Met Office

Meteorology Research and Development

Meteorological components in forecasts of extreme convective rainfall using 12-km and 1-km NWP models: A tale of two storms

Modelling Extreme Rainfall Events project – stage 2 report

The project is half funded by Defra, project code FD2210

Technical Report No. 520

Nigel Roberts

email: nwp_publications@metoffice.gov.uk

Document History

Date	Version	Action/Comments	Approval
14/03/07	1.0	Passed to Brian Golding (Met R&D Programme Manager) and Peter Clark (Head of Mesoscale modelling Group) for comments.	
02/04/07	2.0	Passed back to Brian Golding following requested modifications	
03/04/07	3.0	Final version	Brian Golding, Met R&D Programme Manager

Contents

1 Introduction

This is the second interim report to be delivered from the Modelling Extreme Rainfall Events (MERE) project. It follows the work done in stage 2 of the project.

The overall objective is to investigate the ability of a storm scale configuration of the Met Office Numerical Weather Prediction (NWP) model (Davies et al 2005) to predict extreme rainfall events up to 18-24 hours ahead and to determine what it is about the meteorology of these situations that the model must capture in order to produce useful predictions for flood warning.

Five cases were examined in stage 1. Four were intense convective rainfall events and one was a case of persistent orographically enhanced rain. It was found that short-range 1-km model forecasts (up to 12-hours ahead) performed better than the equivalent 12-km forecasts for all these situations. The amount of improvement differed from case to case, and on some occasions the improvement lead to a much better forecast. This is documented in the stage 1 report (Roberts 2006).

Two of the original five cases have been selected for more detailed study in stage 2.

1. Flooding in the Hawnby area of North Yorkshire following severe convection on the 19th June 2005.
2. Flooding in northwest London following intense convection on the 3rd August 2004.

This report examines the skill of forecasts from 3 to 24 hours old. The accuracy of the forecasts will be assessed both subjectively and using an objective measure, then the particular aspects of the meteorology that the model needed to capture for a successful forecast will be examined. It will be shown that the ability of a forecast to reproduce the storm is dependent on having all the key meteorological components in place. This is a tale of two storms – the one that could be predicted and the one that could not.

1.1 The context of this work

The purpose of this report is to meet the objective of the second stage of the project: to focus on two of the extreme events investigated in stage 1 and investigate whether it is possible to achieve accurate high-resolution (~1 km) simulations of these events and document the reasons why the simulations were successful or otherwise. If the storms can be predicted accurately then further experiments can be performed to examine the sensitivity of forecast accuracy to particular aspects of the model formulation. On the other hand, if the event can not be reproduced, further sensitivity studies are of little value unless the underlying cause of the failure can be corrected. In either instance, the reasons behind the model performance need to be documented.

1.2 Running from different start times

On average, forecasts become more skilful as they become shorter; i.e. 12-hour forecasts are usually more accurate than 24-hour forecasts. However, on an individual forecast by forecast basis, this is not always true. There are reasons why

this may be the case. It may be that the accuracy of the model at the start of a forecast (i.e. how well the analysis matches observations) is worse for a later start time (e.g. the next forecast). This can happen if the later analysis did not have as many observations available in the area of interest (e.g. if the area of interest moves over the sea), or if an observation was wrong, or if the situation was more complex in the area of interest and therefore more difficult for the data assimilation system to deal with (e.g. once convection has broken out). It may be that the forecast of storm or no storm is marginal and any small differences at the start, although essentially in the noise (out of the scope of the data assimilation system), may make the difference between convection occurring or not.

Variability from forecast to forecast is likely to be particularly noticeable for convective situations because convective rain initiates on small scales, and this makes it inherently more difficult to predict because we do not have all the information required about those small scales. In addition, the only way to resolve the dynamics of convective storms (and therefore give realistic forecasts) is through the use of a high resolution grid (e.g. 1 km grid spacing), but a fine-scale grid provides a greater degree of freedom for a divergence of solutions from similar starting points. Nevertheless, the high-resolution model still has the capacity to give the most accurate forecasts because it is capable of representing the local conditions that lead to the storms (e.g. hills, coastal convergence) and the dynamics of the storms themselves.

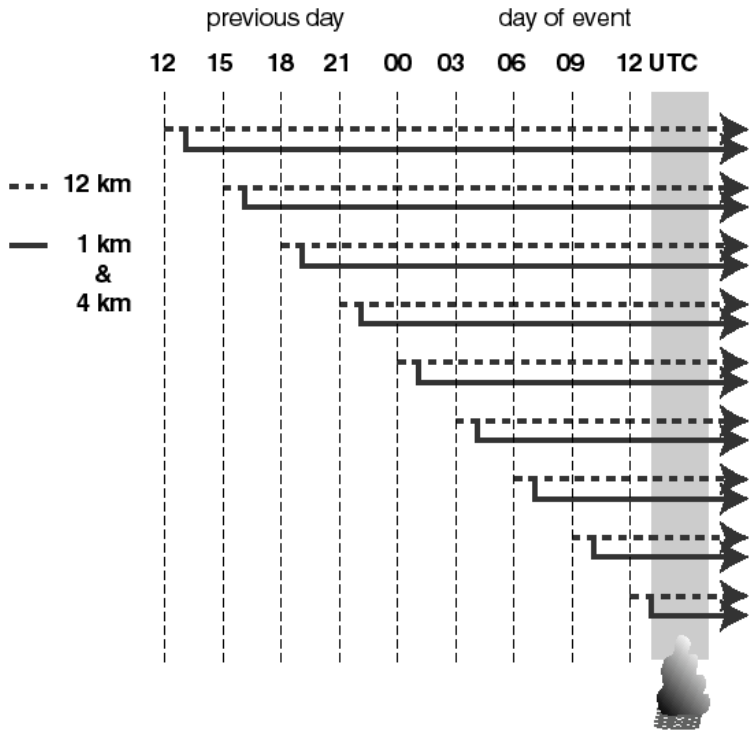

Figure 1. Schematic diagram of the succession of 12, 4 and 1-km forecasts run for a case study.

For the reasons given above, it was decided to initially examine the performance of different-length forecasts as the starting point for finding out if the 1-km model could reproduce the extreme events. This provides a small set (time-lag ensemble) of predictions that will each be more or less accurate depending on how well they were able to capture the initial flow and subsequent mechanisms responsible for the storm

development. By examining the collection of forecasts in detail and making comparisons between them it is possible to determine what aspects were important for a successful forecast. If any of the forecasts are able to reproduce the event to a sufficient level of accuracy, then our understanding of the underlying mechanisms can be used to make decisions about further sensitivity studies that will lead to greater insight into model behaviour. If none of the simulations are successful, then we should have the information to hand to say why.

For both of the case studies, forecasts were run with a grid spacing of 12, (4) and 1 km. The longest forecasts started around 24 hours before the time of the storms and subsequent forecasts began 3-hours later, then a further 3-hours later up to the shortest forecasts which started 1-3 hours before the time of the storms. The 12-km forecasts are forced to start at 0, 3, 6, 9, 12, 15, 18 or 21UTC on any given day to follow the operational system; they are the times for which initial data is available. The 4-km runs are shown in brackets above because the 4-km forecasts will not be examined in this report but were necessary to provide boundary information for the 1-km forecasts.

Figure 1 shows a schematic of the different forecast start times relative to the times of the storms. (This diagram will be reproduced for each of the cases with the specific times and dates for that case). The 12-km model used is the same as the 12-km mesoscale model that was operational at the times of the events. The 1-km model is a still used in a purely research context and was configured to be the same as that implemented in the High Resolution Trial Model (HRTM) suite (Lean et al 2005, Roberts and Lean 2007). Essentially, the 1-km model is the same as used for stage 1. As in stage 1, each of the 1-km forecasts started from 12-km model fields 1-hour into the 12-km model forecast and used the 4-km model to provide information through the boundaries. For example, a 1km forecast starting at 01 UTC would use fields from a 1-hour mesoscale model forecast that started at 00 UTC and the equivalent 4km forecast starting from 01 UTC would supply information through the boundaries during the forecast. The model domains used for each case will be shown in each respective case study.

1.3 Scales of interest

To make a decision about whether a forecast is accurate or not will firstly require a decision about what it is we hope the forecast will achieve. This means deciding on the scale at which we expect a forecast to provide the information we need. We might be interested in the average amount of rain to fall over a particular sized river catchment or it could be the amount of rain at a particular location or less-ambitiously an indication that rain may fall somewhere over northern England this afternoon. If we are going to think about forecast accuracy we need to think in this way – i.e. over what scales does the forecast provide useful information, how does that fit in with my requirements and how should I interpret the forecast. Forecasts are always going to be considered more accurate if we are less demanding about what we want, which put another way means that forecasts will be more successful at forecasting over larger scales than smaller ones. In the context of predicting thunderstorms, it is much easier to predict an area of thunderstorm activity than a particular shower. For this reason a scale-selective technique will be used in the evaluation of the model forecasts in each of the cases. It enables us to discriminate objectively between the forecasts in terms of the scales over which they are sufficiently accurate.

This view of the skill of forecasts over different scales is also useful in another way. The meteorological flow patterns that pre-exist the development of a particular storm, and may be responsible for its development, act on a variety of different scales Larger-scale features such as cyclones and fronts mark out the areas over which thunderstorms are likely to occur or even where extreme storms are possible. This has been shown in the findings of the Extreme Rainfall and Flood Event Recognition project (Hand et al 2004) in which archetypal synoptic situations favourable for particular categories of extreme rainfall events were identified. At the other spectrum, very small scales might also be important. From the same Extreme Rainfall and Flood Event Recognition project (Sleigh and Collier 2004) found that the tipping term in the vorticity equation may be a useful indicator of convective development in a model with a grid spacing of ~1km and recommend that its usage should be 're-visited' in such a model. This is applicable at the scales of the storms themselves, but care must be taken at those scales because any predictability may not last long and forecast accuracy is likely to be poor. At scales in-between there is a whole range of meteorological phenomena that can have an influence on the development of extreme storms, e.g. frontal waves, small cyclones, upper-level vorticity anomalies, jet streaks, storm complexes, slantwise circulations, sea breeze fronts, orographically forced convergence lines. The degree to which these sorts of features are represented accurately in a forecast should have a bearing on the accuracy of the forecast over those scales. Hence, the approach taken in this investigation is to examine the scales at which a forecast was successful or otherwise and then see if it was possible to relate that degree of success to meteorological factors that appeared to be important on that occasion.

2 Case study 1: London Flood 3rd August 2004

2.1 The event

The synoptic pattern is shown in Figure 2. Most of the UK was under the influence of hot humid air from the continent. A cold front through the eastern half of Ireland marked the transition between the hot air and cooler air to the west. In the morning and early afternoon an area of thundery rain moved northwards through some western parts of southern England. During the middle of the afternoon heavy thunderstorms developed further east across southern England. These storms organised into a band along the northern edge of an area of rotating dry air aloft (upper-level vortex). This feature has been marked as a trough line on the synoptic analysis (Figure 2) and will be discussed later. Once the storms had formed they became intense and progressed north into the Midlands as a band moving in association with the upper-level vortex. Over and to the west of London, further storms triggered along northwest-southeast oriented lines behind the main area of activity. These storms too, were very intense and it is the intensity of the rain along with the regeneration of new cells that produced very high rainfall totals and localised flooding.

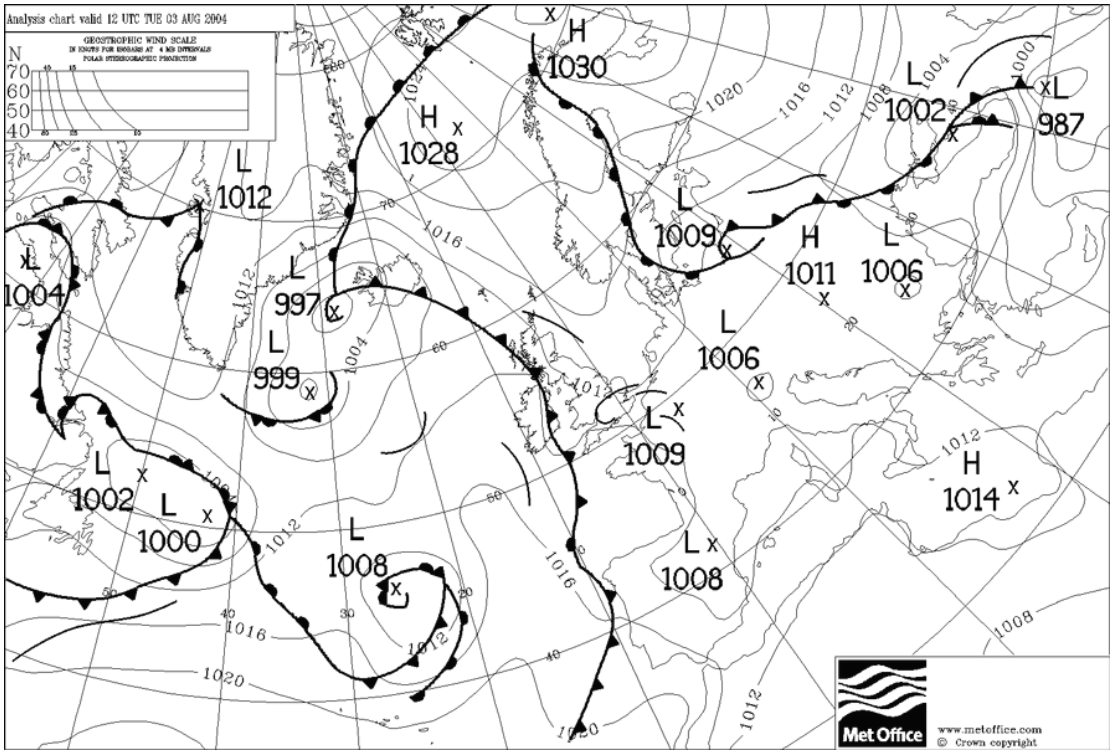

Figure 2. Met Office analysed synoptic chart for 12 UTC 3rd August 2003.

Rainfall accumulations from the network radar processed through the Nimrod system (Golding 1998) are displayed in Figure 3. More than 80mm of rain was measured in parts of northwest London. This was enough to lead to flash flooding and major disruption to commuter traffic. At High Wycombe 42.4mm was collected by a rain gauge in 38 minutes, with peak rainfall rates in excess of 200mm/hour. The value of

42.4mm in an hour gives a return period of ~64 years using the Flood Estimation Handbook (FEH) method. Given that High Wycombe was not in the area of highest accumulations observed by radar, it is likely that more extreme point values would have been obtained in parts of West London if gauge measurements had been obtained.

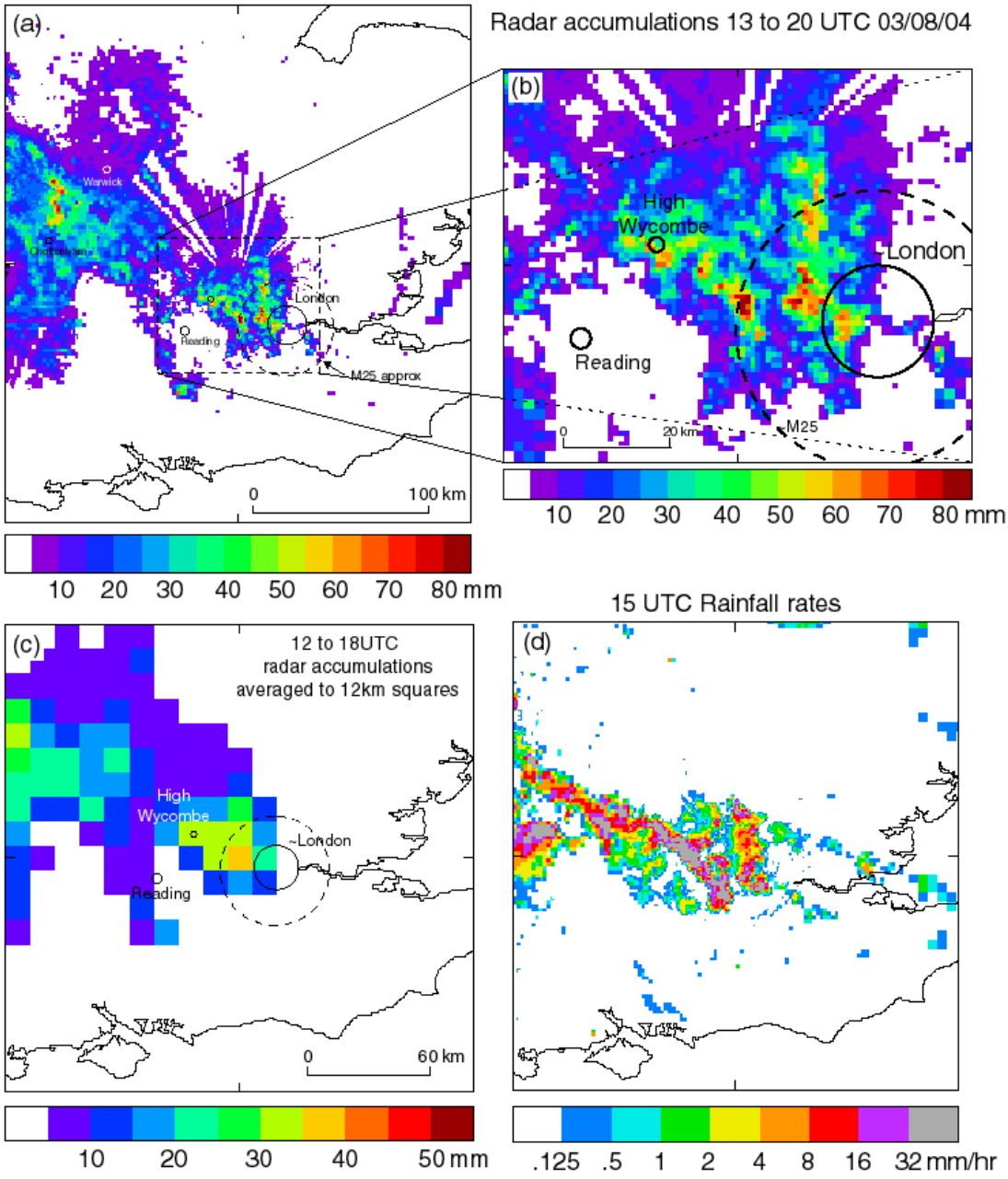

Figure 3. (a & b) Radar accumulations for the period 13 to 20 UTC 3rd August 2004 over Southeast England and a smaller sub area (1km resolution over sub-area). (c) Radar accumulations over the period 12 to 18 UTC averaged to the operational mesoscale model 12km grid. (d) Rainfall rates from radar at 15 UTC on a 1km grid over the area of interest. The circles in (a), (b) and (c) give a guide to the boundaries of inner and outer London.

2.2 The model and forecasts

The model domains used are shown in Figure 4

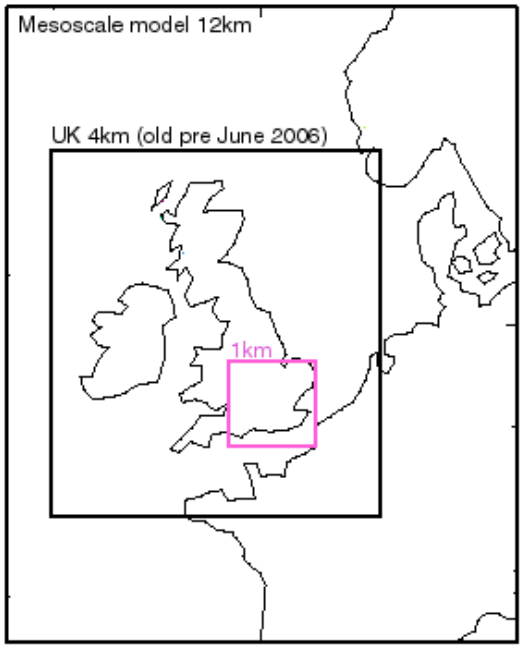

Figure 4. Domains used for the case study.

Forecasts have been run from start times of 18 and 21 UTC on the 2nd and 00, 03, 06, 09 and 12 UTC on the 3rd August 2004 (Figure 5). (The 4 km and 1 km forecasts started 1 hour later using the 12-km information). Each of the forecasts has been given a label shown in Table 1.

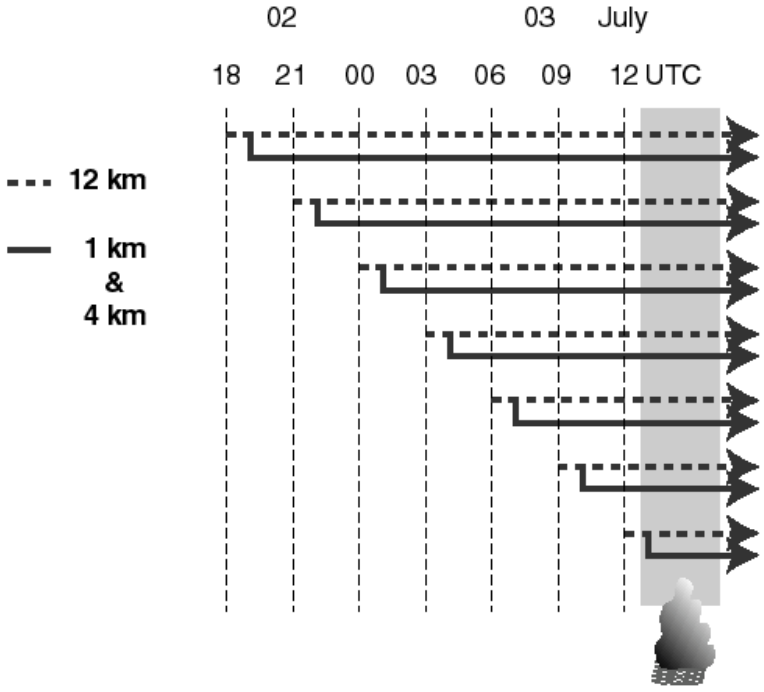

Figure 5. Schematic diagram of the forecasts run for the 3rd July 2004 case.

