# Application and verification of ECMWF products 2009

*Met Eireann, Glasnevin Hill, Dublin 9, Ireland. – J.Hamilton*

## 1. Summary of major highlights

The verification of ECMWF products has continued as in previous years. We verify certain grid-field products [such as mean sea level pressure and 500hPa geopotential] against the corresponding ECMWF analyses. Various scores [such as the correlation coefficient, the rms error and the S1 score] are calculated for a 'large' area [corresponding to Western Europe and the North Atlantic] and a 'small' area [cantred around Ireland]. We also verify the 2-metre temperature and the accumulated precipitation against 4 synoptic stations in Ireland. Currently, we only verify forecasts based on the 12Z run.

## 2. Use and application of products

### 2.1 Post-processing of model output

#### *2.1.1 Statistical adaptation*

Not used to date.

#### *2.1.2 Physical adaptation*

ECMWF fields are used are boundary values for our limited area Hirlam model. The Hirlam model is run out to 48-hours 4-times per day.

#### *2.1.3 Derived fields*

Various products are derived from the deterministic model including forecasts of a potato-blight index. Some work is being done with the EPS model to predict conditions suitable for slurry spreading.

### 2.2 Use of products

The main use of ECMWF products is as guidance in the medium term. The various output fields are made available to the forecaster both as hardcopy output [using large-format ink-jet printers] and *via* an in-house developed interactive graphics system called **xcharts**. [This package runs on SGI workstations and on Linux PC's]. Selected products are also available as web-pages on the Met Eireann intranet.

The EPS products, especially the cluster fields for the North West Europe area, are used increasingly by the operational meteorologists to assess the likelihood of alternative forecast developments. We are also investigating the use of EPS rainfall products. More and more use is being made of the ECMWF member states website.

We continue to use ECMWF fields as boundary conditions for our Hirlam forecasts [with the fields inserted every three hours] and also as boundaries for our runs of the WAM wave-model. Since 2001, we have used frame boundary files for Hirlam. In 2006 we increased the number of vertical levels received from ECMWF as part of the upgrade to the T799 model.

Met Éireann and ICHEC [the Irish Centre for High-End Computing, see www.ichec.ie] signed a collaboration agreement early in 2007. As part of this collaboration, ICHEC provides computational facilities and support to Met Éireann to enable it to run its operational Hirlam model on ICHEC's supercomputers. Initially we used 32 dual-CPU nodes of the cluster "Walton" [an IBM eServer cluster 1350 consisting of 476 IBM e326 compute nodes]. But, at the end of 2008, we switched to using "Stokes" [an SGI Altix ICE 8200EX with 320 compute nodes; each compute node has two Intel (Harpertown) Xeon E5462 quad-core processors and 16GB of RAM]. Met Éireann is currently running version 7 of Hirlam [with 3DVAR] using eight nodes of the stokes cluster [*i.e.* 64 CPU's]. In addition, a backup version of Hirlam is run in-house on a small Linux cluster.

A number of EPS Metgrams and EPS Plumes (atmospheric and wave) for locations in Ireland are generated using the Critical Jobs system and displayed on our intranet for use by forecasters.

Precipitation probability and temperature anomaly charts for 92-hours to 120-hours, derived from the EPS system using Time Critical Jobs, are also displayed

Two MARS extraction processes are run, one for input to the EUMETSAT Nowcasting SAF and the other for the extraction of global forecast fields.

# 3. Verification of products

## 3.1 Objective verification

### *3.1.1 Direct ECMWF model output (both deterministic and EPS)*

Although we also verify various ECMWF fields against the corresponding analyses, this section will only discuss the verification of the direct model output of local weather parameters *viz.* temperature and precipitation.

*Direct Model Output of Local Weather Parameters: Temperature*

Since 1992 we have been verifying the ECMWF forecast of 2-metre temperature against four Irish synoptic stations *viz.* Mullingar, Shannon Airport, Valentia and Dublin Airport. In the case of each station, we interpolate values using the surrounding four grid-points and calculate the mean error, the mean absolute error, and, since July 1994, the rms error. [Previously, we also verified forecasts for Kilkenny and Clones – stations which closed in April 2008]

It is interesting to see how the quality of the forecasts has varied since 1992 and, in this section, we will present results to show that there have been significant improvements.

The model run is for 12Z and we examine the T+12, T+24, T+36, T+48, T+60, T+72, T+84, T+96 and T+108 forecasts. Note that the T+12, T+36, T+60, T+84, and T+108 forecasts verify at midnight and the T+24, T+48, T+72, and T+96 forecasts verify at midday. We have found, looking at the sixteen years of data, that [especially in the early years] there are significant differences in the quality of the forecasts verifying at midday and at midnight. Hence, we will treat these two cases separately.

Figure 1 shows verification scores for the runs verifying at midday for the four synoptic stations. We have plotted monthly means of the absolute error [blue lines] and of the mean error or bias [red lines]. The 4 blue lines and 4 red lines represent the T+24, T+48, T+72, and T+96 forecasts. It is not necessary to distinguish between the various blue lines and red lines to note the following points:

(a) The mean absolute error is typically between 1 and 2 degrees; in the early years it did not vary much with the forecast length; there is a large seasonal variation [although this has become less marked in later years]; the scores for the four stations are of comparable magnitude; and there is a gradual improvement of the scores with time.

(b) The mean bias is almost independent of the forecast length [the various red lines are almost superimposed]; in the early years it was negative for most stations [*i.e.* the forecast values were colder than the observations]; it then became more positive but, for the last few years, it has become generally slightly negative; the size of the bias has become smaller with time and the values for the four stations are similar.

Figure 2 shows the corresponding verification scores for the runs verifying at midnight. The 5 blue lines [mean absolute error] and 5 red lines [mean error or bias] represent the T+12, T+36, T+60, T+84, and T+108 forecasts. Again, it is not necessary to distinguish between the various blue lines and red lines to note the following points:

(a) The mean absolute error is higher for the runs verifying at midnight. Also, the scores for the various stations are quite different -- in particular the scores for Valentia were very poor until 1994. Again, however, the scores did not vary much with the forecast length, they showed a large seasonal variation [at least in the early years] and they showed a gradual improvement with time.

(b) Again, the mean bias is almost independent of the forecast length [the various red lines are almost superimposed]; in the early years it was negative for all stations but nowadays is generally positive. At present, the bias is similar for the four stations but in earlier years there were large variations.

