

# Multi-site simulation of precipitation by conditional resampling

R. L. Wilby<sup>1,\*</sup>, O. J. Tomlinson<sup>2</sup>, C. W. Dawson<sup>3</sup>

<sup>1</sup>Department of Geography, King's College London, London WC2R 2LS, United Kingdom

<sup>2</sup>Division of Geography, University of Derby, Derby DE11 2GB, United Kingdom

<sup>3</sup>Department of Computer Science, Loughborough University, Loughborough LE11 3TU, United Kingdom

**ABSTRACT:** A single-site, regression-based downscaling method is extended to multi-site synthesis of daily precipitation at stations in Eastern England and the Scottish Borders. Area-averaged precipitation series for each region are downscaled using gridded predictor variables selected from a candidate suite representing atmospheric circulation, thickness and moisture content at length scales of 300 km. Three simulation methods are compared: (1) a deterministic model in which daily precipitation occurrence and amounts are conditioned by atmospheric predictor variables (*DET*); (2) a hybrid model in which the unexplained variance of the deterministic model is represented stochastically (*VAR*); and (3) an unconditional resampling procedure (*RND*). Downscaled daily area averages are, in turn, used to resample daily precipitation amounts at multiple sites contributing to the unweighted areal mean. The temporal dependence of precipitation amounts at individual sites was explored using the standard deviation and autocorrelation of daily amounts, and *N*-day winter maxima. Spatial dependency was examined using inter-site correlations, correlation decay lengths, and Kendall's  $\tau_b$  statistic for joint exceedance of *N*-day precipitation totals. The *DET* procedure underestimated the variance of daily amounts and *N*-day totals and over-estimated observed autocorrelation, leading to generally poor representations of spatial dependency. The *RND* procedure reproduced the distribution of daily amounts and inter-site correlations, but yielded poor representations of *N*-day exceedances for large *N*. Overall, the *VAR* procedure was the most successful downscaling approach. Even so, suggestions are made for refinements to *VAR* in order to better capture seasonal variations in decay distances and pairwise correlations of multi-site precipitation amounts.

**KEY WORDS:** Precipitation · Extreme events · Multi-site · Downscaling

— Resale or republication not permitted without written consent of the publisher —

## 1. INTRODUCTION

Realistic simulations of multi-site precipitation sequences are required for many hydrologic studies, including flood estimation and climate-change impact assessment (Hutchinson 1995, Wilks & Wilby 1999). Several approaches have been developed. For example, the normal distribution for single sites may be adapted to the multivariate situation using truncation of negative values to represent dry days (e.g. Bardossy & Plate 1992). Chain-dependent processes, in which the occurrence of precipitation is described by a Markov chain, may be extended to multi-site cases

using transformed, spatially correlated, normal variables to specify precipitation occurrence and amounts (e.g. Wilks 1998). Alternatively, non-homogeneous hidden Markov models simulate distinct patterns of multi-site precipitation occurrence and amounts conditional on a set of atmospheric predictor variables (e.g. Charles et al. 1999a, Hughes et al. 1999). Others employ linear and non-linear regressions on atmospheric circulation indices, geographic and topographic variables, or regionalization techniques to estimate the parameters of generalized spatial-temporal models (e.g. Cowpertwait & O'Connell 1997, Wilby et al. 2002a). Some question whether parametric models

can reproduce all aspects of spatial dependence of daily precipitation and have chosen to develop non-parametric resampling schemes. For example, involving the selection of daily precipitation amounts from observed data using analogue methods (e.g. Zorita et al. 1995) or re-sampling of nearest neighbours, conditional on simulated values for previous days (e.g. Rajagopalan & Lall 1999, Buishand & Brandsma 2001).

This paper deals with multi-site generation of daily precipitation at stations in Eastern England (EE) and the Scottish Borders (SB) using an extension to a hybrid regression/weather-generator model (Wilby & Dettinger 2000, Wilby et al. 2002b). The inter-site distances range from 13 to 291 km, and the site elevations from 2 to 253 m above mean sea level. The weather generator, Statistical DownScaling Model (SDSM), was initially developed for downscaling future climate-change scenarios at single sites given large-scale climate variables supplied by general circulation models (GCMs). Because of intended applications to extreme discharge simulation under present and future climate forcing, attention is now focused on the model's ability to reproduce temporal and spatial dependence of daily and  $N$ -day maximum precipitation amounts across multiple sites. Accordingly, model simulations are assessed using the temporal dependence of daily amounts, correlation decay lengths, and Kendall's  $\tau_b$  statistic for joint exceedance of 1 and 20 d winter-maximum precipitation amounts at pairs of stations. The 20 d event was chosen to interrogate lower-intensity, longer-duration events of the type resulting in the autumn 2000 UK floods (DEFRA 2001). In addition, distributions of observed and synthetic 5, 10, 20 and 60 d annual maxima are compared for critical flood seasons.

## 2. MODEL DESCRIPTION AND DATA

The SDSM approach to multi-site generation of daily precipitation involves 2 distinct steps. First, the generation of a 'marker' precipitation series for areal average amounts across multiple sites conditional on a set of large-scale atmospheric predictor variables. Secondly, the resampling of observed daily precipitation from constituent sites, conditional on the downscaled precipitation series of the marker site. This 2-stage process is described in more detail below, along with the station data and performance measures.

**2.1. Generation of daily precipitation series for the marker site.** Full technical details and split-sample tests of SDSM are provided by Wilby et al. (1999, 2002b), and Wilby & Dettinger (2000). Within the taxonomy of statistical downscaling techniques, SDSM is best described as a hybrid of the stochastic weather generator and regression-based methods, because

large-scale daily circulation patterns and atmospheric moisture variables ( $j = 1, 2, \dots, n$ ) are used to linearly condition local-scale weather-generator parameters (e.g. precipitation occurrence and intensity) at individual sites. In general:

$$\omega_i = \alpha_0 + \sum_{j=1}^n \alpha_j \hat{u}_i^{(j)} \quad (1)$$

where  $\omega_i$  is the conditional probability of precipitation occurrence on day  $i$ ,  $\hat{u}_i^{(j)}$  are normalised atmospheric predictor variables (see below), and  $\alpha_j$  are regression coefficients estimated for each month using least-squares regression. Wet- and dry-spell sequences are determined stochastically by comparing  $\omega_i$  with the output of a linear random-number generator,  $r_i$ . If it is found that precipitation occurs (i.e.  $\omega_i \leq r_i$ ), the conditional distribution consists of regressions of inverse-normal-transformed precipitation amounts at the site on the large-scale atmospheric circulation using:

$$Z_i = \beta_0 + \sum_{j=1}^n \beta_j \hat{u}_i^{(j)} + \varepsilon \quad (2)$$

where  $Z_i$  is the z-score,  $\beta_j$  are regression coefficients estimated for each month using least-squares regression, and  $\varepsilon$  is an error term which is modelled stochastically using a series of serially independent Gaussian numbers,  $\varepsilon \sim N(0, \sigma^2)$ , and:

$$Z_i = \phi^{-1}[F(y_i)] \quad (3)$$

in which  $\phi$  denotes the normal cumulative distribution function and  $F(y_i)$  is the empirical distribution function of  $y_i$ , the daily precipitation amounts (Charles et al. 1999a). The normal cumulative distribution is constructed from all daily amounts. In other words, separate  $\phi$  are *not* derived for each month so as to maximise the sample of large precipitation amounts available for resampling (see below). Note also that the same predictor variables,  $\hat{u}_i^{(j)}$ , are used for downscaling both precipitation occurrence and amounts, and that all predictors have been normalised with respect to their climatological mean,  $\bar{X}^{(j)}$ , and the standard deviations,  $\sigma^{(j)}$  (in this case using the period 1961–1990):

$$\hat{u}_i^{(j)} = \frac{x_i^{(j)} - \bar{X}^{(j)}}{\sigma^{(j)}} \quad (4)$$

Normalised predictors are routinely employed in order that the same models may, if necessary, be applied to future climate-scenario generation using normalised GCM output.

