



The first Met Office Unified Model–JULES Regional Atmosphere and Land configuration, RAL1

Mike Bush¹, Tom Allen¹, Caroline Bain¹, Ian Boutle¹, John Edwards¹, Anke Finnenkoetter¹, Charmaine Franklin², Kirsty Hanley¹, Humphrey Lean¹, Adrian Lock¹, James Manners¹, Marion Mittermaier¹, Cyril Morcrette¹, Rachel North¹, Jon Petch¹, Chris Short¹, Simon Vosper¹, David Walters¹, Stuart Webster¹, Mark Weeks¹, Jonathan Wilkinson¹, Nigel Wood¹, and Mohamed Zerroukat¹

¹Met Office, FitzRoy Road, Exeter, EX1 3PB, UK

²Bureau of Meteorology (BoM), Melbourne, Victoria, Australia

Correspondence: Mike Bush (mike.bush@metoffice.gov.uk)

Received: 9 May 2019 – Discussion started: 14 June 2019

Revised: 25 October 2019 – Accepted: 29 October 2019 – Published: 21 April 2020

Abstract. In this paper we define the first Regional Atmosphere and Land (RAL) science configuration for kilometre-scale modelling using the Unified Model (UM) as the basis for the atmosphere and the Joint UK Land Environment Simulator (JULES) for the land. RAL1 defines the science configuration of the dynamics and physics schemes of the atmosphere and land. This configuration will provide a model baseline for any future weather or climate model developments to be described against, and it is the intention that from this point forward significant changes to the system will be documented in the literature. This reproduces the process used for global configurations of the UM, which was first documented as a science configuration in 2011. While it is our goal to have a single defined configuration of the model that performs effectively in all regions, this has not yet been possible. Currently we define two sub-releases, one for mid-latitudes (RAL1-M) and one for tropical regions (RAL1-T). The differences between RAL1-M and RAL1-T are documented, and where appropriate we define how the model configuration relates to the corresponding configuration of the global forecasting model.

© Crown copyright 2020

1 Introduction

It is becoming standard practice for national meteorological services (NMSs) and those involved in the prediction of high-impact weather to use regional atmospheric and land models with grid lengths of the order of a kilometre as their prediction systems (e.g. Baldauf et al., 2011; Brousseau et al., 2016; Bengtsson et al., 2017; Klasa et al., 2018). While not truly resolving deep convection, kilometre-scale atmospheric models are able to explicitly represent deep convective processes within the resolved dynamics. These models provide valuable information on local weather and high-impact weather that is critical to the core function of NMSs. The representation of convective systems, topographically driven weather and various mesoscale features is generally improved with these regional modelling systems (Clark et al., 2016). In addition to weather forecasting, kilometre-scale simulations are now emerging as a tool for climate projections (e.g. Kendon et al., 2017). While there is significant computational cost to running regional models with a grid length of the order of the kilometre scale for the many long-duration runs needed for climate projections, the value of the far improved representation of weather systems, especially those related to high-impact weather, makes the computational costs worthwhile.

Over the United Kingdom, the Met Office's primary operational deterministic numerical weather prediction (NWP)

Copyright statement. The works published in this journal are distributed under the Creative Commons Attribution 4.0 License. This licence does not affect the Crown copyright work, which is reusable under the Open Government Licence (OGL). The Creative Commons Attribution 4.0 License and the OGL are interoperable and do not conflict with, reduce or limit each other.

forecast system (the UKV; Tang et al., 2013) and ensemble prediction system (MOGREPS-UK; Hagelin et al., 2017) are run with grid lengths of the order of a kilometre. These systems both use the Met Office Unified Model (UM; Brown et al., 2012) as the basis for the atmosphere and the Joint UK Land Environment Simulator (JULES; Best et al., 2011; Clark et al., 2011) for the land. They are run in variable-resolution mode, with horizontal grid lengths in the central regions of their domains of 1.5 and 2.2 km, respectively. In addition, the Met Office also carries out regional kilometre-scale simulations for climate projection, the latest of which have been run with horizontal grid lengths of 1.5 km over a domain covering the southern UK (Kendon et al., 2014), 2.2 km over Europe (Berthou et al., 2018) and 4.4 km over Africa (Stratton et al., 2018). The exact choice of grid length and domain size is a pragmatic one when the aim is to have as good a resolution as possible while allowing the forecasts or climate projections to run in the allotted time on the computer systems available.

Regional modelling in the Met Office is not confined to the UK for weather or climate. For several international collaborations and to meet various commitments, the Met Office also runs kilometre-scale UM simulations in many other regions around the world. In addition, as part of the UM partnership, a range of institutions beyond the Met Office also run the regional model in their areas of interest. With the many regions and many users of the model it has become more important than ever to coordinate its development and have clearly defined science configurations. In this paper we define the first Regional Atmosphere and Land (RAL) science configuration for kilometre-scale modelling using the UM and JULES. RAL1 defines the science configuration of the dynamics and physics schemes of the atmosphere and land. This configuration will provide a model baseline for any future weather or climate model developments to be described against. It is the intention that from this point forward significant changes to the system will be documented in the literature. This reproduces the process used for global configurations of the UM, which was first documented as a science configuration in 2011 (Walters et al., 2011).

While it is our goal to have a single defined configuration of the model that performs effectively in all regions, this has not yet been possible. Currently we define two sub-releases, one for mid-latitudes (RAL1-M) and one for tropical regions (RAL1-T). The differences between RAL1-M and RAL1-T are clearly documented within this paper. Also, where appropriate, we define how the model configuration relates to the corresponding configuration of the global forecasting model defined in Walters et al. (2019).

Prior to the existence of RAL1, there was no single definition for the configuration of the regional UM. As RAL1 is the first formally documented model configuration there is no previous baseline against which to document performance and recent developments. However, it is a goal of this paper to highlight the most recent updates and describe how

these have improved performance over previous versions of the regional UM system. To do this we focus on the UK and describe the model changes against the previous operational weather prediction system. This baseline, known in the Met Office as Operational Suite 37 (OS37), was the operational system from 15 March to 8 November 2016 and will be referred to in this paper as RAL0.

In Sect. 2, we document the RAL0 configuration. In Sect. 3, we highlight the RAL1-M developments which are added to the RAL0 baseline to define RAL1-M. In Sect. 4 we document the tropical version RAL1-T, and in Sect. 5 we evaluate the performance of the RAL1-M and RAL1-T configurations in five parts of the world with different meteorology. Finally, in Sect. 6 we provide some concluding remarks.

2 Defining Regional Atmosphere and Land – version 0 (RAL0)

2.1 Dynamical core: spatial aspects

The primary atmospheric prognostics are the three-dimensional wind components, virtual dry potential temperature, Exner pressure, dry density, five moist prognostics (mixing ratios of water vapour, liquid, ice, rain and graupel) and murk aerosol (operational UK forecasts only). These prognostic fields are discretised horizontally onto a rotated longitude–latitude grid with the pole rotated so that the grid's equator runs through the centre of the model domain. Optionally, the horizontal grid may be specified as being of variable resolution, whereby the grid size varies smoothly from coarser resolution at the outer boundaries to a uniform fine resolution in the interior of the domain as described in Tang et al. (2013). The prognostic variables are stored using Arakawa C-grid staggering (Arakawa and Lamb, 1977) in the horizontal and Charney–Phillips staggering (Charney and Phillips, 1953) in the vertical. A terrain-following hybrid height coordinate is used that it is a mix of both pure height (i.e. flat levels) and terrain-following levels (Davies et al., 2005).

In the vertical, RAL0 uses a 70-level vertical set labelled L70(61_t, 9_s)₄₀, which has 61 levels below 18 km, 9 levels above this and a fixed model lid 40 km above sea level. This naming convention was originally devised for global model simulations to denote the maximum number of levels that could be in the troposphere at its maximum depth of around 18 km (_t) as well as the number above this that would always be in the stratosphere or above (_s). As the mid-latitude tropopause is typically at a height of roughly 9–11 km, this level set concentrates its levels below 9 km, with only 20 of its 70 levels above this.

2.2 Dynamical core: spatio-temporal discretisation

RAL0 uses the UM's ENDGame dynamical core: a semi-implicit (SI) semi-Lagrangian (SL) formulation that solves

the non-hydrostatic, fully compressible, deep-atmosphere equations of motion (Wood et al., 2014). The discrete equations are solved using a nested iterative structure for each atmospheric time step within which some terms are lagged and computed in an outer loop, while others are treated quasi-fully implicitly in an inner loop.

The SL departure point equations are solved within the outer loop using a centred average of the previous time step (time level n) wind and the latest estimates for the current time step (time level $(n+1)$) wind. Appropriate fields are then interpolated to the departure points using Lagrange interpolation with various polynomial degree options. Since pointwise Lagrange interpolation is not a conservative operation, the mass of dry air, the various water species and any other transported tracers can drift due to numerical errors as well as the net fluxes through the lateral boundaries. The lack of enforcement of the correct budget of such fields in RAL0 is the motivation for a change in RAL1 to use of the zero lateral flux (ZLF) scheme of Zerroukat and Shipway (2017), which is outlined in Sect. 3.1.

Within the inner loop, a linear Helmholtz problem is solved to obtain the pressure increment in which the Coriolis, orographic and non-linear terms are evaluated as source terms to this equation: they are averaged in an off-centred semi-implicit fashion along the semi-Lagrangian trajectory using both the known state at time level n and the latest estimated (iterated) values of the fields at time level $(n+1)$. Having solved the Helmholtz problem, the other prognostic variables are obtained from the pressure increment via a back-substitution process (see Wood et al., 2014, for further details). An off-centring of 0.55 is used for all variables (where a value of 0.5 represents a centred scheme and a value of 1.0 would be a fully implicit scheme).

Imposing the lateral boundary conditions (LBCs) within the solution procedure requires special treatment, and details of this are given in Appendix A.

The physical parameterisations are split into slow processes (radiation and microphysics) and fast processes (atmospheric boundary layer turbulence, cloud and surface coupling). The slow processes are treated in parallel and computed using only the previous time level n model state. They are computed once per time step before the outer loop. The source terms from the slow processes are then added explicitly to the appropriate fields before the semi-Lagrangian advection (i.e. interpolation). The fast processes are treated sequentially and are computed in the outer loop using the latest estimate for the model state at the current time step or time level $(n+1)$ (i.e. fast process are treated approximately fully implicitly as the final state $(n+1)$ cannot be known until the end of the iteration process). A summary of the atmospheric time step is given in Algorithm 1 in Appendix A. In practice two iterations are used for each of the outer and inner loops so that the Helmholtz problem is solved four times per time step. Finally, Table 1 contains the typical length of time step used for a range of horizontal resolutions.

Table 1. Typical time step for a range of horizontal resolutions.

Radial resolution	Nominal physical resolution	Typical time step
0.0135°	1.5 km	60 s
0.02°	2.2 km	100 s
0.04°	4.4 km	100 s

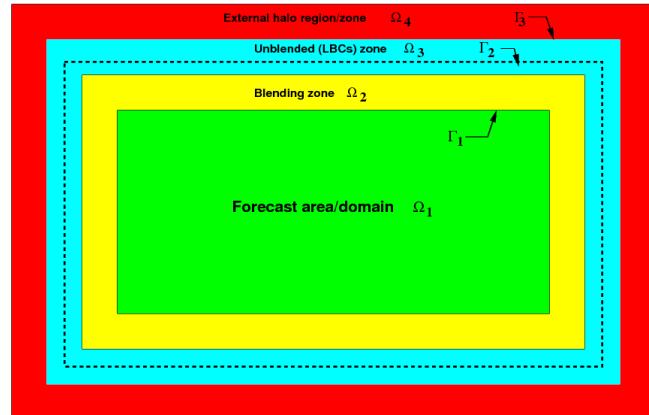


Figure 1. Schematic of the LAM configuration. In this configuration a LAM with a physical (or forecasting) region denoted by Ω_1 is shown in green. On the periphery of the forecasting area there is an extended computational domain ($\Omega_E = \Omega_2 + \Omega_3 + \Omega_4$) that includes a blending (yellow) zone Ω_2 , an unblended (blue) zone Ω_3 and an external halo (red) zone Ω_4 (which arise from the parallel domain decomposition). Note that in general the relative sizes of ($\Omega_2, \Omega_3, \Omega_4$) are a lot smaller than Ω_1 , but they are exaggerated here for clarity. Also, the use of the word RIM refers to the whole size of LBCs, which are all the grid points that lie in the region $\Omega_R = \Omega_2 + \Omega_3$ (yellow and blue).

There are a number of differences between the limited-area model (LAM) formulation of ENDGame and the global version described in Wood et al. (2014). An important one arises due to the iterative nature of the ENDGame algorithm and the requirement, in practice, of applying LBCs over the area covered by Ω_2 and Ω_3 in Fig. 1. Algorithm 1 gives an outline of a typical ENDGame time step, with the primary difference being the addition of the expected updating of the LAM LBCs at the end of each time step but also the addition of an *update-dynamics-only* LBC step during the main iteration. The main purpose of this step is to reset the new time level's velocities to be compatible with the LBCs since these will have been altered in the Helmholtz–inner loop section.

2.3 Lateral boundary conditions (LBCs)

LAMs solve the atmospheric equations on a physical domain Ω_1 subject to LBCs provided by a driving (generally a global) model, imposed on the periphery of Ω_1 (see Fig. 1). The UM's treatment of LBCs uses the method of relaxation

and blending (Davies, 1976; Perkey and Kreitzberg, 1976). The relaxation method requires the LBCs to be a data region (shown in Fig. 1 by the RIM region $\Omega_2 + \Omega_3$) with several grid points so that the driving model (or LBCs) and the LAM solutions are gradually blended to reduce wave reflections from the boundaries (Marbaix et al., 2003). Additionally, for SL models the LBCs are further extended, as a fluid parcel ending up inside the domain Ω_1 may have come from a region outside Ω_1 and far away from its boundary Γ_1 depending on the scale of the horizontal wind and the size of the time step used. The number of points defines the size of the LBCs and depends on the order of interpolation used for SL advection, the size of the blending zone and the maximum (expected) Courant number allowed (Aranami et al., 2014). The UKV model uses $\Omega_2 = 3$, $\Omega_3 = 5$ and $\Omega_4 = 7$.

The solver is identical in structure between LAM and global, with the application of the boundary conditions on the Helmholtz equation being the main difference. The pressure boundary condition is of Dirichlet type with the (hydrostatically balanced) LBC held fixed on the outermost part of Ω_3 . LBC vertical velocity is assumed to be zero, while that obtained from the inner loop will be non-zero. An implicit vertical damping profile is employed whose damping rate is proportional to the blending weights used in regions Ω_2 and Ω_3 . Not only does this help with the model imbalance but it also reduces the iteration count of the linear solver while also improving model stability.

Another difference between the LAM models and global is the calculation of trajectories (departure points) for the SL transport. The absence of the polar singularity allows for a much simpler (less computationally expensive) departure point algorithm compared to Thuburn and White (2013), and it is essentially described in Allen and Zerroukat (2016) but with the additional constraint of the departure points being clipped to Ω_3 in Fig. 1. At excessively large Courant numbers, which can occur sporadically when the jet stream intersects the lid of the model, there is the potential for the data required to interpolate the fields to be off-processor. The solution is derived from observing that for a halo width H and for cubic Lagrange interpolation, the largest westward Courant number allowable is $H - 1$, while the largest eastward Courant number is $H - 2$ and similarly for north and south. This observation allows for the introduction of a trajectory-clipping algorithm which looks at the distance of the departure point (in grid point space) from the arrival point and moves it, depending on the direction of the flow, if the distance is greater than the maximum allowable to the furthest grid point at which there would be no issues. At points that have been moved the interpolation weights are reset to 0.5 to remove any potential biases. Note that, because this calculation is performed in grid point space, the variation of the Courant number with the variable grid resolution is automatically accounted for.

2.4 Solar and terrestrial radiation

Shortwave (SW) radiation from the Sun is absorbed and reflected in the atmosphere and at the Earth's surface and provides energy to drive the atmospheric circulation. Longwave (LW) radiation is emitted from the planet and interacts with the atmosphere, redistributing heat, before being emitted into space. These processes are parameterised via the radiation scheme, which provides prognostic atmospheric temperature increments, prognostic surface fluxes and additional diagnostic fluxes. The SOCRATES (<https://code.metoffice.gov.uk/trac/socrates>, last access: 3 April 2020) radiative transfer scheme (Edwards and Slingo, 1996; Manners et al., 2018) is used with a configuration based on GA3.1 (Walters et al., 2011). Solar radiation is treated in six SW bands and thermal radiation in nine LW bands. In the LW an approximate treatment of scattering is used (Manners et al., 2018) to reduce execution time.

Gaseous absorption uses the correlated- k method with coefficients identical to the GA3.1 configuration. A total of 21 k terms are used for the major gases in the SW bands, with absorption by water vapour (H_2O), carbon dioxide (CO_2), ozone (O_3) and oxygen (O_2). A total of 33 k terms are used for the major gases in the LW bands, with absorption by H_2O , O_3 , CO_2 , CH_4 , N_2O , CFC-11 (CCl_3F) and CFC-12 (CCl_2F_2). Of the major gases considered, only H_2O is prognostic; O_3 uses a climatology, whilst other gases are prescribed using fixed mass mixing ratios and assumed to be well mixed.

Absorption and scattering by aerosols is included based on a simple climatology of five species: water-soluble, dust, oceanic, soot and stratospheric aerosols. The component in the planetary boundary layer is distributed over approximately 3.2 km of the atmosphere (lowest 30 model levels), and the contribution from dust has been scaled by 0.3333 compared to the original climatology of Cusack et al. (1998) as the dust loading of the basic climatology over land (which includes arid areas) is too high for the UK.

The parameterisation of cloud droplets is described in Edwards and Slingo (1996) using the method of “thick averaging”. Padé fits are used for the variation with effective radius, which is computed from the number of cloud droplets calculated in the microphysics scheme (see Sect. 2.5). The parameterisation of ice crystals is described in Baran et al. (2016).

The sub-grid cloud structure is represented using separate cloud fractions for the liquid and ice components, with the liquid water mass mixing ratio scaled by a factor of 0.7 to represent the effect of cloud inhomogeneity as described in Cahalan et al. (1994). Cloud fractions in adjacent layers in the vertical are maximally overlapped, while clouds separated by clear sky are randomly overlapped. Full radiation calculations are made every 15 min using the instantaneous cloud fields and a mean solar zenith angle for the following 15 min period. Corrections for the change in solar zenith an-

gle on every model time step and the change in cloud fields every 5 min are made as described in Manners et al. (2009).

The emissivity and the albedo of the surface are set by the JULES land surface model (see Sect. 2.8). A single frequency-averaged emissivity is specified for each surface type (see Walters et al., 2014, for the numerical values). For the surface albedo, the radiative transfer in plant canopies uses the two-stream radiation scheme and spectral parameters of Sellers (1985).

