

The DYMECS project: a statistical approach for the evaluation of convective storms in high-resolution NWP models

Article

Accepted Version

Stein, T., Hogan, R., Clark, P., Halliwell, C., Hanley, K., Lean, H., Nicol, J. and Plant, R. S. (2015) The DYMECS project: a statistical approach for the evaluation of convective storms in high-resolution NWP models. *Bulletin of the American Meteorological Society*, 96 (6). pp. 939-951. ISSN 1520-0477 doi: <https://doi.org/10.1175/BAMS-D-13-00279.1> Available at <http://centaur.reading.ac.uk/39369/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

To link to this article DOI: <http://dx.doi.org/10.1175/BAMS-D-13-00279.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

**The DYMECS project:
A statistical approach for the evaluation of convective storms in high-resolution NWP
models**

THORWALD H. M. STEIN*

Department of Meteorology, University of Reading, Reading, UK

ROBIN J. HOGAN

Department of Meteorology, University of Reading, United Kingdom

PETER A. CLARK

Department of Meteorology, University of Reading, United Kingdom

CAROL E. HALLIWELL

Met Office@Reading, Reading, United Kingdom

KIRSTY E. HANLEY

Met Office@Reading, Reading, United Kingdom

HUMPHREY W. LEAN

Met Office@Reading, Reading, United Kingdom

JOHN C. NICOL

National Centre for Atmospheric Science, Department of Meteorology, University of Reading, United Kingdom

ROBERT S. PLANT

Department of Meteorology, University of Reading, United Kingdom

ABSTRACT

A new frontier in weather forecasting is emerging by operational forecast models now being run at convection-permitting resolutions at many national weather services. However, this is not a panacea; significant systematic errors remain in the character of convective storms and rainfall distributions. The DYMECS project (Dynamical and Microphysical Evolution of Convective Storms) is taking a fundamentally new approach to evaluate and improve such models: rather than relying on a limited number of cases, which may not be representative, we have gathered a large database of 3D storm structures on 40 convective days using the Chilbolton radar in southern England. We have related these structures to storm life-cycles derived by tracking features in the rainfall from the UK radar network, and compared them statistically to storm structures in the Met Office model, which we ran at horizontal grid length between 1.5 km and 100 m, including simulations with different sub-grid mixing length. We also evaluated the scale and intensity of convective updrafts using a new radar technique. We find that the horizontal size of simulated convective storms and the updrafts within them is much too large at 1.5-km resolution, such that the convective mass flux of individual updrafts can be too large by an order of magnitude. The scale of precipitation cores and updrafts decreases steadily with decreasing grid lengths, as does the typical storm lifetime. The 200-m grid-length simulation with standard mixing length performs best over all diagnostics, although a greater mixing length improves the representation of deep convective storms.

Introduction

Convective storms are the frequent cause of flash floods in many mid-latitude countries and can be accompanied by other threats such as hail, lightning, and severe winds. These events can have wide-ranging impacts on livelihoods and infrastructure, so the timing and location of convective storms, as well as their evolution, are important to forecast accurately. Numerical weather prediction (NWP) models are now run at convection-permitting resolutions at several operational forecasting centres. For instance the UK Met Office runs its forecast model at a 1.5-km grid length and the German weather service at a 2.8-km grid length (Baldauf et al. 2011). At these resolutions, models are run without a convection parameterization, which improves the representation of mesoscale convective systems (Done et al. 2004) and the diurnal cycle of convection (Pearson et al. 2014). In practice at least 4–5 grid lengths are required to represent a cloud (Lean et al. 2008), so clouds that ought to be at smaller scales are either represented at larger scales or are not represented at all. This is at least partly to blame for some of the known problems in km-scale modeling of convection, such as that the timing of storm initiation and onset of rainfall can be delayed or advanced compared to observations (Kain et al. 2008; Lean et al. 2008) and individual storms can fail to organize into larger systems (Pearson et al. 2014). Nevertheless, we perceive a belief in the wider community that resolutions of order 1 km are sufficient for accurate simulations of convection, and any remaining issues will disappear simply by cranking up the resolution.

In order to test this belief, it is paramount to provide NWP models with benchmark observational datasets of convective storms and suitable diagnostic tools, especially as model development turns its focus to even higher resolution and more elaborate microphysics schemes (including the representation of graupel and hail). A number of recent field campaigns have targeted convective systems with radar and had a major focus on evaluating convective-scale models, for instance the Convective Storm Initiation Project (Browning et al. 2007), the Convective and Orographically induced Precipitation Study (COPS; Wulfmeyer et al. 2008), the Midlatitude Continental Convective Clouds Experiment (Tao et al. 2013), the Tropical Warm Pool - International Cloud Experiment (May et al. 2008), and the 2011–2012 MJO field campaign (Yoneyama et al. 2013). However, few studies have evaluated NWP models statistically against systematic observations of the 3D dynamical and microphysical structure of storms to answer key questions for convective-scale NWP. How does the distribution of storm size, lifetime and total

rainfall compare between model and observations? What is the strength and horizontal scale of updrafts in convective clouds and how accurately do models of different resolution capture them? Do models “converge” on the correct behaviour when resolution is increased beyond a certain point?

These were exactly the questions that we targeted in the DYMECS project (DYnamical and Microphysical Evolution of Convective Storms), in which we used automated radar scanning to track the evolution of the 3D structure of more than 1,000 convective storms. The storms were observed over 40 non-consecutive days between July 2011 and August 2012 under various synoptic conditions, incorporating the majority of days with convective weather during this period. Through the development of specialized diagnostic tools, these observations are providing a powerful constraint on simulations by the Met Office model at resolutions between 1.5 km and 100 m, guiding the specification of physical processes. The large number of storms analysed allows us to confidently judge whether model changes lead to real improvements in the statistical representation of convective storms. Future analysis of model predictive skill for individual convective events can thus be analyzed with an improved understanding of a model’s representation of convective storms. The DYMECS observational data are publicly available at the British Atmospheric Data Centre, but much of the approach we propose would be applicable to data from other field campaigns with a substantial radar presence, such as those mentioned above.

Sidebar: A radar scanning strategy for convection

We have developed an automated radar scanning procedure that allows us to target individual convective storms and track their development. This strategy is essential for scanning convective storms with the Chilbolton radar due to its low azimuthal scanning velocity, but the short time scales over which convective storms evolve make this strategy of use to faster scanning weather radars too. The procedure is composed of two algorithms, the first of which identifies and tracks storm features in the 1-km UK Met Office rainfall-radar data, using a rainfall-rate threshold of typically 4 mm hr^{-1} . We use the auto-correlation between consecutive rainfall scenes to calculate velocity vectors for individual storms (Stein et al. 2014). For each rainfall image, this algorithm produces a list of storms recording their size, lifetime, mean rainfall rate, velocity vector, and location relative to Chilbolton, as well as the locations and values of local rainfall maxima, i.e. the convective “core”. The second algorithm, the “scan scheduler”, scores each storm in the list based on its size and rainfall rate, while scores are reduced based on storm location relative to the current scanning position of the radar. The scan scheduler then prioritizes three storms to

*Corresponding author address: Department of Meteorology, University of Reading, Lyle Building, PO Box 243, Reading, RG6 6BX, United Kingdom.

