

Unsupervised Neural Method for Temperature Forecasting

¹Juan M. Corchado* and ²Colin Fyfe

¹Department of Languages and Computing Systems
University of Vigo, Campus Universitario, 32004, Ourense, Spain
Email: corchado@ei.uvigo.es

²Department of Computing and Information Systems
University of Paisley, High St., Paisley, PA1 2BE, United Kingdom
email: fyfe-ci0@wpmail.paisley.ac.uk

Abstract

This paper presents the results of using a novel Negative Feedback Artificial Neural Network for extraction of models of the thermal structure of oceanographic water masses and to forecast time series in real time. The results obtained using this model are compared with those obtained using a Linear Regression and an ARIMA model. The paper presents the Negative Feedback Artificial Neural Network, shows how it extracts the model behind the data set and discusses the Artificial Neural Network's forecasting abilities.

Key Words: Time Series, Forecasting, Modelling, Adaptation, Real time, Oceanography.

* At the time the research presented in in this paper was performed this author was at the Department of Computing and Information Systems of the University of Paisley, UK.

1. Oceanographic Systems.

Every system, even the most complicated ones, such as the oceans of the world, can be modelled if its behaviour is fully known and understood¹. However the current knowledge of the ocean structure is still too weak to create a full model of its behaviour. The goal of the experiments presented here is to create an Artificial Neural Network (ANN) capable of modelling a Thermal Time Series and of forecasting its future values.

Ocean water masses are extremely heterogeneous. Each water mass has certain properties that differentiate it from the others, and the convergence area between different water masses can be very noisy. For example the Arctic and Antarctic convergence zones are extremely heterogeneous and very variable. These water masses have their own general characteristic that can be described and simulated. This has been previously shown with a Neural Network² that produces good results when trained and tested with samples taken from the same province but which does not forecast reliably when tested with data from a different province. This Artificial Intelligence (AI) approach to the problem of describing the ocean environment scene potentially offers advantages over a conventional algorithmic data processing approach, as it is able to deal with uncertain, incomplete and even inconsistent data. This capability appears particularly useful for the problem of describing and predicting the ocean environment.

This paper focuses on a time series (see Figure 1) recorded over almost 15000 km, in 1995, from a vessel going from the UK (latitude: 50.48, longitude: -1.25), to the Falkland Islands (latitude: -51.94, longitude: -58.19). The vessel crossed a number of water masses and well known fronts (See Table 1).

This time series exhibit high multicollinearity and heteroscedastidity³ and therefore requires a model capable of adapting itself to the spatial and temporal changes in the water masses. Figure 1 shows how much the temperature changes from the UK to the Falkland Islands. Fronts (boundaries between different water masses) are easy to spot in the plot but difficult to forecast in real time as will be shown in the following sections. The difficulty in forecasting the shape of a front and in general the thermal structure of any water mass is due to the fact that the water temperature depends on many factors such as atmospheric weather, sea floor geography, currents, seasonal variations, latitude, longitude, depth, etc. In addition the thermal structure of a front can be very different from one year to the next as shown in Figure 2. Changes are always within known limits, but these limits can be very wide (hundreds of kilometres and several degrees Celsius in temperature).

2. Time Series Prediction.

Time series⁴ prediction is based on the assumption that an observable feature of a system is determined by an underlying deterministic system. If the evolution of the system can be described by a set of n ordinary differential equations in n variables, there exists a unique trajectory through every point \mathbf{a} in \mathbb{R}^n . In order to

make a prediction it is desirable to know both the underlying rules of the deterministic system and the current state of the system. For example, consider a system in which we can observe a single scalar quantity, x_t , which is the value of x at time t and which is determined by the state, \mathbf{a}_t , of the system at time t . Now if the underlying system is based on the n dimensional set of differential equations referred to above, then the state, \mathbf{a}_{t+1} , of the system at time $t+1$ can be described by the set of equations $\mathbf{a}_{t+1} = F(\mathbf{a}_t)$, while the observable at time $t+1$ is given by $x_{t+1} = g(\mathbf{a}_{t+1}) = g(F(\mathbf{a}_t))$. It can be shown that there exists a value d such that the vector $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$, consisting of d consecutive observations of x , fully characterises the system. i.e. the evolution of the system may be specified absolutely using the function $F()$ or equally by using the function $H()$ where $x_{t+1} = H(\mathbf{x}_t) = H(x_t, x_{t-1}, \dots, x_{t-d+1})$. Therefore merely by finding a sufficiently long set of consecutive observations of x (d is known as the embedding dimension of the system) a complete specification of the future values of the future observations of the system may be obtained. It is well known, of course, that, for non-linear systems, the underlying dynamics are often such that there will be divergence of trajectories from nearby initial conditions. Thus, since it is not possible to measure observables to infinite precision, such systems may only be predicted a finite length into the future.

