Key Points:

- The representation of turbulence and convection in the convective gray zone is investigated at global 5 km resolution
- In the examined context a reduced mass-flux convection scheme is beneficial in 5 km-resolution global model forecasts
- The assessment is more conclusive with regard to convective organization and tropical variability than extratropical predictability

Correspondence to:
L. Tomassini,
lorenzo.tomassini@metoffice.gov.uk

Author Contributions:

Conceptualization: Lorenzo Tomassini
Formal analysis: Lorenzo Tomassini, Julian Heming, José M. Rodriguez, Duncan Ackerley, Prince Xavier, Charmaine Franklin
Methodology: Lorenzo Tomassini
Software: Lorenzo Tomassini, Martin Willett, Adrian Lock, Michael Whitall, Paul Earnshaw
Visualization: Lorenzo Tomassini, Julian Heming, José M. Rodriguez, Duncan Ackerley, Prince Xavier, Charmaine Franklin

Confronting the Convective Gray Zone in the Global Configuration of the Met Office Unified Model

Lorenzo Tomassini[1] [ID]**, Martin Willett[1], Alistair Sellar[1]** [ID]**, Adrian Lock[1]** [ID]**, David Walters[1], Michael Whitall[1], Claudio Sanchez[1], Julian Heming[1], Paul Earnshaw[1], José M. Rodriguez[1], Duncan Ackerley[1], Prince Xavier[1]** [ID]**, Charmaine Franklin[2]** [ID]**, and Catherine A. Senior[3]** [ID]

[1]Met Office, Exeter, UK, [2]Bureau of Meteorology, Melbourne, VIC, Australia, [3]Met Office Hadley Centre, Exeter, UK

Abstract In atmospheric models with kilometer-scale grids the resolution approaches the scale of convection. As a consequence the most energetic eddies in the atmosphere are partially resolved and partially unresolved. The modeling challenge to represent convection partially explicitly and partially as a subgrid process is called the convective gray zone problem. The gray zone issue has previously been discussed in the context of regional models, but the evolution in regional models is constrained by the lateral boundary conditions. Here we explore the convective gray zone starting from a defined global configuration of the Met Office Unified Model using initialized forecasts and comparing different model formulations to observations. The focus is on convection and turbulence, but some aspects of the model dynamics are also considered. The global model is run at nominal 5 km resolution and thus contributions from both resolved and subgrid turbulent and convective fluxes are non-negligible. The main conclusion is that in the present assessment, the configurations which include scale-aware turbulence and a carefully reduced and simplified mass-flux convection scheme outperform both the configuration with fully parameterized convection as well as a configuration in which the subgrid convection parameterization is switched off completely. The results are more conclusive with regard to convective organization and tropical variability than extratropical predictability. The present study thus endorses the strategy to further develop scale-aware physics schemes and to pursue an operational implementation of the global 5 km-resolution model to be used alongside other ensemble forecasts to allow researchers and forecasters to further assess these simulations.

Plain Language Summary An in-depth exploration of kilometer-scale global atmospheric modeling in the context of the current Met Office modeling system, the Met Office Unified Model, is presented. All simulations were performed using a global atmosphere model at nominal 5 km resolution. The model resolution thus resides in the so-called convective gray zone where the grid length approaches the scale of turbulence and convection, and contributions from both resolved and subgrid convective and turbulent fluxes are non-negligible. A case study approach has been taken in which various testbed cases are defined and model forecasts are evaluated against observations. The focus is on the representation of convection, turbulence and dynamics, the key aspects of the model formulation in the convective gray zone. The main finding is that in the present assessment the configurations which include a scale-aware representation of turbulence and a carefully reduced and simplified convection scheme outperform both the reference model with fully parameterized convection as well as a configuration in which the subgrid convection parameterization is switched off completely. An outlook on further work toward the development of an adequate kilometer-scale resolution global coupled modeling system for use across weather and climate time scales is given.

1. Introduction

Coarse-resolution atmospheric models do not resolve any of the main energy-containing turbulence scales, and therefore turbulence is essentially unresolved and has to be parameterized. Wyngaard (2004) discussed models with resolution in between coarse mesoscale models (grid length 10 km and larger) and large-eddy resolving models (grid length 100 m and smaller) for which the resolution is similar to the length scale of the most energetic eddies in the atmosphere. The range of values for the length scale of these energetic eddies and coherent structures depends on the prevailing meteorological situation, but can vary from less than 100 m in shallow convection to several tens of kilometers in organized convective systems. This has led to the articulation of the "gray zone" problem, the question about the adequate atmospheric model formulation in situations where model resolution

Writing – original draft: Lorenzo Tomassini, Martin Willett, Alistair Sellar, Adrian Lock, David Walters, Michael Whitall, Claudio Sanchez, Julian Heming, Paul Earnshaw, José M. Rodriguez, Duncan Ackerley, Prince Xavier, Charmaine Franklin, Catherine A. Senior

approaches the scale of turbulence and convection (de Roode et al., 2019; Field et al., 2017; Honnert et al., 2020; Tomassini et al., 2017).

A distinction is often made between the gray zone of turbulence and the gray zone of convection, but in practice there is no clear-cut separation and the representation of turbulence has a significant impact on, for example, the modeling of organized deep convective systems (Honnert et al., 2020). Turbulent eddies and convective motions usually occur simultaneously and across a wide range of scales. Ultimately, all atmospheric scales are deeply intertwined. The terms "convection-permitting" or "turbulence-permitting" to designate models operating in the gray zone of turbulence and convection (in contrast to "convection-resolving" and "turbulence-resolving") are adequate as they reflect the fact that convection and turbulence are partly resolved and partly unresolved in these regimes.

The term "convection-permitting" does not imply that the subgrid convection scheme is switched off completely in the model (Kendon et al., 2021). On the contrary, it suggests that part of the convective motion is still unresolved and needs to be parameterized, either through an adequately formulated (ideally scale-aware) and calibrated convection scheme or a non-local term in the turbulence parameterization, or both. In one of the earliest explorations of the convective gray zone problem, Roberts (2003) suggested to use a CAPE-dependent (CAPE for Convective Available Potential Energy) CAPE-closure time scale in the convection scheme (see also Lean et al. (2008)). The idea was that in high-CAPE environments convective systems are more organized and exhibit a larger spatial extent and that the model should be able to resolve these larger convective structures, whereas smaller convection would still need to be parameterized.

Gerard and Geleyn (2005) proposed to use a prognostic closure in the representation of subgrid convection involving prognostic updraught vertical velocities and a prognostic fractional area of a model grid box covered by convective clouds to address the gray zone problem. In related earlier work Pan and Randall (1998) argued that the distinction between convective and large-scale processes is ambiguous anyway. Arakawa et al. (2011) and Arakawa and Wu (2013) framed the problem in a similar way, and various subsequent studies followed an analogous approach (Grell & Freitas, 2014; Kwon & Hong, 2017; Sakradzija et al., 2016; Su et al., 2021; W. Wang, 2022; Zheng et al., 2016). Moreover, Gerard et al. (2009) and Gerard (2015) highlighted the importance of the interaction between convection parameterization, microphysics, and the cloud scheme in the context of the gray zone issue.

Specific modifications to the convection scheme such as a better coupling between subgrid convection and the resolved mesoscale circulation (Becker et al., 2021), the representation of low-CAPE convection (McTaggart-Cowan et al., 2020), or the targeted adjustment of specific parameters in an existing cumulus parameterization (Miyakawa et al., 2018) can also prove important and beneficial in convection-permitting simulations even if these aspects do not address the issue of scale-awareness of the convection parameterization directly from a conceptual perspective.

Convection-permitting regional models have been used with considerable success for more than two decades (Clark et al., 2016; Kendon et al., 2021; Prein et al., 2015), but the dynamics in regional models is constrained and influenced by the prescribed lateral boundary conditions (Dipankar et al., 2020; Kendon et al., 2010; Radermacher & Tomassini, 2012). In the context of the current modeling system at the Met Office, the Met Office Unified Model, the present study reports on a first thorough exploration of convection-permitting global modeling, besides the preliminary study Tomassini (2018). In this work we focus on some of the main and most basic model configuration choices concerning convection, boundary layer processes, turbulence, and to some extent model dynamics. These choices address the gray zone problem in one way or another and are part of what is sometimes discussed under the headline of "gray zone model physics." In order to deal with a tractable and well-defined problem, different options concerning the microphysics and cloud schemes have not been considered here.

The present work is not the first study on kilometer-scale global atmospheric modeling. The field was pioneered primarily by the group who developed the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) (Miyakawa & Miura, 2019; Miyakawa et al., 2018; Satoh et al., 2008, 2019; Tomita & Satoh, 2004). More recently, various other institutions have developed dynamical cores and the necessary infrastructure to run kilometer-scale global models (Caldwell et al., 2021; Harris et al., 2020; Malardel & Bechtold, 2019; Putnam & Suarez, 2011; Skamarock et al., 2012; Stevens et al., 2019; Tomassini, 2018; Tomassini et al., 2017; Zängl et al., 2015).

The development of global convection-permitting models is more than just an incremental enhancement in resolution. Such models allow for, at least partly, resolving fundamental new turbulent and convective phenomena in the atmosphere, studying interactions between convection and the atmospheric circulation across scales, and thus addressing important and novel science questions (Senior et al., 2021; Slingo et al., 2022; Tomassini & Yang, 2022; Tomassini, 2020). This step change in fidelity and realism of weather and climate models implies also a step change in forecasting severe convective storms and in supporting and informing climate change mitigation and adaptation measures for the benefit of societies around the world (Slingo et al., 2021). In the presented work, however, the focus is on developing an appropriate tool for these exciting scientific endeavors, and on better understanding the strengths, limitations, and sensitivities of this tool.

The outline of the paper is as follows. In Section 2 the main model configuration options that are studied in the present work are introduced and briefly described. The results from simulations of different forecasts covering varying meteorological conditions are contained in Section 3. The process-based analysis focusses on different locations, times of the year, and particular phenomena. Section 4 highlights some important aspects of the convective gray zone problem and discusses a few additional sensitivity experiments. Two whole months, July 2016 and January 2018, were simulated based on a particular convection-permitting configuration with a view toward assessing statistics on climate time scales, and the analysis is showcased in Section 5. Finally, in Section 6 the main findings are summarized and an outlook on future planned work is given.

2. Description of Model Configurations

In the present study the reference configuration is Global Atmosphere version 7.0 (Walters et al., 2019), *GA7* in short, as used in deterministic forecasts at the Met Office, that is, with climatological aerosols, without the stochastic physics package, and using 70 levels in the vertical. The simulations are based on the Unified Model code version 11.7. The configurations described in this paper include some of the changes or specifications described in Table 1.

The main model configurations considered are a combination of the configuration options shown in Table 1, namely *GA7*, *MidLevShConv25*, *MidLevShConv25RAturb*, *MidLevShConv15RAturb*, and *ConvOffMoistCons-RAturb*. The latter configuration, *ConvOffMoistConsRAturb*, is chosen in such a way that it shares primary convection, dynamics, and turbulence model formulation features with some of the Unified Model Regional Atmosphere configurations (Bush et al., 2020, 2022). A few important sensitivities and additional configuration tests will be discussed in Section 4.

All simulations are run at global N2560 (nominal 5 km) resolution unless stated otherwise. N2560 refers to a latitude-longitude grid with 1,920 regular latitude lines between the pole and equator, and 5,120 longitude points along each latitude line. The grid length is 5 km in latitude direction everywhere, in longitude direction the grid length is about 7.8 km at the equator, 5 km in the midlatitudes, and 4 m near the poles. In the following the term "5 km-resolution" will be used for the sake of simplicity. Daily varying Operational Sea Surface Temperature and Ice Analysis (OSTIA) sea surface temperatures and sea ice are prescribed. The model time step is 90 s. Soil moisture is initialized from Met Office Unified Model analysis.

In all configurations except *GA7*, the deep convection parameterization is switched off. In the configurations that include the *MidLev* option the midlevel convection scheme is allowed to start from the top of the boundary layer and not lower. The mixing in the boundary layer is left to the boundary layer turbulence and shallow convection schemes. The rationale is to more realistically represent nonequilibrium convection such as the diurnal cycle of convection and the convection tied to advective boundary layers. The approach combines a reduced subgrid convection approach based on a CAPE closure for the free troposphere and a separate representation of boundary layer processes, acknowledging that free-tropospheric adjustment time scales can differ from boundary layer time scales and that there often is an imbalance between boundary layer heating and deep convective overturning (Bechtold et al., 2014). The choice of a CAPE closure time scale of 2,700 s is aimed at reducing the subgrid convective mass flux according to the model resolution, allowing convection to be partly explicit, and at the same time to account for the fact that vertical motion is not fully resolved by a model with 5 km grid spacing. Sensitivity experiments with regard to the CAPE closure time scale are presented in Section 4.1. One should stress that even when the deep convection parameterization is switched off in some of the configurations it is still possible for at least some of the surface-based deep convective processes in the model to be handled by the subgrid schemes.

Table 1
Description of Different Model Configuration Options

Configuration indicators	Description	Configuration details
GA7	Standard GA7.0 configuration	See Walters et al. (2019)
ShConv15/ShConv25	Scaling of the shallow convective mass flux	Shallow convective mass flux scaling: 0.015 or 0.025 Standard value is 0.03
MidLev	Revised formulation of the midlevel convection scheme and deep convection parameterization switched off	Deep convection parameterization switched off Fixed CAPE timescale of 2,700 s Prognostic entrainment for midlevel convection Allow to initiate from the top of the boundary layer Time-smoothing of midlevel convective increments Adjustment of midlevel convective momentum transport setting Forced detrainment-weighted CAPE calculation Additional termination condition for convection
RAturb	Turbulence and boundary layer	Blending scheme as in RA1 (Bush et al., 2020), with fraction of maximum diffusion = 0.75 SHARPEST scheme for stable boundary layers Turbulent mixing based on interactively diagnosed turbulent layer depths
MoistCons	Moisture conservation option	Local moisture conservation scheme for advection as in RA1 (Bush et al., 2020)
ConvOff	Convection parameterization option	No convection parameterization used

Note. See also main body of text in Section 2 and additional information in Appendix A.

The use of a prognostic entrainment rate (Willett & Whitall, 2017) takes into account changing convective entrainment in different stages of the life cycle of convective systems and varying degrees of convective organization. Convective increment time-smoothing in addition improves the coupling between subgrid convection and the resolved circulation, reduces time intermittency of the convection scheme, and undesirable dynamical effects of this intermittency such as spurious gravity waves. One particular advantage of unifying the subgrid convection treatment in the free troposphere is the avoidance of a late afternoon "tea break" in convective rainfall, especially over tropical land. In previous model versions an erroneous minimum in modeled convective rainfall featured in the evening when mainly surface-based convection handled by the deep convection scheme was passed over to the midlevel scheme which treated elevated convection.

In convection-permitting Unified Model simulations, single grid-column updrafts can become unrealistically intense and persistent because a stagnation point forms at the base of the updraught. Equal convergent velocities at either side of an updraught column mean the semi-Lagrangian (SL) advection scheme's back-trajectories from the cell center go straight down, and so fail to advect in what will typically be drier air from neighboring columns. This allows single-point updraught columns to keep creating their own moisture by copying a near-surface moist anomaly upwards. The spuriously created moisture feeds condensation and latent heating, generating stronger ascent and convergence, leading to a positive feedback. To address the problem, all simulations, except *GA7* and *ConvOffMoistConsRAturb*, use the "fountain buster," an implementation of a modification to the SL advection scheme, aimed at making a local correction to the lack of conservation. The scheme is called directly after the SL advection increment has been calculated, by interpolation of the virtual potential temperature and moisture variables to departure points. It works by identifying grid points where the horizontal winds (on the grid cell faces) are converging, and adds onto the standard SL increment a simple linear up-wind advection increment arising from just the locally convergent part of the flow. In this way it is adding in the effects of just the convergent inflow that will have been missed by the SL advection (because SL advection interpolates the cell-face winds to the cell center in order to find the departure point).

The label *ShConv* refers to a multiplicative scaling of the shallow convective subgrid mass flux which is included in its default setting also in the reference configuration *GA7* and is used in some of the convection-permitting configurations to limit the influence of the subgrid shallow convection scheme. The configuration option *RAturb* includes the turbulence blending scheme (Boutle et al., 2014) which blends the one-dimensional turbulence parameterization as used in lower-resolution versions of the global model with a three-dimensional Smagorinsky-Lilly representation of subgrid turbulence as typically applied in large-eddy simulation models. The two schemes are combined via the mixing length and the key parameter is the ratio of grid length to boundary layer depth. The SHARPEST scheme for stable boundary layers as described in Derbyshire (1997) and used in the Regional Atmosphere configuration (Bush et al., 2020, 2022) is employed. The fraction of the maximum allowed value of the diffusion coefficient is set to 0.75 (Hanley et al., 2015). Moreover, the free atmospheric turbulent mixing in *RAturb* is based on interactively diagnosed turbulent layer depths throughout the atmospheric column (as in the tropical version of the Regional Atmosphere configuration described in Bush et al. (2022)).

All simulations, except *GA7* and *ConvOffMoistConsRAturb*, use the parameter choice puns = 1.0 in the non-linearity setting for the boundary layer solver in the case of unstable boundary layers which makes the implicit solver more stable (Wood et al., 2007; see Appendix A). All configurations use the multigrid solver (Maynard et al., 2020).

For Met Office Global Model Evaluation and Development (GMED) tickets associated with the different changes and some more information on the tickets see the Appendix A.

3. Case Studies and Results

A case study approach is taken in which testbed cases are defined and the model is evaluated against observations and reanalysis data. Five model configurations are further investigated in greater detail: *GA7*, *MidLevShConv25*, *MidLevShConv25RAturb*, *MidLevShConv15RAturb*, and *ConvOffMoistConsRAturb*. The focus of the study is on model hindcasts with lead times of up to 10 days because this way the simulations can be directly compared to observations. Note that some errors in the hindcast will come from deficiencies in the initial conditions or a lack of predictability not captured due to running only one hindcast per configuration per initialization time. The

different cases aim at covering different, important meteorological conditions and phenomena in different parts of the world and at various times during the year with a focus on high-impact weather.

3.1. African Easterly Wave Case and Convection-Circulation Coupling Over Africa

An African easterly wave disturbance is detectable starting from 18:00 UTC on 7 July 2010, over North Africa (Tomassini, 2018; Tomassini et al., 2017). The dynamics of the wave is rather weak over the first 30 hr after detection, that is, until about 00:00 UTC on 9 July. Starting around 9 July 03:00 UTC a crucial strengthening phase of the wave occurs, which lasts for about 2 days. Tropical Rainfall Measuring Mission (TRMM) precipitation (Huffman et al., 2007) shows distinct organized precipitation ahead of the trough at around 10°–20° North where the main center of the wave disturbance is located (Figure 1, panel a). The case is of interest because it allows for assessing the coupling between moist convection and atmospheric circulations over tropical land in the different model configurations. The simulations are initialized on 7 July 2010, 00:00 UTC and on 11 July 2010, 00:00 UTC with European Center for Medium-Range Weather Forecasts (ECMWF) operational analyses, and two 5-day hindcasts are performed.

GA7 is not able to develop and predict the mesoscale convective systems associated with the wave in the crucial early, developing stage (Figure 1, panel b). The early stage of the wave development is key for the ability of the model to actually predict the rainfall associated with the wave. The initial phase of the wave evolution makes the sensitivity of the modeled convective rainfall to atmospheric circulations most apparent since the dynamical imprint of the wave is still relatively weak. The interaction between the wave dynamics and organized convective systems at and ahead of the trough are an important ingredient for the subsequent strengthening of the wave (Tomassini, 2018; Tomassini et al., 2017). *GA7* shows hardly any propagating rainfall during the first hindcast simulation but instead exhibits horizontal stripes of precipitation in the Hovmöller plot which stem from an erroneous diurnal cycle in convective rainfall over the area.

Also in *ConvOffMoistConsRAturb* the predictive skill related to the convective systems at and ahead of the wave trough (close to the diagonal from top right to bottom left in the Hovmöller plot; for a snapshot see also Figures 2 and 3 with the approximate location of the wave trough indicated by a vertical red line) is not satisfactory in the earlier stage of the wave (Figure 1, panel f). The detectable precipitating convective systems are mainly behind the wave trough (not shown) and therefore not reinforcing the wave circulation. Moreover, in the early phase, when the dynamical forcing by the wave is weak, the configuration exhibits horizontal stripes in the rainfall Hovmöller plot suggesting issues with the diurnal cycle of convection over tropical land under weak forcing.

In the *MidLevShConv* configurations (panels c–e) the coupling between convection and the wave dynamics is well represented as indicated by the propagating convective systems around the center of the wave, and moist convection supports the dynamical development of the wave (not shown). However, in *MidLevShConv25* the speeds of the convective systems around the wave trough seem to be slightly too slow in the first forecast suggested by the slight discontinuity in the predicted rainfall between the first and the second hindcast (Figure 1, panel c), whereas this aspect is better represented in *MidLevShConv25RAturb* and *MidLevShConv15RAturb* (Figure 1, panel d and e, respectively).

These characteristics are also confirmed in observed and simulated satellite imagery (Figures 2 and 3), showing reflectance based on the visible channel at 0.8 μm on 10 July, 12:00 UTC (Figure 2), and brightness temperature based on the IR channel at 10.8 μm on 10 July, 18:00 UTC (Figure 3), that is, almost 4 days after forecast initialization. The dashed red lines in the plots indicate the approximate position of the wave trough as diagnosed from the ECMWF analysis. The satellite simulated fields from the global model simulations are derived using an offline radiative transfer code (Saunders et al., 2018).

