4 Modelling climate extremes

4.1 Introduction

In addition to the 'pattern' of Mediterranean climate (Chapters 2 and 3), one of the major themes of this study is the cause of extreme climate conditions across the Mediterranean. As discussed at the end of Chapter 3, correlation studies alone are not enough to prove a causal mechanism for any particular form of extreme climate. To that end, a further two statistical downscaling techniques have been employed, enhancing the weight of evidence for links (or the lack of such links) between hemispheric-scale atmospheric circulation and regional (station-scale) extreme climate across the Mediterranean. Although it is very difficult to statistically verify a causal link between one factor and another, the ability to replicate predictands from a set of predictors is one quality that causality implies (Wilks, 1995).

Climate analysis through statistical downscaling is also useful because it allows for an assessment of the amount by which one factor may contribute to another, inherent in the distribution of coefficients or weights. Both of the statistical modelling methods in use throughout this study can be used to directly downscale from the former large-scale circulation to the latter regional scale and have been selected to provide two very different approaches to the statistical downscaling problem. Having ensured the quality of the available data (Section 3.2.1), discussed its properties (Section 3.3), and made decisions about the effective use of several predictors and predictands in Chapter 3.7, this Chapter can go on to discuss the statistical downscalling process. Section 4.2 provides information regarding the method and structure of both statistical models, and provides an introduction to the applicable theories and equations used (rather than a deeply technical discussion of linear vs. non linear modelling or neural networks in general). Section 4.3 assesses the performance of each model, and Section 4.4 provides a discussion of results, including recommendations for future use concerning both methods.

4.2 Downscaling model structure

The methods used in this study are both multivariate direct downscaling techniques, and have a number of common requirements. These are discussed below, before this chapter continues to describe the two model methodologies. In all cases, the downscaling models are run for the target period (1958-2000) for each season separately, so as to avoid the largely solar-dependent effect of a seasonal cycle. The predictors mentioned below take the form of the seasonally dependent indices of circulation and principal components (Chapter 3.3.2, 3.3.3). Predictands are then the indices of climate extremes, discussed in the previous chapter (3.3.1).

Standardized anomalies

Forcing agents that act on very long time scales may alter long-term patterns evident in temperature or precipitation, or dynamically limit the range of values at a particular station (Chapter 2). As seen in the previous chapters, altitude, latitude, and proximity to coastlines may all change the mean climatic response at any given location (Chapters 2 and 3). Having already described these features, it is evident that without explicitly (mathematically) stating model parameters that account for geographic variation, we may unduly attempt to fit consistent but spatially-variable behaviour to the (physically less connected) predictors provided. Assimilating physical data required for a comprehensive analysis is highly time and resource intensive, and in many cases the appropriate data are simply not available over the region or period required (e.g. Mediterranean land use data). In these instances, and where such geographical factors remain largely constant over the period under consideration (i.e. the last 40 years), it can be wise to reduce the station data to a more easily spatially comparable form that focuses on variations or anomalies between the mean and the range. In this fashion, variations in behaviour from station to station can be more easily assessed in an objective fashion. In climatology, station data are frequently 'normalised' using standardized anomalies (Wilks, 1995). These subtract the mean of the data and then divide by the appropriate standard deviation. As this study is interested in relating variations of hemispheric predictors to regional stations in a comparable fashion, both predictors and predictands have been normalised.

Calibration and verification

When constructing a statistical model to relate one set of factors to another, the level of generalisation of the model must be taken into account. If the model has been too tightly constrained to the data used for construction then it will be incapable of providing useful results when applied to other data of the same type. If the data are too generalised then the statistical model will not provide useful results even for the original data set. A solution between these cases of over- and under-fitting must be found. To test for this condition a subset of data (in this case a sub-period of the available 40 years) are reserved for model testing. Models are constructed, or 'calibrated', based upon the data set without the reserved set and then tested, or 'verified', only upon the reserved (verification) data. In this fashion the skill of the model is assessed based upon independent data not included in its construction. If the model is either under-fitted to the generalised relationship, or over-fitted to a specific data set, then skill will suffer.

A particularly robust form of the calibration and verification approach is to 'cross-validate' relationships derived from statistical models (Efron, 1983; Michaelson, 1987). This method involves repeated calibration and verification tests, reserving data as part of a moving window. This window may be any size, and may overlap from iteration to iteration (although the calibration and verification windows remain distinct), but more commonly it is held as one observation with all other (n-1) observations left in the calibration period (the 'leave-one-out' method). This procedure allows for effective verification of small data sets, such as those used for this study. In this study all models are tested using these approaches.

4.2.1 Orthogonal Spatial Regression (OSR)

Orthogonal Spatial Regression (OSR) is a technique adapted from surface pressure reconstruction (Jones *et al.*, 1987), dendroclimatology (Briffa *et al.*, 1986; Briffa *et al.*, 1983), and climate forecasting (Hayden and Smith, 1982). It relies on the construction of a correlation or covariance matrix (see Eqn. 3.18) and its subsequent manipulation. In terms of the reconstruction of tree-ring data, the levels of skill provided by OSR are analogous to those provided by canonical regression (Cook *et al.*,

1994). In OSR each variable to be included in the downscaling model, predictor (input) or predictand (output), is normalised (see above), and then subjected to principal component analysis (see Chapter 3.3), a method that reduces the dimensionality of a dataset (Wilks, 1995) and also avoids some of the issues that arise from strongly inter-correlated predictors (Briffa *et al.*, 1983). The principal components found by this method are retained while the cumulative product of the relevant eigenvalues are greater than 1. This leads to the rejection of roughly 1/3 of the available components, with the remainder explaining around 95% of total variance. This automated discrimination procedure, an alternative to that utilised for the previous PCA study (Chapter 3.3), is called the PVP criterion (Guiot, 1985) and allows for a more simple discrimination than 'scree' plots, although a greater number of components are retained. The resultant 'candidate predictors' (c) and basic predictands (p) are then regressed onto each other to form a principal component (PC) regression (Eqn. 4.1).

$$\beta_{cj} = \sum_{p=1}^{P^*} b^*_{cp} u_{pj}, \quad c = 1, ..., C^*; j = 1, ..., Jyears \tag{4.1}$$

Where b^*_{cp} represents the regression coefficient between the candidate predictor components (β_{cj}) and the predictand components (u_{pj}) for each predictand (p) over time (j). An asterix is used in the above to denote a set that has had some members reduced (using the PVP criterion above). Due to the orthogonality (see Chapter 3.3) of the candidate predictor set there is no need to employ a stepwise regression technique (Wilks, 1995), as commonly included in multivariate climate modelling (Jones *et al.*, 1987). However, based upon their significance in the regression, some further components are rejected. Where the standard t-test value for any given component is less than 1, that component displays less than marginal significance and is removed from the OSR model (Briffa *et al.*, 1983).

Equation 4.1 can be rewritten to more easily show the relationship between predictors (y) and predictands (x) if the appropriate equations for the principal component analysis are substituted, and the resulting additional terms are absorbed into the coefficient d:

$$\hat{y}_{ij} = \sum_{n=1}^{N} d_{in} x_{nj}, \quad i = 1, ..., I; j = 1, ..., J \tag{4.2}$$

d can be enumerated as the product of the eigenvectors (e_{nm}) from the predictor (β_{cj}) and predictand (u_{pj}) PCAs (Eqn. 3.21) and the regression coefficient b (Eqn. 4.1), that exists between them.

OSR summary

The regression model that exists as a result of equations 4.1 and 4.2 has a number of distinct intrinsic qualities:

- The OSR is strictly linear and cannot compensate for nonlinear processes.
- If using the covariance matrix (as this study does), each regression is variance based, rather than centred around the means of a data set (a product of using the correlation matrix). This approach is statistically more appropriate for studying the tails (or extremes) of a data set and provides information additional to that gained from a correlation study (Section 3.6)
- Separate OSR runs are required for each season at least, and may be more appropriate for each month (Jones *et al.*, 1987). However, as each of the STARDEX indices (Section 3.3) are calculated seasonally, this temporal resolution must be used here.
- As a function of the eigenvector core of this method, the correlation between components (i.e. collinearity between candidate predictors) is minimised. No additional pre-selection criterion is required (Jones *et al.*, 1987).

In addition to the dendro-climatological applications for which OSR was originally conceived, the method has also been recently (and successfully) applied to the reconstruction of gridded monthly pressure data from historical station records since 1780. In this application, results are better when the climate is more coherent (i.e. in winter compared to summer) (Jones *et al.*, 1999b). Dependence of performance upon the spatial coherence of climate is common to statistical downscaling methods (Haylock *et al.*, 2006). Recent studies have attempted to reconstruct NAO and SOI indices from

wind-field data, and found a similar division in skill between summer and winter seasons (Jones and Salmon, 2005). This is the first study that applies the OSR technique as a form of statistical downscaling, rather than reconstruction.

4.2.2 Artificial Neural Networking (ANN)

Artificial neural networks (ANN) first appeared in 1943 (McCulloch and Pitts) as an attempt to theorise computational systems that would mimic the structure of the human brain. The resulting networks form relational maps of artificial 'neurons', providing convolutions of linear (and more recently, non-linear) functions between input and output (Harpham, 2005) that are free from *a priori* assumptions regarding the data employed, and produce a particularly efficient method of pattern classification and function approximation (Haykin, 1999). The latter feature has led to ANN deployment in a number of meteorological and climatological studies (Navone and Ceccato, 1994; Tangang *et al*., 1998; Silverman and Dracup, 2000) including dendroclimatology work (Woodhouse, 1999), and statistical downscaling studies such as the one conducted here (Hewitson and Crane, 1992; McGinnis, 1994; Hewitson and Crane, 1996; Cavazos, 1997; Cavazos, 1999; Harpham and Wilby, 2005). Although Trigo (2000b) have found that an ANN approach to statistical downscaling for Portugal compares favourably with a simple linear approach, few studies provide a direct comparison between ANN performance and that of linear regression methodology over a specific target region (Trigo, 2000b) as done in this chapter. Haylock *et al*. (2006) compare downscaling methods over the United Kingdom, finding that neural networks represented the inter-annual variability of climate better than other statistical methods (i.e. Cannonical correlation analysis, multiple linear regression), and dynamical models including HadRM3, but that internal biases lead to a systematic underestimation of extreme climate. A number of additional recommendations for statistical downscaling the climate of sub-regions of the Mediterranean and other European sub-regions can be found on the STARDEX (2006) website, including work that shows ANN methods generally performing comparatively well, particularly in terms of Iberian rainfall indices (Goodess *et al*., 2006).

Neural networks are constructed as collections of 'artificial neurons', themselves composed of 'synapses' (connecting links, each of which possesses its own strength, or 'weight'), 'junctions' (that sum combinations of inputs provided by synapses), external bias (to affect the net output of a junction), and 'activation functions' (that condition junction output amplitude to within predefined limits), as shown in Figure 4.1.

In mathematical terms the output of a single neuron, after activation, can be written:

$$y_k = \phi(u_k + b_k) \tag{4.3a}$$

Where:

$$u_k = \sum_{j=1}^{n} w_{kj} x j \tag{4.3b}$$

Here, j represents the number of inputs (x), affected by synaptic weights w_{k1}, w_{k2},…w_{kn} leading to neuron k. The output of the summing junction (u_k) and the bias (b_k) applied to neuron k, give the 'induced local field' ($u+b$). The final output of the neuron k (y_k) is given by the induced local field term, multiplied by a limiting ('activation') function (ϕ).

Activation functions (ϕ) can be step threshold functions, sloped linear functions, or sigmoid functions (among others), and are generally configured to give output (y) between 0 and 1 or –1 and 1 (Haykin, 1999). As can be seen from Equation 4.3, the ANN is essentially a more flexible variant of a regression system. Figure 4.1 represents a multi-layer perceptron, containing multiple nodes, and in this case Equation 4.3 can be rewritten to give the output at y_1 (Harpham, 2005).

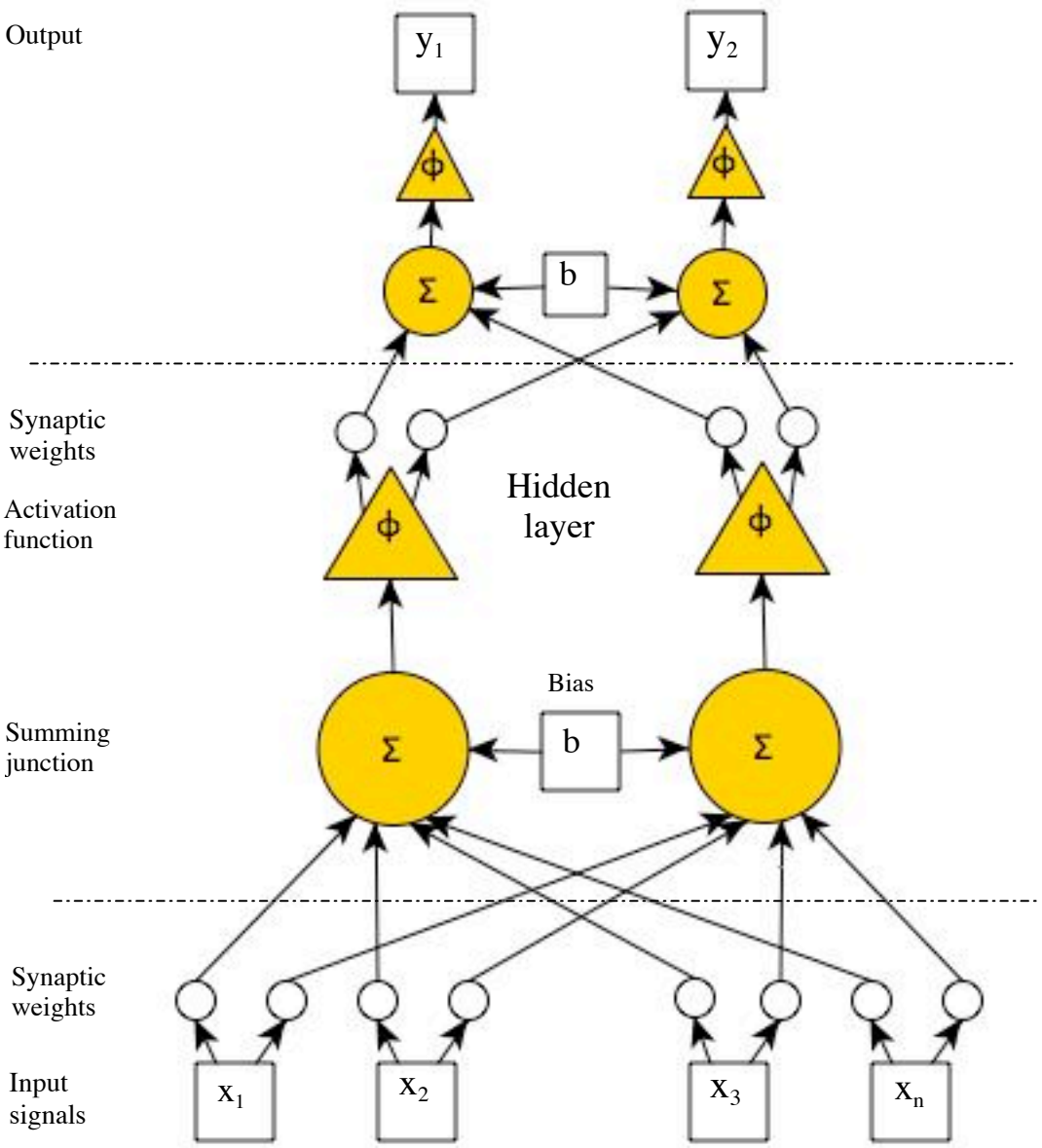

Figure 4.1: A Multi-Layer Perceptron. Interlocking neurons in a multi-layer neural network (Haykin, 1999)

$$y_k = \phi^{(2)}\left(\sum_{j=0}^{M} w_{kj}^{(2)} z_j \phi^{(1)}\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i\right)\right) \qquad k = 1...c \tag{4.4}$$

Each term is described above, with the exception of z_j. The z term represents the output from each node in the hidden layer of Figure 4.1 (given by y_k in Equation 4.3a).

Neural networks seek appropriate weighting values between input and output. This process is iterative, and continues until the network minimises the error (or cost function) between output signal and expected values over the training period, having gained an optimum set of weights through a training algorithm, such as error back-propogation (Rumelhart and McClelland, 1986). Just as with other forms of numerical modelling, networks may be under- (too few nodes, too few hidden layers) or over- (too many nodes, too many hidden layers)- fitted, and there are approaches to both problems (see below). An absolute limit to the number of nodes exists, however, that is equal to the number of predictors entering the network. A summary of these issues is given by Harpham (2005).

Radial Basis Function Neural Networks

The majority of neural networks in active use (Gardner and Dorling, 1998; Harpham, 2005) are of the Multi-layer Perceptron (MLP) type (described above). This study uses a similar one-hidden-layer ANN known as a 'Radial Basis Function' (RBF) neural network (Powell, 1987; Light, 1992). This method overcomes several of the drawbacks inherent in simple MLP models (Harpham, 2005), such as the difficulty in assigning the appropriate number of layers or nodes: there is only one hidden layer, and the number of nodes are constrained by the number of inputs to the RBF model. The paradigm associated with the RBF neural network is also more appropriate for mapping input onto output in a multi-dimensional space, as illustrated below.

The RBF modification changes the overall approach of the network from a stochastic model to a high dimensionality curve-fitting approximation: essentially a 'best-fit' methodology. From Cover (1965), we can assume that any set of binary (i.e. dichotomous) data in a two-dimensional space will possess a plane that separates the dichotomy into homogenous groups. When used for pattern classification (a problem that machine learning is commonly applied to) RBF networks seek to find the appropriate plane, and the higher the dimensionality of the network, the more likely it is to achieve success (i.e. a close approximation). This illustrates the capability of RBF networks to use a nonlinear transformation to map a nonlinear problem into a linear one through high-dimensional 'hidden space'. Rather than the separation problem, the current issue of statistical downscaling is more similar to filtering/interpolation, where a

multidimensional plot of output as a function of input possesses a representative plane (Haykin, 1999).

RBF networks interpolate surfaces (of the required dimensionality) by inserting a conditioning (linear or non-linear) arbitrary 'basis' function (Eqn. 4.5) into the junctions of the hidden layer shown in Figure 4.1, while the second layer shown retains the structure detailed in equations 4.3 above. If the network has a single output the hidden layer nodes then become the function *F(x)*:

$$F(x) = \sum_{i=1}^{N} w_i \phi(||x - x_i||) \quad \textbf{(4.5)}$$

Where ϕ represents N arbitrary radial basis functions with known data points at their centres (Powell, 1988), and $||x\text{-}x_i||$ the Euclidian distance between the input vectors and those functions. For a desired response (*d*), such that $F(x_i)=d_i$ the problem can be restructured in terms of matrices, so that the desired response vector (*x)* becomes the product of the linear weight vector and the interpolation (N-by-N) matrix:

$$\Phi = \phi_{ij} | (j, i) = 1, 2, ...N \quad \textbf{(4.6)}$$

This then gives:

$$\Phi w = x \quad \textbf{(4.7)}$$

And if ϕ is a non-singular matrix the weightings of the network can then be solved:

$$w = \Phi^{-1} x \quad \textbf{(4.8)}$$

As long as ϕ is non-singular and the data points distinct, this equation is valid for multiple outputs in the generalised form, gained by linear superposition of the basis functions used above. However, there is a further modification (Eqn. 4.9) of this equation (Moody and Darken, 1989; Broomhead and Lowe, 1988) that limits under-fitting and enhances the speed of computation. In this scheme, the number of basis functions required decreases (to many less than the number of input vectors), basis function centres are freed from initial constraint and are determined by training, and

biases (Figure 4.1) are introduced that modulate the average basis function values towards the average target values.

$$y_k(\mathbf{x}) = \sum_{j=1}^{M} w_{kj}\phi_j(\mathbf{x}) + w_{k0} \qquad \textbf{(4.9)}$$

Where x represents elements x_i in the d-dimensional input vector (Harpham, 2006).

Michelli's theorem (1986) informs us that there are numerous functions that can be used to produce non-singular matrices in this context, detailed in Bishop (1995), and below:

- Linear: $\phi(x) = x$
- Multiquadratic: $\phi(x) = (x^2 + \sigma^2)^{1/2}$ for $\sigma > 0$ and $x \in \Re$
- Inverse Multiquadratic: $\phi(x) = \frac{1}{(x^2+\sigma^2)^{1/2}}$
- Cubic: $\phi(x) = x^3$
- Gaussian: $\phi(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$ for $\sigma > 0$
- Thin Plate-Spline: $\phi(x) = x^2 \ln(x)$

With terms for constants (c) and a 'basis width' (ϕ), that determines the smoothness of the interpolation function. The appropriate use of basis function is explored further in Section 4.2.3 and 4.4.1.

Training RBF Neural Networks

When 'training' RBF networks to the appropriate architecture, there are two variables that must be optimised, the centres of the basis functions (unsupervised) discussed above, and the synaptic weights of the model (supervised). All other values are either selected initially (such as the basis width variable), or are reflexive (with iteration). The centres of basis functions are determined with the use of an unsupervised k-means clustering approach. This hybrid training method has advantages over entirely

supervised methods, such as error back propogation (Harpham, 2005). The k-means method maintains a high training speed, is not as vulnerable to local minima in error surfaces as error back propagation, and also aids the interpretability of the basis centres. The disadvantage of the k-means method is that the cluster configuration is constrained by the order in which inputs are 'shown' to the RBF model, in turn dependent on initial seed values. Given sufficient time / processing power, a large number of randomised seeds, or a heriarchical clustering method, could be used to overcome this problem and find optimal clustering. Lacking the appropriate resources, comparisons between RBF models are best served by a consistent order of input selection.

The K-means clustering algorithm used for this study divides input vectors into k-groups with maximised homogeneity, the centres of the resulting input clusters are then used to locate the basis function centres in the regions of input space where there are significant volumes of data. Determining both the number of clusters and the centre of each cluster requires iterative experimentation (Anderberg, 1973). Inititally there are *k* clusters, where *k* is the number of input vectors, and each vector determines the centre of one cluster. Each vector is then reassigned to the cluster with the nearest centroid, after each change in the set of inputs the centroids are updated. The reassignment process is repeated for every iteration until the structure of each cluster remains unchanged after one whole cycle.

Here, optimal weights are constructed using singular value decomposition (SVD) to solve for the weights matrix disussed above. SVD (Golub and Van Loan, 1996) represents a particularly robust method used with many other forms of statistical climate modelling (Wilks, 1995). If the matrix (Φ) is a real *NxM* matrix then it possesses orthogonal matrices:

$$U = [u_1, u_2, ..., u_N]$$

and

$$V = [v_1, v_2, ..., v_M] \quad \textbf{(4.10)}$$

such that

$$U^T \Phi V = diag(\sigma_1, \sigma_2, ..., \sigma_k), \quad K = \min(M, N)$$

where

$$\sigma_1 \geq \sigma_2 \geq, ..., \geq \sigma_k > 0$$

The σ values are termed the singular values of the matrix Φ. The *MxN* pseudo-inverse of this matrix is then:

$$\Phi^+ = V\Sigma^+ U^T \tag{4.11}$$

Φ^+ is an *NxN* matrix that can be expressed as singular values of Φ such that:

$$\Sigma^+ = diag(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, ..., \frac{1}{\sigma_K}, 0, ..., 0) \tag{4.12}$$

For a radial basis function neural network:

$$W^T = \Phi^+ T \tag{4.13}$$

Where *W* is the linear weights matrix, *T* is the target output vector, and Φ is the interpolation matrix.

RBF summary

Given the above, Radial Basis Function Neural Networks are distinct from other forms of climate modelling in that:

- ANNs make no *a priori* judgement on the data employed (Haykin, 1999)
- Due to the neuronal weighting system, ANNs do not suffer overly from the use of 'too many' irrelevant or redundant predictors, as they are weighted to zero. Conversely, the more predictors that are used, the higher the dimensional limit of the network, and, as illustrated by Cover (1965), the more likely it is to find appropriate surfaces. Networks are often utilised with hundreds of predictors, rather than tens, although there exists no

specific rule for a network's ideal size (Bishop, 1995; Haykin, 1999; Sarle, 2000).

- Neural networks can suffer from over-fitting. However, the radial basis function neural network limits this problem, and the modification supplied by Moody and Darken (1989) limits it further. Sarle (2000) states that a number of training cases 30 times the number of weights in the network reduces the likelihood of over-fitting, but in climatology such volumes of high-quality data may not be available. In these cases, robust forms of validation are essential.
- RBFs offer a range of solutions to non-linear problems. An example of this is the 'exclusive or' problem, where a network produces a desired response if one predictor or another possesses a given value, but not if both possess that value (Haykin, 1999). RBFs are capable of solving this problem, although it cannot be solved by any linear perceptron (Haykin, 1999).
- If appropriate algorithms are used, RBF neural networks need only optimise values associated with the hidden layer, and do not need to produce a time-consuming full non-linear 'best-fit'.

4.2.3 Model sensitivities

Given the above types of models (OSR and ANN) it then becomes necessary to determine appropriate starting conditions and whether or not multiple variants (i.e. distinct forms of model application with different model structure, or predictor sets, etc.) are required. A number of factors may act to influence the ability of each approach to skilfully reproduce climate predictands. In order to identify the most appropriate configuration(s), a sensitivity test is therefore required. Often, the choice of model starting conditions is subjective, a more objective approach is utilised here. In this test each statistical downscaling model parameter is altered over a range larger than that normally used by the statistical downscaling model to see how end results are affected. In all instances the quantity tested for is model skill. Model skill measures the ability of any given model to replicate observed data, and here is measured by the correlation coefficient (r, Section 3.3.4) between observed and modelled series (Wilks, 1995). The following factors can be altered to give variations upon the models detailed above, and may affect model skill:

- In the case of OSR, the specific combinations of predictors (up to 23) and predictands (up to 14) used may have an effect upon model skill.
- The season of testing may also have an effect upon skill (Section 4.2.1 and 4.2.2)
- There are no other conditions to adjust in the OSR model except the t-test level for correlation rejection. To ensure good validation results for both temperature and precipitation a value of 1.0 has been selected (Briffa *et al.*, 1983) and will not be altered.
- The RBF ANN cannot be improved in skill by removing predictors or predictands. The number of nodes used in construction is, however, limited by the number of predictors available.
- However, the order in which predictors are presented to the RBF model may have some effect upon the clustering algorithm. This is determined via the use of a seed value for randomisation. As discussed in the section above, this seed value will be held constant to aid repeatability at the expense of optimisation, and will not be altered (Harpham, 2005).
- The basis function chosen for each neural network is likely to have a substantial effect upon RBF model performance.
- Intrinsic to this selection is the choice of an appropriate basis width.
- Finally, in addition to the theoretical limits placed upon the number of nodes used in the network (Section 4.2.1), an explicit limit can be imposed (this is referred to as 'early stopping').

So that changes in the above are traceable (i.e. any alteration of model skill can be directly attributed to the change in a single parameter) both approaches are initially set to arbitrary starting conditions. The OSR model is set to utilise all available predictors and predictands. The RBF neural network is constructed with a linear basis function (the most simple of those available), an arbitrary basis width of 10 (as all data is normalised, although the linear model does not require this variable, it is required when testing differences between functions), and no early stopping. All model runs are initially cross-validated as described in Section 4.2.

As a very large number of model runs are required to conduct a comprehensive sensitivity test, the number of stations have been limited for this part of the study to those shown in Figure 2.2 (Bordeaux, Montseny Turo, San Javier, Paganella, Pisa, Prizzi, Calarasi, Tripoli, and Ierapetra). These selected stations (see Table 3.1) provide a wide range of altitudes (0-1706m), latitudes (35.0°N-46.1°N), and longitudes (0.8°W-27.4°E).

OSR sensitivity results

As discussed above, there are few parameters that can be altered in the construction of the OSR model. It was found that rather than the number of predictors used for testing the sensitivity of the OSR model, different combinations of particular predictor indices substantially altered skill. With regard to the correlations found in the previous chapter (Chapter 3.6) and in the literature (Chapter 2), different sets of predictors were created that displayed a large range of correlations with predictands (both very high and very low), and although different stations required different sets of climate predictors, some commonality was found between stations, and particularly between different Mediterranean regions (i.e. west, central, eastern) for each season. Lower limits of skill were generally found with a single, poorly correlated climate predictor (with r often approaching zero). Upper limits of skill were found with climate predictor sets numbering between 9 and 18 varied by both season and region (Table 4.1), to form a Regional OSR predictor set (ROSR), or with a full, Basin wide OSR predictor set (BOSR) (Table 4.2). The predictors in the BOSR set vary only by season, as dictated by seasonally differing principal component predictors (Section 3.5.5).

For winter and spring, full climate predictor sets (BOSR) generally resulted in greater model skill (Figure 4.2). The greatest improvements in skill (over single predictors or a full set) using a climate predictor subset (ROSR) occurred during summer and autumn (Figure 4.2). In many parts of the basin (particularly the east) autumn and summer show similar magnitude temperatures and rainfall values, distinct from behaviour for winter or spring (Chapter 3.4.2 and Chapter 3.7), in addition summer and winter show different sets of climate predictor indices with shared variability and a greater degree of correlation with climate predictands (Chapter 3.5.3, 3.6, and 3.7). Autumn and spring season OSR models were generally more sensitive to

changes in climate predictor sets than those for winter and summer, particularly in terms of the spatial coherency of results. Spring (autumn) results displayed a greater (lesser) spatial coherency for the all-predictors model (BOSR), and less (more) so for a predictor subset (ROSR). Mediterranean climate behaviour is subject to a greater range of influences during these seasons during a changeover from a zonal to a meridional circulation (Chapter 2.7 and 3.7). It can also be seen from Figure 4.2 that for both OSR models seasonal skill results are less coherent (occupy a greater range of values) for summer than other seasons, and more coherent for winter, reflecting the tendency of the Mediterranean climate for increasing regionality during summer (Chapter 2.7, 3.4.2, and 3.7). Seasonal variations in skill are discussed in greater detail in Section 4.3.2.

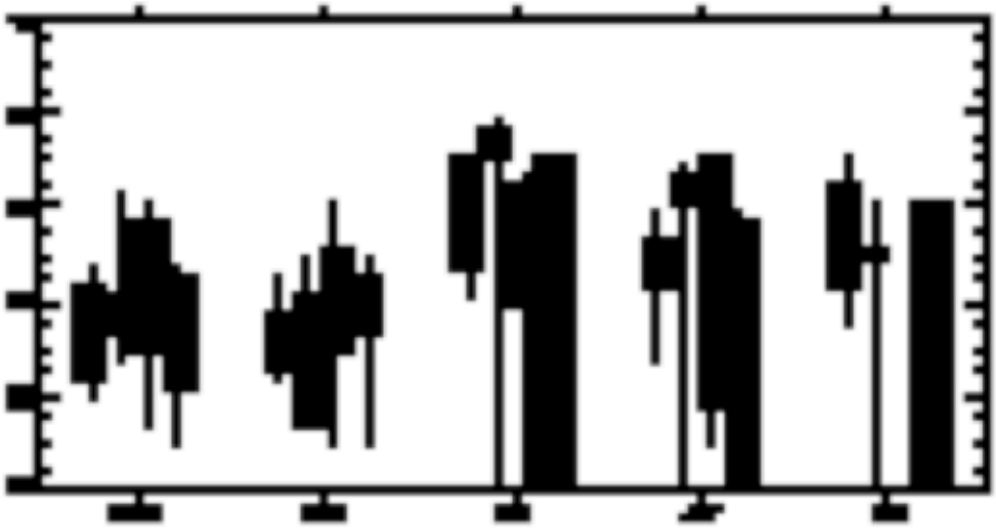

Figure 4.2: Sensitivity test performance (r) of models. Boxes represent winter (dark blue), spring (green), summer (orange), and autumn (light blue) 25^{th}-75^{th} percentiles of performance (r) for all predictand indices at all tested stations, whiskers are limited by the 10^{th} and 90^{th} percentiles of performance (r). BOSR is an OSR model with all available predictors included, ROSR is an OSR model with a sub-set of predictors varied by season and region (west, central, east). LBF, GBF, and TPF, are Linear, Gaussian, and Thin Plate Spline function neural networks.

OSR is a naturally multivariate method, and can model relationships between large numbers of both dependent and independent variables. However, due to the PCA core of the method, skill increased when modelling a single climate predictand index (e.g. TMAX) over multiple stations, rather than multiple predictand indices at a single station (e.g. Paganella). As PCA creates orthogonal components of the original data sets (Section 4.2.1) OSR is unaffected by the multi-collinearity issues that often create

problems for regression modelling of this kind, and the multi-station approach is therefore entirely statistically acceptable (Chapter 1), although certain issues arise from its use (see below).

ANN sensitivity results

As ANN architecture has no implicit means of overcoming issues related to multi-collinearity, each network has been applied on a station by station basis, modelling all indices at one point simultaneously, before moving on to the next (Chapter 1). When applying neural network models, the basis function used to construct the network changes the model at a fundamental level, and is therefore likely to affect ANN model sensitivity. Multi-Quadratic, Inverse Multi-Quadratic, Gaussian, Thin Plate Spline, Cubic, and Linear functions were tested (as discussed in Section 4.2.2). However, Cubic functions were found to systematically underperform, while the functions that returned the greatest overall levels of skill were Linear, Gaussian, or Thin Plate Spline (for all seasons but summer), with Linear ANN models generally outperforming non-linear ANN models. A more detailed discussion of comparative model performance is included in Section 4.3, but only for the latter three basis function networks.

As for the OSR climate models discussed above, seasonal skill can be shown to display varying sensitivities to parameter changes, including alterations to the basis function under consideration. The three functions with greatest overall performance are included in the above plot (Figure 4.2). Although they outperform Cubic or Quadratic models for winter, spring, and autumn, the Thin Plate Spline networks used in this study fail to converge (and therefore produce negligible reliable skill) for Mediterranean summer climate, and results are therefore not presented. Of the remaining results, winter and summer show the least sensitivity to change in basis function, with greater changes in mean skill due to a change in basis function for spring and autumn. This is consistent with a greater sensitivity to OSR parameters in the transition seasons, as discussed above. As also discussed above (for OSR climate models) the range of results for summer are much greater than for winter for both Linear and Gaussian models, Haylock *et al.* (2006) describe similar behaviour for the U.K. as a function of the spatial coherence of the seasonal climate. However, spring ANN results, although highly

sensitive to changes in basis function, and showing an overall greater range of results, are generally (for the central 75% of results) more coherent than results for any other season. Generally high levels of spring coherency can also be seen in the BOSR model (Figure 4.2). Mediterranean spring climate shows a generally less-significant trends (for all indices) than for any other season (Chapter 3.4.4), which may influence results. For each of the ANN models presented above, autumn results show a lack of coherency, potentially as a function of similarities between the summer and autumn Mediterranean climate (Chapters 2.7 and 3.7). In general, although skill is generally higher, results display a much greater spread for ANN models than OSR models. The higher range of values follows from differences in the application of each form of model, as the neural network models developed here consider stations separately, while the OSR models consider each station only as a contribution to a given factor of variance. More detail on seasonal differences in ANN and OSR skill can be found in Section 4.3.

For those basis functions that require the use of a basis width (Section 4.2.2), little change in skill (around -0.01 r) occurred for very large variations in basis width above a value of 10 (tested 10-10,040, incremented in a logarithmic fashion, i.e. 10, 20, 40…) for all ANN models, in all seasons. Below the starting value (i.e. 10), skill was found to decline, more so with very small basis widths (0.1 and below) as a function of the fact that all data are normalised before entering the ANN model, and that the basis width is determined by the order of magnitude of the data (i.e. largely between plus or minus 4). In terms of ANN model structure, skill was also found to be dependent upon the number of internal nodes within each model, limited by the number of climate predictors used (Section 4.2.2). By the 20^{th} node, the skill (and associated errors) gained from any given ANN model had a tendency to approach a constant value, and in most cases values of r flattened out by the 12^{th} node (Figure 4.3).

