JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005343

Precipitation Characteristics and Thermodynamic-Convection Coupling in Global Kilometer-Scale Simulations

Daisuke Takasuka[1], Tobias Becker[2], and Jiawei Bao[3]

[1]Department of Geophysics, Graduate School of Science, Tohoku University, Sendai, Japan, [2]Earth System Modelling, European Centre for Medium-Range Weather Forecasts (ECMWF), Bonn, Germany, [3]Institute of Science and Technology Austria, Klosterneuburg, Austria

Abstract We compare three global kilometer-scale models (ICON, IFS and NICAM) to clarify the advantages and challenges of high-resolution global weather and climate modeling, using different approaches to represent convection, from fully parameterized to fully explicit. Our analysis focuses on tropical precipitation characteristics spanning a wide range of spatio-temporal scales—including the diurnal cycle, extreme precipitation, convective organization, and the Madden-Julian Oscillation (MJO)—along with interactions between convection and the thermodynamic environment. All three models commonly show weaker convective organization with smaller precipitation cells than observed, though the strength of the bias varies by model. This diversity is introduced by differences in the representation of (a) convective initiation affected by the convective sensitivity to moisture and (b) tropospheric moistening associated with deep convection. Models with stronger thermodynamic-convection coupling increase environmental moisture near convection, thereby enhancing convective organization. This has important upscale effects on the MJO; while IFS and NICAM capture its eastward propagation well, ICON has difficulty reproducing it. The amplitudes and phases of precipitation diurnal cycles over land show much greater disagreement among the models than over ocean, influenced by how convection is initiated. Biases in rain evaporation and cold pool formation hinder the propagation of mesoscale convection, leading to errors such as the misrepresentation of nocturnal convection moving off the coast of Sumatra in IFS and ICON. These results highlight the importance of thermodynamic-convection coupling in realistically simulating tropical convection across scales. To improve this coupling, kilometer-scale models require better representation of the interaction between resolved convection and three-dimensional turbulent mixing.

Plain Language Summary Thunderstorms play a key role in tropical weather, affecting everything from local rainfall to large-scale weather patterns. The size of tropical thunderstorms ranges from a kilometer to hundreds of kilometers, meaning that models with kilometer-scale grid sizes can begin to resolve the larger structures and governing processes of most thunderstorms. We analyze how three global kilometer-scale models-ICON, IFS, and NICAM-simulate thunderstorms, daily rainfall, and a major tropical weather system called the Madden-Julian Oscillation (MJO). While all models can represent individual storms, they struggle with storm clustering, often producing too many small, disorganized cells. The extent of this issue varies among models, depending on how well each captures the link between environmental moisture and storm development. Better representation of this moisture-convection relationship leads to larger storm clusters and more realistic MJO signals. Another key factor is the simulation of rain-cooled air that spreads at the surface. Weak cold pools hinder storm propagation, especially at night over tropical coasts, degrading storm organization and movement. These limitations highlight the need for improved representations of storm dynamics in global models. Improving how models handle small-scale turbulence and buoyant plumes will enhance their ability to simulate tropical rainfall variability and large-scale climate patterns more realistically.

Daisuke Takasuka, Tobias Becker and Jiawei Bao contributed equally to this work.

Key Points:

- Km-scale models exhibit common biases: overly intense precipitation, undersized convective cells, and insufficient convective organization
- Magnitude of common biases varies, reflecting differences in initiation of convection, vertical moisture transport, and convection longevity
- Biased thermodynamic-convection coupling affects the representation of atmospheric phenomena such as precipitation diurnal cycles and Madden-Julian Oscillation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. Takasuka, T. Becker and J. Bao,
takasuka@tohoku.ac.jp;
tobias.becker@ecmwf.int;
jiawei.bao@ist.ac.at

Citation:

Takasuka, D., Becker, T., & Bao, J. (2026). Precipitation characteristics and thermodynamic-convection coupling in global kilometer-scale simulations. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005343. https://doi.org/10.1029/2025MS005343

Received 7 JUL 2025
Accepted 13 FEB 2026

1. Introduction

Atmospheric moist convection drives the vertical transport of momentum, heat and moisture, which in turn affects large-scale circulation patterns, cloud formation, and precipitation processes (e.g., Riehl & Malkus, 1958). These processes are fundamental to the Earth's energy balance and hydrological cycle, influencing everything from daily weather patterns to long-term climate variability. Cumulus parameterizations, which constrain the complex three-dimensional nature of atmospheric convection to a 1D column framework (e.g., Arakawa & Schubert, 1974), are of critical importance in current state-of-the-art global numerical weather prediction and climate models, as they

are typically run at resolutions that are too coarse to resolve convection (10–100 km). Despite extensive efforts to improve convection schemes, they continue to face several fundamental issues that challenge their accuracy and reliability, and progress in resolving these issues has been slow (e.g., Randall et al., 2003; Stephens et al., 2010; Stevens, 2024). One crucial issue in cumulus parameterizations is related to the convective triggering assumption, which determines when and where convection initiates, often based on atmospheric instability and moisture content (e.g., Kain & Fritsch, 1990). Such criteria can lead to significant uncertainties as they may not accurately capture the variability and complexity of convective initiation, especially when synoptic-scale forcing is weak and surface heterogeneity becomes an important factor (e.g., J. Chen et al., 2020; Hanesiak et al., 2004). Persistent issues also arise around the closure assumptions. For example, Convective Available Potential Energy (CAPE) closures often rely on prescribed time scales for energy removal (e.g., Moorthi & Suarez, 1992; G. J. Zhang & McFarlane, 1995), which may not dynamically reflect actual atmospheric conditions, resulting in errors in the intensity, diurnal cycle and spatial organization of convection (e.g., Bechtold et al., 2008; Mapes & Neale, 2011; Yang & Slingo, 2001). These biases can have cascading effects on weather predictions and climate projections, underscoring the need for continued research on atmospheric convection.

Due to the uncertain nature of convective parameterizations, many modeling groups have decided to simulate convection explicitly for many decades, using convection-permitting resolutions for example in idealized setups (e.g., radiative-convective equilibrium in three-dimensional models with cyclic boundary conditions, Tompkins & Craig, 1998), or in regional models, nested into a coarser-resolution global model (e.g., Dickinson et al., 1989). Their results have been promising, especially with respect to the timing and distribution of intense precipitation (e.g., Giorgi, 2019). However, while certain characteristics of local convection and precipitation are better represented, their interactions with large-scale atmospheric processes are either missing or remain poorly captured. It is therefore desirable to run global simulations with such a fine resolution, enabling the two-way interaction of convection with its environment across all scales, from local to global. In fact, several pioneering studies performed with the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) (Satoh et al., 2008, 2014; Tomita & Satoh, 2004) showed that the framework with globally explicit convection succeeded in reproducing the multi-scale structure, ranging from individual deep clouds to large-scale circulations (e.g., Miura et al., 2007; Tomita et al., 2005).

With global convection-permitting climate models, since there is no artificial scale separation that originates from the cumulus parameterizations, the multi-scale relationships are easier to understand (Stevens et al., 2020), and because of the upscale impact of resolved convection, the mean climatological state and large-scale climate variability are expected to improve. However, realizing this improvement is not always straightforward (e.g., Takasuka, Kodama, et al., 2024). This is because many processes relevant for convection remain unresolved and/or underresolved at kilometer (km)-scale resolutions without cumulus parameterizations. For example, the boundary layer processes that are relevant for the triggering of convection are mostly unresolved at km-scale resolutions (e.g., Hirt et al., 2020). The representation of the coupling between boundary layer processes and precipitation can influence the Intertropical Convergence Zone (ITCZ), for example, cause a double-peak structure (Bacmeister et al., 2006), which has been known as a long-standing climatological bias in conventional climate simulations (e.g., Dai, 2006). Another example is that the mixing of grid-scale updrafts with their environment is underresolved (e.g., Lebo & Morrison, 2015), meaning that updraft characteristics and thermodynamic-convection relationships can be biased. Takasuka, Kodama, et al. (2024) proposed increasing updraft mixing with the environment by including the Leonard terms (Moeng et al., 2010), which take mixing on the grid scale into account, to reduce the bias of too much mean precipitation in the tropics. The motivation for this model change is that convective precipitation is too intense, too localized and too short-lived and consequently not organized enough in km-scale resolution models with explicit convection (e.g., Becker et al., 2021; Brousseau et al., 2016; Crook et al., 2019).

Given these unresolved processes and their thermodynamic impacts on convection, their representation in global kilometer-scale models is critical for simulating the formation and organization of convective systems across scales. The onset of convection tends to moisten the surrounding environment through vertical advection and mixing, and a moist environment caused by earlier convection promotes the development of new convection in its vicinity, while a dry environment inhibits convection because updrafts that entrain dry air rapidly lose their buoyancy. Thus, the degree to which convection can moisten its surrounding environment plays an important role in the formation and life cycle of organized convective systems. Convective organization can modify this relationship, as environmental humidity is expected to increase around organized convection, thereby decreasing the

efficiency of entrainment in reducing updraft buoyancy (Bao & Sherwood, 2019; Becker et al., 2018). Hence, a stronger coupling of convection and environmental humidity facilitates convective organization. This aspect is crucially important to the realization of large-scale convective organized systems such as the Madden–Julian oscillation (MJO, Madden & Julian, 1972), which has far-reaching impacts on weather and climate patterns and thus should be represented realistically by global models. Another example that influences the thermodynamic-convection relationship is rain evaporation. Realistic rain evaporation is essential for producing realistic cold pools, which in turn is crucial for convective organization, both on the mesoscale and on larger scales, for example, with regard to the width of the ITCZ (Windmiller & Stevens, 2024).

A main objective of our study is thus to better understand how misrepresentations in thermodynamic-convection coupling are responsible for biases in precipitating convection across a wide range of spatio-temporal scales in different global km-scale models. By doing so, we aim to identify which aspects of the remaining moist physics parameterizations need improvement to fully leverage the benefits of increased resolution. Issues with representing convective organization in km-scale models and its relation to model physics have already been discussed within regional modeling frameworks (e.g., Ding et al., 2023; Holloway et al., 2013; Prein et al., 2021), but they cannot fully represent the interaction between mesoscale convection and its environment across the full range of spatial scales. They also depart from strict radiative–convective equilibrium because they are constrained by lateral boundary conditions. In addition, regional models typically focus on limited domains, often over land and outside the deep tropics. In contrast, global kilometer-scale models allow for a seamless representation of the entire tropical belt, enabling self-consistent coupling between convection, radiation, and large-scale circulation. We will primarily focus on both the governing physical processes related to moist convection and their impacts across scales—from the diurnal cycle and extreme precipitation to the MJO and mean precipitation. The present study provides an overview of this convective hierarchy, while detailed analyses of each specific phenomenon are left for future work.

We will compare the models ICON and NICAM, which both explicitly represent both shallow and deep convection, with IFS-FESOM, in which a deep convection parameterization is still weakly active, while the shallow convection parameterization is fully active. The ICON and IFS-FESOM simulations stem from the H2020 Next Generation Earth Modeling Systems (nextGEMS) project (Rackow et al., 2025; Segura et al., 2025). Publications related to nextGEMS have already highlighted the advantages of km-scale models in simulating various aspects of the climate system, including near-surface air temperature (Brunner et al., 2025), precipitation characteristics (Bao et al., 2024; Spät et al., 2024; Wille et al., 2025), tropical cyclone intensification (Baker et al., 2024) and soil-moisture coupling (Lee & Hohenegger, 2024). A key objective of our study is to understand the advantages and challenges of the respective ways in which precipitating convection is represented in the models and to explore potential strategies that improve model performance. A first glance at snapshots of precipitation in these three models indicates that the models develop biases in precipitation characteristics similar to the previous studies mentioned above, relative to the satellite retrieval GPM IMERG (Figure 1).

The paper is organized as follows. Section 2 introduces the model setups of the three global km-scale models that are analyzed, ICON, IFS and NICAM. After Section 3 briefly describes the thermodynamic tropical mean state in the three models, we analyze the main characteristics of tropical deep convection and its organization in Section 4. To understand these characteristics, we investigate the evolution of thermodynamic quantities during deep convection in Section 5, and investigate how thermodynamic-convection coupling affects specific atmospheric phenomena on weather and climate time scales in Section 6. A summary and concluding remarks are provided in Section 7.

2. Models, Simulation Setups, and Observational Data

2.1. Simulation Setups

The ICON and IFS simulations were performed within the nextGEMS project (Rackow et al., 2025; Segura et al., 2025), which aims at running multi-decadal coupled simulations at km-scale resolutions. This ambitious goal required cycles of development. Here we focus on the simulations from Cycle 3, which were performed at a resolution of 5 km for ICON and 4.4 km for the IFS. The simulations are 2020 control runs, meaning that the boundary conditions for 2020, for example, greenhouse gas concentrations and aerosol forcings, are used for each year. The ocean models (FESOM and ICON-Ocean) were spun up for a duration of 5 years in standalone mode, using ERA5 atmospheric forcing (Hersbach et al., 2020), followed by 5 years of fully-coupled free-running simulations, which started on 20 January 2020. Because some of the simulations show a drift in global mean 2 m

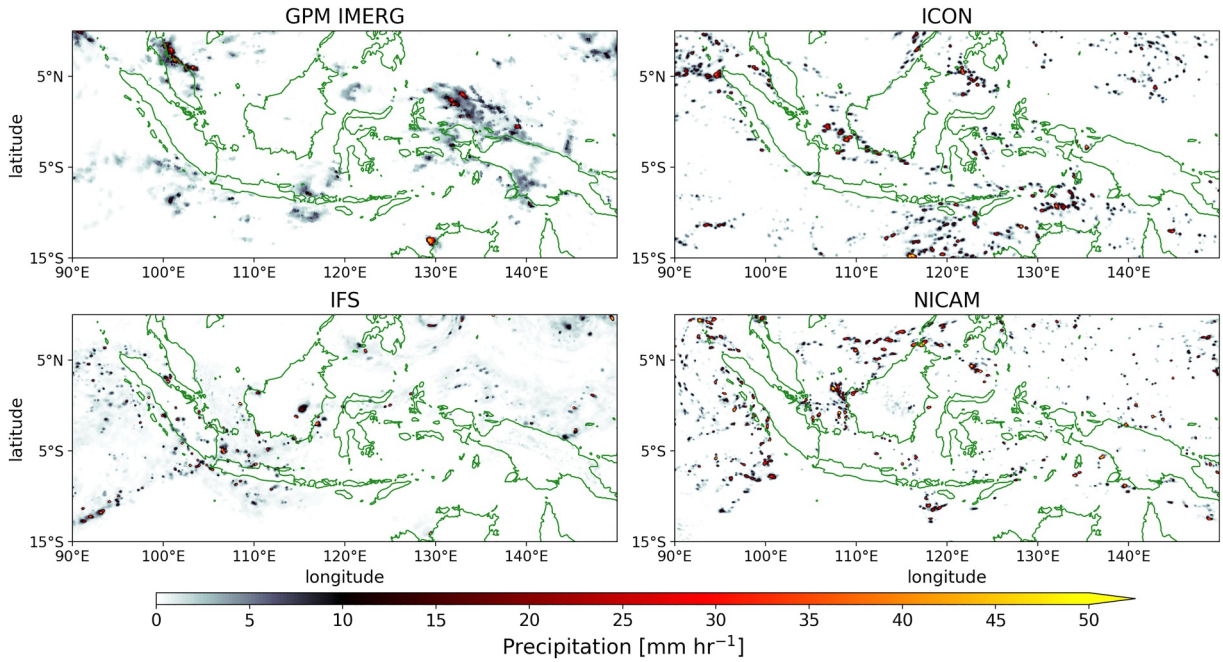

Figure 1. Snapshots of hourly precipitation over the Maritime Continent on 1 December 2012 (NICAM) and 2020 (IMERG, ICON, IFS), after the simulations have run free for months. Snapshots are from 0 to 1 UTC, which locally corresponds to morning hours.

temperature over time (Rackow et al., 2025; Segura et al., 2025), we here only analyze results from the first simulated year, more specifically February–December 2020, unless otherwise noted.

The NICAM simulation is part of an AMIP-type 10-year (2011–2020) simulation at 3.5 km resolution that stems from the DNA climate project (Miura et al., 2023), which aims at bridging the gap between weather and climate modeling. The atmosphere is initialized with ERA-Interim (Dee et al., 2011) on 1 January 2011, and the oceanic initial condition is obtained from the OSTIA product (Good et al., 2020). The sea surface temperature (SST) is predicted by a mixed-layer slab ocean model with a depth of 15 m, and it is nudged to the daily OSTIA data with a relaxation time of 7 days. The greenhouse gas and ozone concentrations and natural aerosol forcing follow the HighResMIP protocol (Kodama et al., 2021), and gas concentrations after 2015 are adopted from the SSP8.5 scenario because of the lack of data. Considering these differences compared to the ICON and IFS simulations, we analyze NICAM results from February to December 2012, unless otherwise noted, since this period is characterized by similar La Niña conditions as 2020.

2.2. Dynamics and Physics Configurations of the Models

The ICON model was originally developed by the Max Planck Institute for Meteorology (MPI-M) for climate simulations at coarse resolutions (100 km) and by the German Weather Service (DWD) for numerical weather forecasts. It is an Earth System Model (ESM) with three components: atmosphere, land and ocean. The high-resolution version of ICON (ICON-sapphire; Hohenegger et al., 2023) has been developed in parallel to better represent the convective processes and storms. The dynamical core of the atmosphere solves the three-dimensional nonhydrostatic version of the Navier-Stokes equations. The horizontal grid discretization is performed on an icosahedral triangular C grid. The parameterizations of ICON-sapphire are limited to radiation (Pincus et al., 2019), cloud microphysics (Baldauf et al., 2011) and turbulence (Dipankar et al., 2015; Lee et al., 2022). Neither a convective parameterization (both deep and shallow) nor a gravity-wave drag scheme is employed. In nextGEMs Cycle 3, the main improvement compared to previous cycles is a reduction in the atmospheric energy leak, which was achieved by fixing bugs in the cloud microphysics and turbulence scheme (Segura et al., 2025).

