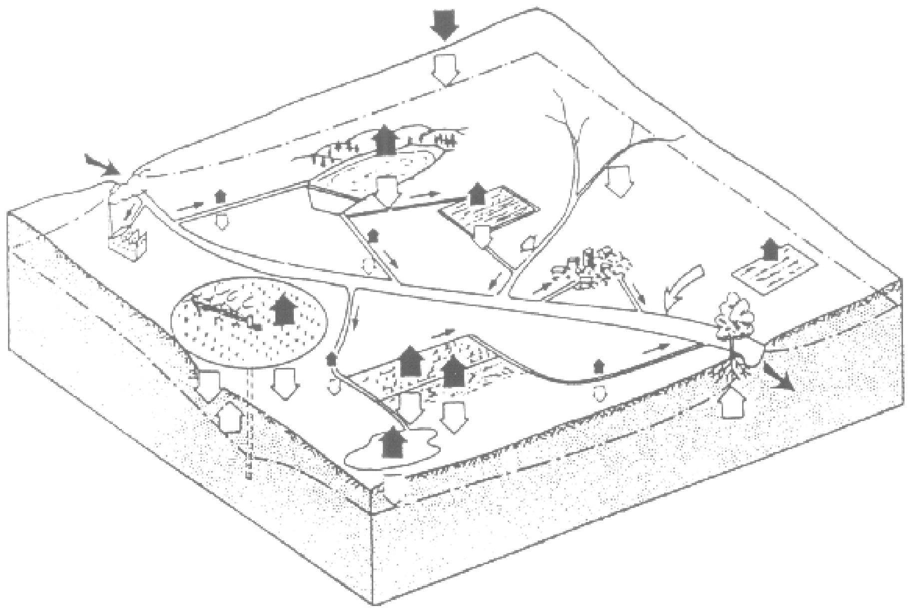
Proceedings of the Sixteenth Annual American Geophysical Union, April 1996
HYDROLOGY DAYS

Forecasting Water Availability by Applying Neural Networks with Global and Solar Indices



Eric Y. Zhang[1], P.E. and Paul Trimble[2]

Contents

- Abstract
- Introduction
- El Nino - Southern Oscillation Event
- Climate Fluctuations Related To Solar Sunspot Cycles
- History of ENSO, Solar Activity, Geomagnetic Activity and Lake Okeechobee Tributary Inflow
- A brief Introduction of Artificial Neural Networks
- Back Propagation
- Choosing a Network Configuration
- Input Data Preparation
- Network Training
- Results And Conclusion
- Future Studies
- References

Figures

- Figure 1. Schematics of a three layer back-propagation neural network
- Figure 2. Predicted versus actual Lake Okeechobee inflow
- Figure 3. Decadal shifts of average annual inflow

Abstract

The ability to forecast changes in water availability associated with climate fluctuations would be a valuable asset to regional water management authorities. These forecasts may provide advanced warnings of extended periods of deficits or surpluses of water availability allowing better regional water management for flood protection, water supply, and environmental enhancement.

In order to achieve this goal, it is necessary to have a global perspective of the oceanic and atmospheric phenomena which may affect regional water resources. However, the complexity involved may hinder traditional analytical approaches because such approaches are usually based upon simplified assumptions. This paper investigates the applicability of neural networks. A neural network is a computational method inspired by studies of the brain and nerve systems in biological organisms. Neural networks have the capability of self-learning and automatic abstracting. Applying this technique may reduce the time of modeling a complex system. Issues such as selecting a suitable back-propagation network configuration, and preprocessing input data are addressed.

The forecasting is focused on the inflows to Lake Okeechobee, the liquid heart of south Florida. Our preliminary results indicate that neural networks are promising tools. When the Southern Oscillation Index (SOI), the solar sunspot number and geomagnetic activity are included together as input for the neural network, the network was able to predict the largest and smallest inflow months of the testing period. Training and testing with the S0I index alone were not successful hinting at the importance of solar and/or geomagnetic activity in climate fluctuations.

