DEPARTMENT for ENVIRONMENT, FOOD and RURAL AFFAIRS **CSG 15**

Research and Development

Final Project Report

(Not to be used for LINK projects)

Two hard copies of this form should be returned to:
Research Policy and International Division, Final Reports Unit
DEFRA, Area 301
Cromwell House, Dean Stanley Street, London, SW1P 3JH.
An electronic version should be e-mailed to resreports@defra.gsi.gov.uk

Project title	Storm scale numerical modelling		
DEFRA project code	**FD2207**		
Contractor organisation and location	Met Office FitzRoy Road, Exeter, Devon EX1 3PB United Kingdom		
Total DEFRA project costs	£ 90000		
Project start date	01/12/2002	Project end date	31/12/04

Executive summary (maximum 2 sides A4)

Background

Short range prediction of precipitation amount is a critical input to flood prediction and hence to the accuracy of flood warnings. The most accurate current methods are based on extrapolation of radar analyses. At best, these provide sufficiently accurate predictions of rainfall for flood warnings up to about three hours ahead and inherent limitations will reduce the scope of long term improvements in quantitative precipitation forecasts by this method.

The next major advance in quantitative precipitation forecasting methods is expected to come from the application of storm scale Numerical Weather Prediction Models. The Met Office is at the forefront in the development of such a system. It is now possible to run with a sufficiently fine grid to resolve the critical scales for flood-producing rainfall, i.e. about 1km grid spacing. At present, this is only possible in a research environment, but as more powerful computer resources become available, operational implementation will become a viable option.

The purpose of this research project was to investigate the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall up to 12 hours ahead and to develop appropriate tools for interpreting and presenting the predictions so that they enhance operational flood prediction capabilities.

The results are to be used to determine whether a storm scale NWP suite should be implemented in support of flood prediction, and if so, what its configuration and outputs should be.

Objectives

1. Investigate and report on the ability of the Met Office Unified Model, running at very fine resolution, to simulate the rainfall evolution for up to 12 hours from initialisation time in a variety of heavy rainfall events.
2. Develop a method for evaluating the fit of rainfall forecasts to radar on a variety of scales; particularly those important for flood prediction.
3. Investigate and report on the ability of the Met Office Unified Model assimilation scheme to fit the rainfall distribution observed by radar in these cases.
4. Investigate and report on the sensitivity of forecast accuracy to the specification of the initial conditions, especially to the accuracy of fit to the radar observed rainfall distribution.
5. Investigate and report on the sensitivity of forecast accuracy to tuneable parameters in the Unified Model.
6. Develop and report on suitable model diagnostics to optimise the usefulness of model output for flood prediction, both by manual interpretation, and through input to hydrological models.
7. Assess the potential benefits for flood prediction from the use of a very fine resolution version of the Unified Model
8. Make recommendations on the most effective way to achieve this, including model configuration, data assimilation, assessment of uncertainty and presentation of results.

Achievements

The model was run successfully with a grid spacing of 1 km for a number of convective events. As it was a new system, there was no guarantee at the outset that this would be the case. Once it became established that a storm-resolving model was capable of producing realistic simulations, a systematic investigation of the sensitivity to a variety of tuneable parameters was carried out. The findings provided the basis for a standard configuration on which to base subsequent experiments. A particular concern was to discover whether it is necessary to include a convection parametrization to represent showers that can not be properly resolved on the grid. Studies showed that this was not essential on a 1km grid, but on the intermediate 4km grid used to supply information to the 1km model, it is necessary. As a result, a modification to the operational convection parametrization was tested and implemented in the 4km model.

A large number of diagnostic products have been developed to optimise the usefulness of storm-scale model precipitation forecasts for flood prediction. They are designed to extract the most useful information on scales that are expected to be predictable and provide guidance on the likelihood of more uncertain events. Since these products require knowledge of the spatial accuracy of the model, a method of evaluating the performance of rainfall forecasts against radar over different spatial scales has been developed. The technique has been used to compare the skill of several storm scale forecasts with the equivalent forecasts from the operational 12km model. It has also been used to assess the performance of operational 12km model precipitation forecasts over a year long period as a control study.

The project has met its objectives and the outcome is a number of significant scientific results that will guide future research and operational implementation of a storm scale model.

Results and Recommendations

A storm-scale model (grid spacing ~ 1 km) is capable of producing significantly more realistic and spatially accurate forecasts of convective rainfall events than is possible with current operational systems. Evidence has come from detailed investigation of selected cases and from performance statistics over a larger sample. There is now a prospect of producing useful forecasts of convective storms on scales applicable for flood prediction.