In *GA7* the convective systems over Africa and their organization are not represented realistically (Figures 2 and 3, top right panel), although it should be mentioned that the satellite simulator does not consider cloud water or ice that is held within the convection scheme. *MidLevShConv25RAturb* and *MidLevShConv15RAturb* develop well organized and realistic mesoscale convective systems at the wave trough and over the African continent. The same could be said of *MidLevShConv25*, however, as in the Hovmöller diagrams of Figure 1, in relation to the wave the main convective systems are located behind the trough (Figures 2 and 3, left panel in the middle row), suggesting that they move somewhat too slowly. The location of the convective storms relative to the wave trough is important because it affects the way moist convection couples and feeds back to the wave circulation.
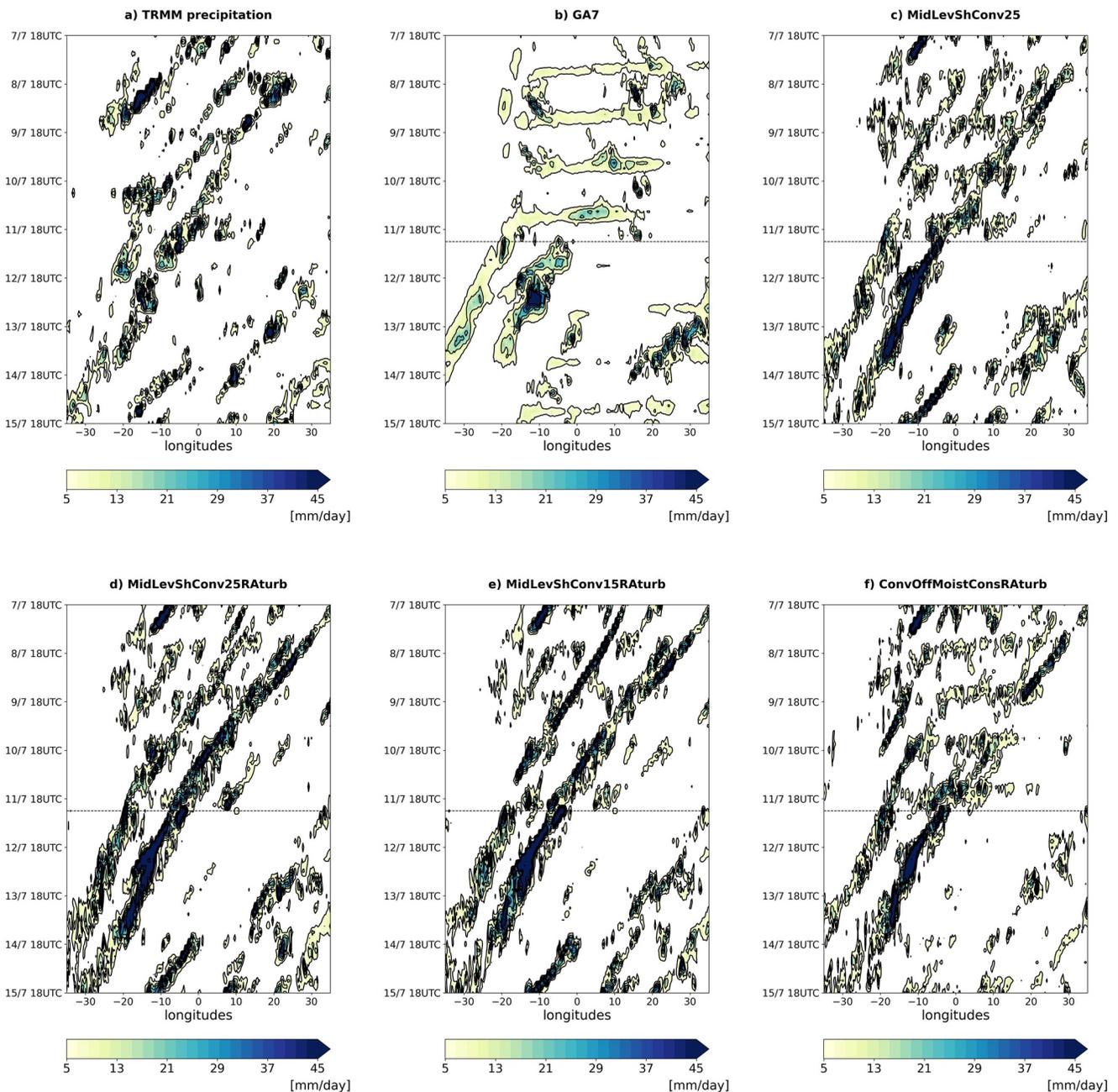
Figure 1. Hovmöller plots of three-hourly mean precipitation rate for the African easterly wave case. (a) Tropical Rainfall Measuring Mission (TRMM) observations; (b) *GA7*; (c) *MidLevShConv25*; (d) *MidLevShConv25RAturb*; (e) *MidLevShConv15RAturb*; (f) *ConvOffMoistConsRAturb*. The model data are interpolated to the TRMM grid using an area-weighted regridding scheme. The data encompass the longitudes 30W–30E and are averaged over the latitudes 10N–20N. The dashed line indicates the time from which on the data of the second hindcasts are used; this time is 24 hr after hindcast initialization.

For *ConvOffMoistConsRAturb* (Figures 2 and 3, bottom right panel), the convective systems are too scattered and widespread across the African land area, a feature that is evident even more clearly in the precipitation field (not shown) and will also be discussed in another case study (Section 3.4) for a tropical ocean region below. Moreover, clouds and rainfall are not very sensitive to the presence of the wave and do not show a distinct response to the wave circulation around the trough in this configuration.

In order to better understand the representation of convective systems over the wider Africa-Atlantic region, convective storms are tracked in the area 40W–40E and 25S–25N over the period from 8 July to 15 July 2010, both for the different model configurations and Global Precipitation Measurement (GPM) observations (Huffman
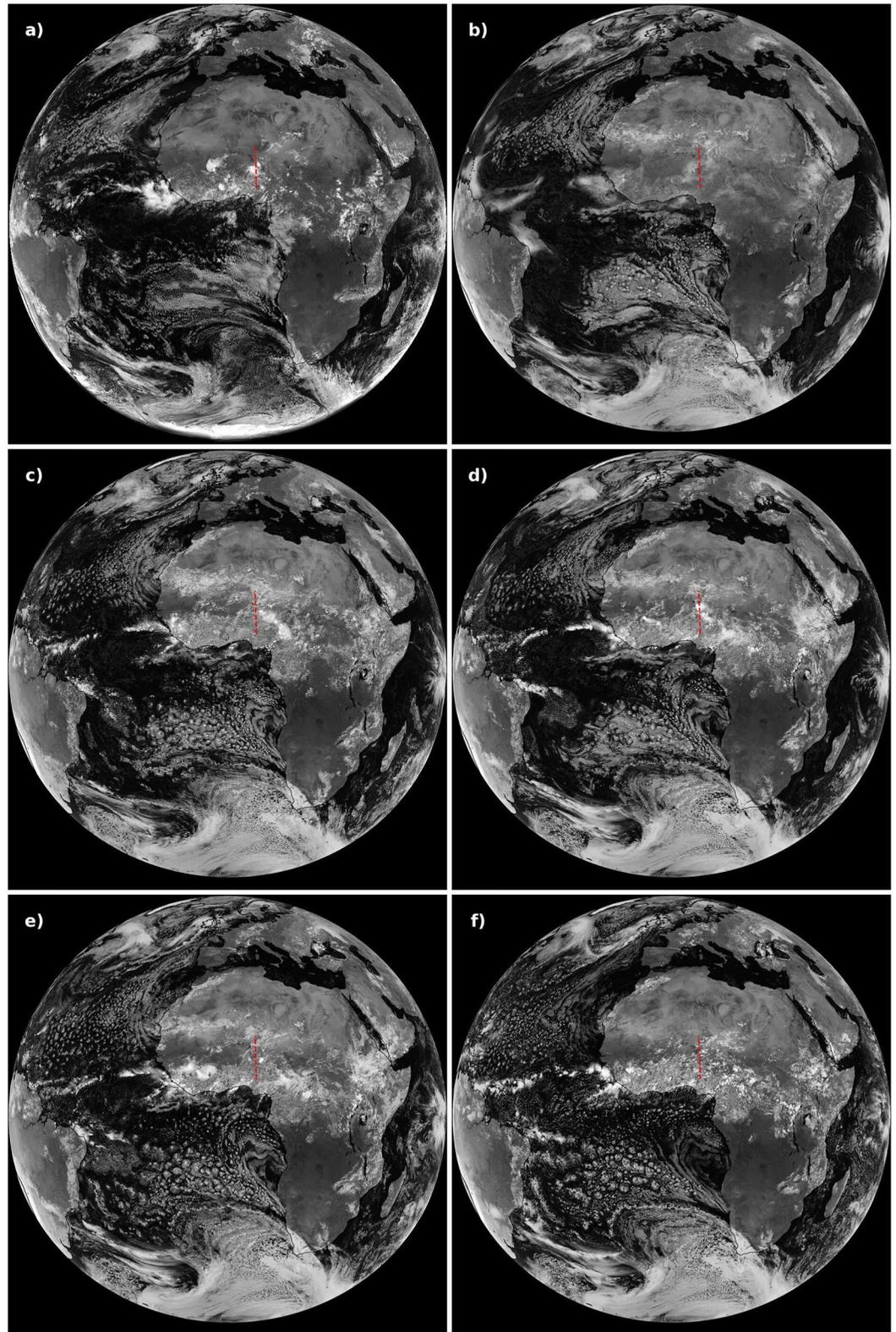
Figure 2. Observed (top left) and simulated reflectance at 0.8 μm for the MSG satellite for 10 July 2010, 12:00 UTC, based on the same configurations as in Figure 1. (a) Observed MSG satellite image; (b) *GA7*; (c) *MidLevShConv25*; (d) *MidLevShConv25RAturb*; (e) *MidLevShConv15RAturb*; (f) *ConvOffMoistConsRAturb*. The dashed red lines in the plots indicate the approximate position of the wave trough as diagnosed from the European Center for Medium-Range Weather Forecasts (ECMWF) analysis.
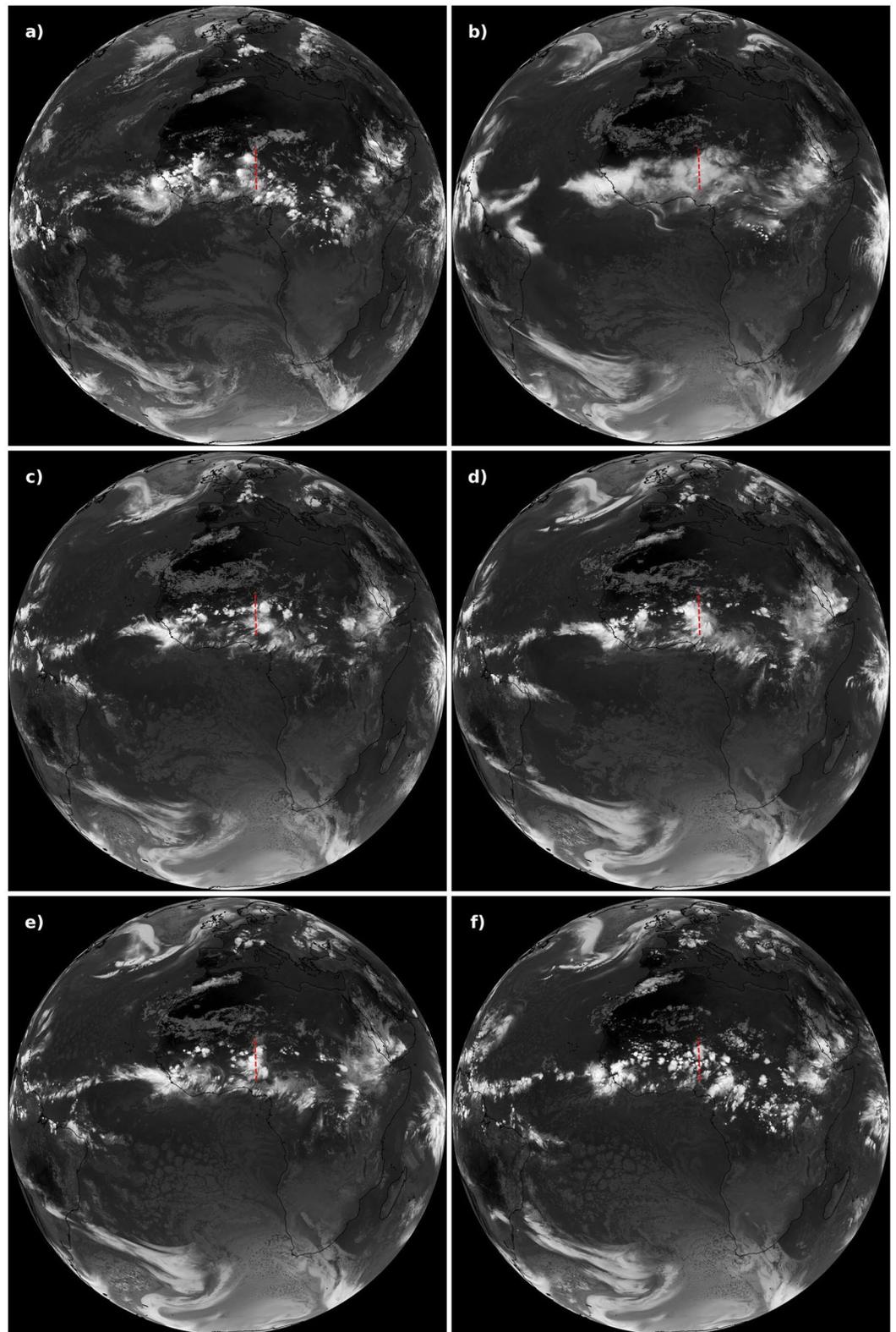
Figure 3. Observed (top left) and simulated brightness temperature at 10.8 μm for the MSG satellite for 10 July 2010, 18:00 UTC, based on the same configurations as in Figure 1. (a) Observed MSG satellite image; (b) *GA7*; (c) *MidLevShConv25*; (d) *MidLevShConv25RAturb*; Panel (e) *MidLevShConv15RAturb*; (f) *ConvOffMoistConsRAturb*. The dashed red lines in the plots indicate the approximate position of the wave trough as diagnosed from the European Centerre for Medium-Range Weather Forecasts analysis.
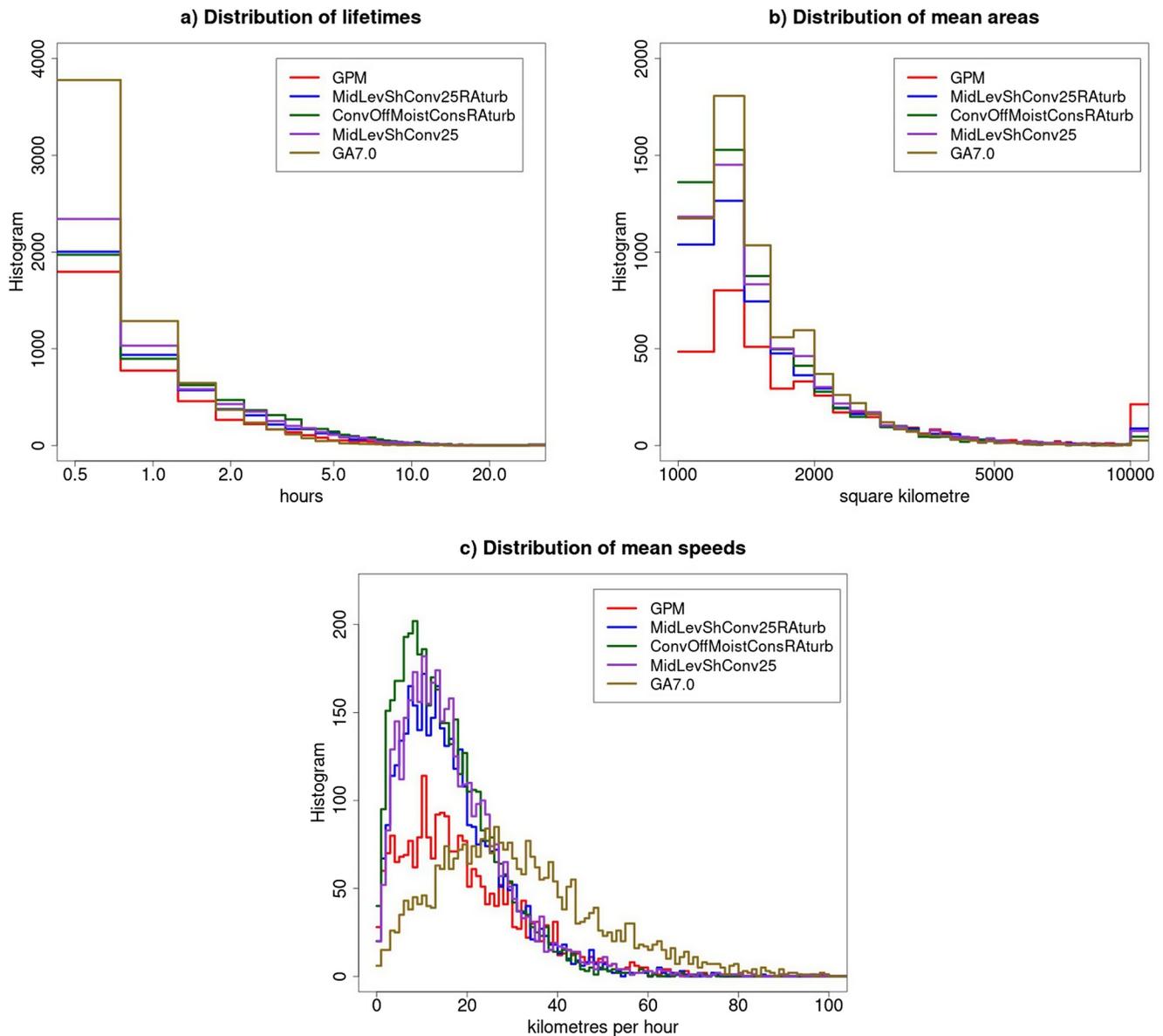
Figure 4. Results of convective storm tracking over the Africa-Atlantic region (40W–40E, 25S–25N) and the period 8 July to 15 July 2010. (Top left panel) histograms of convective storms life times; (top right panel) histograms of convective storms mean areas; (bottom panel) histograms of convective storms mean speeds. The model data are interpolated to the Global Precipitation Measurement grid using an area-weighted regridding scheme before the storm tracking is performed.

et al., 2019). The tracking is based on half-hourly precipitation data, a threshold of 3 mm/hr is chosen to identify convective systems, and an overlap of 50% is required from time stamp to time stamp for propagating systems (Crook et al., 2019; Stein et al., 2014). The histogram of convective storms lifetimes (Figure 4a) indicates that there are overall too many convective storms in all considered model configurations compared to GPM. *GA7* overestimates the number of short-lived storms and underestimates the number of long-lived storms. For the other convection-permitting configurations, the number of long-lived convective storms tends to be overestimated. For the mean size (over the storm life cycle) all models exhibit too many small storms, and too few very large storms compared to GPM (Figure 4b). This holds true for *GA7* in particular. For the number of medium-sized storms it is hard to make a robust statement. The distributions of mean speeds (over the storm life cycle) suggest that the storms in *GA7* move too fast compared to GPM. In the convection-permitting configurations the storms tend to move too slowly, most distinctly so in *ConvOffMoistConsRAturb*. The distributions confirm the visual impression obtained from the Hovmöller plots (Figure 1) that some storms move slightly slower in *MidLevShConv25* than in
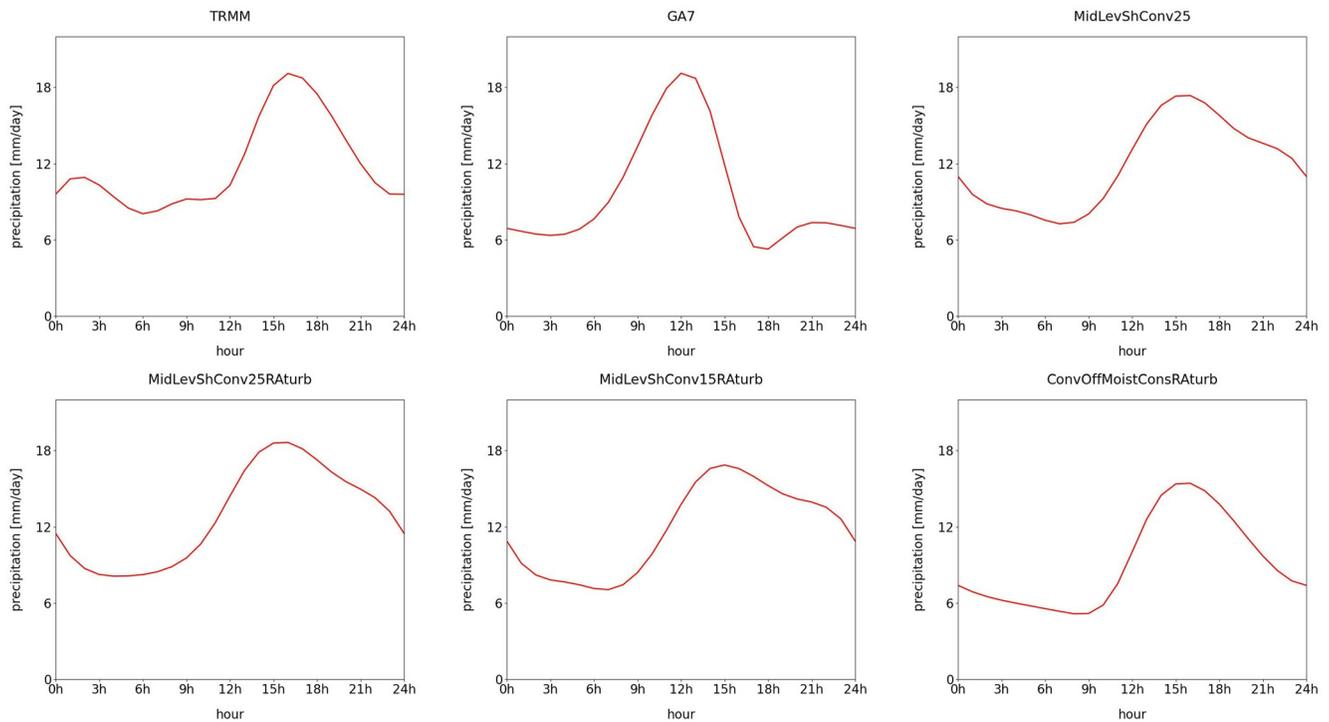
Figure 5. Diurnal cycle of precipitation for 7 July–15 July 2010, averaged over the region 1E to 11E and 5N–10N. The arrangement of the plots is as in Figure 1 with the top left panel showing Tropical Rainfall Measuring Mission observations.

