

i. Motivation of report

Convective-scale phenomena can cause extreme damage in the form of hazards caused by extreme rainfall, such as flooding. Due to difficulty in forecasting such small-scale events, people and services are often left unprepared. With the increase of computer power, research into forecasting these intense, small scale systems is rapidly developing. Extreme convective-scale events have a low probability of occurrence and a high impact, with possible significant economic effects. Therefore diagnosis of their occurrence is vital to many people and services, to allow time to issue warnings and keep people safe. The need for early warnings is important now and may increase in the future, as these extreme rainfall events are predicted to increase in frequency.

Creating models that can accurately represent these damaging systems has been the focus of much research; however, as this is relatively new area, there is still much more work that can be done. Various projects have investigated the effect on forecast skill of increasing resolution and perturbing the model to create ensembles, and so this report investigates the prospects for ensembles at this high resolution.

ii. Aims of report

- To determine at what scale the model shows a selected level of skill
- To investigate which ensemble members deviate most from a control run
- To diagnose a small group of perturbations that show the largest deviation
- To introduce a method that can be used to select ensemble members that may be useful if a 5-6 member ensemble is needed to be generated

iii. Structure of report

The processes leading to, and the effects of convective events are presented in section 1.1, illustrated by a case study of the Boscastle storm (section 1.2). Forecast skill and recent modelling development are introduced in section 1.3. In section 1.3.4, analysis of observations and model output of the Boscastle storm are examined and forecast skill of this event is discussed. An explanation of ensemble forecasting and its added value to forecast output is given in section 1.4. Details of the model used to create forecasts analysed in this report are in section 2, which also introduces the perturbation strategy (section 2.6). The approach of this analysis is discussed in sections 3.1-3.3, with results presented in sections 3.4-3.11. A summary and conclusion, including discussion of further work and a proposed method for choosing a useful, small ensemble, is given in section 4.

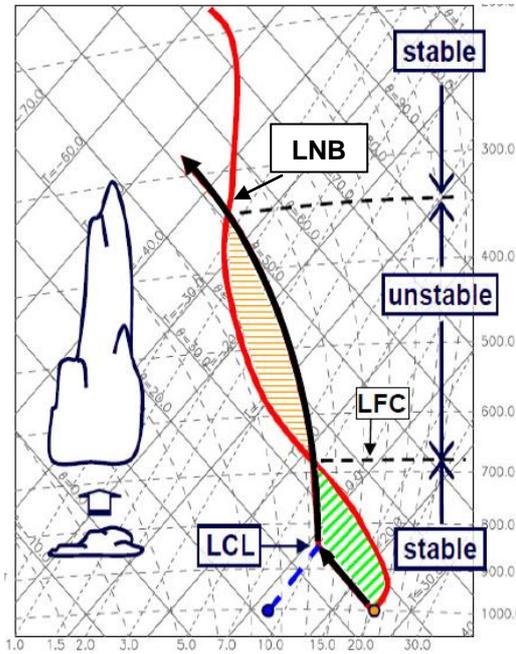
1. Introduction

1.1. Initiation and impacts of convective events in the UK

In the UK, extreme rainfall events generally occur in the form of major convective storms (Bennett *et al*, 2006). These occur in the summer and in early autumn, as insolation is an important factor for convective initialisation (Hand *et al*, 2004). Even though atmospheric instability can be high in spring, the relatively low sea temperatures and colder air mean less available moisture for rain-producing systems (Hand *et al*, 2004). Climate models predict an increase in UK heavy precipitation throughout the year, though more consistently during winter than summer (Maraun *et al*, 2009).

Convective events are more localised and less continuous than frontal events and can be broadly categorised into forcings from a synoptic-scale feature, and forcings from insolation or a mesoscale feature (Hand *et al*, 2004). Hand *et al* (2004) showed that such synoptic features have high Convective Available Potential Energy (CAPE), and forcings from insolation or a mesoscale feature have lower values of CAPE. CAPE is the amount of energy a parcel of air could have if lifted a certain distance vertically through the atmosphere, and this parameter is therefore an indicator of atmospheric stability, valuable for predicting the strength of convection. The tephigram shown in figure 1 can be used to consider stability of the air column represented. The crossing of vertical profiles of the environment and the air parcel create an area of positive energy (red) and an area of negative energy (green). The positive area indicates the CAPE in the air column, and the negative area represents the Convective Inhibition (CIN). CIN is the amount of energy required for an air parcel to overcome the negatively buoyant energy in the environment, and typically a column with large convective inhibition is considered stable and has very little likelihood of developing a thunderstorm. However, a small amount of CIN is needed for extreme convection, as it allows CAPE to build up. Without any CIN, convection would eliminate small amounts of CAPE, inhibiting the development and therefore release of a large amount of energy in the form of a convective storm.

Figure 1. Example tephigram, indicating CAPE and CIN (MSc Atmospheric Physics course notes, 2009). The red line represents the vertical profile of the environment, and the black line represents the vertical profile of the air parcel.



$$\text{CAPE} = \int_{P_{\text{LNB}}}^{P_{\text{LFC}}} R_d (T_p - T_a) d \ln p \quad (1)$$

$$\text{CIN} = \int_{P_{\text{LFC}}}^{P_{\text{LCL}}} R_d (T_p - T_a) d \ln p \quad (2)$$

Equations 1 and 2 use standard meteorological notation.

The Lifting Condensation Level (LCL) is the pressure at which a cloud base would form, as the air parcel reaches saturation. Once the CIN is overcome and the Level of Free Convection (LFC) is reached, the parcel has positive energy and will continue to rise until the Level of Neutral Buoyancy (LNB) is reached. The larger the amount of CAPE stored and released, the deeper the convection.

Bennett *et al* (2006) classified three main areas of important processes for the initiation of convection in the UK: boundary-layer forcing, upper-level forcing, and secondary generation.

1.1.1. Boundary-layer forcing

Boundary-layer forcings determine the specific location where convection is triggered within larger regions of potential instability, created by mesoscale or synoptic-scale features at a higher level (Hand *et al*, 2004) such as dry intrusions (Bennett *et al*, 2006). Convergence lines caused by orography, surface moisture gradients, differential heating, synoptic-scale fronts, gust fronts, roughness contrasts and land/sea contrasts are preferred areas for convective development (Bennett *et al*, 2006), with convergence generated by hills and coasts

usually more important than differences in land surface properties. The topography of the UK has a large influence on the initiation of convection (Collier, 2007). Maraun *et al* (2009) used a model based on extreme value statistics to show that the annual mean of heavy precipitation is highly correlated with orography in the UK.

Coastlines have a mechanical and thermal effect on atmospheric flow. As the UK receives relatively low insolation, sea-breeze-induced convection occurs less frequently and less intensely than elsewhere, and other coastal effects such as frictional convergence could be more important (Bennett *et al*, 2006).

The distribution of moisture in the boundary layer is an important factor in determining where convection might develop. Crook (1996) showed that convection initiation is sensitive to surface temperature and surface moisture, and variations within observational variability made the difference between convection being intense or non-existent, as convection depends sensitively on the amount of CIN.

1.1.2. Upper-level forcing

In the UK when low wet-bulb potential temperature air (frequently associated with a dry intrusion) overruns high wet-bulb potential temperature in the boundary-layer, then deep convection can be triggered once the instability is released either by surface heating or large-scale ascent (Browning & Roberts, 1994). A dry intrusion is associated with an upper-tropospheric maximum in potential vorticity, which induces upward motion ahead of it and downward motion behind. This combination of advection of low wet-bulb potential temperature air above higher wet-bulb potential temperature air and the upward motion ahead of the potential vorticity anomaly is conducive to thunderstorm development.

1.1.3. Secondary generation

Second-generation cells are those formed by the interaction of outflow from convective clouds with the surrounding environmental air. Individual cells or thunderstorms have a lifetime of around an hour, however many can last longer. This is due to a 'cold pool' developing where the downdraught reaches the surface, and secondary convection being triggered as warm environmental air is lifted above the more dense air. These types of storms where new cells are repeatedly triggered on one side of the system are very important as they can produce extreme localised accumulations of rainfall over a long period of time.

1.1.4. Extreme rainfall leading to flooding

Convective precipitation in the UK can range from short-lived, low intensity events, to heavy downpours of which can have significant impacts over small timescales and wide areas.

Short-period convective storms have been identified as the main cause of flash floods in the UK (Hand *et al*, 2004). Such floods are regarded as having a time to peak of less than 3 hours within a catchment of size 5-10km² (Collier, 2007). Of 50 cases of rainfall events that could cause flash floods in the UK, investigated by Hand *et al* (2004), 30 were predominantly convective and a number of the frontal cases had a significant convective element, usually characterised by embedded thunderstorms.

Flash floods occur suddenly and can have massive impacts on the affected area, generating hazards such as landslides and mud flows, causing damage to infrastructure, and put peoples' lives at risk (Collier, 2007). Over the past 60 years, there have been notable examples of extreme rainfall events culminating in flash floods. The Lynmouth storm in 1952 was a summer thunderstorm event in which 34 people died (Collier, 2007; Golding, 2009). The catastrophic results occurred as there was 228mm of rain in 12 hours (Collier, 2007), which built up behind debris that subsequently failed (Hand *et al*, 2004). The 1975 Hampstead summer thunderstorm interrupted rail and underground services in parts of London for several days (Golding, 2009), with 170.8 mm of rainfall in just 3 hours (Golding, 2007). The Boscastle flood in Cornwall on 16th August 2004 (section 1.2) caused severe flooding and associated damage to the village, with efficient responsive action preventing any loss of life. This case study shall be the main focus of this report.

An extreme rainfall event is likely to produce serious flooding, particularly when it occurs over a sensitive catchment or steep orography, or when the ground is very wet from previous rainfalls (Collier & Fox, 2000; Hand *et al*, 2004; Collier, 2007). Such extreme events are often associated with 'quasi-stationary' storms with repeated generation of convective cells in the same area (Kendon, 2010). In the case of a single, intense event, information about the exact location and intensity of rainfall, along with surface runoff properties and response of the catchment will be vital (Hand *et al*, 2004). Therefore accurate warning systems depend on the accurate real-time provision of rainfall information, the use of accurate hydrological models, and on the surface run-off features of the catchment. Numerical Weather Prediction (NWP) models need to provide the spatial resolution required to represent the substantial

variations in precipitation that can occur within a single catchment, and which may be very important for determining catchment response.

The Met Office and Environment Agency have created the Extreme Rainfall Alert launched in July 2008. This alerts emergency services to respond to the probability of rainfall sufficiently heavy to cause flooding somewhere (Golding, 2009; Werner *et al*, 2009).

The Boscastle storm case study shall be presented in the following section (1.2). The application of Numerical Weather Prediction (NWP) (section 1.3) will be discussed, specifically in relation to this event (section 1.3.4). Assessing the potential value of an ensemble approach to NWP (section 1.4.) will be the objective of this study, and will be investigated using various model outputs from the Boscastle storm.

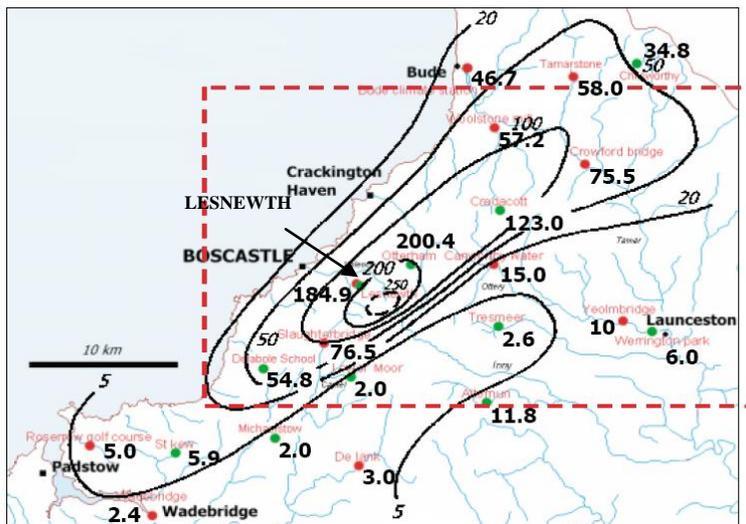
1.2. The Boscastle storm event

Spatial and temporal characteristics of the rainfall and the small-scale mechanisms that focused extreme conditions in the Boscastle area are examined in this section.

During the afternoon of 16th August 2004, prolonged heavy rainfall resulted in flash flooding in the village of Boscastle, on the north Cornwall coast. Around 60 properties were flooded (May *et al*, 2004; Golding, 2005) and a substantial number of people were rescued by the emergency services (Golding *et al*, 2005), preventing any loss of life.

Boscastle is a small Cornish village, close to the high ground of Bodmin Moor and in the valley of the Valency River (figure 2). The rocky river valley collects rainfall efficiently and rapidly adds it to the existing river flow, which is exported to the sea. The main river and its tributaries flow in steep sided valleys, with underlying strata composed mainly of shale and slate, resulting in high levels of runoff (nerc.ac.uk, 2010a). At the beginning of August, the catchment soils were drier than normal, however subsequent weather prior to 16th August was significantly wetter than average and highly variable in distribution (Golding, 2005), resulting in a soil moisture deficit of just over 100mm in the Valency catchment (Golding, 2005).

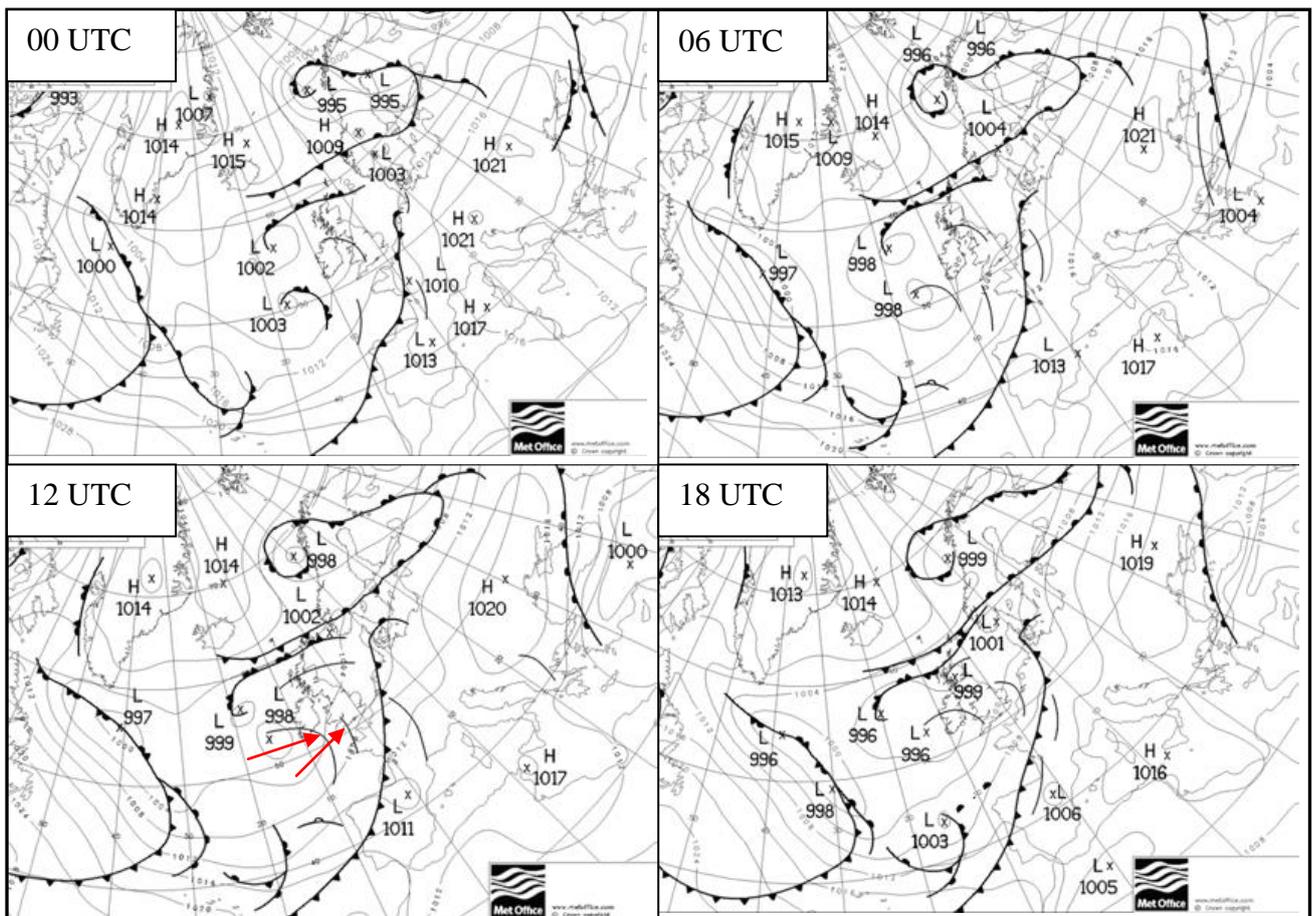
Figure 2. Map of the north Cornwall area, showing the location of Boscastle and rivers in the Valency catchment. Approximate rainfall (mm) on 16th August is shown at points and with isohyets. The dashed rectangle marks the area of intense rainfall; the green circles mark standard daily-read raingauges, and the red circles mark recording gauges (Burt, 2005).



1.2.1. Large-scale synoptic conditions

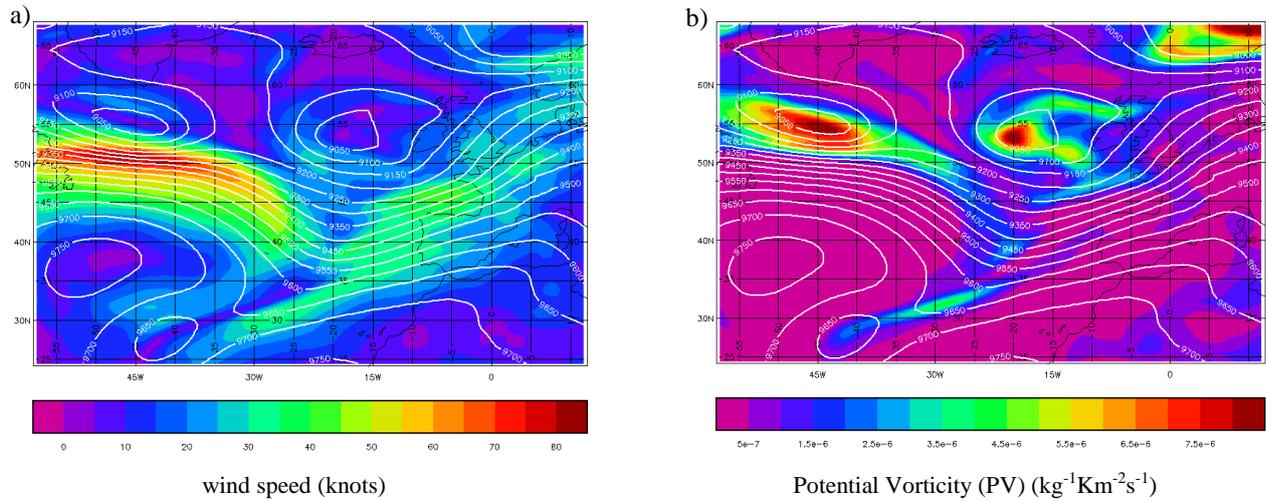
The large-scale synoptic situation over south-west England was very complex, with a slow moving low-pressure system to the west of UK (Golding *et al*, 2005) and two shear zones approaching the southwest of England (Golding, 2005), indicated on figure 3.

Figure 3. The synoptic situation on 16th August 2004 at 6 hour intervals (Golding, 2005). The red arrows indicate the two shear zones.



From 06 UTC, Boscastle was situated in warm moist air ahead of the two troughs (figure 3) and below a jet exit region (figure 4a). On the left hand side of this exit region, there is ascent and a highly active area of surface vorticity (figure 4b), so greatest potential for convective motion and therefore rainfall.

Figure 4. 300hPa height contours at 1200 UTC on 16th August 2004, with a) wind speed (colours), b) Potential Vorticity (colours) (Golding, 2005)



1.2.2. Local air mass analysis

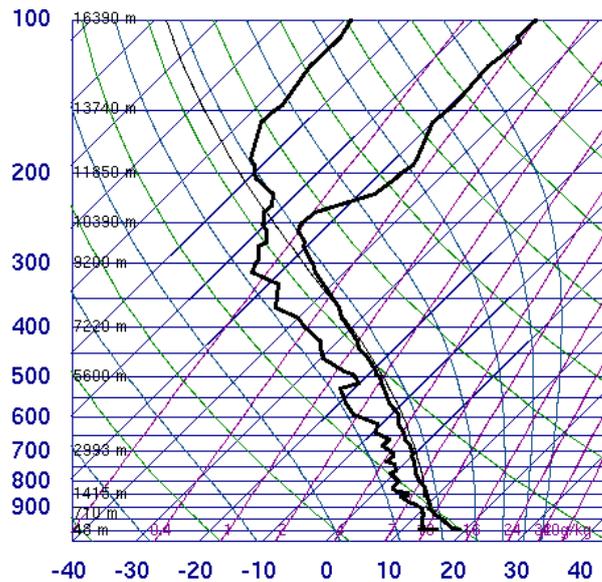
Figure 5 highlights the area of high ground (Hendraburnick Down) close to Boscastle, and marks the location of Camborne, approximately 70 km to the southwest (upwind) of Boscastle.

Figure 5. Map of southwest UK (metoffice, 2010a)



A radiosonde ascent at Camborne at 12UTC on 16th August (figure 6) shows a very unstable airmass, with low CIN (2 Jkg^{-1}) and substantial CAPE for the UK (170 Jkg^{-1}) (Golding, 2005).

Figure 6. A radiosonde ascent at Camborne at 12UTC on 16th August (weather.uwyo.edu, 2010)



Using this value for CAPE, the vertical velocity of the air parcel can be estimated.

Equation 3 calculates the velocity of ascent (w_{\max}):

$$w_{\max} = \sqrt{2 \text{ CAPE}} \quad (3)$$

$$w_{\max} = \sqrt{340} \approx 18 \text{ms}^{-1}$$

This upward motion (18ms^{-1}) is calculated for undilute convection, and the effect of entrainment has to be taken into account, which acts to slow this motion. Golding (2005) stated a mean vertical velocity of 5ms^{-1} is appropriate, and diagnosed that at this speed it would take 15 minutes for air from the boundary layer to reach the Level of Neutral Buoyancy (LNB) (section 1.1), where the parcel is in equilibrium with the environment and does not continue to ascend. Precipitable water in this air column estimated in figure 6 was 25.87mm, and this value can be used to calculate the maximum rain rate, assuming all the moisture in the air column was rained out.

$$\text{Rain rate} = \frac{\text{Precipitable water}}{\text{Lifting time}} \quad (4)$$

$$\text{Rain rate} = \frac{25.87 \text{ mm}}{0.25 \text{ hr}} \approx 100 \text{ mmhr}^{-1}$$

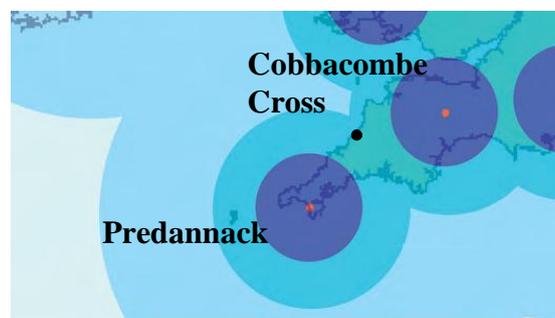
Due to the evaporation of rainfall, efficiency of a typical cloud is $\approx 50\%$, giving a rain rate of $\approx 50 \text{ mmhr}^{-1}$ in this situation. This value shall be compared to the observed rain rate in the Valency catchment in sections 1.2.3 and 1.2.4. Without a radiosonde ascent directly over

Boscastle, it is difficult to deduce how representative this radiosonde ascent over Camborne is, and this issue shall be further discussed in section 1.3.4.