- The development of a storm scale modelling system should continue towards operational implementation. This includes necessary core research activities.

Diagnostic products have been developed to enhance the interpretation of the rainfall forecasts. The use of such products is essential if a storm-resolving model is to be used for flood prediction.

- The generation of output products for flood warning should be developed as an integral part of any operational system.

Further research is required before an operational capability for storm scale NWP can be implemented, especially:

- Assess ability of the model to predict extreme events
- How to blend model forecasts with radar extrapolation forecasts
- How to feed output into hydrological models
- How to incorporate high-resolution observations.

Scientific report (maximum 20 sides A4)

Introduction

The objective of the storm scale modelling project was to investigate the ability of a storm scale configuration (with a 1-km grid spacing) of the Met Office NWP model to predict flood-producing rainfall up to 12 hours ahead and to develop appropriate tools for interpreting and presenting the predictions so that they have the capacity to enhance operational flood prediction capabilities.

To put this in context, the highest resolution that has been run operationally to date by the Met Office is the 12 km gridlength used in the UK mesoscale model. This is insufficient to resolve the majority of convective storms over the UK (or anywhere else). A model with a grid length of 1 km should be capable of representing many more of these storms.

The scientific focus of the project was split into five stages, each comprising a separate area of research within the overall context of the scientific objectives. At the end of each stage, the results were documented in an end of stage report. At the end of the project, a final scientific report was written to tie together the findings. The direction of the project followed a natural progression. At the start, the main concern was to examine whether a storm-scale configuration was able to produce sensible and realistic looking forecasts. Once it became established that the model does indeed have that capability, the characteristics of the model could be examined more closely. Work could then proceed on testing the sensitivity of the model to key parameter changes, generating specialised output products, developing a new approach for evaluating performance and investigating forecast skill and the impact of data assimilation.

Scientific Stages

1. Initial case studies

Four case studies of different types of thunderstorm events were chosen to provide a variety of meteorological situations for testing the model. High resolution simulations (1 or 2 km gridlength nested inside a 4 km gridlength model) were run to see how realistic the forecasts were and to get a subjective assessment of how well the model performed in comparison to the operational 12 km model. In these tests the high-resolution forecasts all started from the same initial conditions as the 12 km forecasts.

The main finding from the subjective assessment of the first four case studies was that the performance of the high-resolution (1 or 2 km grid length) simulations is very encouraging. In three of the cases, the highest resolution forecasts were considerably better than the 12-km forecast model in predicting the location, structure and intensity of convective rainfall events. Even when the 12-km model performed well, the higher resolution was able to produce a more realistic rainfall structure and intensity. Some care is required in the interpretation of these results as it was a subjective assessment of only 4 cases. An objective evaluation of a larger sample of events was carried out later in stage 5 of the project

This assessment was made more objective by constructing a table (Table 1) to provide scores for fundamental characteristics of the forecasts. These have then been combined to give an average score for each of the model resolutions. The 1 and 2-km forecasts were combined in this process because otherwise the figures would be unbalanced for such a small sample.

The individual scores are from 0 to 5.
0 = no skill, 1 = very poor, 2 = poor, 3 = OK, 4 = good, 5 = extremely good

The four categories are:

1. Rainfall accumulations – how well did the forecast produce the observed accumulations somewhere in the area of interest.
2. Spatial accuracy – did the forecast produce the rain in the correct place.
3. Temporal accuracy – did the forecast produce the rain at the correct time.
4. Precipitation structure – how well did the forecast simulate the correct precipitation structures (e.g. squall line, scattered showers, comma cloud, embedded frontal convection etc).

	Model grid spacing															
	12km				4 km				2 km				1 km			
	Cases				Cases				Cases				Cases			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Rainfall accumulations	3	1	2	0	4	2	3	0	4	5	-	-	-	-	4	2
Spatial accuracy	3	1	3	0	3	2	4	0	3	4	-	-	-	-	3	3
Temporal accuracy	5	5	1	0	3	5	3	0	3	5	-	-	-	-	4	2
Precipitation structure	3	3	1	0	4	4	3	0	4	4	-	-	-	-	4	5
Average scores	**1.94**				**2.50**				**3.69**							

Table 1. Subjective scores for various aspects of forecast performance for the resolutions examined in the four case studies.

