RBF AND WAVELET NETWORKS FOR SIMULATION AND PREDICTION OF SULPHUR-DIOXIDE AIR POLLUTION

Vladimír Sýs, Aleš Procházka

Prague University of Chemical Technology
Department of Computing and Control Engineering
Technická 1905, 166 28 Prague 6, Czech Republic
E-mail: Vladimir.Sys@vscht.cz, Ales.Prochazka@vscht.cz
Phone: 0042-2-2435 4198 ∗ Fax: 0042-2-243 11082

ABSTRACT

The paper presents selected problems of the air pollution simulation and prediction using multichannel observations and multirate modelling. Theoretical part of the paper is devoted to time-series analysis based upon the short-time Fourier transform and the use of neural networks for their prediction. The main problem discussed in the paper is connected with the application of various transfer functions including radial basis functions (RBF) and wavelet functions and optimization of such a network by the modified genetic algorithm. All results are presented for the air pollution prediction of the Prague region.

INTRODUCTION

Connectionist or neural network algorithms became a very popular tool for system simulation tasks during the past decade. They are able to tackle highly complicated nonlinear problems. On the other hand they require an enormous number of parameters to be set. Moreover a careful preprocessing of data must precede the simulation itself.

An estimation of future concentration of different atmosphere gases has been a subject of simulation work since the seventies including studies of (9) as well. A pollution of the air caused by augmented sulphur-dioxide exhalations (power stations, individual heating systems, cars) has reached a dangerous level, especially in some countries of the Central Europe. It raises problems in every-day life, e.g. city traffic and energy consumption regulations in crisis situations. Knowing more about at least few-hour evolution of the sulphur-dioxide air concentrations is of public concern and interest.

DATA PREPROCESSING

The data available for this preliminary study consist of twelve one-year (1993) time series measured in twelve different locations in Prague (see Fig. 1). Sampling period has been set to 30 minutes. The basic concept of the system is to be given by spectrum estimation where typical periods are clearly seen however can hardly be explained, except of 24 hours cycle.

The low-pass FIR filtering has been employed in order to reduce the level of a noise. The series has been significantly subsampled along with the spectrum properties. The short-time Fourier transform of a selected data segment is presented in Fig. 2.

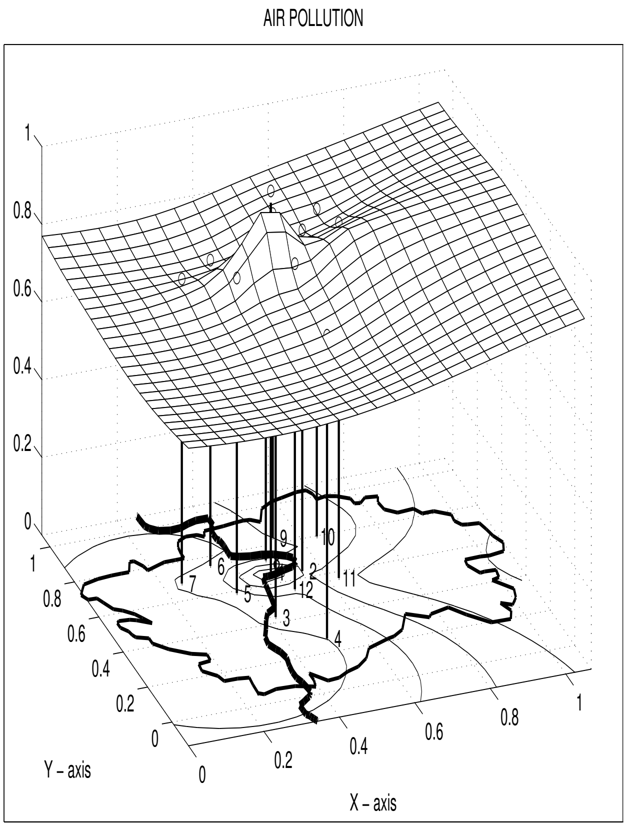

Figure 1: Air pollution monitored on March 11th, 1993 at 6 a.m. at different parts of Prague and its interpolation to the whole region

