# National Climate Information Centre

# Climate Memorandum No 24

# The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK

**by**

**Matthew Perry, Dan Hollis and Margaret Elms**

**National Climate Information Centre**

(June 2009)

**NOTE: This document has not been published. Permission to quote from it or to copy any part should be addressed to**

**Group Head, Climate Research**
**Met Office**
**FitzRoy Road**
**Exeter**
**Devon EX1 3PB**



## Abstract

We have developed gridded datasets of daily rainfall amount and daily maximum, minimum and mean air temperature, covering the UK at 5 km x 5 km resolution, for the period 1960 to 2006.

We used the Met Office archive of daily observations as the main input to the analysis. Geographical effects such as latitude and longitude, altitude, coastal influence, and the effect of urban land use were incorporated through normalisation with respect to monthly 1961-1990 climate normals and, in the case of temperature, a regression model. Station values were then interpolated onto the regular grid using inverse-distance weighted interpolation.

We quality checked both the input station data, and the gridded output. Checking of the output focussed on an analysis of regression residuals for temperature, and on visual inspection of the grids for rainfall.

## Introduction

Gridded datasets of climate variables are required for a wide variety of applications, including verification of climate change models, hydrological modelling, and research into the impact of climate variability and change on ecological systems. The Met Office has already generated gridded datasets on monthly, annual and long-term (30-year) average timescales for a wide range of climatic variables (Perry and Hollis, 2005a, 2005b).

In support of work for UKCP09 (www.ukcip.org), we have developed daily gridded datasets of observed temperature and rainfall. Daily datasets are of particular interest for the analysis of extremes, as well as the investigation of daily relationships with other factors. This memorandum describes the methods used to generate these daily gridded datasets. They cover the UK at 5km x 5km resolution, for rainfall amount, maximum, minimum and mean temperature, for the period 1960 to 2006.

Daily gridded data sets of rainfall amount had previously been developed for the period 1958 to 2003 and made available by the Met Office to individual customers. The new datasets described in this memorandum are of higher quality. They have been extended to cover the period from 1958 and are available to replace the earlier ones. The rainfall data provided for research as part of the UKCP09 capability do not include these daily rainfall grids. Instead, various derived quantities are available, by season and year, as summarised in Appendix 1.

## Data Source

The station data were extracted from the Data Components database, which is a part of the Met Office climate data archive and contains a simplified version of the raw observations generated according to well-defined rules. The data include daily rainfall amount, and daily maximum and minimum air temperature, and the tables contain single values for each day at each station, together with a flag column to indicate whether a value is suspect or estimated, and a column giving the percentage of data which was available for the calculation.

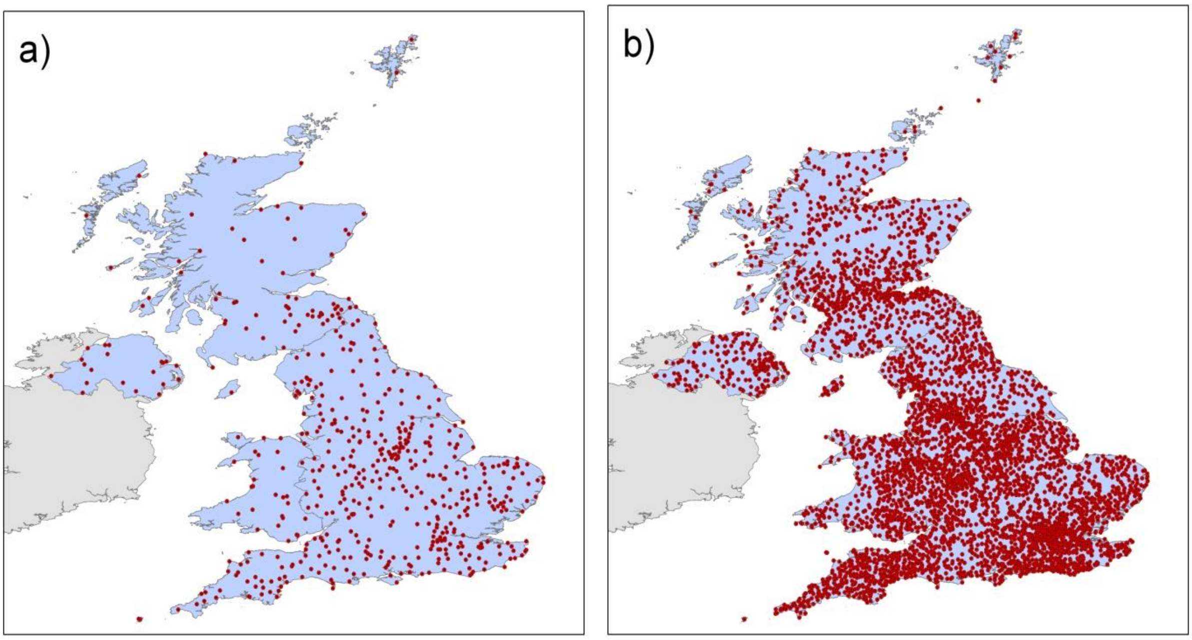
The input station data used related to periods 0900 to 0900 GMT, so the gridded output stored against day D are as follows:

