Advances in
Science & Research
Open Access Proceedings

# Determining the accuracy of gridded climate data and how this varies with observing-network density

**T. P. Legg**

National Climate Information Centre, Met Office, Exeter, UK

**Abstract.** The Met Office National Climate Information Centre regularly produces assessments of mean monthly, seasonal and annual values of weather parameters and their anomalies over the UK. However the gridded values, and corresponding areal-average values, are subject to error. Experiments have been done in an attempt to quantify the mean errors in gridded monthly values and monthly areal averages of temperature and rainfall, and how these errors vary when we artificially thin out the observation network. But there are two additional reasons for this work: firstly we wish to determine how far back we can realistically extend the historical areal series, and secondly, we want to estimate the size of error bars on the historical values. For the UK as a whole, we estimate that error bars of around 0.1 °C would arise for monthly-mean temperatures.

## 1 Introduction

The Met Office National Climate Information Centre regularly produces assessments of mean monthly, seasonal and annual values of weather parameters and their anomalies over the UK[1]. These are based on daily measurements at a network of observing sites, and produced by interpolation to a regular grid as described by Perry and Hollis (2005). Monthly series of gridded temperature and rainfall at 5 km resolution have been produced back to 1910, and of sunshine back to 1929. However the gridded values, and corresponding areal-average values, are subject to error (e.g. due to instrument calibration and reading, site differences, sampling, interpolation). For example, Perry and Hollis (2005) showed that the root-mean-square error for point temperature estimates is around 0.4 °C on average. This paper considers the sampling and interpolation errors, which will vary as a function of network spacing.

Experiments have been done in an attempt to quantify how the mean errors in monthly gridded values and monthly areal averages of temperature and rainfall vary with the size of the observation network. The number of available stations has declined from a peak of around 570 (corresponding to an average horizontal spacing between stations across the UK of approximately 21 km) for temperature and 5400 (7 km) for rainfall in the 1970s, to about 400 (25 km) and 3000 (9 km) respectively in 2010. The objectives of this work were to determine how far back we can realistically extend the historical areal series, and to estimate the size of error bars for the historical values.

## 2 Station data and gridding method

Daily measurements are made of maximum and minimum temperatures, rainfall totals and sunshine amounts. These station observations are then used as the input to our gridding process, which uses regression and interpolation to obtain estimated values on a 5 km × 5 km grid of points to give a better representation of spatial variations than is possible from the station observations alone. There are approximately 10 000 grid-points covering the UK. This process effectively adds information, due to knowledge of orography, exposure etc. even in unobserved areas. Areal averages are then calculated using all grid-point values within each area. The averages should be unbiased overall (a mixture of urban, coastal, upland etc.). Conversion to anomalies is then done using the long-term average for each area.

[1]These UK climate data are available at www.metoffice.gov.uk/climate/uk/index.html.

**Figure 1.** UK station networks for temperature and rainfall; 20 independent verification stations; examples of thinned-out subsets of the network. **(a)** Full station network for temperature observations. **(b)** Full station network for rainfall observations. **(c)** 20 independent verification stations. **(d)** Example of selection of stations left in network when 90% were removed. **(e)** Example of selection of stations left in network when 95% were removed.

## 3  Methodology of this study, and some results

From the full station network (shown in Fig. 1a, b), we take a well-spaced subset of 20 sites as independent verification stations (Fig. 1c). Then we estimate the mean errors at these 20 sites, as interpolated from a gradually decreasing network obtained by artificially removing certain proportions of the original set of stations. Figure 1d and e shows examples of substantially thinned-out subsets of the original network. The network is thinned by removing, in a random manner, fixed percentages of the complete station network (e.g. 4 out of every 10 stations). It is found that the errors increase only slowly as fewer station observations are retained; examples

are given in Figs. 2 and 3. This reflects spatial coherence in the data, especially for temperature: when a station is lost, neighbouring stations provide some, though not all, of the missed information.

