Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model

Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model

Document History

Date	Version	Action/Comments	Approval
21/09/07	1.0	First draft by Nigel Roberts	
24/09/07	2.0	Second draft following comments by Brian Golding and Mark Dixon	Brian Golding, Met R&D Programme Manager

Author:
Nigel Roberts,[1*]
[1] Met Office, Joint Centre for Mesoscale Meteorology, Meteorology Building, University of Reading, PO Box 243, Earley Gate, Reading, Berkshire, RG6 6BB, UK

ABSTRACT:

It is becoming increasingly important to be able to verify the spatial accuracy of precipitation forecasts, especially with the advent of high-resolution Numerical Weather Prediction (NWP) models. In this paper, the Fractions Skill Score (FSS) approach has been used to perform a scale-selective evaluation of precipitation forecasts during 2003 from the Met Office mesoscale model (12 km grid length). The investigation shows how skill varies with spatial scale, the scales over which the data assimilation (DA) adds most skill, and how the loss of that skill is dependent on both the spatial scale and the rainfall coverage being examined. Although these results come from a specific model, they demonstrate how this verification approach can provide a quantitative assessment of the spatial behaviour of new finer-resolution models and data assimilation techniques.

1. Introduction

The perceived accuracy of precipitation forecasts is very dependent on the scales over which they are being assessed. It is easier to predict whether rain will fall somewhere within a large area than a small one. For example, a forecast of showers in the vicinity of a sporting event may have been correct, even if the forecast of rain at the event itself was wrong. Until recently, most verification of rainfall forecasts has been concerned with assessing the performance at point locations (i.e. where there are rain gauges). The problem with this is that in many situations, as with the sporting event example, such a verification approach will fail to recognise when a forecast contains useful information unless it happens to be correct at particular locations. It will also be unable to discriminate between a forecast in which the rain is nearly in the correct place and a forecast that is wrong by a bigger margin, i.e. it knows nothing about the scale of the error.

This deficiency in the point-based verification approach has become much more apparent now that high resolution Numerical Weather Prediction (NWP) models (grid spacing < 5km) are being widely developed. These models are expected to produce more accurate precipitation forecasts, yet it is possible that, even if more realistic, the forecasts may not be more accurate when verified at specific locations. The problem was anticipated by Lorenz (1969) who argued that the ability to resolve smaller scales would result in forecast errors growing more rapidly. So the benefit of finer resolution may be outweighed by faster error growth (especially in convective situations). Put another way, the models will be better equipped to predict the development of localised downpours, but not their exact locations. Evidence for this has already been demonstrated; Mass et al (2002), Zhang et al (2003) and Done et al (2004) have already shown that, although more realistic, higher resolution models did not give better point-specific verification scores. The real benefit of higher resolution should be seen in an improvement in area-averaged rainfall forecasts, which would mean, for example, that useful forecasts could be provided for smaller river catchments than was previously possible. That is what needs to be evaluated.

In response to the problem outlined above, and the need to obtain a more complete assessment of the quality of quantitative precipitation forecasts (QPF), several new verification methods have been developed in recent years. Ebert and McBride (2000), Done et al (2004) and Davis et al (2006) have developed object-based methodologies which assess the positioning, intensity and structure of precipitation forecasts. Briggs and Levine (1997), Zepeda-Arce et al (2000), Casati et al (2004) and Bousquet et al (2006) have used spatial decomposition methods to investigate how forecast error varies with scale. Marzban and Sangathe (2006) have used a cluster analysis method that is both object-based and spatial.

In this paper a spatial verification measure called the Fractions Skill Score (FSS) (Roberts & Lean) is used to compare a year-long set of operational forecasts from the old mesoscale version of the Met Office Unified Model (UM) (12-km grid spacing) with radar. The purpose is to demonstrate how this particular verification approach can provide information about predictability over different spatial scales and how it could be a valuable tool for assessing the behaviour of new high-resolution models and data assimilation (DA) methods.

In section 2 the verification method will be briefly described, in section 3 the model data used will be described, in section 4 the results will be presented and in section 5 conclusions drawn.

2. The verification method

The purpose of this verification method is to obtain a measure of how forecast skill varies with spatial scale. To do this, the fraction of occurrences of specified rainfall accumulations within different sized sampling areas (neighbourhoods) are computed (similar to Theis et al 2005) for both the forecast and radar data, then the forecast and radar fractions are compared. The radar data has been processed through the Nimrod system (Golding 1998), which includes calibration against rain gauges and removes, as much as possible, any spurious artefacts (Harrison et al 2000). Radar data are used, rather than point measurements from gauges, because of their spatial coverage.

