NOAA Technical Memorandum ERL OD-21

A COMPARISON OF NEURAL NETWORK AND LINEAR REGRESSION MODELS FOR PREDICTING EL NIÑO EVENTS

Robert P. Baptist
Christine E. Yeary
Ralph J. Slutz
Vernon E. Derr

Office of the Director
Environmental Research Laboratories
Silver Spring, Maryland
April 1994

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Environmental Research Laboratories

NOAA Technical Memorandum ERL OD-21

A COMPARISON OF NEURAL NETWORK AND LINEAR REGRESSION MODELS FOR PREDICTING EL NIÑO EVENTS

Robert P. Baptist
Christine E. Yeary
Cooperative Institute for Research in Environmental Sciences
University of Colorado/NOAA
Boulder, Colorado

Ralph J. Slutz
Vernon E. Derr
Office of the Director
Boulder, Colorado

Office of the Director
Environmental Research Laboratories
Silver Spring, Maryland
April 1994

UNITED STATES
DEPARTMENT OF COMMERCE

Ronald H. Brown
Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

D. James Baker
Under Secretary for Oceans
and Atmosphere/Administrator

Environmental Research
Laboratories

Alan R. Thomas
Director

NOTICE

Mention of a commercial company or product does not constitute an endorsement by the NOAA Environmental Research Laboratories. Use of information from this publication concerning proprietary products or the tests of such products for publicity or advertising purposes is not authorized.

For sale by the National Technical Information Service, 5285 Port Royal Road
Springfield, VA 22061

Contents

Abstract

Artificial intelligence and statistical techniques can be used in domains where functional relationships between variables are partially or completely unknown. The El Niño phenomenon of the tropical Pacific Ocean consists of major climatic changes which produce significant changes in rainfall, storm systems over much of the two Americas, and various effects on the fisheries off the coast of the Americas. Sea surface temperature is a significant indicator of El Niño events, so this paper analyzes the effectiveness of two methods of forecasting it. One method is the use of a neural network, and the other is the use of linear regression. The methods are developed and tested using data from the Comprehensive Ocean-Atmosphere Data Set, which contains ocean surface temperature, air pressure, and wind data from 1884 to the present. They are compared by determining the root-mean-square differences between forecast and actual values, and by examining the ability of the models to predict the onset, cessation, and continuation of El Niño events. For lead times of one to six months, both methods predict to better than a 1-degree C accuracy the sea surface temperature of a region in the Pacific Ocean that describes the El Niño. The continuation of El Niño events is also forecast with some success. Comparisons of the prediction skills of the two techniques gave mixed results, with the realization that each forecasting method reacts differently to changes in input data representations.

Introduction

The study of neural networks, or connectionist systems as they are often called, arose in an attempt by the Artificial Intelligence (AI) community to use computers to model the processes of cognition in the human brain. These endeavors were buoyed by the discovery of a method of "learning by example" which such systems could employ. In neural networks this is called "back-propagation", originally described by Werbos (1974), but resurrected, popularized, and re-discovered independently by McClelland and Rumelhart (1986). Much has been written recently about the power of such networks to forecast future values of time-series based on patterns gleaned from past data (see Weigend and Gershenfeld,1994). Indeed, neural networks are powerful tools for several reasons. They are efficient and use relatively small amounts of computer time compared to complicated mathematical models. Additionally, to create a successful model, one does not need to know the functional relationships between dependent and independent variables, as the network learns these associations itself. Finally, neural networks have the ability to learn non-linear relationships between input and output vectors.

Despite these advantages, the success of connectionist systems seems to be domain dependent, as they are successful in some applications, but not in others. A sufficiently general study has not yet been completed proving that neural networks are definitively better than more common statistical methods at modeling and prediction. This study does not attempt to do that. Instead, we compare the accuracy of a specific type of neural network with a specific statistical method, namely linear-least-squares regression.

The time-series being predicted is composed of sea surface temperatures in a region of the Pacific Ocean we will call R3 (see Figure 1), and is obtained from the Comprehensive Ocean-Atmosphere Data Set (COADS). These sea surface temperatures (SSTR3) describe a climatic phenomenon known as the El Niño, a recurrent event that has social and economic impacts on the world population because of its effect on rainfall and wind patterns, and thus fisheries.

The work that provided the foundation and inspiration for this study is found in Derr and Slutz (1994), without which this paper would not be possible.

Description of El Niño Phenomenon

Although several definitions of the El Niño phenomenon exist, it is often defined as the appearance of anomalously warm surface water, exceeding in a given month one standard deviation from the long-term temperature mean for that month, along the coast of Ecuador and the Galapagos Islands. The anomalous temperature should persist for four or more months as measured at five coastal stations (Quinn et al 1987). We chose, using the neural network and linear regression methods, to predict warm and cool events, defined as the occasions when SSTR3 exceeds one standard deviation, plus or minus, from the long-term mean. Such warm events we will refer to as El Niño, or as La Niña in the case of cold events.

