Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain

Mikhail A. Semenov*, Roger J. Brooks**

IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol BS41 9AF, United Kingdom

ABSTRACT: The paper describes a method for the spatial interpolation of the site-specific LARS-WG stochastic weather generator to produce 'realistic' daily weather data for the gaps between observed sites. One of the uses of LARS-WG has been site-scale climate change impact assessments. However, such assessments are often applied across regions and so there is a need for an interpolation method to provide input daily weather at many sites or grid-boxes where observed weather data is not available. The interpolation method devised combines the local interpolation of the weather generator parameters from observed sites near the unobserved location with the use of globally interpolated monthly mean statistics for a large number of sites. Thin plate smoothing splines with elevation as an independent variable were used for the global interpolation of mean monthly rainfall and temperature. The data sets used allow daily weather to be generated for any location in Great Britain and the methodology was tested at 3 locations with different local characteristics. The interpolation method showed a good performance at the 3 sites when compared to the observed data, the main differences occurring when the spline method was unable to reproduce closely the observed mean values. The limitations of the interpolation method, its applicability to other regions and its potential use in climate change and other studies are also discussed.

KEY WORDS: Weather generator · Spatial interpolation · Climate modelling

1. INTRODUCTION

A stochastic weather generator is a numerical model which produces synthetic daily time series of a suite of climate variables, such as precipitation, temperature and solar radiation (Richardson 1981, Richardson & Wright 1984, Racsko et al. 1991). Weather generators are site-specific models that require calibration against observed daily weather data for the site. They are mainly used to produce long weather time series in order to assess risk in hydrological, agricultural and other applications since the lengths of observed time series are often insufficient to allow a good estimation of the probability of extreme impact events. Recently,

weather generators have also been used in climate change studies to produce daily site-specific scenarios of future climate (Wilks 1992, Mearns et al. 1997, Semenov & Barrow 1997). This is achieved by applying the predicted changes in climate means and variability obtained from global climate models to the weather generator parameters.

Impact assessments are now often made on high resolution grids or at multiple sites across a region where observed weather records are not available (Harrison et al. 1995). Several interpolation techniques, such as kriging (Phillips et al. 1992), thin plate smoothing splines (Hutchinson 1995a) or the precipitation-elevation regression on independent slopes model (PRISM) (Daly et al. 1994), have been developed to interpolate the monthly means of climate variables, with the emphasis on the interpolation of rainfall in mountainous areas. Rainfall has the greatest variability in time and space of the climatic variables and so makes the greatest demand on such analysis.

*E-mail: mikhail.semenov@bbsrc.ac.uk
**Present address: Department of Management Science, The Management School, University of Lancaster, Lancaster LA1 4YX, United Kingdom

However, many impact models require daily weather data and so a different approach is required. Rather than interpolating the climate variables directly (see, for example, Running et al. 1987), the parameters of a weather generator for each of the observed sites can be interpolated, with the resulting parameters being used by the weather generator to produce synthetic daily data for the unobserved locations. This has been done previously for WGEN in the USA (Richardson & Wright 1984, Hanson et al. 1994), with each individual parameter except for precipitation being spatially interpolated across the USA. For precipitation, parameters from the nearest station or the arithmetic average of parameters from stations within 100 miles (~161 km) were used. However, no relationship between precipitation and elevation was considered.

The aim of this paper is to develop and test a method for the spatial interpolation of the LARS-WG stochastic weather generator (Semenov et al. 1998). LARS-WG has been used in European studies of the effect of climate change on agriculture at the site scale (Harrison et al. 1995). It uses semi-empirical distributions for many of the weather variables rather than pre-defined distributions, which gives it greater flexibility in reproducing different climates. A recent study showed that LARS-WG performs well in diverse climates around the world (Semenov et al. 1998). However, as a consequence of its structure, the number of parameters of LARS-WG is much larger than for WGEN, and there are also several interdependencies between these parameters. Therefore, a direct spatial interpolation of each individual parameter is not a feasible option for LARS-WG, particularly as sites with long records of daily data are often sparsely situated. The paper describes the method for a spatial interpolation of LARS-WG in Great Britain which combines the local interpolation of LARS-WG parameters using neighbouring sites to the target site with a further adjustment based on the global interpolation of monthly weather averages available from an extensive data set covering Great Britain.

2. DESCRIPTION OF LARS-WG STOCHASTIC WEATHER GENERATOR

LARS-WG is based on the series weather generator described in Racsko et al. (1991) with a detailed description being given in Semenov et al. (1998). LARS-WG produces synthetic daily time series of maximum and minimum temperature, precipitation and solar radiation. The weather generator uses input observed daily weather for a given site to determine parameters specifying probability distributions for weather variables as well as correlations between the variables.

The generation procedure to produce synthetic weather data is then based on selecting values from the appropriate distributions using a pseudo-random number generator.

The weather generator distinguishes dry and wet days depending on whether the precipitation is greater than zero. Precipitation is modelled using semi-empirical probability distributions for each month for the lengths of series of wet and dry days and for the amount of precipitation on a wet day. A semi-empirical distribution $Emp = \{e_0, e_i, h_i, i = 1, …, 10\}$ is a histogram with 10 intervals, $[e_{i-1}, e_i)$, where $e_{i-1} < e_i$ and h_i denotes the number of events from the observed data in the ith interval. In the case of precipitation, $e_0 = 0$. The histogram has the effect of slightly smoothing the exact distribution of the empirical values. Since there are typically many small values but also a few large values for these distributions, the interval size increases as i increases.