**2.2. Extension to multiple sites.** For multi-site applications, the single-site model is first used to generate a series of daily precipitation amounts at a 'marker' site. In the present study, the area average is used as the marker series, but precipitation occurrence at individual sites might also be used (as in Palutikof et

al. 2002). Wet-day amounts are resampled from the empirical distribution of area averages ( $y_i$ ), conditional on the large-scale atmospheric forcing and the stochastic error term (Eqs. 2 & 3). The actual amount is determined by mapping the modelled normal cumulative distribution value onto the observed cumulative distribution at the marker site. For example,  $Z_i = 0$  would map to the median value of the observed set of daily precipitation amounts, together with the specific date on which that amount fell. The date is, in turn, used to resample the actual amount from each location in the multi-site array (some of which may return zero amounts, despite the fact that the marker series indicates a wet day). Note that it is possible for a combination of January predictors to produce a value of  $Z_i$  that can only be matched to a precipitation amount in June.

If the marker series is based on an unweighted average of several sites (as here), the resampling procedure preserves both the areal average of the marker series and the spatial pattern of the multi-site array. Furthermore, by using areal averages instead of single sites as the marker series, the risk of employing a non-homogeneous record is reduced, and the signal-to-noise ratio of the predictand is increased. As with all other resampling methods, the maximum *daily* value(s) to be generated cannot exceed the maximum daily amount in the observations. However, synthetic *N*-day totals can exceed observed *N*-day totals if the atmospheric conditioning produces a sequence of  $Z_i$  previously unencountered in the training set.

**2.3. Study areas and data.** Daily rainfall data for 24 UK Meteorological Office stations were selected on the basis of their geographical coverage, completeness and reliability of records for the period 1961–1990 (Fig. 1). Half the sites were distributed across Southern and Eastern England (EE), the remainder across the Scottish Borders (SB). The station elevations were generally lower in EE than in SB (an average of 35 and 126 m above mean sea level, respectively). Conversely, station separations were on average greater in EE than in SB (149 and 107 km, respectively). Unweighted, area-averaged daily precipitation amounts were obtained from the 12 sites in each region. For the purpose of both the spatial analyses and downscaling, wet days were defined as 24 h periods with non-zero precipitation. The transition from imperial units to the metric system is unlikely to affect whether or not a day was classified as wet, given that the area average was derived from multiple sites.

All atmospheric predictor variables originate from the National Center for Environmental Prediction (NCEP) re-analysis data set (Kalnay et al. 1996), but were processed to conform to the  $2.5^\circ$  latitude  $\times$   $3.75^\circ$  longitude grid of the UK SDSM archive (Wilby & Daw-

son 2001), i.e. the Hadley Centre's HadCM3 model grid. The archive contains 32 daily predictors (describing circulation, thickness, and moisture content for 3 atmospheric levels), for 9 regions covering the British Isles, for the period 1961–2000. Although the mean sea-level pressures in the NCEP re-analysis are known to have a positive bias from 1941 to 1967, the worst affected areas lie outside the domain of the present study (see Reid et al. 2001). In addition, there is a lag interval between the timing of the precipitation day (ending at 09:00 h GMT) and the NCEP daily averages (indicative of conditions at midday). The influence on the predictand of the 'missing' 9 h can be accommodated by employing forward-lagged daily predictor variables in the downscaling; however, this refinement was not applied in the present investigation.

**2.4. Experimental design.** Two sets of experiments were undertaken. Firstly, the full 30 yr record was used to explore relationships between large-scale predictors and precipitation amounts at individual sites in order to elucidate the dominant physical controls in each region. This was undertaken via stepwise multiple linear regression *without* seasonal stratification of the data. The purpose of this experiment was to identify,

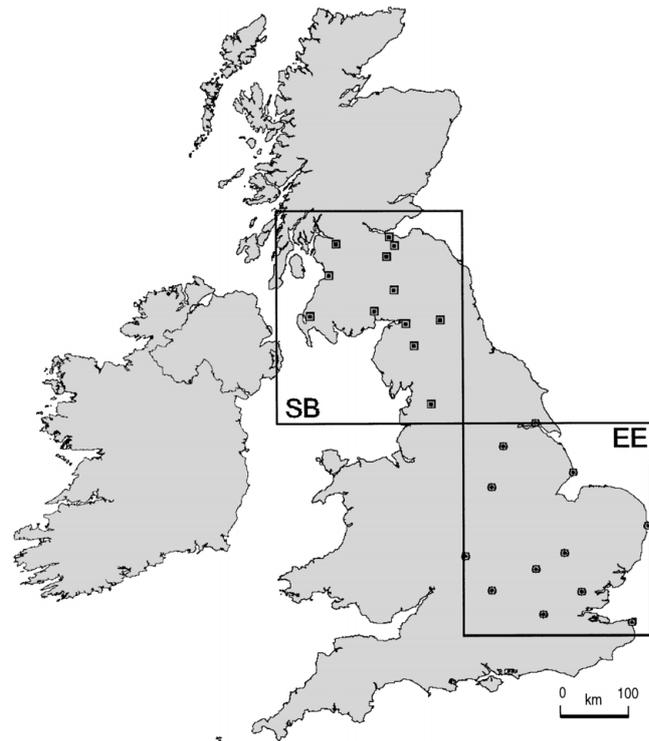


Fig. 1. Location of stations with daily precipitation observations in the Eastern England (EE) and the Scottish Borders (SB) for 1961–1990. The regional boundaries correspond to the SDSM predictor grid-boxes used for downscaling

by consensus, the most promising sub-set of predictors for area-averaged downscaling (below).

Secondly, the models were trained using the entire record of area-averaged daily precipitation amount in each region and predictors drawn from the multi-site analysis (above). For each region and month, large-scale atmospheric predictors were fit by stepwise multiple linear regression to the inverse-normal-transformed areal averages. Three sets of downscaling experiments were then performed: (1) *DET*, in which only the deterministic component of the area-averaged amounts is represented (i.e. the  $\epsilon$  term in Eq. 2 is omitted); (2) *VAR*, with variability in area averages due to both deterministic and stochastic components (i.e. the  $\epsilon$  term in Eq. 2 is included); and (3) *RND*, an unconditional area-average model involving stochastic resampling of observed amounts with replacement. In all 3 cases, the resulting area-averaged amounts were used as the marker series for multi-site resampling.

### 3. MEASURES OF MODEL PERFORMANCE

Multi-site model performance was assessed using 3 sets of measures applied to observed data and model syntheses for the period 1961–1990. (For a discussion of the model's single site, and split-record skill, refer to Wilby & Dettinger [2000].) Use of the full 30 years' data herein is justified on a number of counts. Firstly, the models were fit to the statistics of area-averaged daily precipitation occurrence and amounts, but were assessed in terms of spatial and temporal dependency at individual sites. Secondly, Kendall's  $\tau_b$  statistic returns unreliable estimates from small numbers of station pairs and years, so the available record length had to be maximised. Thirdly, this is the conventional approach to models with stochastic processes and/or re-sampling procedures (see, for example, Wilks 1998, Charles et al. 1999a, Buishand & Brandsma 2001).

