Vol. 10: 83–93, 1998 | CLIMATE RESEARCH Clim Res | Published August 14

Linear versus nonlinear techniques in downscaling

Andreas Weichert[1,*], Gerd Bürger[2]

[1]Fachbereich 8 Physik, Carl von Ossietzki Universität, D-26111 Oldenburg, Germany
[2]PIK, Pf. 601203, D-14412 Potsdam, Germany

ABSTRACT: Standard linear and nonlinear downscaling models are compared using identical atmospheric circulation forcing fields. The target variables chosen were observed daily values of average temperature (TAV), precipitation (PRC), and vapor pressure (HPR) at a Central European station. Being without much sophistication, both models show acceptable performance on this time scale only for TAV and HPR; PRC, which behaves in a predominantly nonlinear fashion, handled very poorly. By considerably refining the evaluation it is nevertheless possible to distinguish significant differences between the 2 models and, with the nonlinear model, to describe specific rainfall conditions. We argue that this difference is caused by the limitations of the linear approach, and discuss how this might affect the downscaling of nonlinear quantities in general.

KEY WORDS: Downscaling · Neural nets · Linear · Nonlinear

1. INTRODUCTION

One of the most serious obstacles for sound and thorough climate impact research is the treatment of extreme, predominantly hydrological, events. Such events have a very great impact on many ecosystems, and can wipe out in a single stroke the agricultural and economic development of a whole year (and longer). Therefore, the adequate modeling of the statistics of extreme events is an indispensable part of downscaling studies.

By definition, an extreme event occurs with small probability, and so its empirical statistical properties have to be determined quite exactly. Small estimation errors create artificial extremes or restrict large model extremes, with enormous consequences for the ecosystem. In the case of a precipitation scenario, for example, the omission of only a few heavy showers can lead to an overall *decrease* of precipitation, even if the general mean trend is *increasing*, and vice versa.

A major problem for today's climate impact studies is how to utilize the climate change information that is produced by General Circulation Models (GCMs). Since those models are not designed to describe the smaller space and time scales of our daily environment, a 'downscaling' mechanism has to be added which transmits information from the global models to the local scales. A number of tools have been formulated in recent years (Bardossy & Plate 1992, Elsner & Tsonis 1992, Hewitson & Crane 1992, Wilson et al. 1992, Hughes et al. 1993, McGinnis 1994, Bürger 1996) which aim at finding a mathematical relation (a 'transfer function') between the global and local fields that is stable in time and, moreover, valid for a variety of different climates. For many quantities, such as temperature and pressure, this relation can be expressed by a simple linear relation between the global and local anomaly fields. Unfortunately, the ecologically most important fields, to which all hydrological phenomena belong, do not exhibit such simple linear relations to the global circulation. For example, the generation of convective storms in a mountainous region or the occurrence of fog over wetlands is too complicated to be linearly predictable from the large-scale flow; these kinds of interaction between circulation and surface cannot be comprised in a simple mathematical form. A standard way to include even complicated, nonlinear interactions in the transfer function is to use a neural net (NN; see Bishop 1995). By decomposing the transfer function into more basic parts, NNs are capable of

*E-mail: weichert@cip.physik.uni-oldenburg.de

© Inter-Research 1998

approximating a wide range of functions, including even strongly nonlinear ones. Although specifically suited for this latter case, a NN still needs a large number of parameters, and it requires large training sets to produce reliable estimates. This usually marks the limit of applicability of NNs.

An alternative to the use of transfer functions is the 'analogue method' (Zorita et al. 1995). By selecting the weather elements from the already observed record (taking, for a given model circulation pattern, an 'analogue' observed circulation pattern) the original time series is re-ordered according to the model circulation. Using this method, no new weather scenarios are generated but, at least, realistic local variability is achieved.

The current study is meant to deepen our understanding of the performance of both methods. It has been demonstrated in other studies, e.g. McGinnis (1994), that, in fact, the use of NNs improves the downscaling of hydrological quantities. (Note that since NNs usually comprise linear models their performance cannot decrease relative to those linear models.) In that study, 5-day averages of the 70 kPa height field and snowfall amount were related, as measured at a number of stations in the western part of the U.S. The use of 5-day averages, of course, improves the downscaling performance, as it does not have to be concerned with, for example, extreme hydrological events, which only occur at the daily time scale.

Following up on McGinnis (1994), we discuss a number of questions which remained unresolved in that study. For example: Does the improvement that is gained by the inclusion of nonlinear methods occur *uniformly* over the whole phase space, or specifically for certain regions, thereby distinguishing processes which are captured by the NN only? What we have in mind is, for example, a convective instability under hot and humid conditions which often leads to convective storms. It is known that, on the physical level, a number of highly nonlinear, small-scale processes are responsible for the evolution of such storms. But these processes cannot be reflected in the coarse resolution of the GCM. On the other hand, if NNs perform better, and our study demonstrates that they do, it is caused by the nonlinear coupling of the global forcing fields.

But which situations are responsible for the different behavior of linear and nonlinear models? Is it possible to distinguish specific circulation patterns which are modeled differently? These questions are very important especially with respect to climate change scenarios. If there is a tendency to more extreme events in a warmer climate (as climate change is often depicted in the media) there is a chance that NN downscaling much better reflects such events, as they are possibly caused by exactly those nonlinear processes.

We emphasize that the models used are designed solely for the described comparison, and have not been streamlined towards any performance optimum. While this is certainly possible, and in fact necessary once they go into practical operation, the main interest of this, mainly theoretical, study lay in the improvement gained from using nonlinear models. But before we enter the comparison in full detail let us first briefly introduce each model.

2. DOWNSCALING

'Downscaling' includes any attempt to model a small-scale (local) process, $\mathfrak{l}$, by means of a series of large-scale (global) states, $\mathfrak{g}$, establishing a mathematical relation of the form

$$\mathfrak{l} = f(\bar{a},\mathfrak{g}) \tag{1}$$

with a transfer function f that has to be determined empirically. Usually, one has an *a priori* knowledge about the main class the function belongs to, such as linear or logarithmic, and so the main purpose of Eq. (1) is to fit the set of parameters, $\bar{a}$. Since the data are usually noisy it is important to keep the number of parameters small to avoid fitting the noise.

