NON-LINEAR CYCLIC REGIMES OF SHORT-TERM CLIMATE VARIABILITY

by

PERRY SIH

B.Sc., The University of British Columbia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Physics and Astronomy)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Physics and Astronomy

The University of British Columbia
Vancouver, Canada

Date October 10, 2003

DE-6 (2/88)

Abstract

The Circular Non-linear Principal Component Analysis (CNLPCA), a variation of the non-linear version of the traditional Principal Component Analysis (PCA), is introduced. It is then applied to monthly-averaged geopotential heights of the NASA Goddard Institute for Space Studies SI2000 Global Circulation Model (GCM). It is shown that height variability in the model troposphere and stratosphere is essentially linear, even with different aerosol forcings. The daily-averaged model output show weak non-linearity. When CNLPCA is applied to observed geopotential height data, cyclic behaviour appears. The preferred states of the climate system can be seen. This cyclic behaviour can be tracked by recording the phase angle, a unique feature of the CNLPCA. By doing so, the preferred direction, as well as the frequency, of the cyclic behaviour can be found.

Contents

List of Figures

Chapter 1

Introduction

Our atmosphere is a non-linear system. Although we can learn a lot from modelling it as a linear system, it is important to investigate its non-linear properties in order to gain more understanding. In 1991, Kramer [1] introduced the non-linear version of the Principal Component Analysis (NLPCA) using a five-layer feed-forward neural network. It is successful in modelling non-cyclical non-linear datasets, such as the Lorentz attractor [2]. However, since it is only capable of producing open, U-shaped curve solutions, the NLPCA is not good at modelling cyclical signals within a dataset. In 1996, Kirby and Miranda [3] modified the NLPCA by replacing the single neuron at the bottleneck layer by the circular neurons. By doing so, closed curve (loops) solutions, as well as open curve solutions, can be obtained. When there exists a closed curve solution, the system is considered cyclical.

When applying the Circular NLPCA to the geopotential height fields, we will come across the Pacific North America (PNA) teleconnection pattern, the North Atlantic Oscillation (NAO), as well as the Arctic Oscillation (AO).

The PNA pattern is one of the most important modes of low-frequency climate variability, especially during the Northern Hemisphere winter [4]. The NAO is a north to south dipole oscillation of geopotential height anomalies across the North Atlantic Ocean [5]. The AO is a circulation pattern in which the atmospheric pressure over the polar regions varies in opposition with that over middle latitudes on time scales ranging from weeks to decades [6]. The NAO and AO are similar patterns [7], but the AO covers more of the Arctic Ocean. These three main spatial patterns of atmospheric variability will combine in a non-linear system and regime behaviour can be observed.

Regime behaviour in our atmosphere has been observed and documented ([8], [9], [10], and [11]). These regimes represent quasi-stationary atmospheric states. Transitions between regimes are fast when compared to the time spent within regimes. The Circular NLPCA is a convenient way to follow the evolution of the atmosphere, and has potential to be used in forecasting.

Chapter 2

The Circular Non-linear Principal Component Analysis

It is best to break down the introduction to CNLPCA into several steps. Since the CNLPCA uses principal components as input, PCA will be discussed briefly. The NLPCA will be introduced next, and then we will modify it with a circular node to get the CNLPCA.

2.1 Principal Component Analysis

For a given dataset that can be expressed as a k-dimensional spatial vector that evolves in time,

$$\mathbf{x}(t_n) = (x_1(t_n), x_2(t_n), ..., x_k(t_n)) \tag{2.1}$$

where t_n is the nth observation time. The Principal Component Analysis looks for a $u_1(t_n)$, which is a linear combination of the components of $\mathbf{x}(\mathbf{t_n})$:

$$u_1(t_n) = \mathbf{x}(t_n) \cdot \mathbf{e_1} \tag{2.2}$$

so that the approximation to $\mathbf{x}$

$$\hat{\mathbf{x}} = u_1 \mathbf{e_1} \tag{2.3}$$

is such that the mean square error between $\mathbf{x}$ and $\hat{\mathbf{x}}$

$$J = \left\langle ||\mathbf{x} - \hat{\mathbf{x}}||^2 \right\rangle \tag{2.4}$$

is minimized. From the residuals, $\mathbf{e_2}$ can be found using the same method. In general,

$$\hat{\mathbf{x}}(t_n) = \sum_{j=1}^{l} [\mathbf{x}(t_n) \cdot \mathbf{e_j}]\mathbf{e_j} \tag{2.5}$$

or if we write it in another way,

$$\mathbf{x}(t_n) = \sum_{j=1}^{l} [\mathbf{x}(t_n) \cdot \mathbf{e_j}]\mathbf{e_j} + \epsilon_n \tag{2.6}$$

where ϵ_n are the residuals. We can see that this is a type of *feature extraction problem*, in which

$$\mathbf{x}(t_n) = \hat{\mathbf{f}}(\mathbf{f}(\mathbf{x}(t_n))) + \epsilon_n \tag{2.7}$$

where $\mathbf{f}$ is a projection function and $\hat{\mathbf{f}}$ is an expansion function. We can see, by comparison, that the projection and expansion functions in the PCA are both linear. This means that the coordinate axes that come from the orthogonal eigenvectors, $\hat{\mathbf{e}}_i$, are straight lines. Therefore, the PCA is optimal only if the feature to be extracted can be characterized by such a set of axes. For example, data clouds that have the shapes of an ellipsoid or a cylinder can be described well by PCA, but data clouds that have the shapes of rings or bows can hardly be characterized by a linear model. Non-linear PCA can provide a better characterization.

2.2 Non-linear PCA

The NLPCA allows the projection function $\mathbf{f}$ and the expansion function $\hat{\mathbf{f}}$ to be non-linear. The solution to the feature extraction problem (equation (2.6) subject to (2.4)) can be implemented using a five-layer feed-forward neural network [1].

A feed-forward neural network is made up of a number of parallel layers of processing units, which are called neurons. The output from each neuron in the ith layer will be used as the input in the $(i+1)$th layer. Let $y_j^{(i)}$ be the output of the jth neuron of the ith layer, then a feed-forward network can be summarized as the following:

$$y_k^{(i+1)} = \sigma^{(i+1)} \left[\sum_j w_{jk}^{(i+1)} y_j^{(i)} + b_k^{(i+1)} \right] \tag{2.8}$$

where $w_{jk}^{(i+1)}$ is the *weight function* or *synaptic strength* and $b_k^{(i+1)}$ is the *bias.* $\sigma^{(i+1)}$ is the *transfer function* that characterizes the $(i+1)$th layer and it can be linear or non-linear. Cybenko [12] discovered that it is possible to approximate to arbitrary accuracy any continuous function from k dimensions to l dimensions by a three-layer neural network with k input neurons, hyperbolic transfer functions in the second layer, and linear transfer functions in the third layer with l neurons. At the third layer, the solution comes from compressing the original data to a one-dimensional time series, therefore $\hat{\mathbf{x}}$ can be considered as the optimal one-dimensional approximation to $\mathbf{x}$, embedded in k-dimensional space. In order to visualize the embedded feature, a second network can be used to map the extracted feature from l dimensions back to k dimensions. In general, we can use one network in place of $\mathbf{f}$ to go from k to l dimensions, and another one in place of $\hat{\mathbf{f}}$ to go from l to k dimensions. We shall use the example in figure 2.1 to demonstrate the details.

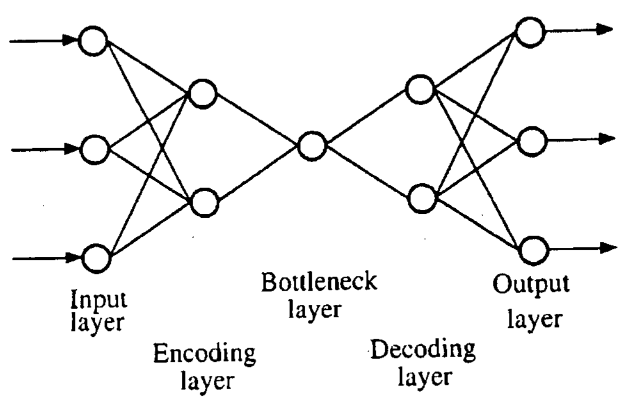

Figure 2.1: The five-layer feed-forward neural network used in NLPCA. This figure can be found in [13].

In this particular model, the neurons in the three middle layers are called *hidden* neurons, so called as they are not physically measurable quantities. There are two hidden neurons in the second and fourth layers, and thus we define the number of hidden neurons, m, to be 2 for this example.

The first layer is the input layer. In our example of figure 2.1, there are three input neurons. These three neurons, $y_j^{(1)}$, receive data (a three-component vector time series $\mathbf{x}(t_n)$ of principal components, for instance) presented to the network, therefore its transfer function, $\sigma^{(1)}$ is just the identity function.

The second layer is the encoding layer with two hidden neurons. The kth neuron in this layer will have the following as its input:

$$y_k^{(2)} = \tanh[w_{1k}^{(2)} y_1^{(1)} + w_{2k}^{(2)} y_2^{(1)} + w_{3k}^{(2)} y_3^{(1)} + b_k^{(2)}] \tag{2.9}$$

The third layer is the bottleneck layer and it consists of one single neuron. This means that the output of this neuron will be a one-dimensional time series. The transfer function is the identity function, and therefore, for this neuron u,

$$y_1^{(3)} = u = \mathbf{w}^{(3)} \cdot \mathbf{y}^{(2)} + b^{(3)} \tag{2.10}$$

We can impose a normalization condition such that we get unit variance, $\langle u^2 \rangle = 1$, by modifying the cost function:

$$J = \left\langle ||\mathbf{x} - \hat{\mathbf{x}}||^2 \right\rangle + \left(\left\langle u^2 \right\rangle - 1 \right)^2 \tag{2.11}$$

The fourth layer is the decoding layer. If we treat the bottleneck layer as the input neuron, then the transfer function will be the hyperbolic tangent:

$$y_k^{(4)} = \tanh(w_k^{(4)}u + b_k^{(4)}) \tag{2.12}$$

The fifth and final layer is the output layer. The transfer function is the identity function, and thus:

$$y_k^{(5)} = w_{1k}^{(5)}y_1^{(4)} + w_{2k}^{(5)}y_2^{(4)} + b_k^{(5)} \tag{2.13}$$

The final output is then:

$$\hat{\mathbf{x}} = \sum_{i=1}^{k} y_i^{(5)}\mathbf{e_i} \tag{2.14}$$

The cost function (equation (2.11)) is minimized by finding the optimal values of all of the weight and bias functions. This process is called "training the network."

One thing of notice is that since the system is non-linear with a large number of degrees of freedom ($2lm+4m+l+1$ without normalization, where l is the number of input neurons, and m is the number of neurons in each of the encoding and decoding layers), there exist many local minima in the space of J (equation (2.11)). Usually the minimization of J will end up in one of the local minima instead of the global minimum because the minimization algorithm is designed to move in the direction of decreasing value of the cost function. Therefore, one must inspect the minimization results from an ensemble of feed-forward neural networks with different initial weights and choose from it the case where the mean square error is the smallest.

Also, the encoding layer and the decoding layer are not required to have the same number of neurons. But generally it is fixed in order to reduce the number of free parameters in the model architecture.

It is possible to obtain better results if the input variables are appropriately scaled. This is because if the values of the variables vary by several orders of magnitudes, some of the weight functions will tend to be negligible when minimizing J. A good way to scale each variable is to first remove its mean and then divide by its standard deviation. If the leading PCs are used, then there is no need to scale the variables because the PCs are already scaled.

If the data cloud is concentrated but of irregular shape, there will be a greater chance of obtaining an overfitted solution. In order to avoid that, a certain fraction (for example, 20%) of the original data can be randomly chosen and excluded from the minimization. After the minimization is complete, the mean square error of the two portions can be compared. An overfitted solution is one that has a mean-square error (MSE) from the training portion lower than the MSE from the randomly selected test samples. Overfitted solutions can be rejected when determining the best solution.

If the underlying feature to be extracted resides in a two-dimensional space instead of a one-dimensional space, a feed-forward neural network with two bottleneck neurons can be used [2]. In general, the number of bottleneck neurons corresponds to the number of dimensions in which the feature being sought resides.

2.3 Circular NLPCA

A cyclic underlying feature containing three or more regimes is best characterized by a closed curve solution to the minimization problem. This is because each regime need to be connected to two regimes in order for the feature to be considered cyclic. In an open curve solution, the regimes at either end of the curve are connected to only one other regime. The system is forced to move towards one regime, and therefore such a feature is not considered cyclic. The NLPCA is not capable of obtaining closed curve solutions because the bottleneck neuron, u, is not an angular variable. It is possible to make the bottleneck variable an angular variable. Kirby and Miranda [3] introduced the circular neurons in the bottleneck layer. It is accomplished by using two bottleneck neurons, p and q, while adding the restriction that they are normalized:

$$p^2 + q^2 = 1 \tag{2.15}$$

This condition constrains the network to have only one degree of freedom with two neurons.

To build the CNLPCA network, we create the following two unnormalized neurons:

$$p_o = w_{1p}^{(3)} y_1^{(2)} + w_{2p}^{(3)} y_2^{(2)} + b_p^{(3)} \tag{2.16}$$

$$q_o = w_{1q}^{(3)} y_1^{(2)} + w_{2q}^{(3)} y_2^{(2)} + b_q^{(3)} \tag{2.17}$$

Using the normalization restriction (equation 2.15), we replace u in equation (2.10) by the following normalized bottleneck neurons:

$$p = \frac{p_o}{\sqrt{p_o^2 + q_o^2}} \tag{2.18}$$

$$q = \frac{q_o}{\sqrt{p_o^2 + q_o^2}} \tag{2.19}$$

Since the bottleneck layer has changed, the decoding layer (equation (2.12)) must change accordingly:

$$y_k^{(4)} = \tanh(w_{pk}^{(4)} p + w_{qk}^{(4)} q + b_k^{(4)}) \tag{2.20}$$

The rest (equations (2.9) and (2.13)) remains unchanged. See figure 2.2 for the schematics of the CNLPCA.

2.4 Non-linearity of a Neural Network

There are two ways to adjust the non-linearity of a neural network. One is to introduce a penalty factor, and the other is to adjust the number of hidden neurons.

The non-linearity of CNLPCA, as well as NLPCA, comes from the non-linear transfer function, the hyperbolic tangent. It has the property that

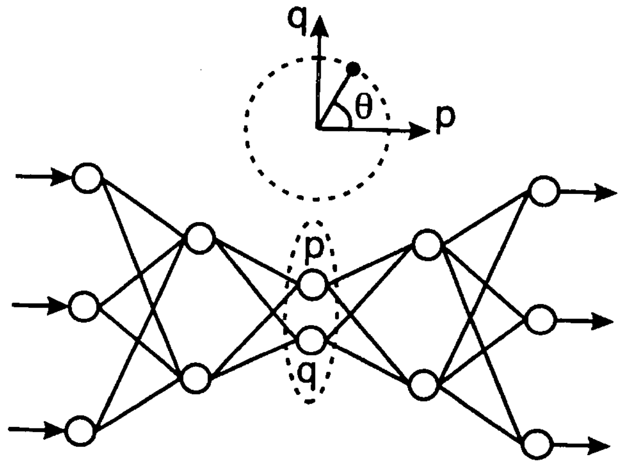

Figure 2.2: The five-layer feed-forward neural network used in CNLPCA. This figure can be found in [13].

for a large weight, w, $\tanh(wx)$ approaches the step function. On the other hand, if w is small, $\tanh(wx) \approx wx$, which is approximately linear. So, if the optimization arrives at a minimum where w is large, we may obtain a highly non-linear result. This can cause overfitting, in which noise is mistaken as signal. Therefore, if we can penalize large w and force w to be sufficiently small, then the non-linearity can be decreased. This can be done by modifying the cost funtion:

$$J = \left\langle ||\mathbf{x} - \hat{\mathbf{x}}||^2 \right\rangle + p \sum_{ik} \left(w_{ik}^{(2)} \right)^2 \tag{2.21}$$

Generally, to choose the appropriate penalty, we can choose a set of values between 0.1 and 1.0 and obtain a solution for each of them, and then choose the solution that yields the smallest MSE to be the solution of our choice.