**3.1. Temporal dependence.** As witnessed by the events of the year 2000, extreme river discharges in the UK can arise from periods of prolonged heavy rainfall in autumn and winter (DEFRA 2001, Marsh 2001). Following Wójcik et al. (2000), sample standard deviations and lag-1 autocorrelation coefficients of daily precipitation values were calculated for each station for the winter half year to compare observed data with model syntheses. In addition,  $N$ -day winter maximum precipitation amounts were investigated for 5, 10, 20, and 60 d totals using observed and model data representing the area averages of the 2 regions, as well as the distributions of maxima at selected sites (Cambridge [EE], Carlisle [SB], Edinburgh [SB] and Oxford [EE]).

**3.2. Correlation decay lengths.** Following the method of Osborn & Hulme (1997) the correlation,  $r$ ,

between the precipitation series at a station and every other station in the grid box was plotted against their separation distance,  $d$ . An exponential decay function was then fitted to the scatter of points, giving:

$$r = e^{-d/d_0} \quad (5)$$

where  $d_0$  is the correlation decay length at the distance where  $r$  falls to  $1/e$ . A major advantage of this method over the estimation of mean inter-station correlation between all pairs of stations ( $\bar{r}$ ) is that regions with different station distributions can be compared. The exponential decay has also been shown to be a reasonable function for describing variations in daily precipitation correlations with separation distance for European sites (Osborn & Hulme 1997).

**3.3. Spatial association of  $N$ -day precipitation maxima.** The final performance measure evaluates the spatial dependence of the  $N$ -day winter maxima at multiple sites by counting the number of joint exceedances of specified thresholds. Buishand & Brandsma (2001) used Kendall's  $\tau_b$  statistic to assess the degree of association from counts of exceedances of extreme rainfall at sites across the Rhine basin. They showed that if  $t_1$  and  $t_2$  are threshold precipitation totals at Sites 1 and 2, and  $\pi_1 = F_1(t_1)$ ,  $\pi_2 = F_2(t_2)$ , and  $\pi_{12} = F(t_1, t_2)$ , then  $\tau_b$  can be defined as:

$$\tau_b = \frac{\pi_{12} - \pi_1\pi_2}{\sqrt{\pi_1(1-\pi_1)\pi_2(1-\pi_2)}} \quad (6)$$

If  $t_1$  and  $t_2$  are the 100  $p$  percentiles of the marginal distributions (i.e.  $F_1[t_1] = p$  and  $F_2[t_2] = p$ ), then  $\tau_b$  becomes:

$$\tau_b(p) = \frac{\pi_{12} - p^2}{p(1-p)} \quad (7)$$

For the median  $p = 0.5$ . As in the case of the correlation coefficient, complete positive dependence is given by,  $\tau_b(p) = 1$ ; conversely, for independent data  $\tau_b(p) = 0$ . For observed, annual winter maxima at 2 stations,  $p$  can be estimated as:

$$\hat{p} = \frac{(K_1 + K_2)}{2K} \quad (8)$$

where  $K$  is the number of years,  $K_1$  is the number of winter maxima at Stn 1 that are less than or equal to  $t_1$ , and  $K_2$  is the corresponding number for Stn 2. the probability  $\pi_{12}$  is estimated as:

$$\hat{\pi} = \frac{K_{\text{joint}}}{K} \quad (9)$$

where  $K_{\text{joint}}$  is the number of years for which the winter maxima are less than or equal to the threshold at both stations.

For the purpose of the present study,  $\tau_b$  was estimated for the median (MED) and 90th percentile (Q90) of the 1 and 20 d winter-maximum precipitation amounts using all combinations of station pairs for the

period 1961–1990. Resulting estimates of  $\tau_b$  for observed and synthetic series were then plotted against separation distance and smoothed using the exponential decay function (see above), to estimate an equivalent correlation decay length for the joint exceedances. Smoothing is necessary because estimates of  $\tau_b$  for individual station pairs have large standard errors (Buishand & Brandsma 2001).

#### 4. RESULTS

##### 4.1. Selection of predictors

Table 1 shows the frequency of predictor variable selection (number of stations), comparing the outcome for EE with that for SB. For each station, the optimal combination of predictors was identified by stepwise multiple linear regression, with the stipulation that all regression coefficients of included predictors be significant at  $p \leq 0.001$ . In both regions, the specific humidity at 500 hPa (Q500) was selected at every station. Near-surface relative humidity (RSUR) and near-surface southerly wind (VWND) were selected at the majority of sites in SB, but rarely at sites in EE. The 500 hPa geopotential height (H500) was the second most frequently selected predictor in EE and, along with the near-surface vorticity (VORT), was chosen at over half the sites in SB. Mean sea-level pressure (MSLP) was the fifth most frequently selected variable. The wind strength at 500 hPa (F500) was of secondary importance at SB stations, but of no consequence to precipitation at stations across EE. Apart from the near-surface specific humidity (QSUR) in EE, the remaining selected predictors represent regional winds at various heights in the atmosphere, mainly of local importance to stations across SB. The larger number of predictors selected for SB stations probably reflects the greater heterogeneity of the topography in this region. Finally, it is worth noting that 15 out of the possible 32 predictors were never selected, and that only 12 predictors were selected on more than 1 occasion.

The above predictors convey information about different aspects of the atmosphere (i.e. thickness, circulation and moisture content), and they have been variously related to precipitation in earlier regional studies (e.g. Kilsby et al. 1998, Charles et al. 1999a, Murphy 1999). Nonetheless, the percentage of explained variance for daily precipitation occurrence at individual stations was relatively low, averaging 25% across EE and 30% across SB. Average levels of explained variance were even lower for precipitation amounts, just 9 and 17%, respectively. This reflects, in part, the

Table 1. Predictor variable description and frequency of selection at individual stations. EE: Eastern England; SB: Scottish Borders

Predictor	Description	EE (n = 12)	SB (n = 12)
Q500 <sup>a,b</sup>	Specific humidity at 500 hPa	12	12
H500 <sup>a,b</sup>	500 hPa geopotential height	10	7
VWND <sup>b</sup>	Near-surface southerly wind	3	9
VORT <sup>b</sup>	Near-surface vorticity	4	7
MSLP	Mean sea-level pressure	4	5
RSUR <sup>b</sup>	Near-surface relative humidity	0	9
QSUR <sup>a</sup>	Near-surface specific humidity	4	1
F500 <sup>b</sup>	Wind strength at 500 hPa	0	5
UWND <sup>a</sup>	Near-surface westerly wind	1	3
U500	Westerly wind at 500 hPa	0	3
U850	Westerly wind at 850 hPa	0	2
V850 <sup>a</sup>	Southerly wind at 850 hPa	2	0
V500	Southerly wind at 500 hPa	0	1
H850	850 hPa geopotential height	0	1
Z850	Vorticity at 850 hPa	1	0
F850	Wind strength at 850 hPa	0	1
D850	Divergence at 850 hPa	1	0

<sup>a</sup>Included in EE area model  
<sup>b</sup>Included in SB area model

decision to evaluate predictors without seasonal stratification. Had predictor-predictand relationships been examined at the seasonal or monthly level, different combinations might have emerged, with higher levels of explained variance anticipated for winter, and lower for summer (Wilby & Wigley 2000).