2.1. Linear downscaling

For linear downscaling, the transfer function is chosen to be a simple linear matrix, **L** (whose entries are the parameters to be fitted):

$$\hat{\mathfrak{l}} = \mathbf{L}\mathfrak{g} \tag{2}$$

It is assumed that $\mathfrak{g}$ and $\mathfrak{l}$ are given as anomalies about some previously determined (daily, monthly, annual) mean values. Depending on how **L** is chosen a whole variety of downscaling methods exist. For example, one way to choose **L** is to determine global and local patterns, $\mathfrak{g}_0$ and $\mathfrak{l}_0$, respectively, such that the respective projections of the global and local fields onto these patterns are maximally correlated. **L** is then given by

$$\mathbf{L} = \rho\,\mathfrak{l}_0\mathfrak{g}_0^T \tag{3}$$

with ρ denoting the correlation. This approach is called 'canonical correlation analysis' and is probably the most frequently applied form of linear downscaling; examples are von Storch et al. (1993), Werner & von Storch (1993), Gyalistras et al. (1994) and Heyen et al. (1996). An alternative is the so-called 'best linear predictor'. Here one tries to minimize the modeling error, $\mathbf{L}\mathfrak{g} - \mathfrak{l}$, in terms of its covariance. The appropriate cost function can be expressed as

$$\Gamma(\mathbf{L}) = tr\left(\langle(\mathbf{L}\mathfrak{g} - \mathfrak{l})\rangle\langle(\mathbf{L}\mathfrak{g} - \mathfrak{l})\rangle^{\mathrm{T}}\right) \tag{4a}$$

(where $\langle\rangle$ denotes expectation); or

$$\Gamma(\mathbf{L}) = tr(\mathbf{L}\mathbf{C}_{\mathfrak{g}\mathfrak{g}}\mathbf{L}^{\mathrm{T}} - \mathbf{L}\mathbf{C}_{\mathfrak{g}\mathfrak{l}} - \mathbf{C}_{\mathfrak{l}\mathfrak{g}}\mathbf{L}^{\mathrm{T}} + \mathbf{C}_{\mathfrak{l}}) \tag{4b}$$

where $\mathbf{C}_{\mathfrak{l}\mathfrak{g}}$ etc. denote the various covariance matrices. The condition of a local extreme, $\partial\Gamma/\partial\mathbf{L} = 0$, gives the solution

$$\mathbf{L} = \mathbf{C}_{\mathfrak{l}\mathfrak{g}} \cdot \mathbf{C}_{\mathfrak{g}\mathfrak{g}}{}^{-1} \tag{5}$$

This approach is taken, for example, in Bürger (1996). So far, significant differences between the various linear techniques have not been observed. In most cases, **L** is allowed to depend on the seasonal cycle. For the following, it will be important to note that the model equation, Eq. (2), describes nothing more than a simple projection of the circulation onto one specific pattern, for each local variable (the rows of **L**).

2.2. Neural nets

Neural networks are used in various domains ranging from concept (Boolean function) learning to function approximation. The task considered here, i.e. approximation of the function f as described in Eq. (1), is a type of supervised function learning. Below, we briefly describe our hierarchical variant of the radial basis function (RBF) approach, which has been tested with excellent results on standard chaotic maps and also on natural time series (Fröhlinghaus et al. 1994). Similar methods have been used for a long time for approximating nonlinear smooth functions. The RBF method was applied first by Moody & Darken (1989) in conjunction with neural networks and consists of 2 main steps:

(1) unsupervised clustering,
(2) supervised function approximation.

Regarding Step 1: To approximate the function f, the f-domain is first partitioned into various regions, using the following, recursive procedure: starting with the whole input space, the f-domain is divided with a hyperplane into 2 regions. Each region is split once again in the next step. This procedure is repeated until a defined break condition is met. The hyperplanes are calculated using only the statistical correlations of the input (predictor) variables. The normal of the hyperplane is the dominant principal component (PC) of the data-points in that region and it points to the corresponding mass-center. The so-generated subsets of the data-points are called clusters, denoted by C_j for cluster j.

It is known that PCs form an orthogonal system which minimizes the coordinate entropy S (Watanabe 1985); the variance of the data distribution is strongly peaked in a few, say d, directions. If in this coordinate system $\mathfrak{g} = (g_1, g_2, ..., g_d)^{\mathrm{T}}$, the coordinate entropy S_j for cluster j is defined as:

$$S_j = -\sum_{i=1}^{d} p_i \log p_i \quad \text{with} \tag{6}$$

$$\sigma_i^2 = \left\langle \left(g_i - \langle g_i \rangle_{\mathfrak{g}\in C_i}\right)^2 \right\rangle_{\mathfrak{g}\in C_j} \quad \text{and} \quad p_i = \frac{\sigma_i^2}{\sum_{l=1}^{d} \sigma_l^2} \tag{7}$$

The dominant PC is the direction with the greatest variance. Hence, our dividing strategy extracts the maximal amount of coordinate information from the data distribution when only a single cut-plane is allowed.

Once a binary tree is constructed up to some predetermined depth, each tree level defines a set of clusters covering the whole f-domain. The tree itself can be regarded as a compact representation of the data distribution. In the hierarchical method all cluster levels are used. Because this approach did not significantly improve the performance in the present task, we use only the cluster at the end of the binary tree.

The clusters are separated by hyperplanes, so that they form polyhedra. This structure can be represented as a multilayer perceptron (Hecht-Nielsen 1990) with 1 hidden layer. The neuron j in the output layer corresponds to a characteristic function $\chi_j(\mathfrak{g}) = 1$, if $\mathfrak{g} \in C_j$, 0 otherwise. This function is further smoothed depending on the chosen form of the sigmoid activation function. The RBF approach uses a similar approach but approximates the data-points in a cluster by a multivariable Gaussian distribution. The RBF uses a characteristic function of form

$$R_j(\mathfrak{g}) \approx \mathrm{e}^{-(\mathfrak{g}-\mathfrak{g}^j)\cdot\eta\cdot\mathbf{C}_j^{-1}\cdot(\mathfrak{g}-\mathfrak{g}^j)} \tag{8}$$

where $\mathfrak{g}^j$ is the mass-center and $\mathbf{C}_j$ the covariance matrix of cluster j. The scaling factor η is a free parameter whose value defines the overlap between different clusters. Characteristic functions of the form of Eq. (8) are known as radial basis functions, which are usually 'normalized' as:

$$P_j(\mathfrak{g}) = \frac{R_j(\mathfrak{g})}{\sum_{j=1}^{M} R_j(\mathfrak{g})} \tag{9}$$

where M denotes the number of clusters.

Regarding Step 2: Once the radial function basis is set up, one proceeds with supervised learning, which means here fitting a linear combination f of the cluster radial functions by minimizing the mean squared error of the training set with N samples:

$$E(f) = \frac{1}{2}\sum_{i=1}^{N}\left(|\mathfrak{l}^i - f(\mathfrak{g}^i)|\right)^2 \tag{10}$$

The function f describes the output of the entire neural network with M clusters as:

$$f(\mathfrak{g}) = \sum_{j=1}^{M} \mathbf{L}_j(\mathfrak{g}) \cdot P_j(\mathfrak{g}) \tag{11}$$

Following Farmer & Sidorowich (1987) the fit might include also (a linear) interpolation term

$$\mathbf{L}_j(\mathfrak{g}) = \alpha_j + \mathbf{B}_j \cdot \mathfrak{g} \tag{12}$$

where α_j is a vector and $\mathbf{B}_j$ a matrix. The training procedure has several global parameters: the number of clusters (the depth of the tree), η, the overlap between clusters, and whether one applies linear interpolation or not. Usually, it is best to start with the simplest networks containing one cluster (which in our case corresponds to a linear regression) and proceed towards more complex ones. By monitoring the improvement in performance of a data-set which has not been used for training, one can avoid overfitting due to an excess of fit parameters. In all cases discussed in this paper we took the fixed value $\eta = 5$, while the parallel use of different tree levels only marginally improved the prediction. Similarly, the Farmer-Sidorowich linear interpolation was not needed. Hence, the only major global network parameter was the number of used clusters. For the downscaling problem we used 113 clusters; a minimal cluster size of 40 points was used as the clustering break condition. This produces a nonbalanced tree with deeper depth (more clusters) in the regions with large point density.