The IFS, developed and maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF), is coupled in nextGEMS to the Finite VolumE Sea Ice-Ocean Model (FESOM2; Danilov et al., 2017; Scholz et al., 2019), developed and maintained by the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine

Table 1
Key Differences in Model Configurations Between the Three Models

Model	ICON	IFS	NICAM
Years	2020–2024 (here 2020)	2020–2024 (here 2020)	2011–2020 (here 2012)
Resolution	5 km	4.4 km	3.5 km
Time step (Dyn/Phys/Rad)	8 s/40 s/720 s	180 s/180 s/3,600 s	10 s/10 s/300 s
Ocean Model	ICON	FESOM	Slab ocean with nudging
Ocean Res.	5 km	<5–13 km	—
Ocean grid	Icosahedral–triangular C	NG5	—
Atm. grid	Icosahedral–triangular C	Cubic-octahedral red. gaussian	Icosahedral A
Dynamics	Two-time-level Predictor–corrector scheme	Spectral transform model + Semi-Lagrangian semi-implicit	Time splitting + Runge-Kutta 3 Upwind-biased conservative adv.
Convection	—	Fully (partly) parameterized shallow (deep) convection	—
Turbulence	Smagorinsky	Eddy-diffusivity mass-flux (EDMF)	Eddy-diffusivity + Leonard terms
Microphysics	Single moment $(q_v, q_l, q_i, q_r, q_s, q_g)$	Single moment $(q_v, q_l, q_i, q_r, q_s)$	Single moment $(q_v, q_l, q_i, q_r, q_s, q_g)$
Clouds	—	Prognostic cloud fraction	—
Radiation	RTE-RRTMGP	ecRad	MSTRNX
Gravity wave drag	—	Orographic + non-orographic	—
Reference	Hohenegger et al. (2023)	Rackow et al. (2025)	Takasuka, Kodama, et al. (2024)

Note. The time step is given for the dynamics (Dyn), all physics schemes excluding radiation (Phys) and radiation scheme (Rad).

Research (AWI). The IFS utilizes a full range of physics parameterizations, as briefly described in Table 1, with more details given in ECMWF (2023) and Rackow et al. (2025). The nextGEMS Cycle 3 simulations are based on IFS Cycle 48r1 but with some modifications to enable km-scale multidecadal simulations. Most importantly, the cloud and microphysics are tuned so that the model provides a realistic top-of-atmosphere radiation balance, and to allow for more explicit convection, the deep convection parameterization (Bechtold et al., 2008, 2014; Becker et al., 2021; Tiedtke, 1989) is tempered by reducing the cloud base mass flux by a factor of 5 compared to its default value at 4.4 km resolution. In addition to analyzing the main IFS-FESOM Cycle 3 simulation at 4.4 km resolution, we will also compare to some IFS sensitivity experiments at 9 km resolution, which were coupled to an ocean at 0.25° spatial resolution (NEMO v3.4). These simulations, which were also started on 20 January 2020, were run for at least one year. They differ in the convection scheme setup. Deep convection is either fully active as in the operational IFS (D_{on}), the cloud base mass flux is reduced as in the 4.4 km simulation (D-RCBMF$_{on}$), the deep convection scheme is switched off (D_{off}) or both shallow and deep convection scheme are switched off ($(D + S)_{off}$). This leads to significant differences in the characteristics of convection, as the precipitation snapshots in Figure S1 in Supporting Information S1.

NICAM is now maintained by some members in Japanese research institutes and universities. It adopts a set of fully compressible nonhydrostatic dynamical equations discretized by an icosahedral A-grid system (Tomita et al., 2002; Tomita & Satoh, 2004). As in ICON, NICAM has not employed any deep- and shallow-cumulus parameterizations and gravity-wave drag schemes. Thus, convective processes are explicitly represented by the cloud microphysics and turbulence schemes coupled to dynamics. Cloud microphysics is calculated using the NICAM Single-Moment Water 6 scheme, which was originally developed by Tomita (2008) and then improved by Roh and Satoh (2014) based on satellite observations. For subgrid-scale turbulent mixing, we adopt the hybrid usage of the Mellor-Yamada-Nakanishi-Niino level 2 (MYNN2) eddy-diffusivity scheme (Nakanishi & Niino, 2006) and the Leonard terms (Moeng et al., 2010), which can partially treat lateral mixing associated with convection. Some parameters in the microphysics scheme are tuned to achieve realistic simulations of both climatological statistics and weather disturbances. Its details and other physics configurations are described in Takasuka, Kodama, et al. (2024).

2.3. Observational Data

To validate the three models, we use the following observational data sets. Note that a conservative remapping method is used when interpolating the original data to a common grid.

1. The hourly mean rainfall data from the satellite-based, gauge-calibrated and globally-gridded Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) V6B (Tan et al., 2019) is used to evaluate characteristics of precipitation and convection (Sections 4 and 5). The original horizontal resolution is 0.1° × 0.1°, and it is interpolated to a 0.25°- and 1° grid in some analyses that focus on the precipitation diurnal cycle (PDC) and mesoscale (i.e., $O(10^2-10^3$ km)) variations. IMERG precipitation is retrieved from passive microwave and infrared sensors on low Earth orbit and geosynchronous Earth orbit satellites, respectively, which enables us to use frequent and high-resolution precipitation data. However, IMERG comes with some significant uncertainties as there are some caveats in the data quality such as the overestimation of the duration and the underestimation of the intensity in strong convective systems (e.g., Maranan et al., 2020).
2. As a second data set for precipitation, we use the hourly climatology of TRMM-3G68 from December 1997 to March 2015, provided by Minobe et al. (2020) to obtain the statistics of PDC (Section 6.1). TRMM-3G68 is retrieved from the satellite-borne precipitation radar, so we can avoid the issue that the infrared-based precipitation retrieval (e.g., IMERG) does not capture the phase of PDC accurately (e.g., Kikuchi & Wang, 2008). This data is interpolated to the 0.25° × 0.25° grid.
3. Daily outgoing longwave radiation (OLR) data from the National Oceanic and Atmospheric Administration (NOAA) satellite (Liebmann & Smith, 1996) is used to locate deep convection. Specifically, this data set is used in a statistical analysis of the MJO in Section 6.2. The horizontal resolution is 2.5° × 2.5°.
4. 6-hourly ERA5 reanalysis data (Hersbach et al., 2020) is used to analyze the thermodynamic variability associated with the life cycle of convection (Section 5) and the MJO propagation (Section 6.2). The original 0.25°-gridded data are interpolated to 1°-, 1.5°-, and 2.5° grids depending on analyses. We use the three-dimensional variables of temperature and specific humidity, spanning from 1,000 to 100 hPa, and 850-hPa zonal winds. Daily mean values are also derived from the 6-hourly snapshots.
5. In-situ surface meteorology observations collected by Tropical Atmosphere Ocean (TAO) buoys are used to examine variations of surface air temperature in precipitation events. We use the data at three buoys deployed at (2°N, 165°E) (8°S, 165°E), and (0°N, 180°E). Precipitation and surface air temperature, which are observed in 10-min intervals, are averaged in time to create hourly means.

Although most of the analysis periods of the observational data follow those in the simulations; that is, February–December 2020 and/or 2012, two exceptions are in the analyses of the MJO and TAO buoy observations. To ensure robust MJO statistics, the OLR and ERA5 data sets span the period from 2011 to 2015, and the analysis method follows Takasuka, Kohyama, et al. (2025). The buoy data set uses the period of February–December 2010, when all precipitation and surface air temperature observations are available for the three buoys. We expect that these deviations from the time periods analyzed in the three models do not impact the interpretation of the results, given the free-running nature of the models and the specifics of our performed analyses, which to a large degree should not be significantly different in another year.

3. Mean Profiles of Tropical Thermodynamic Fields

Since tropical convection both influences and is influenced by the large-scale atmospheric background state, we now briefly compare tropical mean profiles of temperature, water vapor, and hydrometers across the three models (Figure 2), before taking a more local perspective on convection in the following sections. All variables are shown as their differences from ERA5. Note that the amount of hydrometers in ERA5 does not necessarily represent reality, so it is only used as references that make model-to-model differences clearer.

Because ICON and IFS are both coupled to an ocean model, their global mean temperature can slowly drift over time. Figure 2 shows that ICON drifts toward a colder climate, with an amplified cooling in the upper troposphere and an associated decrease in static stability. In line with the cooling, specific humidity is reduced, while there is more liquid water in the lower troposphere than in IFS and ICON, indicating significantly more clouds. In many regards, NICAM's large-scale state is opposite to ICON. Even though the surface temperature over ocean is constrained through the mixed-layer slab ocean model, the temperature is higher in the upper troposphere, indicating a more stable free troposphere. In addition, the whole troposphere is moister, with much more abundant

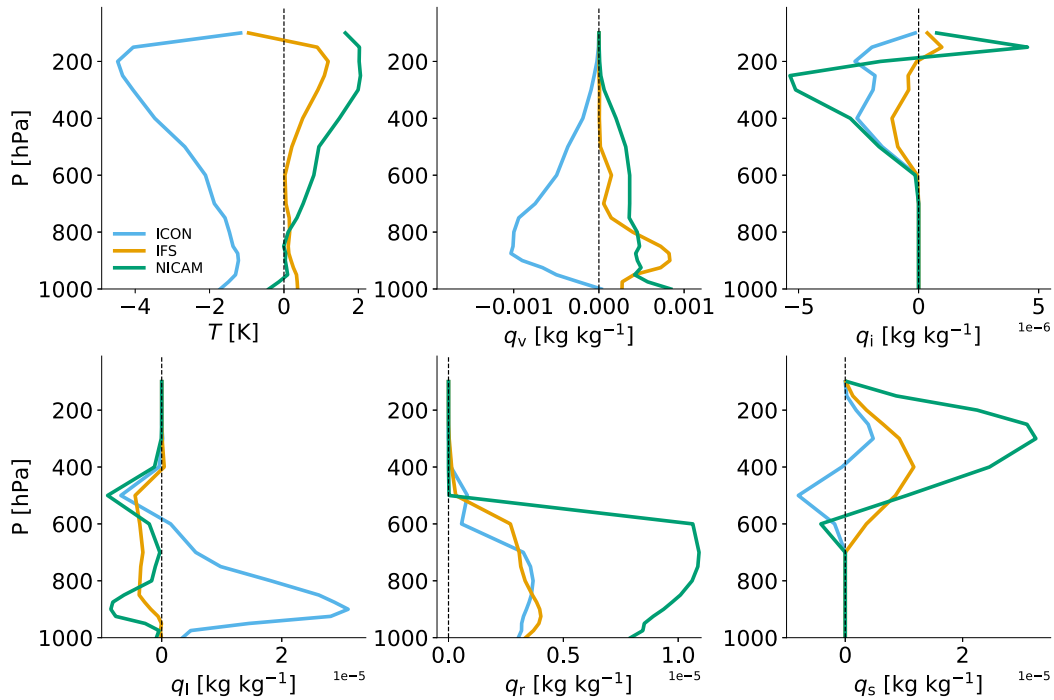

Figure 2. Differences in the tropical (20°S–20°N) mean profiles of temperature (T), specific humidity (q_v), specific cloud ice content (q_i), specific cloud liquid water content (q_l), specific rain water content (q_r) and specific snow water content (q_s) between models and ERA5. The differences are computed for the year 2020 for ICON and IFS and for the year 2012 for NICAM.

rain and ice water content (even total of snow and cloud ice), while cloud liquid water content is slightly lower than in ERA5. IFS lies in between ICON and NICAM for most of the metrics, closest to ERA5. In part, this is because of a realistic top-of-atmosphere radiation balance which ensures that the IFS simulation does not drift much over time (Rackow et al., 2025). The different versions of the IFS in Figure S2 in Supporting Information S1 show that the simulation with $(D + S)_{off}$ has the most significant biases, most importantly a very dry lower troposphere, accompanied by too high temperatures, meaning that relative humidity is very low. This illustrates the important role of the convection scheme in maintaining relatively small mean state biases in the IFS.

4. Characteristics of Precipitating Convection in the Tropics

In this section we analyze a few key characteristics of tropical deep convection that have been shown to be difficult to represent realistically at km-scale resolution in previous studies, particularly when the deep convection parametrization is switched off. In these km-scale studies, convective precipitation was too intense, too localized and too short-lived and consequently not organized enough (e.g., Becker et al., 2021; Brousseau et al., 2016; Crook et al., 2019). To put our study into the context of this previous work, we first discuss precipitation intensity which provides an estimate of updraft velocities (Section 4.1), then the size of precipitation cells which informs about updraft width and spatial coherence (Section 4.2), and finally convective organization, using different organization metrics (Section 4.3).

4.1. Intensity of Precipitation

The distribution of hourly precipitation intensity in the tropics (20°S–20°N) is represented in Figures 3a and 3c on a logarithmic scale, thus allowing investigating precipitation intensities across a wide range. Because the bin sizes increase exponentially and the frequency in each bin is weighted by the corresponding precipitation intensity, the area under the curve represents the contribution of any range of precipitation intensities to the mean. For IMERG in both 2012 and 2020, most precipitation is associated with intensities between 1 and 10 mm hr^{-1}, independent of whether the precipitation is conservatively interpolated to a 0.1° (Figure 3a) or 1° grid (Figure 3c). When analyzed on a 0.1° grid (Figure 3a), ICON peaks at an intensity of about 10 mm hr^{-1} and NICAM at an even higher value.

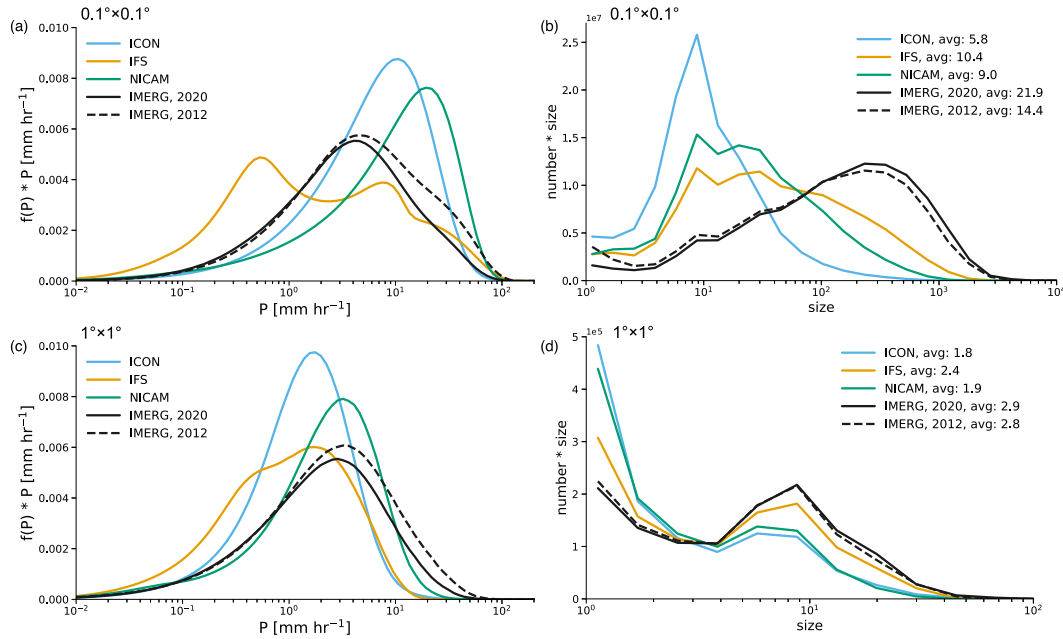

Figure 3. Characteristics of hourly precipitation in the tropics (20°S–20°N) from February to December in 2020 (ICON, IFS and IMERG) and in 2012 (NICAM and IMERG), conservatively interpolated either to a 0.1° grid (a, b) or to a 1° grid (c, d). Panels (a and c) show the frequency times precipitation intensity in each bin. The bin sizes are exponential, meaning that the area under the curve represents the contribution of that intensity range to the mean. Panel (b and d) show histograms of precipitation cell size times bin size. The precipitation cell size is defined as the number of connected grid cells (also considering diagonal neighbors) where precipitation exceeds the 99th percentile. The average number of connected grid cells forming a precipitation cell is given in the legend.

Thus, in both models with explicit convection, intense precipitation (>10 mm hr^{-1}) is too frequent compared to IMERG (in both 2012 and 2020), while intense precipitation is close to IMERG in the IFS. This is because of the reduced cloud base mass flux configuration in the IFS, which removes some vertical instability and helps triggering explicit deep convection (Rackow et al., 2025), thereby reducing updraft velocities. However, the weakly active deep convection scheme causes too much light precipitation, and is responsible for the peak of precipitation at intensities below 1 mm hr^{-1} in the IFS. The sensitivity experiments performed with the IFS at 9 km resolution show that when switching the deep convection scheme off (D_{off}), precipitation intensities look similar to ICON and NICAM (Figure S3a in Supporting Information S1). Interestingly, when also switching the shallow convection scheme off ($(D + S)_{off}$), precipitation intensities are somewhat reduced but still overestimated compared to IMERG.

