The generation of monthly gridded datasets for a range of climatic variables over the United Kingdom

Matthew Perry and Daniel Hollis

Met Office
FitzRoy Road
Exeter
Devon
EX1 3PB

Version 2.0
18 March 2004

The generation of monthly gridded datasets for a range of climatic variables over the United Kingdom

Matthew Perry and Daniel Hollis
Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, United Kingdom
Email: Matthew.Perry@metoffice.com
Tel: +44 (0)1392 884535

Abstract

Monthly or annual 5km x 5km gridded datasets covering the UK are generated for the 1961-2000 period, for 36 climatic parameters. As well as the usual elements of temperature, rainfall, sunshine, cloud, wind speed, and pressure, derived temperature variables such as growing season length, heating degree days, and heat and cold wave durations, and further precipitation variables such as rainfall intensity, maximum consecutive dry days, and days of snow, hail and thunder, are analysed.

The analysis process uses geographical information system (GIS) capabilities to combine multiple regression with inverse-distance weighted interpolation. Geographic and topographic factors such as easting and northing, terrain height and shape, and urban and coastal effects are incorporated either through normalisation with regard to the 1961-90 average climate, or as independent variables in the regression. Local variations are then incorporated through the spatial interpolation of regression residuals.

For each of the climatic parameters, the choice of model is based on verification statistics produced by excluding a random set of stations from the analysis for a selection of months, and comparing the observed values with the estimated values at each point. This gives some insight into the significance, direction, and seasonality of factors affecting different climate elements. It also gives a measure of the accuracy of the method at predicting values between station locations.

The datasets are being used for the verification of climate modelling scenarios and are available via the internet.

Keywords: Climate, United Kingdom, Spatial Interpolation, Regression, Gridded Datasets

1. Introduction

The Met Office has a historical database containing observations of weather elements. These observations come from an irregularly spaced and gradually evolving network of meteorological stations across the United Kingdom.

The aim of this project is to add value to these data by producing a consistent series of climatic statistics which enables comparisons to be made across space and time. In order to do this, methods have been developed to create gridded datasets from the point data.

There is increasing demand for gridded datasets of climate variables from fields such as hydrology, forestry, ecology, agriculture, climate change research, and climate model verification. Consequently there have been numerous attempts made at spatial interpolation, using a variety of methods. Available interpolation methods include inverse-distance weighting, polynomial surfaces, geostatistics e.g. kriging, and smoothing splines. Most authors have found that topographical and other influencing factors need to be incorporated, and this has been done prior to interpolation through regression (Agnew and Palutikof, 2000; Ninyerola et al., 2000) or de-normalisation (Brown and Comrie, 2002; New et al., 1999), or as part of a more complex model, for example co-kriging (Goovaerts, 2000).

Vicente-Serrano et al. (2003) and Jarvis and Stuart (2001) compare several of these methods. Overall, there seems to be no clear preference for any method, and different methods often give similar error statistics. Results depend on the characteristics of the area under study, the data available, and the independent variables used.

This project uses inverse-distance weighted interpolation of regression residuals, and in some cases prior normalisation of the data. These methods have been developed from those described by Lee et al. (2000). The United Kingdom has a wide range of conditions which require terrain shape, coastal and urban effects to be used in a regression model.

There may be potential for improving the results further, for example by including further factors into the regression model. Johansson and Chen (2003) and Kyriakidis et al. (2001) include predictors to model interaction with the lower atmosphere, such as wind and humidity, for precipitation mapping. Daly et al. (2003) incorporate expert knowledge into their regression model. The use of a locally varying (Price et al., 2000) or geographically-weighted regression model (used for precipitation by Brunsdon et al., 2001) also merits further investigation. Agnew and Palutikov (2000) used stepwise regression with a large range of variables, including slope and aspect, maximum and mean elevations within different radii and direction sectors, distance and direction to coast, and proportion of sea within different radii to model temperature and precipitation in the Mediterranean. Slope was chosen for temperature in all seasons, but for precipitation the maximum elevation, usually within 100km in different directions, was more important.

Most authors have only been concerned with temperature and/or precipitation, although New et al. (1999) created global gridded monthly datasets for a range of variables. This project is notable for the range of climate elements tackled: gridded data sets at 5km by 5km resolution over the UK have been produced for 36 monthly or annual climate variables, for the period 1961-2000. This is a total of over 14,000 grids, each comprising 9,700 values across the UK. The start date of 1961 was chosen because there is a significant increase in the availability of digitised data from this point. An extension of the series back beyond 1961 is planned for some key variables.

Twenty-four of the climate variables have been produced with funding from the UK Climate Impact Programme (UKCIP), and are freely available for research purposes via the Met Office web site. The data sets are being used for the verification of climate change scenarios, as well as the creation of areal climatic series against which the recent climate can be compared.

2. Data

The density of the station network varies between elements, from an average over the 40 year period of one station per 59 x 59 km^2 over the UK for pressure, cloud and wind (70 stations), to 29 x 29 km^2 for sunshine (290 stations), 24 x 24 km^2 for hail, thunder, and grass minimum temperature (430 stations), 21 x 21 km^2 for maximum and minimum temperature (540 stations), and 7 x 7 km^2 for rainfall (4400 stations).