MidLevShConv25RAturb. It is to be expected that the characteristics of storms in GPM will have biases over the region, particularly with regard to their areas, and the results should be interpreted with some caution (Rajagopal et al., 2021).

The diurnal cycle of precipitation over the Sahel depends on the exact region considered, and it is not possible to assess the issue very robustly based on just two 5-day forecasts. Nevertheless, choosing the region 1E–11E and 5N–10N over the period from 7 July to 15 July 2010, the picture that emerges is plausible (Figure 5). According to the analysis *GA7* shows an erroneous, well-known peak in precipitation around local noon and too little rainfall during night time. The convection-permitting model configurations fare much better and exhibit an afternoon maximum in agreement with TRMM. In the *MidLev* configurations the rainfall persists somewhat too strongly during the evening, whereas in *ConvOffMoistConsRAturb* rainfall is underestimated during the night. Note that mean rainfall over the course of the day is not generally overpredicted in the model simulations over the area and in fairly good agreement with the observations. This is in contrast to some regions over the tropical oceans where the convection-permitting configurations overestimate mean rainfall quite significantly (see Section 5.1 below).

That rainfall is not generally overestimated by the convection-permitting configurations is confirmed also when it comes to the simulation of heavy rainfall over Africa. Box plots for both TRMM observations and model configurations of 3-hourly mean rainfall rates above the threshold of 30 mm/hr over the region 20W to 45E and 35S to 35N, comprising the whole of the African continent, and the period 7 July–15 July 2010, reveal that the convection-permitting configurations do not overestimate heavy rainfall events over Africa on the 3-hourly time scale (Figure 6). On the contrary, they tend to underestimate extremes compared to TRMM. The box plots show the lower, middle, and upper quartile of the data above the threshold, and the length of the whiskers are 1.5 times the interquartile range. *GA7* does not have any data points above the threshold when considering 3-hourly mean rainfall rates.

In summary, *GA7* does not simulate very realistic mesoscale convective systems in the tropics and the African easterly wave is not well predicted. The configuration *ConvOffMoistConsRAturb* does not represent the interaction between convection and the wave circulation well, the organized convective structures are not very sensitive to the wave trough and the rainfall is too scattered across the wider area. The coupling between convection and
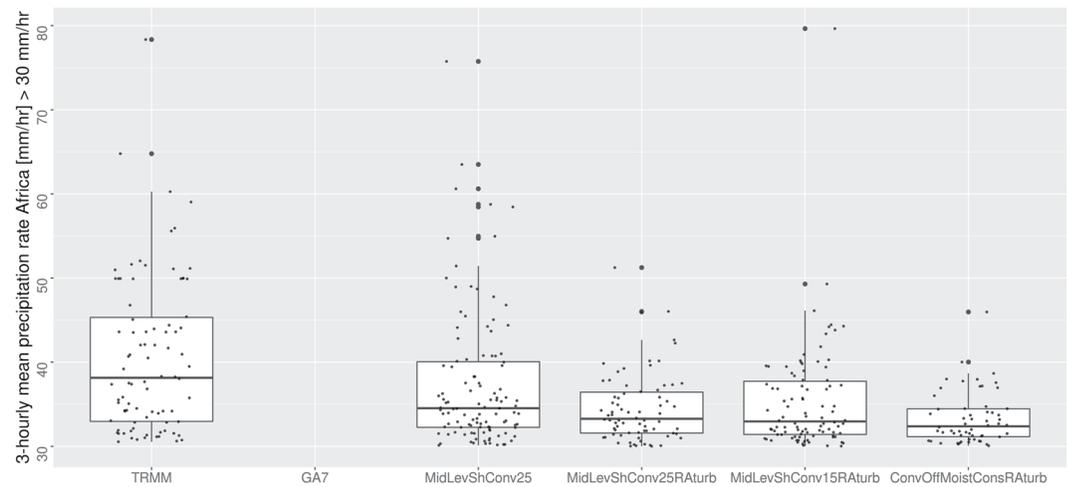
Figure 6. Boxplot of Tropical Rainfall Measuring Mission observations and model simulations of 3-hourly mean rainfall rates above the threshold of 30 mm/hr over the region 20W to 45E and 35S to 35N, comprising the whole of the African continent, and the period 7 July–15 July 2010. The box plots show the lower, middle, and upper quartile of the data above the threshold, and the length of the whiskers are 1.5 times the interquartile range. Individual values are indicated by dots, the points outside the whiskers are also shown as bold dots.

circulation is a key aspect in the context of many phenomena in the tropics. For *MidLevShConv25*, the mesoscale convective systems seem to move a bit too slowly over the African continent. This can be important as it might imply that the interaction between convection and tropical waves is not particularly well captured. The relative location of the convective systems with regard to the circulation features of the waves is an important aspect in tropical wave dynamics. The model configurations *MidLevShConv25RAturb* and *MidLevShConv15RAturb* show the best convection-circulation coupling, speed of convective systems, and degree of convective organization over the African region in the analyses presented in this section. Impressively, the respective simulations can predict the evolution of some individual convective systems several days ahead. And it is these convective systems that impact the livelihoods of people in the area most.

3.2. Hurricane Dorian and Typhoon Goni

The simulation and prediction of tropical cyclones are important applications of weather forecasting because of the devastation these phenomena can cause. Furthermore, changes in tropical cyclones under a warming climate may deliver some of the most significant impacts. Here we present results of forecasts of two tropical cyclones, hurricane Dorian initialized on 30 August 2019, 12:00 UTC, and typhoon Goni initialized on 28 October 2020, at 12:00 UTC. The runs use ECMWF operational analyses as initial conditions except the operational model shown in Figures 7 and 8. The then operational model is GA6.1 (Walters et al., 2017) in the case of Dorian, and GA7.2 in the case of Goni, both run at 10 km resolution and using Met Office analyses as initial conditions.

The predictions of central pressure in the simulations of tropical cyclones Dorian and Goni are significantly improved in the *MidLev* model configurations compared to the operational models, *GA7* at 5 km resolution, and also *ConvOffMoistConsRAturb* (Figure 7). In particular, spells of rapid intensifications for both tropical cyclones are well captured by the *MidLev* model configurations.

The simulations of tropical cyclone tracks, however, are quite consistently degraded in the 5 km-resolution simulations compared to the operational models (Figure 8 in the case of Goni, but also in results that are not shown here). This is true, though, also for *GA7* at 5 km resolution. We do not expect this to be due to the different initial conditions. The reason is not clear at this point and might be related to subtle degradations in regional synoptic scale conditions that affect the steering flow rather than the representation of the local cyclone processes themselves.
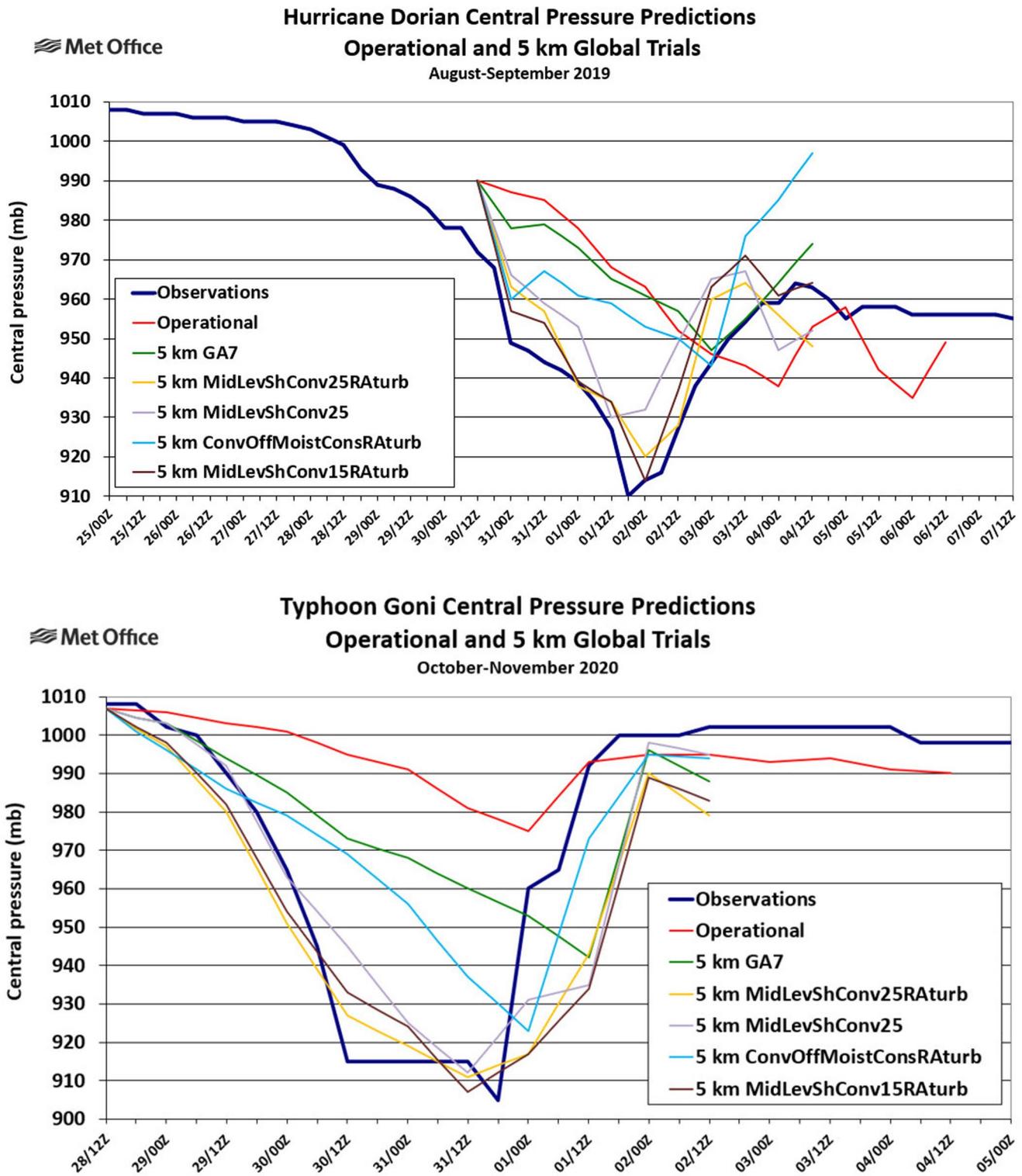
Figure 7. Observed and modeled central pressure developments for tropical cyclones Dorian (top panel) and Goni (bottom panel). The operational model is GA6 in the case of Dorian and GA7 in the case of Goni, both run with fully parameterized convection at 10 km resolution. The observations are from the IBTrACS data base of the National Oceanic and Atmospheric Administration National Climate Data Center.
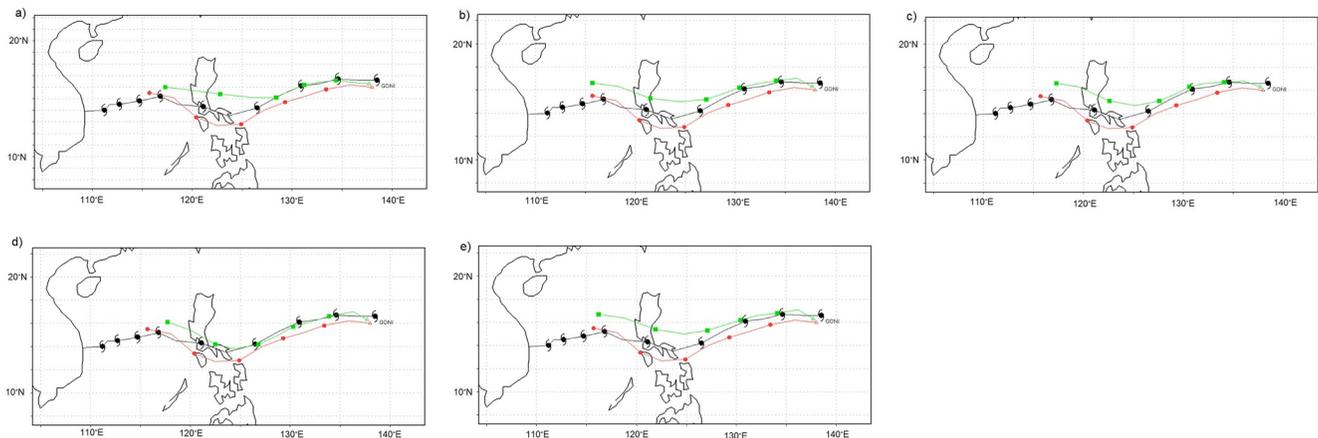
Figure 8. Cyclone track predictions for tropical cyclone Goni in the vicinity of the Philippines. The operational model (OP; red lines) is *GA7* run with fully parameterized convection at 10 km resolution. Black lines are observations and green lines the different 5 km-resolution configurations. The observations are from the IBTrACS data base of the National Oceanic and Atmospheric Administration National Climate Data Center. (a) *GA7*; (b) *MidLevShConv25RAturb*; (c) *MidLevShConv25*; (d) *ConvOffMoistConsRAturb*; (e) *MidLevShConv15RAturb*.

3.3. EUREC⁴A Case Over the Atlantic Trades

One reason why tropical cyclones tracks are not particularly well simulated by the 5 km-resolution models might be related to the structure of the atmosphere in the subtropics which influences the steering flow in the surroundings of the cyclone center. To get a sense of this issue and to investigate shallow cumulus clouds over the trade wind region, a topic of particular importance in climate change studies, a case during the EUREC⁴A field campaign is considered (Stevens et al., 2021). The models are initialized on 1 February 2020, at 00:00 UTC with ECMWF ERA5 reanalysis (Hersbach et al., 2020), and evaluated on February 2.

The GOES16 satellite observed and simulated reflectance demonstrates how trade wind cumulus clouds can be controlled in the model by scaling the shallow convective subgrid mass flux (Figure 9). A reduced subgrid mass flux leads to more, and brighter, shallow clouds. A difficulty here is that the structure and organization of the clouds can thereby get more realistic but the cloud amount or the optical thickness of the clouds might in turn become biased. In the particular snapshot shown in Figure 9 low cloud cover is distinctly underestimated in the *GA7* configuration, especially over the Pacific (Figure 9, top right panel). However, lower-resolution versions of the *GA7* model have been tuned to produce realistic global-mean fluxes, whereas in radiation budget estimates presented in Section 5 *MidLevShConv25RAturb* exhibits a distinctly overpredicted global-mean top-of-the-atmosphere outgoing shortwave radiative flux in the absence of tuning. This shows how seemingly more realistic process representation does not always translate into better agreement with observations, at least not in every respect. Figure 10 shows the same as Figure 9 but zooms in on the subtropical Atlantic of the northern hemisphere which includes the area of the EUREC⁴A field campaign. *MidLevShConv15RAturb* and *ConvOffMoistConsRAturb* manage to reproduce distinct cloud streets and some larger organized low cloud clusters ("flower" clouds) over the ocean in the southwestern part of the domain similar to the GOES16 satellite image scene. Overall the two configurations overestimate cloud cover, an aspect that is, however, also linked to the parameterization of cloud macrophysics and microphysics, an avenue not explored here.

Observed profiles from HALO research aircraft dropsondes sampled over the EUREC⁴A campaign observation area east of Barbados over a time window of 17:50 to 18:10 UTC (George et al., 2021) show that the increased trade wind cloudiness in the *MidLevShConv15RAturb* configuration is associated with a cooler and moister lower troposphere (Figure 11). The overall structure of the lower troposphere is quite well represented, including by *GA7*. The slight cold bias in the subtropical trade wind cloud layer is consistent with anecdotal evidence from other forecasts and might at least partially be due to the somewhat overestimated cloud amount or optical thickness in the lower troposphere of the subtropics. The most notable feature is a dry bias in the subcloud layer (i.e., the part of the boundary layer over which specific humidity is fairly well mixed; see Section 3.3 and Figure 13 in Stevens et al. (2021) for more details regarding the boundary layer structure) of the models which is present in all configurations despite the different turbulence schemes used. The moisture
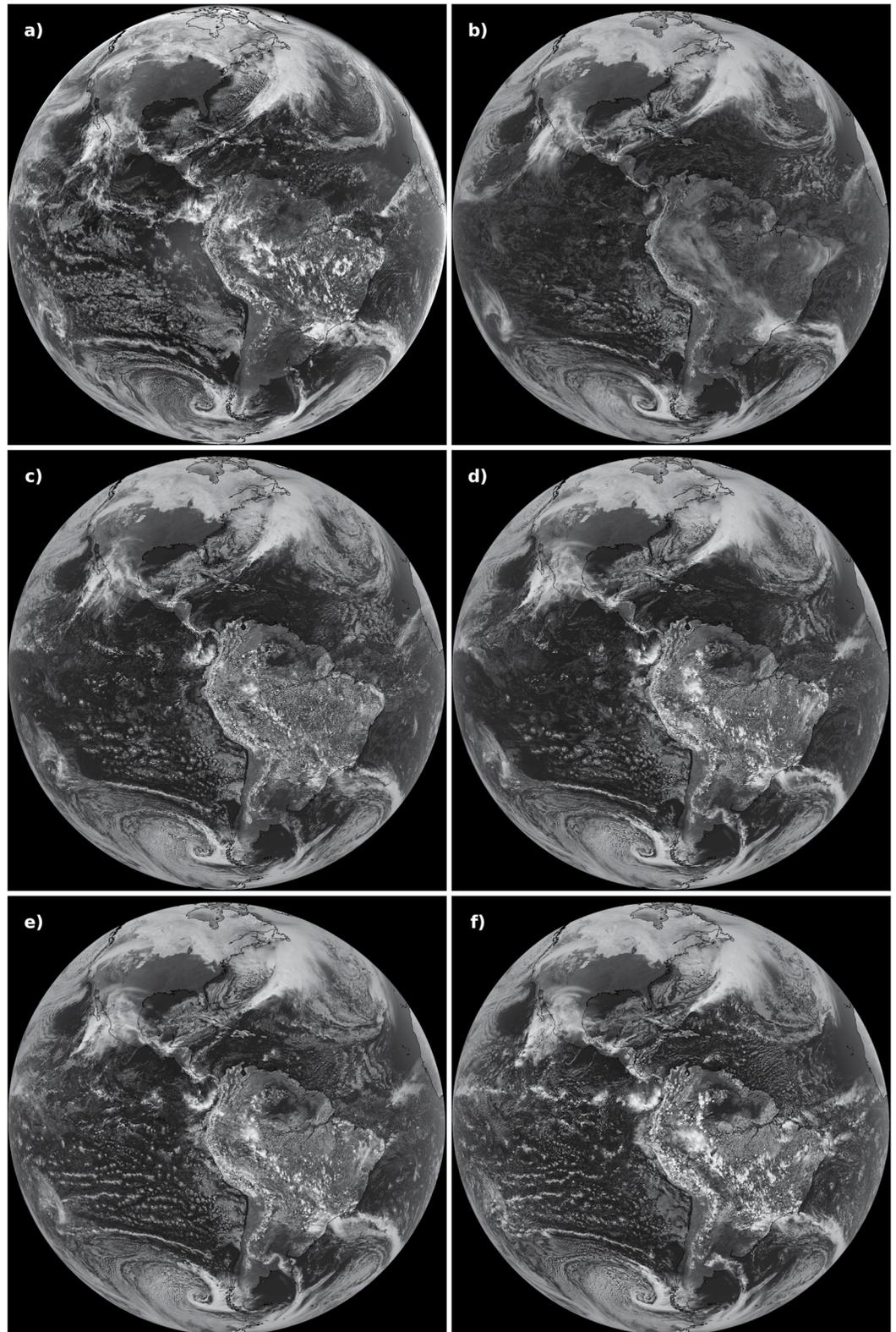
Figure 9. Observed (top left) and simulated reflectance at 0.64 μm for the GOES16 satellite for 2 February 2020, 18:00 UTC, based on the same configurations as in Figure 2, presented in the same order: (a) Observed GOES16 satellite image; (b) *GA7*; (c) *MidLevShConv25*; (d) *MidLevShConv25RAturb*; (e) *MidLevShConv15RAturb*; (f) *ConvOffMoistConsRAturb*.
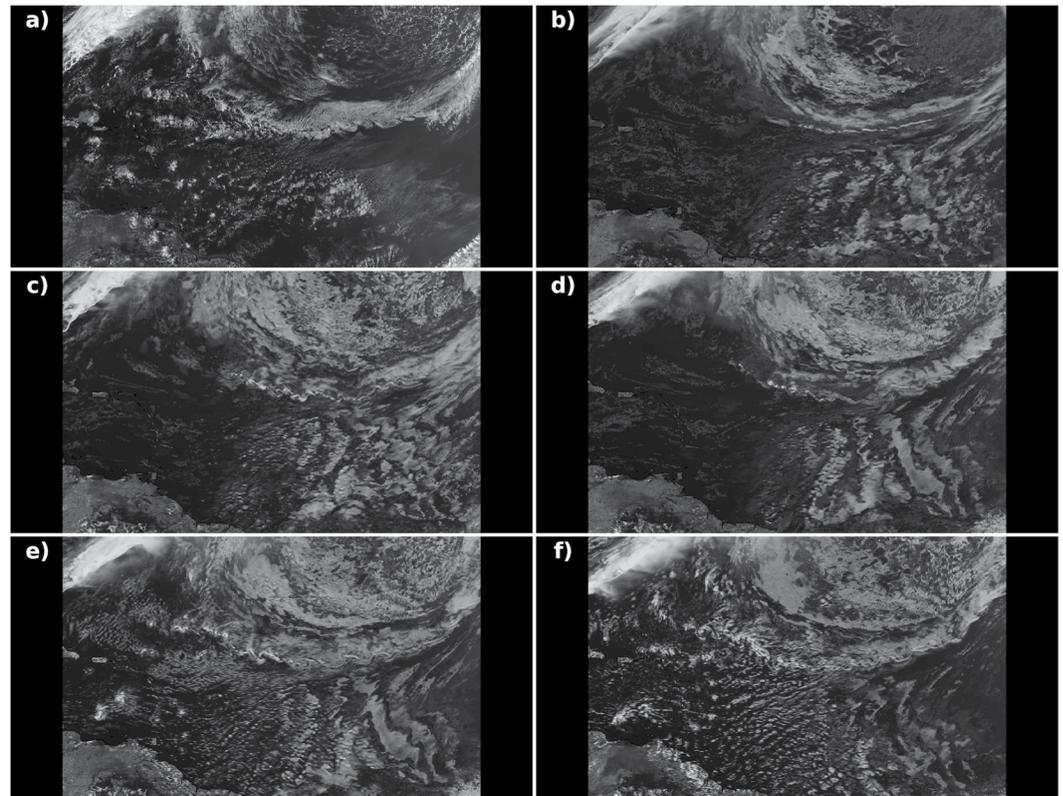
Figure 10. Observed (top left) and simulated reflectance at 0.64 μm for the GOES16 satellite for 2 February 2020, 18:00 UTC, based on the same configurations as in Figure 9, presented in the same order: (a) Observed GOES16 satellite image; (b) *GA7*; (c) *MidLevShConv25*; (d) *MidLevShConv25RAturb*; (e) *MidLevShConv15RAturb*; (f) *ConvOffMoistConsRAturb*. The depicted area is 70W–30W and 5N–30N which includes the area of the EUREC⁴A field campaign.
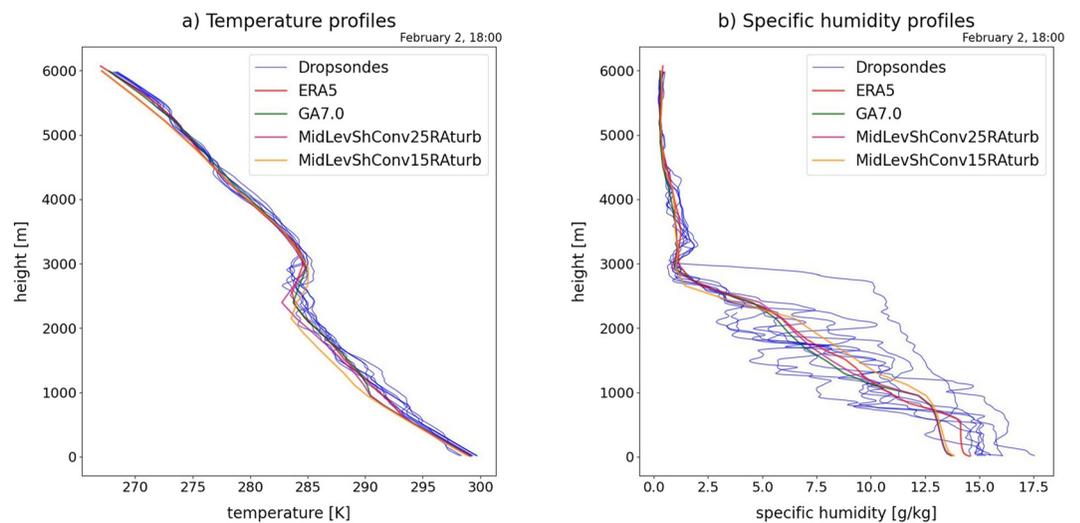