Minimum temperature, maximum temperature and radiation are related to the amount of cloud cover, and so LARS-WG uses separate wet and dry day distributions for each of these variables. The normal distribution is used for the temperature variables with the mean and standard deviation varying daily according to finite Fourier series of order 3. Time auto-correlations used for minimum and maximum temperature are constant through the year for the particular site and the cross-correlation of the standardised residuals from the daily mean is pre-set for all sites at 0.6. Semi-empirical distributions with equal interval size are used for solar radiation (for more details, see Semenov et al. 1998).

For a given set of parameters, LARS-WG produces synthetic data one day at a time by first determining the precipitation status of the day. The data consists of alternate wet and dry series and, when the end of one series is reached, the length of the next is chosen from the wet or dry series semi-empirical distribution for that month. For a wet day, the amount of precipitation is taken from the precipitation distribution for the month, and the temperature and radiation values are taken from the wet day distributions and correlation coefficients applied. The dry day distributions are applied on dry days. Values from the semi-empirical distributions are chosen by first selecting one of the intervals (using the proportion of events in each interval as the selection probability), and then selecting a value within that interval from the uniform distribution.

The use of semi-empirical distributions gives flexibility to the generator, allowing it to model a wide variety of distributions. Analysis of observed data from many sites around the world has shown that the shapes of most of the variables can vary considerably (Semenov et al. 1998). Standard distributions were only consid-

ered to be satisfactory for temperature, of which the residuals are modelled by the normal distribution and the annual variation in mean and standard deviation by Fourier series. The weather generator has been tested for 18 sites in the USA, Europe and Asia chosen to represent a range of climates and was able to reproduce most of the characteristics of the observed data well at each site (Semenov et al. 1998).

3. SPATIAL INTERPOLATION PROCEDURE

3.1. Data available and outline of procedure

Two data sets of observed weather were available. The first consisted of 138 sites in Great Britain with daily values of minimum and maximum temperature, precipitation and radiation or sunshine hours over relatively long periods of between 20 and 40 yr. The second data set contained monthly means of precipitation for 2376 stations and of minimum and maximum temperature for 623 stations for 1961–1990, supplied by the U.K. Met. Office through the Climate Impacts LINK project. The coverage of this latter data set for precipitation, particularly, is extremely good throughout Great Britain.

The spatial interpolation procedure combines global and local interpolation. A certain regularity in the spatial behaviour of monthly means of weather variables can be expected but this is unlikely to be the case for every single parameter of the weather generator. For example, the shape of the distribution of daily precipitation or solar radiation is the result of many factors, of which the local weather circulation patterns can be very important. The global interpolation of each individual parameter (11 for the interval end points and 10 for the frequencies) of the semi-empirical distributions could produce false local effects, particularly as there are few stations with sufficient daily data. It is therefore more appropriate to interpolate the parameters locally with an imposed interpolation method.

Similarity in the nature of the distributions of the weather variables for nearby sites is expected since the sites will normally be subject to the same basic type of weather on each day. However, systematic differences can occur particularly if the sites are at significantly different elevations, with precipitation tending to increase and temperature tending to decrease with elevation. It is difficult to adjust the precipitation parameters of the neighbouring sites for the difference in elevation because the relationship varies depending on the local conditions (Daly et al. 1994).

The interpolation procedure devised consists of an initial local interpolation in which the weighted average of each of the weather generator parameters for 3 neighbouring sites is calculated. The distributions for precipitation and temperature of the target site are then adjusted in a way that accounts for the effects of site elevation. The monthly average precipitation and temperature data are globally interpolated by a thin plate spline function using elevation as an independent variable in addition to the geographical co-ordinates (Hutchinson 1995b). The parameters for precipitation and temperature at the target site are adjusted based on the mean values predicted by the spline function. The resulting parameter file can be used by LARS-WG to generate synthetic weather data series of any length for the target site. The next sections describe the interpolation procedures in detail.

3.2. Local interpolation

This step consists of the local interpolation of all the LARS-WG parameters, namely, the semi-empirical distributions for wet and dry series, precipitation and solar radiation, the Fourier coefficients for temperature and the temperature and radiation correlation coefficients. The procedure devised is described for the general interpolation of N sites with weighting coefficients $w_1, w_2, …, w_N$ where

$$\sum_{i=1}^{N} w_i = 1 \qquad (1)$$

The weightings should reflect the likely relative strength of the relationships between each site and the target site. They will therefore normally depend on the distance and possibly the direction from the target site. For example, the implementation of the procedure described in Section 4 used 3 neighbouring sites weighted proportionally to the inverse distance from the target site.

The interpolation of the semi-empirical distributions aims to interpolate both the shape and the range of the distributions. Consider the interpolation of 2 sites, A and B, with equal weighting (e.g. the same distance from the target but in opposite directions) where, for example, solar radiation at both sites approximately follows a normal distribution. If the mean radiation on dry days in July at Site A is 14 and at Site B is 18 and the standard deviation is small, a simple averaging of the relative frequencies will produce a distribution for the target site with peaks at 14 and 18. However, the most plausible distribution is a normal distribution with mean 16, which, since the variable can only take positive values, is best achieved by re-scaling the distributions. Similar reasoning applies for precipitation and wet and dry series, although the shape of these distributions is usually such that the relative frequencies tend to decrease as the variable increases.