Figure 3 and Figure 4 reinforce these results. They show smoothed monthly midday and midnight scores for the four stations. The lines were smoothed by taking 5-month running means centred on the month in question [*i.e.* the average of values for M-2, M-1, M, M+1, M+2 where M is the month]. Results for the various forecast lengths can now be distinguished. Again, the scores are generally better at midday than at midnight and there is a gradual improvement since 1992. Also, the seasonal variation has become less.

Figure 5 shows the result of averaging the monthly scores for the four stations. The top two plots show results without smoothing, the bottom two show the effects of taking a 5-month running mean. Again there has been a gradual improvement since 1992 and again the scores are better at midday than at midnight.

Next we consider seasonal variation of the scores. For the purpose of this study we divide the year into two 'seasons' called 'winter' [*viz.* Nov to Apr] and 'summer' [*i.e.* May to Oct]. This division is significant because, during the 'winter' season, we run a Vaisala road-ice model to predict road conditions at approximately 50 sites around Ireland. Input, for each of the sites, consists of time series of temperature, dew-point, cloud-cover, rainfall and wind. The forecaster usually starts with Hirlam data as a 'first-guess' and then modifies the data using a graphical editor; [this intervention can sometimes be substantial]. ECMWF data is available as a backup, and the graphical editor can also be applied to this data. Figure 6 compares the average of the four scores for the ECMWF model in 'winter' and in 'summer' and the plot also includes results for the whole calendar year [Jan to Dec]. The results shown are based on the average scores for the four synoptic stations and they confirm the gradual improvement in forecast quality described earlier. Of particular interest is the diagram showing the scores for the 'winter' forecasts verifying at midnight [central plot on right hand side] since these are directly relevant to the road-ice model. It can be seen that the average mean absolute error [for the four stations] is approximately 1-degree [at 36-hours] and the bias is almost zero.

Next we look at how the quality of the forecast varies with the lead time of the forecast. Figure 7 shows results for 2008 and it can be seen that the bias is more or less constant throughout the forecast [but it varies between stations and in some cases shows a strong diurnal variation] but the error [either mean absolute error or rms error] increases with the length of the forecast. More information is provided by the scatter plots of Figure 8 [forecasts verifying at midday] and Figure 9 [midnight]. These figures combine the results for all four stations in 2008. Looking at the two figures we see [again] that the size of the error increases slowly with forecast length. However, we also see a systematic trend in the bias related to the observed temperature. This effect is largest in the forecasts of midnight temperatures. Looking at the right-hand plots [in figure 9] it is clear that the forecast temperatures tend to be too high when it is cold [observed temperatures in the range 0C to 5C] and too low when it is warm [observations in the range 15C to 20C]. Thus the model tends to underestimate extreme events.

To summarise: the quality of the 2-metre temperature forecasts has shown a marked improvement since 1992; forecasts verifying at midday and midnight are of comparable quality [although in the past the midday forecasts were significantly better] and there is a systematic bias [especially for the midnight forecasts] which means the model predictions are too high in cold conditions.

*Direct Model Output of Local Weather Parameters: Precipitation*

Since 1992 we have being verifying the ECMWF forecast of total precipitation against these same four synoptic stations *viz.* Mullingar, Shannon Airport, Valentia and Dublin Airport. We verify the total precipitation for D+1 [36h-12h], D+2 [60h-36h], D+3 [84h-60h] and D+4 [96h-84h]. In the case of each station, we interpolate values using the surrounding four grid-points and calculate the mean error and the mean absolute error. We also carried out a categorical verification of the forecasts based on the three categories 0-0.3mm, 0.3-5mm and greater-than 5mm.

We calculated the Heidke Skill Score for each station. This measure of skill gives 1.0 for a perfect forecast and 0.0 for a forecast which is no better than chance. The results we obtained are summarised in Figure 10. These plots show smoothed values of the mean monthly Heidke score for the four stations. The smoothing is carried out by means of a 5-month centred running mean. The results show that there is skill in the rainfall forecast and that the shorter forecasts are more skilful than the longer. There appears to have been some improvement, in skill, over the past sixteen years. This is most marked in the three-day and four-day forecasts.

### *3.1.2 ECMWF model output compared to other NWP models*

This is only done in an arbitrary subjective manner by operational forecasters on a day to day basis without formal recording of same.

### *3.1.3 Post-processed products*

Post-processed products are not used operationally to date.

### *3.1.4 End products delivered to users*

Only direct model output (DMO) and interpolated DMO outputs are available.

### 3.2 Subjective verification

*3.2.1 Subjective scores (including evaluation of confidence indices when available)*

None.

*3.2.2 Synoptic studies*

None are done due to staff shortages and other priorities.

## 4. References to relevant publications

None.