E-mail: t.h.m.stein@reading.ac.uk

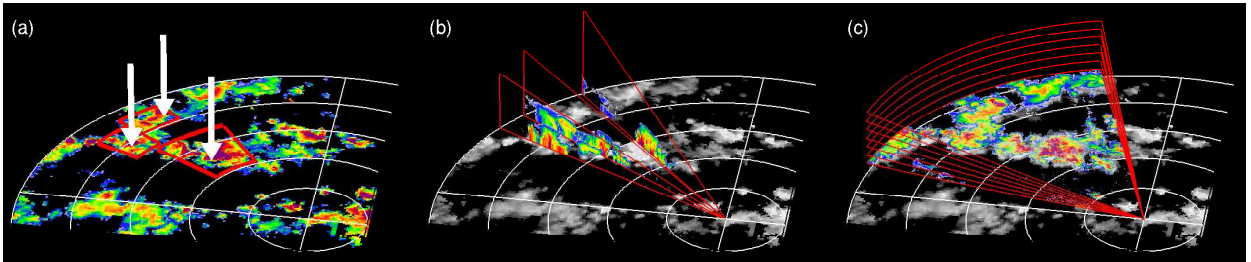


FIG. 1. Illustration of three steps in the scanning strategy. The Chilbolton radar is at the center of the circles, which are 25 km apart. Storms are tracked in the surface rainfall data (panel a, colors) and three storms are prioritized (red boxes), including their locations of rainfall cores (white arrows). The Chilbolton radar is then instructed to first do RHIs through the cores (b, colors show reflectivity) and then stacked PPIs to retrieve three-dimensional structures of storms (c, colors show reflectivity in the bottom scan).

be scanned. Although the automated scanning procedure works well without interference, the user is able to monitor and adjust the list of prioritized storms.

Accounting for the time that has passed since the rainfall-radar data were recorded, the scan scheduler uses the velocity vector to update the storm and core locations, shown in Figure 1a. The scan scheduler issues commands to the radar to automatically scan the storms. Sets of four RHIs target the three most intense cores in the currently prioritized storms (Figure 1b). These commands are issued first as the core locations will be more accurately predicted closer to the time of the new rainfall data, while the cores are also expected to evolve more quickly. The RHIs are followed by stacks of 6–13 PPIs, a minimum of 0.5° apart, that target one or more storm(s) until all prioritized storms have been scanned (Figure 1c). The full cycle can last up to 15 minutes.

Radar observations

During the DYMECS project, we scanned more than 1,000 storms with the 3-GHz Chilbolton Advanced Meteorological Radar in southern England. With a 25-m dish, it is the largest fully steerable meteorological radar in the world. The narrow beamwidth (0.28°) allows for measurements separated by only a few hundred meters, approximately 440 m out to 100 km from the radar, making the instrument ideal for evaluating high-resolution models. The dish size limits the azimuthal scanning velocity to 2°s^{-1} , so an innovative automated scanning procedure was developed to observe individual storms in real-time through their life cycle, as explained in the sidebar “A radar scanning strategy for convection”.

We tracked storms in the Met Office rainfall radar data, which provides radar-derived surface-rainfall rates at 1 km resolution every five minutes (Harrison et al. 2012). Using a rain-rate threshold of 4 mm hr^{-1} to isolate convective storms, storm features with an area of at least 4 km^2 were given a unique identifier and tracked throughout their life cycle, so that life-cycle statistics could be derived for model evaluation.

Using the Chilbolton radar observations, we derived 3D storm volumes from stacks of plan-position indicator scans (PPIs). Following Stein et al. (2014), the storm volumes were re-gridded to a regular Cartesian grid of $333\text{ m} \times 333\text{ m} \times 500\text{ m}$, comparable to the horizontal resolution of the radar data. Doppler winds from the RHI scans were interpolated onto a Cartesian grid with resolution 500 m range by 250 m height and were assumed to be equivalent to the horizontal wind parallel to the plane for scan elevations below 10° . We retrieved vertical velocities in the storm cores from these horizontal winds, using the mass-continuity equation, assuming zero divergence across the plane. The advantage of the mass-continuity method is that its performance can be evaluated using the model vertical wind field, namely by retrieving an updraft velocity from the model horizontal wind fields.

Cloud-resolving model configurations

All DYMECS simulations were run with the Met Office Unified Model (UM), which operates with a non-hydrostatic, deep-atmosphere dynamical core (Davies et al. 2005). The baseline 1500-m simulations were reruns of the UK Met Office operational forecast version of the UM, the UKV, which has 70 vertical levels and is run without a convection parameterization scheme. In these simulations, subgrid mixing is treated using the Lock et al. (2000) non-local boundary layer scheme in the vertical and a Smagorinsky-Lilly scheme in the horizontal with a mixing length of 0.2 times the grid length. Our simulations at 500-m, 200-m and 100-m grid lengths (all with 140 vertical levels) use a local Smagorinsky-Lilly eddy-diffusion scheme, solved explicitly horizontally and implicitly in the vertical direction using the same solution code as the 1D boundary layer scheme.

The subgrid turbulent mixing scheme is an essential component of NWP models with grid lengths of the order 100–1000 m, since such models will partially resolve the inertial subrange. For a Smagorinsky-Lilly type scheme, a ratio of 0.2 between mixing length and grid length was shown by Mason (1994) to best resolve turbulent eddies in