The direct SW flux at the surface is corrected for the angle and aspect of the topographic slope and for shading by surrounding terrain. The net LW flux at the surface is corrected for the resolved sky-view factor due to the surrounding terrain (Manners et al., 2012).

2.5 Microphysics

The formation and evolution of precipitation due to grid-scale processes are the responsibility of the microphysics scheme. The microphysics scheme has prognostic input fields of temperature, moisture, cloud and precipitation from the end of the previous time step, which it modifies in turn. The microphysics used is a single-moment scheme based on Wilson and Ballard (1999), with extensive modifications. We make use of prognostic rain, which allows three-dimensional advection of the rain mass mixing ratio. This has been shown to improve precipitation distributions over and around mountainous regions, especially with the smaller grid spacings used in the RAL configurations (Lean et al., 2008; Lean and Browning, 2013). Prognostic graupel has also been included, and this allows for the explicit representation of a second, more dense ice category which is useful for hail forecasting at kilometre-scale resolutions as well as being a prerequisite for lightning forecasting (Wilkinson and Bornemann, 2014).

The warm-rain scheme is based on Boutle et al. (2014b) and includes an explicit representation of the effect of sub-grid variability on autoconversion and accretion rates (Boutle et al., 2014a). We use the rain-rate-dependent particle size distribution of Abel and Boutle (2012) and fall velocities of Abel and Shipway (2007), which combine to allow a better representation of the sedimentation and evaporation of small droplets. The cloud droplet number concentration can be determined from assuming either (a) a fixed climatological aerosol or (b) using a single-species prognostic aerosol which has been developed for forecasts of visibility (Clark et al., 2008). For the cases in which single-species prognostic aerosol is used, the aerosol concentrations are coupled to the cloud drop number using the methodology described in Wilkinson et al. (2013) and modified following Osborne et al. (2014). In the case of the fixed climatological aerosol, the parameterisation of Jones et al. (1994) is used. In both cases, droplet numbers are reduced near the surface for effective fog simulation, and changes included in RAL1 are described in Sect. 3.3.

Ice cloud parameterisations use the generic size distribution of Field et al. (2007) and mass-diameter relations of Cotton et al. (2013). The fall speed of ice used is the dual fall speed as described in Furtado et al. (2015), wherein the lowest value of two computed fall speed relations is used. This represents the fact that the Field et al. (2007) parameterisation includes contributions from both smaller ice crystals and larger ice aggregates.

Unlike the GA configurations, there is no requirement for multiple sub-time stepping of the microphysics scheme as the model time step in the RAL configurations is shorter than the 2 min period used as a sub-time step in the GA configurations.

As in Stratton et al. (2018), the output taken immediately after the microphysics scheme drives a lightning parameterisation based on McCaul et al. (2009), with the discharge of lightning flashes in the column being determined as described in Appendix A of Wilkinson (2017). This has been shown to be of benefit for a high-profile event (Wilkinson and Bornemann, 2014) and to perform well during the summer months (Wilkinson, 2017).

2.6 Large-scale cloud

Due to sub-grid inhomogeneity, clouds will form well before the humidity averaged over the size of a grid box reaches saturation, and this is still true when the grid box size is at the kilometre scale (Boutle et al., 2016). A cloud parameterisation scheme is therefore required to determine the fraction of the grid box which is covered by cloud and the amount and phase of condensed water contained in those clouds. The formation of clouds will convert water vapour to liquid or ice and release latent heat. The cloud cover and liquid and ice water contents are then used by the radiation scheme to calculate the radiative impact of the clouds and by the microphysics scheme to calculate whether any precipitation has formed.

RAL0 uses the Smith (1990) cloud scheme. This is a diagnostic scheme, in which the cloud cover is calculated only from information available at that moment in time. The scheme relies on a definition of critical relative humidity, RH_{crit} , which is the grid box mean relative humidity at which clouds start to appear. The value of RH_{crit} is set to 0.96 at the surface and decreases monotonically to 0.80 at 850 m (model level 15). It is then held fixed above that.

For liquid cloud, the Smith cloud scheme is built around an assumption that sub-grid temperature and humidity fluctuations can be described by a symmetric triangular probability distribution function (PDF). One consequence of this PDF assumption is that the grid box has 50 % cloud cover when the total relative humidity, $RH_t = (q_v + q_{cl})/q_{sat}$ (where q_v is the vapour, q_{cl} is the liquid content and q_{sat} is the saturation specific humidity), reaches 100 % and that the grid box only becomes overcast when $RH_t >= 2 - RH_{crit}$. However, observations such as in Wood and Field (2000) suggest that the

cloud fraction should be larger than 0.5 when $RH_t = 100\%$. As a result, an empirically adjusted cloud fraction (EACF) is used with the Smith scheme in kilometre-scale models. The relative humidity at which cloud first appears is unchanged, but the smooth function linking cloud fraction to relative humidity increases more rapidly so that cloud fraction is 0.70 when $RH_t = 100\%$.

Forecasts using the EACF still underestimate cloudiness, however, especially the thin clouds forming below a temperature inversion that do not fill the entire depth of a model layer. So an area cloud fraction scheme is also used, which follows a similar approach to that described by Boutle and Morcrette (2010). Each model level is split into three and vertical interpolation is used to find the thermodynamic values in the sub-layers. However, if there is a strong gradient in RH due to the presence of a capping inversion, the thermodynamic properties of the sub-layer are found by extrapolation from above and below instead. This sharpens the inversion and can increase the RH in the sub-layers below it. The Smith cloud scheme, itself modified to use the EACF, is then called on each of the three sub-layers. The cloud fraction for use by the microphysics is set to the mean of the cloud fractions over the three sub-layers, while the cloud fraction seen by radiation is set to the maximum of the values from the three sub-layers.

The ice cloud fraction is parameterised as described by Abel et al. (2017) wherein it is diagnosed from the ice water content.

2.7 Atmospheric boundary layer

The parameterisation of turbulent motions in kilometre-scale models requires special treatment because, although most turbulent motions are still unresolved, the largest scales can be of a similar size as the grid length. The model must therefore be able to parameterise the smaller scales, resolve the largest ones if possible and not alias turbulent motions smaller than the grid scale onto the grid scale. The “blended” boundary layer parameterisation described by Boutle et al. (2014b) is used to achieve this. This scheme transitions from the 1-D vertical turbulent mixing scheme of Lock et al. (2000), suitable for low-resolution simulations such as GA configurations, to a 3-D turbulent mixing scheme based on Smagorinsky (1963) suitable for high-resolution simulations based on the ratio of the grid length to a turbulent length scale. The blended eddy diffusivity, including any non-local contribution from the Lock et al. (2000) scheme, is applied to down-gradient mixing in all three dimensions, whilst appropriately weighted non-local fluxes of heat and momentum are retained in the vertical for unstable boundary layers. The configuration of the Lock et al. (2000) scheme is similar to that of GA7 (Walters et al., 2019), with differences as follows: (i) for stable boundary layers, the “sharp” function is used everywhere but with a parameterisation of sub-grid drainage flows dependent on the sub-grid orography

(Lock, 2012); (ii) heating generated by frictional dissipation of turbulence is not represented; and (iii) the parameterisation of shear-generated turbulence extending into cumulus layers (Bodas-Salcedo et al., 2012) is not used.

The functions that are used to include the effects of stability on turbulence, via the Richardson number (Ri), follow Brown (1999):

$$f_m = (1 - c_{LEM} Ri)^{1/2}, \quad f_h = \frac{1}{Pr_N} (1 - b_{LEM} Ri)^{1/2}, \quad (1)$$

where Pr_N is the neutral Prandtl number (0.7). In RAL0, the constants b_{LEM} and c_{LEM} are both equal to 1.43 (which gives Brown’s “conventional” model). RAL0 uses a mixing length that is a fraction (0.15) of the depth of any layer in which Ri is less than a critical value ($Ri_{crit} = 0.25$) within that layer or 40 m if larger.

In an effort to improve the triggering of explicit convection, stochastic perturbations to temperature are applied. Designed to represent realistic variability that might be seen due to large boundary layer eddies, the perturbation scale for potential temperature, θ , is taken as $\theta_* = \overline{w'\theta'}|_s/w_m$, where $\overline{w'\theta'}|_s$ is the surface turbulent flux of θ , and the turbulence velocity scale w_m is given by $w_m^3 = u_*^3 + c_{ws} w_*^3$. Here, u_* is the friction velocity and w_* the convective velocity scale, with $c_{ws} = 0.25$. Finally θ_* is constrained to be positive and less than 1 K. Loosely based on Munoz-Esparza et al. (2014), the random number field that multiplies the perturbation scale is held constant over eight grid length squares in the horizontal and the perturbations are applied uniformly in the vertical up to the lower of two-thirds of the boundary layer depth and 400 m.

2.8 Land surface and hydrology

Exchanges of mass, momentum and energy between the atmosphere and the underlying land and sea surfaces are represented using the community land surface model JULES (Best et al., 2011; Clark et al., 2011). The configuration adopted in RAL0 largely follows that of GL7.0, as described by Walters et al. (2019). In keeping with the seamless approach to model development, the aim is to minimise the differences between configurations, but different developmental priorities for regional and global modelling can result in differences between the configurations, even if there is no compelling scientific motivation to maintain them. We now list and explain the non-trivial differences.

Because the UKV was developed for short-range forecasting over the UK, the treatment of surface exchange over sea and sea ice has been less of a priority than in the global model so that RAL0 is less advanced in its treatment. A fixed value of Charnock’s coefficient (0.011) is used to determine the surface roughness over open sea, as opposed to the COARE algorithm in GL7.0. GL7.0 also includes a more advanced parameterisation of the sea surface albedo (Jin et al., 2011) that incorporates a dependence on the wind speed and chloro-

phyll concentration. This has not yet been introduced into RAL0, which still uses an earlier scheme based on Barker and Li (1995). Similarly, because the regional model has not yet been used operationally over sea ice, several recent modifications to sea ice parameters have not yet been introduced into the regional configuration. Sea ice is not present in the simulations shown below, so these settings are not relevant to any results presented here.

Although both GL7.0 and RAL0 include the multilayer snow scheme, different densities of fresh snow are specified: in GL7.0 the value is 109 kg m^{-3} , while in RAL0 a value of 170 kg m^{-3} is used as more representative of the conditions in the UK. In the future, it is hoped that it will be possible to relate the density to local meteorological conditions.

Both GL7.0 and RAL0 represent the radiative transfer in plant canopies using the two-stream radiation scheme of Sellers (1985), with the leaf-level reflection and transmission coefficients presented in that paper. However, in GL7.0, an adjustment to these parameters is made as the model runs to make the grid box mean albedo agree more closely with a climatology derived from GlobAlbedo. While developing this adjustment for GL7.0, the simulated direct albedos were found to be unrealistic and the diffuse albedos were used for both the direct and diffuse beams. As implemented in RAL0, there is no adjustment to a climatology and both the direct and diffuse albedos are used. Further discussion of these issues may be found in Sect. 3.5.

Two differences in soil hydrology should be noted. Whereas the more elaborate TOPMODEL scheme is used to represent soil moisture heterogeneity in GL7.0, the simpler PDM scheme is used in RAL0 (consult Best et al., 2011, for details of these schemes). Also, in RAL0, if the simulated soil moisture rises above the saturated water content, the excess is assumed to move upwards and to contribute to surface runoff. This is considered more realistic than the alternative of routing the excess moisture downwards, except in regions of partially frozen soils (Best et al., 2011). In GL7.0 the excess moisture is routed downwards.

In GL7.0 urban surfaces are represented by a single urban tile, but in RAL0 two separate tiles for street canyons and roofs are used (Porson et al., 2010). Currently the two-tile scheme is limited to domains over the UK due to the availability of morphology data.

2.9 Lower boundary condition (ancillary files) and forcing data

In the UM, the characteristics of the lower boundary, the values of climatological fields, and the distribution of natural and anthropogenic emissions are specified using ancillary files. The use of correct ancillary file inputs can potentially play as important a role in the performance of a system as the correct choice of many options in the parameterisations described above. In the future we may consider the source data and processing required to create ancillaries to be part

of the definition of the RAL configurations, as is the case in global configurations. However, we currently leave ancillaries outside the formal definition of RAL as there has been no systematic evaluation of the impact on performance of different ancillary file inputs, and the existence of many country-specific datasets (that are of better quality or higher resolution) means that different applications (especially operational ones such as UKV/MOGREPS-UK) use different source datasets, sometimes even combining different datasets within the model domain. An example of this is described in Sect. 3.5.

Table A1 in the Appendix contains the main ancillaries used in RAL applications as well as references to the source data from which they are created.

2.10 Other differences from GA7 due to horizontal resolution

The high horizontal resolutions used for RAL simulations mean that RAL0 runs with the convection parameterisation switched off, relying on the model dynamics to explicitly represent convective clouds. Although it is acknowledged that not all types of convection are represented with such grid spacing, this choice was made in the current absence of a scale-aware convection scheme which correctly parameterises sub-grid convective motion and hands over to the model dynamics for clouds larger than the model filter scale. Projects are underway to develop convection schemes for use in atmospheric models at all resolutions with grid spacings $\mathcal{O}(1\text{--}100 \text{ km})$, which could be incorporated into a future RAL release.

Also, RAL0 does not include a sub-grid parameterisation scheme for either orographic or non-orographically forced gravity waves. However, for those non-UK-area models that run with a grid length of 0.04° (4.4 km), the inclusion of the effective roughness and gravity wave drag schemes (both as used in GA7) was found to be beneficial to near-surface verification scores.

3 Developments included in RAL1

This section describes the RAL1-M developments which when added to the RAL0 base define RAL1-M. The Regional Model Evaluation and Development (RMED) processes at the Met Office makes use of an online “ticket” tracking system which allows scientists to document changes to the model. A ticket number is assigned to each model development, and thus it is clear to all developers and external collaborators which tickets are included in any one configuration. In this section we discuss the major developments to RAL1 and reference them by ticket number to inform the development community and for future cross-reference. For ease of reference, a complete list of all the RMED tickets included in RAL1 can be found in Table A3.

3.1 Dynamical formulation and discretisation

Conservative advection for moist prognostics (RMED ticket no. 2)

Mass conservation for mixing ratios is achieved with the ZLF (zero lateral flux) scheme (Zerroukat and Shipway, 2017). This scheme is computationally efficient and exploits the relatively large width of LBCs used for semi-Lagrangian-based LAMs (i.e. the size of the extra extended computational zone $\Omega_E = \Omega_2 + \Omega_3 + \Omega_4$ shown in Fig. 1). Assuming that the size of Ω_E is sufficiently large (> 2 points) it can be divided into two regions as shown in Fig. 1 with a dotted-line boundary Γ_2 , which will be referred to as the ZLF boundary. It is also very common that the wind and the time step used are such that the horizontal Courant number in the RIM zone is smaller than half of the RIM size. Under these conditions, the SL advection solution for all the points inside the region $\{\Omega_1 + \Omega_2\}$ (which includes the forecasting zone) will be unaffected by the field beyond the ZLF boundary Γ_2 . Therefore, for convenience, the advection solves a modified problem, whereby inside Γ_2 the advected quantity is the original field, whereas the field beyond Γ_2 is zeroed. This modification does not affect the solution inside the domain $\{\Omega_1 + \Omega_2\}$, and hence it is equivalent to the original problem. However, this modification allows us to impose a simple mass conservation constraint over the whole extended computational domain $\{\Omega_1 + \Omega_2 + \Omega_3\}$ where there is no need to compute lateral fluxes because they are zero by construction (see details in Zerroukat and Shipway, 2017). This is quite an important simplification from the case in which one would like to impose a mass conservation budget for the forecast and/or physical domain Ω_1 , which requires knowledge of mass fluxes through its lateral boundary Γ_1 that are complicated and computationally expensive to compute (Aranami et al., 2014). The ZLF scheme has two components: the first part (just explained above) allows us to avoid computing expensive lateral fluxes, while the second part is the redistribution of the mass conservation error using the optimised conservation scheme (Zerroukat and Allen, 2015). Note that the zeroing is just an intermediate temporary step used during the advection because the zeroed region gets overwritten by the appropriate LBC data at the end of the time step.

3.2 Solar and terrestrial radiation

Improved treatment of gaseous absorption (RMED ticket no. 9)

The treatment of gaseous absorption has been significantly updated to the configuration used with GA7 (Walters et al., 2019).

A total of 41 k terms are used for the major gases in the SW bands, with an improved representation of H_2O , CO_2 , O_3 and O_2 absorption as well as the addition of absorption from

nitrous oxide (N_2O) and methane (CH_4). These changes result in increased atmospheric absorption and reduced surface (clear-sky) fluxes.

A total of 81 k terms are used for the major gases in the LW bands, with an improved representation of all gases. This results in reduced clear-sky outgoing LW radiation and increased downwards surface fluxes.

The method of “hybrid” scattering is used in the LW, which runs full scattering calculations for 27 of the major gas k terms (where their nominal optical depth is less than 10 in a mid-latitude summer atmosphere). For the remaining 54 k terms (optical depth > 10) much cheaper non-scattering calculations are run.

In both spectral regions the band-by-band breakdown of absorption is improved, which should improve interaction with band-by-band aerosol and cloud forcing.

3.3 Microphysics

Improved droplet number profile in the lower boundary layer (RMED ticket no. 1)

Previous work by Wilkinson et al. (2013) discussed a pragmatic method of reducing the cloud droplet number near the surface, often referred to as a “droplet taper”. This reduction accounts for the fact that aerosol activation and cloud droplet numbers measured in fog are often much lower than those found in more elevated clouds, despite the fact that the underlying aerosol concentrations are generally higher. Recent work by Boutle et al. (2018) utilising new observations (Price et al., 2018) has enhanced our understanding of this process, demonstrating that weak updraughts and low supersaturations in fog are the reason for the limited aerosol activation. Boutle et al. (2018) showed that even the droplet number profile of Wilkinson et al. (2013) gave values too high too close to the surface. This triggered a feedback process whereby fog became too deep and well developed too quickly, resulting in significant errors to fog forecasts. Boutle et al. (2018) proposed a modified parameterisation for the near-surface droplet number, which was shown in forecast trials to be of significant benefit. Therefore, RAL1 has adopted the droplet number parameterisation proposed by Boutle et al. (2018); i.e. droplet numbers are held at 50 cm^{-3} throughout the lowest 50 m of the atmosphere, before transitioning to the cloudy values as described in Wilkinson et al. (2013). We note that this is still a pragmatic choice based on model performance, and further work is required to develop an activation scheme which correctly accounts for aerosol effects and is valid in the foggy regime.

