Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom

HUMPHREY W. LEAN, PETER A. CLARK, MARK DIXON, NIGEL M. ROBERTS, ANNA FITCH,
RICHARD FORBES, AND CAROL HALLIWELL

Met Office Joint Centre for Mesoscale Meteorology, Reading, Berkshire, United Kingdom

ABSTRACT

With many operational centers moving toward order 1-km-gridlength models for routine weather forecasting, this paper presents a systematic investigation of the properties of high-resolution versions of the Met Office Unified Model for short-range forecasting of convective rainfall events. The authors describe a suite of configurations of the Met Office Unified Model running with grid lengths of 12, 4, and 1 km and analyze results from these models for a number of convective cases from the summers of 2003, 2004, and 2005. The analysis includes subjective evaluation of the rainfall fields and comparisons of rainfall amounts, initiation, cell statistics, and a scale-selective verification technique. It is shown that the 4- and 1-km-gridlength models often give more realistic-looking precipitation fields because convection is represented explicitly rather than parameterized. However, the 4-km model representation suffers from large convective cells and delayed initiation because the grid length is too long to correctly reproduce the convection explicitly. These problems are not as evident in the 1-km model, although it does suffer from too numerous small cells in some situations. Both the 4- and 1-km models suffer from poor representation at the start of the forecast in the period when the high-resolution detail is spinning up from the lower-resolution (12 km) starting data used. A scale-selective precipitation verification technique implies that for later times in the forecasts (after the spinup period) the 1-km model performs better than the 12- and 4-km models for lower rainfall thresholds. For higher thresholds the 4-km model scores almost as well as the 1-km model, and both do better than the 12-km model.

1. Introduction

Many operational forecast centers including the Met Office are moving toward higher-resolution models for short-range weather forecasting applications. Other examples include the Japan Meteorological Agency (JMA; Narita and Ohmori 2007) and Germany's National Meteorological Service, the Deutscher Wetterdienst (DWD; Steppeler et al. 2003). One motivation for this is to provide improved forecasts of hazardous weather and, in particular, severe convection. In a number of centers regional models with grid length in the range 2–4 km have already been implemented, and it is probable that order 1-km-gridlength models will be common within 5–10 yr. While there is often an as-

sumption that a higher-resolution model will automatically lead to more realistic and more accurate forecasts, it is important to examine if this is the case in practice. In particular for operational forecast centers the central question, as addressed by Kain et al. (2007), is whether the extra computer resources required to run high-resolution models produce a worthwhile increase in forecast accuracy.

There are several reasons why high-resolution models might produce improved forecasts. First, the increased resolution is expected to enable the model to represent mesoscale features that would otherwise not be resolved and to represent convection explicitly rather than by a convection parameterization. There are a number of studies (Weisman et al. 1997; Romero et al. 2001; Speer and Leslie 2002; Done et al. 2004) that demonstrate improved representation of thunderstorms and squall lines, etc., as the grid length is reduced toward 1 km. However there is also evidence (Bryan and Rotunno 2005; Petch 2006) that convection is seriously under resolved at 4 km and this is also the

Corresponding author address: Dr. Humphrey W. Lean, Met Office Joint Centre for Mesoscale Meteorology, Meteorology Building, University of Reading, Reading, Berkshire, RG6 6BB United Kingdom.
E-mail: humphrey.lean@metoffice.gov.uk

case, although less obviously, with grid lengths of 1 km. This is expected to lead to problems (for example, with overprediction of rainfall and with delayed initiation).

Second, higher-resolution models are able to make use of high-resolution input data. This may be through the use of high-resolution datasets in the model (for example orography or land use data) or via assimilation of high-resolution data (e.g., radar or satellite data). The second of these is outside the scope of the current paper although assimilation in our model system is considered elsewhere (M. Dixon 2008, unpublished manuscript).

Although the above give reasons to expect benefits from high-resolution models, there is a fundamental limitation with regard to the short predictability times for small-scale structures (Lorenz 1969; Hohenegger and Schar 2007). For example, if an area of heavy showers is correctly forecast it is still very unlikely that every shower will be correctly predicted. This small-scale randomness will not always be a problem—in many cases lines of showers or even individual showers are forced by orography or larger-scale atmospheric features. However, this aspect must always be borne in mind while interpreting and verifying output from high-resolution models (Roberts and Lean 2008). Care must be taken to avoid methods of interpretation that give the user the impression that the individual small cells seen in the model are likely to be correct. Nevertheless, the small-scale structure is likely to convey useful information about the general morphology of the showers, and so forth.

In the Met Office high-resolution models have been feasible since the advent of the nonhydrostatic version of the Unified Model (UM; Davies et al. 2005). Until early 2005 the highest-resolution model being run operationally was the 12-km "mesoscale model" covering the United Kingdom. At this point a 4-km model was implemented over the United Kingdom embedded in a much expanded 12-km model. The intention is to have moved to a U.K. area model with resolution of order 1 km by 2010.

In anticipation of future plans a 1-km model has been implemented in research mode for evaluation. The purpose of this paper is to examine the performance of the 4- and 1-km models specifically for deep convective rain, and to compare them with the 12-km operational model. We first describe the High-Resolution Trial Model (HRTM) system—a suite of 12-, 4-, and 1-km models. A number of aspects of the model representation of convection will then be discussed.

This paper is concerned only with the high-resolution models running without assimilation; that is, the models use starting data with 12-km grid length. The models were run in this way because the work being reported here was carried out before the assimilation system was set up and optimized (M. Dixon 2008, unpublished manuscript). This has the advantage that the properties of the models are not contaminated by effects from the assimilation system. The downside is the time the model takes to spin up high-resolution structure when starting from a 12-km analysis, and this is discussed in the following sections.

2. Model configuration

A suite of (one way) nested models was run with grid lengths of 12, 4, and 1 km. The 12-km model was run for comparison purposes and also to provide boundaries for the 4-km model.

The Met Office's Unified Model (UM), at version 5.2 onwards, solves nonhydrostatic, deep-atmosphere dynamics using a semi-implicit, semi-Lagrangian numerical scheme (Cullen et al. 1997; Davies et al. 2005). The model includes a comprehensive set of parameterizations, including surface (Essery et al. 2001), boundary layer (Lock et al. 2000), mixed-phase cloud microphysics (Wilson and Ballard 1999), and convection (Gregory and Rowntree 1990, with additional downdraft and momentum transport parameterizations). The model runs on a rotated latitude–longitude horizontal grid with Arakawa C staggering and a terrain-following hybrid-height vertical coordinate with Charney–Philips staggering.

At the time of this study the Met Office ran an operational model with horizontal grid length of 0.11° (approximately 12 km) and 146 × 182 points in the horizontal (as shown in Fig. 1). This was one-way nested inside a global version with horizontal resolution 0.83° × 0.56° (approximately 60 km) resolution at midlatitudes. Both models used the same 38 levels spaced nonuniformly in the vertical.

Most of the model configuration for the 1- and 4-km-gridlength models was taken over from the operational 12-km model. The full list of changes is shown in the table at appendix A. In this section some of the key aspects that have been changed are discussed.

a. Domain

The domains that have been used are shown in Fig. 1. The 4- and 1-km models were both run on square domains that were approximately centered on the Chilbolton radar in central southern England. The 4-km model (190 × 190 grid points) extended south to include a good part of northern France. This 4-km model domain is smaller than the full U.K. domain, which is

FIG. 1. Domains used for the 12-, 4-, and 1-km models.

being used for the new operational 4-km model. The 4-km model used the same 38 levels in the vertical as were used in the operational 12-km model.

The 1-km domain was 300 × 300 grid points, which was the largest that was practical to run with the computer resources available at the time. This model used lateral boundary conditions from the 4-km model and had 76 levels in the vertical (the 38-level set doubled so that every other level was unchanged to minimize interpolation). Previous work (Lean and Clark 2003) showed that enhanced vertical resolution can have definite benefits, in particular, for resolving slantwise structures.

b. Convection

A key issue in these models is the convective parameterization. The 12-km-gridlength operational UM uses a mass flux convection scheme with convectively available potential energy (CAPE) closure (Gregory and Rowntree 1990). This scheme is designed on the assumption that there are many clouds per grid box, an assumption that is already marginal at 12 km but even more questionable at higher resolutions. The 4-km model tends to have a different behavior according to whether or not the convective parameterization is in-

cluded. With no convective parameterization, the large grid length relative to the typical size of developing clouds means that the model tends to delay convective initiation and then produce too few showers, which are then typically much too heavy and as a result there is too much rain overall. The upside, however, is that the organization of showers is often well treated. In situations when only small/shallow showers are expected, explicit convection may not be initiated at all and no rain is produced. With the standard convective parameterization included the convection parameterization may act to remove instability before showers can be explicitly represented by the model dynamics. If this happens the dynamical organization of showers is not properly represented (which impacts the diurnal cycle) and the intensity of rainfall can be grossly underestimated. In some situations the convection scheme can introduce large horizontal temperature (or humidity) gradients, which then feed back on to the dynamics and lead to the generation of spurious rainfall features.