- Maximum temperature between 0900 day D and 0900 Day D+1 (normally occurring on day D)
- Minimum temperature between 0900 day D-1 and 0900 Day D (normally occurring on day D)
- Mean temperature that is the average of the maximum and minimum temperatures
- Rainfall amount between 0900 day D and 0900 day D+1.

Although the data have undergone some quality checking, the extent and effectiveness of this has changed through time since the 1960's. Therefore, we made further checks on the extracted station data in order to identify and exclude data which were clearly in error. The mean and standard deviation across all stations were calculated for each day, and station data which exceeded a set threshold number of standard deviations away from the mean were inspected to see if there was any obvious error, either in relation to other surrounding stations on that day or in relation to the record for that station over surrounding months. The thresholds set were 5 or 6 standard deviations for temperature, and 20 standard deviations above the mean and 1.5 below the mean for rainfall. Sometimes, this led to the discovery of a station record which was wrong for a period of time, for example due to rainfall accumulations, or data being mis-aligned by a day. No attempt was made to correct data and any data found to be in error was excluded.

All available stations were used regardless of record length, in order to make the maximum use of the data. The number of stations used as input to the gridding varies throughout the period, as can be seen in Figures 1 and 2. These variations are due primarily to changes in the size of the observing network.

Figure 1: The station networks for rainfall in a) 1960; b) 1980; c) 2000, and for temperature in d) 1960; e) 1980; f) 2000. Stations included were used for at least half of the days in the respective years.



