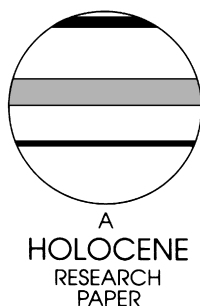# Artificial neural networks and dendroclimatic reconstructions: an example from the Front Range, Colorado, USA

Connie A. Woodhouse

(Institute of Arctic and Alpine Research, University of Colorado, CB450, Boulder, CO 80309, USA)

A
HOLOCENE
RESEARCH
PAPER

Abstract: The feasibility of reconstructing total spring precipitation for the South Platte River basin from tree-ring chronologies using artificial neural networks is explored. The use of artificial neural networks allows a comparison of reconstructions resulting from both linear and nonlinear models. Both types of models produced reconstructions that explained more than 40% of the variation in spring precipitation and were well verified with independent data. Although the nonlinear models produced higher R^2 values than did the linear model for the calibration period, they performed less well in the independent period. This result and other model evaluation statistics suggest that, in this study, the nonlinear models contain a greater degree of overfit than the linear model, and thus, do not offer a clear improvement over the linear model for the reconstruction of spring precipitation in this region. However, neural networks offer an alternative approach to linear regression techniques and may provide improved dendroclimatic reconstructions in other areas.

Key words: Dendrochronology, dendroclimatology, climatic reconstruction, precipitation, artificial neural networks, late Holocene, Colorado Front Range, USA.

Introduction

In reconstructing climate from tree-ring chronologies, traditional approaches for calibrating tree growth with climate have incorporated some type of linear regression technique (multiple regression, principal components regression, discriminant analysis, canonical regression) (e.g., Fritts, 1976; 1991, Briffa *et al.*, 1992; Stahle and Cleaveland, 1993; Cook *et al.*, 1994). The use of this family of statistical techniques assumes that linear relationships exist between climate and tree growth. Although such relationships may exist under some conditions, relationships may be more complex both in term of interactions between climate variables and between tree growth and climate.

Nonlinear relationships between tree growth and climate have long been acknowledged by dendrochronologists (Fritts; 1969, 1991; Graumlich and Brubaker, 1986). A limited number of studies have addressed this issue employing several approaches. These include the use of squares and cross products of predictor variables in linear regression equations to represent interactions between variables, and response surfaces to portray nonlinear or interactive relationships and from which predicted values have been directly interpolated (Graumlich and Brubaker, 1986; Fritts *et al.*, 1991; Graumlich, 1991; 1993). Several recent studies have

used artificial neural networks to model nonlinear relationships between tree growth and climate and to assess the effects of a doubling of CO_2 and pollution on tree growth in the Mediterranean and Alps regions (Guiot *et al.*, 1995; Keller *et al.*, 1997a; 1997b).

The goal of this study is to assess the suitability of nonlinear neural networks for reconstructing precipitation from tree-ring chronologies in the Colorado Front Range region by comparing results from both linear and nonlinear neural network models. There are several reason to focus on the Front Range. There are now over 50 tree-ring chronologies for this region. Dendrochronological studies have shown that tree growth in this region is sensitive to climate (Schulman, 1945; Kienast and Schweingruber, 1986; Kienast *et al.*, 1987; Woodhouse, 1993), and suggest that some of these tree-ring chronologies may be suitable for climate reconstructions (Brown and Shepperd, 1995). However, other studies have suggested that it may be difficult to reconstruct Front Range climate from tree-ring chronologies because of complex and unstable tree-growth/climate relationships for some species and locations (Graybill *et al.*, 1992; Villalba *et al.*, 1994).

The results of the studies cited above suggest that dendroclimatic reconstructions for the Colorado Front Range region are possible, but the task may be complicated by complex, changing and

nonlinear relationships between climate and tree growth. Consequently, dendroclimatic reconstruction efforts may benefit from an alternative statistical approach such as offered by artificial neural network analysis. Successful reconstructions of climate for this region would augment the widening coverage of dendroclimatic reconstructions across the USA. Furthermore, precipitation reconstructions for this area will provide vital information on long-term variability in precipitation and water supply. In this rapidly growing region, population increases and changing water needs are taxing the present water resources. Compounding these problems is the issue of climate uncertainty – due to both natural variability and anthropogenically induced global climate change. Reconstructions of precipitation can provide a longer temporal context than offered by the instrumental record with which to assess current and future climate variability.

Data

Regional precipitation

The focus of this study is the reconstruction of spring precipitation for the South Platte River basin. Spring precipitation was selected on the basis of preliminary correlation analyses which indicate that growth for most trees in this region, especially ponderosa pine, Douglas fir and limber pine, is related to spring moisture conditions. A regional variable was selected over a single station for reconstruction because a regional average reduces the noise and inhomogeneities inherent in a single station record and enhances the regional climate signal, to which trees are more likely to respond (Blasing *et al.*, 1981). A regional average for total spring precipitation (March, April, May) was derived from ten South Platte River basin stations (Figure 1, top). Records for nine of these stations have been screened for homogeneity and have had missing data estimated by personnel at the Colorado Climate Center. The one remaining station record was obtained from the US Historical Climate Network (USHCN; Karl *et al.*, 1990). The USHCN is a high-quality data set of monthly averaged temperature and total monthly precipitation records that have been screened for length of record, percent missing data, number of station moves, and other station changes that may affect the data homogeneity. The regional spring precipitation average spans the years 1913 to present (the 1913–1979 period was used in this study).