Start time	12km	1km
18 UTC 02/08/04	L12km18-02	L1km18-02
21 UTC 02/08/04	L12km21-02	L1km21-02
00 UTC 03/08/04	L12km00-03	L1km00-03
03 UTC 03/08/04	L12km03-03	L1km03-03
06 UTC 03/08/04	L12km06-03	L1km06-03
09 UTC 03/08/04	L12km09-03	L1km09-03
12 UTC 03/08/04	L12km12-03	L1km12-03

Table 1. Note that the 1-km forecasts are labelled as starting at the same time as the 12-km forecasts. This is because they share the same analysis time even though the 1-km forecasts begin 1-hour later (see Figure 5).

2.3 The accuracy of the forecasts

2.3.1 12-km forecasts

Rainfall accumulations over the period from 13 to 18 UTC from radar and from several 12-km forecasts are shown in Figure 6. All the accumulations were projected on to the same 5km grid.

The forecast from 18 UTC on the 2nd was clearly the least accurate in terms of rainfall location (panel (b)). It did not produce any rain in the area where most rain fell, although it did generate rainfall totals that had similar values to those observed. The other forecasts were spatially more accurate because they were able to produce a northwest-southeast oriented band of rain >5mm. There appears to have been a sharp change towards an improved rainfall distribution between L12km18-02 and the other later forecasts. The reason for this will be presented later. The other forecasts, although spatially a better fit, did not however necessarily produce more accurate rainfall totals. L12km00-03 and L12km09-03 produced far too little rain. L12km03-03 and L12km06-03 both produced more realistic totals, although still not enough. L12km06-03 was the best 12-km forecast; it gave a good indication of high rainfall totals in the area of interest.

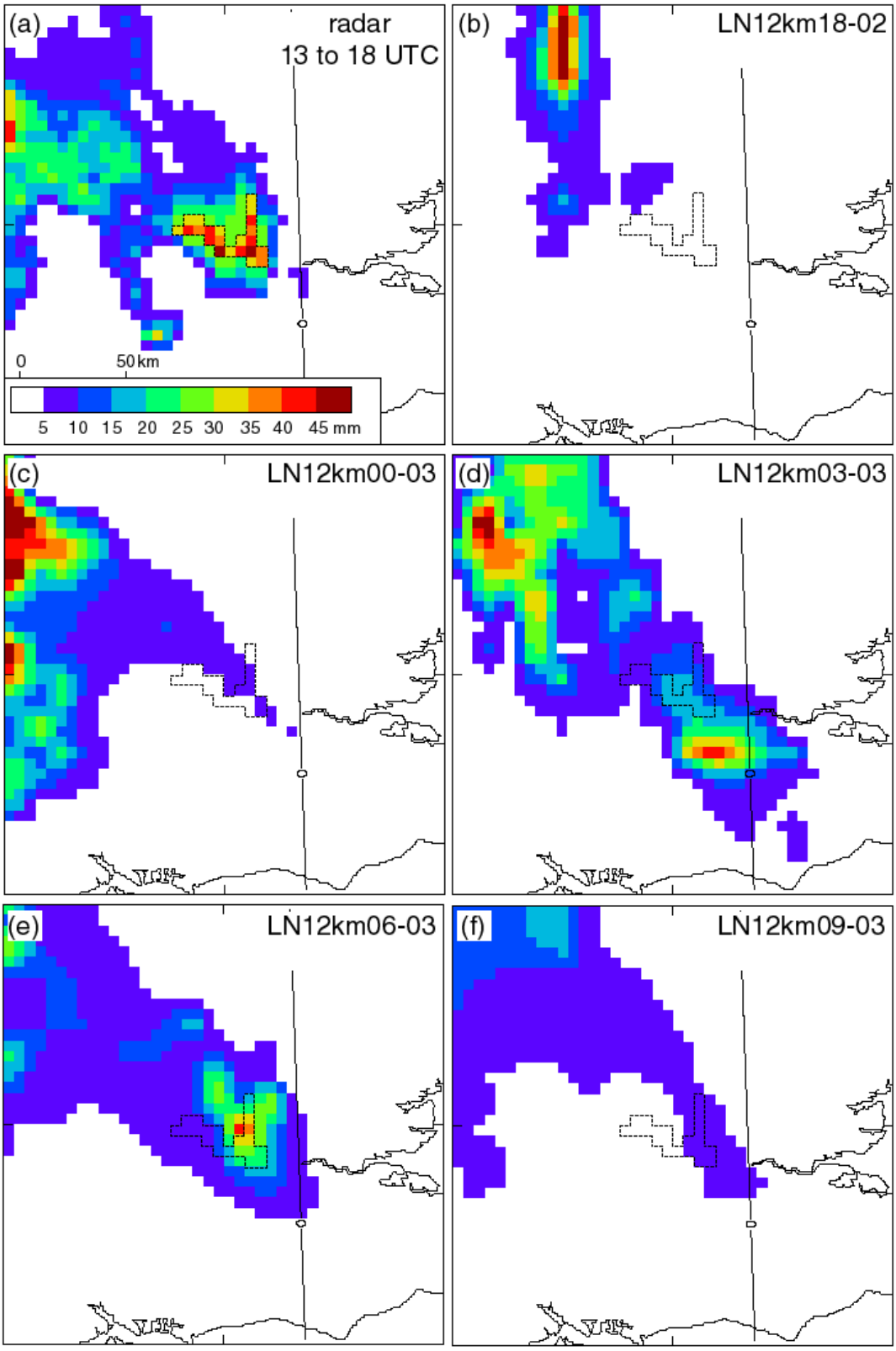

Figure 6. (a) Rainfall accumulations for the period 13 to 18 UTC 3rd August 2004 projected on to a 5-km grid from (a) radar, (b-f) 12-km model forecasts starting from 18 UTC 2nd, 00 UTC 3rd, 03 UTC 3rd, 06 UTC 3rd, 09 UTC 3rd.

2.3.2 1-km forecasts

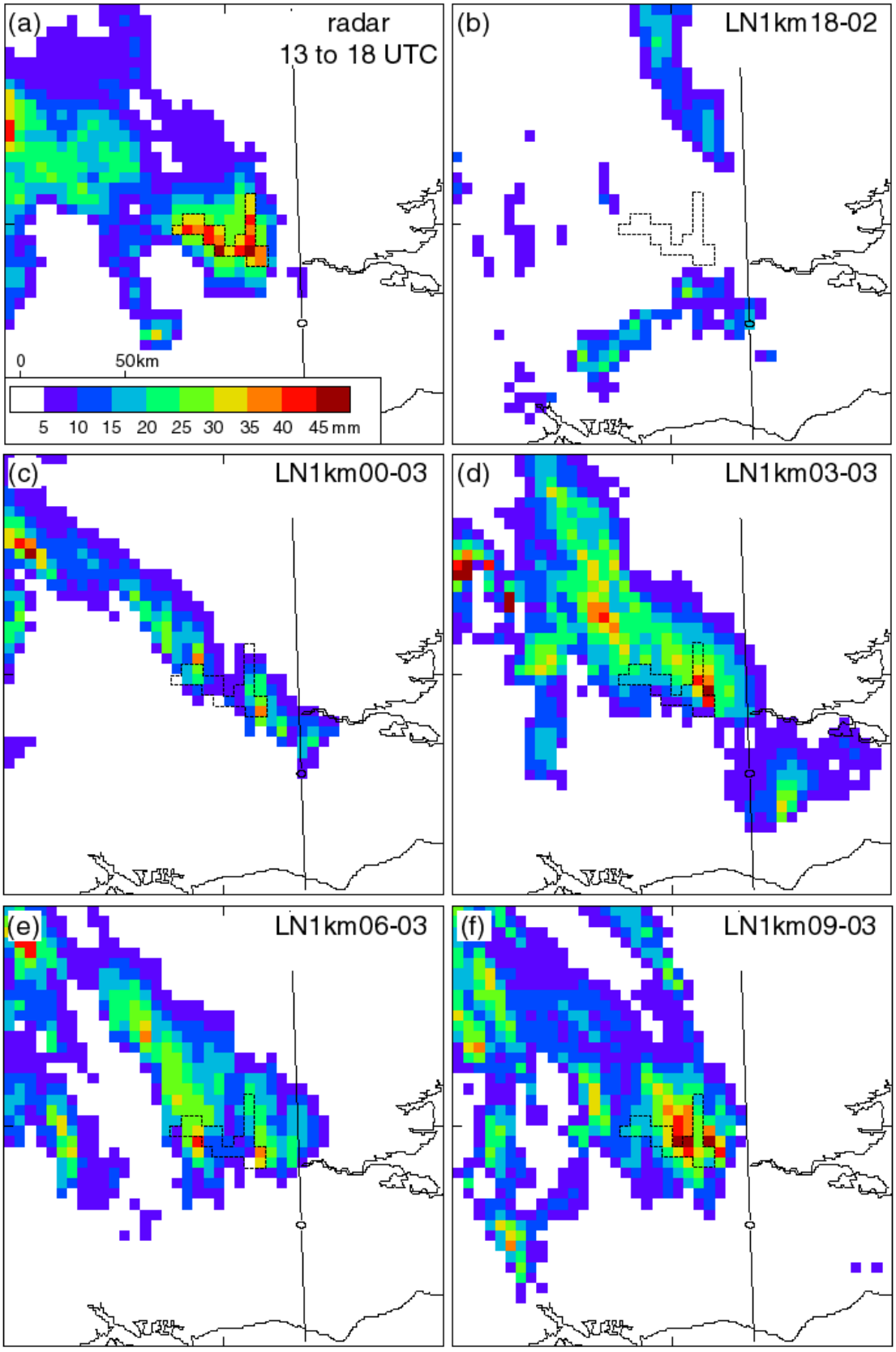

Figure 7. (a) Rainfall accumulations for the period 13 to 18 UTC 3rd August 2004 projected on to a 5-km grid from (a) radar, (b-f) 1-km model forecasts starting from 18 UTC 2nd, 00 UTC 3rd, 03 UTC 3rd, 06 UTC 3rd, 09 UTC 3rd.

Rainfall accumulations over the period from 13 to 18 UTC from radar and from several 1-km forecasts are shown in Figure 7 (equivalent to the 12-km forecasts in Figure 6). All the accumulations were projected on to the same 5km grid. As seen in the 12-km forecasts, the forecast from 18 UTC on the 2nd (L1km18-02) was the least accurate in terms of rainfall location. The other forecasts were spatially more accurate because, like their 12-km counterparts, they were able to produce a northwest-southeast oriented band of rainfall. The improvement between L1km18-02 and the other later forecasts is similar to what was seen in the 12-km model and suggests that a change in the representation of the larger-scale flow was responsible.

The later forecasts were spatially more accurate, and unlike their 12-km counterparts, they also produced rainfall totals very close to those observed by radar. Both the L1km06-03 and L1km09-03 were extremely good forecasts, especially L1km09-03, which gave an excellent prediction of high rainfall totals in the London area.

2.3.3 Objective measure of spatial accuracy

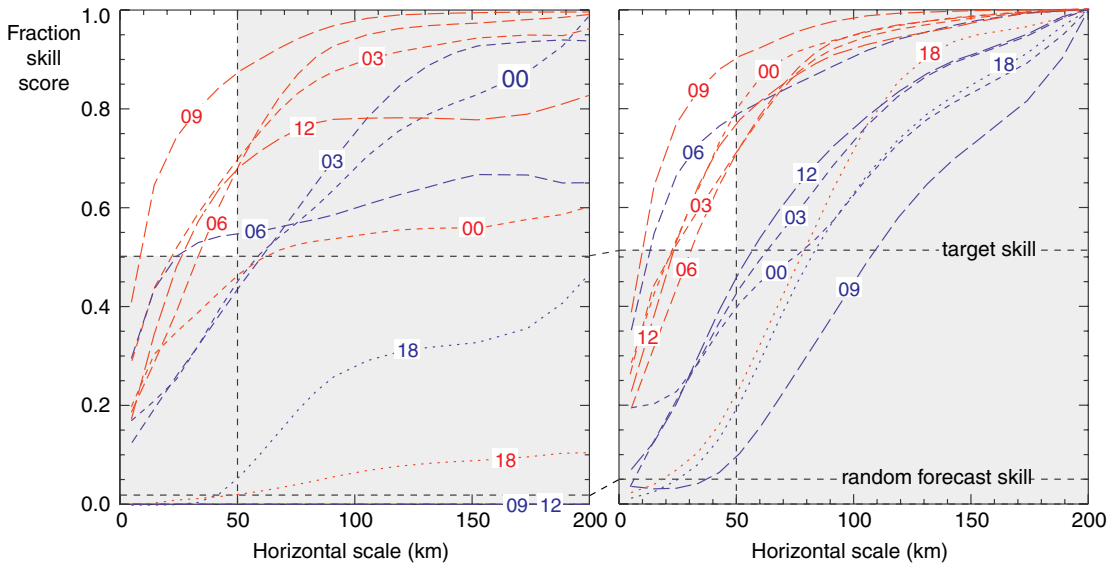

Figure 8. Graphs of forecast skill against spatial scale using the Fractions skill Score (FSS) as a measure for comparing forecast rainfall accumulations against radar. (a) for rainfall pixels exceeding of 30 mm in 5 hours and (b) for rainfall pixels within the top 98th percentile of accumulations over 5 hours. See text for more details. 12-km forecasts are shown with blue lines, 1-km forecasts have red lines. The numbers refer to the start time of the forecast (e.g. 18 = 18 UTC). Forecasts deemed to have sufficient skill have curves that pass through the top-left white section.

A subjective assessment of the 12-km and 1-km forecasts has been given in the previous section. An objective measure of forecast skill is provided here to support the subjective conclusions and quantify the scales over which the forecasts have useful skill. This is done by use of the Fractions Skill Score (FSS), which has been documented in detail in Roberts (2004,) (2005) and Roberts & Lean (2007). The FSS is used to assess forecast skill over different scales by making a comparison

between the fractional coverage of forecast rainfall amounts over varying sized areas (usually squares) with those observed by radar (over the same squares).

Typically, forecasts have lowest skill over small scales (small squares) and skill improves as the square size is increased. This is because small scales are inherently less predictable and small forecast errors can make big differences to the fractional coverage over small areas. We wish to determine the smallest scale/area (e.g. smallest river catchment or forecast region) over which each rainfall prediction has enough skill to be useful. This is done by computing the scale/square size at which a target value of FSS is achieved. This should be made clearer by looking at the graphs of FSS against scale in Figure 8. The FSS has a range of 0.0 to 1.0. A value of 0.0 means no forecast skill. A value of 1.0 means a perfect forecast. The two horizontal dashed lines in Figure 8 represent two reference levels of skill. The lower line is the skill that would be obtained by a purely random forecast. The upper line is our target level of skill and a forecast is deemed to have useful skill above this level. The target is the FSS value that is halfway between the skill of a random forecast and the skill of a perfect forecast. In idealised experiments this target FSS is reached at a length scale that is twice the separation between an observed rain-band and a misplaced forecast rain-band (See Roberts & Lean (2007) for more information). The vertical dashed line marks a scale of 50 km. The value of 50 km was chosen because we would hope that a 1-km model is capable of achieving a sufficient level of skill at scales smaller than 50 km. The un-shaded top left parts of the graphs define where the FSS curves should be found if the forecasts have sufficient skill at scales less than 50 km.

FSS curves for 1 km and 12 km forecasts starting from 18 UTC on the 2nd and 00, 03, 06, 09 and 12 UTC on the 3rd have been plotted. Figure 8(a) shows curves for rainfall accumulations exceeding a 30 mm threshold in the period 13 to 18 UTC on the 3rd. The curves that fall into the un-shaded area (the most skilful) come from the 1-km forecasts starting from 03, 06, 09 and 12 UTC on the 3rd and from the 12-km forecast starting from 06 UTC on the 3rd. These are the forecasts that were also found subjectively to be the most accurate (Figures 6 and 7). The most skilful forecast was the 1km from 09 UTC on the 3rd (L1km09-03), which also agrees with the previous subjective assessment. The 1 km and 12 km forecasts from 18 UTC on the 2nd which were considered the least accurate also have low skill at all scales in Figure 8(a). The 12 km forecasts from 09 and 12 UTC on the 3rd have zero skill at all scales because the 30 mm accumulation threshold was not reached anywhere in the domain. Clearly these forecasts had some spatial skill even though they did not produce enough rain, so in order to focus on the spatial skill, a threshold which compares the locations of the top 5% of rainfall accumulations (rather than a specific rainfall amount) has been presented in Figure 8(b). This essentially tells the same story in that the same forecasts are the most skilful (located in the un-shaded area). The biggest difference is that the 12 km forecast from 06 UTC on the 3rd (L12km06-03) has now become the second most skilful forecast, which makes sense since it had good positioning of the rain, but somewhat under-predicted the totals (Figure 6). The 12 km forecasts from 09 and 12 UTC do now have some skill but are not skilful enough to fall into the un-shaded area.

It may appear surprising that the forecasts from 18 UTC are not the least accurate given that they clearly looked to be the worst. The reason is that when the forecasts were assessed by eye, we looked at the patterns and decided that the 18 UTC forecasts looked less like the radar picture than the other forecasts. We used some knowledge of meteorology. This verification method can not make that kind of judgement, it simply compares the locations of the rain areas, leaving the possibility that a forecast can appear more skilful than it should be if it is ‘lucky’. Nevertheless

the forecasts from 18 UTC did score poorly, as indeed they should have. It is a reminder, however, that any assessment of NWP forecasts should include both a scientific examination of the meteorology alongside measures of forecast skill.

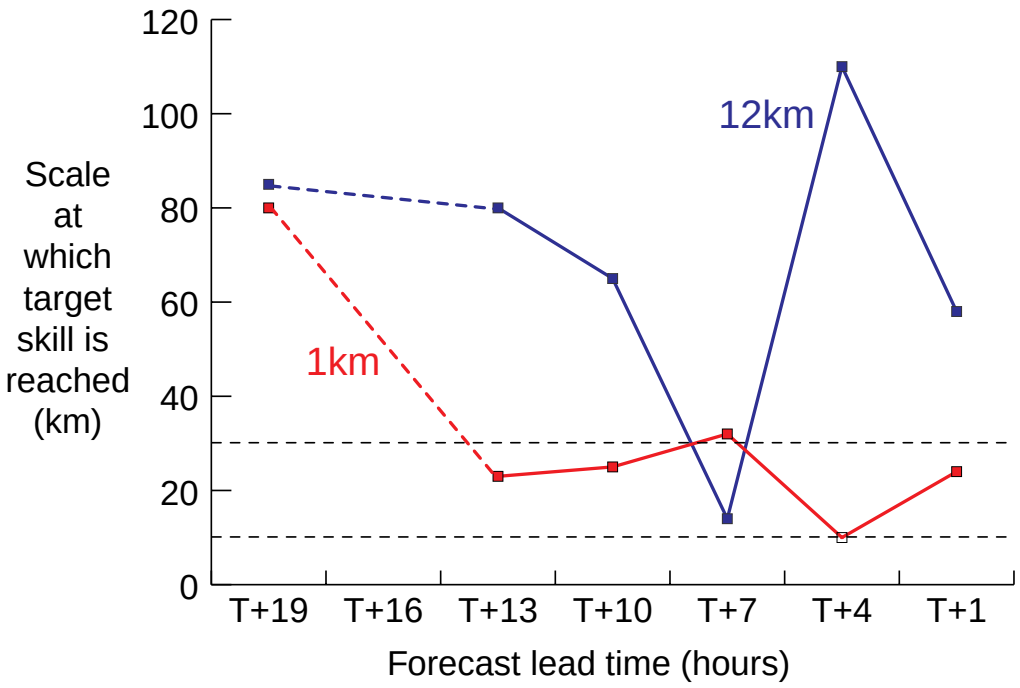

Figure 9. Graph showing the scale at which the target level of skill is achieved for the different forecast lead times for 1-km model forecasts (red) and 12-km model forecasts (blue). The dashed lines mark out scales of 30 and 10 km.

The most important information contained within the graphs in Figure 8 is the scale at which the target level of skill is reached ($scale_{target}$). If that scale is small enough to be comparable to the size of a fast-response river catchment, then that forecast has useful skill. A graph of $scale_{target}$ for each of the forecasts is shown in Figure 9. The 1km forecast from 18 UTC on the 2nd (L1km18-03) only reached a satisfactory level of skill at 80 km, but the forecasts from 00, 03, 06, 09 and 12 UTC were consistently much better and all reached the target skill at scales of around 30 km or shorter, and the 09 UTC forecast (L1km09-03) achieved the target skill at only 10 km. In contrast, the 12-km forecasts were generally poorer and much more volatile. The forecast from 06 UTC (L12km06-03) impressively reached the target skill at only 13 km, but it was the only one to get below 30km and the 12-km model skill oscillated considerably from forecast to forecast.

From the 00 UTC start time onwards, the 1-km model was able to provide a succession of consistently useful forecasts. The 12-km model produced one good forecast, but was unable to give a consistent signal. These results suggest that the spatial distribution of the highest rainfall totals were predictable in this case provided that the model had sufficient resolution, implying that an ability to represent the local dynamics is important. Insufficient resolution resulted in a loss of predictability and an oscillation of solutions.

2.4 Reasons for the differences in forecast accuracy

Now we try to seek an explanation of why some of the forecasts were more accurate than others. The first part of this process will focus on the larger scales and examine how well the forecasts were representing the mesoscale flow pattern by making use of satellite imagery. After that, we will focus on the effect of more local differences between the forecasts and reality as provided by hourly surface reports and radar.

2.4.1 Agreement with larger-scale flow

Each of the 1-km forecasts was initiated from 12-km model fields (1 hour into the 12-km forecast) and the same 12-km forecast provided the 1-km model with information at the boundaries throughout the forecast period. For this reason, the larger-scale flow pattern within the 1-km forecasts (e.g. Atlantic storms, frontal zones) was largely controlled by the 12-km model. This has been shown in Roberts & Lean (2007). Therefore, if we wish to examine whether there are any larger-scale errors present in the 1-km forecasts, we are justified in first looking at the 12-km forecasts. The benefit of doing this is that the 12-km forecasts are less noisy and occupy a larger domain and it is therefore easier to detect larger-scale discrepancies.