The verification is done in two stages. Firstly, the forecast and radar fractions are generated and then those fractions are compared using a measure called the Fractions Skill Score (FSS)

2.1 Stage 1, generating the fractions

Hourly accumulations from radar are re-projected on to the same grid as the model so that they can be directly compared with hourly accumulations from the model forecasts. For every forecast pixel, the fraction of surrounding pixels within a given square area (neighbourhood) that exceed a specified accumulation threshold (e.g. >1mm) is computed. Thus a fraction is assigned to every pixel. Exactly the same is done for the radar data using the same sized neighbourhood and the same accumulation threshold. The same process can then be repeated for different sized neighbourhoods and accumulation thresholds. When complete, radar and model fractions will have been computed for all required thresholds over a range of spatial scales.

In this investigation, percentile rather than accumulation thresholds will be presented. For example, the 95^{th} percentile threshold selects the highest 5% of radar and forecast accumulations (within the whole verification area) for comparison. The purpose of doing this is to remove the impact of any bias in rainfall amounts (since the forecast and radar area of exceedance are forced to be the same) in order to focus on the spatial accuracy of the forecasts.

Figure 1 gives a schematic picture of how fractions are computed over different sized squares. In this example the threshold has been exceeded where the grid squares are shaded and not reached where the grid squares are white. If we focus on the central grid square, then at the grid square itself (i.e. the grid scale) the forecast fraction is 0/1 = 0, but the radar fraction is 1/1 = 1 (the forecast is wrong). Over a 3x3 square the forecast fraction is 4/9 = 0.44 and the radar fraction is 3/9 = 0.33. Over the whole 5x5 domain the forecast fraction is 6/25 = 0.24 and the radar fraction is also 0.24 (the forecast is correct for that specific central grid square over that scale).

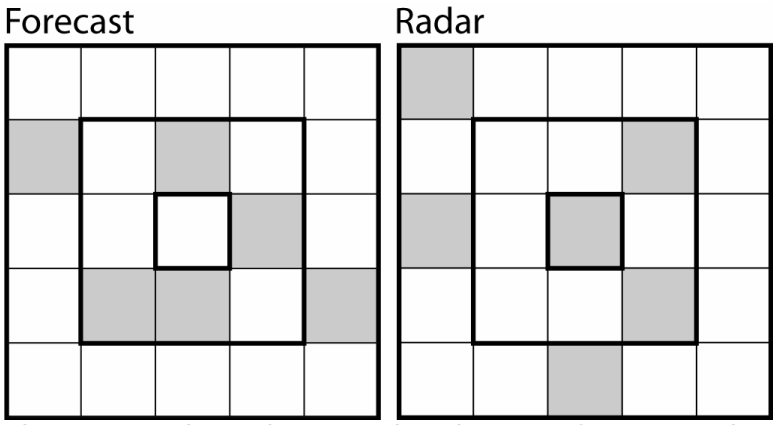

Figure 1. A schematic comparison between forecast and radar (see text)

2.2 Stage 2, computing the FSS

The FSS is a variation on the Brier Skill Score.

$$FSS = 1 - \frac{FBS}{FBS_{worst}}$$

Where *FBS* is the Fractions Brier Score and is a variation on the Brier Score (Brier, 1950) in which both the forecast and observed probabilities (fractions) can have any value between 0 and 1. It is given by:

$$FBS = \frac{1}{N}\sum_{j=1}^{N}(O_j - M_j)^2$$

M_j and O_j are the forecast and radar fractions respectively at each point, with values between 0 and 1. FBS_{worst} is given by:

$$FBS_{worst} = \frac{1}{N}\left[\sum_{j=1}^{N}O_j^2 + \sum_{j=1}^{N}M_j^2\right]$$

It is the largest *FBS* that could be obtained from the forecast and observed fractions when there is no collocation of non-zero fractions and therefore the worst possible *FBS*.

The FSS has the following characteristics:

1. It has a range of 0 to 1; 0 for a complete forecast mismatch, 1 for a perfect forecast.
2. If either there are no forecast grid squares which exceed the threshold and some occur, or some are forecast and none occur, the score is always 0.
3. As the size of the squares used to compute the fractions gets larger, the score will asymptote to a value that depends on the ratio between the forecast and observed frequencies of the event. I.e. the closer the asymptotic

value is to 1, the smaller the forecast bias. The use of percentile thresholds ensures that the FSS tends to 1 as the neighbourhood size approaches the size of the verification area.