Introduction

Lake Okeechobee is located in south central Florida. This body of water is the second largest freshwater lake lying wholly within the boundaries of the United States. It is frequently referred to as the "liquid heart" of south Florida as it is an important source of freshwater for many of the natural ecosystems of south Florida, the primary source of supplemental water supply for over five hundred thousand acres of intensely farmed agricultural land, and is a backup source of water supply for the densely populated urban areas of south Florida. However, south Florida's potential for periods of heavy rains and severe tropical storms and Lake Okeechobee' s large tributary basins require that water levels in the lake be carefully monitored to ensure that they do not rise to levels that would threaten the structural integrity of the levee system surrounding the lake.

The current operation schedule (Trimble and Marban, 1989) was developed considering only the most recent history of water levels in the Lake and the season of the year. The reliability of long term weather forecast and the relationship between global climate fluctuations and local Florida hydrology was seen as, at best, only fair. With an improved understanding of the global climate - south Florida hydrology link and the application of neural networks for hydrologic forecasting, the possibility may exist for a more dynamic operational schedule to be developed that reorder operational priorities of the water management during different climate regimes. Rasmusson and Arkin (1993) did a commendable job in making it clear that a global understanding of climate is needed to understand the reason and causes of local anomalies. They also summarized interdecadal fluctuation in rainfall in the Sahel and India that appear to have very similar trends as those in South Florida.

The purpose of this research is to: l) gain a better understanding of how climate fluctuations within the south and central Florida region may be related to global climate fluctuations or trends, 2) determine if interdecadal fluctuations in the local climate may be explained by global climate indices, and 3) to determine, if such a relationship exists, can it be applied for more effectively managing the water levels and outflows of Lake Okeechobee. A neural network is applied to test the predictability of Lake Okeechobee inflows from global climate indices.

El Nino - Southern Oscillation Event

The signature of an El Nino event is the occurrence of very warm ocean waters at low latitudes located off the west coast of South America. The Southern Oscillation Index (SOl) is the measure of sea level atmospheric pressure difference between Darwin Australia (western Pacific) and Tahiti (eastern Pacific).

There is a strong connection between the El Nino event and the Southern Oscillation Index. The El Nino-Southern Oscillation Event is often referred to by the acronym ENSO. An event of this type affects the climate of a large portion of the planet. The strongest and most reliable effects occur in the tropical Pacific Ocean. Other parts of the world, especially in the middle Latitudes are affected through teleconnections. A negative SOl index is most often associated with a warm sea surface temperature anomaly El Nino event while a positive SOl is synonymous with a cold sea surface anomaly La Nina event.

Teleconnections are represented as statistical associations among climatic variables separated by large distances. Many large rainfall and drought events that occur within the state of Florida are strongly correlated to ENSO events (Hanson and Maul, 1991). This type of relationship is important to investigate farther for both operational and planning concerns. It must also be determined if ENSO events and the global teleconnections are changing as the climate changes due to global heating or the secular fluctuations of the climate.

Evidence that the El Nino existed over four centuries ago is presented by Hanson and Maul (1989) and by Quin et al (1987). Recently, Wang (1995) reported on interdecadal changes of the El Nino onset. It is vital that water managers understand what effects these changes may have on the climate of Florida. In this analysis, we assumed the SOI to be synonymous with ENSO since the period of reliable record available to us was longer than the El Nino sea surface temperature anomalies.

Climate Fluctuations Related To Solar Sunspot Cycles

Global climate fluctuations that occur with a regular frequency may have their origins associated with solar activity. Sunspot activity displays a cyclic pattern with an approximate periodicity of Il years. The period may actually vary between 9 and 14 years. Periods tend to be shorter when the peak of the sunspot activity is more pronounced and longer the peak is less pronounced. Between each 11-year cycle there is a reversal in the direction of the sun's magnetic field. Therefore conditions only repeat themselves every 22-years. The 22-year period is known as the Hale cycle.