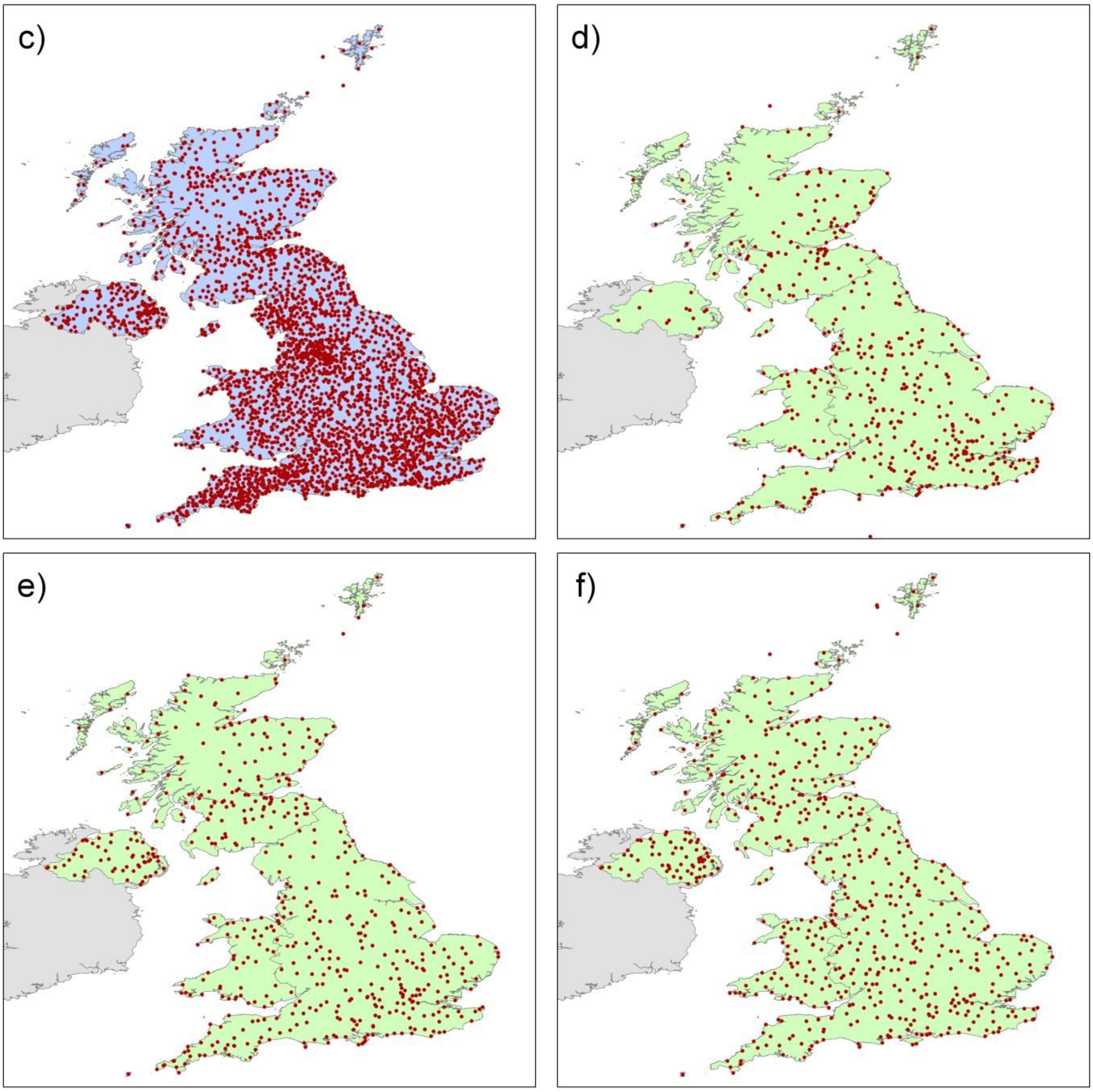
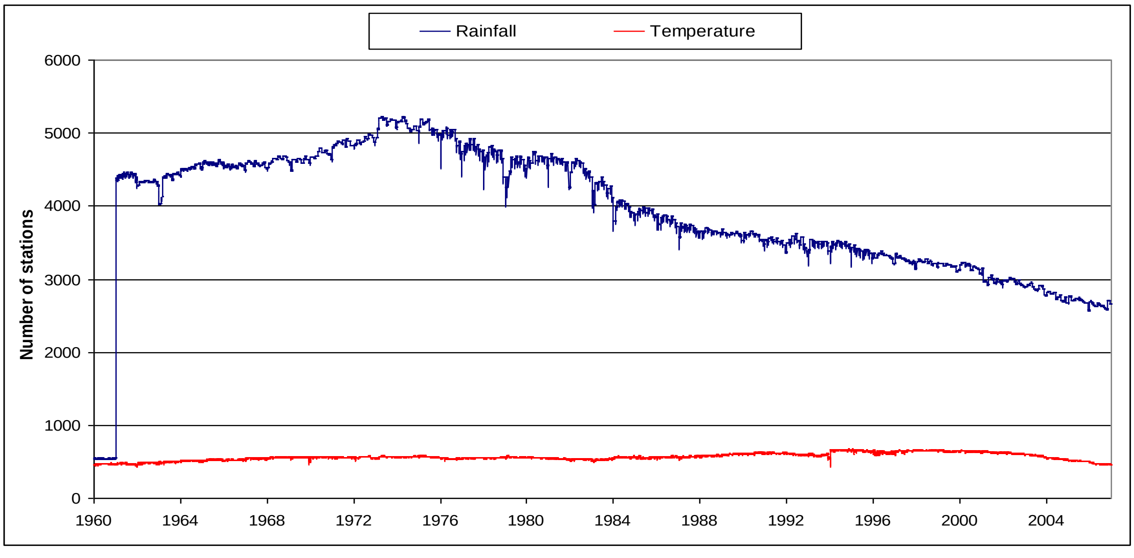



Figure 2: The number of stations used in the gridding analysis for rainfall and temperature from 1960 to 2006.



## Method

The methods used were similar to those used to generate monthly datasets (Perry and Hollis, 2005b). The NCIC climate data analysis software was again used to create the gridded data: this is a customised version of the ArcView Geographical Information System (GIS) software package. The main input was a set of observations made at meteorological stations, irregularly distributed in time and space. The station data were normalised with respect to the monthly 1km x 1km gridded 1961-1990 climate normals described by Perry and Hollis (2005a).

Inverse-distance weighting (IDW) was then used to interpolate the irregular station data to a regular grid. For temperature, the interpolation was done on the residuals from a regression model relating the normalised temperature values to latitude, longitude, altitude, coastal influence and density of urban land use. For rainfall, the interpolation was done on the normalised rainfall values themselves as no regression model was used.

The set-up of the analysis, such as whether to use a regression model, which variables to include in the regression, and the power and radius of the IDW interpolation, were determined by testing on a sample year which was analysed with a set of 10% of the stations omitted for verification. The Root Mean Square Error (RMSE) for each day was calculated from the differences between the observed values at the verification stations and the gridded values at the station locations. The RMSE averaged across all days in the test period was compared for different versions of the analysis. The chosen analysis set-up was that which yielded minimum RMSE.