Tree-ring chronologies and processing

The tree-ring chronologies initially considered for this study are the 31 chronologies located in the Front Range region with the common time period 1700–1979. All chronologies are coniferous species and include bristlecone pine (*Pinus aristata*), Engelmann spruce (*Picea engelmannii*), lodgepole pine (*Pinus contorta*), Douglas fir (*Pseudotsuga menziesii*) and ponderosa pine (*Pinus ponderosa*) chronologies. The raw ring-width measurement series for these chronologies were obtained from the National Geophysical Data Center's International Tree-Ring Data Bank (Grissino-Mayer and Fritts, 1997) and from the unpublished collection of P.M. Brown. Measurement series were standardized using a very conservative detrending method designed to retain as much low frequency variation as possible; a 200-year cubic spline preserved most of the variation at wavelengths of 100 years or less. All the chronologies contain positive autocorrelation, as indicated by one or more significant autocorrelation coefficients at low lags. Because such autocorrelation is considered to result from biological rather than climatic factors (Fritts, 1976), all chronologies were filtered with a low-order autoregressive-moving-average (ARMA) model flexible enough to remove significant autocorrelation at low lags. In this study, an AR(2) model was used for all chronologies.
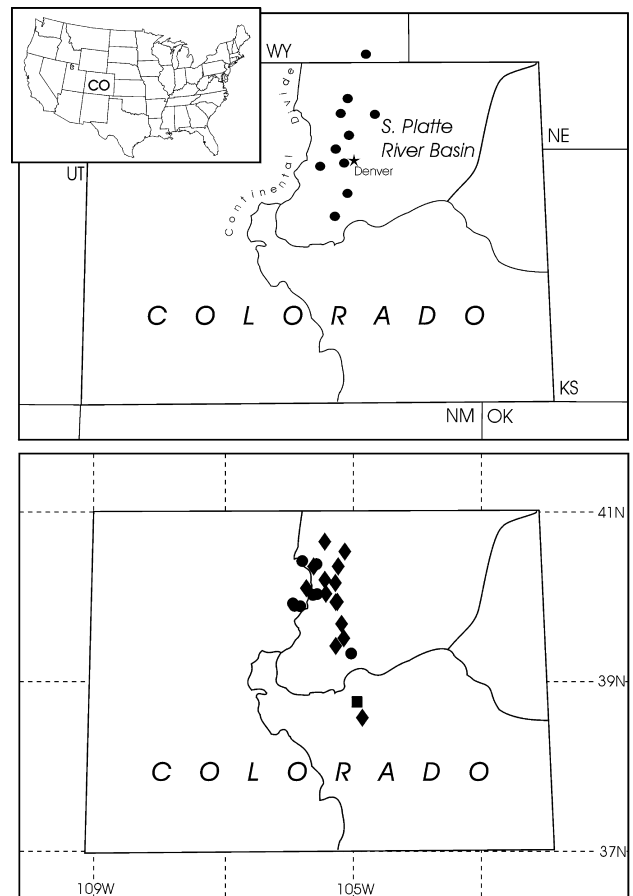
Figure 1 Top: Location of South Platte River basin and precipitation stations used for regional average (black dots). Bottom: Locations of tree-ring chronologies used in reconstructions – diamond = ponderosa pine; dark diamond = Douglas fir; square = limber pine; dot = Engelmann spruce chronologies.

The collection of 31 tree-ring chronologies was reduced to a set of six component scores through a rotated principal components analysis (PCA). In dendrochronology, PCA is commonly used to reduce the full set of original tree-ring chronologies to a more manageable set of transformed variables. These transformed variables, the set of principal component scores, are then used as predictors in climate reconstruction models. The PCA also acts to strengthen the common regional-scale climate response within a group of tree-ring chronologies by concentrating the common signal in the components with the largest eigenvalues. Here, a rotated PCA (Varimax) is used instead of an unrotated PCA as it provides intermediate diagnostic information about the groupings of the chronologies. In the rotated PCA, six components with eigenvalues greater than 1.0 were retained, with the 31 chronologies grouped primarily by species. Correlation analysis indicated that the tree-growth/climate responses for the limber/lodgepole pine component and the bristlecone pine component are unstable over time (i.e., the sign and strength of the correlation between climate and tree growth was not consistent over several time periods) and thus were deemed unsuitable for use in climate reconstruction models. Consequently, only the scores for remaining four components were used in the reconstruction. These four components represented an Engelmann spruce group (PC1), a southern ponderosa/limber pine group (PC2), a low elevation (1700–2000 m) ponderosa pine group (PC5), and a higher elevation (2400–2700 m) ponderosa pine group (PC6), which together totaled 23 chronologies (Figure 1, bottom). This set of PC scores was lagged forward one year (for example, tree growth in 1914 matched with climate in 1913) and the set of lagged scores was added to the original set of PCs to create a set of eight potential

predictor variables. The lagged PC scores were included to account for any lagged response to climate not adjusted for in the prewhitening process (i.e., the ARMA modelling).

Neural network overview

In general, a linear neural network is equivalent to a linear regression model and a nonlinear neural network is a flexible nonlinear regression model. Since dendrochronologists have long used linear regression techniques, the step from linear regression techniques to neural network techniques for dendroclimatic reconstructions is not a great jump. There are many good references on neural networks (e.g., Hewitson and Crane, 1994; Goodman, 1996; Sarle, cited 1998), so only a brief overview of one common type of neural network, the feed-forward backward-propagation network, is given here.

A schematic of a feed-forward backward-propagation network is shown in Figure 2. The main parts of this network include (1) the input variable layer, (2) a hidden layer comprised of a number of units or nodes, and (3) an output layer. There can be any number of hidden units, any number of hidden layers, and numerous output units. In addition, a bias term that acts as an intercept is commonly included. When the model is run, observations from each input variable are weighted and fed into each hidden unit, where they are summed (in this study, although they could be combined in another way) and passed through a function. The function can be a linear, nonlinear, binary or threshold function, and, if a nonlinear function is used, nonlinearity is introduced into the network at this point. The nonlinear function used in this study is the commonly used symmetric logistic function. In this example, the resulting values are then passed to the single output unit, weighted and summed as well. The summed value may be the final output, or it may be passed through a function once again (as shown here). This process is repeated for each set of observations, and a set of predicted values is produced. This set of values is compared to the observed values, and the mean squared error is determined and used to adjust the weights (the original weights are randomly assigned). Feed-forward refers to the direction of the data flow into the model, and backward-propagation refers to the computation of the error gradient in which the weight corrections between the output and hidden layer are calculated first, then those between the hidden layer and the input layer, and

the modified weights are adjusted at the same time for all the weights in the net. This error assessment/weight adjustment process (called training) is iterated until a stable minimum error is reached.

In a linear neural network, there are no hidden layers or units. The model is derived in the same way as the nonlinear model, with initial random weights which are then adjusted with each set of observations passed through the network. A linear neural network is numerically equivalent to a linear regression and results should be the same (Goodman, 1996). My experience has shown this to be true when there is a consistent relationship between the predictors and predictand. In this study, the linear network model is considered to be the equivalent of a linear regression model, as results were almost identical.

The problem of overfit

The search for the minimum error can lead to overfit, one of the greatest pitfalls in working with neural networks. Model overfit results from an overly complex model. The complexity can be due to the number and/or size of weights in the model. The number of weights determines the degrees of freedom lost, and the number of weights in a moderately complex model may fast approach the number of cases in a small data set. For example, the model in Figure 2 takes up 33 degrees of freedom. The size of the weights increases with increased number of training iterations. An overfitted model will be fit to the noise in the series, not to the signal, which will result in poor results when the model is run on independent test data. Fortunately, there are a number of techniques to help prevent model overfit (for more details, see Hewitson and Crane, 1994; Goodman, 1996; Sarle, cited 1998), and in this study the technique of early stopping was used to limit the number of training iterations, and thus help prevent overfitting. Using this technique, training is carried out on a subset of the total observation, while the rest of the values are held out and used to assess the residual error. When the residual error no longer decreases, the minimum error is noted and the training process is stopped. This model fitting and testing process is repeated with a different subset of observations a number of times (five splits in this study). The minimum errors from the five training runs are averaged and the full set of data is then trained until this target error is reached and no further, rather than until a minimum error is reached (thus the term 'early stopping').