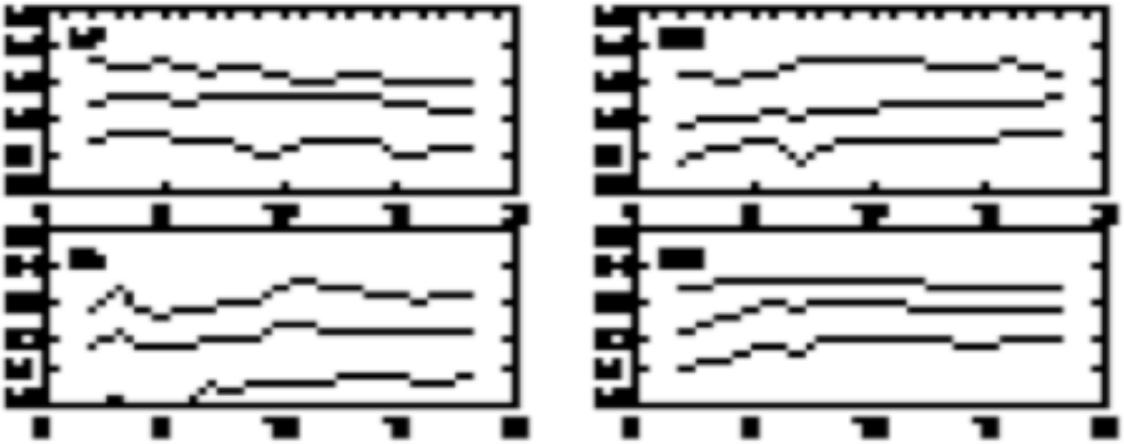

Figure 4.3: Variation of neural network skill (r, y-axis) by season with increasing number of nodes (x-axis) for multiple model types (all 6) and all stations (9, as given above). Solid lines represent upper and lower bounds of skill and the mean.

ANN model testing showed no additional gain in skill through a reduced climate predictor set. In all cases the highest level of skill could be found with the maximum number of available climate predictors. These results are supported by the theory upon which neural networks are constructed, as inappropriate entries are simply weighted to zero as part of the recursive process (Section 4.2.2). Altering the number of climate predictands had no discernible effect upon the results of the ANN model, although theory dictates that model complexity must increase for every additional predictand included (Section 4.2.2).

During sensitivity testing it was found that both models possess issues when it comes to missing or null values. OSR performs array division to produce a variance/covariance matrix (Section 4.2.1), so excessive null values cause arithmetic errors. Similarly the ANN technique finds it difficult to work from an entirely or largely null starting point, yielding floating-point errors. The latter technique is far more tolerant of the null problem than the former is of missing values, but will still not work effectively with the 'frost days' index across the basin. There are therefore no results shown for ANN modelled frost days in this study.

As the OSR model runs are performed on multiple stations simultaneously, even one station with an excessive number of missing or null values (i.e. where over 50% of the validation period is missing or null) can cause the model to fail. Where this is a

problem those stations have been removed from the predictand series used for analysis. In some extreme cases (frost days in the east and some summer rainfall indices) this results in zero skill for an entire region, or for the entirety of the basin.

4.2.4 Summary and model variants

In this section two very different forms of statistical downscaling have been presented, and tested for seasonal and internal sensitivities to model parameters. As a result, two optimal OSR variants and three optimal ANN variants have been constructed (summarised in Table 4.2):

- A basin-wide OSR based model (BOSR) that uses all of the available predictors, applied across the entirety of the Mediterranean simultaneously, to one predictand at a time.
- A regional OSR model (ROSR) that uses a limited number of predictors (Table 4.1) applied separately to western, central, and eastern regions (Section 2.2.1), to one predictand at a time.
- A Linear RBF neural network (LBF NN) applied separately to each station in turn, but for all predictands.
- A Gaussian RBF neural network (GBF NN) applied separately to each station in turn, but for all predictands.
- A Thin Plate Spline neural network (TBF NN) applied separately to each station in turn, but for all predictands.

These variants will be used for the remainder of this Chapter, and between them:

- Represent both linear (OSR, linear ANN), and non-linear (Gaussian and thin plate spline ANN) methods,
- Represent variance based (OSR) and surface fitting (ANN) approaches
- Are applied with respect to issues of multicollinearity (Section 4.2.3)
- Undergo rigorous calibration and validation (Section 4.2)

- Are applied seasonally, justified by previous studies (Jones *et al.*, 1987, Chapter 2.1) and varying sensitivities (Section 4.2.3)
- Use basis functions (where applicable) limited in number by both non-singularity issues (Section 4.2.2) and sensitivity testing (Section 4.2.3)
- Use parameters (such as basis width, early stopping, etc.) and predictor sets identified by the sensitivity testing described in this section, with respect to the theory outlined in Section 4.2.2.

4.3 Model performance

4.3.1 Definitions of model performance

Once the above sensitivities and model variants have been determined, models can be constructed to downscale from hemispheric-scale circulation predictors to station-scale indices of extremes / predictands. This Section (Section 4.3) describes performance differences for the models defined in Section 4.2.4 as applied to the climate predictors and predictands defined in Chapter 3.4.1 and 3.5.2 in an effort to offer a recommendation for 'best use' in terms of statistical downscaling. In addition, the internal weightings of skilful models may offer some insight into the causes of extreme behaviour when taken with the evidence presented in the previous Chapter (See Section 4.2.2). Performance is defined in this Section largely in terms of model skill, as utilised in the sensitivity test conducted above (Section 4.2.3). Where results, or skill, are mentioned in the following discussion they refer to r, categorised as representing moderate (significant at the 0.10 level), good/highly significant (the 0.01 level), or very good/very highly significant (the 0.001 level) correlations (Section 3.3.5) between modelled and observed time series (Wilks, 1995) (Table 3.2). Two other measures of performance, both also based upon the differences between observed and modelled data, are also used.

Ratio of Variance (V)

The ratio of variance is the variance of the modelled data set divided by that of the observed data set:

$$V = \delta_m^2 / \delta_o^2 \quad \textbf{(4.14)}$$

If modelled and observed data are identical V is equal to 1, and the closer to unity V becomes, the more similar the amplitudes of the observed and modelled values. As a ratio this value possesses no dimension.

Root Mean Squared Error (RMSE)

This measure of model performance is self-descriptive, and is the average of the squared differences (x) between each n points of the modelled and observed data, then square rooted:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \quad \textbf{(4.15)}$$

RMSE measures the average difference between observed and modelled data. If modelled and observed data are identical, RMSE will be zero. However, unlike the ratio of variance, RMSE retains the units of measurement of the initial data set, and values are therefore not directly comparable when assessing performance across different predictand indices. In cases where RMSE is compared in such a fashion it can be normalised by the standard deviation of the observed data, or converted to ranks.

The two quantities detailed above may be related to r by the following equation (4.16):

$$RMSE^2 - E_0^2 = 1 + V - 2r\sqrt{V} \quad \textbf{(4.16)}$$

Where E_0 is the overall difference between observed and modelled means.

Equation 4.16 gives the ability to plot the three statistics of performance (r, V, and RMSE) given above on a two dimensional plot, known as a Taylor diagram (Taylor, 2001). Taylor diagrams (Taylor, 2001) of model performance are presented in Figures 4.4 to 4.13 and discussed below (Section 4.3.2). These diagrams display V on a radial co-ordinate, and r on the angular co-ordinate. Where points are closer to the x (y) axis, they show greater (less) model skill, and where they are closer to a line curving between the 1.0 mark on both x and y axes they show a more accurate representation of variance. Points further from (nearer to) the origin than the *V=1.0* line show an overestimation (underestimation) of variance. It follows from Equation 4.16 that RMSE is then the distance between any given point and a reference point, which in the case of this study would be the statistics of the observed data against itself (a perfect model), in all cases represented as a variance ratio (V) of 1.0 and a correlation coefficient (r) of 1.0. In Figures 4.4 to 4.13 Taylor diagrams are presented for each index, and for each model variant. Data points for every climate station are included on each plot, for each season, and are presented with a 'perfect model' reference point on the x-axis.

4.3.2 Variant comparison

The first comparative analysis of performance concerns an overall assessment of how each of the model variants replicates observed data. This analysis is then followed by seasonal and regional analyses. Abbreviations used throughout this section are detailed further in Tables 3.4 (climate predictands) and 4.2 (climate model variants).

Overall performance

It is clear from the Taylor plots (Figs. 4.4-4.13) that both OSR models (Figs 4.4-4.7) systematically underestimate the variance (V is largely between 0.2 and 0.8) of each predictand index. Although the OSR method does reject a number of components of variance as part of the modelling process, over 95% of the predictand variance should be retained. The OSR models used here underestimate predictand variance more for the ROSR (Fig 4.6-4.7) variant than the BOSR (Fig 4.4-4.5) variant, mildly less for temperature than precipitation (by around 0.1), and generally less for spring than autumn (particularly for the BOSR variant). Underestimation of predictand variance is

less evident for TAVG, TMIN, TMAX and TX90 than for TN10, TNFD and HWDI indices. Evidently a larger number of predictors are required to capture a larger proportion of predictand variance, the uneven distribution of precipitation variance (which has a lower bound), causes greater problems for the OSR method than the more normally distributed temperature distribution (Wilks, 1995), and the greater range of both autumn precipitation and temperature values (by comparison with spring values, Chapter 3.4.2) has a negative effect on the OSR method's ability to replicate predictands. It also seems likely that there may be different statistical distributions attached to measures of duration and magnitude with processes behind the former (length of frosts, and heatwaves) more complex than the latter (Chapter 2.4.2), resulting in a difference in performance. The underestimation of predictand variance is not uncommon to modelling methods (Wilby *et al.*, 1998), particularly when modelling rainfall data, and is termed the 'overdispersion' problem (Katz and Zheng, 1999; Goodess 2000). Possible causes for this problem include inadequate representations of high-frequency variability in models, and inadequate representation of low-frequency forcings, both of which may lead to underestimation of both variance and persistence (Katz and Zheng, 1999; Goodess, 2000). As various low frequency forcings are considered within the models utilised here (e.g. the NAO, humidity, sea level pressure) the former explanation seems more likely.

Both OSR models generally produce skill in the 0.20-0.60 range. For the BOSR model, Taylor plot results are roughly diamond shaped, showing that predictand variance estimation is relatively low where skill is either comparatively very low or very high. This variance/skill behaviour is less apparent for TN10 and TX90 than other temperature indices, and more so for precipitation indices than temperature indices. For the ROSR model, variance estimation mildly declines with an increase in skill for all temperature indices but TX90, and all precipitation indices but PREC, PINT, and PN90. These other predictand indices show a relationship between skill and variance closer to the diamond shape given above. In the case where variance estimation declines (by as much as 0.2 V) with high values of skill it is likely that the OSR model's regressive part is more effective than its ability to accurately capture large-scale variability, and that as with many other models (McAvaney *et al.*, 2001) the mean of the temperature or precipitation index under consideration is being more accurately represented than the total variability. It can be seen that the OSR method's theoretical basis in variance

calculation may give little improvement over other regressive methods in practice as improvements in skill reduce the variance of the modelled data.

A larger proportion of precipitation indices (particularly PX5D and PCDD) offer very low levels of skill (near zero) than temperature indices, more so for the ROSR variant than the BOSR variant and more so for summer and autumn than any other season. Otherwise, skill levels are generally highly consistent between seasons, and within a given range, with no one season or predictand index obviously providing a much greater level of skill than any other. Furthermore, from these plots it is difficult to suggest a given circumstance where, overall, the BOSR variant provides greater or lesser levels of performance that the ROSR variant.

Skill and variance for the neural network models (Figs. 4.8-4.13) are far more variable between stations and seasons than the OSR models. The neural network variants utilised here both under- and over- estimate model variance. In the case of the Linear (Fig. 4.8-4.9) and Gaussian (Fig. 4.10-4.11) neural network models, overestimation is more of a problem for temperature indices than precipitation indices. For all ANN variants (but less so for the TBF variant), as the BOSR variant, overestimation of variance is more evident for TAVG, TMIN, TMAX and TX90 than for TNFD and HWDI indices, suggesting that the issue is systematic, and based around the inherent qualities of the predictand data. TN10 behaviour lies somewhere between the two groups. There is no seasonal bias in variance estimation for the former group (magnitude), but for both the latter temperature (duration) indices and the precipitation indices, autumn variances are again routinely underestimated. It is also clear that overestimation of variance is less of a problem for precipitation predictand indices than temperature indices for both the Linear and Gaussian variants. General skill for the Thin Plate Spline variant (Fig 4.12-4.13) is lower than for the other two (discussed further below), and it appears that it is much more prone to overestimation of precipitation indices than either the Linear or Gaussian variants. In terms of temperature indices, the Linear model is more prone to overestimation of variance. As all variants are prone to both over and underestimation the most appropriate model for use, if concerned with the replication of variance, is that which clusters closest and most compactly around the 1.0 V line. As with the OSR variants, no model is ideal, but for magnitude of temperature predictand indices (i.e. TAVG, TMAX, TMIN, TX90) it can be seen that the Gaussian

model is more appropriate, while for duration indices (i.e. TNFD, HWDI), the Thin Plate Spline is better suited. For precipitation indices, although little coherency between indices is displayed, the most balanced estimator of variance is the Thin Plate Spline, as the other variants have a greater tendency toward underestimation.

There appears much less of a discernible, consistent, relationship between skill and variance for the ANN variants than the OSR variants, however. For the magnitude group of temperature indices there is no particular relationship evident, although the distribution of skill is very tight for these indices (see below), which may make such a relationship inherently more difficult to assess. For TNFD and HWDI indices, a slight diamond shape is apparent in results, as discussed above. Precipitation index behaviour is much less consistent between variants, and where a particular distribution/relationship may be apparent for one predictand index and one ANN variant, it may not be for other indices or variants. This lack of apparent relationship between skill and variance estimation, where one exists for the OSR variants, may simply be a facet of the fact that the ANN models have been applied separately for each station, as opposed to collectively for each index.

The difference in application between one and many stations may also contribute to the fact that the ANN variants are more prone to producing negligible skill for both temperature and precipitation indices. Although this behaviour is most evident for JJA TNFD in the Gaussian model (due, again, to an absence of summer frosts), whether or not these very low results are generally consistent to a specific set of stations is explored in greater detail below (Section 4.3.3). Results for the LBF ANN (Fig 4.8-4.9) are the least coherent of all models in terms of skill, but do show higher levels of overall skill than any other model (r=0.5-0.8), even for measures of extreme temperature duration (r=0.4-0.7) and precipitation (r=0.4-0.8). Gaussian temperature results are the most coherent in all seasons but summer, with winter, spring and autumn skill occupying a very small range for the magnitude temperature indices (around 0.1), and summer results much more evenly spread (0-0.8). Levels of skill for both Gaussian and Linear temperature magnitude indices show a consistent seasonal bias across all stations. Seasonal differences in skill are discussed further below. Although performance is generally slightly lower for the Thin Plate Spline variant than the other ANN variants, precipitation skill results show a much greater degree of coherency than

those for either the Linear or Gaussian approach (as well as less of a tendency toward the underestimation of variance). Overall the Linear basis function neural network produces a greater level of skill than either the other ANN variants, or the OSR approaches. However, due to a greater coherency of results the Gaussian or Thin Plate Spline models may be more appropriate for a given season or region. These possibilities are explored below.

Seasonal performance

The OSR Taylor diagram discussion regarding a lack of seasonal variability in results is confirmed when skill distribution (across all stations) is plotted in a box and whisker diagram for each season and index (Figure 4.14). This form of analysis allows for an assessment of the relative performance of each variant between seasons, and whether or not particular indices are well represented in each season for all models.

For all temperature indices the majority (75%) of both BOSR and ROSR skill lies between 0.2 and 0.5. For rainfall indices the same clustering is evident, except where a significant number of stations display very low levels of skill during summer, autumn, or (only for PX5D, PN90, and PF90) either winter or spring. It is certainly evident that the OSR method has difficulty replicating Mediterranean summer (and less so autumn) rainfall, probably as a function of its general absence (Chapter 2.7 and Chapter 3.4.2). Although the range of values is very small, specific indices that show relatively high levels of general skill (for 75% of the distribution) include:

- Winter:
  - Extreme temperatures (TX90 and TN10) for the winter BOSR models,
  - Average and cold (TMIN and TN10) indices for the winter ROSR models,
  - Heavy (PQ90, for both BOSR and ROSR) and intense (PINT, for ROSR) winter rainfall.
- Spring:
  - High temperatures (TX90) for spring for both BOSR and ROSR variants,

    - Average (PREC) and intense rainfall (PINT) for the ROSR models, dry days (PCDD) for the BOSR model.
- Summer:
    - Average (TAVG) and low temperature indices, with relatively high skill for cold temperatures (TMIN) for the ROSR variant, and extreme cold temperatures (TN10) from the BOSR variant,
    - Average precipitation and dry day (PCDD) indices for both variants. Intense rainfall (PINT) for the ROSR variant.
- Autumn:
    - Average temperature (TAVG) for the BOSR variant,
    - Persistent rainfall (PX5D) for the BOSR variant,
    - Average and intense rainfall (PINT) for the ROSR variant.

It can be seen that certain seasonal consistencies occur between the two OSR variants used here: for winter, low temperatures and heavy rainfall; for spring, high temperatures; for summer, persistent dry days. These consistencies in relatively high skill reflect the dominant form of seasonal climate. The Mediterranean spring climate is defined by increasingly large arid regions, and intense rainfall that occurs on a more regional basis. The former behaviour is represented more skilfully in the BOSR variant, and the latter by the ROSR variant. It has already been shown that the Mediterranean summer would be largely arid if not for regions of intense rainfall (Chapter 3.4.2), and intense summer rainfall is well modelled by the ROSR variant. The regional nature of Mediterranean intense rainfall is also evident in the autumn ROSR model's performance, while long periods of persistent rain (generally over larger areas) are well modelled by the BOSR model (Chapter 2.4.2 and 3.4.2). It would appear that although differences in skill are small, both basin-wide and regionally focussed OSR variants perform well when applied to the seasonally (basin-wide, or regional) dominant climate.

It can be seen in Figure 4.14 that the consistency in results for each season, for all of the ANN variants, is much higher than that for the OSR variants for winter, spring, and autumn. The neural network variants display a much greater sensitivity to seasonal climate (as shown in Section 4.2.3), more so for temperature indices than precipitation indices. However, although the Gaussian variant is significantly worse at replicating summer climate than the Linear variant, these two basis functions otherwise

show a substantial similarity in their seasonal performance bias, and where between-season relationships occur for non-summer Linear results, they are also applicable to non-summer Gaussian results. Summer rainfall results for both variants, while generally lower than those for other seasons are always higher than those using the OSR variants. With the exception of TX90, HWDI and PCDD skill (all connected with very high temperatures), the TBF variant systematically presents greater skill for all indices for winter, then autumn, then spring. With these caveats it can then be seen that (as above) certain specific indices show relatively high levels of general skill (for 75% of the distribution) in each season, and that they are highly consistent between ANN variants for most seasons:

- Winter:
  - Low temperatures (TMIN) and frosts (TNFD) for all variants,
  - Rainfall (PREC) and dry days (PCDD), also for all variants.
- Spring:
  - Central temperature indices (TAVG, TMIN, TMAX)) and extreme heat (TX90) for all variants,
  - Intense rainfall (PINT) and dry days (PCDD) for all variants.
- Summer:
  - Low temperatures (TMIN) for Linear and Gaussian variants,
  - Rainfall (PREC) for the Linear variant, dry days (PCDD) for the Gaussian variant.
- Autumn:
  - Central temperature indices (TAVG, TMIN, TMAX) and extreme heat (TX90) for all variants,
  - Intense rainfall (PINT) for the Linear variant, dry days (PCDD) for the Gaussian, and rainfall (PREC) for the Thin Plate Spline variant.

Some similarities can be seen between the above list and the influence evident in the distribution of skill between seasons and indices for the OSR variants. Winter performance is good for indices representative of the dominant climate, and similarly, intense rainfall is skilfully modelled for both spring and autumn, as are central temperatures and high temperatures. Also common to both forms of modelling are the

relatively high levels of skill for summer low temperatures, that of all the rainfall indices PREC, PINT, and PCDD offer high levels of skill, and the fact that during autumn, different rainfall indices require different approaches. It is interesting that unlike the OSR variants, PREC and PCDD indices offer the best skill during summer and winter, and that PINT and PCDD offer the best skill during the more transitive autumn and spring seasons.

From the above, it can be seen that skill for the Linear ANN variant is generally better than other variants in all seasons, and that non-linear ANN variants out-perform OSR variants (with the exception of the TBF variant during summer). However, the set of indices of extremes that are best replicated by models is more dependent on the dominant behaviour of the climate itself than the form of modelling used. The results presented above reflect well on the models used and carry a degree of implication that all the variants used above have developed relationships that are in some way indicative of the processes evident in the underlying climate. If this is the case then it should be expected that skill would not only vary seasonally but also regionally, as Mediterranean climate processes have been seen to display distinct regionality (Chapter 3.4.2). Thus the following discussion concerns differences in skill between the western, central, and eastern basins.

Regional performance

Box and whisker plots of skill arranged by region and index are presented in Figure 4.15. Here the distribution represented is that of results for all applicable stations, for all seasons. This analysis allows for a comparison of how each model variant reproduces values for different large-scale regions, and whether or not certain indices are more accurately replicated in particular regions.

As for the seasonal distribution of OSR variant skill, it can be seen that the differences in distribution for each predictand index between regions is relatively small and that no one region consistently offers greater skill than any other. However, a number of particularly significant results for each region can be identified:

- Western:
  - Very high temperatures (TX90) for the BOSR variant, no clear 'best' for the ROSR variant,
  - Dry days (PCDD) for the BOSR variant, rainfall (PREC) and intense rainfall (PINT) for ROSR.
- Central:
  - High temperatures (TMAX) and frosts (TNFD) for both variants,
  - Rainfall (PREC) and intense rainfall (PINT) for both variants.
- Eastern:
  - High temperatures (TMAX) and frosts (TNFD) for the BOSR variant, average (TAVG) and low (TMIN) temperatures for ROSR,
  - Persistent rainfall (PX5D) for the BOSR variant, Rainfall (PREC) and intense rainfall (PINT) for ROSR.

Throughout the basin, as discussed above, PREC and PINT are generally well modelled. The western basin connection between TX90 and PCDD is clear, as very high temperatures and periods without rainfall are dynamically linked (Chapter 2). The division apparent between skill for western high temperatures, and skill for central and eastern frosts, in addition to high temperatures, is likely to be a facet of the fact that (as discussed in Chapter 2.2.1 and Chapter 3.4.2) the eastern end of the basin experiences more cold (continental) winter weather than the west.

Although the skill for each region is still markedly higher with the neural network variants than the OSR variants, the differences between skill levels for each region do not appear much greater in the linear ANN model. Linear ANN model performance for the western region is similar to that for the eastern region. The same is not true for the non-linear neural networks, and in both the Gaussian and Thin Plate Spline variants, eastern temperature index skill is generally much worse than that for the west. For rainfall indices the contrast between western and eastern skill is much closer to that for the linear ANN variant. Although seasonal results show a greater coherence for the non-linear ANN variants, regionally greater coherence is shown by the Linear ANN variant.

Indices that show particularly high (Table 3.2) general performance (i.e. for 75% of the skill distribution, including all stations and all seasons) include:

- Western:
  - Low (TMIN) and very high (TX90) temperatures for all variants,
  - Rainfall (PREC) and dry days (PCDD) for all variants.
- Central:
  - Low (TMIN) and very high (TX90) temperatures for all variants,
  - Rainfall (PREC) and dry days (PCDD) for all variants.
- Eastern:
  - Low (TMIN) and very high (TX90) temperatures for the LBF and GBF variants, high (TMAX) temperatures for the GBF variant, and average (TAVG), high (TMAX), and very high (TX90) temperatures for the TBF variant,
  - Dry days (PCDD) for all variants.

This set of indices is even more consistent between regions than for the OSR variants. Although the indices listed above differ from those evident for the OSR variants, they are still among those highlighted by the seasonal analysis as generally well represented. Less agreement between variants occurs for the eastern region due to a systematic difference in the response to eastern climate (and generally lower skill), as evident for all temperature indices.

One notable conclusion to come from the discussion of regional performance is that, although non-linear ANN methodologies generally produce lower (higher) skill than the linear method ANN (OSR variants) presented here, they are particularly poor when applied to eastern basin temperature indices. However, from the above it can be seen that for each variant the expected variation in skill with location is generally not as large as the seasonal variation, when results are compared for large regions. In fact, the regional response is surprisingly consistent in terms of the indices that are best represented. The following section explores the possibility that climate model performance varies for smaller regions than those explored above.

4.3.3 Regional comparison and physical dynamics

The results presented in Section 4.3.2 suggest that the model variants applied here (Table 4.2) are replicating some measure of seasonal variability, but that (on a relatively large scale) differences in performance between climate predictand indices and seasons are greater than differences in performance between regions. This section discusses model performance (in terms of skill- r) at station-scale resolution, for each index, variant, and season (Figures 4.16-4.55), in an attempt to assess whether or not this is the case, and to assess whether differences in the level of skill apparent for particular regions are consistent with differing regional climate mechanisms. Where 'spatial coherence' is referred to below, it regards the degree to which performance is the same from one region to the next. As it has already been found that seasonal differences in skill may prove substantial this analysis is conducted on a season-by-season basis.

As in the overall performance part of Section 4.3.2 there are a number of stations that return negligible levels of skill. This part of the analysis allows for the isolation of these stations and an assessment as to whether or not they are consistent between indices or variants. Where negligible results are consistent for a given index across all variants predictability is inherently low or the required predictors are entirely missing from those described in Chapter 3.5 and subsequently utilised in the models. Where negligible results are consistent across all indices for a given variant (but not others) that particular approach is unsuitable for the given station.

Winter

This part of the analysis concerns Figures 4.16-4.25.

It is immediately clear from the performance of Mediterranean areas used in the previous section (west, central, east) sub-regional differences in performance are much greater for the BOSR variant than the ROSR variant in terms of temperature, and generally greater for the OSR models than any of the ANN variants. It is also apparent that sub-regional differences in performance are much larger for the TNFD and HWDI indices (for all variants) than those that represent measures of magnitude.

Although levels of both skill and spatial coherence differ between the BOSR and ROSR variants, similar regional patterns for temperature performance can be seen in both (Figs 4.16 and 4.18). The BOSR variant shows consistently good or very good performance across Portugal, most of Italy, and the western coast of Greece for all temperature indices. Performance is largely moderate for eastern Greece and central Iberia for all temperature indices but TMIN and TX90 (both displaying good levels of performance in those regions). Poor or very poor performance occurs for the north (TMIN, TMAX, TN10) and north west (TAVG, TMIN, TNFD) of Iberia and southern France (TAVG, TMIN, TMAX, HWDI, TNFD) for most temperature indices. For winter temperatures negligible skill for the OSR variants can only be seen for the BOSR frost model (Fig 4.16) at the following stations: Salamanca, Valencia, Alicante, Seville, Alcantarilla, San Javier, Jerez De La Frontera, Zamora, Coruna, Sabiñanigo, Montseny Turo (all Spain), Alghero (Sardinia), Nis (Yugoslavia), Ierapetra, Kythira, and Skyros (all Greece). The majority of these stations are located in the warmer southern part of the basin, or near coasts, where frosts rarely occur.

With the exception of northern Italy, performance is generally highest for low latitude stations on western coasts. The two circulatory phenomena that link these locations are eastward flow from the Atlantic and the subsequent transport of air made relatively warm by the sea (Chapter 3.4.3), and flow north and eastward from Africa of the kind that produces omega-wave heatwave conditions (Colacino and Conte, 1995) during summer (Chapter 2.4.3). Northern Italy is not subject to either of these influences, but is the one location that experiences the same kind of land-ward warming flow travelling into Europe from the Mediterranean that may develop under extreme conditions into a Vb storm track (Section 2.3.1; Ulbrich *et al.*, 2003; Caspary, 2004). There is a substantial amount of sub-regional variation within BOSR temperature index performance, as the regions that show poor skill are distinct from those that show good or very good performance. For the ROSR variant, however, a generally higher level of skill produces a greater degree of spatial coherence for temperature, as although patterns of performance are similar, where skill is poor for the BOSR variant, it is generally moderate for the ROSR variant.

ANN variant results for temperature predictand indices (Figs. 4.20, 4.22, and 4.24) show a high degree of spatial coherence, and the Gaussian variant does not show the difference between eastern and western skill seen in Figure 4.15. Skill is generally very high across the majority of the basin for all indices except HWDI, which displays only moderate or high skill in similar locations (northern and eastern Spain, southern France, eastern Italy, and south eastern Greece) for all variants. These regions are the same as those for which moderate or poor skill is found with the OSR variants, as described above. Temperature predictand index performance for ANN variants differs from that for OSR variants in one other respect, two stations show consistently negligible skill across all temperature indices and variants: Nis in Yugoslavia and Sabiñanigo in northern Spain. The ANN method is clearly not appropriate for these stations, and greater skill for all temperature indices can be found using the ROSR variant. Otherwise the most effective method (in terms of skill) for all winter temperature indices, is the Linear ANN variant.

When spatially disaggregated, winter precipitation indices (Figs 4.17, 4.19, 4.21, 4.23, 4.25), do not perform as well as temperature indices, for all model variants. The difference in performance between temperature and precipitation indices is smaller for ANN variants than OSR variants, but spatial coherence is greater for OSR variants due to large regions of poor skill. For rainfall indices the BOSR model (Fig. 4.17) displays only a small degree of spatial variation, showing moderate to good skill only for northern Iberia and southern France (PREC, PQ90, PINT, PF90, PX5D) and central Greece (PQ90, PF90, PX5D). These areas are either within the Mediterranean region most affected by the NAO (Chapter 2.2.3, Chapter 3.6), or in the lee of the Balkan mountains (affected heavily by westward tracking Mediterranean cyclones, Chapter 2.3.2). Elsewhere skill is poor or very poor. For the OSR variants negligible winter precipitation results show regions of consistency between indices including Salamanca, Valencia, Catalonia, central northern Italy, Sardinia, Nis, Calarasi, and Izmail. Generally far fewer of these negligible stations appear for the ROSR variant (Fig 4.19) than the BOSR variant. As for temperature, spatial patterns of performance are similar for the ROSR variant (Fig. 4.19), which displays slightly higher levels of skill. However where BOSR skill is moderate for the areas described above, it is good for the ROSR variant, which therefore displays a greater degree of sub-regional variations in performance.

The pattern of ANN rainfall performance is common for all ANN variants (Fig. 4.21, 4.23, and 4.25) and is generally higher for PCDD than other indices, for the north west of Iberia and southern France, than the south east of Iberia, and higher for the North west of Greece than the south west. This contrast in performance is more evident for Linear and Thin Plate Spline variants than the Gaussian variant. But for all three it is clear that there are definite sub-regional differences in rainfall performance that are similar to those evident for the OSR variants in the west and east. As for temperature, Nis and Sabiñanigo consistently provide negligible levels of skill for all precipitation indices and the greatest levels of skill are generally provided by the Linear ANN variant.

Spring

This part of the analysis concerns Figures 4.26-4.35

As for winter temperature performance, a substantial difference in the spatial coherence of results is noticeable between those from the OSR variants (Figs. 4.26 and 4.28) and those from the ANN variants (Figs. 4.30, 4.32, 4.34). Again, spatial coherence is much greater for the ANN variants than the OSR variants due to high levels of skill in most locations. As for winter, differences between regions are more noticeable for measures of duration (TNFD, HWDI), than magnitude. However, in contrast to the results for winter, and if disregarding those stations for which TNFD results are very poor, it is the ROSR variant (Fig. 4.28) that shows less spatial coherence than the BOSR variant (Fig. 4.26) for spring, with the former showing relatively high levels of skill across the basin.

The ROSR variant shows three distinct spatial patterns of skill for temperature predictand indices (Fig. 4.28). For TNFD, HWDI, TMAX, and TN10 skill is generally very good across the western coast of Iberia, northern Italy, and Sardinia, either moderate or good to the east of Iberia, and poor for the south of Italy and Greece (although less so for TN10). Performance for these indices generally declines toward the south east of the Mediterranean, where Atlantic influence is weaker and the European influence is more relevant (Chapter 2.2.1). TX90 is the one index that shows

only moderate results for northern Italy, in addition to southern France (both northern, mountainous regions), although skill is good to very good elsewhere. For TAVG and TMIN skill is good to very good in all locations except the north west of Iberia (shown in Chapter 3.4.2 to poses a climate generally unlike that for other nearby locations), the south east of France (subject to the cooling but infrequent Mistral wind), and the south west of Greece (highly affected by occasional regenerated Mediterranean cyclones) (Chapter 2.3). For spring these regions show neither distinctive average patterns of skill (Fig. 3.4), or trend behaviour (Fig. 3.14), and the one predictor that influences all three regions more than others (SHM PC1) is present in the predictor set (Table 4.1), but localised effects on climate may introduce a level of unpredictability. The BOSR variant shows generally better skill in most locations, and a resultant reduction in variations between performance for the central and eastern basins, however, for TMIN, TN10 and TX90 skill is just as poor for northern Iberia and southern France. These regions display poor to negligible skill (Table 3.2) for both winter and spring OSR variants, and may be poorly represented by the OSR method.

As for winter, both Linear (Fig 4.30) Gaussian (Fig. 4.32), and Thin Plate Spline (Fig. 4.34) performance is highly spatially coherent for all temperature indices and regions except for TNFD and HWDI, where skill is moderate or good for those locations where both OSR variants show weakness. The number of stations that show negligible skill across all indices and variants, however, is larger than that for winter, and includes: Bastia (France), Montseny Turo, Sabiñanigo, Zamora, Ponferrada (all Spain), Trevico, Pescara (both Italy), Nis (Yugoslavia), Naxos (Greece). For these stations (during spring) the most appropriate variant presented here is the BOSR variant, otherwise the linear ANN provides better performance in all locations. The one exception to the above is the TNFD index throughout the southern Mediterranean, where the ROSR model gives greater skill.

Also as discussed for winter, spring precipitation (Figs. 4.27, 4.29, 4.31, 4.33, and 4.35) is generally less well represented than temperature. The ROSR variant (Fig. 4.29) shows better performance in all locations for all precipitation indices than the BOSR variant (Fig. 4.27) with the exception of eastern PX5D, where skill is negligible for the ROSR model. As for winter, regions of comparatively good performance are

similar in most instances, but show greater levels of skill in the ROSR model, which therefore displays less in the way of spatial coherence. Regions of high skill are generally toward the north of Iberia, Italy, and Greece, within the path of strongest spring-time jet stream influenced flow (Chapter 2.2.2), but otherwise show little consistency between indices.

ANN variants show a similar relationship to that for winter, with the Linear variant (Fig. 4.31) providing generally better skill, and all variants displaying the same locations of negligible performance for both temperature and rainfall indices. Similar spatial patterns of skill are also evident in all three variants, although they are most visible for the Thin Plate Spline variant, as the relatively greater skill in the other two variants increases spatial coherence. As for winter, the greatest levels of performance can be found across northern Iberia, southern France, northern Italy and (less so) the north west of Greece. These are similar regions to those shown for the OSR variants (although much larger in scale), and are all influenced heavily by westward flow during spring (Chapter 2.2.2).

Summer

This part of the analysis concerns Figs. 4.36-4.43.

Although, as above, both OSR variants show similar spatial patterns of temperature predictand index performance (Fig. 4.36, 4.38), unlike the analyses for spring and winter, west, central, and eastern regions do not universally show one OSR variant as better than the other. For the western region, the ROSR variant (Fig. 4.38) provides higher levels of skill, while for the central and eastern regions, the BOSR variant (Fig. 4.36) gives better results. This difference can only be due to the appropriateness of the selected predictors. Evidently the western summer climate of the Mediterranean benefits from a reduced predictor set (Table 4.1), and the eastern and central summer climates benefit from the use of all the predictors available. Both models show good performance throughout the western basin with the exceptions (as in spring) of central and northern Iberia, and south-western France, and also (different to spring), northern eastern Greece (a region heavily sheltered from western flow). The

region of poor to moderate skill for the first two of these areas is substantially larger for the BOSR variant than the ROSR variant. The reverse is true for northern Greece, where the ROSR variant is more skillful. One or two stations in southern Italy (generally Monte Scuro and Prizzi) provide poor performance for each index for one or the other of the OSR variants, but only for both variants with TNFD and HWDI. With the exception of these two stations, skill is generally better for Italy with the BOSR variant. Negligible results, as for previous seasons, are evident with the BOSR TNFD plot (Fig. 4.36), and more so for the warm southern region than elsewhere, although the generally poorly performing stations of Sabi'anigo and Nis both also show negligible skill.