For example, Fig. 2 shows variations in the correlation between wet-day amounts at Eskdalemuir (the westernmost station in SB) and 2 of the most influential predictors at this site (MSLP and QSUR). Although both predictors could have been selected at this site on the basis of the average correlation throughout the year, it is evident that amounts are most strongly asso-

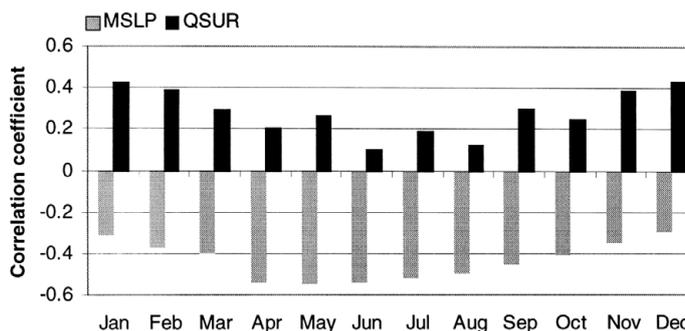


Fig. 2. Monthly variations in the strength of the correlation between daily wet-day amounts at Eskdalemuir (55° 19' N, 3° 12' W) and mean sea-level pressure (MSLP) and near surface specific humidity (QSUR) over the Scottish Borders region, 1961–1990

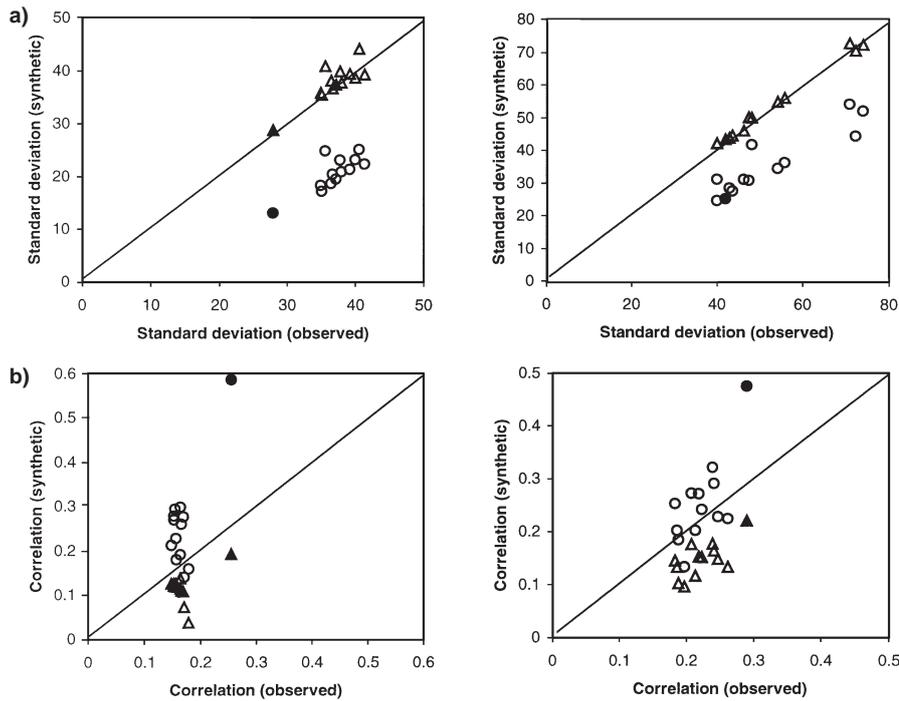


Fig. 3. Comparison of observed versus synthetic (a) standard deviations and (b) lag-1 autocorrelations of daily precipitation amounts at all stations in EE (left) and SB (right), 1961–1990. Open triangles denote the VAR model, and open circles the DET model; solid symbols are the respective results for area averages

ciated with QSUR in winter, and with MSLP in summer. Regional variations in explained variance are also consistent with previous studies showing generally greater skill over wetter western and northern areas than over drier regions to the south and east (e.g. Goodess & Palutikof 1998, Murphy 1999). This probably reflects the greater proximity of moisture sources in the west and the greater significance of sub-grid scale, convective precipitation events in the south and east.

There are a number of practical and theoretical advantages of using area-average series as opposed to individual station series for downscaling (Gregory et al. 1993). Although area averaging reduces the probability of zero rainfall and increases the probability of small non-zero amounts, it also reduces the variability due to very local events and thereby increases the signal attributable to large-scale forcing. This is reflected in the improved levels of explained variance for area-average models when stepwise multiple linear regression was used to identify optimal sets of predictors. In Region EE, the 5 predictor variables Q500, H500, QSUR, UWND and V850 explain 33% of the variance in area-averaged daily precipitation amounts. In Region SB, the 6 predictors Q500, H500, VWND, VORT, RSUR and F500 explain 42% of the variance. These 2 sets of predictors were used to train the monthly models in SDSM and then to generate the

DET and VAR marker series used in the following syntheses of multi-site precipitation.

#### 4.2. Temporal dependence

Fig. 3a compares the standard deviations of observed and synthetic daily precipitation amounts for all stations in EE (left panel) and SB (right panel). The average percentage differences between observed and synthetic series (VAR) were 3 and 2% in EE and SB, respectively. The largest outliers in EE were the stations at Lowestoft ( $52^{\circ} 29' N$ ,  $1^{\circ} 45' E$ ) and Manston ( $51^{\circ} 21' N$ ,  $1^{\circ} 21' E$ )—2 coastal locations situated at the very easterly margin of the grid box (see Fig. 1). In comparison, DET consistently under-represented standard deviations at all stations in both regions, by an average of 16% in EE and by 31% in SB. The greater skill of VAR was also reflected in the area-

averaged results with differences of less than 4% in both regions. As would be expected, the unconditional resampling by RND produced near-perfect syntheses of the standard deviations at all stations (not shown).

Fig. 3b compares the lag-1 autocorrelations of observed, VAR and DET precipitation amounts for all stations in EE (left) and SB (right); RND produced negligible temporal dependency so is not shown. In both regions, VAR slightly under-represented the autocorrelation and, as before, the 2 outliers for EE were Lowestoft and Manston. Conversely, DET generally over-estimated temporal dependency, and for EE exhibited far greater inter-station variability of autocorrelations than in observations. This was because VAR consistently over-estimated the frequency of dry-days (by as much as 6% at some sites), thereby increasing persistence relative to observations. The area-average autocorrelation for VAR reflected the underestimation at the station level, but for DET the difference from observations was even higher. Again, this was attributed to over-representation of dry days by DET: for area averages this was +9%.

Discrepancies in the standard deviation and autocorrelation of daily precipitation amounts are manifested in  $N$ -day winter maximum precipitation amounts. Figs. 4 & 5 compare distributions of observed and synthetic 5, 10, 20 and 60 d area-averaged amounts for EE

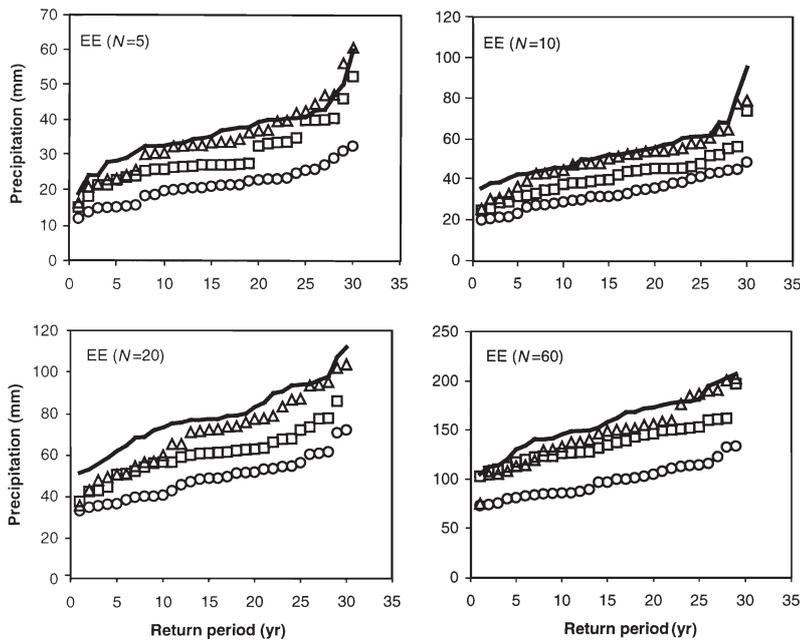


Fig. 4. Distributions of 5, 10, 20, and 60 d annual maximum precipitation for the winter season (September–March) in EE, 1961–1990. Solid lines represent observations; open symbols are model syntheses ( $\Delta$ : VAR;  $\square$ : RND;  $\circ$ : DET)

and SB, respectively. VAR consistently produced the best synthesis of maxima for both regions, for all return periods and values of  $N$ . However, VAR generally over-estimated the magnitude of longer-duration events in SB and under-estimated shorter-duration amounts in EE. DET, and to a lesser extent RND, under-estimated magnitudes of  $N$ -day maxima throughout. Presumably, DET under-estimates maxima because the standard deviation of daily amounts is too low, leading to sequences of moderate-sized daily amounts, whereas RND under-estimates because the model has a temporal persistence too low to synthesise prolonged spells of heavy rainfall. The fact that RND produced more realistic distributions of EE maxima than DET implies that, for this measure, replicating the overall distribution of daily amounts was more important than replicating the serial correlation of amounts.