In spite of techniques such as early stopping, models may still be overfitted to the calibration data. Methods exist to evaluate this overfit. In this study, bootstrapping (Efron and Tibshirani, 1993) was employed to assess the optimistically biased explained variance in the final models and to generate confidence intervals for predicted values. Fifty bootstrapped data sets (with the same number of cases as the original set) are drawn from the original data (with replacement); then the entire model fitting process is repeated for each of the 50 sets of bootstrapped data, each initiated with random weights. For each bootstrap data set and model, the difference between the explained variances obtained when the model was tested with its own bootstrapped data set and the explained variances obtained when the model was tested with the full data set is figured. This difference then is averaged over all bootstrapped data sets, and is the bias that is subtracted from the original model R^2 value. This bias-adjusted R^2 is more conservative than the adjusted R^2 obtained from a stepwise linear regression, which is based on the number of cases and predictor variables (an example from one model: linear neural network model bias adjusted $R^2 = 0.304$; linear regression adjusted $R^2 = 0.363$). Ultimately, the best way to evaluate a model's ability to generalize is to reserve a completely independent set of data for final model testing. Because it is so easy to overfit models using nonlinear neural networks, it is especially important to test these models with independent data.
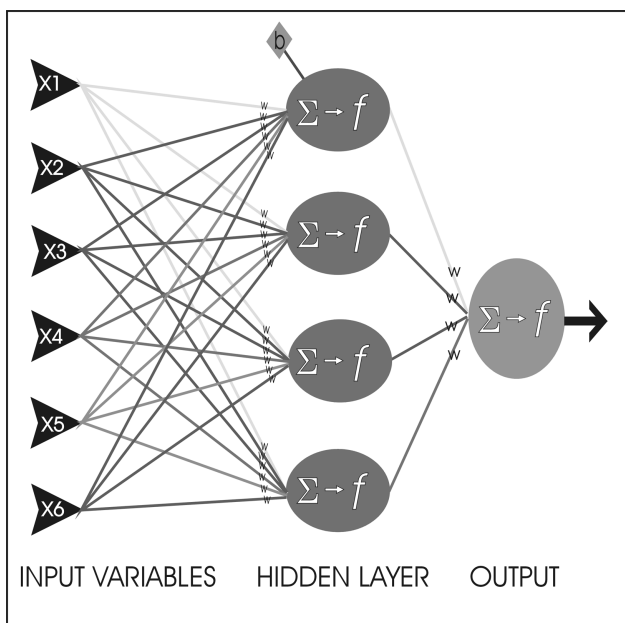

Figure 2 Example of a nonlinear neural network.

Software

The software program used is in this study is NevProp3 (Goodman, 1996). This program is a flexible, easy-to-use software program that includes a number of features that make it especially suitable for dendroclimatic analysis. Some of these features are listed below.

- The relevance (contribution to r^2) of each input variable is determined.
- Bootstrapping is an option, so that confidence intervals can be determined for predicted values and final variable weights as well as a bias-adjusted r^2, allowing an assessment of model fit and comparison of models.
- Several regularization methods are offered to prevent overfitting.
- The use of a independent test data set for model evaluation is easy to incorporate.
- Both linear and nonlinear models can be run, enabling a straightforward comparison between results from the two types of models.

Neural network modelling strategy and results

Strategy

The general strategy for this study was first to develop and select a small set of candidates for the 'best' neural network model, including both linear and nonlinear models. The eight predictor variables (the scores from PCs 1, 2, 5 and 6, plus these PCs lagged forward a year) were the initial input for both linear and nonlinear neural networks. After the models were derived using all potential predictor variables, the predictor variables were evaluated and the set was reduced on the basis of each predictor variable's contribution to the total explained variance. In the nonlinear models, hidden units (each containing a symmetric logistic function) were added incrementally in one, then two, layers, until a notable degree of overfit was indicated by comparing the percent of well-determined parameters, the R^2 values, and bias adjusted R^2 values. This process yielded several candidates for 'best' model which were selected with regard to the efficiency of the model (number of parameters and percent of parameters that were well defined, or effective degrees of freedom used by the model), the variance explained by the model, and the amount of bias in the explained variance (an indication of overfit). Then these candidate 'best' models were compared with each other on the basis of: (1) univariate statistics for observed and predicted precipitation values, (2) the amount of variance explained by the model, (3) a set of statistics used to evaluate the model error, and (4) model ability to effectively estimate values not contained in the original calibration data set.

The full precipitation record (1913–1979) was split into two sets of periods so that the models could be derived from one set of data (calibration period) and tested on an independent set of data (verification or test period). The two sets of periods used were 1933–1979 for the calibration period and 1913–1932 for the verification period, and 1913–1959 for the calibration period and 1960–1979 for the verification period. Results for both the calibration period and the independent verification period were best for the second set of periods, indicating a stronger tree growth/climate relationship during the period 1913–1959, but good model fit during the test period, 1960–1979. Only the results for this model period are reported here.

Results

Results for a progression of models are summarized in Table 1. The candidates selected for the best models were (1) the linear model with five predictor variables (5-0-1, Figure 3, top), (2) a nonlinear model with five predictor variables and one hidden layer with two units (5-2-1, Figure 3, centre), and (3) a nonlinear model with five predictor variables, one hidden layer with three units, and another hidden layer with one unit (5-3-1-1) (Figure 3, bottom). All three models have the same predictor variables (PCs 1, 2, 5, 6, and lagged PC 2), although models with other sets of variables were tested. The models were evaluated with regard to the number of parameters in the model (degrees of freedom used), the percent of well-defined parameters (percent of available degrees of freedom used), the explained variance, and bias-adjusted explained variance (Table 1). For example, although the percent of well-determined parameters in the 5-2-1 model is not outstanding (67%), the number of parameters is relatively low (15), and although the explained variance (0.579) is not a high as some other models', the adjusted explained variance (0.351) is higher than most and the difference between these two values is not great, suggesting a lesser degree of overfit. The 5-3-1-1 model was selected mostly on the basis of high explained variance (0.721) and adjusted explained variance (0.469), although the difference between these values is considerable. The model is well utilizing the parameters in the model (73% well defined), but the number of parameters (24) is fairly high as well and represents a greater number of degrees of freedom lost. In addition, this model has a the lowest average squared error of almost all models tested. The linear model (5-0-1) was selected because of the small number of parameters (6), the high percent of that were well defined (72%), and the smallest difference between the R^2 and the bias adjusted R^2 values. The two nonlinear models offer a greater degree of explained variance over the linear model, but the simpler linear model may have the least amount of overfit. Predicted values for the three models and their associated 95% confidence intervals are shown in Figure 4.

When the three models are compared with respect to how well the predicted values reflect the observed precipitation values, the nonlinear models, especially the 5-3-1-1 model, provide a better fit, overall, for the calibration period data (1913–1959) than the linear model (Figure 5). The nonlinear models do not perform as well for the test period (1960–1979). In particular, the 5-2-1 nonlinear model does a poor job duplicating the dry period of the 1960s, although the nonlinear 5-3-1-1 model does a somewhat better job with this period.

