Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach

Ricardo M. Trigo*, Jean P. Palutikof

Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, United Kingdom

ABSTRACT: Methods to assess the impact of global warming on the temperature regime of a single site are explored with reference to Coimbra in Portugal. The basis of the analysis is information taken from a climate change simulation performed with a state-of-the-art general circulation model (the Hadley Centre model). First, it is shown that the model is unable to reproduce accurately the statistics of daily maximum and minimum temperature at the site. Second, using a re-analysis data set, downscaling models are developed to predict site temperature from large-scale free atmosphere variables derived from the sea level pressure and 500 hPa geopotential height fields. In particular, the relative performances of linear models and non-linear artificial neural networks are compared using a set of rigorous validation techniques. It is shown that even a simple configuration of a 2-layer non-linear neural network significantly improves on the performance of a linear model. Finally, the non-linear neural network model is initialised with general circulation model output to construct scenarios of daily temperature at the present day (1970–79) and for a future decade (2090–99). These scenarios are analysed with special attention to the comparison of the frequencies of heat waves (days with maximum temperature greater than 35°C) and cold spells (days with minimum temperature below 5°C).

KEY WORDS: Downscaling · Artificial neural networks · Climate change scenarios · Portugal

1. INTRODUCTION

The construction of future climate scenarios under conditions of global warming, on a local or regional scale, is necessary for the evaluation of climate change impacts on economic activities such as agriculture or energy production. General circulation models (GCMs) represent the most satisfactory approach to predicting future changes in climate (Wigley et al. 1990, Crane & Hewitson 1998), but their present low spatial resolution (a few hundred km) makes their output problematic to use in impact studies (Karl et al. 1990, Winkler et al. 1997). In recent years, several different methods have been applied to downscale from GCM coarse resolution output to finer spatial scales (e.g. local station or river catchment). These different downscaling techniques can be clustered into 2 conceptually distinct approaches:

(1) Physically based numerical models, i.e. regional dynamical models that work at a finer spatial scale and are nested within a GCM model (Giorgi 1990). Recent applications of this technique have revealed its capacity to reproduce fine-scale features of different regional climates (Jones et al. 1995, Mearns et al. 1995, Kidson & Thompson 1998). Despite significant improvement in computer capacities in recent decades, this physical/numerical model approach remains computationally very demanding (Hewitson & Crane 1996).

(2) Empirical models. This technique is less computationally intensive than that above. However, it requires identification of significant relationships between the observed large-scale atmospheric circulation and local climate, which are subsequently applied to GCM output (Goodess & Palutikof 1998). Multiple regression models have been widely used as an empirical downscaling tool (Wigley et al. 1990, Palutikof et al. 1997). Some other authors have used automated classification procedures to produce downscaled regional scenarios

---

*E-mail: r.trigo@uea.ac.uk

(von Storch et al. 1993, Goodess & Palutikof 1998). Most of the downscaling work done for Portugal has been on the classification of circulation patterns (Zhang et al. 1997, Corte-Real et al. 1998). Circulation patterns have then been applied to GCM output to construct precipitation scenarios, both on monthly (Corte-Real et al. 1995) and daily (Corte-Real et al. 1999, this issue) time scales.

In recent years, a number of papers within the climatological community have adopted artificial neural networks (ANNs) as a tool to downscale from the large-scale atmospheric circulation to local or regional climate variables (Hewitson & Crane 1992, Cavazos 1997). Some applications are for the purpose of constructing climate change scenarios (Hewitson & Crane 1996, Crane & Hewitson 1998). Due to their relevance to this work, a summary of downscaling applications using neural networks is given in Table 1.

Other authors have applied ANN models to forecasting problems at a seasonal scale (Navone & Ceccato 1994, Tang et al. 1994, Tangang et al. 1998) and to short-term (0 to 6 hourly) prediction (Kuligowsky & Barros 1998). Pattern recognition is another area of atmospheric sciences in which the application of ANNs is becoming increasingly important, e.g. cloud classification (Peak & Tang 1992, Pankiewicz 1995) and Atlantic blocking detection (Verdecchia et al. 1996).

With the usual regression analysis, we are constrained to adopt a linear relationship between input and output variables, or at best to use a predefined non-linear function (e.g. logarithmic function). In contrast, ANNs models can be trained to find the best mathematical relationship between the atmospheric circulation and local climate, without predefined restrictions. Thus the method is able to capture some of the non-linear relationships between local climate and the large-scale circulation. In particular, it has been shown that there is always a certain configuration of the multilayer perceptron that can approximate arbitrarily well any continuous function (Hornik et al. 1989,

Hagan et al. 1996). In this sense, it is helpful to regard a multilayer perceptron as a very powerful multiple regression technique (Sarle 1994). Gardner & Dorling (1998) have shown that most atmospheric sciences ANN applications have employed the standard feed-forward configuration of the multilayer perceptron.

Despite the increasingly widespread use of ANNs, most applications do not compare the results obtained from complex ANN models with those from more standard linear techniques such as multiple linear regression. With regard to the few that do, McGinnis (1994) applied a multi-linear stepwise regression and an ANN model to predict snowfall from synoptic circulation. His study showed that 5 d averages of snowfall amounts over the western USA can be better related to the main PCA (principle component analysis) modes of the 700 hPa height field using the ANN model, with a general increase in the explained variance from 44 to 70%. More recently, Weichert & Bürger (1998) compared some important theoretical aspects of both linear and non-linear downscaling techniques. In their study, a linear regression model and a non-linear ANN model were applied to the prediction of daily temperature averages, precipitation and vapour pressure over Potsdam, Germany, and slightly better results were obtained with the non-linear approach (increase of the explained variance between 3% for the precipitation and 10% for the temperature). As shown in Table 1, both these studies do not apply their models to GCM output, thus making a full comparison of their results with those from this paper impossible.

Besides downscaling, other meteorological/climatological applications compare the use of ANNs with other statistical models. Mihalakakou et al. (1998) modelled hourly ambient temperatures using only past values of the predictand. Comparisons of the results between an autoregressive model (ARMA) and a neural network model reveal an important improvement when using the latter approach. Two other applications of ANN models present an explicit comparison

Table 1. Summary of 6 applications that employ ANN (artificial neural network) models to downscale

	Predictand	Comparison with linear models?	Application to GCM data?
Hewitson & Crane (1992)	Daily precipitation in Mexico	NO	NO
McGinnis (1994)	5 d average snowfall in western USA	**YES**	NO
Hewitson & Crane (1996)	Daily precipitation in South Africa	NO	**YES**
Cavazos (1997)	Daily and monthly precipitation in Mexico	NO	NO
Crane & Hewitson (1998)	Daily precipitation in eastern USA	NO	**YES**
Weichert & Bürger (1998)	Daily average temperature precipitation and vapour pressure	**YES**	NO

with linear methods. In both cases there is a significant improvement when using the non-linear technique (Navone & Ceccatto 1994, Gardner & Dorling 1999). Beyond this small number of papers, it is rare to find rigorous comparisons of the performance of linear and non-linear approaches to the same atmospheric sciences problem. Moreover, procedures to rigorously assess model performance, such as cross-validation or computation of skill scores, are not as yet standard.

The main purpose of the present paper is to develop transfer functions to predict local values of minimum and maximum temperatures (hereafter T_{min} and T_{max}) from several large-scale indices. The transfer functions tested here are, on the one hand, linear ANN models, on the other, non-linear ANN models. We test various methods to evaluate quantitatively the improvement (if any) that non-linear ANN models can bring when compared with standard linear models. The method is developed, tested and applied for a site located in central Portugal (Coimbra). Some aspects of the linear models used in this work correspond to a simplified version of the Climatological Projection by Model Statistics (CPMS, Karl et al. 1990) method as developed by Palutikof et al. (1997) and Winkler et al. (1997).

In Section 2 we describe the main characteristics of the observed and model datasets, then in Section 3 we present a brief description of the feed-forward ANN configuration and the training algorithm. The application of several different validation techniques, including skill scores, is described in Section 4. The ANN transfer functions are then applied to GCM output in Section 5. Finally, some conclusions are outlined in Section 6.

2. Datasets and predictor choice

2.1. Observed and model datasets. The following datasets were used in this study:

(1) Daily T_{min} and T_{max} observations for Coimbra in Portugal for the period 1970–93: Coimbra is located in northern Portugal near the Atlantic coast (40.41° N, 7.55° W, altitude 141 m). The site is shown in Fig. 1. The temperature time series were divided into 2 parts. The 1970–79 data were compared with temperature output from the GCM, in order to establish that downscaling is required. The 1979–93 data were used to construct and validate the linear and non-linear downscaling models.

