JM/2

Met O 11 Technical Note No 202

OBJECTIVE ASSESSMENTS OF THE MESOSCALE NWP TRIAL OCT.1984-JAN 1985

by

B W Golding

March 1985

Objective assessments of the Mesoscale NWP trial Oct 1984-Jan 1985

1. Introoduction

The first phase of operational trials of a mesoscale numerical weather prediction system was held from October 1st 1984 to February 1st 1985 with a three week break in December/January. The aim of the trial was to test the usefulness of the model products in the central forecast ofice (CFO) at Bracknell. A combination of subjective and objective verification techniques was employed, concentrating on site specific forecasts for several of the main UK observing stations. The subjective assessment has been discussed elsewhere. In this paper a more detailed summary of the objective results will be presented. The overall conclusions are unaltered but some additional insights into model behaviour are revealed. The statistics are not homogeneous since some forecasts failed before the scheduled twelve hours and also because of changes to the model and analysis. A number of forecasts were also lost for various reasons. The extent of available verification data is shown in Fig.1 together with a histogram of forecast completions. The latter diagram shows that most forecasts reached 6 hours and about 80% reached the full 12 hours.

2. Data collected

Two types of objective verification were performed:- (i) a spot comparison of a specified grid point with a surface station observation (ii) a comparison of grid point values against the objective mesoscale analysis. In both cases the results were grouped into regions for ease of storage. Owing to an error, the stations were not quite grouped as expected. The results are stored as shown in Fig. 2, ie for individual stations except where a group is ringed in which case a composite result is stored for that group. The grid point comparisons were computed over each of the land regions marked in Fig. 3. Note that the figures for the Republic of Ireland represent the whole of that country, not the part shown in the maps presented here.

At the stations in Fig. 2, hourly errors of screen temperature, relative humidity, 10-metre wind speed and accumulated precipitation were stored. Where stations are grouped, the mean error was stored. The hourly accumulated precipitation observations were obtained from the SREW reports. Unfortunately these have a precision of 0.2 mm and nil reports are not made so the results obtained from this comparison were not usable. Hourly station scores were also computed for a categorical forecast of the occurrence or not of specified weather features of importance to customers. The percentage of correct forecasts was accumulated. The features tested were:- any rain, moderate or heavy rain (rate greater than 0.5 mm hr^{-1}), fog (visibility under 1 km), thick fog (visibility under 200 m), low cloud (base below 1000 ft), or any cloud. Unfortunately, although useful to indicate trends in forecasting skill, they are strongly affected by the frequency of occurrence of the feature and so are difficult to interpret on their own. This is particularly noticeable with the fog scores. Overall, thick fog was correctly forecast on 99% of occasions. However, this score could also be obtained by forecasting that thick fog will never occur.

Some specific problems with the station comparison have also come to light. The method of obtaining temperature and relative humidity from the model may be criticised separately from the model performance since humidity is merely the 10 m value and temperature is the average of the surface and 10 m values. Errors in these variables may also be attributed to differences between the elevations of the model grid point and the station. This aspect is also important when considering the wind errors. Many of the stations are coastal and experience markedly different wind climates from locations a few kilometres inland with which the model might be more fairly compared. In addition, about halfway through the trial, certain areas were classified as urban or forested in the model and given increased surface roughness. Where this is representative of the surroundings of a station it has improved the comparison but where it is not, substantial errors have appeared. A specific problem in the cloud scores is caused by interpretation of the observations. Generally the base of the lowest layer of 5 oktas or more is used. However, where none exists, the lowest layer of any cover is given. Thus on some occasions a forecast of no cloud will have been classed as wrong when only 1 okta was observed although model cloud is generally taken to represent 5 oktas or more.

The regional comparison with analyses was performed three-hourly for both the mesoscale forecasts and the interpolated fine mesh forecasts. Since the first guess for each analysis was obtained from the fine mesh interpolation there is a possibility of bias here. However, most of the country is influenced by observations and only the high ground of Wales and N Scotland needs to be treated with caution. Mean errors of screen temperature and wind components together with critical success indices (CSI) of precipitation rate, thick fog, low cloud and frost were stored. The critical success index is defined as

$$CSI = \frac{\text{area over which event successfully forecast}}{\text{total area} - \text{area over which event successfully not forecast}}$$

Note that, unlike the scores recorded in the station comparisons, this is a one- sided test and is not defined if the event does not occur anywhere and is not forecast anywhere (a perfect forecast). It also tends to favour over-forecasting so long as the area over which the event is forecast remains small compared with the total area. Thus predictions of the occurrence of rain scored about 15% with the fine mesh doing better because of its lower resolution while predictions of the occurrence of dry conditions scored about 70% with the mesoscale model doing better because of its higher resolution. It proved impossible to draw useful comparisons from these scores.

Where possible corrupt reports, repeat runs etc have been removed from the data. However it has not been possible to check every suspect error.

3. Results

Mean errors of temperature, humidity and wind speed for all cases and stations are shown in Fig.4 as a function of time of day. Since every forecast started at 0600 GMT this also indicates length of forecast. The temperature shows a clear pattern, which is repeated in individual regions and periods, of a substantial initial bias which grows rapidly in the morning and then returns to a smaller value in the early afternoon. The gross bias indicated here is about one degree cold. The initial bias is related to the time of initialisation which was before sunrise when the surface inversion is strongest, and to difficulties in deriving an acceptable first guess temperature from the interpolated fine mesh data.

The growth of error in the morning will be partly related to the residual effect of the fine mesh interpolation at heights above the influence of the surface analysis. However, there is probably an inherent difficulty in forecasting the temperature during this period because of the rapid changes associated with the growth of the boundary layer. This is supported by the same pattern appearing in the wind speed errors. By contrast, the humidity errors increases steadily until the time of maximum temperature. This can only partly be explained by coldness and must indicate a more general problem of an over-moist boundary layer, as reported in some larger scale models.

Fig. 5 shows the mean and RMS errors for the same elements at Manchester and Marham. In general they show similar trends to those identified in Fig 4. The temperature errors at Marham are slightly worse than at Manchester reflecting a problem which occurred in southerly winds until mid-December. This was caused by a difficulty in the specification of boundary conditions from the fine mesh in the southeast corner of the grid over France and also by the assignment of fine mesh land temperatures to the sea in the Straits of Dover. Both of these combined to give some very cold forecasts over the whole of SE England until remedial action was taken in mid-December. The difference in wind errors is due to the change of surface roughness introduced at the end of November. From a uniform 10 cm roughness length, urban and forested areas were increased to 2 m while arable land was reduced to 5 cm. At Marham the roughness was reduced and a tendency to overpredict was slightly increased. However at Manchester, although the airport is well away from the high-rise urban centre, the grid point used for verification was classified as urban. The result was a substantial negative wind speed bias which, over the whole trial, has cancelled out the previous positive bias.

Fig. 6 shows three-hourly temperature errors for both the fine mesh and mesoscale forecasts from the comparison with analyses. The fine mesh forecasts have a data time of midnight and are therefore six hours older than the mesoscale ones. The regions shown are those containing the two stations shown in Fig. 5. Comparison of the mesoscale results with that figure shows an encouraging consistency between the two verification methods. Some differences in the bias may be attributable to problems of representativeness of the individual stations. The difference between the models is marked. The mesoscale forecasts have little variation in quality through the day while the fine mesh forecasts start very cold (bias greater than 2^{o} at 0900 GMT) and warm up rapidly to give very small errors by late afternoon. It has been established that an error in the fine mesh interactive radiation scheme results in too much cooling overnight and this is the cause of the morning bias.

Figs 7-9 show the 1500 GMT temperature errors in all regions. Fig. 7 gives errors of the mesoscale forecast compared with observations. The variation is dominated by the area of high errors in SE England. As noted before, this resulted from deficiencies in the initialisation of sea temperature and specification of the boundary conditions which gave very cold temperatures in southerly airstreams. Away from this area, mean errors are mostly under 1^{o} and RMS under $1^{1}/_{2}{}^{o}$. The best area is NE England. Figs 8, 9 present equivalent scores from the comparison of both models with analyses. The mean errors often reflect characteristics of individual stations eg Binbrook on the Lincolnshire Wolds in 0.4^{o}C colder than forecast whereas the region as a whole is 0.4°C warmer. This clearly reflects the altitude and exposure of Binbrook which is not adequately represented in the model. In Fig. 8 two areas of poor forecasts stand out - SE England and the Scottish Highlands. Elsewhere mean errors are under 1° and RMS errors under $1^1/_2$°. The fine mesh model (Fig. 9) does well in SE England but badly over Scotland. In remaining areas its mean errors are good but the RMS erros are slightly worse than those from the mesoscale model.

Figs 10--12 show the variation of 1500 GMT temperature errors throughout the trial for two regions. In Fig 10 the mesoscale forecasts are compared with observations at Manchester, Marham, Gatwick, Binbrook, and Edinburgh. The first month shows a general cold bias with several very cold forecasts at each station. This improved after 29th October when the analysis of temperature was improved but this also coincided with a change of weather type to cyclonic with cloudy conditions and hence small temperature changes. Late November and December were characterised by mobile westerly flows. The extreme errors at Marham and Gatwick on 30th November and 3rd December occurred on days of strong southerly flow and were caused by the boundary problem noted earlier. During the cold period of easterly flow in January, the model had a small bias and most forecasts were within 2°. Figs 11, 12 compare the performance of the two models using the verification against analysis. Naturally, this reduces the extremes but the pattern is very similar. In NW England (Fig. 11) the poor performance of the mesoscale model in the first month is emphasised by good results from the fine mesh model. However, early November shows both an improvement in the mesoscale results (as noted above) and a deterioration in the fine mesh. The latter may be related to the removal of the interactive radiation scheme in this period. After some cold forecasts at the end of November in both models, the mesoscale forecasts for December were mostly good while the fine mesh results remained too cold. The radiation scheme had been restored by this time as noted on the figure. Finally, the January period shows fairly good mesoscale results and very poor fine mesh ones, the majority of the latter being over 2^{o} in error. Over E Anglia (Fig.12), the fine mesh model had small errors up to mid-November while the mesoscale model was noticeably too cold throughout this period. The extreme errors on 30th November and 3rd December have already been noted but the mesoscale forecasts remain generally too cold in this period. However there is also a deterioration in the fine mesh results with a generally cold bias and some large errors in December. In January both models have small biases but the mesoscale errors are generally rather smaller than the fine mesh. To further illustrate the comparison for this January period, Figs 13, 14 show mean and RMS temperature errors for 1500 GMT for this period in the two models. The mesoscale model forecasts are much superior to the fine mesh ones except in Kent/Sussex. Over a substantial part of the British Isles, mesoscale RMS errors are under 1^{o} in this period. Only further trials can show whether this improvement in the mesoscale results is due to the changes introduced in December or to the prevailing weather type.

The mean and RMS humidity errors at 1500 GMT are shown in Fig.15 for each station. Reference to Fig 5 shows this to be the time of maximum error. There is much variation between stations, especially in the mean errors. Both mean and RMS errors are generally smallest in NE England and largest in SE England, echoing the temperature errors in Fig.7. This is not unexpected if the absolute humidity error is about the same everywhere because of the relationship between temperature and the saturation moisture content. Further investigation of the humidity errors themselves will have to await removal of the temperature induced component.

The variation of 1500 GMT wind speed error between stations is shown in Fig.16. Mean errors are very small with only London, Bracknell, Eskdalemuir and Oban exceeding 3 knots. These all have exposure problems. London was well predicted until its surface roughness was increased to a realistic urban value. In contrast, Bracknell, which also had its roughness increased as a forested area, had its bias much reduced by the change. Most stations are slightly overforecast, the exceptions being exposed observing sites such as Binbrook on the Lincoln Wolds and St. Mawgan on the Cornish Coast. The RMS errors reflect the size of the mean errors and are under 5 kts except for the stations mentioned above with large mean errors. Fig.17 shows a comparison of RMS vector errors for the two models from the comparison with analyses. The mesoscale model has errors of 5-6 kts in both regions while the fine mesh model has errors of 7 kts in NW England and 5 kts in E Anglia. This suggests that the differences in orography between the two models may be influencing these results. A component of the fine mesh error in NW England is a directional bias of 2 kts from the southwest. No such bias is detectable in E Anglia or in the mesoscale model results. The daily sequence of mesoscale wind speed errors for 1500 GMT is shown in Fig.18 for Manchester, Marham, Gatwick, Binbrook, Edinburgh and London Weather Centre (LWC). It is less easy to interpret than the corresponding temperature sequence except for the dramatic underforecasting at Manchester and LWC after the change in roughness. There is some indication that the reduction of roughness at

Marham at the same time increased the positive bias there. The extreme error at Manchester on 27th November was due to the general forecast development which failed to predict the observed very strong gradient.

As noted in §1, the categorical forecast scores are biased by the frequency of occurrence of the phenomenon. Nevertheless, a brief discussion of results is included for completeness. Table 1 shows overall scores for 4 of the tests in the 6-12 hour period. There was no consistent trend in accuracy with forecast length except for the cloud forecasts for which the accuracy over all stations falls from 89% at 1 hour to 79% at 4 hours for low cloud and from 69% at 1 hour to 64% at 4 hours for all cloud. The apparent success of prediction of moderate rain is due to its infrequent occurrence and this has also influenced the other scores. The regional variation of the two rainfall scores is shown in Fig. 19 for 1500 GMT. All stations correctly forecast rain/no rain on 75% of occasions and in much of England 85% success was exceeded. Eskdalemuir and Oban have particularly poor scores for both tests while SW England and Marham have particularly good scores. Visibility scores are not shown since most are very close to 1. The only significant deviations from this are in northern and eastern parts of England where scores as low as 95% for visibility under 1 km and 97% for visibility under 200 m were produced. Fig.20 shows the accuracy of cloud predictions for 15Z at each station. It should be recalled that the inclusion of observations of less than 5 oktas will have lowered the scores especially for "all cloud". Concentrating on the scores for low cloud, Prestwick is seen to have the best score (93%) while Eskdalemuir has only 56% correct. Since the latter station has a good score for all cloud, this indicates a persistent error in predicting the cloud base, probably related to the local orography.

Conclusions

A number of general results have appeared in the discussion of the results which will be summarised here. The mesoscale forecasts had a bias of about one degree. Steps were taken to improve this during the trial and results for the January period suggest that these were largely successful. The bias was not as bad as in the fine mesh model in the morning but, in the early stages of the trial, the fine mesh scored rather better in the afternoon. However, in the January period, the good results of the mesoscale model contrasted with some very poor results for all forecast times in the fine mesh model. Mesoscale humidities were too high due in part to the cold temperatures. Wind speeds have a negligible bias but RMS errors increase to about 4 kts at 12 hours. RMS vector errors of about 5 kts were recorded for two regions and some indication of superiority over the fine mesh model was evident. The categorical forecast scores and CSI scores yielded little information.