Statistical results also show that the nonlinear models, again especially the 5-3-1-1 model, generally perform better than linear model for the calibration period, but not quite as well for the verification period. Univariate statistics (mean and standard deviation) show that means are duplicated in the calibration period quite well by all models, but models do a poorer job duplicating the verification period mean and the standard deviations for both calibration and verification periods (Table 2). The nonlinear 5-3-1-1 model most closely matches calibration and verification period standard deviations, while the linear model most closely matches the verification period mean. The root mean square error (RMSE) is a measure of error or the difference between the observed values (O) and the values estimated by the models (P) and is defined by:

$$\mathrm{RMSE} = [\mathrm{N}^{-1} \sum_{I=1}^{N} (P_i - O_i)^2]^{0.5}$$

The RMSE indicates that the nonlinear models perform better than the linear model in the calibration period (smaller RMSE), but not as well as the linear model in the verification period (Table 2).

The reduction of error (RE) statistic, the coefficient of determination (R^2), and the index of agreement (d), are all measures of the predictive ability of the models. The RE indicates whether the reconstruction model does a better job of estimating precipitation

Table 1 Model fitting results

Model	No. of Parameters	% Well-defined parameters	r^2	Bias-adjusted R^2	Average squared error
5-0-1 (linear)	**6**	**72%**	**0.431**	**0.304**	**1.77**
5-1-1	8	72%	0.513	0.326	1.52
5-2-1	**15**	**67%**	**0.579**	**0.351**	**1.31**
5-3-1	22	76%	0.699	0.306	0.938
5-4-1	29	63%	0.669	0.349	1.031
5-5-1	36	62%	0.742	0.37	0.803
5-6-1	43	42%	0.616	0.214	1.2
5-7-1	50	36%	0.606	0.249	1.23
5-10-1	71	27%	0.609	0.238	1.22
5-1-1-1	10	65%	0.542	0.377	1.43
5-2-1-1	17	55%	0.548	0.292	1.41
5-3-1-1	**24**	**73%**	**0.721**	**0.469**	**0.87**
5-4-1-1	31	60%	0.694	0.409	0.954
5-5-1-1	38	39%	0.604	0.371	1.23
5-6-1-1	45	55%	0.743	0.445	0.8
5-2-2-1	21	53%	0.597	0.281	1.26
5-3-2-1	29	35%	0.569	0.259	1.34

values in the verification period than does the mean of the calibration period (Fritts, 1976; Cook and Kairiukstis, 1990), and values ranges from negative infinity to positive 1.0. Any positive value indicates that the reconstruction has some predictive skill.

$$RE = 1.0 - \sum_{i=1}^{N'} (O_i - P_i)^2 / \sum_{i=1}^{N'} O_i^2$$

The RE values for all three models are positive, but the RE for the linear model is higher than that for the nonlinear models. Both R^2 and adjusted R^2 show that the nonlinear models explains more of the variation in calibration period precipitation, but the verification period R^2 shows that less of the variance in the verification period is explained by the nonlinear models than by the linear model (Table 2).

A high correlation between two series indicates common variance, but says little about the magnitude of the covariations and differences in proportionality. The index of agreement, d, has been suggested by Willmott (1981; Willmott *et al.*, 1985) as a way to evaluate the degree to which model estimates are error-free. The equation for d is as follows:

$$d = 1 - \frac{\sum_{I=1}^{N} (P_i - O_i)^2}{\sum_{I=1}^{N} [|P_i - \bar{O}| + |O_i - \bar{O}|]^2}$$

The d statistic specifies the degree to which the observed deviations about the observed mean correspond, in magnitude and sign, to the predicted deviations about the observed mean, and is sensitive to both differences in observed and predicted means and changes in proportionality (Willmott, 1981). Values for d range from 0 to 1.0 with a value of 1.0 indicating a perfect association. Although all three d values may be considered good when assessed for the verification period, results indicate that the linear model slightly out-performs the nonlinear model 5-3-1-1 (Table 2).

A comparison between predicted values for the three models and a second set of independent data further tests the predictive skill of the models. The second set of independent data consists of a regional South Platte spring precipitation record derived from a subset (6 of 10) of the original precipitation stations (those that extended back to 1899). The period covered by these stations, 1899–1913, was not used for the formal model calibration and

verification process because it is not representative of the rest of the record (1913–1980). This time period was anomalously wet (Figure 5); in 10-year running means, all 10-year means for 1899–1913 are 16 cm or greater; for rest of period there are only five 10-year periods with values of 16 cm or greater. An initial effort to build a model that performed well on this subset period failed and these results indicate that trees in this region are not very sensitive to very wet conditions. Although none of the models does a very good job at estimating precipitation during the wettest part of this period, the nonlinear neural networks may perform slightly better (Figure 5). This slight edge is also suggested by the fact years with precipitation values ≥ 30% above average (15 years total for the period 1899–1978) tend to be more closely replicated by the nonlinear model estimates than linear model estimates (the nonlinear models had smaller residuals than the linear model in 13 of 15 (5-2-1) and 12 of 15 (5-3-1-1) wet years).

In all three reconstructions, it is likely that the dry years are much more faithfully represented than the wet years. This is commonly the case in dendroclimatic reconstructions that utilize arid site moisture-sensitive trees. In general, it has been noted that narrow rings provide more precise information on limiting conditions (drought, in this case) than wide rings which indicate favorable conditions, but are still limited by some factor perhaps other than precipitation (Fritts, 1976). The reconstructions of dry years are similar for all models in most cases, the exception being the great difference in predicted values during the major dry years in the 1960s, as mentioned previously. Here the extreme years in the beginning of this dry period are better duplicated by the linear model, but the nonlinear 5-3-1-1 model does as well or better on the second part of this dry spell.