The above findings were also valid at the level of individual stations. Once again, VAR showed greater skill than both RND and DET for winter maxima at Cambridge and Oxford (Fig. 6). How-

ever, the superiority of VAR over RND was less pronounced, especially for the 60 d amounts. As for the SB area average, VAR over-estimated the 20 and 60 d amounts at both Carlisle and Edinburgh (Fig. 7). Interestingly, none of the models replicated the marked step-change in amounts at Carlisle associated with 20 and 60 d events of 25+ years return period, although a similar feature in the Oxford 60 d distribution was synthesised by VAR (Fig. 6).

### 4.3. Correlation decay lengths

Comparisons of observed and synthetic inter-site correlations were undertaken for all possible pairs of daily precipitation series. Fig. 8a shows that DET systematically under-estimates the pairwise correlations of station daily precipitation series across the EE region, whereas the VAR and RND procedures are clearly much more successful in simulating this aspect of spatial dependency. The same pattern of behaviour was evident for the SB region (Fig. 8b). As would be expected, RND produces near-perfect spatial correlations,

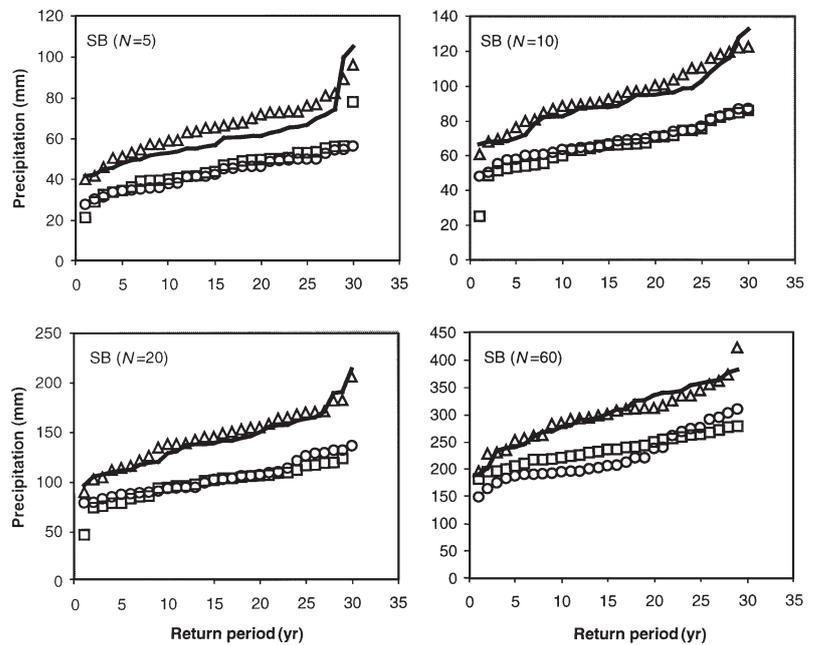


Fig. 5. Distributions of 5, 10, 20, and 60 d annual maximum precipitation for the winter season (September–March) in SB, 1961–1990. Solid lines represent observations; open symbols are model syntheses ( $\Delta$ : VAR;  $\square$ : RND;  $\circ$ : DET)

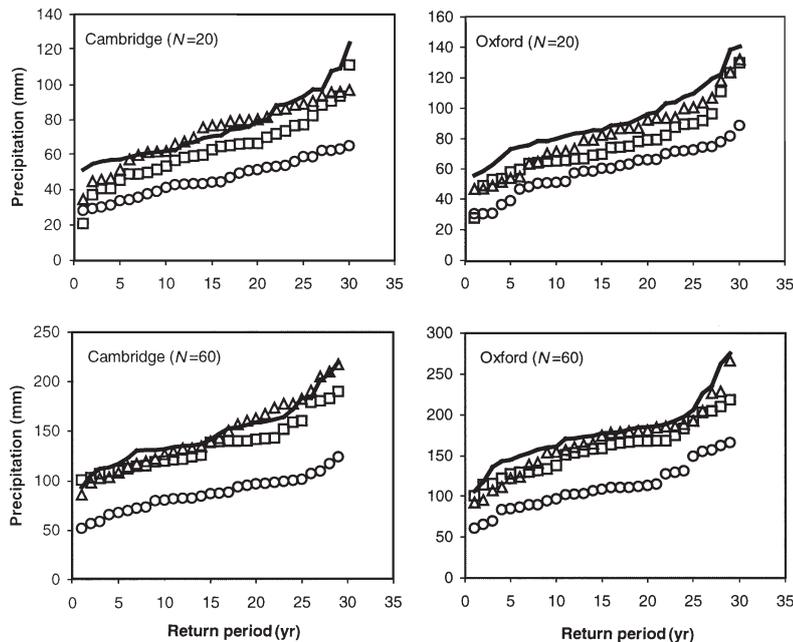


Fig. 6. Distributions of 20 and 60 d annual maximum precipitation for the winter season (September–March) at Cambridge and Oxford, 1961–1990. Solid lines represent observations; other symbols are model syntheses ( $\Delta$ : VAR;  $\square$ : RND;  $\circ$ : DET)

whilst VAR a slight positive bias, and DET significant under-estimation of inter-station correlations. The depressed pairwise correlations of DET was attributed to the relatively narrow sampling range of daily precipitation by this model, as indicated by the under-estimated standard deviation of daily amounts (noted above).

An alternative means of visualising the results in Fig. 8 is to plot each set of pairwise correlations of station daily precipitation series versus station separation distance. For example, Fig. 9 shows the expected decline from the high correlations of nearest neighbours to the low correlations of distant sites across the 2 regions. The structure of the observed decay differs between the regions: EE displays a progressive decline with distance compared with the more heterogeneous pattern for SB, again no doubt reflecting the more complex orography of the latter region. However, the under-estimation by DET of correlation strengths at all separations is clearly evident in both cases.