Meteosat Water Vapour (WV) imagery can be used to infer the dynamics of the upper-troposphere between 5 and 12km (Weldon and Holmes, Roberts 2000, Browning et al 1996). The WV image for 12UTC, 2-3 hours before the storms triggered, is shown in Figure 10(a). Two features of interest have been highlighted in the picture. Firstly, the oval-shaped blue area ‘D’, enclosed by a black arrow. This is where the WV brightness temperature is warmer (despite being shaded blue!) indicating that the upper-troposphere is relatively dry because of a local lowering of the tropopause (boundary between tropospheric air and drier stratospheric air). Such a lowering is typically associated with a local rotation in the upper-tropospheric flow (a vorticity anomaly), which can lead to vertical motion and preferential regions for triggering convection (Roberts 1995, 2000, Browning et al 1996). Rotation is evident in a playback of this feature in the WV imagery. It is therefore important that the model represents this feature correctly if it is to get the general area of convection correct. Secondly, the orange/red areas bounded by a red line. This is where an area of cloud has formed from ascent in association with the vorticity anomaly and was subsequently drawn around in the rotation.

Temperature and humidity fields from NWP model forecasts can be used to generate pseudo WV imagery for comparison with the satellite WV imagery and is also shown in Figure 10 from several of the 12-km forecasts. This allows a direct comparison between model and satellite imagery. It is clear that the forecast from 18UTC on the 2nd (L12km18-02, Figure 10(b)) does not match the WV image particularly well. The dry (blue) region is too stretched and too far south and the hooked area of cloud too far east. Essentially, the upper-level vortex is in the wrong place and probably has the wrong shape. As we move through to the forecast that initiated 6 hours later (Figure 10(c), L12km00-03) there is a noticeable improvement in the position of both the dry region and the hook of cloud.

If the upper-tropospheric dynamics is important, then we should have observed an improvement in the forecast L12km00-03 compared to L12km18-02. A look back at the rainfall accumulation forecasts (Figure 6) reveals that this was indeed the case; the spatial distribution of the rain became more like that observed. What about the 1-km forecasts? The pseudo WV image patterns from the 1-km forecasts are not shown here because they look qualitatively much the same as their 12-km counterparts, showing that the larger-scale evolution of the 1-km forecasts is controlled to a large extent by their equivalent 12-km forecasts. In other words, the story is the same, we should see an improvement in the 1-km forecast from 00UTC on the 3rd (L1km18-02) compared to the forecast from 18UTC on the 2nd (L1km00-03). Again, we can see from Figure 7 that this is the case. The distribution of the rainfall becomes much closer to that observed in L1km00-03 compared to L1km18-02. In fact, it is more noticeable than that seen at 12km.

There is not much difference in the fit to WV imagery at 12UTC on the 3rd between the forecasts starting at 00UTC on the 3rd and subsequent forecasts starting from 03, 06 and 09 UTC on the 3rd. The forecast starting from 09UTC on the 3rd has the best fit. The graph showing the scale at which useful skill is achieved for each 1-km forecast (Figure 9) shows that the forecast from 09UTC on the 3rd is the most skilful. It also shows that the big change in fit to the WV imagery between the 1-km forecast from 18UTC on the 2nd and 00UTC on the 3rd is reflected in a large improvement in the skill. There are only relatively smaller variations in skill with forecasts thereafter. It appears that in this case the 1-km model forecast skill is strongly related to the accuracy of the larger-scale flow pattern.

The dry region in the WV imagery was associated with rotation in the imagery. Figure 11 shows that there was also rotation in the model. There was a vorticity anomaly at 400hPa at 12UTC in the forecast from the 12-km forecast starting at 09UTC. This is coincident with the location of the dry (blue) region in the WV imagery where a local vorticity maxima is expected to occur. In contrast, the larger values of vorticity are more elongated and located too far southeast in the forecast from 18UTC on the 2nd. This indicates that the differences in the pseudo WV imagery are also reflected in very significant differences in the model dynamics.

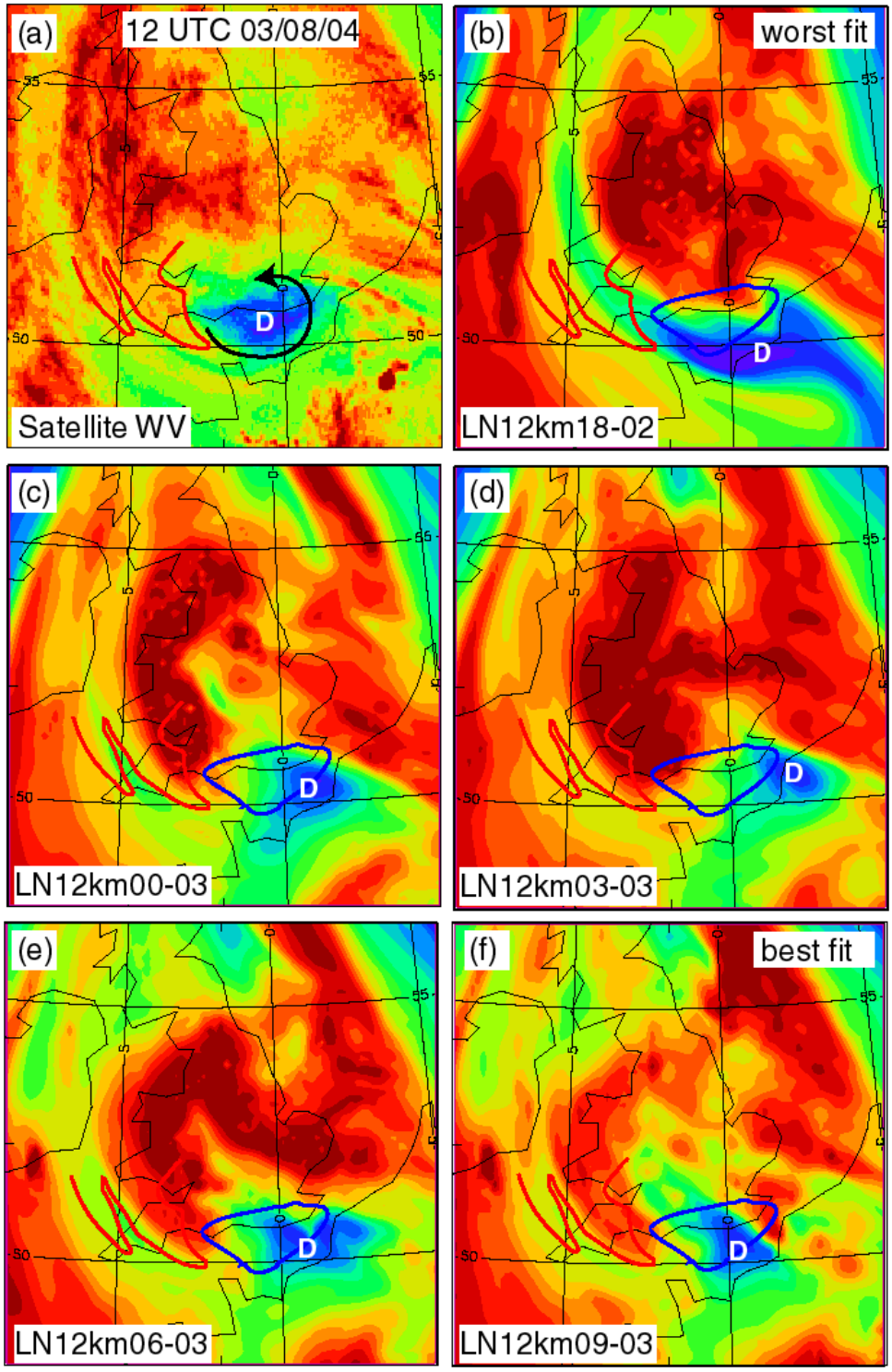

Figure 10. (a) Meteosat Water Vapour image for 12 UTC 03/08/04. Areas with warmer brightness temperatures (dry regions) are blue, areas with cooler brightness temperatures (moist/cloudy regions) are orange/red. 'D' marks the centre of a small dry region which signifies a small region of rotation at upper levels (7-10km). The red lines mark the edge of the moist/cloudy region wrapping round the dry area. (b-f) pseudo Water Vapour imagery extracted from relative humidity fields at 300 to 600 hPa from the 12-km forecasts starting at 18 UTC 2nd, 00, 03, 06 and 09 UTC 3rd August as comparison against the imagery. The blue line markes the dry region in the satellite image. The red line is the same as in (a).

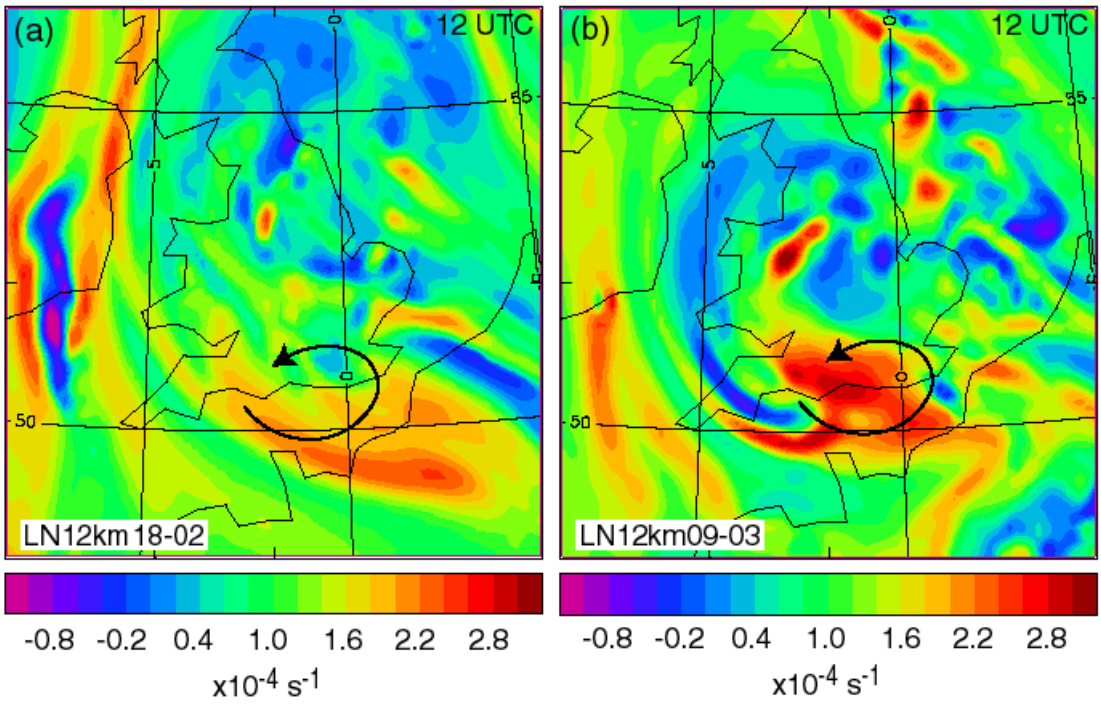

Figure 11. Vorticity at 400 hPa from (a) the 12-km forecast starting from 18 UTC 02/08/04 and (b) the 12-km forecast starting from 09 UTC 03/08/04.

Conclusions about the larger-scale flow

It is necessary for the larger-scale dynamics to be well represented to allow the possibility of a good forecast because the larger-scales control the envelope region within which convection can develop. In this case, when the fit between the WV imagery and the model pseudo WV imagery was poor, the rainfall distribution was also poor. When the fit to the imagery was better, the forecasts produced a better rainfall evolution.

The reason the forecast from 18UTC on the 2nd was so poor was because the upper-level vortex responsible for organising the band of convection was badly misplaced. This occurred in the 12-km model from which the 1-km model was run. It would not be possible for the 1-km model to correct this kind of error in the larger-scale dynamics.

2.4.2 The importance of the local dynamics

We've seen that the position of an upper-level vortex played an important role in the ability of the models to predict the storms. The 1-km forecasts from 00, 03, 06 and 09 UTC were all in reasonable agreement with the imagery and produced reasonable forecasts. However there were more subtle differences in skill between those forecasts, and it was the forecast from 09UTC (L1km09-03) in particular that gave the most accurate local predictions of the rainfall amounts. Why was that?

Surface observations

Figure 12 shows observations from surface synoptic stations at 13UTC, an hour or so before the storms first developed. Most of the southern half of England was under the influence of a warm-moist air-mass near the surface. This is shown by wet-bulb potential temperatures (a combined measure of temperature and humidity) of over 18.5C widely and over 20C in some places. Over Central/Eastern London and along the north Kent coast there was a pool of cooler air with wet-bulb potential temperatures below 17.5C. An examination of the winds (Figure 12(b)) shows that there was a general flow from the west over south-western areas and a flow from the east elsewhere. The region in between, over central southern England, is where these two flows met along two distinct convergence lines (grey shading in Figure 12(b)). One occurred along a wet-bulb potential temperature gradient between the very warm air over central southern England and the cooler air to the southwest. The other occurred along the western and southern edge of the cooler pool of air in the London area. Neither of these convergence lines can be pinpointed exactly because the observation network is not dense enough, but it still provides useful information for comparison with lines of convergence in model forecasts.

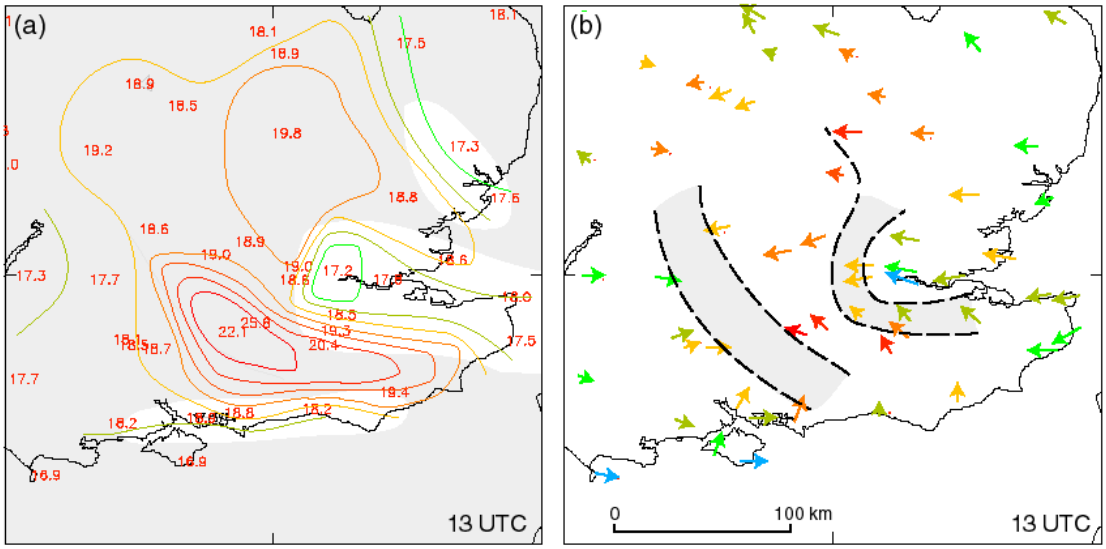

Figure 12. Surface observations from the synoptic network at 13 UTC 03/08/04. (a) Wet-bulb potential temperature displayed and hand contoured. The contouring is more uncertain in the grey shaded areas. (b) Wind vectors shaded according to wet-bulb potential temperature (or likely wet-bulb potential temperature where that measurement was not present). The grey shaded bands are regions of low-level convergence determined from the winds, they are broad because the density of observations is not sufficient to be more precise.

A comparison is first made with the 1-km forecast from 09 UTC on the 3rd because it gave the most accurate rainfall prediction. We can see in Figure 13(a) that the model had the correct range of values for the wet-bulb potential temperature, but the pattern is somewhat different. There was a cooler pool of air of the right dimensions but some 25km too far north. Over the area labelled 'C' the model was colder than observations suggest, because the model had an area of cloud that was not there in reality. It is more difficult to compare other aspects of the wet-bulb potential temperature pattern because the spatial detail that can be extracted from the

observations does not match that of the model. It is of more interest to compare the areas of low-level wind convergence. Around the western side of London the model had a convergence line that was in the same location as that observed within the tolerance of the observations. Further west, the model had two convergence lines, one of which was also correctly located within the tolerance of the observations.

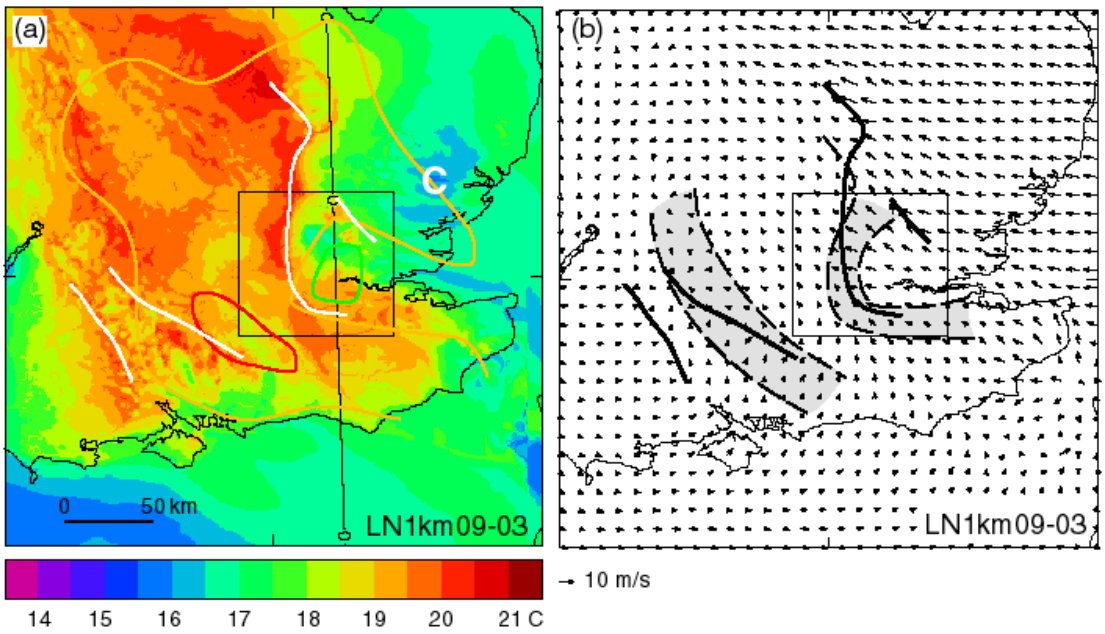

Figure 13. 1-km model forecast from 09 UTC for 13 UTC 3rd August 2004. (a) Colours are Wet-bulb potential temperature at 1000 hPa, White lines indicate low-level convergence of the winds. Coloured contours are the observed wet-bulb potential temperature, green 17.5, orange 18.5, red 20.5°C from Figure 12. The label 'C' is discussed in the text. (b) Wind vectors at 1000 hPa, black line are low-level convergence (same as white lines in (a)), grey shading shows the observed areas within which low-level convergence lines occurred (from Figure 12). The black square is the zoomed area shown in Figure 15.

The importance of the low-level convergence line

The question then, is, how important were these convergence lines? Low-level convergence is associated with ascent. In potentially convective situations this ascent may be the ingredient required for convection to preferentially trigger at that location. Figure 14 shows the observed and forecast convergence lines at 13 UTC before convection in the areas of interest had triggered and superimposed on that the observed and forecast rainfall 30 minutes later. It is striking that both the forecast and observed storms triggered along their respective convergence lines. It appears that in this case, if the convergence lines are correctly represented, and given that everything else is broadly correct, the storms will initiate along those convergence lines. So part of the key to getting the detail in the rainfall correct is to get the local dynamics accurately represented.

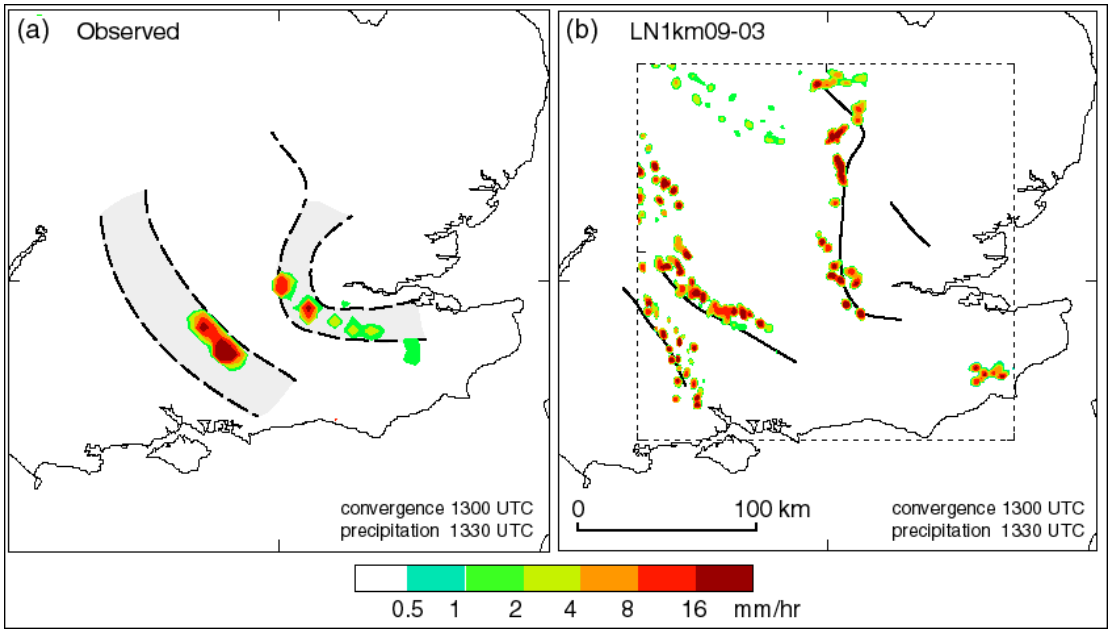

Figure 14. (a) Grey shading shows the observed bands of low-level convergence (from Figure 12) at 13 UTC 3rd August 2004. Shading is the new rainfall 30 minutes later from radar. (b) Black lines are lines of low-level convergence at 13 UTC in the 1-km forecast from 09 UTC (from Figure 13). Shading is the new rainfall in the same forecast 30 minutes later (within the dashed box).

