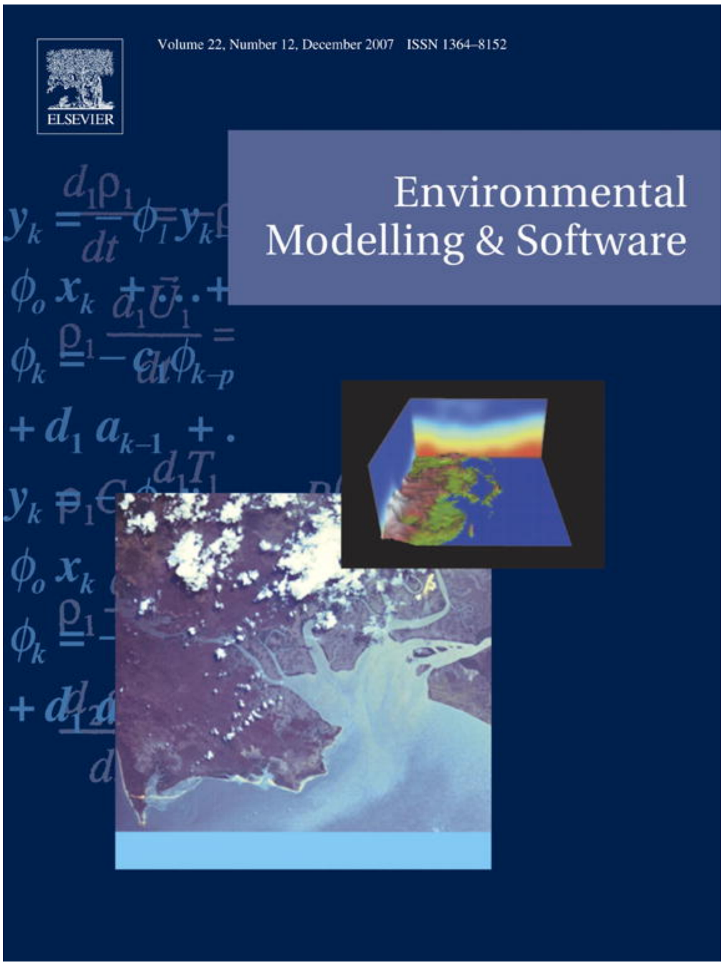
Volume 22, Number 12, December 2007 ISSN 1364-8152
ELSEVIER
Environmental
Modelling & Software

Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software 22 (2007) 1705–1719
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

A daily weather generator for use in climate change studies

C.G. Kilsby [a,*], P.D. Jones [b], A. Burton [a], A.C. Ford [a], H.J. Fowler [a], C. Harpham [b], P. James [a], A. Smith [a], R.L. Wilby [c]

[a] School of Civil Engineering and Geosciences, University of Newcastle upon Tyne, Claremont Road, Newcastle upon Tyne NE1 7RU, UK
[b] Climatic Research Unit (CRU), School of Environmental Sciences, University of East Anglia, Norwich, UK
[c] Environment Agency of England and Wales, Nottingham, UK

Received 19 January 2006; received in revised form 9 February 2007; accepted 12 February 2007
Available online 26 April 2007

Abstract

This paper describes the development of a weather generator for use in climate impact assessments of agricultural and water system management. The generator produces internally consistent series of meteorological variables including: rainfall, temperature, humidity, wind, sunshine, as well as derivation of potential evapotranspiration. The system produces series at a daily time resolution, using two stochastic models in series: first, for rainfall which produces an output series which is then used for a second model generating the other variables dependent on rainfall. The series are intended for single sites defined nationally across the UK at a 5 km resolution, but can be generated to be representative across small catchments (<1000 km^2). Scenarios can be generated for the control period (1961–1990) based on observed data, as well as for the UK Climate Impacts Programme (UKCIP02) scenarios for three time slices (2020s, 2050s and 2080s). Future scenarios are generated by fitting the models to observations which have been perturbed by application of change factors derived from the UKCIP02 mean projected changes in that variable. These change factors are readily updated, as new scenarios become available, and with suitable calibration data the approach could be extended to any geographical region.

Keywords: Weather generator; Stochastic; Rainfall model; Climate change; Climate scenario

Software availability

Name of software: EARWIG (Environment Agency Rainfall and Weather Impacts Generator).
Developer: School of Civil Engineering and Geosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.
Contact: Chris Kilsby, School of Civil Engineering and Geosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK. E-mail address: c.g.kilsby@ncl.ac.uk
Year first available: 2006.
Hardware required: PC.
Software required: Windows 2000 or XP.
Program language: Fortran numerics: interface developed under MS Visual Studio.
Program size: 30 Mb of disk space required: runs within 60 Mb of memory.
Availability: Can be made available to researchers on request to the authors.
Cost: N/A.

1. Introduction

Impact assessments of climate change on hydrology and related fields such as agricultural and water management practice require time series of weather variables for specific catchments or locations at daily or higher resolution. Data are needed for both the current climate and a range of future possible scenarios. These series must be consistent, both between variables, and with a range of observed and projected

* Corresponding author. Tel.: +44 191 222 5614; fax: +44 191 222 6669.
E-mail address: c.g.kilsby@ncl.ac.uk (C.G. Kilsby).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2007.02.005

statistics of the variables in order to account for extremes (floods and droughts) and seasonality. Such series are not directly available from climate models, and this paper describes an alternative approach using a "weather generator" to provide series of rainfall, temperature, humidity, wind and potential evapotranspiration at river catchment scales.

The proposed methodology uses stochastic models of rainfall and weather. Previous weather generators have used simple rainfall models based on either Markov chains (Richardson, 1981) or empirical distributions of wet/dry spells (Semenov and Brooks, 1999). The approach described here uses a more sophisticated Neyman-Scott point process model (Cowpertwait, 1991) capable of more accurately reproducing higher order rainfall statistics. The general approach taken is as follows:

- Observed data of rainfall and other weather variables used to define current climate.
- Regional Climate Model (RCM) rainfall and temperature data used to derive factors of change from current climate state to define climate change scenarios.
- Stochastic model of daily rainfall fitted to current climate, and then re-fitted for possible future climates using future, factored daily rainfall statistics.
- Weather Generator (WG) model based on regression relations between daily climatic variables and daily rainfall: parameterised using current climate data, and then applied for possible future climates, using future factored daily climate variable statistics.
- Software implementation using a map viewer linked to a spatial database allowing the flexible selection of areas for generation of series.

The essentials of this methodology have been previously applied within the BETWIXT project at 17 UK sites (for rainfall) and 5 sites for all variables (see BETWIXT Project website). This paper describes how the method can be extended to continuous UK coverage, providing a capability of generating rainfall and other weather variables representative of any location and a range of possible future climates and time-periods. In principle, and with suitable observed and RCM data, the approach can be extended to any geographical region. The elements of the methodology are described in more detail in the following sections.

2. Climate data

2.1. Observed data

UKMO/UKCIP (UK Meteorological Office/UK Climate Impacts Programme) 5-km gridded weather data were used in this project (Perry and Hollis, 2005a,b). These consist of two 5 km × 5 km gridded datasets covering the UK for the period 1961–2000. The first is of monthly values of mean temperature, daily temperature range, rainfall, sunshine, cloud, and wind speed. The second is daily rainfall for the period 1958–2002. These data were generated in a geographical information system combining multiple regression with inverse distance-weighted interpolation taking account of geographic and topographic factors (Perry and Hollis, 2005a,b). Daily series from 115 sites, with a reasonable national coverage, were used to provide additional information on climatic variables at the daily level. These sites are the same as those used by Osborn et al. (2000).

The 1961–1990 period is taken as a climatological normal for rainfall. However, it is possible to use other periods for this purpose, or even to set the model up for separate decades to explore issues of climatic variability and stationarity of model relationships.

Grids of rainfall statistics derived from this data set are shown in Fig. 1.

Variables apart from rainfall are available only at the daily level for 1995–2000 so an approach combining monthly data (available 1958–2002) with site data has been followed (see Section 5).

2.2. Climate scenarios

The methodology is illustrated using UKCIP02 change factors, but the approach is applicable using change factors taken from any global or regional climate model. The model future scenarios are based on the UKCIP02 scenarios (Hulme et al., 2002) for four emissions scenarios (SRES A1, A2, B1 and B2) and three future time-slices (2020s, 2050s and 2080s as defined below) derived from the HadRM3H integrations. A control scenario simulating the 1961–1990 period is also available. The approach relies on deriving factors of change for various statistics from control to future scenarios and applying these to observed statistics, rather than using the RCM's rainfall climatology directly as it does not reproduce the spatial patterns of mean rainfall or seasonality accurately (Fowler and Kilsby, 2004) and, more importantly, does not accurately represent extreme dry spells or extreme rainfall events (Fowler et al., 2005).