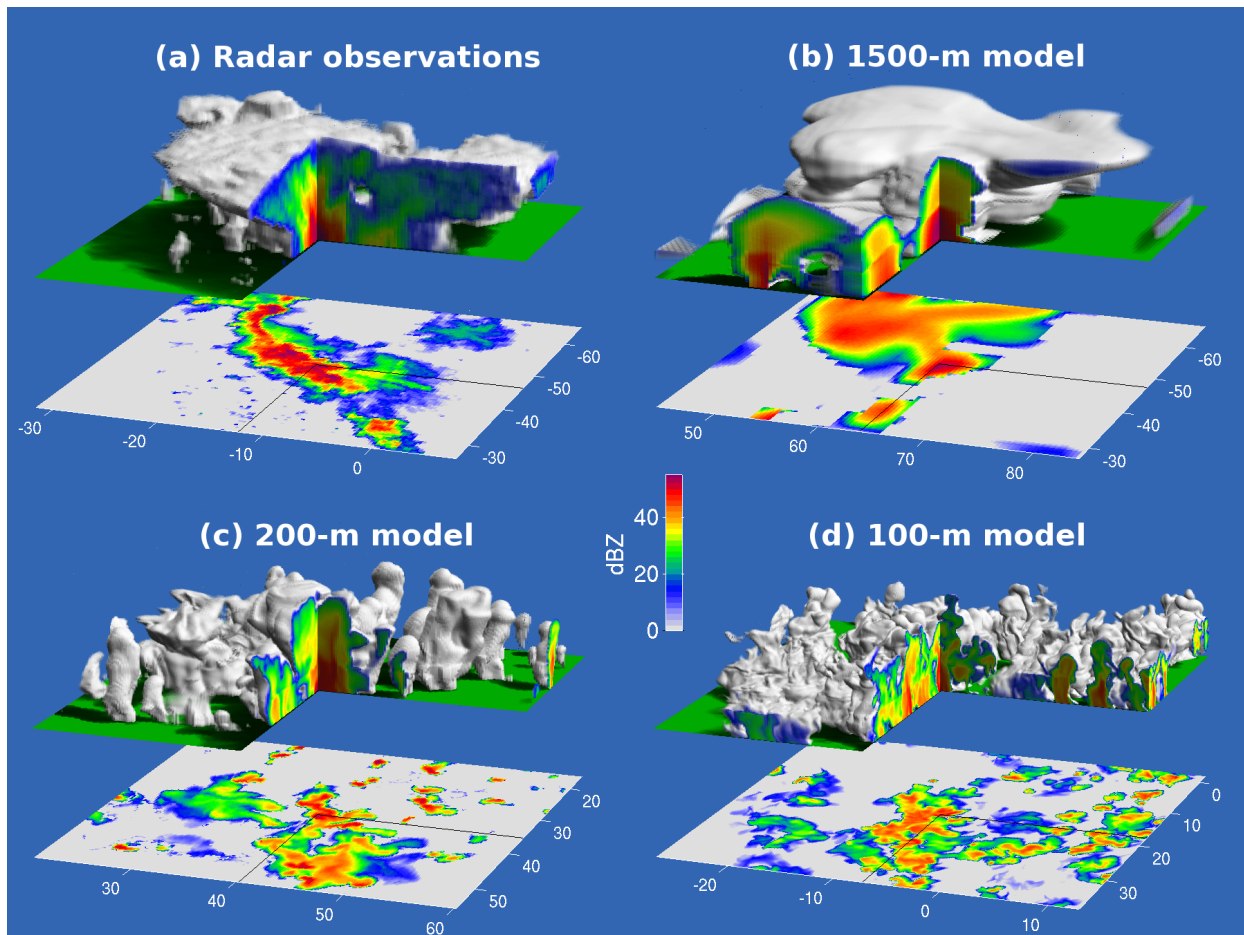


FIG. 2. 3D volume reconstruction of storms observed by the Chilbolton radar on 25 August 2012 and storms of the same day from the 1500-m, 200-m, and 100-m grid-length simulations; time stamps are 1212 UTC, 1500 UTC, 1400 UTC, and 1400 UTC, respectively. The volume is formed by the isosurface of the 0-dBZ reflectivity factor, with slices along the x-axis and y-axis showing the reflectivity structures inside the storm, with distances in kilometers relative to Chilbolton. Underneath each volume, surface reflectivities are shown as a proxy for rainfall rate.

large-eddy simulations (LES); a smaller ratio would lead to grid-scale noise, whereas a larger ratio would lead to overly smoothed flow (e.g., Mason and Callen (1986)). Hanley et al. (2014) showed that the size distribution and intensity of simulated convective storms are sensitive to the mixing-length configuration. We therefore follow their model configurations and test mixing lengths of 300 m, 100 m, and 40 m, which are the default values for the 1500-m, 500-m, and 200-m grid-length simulations, respectively.

In this paper, we primarily focus on convective storms observed on 20 April 2012 and 25 August 2012. The April case will be referred to as the “shower” case as storms did not develop beyond heights of 5 km, while the August case had a large proportion of storms reaching heights above 8 km and will be referred to as the “deep” case; both cases are representative of other DYMECS cases (Hanley et al. 2014). The 1500-m grid-length simulation was initialized

from the 0400 UTC operational UKV analysis, with lateral boundary conditions provided by the Met Office global model. The 500-m simulation was one-way nested in the 1500-m simulation (initialized at 0400 UTC), the 200-m simulation in the 500-m simulation (0700 UTC), and the 100-m simulation in the 200-m simulation (0900 UTC). The UM operates a single-moment microphysics scheme based on Wilson and Ballard (1999), with prognostic treatment of liquid, ice, and rain. Our simulations were run with a diagnostic split between ice crystals and ice aggregates based on cloud-top temperature. Stein et al. (2014) showed that changes to the UM ice microphysics affect the distribution of water contents in the convective storms, but do not greatly affect the overall storm morphology.

To evaluate the simulations against the observed radar reflectivities, a “forward model” was used to calculate reflectivities from the UM hydrometeor fields using the UM microphysics assumptions on the particle size distribution

(see appendix A of Stein et al. (2014)). Examples of the 3D structures thus obtained, as well as cross-sections of reflectivity, are shown in Figure 2 — note that these examples are representative of storm structures observed and simulated, but do not depict a one-to-one correspondence between observed and simulated storms. It is clear from this figure that the 1500-m simulation produces smooth storm structures that vary on scales of several kilometers. The 200-m and 100-m simulations on the other hand show numerous individual convective towers and variations on the sub-kilometer scale, which appear smaller than those in the radar observations, but we need rigorous analysis of the mean morphology and dynamical structures to provide a quantitative evaluation.

Storm life cycles

The analysis of surface rainfall features, including tracking their life cycles, has previously been used to understand model errors in representing convection (e.g. Weusthoff and Hauf (2008); Varble et al. (2011); Caine et al. (2013); Clark et al. (2014)). In their study of multiple DYMECS cases, Hanley et al. (2014) jointly analysed storm-averaged rainfall rate and area and noted that although 500-m and 200-m grid-length simulations generate a similar number of small storms to radar observations, these small storms tend to have rainfall rates a factor two too high. The 1500-m grid-length simulation on the other hand produces storms that are larger and fewer in number than observed.

We build on the Hanley et al. (2014) study by analyzing the evolution of surface rainfall rate and area over storm life times. To isolate “dominant” storms we apply the following rules: (1) for a storm that breaks up into fragments, only the largest fragment maintains the original storm-identifier while all other fragments are given new unique identifiers; and (2) for a merging event, the new storm will maintain the identifier of the largest original storm while all other original fragments are terminated. Thus, we can study the entire storm life cycle from initiation to dissipation, only counting those storms that remain inside a 200×200 km domain centered on Chilbolton. We performed this analysis on the Met Office rainfall radar data and on the five-minute surface-rainfall output from each of the simulations, between 1000 UTC and 1800 UTC, with rainfall rates aggregated onto a 1-km grid.