Now we compare the near-surface winds and convergence zones from the 1-km 09 UTC forecast (L1km09-03) with those from the 1-km forecast from 06 UTC (L1km06-03) and the 12-km forecast from 09 UTC (L12km09-03) (Figure 15). The sharp convergence line evident in the L1km09-03 was not present in L1km06-03 primarily because of a different distribution of the easterly winds. The differences in convergence of the low-level wind between the two forecasts can explain why there were differences between the forecasts of rainfall amounts over the London area, even though the overall evolution of the forecasts was similar. The 12-km forecast had no convergence line because it was incapable of generating such a feature with such a coarse grid. This demonstrates that there are important meteorological pre-cursers to convection that a 12-km model is simply unable to adequately represent.

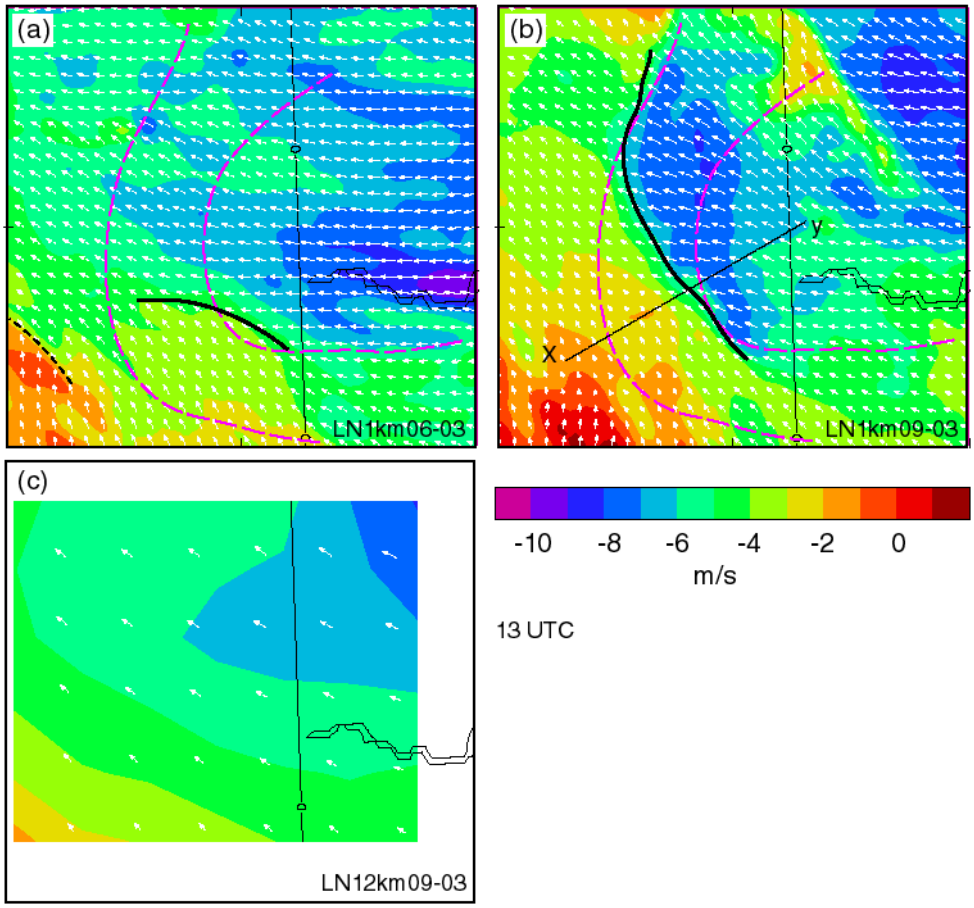

Figure 15. 13 UTC 3rd August 2004. Colours are the u component of the wind at 10 metres, black line are lines of convergence of the winds are 10 metres from (a) 1-km forecast from 06 UTC, (b) 1-km forecast from 09 UTC, (c) 12-km forecast from 09 UTC. Pink dashed lines enclose the observed area within which there was convergence

The role of the low-level convergence line

This section is included to show how the low-level convergence line was responsible for the initial convection. It uses the 1-km forecast from 09 UTC (LN1km09-03). Cross section XY (location shown in Figure 15(b)) is perpendicular to the convergence line at 14 UTC and goes through one of the newly developing storms. Figure 16(a) shows the wet-bulb potential temperature. The atmosphere was potentially unstable to convection developing at that time because the wet-bulb potential temperature at mid-levels (2-8 km) was lower than at low levels (< 1-2 km). This had to have been the case or convection would not have been possible, but we can also see that there was a great deal of potential instability because the difference was up to 5°C. That is enough for severe convective storms. At one particular location (around 28 km from X) a shower had started to develop and the cloud had reached a height of around 4 km. So why had storms not popped up all over the place given that the atmosphere was so potentially unstable? The reason they hadn't is that although potential instability is a pre-requisite for convection, a potentially unstable atmosphere does not necessarily have to trigger convection. There was something inhibiting showers from developing. Figure 16(b) shows that there were two layers where the atmosphere was more 'stable'; the lower one at around 3 km,

the higher one at around 4-5 km. These layers acted like 'lids' or barriers to convective clouds and for the clouds to 'break through' something else needed to happen such as additional heating of the land surface or ascending air motion.

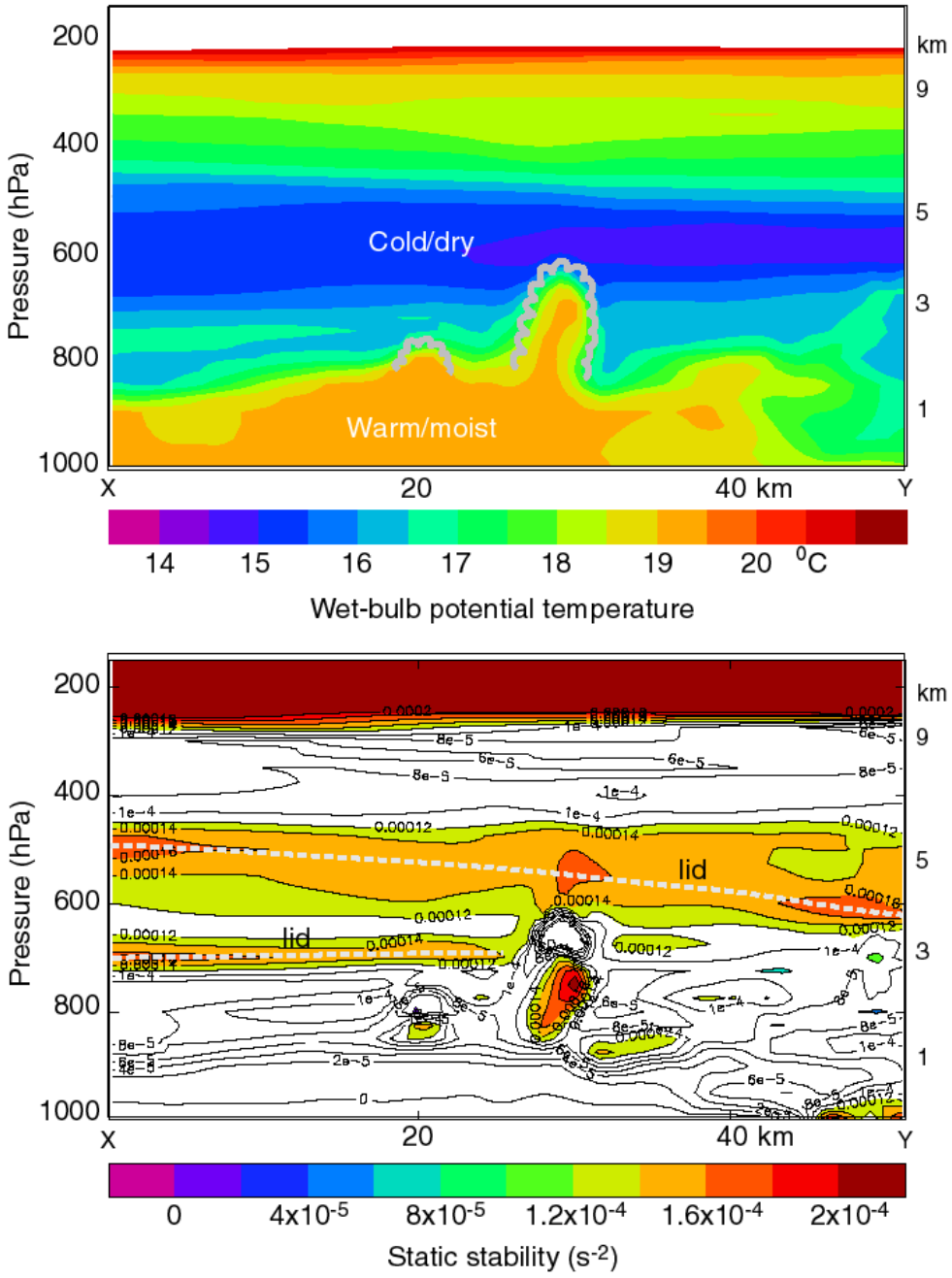

Figure 16. Cross-section along XY in Figure 15 at 14 UTC 3rd August 2004 from the 1-km forecast from 09 UTC. (a) Wet bulb potential temperature. Grey scallop marks schematic cloud edge. (b) Static stability. Dashed lines mark the two bands of high static stability – or 'lids'.

Figure 17(a) shows the flow along the section. At the location of the cloud (~28km from X) the low-level flow from right to left decreased from >6 m/s to <4m/s over a few kilometres. This is the low-level convergence line seen in Figures 12 to 15. The convergence line was associated with vertical motion and it was that vertical motion that either weakened the lid or allowed clouds to form that broke through the lid. Care

is needed here; in Figure 17 we see the low-level flow and vertical motion once the cloud has formed and it could be argued that the convergence is the result of the ascent in the cloud rather than visa versa. However, we have already seen from Figure 14 that the convergence line did pre-exist the storms and examination of nearby cross-sections (not shown) confirms that there was (weaker) ascent associated with the convergence line.

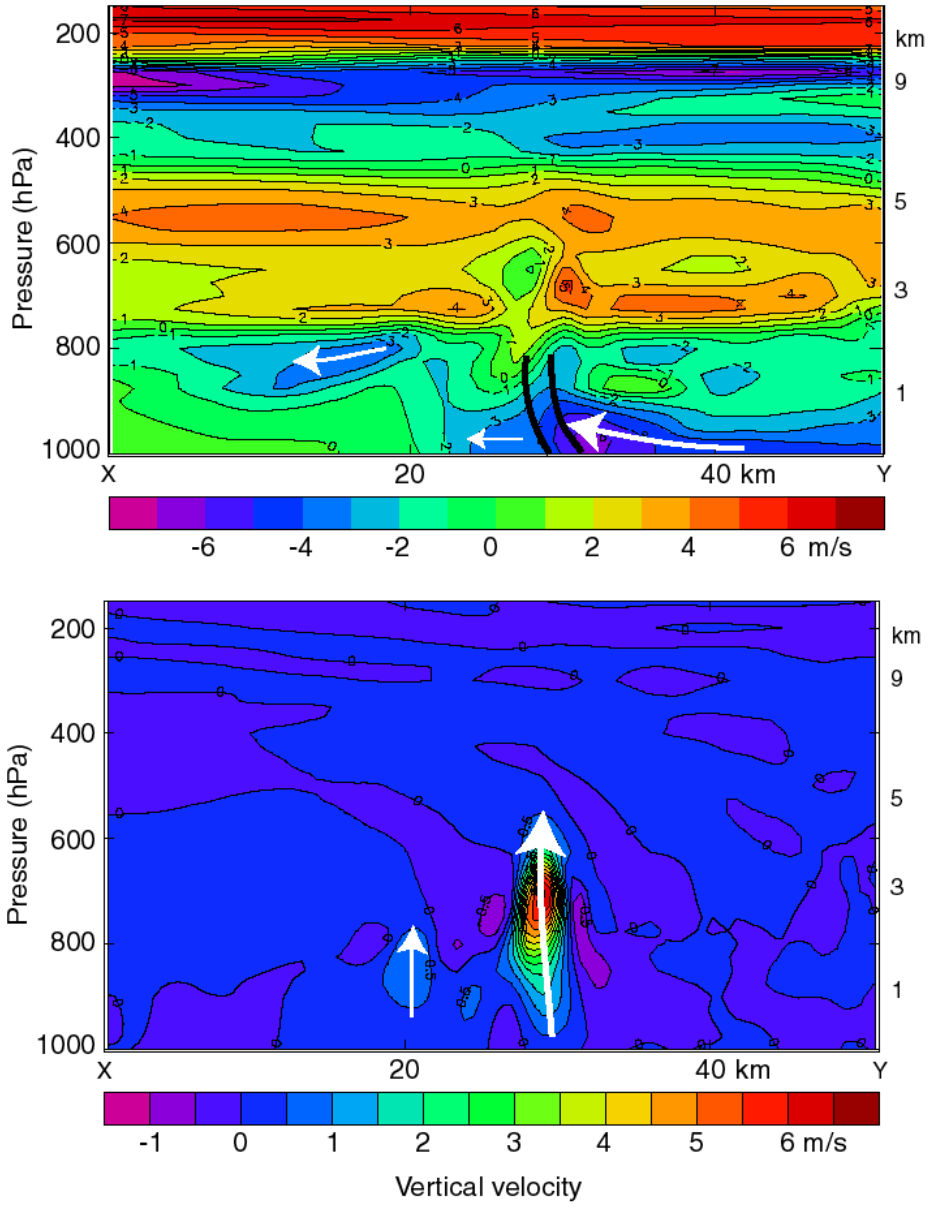

Figure 17. Cross-section along XY in Figure 15 at 14 UTC 3rd August 2004 from the 1-km forecast from 09 UTC. (a) Shading is the along-section wind. Arrows indicate strongest flow from right to left (northeast to southwest). The black lines enclose the zone of largest convergence of the wind. (b) Vertical velocity. Arrows indicate the ascent regions.

Figure 18 is a composite of Figures 16 and 17. The storm was located above the convergence line and had broken through the lower lid, but not yet the upper lid. The fully-developed storms did break through both lids and grew to 9-10 km.

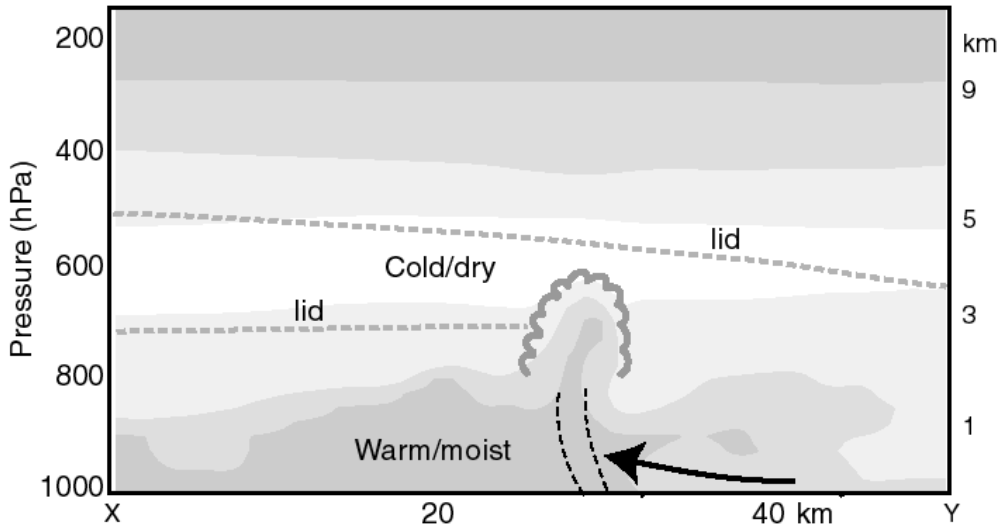

Figure 18. Cross-section along XY in Figure 15 at 14 UTC 3rd August 2004 from the 1-km forecast from 09 UTC. Composite of the important features from Figures 16 and 17.

The origins of the convergence line

Since the convergence line was such an important component in the initial development of the storms to the west of London, it is worth finding out where it came from. At first sight it looks like it may have been to do with the coastline, and Thames Estuary in particular, since the flow was generally from the southeast. In fact, it appears to have occurred as the result of some earlier convective showers.

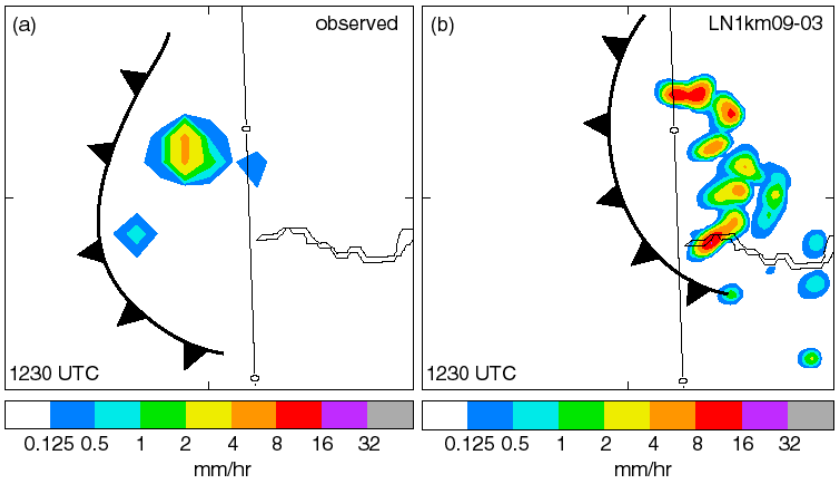

Figure 19. Rainfall rate at 1230 UTC 3rd August 2004 from (a) radar, (b) the 1-km forecast from 09 UTC. The cold front symbol (line with triangles) marks in (a) the approximate edge of the observed convergence line and (b) the forecast convergence line.

At 1230 UTC (30 minutes before the convergence line was analysed in Figure 13) there were some fairly light showers in the London area. A cluster of showers was also present at the same time in the 1-km forecast from 09 UTC (LN1km09-03) that produced the convergence line, but not in the 1-km forecast from 06 UTC that did not. The observed rain at 1230 UTC and the rain from the 09 UTC forecast are

shown in Figure 19, along with a schematic indication of the observed and forecast convergence lines at that time. The convergence line was an outflow boundary from the cold pool created from the downdrafts of the showers. This can be seen by taking a cross-section through the cold pool produced by the showers. The location of the cross-section in relation to the easterly component of the wind and the precipitation is shown in Figure 20 for 1200 UTC (when the cold pool had become well developed). The cross-section itself is shown in Figure 21.

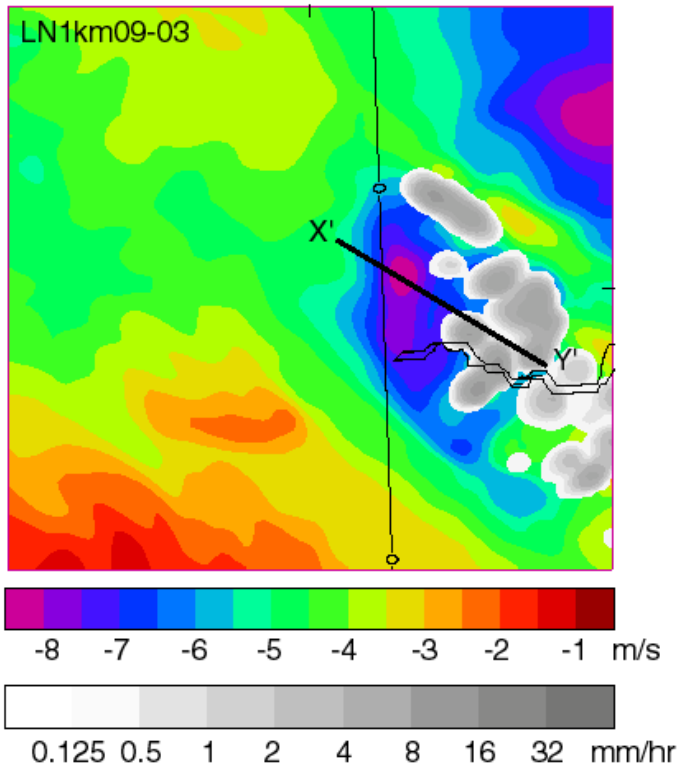

Figure 20. 12 UTC 3rd August 2004. 1-km forecast from 09 UTC. Colours show the u (east-west) component of the wind at 10 metres and grey shading is the surface rainfall rate.

The cross-section shows the difference in wet-bulb potential temperature between 12 UTC when the cold pool was established and 11 UTC when the showers had not yet formed. The first thing to notice is the increased wet-bulb potential temperature where there was a shower at 13 UTC. The cloud base was around 700 hPa, which means that it was a quite a high level and could be termed 'mid-level' convection. It seems to have been associated with a layer of moister air that had advected into the area from the south. Beneath the shower is a region of decreased wet-bulb potential temperature. This is where the cold pool has formed from the downdrafts of several of the showers. The arrows mark the flow associated with the cold pool. Region 'A' is the downdraft linked to the shower shown in the cross-section. Here, the air was descending at more than 5m/s in places. The downdraft was maintained by cooling of the air as rain evaporated below the cloud. Region 'B' is the flow generated from the combined effect of several downdrafts like 'A'. The convergence line was formed at the leading edge of flow 'B'. The decreased values of wet-bulb potential temperature in the cold pool are the result of lower wet-bulb potential temperature air being brought down in the downdrafts.