4. The score is most sensitive to rare events (or for small rain areas).

A more complete discussion of the FSS is given in Roberts and Lean, including a discussion of the link between the verification method and probabilistic post-processing of precipitation forecasts.

For a collection of forecasts the FBS and FBS_{worst} from each forecast are combined and then overall FSS is computed. For a percentile threshold, this is the same as averaging the FSS from each of the forecasts.

3. The model and forecasts

The model examined in this study was the mesoscale version of the Met Office Unified Model (UM) (Davies et al, 2005). This model had a grid spacing of 12km and covered the area shown in Figure 2. It has now been superseded by the North Atlantic European (NAE) model which also has a grid spacing of ~12km and covers much of the North Atlantic and Western Europe. For purposes of this paper it is not important that the model is no longer operational because the intention of the work was not to scrutinise the performance of a particular forecast system, but to demonstrate the use of the methodology as a means of gaining insight into the spatial and temporal variation of skill in Quantitative Precipitation Forecasts.

Operational mesoscale-model forecasts starting at 00 UTC and 12 UTC from the whole of 2003 were used. From those forecasts, hourly precipitation accumulations within the first 24 hours were examined. Although some forecasts were missing from the archive it still represents a large sample of data. The model used 3D-Var (Lorenc et al., 2000) along with the Moisture Observation Preprocessing System (MOPS) (Macpherson et al 1996) with latent heat nudging (LHN) (Jones and Macpherson 1997) data assimilation (DA) techniques to obtain the best fit to observations at the start of each forecast.

The verification was performed on the shaded area shown in Figure 2. This is where there was considered to be sufficiently reliable radar coverage for this purpose.

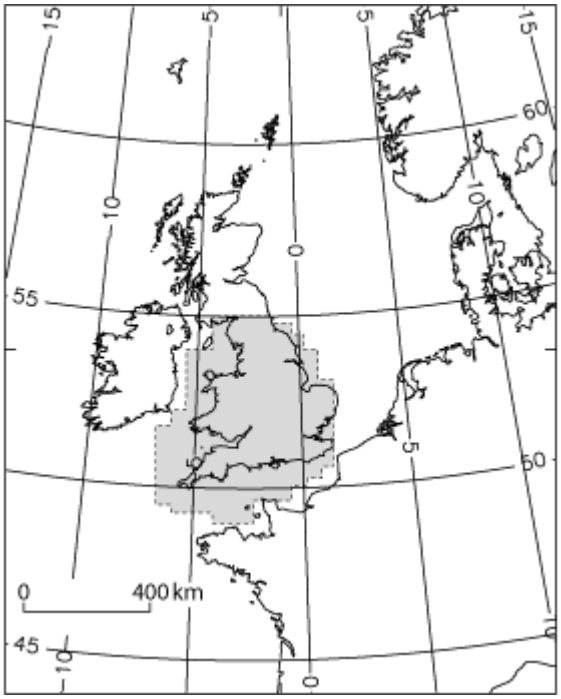

Figure 2. The mesoscale model domain, in which the verification area is shaded grey.

4. Results

4.1 Spatial and temporal variation in skill

The variation of skill with forecast length and spatial scale is shown in
Figure 3. The four panels show the variation of FSS with forecast time for the 75th, 90th, 95th and 99th percentile thresholds. The 99th percentile (top 1% coverage) represents the more localised and generally smaller rainfall events such as showers or the local maximum within larger areas of rain. The 75th percentile represents larger more widespread areas of rain. On average, the thresholds were equivalent to accumulation thresholds of 0.1, 0.3, 0.5 and 1.5 mm. These values are low because high hourly totals are rarely seen over 12x12km squares. The use of percentile thresholds to remove the effect of the bias is valuable for examining spatial information provided that the bias is not too large. For these data the ratio of radar coverage to model coverage (frequency bias) over the domain for selected accumulation thresholds was less than a factor of two for all but the high accumulations (>1mm), giving confidence that the assessment of spatial accuracy is worthwhile, especially given the uncertainties in the radar data (Harrison et al 2000). If the frequency bias had been very large, then the general over prediction of rain would dominate at all scales and an investigation of spatial accuracy would be more difficult to justify.

Two horizontal dashed lines are drawn on each panel. The lower line shows the FSS that would be obtained for that percentile threshold from a purely random set of forecasts. The upper line is the FSS half way between the random skill and perfect skill. It has been shown in idealised experiments to be the equivalent to the FSS that would be obtained from a neighbourhood of length twice the mean spatial error and

can be regarded as a minimum required level of skill for a useful forecast (Roberts and Lean).