The semi-empirical distributions each consist of 10 intervals and frequency values for each interval. The interpolation is carried out by matching the corresponding intervals (i.e. the fifth interval for each site is matched etc.) and averaging both the end points and the relative frequencies. The averaging of the end points approximately re-scales the distributions, which are then mixed by averaging the relative frequencies. This provides a pragmatic estimate of the likely distribution at the target site. The end points and relative frequencies of the interpolated distribution are therefore given by E_j and R_j respectively, where

$$E_j = \sum_{i=1}^{N} w_i e_{ij} \qquad (j = 0, 1, \ldots, 10) \qquad (2)$$

$$R_j = \sum_{i=1}^{N} w_i r_{ij} \qquad (j = 1, 2, \ldots, 10) \qquad (3)$$

where e_{ij} and r_{ij} are the end points and relative frequencies of the jth interval of site i.

Since the shape of the daily temperature distributions for wet and dry days are fixed as the normal distribution, it is sufficient to simply average the mean and standard deviation values. For each site i, these are given by a Fourier series, $f_i(t)$, of order 3, so that

$$f_i(t) = \frac{1}{2} a_{i0} + \sum_{j=1}^{3} [a_{ij} \cos(j\omega t) + b_{ij} \sin(j\omega t)] \qquad (4)$$

where t = Julian day and $\omega = 2\pi/365$.

Each Fourier series for the target site has the same form as Eq. (4) with cosine coefficients, A_j, and sine coefficients, B_j, given by

$$A_j = \sum_{i=1}^{N} w_i a_{ij} \qquad (j = 0, \ldots, 3) \qquad (5)$$

$$B_j = \sum_{i=1}^{N} w_i b_{ij} \qquad (j = 1, \ldots, 3) \qquad (6)$$

The target site time auto-correlation coefficients (for minimum and maximum temperature and radiation) are also the weighted averages of the neighboring site values.

3.3. Global interpolation and adjustment of parameters

This second step in the procedure uses the detailed database of mean values to adjust the distributions of precipitation and temperature to account for the effect of elevation, taking advantage of the better spatial coverage of the mean value data. Typically, precipitation increases with elevation although the amount of the increase depends on the precise local conditions. A linear regression between precipitation and elevation has

been used for several regions (Peck & Brown 1962, Osborn 1984). A better signal can probably be obtained by using the broad-scale (average) elevation over the grid of, say, 2 to 3 km (known as the spatial scale of the orographic effect) rather than the actual point elevation of the station. There are many other important factors and the PRISM model calculated a local regression equation based on relating stations situated on the same slope (Daly et al. 1994).

In order to make the results more generally applicable to regions where detailed terrain topography is not known, a different approach was used here. Partial thin plate smoothing splines, using elevation and geographical co-ordinates as independent variables, were shown to perform well in interpolating mean precipitation in southeastern Australia (Hutchinson 1995b). A recent comparison of the thin plate smoothing spline with the geostatistical method (kriging) showed the close formal connection of both methods (Hutchinson & Gessler 1994). The main advantage of the splines over geostatistical methods is that the splines do not require prior estimation of the spatial auto-covariance structure, which can be difficult to obtain. The general procedure is therefore to fit spline surfaces for the mean values of precipitation and temperature for each month and to use this information to adjust the locally interpolated parameters. This could also be applied to solar radiation if such a data set was available.

The ANUSPLIN computer program (Hutchinson 1995b) was therefore used to fit a spline surface to the observed data points for precipitation and minimum and maximum temperature. The northing and easting grid coordinates and elevation were all considered as independent variables. The spline method balances fitting a smooth surface to the points against closely matching each point, and it is thought that it works best if the typical rate of response of the dependent variable is about the same for each independent variable for the scale used (Hutchinson 1995b). In our situation, the choice of coordinates appeared to make little difference to the fits of the target sites. The grid coordinates here were specified in units of 10 km and the elevation in kilometres. The spline method does not require a regular distribution of data points. When there are a large number of data points, it is sometimes necessary to use a restricted set of points for the spline function to converge to a successful solution. The ANUSPLIN program includes a module for choosing such a sub-set by eliminating points that are close to other points. In calculating the precipitation spline functions, the data set was reduced to 750 stations, whereas all the 623 temperature sites were used for the maximum and minimum temperature functions. Once the spline functions have been calculated the

predicted mean values for any target site can be obtained simply by evaluating the function (using another module of the ANUSPLIN program).

For a given site, the spline monthly mean precipitation values are evaluated and the daily precipitation distribution obtained by the local interpolation adjusted so that the monthly mean of the resulting distribution will equal the spline value. The first step in making the adjustment is to calculate the mean monthly precipitation values for the locally interpolated distribution. The mean of the daily precipitation distribution for a given month, p, is the sum of the relative frequencies multiplied by the mid-points of the intervals

$$ p = \sum_{j=1}^{10} R_j \frac{(E_j + E_{j-1})}{2} \tag{7} $$

The mean lengths of the monthly wet and dry series, w and d, are calculated from their distributions in the same way. The estimated mean precipitation for the month, P, is then given by

$$ P = pl \frac{w}{w+d} \tag{8} $$

where l is the number of days in the month. Provided that the average lengths of the wet and dry series are short, as is the case for the UK, so that each month is not greatly affected by series from the previous month, the estimated mean will be close to the mean precipitation of the data produced by the generator. The adjustment then consists of multiplying the end points of the daily precipitation distribution, E_j, by the spline mean value, S, divided by P.

$$ E_j^{\text{adj}} = E_j \frac{S}{P} \qquad (j = 1, \dots 10) \tag{9} $$

This scales the daily distribution (and therefore the mean) by S/P so that the mean precipitation for the month equals the spline mean value, S.