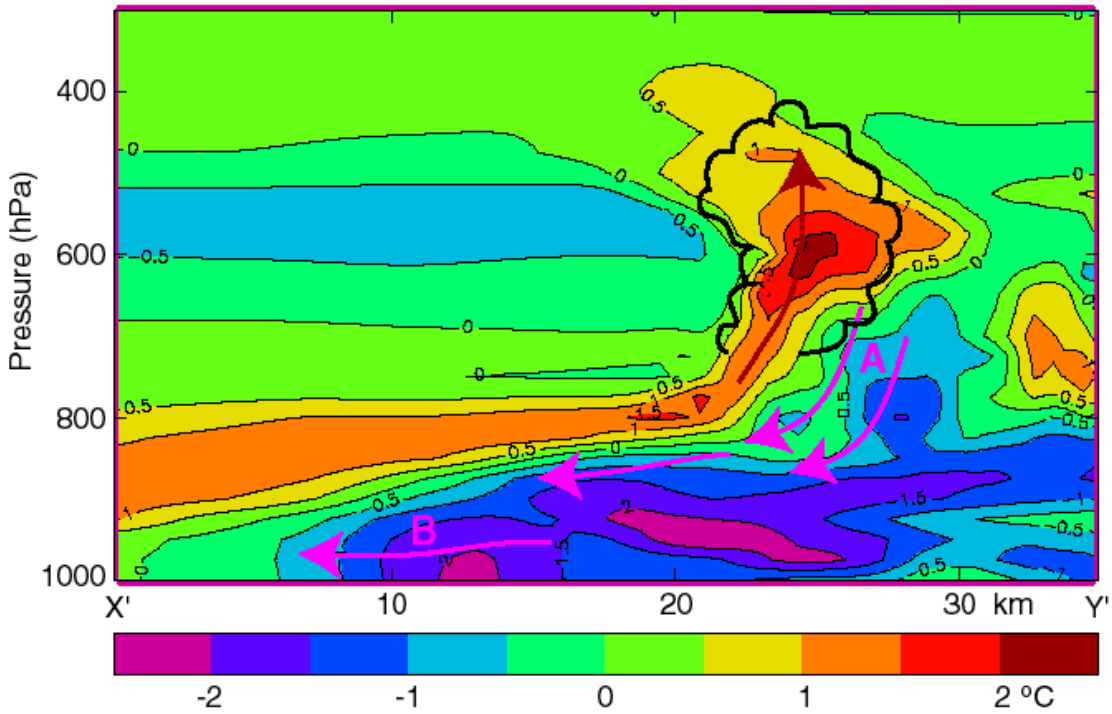

Figure 21. Cross section through X'Y' in Figure 20. Wet-bulb potential temperature at 12 UTC minus the wet bulb potential temperature at 11 UTC from the 1-km forecast from 09 UTC 3rd August 2004. Scalloped line marks the edge of the cloud at 12 UTC. The brown arrow depicts the vertical motion into the cloud. The purple arrows depict the descent beneath the cloud 'A', and the cumulative outflow from several showers 'B'.

The mechanism for the formation of the convergence line has several implications for forecasting the storms that produced the floods. Firstly, it means that the storms were inherently less predictable than at first thought because the location of the triggering of the storms depended on getting some earlier showers correct. The showers in the 09 UTC run that lead to the cold pool and convergence line developed from a layer of moister air that was not present in the 06 UTC run (the run that did not produce the showers). The implication is that, in this case, the additional information about humidity (and perhaps other dynamical aspects) in the later forecast was crucial for giving a more accurate local forecast.

We should notice something else; the showers that lead to the cold pool in the 09 UTC forecast were not as accurately predicted as the location of the subsequent convergence line would suggest. Figure 20 shows that the showers were predicted to be further east than observed, they were also too extensive and probably too heavy. The 'probably' is because there is some uncertainty in the radar rainfall rates in this type of situation. The reason the convergence line was positioned reasonably correctly in the forecast can be attributed to the over-extensive and heavy showers producing a more extensive cold pool and therefore pushing the convergence line further west in relation to the precipitation. This suggests that in addition to getting the location of the showers correct, the characteristics of the showers and the physical processes that lead to the formation of the cold pool also need to be properly represented. There may well be important sensitivities to the way the cloud microphysics is represented in the model and in the way turbulent mixing is applied in

cloudy environments that could have an impact on the shower intensity and downdrafts that lead to the formation of the convergence line. It may have been somewhat fortuitous that the stronger outflow compensated for the error in the location of the showers in the forecast from 09 UTC. The possible variability associated with uncertainties in the representation of cloud and precipitation is worth investigating.

The discussion above raises questions about the inherent predictability of the London storms, i.e. can small changes in sensitive parts of the model formulation (e.g cloud microphysics) or small changes in the initial state (e.g. relative humidity) make a significant difference to the prediction. By significant differences is meant changes to the scale at which the target skill is achieved of around 10-20 km or more. These are important questions and the answer is that we don't know at this stage. However, the evidence from all the forecasts taken together suggests that the storms were reasonably predictable in spite of the dependence of one of the convergence lines on the first initiation (Figure 9). Whilst it is true that the 1-km forecast from 06 UTC (LN1km06-03) was not as accurate as the 1-km forecast from 09 UTC (LN1km09-03), it was still a good forecast even though it did not have the convectively generated cold pool. There are two reasons for this. Firstly, although the cold-pool convergence line was involved in the initial triggering of the convection in the London area, other convergence lines developed later anyway as a result of convection associated with the larger-scale dynamics and these helped to re-trigger new lines of storms. Secondly, the larger-scale dynamics responsible for the broader area of storm activity did not change much from forecast to forecast (from 00 UTC forecast onwards) and was the primary reason for the envelope of convective activity.

2.4.3 More about the larger-scale dynamics.

The Meteosat Water Vapour imagery showed a lens shaped dry area (or 'dark zone') (Figure 10) that signified a local area of rotation in the upper-troposphere (~7-10 km up) moving northwards. This is a so-called vorticity (or potential vorticity) anomaly. These features are often associated with a region of ascending air ahead and descending air behind (Hoskins et al 1985) and this one fits that pattern. This ascent is not convection (which would be several metres per second); instead it is slantwise ascent of a few tens of centimetres per second. Figure 22 shows a band of ascent at 2-3 km ahead of the vorticity anomaly, taken at 13 UTC from the 12-km forecast starting at 06 UTC. A 12-km forecast was used to show this because the equivalent 1-km forecast already had noisy convective updrafts that contaminate the signal of the more gently ascending layer of air. But the 1-km model still contains the same larger-scale dynamics. The 06 UTC 12-km forecast was chosen because it gave the best larger-scale evolution of the storms at both 12 km and 1 km (despite the 1-km 09 UTC forecast giving the most accurate prediction of accumulations for the London area). Continuing to look at Figure 22, we see that in conjunction with the region of ascent was a somewhat larger region in which the atmosphere was potentially unstable to convection. The same area of potential instability was seen in cross section XY in Figure 15 and in that cross-section there was a stable region or 'lid' that was inhibiting storm development. Then it was the low-level convergence that provided the vertical motion which enabled the lid to be broken. In this larger-scale picture it is the ascent associated with the vorticity anomaly that weakens the lid and provides more favourable conditions for convection to develop. The storms develop within the band that has both ascent and potential instability (Roberts 2000).

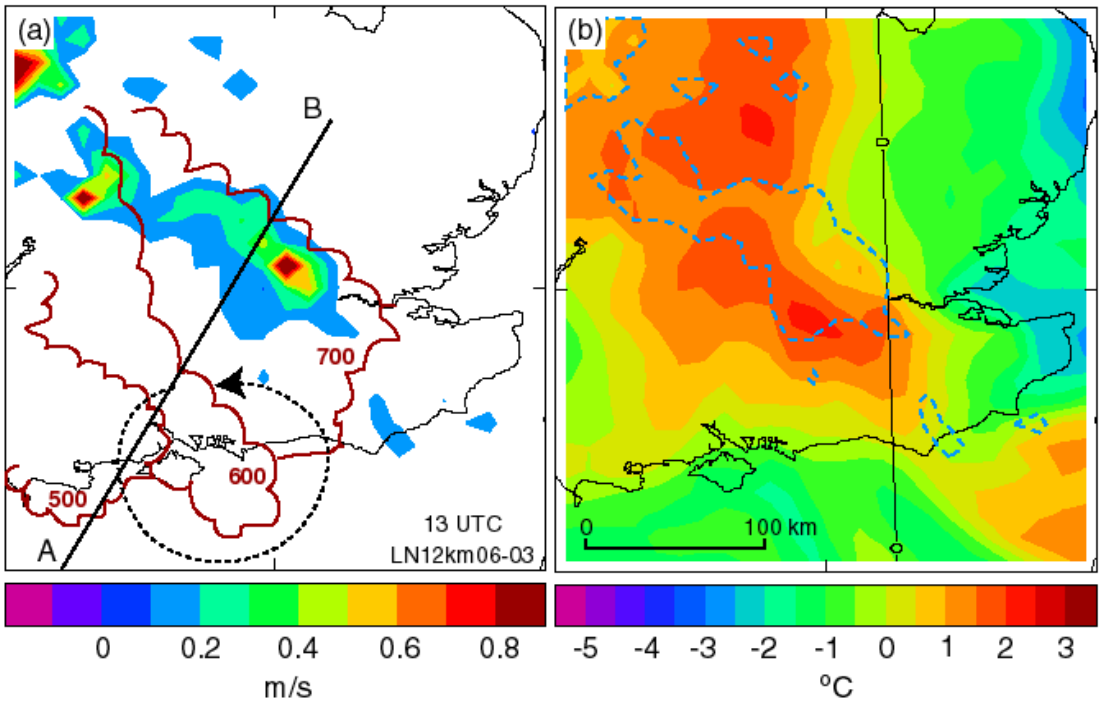

Figure 22. 12-km forecast from 06 UTC for 13 UTC 3rd August 2004. (a) Colours are vertical velocity at 750 hPa, dashed circular arrow depicts the location of the upper-level vortex at around 400 hPA, the brown line encloses areas with relative humidity greater than 80% at 700, 600 and 500 hPa, line AB marks the position of the cross section in Figure 23. (b) The colours show a measure of the possible strength of convection (by subtracting the wet-bulb potential temperature at 650 hPa if the air is saturated from the wet-bulb potential temperature at 975 hPa – orange /red indicates the possibility of strong convection, the dashed blue line is the 10cm/s ascent contour shown in (a).

The cross-section AB has been marked in Figure 22(a) and this is shown in Figure 23 to give a more detailed view of what was happening. Also marked on Figure 22(a) is the sloping outline of cloud/moist air from 700 to 500 hPa. The cloud at 500 hPa (and higher) is the orange/red area outlined in the Water Vapour imagery (and model pseudo imagery) in Figure 10 wrapping around the southwest flank of the lens-shaped dry region. This is picked out in the cross section as the high relative-humidity air sloping upwards from 700 hPa / 150 km to 400 hPa 25 km. Above that moist ascending layer was a drier descended layer associated with the upper-level vorticity anomaly (top +PV on the cross section). The slanted ascending cloudy layer was generated in association with the upper-level vortex and in the process of generating this cloud another potential vorticity anomaly (+PV) was generated, which like the one above, could have the effect of inducing ascent ahead of it as it travelled north.

So what we see is a more complicated picture than the single vorticity anomaly that might be assumed from the Water Vapour imagery. Instead there were two slantwise circulations and two Potential Vorticity anomalies, but the affect was essentially the same as if there were one: they were both associated with ascent ahead (arrows in Figure 23) that acted to de-stabilise the atmosphere.

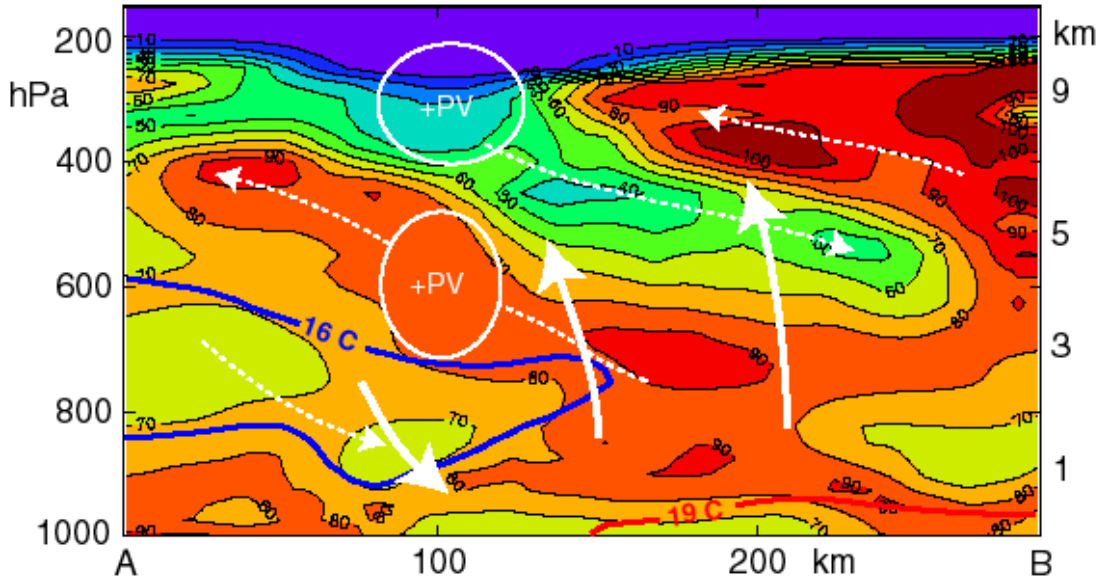

Figure 23. Cross section along the line AB in Figure 22 for 13 UTC 3rd August 2004 from the 12-km forecast from 06 UTC. Colours are relative humidity. Dashed arrows show recent gentle slantwise flow. Solid thick arrows show stronger ascent and descent at 13 UTC. The red/blue contours are wet-bulb potential temperatures of 19 and 16°C. The white ovals enclose areas of locally high potential vorticity (the vorticity anomalies discussed in the text), the upper one is of stratospheric origin, the lower one comes from heating due to the formation of cloud.

The reason for describing some of the details of the larger-scale dynamics is to show that the dynamical interactions that need to be represented can be complex and this can only be done with a Numerical Weather Prediction model. We have seen that the larger-scale dynamical processes can be sufficiently well simulated by a model with a grid spacing of 12 km, as long as the initial conditions at the start of the forecast are sufficiently close to reality. The 1-km model is required for the representation of the convection itself within the context of the larger-scale flow (and the feedback on to the larger scale flow from the convection). Since the 1-km model uses information from the 12-km model at its boundaries (and at the start of the forecast in the setup used here), it is essential that the 12-km model has a good representation of the atmosphere at the start. Otherwise the subsequent evolution of the larger-scale dynamics will be wrong and the 1-km model will then add the wrong detail.

2.5 Summary

The 1-km model was able to produce a good forecast of the convective storms that lead to flooding in parts of northwest London. An objective measure of forecast skill shows that the best forecast (LN1km09-03) was able to achieve the target level of skill over a spatial scale of ~10 km for the higher rainfall accumulations. Several of the forecasts were able to achieve the target level of skill at a scale of ~30 km or less. The ability of the model to predict the storms was dependent on getting two components of the flow correct in the lead up to the storms.

The 1-km model significantly out-performed the 12-km model for all but one of the equivalent forecasts. The 1-km model was more consistent from forecast to forecast than the 12-km model because of a more accurate representation of the storms by the model dynamics. The variability of the 12-km forecasts came from a flip-flop

between how much the convection was ‘resolved’ on the grid or represented by the convection parametrization scheme.

2.5.1 Components

Component 1 – The larger-scale dynamics.

The upper-level vortex and associated dynamics needed to be correct. A comparison of the forecasts with Meteosat Water Vapour imagery (Figure 10) revealed that when the model had a lens-shaped dry region (300 km across) and associated wrap-around of cloud in approximately the right place, a good 1-km forecast was possible. When there was a large miss-match the forecasts were poor. The dry region in the imagery signified the presence of an upper-level vortex. As this vortex moved north, there was a region ahead of it in which ascending motion lead to a destabalisation of the atmosphere (which was already potentially unstable) and provided conditions favourable for convective storms to develop (Figure 22). Thunderstorms developed in a band ahead of the upper-level vortex and moved north, both in reality and in the better forecasts.

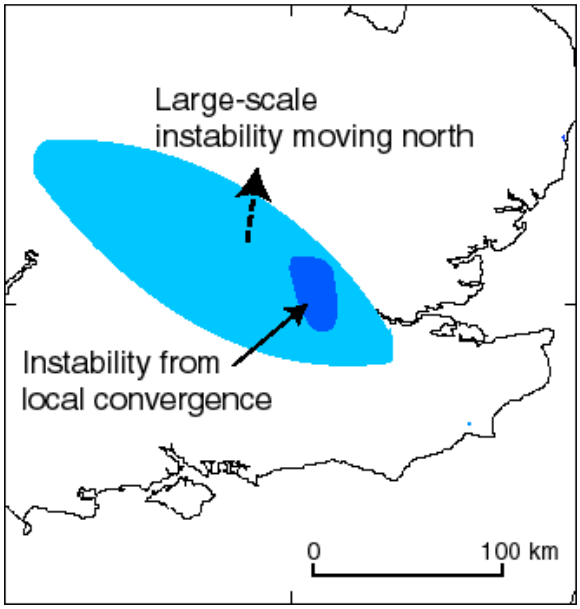

Figure 24. A schematic diagram of the two main components responsible for the development of the storms on the 3rd August 2004.

Component 2 – The local-scale dynamics.

Figure 24 depicts the two components that lead to the large rainfall totals in the London area. The development of storms over the London area involved not only the larger-scale forcing described above, but also initiation along a region of convergence of the low-level winds. The best 1-km forecast (from 09 UTC, LN1km09-03) was more accurate than its predecessor (from 06 UTC, LN1km06-03) because it had a low-level convergence line in approximately the correct location just prior to the time the storms triggered, whereas the previous forecast did not. The

convergence line appears to have been caused by the outflows from earlier less-intense showers. Further storms were triggered in the same area along new convergence lines even when the area of larger-scale forcing had moved away.

The key to getting the most accurate forecast was to have both components in place.

3 Case study 2: Hawnby Flood 19th June 2005

3.1 The event

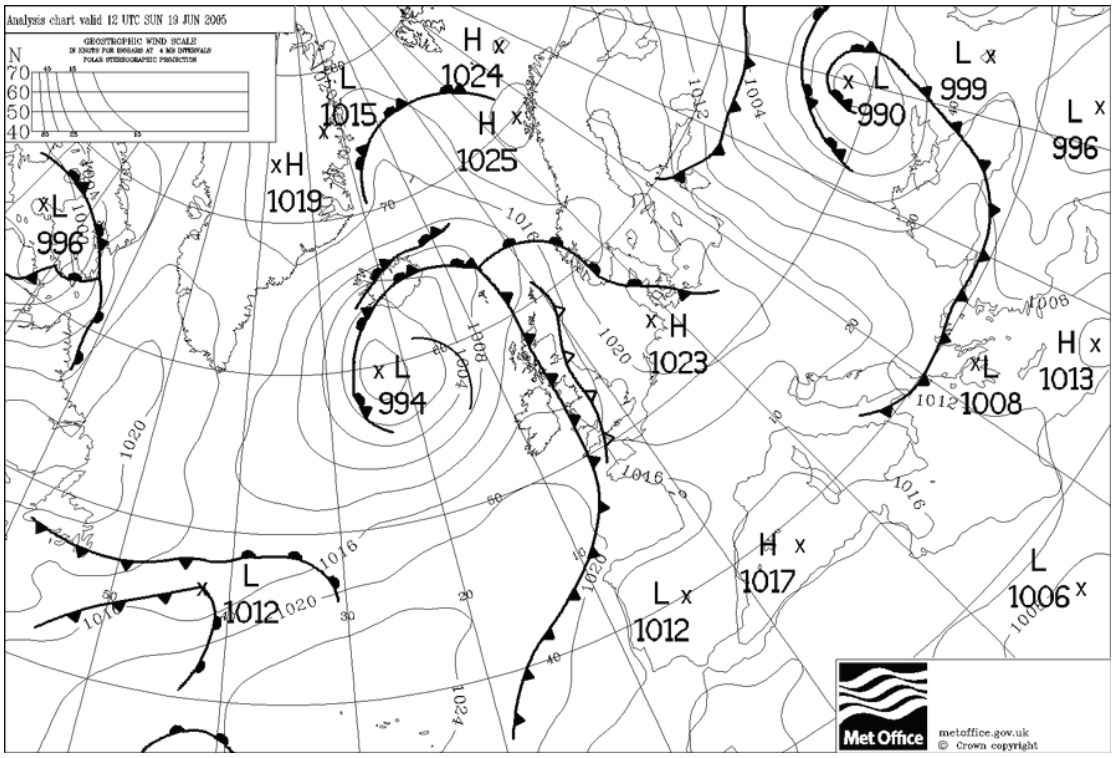

Figure 25. Synoptic chart for 12 UTC 19th June 2005.

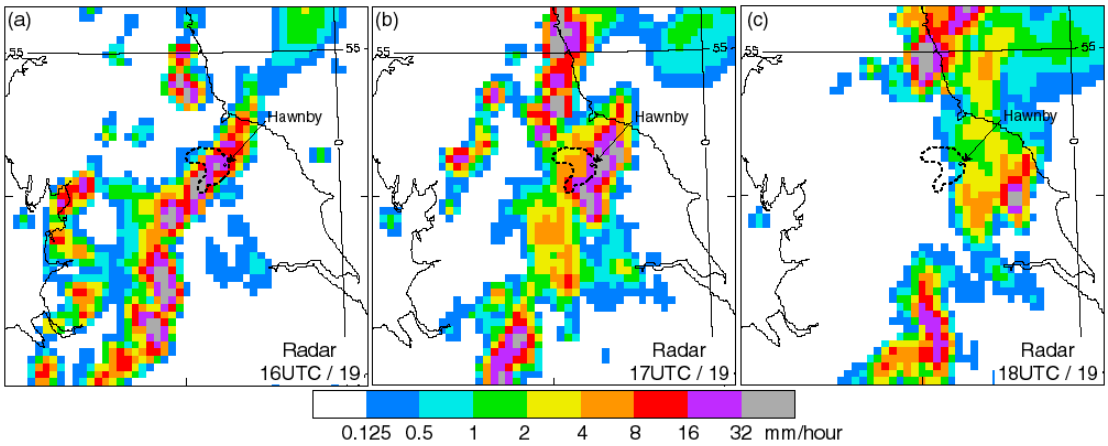

Figure 26. Rainfall rates from radar at 16, 17 and 18 UTC 19th June 2005. The dashed line encloses the area of highest accumulations (see later figures).

