https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Spin-up in humidity and temperature and its consequences for convective diagnostics: a Model Uncertainty Model Intercomparison Project experiment

Edward Groot[1], Hannah Christensen[1], Xia Sun[2], Kathryn Newman[2], Wahiba Lfarh[3], Romain Roehrig[3], Lisa Bengtsson[2], Julia Simonson[2], Keith Williams[4], and Hugo Lambert[5]

[1]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Sherrington Road, Oxford, United Kingdom of Great Britain
[2]National Oceanic and Atmospheric Administration, Boulder, Colorado, USA
[3]Météo-France, CNRS, Univ. Toulouse, CNRM, Toulouse, France
[4]UK MetOffice, Exeter, United Kingdom of Great Britain
[5]University of Exeter (Math & Statistics), Exeter, United Kingdom of Great Britain

Correspondence: Edward Groot (large.edward.simulations@gmail.com; edward.groot@physics.ox.ac.uk)

Abstract. We analyse the evolution of convective diagnostics such as mixed-layer convective available potential energy (CAPE), level of neutral buoyancy and precipitation rate as a function of lead time in the model uncertainty model-intercomparison project. Four model physics packages are exposed to common dynamics to form a large single-column model dataset. We analyse tendencies in an equatorial band over the Indian Ocean out to 6 hr lead time over one month. We prescribe dynamics and initial conditions from an ICON-DYAMOND simulation after coarse-graining to 0.2 degrees. The physics suites represent state-of-the-art global numerical weather and climate prediction models.

Correlation analysis shows that the spatial mean change of CAPE is not associated with precipitation rate, but it correlates very well with mean mixed-layer drying across our suites. This systematic drying occurs below 700 hPa in some suites, especially in the first hour. The sub-grid physics adjusts the initialised ICON state towards the native climate of each physics suite, in particular at low levels.

We apply a column-by-column empirical orthogonal function (EOF) analysis to a two-layer representation of physics and dynamics tendencies, CAPE tendency and precipitation rate. The first EOF is associated with free-tropospheric tendencies and nearly all precipitation variability, with neat compensation between physics and dynamics tendencies. The second and third EOFs of each suite indicate that a imbalance between these terms in the mixed-layer correlates with the CAPE change at least one of them, which are explained by temperature and humidity adjustments, but with little imprint on precipitation.

1 Introduction

Communal model intercomparison exercises may have serve various objectives, such as identifying the range of representations of specific processes, feedbacks or spatiotemporal scales (e.g. Eyring et al., 2016; Wing et al., 2020; Haarsma et al., 2016).

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

They may improve understanding of errors, including those in underlying model parameterisations. Biases between individual models may foster understanding of each model on weather to climate time scales (e.g. Soden and Bretherton, 1994; Sengupta and Boyle, 1998; Rodwell and Palmer, 2007; Williams et al., 2013; Bechtold et al., 2014b; Xavier et al., 2015; Lang et al., 2023). This enterprise serves as learning opportunity for developers (Sengupta and Boyle, 1998; Rodwell and Palmer, 2007; Williams et al., 2013; Schneider et al., 2017).

The biases form an important term in error budgets (e.g. Rodwell et al., 2018). Lead-time-dependent errors converge to lead-time-dependent model biases if they persist in datasets with increasing sample sizes (upon statistical convergence). Insight into the variation in underlying tendencies is critical for learning about model biases, model adjustment and, ultimately, errors. While model tendencies are often utilised to investigate a single model, specific processes, individual parameterisations and/or variables (e.g. Shutts and Palmer, 2007; Lang et al., 2023), the sources of tendency variance can be linked for targeted uncertainty assessment (e.g. Williams et al., 2013; Rodwell et al., 2016, 2018; Christensen, 2020; Bengtsson et al., 2019). Given the wide focus on individual schemes and variables, the linking, budget and constraining opportunities appear under-represented in literature, especially efforts to bridge between the weather-oriented and climate-focused communities.

In this work, we analyse the evolution of convective diagnostics with lead time across multiple models as part of the Model Uncertainty Model Intercomparison Project (MUMIP). We quantify associated adjustments, which occur after intialising each model from the common non-native reference state. When models are initialized from such a non-native state, the models adjust this state during a spin-up phase until they reach some equilibrium between their internal behavior and the external forcing. It is thus important to develop methods that could help identifying such polluting spin-up variability and distinguish it from physically meaningful variability. The aim of this study is to apply such a method to our dataset.

In MUMIP model physics ("sub-grid") tendencies under prescribed common dynamics ("advective, large-scale") tendencies from a parent simulation, with an approach similar to Christensen (2020), but extended to multiple single-column models (SCMs). Furthermore, MUMIP attempts to isolate the sub-grid parameterised physics uncertainty from uncertainty caused by resolved model dynamics. When sub-grid physics (model physics) is not isolated, two-way-feedbacks between model physics and the resolved dynamics are inevitable. By isolating model physics, we explicitly remove the two-way-feedbacks. This artificial isolation is very important because model physics is known to dominate uncertainties in atmospheric modelling (e.g. Groot and Riemer, 2025; Selz et al., 2022; Baumgart et al., 2019; Zhang et al., 2019; Groot and Tost, 2023a; Melhauser and Zhang, 2012). However, consequently, the assumption eliminates uncertainty resolved explicitly by dynamical cores from our assessments, although it is known that they are not flawless.

We have built a comprehensive tendency dataset of the model physics of various models after exposing their common initial conditions and forcing. This forcing is derived from the simulation of the Icosahedral Nonhydrostatic model (ICON) performed at 2.5 km grid spacing within the first phase of the DYAMOND project (Stevens et al., 2019) ("observations", validation of MUMIP). The models that participated in the MUMIP initiative include the Integrated Forecasting System (IFS) developed at the European Centre for Medium-Range Weather Forecasting (ECMWF), the Global Forecast System (GFS) and Rapid Refresh (RAP) system developed at the National Oceanic and Atmospheric Administration (NOAA) in the United States and the ARPEGE-Climat model (hereafter: ARPEGE), which is the atmospheric component of the climate model developed at

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

the Centre National de Rercherches Météorologiques, Météo-France (Forbes et al., 2021; gfs; rap; Roehrig et al., 2020). Each of them runs in as SCM assuming a horizontal resolution of 22 km, whenever relevant. As opposed to other SCMs, NOAA's physics suites have a common SCM host, namely the Common Community Physics Package (CCPP) (Bernardet et al., 2024). Similar to Williams et al. (2013), here we try to understand the structure and links behind tendencies and adjustments.

Finally, we focus on weather time scales as opposed to solely at climate time scales in many of the other "MIPs". In numerical weather forecasts the numerical models are known to be more accurate than in free-running simulations (e.g. Bauer et al., 2015). This is due to initialisation using data assimilation, and to a lesser extent, due to finer grids in horizontal and vertical direction, tailored perturbation techniques and smaller time steps. At weather time scales, the initialisation and perturbation techniques can be validated separately, which supposes more rigorous uncertainty sampling. A major advantage of incorporating weather time scales in our atmospheric analysis is that unknown unsampled uncertainties in a climate context may be isolated in a realistic modeling framework (e.g. Groot et al., 2026b; Christensen, 2020).

Our MUMIP domain covers most of the Indian Ocean, where convective instability and precipitating clouds dominate the (sub-)tropical atmosphere. Therefore, distributions of convective available potential energy (CAPE), both stratiform and convective precipitation need to be known accurately, as well as the cloud outflow levels. This level can be estimated with the level of neutral buoyancy. Constraining them is needed because convective processes like convective initiation may depend on their description by models (e.g. Bechtold et al., 2014b; Mapes, 1993; Schumacher et al., 2004; Houze, 2004; Groot et al., 2024; Groot and Tost, 2023b). Models, reanalysis and observations do not fully agree on their magnitude and distribution (e.g. Buschow, 2024). Hence, we could argue that the true distribution of CAPE and precipitation are not known.

We quantify the distribution convective diagnostics within the MUMIP dataset. Furthermore, we apply a principal component analysis to key physics tendencies and convection diagnostics from a two-layer-perspective to further identify the main contributions their variability over our pseudo-Indian Ocean domain. The dynamics of this area is dominated by deep-convective cloud development and its boundary with quiet high pressure systems in the subtropics (e.g. Sherwood et al., 2010). We will also explore whether, and how, the change of CAPE and state variables are associated with precipitation (see, for instance, Buschow, 2024). Our overall aim is to quantify spin-up and physically meaningful variability with the statistical analyses. Thus, we here identify which spin-up plays an important role in the MUMIP and whether physics tendency patterns can be generalised across multiple models.

In Section 2 we describe our simulation methods, datasets and diagnostics. In Section 3, we describe the results. In Section 4, we discuss implications of the results for the interpretation of the MUMIP dataset, convective diagnostics and the wider modeling context.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

2 Methods

2.1 Model uncertainty model intercomparison project: protocol

In the model uncertainty model intercomparison project (MUMIP) we aim to build a structured dataset of model uncertainty for physics packages utilised across numerical weather prediction and climate models. For this purpose, single column versions of each model's native physics suites are integrated forward in time over an Indian-Ocean-like domain, using the strategies of Christensen et al. (2018) and Christensen (2020). We select the single-column configuration, because this allows for controlling the dynamical part of our tendencies, which is explicitly resolved in regular three-dimensional configurations of atmospheric models. Initial conditions are borrowed from a common benchmark of the DYAMOND project (Stevens et al., 2019).

For each model we integrate 10.780.000 single column simulations over our domain mimicking this DYAMOND simulation. A fixed three-hourly dataset drives the dynamics and initial conditions of single columns from 11 August 2016 until 10 September 2016. This benchmark dataset consists of the ICON convection-permitting output, coarse-grained to a spatial resolution of 0.2 degrees (Christensen et al., 2018; Christensen, 2020). Furthermore, topography (small islands) have been removed by bi-linear interpolation. Hence, our domain represents as an ocean domain, very closely resembling the real Indian Ocean - from 51 degrees east to 95 degrees east and from 35 degrees south to 5 degrees north. Finally, above an altitude of about 45 km, which is the top level available in the DYAMOND archive, profiles are initialised in an arbitrary way selected by each modeling center [Christensen et al., in preparation].

The full dataset is linearly interpolated both in space and time to match the vertical grid and time discretization (5 or 10 min) of each hosting model. The horizontal grid spacing is set to about 0.2 degrees, consistently with the source data. Hence, convection is parameterised in the hosting models. Surface forcing can either be set to prescribed sea surface temperatures - which evolve with time - or, alternatively, prescribed surface fluxes. In each case their values are derived from ICON. The arbitrary choice has an effectively negligible effect on boundary-layer temperature and humidity. An extensive sensitivity test with IFS in each of the settings is in preparation [Groot et al., in preparation]. All further settings are meant to closely resemble the default of the host models.

We initialise the models every three hours and span 31-days. For each initialisation, we run them out to six hours lead time. The pseudo-Indian-Ocean domain consists of 44.000 tiles initiated for up to 245 fixed sets of initial conditions. From all columns we extract major subset that could be cross-validated to mimic ICON dynamically, which systematically excludes the 21 UTC initialisation and, furthermore, one 18 UTC initialisation (see also Groot et al., 2026b). All other columns are investigated.

For more details of the simulation protocol, we refer to (Christensen et al., in preparation, Groot et al., 2026b, and Groot et al., in preparation,).

Since any column is

- connected to its "neighbouring" columns through advective tendencies and geostrophic wind forcing derived from pressure gradients
- experiences a NWP column dynamics reasonably close to realism

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

- experiences realistic vertical advection tendency, coarse-grained from the benchmark - ICON

our configuration constrains the model physics tendencies of each physics suite. This is because the physics tendency needs to be consistent with the preceeding SCM state and benchmark forcing before and after any time step. These constraints may act similar to an active dynamical feedback between an NWP model and a column, but their nature remains common and fixed in our configuration.

Therefore, there may be a statistical misbalance between the dynamics and the physics, stimulating adjustments as if ICON-like "observations" and dynamics are assimilated into our SCMs. We exploit this statistical misbalance in our principal component analysis to identify spin-up and regular, statistically balanced tendency components from physics suites.