Also as for previous seasons, the only substantial regional difference for the ANN variants (Fig. 4.40, 4.42) is between those stations that show high or very high levels of skill, and those that show negligible skill. For summer, however, substantial differences in the number of negligible stations occur between Linear (7) (Fig 4.40), Gaussian (between 7 and allstations depending on index) (Fig. 4.42), and Thin Plate Spline variants (all stations) (not shown). In the former case only Bastia, Clermont-Ferrand (France), Sabiñanigo, Leon (both Spain), Pegoes, Mora (both Portugal), and Nis (Yugoslavia), give very poor performance. For the Gaussiaan variant no skill is returned for the TNFD index at all, very few stations in southern Greece show skill for any index but HWDI, the same is true for Portugal except for TN10, and in addition to Nis and Sabiñanigo, Valencia (Spain) and Tarbes (France) show no skill for any temperature index. As this non-linear method is very poor at representing summer temperature predictand skill in many locations it is less surprising that the Thin Plate Spline variant fails completely. It is evident that, rather than providing additional skill, large parts of the Mediterranean summer climate are not well represented by the non-linear methods utilised in this study. It has also become clear that eastern performance is poor in the previous set of plots (Fig. 4.15) due largely to poor performance in the summer models, rather than those for any other season.

As might be expected, concerning the comparative volume of summer rainfall in the Mediterranean (Chapter 2.2.1, Chapter 3.4.2) and the aggregated results in Figure 4.14, skill for summer rainfall is generally poor but less so for PREC and PCDD than

other indices (Fig. 4.37, 4.39, 4.41, 4.43). Also as for other seasons, the same regions of comparatively good skill exist across all variants, with the Linear ANN variant providing the best performance except in those regions where skill is negligible, in which cases the ROSR variant is the best approach.

Regions of relatively high skill are similar to those for the preceding seasons and include southern France, north eastern Spain, and northern Greece. For PREC, PCDD and PX5D, summer results are also generally higher for coastal locations that face the Gulf of Genoa (i.e. eastern Spain, southern France, and western Italy), regions that (in addition to the northern areas listed previously) experience more summer rainfall than the rest of the Mediterranean. BOSR (Fig 4.37) and ROSR (Fig 4.39) variants both give generally very poor (or negligible) skill for the south of Greece, more so for the ROSR variant than the BOSR variant, and more so for the PX5D index than others. PQ90, PF90, and PN90 also give very poor or negligible skill for the majority of the western basin for both OSR variants. In contrast with previous seasons, and despite a similar ANN pattern between temperature and precipitation indices (as in previous seasons) the ANN variants (Fig 4.41, 4.43) generally produce less negligible skill than the OSR variants. For summer rainfall and temperature indices the Linear ANN (Fig 4.41) produces the best level of skill in almost all instances. Where both the LBF and other ANN methods fail, however, the ROSR gives better skill in the central and western regions, and the BOSR gives better skill in the west.

Autumn

This part of the analysis concerns Figs. 4.44-4.53.

As indicated in Figure 4.14, autumn temperature predictand skill is lower for the OSR variants (Fig. 4.44, 4.46) than the ANN variants (Figs. 4.48, 4.50, 4.52). The differences, however, are less noticeable for TN10, TNFD, and HWDI indices, and the degree of regionality in results is much greater for these indices (and the OSR variants in general) than for others. The spatial coherence of performance is generally much smaller for autumn than other seasons, as is the level of consistency in spatial patterns between indices, as reflected in the differences in best performance given in the

seasonal and regional overview analyses (Section 4.3.2). The patterns shown by the BOSR variant (Figure 4.44) are, however, similar to those for the ROSR variant for most indices (Figure 4.46), as they are for other seasons. For both variants, higher levels of skill are generally found for:

- TAVG and TMAX: Northern and eastern Iberia, southern France, central Italy, northern and western Greece
- TMIN: western Iberia, northern and western Greece
- TN10: Portugal, Sardinia, northern and southern (but not central) Italy, most of Greece (except Aginio, Kozani, Ioannina, and Larissa)
- TX90: Western Iberia, eastern coastal Spain and northern Italy
- TNFD: North eastern Spain, southern France, northern Italy and northern Greece.
- HWDI shows a pattern more similar to TX90 for the BOSR variant, and more similar to that for TNFD for the ROSR variant.

But there is no consistency in indices as to which variant produces better skill. Instead, for western TAVG, TN10 and TX90, the BOSR variant is generally more appropriate, for central and eastern TAVG, TMIN, TMAX, TNFD and HWDI, the ROSR variant gives generally better performance. It is clear that although summer climate is generally the least consistent from region to region, the mix of influences evident for autumn (with similar magnitude indices to summer, but a distribution closer to spring) (Chapter 3.4.2) produces a greater degree of variation in OSR model performance between indices and regions than for other seasons.

As mentioned above, ANN results are consistent and highly spatially coherent for TAVG, TMIN, TMAX and TX90, but less so for the other predictand indices. For these, high levels of skill can be found (consistently between TN10, TNFD and HWDI for all variants) for southern Spain, north eastern Spain/southern France, northern Italy, Sardinia, and north western Greece. In the west of the basin, skill is greatest near Mediterranean facing coastal regions and the moderating influence of the Mediterranean sea. In the case of the central and eastern regions, the contrast in skill is due to a complete lack of skill in the east or south east consistent to all indices and ANN

variants. In the west, those stations that provide negligible temperature predictand skill in summer also do so for autumn. In the central and eastern regions, areas of negligible skill are larger for the LBF variant (Figure 4.48) than in any other season, and are smaller only than summer areas for the GBF variant (Figure 4.50). For TAVG, TMIN, TMAX and TX90 the linear variant produces the best skill except for those regions described as negligible, however, for those regions the 'best' of the two OSR methods varies from index to index, as above.

For rainfall indices, modeling variants (Figs. 4.45, 4.47, 4.49, 4.51, 4.53), do not behave in the disparate fashion described for temperature indices. Levels of spatial coherence are similar for both OSR variants and for all three ANN variants, but are very different between the two different modelling methods. Regions of relatively high skill are, however, common to all five variants. The highest levels of skill can again be found for those areas most influenced by the Atlantic, and westerly flow across the western Greek coast. In addition, for most indices and the ANN variants, skill is at least moderate for south eastern Spain, a region subject to strong easterly flow (Chapter 3.4.2). Rainfall results show a similar level of skill and pattern in autumn to that for other seasons, with the greatest similarlities between winter and autumn. The Linear ANN variant (Fig 4.49) is the best choice for most regions. Where it provides little skill, so do other ANN variants, and the most appropriate choice is the BOSR model (Fig 4.45). However, all five variants produce negligible skill for the south east of Greece for PQ90, PF90, and PN90, as for summer.

4.4 Discussion of results

4.4.1 Recommendations for use

This chapter discusses and explores the performance of two very different climate modelling techniques and five variants on those techniques. Between them they encompass linear and non-linear methods, differing methods of predictor selection, and differing forms of application. From the results of the previous sections a number of

recommendations for the appropriate use of these methods have been identified and are summarised below.

OSR recommendations:

- A methodological advantage lies in the OSR method, in that it may be run on multiple indices and stations without the need for consideration of multi-collinearity (Section 4.2.2).
- However, the components of variance based method provided by the OSR model results in reconstructed data with a restricted variance. This is more evident for precipitation indices than temperature indices (Section 4.3.2).
- Furthermore, the OSR method offers very poor performance for rainfall indices for all seasons and regions (Section 4.3.3), this may be due to the fact that rainfall behaviour across the basin varies over very short distances (Chapter 3.4.2), providing a very large number of components of variance, all of which apply only for relatively small regions.
- OSR performance is good for rainfall indices only where they are all influenced by one or two major forms of strong circulation and the rainfall response is spatially coherent (Jones *et al.*, 1999b). In the case of the target area this region is where Atlantic sourced flow is dominant (Chapter 2.2, Chapter 3.4.2, Chapter 4.3.3).
- The use of pre-selected groups of predictors did, in some cases, improve model skill. It is evident that too many predictors may over-complicate the OSR method and reduce accuracy (Section 4.3.3). However, if the predictor set varies on too small a spatial scale the advantage gained from an immunity to multi-collinearity of spatially distributed indices (e.g. average temperature for multiple locations) may be lost. The balance between a need for an appropriate set of predictors, and the need for a uniform set for a given region is a difficult one to assess.
- Performance is not consistently better for one OSR variant or the other, but instead varies based on season, region, and in some cases, the climate index under consideration (Section 4.3.3). It is not always the case that a pre-selected predictor set improves model performance.

ANN recommendations:

- ANN methods provide several major methodological advantages. They make no prior assumptions about the distribution of the data, consider each predictand individually, may self-select the relevant predictors, and are capable of non-linearity (Section 4.2.2).
- The basis function width required of a given Neural Network model is largely driven by the range of the data considered, and should be larger than the range of data by an order of magnitude. Thus, for normalised data, a width of around 10 is required (Section 4.2.3).
- For some kinds of data, 'early stopping' (Section 4.2.3) may allow for a reduction in Neural Network processing time with little reduction of skill.
- Further, despite the 'no prior judgement' advantage, temperature indices still gave better results than precipitation predictand indices when modelled with the ANN method (Section 4.3.2, and 4.3.3).
- ANN variants generally both over and under-estimate variance for all predictand indices. However, levels of accuracy are slightly higher regarding the variance of the rainfall predictands when compared against OSR methods (Section 4.3.2).
- The spatial coherence of results is much greater for neural network methods when applied to temperature than for the OSR methods. However, this is almost certainly a reflection of the fact that each neural network run optimises its results for every station, whereas the OSR methods optimise their results based on the replication of variance across all stations (Section 4.2.4). Spatial coherence of results is greater for OSR precipitation applications only because model skill is generally very low, and it is greater for the ANN temperature applications only because model skill is generally very high (Section 4.3.3).
- This may mean that dynamically driven regions of high or low performance are lost or 'tuned-out' by the neural network methods, but accurately reproduced by the OSR methods. Where this detail was not lost through very high or very low performance it can be seen, however, that both models show similar (if not entirely common) patterns in their performance.

- With certain exceptions (see below) relationships between the ANN variants remained consistent, if one performed better in a given season, so did the other two, where one variant performed relatively well, the same patterns could generally be seen in the performance for all variants.
- One exception to the above involves stations that consistently (between indices, and in some cases seasons) returned zero or near-zero skill for ANN model application. These regions were generally more numerous for rainfall indices than temperature indices, and more numerous for non-linear networks than the linear variant.
- For the majority of instances, however, non-linear basis function ANN variants did not provide a greater level of skill than the linear variant, nor did they dramatically improve the estimation of predictand variance (Section 4.3.2 and 4.3.3). In almost all cases the Linear variant produced the greatest degree of Neural Network model skill.

Overall recommendations:

- Overall, the linear neural network produces the greatest level of skill for all seasons and most indices.
- However, as there are some stations for which skill is consistently poor, e.g.: Bastia, Clermont-Ferrand (France), Sabiñanigo, Leon (both Spain), Pegoes, Mora (both Portugal), and Nis (Yugoslavia) for summer, and Sabiñanigo and Nis for winter, and as these stations are also inevitably poorly represented by the other neural network models as well, in some cases the OSR approach may be more appropriate (Section 4.2.3).
- The choice of appropriate OSR approach may vary by season or region.

4.4.2 Model weights and the causes of extreme climate behaviour

When examined in fine detail, the structure of the models discussed in this Chapter can be seen to vary from application to application. Each model run may give individual predictors different 'weights', i.e. may assign a greater degree of importance to one predictor over another. Where a skilful model consistently considers a specific

predictor important for the replication of a specific predictand, trends of the two variables are significant and physically consistent (Chapters 3.4.4, 3.5.4, 3.5.7, and 3.7), correlations between them are significant (Chapter 3.6), and physical, synoptic, processes can be identified that support purely statistical relationships (Chapters 2.2.3 and 2.3.1), a weight of evidence for causality may be presented.

From the above section (4.4.1) it can be seen that in the majority of cases the Linear neural network variant most skilfully reproduces the climate predictand indices under consideration (Chapter 3.4.1). The weights for this model have been extracted and plotted in Figures 4.54-4.59. These weights are presented for each predictand, for each season, and for each index for the western, central, and eastern regions.

The western Mediterranean

A link that satisfies all of the above conditions exists between the NAO and western precipitation indices for both winter and autumn. A link is also suggested between temperature indices and the NAO by the literature (Chapter 2.2.3), correlations (Chapter 3.6), and the regions of greatest skill (Section 4.3.3). This link is found for frost days and very high temperatures during winter (two indices which do not correlate well), and for all temperature indices except frost days during autumn. The dry days index shows no link to the NAO, but has been shown to vary in a different fashion to other precipitation extremes (Section 4.3.3). Although a link exists between high temperatures and consecutive dry days, it is at its weakest during winter (Tables 3.5-3.8).

Also evident for winter, are links between both the 4th SLP principal component (an extended Icelandic low component) and 4th SHM principal component (indicative of flow across the Mediterranean from the mid-Atlantic) and cold temperatures (TMIN, TN10). For autumn SLP PC4 (again indicative of an extended Icelandic low), can be linked to cold temperatures (TMIN, TN10, TNFD). It seems likely from the above, and previous discussion (Chapter 2.2.2), that Atlantic flow is highly important for western Mediterranean extreme climate behaviour (particularly extreme cold and rainfall), and that different expressions of that flow may be linked to different forms of extreme climate. For autumn, links can be seen between the 3rd Z500 component (representative

of the NAO centres of action), and extreme rainfall (PF90 and PCDD) in both correlations and weighting.

With certain exceptions (e.g. dry days during winter), the NAO is a powerful driver of extreme climate (particularly precipitation) behaviour in the western basin for winter, autumn, and possibly spring. Furthermore, the more northerly Atlantic centre of action (the Iceland low) persistently shows an influence upon cold extremes when it is extended toward the Mediterranean.

During winter, significant links have been discussed (Chapter 2.3.2) and shown (Chapter 3.5.3) between the MOI and the NAO. From the weightings shown in Figure 4.46 it can be seen that modelled links between the MOI and winter climate are strongest for prolonged heat (HWDI) and indices of extreme cold (TMIN, TN10, TNFD). It has been suggested in the literature (Chapter 2.2.3) that during winter the MO functions as an extension of the NAO, and certainly the results shown here are consistent with a link between North Atlantic pressure systems and the MO. The MOI also appears as a western Mediterranean predictor for extreme cold (TN10) during summer. Correlations exist between these variables, and Maheras *et al.* (1999b) state that in one of its two (non-seasonal) phases the MO may draw cold air down from Europe. Maheras and Kutiel (1999) show that the MO may bring either unusually high or low temperatures, depending on its phase.

The Atlantic Blocking Index (ABI) is shown by model weights (Fig. 4.54) to be important for prolonged cold (TNFD) during the western Mediterranean winter, and for extreme high temperatures during summer (TX90), in both cases correlations (Chapter 3.6) offer supporting evidence. Dynamically, as the ABI measures the strength of a blocking pattern in the Atlantic, and therefore the relative strength (or weakness) of westerly flow, it seems reasonable to suggest links for both of these factors, as they are dependent on stationary or near-stationary circulation systems (Chapter 2.4). Different forms of European blocking are important for both summer high temperatures (EBI, TMAX and TX90) and winter dry spells for very similar reasons. During winter, SLP1 represents a deep European depression and a resulting Winter Dry weather type (Corte-Real *et al.*, 1998a). Both model weights (Fig. 4.55) and correlations indicate a link

between the 1st SLP component and winter PCDD. The 3rd Z500 component also represents a form of blocking in both winter (Euro-Atlantic blocking) and autumn (Siberian blocking) and is shown by both correlations and model weights to be important for temperatures during both seasons and precipitation during autumn. In the latter season, the effect upon temperature and precipitation is likely to be mitigated by the NAO circulation through an extended Icelandic low as correlations between the 3rd Z500 component and the NAO index are highly significant (Table 3.12), and centres of action are similar for both.

During spring the most heavily weighted predictor is the NSCP for all predictand indices. Significant correlations exist between a large number of predictand indices and the NSCP (Chapter 3.6), but the available literature does not discuss impacts of the NSCP upon the western basin climate (Chapter 2.2.3). Also showing strong links with spring climate, however, is the 5th SLP component, which represents pressure centres roughly analogous to those that create the NSCP. In addition, for winter, the 2nd SLP component (an east Atlantic pattern / north Atlantic depression component, also with similar pressure centres) is weighted heavily for all extreme temperature indices. The NSCP is cited as affecting both temperature and precipitation (Kutiel *et al.*, 2002), but until now its influence has only been explored for the eastern Mediterranean climate (Kutiel and Benaroch, 2002). A reasonable amount of statistical evidence is presented here for influence in the western Mediterranean, and the NSCP's potential as a major driver for spring-time (and possibly winter) western Mediterranean climate should be explored further.

African influences can be seen during winter, summer, and autumn. A link is suggested by weights, correlations, and literature (Section 2.2.2) between north east African influences and climate in the western basin. African uplift (indicated by SHM PC4) shows links to both very high temperatures (TX90) and high levels of rainfall (all indices) during winter. For summer, the same African extension (SHM1) can be linked to increased high temperatures (TX90) in the west, as suggested by Colacino and Conte (1995). An autumn diversion of Atlantic flow to the north may allow for the intrusion of African influence, as links between the 3rd SHM component and both HWDI and PCDD are also apparent. African influence is not, however, the only potential driver of

prolonged summer hot and dry conditions, as small pressure or humidity gradients across the basin (Z500 PC7, and SHM PC4) can also be connected to persistent dry days (PCDD) and high temperatures (TMAX).

Although the SOI shows model weights for summer and autumn temperature, there are very few significant correlations, and for temperatures in the western Mediterranean, there is very little in the way of supporting literature.

The central Mediterranean

For the central Mediterranean, links between extremes of climate and Atlantic influence are noticeably less supported by the evidence assessed here, particularly for winter, when few relationships occur in the literature, correlations, trends, or model weightings. It is generally suggested by the literature that strong relationships with the NAO are confined to the western end of the Mediterranean (Chapter 2.2.3). The results shown here support previous work. Statistical links can only be found for winter with winter warm spells (HWDI), and for summer with both high temperatures (TMAX, TX90, HWDI), and the volume of rainfall (PREC), all of which have been shown to exhibit both strong weightings and significant correlations. For autumn, low temperatures (TMIN, TN10) can also be statistically linked to variations in the 2nd Z500 component (as for the west). The latter of these two connections seems plausible given the discussion of western Atlantic influence given above (regarding autumn cold weather and the 2nd Z500 component), but the summer linkages need further investigation.

Far more important in the central basin than the west, is the Southern Oscillation. For winter, links have been shown for low temperatures (TMIN, TN10, TNFD) and the frequency of rainfall (PINT, PF90, PCDD). Although little in the literature details Italian SOI influence, teleconnections have been found with both eastern coastal Spanish rainfall (Rodo and Comin, 2000), and Turkish precipitation (Kadioglu *et al.*, 1999).

Blocking activity is as important in the central Mediterranean as it is in the west, although less so during winter and summer, and more so during autumn and spring.

Euro-Atlantic blocking (indicated by Z500 PC3) displays influences in the central Mediterranean similar to those in the west, with winter average and low temperatures (TMIN, TN10), and summer high temperatures (TMAX), and rainfall amounts (PQ90, PCDD). Similarly the Winter Dry pattern (indicated by SLP1) can be linked to persistent winter temperature extremes (TNFD and HWDI), and a lack of rainfall (PCDD). However, Atlantic pressure centres mainly influence spring climate, in terms of temperature indices (TAVG, TMIN, TMAX, TX90), the number of rainfall events (PN90), and periods without rainfall (PCDD). More persistent conditions (TNFD, HWDI) are also influenced, although more by blocking in the north Atlantic (SLP PC2), than elsewhere. European blocking shows links to the same climate indices during spring. During autumn European blocking links are evident with both high temperatures, and the occurrence of heavy rainfall (PQ90, PF90, PX5D).

As for Euro-Atlantic Blocking, the NSCP shows similar influence in both the west and central Mediterranean, for all temperature indices, mostly during spring. The 5th component of spring Z500 and the 1st component of autumn Z500 both strongly resemble the NSCP pattern and can be linked to rainfall indices during spring in the former case and autumn in the latter. It may be that the influence of the NSCP is wider than previously suggested.

African influences are less supported by statistical evidence for the central basin than the west- the only factors that repeatedly appear in the correlations, trends, and weightings are winter SHM PC2 and PC4, and their effects upon extreme rainfall. Only in the latter case do loadings plots (Fig. 3.29) suggest a dynamic link, and it is the 2nd component of SHM that shows the strongest weightings for both extreme high temperatures (TMAX, TX90) and the majority of rainfall indices (PINT, PF90, PCDD, PN90, PX5D).

As for the western summer regime, central Mediterranean high temperature extremes may also be linked to low pressure gradients across the Mediterranean and continental pressure systems. The 5th and 7th summer components of Z500 can be linked to average (TAVG) and high (TMAX, TX90) summer temperatures, as can the 2nd summer component of SLP, which suggests a weakened Atlantic circulation. For the

central basin, however, the extended warm season is evident in the fact that Z500 PC6 (also representing a small pressure gradient across the Mediterranean) also affects high temperature (TMAX and TX90) indices.

The eastern Mediterranean

For the eastern basin, the Atlantic influence on extreme climate evident in west and central basin is almost entirely absent. Although teleconnections have been found in previous studies in both Turkey and the Middle East (Chapter 2.2.1), any Atlantic influence of the Greek climate is generally not apparent in this study. For winter, the only Atlantic influenced predictor to show a substantial statistical link is the 1st Z500 component (Fig. 3.27), which is highly significantly correlated with the NAO, but also displays high levels of variance throughout the Mediterranean basin and the Arctic circle. As for the NAO in the western basin, links can be found between Z500 PC1 and the majority of winter rainfall indices. As for Atlantic influence in the central basin, links can also be found between Z500 PC1 and indices of extreme high temperature (TMAX, TX90) during winter.

During autumn, however, both the SOI (for all indices but those representing extreme cold), and the MOI (for all indices, but more so for very high temperatures than very low temperatures) have been found to be much more important (Figs. 4.58 and 4.59). As mentioned above (and in Chapter 2.2.1), links between rainfall and the SOI for the eastern basin are not new to this study, although the suggestion that it may have a strong effect upon extreme climate for the Balkan peninsula is a novel finding. The MOI is partly defined by pressure changes in the eastern Mediterranean, so an empirical link may be easily justified, and is referred to by Maheras *et al.* (1999b). Also representing the large-scale pressure regime in the east, a strong Asiatic low (Z5007) may be important for spring dry days (PCDD) and temperatures (Chapter 2.2.2, 2.2.3), although in the latter case its influence is not as important as that of the MOI.

The NSCP is generally cited as having a strong affect upon the eastern Mediterranean (Kutiel and Benaroch, 2002, Kutiel *et al.*, 2002). In this study it has been found to possess links with autumn extreme cold (TN10 and TNFD) and heavy rainfall events (PN90), and summer extreme heat (TX90). The former combination of heavy

rainfall and low temperature, is suggested by Kutiel *et al.* (2002) as a reflection of the negative phase of the NSCP.

For the East Atlantic, blocking shows little to no influence, much as Atlantic circulation is found to be unimportant. As for the central Mediterranean, however, European blocking shows strong statistical links for spring. During spring the EBI influences all precipitation indices, although the heaviest weightings/ strongest correlations are those for heavy and intense rainfall (PQ90, PINT), and all temperature indices except those concerning duration of event. Persistent extremes are instead linked to the 2nd Z500 component, north eastern European blocking, which may also influence spring rainfall. The 3rd Z500 component (indicative of a deep European depression) shows statistical links with the same indices. As the blocking of meridional weather systems is obviously important for the spring climate, it also seems reasonable that continental pressure systems (SLP PC1) might influence spring temperature and rainfall, as suggested by the statistical links drawn from this and the previous chapter. In seasons other than spring, summer depressions to the north of Europe (indicated by SLP PC3) may be linked to the number of rainfall events (PN90 and PCDD) and measures of temperature magnitude. Although there is no evidence for influential blocking action during winter, a low variance humidity field (SHM PC6) is very important for both winter temperature and precipitation indices, although less so for frost days (uncommon for most of the eastern stations) or very high temperatures, rainfall intensities, or persistent dry days.

Much like the climate of the western basin, the east is heavily influenced during summer by incursions of African air. This form of circulation is best represented by the SHM components (Figs. 3.35), which in the east also affect spring climate. During summer, incursions of southern sub-tropical air from both the west and east can be linked to rainfall indices and low summer temperature indices (TMIN, TN10), and a flow of dry air from northern Africa to Turkey (SHM PC2), shows a possible effect upon heatwaves (HWDI). During spring, the SHM PC3 (Asian influence) can be linked to dry days (PCDD), and north east African uplift (indicated by SHM PC4) may influence most forms of spring rainfall and extreme low temperatures (TN10).

4.4.3 Conclusion

In this chapter two different statistical downscaling methods have been applied to Mediterranean climate extremes. In both cases, and for all variants of those methods, the extreme climate patterns described in the previous chapter have had a substantial effect in terms of the regions, seasons, and indices, which have been most skilfully modelled:

- Temperature indices are (in almost every case) more successfully modelled than rainfall indices (Section 4.3.3).
- In terms of both skill and variance performance, there is a definite split between most measures of temperature extreme magnitude, and those of extreme duration. Very low temperatures (TN10), however, are replicated with performances closer to the latter group. Generally measures of magnitude are more skilfully and more coherently represented than measures of duration (Section 4.2.3, and 4.3.3). This may be due to the fact that the influences responsible for the latter forms of extreme behaviour are more complex than those for the former (Chapters 2.4.2 and 2.4.3).
- Both techniques are highly sensitive to the season under consideration, although seasonal sensitivities vary between the OSR variants and are largely consistent for the ANN variants (Section 4.3.2). The number of stations that were very poorly reproduced (i.e. generated negligible skill), however, were generally higher for summer and autumn than winter or spring (Section 4.3.3).
- Taking regions of negligible skill into account, it is difficult to pick (in terms of both the average level of skill produced, and the range of performance) one season for which each variant gives the best performance across the entirety of the area considered in this study. It is much easier to select certain indices that are well represented for all variants in any given season. These are generally reflective of the dominant form of seasonal climate, with the exception of summer low temperatures (TMIN), which are generally well modelled, and Mediterranean dry days (PCDD) and rainfall intensity (PINT), which are relatively well modelled for most seasons and variants.

- Autumn and eastern performance results are less consistent between model variants than other seasons or regions. In the former case this seems likely to be a reflection of a more complex range or seasonal climatic behaviour (Chapter 3.4.2). In the latter case it is due to a large number of regions where Neural Network models returned negligible skill.
- Autumn and summer rainfall shows particularly poor performance for the south east of Greece for all indices and model variants. It seems likely that either the lack of precipitation in the area is negatively affecting skill (Chapter 3.4.2 and Section 4.3.3), or the appropriate predictors for this region are not included in the models.

Of the two methods applied in this study (in most cases but with certain caveats), the neural network method may be the most successful. The addition of particular forms of non-linearity to such a model may not, however, increase skill or overcome certain inherent flaws (the inability to successfully model certain locations), and may introduce new ones (the inability to successfully model summer climate). In the process of developing a neural network downscaling model, a means of discriminating climate predictors useful for the simulation of climate predictands has been constructed. In this way an additional statistical test for links between hemispheric-scale circulation and station-scale extreme events has been conducted. Although the ability to replicate one set of data from another does not alone imply physical causality, a number of different investigative methods have been employed up to this point, and a weight of statistical evidence accrued, supported by empirical evidence within the climatological literature (Chapter 2.4). Section 4.4.2 discusses potential causes for extreme climate behaviour given the literature reviewed in Chapter 2, shared variance, regions of influence or trends found in Chapter 3, and possible reasons for teleconnection, as discussed in both chapters. Whereas many of these factors are known to influence mean climate, little work up until now has been concerned with whether or not they are appropriate for use in the study of extreme Mediterranean climate, and if appropriate, during which seasons and for which regions. Having considered both the patterns and causes of extreme Mediterranean climate, the next three chapters are concerned with the potential impacts of extreme events on socio-economic sectors of activity that are known to be affected by mean climate.

Table 4.1: Climate predictors used for the reduced predictor set OSR variant. All other models use the predictors shown in Table 3.16. Acronyms used below are shown in Table 3.11, PC predictors (SLP, Z500, SHM) are defined in Table 3.17.

	WEST				CENTRAL				EAST			
	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
NAO	X	X	X	X	X	X	X	X				
SOI				X				X			X	
MOI	X	X	X	X	X	X	X	X	X	X	X	X
NSCP		X	X		X	X	X	X	X	X		X
SHI		X			X	X			X	X		
ABI	X	X		X	X	X	X	X		X	X	X
EBI	X	X	X	X	X	X	X	X	X	X	X	X
SLP	X	X	X	X	X	X	X	X		X	X	X
	X		X	X	X		X	X			X	X
		X	X	X	X	X	X	X	X	X	X	X
			X	X							X	X
		X		X								
Z500	X	X			X	X		X	X	X		
	X			X	X							
		X	X			X	X	X		X	X	
		X	X	X		X	X	X				
		X	X	X		X	X					X
		X	X			X	X	X				
		X	X			X	X					
SHM		X	X	X	X	X		X	X	X	X	X
	X				X		X		X	X	X	
	X	X	X	X	X	X	X	X	X	X		
	X	X	X	X	X	X	X		X	X		X
									X			
	X				X				X			

Table 4.2: Model variants selected as a part of sensitivity testing.

Method	Model variant	Notes	
OSR	BOSR: Basin Predictor Set	All available predictors, applied for the whole region, one index of extremes modelled at a time.	
	ROSR: Regional Predictor Set	Seasonally and regionally selected predictors, applied to each sub-region (west, Central, East), as above.	
ANN	LBFNN: Linear Neural Network	Neural network with no preconditioning	12 nodes, Basis width =10, Constant starting seed.
	GBFNN: Gaussian Basis Function	Gaussian preconditioning signal	
	TPFNN: Thin Plate Spline Basis Function	Thin plate spline preconditioning signal	

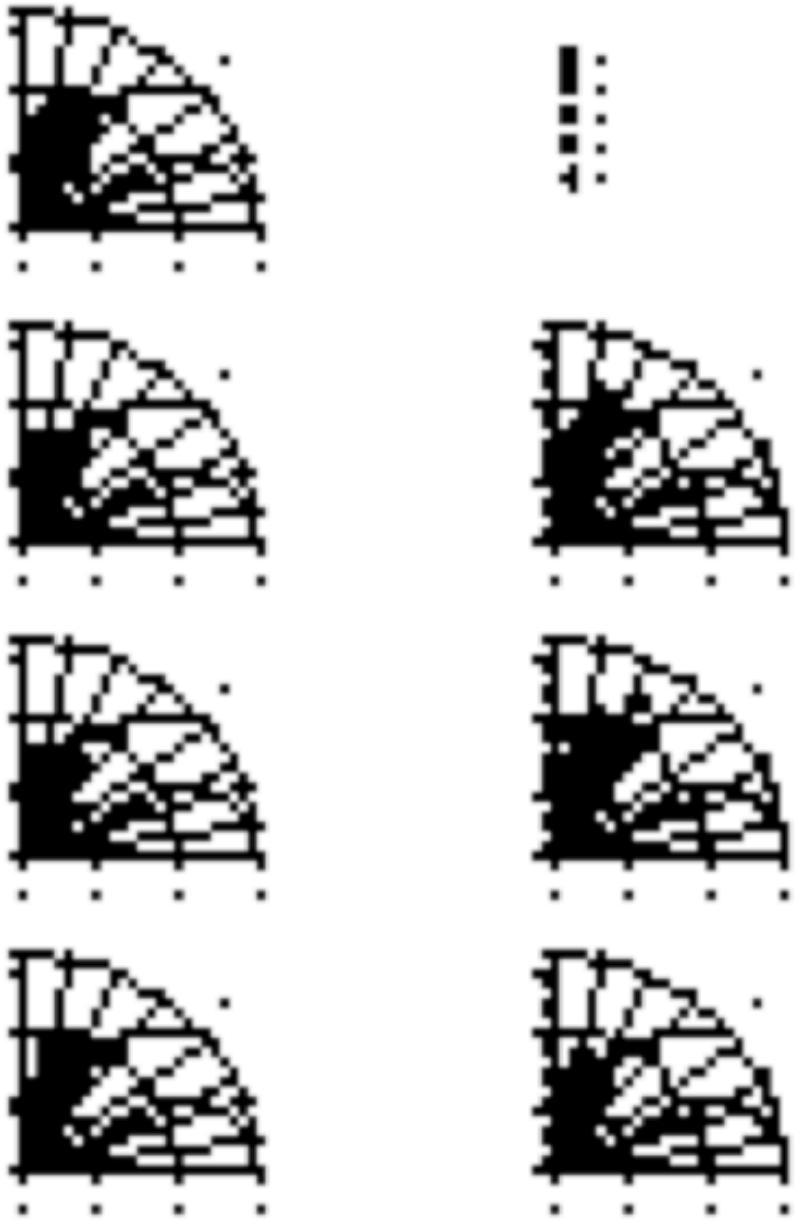

Fig 4.4

Figure 4.4: BOSR temperature predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

Fig 4.5

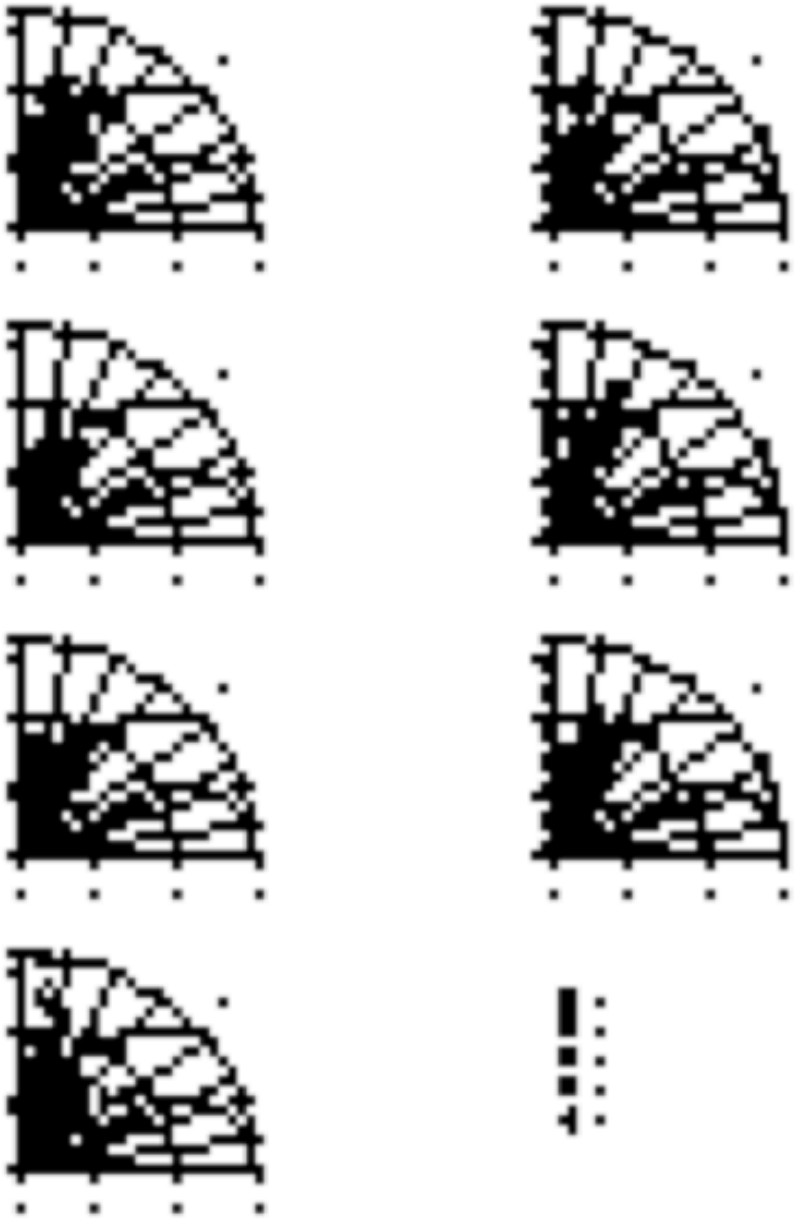

Figure 4.5: BOSR precipitation predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