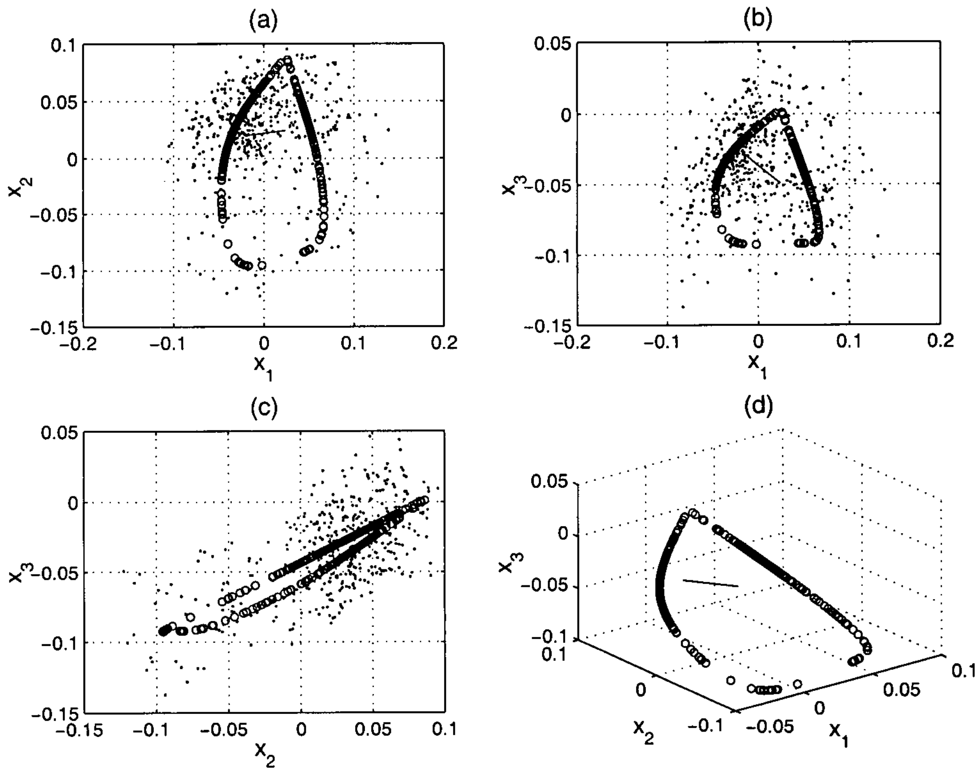

Figure 2.3: CNLPCA with $m = 3$, and a penalty factor of 0. The data are taken from the subsurface temperature of the tropical Pacific Ocean at a depth of 120m. Six PCs were used in the input and output layers, but only the first three are shown. (a) mode 2 vs mode 1; (b) mode 3 vs mode 1; (c) mode 3 vs mode 2; (d) a 3D view of the three modes. The original data are represented by the dots. The linear PCA solution is represented by the line. The CNLPCA solution is represented by the overlapping circles. The vertex at the top of (a) is a symptom of overfitting.

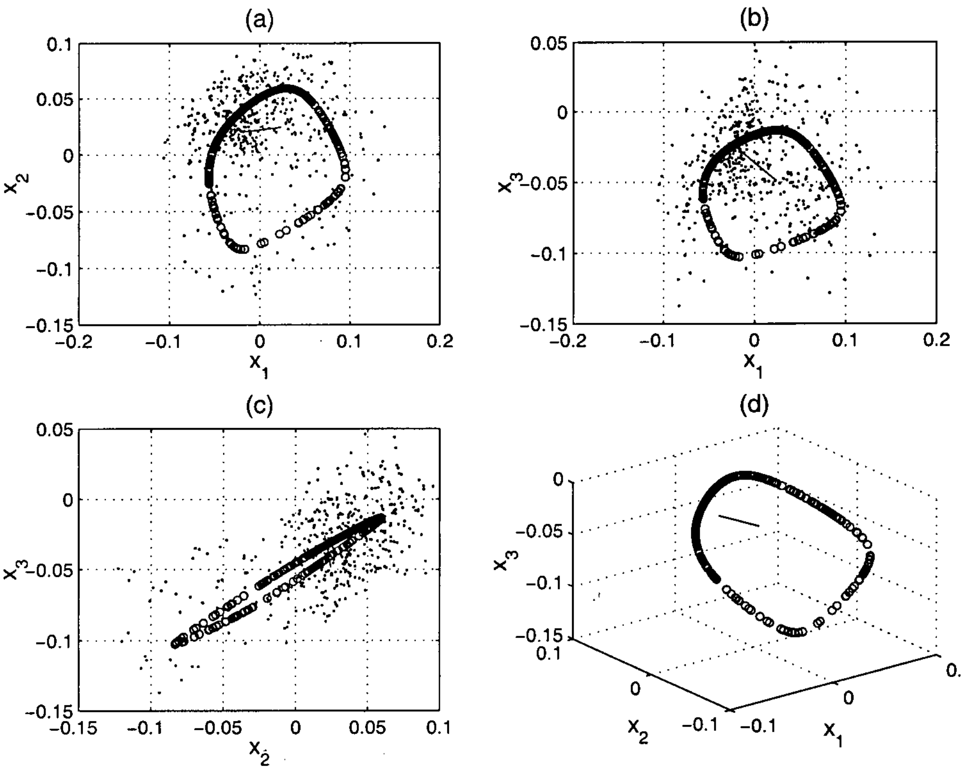

Figure 2.4: CNLPCA with $m = 2$ and a penalty factor of 0.6. Notice that with one fewer hidden neuron and a penalty factor, the solution is smoother than the last case. This is a better solution since it is not overfitted.

Notice that the MSE is always small when the penalty is close to zero because these are the overfitted solutions we are trying to avoid. Also notice that with a large penalty factor, the weights will go to zero in order to minimize the cost function. The neural network will thus become linear.

The number of hidden neurons also affects the non-linearity of the system greatly. In a neural network with l input neurons and m hidden nuerons, adding one extra hidden neuron will introduce $2l + 2$ extra variables to the minimization problem. Typically, more hidden neurons yield smaller mean square errors. But this is usually considered a case of overfitting. Usually, $m = 2$ or $m = 3$ will be sufficient.

An example can be seen in figures 2.3 and 2.4. A small data set containing monthly averaged subsurface sea temperature at 120m below sea surface from the Scripps Institution of Oceanography is analyzed using CNLPCA. In figure 2.3, three hidden neurons in each of the encoding and decoding layers and no penalty factor were used, thus allowing high degree of non-linearity. The result is a closed curve that overfit the data. The vertex at the top of figure 2.3a) is a symptom of overfitting. It is because of noise that moves the curve away from the centre such that it needs to take a sharp turn to fit the next point. The CNLPCA approximation using only two hidden neurons can be seen in figure 2.4. The resulting curve goes through a ring-shaped path defined by a dense upper branch and a scarce lower branch, in which the data points on either side of the curve are about the same distance away from the curve. figure 2.4 is a better fit because the curve is smoother and it takes the shape and the density of the data cloud. The short solid lines offer a comparison between the CNLPCA and the linear PCA. The linear solution simply goes through where the density of the data cloud is the highest, treating all the outliers as noise, whereas the CNLPCA takes some of the outliers into account. But, by how much should we take these outliers into account?

2.5 Search for Suitable Parameters

Since we are trying to fit a loop in multi-dimensional space, there are many local minima in the cost function (equation (2.21)) in which our solutions may be trapped. Therefore, it is important to find suitable parameters that would lead to the best possible solution.

Aside from the penalty factor and a suitable number of hidden neurons, a suitable filter is also useful in finding the best solution. We shall use the daily-averaged output of 500mb geopotential heights from the NASA Goddard Institute for Space Studies Global Climate Model (details in Chapter 4) as an example. In figure 2.5, no filter is used. As a result, the solution with the lowest mean square error is a linear one. However, if the number of hidden neurons is increased to three, as in figure 2.6, to allow for more non-linearity due to the lack of filtering, then the cyclical behaviour can be retrieved. However, a larger number of hidden neurons increase the risk of overfitting. Therefore, a filter is still recommended. This is illustrated by figure 2.7, in which two hidden neurons are used together with filtering. Since some high-

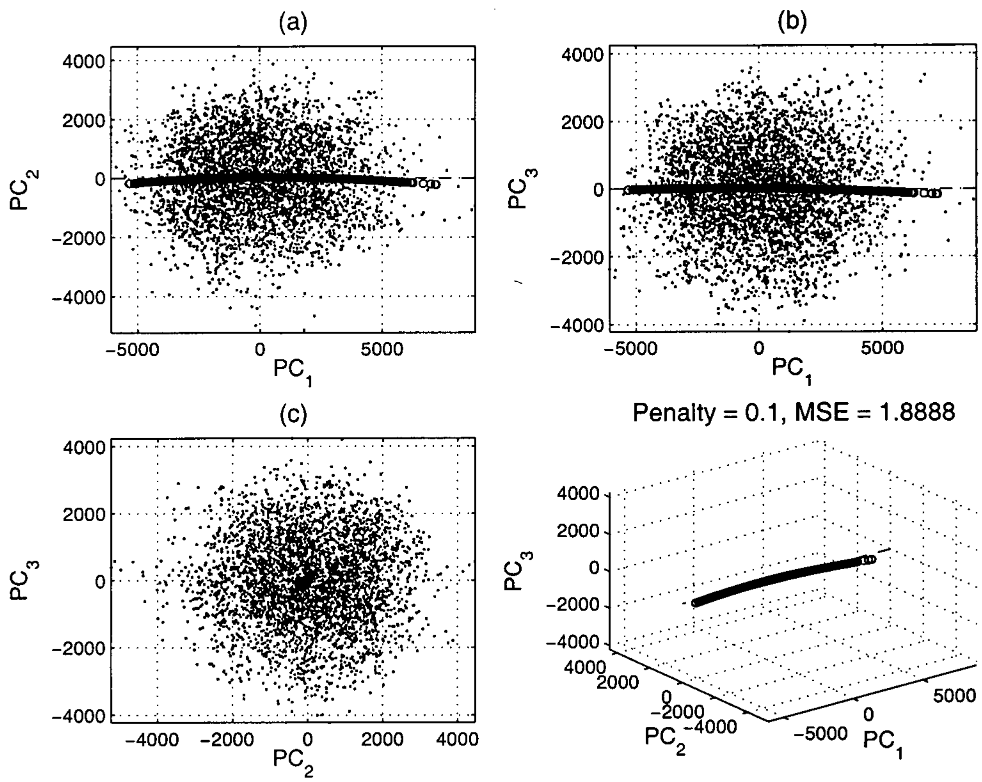

Figure 2.5: CNLPCA mode 1 for 500mb heights Trial A with no filtering. Two hidden neurons are used, and the result yields a linear solution even with a small penalty.

frequency noise is removed here using a 10-day low-pass filter, the averaging effect that would otherwise lead to a linear solution is also removed. A loop solution exists.

In general, non-linear solutions have lower mean square error. Although it is almost impossible to reach the global minimum during the minimization, a loop solution can bring insights into cyclical behaviour, as shown in the Chapter 5 using observed data.

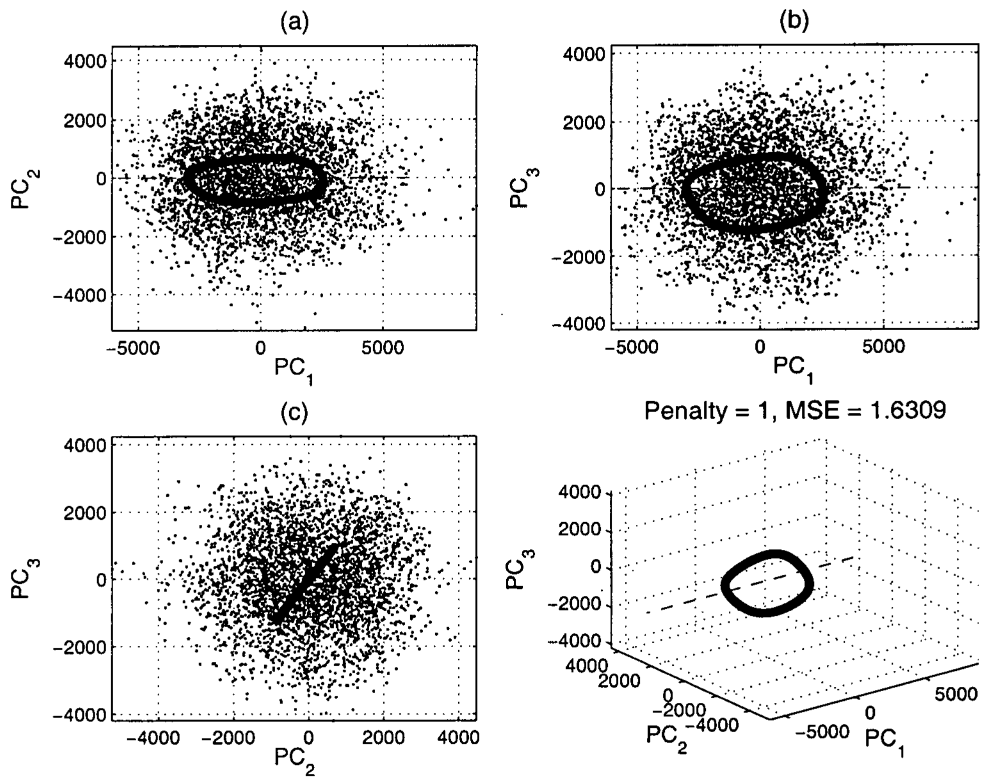

Figure 2.6: CNLPCA mode 1 for 500mb heights Trial A with no filtering. Three hidden neurons are used, and the best results come with penalty of 1.0. A larger penalty is needed to compensate for the non-linearity introduced by adding a third hidden neuron. Since the mean square error is lower than the previous case, figure 2.5, this is a superior result. In general, cyclical results yield lower MSE.

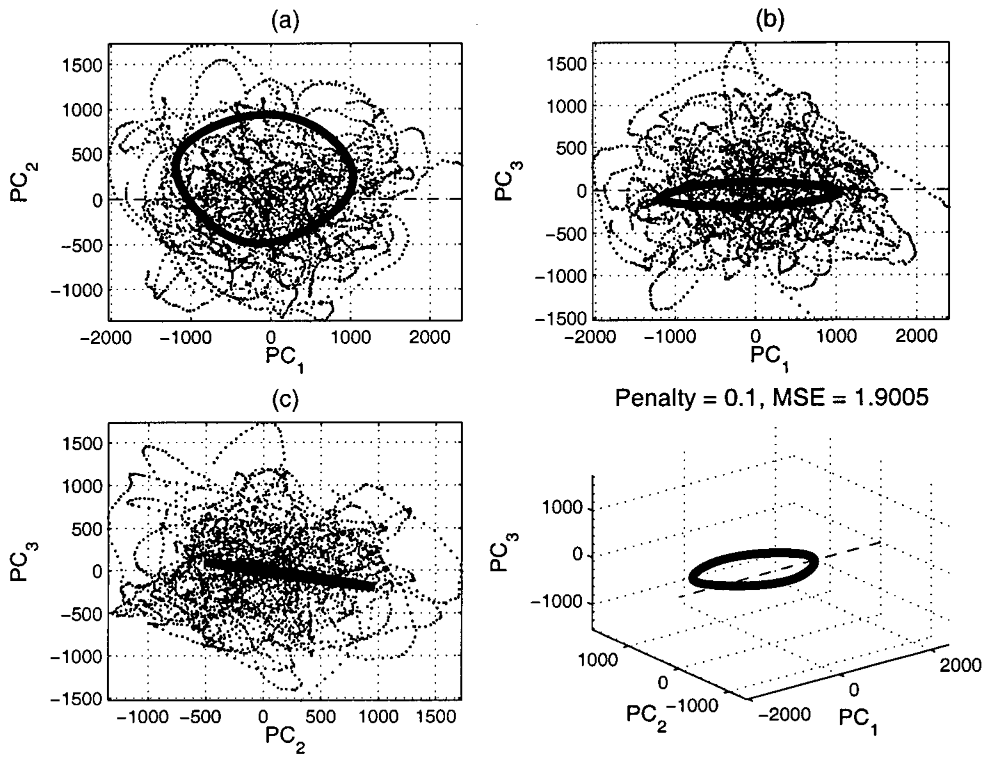

Figure 2.7: CNLPCA mode 1 for 500mb heights Trial A. The heights were filtered using a 10-day low-pass filter before being processed through the CNLPCA neural network. Two hidden neurons are used, but this time a loop solution is obtained. Removing noise reduces the averaging effect, and hence increase the observable non-linearity.