3.4 Atmospheric boundary layer

3.4.1 Updates to stochastic boundary layer perturbations (RMED ticket no. 25)

Several updates were made to the stochastic perturbations in the boundary layer (described in Sect. 2.7) for RAL1 in order to further enhance the triggering of convective activity. The first was also to apply the perturbations to specific humidity using the same formulation for the perturbation scale (based on the surface humidity flux) and constraining the moisture scale to be less than 10 % of the specific humidity itself. Secondly, the random number field was changed from being randomly different every time step to being updated in time following McCabe et al. (2016) using a first-order autoregression model with the autocorrelation coefficient set to give a decorrelation timescale of 600 s, an approximate eddy-turnover timescale. This temporal coherence of the perturbations results in a greater resolved-scale dynamical response. Finally, in the vertical the perturbations are now scaled by a piecewise linear “shape” function equal to unity in the middle of the boundary layer and zero at the surface and top of the sub-cloud layer; this is only applied where a cumulus regime is diagnosed (see Lock et al., 2000). These were pragmatic changes to avoid the perturbations strongly influencing the screen-level temperature diagnostic, which had been found to lead to the degradation of deterministic measures of skill (such a root mean square error).

3.4.2 Revision of free-atmospheric mixing length (RMED ticket no. 12)

In RAL1-M, the free-tropospheric (i.e. above the boundary layer) mixing length is reduced everywhere to its minimum value of 40 m, which was found to give better, more rapid initiation of showers in UKV than the interactive mixing length used in RAL0 (and also kept for RAL1-T; see Sect. 4).

3.4.3 Improved representation of mixing across the boundary layer top (RMED ticket no. 5)

This ticket allows the boundary layer scheme’s explicit entrainment parameterisation to be distributed over a vertically resolved inversion layer, instead of always assuming the inversion to be sub-grid. As a result it allows a smoother transition in the vertical between the boundary layer and free troposphere. More details are given in Walters et al. (2019) (under GA ticket no. 83), noting that the additional representation of “forced cumulus clouds” within a resolved inversion is only included in RAL1-T (see Sect. 4) as that requires the PC2 cloud scheme to be used.

3.4.4 Reductions in sensitivity to vertical resolution (RMED ticket no. 10)

The turbulent mixing and entrainment in cloud-capped boundary layers in the Lock et al. (2000) scheme are parameterised in terms of (among other things) the strength of cloud-top radiative cooling. This is calculated by differencing the radiative flux across the top grid levels of the cloud layer. The complexity of the calculation is increased by making allowance for changes in the height of cloud between radiation calculations (which are not performed on every model time step for reasons of computational efficiency). A new methodology is introduced that identifies where the LW radiative cooling profile transitions from free-tropospheric rates above the cloud to stronger rates within it. It has very little impact at current vertical resolutions (typically greater than 100 m) but has been demonstrated in the single-column version of the model to be robustly resolution independent down to grids of only a few metres.

3.5 Land surface and hydrology

Improvements to land usage and vegetation properties (RMED ticket no. 3)

There are four changes to the representation of the land surface in RAL1.

1. There are updated land use mappings, mainly removing small (< 0.2) bare soil tile fractions from land use categories such as grassland. For UK areas the non-UK source data are changed from IGBP to CCI (for more details, see Table A1). For operational UKV purposes, though, the IGBP land mask is retained to reduce downstream impacts.
2. There is a reduction in the bare soil fraction of short vegetation tiles (given by $F = e^{-k_{ext} \cdot LAI}$, where LAI is the leaf area index) by increasing k_{ext} from 0.5 to 1.
3. There is a reduction in the scalar roughness lengths for the grass tiles, by reducing its ratio to the momentum roughness from 0.1 to 0.01. This enhances the difference between skin and near-surface air temperatures.
4. There are modifications to the canopy radiation model. Two modifications were made in the canopy radiation model. The treatment of direct solar radiation described in Sellers (1985) and originally implemented in JULES applies only in the case of isotropic scattering. It was therefore extended in RAL1 to account for the non-isotropic scattering of direct radiation. Following the assessment of Lawrence et al. (2011), using the CLM4 model, that the leaf-level near-infrared reflection coefficients given by Sellers (1985) for grass are too high, the leaf-level transmission and reflection coefficients for all plant canopies were reviewed and modified. The main

effect of these changes was to reduce the near-infrared albedo of short vegetation, thus increasing daytime temperatures.

The most significant of all these changes is the increase in vegetation cover at the expense of bare soil, which is a combination of (1) and (2). This provides more insulation between the atmosphere and the underlying soil, which results in more rapid evolution of surface and near-surface air temperatures, especially across the diurnal cycle, and a reduction in the diurnal temperature range of the upper soil levels. Both are found to give improved agreement with in situ observations.

4 The tropical configuration RAL1-T

In Sects. 2 and 3 we have described the mid-latitude subversion of RAL1. In this section we describe the tropical subversion of RAL1 known as RAL1-T. Ideally we would prefer to have one configuration for use anywhere in the world, and this is an aspiration for the future. With current parameterisations, however, we find we need two configurations to get good performance in the two different areas.

One of the major reasons why we need two configurations is that convection is sometimes very under-resolved in the UK in kilometre-scale models, particularly in cases of small, shallow showers. This can manifest itself as small showers initiating too late or not at all. In order to cope with this, RAL1-M has relatively weak turbulent mixing and stochastic perturbations to encourage the model fields to be less uniform and help convection initiate. If the model is run with these in the tropics the model initiates too early and convective cells tend to be too small.

4.1 Representation of turbulence (RMED tickets no. 12 and no. 26) and BL stochastic perturbations (RMED ticket no. 25)

There are two differences in the representation of turbulence between RAL1-M and RAL1-T, namely in the form of the stability functions and in the free-atmospheric mixing length. Both give enhanced turbulent mixing in RAL1-T compared to RAL1-M. RAL1-T uses the Brown (1999) “standard” model, whilst RAL1-M uses the Brown (1999) “conventional” model. RAL1-T retains RAL0’s interactive free-atmospheric mixing length, whilst RAL1-M uses a value of 40 m. The other related change is that RAL1-T does not use the stochastic boundary layer perturbations. For more details and a summary of differences between RAL1-T and RAL1-M, see Table 2.

4.2 Improvements to cloud scheme (RMED ticket no. 16)

RAL1-T has three extra prognostic fields (liquid fraction, ice fraction and mixed-phase fraction) as it uses the prognos-

tic cloud prognostic condensate (PC2) cloud scheme (Wilson et al., 2008a). PC2 calculates sources and sinks of cloud cover and condensate and advects the updated cloud fields, hence adding some memory into the system. One advantage of PC2 over the Smith schemes is the looser coupling between variables, hence allowing a cloud to deplete its liquid water content while maintaining high cloud cover. The PC2 scheme performs better than the Smith scheme in climate simulations (Wilson et al., 2008b) and for global numerical weather prediction (Morcrette et al., 2012). It is worth noting that when run in a model using a convection scheme, the detrainment of cloud from convection is a key source of cloudiness (Morcrette and Petch, 2010; Morcrette, 2012b). When run in a model without a convection scheme (such as the RAL configuration), cloud formation from convective motions will be represented by a combination of PC2 initialisation (near the convective cloud base), followed by PC2 pressure forcing through the rest of the updraught. In the PC2 scheme, cloud erosion is a process that accounts for the evaporation and reduction of cloud cover due to unresolved mixing near cloud edges. In the original implementation of PC2 (Wilson et al., 2008a) erosion was carried out as part of the call to the convection scheme, but in RAL1, which has no call to the convection scheme, the erosion process has been moved to occur within the microphysics scheme. In RAL1-T, the PC2 scheme is implemented as in the GA7 global model configuration (Walters et al., 2019). That is, the formulation of cloud erosion accounts for the apparent randomness of cloud fields, as described in Morcrette (2012a), and the RH_{crit} is calculated from the turbulent kinetic energy (Van Weverberg et al., 2016).

Another difference, particularly affecting convection in the tropics, is that the tropopause is deeper than in mid-latitudes. In order to take account of this RAL1-T uses a vertical level set labelled L80(59_t; 21_s)38.5, which adds some additional vertical resolution in the tropical upper troposphere at the expense of resolution in the lower boundary layer.

Figure 2 illustrates the above discussion by showing the effect of running RAL1-M and RAL1-T for a case of small showers in the UK. Unlike RAL1-M, when compared to the radar RAL1-T initiates too late and produces too few showers that are too large.

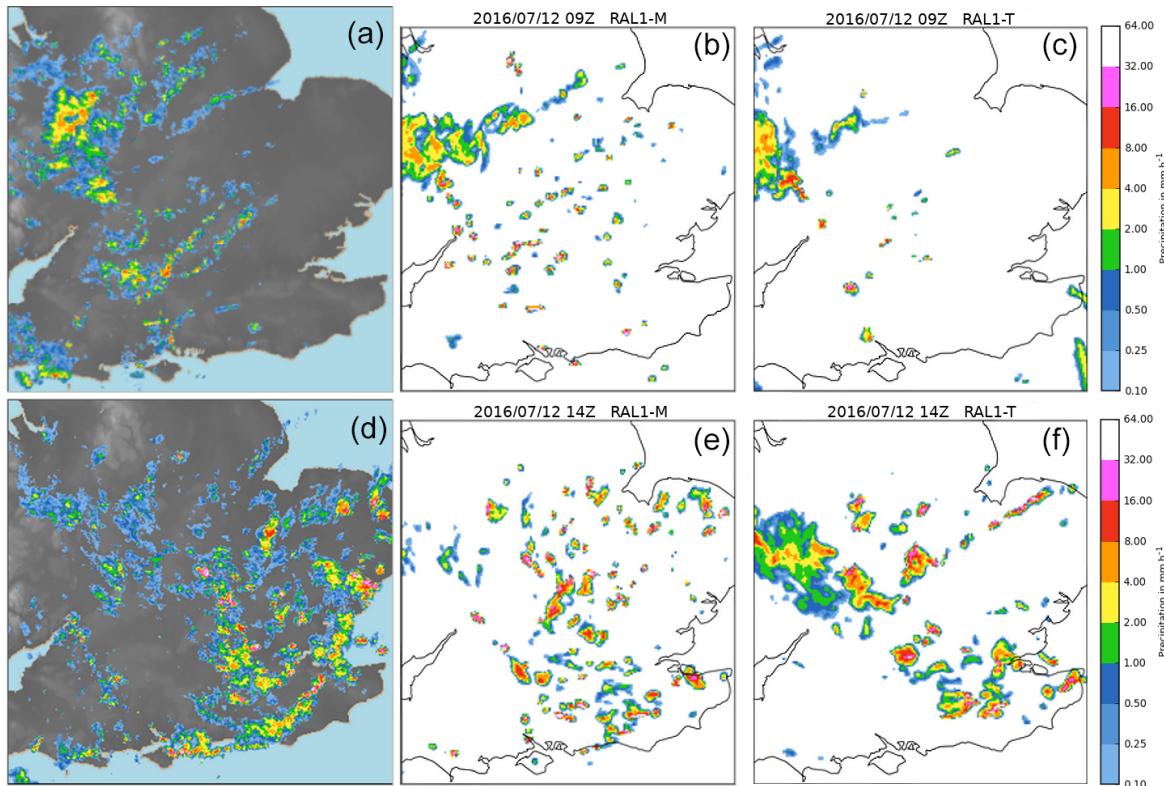
5 Model evaluation

In this section we apply a range of evaluation methods to demonstrate the performance of RAL1. The regional model evaluation process is rapidly evolving and has already benefitted from the multi-institutional UM partnership. The regional model is run by UM partners in a variety of domains worldwide, and RAL1 marks a baseline to which all centres can now focus future evaluation effort.

In this first documentation of the regional model we have focused on the performance of RAL1 over the UK, Singa-

Table 2. RAL1-M and RAL1-T differences.

RMED ticket	Science difference	RAL1-M	RAL1-T
12	BL free-atmospheric mixing length	40 m	interactive mixing length
26	BL stability functions	$b_{LEM} = 1.43, c_{LEM} = 1.43$	$b_{LEM} = 40, c_{LEM} = 16$
25	BL stochastic perturbations to temperature and moisture	on (improved triggering)	off
16	Cloud scheme	Smith (diagnostic)	PC2 (prognostic)

**Figure 2.** A 12 July 2016 case study showing radar (a, d), RAL1-M (b, e) and RAL1-T (c, f) at 09:00 Z (a, b, c) and 14:00 Z (d, e, f) for a case of showers in the UK.

pore, Australia, the western North Pacific (Philippine Area of Responsibility for tropical cyclone forecasting) and the USA. This allows for the inspection of model behaviour in a variety of climatic zones and for different weather phenomena.

A range of evaluation methods are required to assess the performance of models. Verification skill scores, anomaly plots and case studies all provide useful information which builds a picture of model characteristics and skill. Kilometre-scale models behave and look differently to models for which the convection is parameterised. Convection in these models is more likely to look realistic than in a global (parameterised) model and may mimic many of the characteristics seen in satellite images and animations. However, although the detail looks realistic, it may not always be skilful. It is a challenge to create metrics which can truthfully represent the benefit of kilometre-scale models as well as clarify their limitations.

Mittermaier (2014) proposed a new spatial and inherently probabilistic framework for evaluating kilometre-scale models, and Mittermaier and Csima (2017) provide a historical overview of the performance of the 1.5 km model using this new high-resolution assessment (HiRA) framework. The framework uses synoptic observations, but instead of using the single nearest model grid point, it uses a neighbourhood of model grid points centred on the observing location to acknowledge the fact that added detail may not be in the right place at the right time. These points can be treated as a pseudo-ensemble, and we can compute ensemble metrics as it can be assumed that all the forecast values in the neighbourhood are equally likely outcomes at the observing location. One caveat to ensure this assumption holds is that the neighbourhood must not be too large. The framework can be applied to deterministic and ensemble forecasts, including the control member of the ensemble. Whilst it may be

less than intuitive to think that a forecast neighbourhood is required for temperature, it was shown in Mittermaier and Csima (2017) that all variables benefited from the use of at least a 3×3 neighbourhood, but that neighbourhoods which are too large may be detrimental for some variables, including temperature. The HiRA ranked probability score (RPS) is used for non-normally distributed or spatially discrete variables, whilst the continuous ranked probability score (CRPS) is used for temperature.

The fractions skill score (FSS; Roberts and Lean, 2008) requires spatial observation-based analysis. Over the UK this is a radar-based analysis, though more recently a GPM-based product (Skofronick-Jackson et al., 2017) has also been used for evaluating kilometre-grid-scale configurations in the tropics. Analyses based on remotely sensed data may not be accurate in an absolute sense (no observations are perfect and error-free). The FSS is sensitive to the bias (Mittermaier and Roberts, 2010), and for this reason the FSS is generally used in conjunction with percentile thresholds, whereby all the values in the forecast and analysis domains are ranked separately, and the physical value associated with a specific centile is extracted. This quantile transformation removes the bias so that the FSS based on percentile thresholds offers a measure of field texture, pattern and areal extent but *not* intensity.

5.1 Introducing the RMED “toolbox”

To assist the RMED processes, an evaluation toolbox has been created to support model development. The main purpose is to ensure uniformity of the verification and diagnostic output across multiple users and institutions. Version 1 of the toolbox was released in time for the RAL1 assessment. It contains a selection of verification techniques and diagnostic tools with the intention of enabling the comparison with point observations as well as gridded truth sources. One of the outputs of the toolbox is a “scorecard” – a single clear plot with arrows and triangles showing whether the model version being tested is better or worse than a previous incarnation. Triangles pointing upward (green) indicate that the test model is better than the control, and downward (purple) triangles indicate the control model is better. The area of the triangles is proportional to the absolute improvement (or deterioration) of the model, and the triangles are outlined in black if the change is statistically significant at the 0.05 level determined using the Wilcoxon signed-rank test. The maximum triangle size, which occurs when the length of the base of the triangle is equal to the size of the square in which it is contained, is either set automatically (by selecting the maximum difference value from the data being compared) or can be done by manually setting a limit. The figures in this paper have the “max” values set automatically for each model comparison. The scorecards contain a huge amount of information digested into an easy-to-understand summary. This allows fast assessments about model skill to be made,

speeding up the evaluation (and therefore the development) process. The model verification plotting comprises the FSS (score with spatial scale, score with forecast lead time, accumulation equivalent to particular centile with forecast lead time) and HiRA scores including bias (score with neighbourhood size, score with forecast lead time). Plotting of more traditional metrics (e.g. mean error and root mean square error (RMSE) at a grid point) was also included for a range of parameters (surface temperature, wind, relative humidity and 6-hourly precipitation amounts).

The diagnostic methods implemented in RMED toolbox version 1 also included domain (area) average plots (for a comprehensive set of meteorological diagnostics), which are especially useful for considering the diurnal cycle, histograms (for parameters such as screen temperature, wind, 3 h mean rain rates and outgoing longwave radiation) for exploring distributions, and “cell statistics” (Hanley et al., 2015), a method for investigating the texture of a field through the application of a threshold to identify areas of exceedance or “cells”. The number and size of the cells can then be analysed. This was first implemented to compare 3-hourly mean rain rates against GPM IMERG satellite data (Huffman, 2015, 2017) or, if appropriate, UK radar data. The ability to create charts of model fields for a specific set of meteorological variables was also provided.

RAL1 provides a lot of detail due to its use at high resolution, but this can increase noise in traditional verification measures such as the root mean square error, which favours smooth fields over noisy ones. Multiple scores for the same parameter can be a source of confusion, providing different or even contradictory results. The RPS and FSS both evaluate hourly precipitation, but they measure different attributes of the precipitation forecast. The FSS measures pattern, and the HiRA RPS focuses on intensity. It is possible to improve the forecast intensities whilst degrading the spatial pattern or texture of the forecast, and this can lead to verification scores that are difficult to interpret. Murphy and Winkler (1987) stated the need for more than one independent score measuring a range of forecast attributes to get a robust perspective of forecast performance.