A cold front over western areas marked the boundary between the hot air and cooler air from the Atlantic. Thunderstorms developed during the afternoon along the line marked as an upper-cold-front (line with open triangles) on the synoptic chart. It is

debatable whether this really was an upper-cold front or just a means of depicting the line along which convection developed ahead of the cold front. The sequence of hourly snapshots of rainfall rates form radar in Figure 26 shows the line of convective rainfall transferring slowly south-eastwards. The most intense rainfall was concentrated in a small region to the northwest of the village of Hawnby. Figure 28(a) shows rainfall totals in excess of 90mm in the 6-hour period from 12 to 18 UTC; with most of that rain falling between 16 and 17 UTC. At Hawnby itself the radar measured around 70 mm. The rain gauge at Hawnby (information provided by the Environment Agency) recorded 69.4mm in three hours, of which 59.8mm fell in one hour and 50.6mm in just half an hour. All these totals have a return period of more than 200 years (using the Flood Estimation Handbook method). The result of all that rain was flash flooding at Hawnby, the larger village of Helmsley and the surrounding area. Several people were rescued by helicopter, houses were flooded and bridges destroyed.

3.2 The model and forecasts

The model domains used are shown in Figure 27.

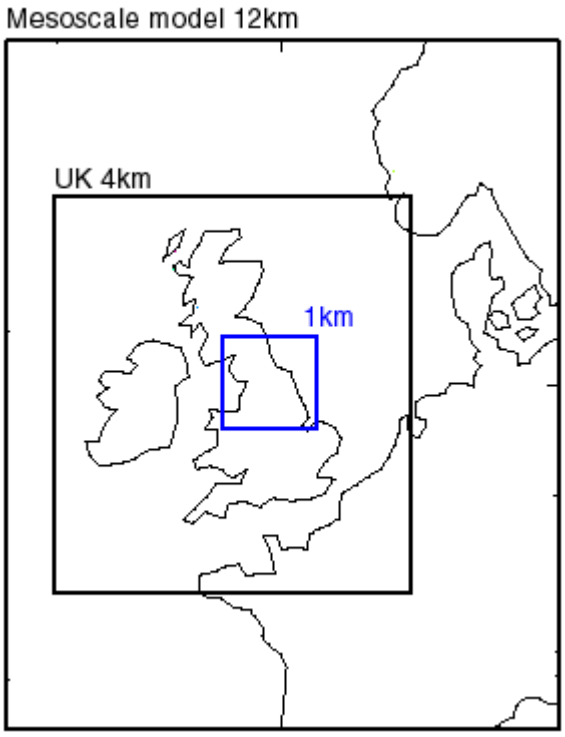

Figure 27. The domains used for case 2, 19th June 2005.

Forecasts have been run from start times of 18 and 21 UTC on the 18th and 00, 03, 06, 09 and 12 UTC on the 19th June 2005. Each of the forecasts has been given a label shown in Table 2.

Start time	12km	1km
12 UTC 18/06/05	Y12km12-18	Y1km12-18
18 UTC 18/06/05	Y12km18-18	Y1km18-18
21 UTC 18/06/05	Y12km21-18	Y1km18-18
00 UTC 19/06/05	Y12km00-19	Y1km18-19
03 UTC 19/06/05	Y12km03-19	Y1km18-19
06 UTC 19/06/05	Y12km06-19	Y1km18-19
09 UTC 19/06/05	Y12km09-19	Y1km18-19
12 UTC 19/06/05	Y12km12-19	Y1km18-19

Table 2. Note that the 1-km forecasts are labelled as starting at the same time as the 12-km forecasts. This is because they share the same analysis time even though the 1-km forecasts begin 1-hour later.

3.3 The accuracy of the forecasts

3.3.1 12-km forecasts

Rainfall accumulations over the period 14 to 17 UTC from the 12-km forecasts initiating at 18 UTC on the 18th and 00, 03, 06 and 09 UTC on the 19th are displayed in Figure 28. None of these forecasts were able to produce anything even vaguely resembling the high rainfall totals observed by radar. Two other forecasts (not shown) from 12 UTC on the 18th and 12 UTC on the 19th were no better. The 12-km model was unable to predict this storm.

3.3.2 1-km forecasts

Forecasts run at 1 km from the same times are shown in Figure 29. The 1-km model was also unable to predict the high rainfall totals. The two other forecasts (not shown) from 12 UTC on the 18th and 12 UTC on the 19th were equally poor. Note, though, that there were still differences between the 12 km and 1 km forecasts. Although the 1-km model was unable to produce the high accumulations seen in the radar, it did manage to produce some local accumulation of 30-40mm in some of the forecasts, which do not show up when averaging to a 5-km grid, and it also generated the rain in southwest-northeast oriented bands as were observed.

In summary, neither the 12-km or 1-km models were able to generate the storm of interest in any of the forecasts. The 1-km model did a better job at predicting the possibility of localised storms and the banded nature of the storm area, but it still failed to produce a forecast that resembled the actual event.

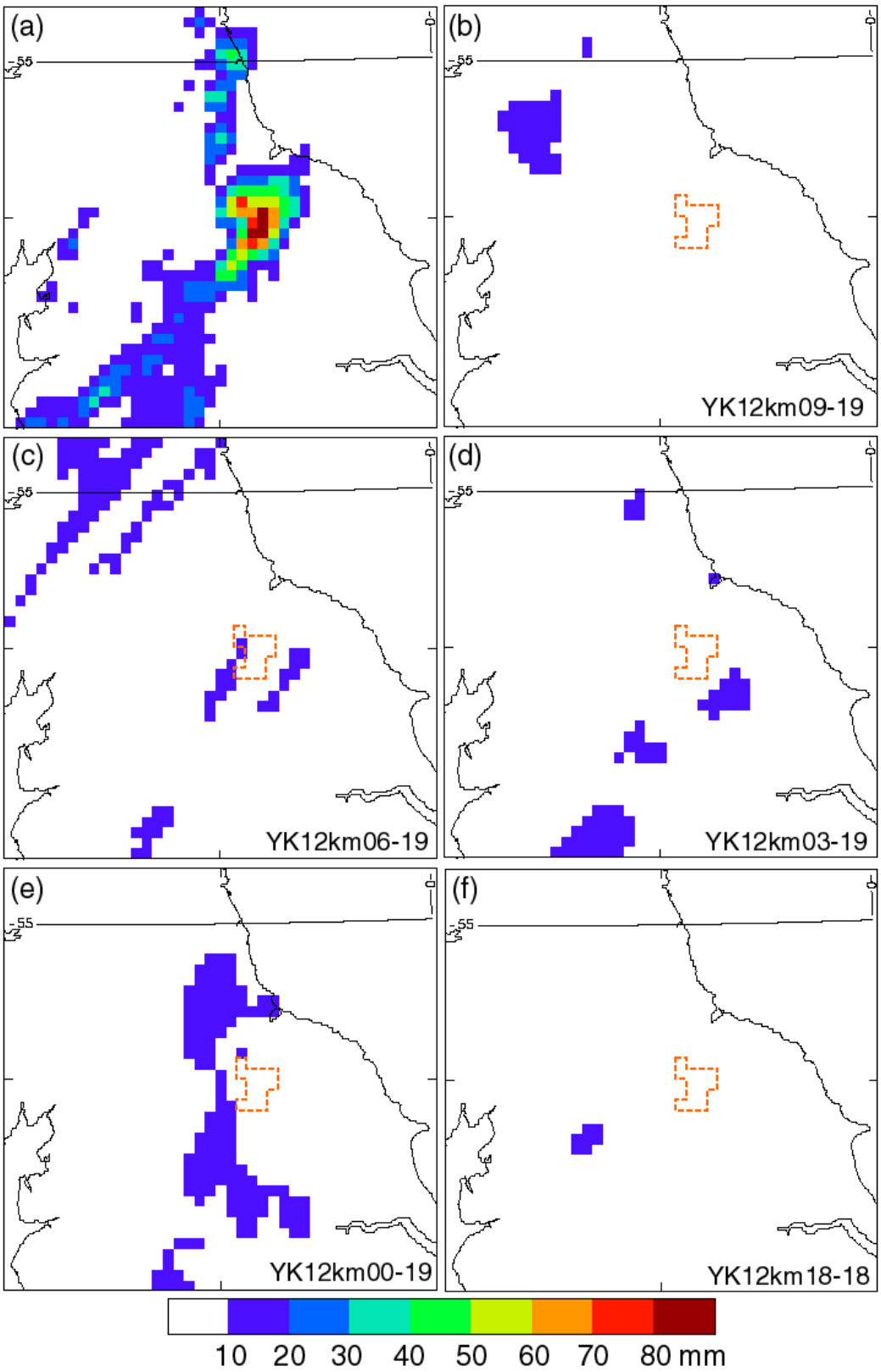

Figure 28. (a) Rainfall accumulations for the period 14 to 17 UTC 19th June 2005 projected on to a 5-km grid from (a) radar, (b-f) 12-km model forecasts starting from 09, 06, 03, 00 UTC, 19th and 18 UTC 18th.

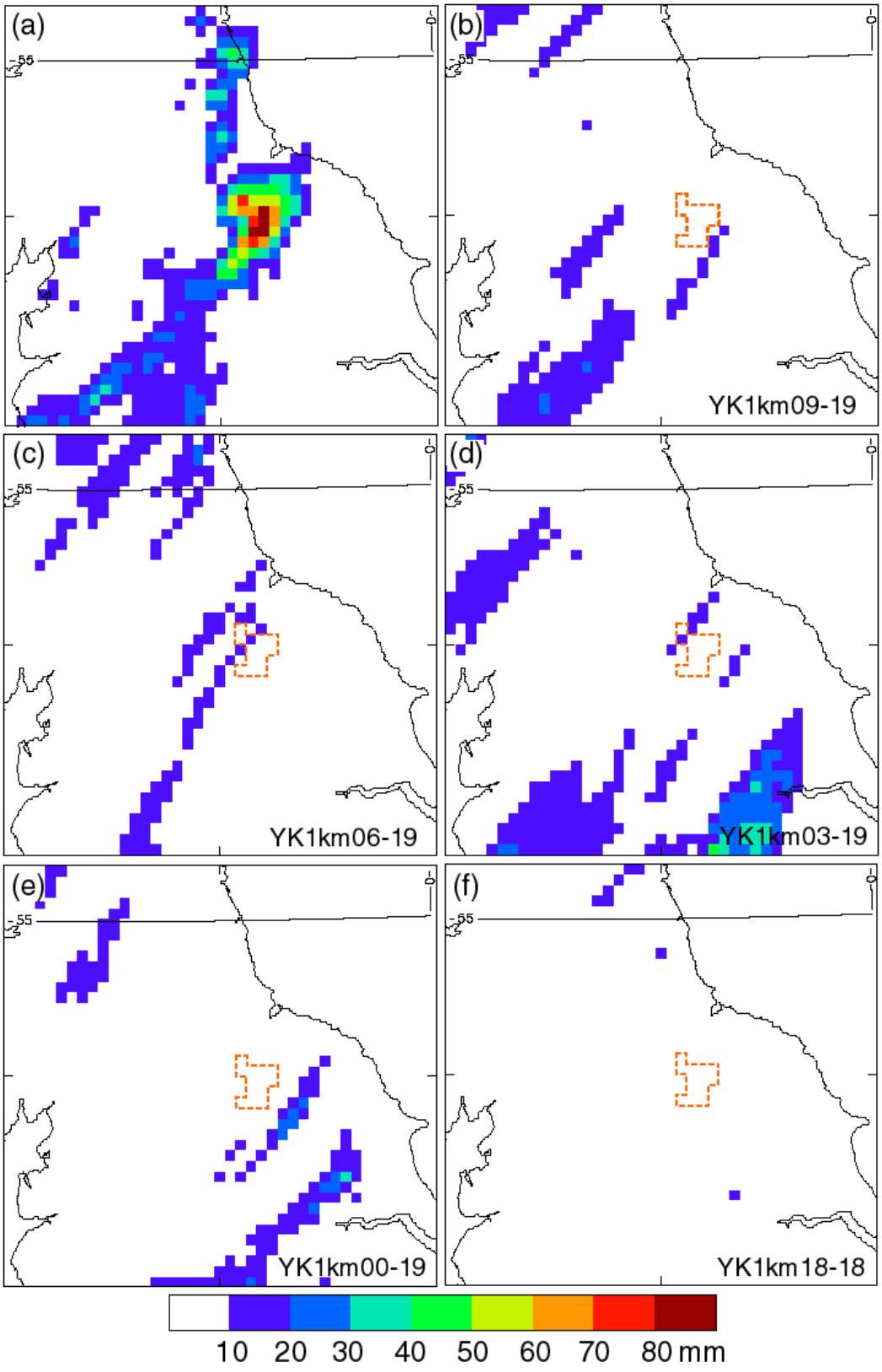

Figure 29. (a) Rainfall accumulations for the period 14 to 17 UTC 19th June 2005 projected on to a 5-km grid from (a) radar, (b-f) 1-km model forecasts starting from 09, 06, 03, 00 UTC, 19th and 18 UTC 18th.

3.3.3 Objective measure of spatial accuracy

The shortcomings of the 12-km and 1-km forecasts can also be seen in the Fractions Skill Score (FSS) verification measure. FSS curves for the 12-km and 1-km forecasts are displayed in Figure 30 (see previous case for more information about the FSS). We can see that none of the curves falls into the top-left (white) area of the graph. This shows that none of the forecasts reached a sufficient level of spatial accuracy (target skill) over scales small enough to be deemed useful (<50km). Indeed, many of the forecasts were poorer than a random forecast over small scales. Other thresholds give similar results.

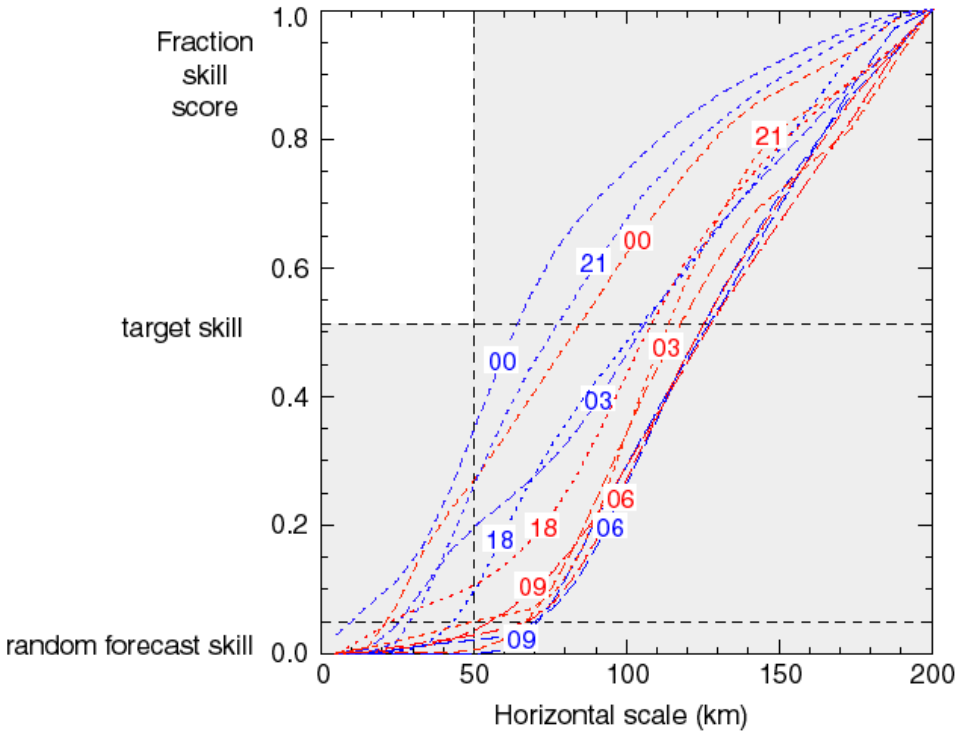

Figure 30. Graph of forecast skill against spatial scale using the Fractions skill Score (FSS) as a measure for comparing forecast rainfall accumulations against radar for rainfall pixels within the top 95th percentile of accumulations over 5 hours. See text for more details. 12-km forecasts are shown with blue lines, 1-km forecasts have red lines. The numbers refer to the start time of the forecast (e.g. 18 = 18 UTC). Forecasts deemed to have sufficient skill have curves that pass through the top-left white section.

3.4 Reasons for the poor forecasts

The immediate question that needs answering is why the forecasts were unable to reproduce the observed storm. The fact that improved resolution did not help much suggests that there may have been a problem with the larger-scale dynamics, so that is what we will examine first.

3.4.1 The importance of the larger-scale flow

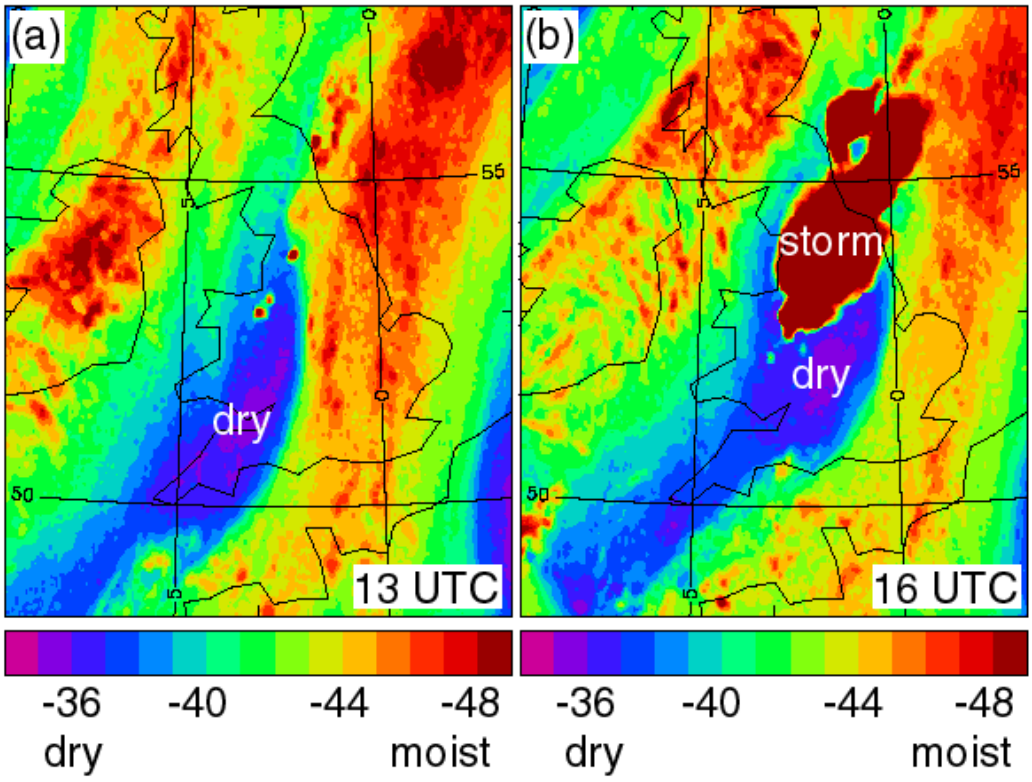

Figure 31. Meteosat Water vapour imagery for 13 and 16 UTC 19th June 2005. The blue colours show high brightness temperatures and therefore dry upper-tropospheric air. The /browns/reds/oranges show moist/cloudy upper-tropospheric air. The storm can be seen at 16 UTC growing from smaller clouds at 13 UTC.

Again, as for the previous case, we turn to Meteosat Water vapour (WV) imagery to give a view of what was happening with the larger-scale dynamics. We can see from Figure 31 that the storms developed within the northern part of a dry zone in the WV imagery and subsequently moved with that dry area. This is typically a favoured region for storm development (Roberts 2000). The WV imagery can tell us a great deal about what was happening with the upper-tropospheric flow at the time. The sharp boundary between the dry region (blue) and the moist/cloudy area (orange/red) marked by a white line in Figure 32(a) indicates the position of the jet stream; the strong core of winds at an altitude of around 10 km. This boundary (the white line) had twisted into an S shape because of rotation at upper-levels. Rotation at upper-levels is typically associated with an area of gradually ascending mid-tropospheric air (few cm/s) ahead of the rotation and descent behind. The effect of the ascent can be to de-stabilise the atmosphere enough to allow the triggering of convection. The other effect of the upper-level rotation is to allow potentially colder air (lower wet-bulb potential temperature) at mid-levels to over-run the hot humid air below and create an environment favourable (or more favourable) for strong convection. We don't know exactly how much ascent and over-running there was over the area of interest because there are no radiosonde observations available at the times of interest, but we do know that if the model was unable to reproduce this upper-level rotation it would have had key ingredients missing.

Figure 32 shows how well pseudo WV images for 12 UTC from the 12-km model compared with the actual imagery. Pseudo WV imagery from the 1-km model is not

shown because the domain is too small to cover the area required for a meaningful comparison. However, the 1-km forecast were very similar to their equivalent 12-km forecasts over the 1-km domain, so what we infer for the 12-km forecasts is also true for the 1-km forecasts. This is because the larger-scale dynamics in the 1-km model is largely a slave to the information passing through the domain boundaries from the 12-km model. It is in the smaller-scales (the representation of the storms) that the differences occur.