Completed ... 28-May-2009

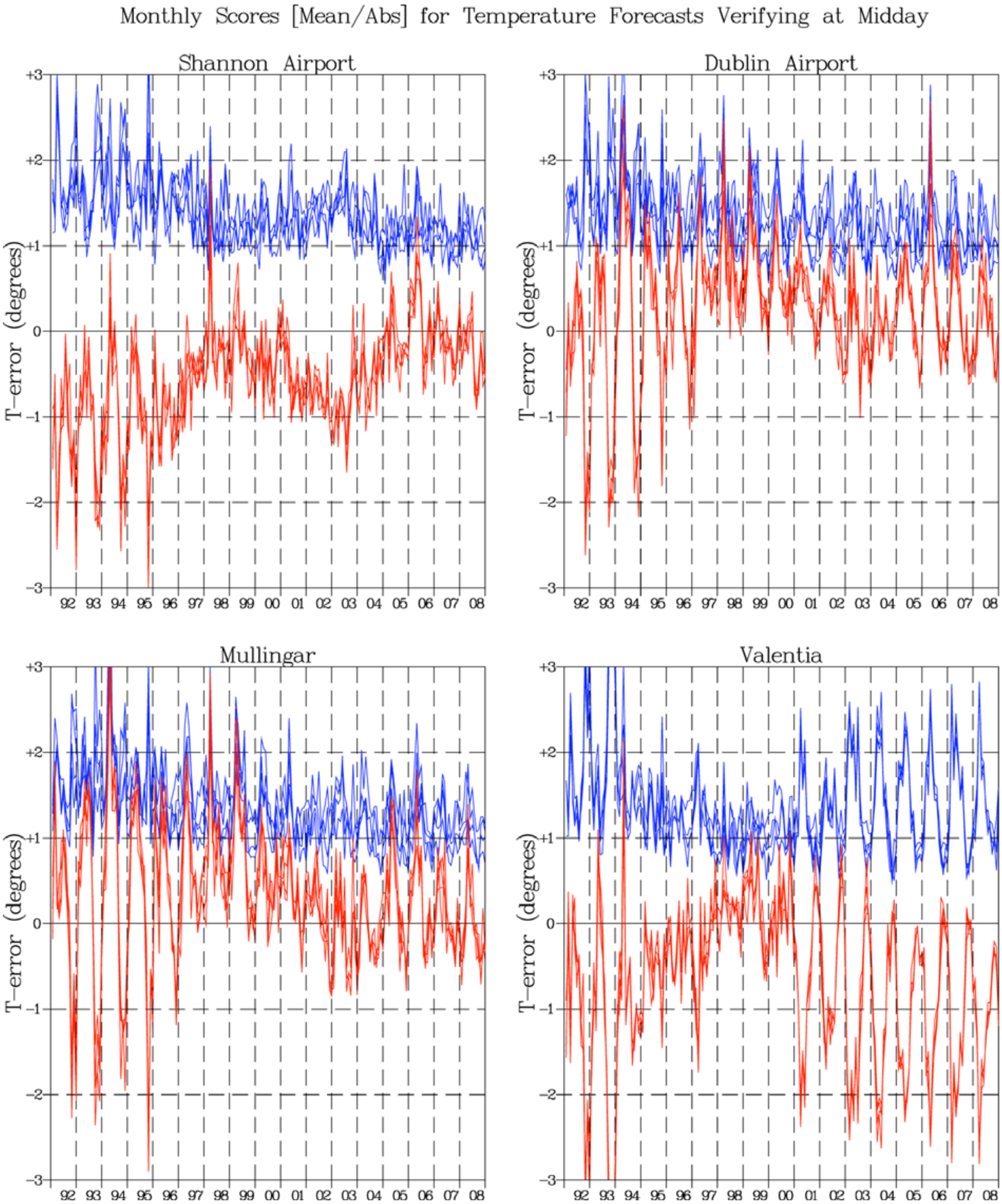


Fig. 1 Scores for the T+24, T+48, T+72, and T+96 ECMWF forecasts of 2-metre Temperature. Note that all these forecasts verify at midday. The blue lines show values of the monthly mean of the absolute error, the red lines values of the monthly mean of the mean error or bias.

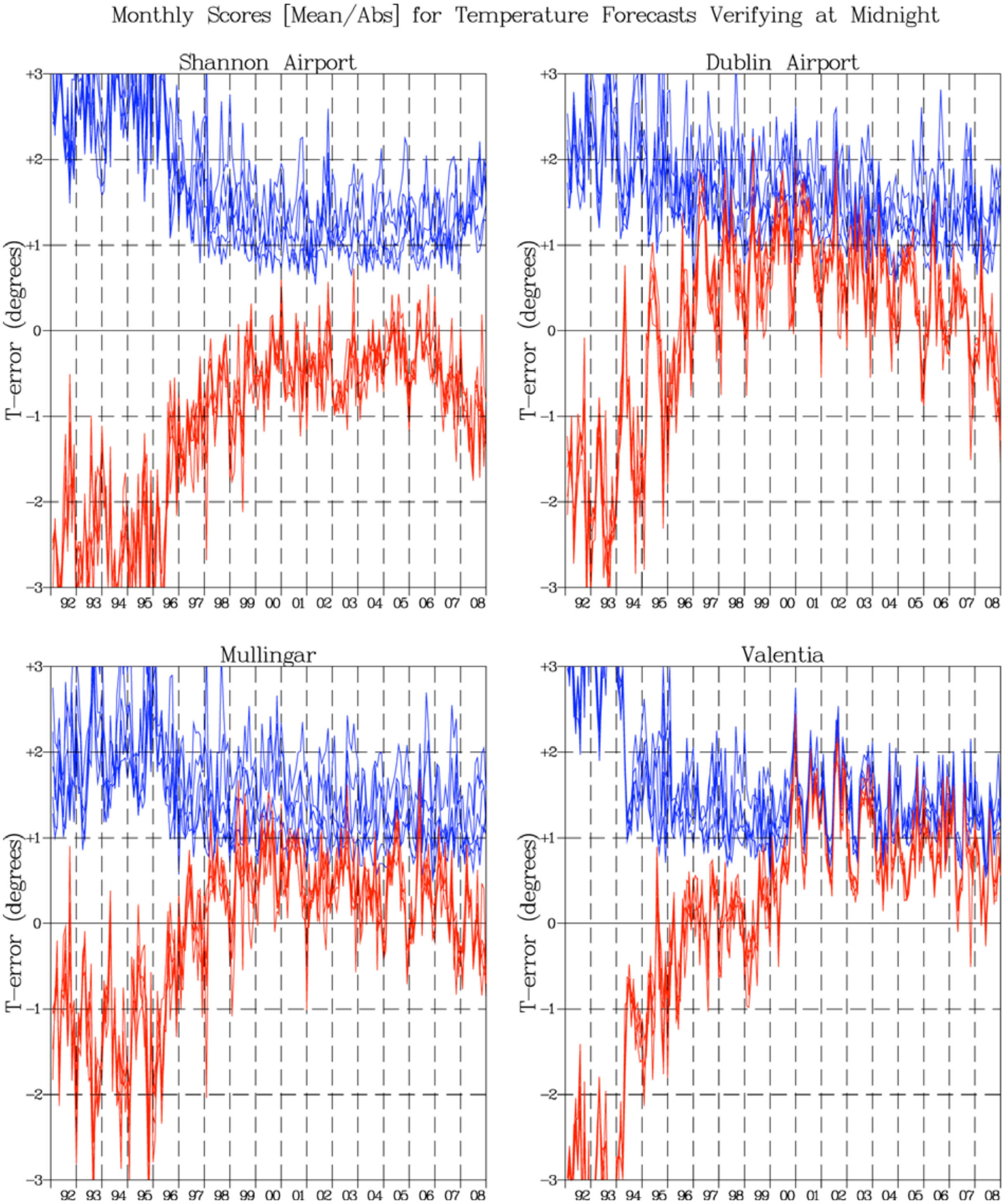


Fig. 2 Scores for the T+12, T+36, T+60, T+84, and T+108 ECMWF forecasts of 2-metre Temperature. Note that all these forecasts verify at midnight. The blue lines show values of the monthly mean of the absolute error, the red lines values of the monthly mean of the mean error or bias.