The archive interval of SCM output is 30 or 60 mins.

2.2 Diagnostics

Key quantities under investigation are the following:

- Convective available potential energy (CAPE), which is the vertical integral of buoyancy of a parcel with respect to its environment where positive. This quantity represents available energy which can be converted into kinetic energy upon vertical ascent, per unit mass, and can therefore be translated into a theoretical maximum of vertical velocity to be reached, $w_{max} = \sqrt{2 \times CAPE}$.
- Level of neutral buoyancy (LNB), which is the highest level of neutral buoyancy found in a column, capped at 18 km.

Convective available potential energy is diagnosed after regridding all the available datasets to a common vertical grid with $dz = 100$ m and applying further interpolation for buoyancy integration. Before we simulate the parcel lifting, a layer of 500 m depth is mixed to derive parcel characteristics before lifting in all cases. Buoyancy is calculated using virtual temperature and the three following assumptions:

1. No dilution
2. A latent heat of condensation independent of temperature, thus neglecting the ice phase
3. Immediate removal of condensed liquid water

Other column diagnostics are derived on the native vertical grids, including precipitation rate, physics and dynamics tendencies in the mixed layer and free troposphere, as defined in Table 1. For these diagnostics, we assume that the atmosphere can be separated into a mixed layer (all model levels up to 500 m) and free troposphere, comprising all model levels above 500 m up to the tropopause, assumed at 16.500 m. Mass weighted bulk tendencies for each of the two layers are computed and utilised.

Finally, we investigate variation across the diurnal cycle, e.g., as a proxy for statistical uncertainties in mean and variance and non-stationarity caused by the diurnal cycle in the sampled PDFs. This only applies to parts of the analysis; in particular

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Table 1. List of model tendencies analysed.

Tendency	Unit	Short name
Total tendency of convective available potential energy	J/kg/6h	DCAPE
Precipitation rate	mm/6h	PREC
Temperature tendency from dynamics in the free troposphere	K/6h	-
Temperature tendency from dynamics in the mixed layer	K/6h	-
Specific humidity tendency from dynamics in the free troposphere	g/kg/6h	-
Specific humidity tendency from dynamics in the mixed layer	g/kg/6h	-
Temperature tendency from physics in the free troposphere	K/6h	-
Temperature tendency from physics in the mixed layer	K/6h	-
Specific humidity tendency from physics in the free troposphere	g/kg/6h	-
Specific humidity tendency from physics in the mixed layer	g/kg/6h	-

to inspect the CAPE and precipitation PDFs. Statistical linking with the principal component analysis is carried out on a column-by-column basis, which by design samples the full statistics within our domain.

3 Results

3.1 Distributions of key diagnostics

3.1.1 Distribution of precipitation rate

A very fundamental property of each weather and climate model is its precipitation rate distribution, which may include global and local biases (e.g. Roehrig et al., 2020). Figure 1 shows the PDF of precipitation rate of all output columns across four model physics suites and the benchmark, the convection-permitting ICON, when represented at a common grid. Millions of columns receive negligible precipitation accumulation (see also Groot et al., 2026b, for a discussion of trace precipitation). On our logarithmic axis, we find a rapidly decaying PDF for low intensities (less than 20 mm/6hrs) in ICON, which looks like an exponential decay on a log-axis. From about 20 mm/6hrs, precipitation rate decays in an approximately loglinear fashion, with virtually no event above 90-100mm/6hrs.

The precipitation rate is broadly similar for five configurations. That means no systematic bias between the convection-permitting benchmark and our physics suites appears, although this might be expected. For an extensive discussion of the joint precipitation PDFs of physics suites and our benchmark, including their conditional biases, we refer to Groot et al. (2026b). This work also suggests a coupling of precipitation to the dynamical forcing, which we also address further in the current work.

Besides the inter-model differences, we can also estimate variability of PDFs over the diurnal cycle from Figure 1. Structural differences between day and night, which most likely occur over land and in coastal areas, remain limited. The diurnal cycle

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

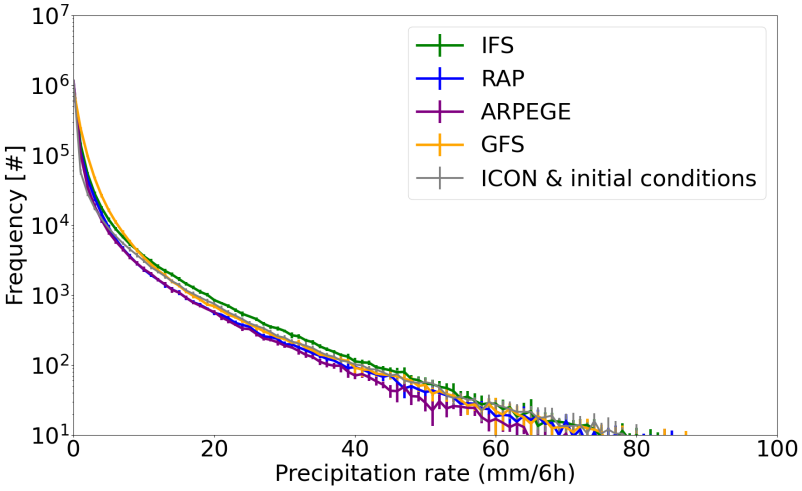

Figure 1. Logarithmic histograms of 6h precipitation rate for convection-paramterised IFS, RAP, ARPEGE, and GFS physics suites, alongside that of coarse-grained convection-permitting ICON in grey. The solid line represents the median and the error bars represent the standard deviation of six PDFs over the diurnal cycle, initialised from 00-15 UTC, every three hours.

shifts the PDFs slightly up and down and the difference between physics suites generally exceeds the variation across the diurnal cycle.

In summary, we find that the precipitation distributions in the MUMIP physics suites are broadly similar. However, whether the convective instability, which is a key driver of tropical convection, has a similar structure across our models is not yet clear. Thus, we now delve deeper into convective diagnostics by comparing CAPE and LNB across physics suites.

3.1.2 Distributions of CAPE and highest LNB

Convective clouds frequently bring precipitation over the Indian Ocean domain. It is well-known that conditional convective instability is an essential ingredient to kick-off the development of deep-convective clouds. We now look at the distributions of mixed-layer convective available potential energy (CAPE, Figure 2), which is an indicator of vertically integrated convective instability. The quantity can be linked to key differences between the profiles of temperature and humidity among suites, later in this Section. The vertical integral signified by CAPE represents the total amount of available potential energy when parcel of low-level air is forced upward until it freely convects up to a clearly stable layer. The parcel may move upward to approximately the tropopause if this is the first clearly stable layer. We have converted CAPE to their corresponding maximum vertical velocity to compress the x-axis of the PDFs. Furthermore, the highest LNB approximately represents where convective outflows and anvils spread out horizontally at the fastest rates.

Figure 2 shows the PDFs of CAPE and level of neutral buoyancy. We find that the benchmark ICON distribution (in grey) has a clear bi-modal distribution of positive CAPE, with a high peak at values below 20 m/s and another lower peak at values at

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

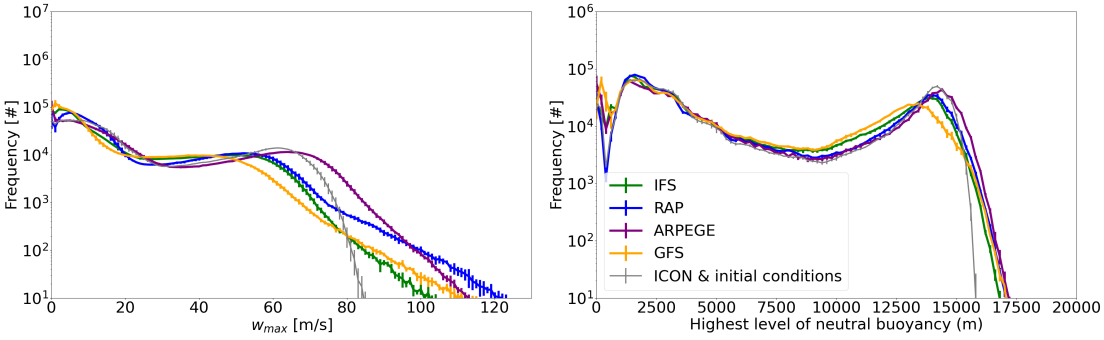

Figure 2. Logarithmic histograms of $\sqrt{2 \times CAPE}$ (where CAPE indicates mixed-layer convective available potential energy), and level of neutral buoyancy (LNB) in forecasts of four convection-parameterised physics suites represented by four colours. Grey lines represent the histograms of the initial conditions, which corresponds with the ICON benchmark distribution of the same variables (after coarse-graining). The solid line represents the median and the error bars represent the standard deviation of six PDFs over the diurnal cycle, initialised from 00-15 UTC, every three hours.

just over 60 m/s. Furthermore, the tail of the distribution of CAPE is rather compact, with a sharp gradient of the PDF at about 80 m/s. The median and standard deviation of six distributions plotted for ICON correspond to the different start times during the day — little variation in CAPE is seen across the diurnal cycle.

CAPE distributions of the four physics suites compare reasonably well to the ICON benchmark, especially at low to intermediate CAPE. At intermediate CAPE,a local minimum is hardly distinguishable between 20 and 40 m/s. If visually distinguishable at all, the secondary high CAPE peak shifts toward lower CAPEs in most physics suites, corresponding to ≈ 50 m/s on the X-axis, but not in ARPEGE forecasts. The shift is the most notable in GFS. At $w_{max} > 80m/s$ the ICON benchmark PDF of CAPE drops off sharply, whereas the other physics suites have an extended tail, and correspondingly, a much more gradually sloping PDF; beyond $w_{max} \approx 100 - 110m/s$ in most physics suites and even $w_{max} \approx 120m/s$ in RAP and ARPEGE.

The level of neutral buoyancy (LNB) shows a clear trimodal distribution, which is nearly indistinguishable between models at very first glance. Therefore, we first describe the shape of the PDF across all configurations. We could define a zeroth mode with stable stratification above the mixed layer. This stably stratified mode consists of the lowest two to three bins with an LNB at or below the mixed-layer top, at 500m (followed by a minimum at $500 < LNB < 1000$ m in the PDFs). For convectively unstable cases, the first mode consists of predominantly subtropical shallow convection and tropical shallow to mid-level convection. This first mode peaks between 850 and 500 hPa, i.e., 1.5-6km altitude. The second mode is a deep tropical mode. In the deep mode LNBs frequently exceeds 14 km and reaches 16-17km occasionally. The separation of the LNB distribution between sub-tropical and tropical bands is not shown explicitly here, but the clear fingerprint of a distinctive deep-convective mode and shallow-convective modes suggests that our domain should be divided into *at least* two dominant cloud and circulation regimes, where one represents the intertropical convergence zone and the other the subtropical latitudes.

Now we will look at the details of the PDFs of LNB and describe model discrepancies. The three modes are rather well cap-

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

tured by all four physics suites. The LNB distribution differs most substantially from that of the ICON benchmark above 10 km altitude, with an overestimation of the LNB frequency for IFS and GFS between 10 and 13 km. Furthermore, the upper tail extends distinctively further upward in all four physics suites than our benchmark, in particular in ARPEGE. In IFS and RAP we observe slight downward shift of the maximum and some erosion of the peak of the PDF, although this is less strong than in GFS.

On the very low end of the distribution, the different physics suites are also sorted in terms of LNB variability, indicating slightly different tendencies in low-level LNBs between the four physics suites. This is likely related to some biases between different suites in stratocumulus representation and that of the coinciding capping inversion. It manifests at levels of 500-1500 m, just above our mixed layer, which is associated with the treatment of the boundary layer and its interactions with the layer just above clouds and/or a capping inversion. It is beyond the scope of this work to investigate boundary layer processes and their interaction with overlying layers in very low CAPE regimes in all their detail. Instead, here, we will focus on describing deeper convective instability in the form of CAPE and on their relation with specific humidity and temperature tendencies.