Biases of the gridded values have been estimated by selecting one independent station at a time and obtaining an estimate of its average error, and relating these errors to location, orography, etc. Individual-station biases tend to be up to ~2 °C, and sometimes (Fig. 4) there is a tendency for similarly-situated stations (e.g. those in valleys) to have similar errors, which constitute elevation-dependent biases.
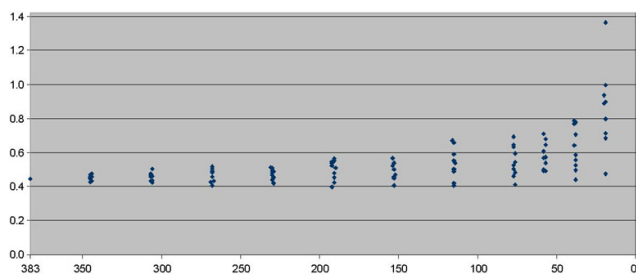
**Figure 2.** Minimum Temperature, April 2007 – Estimated root-mean-square errors (°C) at a subset of 20 independent points, as a function of the number of stations remaining out of the original 403 available (ten runs were made for each proportion of stations removed).
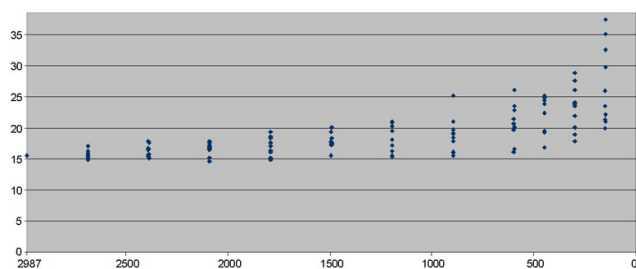


**Figure 3.** Rainfall, July 2007 – Estimated root-mean-square errors (mm) at a subset of 20 independent points, as a function of the number of stations remaining out of the original 3007 available (ten runs were made for each proportion of stations removed).

## 4   Extending our series further back in time

The availability of station data is increasingly limited as we go further back before 1910. But the results presented here show that the mean error of interpolation to the locations of a set of independent stations, increase only slowly as the available station network is artificially thinned out. This is partly because of spatial coherence in the data, especially for temperature. This implies also that errors of areal-average values may increase only slowly as the network is thinned, so that we may be able usefully to extend our series back to earlier dates.

How far back can we extend our series before the errors become unacceptably large? It appears that we need to have approximately 100 temperature stations (average spacing 50 km) (Fig. 2) and 500 rainfall stations (22 km) (Fig. 3) across the UK in order that the estimated mean error at independent sites remains no more than 1.5 times what it is with the current full station network. Currently our historical series for temperatures and rainfall extend back to 1910, but these results, in combination with our information on historical observing networks, suggest that we could reliably extend the series back to at least 1901.

## 5   Errors in areal values

Areal values are calculated for the UK, the four constituent countries, and ten climatological regions. So for temperatures the ten regions each currently contain roughly 40 stations, and for rainfall each region contains about 300 stations.

The areal averages, calculated using all grid-points in each area, effectively smooth out the station errors, so we expect the areal errors to be smaller. Typical point-value errors for monthly-mean temperatures are between 0.3 and 0.6 °C (similar to those in Fig. 2), so errors in areal values will be smaller than this. However individual stations will not be statistically independent of each other, and this is even more true of the grid-points owing to the interpolation. So an error estimate $S/\sqrt{n}$, where $S$ is standard deviation and $n$ is the number of independent grid-points, will require a value of $n$ which is substantially fewer than the actual number of grid-points. Noting the assertion by Parker (2010) that errors in monthly-mean Central England Temperature values are around 0.3 °C, based on measurements from three or four stations, errors for our current monthly UK temperature values, based on station density and geographical coherence, are expected to be of the order of 0.1 °C. However, work is continuing to calculate these, because we wish to determine the total error, and the error due to variations in station density is only a part. Errors would be smaller for seasons and years, but larger for individual regions and counties. Rainfall is less straightforward because of the skew nature of the distribution. Typical point-value error estimates for monthly rainfall are around 10 % of the actual rainfall amount, and larger in magnitude for wet months and in wetter locations. Errors in areal values might therefore be considerably less than this, depending on the station density and the geographical coherence, but again these are likely to be larger in wetter areas, such as western Scotland, than drier areas, such as south-east England.