(2) Observed large-scale data from the ECMWF (European Centre for Medium-range Weather Forecasting) re-analyses for 1979–93 and for the area shown in Fig. 1: Six-hourly values of 1000 and 500 hPa geopotential height were extracted. In order to match the coarser spatial and temporal resolution of the GCM data, the 6-hourly data were averaged into a daily time

series, and interpolated onto a 2.5° latitude by 3.75° longitude grid. Sea level pressure (SLP) was calculated from the 1000 hPa geopotential height field using the hydrostatic balance equation (Peixoto & Oort 1992). The resulting time series were used as predictor variables in the development and validation of the downscaling models.

(3) Temperature output from the GCM grid-point closest to Coimbra for 1970–79 and 2090–99: The 1970–79 data (GHGS1) were compared with observations to establish the need for downscaling. The 2090–99 values (GHGS2) were used with the downscaled scenarios for the same period in order to compare differences in the predicted temperature perturbations.

(4) Daily values of SLP and 500 hPa geopotential height from the GCM and for the field shown in Fig. 1: Data for 1970–79 were used to initialise the downscaling models in order to construct 'present-day' scenarios of T_{max} and T_{min} (DGHGS1). Data for 2090–99 were used to construct the future scenarios (DGHGS2).

The GCM used in this study is the Hadley Centre coupled ocean-atmosphere Unified Model (HadCM2, from the experiment described by Johns et al. 1997). This long run is forced between 1861 and 1990 with the historically recorded atmospheric concentrations of greenhouse gases and sulphate aerosols. Between 1991 and 2099 the model is forced with a constant increase of
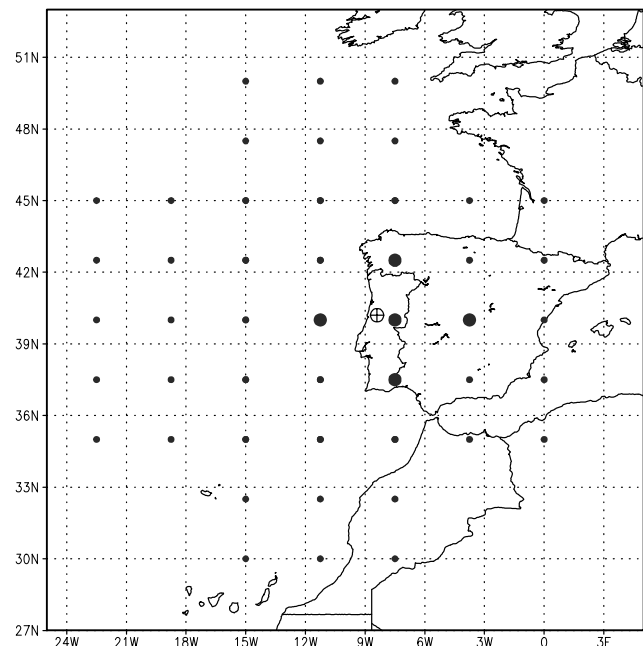

Fig. 1. Grid-points used to compute the geostrophic vorticity of sea level pressure (SLP) (•), the north-south and east-west SLP and 500 hPa height gradients (●) over the NW Atlantic and Iberian Peninsula. (⊕) Coimbra station location
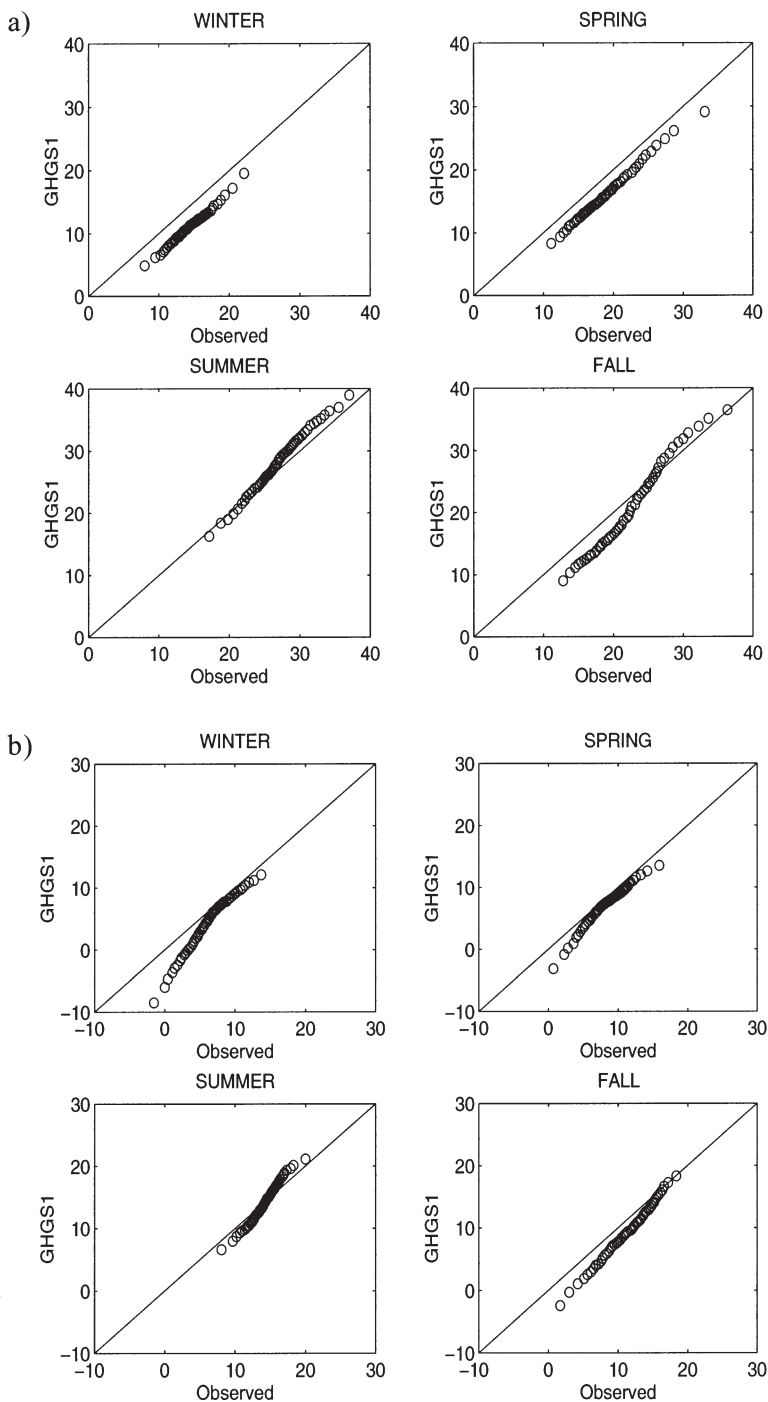
a)

b)

Fig. 2. Quantile-quantile plots of direct GCM output from the nearest grid-point (GHGS1) versus observations in Coimbra (1970–79) for (a) T_{max} and (b) T_{min}

1% yr^{-1} in effective atmospheric carbon dioxide and, in addition, with a representation of the cooling effect of sulphate aerosols. Sea surface temperatures are modelled by the ocean model; they are not prescribed.

2.2. Comparison between observed and HadCM2 T_{min} and T_{max}. The capacity of HadCM2 to successfully reproduce time series of daily T_{max} and T_{min} was evaluated on an annual and seasonal basis. Using GHGS1 (i.e. present day) values from the nearest grid point to Coimbra (40° N, 7.5° W) and observed values for 1970–79, the differences in annual and seasonal means and variances were tested using a 2-tailed *t*-test and an *F*-test respectively. The significance level for rejection of the null hypothesis was set at 5% for both tests. The potentially misleading effects of strong autocorrelation were allowed for by applying the tests to subsamples of the original series. Model and observed time series were sampled once every 10 values to give a sequence less serially correlated (following Palutikof et al. 1997). Results show (Table 2) that the means of both T_{max} and T_{min} from the GHGS1 nearest grid point differ significantly from the observed Coimbra times series in all seasons, except for summer T_{min}. The results for the variance also present significant discrepancies between the seasonal observed and GHGS1 series (the exceptions being T_{max} for winter and T_{max} and T_{min} for spring). On an annual basis, the means and variances of T_{min} and T_{max} also reveal significant differences.