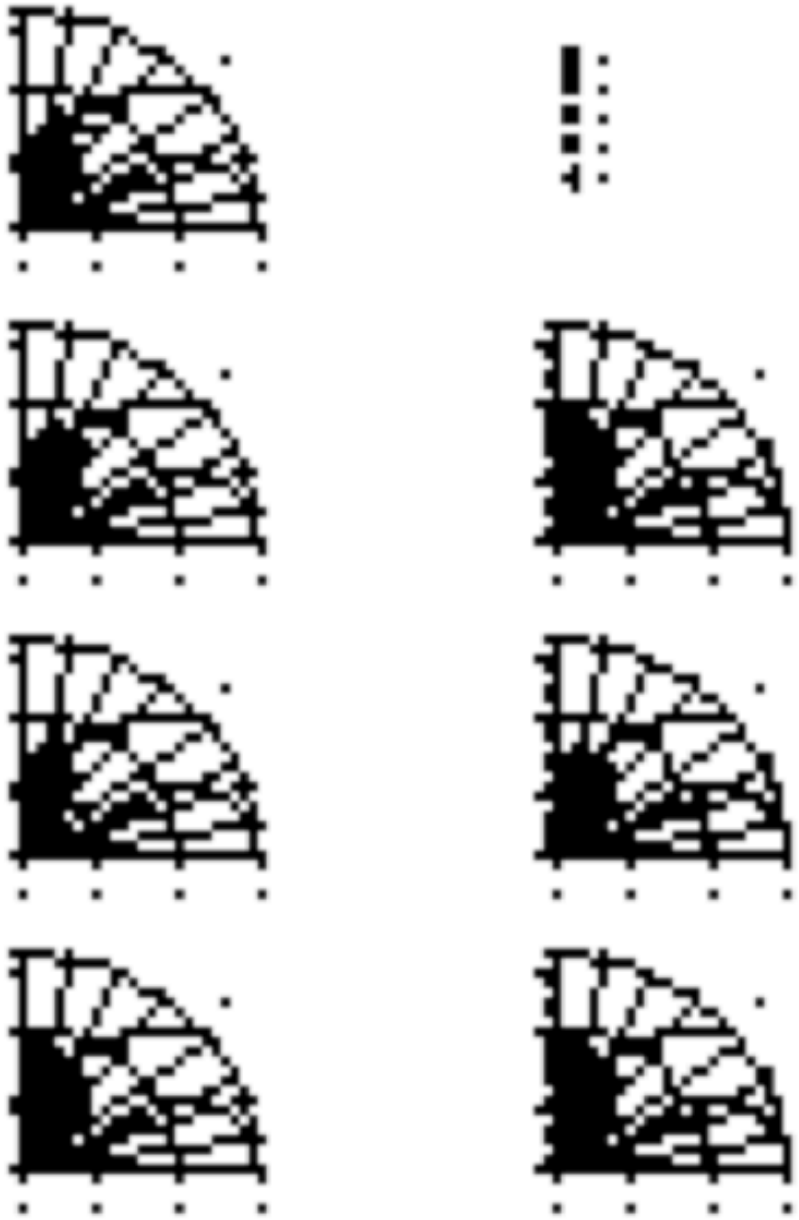

Fig 4.6

Figure 4.6: ROSR temperature predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

Fig 4.7

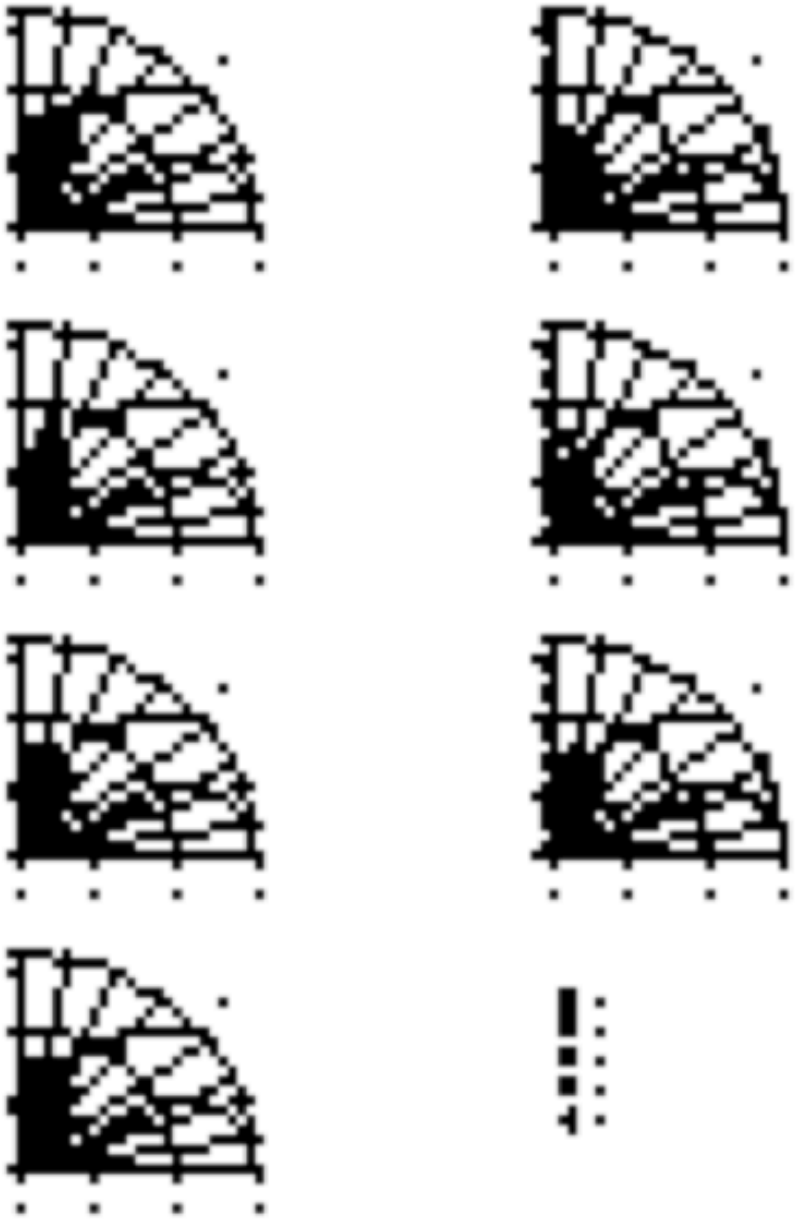

Figure 4.7: ROSR precipitation predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

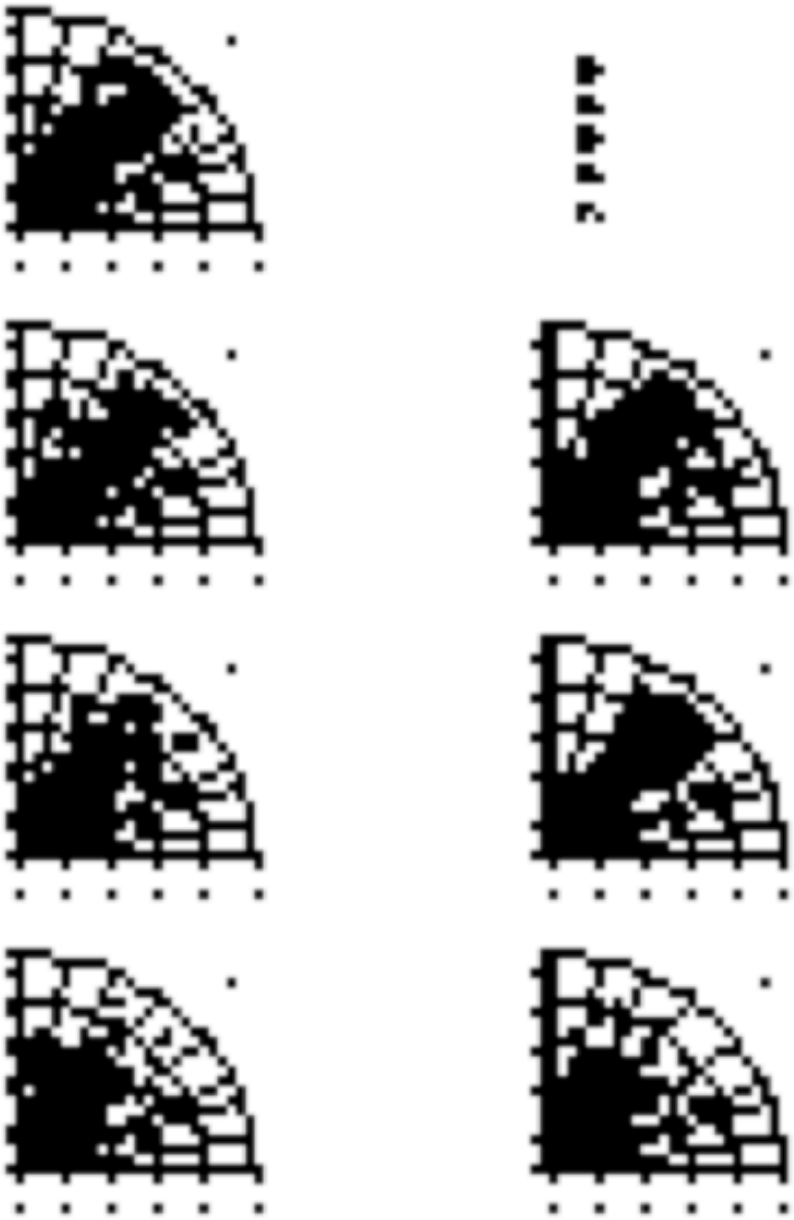

Fig 4.8

Figure 4.8: LBF ANN temperature predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

Fig 4.9

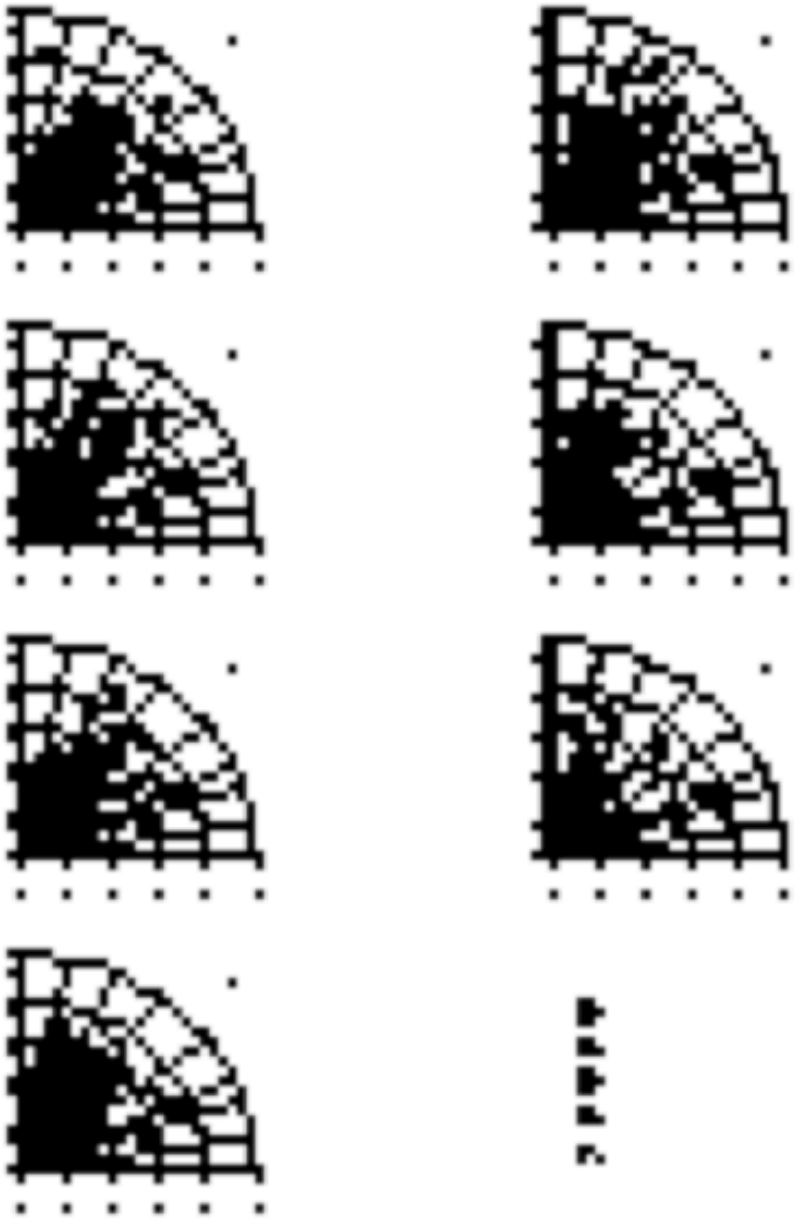

Figure 4.9: LBF precipitation predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

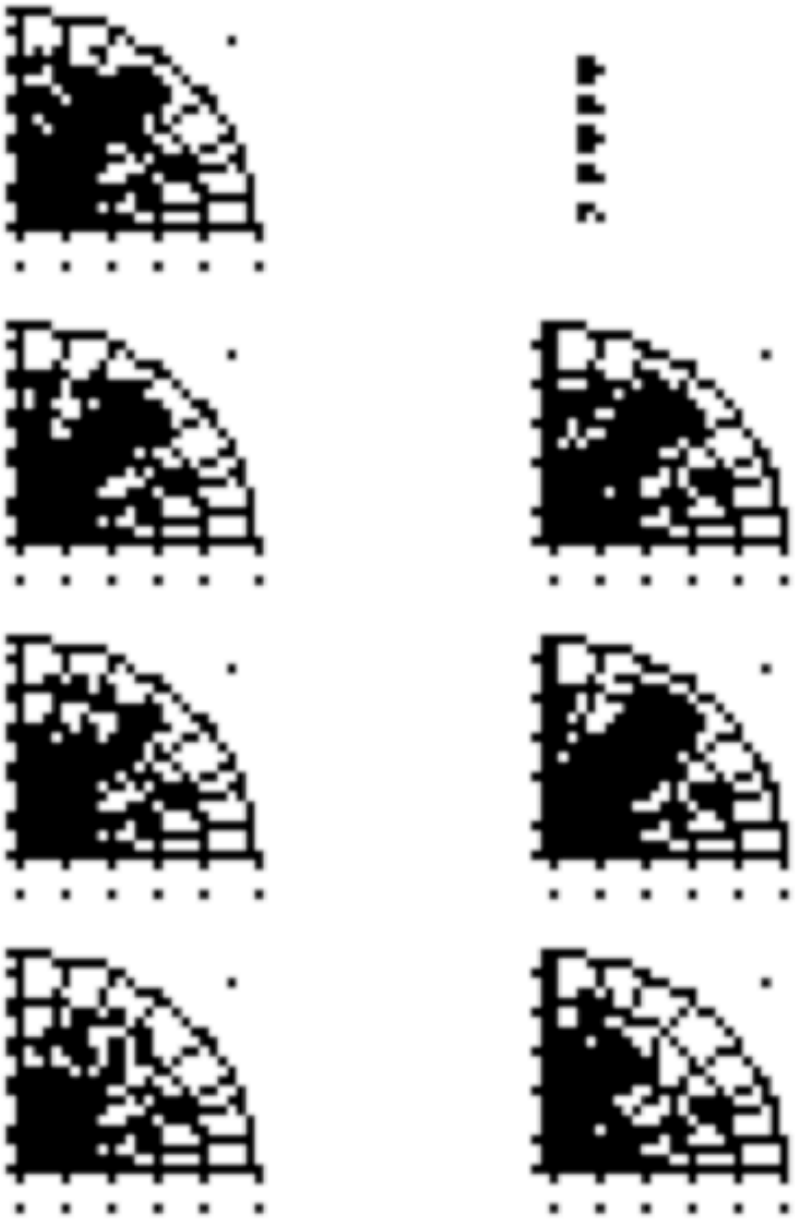

Fig 4.10

Figure 4.10: GBF temperature predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

Fig 4.11

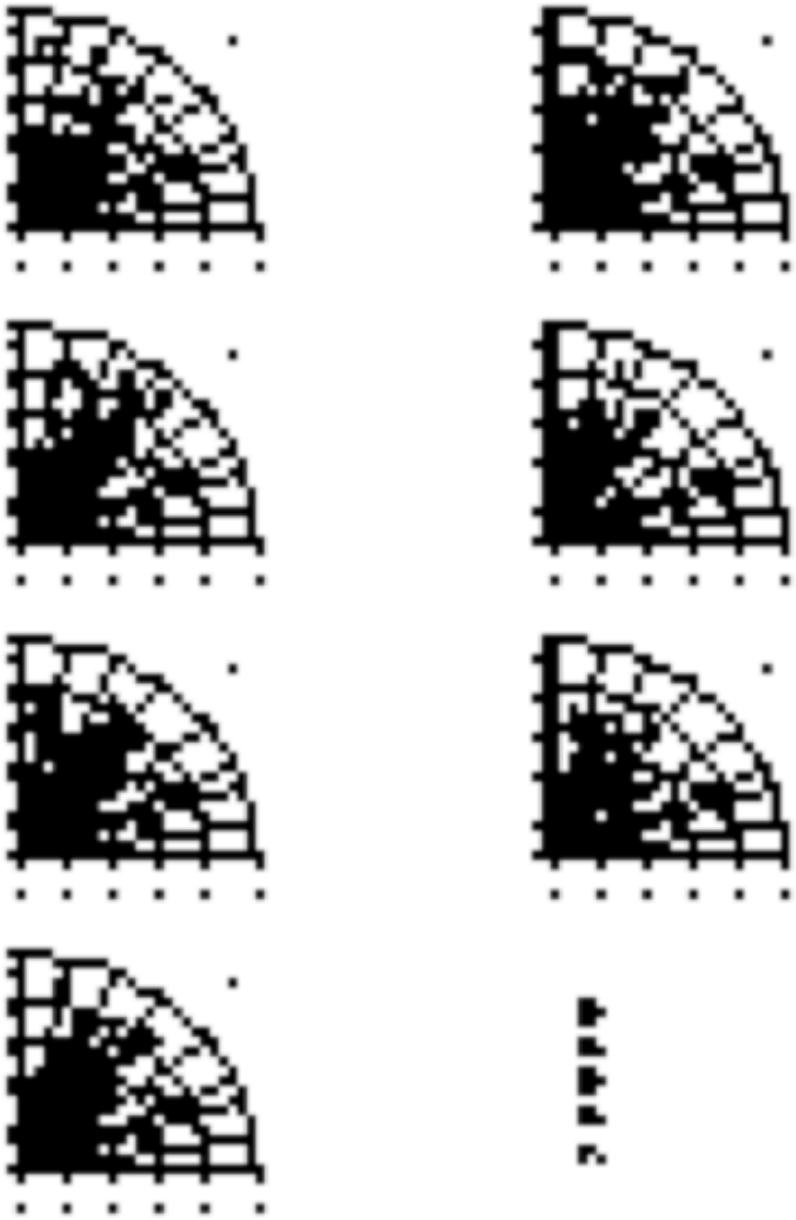

Figure 4.11: GBF precipitation predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

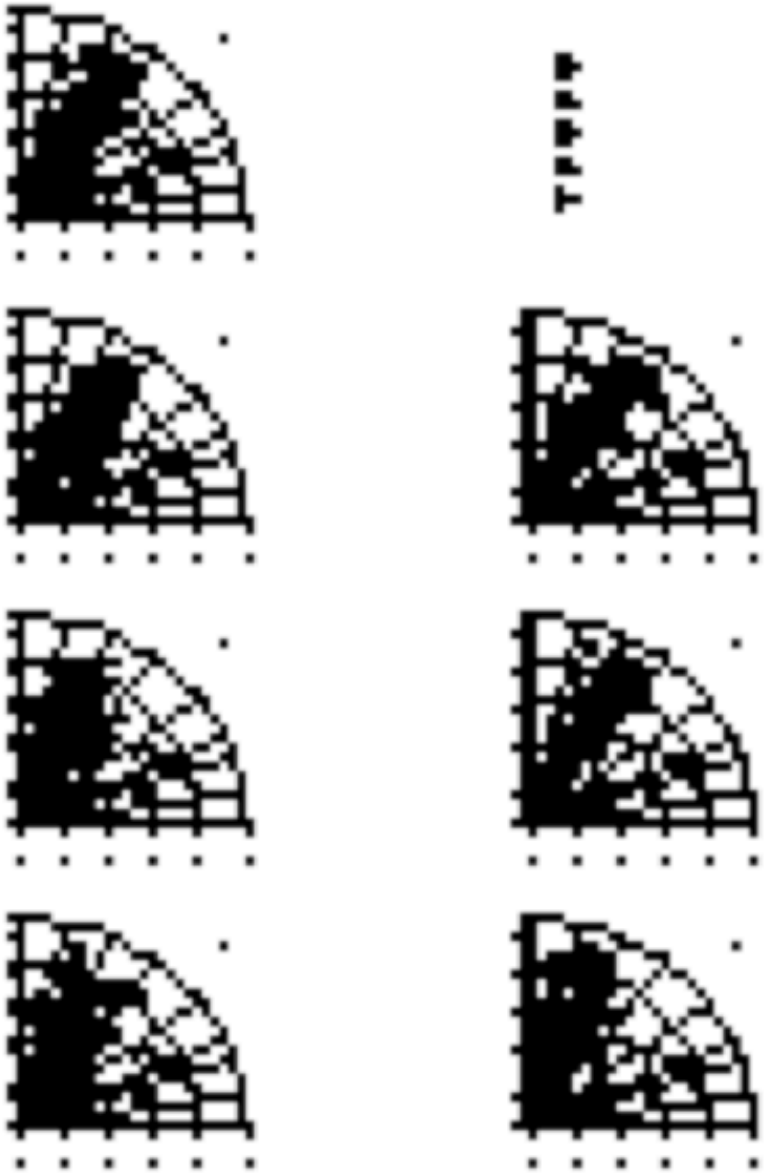

Fig 4.12

Figure 4.12: TBF temperature predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

Fig 4.13

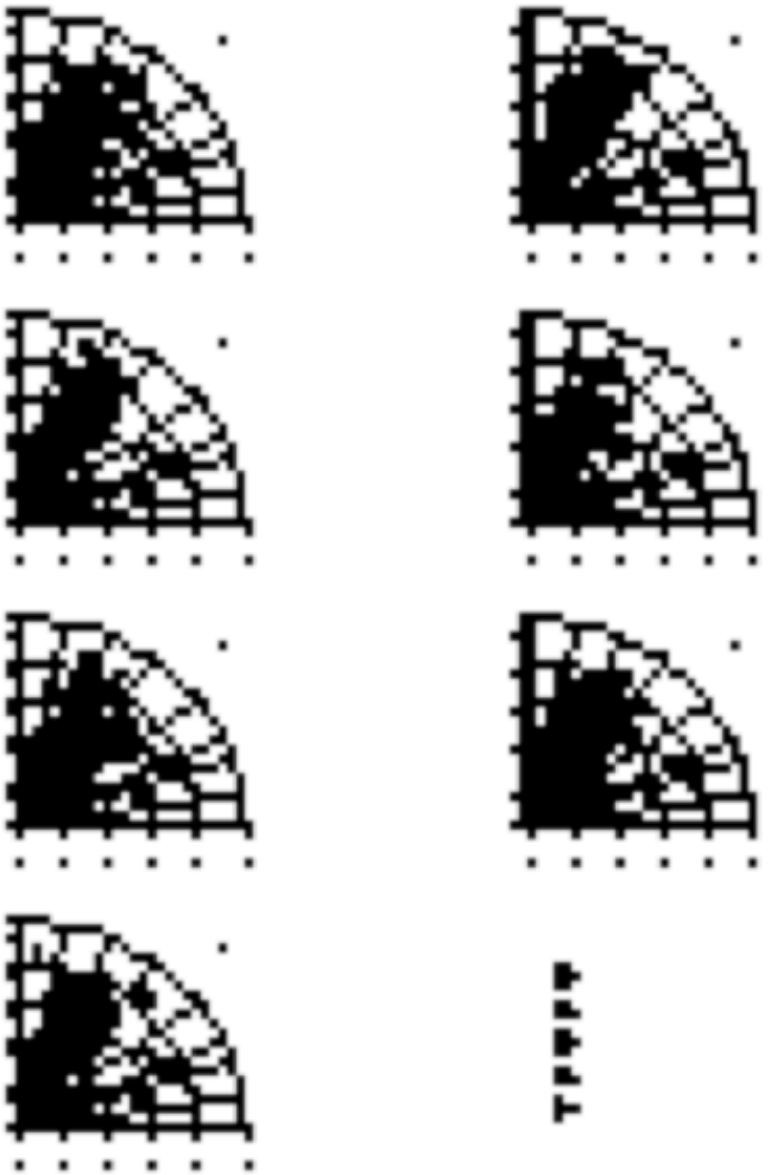

Figure 4.13: TBF precipitation predictand index Taylor diagrams (Section 4.3.1). Representing the estimation of variance (V) on the radial axis and the correlation coefficient (r) on the angular axis for every station, colour coded by season.

BOSR:

ROSR:

LBF:

Fig 4.14

GBF:

TBF:

Figure 4.14: Seasonal box and whisker plots of model skill (r) distribution (all stations) for all climate model variants by predictand index. Seasons are, from left to right, winter (dark blue), spring (green), summer (orange), and autumn (light blue).

Fig 4.15

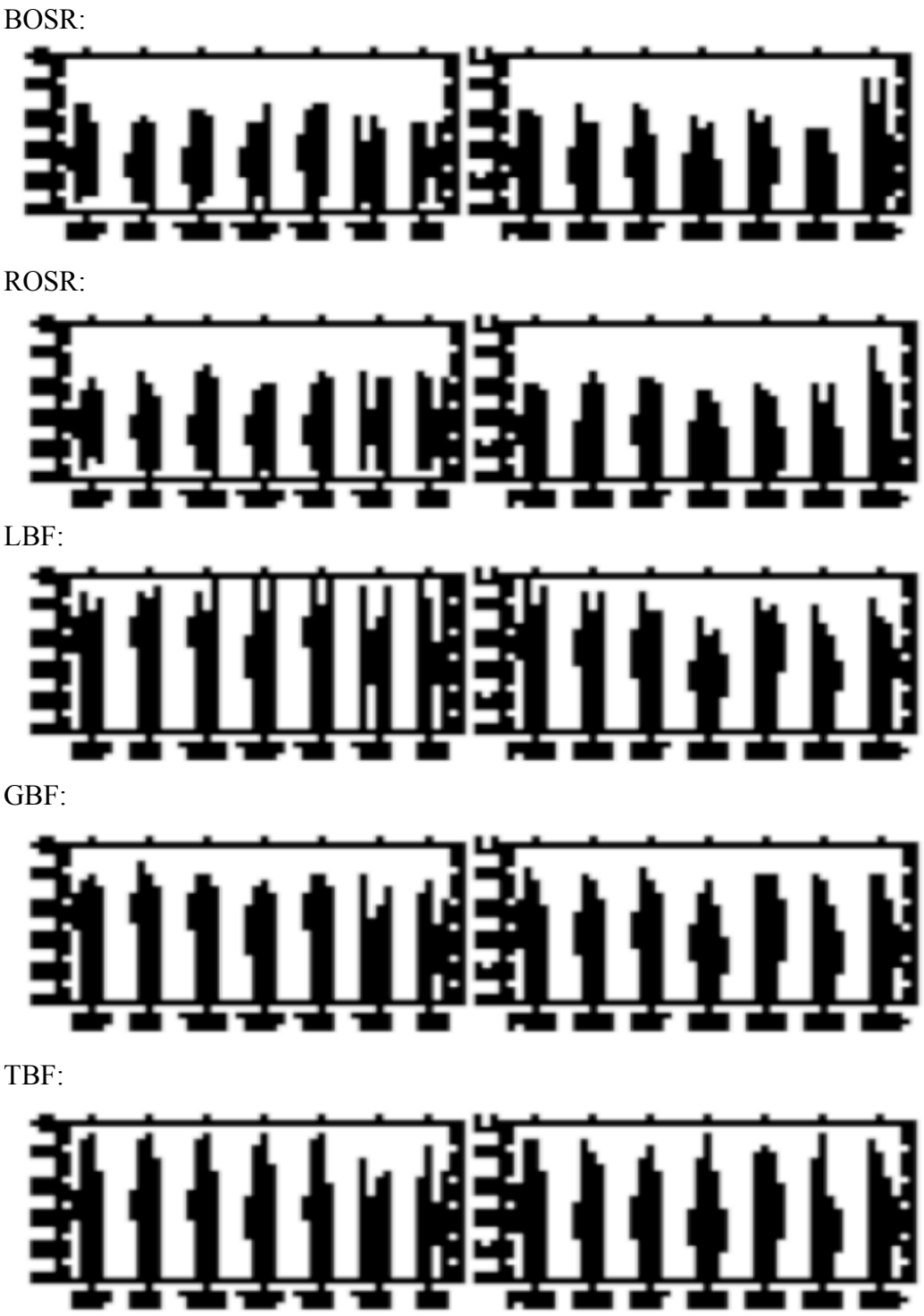

Figure 4.15: Regional box and whisker plots of model skill (r) distribution (all relevant stations, all seasons but summer) for all model variants. Stations are western (blue) between -10° and 6° latitude, central (green) between 6° and 20°, and eastern (orange) between 20° and 30°.

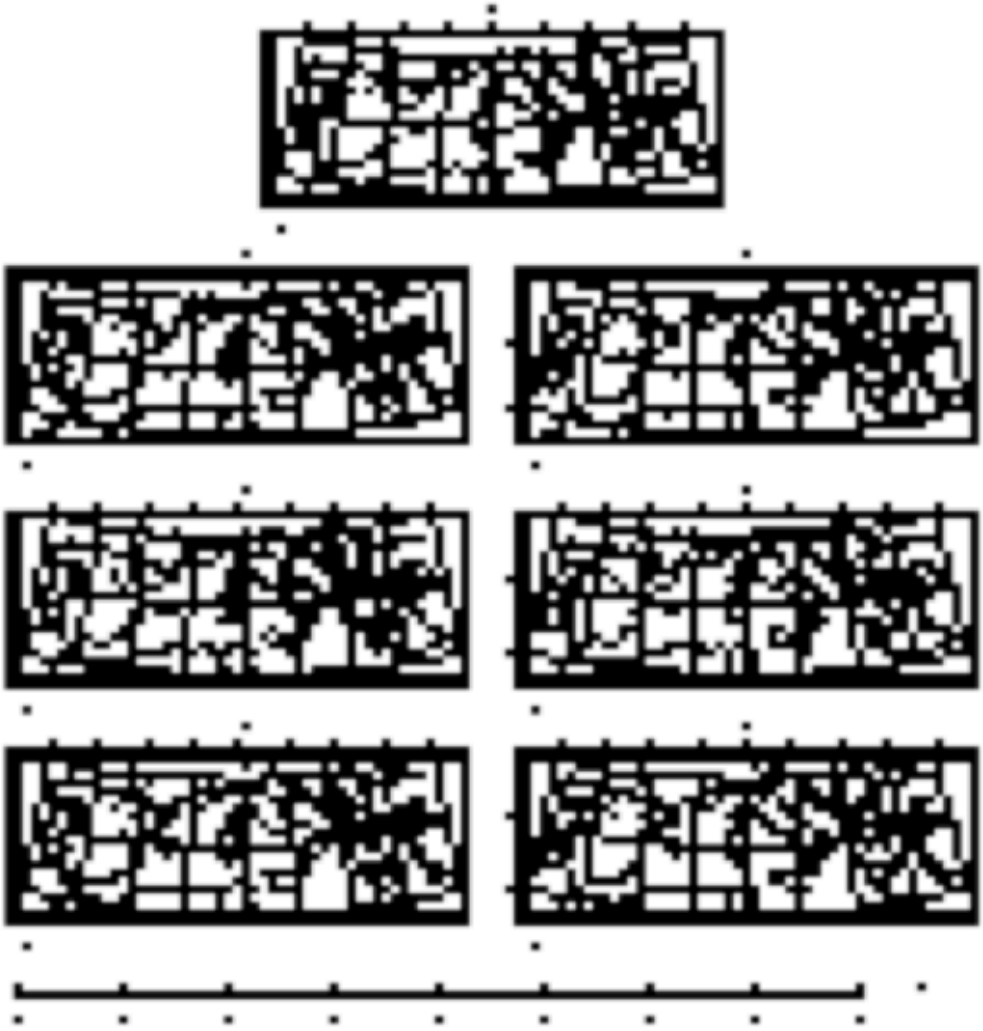

Fig 4.16

Figure 4.16: BOSR winter temperature performance, calculated as model skill (r).

Fig 4.17

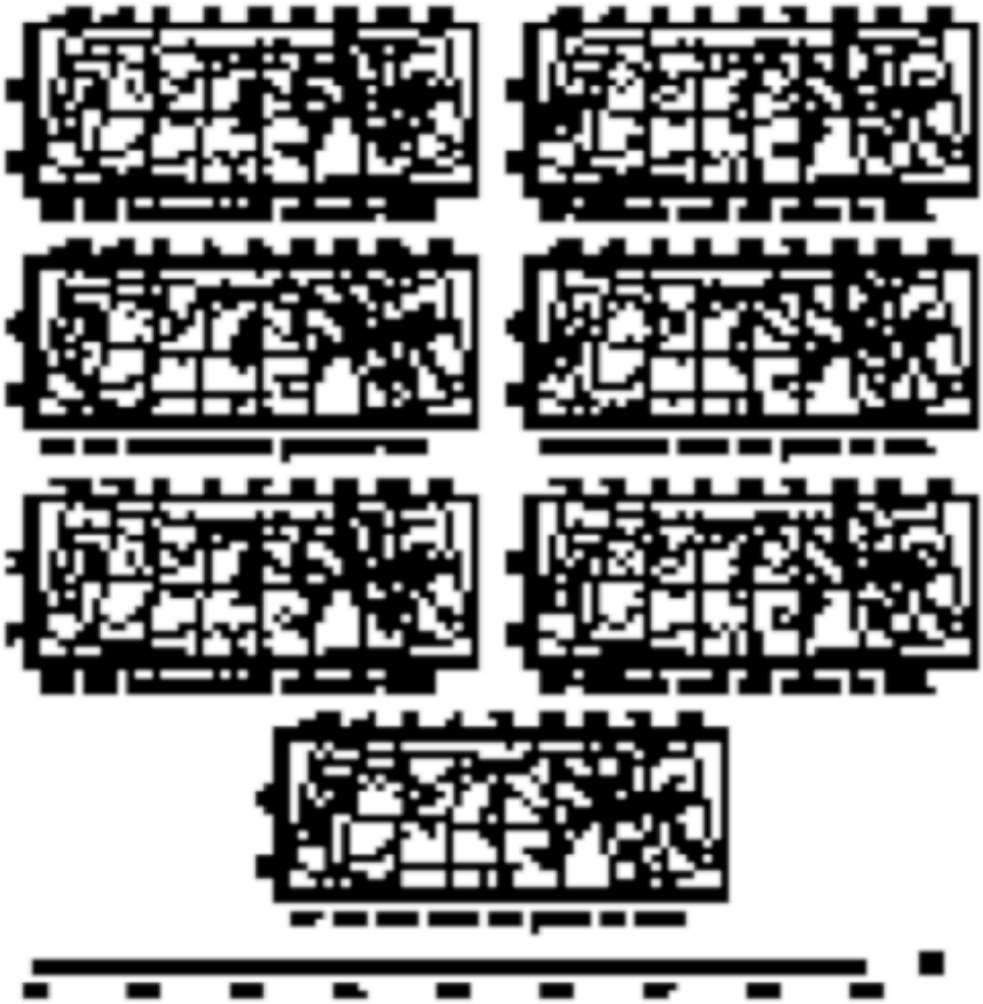

Figure 4.17: BOSR winter precipitation performance, calculated as model skill (r).