To wrap up the LNB distributions, the combination of

- the extended upper tail
- downward shift of the deep-convective maximum and reduced sharpness of the maximum of the PDF
- less distinctive minimum on the low end of the distributions of LNB

overall increase the LNB spread in the SCM simulations with lead time, especially in the deep-convective regime (with modal LNB of 14 km). This may be caused by tendency transitions near stable layers that are overly smooth, partially mixing heat or water vapour across, or at, a boundary where tendencies should decay rapidly. This remains speculative.

We have found subtle multi-modal trends in the PDFs of CAPE and LNB in the MUMIP domain. Upper tails of large convective instability definitely grow massively, while on the contrary, the lower end of the PDF changes differently (i.e. towards decreasing instability), with many columns appearing in this decremental regime (Figure 2, i.e., the differences between physics suites and benchmark). Specifically, we often observe an overall decrease in CAPE within the second peak. Bi-directional contributions to change in CAPE (LNB) imply that these changes cannot be simplified to a single characteristic approximate response occurring everywhere: implicit spatiotemporal variability in the CAPE and LNB tendencies is prevalent.

Therefore, we next investigate the spatial distribution of CAPE and precipitation over our domain and its variation with lead time explicitly (next Subsection).

3.1.3 Subtropics vs. tropics

We have seen that the SCM simulations typically tend to forecast a reduction in the magnitude of CAPE in the high CAPE mode, when initiated with our benchmark and over our domain. However, this is generally not the case for ARPEGE, which appears to have much weaker typical CAPE tendencies. Furthermore, we have seen that the shape of the PDF of the diagnostics is not preserved: a single characteristic (linear) response could not explain the change in CAPE and LNB everywhere. Now,

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

we therefore investigate spatial variability in the diagnostics obtained with the ARPEGE and IFS physics suite, which are representative for all four physics suites and the ICON benchmark.

Figure 3 shows the spatial distribution of precipitation rates over domain in two models. IFS has a significantly larger area with heavy rain of 6-7 mm/6h than ARPEGE (consistently with Fig. 1). The location of the maximum coincides between the SCMs and is located on the eastern flank at about 5 degrees south. Another peak in precipitation occurs at 5 degrees south near 65 degrees east. This second maximum also coincides between the models, and again the amplitude is weaker in ARPEGE than in IFS. Furthermore, we find more subtle maxima right at the northern boundary of our domain.

Figure 4 shows spatial distribution of CAPE in IFS and ARPEGE. The maximum is situated in the same region where our

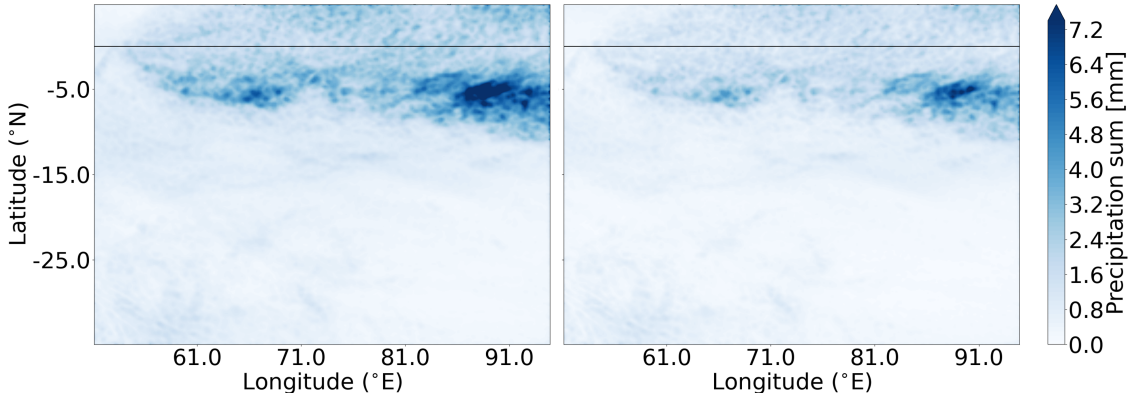

Figure 3. Mean precipitation sum of IFS (left) and ARPEGE (right) for each 6h bin over the validated part of the time series, i.e., 6 to 7 initialisations per day.

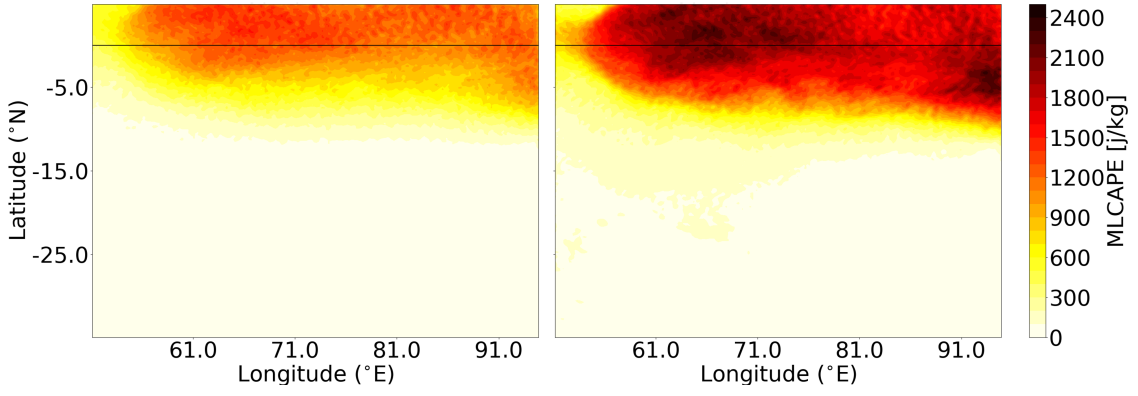

Figure 4. Mean CAPE of IFS (left) and ARPEGE (right) over each 6h bin over the validated part of the time series.

three precipitation maxima are found, generally within about 5 degrees of the equator. South of about 10-15 degrees south, the

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

CAPE is typically marginal, with time averaged values below a 100 J/kg, or locally up to 200 near 15 degrees south. There is some variation in mean CAPE across models near 15 degrees south, especially in the western part of the domain (Fig. 4).

Since the precipitation rate and convective instability is generally large near the equator, we define a 5 degrees band about the equator as the equatorial band, although precipitation and CAPE in the westernmost part is reduced somewhat in all physics suites (e.g., Fig. 3). A strong gradient between about 5 and 15 degrees south separates this inter-tropical convergence zone (ITCZ) from a subtropical band with considerably lower monthly precipitation and CAPE. We define this southern half of the domain, south of 15 degrees south, as subtropical region.

In summary, we can identify the two climate regions dominating our MUMIP domain and we can split up further statistical analysis accordingly. Next, we focus on temporal dependence of CAPE and precipitation rate, which is dominated by lead time dependence.

3.2 Lead time dependency: spin-up?

Figure 5a shows the evolution of the precipitation accumulation as a function of lead time. Both IFS and GFS exhibit a strongly non-linear accumulation of precipitation, which indicates spin-up behaviour in physics suites. Initially, precipitation accumulates rapidly in both models, before the precipitation rate reduces towards the end of the simulation. This pattern is clearly visible in both interquartile range and the mean.

On the contrary, spin-up has been removed from the ICON benchmark and therefore, lead-time dependence should be small, especially after removal of seasonal and diurnal cycles. In the native model climate, each model should have a relatively constant precipitation rate. Note that the curves of especially GFS, IFS and RAP remain relatively parallel beyond about two hours lead time.

Furthermore, the mean precipitation rate exceeds the upper quartile range in most models, which signifies that precipitation accumulation of our domain predominantly occurs in a small number of columns over a narrow region: the equatorial band. However, for GFS, the opposite is true: the upper quartile exceeds the mean precipitation accumulation and this is consistent with the PDFs of Figure 1. Lastly, we can quantify the bias between GFS and ARPEGE precipitation rates more accurately than before (Fig. 1), whereby the first exceeds the latter by about 100%, although this difference seems to be caused mostly in the first hour.

Figure 5b shows the evolution of CAPE within the equatorial band as a function of lead time in each physics suite. Consistently with spin-up patterns in precipitation rate, we find a large net CAPE tendency in the first interval for GFS and IFS in particular, which represents changes over 30 minutes (IFS) or the first hour (other models). On the contrary, the ICON benchmark tendencies would be negligible. The benchmark simulation has been spun-up and should represent its native climate closely, with near-zero mean net tendencies over the entire month.

After a rapid decline in CAPE, the IFS CAPE is nearly stationary. On the contrary, CAPE recovers gradually in ARPEGE and slightly exceeds the initial mean CAPE. Furthermore, the upper quartile range of ARPEGE experiences a negligible decrease in the first hour, which is comparable in RAP. However, on the contrary, the lower quartile of the ARPEGE CAPE decreases as much as in IFS over the first hour, while it recovers later. This indicates that the spatial variance in ARPEGE CAPE increases

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

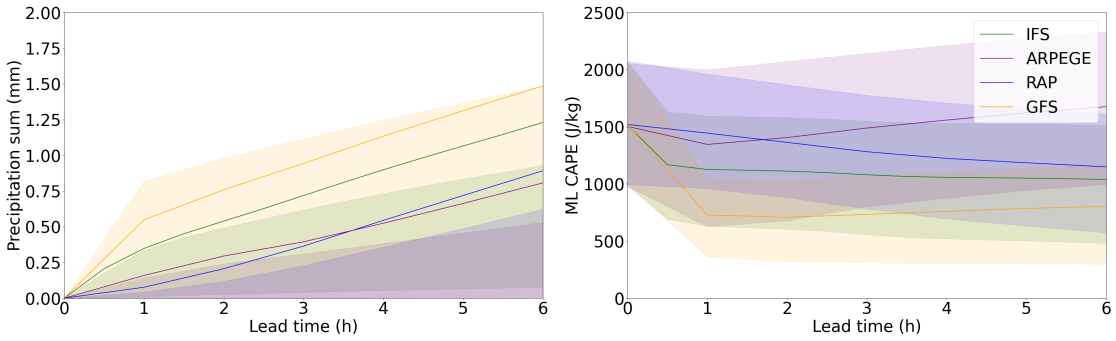

Figure 5. Spin-up pattern of our diagnostics. Left: Cumulative precipitation as a function of lead time for our physics suites. Right: CAPE in the equatorial band as a function of lead time for our physics suites. In each panel the solid lines represent the mean and the shading represents the inter-quartile range.

with time, whereas such a pattern is not obvious in IFS. IFS has an interquartile range that broadly follows the mean.

The subtropical band has negligible CAPE (not shown). While this CAPE is also dependent on lead time, this dependency looks broadly similar to that of the equatorial band. Furthermore, the bulk of columns has a negligible amount of CAPE (below 100 J/kg) and, hence, we refrain from in-depth analysis. We analyse the 31 day evolution of precipitation rate, LNB, CAPE of all simulations in Appendix A.

In summary, our diagnostics depend strongly on lead time, with a substantial imprint on key convective diagnostics, such as ITCZ precipitation and ITCZ CAPE. A substantial portion of the initial CAPE is destroyed in the first hour of our simulations.

We need to understand how exactly such spin-up in the SCMs affects our dataset. Underlying specific humidity and temperature tendencies in the mixed layer and convective environment of the free troposphere must cause this CAPE decrement, but incidental increment. We start by assessing the correlations between the spatial mean tendencies in the next Subsection.