When analyzed on the coarser 1° grid (Figure 3c), results look substantially different. None of the km-scale models overestimate precipitation intensity anymore, similar to the $(D + S)_{off}$ IFS sensitivity experiment, while D_{off} still overestimates intense precipitation. To understand why results on precipitation intensity and possible conclusions so strongly depend on the resolution of the analysis grid, we will analyze the size distribution of precipitation cells in the next subsection.

4.2. Size of Precipitation Cells

To calculate the size of precipitation cells, a precipitation intensity threshold needs to be defined. When comparing models and observations that differ substantially in precipitation intensity, using a fixed threshold (e.g., 3 mm hr^{-1}) can be misleading. In models with more intense precipitation, a larger fraction of grid points exceeds the threshold, making the formation of larger precipitation cells more likely artificially. Thus, we decided to use the 99th percentile of the precipitation distribution as a threshold. This however means that the absolute threshold used to determine precipitation cell sizes can be substantially different. Table 2 shows the 99th percentile of precipitation (p_{99}), on both analysis grids. On the 0.1° grid, p_{99} ranges from 2.6 mm hr^{-1} in the IFS to 5.4 mm hr^{-1} in ICON, while p_{99} is 3.5 and 3.8 mm hr^{-1} for IMERG in 2020 and 2012, respectively. On the 1°

Table 2
99th Percentile of Hourly Precipitation in the Tropics (20°S–20°N) for February to December (2020 or 2012), Conservatively Interpolated Either to a 0.1° Grid or to a 1° Grid

		IMERG				
		2012	2020	ICON	IFS	NICAM
p_{99} (mm hr^{-1})	0.1° × 0.1°	3.8	3.5	5.4	2.6	4.2
	1° × 1°	3.4	3.0	2.8	2.5	3.6

grid, p_{99} is generally smaller, with the most substantial reduction in ICON, which is an indication of small precipitation cell sizes. In the IFS sensitivity simulations, p_{99} is 4.2 mm hr^{-1} in both D_{off} and $(D + S)_{off}$ on the 0.1° grid (Table S1 in Supporting Information S1), and thus the same value as in NICAM, while in D_{on}, p_{99} has a value closer to IMERG.

Using p_{99} as a threshold, the precipitation cell size is defined as the number of connected grid cells (also considering diagonal neighbors) exceeding this threshold. Figures 3b and 3d show the precipitation cell size distribution, using a similar approach as for the precipitation intensity distribution, meaning that they show the bin frequency distribution weighted by precipitation cell size. When analyzed on a 0.1° grid (Figure 3b), ICON has by far the smallest precipitation cell sizes, followed by NICAM and the IFS. But even the IFS has substantially smaller grid sizes than IMERG. While IMERG has a substantial number of precipitation cells that exceed a size of 10^3 grid points, which for example, would correspond to a precipitation object of 5° × 2°, this size is never reached in ICON and NICAM, and almost never in the IFS. The average precipitation cell size is given in the legend of Figure 3b. Unlike the curves in the Figure, these values are not weighted by cell size and thus the small cells have a much bigger impact on the calculated average cell size. The larger number of small cells also explains why the IMERG average is much smaller in 2012 than in 2020, even though the curves look quite similar. Even for IMERG in 2012, the average precipitation cell size is substantially larger than in any of the models, though. This also includes all IFS sensitivity simulations, which show the smallest precipitation cell size distribution in $(D + S)_{off}$, similar to the one in NICAM. The largest cell sizes of all simulations occur in D_{on}, with values slightly higher than with D-RCBMF$_{on}$. Despite that, the average cell size is smallest in D_{on} because that simulation has the largest number of very small precipitation cells, which dominates the calculated average.

When analyzed on the coarser 1° grid (Figure 3d), all simulations peak at the smallest cell sizes, meaning that there is little large-scale coherence in hourly precipitation. Even though all models underestimate cell size compared to IMERG, the difference is substantial in ICON and NICAM, while the IFS is only slightly underestimating cell size compared to IMERG. Again, the IFS sensitivity simulations show that this is only the case with D_{on} and D-RCBMF$_{on}$, while D_{off} and $(D + S)_{off}$ have size distributions similar to the ones in ICON and NICAM.

In summary, in agreement with previous studies, precipitation intensities are overestimated at km-scale resolution in our three models, while precipitation cell sizes are underestimated. Km-scale studies should be careful with the choice of analysis grid, as the biases tend to compensate each other on a coarse analysis grid. Note that the IFS simulations indicate that the precipitation intensity and cell size biases can be partly mitigated by keeping the deep convection scheme still slightly active at km-scale resolutions, however at the expense of getting too much light precipitation.

4.3. Mesoscale Organization of Convection

A key feature of tropical convection is its tendency to cluster and organize, particularly at the mesoscale. On the one hand, organized convective systems often develop within larger-scale regimes such as the ITCZ and MJO, where they are influenced by large-scale dynamics while simultaneously feeding back into them. On the other hand, mesoscale convective systems alone account for more than half of tropical rainfall (Nesbitt et al., 2006; Roca et al., 2014). Therefore, accurately representing convective organization in models is essential for capturing both upscale and downscale interactions in the tropical climate system.

In this study, we apply four metrics to measure the degree of convective organization: the number (N) and size (S) of convective clusters, a spatial clustering metric (I_{org}), and Convective Organization Potential (COP). To compute these four metrics, we first need to define convective clusters. Similarly to Bao et al. (2024), we consider a grid cell as "convective" if its daily precipitation exceeds the 95th percentile of daily precipitation within the tropics (20°N–20°S). If two convective grid cells share a boundary, they belong to one convective cluster (unlike in Figures 3b and 3d, diagonal neighbors are not included). Based on this definition of convective clusters, we compute the four metrics. N is the total number of convective clusters and S is the effective radius of the convective clusters assuming that all convective clusters are spherical. I_{org} is a clustering metric which measures the spatial randomness of convective clusters (Tompkins & Semie, 2017). It first computes the nearest-neighbor distance between centroids of the convective clusters and then compares the cumulative density function of these

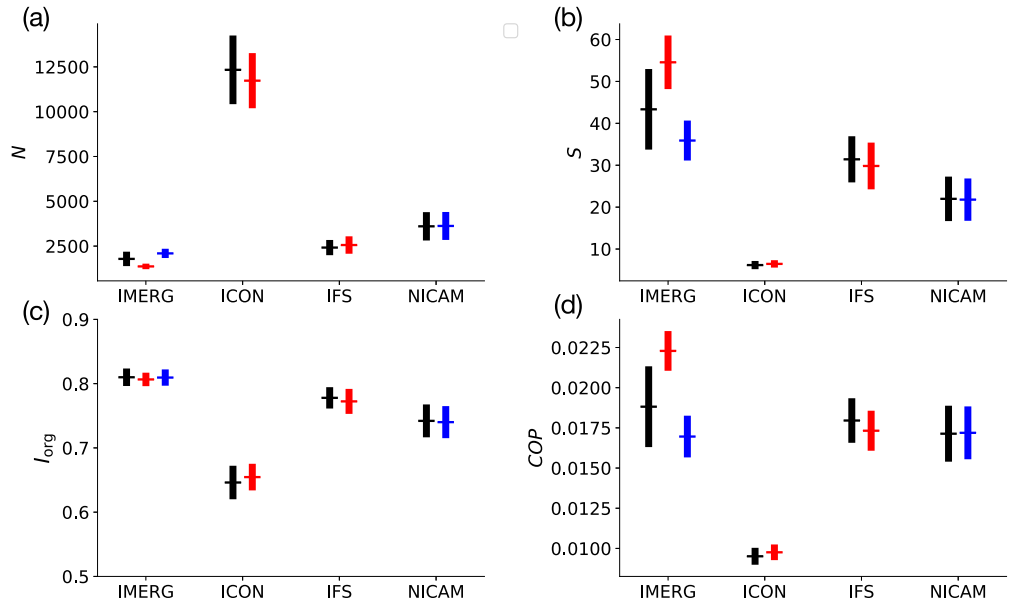

Figure 4. (a) Total number (N), (b) mean size (S, units: km) of convective clusters and (c and d) two convective organization indices (I_{org} and COP) for IMERG, ICON, IFS and NICAM measured over the tropics (20°S–20°N). Convective clusters are identified based on daily precipitation amount (exceeding 95th percentile) on a 0.1° grid. Red, blue and black colors indicate the results for the year 2020, 2012 and the whole time period (5-year simulations for the models, 2011–2020 for IMERG) respectively. Error bars show the standard deviations of daily data for each metric.

nearest-neighbor distances against the distribution assuming that the clusters (with the same number) are randomly distributed. While I_{org} focuses on the spatial distribution of convective clusters and is sensitive to the number, it does not directly depend on the size of the objects. Therefore, we further quantify convective organization with Convective Organization Potential (COP; White et al., 2018). COP measures organization following the concept of 'interaction potential', assuming that two clusters are more likely to interact with each other if they are larger and closer together. Thus, COP takes the size of the objects into account, complementing I_{org}. In a more organized state, convective systems are more spatially clustered, resulting in fewer but larger clusters (N decreases, S increases). Additionally, both I_{org} and COP are expected to increase with organization.

In general, the results indicate that all three models tend to underestimate the degree of convective organization compared to IMERG (Figure 4). The simulated convective clusters are too frequent yet too small. This underestimation is further supported by lower values of both I_{org} and COP. Among the three models, ICON shows the largest discrepancy, substantially underestimating convective organization by producing a large number of very small convective storms. While IFS and NICAM have substantially larger storm sizes than ICON, they still have smaller storms than IMERG, in agreement with Figure 3b. With regards to the other three metrics, N, I_{org} and COP, IFS and NICAM agree relatively well with IMERG, indicating that convective organization is only slightly underestimated. Overall, the characteristics of convective organization are best represented by the IFS. This is mainly due to the partial activation of the deep convective parameterization, because when both shallow and deep convection scheme are turned off in $(D + S)_{off}$, the degree of convective organization becomes substantially weaker and the values of the four metrics are close to those simulated by ICON (Figure S4 in Supporting Information S1). Nevertheless, the convective organization can be enhanced even without cumulus parameterizations, as indicated by the results in NICAM. Interestingly, when deep convection is fully active (D_{on}), I_{org} is larger than in IMERG but COP is very small. This might be related to an unrealistic feature that frequently develops in this setup, where an interaction between the convection scheme and gravity waves causes some unrealistic wave patterns in the precipitation fields. Among the four IFS setups, the D-RCBMF$_{on}$ simulation is overall closest to IMERG, confirming that this setup agrees best with observations in terms of precipitation characteristics and convective organization.

5. Evolution of Thermodynamic Quantities During Convection

To reveal possible reasons that explain the characteristics of precipitation and convective organization in Section 4, we compare how the evolution of mesoscale convection is linked to its thermodynamic environment in the three models. We examine the evolution of precipitation, humidity, and temperature during precipitation events (PEs), using the same methodology as Takasuka, Kodama, et al. (2024). Also, we analyze the simulated convective life cycle in buoyancy space, following Wolding et al. (2022).

5.1. Detection of Precipitation Events

We detect PEs in the tropics, using hourly precipitation on a 1° and 0.25° grid over ocean for ICON, IFS, and NICAM. The focus on 1°-grid scales stems from our interest in mesoscale (∼100 km) variability, which links to synoptic and larger-scale processes that shape climate, rather than individual convective cells. To detect PEs, we simply scan the tropics (20°S–20°N) for local maxima of precipitation with more than 2 mm hrr^{-1}, instead of using 700-hPa vertical velocity as in Takasuka, Kodama, et al. (2024). Note that, on a 1°-grid, the threshold of 2 mm hr^{-1} is close to the maximum of the precipitation distribution in all the models (Figure 3c). Based on this methodology, we detect 407,894, 812,684, 475,673, and 844,266 PEs on 1°-grids from 1-hourly data in IMERG, ICON, IFS (February–December in 2020), and NICAM (February–December in 2012), respectively.

To understand the basic characteristics of the detected PEs, we first compare the amplitudes and duration of PEs. Figure 5a shows the lagged-composite time series of hourly precipitation in the models and IMERG before and after the detection of 1°-grid scale PEs ($t = 0$ hr). In all the models, the amplitudes of precipitation are underestimated compared to IMERG. For ICON and NICAM, this is partially due to less convective organization, as shown in Figure 4; the simulated precipitation is too intense at smaller spatial scales (Figure 3a), but this overestimation is moderated by including non-precipitating areas when analyzing precipitation on a 1°-grid. On the other hand, the underestimated precipitation in IFS can be attributed to the overestimation of weak precipitation cells at 1° resolution (cf., Figure 3c), which includes both intense but small-sized precipitation cells and wide-spread weak precipitation.

The *e*-folding timescale of precipitation from $t = 0$ hr, as indicated by the dots in Figure 5a, provides an estimate of precipitation duration. It is excessively long in IFS, even compared to IMERG. This bias probably comes from effects of the cumulus parameterization, which may correspond to the overestimation of weak precipitation on the 1° grid. In fact, the duration of 0.25°-grid scale PEs, which are expected to have less contributions from wide-spread weak precipitation than on the 1° grid, becomes much shorter and thus close to IMERG (Figure 5c). Meanwhile, even though the duration of 1°-grid scale precipitation appears realistic in ICON and NICAM compared to IMERG, this can result from the error compensation between the larger number of finer-scale precipitation systems (see Figure 4b) and their shorter duration. Consistent with this expectation, the *e*-folding time scale of 0.25°-grid scale precipitation in ICON becomes shorter than IMERG (Figure 5c).

Among the processes associated with deep convection, cold pools play a crucial role by redistributing momentum and moisture, thereby shaping the spatial organization of convective systems. We thus examine the time evolution of lagged-composite 2m temperature during 1°- and 0.25°-grid scale PEs (Figures 5b and 5d). As observations we use the moored buoy data from three TAO stations over the western Pacific (see Section 2.2). Note that the buoy observations are local, meaning that the time evolution is expected to be even sharper than on the 0.25°-grid. To have a comparison that is as meaningful as possible, PEs are defined as local maxima of the precipitation time series that exceed 2 mm hr^{-1} but stay below 20 mm hr^{-1} so that the amplitude of the precipitation composite is consistent with IMERG in Figures 5a and 5c. Although the temperature drop during the PE is simulated in all three models, their amplitude is underestimated compared to the buoy observations even at the finer spatial scales, and differs between the models. This suggests that km-scale models still have uncertainties in representing the cold pool dynamics, such as cloud microphysics, rain evaporation and downdrafts, biasing the physics-dynamics coupling. For example, the excessively weak cold pools in IFS can be related to the single-moment microphysics and underestimated rain evaporation, which causes too weak downdrafts. This bias is much less evident when the convection scheme is fully active (Figures S5b and S5d in Supporting Information S1) because of the downdraft parameterization embedded into the convection scheme, which prescribes downdraft mass flux as 30% of updraft mass flux. In NICAM, relatively strong cold pools probably result from enhanced evaporation introduced by the

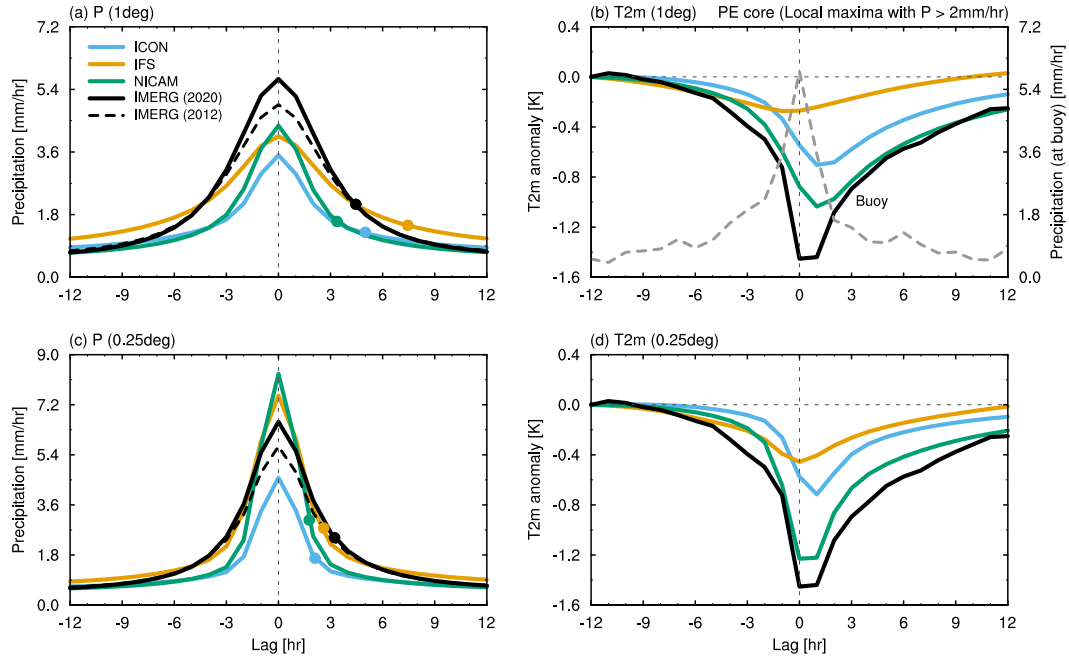

Figure 5. (a) Time series of lagged-composite hourly precipitation during the 24-hr evolution associated with the development of PEs on a 1° grid over tropical ocean (20°S–20°N) for ICON, IFS and NICAM, as well as IMERG in 2020 and 2012. The reference time for composite ($t = 0$ hr) is when deep convective cores are detected. Filled markers denote when precipitation decays at the *e*-folding time scale. (b) Same as (a), but for 2 m-temperature anomalies (solid). Solid black and dashed gray lines denote 2 m-temperature anomalies and precipitation at the three buoy stations over the western Pacific during February–December in 2010, respectively. The anomalies are defined as deviations from values at $t = -12$ hr. (c, d) Same as (a, b), but on a 0.25° grid, again over tropical ocean.

tuning of terminal velocities of raindrops for a realistic seamless representation of weather and climate (Takasuka, Kodama, et al., 2024).