Change factors are derived using multiplicative factors for rainfall statistics and additive ones for other climate variables on a calendar month basis. These are taken directly as ratios for the mean (M), variance (Var) and skewness (S) of daily rainfall, and a logit transformation of proportion of dry days (PDry) to ensure linearity across the range of values.

The following equations are used to apply the calculated change fields (α) for a general variable P (using the suffix GCM to indicate climate model values):

$$\frac{P^{\mathrm{Fut}}}{P^{\mathrm{Obs}}} = \frac{P^{\mathrm{GCMFut}}}{P^{\mathrm{GCMCon}}} \quad (1)$$

where $\alpha = P^{\mathrm{GCMFut}}/P^{\mathrm{GCMCon}}$ and therefore,

$$P^{\mathrm{Fut}} = \alpha P^{\mathrm{Obs}} \quad (2)$$

For PDry however, the following equation is used:

$$\frac{X(\mathrm{PDry}^{\mathrm{Fut}})}{X(\mathrm{PDry}^{\mathrm{Obs}})} = \frac{X(\mathrm{PDry}^{\mathrm{GCMFut}})}{X(\mathrm{PDry}^{\mathrm{GCMCon}})} \quad (3)$$

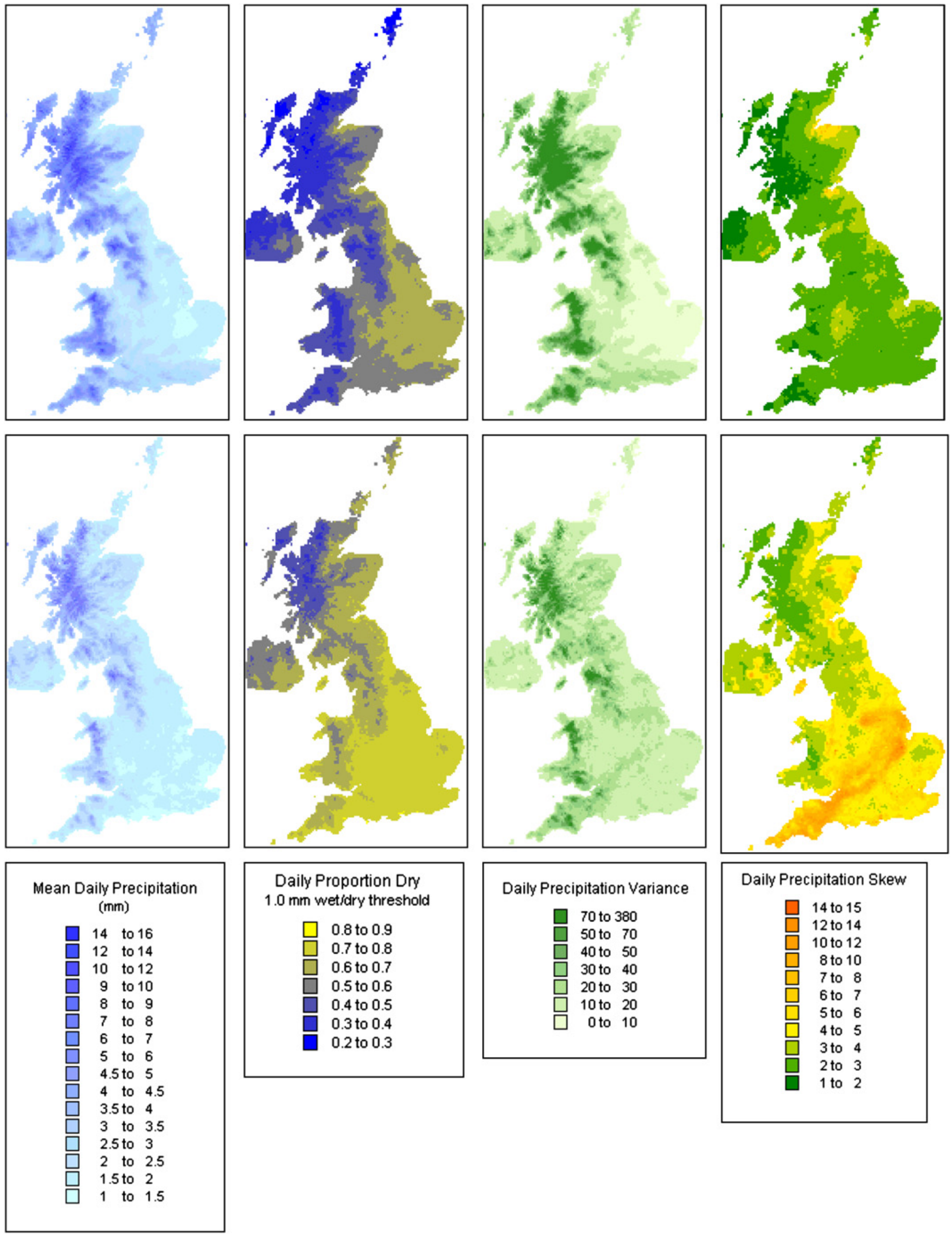

Fig. 1. Daily rainfall statistics from the UKMO 5 km data set. Top row January, second row July. Units are (mm) for mean precipitation and mm^2 for variance. The proportion dry and skew are dimensionless.

where $X(\mathrm{PDry}) = \mathrm{PDry}/1 - \mathrm{PDry}$ and $\alpha = X(\mathrm{PDry}^{\mathrm{GCMFut}})/X(\mathrm{PDry}^{\mathrm{GCMCon}})$ and therefore,

$$\mathrm{PDry}^{\mathrm{Fut}} = X^{-1}\left(\alpha X\left(\mathrm{PDry}^{\mathrm{Obs}}\right)\right) \tag{4}$$

For temperature, the change factor is additive and the following equation is used:

$$T^{\mathrm{Fut}} - T^{\mathrm{Obs}} = T^{\mathrm{GCMFut}} - T^{\mathrm{GCMCon}} \tag{5}$$

where $\alpha = T^{\mathrm{GCMFut}} - T^{\mathrm{GCMCon}}$ and therefore,

$$T^{\mathrm{Fut}} = T^{\mathrm{Obs}} + \alpha \tag{6}$$

Change fields have been derived using all three ensemble members of the HadRM3H SRES A2 Scenario (UKCIP02 Medium-High Scenario), using 1961–1990 for the Control Scenario and 2071–2100 for the Future Scenario to be consistent with the UKCIP02 Scenarios (i.e. change field for M-H 2080s Scenario). To apply the change fields to other Scenarios (Low, Medium-Low and High) and time-slices (2020s, 2050s and 2080s), scaling factors between the global and regional climate models were developed by UKCIP02 to produce four scenarios matching the original SRES emissions scenarios. Time slices are then produced by taking the mean climate for periods conventionally defined as the 2020s (2011–2040), 2050s (2041–2070) and 2080s (2071–2100). The scaling between the 2080s and earlier periods is linked to changes in global mean temperature. These scaling factors are given in Table 1.

These scaling factors were applied to the regional climate changes of the 2080s, α, to produce change fields for all four emissions scenarios and three time-slices. Change fields (α) are shown in Fig. 2a for precipitation statistics, and for temperature in Fig. 2b.

The UKCIP02 data are available on a rotated polar grid with 50 km resolution, so the change factor fields are queried to identify which of the RCM grid squares the contains the centroid of each 5 km grid square.

3. Rainfall model

There is a very extensive literature on stochastic rainfall models, with applications for single- and multi-sites, and for durations ranging from annual through daily, hourly and down to 5-min intervals. Srikanthan and McMahon (2001) review the main methods, including re-sampling models, Markov chain models and two-stage occurrence and amount models. Whilst most models can reproduce low order moments of rainfall (mean and variance) and wet-dry behaviour at the daily level, they cannot accurately reproduce extremes. They additionally do not accurately reproduce the lower order statistics over a wider range of time aggregations such as at the hourly or monthly level. Additionally, most models are not readily adapted for use in future climate applications as they do not explicitly allow re-parameterisation using future projected statistics readily available from climate models.

Stochastic rainfall models using clustering approaches handle occurrence and amount in one process, and are attractive in that they represent the observed clustered nature of rainfall. They have been extensively developed over the last 20 years or so, following Rodriguez-Iturbe et al. (1987). Two main variants, the Neyman-Scott Rectangular Pulses (NSRP) model and Bartlett-Lewis Rectangular Pulses (BLRP) model have been extensively developed and evaluated (Onof et al., 2000; Velghe et al., 1994). Much development has been invested in the NSRP model, which is capable of producing rainfall series of arbitrary length and time resolution down to minutes. NSRP is the basis for standard UK industrial urban drainage design software, and has been regionalised for any site in the UK (Cowpertwait et al., 1996). It has been shown to realistically reproduce extreme values for engineering impact studies, most recently using multi-site data of intense events from Italy (Cowpertwait et al., 2002) and for UK single-site data under present and future climates (Kilsby et al., 2004). Within the BETWIXT project the model was parameterised for 17 sites with hourly rainfall data. The model is run within the RainClim software package, and can generate series at daily, hourly and 5-min resolution for the 17 sites for the current (1961–1990) period as well as for the four emissions scenarios and three time-slices of the UKCIP02 scenarios.