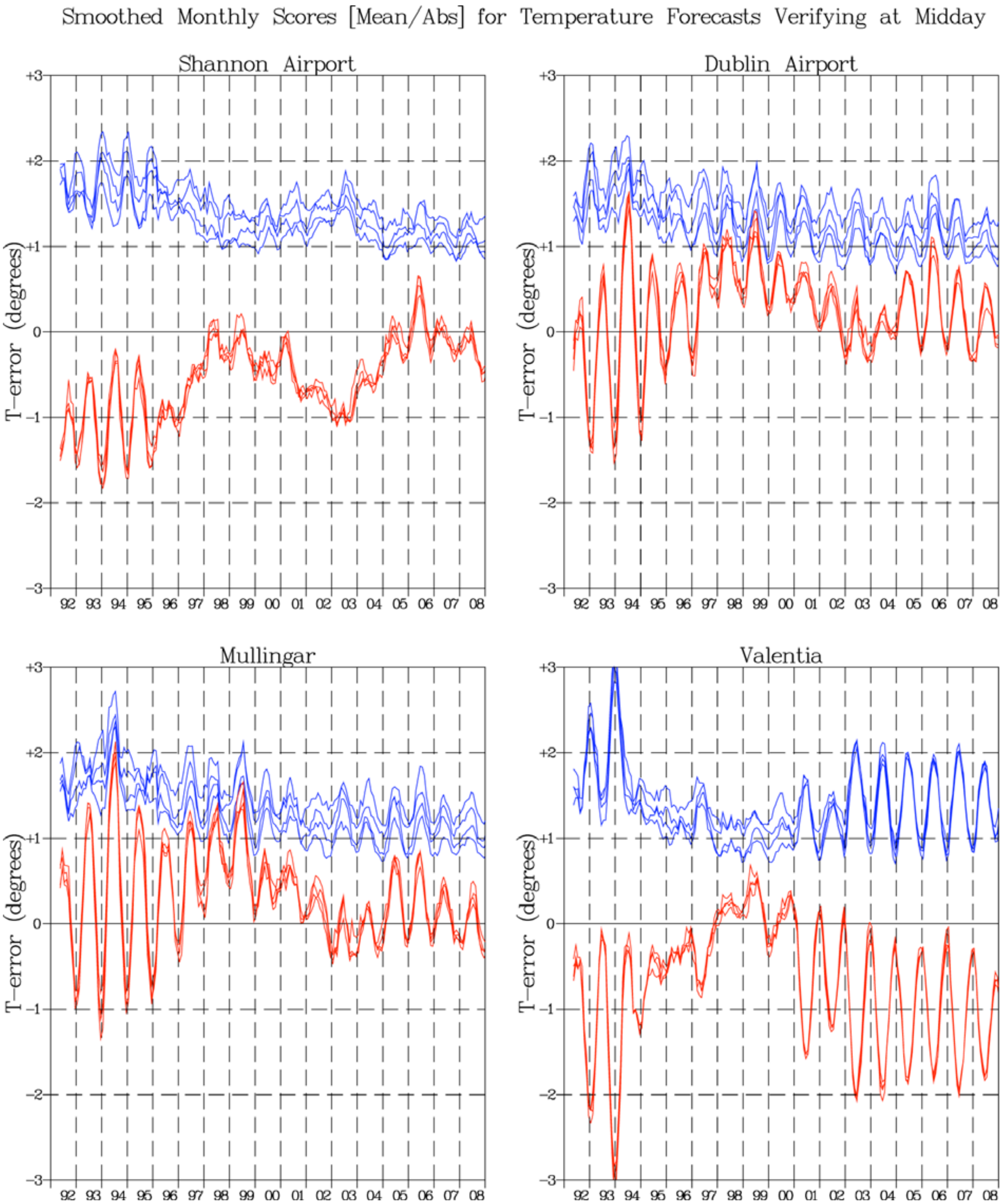


Fig. 3 Smoothed scores for the T+24, T+48, T+72, and T+96 ECMWF forecasts of 2-metre Temperature verifying at midday. The lines were smoothed by taking a 5-month centred running mean. The blue lines show values of the monthly mean of the absolute error, the red lines values of the monthly mean of the mean error or bias. The top blue line corresponds to a T+96 forecast, the one below that to a T+72 forecast *etc.*