5.2. Tropospheric Moisture and Temperature Tendencies During Precipitation Events

The thermodynamic tendencies associated with deep convection determine how convection evolves and how its thermodynamic environment is modulated by convection (e.g., Masunaga, 2013; Raymond, 2000; Takemi et al., 2004). In the context of global convection-permitting models, Takasuka, Kodama, et al. (2024) demonstrated that moisture tendencies before and after the onset of deep convection are strongly influenced by the cloud microphysics and turbulence scheme. Different assumptions in these parameterizations lead to differences in the preferred spatio-temporal scales of moist convective systems. This motivates us to investigate tropospheric moisture and temperature tendencies during PEs in the three models relative to ERA5 reanalysis data, and how this relates to convective organization.

Figure 6a shows the lagged-composite time-height evolution of specific humidity and temperature anomalies for ERA5, based on 1°-grid scale PEs detected from IMERG precipitation in 2020. The anomalies are relative to the time-average from $t = -48$ to $t = 48$ hr, and thus describe the variability associated with the development and decay of deep convection. The gradual shallow-to-deep moistening is observed after $t = -36$ hr, and upward transported moisture remains in the mid-troposphere until $t = 30$ hr. In addition, tropospheric temperature changes especially during the mature stage of deep convection; strong cooling occurs in the planetary boundary layer (PBL) and at the freezing level, while a warm anomaly emerges in the upper troposphere with a cold anomaly directly above, likely in association with the development of anvil clouds. Despite the caveat that ERA5 reanalysis data is affected by the IFS physics parameterizations and thus not exactly consistent with observations, the thermodynamic tendencies in ERA5 are in qualitative agreement with those derived from satellite measurements (Masunaga, 2013).

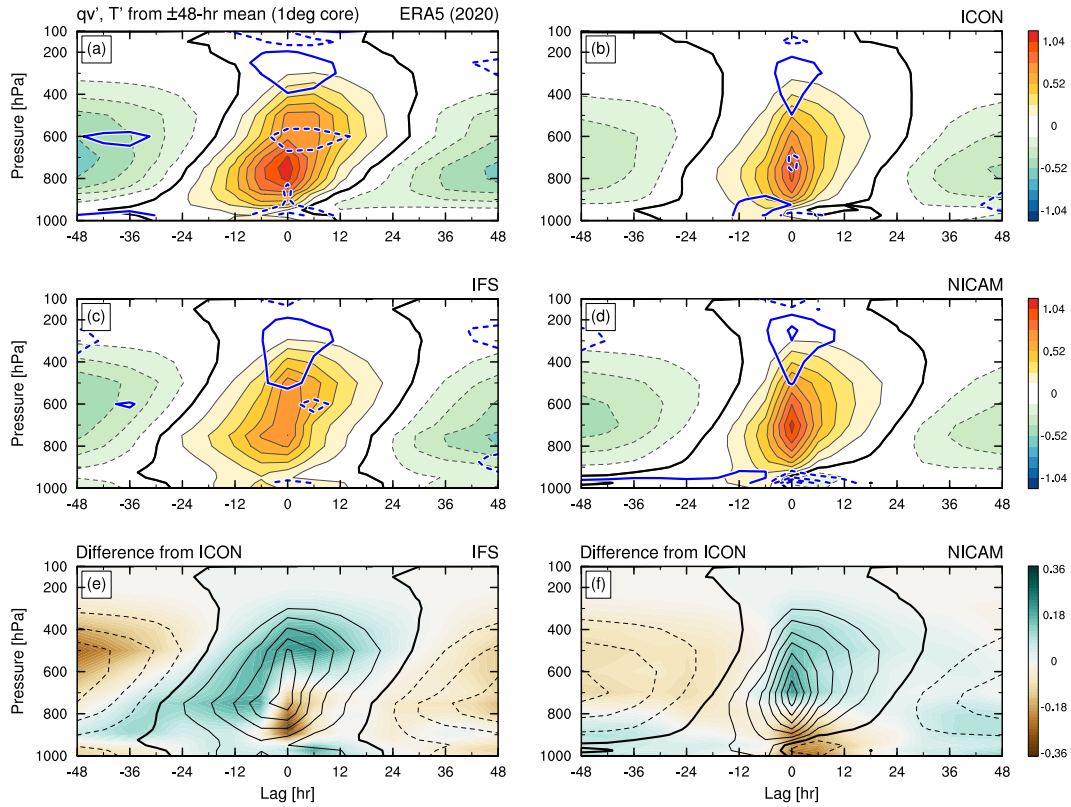

Figure 6. (a–d) Time-height sections of lagged-composite specific humidity (shading; g kg^{-1}) and virtual temperature anomalies (blue contours; K) during the 96-hr evolution associated with the development of 1°-grid scale PEs over tropical ocean (20°S–20°N) for (a) ERA5 (using IMERG to detect the PEs), (b) ICON, (c) IFS, and (d) NICAM. The anomalies are defined as the deviations from the ±48-hr mean. Thick black lines are zero contour lines of specific humidity anomalies, and blue contour lines denote ±0.12 K, 0.36 K, 0.60 K, …, with dashed lines indicating negative values. (e, f) Same as (c, d), but for the difference of specific humidity anomalies from those for ICON (shading).

Although the shallow-to-deep moistening and upper-tropospheric warming are qualitatively captured by all three models (Figures 6b–6d), their timescales and amplitudes do not agree quantitatively with ERA5, and also differ strongly between the models. In ICON (Figure 6b), the shallow-to-deep transition is more rapid than in ERA5, as inferred from the steeper zero-contour line before $t = 0$ hr. In addition, after the onset of deep convection, vertical moisture transport and mid-tropospheric moistening are weaker. This can also impact the large-scale tropical mean state. For example, this could be one of the reasons why the lower free troposphere is too dry in Figure 2, at least to the extend that this cannot be attributed to the too cold mean state. The IFS $(D + S)_{off}$ simulation has both a significant dry bias and a warm bias at 800 hPa, meaning that relative humidity is very low (Figure S2 in Supporting Information S1). A dry bias in the large-scale environment is consistent with the excessive number of small convective clusters that these models simulate (Section 4.3), as the moisture adjustment timescale is too short, which leads to weak mid-tropospheric moistening, and further inhibits large-scale convective organization (e.g., Ahmed & Neelin, 2019; Derbyshire et al., 2004; Takemi et al., 2004). Another potential reason for less organized convection in ICON may be a weaker cloud-radiation interaction due to a shorter lifetime of upper-tropospheric clouds, indicated by the shorter duration of upper-tropospheric warming around $t = 0$ hr (Figure 6b), although this radiative impact on organized convection can be dominated by the other processes described above (e.g., Ahmed & Neelin, 2019).

In IFS and NICAM, the moisture-convection relationship aligns more closely with IMERG/ERA5 than ICON does. For clarity, Figures 6e and 6f illustrate the moisture anomaly differences relative to ICON for IFS and NICAM, respectively. It is notable that both IFS and NICAM simulate more mid-tropospheric moistening after $t = 0$ hr than ICON, although processes that cause this moistening seem to differ between IFS and NICAM. In the IFS, which exhibits the strongest mid-tropospheric convective moistening overall, this moistening is linked to

abundant moisture transport during the shallow-to-deep transition, facilitated by the activation of parameterized shallow convection. This is demonstrated by the moisture-convection relationship in the IFS sensitivity simulations, where convective moistening still peaks in the mid-troposphere with D_{off} but shifts to the lower troposphere with $(D + S)_{off}$ (Figure S6 in Supporting Information S1). In NICAM, the mid-tropospheric moistening results from turbulent mixing supplemented by the Leonard terms and moisture accumulation in the PBL before $t = 0$ hr, as well as more stable stratification around the melting level (cf., Takasuka, Kodama, et al., 2024). The improved representation of moisture detrainment in the mid-troposphere is certainly favorable for organized convection at a larger scale in these models (Figures 3b and 4). Also, deep convection in NICAM is less prone to be initiated, as indicated by moister anomalies in the PBL and drier anomalies in the free troposphere relative to ICON during the shallow-to-deep transition before $t = 0$ hr (Figure 6f), which further helps expand the spatial scale of organized convection.

Nevertheless, IFS and NICAM still struggle to represent the convective development and dissipation realistically, as evidenced by smaller convective cluster sizes than in reality (Figure 3b). In the IFS, the shallow-to-deep transition is faster than in NICAM and IMERG/ERA5 (at $t = -18$ hr, the zero contour line is at 500 hPa in the IFS, but at 700 hPa in NICAM and IMERG/ERA5, Figure 6), which can be linked to the partially active cumulus parameterization (Figure S6 in Supporting Information S1). In addition, the PBL remains moister and warmer than the environment even after the mature stage of deep convection (Figure 6c), indicating the severe underestimation of cold pools, in line with Figure 5b. Meanwhile, compared to IFS and ERA5, NICAM has less vertical moisture transport when convection develops into the deep mode (Figure 6d), which implies the poor representation of cloud eddies in km-scale explicit convection. As for upper-tropospheric temperature tendencies, both IFS and NICAM underestimate the duration of warm anomalies. These results suggest that at km-scale resolution, fundamental difficulties persist in realistically simulating the initiation of convection, mesoscale moistening rate, and longevity of clouds, but the degree of these biases depends on model configurations.

5.3. Life Cycle of Convection in Buoyancy Space

In addition to the direct evaluation of moisture and temperature variations associated with PEs, Wolding et al. (2022) have proposed a more process-oriented diagnosis that focuses on buoyancy variations to describe the life cycle of convection. This diagnosis follows the idea that when tropical convective plumes develop, they consume the instability accumulated in the PBL and experience dilution through the entrainment of environmental air in the lower free troposphere (LFT). By comparing how these processes develop during the phases of convective growth and decay, we aim to reveal differences in thermodynamic-convection coupling across the three models and to deepen our understanding of the moisture-temperature-convection relationships discussed in Section 5.2.

Following Wolding et al. (2022), we calculate the lower-tropospheric buoyancy (B_L) as:

$$B_L = g\left(w_B \frac{\tilde{\pi} e_B - e_L^*}{e_L^*} - w_L \frac{L_v\left(q_L^* - q_L\right)}{e_L^*}\right) \tag{1}$$

where g is the gravitational acceleration, L_v is the latent heat of vapourization, q is the specific humidity, $e = C_p T + L_v q$ is the moist enthalpy (T and C_p are the temperature and specific heat capacity at constant pressure), q^* and e^* represent the saturation values of q and e, respectively, and $\tilde{\pi} = \left(p_L / p_B\right)^{R_d/C_p}$. Subscripts B and L denote mass-weighted average over the deep planetary boundary layer (DBL; surface–850 hPa) and LFT (850–600 hPa), respectively, with $p_B = 925$ hPa and $p_L = 725$ hPa. w_B and w_L measure how DBL and LFT influence plume buoyancy, summing up to 1. We here use $w_B = 0.59$ and $w_L = 0.41$ as in Wolding et al. (2022). B_L can be further decomposed into undiluted and diluted buoyancy as follows:

$$B_L = g\left(\underbrace{\frac{\tilde{\pi} e_B - e_L^*}{e_L^*}}_{\text{Undiluted } B_L} \underbrace{- w_L \frac{\tilde{\pi} e_B - e_L^*}{e_L^*} - w_L \frac{e_L^* - e_L}{e_L^*}}_{\text{Dilution of } B_L}\right) \tag{2}$$

The undiluted B_L represents "pure" buoyancy obtained in DBL under the assumption that a rising air parcel does not mix with its LFT environment, and dilution of B_L measures the loss of plume buoyancy due to LFT entrainment. Less negative dilution of B_L means a smaller impact of LFT entrainment, which corresponds to a more humid environment.

Figure 7 shows the variability of 6-hourly undiluted B_L, dilution of B_L, and precipitation in the phase space constructed by the decomposed buoyancy for IMERG/ERA5 and the three models. In this phase space, all data is averaged in each bin with a width of 0.01 m s^{-2}, and vectors indicate the time tendency of 6-hourly undiluted B_L and dilution of B_L, calculated from their temporal centered differences. For IMERG/ERA5 (Figure 7a), the cyclical coevolution of undiluted B_L and dilution of B_L is clearly captured around the "quasi-equilibrium (QE) point" (black circle), defined as the bin with most samples, in accordance with the development and decay of precipitation. First, undiluted B_L increases during the suppressed phase of precipitation with more negative dilution of B_L (e.g., dry LFT, see lower left corner of Figure 7a). Once convective instability is sufficient to initiate convection, precipitation intensifies (upper left corner), along with an increase in the dilution of B_L, marking the onset of environmental moistening and gradual convective development. Following this, precipitation reaches its peak (upper right corner), leading to a rapid decrease in undiluted B_L as convective instability is released (lower right corner). Consequently, the dilution of B_L becomes more negative due to LFT drying, completing the convective life cycle and returning to the initial state.

Life cycles of convection in buoyancy phase space differ for the three models. In ICON (Figure 7b), the cyclical coevolution of undiluted B_L and dilution of B_L is qualitatively captured, but the variability of dilution of B_L is weaker than in IMERG/ERA5. Specifically, the peak of precipitation and release of undiluted B_L are realized for more negative dilution of B_L. Consistent with this result, the probability distribution of dilution of B_L shifts toward more negative values relative to IMERG/ERA5, which also applies to the change in the quasi-equilibrium (QE) point. These results suggest that the convective development in ICON is less sensitive to moisture variations in the LFT, which agrees with the faster timescale of shallow-to-deep transition found in Figure 6b. Meanwhile, IFS has a significantly different evolution compared to IMERG/ERA5 and the other two models (Figure 7c); the cyclical coevolution of the two-part buoyancy is less clear. The buoyancy phase evolves clockwise in the weak precipitation regime (less than 10 mm day^{-1}), whereas its evolution is reversed and/or very weak in the strong precipitation regime. This unrealistic feature is related to the weakly active deep convective parameterization, as it is not present in the IFS sensitivity experiments with D_{off} and $(D + S)_{off}$ in Figure S7 in Supporting Information S1. Compared to IMERG/ERA5, the probability distribution of dilution of B_L is too narrow at 4 km and in the D-RCBMF$_{on}$ sensitivity experiment, indicating that moisture variations in the LFT are too weak. The distribution of dilution of B_L is more realistic with deep convection scheme fully active (D_{on}). However, D_{on} lacks positive values of undiluted B_L almost completely, indicating that the convection scheme with its CAPE closure releases instability too quickly. The distribution of undiluted B_L looks more similar to ERA5 in the 4 km and D-RCBMF$_{on}$ simulation, in line with the more realistic characteristics of intense precipitation in this configuration (Figure S3 in Supporting Information S1). Meanwhile, the maximum of the probability distribution for undiluted B_L (at the QE point) is slightly more pronounced in IFS than in ICON and NICAM, mostly because the frequency of occurrence of stable conditions (negative undiluted B_L) is slightly underestimated.

NICAM simulates the most realistic buoyancy variations associated with the life cycle of convection (Figure 7d). The distinct cyclic amplification and decay of precipitation are linked to a realistic variability in the dilution of B_L, even though precipitation peaks in an LFT environment with less dilution effects (i.e., less negative dilution of B_L). Compared to ICON and IFS, the convective evolution in NICAM is more sensitive to saturation deficit in the LFT, consistent with the slower shallow-to-deep convective transition (Figure 6d). Note that the rapid decrease of undiluted B_L during the most active precipitation phase can be related to too intense and too short-lived mesoscale precipitation systems (see Figures 3a and 5a).

To sum up, storm-resolving modeling still exhibits diversity in representing the variability of LFT plume buoyancy, depending on model physics that govern temperature and moisture stratification. While simulations with explicit convection (e.g., ICON and NICAM) can capture the connection between cyclical variations of B_L and amplification/decay of precipitation, this coevolution is distorted by the coexistence of parameterized and explicit convection (e.g., IFS), possibly because these two forms of convection are artificially separated and respond to different environmental conditions. There is also significant uncertainty regarding how sensitive convection is to the saturation deficit in the LFT, which affects the timescale at which deep convection is initiated.