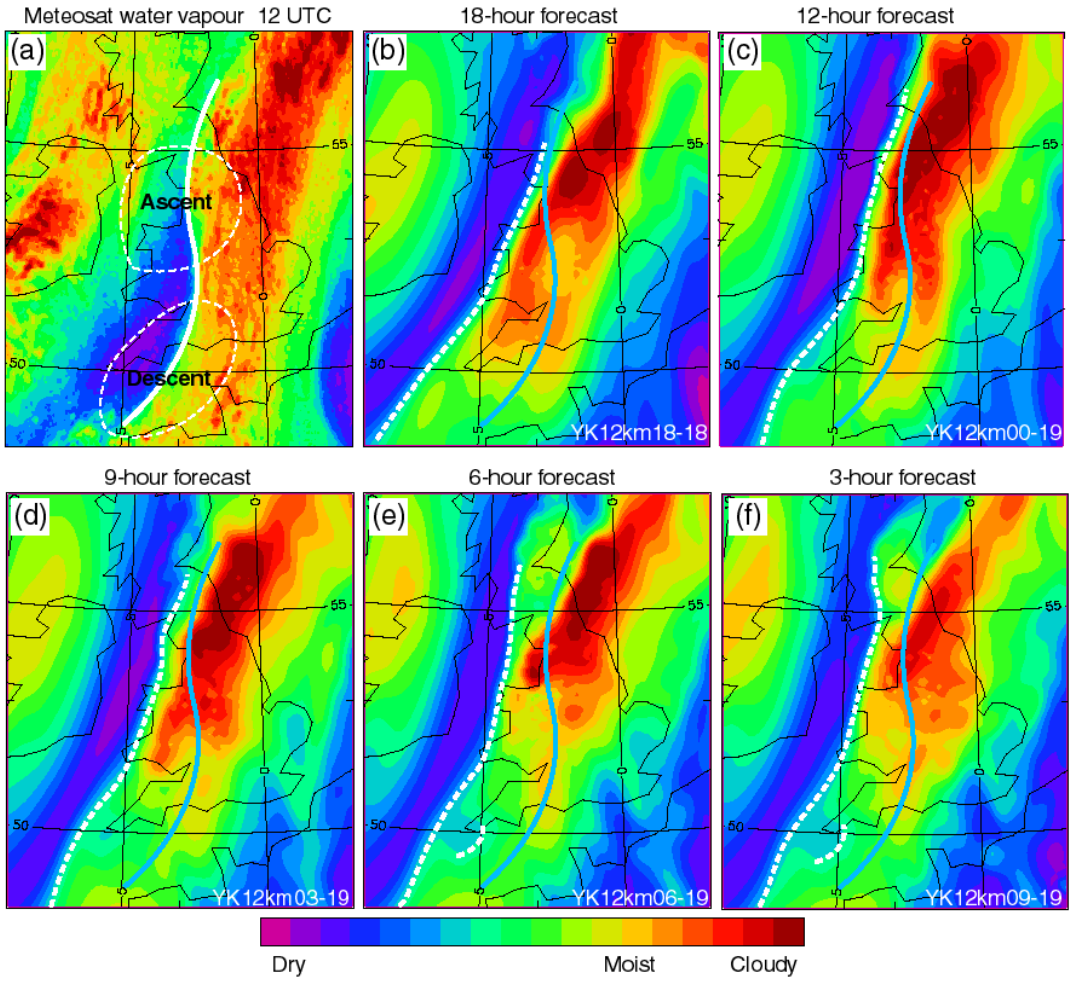

Figure 32. (a) Meteosat Water Vapour image for 12 UTC 19/06/05. Areas with warmer brightness temperatures (dry regions) are blue, areas with cooler brightness temperatures (moist/cloudy regions) are orange/red. The white line marks the sharp transition between the drier and moister air (typically the axis of the jet stream). General regions of possible ascent and descent within the troposphere are depicted by the dashed circles. (b-f) pseudo Water Vapour imagery extracted from relative humidity fields at 300 to 600 hPa from the 12-km forecasts starting at 18 UTC 18th, 00, 03, 06 and 09 UTC 19th June as comparison against the imagery. The blue line is the same as the white line in (a). The dashed white lines mark the moist/dry air transition in the pseudo imagery.

It is immediately evident that none of the 12-km forecasts had the degree of waving (rotation) that was observed. The boundary between the dry and moist air is much straighter. Clearly, the larger-scale upper-tropospheric dynamics was wrong in all the forecasts.

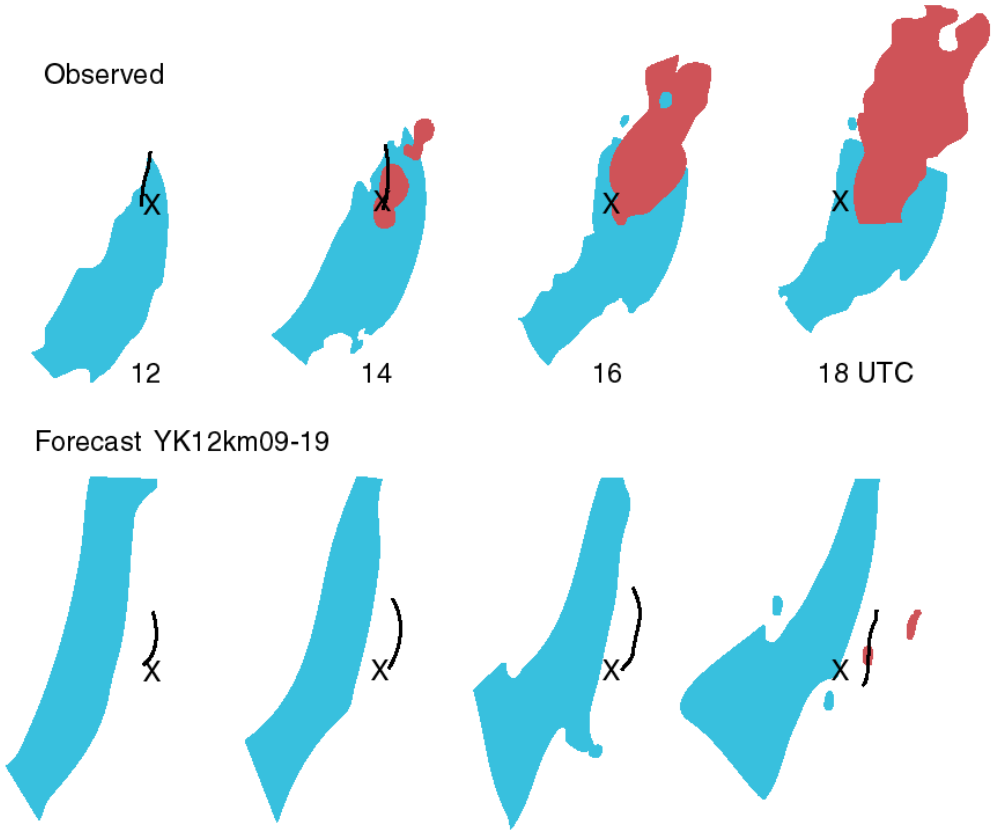

Figure 33. The observed (top) and model pseudo (bottom) water vapour imagery relative to a fixed point 'X' for 12, 14, 16 and 18 UTC. Blue areas are dry, brown areas are convective cloud. The black line in the observed/forecast surface cold front. The forecast is 12-km from 09 UTC 19th June 2005.

Another way to view the disparity between satellite imagery and model forecast is shown in Figure 33. The observed storms formed in the northern part of the dry region and remained there as the whole pattern transferred eastwards. The triggering appears to have occurred where the dry region over-ran the surface cold front (black line). In, contrast, the model dry region had a much straighter more elongated shape and did not over-run the surface cold front (which appears to have been positioned reasonably correctly) and deep convection does not develop.

Figure 33 has introduced a new ingredient into the mix - the surface cold front, which may have been much better predicted. Attention will now turn to the surface front and other aspect of the near-surface meteorology – winds, surface pressure, temperature and convergence.

3.4.2 Agreement with surface observations

Figure 34 shows the observed surface pressure at 14 UTC on the 19th and the surface pressure at the same time from the 1-km forecast starting from 07 UTC. Only one of the forecasts is shown, because there were only small differences in terms of surface parameters between all the 1-km forecasts at that time. The 1-km forecast produced essentially the correct pressure pattern with a trough of low pressure

separating the warm humid southerly gradient to the east from the cooler northerly gradient to the west.

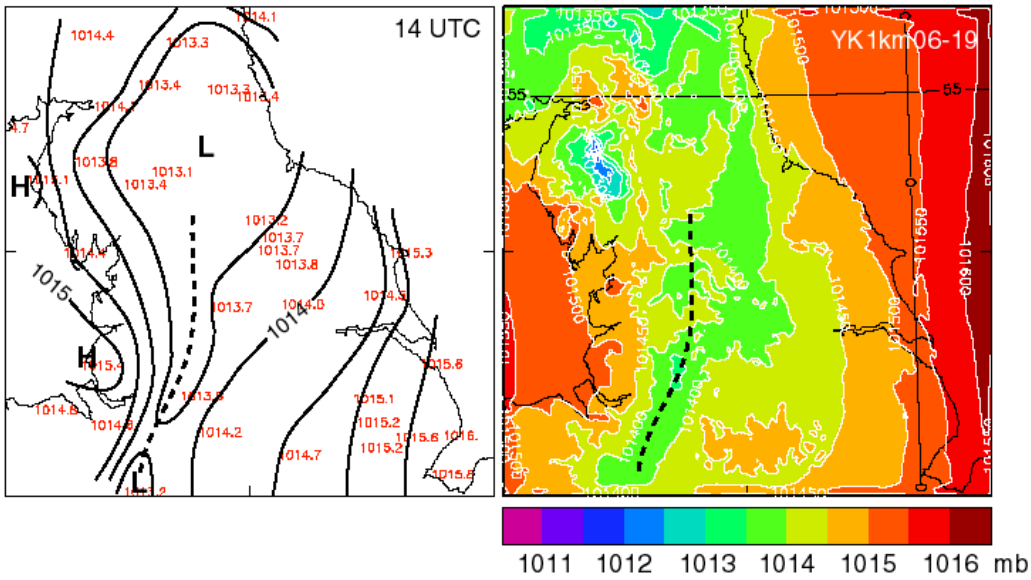

Figure 34. Mean sea level pressure at 14 UTC 19th June 2005 from surface observations (left) and the 1-km forecast from 06 UTC (right).

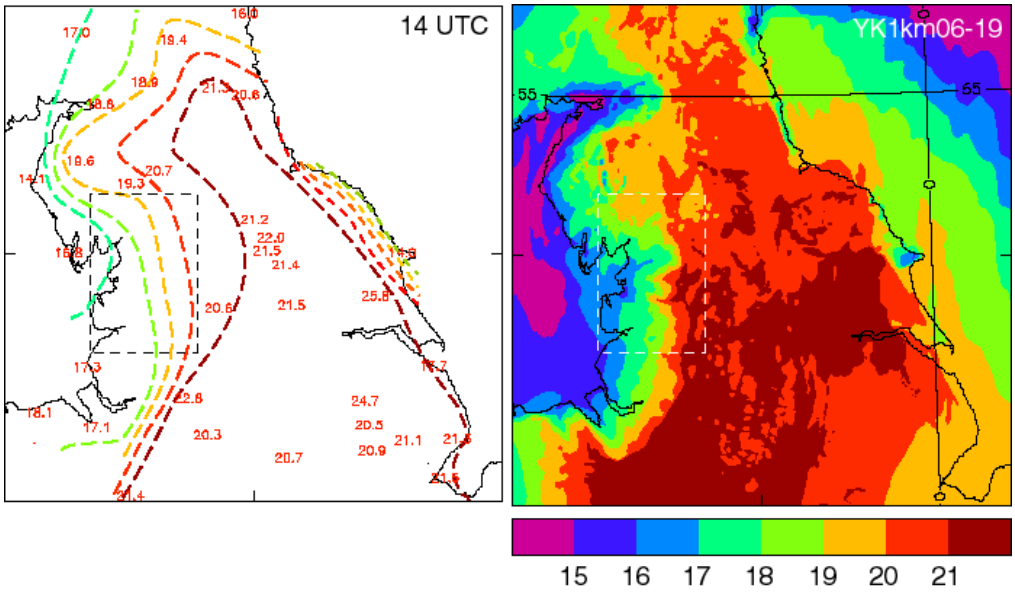

Figure 35. Wet-bulb potential temperature at 14 UTC 19th June 2005 from surface observations (left) and at 1000 hPa in the 1-km forecast from 06 UTC (right).

The 1-km forecast from 06 UTC also had a very good representation of the temperature/humidity pattern shown by the comparison of wet-bulb potential temperature in Figure 35. The strong gradient between the warm moist air (reds/browns) and the cooler air to the west (greens) was almost perfectly positioned (where there were sufficient observations to make that judgement). This is the surface cold front that was marked on Figure 33.

We also see that the near-surface winds were very similar to those observed (Figure 36). There was convergence of the low-level winds along the line of strongest temperature gradient (i.e. the cold front). The observed convergence line could not be pinpointed exactly because of insufficient observations, so it was only possible to mark a broader area within which the convergence line occurred. Nevertheless, the forecast convergence line was located right down the centre of the observed area of convergence.

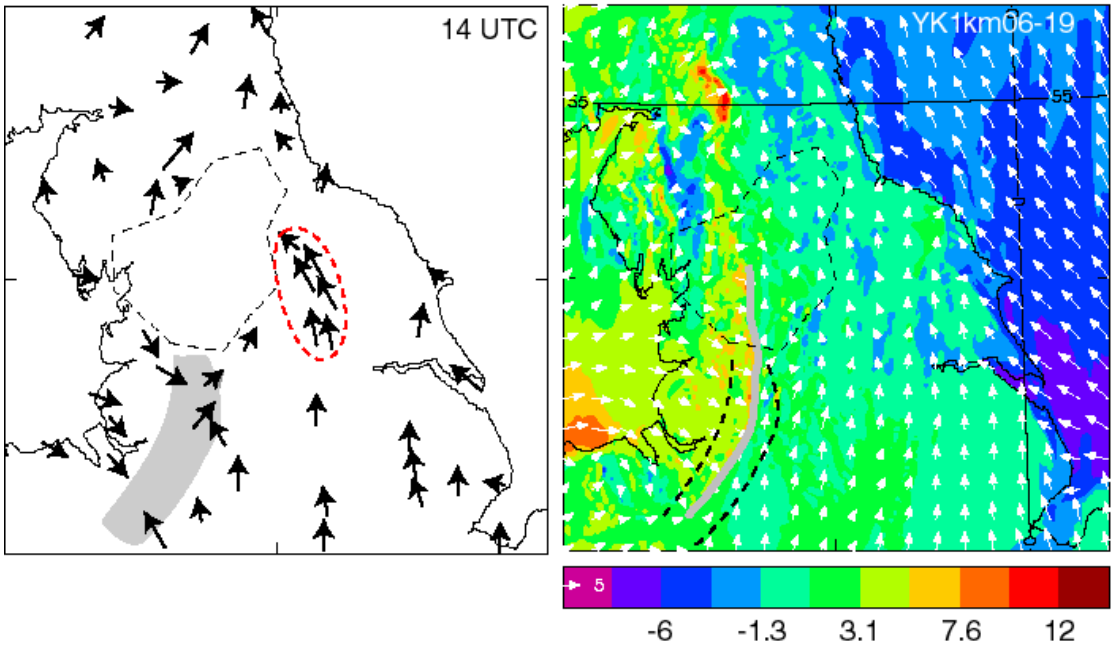

Figure 36. Wind vectors at 14 UTC 19th June 2005 from surface observations (left) and at 10 metres in the 1-km forecast from 06 UTC (right). The colours are the u-component of wind (east-west) in the forecast. The grey shading (left) and dashed lines (right) mark out the observed area of low-level convergence to within the tolerance of the observations. The grey line (right) is the forecast low-level convergence line. The relevance of the dashed red line in (a) is discussed in the text.

It appears then, that the 1-km forecasts were able to predict the surface winds and temperatures accurately (the convergence line in particular) in the run up to the storm development, but were unable to get the larger-scale upper-tropospheric dynamics correct.

The convergence line does seem to have played a key part in the development of storms in the model and probably also in reality, but that is less clear than in the previous case.

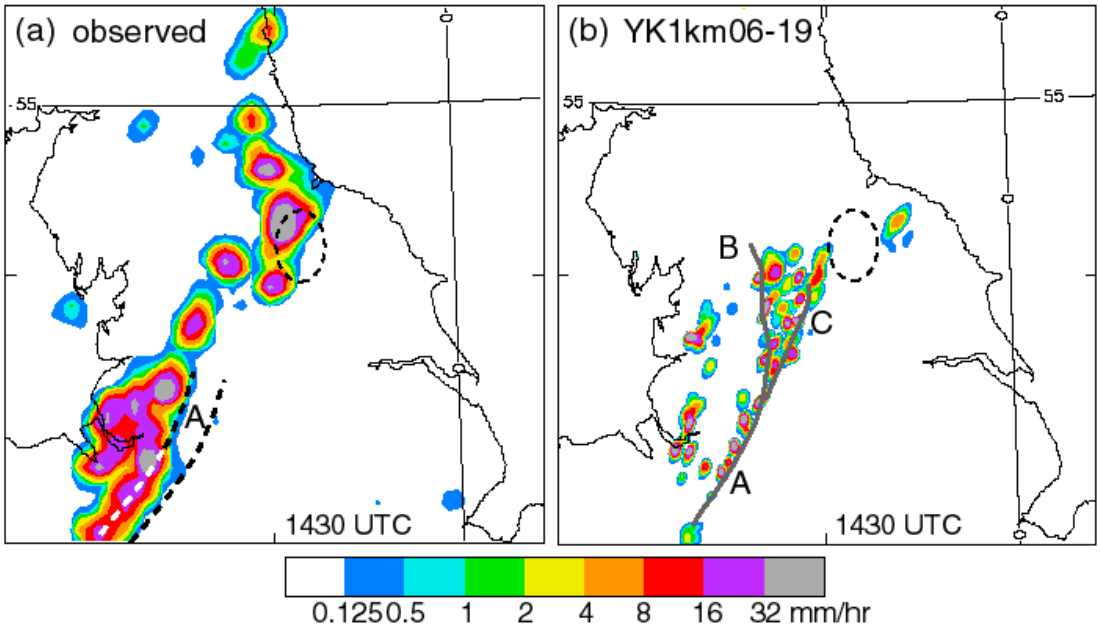

Figure 37. (a) Dashed line labelled 'A' shows the observed band of low-level convergence (from Figure 36) at 14 UTC 19th June 2005. Shading is the new rainfall 30 minutes later from radar. (b) Black lines are lines of low-level convergence at 14 UTC in the 1-km forecast from 06 UTC. Labels A, B and C are discussed in the text. Shading is the new rainfall in the same forecast 30 minutes later. The dashed oval marks roughly the area where the most rain fell over the period 14 to 17 UTC.

Figure 37(b) shows the locations of the convergence lines at 1430 UTC in the 07 UTC 1-km forecast and the model precipitation superimposed on that. The lines have been split into three parts, labelled A, B and C. Along line A the showers had only recently developed and appear to have developed along the convergence line where there was extra local ascent. Line B is a continuation of the cold front. Showers had already existed for around an hour. This line marked the back edge of the shower activity. Line C was not very distinct at earlier times. It was becoming more pronounced during the period 14 to 15 UTC. Showers developed initially along C at around 1330 UTC and then in the region between B and C after that. At 1330 there was little surface convergence along C suggesting that it was not surface convergence that lead to storm triggering. Once the storms developed, the convergence developed as a result of the outflow from the showers. Notice that the 1-km model did develop very heavy showers even though the 3-hour accumulations never became very large.

The observed convergence line, 'A' in Figure 37(a) at 1430 UTC is different from the line A in the model in that a band of storms had already developed and the line is more likely to be a response to those showers. We see some evidence for a convergence line pre-existing the storms over North Wales, but this is not clear because of a lack of observations over the Irish Sea.

By 1430 UTC there was already an active thunderstorm over the area of interest (dashed oval in Figure 37). Observations at 14 and 15 UTC do show local convergence in that area, but again this is more likely to have been as a result of the storm rather than a cause. However it is noticeable that at 1400 UTC (and earlier at 1300 UTC) there was also a deflection of the flow from southerly to south-easterly around the North York Moors (observations within red dashed line in Figure 36(a).

Interestingly, the 1-km model also had the same deflection of the winds and in the model this lead to an enhancement of the convergence line which will now be discussed.

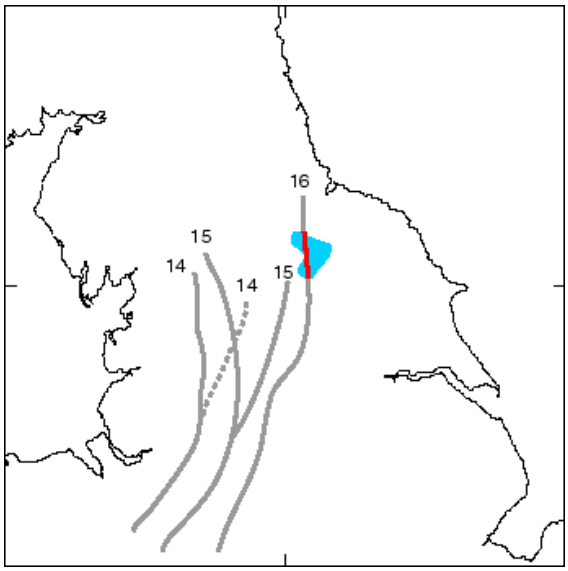

Figure 38. Lines show the hourly progression of the low-level convergence line(s) from14 to 16 UTC 19th June 2005 in the 1-km forecast from 06 UTC. The red section indicates where the convergence was particularly strong. The blue shading shows where the highest rainfall accumulations were observed by radar.

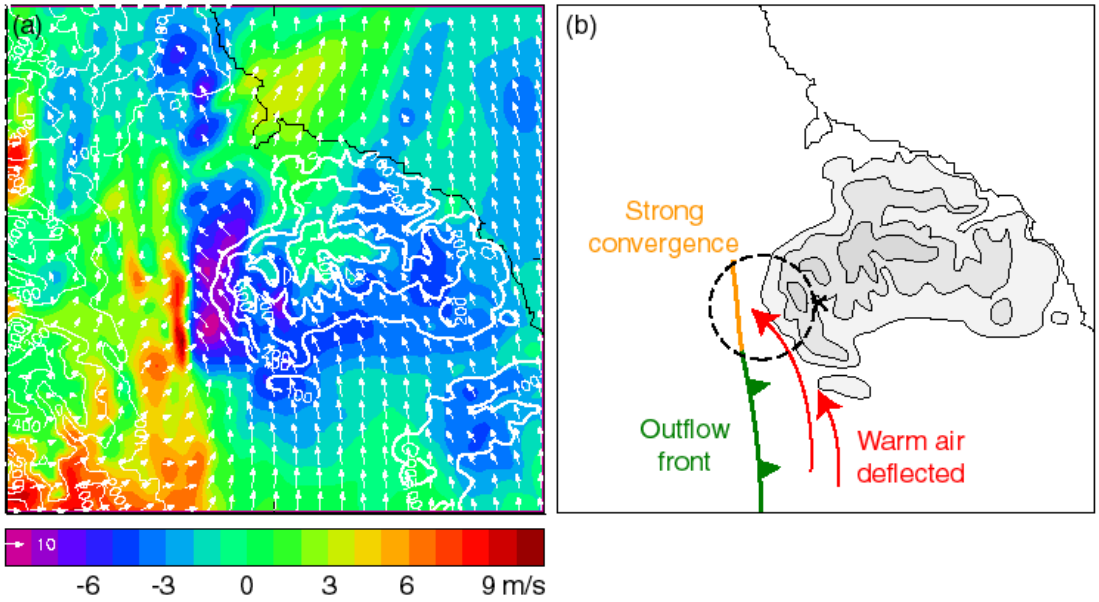

Figure 39. (a) The wind vectors at 975 hPa on top of the u-component of the wind at 975 hPa for 16 UTC 19th June 2005 from the 1-km forecast from 06 UTC. The white contours are height of the North Yorkshire Moors every 100 metres (in the model). (b) A schematic of the deflection of the flow around the North Yorkshire Moors (red arrows), the front marking the outflow from earlier convection (green line) and the line of strong convergence (orange). The dashed circle indicates the general area where the most rain fell (from radar). 'X' locates Hawnby.