The number of small showers and short-lived events increases with decreasing grid length, yet these showers may not contribute much to total rainfall. We therefore show in Figures 3a and b the cumulative fraction of the total rainfall from dominant storms by storm duration. In the 1500-m simulation, approximately 20% of rainfall in the shower cases comes from storms lasting less than 2 hours, compared to 60% in the observations; for the deep cases,

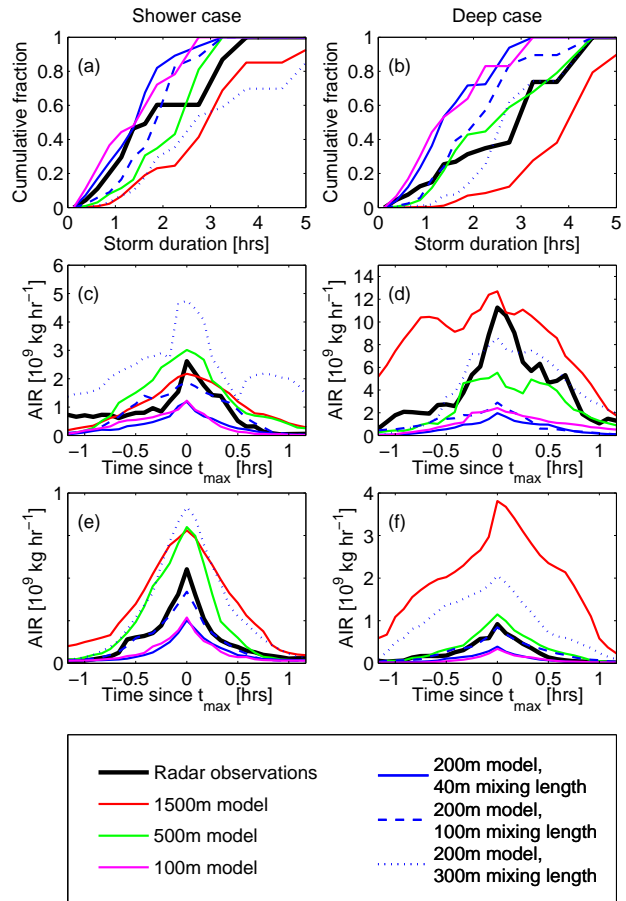


FIG. 3. Life-cycle statistics in simulations with varying grid length and in observations for 20 April 2012 (Shower case, left) and 25 August 2012 (Deep case, right). Only storms that have surface-rainfall rates above 4 mm hr^{-1} over an area of at least 4 km^2 for a duration of at least 30 mins are considered. (a,b) The cumulative fraction of rainfall from storms that have life times shorter than a given storm duration. (c,d) Mean area-integrated rainfall (AIR) with time relative to the time of maximum AIR, t_{max} , with storms weighted by their life-time integrated rainfall and only considering the top 50% by cumulative AIR contribution. (e,f) As in (c,d), but only considering the bottom 50% by cumulative AIR contribution.

these values are 10% and 30%, respectively. Thus, short-lived storms contribute significantly to the observed total rainfall in these cases, yet they are not represented well in the 1500-m simulation. These results agree with previous analyses of storm duration in the UM. Barthlott et al. (2011) have shown that for a COPS case the storms last too long in the UM and that the model overestimates precipitation, while McBeath et al. (2014) showed for a convective cold-air outbreak that the UM storms last 20% longer than those observed. Clearly, the operational forecast model will overestimate storm duration and does not adequately represent short-lived storm life cycles.

The 500-m simulation compares well with observations though it generally underestimates the contribution from short-lived storms in the shower case. The 200-m and 100-m simulations show similar behavior to one another, both underestimating the contribution from storms lasting longer than 2 hours. We note, however, that break-ups and mergers occur more frequently in the 100-m and 200-m simulations than at coarser grid lengths, which may lead to fewer long-lived storms using our metric. Apart from model grid length, storm duration also depends on the treatment of subgrid turbulence. Increasing the turbulent mixing length at 200-m grid length from its default value of 40 m to 100 m or 300 m increases the rainfall contribution from long-lived storms in both cases.

To study the evolution of storm size and intensity, we combine the two variables into a storm-area-integrated rainfall amount, AIR. We center storm life cycles on the time at which they reach their maximum AIR and weight each storm by its life-time integrated AIR. From Figures 3a and b, we note that in the observations long-lived storms dominate total rainfall, while short-lived storms are more numerous (not shown) and will be representative of the “typical” behaviour. We therefore study separately the AIR cycle from the largest AIR contributors in Figures 3c and d and the cycle from the smallest AIR contributors in Figures 3e and f; each weighted-average AIR cycle shown represents 50% of total rainfall from convective storms. Figures 3c and d show that for large AIR storms, the 1500-m simulation compares well with observations around the time of maximum AIR, although this is due to compensating errors of larger but less intense storms, as shown by Hanley et al. (2014). For the shower case, the 500-m simulation behaves comparably to the 1500-m simulation, while for the deep case, it generally underestimates AIR. The 200-m and 100-m simulations again behave similar to one another, both vastly underestimating AIR, likely because they miss the largest storms, which tend to be long-lived. The weighted AIR increases with mixing length, reflecting the increase in rainfall contribution from long-lived storms. The 200-m grid-length with 100-m mixing-length simulation compares very well with the observations for the shower case, while the 300-m mixing-length simulation performs best in the deep case.

The weighted mean from small-AIR storms, shown in Figures 3e and f, clearly indicates that the 1500-m simulation does not represent such storms well, as its peak AIR is a factor 1.5 greater than observed in the shower case and a factor 4 greater in the deep case. The 500-m simulation also has AIR a factor 1.5 greater than observed in the shower case, but compares very well in the deep case. Even for these smaller storms, the 200-m and 100-m simulations underestimate AIR, by about a factor 2 for both cases. This likely relates to these simulations producing numerous storms that are smaller than observed but

more intense, as shown for the 200-m simulation by Hanley et al. (2014). For small-AIR storms, the AIR also increases with mixing length, so that the 200-m grid-length with 100-m mixing length simulation compares exceptionally well with the observations in both cases. The 200-m grid-length simulation with 300-m mixing length, however, has too high AIR by factors 1.5 and 2 for the shower case and deep case, respectively.

The sub-kilometer grid-length simulations clearly outperform the 1500-m grid-length simulation in terms of life-cycle statistics, but we have shown that the results vary with the type of convective storms that we try to simulate, while they are also sensitive to subgrid turbulent mixing length. Furthermore, the 100-m and 200-m simulations have too much of their rainfall contributed by small, but intense, short-lived storms, possibly as storms fail to merge into larger and longer-lived storms. In order to improve these simulations, we will need to better understand and evaluate the microphysical and dynamical structures that are generated.

Storm 3D morphology

In order to evaluate 3D storm structures statistically, we describe them by an area-equivalent diameter at each height, which we then composite over multiple storms to obtain the median storm morphology. Thus, a “typical” storm structure can be represented by the median diameters, shown in Figure 4. Storms are categorized by their maximum height of the 0-dBZ contour (cloud-top height) into three groups with similar numbers of observed storms. Stein et al. (2014) showed how storm structures decrease in width for each individual category as the model grid length is reduced, concluding that the 200-m simulation performs best overall. Since Hanley et al. (2014) showed that the statistics of surface-rainfall areas in high-resolution simulations are sensitive to the turbulent mixing length, we analyse the 3D morphology in the 200-m grid-length simulations for three different mixing lengths and compare these with the 1500-m grid-length simulation and the Chibolton observations.

