Quarterly Journal of the
Royal Meteorological Society

Ensemble forecasts of a flood producing storm: comparison of the influence of model-state perturbations and parameter modifications

Journal:	*QJRMS*
Manuscript ID:	QJ-11-0197.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Leoncini, Giovanni; University of Reading, Meteorology Plant, Robert; University of Reading, Meteorology Gray, Suzanne; University of Reading, Meteorology Clark, Peter; Met Office,
Keywords:	convective-scale ensemble forecasting, parameter perturbations, model-state perturbations, flood, Unified Model

Ensemble forecasts of a flood producing storm: comparison of the influence of model-state perturbations and parameter modifications

G. Leoncini *[a], R. S. Plant [a], S. L. Gray [a], P. A. Clark [b†]

[a]*Department of Meteorology, University of Reading, Reading, UK*

[b]*MetOffice@Reading, MetOffice, Reading, UK*

*Correspondence to: G. Leoncini, MetOffice@Reading, Meteorology Building, University of Reading, PO Box 243, Earley Gate, Reading, Berkshire, RG6 6BB. Email: giovanni.leoncini@metoffice.gov.uk

High-resolution ensemble simulations ($\Delta x = 1$km) are performed with the Met Office Unified Model for the Boscastle (Cornwall, UK) flash-flooding event of 16 August 2004. Forecast uncertainties arising from imperfections in the forecast model are analysed by comparing the simulation results produced by two types of perturbation strategy. Motivated by the meteorology of the event, one type of perturbation alters relevant physics choices or parameter settings in the model's parameterization schemes. The other type of perturbation is designed to account for representativity error in the boundary-layer parameterization. It makes direct changes to the model state and provides a lower bound against which to judge the spread produced by other uncertainties. The Boscastle has genuine skill at scales of approximately 60km and an ensemble spread which can be estimated to within $\sim 10\%$ with only eight members. Differences between the model-state perturbation and physics modification strategies are discussed, the former being more important for triggering and the latter for subsequent cell development, including the average internal structure of convective cells. Despite such differences, the spread in rainfall evaluated at skillful scales is shown to be only weakly sensitive to the perturbation strategy. This suggests that relatively simple strategies for treating model uncertainty may be sufficient for practical, convective-scale ensemble forecasting.

Key Words: convective-scale ensemble forecasting, parameter perturbations, model-state perturbations, flood, Unified Model

Received . . .

Citation: . . .

1. Introduction

Recently, there has been much interest in the prospects for numerical weather prediction at convection-permitting resolutions ($\Delta x \lesssim 4$km). The use of high resolution is necessary for the explicit simulation of weather phenomena that are directly produced by small-scale dynamics or else are sensitive to small-scale features of the surface. Both of these aspects may apply to severe convective storms. In the United Kingdom, for instance, although convection can often be initiated by a complex combination of factors

[†]Now at Department of Mathematics and Department of Civil, Chemical and Environmental Engineering, University of Surrey, UK.

operating over a range of scales (Bennett *et al.* 2006), there are nonetheless good reasons to believe that a convective-scale model may be effective in capturing the initiation (e.g. Lean *et al.* 2009) and subsequent development with genuinely useful skill (Lean *et al.* 2008). A particularly good example is the flash-flooding event at Boscastle on 16 August 2004 (Burt 2005; Golding *et al.* 2005), for which the extreme convective precipitation could be captured by convection-permitting re-forecasts of the event but not by the 12km grid-length forecast model that was available at the time (Lean *et al.* 2005).

Results to date for convection-permitting forecasting have been extremely encouraging, both at the Met Office and elsewhere (e.g. Lean *et al.* 2008, and references therein). However, the appropriate use and interpretation of convective-scale forecasts is a difficult issue, not least because there is currently only a limited understanding of predictability at such scales. The baroclinic wave simulations by Zhang *et al.* (2003, 2007), provide a good example of the different (faster) character of error growth with explicitly-modelled as opposed to parameterized deep convection. Stronger non-linearities and faster error growth at convective scales imply that ensemble strategies may be particularly useful at such scales, as argued by various authors (e.g Hohenegger *et al.* 2008, and references therein). Indeed, despite the computational expense, experimentation with convective-scale ensembles is already well under way (e.g. Marsigli *et al.* 2004; Hohenegger *et al.* 2008; Leoncini *et al.* 2010) even while the operational use of a single forecast at these resolutions remains in its infancy. At the Met Office, for example, a convective-scale model has been run routinely since 2009 and a test system for ensembles at this scale is currently being built with the intention of trialling it during the spring of 2012. In the United States the Spring Forecasting Experiments have been running convective-scale models for a number years and more recently a convective-scale ensemble (Clark and Coauthors 2011).

There is a pressing need for appropriate ensemble strategies designed for the convective scale to be explored. This article is intended as one contribution (necessary but by no means sufficient) towards that end, and investigates aspects of the predictability of the flood that devastated Boscastle (Cornwall, UK) in August 2004. In general, forecast uncertainties arise from the boundary conditions, initial conditions and uncertainties in the physical description provided by the forecast model, commonly termed "model error". All three sources of uncertainty are likely to be relevant for the convective-scale forecasting of severe events. Even at mesoscale resolution, Fujita *et al.* (2007) gave a good example of how both initial-condition and model physics uncertainties are important in producing an ensemble. The spreads produced in their experiments had distinct spatial and temporal structures affecting different variables in different ways: for example, spread arising from initial-condition uncertainty was larger for dynamical variables, while spread arising from physics uncertainty was larger for thermodynamical variables.

Recent studies have investigated various sources of uncertainty at the convective scale (e.g. Marsigli *et al.* 2004; Hohenegger *et al.* 2008; Leoncini *et al.* 2010; van Weverberg *et al.* 2010; Yussouf and Stensrud 2011). The scope of the present study is to compare the uncertainties arising from two classes of forecast model uncertainty. Motivated by the meteorology of the specific event, we will consider uncertainties in what would appear to be important aspects of the parameter settings used by the model (for instance, in the roughness lengths and autoconversion rates), here referred to as parameter modifications. Similar experiments have also been conducted by other researchers for other cases (e.g. van Weverberg *et al.* 2010). We will also consider representativity error in the parameterization of the boundary layer. The methodology has been documented in a previous study of scattered convection in the UK (Leoncini *et al.* 2010) and an overview is provided in

Section 4.2. Small perturbations are applied to boundary-layer potential temperature. These are designed to approximately represent the power in the boundary-layer temperature spectrum at scales larger than the smallest represented well by the model. Such perturbations, applied throughout the forecast evolution, are referred to as model-state perturbations. The study is focussed on contrasting different sources of model error. This should not be taken to imply that initial and boundary condition uncertainties are considered to be unimportant for the event, only that they are outside the scope of the study.

As argued by Leoncini *et al.* (2010), the introduction of boundary-layer state perturbations in a convective-scale forecast addresses the issue that an Reynolds-average-based parametrization of the boundary layer becomes inappropriate if we intend to represent space and/or time averages at fine scales. The Reynolds average is based upon an ensemble of realisations of the turbulent boundary layer and only tends to the space/time average for large space/time scales in steady state. Rather, a space-time filter applied to the underlying equations of motion will leave a residual unsteady behaviour which we represent through a stochastic forcing. The standard Reynolds-average-based parametrization removes variability which can grow, through mechanisms not considered by the convective boundary-layer parametrization, to trigger deep convection which is reasonably well resolved by the model.

The method has something in common with the stochastic kinetic energy backscatter (SKEB) of Shutts (2005), but the latter is aimed at returning some of the energy dissipated by the dynamics rather than at the spatial variability of the area-averaged boundary layer parametrization. SKEB itself derives from the backscatter scheme developed for large eddy simulations (LES) of the boundary-layer, and which first showed success in treating the 'grey-zone' between the fully parametrized surface-layer and the partially resolved eddies away from the surface layer (Mason and Thomson 1992). Our scheme is rather simpler, and focusses on temperature fluctuations which are absent from the LES scheme but have been shown to have substantial impact on development of deep convection (Leoncini *et al.* 2010).

This methodology has two implications. The first is that a control simulation without such forcing may be biased (e.g. through systematically different convection triggering or precipitation initiation). For example, in the extreme case of horizontally homogeneous forcing, a Reynolds-average-based scheme would produce a horizontally uniform state which could never trigger explicit deep convection. The second implication is that any model run must be considered as being sampled from a random population. It is therefore not safe to compare individual model runs with and without a given physics change even with the same stochastic forcing, as the physics change may interact differently in different realisations of the flow; rather an ensemble approach is necessary to compare statistics of the populations (both with stochastic forcing) with and without the physics change.

Just such an approach is followed here. The model response to boundary-layer state perturbations is important in part because this uncertainty may be an important source of forecast spread in its own right (Leoncini *et al.* 2010). However, it also provides a reference against which to judge the spread produced by other uncertainties. In particular, we would conclude that a given uncertainty in physical parametrization is unimportant for the forecast if the resulting differentiation of simulations is substantially less than that caused by random variations in the boundary-layer behaviour. Thus, the comparison of methods will allow us to judge whether or not genuinely meaningful changes to the forecasts are associated with model physics modifications. If the natural variability of the system were ignored by neglecting the stochastic nature of the boundary layer, we may be misled into believing that a physics modification has significant systematic impact on a forecast when in fact it is merely providing a mechanism to enable the model to follow an alternative, but essentially random, trajectory. As an example, a change to cloud microphysics in one

realisation may change the precise location of secondary initiation promoted by the downdraught from a cell; the knock-on effect may result in a very different distribution of cells in subsequent generations and even affect their mesoscale organisation, but the same microphysical change in another equally-valid realisation of the flow may equally well produce no such effects.

