Forecasting Research

Forecasting Research Division
Technical Report No. 156

Report on the First Nimrod Trial

by

Brian Golding

March 1995

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Forecasting Research Division

Technical Report No.156

Report on the First Nimrod Trial

by

Brian Golding

March 1995

Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
ENGLAND

Report on the First Nimrod Trial

Brian Golding, Project Manager

Conduct of the trial

The trial started on 1st January 1995. Since 3rd January, three forecasts per weekday have been assessed subjectively. During January and early February, a number of improvements were made to Nimrod in response to criticisms arising in the subjective assessment. These included use of additional data (METARs) in the anaprop removal, limitations on the long range radar corrections in showers, and default extrapolation vectors for objects which cross the boundary of the radar composite. Objective assessment was introduced on 17th February. Several heavy rain belts crossed the UK during the latter part of February. During March there was a trend to drier weather. Statistics are presented for the period from 17th February to the 14th March, when the trial was suspended.

Assessment Criteria

The original criterion for automation of FRONTIERS was that the results should be as good as the FRONTIERS results. The specific criteria laid down in the project plan to test this did not take account of the possible availability of objective comparisons. For this reason, they rely exclusively on the subjective assessments. The criterion for the actuals was that Nimrod should be 'better on average'. In terms of the assessment method outlined below, this was interpreted as: Nimrod should be significantly better more often that FRONTIERS is significantly better. For the forecasts, the criterion was that FRONTIERS forecasts should be significantly better on less than 10% of occasions (regardless of how often Nimrod forecasts were better).

Actuals

Method of objective assessment

Assessment is against hourly reporting raingauges: SREWs, automatic stations, and radar calibration gauges. 15min composite radar actuals are summed to give hourly accumulations: pixels are only included where at least 3 observations of the pixel have been made in the hour. Due to the uncertainties resulting from a finite tip volume, radar actuals are only compared with gauges when rain is measured and the radar accumulation is at least 0.1mm. Three statistics are calculated: mean, RMS and RMSF.

Results

The period assessed was 17/2/95 - 14/3/95
The number of gauge comparisons was 4450.

	Mean(mm/hr)	RMS(mm/hr)	RMSF
FRONTIERS	0.24	1.19	2.97
Nimrod	0.19	1.10	2.59

The UKON 7 target for RMSF is 1.58 ($\log_{10}1.58 \approx 0.2$), so a reduction of 0.38 in a 1.39 excess represents a significant improvement.

Method of subjective assessment

The first stage of assessment is to compare the FRONTIERS and Nimrod actuals. If there are significant differences, assessment is primarily against synoptic present weather reports, but continuity of the radar, satellite imagery, and experience with similar situations on FRONTIERS are also used. The full FRONTIERS area is assessed. Four scores are possible: No significant differences; Significant differences but neither better; Significant differences and FRONTIERS better; Significant differences and Nimrod better. Comments were recorded to justify the mark.

Results

The period assessed was 17/2/95 - 14/3/95
The number of comparisons was 111

	FRONTIERS better	Nimrod better
Spurious rain left	2	0
Rain removed	1	0
Rain rate	2	1
Satellite rain	3	4
Others	0	1
Total as percentage	7.2%	5.4%

Discussion

The objective scores show that Nimrod has a clear superiority in its rain rate estimates. Given the poor corroborating data available to the forecasters, I suggest we should take the objective score as a better assessment of rain rate differences than the subjective score.

Other differences are better dealt with by the subjective assessment. Remarking of samples of subjective comparisons suggests that there is an uncertainty in the results which should be accounted for by allowing a range of ±10% around each score. On this basis, Nimrod is slightly inferior on average. Taking account of the significantly superior rain rate performance shown by the gauge comparisons, we conclude that the Nimrod actuals are, on balance, rather better than the FRONTIERS actuals.