It is more difficult to make an adjustment for maximum and minimum temperature because, in each case, the spline function models the overall monthly mean whereas the parameter file uses separate distributions for wet and dry days. Mean minimum and maximum temperatures are first estimated for the locally interpolated parameters for each month. For simplicity, the values of the Fourier series at the mid-point of the month are used as approximate estimates of the average daily values on wet and dry days. Denoting these values for maximum temperature for the given month by $t\text{max}_w$ and $t\text{max}_d$, respectively, the estimated mean maximum temperature for the month, $TMAX$, is given by

$$ TMAX = \frac{w \times t\text{max}_w + d \times t\text{max}_d}{w+d} \tag{10} $$

The average of the spline less $TMAX$ values for all 12 months is then calculated (i.e. the difference in annual means) and the Fourier series for mean maximum temperature for both wet and dry days adjusted by this amount (by adding twice the value to the A_0 coefficient). The same procedure is followed for minimum temperature.

There are other ways in which temperature could be adjusted. In particular, average lapse rates could be calculated from the mean value statistics for the relationship between temperature and elevation and the neighbouring sites adjusted before the local interpolation is carried out. However, such rates will tend to vary with climate.

4. EVALUATION OF THE INTERPOLATION METHOD

To test the methodology, 3 sites were chosen from the 138 sites with daily weather data and interpolated from the remaining data using the interpolation procedure described in Section 3. The 3 sites were removed from the data points used to fit the spline surface as, in general, the target sites for the interpolation will usually be locations with no observed data. Synthetic weather data produced by LARS-WG from the interpolated file was then compared both with the observed data for the site and with the synthetic weather data generated by LARS-WG using the parameter file derived directly from the observed data.

4.1. Choice of test sites

The 3 test sites, Buxton, Redesdale and Pershore, were chosen to give a variety of test conditions. In each case, the local interpolation uses 3 neighbouring sites, with the target site being situated within the triangle formed by its neighbours. The database of 138 sites allows 3 such sites to be found for most locations in Great Britain (the main exception being near the coast) and this is considered to be the most appropriate approach in the majority of cases. The same methodology can still be applied if the shape formed by the neighbouring sites does not contain the target site. In some cases, where 3 sites exist surrounding the site, there may be another site much closer to the target than one of the chosen sites so that it may be preferable to use this site instead. The locations of the target sites and their neighbours are shown in Fig. 1. The grid references and summary weather statistics are shown in Table 1.

Buxton was chosen as a site that is likely to be difficult to model closely by interpolation. It is considerably

Fig. 1. Locations of target sites and their neighbours used in the validation of the interpolation method (details in Table 1)

higher than any of its neighbouring sites, being situated on the western side of the Pennine hills, with the prevailing weather coming from the west. Consequently Buxton is considerably wetter and cooler than its neighbours.

Redesdale is also a site at quite a high elevation. However it is on the east side of the Cheviot Hills and, being in a rain shadow, has similar rainfall to Boulmer (a coastal site) and Hayden Bridge (in a valley at the northeast edge of the Pennines) even though they are at much lower altitude. Eskdalemuir is much further west and has much higher rainfall. As with Buxton and its neighbours, the decrease in temperature with elevation is noticeable in the comparison of these sites. The temperature adjustments made in the interpolation method (Section 3.3) are shown in Table 2. There are considerable distances between the sites (the database sites tend to be more sparsely situated further north) and Boulmer may have different characteristics to the others because of its coastal situation. Again, Redesdale would also be expected to be a difficult site to interpolate well.

On the other hand, the successful interpolation of Pershore should be much easier because the neighbouring sites are at similar elevation and the sites are quite close together. As a result, there is little difference in average precipitation or temperature.

Table 1. Geographical, precipitation and temperature characteristics of the target sites and their neighbours used in the validation of the interpolation method. Periods of available data are also shown

Site	Latitude (°N)	Longitude (°E)	Altitude (m)	Annual total rain (mm)	Annual mean temperature (°C)	Distance to target site (km)	Years
Manchester	53.35	−2.27	75	806	9.5	26	1949–1995
Keele	53.00	−2.27	179	787	8.7	36	1959–1995
Warsop	53.22	−1.12	46	629	9.0	53	1959–1995
Buxton	53.25	−1.92	307	1286	7.8	0	1959–1995
Boulmer	55.42	−1.60	23	647	8.6	46	1975–1995
Hayden Bridge	54.97	−2.25	79	724	8.6	31	1959–1995
Eskdalemuir	55.32	−3.20	242	1550	7.0	59	1957–1992
Redesdale	55.25	−2.27	244	873	7.0	0	1970–1995
Preston Wynne	52.12	−2.50	84	660	9.2	31	1959–1995
Cheltenham	51.90	−2.05	65	743	10.1	20	1959–1995
Luddington	52.17	−1.75	56	612	9.5	23	1968–1989
Pershore	52.08	−2.05	40	637	9.9	0	1959–1995

Table 2. Temperature adjustments made to maximum and minimum temperature at Buxton, Redesdale and Pershore

	Max. temperature (°C)	Min. temperature (°C)
Buxton	1.84	0.87
Redesdale	1.38	1.54
Pershore	−0.20	−0.39

4.2. Tests carried out

The interpolation procedure results in an interpolated parameter file that can be used by LARS-WG to produce synthetic weather data for the target site. The best synthetic data that LARS-WG can produce is that using the parameters derived directly from the observed data. Most of the tests therefore compare output data generated by LARS-WG using these 2 files.