The verification of convective-scale forecasts presents some important issues, particularly for quantities such as convective precipitation that exhibit very high variability in both space and time. Simple measures such as the root-mean-square error tend to excessively penalise small displacement errors, which has prompted the development of more specialist techniques to assess the displacements occurring or the fidelity of small-scale features. Similar issues arise when comparing convective-scale forecasts of the same event. It is not sufficient simply to find differences between two simulations, but rather one must establish whether the simulations are meaningfully different based on measures for which the model has been shown to have some skill. Here we adopt and adapt two such specialist techniques: the fractions skill score (FSS) of Roberts and Lean (2008) and the structure, amplitude, and location (SAL) scores of Wernli *et al.* (2008). The FSS is adapted to inform the choice of a suitable spread metric to assess the simulated precipitation, ensuring an evaluation at scales for which the model is skillful in this case. The SAL scores allow us to assess how the precipitation is produced: i.e., whether the storms in one simulation have a systematically different character to those in another.

In summary this work explores the predictability of a convective event utilising two approaches to convective scale ensemble: physics parameter modification and model-state perturbation. Both approaches address aspects of model uncertainty and the former approach is in common use (e.g Gebhardt *et al.* 2011; Clark *et al.* 2009; Yussouf and Stensrud 2011). The main goal of this study is to compare the two approaches. Initial and later boundary conditions are also important but for simplicity their influence on the predictability of this event is not addressed here. While it is demonstrated that the control simulation describes the event reasonably well, a comprehensive verification would certainly provide useful insight into the problem, but it would also require robust statistics, well beyond the one event studied here. However, because model skill varies with scale radar observations are used to make informed decision (see Section 4.1). The article is organised as follows: Section 2 describes the model used and its configuration, Section 3 provides an overview of the observed meteorology of the event and of the control simulation. This provides context and motivates the characterisation of the target area and the ensemble design, as discussed in Section 4. Results are described and discussed in Section 5, while conclusions are presented in Section 6.

2. Model description and configuration

The model ensemble hindcasts were performed using version 6.1 of the Met Office Unified Model (MetUM). The MetUM is an operational finite-difference numerical weather prediction model that solves the non-hydrostatic deep-atmosphere dynamical equations with a semi-implicit, semi-Lagrangian integration scheme (Davies *et al.* 2005). The model uses Arakawa C staggering in the horizontal. The vertical coordinate system is terrain following with a hybrid-height vertical coordinate and Charney-Phillips staggering. The model can be configured either as a global model or as a limited area model with one-way nesting. In the limited area model configuration the horizontal grid is rotated in latitude/longitude.

Parameterization of physical processes includes long- and short-wave radiation (Edwards and Slingo 1996), boundary-layer mixing (Lock *et al.* 2000; Lock 2001), cloud microphysics and large-scale precipitation (Wilson and Forbes 2004) and (if used) convection (Gregory and Rowntree 1990). The large-scale precipitation scheme used in this study is an enhancement of Wilson and Ballard (1999): it remains a single-moment

bulk parameterization but allows for more prognostic species. There are four prognostic species considered: water vapour, liquid water droplets, raindrops and a single species of ice. The Met Office Surface Exchange Scheme (MOSES) is used to model the exchanges of heat, moisture and momentum between the surface and the atmosphere. The version used here is MOSES 2 (Essery *et al.* 2003) in which heterogeneous surfaces may be treated using a tiled representation that allows different surface types to coexist in the same model grid box. Separate surface temperatures, short-wave and long-wave radiative fluxes, sensible and latent heat fluxes, ground heat fluxes, canopy moisture contents, snow masses and snow melt rates are computed for each surface type in a grid box. Nine surface types are defined: broadleaf trees, needleleaf trees, C3 (temperate) grass, C4 (tropical) grass, shrubs, urban, inland water, bare soil and ice. Each type has an associated roughness length and other surface parameters. Air temperature, humidity, and windspeed on model levels above the surface and the temperature and moisture content of each subsurface soil layer are treated as homogeneous across a gridbox.

Very high resolution ensemble simulations were performed using a horizontal grid length of 0.009° (~ 1 km) and 76 vertical levels, arranged with a lid at ~ 39 km and spacings of 200–370 m in the mid-troposphere. The limited area domain used has 300×300 gridpoints and covers the south-west peninsula of England and south Wales. Fig. 1 shows this domain with 25 grid points cropped around the edges to avoid showing any spin-up effects associated with the forced lateral boundaries. The figure also shows a region in the south-east of the domain which is cropped in some of the later analysis (Section 5.4). A large-scale storm crosses this part of the domain in the late morning, but is not the focus of the present study.

Initial and lateral boundary conditions were obtained from a simulation performed using a horizontal grid length of 0.036° (~ 4 km) and 38 vertical levels (comprising every other level of the 76 used at higher resolution). This domain has 190×190 gridpoints, and covers southern and central England, Wales, south-east Ireland and northern France. In turn, the initial and lateral boundary conditions for the 4 km grid-length simulation came from a simulation performed using a horizontal grid length of 0.11° (~ 12 km) and the 38 model level set. This domain has 146×182 gridpoints and covers all of the United Kingdom extending to parts of Scandinavia and north-west Europe; this domain was the operational limited area (mesoscale) model domain for the United Kingdom at the time of the Boscastle storm. Initial and lateral boundary conditions for this simulation were taken, respectively from the operational MetUM mesoscale model analysis and the global forecast.

The configurations used for the MetUM at each resolution match the current operational configurations for 12 km, 4 km and 1.5 km grid spacing as closely as possible. The 8A version of MOSES 2 is used for the configuration with 12 km grid spacing; and the 8B version of MOSES 2 for the configurations with 4 km and 1 km grid spacing. The convection parameterization scheme is turned on for the simulation with 12 km grid spacing, It is also turned on for the simulation with 4 km grid spacing but is subject to the modification of Roberts (2003). Convective parameterization is turned off for the simulation with 1 km grid spacing. The Roberts (2003) method avoids the accumulation of high CAPE at the gridscale which can lead to unphysical "gridpoint storms"; it was specifically designed for the 4 km grid-length configuration of the MetUM and has proved reasonably successful (Lean *et al.* 2005; Roberts and Lean 2008). Ancillary files containing orography, land-sea mask, ozone field, and vegetation distribution (area and structure) were created specifically for the 4 km and 1 km grid-length domains.

All of the simulations presented were initialised at 0100 UTC 16 August 2004, using the same initial and lateral boundary conditions. Results will be shown from this time until 1900 UTC on the same day.