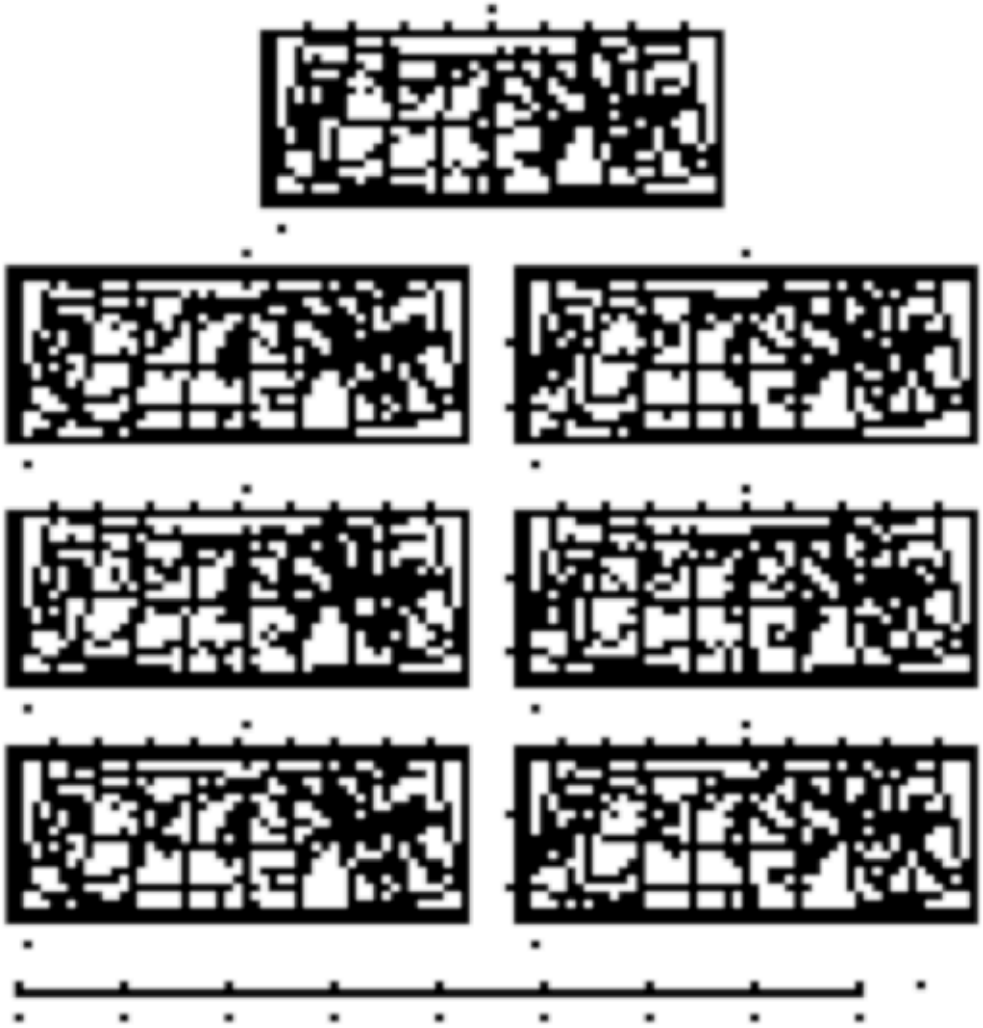

Fig 4.18

Figure 4.18: ROSR winter temperature performance, calculated as model skill (r).

Fig 4.19

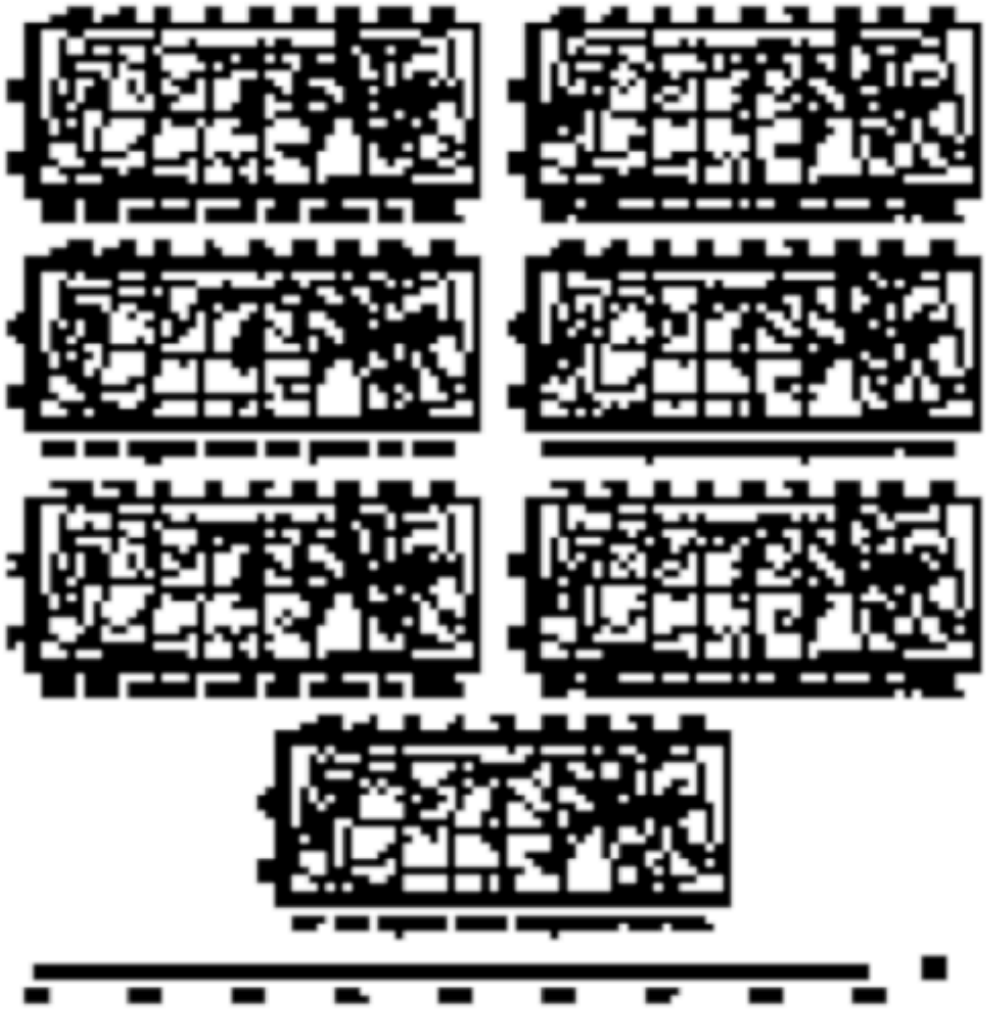

Figure 4.19: ROSR winter precipitation performance, calculated as model skill (r).

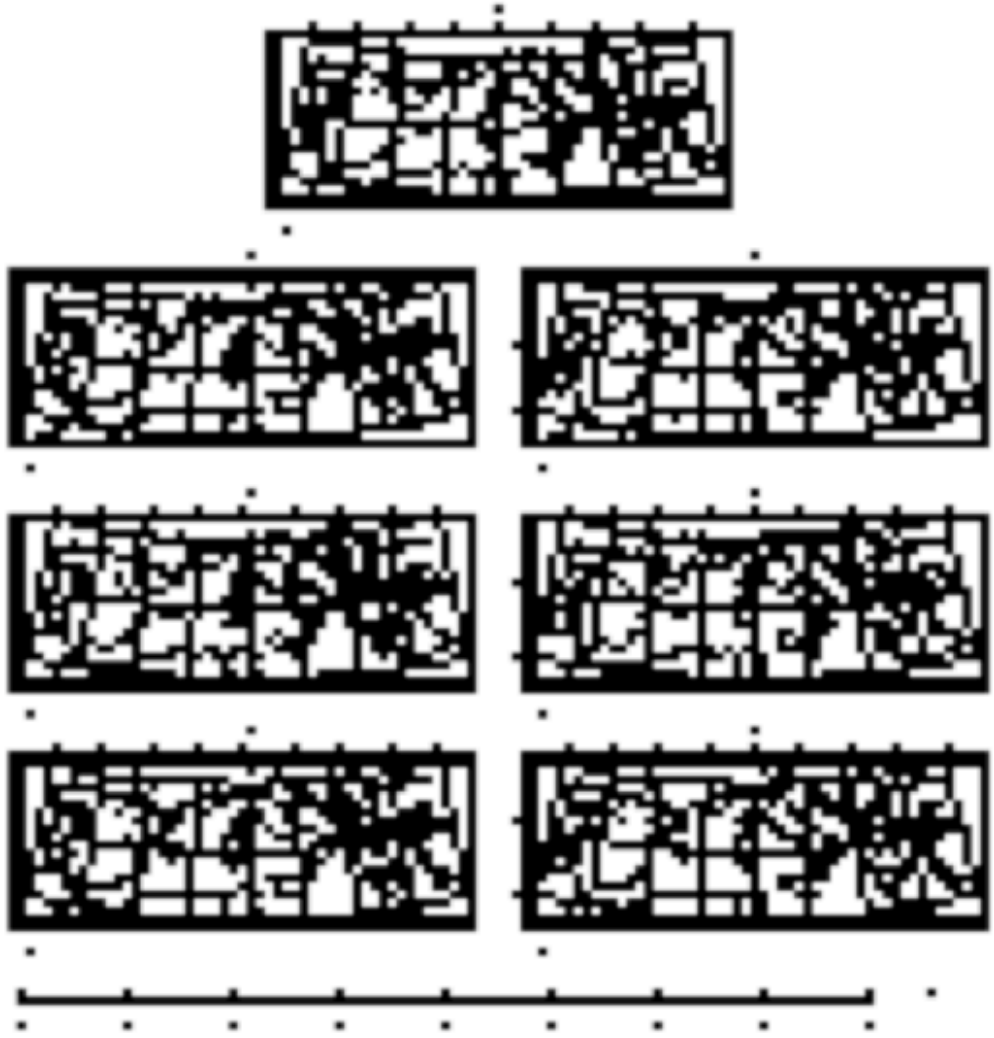

Fig 4.20

Figure 4.20: LBF winter temperature performance, calculated as model skill (r).

Fig 4.21

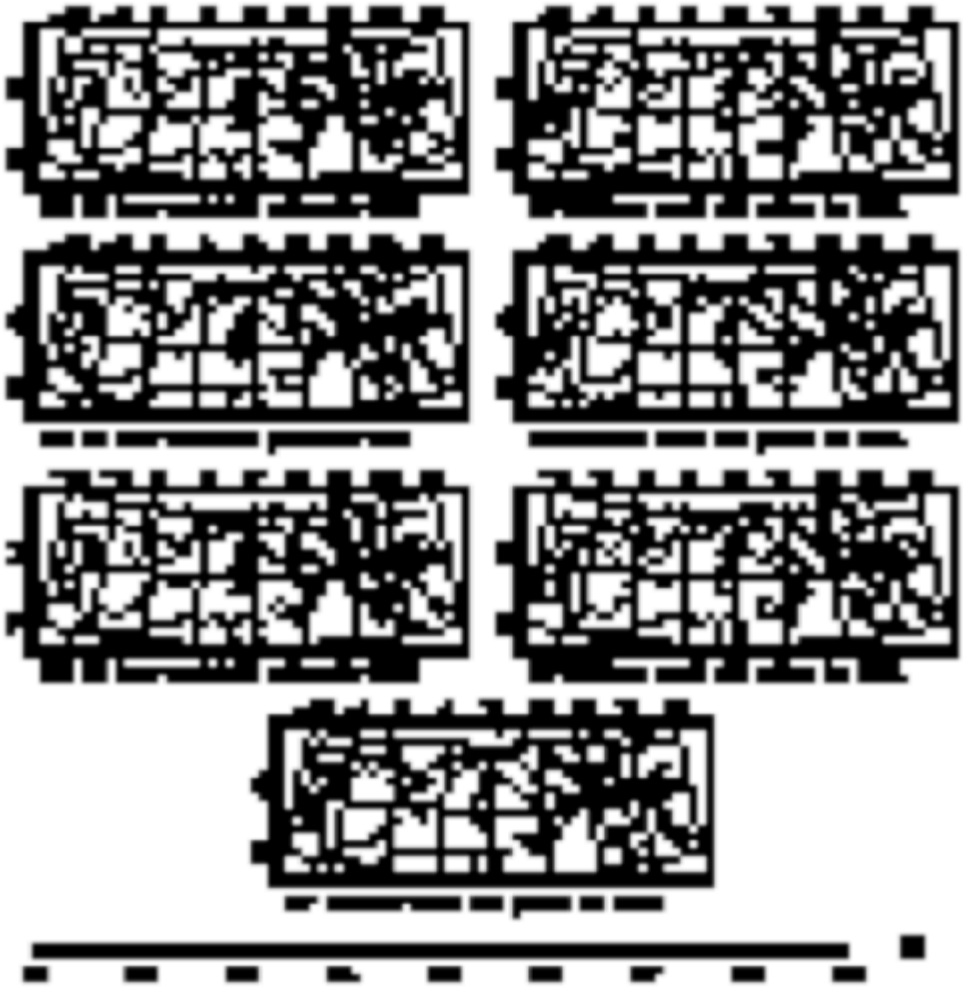

Figure 4.21: LBF winter precipitation performance, calculated as model skill (r).

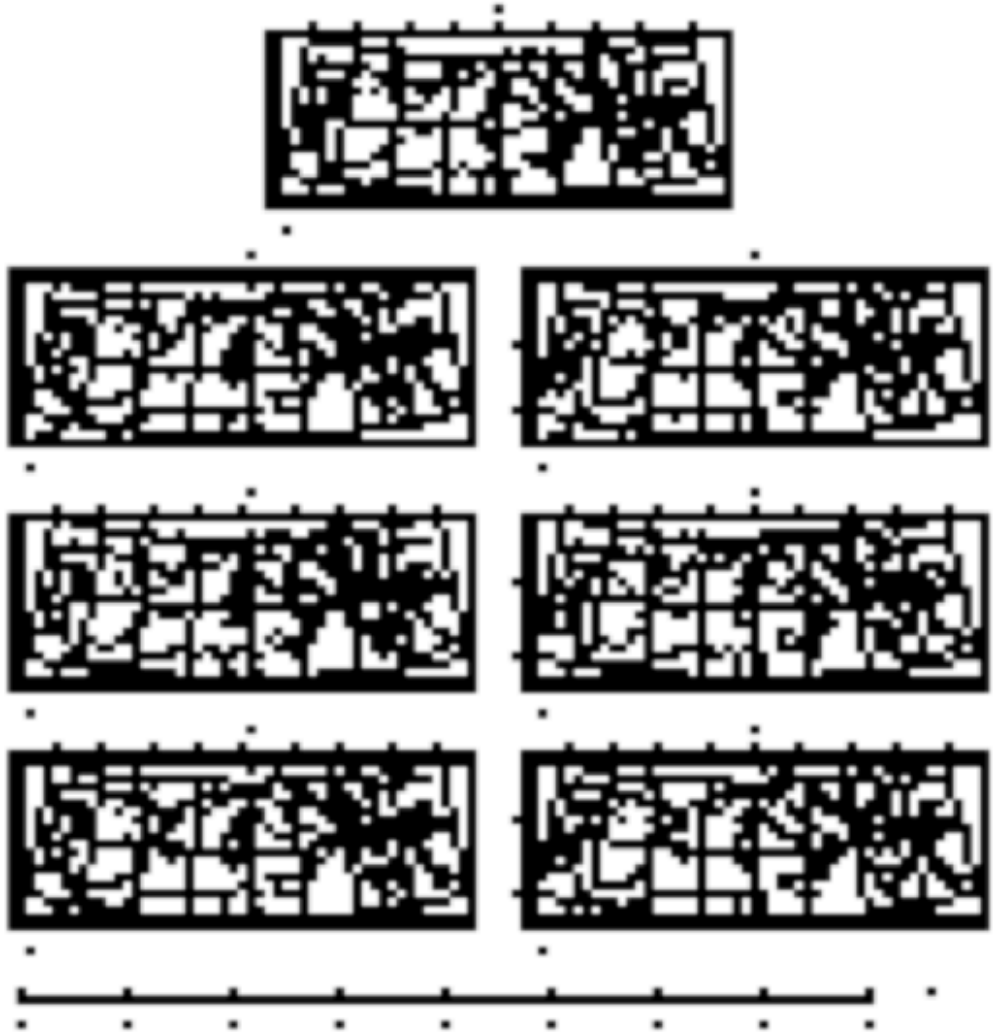

Fig 4.22

Figure 4.22: GBF winter temperature performance, calculated as model skill (r).

Fig 4.23

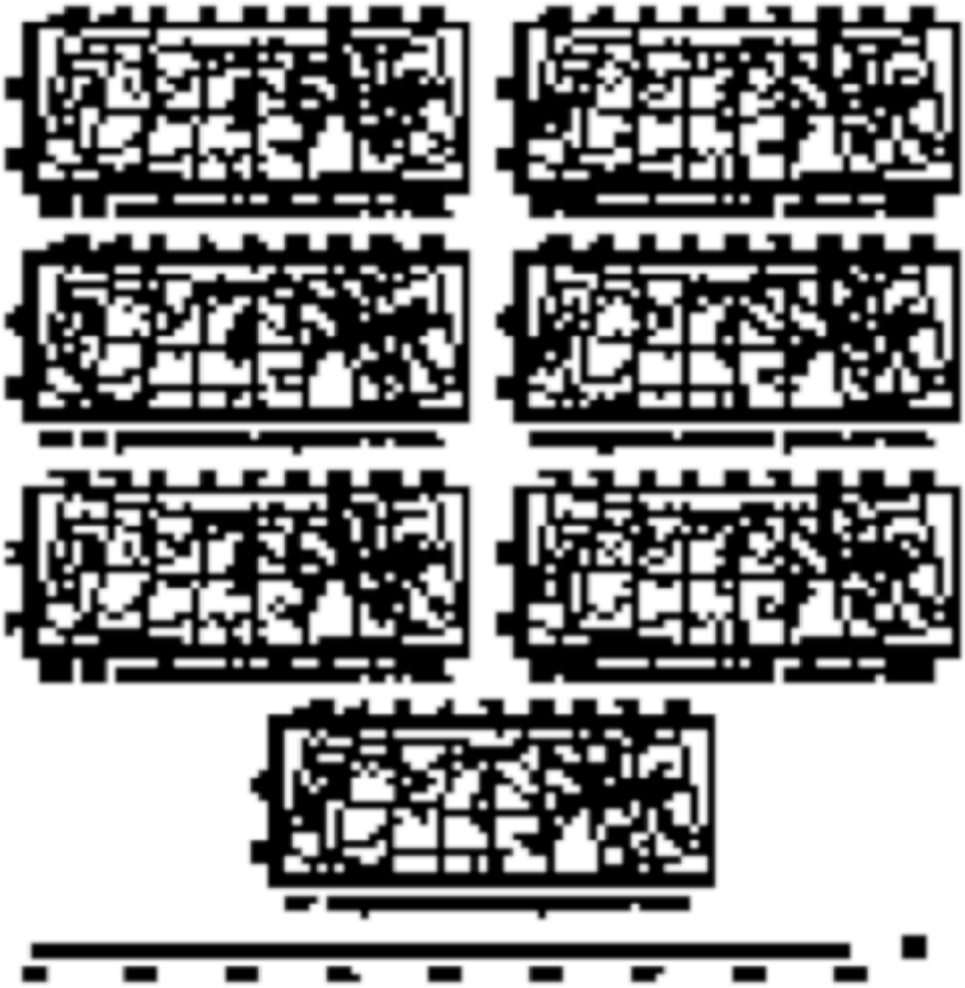

Figure 4.23: GBF winter precipitation performance, calculated as model skill (r).

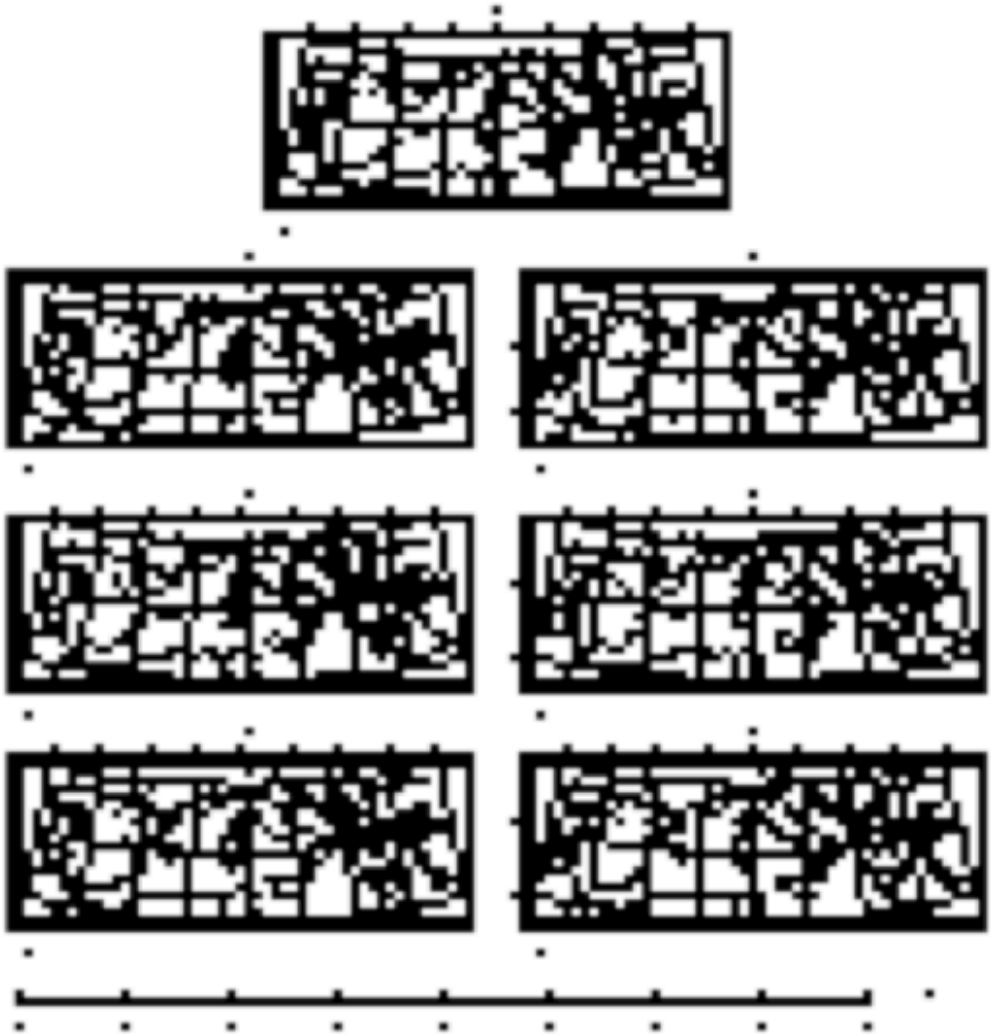

Fig 4.24

Figure 4.24: TBF winter temperature performance, calculated as model skill (r).

Fig 4.25

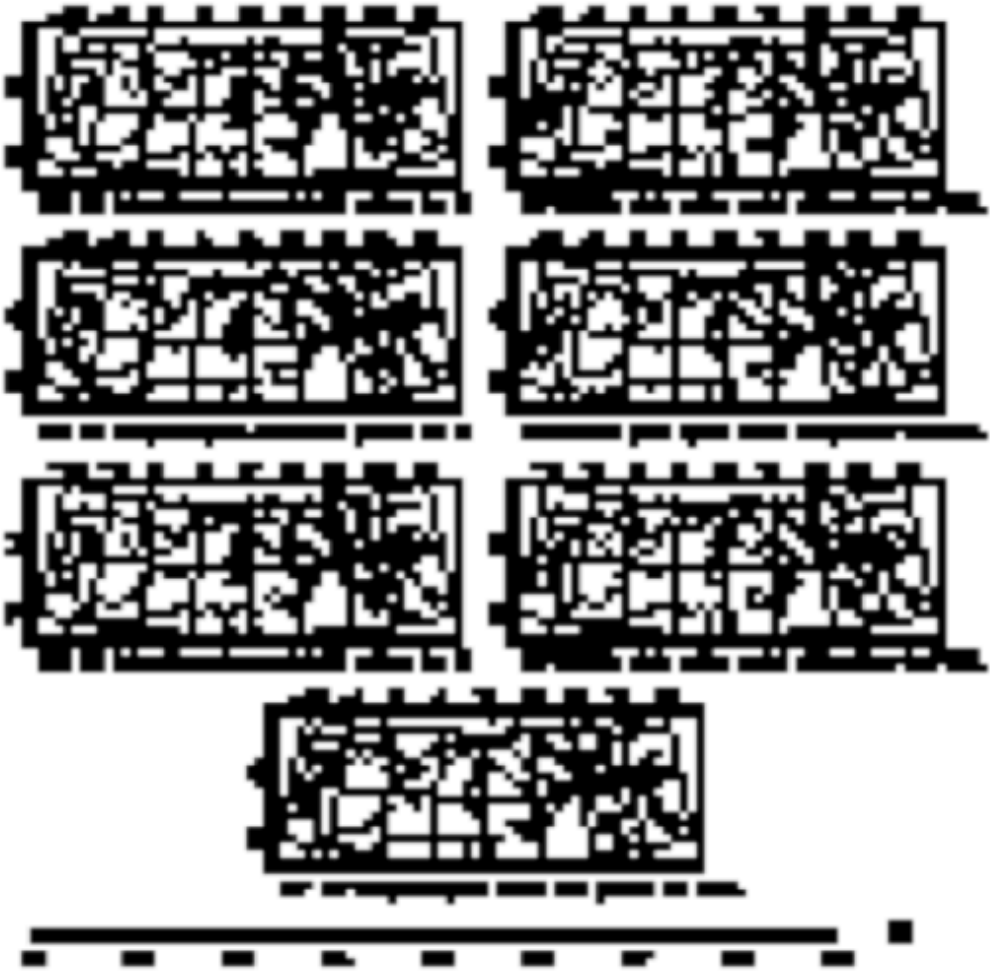

Figure 4.25: TBF winter precipitation performance, calculated as model skill (r).

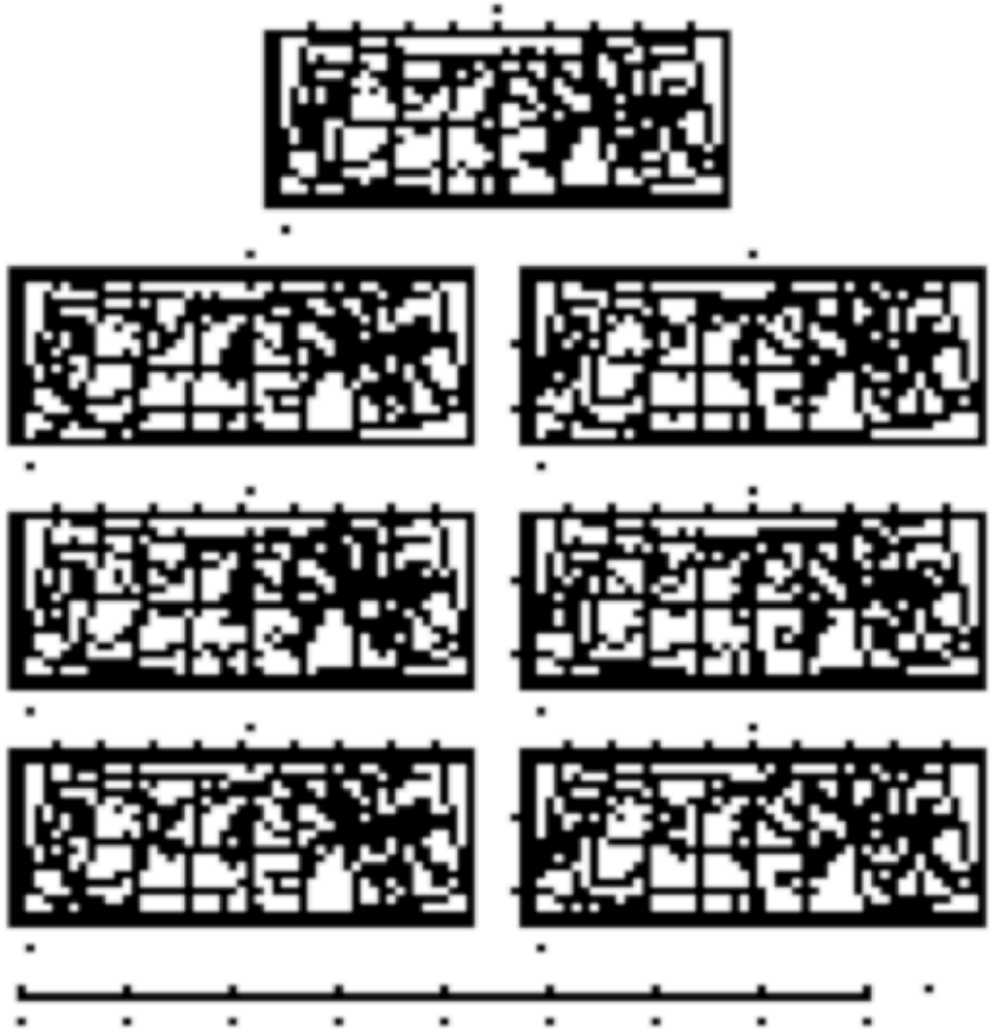

Fig 4.26

Figure 4.26: BOSR spring temperature performance, calculated as model skill (r).

Fig 4.27

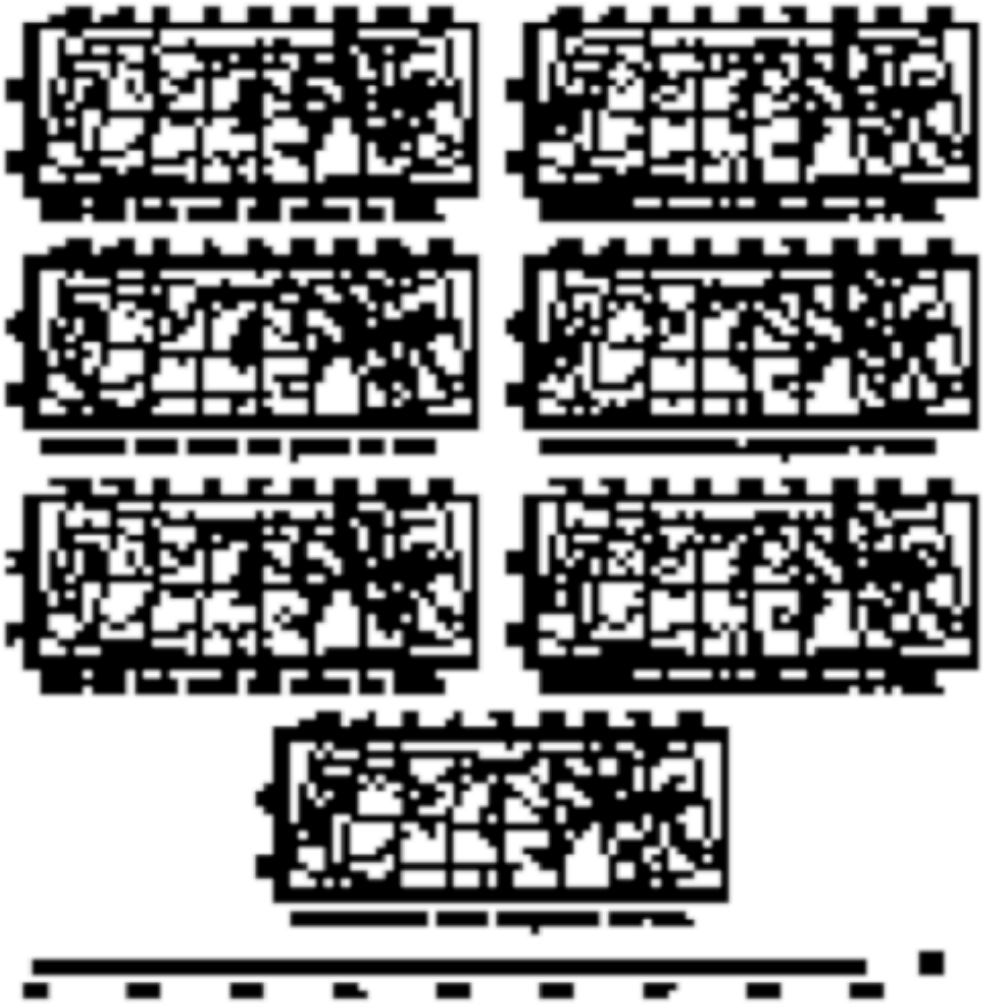

Figure 4.27: BOSR spring precipitation performance, calculated as model skill (r).

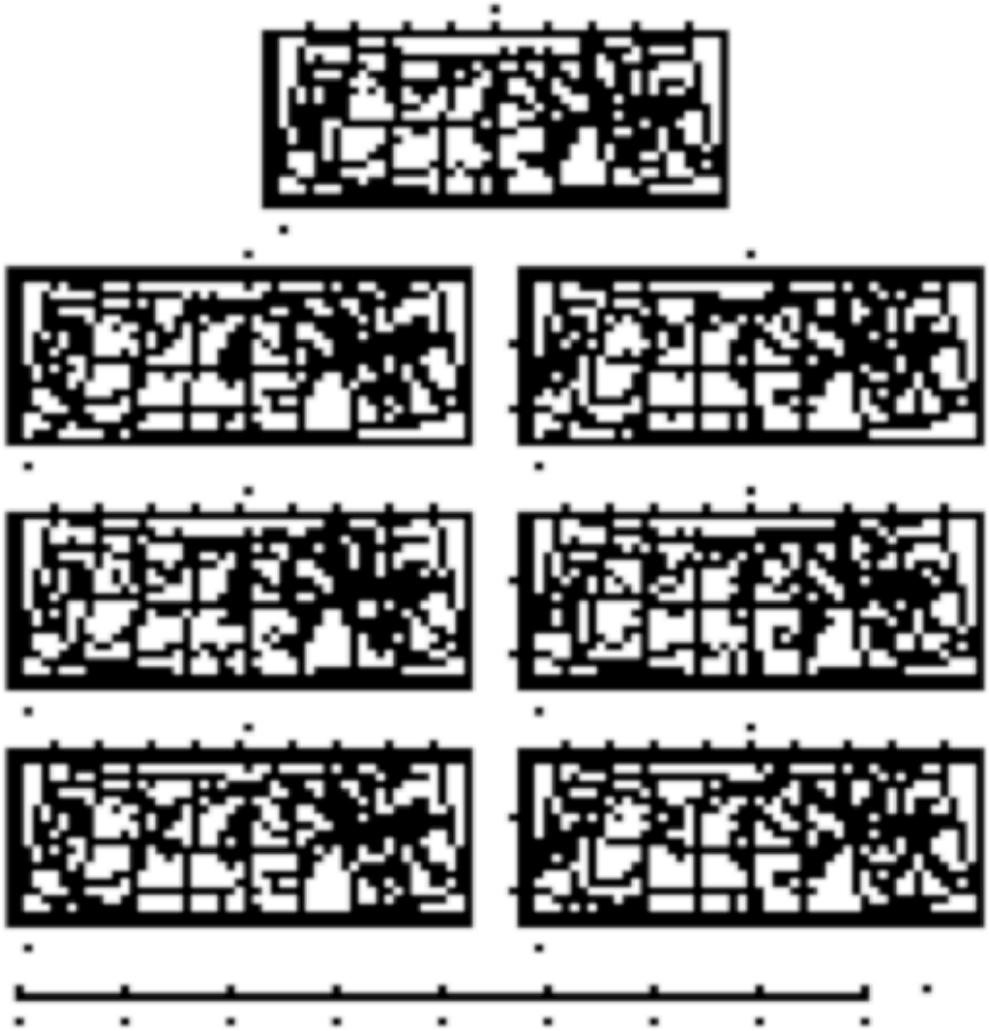

Fig 4.28

Figure 4.28: ROSR spring temperature performance, calculated as model skill (r).