The smallest scales represented by a model tend to be 4–6 times the grid length (Lean et al. 2008), so it is not surprising that features observed at scales smaller than 6–9 km, are wider in the 1500-m simulation. For instance, the 30-dBZ contour is 7–12 km wide over the three storm categories, about twice as wide as observed (Figures 4a–f). A similar conclusion can be drawn studying the 0-dBZ contour, though it is only about 1.5 times as wide as observed. The 200-m grid-length simulations clearly generate more realistic 3D structures than the 1500-m simulation, while both deep and shallow structures tend to reach greater median sizes when mixing length is increased. For broad structures, such as the 0-dBZ contours in the intermediate and deep storms, simulations with 100-m and

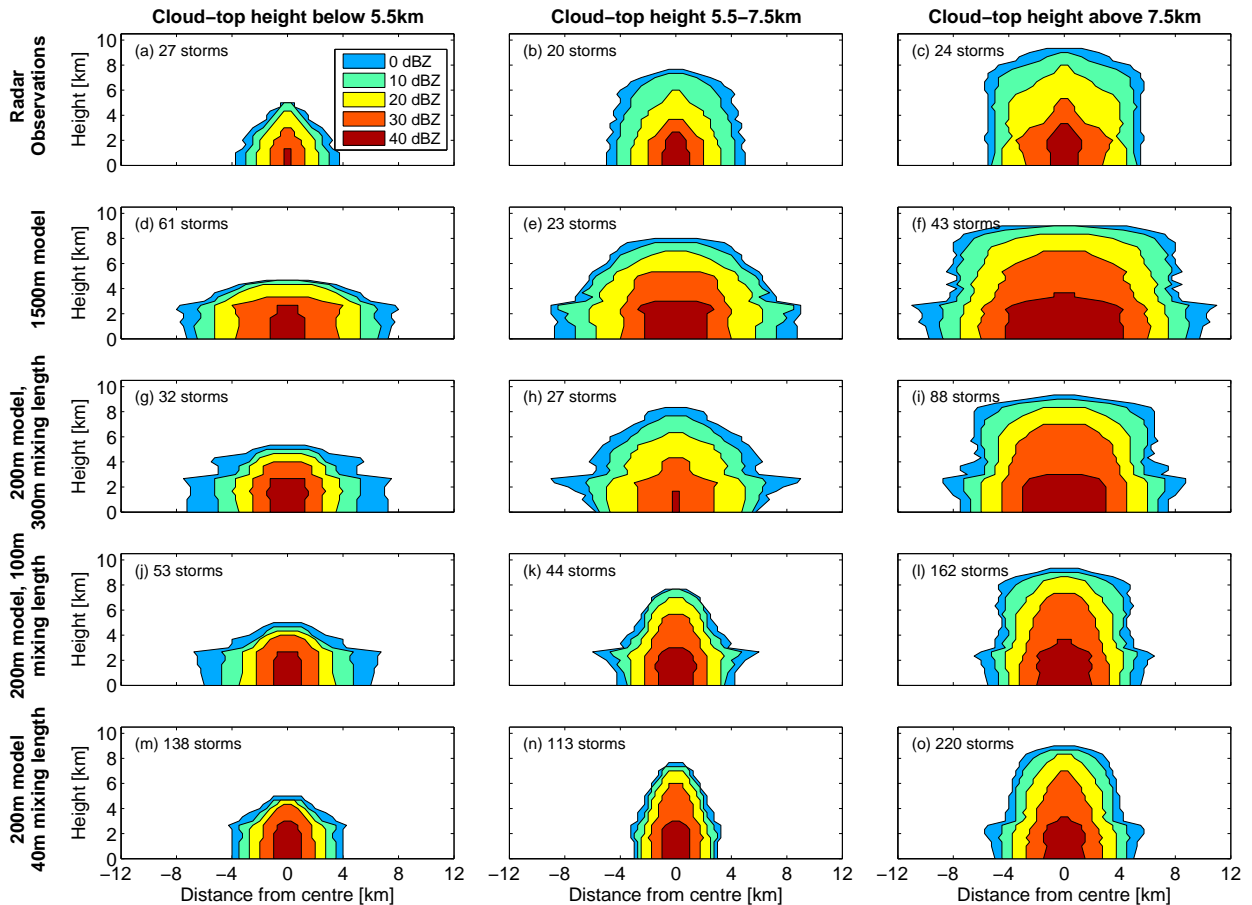


FIG. 4. Storm structures in models and observations for 25 August 2012 between 0900–1700 UTC, represented by the median diameter of reflectivity contours versus height. A rain-rate threshold of 4 mm hr^{-1} and an area threshold of 4 km^2 were used to identify individual storms. Storms are grouped by cloud-top height, namely below 5.5 km (left column), 5.5–7.5 km (middle), and above 7.5 km (right). The number of individual storms in each category is indicated in the top-left corner of each panel. Rows show: observations (a–c), model with 1500-m grid length (d–f), model with 200-m grid length and 300-m mixing length (g–i), model with 200-m grid length and 100-m mixing length (j–l), model with 200-m grid length and 40-m mixing length (m–o).

300-m mixing length compare better with observations. Smaller structures, such as shallow storms and 30-dBZ cores, appear better simulated by the smaller (default) 40-m mixing length. As we increase the mixing length, the total number of storms decreases, which is largely due to a decrease in the number of small storms (Hanley et al. 2014). Thus, it is plausible that the 40-m mixing-length simulation produces median storm structures that are too narrow because too many narrow features develop into tall structures, as we can see for instance in the volume reconstruction in Figure 2c.

We have performed the same analysis for the case of 20 April 2012 (not shown) and note that although the observed storm widths are about 10 km, the 1500-m grid-length simulation produces storm widths still a factor 1.2 too wide. The 200-m grid-length simulations produce storms comparable to the observations, and again, simu-

lations with larger mixing lengths perform best. Both the deep case and the shower case suggest that a smaller grid length produces more realistic storm structures, but that there is no single mixing-length formulation that is satisfactory for all storm classes.

Since the DYMECS simulations were run, the operational UM ice microphysics has changed to the Field et al. (2007) double-moment scheme to determine ice-particle size distribution and changes to the ice-particle fall speed are currently being tested. The effect of changes to model ice microphysics on storm width is small compared to the effect of changes to the model grid length (Bryan and Morrison 2011; Stein et al. 2014). However, these changes may affect the internal reflectivity structure of storms (Stein et al. 2014), so the DYMECS data and methodology provide a useful benchmark for testing such changes.