Table 1
Multiplying factors for conversion from 2080s Medium-High scenario to other scenarios and time-slices

Time-slice	Low emissions SRES B1	Medium-Low emissions SRES B2	Medium-High emissions SRES A2	High emissions SRES A1
2020s	0.24	0.27	0.27	0.29
2050s	0.43	0.50	0.57	0.68
2080s	0.61	0.71	1.00	1.18

Source: UKCIP02 Scientific Report (Hulme et al., 2002).

The fitting method is flexible since it does not use rainfall series directly, but selected characteristic rainfall sample statistics. These statistics can be from observed rainfall, or downscaled from atmospheric circulation variables output from numerical climate models, e.g. global General Circulation Models (GCMs) or Regional Climate Models (RCMs). Here, the parameters will be defined using combinations of the UKMO/UKCIP 5-km observed rainfall statistics together with change factors from the UKCIP02 data as described above. In principle, this flexibility allows other sources of climate model data to be used interchangeably.

3.1. Definition of the Neyman-Scott Rectangular Pulses model

In the NSRP model, rainfall is associated with clusters of "rain cells" making up "storm events". The model rain cells may be thought of, conceptually at least, as loosely representing small-scale rain-bearing meteorological structures. For example, a short intense rain cell could be a convective system (thunderstorm) while a longer less intense cell could be associated with a warm front. The positions of the rain cells are determined by a set of independent and identically distributed random variables representing the time intervals between the storm origin and the birth of the individual cells. The model structure is shown in Fig. 3 and is based upon the following assumptions:

- storm origins arrive in a Poisson process with the arrival rate represented by a parameter λ;
- each storm origin generates a (Poisson) random number C, with mean value ν, of raincells separated from the storm origin by time intervals that are each exponentially distributed with parameter β;
- the duration of each raincell is exponentially distributed with parameter η;
- the intensity of each raincell is exponentially distributed with parameter ξ;
- the rainfall intensity is equal to the sum of the intensities of all the active cells at that instant.

The parameters of the model can be summarised as follows:

1. λ^{-1} the average time between subsequent storm origins (h),

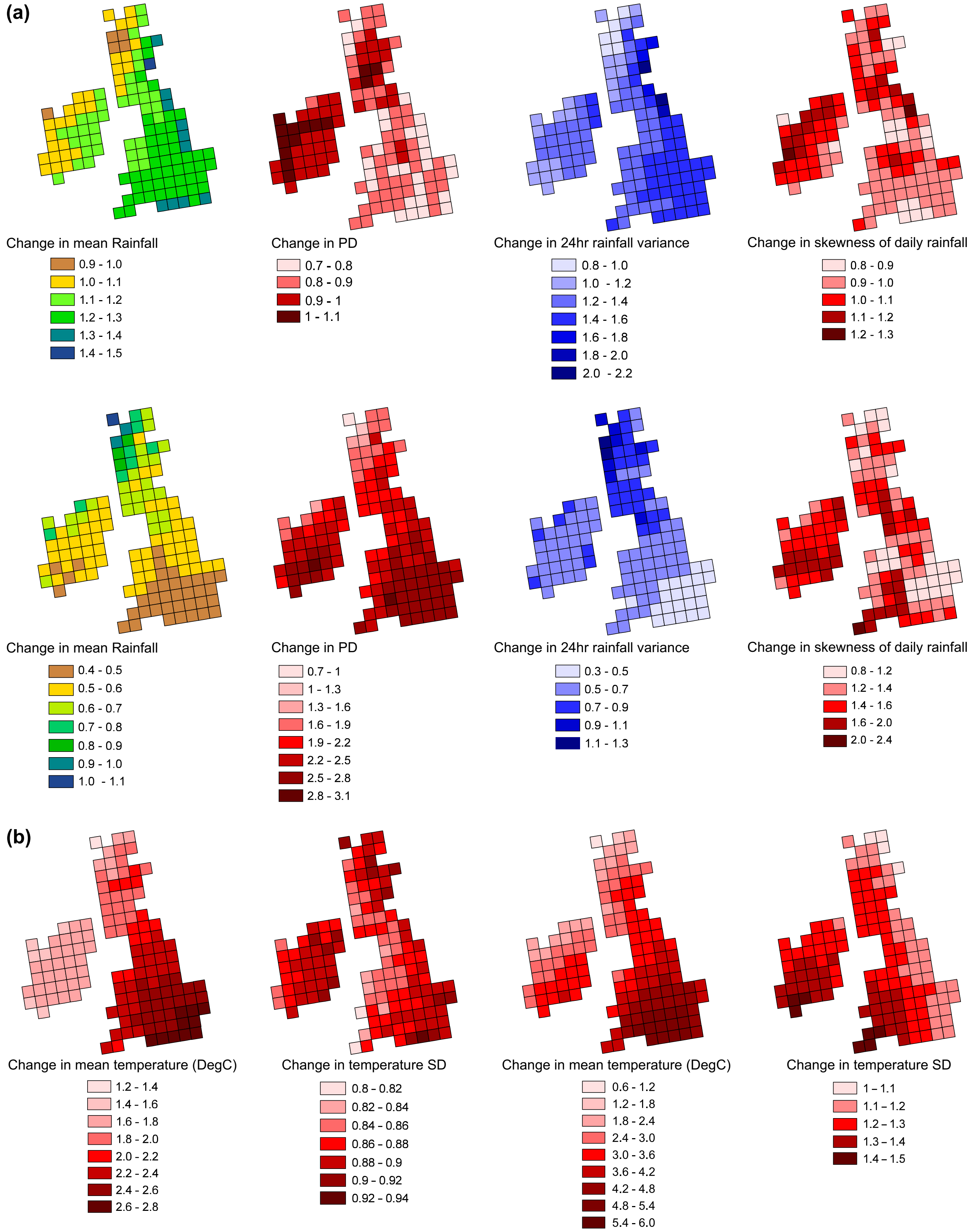

Fig. 2. (a) Change factors for daily rainfall statistics for the HadRM3H A2 SRES 2080s scenario (UKCIP02 Medium-High Scenario). Top row January, second row July. (b) Change factors for the daily temperature statistics, for the HadRM3H A2 SRES 2080s scenario (UKCIP02 Medium-High Scenario). Mean (left panels) and standard deviation (right panels). First and third are for January, second and fourth for July.

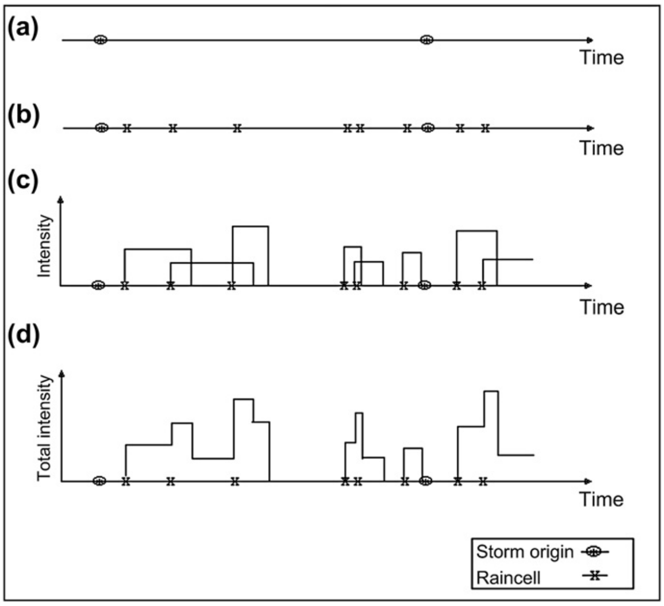

Fig. 3. A schematic representation of the NSRP model.

2. β^{-1} the average waiting time of the raincells after the storm origin (h),
3. η^{-1} the average cell duration (h),
4. ν^{-1} the average number of cells per storm,
5. ξ^{-1} the average cell intensity (mm/h).

3.2. Fitting procedure

There are five parameters of the NSRP model to be estimated, so the usual procedure would be to equate five statistical properties taken from the observed time series with their derived expressions for the model, and to solve the resulting set of simultaneous equations for the parameter estimates. The model would then fit five sample moments exactly, with the fit to other statistics not guaranteed. A more flexible fitting procedure is adopted here which assumes that it is more desirable to fit a larger set of sample moments approximately rather than a smaller set exactly.