3. TESTING THE MODELS

For a comparison of the linear and nonlinear models we used the linear model that is based on the best linear predictor, here denoted as BLP, and the NN which uses radial basis functions, denoted RBF.

Our test case is made up of the following components: As a predictor field, we chose a combination of the geopotential height field at 50 kPa, z_{50}, and the temperature field at 85 kPa, T_{85}, both in a $5° \times 5°$ grid box with corners 55° W, 25° N and 45° E, 75° N. We used daily values from the period 1 October to 30 June 1989. The fields originate from an updated version of an older NCAR analysis project (see Mass et al. 1987). From the PCs of each individual field, the combination uses the first 14 and 25 entries, respectively, which each explain 90% of the total field variance; the annual cycle was removed beforehand. In this way the large-scale North Atlantic/European circulation plus temperature is condensed into a 39-dimensional time-series. On the local side, we experiment with 3 variables: TAV, the average temperature, PRC, the precipitation amount, and HPR, the vapor pressure, as measured at the climate station in Potsdam in Germany (13° E, 52° N). TAV and HPR are examples of variables which behave fairly linearly, with distributions that are essentially normal. PRC, in contrast, shows a quite nonlinear, frequently 'switch-like', behavior and has, correspondingly, a rather skewed distribution function. Consequently, PRC is a quantity whose behavior and importance for ecosystems are not so much determined by the evolution of its mean but by the occurrence or non-occurrence of extreme values or, as for rain, the intermittency characteristics. TAV and HPR are, on the other hand, quantities which are mainly characterized by the evolution of their means.

Both models were calibrated using the first 6000 dates, and validation was performed with the remaining period, consisting of 3770 dates; in this way it was guaranteed that no artificial skill was left in any of the models.

We do not model TAV, PRC, and HPR directly. As mentioned above, PRC is an example of a quantity that is highly non-normally distributed. But since the predictor fields are fairly normal, the modeled PRC would strongly deviate from its observed counterpart. Therefore, it is recommended before any modeling that each of the 3 quantities be normalized. Being rather important, one might as well call this a model step in itself. We do not go into the details of this procedure, especially regarding the treatment of boundaries; for this, see Bürger (1996). We only mention that the procedure is based upon the *probit* (**prob**ability **i**ntegral **t**ransformation) theory, which, in a rather general sense, maps a given random variable 1-1 to a variable which is normally distributed with mean 0 and standard deviation 1 [$N(0,1)$ distribution]. We note that for boundary values, such as 0 for PRC, a *threshold* value is assigned to the normalized variable such that any occurrence of the boundary corresponds to a randomly chosen normal variate below (or above, for an upper bound) the threshold. All seasonality is removed from the normalized fields (it is contained in the monthly probit parameters). However, this does not mean that there is no seasonality left in the normalized fields, as is shown below. Nevertheless, in terms of simple correlations there is practically no improvement if one models each season separately.

In Fig. 1 we see the outcome of the modeling as a scatter-plot of observed and modeled normalized values, together with a quadratic regression line indicating the slope and unresolved nonlinearities. Note that, in 1-dimensional regression for normalized variables, this slope equals the correlation. We note that for both models the quadratic regression part is fairly small, with the exception of PRC (see below). The most obvious feature is the small PRC correlation for both mod-

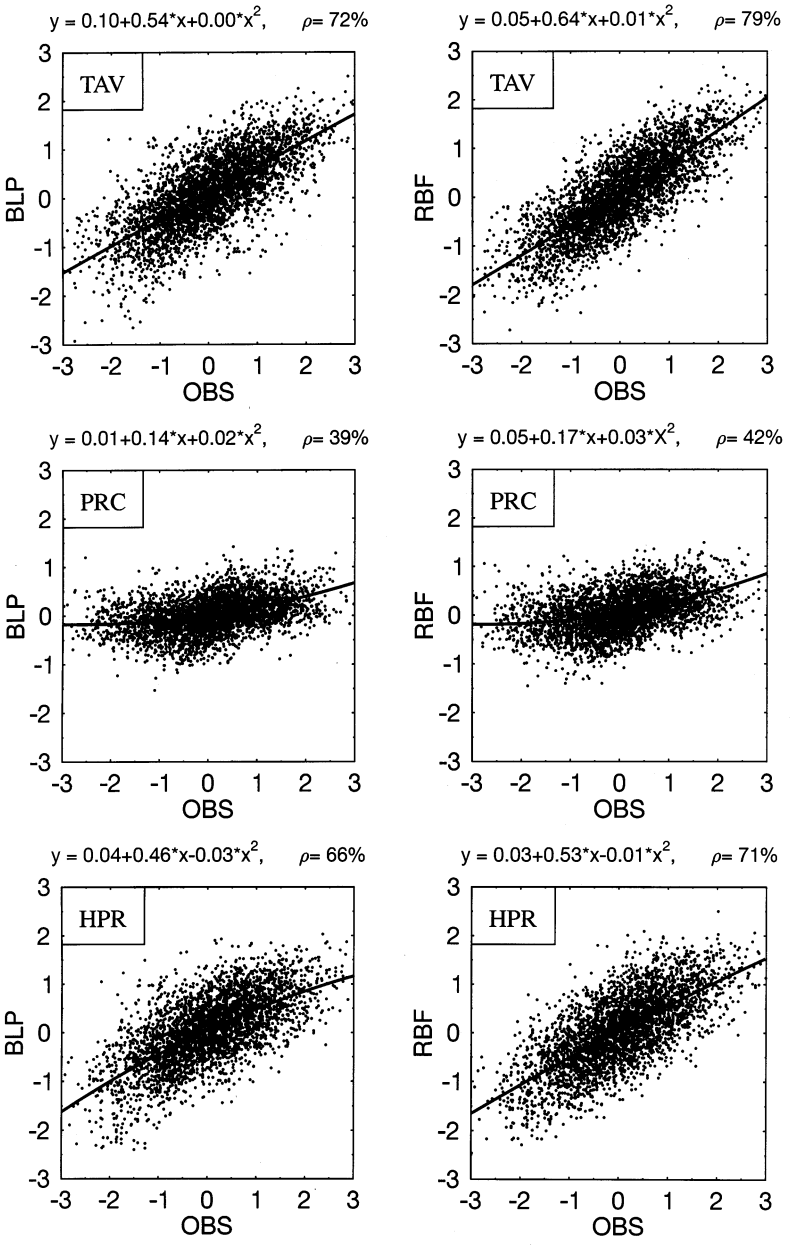


Fig. 1. Scatter plot of modeled and observed normalized variables, together with a regression line. Model performance is generally much better for TAV (upper panels) and HPR (lower panels), with a correlation of 72 and 66%, respectively, for the linear model BLP (left panel) and 79 and 71%, respectively, for the neural network model RBF (right panel). There are some indications that the linear modeling of positive anomalies is better. PRC modeling is much worse than for the other parameters, with correlations of 39% (BLP) and 42% (RBF)

