From Single Storms to Large-Scale Waves: A Multi-Year Kilometer-Scale Global Simulation

Andreas F. Prein[1], Praveen K. Pothapakula[1], Christian Zeman[1], Morgane Lalonde[1], Marius Rixen[1], Anurag Dipankar[2], Matthieu Leclair[2], and Andreas Jocksch[3]

[1]Institute for Atmospheric and Climate Science, ETH Zürich, 8092 Zurich, Switzerland
[2]Center for Climate Systems Modeling (C2SM), ETH Zürich, Zurich, Switzerland
[3]Swiss National Supercomputing Centre (CSCS), ETH Zürich, Lugano, Switzerland

Correspondence: Andreas F. Prein (aprein@ethz.ch)

Abstract. Global kilometer-scale (km-scale) weather and climate models offer new opportunities to unify numerical weather prediction (NWP) and climate modeling by explicitly simulating convection and mesoscale circulations globally within a single modeling framework. We present results from the first multi-year (April 2020–March 2024) global atmosphere-land simulation using the GPU-refactored ICON model at a 2.5 km horizontal grid spacing and 120 vertical levels. The simulation uses NWP physics and observed sea-surface temperatures. We assess its performance against satellite, reanalysis, and in-situ observations using standard statistics and the MOAAP feature-tracking framework to evaluate a wide spectrum of atmospheric phenomena. ICON reproduces global temperature and precipitation patterns, including a realistic single Intertropical Convergence Zone and physically consistent diurnal precipitation cycles. However, ICON exhibits continental summertime warm and dry biases, linked to an overestimation of incoming solar radiation and excessive surface sensible heat fluxes. The model realistically captures the intensity and frequency of hourly precipitation and near-surface winds, as well as the structure and occurrence of tropical cyclones. Mesoscale convective systems (MCSs) exhibit realistic spatial initiation patterns, but their frequency is underestimated over oceans and overestimated over tropical land. Long-lived MCSs are too infrequent and small, while increased rainfall from shallow and mid-level clouds may reflect overactive warm-cloud microphysics or observational deficiencies. These biases might stem in part from a misrepresentation of thermodynamic-convection coupling. Our results demonstrate the feasibility and scientific value of multi-year global km-scale simulations for exploring the weather–climate system and local-scale extreme events, while identifying key directions for future model development.

1 Introduction

The consequences of anthropogenic climate change are increasingly visible through record-breaking extreme events and their associated impacts (e.g., Donat et al., 2016; IPCC, 2023; Seneviratne et al., 2021; Perkins-Kirkpatrick and Lewis, 2020). Reliable, local-scale climate information to support adaptation and mitigation efforts is urgently needed, as many extremes and their impacts occur at spatial scales that are poorly represented in state-of-the-art global climate models (Giorgi and Gutowski Jr, 2015; IPCC, 2023). Kilometer-scale (km-scale) weather and climate models, with horizontal grid spacings of a few kilometers, have emerged as promising tools to bridge this gap by explicitly resolving deep convection, orographic flows, and other mesoscale phenomena (Prein et al., 2015; Kendon et al., 2017; Lucas-Picher et al., 2021). However, the computational cost of km-scale modeling has historically restricted its use to limited-area domains over relatively small regions. Recent advances in model development, numerical efficiency, and high-performance computing now enable month- to multi-year-long km-scale simulations to be conducted globally (Stevens et al., 2019; Schär et al., 2020; World Climate Research Programme, 2022), opening unprecedented opportunities to study mesoscale processes and their interactions with synoptic- and planetary-scale circulations within a physically consistent framework.

Regional km-scale climate modeling has rapidly evolved over the past decade and has now reached a mature stage of scientific and technical development (Prein et al., 2015; Ban et al., 2014; Kendon et al., 2021; Marsham et al., 2013). A central research focus has been to identify the added value of km-scale simulations compared to coarser, convection-parameterized models. Numerous studies have demonstrated that the intermittency, intensity, frequency, and phase of precipitation are substantially improved when deep convection is explicitly resolved (e.g., Holloway et al., 2012; Prein et al., 2013; Ban et al., 2014; Kendon et al., 2017). In particular, sub-daily precipitation extremes are much better captured (e.g., Ban et al., 2014; Prein et al., 2015, 2017; Ban et al., 2021; Pichelli et al., 2021) and the diurnal cycle of convective precipitation is more realistically represented (e.g., Sato et al., 2009; Marsham et al., 2013; Ban et al., 2014; Prein et al., 2013). Furthermore, the spatial organization of convection, including mesoscale convective systems and tropical cyclones, is substantially improved in km-scale models (e.g., Miura et al., 2007a; Gentry and Lackmann, 2010; Clark et al., 2016; Prein et al., 2020, 2021, 2022; Gutmann et al., 2018). Consistent with the explicit treatment of deep convection, km-scale simulations also produce convective wind gusts and straight-line winds that are absent in models with parameterized convection (Prein, 2023; Brown et al., 2024). Improved representation of topography enhances the simulation of orographic precipitation, mountain snow accumulation, and valley wind systems (Liu et al., 2017; Ikeda et al., 2021; Schmidli et al., 2018). Additionally, km-scale grids allow for a more faithful representation of land–atmosphere coupling (Hohenegger et al., 2009; Lee and Hohenegger, 2024; Segura et al., 2022), and lateral groundwater fluxes can be represented explicitly, which can feed back on soil moisture and precipitation characteristics (Schlemmer et al., 2018; Barlage et al., 2021). At km-scale resolution, urban areas can be better represented, enabling more realistic simulations of urban heat islands and their influence on local convection and precipitation (Argüeso et al., 2016; Langendijk et al., 2021). However, dedicated land surface model development efforts are needed to include processes that are typically missing (e.g., shallow groundwater flow, urban parameterizations).

While many of the mesoscale phenomena examined here have already been studied with regional convection-permitting models, the added value of the present study is that it evaluates them within a single, consistent, multi-year, global km-scale framework, enabling direct comparisons across regions and assessments of their links to larger-scale circulation features.

The field of global km-scale modeling was pioneered by the Japanese community (Tomita et al., 2005; Miura et al., 2007b; Satoh et al., 2008) through the development of the Non-hydrostatic ICosahedral Atmospheric Model (NICAM). Over the last two decades, almost every major model-development center has invested in developing next-generation, non-hydrostatic global modeling capabilities. At the same time, advances in computer technology made it feasible to run global km-scale simulations for more than just a few days. The DYAMOND (DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains) intercomparison experiments leveraged these developments and initiated the first global, multi-model, km-scale coordinated intercomparison of simulations of the atmosphere at grid spacings of 3 km–5 km (Stevens et al., 2019; Satoh et al., 2019; Takasuka et al., 2024b). These experiments revealed substantial inter-model differences in the ability to simulate mesoscale convective systems (MCSs) (Feng et al., 2025) and tropical cyclones (Judt et al., 2021). Some modeling systems demonstrated notable progress following dedicated model development efforts. For instance, the U.S. Department of Energy's SCREAM (Simple Cloud-Resolving E3SM Atmosphere Model) initially underestimated convective organization during the first DYAMOND phase but demonstrated marked improvements in MCS and tropical-cyclone structure in DYAMOND-Winter after major revisions to microphysics and numerics (Taylor et al., 2023; Donahue et al., 2024; Feng et al., 2025). Beyond convective systems, global km-scale simulations exhibit more realistic representations of equatorial waves and tropical variability compared to coarser-resolution models (Miura et al., 2007a; Nasuno et al., 2008; Holloway et al., 2012; Weber and Mass, 2019; Judt and Rios-Berrios, 2021). Ongoing developments extend these atmospheric configurations toward multidecadal coupled ocean-atmosphere-land models that include interactive ocean, sea ice, and biogeochemical components (Hohenegger et al., 2023; Segura et al., 2025b; Rackow et al., 2025) as well as the carbon cycle and aerosol emissions (Klocke et al., 2025).

Here, we present an overview of the performance of a global simulation with a 2.5 km horizontal grid spacing spanning four consecutive years, providing a comprehensive assessment of Earth's hydroclimate and associated extremes at km-scales. This simulation represents a major step toward bridging the traditional gap between numerical weather prediction and climate modeling (Miura et al., 2023; Randall and Emanuel, 2024) by explicitly resolving convective processes and mesoscale storm systems within a global, multi-year integration. We employ a graphics processing unit (GPU)–refactored version of the ICON (Icosahedral Nonhydrostatic) model (Zängl et al., 2015; Dipankar et al., 2026) using the physical parameterizations from its operational Numerical Weather Prediction configuration. Such global storm-resolving simulations provide a unified framework for studying the dynamics and statistics of phenomena ranging from single storms to global waves, including tropical cyclones, mesoscale convective systems, and equatorial waves, and for examining their role in shaping large-scale hydroclimatic variability and extremes. We use a model setup that is closely aligned with ICON settings used in numerical weather prediction, which differs substantially in the used physics (i.g, land surface model, surface layer scheme, boundary layer scheme, microphysics, and radiation) compared to previously performed global km-scale ICON simulations (e.g., Segura et al., 2025a). The primary objectives of this study are threefold: (i) to document the experimental design and technical aspects of the four-year simulation, (ii) to evaluate its skill in reproducing climate mean states and key mesoscale phenomena and their contribution to global hydroclimate statistics, and (iii) to identify systematic model deficiencies that can inform future model development. The resulting dataset, which will be openly released through the DYAMOND-III intercomparison initiative (Takasuka et al., 2024b), provides a novel resource for investigating the physical processes governing the Earth's hydroclimate in a high-resolution, fully global context.

2 Data and Methods

2.1 Modeling

With recent advances in exascale high-performance computing (HPC) architectures, the computational throughput of various weather and climate models has substantially improved over the last decade. Especially energy-efficient heterogeneous architectures, leveraging GPU accelerators, have been developed and can make climate model integrations very efficient (Schär et al., 2020). However, substantial investments in enabling km-scale weather and climate model codes are needed to port them and run them efficiently on new HPC systems (Giorgetta et al., 2022; Donahue et al., 2024; Dipankar et al., 2026). Rapidly emerging technologies in hybrid architectures require agile, architecture-agnostic climate model codes. The EXCLAIM (EXtreme scale Computing and data platform for cLoud-resolving weAther and clImate Modeling) project at ETH Zürich addresses this issue by refactoring the ICON code with Domain Specific Language (DSL) with a Python front-end using GridTools for Python, GT4Py, while enabling backend flexibility across architectures (Paredes et al., 2023; Dipankar et al., 2026).

We use the version of ICON Model developed within the EXCLAIM project (Dipankar et al., 2025), which features a complete refactoring of the numerical core into GT4Py (icon-exclaim v0.2.0)(Dipankar, 2025). Two global experiments were conducted on the Swiss National Computing Center's new HPC infrastructure, ALPS. A ten-year (2006–2016) spin-up simulation was first performed at a 10 km horizontal grid spacing. A 2.5 km grid spacing simulation following the DYAMOND protocol (Takasuka et al., 2024b) is conducted for a span of 4 years (January 20, 2020, to April 1, 2024), with its soil state taken from the 10 km course grid simulation integrated for 10 years (Pothapakula et al., 2026). The 2.5 km simulation uses 240 GH200 GPU nodes (960 GPUs in total), achieving a throughput of 0.25 simulated years per day. This setting was chosen for computational efficiency, and throughput can be increased by increasing the number of nodes. More information on the computational aspects of the specific model used for this simulation can be found in Dipankar et al. (2026).

The 2.5 km simulation uses lower boundary conditions from daily updated sea surface temperatures and the sea-ice fraction from the ESA Climate Change Initiative (ESA-CCI) product at a horizontal grid spacing of $1/20°$ (Merchant et al., 2019). Furthermore, the setup includes 120 vertical levels extending up to 85 km with a terrain-following hybrid setup and the smooth level vertical (SLEVE) coordinate (Schär et al., 2002). The single-moment cloud Seifert (2008) bulk microphysics scheme is used to simulate cloud water, ice, snow, graupel, and rain. The second-order turbulent kinetic energy-based surface transfer and planetary boundary layer parameterization are used to represent turbulence (Raschendorfer et al., 2003). Land processes are simulated using the TERRA Multi Layer (TERRA-ML) model (Schrodin and Heise, 2001; Grasselt et al., 2008; Schulz and Vogel, 2020), which employs eight soil layers and governs the exchange of heat, moisture, and momentum. The TERRA-ML model employs a tile approach to accurately estimate cell-averaged surface fluxes and account for deviations in sub-grid-scale surface characteristics. The shallow and deep convection schemes are turned off in the 2.5 km simulation, allowing an explicit treatment of convection. Additionally, subgrid-scale orographic drag and non-orographic gravity wave drag schemes are disabled. We used a numerical time step of 20 seconds, updated cloud cover and the microphysics schemes every time step, and applied the radiation scheme every 6 minutes. The external parameters required by ICON, namely the topographic height of the Earth's surface, surface roughness, vegetation cover, and land/sea/lake distribution, were prepared using the External

Parameters for Numerical Weather Prediction and Climate Application software tool (Asensio et al., 2020). Finally, we use the global Max-Planck-Institute Aerosol Climatology version 2 (MAC-v2) dataset with 1 ° grid spacing at a monthly temporal frequency (Kinne, 2019).

The 2.5 km model configuration is close to the ICON-CLM (Climate Limited-area Modelling) community setup and benefits from the vast experience of the community‘s participation in CORDEX (COordinated Regional Climate Downscaling EXperiment) activities (Giorgi et al., 2009) across multiple CORDEX regional domains. We did not perform any additional model tuning for this simulation. We saved output variables following the DYAMOND III protocol (Takasuka et al., 2024b).

2.2 Observational Datasets

We use a range of observational datasets alongside reanalysis to assess our simulation’s performance.

2.2.1 ERA5

The ERA5 reanalysis is the fifth-generation global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It provides hourly estimates of a large number of atmospheric, land-surface, and ocean-wave variables at a horizontal resolution of about 31 km, spanning from 1940 to the present. ERA5 represents a major improvement over its predecessor, ERA-Interim, particularly in spatial and temporal resolution and in uncertainty estimates (Hersbach et al., 2020). We use ERA5 data to evaluate 2 m above surface temperature.

2.2.2 Brightness Temperature

We use two observational products of brightness temperature (Tb) in this study. The NCEP/CPC Level-3 merged infrared Tb dataset (Janowiak et al., 2017) provides high-resolution data for deep convective cloud identification and tropical wave tracking. The NOAA outgoing longwave radiation (OLR) dataset has a much lower resolution (Liebmann and Smith, 1996). It enables the continuous detection of equatorial waves over the simulation period, including slow-evolving variability such as the Madden-Julian Oscillation.

The NCEP/CPC dataset provides gridded equivalent blackbody temperature fields from merged geostationary infrared satellites (Janowiak et al., 2017). It covers latitudes from 60° S to 60° N, with a 4 km horizontal grid and data every 30 minutes. The product merges observations from multiple geostationary platforms (e.g., METEOSAT, GMS/MTSAT/Himawari, GOES) over their operational periods. We use full-hour data to track MCSs, non-MCS precipitating cold clouds, and equatorial waves, and use 30-minute observations to fill gaps in full-hour records.

The NOAA OLR dataset provides twice-daily estimates of top-of-atmosphere infrared emission from polar-orbiting satellites at 2.5° horizontal resolution (Liebmann and Smith, 1996). The dataset spans from 1974 to the present, covers 60° S–60° N, and provides a consistent long-term record of tropical convective variability. In this study, OLR serves as a large-scale diagnostic for equatorial wave activity.

2.2.3 GPM-IMERGv7 Precipitation

The Integrated Multi-satellite Retrievals for GPM (IMERG) dataset is part of the NASA/JAXA Global Precipitation Measurement mission. It combines observations from a constellation of passive microwave and infrared sensors with gauge data to provide global precipitation estimates. We use IMERG Final Run version 7, which offers high spatio-temporal resolution (0.1°, 30-minute) precipitation fields from the year 2000 to the present (Huffman et al., 2020).

2.2.4 HadISD Station Observations

The HadISD (Hadley-Centre Integrated Surface Database) is a global sub-daily station dataset based on the NOAA Integrated Surface Database (ISD), providing quality-controlled observations of key climatological variables such as temperature, dew point, sea-level pressure, wind speed and direction, and cloud data at individual stations (Dunn et al., 2016; Dunn, 2024). The dataset currently spans from 1931 to the present and has undergone extensive automated quality control to detect and remove erroneous values. However, it is not homogenised, and its station density varies over time. We use hourly precipitation and 10 m above surface wind speed records from HadISD v3.4.1 for model validation. Only stations with more than 50 % valid data coverage during the evaluation period are considered. The location/density of the included station network is shown in the relevant sections. Station data is compared to data from the nearest ICON grid cell on the native model grid.