5.2 Mid-latitude performance over the UK

In this section we illustrate the impact of the RAL1 changes on model performance. The baseline used for the UK and mid-latitudes is RAL0. The UK evaluation consisted of a hierarchy of testing. Firstly, individual science changes (RMED tickets) were tested by running 100 case studies with a 1.5 km horizontal grid length using the same domain as the operational UKV model (Fig. 3). These were simple downscaling runs (from the Met Office global model) with no data assimilation. The cases sampled a wide range of meteorological conditions from the period July 2014 to April 2017 and comprised roughly equal numbers from each season. The cases were a mixture of poor forecasts (as identified by fore-

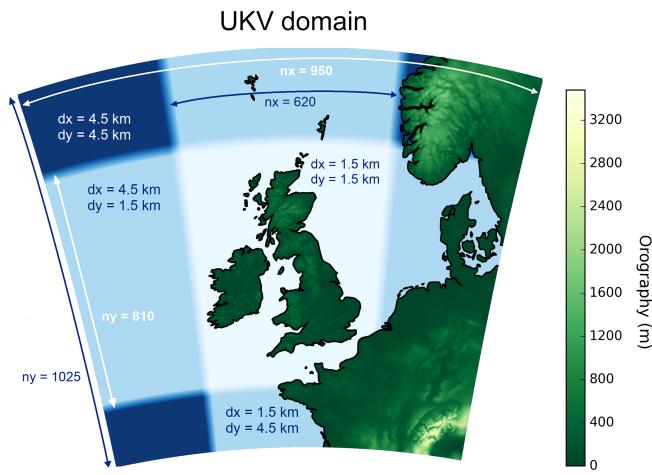


Figure 3. Domain for UK case studies.

casters), high-impact weather and normal everyday weather. The verification results from this stage of testing were used in the decision-making process of whether individual science changes were performing well enough to progress to the next round of testing. Secondly, the tickets were packaged up into a “proto-RAL1” package and the same case study tests repeated. Typically, there may be several “proto” packages trialled before a preferred package is chosen. Thirdly, to test the impact of including data assimilation in RAL1, 1-month-long UKV 3D-Var data assimilation trials were run for summer and winter 2016. The exact choice of dates for the case studies (and indeed the data assimilation trials) can obviously affect the results, but the reason for running the case studies is to provide a relatively cheap and quick test of model changes before moving on to the more expensive data assimilation trials.

Figure 4 shows the HiRA scorecard comparing RAL1 performance with RAL0 for the 100 case studies, and Fig. 5 shows the results for the 3D-Var winter and summer trials. The first thing to note is that there is remarkably good agreement between the case study and the 3D-Var trial results. This shows that the case studies can give a good indication of likely performance in data assimilation trials and that the exact choice of dates is not crucial to the results provided that enough cases are run. The second thing to note is that screen temperature is the variable that is (by far) the most significantly improved in RAL1.

Figure 6 shows the diurnal cycle of 1.5 m temperature bias and RMSE for RAL1-M and RAL1-T against RAL0 for the 100 case studies. The figure shows that RAL1 reduces the bias and RMSE in the diurnal cycle of screen temperature. This addresses a long-standing problem in the UKV model and is reflected in a statistically significant improvement to the temperature RPS at most lead times in both case studies (Fig. 4, top row) and 3D-Var trials (Fig. 5, top row). The improvement is primarily because of an increase in vege-

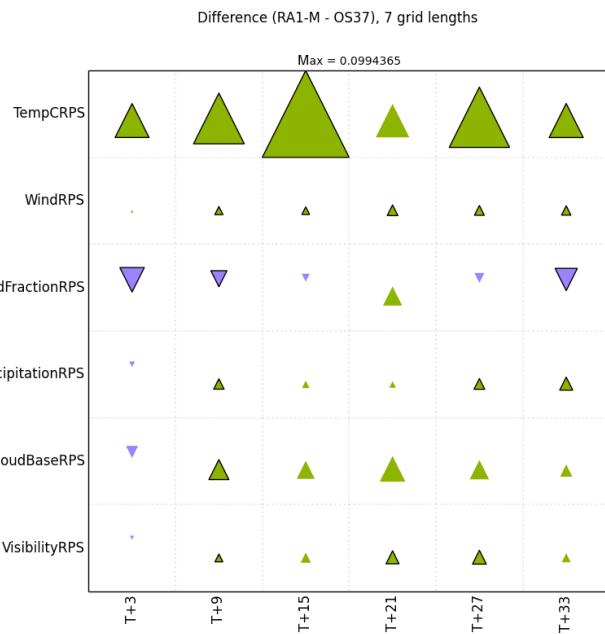


Figure 4. Case studies: RAL1-M vs. RAL0 HiRA summary scorecard at a 10.5 km (seven grid lengths) spatial scale. HiRA uses synoptic observations (see Sect. 5).

tation cover, at the expense of bare soil in RAL1, that reduces the thermal coupling between the atmosphere and soil. The reduction in scalar roughness lengths over grass tiles enhances the difference between skin and air temperatures. These changes lead to an amplified diurnal cycle of screen temperature and are supported by observational studies at the Met Office Research Unit site at Cardington, near Bedford. The albedos of vegetated tiles are also reduced in RAL1 and this results in warmer daytime temperatures. These changes were all components of ticket 3 (see Sect. 3.5). The impact on screen temperature varies according to the amount of vegetation present at a particular location. This is clearly illustrated by temperature differences over the UK shown in Fig. 7. In these plots the imprint of urban areas such as London show up as an area of little change between model versions RAL0 and RAL1. Another impact of the increase in vegetation cover from ticket 3 is that RAL1 reduces wind speeds (through an increase in the roughness length and therefore surface drag). The reduced wind speeds are beneficial at night-time (reducing an overforecasting bias) but detrimental by day (Fig. 8). Overall RAL1 shows statistically significant improvement to the 10 m wind RPS at most lead times in both case studies (Fig. 4) and 3D-Var trials (Fig. 5).

RAL1 gives an improvement to precipitation RPS at most lead times as seen in both case studies (Fig. 4) and the 3D-Var summer trial (Fig. 5b). The 3D-Var winter trial shows even stronger benefit with statistically significant improvements at all lead times (Fig. 5a). These HiRA results are based on rain-gauge data; 1 h FSS results (based on UK radar as truth) for

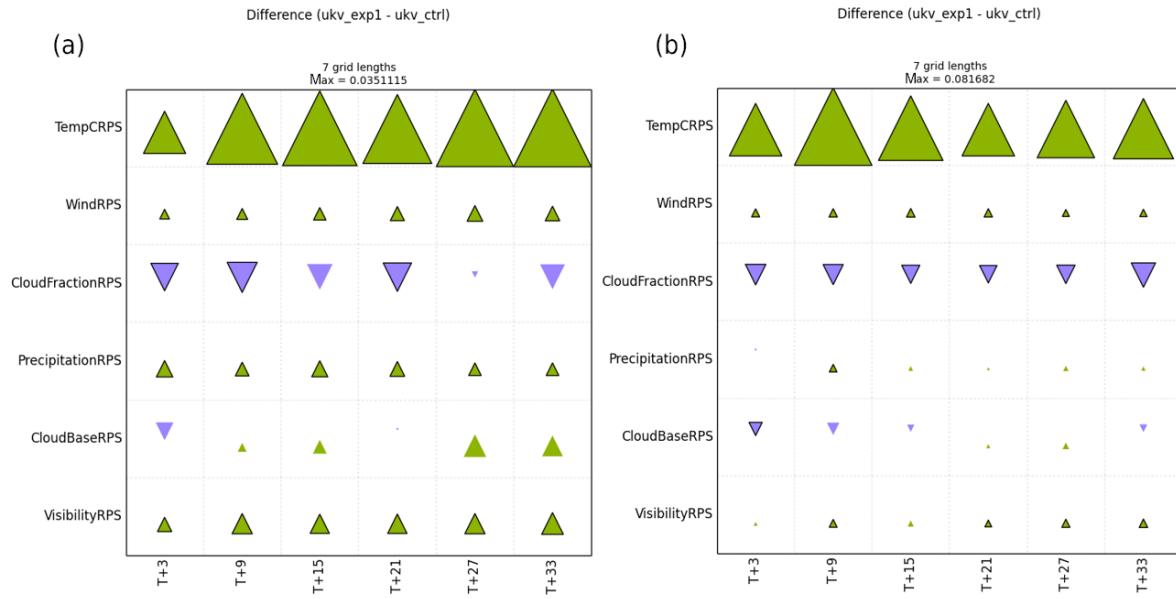


Figure 5. 3D-Var trials: RAL1-M vs. RAL0 HiRA summary scorecard at 10.5 km (seven grid lengths) spatial scale for winter (a) and summer (b). HiRA uses synoptic observations (see Sect. 5).

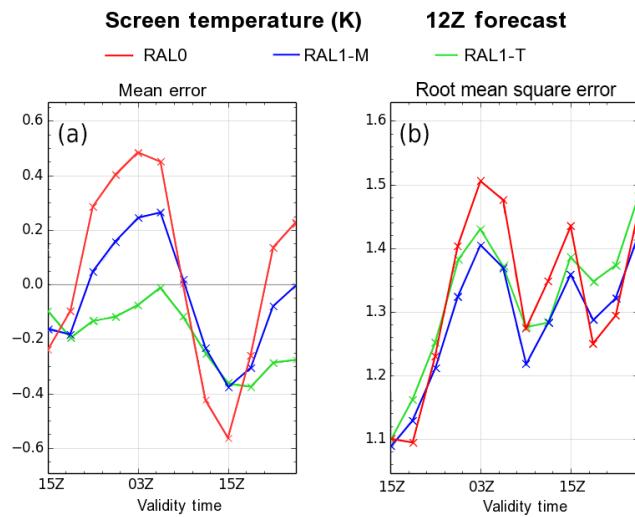


Figure 6. Case studies: diurnal cycle of screen temperature bias (a) and RMSE (b) for 12:00 Z forecasts. RAL0 (red), RAL1-M (blue) and RAL1-T (green).

the case studies (Fig. 10) show improvements to the 90th and 95th percentile results at all forecast ranges. The percentiles contain no bias information. However, the absolute thresholds at 0.5, 1.0 and 4.0 mm in the hour generally show a detriment. The 6 h FSSs for the case studies (Fig. 11) show similar results and point to potentially undesirable changes to bias. The overall precipitation mean error in the case studies is reduced in RAL1-M, and this reduces an overforecasting bias (now shown). The 1 mm frequency bias and 4 mm frequency bias results (not shown) indicate that as we have reduced our

mean error, we now on occasions have a frequency bias that is less than unity. RAL1 reduces the intensity of high precipitation rates (Fig. 9) as a result of the moisture conservation change that removed the spurious generation of precipitation by the semi-Lagrangian advection scheme from ticket 2, but this may have now revealed compensating errors.

RAL1 reduces the optical depth of fog as a result of the droplet taper change (ticket 1), and a further discussion of fog processes and model performance can be found in Boutle et al. (2018). The case study results (Fig. 4) and 3D-Var summer trial (Fig. 5b) show an improvement to visibility RPS at all lead times except for $T + 3$. The 3D-Var winter trial shows even stronger benefit with statistically significant improvement at all forecast ranges (Fig. 5a). Figure 12 shows a fog case study with high pressure centred over N France. RAL1 has less extensive < 100 m fog over England where none is observed.

RAL1 reduces cloud amounts and raises cloud base. This is likely to be related to a drying of the boundary layer as a result of the moisture conservation change. Overall RAL1 shows statistically significant degradation to cloud fraction RPS at most lead times in both case studies (Fig. 4) and 3D-Var winter trials (Fig. 5a). A subjective assessment of RAL1 by forecasters found that whilst largely very similar to RAL0, RAL1 tends to break up lower cloud faster than RAL0, especially where that cloud is fragmented. Whilst on average the reduction in cloud amounts is verified as worse, in some cases it is good. Figure 13 shows a stratocumulus case from 23 June 2015. RAL0 fails to break up the cloud cover through the daytime, leading to excessive low and medium cloud. RAL1-M breaks up the cloud more accurately along with

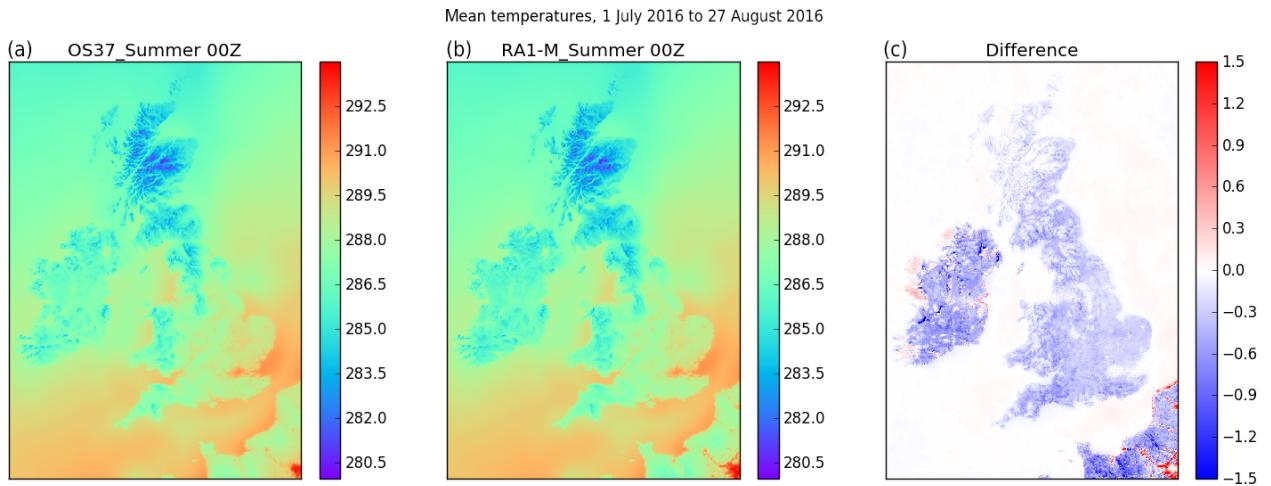


Figure 7. UKV model: summer mean temperature differences at 00:00 Z for RAL0 (**a**), RAL1-M (**b**) and RAL1-M minus RAL0 (**c**).

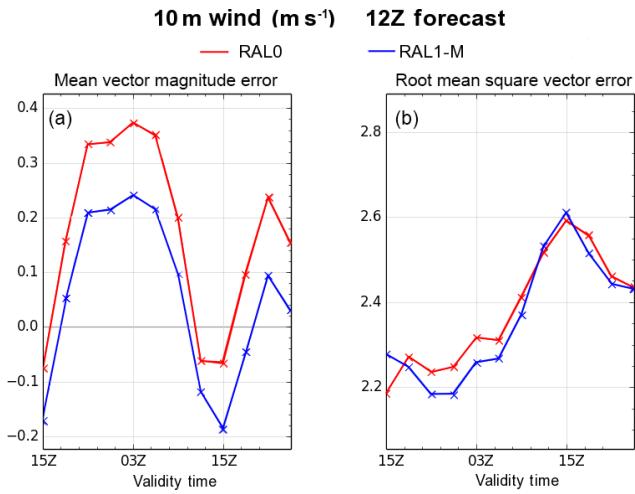


Figure 8. Case studies: diurnal cycle of 10 m wind mean vector magnitude error (**a**) and root mean square vector error (**b**) for 12:00 Z forecasts. RAL0 (red) and RAL1-M (blue).

RAL1-T, with RAL1-T tending to have even less cloud than RAL1-M. RAL1-T uses the PC2 cloud scheme, and this has been found to spuriously break up cloud in the UKV.

5.3 Tropical performance – Singapore

SINGV (Huang et al., 2019) was a 5-year collaborative project between the Met Office and Meteorological Service Singapore, which ran from 2013 to 2018. For the duration of the project the SINGV domain was the focal point for convective-scale model development in the tropics, and it was within this framework that the differences between RAL1-T and RAL1-M were identified, tested and then implemented. In this section we illustrate the impact of the changes implemented over the course of the SINGV project

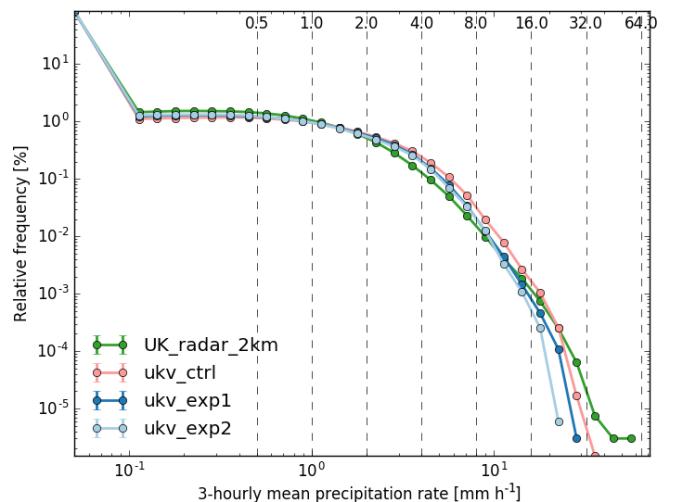


Figure 9. Case studies: relative frequency of 3-hourly precipitation rate. RAL0 (red), RAL1-M (dark blue), RAL1-T (light blue) and 2 km UK radar (dark green).

by comparing the performance over Singapore of the RAL1-T and RAL1-M configurations.

The model development trialling strategy within SINGV focused on downscaling global model forecasts, i.e. using the case study approach described above for UK testing. In order to reduce the potential dependency of the results on the choice of case, a whole month of forecasts were run out to $T + 36$ initialised from every 00:00 and 12:00 Z analysis. This approach ensured that summary measures were as robust as possible, whilst individual forecasts could be assessed in detail.

Figure 14 shows results for November 2016. Three model configurations are shown: (i) RAL1-T, (ii) RAL1-T-mPC2, which is RAL1-T but using the RAL1-M cloud scheme, and (iii) RAL1-T-3xBL, which is RAL1-T but with the RAL1-

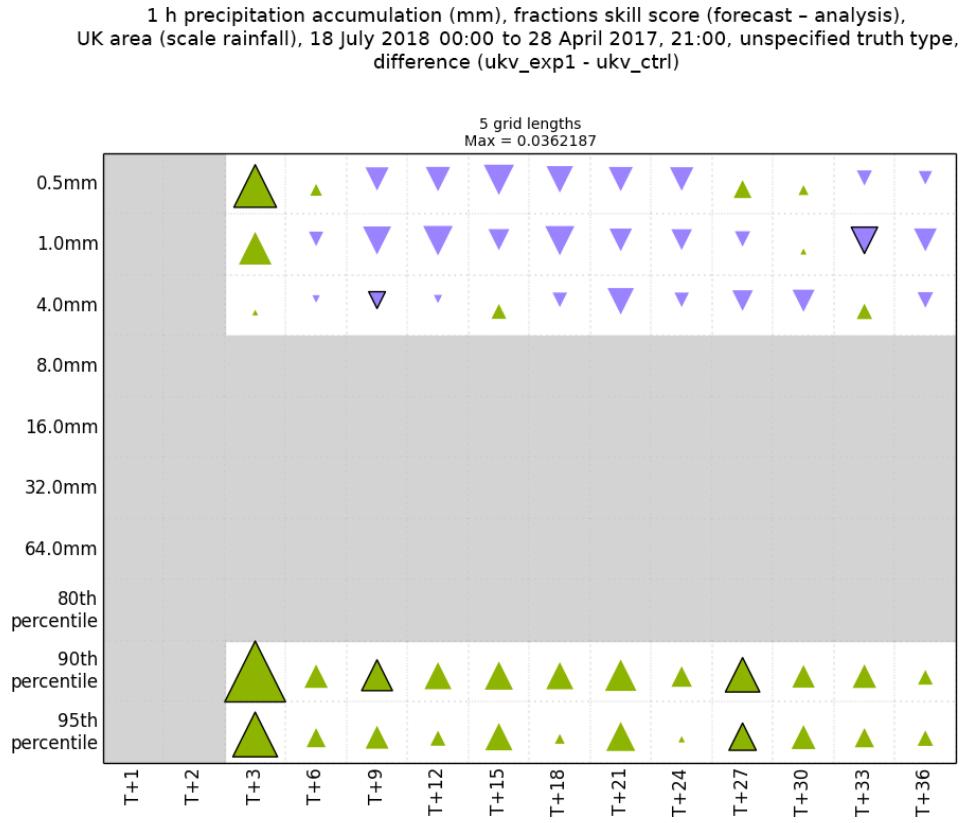


Figure 10. FSS at five grid lengths for 1 h accumulations.

