

**PCMDI**

*Program for Climate Model Diagnosis and Intercomparison*

**PCMDI Report No. 29**

**NONLINEAR PRINCIPAL COMPONENT  
ANALYSIS OF CLIMATE DATA**

by

**Sailes K. Sengupta and James S. Boyle**

**Program for Climate Model Diagnosis and Intercomparison  
Lawrence Livermore National Laboratory, Livermore, CA, USA**

**December 1995**

**PROGRAM FOR CLIMATE MODEL DIAGNOSIS AND INTERCOMPARISON  
UNIVERSITY OF CALIFORNIA, LAWRENCE LIVERMORE NATIONAL LABORATORY  
LIVERMORE, CA 94550**

Nonlinear Principal Component Analysis of Climate Data

by

Sailes K. Sengupta and James S. Boyle

Program for Climate Model Diagnosis and Intercomparison  
Lawrence Livermore National Laboratory  
Livermore, CA USA

December 1995 - (revised February 1996)

## **DISCLAIMER**

**This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. Views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.**

**This report has been reproduced  
directly from the best available copy.**

**Available to DOE and DOE contractors from the  
Office of Scientific and Technical Information  
P.O. Box 62, Oak Ridge, TN 37831  
Prices available from (615) 576-8401, FTS 626-8401**

**Available to the public from the  
National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Rd.,  
Springfield, VA 22161**

## **Abstract**

In traditional principal component analysis (PCA) a few significant linear combinations of the original variables are extracted to arrive at a parsimonious description of a complex data set obtained from climate observations, analysis or from GCM outputs. These are uncorrelated variables which are used in practice to understand the principal modes of variation in the climatological process under study. If we drop the requirement of linearity and uncorrelatedness, a greater data reduction is possible allowing us to deal with fewer modes of variation. These nonlinear functions can in fact be obtained by using a series of auto-associative feed-forward neural networks in which the residuals from the previous network are fed as the contents of the input output pair for the next. It can be shown that in special cases such networks provide ordinary principal components.

We have explored this methodology to gain a better understanding of the precipitation data over the US observed over land and bordering oceans for the 1979 to 1988 decade. A careful comparison with the linear counterpart has been made. The improvement in the data reduction is noticeable but not overwhelming. Certain details in the modes of variation are more pronounced in the nonlinear representation. The leading nonlinear mode captures the seasonal cycles more clearly than the leading linear mode. In the latter, the seasonal cycle is shared by subsequent modes of the PCA. The principal linear and nonlinear modes of the observational data has been intercompared with the corresponding modes of the data obtained from a GCM simulation.

We conclude by observing that nonlinear principal component analysis(NLPCA) based on auto-associative neural networks is potentially a more effective data reduction tool than conventional PCA. Also the principal modes of variation of the precipitation data of the continental US are better differentiated by a NLPCA than by ordinary PCA. It should be tried as an alternative method especially when linear PCA fails to show meaningful patterns in climatological data analysis.

## **1. Introduction**

A Principal Component Analysis (PCA) is concerned with the understanding of the covariance structure through a few 'significant' linear combinations of the original variables, the motivations being twofold. Firstly, it enables us to compress data when significant correlation is present among the variables. Secondly, it allows us to attempt understanding the physical meaning of the data. While in the engineering sciences the primary motivation is the first one, in the geophysical sciences it is particularly important to discern any physical information content that may be present in the extracted principal components (Preisendorfer, 1988). In the work reported here we focus on the meaningful summarization of the large amounts of data output from GCM( General Circulation Model) simulations of the atmosphere and the concomitant observational data.

Given a space-time data set  $\{X(s,t), s=1,2,\dots, m; t=1,2,\dots,n\}$  with  $m$  spatial locations (typically these are grid locations in GCMS) and  $n$  time instants of observations, the spatio-temporal behavior of the physical field  $X$  can be decomposed into an uncorrelated set of  $m$  time series or  $n$  spatial fields depending on the mode of the PCA. If the requirement of uncorrelatedness is dropped and nonlinear functions of the original variables are allowed there is a significant gain in terms of data

compression. Furthermore, the lack of uncorrelatedness per se may not necessarily preclude the understanding of the physical meanings of the fields or time series derived using a nonlinear mode of analysis. In fact, by using an auto-associative neural network with nonlinear processing elements, a task equivalent to the extraction of principal components can be carried out. Such a task has been termed 'nonlinear principal component analysis' (Kramer, 1991) with the corresponding components termed as nonlinear principal components (NLPC).

This paper deals specifically with two geophysical data sets. These are the precipitation values over the North American continent during the decade 1979-1988 obtained from observations and simulated by an atmospheric GCM. In choosing this data set, we anticipated that an NLPC analysis applied to this data set may provide additional insight into the physical process which proved to be true only marginally. The neural network techniques used in this paper are described in the next section. It also contains a short account of how the extraction of ordinary (linear) principal components by multi-layer perceptrons (Bourlard and Kamp, 1987) is done by an auto-associative feed-forward neural network. The final section presents the data description, some results, and conclusions.