In spite of some statistical evidence of a relationship between solar sunspot cycles and the earth's climate fluctuations in certain parts of the world, no completely acceptable theory has been introduced that explains how the very small changes in the ultra-violet energy flux across the outer bounds of the earth's atmosphere due to sunspot cycles can be translated into climatic fluctuations. Willet (1953, 1987) has elaborated that solar flare activity may cause geomagnetic disturbances and strong spot heating that disrupt the zonal weather circulations. This would allow such activity to contribute significantly to climate fluctuations without appreciable changes in energy flux. The aa index of geomagnetic activity was taken by Willet to be the best indicator of solar flare activity. Recent research (Labitzke and van Loon 1989, 1992, 1993) provide additional new evidence that an important connection exist between solar cycles and the earth climate. Enfield and Cid (1991) and Mendoza et al (1991), report on possible connections between solar activity and El Nino's, while Reid and Gage (1988) and Reid (1989) reported on the similarities between the 11-year running means of monthly sunspot numbers and global sea surface temperature.

History of ENSO, Solar Activity, Geomagnetic Activity and Lake Okeechobee Tributary Inflow

The period from 1930 through 1960 contains three sunspot cycles that exhibit increasing sunspot and geomagnetic activity with each cycle. The last cycle exhibits much larger activity than normal. Willet (1987) identified the period of the first three sunspot cycles as being a period of the greatest global warming within the past 500 years. The third sunspot cycle occurred during a period in which Lake Okeechobee received four of the largest inflow years (1957-1960). High levels of geomagnetic disturbances continued throughout this period.

The fourth sunspot cycle which lasted from 1964 until 1978 is one of lesser solar activity. Below normal rainfall and droughts were characteristic of this period in central and southern Florida. Interestingly the geomagnetic activity was delayed during this cycle so the minimum in geomagnetic activity associated with the sunspot minimum of 1977 did not occur until the summer of 1980 and appears to be a precursor of the 1980-1982 drought in south Florida. Lake Okeechobee reached its' lowest recorded water level in July, 1981.

Other extended dry periods including the mid 1940's and the mid 1950's were also periods of low geomagnetic activity which indicates that this activity may be linked to the south Florida climate. Paine (1983) presented a hypothesis that would connect large anomalies in rainfall along the eastern coast of North America during this period to solar activity.

Hanson and Maul (1991) used Superposed Epoch Analysis to examined rainfall the years prior and during moderate to strong El Nino years. These El Nino years were defined as those events in which the El Nino Event lasted 2 years or more and that the year prior to the two years must be an non-El Nino year. The years they determined were strong El Nino years within our study period included: 1939-1940, 1957-1958, 1972-1973 and 1982-1983. Their most significant findings for Florida included: l) below normal rainfall over the entire state of Florida during the winter and spring the year prior to an El Nino event, 2) above normal rainfall over all the state during the winter and spring of the second year of the anomaly. The rainfall anomalies were greatest over the southern half of the state ranging between 145% to 166% of normal.

It is interesting to note that the 1959-1960 period was not accompanied by an ENSO event. The geomagnetic activity, however, remained very active during this period as Lake Okeechobee received it's largest two year inflow. Two separate peaks of large inflow to the Lake appear to correspond to separate peaks in geomagnetic activity. To understand the factors effecting the south and central Florida climate and therefore Lake Okeechobee inflow, the geomagnetic disturbances, sunspot number and ENSO events appear to need consideration in unison. During certain periods the effects of these processes may work together to enhance the likelihood of severe floods or droughts or sometimes to work against each other to lessen the likelihood of an extreme event. In addition to the indices discussed above the suns polarity and month of the year is included as input for the neural network.

A brief Introduction of Artificial Neural Networks

An artificial neural network (hereafter referred as neural network or network) is a computing method inspired by structure of brains and nerve systems. A typical neural network consists of a group of inter-connected processing units which are also called neurons. Each neuron makes its independent computation based upon a weighted sum of its inputs, and passes the results to other neurons in an organized fashion. Neurons receiving input data form the input layer, while those generate output to users form the output layer. A neural network must be trained by data for a problem. The training process is to adjust the connecting weights between each neurons so that the network can generalize the features of a problem and therefore to obtain desired results.