This paper presents a novel ANN method, and an application of this method, to find the embedding function $H()$ for the data set presented in the previous section.

3. The Finite Impulse Response Hebbian Model.

The architecture of the ANN⁵ used for time series prediction, in this experiment, is shown in Figure 3. Input data is fed in from the left of the diagram and in successive time steps is passed onto the further neurons as shown. At time t , $y_i(t) = f_i(x_t)$, for $1 \leq i \leq n$, where n is the number of inputs to the network and the actual values of the inputs are a function of the observed time series.

The feedback lines signify trainable connections; there are no self connections i.e. from a set of inputs to the equivalent y -neuron. The equations governing the activation transfer and learning in the network are given by

$$y_i(t) = x_{i0} - \sum_{j=1, j \neq i}^m \sum_{k=1}^d w_{ijk} x_{jk}$$

where x_{jk} is the value of the j^{th} input at time $(t-k)$, and w_{ijk} is the weight from the i^{th} neuron to this input. The parameter d measures the length of the embedding dimension and m is the number of neurons in parallel in each embedded layer.

Learning is achieved by simple Hebbian learning with momentum:

$$\Delta w_{ijk}(t) = \eta (y_i x_{jk}) + (1 - \eta) \Delta w_{ijk}(t-1)$$

where η is a learning rate which may be decreased during the course of the simulation, and η is a parameter which determines the magnitude of the momentum. The negative feedback in the network ensures that the network

does not suffer from the usual Hebbian problem of weights growing without bound. Now since the expected value of $\Delta w_{ijk} = 0$ only when y_i and x_{jk} are decorrelated, y_i and x_{j0} are consecutively decorrelated and then y_i and x_{j1} etc. (similarly with respect to the time-delayed values of the other inputs). At convergence an approximation to independence exists between the neuron's outputs and the input data stream. To extract the value of the embedding function, $H()$, from the network's weights it is required to create a basis of the function space using predefined $f()$ functions. Several experiments have been done using polynomials and sinusoidal functions. The best results have been found using a function of the type:

$$y_1(t) = x_t \text{ and}$$

$$y_n(t) = \sin((n-1) * x_t) \quad " n > 1$$

So

$$y_1(t) = x_{10} = data(t) \text{ and}$$

$$y_n(t) = x_n = \sin((n-1) * data(t)) \quad " n > 1$$

Thus $x_{1k} = data(t-k), \forall k$ and $x_{nk} = \sin((n-1) * data(t-k)), \forall k$ and $\forall n > 1$

We have previously described this network as a Finite Impulse Response (FIR) network, in which each neuron sees a weighted sum of the inputs but this is not fed back to the other neurons.

4. Application of the FIR model to the Oceanographic Data Set.

The aim of this experiment is to extract a model of the structure of the thermal data set plotted in Figure 1. A window of 500km was used to train the NN and then used to forecast the temperature at 5km, 10km, 15km and 20km ahead.

This window was moved from the beginning of the data set to the end in 2 km steps. Every time that the window was moved, the NN was retrained and the forecast performed. The NN weights were randomly initialised, with a value between 0.01 and -0.01 prior to training.

Once the NN was trained for a sufficient number of iterations of the 500 km data set, the value of the temperature 5 kilometres ahead was forecast and this value was fed back and used to forecast the value of the temperature 10 km ahead, and so on up to 20 km ahead. Initially a NN with 4 layers of embedding ($d=4$) and 4 neurons in parallel ($m=4$) was used. The value of the learning constant was 0.01 and the momentum value was 0.5. The NN was trained for 1000 iteration of the 500km data set or until the average error in the training was smaller than 0.05. The errors or residuals of the forecasting can be seen in Figure 5(a, b, c and d). Figure 4 shows that the values of the weights, w_{ijk} , stabilise after the ANN has been trained for 1900 kilometres and then the ANN

does not seem to respond to changes until the Falkland Islands Front (-48,-56) is reached.