3.3 Relating diagnostics: are physics tendencies and precipitation potential drivers of CAPE change?

The change rate of CAPE may directly inform convection parameterisations for convective mass flux calculations in certain closures, although no uniform formulation of the quasi-equilibrium assumption exists (Arakawa and Schubert, 1974; Yano and Plant, 2012). Therefore, we examine the change rate of CAPE in our comparison of physics suites. The closure of the convective mass flux is particularly important at low latitudes because the ITCZ is located there - our equatorial band mostly overlaps with the ITCZ. Reduction of CAPE is presumably related to precipitation rate within model grid cells, as precipitation reduces convective instability. Even when non-native model states, such as assimilated states or in this study states from another model, are inserted into any model, the relationship presumably holds in the area mean (Buschow, 2024). However, mixed-layer drying, which reduces water vapour content, and cooling as well as upper-level warming by model physics may modify CAPE significantly. This is particularly reasonable during spin-up: physics suites adjust their state to their native model climate, which is likely associated with individually characteristic local temperature and humidity biases between suites.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

In recent versions of IFS, more sophisticated assumptions than the quasi-equilibrium are used (Bechtold et al., 2014b), following Donner and Phillips (2003). In this scheme, the boundary-layer tendencies play an important role in the closure of CAPE, for instance through explicitly accounting for terms like moisture convergence. Non-equilibrium convection is represented better in such schemes (Bechtold et al., 2014b). In addition, the IFS scheme allows for more flexibility in the source layer of convective overturning. Ultimately, both schemes will represent CAPE tendencies, but the mass flux closure differs considerably. Hence, the relation between CAPE change rate and variables like precipitation could vary among the our physics suites.

Nevertheless, we would expect to find some correlation when calculating the regression between the regional average DCAPE (see Table 1) and precipitation rate (Buschow, 2024). DCAPE represents a physically meaningful quantity: the potential energy gain or loss associated with convective overturning, which may be achieved by column water vapour and temperature tendencies. The correlation patterns should be expected within each physics suite and across the suites, because all of the physics suites start from virtually identical initial conditions.

Figure 6 (left) shows the relation between the area-mean precipitation rate in the equatorial band and DCAPE. Each physics suite is represented by a different symbol and colouring follows the diurnal cycle. Little covariance is found between the two variables if we consider variability across physics suites or within each suite; ARPEGE has some covariance between the two variables and uses a DCAPE closure for convection [Christensen et al., in preparation]. However, any relationship fits poorly within the equatorial band.

The right panel of Figure 6 shows the relation between the area mean mixed-layer humidity tendency and the DCAPE. The linear correlation across all SCMs as well as within SCMs is obvious. Furthermore, a residual association of CAPE tendencies with the diurnal cycle is also clear in some physics suites, but not in others. Residuals are particularly well sorted with respect to the mean fit for GFS, which also has the highest correlation between the two variables among all suites. It suggests that area-mean drying of the mixed layer is a dominant mechanism to cause CAPE tendencies, and its variation among suites, within the equatorial band.

Moving to the subtropics (Figure 7), we find a generally stronger association between DCAPE and precipitation rate, although this varies strongly among physics suites: correlations for ARPEGE are 0, whereas IFS shows a very strong correlation. To some extent the residuals are also sorted according to the diurnal cycle. As opposed to Figure 6, the subtropical statistics, which are however only weakly affected by convection (Figure 3), indicate more support for the area mean relationship suggested by Buschow (2024) (this study has assessed a much wider region at low latitudes). If we look at individual model tendencies, the free-tropospheric cooling is well-correlated with DCAPE in the subtropics. Further stratification of our dataset follows the diurnal cycle as indicated by colours in Figure 7 (right panel). Typical mean tropospheric warming or cooling rates are only weak, with values on the order of 0.2K/6hrs, but any tendency can critically affect CAPE. This holds especially in regions with marginal CAPE. Furthermore, one would expect the free-tropospheric mean tendencies to correlate with precipitation rate (as discussed later, in Section 3.5).

From Figures 6 and 7 we can derive clear geographical variation of tendencies and their effects. The main tendency that contributes to CAPE tendencies is consistent with the shape of the Hadley cell: moisture accumulates near the ITCZ, building up

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

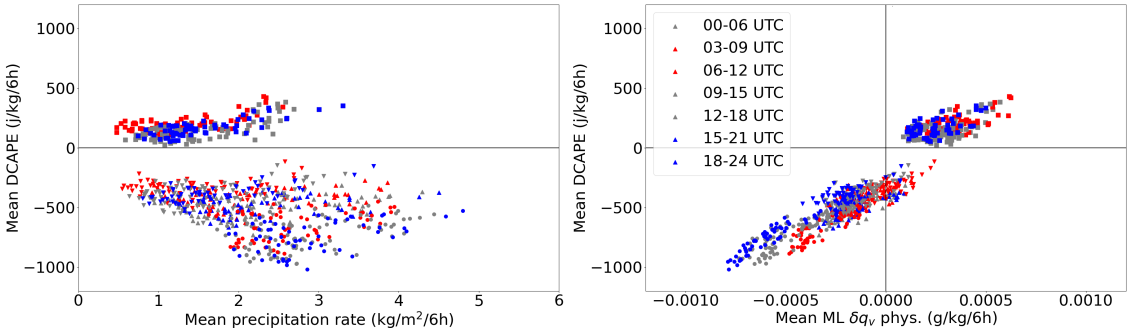

Figure 6. Equatorial mean DCAPE versus (left) mean total precipitation rate and (right) mean mixed-layer physics tendency of specific humidity in four models. Colouring indicates the variation over the diurnal cycle. Squares: ARPEGE; circles: GFS; upward pointing triangle: IFS; downward pointing triangle: RAP.

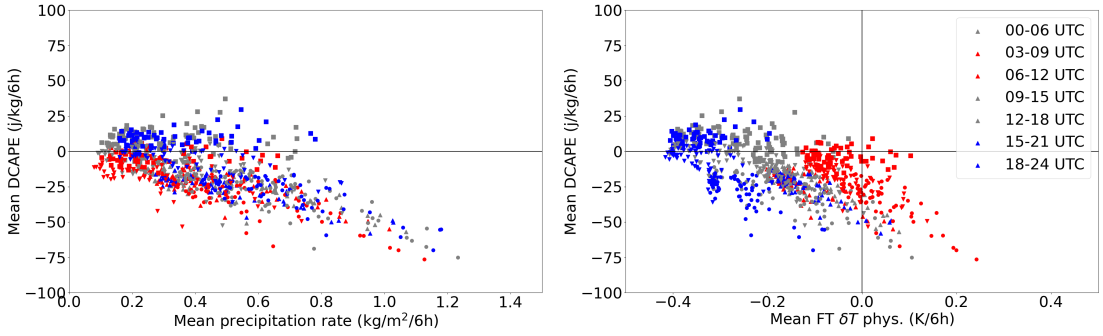

Figure 7. Subtropical mean DCAPE versus (left) mean total precipitation rate and (right) mean free-tropospheric physics tendency of temperature in four physics suites. Colouring indicates the variation over the diurnal cycle. Squares: ARPEGE; circles: GFS; upward triangle: IFS; downward triangle: RAP. δ: tendency of variable.

CAPE maxima which are strongly dependent on the exact humidity, but the CAPE is mostly absent in the subtropics, where subsidence and cooling of the free-troposphere dominate and no significant water vapour is added to the atmosphere at upper levels (e.g. Lang et al., 2023, 2021; Soden and Bretherton, 1994). Furthermore, we can conclude that other physics tendencies than mixed-layer drying and free-tropospheric temperature change must also weakly contribute to CAPE changes in our domain. This is because, following from an estimated fit by eye, zero-tendencies in Figures 6 and 7 (right panels) do not coincide with an expected mean DCAPE of 0. While weak statistical signals are found for the mixed-layer temperature tendency from physics, we find them comparatively unimportant and do not show them explicitly.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

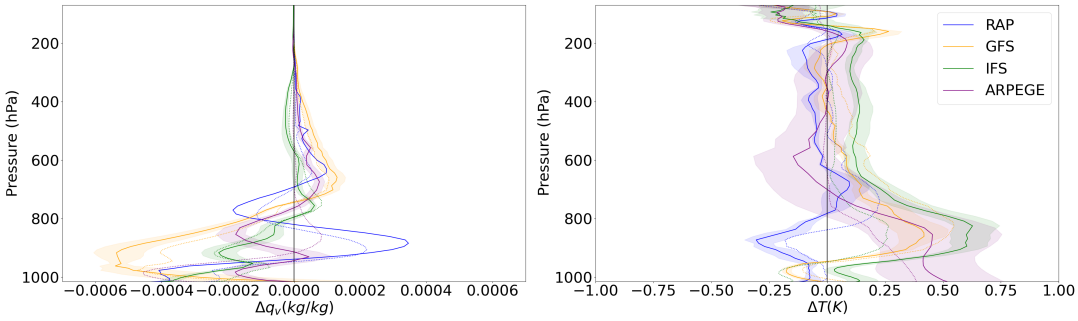

Figure 8. Profile of net change Δ of q_v (left) and temperature (right) profiles for each physics suite (up to 70 hPa). The solid lines represent the 6hrs forecast and the dotted line the 1hr forecast. The envelope of variation of the mean adjustment profile over the diurnal cycle is indicated by light shading, which covers 6 out of 8 initialisations per day.

3.4 Bias and adjustment

A clear relation between mean CAPE tendencies and mean specific humidity tendencies in the mixed layer from the physics has been established across the physics suites. In addition, only a weak connection to precipitation variability has been established, or even none at all. In the following, we first analyse the mean profiles of specific humidity and temperature change after 1hr and at the end of the simulations. This is important because they modulate both the dry-adiabat at low levels and the moist adiabat higher up, each of which project onto our convective diagnostics. The variation in net changes must be dominated by physics differences, because our experiment imposes fixed dynamics by design.

3.4.1 Specific humidity

First, we quantitatively analyse the mean q_v change over the 6 hour window. Starting from near the surface, each of the physics suites in Figure 8 (left) has a net sink of humidity at low levels, which is in the mixed layer (below $\approx$ 970 hPa). The water vapour is extracted from a layer which varies in depth across the suites: in ARPEGE and RAP, the layer is less than 100 hPa deep. In IFS, the thickness over which we find a humidity sink is deeper and in GFS it is the deepest: about 300 hPa, spanning from near the surface up to 700 hPa. On the contrary, we find a net positive change of humidity in RAP between 900 and 800 hPa. Similarly, a much thinner layer with a net humidity gain is found in ARPEGE near 900 hPa. Finally, above 650 hPa, we typically find a slight humidity gain, although IFS has a small humidity sink further up. However, the magnitude of mid-to-upper-tropospheric humidity tendencies is naturally limited by the saturation.

If we look at the temporal evolution of the sink in the mixed layer, it evolves nearly instantaneously in GFS: little difference is found between the profiles at 1hr and 6hrs lead time. Similarly, for IFS, both profiles of humidity adjustment look quite comparable. However, the mixed-layer is still drying between 1hr and 6hrs lead time, especially at the lowest levels. Nevertheless, the rapid adjustment of IFS and GFS is consistent with the evolution of CAPE and precipitation accumulation as a function of

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

lead time in Figure 5. On the contrary, the adjustment profiles of RAP and ARPEGE differ significantly between 1hr and 6hrs lead time. At the lowest levels, the mean humidity adjustments occur very slowly in RAP, although it is rapid near 900 hPa. However, in ARPEGE, the mean mixed-layer drying initially undershoots the drying at 6hrs lead time. In addition, the shape of the adjustment profile varies substantially with lead time.

The adjustment profiles remain very similar when sub-setting for the equatorial band, but generally, the amplitude increases a little bit (not shown). The vertical structure of net change shows similar structures as for the overall mean. Furthermore, generally, the amplitude of net change increases somewhat with increasing initial humidity (not shown). This is expected, such that the relative change between initial time and 6hrs lead time is comparable to Figure 8. Therefore, using the mean profile of specific humidity, absolute mean humidity tendencies can approximately be translated to relative changes.

In summary, significant drying in the lower troposphere occurs rapidly in GFS and IFS, with a strong adjustment in the mixed layer and a decay of this adjustment pattern between 900 and 700 hPa. Instead, RAP and ARPEGE show multi-modal vertical adjustment with contrasting behaviour between the levels of shallow clouds (about 850 hPa) and the lowermost mixed-layer ($\approx$ 970 hPa). The latter dries in all simulations. Furthermore, the evolution of net tendencies is complicated in ARPEGE and slower in RAP compared to IFS and GFS. This adjustment pattern may largely govern the evolution of CAPE over the first hour in Figure 5 (left).