The El Niño has a major effect on people and markets around the world. It is often considered that the Americas face a higher frequency of severe weather during warm event years. The west coast of the Americas will often see more violent storms, while other areas inland may experience unusual droughts. Disruptions in the usual annual upwelling of deeper, colder water in the vicinity of the South American coast are caused by a combination of processes triggered by the warmer waters and by changes in wind patterns in the Pacific which accompany the El Niño. The cool upwelling of the deeper water, rich in nutrients, normally supports extensive fisheries along the South American coast. During warm events, fish catches are radically reduced in the area. The most common fish in the area are anchovies, which are caught and ground into fish meal to be distributed around the world as a protein supplement for chickens and cattle. World markets in fertilizer and chicken feed are disrupted because of the disappearance of the fish. Funding for fisheries is controlled by banking systems that will not lend money if they anticipate that the market will fail in the near future. Accurate predictions of events, therefore, are economically and socially consequential.

Prediction of El Niño events is made more difficult by two of their characteristics. First, the causes of events are not well understood, as it is a truism that every El Niño differs in the details of its occurrence and its strength. Second, El Niño events are statistically rare. Since 1854, when scientific data began to be collected, they have occurred in intervals of four to ten years. From 1950 to 1990, there were approximately twelve El Niños.

The COADS Data Set

In 1981, the Environmental Research Laboratories of the National Oceanic and Atmospheric Administration (NOAA) of the Department of Commerce, the National Climatic Data Center of NOAA, and the National Center for Atmospheric Research began to collect ocean-based weather data for all the world's oceans. These were secured from all nations where they were available. These data became the COADS, which was first issued in 1985 (Slutz et al, 1985) and is currently being updated and extended (Diaz et al, 1992).

COADS contains over 70 million observations available individually or averaged over 2-degree by 2-degree latitude-longitude "squares", and averaged over months or decades. For this project, we used monthly averages from COADS, and combined the area averages into larger areas shown in Figure 1. A 3-month running average was formed to give "seasonal" data for each month. These data were also averaged over the years 1950 through 1979 to give long-term means for each month.

Regression Model Design and Implementation

Statistical regression (or just "regression" as we will call it) allows us to make simplified, quantitative estimates of relationships between a dependent variable and a set of independent variables. Several linear regression models were created to look for a relationship between the dependent variable SSTR3 and several independent variables that would allow us to predict it with some skill. The linear-least-squares technique was used to calculate coefficients in the statistical regression equation. That is, the coefficients $\beta_0, \beta_1, \ldots, \beta_K$ were calculated in the equation

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

(where "i" is an index of the month) so that

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

was minimized. "Y-hat$_i$" represents the predicted value of the dependent variable, while y_i represents the known value of said variable. The x_{ik}'s are the values of the k independent (predictor) variables.

In most tests, different values for $\beta_0, \beta_1, \ldots, \beta_K$ (the "beta coefficients") were calculated for each prediction lead time. The function that computed "y-hat$_i$", therefore, was dependent on how far ahead SSTR3 was predicted. In a special test, different values for the beta coefficients were calculated for each month and prediction lead.

In about half of the tests, a backward stepwise variable selection technique was implemented using subroutine "RSTEP" from the IMSL library of FORTRAN subroutines for statistical analysis. (IMSL, Statistics vol. I) Initially, all variables were entered into the model. An attempt was then made to remove a variable from the model, the "backward step." Variables were removed according to their p-values. The p-value is related to the significance of the coefficient in the regression model as examined by a t-test. A smaller p-value indicates that there is a greater statistical confidence that the coefficient is significant. (Keller, 1991.) In the series examined, the RSTEP routine was instructed to remove a variable if its p-value was greater than .10, a common value in such models. The backward stepping process was continued until all of the variables remaining had a p-value of less than .10. This technique is less computationally expensive than trying all possible combinations of independent variables. Its use, however, does not guarantee that the inputs selected will be the best possible set (the set that produces the lowest error) over all possible subsets of independent variables.

In the other half of the regression tests, the variables were not subjected to the backward stepwise technique. (See Table 3 for descriptions of the regression models and their inputs.) Instead, all variables were "forced" into the equation regardless of significance.

Neural Network Design and Implementation

The neural networks used in this study consisted of three layers of mathematical functions that the computer uses to simulate memory elements, called "neurons." The independent variables were fed as normalized values into the first neuron layer called the "input neurons." The outputs of each input neuron were next multiplied by real numbers, called weights. The sums of these products were then fed into the next neuron layer, composed of "hidden neurons." The outputs of these neurons were then multiplied by a second set of weights. The sum of these products was entered into to the final layer consisting of one "output neuron." The value produced from this output layer was the network's SSTR3 prediction. The nonlinear functions representing the neurons at each of these layers are commonly (and were in this case) of sigmoidal shape. They help to limit values output by the neurons. Since the neurons are non-linear mathematical functions, the process is non-linear. In general, any number of layers of hidden neurons are possible, as are any number of neurons within each layer. The network structure described here is common and is often called a "feed-forward architecture" because inputs are fed into the network and propagated forward to produce outputs. This network is also "fully connected", as each input/hidden and hidden/output neuron pair is assigned a distinct weight. Figure 2 shows a diagram of a fully connected network with arbitrary numbers of thirteen input, five hidden, and three output neurons. The lines connecting neurons in subsequent layers represent the weights in this network. The "activation" arrow on the left side of the diagram represents the forward propagation of values from the input to output layer characteristic of feed-forward networks.

