NONLINEAR MODELLING OF AIR POLLUTION TIME SERIES

Rob Foxall[1]*, Igor Krcmar*[2]*, Gavin Cawley*[1]*, Steve Dorling*[3]*, and Danilo P. Mandic*[1]

1. School of Information Systems, University of East Anglia, Norwich, UK.
email: {rjf,gcc,mandic}@sys.uea.ac.uk
2. Department of Electrical Engineering, University of Banjaluka, Banjaluka, Bosnia–Herzegovina.
email: ikrcmar@etf-bl.rstel.net
3. School of Environmental Sciences, University of East Anglia, Norwich, UK.
email: s.dorling@uea.ac.uk

ABSTRACT

An analysis of predictability of a nonlinear and nonstationary ozone time series is provided. For rigour, the DVS analysis is first undertaken to detect and measure inherent nonlinearity of the data. Based upon this, neural and linear adaptive predictors are compared on this time series for various filter orders, hence indicating the embedding dimension. Simulation results confirm the analysis and show that for this class of air pollution data, neural, especially recurrent neural predictors, perform best.

1. INTRODUCTION

Air pollutants such as surface ozone (O_3) (shown in Figure 1) are the subject of an ongoing analysis. A rigorous analysis of such data requires consideration of a number of meteorological variables (e.g. wind speed) and non-meteorological variables (e.g. traffic density). To obtain an insight into the underlying structure, however, it is worthwhile to look initially at each pollutant time series individually with the standard linear methods, and to do nonlinearity analysis only if it appears that a linear model is inadequate. The first stage in modelling a complex process suspected of being nonlinear is detection of the nonlinearity, otherwise it cannot be known that a linear model would not suffice. Once the nonlinearity is demonstrated, the next stage is to chose the "best" nonlinear model, which is chosen from a finite class of nonlinear models.

Here, the method of Deterministic Versus Stochastic (DVS) plots is used to investigate the nonlinearity of an ozone pollutant time series, and then based upon this linear and neural adaptive models are compared on this ozone series.

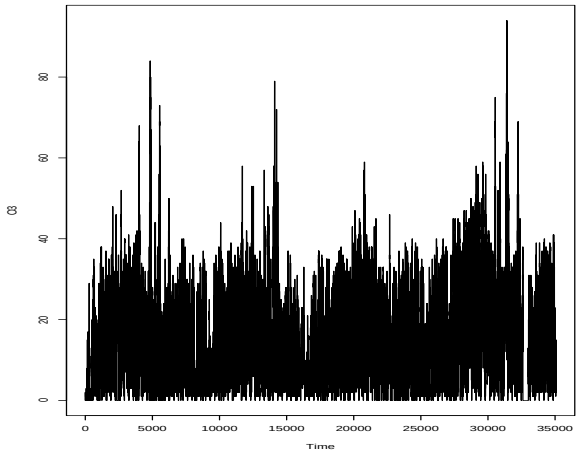

Fig. 1. O_3 time series plot. Time is measured in hours.

2. LINEAR ANALYSIS OF TIME SERIES

An initial analysis of stationary linear time series often involves fitting an ARMA(p, q) model [1]:

$$x(n) = \sum_{i=1}^{p} a_i x(n-i) + \sum_{i=1}^{q} b_i \epsilon(n-i) + \epsilon(n) \qquad (1)$$

with x being the time series, a and b the parameters of the model to be fitted, and ϵ the random noise. For the stationarity requirement of an ARMA model it is sometimes useful to deseasonalise a time series. For the O_3 time series there is evidence of two seasonal components with periods one day and one year respectively.

The obvious extension of the ARMA model to a nonlinear model (the NARMA model) is

$$\begin{aligned} x(n) &= h(x(n-1), \ldots, x(n-p), \\ &\quad \epsilon(n-1), \ldots, \epsilon(n-q)) + \epsilon(n). \end{aligned} \qquad (2)$$

The function h is assumed to be nonlinear and differentiable.

3. THE METHOD OF DVS PLOTS

The general DVS algorithm presented in [2] fits the model

$$x(n) = h\big(x(n-1), x(n-1-\tau), \ldots, x(n-1-(m-1)\tau)\big) + \epsilon(n) \quad (3)$$

i.e. the NAR part of equation (2) when $\tau = 1$. The variable m is the embedding dimension, τ is the lag time. Following the arguments in [2], for which h is a linear function, we have chosen $\tau = 1$ since hourly measurements represent coarse sampling for pollution time series.

