate the LWP with an uncertainty of below 20 g · m^{-2} (Löhnert & Crewell, 2003). For shallow clouds on June 27, the observed lowest values below 20 g · m^{-2} are most likely not robust, although we see quite good agreement between observations and SKM for most of the day. The NWP produces consistently higher LWP compared with SKM and LES, and approaches the two observed values at about 40 g · m^{-2} more closely. Interestingly, while MicroHH has slightly lower cloud cover than ICON-SKM, the reverse is true for LWP. It appears that MicroHH concentrates its condensate in a smaller area compared with ICON-SKM. This is qualitatively consistent with the somewhat deeper clouds in Figure 12b and fewer thin cloud edges (lightest shades) in Figure 11b. We can also see that, while the 3D radiation in MicroHH increases LWP, it has virtually no impact on the cloud cover, in line with Tijhuis *et al.* (2024).

The combined impact of cloud cover and LWP on the global horizontal irradiance at Falkenberg is shown in Figure 14. In Figure 14a, the model irradiance is shown for the area/domain average, together with observations at Falkenberg. Consistent with the greater cloud cover

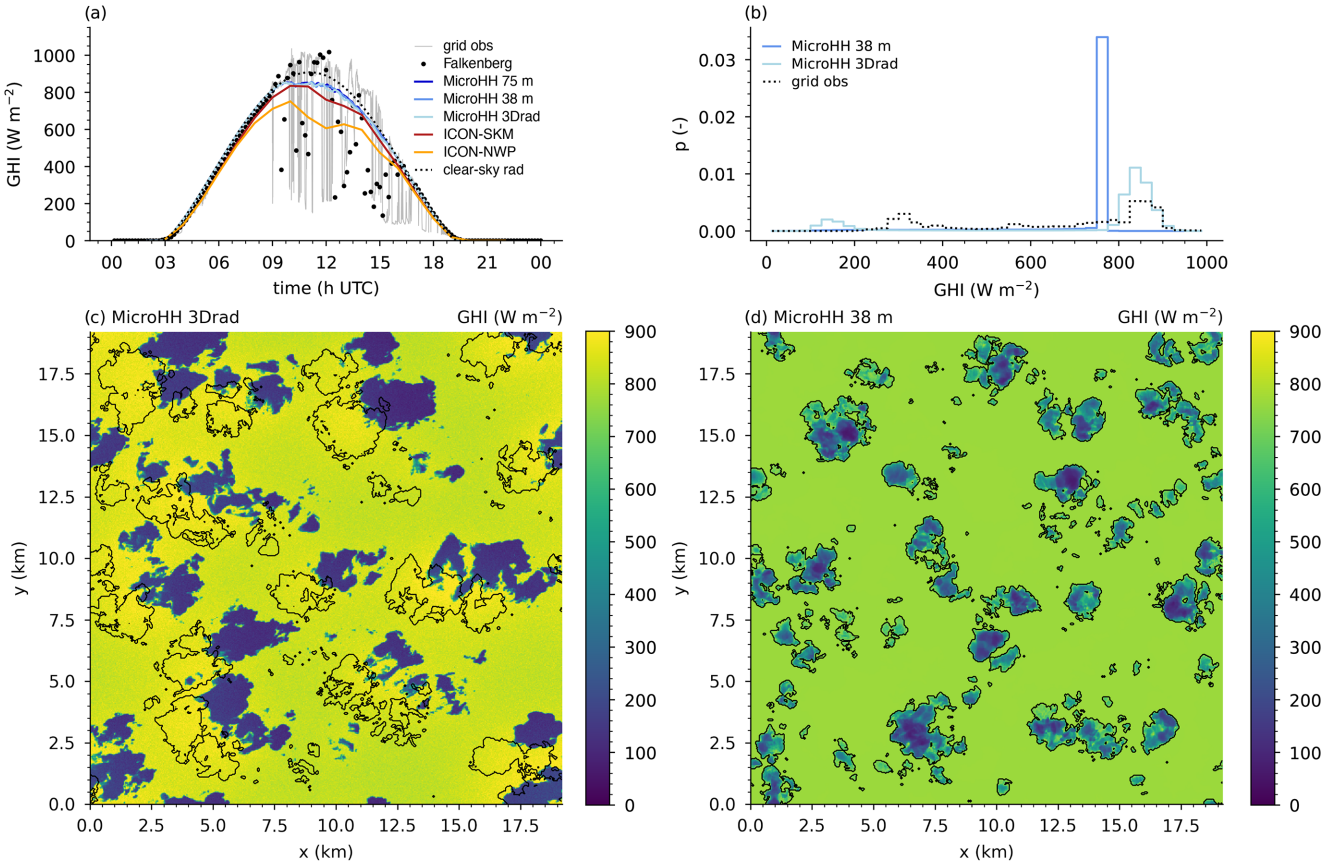

FIGURE 14 (a) Observed (black dots for pyranometer, gray curve for grid) and simulated (colored lines) global horizontal irradiance for June 27, 2021 at Falkenberg. All model versions show an area-averaged radiative flux: the domain average for MicroHH and an area average covering all three supersites for the ICON simulations. (b) Probability density functions of the spatial field of global horizontal irradiance at 1340 UTC taken from the observational grid and simulations with 1D and 3D radiation. (c) Global horizontal irradiance at 1340 UTC in MicroHH simulation with 3D radiation. (d) Global horizontal irradiance at 1340 UTC in MicroHH simulation with 1D radiation. [Colour figure can be viewed at wileyonlinelibrary.com]

and LWP in the NWP configuration, the solar irradiance is about 200 W · m^{-2} lower compared with ICON-SKM. While the SKM and LES configurations are more comparable in terms of solar irradiance, more solar radiation reaches the surface in MicroHH, due to the lower cloud cover, despite MicroHH having a greater LWP. As cloud cover reduces the amount of direct radiation reaching the surface, its impact on the surface irradiance is greater than, for example, differences in microphysical properties, which merely modulate the diffuse radiation.

The probability density function of radiation, extracted from the simulated global horizontal irradiance at 1340 UTC and the grid observations in the 10-min window surrounding this time, show that only the simulation with 3D radiation is able to capture the cloud enhancements that represent the right peak of the probability density function. At the same time, while the simulation with 3D radiation shows the left peak corresponding to the cloud shadows, it is at too low a value. Most likely, this is the result of the absence of aerosols in the simulations and the consequent lack of scattering due to the missing direct effect.