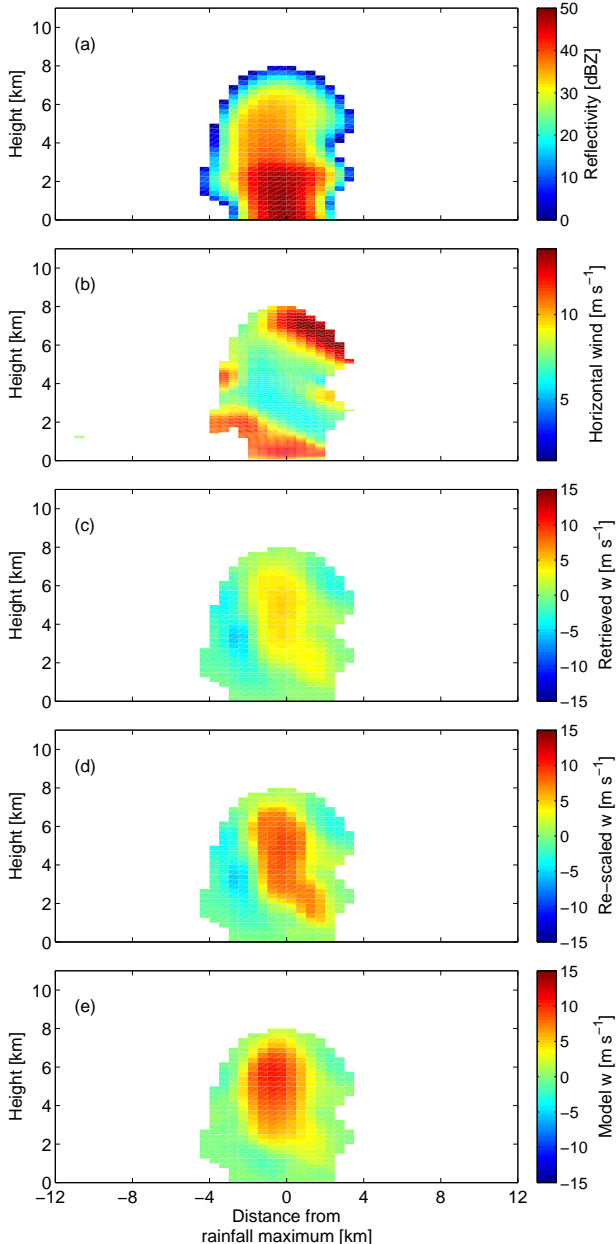


FIG. 5. Illustration of the vertical velocity retrieval for a slice through a storm in the 500-m grid-length simulation, showing forward-modelled reflectivities in panel a. Regions of convergence can be identified as negative gradients in the horizontal winds parallel to the plane in panel b, for instance at the centre of the core below 2 km where the reflectivity is greater than 40 dBZ (see panel a); these regions lead to the greatest vertical-velocity retrievals based on a single component of convergence in panel c. The rescaled single-component velocity estimates in panel d provide an improved representation of the “true” updrafts in panel e.

Convective updrafts

The estimation of 3D wind fields typically requires co-incident observations from two or more Doppler radars;

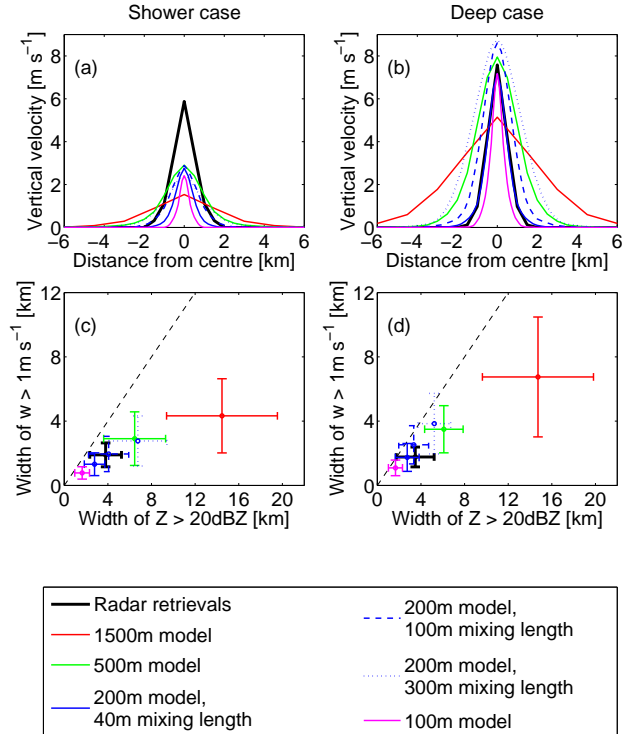


FIG. 6. Updraft statistics in models and observations for 20 April 2012 (shower case, left) for heights between 2–3 km and 25 August 2012 (deep case, right) for heights between 4–5 km (right). For both cases, the heights are about 1–2 km above the 0°C level. (a, b) Mean retrieved vertical velocity centered on distance relative to the location of the peak updraft. (c, d) Mean (symbols) and standard deviation (error bars) of 20-dBZ width versus width of updraft region stronger than 1 m s^{-1} .

such techniques are well established in the literature (e.g. Chong and Testud (1983)). Previous attempts to estimate updrafts from single-Doppler RHI observations require the assumption (e.g., Chapman and Browning (2001)) that all convergence occurs in the line-of-sight, enabling the estimation of vertical velocity based on the mass-continuity equation. When this technique is applied to convective rain cells in particular, updrafts will tend to be underestimated due to the undetected convergence perpendicular to the plane of the RHI. We have developed an innovative new approach to account for this problem statistically, illustrated in Figure 5 using data from the 500-m grid-length simulation. In our method, vertical velocities are retrieved both from the cloud-top down and from the surface upwards, assuming 0 m s^{-1} at the surface and the top; the final estimate shown in Figure 5c is a weighted average between the two retrievals. We found the distribution of radar-derived vertical-velocity estimates to be comparable to the 500-m grid-length simulation, which has the same horizontal resolution as the interpolated radar data. However, the single-component retrieval underesti-

mates the most intense updrafts of the true model vertical velocities. Therefore, we developed a rescaling function derived from the 500-m simulation statistics, which associates the single-component and true cumulative probability functions based on all the storms represented in the 500-m simulations on 25 August 2012. We see in Figure 5c and d that for this example, the strongest vertical velocities are slightly increased and the rescaled velocities are more comparable to the true velocities in Figure 5e; we stress however that the method is not intended to provide the best estimate for an individual slice, but to provide the best estimate of the overall velocity statistics. For the 500-m model, the retrieved peak updraft velocity has a root-mean-squared error of 3.6 m s^{-1} with a standard error of 0.3 m s^{-1} . In this section, radar estimates of vertical velocities are derived and rescaled following this method and are used to statistically evaluate the true model vertical velocities. For a full description of the method, see Nicol et al. (submitted to Quarterly Journal of the Royal Meteorological Society).

In the radar observations, for each RHI scan we consider only the storm with the highest reflectivity observed within 90 km of the radar for inclusion in the velocity statistics presented below. For the target storm, at each height we identify the location of the maximum vertical velocity and calculate the mean vertical velocity as a function of distance from this peak velocity up to the point where the vertical velocity either falls below 1 m s^{-1} or no longer decreases monotonically away from the peak, to avoid any broadening associated with adjacent updrafts. This same methodology was applied to vertical slices through the surface rainfall maxima of each storm extracted from the simulations. We compared the AIR statistics of the storms targeted with RHIs to the storm population in the Met Office rainfall radar data; the storms targeted were distributed almost uniformly amongst the population with above-median AIR. Therefore, only the top 50% of storms according to AIR were included in the statistics for both the observations and simulations. The mean vertical velocity as a function of distance from the updraft core for these storms is shown in Figures 6a and b for 20 April 2012 (shower case) and 25 August 2012 (deep case), respectively.

For the shower case in Figure 6a, the peak of the mean updraft profile in the 1500-m grid-length simulation is about 1.5 m s^{-1} , weaker than observed by a factor 4. The 500-m, 200-m, and 100-m grid-length simulations have their mean updraft peak a factor 2–3 weaker than observed. The mean profile width increases progressively with grid length, from about 1 km in the 100-m to about 3.5 km in the 1500-m grid-length simulation. In the 200-m simulations, this width also increases with increasing mixing length, from 1.5 km with the default 40-m mixing length to about 3 km for the 100-m and 300-m mixing-length simulations. For the deep case in Figure 6b, the

trends with grid length and mixing length are similar to the shower case, but the 200-m grid-length simulation with 40-m mixing length is an excellent match to the observations.