The parameter values themselves were not formally compared because it is the LARS-WG output rather than the parameters that would be used in modelling studies, although a comparison of the parameter values helps to explain any differences. In addition, the interactions of the parameters makes the effects of differences in the parameter values on the output weather data difficult to calculate. The interactions also mean that it is difficult to derive analytically the resulting distributions of the synthetic weather data that LARS-WG will produce. For example, maximum and minimum temperatures for a given month result from the interaction of the choice of wet and dry spells with the separate wet and dry temperature distributions. The spells from one month may continue into the next and the parameters of the temperature distributions vary from day to day. Successive values are also correlated, as are the maximum and minimum values.

Therefore, 100 yr of synthetic weather data were generated by LARS-WG using the interpolated parameter file, and 100 yr generated using the observed data parameter file. These 2 data sets were then statistically compared. To simplify the discussion, these will be described as interpolated data and directly generated data respectively. LARS-WG automatically calculates summary statistics for each month for all 4 weather variables (precipitation, minimum and maximum temperature and radiation) and these values were compared. These consist of the mean, the standard deviation of the daily values for the month (except for precipitation) and the inter-annual standard deviation of the monthly means. Table 3 shows the average of each of these statistics for the 12 months and the root mean square error of the interpolated compared to the directly generated weather data. The spline predictions

Table 3. Comparison of the average annual mean and average of the monthly standard deviation values for the observed data, for 100 yr of synthetic data produced from the directly generated parameter file (generated) and for 100 yr of synthetic data produced from the interpolated parameter file (interpolated). The root mean square error (RMSE) of the differences between the interpolated and generated values for the 12 mo is also shown with percentage values given in parentheses

	Buxton				Redesdale				Pershore			
	Observed	Generated	Interpolated	RMSE	Observed	Generated	Interpolated	RMSE	Observed	Generated	Interpolated	RMSE
Rain (mm)												
Mean	107.1	107.4	99.0	12.38 (12)	72.8	73.0	81.2	9.9 (14)	53.1	54.5	51.0	6.9 (13)
SD monthly	49.2	40.1	42.6	6.07 (15)	34.2	28.0	34.7	7.8 (28)	29.7	25.9	24.9	3.8 (15)
Min. temperature (°C)												
Mean	4.66	4.60	4.58	0.22 (5)	3.41	3.33	3.28	0.22 (7)	6.06	5.93	6.03	0.23 (4)
SD daily	3.05	3.05	3.17	0.18 (6)	3.21	3.20	3.20	0.12 (4)	3.28	3.30	3.38	0.14 (4)
SD monthly	1.26	1.04	1.06	0.14 (13)	1.19	0.98	0.98	0.10 (11)	1.38	1.13	1.11	0.15 (13)
Max. temperature (°C)												
Mean	10.86	10.83	10.88	0.25 (2)	10.63	10.59	10.44	0.43 (4)	13.82	13.70	13.76	0.23 (2)
SD daily	3.35	3.29	3.28	0.08 (2)	3.16	3.12	3.04	0.18 (6)	3.38	3.33	3.33	0.19 (6)
SD monthly	1.60	1.34	1.33	0.11 (8)	1.45	1.17	1.11	0.10 (9)	1.64	1.32	1.37	0.19 (14)
Radiation (MJ m⁻²)												
Mean	8.28	8.37	8.58	0.33 (4)	8.07	8.22	8.42	0.37 (5)	9.20	9.25	9.13	0.22 (2)
SD daily	4.42	4.30	4.47	0.20 (5)	4.38	4.27	4.22	0.15 (4)	4.69	4.57	4.57	0.14 (3)
SD monthly	1.35	1.01	1.05	0.10 (10)	1.11	0.99	0.96	0.12 (12)	1.32	1.03	1.10	0.11 (11)
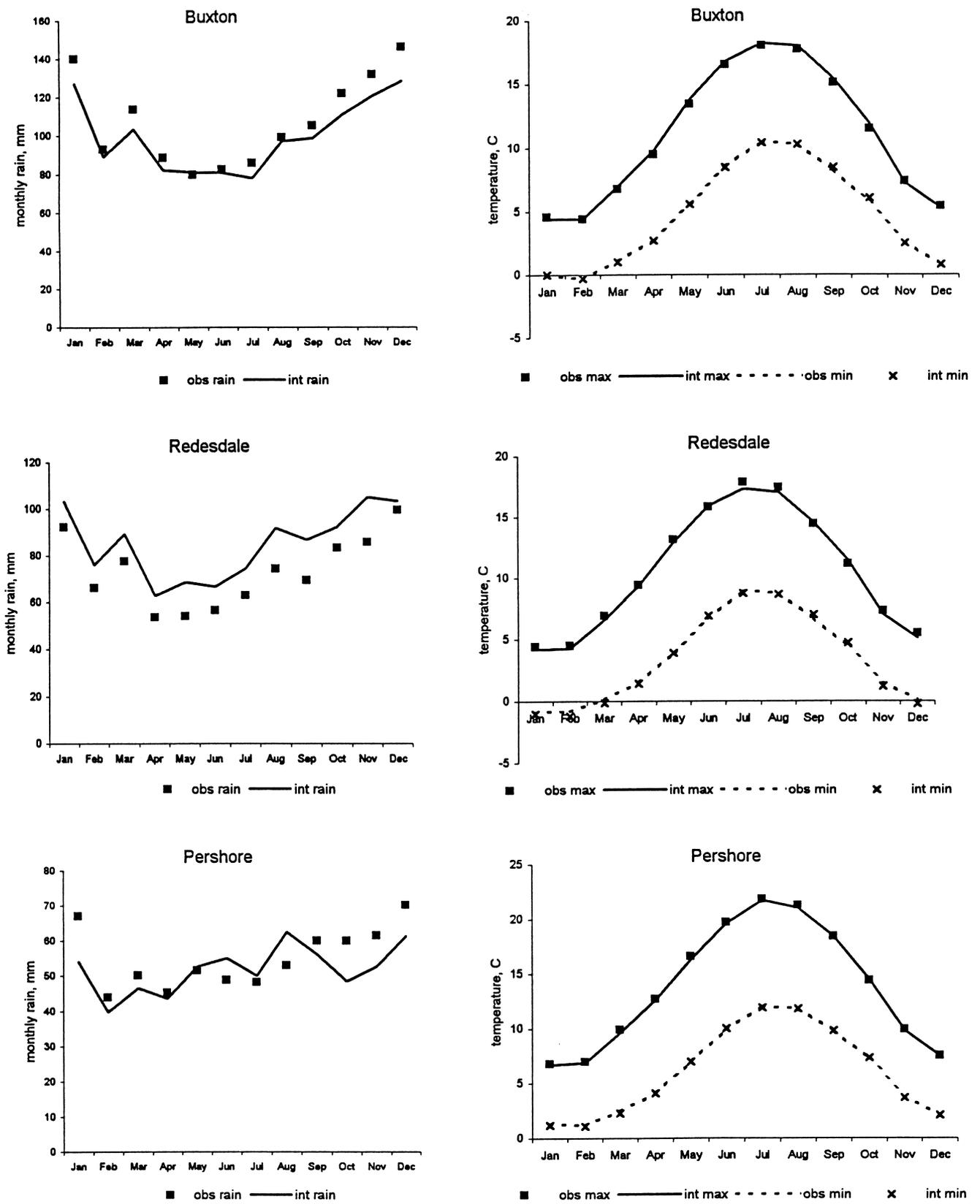
Fig. 2. Comparison of observed and interpolated monthly totals of precipitation and monthly mean maximum and minimum temperature at Buxton, Redesdale and Pershore