els ($\rho \approx 39\%$ for BLP and 42% for RBF), indicating the general difficulty of PRC downscaling. Consequently, the modeled PRC amplitude is very low throughout, for both models. The slope and correlation are much higher for the other quantities, even for the linear model, with $\rho \approx 72\%$ for TAV and 66% for HPR; the performance of RBF is 5 to 7% greater. For these quantities, it can be concluded that model performance, even on a daily basis, is quite satisfactory, although extreme peaks of TAV and HPR certainly cannot be modeled. But as we have mentioned before, this is less important in comparison to the evolution of mean values. A shift in the climatological mean by about 20 to 30%, as is indicated by the correlation of 70 to 80%, can have a dramatic impact on many ecosystems.

Are there specific *scales* in Fig. 1 where modeling is of different quality? There are vague indications that both models, but especially BLP, perform slightly better for higher TAV values, as the clouds are more scattered at the negative end. For HPR, the negative quadratic coefficient for BLP points to some nonlinearities which cannot be resolved by the linear model; BLP performance is slightly better for negative HPR anomalies. For the nonlinear model, scatter and curvature are reduced correspondingly. The PRC situation is different. The visible curvature for both models is caused by the following: below the threshold value (which depends on the month) the ordinate is uncorrelated with the abscissa by construction, hence the slope on that part must be 0. This is a drawback of the (extended) probit transformation when applied to bounded quantities. It is possible to fix this, but this goes beyond the current topic.

3.1. The 'dry' and 'wet' case

Let us now take a short look at precipitation back in the original domain, by rescaling everything via the inverse probit. Precipitation is a typical variable that is usefully grouped into the single parts *occurrence* and *amount*, depending on a certain threshold value (which is usually taken as 0, as above). Modeling the amount is without much value for the models in question since the small amplitudes practically never create amounts larger than 4 to 5 mm. Occurrence, on the other hand, is conveniently measured by giving the rate of correct forecasts (for the validation period). This is 65% for BLP and 70% for RBF. Without any reference to the balance between observed dry and wet

days, these numbers are of little value. If, for example, 70% of all observed days are dry, the trivial forecast 'dry' (or, if it rains more, 'wet') always scores 70%. A measure which respects this observed balance is the so-called true-skill statistic, *TSS*:

$$TSS = q_0 + q_1 - 1 \tag{13}$$

where q_0 and q_1 denote the rate of correct dry (0) and wet (1) forecasts, respectively; it is convenient to give *TSS* in percentages. Note that $-100\% \leq TSS \leq 100\%$, with –100%, 0%, and +100% corresponding to a systematically false forecast, the trivial forecast mentioned above, and a truly perfect forecast, respectively. In our case, above, we find 33% for BLP and 40% for RBF. This is a considerable improvement achieved by the nonlinear model.

It is the existence of this threshold or, in other words, the natural partitioning into 2 separate states (a wet and a dry state) which allows a much more detailed inspection of model performance. To understand this, let us go back to the forcing fields in the multi-dimensional phase-space. The main idea is the following: we are guided by the idea that variations in model performance do not occur uniformly over the whole state space but, rather, are caused by specific physical processes which are captured differently by BLP and RBF and which, therefore, should be identifiable in the global forcing fields. Mathematically, this comes down to the following problem: we are given a set of 3770 points in the 39-dimensional phase-space, each of which bears 3 labels, consisting of the observed value of PRC together with its BLP and RBF version. Now, if we are able to formulate conditions on the point-labels which partition the phase-space into well distinguished regions, then these regions tell us about the physical fields responsible for the different model performances. The following point-labeling is suggested:

Case 1 *(w)*: PRC > 0 and 0 < BLP
Case 2 *(d)*: PRC > 0 and 0 = BLP and 0 < RBF
Case 3: everything else

The '*w*' and '*d*' stand for (1) BLP-wet and (2) BLP-dry + RBF-wet conditions, respectively. Case *d*, for example, simply means that a rainy day is correctly predicted by the nonlinear model, while the linear one misses it. For our question, only the first 2 cases are of interest; in particular we will have to clarify what exactly is meant by 'well-distinguished regions' of the phase space. Each of the above cases can be viewed as a point cloud of a specific color (= label), and the question is: can we identify, in a statistically safe manner, each single cloud, or do we just see a mixed color? From simple 2-dimensional projections of the point clouds (not shown) it is not obvious at all that there exists a significant partition of phase space, since it is very likely that the *w* and *d* clouds are quite mingled. Similarly to the procedure for cluster analysis, we have to measure the 'distance' of 2 clouds and determine if this distance is large enough. A common distance measure is the so-called Mahalanobis distance (Morrison 1967). This is defined to be the following weighted difference between the mean of the *w* cloud and *d* cloud.

$$d^2 = (\bar{d} - \bar{w})\, \mathbf{C}^{-1} (\bar{d} - \bar{w})^{\mathrm{T}} \tag{14}$$

where the bar indicates the mean of the respective cloud, and **C** denotes the pooled covariance of both populations (clouds). We calculated this distance for each season separately, the result of which is shown in Fig. 2 as a vertical black bar.

3.2. Significance of the dry-wet clustering

The question is: is the magnitude of d^2 significantly larger than a d^2 that comes from a random partition, i.e. one that distributes the given set of *w* and *d* labels randomly among the points of the cloud? If the clouds are multivariate, normally distributed clouds, one could use the well-known test statistic $T^2 = n_1 n_2/(n_1 + n_2)d^2$, which is Fisher *F* distributed. If, however, the condition of normality is only approximately satisfied (which is the case with all empirically given samples) the statistic is of less value, especially when the phase-space is large as in our case of 39 dimensions. Therefore, we refrained from using that test. Rather, by running a whole series of random tests we created our own test statistic, in the following manner: starting from the original point cloud with the *w* and *d* labels, we now shuffle the labels randomly and each time calculate the corresponding Mahalanobis distance, d^2. This is done a sufficient number of times (here: 1000) so that the resulting probability distribution of d^2 can be considered reliable and can be compared to the original outcome; the result is shown in Fig. 2 for the 4 seasons. The black bar, i.e. the original d^2 value, is located to the far right for all seasons but fall (SON). Particularly for winter, with an error level ε approaching 0%, it is clear that the corresponding partition of the point cloud is statistically safe; the error level for summer is slightly larger ($\varepsilon \approx 2\%$). For completeness we mention here that, by reversing the role of BLP and RBF in Cases 1 and 2 above, no significant results of this kind can be obtained. Note that the distribution of d^2 exhibits the typical characteristics of the *F* distribution. The difference pattern itself, i.e. $\bar{d} - \bar{w}$, as reconstructed for the z_{50} and T_{85} fields, is shown in Fig. 3 for the 2 seasons. For winter, the most prominent feature is a broad positive z_{50} and T_{85} anomaly in the northern half of the analysis region; below, exactly at the Potsdam site, there is a cell of negative anomalies. In mete-