Chapter 3

Monthly-averaged Model Data

The CNLPCA is used to investigate if there exists any cyclical behaviour in a set of monthly-averaged model data. The model chosen is the Global Circulation Model version SI2000 from the NASA Goddard Institute for Space Studies. In particular, three sets of data with different atmospheric forcings were selected.

The B399 experiment includes no time-varying radiative forcings besides a climatological annual cycle that repeats itself. Boundary forcing comes from the observed sea surface temperature and sea ice, based on the Hadley data set HadISST 1 [14]. The B424 experiment takes into account six types of time-varying radiative forcings. These forcings are caused by variations in the concentrations of greenhouse gases, ozone, solar heating, stratospheric water vapour, stratospheric and tropospheric aerosols, and by variability in the solar heating. The B424o includes the same radiative forcings as B424, but its ocean representation has a specific "q-flux" horizontal heat transport, which adds diffuse mixing with the deep ocean for transient experiments.

All three data sets have a horizontal resolution of $4' \times 5'$, and a vertical resolution of twelve layers (nine in the troposphere and three in the stratosphere).

Each data set uses the same set of data processing techniques and parameters. Height value at each grid point is first multiplied by the squareroot of the cosine of its latitude in order to compensate for the increase in spatial density of the points with increasing latitude. The climatological annual cycle is calculated by averaging each month of the year, which is then subtracted from each grid point. Three-month winters are then extracted out of each year to obtain the winter anomalies.

For the B399 model at 300mb (first two PCs in figure 3.1 and their score in figure 3.2), the first PC is similar to the Arctic Oscillation (AO). The second PC has a ring-like positive structure surrounding an intense negative peak in North America. This is associated with the PNA pattern. At 30mb (first two PCs in figure 3.3 and their score in figure 3.4), the monopole oscillation is the dominant mode and the dipole oscillation is seen in the second PC.

The first PC of the B424o model at 300mb (first two PCs in figure 3.5 and their score in figure 3.6) shows a large positive peak around the Arctic Circle. This is associated with the AO. The second PC has three negative peaks, located on the Pacific and Atlantic Oceans, as well as the Northern Eurasian continent, surrounding a positive peak in the arctic portion of North America. This can be regarded as a cold ocean, warm land pattern. The PCs at 30mb (first two PCs in figure 3.7 and their score in figure 3.8) are

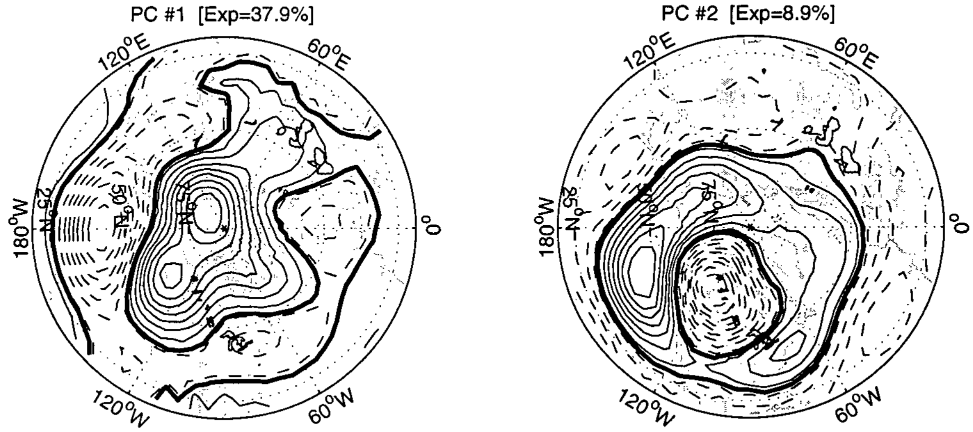

Figure 3.1: First two PCs of SI2000/B399 model at 300mb

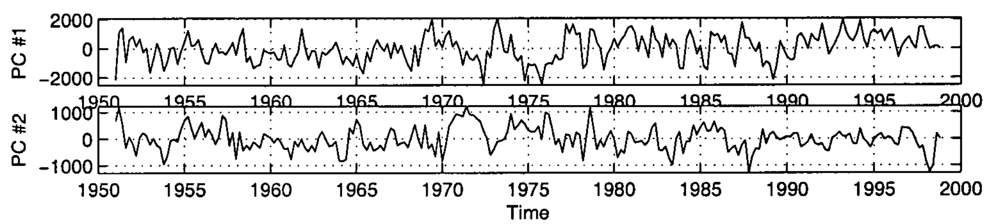

Figure 3.2: Scores of the first two PCs of SI2000/B399 model at 300mb

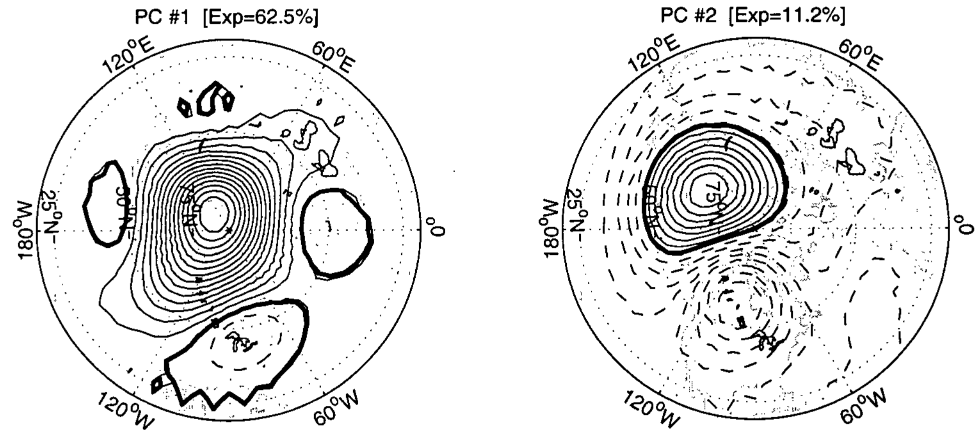

Figure 3.3: First two PCs of SI2000/B399 model at 30mb

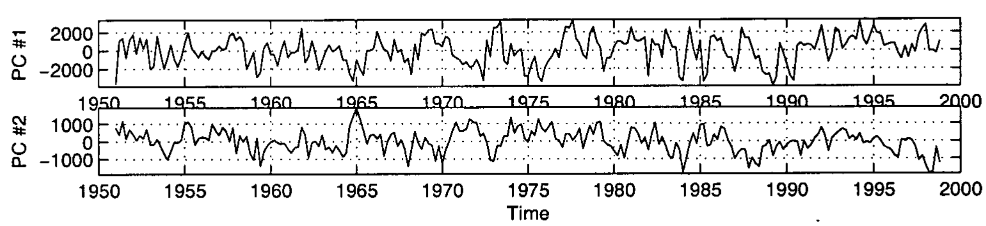

Figure 3.4: Scores of the first two PCs of SI2000/B399 model at 30mb

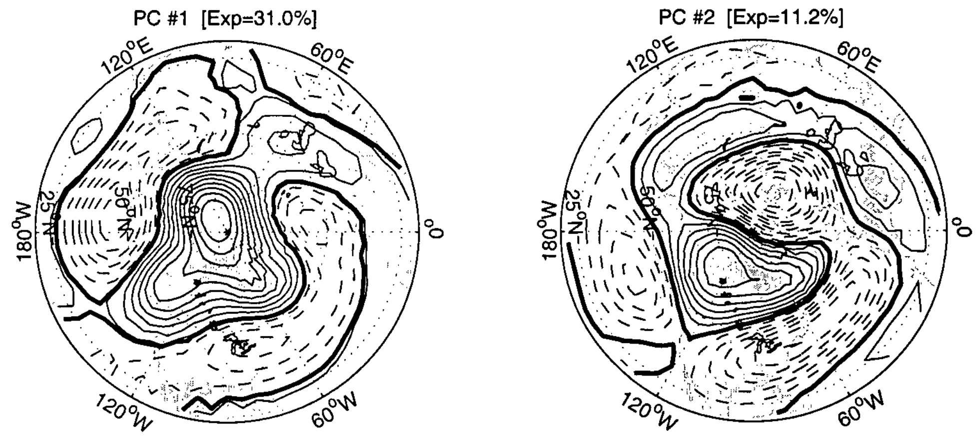

Figure 3.5: First two PCs of SI2000/B424o model at 300mb

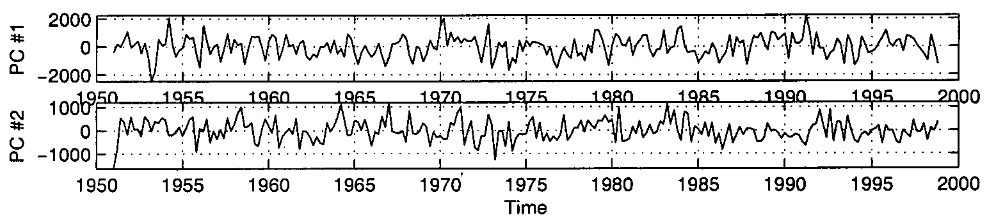

Figure 3.6: Scores of the first two PCs of SI2000/B424o model at 300mb

similar to that in the B399 model.

The first two PCs of the B424 model at 300mb (first two PCs in figure 3.9 and their score in figure 3.11) are similar to those of the B399 model. The PCs at 30mb (first two PCs in figure 3.10 and their score in figure 3.13) again show monopole and dipole oscillations, which are also similar to those of the B399 model. One thing of notice is that the percentage of variance accounted for by the first PC in the stratosphere in all the models are around 60%. This shows that the model stratosphere may be oversimplified.