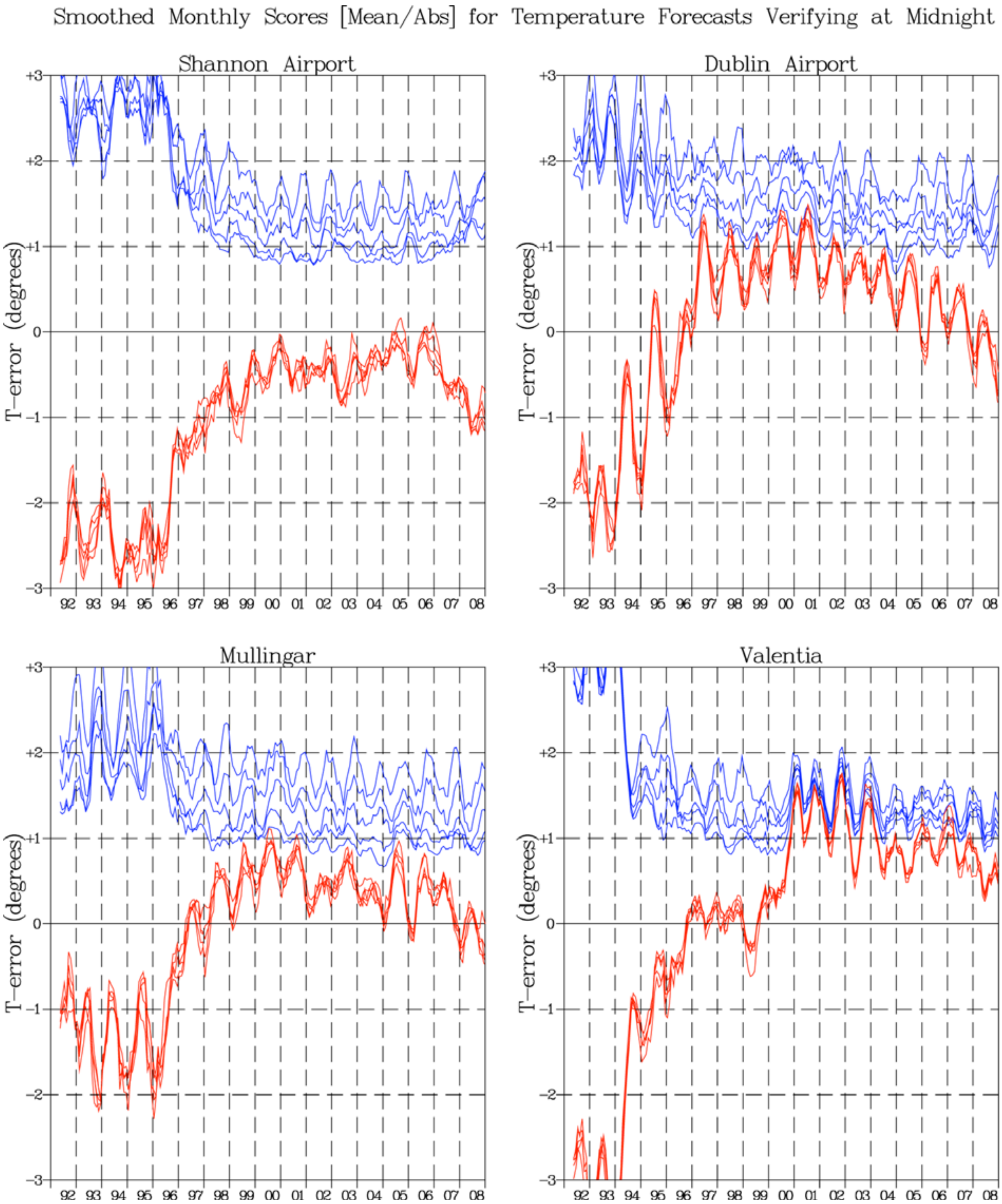


Fig. 4 Smoothed scores for the T+12, T+36, T+60, T+84, and T+108 ECMWF forecasts of 2-metre Temperature verifying at midnight. The lines were smoothed by taking a 5-month centred running mean. The blue lines show values of the monthly mean of the absolute error, the red lines values of the monthly mean of the mean error or bias. The top blue line corresponds to a T+106 forecast, the one below that to a T+84 forecast *etc.*

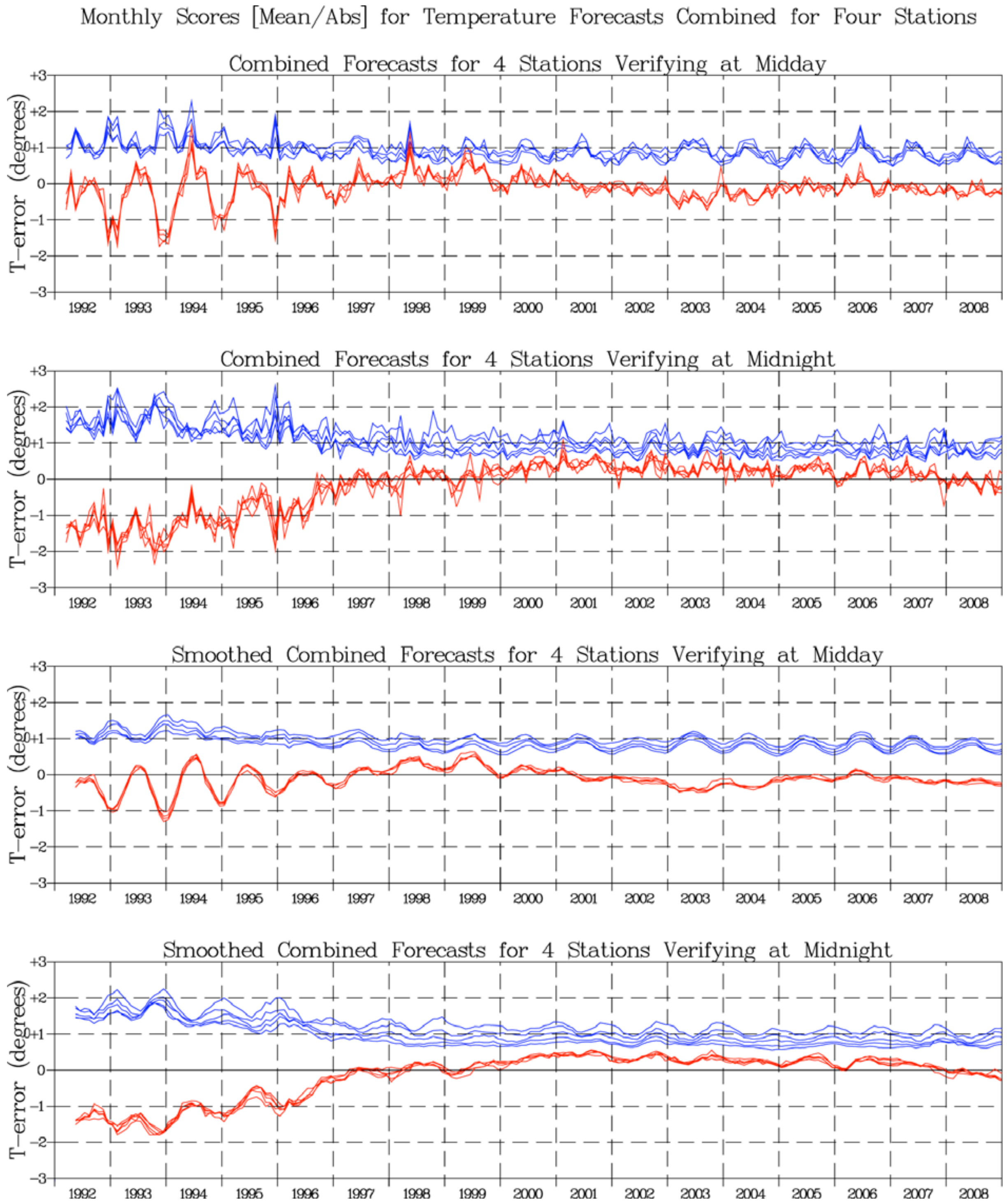


Fig. 5 Mean monthly scores for ECMWF forecasts of 2-metre Temperature averaged over four stations. The top two graphs show the scores for midday [T+24, T+48, T+72, and T+96] and midnight [T+12, T+36, T+60, T+84, and T+108], respectively, without smoothing; the bottom two graphs show the effects of smoothing. The blue lines indicate the error [mean absolute error], the red lines the bias [mean error]. In all cases the top blue line corresponds to the longest forecast, the bottom blue line to the shortest.

