Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 312 (2005) 235–255
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Multi-site downscaling of heavy daily precipitation occurrence and amounts

Colin Harpham, Robert L. Wilby*

Environment Agency, Trentside Offices, Scarrington Road, Nottingham NG2 5FA, UK

Received 1 December 2003; revised 10 January 2005; accepted 11 February 2005

Abstract

This paper compares three statistical models for downscaling heavy daily precipitation occurrence and amounts at multiple sites given lagged and contemporaneous large-scale climate predictors (such as atmospheric circulation, thickness, and moisture content at the surface, 850 and 500 hPa). Three models (a Radial Basis Function (RBF) Artificial Neural Network (ANN), Multi Layer Perceptron (MLP) ANN and a Conditional Resampling Method (SDSM)) were applied to area-average and station daily precipitation amounts in northwest (NWE) and southeast (SEE) England. Predictor selection via both stepwise multiple linear regression and compositing confirmed vorticity and humidity as important downscaling variables. Model skill was evaluated using indices of heavy precipitation for area averages, individual sites and inter-site behaviour.

When tested against independent data (1979–1993), multi-site ANN models correctly simulated precipitation occurrence 80% of the time. The ANNs tended to over-estimate inter-site correlations for amounts due to their fully deterministic forcing, but performance was marginally better than SDSM for most seasonal-series of heavy precipitation indices. Conversely, SDSM yielded better inter-site correlation and representation of daily precipitation quantiles than the ANNs. All models had greatest skill for indices reflecting persistence of large-scale winter precipitation (such as maximum 5-day totals) or dry-spell duration in summer. Overall, predictability of daily precipitation was greater in NWE than SEE.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Downscaling; Precipitation; Multi-site; Diagnostics

1. Introduction

General Circulation Models (GCMs) are instrumental to projections of global climate change. However, their coarse spatial resolution (typically ~300 km) limits their usefulness for regional impact studies. As a consequence techniques have been developed to 'downscale' coarse GCM output to the finer spatial scales required for impact assessment. Overviews of downscaling approaches have been provided elsewhere (see Giorgi and Mearns, 1991; von Storch et al., 1993; Hewitson and Crane, 1996; Wilby and Wigley, 1997). The methods dealt with in this paper are regression-based but, to date, relatively few have

* Corresponding author. Tel.: +44 115 8463727; fax: +44 115 8463710.
E-mail address: rob.wilby@environment-agency.gov.uk (R.L. Wilby).

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.02.020

been explicitly developed for heavy precipitation at multiple sites.

Regression-based downscaling methods use empirical relationships between local scale predictands and regional scale predictor(s). Individual downscaling schemes differ according to the choice of mathematical transfer function, predictor variables or statistical fitting procedure. To date, linear and non-linear regression, artificial neural networks (ANNs), and canonical correlation have all been used to derive predictor–predictand relationships (e.g. Conway et al., 1996; Crane and Hewitson, 1998; Schubert and Henderson-Sellers, 1997). The main strength of regression methods is the relative ease of application, coupled with their use of observable trans-scale relationships. Unfortunately, regression models seldom explain all of the observed climate variability (especially in precipitation series). Regression methods also assume stationarity of model parameters under future climate conditions, and scenarios are known to be highly sensitive to the choice of predictor variables and statistical transfer function (Winkler et al., 1997). Furthermore, downscaling future extreme events using regression methods is problematic since these phenomena, by definition, often lie at the margins or beyond the range of the calibration data set.

To date, there have been relatively few examples of ANNs applied to climate downscaling and even fewer to the task of multi-site precipitation modelling. Table 1 lists examples of previous studies showing mixed success by ANNs for downscaling daily precipitation amounts. Possible explanations include the simplistic treatment of days with zero amounts and/or the highly skewed distribution of wet-day amounts biasing the training process towards small, or even negative, values. In this paper, we address this problem via a two-stage modelling procedure using Radial Basis Function (RBF) ANNs (Broomhead and Lowe, 1988; Moody and Darken, 1989) and Multi Layer Perceptron (MLP) ANNs (Rumelhart and McClelland, 1986). Analogous to conventional weather generation techniques, precipitation is modelled using separate occurrence and amounts processes (Wilks and Wilby, 1999). We are particularly interested in the ability of the ANNs to downscale indices of heavy daily precipitation (defined herein as $\geq$90th percentile amount). By way of a reference point, results from the ANNs are compared with those produced by a conditional resampling method (Wilby et al., 2003).

Having introduced the principle features of the ANNs and resampling procedure, part one of this investigation describes methods of predictor variable selection. Three approaches were compared: stepwise multiple linear regression (SWLR), a compositing procedure, and a Genetic Algorithm (GA). In part two, the three downscaling methods are used to predict daily precipitation at multiple sites in northwest (NWE) and southeast (SEE) England. All models were evaluated against indices of heavy daily precipitation and multi-site behaviour using data that were not employed in model calibration.

2. Model descriptions and data

Unlike previous studies, we model each site simultaneously rather than training downscaling models on a site by site basis (Table 1). The main advantage of this approach is the capture of both the temporal and spatial dependency of multi-site precipitation by model weights. Previous studies have favoured MLP networks so, in order to provide a comparison, the less well-known RBF network has also been used. One advantage of the type of RBF network employed here is the speed of training (a few seconds) in comparison to the MLP.

2.1. Radial basis function (RBF) networks

The RBF network consists typically of two layers, where the hidden layer nodes contain prototype vectors (or basis centres), which are in effect hidden layer weights. The distance between the input and the prototype vector determines the activation level of the hidden layer with the non-linearity provided by a basis function. The activation function in the output layer can be non-linear, however, training is considerably faster if an ordinary linear weighted sum of these activations are performed, and this approach was consequently adopted. Mathematically the output from the final layer node(s) y_k (for the kth output node) is

Table 1
Examples of recent downscaling studies involving the use of ANNs

Region	Predictand	Predictors	ANN type	Authors
Ciapas, Mexico	DP average of 11 stations	Sea level pressure, 500 hPa field	MLP	Hewitson and Crane (1992)
Colorado Plateau, USA	DP at 54 sites split into 5 regions	700 hPa geopotential heights	MLP	McGinnis (1997)
N.E. Mexico	Local winter rainfall, 20 grid points	Sea level pressure, 500 hPa heights, 1000–500 hPa thickness	MLP individual models	Cavazos (1997)
N.E. Mexico and S.E. Texas	DP at 20 grid points	700–500 hPa thickness, 700 hPa moisture	SOM/ MLP individual models	Cavazos (1999)
Susquehanna basin	DP at 16 grid cells	1000 and 500 hPa heights, 1000, 700 and 500 hPa specific humidities	MLP individual models	Crane and Hewitson (1998)
Portugal	Daily temperatures	1000 and 500 hPa geopotential heights	MLP	Trigo and Palutikof (1999)
Iberian Peninsula	Winter DP and MP, 1 site	Daily sea level pressure	MLP	Zorita and von Storch (1999)
Texas and California	Surface wind speeds	Large scale GCM output	MLP	Sailor et al. (2000)
Bucharest, Romania	Winter DP at 12 grid points and Bucharest	500–1000 hPa thickness, 700 hPa geopotential heights, 700 hPa moisture	SOM/MLP	Cavazos (2000)
Mid USA	Surface air temperatures	Max daily temperature	MLP	Snell et al. (2000)
New Zealand	MP and MT anomalies at 5 sites	Zonal and meridional mslp gradients, 850–700 hPa thickness, relative vorticity and horizontal divergence at 500 hPa	MLP	Mpelasoka et al. (2001)
Kyushu Island S. Japan	Extreme precipitation in Chikugo river	Precipitable water and wind speeds at 850 hPa	MLP- classifier	Olsson et al. (2001)
Indianapolis, USA	DP & DT	700 and 500 hPa heights, 850 hPa temperature, sea level pressure and column average relative humidity	MLP	Schoof and Pryor (2001)
Iberia	DP	Sea level pressure	MLP	Trigo and Palutikof (2001)
British Columbia, Canada	Stream flow	500 hPa height, 850 hPa specific humidity, 1000–500 hPa thickness and sea level pressure	MLP	Cannon and Whitfield (2002)
West Antarctica	Surface pressure and temperature	Pressure: 850 and 700 hPa heights, 850–700 hPa thickness, 850 hPa wind speed and direction. Temperature: 850 hPa height, 850–700 hPa thickness, 850 hPa temp advection	MLP	Reusch and Alley (2002)
Reviews	Various			Crane et al. (2002) and Hsieh and Tang (1998)

Key: DP, daily precipitation; MP, monthly precipitation; DT, daily temperature; MT, monthly temperature.

expressed by the following:

$$y_k(\mathbf{x}) = \sum_{j=1}^{M} w_{kj}\phi_j(\mathbf{x}) + w_{k0} \quad (1)$$

Where $\mathbf{x}$ is the d-dimensional input vector with elements x_i, and w_{kj} are the output layer weights together with the bias w_{k0}, and M is the number of hidden layer nodes or basis centres. The basis function $\phi_j(\mathbf{x})$ provides the non-linearity. There are six basis functions that are widely recognised as having useful properties for RBF networks (Bishop, 1995). These are multiquadratic, gaussian, inverse multiquadratic, thin plate spline, cubic and linear. For example, the cubic basis function has the form $\phi_j(\mathbf{x}) = ||\mathbf{x} - \mathbf{u}_j||^3$, where $\mathbf{x}$ is the d-dimensional input vector with elements x_i and $\mathbf{u}_j$ is the vector determining the centre of the basis function ϕ_j and has elements u_{ji}. If the RBF network

has the bias parameters absorbed into the weights then this gives

$$y_k(\mathbf{x}) = \sum_{j=0}^{M} w_{kj}\phi_j(\mathbf{x}) \tag{2}$$

which in matrix form becomes

$$y(\mathbf{x}) = \mathbf{w}\phi \tag{3}$$

where $\mathbf{W} = (w_{kj})$ and $\boldsymbol{\phi} = (\phi_j)$. In matrix form (where T denotes the transpose of a matrix)

$$\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{W}^T = \mathbf{\Phi}^T\mathbf{T} \tag{4}$$

where $\mathbf{\Phi}$ is an $N\times M$ matrix with elements ϕ_j^n, $\mathbf{W}$ is an $C\times M$ matrix with elements W_{kj} and $\mathbf{T}$ is an $N\times C$ matrix with elements t_k^n. For $n=1,\ldots,N\, j=1,\ldots,M\, k=1,\ldots,C$, where N is the number of data points, M is the number of basis centres, and C is the number of output nodes. If $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ is non-singular then $\mathbf{\Phi}* = [\mathbf{\Phi}^T\mathbf{\Phi}]^{-1}\mathbf{\Phi}^T$, where $\mathbf{\Phi}*$ is an $M\times N$ matrix known as the pseudo inverse of $\mathbf{\Phi}$. The solution for the optimal weight vectors can then be expressed as

$$\mathbf{W}^T = \mathbf{\Phi}*\mathbf{T} \tag{5}$$

There are several methods of extracting the best solution for the weight vectors, however, Singular Value Decomposition (SVD) is used here due to its robustness (Press et al., 1993).