1.2.3. Storm analysis using rainfall radar data

The conditions for radar observation were good on 16th August, and the Predannack radar and the Cobbacombe Cross radar both have an unobstructed view of the Boscastle area, at a range close to 100 km (Golding, 2005).

Figure 7. Radar coverage in the southwest of the UK. Circles show where the radar has resolution of 1 km (purple), 2 km (dark blue), and 5 km (pale blue) (metoffice, 2010b). The black dot indicates the location of Boscastle and the red dots show the locations of the radars

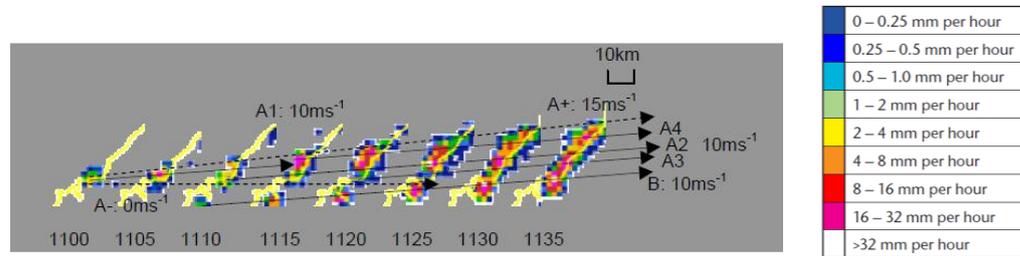


Analysis of the Predannack radar data shows the development of the first cell just before 1100 UTC, and it has been suggested (Golding, 2005) that the trigger for deep cloud convection is associated with uplift due to convergence in the vicinity of the high ground. With extremely small CIN (2Jkg^{-1}) (section 1.2.2), an uplift of 2ms^{-1} is needed to overcome this and access the CAPE. Therefore the mean vertical velocity of ascent of 5ms^{-1} as estimated in section 1.2.2 is more than enough to initiate convection. As cloud development occurs along the whole of this part of the coast, it is also suggested that a mesoscale feature associated with the land-sea contrast is significant in initiating convection (Golding, 2005). However, it is also possible that the synchronised timing is a result of upper level forcing or a gravity wave train (Golding, 2005).

This radar image (figure 8) shows closely packed storm cells along a convergence line that will be shown in section 1.3.4.1, using model data. The radar data shows the cores of these convective cells to be precipitating at a rate of at least $16\text{-}32\text{ mm hr}^{-1}$. The presence of a coastal jet and an onshore pressure gradient caused by solar heating of the land resulted in this strong convergence line, with local enhancement due to topography. The coastal jet was formed as the rough surface of the land caused the wind to back (from southwest to south-

southwest) (Golding, 2005), and over the sea air motion accelerates and turns right due to the Coriolis Effect.

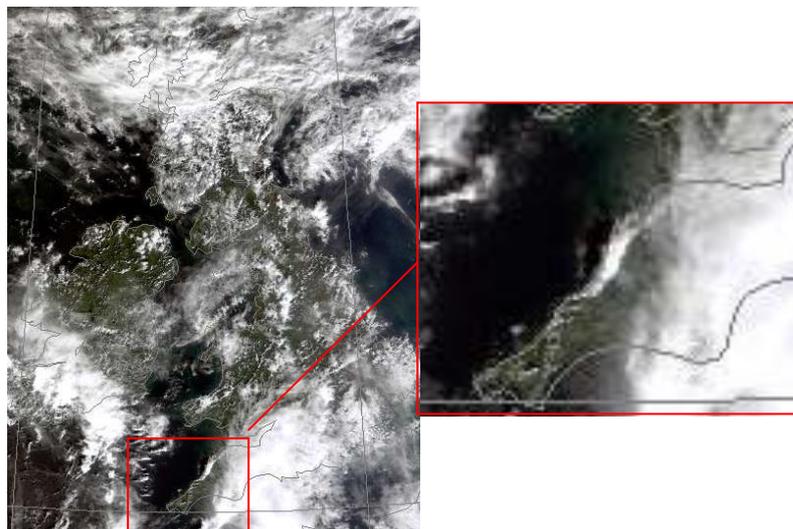
Figure 8. Radar image of the convergence line, with 1st and 2nd storm cells marked, on 5 minute 2 km radar images from the Predannack radar. Each time is shifted right by 25 km for additional clarity (Golding, 2005)



The first cell (A1) spreads out and moves northeast, with its forward edge moving at 15 ms^{-1} (A+) and its back edge almost stationary (A-). Within this area, three secondary cells develop (A2, A3, A4), which are linked to each other and the second storm cell (B). This resulted in a continuous line of precipitation, with each cell moving at 10 ms^{-1} . These multiple secondary elements within each cell are generated by the downdraught from the first cell creating new updraughts along the convergence line itself (section 1.1.3). These downdraughts are not strong enough to modify the convergence line, however as each storm cell decays, the downdraught takes over as the dominant forcing (Golding, 2005). In this case, the downdraughts do not act to destroy the convergence line, but instead, enhance the ascent. In order for this triggering to occur in the same place, downdraughts must be weak, and as they are driven by evaporative cooling, it is important that the lower troposphere should be moist.

The satellite image (figure 9), taken at 1203 UTC on the day of the Boscastle storm shows the line of cloud that was shown to be precipitating in the radar. The position of this cloud follows the convergence line, where there is uplift and continued generation of storm cells.

Figure 9. Satellite image showing line of cloud situated over the Boscastle region at 1203 UTC, 16th August 2004 (sat.dundee, 2010)



1.2.4. Storm analysis using raingauge data

Point measurement of rainfall rates on 16th August 2004 comes mainly from 5 tipping bucket raingauges (figure 2). Figure 2 shows the highest rainfall was recorded at Otterham, with 200.4 mm over the day. The highest autographic recording was at Lesnewth, with 184.9 mm recorded with a standard raingauge, where the adjacent tipping bucket raingauge recorded 155.2 mm (Burt, 2005). This highlights the variation in rainfall recorded at the same site, which may limit the accuracy and reliability of these point observations.

A smoothed profile (figure 10) created from raingauge data at Lesnewth from 12 UTC to 1630 UTC illustrates the highly variable rain rate, reaching a peak of 300 mmhr⁻¹. This period was when the rain was most intense. From approximately 1615, the storm system widened and decayed over Boscastle (Burt, 2005).

Figure 10. Rainfall intensity (mmhr⁻¹) at Lesnewth, Cornwall on 16th August 2004 (Golding, 2007)

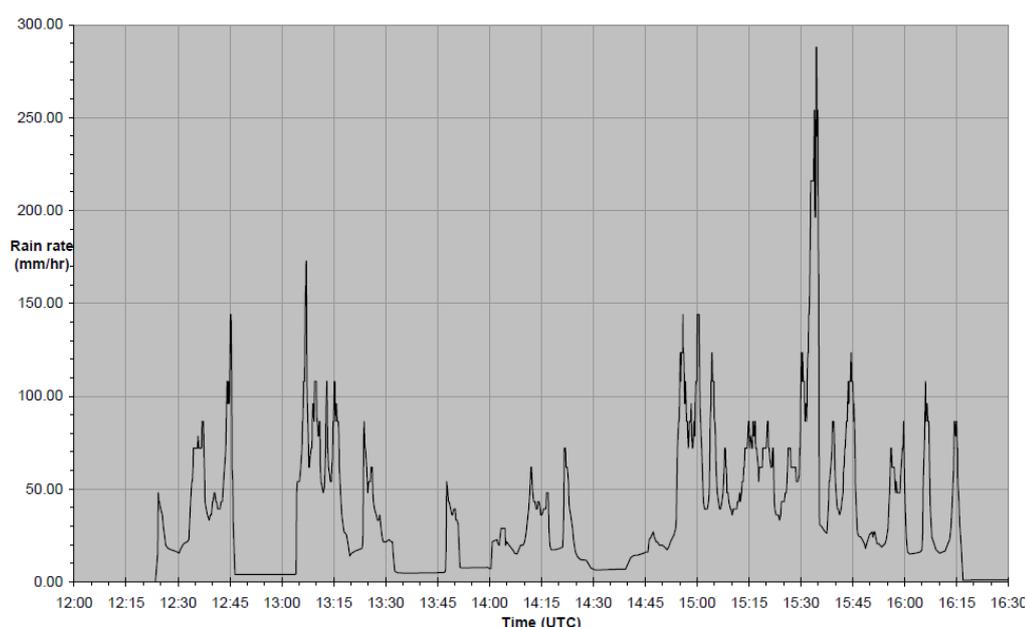


Figure 10 shows that the rain rate frequently reached over 100 mmhr⁻¹. The peak of nearly 300 mmhr⁻¹ at around 1530 UTC and rainfall rate exceeding 100 mmhr⁻¹ for 5 minutes at this time greatly exceeds the rain rate calculated in the local air mass analysis in section 1.2.2. The ‘storm efficiency factor’ at Lesnewth, which is the rainfall (155-185 mm) compared to precipitable water content (25.87 mm) is between 6 and 7 (Burt, 2005). Other 15-minute rainfall rate maxima from raingauges and radar were 80-100 mmhr⁻¹, indicating an unusually high efficiency, while hourly accumulations of up to 60 mm indicate that this efficiency was being maintained over multiple cloud lifecycles without a break (Golding, 2005). Therefore, during the storm event there must have been another source of incoming moisture, as the

extreme rainfall was not due to one storm, but due to a sequence of storms that rained in the same place. As the storms were running along the length of the convergence line, it is possible that the new storms could draw on the moisture rained out from previous storm cells. However, this is unlikely as evaporation is not a fast enough process to account for the high rain rate. A more probable explanation is that there is a convergence of moisture into the area from the surroundings.

1.2.5. Sources of uncertainty in observational data

The air mass sampled may not be representative, as it was ~ 70 km away from the location being investigated (1.2.2.), making this an unreliable source of data.

There are accuracy issues with both rainfall radar and raingauge data, especially at high rain rates. There are a number of reasons why the radar data may not accurately represent the rain amount at the ground, however due to their large sampling area and continuous coverage, they prove effective and are widely of use. The accuracy of rainfall radar measurements (section 1.2.3) can be compromised by the effects of clutter, bright band, and the angle of the radar beam being too high or too low (metoffice, 2010b). In this case, it is unlikely that the rainfall evaporated before the radar sampled the air, as it is probable that there was substantial low-level moisture (section 1.2.3). The power-law equation for converting reflectivity measurements into rainfall rates, the 'reflectivity-rainfall rate (ZR)' may have errors at high rain rates, as there is considerable variation in raindrop size distribution during a rainfall event (Alfieri *et al*, 2010) and limited high rainfall rate data to calibrate against.

The raingauge data (section 1.2.4) shows high spatiotemporal variability, even at the same site, where different types of raingauge were used. By smoothing over many raingauges, the total rates may be accurate; however accuracy at precise locations may be affected. Raingauges may also underestimate rainfall due to turbulence above the gauge, and at high rain rates, tipping gauges may underestimate rainfall during the tipping stage.

Observations from these different sources help to create a picture of the events that occurred on the day of the storm in Boscastle, however, when considering precise details, each has associated uncertainties that should be taken into account.

1.2.6. Summary of Boscastle storm meteorological conditions

- The extreme rainfall accumulations observed in the Valency catchment resulted from prolonged very heavy rain from 12-16 UTC (Golding, 2005)
- A combination of frictionally-backed winds and a developing sea-breeze resulted in a stationary sea-breeze front (Golding *et al*, 2005) made up of a quasi-continuous line of closely packed storm cells (Golding, 2005) along the north coast
- Local processes modulated the intensity of the rainfall; however there were a combination of large scale forcings, arising from the dynamical response in a jet exit region
- Convection was strong enough to generate heavy precipitation and shallow enough to enable the development of closely packed storm cells, with downdraughts weak enough not to distort the coastal convergence line (Golding, 2005, Golding *et al*, 2005)
- Intense, localised rainfall of over 200 mm in 4 hours (Burt, 2005) precipitated into one catchment that led into Boscastle, resulting in flash floods.

It has been proposed that landslips may have contributed to the flooding, and this is currently being investigated by the British Geological Survey (nerc.ac.uk, 2010a).

The severity of the rainfall during this event and the damage that it caused reinforces the need for such an event to be forecast. A forecast lead time of 24 hours could alert emergency services so that necessary action plans could be implemented, therefore investigation of forecasting techniques is a growing area of research.

1.3. Forecast skill of convective events

Improving Quantitative Precipitation Forecasting (QPF) is still one of the major challenges in Numerical Weather Prediction (NWP) (Tibaldi *et al*, 2006), despite the fact that major recent developments in computer power resources have led to a rapid increase in ability. Accurate prediction of extreme weather, in particular intense and localised precipitation, is very difficult beyond two days (Mullen & Buizza, 2001), as this phenomenon has inherently low predictability (Tibaldi *et al*, 2006), generally predictable for only up to a few hours ahead (Pedder *et al*, 2000). As illustrated with the Boscastle event (section 1.2), extreme rainfall can be generated by an interaction of processes on different scales. To produce accurate forecasts, the processes occurring at each scale need to be represented accurately. Improvements in forecast ability can lead to early warnings, used to take action to protect public property and infrastructure, and maintain a stable economy (Dance & Zou, 2010).

1.3.1. Numerical Weather Prediction (NWP)

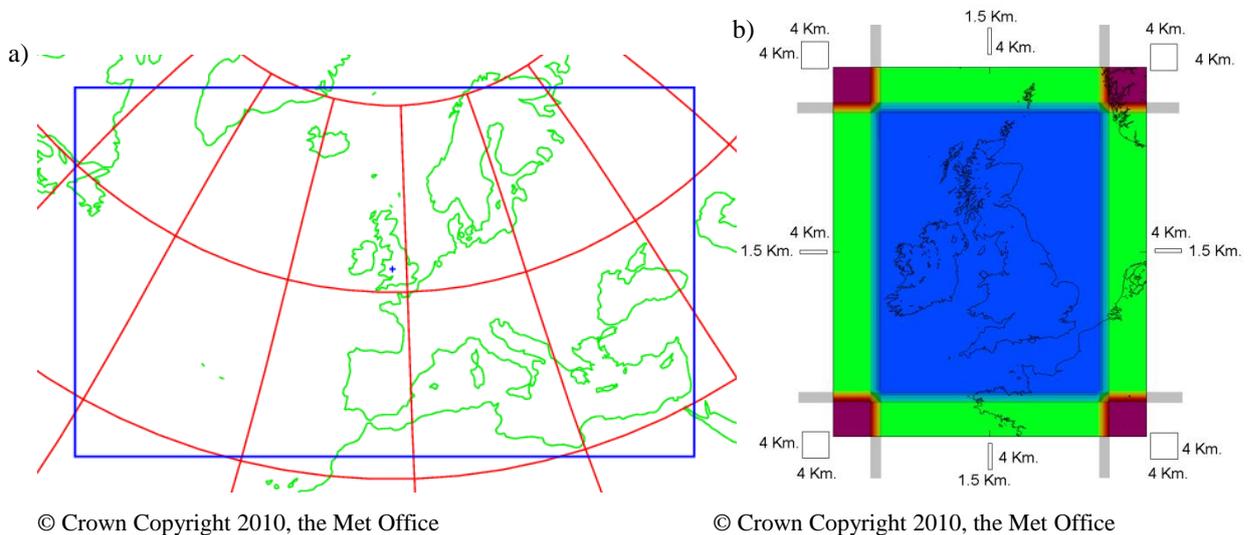
Much of the UK's most damaging weather involves convective clouds, and the Met Office operational mesoscale model (4km grid length) falls short of representing the spatial variability of precipitation produced by these features (Collier, 2007; Wetterhall *et al*, 2009). Parameterization, which represents the bulk effect of subgrid processes (Leutbecher & Palmer, 2007) deals with convective clouds. However with this 4 km grid length, it is possible that some convective clouds may be resolved, leading to an over-sizing of these phenomena. Convection schemes are incorporated into NWP's with grid spaces ≥ 4 km, and as most models until recently have had 10-20 km grid spacing, creating effective convection schemes has been a major focus. However, convection schemes are far from perfect, and often the lack of heavy rain in model output is due to the convection scheme averaging rainfall over a relatively large number of grid boxes.

To obtain a model forecast representative of precipitation variability at scales down to a particular limit, the model should have horizontal resolution that is five to six times finer than this limit. A typical UK thunderstorm is approximately 10 km in all three directions (Bryan *et al*, 2003; Golding, 2009), so a grid length of 2 km or less is needed to resolve this feature (Bousquet *et al*, 2006). In higher resolution models, it has been shown that there are improvements in the correct simulation of orographic enhancement and seeder-feeder mechanisms (Roberts *et al*, 2009; Werner *et al*, 2009), which contribute to some extreme

rainfall events. Higher resolution models are also able to make use of high resolution input data, such as detailed orography and data from radar or satellite (Lean *et al*, 2008). A series of projects have demonstrated the ability of a smaller, 1-2km grid length NWP model to reproduce realistically the structure of convective storms and to forecast them accurately when other conditions such as windspeed and humidity are adequately represented (Golding, 2009; Roberts *et al*, 2008; Lean *et al*, 2008). An important outcome of moving towards higher resolving models is turning the convection scheme off.

The Met Office Unified Model (MetUM, Davies *et al*, 2005) uses different configurations of the same model to produce all weather forecasts and climate predictions (metoffice, 2010c). The MetUM comprises a global 25 km model, a North Atlantic and European 12 km model (NAE) (figure 11a), and a 4km mesoscale model (UK4) (metoffice, 2010d), along with the global and regional ensembles and climate models. The Met Office runs its short range (1-2 days) model over the UK with a 4km grid length. As computer power increases, it is possible to reduce the grid length, and the Met Office is routinely running a ‘convective-scale NWP’ (metoffice, 2010e) with a 1.5 km grid length (Collier, 2009), in preparation for it becoming the operational model.

Figure 11 a) Domain of the NAE model, b) variable resolution at the UKV boundaries (MSc Met Office forecasting module, 2010)



The UKV has 70 vertical levels and lateral boundary conditions (LBCs) are fed in from the 12 km NAE every half hour (Clark, 2009). The 1.5 km UKV model has variable resolution near the boundaries, as the LBCs from the 12 km model are converted at the boundaries into 1.5km grid lengths around the UK. This method takes the boundaries away from the immediate UK area, whilst reducing the heavy disk space required to pass the boundary

conditions from the NAE model. This high-resolution model also has more detailed orography. The Met Office is currently developing an hourly data assimilation cycle for the 1.5 km model (metoffice, 2010f), which would be able to incorporate more detailed observations, and avoid the difficulty of dealing with boundary and initial conditions.

Roberts and Lean (2007) investigated the impact of decreasing model resolution on forecast skill, using the MetUM with 12km, and nested 4km, and 1km grid spacing. They demonstrated that the 1km model showed a great improvement in forecast skill for heavy, localised rain. The 4km model did not improve much on the 12km model, because of the problems with resolving convection at this resolution, which was accentuated by the spin-up from the 12km model fields. They diagnosed that it is reasonable to run the 1km model without a convection parametrisation scheme, but it is less clear whether one is needed for a 4km model (Roberts & Lean, 2007).

1.3.1.1. Model forecast uncertainty

There are three main sources of error in weather forecast models: input data, parametrisations, and model structure (Dance & Zou, 2010). Observations are vital to reducing initialisation errors, and parametrisations are constantly being developed to reduce the error that they may contribute to forecasts. In the UKV model, small-scale features will be resolved, and there will not be the need for convective parameterisations. Other parametrisation schemes will still be required, e.g. boundary layer scheme, although it is debatable whether this should work differently as the resolution increases. Parametrisations are not a precise representation of processes, and therefore have associated uncertainties.

Models with a limited horizontal extent incorporate data from more widespread models at the boundaries, which may be a source of error. Also, using the nested approach, where initial conditions are cascaded down into smaller grid lengths may limit the accuracy of the models. The interaction of scales within convection causes a problem in modelling them, as systems on all scales need to be modelled accurately.

1.3.2. Nowcasting

This is very short-range forecasting that maps the current weather and uses an estimate of its speed and direction of movement, extrapolated forward using estimated winds, to forecast the

weather for a short period ahead (metoffice, 2010g). Nowcasting extrapolates the current weather and its recent trends in such a way as to preserve the most significant features (Pierce *et al*, 2000). In the UK, rainfall nowcasts can be useful up to three or four hours ahead for widespread rain bands in winter, but only one to two hours ahead for thunderstorms (metoffice, 2010g). To extend the period of predictability, nowcasts can be combined with output from NWP models (metoffice, 2010g). The nowcast model applies corrections to the NWP forecast, to deal with orographic enhancement and effects of wind speed; forecasts of which are derived from the Met Office mesoscale NWP model. A blending procedure is also applied at the start of the forecasts with the radar observations, with NWP forecast precipitation amounts having more and more weighting towards the end of the forecast (Werner & Cranston, 2009; Golding, 2000).

The Nimrod Nowcasting system utilised by the Met Office is an automated precipitation nowcasting system, which uses rainfall radar data (from the UK weather radar network), weather satellite observations and Met Office mesoscale model output (Pierce *et al*, 2000) to generate hourly nowcasts with a lead time of 6 hours (Werner & Cranston, 2009). A Probability Distributed Model (PDM) for rainfall-runoff modelling and forecasting is incorporated into MOSES (Met Office Surface Exchange Scheme) for use in the Nimrod nowcasting system. (Roberts *et al*, 2009). Nimrod outperforms the mesoscale model at short lead time (1-4 hours), as the model takes time to spin up high-resolution features. Thereafter, the skill of the mesoscale model matches that of Nimrod, and it has been shown that for lead times beyond a few hours, the mesoscale NWP model precipitation output provides the most reliable forecasts, which also include the development of new rain areas (Collier, 2007).

A major limitation of the Met Office Nimrod system is the way it handles convective initiation and development. To improve the representation of convective storms, GANDOLF was developed (Collier, 2007). This is an automated convective rainfall nowcasting and early warning system with the best available very short-range forecast of rainfall intensity at 2km resolution, which runs during episodes of air mass convection (Pierce *et al*, 2000).

The Met Office also runs STEPS (Short Term Ensemble Prediction System), an ensemble nowcasting system in which the rainfall distribution is separated into different sizes of rainfall feature, so large rainfall events (more predictable) can be nowcast for longer, and small events are only nowcast for a very short time (metoffice, 2010g). STEPS is a stochastic precipitation nowcasting system, which merges an extrapolation nowcast with a downscaled NWP forecast (Bowler *et al*, 2006). This ensemble system captures the uncertainty in the advection velocity and also in the evolution of the precipitation pattern. The advection

velocity is dealt with by using a random field of velocities, and the evolution is modelled using S-PROG (Spectral Prognosis model), which has a multiplicative cascade approach where small-scale features of the forecast precipitation pattern have been randomly generated (Bowler *et al*, 2006). Each ensemble member has a different random cascade of precipitation pattern and a different field of advection velocities. This ensemble approach has a 2 km resolution and provides probabilities of rainfall propagation up to a 6 hour lead time, enabling the location of extreme rainfall to be pinpointed several hours before it occurs. Beyond this predictability limit, information is used from the NWP model.