On the right side of diagram in Figure 2, there is a "weight changes" arrow, pointing in the opposite direction of the activation arrow. This "reverse" component is a characteristic of networks, including the one in this study, that learn by back-propagation. Learning in these systems involves changing the initially random matrices of weights to minimize the errors between what the output neuron produces and what it "should have" produced. During this "training" of the weights, therefore, the user must supply many vectors of inputs and targets associated with these inputs. The system then makes its weight changes by means of "gradient descent" of the error surface. This process begins at the output layer and ends at the input layer, thus the concept of a backward flow. Back-propagation is described in detail in Winston (1992). Typically, and in this study, several different input-output sets are repeatedly presented to the network until it matches all targets plus or minus a preset error tolerance. After training, the network can be tested by feeding it inputs (a validation set) that were not in its training corpus, and comparing its outputs with known values. If this testing is successful, the network can be used to predict the future from current data.

The networks constructed for our tests were composed of varying numbers of input neurons which depended on the set of input data. The number of hidden neurons was chosen after testing several trained networks with numbers of hidden neurons varying between ten and 400. With twelve inputs, it was shown that prediction error was minimized at about thirty hidden neurons. Most tests contained eleven or twelve inputs, therefore thirty hidden neurons were chosen in most instances. For sets that used more than twelve inputs, the number of hidden neurons was increased in proportion with the number of inputs. Testing the number of hidden neurons is best approached experimentally, as the most efficient value depends on the character of the data and structure of the network. Too few hidden neurons can lead to an "overload" of the network causing important information to be lost; too many hidden neurons will cause a network to "overlearn" relationships between variables leading to the learning of noise instead of important associations.

All training and testing of the neural networks were performed on a personal computer with a 486 processor. Typical training times lay between two and twenty-four hours.

Choice and Composition of Data

The data which composed the sets used to train the neural networks and calculate the beta coefficients in the regression were from COADS and were seasonally averaged. These sets contain data from each month of the years 1935 to 1969. Although the composition of data in the neural network and regression models was varied for testing purposes, all models contained the following eleven variables:

1. Southern Oscillation Index (SOI: surface atmospheric pressure at Tahiti minus that at Darwin)
2. Sea Surface Temperatures in regions R12, R3, and R4 (see Figure 1):

SSTR12, SSTR3, SSTR4

3. U-component (east-west) of the surface wind in regions R12, R3, and R4:

   UR12, UR3, UR4

4. Sea Surface Temperatures in regions J1 and J2 (Again refer to Figure 1):

   SSTJ1, SSTJ2, SSTJL1, SSTJL2.

The sea surface temperatures in regions J1 and J2 we will call Joseph data, named for Porathur V. Joseph, who suggested them. SSTJL1 and SSTJL2 are lagged values. JL1 (JL2) is the value of J1 (J2) nine months earlier than the month for which the prediction was being made. These lags were chosen because studies by others (which we verified) have shown that sea surface temperatures in the Joseph regions are highly correlated with later values of the sea surface temperature in region R3.

When making forecasts we used current variables, with the exception of SSTJL1 and SSTJL2. That is, the values of the variables were from the time at which the prediction was made. To introduce the possibility of the significance of trends in the inputs, we created nine additional lagged variables and added them to the neural network and regression models. Values of all the variables named above with the exception of SSTJL1 and SSTJL2 were taken from three months before the time the predictions were made. Note the subtle yet crucial difference between these nine variables and the SSTJL1 and SSTJL2 values: the nine variables were from three months before the time the predictions were being made, and SSTJL1 and SSTJL2 were from nine months before the time *for which* the predictions were being made.

If one or more variables were missing from input or output data in a month, the month was thrown out of the set. Tests we performed showed that neural network predictions were affected less than two percent by deliberate removal of a random ten percent of our training corpus. Data were primarily missing because of wartime conditions and loss of records.

Table 1 provides a complete list of the variables, their units, and their maxima and minima over the period January 1935 to December 1969.

Varying Forecast Formulas for Different Months of the Year

The SSTR3 has a strong seasonal cycle. It varies regularly by an average of about two degrees C from one season to the next. Normally a meteorologist will use a different formula to calculate tomorrow's weather if it is in winter as opposed to summer. In parallel with this, we hypothesized our models might work better if the formulas used changed for different months of the year. We attempted to do this with two techniques.

The first technique involved adding as a variable the number of the month from which SSTR3 was being predicted. In the neural networks, an input neuron with possible discrete input values from one to twelve was added. In the regression models, another independent variable was added which took values from one to twelve depending on the month. Other methods advocated by the literature have been tested with the neural network, most notably a vector with twelve additional binary input neurons as month indicators; these results are not presented here. Methods such as these showed no advantage over the first method described above. Time prevented the testing of more elaborate techniques such as these in the regression models.