Table 2 compares the decay lengths (km) estimated from the best-fit exponential function of pairs of correlations of station daily precipitation series versus separation distance (Eq. 5) for observed and synthetic (DET, VAR, and RND) data. Osborn & Hulme (1997) report that the characteristic scale of observed precipitation-causing disturbances is larger in winter than in summer, due to a higher proportion of precipitation associated with fronts in winter. (Events associated with winter weather fronts are generally more widespread than summer convective storms, leading to greater decay lengths in winter.) Although VAR and RND (but not DET) simulate well the correlation decay length for the year as a whole (Table 2), none of the models capture observed seasonal variations in the decay length. For example, observed decay lengths vary between 189 km (145 km) in summer and 270 km (227 km) in winter for EE (SB). However, the models yield approximately the same decay lengths all year round because the normal cumulative distribution ( $\Phi$ ) used for resampling daily amounts was not stratified by season

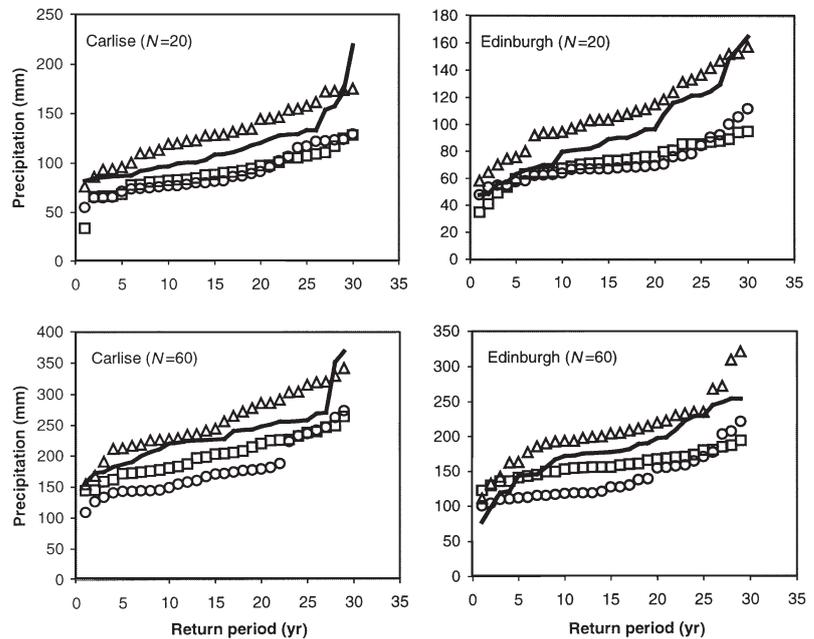


Fig. 7. Distributions of 20 and 60 d annual maximum precipitation for the winter season (September–March) at Carlisle and Edinburgh, 1961–1990. Solid lines represent observations; other symbols are model syntheses ( $\Delta$ : VAR;  $\square$ : RND;  $\circ$ : DET)

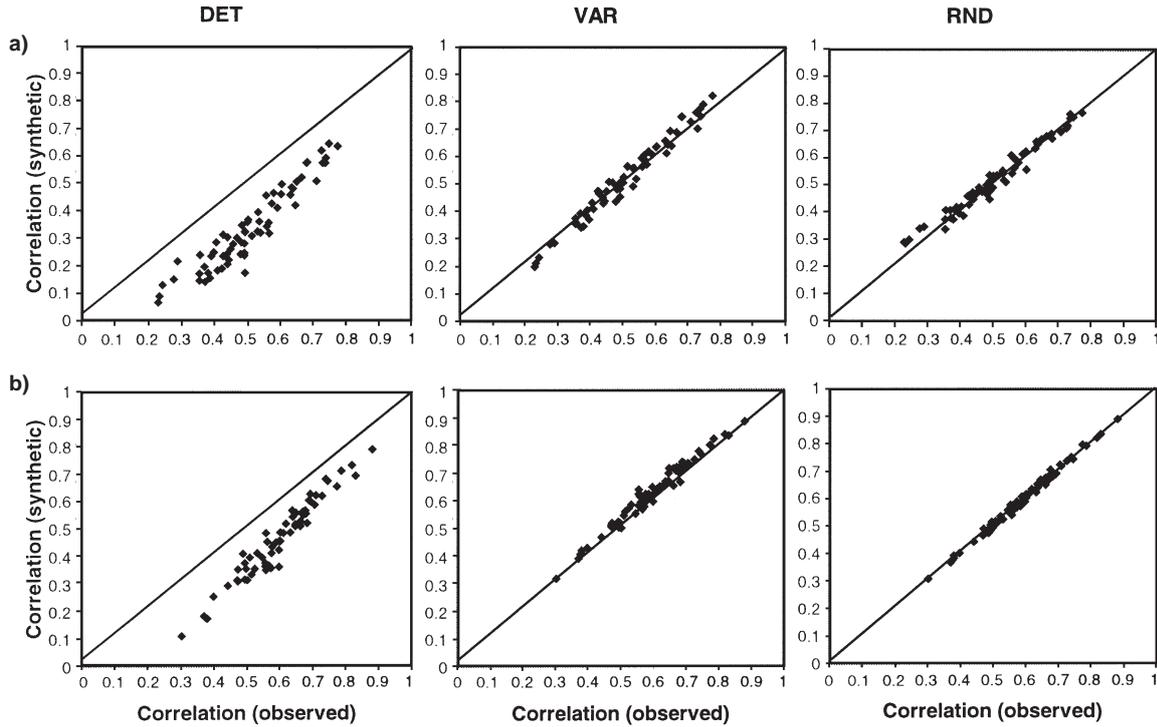


Fig. 8. Comparisons of pairwise correlations of station daily precipitation amounts in the observations (horizontal) and synthetic series (vertical) generated by *DET*, *VAR* and *RND* for (a) EE and (b) SB

Table 2. Correlation decay lengths (km) obtained from all possible pairs of station daily precipitation time series using the entire year. The best model syntheses of observations are given in **bold**

EE				SB			
Obs.	<i>DET</i>	<i>VAR</i>	<i>RND</i>	Obs.	<i>DET</i>	<i>VAR</i>	<i>RND</i>
213	126	<b>215</b>	226	211	135	232	<b>213</b>

by means of Kendall's  $\tau_b$  statistic. For example, Fig. 10 shows estimates of  $\tau_b$  for Q90 of the 1 d winter maxima across EE, together with the exponential function for observed EE and synthetic data. As expected, joint exceedences of observed series are relatively high at short distances, and they decrease for greater separations. This trend is reproduced by the models; however, all 3 under-estimate  $\tau_b$  for nearest neighbours and over-estimate joint exceedences at greater separations.

or month, unlike the regression coefficients in Eqs. (1) & (2). This implies that conditioning  $Z_i$  in Eq. (2) by atmospheric predictors alone (albeit weighted by different coefficients in each month) does not implicitly capture seasonal variations in correlation decay lengths.

#### 4.4. Association of N-day winter maxima

The spatial dependence of winter maxima precipitation totals was assessed for all station pairs

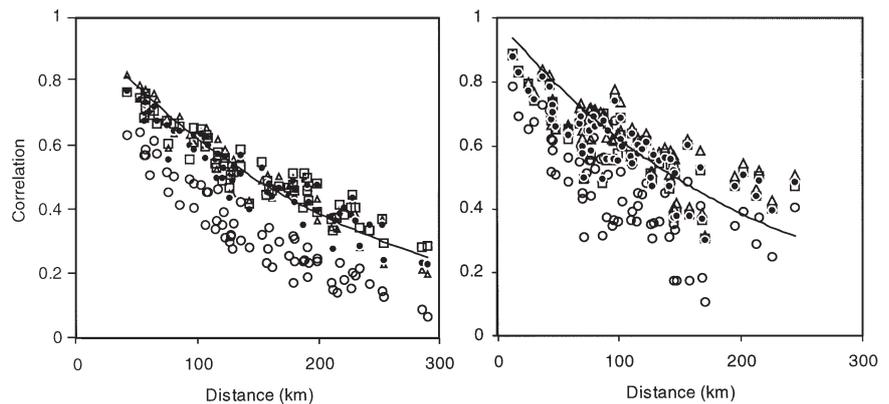


Fig. 9. Correlation between pairs of station daily precipitation amounts vs their separation distance for all possible combinations of station pairs in EE (left) and SB (right), 1961–1990. Solid lines represent exponential decay functions fitted to the observations (solid circles); open symbols are model syntheses ( $\Delta$ : *VAR*;  $\square$ : *RND*;  $\circ$ : *DET*)