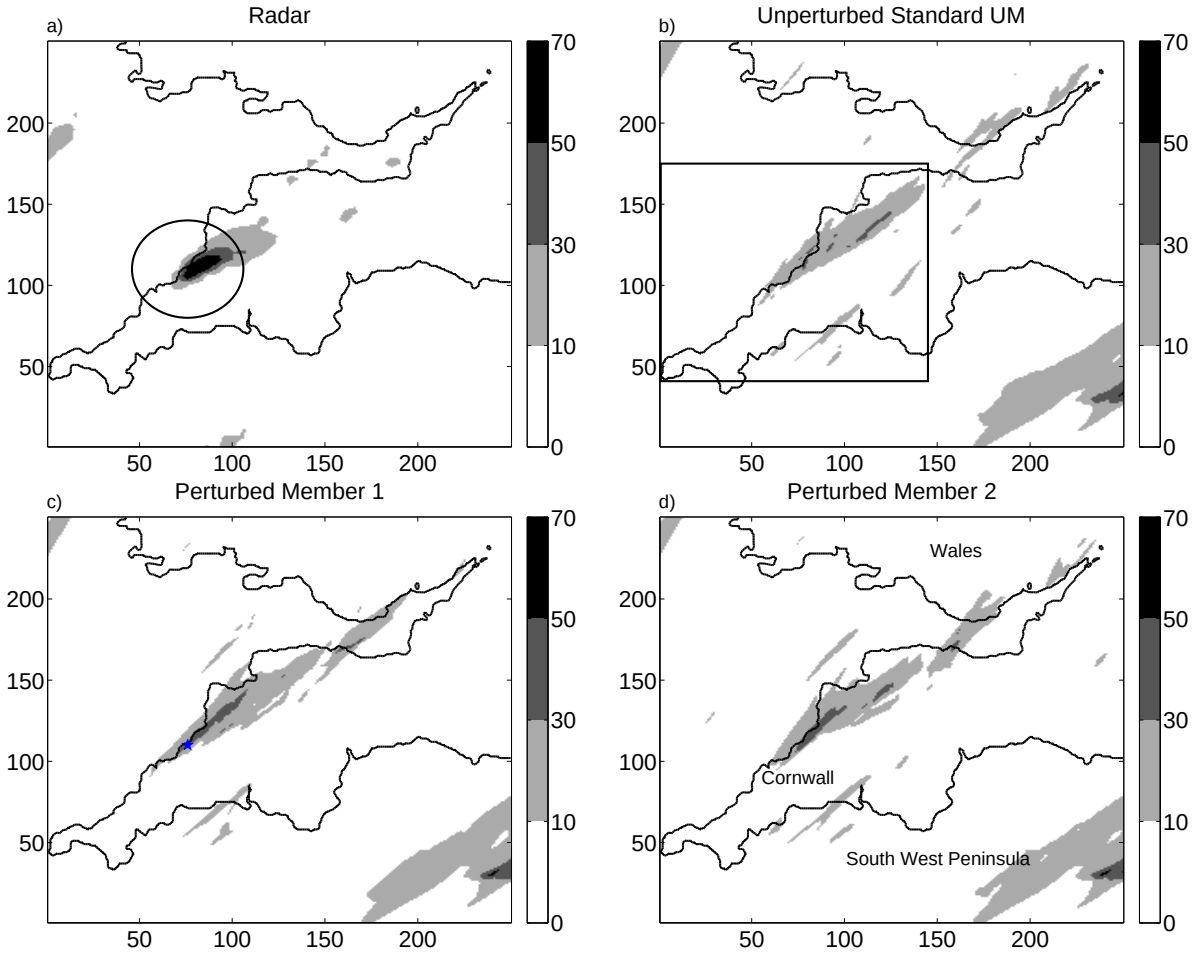


Figure 1. Rainfall accumulations in mm from 1200 to 1700 UTC 16 August 2004 for: a) radar observations (5 km grid length) with the target area circled (see Sec 4.1), b) the control simulation with the area used to compute the cloud profiles marked (as described in Sec 5.4), c) and d) two members of the model-state perturbed control ensemble. The star in c) marks the location of the town of Boscastle. The axis labels indicated the distance in km, from the South West corner.

3. Case Overview

Meteorological analyses of the conditions leading to the Boscastle flood and the development of the observed storm are presented in Golding *et al.* (2005) and Golding (2005) and summarised briefly here. The large-scale flow on the 16 August 2004 was characterised by a large cut-off upper-level vortex to the west of Ireland. At 1200 UTC south-west England was under the left jet exit region of a jet stream maximum on the south-east flank of this vortex. This led to weak uplift maintaining high humidity and supporting the retention of cloud water which contributed to the unusually high rainfall efficiency (conversion of cloud rainfall to surface rainfall). The upper-level vortex was located directly above the complex, slow moving, low pressure system shown in Fig. 2, indicating a non-developing barotropic system. The two troughs marked approaching and over south-west England were also associated with this upper-level vortex and could be linked to maxima in upper-level potential vorticity. The environment was conducive to convective development (from a radiosonde sounding from Camborne at 1200 UTC) with moist unstable lower levels yielding a cloud base at about 900 m. There was moderate convective available potential energy (CAPE) of about 170 Jkg^{-1} and negligible convective inhibition.

3.1. *Observed storm development and evolution*

The first radar echos of the convective cells that later led to floods in the Boscastle area were observed along the northern Cornish coast, south-west of Boscastle at

1100 UTC. By 1130 UTC the observed rain rates already exceeded 32 mm hr^{-1} and intense, small cells (widths not exceeding 10 km) had spread along the coast towards Boscastle. New cells continued to form and move in the same direction at about 10 m s^{-1}, although downstream cell development led to an apparent speed closer to 15 m s^{-1}. By 1200 UTC a distinct line of cells had formed with rain rates still exceeding 32 mm hr^{-1}. Infrared satellite images suggest that the convective clouds extended only to mid-tropospheric depths and were mainly composed of liquid water at this time. However, it is possible that the clouds were being seeded with ice from the outflow cloud shields of the earlier storms over Brittany. At 1330 UTC infrared satellite images show a small area of colder cloud tops downstream of the Boscastle area; here the convective cells reached the tropopause height and the upper parts of the cloud had been moved ahead of the underlying active convection by high-level winds. By 1530 UTC a large cloud shield had developed here as cloud water rapidly turned to ice crystals. At 1630 UTC rainfall associated with an outflow boundary appeared in radar imagery and the storm was clearly decaying; scattered convection was present both over land and over the sea north of Cornwall. The radar accumulations for the most intense period of the storm are shown in Fig. 1.

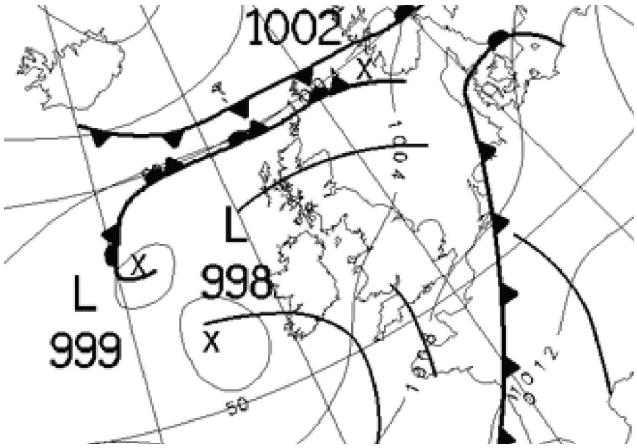


Figure 2. The Met Office surface analysis for 1200 UTC 16 August 2004 (Crown copyright)

Model simulations at 1 km and 4 km grid lengths carried out as part of the Met Office analysis (Golding *et al.* 2005; Golding 2005), show that convection was triggered by a low-level convergence line that formed during the morning along the northern coast of Cornwall. Convective cells continued to form during the morning and the early afternoon because of the wind shear which moved the precipitation ahead of the convection, preventing it from disrupting the convergence line until later in the afternoon. The extreme precipitation over Boscastle occurred because the convection, although strongly precipitating, enabled closely packed storm cells that moved along the same path precipitating over the same small area. High humidity conditions and the speed wind shear helped the downdraughts not to disrupt the convergence line; similarly to a rotational wind shear which has been observed to allow the development of strong low-level mesocyclones in supercells (Gilmore and Wicker 1998).

Sensitivity experiments by Golding (2005) showed that the convergence line was associated with the land-sea contrast in surface heat and moisture fluxes, since it disappeared when both fluxes over land were set to values typical of the sea. However, no strong evidence was found to link the convergence line to a sea breeze, and probably boundary layer circulations were involved.

3.2. *The control run*

The control simulation analysed here is generally consistent with the 1 km grid length simulations described by Golding *et al.* (2005) and Golding (2005). The model environment is much more unstable than indicated by the 1200 UTC radiosonde sounding from Camborne in Western Cornwall. Model CAPE values over land exceed 1200 Jkg^{-1} even at 0900 UTC with strong spatial variability in response to orography, surface temperature and cloud presence. Because of the strong spatial variability, this is not inconsistent with the Camborne sounding mentioned above. Furthermore, as in the observations, no strong lid was detected in the model simulations. The warm surface temperature and the high values of specific humidity are such that cloud base is generally low, often below 500 m,

and clouds are characterised by high water paths, both frozen and liquid.

The evolution of the low-level convergence and precipitation are now described and compared with observations from Golding *et al.* (2005) and Golding (2005) where possible: precipitation accumulations from 1200–1700 UTC are shown in Fig. 1(b); convergence is not shown. A convergence line is present along the northern coast of Cornwall from the start of the simulation at 0100 UTC and sits offshore until about 0900 UTC. It appears to be generated at the western-most tip of Cornwall and it interacts with the inflow from the western boundary of the domain as well as the precipitating cells. It weakens downstream of Boscastle at 1400 UTC but by 1600 UTC has reformed, albeit shifted to the east consistent with a gust front related to the convective cells (it may also merge with a second, newly formed, convergence line slightly inland from the southern coast). After 1700 UTC the intensity of both convergence lines starts diminishing and by 1900 UTC they have disappeared.

At around 0500 UTC the model develops a sharp line of precipitating cells along the convergence line off the northern coast of Cornwall. The extent of the line and its movements are such that weak precipitation in the vicinity of Boscastle is present until 1100 UTC; this weak precipitation was not observed. From 1100 UTC strong accumulations start to occur in both observations (2 km radar data) and the model simulation. During this same period the large-scale storm that crosses the south-east corner of the domain is largely in agreement with observations but is too weak and somewhat displaced to the south east.

At 1100 UTC the model output shows precipitating cells along the convergence line while the radar shows them only to the south west of Boscastle (on the coast) at this time. The development of the simulated storm is nonetheless quite close to the radar analysis: a continuous line of small but intense rain cells moving with the southwesterly mid-tropospheric wind. At this stage between about 1100 and 1600 UTC, the model develops cells that are narrower and tend have less intense peak rates, underestimating the total storm accumulations by roughly a factor of two (Fig. 1). After 1600 UTC the modelled storm decays into a few small cells that keep moving downstream whereas the radar observed broader and weaker cells forming along the convergence lines generated around the two tips of the Cornish peninsula. This difference in the latter part of the storm also contributed to the underestimation of the storm totals.

Three experiments were performed to explore the sensitivity of the generation of the convergence line and provide a context for the physics modifications chosen: the orography was flattened; and, the roughness length of the land points was changed to sea values with both the default and flattened orography. No significant changes in the onset, position and intensity of the convergence line occurred. Therefore, consistent with Golding *et al.* (2005) and Golding (2005), orography and the sea-land roughness contrast were found to modulate, but not to generate, the convergence line.