2.2.5 Surface fluxes: Incoming shortwave and longwave radiation, and latent and sensible heat

We utilize data from 430 flux towers in our model evaluation, sourced from two datasets: FLUXNET (Pastorello et al., 2020) and AmeriFlux (Chu et al., 2023). These towers provide 30-minute measurements of shortwave and longwave radiation, as well as surface latent and sensible heat fluxes. Data from these towers is available from 1991 to 2024, with an average record length of approximately 7 years. The sites are distributed across 43 countries, with denser coverage in North America, Europe, eastern Asia, and Australia, and sparser representation in Africa and western Asia. It is important to note that inter-annual variability in surface fluxes contributes to differences between simulations and observations. For reference, across the 430 observational stations the inter-annual variability, quantified as the standard deviation of annual means, has median values of 5.2 $\mathrm{W\,m^{-2}}$ for SWin, 3.5 $\mathrm{W\,m^{-2}}$ for LWin, 6.7 $\mathrm{W\,m^{-2}}$ for LE, and 5.4 $\mathrm{W\,m^{-2}}$ for H, with interquartile ranges of 3.7–6.8, 2.4–4.6, 4.0–10.5, and 3.4–8.7 $\mathrm{W\,m^{-2}}$, respectively.

Additionally, we use gridded surface-downwelling shortwave radiation data from the EWEMBI (EartH2Observe, WFDEI and ERA-M-B-Ias-corrected) meteorological dataset (lan, 2016; Lange, 2018) during the period 2012–2016. EWEMBI provides globally gridded data at 0.5 ° spatial and daily temporal resolution (calculated as 24-hour averages of hourly fields derived from 3-hourly fluxes). It combines EartH2Observe data (Schellekens et al., 2017), which uses WFDEI (Weedon et al., 2014) over land and ERA-Interim over the ocean (Dee et al., 2011), with radiation fields bias-corrected using Surface Radiation Budget (SRB) observations data released by The GEWEX SRB Project at NASA Langley Research Center (LaRC) (Trenberth, 2011) through daily-scale quantile mapping. Although the EWEMBI period (2012–2016) differs from the simulation period, we include it in the evaluation because interannual variability in surface shortwave radiation is small. The inter-annual standard deviation of the global mean surface-downwelling shortwave radiation is 0.3 W m^{-}2, and the distribution of grid-cell inter-annual variability shows a mean of 4.3 W m^{-}2 (IQR: 2.4 W m^{-}2–5.0 W m^{-}2). Differences exceeding these ranges can therefore be interpreted as significant rather than reflecting natural year-to-year variability.

2.2.6 IBTrACS Tropical Cyclone Characteristics

The IBTrACS (International Best Track Archive for Climate Stewardship) dataset is a comprehensive global compilation of tropical cyclone "best track" data, aggregated from multiple official agencies across all ocean basins. It provides information on storm position, intensity (e.g., maximum sustained winds and/or minimum central pressure), and other parameters, typically at 3-hour or 6-hour intervals, spanning from the mid-19th century to the present (Knapp et al., 2010; Gahtan et al., 2024). The project also records the original reports from contributing agencies and summary statistics (such as the mean track and range across agencies) to facilitate intercomparison. We use U.S. records from IBTrACS v4.01 (specifically the variables 'usa_wind' and 'usa_pres') because they are the most up-to-date and include information on recent global tropical cyclone activity. Recorded speed correspond to maximum sustained wind averaged over 1 minute at 10 m above surface.

2.3 Methods

2.3.1 Atmospheric feature tracking with the MOAAP algorithm

The Multi-Object Analysis of Atmospheric Phenomenon (MOAAP; Prein et al. (2023b)) algorithm is an automated, object-based framework that identifies and tracks a wide range of atmospheric systems using a minimal set of commonly available reanalysis and model variables. MOAAP combines 12 key atmospheric variables from observations (here, ERA5 reanalysis, GPM IMERG precipitation, and GPM_MERGIR cloud-top Tb) and the ICON model to consistently detect and track features such as tropical and extratropical cyclones, cut-off lows, anticyclones, frontal zones, jet streams, moisture streams, mesoscale convective systems (MCSs), and equatorial waves.

Feature detection is based on a connected-component approach applied in space and time. The algorithm first thresholds relevant variables to form binary masks of regions that exceed physically motivated intensity thresholds. It then labels contiguous three-dimensional (latitude–longitude–time) structures using the `ndimage.label` function from SciPy (Virtanen et al., 2020). In a second step, features are segmented based on the presence of multiple minima/maxima within an object using the `skimage.segmentation` and `skimage.feature` functionalities, with watershed segmentation applied. Phenomena are then identified according to temporal persistence rules and phenomenon-specific geometric, temporal, and intensity thresholds. Each identified object is assigned physical characteristics such as area, duration, maximum intensity, and center of mass, which enable subsequent statistical and climatological analyses. MOAAP provides a unified, reproducible framework for quantifying and intercomparing the contributions of different atmospheric phenomena to mean and extreme precipitation across spatial and temporal scales.

The MCS analysis in section 3.7 is performed on a global 0.1 ° regular grid that corresponds to the GPM-IMERG grid. GPM-MERGIR and ICON data are conservatively regridded to this grid. MCSs are defined as in Prein et al. (2024) following four criteria: 1) Continuous Tb $\leq$ 241 K area $\geq$ 40,000 km^2 for $\geq$ 4 hours; 2) maximum hourly precipitation beneath the Tb $\leq$ 241 K area $>$ 10 mm h^{-1} for $\geq$ 4 hours; 3) hourly precipitation volume > 20,000 km^2 mm h^{-1} at least once during the MCS lifetime; 4) minimum Tb $<$ 225 K at least once during the MCS lifetime.

Additionally, the hourly heavy precipitation analysis in section 3.9 is performed on a standard 0.25 ° mesh that corresponds to the ERA5 grid. The only phenomenon not included in MOAAP that we were unable to identify is atmospheric rivers, as ICON lacks an IVT diagnostic. More details about MOAAP can be found in (Prein et al., 2023b, 2024; Feng et al., 2025).

2.3.2 Deriving central pressure, maximum wind, and pressure–wind pairs

We use tropical cyclones (TCs) identified by MOAAP to calculate TC intensity centered on the TC sea level pressure (SLP) minima at each time step (hour). Distances and azimuths are computed using great-circle geometry, and land-sea ocean points are defined via a land-sea mask with a 30 km coastal buffer. The maximum sustained wind $V_{\max}$ is the peak 10-m wind speed from hourly instantaneous model output within the inner-core disk given by the Reference Disk Radius (RDR), defined as

$$\mathrm{RDR} = \min\bigl(\mathrm{RMW}, r_{\partial p}\bigr),$$

where $\mathrm{RMW} = \arg\max_r \overline{|u_t|}(r)$ is the radius of maximum mean tangential wind and $r_{\partial p} = \arg\max_r \bigl|\partial\overline{p}/\partial r\bigr|$ is the radius of strongest radial SLP gradient (Knaff and Zehr, 2007; Schenkel and Hart, 2012).

3 Results

We start this section with a climatological analysis of near-surface temperature, surface fluxes, and precipitation, followed by assessing higher-order statistics of hourly precipitation and surface wind speed, and ending with evaluating the fidelity of simulating atmospheric storms such as TCs, MCSs, and extreme precipitation-producing systems.

3.1 2 m above surface temperature (T2M)

The global average pattern of T2M is well simulated with a Spearman pattern correlation coefficient of 0.998 and a root mean squared error of 1.48 °C for land grid points (Fig. 1). As expected, ocean regions exhibit particularly small differences due to the pre-described sea surface temperature. Striking are cold differences in ICON during polar winters that might be related to the representation of sea ice and snow in addition to differences stemming from uncertainties in ERA5 T2M in these regions (Fig. 1d,f). Larger differences also occur during boreal winters over land areas with up to 3 °C negative differences in eastern Canada and large positive differences in Siberia and Alaska.

During boreal summers, temperatures exceeding 3 °C warm differences emerge in continental regions such as the central U.S. and Eurasia (Fig. 1f). Part of these differences is likely related to a misrepresentation of land-atmosphere interactions, such as the lack of lateral groundwater flow and the corresponding reduced evapotranspiration in land-atmosphere coupling hot-spots such as the central U.S., which can cool down these regions by several degrees Celsius during the summer season (Barlage et al., 2021; Schlemmer et al., 2018). Additionally, misrepresented cloud radiative forcings frequently (Fig. S1) contribute

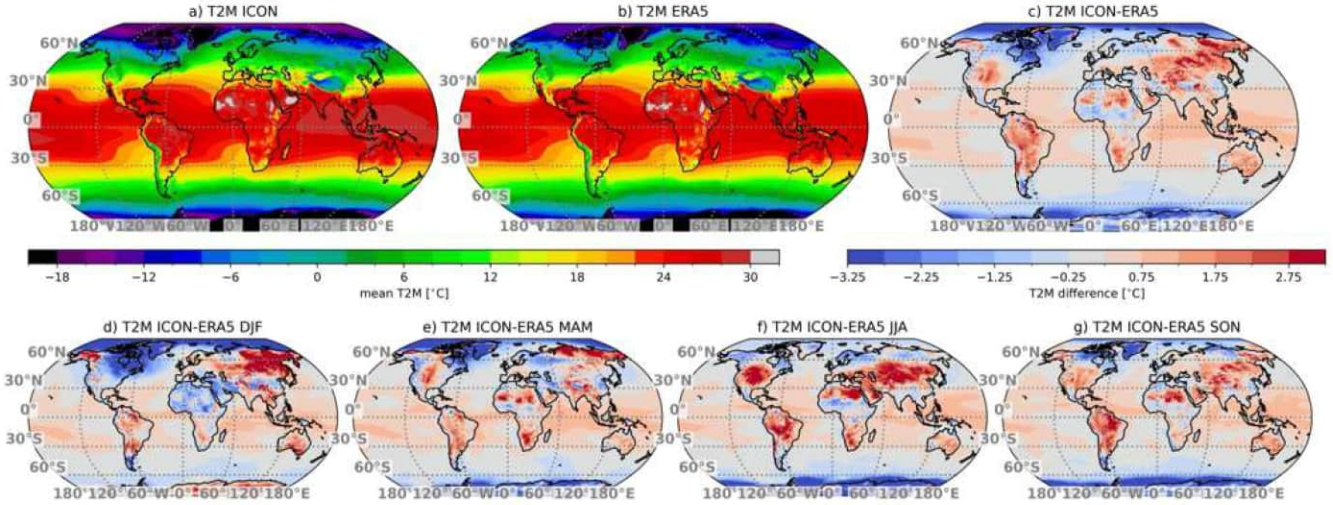

Figure 1. Annual mean daily T2M in ICON (a), ERA5 (b), and their difference (c) during the simulation period. Seasonal average T2M differences are shown in (d–g).

to surface temperature biases in models with convection parameterizations (e.g., Ahlgrimm and Forbes, 2012) and km-scale models (e.g., Sakradzija et al., 2020; Lucas-Picher et al., 2024).

It is important to note that the differences shown here and in the following sections are partly related to the short simulation period of four years. Climate internal variability can cause large differences at local scales (Deser et al., 2012) even though observed sea surface temperatures are used.

3.2 Surface fluxes: Incoming shortwave and longwave radiation, and latent and sensible heat

Fig. 2 shows the daily mean surface incoming shortwave radiation (SWin) for (a) ICON, (b) the EWEMBI observational dataset, and (c) their difference. Overall, SWin is underestimated over oceans, especially along the Intertropical Convergence Zone, and overestimated over land, particularly in mountainous regions. This overestimation over land is also evident when comparing the simulation to flux tower observations (Fig. 2d,e). The median difference in land mean SWin is approximately $6.8\,\mathrm{W\,m^{-2}}$, with a median difference of around $28.2\,\mathrm{W\,m^{-2}}$. The overall median difference is consistent with the mean bias error of approximately $7.8\,\mathrm{W\,m^{-2}}$ reported for CMIP6 models by He et al. (2023). Previous studies have shown that shallow clouds are frequently a main contributor to continental surface energy biases in km-scale models (e.g., Sakradzija et al., 2020; Lucas-Picher et al., 2024).

Fig. 2d and e show the spatial distribution of station median total and noon differences in longwave downward radiation (LWin), latent heat flux (LE), and sensible heat flux (H). LWin is slightly underestimated. LE shows a median difference close to zero across stations, with regionally over- and underestimations of more than $10\,\mathrm{W\,m^{-2}}$ that cancel each other on average. This spatial compensation suggests region-dependent limitations in the representation of land-surface controls on evaporation,

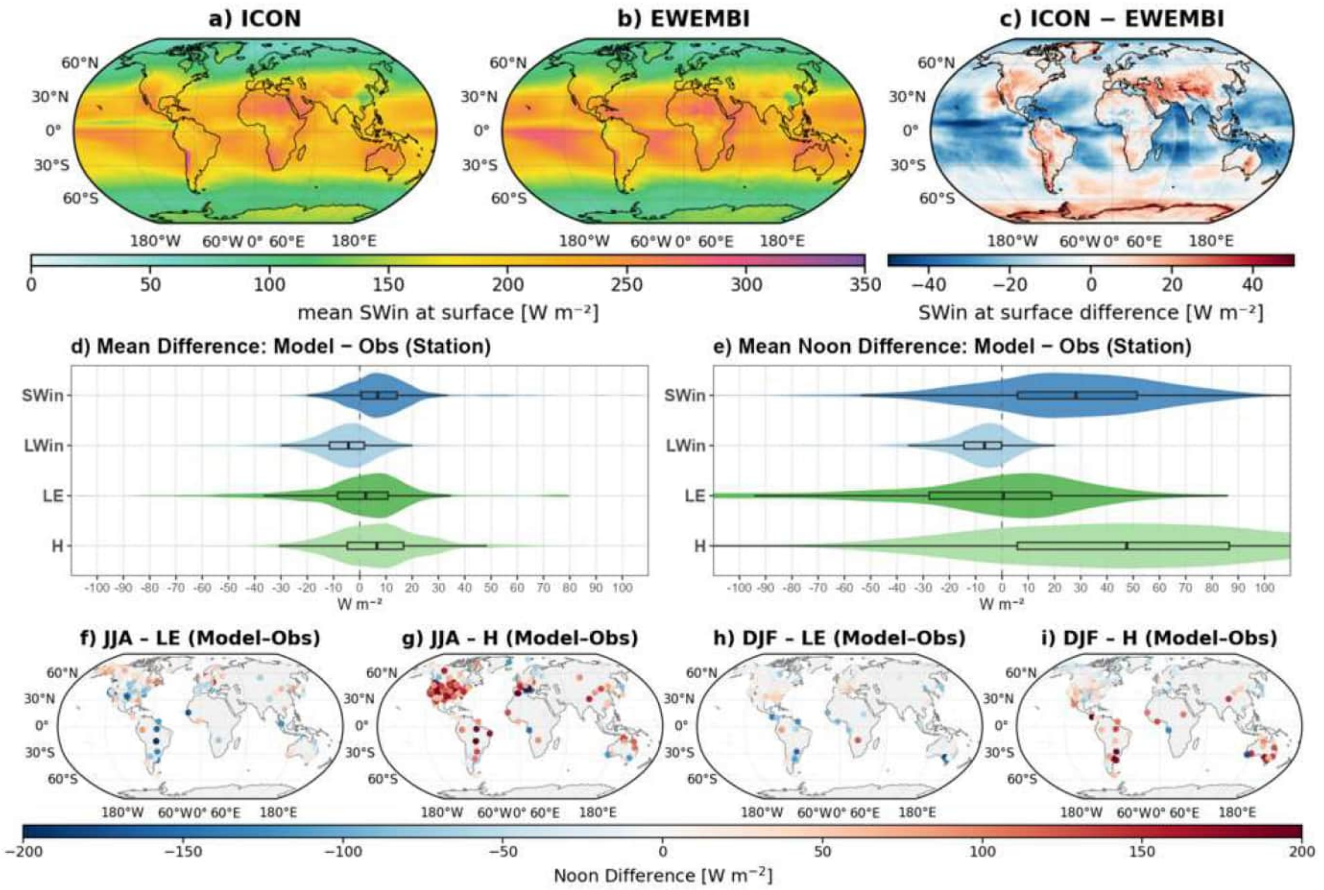

Figure 2. (a–c) Daily mean incoming shortwave radiation at the surface (SWin) from (a) ICON (2020-2024), (b) the EWEMBI (2012-2016) observational dataset, and (c) their difference (ICON - EWEMBI). (d–e) Violin plots of station-based differences (model - observations) for mean (d) and mean local solar time noon (e) surface fluxes: SWin, incoming longwave radiation (LWin), latent heat flux (LE), and sensible heat flux (H). Boxes indicate the interquartile range and median difference. (f–i) Station-based mean noon differences (model - observations) for LE and H during JJA (June, July, August) and DJF (December, January, February).

rather than a uniform model bias. In contrast, H is more consistently overestimated, particularly during noon, agreeing with the overestimation in SWin.