The use of quantile-quantile (qq) plots of the observed and GHGS1 time series helps to identify the range of values that the model is capable (or incapable) of reproducing. These qq plots compare percentiles of the empirical cumulative distribution function of the observations and model data (Wilks 1995). If the match is perfect, the plotted points will fall along the straight line in Fig. 2. In the case of T_{max} (Fig. 2a) the model consistently simulates lower values than observed for both winter and spring. For fall, in the lower range the model data are below the observed, while in the higher range the modelled values are above the observed. Corresponding results for T_{min} are presented in Fig. 2b, where the GHGS1 run fails especially when reproducing the lower range of values for winter, spring and fall. These results show clearly the incapacity of HadCM2 to reproduce accurately station data of T_{min} and T_{max}.

2.3. Predictor choice for downscaling. In order to construct the ANN models, we used observed T_{max} and T_{min} at Coimbra as predictands and several large-scale circulation indices, extracted from the

ECMWF re-analyses, as predictors. Many different circulation indices derived from SLP and 500 hPa height were tested. The 6 indices retained to characterise the large-scale circulation were computed using the grid-points shown in Fig. 1 and are summarised in Table 3.

Besides the circulation indices corresponding to the current day, we used the previous-day indices, giving a total of 12 (6 + 6) predictors describing the state of the atmosphere. To focus attention on the comparison of results between the different model types, all models were constructed on data for the whole year, i.e. we did not use different models for different periods of the year (such as months or seasons). The strong annual cycle, characteristic of most temperature time series, was simulated by introducing 2 additional predictors: the sine and cosine of the Julian day. Thus, the total number of predictors used was 14.

In order to evaluate the best potential predictors we computed simple correlation coefficients between all circulation indices and both predictands. The 500 hPa height from the nearest grid-point is the most important circulation index predictor, explaining 55% of the variance for T_{max} and 38% for T_{min}. The sine variable explains 11% of the T_{max} variance and 16% of the T_{min} variance, whilst the cosine variable explains 53% for T_{max} and 45% for T_{min}. Correlation coefficients of mean SLP from the nearest grid-point and 500 hPa geostrophic vorticity were so weak (R < 0.2) for both T_{max} and T_{min} that we opted to discard these 2 potential predictors.

It is important to mention that temperature at the Coimbra station depends not only on the influence of the circulation indices selected here, but also on conditions at additional levels (e.g. 700 or 850 hPa) and on other variables (e.g. radiation, clouds, humidity). Inclusion of these could improve the general quality of the derived models. Although they can easily be extracted from the ECMWF re-analysis dataset, unfortunately, daily data for these additional relevant variables are not readily available for HadCM2.

Table 2. Means and standard deviations of temperatures at Coimbra, computed from 10 yr of observations and model data (GHGS1) for 1970–79. *Significant differences at the 5% level

	T_{max}		T_{min}	
	Mean	SD	Mean	SD
Annual				
Observations 1970–79	20.72	5.23	9.97	3.51
GHGS1	19.07*	7.21*	8.62*	4.30*
Winter				
Observations 1970–79	14.56	2.81	6.25	3.49
GHGS1	11.21*	2.81	4.15*	4.87*
Spring				
Observations 1970–79	19.05	4.50	8.25	3.18
GHGS1	16.24*	4.51	6.93*	3.51
Summer				
Observations 1970–79	26.99	4.30	14.16	2.32
GHGS1	28.15*	5.18	14.11*	3.38*
Fall				
Observations 1970–79	22.36	5.38	11.23	3.91
GHGS1	20.69*	7.24*	9.28*	4.81*

3. ARTIFICIAL NEURAL NETWORK MODELS

The ANNs used here are based on a feed-forward configuration of the multilayer perceptron (Fig. 3) that has been used by a growing number of authors. The multilayer perceptron is composed of multiple simple processing nodes, or neurons, assembled in several different layers. Each node computes a linear combination of the weighted inputs (including a ***bias*** term) from the links feeding into it. The summed value (***net input***) is transformed using a certain function, either linear or non-linear. The output obtained is then passed as an input to other nodes in the following layer. One important requirement for this type of non-linear function is that it must map any input to a finite output range, usually between 0 and 1 or –1 and 1. It must also be monotonic (Lisboa 1992). Several different functions can be used, namely hard limit, log-sigmoid and the hyperbolic tangent functions. A particular transfer function should be chosen in accordance with the general characteristics of the problem that the ANN is attempting to solve (Hagan et al. 1996). In this work we used a linear transfer function in the only node of the output layer, and the log-sigmoid function in the nodes of the other layers. As an example, the output *y* of a 2-layer network, with ***n*** inputs, ***k*** log-sigmoid nodes in the first layer and one linear node in the output layer is given by:

Table 3. The 6 variables used to characterise the atmospheric large-scale circulation. Values for the same variables, but from the previous day, were also used to account for persistence. The sine and cosine of the respective Julian day were also used

Predictor	500 hPa	SLP
24 h mean (nearest grid-point)	✓	
24 h north-south gradient	✓	✓
24 h east-west gradient	✓	✓
24 h geostrophic vorticity		✓

$$y = \sum_{j}^{k} w_{j}^{(2)} z_{j} + b^{(2)} \tag{1}$$

$$z_j = \frac{1}{1 + \exp\left(\sum_i^n -w_{ij}^{(1)} x_i + b_j^{(1)}\right)} \qquad (2)$$

where x_i corresponds to the ith input, and the coefficients $w^{(2)}$ and $b^{(2)}$ ($w_{ji}^{(2)}$ and $b_j^{(1)}$) are the weights and biases from the second (first) layers. ANNs have to be trained in order to obtain best estimates of the weights that can properly associate predictors with the predictand. There are several different algorithms to train ANNs (Haykin 1994), although they all have in common the capacity to minimise a certain error (cost) function.

The standard training algorithm used in most meteorological applications is the backpropagation algorithm. The full mathematical derivation of this can be found in several neural networks textbooks (e.g. Haykin 1994, Bishop 1995, Hagan et al. 1996). The main steps of the procedure (Fig. 3) are as follows: (1) Initialise network weight value, usually using small numbers obtained from a random number generator. (2) Carry out forward propagation of the first input vector through the whole network. (3) Compute the error by comparing the model output with the target (observations). (4) Backpropagate the error information through the network. (5) Update the weights. (6) Repeat the previous steps for several iterations until some criterion is fulfilled.

Two problems that can arise when using the backpropagation training algorithms are, first, that convergence may be slow and, second, that the final weights may be trapped in local **minima** over the highly complex error surface (Bishop 1995). The specific local minimum where the training process stops is dependent on the adopted initial weight values. Thus slight variations in the initial space phase configuration of the ANN can lead to very different final coefficients. Several recent improvements to the backpropagation algorithm (e.g. variable learning rate and the use of a momentum coefficient) have helped to minimise these problems (Kuligowski & Barros 1998, Tangang et al. 1998). However, an alternative approach has been the introduction of numerically optimised techniques such as the Levenberg-Marquardt method. This is based on an approximation of the Gauss-Newton method. It has the advantage of converging faster and with more reliability than most backpropagation schemes, but on the other hand it requires more memory (Hagan et al. 1996). Here, we trained all non-linear ANN models with the Levenberg-Marquardt scheme. It was found that the cost function decreases very rapidly during calibration, reaching a flat plateau after the first 20 to 50 iterations without further improvement being achievable. Thus, if the cost function did not reach the required threshold, the training process was stopped after a maximum of 100 iterations.

Many different ANN structures are possible, with varying numbers of layers and of nodes per layer. The use of a third layer was tested for this study, but was shown to be redundant, i.e. the same predictive capacity could be obtained with simpler models, using only 2 layers. This is in agreement with other theoretical studies (Bishop 1995). All the non-linear ANN models presented in this work have only 2 layers, 1 input layer with several nodes and 1 output layer with just 1 node. Hereafter, this is referred to as having a $k+1$ structure (e.g. a 5+1 structure implies the use of 5 nodes in the first input layer and 1 node in the output layer).