Nowcast extrapolations contribute positively to the skill of forecasts, despite the uncertainties they contain (Werner & Cranston, 2009). Radar rainfall nowcasts such as those provided by the Met Office Nimrod system provide short range predictions at these small spatial scales and can be used as an input to hydrological models for the prediction of flood flows (Werner & Cranston, 2009). The uncertainty of such forecasts influences the reliability of hydrological forecasts used in flood warning and forecasting (Werner & Cranston, 2009).

1.3.3. Hydrological models

Antecedent conditions of the catchment are critical to diagnosing whether rainfall will result in a flash flood, as the ability of the soil to absorb rainfall during an event depends on the moisture content already in the soil, which is affected by preceding rainfall events and evaporation. Soil moisture affects the Bowen Ratio, which influences how energy is divided between latent heating and sensible heating. This may have an influence on the initiation and properties of convection, so it is a useful input variable for inclusion into the NWP model. Another key aspect to river hydrograph forecasts are their high dependency upon the rainfall data being very accurate at a specific time (Collier, 2007).

The added value of using radar rainfall measurements and nowcasts has been recognised by the hydrological forecasting community (Werner & Cranston, 2009), and a vision for the future is that automated end-to-end systems that feed high-resolution NWP rainfall forecasts into hydrological models will become a standard part of the flood warning procedure (Roberts *et al*, 2009).

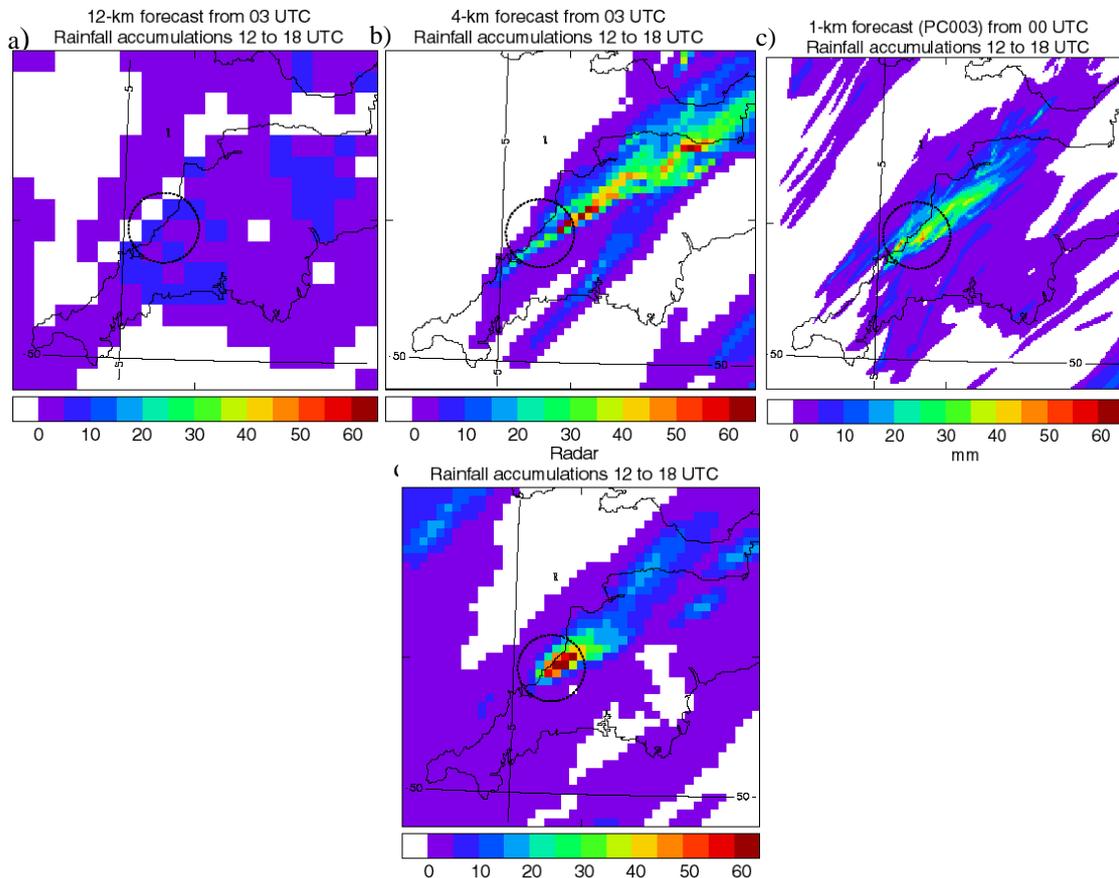
1.3.4. The Boscastle event

This section analyses the forecast skill of this event, with a discussion of literature published by the Met Office (1.3.4.1), and also from examination of model output from the 1 km model (1.3.4.2).

1.3.4.1 Analysis of event from literature

The Boscastle storm event as detailed in section 1.2. was not well forecast by the Met Office. The operational model had a grid length of 12km, which could accurately represent features larger than approximately 60km. Therefore the convection was not represented explicitly, but was encompassed within a parameterisation scheme (Gregory & Rowntree, 1990). In order to investigate the processes that led to this extreme and localised rainfall, the NWP model has been run at higher resolutions. Models with grid lengths of 4km and 1km were used, nested within the 12 km model, with the 1km model also run at double vertical resolution, to resolve better the boundary layer processes (Golding *et al*, 2005).

Figure 12. NWP model output and radar observations of rainfall accumulations (mm) from 1200-1800 UTC a)12km b) 4km c) 1km d) 5 km radar. The circle is a 20km radius from Boscastle (May *et al*, 2004)



The 12km forecast (figure 12a), which was operational at the time of the storm event does not capture the local intensity of the rainfall. The 4km model simulates a convergence line, but creates intense rainfall that is too widespread. This may be attributed to the effect of some deep convection being resolved, whilst also encompassed within the convection parametrisation scheme (1.3.1). The 1km model run captures the event the best, with the local intensity at Boscastle clearly defined. However, the maximum intensity is not reached, and the confined area of intense rainfall is not modelled. The model shows peak accumulation of ~ 30 mm, compared to ~ 60 mm (observed), and spreads the area of intense rainfall too far northeast. The radar shows a maximum 60 mm accumulation over Boscastle, and this has uncertainties associated with it, which have been discussed in section 1.2.5.

The Met Office employs nowcasting techniques to map rainfall, with GANDOLF as the most advanced system as it includes the convective cloud lifecycle. This nowcasting system is good at moving rain along once it has started, but is not effective at triggering storms, so this repeated generation seen in the observations is not predicted. GANDOLF issued heavy rainfall warnings from 1550 UTC onwards and predicted hourly accumulations in excess of 32mm, triggered by the heavy rain advected down wind of Boscastle (May *et al*, 2004). This warning was of little value, as the period of heaviest rain had already started. A 1km forecast would be of use, as it could give a warning from 00 UTC (figure 12c), allowing more time to react and prepare for the forecast event.

Figure 13. 6 hour accumulations for the Valency catchment on 16th August 2004. 2 km resolution radar accumulations from the Cobbacombe Cross radar (figure 7) are compared with GANDOLF, Nimrod and a 50 member ensemble mean accumulation from STEPS (May *et al*, 2004)

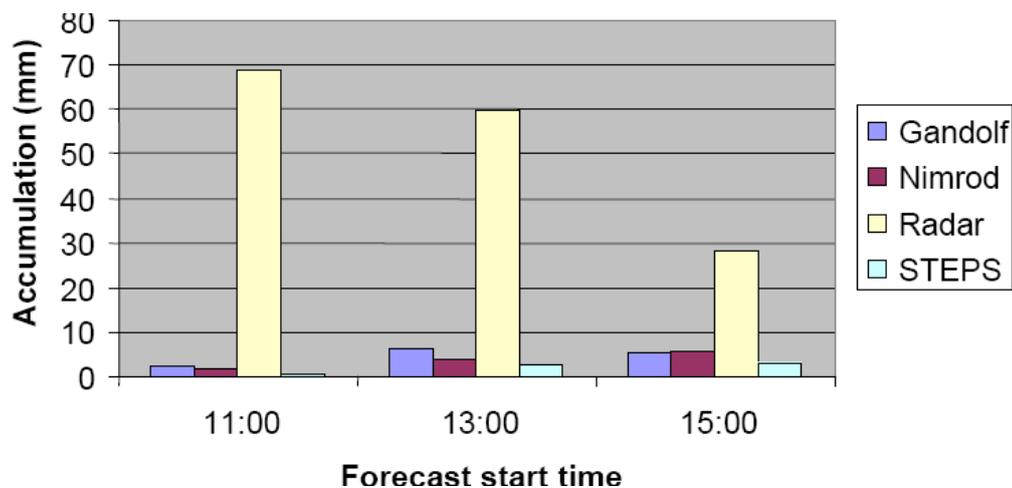


Figure 13 shows that all three nowcasting systems fall short of representing the extreme rainfall accumulation that was recorded on the day. GANDOLF outperforms Nimrod as

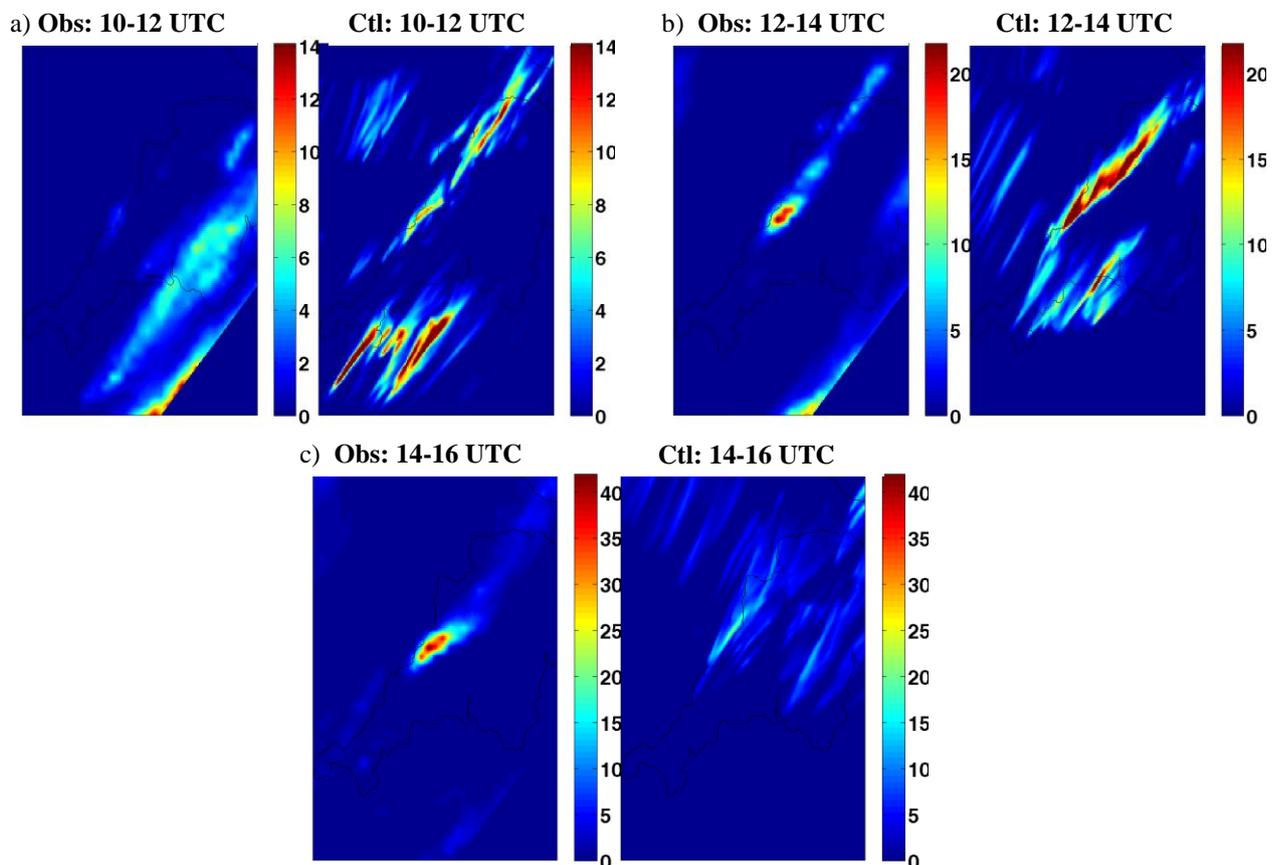
expected, as it includes more detail of the initiation and lifecycle of a convective system. The STEPS system, even with 50 members with different combinations of velocity and precipitation pattern, underestimates the rainfall accumulation more than the other nowcasting systems.

Having discussed analysis of the event from the literature, the following section examines output from the 1km control model of this event. Figures 14 - 17 are output from this 1km control run. Matlab and Jview were used to create some of these plots and some were adapted from a previous study.

1.3.4.1 Analysis of event from 1 km model runs

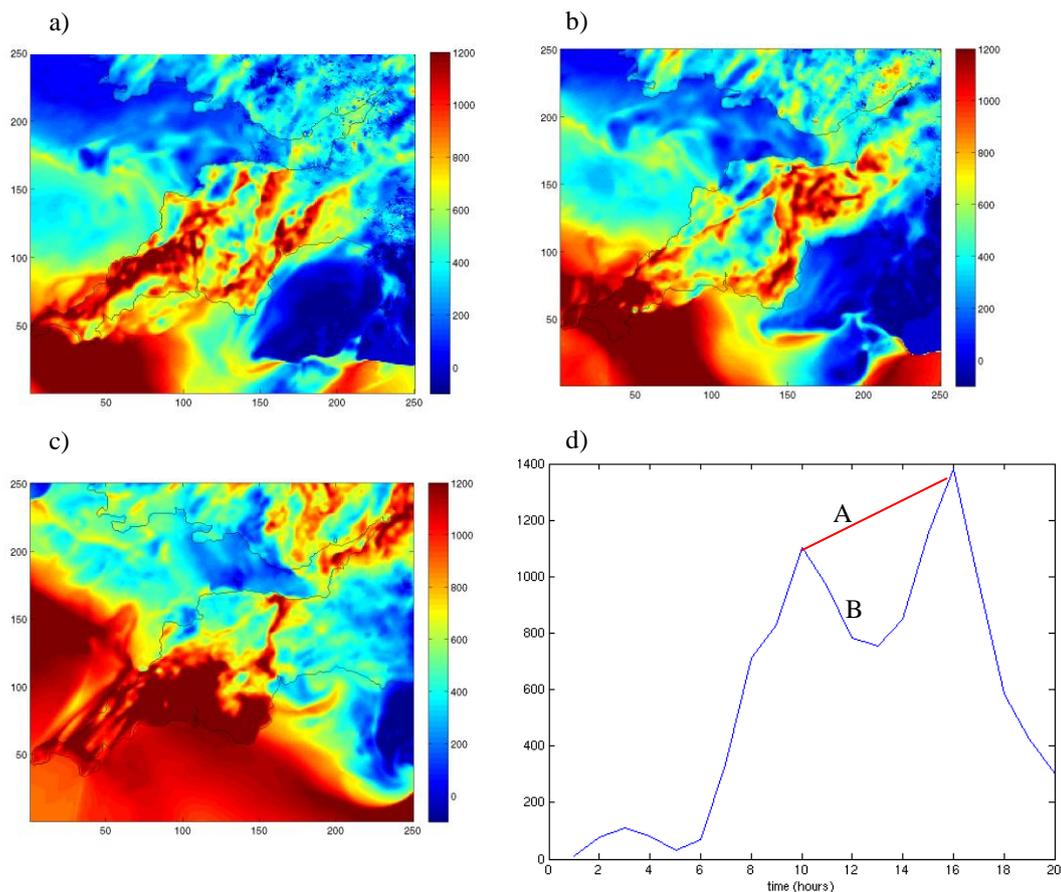
Figure 14 shows that even though the 1km forecast is a large improvement in skill on those run at coarser resolutions (figure 12), the spatial accuracy shows significant variations through different time periods.

Figure 14. Comparison of 2 hour rainfall accumulations (mm) of radar observations ('Obs') and 1km model output ('Ctl') at various times throughout the event. The radar observations come from observations on the 5 km national grid, converted onto a 1km grid for presentation here



Observations show intense rainfall centred over the Boscastle area for at least 4 hours, whereas the model only diagnoses intense convection in the correct location for approximately 2 hours, dissipating prematurely. The intense rainfall began to decay at approximately 16 UTC, however, the control run calculated its decay hours before this, as only ~10 mm rain was shown to fall from 14 – 16 UTC (figure 15c). The control run initiates the convection early and creates storms that are too widespread. This control run also falls short of modelling the generation of new cells in the same place, over the Boscastle region. The error in this model precipitation output is due to the model misdiagnosing different processes. A brief examination of the model fields of CAPE, water paths, divergence and wind, follows, in order to attempt to suggest how well the model performed in other areas, and how errors in these fields may have created error in the rainfall.

Figure 15. CAPE (Jkg^{-1}) development in the 1km model at a) 12 UTC, b) 13 UTC, c) 15 UTC d) over Boscastle, 02-20 UTC



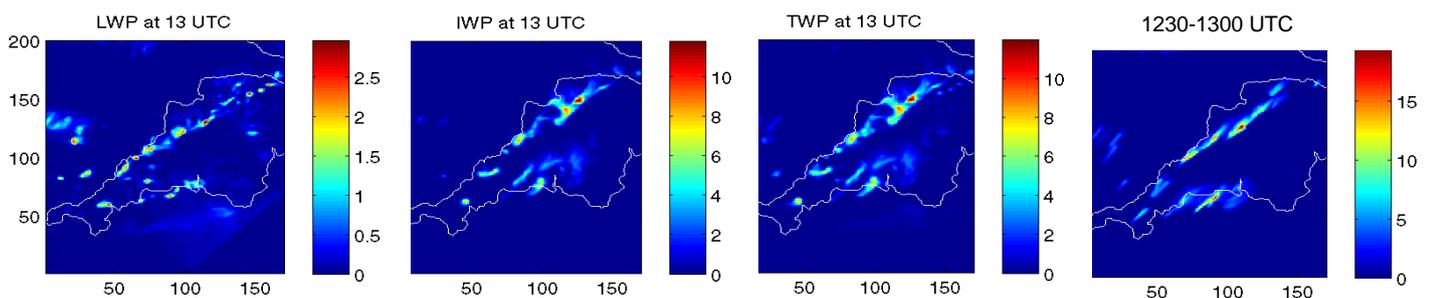
At 1203 UTC, a radiosonde sounding was taken (section 1.2.2), indicating CAPE of 170 Jkg^{-1} in the air column above Camborne (location in figure 5). From figure 15a, it is evident that

the model generated CAPE over Camborne approximately 6 times greater ($\sim 1000 \text{ Jkg}^{-1}$) than that estimated from the sounding. At this time, the model produced a similar value of CAPE over Boscastle. These high values of CAPE in the model could be attributed to the model being too quick to build CAPE, or perhaps the sounding was taken through a cloud, creating an inaccurate representation of the clear air. The higher values of CAPE spreading northeast are associated with an incoming air mass. Changes in CAPE over Boscastle with time (figure 15d), illustrate the movement of this incoming air mass, and also the effect of the deep convection, which produces extreme rainfall. Line A marks a possible profile if the air mass had moved over Boscastle and no convection occurred. B marks the effect of convection, eroding the CAPE. Even though the exact values of CAPE output by the model do not match those in the observations, it models the change in CAPE, which represent the convection around the correct time period.

The distribution of CAPE is patchy over land in the model, and this precise positioning is unlikely to be totally accurate. When the model overestimates CAPE, the effect on the rainfall amount is that storms that occur will be too intense, which is seen in figures 14a and 14b.

The relative quantities of the Liquid Water Path (LWP) and the Ice Water Path (IWP) to create the Total Water Path (TWP) within the model have different effects on rainfall diagnosed by the model.

Figure 16. Water Paths at 1300 UTC. LWP, IWP and TWP are instantaneous ($10^{-4} \text{ kg water/kg air}$), 30 minute rainfall accumulation (mm)



From 1230-1300 UTC, the model outputs rain in the correct place and has a high proportion of ice in the cloud, which is needed for deep convection. The model also creates some artificial rain from more shallow liquid clouds on the south coast.

The observations do not indicate what the relative contributions of liquid water and ice should be; therefore it is difficult to compare these model values to observations. Model simulation

of the contribution of liquid water and ice to the total water path has a significant influence on the intensity of the rainfall. This motivates experimentation with the model to change these relative contributions, and see the effect on the rainfall.

Figure 17. u and v wind at a) 13 UTC, b) 15 UTC. The black dot marks the location of Boscastle

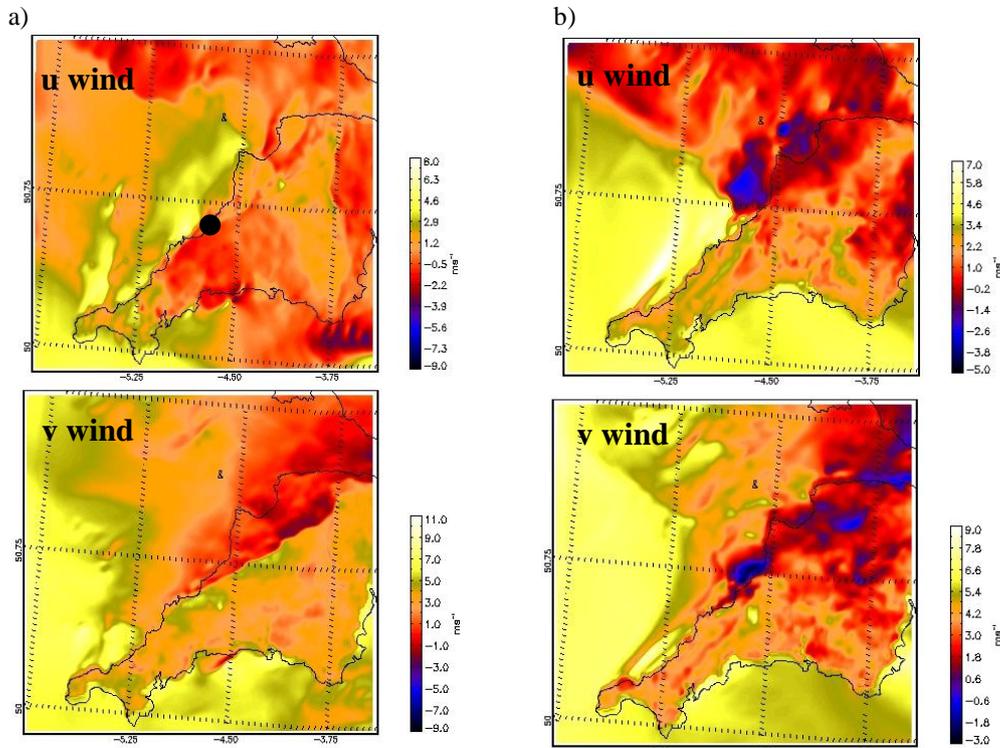
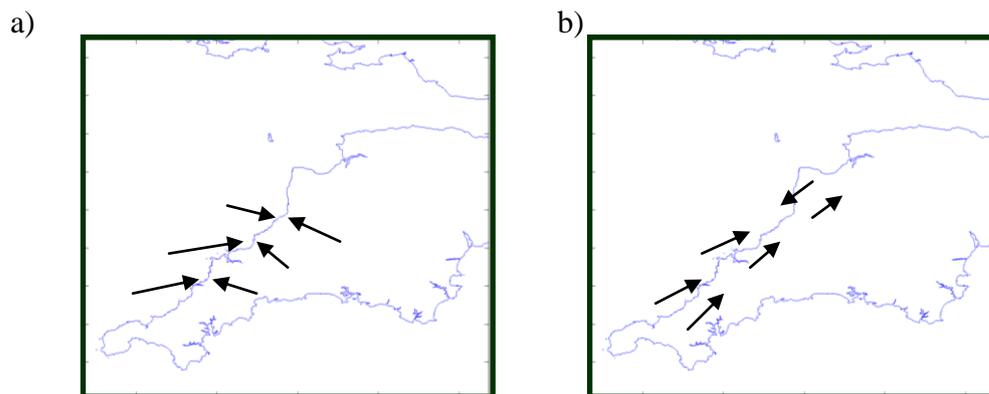


Figure 18. Sketches of wind direction from analysis of figure 20. a) 13 UTC, b) 15 UTC



Analysis of the u and v wind data created by the 1km model illustrates the strong convergence line at 1300, which weakens and disappears over the next 2 hours. This convergence line and related ascent were strong enough to overcome the CIN and release the CAPE, leading to upper level divergence, and rainfall. At 15 UTC, the convergence line and upper level divergence weakened and by 17 UTC, there was no low-level convergence or upper-level

divergence over Boscastle. It is apparent that wind is important for this convergence line and related ascent, and details of this may be important to get rainfall right, therefore it will be useful to run simulations that make convergence line different.