In an attempt to alleviate these problems (described in Roberts 2003) the 4-km model uses a modified version of the convection scheme in which the mass flux at cloud base is limited. (Roberts 2003). The rationale behind this is to allow the model to generate convection explicitly in situations when showers are large enough to be resolved on the grid, but still allow the convection scheme to represent the effects of weaker convective clouds that would otherwise be missed (because they cannot be resolved on the grid). This solution works better than either of the two extreme possibilities (standard convection scheme or no convection scheme); however, there is no single tuning of the modified scheme that is suitable for all convective situations, and a pragmatic decision about the most generally appropriate setting has to be made. In the current work, the 4-km model has been run with the scheme tuned to allow most of the convection to be represented explicitly. Although this results in some weaker showers being missed it means that the larger storms are better represented. In parts of the world where convective storms are typically larger than over the United Kingdom—for example, the central United States—it seems to be generally accepted that it is satisfactory to run at this resolution with no convective parameterization (Kain et al. 2007).

In contrast to the 4-km model, the 1-km model has a small enough grid length to represent many situations without the need for a deep convection scheme and generally produces better results when run in this way. The representation is still not perfect, however, and we are intending to experiment with a shallow convection scheme in the future.
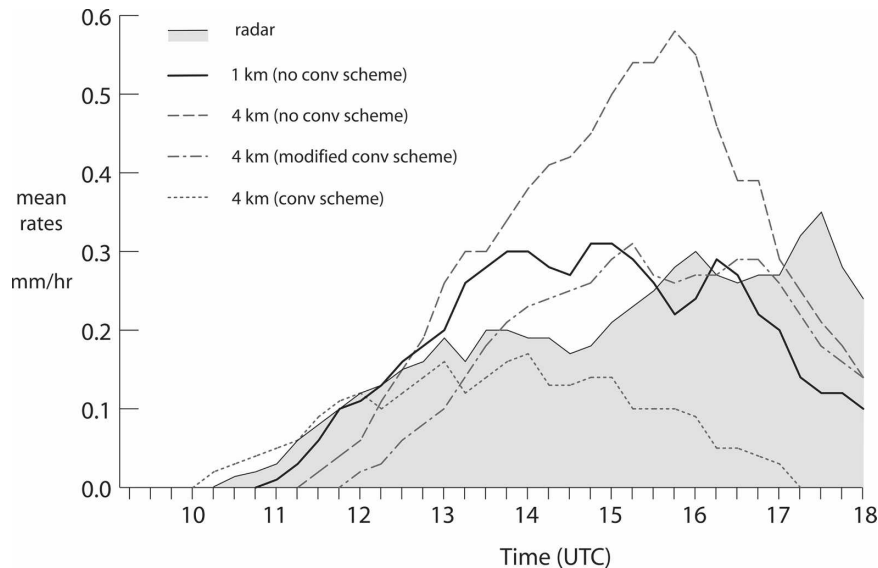
FIG. 2. Domain-averaged rain rates against time for the 3 May 2003 case.

The above points are illustrated by Fig. 2, which shows the 1-km domain area-averaged rainfall rate against time for a case from 3 May 2002 in which scattered showers organized into bands. The 4-km model with convection scheme initially does well but fails to reproduce the organization later in the period resulting in gross underestimates of rainfall. Without a convection scheme the 4-km model has delayed initiation but later produces around a factor of 2 too much rain. The 4-km model with modified convection scheme generally does well but the delay in initiation is increased. The 1-km model with no convection scheme produces a generally good representation; the delay in initiation is less than in the 4-km model and the later phases are captured reasonably well.

c. Advection of rain

While it is well known that the nature of microphysics and turbulence parameterizations has a considerable impact on the evolution of convective-scale models, it is not clear what level of complexity is needed for forecasting (as other factors may dominate forecast error). The microphysics scheme in the convective-scale UM is based on that of Wilson and Ballard (1999). This has been extensively modified to include more prognostic variables. Up to six bulk moisture variables can be used (vapor, cloud water, rainwater, ice, snow, and graupel), with a single moment (the mixing ratio) describing each. However, various options have been implemented to allow diagnostic treatment of some variables. In practice, for U.K. convection, it has been found to be very difficult to show any benefit from

separate treatment of ice and snow, and, instead, a diagnostic split has been retained for the results presented here. Likewise, while it is no doubt important, little benefit has been demonstrated in using graupel. The systematic impact of prognostic rain is, however, clear. Prognostic rain has been shown to improve the spatial distribution of rain relative to the mountains in cases of orographic rainfall and also has a clear systematic impact on lifetime of convective cells. The results presented here thus represent a baseline using a simplified, four prognostic scheme (vapor, cloud water, rainwater, ice + snow). The more detailed assessment of the impact of additional microphysical variables (including higher-order schemes) will be the subject of future work

d. Diffusion–turbulence

The UM does not require additional diffusion for stability. The 12-km model is run without diffusion apart from "targeted diffusion" of moisture, which has been introduced to control a tendency to produce grid-point storms. This essentially applies very localized diffusion of moisture where vertical velocities exceed a given threshold. In practice it operates very rarely.

The 4- and 1-km models benefit from use of near-horizontal diffusion to control the scale of convective cells and prevent them collapsing to the grid scale. In common with many NWP models, the UM uses horizontal "hyperdiffusion" along horizontal model surfaces; terms are added to the model of the form (considering 1D for simplicity)

$$\frac{\partial \Psi}{\partial t} = (-1)^{p+1} K \frac{\partial^{2p}\Psi}{\partial x^{2p}}. \tag{1}$$

This damps a wave with wavelength $n\Delta x$ with e-folding time

$$\tau_n = \left[K\left(\frac{2\pi}{n\Delta x}\right)^{2p} \right]^{-1}. \tag{2}$$

Here $p = 1$ corresponds to conventional Fickian diffusion, while larger integers produce a more scale-selective damping. The choice of diffusivity K has been largely through tuning, but some guidance may be gained by considering the diffusivities that would arise from a more adaptive Smagorinsky approach to diffusion, in which the diffusivity is a function of the local shear; that is, in a shear layer,

$$K = \lambda^{2p} \frac{\partial V}{\partial x}. \tag{3}$$

It follows that

$$\frac{\tau_n}{\Delta t} = \left[C'\left(\frac{2\pi c_s}{n}\right)^{2p} \right]^{-1}, \tag{4}$$

where C' is the shear Courant number, given (in a shear layer) by

$$C' = \left| \frac{\partial V}{\partial x} \right| \Delta t, \tag{5}$$

and $c_s = \lambda/\Delta x$. We anticipate c_s to be constant (typically ~ 0.2) in the Smagorinsky formulation and C' to be less than 1, but, in a well developed turbulent simulation, we would anticipate C' to be of order 1 (otherwise we would be able to run with a much longer time step). The implication of this is that we anticipate that it is appropriate to choose diffusivities that correspond to dissipation measured in time steps, given that the time step is chosen appropriately for the resolution of the model, and that an e-folding time of the smallest wavelength waves of a few time steps would be broadly consistent with the (maximum) diffusivity that would arise from a Smagorinsky approach.

This is a consistency argument rather than a derivation, since it assumes that the time step is chosen to suit the problem but it has provided useful practical guidance. In practice, $p = 2$ (∇^4) diffusion with diffusivities corresponding to an e-folding time for $2\Delta x$ wavelength waves of about eight time steps has been found to give the best results (of those available with this scheme) both in terms of size and structure of convective cells and power spectra. Nevertheless, subsequent sections will show that deficiencies remain.

e. Initial and boundary data

The 4-km model described in this paper was driven by boundary conditions from the 12-km model. The 12-km model was run with the same configuration as the operational 12-km model. This includes a 3-h assimilation cycle including three-dimensional variational data assimilation (3DVAR; Lorenc et al. 2000) for most observations types and also a nudging scheme for cloud data and for precipitation via latent heat nudging (Jones and Macpherson 1997). The assimilation period for each cycle runs from $T - 2$ to $T + 1$ so the best analysis data are available at $T + 1$ (even though this is an hour later than the analysis time). Initial and boundary data for the 12-km model were provided from operationally archived data with the initial data from the 12-km model and the boundary data from the global model. The 4- and 1-km models ran every 3 h from $T + 1$ analyses from the 12-km model to provide start data.

3. Characteristics of the model forecasts

a. Introduction

In this section we cover a number of aspects of the representation of convection in the high-resolution models. We focus on surface rainfall partly because this is easy to compare with the network radar output and partly because this is the main quantity of practical concern.

The models have been run on a number of cases from the summers of 2003, 2004, and 2005 (see table in appendix B). The 2005 cases were intensive observing periods (IOPs) of the Convective Storms Initiation Project (CSIP; Browning et al. 2007). The cases were all convective and ranged from very heavy organized storms to light, scattered showers. For each case four forecasts were run at 3-h intervals covering the period of interest. Since the primary interest of the project was in nowcasting and short-range forecasting, the 1-km model was run out for 6 h from the analysis time at $T + 1$ (i.e., out to $T + 7$). The (less computationally intensive) 4- and 12-km models were run out for 11 h. In this section aggregated statistics are presented including data from 2003, 2004, and 2005. This gave a total of 64 forecasts from 16 cases.

b. Initiation of convection

A key aspect of forecasting severe storms is to represent correctly the initiation of convection. As discussed in section 2b we generally find that our model runs resolving convection explicitly have a delay in initiation compared to those using a convection scheme.