## 2. Artificial neural network

Principal component analysis (PCA) is a technique essential in data compression, feature extraction, compact coding and increasing computational efficiency. In the context of data analysis in climatology, the constraints and interdependency of spatio-temporal data can be identified and redundancy eliminated by the use of PCA. For example, the use of PCA is commonplace in climatological literature (Preisendorfer, 1988) in the efficient summarization of massive amounts of data. Briefly speaking, if  $\underline{x}$  is a centered  $n$ -dimensional vector, PCA extracts  $p$  ( $p \leq n$ ) linear combinations appearing as elements in the product  $W \underline{x}$  of the components of  $\underline{x}$  where  $W$  is a ( $p \times n$ ) matrix of weights subject to the constraints that (a) the variance of each linear combination appearing as elements of  $W \underline{x}$  is maximized and (b) the extracted linear combinations are mutually orthogonal. In practice, the eigensystem for the sample covariance matrix is solved with the resulting  $p$  dominant eigenvectors representing the principal vectors. Several workers in the neural network community (Bourlard and Kamp, 1987; Baldi and Hornik, 1989; Sirat, 1991; Oja, 1992) have related multilayer perceptron learning by back propagation algorithm with principal components extraction in classical statistics. An account of principal components by auto-associative linear networks as proposed by Baldi and Hornik (1989) is presented in an appendix. In their formulation, the network consists of a single hidden layer with linear processing elements (executing the identity function) with as many nodes as the desired number of PCs. The training input-output pairs consist of the same target output as input (hence the name auto-associative). It turns out that the activities at the hidden layer nodes are in some sense equivalent to the  $p$  ordinary principal components obtainable by a singular value decomposition of the covariance. Thus the hidden layer provides a compact representation of the data. Because of this property it has been termed the 'representation layer' in neural network literature. Becker (1991) provides a comprehensive survey relating PCA models to unsupervised learning neural networks. In suggesting a transition from linear to 'nonlinear' principal components (NLPC), Demers and Cottrell (1992) argue that PCA finds an optimal *linear subspace* on which one projects the data with minimum loss of information (in the sense of maximizing the 'explained' variance in the data). However, if the data lie on a *nonlinear submanifold* of the feature space, then the number of

dominant PCs will overestimate the dimensionality. The covariance matrix of sampled points on a *nonplanar* 3-d curve has a full rank and a PCA will generate three distinct eigenvectors. The same for a *nonlinear planar curve* will generate two and for a straight line only one. Yet *all three are intrinsically one-dimensional*.

In order to capture this intrinsic dimension, representations using neural networks have been proposed by several authors (Kramer, 1991; Oja, 1991; Usui, Nakauchi and Nakano, 1991 and Demers and Cottrell, 1992). The addition of hidden layers between the inputs and the representation layer as well as between the representation and the output layer provide a network which is capable of learning nonlinear representation. In the process, one achieves what may be termed a nonlinear analogue of PCA. In the following we present the sketch of such a network, Fig. 2. The network consists of five layers which are fully interconnected. In addition to the input and output layers which are identical since the network is made to be auto-associative, we have a central representation layer (where the principal manifolds or NLPCs are generated as the activities) and two identical layers placed on the two sides of the representation layer. These last two layers as described above, are called the encoding and decoding layer respectively. They are essential in this architecture since the mere addition of a nonlinear (sigmoid) activation function in the representation layer without these layers appears, at most, to be capable of producing a monotone function of the ordinary principal components. We have used the 'back propagation of error' as the training algorithm (Rumelhart and McClelland, 1986).

### Network architecture

Although there are no specific guidelines for the choice of the number of nodes in the encoding and the decoding layer, Kramer (1991) provides bounds based on the principle that the number of weights in the network should be a fraction of the number of constraints imposed by the data set. A few simplifying assumptions then lead to the constraints

$$M_1 + M_2 \ll n, M_1 > f \text{ and } M_2 > f \quad (1)$$

where  $M_1$ ,  $M_2$  are respectively the sizes of the encoding and the decoding layers,  $f$  is the number of nodes in the representation layer and  $n$  is the size of the training set.

In determining the size  $f$  of the representation layer, Kramer (1991) introduces a sequential determination of the NLPCs one at a time similar in spirit to its linear counterpart, namely the PCA extraction algorithm. Applied in this context, it amounts to using in a *recursive manner* the same network in fig. 2 except with a single node in the representation layer. In addition, each recursion feeds the residual matrix obtained from the previous stage as the elements in the training I/O pair. The residual matrix is simply the error matrix obtained by subtracting the output of the trained network from the input. The procedure stops when either a desired number of NLPCs have been extracted or a desired level of accuracy has been attained in the residual matrix. More simply, instead of a sequential procedure, we may simply decide to use a fixed number  $p$  say, of nodes in the representation layer and extract the  $p$  NLPCs. This will however preclude any ranking of these NLPCs.

## **Extraction of nonlinear principal modes and components**

The activities ( output) at the nodes of the representation layer are the NLPCS. This is the analog of the linear case where the actual PCs are generated in the hidden layer of the neural network (see appendix). The principal vectors (modes) no longer have a direct analog in the nonlinear case. We can, however, take a weighted average of the data fields, the weights being the elements of the time series representing a specific nonlinear principal component defined previously. These will be referred to as the nonlinear principal modes (NLPM).

### **3. Data**

The basic data used in this study are monthly averaged precipitation values over the United States for the decade 1979 to 1988 from two sources, observations and the output from a GCM simulation. The precipitation observations were gridded to a 4 degree by 5 degree latitude, longitude lattice. The observations are from surface stations over land (Schemrn et al. 1992) and satellite MSU estimates (Spencer, 1993) over the oceans. The bulk of the analysis grid used here is over the United States where the observational network provides reliable precipitation fields. The data are monthly averages for the 120 months from January 1979 to December 1988.

The model generated data was produced by the AMIP simulation of the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM). The version used here is described by Hack et al. (1993) and is referred to as the CCM2. The model has a spectral formulation and the simulation was run at a resolution of T42. The precipitation monthly means from the model were interpolated to the same 4 x 5 degree grid as the observations.

The sets for input into the PC and neural network were computed by subtracting the 120 month mean from each gridpoint to form deviations. In these data the seasonal cycle is retained. It is of interest to ascertain the relationship of the PCs and NLPCs to the strong seasonal variations. The neural net architecture used was that of Fig. 2 and 3 with 10 nodes in the encoding (MI) and decoding(M2) layers and I node in the representation layer(f). Since there are 120 points in the training set the inequalities in (1) are satisfied.

The PC analysis used the standard routine PRINC from IMSL (1994) to compute the principal components from a covariance matrix computed from the 120 time samples at 95 spatial gridpoints. Since there is no exact correspondence between the PCs and NLPCs beyond the first of each, comparison of the subsequent components is not obvious. The results will focus on the first components of each method since these will be shown to represent the seasonal cycle and allow a fairly direct and unambiguous interpretation.