To summarize, the nonlinear model 5-3-1-1 appears to provide a better fit to the calibration period data than the linear model or the nonlinear 5-2-1 model, but both nonlinear models generally perform better than the linear model in the calibration period. Both nonlinear models perform less well on the independent data than the linear model, although the nonlinear models (especially 5-3-1-1) may replicate wet years somewhat more accurately than the linear model. These results suggest that the more complex nonlinear models may contain a greater degree of overfit than the linear model, a suggestion supported by the greater difference between the calibration period R^2 and bias adjusted R^2 for the nonlinear models (Table 1), (even though this difference is greatest for the 5-3-1-1 model which actually performs slightly better on the independent data than the 5–2-1 model). Statistically, all
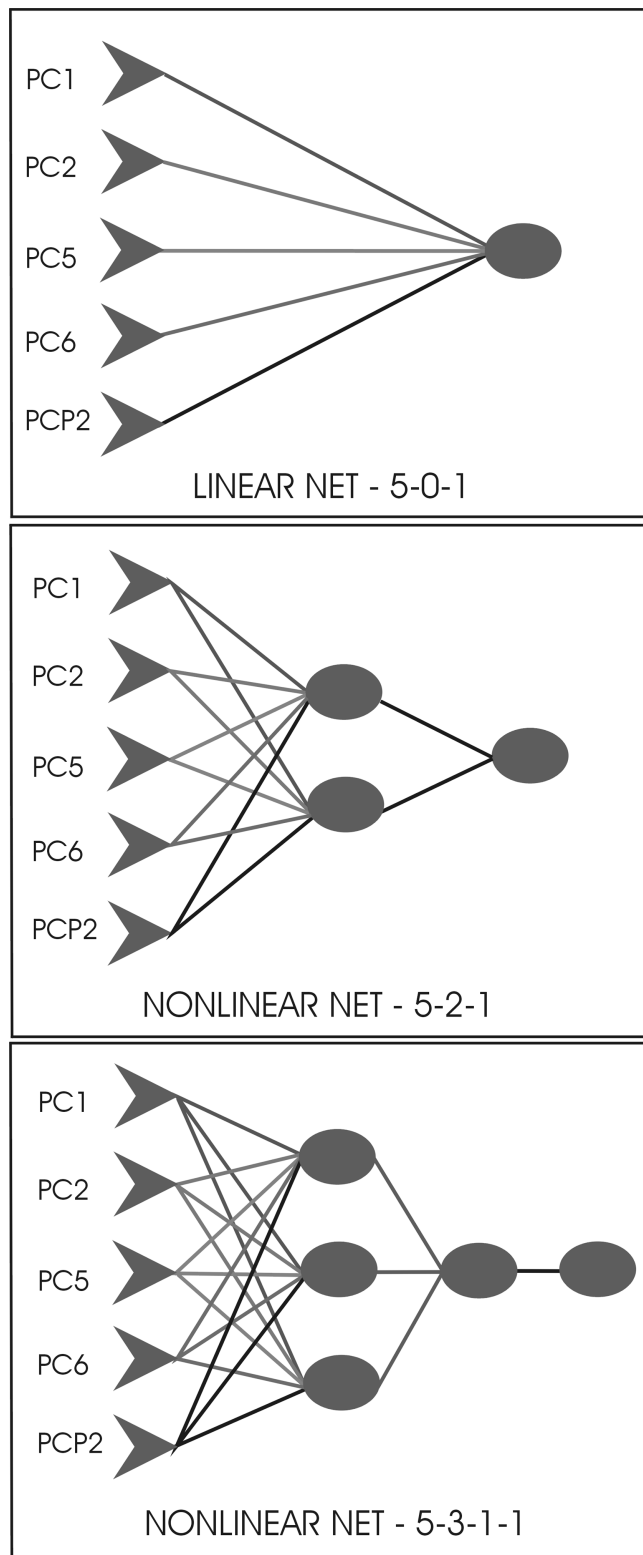
Figure 3 Top: Schematic of the linear neural network used in this study. Centre: Schematic of the 5-2-1 nonlinear neural network used in this study. Bottom: Schematic of the 5-3-1-1 nonlinear neural network used in this study.

three models provide acceptable reconstructions, but the linear model likely provides a more reliable reconstruction.

When these reconstructions are compared to other precipitation-related dendroclimatic reconstructions for the western USA and Great Plains, the amount of variance explained by the reconstructions, averaging about 50% for the calibration period (see Woodhouse and Overpeck, 1998: Table 3), is similar. In Cook *et al.*'s (1996; 1999) reconstructions of summer drought over the cotermi-
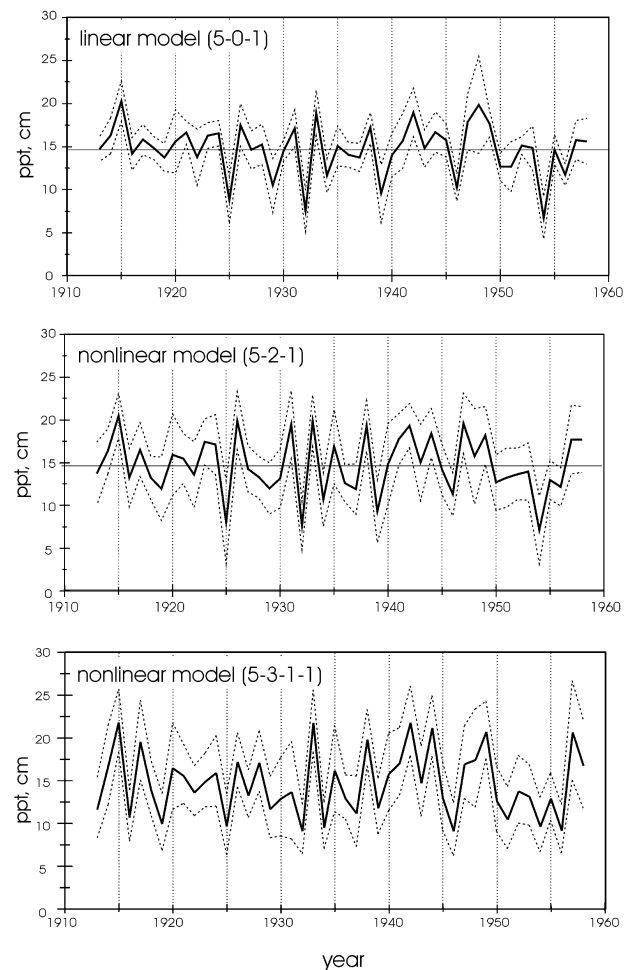

Figure 4 Predicted values for the linear and nonlinear models (solid lines) for calibration period 1913–1959, with 95% confidence intervals (dotted lines).

nous USA, the grid points that cover the South Platte River drainage display a calibration period R^2 of between 50% and 60%, but for the verification period this value drops to around 20%. Verification period explained variance for the neural network models is 44% and 41% for linear and nonlinear models, respectively, showing an improvement over Cook *et al.*'s results. Many of the same chronologies were used for these two sets of reconstructions and the improvement in this study is perhaps due to the exclusion of the early twentieth-century wet period in the model calibration and verification process, while Cook *et al.*'s verification period included this anomalously wet period.

Precipitation reconstructions

The reconstructions of South Platte spring precipitation derived from the three models for the period 1700–1979 are shown in Figure 6, along with smoothed series (5-weight binomial filter) to facilitate comparison. Although the three reconstructions are clearly similar ($r_{501/521} = 0.834$, $r_{501/5311} = 0.810$, $r_{521/5311} = 0.792$), there are some notable differences. One of the most obvious differences is the variability in the yearly values which seems greatest in the nonlinear 5-3-1-1 model. The standard deviation for this reconstruction is greater than for the other two (4.23 versus 3.17 for the linear model and 3.48 for the 5-2-1 model), and is closer to the observed precipitation standard deviations (Table 2). However, the range is less and the distribution of extreme values is also very even over time, to the degree that the distribution looks
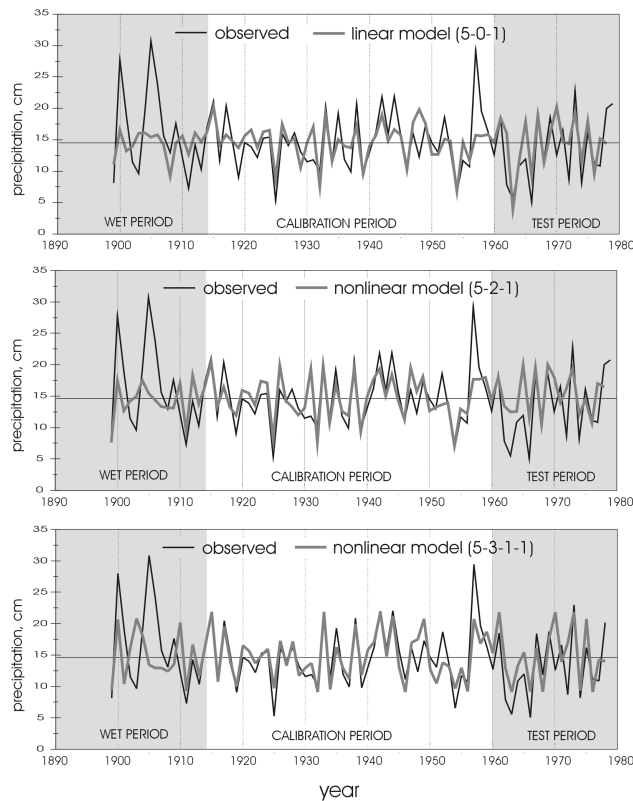
Figure 5 Top: Comparison of observed precipitation and reconstructed precipitation from the linear neural network model for the calibration period (1913–1959), test or verification period (1960–1979), and a second independent period, the wet period of 1899–1912. Centre: the same as above but for the nonlinear 5-2-1 neural network model. Bottom: the same as above but for the nonlinear 5-3-1-1 neural network model.