The cross-sections of global horizontal irradiance at 1340 UTC (Figure 14c,d) provide a deeper explanation of the shape of the PDF. We find that the simulation with 3D radiation has deeper cloud shadows (less than 200 W · m^{-2}) and large regions with radiation exceeding clear-sky radiation, whereas the simulation with 1D radiation has brighter shadows under the smaller clouds, as all diffuse radiation is transferred downwards. Therefore, regions of cloud enhancements surrounding the cloud shadows are absent.

5.4 | Case study June 29: cold pools

On June 29, 2021, the most intense cold pool (CP) of the FESSTVaL campaign was observed, named "Jogi". As described in Section 4, numerous intense convective cells developed in the afternoon along a band stretching from

northwest to southeast across eastern Germany. Using the operational 13-km ICON global forecast with a 6.5-km nest over Europe to provide initial and boundary forcing from June 29, 2021, 0000 UTC until June 29, 2021, 2300 UTC, ICON in NWP mode (2-km horizontal spacing) reliably produces intense convective cells for this case study on the dedicated FESSTVaL limited-area domain (DOM2 in Figure 2c).

Generating a satisfactory LES forcing setup proved to be less straightforward. The setup that worked well for the first two case studies turns out to have some drawbacks: The idealized forcing used in the MicroHH setup applies large-scale tendencies derived from a single ERA5 grid point to the entire double-periodic domain. Deep convection develops rather late and suddenly in the MicroHH LES, as a direct result of strong upward motion in the large-scale forcing.

While the ICON-SKM setup with forcing derived from the operational ICON-D2 forecast model does not suffer this problem, the nested domains reduce in size quite significantly as resolution increases (Figure 2b). Cold pools form in the outer domains and propagate into the inner domains. As a result, precipitation in the inner SKM domains is an unrealistic mixture of "inherited" convective systems advected through the lateral boundaries from the outer domains and convection developing locally within the domain.

We mitigate this effect by increasing the domain size and nesting the domains with a narrower gap between outer and inner domains. In order to keep the computational cost manageable, the outer domain (D1) now has a horizontal grid spacing of 525 m and the inner domain (D2) a spacing of 260 m. Figure 2c shows the new setup for this case, which covers approximately the same area as the innermost domain (DOM3) of the NWP setup (see Figure 2b). With this setup, the evolution of cold pools can be modeled adequately. Figure 15 shows an example snapshot of the 1-h accumulated precipitation and 2-m temperature from the ICON-NWP simulation and the two ICON-SKM simulations. All three simulations have convective cells producing precipitation and cold pools in the early afternoon. As one might expect, the SKM simulations show more numerous, smaller precipitation cells compared with the NWP setup.

Kirsch *et al.* (2024) characterize the morphology of the observed CPs by describing the evolution of their area fraction, object-mean temperature perturbation, and accumulated precipitation. We can analyze individual CPs in the model simulations in a similar manner. Clearly, a single case study is not sufficient to draw robust conclusions on quantitative measures describing cold pools. We use this case study to demonstrate how observed and simulated CPs can be compared in an equitable manner and point out some qualitative results. For a more comprehensive study, Kirsch *et al.* (2024) provide a list of all observed CP objects, and one could simulate all 40 days for a more quantitative and representative analysis.

Cold pools are identified from the ICON simulations by manually defining an area in the northwest quadrant of the domains (marked by an ellipse in Figure 15) and calculating cold-pool properties within this region, starting from the first appearance of significant precipitation in each of the simulations (NWP: 1320 UTC, D1: 1220 UTC, D2: 1200 UTC). The area was selected as all three simulations produce isolated CPs at approximately the same time, and is the equivalent of the roughly circular instrument network deployed during FESSTVaL. The model accumulated precipitation in Figure 16c is the accumulation within this circular region, analogous to the observed accumulated precipitation within the range of the X-band radar used for the observations. The definitions for area fraction and object-mean temperature perturbation follow those from Kirsch *et al.* (2024), considering the area of all points within the circular region with a temperature drop of 2 K or more. The reference temperature field used to calculate the temperature drop is that from the time step of CP initiation, correcting for the diurnal evolution of 2-m temperature (the reference temperature decreasing each time step according to the domain-average temperature evolution). One deliberate difference in the CP detection between model and observations is the size of the detection region. Many of the observed CPs grew to cover an area larger than the FESSTVaL measurement network, such that only the growth phase of the CPs could be tracked but not their decay. In the model we have no such constraints, and we choose a circular region with approximately three times the diameter of the FESSTVaL network. This allows us to also track the CPs throughout their entire life cycle. In Figure 16, the model-derived CP characteristics are shown alongside the observed characteristics of CP "Jogi".

We can see that the simulated cold pools thus identified vary quite dramatically between simulations. The NWP setup (forced with ICON-EU boundary conditions) produces the most precipitation within the first two hours, leading to the largest and most intense cold pool among the simulations, therefore being most like the observed cold pool. SKM (forced with ICON-D2 data) produces less precipitation initially and smaller, weaker cold pools. The CP onset is earliest for the innermost domain and latest in the NWP simulation. As was shown in Kirsch *et al.* (2024), precipitation amount is closely related to CP area and intensity. Figure 16d,e shows CP area and intensity normalized for the growth period of the CPs. For the normalized plots, we select only the time period between CP onset and the end of the growth phase (marked by dots in Figure 16a),