One of the advantages of the RBF network is the ability to choose suitable parameters for the hidden units without having to perform a full non-linear optimisation of the network. This is because of the distinction between the roles of the first and second layer weights. In the first layer the basis centres $\mathbf{u}_j$ can be determined by unsupervised training using the predictor input vector $\{\mathbf{x}^n\}$ without making use of the predictand. This can be accomplished by methods such as a clustering algorithm that is based on a distance measure between the input vectors. The method used here is K-means clustering which assigns data points to the cluster (of k clusters) having the nearest centroid. The distance measure between the data vector and cluster centre used is Euclidean and the algorithm type is the convergent clustering version (Anderberg, 1973).

The main parameter that needs to be set is the number of nodes in the hidden layer. Too many and the generalisation properties of the network will be poor; too few and the network will not 'learn' the problem. In order to explore this effect the network was run with 2–100 nodes, monitoring the error against the validation set. The six basis functions were also compared, and found to have a negligible effect on network performance.

2.2. Multi-layer perceptron (MLP) network

The MLP network is not necessarily restricted to one hidden layer, although it has been demonstrated that any continuous function can be mapped to an arbitrary degree of accuracy by a single hidden layer (Hornik et al., 1989). The configuration used in this investigation is a single hidden layer. Mathematically the output from the final layer node(s) y_k is

$$y_k = \phi^{(2)}\left(\sum_{j=0}^{M} w_{kj}^{(2)}\phi^{(1)}\left(\sum_{i=0}^{d} w_{ji}^{(1)}x_i\right)\right) \quad k = 1,\ldots,C \tag{6}$$

where x_i is the ith input of input vector $\mathbf{x}$, w_{ji} is the weighting factor on the link between the ith input and the jth output, w_{kj} is the weighting factor on the link between the jth input and the kth output, and y_k is the output from the kth output node and the non-linear transformation of the linear sums is catered for by the activation functions $\boldsymbol{\phi}^{(1)}$ and $\phi^{(2)}$. There is no constraint that the same function is used at each layer although this is usually the case. The above equation can be readily generalised to any number of hidden layers. The goal in 'training' a MLP is to find the values of the weight vectors that result in the minimum error between the network output and the desired target value.

The most popular method for training a MLP is the error back-propagation algorithm as popularised by Rumelhart and McClelland (1986). The term back-propagation refers to the signalling back of errors through the MLP network so that adjustments can be made to the weights of the preceding layer. This is often referred to as the credit assignment problem, which refers to the level of responsibility apportioned to previous layers of weights for the output error. The algorithm uses gradient descent search in weight

space to minimise the error defined as

$$E = \frac{1}{N}\sum_{k}(t_k - y_k)^2 \quad (7)$$

where t_k is the target output from the kth node in the output layer and y_k is the actual computed output from the kth node. With respect to each weight the derivative of the error E is set proportional to weight change

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} \quad (8)$$

where the parameter η is the learning rate. This method alone results in very slow progress in weight space and a popular modification is to introduce a momentum term (Plaut et al. 1986) that in effect adds inertia, hence the name. This modified relationship is given by:

$$\Delta w_{jk}(n+1) = \eta(\delta_j y_k) + \alpha \Delta w_{jk}(n) \quad (9)$$

The momentum factor, α, allows the previous weight change to influence the current one and for a constant gradient increases the effective learning rate from η to $\eta/(1-a)$. Inclusion of the momentum term is now so popular it is regarded as a standard feature of most MLP networks.

2.3. Conditional resampling method

The conditional resampling method involves two distinct steps. First, the generation of a 'marker' precipitation series at a key station (or the area averages of several 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.

Full technical details of the Statistical Down Scaling Model (SDSM) and an evaluation of the conditional resampling method are provided by Wilby et al. (2002, 2003). SDSM is best described as a hybrid of stochastic weather generator and regression-based methods, because regional circulation patterns and atmospheric moisture variables are used to condition local weather generator parameters (e.g. precipitation occurrence and intensity) at individual sites. 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.

For multi-site applications, the single site model is first used to generate daily precipitation at a 'marker' site, in this instance the area average amounts (as in Palutikof et al., 2002). Wet-day amounts are resampled from the empirical distribution of area averages conditional on the large-scale atmospheric forcing and a stochastic error term. The actual wet-day amount is determined by mapping the modelled normal cumulative distribution value onto the observed cumulative distribution of amounts at the marker site. The area average is then disaggregated to the constituent amount falling on the same day at each station in the multi-site array.

Resampling in this way preserves both the area average of the marker series and the spatial covariance of the multi-site array (Wilby et al., 2003). Use of area averages (instead of single sites as the marker series), reduces the risk of employing a non-homogeneous or non-representative record, and increases the signal-to-noise ratio of the predictand. As with other resampling methods, the maximum *daily* value 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 previously unseen in the training set.

2.4. Data sets

The precipitation data employed in this study are for sites in the contrasting regions of NWE and SEE. These homogeneous regions were initially defined by Wigley et al. (1984) and Gregory et al. (1991), and are now widely used in analyses of UK precipitation (e.g. Jones and Conway, 1997; Goodess et al., 2003). Daily area averages and multi-site precipitation series were obtained for a network of 13 stations in NWE and 27 stations in SEE for the period 1961–2000 (Fig. 1).

The atmospheric predictor variables originate from the National Center for Environmental Prediction (NCEP) re-analysis data set (Kalnay et al., 1996). The candidate predictor set contained 25 normalised daily predictors (describing atmospheric circulation, thickness, and moisture content at the surface, 850 and 500 hPa), for nine

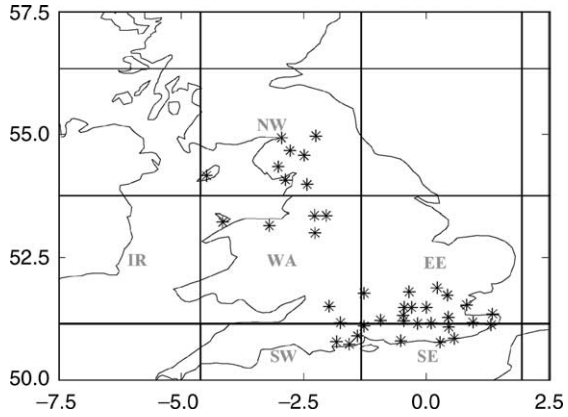

Fig. 1. Distribution of precipitation stations in relation to the grid-boxes used for downscaling in SEE and NWE.

grid-boxes covering the British Isles, for the period 1961–2000 (Table 2). Mean sea level pressures in the NCEP re-analysis are known to have a positive bias from 1941 to 1967, but the worst affected areas lie outside the domain of the present study (see Reid et al., 2001). In addition, the phase difference between the timing of the precipitation day (ending 09GMT), and the NCEP daily averages (indicative of conditions at midday) was accommodated by employing forward lagged daily predictor variables in the downscaling schemes. Following Goodess et al. (2003) predictors were selected from grid boxes overlying target regions (i.e. NWE and SEE), as well as from adjoining westerly grid-boxes (Fig. 1) to capture spatially remote forcing (Wilby and Wigley, 2000; Brinkmann, 2002).

Table 2
Candidate predictor variables available for each grid box

Predictor	Description
TEMP	Mean temperature at 2 m
MSLP	Mean sea level pressure
H850	850 hPa geopotential height
H500	500 hPa geopotential height
USUR	Near surface westerly wind
U850	Westerly wind at 850 hPa
U500	Westerly wind at 500 hPa
VSUR	Near surface southerly wind
V850	Southerly wind at 850 hPa
V500	Southerly wind at 500 hPa
FSUR	Near surface wind strength
F850	Wind strength at 850 hPa
F500	Wind strength at 500 hPa
ZSUR	Near surface vorticity
Z850	Vorticity at 850 hPa
Z500	Vorticity at 500 hPa
DSUR	Near surface divergence
D850	Divergence at 850 hPa
D500	Divergence at 500 hPa
QSUR	Near surface specific humidity
Q850	Specific humidity at 850 hPa
Q500	Specific humidity at 500 hPa
RSUR	Near surface relative humidity
R850	Relative humidity at 850 hPa
R500	Relative humidity at 500 hPa

3. Methodology

Available data were partitioned as follows: 1961–1978 and 1994–2000 for calibration; 1979–1993 for validation (standard periods used by partners in the EU project Statistical and Regional dynamical Downscaling of Extremes for European Regions, STARDEX). Three methods were employed to identify downscaling predictors that have the most significant influence on the predictand. These were SWLR, a compositing procedure and a Genetic Algorithm (GA). A wet-day threshold of 1 mm/day was generally employed, however, other thresholds (5, 10, 15 and 20 mm/day) were also used to test their affect on predictor selection by SWLR.