In future phases of the trial, raw data should be stored from which to calculate the errors so that contingency tables can be drawn up. In order to avoid the uncertainties of using analyses, comparisons should be made of analyses, mesoscale forecasts and fine mesh forecasts with all observations.

TABLE 1 Percentage of correct forecasts, in the 6-12 hr range, of the occurrence or not of specific weather features at the time of observation.

Feature	All Stations	Marham (Norfolk)	Manchester Airport	
Rain	84	87	83	Ignoring Showers
Rain > $^1/_2$mm/hr	95	97	95	Ignoring Showers
Cloud Base Below 1000'	81	78	83	Ignoring Cb
Cloud	64	60	72	Ignoring Cb

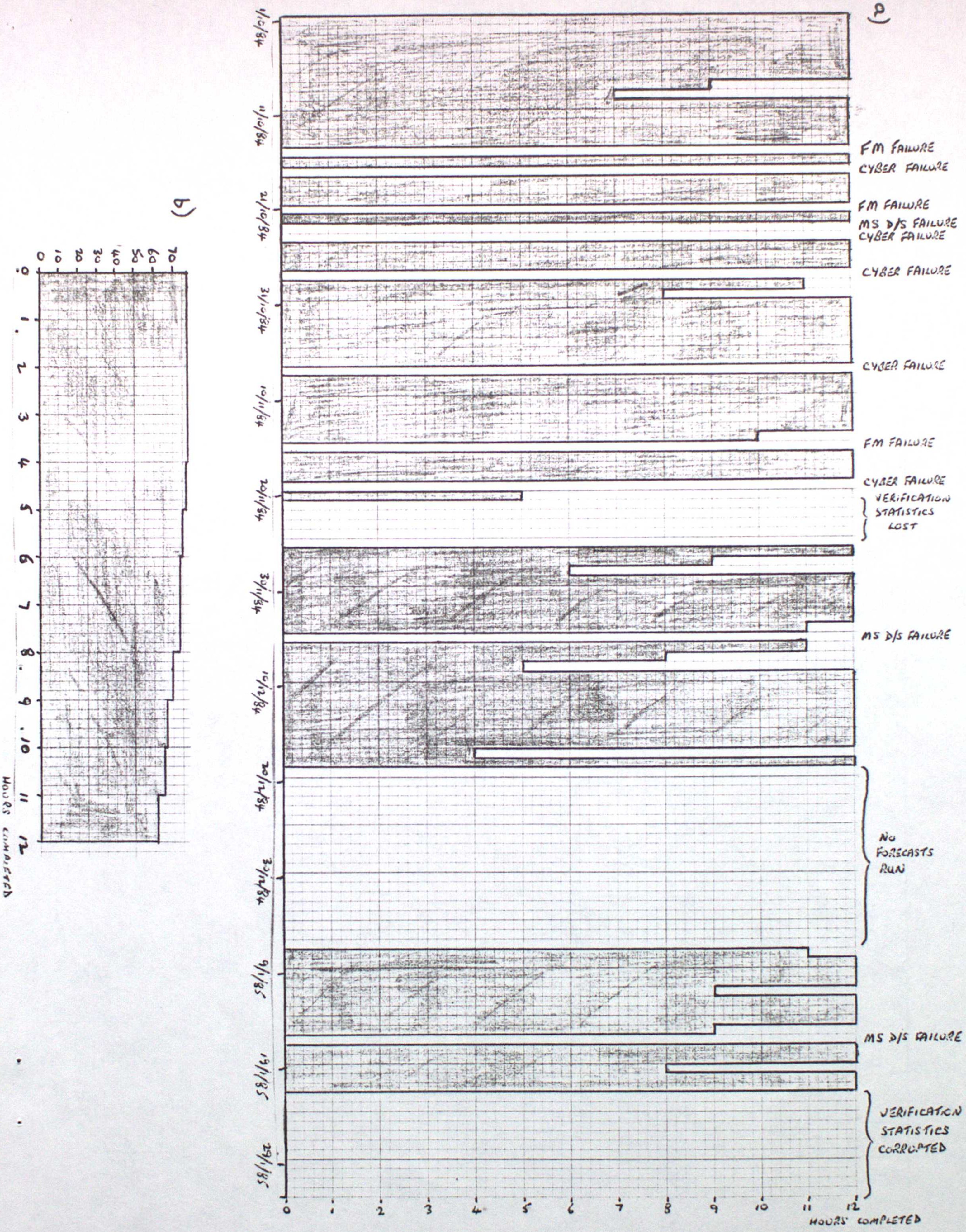


Fig.1 a) Calendar of forecast completions during the trial. All partial forecasts failed due to a model blow-up. Missing runs are annotated

b) Histogram of the frequency of forecasts of given length. All forecasts reached 4 hours.

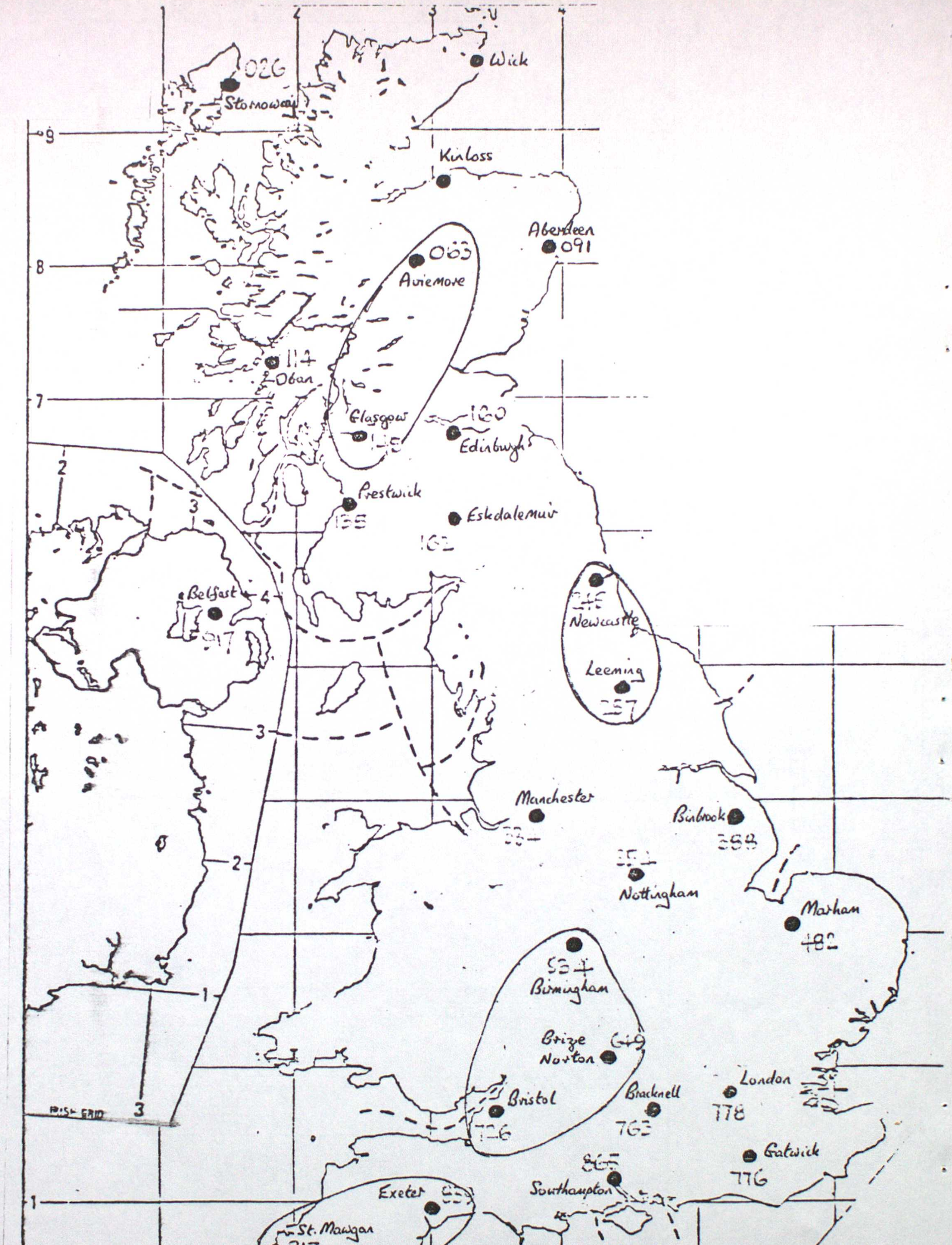


Fig.2 Stations used in the verification. Single scores were stored for the groups for the groups of stations enclosed by rings.

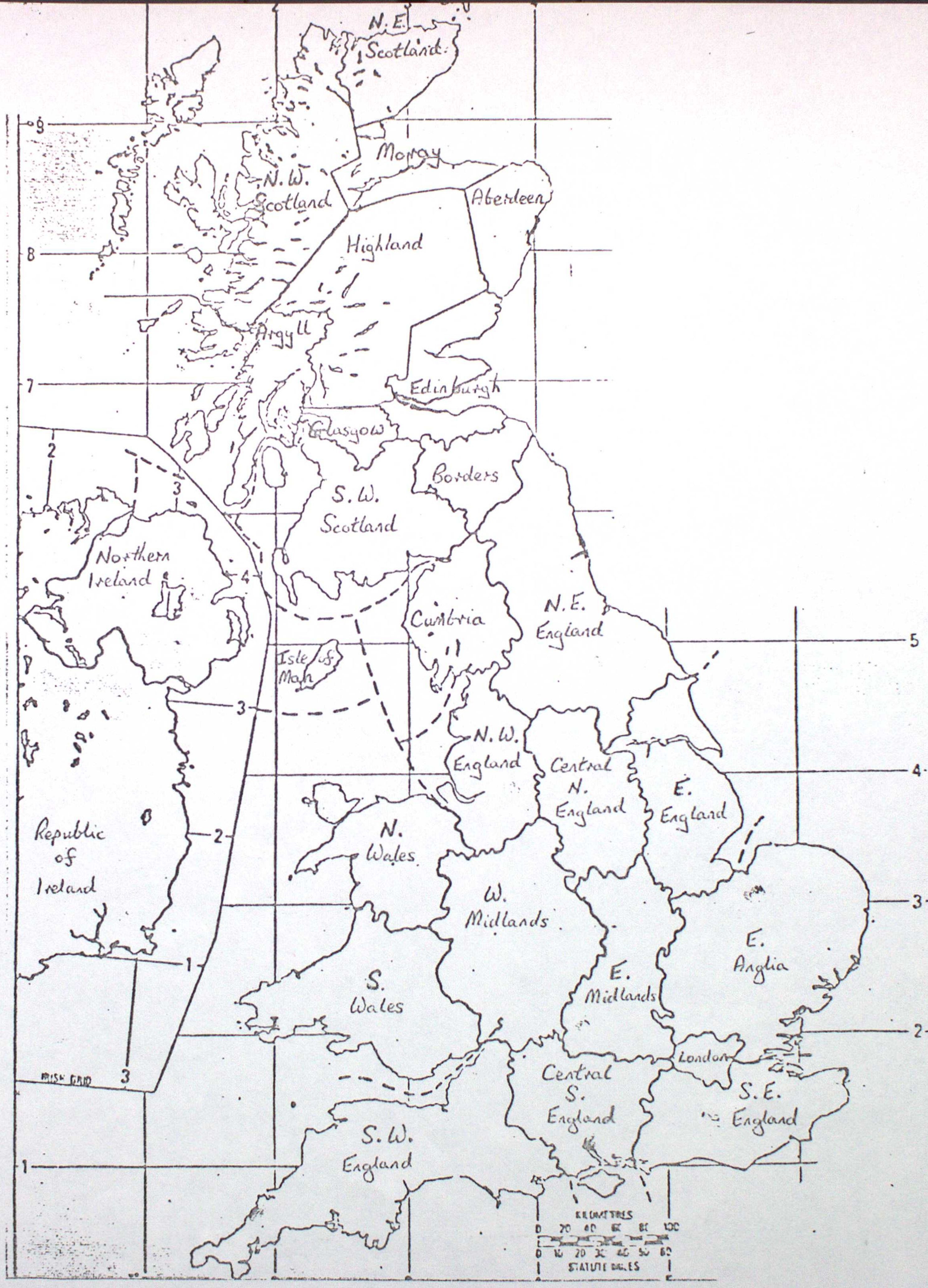


Fig.3 Regional division of the British Isles used in the comparison with analyses.

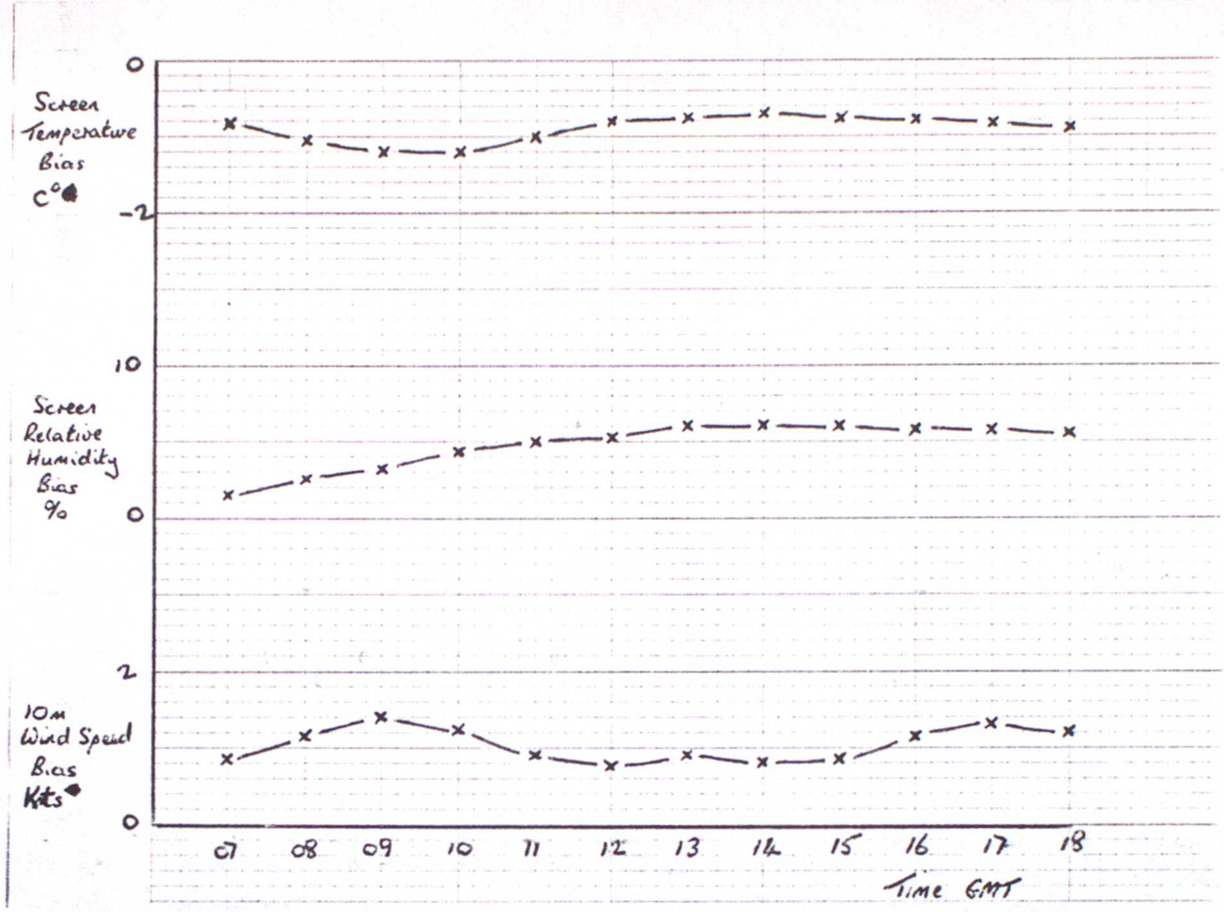


Fig.4 Mean errors of screen temperature, humidity and 10m wind as a function of time of day, for all stations through the whole trial. Mesoscale forecasts compared with observations.