Error bars on our areal values will improve our assessment of their historical context. For example, we will be able to rank values more realistically, e.g. "January 2010 had a mean temperature across the UK of 0.9±0.2 °C, making it the coldest since January 1987 (0.7±0.2 °C)". New graphical ways of illustrating the time-series of areal values could be used, incorporating the error-bars.
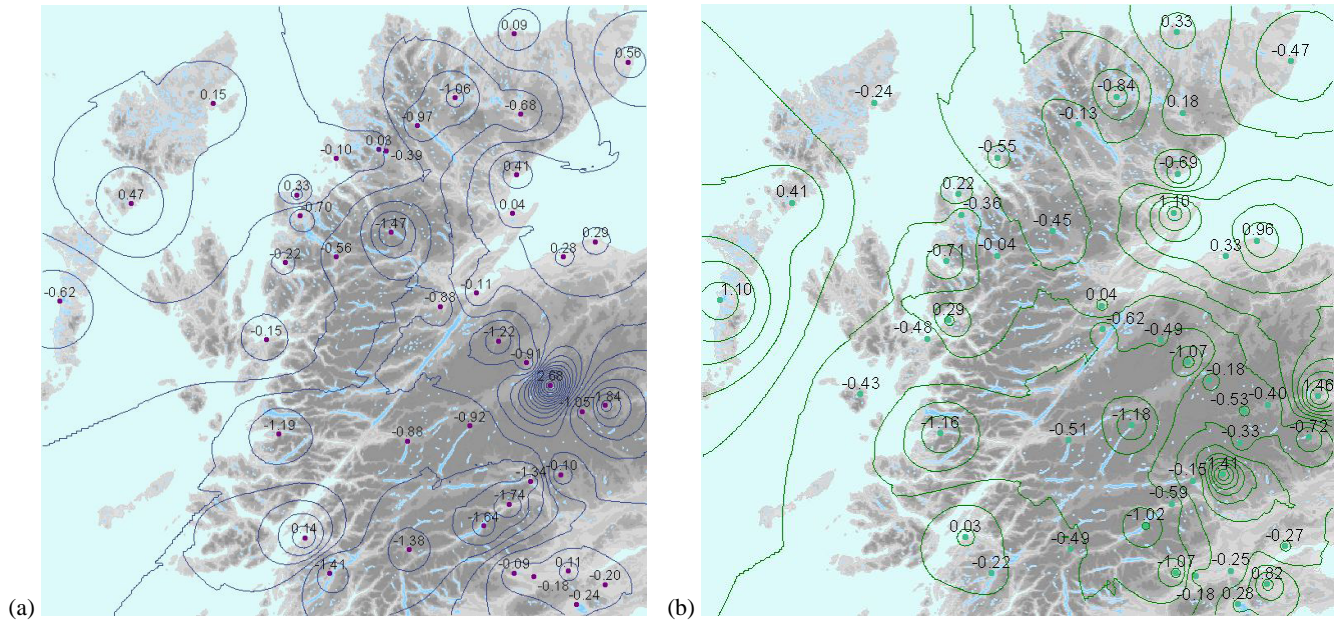
**Figure 4. (a)** Maximum Temperature station biases, October 2007. **(b)** Minimum Temperature station biases, January 2009. Values in °C.

## References

Parker, D. E.: Uncertainties in early Central England Temperatures,
    Int. J. Clim., 30, 1105–1113, 2010.
Perry, M. and Hollis, D.: The generation of monthly gridded
    datasets for a range of climatic variables over the UK, Int. J.
    Clim., 25, 1041–1054, 2005.