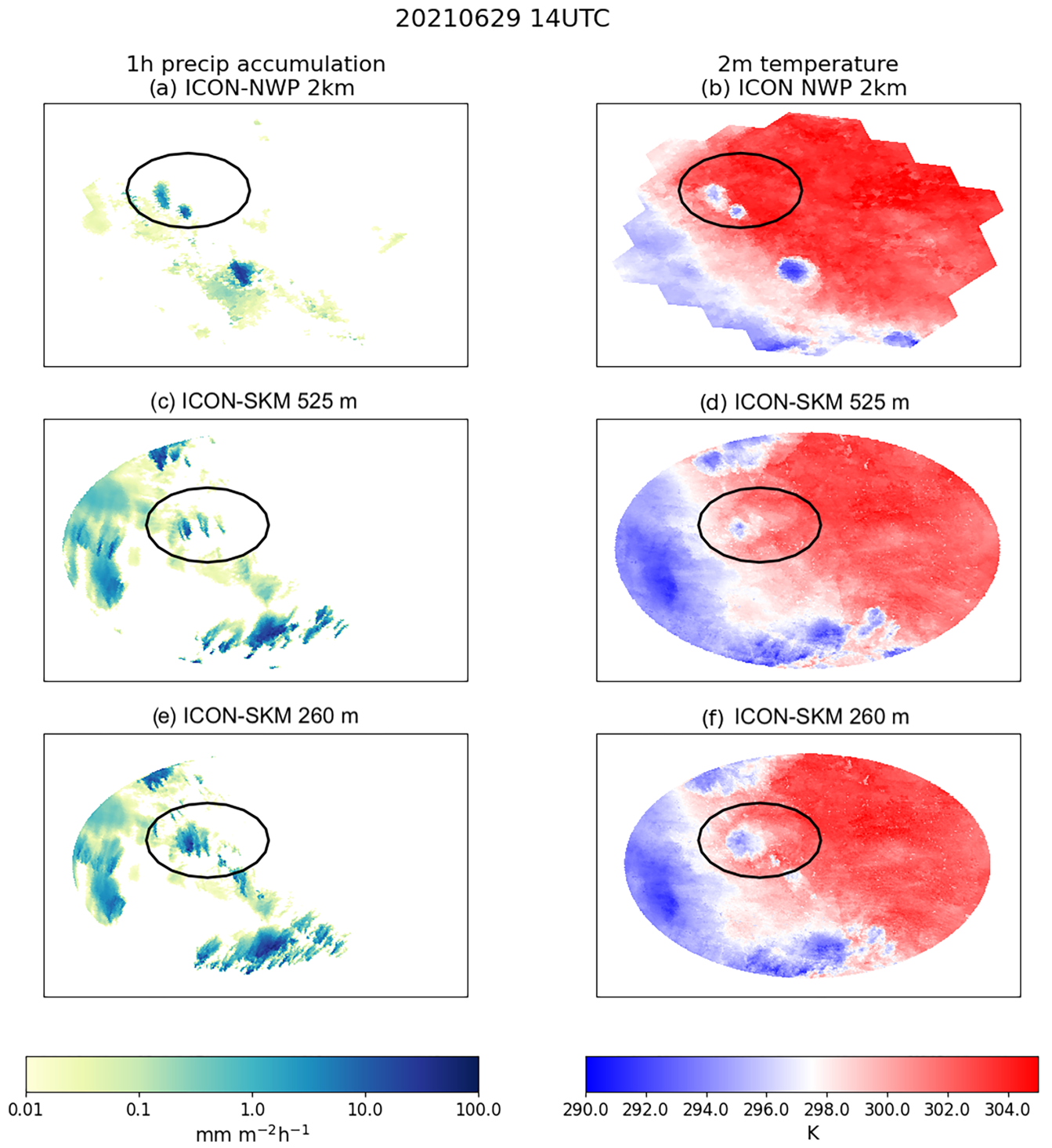

FIGURE 15 Maps of (left) precipitation accumulated over the previous hour and (right) 2-m temperature, from (a,b) ICON 2-km NWP simulation, (c,d) ICON-SKM D1 (525 m), and (e,f) ICON-SKM D2 (260 m) for the FESSTVaL limited-area domain on June 29, 2021 at 1400 UTC. The circles mark the area within which cold-pool properties are analyzed. [Colour figure can be viewed at wileyonlinelibrary.com]

and divide area, intensity, and accumulated precipitation by the maximum value of each property within this time frame before plotting the normalized quantities. All simulations approximate the nearly linear relationship between accumulated precipitation and CP area that was also noted in Kirsch *et al.* (2024). The NWP and D2 SKM simulations also show that the maximum intensity is reached after approximately two-thirds of the growth phase. Qualitatively, these two cold pools show good agreement with the observed morphology.