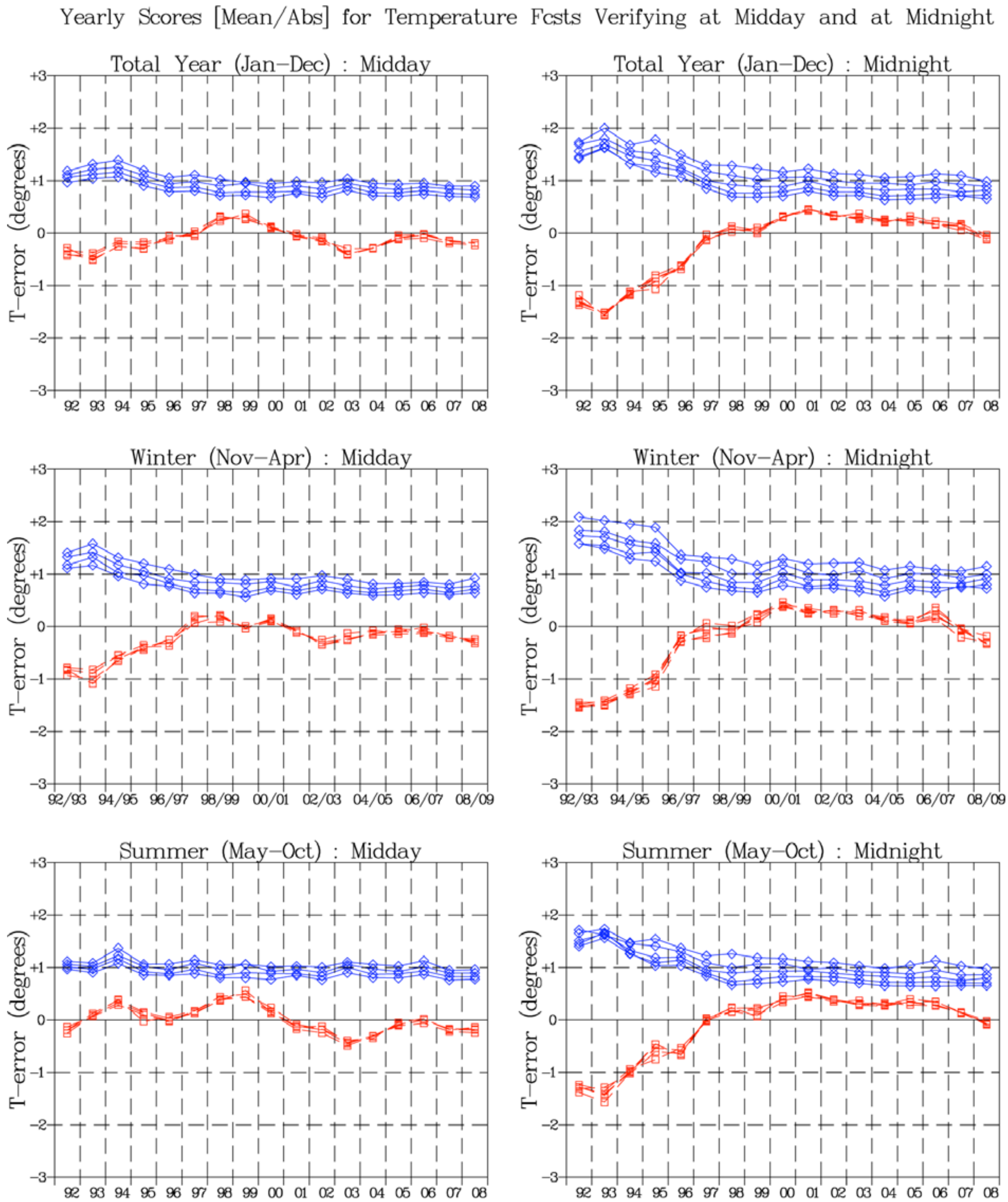


Fig. 6 Mean seasonal and yearly scores for ECMWF forecasts of 2-metre Temperature averaged over four stations. The blue lines indicate the error [mean absolute error], the red lines the bias [mean error]. Note that in 1992 the bias was much larger for the forecasts verifying at midnight rather than at midday but this effect became much less in later years [red lines]. Similarly, the error was greater, in 1992, for the midnight runs but gradually, over time, the difference became less [blue lines]. The errors for the various forecast lengths can be distinguished from the graphs: in all cases the top blue line corresponds to the longest forecast, the bottom blue line to the shortest.