### **4.0 Results**

Figure 4 shows the leading nonlinear and linear modes for the observations of precipitation over the US. The spatial distribution allows for some physical insight into the components. The time series, Fig. 5, of the linear and nonlinear PCs are quite similar overall. Even the succeeding PCs and NLPCs are qualitatively quite similar(not shown). From the time series in Figs. 5, it can be seen that these first components are dominated by the seasonal cycle. The observed variations across the

continent present a real challenge for the GCM simulations. A discussion of the ability of the AMIP GCMs to portray this field will be the subject of a forthcoming report, Boyle and Sengupta (1995). Figure 4 shows a pattern consistent with the central US being out of phase with the two coasts. This type of variation has been documented by Hsu and Wallace (1976), Horn and Bryson (1960) and others. The west and east coast have a wintertime maxima in precipitation attributable to cyclonic storms, while the mid-continent has a summertime maximum which is produced by convective events. The leading PC accounts for 25% of the variance.

The principal modes (linear and nonlinear) in Fig. 4 are very much alike although there are some differences in detail. The NLPM has a larger variation for the far western region and reduced extrema in the Central and Eastern regions compared to the PV. In the NLPM data the maximum off the east coast shifted northeastward, and the minimum in the central US shifted northward with respect to its linear counterpart. The NLPM makes a sharper distinction between the maxima on the Gulf Coast and off the Eastern US. The NLPM has negative values over Florida and to the east.

An RMS and mean absolute difference value was computed for the difference between the input data and the first neural network output for all the 120 months over the grid. Similar values were computed in the linear case using the first principal component approximation. The NLPC RMS and mean absolute difference was 10% less than that of the PC. This indicates that overall the two methods have a similar ability to fit the data, with the NLPC being the slightly better fit. In the time series the PC and NLPC techniques (Fig. 5) evince the dominance of the seasonal oscillation as might be expected from the close match of the plots in Fig. 4, the time series of the first PC and NLPC both exhibit similar temporal behavior.

Although the time series in Fig. 5 are similar, the NLPC tends to have a sharper transition from the winter to summer season. Figure 6 displays the results of a maximum entropy spectral estimation performed on the time series of Fig. 5. The figures show that the NLPCs have the bulk of the power in the first component while the PCs have a sizeable contribution from the second PC in the seasonal frequency. In this sense the NLPC does a much better job in isolating the seasonal cycle in the primary component. Figure 7 is a plot of the mean winter (DJF) data minus the mean summer (JJA) data for the raw observed data set. Figure 7 more closely resembles the data in Fig. 4a, the NLPC, than Fig. 4b, the PC. The ability to be able to isolate the primary seasonal forcing might prove to be quite useful as the GCMs improve and come under closer scrutiny. As the next section will indicate the current simulations are probably not accurate enough to justify this level of analysis.

## **5.0 CCM2 simulation**

Figure 8 displays the leading NLPM and PV for the precipitation data from the CCM2 AMIP simulation. As in the case with observations, the two patterns are quite similar. The region of positive values centered on Arkansas in the NLPM is much reduced in the PV. The zero line east of the Carolinas is moved farther westward in the NLPM compared to the PV field. The RMS computation reveals that the NLPC has a 11% closer fit to the input data than the PC. Although the overall fit to the data are similar, the time series in Fig. 9 indicates that the NLPM and PV are fitting the individual months differently as was the case in the observed analysis. However, as can be seen by the two time series, the model tends to have a great deal more regularity in the seasonal

cycle than did the observations. This results in the patterns in Fig. 8 appearing to be quite similar, the main differences being in the magnitudes of the extrema rather than their positions.

The leading NLPC time series exhibits an extreme regularity for the seasonal cycle. There is an obvious summer and winter regime with sharp transition. The PC data, Fig. 6a, also shows a regularity but it does not plateau about the solstitial seasons like the NLPC. The leading NLPC has isolated the bulk of the seasonal cycle and the remaining modes are involved in describing the interannual variations. In this case the NLPC highlights the regular nature of the CCM2 simulation when compared to the observed variations for the decade. The most obvious interruption of the quite regular pattern in Fig. 9a occurs in the winter of 1982/83. This is the winter in which the strongest El Niño - Southern Oscillation (ENSO) event on record occurred. The nonlinear analysis quite graphically isolates the impact of this event on the precipitation pattern over the US.

Comparing the results of the model in Fig. 8 to the observations in Fig. 4 indicates that the model is very much different from the observations. The difference is great enough that the gain achieved by the NLPC analysis is not really of practical value since even a cursory examination of the linear components is sufficient to show the model problems. However, as the models gain in sophistication and accuracy there will be a need for analysis that captures the more subtle aspects of the phenomena that are modeled.

## **6.0 Conclusions**

The results indicate that the nonlinear mode of analysis for the data reported here offers only a small gain in terms of overall reduction of the data with respect to its conventional linear counterpart. While this is perhaps disappointing, it does indicate that the well known and commonly available linear techniques are providing a viable description of these types of data. The NLPC does fit the data with greater fidelity but at the cost of significantly greater computational complexity. The additional complexity has not yet been shown to yield sufficient benefit to make its regular use worthwhile. A comparison of the leading PV or NLPM of the observations in Fig. 4 and the CCM2 simulation in Fig. 8 shows that for this model the analysis need not be concerned with close fits to demonstrate the model deficiencies. The model is grossly in error along the Eastern US for even such basic variations as can be attributed to the seasonal cycle. Nonetheless, the NLPC does show some promise. Comparing Figs. 4 and 6, the time series of the NLPC dramatically indicate the substantial differences in the nature of the variation of the model and observations. As the models improve, the abilities of the NLPC to discern differences might become more valuable. Acknowledgments. The generosity of the modeling groups involved in AMIP in making their results available is greatly appreciated. This work was performed under the auspices of the Department of Energy Environmental Sciences Division by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48.