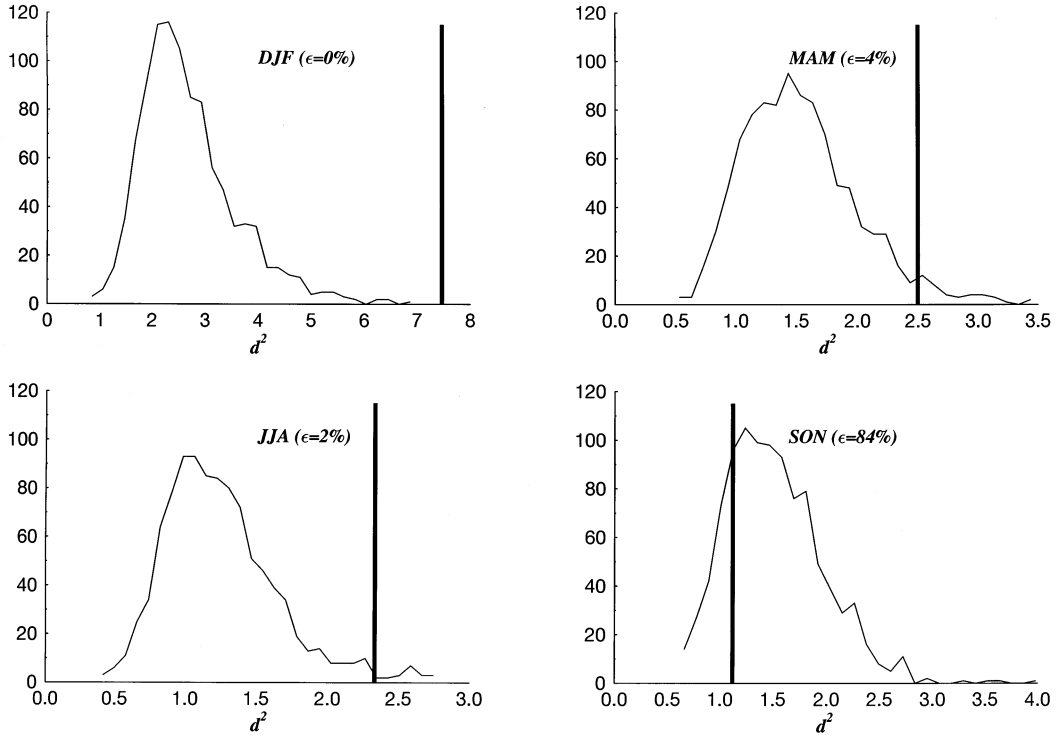


Fig. 2. Distribution of the Mahalanobis distance, d^2, of 2 point clouds that are randomly chosen, together with the value corresponding to the actual choice as determined through the BLP-dry *(d)* and BLP-wet days *(w)* (vertical black bars). The value is significantly non-random at the 5% level for all seasons except fall. For further details see text

orological terms, this means the following: while in the winter season the whole northern part of Europe lies under high pressure and warmer air masses, the Potsdam area is encircled by a body of cold and, probably, by wet air which creates a foggy environment with small amounts of rain that is missed by the BLP model. For summer, hot air to the east of Potsdam and a cold front approaching from the southwest leads to extended convective activity in that area. The spring situation (MAM) is quite similar.

To validate our findings, we examine the local partition of phase-space. For each season, 2 separate histograms of rainfall amounts recorded at Potsdam are shown in Fig. 4, representing the *w* and the *d* case. Since there are many more *w* cases, relative scaling is used to improve visibility, i.e. each histogram adds up to a total of 1. No striking differences are observed for any season. However, we see that for winter, the *d* distribution is a bit more skewed toward smaller PRC values, as compared to case *w*. That means that small PRC amounts are relatively poorly modeled by BLP. For summer (and even more for spring), we find indications of the convective-storm picture: larger amounts of rainfall are missed by the linear model. In fall, where we see a number of strong *w* but no *d* events, the same picture appears, but since the *w*-*d* clustering is not significant in the forcing domain we refrain from drawing any conclusions. As an example of the above, let us take a closer look at the sequence of events of 30 August through 2 September 1977, shown in Fig. 5 (which is not exactly, we know, the calendrical summer). We note that 1 September, with 35.7 mm, showed the 12th largest amount of rain of the entire period of record. A very strong highpressure system slowly moves from the Western Baltic eastward towards Finland, with a kernel anomaly of 150 m. In the west, off the coast of France, low pressure approaches, but is driven backwards as the high-pressure cell is too strong. On 1 September, the heating is maximal (TAV at 22°C), yielding the heavy rainfall event in the Potsdam area. The RBF model captures a small amount of it, 1.3 mm (which is still in the high range of RBF), but BLP misses it completely. After this the situation relaxes for Potsdam.

One might object that the RBF results might have no influence at all on the partitioning of phase-space. But it turns out that, in fact, the patterns become quite different when we merely consider, for example, the cases BLP = 0 and BLP > 0. Nevertheless, that partitioning is significant as well.

This example throws some light on the linear model BLP and one of its caveats: if BLP were able to produce