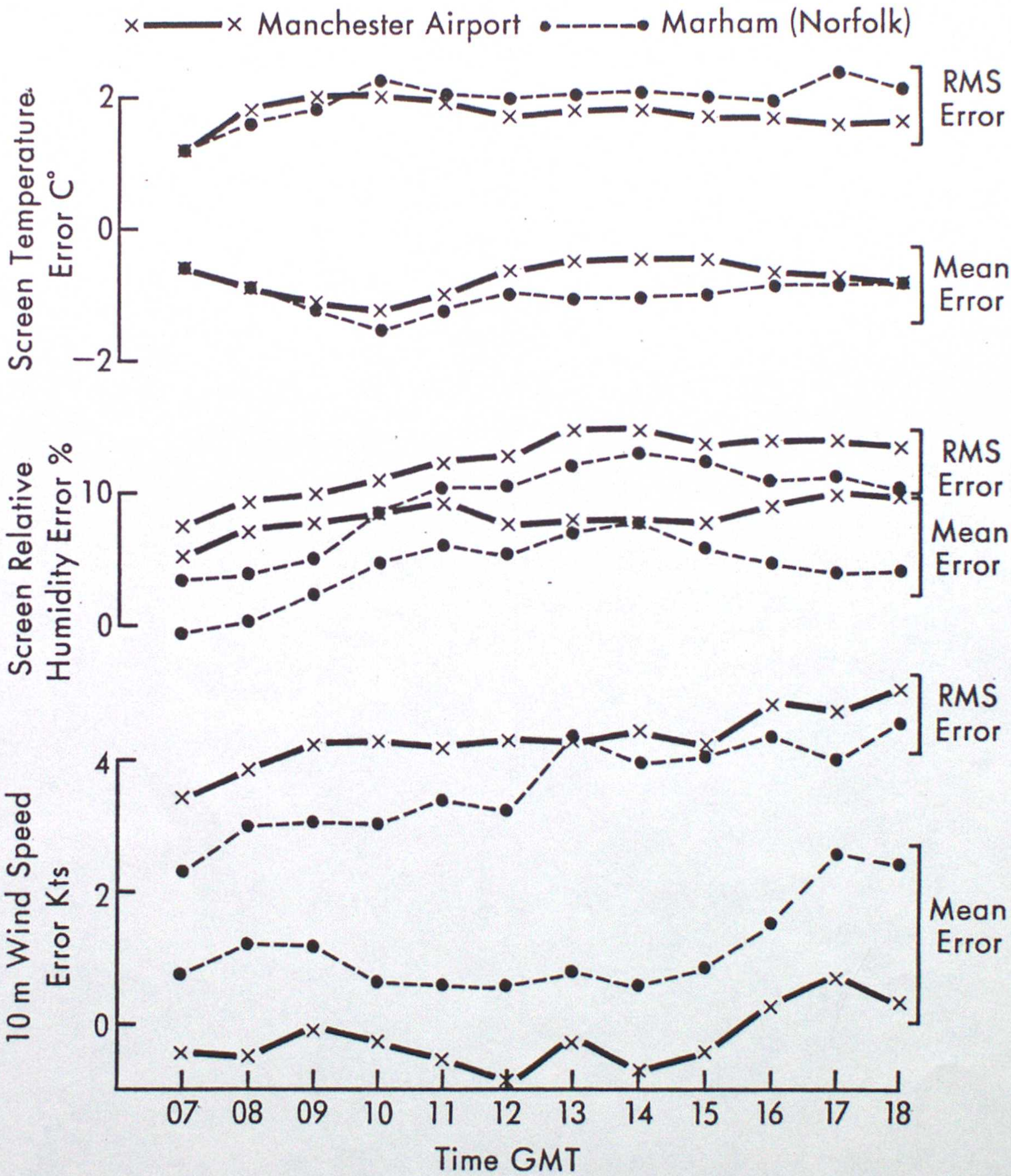


Fig.5 Mean and RMS errors of screen temperature, humidity and 10m wind as a function of time of day through the whole trial. Mesoscale forecasts compared with observations.

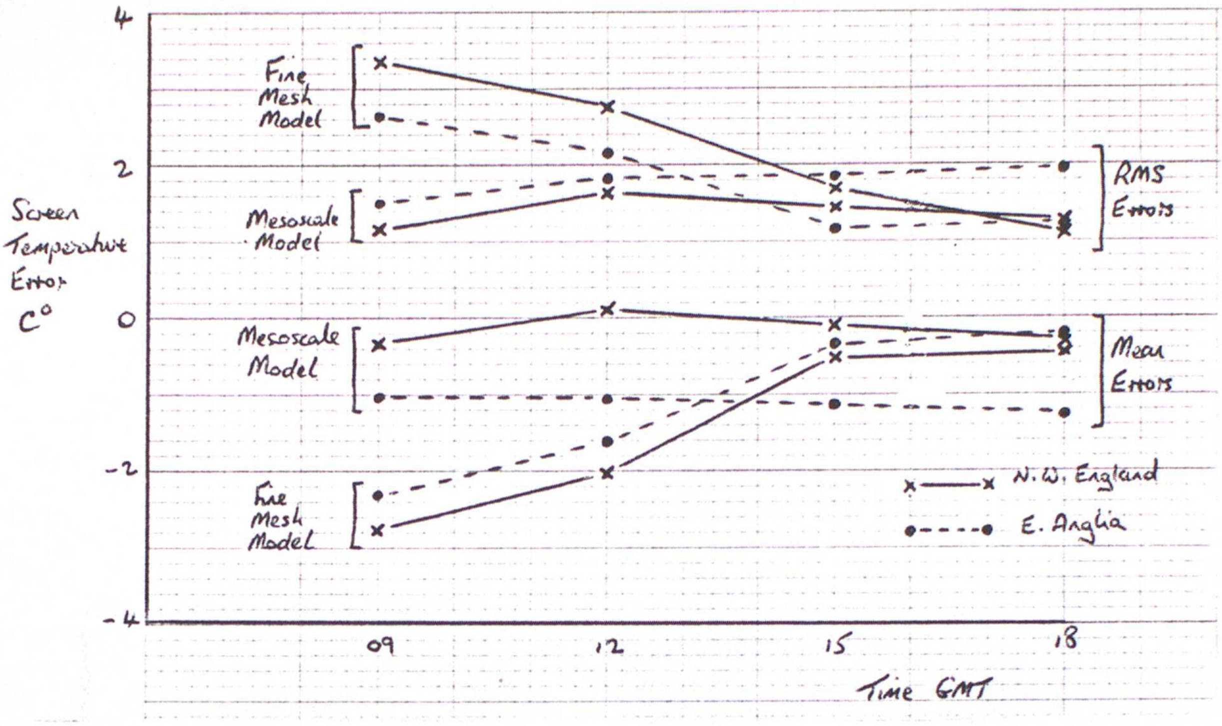


Fig.6 Mean and RMS errors of screen temperature as a function of time of day for the fine mesh and mesoscale models compared with analyses through the whole trial.

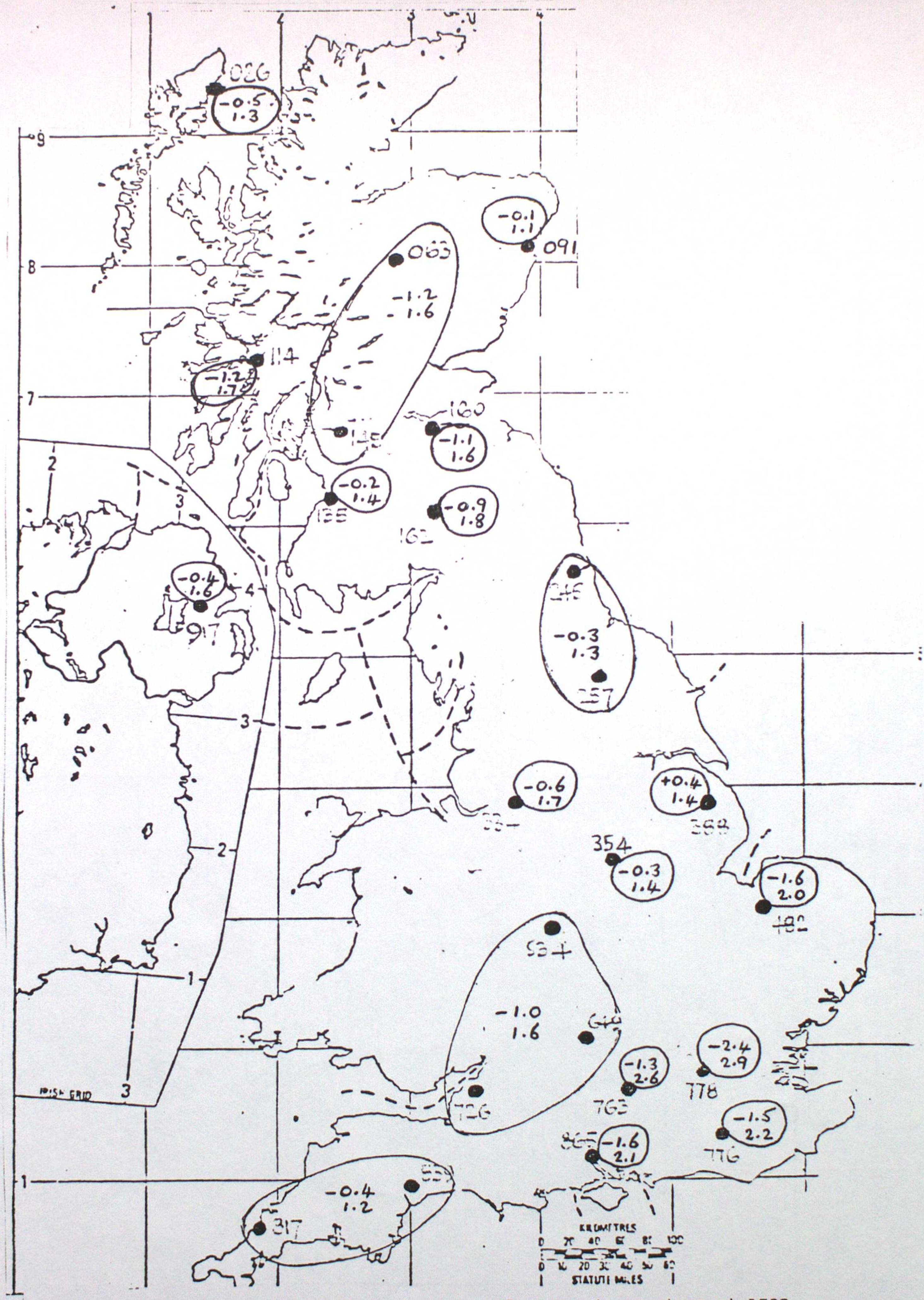


Fig.7 Mean (above) and RMS (below) errors of screen temperature at 1500 GMT for each station through the whole trial. Mesoscale forecasts compared with observations.