Among other advantages when compared with analytical approaches, the neural network approach does not require human expert knowledge in terms of mathematical descriptions of the problem. A neural network is trained from training data sets. This made neural network an appealing tool in dealing with complex systems, especially those of which the analytical descriptions may yet limited while their solutions are more of concerns, such as the problem discussed in this paper. The mathematically descriptive knowledge of the relationship between the solar activities and our regional climate fluctuations are limited. while the outcome of the climate may yield significant impact on water management.

Neural networks have received attention from many professions. In water resources and hydrology, neural network has also been finding its various applications (Karunanithi et al. 1994; Smith and Eli, 1995; Crespo and Mora, 1993; Grubert, 1995; Raman and Sunilkumar, 1995; Derr and Slutz, 1994)

Back Propagation

Among the variety of neural network paradigms, the Back-propagation is the most common in use and has been applied successfully to a broad range of areas such as speech recognition, autonomous vehicle control, pattern recognition, and image classification. Its training procedure is intuitive because of its relatively simple concept: adjust the weights to reduce the error.

Back-propagation networks' topology are usually layered, with each layer fully connected to the layer before it and the one next to it. The input to the network propagates forward from the input layer, through each intermediate layer, to the output layer, resulting in the output response. When the network corrects its connecting weights, the correction process starts with the output units and propagates backward through each intermediate layer to the input layer - hence the term Back propagation.

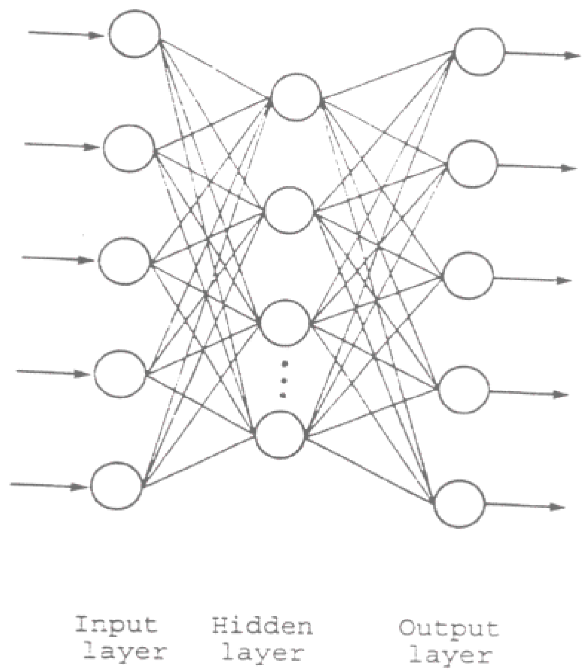


Figure 1. Schematics of a three layer back-propagation neural network

A typical back-propagation neural network has processing units, Figure l shows the topology three or more layers of for a typical three-layer network. The left most layer of the network is the input layer, the only units in the network that receive input data. The middle layer is also called hidden layer, in which the processing units are interconnected to layers right and left. The right most layer is the output layer. Each processing unit is connected to every unit in the right layer and in the left layer, but it is not connected to other units in the same layer. A back-propagation network can have one or more than one hidden layers, although many have one or two hidden layers.

There are two phases in its training cycle, one to propagate the input pattern and the other to adapt the output. It is the errors that are backward propagated in the network iteration to the hidden layer(s).

A detail description of the mechanism of back-propagation neural network can be found in books in the field, such as the one by Rumelhart and MeClelland, 1986.

Choosing a Network Configuration

The size of input layer and output layer are fixed by the number of inputs and outputs our prediction requires, i.e. 5 input layer neurons for all the five input variables, a single output neuron for the predicted change of inflow to the Lake Okeechobee. there is no universally applicable formula to be used for deciding the size of middle layers. In general networks with too many hidden neurons tend to memorize the input patterns and may lack of generalization, while those with too few hidden neurons may not be able to simulate a complex system at all. In the former case, a network responses to the training data very well, but when presented with the data it has not seen before, it falls to generate responsive outputs. In the latter case, a network may not have sufficient dimensions to be trained for the problem and its performance may not be improved no matter how many training it received. A network with more hidden neurons also requires more computing power and more training time needed. The best way to determine the number of middle layers and their sizes is trial-and-error. This can also be helpful to reveal the underlining relationships between variables. The rule of thumb is to start with the smallest size possible for a given problem to allow for generalization, then to increase the size of the middle layer(s), until the optimal results achieved.