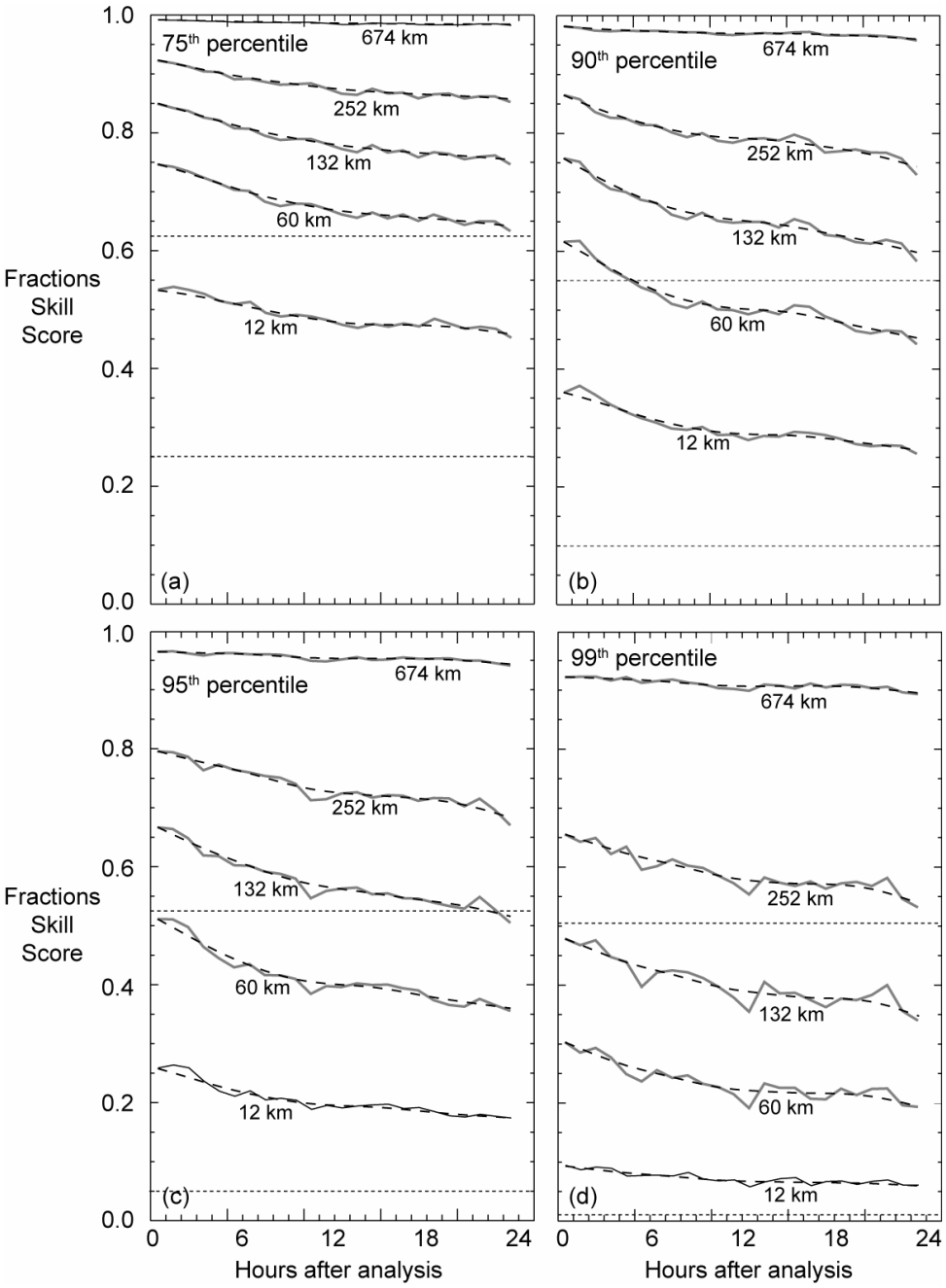

Figure 3. Plots showing the change in FSS for hourly accumulations for (a) 75th, (b) 90th, (c) 95th and (d) 99th percentile thresholds. Each panel shows spatial scales (neighbourhood lengths) of 12, 60, 132, 252 and 674 km. The smoother dashed lines were obtained using a 5 point running mean iterated 3 times. The horizontal dashed lines represent reference levels of skill described in the text.