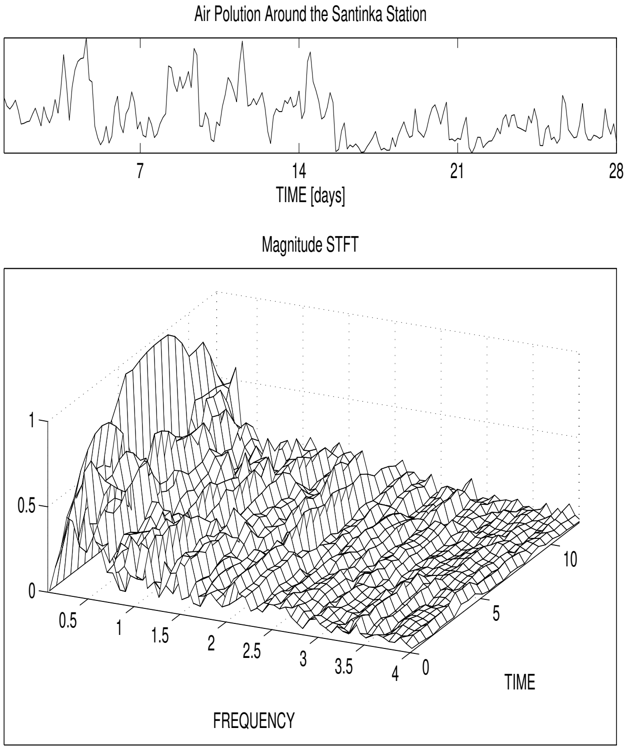

Figure 2: A given signal part of air pollution monitored around the Santinka station from 1.3. to 28.3.1993 decimated from original 30 minutes sampling period to 3 hours and its short-time Fourier transform of 16 days long segments providing information about the main periodic components 1 day and 1 week long including their higher harmonics

The preprocessed data has been tested to discover the nature of the original system. For this task many algorithms evaluating system dimensions exist but their application seems to be very time consuming and not giving always reliable results. In this work a new method has been implemented based upon studies of (1), (20) to decide whether this was the case of a low dimensional (deterministic) chaos or a high dimensional (stochastic) system.

TIME SERIES FORECASTING

Time-series forecasting can be based on finding a function f which maps a known data pattern $\mathbf{x}$ onto a predicted pattern $\mathbf{y}$: $f : \Re^n \rightarrow \Re^m$. The task is to optimize the function f in the sense of the maximum prediction reliability and its accuracy.

When the simulated system presents nonlinear features the function f can be a rather complicated one. Connectionist structures mentioned e.g. (8) have been applied in those cases because they often enable to model a complex dynamics of the studied system. In connectionist terminology the data pattern is identical to the input vector and predicted values corresponds to the output vector. The structure of the network itself represents the mapping f. The mapping optimization is carried out by the network training on known input-output pairs.

RBF NETWORKS

RBF networks presented e.g. by (14), (3) have been given their mathematical background from the multivariable function approximation by radial basis functions (RBF). Besides the multi-layer perceptron described e.g. by (18) they belong to favourite and the most powerful paradigmas.

Generally, the RBF means a function whose argument depends on the distance $r = \|\mathbf{x} - \mathbf{c}\|$ e.g. Euclidean measure between the input vector $\mathbf{x}$ and the centre $\mathbf{c}$. In the simplest case the predictive function f is then represented by the weighted sum of fixed RBFs Φ_i, i.e.

$$f(\mathbf{x}) = \sum_{i=1}^{N} w_i \Phi_i(\mathbf{x}) \tag{1}$$

Unfortunately, the number N of required RBFs grows exponentially with the dimension of the approximated input output mapping. To avoid this problem the RBFs with adjustable parameters are introduced. As a typical choice the Gaussian function is taken

$$\Phi(r) = exp(-r^2/\sigma^2) \text{ where } r = \|\mathbf{x} - \mathbf{c}\| \tag{2}$$

Its peak lies in the centre $\mathbf{c}$ and the value decreases monotonically with the distance r.

The described structure can be easily implemented in the three-layer network structure presented in Fig. 3. Therefore three sorts of parameters are to be set: the weights $w_i \in \Re$, the centres $\mathbf{c}_i \in \Re^n$ where $n = dim(\mathbf{x})$, and the scaling or width parameters $\sigma_i \in \Re$, respectively. The $\mathbf{c}_i$ can be determined as the centres of learning input vector clusters and the σ^2 as a dispersion of data points inside of each cluster. The linear weights w_i are calculated by e.g. LMS method.