The density of the network varies gradually throughout the period 1961-2000, as can be seen in Figure 1. The density of the rainfall network increases to a peak in 1974 before decreasing gradually. For temperature, the density is quite stable, peaking in 1994, and for sunshine the peak is in 1970, followed by a steady decline in station numbers.

There are considerable spatial variations, with comparatively sparse networks in certain areas, especially those which are sparsely populated, e.g. the Scottish Highlands. Figure 2 shows an example of the network of stations used, in this case the temperature network (climate stations reporting daily) for January 2000, and the subset of stations which also recorded pressure (synoptic stations reporting hourly).

The daily observations made at these stations have passed through a rigorous quality control procedure, with substitutions made for poor quality and some missing data. Basic range and consistency checks have been applied to hourly data. These data have then been used to calculate monthly and annual climate averages, totals, and extremes for a range of variables. These statistics have been loaded to an Oracle database of climate statistics and used as the source of input data for the gridding process.

The annual statistics were derived from temperature (T) or rainfall data using the following definitions:

- Growing season length = period (days) bounded by daily T_mean > 5 °C and < 5 °C (after 1st July) for ≥ 6 days.
- Heating Degree Days = $\sum$ 15.5 – daily T_mean for T_mean < 15.5 °C.
- Growing Degree Days = $\sum$ daily T_mean – 5 for T_mean > 5 °C.
- Heat Wave duration = $\sum$ days with daily T_max – 1961-1990 daily normal > 3 °C for ≥ 6 consecutive days.

- Cold Wave duration = $\sum$ days with 1961-1990 daily normal - daily T_min > 3 °C for ≥ 6 consecutive days.
    - Heat Wave and Cold Wave durations are calculated separately for summer and winter half-years.
- Consecutive Dry Days = Maximum number of consecutive days with rainfall < 0.2 mm.
- Rainfall Intensity = Total rainfall on Raindays ≥ 1 mm / Number of Raindays ≥ 1 mm.

For rainfall variables, stations were not used for gridding if they have any missing data in the month, because of the high daily variability of rainfall. For other elements, station months with up to two missing days were used in the analysis.

All available monthly data is used in the analysis in order to make maximum use of the information available and ensure that the most accurate possible representation of the climate can be made for each month. Thus, unlike some studies where a constant set of long-period stations is used (Ninyerola et al. 2000), this study also makes use of data from stations with short and intermittent records. Consequently, the network of stations used changes slightly each month, and the methods used are designed in order to minimise the impact of these changes on the consistency of the datasets through time.

3. Methods

3. 1 Gridding

The production of the monthly and annual grids is carried out using functionality programmed into the ESRI ArcView 3.2 GIS. Code has been written in Avenue (the ArcView scripting language) and Visual Basic that allows the user to take a set of irregularly spaced point data (i.e. station observations) and create a regular grid of values from them.

The method is a two-stage process of multiple regression of the climate variable (which may be normalised first) with geographic factors as the independent variables, followed by interpolation of the model residuals. The regression surface and the interpolated residual surface are then added together to get the final gridded datasets, which are de-normalised if necessary.

Many meteorological elements exhibit dependencies on geographic features, especially topography. Two options are available for reducing or removing geographic effects prior to interpolation, either or both of which can be used.

- normalisation i.e. either subtracting the long-term average (LTA) from the raw values, or dividing the raw values by the LTA (for rainfall and sunshine);
- conversion to regression residuals i.e. creating a model of geographic effects using regression analysis and then subtracting the regression estimate from the raw values.

3. 2 Regression

There are a range of geographic and topographic quantities available to include as independent variables in the regression analysis, as follows:

- easting and northing (to capture spatial variations);
- terrain elevation (to capture altitude effects);
- the mean altitude over a 5km radius centred 10km to the north, east, south and west of the station, or alternatively the mean altitude within a 5km radius of the station (to capture terrain shape effect);
- the percentage of open water within a 5km radius of the station (to capture coastal effects);
- the percentage of urban land use within a 5km radius of the station (to capture urban effects).

Several different methods were tested to represent the shape of the terrain surrounding each point. Slope and aspect and the mean difference in altitude compared with the surrounding area in all directions, generally gave less significant results. The inclusion of altitudes offset in four directions allows the aspect and exposure of the location to be taken into account, and the parameters can vary according to prevailing conditions. Different radii were considered, but the 5km radius at 10km distance gave the best results in most cases. Different radii were also considered for the sea and urban factors.

3. 3 Interpolation

Two versions of inverse-distance weighted (IDW) interpolation were used to interpolate the regression residuals onto a regular grid;

- standard IDW averaging;
- custom IDW averaging (uses a modified weighting function with an option not to go to infinity when station and grid point coincide, and includes an adjustment for variations in station density);

In both cases, the value for each grid cell is calculated as a weighted average of values from surrounding stations. The standard and custom versions of the IDW method use different methods to determine the radius within which stations are used. The standard version has a minimum sample of 12 points, and the radius parameter r is extended to meet this requirement if necessary. The custom version takes all stations within the radius r;

$$(1) \qquad F(x,y) = \sum_{i=1}^{n} w_i f_i \Big/ \sum_{i=1}^{n} w_i$$

where n is the number of stations in the set, f_i are the data values at the stations, and w_i are the weight functions assigned to each station, where;

$$(2) \qquad w_i = \frac{u_i}{d_i^p}$$

or, if infinite weights are not selected ;

$$(3) \qquad w_i = \frac{u_i}{1 + \left(d_i / m\right)^p}$$

where p is the power parameter, d_i is the distance from the station to the interpolation grid centre, and m is a scaling parameter, and u_i is a density weighting parameter.