4. Methodology

4.1. Characterisation of the Target Area

Model skill for precipitation forecasts on scales of the order of a few Δx are notoriously low, and therefore to evaluate the convective scale ensembles it is necessary to focus on areas, rather than grid points. A target area, centred on Boscastle where the flood occurred, was defined over which to analyse the rainfall accumulations of the ensemble members. The area is circular and its diameter is determined using an analysis of the FSS (Roberts and Lean 2008) of the control simulation. The target area was introduced in order to focus the analysis on the flood event, or more specifically on the intense and localised rainfall accumulation. The flood occurred because the rain fell on a very small river catchment: roughly 20km^2. Because the model lacks skill at scales of a few kilometres, the FSS is used to inform a

reasonable choice for the size of the target area, ensuring that the effects of the perturbations applied to the model are assessed in terms of meaningful changes in the simulations and not as alternative realizations of the skill-less aspects of the forecast.

The FSS is a scale-selective measure of skill against observations (for specified thresholds used to convert the data to binary fields) and can be computed for scale lengths ranging from a few grid lengths up to nearly half the domain size. When averaged over a large number of cases, the FSS increases monotonically with scale length. Roberts and Lean (2008) describe how a reference value of FSS may be computed by considering a field uniformly distributed over the averaging area and which has the same frequency (fraction of gridpoints exceeding the threshold over the domain) as the observations. This reference score is denoted $FSS_{uniform}$ and a forecast is said to have skill at those scale lengths for which its FSS is higher than $FSS_{uniform}$.

The choice of accumulation period over which to assess the FSS and other diagnostics is necessarily a compromise. This should be be short enough so that the accumulated field captures the time evolution of the event, but long enough that it is not unduly sensitive to small timing errors in the model. Given that the timescale for a convective updraft is on the order of 30min, an accumulation period of 2h was chosen, with diagnostics being evaluated every 30min. An accumulation period of 30min led to similar (albeit more noisy) results.

Figure 3 shows, as a function of time, the intercept length, defined as the scale length at which the FSS of the control simulation reaches $FSS_{uniform}$. Values larger than 80km are not plotted because a domain width of 170km was used for the calculation (Section 3). The intercept length is shown for three thresholds: the 5th, 50th and 95th percentiles where, for example, the 95th percentile selects the highest 5% of the observed and forecast accumulations over all grid points for comparison. Percentiles are computed for each accumulation period and are preferred here over an absolute value in order facilitate comparison of the rain distributions

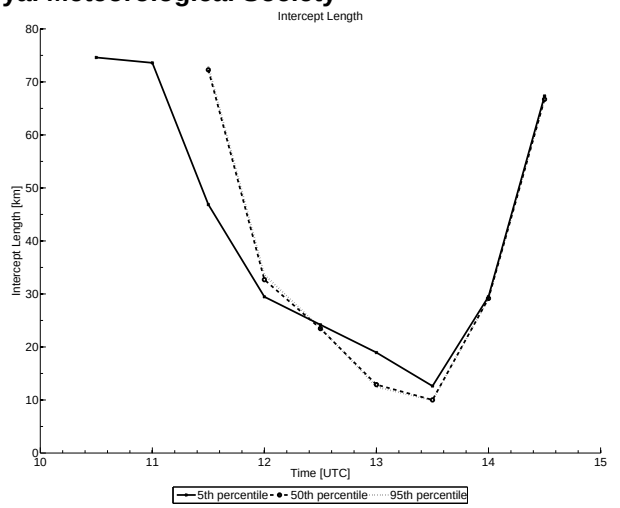


Figure 3. The Intercept Length, computed from the algorithm of Roberts and Lean (2008) using the radar-derived two hour accumulations at 5 km grid spacing as observations. Values larger than 80 km are not plotted. Note that the intercept lengths for the 50th and the 95th percentiles are almost identical.

at different times. The absolute accumulations change significantly during the event, and a fixed absolute threshold would sample quite different and changing portions of the observed and simulated distributions, rendering the interpretation more complex.

The intercept length varies with time and the FSS does not exceed $FSS_{uniform}$ during either the spurious early morning precipitation (up to 1000 UTC) or during the latter part of the storm (from 1500 UTC). In that latter part, compared to observations, the simulated precipitation features extend further downstream and decay too quickly. The shortest intercept lengths are obtained during the period preceding and during the early part of the storm, between 1230 and 1400 UTC. The intercept lengths at this time are close to the minimum possible values given that the original radar-derived rainfall observations had a grid spacing of 5 km. Note also that the intercept lengths at this time are considerably smaller than the typical range of 40–70km quoted by Roberts and Lean (2008). For much of the time, similar values of intercept length are obtained for the three percentiles, which reflects the localised and very intense character of the event. However, before 1230 UTC the 5th percentile is forecast more skilfully (shorter intercept length) than the 50th and the 95th:

during this period, the model forecasts the widespread and less intense precipitation better than the localised intense precipitation. Only when the storm becomes really intense do the 50th and the 95th percentiles produce lower values for the intercept length.

Based on Figure 3, the diameter of the target area has been chosen to be 60km; this exceeds the intercept length during the most active period of the event. The robustness of the results to be presented has been checked by also performing calculations for other target area sizes. Very similar results were obtained for diameters between 20 and 100km, although the plots become much more noisy when using 20km.

4.2. *Model-State Perturbation Ensemble Strategy*

The model-state perturbation method consists of the sequential application, throughout a simulation, of a randomly-generated two-dimensional perturbation in potential temperature on a specific model level. The perturbation is constructed from a linear superposition of Gaussian distributions with random amplitudes. The method is designed to account for inherent model uncertainty in the representation of the boundary layer, and is presented in full by Leoncini *et al.* (2010) who also discuss suitable choices of the parameters defining the perturbations. In the present study, a perturbation is applied every 30min, starting from 30min into the simulation. Perturbations are uncorrelated in time. The perturbation field is described by an amplitude, and by a scale length that represents the standard deviation of the Gaussian distribution.

The perturbed model level is 1280m above ground level, a choice to which there is little sensitivity since for a convective boundary layer the perturbation is rapidly mixed in the vertical. The maximum amplitude of the perturbation is set to 0.1 K which is on the order of typical fluctuations within the convective boundary layer at the smallest scales represented well by the model. Perturbations of this strength were shown by Leoncini *et al.* (2010) to be sufficient to generate considerable growth in the root-mean-square error of simulated precipitation. The standard deviation of the Gaussian is set to 8km, which will be well-resolved on the 1km model grid length. Note also that the perturbation field contains a number of maxima and minima within the 300km-wide domain, so that the model response to a perturbation will not be dominated by interactions between the background environment and one specific extremum.

Examples of the storm accumulations for two simulations with model-state perturbations are shown in Fig. 1(c) and (d), alongside the corresponding results for the radar-derived accumulations and the control simulation (Fig. 1(a) and (b) respectively).

4.3. *Physics Modification Ensemble Strategy*

Given the convective character of the event several microphysics parameters have been altered to explore the associated uncertainty and its effect on both cloud structure and precipitation.

The first set of changes concerns the autoconversion process, or more specifically the threshold liquid water content for its activation. The computation of this threshold, in the control simulation, depends on the number concentration of Cloud Condensation Nuclei (CCN) which is set to 3.0×10^8 m^{-3} over the land grid points and 1.0×10^8 m^{-3} over the sea, at all times and on all levels. The chosen values are realistic, but the small grid length used here and the coastal character of the event undermine the realism of discontinuous changes in CCN across the coast. Thus, two modifications have been implemented: land value of CCN concentration used everywhere (simulations labelled *Land*) and sea values used everywhere (simulations labelled *Sea*). In the *Aerosol* simulations CCN concentrations remain unaltered, but computation of the liquid water threshold makes use of an alternative formula for the threshold (Wilson and Forbes 2004), based on a previous version of the parameterization.

Ice processes may also play an important role in the development as described in Section 3.1. The threshold temperature for heterogeneous ice nucleation

has been changed from its standard value of $-10°$C to $-15°$C (simulations *Tnuc15*) and -5C (simulations *Tnuc5*). Such changes are large, but by no means unrealistic (e.g. Meyers *et al.* 1992). Other parameters relevant for heterogeneous nucleation are critical values of relative humidity and liquid water content (Fletcher 1962; Heymsfield and Miloshevich 1995) but these have not been altered.

Other modifications have been implemented to investigate the influence of the land/sea roughness contrast in modulating the convergence line that is responsible for the triggering of most of the convective cells (Section 3.1). The roughness lengths for C3 and C4 grasses, which together account for roughly 80% of the land cover, have been multiplied (simulations *Rough*2*) or divided (simulations *Rough/2*) by a factor of two. These changes are entirely plausible (Oke 1987) and are intended to represent uncertainty in the roughness of grass per se, as well as in the variability of other surface features (hedges, trees etc) that are normally present within areas designated as grass. The soil moisture has also been increased and decreased by 20% (simulations *SMup* and *SMdown* respectively) at all four levels. The control values are interpolated from the parent model. Such changes can affect convective initiation (e.g. Trier *et al.* 2004) and alter the thermodynamics of the land-sea contrast.