WAVELET NETWORKS

A theoretical work devoted to so called wavelet functions presented e.g. by (5), (19) has allowed the modification of RBF network named wavelet network or wavenet (24), (17).

The set of continuous wavelet functions is usually derived from the initial (mother, basis) wavelet $h(t)$ which is dilated by value $a = 2^m$, translated by constant $b = k\ 2^m$ and normalized so that

$$h_{m,k}(t) = \frac{1}{\sqrt{a}}\, h(\frac{t-b}{a}) = \frac{1}{\sqrt{2^m}}\, h(2^{-m}\ t - k) \tag{3}$$

where m and k are integers. Initial wavelet can be defined either by the solution of a dilation equation

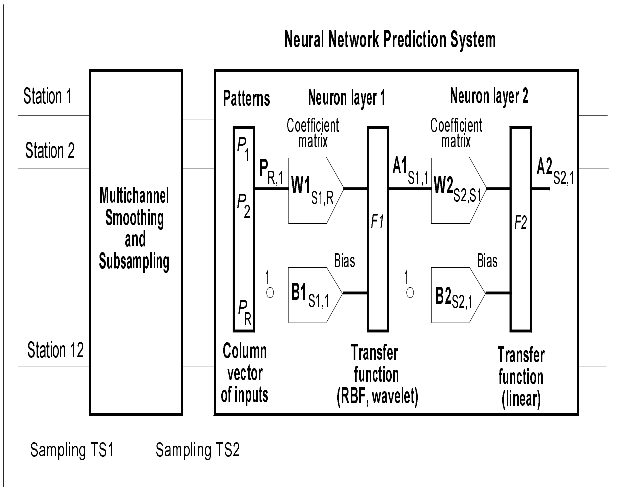

Figure 3: Multirate system for multichannel air pollution prediction using a neural network having output vector $\mathbf{A2} = F2(\mathbf{W2} * \mathbf{A1} + \mathbf{B2})$ where $\mathbf{A1} = F1(\mathbf{W1} * \mathbf{P} + \mathbf{B1})$ and the activation function $F1$ in the first layer chosen either as radial basis or wavelet functions and a linear output layer function $F1$

(5) or in an analytical form presented both by (5) and (13).

Using the digital signal processing point of view it is possible to consider the initial wavelet as a pass-band filter and in most cases half-band filter covering frequency band $\langle 0.25, 0.5 \rangle$. A wavelet dilation by the factor $a = 2^m$ corresponds to a pass-band compression. This feature is very useful one in system modelling and signal approximation providing fundamental idea of their use as transfer functions in neural networks as an alternative to sigmoidal and radial basis functions.

The basic application of wavelet functions for a given vector $\{x(n)\}_{n=0}^{N-1}$ assumes a half-band low-pass scaling sequence

$$\{l(n)\}_{n=0}^{L-1} = [l(0), l(1), l(2), \cdots, l(L-1)]$$

of an even length L and the complementary orthogonal wavelet sequence

$$\{h(n)\}_{n=0}^{L-1} = [l(L-1), -l(L-2), l(L-3), \cdots, -l(0)]$$

This methods enables signal subsampling by two in each stage and subsequent application of the same method at further scales.

Wavelet function defined in the analytical form are more suitable for their application in neural networks as they allow more convenient coefficients optimization. There are various such functions described in literature including harmonic wavelet functions introduced by (13) or Gaussian functions used by (5).

SYSTEM OPTIMIZATION

The training of the network exhibits always a major obstacle in introduction of connectionist models to real applications. As a classical back-propagation (18) method has failed here a rather novel approach has been tested, namely the genetic algorithms.

The genetic algorithms provide an effective tool for the robust multi-parameter optimization. A principal advantage seems to be an insensitivity to local minimas of error surface during the search for ideal values in the parameter space. However the algorithm imposes higher requirements on the computer memory and the machine time.

The principle of genetic algorithms (GA) is generally known and widely studied and used. The most famous book on the subject seems to be the Goldberg's book (7) and there is also a lot of papers devoted to the optimization of feedforward neural networks published recently e.g. (11), (23), (21), (12), (4). We concentrate on implementation concepts here and on a brief discussion of setting parameters for the learning process.