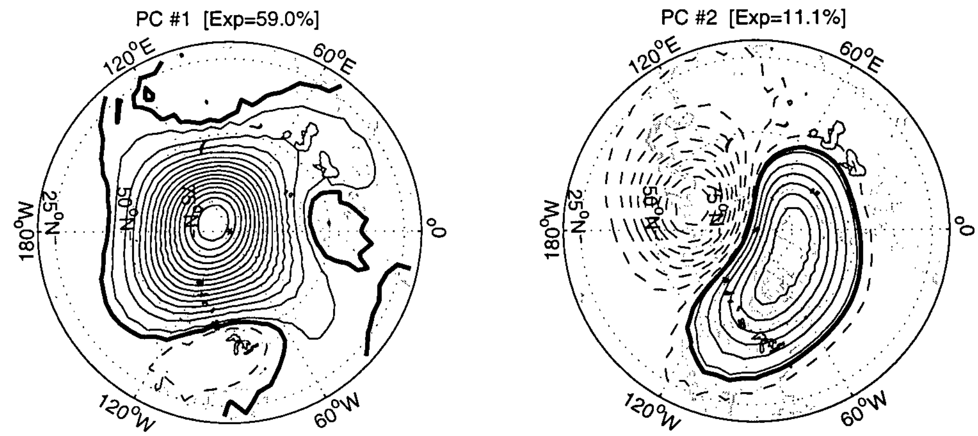

Figure 3.7: First two PCs of SI2000/B424o model at 30mb

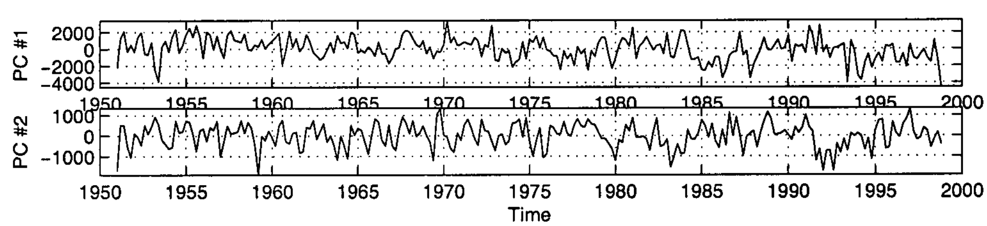

Figure 3.8: Scores of the first two PCs of SI2000/B424o model at 30mb

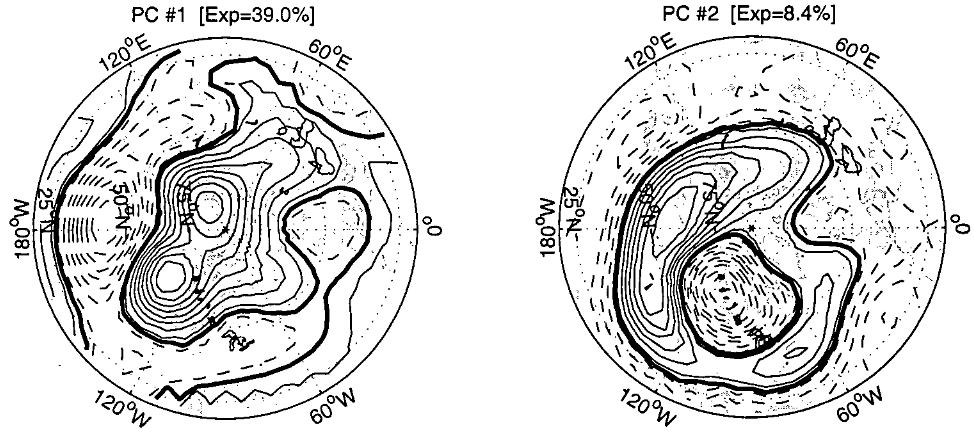

Figure 3.9: First two PCs of SI2000/B424 model at 300mb

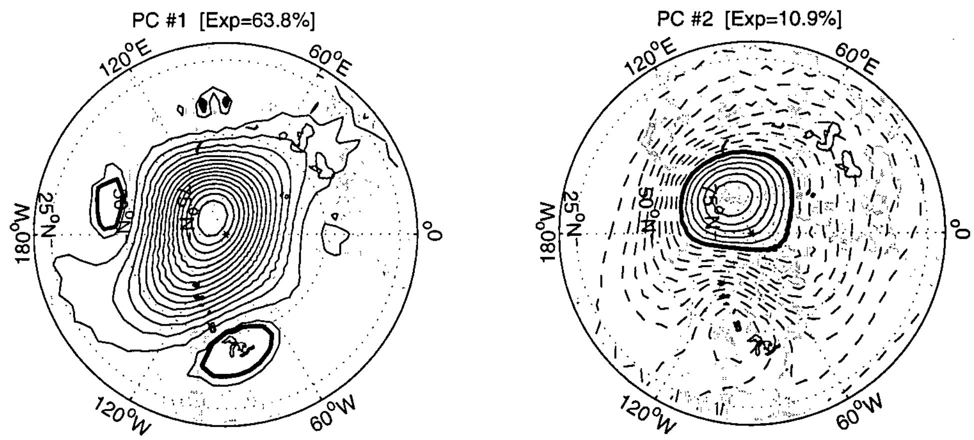

Figure 3.10: First two PCs of SI2000/B424 model at 30mb

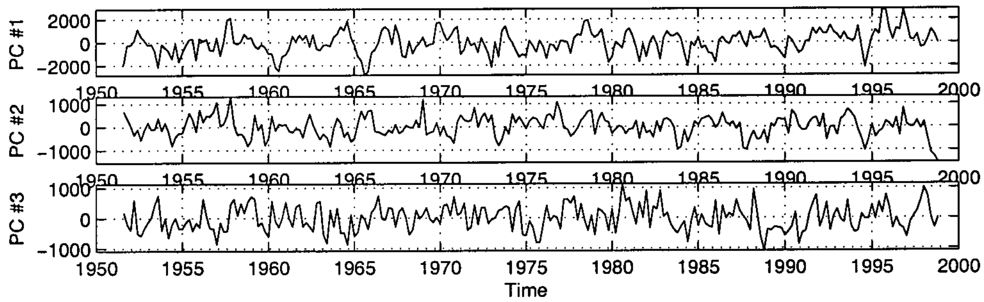

Figure 3.11: Scores of the first three PCs of SI2000/B424 model at 300mb

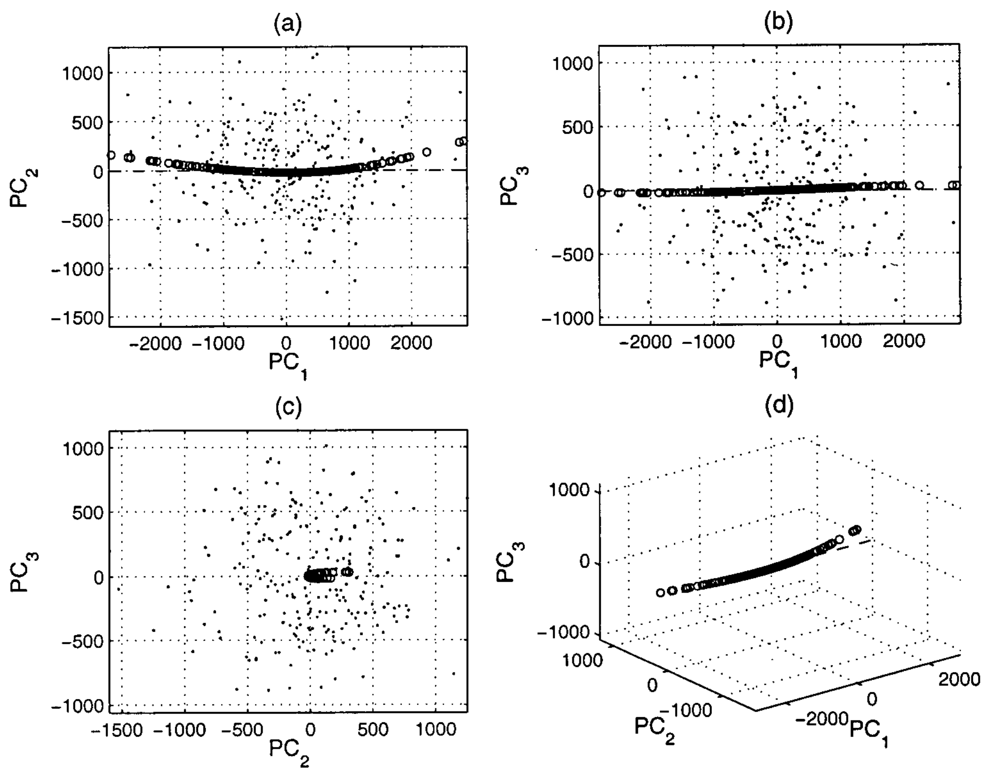

Figure 3.12: CNLPCA results of SI2000/B424 model at 300mb, three hidden neurons, penalty 1.2. The broken lines represent the linear PCA solutions. Slight non-linearity and no cyclic behaviour can be seen.

The scores of the first eight PCA modes ($l = 8$) are used as inputs to the CNLPCA neural network. Three hidden neurons ($m = 3$) are used in the hidden layers. A series of runs are conducted with different penalty factors, and the optimal ones are found for each model. We find that even though the PCs look different in some cases, the CNLPCA results are similar. Therefore, we shall use the B424 run to illustrate the results.

The scores of the first three PCA modes for the B424 runs can be seen in figures 3.11 and 3.13. The respective results can be see in figures 3.12 and 3.14. It can be seen that the CNLPCA solutions are very close to the PCA solutions. The PC2 magnitude in the 300mb run is only 10% of the PC1 magintude, and 20% in the 30mb run. Therefore we conclude that these models are essentially linear. Also, since the CNLPCA solutions do not yield closed loops, we conclude that these monthly-averaged models do not show cyclic behaviour.

If monthly-averaged model output data are essentially linear, then perhaps there is some non-linear behaviour when we shorten the averaging time interval [15]. So, next we look at the daily-averaged models.

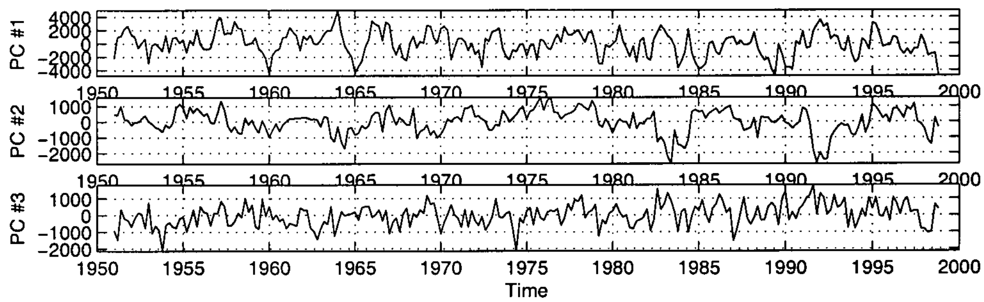

Figure 3.13: Scores of the first three PCs of SI2000/B424 model at 30mb

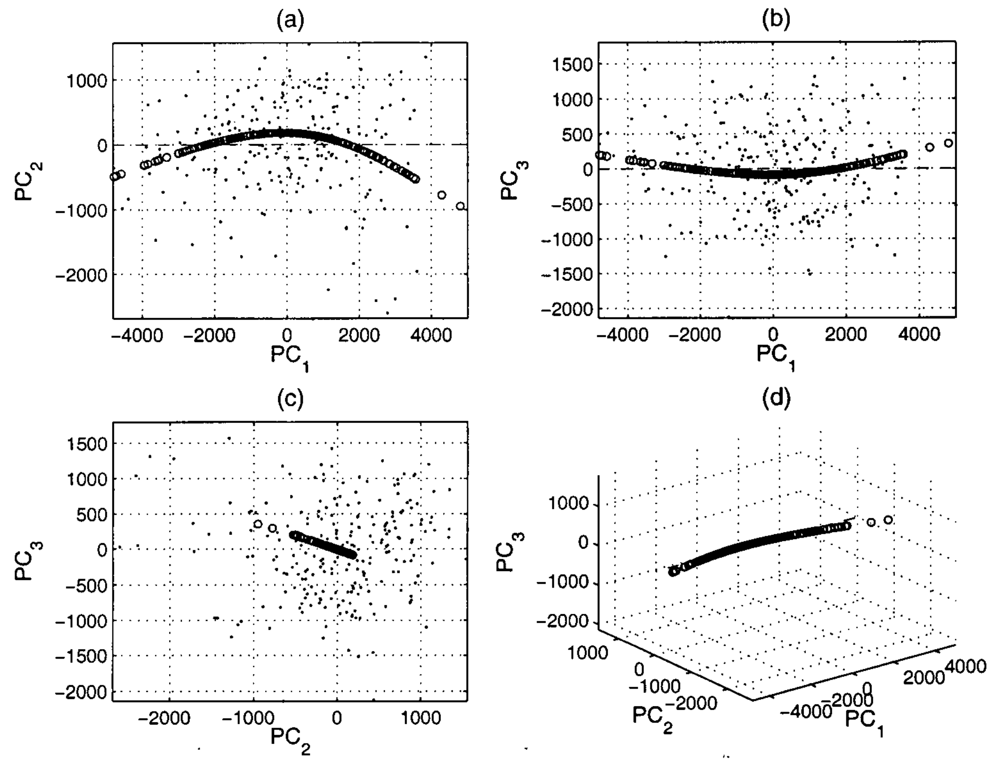

Figure 3.14: CNLPCA results of SI2000/B424 model at 30mb, three hidden neurons, penalty 1.1. The broken lines represent the linear PCA solutions. There is slightly more non-linearity than at 300mb, but there is still no cyclic behaviour.

Chapter 4

Daily-averaged Model Data

The SI2000/B424 model is used to obtain daily-averaged geopotential height data at 500mb and 50mb. An ensemble of five runs is produced. Each model integration is started from slightly different initial conditions. Initially, each run is analyzed separately. Similar results are obtained for each run at the same pressure level, therefore the Trials A for each level are used in illustration.

The data processing and parameters used in individual data sets are the same. For each grid point, height value is first multiplied by the square-root of the cosine of its latitude in order to compensate for the increase in spatial density of the points with increasing latitude. Then, the climatological annual cycle is calculated by averaging similar days of successive years. Anomalies are created by subtracting the climatological annual cycle from each grid point. A ten-day low-pass filter is applied to the anomalies to remove high-frequency noise, such as the synoptic-scale storms. Finally, three-month winters (December, January and February) are extracted out of each year.

The principal components are found for the processed data using the Principal Component Analysis (PCA). The first eight PCs become the input to the CNLPCA neural network. Three hidden neurons are used in each of the hidden layers to allow sufficient non-linearity. A variety of penalty factors are used before the one that yields the lowest mean square error is determined the appropriate result.

4.1 500mb Geopotential Height

Figure 4.1 shows the first two principal components at 500mb. The first PC is similar to the Arctic Oscillation, where there is a maximum near the polar area. The pattern of the anomaly shows a different sign in the polar area and the area around it, with respect to the Pacific Ocean. The second PC is similar to the North Atlantic Oscillation, characterized by the strong maximum in the Atlantic Ocean. The positive anomaly extends around that latitude, enclosing a minimum centered in the northern part of Canada.

Figure 4.2 shows the time series of the first three principal components. The first eight principal components, accounting for 68.8% of the total variance, form the input to the CNLPCA neural network, and the result can be seen in figure 4.3. The dots that represent the input data follow a path. This is because a ten-day low-pass filter is applied in order to remove high-frequency noise, hence the more obvious trajectory. The loop solution suggests cyclical behaviour. It is interesting to investigate the time spent in

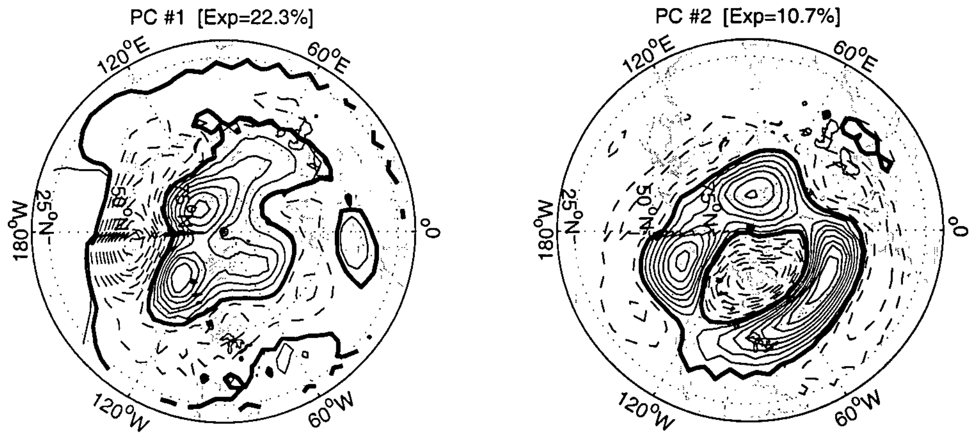

Figure 4.1: First two PCs for the B424 Trial A at 500mb between 20°N and 90°N. "Exp" is the percentage of variance accounted for by the corresponding PC.

each area of the loop. A histogram of the angular distribution can be seen in figure 4.4. The angle is calculated by:

$$\theta = \tan^{-1} \frac{PC2}{PC1} \tag{4.1}$$

The system spends the most time in the positive PC1 direction. The negative PC1 direction is the second most frequent. The system also spends a large proportion of time at the $-85°$ and the $110°$ directions. These are close to the positive and negative PC2 directions, but there is a small PC1 component mixed in them.

The spatial patterns at each point within the central bin, as well as the two bins on each side, are averaged to obtain spatial patterns for each regime. Figure 4.5 shows the spatial patterns of the high frequency regimes mentioned

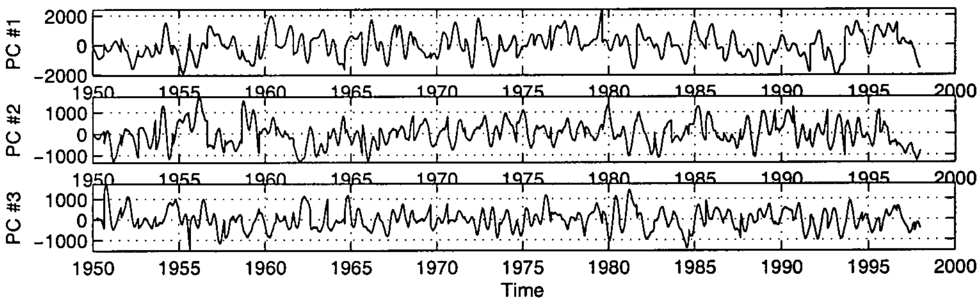

Figure 4.2: The scores of the first three PCs for the Trial A at 500mb.

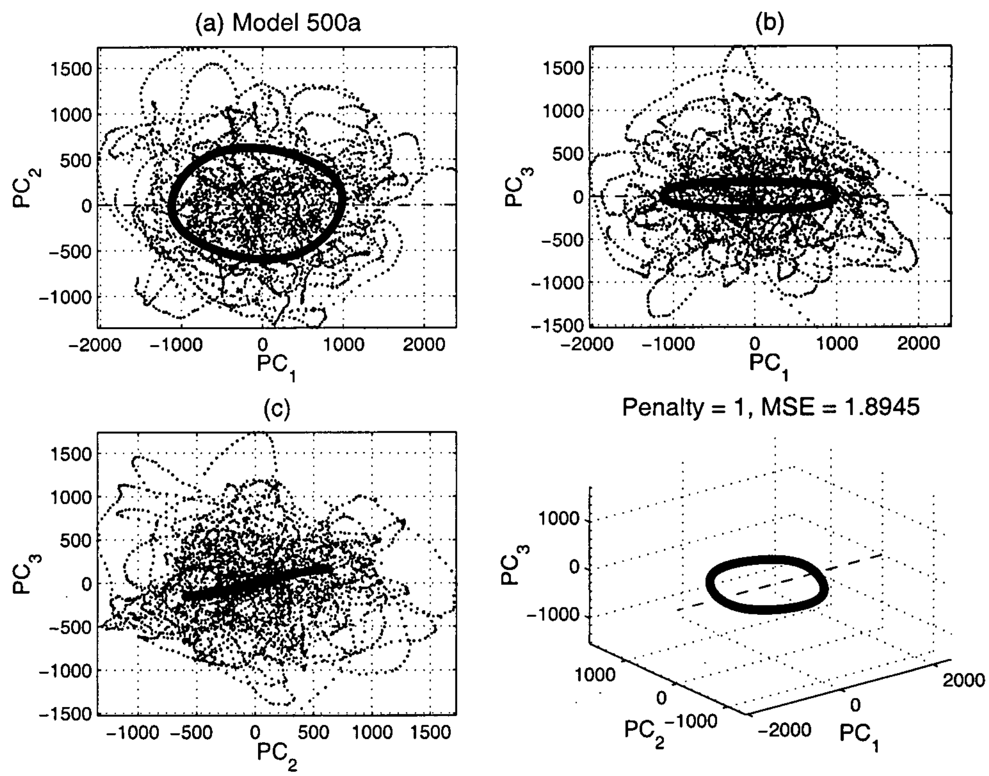

Figure 4.3: CNLPCA results for trial A at 500mb in PC space. Three hidden neurons are used. The dots are the input data into the CNLPCA neural network. The dashed line represents the linear PCA results.

above. The two patterns at $2°$ and $-172°$ are similar to the PC1 and -PC1 patterns (the AO) respectively, as expected. Also, the two patterns at $110°$ and $-85°$ are similar to the PC2 and -PC2 patterns (the NAO) respectively.