Even with the relatively simple ANNs proposed here, the risk of overprediction must be borne in mind, because the number of parameters increases dramatically with the number of internal nodes. With the most complex neural network proposed here (the 5 + 1) the number of parameters is estimated at 81 (75 weights and 6 biases). The number of cases used in model development is always at least
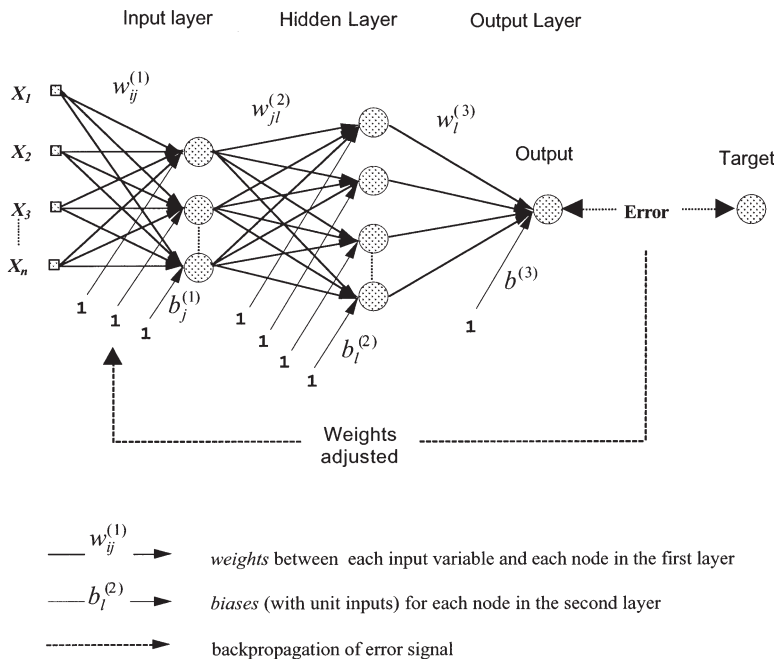

Fig. 3. An example of the structure for a feed-forward artificial neural network model with 3 layers. In this work, the inputs X_1 to X_n correspond to the 14 predictors used to predict both T_{max} and T_{min}

Table 4. Main characteristics for the 3 types of validation used during the testing period

Type of validation	Linear network	Non-linear network
Simple	10 yr calibration (3650 d) 5 yr validation	10 yr calibration (3650 d) 5 yr validation (20 different starting conditions)
Cross	14 yr calibration (5110 d) 1 yr validation with a moving window	14 yr calibration (5110 d) 1 yr validation with a moving window
Bootstrap	Random subsample of 3000 d to calibrate and 5 yr to validate (20 runs with different subsamples)	Random subsample of 3000 d to calibrate and 5 yr to validate (20 runs with different subsamples and different starting conditions)

3000 (see Table 4). However, the data are strongly autocorrelated, thus reducing the number of independent values. The fact remains that the models are demonstrated to work in independent validation, which suggests that overprediction by the models is not a problem here.

In this work we used a perceptron with a linear activation function to construct the linear models. Such linear ANN models produce the same results as a linear regression model (Weisberg 1985, Myers 1986), which here has all 14 predictor variables incorporated. The training algorithm for these linear ANN models is based on the Widrow-Hoff learning-rule, also known as the least mean square (LMS) rule. Unlike the back-propagation algorithm (and its variations) that can be trapped in local *minima*, the Widrow-Hoff rule will give a unique solution corresponding to the absolute minimum value of the error surface.

4. MODEL VALIDATION

As with any other statistical model, ANNs should generally be trained (calibrated) and tested (validated) using 2 independent data sets. Due to the potential instability of ANN models, some authors have recently suggested that validation should be performed on ANN model ensembles, which may be created using bootstrapping techniques (Breiman 1997). This should lead to greater stability of the final results (Tangang et al. 1998).

Here, we applied 3 different procedures to derive data sets for calibration and validation (including bootstrapping). The main characteristics of the 3 techniques are presented in Table 4, and can be summarised as follows:

(1) Simple validation. A single 10 yr period is selected for calibration (here 1979–87) and a single 5 yr period is then used for validation (1988–93). For the linear model, calibration and validation are carried

out only once. For the non-linear ANN, both calibration and validation are carried out 20 times, each time with a different starting condition, resulting in 20 different solutions. The mean of the validation results from the 20 runs is displayed in Table 5.

(2) Cross validation. Of the 15 yr of data available, 14 yr are used for construction and 1 for validation. A moving window is applied, such that the first run is performed using data for 1979–92 for construction and data from 1993 for validation; on the second run data for 1980–93 are used for construction and data for 1979 for model development, and so on. This procedure is applied to both the linear and non-linear cases. Once again, the means of the validation results from the 15 runs are displayed in Table 5.

(3) Bootstrap validation. A random sample of 3000 d is selected from the period 1979–88 and used for calibration. Validation is performed using the last 5 yr of data (1989–93). The sample selection is repeated 20 times in the case of the linear model. In the case of the non-linear model the sample selection is also repeated 20 times, each time with a different starting condition. The means of the validation results from the 20 runs are displayed in Table 5.

One advantage of using varying periods for validation and calibration is that we ensure the models are constructed from, and validated on, representative data. If a particular time period were not representative of the whole series, the resulting model would have very different parameters from a model constructed using representative data.

Having created the data sets of observed and modelled results for validation, it then remains to decide upon the tests to be applied in order to evaluate model performance. As mentioned previously, one of the main objectives of this paper is to show the improvement (if any) that non-linear ANNs can bring when compared to linear models. The performance measures used here are standard measures: the explained variance (i.e. the square of the correlation coefficient between the ob-

Table 5. Values of explained variance (EV) and root mean square error (RMSE) obtained with the linear and non-linear ANN models, using 3 different validation techniques. Values of 2 skill scores (SS_c and SS_p) are also presented for the simple validation technique. The non-linear ANN models are described with the $k+1$ notation, denoting the number of nodes in the 1st and 2nd layers (see Section 3 for explanation)

		Simple validation			Cross validation		Bootstrap validation	
	EV	RMSE	SS_c	SS_p	EV	RMSE	EV	RMSE
T_{max}								
Linear	83	2.73	83	3	83	2.67	84	2.67
1+1	86	2.50	86	18	86	2.46	86	2.50
2+1	90	2.08	90	43	90	2.07	89	2.23
3+1	90	2.03	91	46	91	1.94	90	2.19
4+1	91	1.99	91	49	91	1.92	90	2.15
5+1	91	1.97	91	50	91	1.93	90	2.14
T_{min}								
Linear	78	2.16	78	12	78	2.21	79	2.36
1+1	79	2.13	79	15	78	2.20	79	2.36
2+1	82	1.99	81	25	80	2.10	81	2.16
3+1	83	1.94	83	31	82	1.97	81	2.15
4+1	84	1.86	84	35	82	1.94	82	2.13
5+1	85	1.82	84	36	83	1.90	82	2.12

served and modelled daily data sets), the root mean square error (RMSE), and 2 skill scores (which are more commonly employed in evaluating weather forecasts).

Usually a *skill score* is interpreted as the percentage improvement over a certain reference model (Murphy 1988, Wilks 1995). For a particular measure of accuracy, A (e.g. the mean square error or the mean absolute error), the skill score is defined as:

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}} \times 100\% \qquad (3)$$

where A_{ref} is the accuracy of the reference model and A_{perf} is the accuracy that would be achieved by a perfect model.

The 2 skill scores applied in the present work are based on the mean square error (MSE), and use the climatological mean (MSE_{clim}) and the persistence (MSE_{pers}) as reference models:

$$SS_c(MSE) = \frac{MSE - MSE_{clim}}{0 - MSE_{clim}} \times 100\% \qquad (4)$$

$$SS_p(MSE) = \frac{MSE - MSE_{pers}}{0 - MSE_{pers}} \times 100\% \qquad (5)$$

where persistence is simply the assumption that tomorrow's temperature will be the same as today's. Due to the strong autocorrelation of daily temperature, the SS_p score is probably more important to compute than the SS_c, since persistence is intrinsically of higher quality as a temperature forecast model.

The validation tests presented here are based on the use of models (both linear and non-linear) in which all 14 predictor variables are incorporated. Validation

results for T_{max} and T_{min}, obtained with the linear model and 5 different non-linear models (1+1 up to 5+1), are shown in Table 5. Non-linear ANN results were averaged over the ensemble of models constructed for validation purposes (Table 4). The skill scores (SS_p and SS_c) were computed only for the simple validation technique. (Note that in Table 5 comparisons can be made only between different models, i.e. in the vertical, and not between different validation techniques, i.e. in the horizontal. This is because the different validation techniques are based on data sets of slightly different sizes).