In each case the user can control the parameters of the interpolation e.g. the power of the IDW relationship, the radius, and the scaling parameter, as well as whether to include the density weighting function, and whether to allow infinite weights. If the density function is not used then $u_i = 1$, and otherwise it is obtained from a density surface created using an Avenue function which adds together kernel density functions (extending to r km) fitted over each station.

$$(4) \qquad u_i = \frac{1}{1 + \sum_{i=1}^{n} k_i}$$

where the k_i are the values of the kernel function from each station.

3.4 Optimisation

In order to select the regression model and interpolation method to be used for each climate variable, several different methods were tested. The initial hypothesis was a model based on prior knowledge of the topographical and geographical factors affecting each climate variable, and the statistical distribution of the variable. Subsequent test runs were altered by adding or removing regression variables, changing the interpolation settings, or changing the input data (actual or anomaly).

The optimisation analyses were carried out with a 10% random sample of stations excluded from the analysis as if they were missing. Estimated values for these stations were then interpolated from the final grid, and compared with the corresponding observed values. Each test was made using twelve months in the period 1961-2000, a random sample including one for each month of the year. Annual statistics used six randomly selected years.

Various statistics were automatically calculated, both for the set of withheld stations, and for the analysis stations. These were used to compare the analysis profiles, with particular importance attached to the root mean square error (RMSE) of the verification stations (the error of the predicted value at the station location from the observed value at the station). The Mean Error (ME) was used to detect any bias in the estimates.

4. Results

4. 1 Choice of Methods

Long Term Average grids for the 1961-1990 period were available for most variables (the methods used to create these are described by Hollis and Perry (submitted for publication), and it was generally found to be preferable to normalise the data (i.e. create anomalies from average) prior to its analysis, in order to generate a smoother surface to analyse. This was especially the case where the variable showed complex geographical patterns. For some variables, long term average grids were not yet available, and for others normalisation was not necessary: in these cases, actuals were analysed.

Regression was always used, even when the data had first been normalised. This is because there are usually broad spatial trends even in the normalised monthly data which can be removed by regression on latitude and longitude. There may also be monthly variations in the strength and pattern of other geographic effects. Thus the quantity which was interpolated was residuals obtained from a regression analysis of either actual or normalised data.

4.2 Regression Model

Cross-polynomial terms in easting and northing were used for all variables to model the pattern of spatial variation over the UK. For elements with a sparse network of stations, a low power of 1 (3 parameters) or 2 (5 parameters) was used, to avoid unrealistic extrapolation of the surface. Terms up to a power of 3 (10 parameters) were used when greater flexibility in the spatial trend surface was required, e.g. for rainfall and temperature.

When normalised data was used, any additional terms were often of little significance, because most of the topographical variations were being described through the climate average. It was difficult to determine whether topographical effects introduced to the temperature anomalies were real effects, or were simply adding noise to the result. It was decided for the most part (rainfall, days of rain, pressure, thunder and maximum and minimum temperature) to use only the spatial terms. For 'days of frost' however, except for the summer months, altitude, coastal and urban effects were found to reduce the grid error at test stations by a small amount (about 0.1 days). For sunshine, the coastal effect was used for the summer months.

The elements for which actual data was used have more intuitive regression models. Growing season length, and growing degree days have a strong altitude dependency, and quadratic altitude terms were used. Heat wave and cold wave durations have especially strong coastal effects.

Topographic effects on cloud cover were difficult to model due to the lack of high-altitude stations. However, altitude was shown to make a significant improvement to the model in summer and autumn months, and the coastal effect was important except for in the summer months.

Altitude and terrain shape variables were obviously very important for days of snow lying and falling, as well as for rainfall intensity and the greatest annual 5-day rainfall total.

The independent variables used in the regression model for each element, and an indication of whether the data was normalised first, can be seen in Table 1.

4.3 Interpolation Methods

The power of the weighting function with distance has a significant effect on the result: decreasing the power has the effect of smoothing the interpolation surface around stations with high residuals, while increasing the power brings the surface closer to the actual station values. Inverse distance cubed was used as the weighting function for most variables, as this gave a good degree of closeness to the observed data, while taking surrounding values into account. Maximum, minimum, mean, and grass minimum temperature were analysed using a power of 2, as a means of reducing localised variations that were unable to be modelled. For these variables, as well as for cloud cover, the use of the squared weighting gave a slightly lower grid RMSE for the verification stations, while increasing the grid RMSE for the analysis stations, compared with a power of 3.