Figure 11. Temperature and specific humidity profiles for HALO aircraft dropsondes, ERA5, *GA7*, *MidLevShConv25RAturb*, and *MidLevShConv15RAturb* on 2 February 2020, 18:00 UTC, averaged over the EUREC⁴A observation circle east of Barbados (Stevens et al., 2021). The center of the observation circle is at 13.3N and 57.72W, the radius is 110 km. The observed dropsonde profiles are sampled between 17:50 UTC and 18:10 UTC.
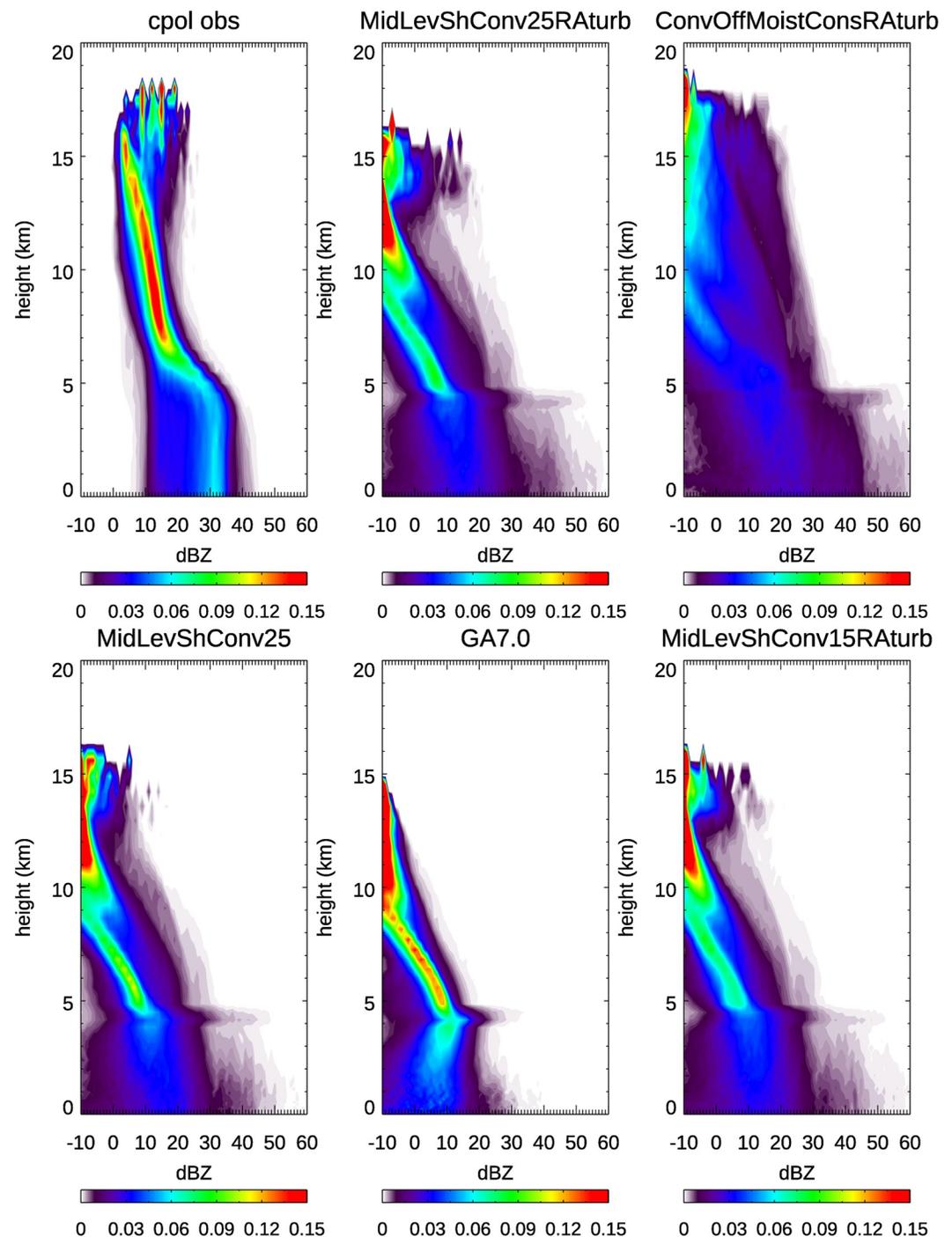
Figure 12. Observed (top left) and simulated radar reflectivity profiles (18 February, 12–24 UTC mean) over the area within reach of the Darwin C-band radar. Color shaded contours represent the frequency of occurrence.

profile in the cloud layer is probably most realistic in *MidLevShConv15RAturb* while *GA7* tends to be too dry. The case shows quite clearly how none of the configurations is perfect, and a subjective, informed judgment has to be made when assessing the different model configurations with regard to various aspects and features of the atmosphere.

3.4. Darwin Mesoscale Convective Systems Case

To conclude the analysis of tropical and subtropical cases, advantage is taken of the C-band radar at Darwin on the northern tip of Australia. The models are initialized on 17 February 2014, at 00:00 UTC with ECMWF operational analysis (Franklin et al., 2016).

All models show substantial deficiencies in the profiles of radar reflectivities (Figure 12). As with the simulated satellite imagery previously, the simulated radar reflectivity does not include cloud liquid and ice that are held within the convection scheme, somewhat penalizing the model with parameterized convection (*GA7*). Most model configurations tend to underestimate radar reflectivity at upper levels. The configuration with convection switched off, *ConvOffMoistConsRAturb*, reaches highest, but seems to overestimate high radar relectivities at middle levels. All models simulate an unrealistically strong outflow around the freezing level at around 4 km height. The shortcomings of the models will at least partly be related to the representation of microphysics, an aspect that has not been considered in detail in the present study and could in principle be tuned for a better agreement with observations. The comparison makes the high-dimensional nature of the model development problem obvious, and the extent of the challenge that is due to the fact that only a limited number of sensitivity experiments and cases can be run and assessed with a high-resolution global model. The profiles in Figure 12 show an area mean over 12 hr, so part of the errors may also be caused by inaccuracies in the exact location of the convective systems and not only by their structure.

Despite the challenges in the assessment, there are robust features across the different tropical cases, for example, the structure of precipitation, here shown around Darwin (Figure 13). Our judgment is that the *MidLev* configurations overall exhibit the most realistic characteristics in terms of structure and organization (Figure 13d). The broader pattern of a precipitating band of stronger rainfall off the coast of Australia, oriented from west to east, with secondary centers of precipitation over some of the islands of the Maritime Continent, is best reproduced by the *MidLev* configurations. Moreover, based on evidence from the radar rainfall estimate (not shown), the GPM product likely overestimates the extent of the rainfall areas of the largest clusters (see also the discussion in Rajagopal et al. (2021)). And the model *ConvOffMoistConsRAturb* exhibits a too scattered and blobby rainfall field and misses lighter rainfall. *GA7* shows too widespread areas of light rainfall and very little precipitation over land. Similar conclusions can be drawn also when investigating other cases like the African easterly wave case (Section 3.1) over tropical land (not shown).

3.5. NAWDEX Cases: The Atlantic Extratropics

To understand the behavior of the convection-permitting model configurations in the midlatitudes, cases from the North Atlantic Waveguide and Downstream Impact Experiment (NAWDEX) field campaign (Schäfler et al., 2018) are examined. The results from two sets of forecasts are presented here. For one set the models are initialized on 22 September 2016, 00:00 UTC with ECMWF operational analysis, for the other set the initialization time is 5 October 2016, 00:00 UTC, again using ECMWF operational analysis.

The ex-tropical storm Karl reintensified on 26 September 2016, as it approached western Europe. In the subsequent development the jet stream was unusually strong on its southern flank, forming a jet streak that propagated ahead from Karl. The impact on European weather occurred through the formation of a new cyclone, Walpurga. Moisture-laden air was drawn around the Atlantic subtropical high. A particularly moist boundary layer was observed in this atmospheric river-type flow that extended to Norway, where it caused heavy, persistent rainfall (Schäfler et al., 2018).

A similar moisture transport pattern was also involved in the second case considered. On 5 October 2016, there was a high-pressure block over Europe. Two days later midlatitude cyclone Sanchez developed over the Atlantic. Nine days after forecast initialization, on 13 October, Sanchez caused heavy rainfall over southeastern France (Schäfler et al., 2018).

One could argue that the 5-day rainfall forecast for the September case is most accurate in the *GA7* configuration (Figure 14). Although the front-like structure in rainfall moved slightly too fast over northern Europe in *GA7*, the heavy rainfall over Norway is well reproduced. In the convection-permitting *MidLevShConv25RAturb* configuration, the main rainband has moved slightly too slowly, although some of the rainfall reaches the west coast
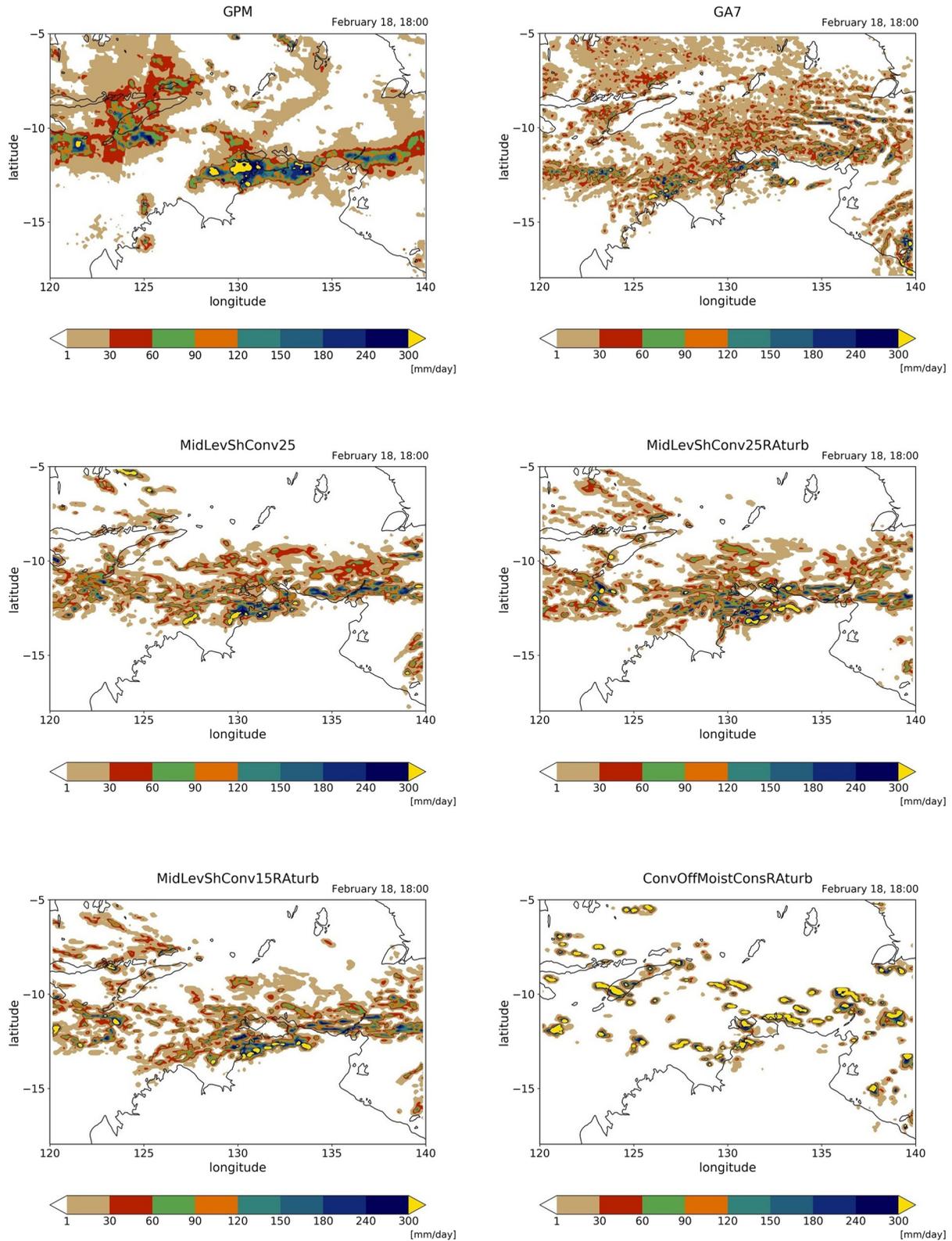
Figure 13. Precipitation rate for 18 February 2014, 18:00 UTC, over the wider Darwin region based on the same configurations as in Figure 2, presented in the same order. Top left is for Global Precipitation Measurement (GPM) observations. The model data are interpolated to the GPM grid using an area-weighted regridding scheme.

Figure 14. North Atlantic Waveguide and Downstream Impact Experiment case precipitation for 27 September 2016, 18:00 UTC, based on the same configurations as in Figure 13, presented in the same order. Top left is for Global Precipitation Measurement (GPM) observations. The model data are interpolated to the GPM grid using an area-weighted regridding scheme.
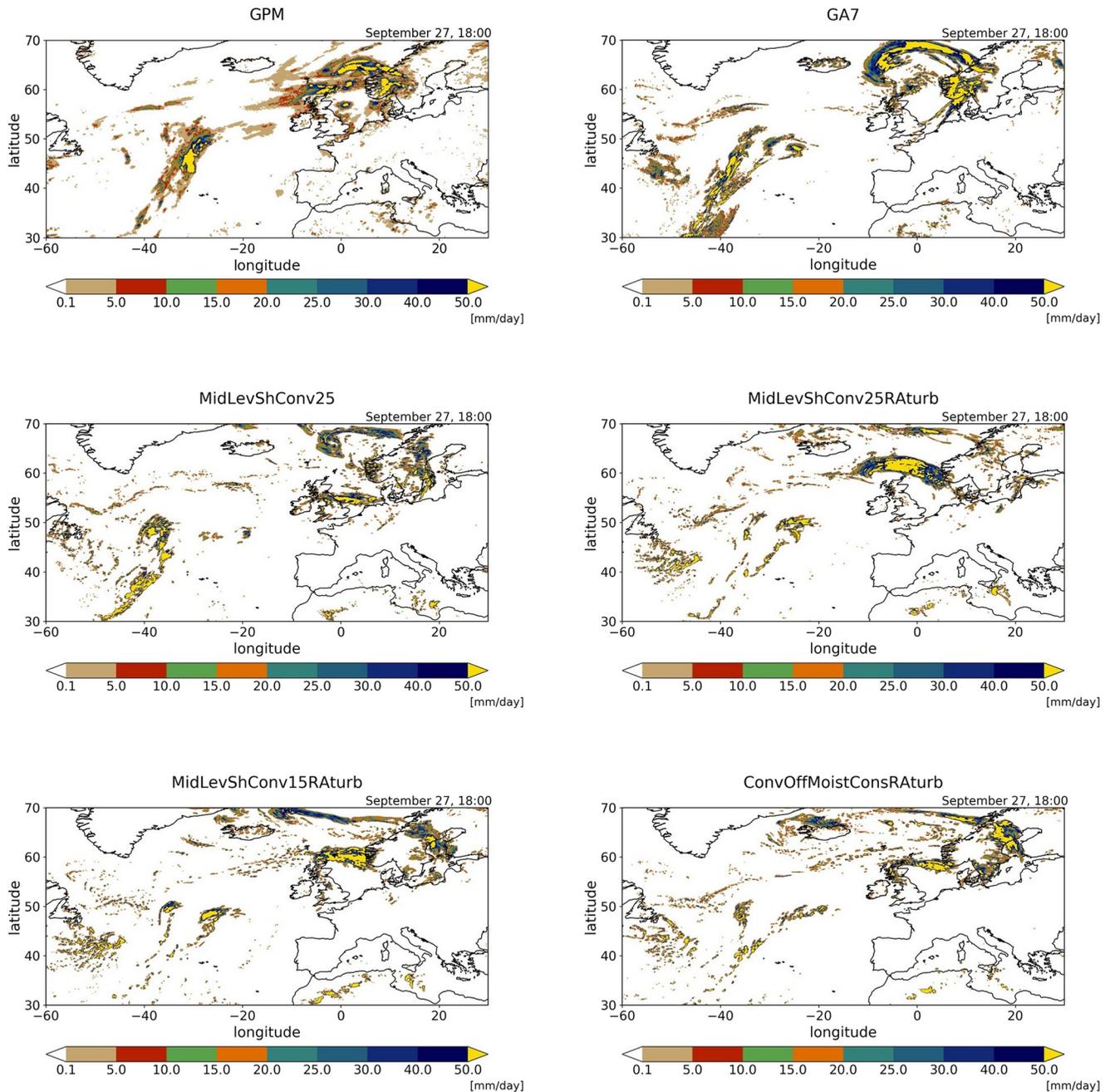
of Norway. *ConvOffMoistConsRAturb* shows more deficiencies and places the main precipitation area further inland.