for monthly mean precipitation and temperature are compared with the observed values in Fig. 2.

The interpolated data can also be compared with the observed data. However, some of the differences will be due to the limitations of LARS-WG, rather than being due to the interpolation process. Generally, LARS-WG reproduces most of the characteristics of input observed data well so that comparisons with the observed or directly generated data will give similar results. The main shortcoming found when LARS-WG was tested for 18 sites worldwide was that the synthetic data generated tend to have a lower standard deviation of monthly means than the observed data (Semenov et al. 1998). Circulation patterns persisting over several days or weeks will result in the observed data containing complex time auto-correlations that are not modelled by the weather generator. Such correlations increase the variability of monthly means.

LARS-WG implements a number of statistical tests to compare the synthetic data produced by the weather generator with the observed data on which it is based. These are designed for the direct generation of synthetic from observed data to test how well the weather generator is performing in reproducing the characteristics of the observed data. They are described in detail in Semenov et al. (1998) and, as well as comparing the mean and variances, they also test the distributions of wet and dry series and precipitation, as well as the agriculturally important extreme events of frost and high temperature. These tests were carried out here by comparing both the directly generated and the interpolated data with the observed data. The tests identify cases in which the synthetic data would be unlikely to come from the same distribution as the observed using a 5% significance level. Even if the distributions were identical, therefore, 1 out of 20 tests would be expected to give a significant result. The number of such failures is shown in Table 4 and a comparison of the interpolated and directly generated values gives the additional number of failures resulting from the interpolation process. However, the direct comparison of the interpolated and directly generated data is a better test in that it measures the magnitude of the differences rather than the existence of statistically significant differences, since the interpolation process would be expected to introduce some such differences (Table 3).

4.3. Results

4.3.1. Buxton

The average annual precipitation for the interpolated data is less than both the observed and the directly generated data by about 8%. This is a result of the spline under-estimating precipitation, mainly in the winter months (Fig. 2). Buxton receives particularly high rainfall because of its situation on the windward side of the Pennines and so the rainfall is higher than would be expected based just on its elevation.

Examination of the parameters also shows that the average length of the interpolated wet series is less and of the interpolated dry series is more than that of the directly generated parameters. The series are obtained just by local interpolation and so this is a result of the drier neighbouring sites having longer dry

Table 4. Results of the statistical tests comparing the observed data for 3 sites with 2 datasets of 100 yr of synthetic data generated by LARS-WG. The first dataset was generated using a parameter file fitted to the observed weather (Gen) and the second used a parameter file interpolated from neighbouring sites (Int). The variables compared for rainfall are the seasonal distributions of wet and dry series, the distributions of daily rainfall for each month, monthly total mean rainfall and its variance (var.). For minimum (Tmin) and maximum (Tmax) temperature and for solar radiation (rad.), the monthly means, the inter-annual variances of the monthly means (inter-ann. var.) and the within-month variances of the daily values were compared, as well as seasonal distributions of spells of temperature below 0°C and above 30°C (frost and hot spells). Values shown are the numbers of tests giving significant results at the 5% significance level