3.1. Diagnostics

To compare the occurrence process for the three models the Percentage of Forecasts Correct (PFC) and the Critical Success Index (CSI) were employed (Wilks, 1995). The CSI measure is required because a 1 mm threshold gives the proportion of dry days as 60% for the NWE and 70% for the SEE, so clearly a model could predict all dry days and still have a percentage accuracy of 70% for the SEE. The CSI compares the observed and modelled time series in the following way

$$\mathrm{CSI} = M_{11}/(M_{11} + M_{10} + M_{01}) \quad (10)$$

and the PFC is defined as:

$$\mathrm{PFC} = (M_{11} + M_{00})/(M_{00} + M_{10} + M_{01} + M_{11}) \quad (11)$$

Where M_{11} is the number of correct wet-days, M_{10} is modelled wet and observed dry days, M_{01} is observed

Table 3
STARDEX diagnostic tests for daily precipitation

90% Quantile of rain day amounts (mm/day) [PQ90]	
Maximum number of consecutive dry days [PXCDD]	Let R_{ij} be the daily precipitation amount for day i of period j. Then counted is the largest number of consecutive days, where: $R_{ij} \leq$ threshold, where threshold is a user specified variable
Greatest 5 day total rainfall [Px5D]	Let R_{kj} be the precipitation amount for the N day interval k of period j, where k is defined by the last day. Then the maximum N day values for period j are: $\mathrm{PxND}_j = \max(R_{kj})$
Simple Daily Intensity (rain per rain day) [PINT]	Let R_{wj} be the daily precipitation amount for wet-day $w(R >$ wd-cutoff) of period j, where threshold is a user specified variable. Then the mean precipitation amount at wet-days is: $\mathrm{PINT}_j = \sum_{w=1}^{w} R_{wj}/W$
No. of events >90th percentile [PN90]	Let R_{wj} be the daily precipitation amount at wet-day $w(R >$ threshold) of period j and let $R_{wn}90$ be the 90th percentile of precipitation at wet-days in the specified period. Then the percent of time is determined, where: $R_{wj} > R_{wn}90$
The % of total rainfall from events >90th percentile [PF90]	Let R_j be the sum of daily precipitation amount for period j and let R_{wj} be the daily precipitation amount at wet-day w(R(threshold) of period j and $R_{wn}90$ the 90th percentile of precipitation at wet-days in the specified period. Then $PF90_j$ is determined as: $\mathrm{PTR90}_j = \frac{\sum_{w=1}^{w} R_{wj}}{R_j}$, where $R_{wj} > R_{wn}90$

dry and modelled wet-days, and M_{00} is correct dry days.

Model performance was further evaluated using six STARDEX diagnostics of daily precipitation (see Table 3 for definitions), the Root Mean Squared Error (RMSE) and percentage of explained variance (E%). The STARDEX indices focus on extreme events, such as the maximum number of consecutive dry days (PXCDD). The standard deviation of wet-day amounts, lag−1 autocorrelation for individual precipitation series, and inter-site correlation coefficient were also computed for multi-site applications. Finally, quantile-quantile plots of observed and downscaled daily precipitation amounts were produced for representative sites in each region.

3.2. Stepwise multiple linear regression procedure (SWLR)

Series of area-average precipitation occurrence were prepared for the various wet-day exceedance thresholds (see above). SWLR was then applied using all predictor variables for coincident, lag−1 and lag+1 daily time steps. The selection process was halted once the improvement in the correlation between observed and fitted series was less than 1%. Independent validation data were used to verify the skill of predictors selected from the calibration period. Wet-day amounts modelling was undertaken using the same thresholds, noting that increasing wet-day thresholds led to progressively smaller training sets. For example, the NWE calibration set had only 67 days greater than the 20 mm threshold equating to just 0.9% of the data.

3.3. Compositing procedure

SWLR predictor selection is biased towards reproducing *mean* wet-day amounts. In order to weight the predictor variable selection towards *heavier* events a compositing method was trialed. For each region, area-average amounts were ranked in descending order from largest to smallest daily precipitation totals. Averages of the associated predictor variables were then calculated for the top 100 events. The statistical significance of each predictor average was tested using a bootstrap method in which 10,000 means were calculated for 100 randomly selected (with replacement) predictor values. Composited means lying outside the 2.5 and 97.5 percentiles of the bootstrapped distribution were deemed to be significantly different from chance.

3.4. Genetic algorithm (GA)

The pioneering work of Holland (1975) illustrated how the Darwinian evolution process could be applied, in the form of an algorithm, to solve a wide variety of problems. Due to the biological motivation, this highly parallel adaptive system is now called the genetic algorithm (GA). The GA has a population of individuals competing against each other in relation to a measure of fitness, with some individuals reproducing, others dying, and new individuals arising

Table 4
Selection of predictor variables for NWE precipitation using the sub-periods 1961–1978 and 1994–2000

Step	Occurrence			Amounts					
	Predictor	Lag	*r*	Predictor (SWLR)	Lag	*r*	Predictor (composited)	Lag	Weight
1	**U850WA**	**+1**	**0.49**	**ZSURWA**	**+1**	**0.30**	**ZSURWA**	**+1**	**1.38**
2	**Z850WA**	**+1**	**0.63**	**Q500WA**	**0**	**0.42**	**Q500WA**	**0**	**1.28**
3	**ZSURWA**	**0**	**0.65**	**F850WA**	**+1**	**0.49**	**MSLPWA**	**+1**	**1.16**
4	**RSURIR**	**0**	**0.67**	**ZSURWA**	**0**	**0.50**	**Q500IR**	**0**	**1.14**
5	**USURWA**	**0**	**0.68**	**U500WA**	**+1**	**0.52**	**Z850WA**	**+1**	**1.10**
6	F500WA	+1	0.68	**Z500WA**	**+1**	**0.53**	**F500WA**	**+1**	**1.08**
7	Q500WA	0	0.68	**S500IR**	**−1**	**0.54**	**F850WA**	**+1**	**1.07**
8	H500IR	0	0.69	**ZSURIR**	**+1**	**0.55**	**U500WA**	**+1**	**1.01**
9	RSURWA	+1	0.69	QSURIR	0	0.55	U850WA	+1	0.99
10	Q850IR	0	0.70	TEMPWA	+1	0.56	H850WA	+1	0.98

The first four characters in each predictor name refer to Table 2, the last two characters specify the grid box (WA, Wales; IR, Ireland). The wet-day threshold was 1.0 mm/day. Note: bold type indicates the predictor variables used.

through combination and mutation. In this instance, the GA uses the trained RBF network as a fitness function so that the modelled/observed error could be minimised to evolve an optimal subset of predictor variables, which had been encoded into each member of the population.

4. Results

The results are split into four parts. Section 4.1 deals with predictor selection, Section 4.2 with precipitation occurrence, Section 4.3 with variability, autocorrelation and quantiles of daily amounts at individual sites, and Section 4.4 with diagnostics of multi-site behaviour.

4.1. Predictor selection

Tables 4 and 5 show the leading predictor variables identified by SWLR (without seasonal stratification) for NWE and SEE, respectively, using a 1 mm/day threshold. Tables 4 and 5 also report the ten most heavily weighted predictors identified by the compositing procedure for each region. The results highlight the gain of employing forward lagged predictors, which are thought to compensate for the mismatch between the NCEP day and the precipitation day.

Table 5
As Table 4 but for SEE

Step	Occurrence			Amounts					
	Predictor	Lag	*r*	Predictor (SWLR)	Lag	*r*	Predictor (composited)	Lag	Weight
1	**ZSURSW**	**1**	**0.53**	**ZSURSE**	**1**	**0.36**	**ZSURSE**	**1**	**1.87**
2	**MSLPSW**	**0**	**0.59**	**Q500SE**	**0**	**0.43**	**Z850SE**	**1**	**1.66**
3	**MSLPSE**	**0**	**0.61**	**VSURSE**	**0**	**0.48**	**Z850SW**	**1**	**1.54**
4	**RSURSE**	**1**	**0.62**	**FSURSE**	**1**	**0.5**	**ZSURSW**	**1**	**1.51**
5	**Z500SW**	**1**	**0.64**	**Z850SE**	**1**	**0.51**	**MSLPSE**	**1**	**1.39**
6	**ZSURSE**	**0**	**0.65**	RSURSE	1	0.51	MSLPSE	1	1.33
7	F500SE	1	0.65	DSURSE	0	0.52	H850SE	1	1.19
8	USURSW	0	0.66	F500SE	1	0.52	H850SW	1	1.17
9	ZSURSE	1	0.66	F850SE	0	0.52	Z500SW	1	1.09
10	Q500SE	0	0.67	QSURSE	0	0.53	Z500SE	1	1.06

SW, South West England; SE, Southern England.

In contrast, backward lagged predictors were noteworthy by their absence, possibly reflecting the use of grid boxes to the west of the target regions (i.e. spatio-temporal substitution). Overall, secondary (airflow) variables tended to dominate over moisture variables for both selection methods and regions, with NWE exhibiting slightly more predictability than SEE.

Vorticity at the surface (ZSUR) and at 850 hPa (Z850) were the most frequently selected variables for precipitation occurrence and amounts in both regions, consistent with previous analyses for the UK (Conway et al., 1996). Specific humidity at 500 hPa (Q500) was second for both methods in NWE and for the SWLR set in SEE. Beyond this, there was little consistency amongst the variables chosen by the two methods. Mean sea level pressure (MSLP) was ranked in third and fifth positions for NWE and SEE, respectively, by compositing, but was not selected by SWLR. Relative humidity at 500 hPa (R500) and 850 hPa (R850) were notable omissions given their prominence in other studies (e.g. Beckman and Buishand, 2001).

Predictor variables from the Wales (WA) grid box tend to influence NWE amounts; whereas the balance was more even between WA and IR (Ireland) for occurrence (Fig. 1). Similarly, for SEE the southern England (SE) grid box dominates. This supports the view that downscaling should involve the use of both propinquitous and remote predictors of precipitation (Wilby and Wigley, 2000; Brinkmann, 2002). This is an implicit assumption of traditional methods involving airflow indices or circulation patterns covering large spatial domains. An added advantage is that mixtures of grid-boxes could help reduce anomalies in GCM predictors due to smoothing of the underlying orography.

The wet-day threshold was also found to influence SWLR predictor selection (not shown). For example, a threshold of 5 mm/day yielded only two predictors in common with the 1 mm/day threshold model for NWE for both the occurrence (Q500WA, U850WA) and amounts (Q500IR, ZSURWA). However, there was greater overlap between the SWLR amounts model and the predictors selected by compositing (Q500IR, H850WA, ZSURWA, F500WA) suggesting some convergence between the methods at higher thresholds. This might be expected because the smallest wet-day total in the NWE (SEE) composited set was 18.48 (17.74) mm, and the average daily total was 22.80 (23.43) mm. In contrast, results for SEE showed more agreement between the 1 mm/day and 5 mm/day SWLR amount models (ZSURSE, Q500SE, FSURSE, QSURSE) and less with the compositing (only ZSURSE). Overall, models with predictors obtained for higher thresholds yielded less explained variance than the 1 mm/day models in both NWE and SEE.