The spatial pattern of station-based mean noon differences in LE and H during JJA shows a pronounced underestimation of LE and an overestimation of H at most stations in ICON (Fig. 2f,h) in agreement with the T2M differences shown in Fig. 1. We suspect that parts of the systematic dry-summer biases are due to the absence of a lateral subsurface water-flux scheme in TERRA. Previous studies have shown that the lack of such lateral groundwater redistribution can lead to substantial surface flux and temperature biases at km-scale grid spacing (Barlage et al., 2021; Soares et al., 2024). In contrast, there is more

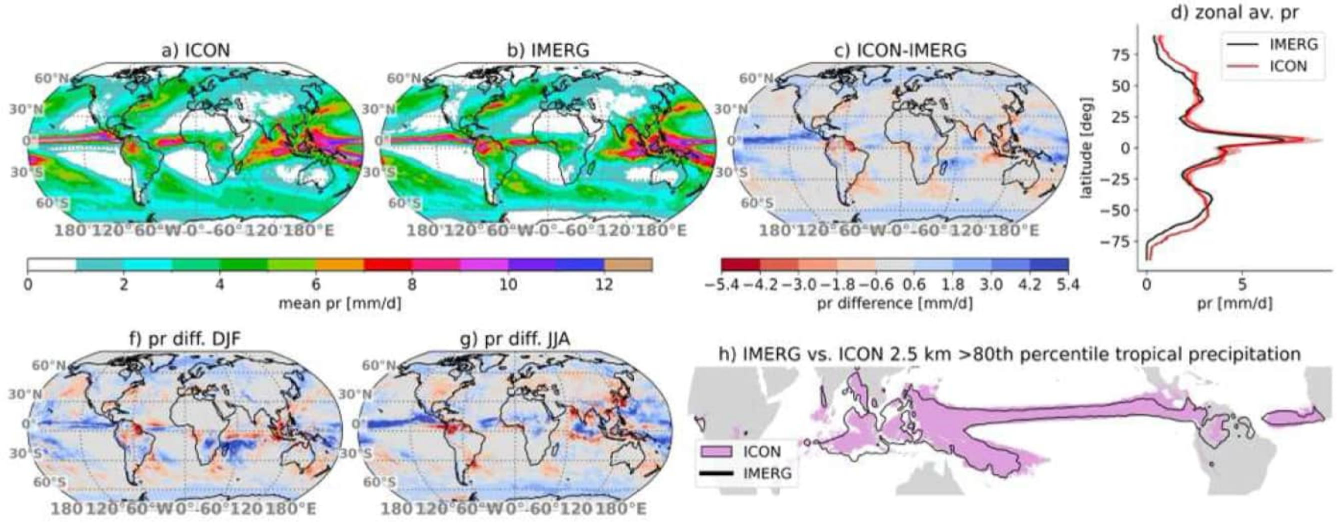

Figure 3. Annual mean daily precipitation in ICON (a), IMERG (b), and their difference (c) during the simulation period. DJF and JJA average precipitation differences are shown in (f) and (g), respectively. Mean zonal average precipitation is shown in (d) with inter-annual spread shown in contours. The area that covers the 80th percentile of tropical average precipitation ($\pm 20^\circ$) from ICON (purple) and IMERG (black contour) is shown in (h).

station-to-station variability for winter biases. The spatial patterns and characteristics of surface flux differences are consistent with biases in temperature and precipitation.

3.3 Precipitation (PR)

Annual and seasonal differences between ICON and GPM-IMERG are shown in Fig. 3. The Pearson pattern correlation of annual average precipitation is 0.91 (0.87 over land cells) and the corresponding relative RMSE is 41 %. Most noticeable are wet differences over most tropical ocean regions except for the west Maritime Continent (Fig. 3c). This is consistent with the negative SWin differences shown in Fig. 2c. Land precipitation is relatively well simulated overall, although dry differences emerge in several continental regions during boreal summer (Fig. 3g). In some of these regions, including parts of the central U.S. and Eurasia, the dry differences coincide with warm T2M differences (Fig. 1f), higher SWin, lower LE, and higher H, suggesting a contribution from land-atmosphere coupling and cloud radiation biases.

Zonal average precipitation agrees very well between GPM-IMERG and ICON, partly because of error cancellation effects (Fig. 3d). The differences are typically smaller than the inter-annual variability (shading in Fig. 3d) except for high-latitudes where IMERG is known to underestimate precipitation accumulations due to issues in remote sensing solid precipitation (Song et al., 2021) complicating a reliable evaluation. The location of the wet regions in the tropics (defined as the 80th percentile of tropical average precipitation; Fig. 3h) is remarkably well captured with a single tropical rainband.

This is a big achievement since simulating a realistic rainband is challenging, even at a km grid spacing, as Segura et al. (2022, 2025a, b) found a double ITCZ in the Indo-Pacific using km-scale ICON simulations with model physics different from ours. They found a lack a surface fluxes (caused by too little surface winds) in this region, which they resolved by increasing the minimum surface wind speed in the lowest model level from $1\,\mathrm{m\,s^{-1}}$ to $4\,\mathrm{m\,s^{-1}}$ (Segura et al., 2025a). Our model settings, which are similar to ICON‘s NWP setup, do not feature this issue. While it is unclear exactly what causes the differences in model performance among the ICON setups, a key difference lies in the turbulence parameterizations and surface-layer schemes, while Takasuka et al. (2024b) found large sensitivities to the formulation of microphysics. The location of mid-latitude precipitation bands associated with the storm tracks is reasonably well simulated (Fig. 3a,b).

Comparing peak hourly precipitation over the four-year evaluation period between GPM-IMERG and ICON shows that our simulation yields substantially larger values, particularly in the tropics and subtropics (Fig. 4a,b). This is predominantly caused by too low heavy hourly precipitation in GPM-IMERG since comparisons with hourly precipitation gauge observations from the HadISD dataset show that ICON reliably reproduces observed intense precipitation rates across various climate regions. At the same time, GPM-IMERG fails to capture intense rates, particularly at lower latitudes (Fig. 4c–f), which is in agreement with previous research (Guilloteau and Foufoula-Georgiou, 2020; Dominguez et al., 2024; Yu et al., 2025). However, a proper evaluation of hourly precipitation using station data is not possible in many parts of the globe due to insufficient data coverage and a lack of data sharing (e.g., Prein and Gobiet, 2017). Our presented analysis is primarily based on data from the U.S. and Europe (see Fig. S2). Results for additional climate regions, often based on only a few stations, are shown in Fig. S3.

Using the same station network, we assess the June-July-August (JJA) season-average precipitation diurnal cycle (Fig. 5; we use $0.1\,\mathrm{mm\,h^{-1}}$ as the threshold for precipitation intensity and frequency statistics). The performance of ICON is regionally dependent and shows good agreement for capturing precipitation amount, frequency, and intensity in the North Central America (NCA) region that is heavily influenced by the North American Monsoon (Fig. 5g–i). In Eastern North America (ENA), precipitation frequency is well captured compared to station data. Still, precipitation intensity is overestimated, resulting in an overly amplified evening and nocturnal peak in precipitation amplitudes (Fig. 5a–c). Western Central Europe (WCE) shows a too-early peak in precipitation that decays too quickly. The precipitation amount is also low-biased due to the too low precipitation frequency. Generally, ICON shows greater fidelity in capturing the diurnal cycle of convective precipitation in southern regions (ENA and NCA) than GPM-IMERG.

From Fig. 5, we see that the peak timing of convective precipitation amount is relatively well captured by GPM-IMERG, consistent with previous findings (e.g., Dominguez et al., 2024) although IMERG has been reported to have a phase lag relative to Ku-band radar in many regions which should be kept in mind (Hayden and Liu, 2021). This allows us to evaluate ICON on a near-global scale (Fig. 6). We perform this analysis during seasons dominated by convective precipitation, namely JJA at $>30\,^{\circ}$N, December-January-February (DJF) at $<30\,^{\circ}$S, and annually in the tropics. GPM-IMERG and ICON feature afternoon precipitation peaks over land. In contrast, tropical ocean regions feature predominantly nighttime and early-morning peaks. Some land regions also feature nocturnal precipitation peaks, particularly in the lee of mountain ranges (e.g., the central U.S., the La Plata basin, eastern China, and the Sahel and Western Africa), which are related to nocturnal mesoscale convective system activity. ICON can capture these nocturnal peaks reasonably well, but the spatial extent of nocturnal regions is often

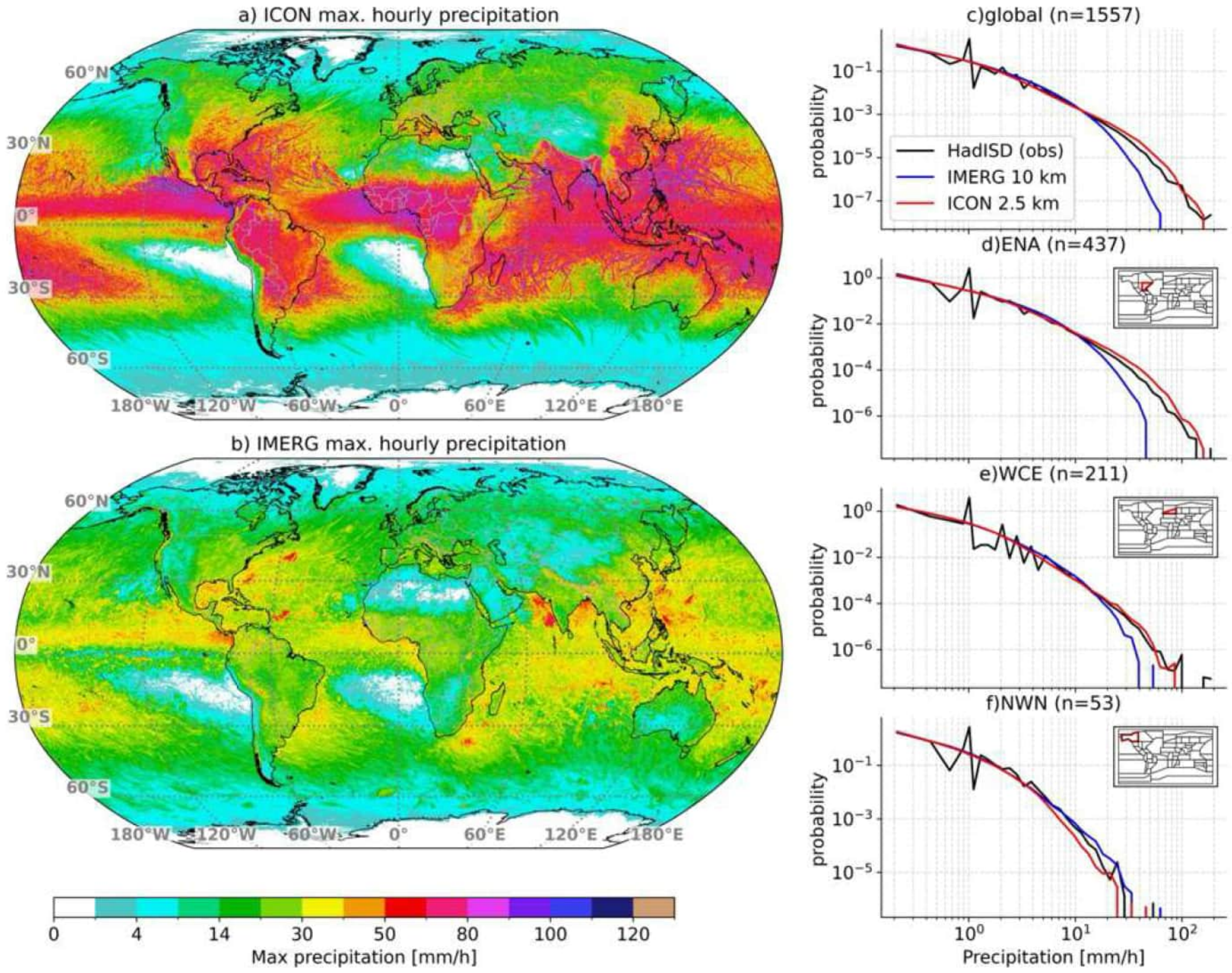

Figure 4. Hourly maximum precipitation accumulation during the simulation period from ICON (a) and IMERG (b) on the 0.1° IMERG grid. Probability density functions for hourly precipitation comparing ICON (red) and IMERG (blue) against HadISD station observations (black) globally (c) and in the Eastern North America (ENA, d), Western Central Europe (WCE, e), and North Western North America (NWN, f) region. The locations of these regions are shown in the inlet maps in (b–f). The underlying station density in the HadISD dataset is shown in Fig. S2.

too small, particularly in the Amazon Basin and the central U.S. The relative amplitude of the convective precipitation diurnal cycle (diurnal amplitude divided by mean precipitation) is also relatively similar, with a tendency for larger values in ICON compared to GPM-IMERG.

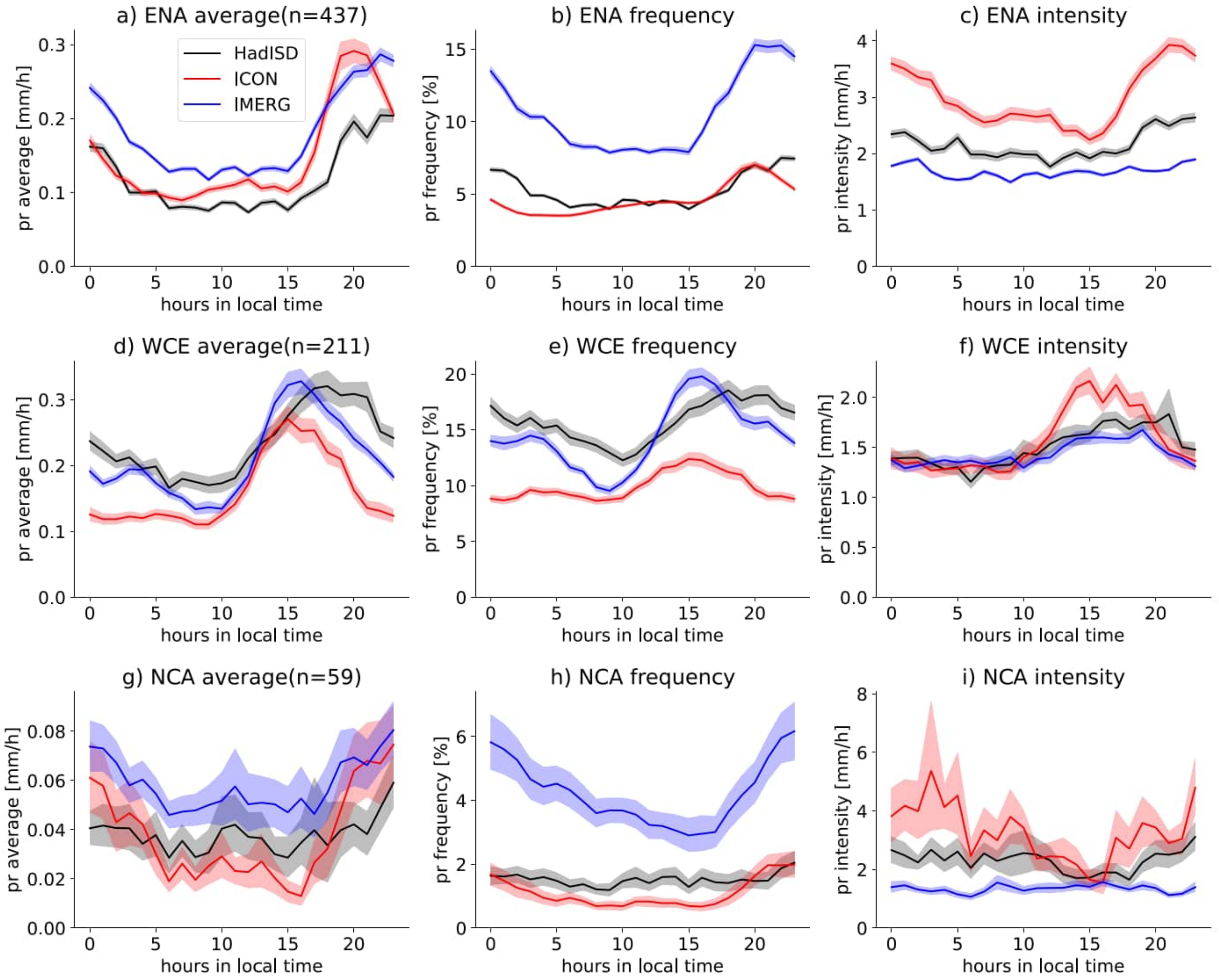

Figure 5. Hourly precipitation diurnal cycle during June, July, and August of average precipitation (a, d, g), precipitation frequency (b,e,h), and precipitation intensity (c,f,i). Results are shown for the Eastern North America (a–c), Western Central Europe (WCE), and North Central America (NCA) region. Precipitation frequency and intensity statistics are based on a precipitation threshold of $\geq 0.1\,\mathrm{mm\,h^{-1}}$. Data are based on HadISD records (black lines) with more than 50 % coverage. Missing values in HadISD are also removed in the ICON (red ines) and IMERG (blue lines) data. Contours show station sampling uncertainties based on the 10th to 90th percentile values from 10,000 bootstrap samples (with replacement).