TABLE 1. Initiation delay relative to radar in various runs in hours (to nearest 15 mins). Negative values mean the model initiated before the radar. The numbers given under standard error are the standard deviation of the estimate of the mean calculated as $\sigma/\sqrt{N}$.

Run	12 km	4 km	1 km
0600 UTC 28 Aug 2003	−2.00	2.75	1.25
0900 UTC 27 Apr 2004	−2.00	0.50	−0.75
0600 UTC 20 Jul 2004	−1.00	−0.25	−0.75
0900 UTC 20 Jul 2004	−1.25	0.00	−0.75
0900 UTC 29 Jun 2005	−1.25	1.25	−0.75
0300 UTC 28 Jul 2005	−0.50	0.50	0.00
0600 UTC 28 Jul 2005	−1.00	0.75	0.25
Average	−1.29	0.79	−0.23
Standard error	0.21	0.38	0.28

To get a more systematic idea of the initiation times in the three models each run of the 2003, 2004, and 2005 cases was examined to look for clean comparisons of initiation. Initiation was defined as when the first rain (over 0.05 mm h^{-1}) appeared in the model or radar field. The only runs used were those that met the following criteria:

1) All three models and the radar-initiated showers during the run and the model showers in all models were clearly identifiable with the showers that initiated in the radar.
2) The initiation was neither so close to the start of the run (in the first 2 h) or so close to boundaries (within approximately 75 km) that it might have been contaminated by spinup effects.

Only 7 runs were found that met these criteria, and the results are listed in Table 1. Because of the small-

ness of the sample it is reasonable to ask if these results have any statistical significance. T tests have been performed on the data and are found to lead to the same conclusions as the use of standard errors as shown in the table. On average the 12-km model initiates too early by around 1.5 h. This is because, despite the trigger function in the Gregory Rowntree scheme, the convection scheme fires instantly when CAPE appears with little convective inhibition (CIN). In the case of the explicit convection in the 4- and 1-km models it is known from idealized studies that initiation takes place more rapidly as the grid length is reduced (Petch 2006). One aspect is that the longer grid length models tend to have a great deal of diffusion compared to the real atmosphere. Larger grid length models are also likely to be less effective at eroding lids because the effect is less concentrated at one point. The figures shown in the table imply that the 1-km model is the best of the three in terms of initiation time—it is the only model whose initiation time is within a standard error of being correct.

There is a second aspect to the initiation issue that concerns the time convection takes to initiate when air enters the domain of the model from the boundary. An example of this is shown in Fig. 3, which shows the 4-km model rainfall compared to the radar for a case of showers in a westerly flow. Although the showers over England are represented, there is a complete absence over Ireland and the western third of the domain generally. This effect has been observed on numerous occasions. It is caused by air entering the domain from the boundary conditions that were provided by the 12-km model, which is running with a convection scheme. It
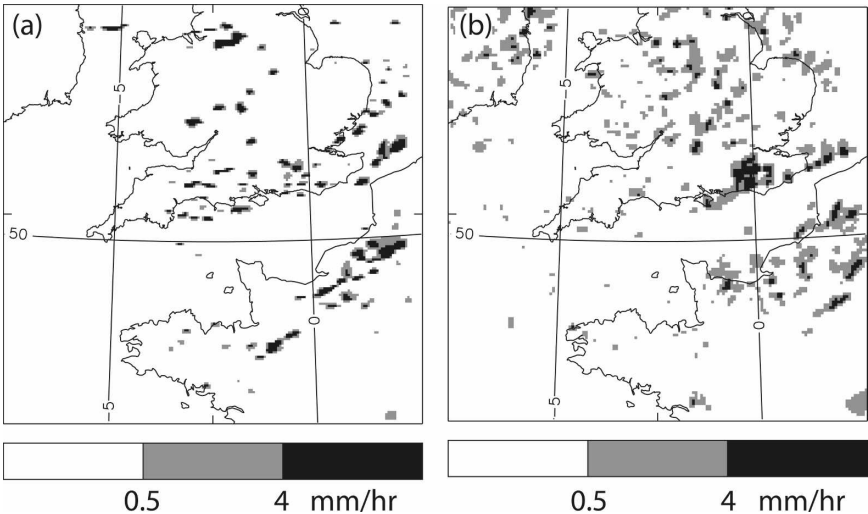

FIG. 3. Rainfall rates at 1300 UTC 25 Aug 2005 from (a) 4-km model run started at 0600 UTC and (b) radar.
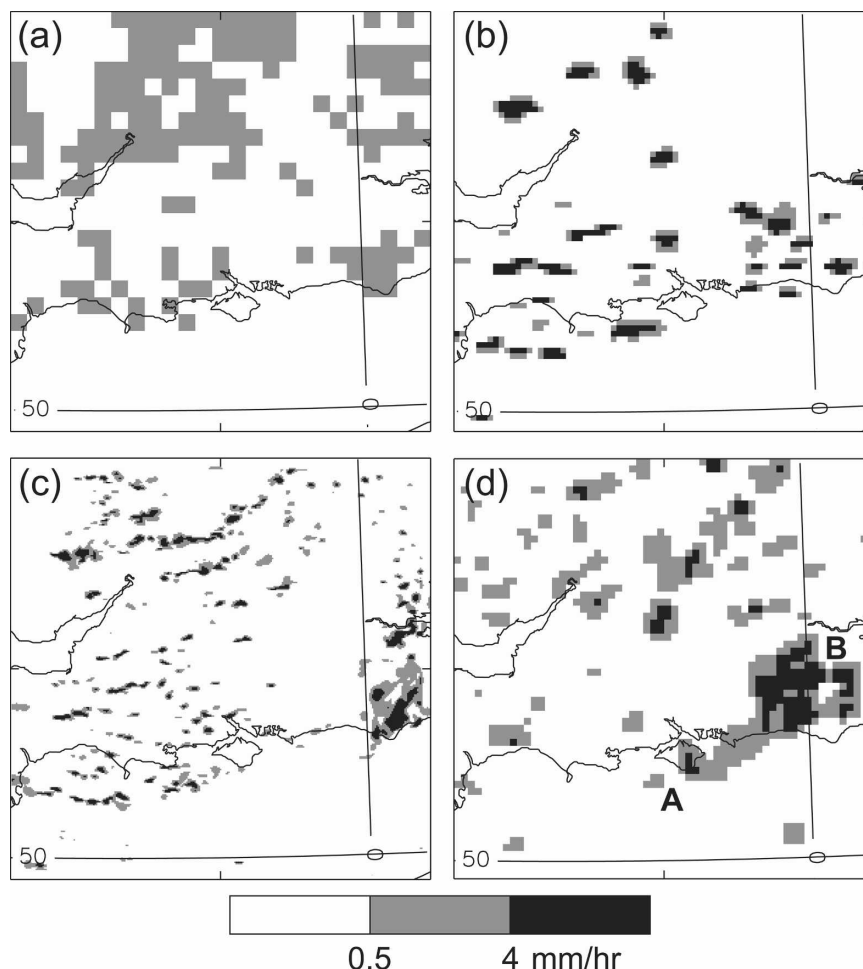
FIG. 4. Instantaneous rainfall rates for 1300 UTC 25 Aug 2005 with model runs started at 0700 UTC using 0600 UTC data: (a) 12-km model, (b) 4-km model, (c) 1-km model, and (d) 5-km radar data. The main squall line is the feature between the letters "A" and "B."

takes a finite time for the explicit convection to initiate as the air enters the domain, and this can be seen as a strip along the inflow boundary without any showers. The width of this region will depend on factors such as the wind speed and the forcing of convection, but it is clear that forecasts in areas close to upwind boundaries should be treated with caution.