The given signal modelling with the feedforward network and activation functions described in the previous sections can be considered as an multi-dimensional optimization problem. The hidden layer parameters evolve through the GA and provide a solution of the approximation task with a certain accuracy and corresponds to an "individual" according to the GA terminology. Particular steps of the algorithm are as follows:

1. Generation of an initial population of individuals which means that several network sets of coefficients are chosen at random in given ranges
2. Parameters coding into binary strings of a predefined length in order to allow that genetic operators might be applied
3. Evaluation of each individual's fitness
4. Probabilistic selection of individuals for the next generation according to their fitness values preferring "better" individuals
5. Random coefficients pairs selection to mate each other changing randomly their strings segments (crossover)
6. String mutation turning over values of every bit in strings with relatively small probability
7. Repeating the loop until the required accuracy is reached

The encoding is viewed as problem which resists to a wider application of GAs. Usually either binary or Gray code is employed. However we have not noted any dramatic change of the performance between these two codings.

The occurrence probabilities of the genetic operators seem to be critical for the efficiency of the learning process. Studies of (22) propose to increase the mutation probability in time to sustain a population diversity

TABLE 1: THE LIST OF MEASURING SPOTS IN THE PRAGUE AREA USED FOR AIR POLLUTION ANALYSIS

1	Square of Republic	7	Veleslavín
2	Rieger's Park	8	Exhibition Place
3	Bráník	9	Kobylisy
4	Libuš	10	Vysočany
5	Mlynářka	11	Počernice
6	Santinka	12	Vršovice

while (6) suggests to reduce the mutation rate and to augment the crossover probability in next generations. In (15), (16) three strategies are tested:

- Exponentially decreasing mutation and increasing cross-over probability
- Exponentially increasing mutation and decreasing cross-over probability
- The constant probabilities

These three different approaches cause completely different evolution strategies but the results do not distinct very much. Our experiments do not prove a significantly positive influence of the inversion operator on the learning process quality.

A relatively poor performance of GAs to train the feedforward net caused a more closer look on the purpose of the crossover operator. Unfortunately, our experiments proved Kitano's conclusion (10) that the network training lacks the compositionality property that makes the crossover worthwhile.

The remove of the crossover operator enables to simplify the algorithm while the only mutation is being applied which further allow for rejection of the principal idea consisting of a binary string and its substitution by a real number for each parameter. The mutation is then carried out by adding some random number suitably scaled.

RESULTS

Analysis presented in the paper were applied to 12 measuring stations in the Prague area listed in Tab. 1. Graphical and numerical results are given in Fig. 4 and in Tab. 2 respectively.

Several methods have been tested for simulation. The models have been set on six hundred pattern segment of data. Then the mean square errors for one step prediction of 190 pattern time series have been evaluated. The results are shown in Tab. 2.

CONCLUSION

The results have proven the capability of feedforward networks to treat a highly nonlinear and noised time series. The forecasts are much more reliable when the simulated system occurs in normal conditions. The described methods fail to predict a sudden out-of-range events which is however always a hard task.

TABLE 2: COMPARISON OF MEAN SQUARE ERRORS OF SINGLE SERIES PREDICTION FOR VARIOUS METHODS

Autoregressive	1.351
Back Propagation	0.508
Levenberg-Marquardt	0.511
RBF network	0.514
GA wavenet	0.601
GA wavenet (mutation)	0.396

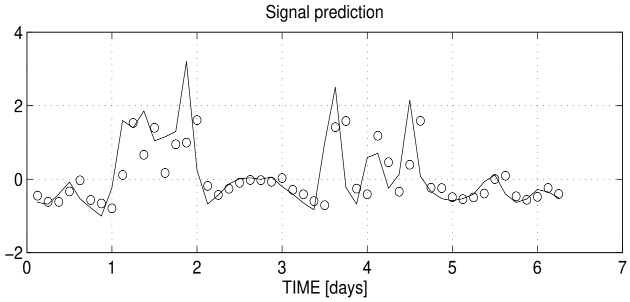

Figure 4: A given one step ahead prediction (standing for 3 hours) based upon the single channel neural network prediction using radial basis transfer functions in the first layer

The best solution has been provided by the wavenet structure optimized with the simplified genetic algorithm which seems to offer a powerful alternative to the standard learning strategies.