The main conclusions to be drawn from Table 5 are as follows:

(1) Results from the simple validation procedure show that the amount of explained variance increases for T_{max} and T_{min} between the linear and the 5+1 non-linear model, adding an extra 8 and 7% of explained variance respectively. The RMSE values also show a significant and steady decrease between the linear model and the 5+1 non-linear model, this decrease being higher for T_{max} (~0.7°C) than for T_{min} (~0.3°C). These increments are of similar magnitude to the improvements obtained by Weichert & Bürger (1998), although with a very different approach. Similar trends of increasing values of explained variance and decreasing values of RMSE can be observed with the other 2 validation techniques for both T_{max} and T_{min}.

(2) There is a very large increase in SS_p between the linear and the 5+1 model for both T_{max} and T_{min}. The change is from 3 to 50% for T_{max} and from 12% to 36% for T_{min}. It is important to mention that, for both T_{max} and T_{min}, the SS_p score for the linear model is quite low,

showing that such a model is only slightly more useful than persistence.

(3) The SS_c increases, although only moderately, between the linear and the 5+1 non-linear models. This might reflect the fact that a forecast based on climatology is easily outperformed by a linear model, thus not allowing much room for further improvement.

(4) We studied the possibility that the improvement obtained by using non-linear models could be just a simple consequence of modelling the raw values for T_{max} and T_{min}, instead of the anomalies from the monthly means. In fact, even when using anomaly values, the linear model results are always worse than those obtained with non-linear models. For example, when we validated (by simple validation) the anomaly T_{max} models, the explained variance increased from 46% with the linear model to 67% with the 5+1 ANN model. The corresponding RMSE decreased from 2.79 to 2.19.

These results do suggest a positive answer to our main question, i.e. a substantial improvement in all accuracy and skill measures can be observed when we replace a linear model with a non-linear model. This result is independent of the validation technique. However, despite these improvements, 2 important issues remain unanswered: (1) Why should we use 'only' 5 nodes in the first layer and not 10 or even 100 nodes? (2) Non-linear models show, on average, an improvement over linear ones. But, bearing in mind that ensemble techniques have been used in the validation, how large is the dispersion of results around the mean improvement?

These 2 questions are not independent. We see from Table 5 that the rate of result improvement decreases from the 1+1 to the 5+1 model. In addition, although the dispersion of values (not shown) around the mean is virtually zero for the 1+1 model, it grows steadily as we increase the number of nodes. The dispersion of values for the 5+1 model is still sufficiently small when compared to the difference between the linear and 5+1 model mean results that we can regard each individual run as being representative of the 5+1 ANN model.

We conclude that, despite the possibility of using many nodes, the marginal improvement they can bring will be offset by an increasing dispersion of the results. This reflects the fact that a more complex model has to be built with many more weights (coefficients) to be fit. In this work, for example, each additional node in the first layer requires an extra 15 weights (14 inputs and 1 bias) to be calibrated. The use of any additional node **per** layer (or additional layer) requires, always, objective consideration of the balance between the advantages and disadvantages.

Comparison of results for T_{max} and T_{min} derived with the linear model and the 5+1 non-linear model are pre-sented on a seasonal basis in Fig. 4. The temperature values for the percentiles of the empirical cumulative frequency distribution are shown. For T_{max} the differences between the observed and modelled distributions decrease substantially, for all seasons, between the linear and the 5+1 non-linear models. This improvement is particularly important in summer and fall, when the 5+1 model shows a good capacity to reproduce the high and low extremes of the observed distribution. Results for T_{min} obtained with the linear model are better than those obtained for T_{max}; therefore the improvement provided by the T_{min} 5+1 model, although present, is less evident.

5. DOWNSCALED RESULTS USING GCM DATA

The procedures of generating reliable climate change scenarios using empirical downscaling techniques are based on a number of testable assumptions, namely: (1) The GCM is unable to simulate the surface climate variables (such as temperature and rainfall) at the required temporal (e.g. daily) and spatial (e.g. local) scales with the necessary level of accuracy. (2) The local surface climate variables are mainly forced by synoptic-scale features. (3) The GCM is able to simulate these synoptic-scale features successfully, so that the modelled present-day predictor variables are in good agreement with observations. (4) The transfer functions derived with observed data to relate the synoptic field to surface variables remain valid under a likely future warmer atmosphere.

The first and second points were the objects of discussion in Sections 2 and 4, respectively. The third point is briefly analysed here. The last point is impossible to prove; most analysts implicitly assume that the relations between the large-scale circulation and the local surface climate will not suffer any major change under a future climate change scenario (Hewitson & Crane 1996, Palutikof et al. 1997, Winkler et al. 1997). However, temperature depends not only on the large-scale circulation but also on variables such as radiation, cloud type and cover. These relationships are not considered here, but may change in a high CO_2 world, thus affecting the size of temperature perturbations.

The successful application of the previously derived ANN models to the problem of constructing climate change scenarios requires that the predictor variables should be well modelled by the GCM. A comparison between observed (from the ECMWF re-analyses) and GHGS1 means and standard deviations of the 6 circulation indices was carried out on an annual and seasonal basis (not shown). Results were, in general, reasonable for both measures. Thus we opted not to adjust
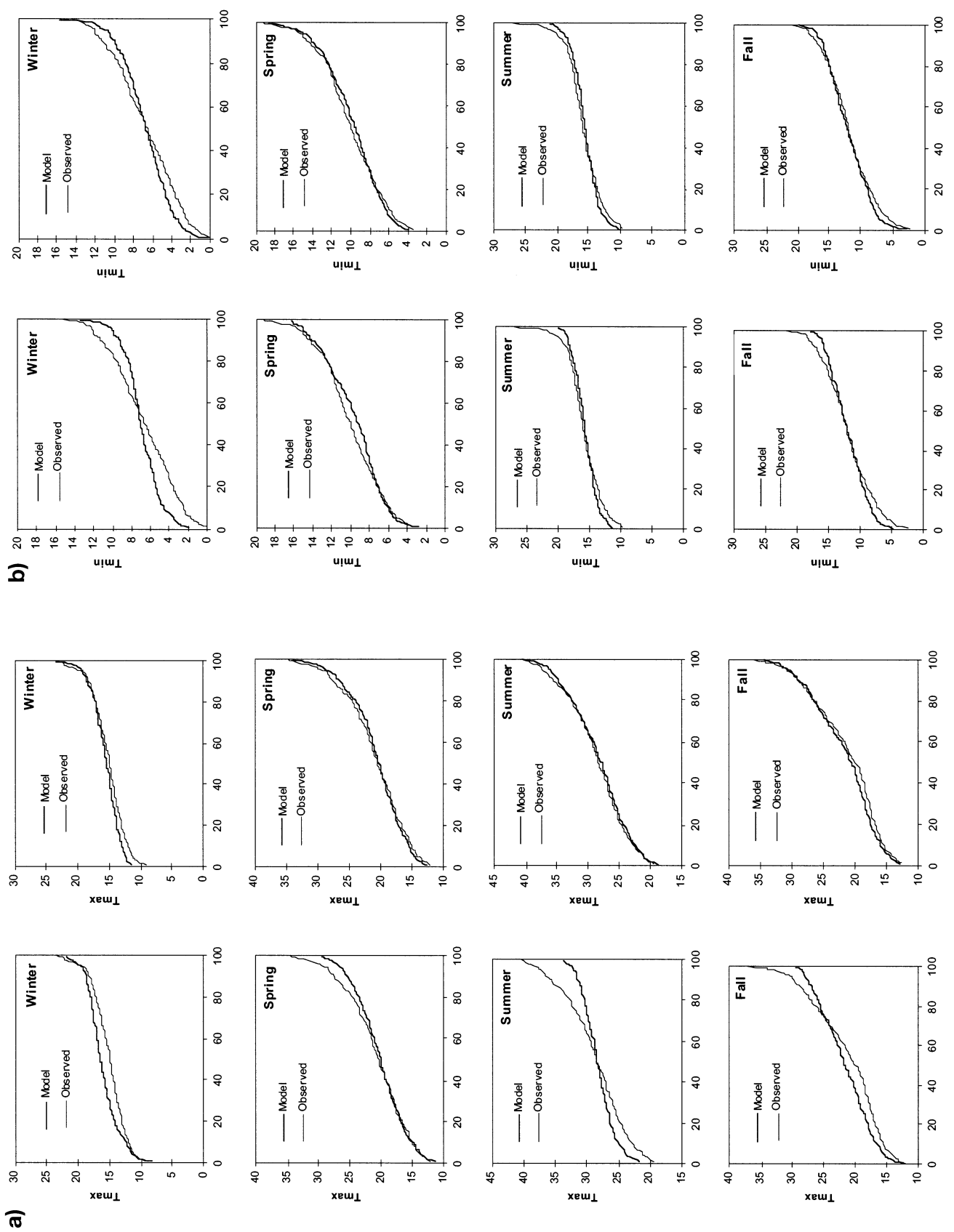
Fig. 4. Comparison between the observed and modelled empirical cumulative frequency distributions for the linear (left column) and 5+1 (right column) models for (a) T_{max} and (b) T_{min}

in any way (e.g. by removing bias) the 500 hPa and SLP fields obtained from the HadCM2 model.