The above discussion shows that the treatment of convection in the larger-scale model providing the boundary conditions may also have a large effect on the representation of convection in the high-resolution model. This is revealed by a detailed study of another set of models that has been presented by Warner and Hsu (2000). We have also found that the representation is sensitive to the treatment of convection in the driving model. For example, if we drive the 1-km model with a 4-km model without a convection scheme the initiation at the boundary problem is changed, with the large

shower cells in the 4-km model having to adjust to a smaller scale. Unfortunately, in an operational system there is likely to be very little flexibility in this regard, since the 4-km model will, itself, be used to produce forecast output.

c. Evolution of convection

We now look at the evolution of the rainfall pattern once the convection has initiated. Figure 4 shows an example of the rainfall fields from the three models compared to the radar for CSIP IOP18, which was a case of a squall line (marked AB) in a westerly flow. Model results from this case are discussed in more detail by P. A. Clark et al. (2008, unpublished manuscript). Observations show that the squall line is associated with a convectively generated cold pool. This appears in the 4- and 1-km models and is shown in Fig.
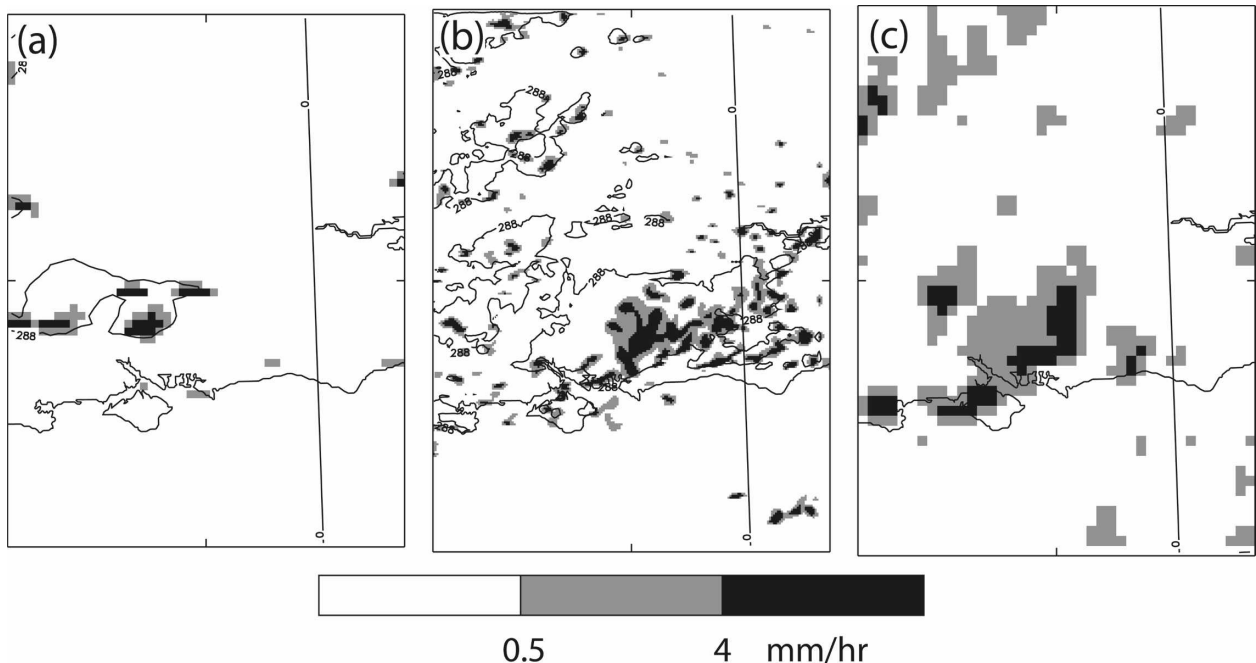
FIG. 5. Instantaneous rainfall rates for 1200 UTC 25 Aug 2005 with model runs started at 0700 UTC for the (a) 4-km model, (b) 1-km model, and (c) radar data. 288-K contours of the 1.5-m temperature field in the two models are also shown in order to show cold pool.

5 as one contour of the surface temperature fields along with the precipitation fields at 1200 UTC. Along with the main squall line there are also lighter showers further west, behind the main feature. The 12-km model does have a feature in the same place as the heaviest part of the squall line (associated with an upper trough) but gives no hint of the very heavy rain. The 4-km model has a cluster of large cells in place of the squall line. The 1-km model also correctly picks out the very heavy feature and has it as a continuous area rather than a number of cells. It has, however, moved it somewhat too far forward by 1300 UTC (although it is more correct at 1200 UTC, implying that the error could be related to the proximity of the boundary). The showers behind the heavy feature in Fig. 4 show the characteristic properties of the models. The 12-km model has a general area of convective rain but no realistic shower structure. The 4-km model has large showers that are too heavy, and the 1-km model has too many showers that tend to have high peak rain rates but are too small. These differences in the sizes of the convective cells reflect the difference in the model grid length and will be discussed later. An examination of a time sequence of rainfall fields would show that, unlike in the 12-km model, individual convective cells are advected with the flow. The more realistic organization in the 4- and 1-km models is a consequence of the convective rain being produced explicitly. In the rest of this section we look in

more detail at some aspects of the representation of convection.

We start our analysis with the overall amount of precipitation being produced by the model. Figure 6 shows the average precipitation rate over the area of the 1-km domain for the various models as a function of time after the analysis, averaged over all the forecasts run on all the cases. The 12-km model produces nearly a factor of 2 too much rain initially, but this gradually reduces to around the correct value by about $T + 8$ (and is some-
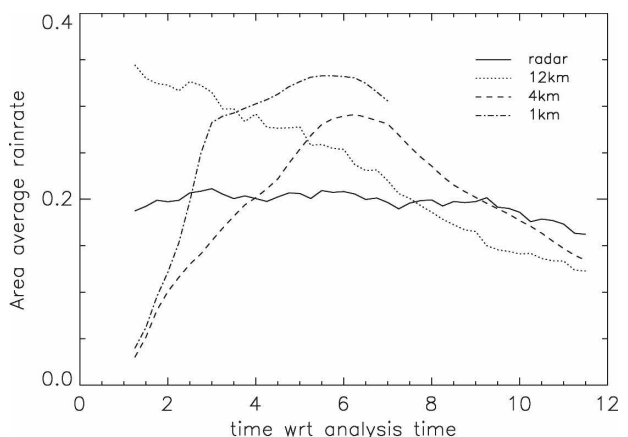

FIG. 6. The 1-km domain area-averaged rainfall rates as a function of time after analysis averaged over all runs of the 2003, 2004, and 2005 cases.
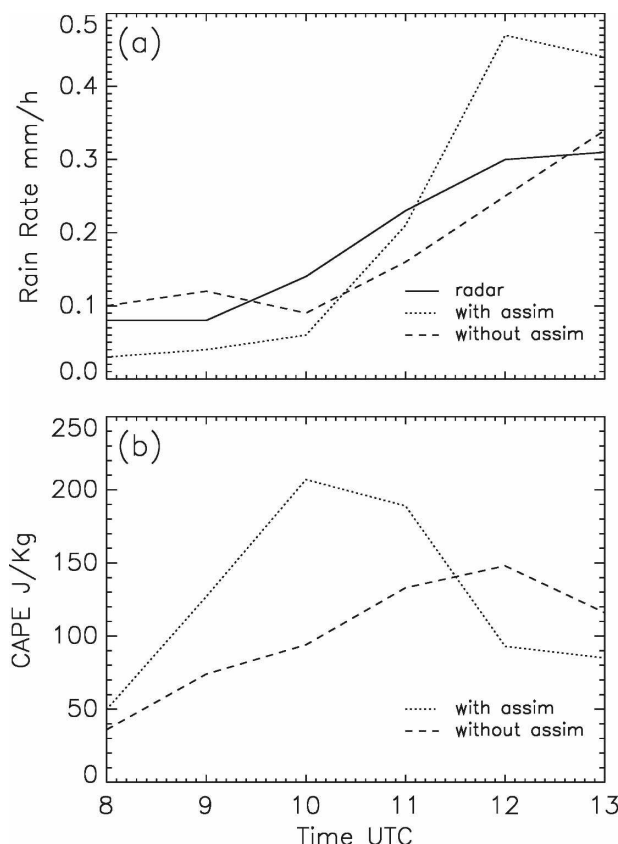
FIG. 7. The 1-km domain area-averaged (a) rain rates and (b) CAPE from the 4-km model plotted against time (UTC) for 0600 UTC forecasts on 13 May 2003 comparing a run with assimilation and a run without spinning up from the 0700 UTC $T+1$ 12-km analysis. The corresponding average rain rate from the radar data is also shown in (a) for comparison.

what too low thereafter). It is thought that the initial overprediction is related to the data assimilation (M. Dixon 2008, unpublished manuscript). As discussed in section 3b this overprediction of rain by the 12-km model is likely to have an impact on the convection in the 4- and 1-km models nested inside it.

The 4- and 1-km models both start with very small precipitation rates at $T+1$ when the models are initiated. This is because, in these convective situations, all of the rain in the 12-km model is represented in the convection scheme, whereas the 4- and 1-km models represent the convection explicitly. The curves generally increase to $T+6$ as the explicit convection spins up. By $T+6$ the values are around 50% higher than those observed by the radar. It is thought that this overshoot is related to the spinup in that extra CAPE builds up in the time when there is insufficient convection, which is then released. Figure 7 shows the CAPE and rain rates in one of the cases (13 May 2003) comparing

the 4-km model using a 12-km analysis as start data with a second 4-km model running with an assimilation cycle. Although the model including assimilation is not discussed in this paper, it is shown here as a point of comparison because the starting analysis already includes the showers at 4-km resolution. This model has no spinup effects and reproduces the rainfall rates much more accurately, whereas the spinning up model has the low rain rates initially followed by an overshoot. The lower frame of the figure compares the domain-averaged CAPE for the same two runs and shows the early build up of CAPE in the spinup run, which is then released later. Returning to Fig. 6, the 1-km has a faster spinup and reaches higher values than the 4-km model. The faster spinup is associated with the smaller grid length allowing convection to start with smaller cells, and this leads, at least initially, to higher values. Unfortunately, we were unable to run the 1-km model for long enough to see if the 1-km model rain rates reduce to below those of the 4-km model after $T+7$, but looking at the curves it seems unlikely. This would imply that the 1-km model has an inherent tendency to produce more rain than the 4-km model, which may be related to the tendency to produce too many cells as discussed below.