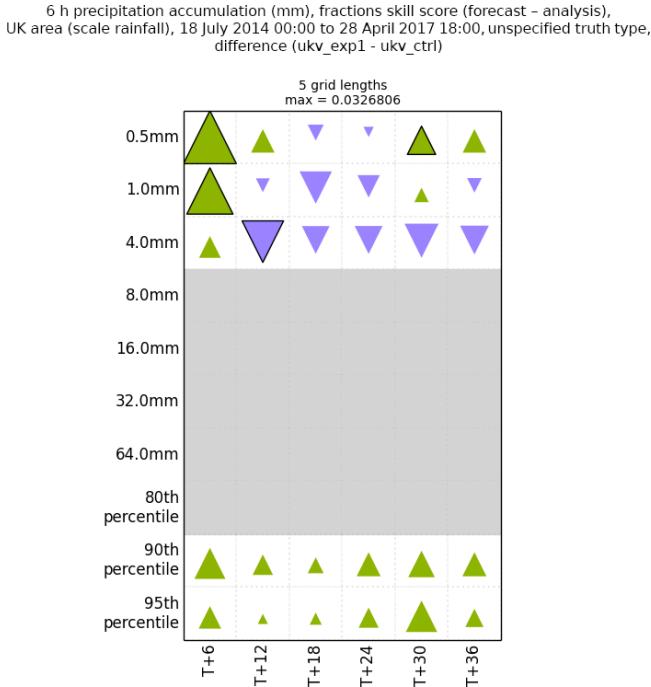


Figure 11. FSS at five grid lengths for 6 h accumulations.

M boundary layer settings. With these configurations we are able to illustrate the impact of the key differences between RAL1-T and RAL1-M. In Fig. 14a it is evident that the peak in the diurnal cycle of rainfall is too early compared to GPM for all three configurations. However, the time of convective initiation (when the rainfall first begins to increase, i.e. at $T + 15$) is well captured by RAL1-T and RAL1-T-mPC2. In contrast, RAL1-T-3xBL initiates even earlier (approximately 2 h) than RAL1-T. Other experiments (not shown) indicate that both the activation of the stochastic perturbations and the change to the convective BL stability functions contribute to this degradation in performance.

Figure 14b shows the impact on the rainfall FSS of removing PC2 from the RAL1-T configuration. The impact is large and shows that switching from the PC2 cloud scheme and reverting back to the Smith cloud scheme significantly reduces the ability of the model to skilfully predict high-impact rainfall events. Figure 14c shows that the impact of the BL differences also reduces the skill of the model, and this signal is significant for the high-percentile threshold for the majority of lead times.

An illustration of the differences in the RAL1-T and RAL1-M rainfall distribution over Singapore is shown in Fig. 15, which shows snapshots of the model forecasts for a single case study for 18 August 2016 compared to the Changi

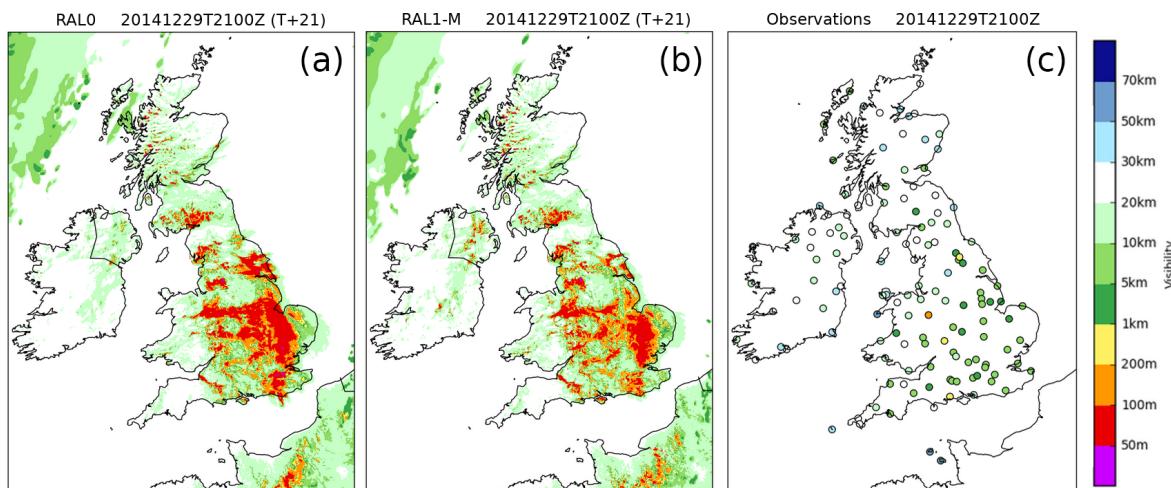


Figure 12. Fog case study for 29 December 2014: visibility from RAL0 (a) and RAL1-M (c) at $T + 21$ VT 21:00 Z on 29 December 2014 and corresponding station observations.

radar. The rainfall maps for early afternoon local time shown in Fig. 15a–c further illustrate the benefit of deactivating the stochastic perturbations in RAL1-T. The RAL1-T rainfall map compares favourably with the radar-estimated rainfall hourly accumulation with, in both cases, isolated convection just starting to develop over the Malay Peninsula. In contrast, the RAL1-M rainfall map shows the spurious localised convection has been initiated over a large area. This spurious convection has been triggered by the combined effect of the stochastic perturbations and the change to the convective boundary layer stability functions (as confirmed by additional experiments; not shown).

The rainfall maps for early the following morning local time (Fig. 15d–f) show a Sumatran squall passing through Singapore. The improved location of the squall in the RAL1-T forecast is typical of the impact found when the PC2 cloud scheme is implemented in SINGV. The impact of PC2 increases light rain amounts and decreases very heavy rain amounts compared to the Smith scheme. Effectively this makes the model more dissipative, and this leads to a reduction of small-scale structures, which enables the large-scale envelope of features like Sumatran squalls to be better handled and hence to propagate more realistically. The increased free-atmospheric mixing further increases the dissipation, and the two together were found to improve the ability of the model to propagate Sumatran squalls faster and further, rather than have them not develop or dissipate prematurely.

5.4 Tropical performance – Darwin MCS case

The Australian evaluation was carried out by the Bureau of Meteorology in Australia and consisted of running eight case studies over various domains with a 1.5 km horizontal grid

length. Here, we discuss one of the eight cases and compare both RAL1-T and RAL1-M against radar observations.

The observations come from the Darwin C-band polarimetric radar, which collects 3-D observations out to a range of 150 km (Louf et al., 2018); this allows for a detailed evaluation of simulated tropical convection. (Figure 16 shows the domain the radar covers and the area over which the comparison with the model is done.) The case studied is 18 February 2014 when active monsoon conditions produced a mesoscale convective system (MCS). The monsoon trough was stalled at the base of the Top End (geographical region encompassing the northernmost section of the Australia's Northern Territory), and there was a deep moisture layer and low-level convergence. The observed and modelled MCS life cycle is illustrated in the time series plots in Fig. 17, which shows the fractional area of the radar domain covered by reflectivities greater than 10 dBZ as a function of height and time over a 12 h period. From 12:00 to 15:00 UTC scattered convection was observed around Darwin, and the observed spatial coverage of cloud and rain within the radar domain increased from 20 % to 40 %. By 17:00 UTC the convection had become organised with numerous cells and a cloud shield exceeding 200 km in diameter. At 18:00 UTC the deepest convection was observed, with 10 dBZ cloud-top heights around 13 km. After this time, the mostly oceanic MCS matured and was composed of an extensive stratiform cloud region. The 1.5 km horizontal grid length simulations using RAL1-M and RAL1-T show deeper clouds and more extensive cloud and rain area coverage at earlier times than the radar observations. The cloud-top heights peak at 15:00 UTC in the RAL1-M simulation at a height greater than 14 km, which is 3 h earlier than the observed cloud-top height maximum and about 1.5 h earlier than the RAL1-T simulation. RAL1-M fails to produce significant fractional areas of cloud and rain greater than 0.8 throughout the MCS

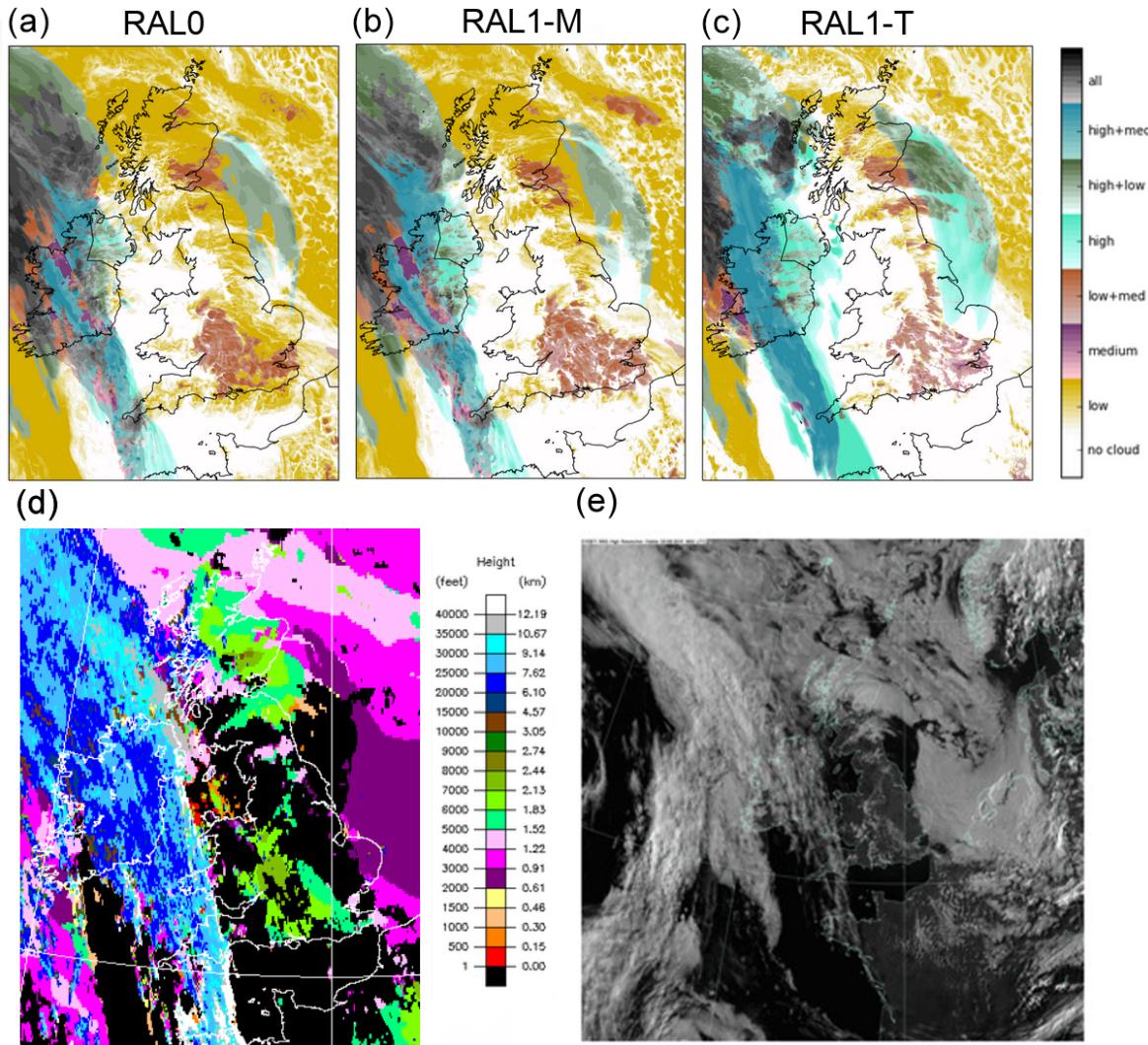


Figure 13. Cloud cover at $T + 18$ (stratocumulus case) showing RAL0 (a), RAL1-M (b) and RAL1-T (c) at 18:00Z on 23 June 2015 with corresponding satellite imagery; false colour and visible in panels (d) and (e), respectively. Imagery: © Crown Copyright, Met Office, Data: EUMETSAT.

life cycle, whereas RAL1-T shows a better representation of extensive stratiform cloud and rain areas, although it is a couple of hours too early. Both simulations overestimate the rainfall at the surface across the radar domain (Fig. 16), which is due to too many areas of heavy rain $> 8 \text{ mm h}^{-1}$ (not shown). The timing of the observed domain mean rainfall maximum occurs about an hour after the deepest clouds and a couple of hours before the hydrometeor spatial cover is maximal. While the simulations capture the same sequence of events, the rate of change in the domain mean rain rate as the system evolves from a developing to a mature MCS is amplified. This is primarily due to the model overestimating rainfall during the developing stages that are dominated by deep convection. RAL1-T produces a larger overestimate in total precipitation than RAL1-M but more accurately represents the timing of the MCS life cycle of precipitation.

5.5 Tropical performance – tropical cyclones in the western North Pacific

Evaluation of RAL1 for tropical cyclone (TC) forecasting concentrated on the Philippines since this is the most exposed country in the world to TCs. Figure 18 shows the regional model domain used. This has a large extent to the east of the Philippines to ensure that TCs travelling northwest towards the islands are captured in the domain long before making landfall.

A total of 130 TC forecasts (initialisation times between 15 March and 16 December 2015) were produced with both RAL1-T and RAL1-M using the domain shown in Fig. 18 with a horizontal grid length of 4.4 km and the L80($59_t, 21_s$)_{38.5} vertical level set. Storms were tracked in model output using the Met Office TC tracker (Heming,

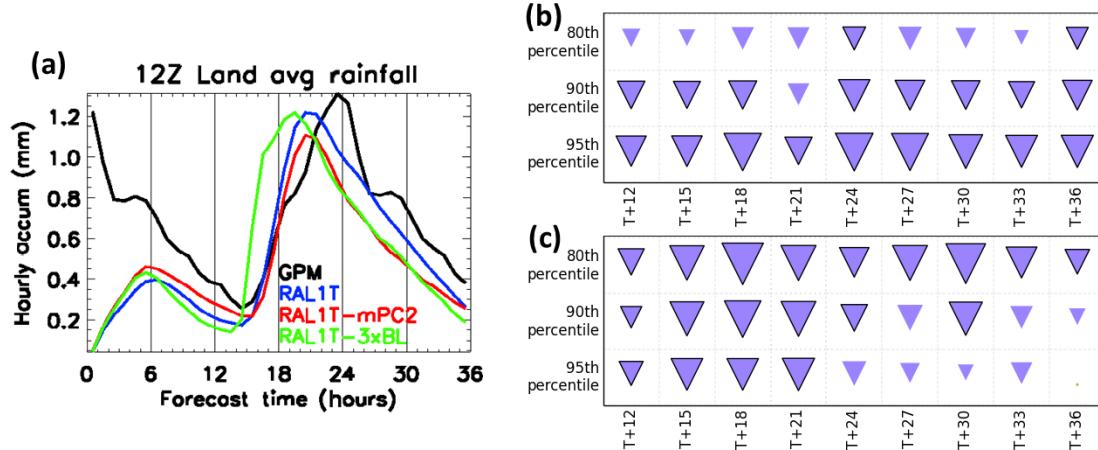


Figure 14. (a) Domain-averaged rainfall time series for all forecasts initialised at 12:00Z for November 2016. (b) Hinton plot showing the difference in FSS due to removing ticket 16 (RAL1-T-mPC2 minus RAL1-T) for a 250 km length scale compared to 3-hourly GPM. Panel (c) is as panel (b), but showing the impact on FSS of reverting the three BL changes back to the RAL1-M settings (RAL1-T-3xBL minus RAL1-T). Note that on the Hinton plots downward-pointing purple triangles indicate that RAL1-T is more skilful, whilst solid lines around the triangles denote that the difference is statistically significant.

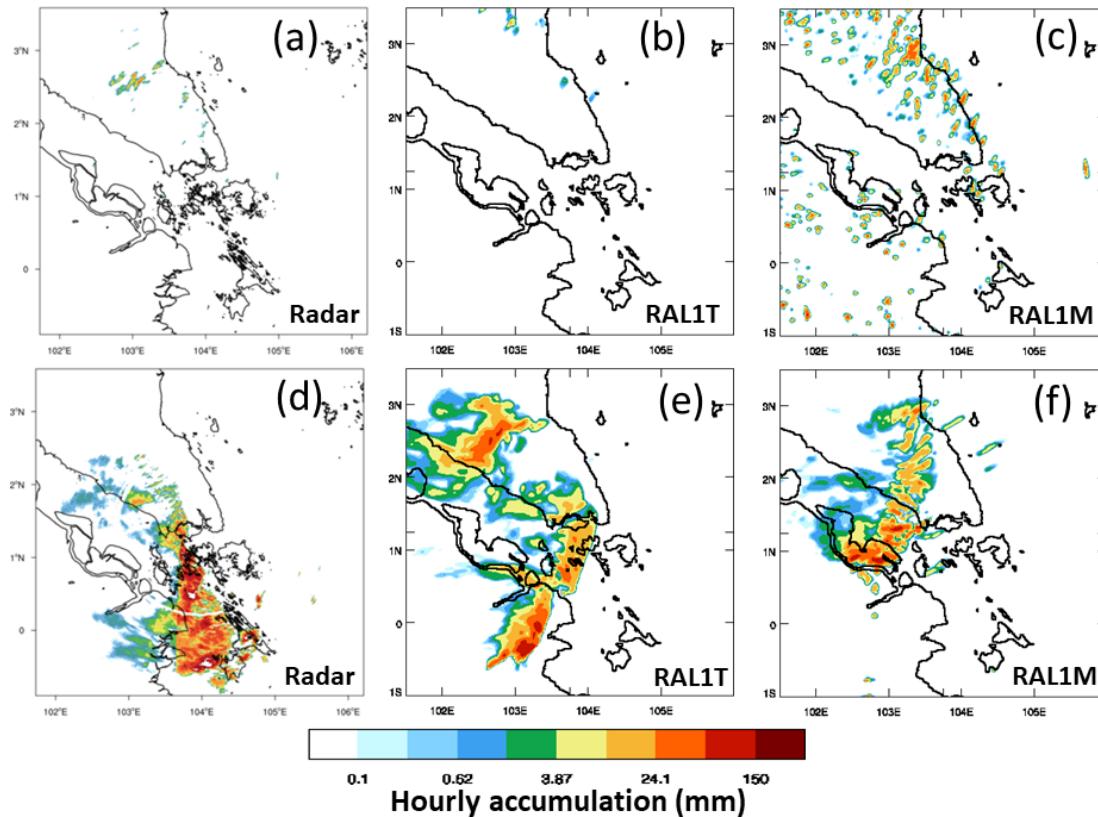


Figure 15. (a) Hourly rainfall accumulation for the Singapore radar at 06:00Z on 18 August 2016. Panel (b) is as panel (a), but showing the $T + 18$ RAL1-T forecast initialised at 12:00Z on 17 August 2016 (so valid at the same time as the radar). Panel (c) is as panel (b), but showing the equivalent RAL1-M imagery. Panels (d–f) are as panels (a–c), but for the hourly accumulation at 00:00Z on 19 August 2016, which is $T + 36$ for the same model forecasts as shown in (b) and (c).