In summary, we recommend using GPM-IMERGv7 hourly precipitation with caution in model evaluation studies, given its tendency to overestimate precipitation frequency and underestimate precipitation intensity. Where possible, in-situ hourly rain gauge observations should be used.

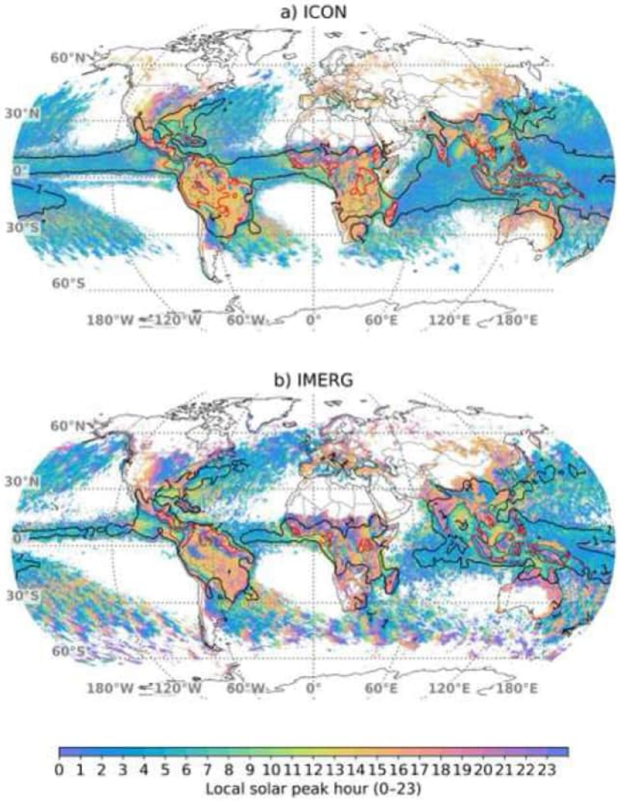

Figure 6. Color-coded local peak time of precipitation diurnal cycle in ICON (a) and IMERG (b). Colored regions show areas where the precipitation diurnal cycle amplitude is larger than 40 % of the mean precipitation. Black/red contours show regions where the amplitude is larger than 100 %/200 % of the mean precipitation. Data from all months is used in the tropics ($\leq\pm30^\circ$), data from June, July, and August is used for latitudes $>30^\circ$, and data from December, January, and February is used for latitudes $<$-30°.

3.4 Near surface winds (UV10)

Tropical cyclones cause the most intense UV10 extremes during the evaluation period (Fig. 7a). Storm tracks of extratropical cyclones are also clearly visible in both hemispheres. Over land, mountain regions and plateaus such as the Tibetan Plateau and the Andes stand out, as do mid-latitude plains that feature straight-line windstorms (e.g., the central U.S., the La Plata Basin, and the Sahel). Also, the coastal regions of Greenland and Antarctica feature peak wind speeds exceeding 40 $\mathrm{m\,s^{-1}}$.

Evaluations against UV10 HadISD station observations show overall good agreement with a tendency of heavy wind speeds (>20 $\mathrm{m\,s^{-1}}$) to be underestimated in ICON (Fig. 7b). However, regional differences are observed, with lower differences at high wind speeds, mostly in Mid-Latitude regions such as CNA and WCE (Fig. 7c,d). The region with the most pronounced underestimation of UV10 extremes is Eastern North America (Fig. S5), which might be partly related to a low difference in tropical cyclone frequency as we will see in section 3.5, but regional differences in the simulation of extra-tropical cyclones

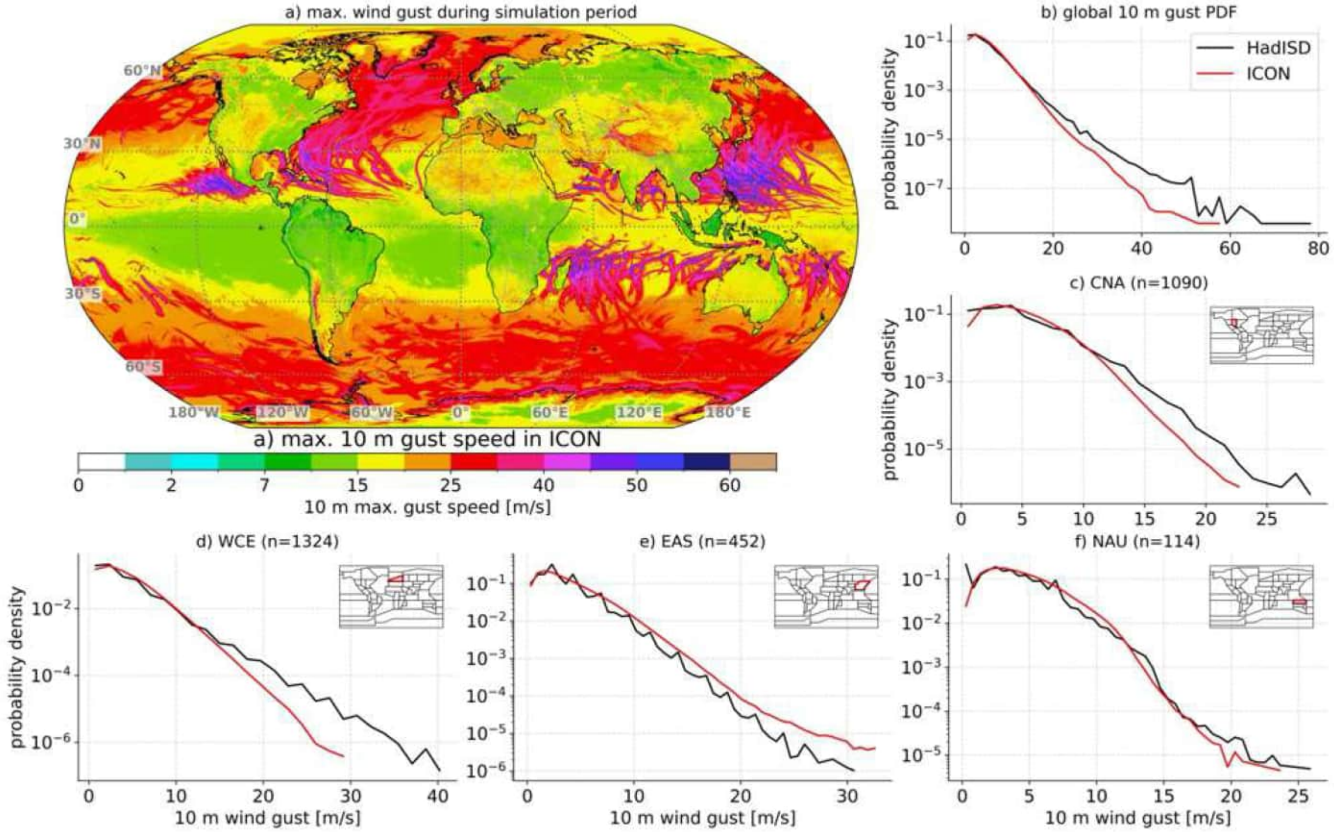

Figure 7. Maximum instantaneous simulated 10 m above surface wind from hourly model output during the simulation period (a). Probability density functions for hourly sampled winds from HadISD (black line) and ICON (red line). Statistics are shown for all stations globally (b; see Fig. S4 for station locations) and for the Central North America (CNA; c), Western Central Europe (WCE; d), East Asia (EAS; e), and North Australia (NAU; f) region. The number of stations (n) in each region is shown in the panel titles.

could also contribute. Intense winds are well simulated in most tropical and high latitude areas such as East Asia (EAS) and Northern Australia (NAU) (Fig. 7e,f). The station density that is available for comparison in HadISD is shown in Fig. S4, and statistics for additional IPCC regions are shown in Fig. S6. Also, ICON reproduces the temporal autocorrelation of hourly winds well (Fig. S5), indicating that it captures the persistence and day-to-day temporal variability of near-surface wind fluctuations, including both the rapid decorrelation at short lags and the diurnal recurrence seen in the observations.

The wind speed evaluation should be interpreted with caution, as we compare station data with 2.5 km grid-cell averages. Additionally, short-term average wind speeds (e.g., 2-minute averages at U.S. stations, 10-minute averages at many WMO stations) are compared with hourly instantaneous model winds. Instantaneous model snapshots tend to produce broader, higher-tailed wind-speed PDFs than temporally averaged station observations, while spatial averaging over 2.5 km grid cells damps small-scale wind maxima. The combined effect partially compensates for differences. Km-scale modes are generally more comparable to station observations because the scale differences between the simulated processes on the grid and point-based observations are smaller than in coarser-resolution models.

3.5 Tropical cyclones

Tropical cyclone tracks compare well in frequency and location between ICON and IBTrACS, except for the North Atlantic Basin, which features a low bias (Fig. 8a,b). Note that the longer tracks in IBTrACS are caused by identifying TCs already as tropical depressions and following them after the extra-tropical transition. In contrast, MOAAP only identifies the main TC phase. The inter-annual variability of basin total tropical cyclone frequencies is also well captured in ICON, except for the Southwest Pacific and South Indian Ocean, where the variability is smaller (Fig. 8b).

The TC pressure wind relationship is relatively well captured, but ICON wind speeds are too low (Fig. 8c), consistent with the UV10 analysis in the previous section and simulations with the NICAM model at km-scale (Takasuka et al., 2024a). The low difference is also reflected in the probability histogram of tropical cyclone lifetime peak wind speeds, which drops off too rapidly at high wind speeds (Fig. 8d). The IBTrACS distribution shows a bi-modal distribution with a minimum around $40\,\mathrm{m\,s^{-1}}$ and a primary peak at low wind speeds. The reason for this primary peak is that the tropical depression and tropical storm phases of TC are included in IBTrACS. In contrast, the MOAAP tracker is calibrated to identify only storms of tropical storm intensity. Similar to the comparison to station-based wind observations in the previous chapter, comparing modeled instantaneous wind fields to IBTrACS observations is challenging. The simulation provides gridded 10 m near-surface winds, while IBTrACS reports storm-level maximum sustained winds from observational best-track analyses. They are therefore not directly equivalent, since the model resolves spatial wind structure and average wind speeds within grid cells, whereas IBTrACS summarizes cyclone intensity. This could explain parts of the low difference in wind speed seen in ICON.

3.6 Equatorial waves and the Madden-Julian oscillation

Fig. 9 shows Hoffmöller diagrams of tropical average Tb in ICON and GPM-MERGIR, with labels indicating equatorial waves from MOAAP output for May 2020. While Rossby waves are well simulated in ICON, other waves that are strongly coupled to deep convection are underestimated in both frequency and amplitude. Kelvin and inertia-gravity waves (IGWs) are very infrequent, and Mixed Rossby-Gravity Waves (MRGs) are only regularly observed over Africa.

Global kilometer-scale (storm-resolving) models substantially improve the simulation of convectively coupled equatorial waves by explicitly representing deep convection, leading to more realistic precipitation spectra, wave propagation, and dynamical structures compared to models with parameterized convection (Satoh et al., 2019; Hohenegger et al., 2020; Jung and Knippertz, 2023). However, despite these advances, biases remain in wave–convection coupling strength, phase speed, and intermittency, indicating that the representation of equatorial waves at km-scale resolution is improved but still model- and configuration-dependent rather than fully resolved (Hohenegger et al., 2020; Takasuka et al., 2024a; Jung and Knippertz, 2023).

Wheeler–Kiladis space–time spectra computed from the four-year evaluation period confirm these results (Fig. 10). We follow the approach of Takayabu (1994) and Takayabu (1994); Wheeler and Kiladis (1999) and apply it to outgoing longwave

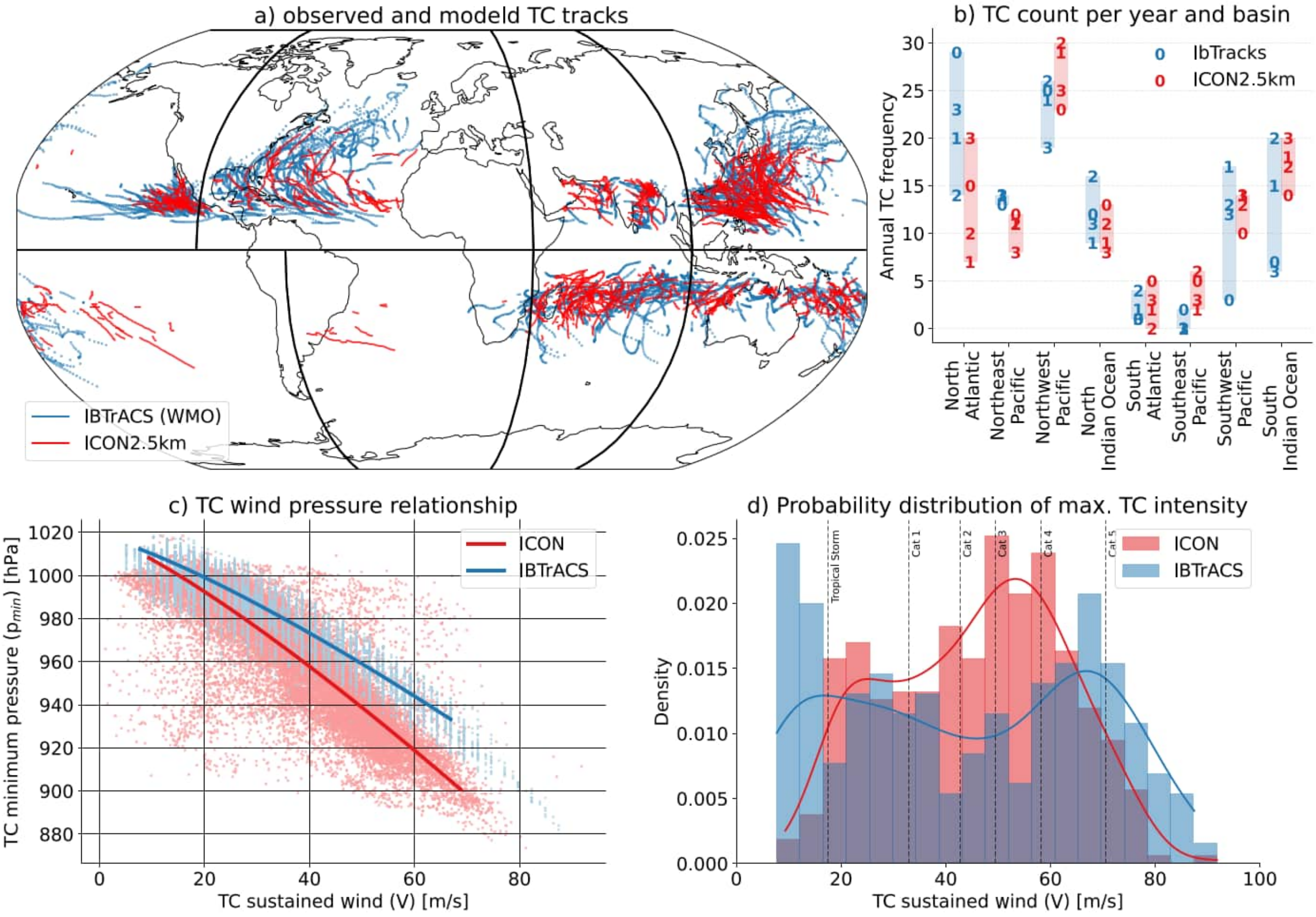

Figure 8. Tropical cyclone (TC) tracks from IBTrACS (blue) and ICON (red) during the simulation period (a). Basin annual frequency of tropical cyclones (b) in 2020 (0; only April to December), 2021 (1), 2022 (2), and 2023 (3). Tropical cyclone pressure wind relationship (c) from IBTrACS (blue) and ICON (red). Thick lines show the fitted figure of $V\left(p_{\min}\right)=a\left(p_{\mathrm{ref}}-p_{\min}\right)^{b}$ with p_min being the minimum cyclone pressure, V being the sustained wind speed, and p_ref being the reference pressure at zero wind speed. Probability density of lifetime peak tropical cyclone sustained wind speed in IBTrACS and ICON (d).

radiation (OLR) from the ICON simulation and the NOAA-interpolated OLR dataset. Overall, ICON reproduces the observed spectral characteristics of equatorial waves reasonably well. Rossby wave activity shows the closest agreement with observations, while other modes—such as Kelvin waves, the Madden–Julian Oscillation (MJO), and IGW—are captured but exhibit lower amplitudes than in the NOAA OLR data.