Table 2 Model comparison statistics for calibration period (1913–1959) (subscripted c) and verification or test period (1960–1979) (subscripted v). RMSE = root mean square error; RE = reduction of error statistic; R^2 = explained variance; d = index of agreement

Statistics	Observed values	Linear model 5-0-1	Nonlinear model 5-2-1	Nonlinear model 5-3-1-1
$mean_c$	14.73	14.73	14.84	14.76
$mean_v$	13.45	14.53	15.32	14.74
standard deviation$_c$	4.44	2.92	3.36	3.77
standard deviation$_v$	5.09	4.08	3.17	4.92
$RMSE_c$		3.35	2.89	2.37
$RMSE_v$		4.06	4.44	4.34
RE		0.393	0.268	0.306
R^2_c		0.431	0.579	0.721
bias adjusted R^2_c		0.304	0.351	0.469
R^2_v		0.441	0.411	0.407
d_v		0.787	0.704	0.778

artificial, perhaps an artifact of the model. This is less noticeable in the smoothed series. Other differences are apparent during particular periods of time, but it is difficult to judge which reconstruction is more accurate as there are no other record of this precision with which to closely assess these reconstructions. However, the reconstructions can be qualitatively assessed in terms of extreme events recorded in other dendroclimatic reconstructions and other types of proxy data for this general region.

All three precipitation reconstructions capture periods of his-
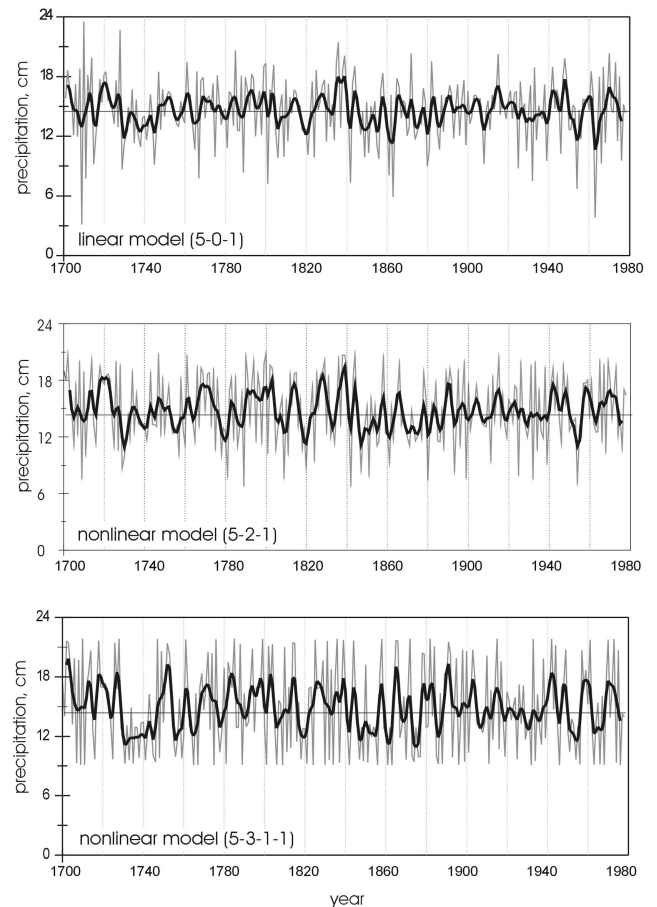
Figure 6 Precipitation reconstructions for 1700–1979, from the linear model (top), the 5-2-1 nonlinear model (centre), and the 5-3-1-1 nonlinear model (bottom), with smoothed series (5-weight binomial filter), in heavy lines.

torically documented drought as well as several droughts found in other tree-ring reconstructions for the western and central USA (Figure 7). Nineteenth-century newspapers, travel accounts and historical fort climate data indicate the occurrence of drought in the early 1860s and around 1820 across the Great Plains (Ludlum, 1971; Bark, 1978; Lawson and Stockton, 1981; Mock, 1991; Muhs and Holliday, 1995). Additionally, tree-ring reconstructions of drought in the western USA show a drought in the late 1840s (Fritts, 1965; Stockton and Meko, 1975). These three periods of drought are reflected in the precipitation reconstructions, and, although the three models portray each somewhat differently, these are clearly major droughts for the nineteenth century. In the eighteenth century, tree-ring reconstructions record several more severe droughts across the western USA and parts of the Great Plains. A severe drought in the late 1770s is recorded in reconstructions of Texas drought (Stahle and Cleaveland, 1988), as well as in regional drought reconstructions for the western and southwestern USA (Fritts, 1965; Stockton and Jacoby, 1976; Meko *et al.*, 1995). Another period of drought centred around 1740 is found in a reconstruction of Iowa drought (Cleaveland and Duvick, 1992), and is also extensively reflected in a network of drought reconstructions for the central and western USA (Stockton and Meko, 1975). These eighteenth-century droughts are evident in all of the South Platte regional precipitation reconstructions.