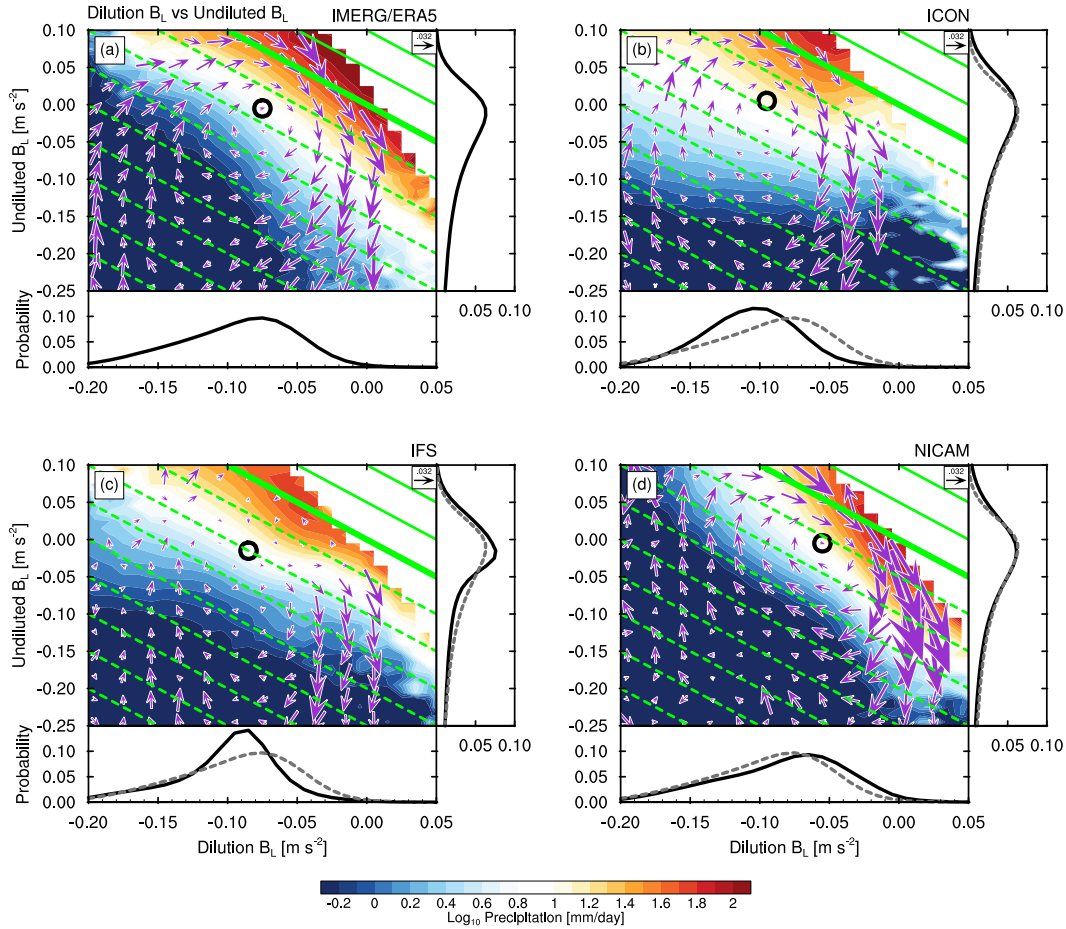

Figure 7. Two-dimensional diagrams constructed with undiluted B_L (y-axis) and dilution of B_L (x-axis) over the warm pool (60°E−180°, 15°S–15°N) for (a) IMERG/ERA5, (b) ICON, (c) IFS, and (d) NICAM, following Wolding et al. (2022). Data is plotted for each bin of width 0.01 m s^{-2} along the x and y axes. Shading represents bin-mean 6-hourly precipitation rate (log scale), and vectors indicate the time evolution of undiluted B_L and dilution of B_L. The time evolution is calculated as the bin-mean temporal centered differences of those two variables using 6-hourly data, and for visibility, arrow length is multiplied by 2. Circles denote the bin with the most samples. Green lines indicate total B_L, and their interval is 0.05 m s^{-2}, with negative contours dashed and zero contour thickened. Probability distributions of undiluted B_L and dilution of B_L are also plotted along each of the axes, and gray dashed lines in these subplots in (b–d) indicate the results from ERA5.

6. Atmospheric Phenomena Connected to Thermodynamic-Convection Coupling

In this section, we evaluate how the three models reproduce key weather and climate phenomena that are closely tied to thermodynamic–convection coupling. First, we consider the precipitation diurnal cycle (PDC)—a fundamental aspect of tropical convection—and its relationship to mean precipitation and the propagation of convection at coastlines. Next, we examine the MJO, the leading mode of intraseasonal variability in the tropics, marked by the slow eastward propagation of convective envelopes on the order of $O(10^3)$-km across the Indo-Pacific warm pool. Finally, we explore how convective organization influences the characteristics of precipitation extremes across the whole tropics.

6.1. Precipitation Diurnal Cycle

Figures 8a and 8b show the composite hourly time series of PDC on a 0.25° grid over ocean and land, averaged over 20°S–20°N. Here, the models are compared to the two satellite-based products: TRMM-3G68 climatology (black solid) and IMERG in 2020 (black dashed). However, the observed phase of the PDC should be assessed using TRMM-3G68, which relies solely on active radar sensors. This is important because IMERG precipitation estimates, which incorporate passive microwave and infrared sensor data, are known to exhibit a systematic phase

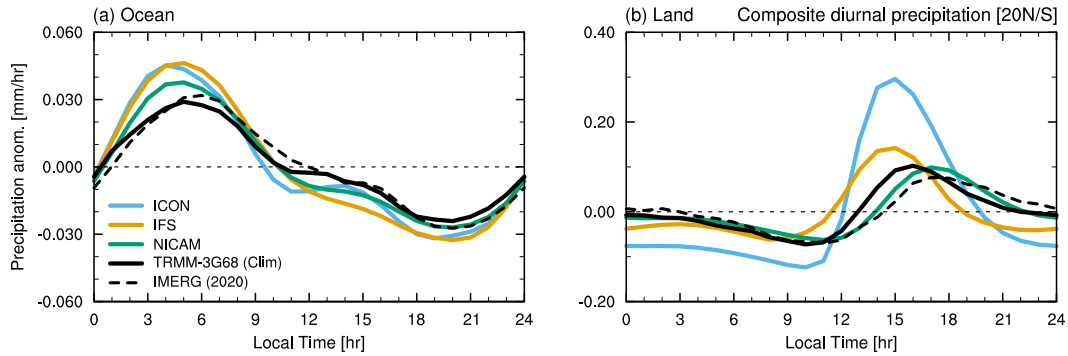

Figure 8. Composite diurnal cycle of hourly mean precipitation anomaly averaged over 20°S–20°N over (a) ocean and (b) land for ICON, IFS, NICAM, the TRMM-3G68 climatology, and IMERG in 2020.

delay in the PDC (e.g., Kikuchi & Wang, 2008; Sato et al., 2009). In fact, the peak time of PDC is later by 1–2 hr for IMERG than TRMM-3G68 (Figure 8).

Over ocean, the PDC simulated by the three models agrees well with each other, and its phase is consistent with the observation (TRMM-3G68), although PDC in ICON peaks 1 hr earlier than in the IFS and NICAM, and all the models overestimate the PDC amplitude. The main driver of the PDC over ocean is nocturnal radiative cooling at cloud tops, which enhances convective instability (e.g., Randall et al., 1991). This process is generally well captured by models, which explains why inter-model PDC differences are small over ocean.

The representation of the PDC over land exhibits significant uncertainties, both in terms of amplitude and timing (Figure 8b). The amplitude is strongest and most overestimated in ICON. A potential explanation for the differences in amplitude is the discrepancy in how mixing effects are represented. ICON uses the Smagorinsky scheme for turbulent diffusion, which, at the km-scale, tends to underestimate mixing. This leads to exaggerated convection due to a reduced likelihood of buoyancy loss. In case of NICAM, the lateral mixing associated with deep convection is partly supplemented by the Leonard terms (see Section 2.1), which contribute to moderating excessively intense convection (Takasuka, Kodama, et al., 2024). As for the timing of PDC, the TRMM-3G68 climatology has an afternoon peak at 16:00 LT, and the peak is simulated 1 hr earlier in ICON/IFS and 1 hr later in NICAM. Also, the timing at which precipitation begins to increase in the morning is about 2 hr earlier in the IFS (08:00 LT) and 1 hr later in NICAM (11:00 LT) than in TRMM-3G68 (10:00 LT). These large differences in the simulated PDC phase over land are consistent with those in the timescale of the shallow-to-deep convective transition in Figure 6; convection is more prone to be initiated quickly in ICON and IFS than in NICAM. Hence, it remains a challenge at km-scale resolution to represent convective initiation in response to destabilization, which critically controls precipitation over land.

The connection between mean precipitation and the precipitation diurnal cycle is particularly strong in the Maritime Continent, where rainfall processes are driven more by the complex interplay of terrain and land-sea contrasts than by large-scale forcing (Yoneyama & Zhang, 2020). In fact, in coastal regions with climatologically high precipitation (IMERG in 2020; Figure 9a), the first harmonic of the PDC normalized by mean precipitation is larger than over ocean (Figure 9b).This suggests an upscale effect of the PDC on the mean precipitation field. This relationship holds true for other years (e.g., IMERG in 2012) and the TRMM-3G68 climatology (not shown). Figures 9c–9h compare the differences in mean precipitation (left column) and PDC amplitudes (right column) from IMERG for ICON, IFS, and NICAM. The most notable feature is that, with the exception of the Java Sea in ICON where the precipitation band in the Southern Hemisphere is exaggerated by the double ITCZ bias, all models underestimate the mean precipitation in coastal regions, for example, near Sumatra, Borneo, and New Guinea. In agreement with this, PDC amplitudes are weaker than in IMERG in these regions. The extent of these biases depends on the model. Biases are most significant in the IFS (see the affected regions marked by pink and green boxes in Figures 9e and 9f, respectively), followed by ICON and NICAM. Interestingly, this model ranking based on bias magnitude aligns with that observed for cold-pool intensity (see Section 5.1; Figures 5b and 5d). This relationship reinforces the idea of Yokoi and Kajikawa (2024) that the poor representation of cold pools is a source of biases in PDCs in coastal regions. Nevertheless, this is just speculation; other factors such as orography and low-level vertical wind shear associated with land-sea breezes also truly regulate PDCs in coastal regions, as

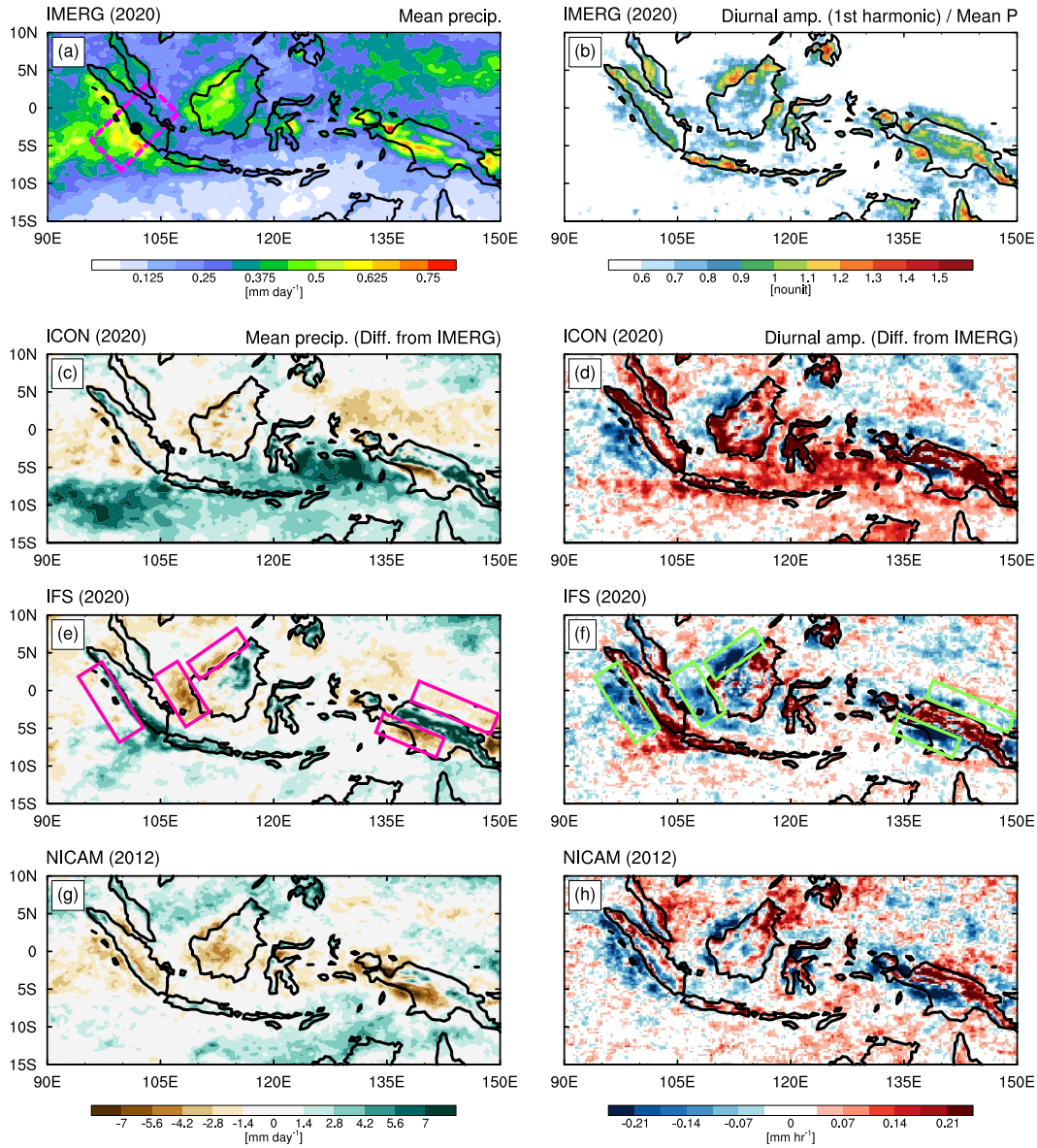

Figure 9. (a, b) Horizontal maps over the Maritime Continent of (a) annual-mean precipitation and (b) the ratio of precipitation diurnal cycle (PDC) amplitude calculated as the first harmonic component of composite PDCs to the mean precipitation for IMERG in 2020. The dashed magenta box indicates the region plotted in Figure 10. (c, e, g) Horizontal maps over the Maritime Continent of the difference in annual-mean precipitation for ICON/IFS from IMERG in 2020 and for NICAM from IMERG in 2012. (d, f, h) Same as in (c, e, g), but for the difference in the PDC amplitude. The magenta/green boxes in (e, f) highlight regions where dry biases of mean precipitation and too low PDC are prominent.

exemplified by the dependence of bias magnitude on local areas (e.g., large bias in NICAM especially in coastal regions of Sumatra). These results support the notion that mesoscale convective processes play an important role in rectifying large-scale precipitation patterns over the Maritime Continent.

PDCs in coastal regions are characterized by the offshore migration of organized convective systems. To further examine how this evolution is simulated by the three models, we look at the details of PDCs that are prominent around Sumatra (e.g., Mori et al., 2004). Figure 10a shows the composite time series of PDCs captured by TRMM-3G68 climatology, along the orthogonal direction of the western coastlines of Sumatra. For orientation, see the dotted pink box in Figure 9a which depicts the analyzed area. Convection develops over land close to the coast in the afternoon (around 16 LT), and then organized precipitation systems propagate away from the coastline during nighttime, both inland and offshore. The offshore-propagating systems are more significant and

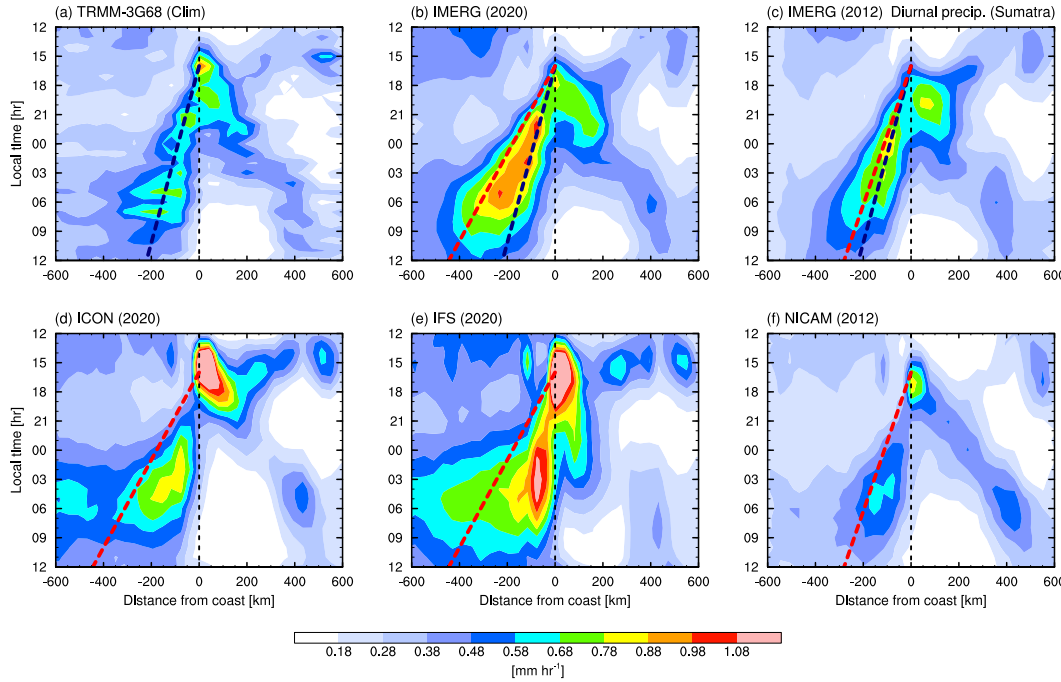

Figure 10. Time series of composite hourly precipitation along the direction perpendicular to the western coastline of Sumatra Island for the (a) TRMM-3G68 climatology, (b) IMERG in 2020, (c) IMERG in 2012, (d) ICON, (e) IFS, and (f) NICAM. Negative values of distance mean the offshore direction. Black and red dashed lines indicate the direction of the offshore propagation of precipitation systems in TRMM-3G68 climatology and IMERG 2020/2012, respectively.

survive until early morning, as in several in situ observations (Yokoi et al., 2017, 2019). This feature is also captured by the IMERG products in 2020 and 2012 (Figures 10b and 10c). In IMERG 2020, the speed and amplitude of offshore propagation of precipitation systems are faster and stronger than those in the TRMM-3G68 climatology, while in IMERG 2012, they closely match the climatological values.