Correlation coefficients between downscaled and observed STARDEX indices for area-averages indicated that the SWLR had an advantage over the compositing procedure in the vast majority of cases although gains were marginal in the case of NWE (Tables 6 and 7). This view is supported by the overall

Table 6
Correlations between observed and downscaled area average precipitation diagnostics in NWE using 1979–1993

Diagnostic	SWLR			Composited		
	Winter	Summer	Annual	Winter	Summer	Annual
PQ90	**0.34**	**0.44**	-0.04	0.24	0.27	-0.02
PXCDD	0.77	0.7	0.86	0.77	0.7	0.86
PX5D	0.29	**0.22**	-0.23	**0.41**	**0.22**	-0.26
PINT	**0.44**	**0.21**	0.37	0.43	**0.21**	**0.54**
PN90	**0.36**	**0.41**	**0.17**	0.26	0.16	-0.18
PF90	**0.63**	**0.38**	**0.09**	0.57	0.18	-0.07
Average r	**0.47**	**0.39**	**0.2**	0.45	0.29	0.14
RMSE (mm)	**3.1**			3.13		
E (%)	**48**			47		

Results are stratified by season and by method of predictor variable selection. The overall Root Mean Squared Error (RMSE) and percentage explained variance (E %) of daily amounts are also shown. Note: the best values are in bold, ties are greyed. PXCDD values are the same because identical occurrence series were employed.

Table 7
Correlations between observed and downscaled area average precipitation diagnostics in SEE using 1979–1993

Diagnostic	SWLR			Composited		
	Winter	Summer	Annual	Winter	Summer	Annual
PQ90	**0.6**	**0.15**	**0.32**	0.28	-0.21	0.18
PXCDD	0.57	0.29	0.03	0.57	0.29	0.03
PX5D	**0.82**	**0.31**	**0.58**	0.53	0.18	0.29
PINT	**0.8**	**0.07**	**0.39**	0.48	-0.31	0.11
PN90	**0.44**	-0.01	**0.29**	0.19	-0.37	0.03
PF90	**0.68**	**0.37**	**0.39**	0.49	0.24	0.11
Average *r*	**0.65**	**0.2**	**0.33**	0.42	-0.03	0.12
RMSE (mm)	**2.86**			3.03		
E (%)	**44**			37		

RMSE and *E*% statistics. Therefore, predictor variables identified by SWLR were utilised for subsequent multi-site investigations. As an aside, the GA approach yielded far too many predictors for statistical regression applications despite efforts to constrain selections via penalty functions and, consequently, was pursued no further.

4.2. Precipitation occurrence

The RBF configuration for NWE comprised 30 nodes with an inverse quadratic transfer function and, for the MLP, a 13-node network trained for 2000 epochs. For SEE the RBF comprised 19 nodes with an inverse quadratic function and, the MLP had 11 nodes and 1000 epochs. These configurations did not yield the smallest overall RMSE in the validation set. Rather, to avoid over-fitting, the network with the least number of nodes and stable RMSE was chosen (i.e. just after the error measure begins to equalise for the validation set).

The results for NWE and SEE (Tables 8 and 9, respectively) indicate that the occurrence skill of the ANNs was very similar. However, SDSM was less skillful in both regions, achieving for NWE an average percentage correct of 70% and CSI of 0.49, compared with 77% and 0.59 for the RBF. Similarly for SEE, SDSM achieved an average percentage correct of 72% and CSI of 0.35 compared with 80% and 0.35 for the RBF. All models had less skill downscaling precipitation occurrence in SEE, where east-coast stations such as Manston proved to be most problematic.

Table 8
Model comparison for multi-site precipitation occurrence in NWE

Station	PFC			CSI		
	RBF	MLP	SDSM	RBF	MLP	SDSM
HAYDON BRIDGE	75	70	67	0.50	0.51	0.42
LYME PARK RESR	77	77	70	0.61	0.61	0.50
MORECAMBE	78	79	70	0.61	0.61	0.50
GRIZEDALE	79	79	71	0.67	0.67	0.54
APPLEBY CASTLE	77	75	70	0.58	0.58	0.48
DOUGLAS	76	76	69	0.59	0.60	0.47
KEELE	77	76	69	0.56	0.56	0.46
PEN-Y-FFRIDD	76	74	71	0.60	0.60	0.50
LOGGERHEADS	76	75	70	0.56	0.58	0.48
RINGWAY	77	71	69	0.57	0.55	0.47
SLAIDBURN	80	79	72	0.67	0.67	0.54
NEWTON RIGG	79	79	71	0.60	0.60	0.50
CARLISLE	77	78	70	0.56	0.56	0.47
Mean	77	76	70	0.59	0.59	0.49

Table 9
Model comparison for multi-site precipitation occurrence in SEE

Station	PFC			CSI		
	RBF	MLP	SDSM	RBF	MLP	SDSM
Greenwich	79	79	72	0.44	0.48	0.34
Falconhurst	81	80	72	0.52	0.52	0.38
Dover W.Wks	78	78	70	0.45	0.47	0.33
Bedford (RAE)	79	79	72	0.42	0.40	0.32
Writtle	79	77	72	0.40	0.30	0.31
Shoeburyness	79	79	72	0.36	0.39	0.30
Stansted	78	79	71	0.44	0.47	0.33
Heathrow	81	81	73	0.45	0.48	0.34
Oxford	81	81	73	0.49	0.47	0.35
Long sutton	81	80	72	0.52	0.48	0.37
Rothamsted	80	79	72	0.48	0.45	0.36
Wisley	81	80	73	0.46	0.43	0.35
Gatwick	81	80	73	0.51	0.51	0.38
Kew (Roy Bot Gdns)	80	80	73	0.46	0.48	0.35
Goudhurst	81	80	73	0.51	0.45	0.38
East malling	80	80	72	0.45	0.48	0.34
Manston	78	78	71	0.37	0.42	0.28
Wye	78	78	72	0.44	0.36	0.34
Hastings	80	80	72	0.46	0.43	0.34
Eastbourne	80	79	72	0.48	0.43	0.35
Rustington	82	82	73	0.51	0.53	0.36
Martyr worthy	82	80	73	0.52	0.46	0.37
Southampton	82	82	73	0.51	0.52	0.35
Everton	82	81	72	0.51	0.47	0.35
Boscombe down	82	81	72	0.53	0.48	0.37
Hurn	82	82	73	0.53	0.53	0.36
Lyneham	82	81	72	0.55	0.50	0.35
Mean	80	80	72	0.47	0.46	0.35

4.3. Variability, autocorrelation and quantiles of amounts at individual sites

Figs. 2 and 3 compare the standard deviations of observed and downscaled daily precipitation amounts for winter, summer and annual for all stations in NWE and SEE. The ANNs consistently under-represented standard deviations (most notably at the high elevation sites of Grizedale and Slaidburn), whereas SDSM consistently over-represented variability in amounts (but to a lesser degree). The seasonal biases in the standard deviations of observed and modelled daily precipitation amounts were up to 35% for the ANN models, compared with ~14% for SDSM. Conversely, for SEE winter and annual amounts the MLP returned biases of 20 and 14%, respectively, compared with 35 and 45% for SDSM. The situation was reversed for summer with biases of 33% for the MLP and 9% for SDSM. In both regions, and all seasons, the RBF performance was worse than the MLP.

Figs. 2 and 3 also compare the lag−1 autocorrelations for which the ANNs show consistent over-estimation. This probably reflects the lack of backward-lagged predictors (see Tables 4 and 5) or an explicit autoregressive mechanism in the ANNs. In comparison, SDSM provides better skill, but tended to under-estimate serial correlation, in particular for winter in SEE. This was attributed to the large stochastic component of modelled precipitation amounts in SDSM.

Downscaling model skill for precipitation amount distributions was assessed at representative sites in NWE (Grizedale [upland], Douglas [maritime], Ringway [lowland]) and in SEE (Rothamstead [lowland], Eastbourne [coastal], Oxford [lowland]).

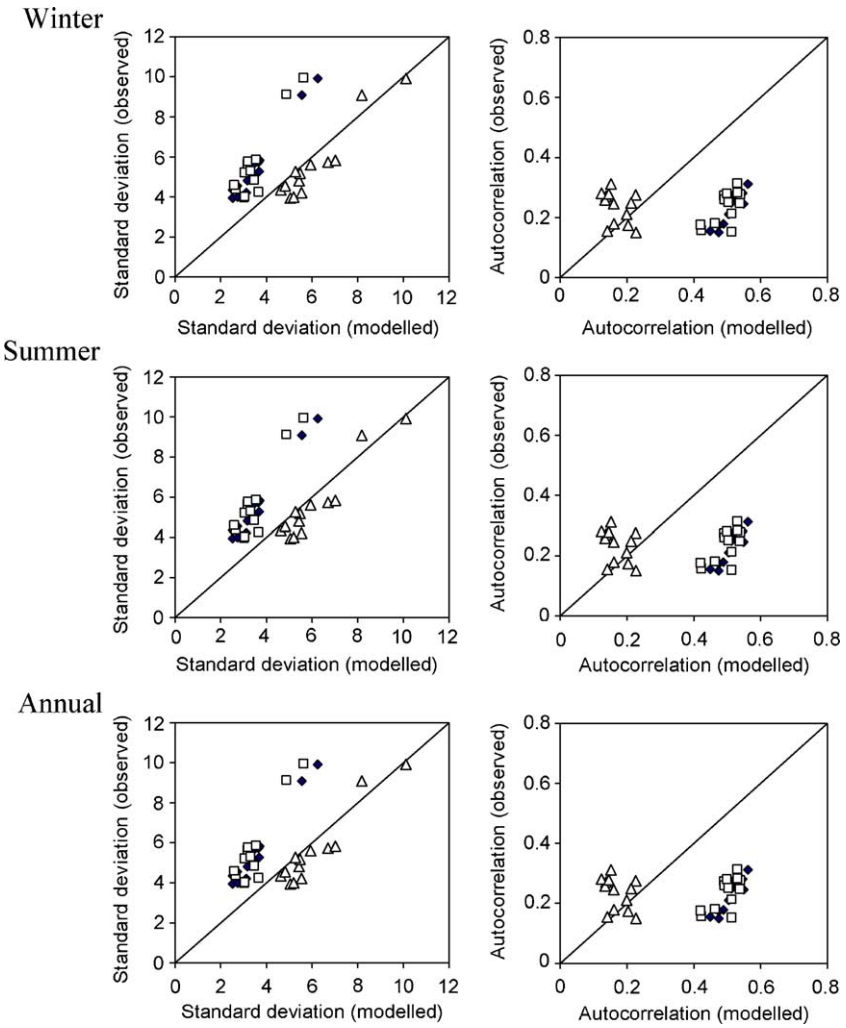

Fig. 2. Multi-site standard deviations and autocorrelations for NWE. Key: Black diamonds (RBF); squares (MLP); triangles (SDSM).