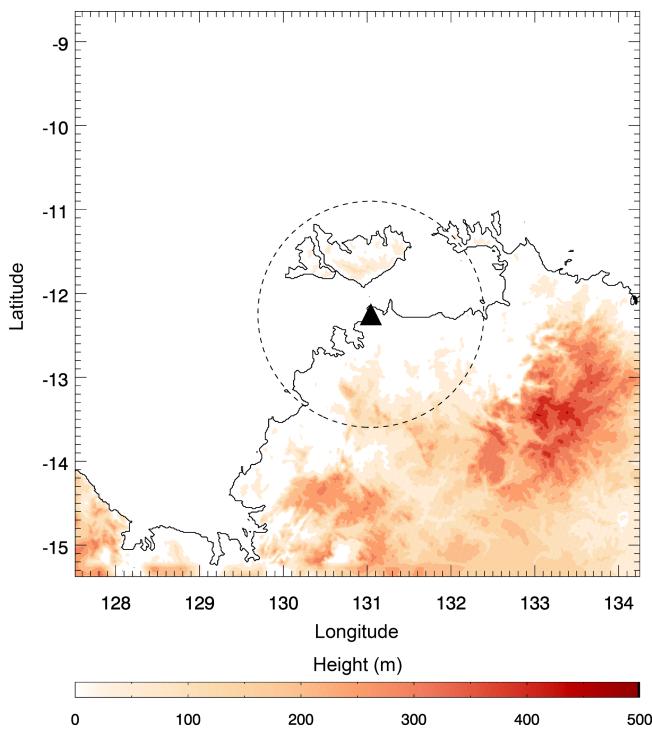


Figure 16. Domain of Australian MCS case study showing the Top End of Australia’s Northern Territory (which includes Darwin) and the Tiwi Islands. The CPOL radar location is denoted by the black triangle and its coverage by the area within the circle of dashed lines.

2017), and only storm cases appearing in both experiments were kept to ensure a fair comparison. A number of cases had two storms present in the domain at $T + 0$.

Figure 19 shows the mean bias (model – obs) in TC maximum surface wind speed and central pressure as a function of forecast lead time for the two RAL1 models. It is clear that both configurations give very similar intensity predictions. There is a protracted spin-up period as the regional models adjust from the weak initial state inherited from the driving global model. During this time, intensity errors are steadily reduced and, beyond $T + 36$, the bias in wind speed is close to zero (although this is the result of compensating errors: surface winds are typically underestimated in storms of category 3 and above but overestimated in weaker storms). However, RAL1 has a tendency to over-deepen storms, with central pressures dropping below those observed at about $T + 24$, asymptoting to a value approximately 10–15 hPa too low beyond $T + 48$. This could be due, at least in part, to the lack of ocean feedback on the atmosphere in the model. The differences in mean intensity biases visible beyond $T + 72$ are not statistically significant owing to the declining sample size with lead time.

It follows from Fig. 19 that the dynamical relationship between the wind and pressure fields in the model must be different to that observed. To highlight this, Fig. 20 shows scat-

terplots of maximum surface wind speed and central pressure for the RAL1 configurations, along with the observed wind–pressure relation (WPR) derived from Joint Typhoon Warning Center (JTWC) best-track data.

The RAL1 relations are a good match to the observed WPR up to wind speeds ~ 100 knots but are too steep beyond this. In other words, wind speeds in strong storms are too slow for their central pressure. This is likely because air–sea drag is currently overestimated in the model at high wind speeds. Plans for RAL2 include a reduction of the drag coefficient at high wind speeds, consistent with available observations.

Figure 21 displays the mean error in storm position relative to observations (as measured by the direct positional error, DPE) as a function of forecast lead time for the RAL1 models. Track errors in RAL1-T and RAL1-M are broadly comparable. In both cases the DPE increases by approximately 36 km per day of forecast, reaching a maximum of around 200 km at $T + 120$. There is a hint that RAL1-T may give more accurate track predictions, but the current sample is too small for this to be a statistically significant result.

5.6 Regional model ensemble performance for USA Hazardous Weather Testbed

The Met Office has been involved in the US Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (Kain et al., 2017), held annually in Norman, Oklahoma, for a number of years. UM kilometre-grid-scale regional models have been run and their performance has been found to be very competitive with the locally developed models (Kain et al., 2017). The meteorology of the Midwest USA with its severe weather (tornadoes, hail etc.) is different from that of the mid-latitudes (Sect. 5.2) and the tropics (Sect. 5.4). This is a good test for the regional model and ensures that we do not tune the model for a narrow set of meteorology. In addition the expertise of the HWT forecast team and the excellent observational network allow for a robust assessment of model performance.

After the 2017 HWT, a 12-member 2.2 km grid length UM ensemble was generated by one-way nesting the US 2.2 km domain (run routinely for the HWT) within the 12-member global ensemble (MOGREPS-G). The case studied was 16 May 2017, with MOGREPS-G initialised at 00:00 UTC on this day. MOGREPS-G had initial condition perturbations and used the random parameter (RP) scheme (McCabe et al., 2016) to perturb the model physics. Initial conditions and LBCs for each 2.2 km ensemble member were obtained from the corresponding global member. The RP scheme was not used in the 2.2 km ensemble; members were purely downscaled from the global. Each global ensemble member drove two 2.2 km ensemble members, each with a different science configuration: RAL1-M and RAL1-T. On this day there was a trough situated over the southern Rockies which was moving eastward, with a converg-

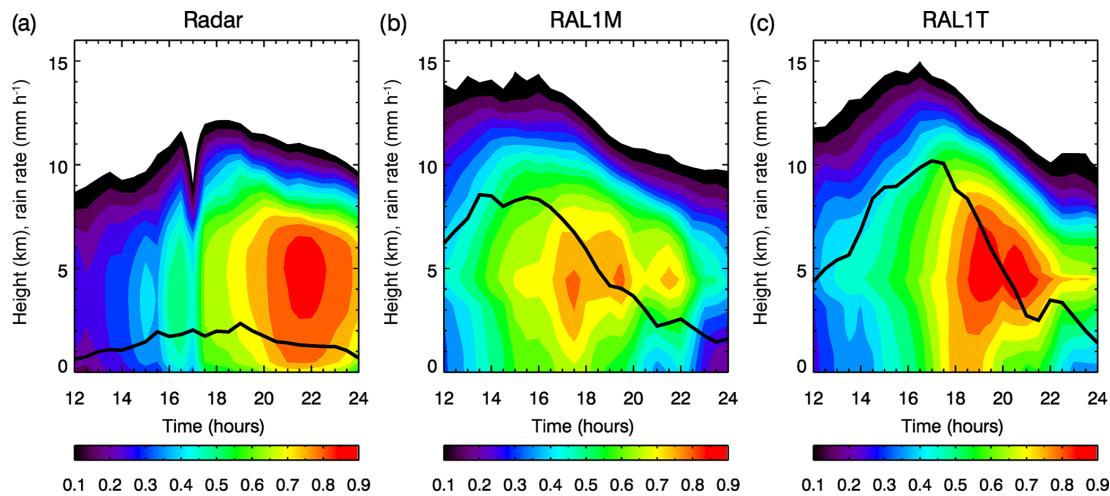


Figure 17. Fraction of radar area covered by reflectivities greater than 10 dBZ as a function of height and time (coloured contours) from 12:00 to 24:00 UTC on 18 February 2014. Solid lines are the time series of the domain mean rain rate (mm h^{-1}).

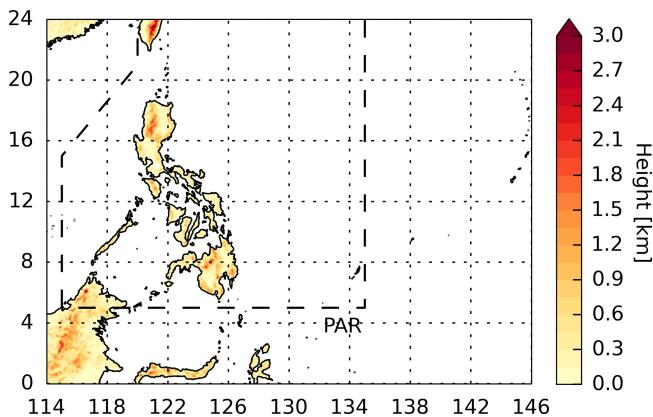


Figure 18. Philippine regional model domain and orography. The dashed black line shows the portion of the Philippine Area of Responsibility (PAR) inside the domain.

ing dry line across the Midwest and a strengthening low-level jet. Convection initiated over Texas at around 18:00 Z (13:00 CDT) and upscaled very quickly, with supercells observed over Oklahoma. Figure 22 shows the hourly accumulated precipitation averaged over the Texas–Oklahoma region for 16–17 May 2017 for the RAL1-M and RAL1-T ensembles, respectively. These figures highlight differences in the convection initiation time between the two configurations. Compared with the radar observations the RAL1-M members tended to initiate too early and produced a peak in precipitation at around 20:00–21:00 UTC that was not observed by the radar. Conversely, the RAL1-T ensemble members tended to initiate too late. Switching off the stochastic perturbations in the RAL1-M ensemble resulted in about a 1 h delay in the onset of precipitation and reduced the precipitation peak (not shown). However, the onset of precipitation was still not as delayed as it was in the RAL1-T ensem-

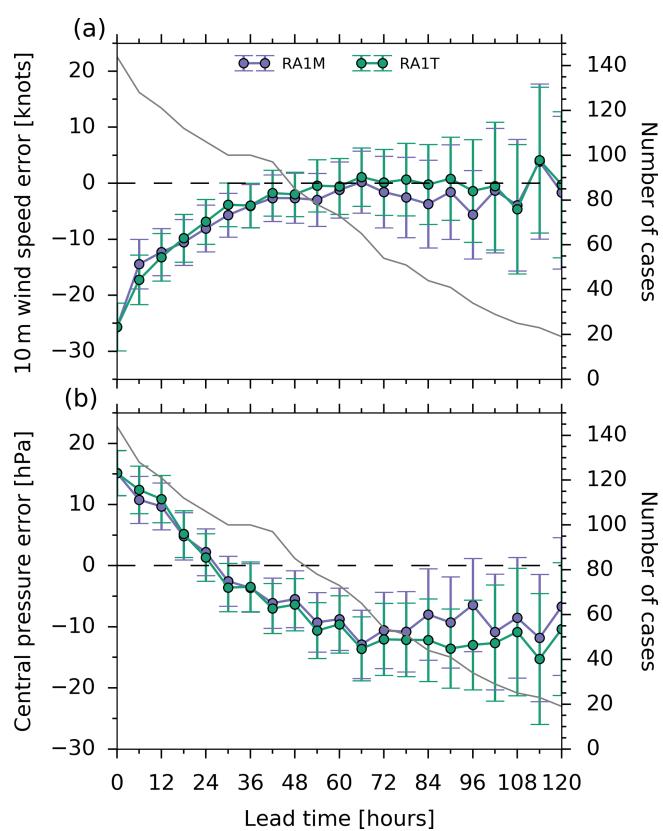


Figure 19. Mean bias in (a) maximum surface wind speed and (b) central pressure as a function of lead time for the RAL1-T and RAL1-M regional models. Error bars are 95 % confidence intervals on the mean. The solid grey lines indicate the number of cases at each lead time (see the right-hand axis of each plot).

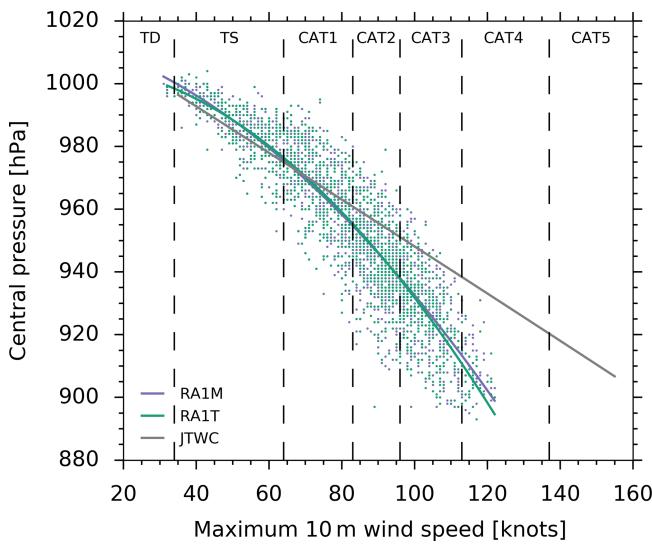


Figure 20. Wind–pressure relations for the RAL1-T and RAL1-M regional models. The corresponding observed relation from JTWC is shown for comparison. The solid lines are second-order polynomial fits to the data points.

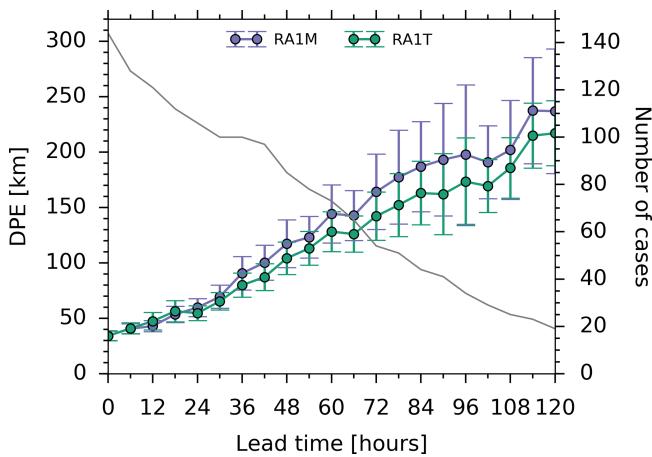


Figure 21. Error in forecast storm position relative to observations (direct positional error, DPE) as a function of lead time for the RAL1-T and RAL1-M models. Error bars are 95 % confidence intervals on the mean. The solid grey lines indicate the number of storm cases (see the right-hand axis of the plot).

ble, suggesting that the mixing length differences also contribute to the initiation time differences between RAL1-M and RAL1-T. Overall, the RAL1-T ensemble seems to better capture the supercells on this day, with more members simulating supercell-like features (Hanley and Lean, 2020).

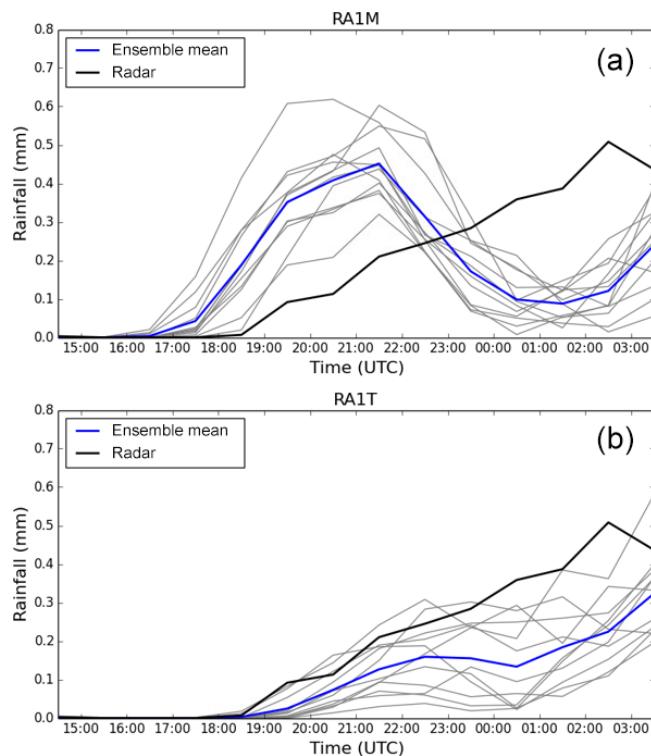


Figure 22. Domain-averaged precipitation from the RAL1-M ensemble (a) and RAL1-T ensemble (b) for 16–17 May 2017. Averaged over Texas and Oklahoma.

6 Conclusions

The definition of RAL1 is an important step in the development of kilometre-grid-scale configurations of the Unified Model. By concentrating the model development effort on a well-defined system, model users are better placed to learn from each other and to identify and resource the main priorities for future model development. In this paper we have defined configurations of the regional Met Office Unified Model, described a “toolbox” that allows us to evaluate its performance and provided some baseline tests to give a benchmark of performance. Performance is tested in simulations both with data assimilation and without – the latter we refer to as case studies.

While it remains an ambition to have a single configuration of the model that works across all regions, at this stage we have defined two: RAL1-M for mid-latitudes and RAL1-T for the tropics. Both are clearly documented in terms of the model physics and their performance in relevant regions. For the mid-latitude system the most recent developments are described in more detail, and the NWP performance changes due to these recent changes are shown. To do this we have defined a previous operational NWP version of the Unified Model, which we refer to here as RAL0. The performance of the tropical system is presented as a benchmark for future developments.

Table 3. Identifiers for a set of RAL1 reference simulations across a number of systems and/or applications. These suites are held on the Met Office Science Repository Service, which also holds the UM and JULES code.

Application	Suite ID	UM version/JULES version
UK case studies	u-ao109	UM10.6/JULES4.7
Singapore case studies	u-av356	UM10.9/JULES5.0
Darwin MCS case study	u-ax904	UM10.6/JULES4.7
Tropical cyclone case studies	u-aq686	UM10.6/JULES4.7
USA Hazardous Weather Testbed case study	u-ao861	UM10.6/JULES4.7

The recent science developments included in RAL1-M are shown to significantly improve two long-standing issues with model performance in NWP. The inclusion of moisture conservation reduces overly intense local precipitation rates, and the changes to land use and vegetation properties improve a damped diurnal cycle in near-surface temperatures. We also see modest improvements to forecasts of low visibilities. The conservation of moisture was of particular importance to the tropical configuration of the model, although this was not shown in the paper.