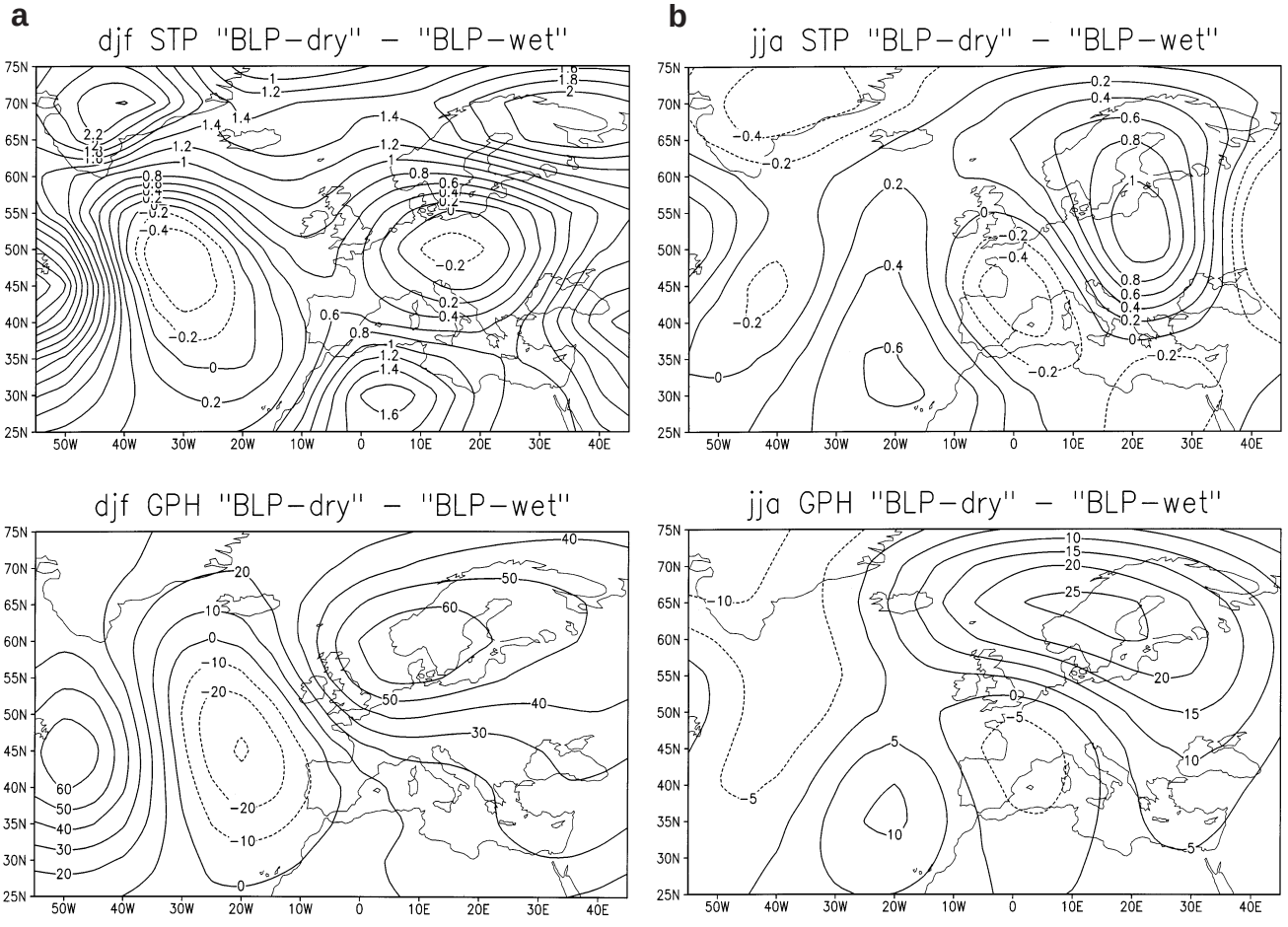


Fig. 3. (a) Difference pattern for winter between the 2 clusters of forcing fields for PRC modeling (see text); 'STP' and 'GPH' correspond to T_{85} and z_{50}, respectively. Positive STP anomalies dominate the entire area, with the exception of 2 negative cells in the Central Northern Atlantic and over Central Europe. The pressure field shows a large dipole with a positive cell over southern Norway and a negative cell off the coast of Spain. This pattern represents a typical cold weather type associated with small amounts of precipitation in the form of snow. (b) Same as (a), but for summer. Anomalously hot conditions over Poland and an approaching disturbance from the southwest favor convective activity over the area of interest, with heavy showers

rainfall for this specific pattern, the opposite situation — that is the one with reversed signs — will inevitably show less or no rain. Yet that situation would be a northwesterly weather regime, which has a well-known strong tendency toward wet conditions for central Europe that would outweigh the pattern of 1 September. While the nonlinear model, RBF, captures this by just adding the circulation pattern to its set of clusters, nothing similar is possible for the linear model: as mentioned in Section 2, linear models are restricted to dealing with a single pattern only.

Summarizing, we might say that, while the nonlinear model does not produce significantly larger amounts of rain than the linear model—and these are much too low—it seems to have a greater sensitivity for the *occurrence* of rain. And, strangely enough, the combination 'RBF-wet' and 'BLP-dry' seems to occur preferably on specific days: days with little precipitation, probably snow, in winter and heavy rainfall events in spring and summer.

4. CONCLUSIONS

In an overall effort to improve the modeling of extreme events for downscaling purposes, this study was meant to deepen our understanding of various obstacles. We have focused on the improvement that nonlinear methods yield in comparison to linear ones. As this is a more theoretical undertaking, we have kept the 2 participating models as simple as possible; each one of them can certainly be improved. Using daily values of 3 crucial meteorological quantities, mean temperature, TAV, precipitation, PRC, and vapor pressure, HPR, the linear model, which is based on the best linear predictor, BLP, is compared to the neural network that uses