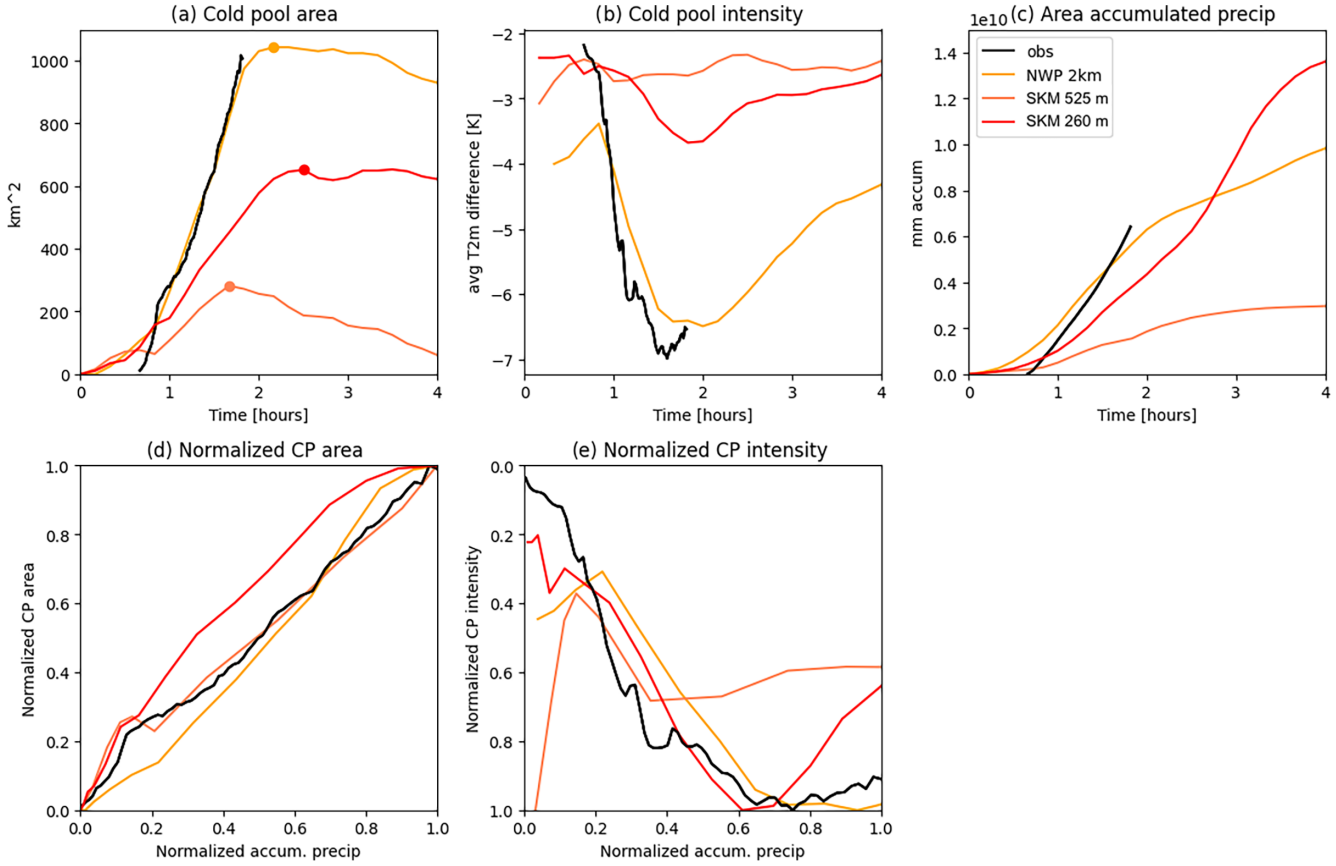

FIGURE 16 (a) CP area, (b) CP intensity, and (c) accumulated precipitation for the observed CP "Jogi" (black) and simulated cold pools from the ICON-NWP and SKM simulations on the same day, June 29, 2021. The bottom row shows (d) CP area and (e) intensity versus accumulated precipitation, normalized over the growth phase of the CPs only. [Colour figure can be viewed at wileyonlinelibrary.com]

6 | DISCUSSION ON RESOLUTION DEPENDENCES

Starting from the land surface, ICON-SKM includes finer-resolved properties, such as orography, land cover, and land-use classes, compared with a coarser-resolution NWP configuration. Better resolved land-surface properties coupled directly to the resolved convective circulations and clouds allow for the feedback of these processes into the surface heat flux and soil moisture. Thus, even though the formulation of the surface turbulent fluxes of sensible and latent heat is identical in the ICON-SKM and -NWP configurations, we simulate a lower sensible and higher latent heat flux during the convective daytime hours in ICON-SKM (Figure 7a,b).

Atmospheric large eddies and cloud dynamics are resolved explicitly in ICON-SKM, which eliminates the simplifications and biases introduced by a convective parameterization. Turbulent eddies are, however, highly dependent on model resolution (Figure 8). The results shown in Figure 8 remind us of the need to develop scale-aware turbulence parameterizations to bridge fully resolved and fully parameterized turbulent scales. One aspect of a scale-aware parameterization of turbulence is the transition from a 1D vertical to a 3D representation of mixing with the refinement of model resolution (Honnert *et al.*, 2020). As our case demonstrates, the variance at the highest model resolution is not lacking when we employ a 1D hybrid turbulence scheme. Still, the partitioning between resolved and subgrid flow has to be adapted, as the model grid spacing increases from about 78 m to 313 m. In our case, it is most likely that the subgrid contribution to the TKE has to increase more than it currently does with the coarsening of resolution at all heights in the mixed layer to bring the total variance profiles together, that is, to achieve scale awareness. In addition, the scale dependence of the resolved convective flow should be tackled, as the resolved flow due to nonlinear convective dynamics forms spurious convective circulation patterns with spatial dimensions set by the model resolution instead of, for example, boundary-layer depth. It is an open question whether one should develop a subgrid-scale parameterization that would eliminate the spurious circulations, or instead compensate for their deficiencies in some way (e.g., Ching *et al.*, 2014; Sakradzija *et al.*, 2016).

Clouds resolved by ICON-SKM are of a more fractional and compact formation, lower liquid water path, and lower fractional coverage compared with NWP (Figures 11, 13, and 12), which reflects in a better match of the total radiative flux at the surface with the observed one (Figure 7c). In terms of spatial patterns of light and shadows, only MicroHH-LES with 3D radiation is able to reproduce the field observations. This is because horizontal transfer of diffuse radiation is essential in setting the darkness of the shadows, as well as in creating regions with surface solar irradiance with magnitudes larger than clear-sky radiation (cloud enhancements) surrounding cloud shadows. Solving the shortwave radiation in 3D impacts cloud properties, in line with earlier studies (e.g., Jakub & Mayer, 2015; Tijhuis *et al.*, 2024; Veerman *et al.*, 2022).

Modeled convective clouds, originating from buoyant updrafts starting near the land surface, are already highly influenced by the model grid resolution at sub-km scales. The impact of the grid resolution on the resolved cloud dynamics is considerable (Figures 12–13), as the model resolution crosses the convective "gray zone", starting in our case from a resolution of 78 m to about 2.2 km in NWP. Clouds resolved in ICON-SKM are shallower, fractional, and initiate later in the morning (at about 0900 UTC) compared with the clouds in ICON-NWP, where we also already see a non-convective cloud layer developing in the early morning hours before sunrise (Figure 12). The cloud-base height is reduced by up to several hundred meters, while the liquid water path peaks up to four times higher in NWP compared with ICON-SKM (Figure 13). As the model resolution coarsens, the clouds become wider, with lower water content and less well-defined cloud boundaries, until they become smoothed and form a cloud layer without clear-sky gaps at about 2.2-km resolution (Figure 15). These results demonstrate the need to develop and improve further the parameterizations of convection and clouds for use in sub-km NWP and climate models.

Deep convective dynamics are resolved in higher detail in ICON-SKM compared with the NWP model, where SKM and LES show more numerous and smaller precipitation cells (Figure 15). This resolved dynamics most likely influences the correct timing of the cold-pool onset in SKM (Figures 4 and 6). However, deep clouds and associated cold pools are highly influenced by the finite domain sizes of the SKM. Proper treatment of deep convection and cold pools requires about 10 times larger spatial domains in SKM compared with shallow ABL convection cases. The domain size puts a constraint on the horizontal grid spacing, which, in our case, is limited to 260 m. The cold pools developed in D2 at 260-m resolution are broader and stronger than the cold pools of the coarser domain D1 (Figure 16); however, they still do not reach the absolute intensity of the observed CP. When normalized by the growth phase of the CP, the match of the modeled CP with the observed characteristics is qualitatively independent of the model resolution.

In overview of the scale dependence, we find that the domain-average temperature, humidity, winds, and surface turbulent and radiative fluxes are invariant to model resolution. The sub-mesoscale and temporal variability of these properties, however, is directly influenced by resolution-dependent turbulent and convective dynamics, and, in the second place, the ability of the model configuration to introduce spatial variability of the land-surface properties.