In general, the parameters of the NSRP model can be estimated by selecting a set that matches, as closely as possible, the expected statistics of the generated time series with the corresponding statistics estimated from observed rainfall time series. To implement this, the model parameters are estimated by minimising the weighted sum of squared differences, D_i, between the statistics of the observed time series and the expected model statistics for each month, i, in turn:

$$D_i = \sum_{h \in G} w_h \left(1 - \frac{\widehat{f}_{i,h}(\lambda_i, \beta_i, \eta_i, \nu_i, \xi_i)}{f_{i,h}} \right)^2 \quad (7)$$

where G is a set of statistics, h, each with a specified aggregation level, w_h is the weight applied to statistic h chosen to give better fits to particular statistics according to the requirements of different applications, $\widehat{f}_{i,h}(\ )$ is the expected value of h for the NSRP model using a given set of parameters and $f_{i,h}$ denotes the sample estimate of h evaluated from observed data for month i. A numerical optimizing routine, such as the Simplex algorithm, is used to find the parameter set that minimizes the D_i function subject to fixed upper and lower bounds applied to the parameters.

4. Weather generator

4.1. Background to weather generators

Weather generators generally have a similar structure, with precipitation considered to be the primary variable (Wilks and Wilby, 1999). Depending on whether the day is wet or dry, other meteorological variables such as maximum and minimum temperatures, sunshine/cloudiness, vapour pressure and windspeed are determined by regression relationships with precipitation and values of the variables on the previous day. The regression relationships maintain both the cross- and auto-correlations between and within each of the variables. The success of the procedure in producing realistic weather sequences, therefore, rests primarily on the method of precipitation generation (Hutchinson, 1986). There is a vast literature on stochastic rainfall simulation (see previous section) which is ignored in much of the recent developments in weather generation, which therefore do not always use the best available precipitation generation models.

The well-known weather generator developed by Richardson (1981) is available as computer software (Richardson and Wright, 1984). Although correctly stating that any daily type of precipitation generator can be used, the Richardson model incorporates the simplest, a first order Markov chain with only two states, wet or dry, and an exponential distribution to select precipitation amounts on wet days. The main rationale behind this choice was simplicity: few parameters, all of which are capable of being estimated from both neighbouring sites and smoothed for each day of the year from monthly estimates.

In a previous version, the WG developed here used a two state Markov Chain to determine whether a day is wet or dry, with the two states being conditional on whether the previous day was wet or dry. The amount of rain on wet days was determined using a gamma distribution derived from wet day rainfall totals. However, for most rainfall regimes, it is widely recognised that the clustered nature of rainfall occurrence is better modelled by the structure of the more sophisticated and complex clustered point process models such as Neyman-Scott and Bartlett-Lewis; "the Markov model is inappropriate due to event clustering or other phenomena" (Srikanthan and McMahon, 2001). The LARS-WG approach (Semenov and Brooks, 1999) uses a semi-empirical approach that, although an improvement upon a basic Markov process, is still unable to provide a correlation between precipitation amounts on successive wet days, the reproduction of dry spells or extreme rainfall events. Clustered point process based models such as the NSRP are able to provide this "event clustering" and thus also reproduce daily variability more reliably. A comparison is presented in an example application in

Section 6. In this latest development, the WG is being driven by the NSRP rainfall model, providing a powerful combination of the two models.

Although the present tool uses the NSRP rainfall model (see Section 3), it is important to put the generation of the other variables in context. The most important issue is not that the model should fit the data distribution well, but that the model should *validate* well (i.e. produce realistic sequences, particularly when extremes are considered for data not used in model calibration). Normally, statistical properties (means, variability on different time scales, frequencies of extremes, length of dry and wet spells, etc.) are compared between the observed and simulated data. Inadequate validation may lead to a false sense of security regarding the model's veracity and realism. For example, although synthetic data may agree well in the distribution of precipitation amounts on wet days, giving what appear to be good fits in terms of monthly precipitation and wet day totals, low frequency aspects of the synthetic data are often poorly reproduced. Both Gregory et al. (1992, 1993) and Cowpertwait (1991) have found that the variability of seasonal totals and the autocorrelation of daily amounts are both lower when compared with similar measures from observational series. These are serious deficiencies that affect other weather variables, which are ultimately dependent upon the accuracy of the precipitation generation.

4.2. Generating the other variables

Once the precipitation sequence has been generated, other weather variables can be generated, maintaining, where possible, observed relationships between the variables. In the Richardson generator this is achieved by having two states, wet and dry. Each secondary variable is normalized by removing the appropriate mean and standard deviation for that time of year; there being two different distributions, one for wet days and one for dry days. Regression relationships then generate the variables for each of the states, maintaining the cross-correlations and the lag auto-correlations between the variables.

Nicks and Harp (1980) had earlier introduced additional states by considering four types of days determined by the wet/dry status of the preceding and the current day, i.e., wet-wet, dry-dry, wet-dry and dry-wet. This introduces many more parameters and thus attempts have been made to reduce the number of parameters (Bruhn et al., 1980) and to use sine curves (Larsen and Pense, 1981) to model the seasonal dependence. In this development the four states defined by Nicks and Harp (1980) will be used.

Daily precipitation series are generated separately from the other variables because of the fundamental difference in character between precipitation and other variables. Other meteorological variables have continuous ranges of variation that are more reliably generated as autoregressive processes.

The program generates meteorological data for the following five variables:

T daily mean temperature
R daily temperature range
VP vapour pressure
S sunshine duration
W wind speed

These five variables are sufficient to calculate potential evapotranspiration (PET) using the FAO-modified (Food and Agriculture Organization of the United Nations) version of the Penman method (described in Ekström et al., in press). Modelling mean daily temperature (T) and temperature range (R) has advantages compared to the equivalent method of modelling maximum and minimum temperature separately. For instance, the temperature range is likely to be high on a dry, sunny day and low when wet and cloudy. These temperatures are related by $T_{MAX} = T + 0.5R$; $T_{MIN} = T - 0.5R$. Vapour pressure is easier to model than relative humidity as its distribution is near normal. If relative humidity is required, it is easily calculated from vapour pressure using the saturation vapour pressure at the mean temperature.

The model development procedure for these variables is best described as a stepwise procedure:

1. The seasonal cycles of both the mean and the standard deviation of all five variables are removed. This is achieved by dividing each month into two parts with the mean and daily standard deviation calculated for 24 (12×2) half monthly periods. This is performed separately for the four transition states (DD, WW, DW and WD) using all days in each half month. All variables are then reduced to time series of normalised values (residuals) that have a mean of zero and a standard deviation of unity. The mean and standard deviation is based on all available data in each transition state for each half month: for this application typically 30 years were available. Note that a normal distribution may not be ideal for sunshine duration, but experimentation has not yet revealed a better approximation.

2. For both daily mean temperature and the temperature range, the residual time series are modelled as first-order autoregressive processes (the adequacy of which has been amply demonstrated in the literature, even for the simulation of temperature extremes – see for example, Mearns et al., 1984). To accommodate the four transition states, four equations and associated regression and correlation coefficients are calculated. The models are:

Dry periods (current day dry, previous day dry; DD):

$$T_i = a_1 T_{i-1} + b_1 + e \tag{8}$$

$$R_i = a_2 R_{i-1} + b_2 + e \tag{9}$$

Wet periods (current day wet, previous day wet; WW):

$$T_i = a_3 T_{i-1} + b_3 + e \tag{10}$$

$$R_i = a_4 R_{i-1} + b_4 + e \tag{11}$$

Dry/wet transition (current day wet, previous day dry; DW):

$$T_i = a_5 T_{i-1} + a_6 P_i + b_5 + e \tag{12}$$

$$R_i = a_7 R_{i-1} + a_8 P_i + b_6 + e \tag{13}$$

Wet/dry transition (current day dry, previous day wet; WD):

$$T_i = a_9 T_{i-1} + a_{10} P_{i-1} + b_7 + e \tag{14}$$

$$R_i = a_{11} R_{i-1} + a_{12} P_{i-1} + b_8 + e \tag{15}$$

All the regression weights (a_1 to a_{12}, b_1 to b_8) have been determined by regression analysis using the observed data. T_i, R_i and P_i are respectively mean temperature (°C), daily temperature range (°C) and mean precipitation (mm) on day i, and suffix $i-1$ indicates the previous day's value. All the e's are independent standard normal (Gaussian) variables (scaled by the degree of fit or explained variance of each regression) and are selected randomly when the models are used in simulation or "weather generation" mode. The random numbers were chosen using the machine independent generator RAN1 given by Press et al. (1992).