The structure of the DVS algorithm is as follows: the time series is firstly divided up into a *training* set and a *test* set. For a given m, from the training set construct the *delay vectors*

$$\mathbf{x}(n) = [x(n-1), x(n-2), \ldots, x(n-m)].$$

For each delay vector calculate its Euclidean distance from all the other delay vectors, then choose the k nearest neighbours to fit the best linear model from equation (3). This fitted model is used for prediction purposes on the test set, and the prediction error calculated. Finally, the number of nearest neighbours k is varied over representative values up to the number of delay vectors in the training set. The DVS plot gives the prediction error for the test set versus k. The value of k which gives the optimal prediction error indicates the nonlinearity of the time series. If the optimal k is at or close to the total number of delay vectors in the training set, then globally linear models perform best and there is no indication of nonlinearity. In this case the model is equivalent to an AR model of order m when $\tau = 1$. Small or intermediate optimal k suggests local linear models perform best, indicating nonlinearity and/or chaotic behaviour. For rigour, a high-order AR model was fitted to the O_3 series and used to simulate a linear series with similar properties [3]. This was done for both the raw and the deseasonalised series, giving four DVS plots in total. For each plot, the final 500 elements of the series were used to construct the test set. In Figure 2 the DVS plot for $m = \{2, 4, 6, 8, 10\}$ is shown. For each value of m, the optimal k is less than the maximum, but the difference in the prediction error is minimal. Figure 3 displays the equivalent DVS plot for a surrogate dataset simulated from an $AR(45)$ model fit to the series. The behaviour for the surrogate data is similar to the original data, but clearly has the optimal k at $k_{\max}$ suggesting at least a certain degree of nonlinearity.

The deseasonalised DVS plots, (omitted due to space constraints), corroborate with the seasonal DVS plots, although a higher embedding dimension is indicated. We now investigate the quality of prediction of this time series employing neural algorithms.

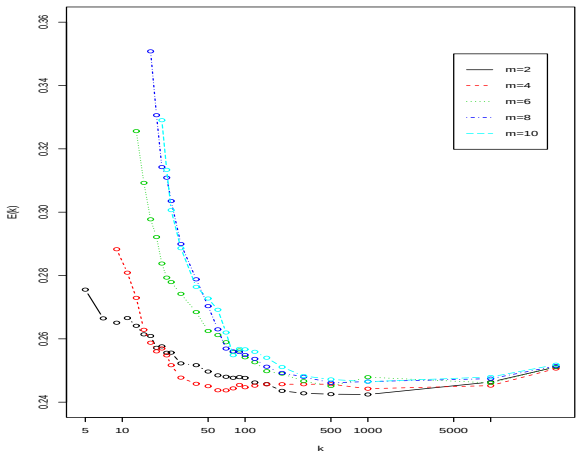

Fig. 2. DVS plot for the O_3 time series with lag delay $\tau = 1$.

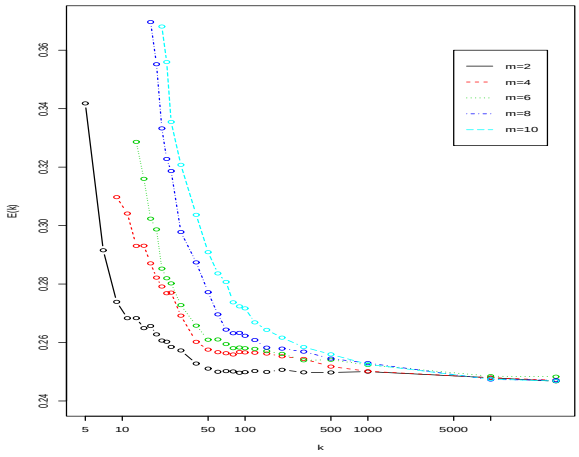

Fig. 3. DVS plot for the simulated O_3 time series with lag delay $\tau = 1$.

4. NEURAL ADAPTIVE FILTERS IN THE AIR POLLUTION TIME SERIES PREDICTION

The DVS plots indicate a certain degree of nonlinearity of the air pollution time series. Thus, in order to obtain good prediction of the future value of the time series at hand, based on the past measurements, an efficient algorithm should be employed, i.e. an algorithm that is inherently nonlinear and/or adaptive. Gradient-descent (GD) based neural adaptive filters, due to their simplicity and nonlinearity are an adequate choice for the prediction of time series that represents atmospheric pollution data. Furthermore, the structure of neural adaptive filters could be chosen to reflect the nature of the underlying process, i.e. it could be feedforward or recurrent.