We experimented with both one and two hidden layer configurations, with the size ranging from 3 neurons to II neurons, and found the one hidden layer with 6 neurons most suitable to the problem.

Input Data Preparation

This procedure is crucial to the success of applying neural network approach. The performance of a neural network largely depend upon the data set it was trained. in general, the better the training data sets represent the objective system, the better performance of the neural network. The preparation includes the selection of input variables, the examination of the data to eliminate bad data points, averaging, and normalizing.

The selection of input variables is solely problem dependent. After analyze the problem. five variables were chosen for this study. They are: Southern Oscillation Index (SOI), Sun Spot Number (SSN), aa-index, polarity index, and month index. Data source for the monthly SOI was the Climate Analysis Center[3] while the monthly aa indices and smoothed sunspot number were obtained from the National Geophysical Data Center[4]. The Lake Okeechobee inflows include net rainfall on the Lake and are computed by adding historical outflows to the storage change estimated from water level fluctuations for a particular time period. Prior to 1963 the computed inflow values were obtained from the United States Army Corps of Engineers (1978). After 1963 values were computed from hydrologic data obtained from the South Florida Water Management District.

Our goal is to use the information of past 6 months, including current month, to predict future 6 months inflow to the Lake Okeechobee. Therefore, a six month running averaging is applied to the raw data. All input variables were averaged for past six months, including current month, and the observed inflow data was averaged for the future 6 months. This is also necessary to farther factor out local noises of the data (Derr and Slutz, 1994).

Because neurons at the middle layer fire when their input data exceeds a threshold, neural network are more responsive to a particular range of input data, it is necessary to normalize the data to the range from 0 to 1. This was done in two steps. Step one, normalize each variable by using their respective mean and standard deviation as following:

normalized value = (value - mean)/standard deviation

Step two, use Sigmoid function to further normalize the data to the range from 0 to 1.

Network Training

The prepared data are 6 time series data sets, 5 for input variables and one for the target values. The duration of this time series ranging from March, 1933 to July, 1995. Each set was divided into two sections, one for training and the other for testing. An assumption on which this prediction is based is that the past data provide adequate patterns from which future events may be deduced. The duration for training data is from March 1933 through April 1987, total 650 averaged monthly data points. The duration for the testing data is from May 1987 through July 1995, total 99 data points.

All the training and testing of the neural network was done on a SPARC 20 workstation. Typical training times located between one to five hours.

Results And Conclusion

After training, the testing data were presented to the network to generate the forecasting results as shown in Figure 2.

The testing period contained a moderately severe dry period from September 1988 through May 1990 and the very wet year of 1994. The neural network was able to predict both of these events illustrating the sensitivity of south central Florida's hydrologic conditions to the global climate factors.

The best global indicator of a possible drought during the 1988-1989 period was a very strong La Nina that occurred during this period. The geomagnetic activity was high during this period and appeared to be out of phase with the SOI as an indicator of drought for the region. This likely explains why the network did not predict as severe a drought as the one that actually occurred and might be expected by the strong La Nina event. The magnitude of the peak of the 1994 period was better predicted by the network.