3. The remaining variables (X) are determined by regression analyses of the form:

$$X_{ij} = c_j + d_j P_i + e_j T_i + f_j R_i + g_j X_{i-1,j} + e$$

where $j = 1{,}3$ corresponds to vapour pressure, sunshine duration, and wind speed. This general form ensures that the simulated data will have the correct autocorrelation structure. Correlations between these three variables, temperature and precipitation (which are generally quite high) will also be correctly simulated, and correlations between the different X_{ij} will arise naturally through the common dependences on P_i, T_i and R_i. All generated variables are then transformed back to absolute values using the appropriate means and standard

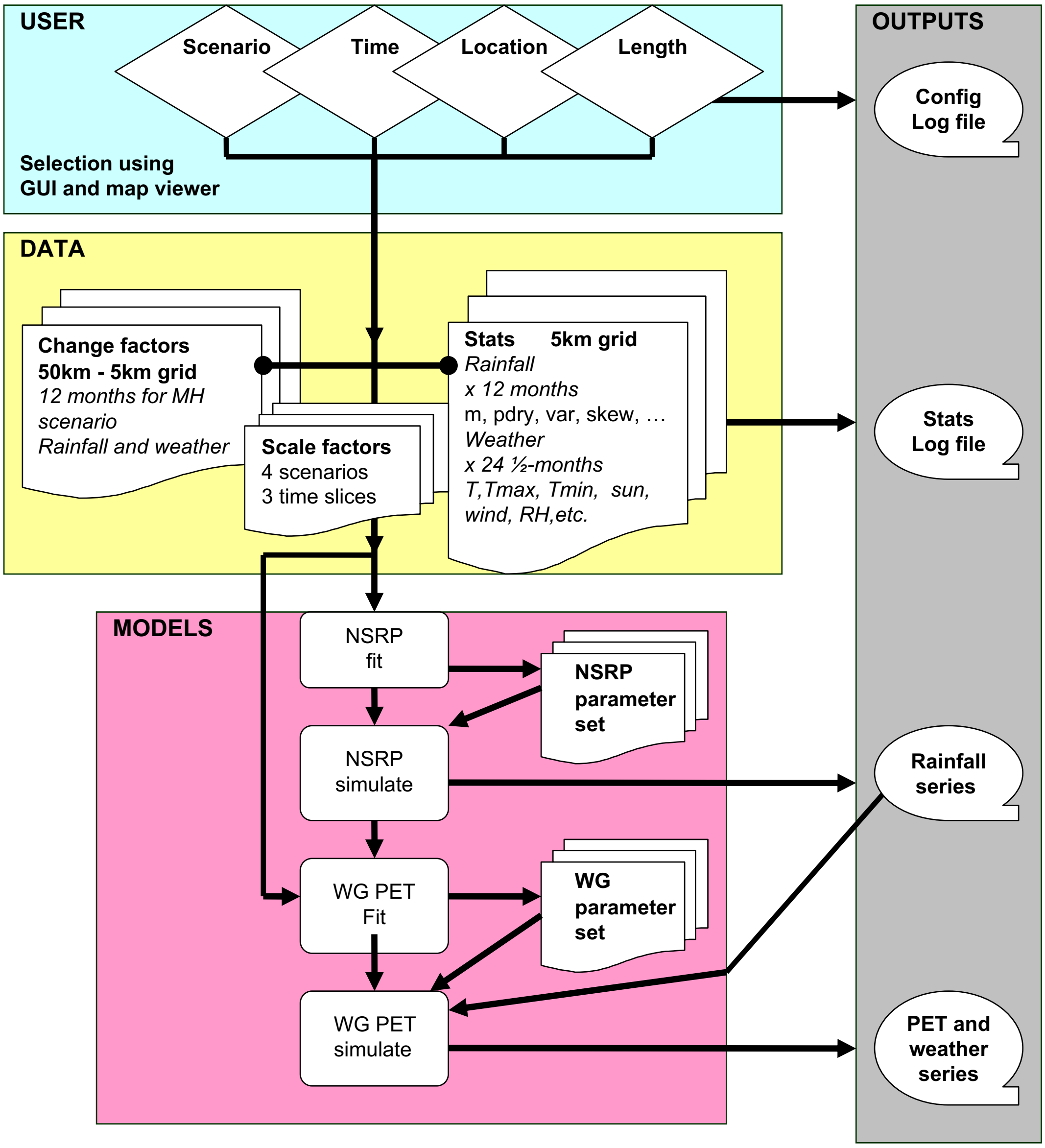

Fig. 4. Schematic of operation of weather generator including map viewer.

deviations. Finally, PET is calculated using the FAO-modified version of the Penman's formula.

The method is illustrated by an example application in Section 6.

4.3. Parameter adjustment for future climate applications

For the future, all regression weights and explained variances (which determine the size of the error terms) are assumed not to change. The only parameters that are changed are the mean and standard deviation for each half month. Both GCMs and RCMs indicate little change in future temperature range, vapour pressure or wind speed for the UK (see e.g. Hulme, 2002 using HadRM3H). Potential future changes in these variables are also likely to differ between the many available RCM integrations. As maximum sunshine hours cannot increase, it is only necessary to perturb the mean and standard deviation of the temperature. The mean is changed by a simple increment: the difference between the future and control RCM temperature (scaled appropriately for the scenario; see Table 1). The standard deviation (SD) is perturbed by the ratio of the future to the control SD values from the RCM (then scaled for the scenario according to Table 1).

4.4. Implementation

The application of the weather generator at any location uses the generated rainfall sequence from Section 3, together with the parameters determined from Section 4.2. To extend this to 5 km resolution across the country, daily climate data for 115 stations across the UK are used. Spatial regression relationships were developed between each of the half-month means and standard deviations, regression weights and explained variances (the a's to g's in Section 4.2) for the 115 stations and geographic variables. The geographic variables used were elevation, easting and northing (British National Grid co-ordinates) and distance from coast. All the necessary information required to run the weather generator for any 5 km square can then be developed from the same geographic data for each square. As the 5-km resolution data are available gridded at monthly resolution for 1958–2002, comparison can be undertaken at this timescale for all the variables, in terms of monthly means and standard deviations.

For spatial applications, a simple averaging of the weather generator parameters (the means, standard deviations, regression weights and explained variances for each half month) across a number of 5 km squares will then generate areally representative weather. As the principal determinant of the quality of the generated series is the precipitation sequence, an upper limit to the area must be imposed, because use of point rainfall as an approximation to an areal average will incur errors with increasing catchment area. Choice of this upper limit is difficult to define strictly, and depends on the nature of the application and topographic variation within the catchment. However, variogram analysis of observed rainfall (Skøien et al., 2003) showed e-folding distances of about 45 km for point precipitation, which suggests 1000 km^2 is a conservative estimate for the upper limit.

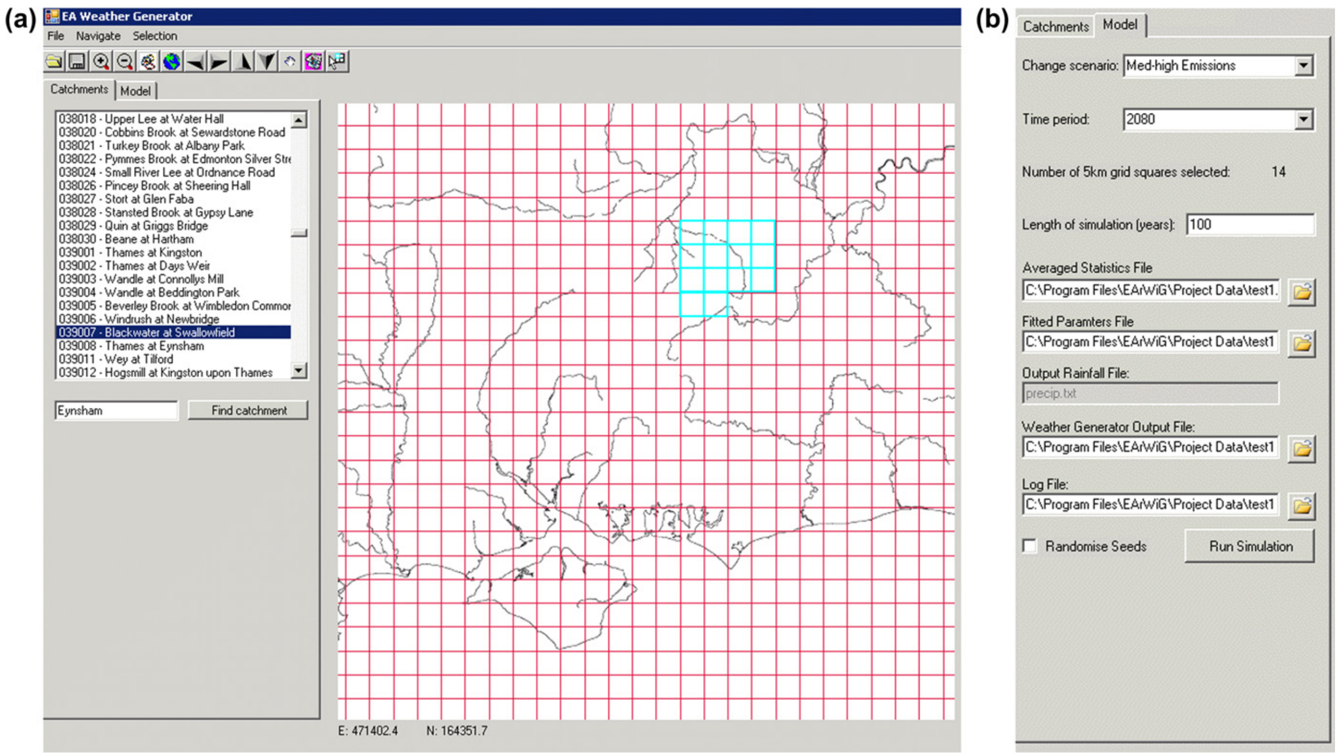

Fig. 5. (a) Screenshot of map viewer with 5 km grid and main river overlay, showing selection of a catchment. (b) Screenshot of map viewer showing scenario selection.