Figures 10d–10f compare the aforementioned aspects between ICON, IFS, and NICAM. Note that thick dashed red lines indicate the propagation on the offshore side based on IMERG, and that the phase delay is modified by assuming that precipitation begins to develop at 16 LT as in the TRMM-3G68 climatology. In ICON (Figure 10d), organized precipitation systems do not propagate as far offshore as in the observations during nighttime, and their amplitude is underestimated. In the IFS, most of the organized convection stays too close to the coastline, leading to too strong PDC amplitudes and too much mean precipitation at the coastline, but too weak precipitation 200 km off the coast and beyond (Figure 10e). In NICAM (Figure 10f), the phase of the offshore propagation of convection during nighttime is relatively well-captured, but precipitation intensities are underestimated in comparison with IMERG in 2012.

In summary, all models cannot fully represent the continuous regeneration of mesoscale convective systems migrating from land to open ocean, with consequences both locally and upscale. Notably, the magnitudes of the biases associated with offshore propagation that are evident in PDC amplitudes in coastal regions (right in Figure 9) are consistent with those for simulated cold pool intensity (Figures 5b and 5d), which is one of possible factors affecting mesoscale convective organization. This aligns with the idea that the interaction between cold pools and low-level vertical wind shear plays a crucial role in triggering new convective cells within mesoscale convective systems (e.g., Houze, 1981; Rotunno et al., 1988), which should be represented at km-scale resolutions to some extent depending on model physics (Okugawa et al., 2024).

6.2. MJO

We here compare the observed and simulated eastward propagation of large-scale convective envelopes associated with the MJO. We detect MJO events initiated in the Indian Ocean based on the evolution of the Real-time Multivariate MJO (RMM) index (Wheeler & Hendon, 2004), which is constructed from two principal

components (PC1 and PC2) of equatorial intraseasonal OLR and zonal wind anomalies at 850 and 200 hPa. The MJO detection methodology follows Takasuka, Kohyama, et al. (2025), a modified version of Suematsu and Miura (2018). It takes the following two steps: (S_1) detecting basin-scale convective events (BCEs) over the Indian Ocean; and (S_2) classifying the detected BCEs into MJO and non-MJO events by tracking RMM amplitudes ($A = \sqrt{PC1^2 + PC2^2}$) and phase angles ($\alpha = \tan^{-1}(PC2/PC1)$). Step S_1 identifies periods of BCEs over the Indian Ocean (60°–90°E) based on areal-mean OLR anomalies relative to their 61-day running mean. Step S_2 then assesses whether the RMM index exhibits sufficiently large amplitudes during these BCE periods and whether it shows a coherent phase progression. The criteria for detecting BCEs and evaluating the RMM index follow those outlined in Takasuka, Kohyama, et al. (2025), where the detection procedure is described in detail. Note that this detection has been performed to evaluate the 9-km D_{on} and 4.4 km IFS simulations (Rackow et al., 2025), although the present study adopts larger criteria imposed on A when tracking RMM amplitudes to detect more robust MJO events.

Figure 11 shows the lagged-composite time evolution of equatorial (15°S–10°N) OLR and lower-tropospheric westerly anomalies associated with the detected MJO events. Note that Day 0 corresponds to the date when MJO convection is initiated over the Indian Ocean, and that the anomalies are defined as the deviations from a 61-day running mean. In the observational reference (Figure 11a), large-scale convective envelopes are co-located with westerly wind anomalies. They move across the Maritime Continent (MC) with slightly reduced amplitude, and subsequently propagate into the western Pacific (WP). Even in global km-scale models, these MJO features are not necessarily captured; in fact, there is a large discrepancy in MJO reproducibility between ICON and the other two models. In ICON (Figure 11b), the OLR anomaly associated with MJO initiation is not as strong as in the observations, and it dissipates over the MC. This behavior reflects the overestimation of the barrier effects of the MC on MJO propagation (C. Zhang & Ling, 2017) or "standing" MJOs (Wang et al., 2019), which has also been a long-standing issue in conventional GCMs such as CMIP6 models (e.g., Back et al., 2024; G. Chen et al., 2022). In contrast, this issue is largely resolved in IFS and NICAM irrespective of the presence or absence of ocean coupling; MJO convection migrates to the WP successfully (Figures 11c and 11d), although several biases remain (e.g., weaker convection in the WP and less prominent westerly wind anomalies in IFS; and overestimated weakening of OLR and wind anomalies over the MC in NICAM).

Large-scale convective disturbances, such as the MJO, are marked by pronounced contrasts between moist and dry regions in the tropics. These disturbances are partially sustained by moisture–convection feedback mechanisms, in which radiative cooling within the dry regions plays a key role (e.g., Grabowski & Moncrieff, 2004). Therefore, the way convection manifests may depend on how large-scale subsidence associated with radiative cooling is balanced by ascending motions. In this context, we can apply the insights gained from the mesoscale convection characteristics presented in Section 5. In ICON, deep convection is developed more rapidly and abruptly, and moisture overturns relatively quickly in response to deep convection, as indicated by the underestimated moisture detrainment in the mid-troposphere (Figure 6b). With these features, intense but short-lived updrafts can easily be balanced through compensating subsidence and downdrafts locally, implying no need for well-organized and slowly varying convective systems. Meanwhile, mesoscale convection in NICAM is less readily initiated and moistens the mid-troposphere (Figure 6f), as indicated by its higher sensitivity to lower free-tropospheric (LFT) moisture variations (Figure 7d) and a more stable mid-tropospheric layer (Figure 2; Takasuka, Kodama, et al., 2024). Hence, it is difficult to remove the convective instability locally and thus the moisture-convection feedback becomes more efficient at larger spatial scales, which is expected to be favorable for the MJO realization. Similarly, the good representation of MJO in IFS is related to more significant moistening by vertical moisture transport (assisted by the shallow convection scheme) and by more mid-tropospheric moisture detrainment than in ICON (Figure 6e). These insights are consistent with previous results that a strong coupling between precipitation and moisture enhances the MJO variability (e.g., Hannah & Maloney, 2011; Holloway et al., 2013; Kim et al., 2014).

While the characteristics of convective organization and associated moisture transport explain the model diversity of MJO reproducibility to some extent, we should note a caveat that background SSTs against the MJO, which also modulate the MJO activity (e.g., Suematsu & Miura, 2018; Wang et al., 2019), are different between the models. This is an issue when evaluating coupled simulations in ICON and IFS. Specifically, ICON has a strong bias in the La Niña-like SST pattern (Segura et al., 2025), exaggerating the Walker circulation and thus suppressing

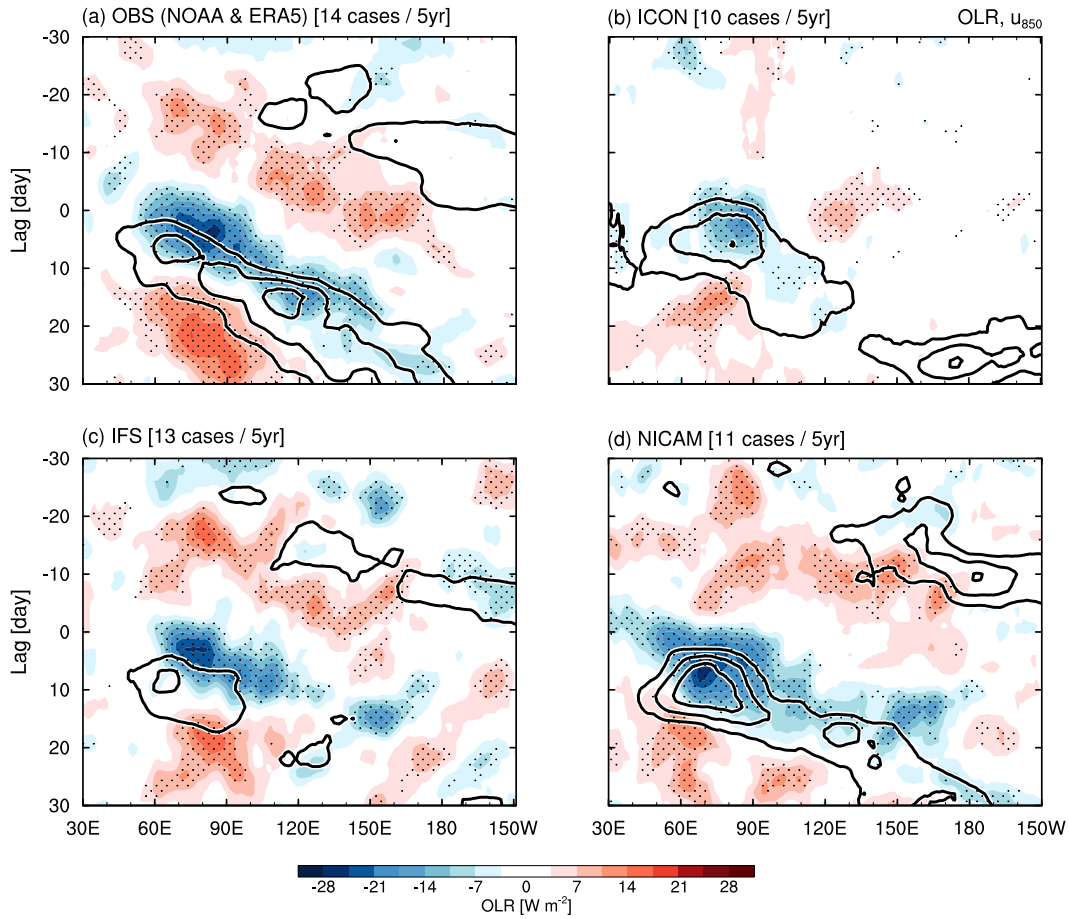

Figure 11. Time-longitude diagrams of lagged composite OLR (shading) and westerly wind anomalies (contours) in the lower-troposphere (850 hPa for IFS and NICAM and about $z = 1500$ m for ICON) associated with the detected Madden-Julian Oscillation (MJO) events for (a) NOAA-OLR and ERA5 (2011–2015), (b) ICON, (c) IFS, and (d) NICAM. All fields are averaged over 15°S–10°N. The number of the detected MJO events is denoted at the top-left corner. Contour interval is 0.8 m s^{-1}, with zero contours omitted. Stippling denotes regions where OLR anomalies are statistically significant at the 90% confidence level.

background convective activities over the WP. This is unfavorable for the realization and propagation of MJO convection (Takasuka, Kohyama, et al., 2025), which contributes additionally to the lack of MJOs in ICON.

6.3. Extreme Precipitation

The strength of moisture–convection coupling potentially also affects the relationship between convective organization and extreme precipitation. Enhanced coupling fosters stronger convective organization, which subsequently amplifies extreme precipitation. To investigate this relationship, we measure the degree of organization over the whole tropics and compare its relationship with extreme precipitation. We focus on events with a relatively high degree of convective organization (90th percentile) and analyze the time evolution of deseasonalized convective organization (δI_{org}) and the deseasonalized 99th percentile of daily precipitation amount (δP_{99}) during composite strong-organization events, spanning 14 days before to 14 days after the peak organization day (Figure 12). Each variable is deseasonalized by subtracting the corresponding 30-day running-averaging values.

In IMERG, daily variations in extreme precipitation closely follow changes in convective organization, peaking at day 0. To further understand this relationship, we decompose P_{99} into precipitation intensity (I_{99}) and duration (D_{99}). D_{99} is the total rainy hours of the day which are the time when the hourly precipitation intensity >3 mm hour^{-1}. Then, I_{99} is diagnosed by dividing P_{99} by D_{99} which is the mean precipitation rate averaged only over the rainy period. The results indicate that while I_{99} also peaks at day 0, the dominant factor driving changes in

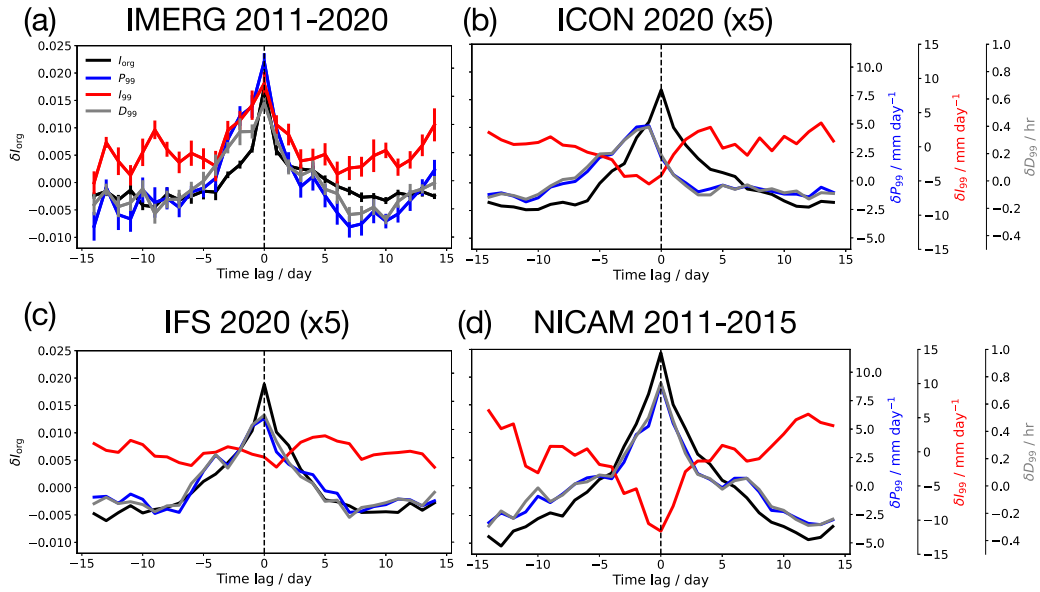

Figure 12. Time evolution of deseasonalized convective organization (I_{org}), 99th percentile of daily precipitation amount (P_{99}), intensity (I_{99}) and duration (D_{99}) during composite strong organization events from 14 days before to 14 days after the peak organization day. Both extreme precipitation statistics and convective organization are calculated over the whole tropics (20°S–20°N). The peak organization days are identified as δI_{org} larger than the 90th percentile of δI_{org} for each data set. Each variable is deseasonalized by subtracting the corresponding 30-day running-averaging values.

extreme precipitation is D_{99}. This suggests that the increase in total precipitation during periods of strong convective organization is primarily due to prolonged event duration rather than enhanced intensity. The models generally capture this relationship; however, ICON exhibits a timing mismatch, with extreme precipitation peaking a few days earlier than peak organization. This suggests that in ICON, extreme precipitation is not very sensitive to the degree of convective organization. This discrepancy is likely driven by weaker lateral mixing between convection and the environment. As a result, the onset of convection is less sensitive to ambient moisture, while convection itself contributes less to moistening the atmosphere. The reduced moisture-convection coupling further weakens mesoscale organization, which may in turn hinder larger-scale organization—potentially contributing to the underrepresentation of the MJO, as discussed in Section 6.2. These findings are consistent with earlier results and further highlight the critical role of moisture–convection coupling. Despite this difference, all models consistently show that the increase in extreme precipitation with organization is primarily driven by longer event duration, while intensity exhibits a weak (IFS) or even negative (ICON and NICAM) correlation with organization.

7. Summary and Concluding Remarks

The recent advancement of global kilometer-scale models enabling multi-year simulations has provided a powerful new tool for improving our understanding of the climate system. A major strength of these models lies in their high spatial resolution, which allows for a more realistic representation of convective processes and directly influences precipitation and a range of larger-scale climate phenomena. In the present study, we analyzed three global km-scale models, ICON, IFS and NICAM, which have been run for multiple years, and investigated the representation of convection and precipitation characteristics in these models. Unlike ICON and IFS, which are both fully coupled to ocean models, the SST in NICAM is predicted by a mixed-layer slab ocean model with a depth of 15 m. The three models also differ in the treatment of convection: while ICON and NICAM do not use convective parameterizations (both shallow and deep), IFS uses its default shallow convective parameterization and a modified deep convective parameterization, in which the cloud base mass flux is significantly reduced. To understand the impact of the convective parameterization on the discussed results, we also analyzed four IFS

sensitivity simulations at a slightly coarser resolution (9 km), which modify the activity of the shallow and deep convective schemes, from both on to both off.

By comparing model results with observational and reanalysis data sets, we have identified common biases in the precipitation intensity and convective cluster size across the three models. Compared to hourly mean rainfall data from the Integrated Multi-satellite Retrievals for GPM (IMERG), km-scale models with explicit convection tend to produce deep convective systems with excessively intense precipitation (Figure 3a), and weaker organization (Figure 4). The weak organization is associated with a larger number of smaller convective clusters (Figure 4). However, the extent of this bias varies significantly across models: IFS captures the organization metrics most closely aligned with observations, while ICON deviates most from the observed characteristics. It should also be noted that NICAM achieves stronger convective organization than ICON, even comparable to IFS, despite not using any cumulus parameterizations.