## Appendix

### Auto-associative linear networks and principal components

In a linear network (see Fig. 1) there are three layers, the input layer, say with  $n$  nodes, the output layers, say also with  $n$  nodes (for the purpose of this work) and a single hidden layer with  $p$  ( $p \leq n$ ) linear nodes each executing the identity function without any loss of generality. Accordingly, the relationship between the input vector  $\underline{x} = (x_1, x_2, \dots, x_n)'$  and the output vector  $\underline{y} = (y_1, y_2, \dots, y_n)'$  may be expressed in a matrix form:  $\underline{y} = A B \underline{x}$ , where  $A$ , a  $(n \times p)$  matrix and  $B$ , a  $(p \times n)$  matrix represent the weights of the interconnections between the input and the hidden layer and between the hidden layer and the output layer respectively. Letting  $W = A B$ , we get  $\underline{y} = W \underline{x}$ . Thus if we have  $T$  training pairs  $(x_t, y_t)$ ,  $t=1, 2, \dots, T$  we can define the error function as

$$E(A, B) = \sum_{t=1}^T \|y_t - Wx_t\|^2$$

Notice  $E(A, B)$  is simply the sum of the squared residual for a linear regression of  $Y$  on  $X$ . It should be borne in mind that for any invertible matrix  $C$  the matrix product  $AB$  can be replaced by  $A C^{-1} C B$  and hence the matrices  $A, B$  are unique only up to an equivalence. Also,  $\text{rank}(W)$  is at most  $p$  and the expression  $E(A, B)$  is minimized by the regression 'slope' matrix  $S_{yx} S_{xx}^{-1}$  of the random vector  $Y$  on  $X$ , where  $S_{xx}$ ,  $S_{xy}$ , and  $S_{yx}$  are the sample covariances given by

$$S_{xx} = \sum_t x_t x_t', S_{xy} = \sum_t x_t y_t', \text{ and}$$

$$S_{yx} = \sum_t y_t x_t'$$

The following general result about the critical points and the global minimum of the error function has been established in Baldi and Hornik (1989):

Let  $S_{xx}$  be invertible and  $S$  be the  $(n \times n)$  matrix given by:

$$S = S_{yx} S_{xx}^{-1} S_{xy}$$

Assume that  $S$  is full rank with  $n$  distinct eigenvalues  $\lambda_1 > \lambda_2 > \dots > \lambda_n$ . Let  $I = \{i_1, i_2, \dots, i_p\}$  ( $1 \leq i_1 < i_2 < \dots < i_p \leq n$ ) be an ordered index set and  $U_I$  be the matrix of orthonormal eigenvectors of  $S$  associated with the eigenvalues  $\lambda_i, i \in I$ . Then two full rank matrices  $A$  and  $B$  define a stationary point of the error function  $E$  if and only if there exist an ordered index set  $I$  with  $p$  elements and an invertible  $(p \times p)$  matrix  $C$  such that

$$A = U_I C$$

and

$$B = C^{-1} U_1' S_{yx} S_{xx}^{-1}.$$

For such a stationary point the net weight matrix  $W$  is given by

$$W = P_{U_1} S_{yx} S_{xx}^{-1}$$

and the corresponding error by

$$E(A, B) = \text{Trace}(S_{yy}) - \sum_{i \in I} \lambda_i.$$

Here the notation  $P$  with a subscript  $U$  is used to mean the orthogonal projection operator corresponding to the  $(n \times p)$  matrix  $U$  (that is projection onto the subspace spanned by the column vectors of  $U$ ) and is given by

$$P_U = U(U'U)^{-1}U'.$$

From the stationary value of  $E$  above it is clear that the globally maximizing stationary value is obtained when the  $p$  largest eigenvalues are chosen corresponding to the ordered index set  $\{1, 2, \dots, p\}$ .

Specializing to the auto-associative case, we have  $y_t = x_t, t = 1, 2, \dots, T$ . Hence  $S_{yx} = S_{xx}$  and we have the global minimum when

$$A = U_1 C,$$

$$B = C^{-1} U_1', \text{ and } W = P_{U_1}$$

In summary, the unique global minimum value of the error function is obtained when the overall weight matrix  $W$  is the orthogonal projection onto the space spanned by the  $p$  dominant eigenvectors of the matrix  $S_{xx}$ . Furthermore, when  $C$  is chosen to be the identity matrix  $I_p$  the activities  $b_{x_t}, t=1, 2, \dots, T$  at the nodes of the hidden layer are given by  $u_1' x_t, t=1, 2, \dots, T$ , the first  $p$  principal components of  $x_t$ . When  $C$  is not the identity matrix  $I_p$  however, the activities at the hidden layer nodes are no longer individually identifiable as the principal components, although these activities will still form the projection on the same linear subspace.

## References

- Baldi, P. and K. Hornik, 1989: Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2, 53-58.
- Becker, S., 1991: Unsupervised learning procedures for neural networks. *International Journal of Neural Systems*, 2, 17-33.
- Bourlard, H. and Y. Karp, 1992: Auto-association by multi-layer perceptron and singular value decomposition. *Biological Cybernetics*, 59, 291-294.
- DeMers, D. and G. Cottrell, 1992: Non-Linear dimensionality reduction. In *Advances in Neural Processing Systems 5*, (eds. Hansen et al.), Morgan Kaufmann.
- Hack, J.J., B. A. Boville, B. P. Briegleb, J. T. Kiehl, P. Rasch and D. L. Williamson, 1993: Description of the NCAR Community Climate Model (CCM2). NCAR/TN-382+STR, National Center for Atmospheric Research, Boulder, CO 108pp.
- Hom, L. H. and R. A. Bryson, 1960: Harmonic analysis of the annual march of precipitation over the United States. *Ann. Assoc. Amer Geogr*, 50, 157-171.
- Hsu, C.-P., and J. M. Wallace, 1976: The global distribution of the annual and semiannual cycles in precipitation. *Mon. Wea. Rev.*, 104, 1093-1101.
- Kramer, M. A., 1991: Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37, 233-243.
- Oja, E., 1992: Principal components, minor components, and linear neural networks. *Neural Networks*, 5, 927-935.
- Preisendorfer, R. W., 1988: *Principal component analysis in meteorology and oceanography*, *Developments in Atmospheric Science* 17, 425pp., Elsevier, Amsterdam.
- Rumelhart, D. and J. McClelland, 1986: *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol 1, MIT Press., Cambridge, MA, 547 pp.
- Schemm, J. S., S. Schubert, J. Terry and S. Bloom, 1992: Estimates of monthly mean soil moisture for 1979-1989. *NASA Technical Memorandum* 104571, 254p.
- Sirat, J. A., 1991: A fast neural algorithm for principal component analysis and singular value decomposition. *International Journal of Neural Systems*, 2, 147-155.
- Spencer, R. W., 1993: Global oceanic precipitation from the MSU during 1979-91 and comparisons to other climatologies. *J. Climate*, 6, 1001-1326.