Since the aim of this experiment is to extract a general model of the whole data set, it was crucial to ensure that the network did not over-learn any particular section of data and thus memorise the local characteristics of the data in the training set. As can be seen from Figure 4, the absolute value of the means of w_{11} (1.95), w_{12} (1.04), w_{21} (0.40) and w_{22} (0.54) are much higher than the absolute value of the other weights which are always below 0.1 except for w_{14} and w_{32} , whose means are 0.21 and 0.20. From these results it would be easy to conclude that the data set presented in Figure 1 can be modelled by a polynomial based on the functions $y_1(t) = data(t)$ and $y_1(t) = \sin(data(t))$ for the two first levels of embedding, but if only these functions are used the error of the forecast (for 5 km) increases by a factor of 2.5. Data presented in Figure 1 can be modelling by this ANN. Comparing Figures 5(a to d) with Figure 1, it is observed that the higher errors in the forecast are exactly in areas where the temperature changes more. So the error in the forecast is directly related to the magnitude of the change in temperature.

A second FIR network has been reviewed for comparison purposes. In this case a NN with 8 embedded layers ($d=8$) and 8 neurons in parallel per layer ($m=8$) was used. Again the learning constant was 0.01 and the momentum was 0.5. Figure 6 shows the evolution on the weights, and the errors in the forecast are shown in Figure 7(a to d). The training time was longer (2000 iterations of the

500 km training vector or until the average error in the training was smaller than 0.02). Figure 6 shows how the ANN responds to all the fronts by adjusting its weights in frontal areas and stabilising them in areas where the structure of the water maintain its properties. Looking at the changes in the weights it is possible to determine the limits of the water masses. Fronts between water masses are identifiable by changes in the weights (compare Figure 1 and Figure 6).

In this case, as can be seen by comparing Figure 7(a to d) and Figure 5(a to d), both the error in the forecasting and its Standard Deviation are higher for the 8*8 ANN than for the 4*4 ANN. This increment in the error may suggest that increasing the number of neurons in the ANN does not improve its performance, but when the forecast is analysed on a case by case basis from a qualitative point of view, it can be observed that the 8*8 ANN reacts to changes before the 4*4 ANN. It reacts between 1 and 2 km before the 4*4 ANN to major changes in the temperature; in other words, the 8*8 ANN gives warning of a change in the behaviour of the water mass before the 4*4 ANN. This reaction has the disadvantage that the 8*8 ANN overreacts to changes and misleads the forecasting. This is an instance of the well known bias-variance trade off⁶. If the network has too many weights, it will tend to model the noise as well as the underlying structure of the data set. This increases the variance of the errors from such a network. If it has too few weights, it will not be powerful enough to model the structure of the data set. This increases the errors and the forecast will be perceived to have a bias.

So even though the averaged error in the forecasting increases, the ANN improves its ability to predict major changes. Thus this 8*8 ANN can complement the 4*4 ANN because the first performs better in areas with big changes in the water temperature, where the second gives the higher errors in the forecast. With the 8*8 ANN the bigger errors occur after a big change in the temperature has been forecast, because it over reacts to these changes due to the excess of neurons.

5. Linear Regression Forecasting.

Linear Regression is a method that can be used for forecasting⁷. In this particular case it is only used as a means of comparing its performance with the results obtained using the FIR model. Surprisingly, no method has been developed for forecasting this type of data, except the ORKA system⁸. The ORKA system is based on a Linear Regression Algorithm for forecasting up to 5 km ahead of an ongoing submarine. Since results of this system have not been published, it was decided to use a Linear Regression for comparison.

In this experiment a prediction of the value of the temperature x at point y can be obtained using the vector of temperatures (dependent variable) recorded between y_{-1} and y_n (independent variable), where $n-1$ is the number of kilometres of temperature used to carry out the regression.