	Wet and dry series	Rain distri- bution	Mean rain	Var. rain	Mean Tmin	Inter- ann. var. Tmin	Var. Tmin	Mean Tmax	Inter- ann. var. Tmax	Var. Tmax	Mean rad.	Inter- ann. var. Rad	Var. rad.	Frost and hot spells
Number of tests	8	12	12	12	12	12	12	12	12	12	12	12	12	8
Statistical test	χ^2	χ^2	t-test	F-test	t-test	F-test	F-test	t-test	F-test	F-test	t-test	F-test	F-test	χ^2
Buxton, Gen	0	0	0	5	0	5	1	0	4	1	1	5	4	2
Buxton, Int	2	9	2	2	0	5	6	1	4	4	6	4	2	1
Redesdale, Gen	0	1	0	4	0	2	1	1	5	3	1	1	1	0
Redesdale, Int	4	10	2	1	1	2	2	3	7	4	7	1	4	2
Pershore, Gen	0	0	0	2	0	3	1	0	5	1	0	4	2	1
Pershore, Int	1	4	3	6	0	5	5	1	5	6	1	2	3	2

spells (due to more dry days) than Buxton. There is no mechanism in the methodology to adjust the lengths of the locally interpolated dry and wet series, but the series lengths are taken into account in adjusting the daily precipitation distribution (Eq. 8). In the summer months, for example, the mean monthly rainfall of the interpolated data matches the observed mean rainfall well since the spline values are quite accurate (Fig. 2). The methodology achieves this despite the smaller ratio of average lengths of wet to dry spells by having higher average daily precipitation. For example, in May the calculated mean precipitation values for the directly generated and interpolated parameter files are 82.3 and 82.1 mm, respectively (the observed value is 79.7 mm). However, these values result from average values of the wet spells, dry spells and daily precipitation of 3.46 d, 3.30 d and 5.19 mm for the directly generated file, but from values of 3.14 d, 3.46 d and 5.57 mm for the interpolated file. The ratio of the average lengths of wet series to dry series will tend to increase as monthly precipitation increases, but a general relationship is unlikely to hold, making it difficult to derive a robust adjustment on this basis.

The temperature statistics of the interpolated data are all close to those of the directly generated data since the spline does an excellent job in reproducing the observed monthly means (Fig. 2). As already explained, data generated by LARS-WG tends to have a lower inter-annual standard deviation of monthly means compared with the observed data and this is the case here. Radiation data was not available at a sufficiently large number of sites and so the interpolation for this variable is purely local. The drier nature of the interpolating sites (and, hence, less cloud) means that radiation is slightly over-estimated in the interpolated data in most months. Again, it might be possible in future developments of the methodology to incorporate an adjustment based on precipitation levels.

In the statistical tests against the observed data (Table 4), the main difference between the directly generated and interpolated values is the increased number of failures for the daily rainfall distribution, monthly mean radiation and daily variability in temperature. As already discussed, the interpolated rainfall parameters differ from the directly generated parameters (and, hence, the observed data) in both the spell lengths and the daily rainfall distribution, although the procedure ensures that the resulting monthly mean values match the spline values. This results in 9 failures for the test of the daily rainfall distribution for the interpolated data but only 2 for the monthly means. The purely local nature of the radiation interpolation procedure leads to 6 failures for the tests on mean radiation. The daily temperature standard deviations are also obtained simply by local inter-

polation and tend to be slightly under-estimated, resulting in more significant values in statistical tests. For most statistics, however, the interpolated data matches the observed data well.

4.3.2. Redesdale

The spline overestimated precipitation in all months at Redesdale. This is the opposite situation to Buxton, since Redesdale is in a rain shadow and so has less precipitation than would be expected based just on its elevation. As with Buxton, the lengths of the wet series tend to be under-estimated and daily precipitation over-estimated because 2 of the neighbouring sites are drier. The third site, Eskdalemuir, has longer wet spells (and much larger annual precipitation), but it is farther from Redesdale than the other 2 sites and therefore receives a smaller weighting. The results for temperature and radiation are very similar to those for Buxton with a close match of the interpolated temperature statistics to those of the directly generated data. Radiation values are slightly high because 2 of the sites are drier, and the observed data have higher inter-annual variability of means than the weather generated data. The over-estimation of precipitation results in many failures for the χ^2 test on the daily precipitation distribution. The high variance in the monthly precipitation means prevents the differences in monthly means from being significant at the 5% level for this length of run (only 2 significant values).

4.3.3. Pershore

As with Buxton, the spline under-estimates precipitation from October to January but gives a good match for the rest of the year. The slightly wetter nature of 2 of the neighbouring sites results in the interpolated mean wet spells tending to be a little longer than those for the directly generated parameter file. Temperature and radiation are both modelled well, giving results similar to the other 2 test sites.

5. SUMMARY AND DISCUSSION

An interpolation procedure for the LARS-WG weather generator, based on a combination of local and global techniques, has been described. The local method constructs the distribution of the weather variables using the distributions from neighbouring sites. The sites used for this purpose must have relatively long daily records of weather in order to enable calculation of reliable distributions. Such weather data are

usually available only at a low spatial resolution. The interpolation of the neighbouring sites can be used to provide an estimate of the shape of the distribution at an unobserved location. The global method, thin plate spline interpolation, is used to calculate a spatial trend in monthly statistics of a weather variable such as, for example, monthly mean rainfall, taking into account the elevation of the sites. To be successful, the thin plate spline interpolation requires higher spatial coverage but with lower temporal resolution, such as monthly means of the weather variables. The interpolated distributions can then be re-scaled so that their means match as closely as possible the values predicted by the spline.