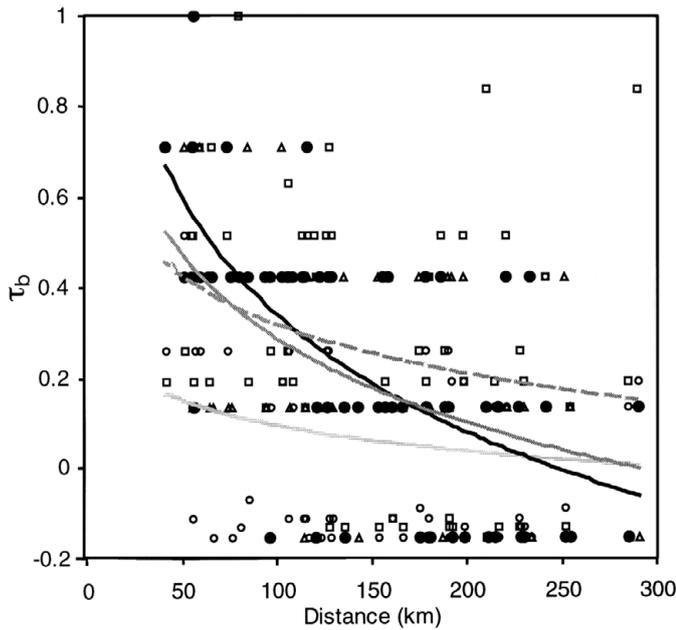


Fig. 10. Estimates of Kendall's  $\tau_b$  for the 90th percentile (Q90) 1 d winter maxima precipitation amounts across EE, 1961–1990. The solid black line represents the exponential decay function fitted to the observations (●); grey dotted line and ○: DET; solid grey line and △: VAR; grey dashed line and □: RND

Over-estimation of long-distance associations was particularly evident in RND; conversely, DET performed worst for nearest neighbours.

As in Section 4.3, decay lengths were estimated from the best-fit exponential function of  $\tau_b$  for all station pairs. Table 3 reports the decay lengths arising from the Q90 and MED of the 1 and 20 d winter maxima. Comparing the results for 1 and 20 d maxima, it is evident that the average strength of association is typically greater for longer-duration events. These results are consistent with the findings of Buishand & Brandsma (2001) for 1 and 10 d winter maximum precipitation in the Rhine basin. Overall, VAR yielded the best estimates of the 20 d maxima in EE, whereas DET

Table 3. Decay lengths (km) obtained from Kendall's  $\tau_b$  using all possible pairs of stations. The best model syntheses of observations are given in **bold**. na: no relationship with distance

	EE				SB			
	Obs.	DET	VAR	RND	Obs.	DET	VAR	RND
Q90 <sub>20</sub>	113	162	<b>110</b>	154	37	<b>5</b>	>300	>300
Q90 <sub>1</sub>	94	4	<b>75</b>	<b>75</b>	24	32	<b>18</b>	32
MED <sub>20</sub>	205	>300	<b>270</b>	107	97	<b>112</b>	>300	47
MED <sub>1</sub>	79	na	28	<b>57</b>	5	<b>4</b>	33	33

Table 4. Average Kendall's  $\tau_b$  using all possible pairs of stations. The best model syntheses of observations are given in **bold**

	EE				SB			
	Obs.	DET	VAR	RND	Obs.	DET	VAR	RND
Q90 <sub>20</sub>	0.30	0.35	<b>0.32</b>	0.40	0.19	0.28	0.44	<b>0.20</b>
Q90 <sub>1</sub>	0.23	0.07	<b>0.21</b>	0.27	0.12	0.18	<b>0.11</b>	0.19
MED <sub>20</sub>	0.45	0.53	<b>0.46</b>	0.29	0.37	<b>0.38</b>	0.53	0.23
MED <sub>1</sub>	0.28	0.15	<b>0.16</b>	0.15	0.18	<b>0.18</b>	<b>0.18</b>	0.23

was most proficient for SB. The pattern for 1 d maxima was less clear-cut: RND performed best in EE but not in SB.

Finally, it is evident from Table 3 that the spatial coherence (i.e. decay lengths) of joint exceedances was much greater across EE than SB. This is confirmed by the regional average Kendall's  $\tau_b$  shown in Table 4. With the exception of the MED of 1 d maxima, VAR produced a close approximation of the average level of association across EE. Again, the results for SB are less consistent, with no model emerging as overall leader. However, all models correctly signified that the joint exceedances were strongest for the MED of 20 d maxima, and least for the Q90 of the 1 d maxima. In other words, the 20 d event that occurs on average once in 2 yr (MED) has greater spatial coherence than the single-day event that occurs on average once in 10 yr (Q90).

### 5. DISCUSSION AND CONCLUSIONS

Multi-site downscaling of daily precipitation using conditional resampling methods was explored for Eastern England (EE) and the Scottish Borders (SB). Following an investigation of the relative importance of a host of atmospheric predictors to station precipitation series, the Statistical DownScaling Model (Wilby et al. 2002b) was used to synthesise area-averaged daily precipitation amounts in each region. Three methods were compared: (1) a fully deterministic procedure involving atmospheric conditioning of precipitation amounts (DET); (2) a semi-stochastic procedure involving atmospheric forcing and stochastic representation of model errors (VAR); and (3) an unconditional procedure involving random sampling of daily precipitation amounts with replacement (RND). Multi-site series were then constructed for each method and region by sampling the days at constituent stations corresponding to the days represented by the area-average 'marker' series. Finally, the realism of multi-site precipitation amounts was assessed using a range of statistical measures of temporal and spatial dependency.

The analysis of large-scale predictor-predictand relationships supported the view that atmospheric humidity, in addition to circulation patterns, is needed for effective downscaling of daily precipitation amounts (Charles et al. 1999b, Wilks & Wilby 1999, Beckmann & Buishand 2001). Comparisons of the dominant predictor variables of EE with those of SB further underlined the level of sub-regional variability of predictors. This reflects the interaction of regional airflows with local variations in station elevation, aspect, proximity to coast, etc. (see Kilsby et al. 1998). Correlations between grid-box circulation indices and precipitation are typically stronger at locations near oceanic sources in winter, where the advection of atmospheric moisture is more direct (Robertson & Ghil 1999, Wilby & Wigley 2000). This was reflected by the higher explained variance for area-averaged precipitation in SB (42%) than in EE (33%). Furthermore, westerly winds figured more prominently as predictors of precipitation amounts at stations in the 'maritime' SB grid box than in the 'near-continental' EE grid box.

The preceding analyses indicated that some, but by no means all, of the variance in daily area-averaged precipitation amounts was explained by large-scale atmospheric predictor variables. In fact, the fully deterministic *DET* procedure for generating area-averaged daily precipitation amounts was the least successful of the 3 methods for multi-site resampling. This was because the variance of the daily area averages was too low (resulting in under-estimated standard deviation of daily amounts and winter *N*-day maxima at individual stations), whereas the lag-1 autocorrelation coefficient was too high (due to over-representation of dry-days). The *DET* procedure also significantly under-estimated spatial dependency amongst stations for both daily series and joint exceedance of *N*-day amounts. Without artificial variance inflation *DET* resamples from too narrow a range of observed daily amounts to produce credible syntheses of observed data.