In the second technique, we trained separate sets of weights for the neural network and calculated separate sets of beta coefficients for the regression model for each month of the year. We hypothesized that using different functions for the different months would give both techniques a clear way to distinguish the month of the year. This was not pursued as vigorously as the month variable technique because it yielded no increase in skill.

Evaluation of Predictive Skill: Root-Mean-Square Error

The reader will be better able to understand the discussion in this and the next section by keeping Tables 2 and 3 handy. The characteristics of four sets of input variables used by the neural network to predict SSTR3 are described in Table 2. The input sets varied by which variables were used and by the way they were altered to distinguish different months for making forecasts. Four sets of input variables used in the regression tests were similarly diversified and outlined in Table 3. There was one additional variation in the regression tests. Two of the input sets were tested twice, once with the technique of forcing all variables into the regression equation, and again with the backward stepwise variable selection technique. These additional tests were performed to test the validity of using the backward stepwise method in this study.

Forecast formulas were developed with COADS data from January 1935 to December 1969, and all tests used COADS data from January 1970 to December 1992 for validation purposes. This validation period was chosen to permit comparisons between neural network and regression tests during a period where several warm events occurred. Root-mean-square error (rms) differences between the forecast value and actual SSTR3 values were calculated for all months in the validation set for each prediction lead and presented, averaged by prediction lead, in Tables 4 and 5. The mean of these results for all leads is also shown, and for leads four through six. Smaller numbers indicate greater skill.

Table 4 exhibits rms error scores for two groups of regression tests. One series in each group used backward stepwise regression and the other forced all variables into the regression equations. They were not different in any other respect. Of the twelve pairs of results for individual leads, the backward stepwise method was slightly better seven times and the forced method was slightly better two times, but the differences are too small to give a clear indication that either is better than the other. We can conclude in these cases that the backward stepwise regression model was not advantageous to the regression model that forced the use of all variables.

Table 5 presents scores comparing groups of neural network and regression tests that have similar input sets. The first comparison, NNX2 versus r2b1, shows the regression technique to be superior at predicting sea surface temperatures at every prediction lead. The input sets that comprised these two series contained the "basic" group of eleven predictors with no month number variable. One set of weights for the neural network and one set of beta coefficients for the regression model was created for each prediction lead time.

The next comparison, NNX versus r2b2, shows that the addition of a month indicator into the input set notably helped the neural networks at most prediction leads, but hindered the regression techniques' performance at most leads. This suggests that networks may be better at handling integer month indicators. It is possible that regression input sets containing eleven binary or "dummy" indicators for month distinction (where the twelfth month is represented when the eleven binary values are "off") might have fared better, but as mentioned, the neural networks we tested did not benefit from using this input representation.

The third group compares series where seventy-two sets of weights and beta coefficients, one for each month and prediction lead, were formed to predict SSTR3. Neither neural networks nor the regressions benefited from this type of differentiation by month. The method particularly hurt the regression predictions.

The final comparison in Table 4 involves series containing the additional inputs aimed at using input trends. These two series did better than their counterparts with less information. Adding additional information about input trends, in the cases that were tested here, was effective.

Skill of Warm Event Predictions

Measuring the accuracy of forecasting temperatures is not the only tool to gauge the success of these models. Imagine a system that is poor at predicting absolute temperatures (has a high rms difference between predicted and actual temperatures) but skilled at predicting the onset and cessation of warm events. Banks loaning money to fisherman might embrace such a model, only being concerned about a reliable "yes or no" answer to the question of whether an event is imminent.

Tests were performed to compare the skill of two sets of neural network and regression tests at predicting warm events. NNX was compared with r2b2, and NNR3 with r2a2. These series were chosen based on their similarities; refer to Tables 2 and 3 for descriptions. These tests were done on prediction lead times of three to six months into the future. They were performed using warm events in the validation set January 1970 to March 1993. There were five warm events in this period by the criteria of Quinn et al (1987).

Tables 6 and 7 contain examples of these results, the prediction of the 82-83 warm event four months into the future. (Table 6 displays the NNX-r2b2 comparison, and Table 7 the NNR3-r2a2 comparison.) The leftmost column in each Table shows the month the models were predicting for. The next three columns have three possible values per month. A "C", for a cold event, a "W" for a warm event, or an "O" for neither a warm nor cold event can appear in these columns. Under the "actual occurrence", the "C", "W", or "O" represents what was measured for SSTR3 during each month. The "NNX" (or "NNR3") and the "r2b2" (or "r2a2") columns contain which state of SSTR3 the models predicted for a particular month. Below each event is the tally of the number of months each model predicted a correct state for the period shown.