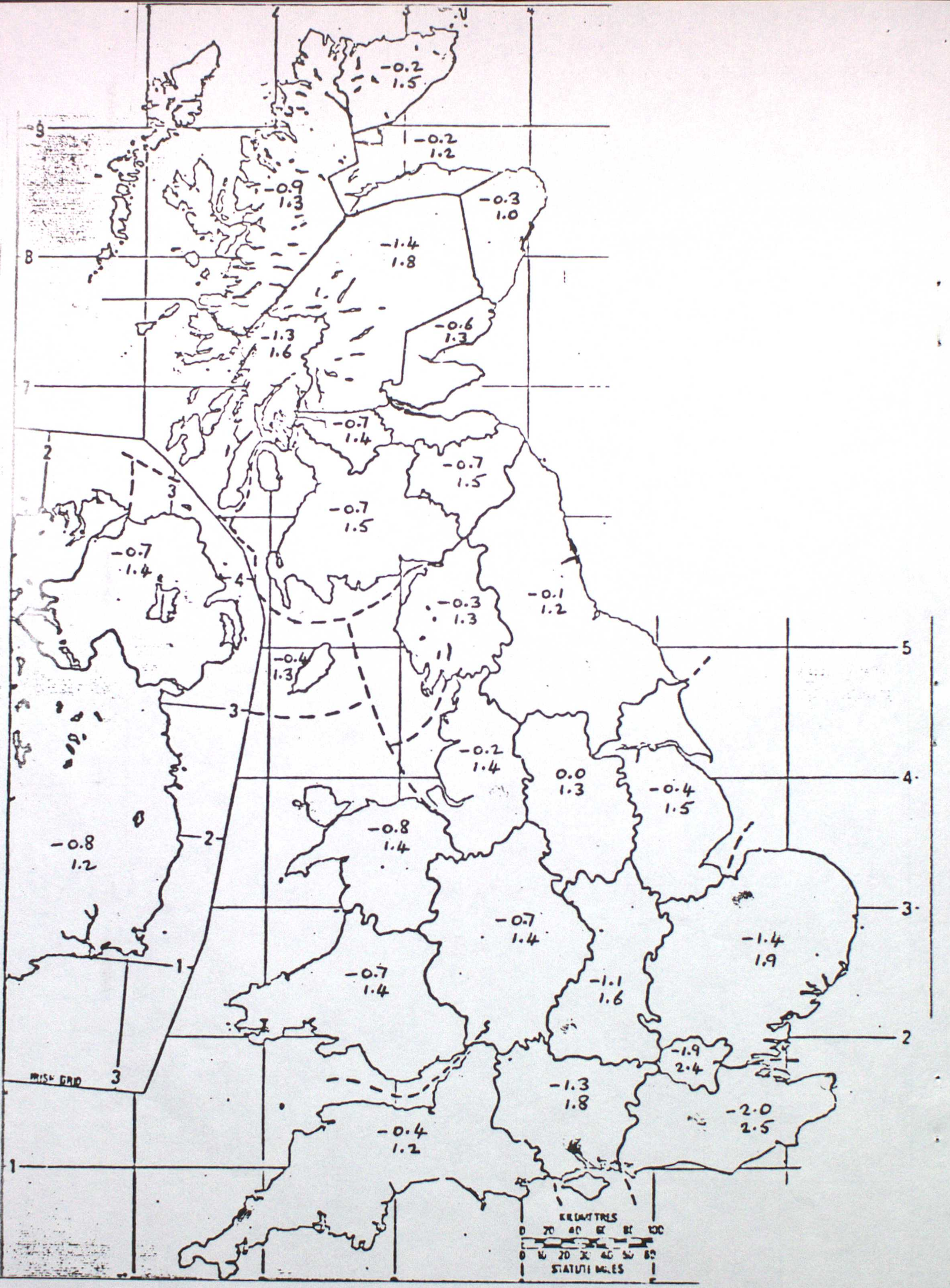


Fig.8 Mean (above) and RMS (below) errors of screen temperature at 1500 GMT through the whole trial. Mesoscale forecasts compared with analyses.

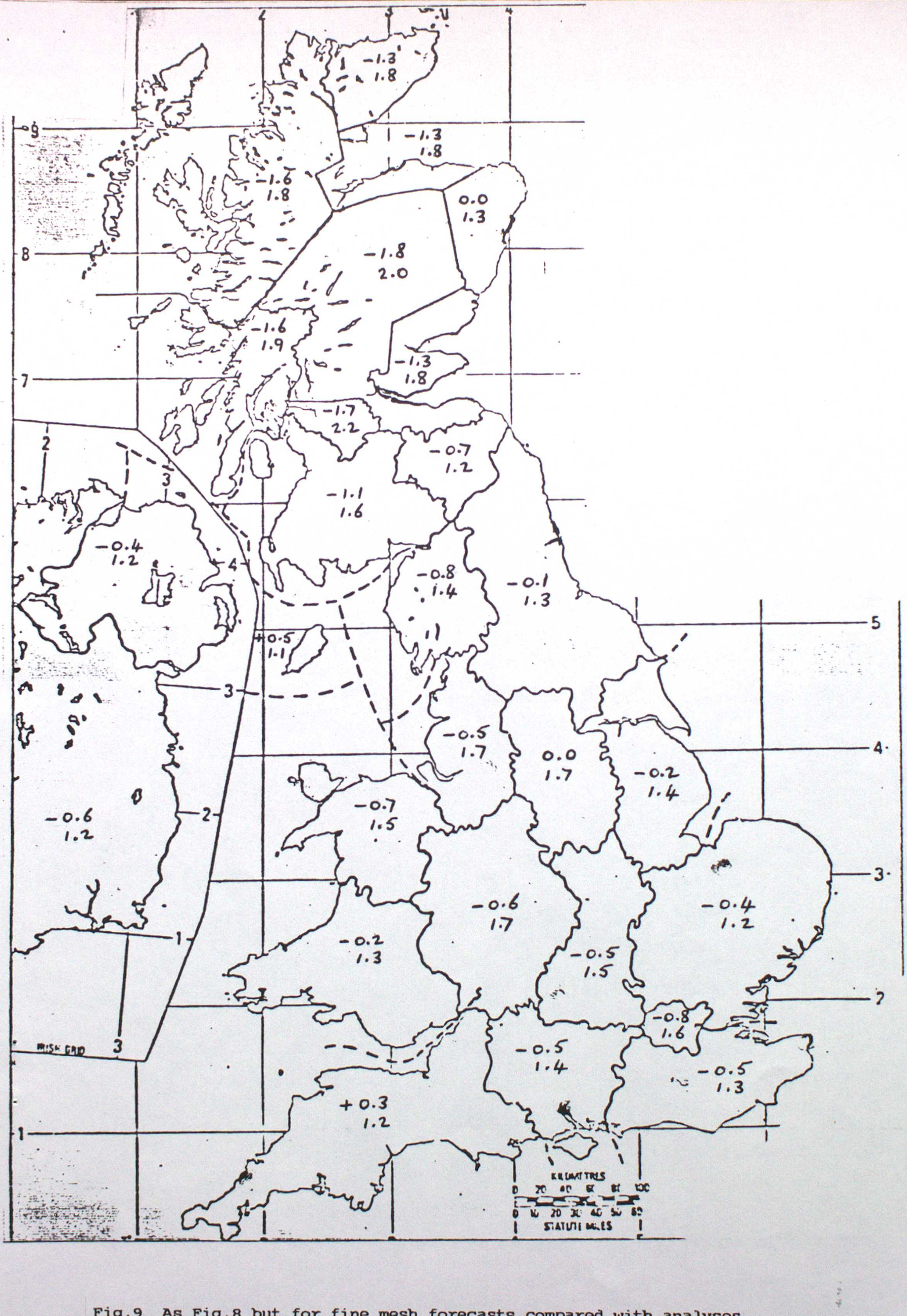


Fig.9 As Fig.8 but for fine mesh forecasts compared with analyses.

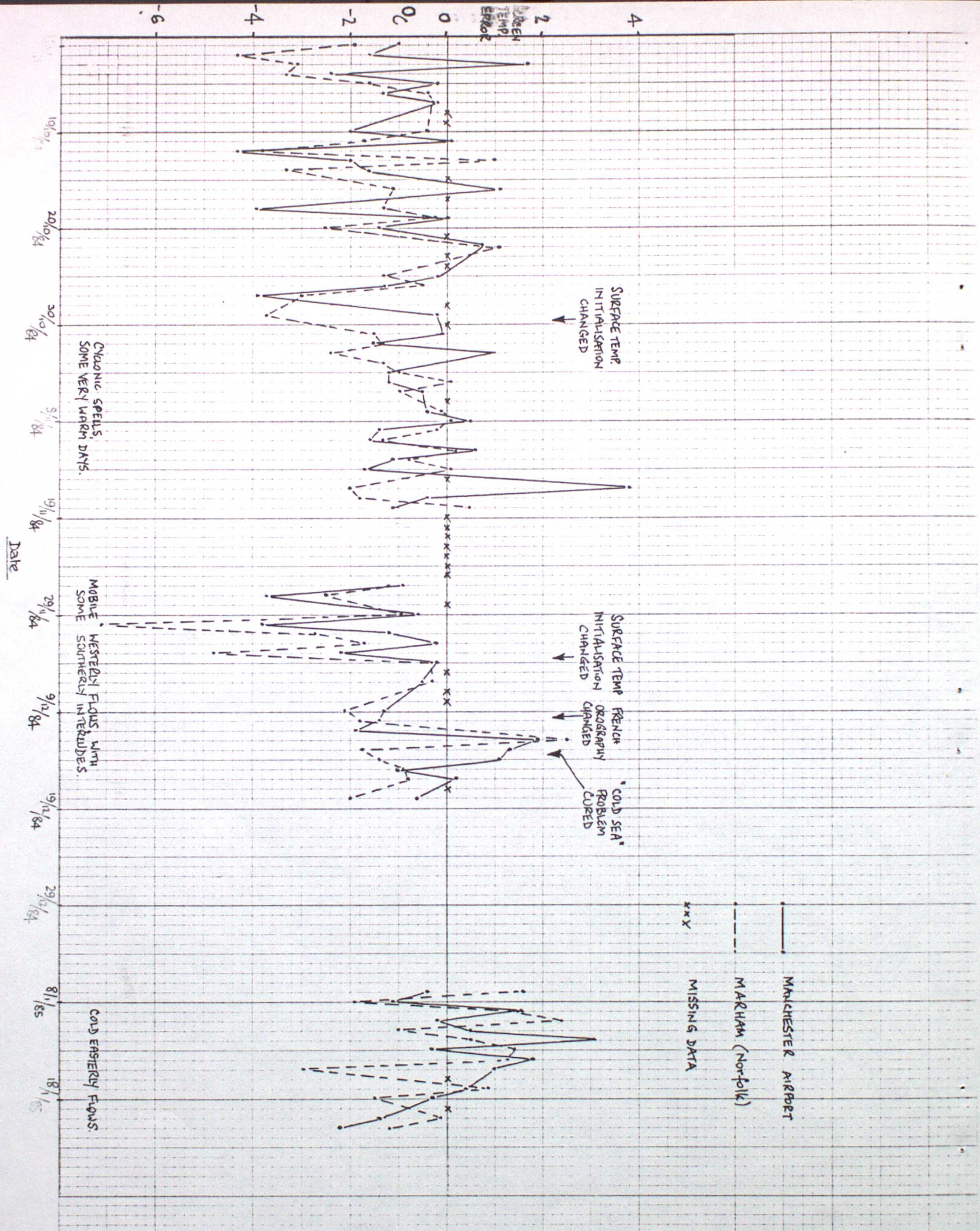
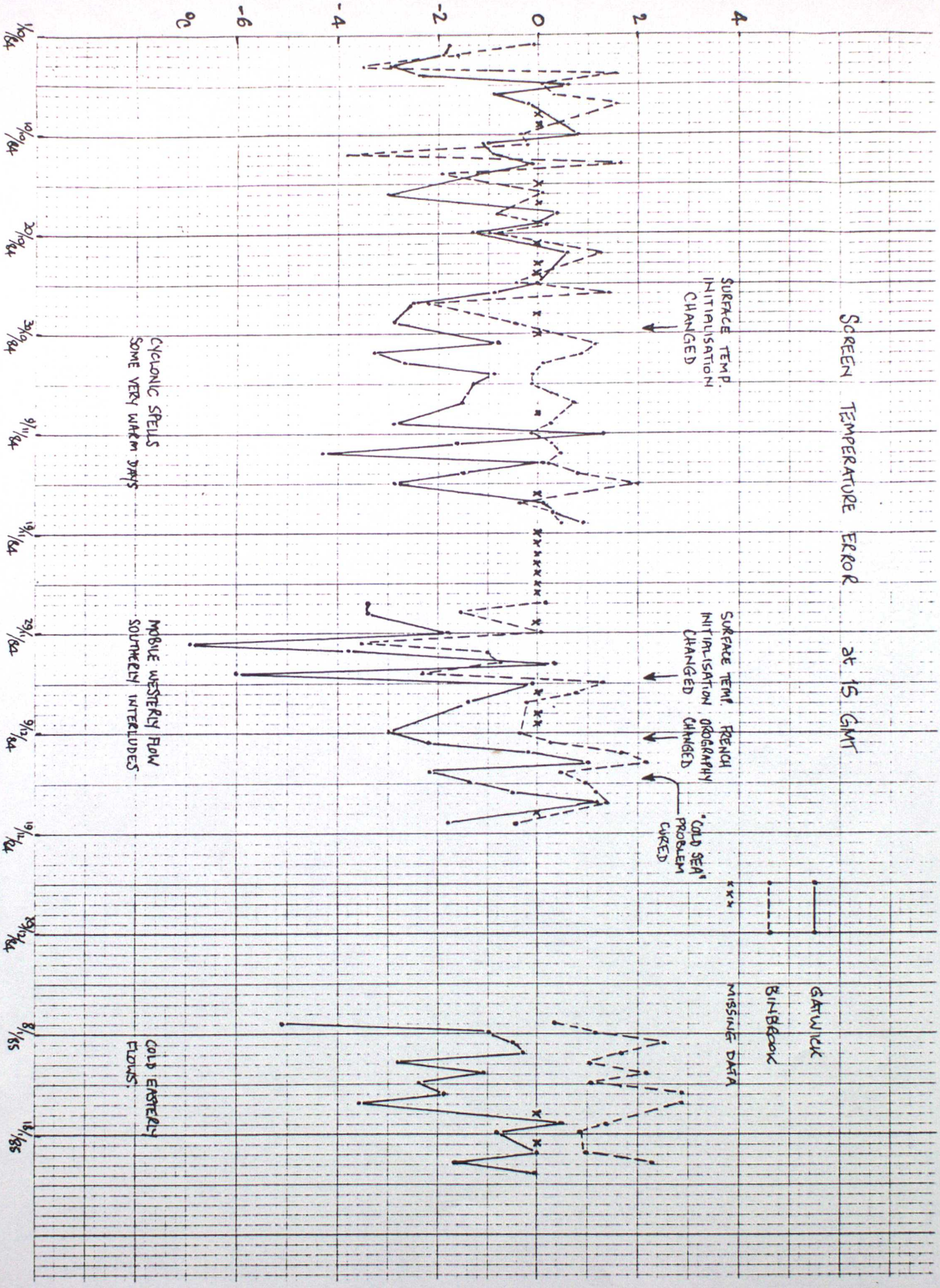
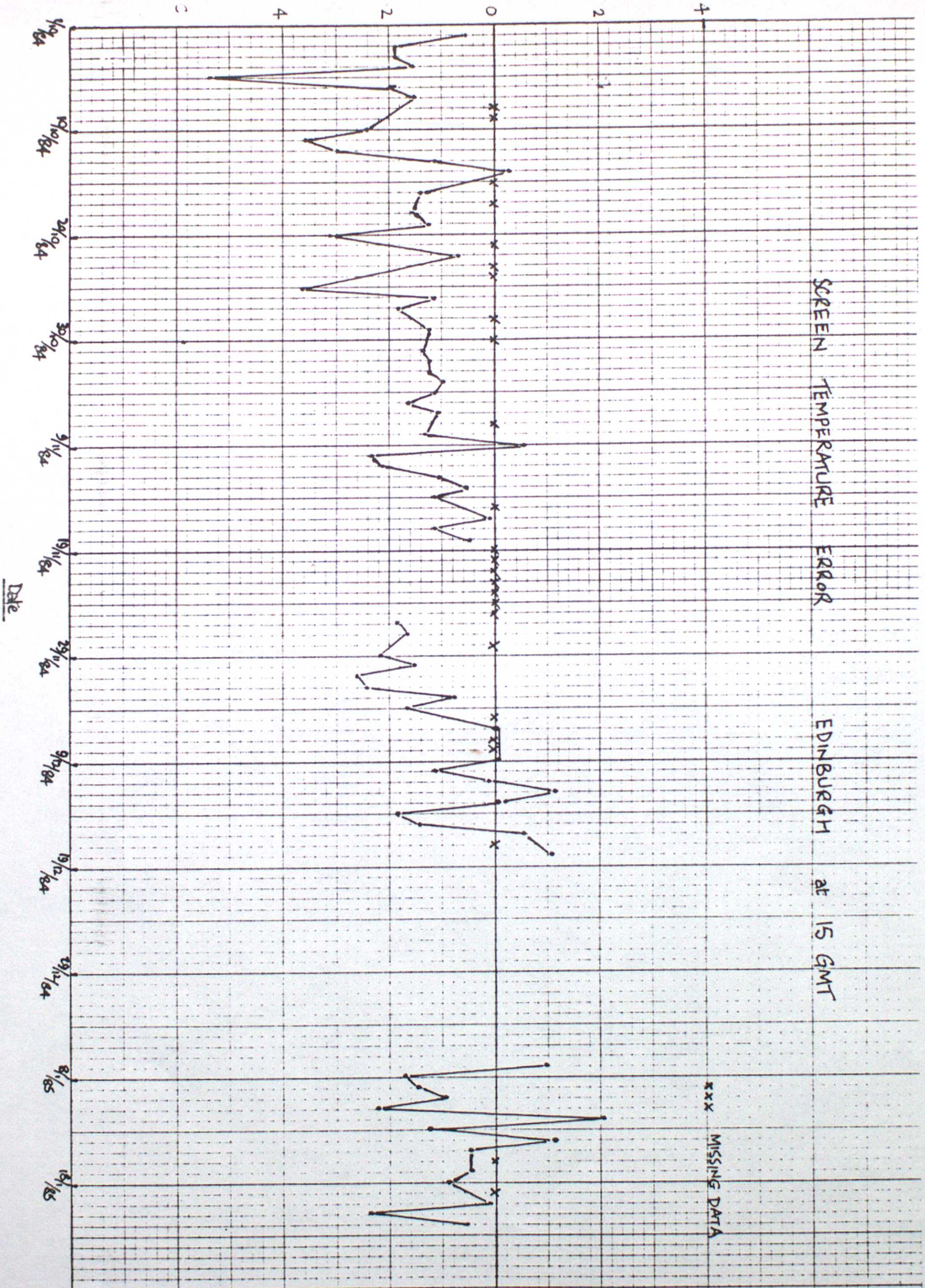