Fig 4.29

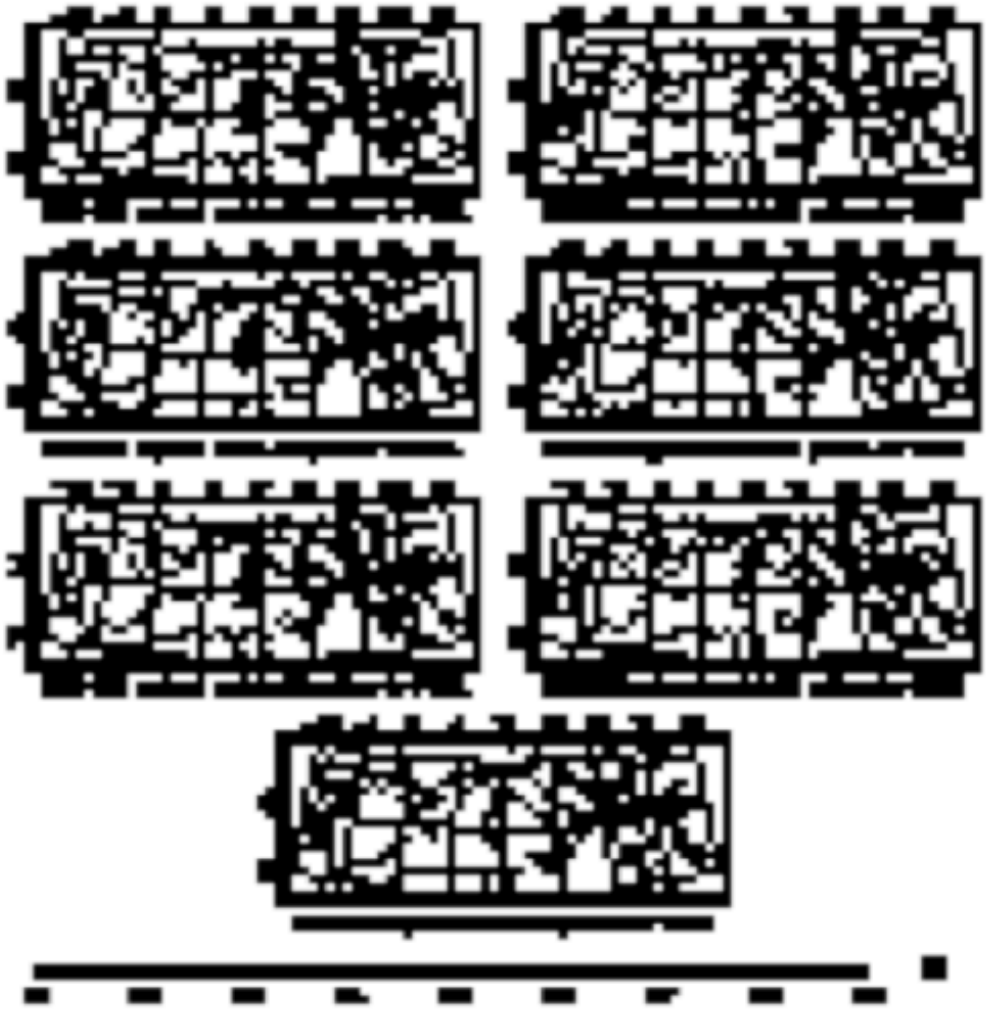

Figure 4.29: ROSR spring precipitation performance, calculated as model skill (r).

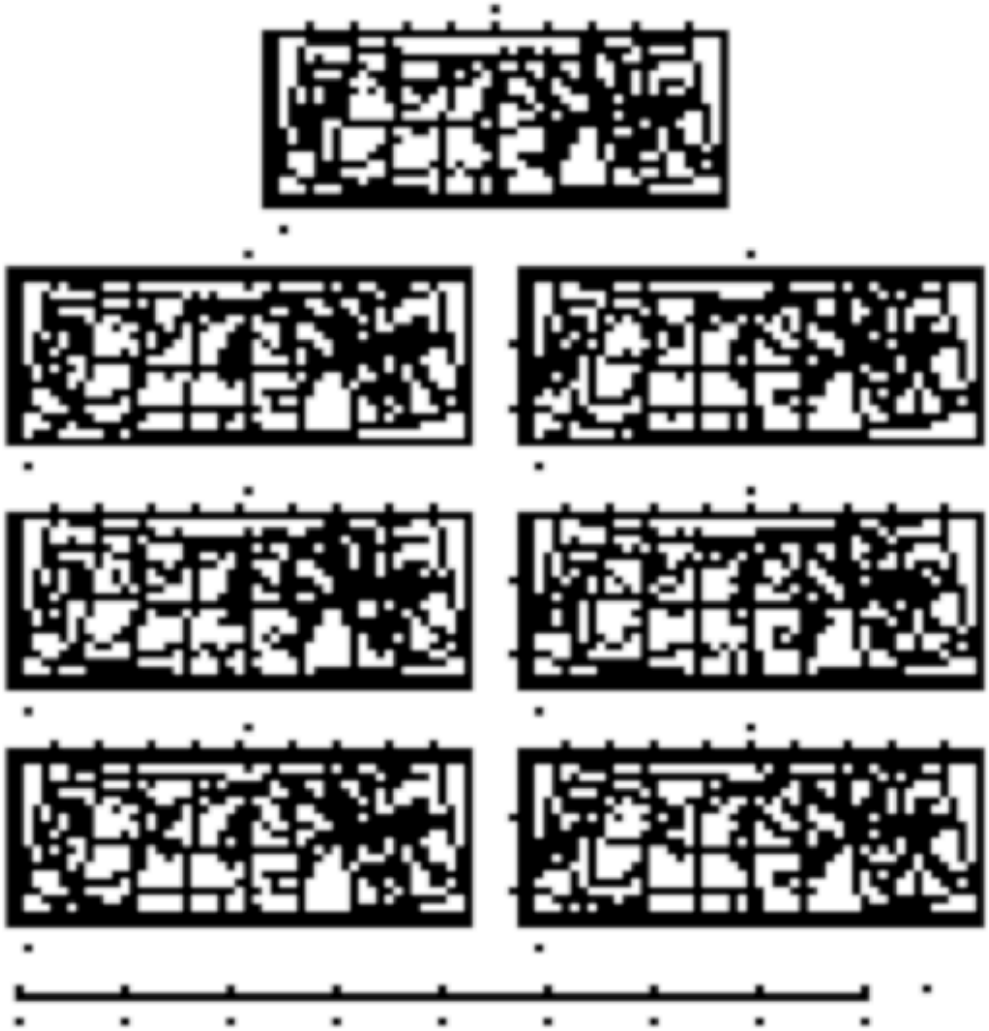

Fig 4.30

Figure 4.30: LBF spring temperature performance, calculated as model skill (r).

Fig 4.31

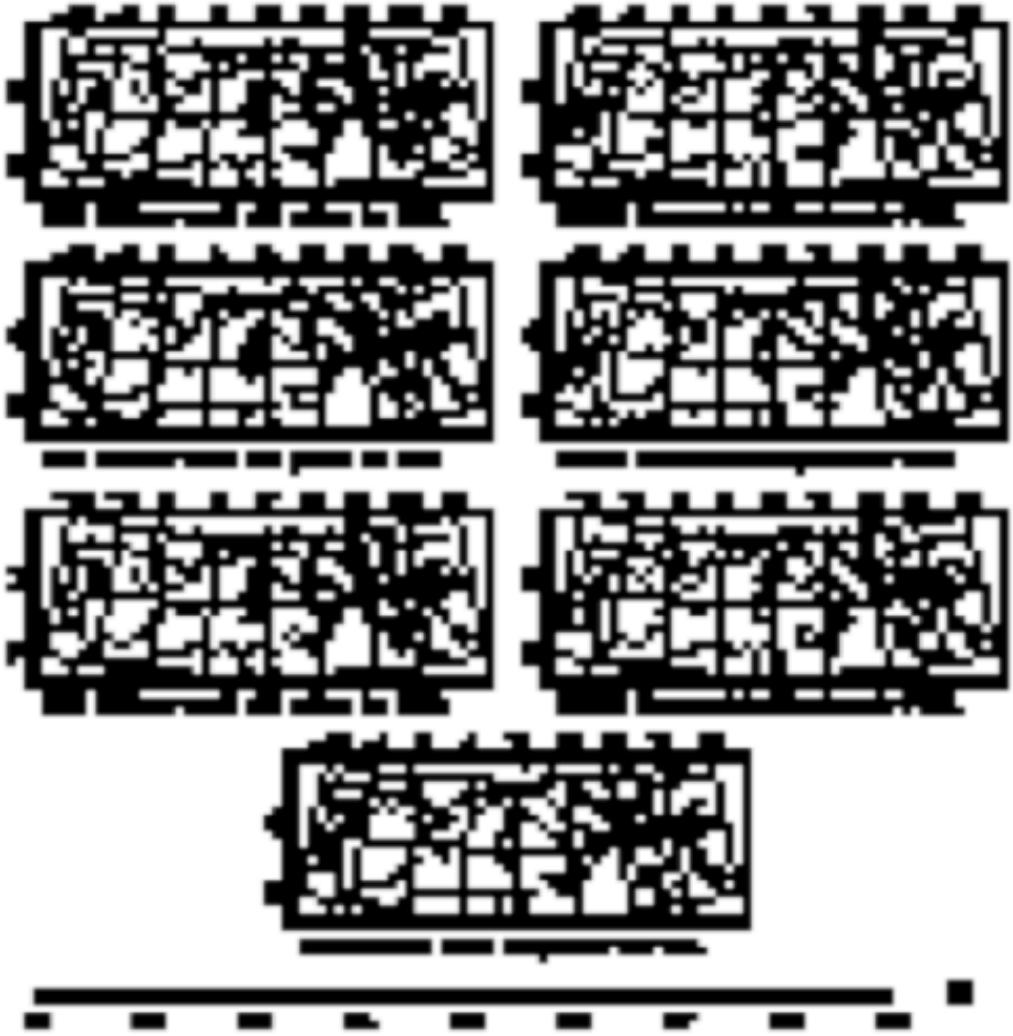

Figure 4.31: LBF spring precipitation performance, calculated as model skill (r)

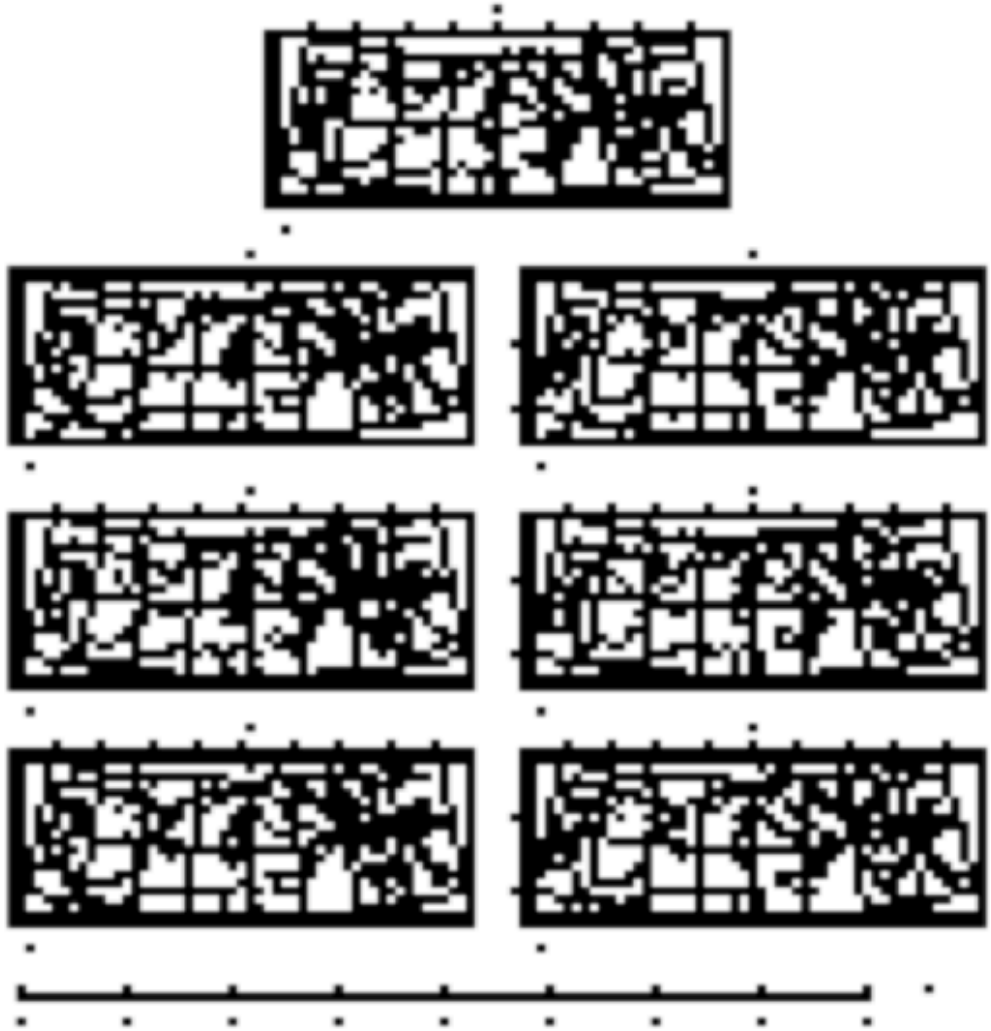

Fig 4.32

Figure 4.32: GBF spring temperature performance, calculated as model skill (r).

Fig 4.33

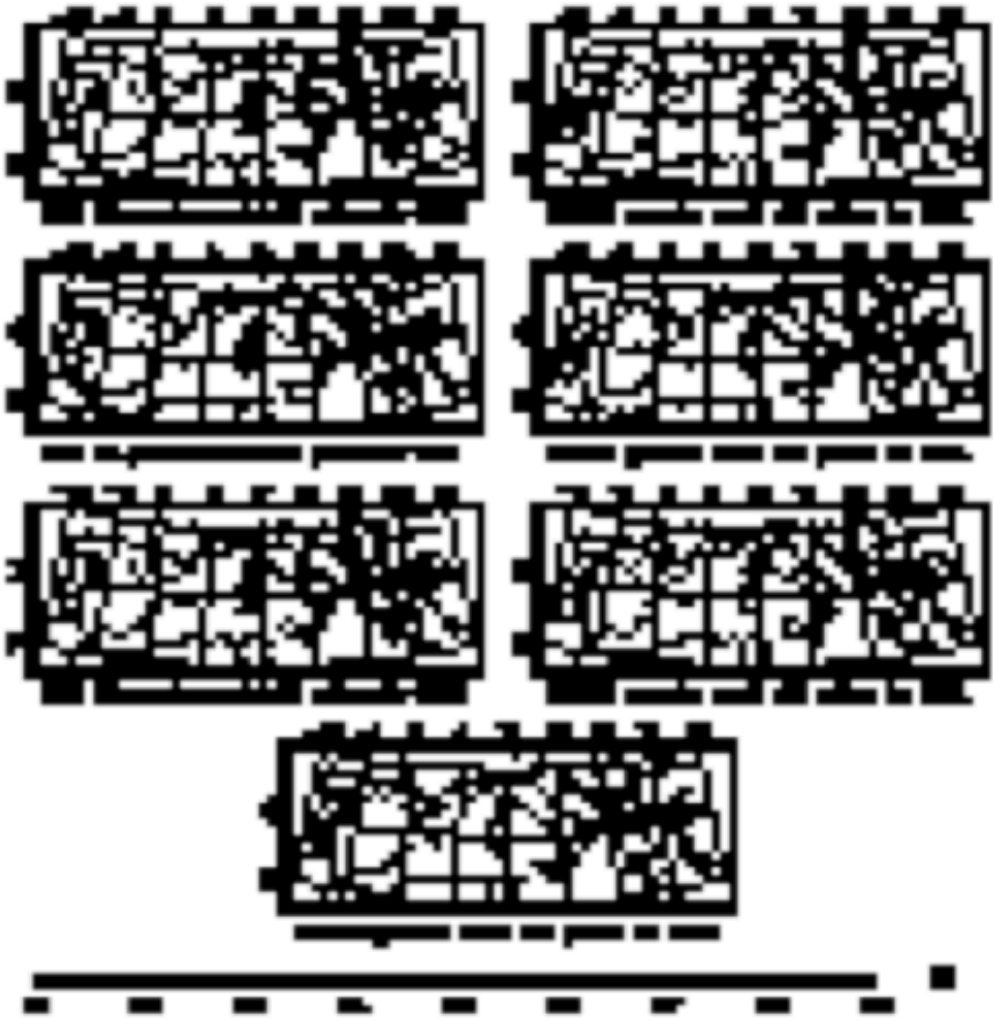

Figure 4.33: GBF spring precipitation performance, calculated as model skill (r).

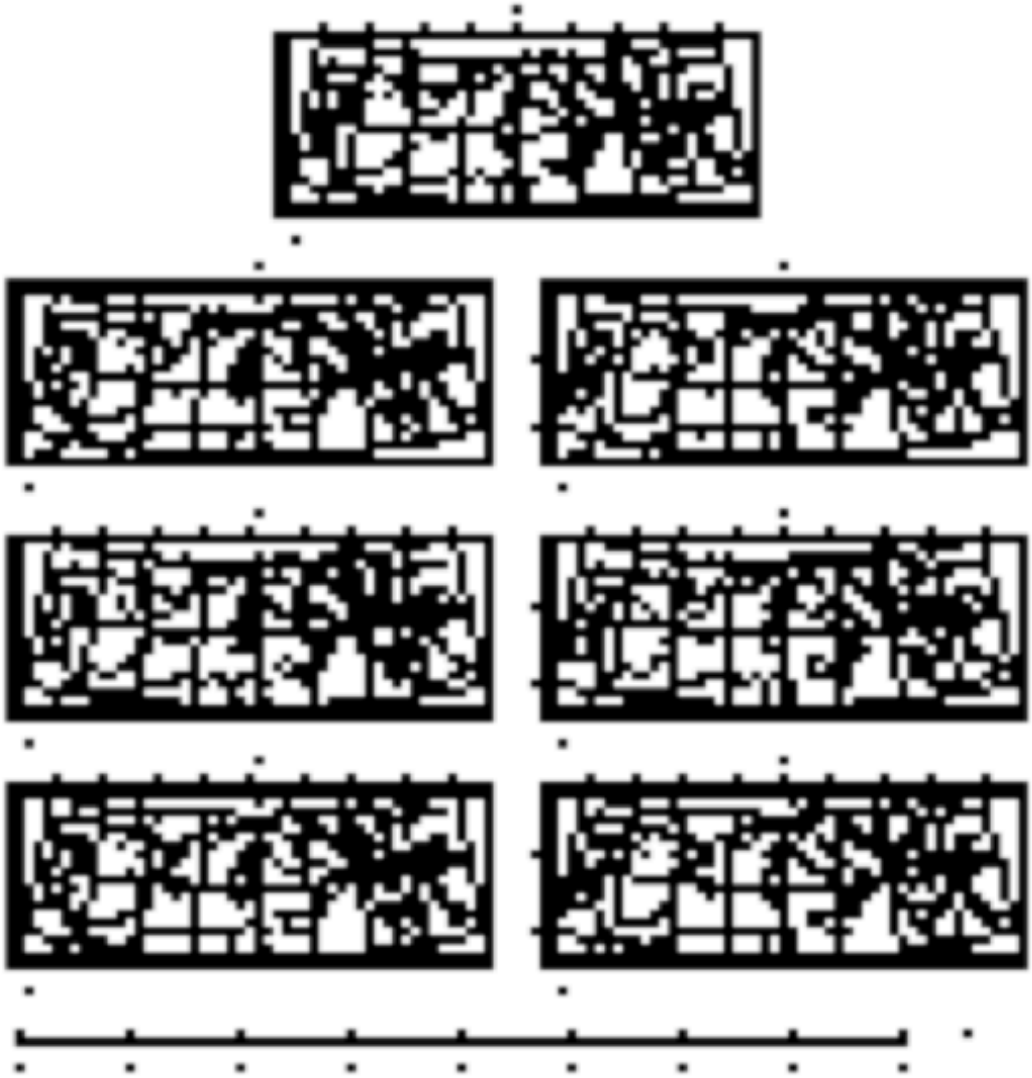

Fig 4.34

Figure 4.34: TBF spring temperature performance, calculated as model skill (r).

Fig 4.35

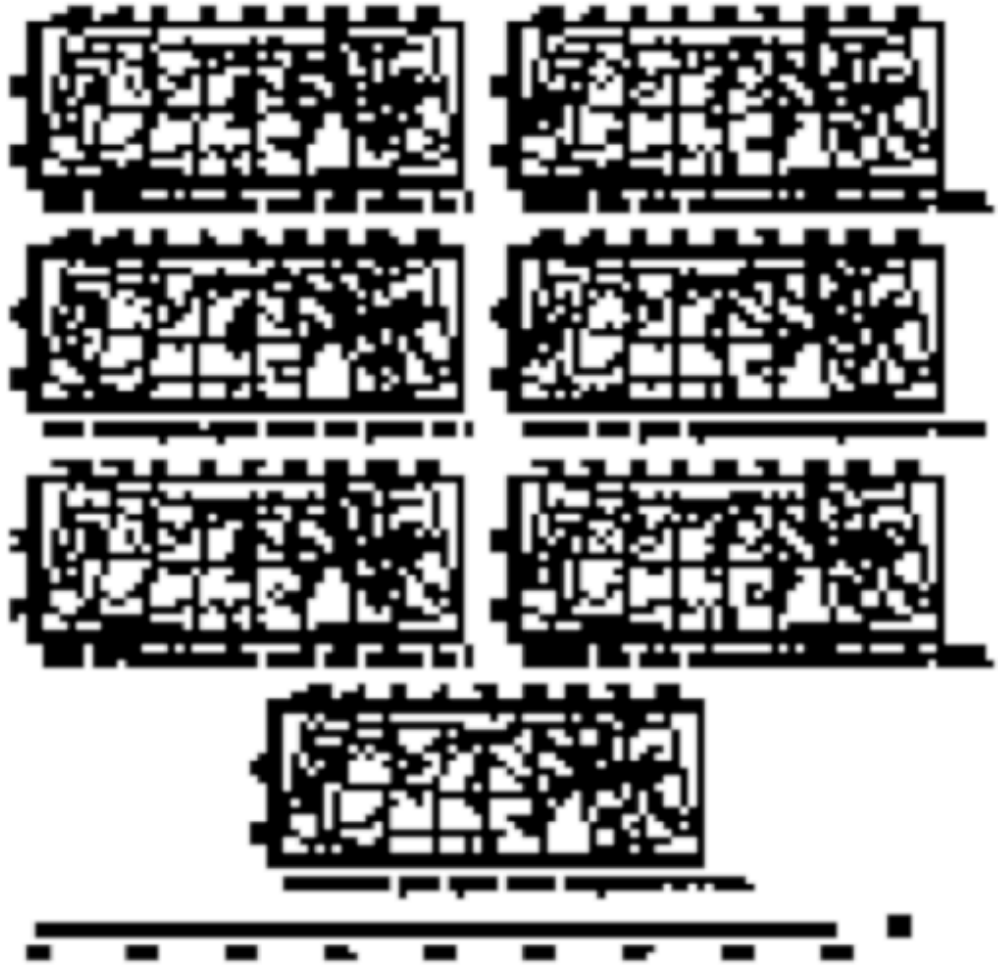

Figure 4.35: TBF spring precipitation performance, calculated as model skill (r).

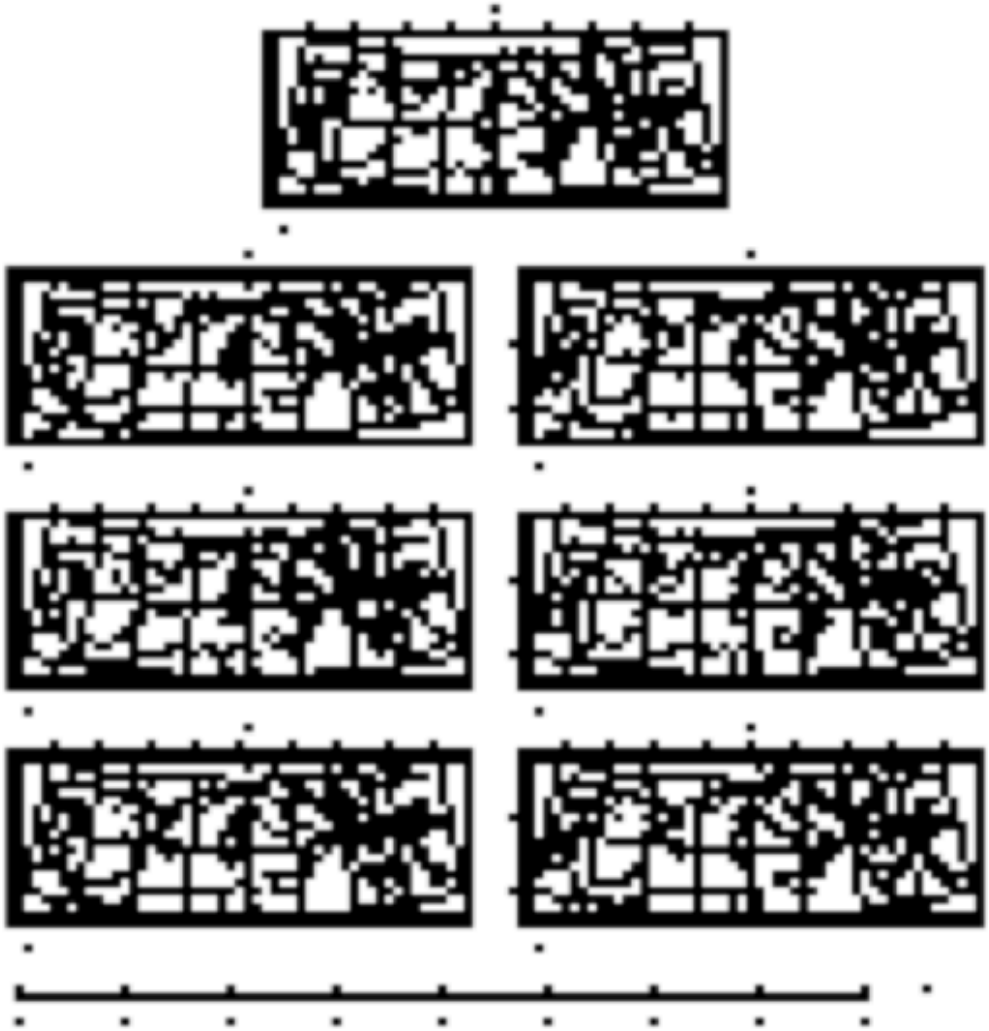

Fig 4.36

Figure 4.36: BOSR summer temperature performance, calculated as model skill (r).

Fig 4.37

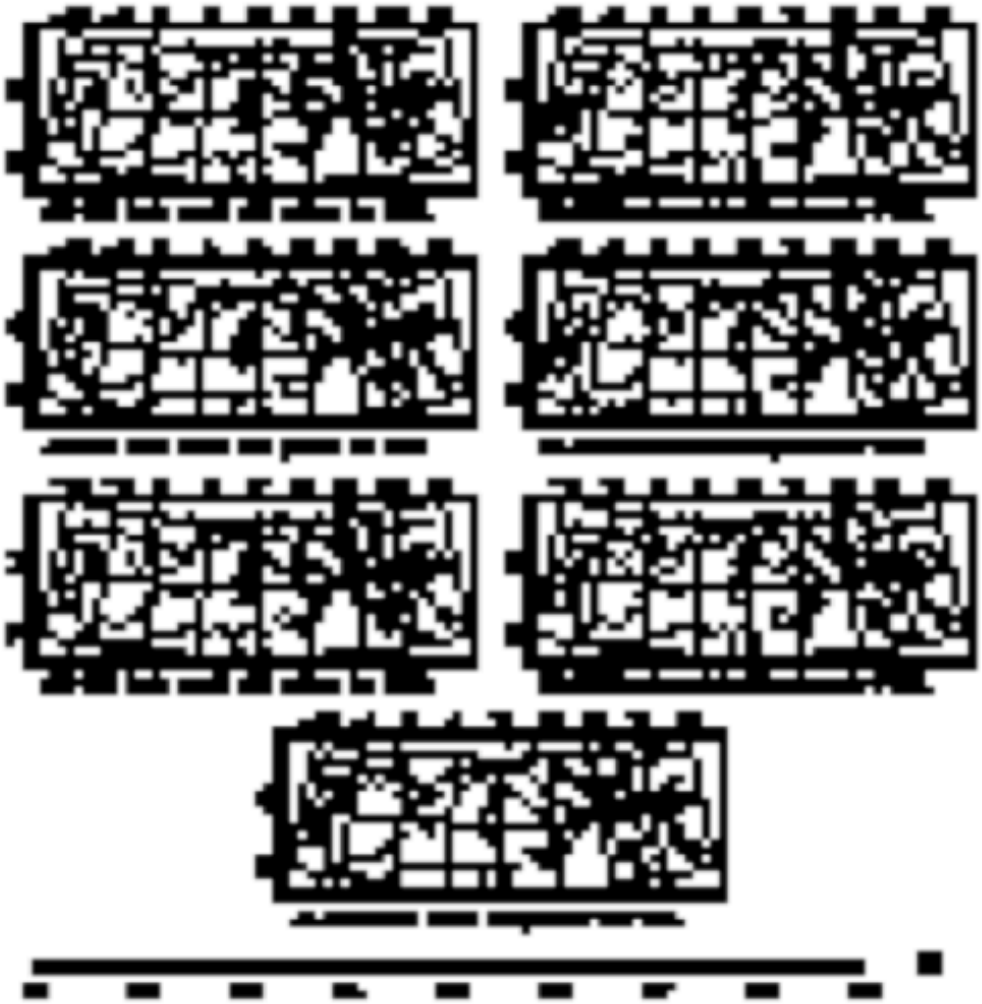

Figure 4.37: BOSR summer precipitation performance, calculated as model skill (r).

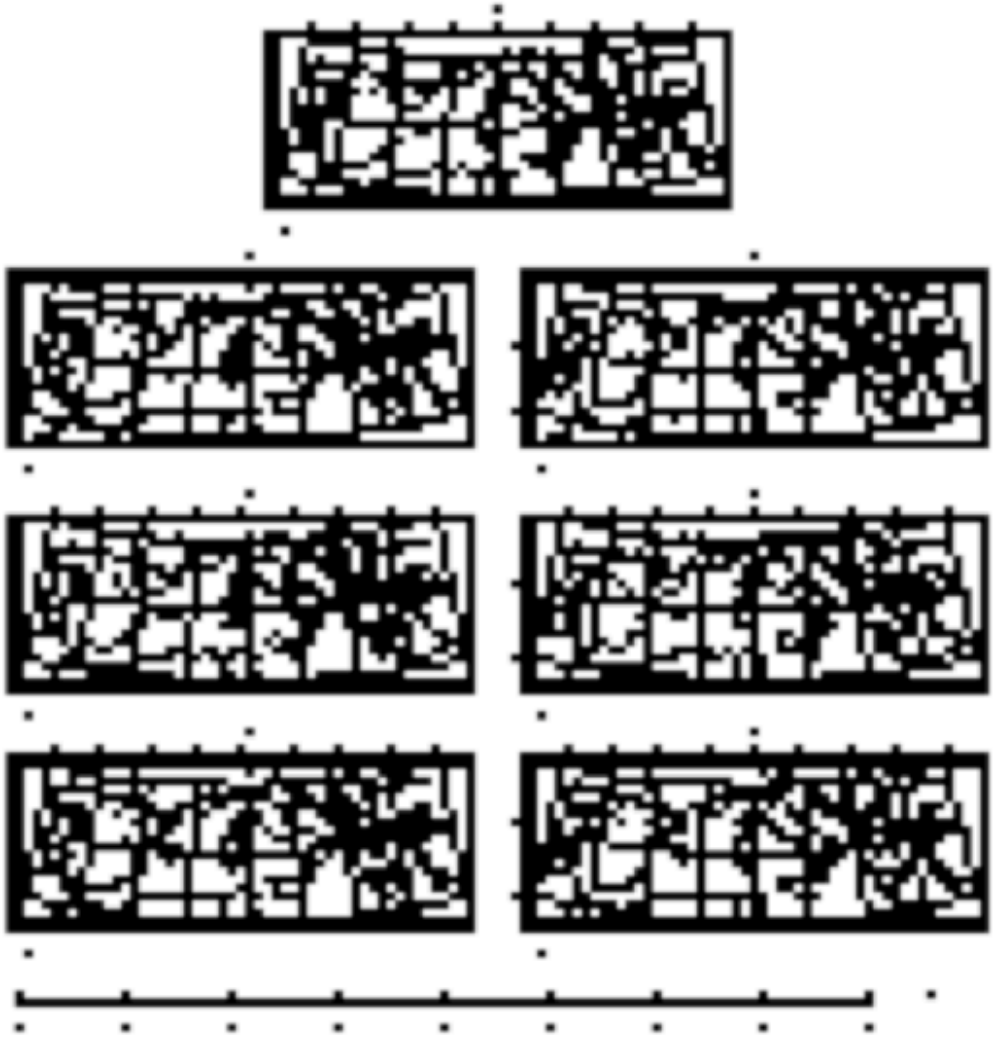

Fig 4.38

Figure 4.38: ROSR summer temperature performance, calculated as model skill (r).

Fig 4.39

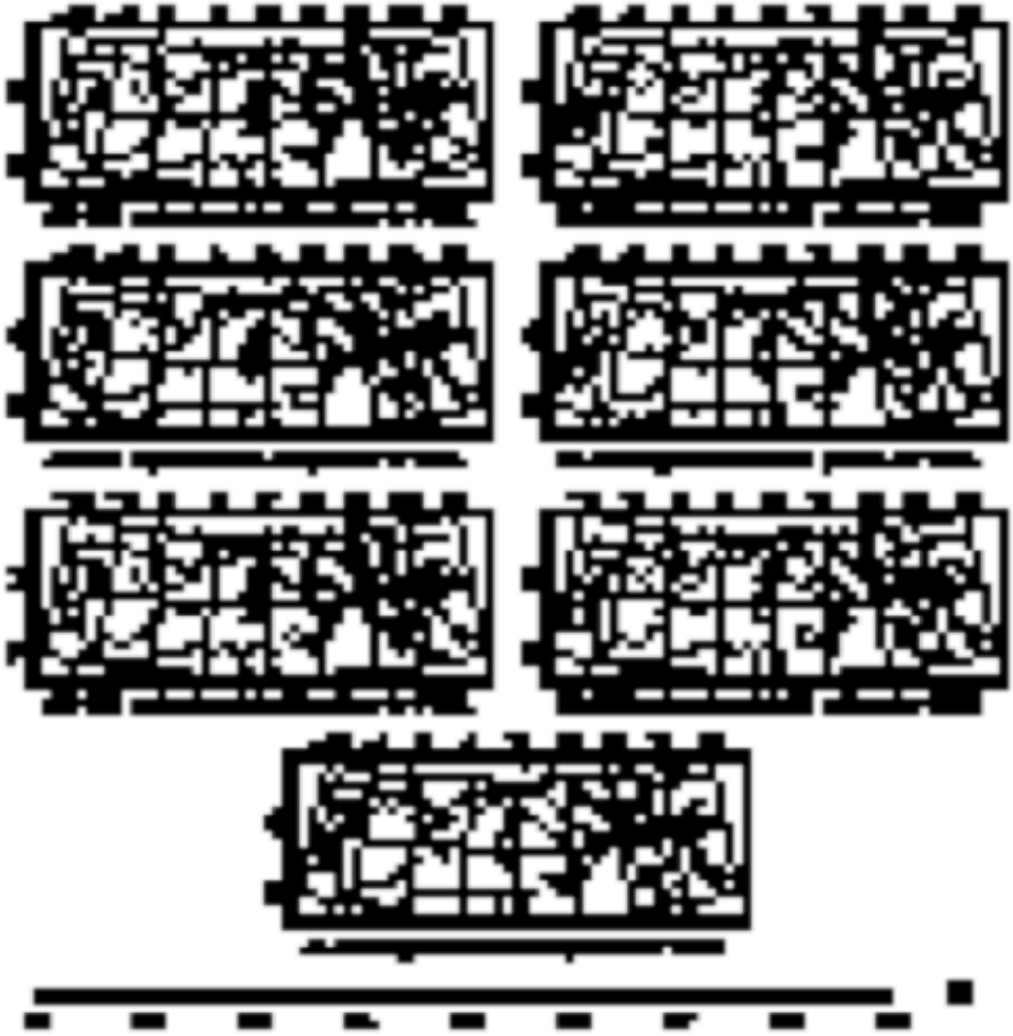

Figure 4.39: ROSR summer precipitation performance, calculated as model skill (r).