4.4. *Ensemble Size*

A set of 50 model-state perturbation simulations was produced using the default model physics to determine an appropriate ensemble size. A reasonable estimate of the ensemble-mean precipitation accumulation for the target area is required. Standard sampling theory indicates that error of the mean can be estimated by the standard deviation of the distribution divided by $\sqrt{N}$ where N is the ensemble size. However, it also requires the ensemble members to be independent and identically distributed. Therefore, a simple re-sampling approach has been adopted, in which 50 estimates of the target-area accumulation have been constructed from ensemble means using three possible ensemble sizes (5, 8 and 20) sampled from the full 50 member ensemble. Each such estimate of target-area accumulation is then expressed as an anomaly with respect to the 50-member ensemble-mean. If it can be assumed that 50 is indeed a large enough number to capture the variability of the model, then the anomalies provide an indication of the error associated with estimations of the accumulation based on a smaller number of realisations.

The time evolution of the means, standard deviations and ranges of these anomalies are shown in Fig. 4 to have some common features independent of sampling size. The spreads (standard deviations and ranges) of the anomalies grow slowly until 0500 UTC (during model spin-up) but then increase more rapidly. It is encouraging to note that the spreads are relatively small and static between 1300 and 1500 UTC, at the height of the storm. The largest spreads occur shortly before and during the onset of the main storm (peaking around 1100 UTC), and during the decay of the storm (after 1600 UTC). As expected, the spreads are smaller, at all times, with more ensemble members. Based on these results, 8 members was deemed to be a practical choice for this study. It is sufficient to be able to estimate accumulations to better than 10% for almost the full length of the simulations. This result is consistent with Clark *et al.* (2011) which found that convective scale ensembles of nine members had similar skill to the full ensemble of 17 members for the the median ROC curve of probabilistic quantitative precipitation forecasts of 6 hourly accumulations. The standard deviation of the anomalies is remarkably consistent with the error of the mean when estimated using sampling theory (Fig. 4), suggesting that the members are indeed independent and identically distributed. However, this does not address the issue of equal likelihood of the individual members.

Ten different configurations of the model physics are used in total, the default model physics and the nine modifications described in Section 4.3. A nine member ensemble was produced for each physics configuration

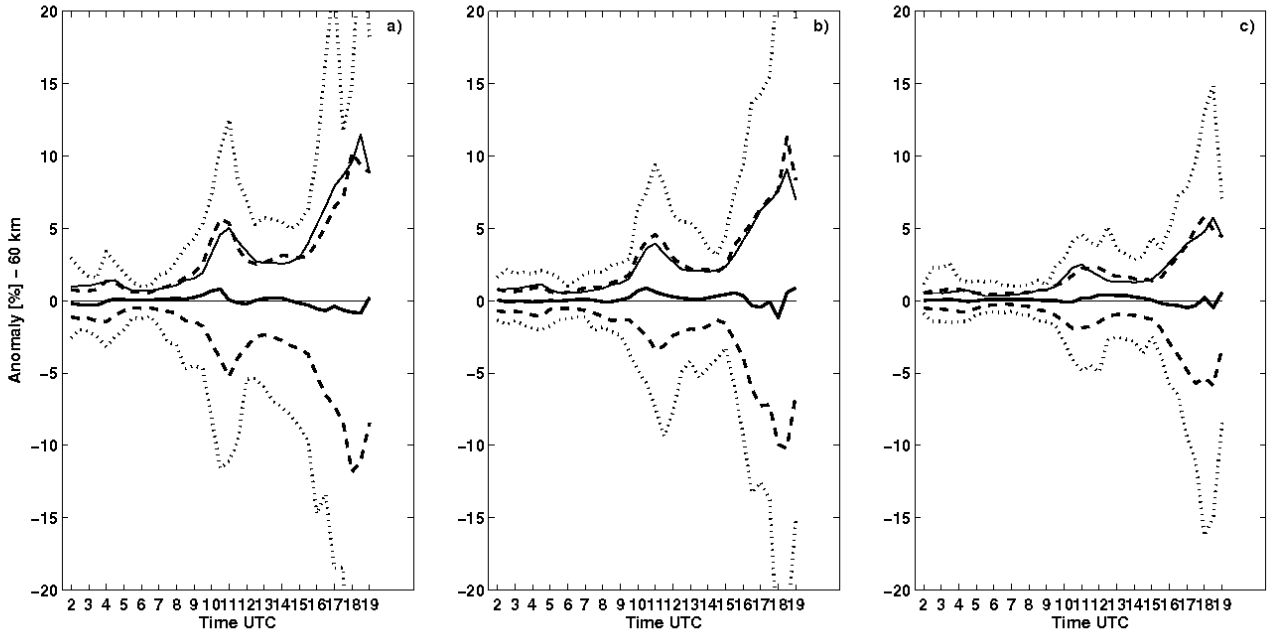


Figure 4. Time evolution of the anomaly (in %) relative to the 50-member ensemble mean of the target-area accumulation estimated with (a) 5, (b) 8, and (c) 20 ensemble members randomly chosen from the full 50 member ensemble. Solid thick lines are anomalies averaged over 50 samplings from the full ensemble, thin solid lines are the error of the mean estimated as $\sigma_{50\mathrm{mem}}/\sqrt{N}$, dashed lines denote one standard deviation of those samples, dotted lines denote their range (see text for more details).

consisting of eight *model-state runs* obtained from independent realisations of the model-state perturbation strategy presented in Section 4.2 and one simulation without model-state perturbations. Each set of nine simulations, with fixed model physics, and the same eight perturbations, is referred to as a *model-state physics ensemble*, and is labelled according to the model-physics configuration used. The *model-state standard physics ensemble* is the nine-member ensemble associated with the default MetUM physics. One more piece of nomenclature arises because it is useful to be able to refer to the set of runs without model-state perturbations. There are ten of these in total: one with default physics and nine *physics runs*, with modified physics. This set of ten is termed the *physics modification ensemble*. All simulations used in this study are summarised in Table I.

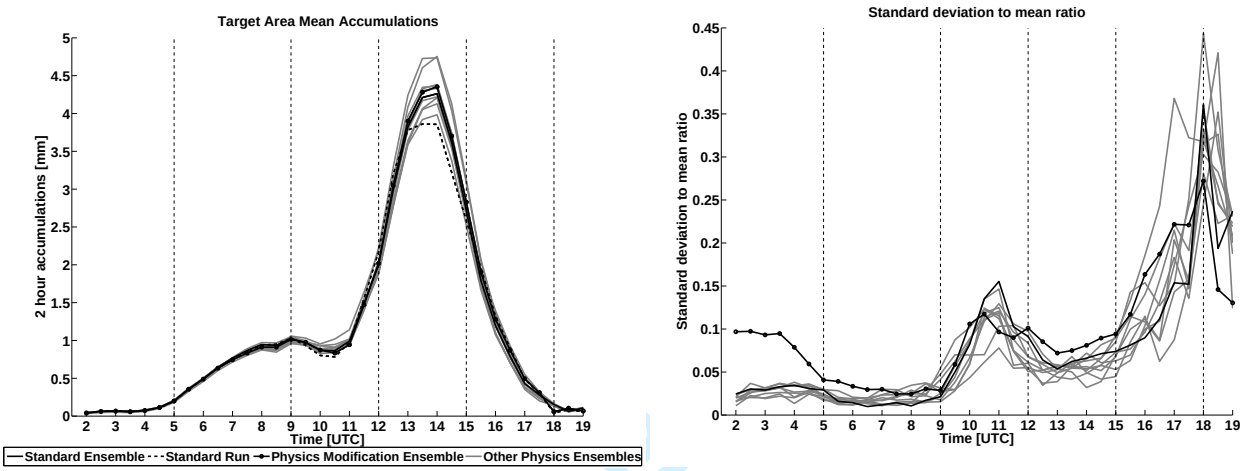
5. Analysis of ensembles

5.1. *Target Area Means and Spread*

Figure 5(a) shows the evolution of the two-hourly rainfall accumulations over the target area from the control run and the ensemble mean values from the nine model-state physics ensembles and the physics modification ensemble. The main peak of rainfall has some uncertainty, but in general the ensemble means are in close agreement with each other and with the control run. Radar observations peak later because of the different model behaviour during the latter part of the storm, mentioned in Section 3.2. Also, the earlier peak between 0700 and 1000 UTC was not observed in reality (Section 3.1). Individual ensemble members share a similar behaviour (not shown) although with slightly larger variations than the corresponding means. Sensitivity simulations with autoconversion disabled did not produce this secondary peak, which is indeed due to warm rain processes in the model simulations. However, this change (disabling the

Table I. Physics configurations used. The first two columns with sideways text contain the names of the two corresponding groups of runs without model-state perturbations. All the individual runs are part of a model-state perturbation ensemble. See text for further details.

		Name	Description
physics runs	physics modification perturbation ensemble	Standard phys.	Default physics
		Aerosol	Old version for autoconversion threshold
		Land	Land value of CCN concentration used everywhere
		Sea	Sea value of CCN concentration used everywhere
		Tnuc15	Ice heterogeneous nucleation temperature set to -15°C
		Tnuc5	Ice heterogeneous nucleation temperature set to -5°C
		Rough*2	Roughness of grasses doubled
		Rough/2	Roughness of grasses halved
		SMup	Initial soil moisture is increased by 20%
		SMdown	Initial soil moisture is decreased by 20%



(a) (b)

Figure 5. (a) Ensemble means of the running two-hourly rainfall accumulation over the target area for the model-state standard physics ensemble (black solid line), the physics modification ensemble (solid black line with circles) and all of the model-state physics ensembles (solid grey lines). The dotted line shows accumulations for the control simulation. (b) The ratio of the standard deviation across each ensemble to its ensemble-mean accumulation (solid black and grey lines as for (a)).

autoconversion) did not affect the main evolution of the storm, it only accelerated its decay.