Quantile-quantile plots reveal that the ANNs consistently under-estimated daily precipitation amounts exceeding 10 mm/day in both regions (Figs. 4 and 5). This pattern of behaviour was also exhibited when data were stratified by season (not shown), highlighting the influence of the large number of smaller events on ANN weights. Conversely, SDSM slightly over-estimated quantiles except for the maximum daily totals at Ringway and Eastbourne. The sign of the anomalies of the largest events depend critically on the magnitude of the heaviest precipitation event resampled from the training set. SDSM performs relatively well because the whole distribution is being evaluated, rather than the time-series attributes (see below).

4.4. Diagnostics of multi-site behaviour

Both ANNs performed relatively poorly with respect to the inter-site correlation (Figs. 6 and 7) due to the fully deterministic nature of the models leading to over-estimation of inter-site correlation. In comparison, the conditional resampling method of SDSM has an 'adjustable' degree of randomness yielding more realistic inter-site behaviour (Wilby et al., 2003). Even so, there is a seasonal bias in

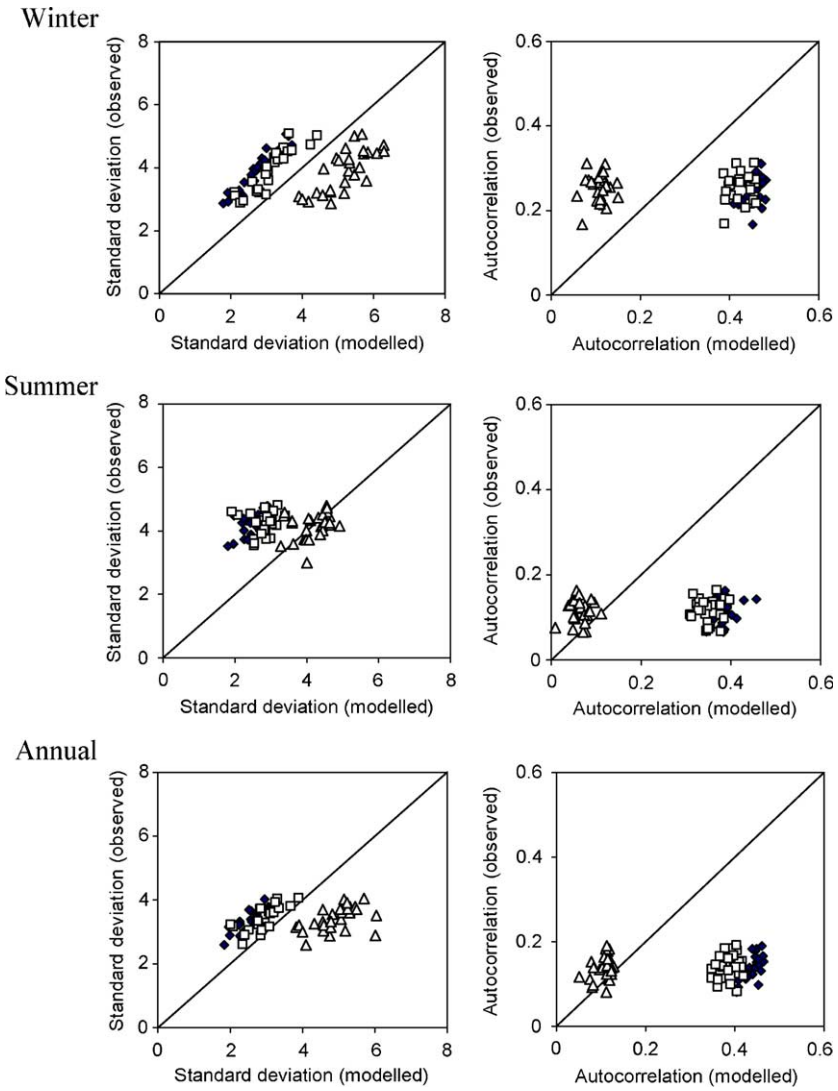

Fig. 3. Multi-site standard deviations and autocorrelations for SEE. Key: Black diamonds (RBF); squares (MLP); triangles (SDSM).

the results with inter-site correlation too weak in winter and too strong in summer. The skill of SDSM was also generally greater in NWE than SEE, presumably due to the greater influence of large-scale Atlantic weather systems in the former region.

Box and Whisker plots (Figs. 8 and 9) indicate that seasonal series of STARDEX diagnostics (with the exception of PXCCD) are modelled marginally better in SEE than NWE although the regional differences are less pronounced in the case of SDSM. This is partly an artefact of the different sample sizes, the chosen metric of skill, and range of behaviour in each region. Interpretations are further tempered by the very low skill for most indices (other than PXCDD and PX5D). Nonetheless, all models performed better in winter than summer, presumably due to the more vigorous circulation and scale of rain-bearing systems at this time.

For NWE all models consistently performed best for the maximum number of consecutive dry days (PXCDD) with both ANNs outperforming SDSM. This is to be expected because PXCDD

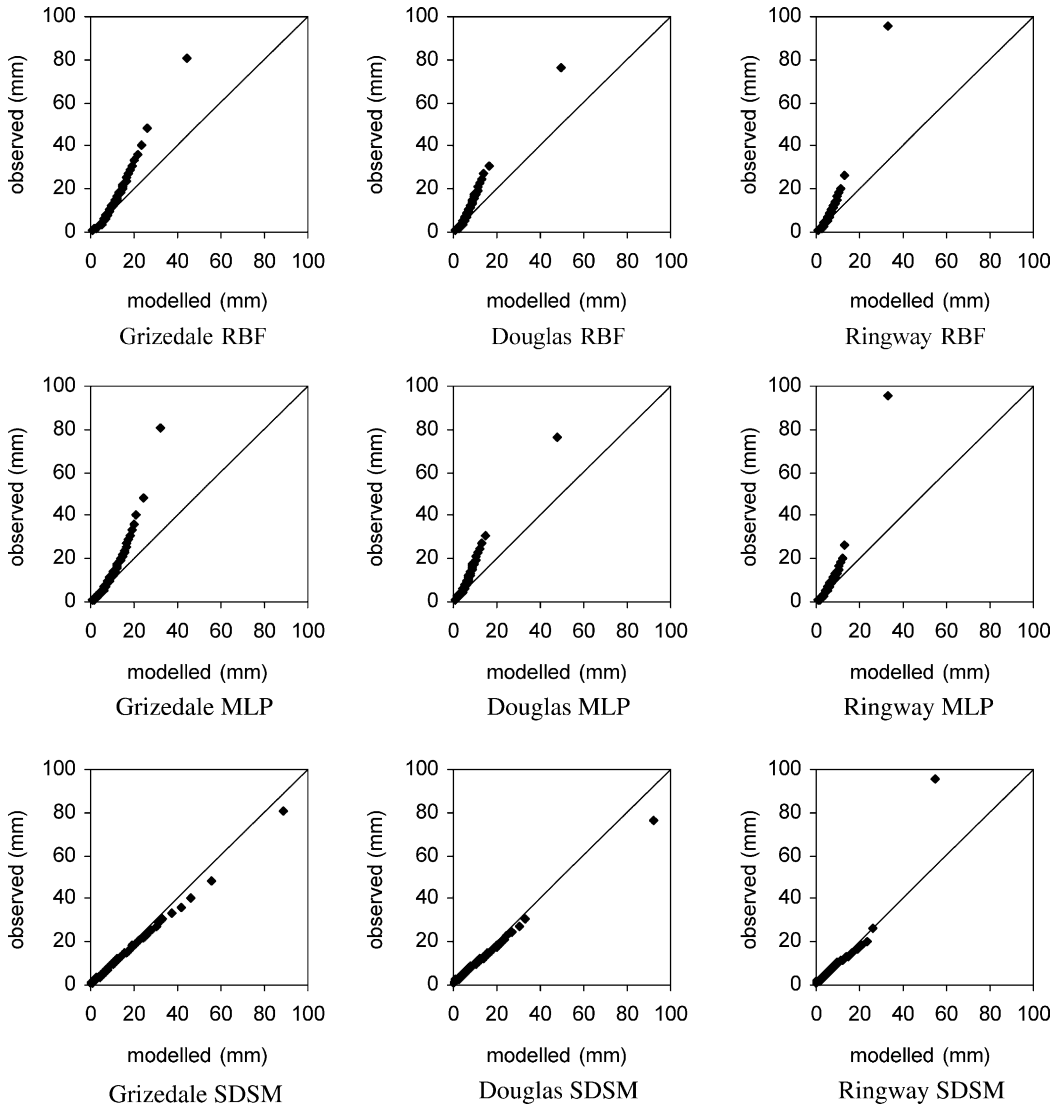

Fig. 4. Quantile-quantile plots for wet-day amounts at selected sites in NWE for the period 1979–1993.

depends on the realism of the daily occurrence process which was captured best by the ANNs. In contrast, for SEE, the only skillful results were for PX5D and PF90 in winter. Overall, the ANNs performed marginally better than SDSM, although the strength of the correlation between observed and downscaled annual series was insignificant for many STARDEX diagnostics, most notably PINT and PX5D in summer.