Another interesting thing to investigate is the extreme cases, as seen in figure 4.6. Because of the tilt of the loop, the maxima and minima PC1 and PC2 values of the CNLPC do not occur on the axes, but rather at an angle. Notice that the maximum and minimum for PC1 occur close to the PC1-axis where the frequency is at the highest, but the maximum and the minimum for PC2 occur at a direction off the PC2-axis where the frequency is near the lowest. The PCs do not capture the directional variabilities. The greater the tilt, the more different the spatial patterns look.

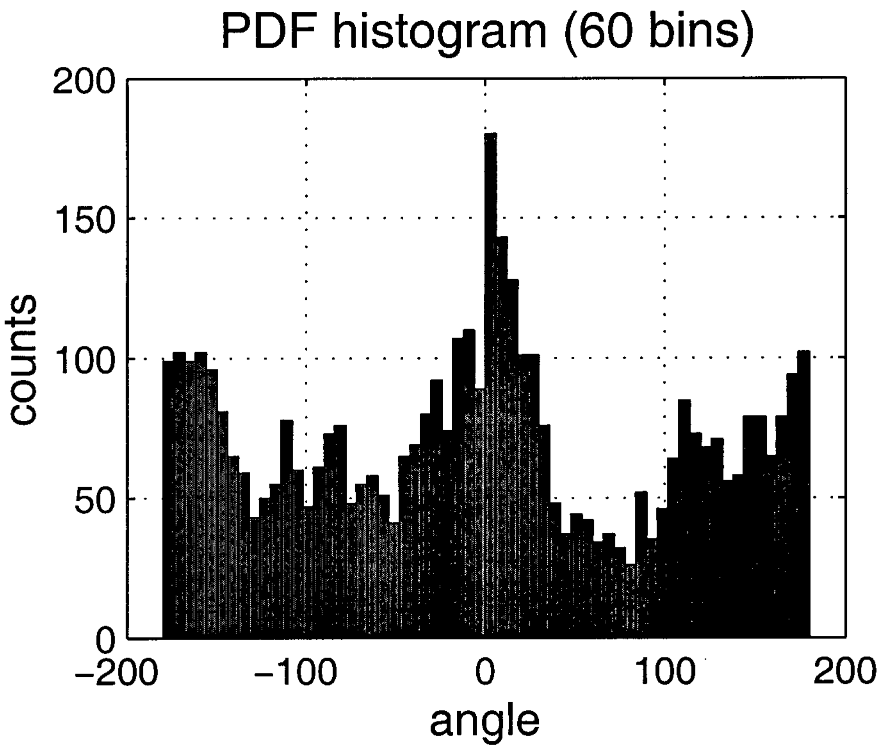

Figure 4.4: Angular distribution of the results at 500mb. The horizontal axis is the angle obtained from taking $\arctan(\frac{PC2}{PC1})$ for each of the 4320 points on the CNLPCA output. Each bin is 6° in size, for 60 bins in total.

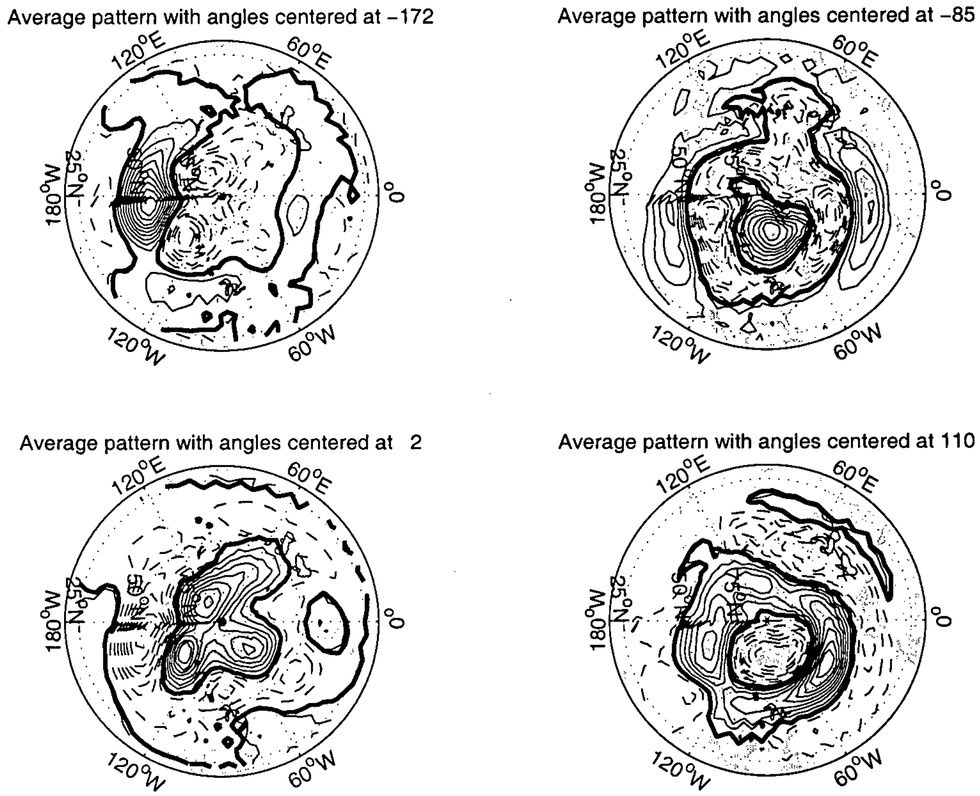

Figure 4.5: The average spatial patterns for each regime at 500mb, obtained from original processed data. The angles are determined by the peaks in the angular distribution, figure 4.4.

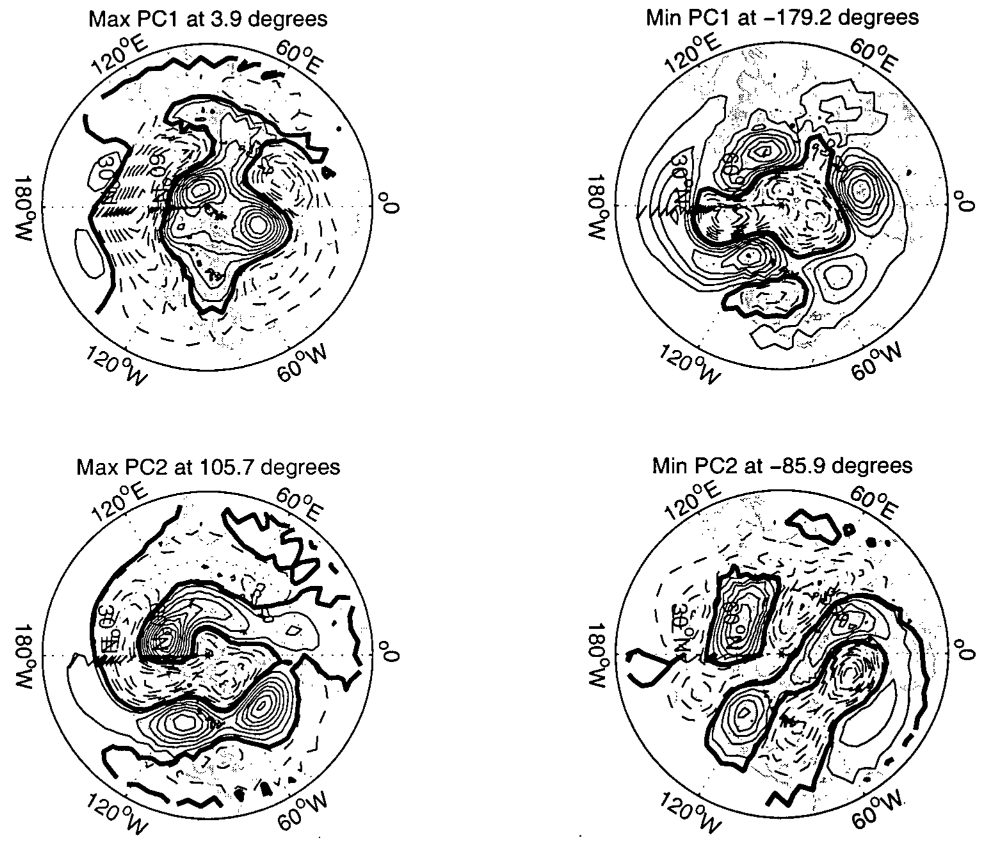

Figure 4.6: The spatial patterns at the points on figure 4.3 with (top left) Maximum PC1, (top right) Minimum PC1, (lower left) Maximum PC2, and (lower right) Minimum PC2.

4.2 50mb Geopotential Height

Figure 4.7 shows the first two principal components of Trial A at the 50mb geopotential height. The structure is simpler than that at the 500mb geopotential height. The first PC is an oscillation similar to the tropospheric Arctic Oscillation. This is a symptom of an oversimplified representation of the stratosphere. The second PC is the dipole oscillation. Figure 4.8 shows the score of the first three PCs.

The CNLPCA output of the scores (figure 4.9) reveals a loop that lies on the PC1-PC2 plane. This is similar to the tropospheric case, which supports the idea of an oversimplified stratosphere [16]. The penalty factor used to obtain an optimal solution is only 0.1, compared to 1.0 at the 500mb level. This shows that there is less non-linearity involved in the stratosphere of the model.

The angular distribution, shown in figure 4.10, has a similar structure as that in the 500mb geopotential height. This shows that even though the spatial patterns are oversimplified, the model successfully captured the evolution of the score.

The spatial patterns of the regimes, figure (4.11), show that the stratosphere changes between oscillating as a monopole (top left and bottom left) and a dipole (top right and bottom right). The patterns at the extrema (figure 4.12), however, show more wavelike structures than the regimes.

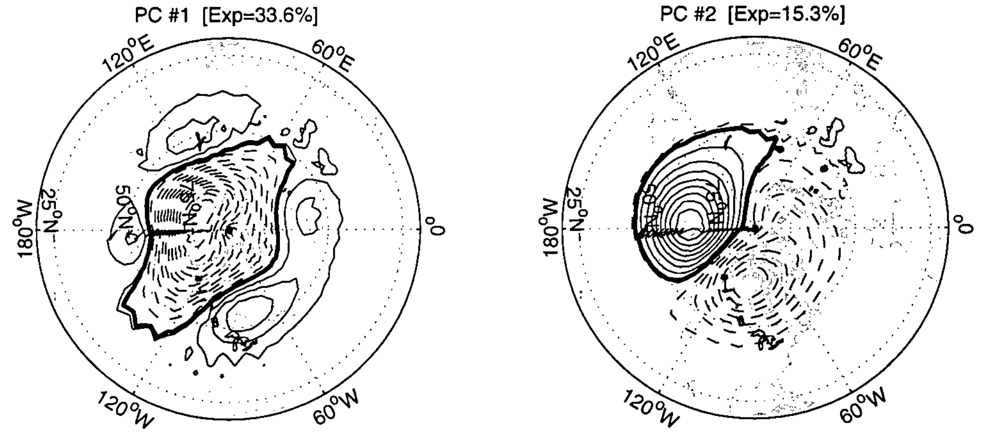

Figure 4.7: First two PCs for the Trial A at 50mb between 20°N and 90°N. "Exp" is the percentage of variance.

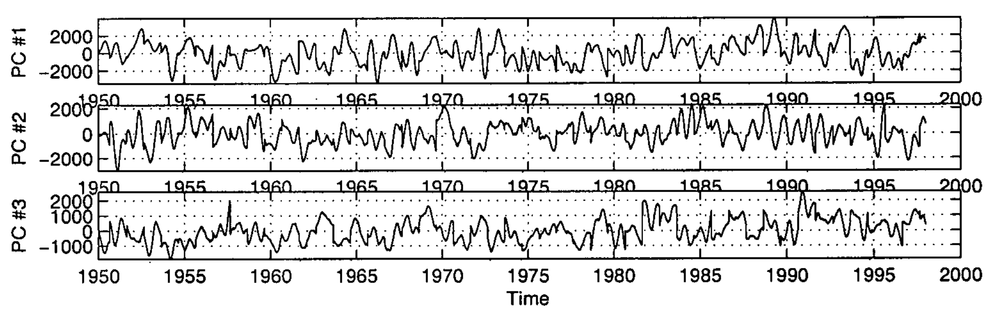

Figure 4.8: The scores of the first three PCs for the Trial A at 50mb.

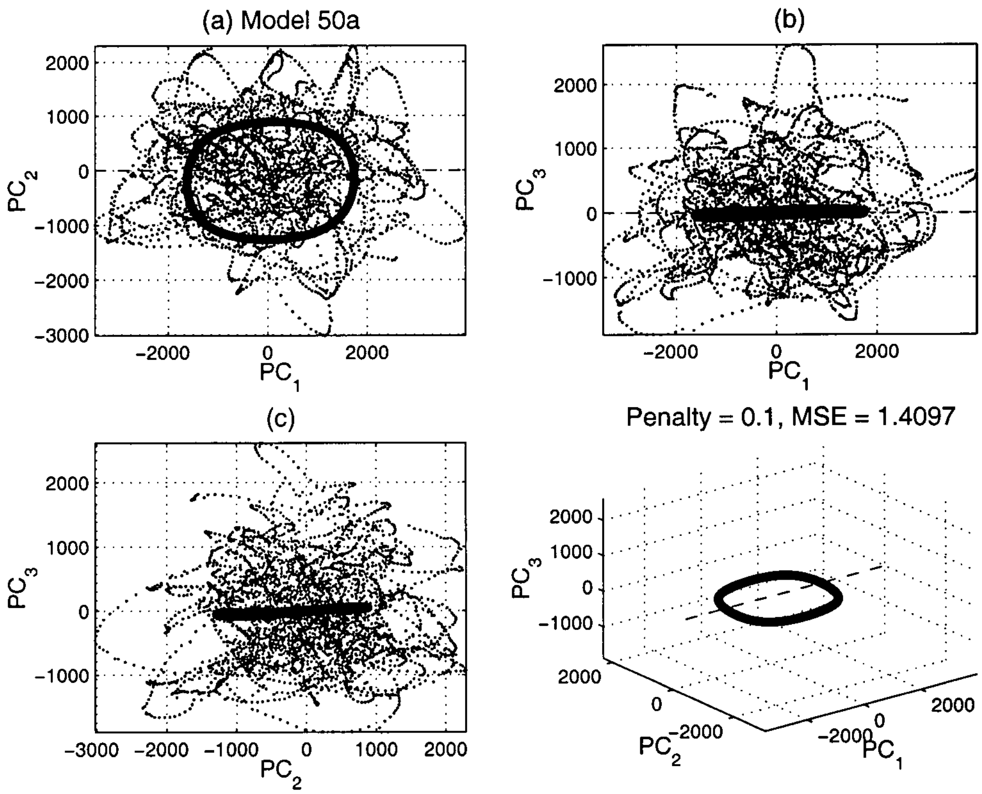

Figure 4.9: CNLPCA results for Trial A at 50mb in PC space are the overlapping circles. The dots are the input data into the CNLPCA neural network. The dashed line represents the linear PCA results.