Forecasts

Methods of objective assessment

Assessment is against the corresponding actual, within the Nimrod 'area of radar coverage' and excluding pixels which have satellite derived rain in either the forecast or actual. For rain location, hit rate, false alarm rate and critical success index (CSI) are computed after both forecast and actual have been thresholded at 0.125mm/hr, and the resulting detection field smoothed with a 25km half width Gaussian to remove 'speckle' at the pixel scale. The effect of the smoother is shown in fig.1. The 25km version (fig.1c) removes the individual showers in the original (fig.1a), leaving coherent rain areas of at least 50km width. In order to exclude scores comparing two worthless forecasts, comparison of a forecast sequence is terminated once both methods have a CSI less than 0.2, this being approximately the expected value for a random forecast. This also yields an assessment of the percentage of forecasts at each lead time for which at least one of the systems produces 'useful' results. The number of forecasts at each lead time, differing by at least 0.2 in CSI is also computed.

hit rate = correct forecasts over threshold / actual occurrences over threshold

false alarm rate = spurious forecasts over threshold / forecasts over threshold

CSI = correct forecasts over threshold / (actual occurrences + spurious forecasts)

For rain rate, 15km average values of forecast and actual are compared where both exceed 0.125mm/hr. Both the number of successful comparisons and the RMSF are calculated.

$$RMSF = \exp (\mathrm{sqrt} (\ln (\text{forecast} / \text{actual}) / \text{number of comparisons}))$$

<u>Results</u>

The period assessed was 17/2/95 - 14/3/95

Percentage of useful forecasts:

Lead time	T+1	T+2	T+3	T+4	T+5	T+6
%	89	72	49	32	19	12

The number of useful forecasts drops rapidly with forecast time. A customer is unlikely to make good use of a product which is misleading on more than 50% of occasions. Both schemes fail on this measure after T+3.

Rain location:

		T+1	T+2	T+3	T+4	T+5	T+6	All
Nimrod	Hit rate	65	47	37	32	30	28	47
FRONTIERS	Hit rate	63	47	38	34	34	34	48
Nimrod	False alarm rate	32	43	48	51	52	52	42
FRONTIERS	False alarm rate	33	44	48	51	50	48	43
Nimrod	No of sig better f/c's	27	19	14	6	3	1	
FRONTIERS	No of sig better f/c's	10	13	21	25	25	21	
Number forecasts assessed		845	666	447	283	166	105	

Nimrod yields better hit rates at T+1 and worse from T+3; better false alarm rates up to T+2 and worse from T+5 Average performance across all forecast times puts FRONTIERS one point better on hit rate and Nimrod one point better on false alarm rate. Nimrod has more 'wins' at T+1,2 and FRONTIERS has more at T+3-6. Overall, Nimrod performance is better at T+1,2 and FRONTIERS performance is better at T+4-6.

Rain rate:

		T+1	T+2	T+3	T+4	T+5	T+6
Nimrod	correct hits($x10^6$)	1.7	1.0	0.6	0.4	0.2	0.1
FRONTIERS	correct hits($x10^6$)	1.3	0.9	0.6	0.4	0.2	0.1
Nimrod	RMSF	2.83	3.18	3.30	3.40	3.37	3.34
FRONTIERS	RMSF	2.74	3.14	3.34	3.42	3.39	3.45

Nimrod has higher numbers of hits and a worse RMSF at T+1,2. Otherwise the numbers of hits are the same (to the precision given) and Nimrod has slightly better RMSF scores. Unfortunately, the absence of a penalty for a miss or false alarm makes it impossible to determine which system is better since the judgement based on 'number of hits' conflicts with the judgement based on RMSF error. It is planned to introduce a modified score which deals with this problem.

Method of subjective assessment

The first stage of assessment is to compare the FRONTIERS and Nimrod forecasts for T+1,3,6 with their corresponding actuals. If there are significant differences, assessment is by comparison with the actuals, taking all three forecast times into account. The full FRONTIERS area is assessed. Four scores are possible: No significant differences; Significant differences but neither better; Significant differences and FRONTIERS better; Significant differences and Nimrod better. Comments were recorded to justify the mark.