The radius was set at between 100 and 150 km for most variables. Precipitation variables, with their higher spatial variability, were analysed using a lower radius of between 50 and 75 km. The test grid RMSE was fairly insensitive to small changes in radius.

The custom IDW method gave lower test grid RMSE values than the standard method for most temperature variables, as well as wind speed, pressure and thunder, while the standard method performed better for other elements, including precipitation and sunshine. The density weighting function was used with the custom method, but did not have a significant impact on the results. Infinite weighting at station locations was only removed for maximum, minimum and mean temperature; for other variables, the surface was forced to pass through the station values (depending on how close the station location is to the grid centre).

4.4 Grid Accuracy

The verification statistics also give a good indication of the accuracy of the gridding process at estimating values of each climate variable at locations where we do not have observed data. Table 2 shows the average value of the RMS Error of the grid estimates for test stations, using the best method for each variable. It also shows variation across the test period, showing especially the strong seasonality present in some variables.

The RMSE has also been expressed as a percentage of the range of grid values over the UK, averaged across the test months/years. Using this measure, the lowest errors are for MSL Pressure, Heating Degree Days, and Mean Temperature. These are all fairly smoothly varying quantities, and the analysis is able to closely predict values at points between stations. Precipitation also has a low percentage error of 3.6, although this may be partly due to the highly skewed distribution of values inflating the range.

Heat wave and cold wave duration, and days of hail, thunder, and ground frost were especially difficult to model, with a high degree of spatial variability which was not covered by the geographical or topographical variables used in the regression.

4.5 Regression Results

Table 1 shows the data and regression variables used in the analysis of each climatic parameter (in addition to the spatial terms, easting and northing, which are used for all elements). It also shows the average R^2 values for each climatic parameter over the 40-year period, indicating how much of the variation in the actual or anomaly data is explained by the variables used in the regression analysis. This will generally be lower when anomaly data is used, because normalisation by the long term average removes much of the explainable variation.

The highest R^2 values were achieved for mean sea level pressure, daytime maximum and minimum temperatures, and heating and growing degree days, with over 90% of the variation explained by the model. These elements also had low verification RMSE as a percentage of the range, and are smoothly varying or have strong geographic dependencies. The lowest R^2 values were for ground frost and thunder (from anomaly data), and hail and cold wave duration (from actual data). Ground frost data is subject to inconsistencies caused by different measuring periods (0900 to 0900 or dusk to 0900) and soil type. Along with cold wave duration, it is also a 'days exceeding a threshold' variable, which is subject to increased random variations. Days of hail and thunder are also difficult to model as they occur in small quantities, and without a strong geographical pattern.

5. Output

Inspection of regression residuals from the analysis led to some dubious data values being investigated further through inspection of maps of the gridded datasets, and an investigation of the daily data from which the monthly values were derived. Erroneous data was removed and the affected months / variables were re-run. A sample of the final maps was inspected to check that the patterns and levels of the variables fitted with our knowledge of the UK climate. The datasets were converted to ASCII grid format and those for 24 variables were made available on the internet. Figure 3 shows some examples of the maps produced. The first map shows the number of days of hail for March 1995 (a month when hail was especially common) and is an example of a variable which was difficult to model; there are apparently random variations over England and Wales, with very high levels of hail in the western and northern fringes of Scotland, but low levels over the rest of Scotland. The map of rainfall intensity for 1978 shows a similar pattern to that of rainfall, with a strong orographic effect. Variations in annual Heating Degree Days (map for the year 2000 shown) were very well modelled, and the strong altitude effect as well as the urban effect for London and Manchester can be seen. The strong coastal effect can be seen in the temperature range map for August 1995 (this month had large temperature ranges, especially in central and southern England).

For each month and variable, the grid values within each of ten climate districts were averaged, as well as for the countries of the UK, and the UK as a whole. These areal averages provide a series against which the most recent month's weather can be put into context. Looked at over the period from 1961 onwards, they can also provide an interesting summary of temporal patterns in the climate, including seasonality, trends, and random variability. Figures 4 and 5 show a time series of growing season length and rainfall intensity by country. Linear regression models for each of the ten climate districts show increasing trends in growing season length and rainfall intensity, with increases in growing season length over the 40-year period varying from 0.4 days per year (N Scotland) to 1.0 days per year (SE England), and increases in rainfall intensity between 0.002 mm per rainday per year (NW England and N Wales) and 0.026 per rainday per year (W Scotland). The graphs show a high degree of year-to-year variability, especially for growing season length, although there has been a fairly steady increase since the particularly short growing season of 1979.

The range of variables gives a fairly complete picture of the climate. Patterns of variables which are not often analysed, such as cloud cover, humidity, hail, and temperature range can be seen. Temperature indices such as growing season length and growing and heating degree days give objective assessments of the impacts of climate change on areas such as agriculture, ecology and energy use. Changes in the distribution of rainfall amounts, extremes, and intensity can be investigated with impacts for flood risk. The results of any such analyses will not be presented in this paper, but further work will be done to analyse the datasets produced.

6. Conclusions

Gridded datasets can be created from station data and grids of explanatory variables, for a wide range of climate variables, using the described methodology which uses GIS capabilities to combine multiple regression and inverse-distance weighted interpolation.