7 | SUMMARY

Our study accompanies the FESSTVaL observational campaign, with the aim of selecting three unique golden days (clear sky, shallow convection, deep convection), allowing a comparison of LES and NWP under different conditions. The analysis focuses on three days in June 2021, representative of clear-sky ABL turbulence (June 14), shallow convective clouds and surface radiation (June 27), and deep convective cold pools (June 29). The state-of-the-art operational ICON-NWP model is challenged to approach the spatial resolutions of MicroHH, a cutting-edge LES model developed and used for ABL research. We find that the operational ICON model used at sub-km resolutions (ICON-SKM) meets the performance of MicroHH-LES in the daytime ABL, and, although not surprisingly, it outperforms the idealized LES in cases where realistic spatial heterogeneities are relevant.

In a test of basic requirements, the SKM model reproduces mean thermodynamic states and winds with high fidelity and is comparable with MicroHH-LES. Improvements brought by ICON-SKM with respect to ICON-NWP are found mainly in total surface radiation flux, due to better-resolved ABL dynamics, clouds, and land-surface properties. This further causes differences in surface turbulent heat flux, near-surface temperature, and humidity between the SKM and NWP configurations. However, a direct evaluation of the surface turbulent heat flux compared with the observed estimates using the eddy-covariance method and scintillometry remains elusive, due to a treatment of surface energy balance closure in the models that does not correspond to the observed gap in the closure.

Sub-mesoscale variability of near-surface temperature during convective daytime is reproduced well by ICON-SKM on the day with shallow cumulus clouds, and less so in clear-sky conditions and non-convective periods. The difference in the reproduced spatial variability between the SKM and LES models demonstrates the

importance of a realistic representation of the land-surface properties in SKM runs. Cloud shadows introduce significant additional spatial variability, while the LES grid spacing of 75 m already resolves most of the variability, since no further contribution to the spatial variation in temperature is found in our highest resolution LES at 38 m.

The SKM and LES models resolve a slightly higher spatially averaged variance of vertical velocity compared with the estimated variance from a vertically pointed Doppler lidar, confirming the findings of similar previous studies. The estimates of TKE using a conically scanning Doppler lidar are in good agreement with observations, showing equal importance of resolved and subgrid contributions to the modeled TKE at sub-km resolutions.

The increase in the resolution of the models brings significant benefits to the representation of convective clouds and their effects on radiation transfer. Although the SKM model shows differences in the cloud depths and liquid-water content near cloud base compared with LES, the two models are consistent in the timing of cloud onset and decay. In contrast to SKM and LES, the onset and decay of clouds occurs in the early morning hours and extends to late evening in ICON-NWP. The cloud dynamics and spatial distributions of the cloud fields are sufficiently resolved at SKM and LES resolutions of <100 m. Furthermore, the cloud-base height modeled by SKM and LES closely matches the estimate from the ceilometer measurements.

Challenges in simulating realistic deep convective cold pools are related to resolution, domain size, nesting strategy, and specification of the lateral boundary conditions. Large-scale forcing plays a dominant role in all configurations. When cyclic boundary conditions are used, point-wise forcing dictates the rapid onset of cold pools and their expansion until the whole domain is saturated. In the nested SKM setup, propagation of systems from parent to child domains overtakes the cold-pool dynamics, so additional considerations have to be made to allow undisturbed local cold-pool development. Our setup reaches a qualitatively realistic cold-pool development, with a prospect of further improvements at higher grid resolutions.

The dependence on the model resolution as an inevitable modeling feature (e.g., Jung & Arakawa, 2004; Weisman *et al.*, 1997) is outlined and quantified in our study for several micro- and sub-mesoscale processes, aiming to provide guidelines for future studies. Our analysis reaches convergence for most turbulent and shallow convective properties and spatial variability during the daytime. However, such representativeness of non-convective and stable boundary layers has not been achieved at sub-km resolutions, as shown in Figures 4 and 6. Stable conditions have been challenging for both LES and NWP in the past (Beare *et al.*, 2006; Sandu *et al.*, 2013). However, we demonstrate that both SKM and LES models are reliable in reproducing the daytime average thermodynamical and dynamical ABL states, sub-mesoscale temperature variability, turbulent and convective variances in the observed range, and a qualitatively good representation of the processes, such as convective cold pools, for which further refinement in the model resolution is still needed.

With the three case studies presented here, we would like to propose a modeling benchmark based on FESSTVaL. We have demonstrated the value of the comprehensive observational dataset collected during this experiment for model studies on the evolution of convective boundary layers in different weather situations. Similarly, the dataset appears to be suitable for studying the night-time stable boundary layer. By focusing only on a subset of observations conducted during FESSTVaL, we did not make use of all the spatially distributed datasets from the three supersites, the X-band radar, and additional data from the surface networks. Moreover, UAV flight data are available for a number of days during the IOP. Our modeling setup can easily be expanded to cover the entire FESSTVaL campaign, thus achieving more robustness in the representation of specific processes, which will be the subject of dedicated follow-up studies.

ACKNOWLEDGEMENTS

Thanks for performing and carefully analyzing the FESSTVaL measurements used in this article go to Udo Rummel, Ronny Leinweber, Carola Detring, Nadja Samtleben, and Larissa Scholz (all DWD Lindenberg), and to Wouter Mol and Bert Heusinkfeld (WUR). We thank Arianna Valmassoi (DWD) for providing the Corine Land Cover maps in ICON Extpar and Daniel Klocke (MPI-M) for the initial setup of the LES for FESSTVaL, upon which current experiments are built. We appreciate the insightful comments of two anonymous reviewers, which helped us to refine our work. The statistical analysis is performed using the Python programming language (Python Software Foundation, https://www.python.org/). This work used the resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID bb1096. This research was supported through the Hans-Ertel-Centre for Weather Research (Hans-Ertel-Zentrum für Wetterforschung; HErZ) and German Meteorological Service (Deutscher Wetterdienst), funded by the Federal Ministry of Transport (Bundesministerium für Verkehr; BMV; Grant number 4823DWDP3). Chiel van Heerwaarden is funded by the Dutch Research Council (grant VI.Vidi.192.068). The Cloudnet data used in this study are generated by the Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS) and are available from

the ACTRIS Data Centre using the following link: https://hdl.handle.net/21.12132/1.edbbe76762ba4f66. We acknowledge ACTRIS and Finnish Meteorological Institute for providing the Cloudnet data set which is available for download from https://cloudnet.fmi.fi. We acknowledge ECMWF for providing IFS model data. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The model output is generated using ICON, which is available to the community under a permissive open-source licence (BSD-3C): see https://icon-model.org. The observational data are available as a SAMD archive at https://www.fdr.uni-hamburg.de/search?page=1&size=20&q=fesstval with open access (last accessed on July 28, 2025). MicroHH and accompanying documentation can be found on the model's website at http://microhh.org (last accessed on July 28, 2025). The code can be accessed via the website or directly at GitHub from https://github.com/microhh/microhh (last accessed on July 28, 2025). The GitHub repository is coupled to Zenodo, which provides DOIs for released software. The set of ICON and MicroHH model outputs used in this study is stored at the German Climate Computation Center (DKRZ). The final model output used in the analysis is available in the Long-term Archive of Climate Model Data at DKRZ (DOKU): https://www.wdc-climate.de/ui/entry?acronym=DKRZ_LTA_1096_ds00036.