5. Discussion and conclusions

Models of area-average and multi-site daily precipitation were compared for northwest (NWE) and southeast (SEE) England with particular emphasis on heavy precipitation events. It was found that stepwise linear regression (SWLR) was generally more effective than compositing for predictor variable selection when a Radial Basis Function (RBF) model

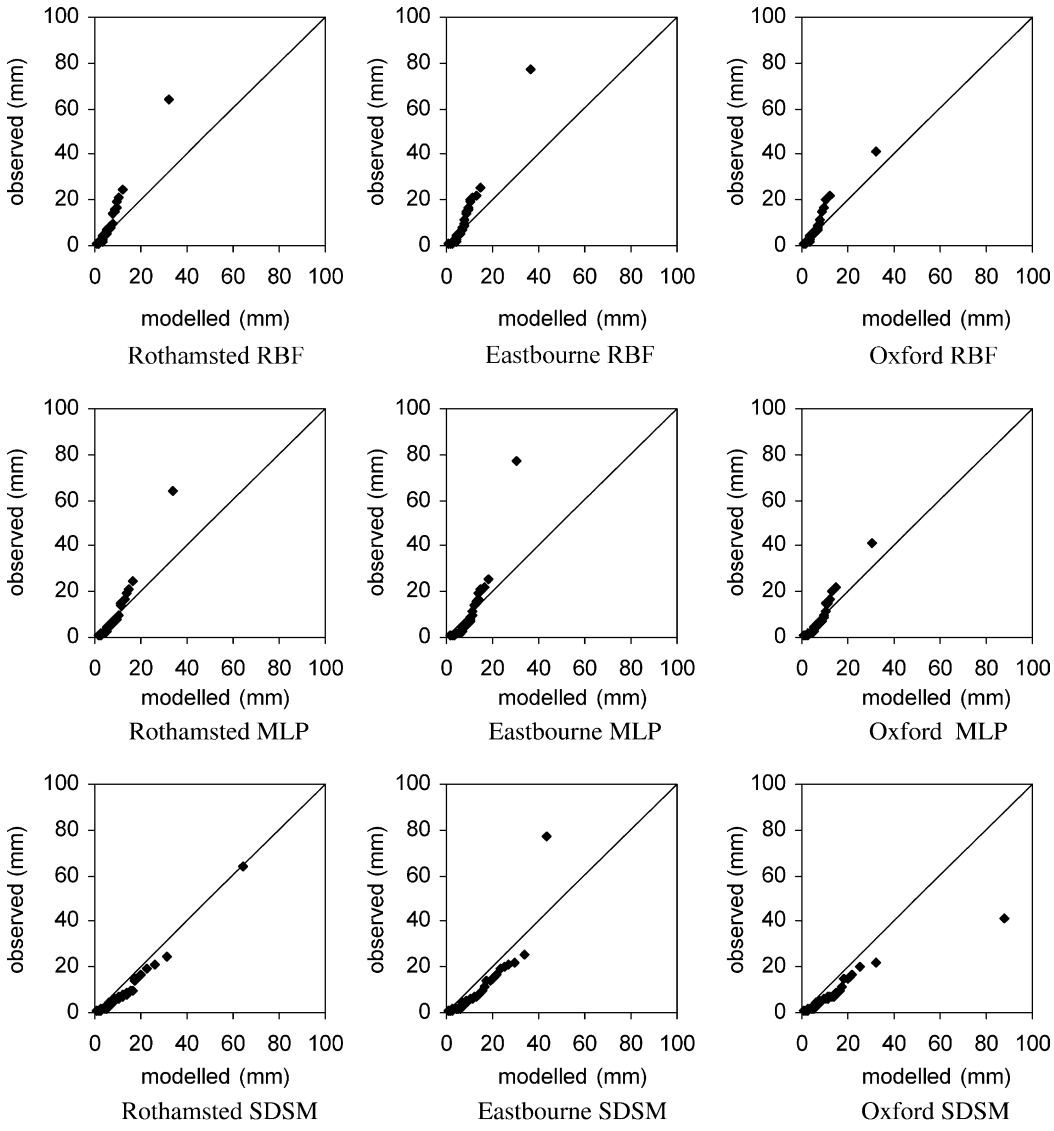

Fig. 5. Quantile-quantile plots for wet-day amounts at selected sites in SEE for the period 1979–1993.

of area average precipitation was used to assess predictor skill against data not used for calibration (1979–1993). However, the overall gain was marginal. The importance of performing separate SWLR for precipitation occurrence and amounts series was shown by differences in the resulting predictor sets. The SWLR and compositing analyses further confirm the over-riding importance of vorticity and specific humidity as downscaling predictors for daily precipitation. Predictor selection was also sensitive to the choice of wet-day threshold.

Having selected the predictors, RBF and Multi Layer Perceptron (MLP) models were re-calibrated using multi-site daily precipitation amounts in each region. A conditional resampling model (SDSM) was also trained using daily precipitation series at the same sites. The realism of downscaled precipitation amounts generated by the three models was then

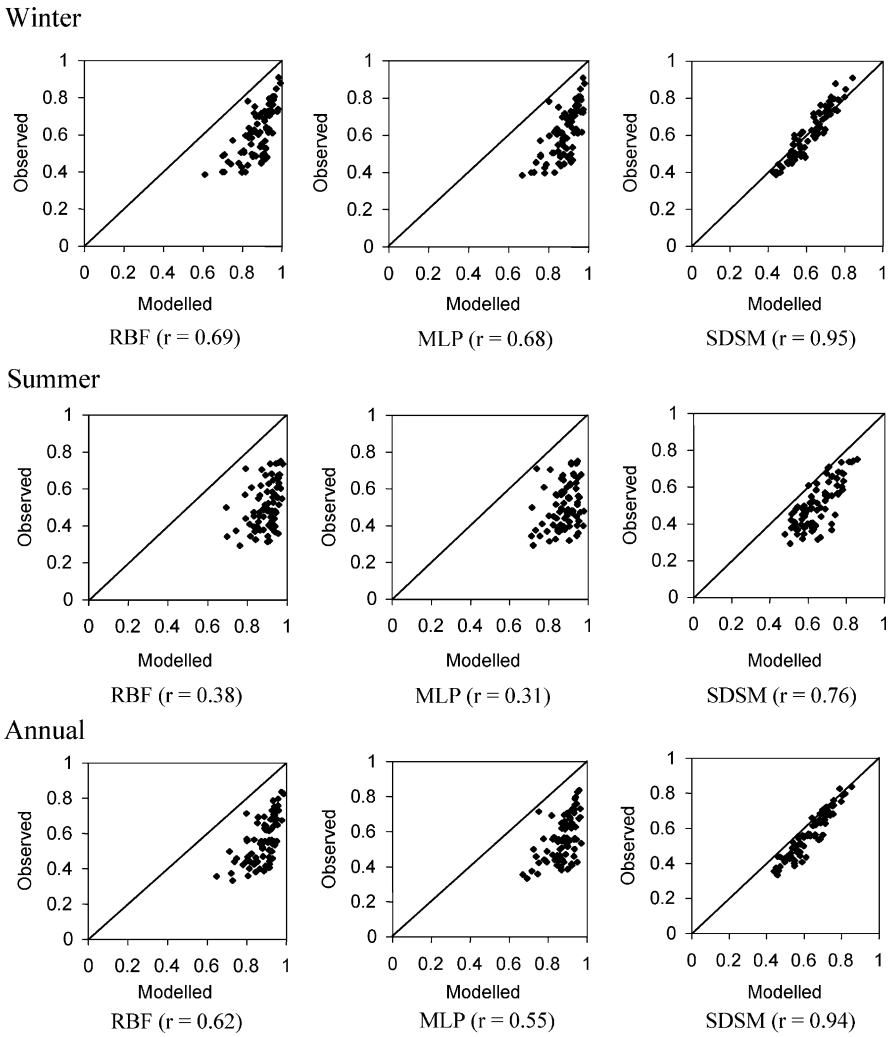

Fig. 6. Inter-site correlations for NWE.

compared using a range of diagnostics for heavy precipitation events, as well as for temporal and spatial dependency. None of the models performed consistently well for all STARDEX indices. Overall, the best results were returned by the ANNs for the maximum number of consecutive dry days (PXCDD) in NWE (all seasons) and greatest 5 day total rainfall (PX5D) in SEE (winter). This suggests that the ANNs captured some attributes of persistent, large-scale events with reasonable skill. In contrast, annual series of quantiles (PQ90) and exceedance thresholds (PN90) for individual daily amounts were downscaled with least skill. For SDSM, this is partly explained by the large stochastic component to the amounts modelling which means that any downscaled precipitation *series* would not be expected to match observations, although summary statistics should concur. This was confirmed by an analysis of amount quantiles downscaled by SDSM.