Both *GA7* and *MidLevShConv25RAturb* reproduce the river-like moisture band from the subtropics toward Scandinavia in the daily-mean vertically integrated water vapor fields (Figure 15, left column for the September case). *MidLevShConv25RAturb* shows a stronger gradient in moisture in the northern part of Great Britain which could be responsible for the rainfall in this region. The differences in rainfall over the Norwegian coast seem to be related also to differences in the dynamical fields where *GA7* exhibits a stronger bend in the upper-level potential vorticity (PV) in the area (not shown).
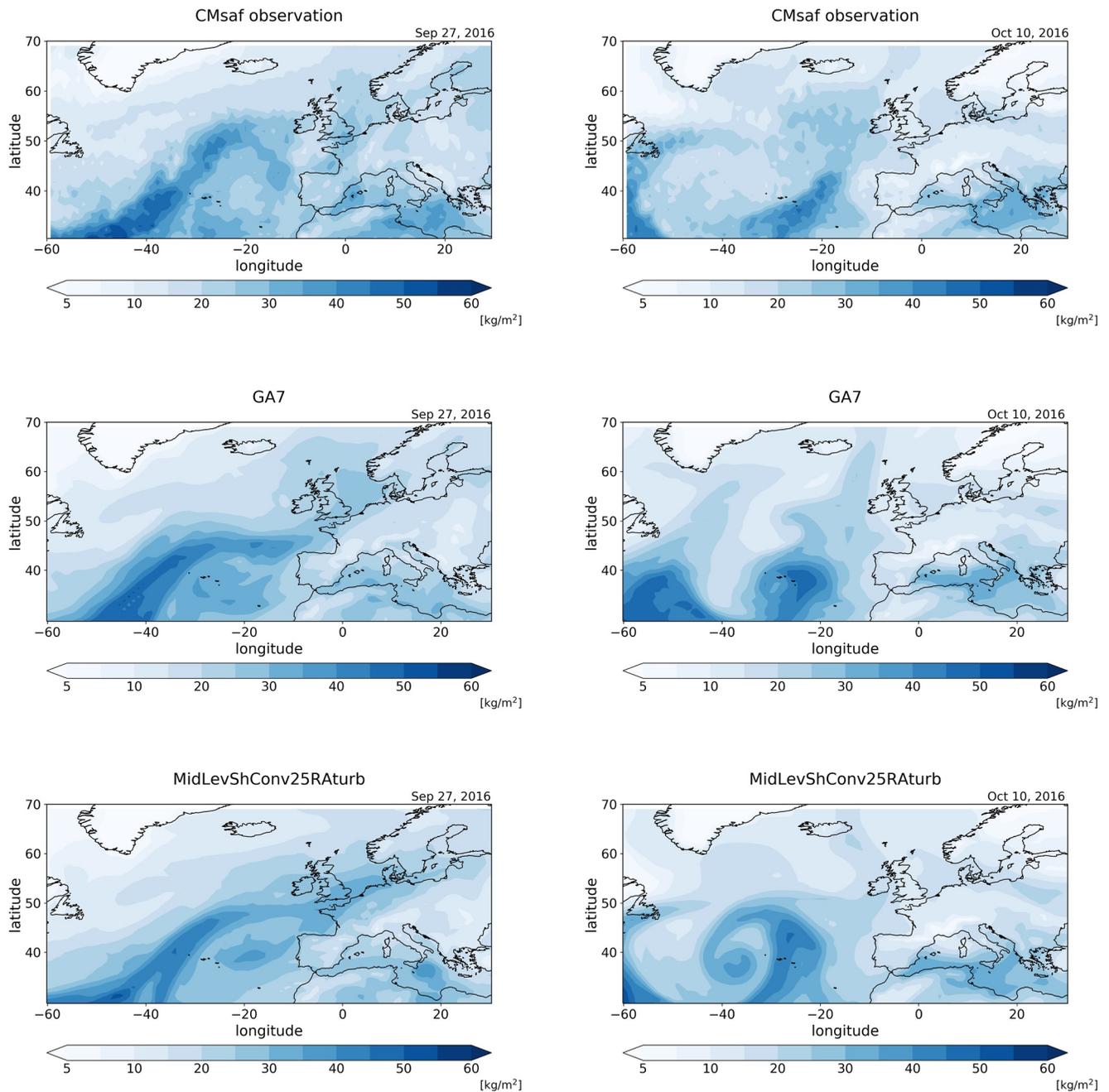
Figure 15. Daily mean integrated water vapor path for 27 September 2016 (left), and 10 October 2016 (right). (Top row) Satellite Application Facility on Climate Monitoring (CMsaf) observations. (Middle row) *GA7* configuration; (bottom row) *MidLevShConv25RAturb*. The model data are interpolated to the CMsaf grid using an area-weighted regridding scheme.

In the October case, *GA7* performs distinctly worse than *MidLevShConv25RAturb* in terms of rainfall prediction 9 days ahead after initialization (Figure 16). The rainfall field on October 13, when the heavy rainfall event over southern France occurred (red circles in the panels of Figure 16), shows that *GA7* moves the rainfall too quickly over the continent. Only the *MidLev* configurations predict the location of the heavy rainfall area within the indicated red circle correctly. Comparing the integrated water vapor fields on 10 October (Figure 15 right column) suggests that *MidLevShConv25RAturb* manages to draw the moisture further north toward the British Isles. However, it is not obvious whether this difference is key in this case. Inspection of the surface pressure fields of the different configurations and a comparison to ERA5 (Figure 17) confirms that the configurations
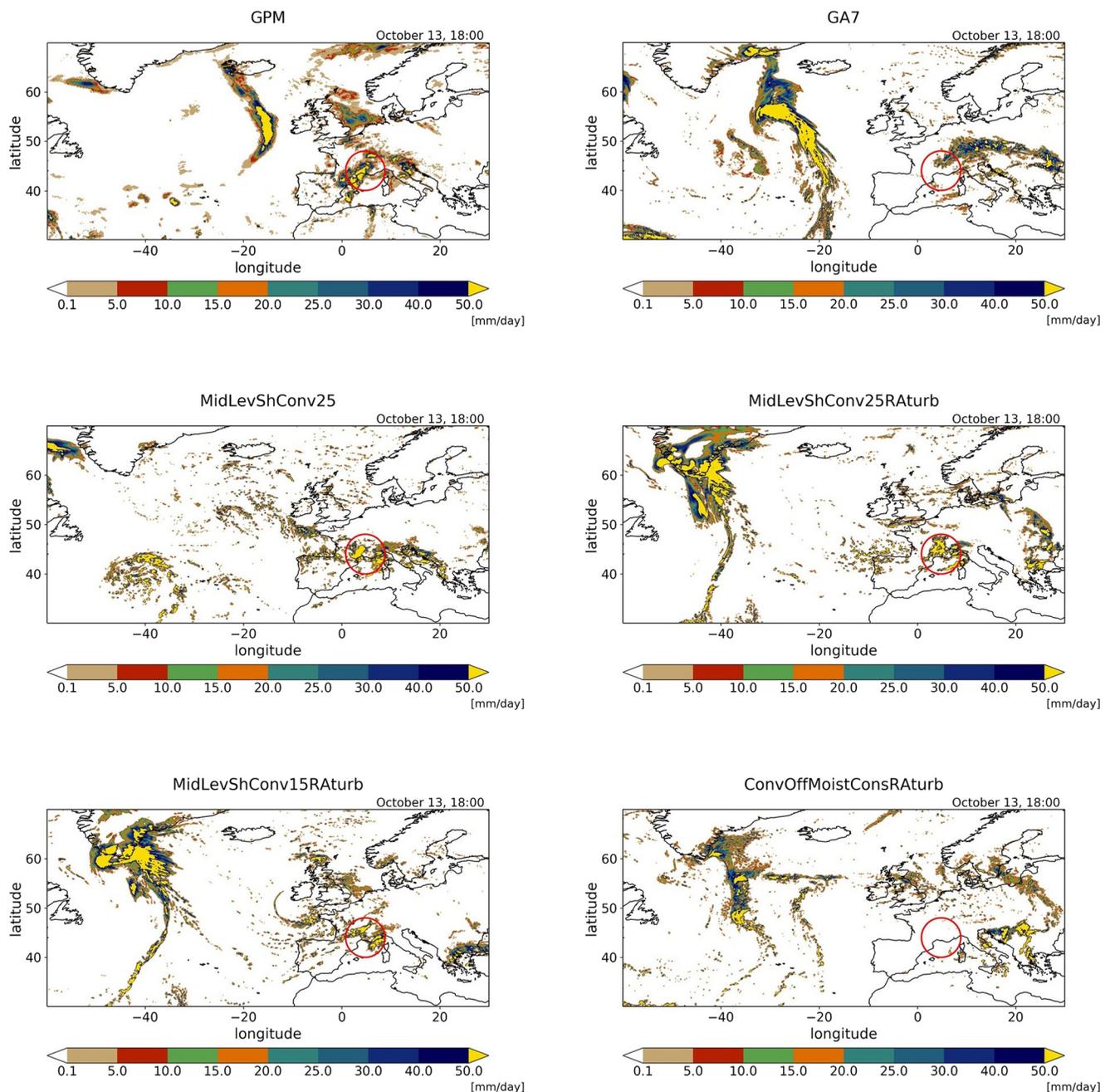
Figure 16. North Atlantic Waveguide and Downstream Impact Experiment case precipitation for 13 October 2016, 18:00 UTC, based on the same configurations as in Figure 14, presented in the same order. Top left is for Global Precipitation Measurement (GPM) observations. The red circle highlights the area of heavy rainfall in southern France. The model data are interpolated to the GPM grid using an area-weighted regridding scheme.

that predict the heavy rainfall over southern France accurately correctly exhibit a low-pressure signature over western Europe, in agreement with ERA5. This suggests that the quality of the rainfall prediction is indeed tied to the pressure distribution over western Europe on the synoptic scale, a more robust field. The exact reason for the superior performance of the *MidLev* configurations is, however, not obvious, and ensemble simulations of various extratropical cases would be required to better understand the physical processes underlying the behavior of the different model variants.

It is also interesting to examine the PV pattern at the 320 K isentrope in the different configurations (Figure 18). It would be difficult to derive the differences in rainfall prediction from the differences in the upper-level PV
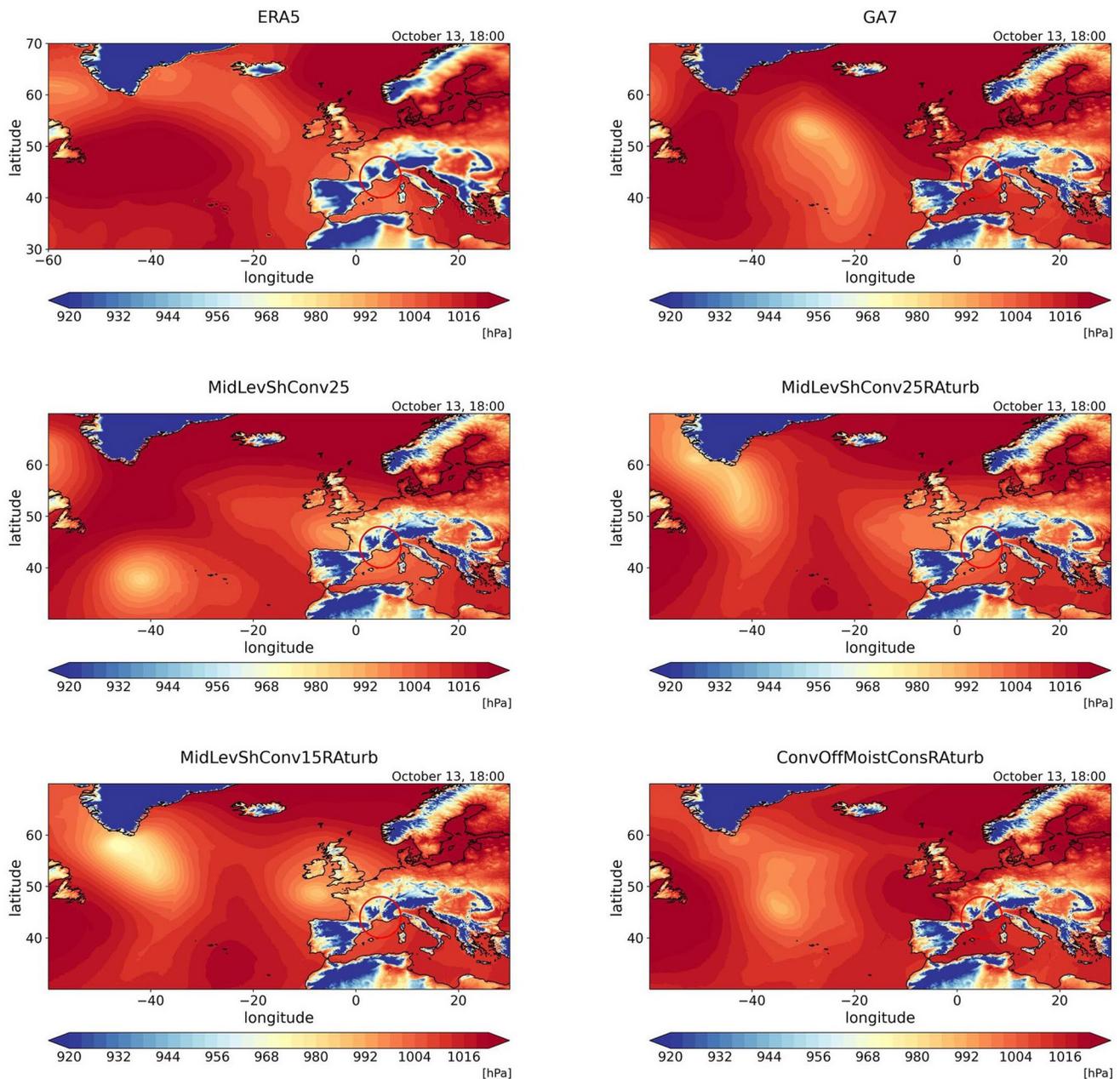
Figure 17. North Atlantic Waveguide and Downstream Impact Experiment case surface pressure for 13 October 2016, 18:00 UTC, based on the same configurations as in Figure 14, presented in the same order. Top left is for ERA5. The red circle highlights the area of heavy rainfall in southern France.

field. In particular, it is not clear whether the model with the smallest root-mean-square error, for example, in upper-level PV produces the most realistic rainfall prediction, or whether the position and bend of specific PV filaments are more important than any conventional statistical measure over the wider area. What is striking is how different the upper-level PV fields are in *MidLevShConv25RAturb* compared to *MidLevShConv25* (Figure 18, panels c and d). The two configurations differ only in the representation of turbulence, suggesting that the subgrid turbulence scheme can have a strong impact on upper-level dynamics. However, as noted before, the predictions of rainfall over Europe disagree less, although there are some differences.

Based on the two presented cases it is possible that the convection-permitting configurations tend to move the fronts related to extratropical cyclones somewhat too slowly compared to observations, perhaps due to increased cumulus friction or differences in meridional geopotential height gradients resulting in changes in geostrophic
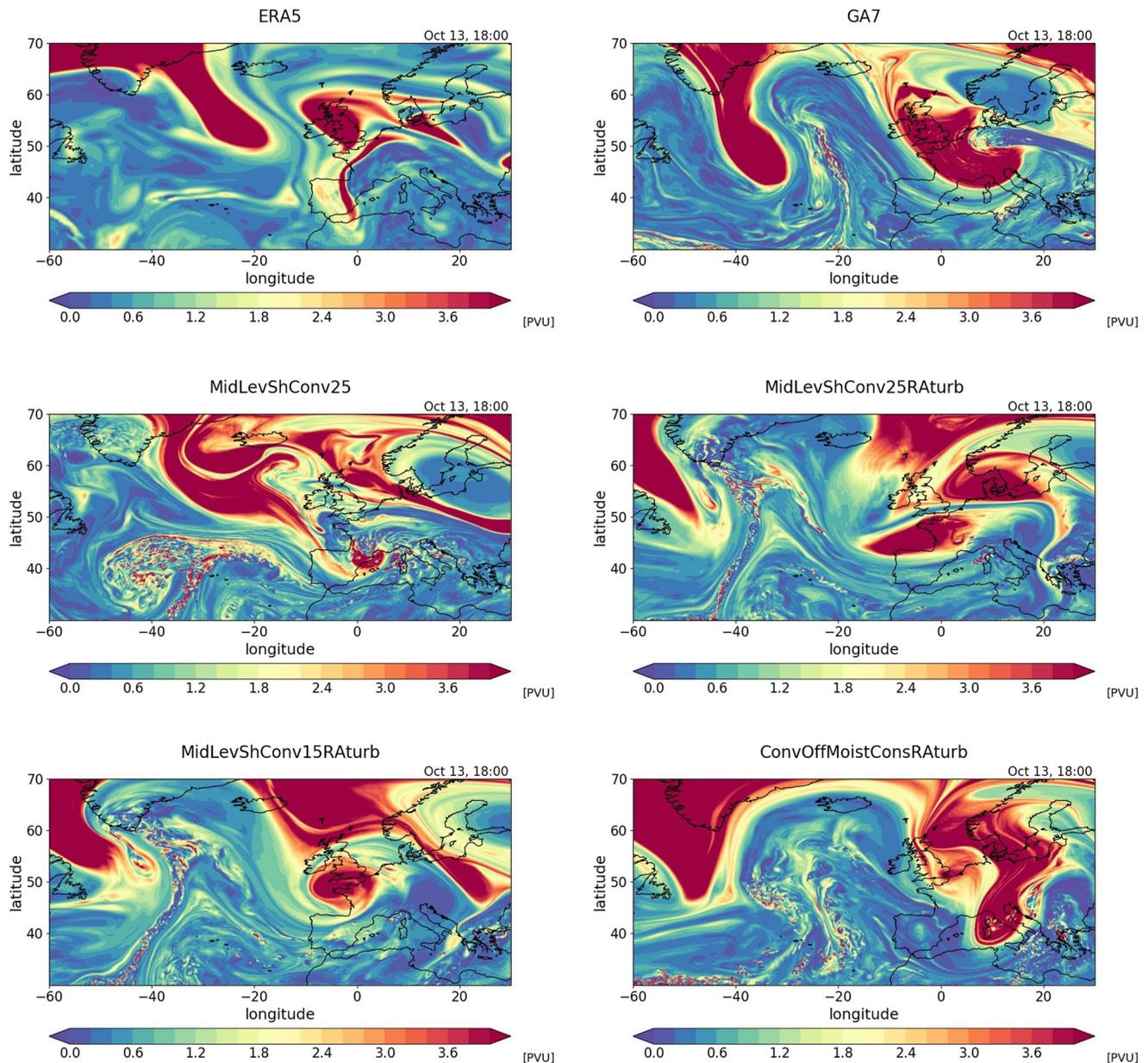
Figure 18. North Atlantic Waveguide and Downstream Impact Experiment case PV at 320 K for 13 October 2016, 18:00 UTC, based on the same configurations as in Figure 16, presented in the same order. Top left is for ERA5.

jet strength. However, this does not necessarily translate into worse performing rainfall forecasts compared to the model with parameterized convection. The key features in large-scale dynamical fields that lead to better rainfall prediction can be difficult to identify, and over the whole of the Atlantic area the differences in these fields are typically large in individual forecasts. This poses significant challenges in the assessment of the performance of the various configurations over the midlatitudes which will require a much more extensive set of statistics and a considerable ensemble size. Moreover, the results of such an assessment will depend on the considered quantity and the specific application.

4. Important Sensitivities and Additional Configuration Tests

A few additional sensitivity tests regarding model configuration options are briefly described here. They are generally based on a rather limited amount of evidence, but some of them concern important aspects directly
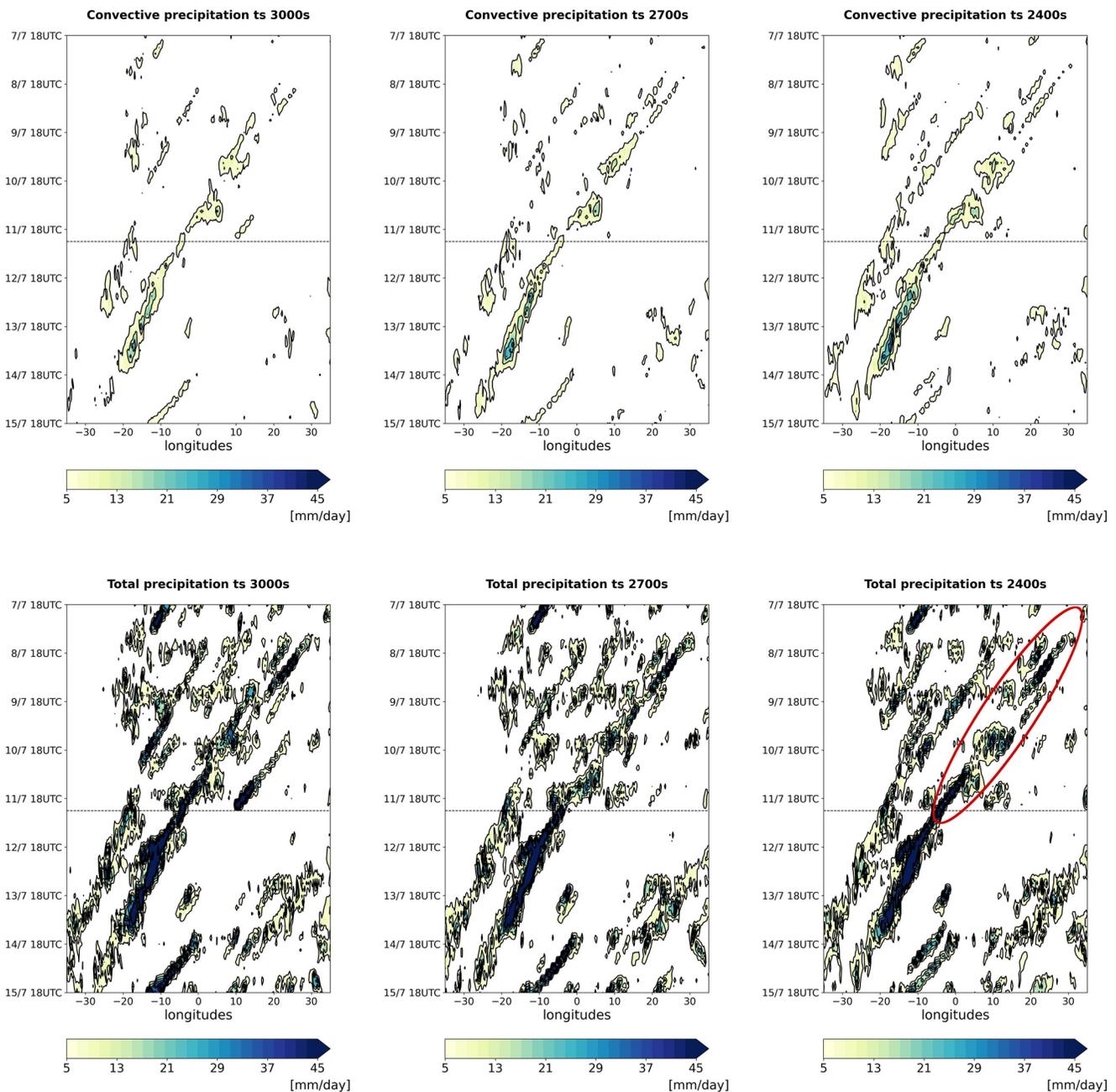
Figure 19. Model convective rainfall for *MidLevShConv25* with Convective Available Potential Energy (CAPE) time scale 3,000 s (top left), *MidLevShConv25* with CAPE time scale 2,700 s (top middle), and *MidLevShConv25* with CAPE time scale 2,400 s (top right). (Bottom row) Total precipitation for the same configurations. Although the convective rainfall is only moderately increased in the simulation with CAPE time scale 2,400 s compared to the other ones, the more active convection scheme starts to disrupt the prediction of the mesoscale convective systems during the early strengthening phase of the African easterly wave. The model data are interpolated to the TRMM grid using an area-weighted regridding scheme.

related to the convective gray zone problem and are not incidental. Not all sensitivity experiments that were conducted are presented and discussed, though, in this section.