Inspection of
Figure 3 shows that the forecast skill improves with spatial scale for all the thresholds, or put another way, spatial errors in the forecasts become less important

as the area of interest is increased. It is also evident that skill decreases with forecast length. An exception to this is during the second hour over small scales, where an increase in skill may be attributed to the addition of increments from LHN for two hours after the nominal analysis time. We also see that the model was more skilful at predicting the spatial distribution of widespread rainfall events (75th percentile) than the more localised events (99th percentile). For example, the 60km sampling square gives a FSS of between 0.2 and 0.3 for the 99th percentile and between 0.65 and 0.75 for the 75th percentile. The more jagged lines for the 99th percentile (smoothest for the 75th percentile) show the greater sensitivity of the FSS to the smaller more localised rainfall features.

At the grid scale (12km), the lines tend to flatten off with forecast time, particularly the 99th percentile threshold, for which the error growth had largely saturated and the forecasts were becoming close to behaving like random noise (FSS ~0.05). At the largest scale shown (674km), the forecasts did not lose much skill through the forecast or gain much skill from the DA and the lines are almost flat throughout. The biggest drop in skill occurred at the intermediate scales (60 to 252km). This suggests that these are amongst the scales at which the DA adds most skill (discussed more later), but are also the scales at which the largest proportion of the skill is lost (because the DA has provided the skill to lose, unlike at the largest and smallest scales).

4.2 The smallest scales over which the model has sufficient skill

By looking at where the curves intercept the upper dashed line (target skill), the smallest scale at which the model has 'useful' skill can be found. For the 95th percentile threshold this occurs at around 60km for the first hour of the forecasts and increases to around 130km for the 24th hour, which means that the minimum length scale over which the model is deemed to have achieved the target level of skill has roughly doubled in 24 hours.
Figure 3 only shows curves from selected scales to avoid too cluttered a diagram, so it can only be used to provide a rough estimate for the other thresholds. Table 1 lists the scales at which the target skill is achieved at 0-1 and 23-24 hours for each of the thresholds using the full set of curves that haven't been shown. The values in Table 1 show that the scale at which the model reaches the target skill approximately doubles over 24 hours for all the thresholds.

Percentile threshold	Scale at which the target skill is reached (km) +/- 10km		Approximate ratio of scales
	First hour (0-1)	Last hour (23-24)	
99th (localised rain)	140	230	~1.6
95th	75	140	~1.9
90th	40	85	~2.1
75th (widespread rain)	30	65	~2.2

Table 1

4.3 Variations in skill over the first six hours

The amount of skill lost at each scale can be examined in more detail. In particular, the loss of skill over the first few hours is of interest because it can be used to infer

the gain in skill from DA. The DA operates by updating a previous short forecast with new information to create a new analysis with which to start the next forecast. Provided that the average skill of the model remains unchanged over the verification period, the average loss of skill early on in the forecasts can be considered equivalent to the average gain in skill from the DA. Figure 4 shows the reduction in skill over the first six hours of the forecasts for each of the thresholds using the running means (dashed lines) in
Figure 3. The change in skill has a peak at spatial scales of between 40 and 100 km for the 75th, 90th and 95th percentile thresholds. They are the scales over which the DA operated most effectively and close to the scales at which the DA is expected to add most skill (background error covariances used in VAR had length-scales of 80-120km, MOPS/LHN ~20-80km) . The 99th percentile threshold is somewhat different because it has a double peak. It is possible that this is a result of the particular meteorological situations this threshold might examine. On occasions when the rainfall coverage was too small to be included in the other thresholds (<5% coverage) the rain is likely to have been the result of either localised convection or small areas of drizzle. Either of these may have a unique behaviour that is only detected with the 99th percentile threshold, but further investigation is needed.

The skill change approaches zero at the largest scales. This is a consequence of using percentage thresholds to remove the bias, so, by definition FSS=1 over the whole verification area at all forecast times. It should be noted again that any impact DA has on the bias would affect all scales. The skill change at 12km (grid scale) was much less than over scales in the range 30 to 200km, showing that there was relatively less benefit from DA at the grid scale.