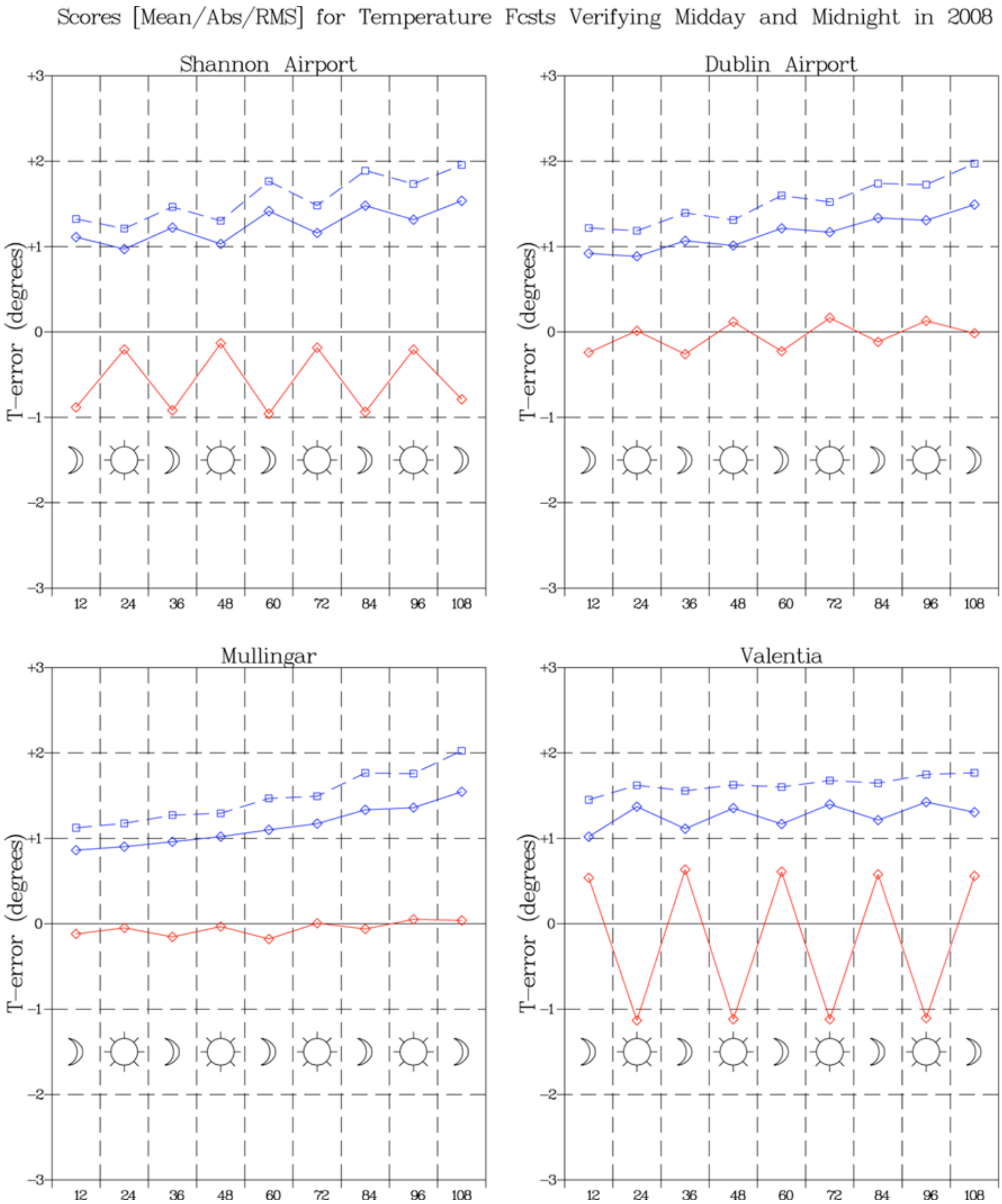


Fig. 7 Mean scores for ECMWF forecasts of 2-metre Temperature, in 2008, averaged over a year of data. The solid blue line is the mean absolute error, the dashed blue line the rms error and the red line the mean error or bias. It can be seen that the quality of the forecast decreases with the length of the forecast. The 'sun' and 'moon' symbols indicate forecasts verifying at midday and midnight, respectively. The length of the forecast [in hours] is indicated on the x-axis.

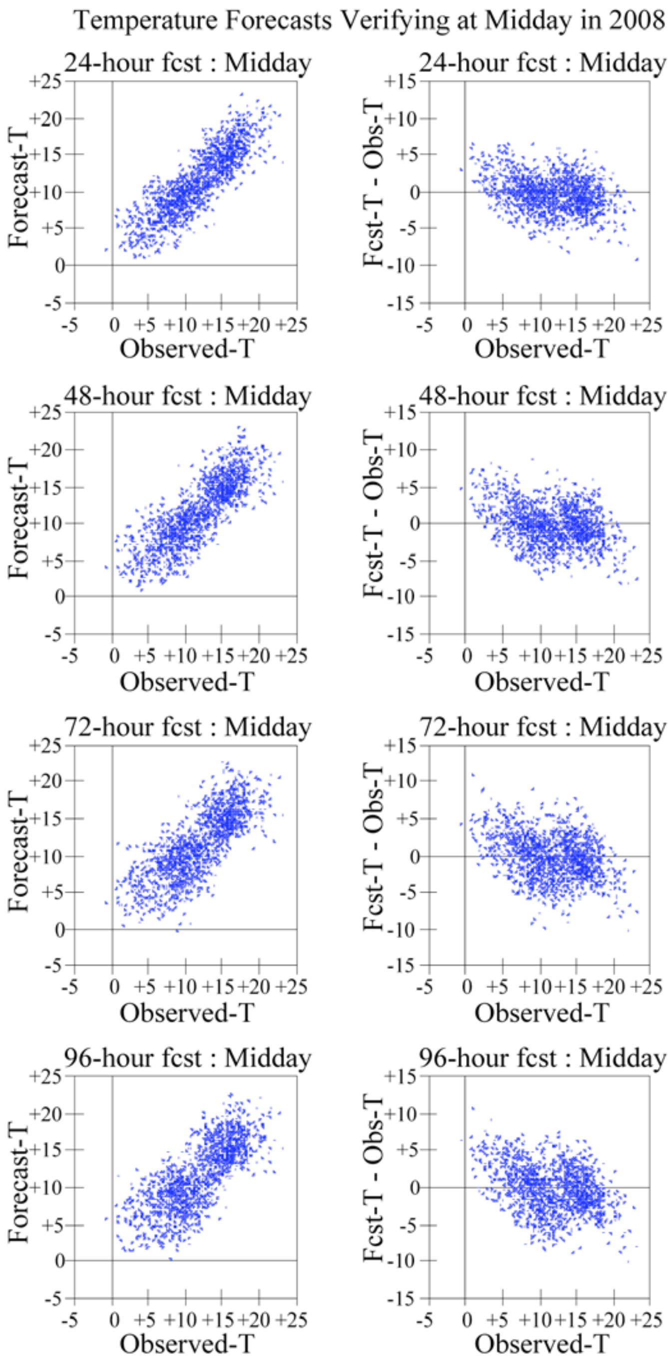


Fig. 8 Scatter plots for ECMWF forecasts of 2-metre Temperature, in 2008. All forecasts verify at midday and the four stations have been combined.