This Boscastle storm is the type of event that would be hoped to be forecast with a high resolution ensemble. This storm created widespread damage, and so an early and accurate warning of its intensity and location is vital to avoid damage to property and life. This is therefore a useful event to use for the rest of this study.

The past lack of skill in forecasting these extreme convective events ahead of their initialisation has prompted the development of methods for forecasting these more skilfully. The use of ensembles with low resolution models at the medium range has proved effective, and it is hoped that this approach will also add value to short-range convective scale forecasts.

1.4. Importance of ensemble forecasting

Weather forecasting is primarily an initial value problem, in that small differences in initial conditions can result in large differences in forecast outcome. Deterministic forecasts of extreme rainfall are limited to a few hours ahead by the chaotic nature of the atmosphere. Developments in probabilistic forecasting over the past decade offer the possibility of useful information at more extended forecast ranges (Golding, 2009).

Ensemble forecasting provides a range of possible solutions, whose average is generally more accurate than the single deterministic forecast, and whose spread gives information about the forecast uncertainty (Kalnay *et al*, 2006). The ensemble mean is generally more accurate than any single forecast because the unpredictable scales of motion have been filtered out in the ensemble mean and only the signal of the predictable scales remain (Leutbecher & Palmer, 2007). Ensemble spread, defined as the standard deviation of ensemble members relative to the ensemble mean is a measure of forecast variability (Walser & Schar, 2003; Kong *et al*, 2007). Ensemble forecasts can be used to distinguish and forecast an extreme event, enabling warnings to be issued with an understood level of confidence (Golding, 2009). The predictability of precipitation amounts differs strongly depending upon the weather type and the spatiotemporal scales considered (Walser *et al*, 2004).

Probability forecasts of weather and climate have greater potential economic value than corresponding single deterministic forecasts with uncertain accuracy (Palmer, 2000), and are a valuable tool for decision making, implementing a probabilistic risk-based decision making approach (Ryder, 2009). The model rarely outputs a severe weather event as the most likely scenario. However even a small probability of such an event occurring is valuable to the forecasters, as an early warning will help to monitor the situation when the weather comes closer (Lalurette & van de Grijn, 2006). It is possible to define threshold probabilities for prescribed events at which specific actions should be initiated, and this can be as low as 20% in extreme circumstances, when the result of not mitigating may be extremely costly or politically unacceptable (Golding, 2009). These low probability, high impact events can have significant negative economic and social repercussions, therefore there is a need for accurate forecasts.

Ensemble forecasting uses forecasts from slightly different initial conditions, different models, the same model initialised at different times, and/or the use of different physics options within the same or multiple models (Kong *et al*, 2006). Ideally, ensemble members

should span the entire space of possible solutions, but in practice the atmosphere has too many degrees of freedom (Walser & Schar, 2003). Combining probabilistic and deterministic forecasts, given by the control or ensemble mean, makes it possible to assess the predictive skill of the deterministic forecast itself (Molteni *et al*, 2001).

To reduce biases, any ensemble should be the same resolution as the deterministic forecast (Clark, 2009), however the spatial resolution of ensembles typically lags deterministic forecasts (Golding, 2009). To run a large ensemble at the same resolution as the deterministic forecast is expensive. However there may be benefits in using a small ensemble, perhaps combined with lagged ensembles (Clark, 2009).

1.4.1. Medium-range ensembles

Ensemble prediction is an established technique in medium-range prediction, having been operational for over 15 years (Leutbecher & Palmer, 2007). Ensemble methods for medium-range forecasting have shown to have value, addressing the issue of limited predictability within 6-10 days lead time. Medium-range ensembles for the mid-latitudes are mainly concerned with the evolution of baroclinic perturbations to the initial conditions (Walser *et al*, 2004), but model physics perturbations at the medium-range are also useful.

The Met Office operational medium-range ensemble forecasts are provided by the European Centre for Medium-range Weather Forecasting (ECMWF); a well-established ensemble forecasting centre. The Met Office currently uses the Met Office Global and Regional Ensemble Prediction System (MOGREPS) to produce experimental medium-range forecasts (metoffice, 2010h).

1.4.2. Short-range ensembles

In recent years, the regional MOGREPS has been reduced from a 25km to a 18km grid length (metoffice, 2010d). This ensemble system produces uncertainty information for short-range forecasts up to two days ahead, giving earlier and more reliable warnings of extreme events, such as rapid storm development. This is a 24 member ensemble for the NAE model, where the members are obtained by perturbing both the initial conditions and some aspects of the physics schemes within the model (metoffice, 2010d). The Met Office want to expand their ensemble system to produce forecasts at the convective scale for 6-36 hours ahead. A

demonstration system of this type is currently under development in the Joint Centre for Mesoscale Meteorology (JCMM), a collaborative research centre. For lead times of less than 6 hours, nowcasts are still the most accurate (section 1.3.2).

1.4.3. Convective-scale ensembles

Predictability of the atmosphere at the convective scale is less than that at the synoptic scale (Leutbecher & Palmer, 2007, Leoncini *et al*, 2010). During research on the feasibility of ensemble prediction at convective scales, Leoncini *et al* (2010) concluded that moist convection and nonlinearities in general strongly favour rapid error growth. Work from Hohenegger & Schar (2007) shows error growth rates are about 10 times larger on cloud-resolving scales, where perturbation growth primarily relies upon convective (instead of baroclinic) instabilities (Hohenegger & Schar, 2007). With forecast errors growing more rapidly at these scales, it is important to avoid taking literally deterministic information on scales that are expected to be unpredictable for the forecast lead time, and for that reason a probabilistic approach is more desirable (Roberts & Lean, 2007). This inherent reduction in predictability at these small-scales, with the significant nonlinearities of the atmosphere at small scales results in poor convective-scale predictability (Leoncini *et al*, 2010). The use of forecast ensembles for rainfall prediction with high-resolution models and the subsequent flow forecasts may limit the impact of errors (Collier, 2007).

These high resolution ensembles are new and still in the development stage, and there have been a number of projects to test their skill.

1.4.4. Exploratory studies of ensembles

Results from Molteni *et al* (2001), investigating very intense rainfall leading to localised floods in the Alpine region, show that smallest precipitation errors are generally achieved by the high resolution forecast, and that a six-member high-resolution configuration based on only five Representative Members (RMs) and the control, can be a good substitute for a full 51-member high-resolution ensemble. These results show evidence that using a Targeted Ensemble Prediction System (TEPS) could be beneficial, which would reduce the computer power needed to process a large ensemble. Walser *et al* (2004) showed that forecast skill based on the precipitation field is only weakly sensitive to ensemble size, provided that at

least six ensemble members were used. This question of number of members required shall be explored in the results section (section 3).

Kong *et al* (2006) found that for 24km, 6km, and 3km grid length, ensembles showed both qualitative and quantitative improvement relative to their respective deterministic control forecasts of a tornadic thunderstorm system. The 6km grid spacing has the problem that clouds are neither implicitly nor explicitly represented, and the model vastly overpredicted precipitation, as it generated 'grid point storms'. In the 3km model, the assimilation of Doppler radar and the use of shorter forecast lead times improved ensemble precipitation forecasts. However, even at longer lead times and in certain situations without assimilated radar, the ensembles captured storm-scale features when the deterministic control forecast did not (Kong *et al*, 2007)

Walser *et al* (2004) found that the predictability of convective summer precipitation is lower in regions of flat terrain. Walser & Schar (2003) showed that complex topography increases the predictability of quantitative precipitation, due to topographic control of both the convection triggering mechanism and the larger-scale uplift on the Alpine slopes.

Even a perfect synoptic forcing and a perfect model would not necessarily allow skilful deterministic short-range forecasts, as growing small-scale uncertainties and nonlinear atmospheric interactions may quickly disrupt predictability (Walser *et al*, 2004). Once the Met Office is running the UKV 1.5km model operationally, within 3-4 years they aim to have ensembles of these convective-scale forecasts.

Therefore the use of ensembles for extreme events has been widely noted, providing a range of solutions and a probability value, which can be used by decision makers to take appropriate action. However, they are expensive and take longer to run, therefore it is important to run members that add value to the forecast, prompting the use of a targeted members approach.

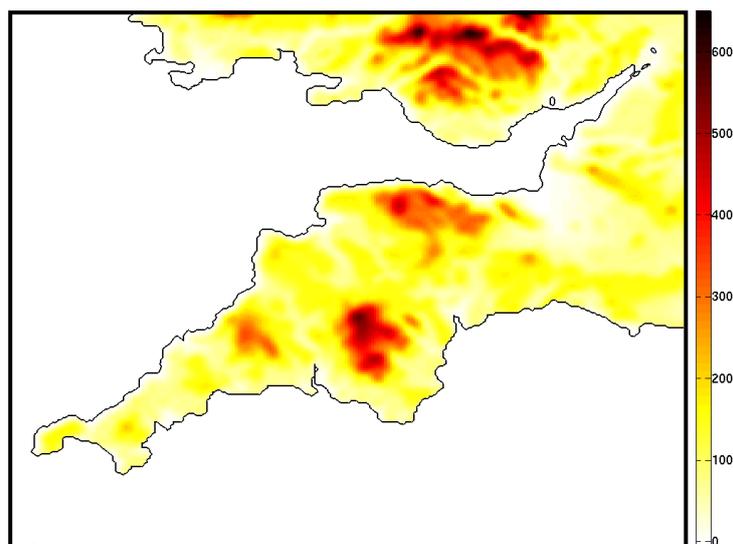
2. Model configuration

The Met Office Unified Model (MetUM), version 6.1 was used to create the ensemble forecasts used in this study. This is a non-hydrostatic, deep-atmosphere dynamics model using a semi-implicit, semi-Lagrangian numerical scheme (Leoncini *et al*, 2010). For a detailed description of the standard UM dynamics, see Davies *et al* (2005).

2.1. Resolution

The model has a grid length of 1 km, with 200 x 170 grid points, and a 38 level staggered vertical grid, closely packed near the surface and increasingly spaced higher up. Deep convection should be resolved, as typically this feature is more than 5 times the model grid length (section 1.3.1).

Figure 19. The domain of the 1 km model, with orographic detail (m) (courtesy of G. Leoncini)



2.2. Boundary Conditions

The fields required at the lower boundary are known as ancillary fields, and include: land/sea mask, soil type, vegetation type, grid-box mean and variance of orography, sea-surface temperature, proportion of sea-ice cover, sea-ice thickness, sea surface currents. As this model is run with a limited horizontal extent, lateral boundary conditions (LBCs) must also be specified. The 1km model used receives its boundary conditions from the NAE model, which is where the data assimilation process occurs. For up to date data for the NAE model, the global model must be run first.

2.3. Parameterisation scheme

In most operational models, many physical processes occur on a scale smaller than that which can be resolved by the model. As these processes are not explicitly resolved, they must be represented by parameterisation, which simulates their overall effect on the resolved scale. However, a principal aim of increasing resolution is to reduce the need for such schemes, and in the 1km model, deep and shallow convection is explicitly resolved, not parameterised. Other processes that are encompassed within parameterisation schemes include radiative processes and boundary layer processes.

2.4. Data Assimilation

The addition of observational data occurs in the NAE (12 km) model. These initial conditions are cascaded down through 4km and then to 1km grid length. With a full forecasting model operating at a 1km grid length, there would be the advantage of including extra high-resolution observations, bringing the model starting point further in line with reality.

For detailed information on this model setup, refer to Leoncini *et al* (2010), who produced the forecasts analysed in this paper.

2.5. Control run

The control run refers to the standard model run, without any changes to the physics schemes or potential temperature (section 2.6). The model run is initialised at 0100 UTC on the 16th August, and within the first few hours spin-up occurs and small-scale features are generated. The model produces forecasts to 18 UTC. Changes to the control run, with regards to the physics scheme and /or the potential temperature, as detailed below, create the ensembles that shall be investigated. The control forecast is perturbed in order to ascertain if realistic changes to the model create significant changes to the precipitation output (1.3.4.2).

2.6. Perturbation structure

Two types of perturbation are used in the 1 km model, to reflect uncertainty in model evolution, which principally arises from model error and initial condition error. Changes to the model physics will be implemented to account for uncertainties in the parametrised processes. Perturbations to the potential temperature field during the forecast will be applied, representing observational uncertainty. Initial conditions are not altered within the production of these ensembles.

2.6.1. Model physics perturbations

The following physics perturbations will affect the processes that appear to be vital to forecasting this extreme rainfall over Boscastle, such as the microphysics scheme and mechanisms that generate the convergence line (section 1.3.4.2).

Ten physics perturbations are considered, affecting aerosol concentration, roughness length, freezing temperature, rain production and soil moisture.

The microphysics scheme, detailed in Wilson & Ballard (1999), is perturbed, as liquid and ice contributions seem to be important (section 1.3.4.2).

- ‘*Aero 3D to 3B*’ uses the aerosol scheme previous to that currently in use. The 3B scheme allows for only one sort of ice, whereas the 3D scheme incorporates the existence of both ice crystals and snow aggregates. All other simulations use the current 3D scheme

- ‘*Nland*’ uses the typical aerosol concentration over land everywhere in the domain, which is approximately $3.0 \times 10^8 \text{m}^{-3}$ (Leoncini, 2009)
- ‘*NSea*’ assumes that the aerosol amount everywhere is the same as a typical concentration found over the sea, which is approximately $1.0 \times 10^8 \text{m}^{-3}$. The sea and land areas are approximately equal within the domain
- ‘*tnuc -15C*’ changes the -10^0C threshold in the control, where cloud droplets are set to freeze, to allow them to freeze at -15^0C
- ‘*tnuc -5C*’ allows cloud droplets to freeze at -5^0C .

Changes to the roughness length could affect the convergence line that is due to coastal effects (section 1.3.4.2).

- ‘*Rough*2*’ doubles the conventional roughness length over land considered in the control
- ‘*Rough/2*’ halves the roughness length over land

Altering the soil moisture affects the Bowen ratio, which may influence convection.

- ‘*SM up 20%*’ increases the soil moisture is by 20%
- ‘*SM down 20%*’ decreases the soil moisture is by 20%

A dramatic change to the microphysics scheme, preventing the occurrence of a process that occurs in the real atmosphere.

- ‘*No Auto*’, the no autoconversion scheme inhibits the production of warm rain (rain produced without ice processes).

The ‘*Aero 3D to 3B*’ scheme doubles the aerosol concentration over land, and increases aerosol concentration over the sea by 50%, compared to the control (table 1).

Table 1. Aerosol concentrations used in the different schemes (Leoncini, 2009)

Scheme	Land	Sea
3B	$6.0 \times 10^8 \text{m}^{-3}$	$1.5 \times 10^8 \text{m}^{-3}$
3D	$3.0 \times 10^8 \text{m}^{-3}$	$1.0 \times 10^8 \text{m}^{-3}$

The effect of an increase aerosol concentration is an increased number of Cloud Condensation Nuclei (CCN), resulting in more droplets, which are smaller in size. The effects of aerosol on precipitation can act to either increase or decrease total amount of rainfall as it may delay precipitation, creating a large supply to be rained out, or may totally suppress precipitation. The effect of changing aerosol concentration depends strongly on other conditions. The

MetUM handles the effects of aerosol in a basic manner, with one size of aerosol assumed. Boscastle is close to the coast, with prevailing winds coming from the southwest, over the sea, so aerosol concentrations over Boscastle are more likely to be like that over the sea. Therefore, with fewer, larger droplets, the growth of precipitation by the collision-coalescence mechanism is more efficient than in similar clouds over land (Choullarton *et al*, 1998). It could also be possible that a backing of the wind, to a more southerly direction could bring more land like aerosol concentrations up from France, leading to a mix of land and sea aerosol concentration. Therefore, at such coastal locations, it can be difficult to model correctly the aerosol concentration. Aerosol can also have an effect on the temperature of nucleation, which is normally set at a -10°C threshold in the MetUM.

A key problem in cloud physics that needs to be solved is to explain quantitatively how precipitation develops in convective clouds through the ice process (Huang *et al*, 2008).

By altering the temperature threshold at which nucleation occurs, from -10°C to -15°C (*'tnuc-15'*) and -5°C (*'tnuc-5'*), the formation of ice within the cloud will be delayed or accelerated. This parameter has a combined effect with aerosol on cloud formation and precipitation processes. Setting the nucleation temperature at -15°C or -5°C are both realistic perturbations, and in the real atmosphere this temperature depends on a variety of conditions that are dealt with simplistically in the model. There are many uncertainties within the microphysics scheme, and to run a detailed model reflecting all of the processes and interactions of different parameters would be expensive and time consuming and has been shown to add little value (Choullarton *et al*, 1998).

Roughness length represents surface drag, and over the sea, is affected by windspeed, and therefore interaction with waves (metoffice, 2010e). In this case, roughness length remains constant over the sea, but it is changed over the land. Therefore, change in roughness length changes the land/sea contrast. When roughness length over land is halved (*'Rough/2'*), the wind is able to flow faster near the surface, and this may alter the surface processes, such as moisture fluxes. Such changes in the land/sea contrast are likely to influence the mechanism for the convergence line.

Soil moisture affects the Bowen Ratio, which influences how energy is divided between latent heating and sensible heating (section 1.3.3), therefore influences the initiation and properties of convection.

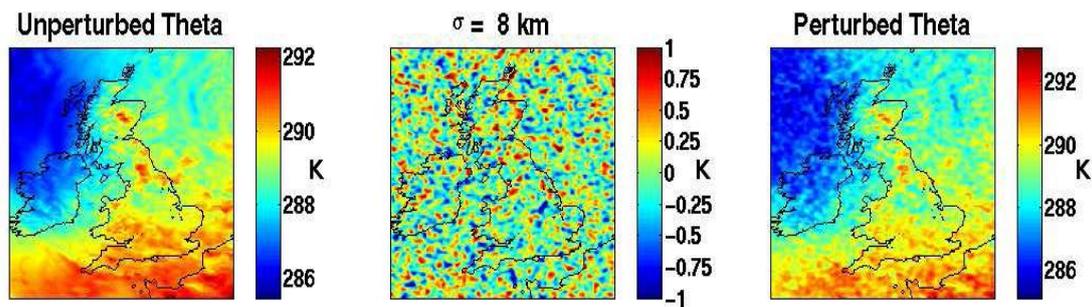
The *'No auto'* perturbation in which autoconversion is inhibited should reduce the artificial rain on the south coast (section 1.3.4.2).

The ‘*Control physics*’ ensemble is comprised of 10 members, each one having one of these physics perturbations applied. The control physics ensemble has also been analysed with the ‘*No auto*’ member removed, producing a 9 member ensemble, ‘*Ctl phys (-No auto)*’, for reasons given in the section 3.5.1.

2.6.2. Potential temperature perturbations

These perturbations are applied with regards to the control run, and also within the runs with altered physics schemes. The potential temperature is perturbed within a range consistent with surface temperature measurement errors and typical turbulent fluctuations in the convective boundary layer (Leoncini *et al*, 2010).

Figure 20. Application of potential temperature (theta) perturbations on a scale of 8km. (Plant, 2009) This example shows larger perturbations for illustrative purposes.



A random potential temperature perturbation between +0.1K and -0.1K is applied at each grid point. These perturbations are smeared out to give spatial coherence on a scale of about 8km, using Gaussian distribution, and then added to unperturbed theta, to create the perturbed theta field (figure 20). The perturbations are applied every 30 minutes, which is just longer than the typical equilibrium time-scale for a well-mixed boundary layer, allowing time for the layer to adjust to each perturbation, and frequent enough so that strong perturbation growth can occur (Leoncini *et al* 2010). No perturbations to the initial conditions have been made.

For further details, see Leoncini *et al* (2010).

There are 12 ensembles with potential temperature perturbations. The ‘*Control ensemble*’ (8 members) and ‘*50 member ensemble*’ are those where the standard physics scheme is used and the potential temperature has been perturbed (see section 2.6.2). Each physics perturbation described (2.6.1), has its own ensemble of 8 members with potential temperature

perturbations, within its altered physics scheme. For a list of all of the perturbation combinations used, see table 2 (section 3.5).

3. Results

This section discusses analysis of output from the control and perturbed ensembles, based on the Fractions Skill Score (section 3.1). Forecast skill against observations is calculated for the control run (section 3.4), providing a reference length scale to use in analyses of perturbed ensembles against the control run (section 3.5). Further examination involves comparing selected ensemble members to all other members (sections 3.6-3.10). A summary of the selection process to create an ensemble of most interest is given in section 3.11.

3.1. Fractions Skill Score (FSS)

This is a scale-selective method for evaluating precipitation forecasts, so that a scale at which the forecast becomes skillful can be determined. The concept of nearest neighbours is used to select the scales of interest, and is applied to thresholds, resulting in a measure of forecast skill against spatial scale for each threshold (Roberts and Lean, 2007).