Results in Section 4 give confidence to the use of a 2-layer feed-forward ANN model to downscale from GCM output to local values of T_{max} and T_{min} at Coimbra. Thus, all the downscaled results presented in this section were obtained with 1 non-linear ANN model (5+1, with 5 nodes in the input layer and 1 in the output layer, see Section 3) calibrated and validated with observed data. Once again we have employed qq plots to compare the observed (1970–79) and downscaled (DGHGS1) time series (Fig. 5). In general, both T_{max} and T_{min} values tend to fall near the 1:1 line, showing an improvement when compared to the performance of the direct GCM output (Fig. 2).

Daily means of observed, DGHGS1 and direct GHGS1 output (taken from the nearest gridpoint) T_{max} and T_{min} are shown in Fig. 6. In both cases all time series show a good phase coherence throughout the seasonal cycle. As mentioned before, T_{max} values from the nearest GCM grid-point tend to be below the observed throughout the year, except in summer (Fig. 6a), while downscaled T_{max} values show a much better agreement with the annual evolution of observed temperatures. DGHGS1 results for T_{min} also show an improvement when compared with direct GCM mean daily values, with the exception of spring (Fig. 6b). Annual and seasonal means and standard deviations of downscaled and direct GCM output values (again for the period 1970–79) are compared in Table 6. There are no significant differences between the observed and DGHGS1 T_{max} series (except for the summer standard deviation), revealing a considerable improvement over the comparison of observed and direct GHGS1 output in Table 2. Although the results for T_{min} do not show such a large improvement, there is still a general decrease in the annual and seasonal differences between the observed and DGHGS1 series, especially for the mean.

The application of the same ANN model to the construction of future scenarios (DGHGS2) for T_{max} (Fig. 6c) and T_{min} (Fig. 6d) shows a different annual evolution when compared with the corresponding direct GHGS2 output. As expected, both are generally above the observed. Annual and seasonal statistics compared with present-day values are presented in Table 6. Summer values of DGHGS2 for both T_{max} and T_{min} are much smaller than the equivalent temperatures simulated by the GCM directly, but
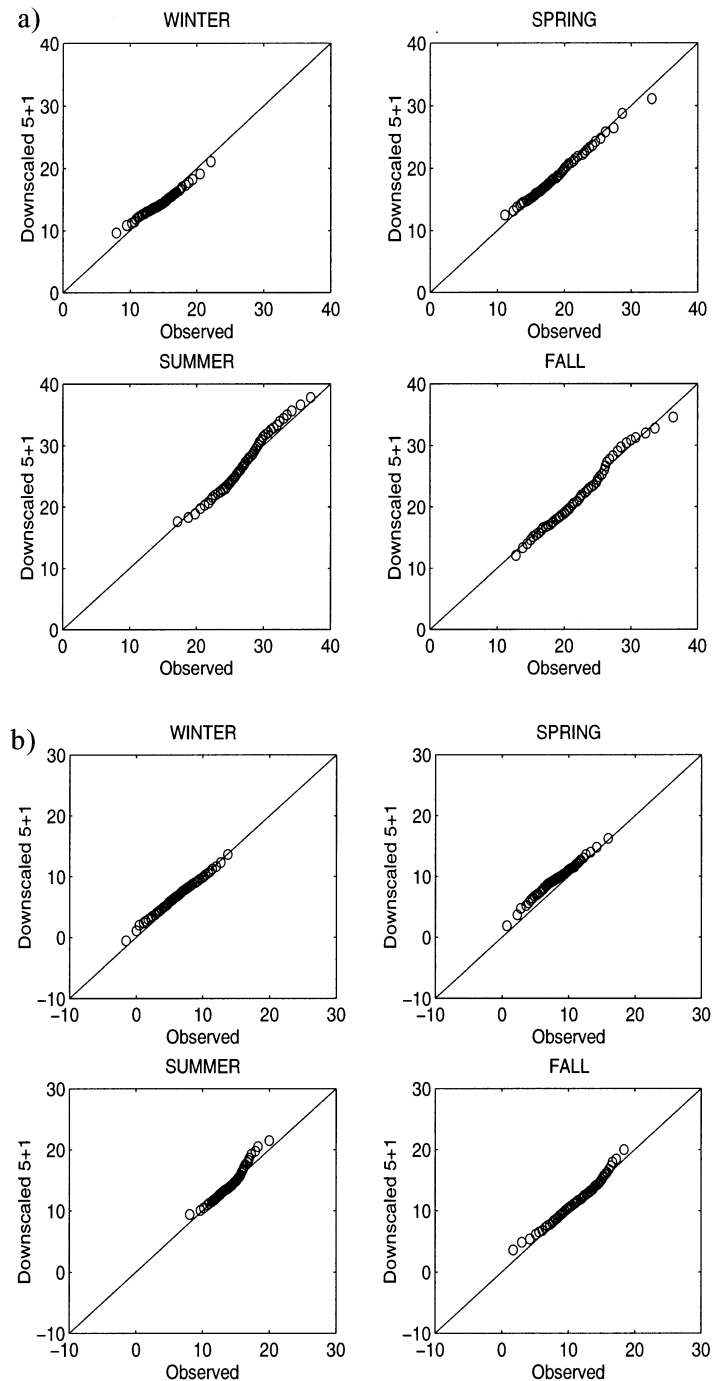

Fig. 5. Quantile-quantile plots of 5+1 downscaled results for present-day scenarios (1970–79) versus observations in Coimbra (1970–79) for (a) T_{max} and (b) T_{min}