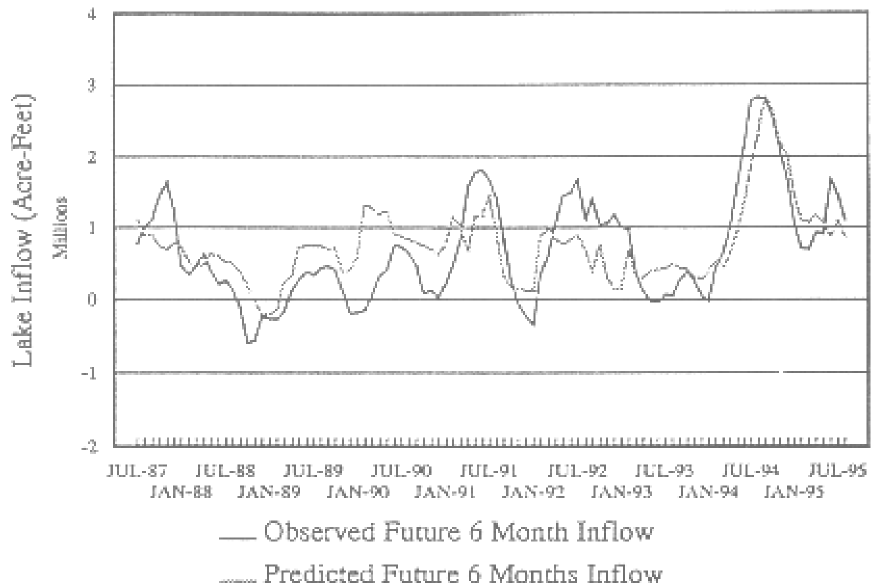


Figure 2. Predicted versus actual Lake Okeechobee inflow

Experiments attempting to train the neural network with only the SOI failed. This seemed somewhat surprising at first since the two extreme events of the period seemed adequately explained by ENSO alone. However, two of the wettest (1959-1960) and driest (1980-1981) periods on records and several other episodes during the training periods could not be explained in terms of the ENSO.

The increase in geomagnetic activity in 1989 may have been a precursor of things to come. This high level of activity has continued through the 1990s. Inflows to Lake Okeechobee have returned to more normal levels as illustrated in Figure 3. There has also been an extended El Nino event that enhanced flows during this period. The last three decades have been very dry for southern and central Florida as indicted in this same Figure. The neural network was able to indicate the return to a wetter conditions.

The predictor should be useful for operation purposes of Lake Okeechobee when used in conjunction with existing hydrologic conditions in the Lake Okeechobee tributary basins and the Lake Okeechobee water level.

Future Studies

Experiments including other global inputs such as the Pacific-North American (PNA) index. the Quasi-Biennial Oscillation, and the North Atlantic Oscillation and local antecedent hydrologic conditions need to be considered. The forecasting may be improved by also training the neural network with t:rends of global indices. Comparison to the predictions of traditional methods such as statistical analysis is also desirable.

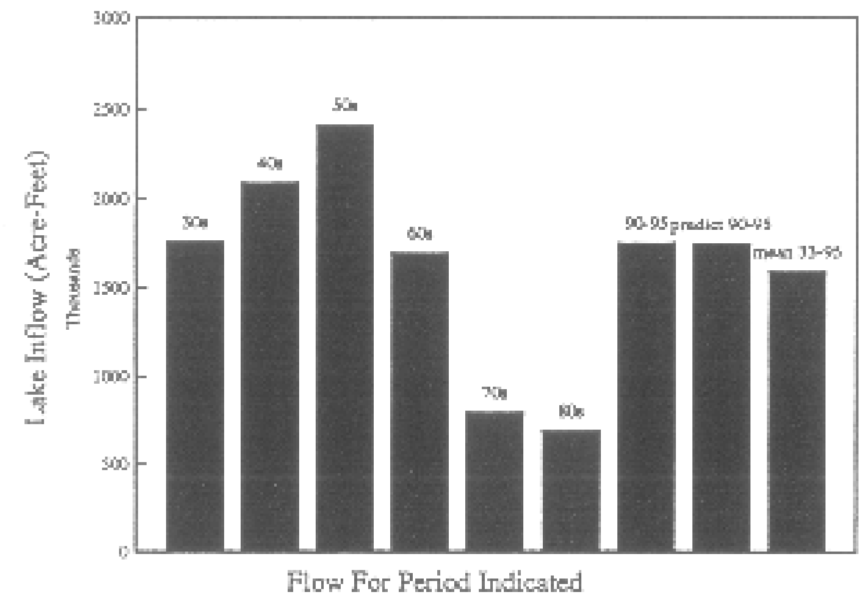


Figure 3. Decadal shifts of average annual inflow

Footnotes

l: South Florida Water Management District, Department of Planning, 3301 Gun Club Road. West Palm Beach, FL 33406