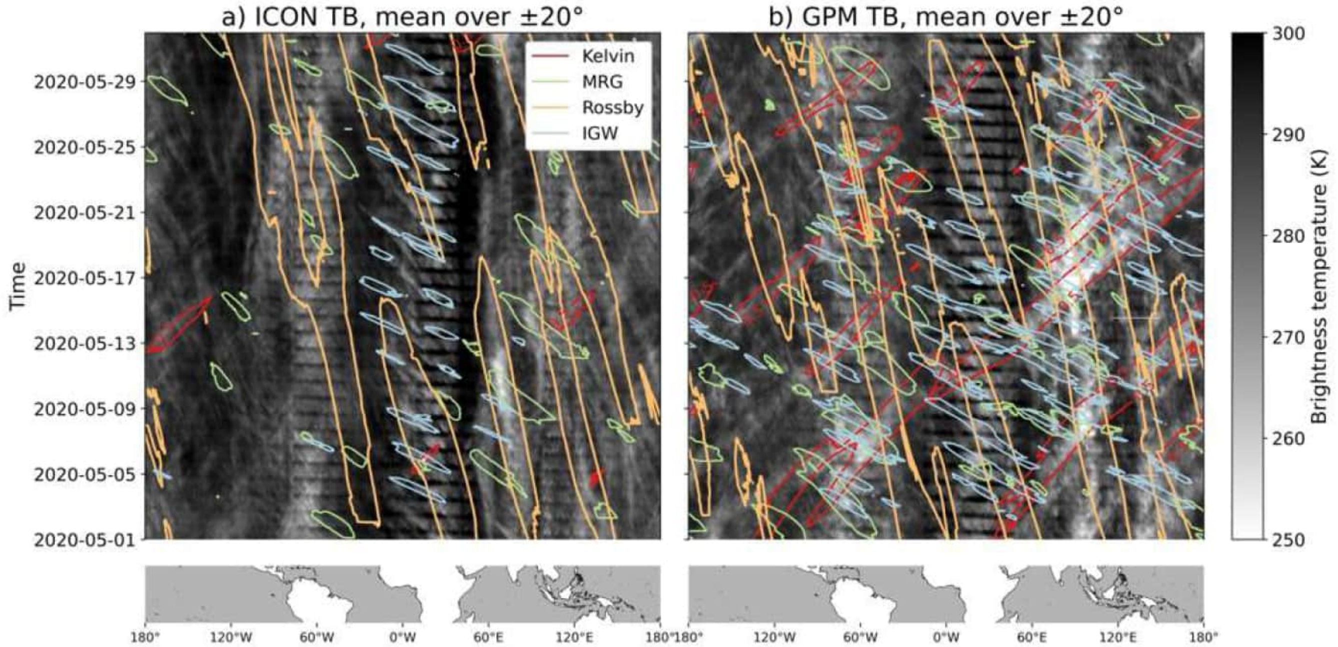

Figure 9. Hoffmöller diagrams of tropical average (±20°) brightness temperature from ICON (a) and GPM-MERGIR (b) shown in gray contours for May 2020. Contours lines show Kelvin (red), Mixed Rossby Gravity (MRG, green), Rossby (orange), and Inertia Gravity (IGW, blue) waves.

3.7 Mesoscale convective systems (MCSs)

Equatorial waves often organize convection into clusters of MCSs. The latent heating from these MCSs can, in turn, amplify and sustain the wave, creating a feedback loop between the large-scale disturbance and convection (Kiladis et al., 2009). The general pattern of MCS initiation frequency is well captured in ICON, but initiation is underestimated over tropical oceans and overestimated over tropical land regions (Fig. 11a–c). The relative underestimation of oceanic MCS frequency is largest in the northern hemisphere, followed by the Southern Hemisphere, and tropics (Fig. 11g–i). In the tropics, oceanic MCS frequencies are well simulated between December and March but have a systematic low bias in the other months of the year (Fig. 11h).

Simulated land-based MCSs frequencies are most similar to observations in the northern hemisphere. Frequencies are overestimated in the Southern Hemisphere during summer and in the tropics year-round, except in June and July.

These frequency biases result in a systematic underestimation of the MCS to total precipitation ratio over almost all ocean regions, while this ratio is overestimated over Africa (Fig. 11d–f). The reasons for the MCS frequency biases are unclear, but they might be related to biases in tropical waves and the thermodynamic-convection coupling (**?**), which will be the topic of future research.

In addition to frequency statistics, we also performed lifetime analysis of oceanic and land-based MCSs in the tropics, northern, and Southern Hemispheres (Fig. 12 and Fig. S7 and S8 respectively). We discuss only the results from the tropics

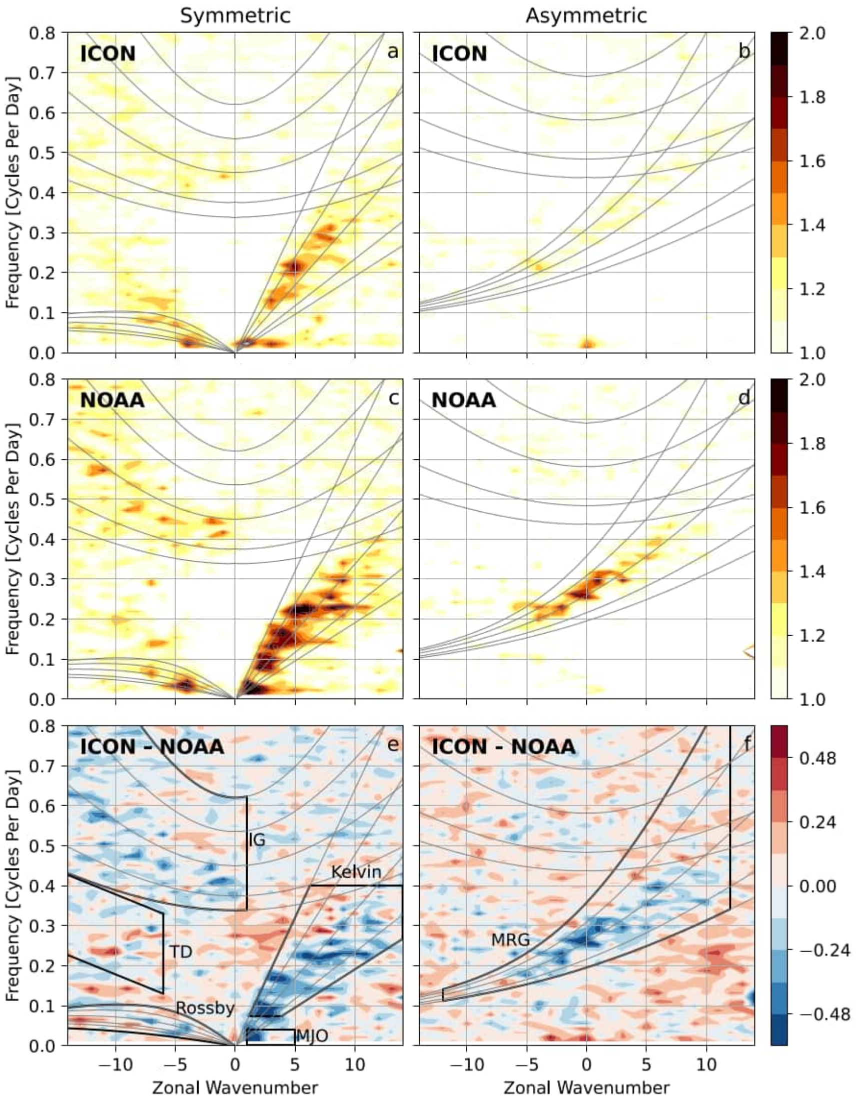

Figure 10. Wavenumber–frequency spectra of outgoing longwave radiation (OLR), divided by the background spectrum, from the ICON simulation (a, b) and NOAA satellite observations (c, d), along with their differences (e, f). Spectra are separated into symmetric (a, c, e) and antisymmetric (b, d, f) components. Gray curves indicate theoretical dispersion relations, while bold black outlines denote spectral windows associated with different equatorial wave modes: IG (inertia–gravity waves), TD (tropical depressions), Rossby, MJO (Madden–Julian Oscillation), Kelvin, and MRG (mixed Rossby–gravity waves).

here, since they are similar in both hemispheres. Observed MCS sizes are similar over tropical oceans and land (Fig. 12a). Simulated MCSs are generally smaller than observed systems. The simulated MCS movement speed is systematically higher for most quintiles (Fig. 12b), while the mean precipitation under the MCS cloud shields is lower (Fig. 12c), particularly for

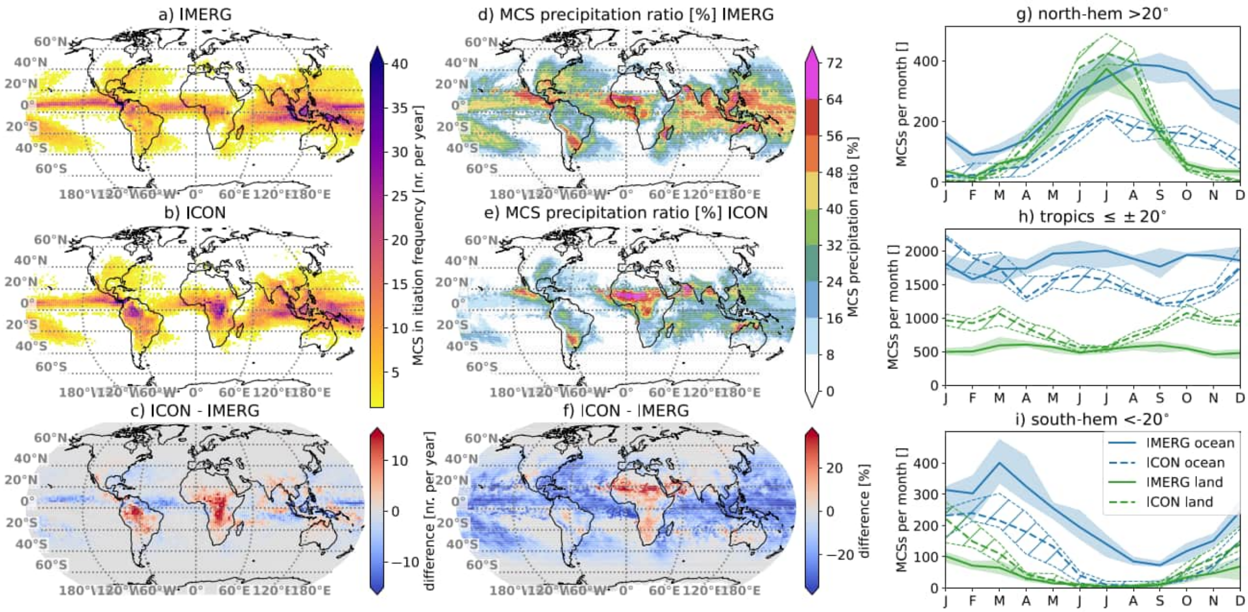

Figure 11. Annual average initiation frequency of MCSs in IMERG (a), ICON (b), and their difference. Ratio of MCS precipitation compared to total precipitation averaged over the simulation period from IMERG (d), ICON (e), and their difference (f). Monthly mean MCS frequency over land (green) and ocean (blue) in the northern hemisphere (g; >20°), the tropics (h; $\leq \pm 20°$), and the Southern Hemisphere (i; <20°). IMERG results are shown with solid lines, while dashed lines show ICON results. The thick line represents the median frequency, while the contours indicate the maximum and minimum spreads over the four years.

land-based MCSs. Finally, modeled MCS durations are too short (2–4 hours median difference), with larger differences over land than over ocean (Fig. 12d).

To compare the temporal evolution of MCSs, we differentiate between short-lived (4-9 hours) and long-lived (>14 hours) tropical MCSs, given lifetime-dependent model biases. The growth and decay of MCS cloud shields are well simulated for oceanic and land-based tropical MCSs (Fig. 12e,f). However, long-lived storms do not grow fast enough and reach a peak size that is about 20–30 % too small. On the other side, long-lived storms feature more accurate movement speeds while small storms move about 20 % too fast (Fig. 12g,h).

Mean MCS precipitation starts high during the initiation phase of MCSs and steadily decreases during their lifetime (Fig. 12i,j). This is because most MCSs start from a burst of small-scale convection and develop stratiform precipitation areas as they grow upscale until the convective areas die off in the decaying phase, resulting in mostly stratiform precipitation (e.g., Mapes et al., 2006). ICON simulates mean precipitation within ±20 % of observed values during the growth phase, with the exception of land-based long-lived systems, which show a high difference of up to 40 %. Long-lived MCSs maintain high mean precipitation rates during the growth phase of the storm (especially over tropical land), while simulated MCSs show a decline in mean

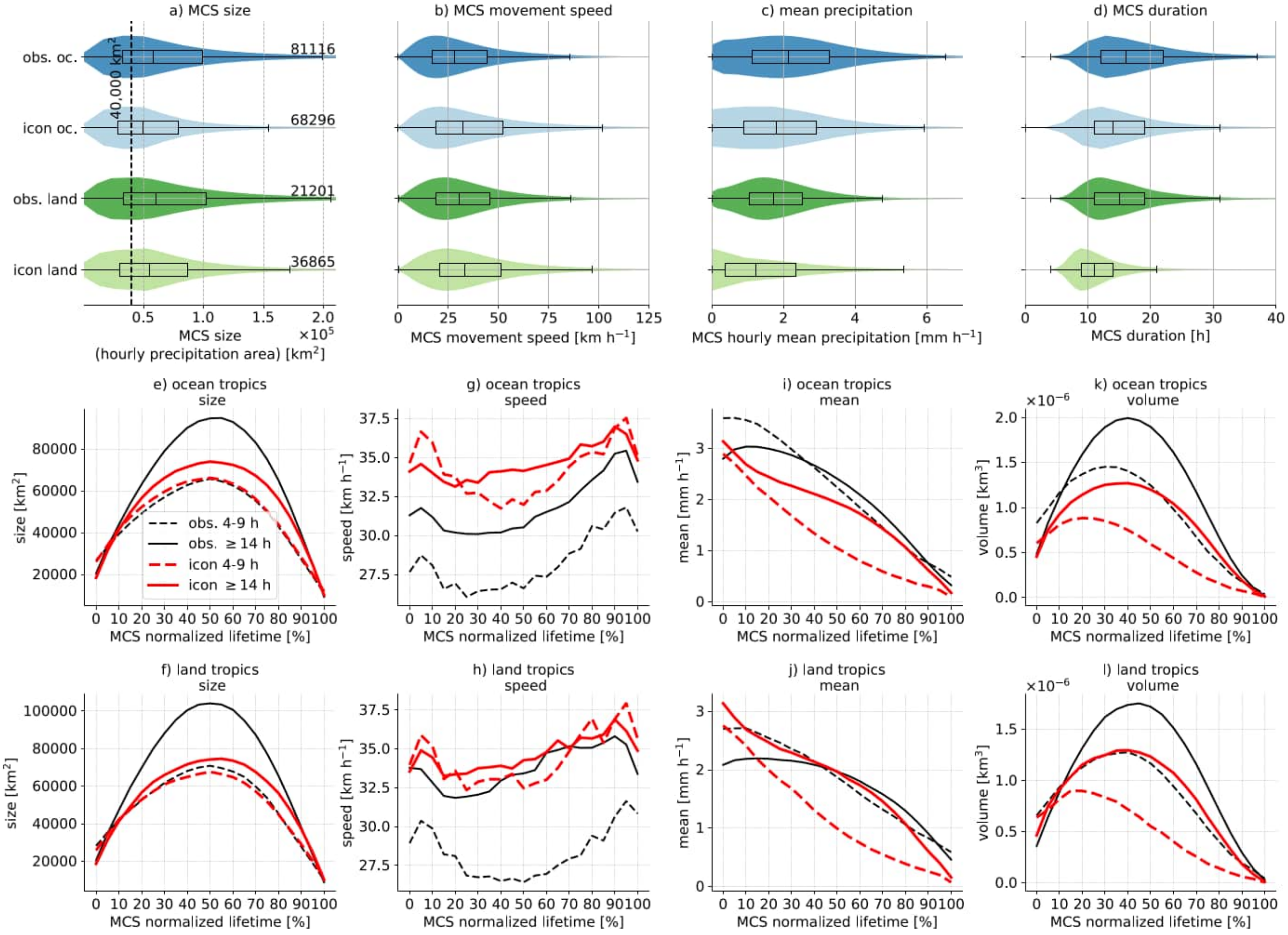

Figure 12. Violin and box-whisker plots for the distribution of tropical MCS sizes (a), speed (b), mean precipitation (c), and duration (d). The numbers right to the violins in (a) show the number of MCSs in each distribution. The lower panels show the normalized median evolution of short-lived (4-9 hours; dashed) and long-lived ($\geq$14 hours; solid) MCSs in GPM observations (black lines) and in ICON (red lines). The middle panels show results for tropical oceans while the lower panel shows tropical land regions.

precipitation even early on during their lifetime. This results in simulated mean MCS precipitation becoming increasingly low-biased over the life cycle of storms ending at low differences of up to -60 % during the decaying phase. This indicates difficulties in simulating MCS stratiform precipitation, which is a common issue in many km-scale models (e.g., Prein et al., 2023a; Zhang et al., 2024).