3.4.2 Temperature

Moving on to the temperature adjustment of Figure 8 (right) we find weak net changes above 600 hPa. However, the mixed layer warms by about 0.5K in ARPEGE, with a considerable variation over the diurnal cycle. The warming extends upward to about 800 hPa. In combination with the humidity adjustment, which strongly varies with lead time for ARPEGE (left panel), the warming could explain the CAPE increase in this suite (Figure 5, left). Both mean humidity and mean temperature of the mixed layer increase beyond 1hr lead time in this suite. In combination with weak temperature tendencies, which are predominantly negative further up in the free troposphere, the mixed-layer signal implies CAPE must increase.

On the contrary, GFS and IFS indicate a strong warming of similar magnitude above the mixed layer, between 900 and 800 hPa. This may enhance stable stratification, and it could even create an inversion under specific conditions. Overall, the convective stability will increase near 850 hPa. This further favours a decrease of CAPE. Finally, temperature adjustment in RAP is much weaker than in the other physics suites. If attributable to any specific process, it suggests the fingerprint of RAP CAPE is more likely dominated by the mixed-layer drying.

3.4.3 Synthesis

The differences in adjustment profiles indicate that the boundary layer physics and stratocumulus and shallow convection are likely represented differently across physics suites. Shallow cumulus and stratocumulus would mix the lowest 2-3 km of the atmosphere effectively, and hence, should increase the humidity near 800-850 hPa and cool this layer. Differences across models suggest potential over-representation in RAP compared to the benchmark, but under-representation in GFS - and less so - IFS. These patterns will affect convective instability metrics like CAPE. The exact interpretation remains very speculative for

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

now and goes mostly beyond the scope of this work. For an extensive discussion regarding a rationale and consistent findings, we refer to Bechtold et al. (2014a) and references therein, who present a previous version of IFS, and the more recent cycle 47 (see Forbes et al., 2021; Bechtold et al., 2020).

We have found a noticable spin-up of CAPE and precipitation. Furthermore, we have identified that specific humidity and, to a lesser extent, temperature tendencies seem to directly explain area mean change rate of CAPE. In the following, we are going to inspect the EOFs of physics and dynamics tendencies in the mixed-layer and free-troposphere, and their association with CAPE and precipitation variability. From here, we focus our column-by-column comparison in the equatorial band, where precipitation and CAPE depend strongly on lead time (Figure 5). We can quantify to which extent the profile adjustments, which cause the profiles of Figure 8, imprint on CAPE and precipitation evolution. Subsequently, we revisit the suggested relationship in Figure 6 (right). Furthermore, the EOFs objectively quantify the similarities of tendencies across physics suites, in response to a common forcing, providing an opportunity to estimate the importance of dominant variance modes . Our approach is comparable with Lambert et al. (2020), but retains two layers instead of detailing vertical profiles. Finally, we overcome the limitation of comparing area means.

3.5 Empirical Orthogonal Function analysis: physics relations and spin-up

We apply the classical empirical orthogonal function (EOF) analysis (Lorenz, 1956) and investigate a dataset of SCM tendencies (see table 1). Thereby, we normalise variables with respect to their standard deviation. Furthermore, we assume the following:

- the long-term mean of the sum of physics and dynamics tendencies should go to zero in a stationary climate simulation with (near-)stationary forcing
- the long-term mean of DCAPE should go to zero under (near-)stationary forcing
- any net physics tendency is expected to be compensated by the opposite dynamical tendencies. Theoretically, this is valid in the long run and any model in its native climate should closely follow this assumption on much shorter time scales.

Therefore, we expect certain EOFs of tendencies in our approach to project onto near-zero CAPE change rate and we expect nearly perfect compensation between physics and dynamics tendencies for EOFs that behave stably in our SCMs (i.e., $\Delta X_{physics} \approx -\Delta X_{dynamics}$). Accordingly, the tendencies would statistically cancel out in the long run. On the contrary, EOFs that project strongly onto CAPE change, or with a strong deviant amplitude between physics and dynamics, are suggested to reflect spin-up patterns, which drive initial adjustment, such as shown in Figures 5 and 8. Furthermore, some associated spin-up variability might be projected onto GFS and IFS precipitation, since we find precipitation rate is lead-time dependent (Figure 5). In addition, dipoles of free-tropospheric and mixed-layer tendency pairs indicate vertical exchange processes like convective mixing. In such cases, the dominant balance of physics and dynamics tendencies might involve four tendencies rather than two.

Finally, we may hypothesise that a misbalance between physics and dynamics in the low-level humidity tendencies could cause the CAPE and precipitation evolution of Figure 1.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

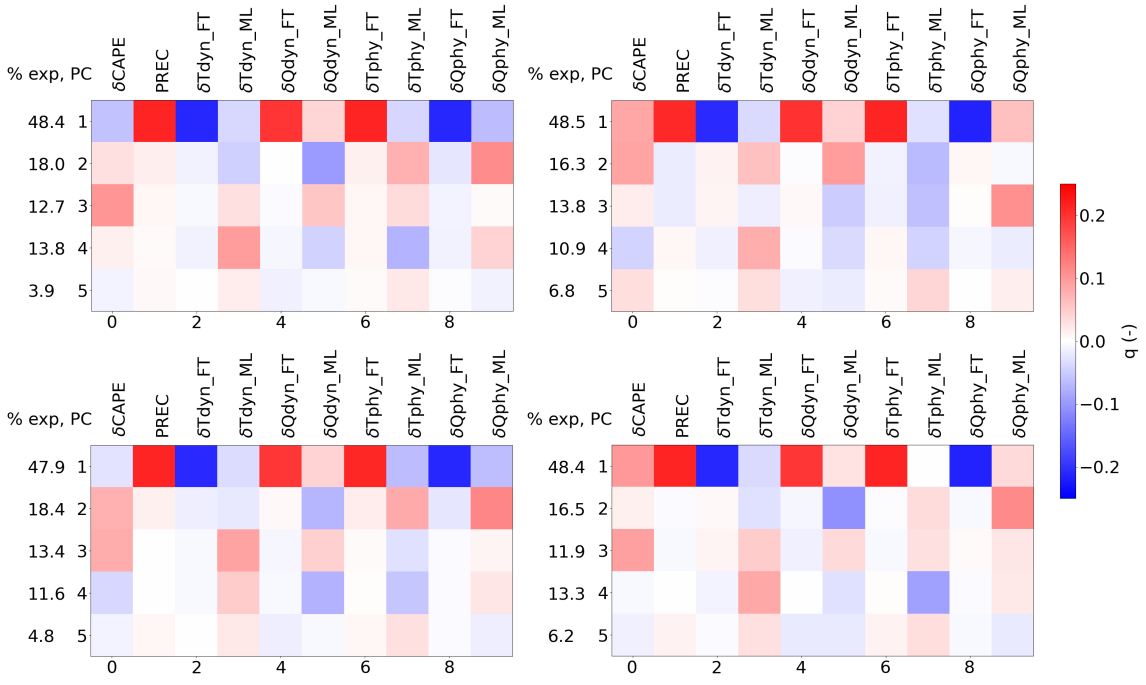

Figure 9. Empirical orthogonal functions of column-by-column tendency dataset in the equatorial band. The EOF amplitude is multiplied by the corresponding normalised explained variance to obtain q, whereby variability in dominant EOFs is emphasised implicitly. Left top: IFS; Right top: ARPEGE; Left bottom: GFS; Right bottom: RAP. PC: Principal component, with no.; Exp: explained; Q: specific humidity; T: temperature; phy: physics tendency; dyn: dynamics tendency; ML: mixed layer; FT: free troposphere; δ: tendency of variable.

The first five EOFs of each physics suite are visualised in Figure 9. Together, these EOFs represent 96-97% of the total variability in the dataset of each suite, which means that the least important EOF in this visualisation (EOF5 of IFS, left top) represents about as much variability as do the ranks 6 to 10 combined for each physics suite.

3.5.1 EOF1

The first EOF is associated with 48% of the variability in each physics suite. Being dominated by precipitation, which is

- strongly anti-correlated with physics humidity tendencies of the free troposphere (hence, as expected, strongly correlated with dynamics humidity tendencies)
- strongly anti-correlated with dynamics temperature tendencies of the free troposphere (hence, as expected, strongly correlated with physics temperature tendencies)

this EOF represents the main physics of precipitation variability. It is nearly uncorrelated with DCAPE in GFS. In other models, the association with DCAPE is of changeable sign and slightly stronger, but still very weak compared to the leading order

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

variables. Furthermore, a balanced compensation between physics and dynamics tendencies is of leading-order importance. The combination of balanced compensation and weak DCAPE signal in this EOF indicates it manifests little, if any, spin-up tendencies. This EOF is apparently also emerging strongly in Groot et al. (2026b), where it is effectively demonstrated that our SCM tendencies in the free troposphere associate strongly with precipitation rate beyond 3hrs lead time.

Upon very close numerical inspection of the exact amplitude of free-tropospheric humidity physics tendency in EOF1, slight overcompensation of the corresponding dynamics tendency appears: the physics tendencies and precipitation have roughly 5-10% stronger amplitude than the corresponding dynamics tendencies across SCMs in this EOF. However, upon assessing EOF1 derived for 1hr lead time, this offset across SCMs does not appear to be correlated with the higher initial precipitation rates in GFS and IFS (Figure 5), although one may expect so. This could suggest that the precipitation signal causing spin-up in GFS and IFS is not coupled the usual way to the prescribed dynamics, but instead a removal of some other sort of excess humidity, although this remains speculative.

Finally, each model shows a rather weak association of EOF1 with mixed-layer tendencies.

From the experimental perspective, EOF1 is effectively derived from prescribed dynamics of the free troposphere and has strong commonality between all the models. Lower order EOFs associate only very weakly with tendencies in the free troposphere.

3.5.2 EOFs 2-4

EOFs 2-4 represent 11 to 18 percent of the total variability in the dataset. For each EOF we effectively find substantial amplitude in mixed-layer tendencies and DCAPE only. Correspondingly, their assocaition with precipitation rate is negligible. In three out of four SCMs, namely IFS, GFS and RAP, EOF2 is dominated by humidity tendencies from the physics. However, in ARPEGE, this variable dominates EOF3. It is very plausible that variability of EOF3 of ARP mostly overlaps with EOF2 in other models. This is a plausible consequence of EOF analysis, where EOFs are often sorted by the magnitude of their explained variance. Furthermore, EOF2 (EOF3 in ARPEGE) still associates rather weakly with DCAPE in three out of four models. In GFS, however, mixed-layer drying from the physics projects considerably onto CAPE destruction, which is consistent with Figures 5, 6 and 8.

If we look at EOF2 (EOF3 in ARPEGE) and assess the balance between physics and negative dynamics tendencies, such balance appears to hold very well for IFS and RAP. Instead, mixed-layer temperature and humidity tendencies of the dynamics appear to be significantly weaker in ARPEGE and GFS than their physics counterparts. Hence, it appears as if the physics is overcompensating the dynamics tendencies in this EOF (no. 2 for GFS, 3 for ARPEGE). Despite overcompensation, ARPEGE does indicate a strong effect on DCAPE, which presumably underlies vertical shifts of the tendencies between physics suites and their measure of deviation from balance.

EOF3 (EOF2 in ARPEGE) has a pronounced association with DCAPE across all suites. The DCAPE has dominant positive correlation with dynamics tendencies in the mixed layer within this EOF, which is dominated by humidity (IFS, ARPEGE), temperature (GFS) or neither of the two (RAP). Interestingly, its dominance by humidity in IFS and ARPEGE (EOF2) is not balanced with compensating physics tendencies - their amplitude is negligible. Furthermore, IFS shows a weak relation be-

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

tween forced dynamical warming and warming by the physics. On the contrary, GFS indicates that both dynamical temperature and humidity tendencies are correctly compensated, albeit by weaker physics tendencies than their dynamics counterparts. Finally, in RAP, we find a pattern very similar to IFS.

EOF3 (EOF2 in ARPEGE) has an imprint on CAPE tendencies across all physics suites, but no significant relation with precipitation.