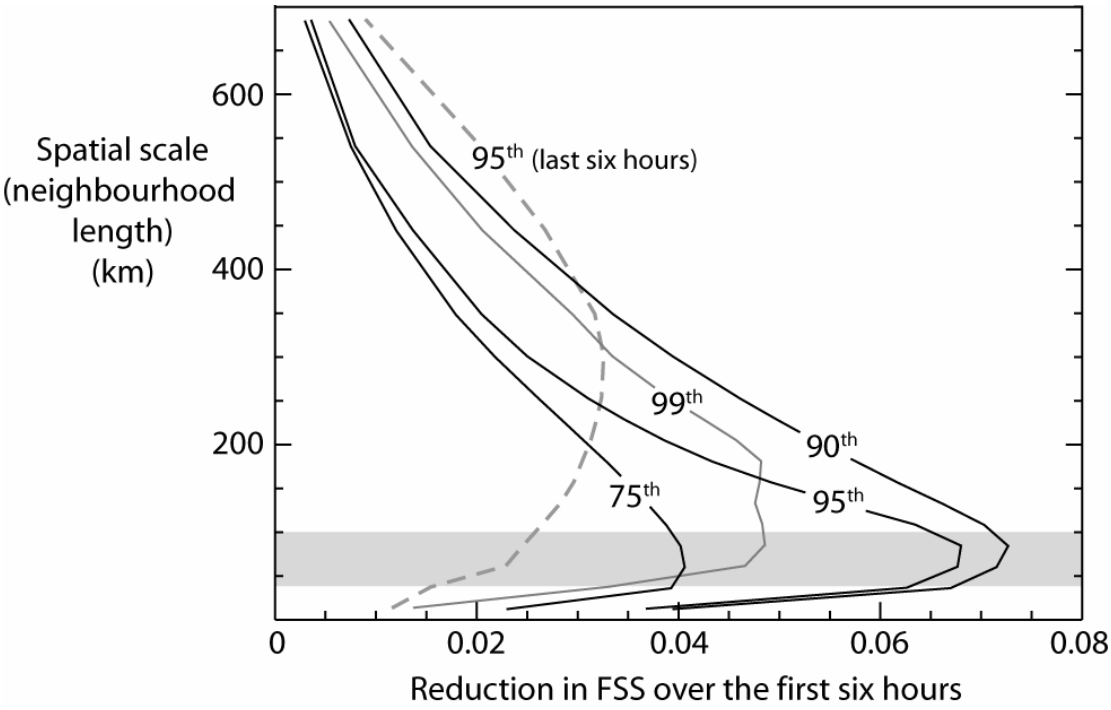

Figure 4. Graph of the amount of reduction in FSS over the first six hours against spatial scale for the 75th , 90th, 95th and 99th percentile thresholds. The grey shading indicates the scales at which the loss of skill peaked for the 75th, 90th and 95th percentile thresholds. A similar picture was obtained for the first 3 hours.

The intermediate percentiles (90th and 95th) have a bigger peak than the other two, which reflects the larger drops in skill for those thresholds (
Figure 3b & c) and therefore the larger impact of DA. In comparison, for 75th percentile (widespread rain), skill falls away more slowly and it does not require as much work from the DA to regain that skill. For the 99th percentile (localised rain), the initial skill is less because it is more difficult for DA to add skill, and so the amount of skill that can be lost is restricted. This reveals a trade off between the loss of forecast skill and the ability of DA to add skill that depends, not only on the spatial scale being examined, but also on the sizes of the rainfall features themselves.

During the last 6 hours of the 24-hour forecast period, the overall loss of skill was less and the peak had up-scaled and broadened (dashed line in Figure 4). This shows that later in the forecasts the biggest loss of skill had transferred to larger scales as more of the skill at smaller scales had already been lost (the flattening of curves at small scales discussed in section 4.1).

4.4 Skill half-life

Another pattern that emerged from Figure 3, but has not yet been discussed, is that there seems to be a greater proportion of the skill lost earlier in the forecasts at small scales for the 90th, 95th and 99th percentile thresholds, i.e. there is a poorer immediate retention of skill at small scales. To investigate this further, the time taken to lose half of the full 24-hour skill loss is plotted in
Figure 5. It shows that, for the 90th, 95th and 99th percentiles, the time taken to lose half of the total skill loss gets longer with spatial scale from around 35 km up to around 550 km. The implication is that the skill added by DA (however much that might be) was retained best in the forecasts over the largest spatial scales and lost most quickly at scales around 36km. All thresholds showed that the poorest retention of analysis skill occurred at 36km (3 grid lengths) rather than at the grid scale as might be expected. This may well be an indication that the information from the MOPS part of the DA (particularly LHN which is applied at scales of a few grid-lengths) was not well retained. The better retention of skill at the grid scale compared to scales up to ~100km is probably a result of less skill being added in the first place at the grid scale as shown in Figure 4.

The 75th percentile threshold showed little variation in skill retention with spatial scale. This makes intuitive sense, since for widespread rain events the retention of skill should become less dependent on spatial scale if there is no bias in rainfall coverage (seen by the lines becoming closer together in Figure 3d than in a, b or c). In the extreme of a 0th percentile threshold (forecast and observed rain filling the whole verification domain) the skill is always perfect at all scales.