Fig.10 Daily series of screen temperature errors at 1500 GMT. Mesoscale forecasts compared with observations.

a) Manchester and Marham;



b) Gatwick and Binbrook;


SCREEN TEMPERATURE ERROR
EDINBURGH at 15 GMT
xxx MISSING DATA
Date
4
2
0
2
4
1/10/84
10/10/84
20/10/84
30/10/84
9/11/84
19/11/84
29/11/84
9/12/84
19/12/84
29/12/84
8/1/85
18/1/85

c) Edinburgh.

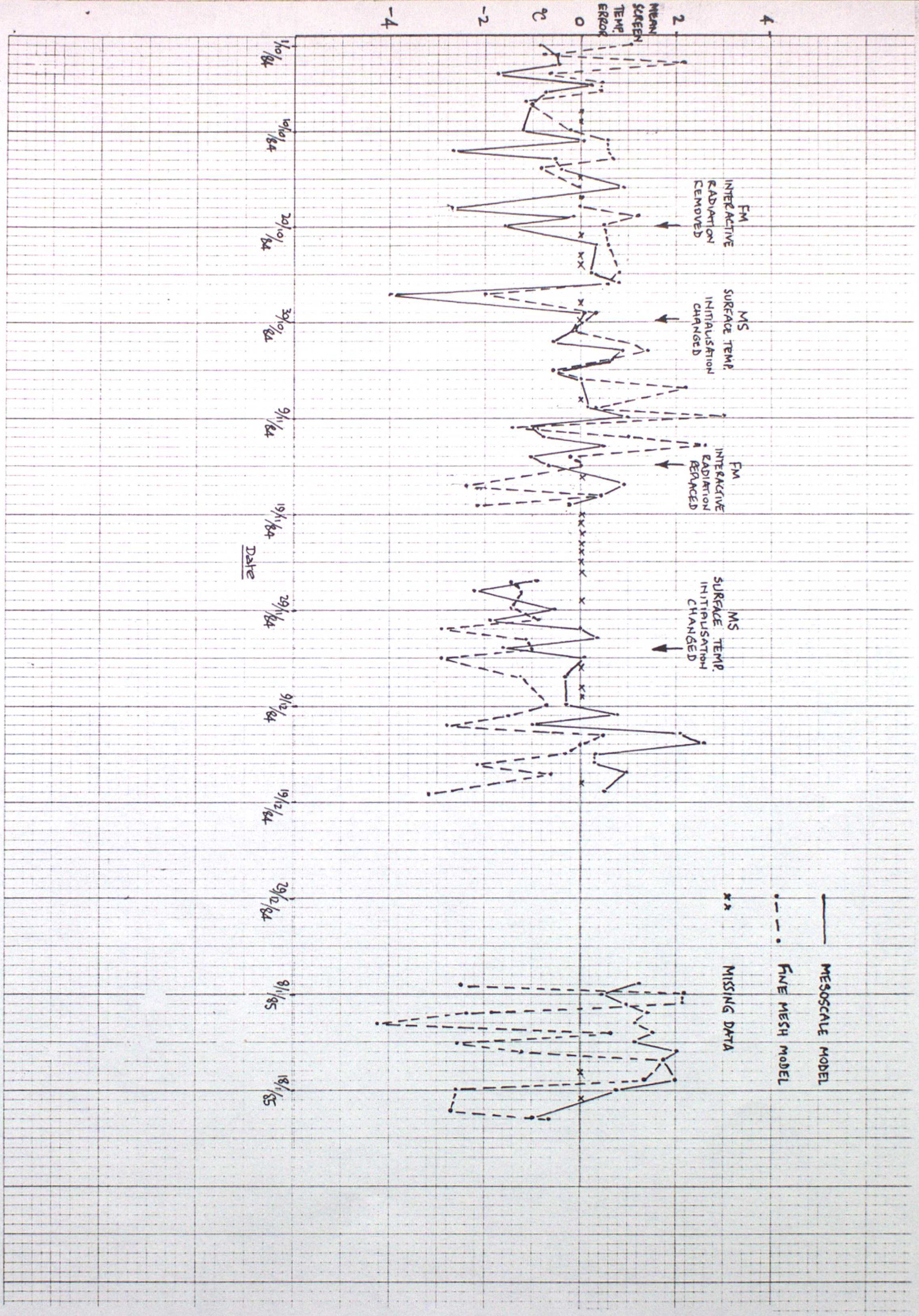


Fig.11 Daily series of screen temperature errors at 1500 GMT for NW England compared with analyses.

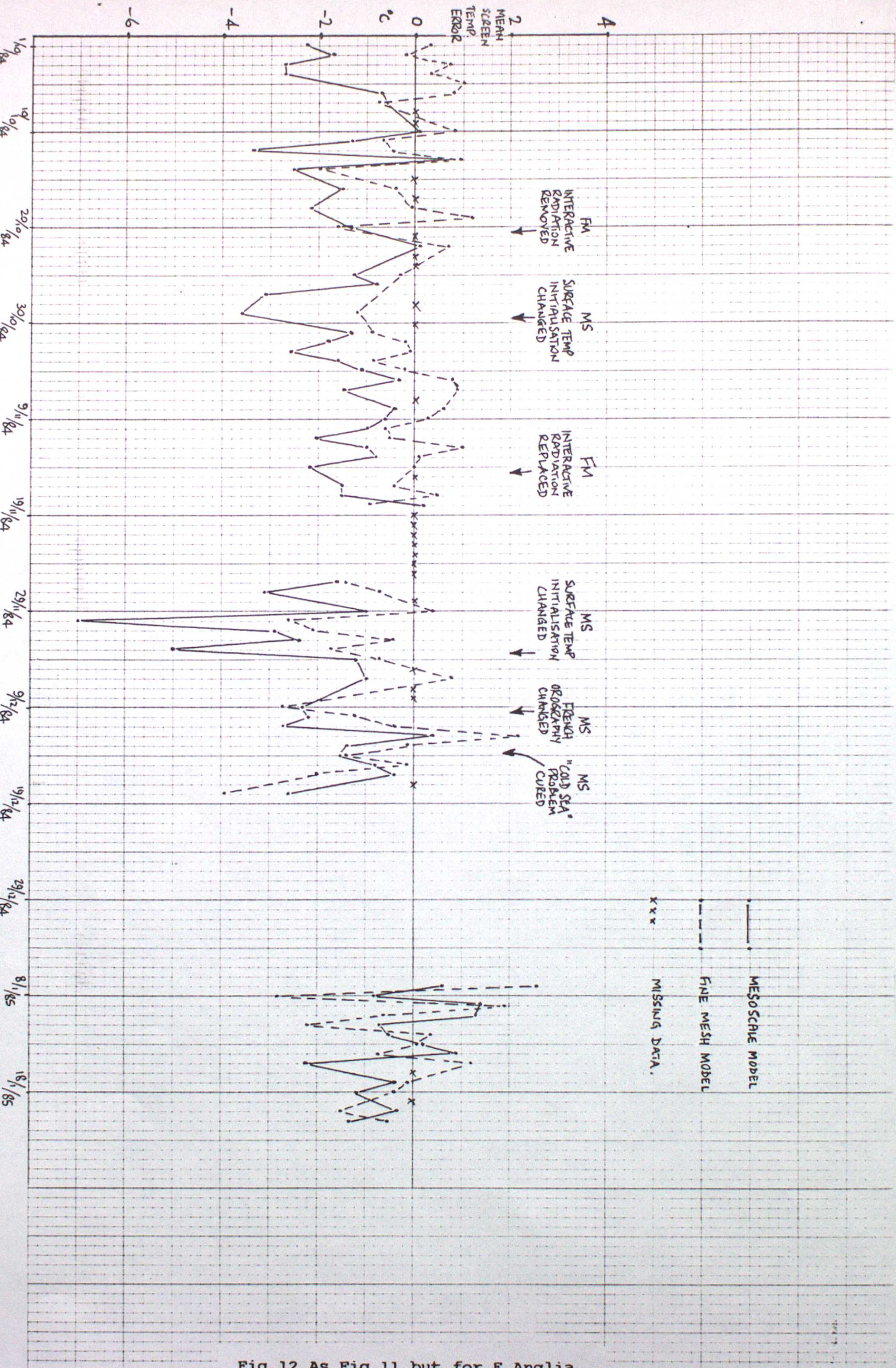


Fig.12 As Fig.11 but for E Anglia.

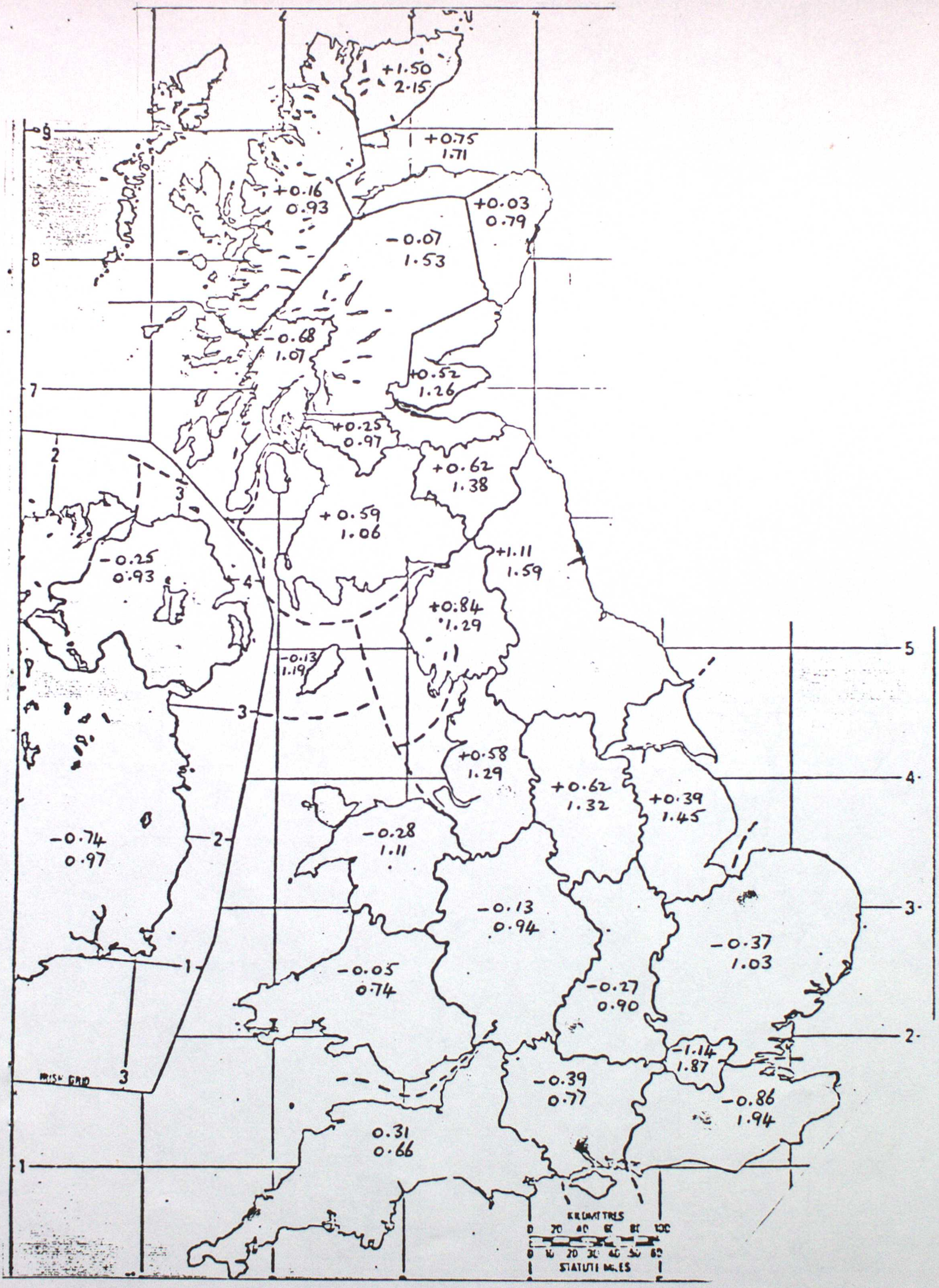


Fig.13 Mean (above) and RMS (below) errors of screen temperature at 1500 GMT for January. Mesoscale forecasts compared with analyses.