In EOF4, which is common between all SCMs, we find anti-correlation between dynamics temperature and humidity tendencies in the mixed layer. This contrasts with EOFs 2-3 for each model, although EOF2 in GFS has very little association with temperature tendencies. The physics and dynamics tendencies appear to balance relatively well in this component, except for physics tendencies of humidity in ARPEGE, with too low amplitude. This results in weak association with DCAPE for most models, although less so for ARPEGE.

EOFs 5 (and further) represent less than 7 to 11% of the variance. While EOF5 is interesting for its representation of connectivity between free-tropospheric and mixed-layer tendencies, it explains 4 to 7% of the variance, which is about half to a third of EOFs 2-4. Hence, we refrain from analysing EOFs 5-10 in detail.

Summarised, EOF3 (EOF2 in ARPEGE) is the dominant mixed-layer component that associates with DCAPE variability in three out of four models. In GFS, EOF2 and EOF3 jointly dominate the DCAPE variability. Generally, too much drying by the physics (EOF2, GFS) or insufficient moistening by physics tendencies (EOF2 of ARPEGE and EOF3 of IFS) in the mixed layer causes considerable CAPE changes. Similarly, off-balance cooling or warming by both physics and dynamics may lead to CAPE changes (EOF3 in RAP, IFS). Neither humidity adjustment, which may dominate low levels in Figure 8 (left, especially GFS), nor temperature adjustment (right, ARPEGE) controls DCAPE in isolation from other tendencies.

While CAPE changes cannot be attributed solely to mixed-layer drying by the physics, the CAPE balance involves tendencies restricted to the mixed layer. Furthermore, the mechanism behind lead time dependence of precipitation rate in GFS and IFS remains unclear. Finally, the free-tropospheric tendencies form a single EOF. However, three comparable orthogonal components blend into net mixed-layer tendencies.

4 Implications

4.1 Summary

The choices made in physics suites and, more broadly, parameterised model physics have a profound effect on convective diagnostics in our experiment. The fingerprints of four physics suites are compared against a benchmark - convection-permitting ICON - over the Indian Ocean. The precipitation PDFs are very similar across the physics suites. Non-stationary evolution of CAPE and precipitation rate with lead time is found in IFS and GFS, as well as ARPEGE (CAPE). RAP does not have a clear non-linear CAPE evolution with lead time. On average, we find a reduction of CAPE in the moderate-to-high part of the CAPE PDF, which is associated with the ITCZ. Nevertheless, ARPEGE recovers this CAPE, after an initial decrease. Furthermore, extreme CAPE values are also more probable for each physics suite than in the ICON benchmark. Similarly, the level of neutral buoyancy (LNB), an estimate for the level of maximum outflow rate of convective anvils, tends to decrease in the ITCZ, where

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

values of 14km are typical. Conversely, as for CAPE, the frequency of extremes increases in all physics suites. Furthermore, as for CAPE, ARPEGE shows no clear shift of the secondary peak of deep-convective LNBs, or perhaps a slight upward shift.

The area mean CAPE change in the ITCZ is very well correlated with area-mean mixed-layer drying, despite a weak correlation with precipitation rate. This holds across all physics suites and stands out especially for GFS, which has the most profound drying. Furthermore, the diurnal cycle modulates this correlation patterns, as do other tendencies (Figure 7). At subtropical latitudes, the free-tropospheric temperature tendencies correlate with the mean CAPE tendencies, but they associate weakly with the mean CAPE tendencies there.

The mean adjustment profile of humidity indicates that mixed-layer drying occurs in all suites (Figure 8). Substantial and rapid drying occurs in IFS and GFS from the surface up to almost 700 hPa. On the contrary, the drying evolves gradually in RAP at a similar rate, albeit over a shallower layer. Finally, ARPEGE initially dries, but then re-moistens. In addition, the layer above (900 hPa) moistens in ARPEGE and RAP.

Furthermore, ARPEGE has distinctive warming from the surface up to 800 hPa. While other models reveal comparatively weak adjustment directly above the surface (± 0.15 K), warming by 0.5-0.6 K is found for IFS and GFS at about 850 hPa, and cooling by about 0.3K is found in RAP at similar levels.

In a column-by-column EOF analysis of tendencies, change rate of CAPE and precipitation, we find that precipitation is strongly associated with free-tropospheric humidity and temperature tendencies, which explains about 48% of the variance in each dataset. These tendencies correspond strongly with anti-correlation between physics and dynamics. On the contrary, EOFs 2 to 4 project strongly on the mixed-layer tendencies, some of which are less balanced than EOF1. The second EOF in most models (third in ARPEGE) reveals the strongest signal in mixed-layer drying. This is associated with negative CAPE change in GFS, but less so in other models. The third EOF (second in ARPEGE) shows a strong association with CAPE change across the physics suites. The mixed-layer physics humidity tendency under-compensates the dynamics tendencies, hence this reduces the CAPE for the typical situation of net low-level drying. However, temperature tendencies in the mixed-layer also appear distinctively out-of-balance, especially in IFS and RAP. This indicates that both drying and warming with an origin in the mixed layer lead to CAPE destruction; temperature and humidity adjustment reinforce each other. Finally, a fourth mixed-layer EOF projects less strongly on CAPE change rate, which therefore uniquely represents a mode of anti-correlation between dynamics tendencies of temperature and humidity in the mixed-layer.

The remaining EOFs 5, and lower, are an order of magnitude less important: they explain less than 7% of the variance each, or between 7 and 11% for each SCM when combined.

4.2 Contribution of adjustment profiles to in changes in convective diagnostics

As indicated by the EOF analysis (Figure 9), most changes in the CAPE are realised with net tendencies in the mixed layer. Correspondingly, cooling and warming tendencies in the mid-to-upper troposphere remain on average small, within ± 0.15K (Figure 8, right) and vary by model. Within the mixed-layer, both humidity and temperature tendencies off-balanced between dynamics and physics cause CAPE change. We can separate this mixed-layer contribution into the following terms, *under our buoyancy assumptions* (defined in Sect. 2):

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

- pre-condensation humidity changes
- pre-condensation temperature changes, or more specifically virtual temperature changes
- corresponding net effect on the potential moist adiabat

Specific humidity in the mixed layer changes by up to of 0.5 g/kg in GFS, which translates to roughly 3% relative change. For IFS and RAP this is effect a little smaller ($\approx 2\%$) and for ARPEGE this adjustment is almost erased between 1hr and 6hrs lead time (Figure 8). These tendencies increase the lifting condensation level, which, *under our buoyancy assumptions*, largely control CAPE and LNB (via the moist adiabat). The moist adiabat thereby cools, which reduces the buoyancy above condensation level, hence CAPE and highest level of neutral buoyancy. The lifting condensation level height difference of GFS translates to about 50-60m in the mean (using McDonald, 1963; Lawrence, 2005; Romps, 2017) if assuming our domain is moist in our equatorial band. If we consider pure humidity adjustment and ignore any temperature adjustment, the specific humidity adjustment, in combination with a difference between dry and moist adiabatic lapse rate on the order of 4-5 K per km, leads to an average cooling of the moist adiabat by about 0.3 K. This cooling affects buoyancy tendencies considerably. Furthermore, the lifting condensation level change also reduces the probability of cloud cover.

However, to assess the total effects of profile adjustment, we need to include the virtual temperature adjustment. It will be dominated by temperature adjustments. The humidity adjustment, now more specifically water vapor mixing ratio adjustment, has a comparatively negligible effect on the virtual potential temperature. An average ≈ 0.5 K warming in isolation, corresponding to ARPEGE mixed-layer adjustment in Figure 8 (right), implies virtual temperature changes of 0.5 K in the mixed layer. This translates to rises of the lifting condensation level of about 60m (using McDonald, 1963; Lawrence, 2005; Romps, 2017). Correspondingly, the moist adiabat will be slightly warmer (less than 0.5K) and, hence, mixed-layer air becomes more buoyant, thereby increasing CAPE. In ARPEGE a slight drying of 1%-2% in terms of relative humidity occurs initially, which is smaller than for GFS. However, the humidity recovers and, hence, CAPE recovers after the initial decrease. The upper-troposphere undergoes negligible temperature adjustment, indicating that a recovery of the CAPE (Figure 5, right, after $\approx$ 4hrs lead time) is caused by a good balance between mean mixed-layer warming and mixed-layer drying, which would approximately retain the moist adiabat.

However, in other physics suites, mixed-layer temperature tendencies are much smaller than in ARPEGE and the mixed-layer mean temperature tendencies has a negligible impact on CAPE.

Altogether, following the mean adjustment profile dominates CAPE changes in RAP, IFS and GFS. In ARPEGE, mean temperature and humidity adjustment balances. Nevertheless, the lifting condensation level is expected to increase in all physics suites following the mean adjustment profile, with the largest tendency in GFS ($\approx$ 60m). Furthermore, the mean moist adiabat of parcels within the equatorial band is expected to decrease in all models but ARPEGE (most strongly in GFS), also decreasing the level of neutral buoyancy under a near-stationary free troposphere. This is consistent with Figures 2.

In Section 3.4 we have argued that the mean adjustment scales well with most samples and conditional spread among samples is expected to be small. Accordingly, we can argue that the effects of mean adjustment of specific humidity and temperature represent their typical imprint on lifting condensation level and level of neutral buoyancy properly. In addition, we can prop-

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

erly assess the impact of spread by investigating the changes of the tails of the distribution, which represent rare behaviour (or, alternatively: by quantifying the combined multivariate impact of each of the three mixed-layer EOFs of Figure 9). This tail of the CAPE and LNB indicate increases across all physics suites. Here, temperature tendencies may be more impactful, and like in ARPEGE, dominate the mixed-layer adjustment, as well as possibly certain low-level layers within the free troposphere. Under rapid heating or moistening conditions, for instance, a capping inversion might build up given profiles of GFS and IFS of Figure 8 (right), or spatial organisation of shallow convection, in concert with potentially lagged deep-convective overturning, might impact local convective near-grid-scale circulations Figure 8 (left). This may be followed by sudden release of accumulated convective instability, although this would be rare over tropical ocean in reality (see also Tuckman and Emanuel, 2024). The exact dynamics of infrequent events and shallow overturning could be investigated in more detail by applying our approach and by separating the vertical profiles into additional modes, such as a shallow-convection layer (e.g., 2000-3000m) and a transition layer between the mixed layer and this layer (500m to 2000m).

4.3 General implications

While physics packages should aim to define uniform model-independent rules that connect larger scale tendencies with the precursor state variables, it has often been argued that truly correct formulation of such rules is not possible, since subgrid scales are not truly separated from resolved scales in each numerical model (e.g. Craig and Cohen, 2006; Schemann et al., 2013; Palmer, 2019). This means a single state at the larger mesoscales, beyond 100km, could envelope many states at the scale of individual convective updrafts, and smaller. This idea has directed development of stochasticity in physics (e.g. Plant and Craig, 2008; Berner et al., 2017; Leutbecher et al., 2017; Ollinaho et al., 2017) and has directed MUMIP to possibly its main quests: to which degree is uncertainty addressed in deterministic physics suites, and why? Our analysis indicates that variability among different physics suites in the free troposphere is highly similar, consistently with Groot et al. (2026b). However, the mixed-layer tendency patterns do not summarise into one mode of variability. They are multivariate of nature, along with prominent spin-up.

Our findings have indirect implications related to data assimilation and bias adjustment, at least based on our benchmark of initial states. A coupling between regional CAPE tendencies and parameterised convective precipitation has been suggested by Buschow (2024). Our experiment links to the analysis of Buschow (2024), because we impose our benchmark state and its full dynamics evolution into IFS and GFS physics. This triggers a strong adjustment process, similarly to analysis increments in data assimilation problems, which also affect products like reanalyses, as investigated by the former study.