5. Software implementation

5.1. Overall structure

The system is implemented in Microsoft Windows compatible software with a graphical user interface (GUI). The NSRP and WG models are incorporated as separate executables, allowing individual development. The model parameters are contained in a database (Access) allowing rapid modification or replacement for new climate scenarios, and flexible reading by a variety of applications.

The software structure is as shown in Fig. 4. A key point is the modularity of the data structure and model executables, allowing flexible upgrade and replacement of data as new

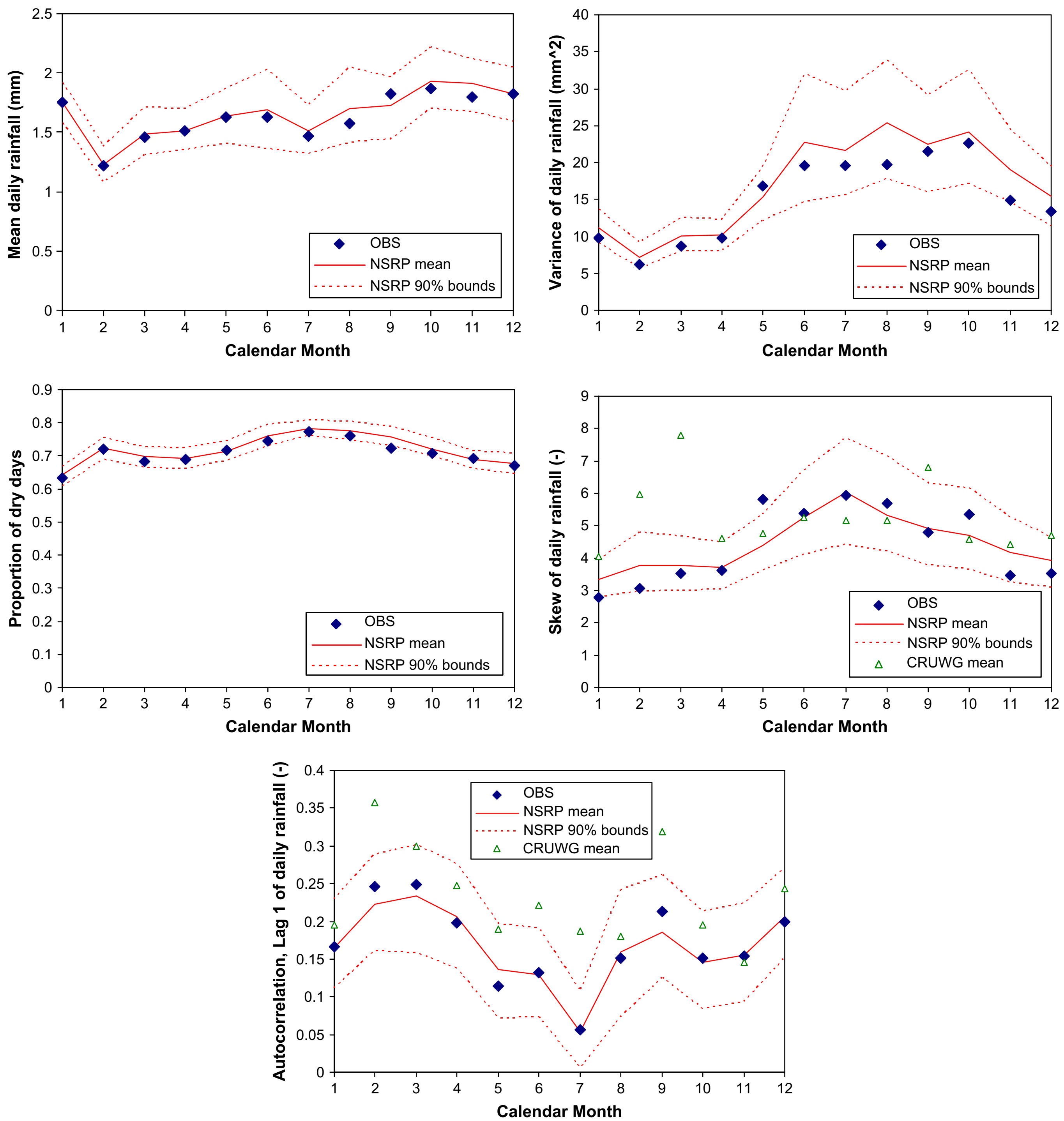

Fig. 6. Performance of NSRP rainfall model in reproducing observed rainfall statistics for Heathrow. Calculated from 100 30-year simulated series: mean and 10 and 90 percentile bounds are plotted. The CRU WG simulated statistics are also plotted for skew and autocorrelation at lag1, where the NSRP is markedly superior: performance of the CRU WG rainfall model is similar to the NSRP model for other statistics.

scenarios become available. This flexibility is promoted by the inclusion of rainfall and weather variable statistics in the package (in standard database format). These are used to estimate model parameters at run-time. An alternative approach would be to pre-calculate the model parameters for each scenario/time-slice/location. This is restrictive on two counts:

- scenarios must be pre-defined, so that for example, a later extension to probabilistic scenarios would not be supported (e.g. the UKCIPnext initiative);
- model parameters cannot in general be area-averaged, so catchment-averaged outputs are not supported. On the other hand it is possible to area-average (some) rainfall and weather variable statistics, and work will be done to implement this (e.g. by calculating variation of rainfall Pdry with catchment area).

5.2. Map viewer

A map viewer with pan and zoom capability is provided to allow easy selection of the area of interest for scenario generation. The viewer shows a national map overlaid by major rivers, catchment boundaries, major towns and the 5 km grid as appropriate (see Fig. 5).

The mapviewer provides a number of intuitive methods for selecting an area of interest. Users' selections can be named and saved in a *config* file for future repeated use. Areas of interest for catchments have been predefined. These can be loaded and edited by the end user.

A simple rule for selecting the appropriate 5 km grid squares for catchments is used: if the centroid of the square is within the catchment polygon. This is adequate given the interpolation of the UKMO 5 km data, the 50 km resolution of the HadRM3 change factors and the spatial interpolation of many climate variables (e.g. wind).

Catchment definition and averaging are performed, with an upper limit (currently set at 1000 km^2) to preclude the use of very large areas with too much heterogeneity of rainfall or other weather variables to define a meaningful average.

6. Example application

The performance of the NSRP rainfall model is illustrated in Figs. 6 and 7 for Heathrow. Fig. 6 shows validation against rainfall statistics used in fitting of the model, and performance is good throughout. The weighting of statistics can be changed in the fitting procedure to improve the reproduction of any particular statistic, usually at the expense of worse fits elsewhere.

Comparisons are presented with the Markov rainfall model in cases where the NSRP model performs significantly better. The reproduction of autocorrelation and skewness is compared in Fig. 6 where it can be seen that the NSRP model is superior. Fig. 7 shows an assessment of performance in estimating extreme values which are not explicitly used in the fitting procedure. It can be seen that the NSRP model performs better

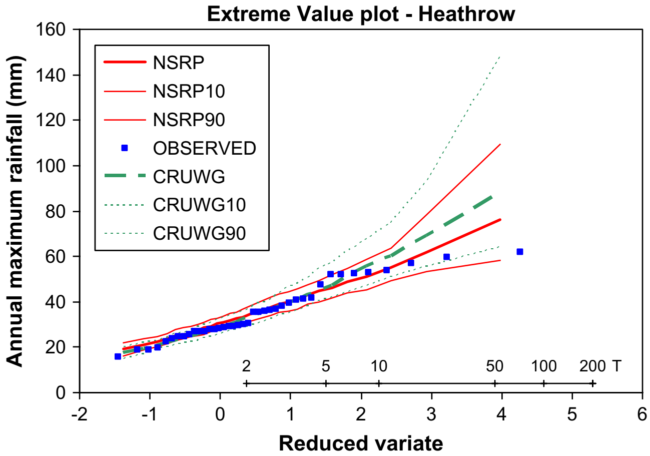

Fig. 7. Extreme value plot of annual maximum rainfall for Heathrow. 100 series of 30 years data were generated for each rainfall model (NSRP and CRU WG) and the 10 percentile, mean and 90 percentile curves plotted. The CRU WG rainfall model underestimates the observed maxima at higher return periods, and the high upper bound indicates that some unrealistically large maxima are being generated.

than the Markov model which overestimates the annual maxima, generating some unrealistically large values.