ORCID

Mirjana Sakradzija https://orcid.org/0000-0002-1544-8268
Maike Ahlgrimm https://orcid.org/0000-0002-8482-0668
Frank Beyrich https://orcid.org/0009-0006-7554-3634
Chiel van Heerwaarden https://orcid.org/0000-0001-7202-3525
Eileen Päschke https://orcid.org/0009-0001-5553-7900
Bastian Kirsch https://orcid.org/0000-0002-9682-0171
Juerg Schmidli https://orcid.org/0000-0002-6322-6512
Noviana Dewani https://orcid.org/0000-0001-7990-4704

REFERENCES

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M. et al. (2009) A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *Journal of Hydrometeorology*, 10, 623–643.

Bašták Ďurán, I., Köhler, M., Eichhorn-Müller, A., Maurer, V., Schmidli, J., Schomburg, A. et al. (2021) The ICON single-column mode. *Atmosphere*, 12, 906.

Bauer, H.-S., Späth, F., Lange, D., Thundathil, R., Ingwersen, J., Behrendt, A. et al. (2023) Evolution of the convective boundary layer in a WRF simulation nested down to 100 m resolution during a cloud-free case of LAFE, 2017 and comparison to observations. *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037212.

Beare, R.J., Macvean, M.K., Holtslag, A.A.M., Cuxart, J., Esau, I., Golaz, J.-C. et al. (2006) An intercomparison of large-eddy simulations of the stable boundary layer. *Boundary-Layer Meteorology*, 118, 247–272.

Beyrich, F. & Adam, W.K. (2007) *Site and data report for the Lindenberg reference site in CEOP - Phase I*, Vol. 230. Offenbach am Main, Germany: Berichte des Deutschen Wetterdienstes, p. 55.

Beyrich, F. & Mengelkamp, H. (2006) Evaporation over a heterogeneous land surface: EVA_GRIPS and the LITFASS-2003 experiment - An Overview. *Boundary-Layer Meteorology*, 121(1), 5–32.

Beyrich, F., Hartogensis, O., de Bruin, H. & Ward, H. (2021) Scintillometers. In: Foken, T. (Ed.) *Springer handbook of atmospheric measurements*, 1st edition. Germany: Springer, pp. 969–997.

Boutle, I.A., Eyre, J.E.J. & Lock, A.P. (2014) Seamless stratocumulus simulation across the turbulent gray zone. *Monthly Weather Review*, 142, 1655–1668.

Ching, J., Rotunno, R., LeMone, M., Martilli, A., Kosovic, B., Jimenez, P.A. et al. (2014) Convectively induced secondary circulations in fine-grid mesoscale numerical weather prediction models. *Monthly Weather Review*, 142, 3284–3302. Available from: https://journals.ametsoc.org/view/journals/mwre/142/9/mwr-d-13-00318.1.xml

CORINE Land Cover 2018 (raster 100 m). (2020) Europe, 6-yearly - version 2020_20u1. European Union's Copernicus Land Monitoring Service information. Available at: https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac [Accessed 16th September 2025].

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

Dewani, N., Sakradzija, M., Schlemmer, L., Leinweber, R. & Schmidli, J. (2023) Dependency of vertical velocity variance on meteorological conditions in the convective boundary layer. *Atmospheric Chemistry and Physics*, 23, 4045–4058.

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

Dudhia, J. (2022) Challenges in sub-kilometer grid modeling of the convective planetary boundary layer. *Meteorology*, 1, 402–413.

Efstathiou, G. & Beare, R. (2015) Quantifying and improving sub-grid diffusion in the boundary-layer grey zone. *Quarterly Journal of the Royal Meteorological Society*, 141, 3006–3017.

Fast, J.D., Berg, L.K., Feng, Z., Mei, F., Newsom, R., Sakaguchi, K. et al. (2019) The impact of variable land-atmosphere coupling on convective cloud populations observed during the 2016 HI-SCALE field campaign. *Journal of Advances in Modeling Earth Systems*, 11, 2629–2654.

Field, P.R., Brozková, R., Chen, M., Dudhia, J., Lac, C., Hara, T. et al. (2017) Exploring the convective grey zone with regional

simulations of a cold air outbreak. *Quarterly Journal of the Royal Meteorological Society*, 143, 2537–2555.

Gassmann, A. & Herzog, H.-J. (2008) Towards a consistent numerical compressible non-hydrostatic model using generalized Hamiltonian tools. *Quarterly Journal of the Royal Meteorological Society*, 134, 1597–1613.

Giorgetta, M.A., Brokopf, R., Crueger, T., Esch, M., Fiedler, S., Helmert, J. et al. (2018) ICON-A, the atmosphere component of the ICON earth system model: I. model description. *Journal of Advances in Modeling Earth Systems*, 10, 1613–1637.

Goger, B. & Dipankar, A. (2024) The impact of mesh size, turbulence parameterization, and land-surface-exchange scheme on simulations of the mountain boundary layer in the hectometric range. *Quarterly Journal of the Royal Meteorological Society*, 150, 3853–3873. Available from: https://doi.org/10.1002/qj.4799

Goger, B., Rotach, M.W., Gohm, A., Fuhrer, O., Stiperski, I. & Holtslag, A.A. (2018) The impact of three-dimensional effects on the simulation of turbulence kinetic energy in a major alpine valley. *Boundary-Layer Meteorology*, 168, 1–27.

Grant, L.D., Kirsch, B., Bukowski, J., Falk, N.M., Neumaier, C.A., Sakradzija, M. et al. (2024) How variable are cold pools? *Geophysical Research Letters*, 51, e2023GL106784.