Finally, the combined differences in mean precipitation rates and MCS size are affecting the differences in MCS precipitation volume, which is better simulated during the growth phase but becomes increasingly negatively biased over their lifetime, especially for short-lived systems (Fig. 12k,l). The dominant driver of this bias is the low mean precipitation difference for short-lived systems, as well as the low size difference for long-lived storms. We refrain from analyzing differences in the development of heavy precipitation rates due to the deficiencies in capturing these in GPM-IMERG (see Fig. 4.)

3.8 Precipitating cloud characteristics

We found some of the largest model wet biases in tropical ocean ITCZ precipitation (Fig. 3). At the same time, MCSs are too infrequent in this region. To better understand the causes of this bias we also tracked smaller precipitating cloud systems that have cold cloud tops in addition to MCSs (see section 2.3.1 and Fig. 13a–c). As previous studies have shown, cloud sizes (i.e., cold cloud top areas) follow a power-law frequency distribution (e.g., Savre and Craig, 2023). Observations show that cold cloud shields that exceed our MCS threshold (>40,000 km^2) contribute the most to tropical precipitation with a maximum contribution from clouds with ∼75,000 km^2 (Fig. 13a). This maximum is shifted towards smaller cloud sizes and lower volumes in ICON (Fig. 13b) with a low contribution bias for large clouds and an overestimation from small clouds (Fig. 13c). Not visible in Fig. 13 are the low-frequency biases in MCSs over land and ocean, and in non-MCS precipitating cold clouds over land (Fig. S9), since only normalized distributions are shown. These results are broadly consistent with Segura and Hohenegger (2024), who showed that a large fraction of tropical precipitation originates from congestus clouds. GPM-IMERGv7 is not well suited to robustly quantify precipitation specifically from shallow or congestus clouds due to deficiencies in detecting precipitation from warm rain clouds (North et al., 2022; Huffman et al., 2023).

A key question is: which clouds cause the excess simulated precipitation in the tropics if deep convective clouds are too infrequent and too small? Fig. 13d–f show that in observations, most tropical precipitation originates beneath cold cloud tops, while ICON simulations have a too large fraction of rainfall coming from warm clouds. This issue is particularly evident in tropical land and ocean regions, but is also evident to a lesser extent in the Northern and Southern Hemispheres (Fig. S9 and Fig. S10). Excess precipitation from clouds with warm cloud tops indicates overly active warm-cloud processes, too slow glaciation in tropical clouds in the ICON microphysics scheme, or errors in GPM-IMERGv7 in capturing warm rain. This is consistent with the underestimated surface shortwave radiation over tropical oceans (Fig. 2c) that could be caused by overly reflective or overly abundant water clouds.

3.9 Drivers of heavy precipitation

The output from the MOAAP algorithm allows us to analyze the interactions among atmospheric phenomena in creating hourly heavy precipitation events in observations and in the ICON model (Fig. 14). We decided to exclude equatorial waves and fronts from the analysis in Fig. 14a–f since equatorial waves dominate tropical heavy precipitation statistics due to their large size, and including them obscures details from smaller-scale phenomena (Fig. S12). We decided to exclude fronts because they are more frequent in ICON, likely due to its higher model resolution, which allows sharper gradients to be simulated. Results including fronts and equatorial waves are shown in Fig. S12.

ICON can effectively simulate the processes that cause extreme precipitation across different climate regions, though it has some notable deficiencies. In observations, MCSs are more often the most frequent producers of heavy precipitation in tropical and subtropical ocean regions (Fig. 14a,b). This agrees with our previous analysis, which showed low-frequency and

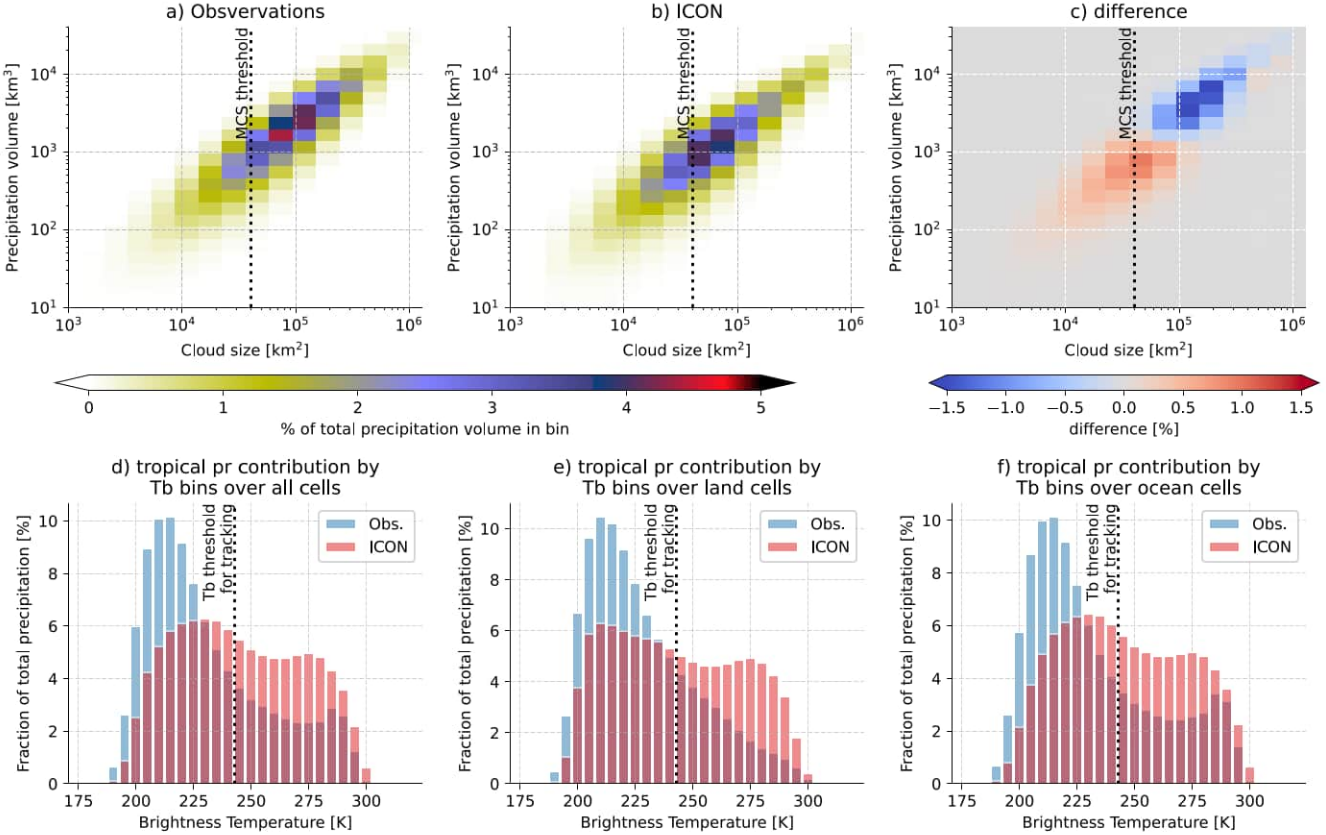

Figure 13. Relative contribution of tropical (latitude < ±20 °) total precipitation from storms with cold cloud shields (Tb < 243 K) as a function of cloud shield size and precipitation volume under the cloud from observations (a), ICON (b), and their difference (c). The vertical dashed line shows the cloud shield threshold used for MCS classification. The fraction of total precipitation conditioned on Tb over the tropics (d), tropical land (e), and tropical ocean (f). The vertical dashed line shows the Tb threshold used for cold cloud identification.

size differences in simulated MCSs. In contrast, simulated MCSs are more dominant over tropical and subtropical land regions, where observations highlight non-MCS cold clouds as most important. At the same time, moisture streams (areas of low-level strong and coherent horizontal moisture transport) are frequently more relevant in ICON than in observations, particularly in subtropical ocean regions.

Fig. 14g–i show frequency statistics that visualize how often a phenomenon was co-located with hourly heavy precipitation in the Northern European (NEU), Southeast South America (SES), and the Equatorial Pacific Ocean (EPO) region. In the high-latitude NEU region, observations and ICON agree that surface cyclones and mid-level cyclones are the most frequent drivers with secondary contributions from jet streams and moisture streams (Fig. 14g). ICON features more heavy precipitation events from non-MCS clouds than observed and a larger contribution of frontal systems. Cut-off lows are less frequently involved in simulated heavy precipitation events than in observations.

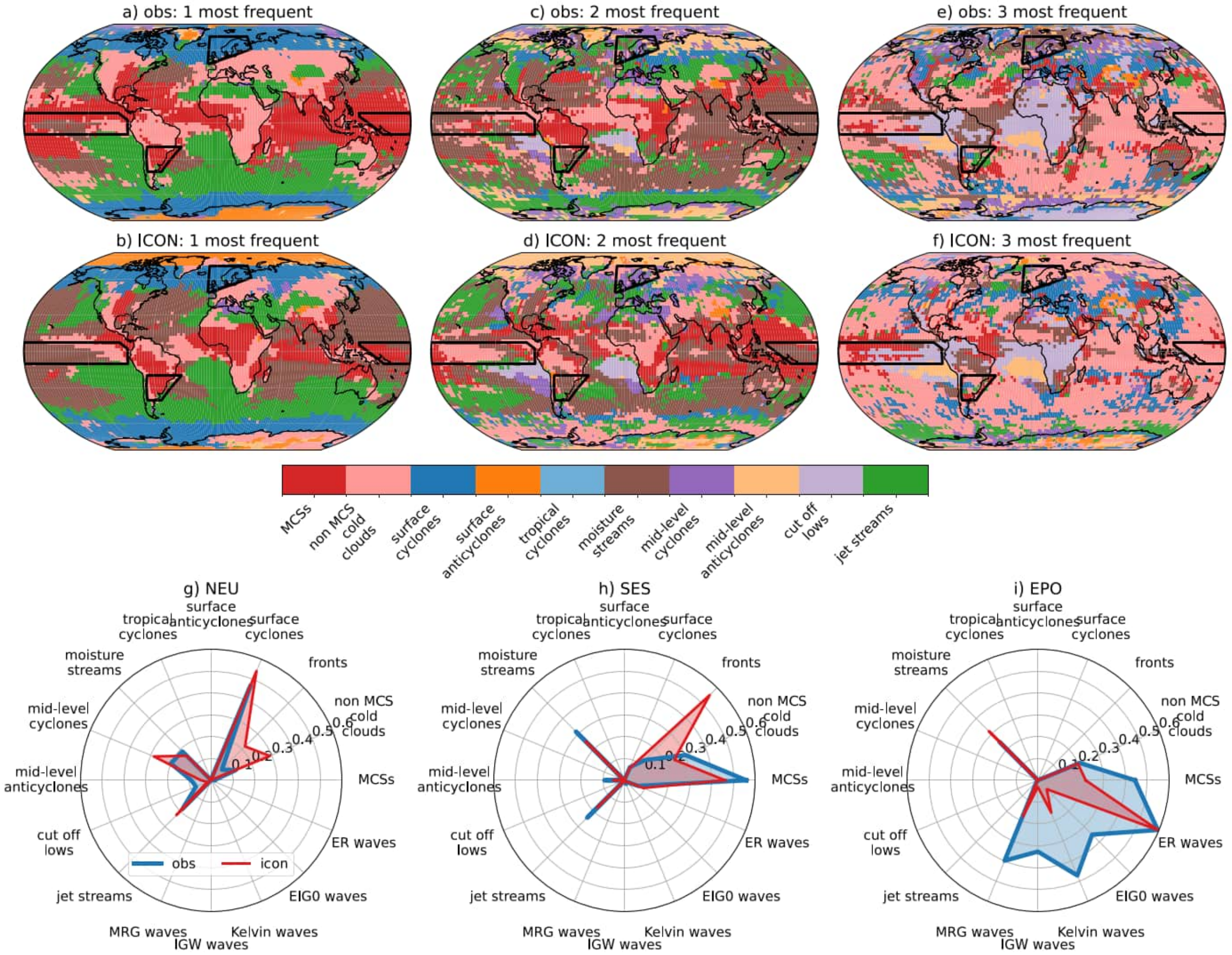

Figure 14. The most frequent (a,b), second most frequent (c,d), and third most frequent (e,f) atmospheric phenomena identified by the MOAAP algorithm that is co-located during the top 20 hourly precipitation events in the evaluation period. Observed/ICON results are shown in the top/middle panel. The data were upscaled to a 2.5 ° grid to reduce the signal-to-noise ratio. The same data is shown in panel (g–i) as the fractional frequency of the phenomenon in the Northern European (NEU), Southeast South America (SES), and Equatorial Pacific Ocean (EPO) regions (these regions are highlighted in the maps in (a–f)). A value of 0.5 means that 50 % of all grid cells in a region had the corresponding phenomena co-located during a heavy precipitation event.

In the SES region, ICON agrees with observations that MCSs and non-MCS clouds are the dominant heavy precipitation-producing phenomena (Fig. 14h). ICON also captures the importance of moisture streams and jet streams well, while frontal involvement is much higher than in ERA5.

We find a relatively poor agreement in the tropical EPO region (Fig. 14i) where ICON substantially underestimates the importance of MCSs and all equatorial waves except for Equatorial Rossby waves, in agreement with previously shown results. However, ICON can capture the importance of non-MCS cold clouds and moisture streams in the production of heavy precipitation.

4 Summary and Conclusion

We present results from the first multi-year (April 2020–March 2024), global prescribed SST km-scale simulation performed with the GT4Py GPU-refactored ICON model (Dipankar et al., 2026) at a horizontal grid spacing of 2.5 km and 120 vertical levels. This experiment, conducted within the EXCLAIM project, represents a major milestone toward bridging the long-standing divide between numerical weather prediction (NWP) and climate modeling (Randall and Emanuel, 2024), using a model configuration based on operational forecasting without empirical tuning. By explicitly resolving deep convection, orographic drag, mesoscale circulations, and local land heterogeneity globally, the model enables the consistent study of weather and climate processes within a unified physical framework.

We conducted a comprehensive evaluation using satellite, reanalysis, and in-situ observations to assess the fidelity of the simulation of atmospheric phenomena at scales ranging from global to local. The following main points summarize our study:

- The ICON model reproduces observed large-scale patterns of temperature, precipitation, and surface energy fluxes with high fidelity, showing good agreement with reanalysis, satellite, and in situ observations, although biases, such as overestimation of sensible heat fluxes over land during the warm season, persist. Importantly, the model can realistically simulate the spatial pattern of tropical precipitation with a single ITCZ. However, there is an overestimation of tropical oceanic clouds (up to -40 % in incoming shortwave radiation at the surface) and precipitation (up to 50 %) and cold biases in polar regions (more than -3 °C) during winter. Additionally, warm biases over mid-latitude continents during the warm season are related to too much incoming solar radiation, an underestimation in latent heat fluxes, and an overestimation of sensible heat fluxes.

- Explicitly resolving convection yields a realistic convective diurnal cycle and an excellent representation of the hourly rainfall probability density distribution, including extremes. Still, some land regions, such as those over southern Africa, Australia, and parts of the Amazon basin, feature a diurnal precipitation peak that is too early. Importantly, model analyses that focus on the intensity or frequency of hourly precipitation should use in-situ observation records where possible, since these precipitation characteristics are not well captured in GPM-IMERG-v7 and reanalysis products, as we show here and in previous work (e.g., Dominguez et al., 2024).

- Near-surface winds and tropical cyclone statistics are generally well simulated, but maximum wind speeds are underestimated by $\sim$20 %. Tropical cyclone frequencies are well captured, except in the North Atlantic basin, where low biases are observed, similar to other high-resolution climate model simulations (Roberts et al., 2020).

- Spatial initiation patterns of MCSs are realistic, but ICON underestimates oceanic MCS frequencies (especially outside DJF) and overestimates tropical land MCSs. Simulated systems are smaller, slightly too fast, and too short-lived, resulting in an underestimation of MCS precipitation over oceans and an overestimation over some tropical land areas. The low bias in tropical MCS precipitation and the underestimation of MCS size are common among km-scale global models (Feng et al., 2025). The low-frequency bias in long-lived, large MCSs contributes to the earlier peak timing in the convective diurnal cycle.

- Since both precipitation from MCS and non-MCS cold clouds is underestimated over tropical ocean regions, it is unclear which clouds are causing the oceanic wet tropical precipitation difference in ICON. The primary cause is larger simulated rainfall from shallow and mid-level clouds (e.g., cumulus congestus) in the model. The reason for these differences is unclear, but an overactive warm-rain process in the model and deficiencies in warm-rain detection in GPM-IMERG-v7 might contribute.