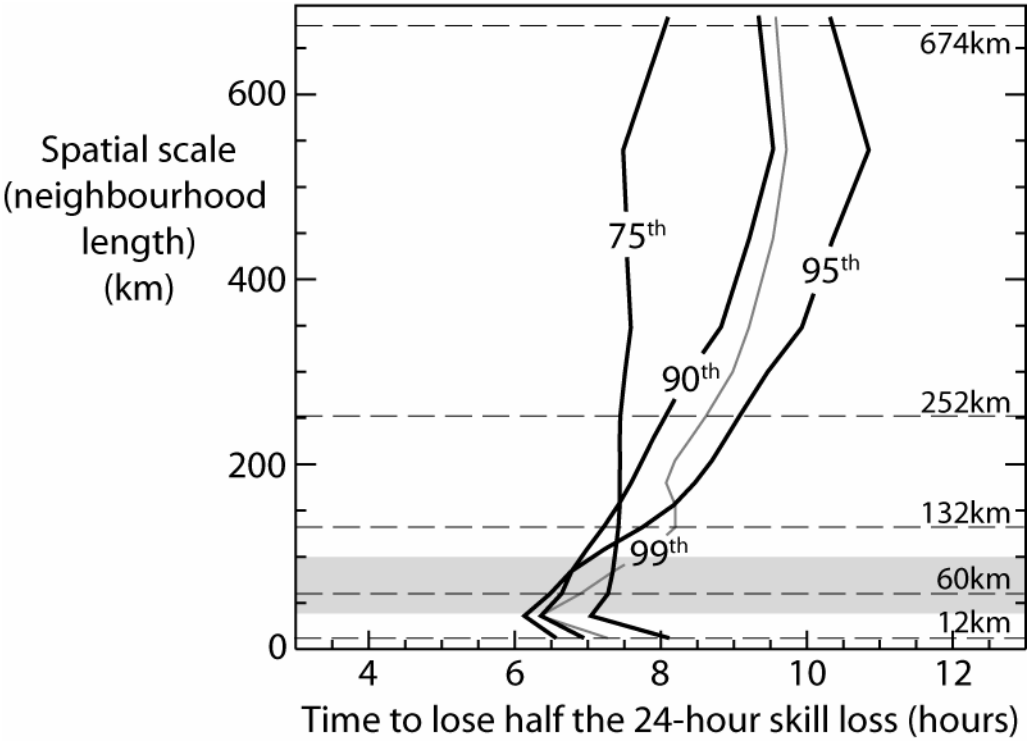

Figure 5. Graph of the time taken to lose half of the total 24-hour loss of skill against spatial scale for the 75th, 90th, 95th and 99th percentile thresholds. The grey shading is taken from Figure 4.

At the scales where DA adds most of the information (within or close to grey shading) the time taken to lose half of the 24-hour skill loss is between 6 and 8 hours. In addition, there appears to be a monotonic change in the shape of the curves between the 75th, 90th and 95th percentile thresholds, with more of the skill lost earlier in the forecasts as the rainfall is more localised. However, the 99th percentile curve does not quite fit that progression - perhaps for the reason discussed for Figure 4.

5. Conclusions

There is a need to be able to evaluate the spatial and temporal variation in skill of Quantitative Precipitation Forecasts (QPFs), particularly now, with the advent of high-resolution NWP models. Users of these forecasts will want to identify the scales over which there is sufficient skill for particular applications, such as flood warning, so that appropriate outputs can be developed. Model developers want to know about the scale-selective sensitivity of forecast skill to changes in DA or model formulation.

In this paper, the Fractions Skill Score (FSS) approach has been used to perform a scale-selective evaluation of precipitation forecasts from the Met Office mesoscale model. The purpose was to make use of the relatively large dataset to gain insight into the behaviour of an NWP forecast model and demonstrate the value of the method for assessing new models.

The results have shown that the skill of the model is dependent on both the scale over which the forecasts are being examined and the spatial coverage of the rain itself. I.e. it is easier to predict with reasonable accuracy the probability of it raining

over a large area than a small one, and when the rainfall is widespread rather than localised (given no serious systematic over/under prediction of the rain). Unfortunately, it is often the more localised rainfall events that produce the heaviest rain and it is therefore desirable to be able to find out how localised is 'too localised' to be generally predictable. The FSS has been used to determine the smallest scale, on average, over which the model has useful skill. It was found that even at the start of the forecasts the smallest useful scale for the very-localised rain (99^{th} percentile threshold) is around 140 km (12 grid lengths) and given that the rain areas themselves would be much smaller, it indicates that the model seriously misplaced the locations of the local rainfall maxima. However, for larger rainfall coverage (90^{th} percentile) the smallest useful scale is around 40 km (3-4 grid lengths) increasing to 85 km (7 grid lengths) after 24 hours, which is a considerable improvement. Such information is valuable for understanding how model output should be interpreted and from that how it should be presented. It is hoped that the current development of high-resolution NWP models will lead to a reduction in the scales over which forecasts are useful and this verification approach can help to answer that question. Evidence from Roberts and Lean suggests that this may well be the case.