In the NNX-r2b2 comparison (on the full validation set), results are mixed. The neural net was more consistent at leads three and four, while the regression was slightly, but not significantly, better at leads five and six. Neither method was good at predicting the exact time of the beginning and completion of warm events, but there is clearly a correlation between actual and predicted states in most cases. At lead six, predictions appear random at times. Additionally, some events were predicted more easily than others. The shorter event from July 1976 to March 1977 was missed by both methods at most leads, while the longer event from June 1982 to October 1983 was predicted earlier and more consistently.

The NNR3-r2a2 results (on the whole validation set) are also mixed, but are generally better than those in the NNX-r2b2 comparison. Recall that NNR3 and r2a2 both contain a greater number of trend variables than NNX and r2b2. The additional variables improved event prediction. Events are generally predicted with greater accuracy and hardly ever missed altogether. Lead three shows a slight advantage to the neural networks, while leads four through six show the regression models to be moderately better event predictors.

Conclusions and Future Directions

Three of the four groups of rms error comparisons in Table 5 revealed lower mean scores for neural network tests compared with the statistical regression tests. The best choice of input data for both predictive methods, NNR3 and r2a2, produced rms errors, averaged over all leads, of .51 degrees and .57 degrees C respectively. This shows significant skill, since SSTR3 itself ranges over 8.5 degrees C. Neither technique consistently predicted the onset or cessation of any of the five events of the period January 1970 to March 1993 at any lead time. Despite this, each technique often showed skill at predicting the continuation of events some number of months after their inception. The definition of an event chosen here is somewhat arbitrary, but has the advantage of clarity and objectivity.

The usefulness of a forecast depends on who is using it and for what purpose. For fishermen who borrow money to outfit their boats and hire their crews, and for banks who loan them money, a three to five month lead time is probably sufficient. Though clearly not close to being foolproof, both regression and neural network models show skill at predicting sea surface temperatures. There is no clear "winner" between the techniques. RMS differences between predicted and actual temperatures are generally lower for the networks tested in this study, while event prediction results are mixed. It is apparent that slight changes in variables, month representation, and the way that the models are constructed can all affect results significantly. Other combinations of input data from other areas of the Pacific Ocean should be tested in a search for the most effective forcing functions for the El Niño phenomenon. This, along with the use of more advanced neural network and statistical models should be tested in the search for better forecasts of El Niño events.

Acknowledgments

The authors are highly grateful to Maurice L. Blackmon and Donald R. Mock, both of the NOAA Climate Diagnostics Center, for the use of their computer facilities. This project could not have been completed without them.

We also acknowledge the support of the staff at California Scientific Software in using their product, Brainmaker Professional, for this project. Although we cannot give a general endorsement, it worked well for us.

References

Derr, V.E., and R.J. Slutz. 1994. Prediction of El Niño Events in the Pacific by Means of Neural Networks. AI Applications, (in press).

Diaz, H.F., K. Wolter, and S.D. Woodruff. 1992. Proceedings of the International COADS Workshop. National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Boulder, Colorado.

IMSL STAT/LIBRARY Volume I: FORTRAN Subroutines for Statistical Analysis. IMSL, Incorporated, Houston, Texas.

Keller, G., and B. Warrack. 1991. Essentials of Business Statistics. Wadsworth Publishing Company, Belmont, California.

Quinn, W.H., V.T. Neal, and S.E. Antunez de Mayolo. 1987. El Niño Occurrences over the past four and a half centuries. Journal of Geophysical Research 92:14, 449-14, 461.

Rumelhart, D.E., and J.L. McClelland. 1986. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. MIT Press, Cambridge, Massachusetts.

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steurer, and J.D. Elms. 1985. Comprehensive Ocean Atmosphere Data Set, Release 1. National Oceanic and Atmospheric Administration, Environmental Research Laboratories, Boulder, Colorado.

Weigend, A.S, and N.A. Gershenfeld. 1994. Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley Publishing Company, Reading, Massachusetts.

Werbos, P.J. 1974. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. Thesis, Harvard University.

Winston, P.H. 1992. Artificial Intelligence. Addison-Wesley Publishing Company, Reading, Massachusetts.

Figure 1. Map of the El Niño-Southern Oscillation regions of the Pacific Ocean used for predictions of sea surface temperatures and warm and cool events. Observations of wind and sea surface temperature are spatially and temporally averaged over these regions.