The FSS is computed using equation 5, with its components derived in equations 6, 7, and 8.

$$\text{FSS}_{(n)} = \frac{\text{MSE}_{(n)} - \text{MSE}_{(n)\text{ref}}}{\text{MSE}_{(n)\text{perfect}} - \text{MSE}_{(n)\text{ref}}} \quad (5)$$

$$\text{MSE}_{(n)\text{ref}} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{(n)i,j}^2 \right] \quad (6)$$

$$\text{MSE}_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)i,j} - M_{(n)i,j}]^2 \quad (7)$$

$$\text{MSE}_{(\text{perfect})} = 0 \quad (8)$$

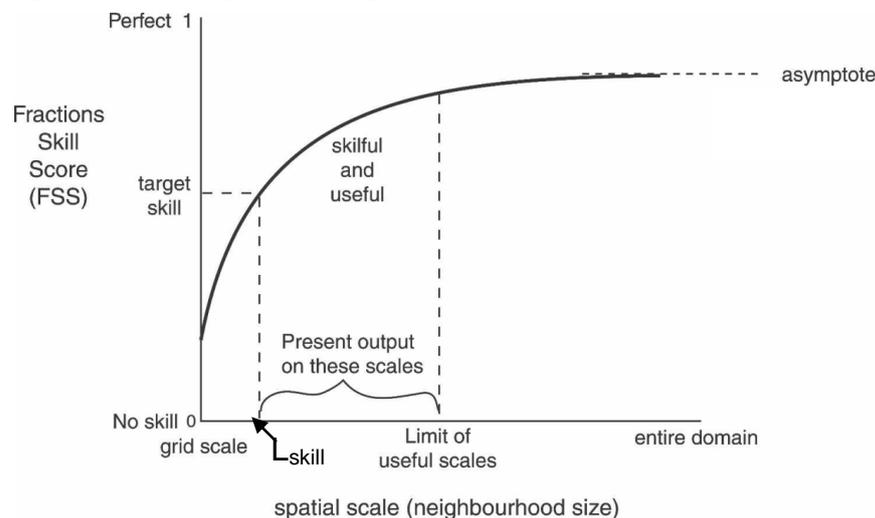
(Roberts & Lean, 2007)

Subscript (n) refers to the grid length or ‘neighbourhood length’ (Roberts & Lean, 2007).

' $MSE_{(n)}$ ' is the Mean Square Error for the observed and forecast fractions from a neighbourhood of length 'n', and is not useful on its own as it is highly dependent on the frequency of the event itself (Roberts & Lean, 2007). ' $MSE_{(n)ref}$ ' is the largest possible MSE that can be obtained from the forecast and observed fractions, whereas ' $MSE_{(n)perfect}$ ' is that of a perfect forecast. The combination of these produce FSS (equation 5), which is an MSE skill score relative to a low-skill reference forecast (Roberts & Lean, 2007). ' $O_{(n),j}$ ' is the field of observed fractions for a square of length 'n' and ' $M_{(n),j}$ ' is the field of model forecast fractions. Refer to Roberts & Lean (2007) for details of the method used to generate these fractions.

FSS has a range from 0 (no skill) to 1 (perfect skill). At the grid-scale, FSS for precipitation is the lowest, reflecting that the model has very little skill there. A selected target skill (section 3.1.1) can be interpreted as a spatial scale at which the forecast is skilful (figure 21). This skill level increases as spatial scale increases; however, usefulness of these high skill forecasts is then restricted at larger scales, where the content of the forecast is limited by smoothing. Figure 21 is a schematic showing an idealised FSS curve against spatial scale. The length from where the target skill is reached to where the output is no longer of use can therefore be determined, and so the forecast can be interpreted with a certain level of confidence at different scales.

Figure 21. Schematic graph of skill against spatial scale (adapted from Roberts & Lean, 2007).



Length scale ' L_{skill} ' has been added to this schematic (figure 21), and this is discussed in the following text.

3.1.1. FSS_r and FSS_u

On computing the FSS, two other values are derived, which have a constant value with length: FSS_r and FSS_u .

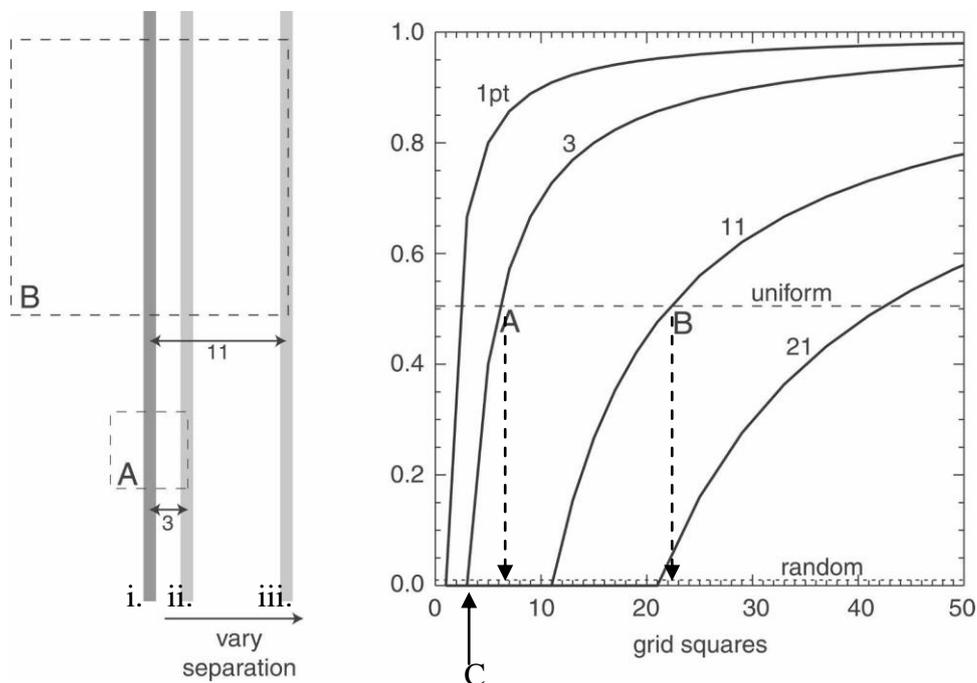
FSS_{random} (FSS_r) is the FSS that would be achieved from a random forecast, which has the same fractional coverage over the domain as the rainband itself.

FSS_{uniform} (FSS_u) is a skill score halfway between random forecast skill and perfect skill (equation 8).

$$FSS_{\text{uniform}} (FSS_u) = 0.5 + FSS_r/2 \quad (9)$$

At FSS_u the hit rate is 0.5, which can be considered a target skill. Therefore the length scale where the FSS reaches this target (L_{skill}) represents the smallest scale at which the forecast output contains useful information (figure 21).

Figure 22. Idealised example of application of FSS to determine neighbourhood size. The grey vertical lines represent the rain band i) observations, ii) forecast with 3 pixel displacement ii) forecast with 11 pixel displacement. The line graph is FSS against neighbourhood length. Adapted from Roberts & Lean (2007)



In the idealised situation illustrated above, both forecasts compute a rain band with identical structure, alignment and coverage as the observed rainband (i). However, one rainband is misplaced by 3 pixels (ii), and one is misplaced by 11 pixels (iii). The FSS plot shows skill scores for bands displaced by 1, 3, 11, and 21 pixels. A and B show where the forecast skill

for those bands displaced by 3 and 11 pixels cross the FSS_u . This is interpreted into grid square size on the FSS plot (dashed vertical arrows), and creates the neighbourhood sizes for A and B (dashed squares). When the length of the sampling square is less than or equal to the displacement error, there is no skill and the $FSS = 0$ (C). For spatial scales (grid length) longer than the displacement error, the FSS increases with spatial scale, and the smaller the forecast error, the more rapidly the skill increases with scale.

As well as finding which forecasts show greatest similarity to observations, this method can also be used to discover which forecasts show the greatest divergence from a control. The FSS can be computed against a standard control run at a defined length or range of lengths, and the lower the FSS score, i.e. low skill, the more different the forecast in terms of precipitation accuracy over the given lengths. This can be useful in determining which ensemble runs may be most effective at providing forecasts which diverge significantly from a control run.

When the data is from observations, the term ‘FSS’ shall be used (section 3.4), as in Roberts & Lean (2007). However, when the reference is the standard control run, or other perturbed members (sections 3.5-3.11), Fractions Difference Score, ‘FDS’ = $(1 - FSS)$ will be used. The FSS equation (equation 5) gives “ $FSS = 1 - (MSE_n / MSE_{(n)ref})$ ”, so ‘FDS’ is the same as examining $(MSE_n / MSE_{(n)ref})$, analysing difference rather than skill. High values of FDS will therefore identify simulations that are most different from the control.

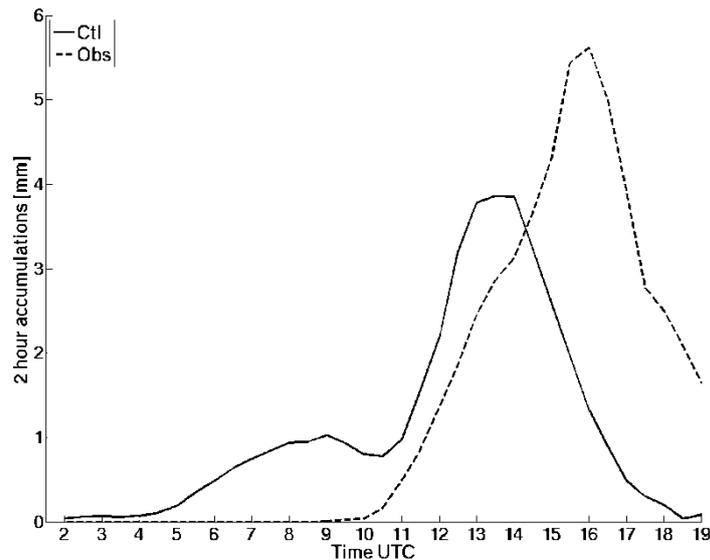
For further detailed information on the Fractions Skill Score, refer to Roberts & Lean (2007).

The FSS is effective at illustrating how skill (or difference) in the precipitation field varies with spatial scale. However, it is less intuitive than object-based methods, and the evaluation of skill depends on the metric used, and it is thus important that the metric used addresses the question of interest.

3.2. Time period

To examine model performance for extreme convection, data from the appropriate time period shall be used.

Figure 23. Mean rainfall 2 hour accumulations within a circle with 60 km diameter centred over Boscastle. Standard control forecast 'Ctl' and observations 'Obs' (courtesy of G. Leoncini).



The main period of intense rainfall over Boscastle was from 12-16 UTC. The later peak in the observations (figure 23) is due to the rainband widening and stretching northeast as it decayed, with this rainfall still encompassed within the 60 km diameter area considered above. This intense event over Boscastle can be considered the most important to forecast throughout the day, as it brought high rainfall accumulations to the area being investigated, causing flash floods. The peak in the model accumulations occurs within this time period of 12-16 UTC, which shall be investigated. The lesser peak at approximately 8-9 UTC is the artificial warm rain from shallow liquid clouds (section 1.3.4.2).

3.3. Thresholds

The FSS score can be calculated for different threshold values of rainfall, in terms of accumulation, or percentiles. By setting an accumulation threshold, the same proportion of data may not be compared in each forecast, and comparison of useful data may be omitted. If a model run under-predicted rain at all scales and a high accumulation threshold was used,

there would be a low FSS at all scales, even if the forecast showed accurate spatial distribution but inaccurate rainfall totals. However, accumulation thresholds are useful when setting actual limits, such as a rainfall threshold that when reached, the emergency services are alerted.

By using a percentile threshold, the same proportion of data is sampled in each case and therefore this method focuses on the spatial accuracy of the forecasts, and the impact of any under-prediction (or over-prediction) is removed. Higher percentile thresholds sample less extensive rain areas, with a 95th percentile picking out the localised features in the rainfall pattern that occupy only 5% of the domain (Roberts & Lean, 2007). Initial investigation into sampling different thresholds for 12-16 UTC storm total against observations for a variety of ensembles showed that examining the 95th percentile gave higher skill scores (FSS) than when analysing the 80th percentile. Therefore, the model is more accurate at positioning the smaller-scale rainfall than that on a larger-scale. This contradicts results found by Roberts & Lean (2007), when using 40 aggregated forecasts, demonstrated that it is more usual for forecasts to grasp the larger scale precipitation pattern, than the more localised features. However, this may depend on the type of rainfall and localised features and processes.

To investigate the forecast skill of extreme, localised rainfall, it is appropriate to examine the skill at the 95th percentile threshold.

The results presented in the following sections (3.4 – 3.11), are those found using the previously discussed 1km model (section 2) with analysis of the 95th percentile of the storm total from 12-16 UTC.

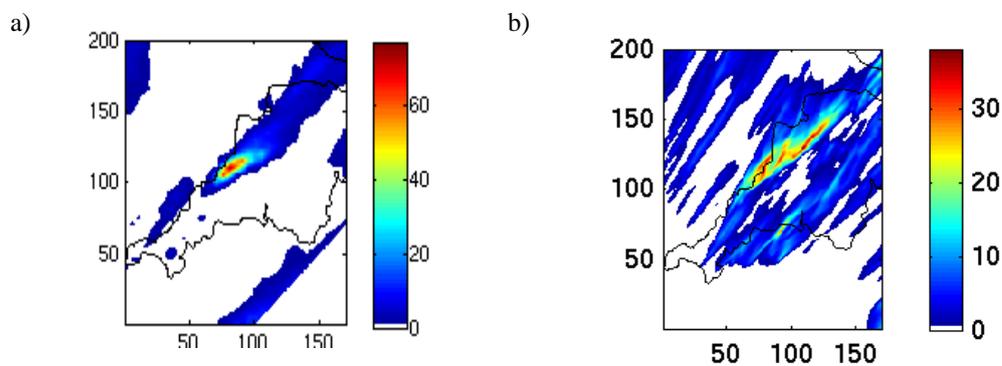
FSS was calculated using an adapted Matlab code (original code courtesy of G. Leoncini).

3.4. FSS: Observations vs. Standard Control

To ascertain L_{skill} , as defined in section 3.1.1, the observations are compared to the standard control run.

The observations of rainfall used in this analysis come from radar data on a 5 km grid (section 1.3.4), which have been interpolated onto a 1 km grid (figure 24a). The figures below show the rainfall accumulation from the observations and the standard control run from 12-16 UTC, over the entire domain.

Figure 24. Storm total map (mm) 12 – 16 UTC a). Radar b) Standard unperturbed control run



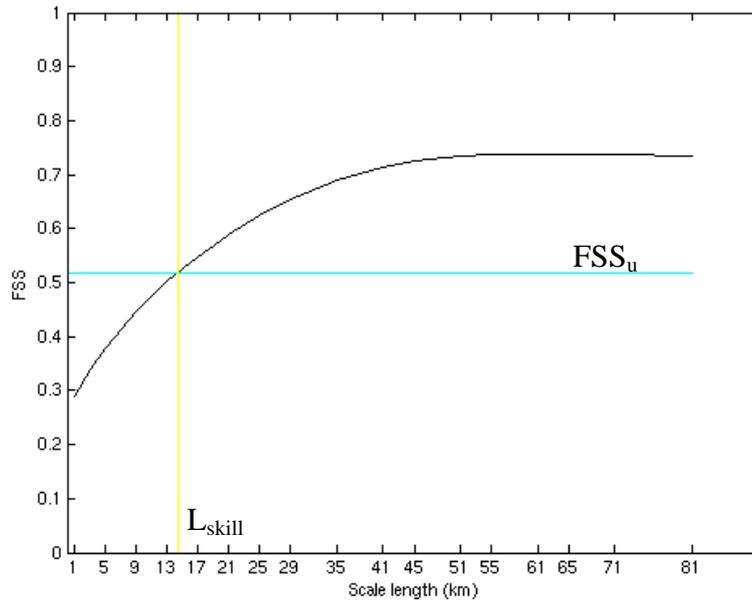
The accumulation of rainfall concentrated over Boscastle totalled 67 mm (figure 24a) during this period, however the standard control run simulated only 32 mm of rainfall over Boscastle (figure 24b). The radar observations show that the rainfall was concentrated over a small area encompassing Boscastle, whereas the control run simulated its peak intensity rainfall over a much more widespread area. There are errors in the control run in intensity and location of rainfall, but as the percentile threshold is examined, the impact of this under-prediction is removed and spatial accuracy of the precipitation pattern will be considered.

Figure 25 shows the FSS plot for the observations against the standard, unperturbed control run. The computed FSS_u (0.5181) is marked, and it crosses the FSS curve to give L_{skill} . This length (14.33 km) is the lowest limit where the model presents skilful results. This has been rounded up to the nearest length where FSS was calculated, giving $L_{skill} = 17$ km. Within the Matlab code used, FSS could only be computed at lengths of odd number, because of the way that the code handled grid point averaging.

Roberts and Lean (2007), who developed this skill score, tested it on a number of different cases including organised thunderstorms on 3rd August 2004 over London. In their analysis against this one thunderstorm event in which they examined the 95th percentile of rainfall in a

1 km model, they concluded $\text{scale}_{\min} (L_{\text{skill}}) \sim 15$ km, compared with ~ 40 km for the 12 km and 4 km forecasts. However, L_{skill} calculated over 40 aggregated cases for this model was 40-70 km (Roberts & Lean, 2007). Forecasts of thunderstorms of this sort therefore have skill at scales much lower than 40-70 km, and so this control run is a good simulation in terms of general forecast.

Figure 25. FSS curve for the standard control run against the observations



The plot of skill against scale length (figure 25), follows the classic FSS idealised curve (section 3.1), with skill increasing with scale up to a certain limit, where a maximum skill is reached. This maximum skill value is approximately 0.75, illustrating that there are still differences in spatial accuracy of storm total over the entire domain.

In order to discover the added value of a variety of ensembles, the following section (3.5) examines FDS against the standard control run, and sections 3.6-3.10 show FDS against selected perturbed members. Plots of FSS against scale length are still used in the following sections, for illustrative purposes and ease of comparison to previous results and the idealised example, however, values considered will be in terms of FDS (difference, rather than skill).

3.5. FDS: Standard control run vs. Perturbed ensembles

The following FSS are computed against the standard control run, for which L_{skill} has been diagnosed (section 3.4). Comparison of ensemble skill against the standard control run makes these results more applicable to general model output, rather than comparing ensemble runs to observations, which would restrict analysis to just this one event. L_{skill} is marked on all FSS plots in this section. At lengths greater than this, the standard control run showed skill, and so in this analysis, only results on scales equal to or greater than this length shall be discussed.

Table 2 details all of the ensembles that were used in this report. The ensemble names that shall be used throughout the results correspond to that within the table, which shows the perturbation strategy applied.

Table 2. List of ensembles and detail of the perturbations within them

Ensemble Name	Perturbation		No. of members
	Model physics	Potential temperature	
'Control Physics'	All physics	-	10
'Ctl physics (-No auto)'	Physics, without 'No auto'	-	9
'Control ensemble'	-	Yes	8
'Aero 3D to 3B'	Aero 3D to 3B	Yes	8
'Nland'	Nland	Yes	8
'NSea'	NSea	Yes	8
'tnuc-15'	tnuc-15	Yes	8
'tnuc-5'	tnuc-5	Yes	8
'Rough*2'	Rough*2	Yes	8
'Rough/2'	Rough/2	Yes	8
'SM up 20%'	SM up 20%	Yes	8
'SM down 20%'	SM down 20%	Yes	8
'No auto'	No auto	Yes	8
'50 member ensemble'	-	Yes	50

Details of these perturbations are in section 2.6.

Figure 26. FSS for ‘Control physics’ against the standard control, a) all members, b) mean and standard deviation

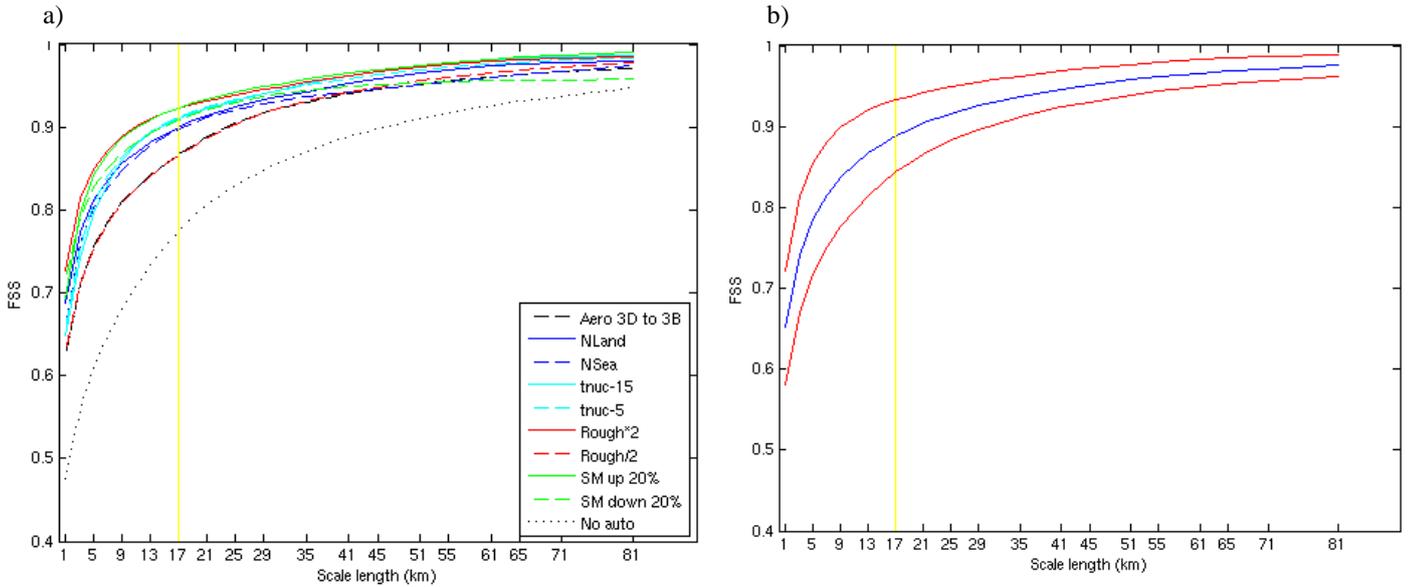
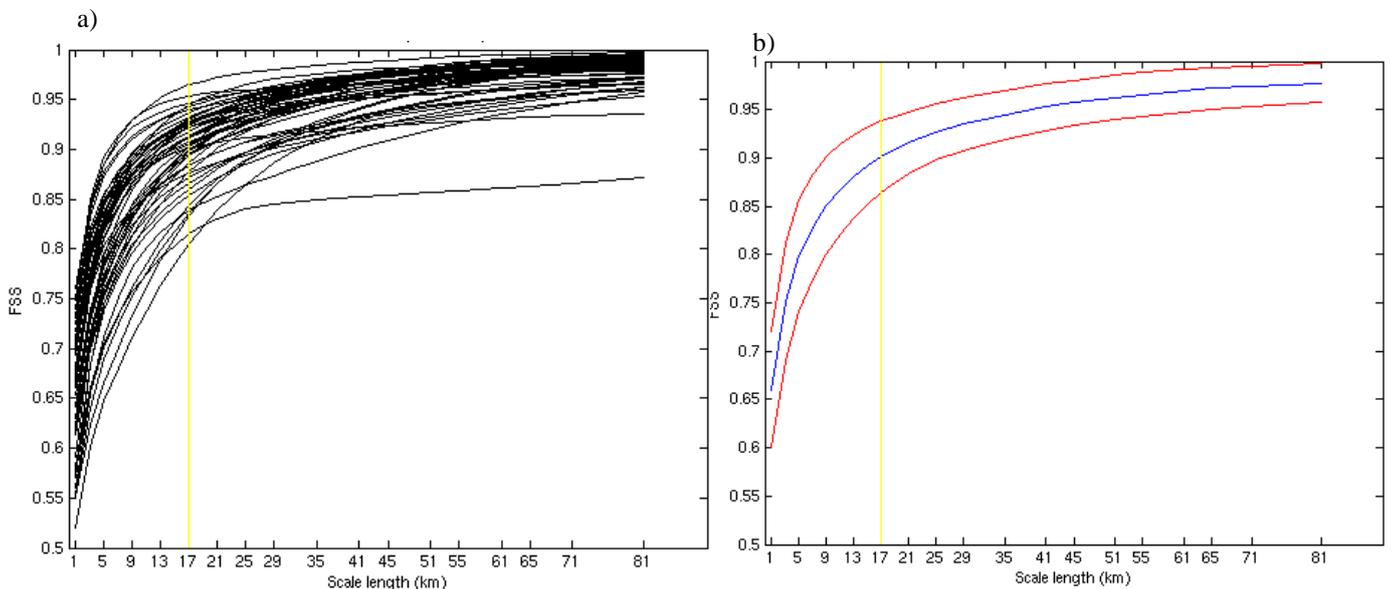


Figure 26a shows that the ‘No auto’ member is a clear outlier when the ‘Control physics’ ensemble is run against the standard control. This member shows low ‘skill’, i.e. large differences in spatial pattern of rainfall from the standard control run. All curves show the same trend, with lowest skill (largest difference) at the grid-scale, increasing towards an asymptote at larger lengths. Figure 26b shows the mean (blue) \pm standard deviation (red) for this ‘Control physics’ ensemble. The standard deviation represents the spread of the members within the ensemble, and is a measure of forecast uncertainty (section 1.4), clearly showing the convergence of skill with length, i.e. less uncertainty in precipitation pattern over the entire domain. Over all members, skill levels off at approximately 0.95, a value significantly higher than FSS of the standard control against observations. Therefore, these forecast runs are more similar to each other than to the observations.