The equation used for the forecast is $y=a+bx$, where:

$$a = \bar{Y} - b\bar{X}$$

and:

$$b = \frac{n\sum xy - (\sum x) - (\sum y)}{n\sum x^2 - (\sum x)^2}$$

This method has been used to calculate the value of the temperature at 5, 10, 15 and 20km. In each cases, the optimum dimension for the vector used to carry out the regression has been chosen empirically. In this case, the optimum number of temperature values and distance are those that give the smallest error in the forecasting. Table 2 shows the results obtained with the Linear Regression:

As is shown in Figure 8(a, b, c and d) the absolute value of the error of the forecast is higher using the Linear Regression than the ANN. Figure 9 shows how the Linear Regression tends to react to changes a few km after the real change occurs.

The error in the forecast of the Linear Regression is also smother than the one obtained with the ANNs because this is a much more conservative method. The changes in the absolute value of the error obtained with the Linear Regression are very smooth, therefore even though the average value of the errors is higher than those obtained with the FIR ANN, the Standard Deviation of the error is less than that of the FIR methods when forecasting at 10 and 15 km.

6. ARIMA model.

Modeling and forecasting involve knowledge about the mathematical model of the processes. The ARIMA methodology was developed by Box and Jenkins⁹ and it allows the modelling and forecasting of complex data sets. It has gained enormous popularity in many areas and research practice¹⁰ confirms its power and flexibility.

To find the ARIMA parameters a stationary data set is required as input, the data set presented in Figure 1 needs to be differentiated in order to obtain a stationary data set. The average forecasting errors and the standard error of the forecast using the ARIMA(1,1,0) model are shown in Table 3. The shape of the evolution of the error is very similar to the one obtained with the linear regression.

The high heteroscedasticity and multicollinearity of the data makes very difficult to obtain an accurate global statistical forecasting system even using ARIMA models.

7. Conclusions

The FIR model can be used (i) for modelling this system, (ii) as a method to forecast with reasonable accuracy (FIR 4*4) and (iii) for detecting changes accurately (FIR 8*8), in the behaviour of the thermal structure of the data set shown in Figure 1. Each ANN complements each other; they both forecast better than the method of Linear Regression. The FIR ANN is capable of

adapting its weights to the spatial and temporal variations of the continuous time series presented to it.

The FIR ANN are capable of accurately forecasting whether the temperature of the water 5 km ahead of the vessel is going to ascend or descend more than 90% of the times for the time series presented in Figure 1. This percentage is reduced with the distance ahead of the forecast, so that when forecasting the temperature of the water 20 km ahead the percentage is 65%.

Research is being carried out in order to study if different basis functions provide more accurate prediction. Different methods to automate the process of defining the optimum number of neurons are also under investigation.

References

1. Tomczak, M. & Godfrey, J. S. Regional Oceanography: An Introduction. Pergamon. Nueva York. 1994. ISBN: 0080410219.
2. Corchado, J., Fyfe, C., Lees, B. & García, R. Multi Agent Adaptive Architecture for Forecasting. SOCO'97. 1997. 94-100.
3. Venugopal V. & Baets W. Marketing Intelligence and Planning. Vol 122 No 7. 1994. 30-38.
4. Weigend A. S. & Gershenfeld N. A., Time Series Prediction: Forecasting the Future and Understanding the Past. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wiley, 1995.

5. Fyfe, C. An Unsupervised Neural Method for Time Series Analysis, Characterisation and Prediction. ICANNGA. 1997. 464-467.
6. Geman, S. Bienenstock, E. & Doursat, R.. Neural Networks and the Bias/Variance Dilemma. Neural Computation. 1992. Vol 4, No 2. 1-58.
7. Pankratz, A. Forecasting with Dynamic Regression Models. Wiley - Interscience Publication, 1983.
8. Lees, B., Aiken, J. & Rees, N., Knowledge based oceanographic data analysis. Expertsys-92, 1992, 561-566.
9. Box, G. E. P. & Jenkins, G. M., Time Series Analysis: Forecasting and Control. Rev. ed., San Francisco, 1976, Holden-Day.
10. Vandaele W., Applied Time Series and Box-Jenkins Models. New York, 1983, Academic Press.