In Buschow (2024), it has been suggested that differences in CAPE within a wider tropical band relate to a varying proportionality of parameterised convective precipitation in reanalysis data. Thereby, the convection parameterisation is informed about CAPE variability through a partial dependence on CAPE in the convective closure. This imprints on CAPE during the data assimilation. Isolating the EOFs by column, our analysis suggests a wide tropical mean does not represent the associated adjustments comprehensively. After starting with an unbalanced ICON benchmark state, we find mostly tendency variability that decouples precipitation and CAPE within our PCs. In fact, at large scales, the mixed-layer drying tendencies appear to explain our CAPE variability very well. Considerable orthogonality between CAPE and precipitation certainly suggests that

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

the link is much more complicated (multivariate) in realistic experiments than suggested by Buschow (2024), and certainly in IFS. GFS indicates stronger statistical relationships than IFS, which might be attributable to their convective closure, but this remains speculative and probably involves links far beyond those we investigate here.

Nevertheless, one may wonder if the correlations between CAPE and precipitation rate would improve with a different CAPE algorithm, for instance when including entrainment. Entrainment refers to the turbulent mixing and dilution of an initial parcel with its environment. We only mix parcels below 500m. In real-world convection may differently entrain depending on factors as parcel size, altitude, relative humidity, convective organisation and other factors. An implementation of an outdated IFS convection scheme in GFS shows that IFS has a bottom-heavier imprint of entrainment on the tendencies, with most mixing in a vertically restricted layer (Bengtsson et al., 2019, see Figure 2). This is consistent with our deeper drying in GFS than in IFS (Figure 3.4). GFS will weigh more heavily towards a vertical average parcel over the lower 200-400 hPa, which means our CAPE matches better with IFS. This implies a closer association of GFS CAPE with dynamics tendencies above the mixed-layer and, similarly, precipitation rate. We could argue that CAPE of GFS would be a hybrid of ours and one purely associated with EOF1. Nevertheless, the GFS and IFS schemes have been revised since (Bechtold et al., 2020; Bengtsson and Han, 2024). Finally, our mixed-layer definition likely over-emphasises the EOFs, which probably become more convoluted when our CAPE would resemble the GFS algorithm.

Humidity redistribution by physics occurs by design in the cloud, convection and turbulence parameterisations, which includes exchange with liquid water and ice in the cloud reservoirs. Their combined tendencies appear under-dispersed at 20-100km scales in our suites (Groot et al., 2026b) and this notably affects uncertainty in the free-troposphere. In addition, we find systematic drying in the mixed layer, which ultimately affects convective diagnostics like CAPE and LNB. Our mixed-layer either dries through upward mixing of water vapour or through warming (ARPEGE). It appears that the physics suites cannot hold their mixed-layer moisture, although this might be an ICON-relative artifact. Further MUMIP experiments would reveal this. Nevertheless, our results and Groot et al. (2026b) indicate humidity is a difficult quantity to represent accurately with physics suites.

Apparently enhanced, or over-, diffuseness seems to fit a broader picture of specific humidity and relative humidity biases in convection-parameterised weather and climate models. Relative humidity profiles remain very major source of cross-model uncertainty across various types of numerical models and configurations (see also, for instance, Wing et al., 2020), although theoretical arguments for the approximate shape of humidity profiles in the tropics exist (Romps, 2014). While convection-permitting models converge on relative humidity and tend to show a similar negative low-level bias (Lang et al., 2021, 2023), their impact induces profound uncertainty all across the field: on climate time scales via radiation budgets of the troposphere and composition of the stratosphere, but also on weather forecasting time scales through, for instance, tropopause sharpness at higher latitudes (e.g. Saffin et al., 2017; Krüger et al., 2022). Similarly to the latter, our LNB distribution also seems to suggest that the sharpness of the PDF of the vertical stratification gradient near the tropopause in the equatorial band reduces in all physics suites. Similarly, vertical temperature and moisture gradients are very different across our models near the shallow-convection and stratocumulus levels (800-900 hPa), another region with stratification challenges, which suggests that strong temperature gradients are also problematic to represent and stabilise in numerical models.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Code and data availability. The dataset used in this work, along with README instructions, are available via Zenodo (Groot et al., 2026a):
https://doi.org/10.5281/zenodo.18174141 (retrieved: 15-03-2026)

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Appendix A: Time evolution of diagnostics

Figure A1 shows the time evolution of precipitation accumulation, CAPE and LNB over our 31 days experiment. We can see that precipitation accumulation is generally higher over the first few days, around day 15 and on the final days than around day 5-10 and 20-25. CAPE increases during those convective bursts, but it enhances particularly towards the end of the month. The median evolution of LNB has a very similar evolution.

ICON initials are close to the other models in their median LNB, but the diurnal cycle seeems more pronounced in SCMs, especially in ARPEGE. Finally, the upper quartile of CAPE and LNB provide a good proxy for the values in the ITCZ. These are minimal around day 8, with ARPEGE and ICON dropping to 12.000m and the other models even to about 8000m for LNB. CAPE also has low values at 200 j/kg in RAP and 500 in ARPEGE. In the final days, CAPE increases to 1400 j/kg in ARPEGE and about 500-1000 in the other SCMs.

At the middle of our experiment, near day 15, the lower quartile of LNB also tips near 3000m in all models, indicating that the subtropical part of our domain also becomes somewhat convective. Conversely, in the final days, despite the median increasing, this does not happen.

Finally, we can see that all of the physics suites have a mean precipitation accumulation that constantly exceeds their upper quartile, except for GFS, as mentioned in the description of Figure 5. The two are relatively close to each other throughout the month.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

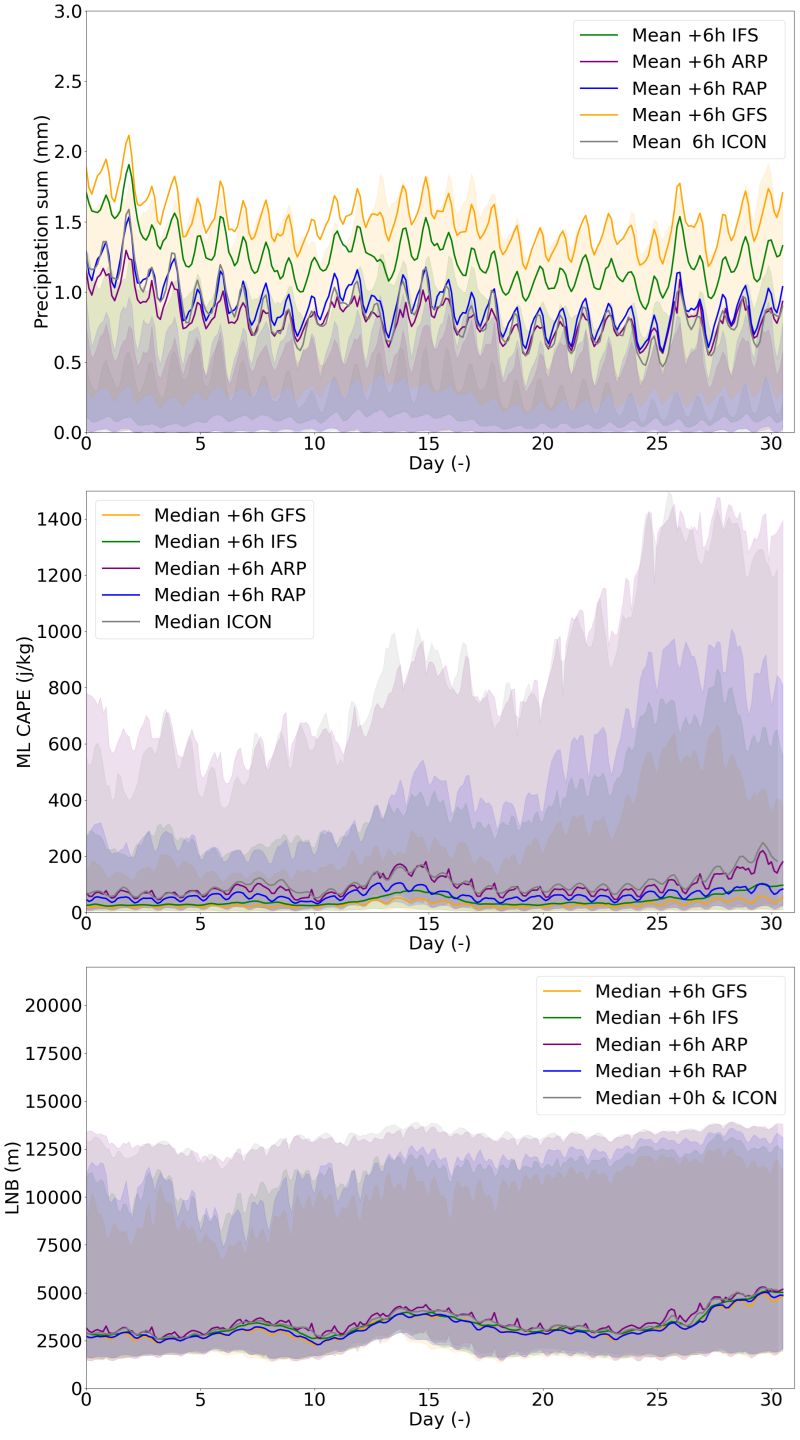

Figure A1. Mean and interquartile range of (top) precipitation accumulation, median and interquartile range of (middle) CAPE and median and interquartile range of (bottom) LNB over the coarse of our 31 days experiment.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Author contributions. Conceptualisation, formal analysis and writing of original draft: EG; Project data curation: EG, HC, XS, KN, WL, RR; Reviewing: EG, HC, XS, WL, LB, JS, KW, HL, RR; Funding acquisition: HC.

Competing interests. The authors are not aware of any competing interests.

Acknowledgements. EG, WL, and HC acknowledge funding from the Leverhulme Trust, Grant Number: RPG-2022-192.
XS and JS were supported in part by NOAA cooperative agreement NA22OAR4320151, for the Cooperative Institute for Earth System Research and Data Science (CIESRDS). The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.
The research of LB is supported by US federally appropriated funds.
The authors would like to thank Ligia Bernardet for checking the original draft and her role in the initial development of the Model Uncertainty Model Intercomparison Project.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

References

Documentation GFS Physics suite, https://dtcenter.ucar.edu/GMTB/v6.0.0/sci_doc/_g_f_s_v16_page.html.

Documentation RAP Physics suite, https://dtcenter.ucar.edu/GMTB/v6.0.0/sci_doc/rap_suite_page.html.

Arakawa, A. and Schubert, W. H.: Interaction of a cumulus cloud ensemble with the large-scale environment, Part I, Journal of Atmospheric Sciences, 31, 674–701, 1974.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, 2015.

Baumgart, M., Ghinassi, P., Wirth, V., Selz, T., Craig, G., and Riemer, M.: Quantitative View on the Processes Governing the Upscale Error Growth up to the Planetary Scale Using a Stochastic Convection Scheme, Monthly Weather Review, 147, 1713–1731, https://doi.org/10.1175/mwr-d-18-0292.1, 2019.

Bechtold, P., Sandu, I., Klocke, D., Semane, N., Ahlgrimm, M., Beljaars, A., Forbes, R., and Rodwell, M.: The role of shallow convection in ECMWF's Integrated Forecasting System, European Centre for Medium-Range Weather Forecasts, 2014a.

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J., Beljaars, A., and Bormann, N.: Representing equilibrium and nonequilibrium convection in large-scale models, Journal of the Atmospheric Sciences, 71, 734–753, 2014b.

Bechtold, P., Forbes, R., Sandu, I., Lang, S., and Ahlgrimm, M.: A major moist physics upgrade for the IFS, ECMWF Newsletter, 164, 24–32, 2020.

Bengtsson, L. and Han, J.: Updates to NOAA's Unified Forecast System's Cumulus Convection Parameterization Scheme between GFSv16 and GFSv17, Weather and Forecasting, 39, 1559 – 1570, https://doi.org/10.1175/WAF-D-23-0232.1, 2024.

Bengtsson, L., Dias, J., Gehne, M., Bechtold, P., Whitaker, J., Bao, J.-W., Magnusson, L., Michelson, S., Pegion, P., Tulich, S., and Kiladis, G. N.: Convectively Coupled Equatorial Wave Simulations Using the ECMWF IFS and the NOAA GFS Cumulus Convection Schemes in the NOAA GFS Model, Monthly Weather Review, 147, 4005 – 4025, https://doi.org/10.1175/MWR-D-19-0195.1, 2019.