Figure 27. FSS for ‘50 member’ ensemble against standard control a) all members, b) mean and standard deviation

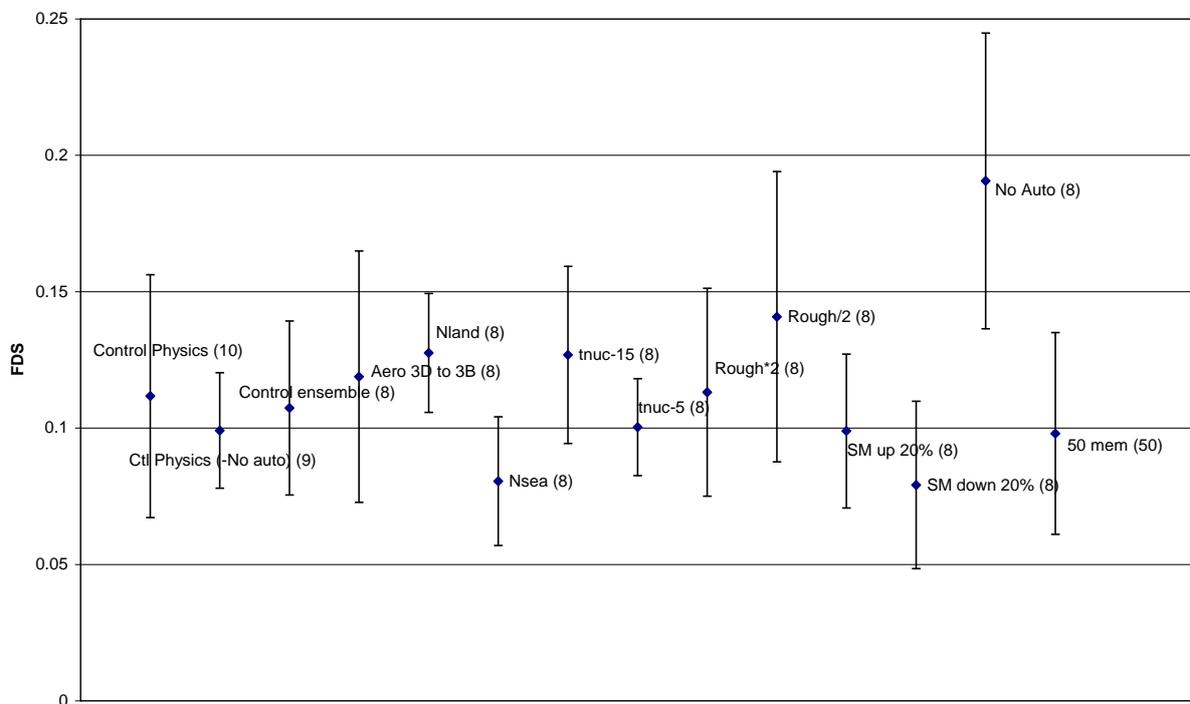


The FSS against scale length for all 50 members within the ‘50 member ensemble’ is shown in figure 27a, and all lines show the same trend. At larger length scales, the skill of the members converge, however there is one outlying member. This member reaches a skill of 0.87 at 81 km, considerably lower than the other 49 members. As all of the potential temperature perturbations within this ensemble are random, this one member cannot be picked out from the rest, and the ensemble has to be treated as a whole. This illustrates one member within a large ensemble deviating significantly from all others, highlighting the use of a large ensemble to pick out a low probability, extreme event.

3.5.1. Analysis at L_{skill}

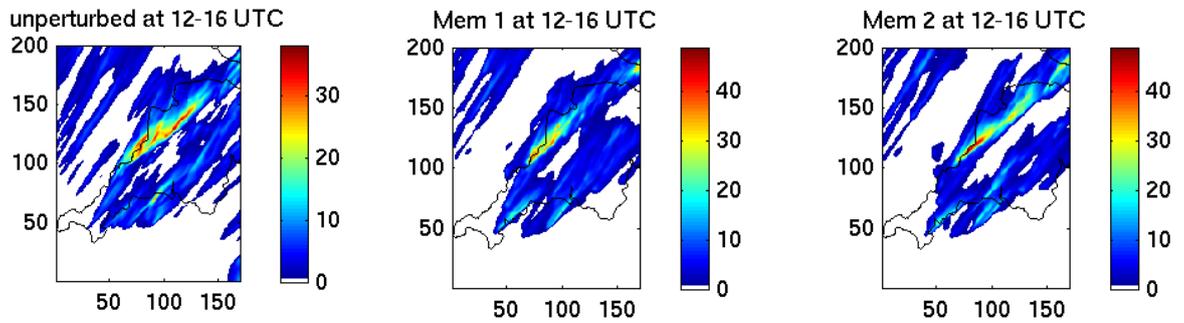
The mean FDS score \pm standard deviation at L_{skill} is calculated for all ensembles against the standard control run (figure 28). The mean values indicate how different the spatial pattern of rainfall at L_{skill} in each ensemble is from the standard control, and the standard deviation gives an idea of the FDS spread within each ensemble.

Figure 28. Mean FDS and standard deviation for standard control at L_{skill} for all ensembles. Values in brackets are number of members within the ensemble



The ensemble with the largest mean FDS and greatest standard deviation is ‘*No auto*’. However, this change to the physics scheme which restricts autoconversion, i.e. warm rain, is not a feasible alteration, as this process certainly does occur in the atmosphere.

Figure 29. Comparison of storm total maps (mm). Standard control and members 1 and 2 of ‘*No auto*’ ensemble



Even though this perturbation has an unrealistic physical basis, the model output forecasts a more intense peak rainfall over Boscastle (38-40 mm) than the standard control run (32 mm) (figure 29), however this may only be in a small number of grid points. The spatial accuracy of the forecast compared to the standard control was the lowest because the warm rain process was inhibited; a process allowed in all other forecasts. Visual analysis of figure 29 suggests that the warm rain on the south coast has been reduced in the ‘*No auto*’ ensemble.

The effect of a ‘*No auto*’ perturbation can also be seen in the ‘*Control physics*’ ensemble, which holds the control member of this perturbation (i.e. no potential temperature perturbation). When this control ‘*No auto*’ member is removed from the ‘*Control physics*’ ensemble, the mean FDS and spread of the ensemble members are considerably reduced (figure 28). As altering the physics scheme to inhibit autoconversion is not of use operationally, the full ‘*Control physics*’ ensemble is also invalid; however, this ensemble without the ‘*No auto*’ member (‘*Ctl physics (-No auto)*’) will still be examined. From sections 3.6-3.10, the ‘*Control physics*’ and ‘*No auto*’ ensembles will be excluded from analysis.

The ranked bar chart (figure 30) shows clearly the order of ensembles in relation to mean FDS, and also the relationship between mean FDS and standard deviation. The ‘*Rough/2*’ ensemble shows the highest mean FDS and largest standard deviation of those ensembles that could be used operationally. A member of this ensemble would therefore be the first choice to use in an ensemble aiming to include members with the greatest divergence from the control.

Figure 30. Ranked bar chart of all ensembles showing mean FDS (blue) and standard deviation (red). ‘No auto’ and ‘Control Physics’ are not of use operationally, and so have been faded out. Number of members within each ensemble shown in brackets

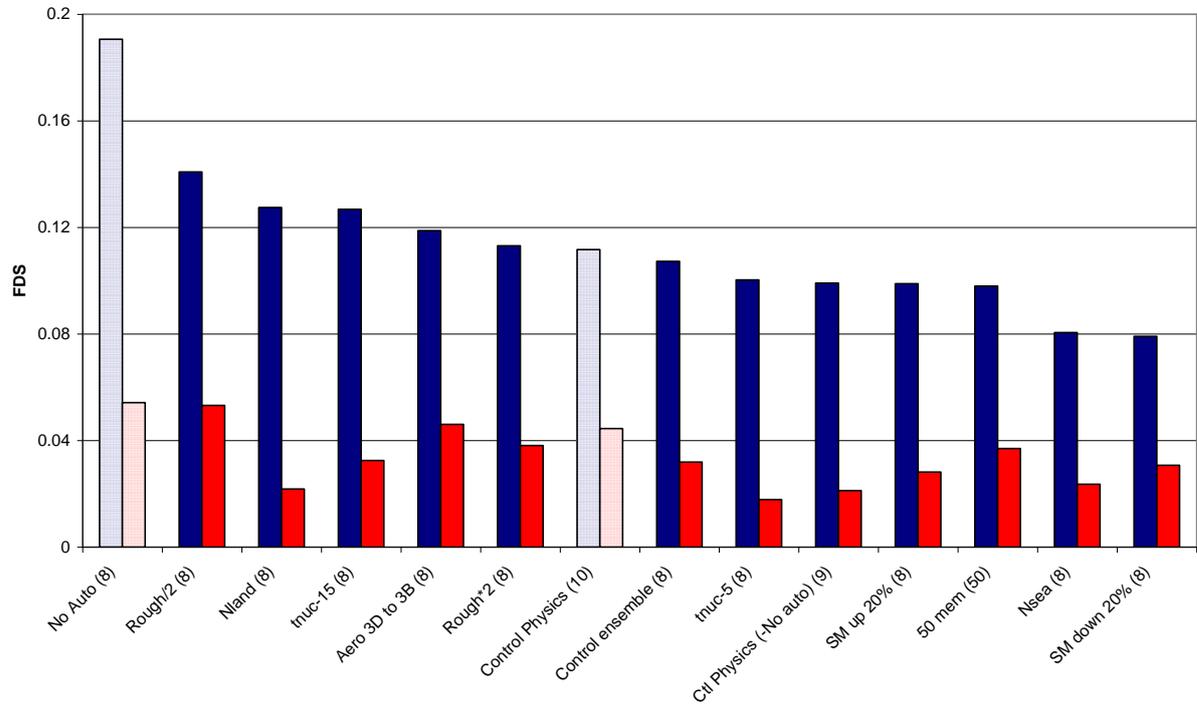


Figure 31. Storm total maps (mm) 12-16 UTC, of the unperturbed standard control and the 8 members of the ‘Rough/2’ ensemble

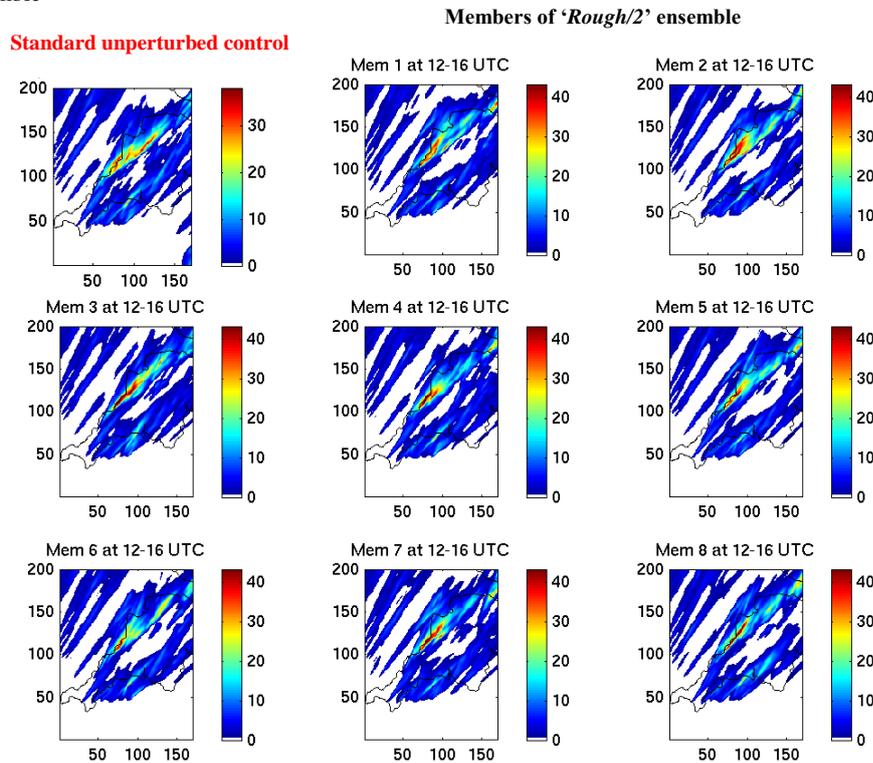


Figure 31 shows the maximum accumulation in the standard control run is 32 mm, however, the ‘*Rough/2*’ ensemble members produce maximum accumulations of 35-29 mm. Extensive investigation into these model runs may lead to conclusions as to why larger accumulations are simulated with this perturbation, such as effects of wind speed and upper level divergence, however, it may still be difficult to ascertain specific reasons, as it is likely to be due to a combination of factors.

Other ensembles which have large mean FDS and standard deviations (figure 30) are ‘*Aero 3D to 3B*’, ‘*Nland*’, ‘*tnuc-15*’ and ‘*Rough*2*’. Those ensembles which compute a smaller FDS in storm total when compared to the standard control run, are ‘*Nsea*’, ‘*tnuc-5*’, ‘*SM up 20%*’, ‘*SM down 20%*’, and ‘*Ctl Physics (-No auto)*’ (figure 30). This first step in the analysis indicates that a member from each of these ensembles would not produce the most different forecast from the standard control run in terms of spatial pattern of storm accumulation from 12-16 UTC.

The ‘*control ensemble*’ has a higher mean FDS than the ‘*50 member ensemble*’ and only slightly lower standard deviation, 0.0319 and 0.037, respectively (figure 30). This questions the value of ensembles made up of potential temperature perturbations with more than 8 members. The added expense and time of running a 50 member ensemble compared to an 8 member ensemble does not appear worthwhile from these initial results. However, figure 27a showed one member of the ‘*50 member ensemble*’ with a significantly different precipitation pattern from the other 49 members, which may be important for one member of a large ensemble to simulate an extreme event.

When comparing the ensembles of opposite pairs, there are clear differences between those with larger mean FDS and standard deviations to those with lower values. ‘*Nland*’ has the second highest mean FDS (figure 30), whereas ‘*Nsea*’ has the second lowest mean FDS. The standard deviation within each ensemble is similar, with 0.0218 and 0.0236 respectively. Therefore, to compute large differences in precipitation pattern, increasing the aerosol over the sea (‘*Nland*’) is more important than decreasing aerosol over the land (‘*NSea*’). When it is more difficult to make ice (‘*tnuc-15*’), there is a higher mean FDS and larger standard deviation than when this process is made easier (‘*tnuc-5*’). This suggests that whether ice is present or not is more important than how much there is. ‘*Rough/2*’ has the largest mean FDS and standard deviation of any of the ensembles of use, and its opposite, ‘*Rough*2*’, has the fifth highest mean FDS, although the third largest standard deviation. The ensemble in which the soil moisture has been increased, ‘*SM up 20%*’, has a higher mean FDS than the ‘*SM down 20%*’ ensemble, which have standard deviations just 2.5×10^{-3} different, 0.0282 and

0.0307 respectively. This comparison of opposite ensembles suggests that in order to get a greater difference in members from the standard control, perhaps it is more worthwhile pushing the physics scheme in one direction. This initial analysis also suggests that there is some similarity in the standard deviation of ensemble pairs.

The ranked bar chart of mean FDS (figure 30) highlights the similar ranking of ensembles when standard deviations are compared. Apart from ‘*Nland*’, the 6 ensembles (of use operationally) with the largest standard deviations match the ensembles with the largest mean FDS. The same can therefore be said about the 6 ensembles with the lowest mean FDS, having the lowest standard deviations, apart from one ensemble (‘*50 member ensemble*’). This larger standard deviation is one reason large ensembles may be of use, as in figure 27, as already discussed; one member may pick out an extreme event that is not computed by the rest of the ensemble.

This analysis of mean FDS and standard deviation is at $L_{\text{skill}} = 17$ km. However, the lengths at which these forecasts will be used would be up to approximately 45 km. It is not necessary to examine skill or difference at scales larger than this, as the 4km model from which this 1km model data is downscaled already has skill here. The use of the 1 km model is to find skilful data at the smaller scales. To verify whether analysis at this one length scale is representative of that over 17 km – 45 km, the mean FDS and standard deviation of skill averaged from 17 km to 45 km has been calculated (section 3.5.2).

3.5.2. Analysis at 17 km – 45 km

Figure 32 has been output on the same scale as figure 28, and shows that there is a clear decrease in mean FDS for all ensembles. This is as expected, as figures 26 and 27 show that at larger scales the FSS curves of all members converge to a more similar FSS. The standard deviation within each ensemble is therefore decreased when examining FSS (or FDS) over this wider range of scales. Apart from the slight transposition of values, the same trend is seen when examining FDS at L_{skill} or over the range of lengths that are relevant. The ranking of mean FDS for 17 – 45 km (not shown) only has one difference to the ranking at 17 km (figure 30), a reversal of the order of the ‘*50 member ensemble*’ and ‘*SM up 20%*’. Therefore, for the remainder of this analysis, FDS at L_{skill} will be taken as representative for the relevant length range.

Figure 32. Mean FDS and standard deviation from 17 km to 45 km for all ensembles. Values in brackets are number of members within each ensemble

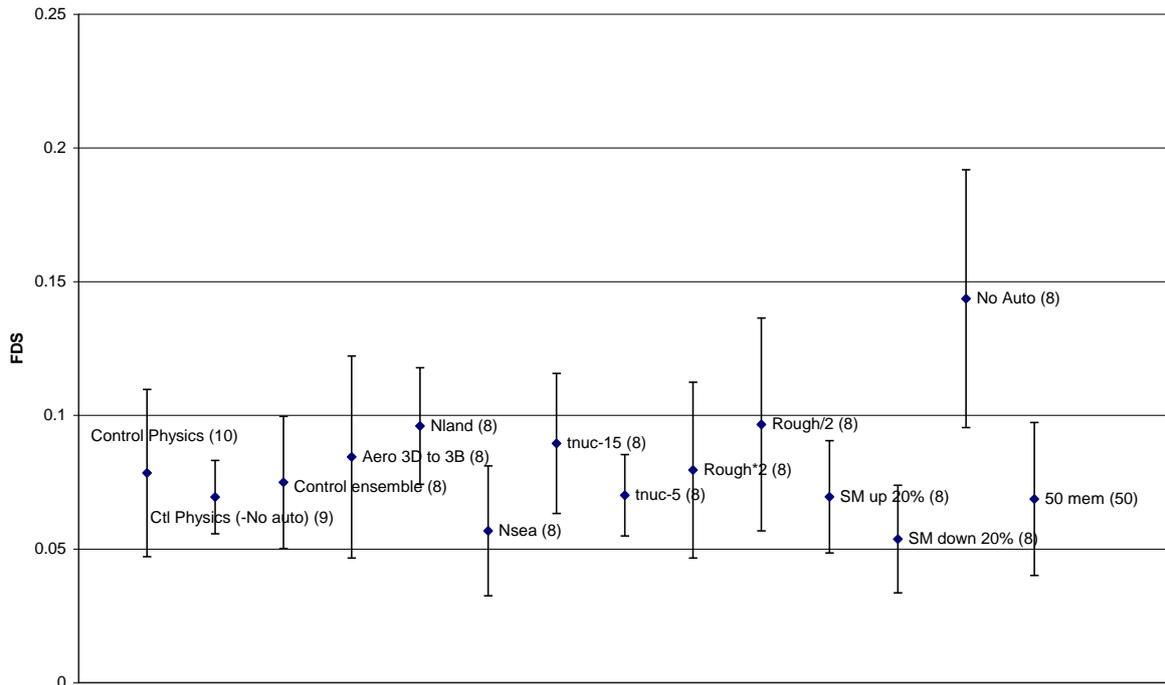


Figure 33 illustrates that there is little change of the order of ensembles' mean FSS (FDS) over the length scales within the domain. All ensembles reach a similar skill (difference) at domain scale, of ~ 0.975 (~ 0.025), however at L_{skill} the skill (difference) ranges from 0.82 (0.28) ('No auto') to 0.92 (0.018) ('SM down 20%').