2: South Florida Water Management District, Department of Planning, 3301 Gun Club Road. West Palm Beach, FL 33406

3: Climate Analysis Center, Camp Springs, Maryland

4: National Geophysical Data Center, Boulder, Colorado

References

Crespo, J.L. and Mora . E.. Drought estimation with neural networks, Advances in Engineering Software, 18 (1993)

Derr, Vernon E. and Slutz. Ralph J.. Prediction of El Nino Events in the Pacific by Means of Neural Networks, Al Applications. 8. 51(1994)

Grubert John P.. Predication of Estuarine lnstabilities with Artificial Neural Networks, .Journal of Computing Civil Engineering. 226(995)

Hanson, K. and G.A. Maul, Florida Precipitation and The Pacific El Nino, 1859-1989, Florida Scientist,54, pp. 160-168 (1991)

Jason Smith and Robert N. Eli, Neural Networks Models of Rainfall-Runoff Process, Journal of Water Resources Planning and Management, 121, 56 (1995)

Labitzke, K. and H. van Loon, Association Between the 11-Year Cycle, the QBO and the Atmosphere, Journal of Climate, 2(1989)

Labitzke, K. and H. van Loon, Association between the 11-Year Cycle and the Solar Cycle, Journal of Climate. Part 5,5(1992)

Labitzke, K. and H. van Loon, Some recent studies of probable connections between solar and atmospheric variability, Ann. Geophysicae,11(1993)

Mendoza, B., R. Perez-Enriquez, and M. Alvarez-Madrigal, Analysis of solar activity conditions during periods of El Nino events, Ann. Geophysicae, 9, 50-54,(1991)

Nachimuthu Karunanithi et al, Neural Networks for River Flow Prediction, Journal of Computing in Civil Engineering, 8,201(1994)

Paine, D.A., Climate hypothesis Describing the Terrestrial System as a Frequency Domain With Specific Response Characteristics, American Geophysical Union, vol. 64, No. 26, (1983)

Quin, W.H., V. T. Neal, and S.E. Antunez De Moyola, El Nino Occurrences over the past four and a half centuries. Journal of Geophysical Research, 92(1987)

Raman, H. and Sunilkumar, N., Multi-variate modeling of water resources time series using artificial neural networks, Hydrological Sciences, 40(1995)

Ramusson E. U. and P. A. Arkin, A Global View of Large Scale Precipitation Variability, Journal of Climate, 6 (1993)

Reid, G.C., Solar irradiance variations and global Ocean Temperature, Journal of Geomagn. Geoelectr., 43,pp. 795-801 (1991).

Reid,G.C. and K.S. Gage, The climatic impact of secular variations in solar irradiance, in Secular Solar and geomagnetic Variations in the Last 10000 years, Eds. F.R. and A.W. Wplfendale, Nato AS] Series, Kluwer, Dordrecht, 225-243 (1988)

Rumelhart, D.E. and J.L.MeClelland, Parallel Distributed Processing, MIT Press, Cambridge, Mass (1986)

Trimble P.J. and J.A. Marban, A Proposed Modification To Regulation Of Lake Okeechobee, Water Resources Bulletin, 25 (1989)

United States Army Corps of Engineers, Central and Southern Florida Project, Rules and Key Operating Criteria, Master Manual, Jacksonville District, Jacksonville, Florida (1978)

Wang, B., Interdecadal Changes In El Nino in the Last Dour Decades, Journal of Climate, 8(1995)

Willet, H. C., Atmospheric and Ocean Circulations as Factors in Glacial-Interglacial Changes of Climate, Climate Change, Harvard University Press, Cambridge Mass., U.S.A.,pp. 51-71, (1953)

Willet, H.,C., Climatic Responses to Variable Solar Activity-Past Present and Predicted MIT, Cambridge, in Rampino, MIchael R, N., Climate: History, Periodicity, Predictability Van Nostrand Reinhold Company, Inc. (1987)