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Heusinkveld, B.G., Mol, W.B. & van Heerwaarden, C.C. (2023) A new accurate low-cost instrument for fast synchronized spatial measurements of light spectra. *Atmospheric Measurement Techniques*, 16, 3767–3785. Available from: https://doi.org/10.5194/amt-16-3767-2023

Hill, R.J. (1997) Algorithms for obtaining atmospheric surface-layer fluxes from scintillation measurements. *Journal of Atmospheric and Oceanic Technology*, 14, 456–467.

Hohenegger, C., Ament, F., Beyrich, F., Löhnert, U., Rust, H., Bange, J. et al. (2023) FESSTVaL: the field experiment on submesoscale spatio-temporal variability in Lindenberg. *Bulletin of the American Meteorological Society*, 104, E1875–E1892.

Honnert, R., Efstathiou, G.A., Beare, R.J., Ito, J., Lock, A., Neggers, R. et al. (2020) The atmospheric boundary layer and the gray zone of turbulence: a critical review. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD030317.

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

Ito, J., Niino, H., Nakanishi, M. & Moeng, C.-H. (2015) An extension of the mellor-yamada model to the terra incognita zone for dry convective mixed layers in the free convection regime. *Boundary-Layer Meteorology*, 157, 23–43.

Jakub, F. & Mayer, B. (2015) A three-dimensional parallel radiative transfer model for atmospheric heating rates for use in cloud resolving models–the TenStream solver. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 163, 63–71. Available from: https://doi.org/10.1016/j.jqsrt.2015.05.003

Jung, J.-H. & Arakawa, A. (2004) The resolution dependence of model physics: illustrations from nonhydrostatic model experiments. *Journal of the Atmospheric Sciences*, 61, 88–102.

Kealy, J.C., Efstathiou, G.A. & Beare, R.J. (2019) The onset of resolved boundary-layer turbulence at grey-zone resolutions. *Boundary-Layer Meteorology*, 171, 31–52. Available from: https://doi.org/10.1007/s10546-018-0420-0

Kirsch, B., Hohenegger, C. & Ament, F. (2024) Morphology and growth of convective cold pools observed by a dense station network in Germany. *Quarterly Journal of the Royal Meteorological Society*, 150, 857–876.

Kirsch, B., Hohenegger, C., Klocke, D., Senke, R., Offermann, M. & Ament, F. (2022) Sub-mesoscale observations of convective cold pools with a dense station network in Hamburg, Germany. *Earth System Science Data*, 14, 3531–3548.

Kohsiek, W., Meijninger, W., Debruin, H. & Beyrich, F. (2006) Saturation of the large aperture scintillometer. *Boundary-Layer Meteorology*, 121, 111–126.

Kooijmans, L.M.J. & Hartogensis, O.K. (2016) Surface-layer similarity functions for dissipation rate and structure parameters of temperature and humidity based on eleven field experiments. *Boundary-Layer Meteorology*, 160, 501–527. Available from: https://doi.org/10.1007/s10546-016-0152-y

Kosović, B., Jimenez Munoz, P., Juliano, T.W., Martilli, A., Eghdami, M., Barros, A.P. et al. (2020) Three-dimensional planetary boundary layer parameterization for high-resolution mesoscale simulations. *Journal of Physics: Conference Series*, 1452, 012080. Available from: https://doi.org/10.1088/1742-6596/1452/1/012080

Kracher, D., Mengelkamp, H.-T. & Foken, T. (2009) The residual of the energy balance closure and its influence on the results of three SVAT models. *Meteorologische Zeitschrift*, 18, 647–661. Available from: https://doi.org/10.1127/0941-2948/2009/0412

Leroyer, S., Bélair, S., Husain, S.Z. & Mailhot, J. (2014) Subkilometer numerical weather prediction in an urban coastal area: a case study over the vancouver metropolitan area. *Journal of Applied Meteorology and Climatology*, 53, 1433–1453.

Leuenberger, D., Koller, M., Fuhrer, O. & Schär, C. (2010) A generalization of the SLEVE vertical coordinate. *Monthly Weather Review*, 138, 3683–3689.

Lilly, D.K. (1962) On the numerical simulation of buoyant convection. *Tellus*, 14, 148–172.

Löhnert, U. & Crewell, S. (2003) Accuracy of cloud liquid water path from ground-based microwave radiometry 1. Dependency on cloud model statistics. *Radio Science*, 38, 8041. Available from: https://doi.org/10.1029/2002RS002654

Lorenz, E.N. (1960) Energy and numerical weather prediction. *Tellus*, 12, 364–373.

Louis, J.-F. (1979) A parametric model of vertical eddy fluxes in the atmosphere. *Boundary-Layer Meteorology*, 17, 187–202.

Lüdi, A., Beyrich, F. & Mätzler, C. (2005) Determination of the turbulent temperature–humidity correlation from scintillometric measurements. *Boundary-Layer Meteorology*, 117, 525–550. Available from: https://doi.org/10.1007/s10546-005-1751-1

Mason, P.J. (1989) Large-eddy simulation of the convective atmospheric boundary layer. *Journal of Atmospheric Sciences*, 46, 1492–1516.

Mauder, M., Foken, T. & Cuxart, J. (2020) Surface-energy-balance closure over land: a review. *Boundary-Layer Meteorology*, 177, 395–426.

Maurer, V., Kalthoff, N., Wieser, A., Kohler, M., Mauder, M. & Gantner, L. (2016) Observed spatiotemporal variability of boundary-layer turbulence over flat, heterogeneous terrain. *Atmospheric Chemistry and Physics*, 16, 1377–1400.

Moeng, C.-H. & Sullivan, P.P. (2015) Large-eddy simulation. *Encyclopedia of Atmospheric Sciences*, 2, 232–240.

Mol, W., Heusinkveld, B., Mangan, M.R., Hartogensis, O., Veerman, M. & van Heerwaarden, C. (2024) Observed patterns of surface solar irradiance under cloudy and clear-sky conditions. *Quarterly Journal of the Royal Meteorological Society*, 150, 2338–2363. Available from: https://doi.org/10.1002/qj.4712

NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team. (2019). *ASTER Global Digital Elevation Model V003 [Data set]*. NASA Land Processes Distributed Active Archive Center. Available at: https://doi.org/10.5067/ASTER/ASTGTM.003 [Accessed 16th September 2025].

Päschke, E. & Detring, C. (2024) Noise filtering options for conically scanning doppler lidar measurements with low pulse accumulation. *Atmospheric Measurement Techniques*, 17, 3187–3217. Available from: https://amt.copernicus.org/articles/17/3187/2024/

Pincus, R., Mlawer, E.J. & Delamere, J.S. (2019) Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems*, 11, 3074–3089.