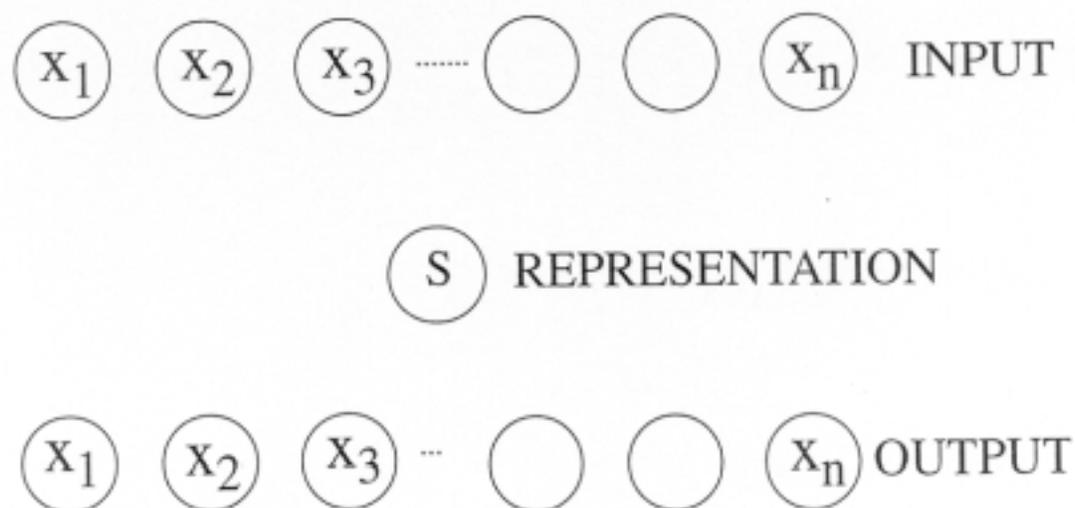


Figure 1. A schematic of the architecture of a simple 3 layer auto-associative neural network with a single node(S) in the hidden layer. The input and output are the same data ( $x_n$ ), thus it is auto-associative. This architecture is suitable for a single PC extraction. The x's represent the input data and output of the net

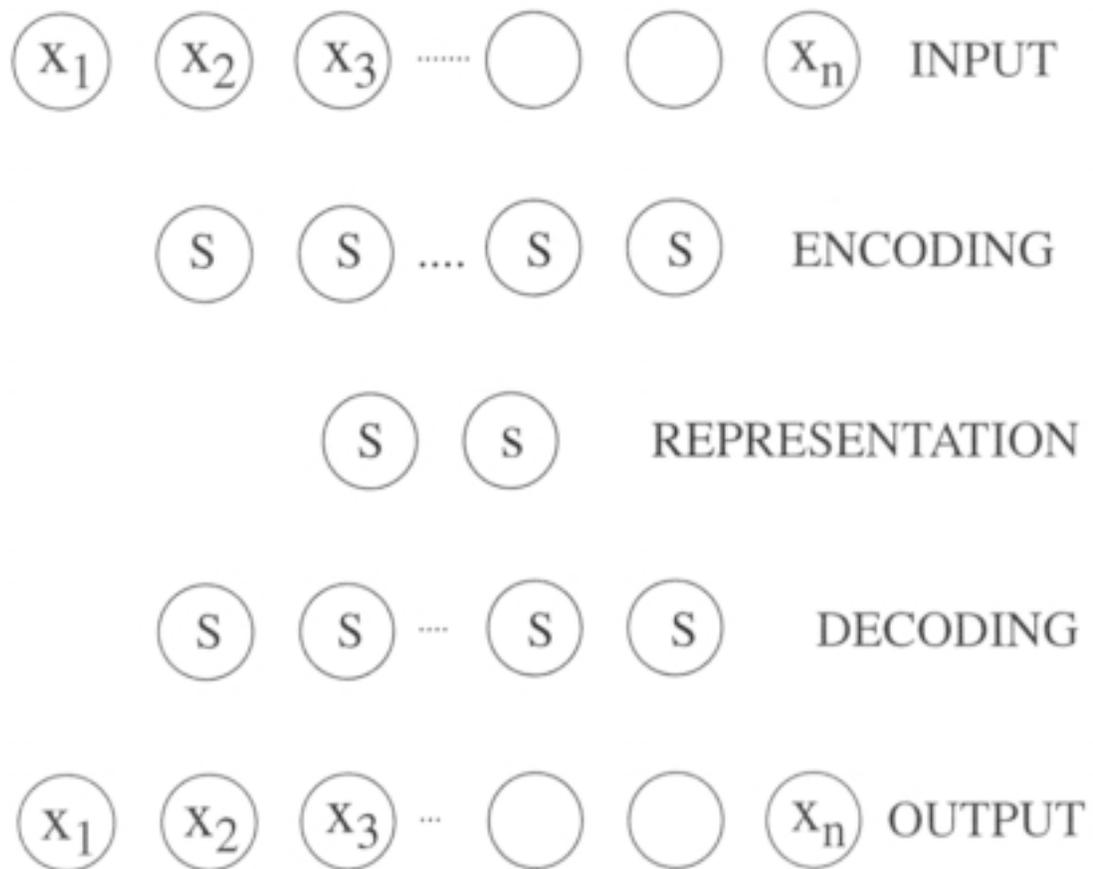


Figure 2. A fully interconnected ( connections not shown) auto-associative neural network architecture with three hidden layers. Showing from top to bottom: the input layer, encoding layer, the representation layer ( with two NLPC nodes ), the decoding and output layers. In the work reported here there is only a single node in the representation layer. The X's represent the input and output data, the S's are the network nodes.