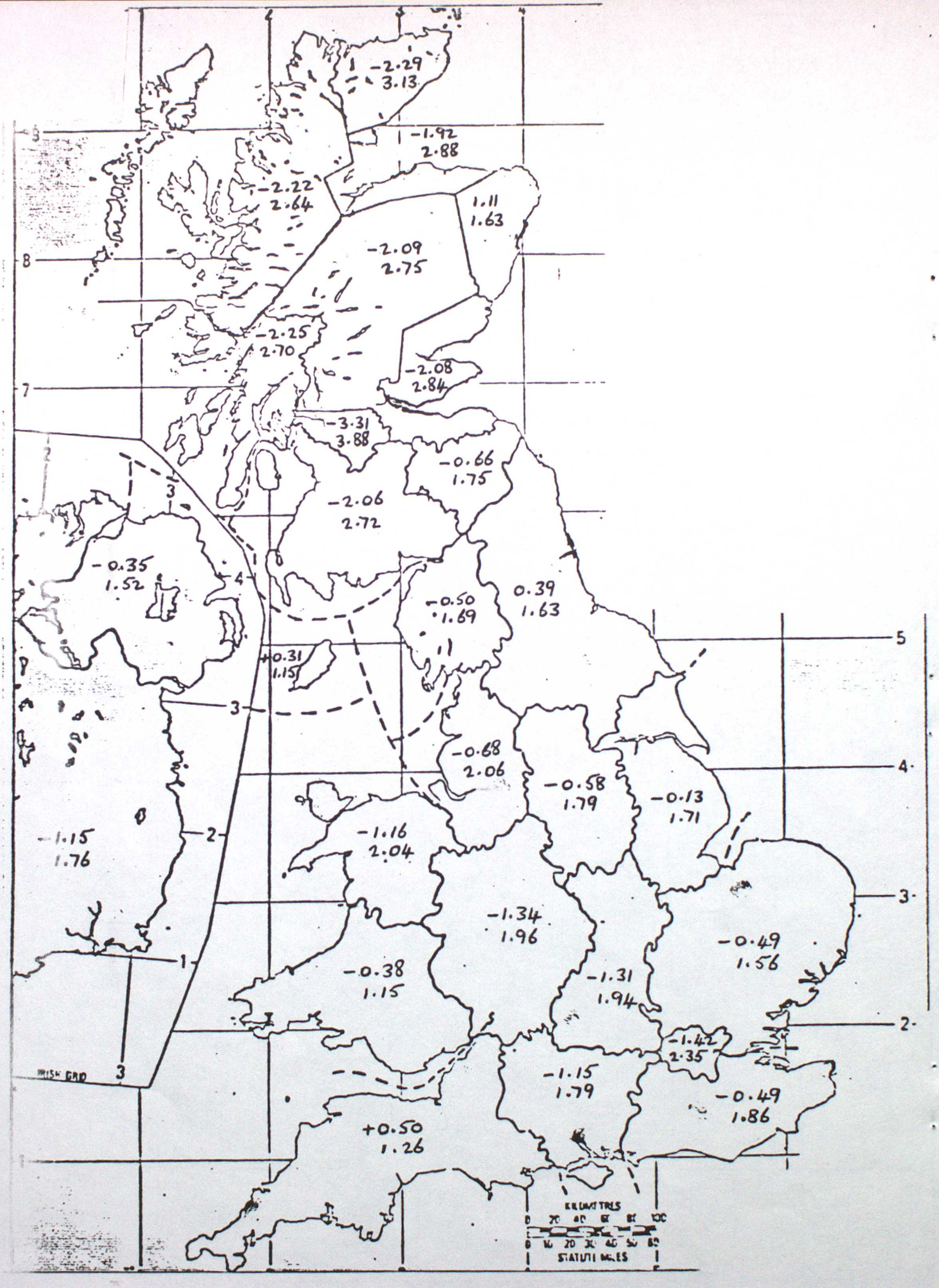


Fig.14 As Fig.13 but fine mesh forecasts compared with analyses.

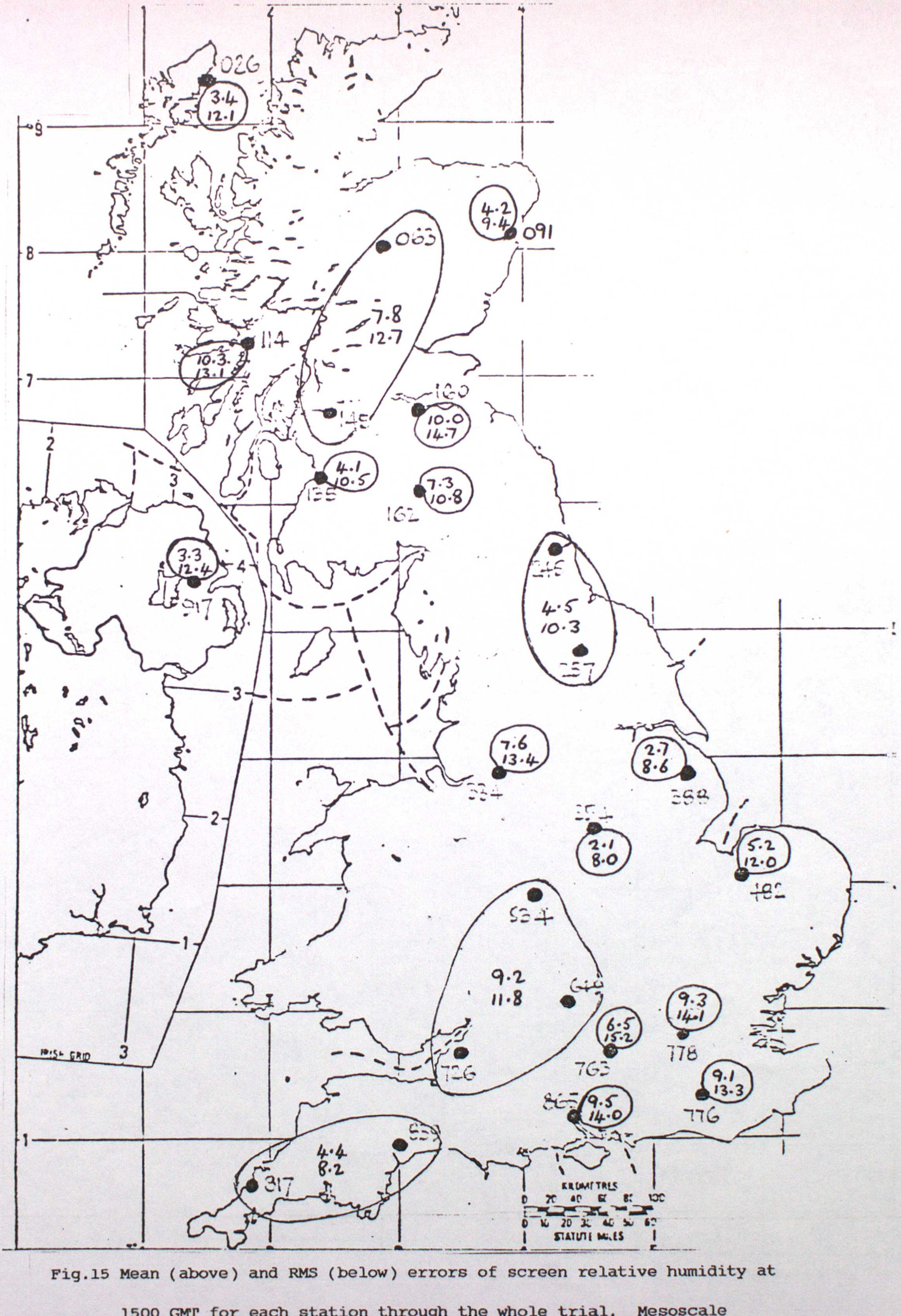


Fig.15 Mean (above) and RMS (below) errors of screen relative humidity at 1500 GMT for each station through the whole trial. Mesoscale forecasts compared with observations.

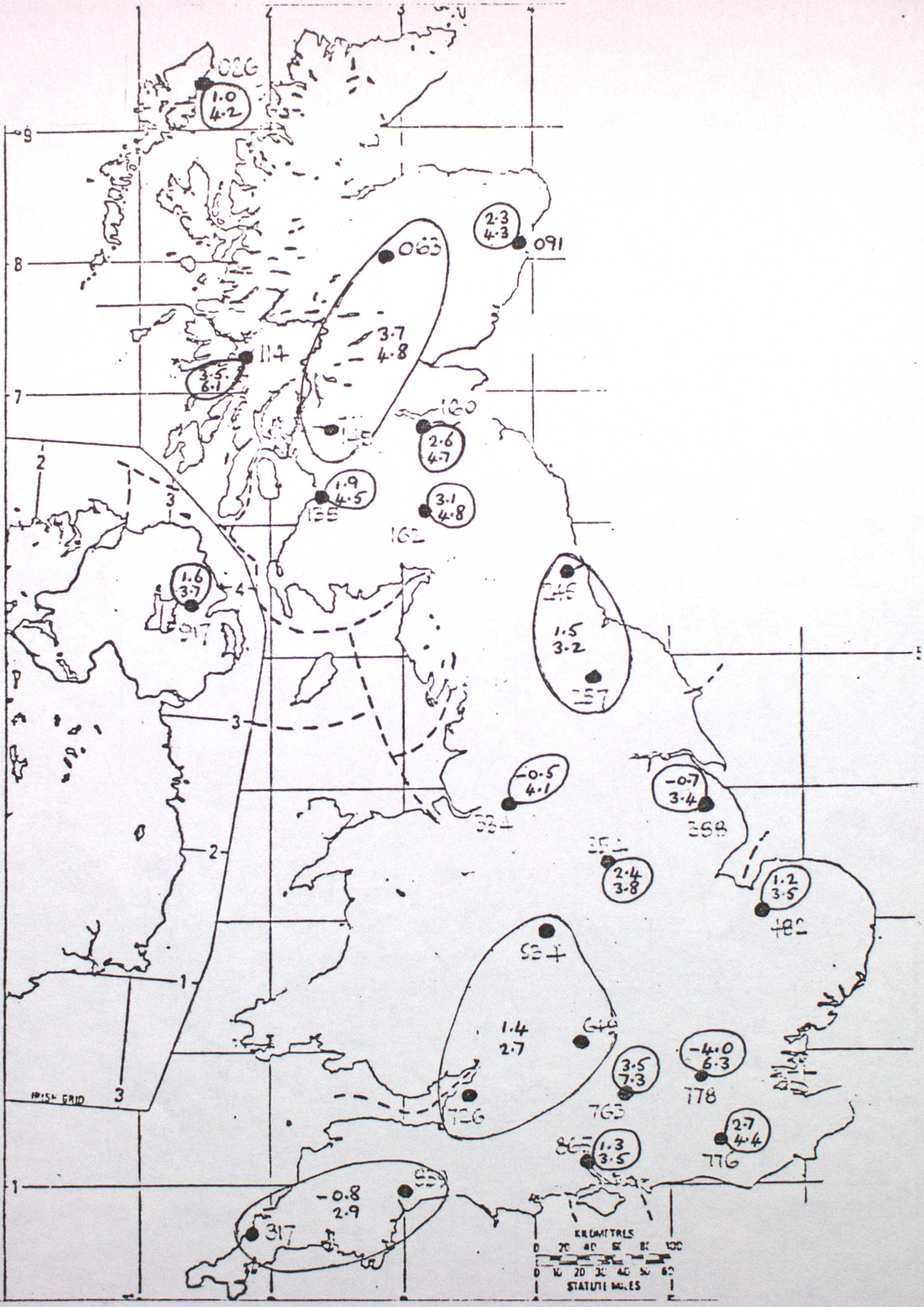


Fig.16 Mean (above) and RMS (below) errors of 10m wind speed at 1500 GMT for each station through the whole trial. Mesoscale forecasts compared with observations.

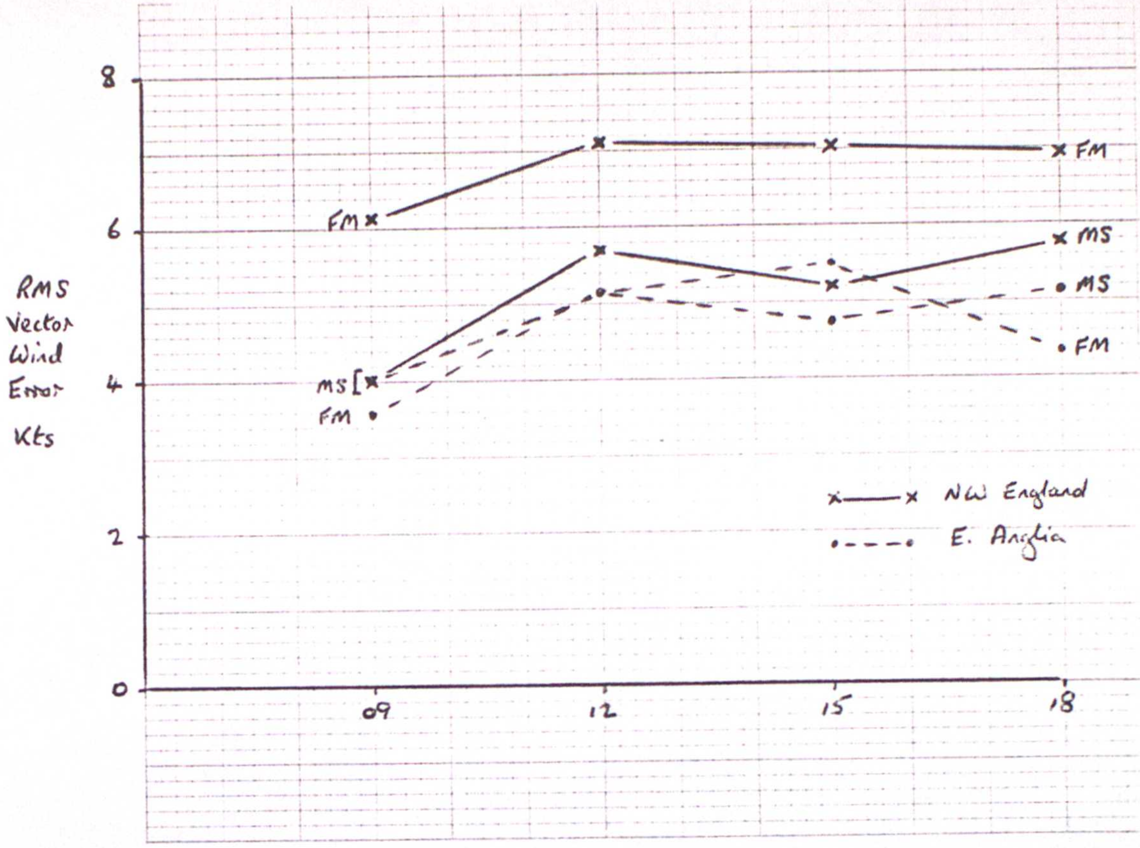


Fig.17 RMS vector errors for 10m wind as a function of time of day for the fine mesh and mesoscale models compared with analyses through the whole trial.