Results

The period assessed was 17/2/95 - 14/3/95
The number of comparisons was 38

	FRONTIERS better	Nimrod better
Spurious advection of showers	6	0
Satellite rain	4	0
Others	5	3
Total as percentage	39%	8%

Discussion

The objective scores suggest that Nimrod performs better early in the forecast and FRONTIERS performs better at the end. This pattern is not contradicted by the subjective scores, which measure the whole forecast - since the forecasts are qualitatively similar at T+1 on most occasions, the subjective score is primarily based on the T+3 and T+6 comparisons. However, the percentage of forecasts with significant differences is considerably larger, suggesting a smaller threshold than the 0.2 CSI difference taken in the objective results. Both of the main reasons identified for Nimrod being inferior would largely be a problem for the latter part of the forecast.

In order to assess the relative importance of the improvements early in the forecast as against those towards the end, a value must be placed on each forecast range. It is suggested that these values are well represented by the objectively derived percentages of useful forecasts at each lead time - the lower the percentage of useful forecasts, the less valuable the forecast, even if it is right. With such a weighting, it is seen that the improvements made by Nimrod in the early part of the forecast outweigh the deficiencies later. However, for some purposes the later part of the forecast is more important despite its unreliability. For these purposes, the subjective assessment can be taken at face value, indicating that performance of the Nimrod system is not yet adequate.

Other factors

Lead time

The comparisons have been performed as if the products were available at the same time. However, current timings give an average availability for Nimrod of T+0:12 for the actual (against T+0:20 for FRONTIERS) and T+0:23 for the forecast (against T+0:33 for FRONTIERS). Given the steep reduction in quality with lead time shown above, the resulting gain in lead time is significant.

Cost

A major justification for the Nimrod development is to save much of the cost of the FRONTIERS system: ie the cost of human intervention and much of the hardware maintenance cost. This amounts to at least £150k p.a..

Upgrades

The scope for upgrading FRONTIERS is limited by the structure of the human interface. It is also very expensive to make changes. By contrast, significant improvements to Nimrod are already under test for planned implementation this autumn.

Reliability

There were two outages of Nimrod in the period 15/2 - 14/3 giving an overall product availability of 89%. Both outages were caused by failures of the CDN which also affected Radarnet and FRONTIERS. The first was of 51 hours and the second of 22 hours. FRONTIERS recovered more quickly due to call-out arrangements. Given equivalent arrangements, Nimrod would have recovered equally quickly. It is not known whether any shorter failures occurred on FRONTIERS.

Acknowledgements
The automated FRONTIERS system was developed by B.Conway, G.Ryall, N.Thomas, R.Brown, M.Kitchen, A.Davies, W.Hand, S.Higginson and implemented by D.Harrison, D.Stone, P.Newcomb. The subjective assessment was carried out by CFO forecasters on the General Aviation G2 position.

References
The Nimrod plan was originally stated in the Project Initiation Document, agreed in June 1993. The current update is version 3.1, agreed in June 1994.
The plan for development of the automation algorithms (stage 2) was agreed in September 1993 and completed in January 1994.
The plan for implementation of the automated FRONTIERS (stage 3) was agreed in September 1993.
These documents are availale from the project manager on request.
An overview of the project, periodically updated, is also available from the project manager.
Development to December 1994 was reported in Golding,B.W., 1995, "Rainfall - analysis and short range prediction", in *Hydrological Uses of Weather Radar*, Ed. K.A.Tilford. British Hydrological Society, Occasional Paper No.5.