The values of climate variables at locations between observing stations can be estimated to a good degree of accuracy, producing detailed and representative maps of the UK climate. Spatial and temporal variability and trends in the UK climate over the last 40 years can be investigated using the results of the gridding, which provide a consistent set of data.

The accuracy varies, however, and is dependent on the nature of the variable, and the density and representativity of the station network. Errors will be highest in areas of sparse station coverage, particularly the highlands of Scotland which are also areas of complex mountainous terrain. The average proportion of variance in the data explained by the regression models varies between 21% (normalised days of ground frost) and 95% (mean sea level pressure). The verification RMSE as a percentage of the range of values is also a useful guide to the level of error, and it ranges from 2.7% (mean sea level pressure) to 19.3% (cold wave duration). It is below 10% for 29 of the 36 variables.

Localised effects on the climate such as frost hollows, and effects caused by soil type and forests, have not been taken into account. Future work may include investigating the impact of incorporating these and other effects into the model, through the development and use of further gridded independent variables. The use of a local or geographically-weighted regression model could improve the result, and would also merit further investigation, as would the use of stepwise regression for choosing independent variables.

Acknowledgements

This work was partly funded by the UK Climate Impact Programme (UKCIP). The authors would like to thank Malcolm Lee, Eddie Spackman, and Richard Tabony for their input into the development of methods and comments on the draft paper.

References

Agnew MD, Palutikof JP. 2000. GIS-based construction of baseline climatologies for the Mediterranean using terrain variables. *Climate Research* **14**: 115-127.

Brown DP, Comrie AC. 2002. Spatial modelling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research* **22**: 115-128.

Brunsdon C, McClatchey J, Unwin DJ. 2001. Spatial variations in the average rainfall-altitude relationship in Great Britain: and approach using geographically weighted regression. *International Journal of Climatology* **21**: 455-466.

Daly C, Helmer EH, Quinones M. 2003. Mapping the climate of Puerto Rico, Vieques and Culebra. *International Journal of Climatology* **23**: 1359-1381.

Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228**: 113-129.

Jarvis CH, Stuart N. 2001. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: the interaction between number of guiding variables and the type of interpolation method. *Journal of Applied Meteorology* **40**: 1075-1084.

Kyriakidis PC, Kim J, Miller NL. 2001. Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *Journal of Applied Meteorology* **40**: 1855-1877.

Lee MJ, Hollis DM, Spackman E. 2000. From raw data to the internet – producing quality climatological services. Proceedings of the 3rd European Conference on Applied Climatology, Pisa.

New M, Hulme M, Jones P. 1999. Representing twentieth century space-time variability. Part II: Development of a 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate* **13**: 2217-2238.

Ninyerola M, Pons X, Roure JM. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology* **20**: 1823-1841.

Price DT, McKenney DW, Nalder IA, Hutchinson MF, Kesteven JT. 2000. A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology* **101**: 81-94.

Vicente-Serrano SM, Saz-Sanchez MA, Cuadrat JM. 2003. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research* **24**: 161-180.

Figure 1: The changing density of the station network from 1961 to 2001, for precipitation, temperature and sunshine.

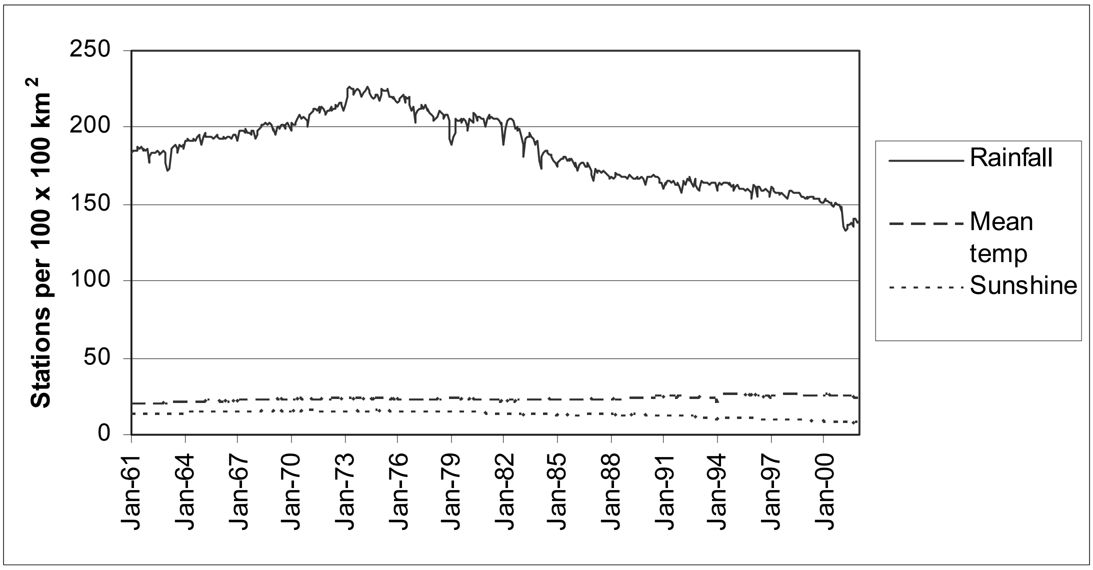

Figure 2: Spatial distribution of Temperature and Pressure stations, January 2000