To understand the common biases in precipitation characteristics and convective organization—as well as their inter-model differences—we examine how thermodynamic–convection coupling is represented in the simulations. This coupling is crucial for the initiation of convection and subsequent tropospheric moistening, affecting the transition from shallow to deep convection and the overall moisture-convection feedback. Reanalysis data from ERA5 indicates that this transition unfolds gradually, as shallow convection moistens the lower troposphere and sets the stage for deep convection (Figure 6a). Although models capture this process qualitatively, they exhibit quantitative differences mainly in the initiation of convection (i.e., shallow-to-deep transition timescale) and moistening rate (Figures 6b–6f). These differences and their impacts on several atmospheric phenomena are summarized as follows:

1. ICON, and to a lesser extent IFS, initiate deep convection more rapidly than ERA5 and NICAM (see the slope of zero lines in Figure 6). This probably contributes to weak convective organization because convective instability is eliminated by fast local circulations and thus is hard to accumulate on wider spatial scales. The more rapid shallow-to-deep transition in ICON and IFS can be attributed to less convective sensitivity to moisture variations in the lower-free troposphere that control the dilution of buoyancy (Figure 7). In direct agreement with the convective initiation bias, the phase of the precipitation diurnal cycle (PDC) over land is leading in ICON/IFS and lagged in NICAM, relative to the TRMM-3G68 climatology (Figure 8b).
2. Compared to ERA5, all models underestimate tropospheric convective moistening (Figures 6b–6d), even though this bias is mitigated in IFS and NICAM. Compared to ICON, IFS exhibits greater column moistening prior to the peak of deep convection and increased moisture detrainment in the mid-troposphere afterward (Figure 6e). The stronger mid-tropospheric moistening is also simulated in NICAM (Figure 6f). The enhanced moistening of the environment surrounding deep convection—reducing the impact of entrainment on updraft buoyancy—provides a plausible explanation for the stronger convective organization in IFS and NICAM compared to ICON. Together with the first point, this effect can also influence planetary-scale phenomena such as the MJO. While both IFS and NICAM successfully simulate this propagation, ICON struggles, with convection stagnating around the Maritime Continent and exhibiting a weak propagation signal (Figure 11).

In addition to tropospheric processes involved in thermodynamic-convection coupling, we have also examined near-surface processes that influence precipitation characteristics. The intensity of cold pools is generally underestimated by all three km-scale models, although the degree of underestimation varies (Figures 5b and 5d); NICAM produces the strongest cold pools, followed by ICON and IFS. These biases subsequently affect how well the models capture the initiation and propagation of organized convective systems. Observations show that mean precipitation in coastal regions over the Maritime Continent is primarily characterized by strong PDCs associated with offshore-propagating convective systems (Figures 9a, 9b, and 10a–10c). Among the models, IFS exhibits the largest underestimation of PDC amplitudes in coastal oceanic regions, leading to deficient mean precipitation there, while NICAM shows the smallest biases (Figures 9c–9h). This underestimation arises from both the limited offshore propagation distance and the weaker intensity of offshore convection (Figures 10d–10f), both of which can be linked to the poor representation of interactions between cold pools and preexisting convective cells.

Our model intercomparison highlights two important implications for the development of km-scale models. First, it underscores the substantial uncertainties that persist even within explicit convection frameworks, so between ICON and NICAM. This mainly originates from differences in microphysics and turbulent mixing. In NICAM, rain and snow processes have been tuned to make convective initiation less likely and to promote moisture detrainment near the melting level (Takasuka, Kodama, et al., 2024), as shown in Section 5.1. This implies that

refining microphysics alters the heating rate and static stability, thereby significantly influencing the sensitivity of convection to its surrounding environment. Additionally, incorporating partial lateral mixing associated with deep convection (i.e., the Leonard terms) can improve the representation of entrainment and detrainment processes. In fact, NICAM simulates convective development and organization more realistically than ICON, likely due to stronger turbulent diffusion, whereas ICON relies solely on the Smagorinsky scheme. Nevertheless, km-scale resolutions with explicit convection are obviously too coarse to represent horizontal and vertical mixing caused by convective motions, as exemplified by too intense precipitation (Figure 3a) and less moistening during shallow-to-deep transition (Figure 6). Incorporating convective mixing into km-scale equation systems in a physically consistent manner remains a significant challenge (e.g., Schneider et al., 2024), in addition to selecting appropriate turbulent diffusion coefficients (Watanabe & Ito, 2024). Furthermore, the different convection characteristics could also be related to differences in the physics-dynamics coupling. As shown in Table 1, in NICAM, dynamics and non-radiative physics are tightly coupled in time, whereas in ICON they are updated on different time scales. This implies that the timescale of the resolved convective response to changes in heating and moistening differs between the models, which could affect mesoscale convection. Examining this aspect quantitatively would be the subject of future work.

For km-scale model development, another implication is related to the strengths and limitations of using cumulus parameterizations. Results from IFS suggest that a cumulus parameterization mitigates several biases of weak convective organization, intense precipitation, and weak column moistening associated with convective development. These good aspects probably result from cumulus eddies supplemented by parameterizations. Meanwhile, there are also negative impacts on convection characteristics, such as too frequent weak precipitation (Figure 3a), artificial separation between weak and strong precipitation regimes, and small deviations from quasi-equilibrium states (Figure 7c). These biases seem to be a legacy of well-known issues of cumulus parameterizations. It is plausible that parameterized cumulus eddies can supplement turbulent motions deficient at km-scale resolutions, but it is important to take care of the coupling between explicit dynamics and parameterized physics so as to avoid artificial scale breaks.

Typical model issues that originate from insufficient turbulent mixing and cloud microphysics uncertainties have long been recognized in regional km-scale simulations. Our results demonstrate that these local deficiencies do not remain confined to small scales. Instead, they propagate upscale, modulating the mesoscale convective organization (Figure 4) and altering the large-scale thermodynamic environment (Figure 2). The impacts accumulate and ultimately influence planetary-scale phenomena such as the MJO (Figure 11). Moreover, a robust correlation between PDC amplitudes and mean precipitation over the Maritime Continent reproduced by global km-scale models (Figure 8) provides strong evidence of the rectification of diurnal variability onto the mean state, as the large-scale precipitation patterns in these models emerge self-consistently from the internal energy redistribution without constraints by lateral boundaries. These aspects highlight advantages that are uniquely accessible within the global modeling framework, as opposed to regional frameworks.

The mesoscale thermodynamic-convection coupling and its cross-scale dependencies warrant further investigation through intercomparisons of a broader suite of global km-scale models with a wide range of physical configurations, including microphysics, turbulent mixing, and cumulus parameterizations. Beyond the topics addressed in the present study, representative phenomena involving multi-scale interactions that serve as natural testbeds for examining these aspects include the newly identified tropics-wide intraseasonal oscillation (Bao et al., 2025), the geographical variability of convectively coupled equatorial waves (Takayabu, 1994; Wheeler & Kiladis, 1999), and the mean state of tropical precipitation. All of these can be meaningfully examined using global kilometer-scale models. For example, even among the models examined here, there is substantial diversity in the structure, position, and amplitude of the ITCZ (Figure S8 in Supporting Information S1). Notably, ICON and IFS with $(D + S)_{off}$ exhibit a pronounced double ITCZ bias, and both models display similar thermodynamic–convection coupling, with reduced convective sensitivity in the LFT (Figure 7b and Figure S7d in Supporting Information S1) and weaker mid-tropospheric moistening associated with deep convection (Figure 6b and Figure S6d in Supporting Information S1). The robustness of this relationship should be further evaluated to clarify the impacts of convection on the climatological states (e.g., Bacmeister et al., 2006; Hirota et al., 2011). In this context, the third phase of the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains (DYAMOND) model intercomparison project—with extended simulation periods of 1 year or more (Takasuka, Satoh, et al., 2024)—offers a great opportunity to investigate the diversity of thermodynamic–convection coupling and its upscale impacts.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The IMERG data can be downloaded from https://gpm.nasa.gov/data/directory. The hourly climatology of the TRMM-3G68 product is available at https://geodynamics.sci.hokudai.ac.jp/poc/minobe/data/TRMM3G68_diurnal_clim/index.html. The daily OLR data from NOAA satellite is found at https://psl.noaa.gov/data/gridded/data.olrcdr.interp.html. The ERA5 reanalysis data are archived at https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels and https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels-monthly-means. The TAO buoy data are available online (https://tao.ndbc.noaa.gov/data-download/). Data from the ICON and IFS simulations are publicly accessible through the DOI provided below or directly from the DKRZ supercomputer Levante, subject to prior registration (https://luv.dkrz.de/register/, German Climate Computing Center (DKRZ), 2024). The nextGEMS Cycle 3 data for ICON and IFS can be found at WDCC under https://doi.org/10.26050/WDCC/nextGEMS_cyc3 (Koldunov et al., 2023). The NICAM data used for creating the figures can be downloaded from https://doi.org/10.6084/m9.figshare.29497769 (Takasuka, Becker, & Bao, 2025). The scripts used in the present work are available (Takasuka, Becker, & Bao, 2025). Figures 5–11 and Figures S6 and S7 in Supporting Information S1 are created by NCAR Command Language version 6.6.2, which can be installed via UCAR/NCAR/CISL/TDD (2019). The other figures are created with Python.

Acknowledgments

We thank Peter Bechtold, Lukas Brunner, Peter Dueben, Richard Forbes, Estibaliz Gascon, and Benoit Vanniere for providing insightful comments on the present study. We also thank Sebastian Milinski, Xabier Pedruzo and Thomas Rackow for their contributions to setting up IFS-FESOM for nextGEMS. We are also grateful to Dr. Walter Hannah and an anonymous reviewer for their constructive comments, which improved the original version of the manuscript. D. Takasuka was supported by JSPS KAKENHI Grants 20H05728 and 24K22893 and by JSPS Core-to-Core Program, "International Core-to-Core Project on Global Storm Resolving Analysis" (Grant Number: JPJSCCA20220001). T. Becker was supported by the Horizon 2020 project nextGEMS under grant agreement number 101003470. J. Bao acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant (grant agreement No 101034413). The ICON and IFS simulations were performed with supercomputing resources of the German Climate Computing Centre (Deutsches Klimarechenzentrum, DKRZ) granted by its Scientific Steering Committee (WLA) under project ID 1235. The NICAM simulation was performed on the supercomputer Fugaku (proposal numbers hp220132, hp230078, hp230108, hp230278, and hp240267).

References

Ahmed, F., & Neelin, J. D. (2019). Explaining scales and statistics of tropical precipitation clusters with a stochastic model. *Journal of the Atmospheric Sciences*, *76*(10), 3063–3087. https://doi.org/10.1175/JAS-D-18-0368.1

Arakawa, A., & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *Journal of the Atmospheric Sciences*, *31*(3), 674–701. https://doi.org/10.1175/1520-0469(1974)031⟨0674:IOACCE⟩2.0.CO;2

Back, S.-Y., Kim, D., & Son, S.-W. (2024). MJO diversity in CMIP6 models. *Journal of Climate*, *37*(18), 4835–4850. https://doi.org/10.1175/jcli-d-23-0656.1

Bacmeister, J. T., Suarez, M. J., & Robertson, F. R. (2006). Rain reevaporation, boundary layer–convection interactions, and Pacific rainfall patterns in an AGCM. *Journal of the Atmospheric Sciences*, *63*(12), 3383–3403. https://doi.org/10.1175/JAS3791.1

Baker, A. J., Vannière, B., & Vidale, P. L. (2024). On the realism of tropical cyclone intensification in global storm-resolving climate models. *Geophysical Research Letters*, *51*(17), e2024GL109841. https://doi.org/10.1029/2024GL109841

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., & Reinhardt, T. (2011). Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, *139*(12), 3887–3905. https://doi.org/10.1175/MWR-D-10-05013.1

Bao, J., Bony, S., Takasuka, D., & Muller, C. (2025). Tropics-wide intraseasonal oscillations. *Proceedings of the National Academy of Sciences*, *122*(48), e2511549122. https://doi.org/10.1073/pnas.2511549122

Bao, J., & Sherwood, S. C. (2019). The role of convective self-aggregation in extreme instantaneous versus daily precipitation. *Journal of Advances in Modeling Earth Systems*, *11*(1), 19–33. https://doi.org/10.1029/2018ms001503

Bao, J., Stevens, B., Kluft, L., & Muller, C. (2024). Intensification of daily tropical precipitation extremes from more organized convection. *Science Advances*, *10*(8), eadj6801. https://doi.org/10.1126/sciadv.adj6801

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M. J., et al. (2008). Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, *134*(634), 1337–1351. https://doi.org/10.1002/qj.289

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A., & Bormann, N. (2014). Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, *71*(2), 734–753. https://doi.org/10.1175/JAS-D-13-0163.1

Becker, T., Bechtold, P., & Sandu, I. (2021). Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4388–4407. https://doi.org/10.1002/qj.4185

Becker, T., Bretherton, C. S., Hohenegger, C., & Stevens, B. (2018). Estimating bulk entrainment with unaggregated and aggregated convection. *Geophysical Research Letters*, *45*(1), 455–462. https://doi.org/10.1002/2017gl076640

Brousseau, P., Seity, Y., Ricard, D., & Léger, J. (2016). Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, *142*(699), 2231–2243. https://doi.org/10.1002/qj.2822

Brunner, L., Poschlod, B., Dutra, E., Fischer, E. M., Martius, O., & Sillmann, J. (2025). A global perspective on the spatial representation of climate extremes from km-scale models. *Environmental Research Letters*, *20*(7), 074054. https://doi.org/10.1088/1748-9326/ade1ef

Chen, G., Ling, J., Zhang, R., Xiao, Z., & Li, C. (2022). The MJO from CMIP5 to CMIP6: Perspectives from tracking MJO precipitation. *Geophysical Research Letters*, *49*(1), e2021GL095241. https://doi.org/10.1029/2021GL095241

Chen, J., Hagos, S., Xiao, H., Fast, J. D., & Feng, Z. (2020). Characterization of surface heterogeneity-induced convection using cluster analysis. *Journal of Geophysical Research: Atmospheres*, *125*(20), e2020JD032550. https://doi.org/10.1029/2020JD032550

Crook, J., Klein, C., Folwell, S., Taylor, C. M., Parker, D. J., Stratton, R., & Stein, T. (2019). Assessment of the representation of West African storm lifecycles in convection-permitting simulations. *Earth and Space Science*, *6*(5), 818–835. https://doi.org/10.1029/2018EA000491

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, *19*(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1

Danilov, S., Sidorenko, D., Wang, Q., & Jung, T. (2017). The finite-volume sea ice–ocean model (FESOM2). *Geoscientific Model Development*, *10*(2), 765–789. https://doi.org/10.5194/gmd-10-765-2017

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Derbyshire, S., Beau, I., Bechtold, P., Grandpeix, J.-Y., Piriou, J.-M., Redelsperger, J.-L., & Soares, P. (2004). Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3055–3079. https://doi.org/10.1256/qj.03.130

Dickinson, R. E., Errico, R. M., Giorgi, F., & Bates, G. T. (1989). A regional climate model for the western United States. *Climatic Change*, *15*(3), 383–422. https://doi.org/10.1007/BF00240465

Ding, T., Guo, Z., Zou, L., & Zhou, T. (2023). Impact of convection-permitting and model resolution on the simulation of mesoscale convective system properties over East Asia. *Journal of Geophysical Research: Atmospheres*, *128*(24), e2023JD039395. https://doi.org/10.1029/2023jd039395

Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., & Brdar, S. (2015). Large eddy simulation using the general circulation model icon. *Journal of Advances in Modeling Earth Systems*, *7*(3), 963–986. https://doi.org/10.1002/2015MS000431

ECMWF. (2023). ECMWF IFS documentation CY48R1 – Part IV Physical processes. *IFS Documentation CY48R1*. https://doi.org/10.21957/02054f0fbf

Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, *124*(11), 5696–5723. https://doi.org/10.1029/2018JD030094

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., et al. (2020). The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, *12*(4), 720. https://doi.org/10.3390/rs12040720

Grabowski, W. W., & Moncrieff, M. W. (2004). Moisture–convection feedback in the tropics. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3081–3104. https://doi.org/10.1256/qj.03.135

Hanesiak, J. M., Raddatz, R. L., & Lobban, S. (2004). Local initiation of deep convection on the Canadian prairie provinces. *Boundary-Layer Meteorology*, *110*(3), 455–470. https://doi.org/10.1023/B:BOUN.0000007242.89023.e5

Hannah, W. M., & Maloney, E. D. (2011). The role of moisture-convection feedbacks in simulating the Madden-Julian oscillation. *Journal of Climate*, *24*(11), 2754–2770. https://doi.org/10.1175/2011JCLI3803.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hirota, N., Takayabu, Y. N., Watanabe, M., & Kimoto, M. (2011). Precipitation reproducibility over tropical oceans and its relationship to the double ITCZ problem in CMIP3 and MIROC5 climate models. *Journal of Climate*, *24*(18), 4859–4873. https://doi.org/10.1175/2011jcli4156.1

Hirt, M., Craig, G. C., Schäfer, S. A. K., Savre, J., & Heinze, R. (2020). Cold-pool-driven convective initiation: Using causal graph analysis to determine what convection-permitting models are missing. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 2205–2227. https://doi.org/10.1002/qj.3788

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., et al. (2023). ICON-Sapphire: Simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, *16*(2), 779–811. https://doi.org/10.5194/gmd-16-779-2023

Holloway, C. E., Woolnough, S. J., & Lister, G. M. S. (2013). The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. *Journal of the Atmospheric Sciences*, *70*(5), 1342–1369. https://doi.org/10.1175/JAS-D-12-0227.1

Houze, R. A., Jr. (1981). Structures of atmospheric precipitation systems: A global survey. *Radio Science*, *16*(5), 671–689. https://doi.org/10.1029/RS016i005p00671

Kain, J. S., & Fritsch, J. M. (1990). A one-dimensional entraining/detraining plume model and its application in convective parameterization. *Journal of the Atmospheric Sciences*, *47*(23), 2784–2802. https://doi.org/10.1175/1520-0469(1990)047⟨2784:AODEPM⟩2.0.CO;2

Kikuchi, K., & Wang, B. (2008). Diurnal precipitation regimes in the global tropics. *Journal of Climate*, *21*(11), 2680–2696. https://doi.org/10.1175/2007JCLI2051.1

Kim, D., Xavier, P., Maloney, E., Wheeler, M., Waliser, D., Sperber, K., et al. (2014). Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *Journal of Climate*, *27*(14), 5379–5395. https://doi.org/10.1175/JCLI-D-13-00497.1

Kodama, C., Ohno, T., Seiki, T., Yashiro, H., Noda, A. T., Nakano, M., et al. (2021). The Nonhydrostatic ICosahedral Atmospheric Model for CMIP6 HighResMIP simulations (NICAM16-S): Experimental design, model description, and impacts of model updates. *Geoscientific Model Development*, *14*(2), 795–820. https://doi.org/10.5194/gmd-14-795-2021

Koldunov, N., Kölling, T., Pedruzo-Bagazgoitia, X., Rackow, T., Redler, R., Sidorenko, D., et al. (2023). nextGEMS: Output of the model development cycle 3 simulations for ICON and IFS [Dataset]. *World Data Center for Climate (WDCC) at DKRZ*. https://doi.org/10.26050/WDCC/nextGEMS_cyc3

Lebo, Z. J., & Morrison, H. (2015). Effects of horizontal and vertical grid spacing on mixing in simulated squall lines and implications for convective strength and structure. *Monthly Weather Review*, *143*(11), 4355–4375. https://doi.org/10.1175/MWR-D-15-0154.1

Lee, J., & Hohenegger, C. (2024). Weaker land–atmosphere coupling in global storm-resolving simulation. *Proceedings of the National Academy of Sciences*, *121*(12), e2314265121. https://doi.org/10.1073/pnas.2314265121

Lee, J., Hohenegger, C., Chlond, A., & Schnur, R. (2022). The climatic role of interactive leaf phenology in the vegetation-atmosphere system of radiative-convective equilibrium storm-resolving simulations. *Tellus B: Chemical and Physical Meteorology*, *74*(1), 164. https://doi.org/10.16993/tellusb.26

Liebmann, B., & Smith, C. (1996). Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society*, *77*, 1275–1277.