The ability of downscaling models to capture prolonged heavy rainfall events is particularly important given observed trends in such extremes (Fowler and Kilsby, 2003), and recent severe autumn/winter flooding in the UK (Lamb, 2001). Models should ideally replicate joint occurrence of such events at multiple sites. The present study

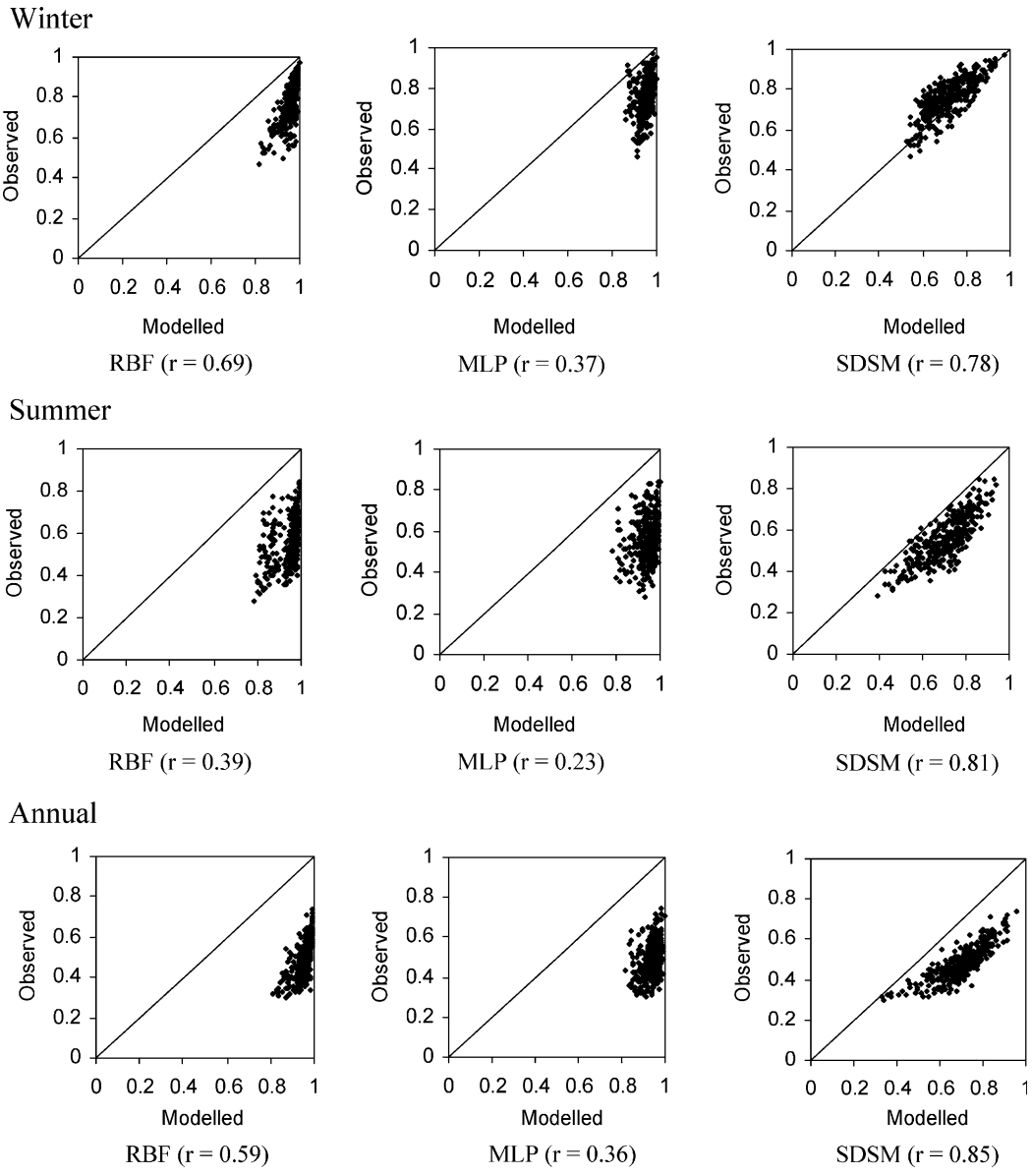

Fig. 7. Inter-site correlations for SEE.

suggests that these features can be downscaled by SDSM with moderate success in winter, but less so in summer. However, further research is needed to account for seasonal and regional variations in model skill, and to assess the capability of dynamical models under the same conditions. Other work is beginning to reveal how statistically and dynamically downscaled heavy rainfall might translate into changing flood frequencies in large river catchments (Reynard et al., 2004).

The ANNs were more skillful than SDSM regarding occurrence modelling at individual sites, returning percent correct forecasts ~80% for the MLP. However, the ANNs significantly and consistently over-estimated inter-site correlations of daily amounts. This systematic bias was attributed to

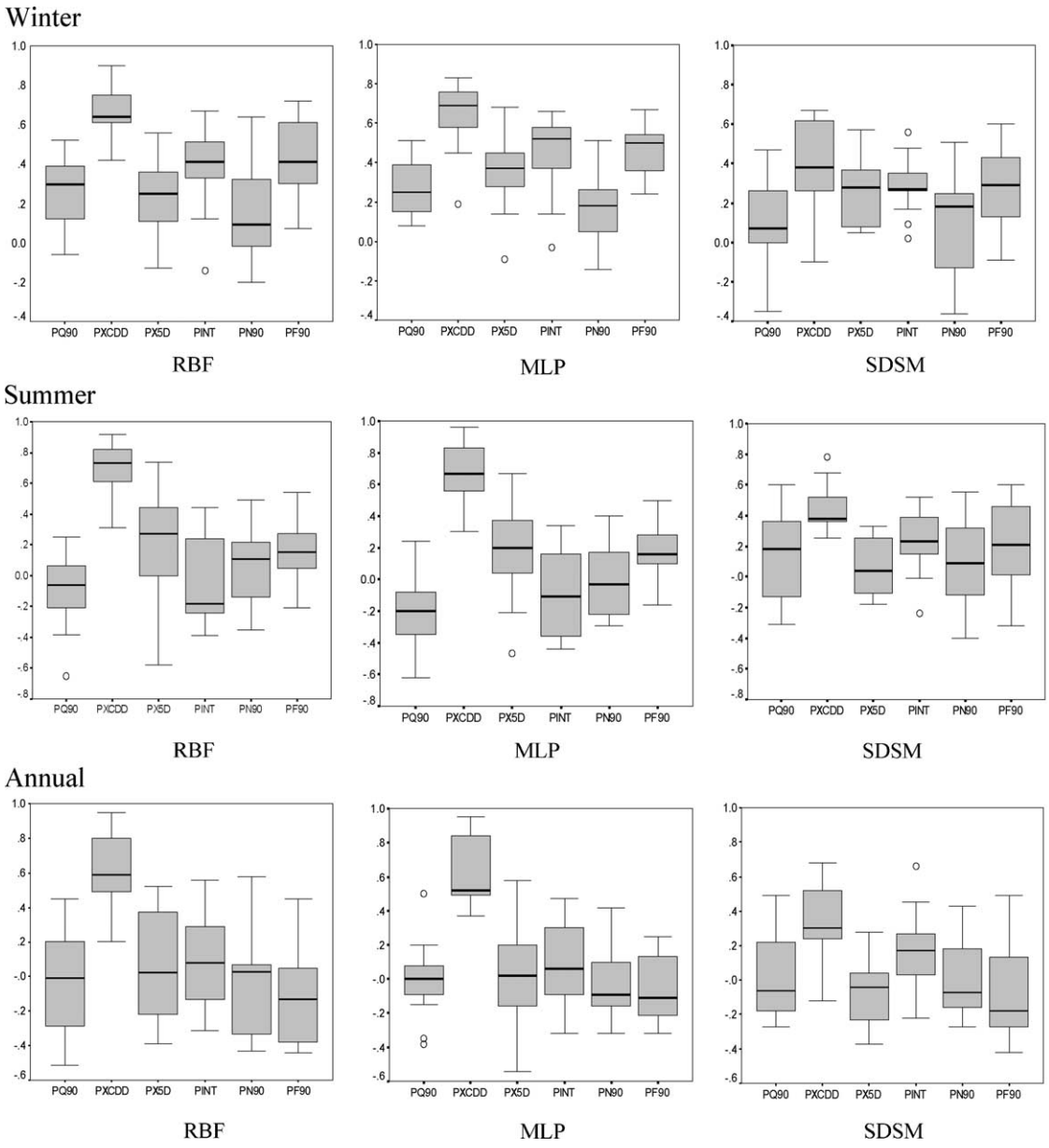

Fig. 8. Distribution of correlation coefficients for observed and downscaled multi-site STARDEX diagnostics. The horizontal line shows the median value, the box represents the interquartile range, the whiskers include 95% of values, and the circles denote outliers. These results are based on all sites in NWE.

the fully deterministic forcing of amounts, whereas the stochasticity of SDSM leads to greater heterogeneity in the response surface and hence more realistic (lower) inter-site relations. Even so, spatial autocorrelation was still too high amongst downscaled summer series due, presumably, to the difficulty of resolving isolated convective storms.

Unlike previous studies, separate ANN models were not constructed for each station. Rather, multi-site behaviour was represented by a single set of