4.1. CAPE Closure Time Scale

The CAPE closure time scale of the convection scheme in the *MidLev* configurations is a key parameter (see Table 1) that controls the amount of subgrid convective mass flux relative to the resolved, explicit vertical motion

and related rainfall in the model. Figure 19 shows the rainfall Hovmöller plots of total rainfall and convective rainfall for the configuration *MidLevShConv25* in the African easterly wave case discussed in Section 3.1 using three different CAPE closure time scales: 3,000, 2,700, and 2,400 s. One can see how the amount of convective rainfall increases when reducing the CAPE closure time scale. When the CAPE closure time scale is set to 2,400 s, the precipitation around the wave trough starts to break up in the early phase of the wave (area indicated by the red ellipse in bottom right panel of Figure 19), particularly between 9 July and 10 July, because the coupling between convection and wave circulation weakens. This trend continues when further shortening the CAPE closure time scale (not shown).

The effect of changing the CAPE closure time scale can also be identified in global mean values of total rainfall versus convective rainfall. For the *MidLevShConv25* configuration and a mean over the days 12 July–15 July (the second forecast of the African easterly wave case) the global mean of total precipitation is 3.689, 3.666, and 3.638 mm/day for the simulations with CAPE time scales 3,000, 2,700, and 2,400 s, respectively. In other words, there is relatively little change in global mean total rainfall. The global mean convective rainfall, however, increases from 0.909 to 1.032–1.178 mm/day, implying that the percentage of convective rainfall as part of the total rainfall increases from 24.7% to 28.1% to 32.5% when reducing the CAPE closure time scale.

4.2. Initial Perturbations in the Convection Scheme

It is common in convection schemes to use initial perturbations in order to represent subgrid variability, where some regions will be buoyant enough to trigger convection not achievable with the grid-box mean profiles, and this is also the case for the convection scheme used in the *MidLev* configurations. The original and default option is to apply perturbations in the temperature field, but in the deep convection scheme used in *GA7* humidity perturbations are added. Whether temperature or humidity perturbations are used in the *MidLev* configurations turns out to have quite a substantial effect. Using humidity perturbations makes the *MidLev* simulations look somewhat more similar to *GA7*, the model with fully parameterized convection (Figure 20). The buoyancy perturbation is fixed when shifting from temperature to humidity perturbations, but the perturbation in terms of moist static energy becomes somewhat larger, a circumstance that could also play a role.

In some of the diagnostics used in the African easterly wave case (Section 3.1), for example, one can see that the coupling between rainfall and wave becomes weaker in the early phase of the wave when applying humidity perturbations, and the clouds have a somewhat more blurred appearance compared to the simulations using temperature perturbations (Figure 20). A similar impression, namely that the simulations tend to become a bit more similar to *GA7* when using humidity perturbations, results also from the extratropical NAWDEX cases (Section 3.5, not shown).

4.3. Sensitivity to Turbulence Formulation and Miscellanea

Given that in Section 3.5 it was shown that the turbulence scheme has a substantial influence on the large-scale dynamical fields in the troposphere, two more turbulence options were tested and investigated for the NAWDEX cases on top of the *MidLevShConv25RAturb* configuration. In one test, instead of relaxing the mixing length toward the Smagorinsky length scale away from turbulent layers, a background mixing length scale of 40 m is used if the layer is stable or weakly turbulent. In another test a different stability function is used for unstable Richardson numbers, which produces significantly more mixing and which has previously been employed in the tropical version of the Unified Model regional configuration (Bush et al., 2022). In both tests the rainfall forecasts look somewhat degraded. For example, in the October case (see Section 3.5), the rainfall is placed mainly over the Alps in the first test, and mainly east of the Alps in the second test on day nine of the forecast and therefore showing similar deficiencies as *GA7* (not shown).

A few additional targeted, but limited, sensitivity tests have been conducted also with regard to the fountain buster, the time-smoothing of midlevel convective increments, the convective momentum transport settings, and different moisture conservation schemes (not shown). For example, the impact of the convective increment smoothing and the convective momentum transport settings on the speed of convective systems was assessed in the African easterly wave case, but the impact was found to be small. The use of the fountain buster was judged to be overall beneficial. And the effect of applying different available local moisture conservation schemes was concluded to be minor on the considered time scales.
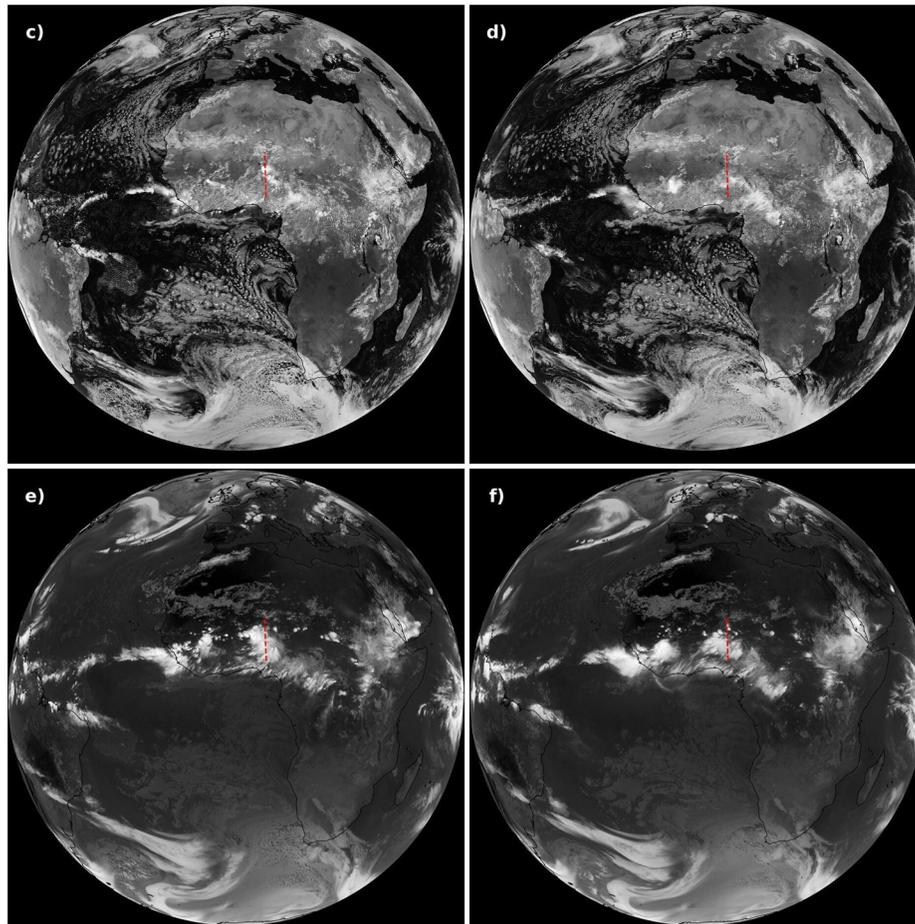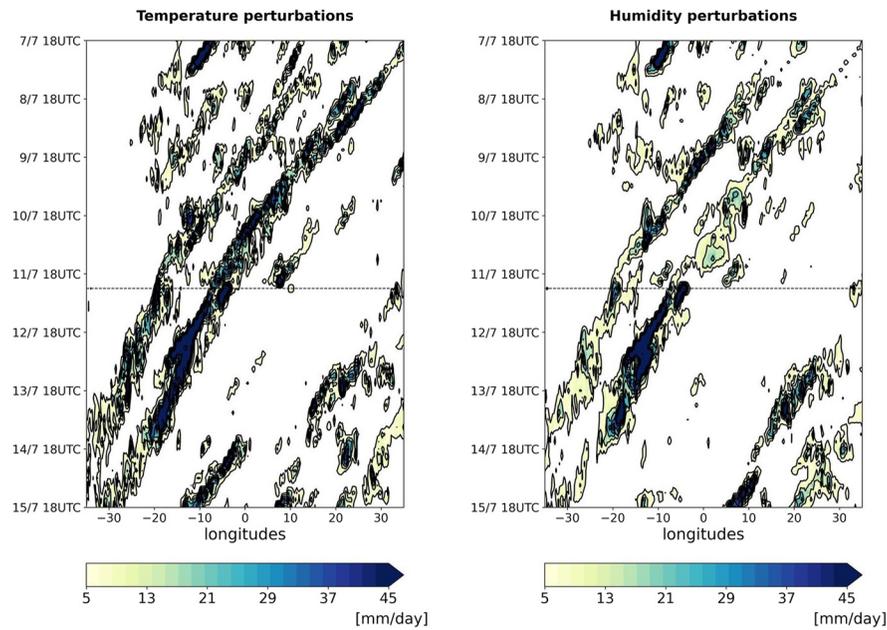
Figure 20. Hovmöller plots of precipitation for the African easterly wave case (top row). (Top left) *MidLevShConv25RAturb*. (Top right) the same configuration, but with humidity perturbations in the midlevel scheme instead of temperature perturbations. The model data are interpolated to the Tropical Rainfall Measuring Mission grid using an area-weighted regridding scheme. (Middle row) the same comparison, but showing simulated reflectance at 0.8 μm for the MSG satellite for 10 July 2010, 12:00 UTC. Bottom row: the same comparison but showing simulated brightness temperature at 10.8 μm for the MSG satellite for 10 July 2010, 18:00 UTC.

5. Months of July 2016 and January 2018

To estimate features that are of importance in climate studies, in particular the top-of-the-atmosphere energy budget, two full months, July 2016 and January 2018, are covered by 5-day forecasts with one overlapping day between forecasts using the convection-permitting *MidLevShConv25RAturb* configuration. The first 24 hr of each forecast are not included in the analysis. The first initial times are 30 June 2016, 00:00 UTC, and 31 December 2017, 00:00 UTC, respectively. Due to a model failure, all forecasts that cover January 2018 were run with a so-called polar cap for advection in which the SL advection is replaced by an interpolation in an area close to the poles (see Appendix A). Low-resolution tests have shown that on time scales of days the impact of the polar cap scheme on the quality of forecasts is neutral.

5.1. Mean Rainfall and Top-Of-The-Atmosphere Radiative Fluxes

Figures 21 and 22 show mean top-of-the-atmosphere outgoing shortwave (top row) and longwave (middle row) radiation for both months alongside the corresponding CERES EBAF 4.1 observations (Loeb et al., 2018). Moreover, monthly mean precipitation is depicted for the model simulations and GPM rainfall observations (bottom rows). Overall, the clouds look very realistic in the simulations and it would be difficult to distinguish between simulations and observations by eye. There are, however, some quite distinct biases in the precipitation fields, particularly over the tropical oceans. Rather strong rainfall biases can be identified over the tropical Pacific and Indian Ocean during boreal winter (Figure 22) and the southern branch of the ITCZ in the western tropical Pacific during boreal summer (Figure 21). Over tropical land and over the extratropics the biases are much less pronounced. Similar tropical rainfall biases have also been described in other convection-permitting global atmosphere models (Caldwell et al., 2021; Hohenegger et al., 2022).

In terms of the global-mean top-of-the-atmosphere radiation budget, the outgoing longwave radiation (OLR) is fairly close to observations without any tuning, while the model is too reflective when it comes to the top-of-the-atmosphere outgoing shortwave radiative fluxes, especially during boreal summer. The results are summarized in Table 2. Some cloud tuning would be required before coupling the *MidLevShConv25RAturb* configuration to an ocean model.

5.2. Madden-Julian Oscillation

A strong Madden-Julian Oscillation (MJO) passed from the Indian Ocean over the Maritime Continent to the West Pacific during the month of January 2018 (Figure 23). The *MidLevShConv25RAturb* model reproduces the propagation of the rainfall very well and the 3-hourly mean rainfall rates are in good agreement with GPM observations. The then-operational GA6.1 global model with parameterized convection shows much weaker eastward propagation with a significant reduction of rainfall when it reaches the Maritime Continent land which is often referred to as the barrier effect (Zhang & Ling, 2017). The propagation of the MJO does not appear to suffer from the barrier effect over the Maritime Continent in the convection permitting simulations and its regions of impacts are fairly well captured. Figure 23 also shows evidence of faster eastward Kelvin-like disturbances and westward Rossby or Mixed Rossby-Gravity wave features (as, e.g., indicated in the figure by red ellipses) that are better captured in the convection-permitting model configuration *MidLevShConvRAturb* compared to the parameterized convection model. The result shows that the convection-permitting model configuration *MidLevShConv25RAturb* is able to accurately predict MJO-related high impact weather in the region several days ahead, though not with lead times on MJO time scales of weeks or months.

5.3. The West Pacific Subtropical High

The West Pacific subtropical high (WPSH), a distinct anticyclone in the middle and lower troposphere over the western North Pacific, is a key component of the East Asian summer monsoon system. It affects the regional hydrological cycle and has a significant influence on tropical cyclone activity in the Western North Pacific (B. Wang et al., 2013; Yihui & Chan, 2005). Correct simulations of the WPSH are important for seasonal forecasting and climate projections. In low-resolution versions of the Unified Model, as in many other models, the mean-state summer WPSH is too weak and located too far east (Rodríguez et al., 2017). These biases lead to an underestimation of the southwesterly monsoon flow over East Asia, affecting its representation of the seasonal
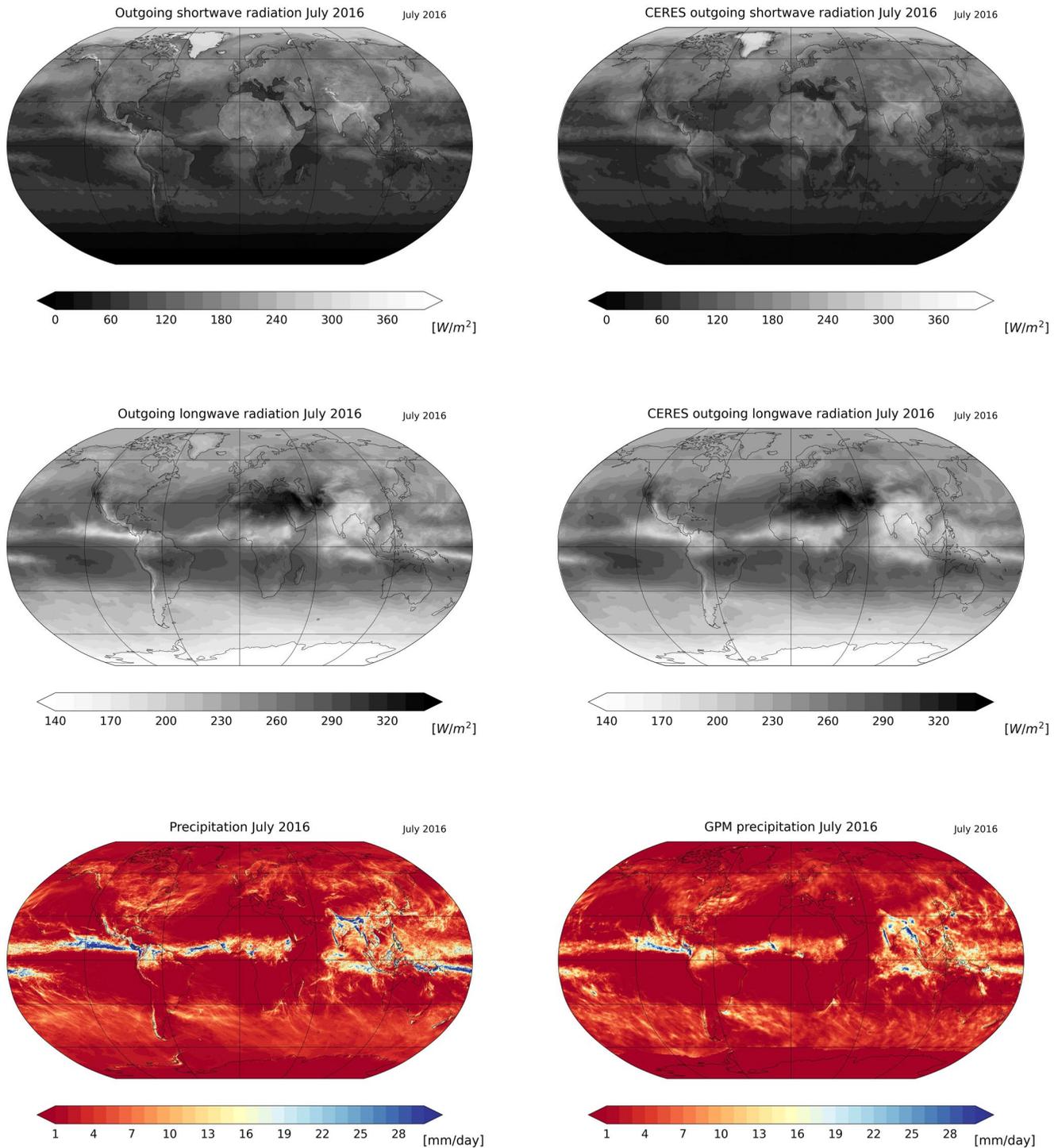
Figure 21. Mean outgoing shortwave radiation for July 2016 from the *MidLevShConv25RAturb* simulation (top left) and CERES EBAF 4.1 (top right). Mean outgoing longwave radiation for July 2016 from the *MidLevShConv25RAturb* simulation (middle left) and CERES EBAF 4.1 (middle right). Mean precipitation for July 2016 based on the *MidLevShConv25RAturb* simulation (bottom left) and Global Precipitation Measurement (bottom right).

water cycle in the area. Studies of systematic errors in climate models have shown that they can develop in the first few days of a simulation, and then persist to climate time scales (Martin et al., 2021). We can therefore gain an understanding of the emergence of such biases by analyzing errors in initialized hindcasts (Rodríguez & Milton, 2019).
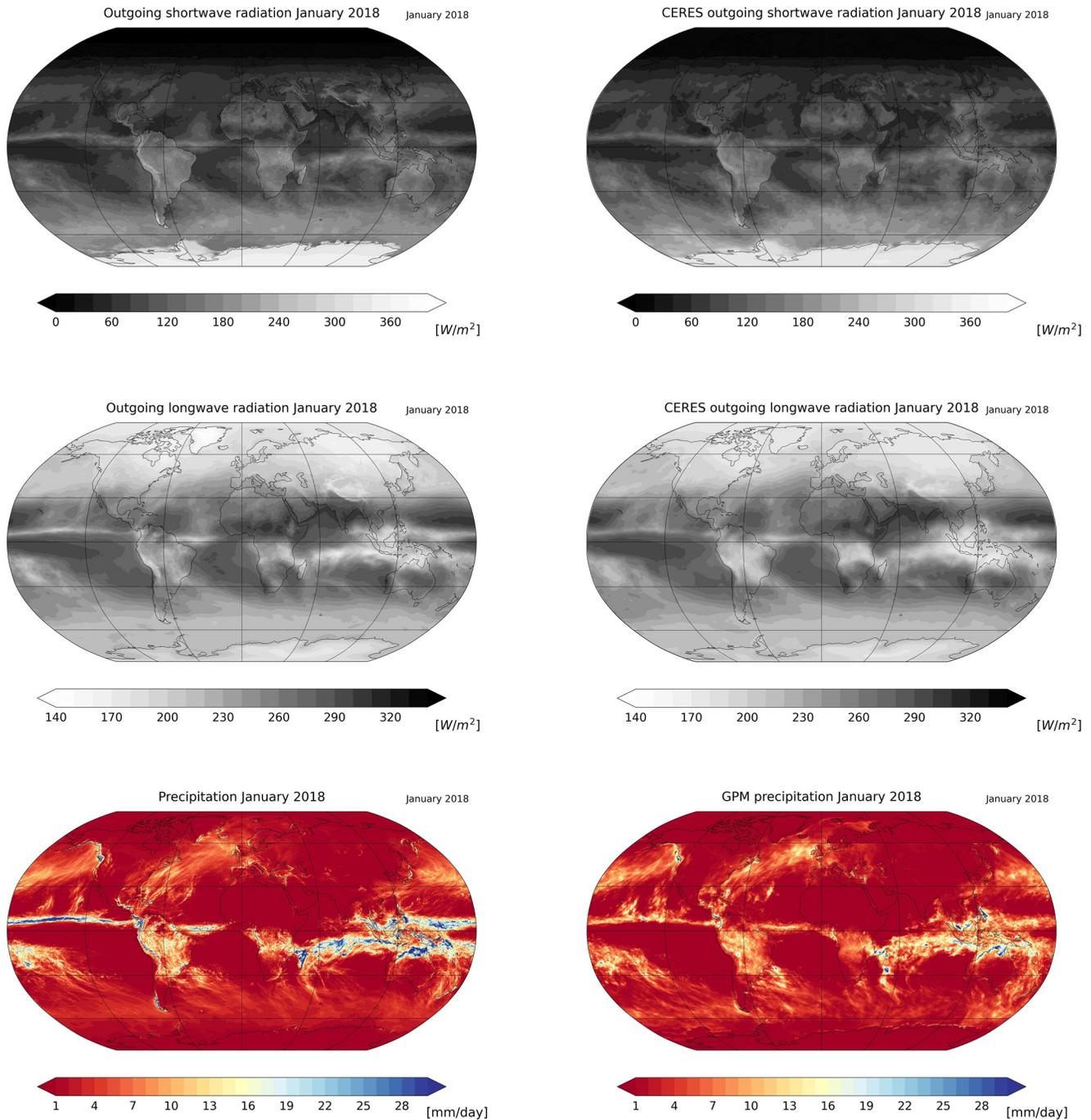
Figure 22. Mean outgoing shortwave radiation for January 2018 from the *MidLevShConv25RAturb* simulation (top left) and CERES EBAF 4.1 (top right). Mean outgoing longwave radiation for January 2018 from the *MidLevShConv25RAturb* simulation (middle left) and CERES EBAF 4.1 (middle right). Mean precipitation for January 2018 based on the *MidLevShConv25RAturb* simulation (bottom left) and Global Precipitation Measurement (bottom right).