Figure 3. Sequential selection of non-linear principal components showing a sequence of  $p$  auto-associative networks  $A_0, A_1, \dots, A_{p-1}$  with each  $A_i$  configured like  $A$  in Fig. 1 (except with a single node in each representation layer). The I/O pair within each  $A_i$  during training are the successive residuals  $R_0, R_1, \dots, R_{p-1}$  with the original data matrix in  $R_0$ .

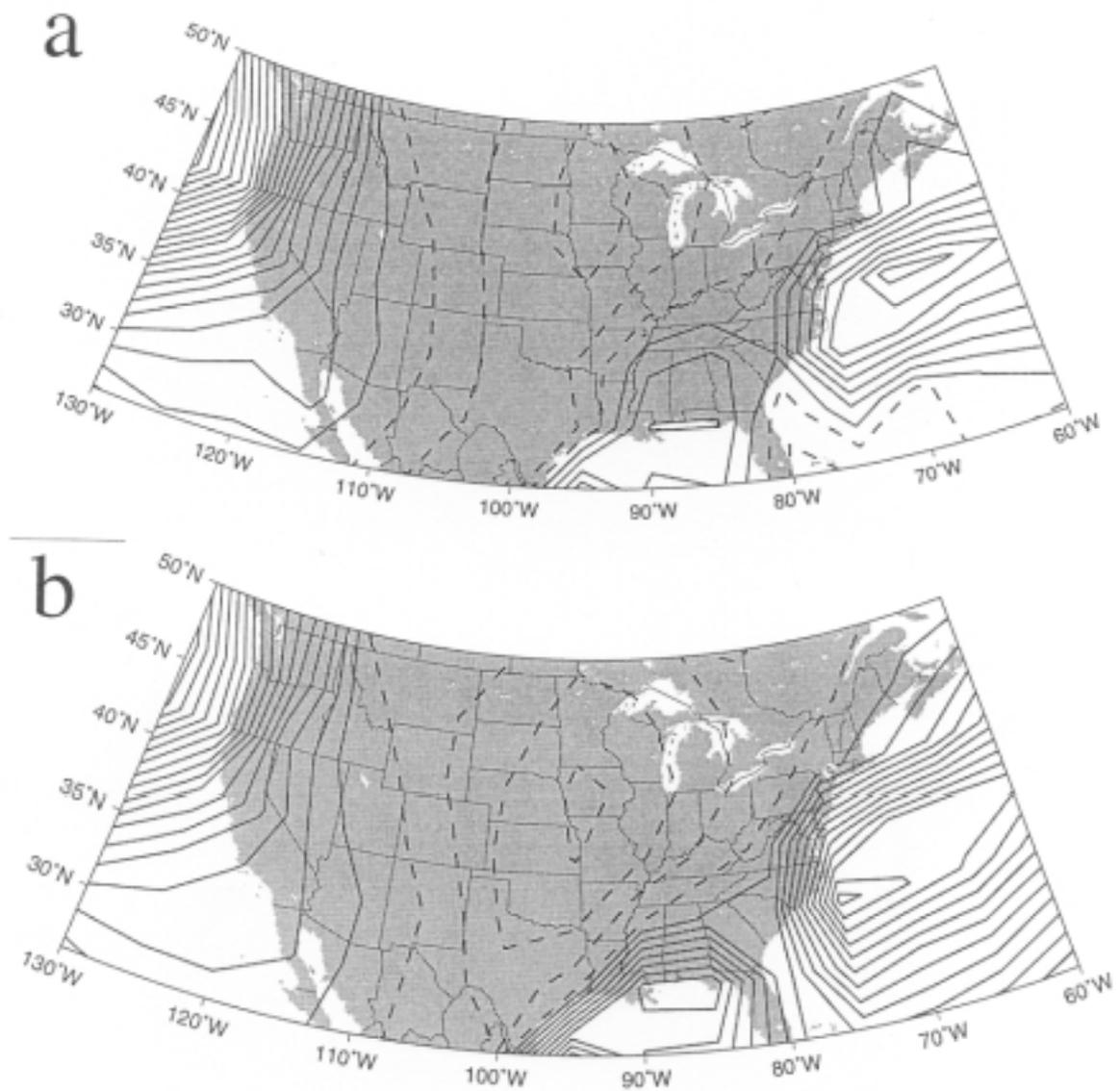


Figure 4. (a) The leading nonlinear principal mode of observed precipitation for the 1979-88 decade. This corresponds to the output of the initial sequence in Fig. 3. Dashed contours indicate negative values, the solid contours are zero and positive. The contour interval is 0.25. (b) As in (a) except for the first linear principal vector.