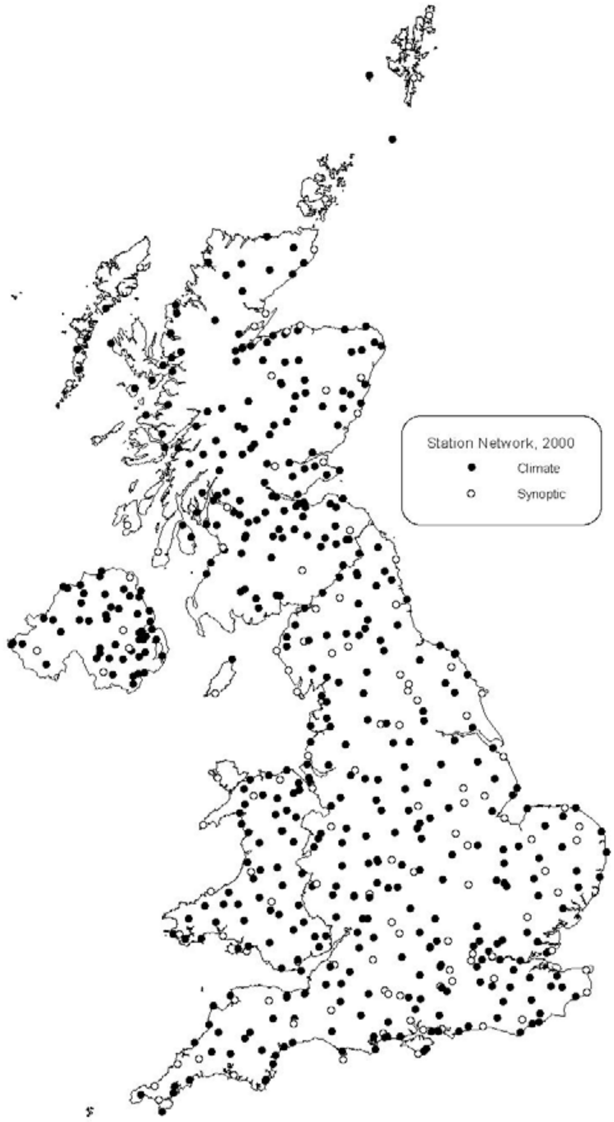

Figure 3: Examples of maps produced; a) Days of Hail for March 1995, b) Rainfall Intensity (mm per rainday) for 1978, c) Mean Diurnal Temperature Range (°C) for August 1995, d) Heating Degree Days for 2000.