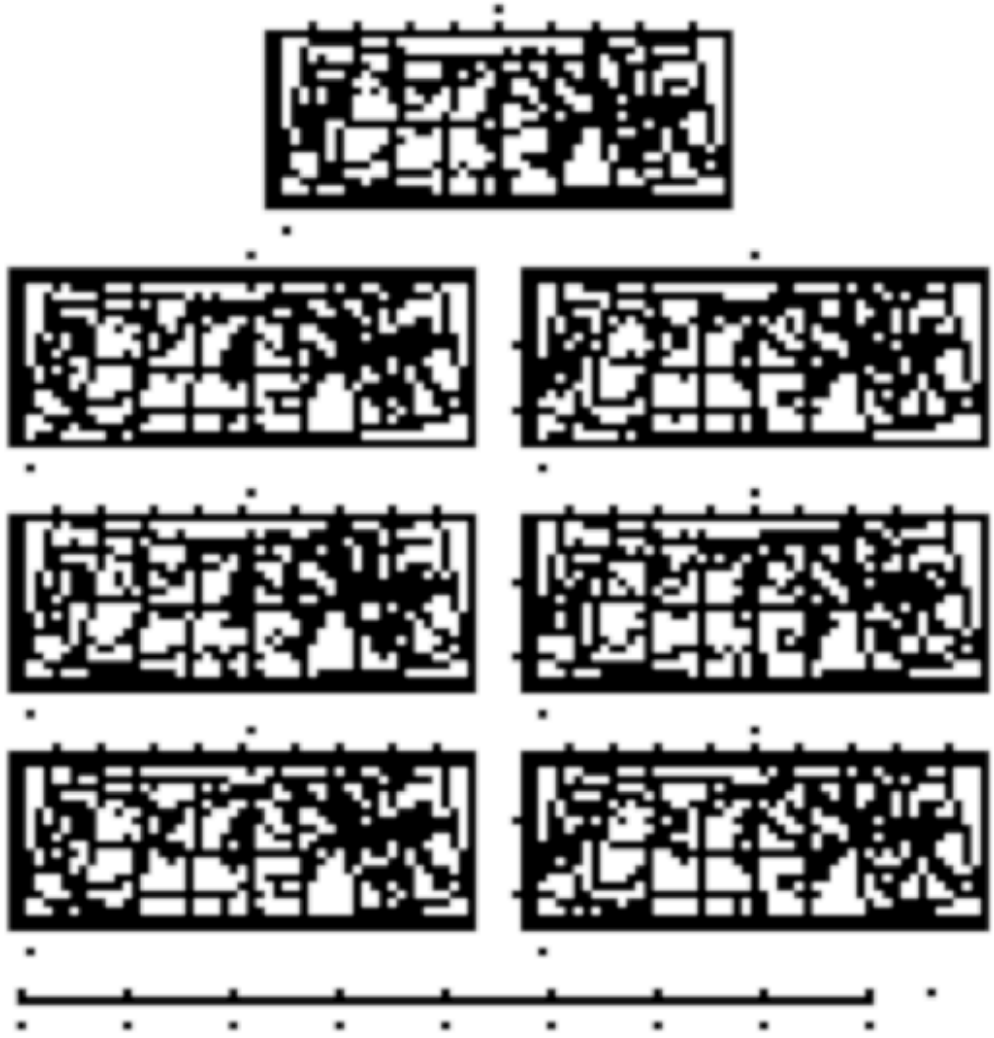

Fig 4.40

Figure 4.40: LBF summer temperature performance, calculated as model skill (r).

Fig 4.41

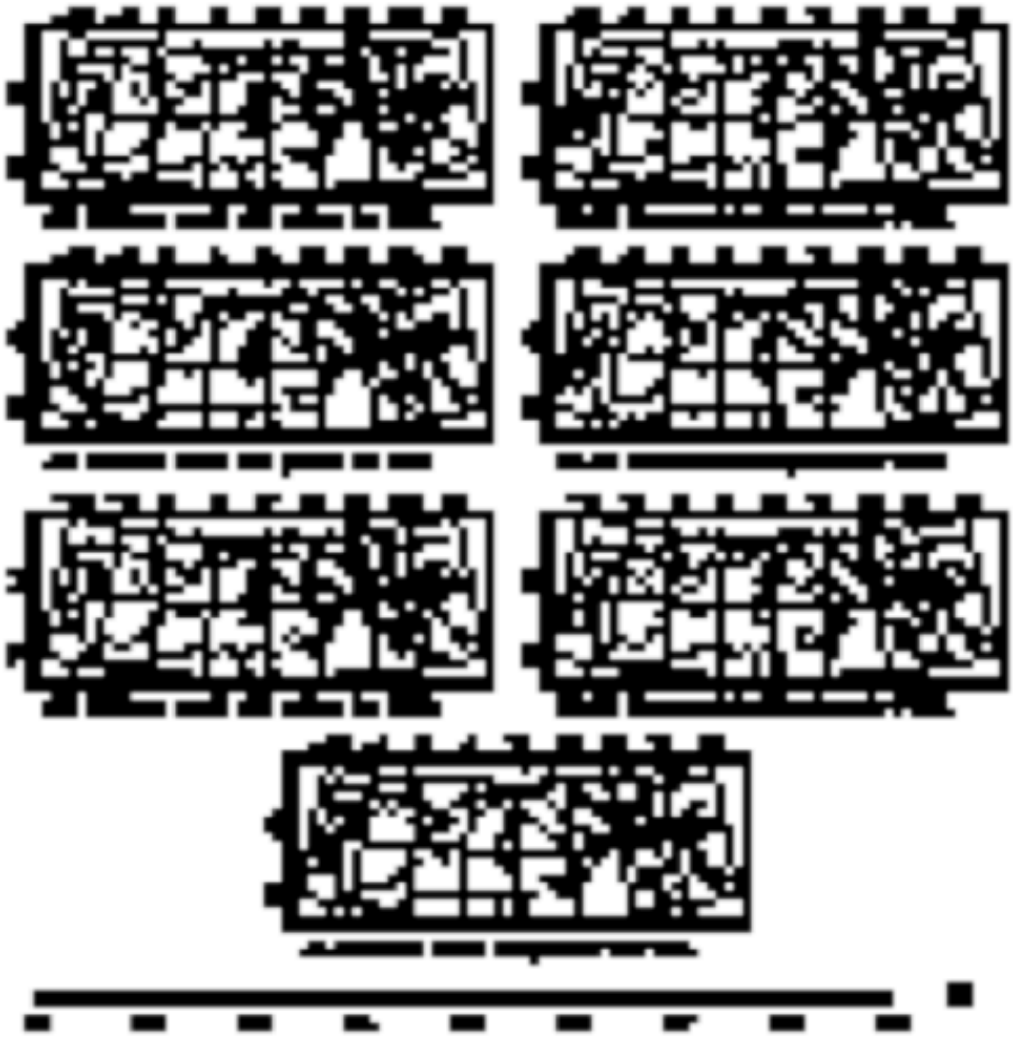

Figure 4.41: LBF summer precipitation performance, calculated as model skill (r).

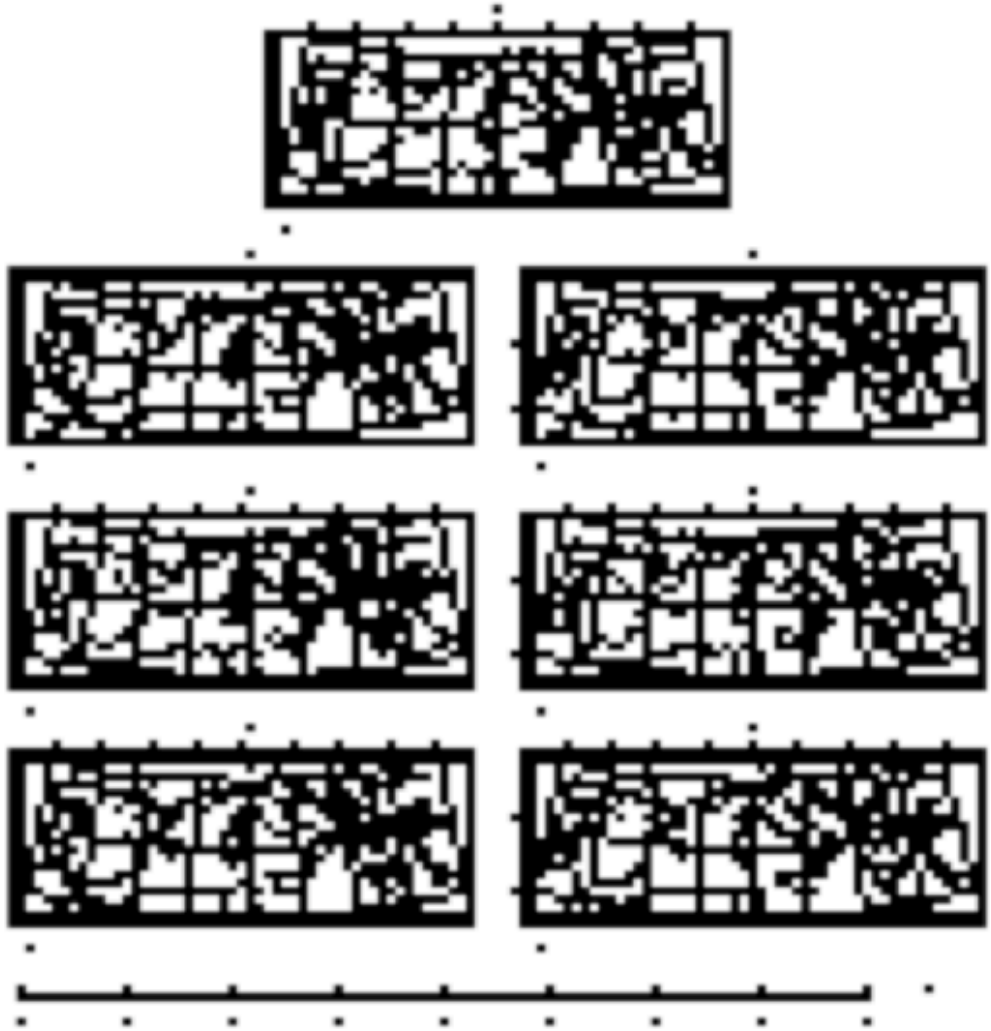

Fig 4.42

Figure 4.42: GBF summer temperature performance, calculated as model skill (r).

Fig 4.43

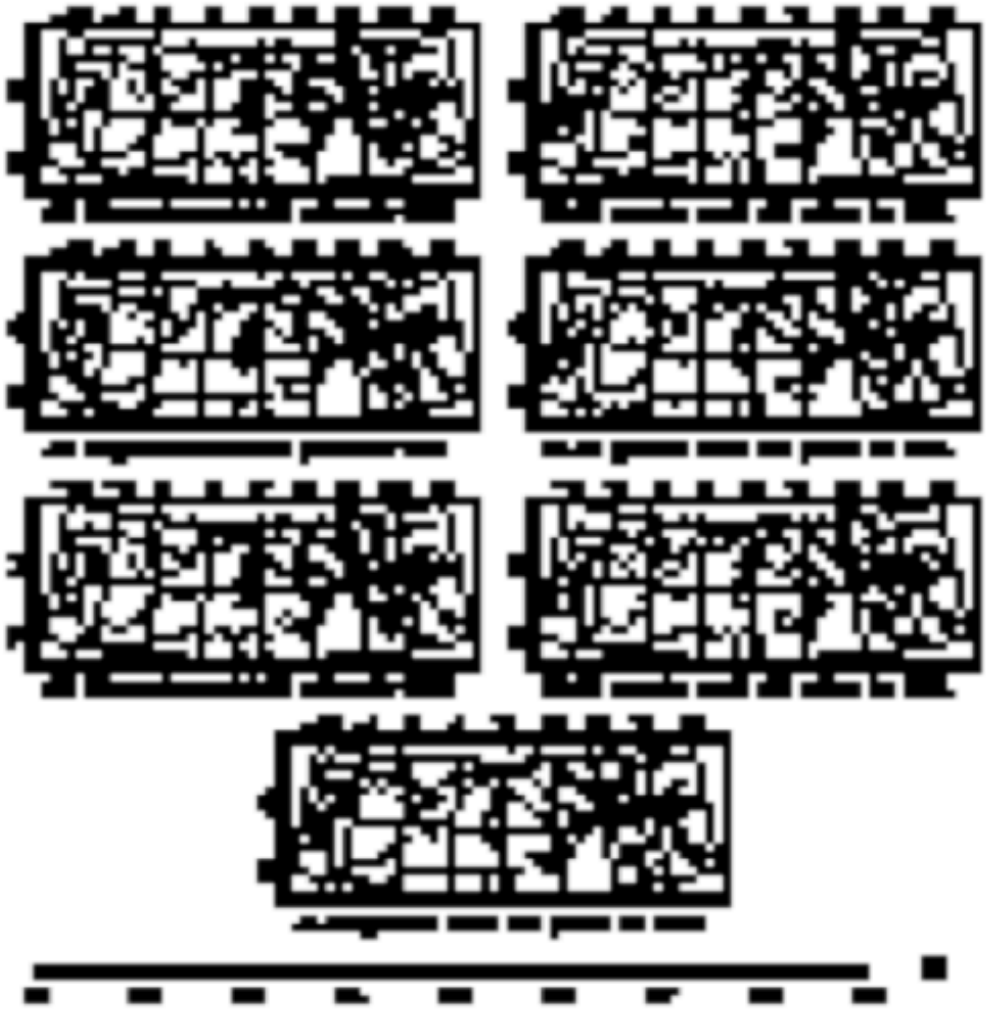

Figure 4.43: GBF summer precipitation performance, calculated as model skill (r).

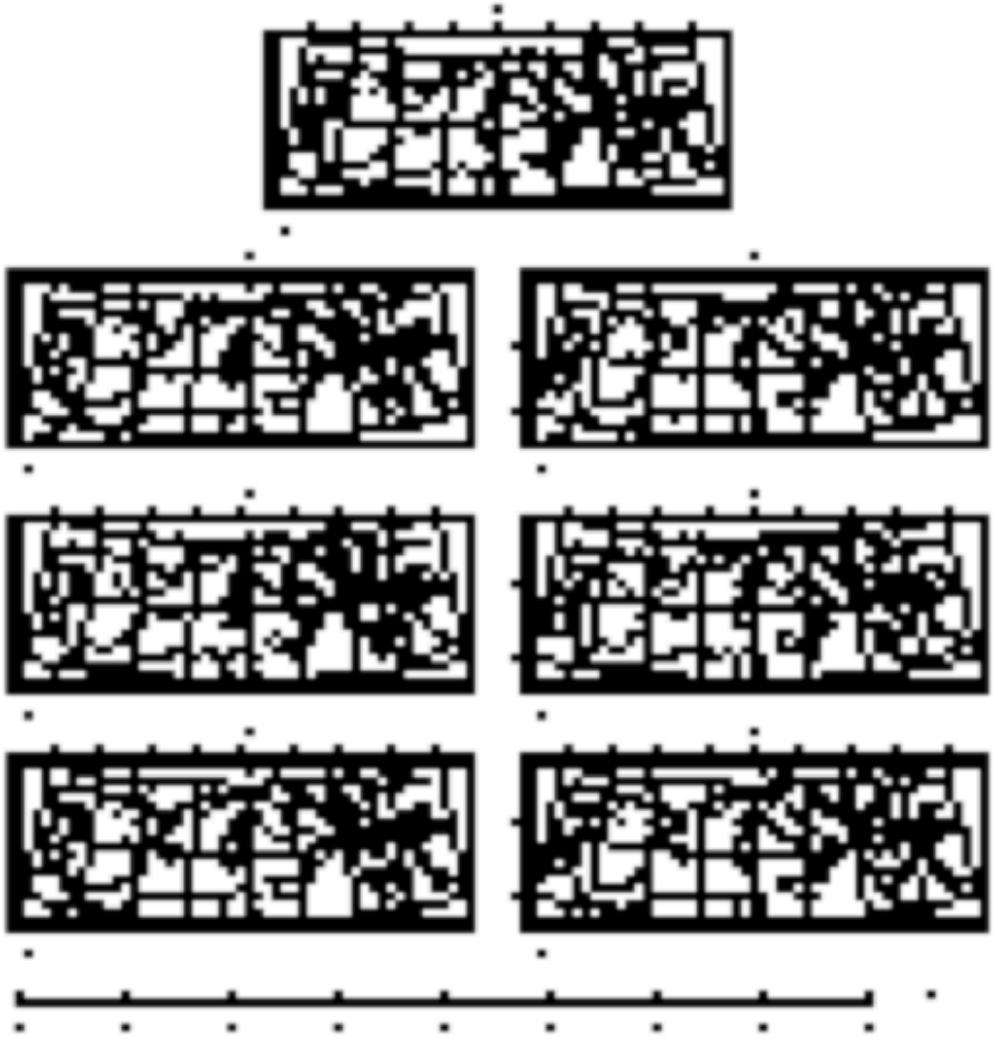

Fig 4.44

Figure 4.44: BOSR autumn temperature performance, calculated as model skill (r).

Fig 4.45

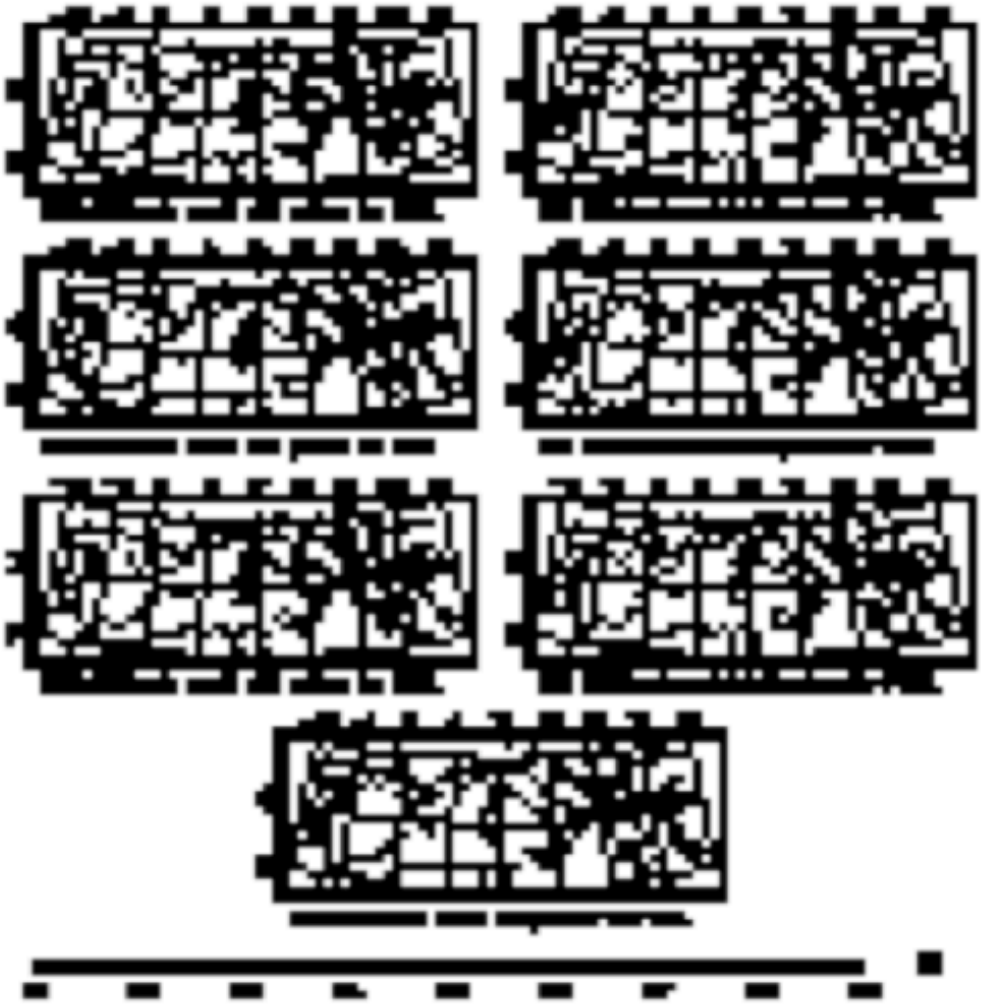

Figure 4.45: BOSR autumn precipitation performance, calculated as model skill (r).

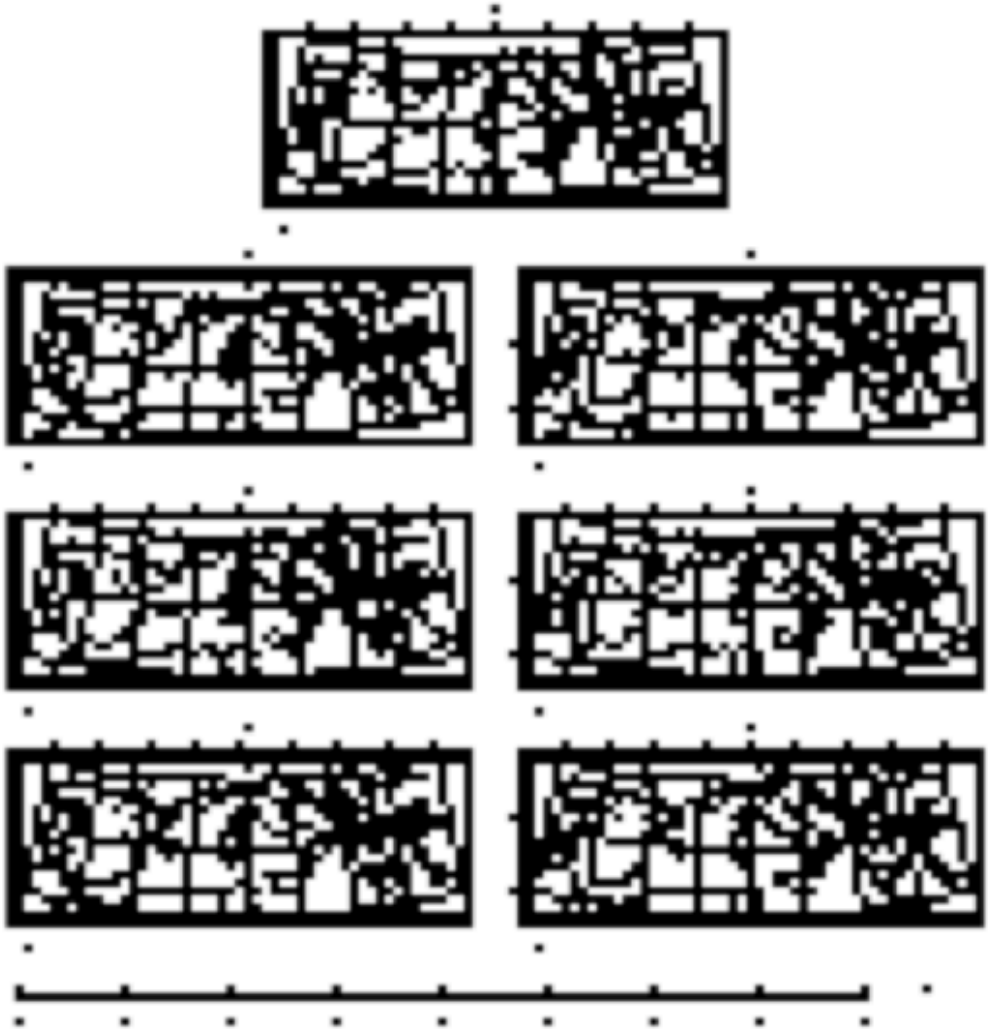

Fig 4.46

Figure 4.46: ROSR autumn temperature performance, calculated as model skill (r).

Fig 4.47

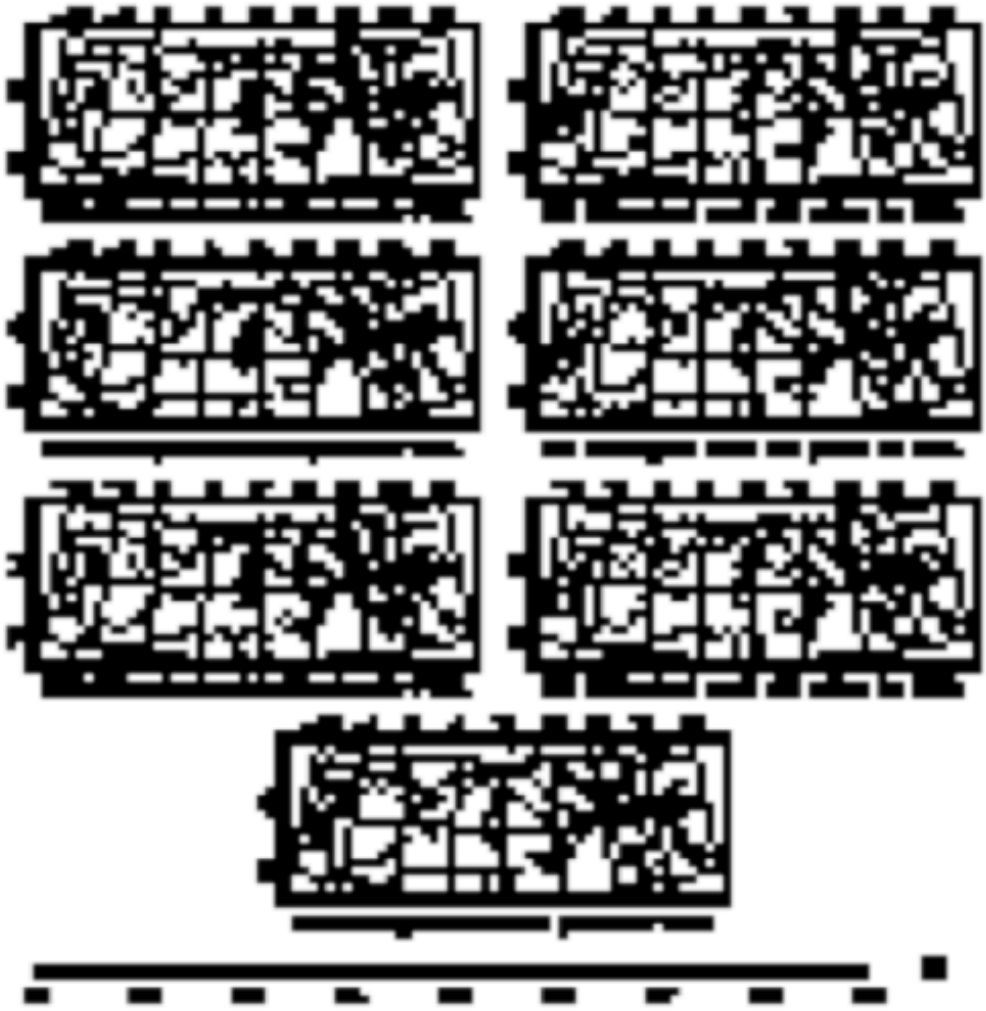

Figure 4.47: ROSR autumn precipitation performance, calculated as model skill (r).

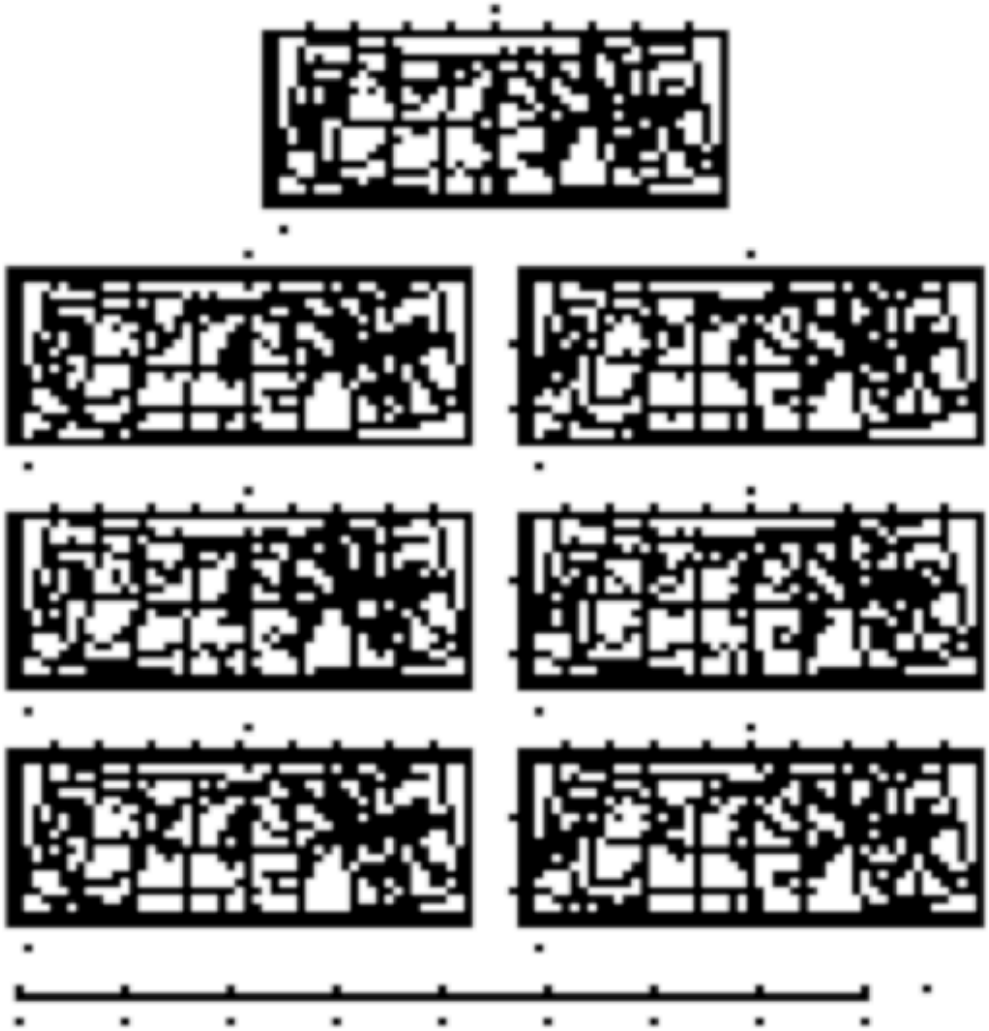

Fig 4.48

Figure 4.48: LBF autumn temperature performance, calculated as model skill (r).

Fig 4.49

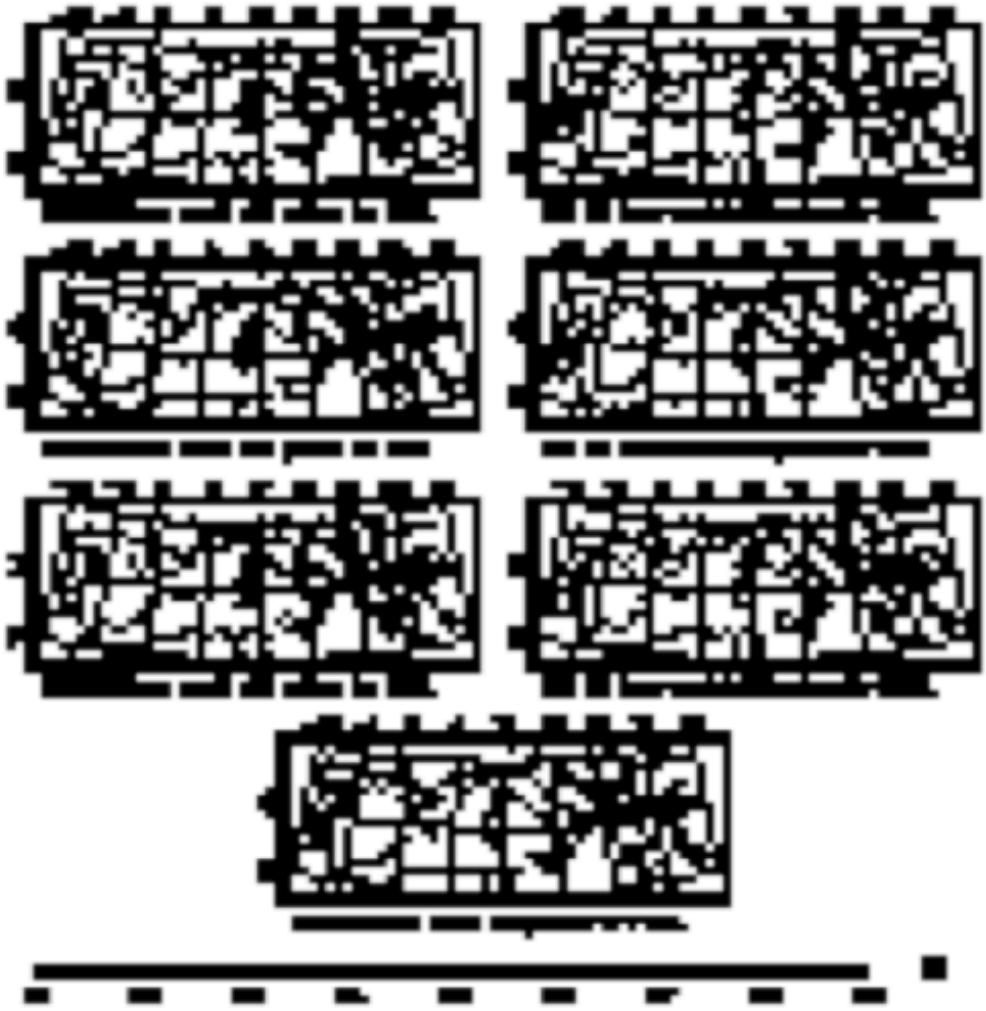

Figure 4.49: LBF autumn precipitation performance, calculated as model skill (r).