during winter and spring DGHGS2 values tend to be higher than the direct GCM output. Thus we can state that the downscaling method gives, for both variables, a smaller range of annual variability than the direct GCM output. Looking at the changes from the present-
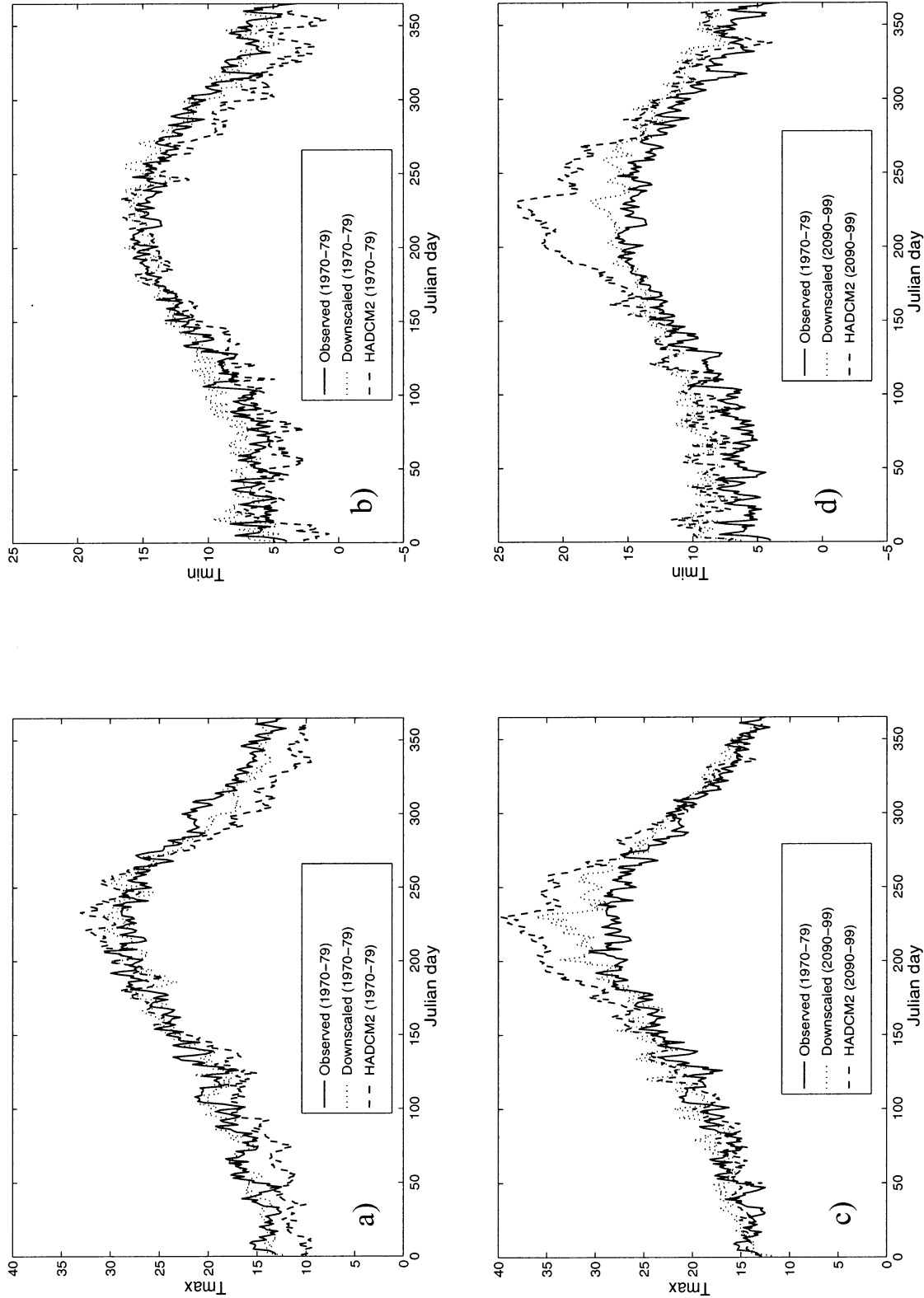
Fig. 6. Ten year daily means of downscaled and direct GCM output temperatures from the nearest grid-point to Coimbra. Present-day scenarios (1970–79) for (a) T_{max} and (b) T_{min}, and future (2090–99) scenarios for (c) T_{max} and (d) T_{min}. Coimbra observations (1970–79) are shown to allow comparison

Table 6. Annual and seasonal statistics of temperature (°C) for observations (1970–79) and for downscaled scenarios and direct GCM output. *Significant differences between DGHGS1 and observations at the 5% level

| | T_{max} | | T_{min} | |
	Mean	SD	Mean	SD
Annual				
Observed, 1970–79	20.72	5.23	9.97	3.51
DGHGS1	20.57	5.34	10.62	3.35
DGHGS1–observed	−0.15	0.11	0.65	−0.16
GHGS1–observed	−1.65	1.97	−1.34	0.79
DGHGS2–DGHGS1	1.79	0.55	1.44	−0.53
GHGS2–GHGS1	4.03	0.68	4.33	0.54
Winter				
Observed, 1970–79	14.56	2.81	6.25	3.49
DGHGS1	14.39	2.18	6.79	3.08*
DGHGS1– observed	−0.17	−0.63	0.54	−0.41
GHGS1–observed	−3.36	0.00	−2.10	1.37
DGHGS2–DGHGS1	1.24	0.47	2.21	−0.11
GHGS2–GHGS1	3.54	−0.05	4.28	−0.67
Spring				
Observed, 1970–79	19.05	4.50	8.25	3.18
DGHGS1	18.98	4.10	9.56*	2.83
DGHGS1– observed	0.07	−0.40	1.30	−0.35
GHGS1–observed	−2.81	0.01	−1.32	0.33
DGHGS2–DGHGS1	1.33	1.21	1.25	0.02
GHGS2–GHGS1	2.94	0.65	3.14	0.06
Summer				
Observed, 1970–79	26.99	4.30	14.16	2.32
DGHGS1	27.00	5.12*	14.49	2.69*
DGHGS1–observed	0.01	0.82	0.33	0.37
GHGS1–observed	1.16	0.89	−0.05	1.05
DGHGS2–DGHGS1	2.01	0.78	0.64	−0.14
GHGS2–GHGS1	4.51	1.43	4.74	1.34
Fall				
Observed, 1970–79	22.36	5.38	11.23	3.91
DGHGS1	21.91	5.63	11.66	3.77
DGHGS1– observed	−0.45	0.25	0.43	−0.14
GHGS1–observed	−1.67	1.87	−1.94	0.91
DGHGS2–DGHGS1	2.57	0.10	1.68	−0.49
GHGS2–GHGS1	5.13	−0.03	5.16	−0.02

day decade (1970–99) to the future (2090–99), we see that the raw model mean changes in temperature (GHGS2 – GHGS1) are always at least double the downscaled mean changes (DGHGS2 – DGHGS1). With respect to the standard deviations, there is no consistent pattern of differences between the raw and downscaled scenarios.

Downscaled results from the 5+1 ANN model show a significant improvement over direct GCM output in their ability to reliably simulate contemporary daily values of T_{max} and T_{min} for Coimbra (DGHGS1 results). Thus, it seems more reasonable to adopt future climate scenarios obtained with this downscaling technique (DGHGS2 results) than to accept direct GCM output (GHGS2 results). However, this statement is based principally on a comparison of the first 2 moments of

the frequency distribution, the mean and the variance. Most sectors of potential end users of climate change impact studies (e.g. agricultural, tourism and energy production companies) would also require reliable scenarios of extremes episodes. Typically, those end users would like to know for example if we can expect more heat waves in summer or fewer cold spells in winter.

We computed the number of extreme situations, defining heat waves (cold spells) as being all episodes of 3, 5 or 7 consecutive days with T_{max} above 35°C (T_{min} below 5°C). Table 7 shows the main results for both the present-day (observed, DGHGS1 and GHGS1) and the future (DGHGS2 and GHGS2). As expected, the results show for both approaches an increase in the number of heat waves and a decrease in the number of cold spells in future. Table 7 also reveals the inability of HadCM2 to simulate with accuracy the present-day frequency of extreme episodes, with both heat waves and cold spells being over-predicted by GHGS1. These results also show that the downscaling technique does not reproduce the observed frequency of extreme episodes with sufficient quality, especially in the case of T_{max} (heat waves). Notwithstanding, for both types of extreme there is a real improvement over the direct model output, especially in the case of T_{min} (cold spells). It is of very real interest (and concern) that the improvement achieved by the downscaling technique in the simulation of the mean and variance of the observed distribution is not fully sustained at the extremes.

6. SUMMARY AND CONCLUSIONS

There is an increasing demand from many sectors of society for reliable climate scenarios on highly resolved temporal (e.g. daily) and spatial (e.g. station)

Table 7. Number of heat waves and cold spells (of 3, 5 or 7 d duration) **per** decade obtained from observations, direct GCM (GHGS1, GHGS2) and downscaled (DGHGS1, DGHGS2) time series

	3 d	5 d	7 d
Heat waves (T > 35°C)			
Observed (1970–79)	4	0	0
DGHGS1 (1970–79)	24	9	1
GHGS1 (1970–79)	54	23	8
DGHGS2 (2090–99)	85	26	8
GHGS2 (2090–99)	374	295	238
Cold spells (T < 5°C)			
Observed (1970–79)	214	106	64
DGHGS1 (1970–79)	153	82	45
GHGS1 (1970–79)	491	333	239
DGHGS2 (2090–99)	33	8	1
GHGS2 (2090–99)	115	51	27

scales. It is a generally accepted argument within the climate research community that, in the long term, direct output from GCMs will fulfil this demand. In the meantime, impact analysts have devoted much of their attention to the development of downscaling techniques. Many different aspects within a downscaling procedure can be the object of improvement, including the number of predictors used, the length of the calibration series, and the use of monthly or seasonal models rather than an all-year model. In the present work, our main efforts have been concentrated on the development of an improved empirical downscaling model for T_{max} and T_{min} at Coimbra, based on the comparison of linear and several non-linear ANN models.