The operation of a single-neuron neural adaptive filter can be described as

$$\begin{aligned} y(k) &= \Phi(v(k)) \\ v(k) &= \mathbf{w}^T(k)\mathbf{u}(k) \end{aligned} \quad (4)$$

where k denotes a discrete time instant, $y(k)$ is the output of the filter, $v(k)$ is the neuron activation function, $\mathbf{w}(k) = [w_1(k), ..., w_N(k)]^T$ is the weight vector, $(\cdot)^T$ denotes the vector transpose and $\Phi(\cdot)$ represents a nonlinear activation function of a neuron. Definition of the vector $\mathbf{u}(k)$ depends on the structure of a neural adaptive filter. In the case of the feedforward filter $\mathbf{u}(k)$ contains only samples of the input signal $x(k)$, and is defined as $\mathbf{u}(k) = [x(k-1), ..., x(k-N)]^T$, where N denotes the order of the filter. Such a neural adaptive filter is of the finite impulse response (FIR) type. On the other hand, in the case of the recurrent filter, it contains samples of the input signal $x(k)$, as well as samples of the delayed output of the filter $y(k)$, i.e. input vectors defined as $\mathbf{u}(k) = [y(k-1), \ldots, y(k-q), 1, x(k-1), ..., x(k-p)]^T$, where unity stands for bias term. This recurrent filter is known as a nonlinear ARMA(p, q) recurrent perceptron, where p and q denote respectively the order of autoregressive (AR) and moving average (MA) part of the filter.

The adaptation of a GD based neural adaptive filter can be described by the following set of equations

$$e(k) = d(k) - \Phi(v(k)) \quad (5)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \eta \nabla_{\mathbf{w}} E(e(k)) \quad (6)$$

where $d(k)$ is some training (desired) signal, $e(k)$ is the instantaneous error at the output neuron, $E(\cdot)$ is the filter cost function, and η denotes the learning rate parameter. The most common choice for the cost function $E(\cdot)$ is

$$E(e(k)) = \frac{1}{2} e^2(k). \quad (7)$$

Obviously, computation of the gradient of the cost function, denoted by $\nabla_{\mathbf{w}} E(e(k))$, depends on the structure of a neural adaptive filter. For the feedforward type of a filter, this gradient is given by

$$\begin{aligned} \nabla_{\mathbf{w}} E(e(k)) &= e(k)\mathbf{u}(k) \\ &= e(k)[x(k-1), ..., x(k-N)]^T. \end{aligned} \quad (8)$$

The algorithm described by equations (4) - (8) is usually referred to as the nonlinear gradient-descent (NGD) algorithm. The gradient of the cost function for a nonlinear ARMA(p, q) recurrent perceptron is defined as

$$\nabla_{\mathbf{w}} E(e(k)) = e(k)\mathbf{\Pi}(k) \quad (9)$$

where $\mathbf{\Pi}(k) = \left[\frac{\partial y(k)}{\partial w_1(k)}, \ldots, \frac{\partial y(k)}{\partial w_N(k)}\right]$ represents the gradient at the output of the neuron. The procedure of gradient computation according to equation (9) is more complex than in the case of a feedforward neural adaptive filter, due to functional dependence between $w_i, i = 1, \ldots, N$ and $y(k-j), j = 1, 2, \ldots$. Hence, gradient computation is based upon the approximation [4] that for small learning rate η, the following relationship holds

$$\frac{\partial y(k-j)}{\partial w_i(k)} \approx \frac{\partial y(k-j)}{\partial w_i(k-j)}, j = 1, 2, \ldots, q. \quad (10)$$

It is important to note that the class of algorithms described by the equations (4) - (7), with fixed learning rate η, might suffer from slow convergence and local minima. A constant learning rate parameter η can be considered as one of the factors contributing to these problems [4, 5]. Equation (7) gives an instantaneous estimate of the ensemble average $\langle e^2(k) \rangle$, thus introducing a gradient noise in the operation of an algorithm [6]. This gradient noise will help an algorithm to escape from local minima, but unfortunately will also reduce convergence rate. Further, a recent result [7] indicates an inherent relationship between the learning rate parameter η and slope of the nonlinear activation function of an output neuron β, which also has a negative impact on the convergence properties of the algorithm with fixed η.

A successful design of an algorithm for adaptation of GD based, single-neuron neural adaptive filter is given in [8]. The normalised nonlinear gradient-descent (NNGD) algorithm exhibits optimal behaviour in the sense that it minimises the instantaneous prediction error, thus providing adaptive learning rate η. In the case of the linear activation function of an output neuron, the NNGD algorithm reduces to the normalised least mean squares (NLMS) algorithm.

5. EXPERIMENTAL RESULTS

To acquire the true nature of the ozone time series we performed several experiments. Air pollution data represent hourly measurements of the concentration of ozone (O_3), in the period 1994 - 1997. These data were provided to us, courtesy of the Leeds meteo station.

In the performed experiments the logistic function was chosen as the nonlinear activation function of an output neuron. The quantitative performance measure was the standard prediction gain, a logarithmic ratio between the expected signal and error variances $R_p = 10 \log(\hat{\sigma}_s^2 / \hat{\sigma}_e^2)$. The slope of the nonlinear activation function of the neuron β was set to be $\beta = 4$, since this value makes Φ close to the linear function in the vicinity of the origin. Due to the saturation type of the logistic nonlinearity, input data were prescaled to match the range of an output neuron activation function.