The goal of having a clearly defined version of a regional model, and perhaps more importantly a series of tests for that model to give confidence that changes are generally improving the system, is hugely challenging. In this paper we have shown a series of tests in a small number of regions that require substantial computational effort. Yet, we have only sampled a small fraction of the types of meteorology that the model should be expected to represent. Looking ahead, we need to consider other regions such as the poles and more broadly sampling the range of weather types seen in the regions we have considered. One very specific area which is not covered in this paper is the performance of the model in climate simulations. It remains a high priority to include climate testing in the development process of the regional model, although with the high computing costs involved in regional climate runs at the kilometre-grid-scale system, the test will need careful design.

Looking ahead, in addition to improving the modelling system, consolidating regional differences and documenting this, we also aim to substantially improve the evaluation process. This will include climate testing, increased use of ensembles and testing in more regions. This will require concerted effort and coordination from the partnership developing the RAL configuration, but this should lead to a better understanding of its strengths and weaknesses and lead to the more efficient development of further improvements.

Appendix A

Algorithm 1 Iterative structure of time step $(n + 1)$. Here, we use two inner and two outer loops ($L = 2, M = 2$).

- 1: Given the model's state at the previous time step, let the first guess of the prognostic variable F^{n+1} at time level $(n + 1)$ be

$$F^{n+1} = F^n$$
 - 2: For LAMs the appropriate LBCs are also provided for time step $(n + 1)$ (note that LBCs are interpolated in time and space and hydrostatically balanced)
 - 3: If LAM, fill external halos with the appropriate data (external halos contain constant values extended outward from adjacent LBC data, e.g. zero gradient; see text)
 - 4: Compute slow parametrised processes and time level n forcings R_F^n
 - 5: **for** $m = 1, M$ **do** {departure (outer-loop) iteration}
 - 6: For LAM, update dynamics (or wind) only LBCs to avoid the back-substitution wind diverging from LBCs (see text)
 - 7: Compute departure points using wind at time level n and the latest estimate at time level $(n + 1)$
 - 8: SL advect all prognostic variables (requires interpolation to departure points)
 - 9: SL advect moisture fields
 - 10: Apply ZLF to restore moisture mass conservation after advection (if required)
 - 11: If ($m = M$), SL advect all other tracers and apply ZLF (if required)
 - 12: Compute fast parametrised processes using the latest F^{n+1}
 - 13: Evaluate all Helmholtz terms involving time level n (this includes terms interpolated at departure points)
 - 14: **for** $l = 1, L$ **do** {non-linear (inner-loop) iteration}
 - 15: Evaluate non-linear terms involving the latest state $(n + 1)$
 - 16: Evaluate time level $(n + 1)$ components of Helmholtz right-hand side RHS^{n+1}
 - 17: For LAM, adjust RHS^{n+1} to satisfy the LBCs pressure values (note also that the Helmholtz coefficients are also modified to enforce a Dirichlet-type boundary condition (known pressure at the LBC at time level $(n + 1)$))
 - 18: Solve the Helmholtz problem for the pressure increment π' and evaluate $\pi^{n+1} \equiv \pi^n + \pi'$
 - 19: Compute the other prognostic variables at time level $(n + 1)$ via the back-substitution
 - 20: **end for**
 - 21: **end for**
 - 22: For LAM, update all LBCs (i.e. blend LBC data and model's solution)
-

Table A1. Source datasets used to create standard ancillary files used in RAL0.

Ancillary field	Source data	Notes
Land–sea mask	IGBP; Loveland et al. (2000) CCI; Hartley et al. (2017)	Used for UKV/MOGREPS-UK CCI mask lacking in inland lakes definition
Mean/sub-grid orography	DTED 1 km; GLOBE 30''; Hastings et al. (1999) SRTM; Bunce et al. (1996)	Used for UKV/MOGREPS-UK Fields filtered before use Shuttle Radar Topography Mission; mean orography only; available up to 60° north
Land usage	IGBP; Loveland et al. (2000) ITE; Bunce et al. (1996) CCI; Hartley et al. (2017)	Mapped to nine tile types UK only
Soil properties	HWSD; Nachtergaele et al. (2008) STATSGO; Miller and White (1998) ISRIC-WISE; Batjes (2009)	European Space Agency Land Cover Climate Change Initiative Three datasets blended via optimal interpolation
Leaf area index	MODIS collection 5	4 km data (Samanta et al., 2012) mapped to five plant types
Plant canopy height	IGBP; Loveland et al. (2000)	Derived from land usage and mapped to five plant types
Bare soil albedo	MODIS; Houldcroft et al. (2008)	
SST/sea ice	System/experiment dependent	
Ozone	Li and Shine (1995)	
Murk aerosol	NAEI, ENTEC and EMEP emission inventories	
CLASSIC aerosol climatologies	System/experiment dependent	Used when prognostic fields not available

We document a list of acronyms in Table A2.

Table A2. Acronym list.

Acronym	Meaning	Notes
EACF	Empirically adjusted cloud fraction	
ENDGame	Even Newer Dynamics for General atmospheric modelling of the environment	Dynamical core used in RAL0 and RAL1
GA	Global Atmosphere	Global atmosphere science configuration
GA3.1	Global Atmosphere 3.1	A specific GA science configuration
GA7.0	Global Atmosphere 7.0	A specific GA science configuration
GL	Global Land	Global land science configuration
GL7.0	Global Land 7.0	A specific GL science configuration
JULES	Joint UK Land Environment Simulator	Community land surface model
LAM	Limited-area model	
LBCs	Lateral boundary conditions	
LW	Longwave	
MOGREPS-UK	Met Office Global and Regional Ensemble system – UK	UK NWP operational ensemble system
NMS	National Met Services	
NWP	Numerical weather prediction	
RAL	Regional Atmosphere and Land	
RAL0	Regional Atmosphere and Land 0	Baseline RAL science configuration
RAL1	Regional Atmosphere and Land 1	First RAL science configuration
RAL1-M	Regional Atmosphere and Land 1 – Mid Latitudes	
RAL1-T	Regional Atmosphere and Land 1 – Tropics	
RMED	Regional Model Evaluation and Development	
SI	Semi-implicit	
SL	Semi-Lagrangian	
SOCRATES	Suite Of Community RADiative Transfer codes based on Edwards and Slingo	Radiative transfer scheme
SW	Shortwave	
UKV	UK variable (resolution)	UK NWP operational deterministic model
UM	Unified Model	
ZLF	Zero lateral flux	

We document the list of RMED tickets included in RAL1 in Table A3.

Table A3. RMED tickets included in RAL1.

RMED ticket number	RAL1-M/RAL1-T	Description of RAL1 change
1	RAL1-M and RAL1-T	Improved droplet number profile in the lower boundary layer
2	RAL1-M and RAL1-T	Conservative advection for moist prognostics
3	RAL1-M and RAL1-T	Improvements to land usage and vegetation properties
5	RAL1-M and RAL1-T	Improved representation of mixing across the boundary layer top
9	RAL1-M and RAL1-T	Improved treatment of gaseous absorption
10	RAL1-M and RAL1-T	Reductions in sensitivity to vertical resolution
11	RAL1-M and RAL1-T	Change method of PMSL calculation to be more efficient
12	RAL1-M	Revision of free-atmospheric mixing length
15	RAL1-M and RAL1-T	Retuning BL mixing across the LCL in cumulus
16	RAL1-T	Cloud scheme upgrades
19	RAL1-M and RAL1-T	Correct inappropriate treatment of graupel as fresh snow in JULES
25	RAL1-M	Updates to stochastic boundary layer perturbations
26	RAL1-T	Revised unstable stability functions

Code availability. Due to intellectual property right restrictions, we cannot provide either the source code or documentation papers for the UM or JULES.

Obtaining the UM. The Met Office Unified Model is available for use under licence. A number of research organisations and national meteorological services use the UM in collaboration with the Met Office to undertake basic atmospheric process research, produce forecasts, develop the UM code, and build and evaluate Earth system models. For further information on how to apply for a licence, see <http://www.metoffice.gov.uk/research/modelling-systems/unified-model> (last access: 3 April 2020).

Obtaining JULES. JULES is available under licence free of charge. For further information on how to gain permission to use JULES for research purposes, see http://jules-lsm.github.io/access_req_JULES_access.html (last access: 3 April 2020).

Details of the simulations performed. UM–JULES simulations are compiled and run in suites developed using the Rose suite engine (<http://metomi.github.io/rose/doc/html/index.html>, MetOffice, 2020) and scheduled using the cylc workflow engine (<https://cylc.github.io/>, Oliver et al., 2019). Both Rose and cylc are available under v3 of the GNU General Public License (GPL). In this framework, the suite contains the information required to extract and build the code as well as configure and run the simulations. Each suite is labelled with a unique identifier and is held in the same revision-controlled repository service in which we hold and develop the model code. This means that these suites are available to any licensed user of both the UM and JULES. We document a set of reference RAL1-based simulations in Table 3.

Author contributions. MB led the RAL1 testing and evaluation process and prepared the paper with contributions from all co-authors. TA, IB, JE, AL, JM, CM, JW, NW and MZ are either code owners and/or developers of the model code included in RAL1. CB, AF, CF, KH, HL, MM, RN, CS, SW and MW performed the evaluation. JP, SV and DW contributed to the writing of the Introduction and Conclusions sections.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The development and assessment of the Regional Atmosphere Land configurations are possible only through the hard work of a large number of people that exceeds the list of authors. Specifically, we would like to thank Belinda Roux and Hongyan Zhu (Bureau of Meteorology (BoM), Melbourne, Victoria, Australia), Magdalena Gruziel-Slomka and Małgorzata Melonek (Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), Warsaw, Poland), Seungwoo Lee (Korea Meteorological Administration (KMA), Seoul, South Korea), Jayakumar (National Centre for Medium Range Weather Forecasting (NCMRWF), Noida, India), Stuart Moore (National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand), and Estelle Marx and Elelwani Phaduli (South African Weather Service (SAWS), Pretoria, South Africa) for running case studies and evaluating model performance. Douglas Boyd (Met Office, Exeter, UK), George Pankiewicz (Met Office, Exeter, UK), Charmaine Franklin

(Bureau of Meteorology (BoM), Melbourne, Victoria, Australia), Magdalena Gruziel-Slomka (Interdisciplinary Centre for Mathematical and Computational Modelling (ICM), Warsaw, Poland), Seungwoo Lee (Korea Meteorological Administration (KMA), Seoul, South Korea), Jeff Lo (Meteorological Service Singapore (MSS), Singapore), Saji Mohandas (National Centre for Medium Range Weather Forecasting (NCMRWF), Noida, India), Stuart Moore and Trevor Carey-Smith (National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand), Stephanie Landman (South African Weather Service (SAWS), Pretoria, South Africa), and Steven Rugg (United States Air Force (USAF), 557th Weather Wing, Offutt Air Force Base, Nebraska, United States of America) are thanked for the coordination, led by Mike Bush (Met Office, Exeter, UK), of RAL1 testing and evaluation around the UM partnership.

The research and project work of Charmaine Franklin was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

The GPM IMERG Late Precipitation L3 Half Hourly $0.1^\circ \times 0.1^\circ$ V04 data were provided by the NASA/Goddard Space Flight Center's Goddard Earth Sciences Data and Information Services Center and PPS, which develop and compute the GPM IMERG Late Precipitation L3 Half Hourly $0.1^\circ \times 0.1^\circ$ as a contribution to GPM, and are archived at the NASA GES DISC.

Review statement. This paper was edited by Jatin Kala and reviewed by two anonymous referees.

References

- Abel, S. J. and Boutle, I. A.: An improved representation of the raindrop size distribution for single-moment micro-physics schemes, *Q. J. Roy. Meteor. Soc.*, 138, 2151–2162, <https://doi.org/10.1002/qj.1949>, 2012.
- Abel, S. J. and Shipway, B. J.: A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO, *Q. J. Roy. Meteor. Soc.*, 133, 781–794, <https://doi.org/10.1002/qj.55>, 2007.
- Abel, S. J., and K. Waite, I. A. B., and P. R. A. Brown, S. F., Cotton, R., Lloyd, G., Choularton, T. W., and Bower, K. N.: The role of precipitation in controlling the transition from stratocumulus to cumulus clouds in a northern hemisphere cold-air outbreak, *J. Atmos. Sci.*, 74, 2293–2314, <https://doi.org/10.1175/JAS-D-16-0362.1>, 2017.
- Allen, T. and Zerroukat, M.: A deep non-hydrostatic compressible atmospheric model on a Yin-Yang grid, *J. Comput. Phys.*, 319, 44–60, 2016.
- Arakawa, A. and Lamb, V. R.: Computational design of the basic dynamic processes of the UCLA general circulation model, *Methods Comput. Phys.*, 17, 173–265, 1977.
- Aranami, K., Davies, T., and Wood, N.: A mass restoration scheme for limited-area models with semi-Lagrangian advection, *Q. J. Roy. Meteor. Soc.*, 141, 1795–1803, 2014.
- Baldauf, M., Seifert, A., Forstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Descrip-

- tion and Sensitivities, *Mon. Weather Rev.*, 139, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>, 2011.
- Baran, A. J., Hill, P., Walters, D., Hardman, S. C., Furtado, K., Field, P. R., and Manners, J.: The impact of two coupled cirrus microphysics-radiation parameterizations on the temperature and specific humidity biases in the tropical tropopause layer in a climate model, *J. Climate*, 29, 5299–5316, <https://doi.org/10.1175/JCLI-D-15-0821.1>, 2016.
- Barker, H. and Li, Z.: Improved simulation of clear-sky radiative transfer in the CCC-GCM, *J. Climate*, 8, 2213–2223, [https://doi.org/10.1175/1520-0442\(1995\)008<2213:ISOCSS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<2213:ISOCSS>2.0.CO;2), 1995.
- Batjes, N. H.: Harmonized soil profile data for applications at global and continental scales: updates to the WISE database, *Soil Use Manage.*, 25, 124–127, <https://doi.org/10.1111/j.1475-2743.2009.00202.x>, 2009.
- Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemela, S., Nielsen, K. P., Onvlee, J., Rontu, L., Samuelsson, P., Munoz, D. S., Subias, A., Tijm, S., Toll, V., Yang, X., and Koltzow, M. O.: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system, *Mon. Weather Rev.*, 145, 1919–1935, <https://doi.org/10.1175/MWR-D-16-0417.1>, 2017.
- Berthou, S., Kendon, E., Chan, S., Ban, N., Leutwyler, D., Schar, C., and Fosser, G.: Pan-European climate at convection-permitting scale: a model intercomparison study, *Clim. Dynam.*, <https://doi.org/10.1007/s00382-018-4114-6>, online first, 2018.
- Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, *Geosci. Model Dev.*, 4, 677–699, <https://doi.org/10.5194/gmd-4-677-2011>, 2011.
- Bodas-Salcedo, A., Williams, K. D., Field, P. R., and Lock, A. P.: The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: the role of midlatitude cyclone clouds, *J. Climate*, 25, 7467–7486, <https://doi.org/10.1175/JCLI-D-11-00702.1>, 2012.
- Boutle, I., Price, J., Kudzotsa, I., Kokkola, H., and Römkänniemi, S.: Aerosol-fog interaction and the transition to well-mixed radiation fog, *Atmos. Chem. Phys.*, 18, 7827–7840, <https://doi.org/10.5194/acp-18-7827-2018>, 2018.
- Boutle, I. A. and Morcrette, C. J.: Parametrization of area cloud fraction, *Atmos. Sci. Lett.*, 11, 283–289, <https://doi.org/10.1002/asl.293>, 2010.
- Boutle, I. A., Abel, S. J., Hill, P. G., and Morcrette, C. J.: Spatial variability of liquid cloud and rain: observations and microphysical effects, *Q. J. Roy. Meteor. Soc.*, 140, 583–594, <https://doi.org/10.1002/qj.2140>, 2014a.
- Boutle, I. A., Eyre, J. E. J., and Lock, A. P.: Seamless stratocumulus simulation across the turbulent gray zone, *Mon. Weather Rev.*, 142, 1655–1668, <https://doi.org/10.1175/MWR-D-13-00229.1>, 2014b.
- Boutle, I. A., Finnenkoetter, A., Lock, A. P., and Wells, H.: The London Model: forecasting fog at 333 m resolution, *Q. J. Roy. Meteor. Soc.*, 142, 360–371, <https://doi.org/10.1002/qj.2656>, 2016.
- Brousseau, P., Seity, Y., Ricard, D., and Leger, J.: Improvement of the forecast of convective activity from the AROME-France system, *Q. J. R. Meteor. Soc.*, 142, 2231–2243, <https://doi.org/10.1002/qj.2822>, 2016.
- Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., and Shelly, A.: Unified modeling and prediction of weather and climate: a 25 year journey, *B. Am. Meteorol. Soc.*, 93, 1865–1877, <https://doi.org/10.1175/BAMS-D-12-00018.1>, 2012.
- Brown, A. R.: The sensitivity of large-eddy simulations of shallow cumulus convection to resolution and subgrid model, *Q. J. Roy. Meteor. Soc.*, 125, 469–482, <https://doi.org/10.1002/qj.49712555405>, 1999.
- Bunce, R., Barr, C., Gillespie, M., and Howard, D.: The ITE Land Classification: Providing an Environmental Stratification of Great Britain, *Environ. Monit. Assess.*, 39, 39–46, <https://doi.org/10.1007/BF00396134>, 1996.
- Cahalan, R., Ridgway, W., Wiscombe, W., Bell, T., and Snider, J.: The Albedo of Fractal Stratocumulus clouds, *J. Atmos. Sci.*, 51, 2434–2455, [https://doi.org/10.1175/1520-0469\(1994\)051<2434:TAOFSC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<2434:TAOFSC>2.0.CO;2), 1994.
- Charney, J. G. and Phillips, N. A.: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows, *J. Meteorol.*, 10, 71–99, [https://doi.org/10.1175/1520-0469\(1953\)010<0071:NIOTQG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1953)010<0071:NIOTQG>2.0.CO;2), 1953.
- Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, 4, 701–722, <https://doi.org/10.5194/gmd-4-701-2011>, 2011.
- Clark, P., Roberts, N., Lean, H., Ballard, S., and Charlton-Perez, C.: Convection-permitting models: a step-change in rainfall forecasting, *Meteorol. Appl.*, 23, 165–181, <https://doi.org/10.1002/met.1538>, 2016.
- Clark, P. A., Harcourt, S. A., Macpherson, B., Mathison, C. T., Cusack, S., and Naylor, M.: Prediction of visibility and aerosol within the operational Met Office Unified Model. I: Model formulation and variational assimilation, *Q. J. Roy. Meteor. Soc.*, 134, 1801–1816, <https://doi.org/10.1002/qj.318>, 2008.
- Cotton, R. J., Field, P. R., Ulanowski, Z., Kaye, P. H., Hirst, E., Greenaway, R. S., Crawford, I., Crosier, J., and Dorsey, J.: The effective density of small ice particles obtained from in situ aircraft observations of mid-latitude cirrus, *Q. J. Roy. Meteor. Soc.*, 139, 1923–1934, <https://doi.org/10.1002/qj.2058>, 2013.
- Cusack, S., Slingo, A., Edwards, J. M., and Wild, M.: The radiative impact of a simple aerosol climatology on the Hadley Centre atmospheric GCM, *Q. J. Roy. Meteor. Soc.*, 124, 2517–2526, <https://doi.org/10.1002/qj.49712455117>, 1998.
- Davies, H. C.: A lateral boundary formulation for multi-level prediction models, *Q. J. Roy. Meteor. Soc.*, 102, 405–418, 1976.
- Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, *Q. J. Roy. Meteor. Soc.*, 131, 1759–1782, <https://doi.org/10.1256/qj.04.101>, 2005.
- Edwards, J. M. and Slingo, A.: Studies with a flexible new radiation code. I: Choosing a configuration for a large-