The time evolution of the ratio of the standard deviation of each ensemble to its mean is plotted in Fig. 5(b) as a measure of the relative spread. The main features of the time evolution of the spread are common to all of the ensembles. A notable difference is the much larger spread from the physics modification ensemble prior to about 0500 UTC. This reflects the different character of its members, which are from slightly different configurations of the model, rather than different realisations of the same simulation. Spin-up may be achieved slightly differently in each configuration, resulting in the higher ratio at early times. The two main periods of greatest relative spread occur between 0900–1200 UTC, during the transition from the warm rain period to the peak of the main storm, and after 1500 UTC when the main convective elements are decaying. Large variabilities at these times are consistent with the increased uncertainty in the ensemble-mean from the model-state standard physics ensemble (Fig. 4). Within the latter period, the large variations of the relative spread over time and across ensembles can be attributed to the low values of mean precipitation and to the fact that each physics modification or model-state perturbation has had more time in which to exert an influence on the evolution of the simulations. However, the earlier period of large relative spread occurs when there is stronger precipitation, at the time when the cells associated with the Boscastle storm are

initiated. One likely cause of the increased relative spread at this time is that the model-state perturbations within the boundary layer will affect precisely where and when cells are initiated, as was found by Leoncini *et al.* (2010).

This analysis of Fig. 5(b) has implicitly identified five distinct periods during the simulations which it is convenient to name here for future ease of reference (these are marked on Fig. 5): a *spin-up* period from the outset to 0500 UTC; a *warm rain period* between 0500 and 0900 UTC; a *transitional* period between 0900 and 1200 UTC; a *storm* period between 1200 and 1500 UTC; and a *decay* period between 1500 and 1800 UTC.

5.2. *Relative Dispersion of the Model-State Perturbation and Physics Modification Ensembles*

At the synoptic scale, ensembles constructed from atmospheric models are often under-dispersed (i.e. the ensemble standard deviation is smaller than the root-mean-square error) (Buizza *et al.* 2005). It is useful to compare ensembles using their relative variance, computed here as a normalised variance difference (NVD) (Gebhardt *et al.* 2011)

$$\mathrm{NVD} = \frac{\sigma 2 - \sigma 2_{\mathrm{ref}}}{\sigma 2 + \sigma 2_{\mathrm{ref}}}. \qquad (1)$$

where $\sigma 2$ is the variance of the ensemble and $\sigma 2_{\mathrm{ref}}$ is the variance of the reference ensemble. Thus a positive value indicates an ensemble with more spread than the reference ensemble.

Values of NVD for the target-area two-hour mean accumulations are shown in Fig. 6, taking the model-state standard physics ensemble as the reference and averaging over each of the five periods identified in the previous subsection. Each model-state physics ensemble has a broadly similar time evolution of NVD, which contrasts with that of the physics modification ensemble. During spin-up, the physics modification ensemble has a strongly positive NVD, while most of the other ensembles have negative NVD. The warm rain period is characterised by positive values for all of the ensembles with the physics

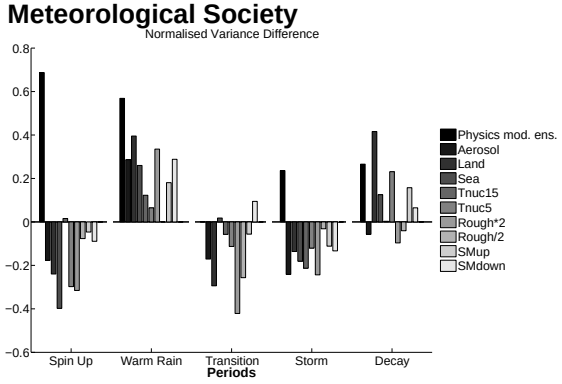


Figure 6. NVD of two-hourly rainfall accumulation over the target area. The NVD is shown for each period of the storm. The identity of each ensemble plotted is indicated by the legend and the reference ensemble in each case is the standard physics ensemble.

modification ensemble having the largest values. NVD is mainly negative during the transition period (with large variability amongst the ensembles). The NVD is positive for the physics modification ensemble but negative for all of the model-state physics ensembles during the storm period. Finally, NVD values are generally positive during the decay period. Thus the NVD of the physics modification ensemble is larger than that of the model-state physics ensembles during both of the peak rain periods: warm rain and storm. Note also that none of the ensembles has a consistent sign throughout.

5.3. *Response of the Different Physics Configurations to Model-State Perturbations*

The NVD of the different ensembles indicates that the model can acquire different sensitivities to model-state perturbations when the model physics is changed (Section 5.2). To investigate this issue further, the members of each model-state physics ensemble are compared with their equivalent-physics unperturbed simulations. The comparison is made in terms of a slightly-modified version of the SAL method of Wernli *et al.* (2008, 2009).

The SAL method was developed to evaluate precipitation forecasts against observations, and has also been used to evaluate air pollution forecasts (Dacre 2010). Here, it is used to provide measures of difference between simulations with and without model-state perturbations. The SAL

method provides three scores – structure, amplitude and location – which together give a comprehensive and quantitative view of forecast difference. The amplitude score (A) is the normalised difference of precipitation accumulation within the domain: if positive then the model-state perturbed simulation precipitates more strongly. The structure score (S) depends on how precipitation (above some threshold) is distributed amongst spatially-coherent objects: if positive then objects in the model-state perturbed simulation are broader and flatter. Both S and A range from -2 to 2, with 0 indicating simulations that are indistinguishable by that measure. The location score (L) ranges from 0 to 2 and measures displacements of the "centre of mass", both for the full precipitation field and for each object.

The focus of this work is the flood-producing precipitation from convective elements in a specific event. Hence, the scores A and L have been slightly modified here, so that they are based on precipitation objects only, rather than the entire precipitation field. This means that all three scores depend upon the threshold used to define objects. The SAL plots in Fig. 7 are for the 95th percentile of storm accumulation from 1200–1700 UTC, and show scores for simulations with model-state perturbations in comparison to the run with corresponding model physics but without model-state perturbations. The threshold is chosen as a percentile to avoid bias. Not surprisingly, the values of A and L are smaller than those typically obtained when comparing forecasts and observations, which can often be larger than 1 (Wernli *et al*. 2008, 2009).

The eight perturbed members in each of the model-state physics ensembles are not randomly distributed in S, A and L. Rather they tend to cluster within a limited area of SAL space, the location of that area depending upon the physics configuration. For example, all of the model-state perturbed members with the reduced soil moisture (panel labelled *SMdown*) have $A < 0$, indicating that they precipitate more weakly than the corresponding unperturbed simulation. They also tend to have negative values of S, corresponding to smaller and more peaked precipitation objects. This implies that these physics modifications alter the overall sensitivity of the model to the model-state perturbations. We note also that changes to the threshold can alter significantly the position of clusters (not shown). Thus, the sensitivity to model-state perturbations changes with both the threshold and the model physics.

5.4. *Effects of physics modification on cloud structure*

The ratio of the ensemble rainfall standard deviation to its mean has a similar time evolution for the model-state physics ensembles and the physics modification ensemble (after the spin-up time, Fig. 5(b)) but this spread is associated with different changes in the structure, amplitude and location of precipitation structures (Fig. 7). One possible source for these changes, systematic effects of model physics modifications on the averaged cloud structure, is examined here.

To construct suitable cloud-average profiles, first, cloudy grid columns were defined to be those with either a liquid water path exceeding 1kgm^{-2} or a frozen water path exceeding 2kgm^{-2}. These choices allow for clouds to be identified both at their time of formation within the boundary layer and also in their final stages, when the wind shear has advected the ice further downstream than the liquid water. Clearly, the specific values of the thresholds are somewhat arbitrary. Tests with lower thresholds tended to detect larger clouds less representative of the convective cells that characterise the event. Tests with higher thresholds focussed the analysis on fewer, smaller clouds that were representative only of the strongest convective cells. In-cloud profiles were computed every hour over the area shown in Fig. 1 and averaged over the cloudy grid points and then over the transition and storm periods. Some checks with five-minute data from the control simulation confirmed that hourly data is sufficient to construct these profiles reliably.