In addition to investigating the scale-dependence of the skill, the evolution with time was also examined. It was found that the smallest useful scale roughly doubled in 24 hours. It was also possible to identify the scales at which DA operated most; they were between 40 and 100km with a big drop towards the grid scale. Skill was lost most rapidly early on at the smaller scales and this was particularly noticeable at 3 grid lengths (36km) which points to a particular difficulty in retaining information at that scale in this model. It's also notable that DA had the largest impact on the intermediate thresholds (90^{th}, 95^{th} percentiles) and was less effective for the localised (99^{th} percentile) and widespread rain (75^{th} percentile). For widespread rain the skill remained high and that restricted how much could be added, for the localised rain the skill was low because it dropped more quickly and was difficult to add. The balance between loss of skill with time and addition of skill from DA therefore depends on both the sizes of the rainfall features and the scale being verified.

It will now be interesting to examine what happens in higher-resolution models. Presumably there will be a more rapid loss in skill at the smaller scales. However, if DA is not able to add information at those scales there will be little skill to retain. If DA can add skill at the small scales, then it will be important to see if a more rapid loss of skill undermines that benefit and how much it may affect the larger scales of interest.

Acknowledgements

I would like to thank Peter Clark, Mark Dixon and Brian Golding at the Met Office for valuable comments about this work.

References

Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather. Review* 78: 1-3

Briggs WM, Levine RA. 1997. Wavelets and field forecast verification. *Monthly Weather. Review* 125: 1329-1341

Bousquet O, Lin CA, Zawadzki I. 2006. Analysis of scale dependence of quantitative precipitation forecast verification: A case-study over the Mackenzie river basin. *Quarterly Journal of the Royal Meteorological Society* 620: 2107-2125

Casati B, Ross G, Stephenson DB. 2004. A new intensity-scale approach for the verification of spatial precipitation forecasts. Meteorological Applications 11: 141-154

Davies T, Cullen, MJP, Malcolm, AJ, Mawson MH, Staniforth A., White, AA, Wood, N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Soc*iety 608: 1759-1782

Davis C, Brown B, Bullock R. 2006. Object-based verification of precipitation forecasts. Part I. Methodology and application to mesoscale rain areas. Monthly Weather Review 134: 1772-1784

Done J, Davis CA, Weisman M. 2004. The next generation of NWP: explicit forecasts of convection using the weather research and forecasting (WRF) model. Atmospheric. Science. Letters 5: 110-117

Ebert EE, McBride JL. 2000. Verification of precipitation in weather systems: determination of systematic errors. Journal of Hydrology 239: 179-202

Golding BW. 1998. Nimrod: A system for generating automated very short range forecasts. *Meteorological Applications* 5(1): 1-16.

Harrison DL, Driscoll SJ, Kitchen M. 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. Meteorological. Applications 7: 135-144.

Lorenc, AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker, DM, Bray JR, Clayton AM, Dalby TD, Li D, Payne, TJ, Saunders FW. 2000. The Met. Office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Soc*iety 126: 2991-3012.

Lorenz EN. 1969. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences 26: 636-646.

Mass CF, Ovens D, Westrick K, Colle BA. 2002. Does Increasing Horizontal Resolution Produce More Skilful Forecasts? Bulletin of the.American Meteorological Society 83: 407-430.

Marzban C, Sandgathe S. 2006. Cluster analysis for verification of precipitation fields. Weather and. Forecasting 21(5): 824-838.

Theis SE, Hense A, Damrath U. 2005. Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. Meteorological Applications 12: 257-268.

Roberts NM, Lean HW. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*: Accepted for publication.

Zepeda-Arce J, Foufoula-Georgiou E, Droegemeier KK. 2000. Space-time rainfall organization and its role in validating quantitative precipitation forecasts. Journal of Geophysical Research (Atmospheric) 105: 10129-10146

Zhang F, Snyder C, Rotunno R. 2003. Effects of moist convection on mesoscale predictability. Journal of the Atmospheric Sciences 60: 1173-1185