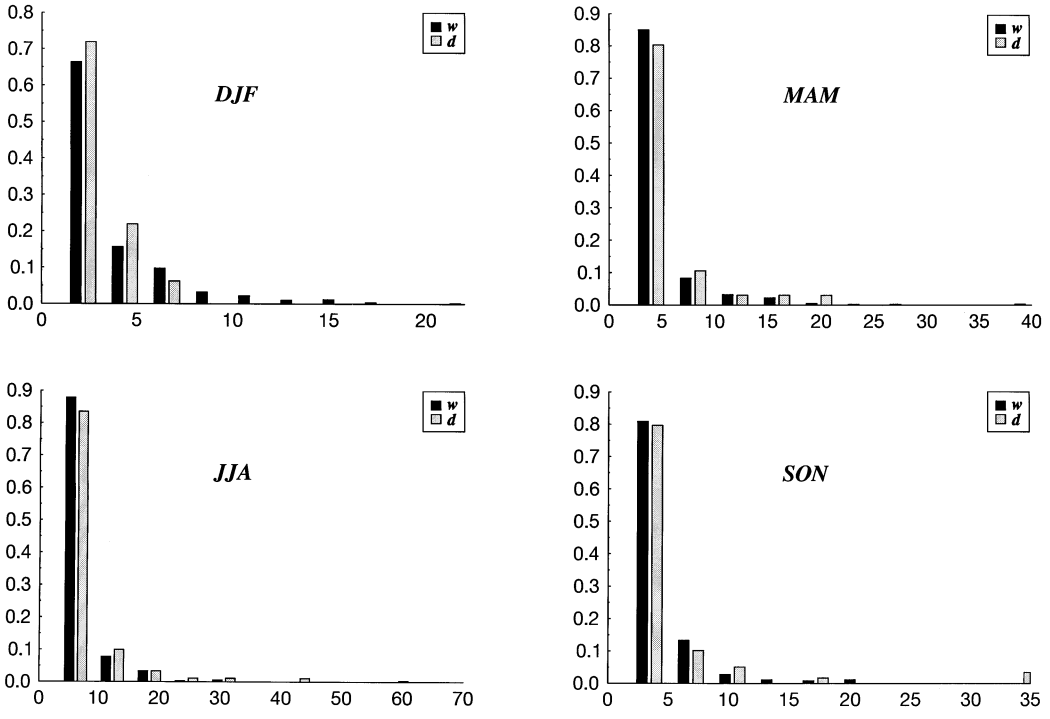


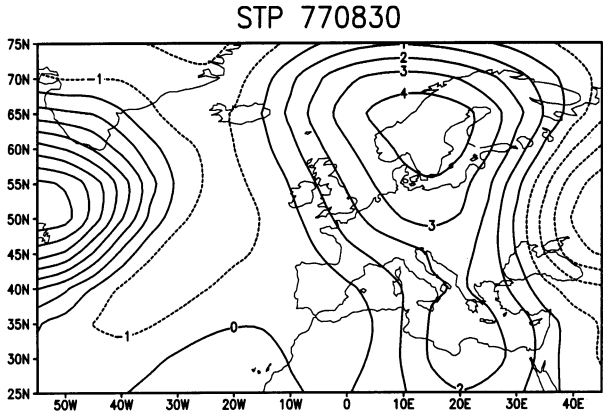
Fig. 4. Histogram of the *d* and *w* days, for each season, in relative units (each adding up to 1). For winter, *d* days show smaller amounts compared to *w* days. In contrast, any other season shows relatively large amounts for *d* days. This indicates that BLP is inherently unable to model heavy rainfall events

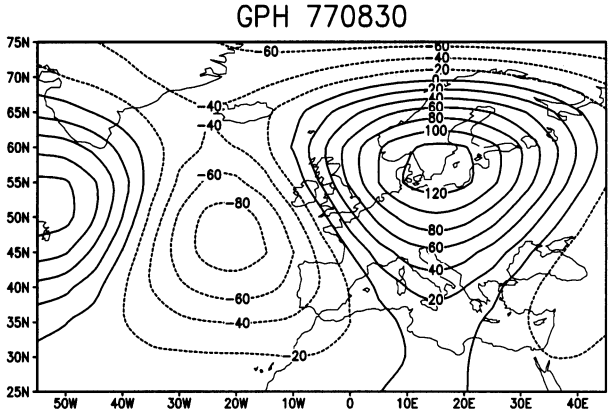
radial basis functions, RBF. The comparison is done in a normalized space since precipitation, in particular, is a highly non-normal quantity while the global predictor variables are fairly normal. The seasonal cycle is incorporated only in the normalization, which itself is based upon a generalization of the probit transformation. The models themselves do not contain any external seasonal parameters.

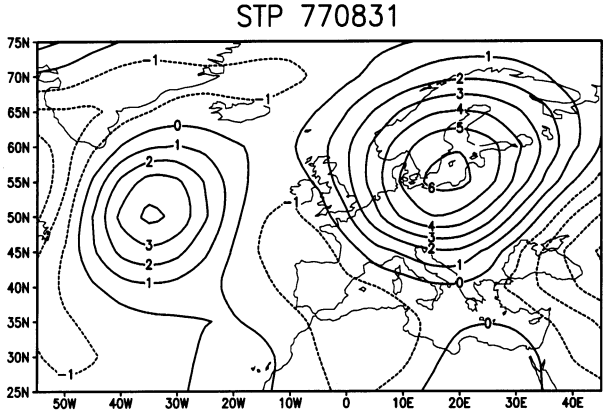
Since the RBF model is more general than the BLP model, it can only improve the overall performance. This is verified in the context of our study through the most simple model-observation correlations, all of which are slightly higher for the RBF. This has the consequence that the modeled variability is larger for this model, but it is still too small to produce realistic extreme events. Although there were some indications that the models are sensitive to different scales, this claim cannot be made consistently and reliably.

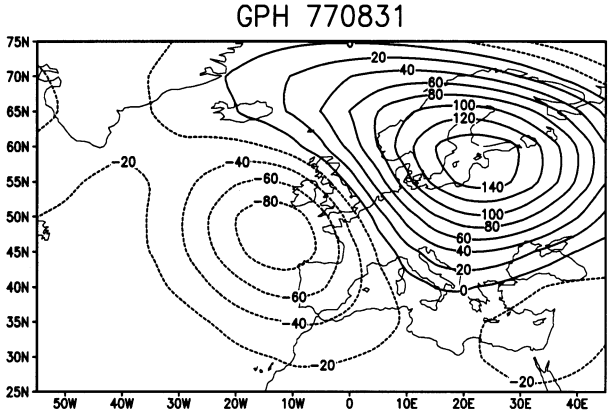
For the most crucial variable, precipitation, modeling is very poor for both models. In the normalized space, one finds a correlation of about 39% for BLP and 42% for RBF. Hence, on this basis, there is no essential difference between the linear and nonlinear approaches. However, by going back to the original forcing fields and local quantities and systematically tracing back the model errors, one finds surprisingly clear differences between the 2 models. As it turns out, these differences are statistically significant for each individual season except fall. In summer, for example, heavy rainfall events which mainly occur during convective storms and which are predominantly nonlinear are often *detected* by the nonlinear model RBF, while the linear model, BLP, misses them completely. Nevertheless, once detected, the modeling of the shower's rainfall amount is still very poor for the RBF model, rarely reaching more than a few millimeters. This renders the model and all related models—in its current form—basically useless for hydrological impact studies which are based on a short time resolution. On the other hand, the RBF performance is very likely to be close to the theoretical limit of all deterministic models, i.e. models that simulate complex dynamical processes by using a few parameters fitted to observations; such models suffer from correlations between the acting fields that are too low, and these models are inherently unable to produce sufficient variability. This can only be remedied by incorporating greater regional detail about the state of the atmosphere, such as humidity and turbulence, into the models.

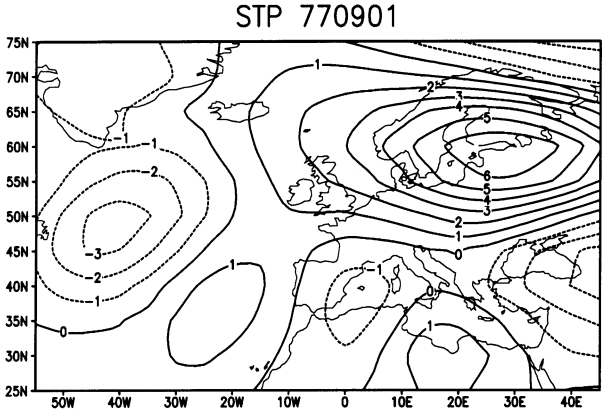
We do not close without a more optimistic conclusion that focuses on 2 things which are related. First, even if the intensification of amplitudes is small it has a greater influence, especially in the long run, for which even a small climate trend can have considerable con-