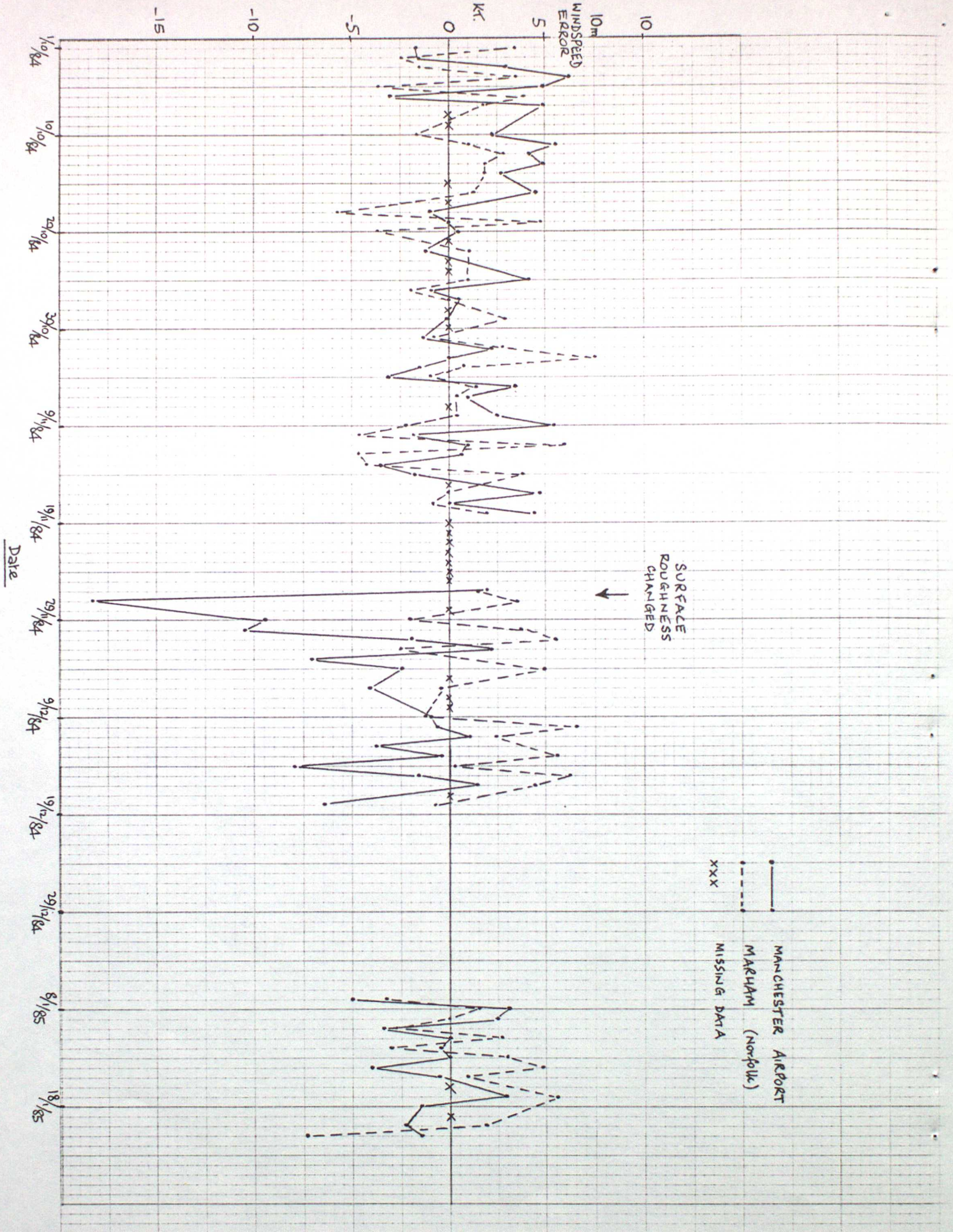
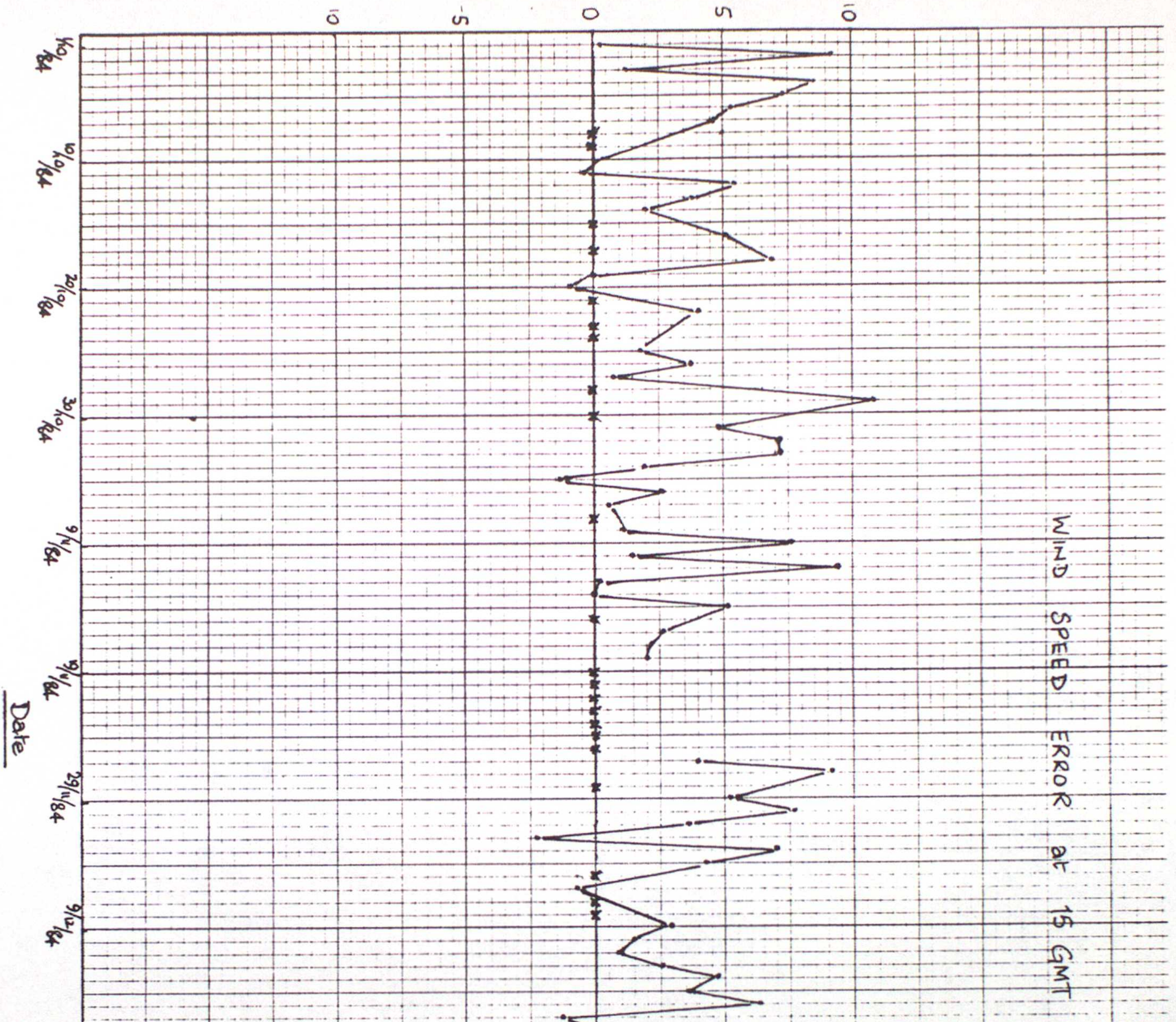
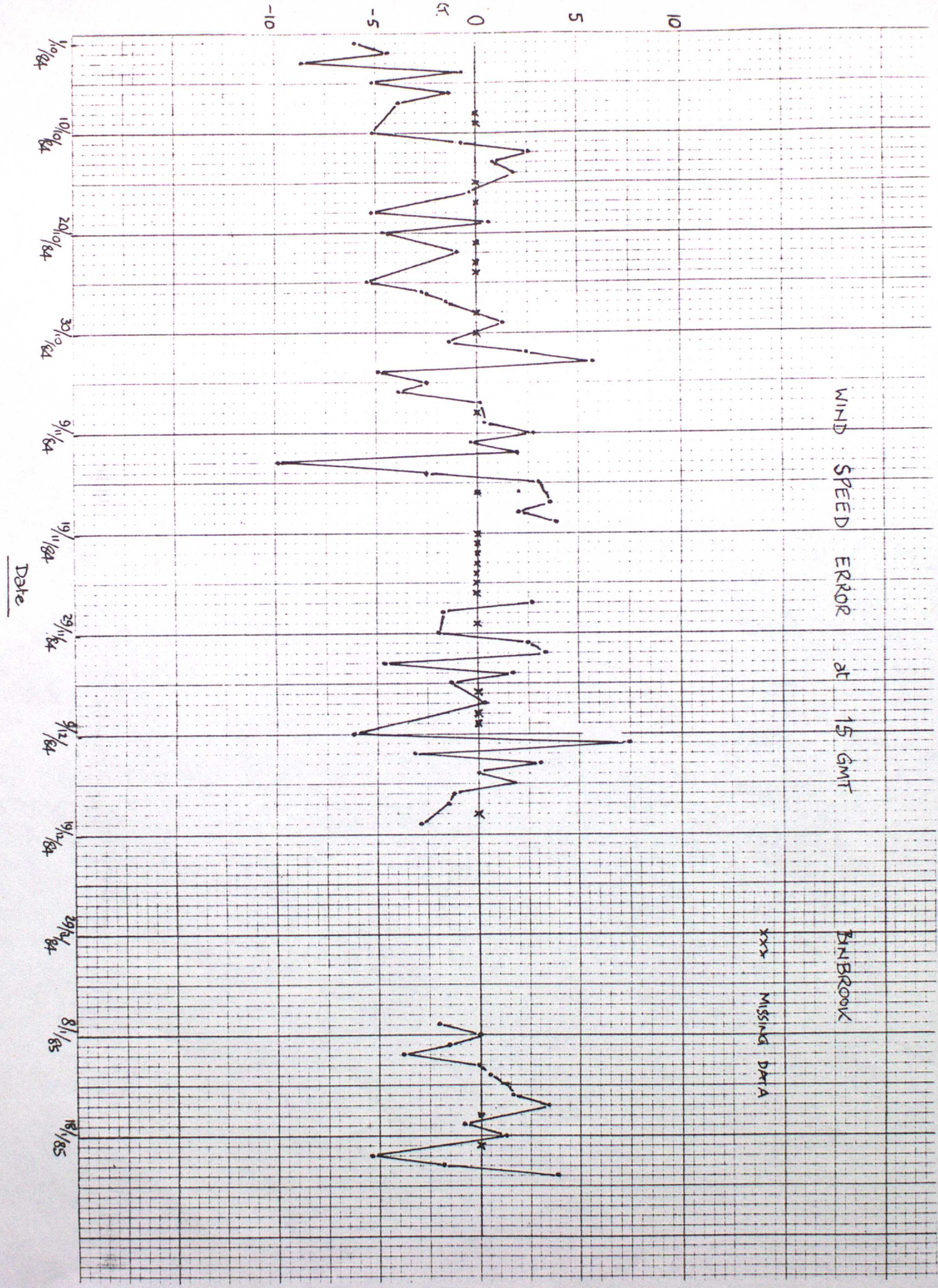


Fig.18 Daily series of 10m wind speed errors at 1500 GMT. Mesoscale forecasts compared with observations.