Having constructed profiles for each member of the physics modification ensembles, Figure 8 shows the results in terms of a mean across each physics configuration

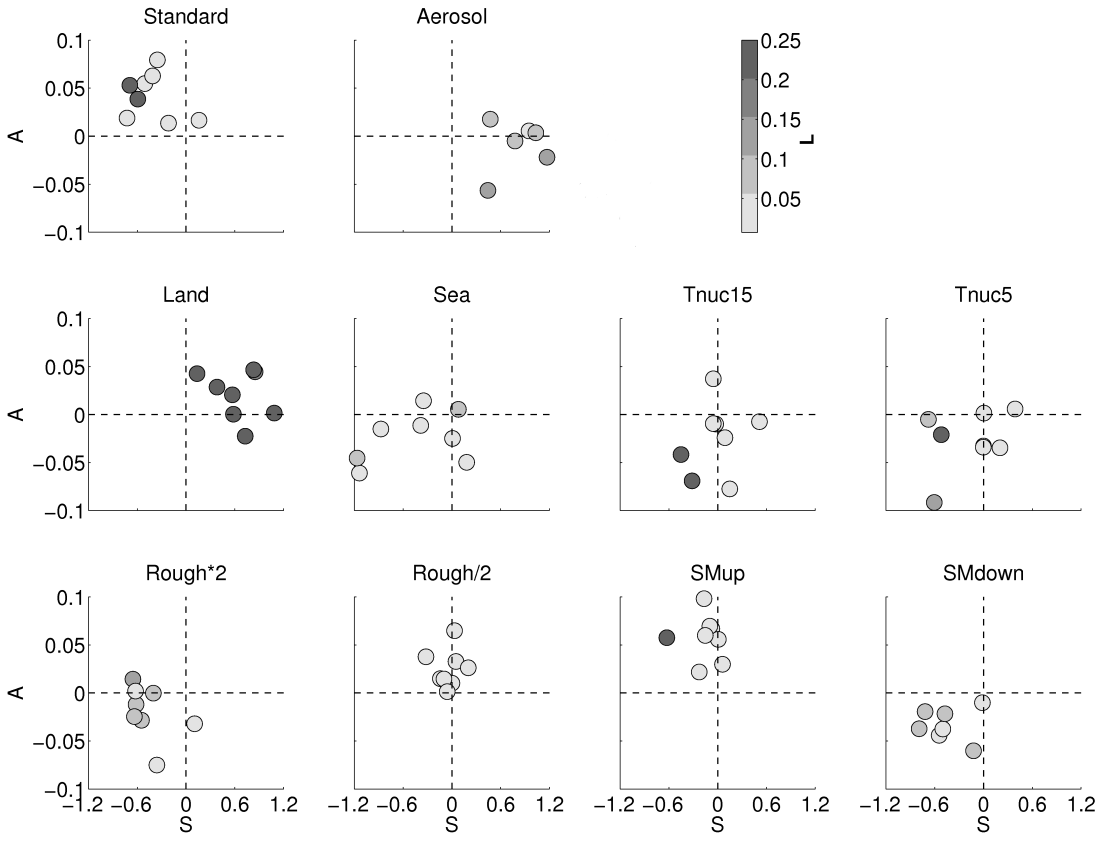


Figure 7. SAL scores (scores for L shaded) for all of the model-state perturbation runs, each being compared against the run with the same model physics, but no model-state perturbations. The 95th percentile of precipitation accumulation between 1200 UTC and 1700 UTC has been used as a threshold.

(plus and minus one standard deviation). Also shown are the profiles for the control simulation. The evolution is characterised by the growth of ice content and increases in vertical velocity during the transition period, leading to a secondary vertical velocity maximum at roughly 8 km during the storm period (Fig. 8). While the primary maximum lies just above cloud base, the secondary maximum is associated with a secondary maximum in ice content, consistent with Heymsfield *et al.* (2010) and Fierro *et al.* (2009).

The physics modifications have little effect on the cloud profiles (comparing the mean of the runs with modified physics to the control simulation and considering the spread of the runs with modified physics). At least for the simulations in which parameters for the autoconversion process are altered (*Aerosol*, *Land* and *Sea*), this can be partially attributed to the large availability of water vapour which results in high liquid water content, regardless of the concentration of cloud condensation nuclei and the autoconversion threshold. Similar mean cloud profiles could also be achieved through slightly different physical balances. This is certainly plausible for the simulations with changes to the heterogeneous ice nucleation threshold (*Tnuc15* and *Tnuc5*). Lowering the threshold to $-15°$C will tend to reduce the ice and hence increase the water content of a cloud. When the ice precipitates it acts as a seed for the collision and coalescence process, which now has fewer seeds, but more water content available for collision. Analogous reasoning applies for the *Tnuc5* simulations. Overall, the model responds to these physics perturbations by limiting their effect. Another example of a negative microphysical feedback is the response to the doubling

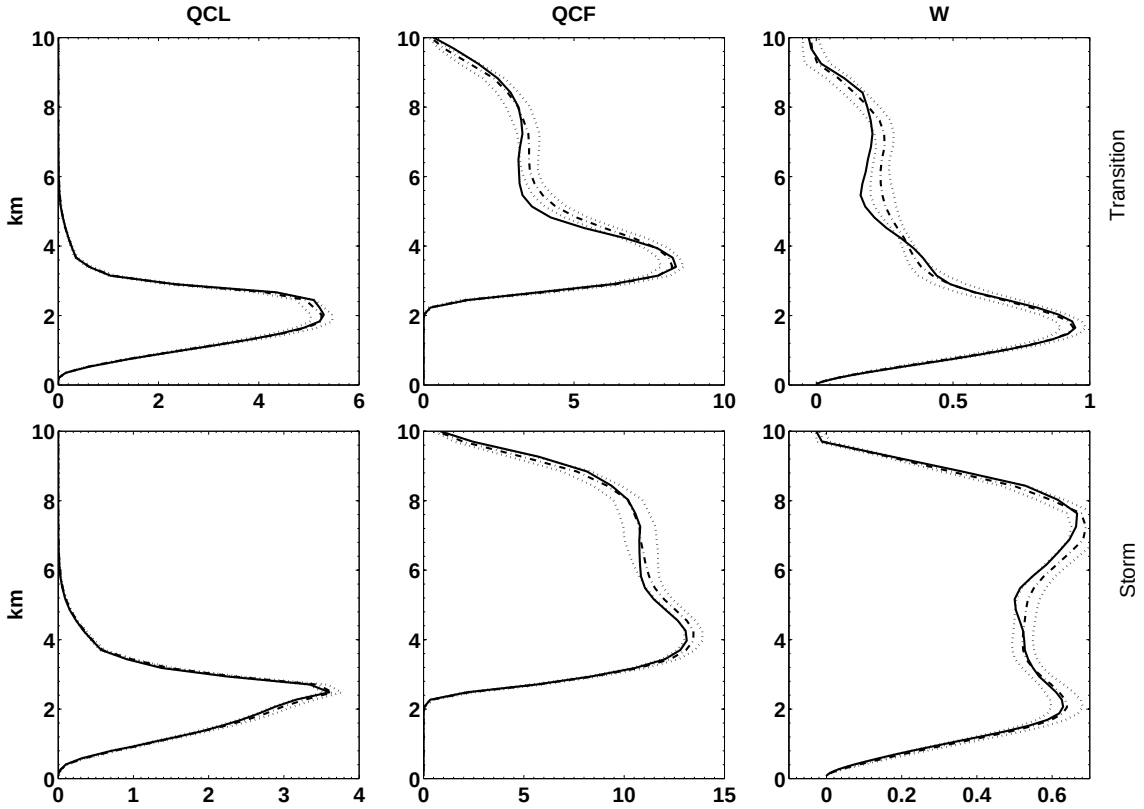


Figure 8. In-cloud vertical profiles, averaged over the transitional period (top row panels) and the storm period (bottom row panels) of liquid water content (in units 10^{-4}kg kg^{-1}, left column), ice content (in units 10^{-4}kg kg^{-1}, central column) and vertical velocity (in units of ms^{-1}, right column). The solid line represents the standard run, the dashed-dotted line represents the mean of the physics modification ensemble members, and the dotted lines show the one standard deviation departures from the mean.

of ice terminal fall speed and the deposition/sublimation rate that Forbes and Clark (2003) found for three of the FASTEX cases. However, their ice content changes for the ice terminal fall speed were significantly larger, highlighting a stronger sensitivity. It is less obvious (and outside the scope of this work) to determine how other physics changes affect the cloud profiles. However, we note that model physics changes not only affect the cloud structure instantaneously, but also affect the environment within which clouds form and develop, as evidenced for example by the variability in spin-up (Section 5.1) and in the different sensitivities to model-state perturbations (Section 5.3). The vertical velocity profiles show that the larger standard deviations are associated with the larger standard deviation of both the ice and water cloud content, highlighting the coupling between the microphysics and dynamics of the cloud.

6. Conclusions

The development of ensemble strategies for convective-scale forecasting is currently an area of active research. Here we have investigated simulations of the Boscastle storm of 16 August 2004 with a model grid length of 1km, in an analysis that considers the initial and boundary conditions to be given, but which admits imperfections in the model physics. Specifically, we have contrasted uncertainties arising from structural representativity errors in parameterizing the boundary layer (Section 4.2) with uncertainties associated with some key parameter and physics choices in the model parameterizations (Section 4.3). Boundary layer perturbations suitable for

describing local fluctuations were applied following the superposition-of-Gaussians methodology of Leoncini *et al.* (2010), while physics modifications were motivated from considerations of the meteorology of the specific event.

While all the model forecasts showed a significant bias in total storm accumulations, the overall evolution and spatial pattern of precipitation are consistent with the observations and relatively insensitive to perturbations applied. The Boscastle event is found to have a high degree of predictability: rainfall accumulations in the area around Boscastle are very accurately positioned (Section 4.1) and all of the essential qualitative features of the event (Section 3) are robust to all of the model-state perturbations and physics modifications considered here. Key to the event is the repeated triggering and propagation of cells along a convergence line, located a little inland of the northern coast of the Cornish peninsula. That line is dynamically produced by the land-sea contrast within the prevailing synoptic flow, and while its details may change somewhat across simulations, such changes are not sufficient to alter the character of the event.