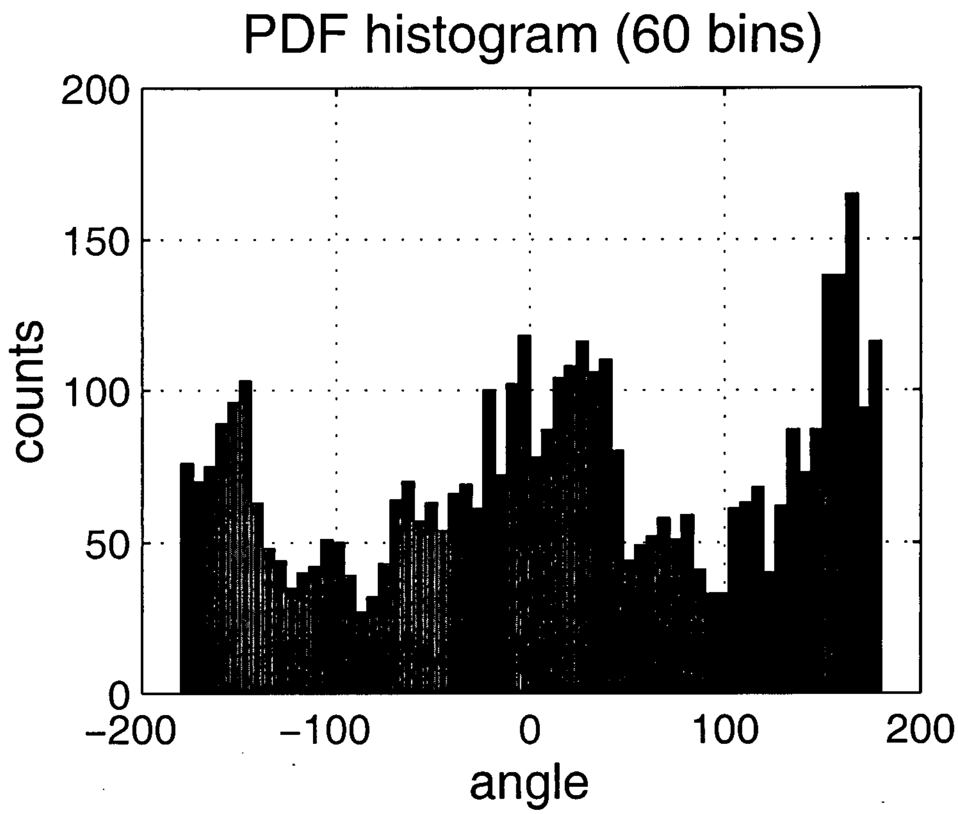

Figure 4.10: Angular distribution of the results at 50mb. The horizontal axis is the angle obtained from taking $\arctan(\frac{PC2}{PC1})$ for each point on the CNLPCA output. Each bin is 6° in size, for 60 bins in total.

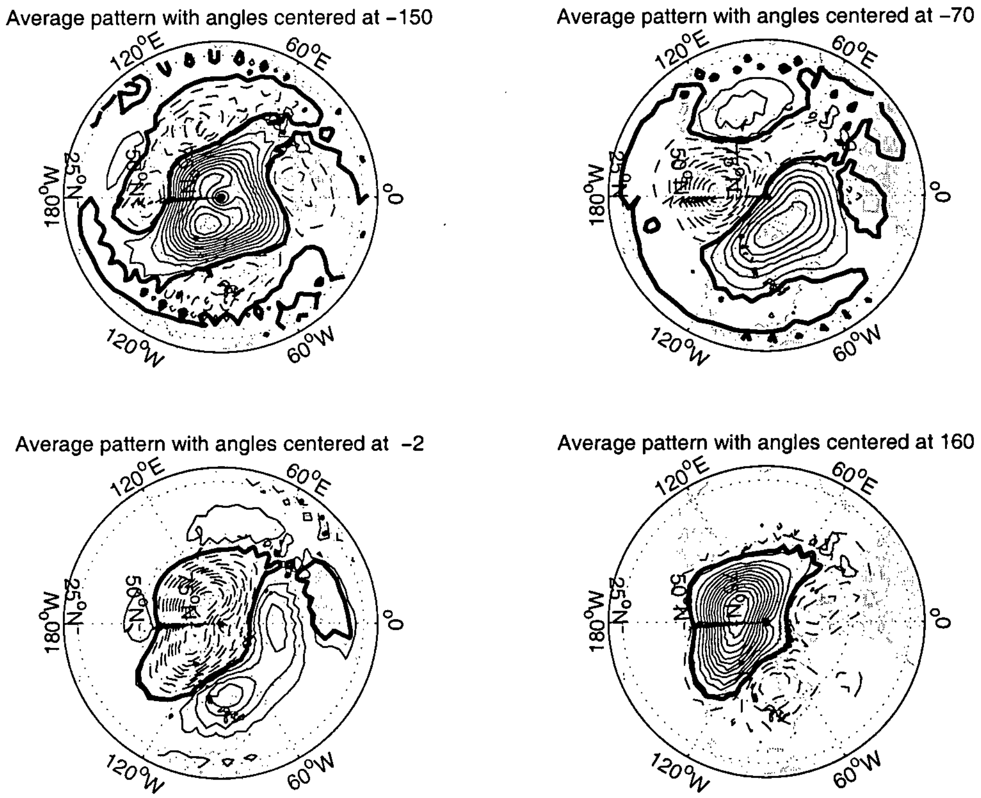

Figure 4.11: The average spatial patterns around each regime at 50mb. The angles are determined by the well-separated peaks in the angular distribution, figure 4.10.

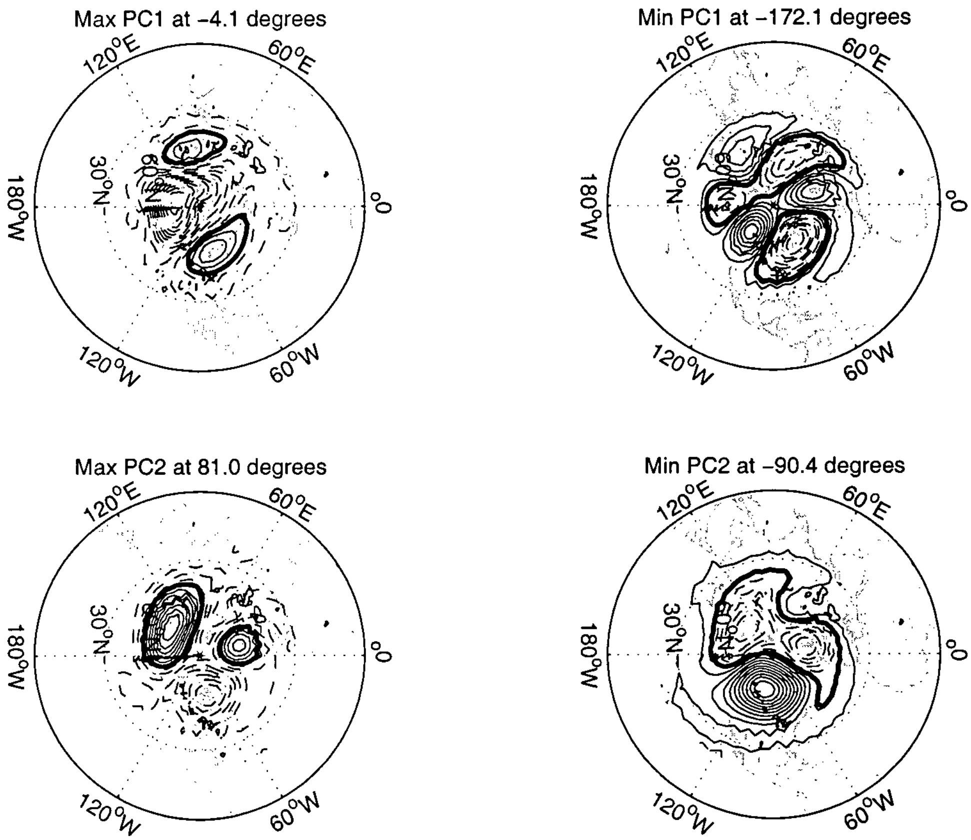

Figure 4.12: The spatial patterns at the points on figure 4.9 with (top left) Maximum PC1, (top right) Minimum PC1, (lower left) Maximum PC2, and (lower right) Minimum PC2.

4.3 Ensemble Averages

When taking an ensemble average to average out the chaotic dynamics and amplify the forced signal, some non-linearity is lost. Therefore, some parameters need to be adjusted for this change. When analyzing the five-member ensemble average, two hidden neurons are used in each of the encoding and decoding layers, and the first ten PCs are used. No low-pass filter is used because taking an average acts like a filter, but the data are taken in seven-day segments to remove noise with synoptic timescales.

Figure 4.13 shows the first two PCs of the five-member ensemble average at 500mb geopotential height. The first PC has the same features as the first PC in Trial A, seen in figure 4.1. The second PC is the PNA teleconnection pattern. The scores of the first three PCs are shown in figure 4.14.

Figure 4.15 shows an open curve CNLPCA approximation as a result of averaging. The U-shaped curve suggests that although there is no cyclic behaviour, there is still non-linearity. However, an inspection of the point distribution along the curve and its cumulative distribution function, shown in figure 4.16, tells us that the system spends most of the time between 0.4 and 0.8 of the curve, normalized to unit length. This translates to the area close to the PC1 axis. Thus, the system is mostly linear after taking an ensemble average.

Figure 4.17 shows the first two PCs of the five-member ensemble average at 50mb geopotential height. The first PC is very similar to the first PC in

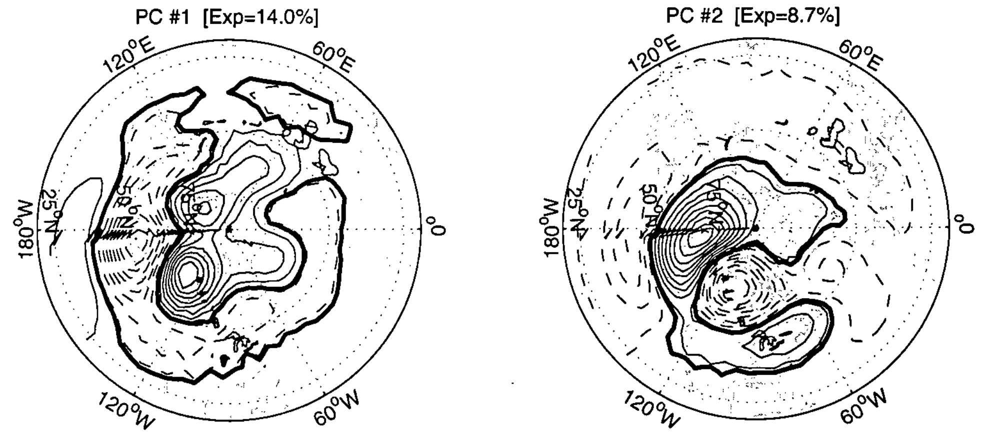

Figure 4.13: First two PCs of the ensemble average at 500mb.

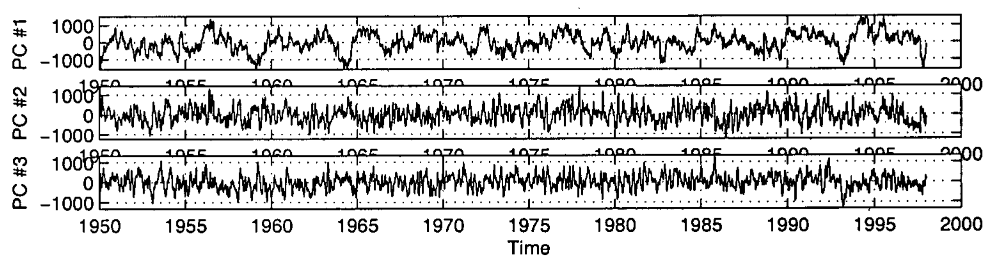

Figure 4.14: Scores of the first three PCs of the ensemble average at 500mb.

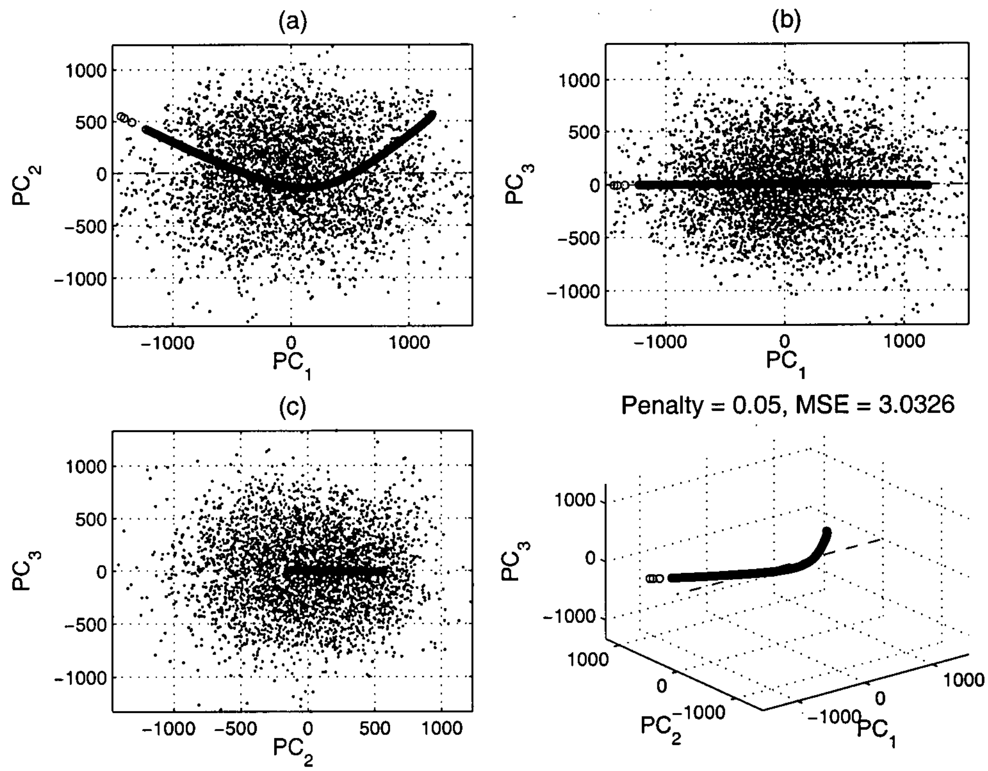

Figure 4.15: CNLPCA results for ensemble average at 500mb in PC space. The dots are the input data into the CNLPCA neural network. The dashed line represents the linear PCA results. An open curve solution is obtained even with a small penalty of 0.05.

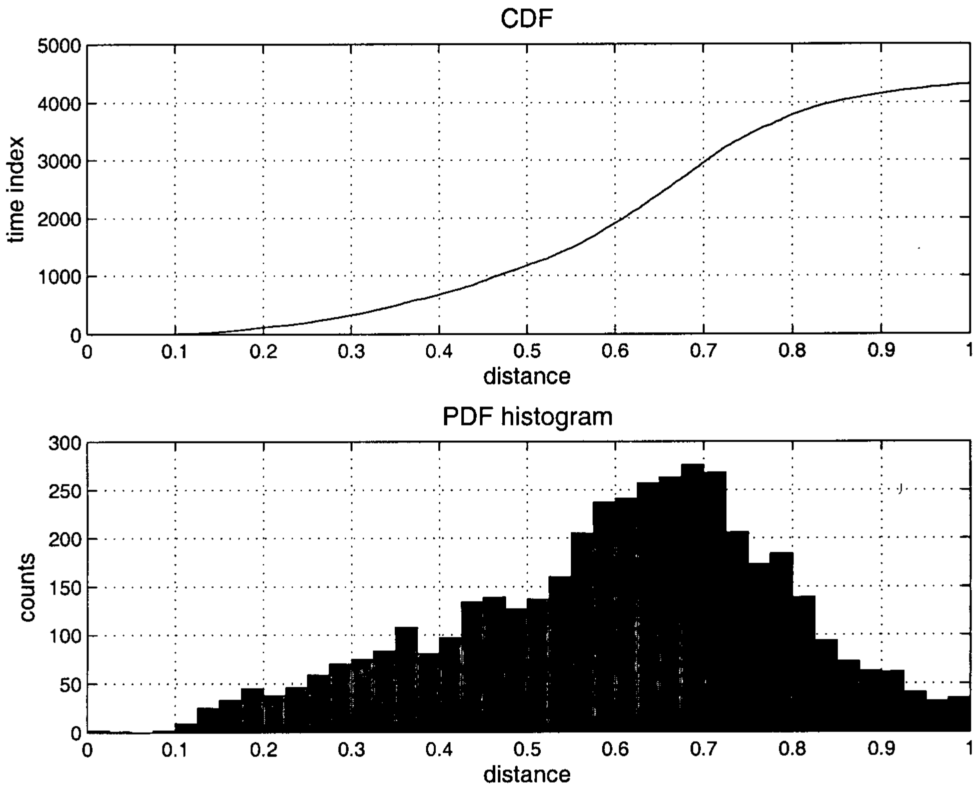

Figure 4.16: Top: cumulative distribution function (CDF) of the results of the ensemble average at 500mb. Bottom: distribution of points along the curve, normalized to unit length. The curve starts at 0 on the left side, and ends at 1 on the right side. Each bin is 9° in size, for 40 bins in total.