After $T+6$ in Fig. 6 the 4-km average reduces as the forecast length is increased to eventually become comparable to the radar values by $T+9$ and onward. The simplistic conclusion from this would be that the overprediction is entirely a spinup issue. Caution must be used here, however, since the radar rainfall rate can be seen to be significantly falling toward $T+12$. The times of the forecasts were generally chosen in each case so that the period of convection was covered by the 6-h 1-km model runs. Hence the runs out to $T+12$ are likely to disproportionally include times when the convection was decaying. Previous tests with longer 4-km runs in convective situations imply that the overprediction may be proportional to the overall rainfall rate. It is, therefore, still possible that there is some inherent predisposition of the high-resolution models toward overprediction because the convection is under resolved.

The conclusion from Figs. 6 and 7 is that a real forecast system in which useful forecasts are required before $T+9$ must include some method to allow high-resolution features to propagate from one forecast cycle to the next. This will avoid the spinup and overshoot problems mentioned above. A satisfactory way to do this would be to include an assimilation system in the high-resolution models. Some work on this is described by M. Dixon (2008, unpublished manuscript).
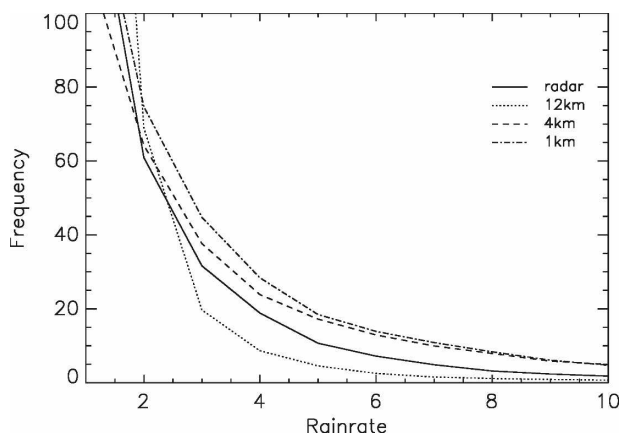
FIG. 8. Histograms of rainfall rates for all runs of 2003, 2004, and 2005. The 1 mm h^{-1} bins were used and the histograms were calculated only for times from $T + 6$ to $T + 7$.

Figure 8 shows a histogram of rainfall rates. The rainfall-rate data from the three models has been interpolated or aggregated onto the 5-km radar grid and cut down to the area of the 1-km model. To avoid contamination by the early period of the forecast where the rain rates are very low during the spinup of the 4- and 1-km forecasts, the histograms are only calculated for times between $T + 6$ and $T + 7$. The histogram shows that the 12-km model has too much light rain, but for rates over about 3 mm h^{-1} it then has too little. The lack of heavy rain is probably due to the convection scheme effectively averaging over the relatively large grid boxes. The 4- and 1-km models, in contrast, have too much heavy rain, which fits with the explicit convection still being under resolved in these models. The 1- and 4-km models have similar amounts of the highest rain rates, but the 1-km model has more of an excess problem at intermediate rain rates. The 4-km model has too little light rain, which is again a characteristic of trying to represent the convection explicitly with a too large grid length. In contrast the 1-km model seems to tend to approximately the correct value at the lower rates.

To understand further the properties of showers in the models, it is useful to look at the properties of the convective cells. Cell statistics were calculated by searching for contiguous areas with values above a threshold. Two sets of calculations were done—one where the data were first aggregated or interpolated onto the 5-km grid that the radar data was on and a second on the original model grids. The calculations were carried out over the 1-km model area at 15-min intervals for each run from the 2003, 2004, and 2005 set and averaged. Statistics were produced for the average number and area of the cells with the areas then being converted into radii assuming the cells were circular.
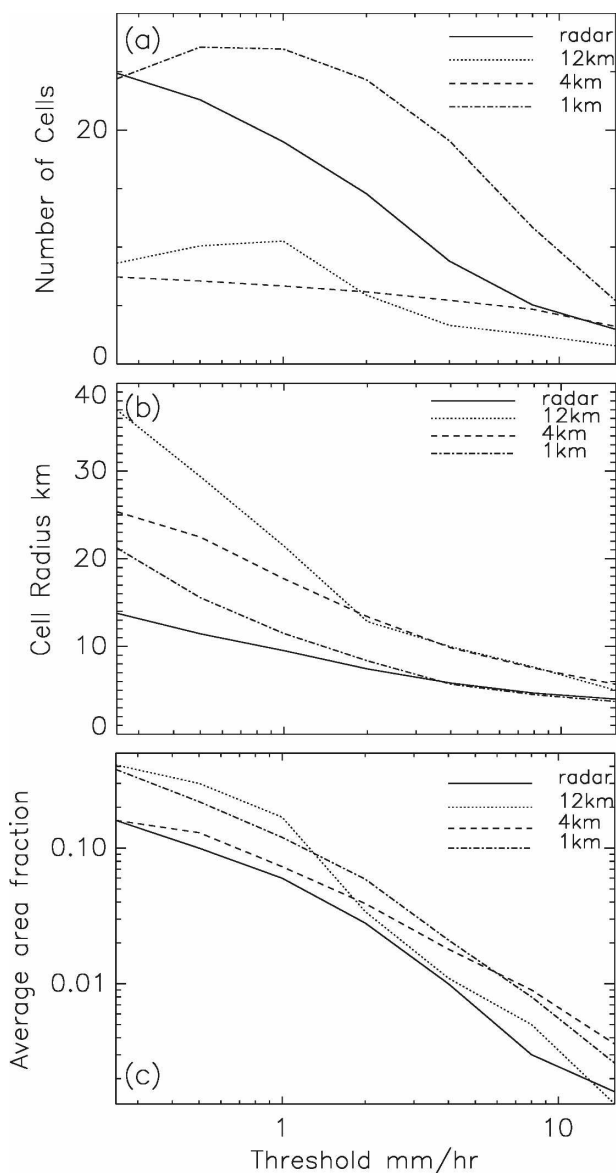

FIG. 9. Cell statistics as function of rainfall-rate threshold. For all models the statistics were calculated using data interpolated or aggregated onto the 5-km grid of the radar data.

The average fractional coverage was also calculated by multiplying the number and area together and dividing by the total domain area.

Figure 9 shows the results of the calculations on the 5-km grid for seven thresholds ranging from 0.25 to 16 mm h^{-1} each separated by a factor of 2. Generally both the radar data and all the models have the expected trends with threshold, namely that increasing the threshold reduces the number of cells found and also reduces the size of the cells. The 1-km model has too many cells for all thresholds in contrast to the 4- and 12-km models, which have too few, although the 4-km

model has about the correct number at the highest thresholds.

The number of cells *increases* with increasing threshold for the lowest thresholds in the 12- and 1-km models. This is due to large cells with several maxima splitting up as the threshold is increased and implies there is light rain over a relatively large area. This is seen in the 12-km model because of the convection scheme and in the 1-km model is likely to be an artifact of the aggregation to the 5-km grid where many smaller cells contribute to an overall low value. The 4-km model does not have this effect since its grid length is close to the grid on which the calculation was carried out.

For the cell radii (Fig. 9b) the 1-km model is closest to the radar over the whole range of thresholds although the values are too large for lower thresholds. The 12- and 4-km models both have larger radii cells. A striking feature is the large cells seen at lower thresholds in the 12-km model. This reflects the tendency of the model to produce widespread light rain as also shown in the histogram in Fig. 8. The 12- and 4-km models both have cell radii that reduce to around 6 km at the highest threshold (16 mm h^{-1}). In the 12-km model this corresponds to just one or two model grid points. This is possible because the convection is being represented by the convection scheme, which tends to turn on and off on individual grid points. The 4-km model produces convection explicitly and the cells are roughly the same absolute size as the 12-km model but now represent cells approximately 3–4 grid points across. This is about the smallest size cell that a model can be expected to represent. Although the cells are too large compared to reality, it is not desirable for models to produce features much smaller than this because of the likelihood of numerical inaccuracy and possible instability (gridpoint storms). Although detailed sensitivity work has not been carried out the value of the cell sizes obtained here is likely to depend on the amount of horizontal diffusion applied in the model (section 2d), which was chosen partly to reduce the gridscale structure in the rainfall field.

Looking at the average area fraction (Fig. 9c) the 4-km model is closest to being correct at low thresholds and the 12-km model is closest at higher ones. The 1-km model is always too high. These results mirror the overall rain amounts as shown in the histogram in Fig. 8.