- The frequency and amplitude of all equatorial waves except for Rossby waves are underestimated. This is consistent with the underestimation of MCSs and the overestimation of precipitation from shallow clouds in the tropics since equatorial waves are frequently coupled to deep convection (Kiladis et al., 2009; Prein et al., 2023b). Km-scale global models tend to capture equatorial Rossby waves well since their dynamics are more strongly controlled by large-scale vorticity gradients rather than the details of convective heating, although column water vapor also contributes (Yasunaga and Mapes, 2012; Yasunaga et al., 2019; Nakamura and Takayabu, 2022). Previous global km-scale modeling studies have shown that uncoupled models can capture convective-coupled equatorial waves (e.g., Judt and Rios-Berrios, 2021; Takasuka et al., 2024a; Ortega et al., 2026). The biases presented here might also stem in part from a misrepresentation of thermodynamic–convection coupling, which can strongly influence convective organization and its coupling to larger-scale tropical variability (Takasuka et al., 2026). Additionally, ocean-atmosphere interactions are important for simulating connectively coupled equatorial waves, which our forced SST setup does not capture (e.g., DeMott et al., 2015).

- An object-based evaluation using the MOAAP feature tracker (Prein et al., 2023b) reveals that ICON realistically captures the main meteorological drivers of heavy hourly precipitation, including cyclones, MCSs, and moisture streams. We consistently find deficiencies in simulating heavy rainfall associated with tropical oceanic MCSs and equatorial waves.

In summary, the EXCLAIM 2.5 km global ICON simulation provides a unique benchmark for studying global hydroclimatic extremes and mesoscale dynamics. It will serve as a contribution to the DYAMOND-III intercomparison initiative (Takasuka et al., 2024b), allowing for comparison with an ensemble of km-scale global models. The simulation offers novel insight into the interplay between different modes of convection and large-scale circulation. Current ICON developments within the EXCLAIM project aim to optimize the code for faster runtime and implement ocean-atmosphere coupling at km-scales (Dipankar et al., 2026). This framework also provides a basis for follow-up studies investigating how mesoscale processes, such as convective outbreaks, feed back onto and modify the larger-scale circulation. Future studies will also focus on longer integration periods, as four years are insufficient to capture climate variability.

While notable challenges remain in representing tropical variability, mesoscale convective organization, and land-atmosphere coupling, the demonstrated capability of our simulation and other km-scale global modeling efforts establishes a new foundation for next-generation Earth system modeling and the transition toward truly unified weather–climate simulations.

Code and data availability. The global 2.5 km ICON simulation was produced within the EXCLAIM project following the DYAMOND-III protocol (Prein, 2026).

The ICON model version used in this study corresponds to ICON-EXCLAIM v0.2.0. The complete model source code, including the EXCLAIM extensions, is publicly archived at Zenodo (Dipankar, 2025): https://doi.org/10.5281/zenodo.17255275

The configuration files, namelists, and run scripts used to perform the global 2.5 km simulation are archived separately at Zenodo: https://doi.org/10.5281/zenodo.17250248

ICON global 2.5 km simulation and analysis data can be accessed from Figshare: https://doi.org/10.6084/m9.figshare.31341982.

The analysis and visualization code used to process the model output and observational datasets, and to generate all figures in this manuscript, is archived on Zenodo (Prein, 2025): https://doi.org/10.5281/zenodo.18648539.

ERA5 reanalysis data were obtained from the Copernicus Climate Data Store (Hersbach et al., 2020): Copernicus Climate Change Service Climate Data Store. https://doi.org/10.24381/cds.adbb2d47.

NOAA Interpolated Outgoing Longwave Radiation (OLR) data were obtained from NOAA Physical Sciences Laboratory (PSL), dataset "NOAA Interpolated OLR" (accessed 2026.02.15): https://psl.noaa.gov/data/gridded/data.olrcdr.interp.html. The interpolation method is described by Liebmann and Smith (1996).

GPM IMERG precipitation data were obtained from the NASA Goddard Earth Sciences Data and Information Services Center (Huffman et al., 2020): https://doi.org/10.5067/GPM/IMERG/3B-HH/07; https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_07/summary.

Merged infrared brightness temperature data were obtained from the NOAA Climate Prediction Center merged IR dataset archive: https://doi.org/10.5067/P4HZB9N27EKU.

HadISD station observations were obtained from the Met Office Hadley Centre archive (Dunn et al., 2016): https://www.metoffice.gov.uk/hadobs/hadisd/.

FLUXNET2015 surface flux data were obtained from the FLUXNET data archive (Pastorello et al., 2020): https://doi.org/10.18140/FLX/1440160.

AmeriFlux surface flux data were obtained from the AmeriFlux data archive: https://ameriflux.lbl.gov/data/.

IBTrACS tropical cyclone data were obtained from NOAA National Centers for Environmental Information (Gahtan et al., 2024): https://doi.org/10.25921/82ty-9e16; https://www.ncei.noaa.gov/products/international-best-track-archive.

EWEMBI surface radiation data were obtained from the ISIMIP data archive (Lange, 2018): https://www.isimip.org/gettingstarted/input-data-bias-adjustment/details/27/.

Author contributions. AFP designed, conceptualized, performed most of the analyses, and wrote the first draft of the manuscript. PP ran the ICON simulations and archived the data with support from CZ. ML performed the surface flux analyses, and MR performed the Wheeler

and Kiladis analysis. AD, ML, and AJ supported the development and configuration of the applied ICON model. All authors contributed to the writing of the paper.

Competing interests. We do not have any competing interests.

Acknowledgements. This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under the EXCLAIM project on Alps. The authors thank the NOAA Physical Sciences Laboratory (PSL), Boulder, Colorado, USA, for providing the Interpolated Outgoing Longwave Radiation (OLR) dataset, available at https://psl.noaa.gov. ERA5 reanalysis data were obtained from the Copernicus Climate Change Service (C3S) Climate Data Store. The GPM IMERG data were provided by the NASA/Goddard Space Flight Center's Global Precipitation Measurement and PPS, which develop and compute the GPM IMERG as a contribution to GPM, and archived at the NASA GES DISC. Merged infrared brightness temperature data were provided by NASA and NOAA. We acknowledge the UK Met Office for HadISD station observations, the FLUXNET and AmeriFlux networks and their contributing site investigators, NOAA/NCEI for IBTrACS tropical cyclone data, and ISIMIP for providing the EWEMBI surface radiation dataset. We used Grammarly and ChatGPT to support the writing of this manuscript.

References

EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI), author=Lange, Stefan, (No Title), 2016.

Ahlgrimm, M. and Forbes, R.: The impact of low clouds on surface shortwave radiation in the ECMWF model, Monthly Weather Review, 140, 3783–3794, 2012.

Argüeso, D., Di Luca, A., and Evans, J. P.: Precipitation over urban areas in the western Maritime Continent using a convection-permitting model, Climate dynamics, 47, 1143–1159, 2016.

Asensio, H., Messmer, M., Lüthi, D., and Osterried, K.: External Parameters for Numerical Weather Prediction and Climate Application EXTPAR v5_0, User and Implementation Guide. Available online: http://www. cosmo-model. org/content/support/software/ethz/EXTPAR_user_and_implementation_ manual_202003. pdf (accessed on 16 November 2018), 2020.

Ban, N., Schmidli, J., and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, Journal of Geophysical Research: Atmospheres, 119, 7889–7907, 2014.

Ban, N., Caillaud, C., Coppola, E., Pichelli, E., Sobolowski, S., Adinolfi, M., Ahrens, B., Alias, A., Anders, I., Bastin, S., et al.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation, Climate Dynamics, 57, 275–302, 2021.

Barlage, M., Chen, F., Rasmussen, R., Zhang, Z., and Miguez-Macho, G.: The importance of scale-dependent groundwater processes in land-atmosphere interactions over the central United States, Geophysical Research Letters, 48, e2020GL092 171, 2021.

Brown, A., Dowdy, A., and Lane, T. P.: Convection-permitting climate model representation of severe convective wind gusts and future changes in southeastern Australia, Natural Hazards and Earth System Sciences, 24, 3225–3243, 2024.

Chu, H., Christianson, D. S., Cheah, Y.-W., Pastorello, G., O'Brien, F., Geden, J., Ngo, S.-T., Hollowgrass, R., Leibowitz, K., Beekwilder, N. F., et al.: AmeriFlux BASE data pipeline to support network growth and data sharing, Scientific Data, 10, 614, 2023.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C.: Convection-permitting models: A step-change in rainfall forecasting, Meteorological Applications, 23, 165–181, 2016.

Dee, D. P., Uppala, S., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, https://doi.org/https://doi.org/10.1002/qj.828, 2011.

DeMott, C. A., Klingaman, N. P., and Woolnough, S. J.: Atmosphere-ocean coupled processes in the Madden-Julian oscillation, Reviews of Geophysics, 53, 1099–1154, 2015.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: the role of internal variability, Climate dynamics, 38, 527–546, 2012.

Dipankar, A.: EXCLAIM use cases, https://doi.org/10.5281/zenodo.17250248, 2025.

Dipankar, A., Bianco, M., Bukenberger, M., Ehrengruber, T., Farabullini, N., Gopal, A., Hupp, D., Jocksch, A., Kellerhals, S., Kroll, C. A., Lapillonne, X., Leclair, M., Luz, M., Müller, C., Ong, C. R., Osuna, C., Pothapakula, P., Röthlin, M., Sawyer, W., Serafini, G., Vogt, H., Weber, B., and Schulthess, T.: Toward Exascale Climate Modelling: A Python DSL Approach to ICON's (Icosahedral Non-hydrostatic) Dynamical Core (icon-exclaim v0.2.0), EGUsphere, 2025, 1–26, https://doi.org/10.5194/egusphere-2025-4808, 2025.

Dipankar, A., Bianco, M., Bukenberger, M., Ehrengruber, T., Farabullini, N., Fuhrer, O., Gopal, A., Hupp, D., Jocksch, A., Kellerhals, S., Kroll, C. A., Lapillonne, X., Leclair, M., Luz, M., Müller, C., Ong, C. R., Osuna, C., Pothapakula, P., Prein, A., Röthlin, M., Sawyer, W., Schär, C., Schemm, S., Serafini, G., Vogt, H., Weber, B., Wills, R. C. J., Gruber, N., and Schulthess, T. C.: Toward exascale climate modelling: a python DSL approach to ICON's (icosahedral non-hydrostatic) dynamical core (icon-exclaim v0.2.0), Geoscientific Model Development, 19, 713–729, https://doi.org/10.5194/gmd-19-713-2026, 2026.

Dominguez, F., Rasmussen, R., Liu, C., Ikeda, K., Prein, A., Varble, A., Arias, P. A., Bacmeister, J., Bettolli, M. L., Callaghan, P., et al.: Advancing South American water and climate science through multidecadal convection-permitting modeling, Bulletin of the American Meteorological Society, 105, E32–E44, 2024.

Donahue, A. S., Caldwell, P. M., Bertagna, L., Beydoun, H., Bogenschutz, P. A., Bradley, A., Clevenger, T. C., Foucar, J., Golaz, C., Guba, O., et al.: To exascale and beyond—The Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM), a performance portable global atmosphere model for cloud-resolving scales, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 314, 2024.

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., and Maher, N.: More extreme precipitation in the world's dry and wet regions, Nature Climate Change, 6, 508–513, 2016.

Dunn, Robert J. H.: HadISD.3.4.0: Product User Guide, Met Office Hadley Centre, Exeter, UK, https://hadleyserver.metoffice.gov.uk/hadisd/hadisd_v340_2023f_product_user_guide.pdf, version 3.4.0 (2023f) of the HadISD dataset, updated 12 Jan 2024., 2024.

Dunn, R. J., Willett, K. M., Parker, D. E., and Mitchell, L.: Expanding HadISD: Quality-controlled, sub-daily station data from 1931, Geoscientific Instrumentation, Methods and Data Systems, 5, 473–491, 2016.

Feng, Z., Prein, A. F., Kukulies, J., Fiolleau, T., Jones, W. K., Maybee, B., Moon, Z. L., Núñez Ocasio, K. M., Dong, W., Molina, M. J., et al.: Mesoscale convective systems tracking method intercomparison (MCSMIP): Application to DYAMOND global km-scale simulations, Journal of Geophysical Research: Atmospheres, 130, e2024JD042 204, 2025.

Gahtan, J., Knapp, K. R., Schreck, C. J. I., Diamond, H. J., Kossin, J. P., and Kruk, M. C.: International Best Track Archive for Climate Stewardship (IBTrACS) Project, Version 4.01, NOAA National Centers for Environmental Information, https://doi.org/10.25921/82ty-9e16, 2024.

Gentry, M. S. and Lackmann, G. M.: Sensitivity of simulated tropical cyclone structure and intensity to horizontal resolution, Monthly Weather Review, 138, 688–704, 2010.

Giorgetta, M. A., Sawyer, W., Lapillonne, X., Adamidis, P., Alexeev, D., Clément, V., Dietlicher, R., Engels, J. F., Esch, M., Franke, H., et al.: The ICON-A model for direct QBO simulations on GPUs (version icon-cscs: baf28a514), Geoscientific Model Development, 15, 6985–7016, 2022.

Giorgi, F. and Gutowski Jr, W. J.: Regional dynamical downscaling and the CORDEX initiative, Annual review of environment and resources, 40, 467–490, 2015.

Giorgi, F., Jones, C., Asrar, G. R., et al.: Addressing climate information needs at the regional level: the CORDEX framework, World Meteorological Organization (WMO) Bulletin, 58, 175, 2009.

Grasselt, R., Schuttemeyer, D., Warrach-Sagi, K., Ament, F., and Simmer, C.: Validation of TERRA-ML with discharge measurements, Meteorologische Zeitschrift, 17, 763, 2008.

Guilloteau, C. and Foufoula-Georgiou, E.: Multiscale evaluation of satellite precipitation products: Effective resolution of IMERG, in: Satellite Precipitation Measurement: Volume 2, pp. 533–558, Springer, 2020.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Bruyere, C. L., Done, J. M., Garrè, L., Friis-Hansen, P., and Veldore, V.: Changes in hurricanes from a 13-yr convection-permitting pseudo–global warming simulation, Journal of Climate, 31, 3643–3657, 2018.

Hayden, L. and Liu, C.: Differences in the diurnal variation of precipitation estimated by spaceborne radar, passive microwave radiometer, and IMERG, Journal of Geophysical Research: Atmospheres, 126, e2020JD033 020, 2021.

He, J., Hong, L., Shao, C., and Tang, W.: Global evaluation of simulated surface shortwave radiation in CMIP6 models, Atmospheric Research, 292, 106 896, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly journal of the royal meteorological society, 146, 1999–2049, 2020.

Hohenegger, C., Brockhaus, P., Bretherton, C. S., and Schär, C.: The soil moisture–precipitation feedback in simulations with explicit and parameterized convection, Journal of Climate, 22, 5003–5020, 2009.

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., Schulzweida, U., and Stevens, B.: Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing, Journal of the Meteorological Society of Japan. Ser. II, 98, 73–91, 2020.

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., et al.: ICON-Sapphire: simulating the components of the Earth system and their interactions at kilometer and subkilometer scales, Geoscientific Model Development, 16, 779–811, 2023.

Holloway, C., Woolnough, S., and Lister, G.: Precipitation distributions for explicit versus parametrized convection in a large-domain high-resolution tropical case study, Quarterly Journal of the Royal Meteorological Society, 138, 1692–1708, 2012.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C., Nelkin, E. J., Sorooshian, S., Stocker, E. F., Tan, J., et al.: Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG), in: Satellite precipitation measurement: Volume 1, pp. 343–353, Springer, 2020.

Huffman, G. J., Bolvin, D. T., Joyce, R., Kelley, O. A., Nelkin, E. J., Tan, J., Watters, D. C., and West, B. J.: Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation, NASA Goddard Space Flight Center, https://gpm.nasa.gov/sites/default/files/2023-07/IMERG_TechnicalDocumentation_final_230713.pdf, version 07, dated 13 July 2023, 2023.

Ikeda, K., Rasmussen, R., Liu, C., Newman, A., Chen, F., Barlage, M., Gutmann, E., Dudhia, J., Dai, A., Luce, C., et al.: Snowfall and snowpack in the Western US as captured by convection permitting climate simulations: Current climate and pseudo global warming future climate, Climate Dynamics, 57, 2191–2215, 2021.

IPCC: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland, https://doi.org/10.59327/IPCC/AR6-9789291691647, 2023.

Janowiak, J., Joyce, B., and Xie, P.: NCEP/CPC L3 Half Hourly 4km global (60S–60N) Merged IR V1. Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), 2017.

Judt, F. and Rios-Berrios, R.: Resolved convection improves the representation of equatorial waves and tropical rainfall variability in a global nonhydrostatic model, Geophysical Research Letters, 48, e2021GL093 265, 2021.

Judt, F., Klocke, D., Rios-Berrios, R., Vanniere, B., Ziemen, F., Auger, L., Biercamp, J., Bretherton, C., Chen, X., Düben, P., et al.: Tropical cyclones in global storm-resolving models, Journal of the Meteorological Society of Japan. Ser. II, 99, 579–602, 2021.

Jung, H. and Knippertz, P.: Link between the time-space behavior of rainfall and 3d dynamical structures of equatorial waves in global convection-permitting simulations, Geophysical Research Letters, 50, e2022GL100 973, 2023.