The major twentieth-century droughts are also well reflected in the South Platte spring precipitation reconstructions, especially the 1950s drought, which was more severe, but shorter in duration than the 1930s drought in this area. An inspection of the full reconstruction period helps place these major twentieth-century
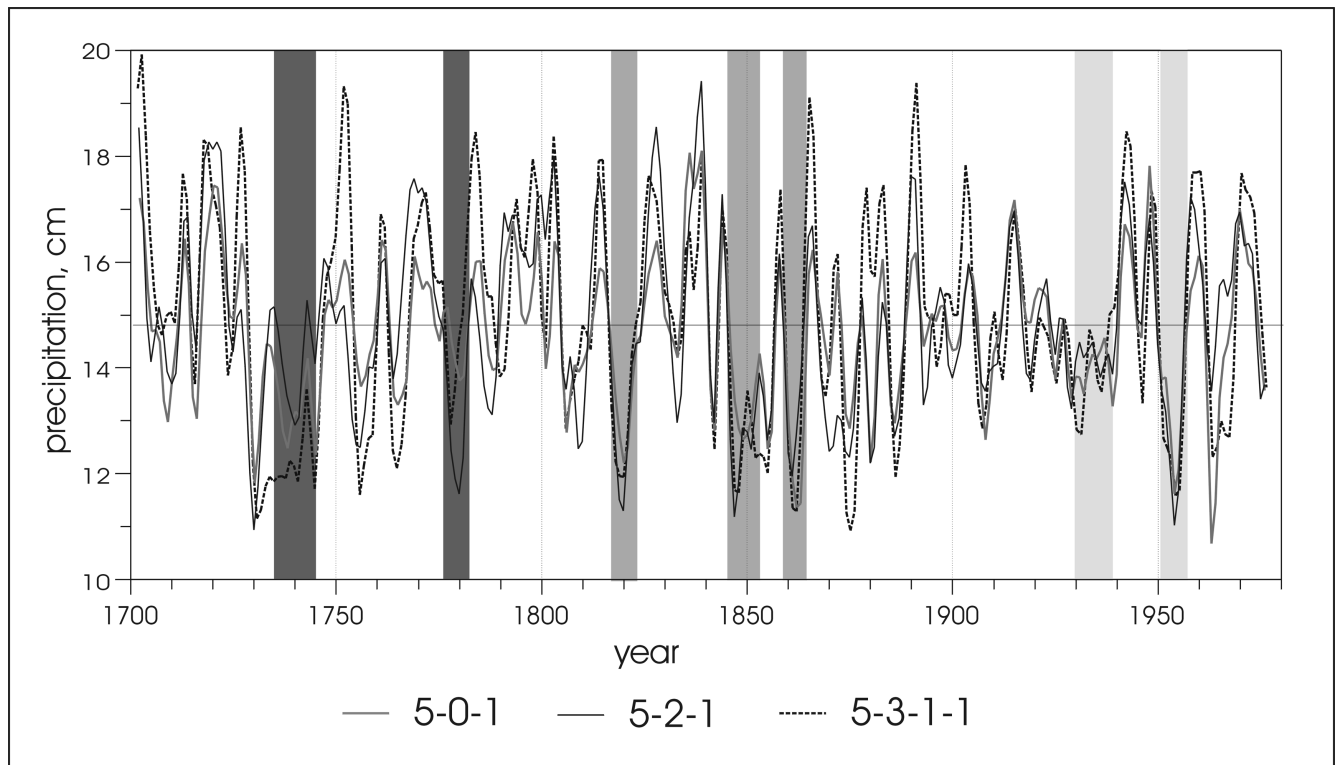
Figure 7 Smoothed reconstructions for all three models (linear 5-0-1 model in grey line; nonlinear 5-2-1 model in black line; nonlinear 5-3-1-1 model in dotted line), along with droughts documented by instrumental data (twentieth century), historical accounts and records, and other tree-ring reconstructions (eighteenth and nineteenth centuries) shown by vertical bars.

droughts in perspective (Figure 7). These reconstructions suggest that the Front Range was affected by severe and widespread droughts of the magnitude at least equal to those of the twentieth century several times a century over the past 300 years, confirming what other proxy data have also suggested for other parts of the central and western USA (Woodhouse and Overpeck, 1998). The reconstructions from the nonlinear 5-3-1-1 model shows the 1740s drought to be particularly severe, but results of this comparative study suggest the estimates from this model may not be the most reliable.

Conclusions

Does the use of nonlinear neural networks offer improved reconstructions of spring precipitation for the South Platte River basin? There is reason to believe that nonlinear models should better duplicate the nonlinear response of tree growth to precipitation (i.e., low growth in dry years, but not necessarily high growth in wet years) and thus provide a better reconstruction model. Although it is possible to produce nonlinear models that more closely duplicate calibration period values than do linear models, this improved fit appears to be at the cost of incorporating more noise into the model, with the consequence that these models do not perform as well on the independent data (although in this study, the nonlinear models do appear to offer some improvement in the reconstructions of extremely wet years). All three models generated acceptable reconstructions that are at least comparable in quality to other reconstructions for the central USA, but the linear model may produce a more reliable reconstruction. The apparently restricted range of reconstructed values in the nonlinear 5-3-1-1 model is of particular concern.

In assessing the results of this research, it is not possible to say that nonlinear neural networks offer an unequivocal improvement over linear reconstruction techniques (neural network or regression) in this particular study; however, they do present an

alternative statistical technique to linear regression techniques. Further work using this technique in other regions and with other dendrochronological data is needed to fully evaluate the usefulness of neural networks for dendroclimatic reconstructions. In some cases, they may well provide improved dendroclimatic reconstructions, especially when using tree-ring chronologies with a higher signal-to-noise ratio than contained in Front Range chronologies.

The use of nonlinear neural networks is rapidly gaining acceptance as a new method for investigating relationships in multivariate data. As software becomes more available and easily used, utilization of this technique will increase. The leap from linear regression techniques to nonlinear neural networks is not great, but should be made with caution as these models can easily become more complex than the data warrant. Model overfit, with the consequence of modelling the noise in the data instead of the signal is a problem that may come with an incomplete understanding of the limitations and pitfalls of the technique as well as a lack of familiarity with the data being investigated. Neural network models appear to be more sensitive to noise in the data, demanding a better understanding of data set characteristics than regression-based techniques, and careful validation of the model with completely independent data.

Acknowledgements

References

Bark, L.D. 1978: History of American drought. In Rosenberg, N.J., editor, *North American droughts*, Boulder, Colorado: Westview Press, 9–23.

Blasing, T.J., Duvick, D.N. and **West, D.C.** 1981: Dendroclimatic calibration and verification using regionally averaged and single station precipitation data. *Tree-Ring Bulletin* 41, 37–43.

Briffa, K.R., Jones, P.D. and **Schweingruber, F.H.** 1992: Tree-ring density reconstructions of summer temperature patterns across western North America since 1600. *Journal of Climate* 5, 735–54.

Brown, P.M. and **Shepperd, W.D.** 1995: Engelmann spruce tree-ring chronologies from Fraser Experimental Forest, Colorado: potential for a long-term temperature reconstruction in the central Rocky Mountains. In Tinus, R.W., editor, Interior West Global Change Workshop, 25–27 April, Fort Collins, CO, *USDA Forest Service General Technical Report RM-GTR-262,* 23–26.

Cleaveland, M.K. and **Duvick, D.N.** 1992: Iowa climate reconstructed from tree rings, 1640–1982. *Water Resources Research* 28, 2607–15.

Cook, E.R. and **Kairiukstis, L.A.** 1990: *Methods of dendrochronology: applications in the environmental sciences.* Dordrecht: Kluwer Academic Press, 394 pp.

Cook, E.R., Briffa, K.E. and **Jones, P.D.** 1994: Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *International Journal of Climatology* 14, 379–402.