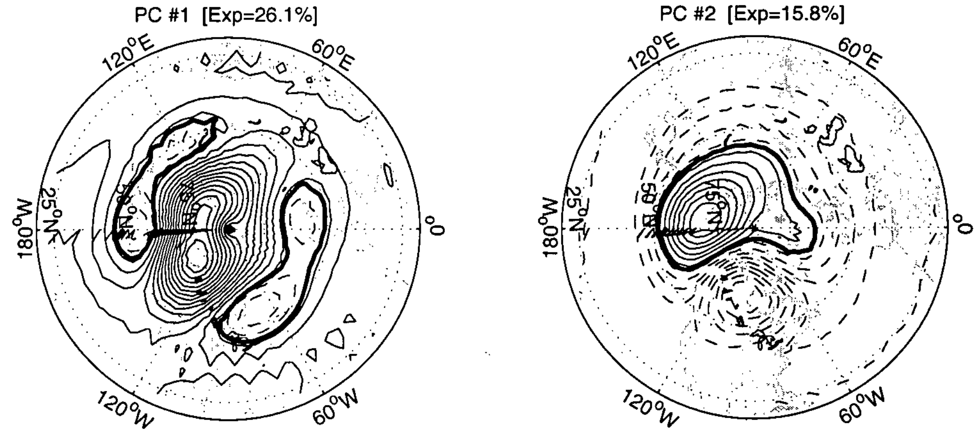

Figure 4.17: First two PCs of the ensemble average at 50mb.

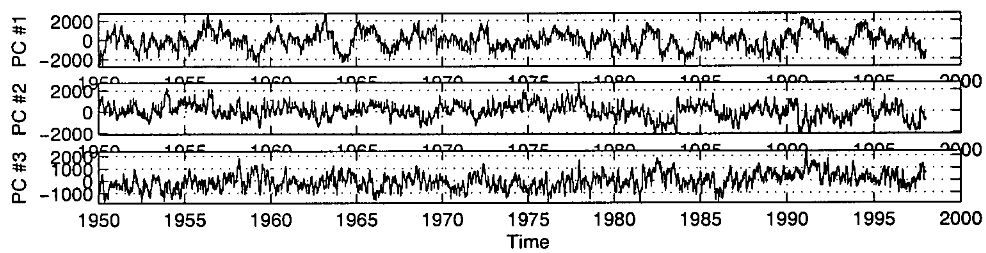

Figure 4.18: Scores of the first three PCs of the ensemble average at 50mb.

figure 4.7. Also, the positive regions over the Pacific and Asia are connected in the ensemble. The second PCs shows a dipole oscillation similar to the second PC in figure 4.7. In the ensemble, the positive region extends accross the North Pole to cover all of the Arctic Ocean.

Figure 4.19 shows a loop solution. A penalty factor of only 0.05 is needed because the ensemble averaging has reduced the level of non-linearity. A larger factor would yield non-cyclical results similar to the previous case at

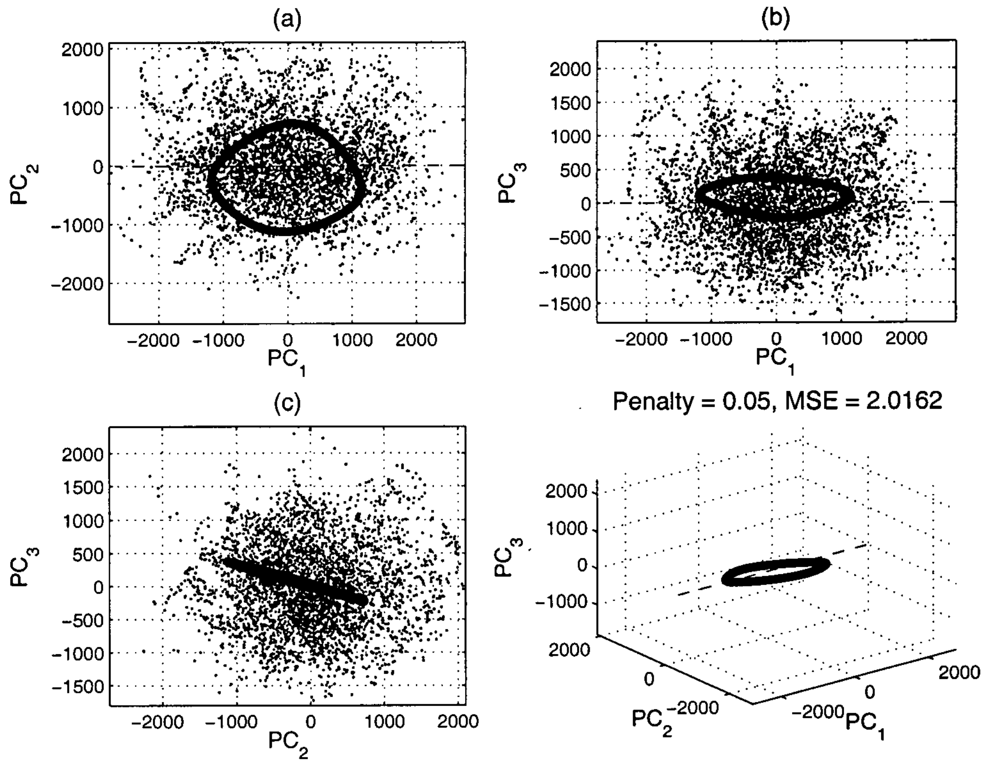

Figure 4.19: CNLPCA results for ensemble average at 50mb in PC space. The dots are the input data into the CNLPCA neural network. The dashed line represents the linear PCA results.

500mb.

The angular distribution of the loop in figure 4.19 is shown in figure 4.20. There are three distinct peaks, one at around $-175°$, one at around $-20°$ and one at around $80°$. The system seldom stays in the region between $-50°$ and $-150°$.

Figure 4.21 shows similar spatial patterns to the first two PCs. However, there are some differences. For example, at the $-20°$ direction, the negative is very weak. In fact, most of that pattern is positive. Also, in the $-75°$ and $80°$ directions, the Greenland region remain in the negative, instead of oscillating between positive and negative in a linear fashion. These non-linear oscillations are what the PCA cannot capture.

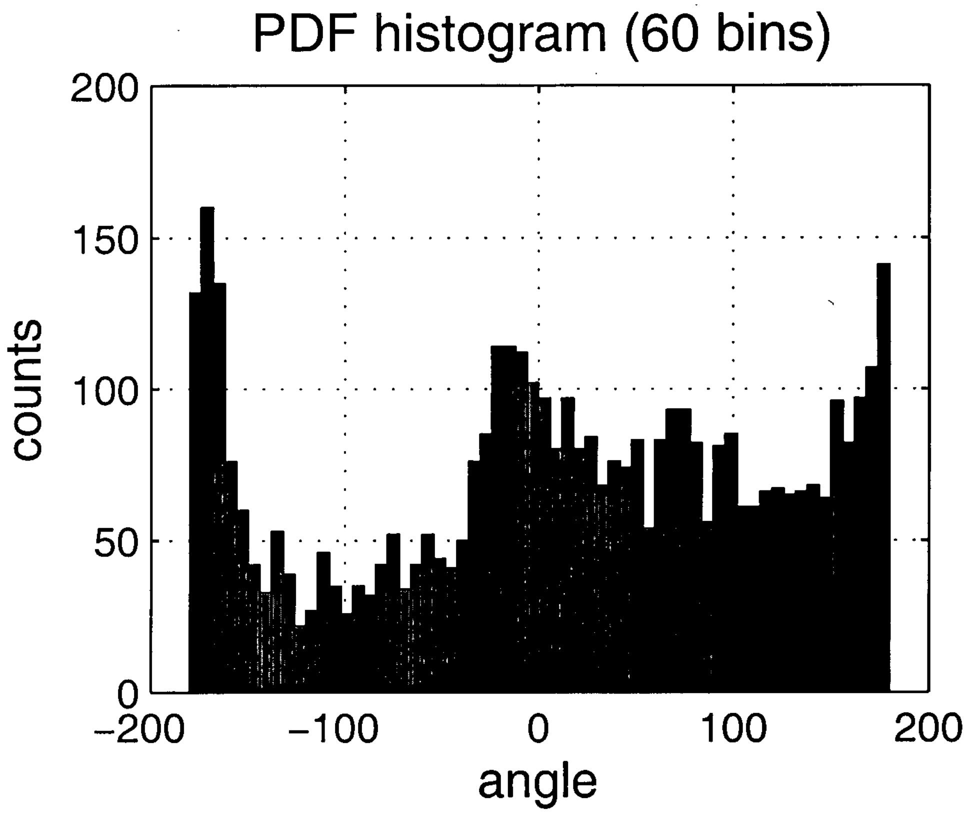

Figure 4.20: Angular distribution of the results of the ensemble average at 50mb. Each bin is 6° in size, for 60 bins in total.

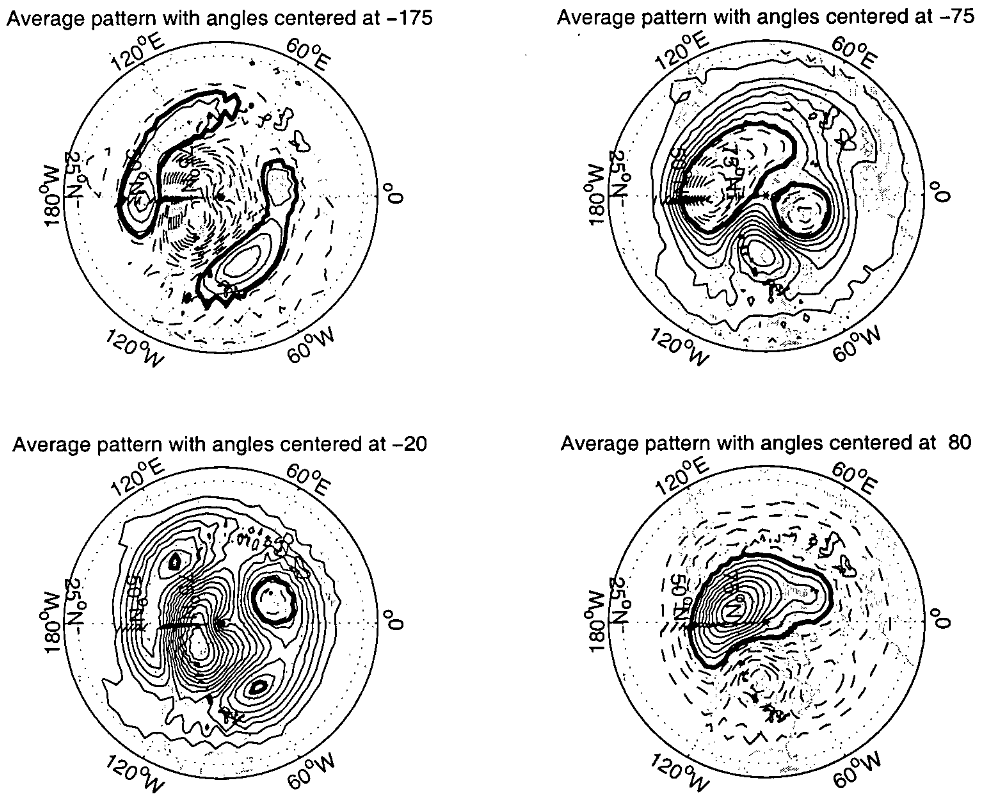

Figure 4.21: The average spatial patterns around each regime of the ensemble average at 50mb. The angles $-175°$, $20°$ and $80°$ are determined by the peaks in the angular distribution, figure 4.20. The trough of the distribution at $-75°$ is also shown.

Chapter 5

Observed Data

When loop solutions are obtained, it is interesting to know which direction, clockwise or counterclockwise in two-dimensional projection, the system moves. If a trend is found, it may provide useful insight on our climate system. In this chapter, the National Centers for Environmental Prediction-National Center for Atmospheric Reserach (NCEP-NCAR) reanalysis [17] for the 500mb geopotential height will be used for illustration.

The data, which have a $2.5° \times 2.5°$ horizontal resolution, are pre-processed the same way as the monthly-averaged and daily-averaged model data. A 10-day low-pass filter is applied. The first ten principal components are used as the input to a CNLPCA neural network (figure 2.2) with two hidden neurons ($m = 2$).

5.1 Regime Behaviour

Figure 5.1 shows the first two PCs of the observed data. PC1 has a positive peak in the Pacific coast of North America and a strong negative peak in the

Pacific Ocean. This pattern resembles the positive phase of the Pacific/North America oscillation (PNA) pattern. PC2 has a strong negative peak over Greenland, and the negative anomaly spreads into the Pacific Ocean. There is also a strong positive peak over the Atlantic Ocean. This meridional dipole in the North Atlantic is identified as the NAO.

Figure 5.2 shows the score of the first three PCs, which are plotted in dashed lines in figure 5.3. A penalty factor of 0.1 is found to give a mean square error smaller than those given by nearby penalty factors. A loop solution is obtained in the PC1-PC2 plane. The probability density function around the loop is plotted in figure 5.4 and several well-separated peaks can be seen. The spatial patterns of these frequently visited regimes are shown in figure 5.5.

The most frequently visited regime is located near the negative PC1 axis. Its spatial pattern, top left in figure 5.5 is the negative phase of the PNA. This is consistent with the regime A identified in Monahan et al. [10] using NLPCA. Cheng and Wallace [18], as well as Corti et al. [11], also found this pattern using cluster analysis (cluster B in figure 3 of [11]).

The spatial pattern of a weak peak near the negative PC2 axis is shown in the top right of figure 5.5. It is associated with the negative phase of the NAO, and it is named the "G" pattern after the peak over Greenland. This is consistent with the NLPCA results, as well as that of the cluster analysis (negative of cluster A). This peak is chosen over the apparent peak at $-126°$ because the bin next to the latter is one of the shortest.

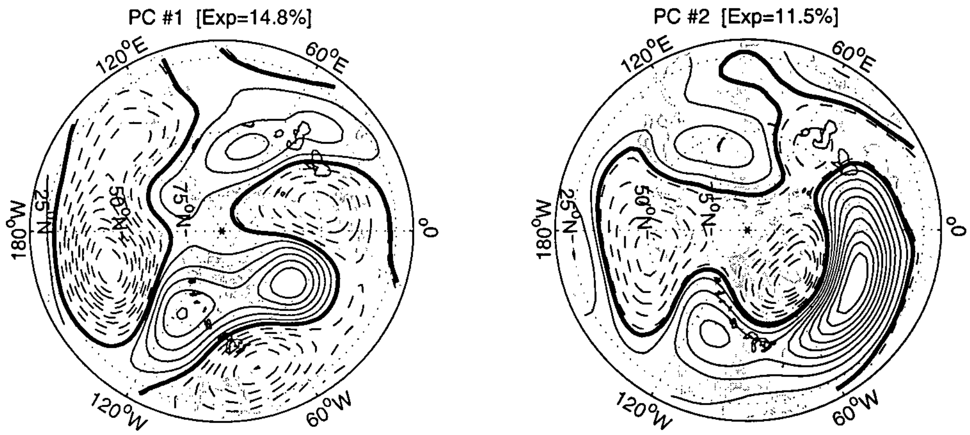

Figure 5.1: First two PCs for observed data at 500mb between 20°N and 90°N.

The next peak occurs at −32°. Its spatial pattern (bottom left of figure 5.5 is positive in the high lattitudes and mostly negative in lower lattitudes. This resembles the Arctic Oscillation (AO), which is not captured by the three-regime NLPCA, but is consistent with cluster analysis (cluster D).