The differences in mean updraft profile between the simulations and the observations have implications for the mass flux of individual updrafts. Assuming that storm width is about the same in both horizontal dimensions, we can estimate that updrafts in the 1500-m grid-length simulation have a mass flux at least an order of magnitude too large. To obtain the same averaged mass flux over all storms, fewer storms will be required in the 1500-m grid-length simulation than in the sub-kilometer simulations and than in the observations. Hanley et al. (2014) found that domain-averaged rainfall rates were largely insensitive to changes in grid length and mixing length for the cases considered here. This suggests that the misrepresentation of updraft size and strength is closely tied to an inaccurate representation of the number of storms in the simulations.

A direct comparison between updraft and cloud morphology is obtained by comparing the width of the updraft profiles (as defined above) to the width of reflectivity profiles (determined in the same manner with a 20-dBZ threshold) in the same vertical slice or RHI scan. From Figure 6c and d it is clear that both the mean updraft and reflectivity widths decrease with model grid length, with the latter result expected from Figure 4. The simulations exhibit a strong correlation between the two widths. The joint comparison of updraft width and storm reflectivity structure shows that the 200-m grid-length simulations perform better than the simulations at other grid lengths. Both the updraft and reflectivity widths increase with increasing mixing length and the 200-m grid-length simulations with greater mixing length behave similarly to the 500-m grid-length simulation by this comparison. We note that when we relax the monotonicity condition on the updraft and reflectivity widths, the updraft widths remain essentially unchanged (not shown), but the reflectivity widths increase by up to a factor 3. This change due to “multi-peaked” updrafts is most pronounced in the 100-m grid-length simulations and in the radar observations, so that the 500-m grid-length simulation and the 200-m simulations with 300-m mixing length perform best in terms of reflectivity width when multi-peaked profiles are considered, while the 200-m grid-length simulations with 40-m and 100-m mixing length consistently perform best in terms of updraft width, regardless of this condition.

Is 200-m resolution good enough?

An accurate representation of the size spectrum of storms is an essential pre-requisite for simulating convective storms, whether achieved by parameterization or explicit representation. A systematic over-forecast of large

storms leads to excessive false alarms. Predictability is a separate issue, and can be addressed either through assimilation of cloud-scale data such as radar (or nowcasting applications) or ensemble techniques, or, ideally, a combination of the two. However, either approach require a faithful representation of storm sizes.

Our diagnostics for convective-storm evaluation show improvements with decreasing grid length, with the 200-m grid-length simulation with 40-m mixing length performing best overall. Nonetheless, there are two aspects of this simulation that have scope for improvement. First, storms are predominantly small and short-lived but intensely precipitating, which points to the need for further refinements in the microphysical assumptions and process rates. Second, deep convective clouds tend to be too narrow. Figures 3 and 4 indicate that both issues can be addressed by increasing the mixing length in the subgrid mixing scheme, but the fact that the default mixing length produces shallow convective clouds of around the right width supports the suggestion of Canuto and Cheng (1997) that the optimum mixing length is flow dependent.

An interesting finding is that storms and updrafts in the 100-m grid-length simulations are by most metrics less realistic than those in the 200-m simulations, being systematically too short-lived and too narrow (see also Stein et al. (2014)). When used in large-eddy models, the mixing length in the Smagorinsky-Lilly scheme relates to the filter scale of eddies much smaller than the large, energy-containing eddies. If the filter scale lies in the inertial subrange, its precise value is a matter of choice (provided we resolve the resulting flow) and it is usual to reduce the filter scale with decreasing grid-length as more of the flow can be explicitly resolved, though it would be equally valid to hold it fixed to demonstrate numerical convergence. However, the finding that updrafts continue to decrease in size down to 100-m grid length demonstrates that the largest energy-containing eddies are not properly resolved or, at least, that the impact of the unresolved eddies on the resolved flow has substantial deficiencies.

Previous studies differ on the question of the resolution at which convergent behavior occurs: Bryan et al. (2003) reported that statistical properties of squall lines simulated at 250-m and 125-m grid length had not converged, and it has been reported that grid lengths smaller than 100 m is required for convergent behavior in simulations of cumulus (Petch et al. 2002) and thermals (Craig and Dörnbrack 2008). Matheou et al. (2011) claimed that a grid length of 20-m was needed to obtain convergence of cloud variables. Conversely, Khairoutdinov and Randall (2006) concluded that their idealised simulations of convection over Amazonia at 100-m and 250-m grid length already showed similar behavior, although differences in interpretation will have arisen because these papers did not use the same definition of “convergence”.

A key difference between previous studies and this one is that we have observational evidence to show which resolution model produces more realistic storms. However, our finding that the 200-m model performs better than the 100-m may be partially due to the fact that the *vertical* resolution of these simulations is constant for all horizontal grid-lengths of 500-m or less (at around 100 m at 1 km altitude and 300 m at 8 km altitude), so the 100-m horizontal grid-length model may not be any better able to resolve large eddies than the 200-m model. Therefore, further work is needed to investigate the characteristics of storms when vertical resolution is improved as well. From a practical forecasting point of view, it is also important to further diagnose what similar behavior to the 200-m model is obtained by running at a resolution of around 500-m but *reducing* the mixing length from the default value at this resolution (Hanley et al. 2014).

Outlook

Sub-kilometer scale models are emerging from being run experimentally to actually aiding forecasts, for instance the London and Weymouth 333-m models (Golding et al. 2013; Wang et al. 2013), model development for the High Definition Clouds and Precipitation for Climate Prediction project, and Environment Canada’s high-resolution model development for outdoor venues at the 2015 Toronto Pan-Am Games. We therefore require novel observational strategies and diagnostic tools for model evaluation. In DYMECS, we have developed a modern test bed for model evaluation with statistics of over 1,000 storms observed with the Chilbolton radar and developed new diagnostics to be used in tandem to highlight model strengths and weaknesses in storm dynamics and microphysics. We have found that for the Met Office Unified Model (UM), a grid length of 200 m performs best in all diagnostics — life cycle statistics, storm morphology, and convective updrafts — but the results are sensitive to the choice of mixing length in the subgrid turbulence scheme. For instance we found that shallow storm structures are better represented by a smaller mixing length, whereas deep storm structures are better represented with a larger mixing length. Also, updraft cores respond differently to changes in mixing length compared to the cores of high reflectivity.

The DYMECS approach could be applied to other radar data sets and models, provided that these include a component at high temporal resolution, e.g. five-minute intervals for storm tracking and a narrow radar beamwidth to resolve storm structures. A key innovation of the DYMECS project has been to show how updraft width and intensity can be estimated from RHIs measured by a single Doppler radar and used to test these crucial aspects in cloud-resolving models. The next challenge is to establish whether better simulations of the character of convective

storms leads to more accurate forecasts, in particular the timing and location of the most intense flood-producing storms.