2. Sensitivity studies and parameter selection

An examination of the sensitivity to various parameter changes was carried out in parallel with a systematic testing of possible model configurations within the High Resolution Trial Model (HRTM) project to establish the most appropriate setup to adopt as a standard for further experiments. Decisions were made about the following:

Domain size and location

The domains for the 12km (operational), 4km and 1km models are shown in Figure 1. Information is fed from the 12km domain through the boundaries of the 4km domain, and in turn from the 4km domain through the boundaries of the 1km domain. These domains were used for all further experiments, except when an extreme event such as the flash flood at Boscastle Cornwall warranted a shift of the 1km domain.

Number of gridpoints
12km 146 x 182
4km 190 x 190
1km 300 x 300

Vertical resolution

Two sets of vertical levels were tested for both the 4 and 1-km models. Firstly, the set of 38 levels used operationally in the ~60-km global model and 12-km mesoscale model was used, then a doubling to 76 levels (an extra level between each existing level) was examined. The levels are not equally spaced; both sets have many more levels low down (a few tens of metres spacing near the ground) than towards the top (a few kilometres spacing in the stratosphere). It was decided to use 38 levels as the default in the 4-km model so that data assimilation could more easily be applied and because it is cheaper to run. However, for the 1-km model the use of 76 levels is regarded as physically more appropriate.

Time step

The time step is the time interval over which the equations in the dynamics part of the model are solved. A longer time step is more economical but less accurate. The higher the resolution the shorter time step must be to maintain sufficient accuracy. It is partly why a high resolution model is more expensive to run. The choices were 1 minute for the 4km model and 30 seconds for the 1km model.

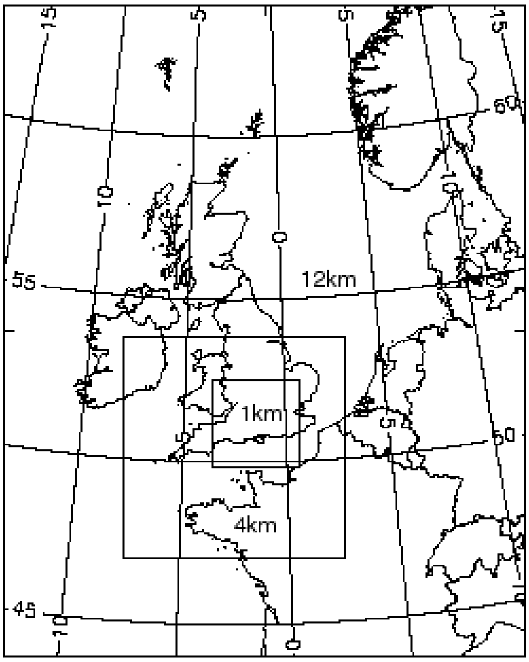

Figure 1. The domains used for the 12, 4 and 1km grid-length models.

Diffusion

The diffusion equation can be used to apply smoothing to temperature, humidity and wind fields. However, the use of diffusion is sometimes contentious. The operational 12-km model is run without any diffusion because it is thought that it should not be needed with a semi-lagrangian formulation for the dynamics and we do not wish to remove important detail from the forecasts. At higher resolution the addition of a small amount of horizontal diffusion has been necessary to reduce the occurrence of unrealistic grid-scale precipitation patterns associated with convection.

Boundary updating

The operational 12-km grid-length model supplies the values that are needed at the boundaries of the 4-km model, which in turn supplies the boundary information for the 1-km model. Ideally, the boundaries should be updated as often as possible (every time step) to eliminate significant mismatches between what the inner and outer models think is happening at the inner-model boundary. Boundary update intervals of 30 minutes for the 4km model and 15 minutes for the 1km model were chosen because they are the longest that give satisfactory results in most situations.

Critical Relative Humidity

This is the relative humidity at which the model diagnoses that some cloud has formed in a grid box and is less than 100%. Changes to RH-crit made some differences to the detail, but not the essence of the forecasts in the convective cases examined. For that reason, the values of 85% at most model levels used in the 12-km mesoscale model were retained for the higher resolution grids.