In the first experiment we compared the performance of the NLMS, NGD, and NNGD algorithms in the prediction of the O_3 time series. The learning rate parameter η in the NGD algorithm, was set to be $\eta = 0.3$, and the constant C in the NNGD algorithm, was set to be $C = 0.2$.

The order of the feedforward filter N varied in the range $N = 1, 2, \ldots, 30$. A summary of the performed experiment is given in Figure 4. In the second experiment we

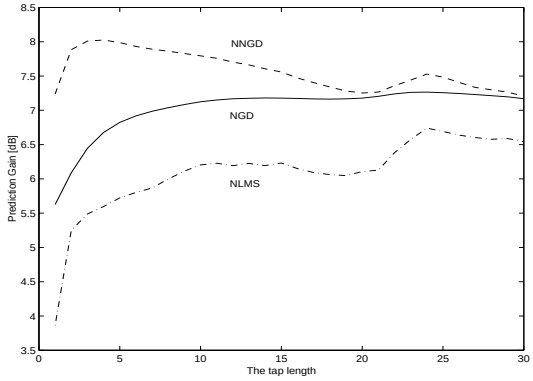

Fig. 4. Performance of the NLMS, NGD, and NNGD algorithms in the prediction of the O_3 time series

investigated the performance of the ARMA recurrent perceptron in the prediction of the O_3 time series. The order of the MA part q, and the AR part p of the nonlinear ARMA recurrent perceptron varies in the range $p, q = 1, 2, \ldots, 10$. A summary of results of the experiment is shown in Figure 5.

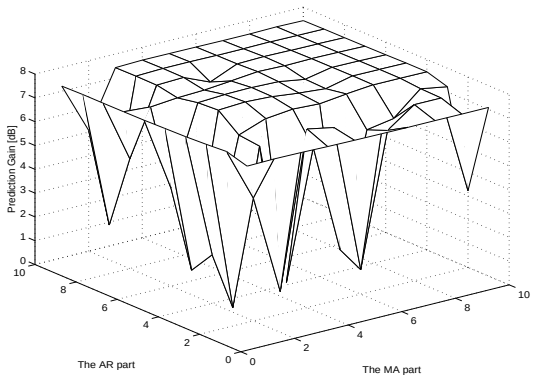

Fig. 5. Performance of the NARMA recurrent perceptron in the prediction of the O_3 time series

It is obvious that nonlinear NGD and NNGD algorithms for adaptation of a neural adaptive filter have better performance comparing to the linear algorithm (NLMS). A recurrent perceptron outperforms the NLMS, NGD, and NNGD algorithms for a wide range of p and q, suggesting the recursive nature of the generating process of ozone [9]. This also corroborates with the DVS plots shown before.

6. CONCLUSION

An analysis of predictability of the ozone time series has been performed. This has been done rigorously, starting from the detection of nonlinearity and embedding dimension through to the best NARMA(p, q) recurrent predictor. Simulation results confirm that this class of time series exhibits inherent nonlinearity and that a recurrent NARMA(p, q) perceptron is the best model among the single-neuron models to describe this time series.

7. ACKNOWLEDGEMENTS

This work was supported by the European Commission, grant number IST-99-11764, as part of its Framework V IST programme.

8. REFERENCES

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden-Day, Inc., 1970.

[2] M. C. Casdagli and A. S. Weigend, *Time Series Prediction : Forecasting the Future and Understanding the Past*, ch. Exploring the Continuum Between Deterministic and Stochastic Modeling, pp. 347–366. Addison Wesley, 1993.

[3] J. Theiler, P. S. Linsay, and D. M. Rubin, *Time Series Prediction : Forecasting the Future and Understanding the Past*, ch. Detecting Nonlinearity in Data with Long Coherence Times, pp. 429–455. Addison Wesley, 1993.

[4] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New Jersey: Prentice Hall, 1994.

[5] C. Darken and J. Moody, "Towards faster stochastic gradient search," *Neural Information Processing Systems*, vol. 4, pp. 1009–1016, 1992.

[6] P. E. An, M. Brown, and C. J. Harris, "Aspects of instantaneous on-line learning rules," in *Proceedings of Control '94*, pp. 646–651, 1994.

[7] D. P. Mandic and J. Chambers, "Relationship between the slope of the activation function and the learning rate for the rnn," *Neural Computation*, vol. 11, no. 5, pp. 1069–1077, 2000.

[8] D. P. Mandic, "NNGD algorithm for neural adaptive filters," *Electronic Letters*, vol. 36, no. 9, pp. 845–846, 2000.

[9] G. Nunnari, A. Nucifora, and C. Randieri, "The application of neural techniques to the modelling of time-series of atmospheric pollution data," *Ecological Modelling*, vol. 111, pp. 187–205, 1998.