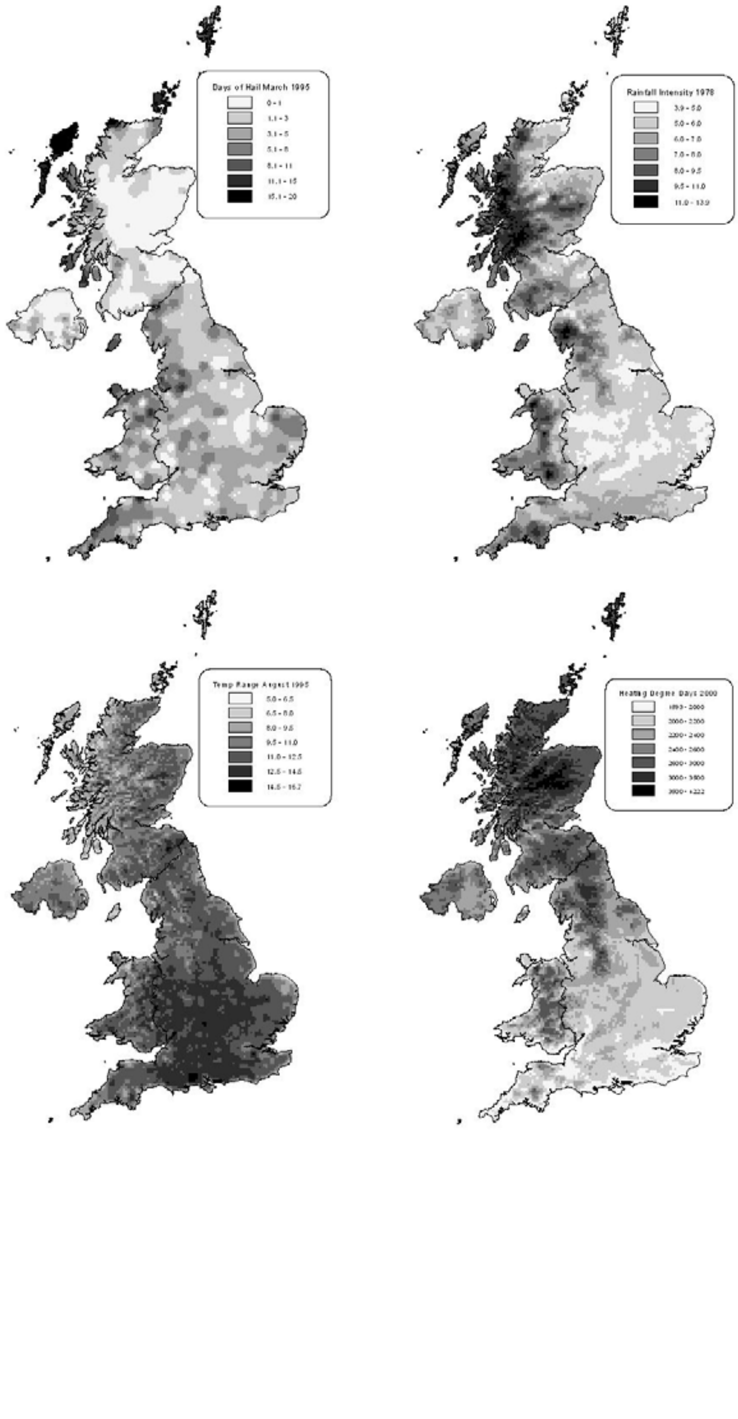

Figure 4: Country averages of Growing Season Length; time series for 1961-2000 period.

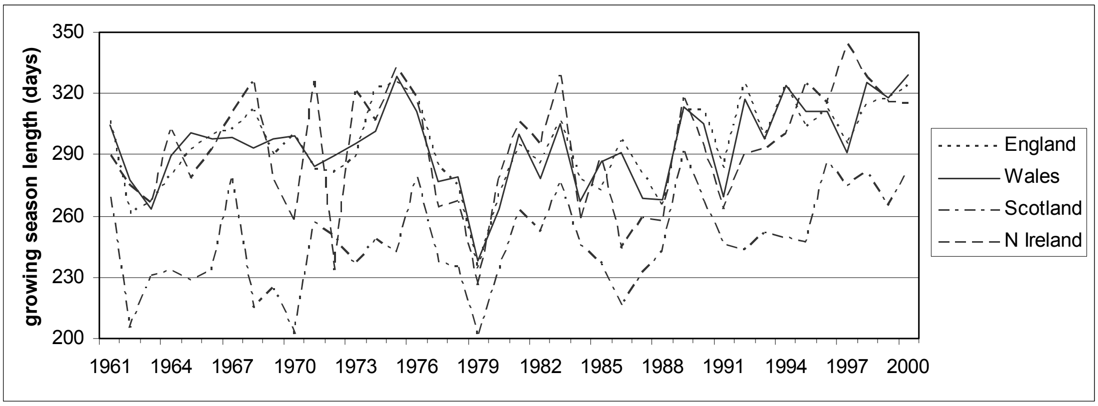

Figure 5: Country averages of Rainfall Intensity; time series for 1961-2000 period.

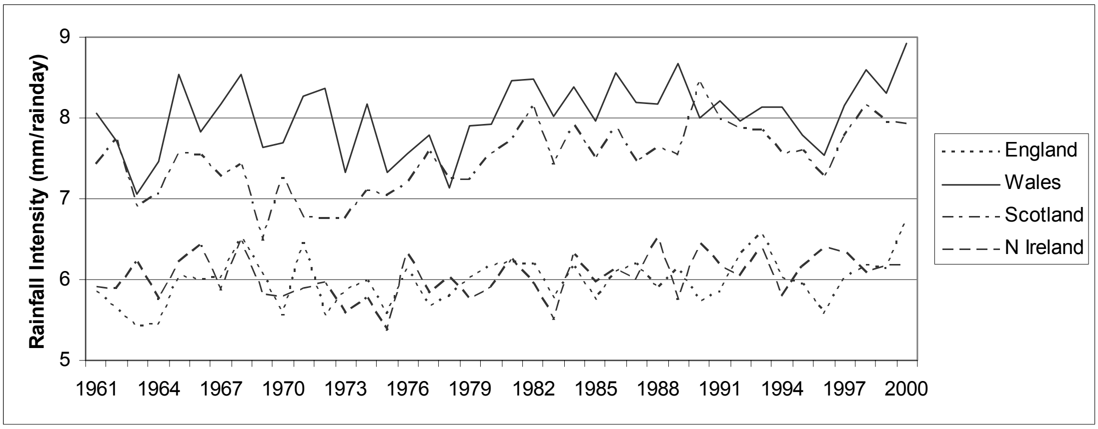

Table 1: Regression Variables used, and average R^2 values.