Kendon, E., Prein, A. F., Senior, C., and Stirling, A.: Challenges and outlook for convection-permitting climate modelling, Philosophical Transactions of the Royal Society A, 379, 20190 547, 2021.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M.: Do convection-permitting regional climate models improve projections of future precipitation change?, Bulletin of the American Meteorological Society, 98, 79–93, 2017.

Kiladis, G. N., Wheeler, M. C., Haertel, P. T., Straub, K. H., and Roundy, P. E.: Convectively coupled equatorial waves, Reviews of Geophysics, 47, 2009.

Kinne, S.: The MACv2 aerosol climatology, Tellus B: Chemical and Physical Meteorology, 71, 1–21, 2019.

Klocke, D., Frauen, C., Engels, J. F., Alexeev, D., Redler, R., Schnur, R., Haak, H., Kornblueh, L., Brüggemann, N., Chegini, F., et al.: Computing the Full Earth System at 1km Resolution, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 125–136, 2025.

Knaff, J. A. and Zehr, R. M.: Reexamination of tropical cyclone wind–pressure relationships, Weather and Forecasting, 22, 71–88, 2007.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data, Bulletin of the American Meteorological Society, 91, 363–376, 2010.

Lange, S.: Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset, Earth System Dynamics, 9, 627–645, 2018.

Langendijk, G., Rechid, D., Sieck, K., and Jacob, D.: Added value of convection-permitting simulations for understanding future urban humidity extremes: case studies for Berlin and its surroundings, Weather and Climate Extremes, 33, 100 367, 2021.

Lee, J. and Hohenegger, C.: Weaker land–atmosphere coupling in global storm-resolving simulation, Proceedings of the National Academy of Sciences, 121, e2314265 121, 2024.

Liebmann, B. and Smith, C. A.: Description of a Complete (Interpolated) Outgoing Longwave Radiation Dataset, Bulletin of the American Meteorological Society, 77, 1275–1277, http://www.jstor.org/stable/26233278, 1996.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., et al.: Continental-scale convection-permitting modeling of the current and future climate of North America, Climate Dynamics, 49, 71–95, 2017.

Lucas-Picher, P., Argüeso, D., Brisson, E., Tramblay, Y., Berg, P., Lemonsu, A., Kotlarski, S., and Caillaud, C.: Convection-permitting modeling with regional climate models: Latest developments and next steps, Wiley Interdisciplinary Reviews: Climate Change, 12, e731, 2021.

Lucas-Picher, P., Brisson, E., Caillaud, C., Alias, A., Nabat, P., Lemonsu, A., Poncet, N., Cortés Hernandez, V. E., Michau, Y., Doury, A., et al.: Evaluation of the convection-permitting regional climate model CNRM-AROME41t1 over Northwestern Europe, Climate Dynamics, 62, 4587–4615, 2024.

Mapes, B., Tulich, S., Lin, J., and Zuidema, P.: The mesoscale convection life cycle: Building block or prototype for large-scale tropical waves?, Dynamics of atmospheres and oceans, 42, 3–29, 2006.

Marsham, J. H., Dixon, N. S., Garcia-Carreras, L., Lister, G. M., Parker, D. J., Knippertz, P., and Birch, C. E.: The role of moist convection in the West African monsoon system: Insights from continental-scale convection-permitting simulations, Geophysical Research Letters, 40, 1843–1849, 2013.

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., Good, S. A., Mittaz, J., Rayner, N. A., Berry, D., et al.: Satellite-based time-series of sea-surface temperature since 1981 for climate applications, Scientific data, 6, 223, 2019.

Miura, H., Satoh, M., Nasuno, T., Noda, A. T., and Oouchi, K.: A Madden-Julian oscillation event realistically simulated by a global cloud-resolving model, Science, 318, 1763–1765, 2007a.

Miura, H., Satoh, M., Tomita, H., Noda, A. T., Nasuno, T., and Iga, S.-i.: A short-duration global cloud-resolving simulation with a realistic land and sea distribution, Geophysical research letters, 34, 2007b.

Miura, H., Suematsu, T., Kawai, Y., Yamagami, Y., Takasuka, D., Takano, Y., Hung, C.-S., Yamazaki, K., Kodama, C., Kajikawa, Y., et al.: Asymptotic matching between weather and climate models, Bulletin of the American Meteorological Society, 104, E2308–E2315, 2023.

Nakamura, Y. and Takayabu, Y. N.: Convective couplings with equatorial Rossby waves and equatorial Kelvin waves. Part I: Coupled wave structures, Journal of the Atmospheric Sciences, 79, 247–262, 2022.

Nasuno, T., Tomita, H., Iga, S., Miura, H., and Satoh, M.: Convectively coupled equatorial waves simulated on an aquaplanet in a global nonhydrostatic experiment, Journal of the atmospheric sciences, 65, 1246–1265, 2008.

North, R. C., Mittermaier, M. P., and Milton, S. F.: Using SEEPS with a TRMM-derived climatology to assess global NWP precipitation forecast skill, Monthly Weather Review, 150, 135–155, 2022.

Ortega, S., Segura, H., Mayta, V. C., Fiévet, R., Bravo, A. P., Lee, J., Giorgetta, M. A., and Stevens, B.: Convectively Coupled Equatorial Waves in a Global Storm-Resolving Model, Authorea Preprints, 2026.

Paredes, E. G., Groner, L., Ubbiali, S., Vogt, H., Madonna, A., Mariotti, K., Cruz, F., Benedicic, L., Bianco, M., VandeVondele, J., et al.: Gt4py: High performance stencils for weather and climate applications using python, arXiv preprint arXiv:2311.08322, 2023.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific data, 7, 225, 2020.

Perkins-Kirkpatrick, S. and Lewis, S.: Increasing trends in regional heatwaves, Nature communications, 11, 3357, 2020.

Pichelli, E., Coppola, E., Sobolowski, S., Ban, N., Giorgi, F., Stocchi, P., Alias, A., Belušić, D., Berthou, S., Caillaud, C., et al.: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation, Climate Dynamics, 56, 3581–3602, 2021.

Pothapakula, P. K., Prein, A. F., Sunkisala, A., and Dipankar, A.: Global Monsoon in ICON: The Scale-Dependent Response of Northern Hemisphere Monsoons, EGUsphere, 2026, 1–41, 2026.

Prein, A.: andreas-prein/icon2.5_dyamond3_paper: vo.1.0.0, https://doi.org/10.5281/zenodo.18007290, 2025.

Prein, A. F.: Thunderstorm straight line winds intensify with climate change, Nature Climate Change, 13, 1353–1359, 2023.

Prein, A. F.: Data used in the publication: From Single Storms to Global Waves: A Global 2.5 km ICON Simulation of Weather and Climate, https://doi.org/10.6084/m9.figshare.31341982, 2026.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, International Journal of Climatology, 37, 305–327, 2017.

Prein, A. F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N. K., Keuler, K., and Georgievski, G.: Added value of convection permitting seasonal simulations, Climate Dynamics, 41, 2655–2677, 2013.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., et al.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Reviews of geophysics, 53, 323–361, 2015.

Prein, A. F., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., and Holland, G. J.: The future intensification of hourly precipitation extremes, Nature climate change, 7, 48–52, 2017.

Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., and Clark, M.: Simulating North American mesoscale convective systems with a convection-permitting climate model, Climate Dynamics, 55, 95–110, 2020.

Prein, A. F., Rasmussen, R., Wang, D., and Giangrande, S.: Sensitivity of organized convective storms to model grid spacing in current and future climates, Philosophical Transactions of the Royal Society A, 379, 20190 546, 2021.

Prein, A. F., Ge, M., Valle, A. R., Wang, D., and Giangrande, S. E.: Towards a unified setup to simulate mid-latitude and tropical mesoscale convective systems at kilometer-scales, Earth and Space Science, 9, e2022EA002 295, 2022.

Prein, A. F., Ban, N., Ou, T., Tang, J., Sakaguchi, K., Collier, E., Jayanarayanan, S., Li, L., Sobolowski, S., Chen, X., et al.: Towards ensemble-based kilometer-scale climate simulations over the third pole region, Climate Dynamics, 60, 4055–4081, 2023a.

Prein, A. F., Mooney, P. A., and Done, J. M.: The multi-scale interactions of atmospheric phenomenon in mean and extreme precipitation, Earth's Future, 11, e2023EF003 534, 2023b.

Prein, A. F., Feng, Z., Fiolleau, T., Moon, Z. L., Núñez Ocasio, K. M., Kukulies, J., Roca, R., Varble, A. C., Rehbein, A., Liu, C., et al.: Km-scale simulations of mesoscale convective systems over South America—A feature tracker intercomparison, Journal of Geophysical Research: Atmospheres, 129, e2023JD040 254, 2024.

Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., Bechtold, P., Beyer, S., Bidlot, J., Boussetta, S., et al.: Multi-year simulations at kilometre scale with the Integrated Forecasting System coupled to FESOM2. 5 and NEMOv3.4, Geoscientific Model Development, 18, 33–69, 2025.

Randall, D. A. and Emanuel, K.: The weather–climate schism, Bulletin of the American Meteorological Society, 105, E300–E305, 2024.

Raschendorfer, M., Simmer, C., and Gross, P.: Parameterisation of turbulent transport in the atmosphere, in: Dynamics of multiscale earth systems, pp. 167–185, Springer, 2003.

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., et al.: Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble, Journal of Climate, 33, 2557–2583, 2020.

Sakradzija, M., Senf, F., Scheck, L., Ahlgrimm, M., and Klocke, D.: Local impact of stochastic shallow convection on clouds and precipitation in the tropical Atlantic, Monthly Weather Review, 148, 5041–5062, 2020.

Sato, T., Miura, H., Satoh, M., Takayabu, Y. N., and Wang, Y.: Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model, Journal of Climate, 22, 4809–4826, 2009.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.-I.: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations, Journal of Computational Physics, 227, 3486–3514, 2008.

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

Savre, J. and Craig, G.: Fitting cumulus cloud size distributions from idealized cloud resolving model simulations, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 360, 2023.

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

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Di Girolamo, S., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., et al.: Kilometer-scale climate models: Prospects and challenges, Bulletin of the American Meteorological Society, 101, E567–E587, 2020.

Schellekens, J., Dutra, E., Martínez-de La Torre, A., Balsamo, G., Van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., et al.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth System Science Data, 9, 389–413, https://doi.org/10.5194/essd-9-389-2017, 2017.

Schenkel, B. A. and Hart, R. E.: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets, Journal of Climate, 25, 3453–3475, 2012.

Schlemmer, L., Schär, C., Lüthi, D., and Strebel, L.: A groundwater and runoff formulation for weather and climate models, Journal of Advances in Modeling Earth Systems, 10, 1809–1832, 2018.

Schmidli, J., Böing, S., and Fuhrer, O.: Accuracy of simulated diurnal valley winds in the Swiss Alps: Influence of grid resolution, topography filtering, and land surface datasets, Atmosphere, 9, 196, 2018.

Schrodin, R. and Heise, E.: The multi-layer version of the DWD soil model TERRA_LM, DWD Offenbach, Germany, 2001.

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

Segura, H. and Hohenegger, C.: How do the tropics precipitate? Daily variations in precipitation and cloud distribution, Journal of the Meteorological Society of Japan. Ser. II, 102, 525–537, 2024.

Segura, H., Hohenegger, C., Wengel, C., and Stevens, B.: Learning by doing: Seasonal and diurnal features of tropical precipitation in a global-coupled storm-resolving model, Geophysical Research Letters, 49, e2022GL101 796, 2022.

Segura, H., Bayley, C., Fievét, R., Glöckner, H., Günther, M., Kluft, L., Naumann, A., Ortega, S., Praturi, D., Rixen, M., et al.: A single tropical rainbelt in global storm-resolving models: The role of surface heat fluxes over the warm pool, Journal of Advances in Modeling Earth Systems, 17, e2024MS004 897, 2025a.

Segura, H., Pedruzo-Bagazgoitia, X., Weiss, P., Müller, S. K., Rackow, T., Lee, J., Dolores-Tesillos, E., Benedict, I., Aengenheyster, M., Aguridan, R., et al.: nextGEMS: entering the era of kilometer-scale Earth system modeling, EGUsphere, 2025, 1–39, 2025b.

Seifert, A.: A revised cloud microphysical parameterization for COSMO-LME, COSMO Newsletter 7, Consortium for Small-Scale Modelling, http://www.cosmo-model.org, proceedings from the 8th COSMO General Meeting, Bucharest, 2006, 2008.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A. D., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., et al.: Weather and climate extreme events in a changing climate, 2021.

Soares, P. M., Careto, J. A., Cardoso, R. M., Goergen, K., Katragkou, E., Sobolowski, S., Coppola, E., Ban, N., Belušić, D., Berthou, S., et al.: The added value of km-scale simulations to describe temperature over complex orography: the CORDEX FPS-Convection multi-model ensemble runs over the Alps, Climate Dynamics, 62, 4491–4514, 2024.

Song, Y., Broxton, P. D., Ehsani, M. R., and Behrangi, A.: Assessment of snowfall accumulation from satellite and reanalysis products using SNOTEL observations in Alaska, Remote Sensing, 13, 2922, 2021.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., et al.: DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains, Progress in Earth and Planetary Science, 6, 1–17, 2019.

Takasuka, D., Kodama, C., Suematsu, T., Ohno, T., Yamada, Y., Seiki, T., Yashiro, H., Nakano, M., Miura, H., Noda, A. T., et al.: How can we improve the seamless representation of climatological statistics and weather toward reliable global K-scale climate simulations?, Journal of Advances in Modeling Earth Systems, 16, e2023MS003 701, 2024a.

Takasuka, D., Satoh, M., Miyakawa, T., Kodama, C., Klocke, D., Stevens, B., Vidale, P. L., and Terai, C. R.: A protocol and analysis of year-long simulations of global storm-resolving models and beyond, Progress in Earth and Planetary Science, 11, 66, 2024b.

Takasuka, D., Becker, T., and Bao, J.: Precipitation characteristics and thermodynamic-convection coupling in global kilometer-scale simulations, Journal of Advances in Modeling Earth Systems, 18, e2025MS005 343, 2026.

Takayabu, Y. N.: Large-scale cloud disturbances associated with equatorial waves Part I: Spectral features of the cloud disturbances, Journal of the Meteorological Society of Japan. Ser. II, 72, 433–449, 1994.

Taylor, M., Caldwell, P. M., Bertagna, L., Clevenger, C., Donahue, A., Foucar, J., Guba, O., Hillman, B., Keen, N., Krishna, J., et al.: The simple cloud-resolving E3SM atmosphere model running on the Frontier exascale system, in: Proceedings of the international conference for high performance computing, networking, storage and analysis, pp. 1–11, 2023.

Tomita, H., Miura, H., Iga, S.-i., Nasuno, T., and Satoh, M.: A global cloud-resolving simulation: Preliminary results from an aqua planet experiment, Geophysical Research Letters, 32, 2005.

Trenberth, K. E.: 24.5-Year Surface Radiation Budget Data Set Released, 2011.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature methods, 17, 261–272, 2020.

Weber, N. J. and Mass, C. F.: Subseasonal weather prediction in a global convection-permitting model, Bulletin of the American Meteorological Society, 100, 1079–1089, 2019.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505–7514, https://doi.org/https://doi.org/10.1002/2014WR015638, 2014.

Wheeler, M. and Kiladis, G. N.: Convectively Coupled Equatorial Waves: Analysis of Clouds and Temperature in the Wavenumber–Frequency Domain, Journal of the Atmospheric Sciences, 56, 374 – 399, https://doi.org/10.1175/1520-0469(1999)056<0374:CCEWAO>2.0.CO;2, 1999.

World Climate Research Programme: Report of the WCRP km-scale modeling workshop, 3–7 October 2022, hybrid format, Tech. Rep. 08/2022, World Climate Research Programme (WCRP), Geneva, Switzerland, https://www.wcrp-climate.org/WCRP-publications/2022/WCRP_Report_08-2022_k-scale-report-final.pdf, 2022.

Yasunaga, K. and Mapes, B.: Differences between more divergent and more rotational types of convectively coupled equatorial waves. Part I: Space–time spectral analyses, Journal of the atmospheric sciences, 69, 3–16, 2012.

Yasunaga, K., Yokoi, S., Inoue, K., and Mapes, B. E.: Space–time spectral analysis of the moist static energy budget equation, Journal of Climate, 32, 501–529, 2019.

Yu, H., Prein, A. F., Qi, D., and Wang, K.: Mesoscale convective systems in Northeast China from satellite products, global reanalysis, and kilometer-scale modeling, Geophysical Research Letters, 52, e2024GL112 349, 2025.

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

Zhang, Z., Varble, A. C., Feng, Z., Marquis, J. N., Hardin, J. C., and Zipser, E. J.: Dependencies of simulated convective cell and system growth biases on atmospheric instability and model resolution, Journal of Geophysical Research: Atmospheres, 129, e2024JD041 090, 2024.