Three different validation schemes have been applied to evaluate the results from ensembles of results from the non-linear ANN model, the ensembles being created by using different initial conditions. The 3 different validation procedures indicate an effective improvement in terms of the explained variance and RMSE when we substitute the 2-layer non-linear ANN model (5+1, with 5 nodes in the input layer and 1 in the output layer) for a linear model. The use of skill scores gives a better assessment of the real improvement achieved. The improvements are particularly impressive for T_{max}. These improvements demonstrate the ability of the non-linear ANNs to capture the relationships between the large-scale circulation and local values of T_{max} and T_{min}.

We applied the non-linear ANN models to GCM present-day output. The downscaled results from the 5+1 ANN model show a considerable improvement over direct GCM output. The application of the same ANN model to the development of a future scenario for T_{max} and T_{min} produces a climatology with a smaller seasonal cycle when compared with direct GCM output. Summer values, for both T_{max} and T_{min}, are much smaller than the equivalent temperatures simulated by the GCM alone. It is difficult to compare the results of this analysis with those of other authors. Indeed, recent papers using ANNs for downscaling purposes have generally failed to compare their ANN model results with those achieved by standard linear regression techniques, as we do here. The papers which do carry out this procedure do not carry out the next step of applying their models to GCM output. Few papers are concerned with downscaling temperature (see Table 1). One study equivalent to ours (by Weichert & Bürger 1998) obtained an improvement of 10% in the explained variance of average temperature by using non-linear models, a result similar to that which we have achieved here.

Finally, we computed the frequency of extreme episodes (heat waves and cold spells) from the downscaled scenarios and from the direct GCM output. When compared with the observations, the direct GCM results are very poor. Although substantial improvement is achieved by the downscaling procedure, it is not sufficient to allow us to accept blindly any future scenarios of extreme situations derived with this approach.

Our main conclusions can be summarised as follows:

(1) The current generation of GCM (HadCM2) is still unable to simulate reliable daily time series of T_{max} and T_{min} for a local site in south-western Europe.

(2) Linear and non-linear transfer function methods can be used to downscale from large-scale circulation indices to local temperatures, but all validation techniques employed show an improvement when using a simple 2-layer feed-forward ANN model.

(3) Downscaled results of T_{max} and T_{min}, obtained with the 5+1 non-linear ANN model, are in better agreement with observations than the corresponding direct GCM outputs.

(4) It seems more reasonable to adopt future climate scenarios obtained with this downscaling technique rather than using direct GCM output.

(5) More work is required to further develop these downscaling techniques for analysis of changes in the frequency of extreme situations such as cold spells and heat waves.

These results do suggest some important limitations to this type of empirical downscaling technique. The use of stochastic models (or weather generators) to simulate extremes episodes (Corte Real et al. 1999) has to be considered as an alternative approach to downscaling. On the other hand, we are confident that the use of additional predictors, especially related to the moisture content of the atmosphere and to radiation, could significantly increase the general quality of these downscaled results, especially in the winter wet season. Such work will be carried out with the next generation of Hadley Centre Model (HadCM3) currently being developed.

LITERATURE CITED

Bishop CM (1995) Neural networks for pattern recognition. Clarendon Press, Oxford

Breiman L (1997) Bagging predictions. Machine Learning 24: 123–140

Cavazos T (1997) Downscaling large-scale circulation to local winter rainfall in north-eastern Mexico. Int J Climatol 17: 1069–1082

Corte-Real J, Zhang X, Wang X (1995) Downscaling GCM information to regional scales: a non-parametric multivariate approach. Clim Dyn 11:413–424

Corte-Real J, Qian B, Xu H (1998) Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation. Int J Climatol 18:619–635

Corte-Real J, Xu H, Qian B (1999) A weather generator for obtaining daily precipitation scenarios based on circulation patterns. Clim Res 13:61–75

Crane RG, Hewitson BC (1998) Doubled CO_2 climate change scenarios for the Susquehanna Basin: precipitation. Int J Climatol 18:65–76

Gardner MW, Dorling SR (1998) Artificial Neural Networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ 32:2627–2636

Gardner MW, Dorling SR (1999) Neural Network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air London. Atmos Environ 33:709–719

Giorgi F (1990) Simulation of regional climate using a limited area model in a general circulation model. J Clim 3: 941–963

Goodess CM, Palutikof JP (1998) Development of daily rainfall scenarios for Southeast Spain using a circulation type approach to downscaling. Int J Climatol 18:1051–1081

Hagan MT, Demuth HB, Beale MH (1996) Neural Networks design. PWS Publishing Company, Boston

Haykin S (1994) Neural Networks: a comprehensive foundation. Macmillan College Publishing Company, New York

Hewitson BC, Crane RG (1992) Large-scale atmospheric controls on local precipitation in tropical Mexico. Geophys Res Lett 19:1835–1838

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–95

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

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre coupled ocean-atmosphere GCM: model description, spinup and validation. Clim Dyn 13:103–134

Jones RG, Murphy JM, Noguer M (1995) Simulation of climate change over Europe using a regional climate model. Part 1: Assessment of control climate, including sensitivity to location of lateral boundaries. Q J R Meteorol Soc 121: 1413–1449

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to the observed local climate. Part 1: Seasonal statistics. J Clim 3:1053–1079

Kidson JW, Thompson CS (1998) A comparison of statistical and model-based downscaling techniques for estimating local climate variations. J Clim 11:735–753

Kuligowski RJ, Barros A (1998) Experiments in short-term precipitation forecasting using Artificial Neural Networks. Mon Weather Rev 126:470–482

Lisboa PGJ (1992) Neural Networks: current applications. Chapman & Hall, New York

McGinnis DL (1994) Predicting snowfall from synoptic circulation: a comparison of linear regression and neural network. In: Hewitson B, Crane R (eds) Neural Nets: applications in geography. Kluwer Academic Publishers, Dordrecht, p 79–99

Mearns LO, Giorgi F, McDaniel L, Shields C (1995) Analysis of variability and diurnal range of daily temperature in a nested regional climate model: comparison with observations and doubled CO_2 results. Clim Dyn 11:193–209

Mihalakakou G, Santamouris M, Asimakopoulos D (1998) Modeling time series using neural networks. J Geophys Res 95:1943–1953

Murphy AH (1988) Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon Weather Rev 116:2417–2424

Myers RH (1986) Classical and modern regression with applications. Duxbury Press, Boston

Navone HD, Ceccatto HA (1994) Predicting Indian monsoon rainfall: a neural network approach. Clim Dyn 10: 305–312

Palutikof JP, Winkler JA, Goodess CM, Andresen JA (1997) The simulation of daily temperature time series from GCM output. Part I: Comparison of model data with observations. J Clim 10:2497–2513

Pankiewicz GS (1995) Pattern recognition techniques for the identification of cloud and cloud systems. Meteorol Appl 2:257–271

Peak JE, Tang PM (1992) Towards automated interpolation of satellite imagery for navy shipboard applications. Bull Am Meteorol Soc 73:955–1008

Peixoto JP, Oort AH (1992) Physics of climate. American Institute of Physics, New York

Sarle WS (1994) Neural Networks and statistical models. Proceedings of the 19th Annual SAS Users Group International Conference. SAS Institute, Cary, NC

Tang B, Flato GM, Greg H (1994) A study of artic ice and sea-level pressure using POP and Neural Network models. Atmos Ocean 32:507–529

Tangang FT, Tang B, Monaham AH, Hsieh WW (1998) Forecasting ENSO events: a Neural Network-Extended EOF approach. J Clim 11:29–41

Verdecchia M, Visconti G, D'Andrea F, Tibaldi S (1996) A neural network approach for blocking recognition. Geophy Res Lett 23:2081–2084

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

Weichert A, Bürger G (1998) Linear versus nonlinear techniques in downscaling. Clim Res 10:83–93

Weisberg S (1985) Applied linear regression. John Wiley & Sons, New York

Wigley TML, Jones PD, Briffa KR, Smith G (1990) Obtaining sub-grid-scale information from coarse-resolution General Circulation Model output. J Geophys Res 95:1943–1953

Wilks DS (1995) Statistical methods in the atmospheric sciences: an introduction. Academic Press, San Diego

Winkler JA, Palutikof JP, Andresen JA, Goodess CM (1997) The simulation of daily temperature time series from GCM output. Part II: Sensitivity analysis of an empirical transfer function methodology. J Clim 10:2514–2532

Zhang X, Wang XL, Corte-Real J (1997) On the relationships between daily circulation patterns and precipitation in Portugal. J Geophys Res 102:13495–13507