Acknowledgment. The Chilbolton radar is operated and maintained by the Rutherford Appleton Laboratory. We are especially grateful to Darcy Ladd, Mal Clarke, and Alan Doo at the Chilbolton Observatory for their invaluable assistance with gathering the radar data. We acknowledge use of the Met Office Unified Model, the Met Office rainfall radar data, and the MONSooN system, which is a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council. The DYMECS project is funded by NERC (grant NE/I009965/1).

References

- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Monthly Weather Review*, **139** (12), 3887–3905.
- Barthlott, C., and Coauthors, 2011: Initiation of deep convection at marginal instability in an ensemble of mesoscale models: A case-study from COPS. *Quarterly Journal of the Royal Meteorological Society*, **137** (S1), 118–136.
- Browning, K. A., and Coauthors, 2007: The Convective Storm Initiation Project. *Bulletin of the American Meteorological Society*, **88** (12), 1939–1955.
- Bryan, G. H., and H. Morrison, 2011: Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. *Monthly Weather Review*, **140** (1), 202–225.
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, **131** (10), 2394–2416.
- Caine, S., T. P. Lane, P. T. May, C. Jakob, S. T. Siems, M. J. Manton, and J. Pinto, 2013: Statistical assessment of tropical convection-permitting model simulations using a cell-tracking algorithm. *Monthly Weather Review*, **141** (2), 557–581.
- Canuto, V. M., and Y. Cheng, 1997: Determination of the Smagorinsky–Lilly constant CS. *Physics of Fluids (1994-present)*, **9** (5), 1368–1378.
- Chapman, D., and K. Browning, 2001: Measurements of dissipation rate in frontal zones. *Quarterly Journal of the Royal Meteorological Society*, **127** (576), 1939–1959.
- Chong, M., and J. Testud, 1983: Three-dimensional wind field analysis from dual-Doppler radar data. Part III: The boundary condition: An optimum determination based on a variational concept. *Journal of Climate and Applied Meteorology*, **22** (7), 1227–1241.
- Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of object-based time-domain diagnostics for tracking precipitation systems in convection-allowing models. *Weather and Forecasting*, **29** (3), 517–542.
- Craig, G. C., and A. Dörnbrack, 2008: Entrainment in cumulus clouds: What resolution is cloud-resolving? *Journal of the Atmospheric Sciences*, **65** (12), 3978–3988.
- Davies, T., M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N. Wood, 2005: A new dynamical core for the Met Office’s global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **131** (608), 1759–1782.
- Delanoë, J., R. J. Hogan, R. M. Forbes, A. Bodas-Salcedo, and T. H. M. Stein, 2011: Evaluation of ice cloud representation in the ECMWF and UK Met Office models using CloudSat and CALIPSO data. *Quarterly Journal of the Royal Meteorological Society*, **137** (661), 2064–2078.
- Done, J., C. A. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmospheric Science Letters*, **5** (6), 110–117.
- Field, P. R., A. J. Heymsfield, and A. Bansemer, 2007: Snow size distribution parameterization for midlatitude and tropical ice clouds. *Journal of the Atmospheric Sciences*, **64** (12), 4346–4365.
- Golding, B. W., and Coauthors, 2013: Forecasting capabilities for the London 2012 Olympics. *Bulletin of the American Meteorological Society*.
- Hanley, K. E., R. S. Plant, T. H. M. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2014: Mixing length controls on high resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.2356.
- Harrison, D. L., K. Norman, C. Pierce, and N. Gaussiat, 2012: Radar products for hydrological applications in the UK. *Proceedings of the ICE-Water Management*, **165** (2), 89–103.
- Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Weather and Forecasting*, **23** (5), 931–952.
- Khairoutdinov, M., and D. Randall, 2006: High-resolution simulation of shallow-to-deep convection transition over land. *Journal of the atmospheric sciences*, **63** (12), 3421–3436.
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review*, **136** (9), 3408–3424.
- Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith, 2000: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Review*, **128** (9), 3187–3199.
- Mason, P. J., 1994: Large-eddy simulation: A critical review of the technique. *Quarterly Journal of the Royal Meteorological Society*, **120** (515), 1–26.
- Mason, P. J., and N. S. Callen, 1986: On the magnitude of the subgrid-scale eddy coefficient in large-eddy simulations of turbulent channel flow. *Journal of Fluid Mechanics*, **162**, 439–462.
- Matheou, G., D. Chung, L. Nuijens, B. Stevens, and J. Teixeira, 2011: On the fidelity of large-eddy simulation of shallow precipitating cumulus convection. *Monthly Weather Review*, **139** (9), 2918–2939.

- May, P. T., J. H. Mather, G. Vaughan, K. N. Bower, C. Jakob, G. M. McFarquhar, and G. G. Mace, 2008: The Tropical Warm Pool International Cloud Experiment. *Bulletin of the American Meteorological Society*, **89** (5), 629–645.
- McBeath, K., P. R. Field, and R. J. Cotton, 2014: Using operational weather radar to assess high-resolution numerical weather prediction over the British Isles for a cold air outbreak case-study. *Quarterly Journal of the Royal Meteorological Society*, **140** (678), 225–239.
- Pearson, K. J., G. M. S. Lister, C. E. Birch, R. P. Allan, R. J. Hogan, and S. J. Woolnough, 2014: Modelling the diurnal cycle of tropical convection across the ‘grey zone’. *Quarterly Journal of the Royal Meteorological Society*, **140** (679), 491–499.
- Petch, J. C., A. R. Brown, and M. E. B. Gray, 2002: The impact of horizontal resolution on the simulations of convective development over land. *Quarterly Journal of the Royal Meteorological Society*, **128** (584), 2031–2044.
- Stein, T. H. M., R. J. Hogan, K. E. Hanley, J. C. Nicol, H. W. Lean, R. S. Plant, P. A. Clark, and C. E. Halliwell, 2014: The three-dimensional morphology of simulated and observed convective storms over southern England. *Monthly Weather Review*, doi: 10.1175/MWR-D-13-00372.1.
- Tao, W.-K., and Coauthors, 2013: Precipitation intensity and variation during MC3E: A numerical modeling study. *Journal of Geophysical Research: Atmospheres*, **118** (13), 7199–7218.
- Varble, A., and Coauthors, 2011: Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure. *Journal of Geophysical Research: Atmospheres* (1984–2012), **116** (D12).
- Wang, C., R. Jones, M. Perry, C. Johnson, and P. Clark, 2013: Using an ultrahigh-resolution regional climate model to predict local climatology. *Quarterly Journal of the Royal Meteorological Society*, **139** (677), 1964–1976.
- Weusthoff, T., and T. Hauf, 2008: The life cycle of convective-shower cells under post-frontal conditions. *Quarterly Journal of the Royal Meteorological Society*, **134** (633), 841–857.
- Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quarterly Journal of the Royal Meteorological Society*, **125** (557), 1607–1636.
- Wulfmeyer, V., and Coauthors, 2008: Research campaign: The Convective and Orographically induced Precipitation Study. *Bulletin of the American Meteorological Society*, **89** (10), 1477–1486.
- Yoneyama, K., C. Zhang, and C. N. Long, 2013: Tracking pulses of the Madden–Julian Oscillation. *Bulletin of the American Meteorological Society*, **94** (12), 1871–1891.