With regard to the two types of model uncertainty considered, we have found that the spread associated with model-state perturbations is similar to that associated with the physics modifications. Thus, we conclude that the model-state perturbation strategy proposed by Leoncini *et al.* (2010) is indeed capable of producing physically plausible ensemble simulations, broadly consistent with credible changes to the model parameter and physics settings. Moreover, it was found that the ensemble spread could be well estimated from a modest number of model-state perturbation members (Section 4.4), eight members having been found to be sufficient for the purposes of the present study.

These remarks are based on analysis of the rainfall associated with the Boscastle event, accumulated over space and timescales for which the model was shown to have predictive skill (Section 4.1). For this diagnostic, the model-state perturbation strategy alone may be considered a good-enough approach for capturing broad aspects of the sensitivity of the simulation to model uncertainties.

Nonetheless, a number of differences between the model-state perturbation and physics modification strategies have been demonstrated, such that both methods are required for a more complete description of the simulation uncertainties. For example, model-state perturbations produced the largest relative spreads during the transitional and decay periods of the storm (Section 5.1), whereas physics modifications were more effective at generating spread during the period of the main rainfall event (Section 5.2). In essence, boundary-layer fluctuations were somewhat more important for the initial triggering of the storms whereas physics modifications were more important for the development of the triggered storms.

Differences between the ensemble-generation strategies were most evident, however, in the SAL diagnostics describing the morphology and location of the convective cells. While different ensemble strategies produced similar spread in terms of the rainfall accumulations, they produced that spread by altering the character of individual cells in different ways (Section 5.3). Different model physics produced different sensitivities to model-state perturbations and systematic differences in the individual convective cells, notwithstanding the fact that averaging over cells produced similar in-cloud profiles (Section 5.4).

Of course, there must be the caveat that a single convective case has been investigated here, and that it is one for which the trigger, and the main features of the environment, are dictated by dynamics and are not sensitive to details of the physics. Further studies would therefore be highly desirable, but it is nonetheless tempting to speculate about the implications should similar conclusions prove to hold more broadly. Certainly the model-state perturbation approach has valuable practical advantages in that it is a simple, generic approach that can be applied without any necessity for a careful prior consideration of the physics that may be important for a particular event. If the approach does give a good indication of model uncertainties in a

wider range of cases, then the problem of accounting for model uncertainties in short-range convective-scale forecasts might actually prove to be simpler and more tractable than is often supposed.

Acknowledgement

The authors would like to acknowledge Prof Heinli Wernli for the SAL code, Mr Nigel Roberts for the FSS code and the NCAS Computational Modelling Services team for their invaluable help. This work was funded by the Flood Risk from Extreme Events (FREE) program of the National Environment Research Council, grant no. NE/E002137/1.

References

Bennett LJ, Browning KA, Blyth AM, Parker DJ, Clark PA. 2006. A review of the initiation of precipitating convection in the United Kingdom. *Q. J. R. Meteorol. Soc.* **132**: 1001–1020.

Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, Zhu Y. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.* **133**: 1076–1096.

Burt S. 2005. Cloudburst upon Hendraburnick down: The Boscastle storm of 16 August 2004. *Weather* **60**: 219–227.

Clark AJ, Coauthors. 2011. An overview of the 2010 hazardous weather testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.* : In press.

Clark AJ, Gallus WA, Xue M, Kong F. 2009. A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Weather and Forecasting* **24**: 1121–1140.

Clark AJ, Kain JS, Stensrud DJ, Xue M, Kong F, Coniglio MC, Thomas KW, Wang Y, Brewster K, Gao J, Wang X, Weiss SJ, Du J. 2011. Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.* **139**: 1410–1418.

Dacre HF. 2010. A new method for evaluating regional air quality forecasts. *Atmos. Env.* doi:10.1016/j.atmosenv.2010.10.048.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Edwards J, Slingo A. 1996. Studies with a flexible new radiation code. Part I: Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc.* **122**: 689–719.

Essery RLH, Best MJ, Betts RA, Cox PM, Taylor CM. 2003. Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *J. Hydromet.* **4**: 530–543.

Fierro AO, Simpson JM, LeMone MA, Straka JM, Smull BF. 2009. On how hot towers fuel the Hadley cell: An observational and modeling study of line-organized convection in the equatorial trough from TOGA COARE. *J. Atmos. Sci.* **66**: 2730–2746.

Fletcher NH. 1962. *The physics of rainclouds*. Cambridge University Press.

Forbes RM, Clark PA. 2003. Sensitivity of extratropical cyclone mesoscale structure to the parametrization of ice microphysical processes. *Q. J. R. Meteorol. Soc.* **129**: 1123–1148.

Fujita T, Stensrud DJ, Dowell DC. 2007. Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.* **135**: 1846–1868.

Gebhardt C, Theis SE, Paulat M, Bouallegue ZB. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbation and variations of later boundaries. *Atmos. Res.* **100**: 168–177.

Gilmore MS, Wicker LJ. 1998. The influence of midtropospheric dryness on supercell morphology and evolution. *Mon. Wea. Rev.* **126**: 943–958.

Golding B. 2005. Boscastle and North Cornwall post flood event study - meteorological analysis of the conditions leading to flooding on 16 August 2004. Technical report, Met Office. Forecasting Research Technical Report No. 459.

Golding B, Clark P, May B. 2005. The Boscastle flood: Meteorological analysis of the conditions leading to flooding on 16 August 2004. *Weather* **60**: 230–235.

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

Heymsfield AJ, Miloshevich LM. 1995. Relative humidity and temperature influences on cirrus formation and evolution: observations from Wave Cloud and FIRE II. *J. Atmos. Sci.* **52**: 4302–4326.

Heymsfield GM, Tian L, Heymsfield AJ, Li L, Guimond S. 2010. Characteristics of deep tropical and subtropical convection from nadir-viewing high-altitude airborne doppler radar. *J. Atmos. Sci.* **67**: 285–308.

Hohenegger C, Walser A, Langhans W, Schar C. 2008. Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quart. J. Roy. Meteorol. Soc.* **134**: 889–904.

Lean HW, Ballard SP, Clark PA, Dixon MA, Li Z, Roberts NM. 2005. The summer 2004 reruns with the High Resolution Trial Model. Technical Report 466, MetOffice R&D.

Lean HW, Clark PA, Dixon M, Roberts NM, Fitch A, Forbes R, Halliwell C. 2008. Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.* **136**: 3408–3424.

Lean HW, Roberts NM, Clark PA, Morcrette C. 2009. The surprising role of orography in the initiation of an isolated thunderstorm in southern England. *Mon. Wea. Rev.* **137**: 3026–3046.

Leoncini G, Plant RS, Gray SL, Clark PA. 2010. Perturbation growth at the convective scale for CSIP IOP18. *Q. J. R. Meteorol. Soc.* **136**: 653–670.

Lock AP. 2001. The numerical representation of entrainment in parameterizations of boundary layer turbulent mixing. *Mon. Wea. Rev.* **129**: 1148–1163.

Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB. 2000. A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev.* **128**: 3187–3199.

Marsigli C, Montani A, Nerozzi F, Paccagnella T. 2004. Probabilistic high-resolution forecast of heavy precipitation over central Europe. *Nat. Hazards Earth Syst. Sci.* **4**: 315–322.

Mason PJ, Thomson DJ. 1992. Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.* **242**: 51–78.

Meyers MP, DeMott PJ, Cotton WR. 1992. New primary ice-nucleation parameterizations in an explicit cloud model. *Journal of Applied Meteorology* **31**: 708–721.

Oke TR. 1987. *Boundary layer climates*. Routledge, second edn. Pp. 435.

Roberts NM. 2003. Stage 2 report form the storm-scale numerical modelling project. Technical Report 407, MetOffice R&D, URL http://www.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/index.html.

Roberts NM, Lean HW. 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.* **136**: 78–97.

Shutts GJ. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **612**: 3079–3102.

Trier SB, Chen F, Manning KW. 2004. A study of convective initiation in a mesoscale model using high-resolution land surface initial conditions. *Mon. Wea. Rev.* **132**: 2954–2976.

van Weverberg K, van Lipzig NPM, Delobbe L, Lauwaet D. 2010. Sensitivity of quantitative precipitation forecast to soil moisture initialization and microphysics parameterization. *Q. J. R. Meteorol. Soc.* **136**: 978–996.

Wernli H, Hofmann C, Zimmer M. 2009. Spatial forecast verification methods intercomparison project: Application of the SAL technique. *Weather and Forecasting* **24**: 1472–1484.

Wernli H, Paulat M, Hagen M, Frei C. 2008. SAL - a novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Wea. Rev.* **136**: 4470–4487.

Wilson D, Forbes R. 2004. Unified model documentation paper 26: The large-scale precipitation parametrization scheme. Technical Report 26, MetOffice R&D.

Wilson DR, Ballard SP. 1999. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **125**: 1607–1636.

Yussouf N, Stensrud DJ. 2011. Comparison of single-parameter and multi-parameter ensembles for assimilation of radar observations using the ensemble kalman filter. *Mon. Wea. Rev.* : In press.

Zhang F, Bei N, Rotunno R, Snyder C, Epifanio CC. 2007. Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.* **64**: 3579–3594.

Zhang F, Snyder C, Rotunno R. 2003. Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.* **60**: 1173–1185.