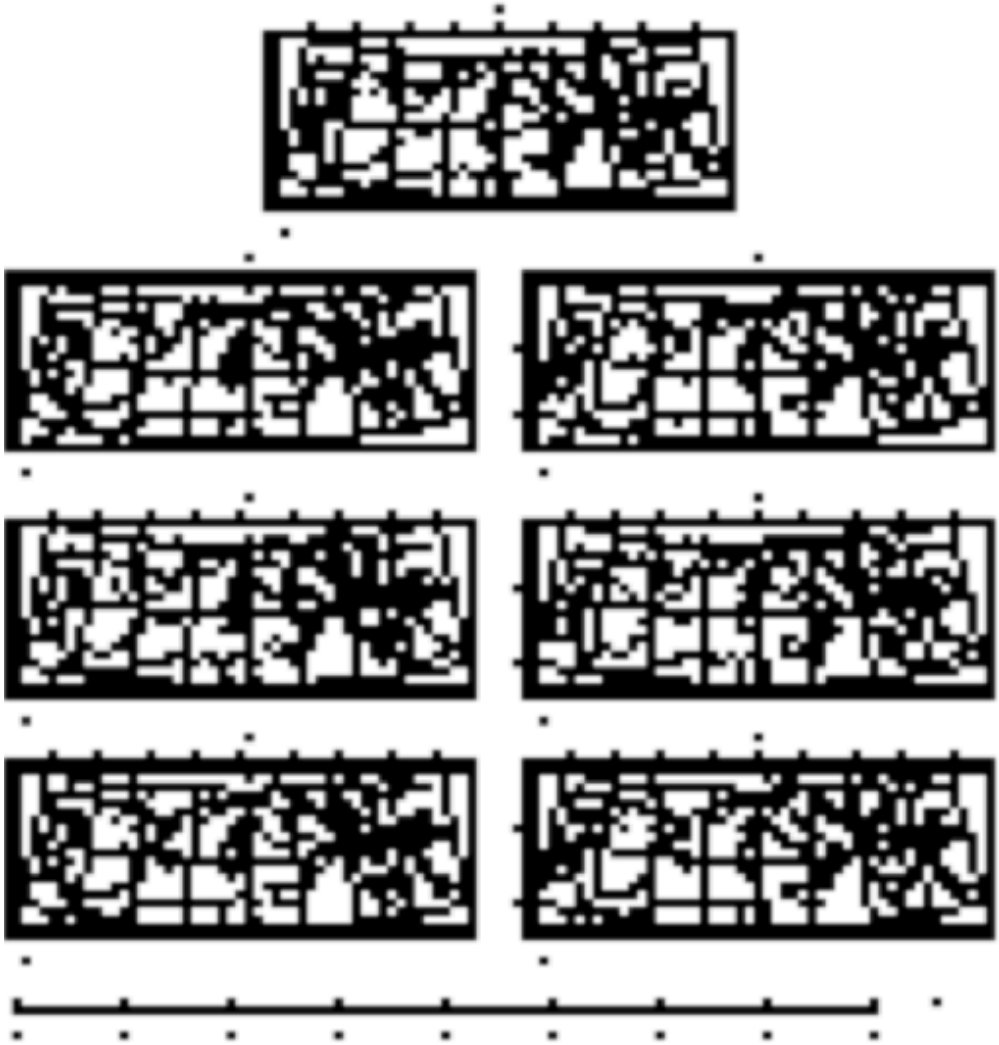

Fig 4.50

Figure 4.50: GBF autumn temperature performance, calculated as model skill (r).

Fig 4.51

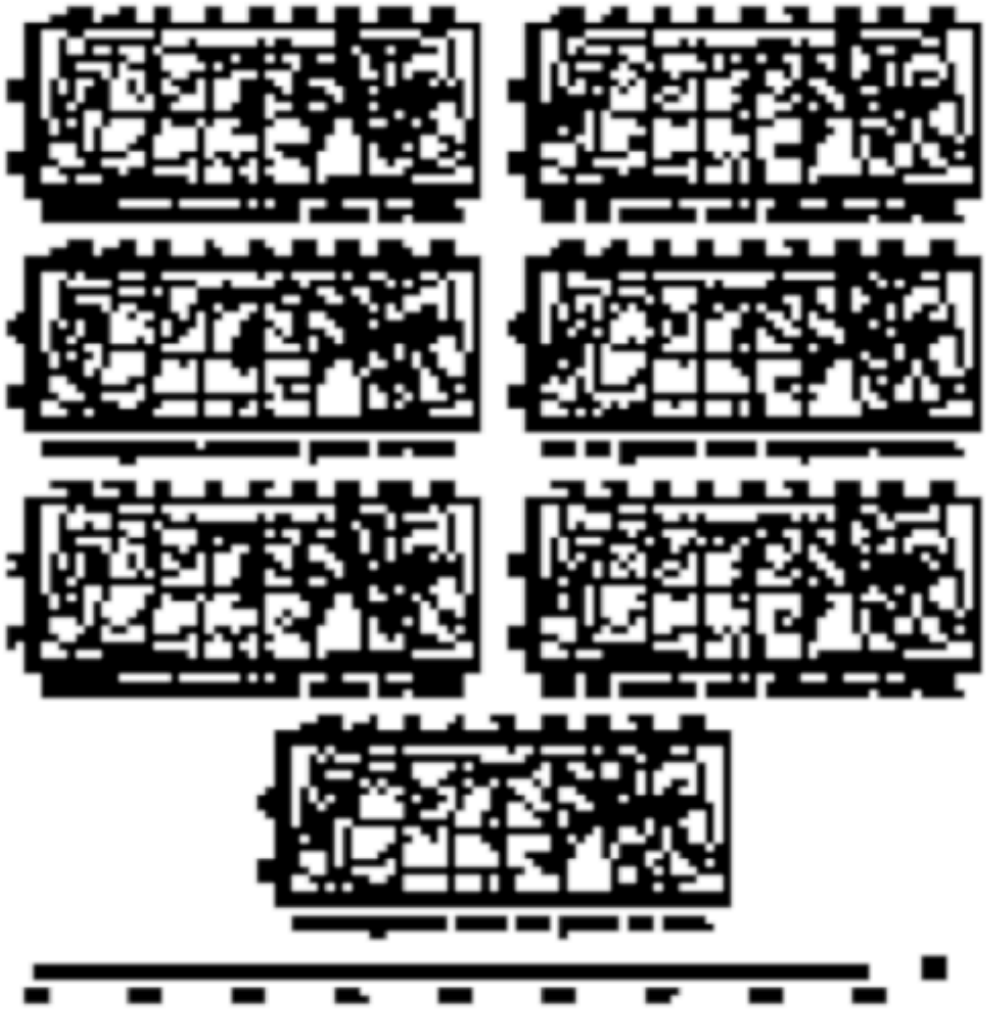

Figure 4.51: GBF autumn precipitation performance, calculated as model skill (r).

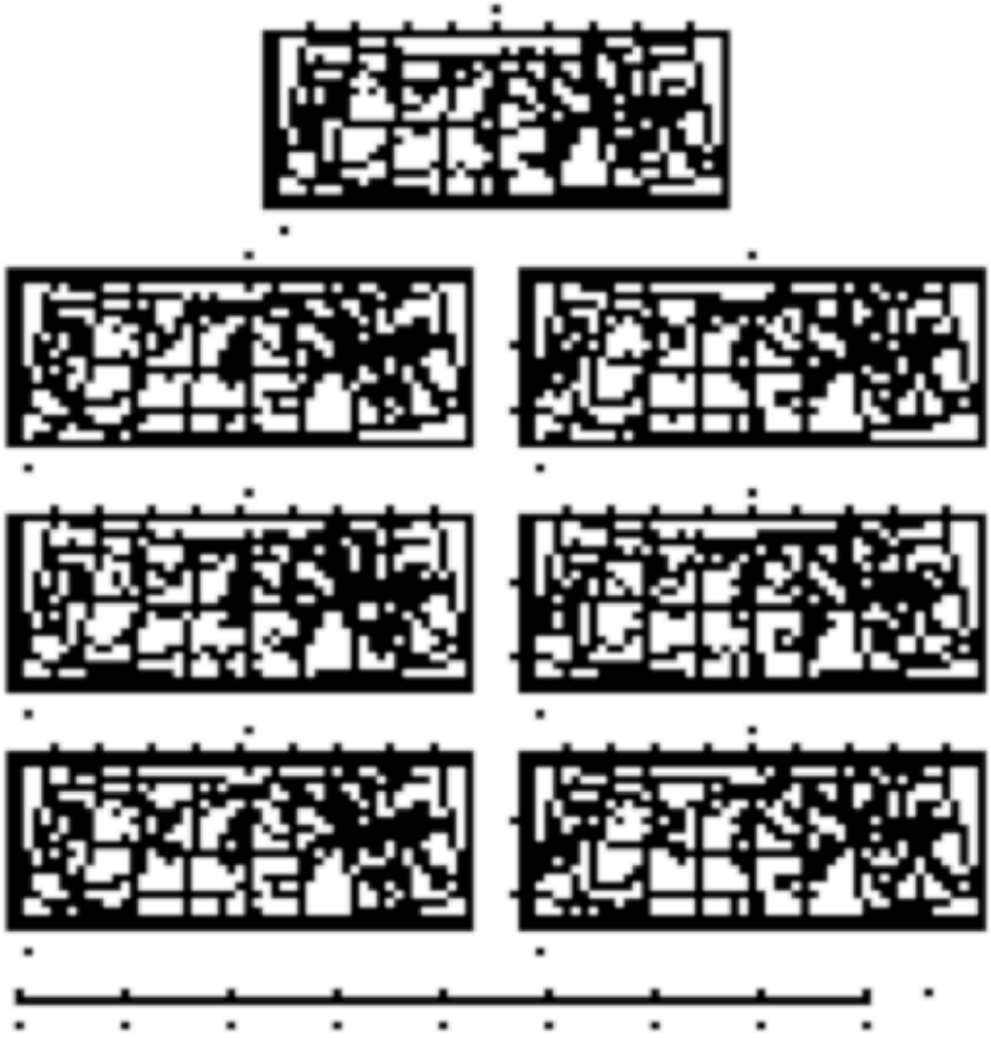

Fig 4.52

Figure 4.52: TBF autumn temperature performance, calculated as model skill (r).

Fig 4.53

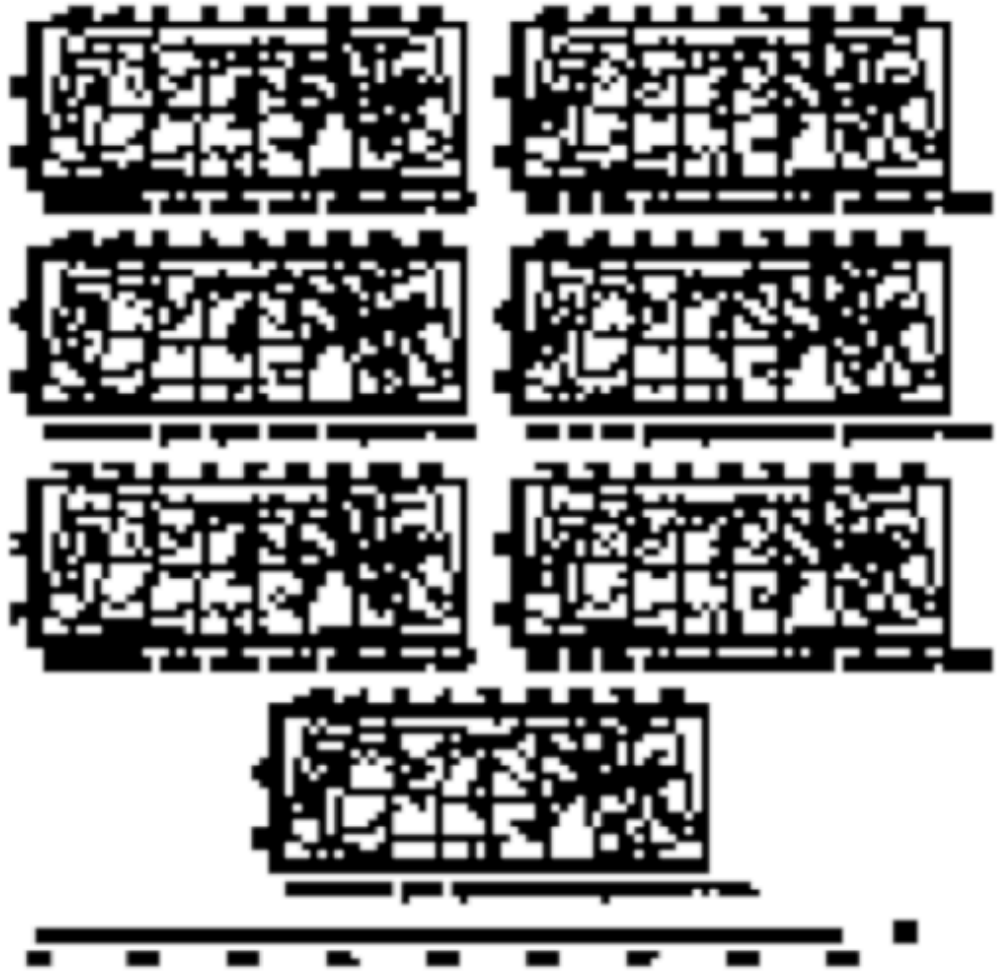

Figure 4.53: TBF autumn precipitation performance, calculated as model skill (r).

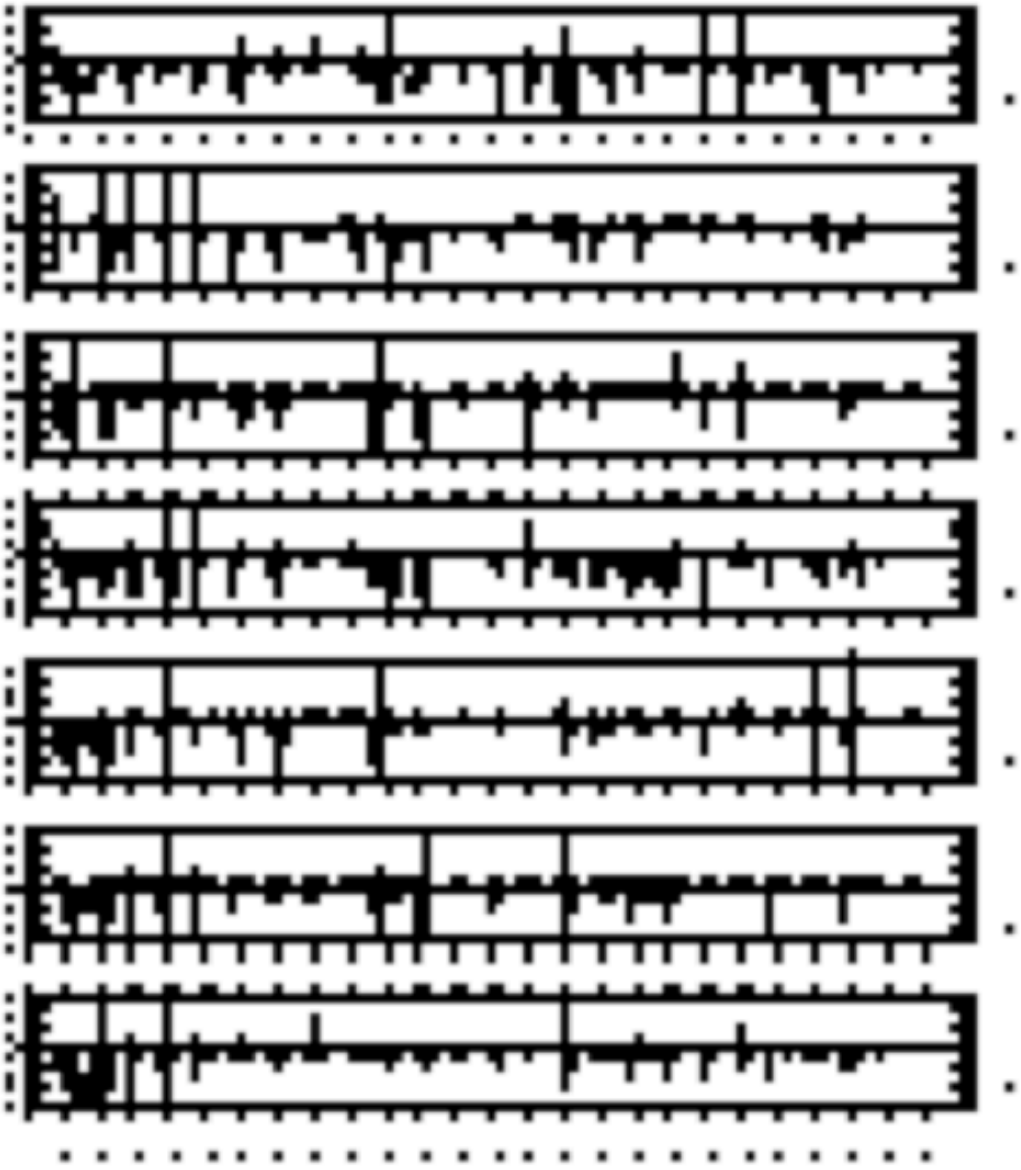

Fig 4.54

Figure 4.54: LBF ANN climate model predictor weights for the western region and temperature predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for TAVG are normalised by the overall winter TAVG mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.

Fig 4.55

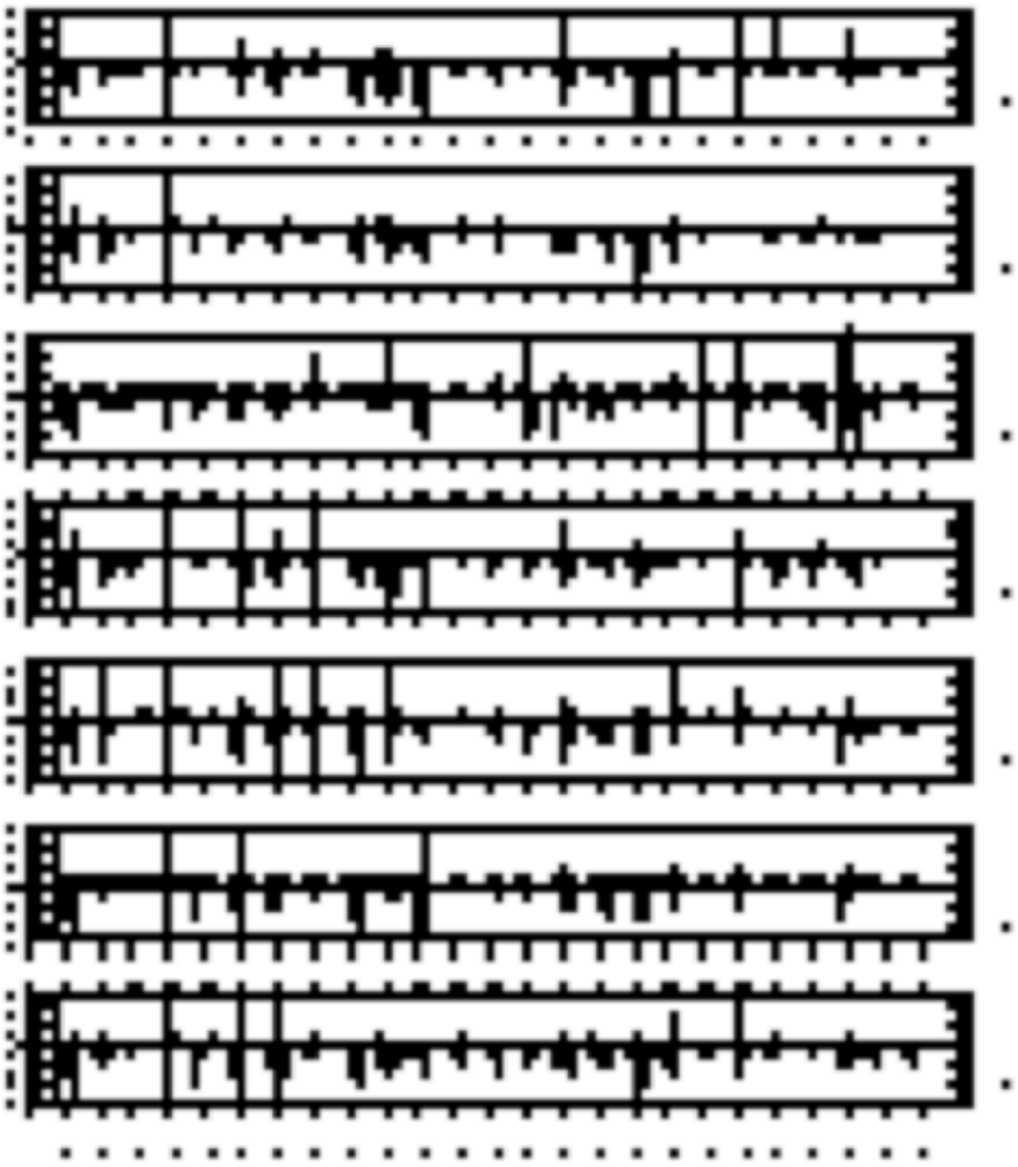

Figure 4.55: LBF ANN climate model predictor weights for the western region and precipitation predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for PREC are normalised by the overall winter PREC mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.

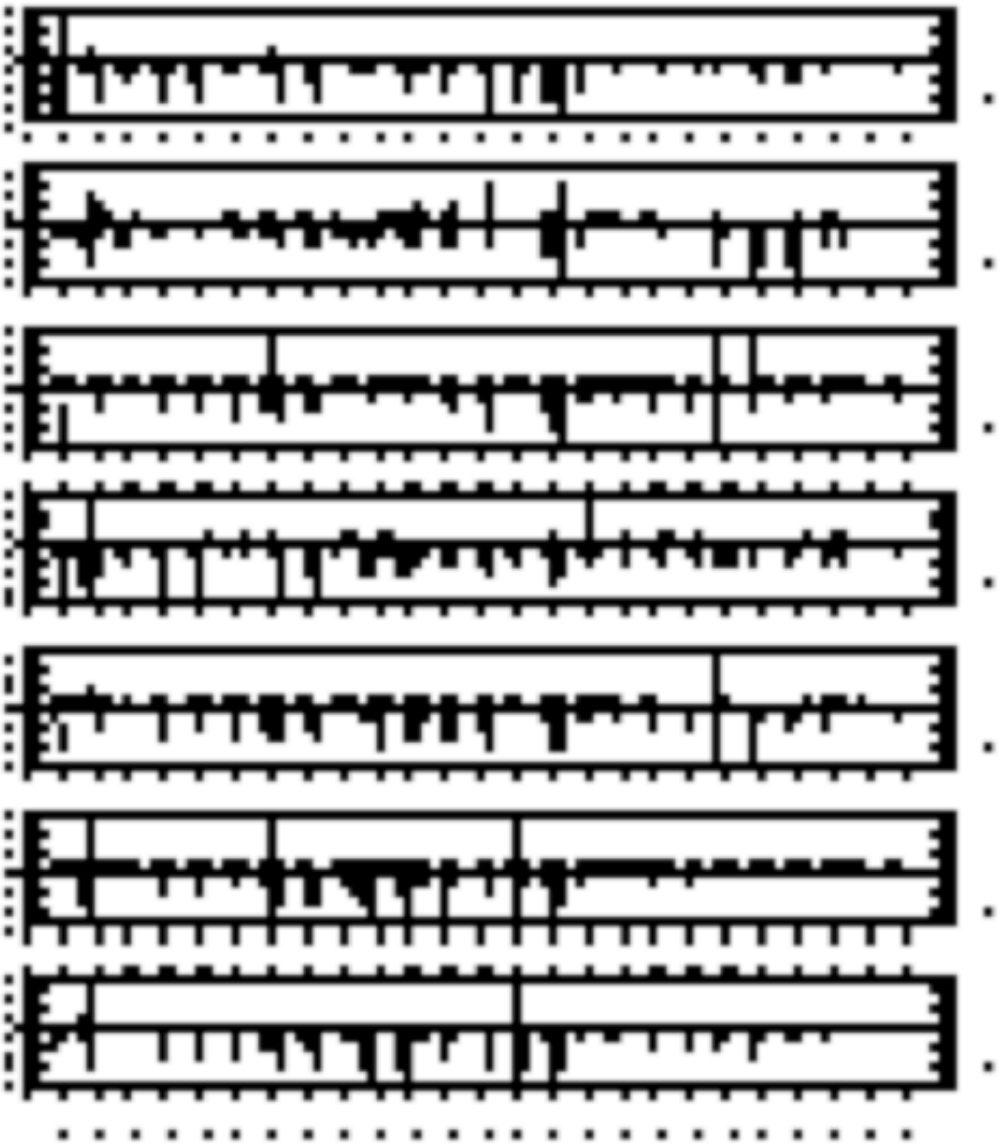

Fig 4.56

Figure 4.56: LBF ANN climate model predictor weights for the central region and temperature predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for TAVG are normalised by the overall winter TAVG mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.

Fig 4.57

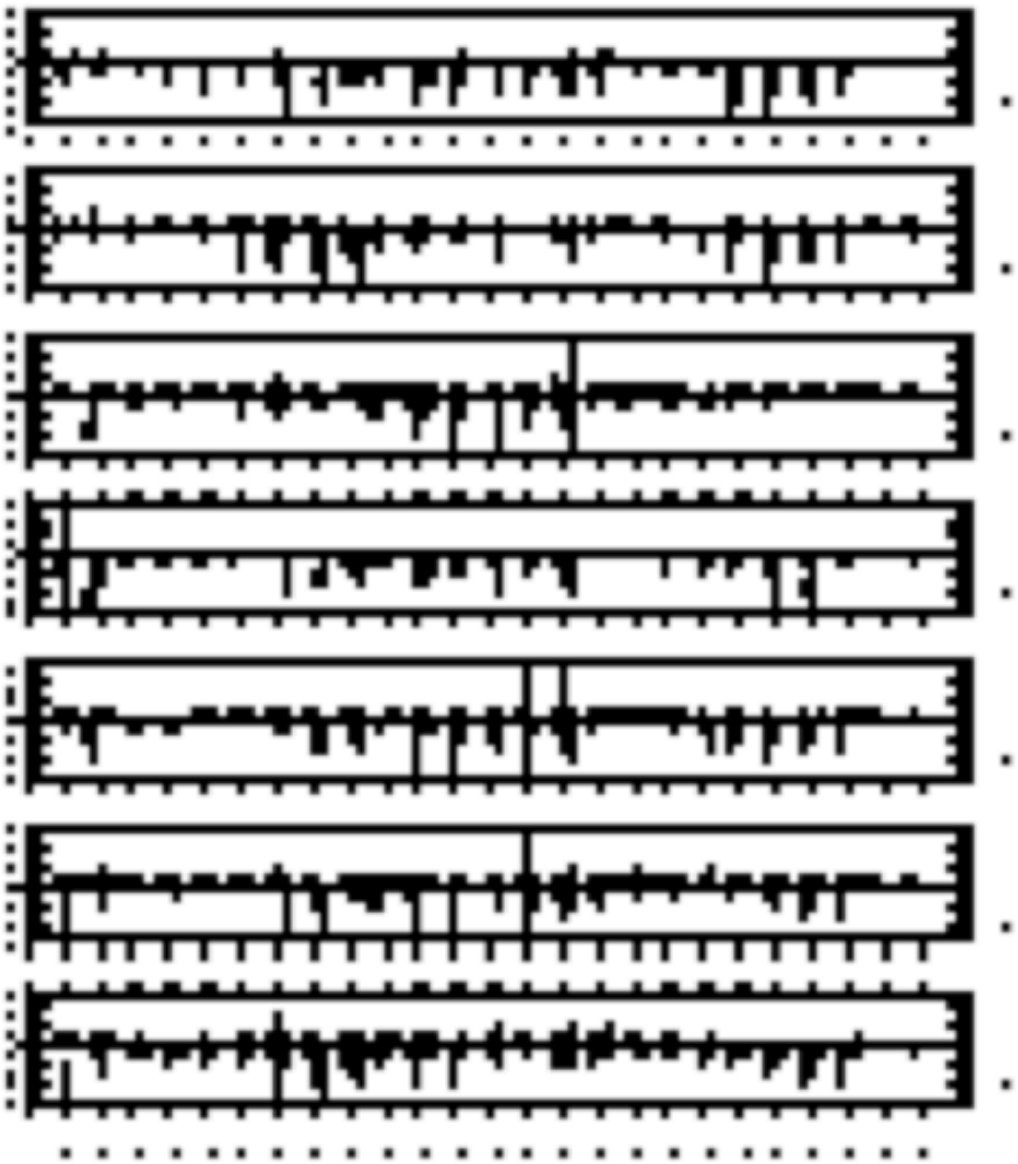

Figure 4.57: LBF ANN climate model predictor weights for the central region and precipitation predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for PREC are normalised by the overall winter PREC mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.

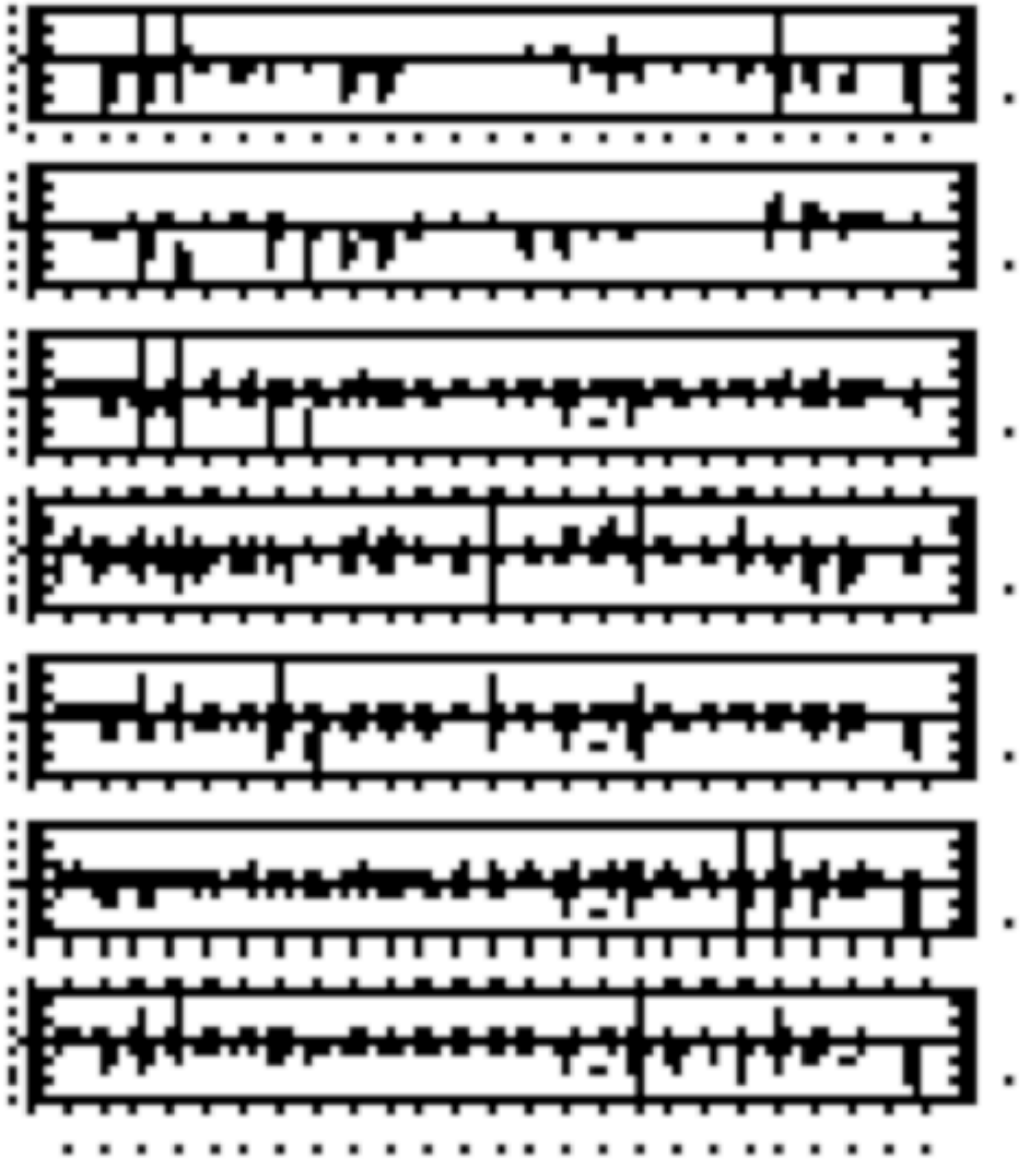

Fig 4.58

Figure 4.58: LBF ANN climate model predictor weights for the eastern region and temperature predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for TAVG are normalised by the overall winter TAVG mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.

Fig 4.59

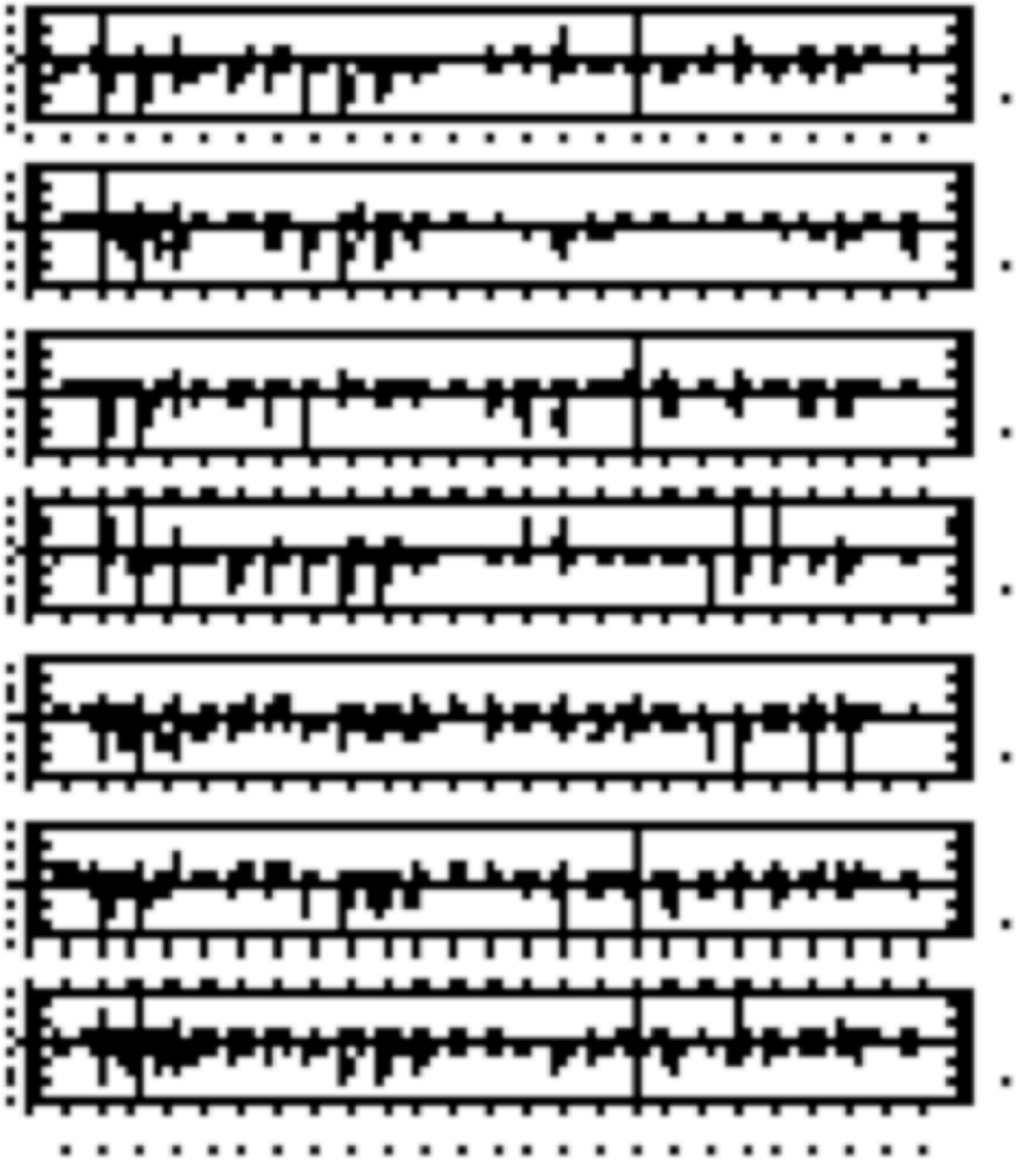

Figure 4.59: LBF ANN climate model predictor weights for the eastern region and precipitation predictand indices. Box and whisker bars include weights for all western stations (between -10° and 6° latitude), and are shown individually for winter (dark blue), spring (green), summer (orange), and autumn (light blue), and each predictor (as shown on the x-axis). For ease of comparison all values are normalised seasonally (e.g. all winter predictor weight values for PREC are normalised by the overall winter PREC mean and standard deviation). See Tables 3.4 and 3.11 for acronyms.