Bernardet, L., Bengtsson, L., Reinecke, P. A., Yang, F., Zhang, M., Hall, K., Doyle, J., Martini, M., Firl, G., and Xue, L.: Common Community Physics Package: Fostering Collaborative Development in Physical Parameterizations and Suites, Bulletin of the American Meteorological Society, 105, E1490 – E1505, https://doi.org/10.1175/BAMS-D-23-0227.1, 2024.

Berner, J., Achatz, U., Batte, L., Bengtsson, L., Cámara, A. d. l., Christensen, H. M., Colangeli, M., Coleman, D. R., Crommelin, D., Dolaptchiev, S. I., et al.: Stochastic parameterization: Toward a new view of weather and climate models, Bulletin of the American Meteorological Society, 98, 565–588, 2017.

Buschow, S.: Tropical convection in ERA5 has partly shifted from parameterized to resolved, Quarterly Journal of the Royal Meteorological Society, 150, 436–446, 2024.

Christensen, H.: Constraining stochastic parametrisation schemes using high-resolution simulations, Quarterly Journal of the Royal Meteorological Society, 146, 938–962, 2020.

Christensen, H. M., Dawson, A., and Holloway, C. E.: Forcing single-column models using high-resolution model simulations, Journal of Advances in Modeling Earth Systems, 10, 1833–1857, 2018.

Craig, G. and Cohen, B.: Fluctuations in an Equilibrium Convective Ensemble. Part I: Theoretical Formulation, Journal of the Atmospheric Sciences, 63, 1996 – 2004, https://doi.org/10.1175/JAS3709.1, 2006.

Donner, L. J. and Phillips, V. T.: Boundary layer control on convective available potential energy: Implications for cumulus parameterization, Journal of Geophysical Research: Atmospheres, 108, 2003.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

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

Forbes, R., Laloyaux, P., and Rodwell, M.: Ifs upgrade improves moist physics and use of satellite observations, ECMWF Newsletter, 169, 17–24, 2021.

Groot, E. and Riemer, M.: When tiny convective spread affects a midlatitude jet: Spread sequence, Quarterly Journal of the Royal Meteorological Society, 9999, 2025.

Groot, E. and Tost, H.: Evolution of squall line variability and error growth in an ensemble of large eddy simulations, Atmospheric Physics and Chemistry, 2023. *Note: Citation includes its supplement.*, https://doi.org/https://doi.org/10.5194/acp-23-565-2023, 2023a.

Groot, E. and Tost, H.: Divergent convective outflow in large-eddy simulations, Atmospheric Chemistry and Physics, 23, 6065–6081, https://doi.org/https://doi.org/10.5194/acp-23-6065-2023, 2023b.

Groot, E., Kuntze, P., Miltenberger, A., and Tost, H.: Divergent convective outflow in ICON deep-convection-permitting and parameterised deep convection simulations, Weather and Climate Dynamics, 5, 779–803, 2024.

Groot, E., Christensen, H., Sun, X., Newman, K., Lfarh, W., and Roehrig, R.: Spin-up in the Model Uncertainty Model Intercomparison Project: humidity and temperature adjustment and its consequences for convective diagnostics, https://doi.org/10.5281/zenodo.18174141, 2026a.

Groot, E., Christensen, H., Sun, X., Newman, K., Lfarh, W., Roehrig, R., Bengtsson, L., and Simonson, J.: How different are deterministic physics suites when coupled to fixed model dynamics and why?, arXiv preprint arXiv:2601.03393, 2026b.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Houze, R.: Mesoscale convective systems, Reviews of Geophysics, 42, https://doi.org/10.1029/2004rg000150, 2004.

Krüger, K., Schäfler, A., Wirth, M., Weissmann, M., and Craig, G. C.: Vertical structure of the lower-stratospheric moist bias in the ERA5 reanalysis and its connection to mixing processes, Atmospheric Chemistry and Physics, 22, 15 559–15 577, https://doi.org/10.5194/acp-22-15559-2022, 2022.

Lambert, F., Challenor, P., Lewis, N., McNeall, D., Owen, N., Boutle, I., Christensen, H., Keane, R., Mayne, N., Stirling, A., et al.: Continuous structural parameterization: A proposed method for representing different model parameterizations within one structure demonstrated for atmospheric convection, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 085, 2020.

Lang, T., Naumann, A. K., Stevens, B., and Buehler, S. A.: Tropical free-tropospheric humidity differences and their effect on the clear-sky radiation budget in global storm-resolving models, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 514, 2021.

Lang, T., Naumann, A. K., Buehler, S. A., Stevens, B., Schmidt, H., and Aemisegger, F.: Sources of uncertainty in mid-tropospheric tropical humidity in global storm-resolving simulations, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 443, 2023.

Lawrence, M. G.: The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications, Bulletin of the American Meteorological Society, 86, 225–234, 2005.

Leutbecher, M., Lock, S., Ollinaho, P., Lang, S., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R., Goddard, J., Haiden, T., Hogan, R., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P., Subramanian, A., Vitart, F., Wedi, N., and Weisheimer, A.: Stochastic representations of

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

model uncertainties at ECMWF: state of the art and future vision, Quarterly Journal of the Royal Meteorological Society, 143, 2315–2339, https://doi.org/https://doi.org/10.1002/qj.3094, 2017.

Lorenz, E. N.: Empirical orthogonal functions and statistical weather prediction, vol. 1, Massachusetts Institute of Technology, Department of Meteorology Cambridge, 1956.

Mapes, B.: Gregarious Tropical Convection, Journal of Atmospheric Sciences, 50, 2026 – 2037, https://doi.org/10.1175/1520-0469(1993)050<2026:GTC>2.0.CO;2, 1993.

McDonald, J. E.: James Espy and the Beginnings of Cloud Thermodynamics, Bulletin of the American Meteorological Society, 44, 634 – 641, https://doi.org/10.1175/1520-0477-44.10.634, 1963.

Melhauser, C. and Zhang, F.: Practical and intrinsic predictability of severe and convective weather at the Mesoscales, Journal of the Atmospheric Sciences, 69, 3350–3371, https://doi.org/10.1175/jas-d-11-0315.1, 2012.

Ollinaho, P., Lock, S., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R., Haiden, T., Hogan, R., and Sandu, I.: Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble, Quarterly Journal of the Royal Meteorological Society, 143, 408–422, 2017.

Palmer, T.: Stochastic weather and climate models, Nature Reviews Physics, 1, 463–471, 2019.

Plant, R. S. and Craig, G. C.: A Stochastic Parameterization for Deep Convection Based on Equilibrium Statistics, Journal of the Atmospheric Sciences, 65, 87–105, https://doi.org/10.1175/2007jas2263.1, 2008.

Rodwell, M., Lang, S. T., Ingleby, N., Bormann, N., Hólm, E., Rabier, F., Richardson, D., and Yamaguchi, M.: Reliability in ensemble data assimilation, Quarterly Journal of the Royal Meteorological Society, 142, 443–454, 2016.

Rodwell, M. J. and Palmer, T. N.: Using numerical weather prediction to assess climate models, Quarterly Journal of the Royal Meteorological Society, 133, 129–146, https://doi.org/https://doi.org/10.1002/qj.23, 2007.

Rodwell, M. J., Richardson, D. S., Parsons, D. B., and Wernli, H.: Flow-dependent reliability: A path to more skillful ensemble forecasts, Bulletin of the American Meteorological Society, 99, 1015–1026, 2018.

Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Abdel-Lathif, A. Y., Bazile, E., Belamari, S., et al.: The CNRM global atmosphere model ARPEGE-Climat 6.3: Description and evaluation, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 075, 2020.

Romps, D. M.: An Analytical Model for Tropical Relative Humidity, Journal of Climate, 27, 7432 – 7449, https://doi.org/10.1175/JCLI-D-14-00255.1, 2014.

Romps, D. M.: Exact expression for the lifting condensation level, Journal of the Atmospheric Sciences, 74, 3891–3900, 2017.

Saffin, L., Gray, S. L., Methven, J., and Williams, K. D.: Processes Maintaining Tropopause Sharpness in Numerical Models, Journal of Geophysical Research: Atmospheres, 122, 9611–9627, https://doi.org/https://doi.org/10.1002/2017JD026879, 2017.

Schemann, V., Stevens, B., Grützun, V., and Quaas, J.: Scale Dependency of Total Water Variance and Its Implication for Cloud Parameterizations, Journal of the Atmospheric Sciences, 70, 3615 – 3630, https://doi.org/10.1175/JAS-D-13-09.1, 2013.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, Geophysical Research Letters, 44, 12–396, 2017.

Schumacher, C., Houze, R., and Kraucunas, I.: The Tropical Dynamical Response to Latent Heating Estimates Derived from the TRMM Precipitation Radar, Journal of the Atmospheric Sciences, 61, 1341 – 1358, https://doi.org/10.1175/1520-0469(2004)061<1341:TTDRTL>2.0.CO;2, 2004.

https://doi.org/10.5194/egusphere-2026-1445
Preprint. Discussion started: 4 May 2026

Selz, T., Riemer, M., and Craig, G. C.: The transition from practical to intrinsic predictability of midlatitude weather, Journal of the Atmospheric Sciences, 79, 2013–2030, 2022.

Sengupta, S. and Boyle, J. S.: Using Common Principal Components for Comparing GCM Simulations, Journal of Climate, 11, 816 – 830, https://doi.org/10.1175/1520-0442(1998)011<0816:UCPCFC>2.0.CO;2, 1998.

Sherwood, S., Roca, R., Weckwerth, T., and Andronova, N.: Tropospheric water vapor, convection, and climate, Reviews of Geophysics, 48, 2010.

Shutts, G. and Palmer, T.: Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem, Journal of climate, 20, 187–202, 2007.

Soden, B. J. and Bretherton, F. P.: Evaluation of water vapor distribution in general circulation models using satellite observations, Journal of Geophysical Research: Atmospheres, 99, 1187–1210, https://doi.org/https://doi.org/10.1029/93JD02912, 1994.

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

Tuckman, P. J. and Emanuel, K.: Origins of Extreme CAPE Around the World, Journal of Geophysical Research: Atmospheres, 129, e2024JD041 833, https://doi.org/https://doi.org/10.1029/2024JD041833, e2024JD041833 2024JD041833, 2024.

Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S. A., Senior, C. A., and Williamson, D. L.: The Transpose-AMIP II Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models, Journal of Climate, 26, 3258 – 3274, https://doi.org/10.1175/JCLI-D-12-00429.1, 2013.

Wing, A. A., Stauffer, C. L., Becker, T., Reed, K. A., Ahn, M.-S., Arnold, N. P., Bony, S., Branson, M., Bryan, G. H., Chaboureau, J.-P., et al.: Clouds and convective self-aggregation in a multimodel ensemble of radiative-convective equilibrium simulations, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 138, 2020.

Xavier, P. K., Petch, J. C., Klingaman, N. P., Woolnough, S. J., Jiang, X., Waliser, D. E., Caian, M., Cole, J., Hagos, S. M., Hannay, C., Kim, D., Miyakawa, T., Pritchard, M. S., Roehrig, R., Shindo, E., Vitart, F., and Wang, H.: Vertical structure and physical processes of the Madden-Julian Oscillation: Biases and uncertainties at short range, Journal of Geophysical Research: Atmospheres, 120, 4749–4763, https://doi.org/https://doi.org/10.1002/2014JD022718, 2015.

Yano, J.-I. and Plant, R.: Convective quasi-equilibrium, Reviews of Geophysics, 50, 2012.

Zhang, F., Sun, Y., Magnusson, L., Buizza, R., Lin, S., Chen, J., and Emanuel, K.: What Is the Predictability Limit of Midlatitude Weather?, Journal of the Atmospheric Sciences, 76, 1077 – 1091, https://doi.org/10.1175/JAS-D-18-0269.1, 2019.