Element	1961-00 Average R^2	Data analysed	Altitude	Terrain Shape	Coast	Urban
Monthly						
Minimum Temperature	0.39	Anomaly				
Mean Temperature	0.57	Anomaly	*	*		
Maximum Temperature	0.53	Anomaly				
Days of Ground Frost	0.21	Anomaly	*	*	*	*
Days of Air Frost	0.31	Anomaly	*	*	*	*
Sunshine	0.44	Anomaly			*	
Wind Speed	0.82	Actual	*	*	*	
Precipitation	0.44	Anomaly				
Days of rain ≥ 0.2 mm	0.57	Actual	*	*		
Days of rain ≥ 1mm	0.42	Anomaly				
Days of heavy rain ≥ 10mm	0.40	Actual	*	*		
Hours of relative humidity > 95%	0.46	Actual	*		*	
Vapour Pressure	0.81	Actual	*		*	
Mean Sea Level Pressure	0.95	Actual				
Cloud Cover	0.58	Actual			*	*
Days of Snow Falling	0.36	Anomaly	*	*		
Days of Snow Lying	0.48	Actual	*	*		
Daytime maximum temperature	0.93	Actual	*	*	*	*
Daytime minimum temperature	0.92	Actual	*	*	*	*
Night-time maximum temperature	0.83	Actual	*	*	*	*
Night-time minimum temperature	0.87	Actual	*	*	*	*
Concrete minimum temperature	0.74	Actual	*		*	
Grass minimum temperature	0.28	Anomaly	*		*	
Days of Concrete Frost	0.47	Actual	*		*	
Mean Diurnal Temp Range	0.64	Actual	*	*	*	*
Days of Thunder	0.22	Anomaly				
Days of Hail	0.24	Actual				
Annual						
Extreme Temperature Range	0.66	Actual	*	*	*	
Growing Season Length	0.68	Actual	*		*	
Heat wave Duration	0.41	Actual			*	*
Cold wave Duration	0.25	Actual	*	*	*	
Heating Degree Days	0.92	Actual	*	*	*	*
Growing Degree Days	0.93	Actual	*	*	*	*
Greatest 5-day Rainfall	0.46	Actual	*	*	*	
Rainfall Intensity	0.54	Actual	*	*		
Consecutive Dry Days	0.44	Actual	*		*	

Table 2: Error Statistics from the Verification Stations

Element	RMS Error at Verification Stations	Variation across test period			Average range of grid values over UK	RMSE as a percentage of the range of values
		Standard Deviation	Min	Max		
Monthly (12 test months)						
Mean Daily Minimum Temperature (°C)	0.45	0.09	0.3	0.6	9.4	4.8
Mean Temperature (°C)	0.36	0.04	0.3	0.45	10.4	3.5
Mean Daily Maximum Temperature (°C)	0.66	0.12	0.38	0.79	12.2	5.4
Days of Ground Frost	2.24	0.78	0.7	3.4	17.4	12.9
Days of Air Frost	1.4	1.12	0	3.2	15.2	9.2
Sunshine (hours per day)	0.33	0.1	0.2	0.54	4.7	7.0
Mean Wind Speed (knots)	1.41	0.44	1.0	2.2	25.7	5.5
Total Precipitation (mm)	16	4.9	11	27	443	3.6
Days of Rain ≥ 0.2 mm	1.7	0.3	1.2	2.1	20	8.5
Days of Rain >1 mm	1.3	0.15	1.1	1.5	21.2	6.1
Days with Heavy Rain (>10 mm)	0.78	0.22	0.3	1.2	12.5	6.2
Hours of Relative Humidity > 95%	48	22	17	98	536	9.0
Vapour Pressure (HPa)	0.26	0.13	0.08	0.53	5.2	5.0
Mean Sea Level Pressure (HPa)	0.27	0.14	0.16	0.67	9.9	2.7
Cloud Cover (%)	2.69	0.86	1.6	4.6	40.7	6.6
Days of Snow Falling (not inc. Jul – Aug)	1.35	1.3	0	3.9	13.8	9.8
Days of Snow Lying (not inc. Jun – Sep)	1.1	1.1	0	2.7	15.3	7.2
Daytime maximum temperature (°C)	0.47	0.15	0.26	0.7	12.0	3.9
Daytime minimum temperature (°C)	0.47	0.14	0.31	0.79	11.2	4.2
Night-time maximum temperature (°C)	0.75	0.35	0.33	1.57	10.9	6.9
Night-time minimum temperature (°C)	0.61	0.19	0.37	0.92	10.3	5.9
Concrete minimum temperature (°C)	0.65	0.19	0.43	1.02	10.7	6.1
Grass minimum temperature (°C)	0.85	0.17	0.46	1.05	9.8	8.7
Days of Concrete Frost (not inc. July)	1.8	1.3	0.3	3.6	16.1	11.2
Mean Diurnal Temp Range (°C)	0.66	0.17	0.42	0.99	6.9	9.6
Days of Thunder	0.48	0.41	0.08	1.43	4.0	12.0
Days of Hail	1.3	0.9	0.2	3.5	7.6	17.1
Annual (6 test years)						
Extreme Temperature Range (°C)	2.15	0.51	1.6	2.7	23.6	9.1
Growing Season Length (days)	19.5	6.2	11	28	270	7.2
Heatwave Duration (days) [4 per half-year]	3.6	1.4	1.8	5.9	29.1	12.4
Coldwave Duration (days) [4 per half-year]	2.9	1.7	0	5.8	15.0	19.3
Heating Degree Days	106	16	80	125	3157	3.4
Growing Degree Days	100	10	87	116	2108	4.7
Greatest 5-day Rainfall (mm)	12.8	2.7	9.2	16.1	250	5.1
Rainfall Intensity (mm per rain day)	0.53	0.11	0.44	0.75	12.7	4.2
Consecutive Dry Days	3	1.2	1.4	5	29	10.3