Cook, E.R., Meko, D.M., Stahle, D.W. and **Cleaveland, M.K.** 1996: Tree-ring reconstructions of past drought across the coterminous United States: tests of a regression method and calibration/verification results. In Dean, J.S., Meko, D.M. and Swetnam, T.W., editors, *Tree rings, environment, and humanity,* Tucson, Arizona: Radiocarbon, 155–69.

—— 1999: Drought reconstructions for the continental United States. *Journal of Climate* 12, 1145–62.

Efron, T. and **Tibshirani, R.** 1993: *An introduction to the bootstrap.* London: Chapman and Hall.

Fritts, H.C. 1965: Tree-ring evidences for climatic changes in western North America. *Monthly Weather Review* 93, 421–43.

—— 1969: Bristlecone pine in the White Mountains of California; growth and ring-width characteristics. *Papers of the Laboratory of Tree-Ring Research* 4, Tucson: University of Arizona Press, 44 pp.

—— 1976: *Tree rings and climate.* London: Academic Press.

—— 1991: *Reconstructing large-scale climatic patterns from tree-ring data.* Tucson: University of Arizona Press, 567 pp.

Fritts, H.C., Vaganov, E.A., Sviderskaya, I.V. and **Shashkin, A.V.** 1991: Climatic variation and tree-ring structure in conifers: empirical and mechanistic models of tree-ring width, number of cells, cell size, cell-wall thickness and wood density. *Climate Research* 1, 97–116.

Goodman, P.H. 1996: NevProp software, version 3, Users' Manual. Reno, Nevada: University of Nevada (ftp://ftp.scs.unr.edu/pub/cbmr/nevpropdir).

Graumlich, L.G. 1991: Subalpine tree growth, climate, and increasing CO^2: an assessment of recent growth trends. *Ecology* 72, 1–11.

—— 1993: A 1000-year record of temperature and precipitation in the Sierra Nevada. *Quaternary Research* 39, 249–55.

Graumlich, L.G. and **Brubaker, L.B.** 1986: Reconstruction of annual temperature (1590–1979) for Longmire, Washington, derived from tree rings. *Quaternary Research* 25, 223–34.

Graybill, D.A., Peterson, D.L. and **Arbaugh, M.J.** 1992: Coniferous forests of the Colorado Front Range: the response of western forests to air pollution. In Olson, R.K., Binkley, D., Bohm, M., and Arbaugh, M., editors, *Ecological Studies* 97. New York: Springer-Verlag, 366–401.

Grissino-Mayer, H.D. and **Fritts, H.C.** 1997: The International Tree-Ring Data Bank: an enhanced global database serving the global scientific community. *The Holocene* 7, 235–38.

Guiot, J., Keller, T. and **Tessier, L.** 1995: Relational database in dendroclimatology and new non-linear methods to analyze the tree response to climate and pollution. In Ohta, S., Fujii, T., Okada, N., Hughes, M.K. and Eckstein, D., editors, Tree rings: from the past to the future. Proceedings of the International Workshop on Asian and Pacific Dendrochronology. *Forestry and Forest Products Research Institute Scientific Meeting Report* 1, 17–23.

Hewitson, B.C. and **Crane, R.G.** 1994: *Neural nets: applications in geography.* Dordrecht: Kluwer Academic Press, 194 pp.

Karl, T.R., Williams, C.N., Jr., Quinlan, F.T. and **Boden, T.A.** 1990*: United States historical climatology network (HCN) serial temperature and precipitation data.* Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Keller, T., Guiot, J. and **Tessier, L.** 1997a: The artificial neural networks: a new advance in response function calculation. In Urbinati, C. and Carrer, M., editors, *Dendrocronologia: una scienza per l'ambiente tra passato e presente.* Atti del XXXIV Corso di Cultura in Ecologia, San Vito di Cadore, Italy, 1–5 September 1997. Dipartimento Territorio e Sistemi Agroforestali, Universita degli Studi di Padova, 43–53.

—— 1997b: Climatic effect of atmospheric CO_2 doubling on radial tree growth in southeastern France. *Journal of Biogeography* 24, 857–64.

Kienast, F. and **Schweingruber, F.H.** 1986: Dendro-ecological studies in the Front Range, Colorado, USA. *Arctic and Alpine Research* 18, 277–88.

Kienast, F., Schweingruber, F.H., Bräker, O.U. and **Schär, E.** 1987: Tree-ring studies on conifers along ecological gradients and the potential of single-year analysis. *Canadian Journal of Forest Research* 17, 683–96.

Lawson, M.P. and **Stockton, C.W.** 1981: Desert myth and climatic reality. *Annals of the Association of American Geographers* 71, 527–35.

Ludlum, D.M. 1971: *Weather record book.* Princeton, New Jersey: Weatherwise, Inc., 98 pp.

Meko, D., Stockton, C.W. and **Boggess, W.R.** 1995: The tree-ring record of severe sustained drought. *Water Resources Bulletin* 31, 789–801.

Mock, C.J. 1991: Drought and precipitation fluctuations in the Great Plains during the late nineteenth century. *Great Plains Research* 1, 26–56.

Muhs, D.R. and **Holliday, V.T.** 1995: Evidence of active dune sand on the Great Plains in the 19th century from accounts of early explorers. *Quaternary Research* 43, 198–208.

Sarle, W.S., editor, cited 1998: Neural Network FAQ, posted to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html.

Schulman, E. 1945: Tree-rings and runoff in the South Platte River basin. *Tree-Ring Bulletin* 11, 18–24.

Stahle, D.W. and **Cleaveland, M.K.** 1988: Texas drought history reconstructed and analyzed from 1698 to 1980. *Journal of Climate* 1, 59–74.

—— 1993: Southern Oscillation extremes reconstructed from tree rings of the Sierra Madre Occidental and southern Great Plains. *Journal of Climate* 6, 129–410.

Stockton, C.W. and **Jacoby, G.C.** 1976: Long-term surface water supply and streamflow levels in the upper Colorado River basin. *Lake Powell Research Project Bulletin* No. 18, Los Angeles, California: Institute of Geophysics and Planetary Physics, 67 pp.

Stockton, C.W. and **Meko, D.M.** 1975: A long-term history of drought occurrence in western United States as inferred from tree rings. *Weatherwise* Dec., 244–49.

Villalba, R., Veblen, T.T. and **Ogden, J.** 1994: Climatic influences on the growth of subalpine trees in the Colorado Front Range. *Ecology* 75, 1450–62.

Willmott, C.J. 1981: On the validation of models. *Physical Geography* 2, 184–94.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J. and **Rowe, C.M.** 1985: Statistics for the evaluation and comparison of models. *Journal of Geophysical Research* 90 (C5), 8995–9005.

Woodhouse, C.A. 1993: Tree-growth response to ENSO events in the central Colorado Front Range. *Physical Geography* 14, 417–35.

Woodhouse, C.A. and **Overpeck, J.T.** 1998: Two thousand years of drought variability in the central United States. *Bulletin of the American Meteorological Society* 79, 2393–414.