Table 2 presents the difference between the calculation on the 5-km grid and on the model grids for the 4 mm h^{-1} threshold. The results for the 12- and 4-km models are similar; however, the 1-km model shows marked differences. There are over a factor of 2 more cells when calculated on the model grid, and the cell radius (which agreed reasonably well with the radar cell

TABLE 2. Comparison of cell statistics on 5-km grid and raw models grids for a threshold of 4 mm h^{-1}.

	On 5-km grid	On model grids
Average radar No. of cells	8.80	8.80
Average 12-km No. of cells	3.30	4.18
Average 4-km No. of cells.	5.45	6.15
Average 1-km No. of cells	19.09	46.21
Average radar cell radius (km)	5.82	5.82
Average 12-km cell radius (km)	9.96	9.19
Average 4-km cell radius (km)	9.84	9.42
Average 1-km cell radius (km)	5.69	3.81

radius on the 5-km grids) is reduced by a factor of about 1.5. Although we cannot compare the 1-km model grid results with 1-km radar data, the 5-km averaged results imply that there is a problem with too many cells in the 1-km model, and inspection of the model fields in many cases implies these cells are too small. This often appears to be worst during/immediately after the initiation of showers. An example is shown in Fig. 10 from 29 June 2005 during a phase of initiation of deep convection. Although the initiating line of rain can be picked out in the 1-km model, it clearly has too many cells when shown on either the 5-km grid (Fig. 10a) or the model grid (Fig. 10b). There is evidence that the problem of too many cells is caused by the choice of turbulence scheme in the 1-km model. Work is under way to investigate alternative turbulence schemes, though at present the standard 1D boundary layer scheme plus horizontal diffusion outperforms other approaches.

Figure 11 shows the dependence of the cell statistics above on forecast length. The most prominent feature in the 4- and 1-km data is the spinup effects at the start of the runs. Once again it is clear that the models take until about 2–3 h after analysis time to spin up (when starting from an analysis at $T + 1$). The properties appear mostly to be constant after $T + 3$. The 4-km model has the number of cells increasing throughout the time range shown—this may reflect the fact that this model often produces very late initiation of cells in more weakly forced situations.

d. Summary of model characteristics

The 12-km model has problems resulting from representing convection via a parameterization rather than explicitly. One of these is the model missing organization of convection, which often leads to underestimation of the peak rainfall rates. A second is convection often initiating too early. The 12-km model is still useful if the convective rain is interpreted as an indication that convection is likely to take place in an area larger than several grid squares rather than expecting the gridscale
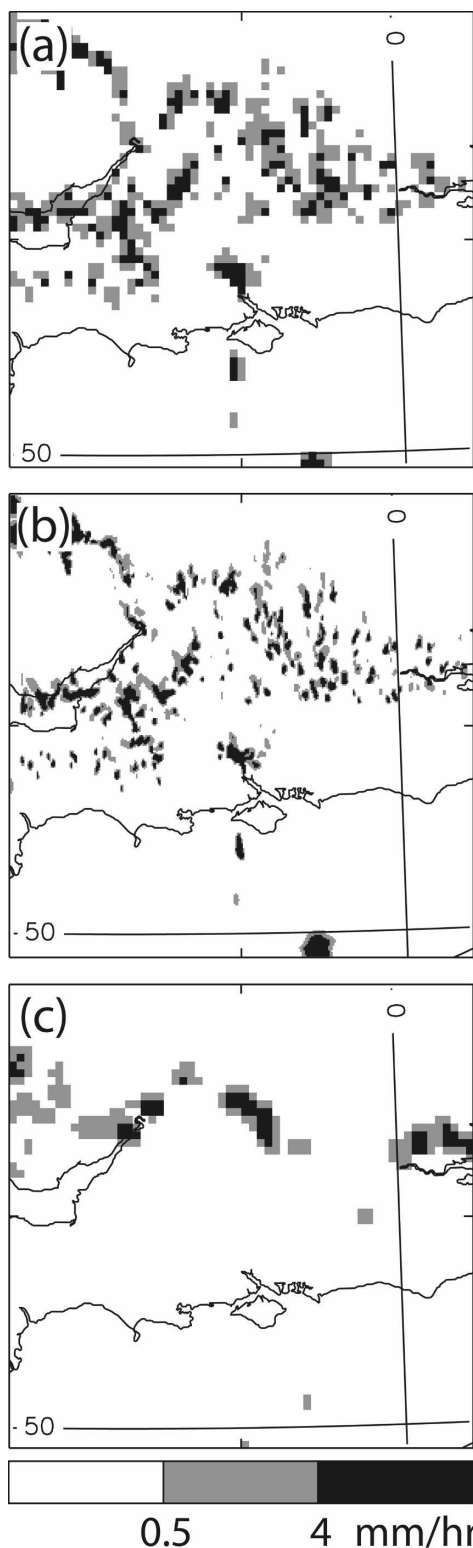
FIG. 10. Instantaneous rainfall rates for 1400 UTC 29 Jun 2005 from 0900 UTC model runs. (a) The 1-km model data interpolated onto 5-km grid, (b) 1-km model on its original grid, and (c) (5 km) radar data.
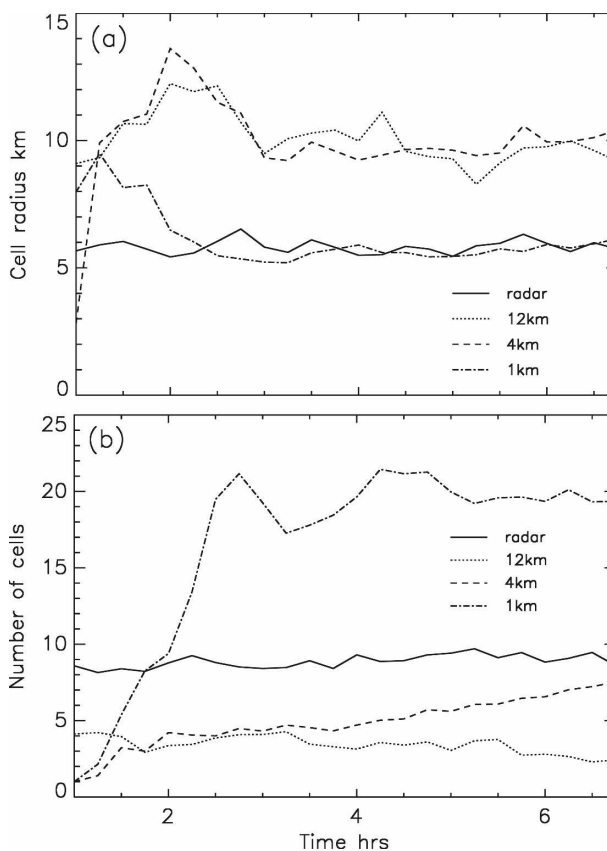

FIG. 11. (a) Cell size and (b) number of cells on model grids as a function of forecast length averaged over all 2003, 2004, and 2005 cases.

distribution of rainfall to be correct. In contrast, the 4- and 1-km models benefit from the convection being explicitly represented. The 4-km model has problems that appear to result from the grid length being too large to properly represent the explicit convection, namely a delay in initiation and too large, too heavy convective cells. The 1-km model has the advantages of not parameterizing the convection and also has a small enough grid length to avoid many of the problems seen in the 4-km model. As a result it often produces the qualitatively best-looking representation of the three models. The main problem that remains in the current implementation of the 1-km model is that the convective cells are sometimes too small and numerous, which may be the result of the choice of dissipation.

4. Skill score verification of precipitation

The precipitation fields from the summer 2003, 2004, and 2005 runs have been analyzed with a scale-dependent verification method. For details of this analysis technique the reader is referred to Roberts and

Lean (2008). The verification technique will now be briefly described.

The precipitation accumulation fields from the three models are interpolated or aggregated onto the same 5-km grid as the radar data and the verification method carried out on that grid. Fractions are generated using the neighborhood approach (Theis et al. 2005). For every 5-km pixel, we compute the fraction of surrounding pixels within a given sized square "neighborhood" that exceed a particular accumulation threshold (e.g., 8 mm in a 3-h period). This is done to both the radar and forecast fields. As a result, every pixel in the forecast field has a fraction that can be compared to its equivalent pixel in the radar field. The fractions are generated for different spatial scales by changing the size of the neighborhood squares. As the neighborhoods become larger, the forecast and radar fractions will become more alike because the spatial errors in the forecast (e.g., misplaced rainbands) will have less significance.

The fractions are compared by using the fractions skill score (FSS). This is given by

$$FSS = 1 - \frac{FBS}{FBS_0}, \quad (6)$$

where FBS is a version of the Brier skill score that compares fractions with fractions and is given by

$$FBS = \frac{1}{N} \sum_{j=1}^{N} (o_i - p_i)^2, \quad (7)$$

where o_j and p_j are the radar and forecast fractions, respectively, at each point j.