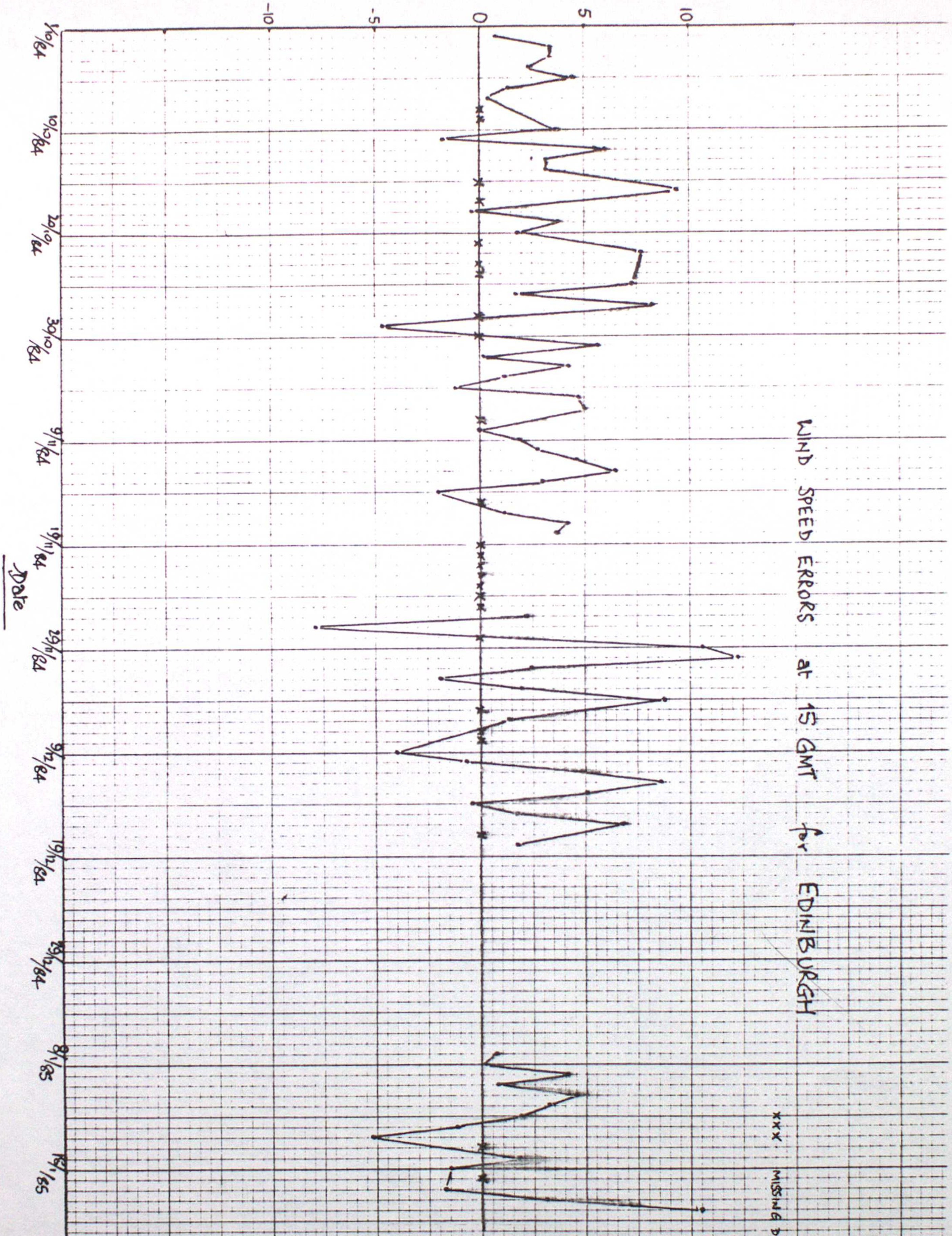
a) Manchester and Marham;


WIND SPEED ERROR at 15 GMT
GATWICK
xxx MISSING DATA
Date
-10
-5
0
5
10

b) Gatwick;

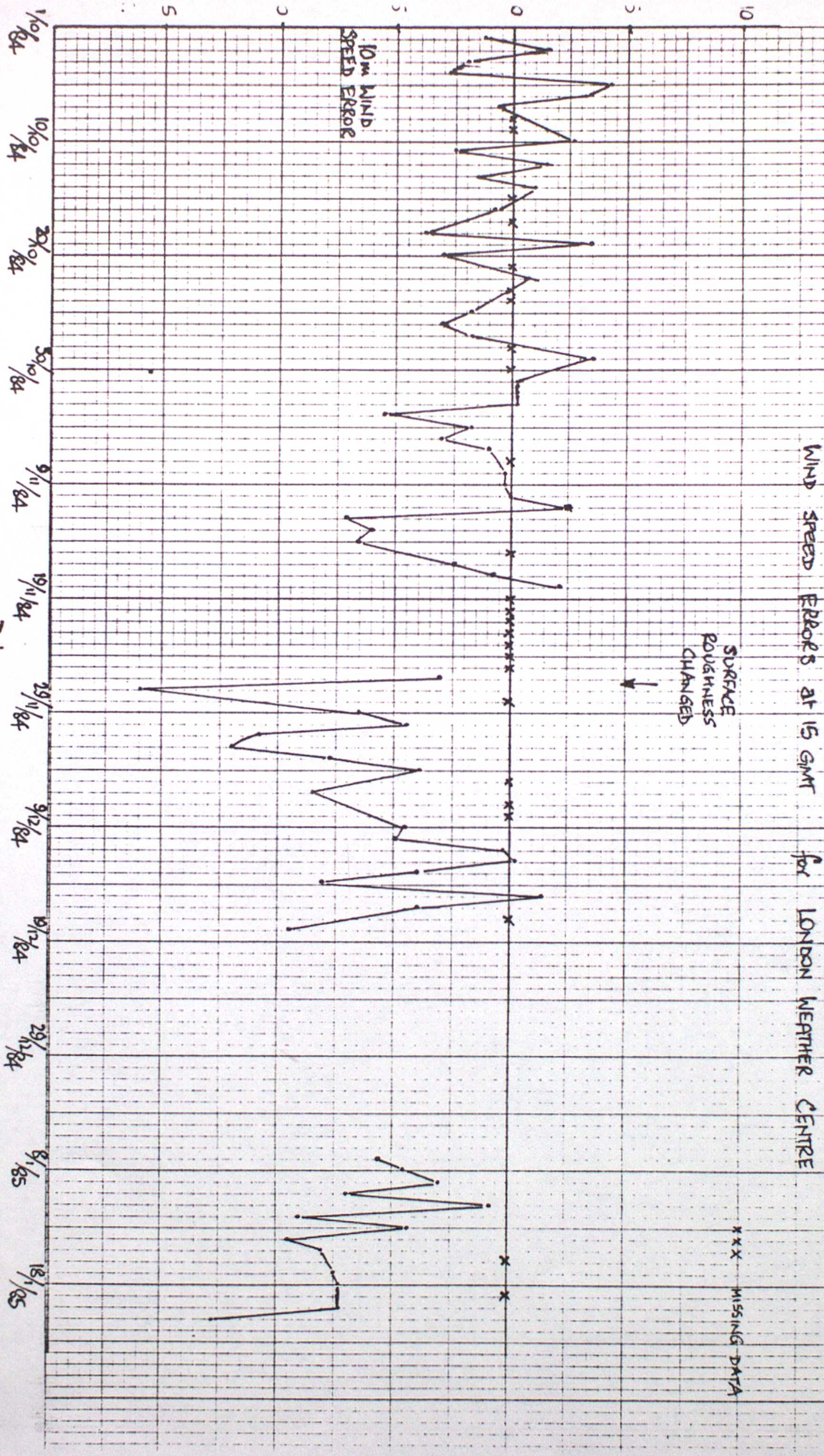

WIND SPEED ERROR at 15 GMT
BINBROOK
xxx MISSING DATA
10
5
0
-5
-10
1/10/84
10/10/84
20/10/84
30/10/84
9/11/84
19/11/84
29/11/84
9/12/84
19/12/84
29/12/84
8/1/85
18/1/85
Date

c) Binbrook;



WIND SPEED ERRORS at 15 GMT for EDINBURGH
xxx MISSING DATA
10
5
0
-5
-10
1/10/84
10/10/84
20/10/84
30/10/84
9/11/84
19/11/84
29/11/84
9/12/84
19/12/84
29/12/84
8/1/85
18/1/85
Date

d) Edinburgh;



e) London Weather Centre.

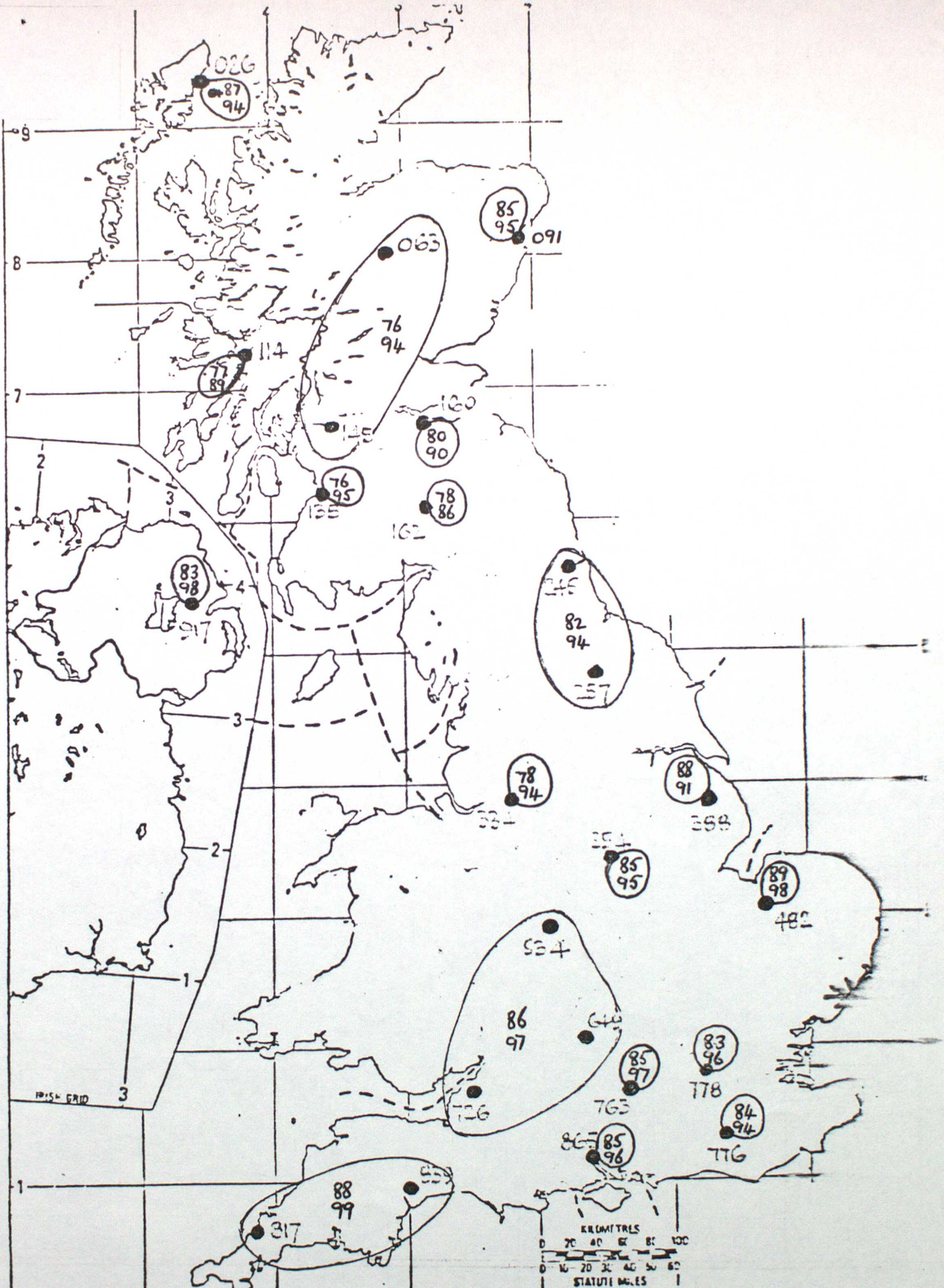


Fig.19 Percentage success rate of forecasts of rain/no rain (above) and moderate or heavy rain/light rain or no rain (below) for 1500 GMT through the whole trial. Mesoscale forecasts compared with observations

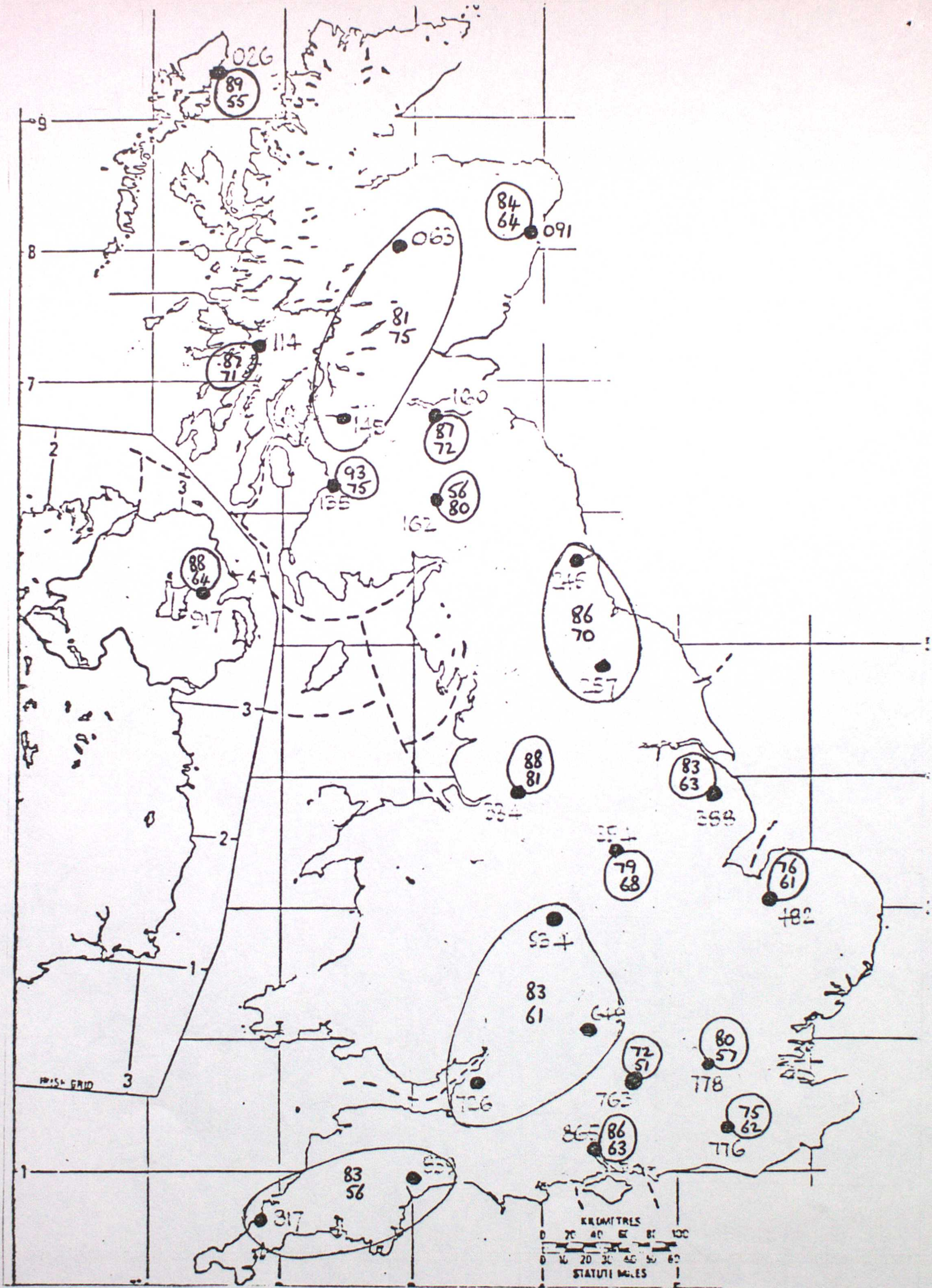


Fig.20 Percentage success rate of forecasts of cloud base below 1000 ft/ cloud base above 1000 ft (above) and cloud/no cloud (below) for 1500 GMT through the whole trial. Mesoscale forecasts compared with observations.