The model has been validated at higher time aggregations, e.g. monthly and annual. For rainfall some reduction in variance is found for annual totals, which has a concomitant effect on annual mean temperature and evaporation, etc. This may be significant for some applications, and it may be necessary to model these time scales separately and condition the model externally. It should of course be noted that climate models cannot in general reproduce low frequency variability either.

Fig. 8a,b illustrate the performance of the weather generator for average values of maximum and minimum temperature, sunshine duration, wind speed, vapour pressure and PET. The cross indicates the observed average value for 1961–1990 for each half month. The error bar range encompasses 90% of the same averaged values from 100 generated sequences of 30 years. Most of the observed values fall within the generated ranges, although winter sunshine durations tend to be too great in the model.

Fig. 8c shows the performance of the generated data for the estimation of temperature extremes. A software package for the calculation of a range of extremes from daily precipitation and temperature has been developed through international collaboration over the last 5 years. The software can be downloaded from http://cccma.seos.uvic.ca/ETCCDMI/software.shtml and the indices are also described by Alexander et al. (2006). Here average values for 1961–1990 for four indices are shown. The first three are the number of hot days (estimated from maximum temperatures) and warm and cold nights (from minimum temperatures). Site specific thresholds are estimated for each season based on the 10th and 90th percentiles of all observed days in each season for the 30 years. As each season has approximately 90 days, the average number of days above/below these percentiles is roughly nine for

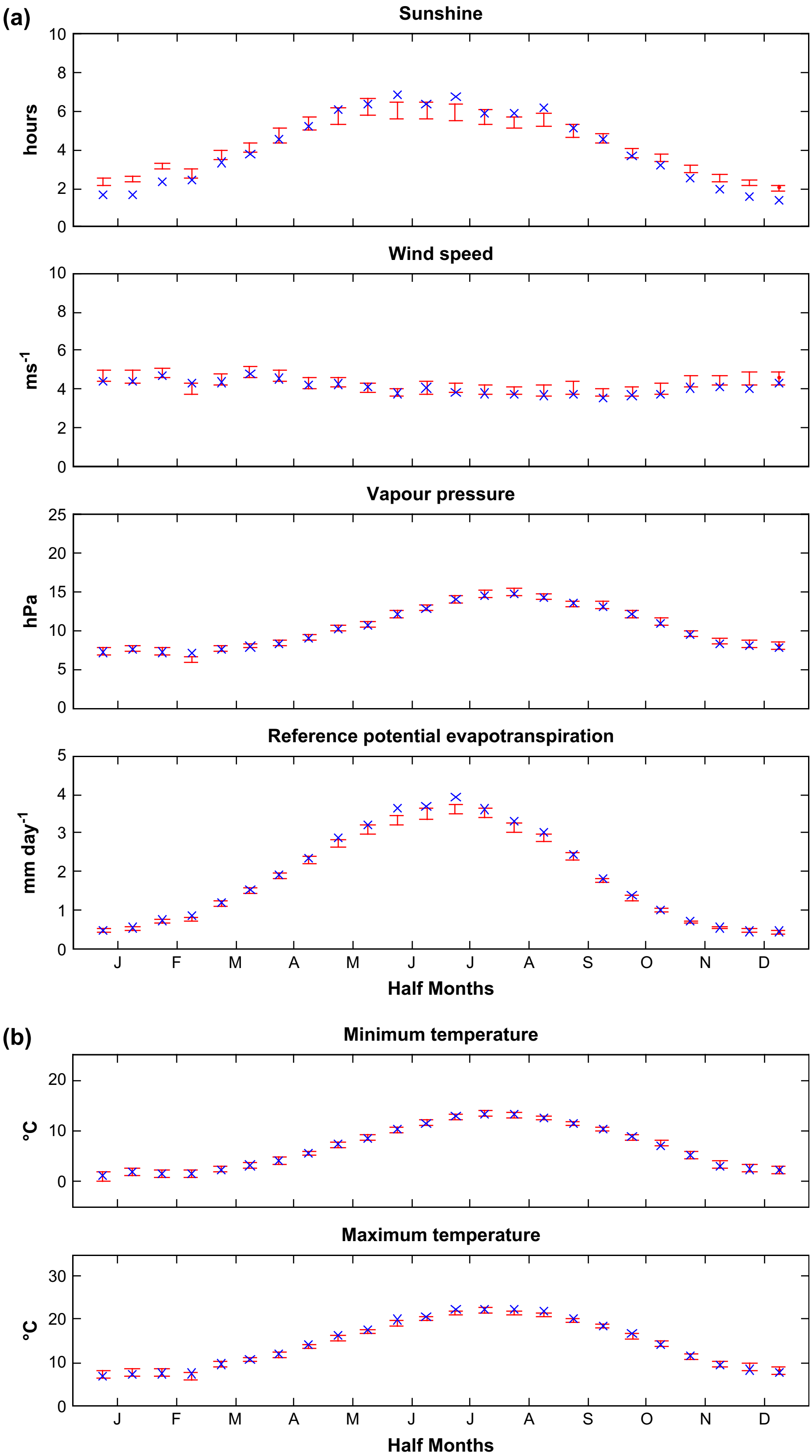

Fig. 8. (a) Validation of weather generator for Heathrow (1961–1990). Crosses are the observed values, bars show mean and 5 to 95 percentile range for 100 simulated 30 year series. (b) Validation of weather generator for Heathrow (1961–1990) minimum and maximum temperature. Crosses are the observed values, bars show 5–95 percentile range for 100 simulated 30 year series. (c) Validation of the extremes of the weather generator for Heathrow (1961–1990). Crosses are the observed values, bars show mean and 5–95 percentile range for 100 simulated 30 year series.

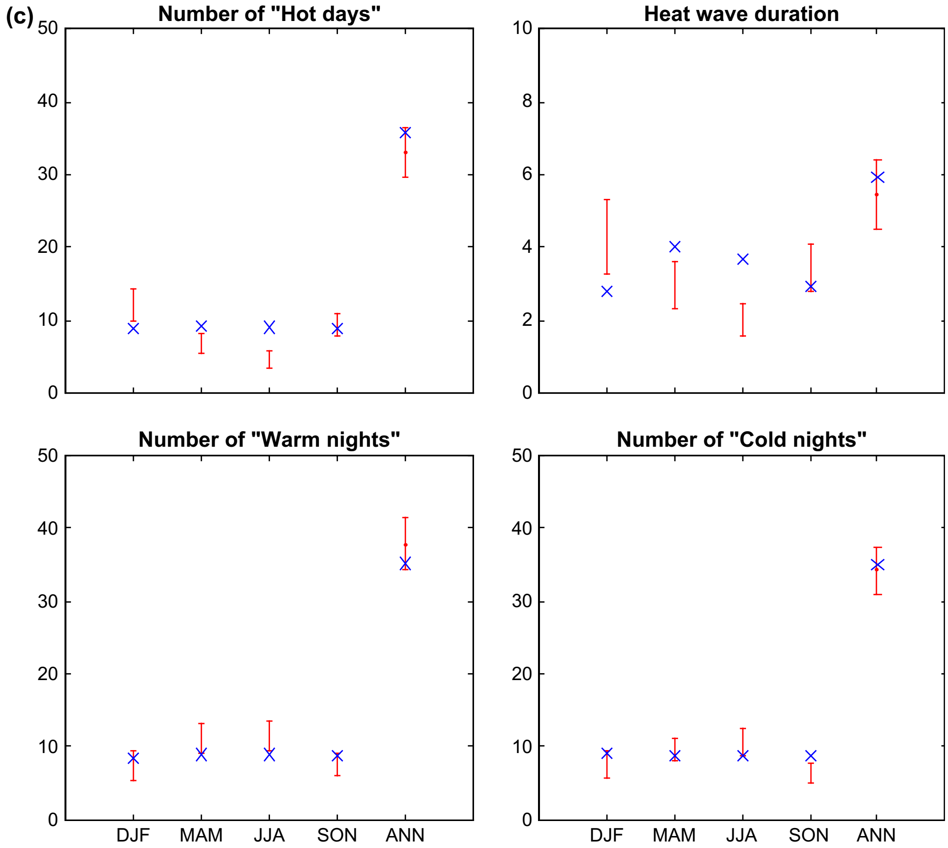

Fig. 8. (*continued*)

the observed data (plotted with crosses). The same thresholds are used for the generated sequences, with the ranges shown with similar error bars to Fig. 8a,b. The annual counts are the sum of the seasonal counts. The fourth index looks at spells of warm days for the time of year and is called the Heatwave Duration Index. Here spells of maximum temperature above the 90th percentile threshold for each season are counted if the duration is 6 days or more. Here the average value of this index for the 30 years is plotted. The index has a value less than 6 as many seasons do not have any durations above 5 days. The annual value of the Heatwave Duration Index is based on the seasonal counts, but takes the annual maximum value rather than a summation as for the other three indices. For example, if one season has a value of 6 and the other three are zero, the annual value will be 6.

When the weather generator is used for future extremes, the same software is used, but with the observed thresholds from the 1961–1990 observations. Future simulations can then be compared with present conditions, for both average as well as a range of extreme indices.