FBS_0 is the worst possible value of FBS in which there is no collocation of nonzero fractions and is given by

$$FBS_0 = \frac{1}{N} \left(\sum_{j=1}^{N} p_j^2 + \sum_{j=1}^{N} o_j^2 \right). \quad (8)$$

The FSS can have a value that varies between 1.0 for a perfect forecast and 0.0 for a forecast with no skill. The FSS typically increases with neighborhood size (it always increases with neighborhood size for a large enough sample). An unbiased forecast will tend toward a FSS of 1.0 as the sampling square approaches the size of the verification domain.

Percentile thresholds (e.g., accumulations exceeding the 90th percentile value, that is, the top 10%) are used in addition to absolute thresholds (e.g., 4 mm h^{-1}). By definition, these make the forecast and observed frequencies the same in the sample, which means that we can focus more on the spatial accuracy (the bias should always be borne in mind though).

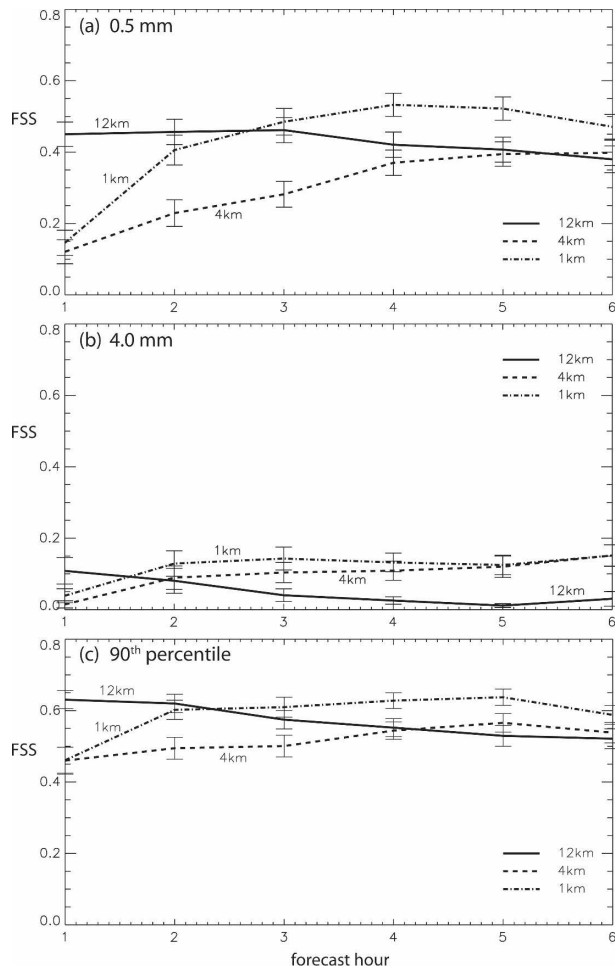Figure 12 shows the fraction skill scores for hourly

FIG. 12. Fraction skill scores for hourly accumulations aggregated over all cases from summer 2003, 2004, and 2005. The hourly fraction skill scores are shown as a function of forecast time for a fixed sampling radius of 75 km (each point is the score for the accumulation from that time to an hour later). The threshold was (a) 0.5 mm h^{-1}, (b) 4.0 mm h^{-1}, and (c) 90th percentile. The error bars are calculated by a bootstrapping technique as described in the text.

accumulations. The scores from every forecast with the same forecast length have been aggregated together and the figure shows the scores as a function of forecast length from $T + 1$ to $T + 6$. To plot these curves a fixed horizontal scale (neighborhood length) must be chosen and a value of 75 km has been used. For short scales there is no skill (because of the short-scale errors) and for long scales there is no spatial information so an intermediate value is required. The radius of 75 km is chosen because, as shown in Roberts and Lean (2008), 6-h accumulation forecasts have useful skill at that scale in all models. Absolute thresholds of 0.5 and 4.0 mm h^{-1} are shown (Figs. 12a,b) and also a relative one of the 90th percentile (12c). (The 90th percentile thresh-
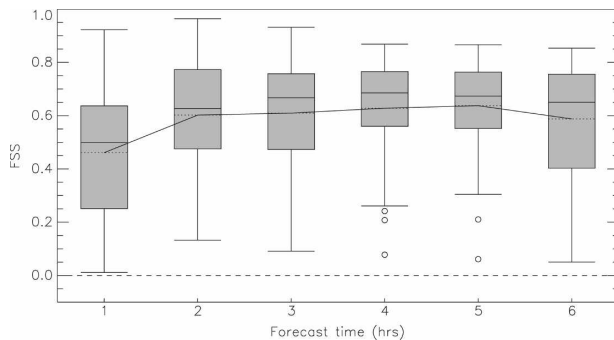
FIG. 13. Box plots for 1-km model hourly scores for a 75-km horizontal scale showing spread over all forecasts of 2003, 2004, and 2005 cases for the 90th percentile threshold. The solid lines indicate the mean values of the distributions; boxes stretch from the 25th to 75th percentiles, that is, the inner-quartile range (IQR). Extreme outliers (i.e., those located more than 3 times the IQR above or below the box) are represented by black circles; mild outliers (located more than 1.5 times but less than 3 times the IQR) are represented by open circles. Smallest nonoutlier data values are represented by whiskers; the solid horizontal bar within in each box indicates the median.

old typically corresponds to an absolute threshold of around 1 mm h^{-1}.) For thresholds much higher than 4.0 mm h^{-1} the curves become noisy because of the small number of points being sampled. For all three thresholds the 12-km model shows a gradual reduction of skill with forecast length. In contrast, the 4- and 1-km models have low values of skill scores at the start of the forecasts and improve with forecast time because of the spinup discussed in previous sections. The absolute value of the scores is very low for the 4 mm h^{-1} threshold. This is partly due to the bias (the overprediction of rain). In addition, if the score of a random distribution of rain with the same bias is calculated it is typically found to be around 0.2 for the 0.5 mm h^{-1} threshold but 0.01 for 4 mm h^{-1}. Although the absolute scores are lower for the 4 mm h^{-1} threshold they still represent an improvement over the score from a random forecast.

These curves hide a great deal of variability between the 64 forecasts that make up the sample. Figure 13 shows box plots of the distribution of the scores for the 1-km model only for the 90th percentile threshold. There is a great deal of spread with 50% of the forecasts giving scores within a range larger than 0.3. This raises the question of whether the differences between the models seen in Fig. 12 are significant. To address this, error bars are shown in Fig. 12, which show the estimate of the standard error of the mean from a bootstrapping technique (Wilks 1995). The bootstrapped estimations of the standard error were calculated using 30 000 bootstrap members in each case. The resulting member distributions confirmed that the standard error

in each case was well approximated by a normal distribution. Because of the large spread of scores it is important to realize that even where there is no overlap between the error bars (implying a robust statistically significant difference) one would still expect a significant number of individual occasions where the relative scores are reversed.

Taking into account the error bars in Fig. 12, and looking at the later parts of the forecasts after the spinup, the 1-km model does better than the 4- and 12-km models in a statistically significant sense with the 0.5 mm h^{-1} and 90th percentile thresholds. For the 4 mm h^{-1} threshold the 1- and 4-km models do roughly as well as each other and both do better than the 12-km model, which often struggles to produce heavier rain. These statistics have also been calculated for other accumulation periods (3 and 6 h), scale lengths, and thresholds, but these results are not shown since they do not add anything to the discussion. The absolute values of the scores change but the trends with time and the relative position of the three models are still consistent with those that have already been shown.

5. Conclusions

We have described an experimental configuration of the UM at 4- and 1-km resolutions. These models have been run for a number of convective cases from the summer 2005 CSIP project, and also from summers of 2004 and 2003, in a suite of models that also included the 12-km model (for comparison and in order to provide boundary conditions). The configurations of the 4- and 1-km models are generally very similar to that of the operational 12-km model; the major change being that the 1-km model is run without a convection parameterization and the 4-km model has the parameterization modified to greatly reduce the convective mass flux. We have presented results only from 4- and 1-km models without data assimilation; that is, each forecast was started from a 12-km analysis with no carryover of high-resolution information from one cycle to the next. Although this leads to reduced performance during the "spinup" period as the high-resolution model develops structure from the low-resolution analysis, it avoids issues due to data assimilation and allows the general character of the models to be evaluated. This evaluation has included both subjective analysis of the model precipitation fields and objective statistical data.

It is noticeable from subjective examination that the 4- and 1-km models with explicit convection tend to initiate convection later than the 12-km model. Analysis of systematic statistics from a subset of the cases run has revealed that the 12-km model tends to initiate precipitation too early by 1–2 h and the 4- and 1-km mod-

els are closer to reality. The 4-km model tends to initiate somewhat too late and the 1-km model is close to being correct.

Once convection has initiated it has, subjectively, very different characteristics in 4- and 1-km models compared to the 12-km model. These differences arise from the explicit representation of convection, which gives more realistic-looking features in many cases. There are, however, obvious problems in the 4- and 1-km models, including too large peak rainfall rates in the showers, too few but too large cells (4 km), and too many small cells in the 1-km model. These observations are confirmed by statistics aggregated from all the cases including rain-rate histograms and cell statistics. The problem with large cells in the 4-km model stems from the convection being seriously under resolved at this grid length. It is thought that the issue with too many cells in the 1-km model results from the choice of turbulence scheme.