Convection parametrization scheme

The vast majority of convective storms can not be resolved properly by the operational 12-km grid-length mesoscale model. This has meant that a convection parametrization scheme has been required to represent the average effect of showers within each grid box. The big advantage of going to a 1-km grid-length model is that most convective showers can then be resolved by the model dynamics. It should be possible for the model to simulate the initiation, structure and decay of individual convective events. However, it is still not be possible at that resolution to resolve the smaller showers or storm cells and it is conceivable that there will be a particular difficulty in simulating the very first stages of shower development. The question that then arises is whether there is still a need to use a convection scheme in a 1-km grid-length model? The current view is that it is better to switch it off, as we want the model dynamics to do the work unhindered.
Results from case studies have shown that a 1-km grid-length model with no convection scheme (but with numerical diffusion) does produce realistic forecasts. For that reason, it is currently the default to switch off the convection scheme for 1-km simulations.

The situation becomes very different when the grid spacing is 4-km. At that resolution, larger convective systems will be resolved by the model, but most convective events can not be resolved properly, particularly at the early stages of development. Tests at 4km revealed a problem; the convection scheme was too active and tended to remove the convective instability before the dynamics had any opportunity to represent any showers on the grid. A modification to the convection scheme has been made so that a limit is applied to how active it is allowed to be. This was presented in the stage 2 report. It has improved test simulations and is now used in the default 4-km configuration.

It might seem strange to have concentrated so much on modifying the convection scheme in the 4-km model when it is the 1-km model we are primarily interested in, but there are good reasons for doing so. Firstly, the 4-km model supplies the information at the boundaries of the 1-km domain and if the wrong information is coming through the boundaries the 1-km forecast will give poor forecasts. Secondly, the approach can also be applied to the 1-km model (or any model with a grid-length greater than ~0.5 km and less than ~ 20 km). Thirdly, we should bear in mind that a ~4-km model will be implemented operationally before a 1km model..

3. The development of products

The case studies have indicated that a storm-scale model does indeed have the potential to deliver more accurate forecasts of high-impact rainfall events than our current operational systems can achieve. However, we should be realistic about the remaining difficulties. One of the major concerns often raised about a storm-scale model is that it will be attempting to resolve features that are inherently unpredictable within the time period over which forecasts will be run. In other words, a model with a grid spacing of 1 km will have the capacity to resolve quite small showers, but the exact location of any individual shower is not very predictable beyond a short period of time (perhaps minutes). Further uncertainty is introduced because we can never know the exact state of the atmosphere at the start of a forecast and errors will grow with forecast time.

Despite these difficulties we still expect a storm scale model to be noticeably more accurate over larger more predictable scales than current operational systems. If those scales coincide with the sizes of sensitive river catchments, then a storm-scale model becomes a valuable tool for flood-prediction, but only if the output is interpreted appropriately. It means that there is a need to develop diagnostic products to present the information on scales that are both reliable and meet customers' requirements.

Several diagnostic products have been developed; they are presented in the stage 3 project report. Further examples are presented in the final scientific (stage 6) report.

4. A strategy for forecast assessment

The performance of NWP models can not be assessed properly by just looking at the output and judging 'by eye'. Some kind of objective evaluation is also required if we wish to reveal systematic model behaviours.

The problem is that traditional methods for evaluating model performance have generated statistics or scores on a gridpoint by gridpoint or observation point by observation point basis alone. That is not a viable approach here because we do not expect the model to be skilful at the grid scale (because of reasons presented above) and it says nothing about skill over any other scales. The spatial scales we are interested in, and the spatial scales we verify on, should be compatible. We judge forecasts by eye over a variety of scales by saying that they are good at getting a large feature like a front in the right place, but are not quite so accurate for organised thunderstorms and poorer still for individual showers. But how can that be determined objectively?

There is a need to be able to assess the performance of precipitation forecasts from a high-resolution NWP model in a way that measures the accuracy of a forecast over different spatial scales. A storm-scale model is not likely to be very skilful at the grid scale, but could nevertheless provide very useful information over say an area the size of a river catchment. A technique for evaluating the skill of rainfall forecasts over different spatial scales has been developed and used successfully.

5. Assessment of model performance

Objective performance scores were computed using the scale-selective verification method. They were obtained from 16 7-hour precipitation forecasts over four separate days in spring/ summer 2003. The four days were chosen purely because significant convection occurred in the 1-km model domain, and are different to the initial case studies. They were not selected on the basis of the operational 12-km model's performance. For each of the four days, forecasts were run that started at 06, 09, 12 and 15 UTC.

The days chosen were:
1. 13th May 2003
2. 25th May 2003
3. 1st July 2003
4. 28th August 2003

The aim was to examine both the impact of changing model resolution and the impact of using data assimilation at high resolution on forecast skill over different spatial scales. Verification scores were obtained for hourly and 6-hourly rainfall accumulations.