7. Discussion and conclusions

A combination and implementation of models, data and methodology has been described, which is capable of generating self-consistent series of meteorological variables. These variables comprise precipitation, temperature, vapour pressure, windspeed, and sunshine hours. Additionally, potential evapotranspiration is derived using these variables. The system produces time series at a daily resolution using two stochastic models in series. First, a rainfall model produces a series which is then used as the input for a weather generator producing temperature, vapour pressure, sunshine duration and windspeed values. These series are for single sites (points) defined across the UK at a 5-km resolution. However, by fitting the model to area-average statistics rather than point statistics, series representative of small catchments (<1000 km^2) can be generated.

Weather time series can be generated for a control period (1961–1990) based on observed data, and for future emissions scenarios at various time slices (2020s, 2050s and 2080s). Future scenarios are generated by fitting the models to observations that have been perturbed by application of change factors derived from mean projected changes in that variable from climate models. These change factors can readily be updated, as new scenarios and simulations from different climate models become available.

Initial development of the tool is customised for the UK, but the overall framework is readily transportable to other regions. A version with a simpler rainfall model has been applied successfully across Europe, suggesting general applicability in mid-latitudes. The approach may however be expected to be limited geographically for the following reasons:

- poor data availability in some regions, particularly since consistent daily records of multiple variables are needed;
- precipitation occurring with heavier-tailed distributions (e.g. tropical storms) where alternative rain cell intensity distributions can be used;
- precipitation occurring in mixed (or bimodal) distributions, e.g. from distinct mechanisms or directions: a more sophisticated two-cell type rainfall model can be used in these cases;
- long dry seasons where conditioning by wet/dry state is ineffectual (too few samples within the available precipitation record).

A number of further developments and extensions to the WG framework is envisaged beyond this project, and readily achievable. These include:

- *Further or new climate scenarios:* the system is now being further developed to provide time series for the ''UKCIP08'' set of future climate scenarios for the UK developed by the UK Met Office Hadley Centre. This development implements a probabilistic framework, where users will be able to select scenarios by their likelihood of occurrence from calculated probability density functions of a given weather variable, and so are not limited to scenarios associated with a particular single RCM simulation.
- *Hourly variables:* the rainfall and WG models already have an hourly capability for a limited number of sites, so extension of the UK model to the hourly level is straightforward provided that hourly meteorological data can be sourced for model calibration.
- *Incorporation into GIS:* operational use would be facilitated by embedding the modelling system within a GIS such as ARCGIS™, where full use can be made of outputs within catchment modelling frameworks using other live data sets.
- *Spatial models:* whilst the models described here can be used for small catchment-average applications (e.g. less than 1000 km^2), they are essentially single-site models applied over aggregated areas. True spatial or multi-site models would produce multiple series at sites within a catchment and allow simulation of larger catchment areas with spatially consistent inputs of, for example, rainfall. This would be a requirement for simulation of flood events where the spatial distribution and timing of rainfall within a catchment or collection of sub-catchments is important.

Acknowledgements

Development of this system was supported by UKCIP/EPSRC as part of the BETWIXT project and by the Environment Agency under Science Project SC030301 ''Environmental Effects of Agriculture & Land Use: Weather Generator Tool''. The HadRM3H data were supplied by the Climate Impacts LINK project (DEFRA Contract EPG 1/1/154) on behalf of the Hadley Centre and UK Meteorological Office. The 5-km gridded rainfall and weather variable data set has been used under license from the UK Met Office. We gratefully acknowledge the constructive comments of the reviewers.

References

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Ambenje, P., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., Vazquez-Aguirre, J.L., 2006. Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res. 111 (d5), D05109, doi:10.1029/2005JD006290.

BETWIXT Project website. http://www.cru.uea.ac.uk/cru/project/betwixt.htm.

Bruhn, J.A., Fry, W.E., Fick, G.W., 1980. Simulation of daily weather data using theoretical probability distributions. J. Appl. Met 19, 1029–1036.

Cowpertwait, P.S.P., 1991. Further developments of the Neyman-Scott clustered point process for modelling rainfall. Water Resour. Res. 27, 1431–1438.

Cowpertwait, P.S.P., O'Connell, P.E., Metcalfe, A.V., Mawdsley, J.A., 1996. Stochastic point process modelling of rainfall. 2. Regionalisation and disaggregation. J. Hydrol 175, 47–65.

Cowpertwait, P.S.P., Kilsby, C.G., O'Connell, P.E., 2002. A space-time Neyman-Scott model of rainfall: empirical analysis of extremes. Water Resour. Res. 38 (8), 1131, doi:10.1029/2001WR000709.

Ekström, M., Jones, P.D., Fowler, H.J., Lenderink, G., Buishand, A., Conway, D. Regional climate model data used within the SWURVE project. 1: Projected changes in seasonal patterns and estimation of PET. Hydrol. Earth Syst. Sci., in press.

Fowler, H.J., Kilsby, C.G., 2004. Future increases in UK water resource drought projected by a regional climate model. Proceedings of the BHS International Conference on Hydrology: Science & Practice for the 21st Century, Volume 1, London, 12–16 July 2004, pp. 15–21.

Fowler, H.J., Ekström, M., Kilsby, C.G., Jones, P.D., 2005. New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1: Assessment of control climate. J. Hydrol 300, 212–233.

Gregory, J.M., Wigley, T.M.L., Jones, P.D., 1992. Determining and interpreting the order of a two-state Markov chain: application to models of daily precipitation. Water Resour. Res. 28, 1443–1446.

Gregory, J.M., Wigley, T.M.L., Jones, P.D., 1993. Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. Clim. Dyn 8, 299–310.

Hulme, M., Jenkins, G.J., Lu, X., Turnpenny, J.R., Mitchell, T.D., Jones, R.G., Lowe, J., Murphy, J.M., Hassell, D., Boorman, P., McDonald, R., Hill, S., 2002. Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report, Tyndall Centre for Climate Change Research. School of Environmental Sciences, University of East Anglia, Norwich, UK, 120 pp.

Hutchinson, M.F., 1986. Methods of generation of weather sequences. In: Bunting, A.H. (Ed.). Agricultural Environments C.A.B. International, Wallingford, pp. 149–157.

Kilsby, C.G., Moaven-Hashemi, A., O'Connell, P.E., September, 2004. Simulation of rainfall extremes: fitting to observed annual maxima, First International Conference on Flood Risk. Institute of Mathematics and its Applications, University of Bath, UK. 7–8.

Larsen, G.A., Pense, R.B., 1981. Stochastic simulation of daily climatic data. SRS Staff Report AGE810831. Statistical Reporting Service, Washington, DC, USDA.

Mearns, L.O., Katz, R.W., Schneider, S.H., 1984. Extreme high-temperature events: changes in their probabilities with changes in mean temperature. J. Clim. App. Meteorol 23, 1601–1613.

Nicks, A.D., Harp, J.F., 1980. Stochastic generation of temperature and solar radiation data. J. Hydrol 48, 1–17.

Onof, C., Chandler, R.E., Kakou, A., Northrop, P., Wheater, H.S., Isham, V., 2000. Rainfall modelling using Poisson-cluster processes: a review of developments. Stoch. Environ. Res. Risk Assess 14 (6), 384–411.

Osborn, T.J., Hulme, M., Jones, P.D., Basnett, T.A., 2000. Observed trends in the daily intensity of United Kingdom precipitation. Int. J. Climatol 20, 347–364.

Perry, M., Hollis, D., 2005a. The development of a new set of long-term climate averages for the UK. Int. J. Climatol 25 (8), 1023–1039, doi:10.1002/joc.1160.

Perry, M., Hollis, D., 2005b. The generation of monthly gridded datasets for a range of climatic variables over the UK. Int. J. Climatol 25 (8), 1041–1054, doi:10.1002/joc.1161.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing, second ed. Cambridge University Press, New York.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190.

Richardson, C.W., Wright, D.A., 1984. WGEN: A Model for Generating Daily Weather Variables. US Dept. of Agriculture, 83 pp. AR5–8.

Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1987. Some models for rainfall based on stochastic point processes. Proc. R. Soc. Lond. A410, 269–288.

Semenov, M.A., Brooks, R.J., 1999. Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain. Clim. Res. 11, 137–148.

Skøien, J.O., Bloschl, G., Western, A.W., 2003. Characteristic space scales and timescales in hydrology. Water Resour. Res. 39 (10), 1304, doi:10.1029/2002WR001736.

Srikanthan, R., McMahon, T.A., 2001. Stochastic generation of annual, monthly and daily climate data: a review. Hydrol. Earth Syst. Sci. 5, 653–670.

Velghe, T., Troch, P.A., De Troch, F.P., Van de Velde, J., 1994. Evaluation of cluster-based rectangular pulses point process models for rainfall. Water Resour. Res. 30, 2847–2857.

Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic weather models. Prog. Phys. Geog 23, 329–357.