The interpolation method for LARS-WG was tested in 3 different areas in Great Britain and showed a good performance when compared to the observed data. The main differences occurred when the spline was unable to reproduce closely the observed mean values of precipitation. Differences were also observed for mean radiation since, in the absence of a spatially detailed data set, only a local interpolation was carried out. It may be possible to adjust radiation by relating it to the amount of precipitation.

At present, the thin spline interpolation uses the actual elevation of sites when calculating the relationship between precipitation and elevation. The substitution of actual elevation with 'orographic' elevation averaged over 5 to 10 km^2 may potentially improve the precipitation-elevation relationship and, as a result, improve the performance of the method (Daly et al. 1994, Hutchinson 1995b). A digital elevation model (DEM) will be required to calculate average elevations.

One of the limitations of the described methodology is that the generated weather time series at different sites are not spatially correlated. For most agricultural regional assessments, the analysis can be done separately at different sites or grid-boxes without taking into account inter-site correlation, because the interactions between processes at different sites are very weak. On the other hand, for some hydrological applications, such as flood prediction, the lack of spatial correlation between daily rainfall would be crucial, because the total amount of precipitation over the whole watershed needs to be estimated. The methodology for developing a regional weather generator to produce spatially correlated synthetic weather data with the same characteristics as the observed data is unknown. The current method takes into account the effects of elevation, which are incorporated into the spline interpolation. Local irregular weather effects, such as coastal effects, cannot be incorporated and so will only be taken into account if one of the neighbouring sites has similar characteristics.

The spatial interpolation of the LARS-WG generator allows baseline daily weather data to be provided for any location in Great Britain, which can then be used in conjunction with impact models for impact assessment studies. The interpolation procedure produces a LARS-WG parameter file for the target location and this also enables climate change scenarios to be constructed for any location, using the methodology described in Semenov & Barrow (1997). This approach takes daily data from a global climate model (GCM) for the appropriate grid box and a comparison of control and perturbed runs is used to calculate changes in precipitation intensity, duration of wet and dry spells and temperature means and variances. These values are then used to adjust the parameters of the weather generator at the location of interest in order to simulate a site-specific daily climate change scenario. Changes in the mean and variability can therefore be incorporated into climate change scenarios in a consistent and computationally inexpensive way. This also avoids any interpolation of the predicted future weather data.

The spatial interpolation of the LARS-WG generator was applied here for Great Britain and showed a good performance for the test sites. The interpolation method could be applied in principle to any regions or countries where sufficient weather data sets are available. Although the method is based on 2 models, LARS-WG and ANUSPLIN, which have both been shown to perform well for diverse climates and different regions (Hutchinson 1995b, Semenov et al. 1998), it does not necessarily mean that the spatial interpolation will also perform well for every region. The performance of the interpolation method needs to be tested for a new region before it is used in impact studies.

The interpolation method is very flexible from a practical point of view. A new site with a long weather record can easily be added to the database, by calculating the parameters of LARS-WG for the site, which allows the site to be used in the local interpolation without having to re-calculate the global characteristics.

The LARS-WG stochastic weather generator is available from the authors (a Windows 95/NT version of theLARS-WG stochastic weather generator can be downloaded from http://www.lars.bbsrc.ac.uk/model/larswg.html). A new version, which includes interpolation for Great Britain, is under development and testing and will be available soon.

LITERATURE CITED

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J Appl Meteorol 33:140–158

Harrison PA, Butterfield RE, Downing TE (eds) (1995) Climate change and agriculture in Europe. Assessment of impacts and adaptation. Environmental Change Unit, University of Oxford

Hanson CL, Cumming KA, Woolhiser DA, Richardson CW (1994) Microcomputer program for daily weather simulation in the contiguous United States. Agricultural Research Service ARS-114, US Department of Agriculture, Washington, DC

Hutchinson M (1995a) Stochastic space-time weather models from ground–based data. Agric For Meteorol 73:237–265

Hutchinson MF (1995b) Interpolating mean rainfall using thin plate smoothing splines. Int J Geogr Inf Sys 9:385–403

Hutchinson MF, Gessler PE (1994) Splines—more than just a smooth interpolator. Geoderma 62:45–67

Mearns LO, Rosenzweig C, Goldberg R (1997) Mean and variance change in climate scenarios: methods, agricultural applications, and measures of uncertainty. Clim Change 35:367–396

Osborn HB (1984) Estimating precipitation in mountainous regions. J Hydraul Eng 110:1859–1863

Peck EL, Brown MJ (1962) An approach to the development of isohyetal maps for mountainous areas. J Geophys Res 67:681–694

Phillips DL, Dolp J, Marks D (1992) A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agric For Meteorol 58:119–144

Racsko P, Szeidl L, Semenov M (1991) A serial approach to local stochastic weather models. Ecol Model 57:27–41

Richardson CW (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour Res 17:182–190

Richardson CW, Wright DA (1984) WGEN: a model for generating daily weather variables: Agricultural Research Service ARS-8, US Department of Agriculture, Washington, DC

Running SW, Newmani RR, Hungerford RD (1987) Extrapolation of synoptic meteorological data in mountainous terrain and its use for simulating forest evapotranspiration and photosynthesis. Can J For Res 17:472–483

Semenov MA, Barrow EM (1997) Use of a stochastic weather generator in the development of climate change scenarios. Clim Change 35:397–414

Semenov MA, Brooks RJ, Barrow EM, Richardson CW (1998) Comparison of the WGEN and LARS-WG stochastic weather generators in diverse climates. Clim Res 10:95–107

Wilks DS (1992) Adapting stochastic weather generation algorithms for climate changes studies. Clim Change 22:67–84