Figure 33. Mean FSS with length, for each ensemble against the standard control run

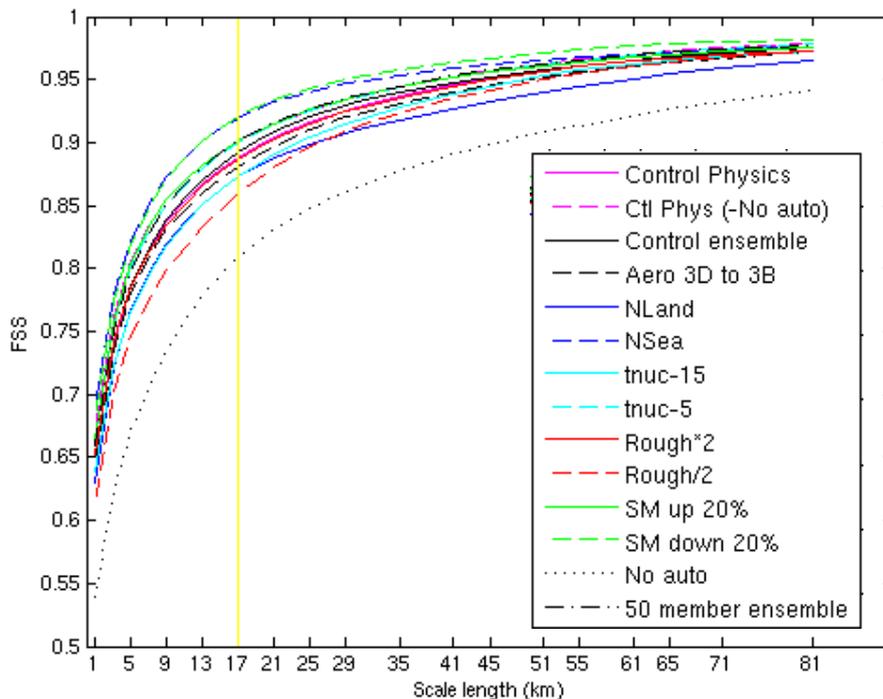
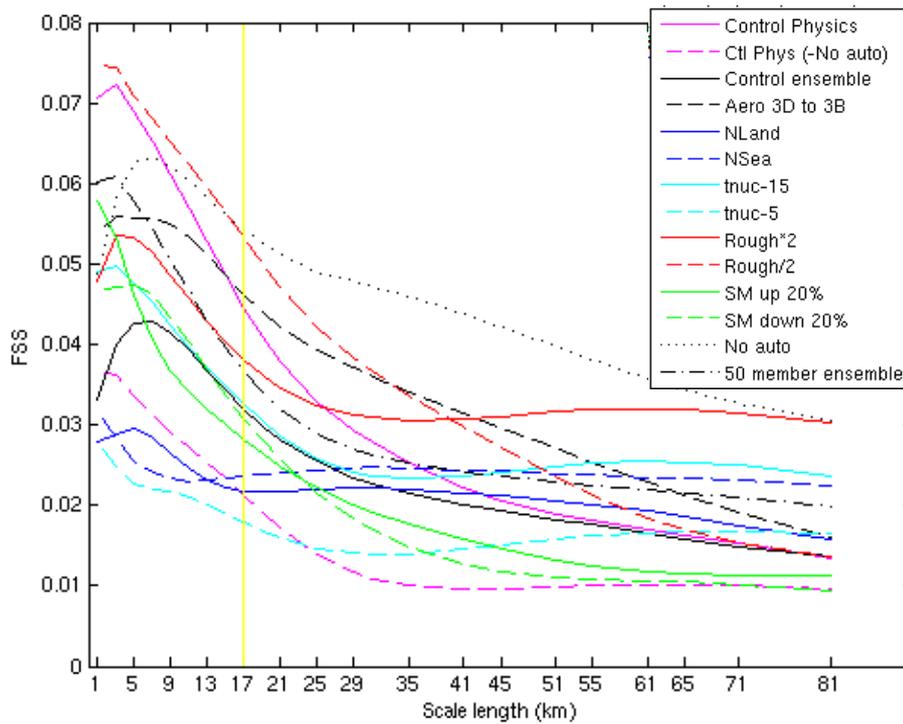


Figure 34. Standard deviation against length, for each ensemble against the standard control run



Over all length scales, the standard deviation of the ensembles changes in a less uniform manner than mean FSS. The spread in standard deviation changes little over all lengths, from a range of 0.033 at L_{skill} (17 km), to a range of 0.031 at 81 km. With increasing length, the general trend for all ensembles is a decreased standard deviation. At the largest length scales, the standard deviation within most ensembles appears to be levelling off, reaching a minimum. Comparison of pairs of ensembles shows that the change in the standard deviation with shows a similar trend. For example, the standard deviation of ‘*tnuc-15*’ and ‘*tnuc-5*’ reach a minimum at approximately 31 km, followed by a slow upward trend, and then levelling off at 81 km. This illustrates the trend discussed in section 3.5.1, where the standard deviation of values at L_{skill} appeared to show similarities in ensemble pairs.

These results also show that different model physics perturbations produce different sensitivities to the small-scale potential temperature perturbations, with some ensemble pairs (e.g. ‘*NSea*’ & ‘*Nland*’, ‘*SM up 20%*’ & ‘*SM down 20%*’) showing very similar standard deviations.

3.5.3. Summary of analysis of standard control run vs. all ensembles:

- ‘*No auto*’ ensemble shows the greatest mean FDS and spread within its ensemble, but is not of use operationally

- ‘*Rough/2*’ ensemble shows the greatest mean FDS and standard deviation of useful ensembles
- ‘*50 member ensemble*’ is of similar value to the ‘*Control ensemble*’, which has 8 members
- Analysis at L_{skill} (17 km) is sufficient to characterise model difference over all lengths that are relevant (17 – 45 km)

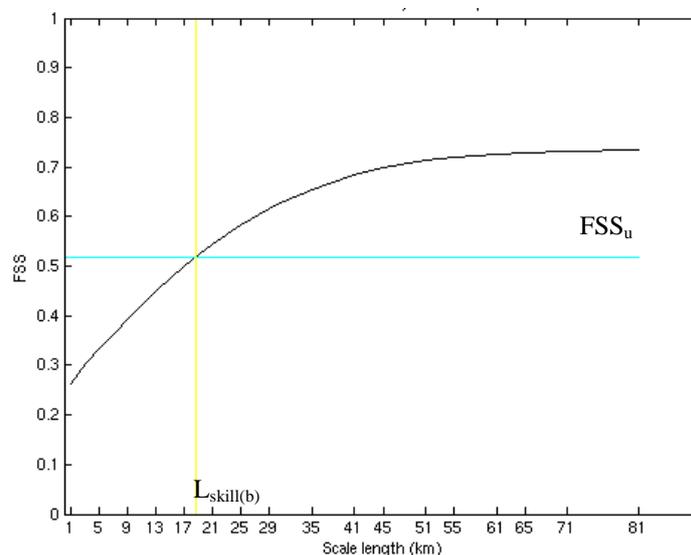
The following section continues analysis with a random member from the ‘*Rough/2*’ ensemble used as the reference forecast, as it had the largest FDS against the standard control run.

3.6. FDS: Second member vs. perturbed ensembles

A member of the ‘*Rough/2*’ 8 member ensemble was randomly chosen (member 4), as this ensemble showed the greatest mean FDS against the standard control run. FSS was calculated for all ensembles against this run, in the same way as it was calculated against the standard control run (section 3.5).

In order to continue FDS analysis at one length (L_{skill}), the assumption is made that L_{skill} varies little between the ensemble members. In order to verify this assumption and continue to use 17 km as a representative length, the FSS for the 4th member of ‘*Rough/2*’ was run against the observations (figure 35).

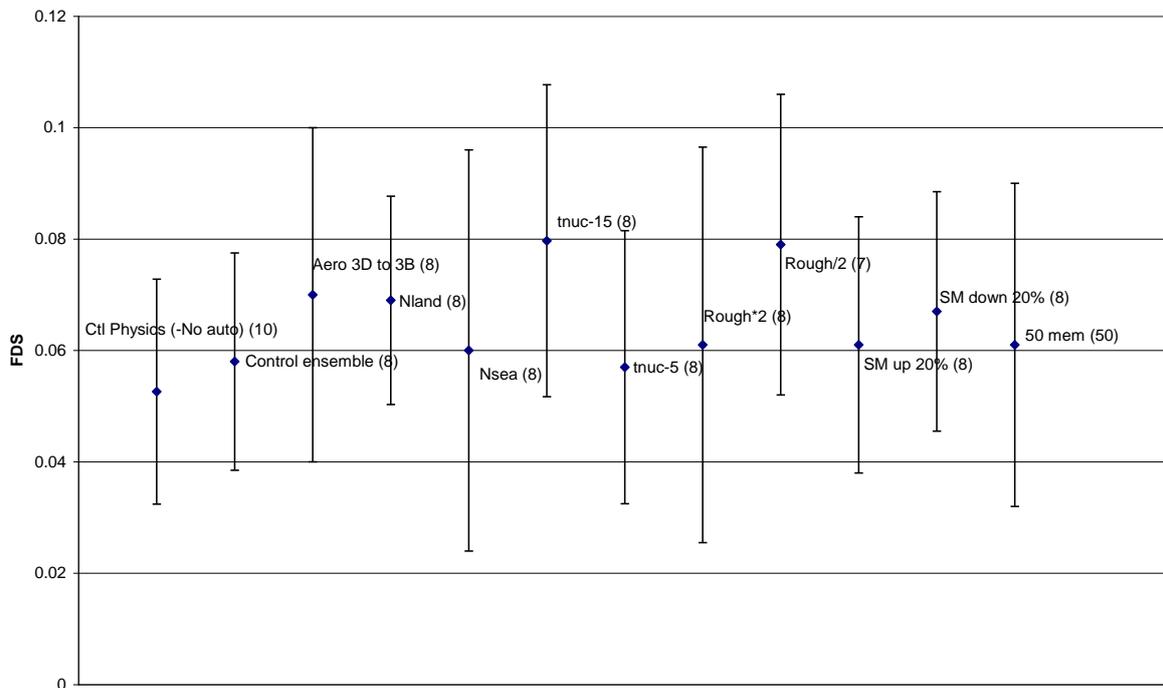
Figure 35. FSS for ‘*Rough/2*’ (member 4) against observations.



FSS_u is the same as for the standard control (0.5181), giving L_{skill(b)} of 18.5 km. This is between two lengths where FSS is computed (17 km and 19 km), and so it is a reasonable assumption to continue to use L_{skill} = 17 km, as derived from FSS of the standard control against observations (section 3.4). The use of FSS in an operational environment would require the determination of L_{skill} using a large number of past runs against observations of similar cases, as the matching observations would not be available in a forecast situation.

In this set of results, the complete control physics ensemble has 11 members, as the standard control is included. The ‘*Rough/2*’ ensemble is composed of only 7 members, as FSS is calculated against the 4th member. ‘*No auto*’ and ‘*Control physics*’ are no longer included, for reasons given in section 3.5.1.

Figure 36. Mean FDS and standard deviation at L_{skill} for all ensembles against ‘*Rough/2*’ member 4. Values in brackets are number of members within each ensemble



The mean FDS between all ensembles and ‘*Rough/2*’ (member 4) (figure 36) is lower than when compared to the standard control run (figure 28). The range of mean FDS against ‘*Rough/2*’ (member 4), is 0.0797 (‘*tnuc-15*’) to 0.0526 (‘*Ctl physics – no auto*’), whereas in section 3.5.1, analysed against the standard control, the range is 0.1408 (‘*Rough/2*’) to 0.0791 (‘*SM down 20%*’). Along with a decrease in mean FDS, standard deviation values are also decreased, with a maximum of only 0.036 (‘*Nsea*’ against ‘*Rough/2*’ (member 4)), compared to a maximum of 0.0532 (‘*Rough/2*’ ensemble against the standard control).

The ‘*tnuc-15*’ has the largest FDS against ‘*Rough/2*’ (member 4) (figure 36), and the remaining 7 members of the ‘*Rough/2*’ ensemble have the second largest mean FDS (figure 36). This may be attributable to large differences in the storm total pattern of individual members, leading to a large standard deviation within the ‘*Rough/2*’ ensemble (figures 28, 30, 34).

As in section 3.5.2, this analysis was also computed from 17 – 45 km, and the ensembles ranked in the same order as at 17 km. This reaffirms the use of analysis at just one scale length (L_{skill}), and this length shall be used for the remainder of the analysis.

3.6.1. Summary of analysis for second member vs. all ensembles:

- L_{skill} varies little and so analysis at 17 km is justified
- There is a decreased mean FDS in all ensembles when compared to ‘*Rough/2*’ (member 4), than when compared to the standard unperturbed control
- The ‘*Rough/2*’ ensemble continues to have a large mean FDS and standard deviation when compared to one of its members

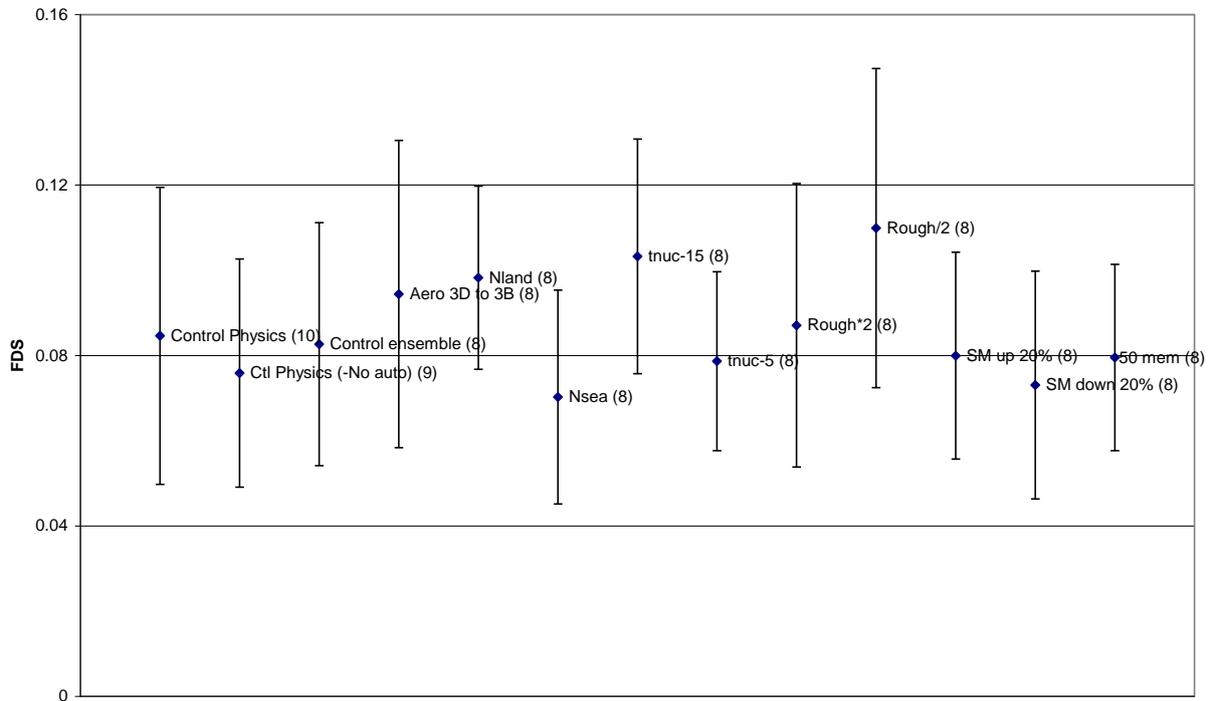
The third member is chosen from a combination of these results gathered in sections 3.5 and 3.6, as this next member needs to be most different from these two runs already used.

3.7. Selection of third member: Calculating averages

In order to ascertain which member should now be chosen, the average mean FDS and standard deviation from sections 3.5 and 3.6, against the standard control and ‘*Rough/2*’ (member 4) is calculated.

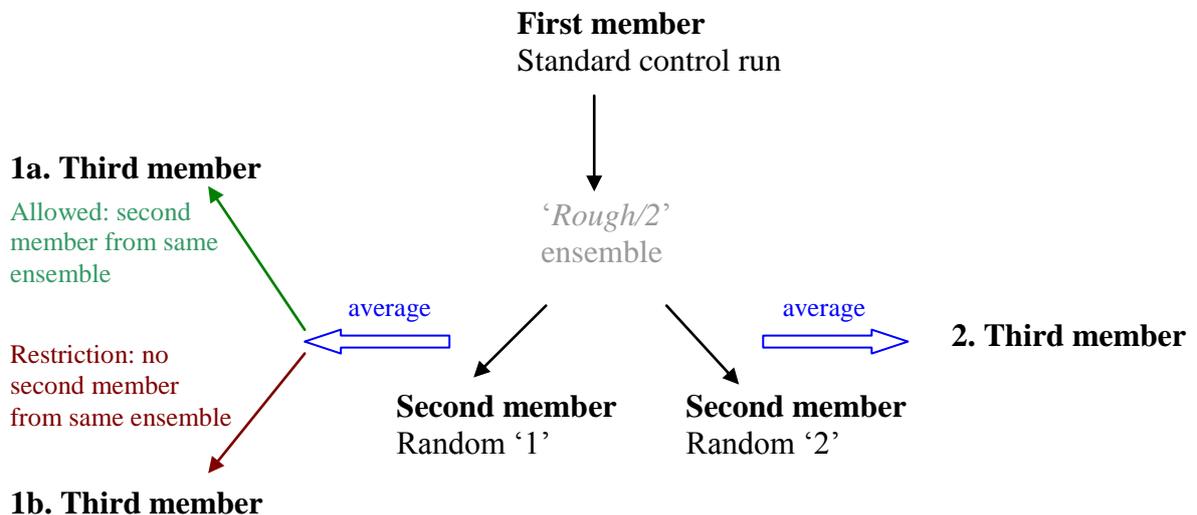
Even though ‘*tnuc-15*’ showed a larger difference in section 3.6 (against ‘*Rough/2*’ (member 4)), by including also the mean FDS when against the standard control (figure 28), the ‘*Rough/2*’ ensemble continues to have the highest mean FDS and standard deviation of all ensembles (figure 37). These results suggest that another ‘*Rough/2*’ ensemble member should be used, as the ensemble continues to have a large mean FDS and standard deviation. However, as one member has already been used, it may be more effective to choose a member from a different ensemble.

Figure 37. Average mean and standard deviation at L_{skill} for FDS against standard control and 'Rough/2' (member 4). Values in brackets are number of members within each ensemble



Pathway 1a (section 3.8), demonstrates results that allow a second member from this ensemble to be used, whereas pathway 1b (section 3.9), applies a restriction, allowing only one member from the same ensemble (figure 38). Within the following analysis, this cascade of averaging all previous FDS results to obtain the next member shall be used.

Figure 38. Schematic of various methods used to select a few, most useful ensemble members



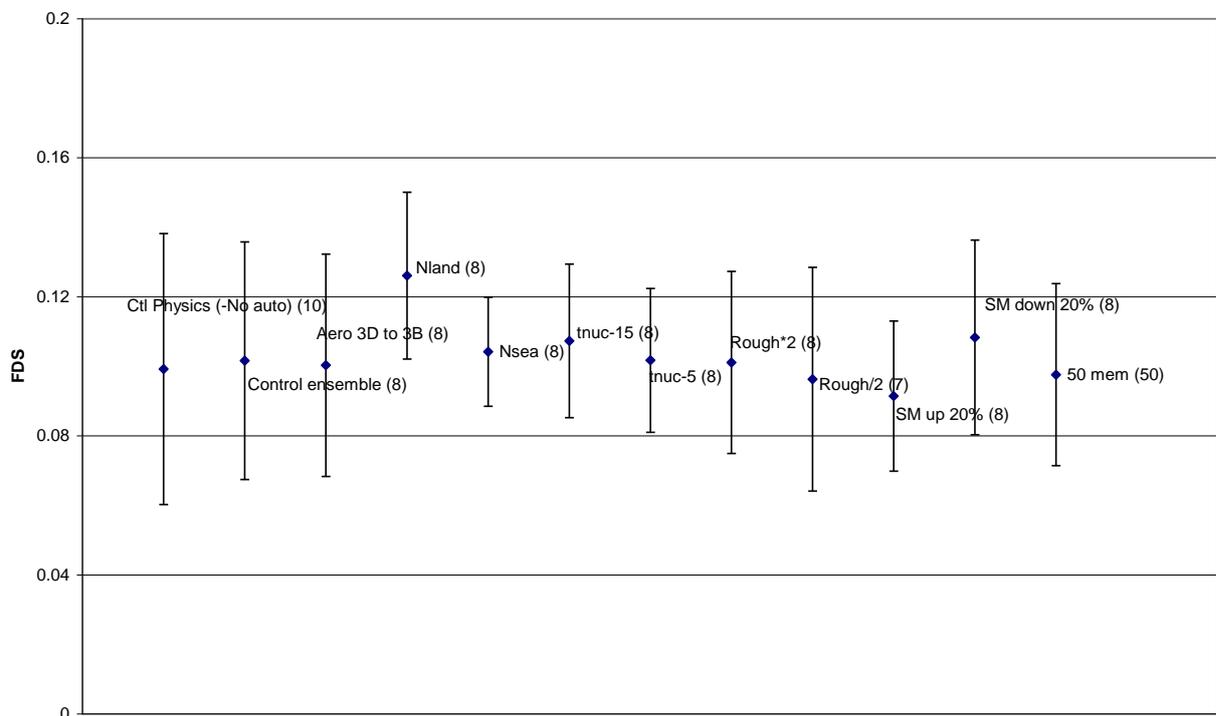
The effect of using a different second member from the ‘*Rough/2*’ ensemble (pathway 2) is investigated in section 3.10 (figure 38).

Following the different pathways, once a random member of a selected ensemble has been used as the reference, this same member shall be used if this ensemble is chosen within a different pathway.

3.8. Pathway 1a: allow 2 members from the same ensemble

The ‘*Rough/2*’ ensemble was shown to still have the highest mean FDS and standard deviation after averaging scores (section 3.7) against the standard unperturbed control (section 3.5) and against ‘*Rough/2*’ (member 4) (section 3.6), so another member (member 1) was randomly chosen from the remaining 7, against which FSS values were computed.

Figure 39. Mean FDS and standard deviation of all ensembles against ‘*Rough/2*’ member 1. Values in brackets are number of member within ensemble



FDS against ‘*Rough/2*’ (member 1) (figure 39) shows significant differences to FSS against ‘*Rough/2*’ (member 4) (figure 36). This prompts the question as to whether it matters which member of the ensemble is used, which shall be explored in section 3.10. The FDS values are higher against member 1 than against member 4. The ‘*Rough/2*’ ensemble has moved further down the ranking of mean FDS, and this suggests that member 4 was not a very typical member compared to the other 7. The ‘*Nland*’ ensemble shows the largest FDS against ‘*Rough/2*’ (member 1), and averaging this result with the previous two, leads to choosing ‘*Nland*’ as the next ensemble. Member 2 from the ‘*Nland*’ ensemble was randomly chosen, and following the cascade of averaging previous FDS results, the choice of ‘*tnuc-15*’ followed (member 6 randomly chosen), and then the ‘*Aero 3D to 3B*’ ensemble was chosen. This pathway with 2 members from the same ensemble has only been allowed once, to compute the third member.

3.9. Pathway 1b: only 1 member from the same ensemble

The ensemble with the second highest mean FDS in section 3.7, after the ‘*Rough/2*’ ensemble, is the ‘*tnuc-15*’ ensemble (figure 37). FSS was computed against member 6 of the ‘*tnuc-15*’ ensemble (member already chosen in section 3.8). The same method was followed, with an average FDS and standard deviation calculated by averaging the FDS data against this run with the data obtained from the preceding analysis against first and second members. ‘*Rough/2*’ remained the most different, with ‘*tnuc-15*’ second, and the ‘*Aero 3D to 3B*’ ensemble third. Following the strategy in which members are always chosen from different physics ensembles, a random member from the ‘*Aero 3D to 3B*’ ensemble was chosen as the fourth member. Following FDS against ‘*Aero 3D to 3B*’ (member 8), and averaging all previous FDS scores gave the ‘*Nland*’ ensemble as next highest FDS, therefore the final member should be chosen from this ensemble.