The sequence of positions of the convergence lines in the 1-km forecast from 06 UTC are shown in Figure 38. There was a transition in the northern part of the line as the convergence along the old cold front part weakened (gone at 16 UTC) and the outflow line ahead became more pronounced (weak at 14 UTC). We also see that at 16 UTC the convergence line became much stronger (red line) in the region where the most rainfall was observed to have occurred (blue shading). Partly this enhanced convergence was a response to thunderstorm development, but it also appears that the storm developed in response to the convergence line and because of a local deflection of the flow around the North York Moors.

The deflection of the flow around the North York Moors and the interaction with the convergence line is shown in Figure 39. As the convergence line progressed eastwards it encountered the southerly flow ahead that had been turned more towards the southeast to the east of the high ground. This produced a stronger convergence line, which, as stated earlier, happens to be located in the area where the largest rainfall totals were measured. Storms were triggered there in the model, but they were smaller and more transitory than observed and rainfall accumulations were too small.

We can speculate that the peak in rainfall amounts observed in that area did have something to do with the flow around the North York Moors and the development of an enhanced localised convergence zone. A deflection of the flow is supported by some of the wind observations (Figure 36). If this was an important element for focusing activity in that area, the evidence suggests that the model was able to reproduce that mechanism. We can suppose that this event may have been predictable if the larger-scale dynamics were correct. The problem, of course is that the model failed to get the larger-scale flow correct and that made any skill in predicting the local effects redundant.

3.5 Summary

Both the 1-km or 12-km models failed to reproduce the location and quantity of rainfall that was observed. Some of the 1-km model forecasts did manage to generate heavy showers in the right general area, but the showers were too small and transitory to produce very high rainfall accumulations. As for the previous case, the accuracy of the forecasts depended on getting both the larger-scale and local-scale components of the flow correct.

3.5.1 Components

Component 1 – The larger-scale dynamics.

Water-vapour imagery revealed that the jet stream was twisted into an ‘S’ shape over a length of several hundred kilometres. This was indicative of an upper-level rotation that would have allowed ascending air in the mid-troposphere with low wet-bulb potential temperature to overrun the hot humid air near the surface and create an environment favourable for the development of severe convection. This larger-scale twisting of the jet stream did not occur in the forecasts and this meant that the

environment for convection was absent in the forecasts over the whole of northern England.

Component 2 – The local-scale dynamics.

The surface pressure, wet-bulb potential temperature and wind fields near the surface were all well represented by the 1-km forecasts. In addition the 1-km model was able to generate a line of enhanced convergence in the area where the high rainfall totals were observed because its ability to reproduce a deflection of the near-surface winds around the south-eastern flank of the North York Moors. Although speculative, it is thought likely that this local effect may have been very significant in the evolution of the observed storms.

The forecasts failed to reproduce the rainfall event because of a significant error in the larger-scale dynamics even though the representation of the surface fields was good. The forecasts were poor because one of the required components was wrong.

4 Conclusions

4.1 A tale of two storms

In stage 2 of the Modelling Extreme Rainfall Events project, the ability of a 1-km version of the Met office Unified Model to predict two separate cases of flood-producing convective storms has been investigated. These two cases were narrowed down from the original five cases documented in stage 1. They were:

Case 1: Flooding in London on 3rd July 2004.

Case 2: Flooding in Hawnby, North Yorkshire and surrounding places on the 19th June 2005.

4.1.1 Case 1, storm 1, 3rd August 2004

Both 1-km and 12-km forecasts with lead times of 18-24 hours were poor. They were unable to predict storms in the correct area, although they did predict convection in some places over southern England. All the later 1-km forecasts (lead times 3-15 hours) were much more accurate; they produced a northwest-southeast oriented band of high rainfall accumulations that was similar to the observed rainfall pattern. The reason for the improvement in the later forecasts was a better representation of the relevant larger-scale dynamics. In particular, the upper-tropospheric vortex associated with the broader area of convective development, more closely matched the positioning of the vortex inferred from Water Vapour imagery than was the case in the earlier forecasts.

Of the later more accurate 1-km forecasts, one stood out as being significantly better than the rest in the area of interest. This was the 1-km forecast from 09 UTC, which from both an objective measure and visual assessment was deemed to be a very good forecast. The reason it was so good was that, as well as capturing the larger-scale dynamics, it also had the convergence of the low-level winds that was responsible for the initial triggering of the storms of interest.

In summary: The 1-km model was able to produce a very good forecast of the storms on the 3rd August 2004 when both the larger-scale dynamical forcing and the local convergence line were in place.

4.1.2 Case 2, storm 2, 19th June 2005

None of the 1-km or 12-km forecast could reproduce the high rainfall totals or get the heaviest rain in the correct place. Even forecast just a few hours old were poor. Nevertheless, the 1-km forecasts, unlike the 12-km forecasts, did produce some very heavy showers and therefore gave more of an indication that heavy storms were possible. The problem was that the showers were too small and short-lived to give large rainfall amounts.

A comparison of surface observations with the 1-km forecasts showed that, despite the poor rainfall forecasts, the 1-km model did have a good representation of the surface temperature, humidity and wind fields. Most notably it had the position of the

surface cold front very well positioned, and it was also able to produce a zone of strong low-level convergence in precisely the region where the largest amounts of rain fell. The enhanced convergence in the 1-km model was due to a deflection of the southerly flow around the North Yorkshire Moors. Observations suggest that the same deflection of the wind occurred in reality. This may have been the reason why the actual storms were particularly heavy and persistent in that one area, but that is speculative.

The reason the rainfall forecasts were poor even though the surface conditions appear to have been well represented is because there was a significant error in the larger-scale flow in the mid and upper troposphere. Water Vapour imagery revealed that the jet stream above the frontal zone had started to twist because of imposed rotation. This would have made the atmosphere more unstable to deep convection as drier air overran the surface cold front. None of forecasts (even the shortest) had that larger-scale twisting, and therefore the environment in which the 1-km model was producing showers was very unlike the actual environment in which the storms occurred and so it is not surprising that the forecasts had deficiencies.

In summary: None of the 1-km or 12-km forecasts were able to produce the large rainfall totals because there was a major error in the mid and upper-tropospheric flow in every run. This was despite having the surface conditions well represented. The 1-km model was more realistic.

4.2 Components for a good forecast

We have seen that it is possible for a 1-km version of the Unified Model to accurately predict 'extreme' convective storms. There are no inherent reasons why an 'extreme' storm can not be predicted with an NWP model if it can simulate the physical processes and it is given the correct information. The accuracy of the forecasts depends on whether critical aspects of the meteorology have been represented correctly. These aspects can be thought of as the components necessary for a successful forecast. Each component operates on a particular scale and has most impact on forecast accuracy at that scale. Tables 3 and 4 summarise the important meteorological components involved in each of the cases and how they affect the performance of the forecasts.

Case 1: 3rd August 2004

~200-400 km	~10-30km	Performance
Upper-level vortex, poor		Poor forecast. Storms in wrong location. Too little rain.
Upper-level vortex, good	Low-level convergence line, poor	Generally good forecast. Useful skill at scales of ~20-35 km
Upper-level vortex, good	Low-level convergence line, good	Very good forecast. Useful skill at scales of ~10 km

Table 3

Case 2: 19th June 2005

~200-500 km	~20-100km	Performance
Rotation at upper-levels, poor	Surface cold front. Low-level convergence, good.	Poor forecast. Storms in wrong location. Too little rain.
Rotation at upper-levels, poor	Convergence due to hills, good.	Poor forecast. Storms in wrong location. Too little rain.

Table 4

The findings in tables 3 and 4 have been generalised in table 5. A good forecast depends firstly on getting the larger-scale flow correct. If that is wrong the forecast will be poor even if the local dynamics is well represented. Even in situation where local effects appear to be of primary importance, it is the larger-scale flow that sets up the correct environment in which the local dynamics operates. If the larger-scale flow is well represented and the local dynamics is also well captured then the forecast can be accurate even at local scales. If the larger-scale flow is well represented, but the local dynamics is not then the accuracy of the forecast will depend on the relative importance of those components (and the degree of accuracy that is required). Morcrette et al 2007 describe another example of the way in which a number of different components combined for the initiation of an isolated convective storm during the observational phase of the Convective Storms Initiation Project (CSIP) (Browning et al 2006).

~ 100-500 km	~ < 100 km	Performance
Larger-scales poor	Local scales poor	Forecast poor
Larger-scales poor	Local scales good	Forecast poor
Larger-scales good	Local scales poor	Forecast may be good at larger scales
Larger-scales good	Local scales good	Forecast good at all scales

Table 5. The meaning of larger and local scales in the context of forecasting convection using this model setup is described in appendix 1.

All this may seem intuitively obvious, but it is important to emphasise because it contradicts a view often held about the nature of convection over the UK. It is typical for scientists and forecasters to focus on just the local effects as if that is all we need to get right and assume that there is no variability on larger scales or that it is automatically correct or of little consequence. That is not an appropriate view to take.

4.3 Implications for research and development

4.3.1 Coarse resolution data assimilation

The first thing is to reiterate that when trying to predict severe convective storms with a 'storm-resolving' model (grid-length ~1 km), the larger scale dynamics must be sufficiently correct or the forecast is very likely to be poor. It follows that it is just as important to improve data assimilation systems in the coarser resolution model that provides the boundary information, as it always was before storm-resolving models were developed. This is particularly true for the UK where dynamical features on scales of ~100 to 500 km can spin up in data sparse areas over the Atlantic. It may be less of an issue for nowcasting systems predicting up to 6 hours ahead, but even then it will matter.

4.3.2 Ensembles

The first case showed that as the forecasts got closer and closer to the event, the upper-level vortex became more accurately represented and this had a big impact on forecast skill. We were, in a sense, looking at an ensemble of forecasts in which the larger-scale dynamics played a prominent role. This demonstrates the value an ensemble forecast system might have. It may well have been the case that with an ensemble some of the forecasts with longer lead times would have had the correct upper-level vortex, or even some of the forecasts in the second case might have had the correct upper-level flow. Also, confidence in the 1-km solution was increased because of the run-to-run consistency. In the same way consistency amongst members of an ensemble can raise confidence in the solution (provided the ensemble has a history of representing uncertainty well). The Met Office now has an ensemble forecast system (MOGREPS) with 24 members and a grid spacing of 24 km running for 54 hours (Bowler et al 2007). The problem for predicting extreme convective events is that a grid spacing of 24 km is not sufficient to represent the storms themselves or local dynamical precursors, which is precisely why a ~1 km model has been developed.

A useful way forward would now be to combine the two and nest the ~1 km model inside ensemble members. At present, computer resources are insufficient to run a large ensemble of high resolution forecasts (even in research mode), so there will need to be a methodology for selecting appropriate members from the ensemble. In terms of predicting the possibility of intense convection this will have to involve an automatic way of differentiating coarser-resolution forecasts that are more likely to produce strong convection from those that are not and just running from a select few of the members. This needs to be an area of research. It may be reasonably easy in situations when convection is strongly forced from the larger-scales, but in other situations when variations in the larger-scale environment are important but the storms themselves are triggered and maintained by local effects (e.g. the Boscastle storm, Golding 2005) it may not be trivial. The aim would be to have a system that could provide plausible alternative scenarios for convective storms and use a 'storm-resolving' model to represent the convection under those different scenarios. A further area of research would then be needed to examine appropriate ways of presenting risk and uncertainty from such a system.

4.3.3 High resolution data assimilation

In both cases we saw the importance of low-level convergence lines for the initiation of convection. If those lines had not been present or were wrongly located, the initiation of convection would have been wrong. Since these features can only be properly represented in the ~1-km model (not at 12 km) there is a need for a data assimilation system that can adjust the model dynamics on those scales. It will require the introduction of advanced data assimilation methods using high-spatial-density observations (e.g. Doppler winds from radar).

Not only is it important to capture the local winds and any convergence lines at the start of a forecast, it is also important to represent the causes if the correct behaviour of the flow is to be maintained into the forecast. When the cause is to do with orography or coastlines (example in case 2) it is not so problematic because they don't move and are represented in sufficient detail in a ~1-km model (provided things like sea-surface temperatures, surface roughness and soil moisture are accurate enough). However, in both of the cases, there were convergence lines that existed because of the outflow from earlier showers. This is more problematic because the correct thermal, moisture, cloud and rainfall structures are also needed. In addition, the earlier showers were themselves dependent on having the correct 3-D moisture and temperature patterns (e.g. mid-level relative humidity in case 1). Other factors like cloud shadowing require an accurate cloud distribution. Frontal convergence requires an accurate representation of the frontal structure.

New developments in high-resolution data assimilation could be instrumental in improving short-range forecasts of severe storms in high-resolution models at local scales, but it will be challenging.

4.3.4 Turbulence and cloud microphysics

So far we have discussed the importance of getting the environment correct before the initiation of the storms. This has neglected the representation of the showers themselves. If the rainfall predictions are to be accurate, then the development of the showers from non-precipitating clouds through to full-blown thunderstorms needs to be modelled correctly. The timing of initiation, growth of the showers, intensity of the showers, interactions of the showers and decay of the showers are all important. For example, a forecast of storms that are in the correct location but the wrong size and intensity will not be as useful as it should be.

One of the problems we have seen in the two cases (and others from stage 1, Roberts 2006 and from the Storm Scale Modelling Project, Roberts 2005) is the 'measles' effect. Showers appear to initiate too soon and appear as a rash of small intense cells that are too extensive and do not readily upscale into larger storms. It is thought that this behaviour is partly due to insufficient sub-grid-scale mixing within clouds because the only mixing above the boundary layer comes from numerical diffusion (the convection parametrization scheme provides additional mixing in the 12-km model but is not used at ~1km). This is about to be addressed by the introduction of a turbulence parametrization that essentially mixes the air between grid points where the vertical velocity is high – i.e within the clouds. The intention is to examine this in the final stage of this project using case 1 (3rd August 2004).

The other factor to consider is the way cloud and precipitation is represented in the model. This too will have a bearing on the structure of the showers and the amount of rain produced and will clearly interact with the new turbulence parametrization. There are unknowns and limitations in the way the cloud microphysics is modelled and realistic sensitivity experiments to examine the impact on forecasts of extreme convective rainfall events, starting with case 1, would be useful.

Furthermore, the representation of the cloud and precipitation will play an important role in the generation of cold-pool outflows because the downdraughts are driven by cooling from evaporating rain. In case 1, this was crucial for the formation of the convergence line that lead to the initial triggering of the deep convection and it seems that an error in the location of the earlier showers was offset by an over-prediction of their intensity.

4.3.5 Verification and post-processing

The scale-selective verification approach used to obtain an objective measure of forecast skill (the Fractions skill Score) has proved to be useful for determining the scales over which a forecast is deemed to have sufficient accuracy. It is now being adopted for evaluating the UK 4-km model and should also be used for higher-resolution models both in a research context and then operationally.

The method can also be used to investigate the scales at which changes to model parameters have an impact and will be used for that purpose in the next stage of this project.

The results from this project and from its predecessor the Storm Scale modelling project have shown that although a ~1-km model has the capability of producing very good forecasts it will not always do so and there is inherently less predictability at smaller scales. There is a need for a presentation of the output in terms of probabilities that reflect the expected scale-dependence of the forecast accuracy. The variation of forecast skill with spatial scale can be determined from the scale-selective verification method.

4.4 Final Comments

The Met Office Unified model has been run with a grid spacing of 1 km with varying forecast lengths for two flood-producing convective rainfall events.

One of the storms could be predicted very accurately (3rd August 2004), the other could not (19th June 2005).

The ability of the model to predict the storms depended on having particular meteorological components in place. The accuracy we require for flood warning requires the smaller-scale components (e.g. low-level convergence lines) to be correctly represented, but this depends on the larger-scale dynamics being sufficiently correct first. The failure of the forecasts of the 19th June 2005 storms was due to errors in the larger scales dynamics, the successful forecasts of the 3rd August 2004 storms was due to a good representation of all scales down to the local convergence lines.

The challenge for high-resolution Numerical weather Prediction models is to represent the key meteorological components at all relevant scales. This means a multi-faceted approach to research and development which focuses on the problem of larger-scale initial errors and the possible application of ensembles then down to data assimilation at fine scales, the representation of turbulence, cloud and precipitation process that impact all scales and the local representation of the land/sea surface. Forecast evaluation and presentation are also key areas that require further work.

Provided that the larger-scale meteorological components are in place, a ~1-km model has the capability to produce substantially better forecasts of 'extreme' convective rainfall events than a 12-km model. However, since there will always be some uncertainty in the larger-scale dynamics and a random nature to the smallest scales, care is needed to make sure this uncertainty is somehow included in the final product.

5 Appendix 1

Larger scales:
The flow that can be accurately represented by a 12-km model. This means only dynamical features that are significantly larger than the convective storms and not dependent on convective storms. Scales are ~ 60 to 500 km.
e.g. Cyclones, fronts (but not rainbands or convection within fronts), upper-level vorticity features, jet streaks,

Local scales:

(i) Non-convective: The flow that can not be represented by a 12-km model but can by a 1-km model, and is not associated with convective storms. Scales are ~ 5 to 60 km. e.g. low-level convergence lines, cloud heads, mesoscale low pressure areas, gravity waves, frontal sub-structure

(ii) Convective: The flow associated with convective storms that can not be represented by a 12-km model but can by a 1-km model. Scales are ~ 5 to 60 km. e.g. convective outflows, storm-scale circulations, cloud shadowing, daughter cells, convectively generated gravity waves

Intermediate scales (including convection):
The flow associated with an area of convection that is large enough to be represented by a 12-km model but is dependent on convection that can not be represented by the 12-km model. Scales are ~ 60 to 200 km.
e.g. circulations associated with storm clusters, larger-scale cloud shadowing

6 References

Browning et al. (2006). A summary of the Convective Storm Initiation Project Intensive Observation Periods. Met Office NWP technical report 474.

Browning, K.A. and N.M. Roberts (1994) Use of satellite imagery to diagnose events leading to frontal thunderstorms. Part I of a case study. *Meteorol.Appl*. 1, 303-310.

Browning, K.A., N.M. Roberts and C.S.Sim, (1996) A mesoscale vortex diagnosed from combined satellite and model data. Meteorol.Appl., 3, 1-4

Bowler, N. E., A. Arribas, K. R. Mylne and K. B. Robertson, (2007): The MOGREPS short-range ensemble prediction system Part I: system description, Met Office NWP Technical Report No 497

Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, S. E. John and T. P. Legg (2007): The MOGREPS short-range ensemble prediction system Part II: Performance and Verification, Met Office NWP Technical Report No 498

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q. J. R. Meteorol. Soc., 131, 1759–1782.

Golding, B. (1998). Nimrod: a system for generating automated very short range forecasts. Meteorol. Appl., 5, 1–16.

Golding, B. W (editor)., 2005: Boscastle and North Cornwall Post Flood Event Study - Meteorological Analysis of the Conditions Leading to Flooding on 16 August 2004. Met Office *NWP Technical Report.,* 459

Hand, W. H., and Fox, N. I. and Collier, C. G. (2004), A study of twentieth-centuary extreme rainfall events in the United Kingdom with implications for forecasting. Meteor. Apps., 11, 15-31.

Hoskins, B. J., McIntyre, M. E., and Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. Q. J. R. Meteorol. Soc., 111, 877–946.

Morcrette, C., H. W. Lean, K. A. Browning, J. Nicol, N. M. Roberts, P. A. Clark, A. Russell, A. Blyth. (2007). Combination of mesoscale and synoptic mechanisms for triggering an isolated thunderstorm: observational case
study of CSIP IOP 1. Accepted Q. J. R. Meteorol. Soc

Lean H. W. S.P. Ballard, P.A. Clark M.A.G. Dixon, Zhihong Li and N. M. Roberts (2005). The Summer 2004 Reruns with the High Resolution Trial Model. Met Office NWP technical report 466

Roberts, N.M., (1995) Association of sferics with a dry intrusion in Meteosat imagery. *Meteorol.Appl*. 2, 109-111.

Roberts, N. M. (2000). A guide to aspects of water vapour imagery interpretation: the significance of dry regions. Met Office NWP technical report 294

Roberts, N. M. (2000). The relationship between water vapour imagery and thunderstorms. Met Office NWP technical report 300

Roberts, N.M., (2004). Measuring the fit of rainfall analyses and forecasts to radar: Strategy for stage 5 evaluation. Stage 4 report from the Storm-Scale numerical. Met Office NWP technical report 432.

Roberts, N.M., (2005) An Investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood producing rainfall. Final scientific report from the storm scale modelling project. Met Office NWP technical report 455.

Roberts, N. M. (2006). Simulations of extreme rainfall events using the Unified Model with a grid spacing of 12, 4 and 1 km. Met Office NWP technical report 486.

Roberts, N. M. and H. W. Lean (2007). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Submitted to Monthly Weather Review. Accepted following minor revisions.

Sleigh, M. W. and Collier, C. G. (2004). An investigation of the tipping term as a prognostic tool in short-range thunderstorm forecasting. Atmos. Res., 69, 185-198

Weldon, R.B. and S.J. Holmes, Water vapour imagery: interpretation and Application to weather analysis and forecasting.
NOAA Technical Report. NESDIS 57, National Oceanic and Atmospheric Administration, US Dept of Commerce, Washington DC, 213pp.

Met Office technical reports are available from the Met Office web site.