By examining the domain-averaged precipitation rates against time for the 4- and 1-km models it is found that there is a deficit of rain at the start of the runs followed by an overshoot peaking at around $T + 6$. These features are a result of starting the model from lower-resolution analyses from a model running with a convection scheme. This "spinup" time is shorter in the 1-km model than the 4-km model. It is clear that a real forecasting system for times shorter than 9–12 h will need to have some method of propagating high-resolution explicit convection information from one forecast to the next (for example, assimilation).

We have used a scale-selective verification technique to verify precipitation accumulations against radar data. The results show that, although there is a great deal of variation from run to run in the scores, the 1-km model gives statistically significant improvements to the scores from the 12-km model after the initial spinup period when hourly accumulations are considered on scales of greater than 75 km. By the same measure the 4-km model does significantly worse than the 1-km model for lower thresholds, partly because it appears to take longer to spin up at the start of the run. However for higher thresholds it does nearly as well as the 1-km model.

The basis for carrying out this work was to determine whether running the UM with 4- and 1-km grid lengths would provide improvements to the precipitation forecasts over the 12-km model. Similarly we wanted to determine whether the 1-km model gives further benefits over the 4-km model. Despite the spinup problems observed in the current work, we have evidence both subjective and statistical that the 4- and 1-km-resolution models do provide benefits over the 12-km model. Although there are problems in some situations with too small convective cells, the 1-km model generally performs better than the 4-km model with regard to convective initiation and the general scales evident in the precipitation fields. There is therefore every reason to expect that future work will realize this potential and order 1-km-gridlength models will, in time, become an important part of the Met Office operational forecast system.

APPENDIX A
Summary of Configuration Differences

TABLE A1. Summarizing differences in configuration between 4- and 1-km models and the operational 12-km model.

	12 km	4 km	1 km
Horizontal domain			
Approx grid length (km)	12	4	1
Grid length (°)	0.11	0.036	0.009
Lat bottom left corner (BLC) (°)	−8.41	−5.45	−2.630
Lon BLC (°)	352.95	356.6	359.3
Grid size	146 × 182	190 × 190	300 × 300
Pole lat (°)	37.5	37.5	37.5
Pole lon (°)	177.5	177.5	177.5
Vertical levels			
No. of levels	38	38	76
Top of model (m)	Approx 40 000	Approx 40 000	Approx 40 000
No. of boundary layer levels	13	13	26
No. of ozone levels	11	24	48
Boundary conditions			
Driving model	Global (60 km)	12 km	4 km
Rim width	8	8	8
Time frequency	60 min	30 min	15 min
Aerosol boundary values from	UK mesoscale model (UKmes) boundary model	12 km	4 km
Timings			
Time step	5 min	100 s	30 s
Radiation time step	60 min	15 min	5 min
Parameterizations			
Convection scheme	Mass flux with CAPE closure time scale 1800s	Mass flux CAPE dependent CAPE closure (see text)	No convection scheme
Microphysics	Dual phase including iterative melting	Dual phase with prognostic rain	Dual phase with prognostic rain
Gravity wave drag	On	Off	Off
Boundary layer	13 levels	13 levels	26 levels
Other			
Horizontal diffusion	None	∇^4, 8 time steps (see text)	∇^4, 8 time steps

APPENDIX B
Summary of Cases

TABLE B1. Summary of cases investigated from summer 2003 and 2004.

Date	CSIP IOP	Model runs (UTC)	Description
13 May 2003	—	0600, 0900, 1200, 1500	Line of thunderstorms develops around 1500 UTC
25 May 2003	—	0600, 0900, 1200, 1500	Scattered convection
1 Jul 2003	—	0600, 0900, 1200, 1500	Line of convection
28 Aug 2003	—	0600, 0900, 1200, 1500	Bands of convective rain
27 Apr 2004	—	1200, 1500, 1800	Heavy storms initiating over London at about 1530 UTC and subsequently moving west
8 Jul 2004	—	0300, 0600, 0900, 1200	Bands of rain around a cyclone over the English Channel
10 Jul 2004	—	0300, 0600, 0900, 1200	Gust fronts initiating showers downstream from initial development over south Wales at 0600 UTC
20 Jul 2004	—	0600, 0900, 1200, 1500	Showers initiated at around 1300 UTC in southerly flow
3 Aug 2004	—	0600, 0900, 1200, 1500	Very heavy rain—High Wycombe floods
20 Aug 2004	—	0300, 0600, 0900, 1200	Bands of heavy showers moving east
15 Jun 2005	1	0300, 0600, 0900, 1200	Single thunderstorm along southwest convergence line
29 Jun 2005	5	0600, 0900, 1200, 1500	Initiation and development of organized east–west line in southwest flow
4 Jul 2005	6	0600, 0900, 1200, 1500	Stream of cells in northwest flow with triggering and organization
18 Jul 2005	9	0600, 0900, 1200, 1500	Narrow line of convective showers
28 Jul 2005	12	0300, 0600, 0900, 1200	Initiation by low-level convergence lines
25 Aug 2005	18	0600, 0900, 1200, 1500	Cold pool and secondary convective initiation

REFERENCES

Browning, K. A., and Coauthors, 2007: The Convective Storm Initiation Project. *Bull. Amer. Meteor. Soc.,* **88,** 1939–1955.

Bryan, G. H., and R. Rotunno, 2005: Statistical convergence in simulated moist absolutely unstable layers. Preprints, *11th Conf. on Mesoscale Processes,* Albuquerque, NM, Amer. Meteor. Soc., 1M.6.

Cullen, M. J. P., T. Davies, M. H. Mawson, J. A. James, S. C. Coulter, and A. Malcolm, 1997: An overview of numerical methods for the next generation UK NWP and climate model. *Numerical Methods in Atmospheric and Ocean Modelling: The André J. Robert Memorial Volume,* C. A. Lin, R. Laprise, and H. Ritchie, Eds., Canadian Meteorological and Oceanographic Society, 425–444.

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. *Quart. J. Roy. Meteor. Soc.,* **131,** 1759–1782.

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. *Atmos. Sci. Lett.,* **5,** 110–117.

Essery, R., M. Best, and P. Cox, 2001: MOSES 2.2 Tech. Doc. Hadley Centre Tech. Rep. 30, Met Office Hadley Centre, 30 pp.

Gregory, D., and P. R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.,* **118,** 1483–1506.

Hohenegger, C., and C. Schar, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.,* **88,** 1783–1793.

Jones, C. D., and B. Macpherson, 1997: A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteor. Appl.,* **4,** 269–277.

Kain, J. S., and Coauthors, 2007: Some practical considerations for the first generation of operational convection-allowing NWP: How much resolution is enough? Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction,* Park City, UT, Amer. Meteor. Soc., 3B.5.

Lean, H. W., and P. A. Clark, 2003: The effects of changing resolution on the mesoscale modelling of line convection and slantwise circulations in FASTEX IOP16. *Quart. J. Roy. Meteor. Soc.,* **129,** 2255–2278.

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. *Mon. Wea. Rev.,* **128,** 3187–3199.

Lorenc, A., and Coauthors, 2000: The Met Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **126,** 2991–3012.

Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.,* **26,** 636–646.

Narita, M., and S. Ohmori, 2007: Improving precipitation forecasts by the operational nonhydrostatic mesoscale model with the Kain–Fritsch convective parameterization and cloud microphysics. Preprints, *12th Conf. on Mesoscale Processes,* Waterville Valley, NH, Amer. Meteor. Soc., 3.7.

Petch, J. C., 2006: Sensitivity studies of developing convection in a cloud-resolving model. *Quart. J. Roy. Meteor. Soc.,* **132,** 345–358.

Roberts, N. M., 2003: The impact of a change to the use of the convection scheme to high-resolution simulations of convective events. Met Office Tech. Rep. 407, 30 pp.

——, and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high resolution forecasts of convective events. *Mon. Wea. Rev.,* **136,** 78–97.

Romero, R., C. A. Doswell, and R. Riosalido, 2001: Observations and fine-grid simulations of a convective outbreak in northeastern Spain: Importance of diurnal forcing and convective cold pools. *Mon. Wea. Rev.,* **129,** 2157–2182.

Speer, M. S., and L. M. Leslie, 2002: The prediction of two cases of severe convection: implications for forecast guidance. *Meteor. Atmos. Phys.,* **80,** 1–4.

Steppeler, J., G. Doms, U. Schattler, H. W. Bitzer, A. Gassmann, and U. Damrath, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.,* **82,** 75–96.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.,* **12,** 257–268.

Warner, T. T., and H.-M. Hsu, 2000: Nested-model simulation of moist convection: The Impact of coarse-grid parameterized convection on fine-grid resolved convection. *Mon. Wea. Rev.,* **128,** 2211–2231.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.,* **125,** 527–548.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.

Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.,* **125,** 1607–1636.