- scale model, *Q. J. Roy. Meteor. Soc.*, 122, 689–719, <https://doi.org/10.1002/qj.49712253107>, 1996.
- Field, P. R., Heymsfield, A. J., and Bansemer, A.: Snow Size Distribution Parameterization for Midlatitude and Tropical Ice Clouds, *J. Atmos. Sci.*, 64, 4346–4365, <https://doi.org/10.1175/2007JAS2344.1>, 2007.
- Furtado, K., Field, P. R., Cotton, R., and Baran, A. J.: The sensitivity of simulated high clouds to ice crystal fall speed, shape and size distribution, *Q. J. Roy. Meteor. Soc.*, 141, 1546–1559, <https://doi.org/10.1002/qj.2457>, 2015.
- Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N., and Tennant, W.: The Met Office convective-scale ensemble, MOGREPS-UK, *Q. J. Roy. Meteor. Soc.*, 143, 2846–2861, <https://doi.org/10.1002/qj.3135>, 2017.
- Hanley, K. and Lean, H.: Elucidating the causes of errors in the 2.2 km Met Office Unified Model simulations of a convective case from the 2017 Hazardous Weather Testbed, in preparation, 2020.
- Hanley, K., Plant, R., Stein, T., Hogan, R., Nicol, J., Lean, H., Halliwell, C., and Clark, P.: Mixing-length controls on high-resolution simulations of convective storms, *Q. J. Roy. Meteor. Soc.*, 141, 272–284, <https://doi.org/10.1002/qj.2356>, 2015.
- Hartley, A., MacBean, N., Georgievski, G., and Bontemps, S.: Uncertainty in plant functional type distributions and its impact on land surface models, *Remote Sens. Environ.*, 203, 71–89, <https://doi.org/10.1016/j.rse.2017.07.037>, 2017.
- Hastings, D. A., Dunbar, P. K., Elphingstone, G. M., Bootz, M., Murakami, H., Maruyama, H., Masaharu, H., Holland, P., Payne, J., Bryant, N. A., Logan, T. L., Muller, J.-P., Schreier, G., and MacDonald, J. S.: The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0, Digital data base on the World Wide Web, available at: <http://www.ngdc.noaa.gov/mgg/topo/globe.html> (last access: 25 October 2017), 1999.
- Heming, J. T.: Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models, *Meteorol. Appl.*, 24, 1–8, <https://doi.org/10.1002/met.1599>, 2017.
- Houldcroft, C., Grey, W., Barnsley, M., Taylor, C., Los, S., and North, P.: New vegetation albedo parameters and global fields of background albedo derived from MODIS for use in a climate model, *J. Hydrometeorol.*, 10, 183–198, <https://doi.org/10.1175/2008JHM1021.1>, 2008.
- Huffman, G.: GPM IMERG Late Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V05, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), <https://doi.org/10.5067/GPM/IMERG/3B-HH-L/05>, 2015.
- Huffman, G.: GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V05, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), <https://doi.org/10.5067/GPM/IMERG/3B-HH/05>, 2017.
- Jin, Z., Qiao, Y., Wang, Y., Fang, Y., and Yi, W.: A new parameterization of spectral and broadband ocean surface albedo, *Opt. Express*, 19, 26429–26443, <https://doi.org/10.1364/OE.19.026429>, 2011.
- Jones, A., Roberts, D. L., and Slingo, A.: A climate model study of indirect radiative forcing by anthropogenic sulphate aerosols, *Nature*, 370, 450–453, <https://doi.org/10.1038/370450a0>, 1994.
- Kain, J. S., Willington, S., Clark, A. J., Weiss, S. J., Weeks, M., Jirak, I. L., and Suri, D.: Collaborative efforts between the United States and United Kingdom to advance prediction of high-impact weather, *B. Am. Meteorol. Soc.*, 98, 937–948, <https://doi.org/10.1175/BAMS-D-15-00199.1>, 2017.
- Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, *Nat. Clim. Change*, 4, 570–576, <https://doi.org/10.1038/nclimate2258>, 2014.
- Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J., Fosser, G., and Wilkinson, J.: Do convection-permitting regional climate models improve projections of future precipitation change?, *B. Am. Meteorol. Soc.*, 98, 79–93, <https://doi.org/10.1175/BAMS-D-15-0004.1>, 2017.
- Klasa, C., Arpagaus, M., Walser, A., and Wernli, H.: An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland, *Q. J. Roy. Meteor. Soc.*, 144, 744–764, <https://doi.org/10.1002/qj.3245>, 2018.
- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in version 4 of the community land model, *J. Adv. Model Earth Sy.*, 3, M03001, <https://doi.org/10.1029/2011MS00045>, 2011.
- Lean, H. and Browning, K. A.: Quantification of the importance of wind drift to the surface distribution of orographic rain on the occasion of the extreme Cockermouth flood in Cumbria, *Q. J. Roy. Meteor. Soc.*, 139, 1342–1353, <https://doi.org/10.1002/qj.2024>, 2013.
- Lean, H., Clark, P. A., Dixon, M., Roberts, N. M., Fitch, A., Forbes, R., and Halliwell, C.: Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom, *Mon. Weather Rev.*, 136, 3408–3424, <https://doi.org/10.1175/2008MWR2332.1>, 2008.
- Li, D. and Shine, K.: A 4-D ozone climatology for UGAMP models, Tech. rep., UGAMP, University of Reading, 1995.
- Lock, A. P.: Stable boundary layer modelling at the Met Office, in: ECMWF/GABLS workshop on “Diurnal cycles and the stable atmospheric boundary layer”, available at: <https://www.ecmwf.int/sites/default/files/elibrary/2012/10770-stable-bounday-layer-modelling-met-office.pdf> (last access: 3 April 2020), 2012.
- Lock, A. P., Brown, A. R., Bush, M. R., Martin, G. M., and Smith, R. N. B.: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests, *Mon. Weather Rev.*, 128, 3187–3199, [https://doi.org/10.1175/1520-0493\(2000\)128<3187:ANBLMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2), 2000.
- Louf, V., Protat, A., Jakob, C., Warren, R., Rauniar, S., Petersen, W., Wolff, D., and Collis, S.: An integrated approach to weather radar calibration and monitoring using ground clutter and satellite comparisons, *J. Atmos. Ocean. Tech.*, 36, 17–39, <https://doi.org/10.1175/JTECH-D-18-0007.1>, 2018.
- Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, 21, 1303–1330, <https://doi.org/10.1080/014311600210191>, 2000.
- Manners, J., Thelen, J.-C., Petch, J., Hill, P., and Edwards, J. M.: Two fast radiative transfer methods to improve the temporal sampling of clouds in numerical weather prediction

- and climate models, *Q. J. Roy. Meteor. Soc.*, 135, 457–468, <https://doi.org/10.1002/qj.956>, 2009.
- Manners, J., Vosper, S. B., and Roberts, N.: Radiative transfer over resolved topographic features for high-resolution weather prediction, *Q. J. Roy. Meteor. Soc.*, 138, 720–733, <https://doi.org/10.1002/qj.956>, 2012.
- Manners, J., Edwards, J. M., Hill, P., and Thelen, J.-C.: SOCRATES (Suite Of Community RADiative Transfer codes based on Edwards and Slingo) Technical Guide, Met Office, UK, available at: <https://code.metoffice.gov.uk/trac/socrates>, last access: 29 June 2018.
- Marbaix, P., Galée, H., Brasseur, O., and Ypersele, J. V.: Lateral Boundary Conditions in Regional Climate Models: A Detailed Study of the Relaxation Procedure, *Mon. Weather Rev.*, 131, 461–479, 2003.
- McCabe, A., Swinbank, R., Tennant, W., and Lock, A.: Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting, *Q. J. Roy. Meteor. Soc.*, 142, 2897–2910, <https://doi.org/10.1002/qj.2876>, 2016.
- McCaull, E. W., Goodman, S. J., LaCasse, K. M., and Cecil, D. J.: Forecasting Lightning Threat Using Cloud-Resolving Model Simulations, *Weather Forecast.*, 24, 709–729, <https://doi.org/10.1175/2008WAF2222152.1>, 2009.
- Met Office Modelling Infrastructure Support Systems Team, MetOffice: Rose, available at: <http://metomi.github.io/rose/doc/html/index.html>, last access: 3 April 2020.
- Miller, D. A. and White, R. A.: A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling, *Earth Interact.*, 2, 1–26, [https://doi.org/10.1175/1087-3562\(1998\)002<0001:ACUSMS>2.3.CO;2](https://doi.org/10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2), 1998.
- Mittermaier, M.: A strategy for verifying near-convection-resolving forecasts at observing sites, *Weather Forecast.*, 29, 185–204, <https://doi.org/10.1175/WAF-D-12-00075.1>, 2014.
- Mittermaier, M. and Csima, G.: Ensemble versus deterministic performance at km-scale, *Weather Forecast.*, 32, 1697–1709, <https://doi.org/10.1175/WAF-D-16-0164.1>, 2017.
- Mittermaier, M. and Roberts, N.: Inter-comparison of spatial forecast verification methods: Identifying skillful spatial scales using the Fractions Skill Score, *Weather Forecast.*, 25, 343–354, <https://doi.org/10.1175/2009WAF2222260.1>, 2010.
- Morcrette, C. J.: Improvements to a prognostic cloud scheme through changes to its cloud erosion parametrization, *Atmos. Sci. Lett.*, 13, 95–102, <https://doi.org/10.1002/asl.374>, 2012a.
- Morcrette, C. J.: Prognostic-cloud-scheme increment diagnostics: a novel addition to the case-study tool kit, *Atmos. Sci. Lett.*, 13, 200–207, <https://doi.org/10.1002/asl.380>, 2012b.
- Morcrette, C. J. and Petch, J. C.: Analysis of prognostic cloud scheme increments in a climate model, *Q. J. Roy. Meteor. Soc.*, 136, 2061–2073, <https://doi.org/10.1002/qj.720>, 2010.
- Morcrette, C. J., O'Connor, E. J., and Petch, J. C.: Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations, *Q. J. Roy. Meteor. Soc.*, 138, 964–979, <https://doi.org/10.1002/qj.969>, 2012.
- Munoz-Esparza, D., Kosovic, B., Mirocha, J., and van Beeck, J.: Bridging the Transition from Mesoscale to Microscale Turbulence in Numerical Weather Prediction Models, *Bound.-Lay. Meteorol.*, 153, 409–440, <https://doi.org/10.1007/s10546-014-9956-9>, 2014.
- Murphy, A. and Winkler, R.: A general framework for forecast verification, *Mon. Weather Rev.*, 115, 1330–1338, 1987.
- Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Wiberg, D., and Shi, X.: Harmonized World Soil Database (version 1.0), FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2008.
- Oliver, H., Shin, M., Matthews, D., Sanders, O., Bartholomew, S., Clark, A., Fitzpatrick, B., van Haren, R., Hut, R., and Drost, N.: Workflow Automation for Cycling Systems: The Cyclops Workflow Engine, *Comput. Sci. Eng.*, 21, 7–21, <https://doi.org/10.1109/MCSE.2019.2906593>, 2019.
- Osborne, S., Abel, S., Boutle, I., and Marengo, F.: Evolution of Stratocumulus Over Land: Comparison of Ground and Aircraft Observations with Numerical Weather Prediction Simulations, *Bound.-Lay. Meteorol.*, 153, 165–193, <https://doi.org/10.1007/s10546-014-9944-0>, 2014.
- Perkey, D. J. and Kreitzberg, C. W.: A time-dependent lateral boundary scheme for limited-area primitive equation models, *Mon. Weather Rev.*, 104, 744–755, 1976.
- Porson, A., Clark, P., Harman, I., Best, M., and Belcher, S.: Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations, *Q. J. Roy. Meteor. Soc.*, 136, 1514–1529, <https://doi.org/10.1002/qj.668>, 2010.
- Price, J. D., Lane, S., and Boutle, I. A.: LANFEX: A Field and Modeling Study to Improve Our Understanding and Forecasting of Radiation Fog, *B. Am. Meteorol. Soc.*, 99, 2061–2077, <https://doi.org/10.1175/BAMS-D-16-0299.1>, 2018.
- Roberts, N. and Lean, H.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Mon. Weather Rev.*, 136, 78–97, <https://doi.org/10.1175/2007MWR2123.1>, 2008.
- Samanta, A., Ganguly, S., Schull, M. A., Shabanov, N. V., Knyazikhin, Y., and Myneni, R. B.: Collection 5 MODIS LAI/FPAR Products, presented at AGU Fall Meeting, San Francisco, USA, 15–19 December 2008, 2012.
- Sellers, P. J.: Canopy reflectance, photosynthesis and reflection, *Int. J. Remote Sens.*, 6, 1335–1372, <https://doi.org/10.1080/01431168508948283>, 1985.
- Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., and Wilheit, T.: The Global Precipitation Measurement (GPM) Mission for Science and Society, *B. Am. Meteorol. Soc.*, 98, 1679–1695, <https://doi.org/10.1175/BAMS-D-15-00306.1>, 2017.
- Smagorinsky, J.: General circulation experiments with the primitive equations: I. the basic experiment, *Mon. Weather Rev.*, 91, 99–164, [https://doi.org/10.1175/1520-0493\(1963\)091<0099:GCEWTP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2), 1963.
- Smith, R. N. B.: A scheme for predicting layer cloud and their water content in a general circulation model, *Q. J. Roy. Meteor. Soc.*, 116, 435–460, <https://doi.org/10.1002/qj.49711649210>, 1990.
- Stratton, R. A., Senior, C. A., and Vosper, S. B.: A Pan-African Convection-Permitting Regional Climate Simulation with the Met Office Unified Model: CP4-Africa, *J. Climate*, 31, 3485–3508, <https://doi.org/10.1175/JCLI-D-17-0503.1>, 2018.

- Tang, Y., Lean, H., and Bornemann, J.: The benefits of the Met Office variable resolution NWP model for forecasting convection, *Meteorol. Appl.*, 20, 417–426, <https://doi.org/10.1002/met.1300>, 2013.
- Thuburn, J. and White, A. A.: A geometrical view of the shallow-atmosphere approximation, with application to the semi-Lagrangian departure point calculation, *Q. J. Roy. Meteor. Soc.*, 139, 261–268, <https://doi.org/10.1002/qj.1962>, 2013.
- Van Weverberg, K., Boutle, I. A., Morcrette, C. J., and Newsom, R. K.: Towards retrieving critical relative humidity from ground-based remote-sensing observations, *Q. J. Roy. Meteor. Soc.*, 142, 2867–2881, <https://doi.org/10.1002/qj.2874>, 2016.
- Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D., Harris, C. M., Lock, A. P., Manners, J. C., Morcrette, C. J., Roberts, M. J., Stratton, R. A., Webster, S., Wilkinson, J. M., Willett, M. R., Boutle, I. A., Earnshaw, P. D., Hill, P. G., MacLachlan, C., Martin, G. M., Moufouma-Okia, W., Palmer, M. D., Petch, J. C., Rooney, G. G., Scaife, A. A., and Williams, K. D.: The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations, *Geosci. Model Dev.*, 4, 919–941, <https://doi.org/10.5194/gmd-4-919-2011>, 2011.
- Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Stratton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copsey, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, *Geosci. Model Dev.*, 7, 361–386, <https://doi.org/10.5194/gmd-7-361-2014>, 2014.
- Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, *Geosci. Model Dev.*, 12, 1909–1963, <https://doi.org/10.5194/gmd-12-1909-2019>, 2019.
- Wilkinson, J. M.: A Technique for Verification of Convection-Permitting NWP Model Deterministic Forecasts of Lightning Activity, *Weather Forecast.*, 32, 97–115, <https://doi.org/10.1175/WAF-D-16-0106.1>, 2017.
- Wilkinson, J. M. and Bornemann, F. J.: A lightning forecast for the London 2012 Olympics opening ceremony, *Weather*, 69, 16–19, <https://doi.org/10.1002/wea.2176>, 2014.
- Wilkinson, J. M., Porson, A. N. F., Bornemann, F. J., Weeks, M., Field, P. R., and Lock, A. P.: Improved microphysical parametrization of drizzle and fog for operational forecasting using the Met Office Unified Model, *Q. J. Roy. Meteor. Soc.*, 139, 488–500, <https://doi.org/10.1002/qj.1975>, 2013.
- Wilson, D. R. and Ballard, S. P.: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model, *Q. J. Roy. Meteor. Soc.*, 125, 1607–1636, <https://doi.org/10.1002/qj.49712555707>, 1999.
- Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., and Morcrette, C. J.: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description, *Q. J. Roy. Meteor. Soc.*, 134, 2093–2107, <https://doi.org/10.1002/qj.333>, 2008a.
- Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., Morcrette, C. J., and Bodas-Salcedo, A.: PC2: A prognostic cloud fraction and condensation scheme. II: Climate model simulations, *Q. J. Roy. Meteor. Soc.*, 134, 2109–2125, <https://doi.org/10.1002/qj.332>, 2008b.
- Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., and Thuburn, J.: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations, *Q. J. Roy. Meteor. Soc.*, 140, 1505–1520, <https://doi.org/10.1002/qj.2235>, 2014.
- Wood, R. and Field, P. R.: Relationships between Total Water, Condensed Water and Cloud Fraction in Stratiform Clouds Examined Using Aircraft Data, *J. Atmos. Sci.*, 57, 1888–1905, 2000.
- Zerroukat, M. and Allen, T.: On the monotonic and conservative transport on overset/Yin-Yang grids, *J. Comput. Phys.*, 302, 285–299, <https://doi.org/10.1016/j.jcp.2015.09.006>, 2015.
- Zerroukat, M. and Shipway, B.: ZLF (Zero Lateral Flux): a simple mass conservation method for semi-Lagrangian-based limited-area models, *Q. J. Roy. Meteor. Soc.*, 143, 2578–2584, <https://doi.org/10.1002/qj.3108>, 2017.