The graph in figure 2 shows the skill of 12km and 1km model forecasts at predicting locations of the highest 6-hour rainfall accumulations (after removal of bias). In both models the skill increases at larger scales, but the 1km model is more skilful overall, particularly at the intermediate scales from 20-60km. The verification results can be used to determine the most appropriate scales on which to present rainfall forecasts.

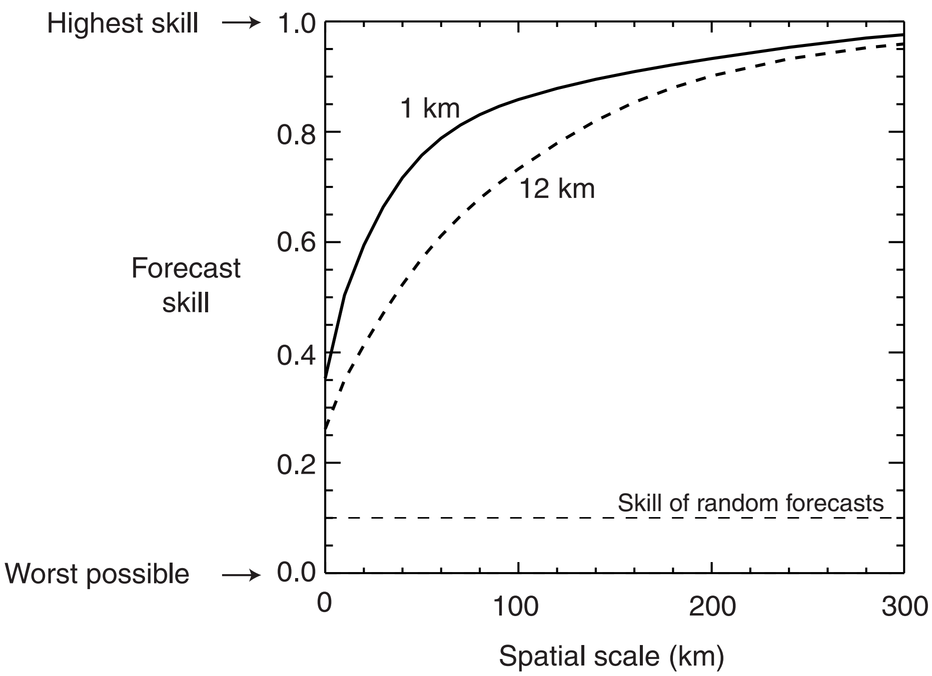

Figure 2. Skill, compared with radar, in predicting the location of the top 10% rainfall accumulations over different spatial scales for 16 cases

Results

The storm-scale model produced more realistic forecasts

- A high-resolution NWP model (grid spacing 4km or less) is capable of producing much more realistic simulations of convective rainfall events than the current operational mesoscale model (grid spacing 12 km). The 'storm scale' model forecasts (1-km grid spacing) gave the most realistic representation of rainfall patterns and intensity. The evidence for this conclusion comes primarily from the visual examination of the 9 events presented in the project reports, and also from inspection of further cases run within the High Resolution Trial Model (HRTM) project. This result was anticipated because a storm scale model by design is able to resolve features that can only be represented by a convection scheme in a 12-km grid-length model. Even so, the results are impressive.

The storm-scale model produced more accurate forecasts on scales applicable for flood prediction

- A new verification methodology for comparing rainfall accumulation forecasts with radar in a scale-selective way has been used to assess forecast performance over a number of events. The results show that the forecasts performed better at all resolutions when the target area was larger, i.e. it was easier to predict that it would rain somewhere within a large area than a small one. Most important though, was the discovery that the high-resolution (1 and 4km) forecasts out-performed the 12-km forecasts over nearly all spatial scales. In particular, they gave more accurate predictions of higher rainfall accumulations over areas the size of small to medium sized river catchments. The best scores for these criteria were obtained from the 1-km grid-length model. Since this project is concerned with the use of a storm-scale model for flood forecasting, these are encouraging results. Not only can a high-resolution model produce more realistic simulations, but it is also more accurate, at least up to 7 hours ahead, over scales that are important for flood prediction.