Here, a comparison of the development of circulation biases associated with the WPSH in global 5 km-resolution *MidLevShConv25RAturb* configuration simulations and low-resolution N216 (about 80 km-resolution) GA6 configuration forecasts initialized from Met Office analyses is made for July 2016 (Figure 24). The 850 hPa geopotential height fields at 5 days lead time indicate that in the 5 km-resolution *MidLevShConv25RAturb* simulations the WPSH is developed more strongly and extends further to the west and south, in better agreement with ERA5 compared to the low-resolution GA6 forecasts (Figure 24, panels a–c), although other aspects are slightly degraded. This is confirmed by comparing three indices, I_{in}, I_W and I_N in ERA5 and the model forecasts at a

Table 2
Observed and Modeled Global-Mean Top-Of-The-Atmosphere (TOA) Radiative Fluxes for July 2016 and January 2018

	TOA outgoing longwave flux	TOA outgoing shortwave flux
Observed CERES EBAF 4.1 July 2016	244.51 (W/m^2)	92.96 (W/m^2)
Modeled *MidLevShConv25RAturb* July 2016	243.92 (W/m^2)	97.67 (W/m^2)
Observed CERES EBAF 4.1 January 2018	237.30 (W/m^2)	105.03 (W/m^2)
Modeled *MidLevShConv25RAturb* January 2018	237.59 (W/m^2)	106.47 (W/m^2)

lead time of 5 days, following Chen et al. (2010) and Lu and Dong (2001). The indices represent, respectively, the intensity, westward extension and northern edge of the WPSH (Figure 24, panel d). The 5 km-resolution *MidLevShConv25RAturb* forecasts show an improvement in the representation of the WPSH mainly in terms of the intensity and the westward extension, important aspects in the context of the regional weather and climate. However, biases in the region of the Kamchatka Peninsula and some other areas increase.

5.4. Extratropical Cyclones

Extratropical cyclones are tracked in the *MidLevShConv25RAturb* configuration simulations for both months, July 2016 and January 2018, using the TRACK (Hodges, 1995) algorithm on the 850 hPa relative vorticity field. The same tracking algorithm is also applied to ECMWF analyses and the then-operational Met Office Unified model for both July 2016 and January 2018. The operational model used by the Met Office in July 2016 and January 2018 was the GA6 configuration with fully parameterized convection (Walters et al., 2017), at N768 resolution for July 2016 and at N1280 resolution for January 2018.

Cyclones identified in the *MidLevShConv25RAturb* simulation are matched (location and intensity) against those in ECMWF analyses using the method of Froude et al. (2007a, 2007b). Then, for comparison, the matching is re-run on the then-operational GA6.1 forecast data and ECMWF analyses. The mean error in extratropical cyclone location (top row) and intensity (bottom row) for *MidLevShConv25RAturb* (black lines) and GA6 (red lines) relative to ECMWF analyses, are shown for the NH winter (Figures 25a and 25b) and SH winter (Figures 25c and 25d). The performance of the *MidLevShConv25RAturb* simulation is near neutral to that of both resolutions of the GA6 operational model, relative to ECWMF analyses (Figure 25). Therefore, in the context of extratropical cyclone mean location and intensity, there is no improvement from running the *MidLevShConv25RAturb* configuration at 5 km grid spacing compared to the then-operational model at 10 km resolution, but also no degradation.
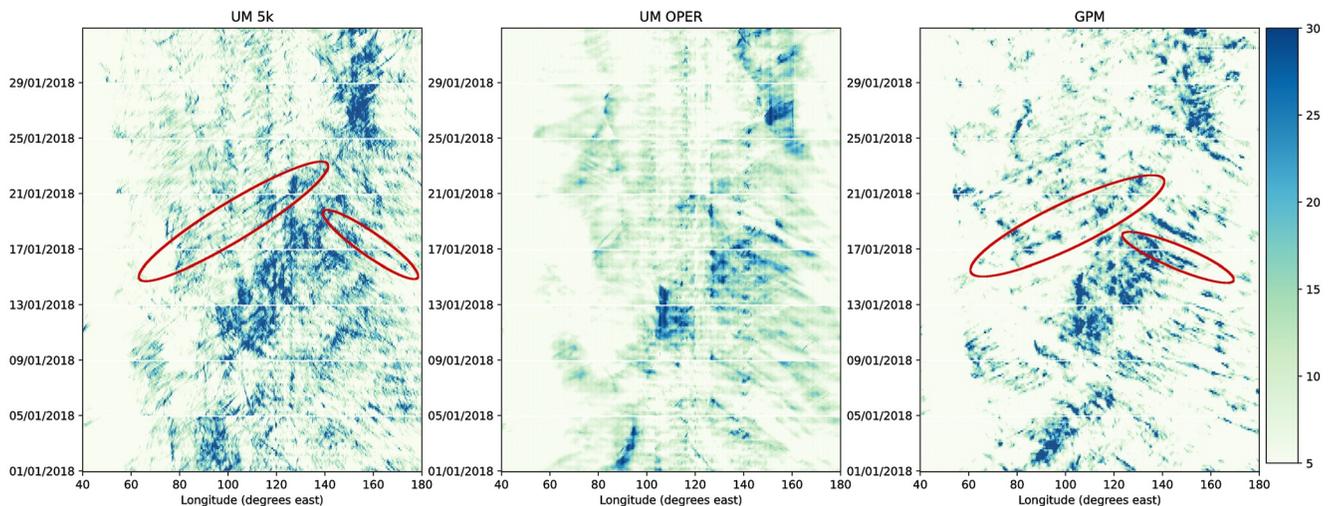
Figure 23. (Left panel) Hovmöller plot of 3-hourly mean rainfall rate based on the *MidLevShConv25RAturb* configuration simulations for January 2018. Middle panel: Hovmöller plot of 3-hourly mean rainfall rate based on the then operational model GA6.1 (Walters et al., 2017) with fully parameterized convection. (Right panel) Hovmöller plot of rainfall based on the Global Precipitation Measurement rainfall data averaged to 3-hourly mean values. For all plots the rainfall was averaged over the latitudes 10 South to 10 North, units are mm/day.

Figure 24. (a) July 2016 mean 850 hPa geopotential height and horizontal wind in ERA5. (b, c) Difference of July 2016 mean 850 hPa geopotential height and horizontal wind to ERA5 for the 5 km-resolution *MidLev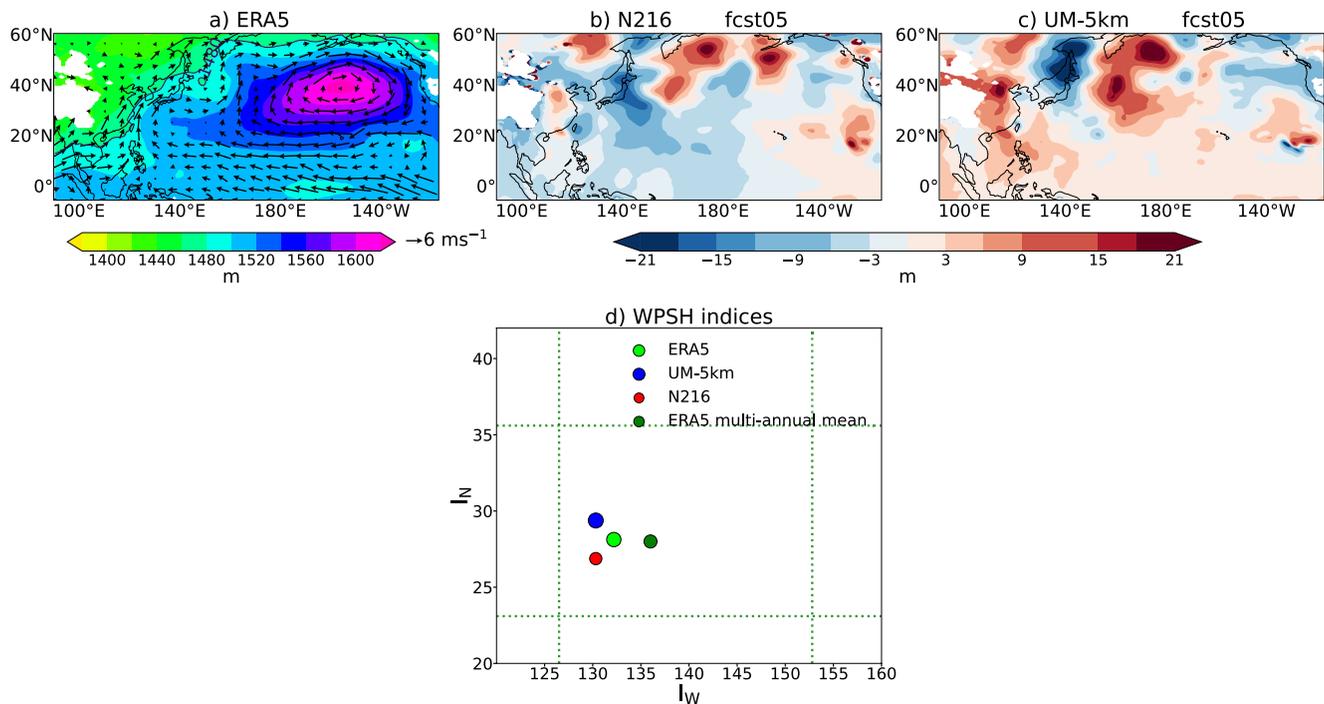ShConv25RAturb* configuration and GA6 at N216 resolution, respectively. Geopotential height is in m and wind in m s^{-1}. West Pacific subtropical high intensity, west extension and northern edge indices in ERA5 and model hindcasts are shown in (d). Axes in the scatter plot are as follows: abscissa displays values of the west extension index, I_W (°E) and ordinate shows values of the northern edge index, I_N (°N). The radii of scatter circles are proportional to the intensity index, I_{in}. Indices are shown for July 2016 ERA5 (light green), 5 km-resolution (blue) and N216-resolution (red). To place 2016 data in context, the following information has been added: dotted lines indicate maximum and minimum values of July ERA5 I_W and I_N for the 1979–2019 period, and a dark green dot shows the period mean value.

The statistics in the analysis is limited and includes only 1 month and eight 5-day forecasts for each hemisphere. Small differences between model configurations should not be overinterpreted, even if they lie outside of the 95% confidence intervals indicated by the shaded areas. Nevertheless, the structure of individual cyclones (e.g., fronts) may be simulated better at 5 km resolution and will be an area of future work.

6. Summary and Conclusions

In the present work the challenges of the convective gray zone are discussed in the context of a 5 km-resolution global atmospheric model, and a few fundamental ideas concerning the representation of turbulence and convection are proposed and assessed. Different questions regarding the formulation of subgrid turbulence, convection, and model dynamics are considered: should the convection parameterization be switched off at a particular grid length in convection-permitting models or is a more seamless approach to be preferred? Are one-dimensional turbulence parameterizations sufficient in kilometer-scale models or is a three-dimensional representation of turbulence beneficial, or a compromise between the two? What are the impacts of non-conservation characteristics of the advection scheme and physical parameterizations? Most of these questions are not in themselves binary in nature (convection parameterization switched on or off). On the contrary, as the model resolution is increased continuously, the resolved convective and turbulent fluxes are supposed to increase continuously and the subgrid, parameterized fluxes should decrease continuously.

The investigation of the aforementioned questions always takes place in a certain context, with regard to a particular chosen reference configuration and taking specific representations of subgrid processes into account. For example, if the considered representation of subgrid convection is distinctly deficient, then in practice it might be advisable to switch off such a subgrid convection scheme in kilometer-scale models completely, even though in theory convection is not well resolved. Moreover, it is clear that the particular application of the model and the quantities of interest are an important factor in the assessment. There is not one model configuration which
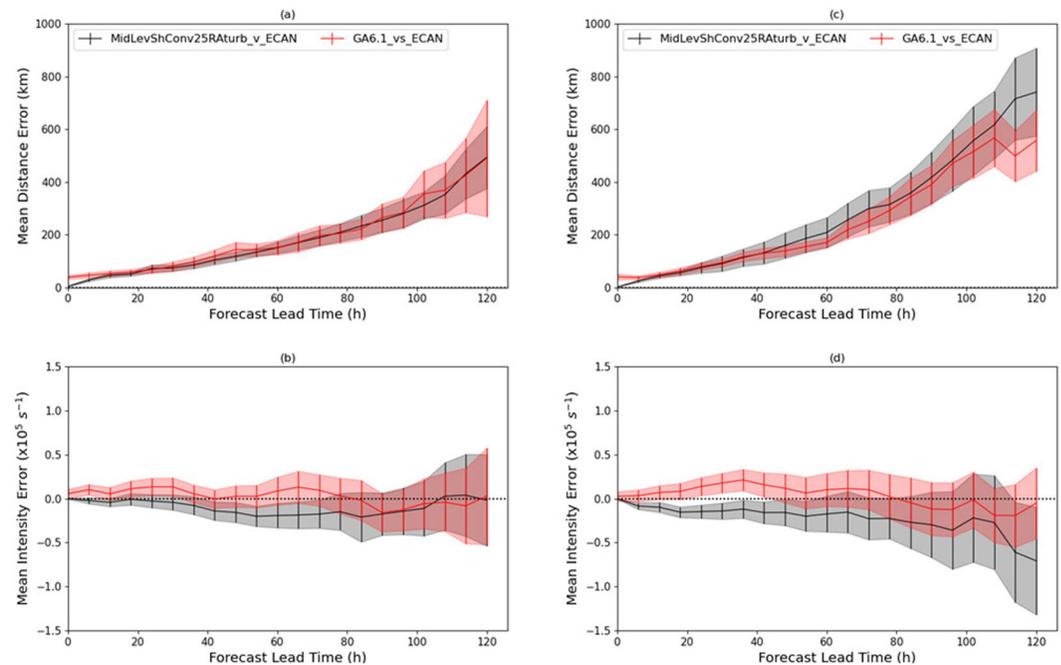
Figure 25. Results of extratropical storm tracking for January 2018 in the northern hemisphere (two left panels) and July 2016 in the southern hemisphere (two right panels). (a) Mean storm track position error of the global 5 km-resolution *MidLevShConv25RAturb* forecasts against European Centerre for Medium-Range Weather Forecasts (ECMWF) analysis (black line), and the then-operational Unified Model GA6 at 10 km (N1280) resolution against ECMWF analysis (red line) for January 2018 and the northern hemisphere. (b) The same models and comparisons as in (a), but for the mean intensity error of extratropical cyclones. The intensity measure is 850 hPa vorticity. (c) As for (a), but for July 2016 and the southern hemisphere. Here the then-operational model is run at N768 resolution. Panel (d) The intensity error results for July 2016 over the southern hemisphere. The shaded areas indicate 95% confidence intervals.

performs better than any other configuration in all meteorological situations and geographical locations. The particular metrics used in the evaluation will always influence the conclusions. In numerical weather prediction, for instance, an issue is how to weight more traditional error indices that target large-scale fields in the middle troposphere against more impact-based, near-surface measures.

Here a case study approach is taken in which selected test cases are compared to observations. A robust conclusion of the assessment in the present study, conditional on the assumed framework, is that the convection-permitting configurations which include scale-aware turbulence and a carefully reduced and simplified mass-flux convection scheme outperform both the model with fully parameterized convection as well as a configuration in which the subgrid convection parameterization is switched off completely with regard to the organization of convection and related features of tropical variability. The key idea with regard to the representation of subgrid convection in the convective gray zone is to use a unified scheme in the free troposphere which is allowed to trigger from the top of the boundary layer and not below, and to limit the subgrid convective mass flux, both in the free troposphere as well as the shallow-convective mass flux in the boundary layer. The assessment in the extratropics is more ambiguous and challenging than in the tropics, and better statistics and larger ensemble sizes are needed. The configuration with fully parameterized convection is more competitive in the midlatitudes, whereas for tropical variability it shows distinct deficiencies. The results are in broad agreement with the studies by Judt and Rios-Berrios (2021) and Judt (2018). Moreover, as already mentioned, the conclusion might differ depending on the considered variable and metric, for example, rainfall or upper-tropospheric wind and temperature. Also the comparison of longer-term mean statistics can be less straightforward because the model with parameterized convection is rather resolution insensitive and can therefore be tuned using low-resolution versions of the model. This does not apply the same way to the convection-permitting configurations, especially given that physical parameterizations are generally not scale-aware. Nevertheless, for the convection-permitting configurations a further effort in cloud tuning will be required, for instance, to bring global-mean top-of-the-atmosphere radiative fluxes more inline with observations, an aspect that is important in climate applications. A process-based

approach to evaluation, however, as taken in the present study, is certainly a valuable and integral part of any assessment.

Based on the presented work the further development of scale-aware physical parameterizations is endorsed, particularly with regard to convection (Holloway et al., 2014), not only because of the expected better performance, fidelity of process representation, and agreement with observations, but also because it enables better traceability and seamlessness of the model development process across resolutions, time scales, and applications. In the context of the Met Office and its partner institutions, we plan to operationally implement the 5 km-resolution global model as an experimental forecast to supplement the main lower-resolution global ensemble. This could contribute to a limited number of operational outputs, likely focused on near-surface variables and high-impact weather, as well as to the ongoing assessment and development and act as a stepping-stone toward a future convection-permitting global ensemble. This step-by-step approach allows for better identifying and understanding the strengths and limitations of a convection-permitting global model under various meteorological conditions. Future plans also include season- and year-long simulations and the coupling to the NEMO ocean model. But as the radiative flux estimates in the case of the *MidLevShConv25RAturb* configuration show, some cloud tuning will be required as a prerequisite.

The Met Office modeling system will be overhauled over the next few years with the introduction of a new dynamical core, a new two-moment microphysics scheme, a new convection scheme, and improvements to the cloud scheme. The present work is therefore not intended to be the final word on convection-permitting global model development. Nevertheless, the results of the study will help to guide some fundamental choices in the future development of an adequate convection-permitting global coupled modeling system for use across weather and climate time scales.

Appendix A: Associated Global Model Development Tickets

Met Office Global Model Evaluation and Development tickets associated with the discussed model configurations are listed and described in some more detail in Table A1. The information is supplementary to Table 1 and the configuration description in Section 2. A more in-depth description of the tickets is accessible only to registered Unified Model developers and users, but the ticket numbers are nevertheless provided here as an additional potential reference.

Table A1
Met Office Global Model Evaluation and Development (GMED) Tickets and Descriptions Related to the Model Configurations Discussed in Section 2 and Table 1

Short description	GMED ticket number	Notes
Fountain buster	492	Modification of the semi-Lagrangian advection scheme, see Section 2
Additional termination condition for convection	417	Addition termination condition for subgrid convection whereby convection terminates when it reaches the level of neutral buoyancy of an undilute parcel
Adaptation of convective momentum transport settings	487	The pressure gradient term in the Gregory-Kershaw scheme (Gregory et al., 1997) is reduced: cpress_term = 0.3 instead of 0.7
Forced detrainment weighted CAPE calculation	508	The CAPE used in the convection scheme closure is a dilute CAPE, that is, it is reduced by entrainment. The convection scheme represents an ensemble of plumes through adaptive forced detrainment (Derbyshire et al., 2011). The change modifies the CAPE calculation so that it uses a forced detrainment weighted CAPE
Time-smoothing of convective increments	191	To improve physics-dynamics coupling time-damped convective increments to theta and humidity are passed out of the convection scheme rather than instantaneous values. Damping time scale is set to 1,800 s

Table A1
Continued

Short description	GMED ticket number	Notes
Switch off deep convection parameterization and fix CAPE timescale for midlevel scheme	557	Switch off deep convection scheme and rely on midlevel and shallow convection parameterizations to represent subgrid convection in the model. The CAPE closure time scale in the midlevel parameterization is set to a fixed value
Setting for implicit boundary layer solver	290	Increase parameter for implicit boundary layer solver in unstable boundary layer columns from 0.5 to puns = 1.0. The higher the value the more stable the scheme
Polar cap for advection	471	The regions close to the poles have very small longitudinal grid lengths at high resolution. This leads to crossing of trajectories in the advection scheme. To avoid these problems polar cap regions are introduced across which advected quantities are interpolated over the poles from advected values at the edges of the polar caps
Prognostic entrainment for midlevel convection scheme and allow midlevel convection to initiate from the top of the boundary layer	507	Use of the prognostic entrainment scheme in the midlevel convection parameterization (Willett & Whitall, 2017). Allow the midlevel scheme to start from the top of the boundary layer instead of two model levels above

Note. CAPE, Convective Available Potential Energy.