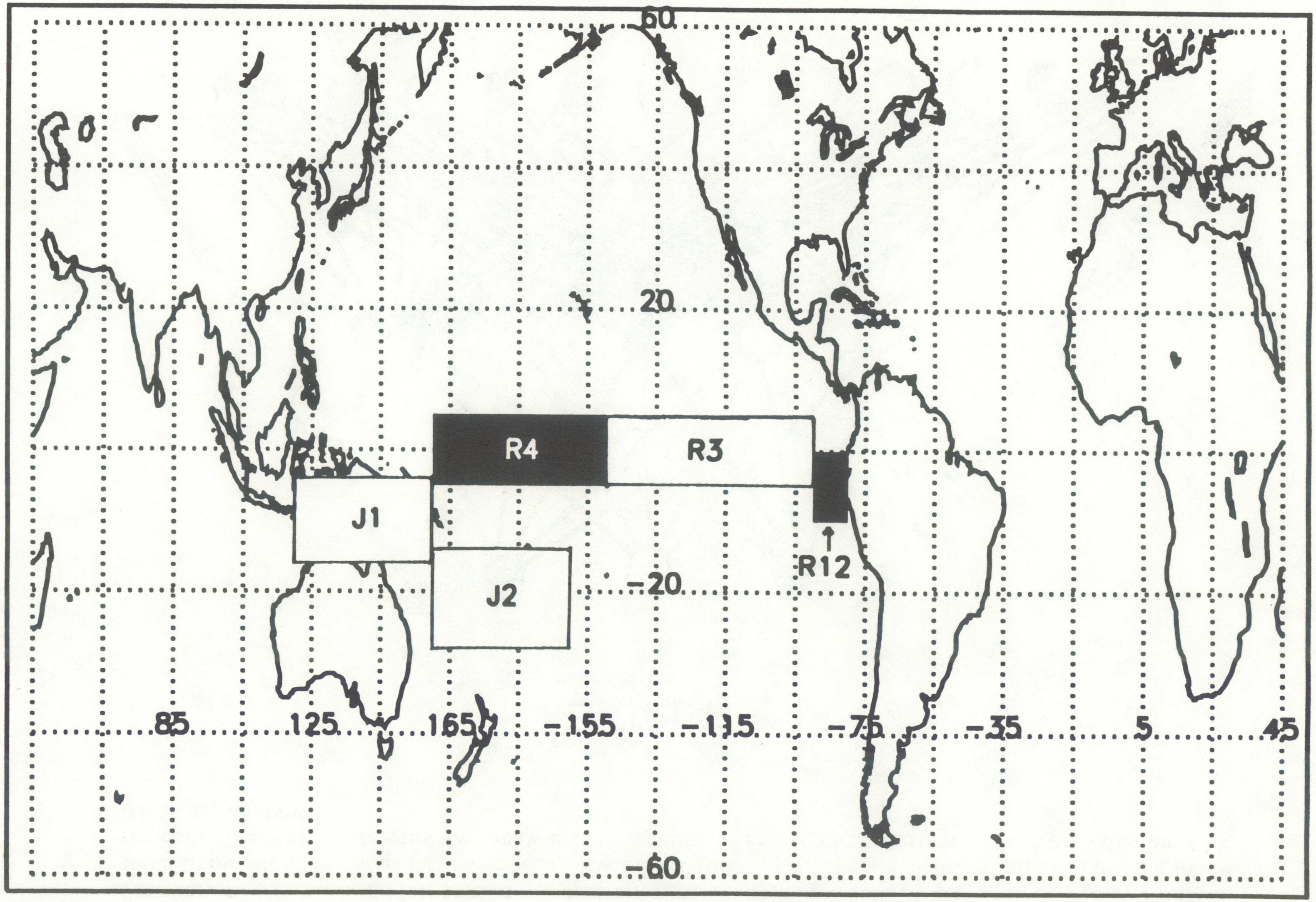

Figure 2. Diagram of a fully connected network with thirteen input, five hidden, and three output neurons. The lines connecting neurons in subsequent layers represent the weights in this network. The "activation" arrow on the left side of the diagram represents the forward propagation of values from the input to output layer characteristic of feed-forward networks. The "weight changes" arrow, pointing in the opposite direction, represents the backward propagation of errors, characteristic of networks that learn by back-propagation.