Raschendorfer, M. (2001) The new turbulence parametrization of LM. *COSMO Newsletter*, 1, 89–97.

Raschendorfer, M. (2011) Further steps towards a scale separated turbulence scheme. In: *Proceedings of the 13th COSMO General Meeting in Rome 2011, Rome, Italy, 5-9 September 2011*. Available at: http://www.cosmo-model.org/content/consortium/generalMeetings/general2011/wg3a.htm [Accessed 16th September 2025].

Raschendorfer, M. (2016) *Common parameterisation of turbulence and surface-to-atmosphere transfer in COSMO and ICON*. Germany: Deutscher Wetterdienst, Offenbach am Main. Available from: https://www.cosmo-model.org/content/model/cosmo/development/turbulence_in_COSMO_and_ICON-Raschendorfer-2016.doc

Raschendorfer, M., Simmer, C. & Gross, P. (2003) Parameterization of turbulent transport in the atmosphere. *Dynamics of Multiscale Earth Systems, Lecture Notes in Earth Science*, 97, 167–185.

Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraf, C. et al. (2025) *DWD database reference for the global and regional ICON and ICON-EPS forecasting system*. Germany: Deutscher Wetterdienst, Offenbach am Main.

Rotunno, R., Chen, Y., Wang, W., Davis, C., Dudhia, J. & Holland, G.J. (2009) Large-eddy simulation of an idealized tropical cyclone. *Bulletin of the American Meteorological Society*, 90, 1783–1788.

Sakradzija, M., Seifert, A. & Dipankar, A. (2016) A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, 8, 786–812. Available from: https://doi.org/10.1002/2016MS000634

Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T. & Balsamo, G. (2013) Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *Journal of Advances in Modeling Earth Systems*, 5, 117–133.

Schär, C., Leuenberger, D., Fuhrer, O., Lüthi, D. & Girard, C. (2002) A new terrain-following vertical coordinate formulation for atmospheric prediction models. *Monthly Weather Review*, 130, 2459–2480.

Schmidli, J. & Quimbayo-Duarte, J. (2023) Diurnal valley winds in a deep alpine valley: model results. *Meteorology*, 2, 87–106.

Schulz, J.-P. & Vogel, G. (2020) Improving the processes in the land surface scheme TERRA: bare soil evaporation and skin temperature. *Atmosphere*, 11, 513.

Schulzweida, U. (2023) *CDO User Guide (2.3.0)*. Zenodo. Available from: https://doi.org/10.5281/zenodo.10020800

Seifert, A. & Beheng, K.D. (2001) A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, 59, 265–281.

Shin, H.H. & Hong, S.-Y. (2015) Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Monthly Weather Review*, 143, 250–271.

Smagorinsky, J. (1963) General circulation experiments with the primitive equations: I. the basic experiment. *Monthly Weather Review*, 91, 99–164.

Smalikho, I.N. & Banakh, V.A. (2017) Measurements of wind turbulence parameters by a conically scanning coherent Doppler lidar in the atmospheric boundary layer. *Atmospheric Measurement Techniques*, 10, 4191–4208.

Smith, D.K., Renfrew, I.A., Dorling, S.R., Price, J.D. & Boutle, I.A. (2021) Sub-km scale numerical weather prediction model simulations of radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 147, 746–763.

Stevens, B. & Seifert, A. (2008) Understanding macrophysical outcomes of microphysical choices in simulations of shallow cumulus convection. *Journal of the Meteorological Society of Japan*, 86, 143–162.

Sullivan, P.P. & Patton, E.G. (2011) The effect of mesh resolution on convective boundary layer statistics and structures generated by large-eddy simulation. *Journal of the Atmospheric Sciences*, 68, 2395–2415.

Tijhuis, M., van Stratum, B.J.H. & Heerwaarden, C.C. (2024) The impact of coupled 3D shortwave radiative transfer on surface radiation and cumulus clouds over land. *Atmospheric Chemistry and Physics*, 24, 10567–10582. Available from: https://doi.org/10.5194/acp-24-10567-2024

Valkonen, T., Stoll, P., Batrak, Y., Køltzow, M., Schneider, T.M., Stigter, E.E. et al. (2020) Evaluation of a sub-kilometre NWP system in an arctic fjord-valley system in winter. *Tellus. Series A, Dynamic Meteorology and Oceanography*, 72, 1–21.

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

van Stratum, B.J.H., van Heerwaarden, C.C. & Vilà-Guerau de Arellano, J. (2023) The benefits and challenges of downscaling a global reanalysis with doubly-periodic large-eddy simulations. *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003750.

Veerman, M.A., van Stratum, B.J.H. & Heerwaarden, C.C. (2022) A case study of cumulus convection over land in cloud-resolving

simulations with a coupled ray tracer. *Geophysical Research Letters*, 49, e2022GL100808. Available from: https://doi.org/10.1029/2022GL100808

Vionnet, V., Bélair, S., Girard, C. & Plante, A. (2015) Wintertime subkilometer numerical forecasts of near-surface variables in the Canadian rocky mountains. *Monthly Weather Review*, 143, 666–686.

Wan, H., Giorgetta, M.A., Zängl, G., Restelli, M., Majewski, D., Bonaventura, L. et al. (2013) The ICON-1.2 hydrostatic atmospheric dynamical core on triangular grids – Part 1: formulation and performance of the baseline version. *Geoscientific Model Development*, 6, 735–763.

Weisman, M.L., Skamarock, W.C. & Klemp, J.B. (1997) The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125, 527–548.

Wyngaard, J.C. (2004) Toward numerical modeling in the "terra incognita". *Journal Of The Atmospheric Sciences*, 61, 1816–1826.

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

Zhang, X., Bao, J.-W., Chen, B. & Grell, E.D. (2018) A three-dimensional scale-adaptive turbulent kinetic energy scheme in the WRF-ARW model. *Monthly Weather Review*, 146, 2023–2045.

How to cite this article: Sakradzija, M., Ahlgrimm, M., Beyrich, F., Heerwaarden, C., Päschke, E., Görsdorf, U. *et al*. (2025) Where numerical weather prediction meets large-eddy simulations in a convective boundary layer during the FESSTVaL field experiment. *Quarterly Journal of the Royal Meteorological Society*, e70037. Available from: https://doi.org/10.1002/qj.70037