Data Availability Statement

ERA5 data (Hersbach et al., 2020) were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels). Tropical Rainfall Measuring Mission (TRMM) (Huffman et al., 2007) data (TMPA/3B43, V7) are available from the Goddard Earth Sciences Data and Information Services Center (GES DISC), https://disc.gsfc.nasa.gov/datasets/TRMM_3B43_V7/summary. The GPM data (Huffman et al., 2019) were provided by the NASA/Goddard Space Flight Center, which develop and compute the GPM IMERG data set as a contribution to the GPM project, and archived at the NASA GES DISC (https://disc.sci.gsfc.nasa.gov/). The EUREC⁴A dropsonde data set (George et al., 2021) is publicly available at AERIS (https://doi.org/10.25326/221). The Darwin radar data can be accessed via the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) programme data portal at https://www.arm.gov/data/data-sources/cpol (Raut et al., 2021). The daily mean CMsaf water vapor path data for the year 2016 can be obtained from the CMsaf team on request through the web page at https://wui.cmsaf.eu/safira/action/viewProduktSearch (Courcoux & Schröder, 2013). The observations related to tropical cyclones can be downloaded from the IBTrACS data base of the NOAA data center at https://www.ncdc.noaa.gov/ibtracs/ (Levinson & Diamond, 2009). The CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data (Loeb et al., 2018) in netCDF Edition 4.1 are available from the NASA Langley Atmospheric Science Data Center DAAC at https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF_L3B.004.1. The model data on which the figures are based are publicly available at https://doi.org/10.5281/zenodo.7096991 (Tomassini, 2022).

References

Arakawa, A., Jung, J.-H., & Wu, C.-M. (2011). Toward unification of the multiscale modeling of the atmosphere. *Atmospheric Chemistry and Physics*, *11*(8), 3731–3742. https://doi.org/10.5194/acp-11-3731-2011

Arakawa, A., & Wu, C.-M. (2013). A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I. *Journal of the Atmospheric Sciences*, *70*(7), 1977–1992. https://doi.org/10.1175/jas-d-12-0330.1

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

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

Boutle, I. A., Eyre, J. E. J., & Lock, A. P. (2014). Seamless stratocumulus simulation across the turbulent gray zone. *Monthly Weather Review*, *142*(4), 1655–1668. https://doi.org/10.1175/mwr-d-13-00229.1

Bush, M., Allen, T., Bain, C., Boutle, I., Edwards, J., Finnenkoetter, A., et al. (2020). The first Met Office unified model-JULES regional atmosphere and land configuration, RAL1. *Geoscientific Model Development*, *13*(4), 1999–2029. https://doi.org/10.5194/gmd-13-1999-2020

Bush, M., Boutle, I., Edwards, J., Finnenkoetter, A., Franklin, C., Hanley, K., et al. (2022). The second Met Office unified model-JULES regional atmosphere and land configuration, RAL2. *Geoscientific Model Development*, *16*(6), 1713–1734. https://doi.org/10.5194/gmd-2022-209

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., et al. (2021). Convection-permitting simulations with the E3SM global atmosphere model. *Journal of Advances in Modeling Earth Systems*, *13*(11), e2021MS002544. https://doi.org/10.1029/2021MS002544

Chen, H., Zhou, T., Neale, R. B., Wu, X., & Zhang, G. J. (2010). Performance of the new NCAR CAM3.5 in East Asian summer monsoon simulations: Sensitivity to modifications of the convection scheme. *Journal of Climate*, *23*(13), 3657–3675. https://doi.org/10.1175/2010jcli3022.1

Clark, P., Roberts, N., Lean, H., Ballard, S. P., & Charlton-Perez, C. (2016). Convection-permitting models: A step-change in rainfall forecasting. *Meteorological Applications*, *23*(2), 165–181. https://doi.org/10.1002/met.1538

Courcoux, N., & Schröder, M. (2013). Vertically integrated water vapour, humidity and temperature at pressures levels and layers from ATOVS—Daily Means Monthly Means [Dataset]. Satellite Application Facility on Climate Monitoring. https://doi.org/10.5676/EUM_SAF_CM/WVT_ATOVS/V001

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

Derbyshire, S. H. (1997). *Recommendations for UM parametrization of stable boundary layer (technical report no. 38)*. MRU Cardington.

Derbyshire, S. H., Maidens, A. V., Milton, S. F., Stratton, R. A., & Willett, M. R. (2011). Adaptive detrainment in a convective parametrization. *Quarterly Journal of the Royal Meteorological Society*, *137*(660), 1856–1871. https://doi.org/10.1002/qj.875

de Roode, S. R., Frederikse, T., Siebesma, A. P., Ackerman, A. S., Field, P. R., Hill, A., et al. (2019). Turbulent transport in the grey zone: A large-eddy simulation intercomparison study of the CONSTRAIN cold air outbreak case. *Journal of Advances in Modeling Earth Systems*, *11*(3), 597–623. https://doi.org/10.1029/2018MS001443

Dipankar, A., Webster, S., Sun, X., Sanchez, C., North, R., Furtado, K., et al. (2020). SINGV: A convective-scale weather forecast model for Singapore. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 4131–4146. https://doi.org/10.1002/qj.3895

Field, P. R., Brodzkova, R., Chen, M., Dudhia, J., Lac, C., Hara, T., et al. (2017). Exploring the convective grey zone with regional simulations of a cold air outbreak. *Quarterly Journal of the Royal Meteorological Society*, *143*(707), 2537–2555. https://doi.org/10.1002/qj.3105

Franklin, C. N., Protat, A., Leroy, D., & Fontaine, E. (2016). Controls on phase composition and ice water content in a convection-permitting model simulation of a tropical mesoscale convective system. *Atmospheric Chemistry and Physics*, *16*(14), 8767–8789. https://doi.org/10.5194/acp-16-8767-2016

Froude, L. S. R., Bengtsson, L., & Hodges, K. I. (2007a). The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Monthly Weather Review*, *135*(2), 315–333. https://doi.org/10.1175/mwr3274.1

Froude, L. S. R., Bengtsson, L., & Hodges, K. I. (2007b). The prediction of extratropical storm tracks by the ECMWF and NCEP ensemble prediction systems. *Monthly Weather Review*, *135*(7), 2545–2567. https://doi.org/10.1175/mwr3422.1

George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., et al. (2021). JOANNE: Joint dropsonde observations of the atmosphere in tropical North Atlantic meso-scale environments. *Earth System Science Data*, *13*(11), 5253–5272. https://doi.org/10.5194/essd-13-5253-2021

Gerard, L. (2015). Bulk mass-flux perturbation formulation for a unified approach of deep convection at high resolution. *Monthly Weather Review*, *143*(10), 4038–4063. https://doi.org/10.1175/mwr-d-15-0030.1

Gerard, L., & Geleyn, J.-F. (2005). Evolution of a subgrid deep convection parametrization in a limited area model with increasing resolution. *Quarterly Journal of the Royal Meteorological Society*, *131*(610), 2293–2312. https://doi.org/10.1256/qj.04.72

Gerard, L., Piriou, J.-M., Brožková, R., Geleyn, J.-F., & Banciu, D. (2009). Cloud and precipitation parameterization in a meso-gamma scale operational weather prediction model. *Monthly Weather Review*, *137*(11), 3960–3977. https://doi.org/10.1175/2009MWR2750.1

Gregory, D., Kershaw, R., & Inness, P. M. (1997). Parametrization of momentum transport by convection II: Tests in single-column and general circulation models. *Quarterly Journal of the Royal Meteorological Society*, *123*(541), 1153–1183. https://doi.org/10.1002/qj.49712354103

Grell, G. A., & Freitas, S. R. (2014). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, *14*(10), 5233–5250. https://doi.org/10.5194/acp-14-5233-2014

Hanley, K. E., Plant, R. S., Stein, T. H. M., Hogan, R. J., Nicol, J. C., Lean, H. W., et al. (2015). Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, *141*(686), 272–284. https://doi.org/10.1002/qj.2356

Harris, L., Zhou, L., Lin, S.-J., Chen, J., Chen, X., Gao, K., et al. (2020). GFDL SHiELD: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2020MS002223. https://doi.org/10.1029/2020MS002223

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

Hodges, K. I. (1995). Feature tracking on the unit sphere. *Monthly Weather Review*, *123*(12), 3458–3465. https://doi.org/10.1175/1520-0493(1995)123<3458:ftotus>2.0.co;2

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., et al. (2022). ICON-sapphire: Simulating the components of the Earth system and their interactions at kilometer and subkilometer scales. In *Geoscientific Model Development Discussions*. https://doi.org/10.5194/gmd-2022-171

Holloway, C. E., Petch, J. C., Beare, R. J., Bechtold, P., Craig, G. C., Derbyshire, S. H., et al. (2014). Understanding and representing atmospheric convection across scales: Recommendations from the meeting held at Dartington Hall, Devon, UK, 28–30 January 2013. *Atmospheric Science Letters*, *15*, 348–353. https://doi.org/10.1002/asl2.508

Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., et al. (2020). The atmospheric boundary layer and the gray zone of turbulence: A critical review. *Journal of Geophysical Research*, *125*(13), e2019JD030317. https://doi.org/10.1029/2019JD030317

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., et al. (2007). The TRMM multi-satellite precipitation analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/jhm560.1

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). *GPM IMERG final precipitation L3 half hourly 0.1 degree x 0.1 degree V06. Greenbelt, MD*. Goddard Earth Sciences Data and Information Services Center (GES DISC). https://doi.org/10.5067/GPM/IMERG/3B-HH/06

Judt, F. (2018). Insights into atmospheric predictability through global convection-permitting model simulations. *Journal of the Atmospheric Sciences*, *75*(5), 1477–1497. https://doi.org/10.1175/jas-d-17-0343.1

Judt, F., & Rios-Berrios, R. (2021). Resolved convection improves the representation of equatorial waves and tropical rainfall variability in a global nonhydrostatic model. *Geophysical Research Letters*, *48*(14), e2021GL093265. https://doi.org/10.1029/2021GL093265

Kendon, E. J., Jones, R. G., Kjellström, E., & Murphy, J. M. (2010). Using and designing GCM-RCM ensemble regional climate projections. *Journal of Climate*, *23*(24), 6485–6503. https://doi.org/10.1175/2010JCLI3502.1

Kendon, E. J., Prein, A. F., Senior, C. A., & Sirling, A. (2021). Challenges and outlook for convection-permitting climate modelling. *Philosophical Transactions of the Royal Society A*, *379*(2195), 20190547. https://doi.org/10.1098/rsta.2019.0547

Kwon, Y. C., & Hong, S. Y. (2017). A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Monthly Weather Review*, *145*(2), 583–598. https://doi.org/10.1175/mwr-d-16-0034.1

Lean, H. W., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., & Halliwell, C. (2008). Characteristics of high-resolution versions of the Met Office unified model for forecasting convection in the United Kingdom. *Monthly Weather Review*, *136*(9), 3408–3424. https://doi.org/10.1175/2008mwr2332.1

Levinson, D. H., & Diamond, H. J. (2009). Combining tropical cyclone data sets worldwide. *Eos*, *90*(35), 301. https://doi.org/10.1029/2009EO350006

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth's radiant energy system (CERES) energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) Edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

Lu, R., & Dong, B. (2001). Westward extension of North Pacific subtropical high in summer. *Journal of the Meteorological Society of Japanese Series II*, *79*(6), 1229–1241. https://doi.org/10.2151/jmsj.79.1229

Malardel, S., & Bechtold, B. (2019). The coupling of deep convection with the resolved flow via the divergence of mass flux in the IFS. *Quarterly Journal of the Royal Meteorological Society*, *145*(722), 1832–1845. https://doi.org/10.1002/qj.3528

Martin, G. M., Levine, R. C., Rodriguez, J. M., & Vellinga, M. (2021). Understanding the development of systematic errors in the Asian summer monsoon. *Geoscientific Model Development*, *14*(2), 1007–1035. https://doi.org/10.5194/gmd-14-1007-2021

Maynard, C., Melvin, T., & Müller, E. H. (2020). Multigrid preconditioners for the mixed finite element dynamical core of the LFRic atmospheric model. *Quarterly Journal of the Royal Meteorological Society*, *146*(733), 3917–3936. https://doi.org/10.1002/qj.3880

McTaggart-Cowan, R., Vaillancourt, P. A., Separovic, L., Corvec, S., & Zadra, A. (2020). A convection parameterization for low-CAPE environments. *Monthly Weather Review*, *148*(12), 4917–4941. https://doi.org/10.1175/mwr-d-20-0020.1

Miyakawa, T., & Miura, H. (2019). Resolution dependencies of tropical convection in a global cloud/cloud-system resolving model. *Journal of the Meteorological Society of Japan*, *97*(3), 745–756. https://doi.org/10.2151/jmsj.2019-034

Miyakawa, T., Noda, A. T., & Kodama, C. (2018). The impact of hybrid usage of a cumulus parameterization scheme on tropical convection and large-scale circulations in a global cloud-system resolving model. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2952–2970. https://doi.org/10.1029/2018MS001302

Pan, D. M., & Randall, D. A. (1998). A cumulus parameterization with a prognostic closure. *Quarterly Journal of the Royal Meteorological Society*, *124*(547), 949–981. https://doi.org/10.1002/qj.49712454714

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, *53*(2), 323–361. https://doi.org/10.1002/2014RG000475

Putnam, W. M., & Suarez, M. (2011). Cloud-system resolving simulations with the NASA Goddard Earth Observing System global atmospheric model (GEOS-5). *Geophysical Research Letters*, *38*, L16809. https://doi.org/10.1029/2011GL048438

Radermacher, C., & Tomassini, L. (2012). Thermodynamic causes for future trends in heavy precipitation over Europe based on an ensemble of regional climate model simulations. *Journal of Climate*, *25*(21), 7669–7689. https://doi.org/10.1175/JCLI-D-11-00304.1

Rajagopal, M., Zipser, E., Huffman, G., Russell, J., & Tan, J. (2021). Comparisons of IMERG Version 06 precipitation at and between passive microwave overpasses in the tropics. *Journal of Hydrometeorology*, *22*, 2117–2130. https://doi.org/10.1175/jhm-d-20-0226.1

Raut, B., Jackson, R., Picel, M., Collis, S., Bergemann, M., & Jakob, C. (2021). An adaptive tracking algorithm for convection in simulated and remote sensing data. *Journal of Applied Meteorology and Climatology*, *60*(4), 513–526. https://doi.org/10.1175/JAMC-D-20-0119.1

Roberts, N. M. (2003). *The impact of a change to the use of the convection scheme to high-resolution simulations of convective events (technical report no. 407)*. Met Office.

Rodríguez, J. M., & Milton, S. F. (2019). East Asian summer atmospheric moisture transport and its response to interannual variability of the West Pacific subtropical high: An evaluation of the Met Office Unified Model. *Atmosphere*, *10*(8), 457. https://doi.org/10.3390/atmos10080457

Rodríguez, J. M., Milton, S. F., & Marzin, C. (2017). The East Asian atmospheric water cycle and monsoon circulation in the Met Office Unified Model. *Journal of Geophysical Research: Atmospheres*, *122*(19), 10–246. https://doi.org/10.1002/2016jd025460

Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, *8*(2), 786–812. https://doi.org/10.1002/2016MS000634

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., & Iga, S. (2008). Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, *227*(7), 486–514. https://doi.org/10.1016/j.jcp.2007.02.006

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, *5*(3), 172–184. https://doi.org/10.1007/s40641-019-00131-0

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., et al. (2018). An update on the RTTOV fast radiative transfer model. *Geoscientific Model Development*, *11*(7), 2717–2737. https://doi.org/10.5194/gmd-11-2717-2018

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., et al. (2018). The North Atlantic waveguide and downstream impact experiment. *Bulletin America Meteorology Social*, *99*(8), 1607–1637. https://doi.org/10.1175/bams-d-17-0003.1

Senior, C. A., Marsham, J. H., Berthou, S., Burgin, L. E., Folwell, S. S., Kendon, E. J., et al. (2021). Convection-permitting regional climate-change simulations for understanding future climate and informing decision making in Africa. *Bulletin America Meteorology Social*, *102*(6), E1206–E1223. https://doi.org/10.1175/bams-d-20-0020.1

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., & Ringler, T. D. (2012). A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Monthly Weather Review*, *140*(9), 3090–3105. https://doi.org/10.1175/mwr-d-11-00215.1

Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., et al. (2022). Ambitious partnership needed for reliable climate prediction. *Nature Climate Change*, *12*(6), 499–503. https://doi.org/10.1038/s41558-022-01384-8

Slingo, J., Bates, K., Nikiforakis, N., Piggott, M., Roberts, M., Shaffrey, L., et al. (2021). *Next generation climate models: A step change for net zero and climate adaptation (climate change science and solutions: Briefing 1)*. The Royal Society. Retrieved from https://royalsociety.org/-/media/policy/projects/climate-change-science-solutions/climate-science-solutions-modelling.pdf

Stein, T. H. M., Hogan, R. J., Hanley, K. E., Nicol, J. C., Lean, H. W., Plant, R. S., et al. (2014). The three-dimensional morphology of simulated and observed convective storms over Southern England. *Monthly Weather Review*, *142*(9), 3264–3283. https://doi.org/10.1175/MWR-D-13-00372.1

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., et al. (2021). EUREC^{4}A. *Earth System Science Data Discussions*, *13*(8), 4067–4119. https://doi.org/10.5194/essd-13-4067-2021

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, *6*(1), 61. https://doi.org/10.1186/s40645-019-0304-z

Su, C.-Y., Wu, C.-M., Chen, W.-T., & Chen, J.-H. (2021). Implementation of the Unified Representation of deep moist convection in the CWBGFS. *Monthly Weather Review*, *149*(10), 3525–3539. https://doi.org/10.1175/mwr-d-21-0067.1

Tomassini, L. (2018). Mesoscale circulations and organized convection in African easterly waves. *Journal of the Atmospheric Sciences*, *75*(12), 4357–4381. https://doi.org/10.1175/JAS-D-18-0183.1

Tomassini, L. (2020). The interaction between moist convection and the atmospheric circulation in the tropics. *Bulletin of the American Meteorological Society*, *101*(8), 1378–1396. https://doi.org/10.1175/bams-d-19-0180.1

Tomassini, L. (2022). Supporting data for the manuscript "Confronting the convective gray zone in the global configuration of the Met Office Unified Model" by L. Tomassini et al. [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7096991

Tomassini, L., Field, P. R., Honnert, R., Malardel, S., McTaggart-Cowan, R., Saitou, K., et al. (2017). The "grey zone" cold air outbreak global model intercomparison: A cross-evaluation using large-eddy simulations. *Journal of Advances in Modeling Earth Systems*, *9*(1), 39–64. https://doi.org/10.1002/2016MS000822

Tomassini, L., & Yang, G.-Y. (2022). Tropical moist convection an important driver of Atlantic Hadley circulation variability. *Quarterly Journal of the Royal Meteorological Society*, *148*(748), 3287–3302. https://doi.org/10.1002/qj.4359

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

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., et al. (2019). The Met Office unified model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geoscientific Model Development*, *12*(5), 1909–1963. https://doi.org/10.5194/gmd-12-1909-2019

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., et al. (2017). The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, *10*(4), 1487–1520. https://doi.org/10.5194/gmd-10-1487-2017

Wang, B., Xiang, B., & Lee, J.-Y. (2013). Subtropical high predictability establishes a promising way for monsoon and tropical storm predictions. *Proceedings of the National Academy of Sciences*, *110*(8), 2718–2722. https://doi.org/10.1073/pnas.1214626110

Wang, W. (2022). Forecasting convection with a "scale-aware" Tiedtke cumulus parameterization scheme at kilometer scales. *Weather and Forecasting*, *37*(8), 1491–1507. https://doi.org/10.1175/waf-d-21-0179.1

Willett, M. R., & Whitall, M. A. (2017). *A simple prognostic based convective entrainment rate for the unified model: Description and tests (technical report no. 617)*. Met Office.

Wood, N., Diamantakis, M., & Staniforth, A. (2007). A monotonically-damping second-order-accurate unconditionally-stable numerical scheme for diffusion. *Quarterly Journal of the Royal Meteorological Society*, *133*(627), 1559–1573. https://doi.org/10.1002/qj.116

Wyngaard, J. C. (2004). Toward numerical modelling in the "Terra Incognita". *Journal of the Atmospheric Sciences*, *61*(14), 1816–1226. https://doi.org/10.1175/1520-0469(2004)061<1816:tnmitt>2.0.co;2

Yihui, D., & Chan, J. C. (2005). The East Asian summer monsoon: An overview. *Meteorology and Atmospheric Physics*, *89*(1), 117–142. https://doi.org/10.1007/s00703-005-0125-z

Zängl, G., Reinert, D., Ripodas, P., & Baldauf, M. (2015). The ICON (ICOsahedral non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 563–579. https://doi.org/10.1002/qj.2378

Zhang, C., & Ling, J. (2017). Barrier effect of the Indo-Pacific Maritime continent on the MJO: Perspectives from tracking MJO precipitation. *Journal of Climate*, *30*(9), 3439–3459. https://doi.org/10.1175/jcli-d-16-0614.1

Zheng, Y., Alpaty, K., Herwhe, J. S., Genio, A. D., & Niyogi, D. (2016). Improving high-resolution weather forecasts using the Weather Research and Forecasting (WRF) model with an updated Kain-Fritsch scheme. *Monthly Weather Review*, *144*(3), 833–860. https://doi.org/10.1175/MWR-D-15-0005.1