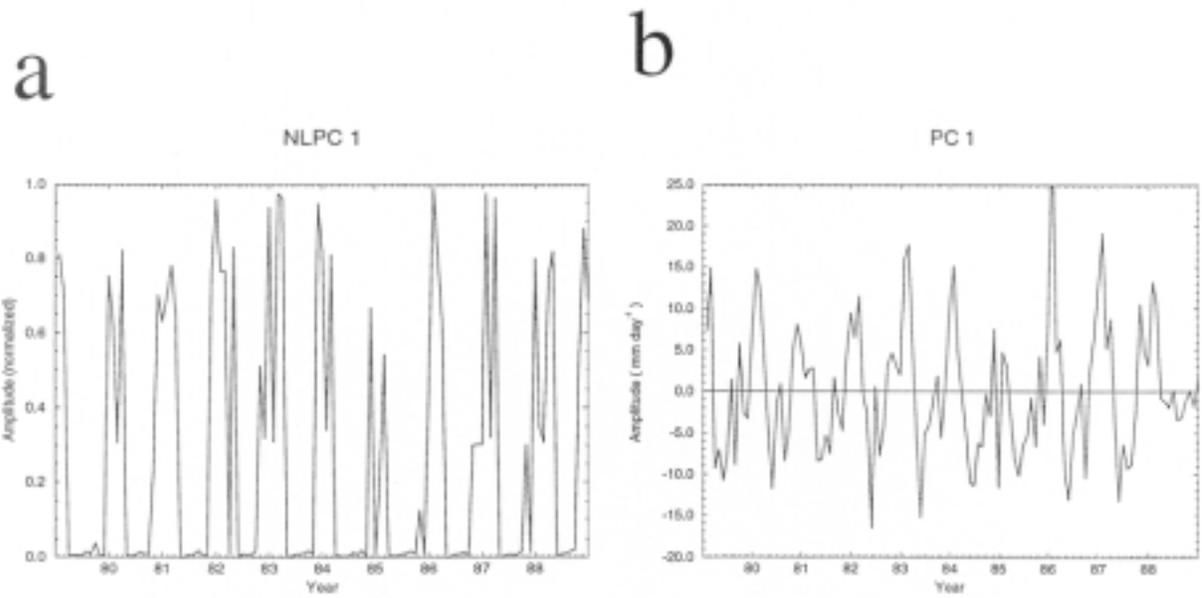
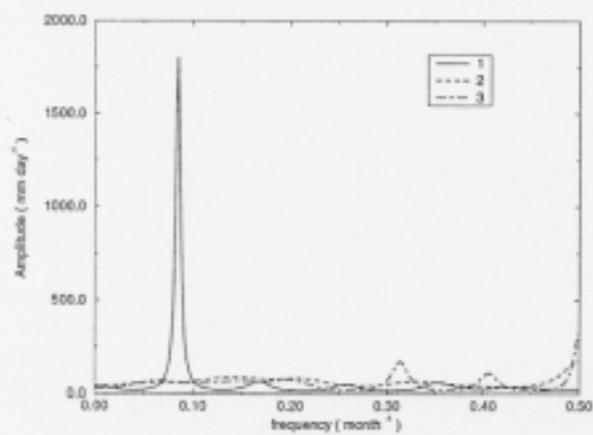


Figure 5. (a) Time series of the projection of the data on the leading PV depicted in Fig. 4b for the observed precipitation. The year indicated along the abscissa is plotted on the January of that year.  
 (b) The leading NLPC computed for the observed precipitation.

a



b

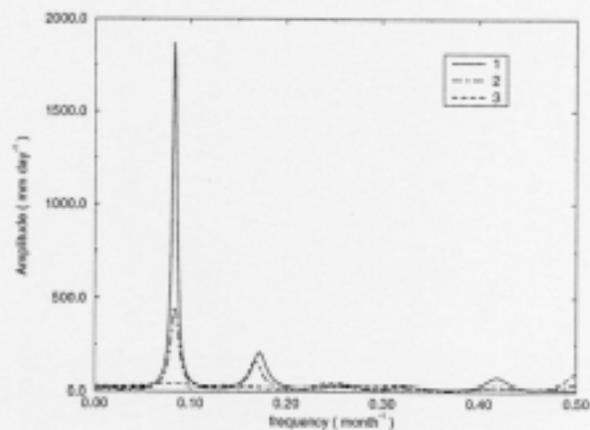


Figure 6. (a) Spectra of the three leading PCs using maximum entropy estimation. The leading PC shown in Fig. 5 is depicted by the heavy, solid line. (b) As in (a) except for the NLPCs.

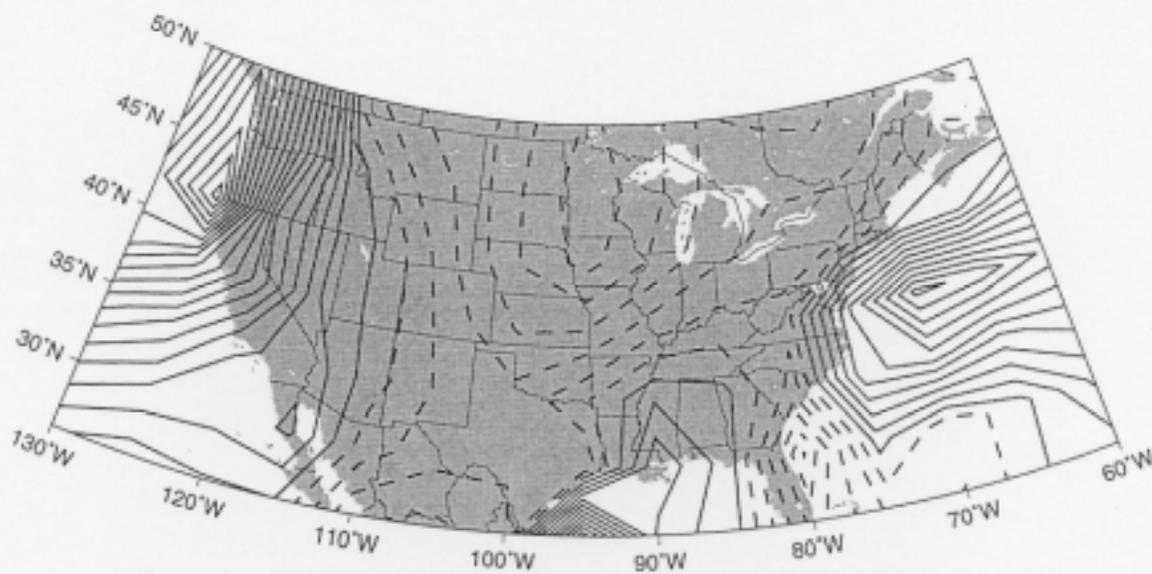


Figure 7. Contours of the difference between the mean winter (DJF) precipitation and mean summer (JJA) precipitation for the observed data set. Contour interval is 10 mm/month, the dashed lines indicate negative values (JJA > DJF), the solid contours are zero and positive.