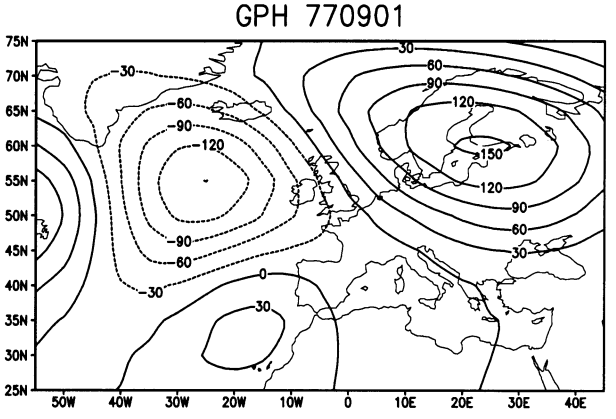

STP 770830
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E

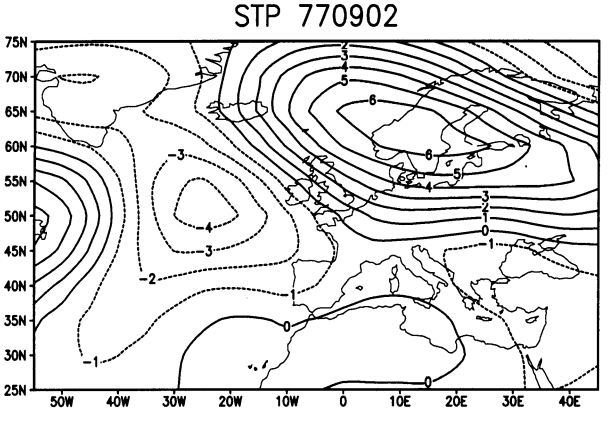

GPH 770830
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E


STP 770831
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E


GPH 770831
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E


STP 770901
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E


GPH 770901
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E


STP 770902
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E

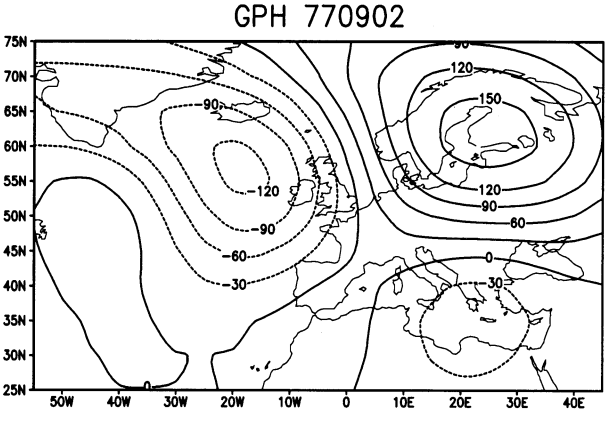

GPH 770902
75N
70N
65N
60N
55N
50N
45N
40N
35N
30N
25N
50W 40W 30W 20W 10W 0 10E 20E 30E 40E

Fig. 5. Illustration of a typical *d* day, 31 August 1977 (data shown as yymmdd). The sequence 30 August through 2 September depicts the development of a cold cyclone southwest of a strong high-pressure system over the Baltic, yielding warmer temperatures. On 31 August, the heating at Potsdam is maximum, with a STP (= T_{85}) anomaly of 5 to 6°C, and the cold front is at its closest. After this, both the front and the high-pressure cell retreat and the conditions in Potsdam return to normal

sequences. The second optimistic aspect applies to quantities whose appearance can be categorized by occurrence and amount, such as precipitation as in our example. We found that the nonlinear model mainly improves the occurrence prediction. Moreover, the above partitioning of the phase-space can be used as an additional forecasting and downscaling scheme; for spring and summer in particular, it should have a certain skill for predicting heavy rainfall events.

Since the performance of deterministic regression models is hopeless regarding the magnitude of extreme events, one has to rely on analogue or stochastic methods. As for the latter, the phenomenon of extreme events is best described by skewed distributions. In our opinion, the probit transformation represents the most general method of dealing with such distributions, and it usually needs no more than 2 or 3 parameters to describe the probit mathematically. However, the transformation depends very heavily on those parameters, and only small changes in them can have a huge impact, easily turning an event of 100 mm into one of 200 mm. Hence, for safe estimation one needs a considerable empirical basis. Unfortunately, under climate change, exactly those parameters are affected, so that a local assessment with respect to extreme events is a highly risky undertaking. This is the problem of all currently applied downscaling schemes, mentioned in the introduction, which utilize stochastic elements. Analogue schemes are not in danger of producing unrealistic events as they simply rearrange events that have happened already. There is, however, no guarantee that climatic change happens this way and does not bring about new and unforeseen scenarios. The best future downscaling approach might lie somewhere between the more 'conservative', analogue methods and the more 'progressive' schemes which employ stochastic elements.

Acknowledgements. A.W. was supported by DFG Grant EC 102/5-2 DFG. He thanks P. Ruján for discussions and a critical reading of the manuscript. In all calculations we extensively used the Data-Stream-Network Tool (www.neuro.uni-oldenburg.de/~weichert) implemented in Oldenburg, which has graphic-editing capabilities that allow for an easy analysis of data and control of the main parameters involved in numerical calculations.

LITERATURE CITED

Bardossy A, Plate E (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Wat Resour Res 28:1247–1259

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

Bürger G (1996) Expanded downscaling for generating local weather scenarios. Clim Res 7:111–128

Elsner JB, Tsonis AA (1992) Nonlinear predictions, chaos and noise. Bull Am Meteorol Soc 78:49–60

Farmer JD, Sidorowich JJ (1987) Predicting chaotic time series. Phys Rev Lett 59:845–849

Fröhlinghaus T, Weichert A, Ruján P (1994) Hierarchical neural networks for time-series analysis and control. Network 6:101–116

Gyalistras D, von Storch H, Fischlin A, Beniston M (1994) Linking GCM-simulated climatic changes to ecosystem models: case studies of statistical downscaling in the Alps. Clim Res 4:167–189

Hecht-Nielsen R (1990) Neurocomputing. Addison-Wesley, Boston

Hewitson BC, Crane RG (1992) Large-scale controls on local precipitation in tropical Mexico. Geophys Res Lett 19:18, 1835–8

Heyen H, Zorita E, von Storch H (1996) Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. Tellus 48A: 312–323

Hughes JP, Lettenmaier DP, Guttorp P (1993) A stochastic approach for assessing the effect of changes in regional circulation patterns on local precipitation. Water Resour Res 29:3303–3315

Mass CF, Edmon HJ, Friedman HJ, Cheney NR, Recker EE (1987) The use of compact discs for the storage of large meteorological and oceanographic data sets. Bull Am Meteorol Soc 68:1556–1558

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

Moody J, Darken CH (1989) Fast learning in networks of locally-tuned processing units. Neural Comput 1:281–294

Morrison DF (1967) Multivariate statistical methods. McGraw-Hill, New York

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

Watanabe S (1985) Pattern recognition: human and mechanical. John Wiley & Sons, New York

Werner PC, von Storch H (1993) Interannual variability of Central European mean temperature in January–February and its relation to large-scale circulation. Clim Res 3: 195–207

Wilson LL, Lettenmaier DP, Skyllingstad E (1992) A multiple stochastic daily precipitation model conditional on large-scale atmospheric circulation patterns. J Geophys Res 97: 2791–2809

Zorita E, Hughes JP, Lettemaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 8:1023–1042

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: August 12, 1997; Accepted: February 4, 1998
Proofs received from author(s): April 16, 1998