Diagnostic products are vital for forecast interpretation

- The use of diagnostic products is essential for a meaningful presentation of rainfall forecasts from a storm-scale model, especially when used in a flood forecasting context. The objective verification results showed that over smaller spatial scales, even the high-resolution simulations can sometimes have little skill. Certainly, any scale close to the grid-scale of the model can not be relied upon. Sensitivity studies have revealed that small changes to the values of some parameters will have an impact on rainfall patterns in a forecast. The problem in interpretation arises because high-resolution forecasts can look so realistic and it is tempting to believe the detail, when in fact a different forecast outcome could look just as realistic. Diagnostic products that were designed to exploit the advantages of the storm-scale model for the purpose of rainfall prediction, but still take forecast uncertainty into account, have been shown to add considerable benefit to the raw model output. Any storm scale forecast model should have an appropriate post-processing system built in, otherwise the full potential of the model as a forecasting tool can not be realised and there is a danger of misinterpretation. There is considerable scope for products to be developed that can generate different forecast scenarios for input into hydrological models.

Data assimilation at high resolution requires further development to give consistent benefit

- Scale-selective verification was used to examine the impact of data assimilation. The results showed that data assimilation at high resolution gave a better fit to radar at the early stages of the forecasts, but tended to lead to a poorer distribution of rainfall after a few hours compared to forecasts that had to 'spin up' from the 12-km grid-length model fields. On the other hand, the spin-up forecasts produced too much rain after a few hours and this was improved by the data assimilation. The 1-km forecasts with data assimilation scored higher than the other forecasts for a rainfall accumulation threshold of 8 mm (the highest accumulation it was possible to examine over these events). These are very preliminary results, though nevertheless a useful measure of current capability. The data assimilation used at 4 km was essentially the same as that used operationally at 12 km (3DVAR and MOPS cloud analysis and latent heat nudging, see stage 4 report). The data assimilation at 1 km involved the addition of increments from 3DVAR at 4 km. More appropriate methods for high-resolution models are being developed at the Joint Centre for Mesoscale Meteorology (JCMM). It is clear that some kind of data assimilation on the high-resolution model grid is essential if we are intending to use such a model for very short range forecasting up to a few hours ahead. It is not clear whether the same can be said for forecasts beyond a few hours. Care should be taken with these findings however, they have been obtained from a small sample and further objective testing is required.

The results have provided evidence that a storm-scale model does indeed have the potential to deliver a significant improvement in our ability to predict high-impact convective rainfall events.

Recommendations

- The development of a storm scale modelling system should continue towards operational implementation. This includes necessary core research activities.

- The generation of output products for flood warning should be developed as an integral part of any operational system.

Three new avenues of research have been identified as necessary if the potential of a storm scale modelling system for flood prediction is to be fully realised. They are:

1. To assess and optimize the ability of the next generation Numerical Weather Prediction (NWP) model to predict extreme rainfall events, i.e. those with return

periods in excess of 100 years. It is important to understand how well the model is able to represent such events and the reasons behind aspects of model performance.

2. To optimize short range rainfall predictions by blending ensemble nowcasting methods with convective scale numerical weather prediction forecasts. If successful, this could result in a substantial improvement in rainfall predictions over a few hours.
3. To incorporate rainfall predictions from convective scale Numerical Weather Prediction (NWP) models into hydrological models for flood prediction.

They are complementary to core research activities into NWP model development and data assimilation that also need to be maintained.

R&D Outputs

Seven technical reports present technical information and research findings from the project. 1 to 5 are the end of stage reports. 6 is the final scientific overview. 7 is a review of the project.

1. ***Results from High Resolution Simulations of Convective Events*** Published February 2003, Met Office Forecasting Research Technical Report 402 (JCMM Internal report 140)
2. ***Report on Sensitivity of Case Studies to Model Parameters and Options*** Published March 2003, Met Office Forecasting Research Technical Report 407 (JCMM Internal report 142)
3. ***Precipitation Diagnostics for a High Resolution Forecasting System*** Published September 2003, Met Office Forecasting Research Technical report 423 (JCMM Internal Report 143)
4. ***Measuring the Fit of Rainfall Analyses and Forecasts to Radar*** Published January 2004, Met Office Forecasting Research Technical report 432 (JCMM Internal Report 146)
5. ***Verification of the Fit of Rainfall Analyses and Forecasts to Radar*** Published April 2004, Met Office Forecasting Research Technical Report 442 (JCMM Internal Report 148)
6. ***An Investigation of the Ability of a Storm Scale Configuration of the Met Office NWP Model to Predict Flood-producing Rainfall*** Published December 2004, Met Office Forecasting Research Technical Report 455 (JCMM Internal Report 150)
7. ***Review of the Storm Scale Modelling Project and Proposals for Future Research*** Published January 2005