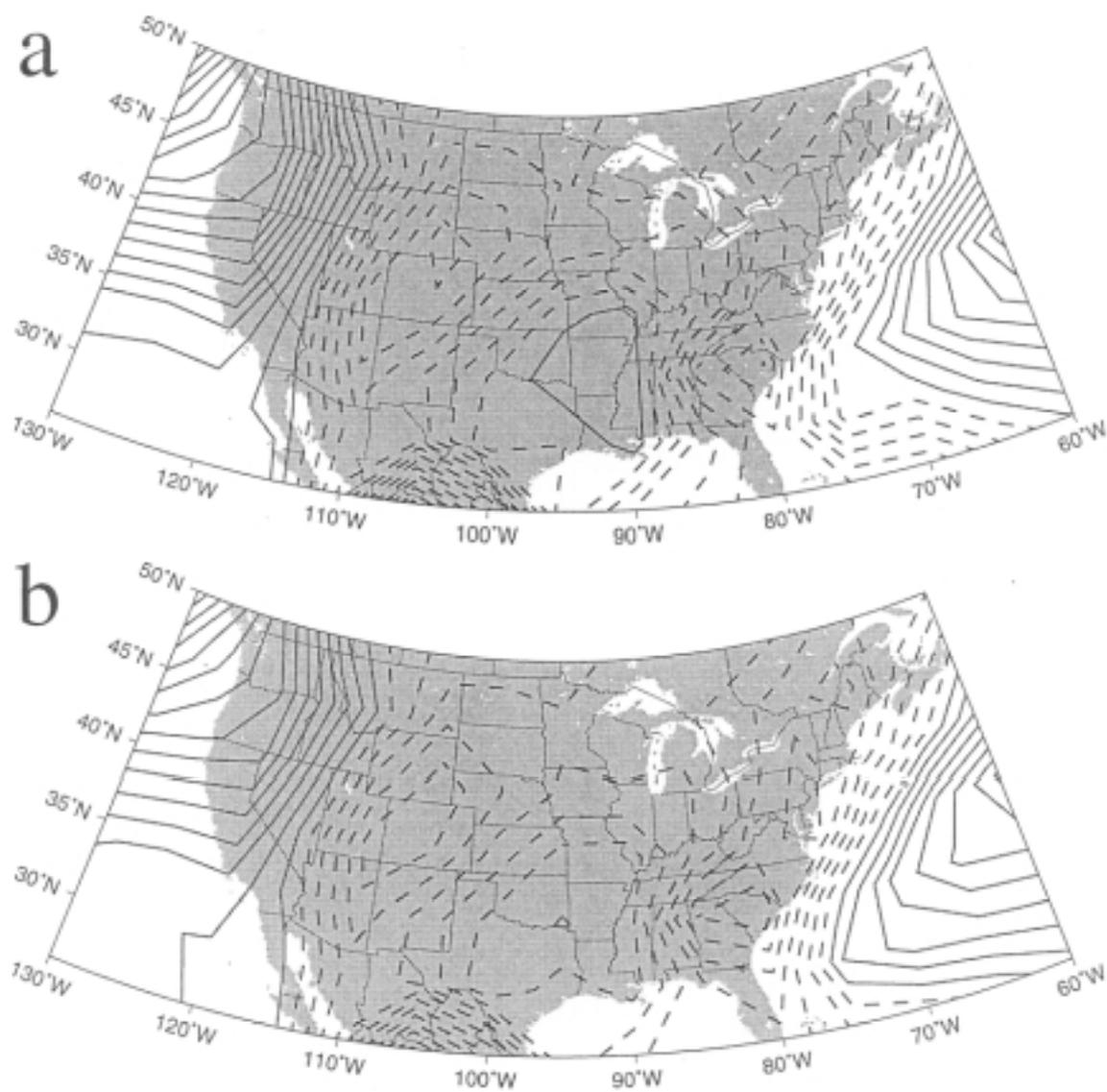
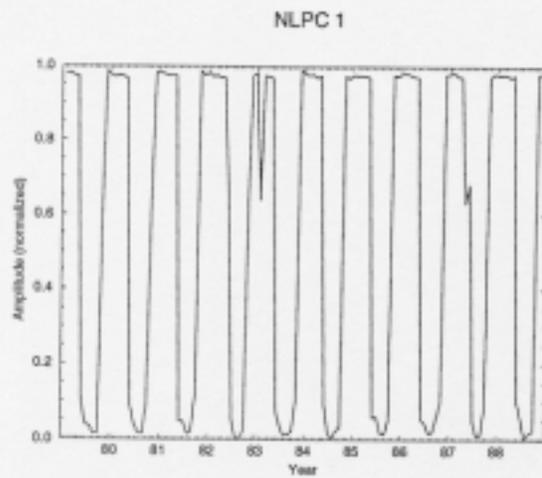


Figure 8. As in Fig 4. except for the CCM2 model simulation.

a



b

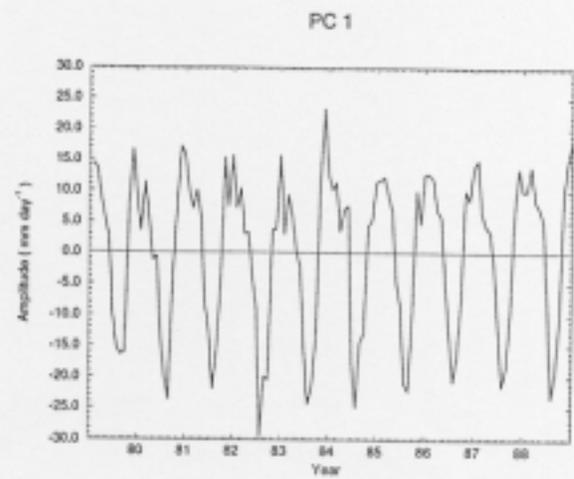


Figure 9. As in Fig. 5 except for the CCM2 simulation.