The last peak is located at 50°. Its spatial pattern (bottom right of figure 5.5 contains three positive peaks and three negative peaks, positioned in a wavelike fashion. Comparing with the top left pattern, this pattern resembles the PNA pattern. It is captured in the NLPCA (regime "R"), as well as the cluster analysis (negative of cluster C).

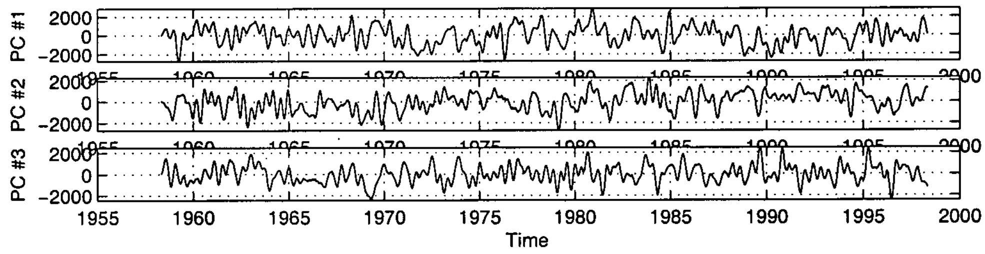

Figure 5.2: The scores of the first three PCs for observed data at 500mb.

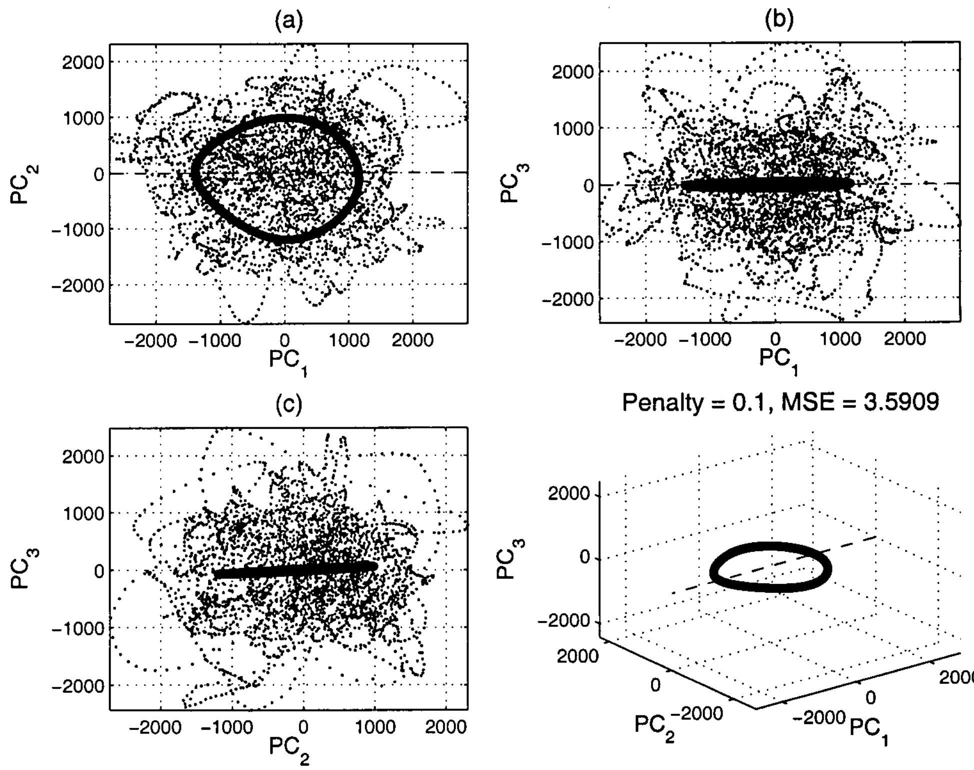

Figure 5.3: CNLPCA results for observed data at 500mb in PC space are the overlapping circles. The dots are the input data into the CNLPCA neural network. The dashed lines represent the linear PCA results.

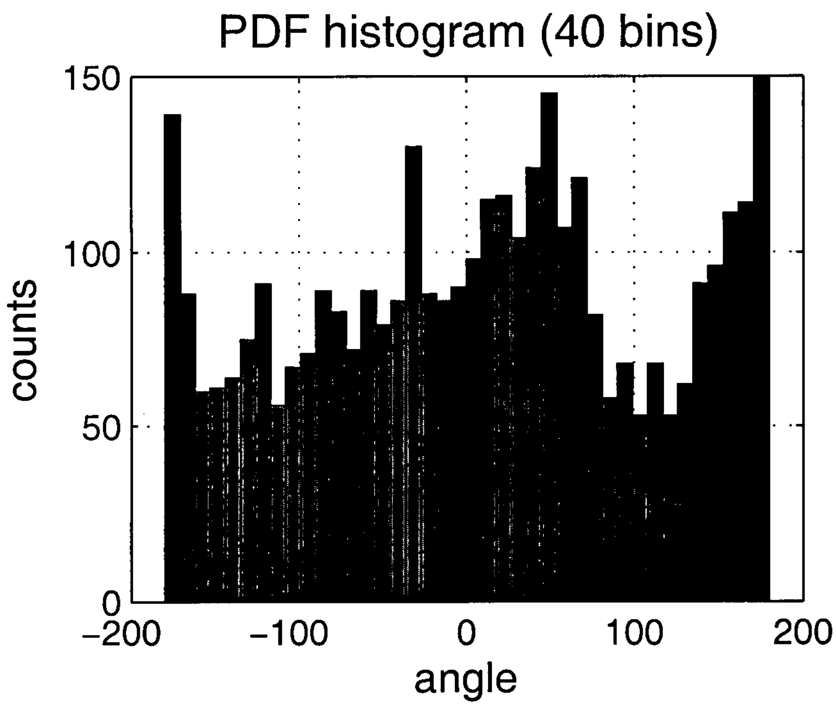

Figure 5.4: Angular distribution of the results at 500mb. The horizontal axis is the angle obtained from taking $\arctan(\frac{PC2}{PC1})$ for each of the 3600 points of the CNLPCA output. Each bin is 9° in size, for 40 bins in total.

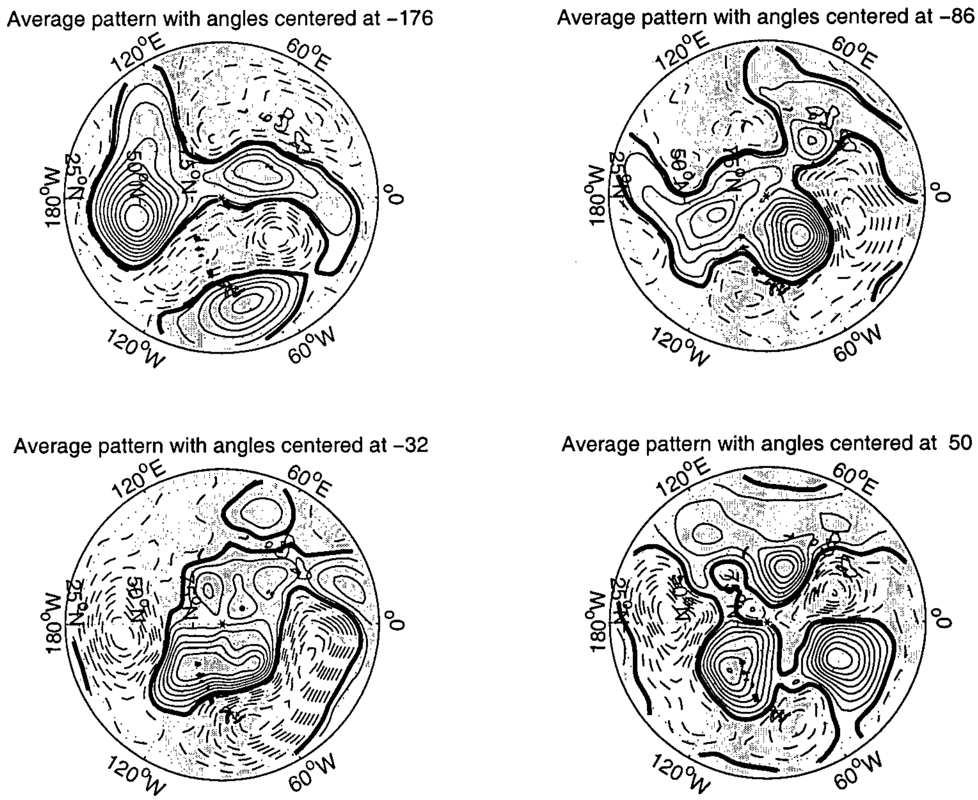

Figure 5.5: The regime behaviour is determined from the most frequently visited states, seen in figure 5.4.

5.2 Time Series of Phase Angle

When the arctangent of each point on the CNLPCA solution is taken, the PC1 and PC2 scores can be combined into one single time series, shown in figure 5.6. This graph shows the evolution of the system. The long-term behaviour of the system was to move counterclockwise before the mid-1970s, and then clockwise between 1976 and 1983. In 1983-84, it suddenly turned around and moved counterclockwise rapidly, and this happened again in 1987-88. The system turned around again in the mid-1990s, before moving counterclockwise rapidly once again in 1998.

Some of these changes in directions coincide with important happenings in our climate system. In the mid-1970s, there was a sudden climate change, resulting in the less frequent visits of regime "G" [10]. A lot of studies involving observed data tend to separate datasets into sub-periods before and after the mid-1970s because of this climate change [19]. The three rapid counterclockwise motions coincide with three of the most intense El Niño Southern Oscillation (ENSO) events ever recorded. On the other hand, La Niña events tend to make the system move in a clockwise direction, although the intense event in 1989 only move the system clockwise by one revolution, compared to the many revolutions during the El Niño events. There was an intense La Niña event in 1999, but the available observed data end just before that. It is expected that the time series will turn around and move in the clockwise direction.

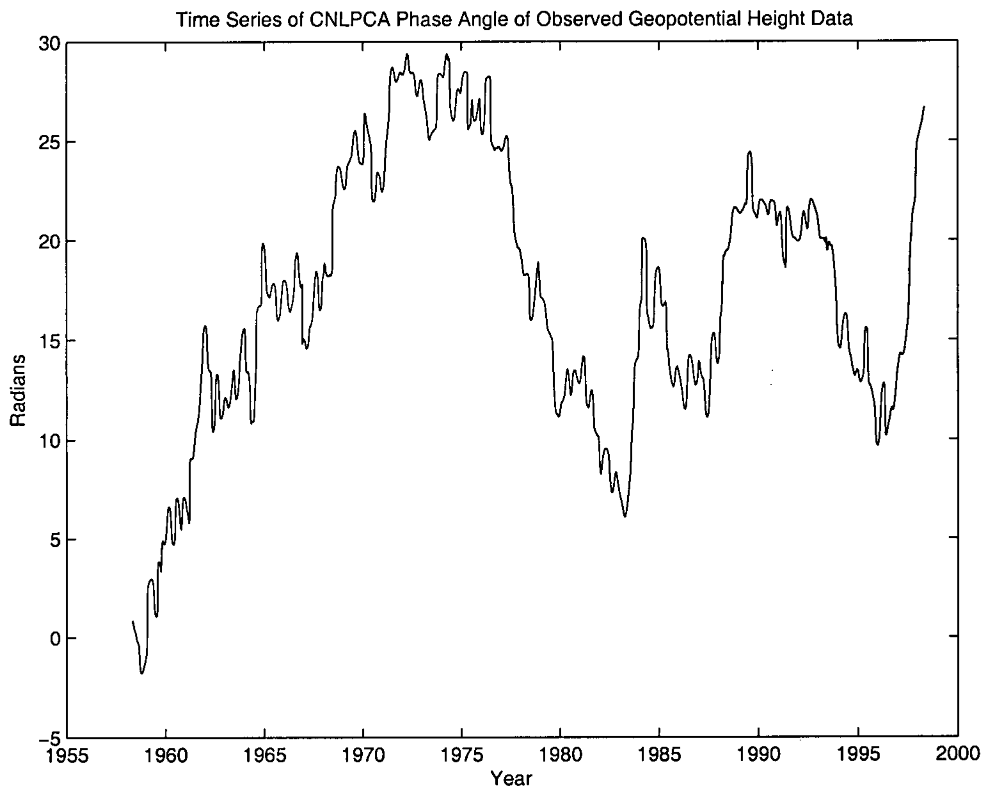

Figure 5.6: Time series of the phase angle of observed geopotential height data. The time series of the phase angle contains several abrupt changes. It changes directions in mid-1970s. There are also three rapid counterclockwise rotations in 1983-84, 1987-88 and 1996-98.

Chapter 6

Conclusions

The Circular Non-linear Principal Component Analysis is a versatile non-linear data analysis tool. It can be used to determine the level of non-linearity in a given set of data and extract non-linear features. However, the most important value of the CNLPCA is its ability to extract features that contain cyclical behaviour that involve three or more regimes.

Further investigations include:

- full-year analysis on the effect of discontinuous data between winters,
- recent data (1999-current) to investigate the La Niña event in 1999,
- usage of the time series of phase angles, for example, in forecasting,
- consistency after subdividing a dataset into decadal timescales, and
- comparisons with the results obtained from other data analysis methods.

The CNLPCA does contain shortcomings. For example, the results are subjective in the sense that the mean square error tolerance level (that constitutes overfitting) depends on the researcher. However, this is a universal problem in the field of non-linear data analysis. Also, the cost function minimization is very time-consuming when compared to the linear PCA. But with increasing computing power and the need to understand non-linear features, computing time should be reduced greatly in the future.

Bibliography

[1] M.A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37:233–243, 1991.

[2] A. Monahan. Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. *Journal of Climate*, 13:821–835, 2000.

[3] M. Kirby and R. Miranda. Circular nodes in neural networks. *Neural Computation*, 8:391–402, 1996.

[4] J.M. Wallace and D.S. Gutzler. Teleconnections in the geopotential height field during the Northern hemisphere Winter. *Mon. Wea. Rev.*, 109:784–812, 1981.

[5] J.W. Hurrell and H. van Loon Decadal Variations associated with the North Atlantic Oscillation. *Climatic Change*, 36:301-326, 1997.

[6] D.W. Thompson and J.M. Wallace. The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, 25:1297-1300, 1998.

[7] J.M. Wallace. North Atlantic Oscillation/Annular Mode: Two paradigms – One phenomoenon. *Quart. J. Roy. Meteor. Soc.*, 126:791-806, 2000.

[8] M. Kimoto and M. Ghil. Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.*, 50:2625–2643, 1993.

[9] P. Smyth, K. Ide, M. Ghil. Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, 56:3704–3723, 1999.

[10] A. Monahan, L. Pandolfo and J. Fyfe. The preferred structure of variability of the Northern Hemisphere atmospheric circulation. *Geophysical Research Letters*, 28(6):1019–1022, March 2001.

[11] S. Corti, F. Molteni and T.N. Palmer. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, 398:799–802, 1999.

[12] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314, 1989

[13] W.W. Hsieh. Nonlinear principal component analysis by neural networks. *Neural Networks*, 13:1095–1105, 2000..

[14] N. Rayner. HadISST1 Sea Ice and Sea Surface Temperature Data Files. *Geophysical Research Letters*, Hadley Center, Bracknell, U.K., 2000.

[15] Yuval, and W.W. Hsieh The impact of time-averaging on the detectability of nonlinear empirical relations. *Quart. J. Roy. Met. Soc.*, 128: 1609–1622., 2002.

[16] A. Monahan, J. Fyfe, L. Pandolfo. The Vertical Structure of Wintertime Climate Regimes of the Northern Hemisphere Extratropical Atmosphere. *Journal of Climate*, 16(12):2005–2021, 2003.

[17] E. Kalnay and Coauthors. The NCEP/NCAR 40-year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, 77:437-471.

[18] X. Cheng and J.M. Wallace. Cluster analysis of the Northern Hemisphere wintertime 500-hpa height field: Spatial patterns. *J. Atmos. Sci.*, 50:2674-2696, 1993.

[19] J.L. Knox, K. Higuchi, A. Shabbar, E. Neil. Secular Variation of Northern Hemisphere 50 kPa Geopotential Height. *Journal of Climate*, 1(5):500-511, 1988.