Madden, R. A., & Julian, P. R. (1972). Description of global-scale circulation cells in the tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, *29*(6), 1109–1123. https://doi.org/10.1175/1520-0469(1972)029⟨1109:DOGSCC⟩2.0.CO;2

Mapes, B., & Neale, R. (2011). Parameterizing convective organization to escape the entrainment dilemma. *Journal of Advances in Modeling Earth Systems*, *3*(2), M06004. https://doi.org/10.1029/2011MS000042

Maranan, M., Fink, A. H., Knippertz, P., Amekudzi, L. K., Atiah, W. A., & Stengel, M. (2020). A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the West African forest zone. *Journal of Hydrometeorology*, *21*(4), 729–749. https://doi.org/10.1175/JHM-D-19-0257.1

Masunaga, H. (2013). A satellite study of tropical moist convection and environmental variability: A moisture and thermal budget analysis. *Journal of the Atmospheric Sciences*, *70*(8), 2443–2466. https://doi.org/10.1175/JAS-D-12-0273.1

Minobe, S., Park, J. H., & Virts, K. S. (2020). Diurnal cycles of precipitation and lightning in the tropics observed by TRMM3G68, GSMaP, LIS, and WWLLN. *Journal of Climate*, *33*(10), 4293–4313. https://doi.org/10.1175/JCLI-D-19-0389.1

Miura, H., Satoh, M., Nasuno, T., Noda, A. T., & Oouchi, K. (2007). A Madden-Julian oscillation event realistically simulated by a global cloud-resolving model. *Science*, *318*(5857), 1763–1765. https://doi.org/10.1126/science.1148443

Miura, H., Suematsu, T., Kawai, Y., Yamagami, Y., Takasuka, D., Takano, Y., et al. (2023). Asymptotic matching between weather and climate models. *Bulletin of the American Meteorological Society*, *104*(12), E2308–E2315. https://doi.org/10.1175/BAMS-D-22-0128.1

Moeng, C.-H., Sullivan, P., Khairoutdinov, M., & Randall, D. (2010). A mixed scheme for subgrid-scale fluxes in cloud-resolving models. *Journal of the Atmospheric Sciences*, *67*(11), 3692–3705. https://doi.org/10.1175/2010JAS3565.1

Moorthi, S., & Suarez, M. J. (1992). Relaxed Arakawa-Schubert. A parameterization of moist convection for general circulation models. *Monthly Weather Review*, *120*(6), 978–1002. https://doi.org/10.1175/1520-0493(1992)120⟨0978:RASAPO⟩2.0.CO;2

Mori, S., Jun-Ichi, H., Tauhid, Y. I., Yamanaka, M. D., Okamoto, N., Murata, F., et al. (2004). Diurnal land–sea rainfall peak migration over Sumatera Island, Indonesian Maritime Continent, observed by TRMM satellite and intensive rawinsonde soundings. *Monthly Weather Review*, *132*(8), 2021–2039. https://doi.org/10.1175/1520-0493(2004)132⟨2021:DLRPMO⟩2.0.CO;2

Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*(2), 397–407. https://doi.org/10.1007/s10546-005-9030-8

Nesbitt, S. W., Cifelli, R., & Rutledge, S. A. (2006). Storm morphology and rainfall characteristics of TRMM precipitation features. *Monthly Weather Review*, *134*(10), 2702–2721. https://doi.org/10.1175/MWR3200.1

Okugawa, R., Yasunaga, K., Hamada, A., & Yokoi, S. (2024). Numerical study on the precipitation concentration over the western coast of Sumatra Island. *Monthly Weather Review*, *152*(3), 689–704. https://doi.org/10.1175/MWR-D-23-0037.1

Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3074–3089. https://doi.org/10.1029/2019MS001621

Prein, A. F., Rasmussen, R., Wang, D., & Giangrande, S. (2021). Sensitivity of organized convective storms to model grid spacing in current and future climates. *Philosophical Transactions of the Royal Society A*, *379*(2195), 20190546. https://doi.org/10.1098/rsta.2019.0546

Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., et al. (2025). Multi-year simulations at kilometre scale with the integrated forecasting system coupled to FESOM2.5 and NEMOv3.4. *Geoscientific Model Development*, *18*(1), 33–69. https://doi.org/10.5194/gmd-18-33-2025

Randall, D., Harshvardhan, & Dazlich, D. (1991). Diurnal variability of the hydrologic cycle in a general circulation model. *Journal of the Atmospheric Sciences*, *48*(1), 40–62. https://doi.org/10.1175/1520-0469(1991)048⟨0040:DVOTHC⟩2.0.CO;2

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Raymond, D. J. (2000). Thermodynamic control of tropical rainfall. *Quarterly Journal of the Royal Meteorological Society*, *126*(564), 889–898. https://doi.org/10.1002/qj.49712656406

Riehl, H., & Malkus, J. S. (1958). On the heat balance in the equatorial trough zone. *Geophysica*, *6*, 503–537.

Roca, R., Aublanc, J., Chambon, P., Fiolleau, T., & Viltard, N. (2014). Robust observational quantification of the contribution of mesoscale convective systems to rainfall in the tropics. *Journal of Climate*, *27*(13), 4952–4958. https://doi.org/10.1175/JCLI-D-13-00628.1

Roh, W., & Satoh, M. (2014). Evaluation of precipitating hydrometeor parameterizations in a single-moment bulk microphysics scheme for deep convective systems over the tropical central Pacific. *Journal of the Atmospheric Sciences*, *71*(7), 2654–2673. https://doi.org/10.1175/JAS-D-13-0252.1

Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences*, *45*(3), 463–485. https://doi.org/10.1175/1520-0469(1988)045⟨0463:ATFSLL⟩2.0.CO;2

Sato, T., Miura, H., Satoh, M., Takayabu, Y. N., & Wang, Y. (2009). Diurnal cycle of precipitation in the tropics simulated in a global cloud-resolving model. *Journal of Climate*, *22*(18), 4809–4826. https://doi.org/10.1175/2009JCLI2890.1

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., & Iga, S. (2008). Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, *227*(7), 3486–3514. https://doi.org/10.1016/j.jcp.2007.02.006

Satoh, M., Tomita, H., Yashiro, H., Miura, H., Kodama, C., Seiki, T., et al. (2014). The Non-hydrostatic Icosahedral Atmospheric Model: Description and development. *Progress in Earth and Planetary Science*, *1*(1), 1–32. https://doi.org/10.1186/s40645-014-0018-1

Schneider, T., Leung, L. R., & Wills, R. C. J. (2024). Opinion: Optimizing climate models with process knowledge, resolution, and artificial intelligence. *Atmospheric Chemistry and Physics*, *24*(12), 7041–7062. https://doi.org/10.5194/acp-24-7041-2024

Scholz, P., Sidorenko, D., Gurses, O., Danilov, S., Koldunov, N., Wang, Q., et al. (2019). Assessment of the Finite-volumE Sea ice-Ocean Model (FESOM2.0) – Part 1: Description of selected key model elements and comparison to its predecessor version. *Geoscientific Model Development*, *12*(11), 4875–4899. https://doi.org/10.5194/gmd-12-4875-2019

Segura, H., Pedruzo-Bagazgoitia, X., Weiss, P., Müller, S. K., Rackow, T., Lee, J., et al. (2025). nextGEMS: Entering the era of kilometer-scale Earth system modeling. *EGUsphere*, *2025*, 1–39. https://doi.org/10.5194/egusphere-2025-509

Spät, D., Biasutti, M., Schuhbauer, D., & Voigt, A. (2024). Autocorrelation—A simple diagnostic for tropical precipitation variability in global kilometer-scale climate models. *Geophysical Research Letters*, *51*(17), e2024GL108856. https://doi.org/10.1029/2024GL108856

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., et al. (2010). Dreary state of precipitation in global models. *Journal of Geophysical Research*, *115*(D24), D24211. https://doi.org/10.1029/2010JD014532

Stevens, B. (2024). A perspective on the future of CMIP. *AGU Advances*, *5*(1), e2023AV001086. https://doi.org/10.1029/2023AV001086

Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., et al. (2020). The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan. Series II*, *98*(2), 395–435. https://doi.org/10.2151/jmsj.2020-021

Suematsu, T., & Miura, H. (2018). Zonal SST difference as a potential environmental factor supporting the longevity of the Madden-Julian oscillation. *Journal of Climate*, *31*(18), 7549–7564. https://doi.org/10.1175/JCLI-D-17-0822.1

Takasuka, D., Becker, T., & Bao, J. (2025). Data and scripts for Takasuka et al. (2025) submitted to JAMES [Dataset]. *Figshare*. https://doi.org/10.6084/m9.figshare.29497769

Takasuka, D., Kodama, C., Suematsu, T., Ohno, T., Yamada, Y., Seiki, T., et al. (2024). 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*(2), e2023MS003701. https://doi.org/10.1029/2023MS003701

Takasuka, D., Kohyama, T., Suematsu, T., & Miura, H. (2025). ENSO and QBO controls the favorableness of the MJO realization cooperatively. *Journal of Geophysical Research: Atmospheres*, *130*(2), e2024JD042116. https://doi.org/10.1029/2024JD042116

Takasuka, D., Satoh, M., Miyakawa, T., Kodama, C., Klocke, D., Stevens, B., et al. (2024). A protocol and analysis of year-long simulations of global storm-resolving models and beyond. *Progress in Earth and Planetary Science*, *11*(1), 66. https://doi.org/10.1186/s40645-024-00668-1

Takayabu, Y. N. (1994). Large-scale cloud disturbances associated with equatorial waves Part I: Spectral features of the cloud disturbances. *Journal of the Meteorological Society of Japan. Series II*, *72*(3), 433–449. https://doi.org/10.2151/jmsj1965.72.3_433

Takemi, T., Hirayama, O., & Liu, C. (2004). Factors responsible for the vertical development of tropical oceanic cumulus convection. *Geophysical Research Letters*, *31*(11), L11109. https://doi.org/10.1029/2004GL020225

Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). IMERG V06: Changes to the morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, *36*(12), 2471–2482. https://doi.org/10.1175/JTECH-D-19-0114.1

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, *117*(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117⟨1779:ACMFSF⟩2.0.CO;2

Tomita, H. (2008). New microphysical schemes with five and six categories by diagnostic generation of cloud ice. *Journal of the Meteorological Society of Japan*, *86A*, 121–142. https://doi.org/10.2151/jmsj.86A.121

Tomita, H., Miura, H., Iga, S.-I., Nasuno, T., & Satoh, M. (2005). A global cloud-resolving simulation: Preliminary results from an aqua planet experiment. *Geophysical Research Letters*, *32*(8), L08805. https://doi.org/10.1029/2005GL022459

Tomita, H., & Satoh, M. (2004). A new dynamical framework of nonhydrostatic global model using the icosahedral grid. *Fluid Dynamics Research*, *34*(6), 357–400. https://doi.org/10.1016/j.fluiddyn.2004.03.003

Tomita, H., Satoh, M., & Goto, K. (2002). An optimization of the icosahedral grid modified by spring dynamics. *Journal of Computational Physics*, *183*(1), 307–331. https://doi.org/10.1006/jcph.2002.7193

Tompkins, A. M., & Craig, G. C. (1998). Radiative–convective equilibrium in a three-dimensional cloud-ensemble model. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 2073–2097. https://doi.org/10.1002/qj.49712455013

Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, *9*(2), 1046–1068. https://doi.org/10.1002/2016MS000802

UCAR/NCAR/CISL/TDD. (2019). The NCAR Command Language (Version 6.6.2) [Software]. *UCAR/NCAR/CISL/TDD*. https://doi.org/10.5065/D6WD3XH5

Wang, B., Chen, G., & Liu, F. (2019). Diversity of the Madden-Julian oscillation. *Science Advances*, *5*(7), 2524–2536. https://doi.org/10.1126/sciadv.aax0220

Watanabe, S.-I. I., & Ito, J. (2024). A priori test of scale-similarity parameterizations for subgrid-scale fluxes in convective storms. *Monthly Weather Review*, *152*(10), 2229–2246. https://doi.org/10.1175/MWR-D-24-0059.1

Wheeler, M., & Hendon, H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, *132*(8), 1917–1932. https://doi.org/10.1175/1520-0493(2004)132⟨1917:AARMMI⟩2.0.CO;2

Wheeler, M., & Kiladis, G. N. (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *Journal of the Atmospheric Sciences*, *56*(3), 374–399. https://doi.org/10.1175/1520-0469(1999)056<0374:ccewao>2.0.co;2

White, B. A., Buchanan, A. M., Birch, C. E., Stier, P., & Pearson, K. J. (2018). Quantifying the effects of horizontal grid length and parameterized convection on the degree of convective organization using a metric of the potential for convective interaction. *Journal of the Atmospheric Sciences*, *75*(2), 425–450. https://doi.org/10.1175/JAS-D-16-0307.1

Wille, J. D., Koch, R., Becker, T., & Fischer, E. (2025). Extreme precipitation depiction in convection-permitting Earth System Models within the nextGEMS project. *Journal of Advances in Modeling Earth Systems*, *17*(7), e2024MS004840. https://doi.org/10.1029/2024MS004840

Windmiller, J. M., & Stevens, B. (2024). The inner life of the Atlantic Intertropical Convergence Zone. *Quarterly Journal of the Royal Meteorological Society*, *150*(758), 523–543. https://doi.org/10.1002/qj.4610

Wolding, B., Powell, S. W., Ahmed, F., Dias, J., Gehne, M., Kiladis, G., & Neelin, J. D. (2022). Tropical thermodynamic–convection coupling in observations and reanalyses. *Journal of the Atmospheric Sciences*, *79*(7), 1781–1803. https://doi.org/10.1175/JAS-D-21-0256.1

Yang, G.-Y., & Slingo, J. (2001). The diurnal cycle in the tropics. *Monthly Weather Review*, *129*(4), 784–801. https://doi.org/10.1175/1520-0493(2001)129⟨0784:TDCITT⟩2.0.CO;2

Yokoi, S., & Kajikawa, Y. (2024). Precipitation diurnal cycle over tropical coastal regions represented in climate experiments with a global cloud-system resolving model. *SOLA*, *20*, 145–151. https://doi.org/10.2151/sola.2024-020

Yokoi, S., Mori, S., Katsumata, M., Geng, B., Yasunaga, K., Syamsudin, F., et al. (2017). Diurnal cycle of precipitation observed in the western coastal area of Sumatra Island: Offshore preconditioning by gravity waves. *Monthly Weather Review*, *145*(9), 3745–3761. https://doi.org/10.1175/MWR-D-16-0468.1

Yokoi, S., Mori, S., Syamsudin, F., Haryoko, U., & Geng, B. (2019). Environmental conditions for nighttime offshore migration of precipitation area as revealed by in situ observation off Sumatra Island. *Monthly Weather Review*, *147*(9), 3391–3407. https://doi.org/10.1175/MWR-D-18-0412.1

Yoneyama, K., & Zhang, C. (2020). Years of the Maritime Continent. *Geophysical Research Letters*, *47*(12), e2020GL087182. https://doi.org/10.1029/2020GL087182

Zhang, C., & Ling, J. (2017). Barrier effect of the Indo-Pacific Maritime Continent on the MJO: Perspectives from tracking MJO precipitation. *Journal of Climate*, *30*(9), 3439–3459. https://doi.org/10.1002/2016JD025665

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. https://doi.org/10.1080/07055900.1995.9649539