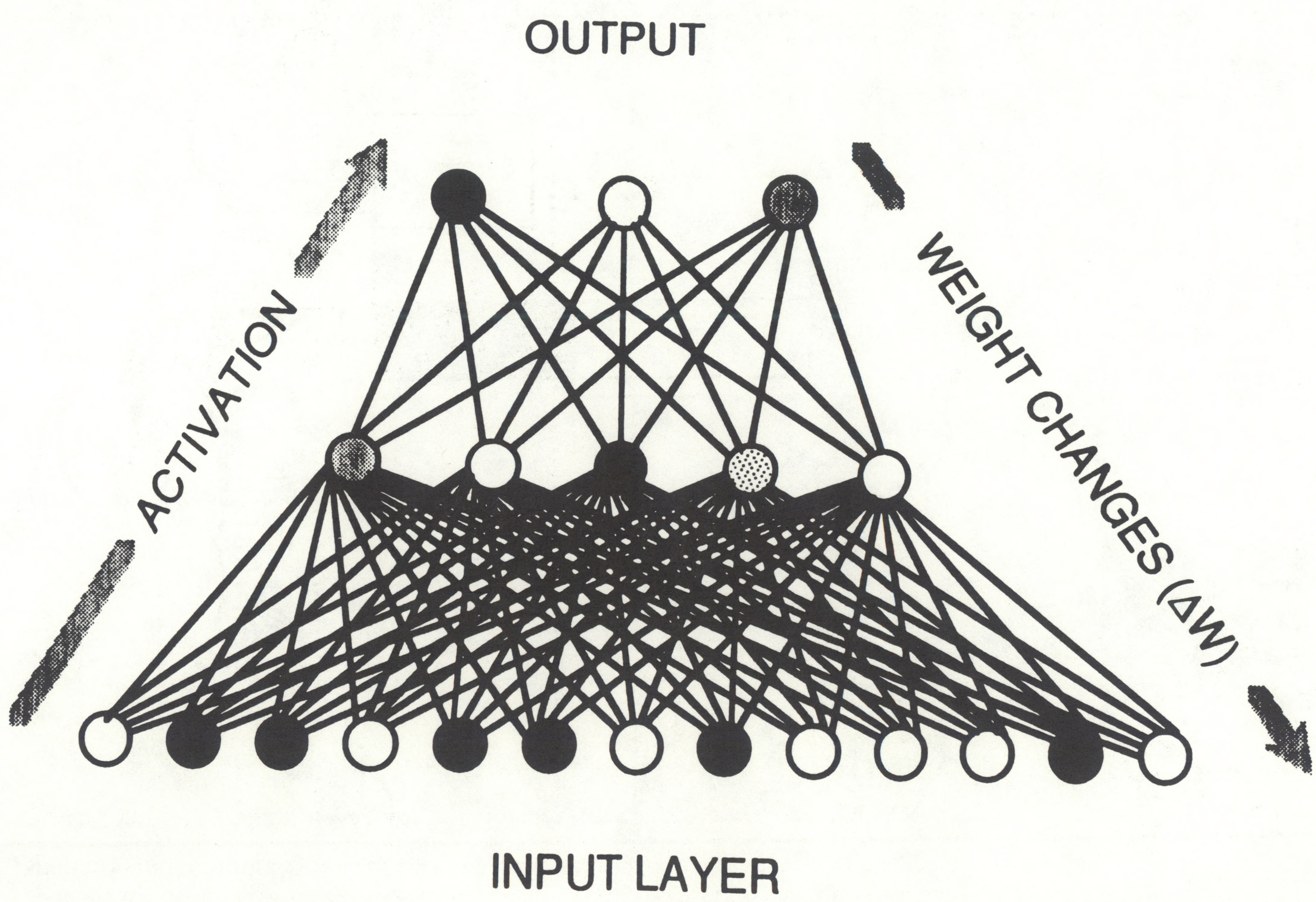

Table 1. The twelve variables used in this study, their maxima and minima over the period 1935 to 1969, and their units. Note: SST= sea surface temperature. All lagged variables referenced in the text retain the same maximum and minimum values as their non-lagged counterparts.

VARIABLE	MAXIMUM VALUE	MINIMUM VALUE	UNITS
Month number	12	1	
Southern Oscillation Index	2.27	-3.37	millibars
Horizontal wind, R2 (UR2)	2.93	-3.12	meters/second
Horizontal wind, R3 (UR3)	0.83	-5.27	"
Horizontal wind, R4 (UR4)	0.87	-7.03	"
SST R2	28.49	18.93	Celsius
SST R3	28.88	20.35	"
SST R4	29.57	26.25	"
SST J1	29.55	24.95	"
SST J2	27.54	22.22	"

Table 2. Descriptions of neural network tests.

Name	Characteristics	
NNX	12 Independent variables:	SOI,SSTR12,SSTR3,SSTR4,UR2,UR3,UR4, SSTJ1,SSTJ2 SSTJ1 and SSTJ2 as observed 9 months before the time for which the prediction was made Month number
	30 hidden neurons 6 sets of weights created; a different one was trained and tested for each prediction lead	
NNX2	11 Independent variables:	Same as NNX, except the month number was excluded
	30 hidden neurons 6 sets of weights created; a different one was trained and tested for each prediction lead	
NNM	11 Independent variables:	Same as NNX2
	30 hidden neurons 72 sets of weights created; a different one was trained and tested for each month of the year and prediction lead	
NNR3	21 Independent variables:	SOI,SSTR12,SSTR3,SSTR4,UR2,UR3,UR4, SSTJ1,SSTJ2 also all above variables as observed 3 months before the time at which the prediction was made SSTJ1 and SSTJ2 as observed 9 months before the time for which the prediction was made Month number
	52 hidden neurons 6 sets of weights created; a different one was trained and tested for each prediction lead	

Table 3. Descriptions of regression tests.