Table 1 shows the RMSE at verification stations, averaged over each day of the test year, for each of the variables for the chosen analysis set-up. This gives an indication of the quality and accuracy of the grid values. Table 2 shows the chosen analysis set-ups. Latitude and longitude are represented by a cubic cross-polynomial. The coastal influence variables used are the percentage of land within a radius of either 10 km or 30 km. The urban variable used was the proportion of urban land use within a 5 km radius.

Table 1: RMSE at verification stations for each climate variable.

| Variable | Verification RMSE |
|---|---|
| Rainfall amount | 1.23 mm |
| Maximum Temperature (Tmax) | 1.06 deg C |
| Minimum Temperature (Tmin) | 1.27 deg C |
| Mean Temperature | see note |

Note: The mean temperature grids were calculated from 0.5 (Tmax + Tmin)

Table 2: Analysis regression model and interpolation settings used

| Variable | Regression variables | | | | IDW settings | |
|---|---|---|---|---|---|---|
| | Lat / Long | Altitude | Coastal | Urban | Power | Radius |
| Rainfall amount | n/a | n/a | n/a | n/a | 3 | 50 |
| Mean Temperature | X*Y^3 | Yes | %land 30km | No | 2 | 100 |
| Maximum Temperature | X*Y^3 | Yes | %land 30km | No | 2 | 100 |
| Minimum Temperature | X*Y^3 | Yes | %land 10km | 5km | 2 | 100 |

The results for mean temperature are for gridded datasets generated by interpolating station values of daily mean temperature, obtained by averaging the observations of maximum and minimum temperature, where both are available. However, tests showed that accurate mean temperature grids could be generated by averaging the gridded datasets for maximum and minimum temperature and this approach was adopted.

## Quality-checking of Gridded Outputs

**Temperature**. The average value of the coefficient of determination, $R^2$, from the regression of anomaly values was 0.55 for maximum temperature and 0.51 for minimum temperature. The checking for these variables focussed on an analysis of the regression residuals at each station for each day. High regression residuals are often caused by data that are in error. However, they may also be caused by the regression surface being a poor representation of the actual temperature, especially around the edges of the UK, in areas of complex terrain, or on days with unusual temperature patterns, so care needed to be taken before excluding values. Two tests were used to identify suspect values:

- A large absolute value of the regression residual (≥ 5 °C)
- A regression residual that is an outlier based on the intervals between the 5 highest and lowest regression residual values on each day

Station values fulfilling both criteria were inspected in the light of neighbouring station data before deciding whether to exclude them from the analysis.

**Rainfall**. Regression was not used in the analysis, so software was written in ArcView to enable the relatively rapid and efficient visual checking of gridded output for each day. When data “bulls-eyes” or clear inconsistencies were spotted, station values could be plotted in a zoomed view, to enable a decision to be made about whether to exclude a station from the analysis. The time-series array of station data could also be inspected to help with the decision, for example to identify rainfall accumulations.

The gridding was then re-run without the excluded data, and the final version of the output was archived ready for further analysis and the generation of derived products.

## References

Perry MC, Hollis DM. 2005a. The development of a new set of long-term average climate averages for the UK. *International Journal of Climatology* **25:** 1023-1039.

Perry MC, Hollis DM. 2005b. The generation of monthly gridded datasets for a range of climatic variables over the UK. *International Journal of Climatology* **25:** 1041-1054.

## Appendix 1: Rainfall metrics produced for UKCP09

The following have been calculated for 5km x 5km grids for each season and year in the period 1960 to 2006.

- Precipitation total (mm)
- Amount (mm) and date* of the greatest single-day precipitation
- Amount (mm) and date* of the greatest two-day precipitation
- Amount (mm) and date* of the greatest ten-day precipitation
- 1961-1990 90th percentile for single-day precipitation amount
- 1961-1990 99th percentile for single-day precipitation amount
- Number of single-day rainfall events that exceed the 1961-1990 90th percentile
- Number of single-day rainfall events that exceed the 1961-1990 99th percentile
- Proportion (%) of precipitation total from events that exceed the 1961-1990 90th percentile
- Proportion (%) of precipitation total from events that exceed the 1961-1990 99th percentile
- Maximum number of consecutive wet days (>=10mm)
- Maximum number of consecutive dry days (<=0.2mm)

* where the greatest amount occurs more than once in a season or year, the first date of occurrence is provided