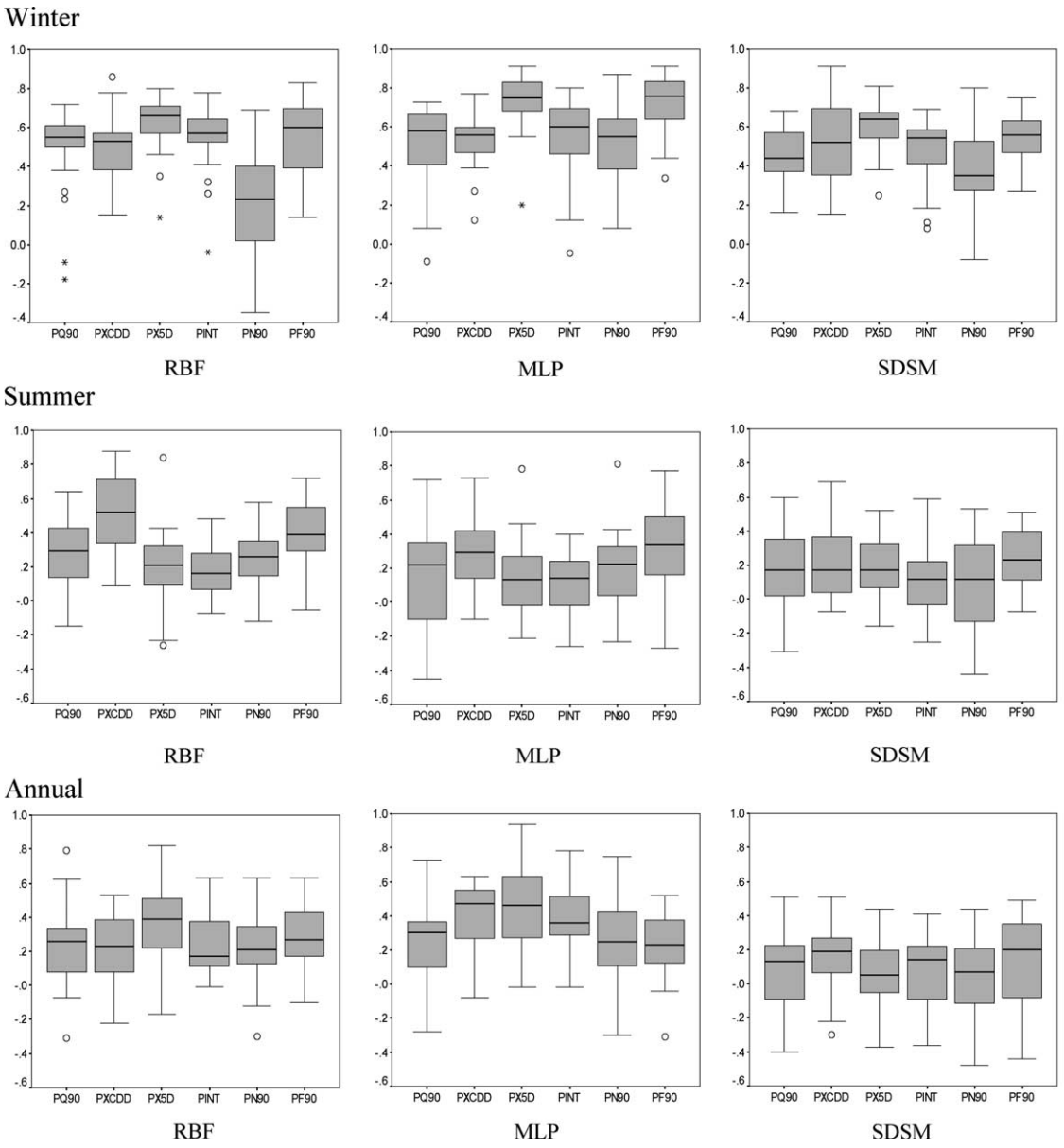

Fig. 9. Distribution of correlation coefficients for observed and downscaled multi-site STARDEX diagnostics. The horizontal line shows the median value, the box represents the interquartile range, the whiskers include 95% of values, and the circles denote outliers. These results are based on all sites in SEE.

model weights and transfer functions for each regional array. This raises the question as to what extent such generalised models can resolve fixed (e.g. topographic), variable (e.g. soil moisture) and random (i.e. unexplained) local forcing in transfer function weights (Hewitson and Crane, *pers. comm.*). Training a single ANN against multiple outputs has the advantage of parsimony but it is suspected that only large-scale forcing is being captured in downscaled precipitation series. (On this basis, predictability was marginally greater in NWE than SEE). The realism of downscaled precipitation changes will, therefore,

hinge on the extent to which the variable local forcing or large-scale forcing predominates under future climate conditions. These mechanisms could be explored by interrogating inter-variable relationships amongst the precipitation, atmospheric and land-surface schemes of dynamical models under current and future climate conditions. Ultimately, however, all downscaling methods depend on the faithful reproduction of large-scale climatology by the host GCM.

Acknowledgements

This research was supported in part by STARDEX (STAstical and Regional Dynamical downscaling of EXtremes for European regions) under the European Community Research Programme (EVK2-CT-20010015). The authors also thank Tim Osborn for initial guidance on the choice of predictor variables in the two regions.

References

Anderberg, M.R., 1973. Cluster Analysis for Applications. Academic Press, New York.

Beckmann, B.R. Buishand, T.A. 2001. KNMI contribution to the European project WRINCLE: downscaling relationships for precipitation for several European sites. Technical Report TR-230, KNMI, De Bilt.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Brinkmann, W.A.R., 2002. Local versus remote grid points in climate downscaling. Climate Research 21, 27–42.

Broomhead, D.S., Lowe, D., 1988. Multivariable function interpolation and adaptive networks. Complex Systems 2, 321–355.

Cannon, A.J., Whitfield, P.H., 2002. Downscaling recent streamflow conditions in British Columbia Canada using ensemble neural network models. Journal of Hydrology 259, 36–151.

Cavazos, T., 1997. Downscaling large-scale circulation to local winter rainfall in north-eastern Mexico. International Journal of Climatology 17, 1069–1082.

Cavazos, T., 1999. Large scale circulation anomalies conductive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. Journal of Climate 12, 1506–1523.

Cavazos, T., 2000. Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. Journal of Climate 13, 1718–1732.

Conway, D., Wilby, R.L., Jones, P.D., 1996. Precipitation and air flow indices over the British Isles. Climate Research 7, 169–183.

Crane, R.G., Hewitson, B.C., 1998. Doubled CO_{2-} precipitation changes for the Susquehanna basin: down-scaling from the genesis general circulation model. International Journal of Climatology 18, 65–76.

Crane, R.G., Yarnal, B., Barron, E.J., Hewitson, B.C., 2002. Scale interactions and regional climate: examples from the Susquehanna River basin. Human and Ecological Risk Assessment 8, 147–158.

Fowler, H.J., Kilsby, C.G., 2003. A regional frequency analysis of UK extreme rainfall from 1961 to 2000. International Journal of Climatology 23, 1313–1334.

Giorgi, F., Mearns, L.O., 1991. Approaches to the simulation of regional climate change A review. Rev. Geophys. 29, 191–216.

Goodess, C., Osborn, T., Hulme, M.The identification and evaluation of suitable scenario development methods for the estimation of future probabilities of extreme weather events. Tyndall Centre for Climate Change Research, Technical Report.4 2003.

Gregory, J.M., Jones, P.D., Wigley, T.M.L., 1991. Precipitation in Britain: an analysis of area-average data updated to 1989. International Journal of Climatology 11, 331–345.

Hewitson, B.C., Crane, R.G., 1992. Large scale atmospheric controls on local precipitation in tropical Mexico. Geophysical research letters 19, 1835–1838.

Hewitson, B.C., Crane, R.G., 1996. Climate downscaling: techniques and application. Climate Research 7, 85–95.

Holland, J., 1975. Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximaters. Neural Networks 2, 359–366.

Hsieh, W.W., Tang, B., 1998. Applying neural network models to prediction and data analysis in meteorology and oceanography. Bull. Am. Meteorol. Soc. 79, 1855–1870.

Jones, P.D., Conway, D., 1995. Precipitation in the British Isles: an analysis of area-average data updated to 1995. International Journal of Climatology 17, 427–438.

Kalnay, E., Kanamitsu, M., Kistler, R., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437–471.

Lamb, R. 2001. To what extent can the October/November 2000 floods be attributed to climate change? Defra FD2304 Final Report.

McGinnis, D.L., 1997. Estimating climate-change impacts on Colorado Plateau snowpack using downscaling methods. Professional Geogarapher 49, 117–125.

Moody, J.E., Darken, C.J., 1989. Fast learning in networks of locally-tuned processing units. Neural Computation 1, 281–294.

Mpelasoka, F.S., Mullan, A.B., Heerdegen, R.G., 2001. New Zealand climate change information derived by multivariate statistical and artificial neural network approaches. International Journal of Climatology 21, 1415–1433.

Olsson, J., Uvo, C.B., Jinno, K., 2001. Statistical atmospheric downscaling of short-term extreme rainfall by neural networks. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere 26, 695–700.

Palutikof, J.P., Goodess, C.M., Watkins, S.J., Holt, T., 2002. Generating rainfall and temperature scenarios at multiple sites: examples from the Mediterranean. Journal of Climate 15, 3529–3548.

Plaut, D., Nowlan, S., Hinton, G.E., 1986. Experiments on learning by back propagation. Technical Report CMU-CS-86-126, Department of Computer Science. Carnegie Mellon University, Pittsburgh, PA.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling W.T. Press. 1993. Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press.

Reid, P.A., Jones, P.D., Brown, O., Goodess, C.M., Davies, T.D., 2001. Assessments of the reliability of NCEP circulation data and relationships with surface climate by direct comparisons with station based data. Climate Research 17, 247–261.

Reusch, D.B., Alley, R.B., 2002. Automatic weather stations and artificial neural networks: Improving the instrumental record in West Antarctica. Monthly Weather Review 130, 3037–3053.

Reynard, N., Crooks, S., Wilby, R.L. Kay, A. 2004. Climate change and flood frequency in the UK. Proceedings of the Defra National Conference, University of York, UK.

Rumelhart, D.E., McClelland, J.L., 1986. Parallel Distributed Processing, Explorations in the Microstructure of Cognition, vol. 1. MA:MIT Press, Cambridge.

Sailor, D.J., Hu, T., Li, X., Rosen, J.N., 2000. A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. Renewable Energy 19, 359–378.

Schoof, J.T., Pryor, S.C., 2001. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. International Journal of Climatology 21, 773–790.

Schubert, S., Henderson-Sellers, A., 1997. A statistical model to downscale local daily temperature extremes from synoptic-scale atmospheric circulation patterns in the Australian region. Climate Dynamics 13, 223–234.

Snell, S.E., Gopal, S., Kaufmann, R.K., 2000. Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling GCMs. Journal of Climate 13, 886–895.

Trigo, M., Palutikof, J.P., 1999. Simulation of daily temperatures for climate change scenarios over Portugal: A neural network approach. Climate Research 13, 45–59.

Trigo, M., Palutikof, J.P., 2001. Precipitation scenarios over Iberia: A comparison between direct GCM output and different downscaling techniques. Journal of Climate 14, 4422–4446.

von Storch, H., Zorita, E., Cubasch, U., 1993. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. Journal of Climate 6, 1161–1171.

Wigley, T.M.L., Lough, J.M., Jones, P.D., 1984. Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series. International Journal of Climatology 4, 1–25.

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography 21, 530–548.

Wilby, R.L., Wigley, T.M.L., 2000. Precipitation predictors for downscaling: observed and General Circulation Model relationships. International Journal of Climatology 20, 641–661.

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002. SDSM - a decision support tool for the assessment of regional climate change impacts. Environmental Modelliing & Software 17, 147–159.

Wilby, R.L., Tomlinson, O., Dawson, C.W., 2003. Multi-site simulation of precipitation by conditional resampling. Climate Research 23, 183–194.

Wilks, D.S., 1995. Statisistical Methods in the Atmospheric Sciences. Academic Press, San Diego, California.

Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic weather models. Progress in Physical Geography 23, 329–357.

Winkler, J.A., Palutikof, J.P., Andresen, J.A., Goodess, C.M., 1997. The simulation of daily temperature series from GCM output Part II: Sensitivity analysis of an empirical transfer function methodology. Journal of Climate 10, 2514–2532.

Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. Journal of Climate 12, 2474–2489.