In contrast, the *RND* procedure for synthesising area-average marker series resamples from the full range of daily amounts at the station level, and hence it yields excellent results for the standard deviation. Unsurprisingly, the lag-1 autocorrelation coefficient of observed series is not replicated by this method. The absence of conditioning by atmospheric predictors also meant that the distribution of *N*-day winter maxima was generally too low (i.e. without atmospheric modes leading to clusters of heavy rain-days, winter maxima are less likely to be represented stochastically). However, as expected, *RND* did yield near-perfect results for pairwise correlations of station daily precipitation amounts and decay lengths for 1 d joint exceedances of MED and Q90 amounts. Decay dis-

tances for the 20 d exceedances of MED were significantly under-estimated because of the non-existent autocorrelation. The spatial associations for the 20 d Q90 amounts were more ambiguous, possibly reflecting larger standard errors in the estimate of  $\tau_b$  for rarer events.

The *VAR* procedure combines the features of both the *DET* and *RND* methods and, as such, represented the most promising basis for resampling. Standard deviations of daily amounts, area-averaged *N*-day winter precipitation maxima, pairwise correlations of station daily amounts, and decay distances (for daily amounts and joint exceedances) were all successfully synthesised. However, *VAR* was less successful at simulating distributions of station *N*-day winter precipitation maxima and 20 d joint exceedances, particularly across SB. This is because the inclusion of the stochastic component slightly degrades the autocorrelation of synthetic daily series, a deficiency that might be addressed by further conditioning of the error term  $\epsilon$  in Eq. (2). Alternatively, the relative significance of the  $\epsilon$  term could be reduced through improved conditioning of the daily amounts by a more judicious choice of lagged predictors and/or grid-point location. For example, a number of studies have indicated that the optimum grid point for downscaling does not necessarily coincide with the target region, but may be located to the west or northwest of the station(s) (Wilby & Wigley 2000, Brinkmann 2002).

An acknowledged weakness of the downscaling procedures was their collective failure to reproduce seasonal variations in the pairwise daily correlation and decay lengths. This was traced to the unstratified resampling of daily amounts by  $Z_i$  (Eq. 3). A diagnostic statistic such as the Kolmogorov-Smirnov could, in future, be applied to test for significant differences between the sub-samples of seasonally stratified and unstratified amounts. Although this presents an opportunity for further model refinement, resampling stratified by month or season might not be helpful in all situations. For example, in semi-arid climates there may be insufficient rain days to properly characterise amounts at the monthly scale. Alternatively, Charles et al. (1999a) demonstrated that the worst results from a spatio-temporal model were linked to particular weather states with limited power for resolving precipitation occurrence at multiple sites. In other words, it may be informative to further disaggregate downscaling model output to examine synoptic conditions yielding the largest anomalies in multi-site precipitation amounts. Ongoing research will focus explicitly on *N*-day precipitation maxima to identify the most useful predictors for downscaling such events, with a view to constructing climate-change scenarios of future flood risk in the UK.

*Acknowledgements.* The authors wish to thank the 3 anonymous reviewers for their constructive comments. The insightful comments made by Clare Goodess are also gratefully acknowledged. This research was supported in part by STARDEX (STAtistical and Regional Dynamical downscaling of EXtremes for European regions) under the European Community Research Programme (contract: EVK2-CT-20010015).

#### LITERATURE CITED

- Bárdossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour Res* 28:1247–1259
- Beckmann BR, Buishand TA (2001) KNMI contribution to the European project WRINCLE: downscaling relationships for precipitation for several European sites. Tech Rep TR-230, KNMI, De Bilt
- Brinkmann WAR (2002) Local versus remote grid points in climate downscaling. *Clim Res* 21:27–42
- Buishand TA, Brandsma T (2001) Multi-site simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbour resampling. *Water Resour Res* 37: 2761–2776
- Charles SP, Bates BC, Hughes JP (1999a) A spatio-temporal model for downscaling precipitation occurrence and amounts. *J Geophys Res* 104:31657–31669
- Charles SP, Bates BC, Whetton PH, Hughes JP (1999b) Validation of downscaling models for changed climate conditions: case study of southwestern Australia. *Clim Res* 12: 1–14
- Cowpertwait PSP, O'Connell PE (1997) A regionalised Neyman-Scott model of rainfall with convective and stratiform cells. *Hydrol Earth Syst Sci* 1:71–80
- DEFRA (2001) To what degree can the October/November 2000 flood events be attributed to climate change? DEFRA FD2304, Final Report, March 2001 (revised June 2001), Centre for Ecology and Hydrology, Wallingford, & The Meteorological Office, Bracknell
- Goodess CM, Palutikof JP (1998) Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *Int J Climatol* 18:1051–1083
- Gregory JM, Wigley TML, Jones PD (1993) Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. *Clim Dyn* 8: 299–310
- Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov model for precipitation occurrence. *Appl Stat* 48:15–30
- Hutchinson MF (1995) Stochastic space-time weather models from ground-based data. *Agric For Meteorol* 73:237–264
- Kalnay E, Kanamitsu M, Kistler R, Collins W and 18 others (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471
- Kilsby CG, Cowpertwait PSP, O'Connell PE, Jones PD (1998) Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *Int J Climatol* 18:523–539
- Marsh TJ (2001) The 2000/01 floods in the UK—a brief overview. *Weather* 56:343–345
- Murphy J (1999) An evaluation of statistical and dynamical techniques for downscaling local climate. *J Clim* 12: 2256–2284
- Osborn TJ, Hulme M (1997) Development of a relationship between station and grid-box rainfall frequencies for climate model evaluation. *J Clim* 10:1885–1908
- Palutikof JP, Goodess CM, Watkins SJ, Holt T (2002) Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean. *J Clim* 15:3529–3548
- Rajagopalan B, Lall U (1999) A k-nearest neighbor simulator for daily precipitation and other weather variables. *Water Resour Res* 35:3089–3101
- Reid PA, Jones PD, Brown O, Goodess CM, Davies TD (2001) Assessments of the reliability of NCEP circulation data and relationships with surface climate by direct comparisons with station based data. *Clim Res* 17:247–261
- Robertson AW, Ghil M (1999) Large-scale weather regimes and local climate over the western United States. *J Clim* 12:1796–1813
- Wilby RL, Dawson CW (2001) Using SDSM—a decision support tool for the assessment of regional climate change impacts. User Manual prepared on behalf of the Environment Agency Risk Analysis Section, London
- Wilby RL, Dettinger MD (2000) Streamflow changes in the Sierra Nevada, CA simulated using a statistically downscaled General Circulation Model scenario of climate change. In: McLaren SJ, Kniveton DR (eds) Linking climate change to land surface change. Kluwer Academic, Publishers, Dordrecht, p 99–121
- Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: observed and General Circulation Model relationships. *Int J Climatol* 20:641–661
- Wilby RL, Hay LE, Leavesley GH (1999) A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. *J Hydrol* 225:67–91
- Wilby RL, Conway D, Jones PD (2002a) Prospects for downscaling seasonal precipitation variability using conditioned weather generator parameters. *Hydrol Proc* 16: 1215–1234
- Wilby RL, Dawson CW, Barrow EM (2002b) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ Model Software* 17:145–157
- Wilks DS (1998) Multisite generalization of a daily stochastic precipitation generation model. *J Hydrol* 210:178–191
- Wilks DS, Wilby RL (1999) The weather generation game: a review of stochastic weather models. *Prog Phys Geogr* 23: 329–357
- Wójcik R, Beersma JJ, Buishand TA (2000) Rainfall generator for the Rhine Basin: multi-site generation of weather variables for the entire drainage area. 186-IV, KNMI, De Bilt
- Zorita E, Hughes JP, Lettenmaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J Clim* 8:1023–1042

*Editorial responsibility: Clare Goodess, Norwich, United Kingdom*

*Submitted: December 10, 2001; Accepted: August 16, 2002  
Proofs received from author(s): February 4, 2003*