3.10. Pathway 2: Second member: different member from ‘*Rough/2*’ ensemble

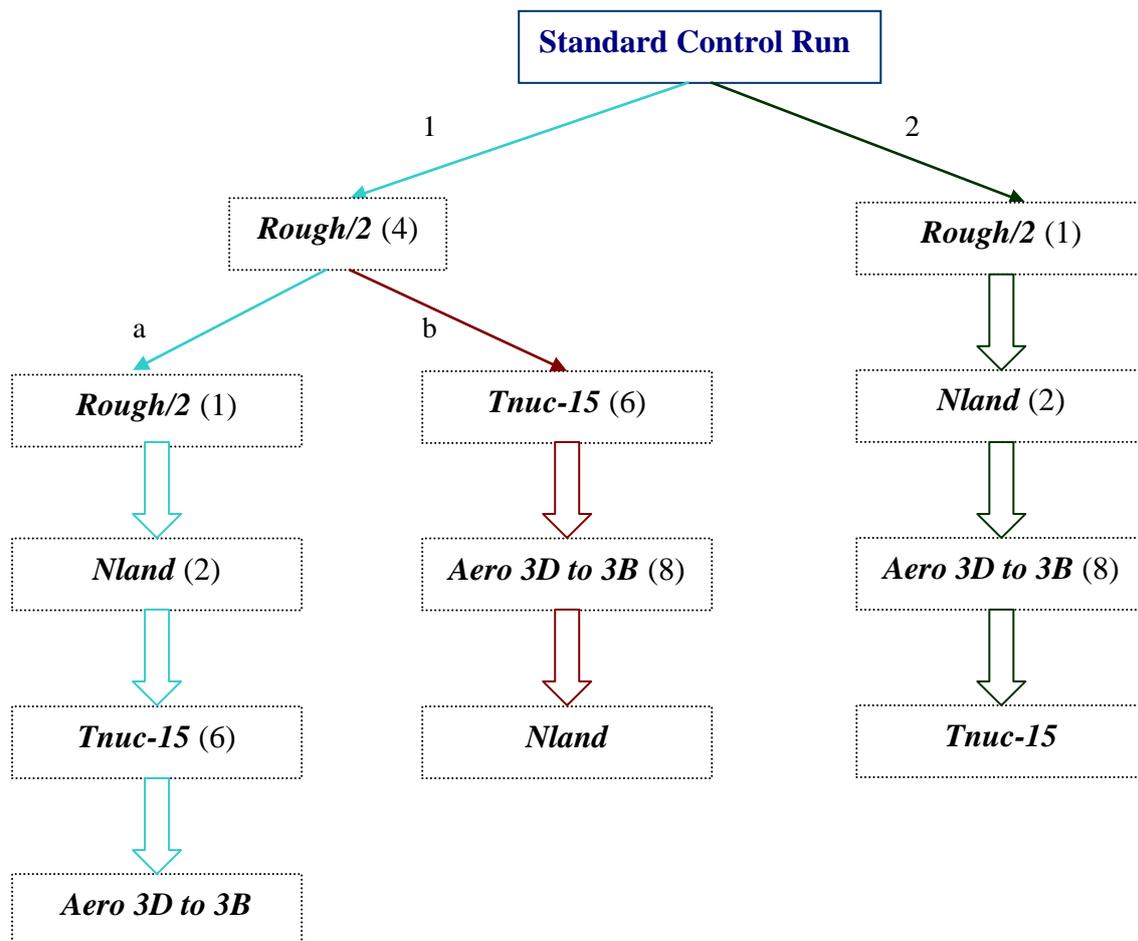
This pathway goes back to the first step (section 3.6), where a member from the ‘*Rough/2*’ ensemble was randomly chosen. This ensemble, which showed greatest FDS against the standard control (section 3.5) shall be used, however a different member from the remaining 7

members (member 4 excluded) will be used as the reference forecast. This is in order to discover the importance of which member is randomly selected from a chosen ensemble.

FSS against ‘*Rough/2*’ (member 1) was computed in section 3.8, and so is relevant for this analysis, in which member 1 shall be used as the second member. As shown in figure 39, the most different ensemble from ‘*Rough/2*’ (member 1), is ‘*Nland*’. After analysis against ‘*Nland*’ (member 2), and averaging FDS scores over previous results, ‘*Rough/2*’ was the most different, followed by ‘*Aero 3D to 3B*’. ‘*Aero 3D to 3B*’ was therefore chosen as the next ensemble to take a member from, then the further averaging of preceding scores shows a member from the ‘*tnuc-15*’ should be used as the fifth member.

3.11. Summary of different pathways

Figure 40. Flow chart of different pathways for choosing ensemble members



The summary of different pathways (figure 40) shows that whichever pathway was chosen, ultimately produced results that would be interpreted with the same ensembles being selected, although in a different order. These ensembles that were chosen using a cascade of FDS averages were the same as those with largest mean FDS in the first section (section 3.5), when FSS was computed against the standard control run.

This is an encouraging result, as it suggests that the overall approach to choosing ensemble members should not be too sensitive to details of how the choices are made. Once a good approach for picking out differences is established, it is hoped that it will identify the same simulations as being of most interest for a reduced ensemble of 5 or 6 members overall.

4. Summary and Conclusions

The motivation of this report was to investigate the effect of various perturbations on high-resolution model output, and to suggest how a useful small ensemble could be generated. Observations and forecasts for the Boscastle storm in 2004 were examined, and the Fractions Skill Score (FSS) was used to compute scores of skill and difference of precipitation pattern against scale.

Model runs at 1 km resolution were computed with various perturbations, and the related impacts on the spatial accuracy of the precipitation forecasts were investigated. This analysis has introduced a process in which to find a 5-6 member ensemble using these differences in precipitation pattern simulated in the perturbed runs. This method could be applied over numerous cases like Boscastle, with many members each time, sampling different uncertainties. The aim of such a procedure is to find out which methods of perturbation produce the most different precipitation patterns, and to discard any perturbations that are never selected in the preferred 5/6 members. There are many more ways of perturbing the model than there are of ensemble members that could be used operationally, therefore establishing the most useful ensembles will result in the best use of computer power and time.

From the many different ways of creating ensembles, this analysis has examined two strategies: perturbing model physics and perturbing the potential temperature within the boundary layer (section 2). The changes to the physics scheme deal with some uncertainty in the model itself, and how well it simulates real processes. The changes to the potential temperature take into account uncertainty in observational data. These results have shown that perturbing the potential temperature alone does not create an ensemble with the greatest mean FDS from the control, or from the other members that were used as a reference during the analysis. Results also show that an ensemble with 8 or 50 members with this potential temperature perturbation have little difference in mean FDS and standard deviation, suggesting that this ensemble strategy with more than 8 members does not add value. This agrees with the results found by Walser *et al* (2004) and Molteni *et al* (2001) (section 1.4.3),

who found that small ensembles of at least 5 targeted members were as useful as ensembles of up to 51 members. However, within the '*50 member ensemble*', one run diverged significantly from all others, illustrating the value of using large ensembles, as one member may pick out an extreme event that all other missed.

The ensemble '*Ctl physics (-No auto)*' did not compute the highest mean FDS against any reference forecast, suggesting that it would be an unsuitable choice for creating an ensemble with most different members. However, unlike all other ensembles, the members within this ensemble do not all have the same physics scheme. A more useful interpretation would be to examine each control physics perturbation individually, and if a member had high FDS, it may be a suitable choice. Another strategy would be to group the members into smaller groups, as investigation of this found large differences in mean FDS and standard deviation depending on which members were present (not shown). When those with large FDS against the standard control, such as '*Rough/2*', '*Nland*' and '*Aero 3D to 3B*' were grouped together, a high FDS was found, whereas when those with lower ranking FDS were grouped, such as '*tnuc-5*', '*SM up 20%*', the mean FDS was considerably lower.

It may also be feasible to include the relevant '*Ctl physics (-No auto)*' member when randomly choosing a member out of the potential temperature ensemble with the same perturbed physics scheme. However, an issue for operational application is that the control physics run is usually a higher resolution, so it is difficult to compare skill scores.

The interaction of the microphysics variables, such as temperature of nucleation and changes in aerosol concentration make extreme convective rainfall very difficult to forecast, as all parameters on all scales need to be modelled accurately. The Met Office employs a very basic aerosol scheme, as it is very expensive and time consuming to run a more detailed model, which includes all realistic variants. Some inaccuracies may cancel each other out, as may have occurred with the '*No auto*' perturbation, where peak rainfall increased compared to the standard control run. This highlights the need for a strong understanding of the physical mechanisms which contribute to this extreme rainfall, rather than just relying on results from a skill score.

Such changes that I have investigated are realistic, however it is difficult to diagnose specific impacts of these changes, as many are co-dependent, such as aerosol and nucleation temperature (section 2.6.1), and aerosol concentration and wind speed and direction. Including the detail of these processes, such as changing aerosol concentration with the winds

within the model would be very expensive and timely, and by perturbing various schemes to cover a variety of uncertainties, it is hoped that some effects of these interactions are sampled.

Computing FSS against the standard control showed the same ensemble to be chosen following different pathways (figure 40), suggesting that the process of choosing these ensembles is not highly sensitive to the way in which these members have been selected. The scale at which model output is useful was found to be approximately 17 km, and this agreed with results found by Roberts & Lean (2007), for an extreme rainfall event (section 3.4). The value of a small ensemble was shown, as the 50 member potential temperature perturbation did not rank higher than any other 8 member ensembles.

The three main sources of uncertainty are input data, parameterisations and model structure (section 1.3.1.1). Other perturbations could be used, such as changes to the initial conditions, boundary conditions, and time of initialisation, e.g. lagged ensembles, and also more extensive model physics perturbations. The move towards the high resolution, ‘convective-scale’ (metoffice, 2010e) 1.5 km model has the benefit that convective parametrisation will not be required, therefore eliminating the errors associated with these schemes. The input data in this investigation came from the 12 km NAE model, and it is hoped, with future development that the operational 1.5 km Met Office model will have its own data assimilation, increasing the accuracy of input data. High resolution models add value, such as the potential to include more high resolution observations (Collier, 2009) and more detailed orography data, and so running these models with ensembles would be a valuable progression.

The standard deviation of each ensemble shows the spread of the members and uncertainty of the forecast. This could prove a useful value if the same ensemble was chosen more than once. This may be due to the strong weighting of one member, and perhaps a method could be developed so that more than one member can be chosen from those ensembles with a large standard deviation, and those with a small standard deviation would not be used more than once. This in effect would set a threshold where pathway 1a or 1b would be chosen.

Ensembles allow a probabilistic risk-based approach (Ryder, 2009), and the economic value of a reliable probability forecast exceeds the value of a single deterministic forecast with uncertain accuracy (Palmer, 2000). By using an ensemble strategy and defining the probability of an event occurring, one is able to select a limit for a certain extreme event occurring, in which to implement an action plan. This adds value, as even a small probability of an event occurring may be worth taking action on; therefore ensembles are a valuable tool

for decision making. Different users have different criteria for judging usefulness, and so the application of ensembles provides a variety of results that can be interpreted in a variety of ways.

Knowledge about present day occurrence of heavy precipitation is essential to assess the impact of future changes. Accurate precipitation forecasts need to be combined with accurate hydrological models in order to forecast events such as the Boscastle storm, which generated damaging flash floods. Convective scale ensembles have been shown to add value to the forecast, however much more research needed. The Met Office is currently developing their 1.5 km model and there are numerous studies on this research area currently in progress.

3.10. Recommendation

A recommendation for a generic method that can be adapted to create an ensemble with different members:

- 1). Determine L_{skill} for the model in each type of situation (e.g. intense convective rainfall), using many runs of FSS observations against standard control run.
- 2). Compare the standard control run to all perturbed ensembles using FSS
- 3). Analyse these results at L_{skill} or a range of skills above L_{skill} , up to where the model is of use.
- 4). Choose the perturbation strategy that has largest ensemble mean FDS
- 5). Randomly choose a member of this ensemble and run FSS for all other runs against this run.
- 6). Repeat (4) and (5) with a new strategy chosen from averages of all FDS calculations so far.
- 7). Continue until number of allowed members reached.

To apply this method, it needs to be previously determined if more than one member from the same ensemble is allowed, or if there is a restriction. If more than one member is allowed, certain limits may need to be applied, perhaps by analysing the standard deviation within the specific ensemble.

A wider variety of perturbations than those described could be used, such as different boundary conditions, initial conditions, and combined or more complex physics schemes. As

well as computers calculating skill scores, some human intervention is needed to judge the value of a member in relation to time and expense. It is hoped that over many hundreds of runs and comparison of FDS, a certain selection of perturbed members stands out from the rest, and these could be most suitable to use in a small, convective-scale ensemble.

3.11 Limitations and Further Work

This investigation has provided some useful initial results, whilst containing some assumptions and limitations.

The altered physics schemes were all run separately; however, it may be useful to analyse results when combined physics perturbations are used, as many variables interact, such as aerosol and temperature of nucleation. The range of potential temperature perturbation that was applied within the boundary layer of some ensembles was selected due to investigation by Leoncini *et al* (2010). Further investigation into this, perhaps with different weather regimes and local effects, may suggest a different range of perturbations to be used.

All of this data analysis has used model output of the 95th percentile of storm total from 12-16 UTC, for reasons justified in sections 3.2 and 3.3. It may be useful to analyse just the period of convection initiation, and use a smaller percentile, such as 99th percentile, to capture the pattern of even more localised rainfall. Through analysing a percentile threshold, I have eliminated the impact of under-prediction (or over-prediction), and concentrated on the spatial details of forecasts. To adapt this method to include analysis of precipitation amount, an accumulation threshold can be applied, as discussed in section 3.3.

Results from this study revealed that the members chosen were those initially most different from the standard control run, which was shown produce the largest FDS in ensembles when used as the reference forecast. This suggests that all of the perturbed runs were more similar to themselves than to the control ensemble, and so calculating the cascade of averages whilst omitting the standard control run, may produce different results and is worth investigating.

This report has been based on model runs from one event and comparison was made to the control run to make the analysis applicable to general model output, rather than this specific case. It would therefore be useful to use this approach for many cases of this sort, i.e. extreme convective rainfall, as FSS has been shown to be a useful tool for this application.

Leoncini *et al* (2010) stated that predictability also has dependence on the weather regime, and that model uncertainties dominate for the first few hours, but thereafter lateral boundary conditions (LBCs) dominate. Therefore, in other situations a different variable may be considered, and the benefit of this skill score is that any percentile or accumulation threshold can be applied to adapt analysis specific to needs. For example, this method may be applied to obtain useful members for the forecast of fog, as this is a difficult phenomenon to forecast with low resolution models.

When comparing the different model runs and selecting which may be most useful in creating a diverse ensemble, there has been the assumption that all runs are as efficient in terms of time to run and expense. These practical considerations could be investigated further in an operational environment, where hundreds of runs can be computed and compared, and detailed knowledge of the computer system can be used to assess cost and time benefits of members and combine this with FSS analysis.

5. References

Alfieri, L., Claps, P., Laio, F., 2010. Time-dependent Z-R relationships for estimating rainfall fields from radar measurements. *Nat. Hazards Earth Syst. Sci.* **10**: 149-158

Bennett, L.J., Browning, K.A., Blyth, A.M., Parker, D.J., Clark, P.A., 2006. A review of the initiation of precipitating convection in the United Kingdom. *Q. J. R. Meteorol. Soc.* **136**: 1001-1020

Bousquet, O., Lin, C.A., Zawadzki, I., 2006. Analysis of scale dependence of quantitative precipitation forecast verification: A case-study over the Mackenzie river basin. *Q. J. R. Meteorol. Soc.* **132**: 2107-2125

Bowler, N.E., Pierce, C.E., Seed, A.W., 2006. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.* **132**: 2127-2155

Browning, K.A. & Roberts, N.M., 1994. Use of satellite imagery to diagnose effects leading to frontal thunderstorms: Part 1 of a case-study. *Meteorol. Appl.* **1**: 303-310

Bryan, G.H., Wyngaard, J.C., Fritsch, J.M., 2003. Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review* **131**: 2394-2416

Burt, S., 2005. Cloudburst upon Hendrabortnick Down: The Boscastle storm of 16 August 2004. *Weather* **60(8)**: 219-227

Choulaton, T.W., Bower, K.N., Beswick, K.M., Parkin, M., Kaye, A., 1998. A study of the effects of cloud processing of aerosol on the microphysics of cloud. *Q. J. R. Meteorol. Soc.* **124**: 1377-1389

Clark, P., 2009. Issues with high-resolution NWP. MOSAC-14 Paper No. 14.6, Met Office

- Collier, C.G., 2007. Flash flood forecasting: What are the limits of predictability? *Q. J. R. Meteorol. Soc.* **133**: 3-23
- Collier, C.G., 2009. On the propagation of uncertainty in weather radar estimates of rainfall through hydrological models. *Meteorol. Appl.* **16**: 35-40
- Collier & Fox, 2000. On the impact of spatial variations of orographic frontal rainfall upon the probability matching method of radar rainfall adjustment. *Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere.* **25**(10-12): 939-942
- Crook, N.A., 1996. Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Monthly Weather Review.* **124**: 1767-1785
- Dance, S.L., & Zou, Q.P., 2010. HESS Opinions: “Ensembles, uncertainty and flood prediction”. *Hydrology and Earth System Sciences Discussions.* **7**: 3591-3611
- Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A., Wood, N., 2005. A new dynamical core for the Met Office’s global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **135**: 1759-1782
- Golding, B., 2005. Numerical Weather Prediction: Boscastle and North Cornwall post flood event study – Meteorological analysis of the conditions leading to flooding on 16 August 2004. *Met Office Forecasting Technical Report* **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 (8)**: 230-235
- Golding, B., 2007. Severe rainfall in London. Talk: May 2007
- Golding, B.W., 2009. Review: Long lead time flood warnings: Reality or fantasy? *Meteorol. Appl.* **16**: 3-12
- Gregory, D. & Rowntree, P.R., 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Review.* **118**: 1483-1506

Hand, W.H., Fox, N.I., Collier, C.G., 2004. A study of twentieth-century extreme rainfall events in the United Kingdom with implications for forecasting. *Meteorol. Appl.* **11**: 15-31

Hohenegger, C., Schar, C., 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. *BAMS* 1783-1793

Huang, Y., Blyth, A.M., Brown, P.R.A., Choullarton, T.W., Connolly, P., Gadian, A.M., Jones, H., Latham, J., Cui, Z., Carslaw, K., 2008. The development of ice in a cumulus cloud over southwest England. *New Journal of Physics.* **10**: 1-23

Kalnay, E., Hunt, B., Ott, E., Szunyogh, I., 2006. Ensemble forecasting and data assimilation: two problems with the same solution? Pages 157-180 in Palmer, T. & Hagedorn, R. (Ed.). Predictability of weather and climate. Cambridge University Press

Kendon, E., 2010. Understanding extreme rainfall processes and the reliability of future projections. Talk at: University of Reading, 10th May

Kong, F., Droegemeier, K.K., Hickmon, N.L., 2006. Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparison of coarse- and fine-grid experiments. *Monthly Weather Review.* **134**: 807-833

Kong, F., Droegemeier, K.K., Hickmon, N.L., 2007. Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Monthly Weather Review.* **135**: 759-782

Lalaurette, F. & van der Grijn, G., 2006. Ensemble forecasts: can they provide useful early warnings? Pages 614-627 in Palmer, T. & Hagedorn, R. (Ed.). Predictability of weather and climate. Cambridge University Press

Lean, H.W., Clark, P.A., Dixon, M., Roberts, N.M., Fitch, A., Forbes, R., Halliwell, C., 2008. Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review* **136**: 3408-3424

Leoncini, 2009. Workshop on short-range ensemble prediction systems. Talk at: 4th SRNWP Workshop, 23-25 June, Exeter

- Leoncini, G., Plant, R.S., Gray, S.L., Clark, P.A., 2010. Perturbation growth at the convective scale for CSIP IOP18. *Q. J. R. Meteorol. Soc.* **136**: 653-670
- Leutbecher, M., Palmer, T.N., 2007. Ensemble forecasting. *Journal of computational physics* **227**: 3515-3539
- Maraun, D., Rust, H.W., Osborn, T.J., 2009. The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics. *Int. J. Climatol.* **29**: 1731-1744
- May, B., Clark, P., Cooper, A., Forbes, R., Golding, B., Hand, W., Lean, H., Pierce, C., Roberts, N., Smith, R., 2004. Numerical Weather Prediction: Flooding at Boscastle, Cornwall on 16 August 2004 – A study of Met Office Forecasting Systems. *Met Office Forecasting Research Technical Report* **429**
- metoffice, 2010a
http://www.metoffice.gov.uk/education/teens/casestudy_boscastle.html
- metoffice, 2010b
www.metoffice.gov.uk/corporate/library/factsheets/factsheet15.pdf
- metoffice, 2010c
www.metoffice.gov.uk/science/creating/daysahead/nwp/um.html
- metoffice, 2010d
www.metoffice.gov.uk/research/modelling-systems/unified-model/weather-forecasting
- metoffice, 2010e
www.metoffice.gov.uk/science/creating/hoursahead/mesoscale.html
- metoffice, 2010f
www.metoffice.gov.uk/research/areas/data-assimilation-and-ensembles/high-resolution
- metoffice 2010g
<http://www.metoffice.gov.uk/science/creating/hoursahead/nowcasting.html>
- metoffice, 2010h
<http://www.metoffice.gov.uk/science/creating/daysahead/ensembles/MOGREPS.html>
- metoffice, 2010i
http://www.metoffice.gov.uk/science/creating/first_steps/atmos_surface.html

Molteni, F., Buizza, R., Marsigli, C., Montani, A., Nerozzi, F., Paccagnella, T., 2001. A strategy for high-resolution ensemble prediction. I: Definition of representative members and global-model experiments. *Q. J. R. Meteorol. Soc.* **127**: 2069-2094

Mullen, S.L. & Buizza, R., 2001. Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. *Monthly Weather Review.* **129**: 638-663

nerc.ac.uk (2010). National Environment Research Council. Landslides, flooding and climate change in the UK. www.nerc.ac.uk/press/releases/2004/25-landslides.asp

Palmer, T.N., 2000. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.* **63**: 71-116

Pedder, M.A., Haile, M., Thorpe, A.J., 2000. Short period forecasting of catchment-scale precipitation. Part I: the role of Numerical Weather Prediction. *Hydrology and Earth System Sciences* **4(4)**: 627-633

Plant, 2009. Flash-flood forecasting. Non-technical talk.

Pierce, C.E., Hardaker, P.J., Collier, C.G., Haggett, C.M., 2000. GANDOLF: a system for generating automated nowcasts of convective precipitation. *Meteorol. Appl.* **7**: 341-360

Roberts, N.M., Lean, H.W., 2007. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* **136**: 78-97

Roberts, N.M., Cole, S.J., Forbes, R.M., Moore, R.J., Boswell, D., 2008. Use of high-resolution NWP rainfall and river flow forecasts for advance warning of the Carlisle flood, north-west England. *Meteorol. Appl.* **16**: 23-34

Ryder, P., 2009. Editorial: Flood Forecasting and Warning. *Meteorol. Appl.* **16**: 1-2

sat.dundee, 2010.

<http://www.sat.dundee.ac.uk/abin/piccyjpeghtml/modis/2004/8/16/1203+42/ch38.jpg>

Tibaldi, S., Paccagnella, T., Marsigli, C., Montani, A., Nerozzi, F., 2006. Limited-area ensemble forecasting: the COSMO-LEPS system. Pages 489-513 in Palmer, T. & Hagedorn, R. (Ed.). Predictability of weather and climate. Cambridge University Press

weather.uwyo.edu, 2010. <http://www.weather.uwyo.edu/cgi-bin/sounding?region=europe&TYPE=GIF%3ASKEWT&YEAR=2004&MONTH=08&FROM=1612&TO=1612&STNM=03808>

Walser, A., Schar, C., 2003. Convection-resolving precipitation forecasting and its predictability in Alpine river catchments. *Journal of Hydrology* **288**: 57-73

Walser, A., Luthi, D., Schar, C., 2004. Predictability of precipitation in a cloud-resolving model. *Monthly Weather Review*. **132**: 560-577

Werner, M. & Cranston, M., 2009. Understanding the value of radar rainfall nowcasts in flood forecasting and warning in flashy catchments. *Meteorol. Appl.* **16**: 41-55

Werner, M., Cranston, M., Harrison, T., Whitfield, D., Schellekens, J., 2009. Review: Recent developments in operational flood forecasting in England, Wales and Scotland. *Meteorol. Appl.* **16**: 13-22

Wetterall, F., He, Y., Cloke, L., Pappenberger, F., Wilson, M., Freer, J., McGregor, G., 2009. Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions. *Meteorol. Appl.* **16**: 91-101

Wilson, D.R. & Ballard, S.P., 1999. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **125**: 1607-1636