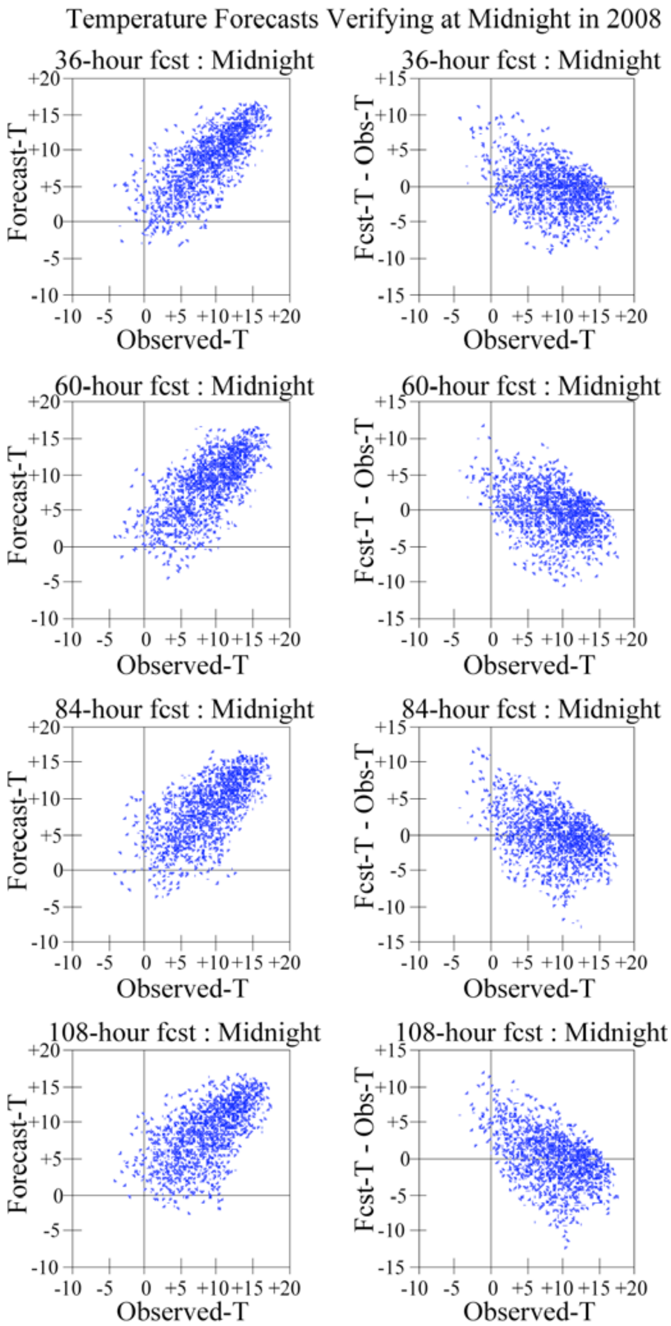


Fig. 9 Scatter plots for ECMWF forecasts of 2-metre Temperature, in 2008. All forecasts verify at midnight and the four stations have been combined.

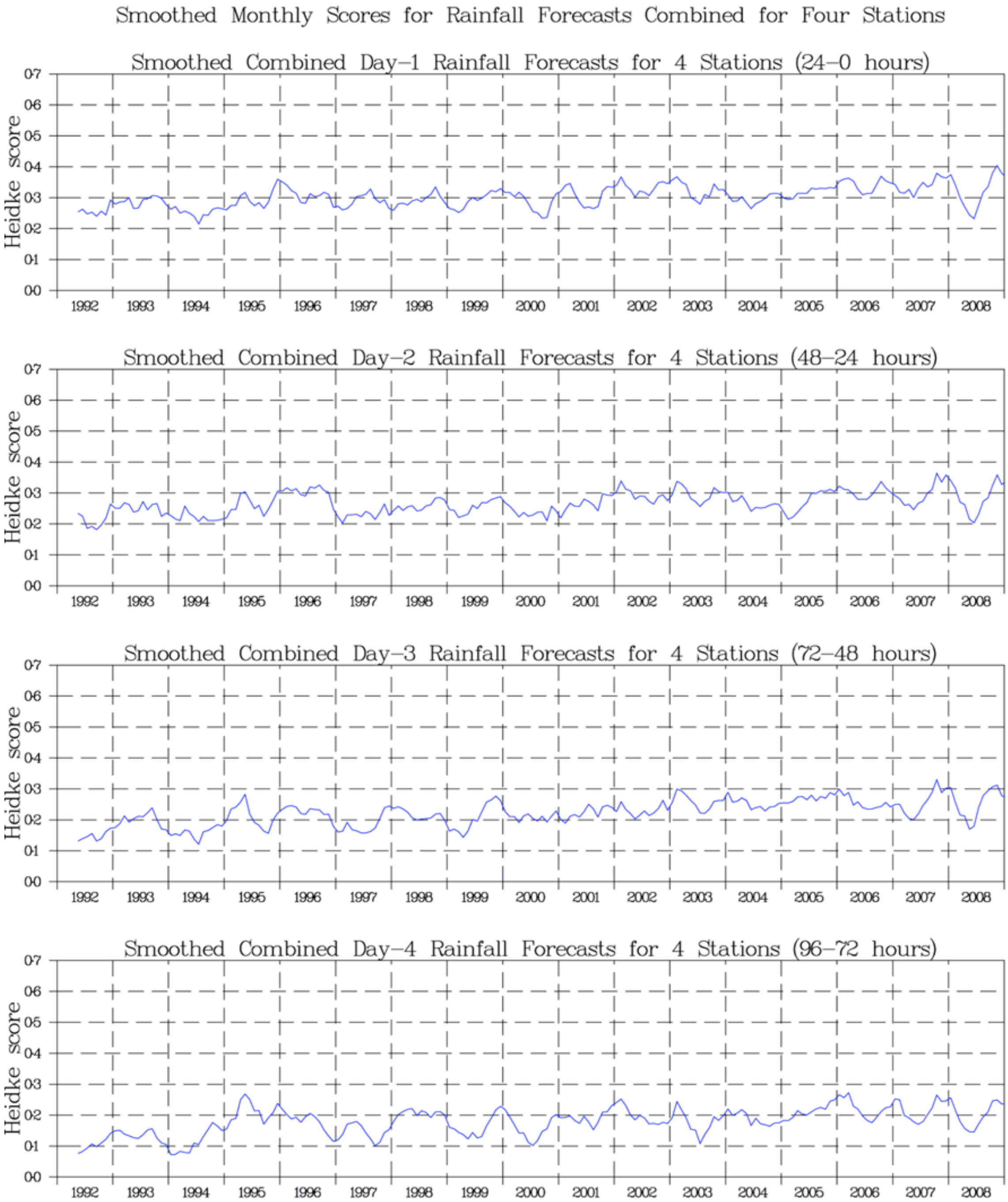


Fig. 10 Verification of precipitation forecasts: The plots show smoothed values of the mean monthly Heidke score for four stations. The smoothing is carried out by means of a 5-month centered running mean. The larger the value of the score the better the forecast. The results show that there is skill in the rainfall forecast and that the shorter forecasts are more skilful than the longer.