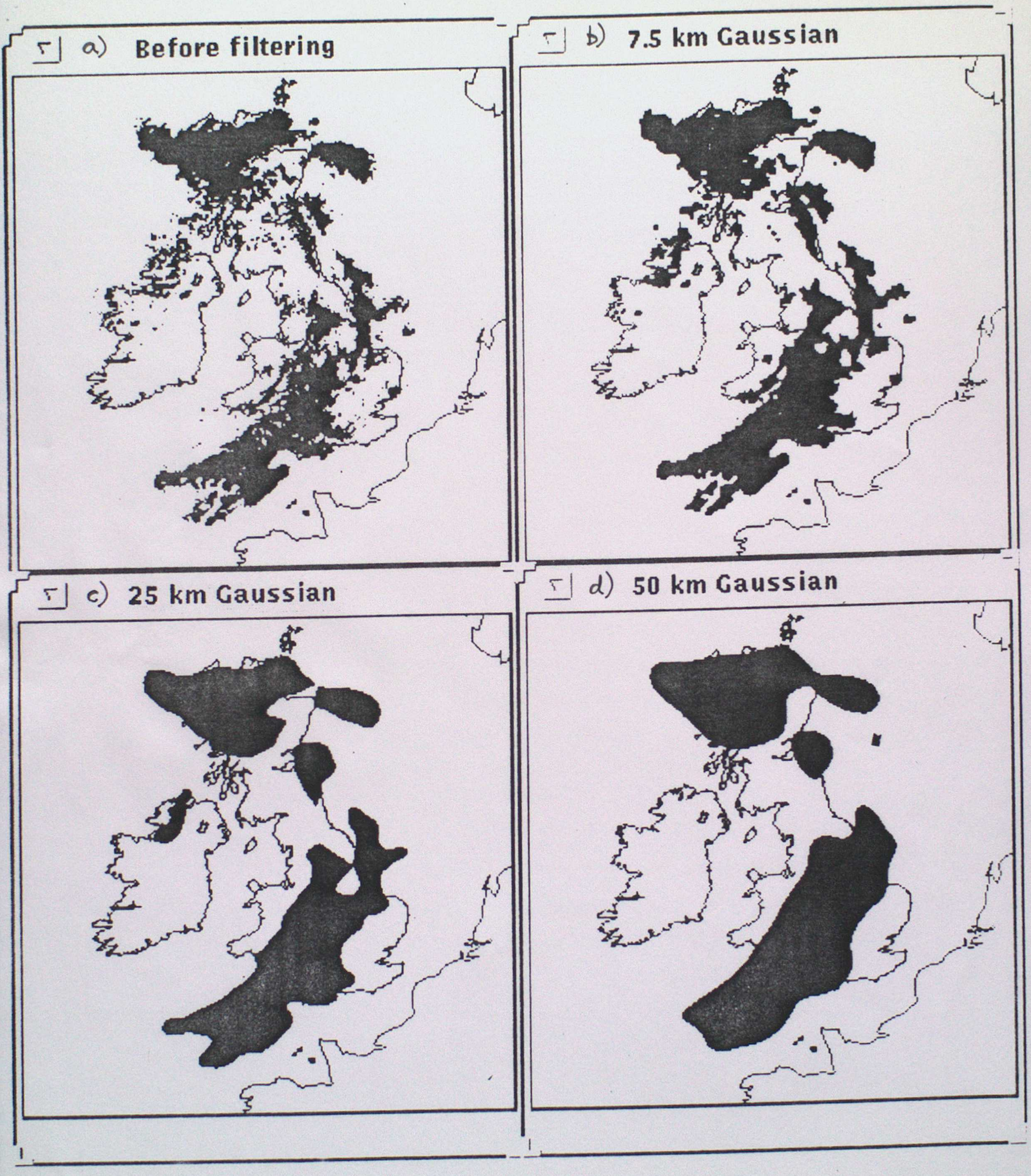
a) Before filtering
b) 7.5 km Gaussian
c) 25 km Gaussian
d) 50 km Gaussian

Figure 1