Name	Characteristics	
r2b2f	12 Independent variables:	SOI,SSTR12,SSTR3,SSTR4,UR2,UR3,UR4, SSTJ1,SSTJ2 SSTJ1 and SSTJ2 as observed 9 months before the time for which the prediction was made Month number
	No backward stepwise regression; all variables were forced into the model 6 sets of beta coefficients were created; a different set was calculated and tested for each prediction lead	
r2b2	Same as r2b2f, except the backward stepwise regression technique was implemented	
r2b1	11 Independent variables: Same as r2b2 except no month number included	
r1a1	Same as r2a1 except 72 sets of beta coefficients were created; a different set was calculated and tested for each month of the year and prediction lead	
r2a2f	21 Independent variables:	SOI,SSTR12,SSTR3,SSTR4,UR2,UR3,UR4, SSTJ1,SSTJ2 all above variables as observed 3 months before the time at which the prediction was made SSTJ1 and SSTJ2 as observed 9 months before the time for which the prediction was made Month number
	No backward stepwise regression; all variables were forced into the model 6 sets of beta coefficients were created; a different set was calculated and tested for each prediction lead	
r2a2	Same as r2a2f except the backward stepwise regression technique was implemented	

Table 4. Comparison of backward stepwise regression and "all forced" regression where all input variables were forced to be included in the formulas. RMS errors were calculated from the validation tests on data for January 1970 through December 1992. The two columns on the right give the means of the six different lead times, and the means of the leads for four through six months.

	RMS Difference (degrees Celsius)							
Series Tested	Lead1	Lead2	Lead3	Lead4	Lead5	Lead6	Mean	Mean(4-6)
r2b2f (bckwrd. step)	.29	.57	.77	.93	.86	.88	.72	.89
r2b2 (all forced)	.29	.58	.78	.91	.87	.90	.72	.89
r2a2f (bckwrd. step)	.20	.39	.51	.62	.75	.88	.56	.75
r2a2 (all forced)	.21	.42	.54	.62	.74	.88	.57	.75

Table 5. RMS error comparison of neural network and regression tests for forecasting El Niño sea surface temperatures. Results are grouped into pairs containing similar input sets to the neural network and regression models.

	RMS Difference (degrees Celsius)							
Series Tested	Lead1	Lead2	Lead3	Lead4	Lead5	Lead6	Mean	Mean(4-6)
NNX2	.36	.69	.79	.75	1.23	.92	.79	.97
r2b1	.27	.53	.69	.83	.87	.90	.68	.87
NNX	.45	.59	.53	.62	.65	.84	.61	.70
r2b2	.29	.58	.78	.91	.87	.90	.72	.89
NNM	.41	.49	.59	.60	.66	.72	.58	.66
r1a1	.45	.64	.78	1.01	1.46	.96	.88	1.14
NNR3	.37	.45	.48	.50	.61	.66	.51	.59
r2a2	.21	.42	.54	.62	.74	.88	.57	.75

Table 6. The abilities of the neural network model NNX (see Table 2) and the statistical regression model r2b2 (see Table 3) to predict warm events are compared. This Table is an example of tests done on the five warm events from January 1970 to March 1993 at prediction leads three through six. This Table shows results for the 82-83 El Niño with predictions made four months into the future.

Key:
W : Warm Event
O : No Event
C : Cold Event

Date Predicting For	Actual Occurrence	Lead 4 r2b2 Predicted	Lead 4 NNX Predicted
Feb 82	O	O	O
Mar 82	O	O	O
Apr 82	O	O	O
May 82	O	O	O
Jun 82	W	O	O
Jul 82	W	O	O
Aug 82	W	O	O
Sep 82	W	O	O
Oct 82	W	O	O
Nov 82	W	W	W
Dec 82	W	W	W
Jan 83	W	W	W
Feb 83	W	O	W
Mar 83	W	O	W
Apr 83	W	W	W
May 83	W	W	W
Jun 83	W	W	W
Jul 83	W	W	W
Aug 83	W	W	W
Sep 83	W	W	W
Oct 83	W	W	W
Nov 83	O	W	W
Dec 83	O	O	W
Jan 84	O	O	O
Feb 84	O	W	O
Mar 84	O	O	O
Apr 84	O	O	O
May 84	O	O	O
		19/28 Correct	21/28 Correct

Table 7. The abilities of the neural network model NNR3 (see Table 2) and the statistical regression model r2a2 (see Table 3) to predict warm events are compared. This Table is an example of tests done on the five warm events from January 1970 to March 1993 at prediction leads three through six. This Table shows results for the 82-83 El Niño with predictions made four months into the future.

Key:
W : Warm Event
O : No Event
C : Cold Event

Date Predicting For	Actual Occurrence	Lead 4 r2a2 Predicted	Lead 4 NNR3 Predicted
Feb 82	O	O	O
Mar 82	O	O	W
Apr 82	O	O	O
May 82	O	O	O
Jun 82	W	O	O
Jul 82	W	O	O
Aug 82	W	O	O
Sep 82	W	O	W
Oct 82	W	O	W
Nov 82	W	W	W
Dec 82	W	W	W
Jan 83	W	W	W
Feb 83	W	W	W
Mar 83	W	W	W
Apr 83	W	W	W
May 83	W	W	W
Jun 83	W	W	W
Jul 83	W	W	W
Aug 83	W	W	O
Sep 83	W	W	O
Oct 83	W	O	O
Nov 83	O	O	O
Dec 83	O	O	O
Jan 84	O	O	O
Feb 84	O	O	O
Mar 84	O	O	O
Apr 84	O	C	O
May 84	O	O	O
		22/28 Correct	21/28 Correct

*U.S. GOVERNMENT PRINTING OFFICE:1994-573-013/00009