Of course, a more serious analysis of data and more profound experimental work on learning of networks would be necessary to decide finally on feasibility of these approaches.

ACKNOWLEDGEMENT

The contribution has been supported by a grant No. 201/94/013 of the Grant Agency of Czech Republic. All real signals were kindly provided by Dr. M. Fiala, Head of the Department of Air Pollution Control of the Prague Institute of Hydrometeorology.

REFERENCES

Casdagli, M. and A.S. Weigend. 1992. "Exploring the Continuum Between Deterministic and Stochastic Modeling". In *Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis*, Santa Fe, New Mexico.

Chen, S.; C. F. N. Cowan; and P.M.Grant. 1994. "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks". *IEEE Trans. on Neural Networks* 2, no. 2: 302–309.

Chng, E. S.; S. Chen; and B. Mulgrew. 1994. "Improving the Radial Basis Function Networks for Homogeneous Nonstationary Time Series Prediction". In *Signal Processing VII*, European Association for Signal Processing.

Chu, C.H. and C.R. Chow. 1993. "A Genetic Algorithm Approach to Supervised Learning for Multilayered Networks". In *Proceedings of World Congress on Neural Networks*, pages IV/744–IV/747, Portland, Oregon.

Daubechies, I. 1990. "The Wavelet Transform, Time-Frequency Localization and Signal Analysis". *IEEE Trans. Inform. Theory* 36: 961–1005.

Etter, D.M. and D.C. Dayton. 1983. "Performance Characteristic of a Genetic Algorithm in Adaptive IIR Filter Design". In H.W.Schussler, editor, *Signal Processing II*. Elsevier Science Publishers B.V. (North-Holland).

Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, Massachusetts.

Hertz, J.; A. Krogh; and R. G. Palmer. 1991. *Introduction to the Theory of Neural Computing*. Addison-Wesley Publishing Company, Redwood City, California.

Karayiannis, N. B. and A. N. Venetsanopoulos. 1993. *Artificial Neural Networks*. Kluwer Academic Publishers, Norwell, MA.

Kitano, H. 1990. "Empirical Studies on the Speed of Convergence of Neural Network Training Using Genetic Algorithms". In *Proc. of the 8th Conf. on AI.*

Maniezzo, V. 1994. "Genetic Evolution of the Topology and Weight Distribution of Neural Networks". *IEEE on Neural Networks* 5, no. 1: 39–53.

McDonnell, J. R. and D. Waagen. 1994. "Evolving Recurrent Perceptrons for Time-Series Modelling". *IEEE Trans. Neural Networks* 5, no. 1: 24–38.

Newland, D. E. 1994. "Wavelet Theory and Applications". In *Proceedings of the Third International Congress on Air- and Structure- Borne Sound and Vibrations*, Montreal, Canada.

Poggio, T. and F. Girosi. 1990. "Networks for Approximation and Learning". In *Proceedings of the IEEE.*

Procházka, A. and V. Sýs. 1994. "Application of Genetic Algorithms for Wavelet Networks Signal Modelling". In *7th European Signal Processing Conference*, Edinburgh, Scotland.

Procházka, A. and V. Sýs. 1994. "Time Series Prediction Using Genetically Trained Wavelet Networks". In *4th Workshop on Neural Networks for Signal Processing*, Porto Hydra, Greece.

Rao, S. S. and R. S. Pappu. 1993. "Nonlinear Time Series Prediction Using Wavelet Networks". In *Proceedings of World Congress on Neural Networks*, pages IV/613–IV/61 6, Portland, Oregon.

Rumelhart, D. E.; G. E. Hinton; and R. J. Williams. 1986. "Learning internal representations by error propagation" *Nature* 323: 533-536.

Strang, G. 1989. "Wavelets and Dilation Equations: A brief introduction". *SIAM Review* 31 no. 4: 614–627.

Weigend, A. S. and N. A. Gershenfeld. 1992. "The Future of Time Series: Learning and Understanding the Past". In *Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis*, Santa Fe, New Mexico.

Whitehead, A. and T. D. Choate. 1994. "Evolving Space-Filling Curves to Distribute Radial Basis Functions Over an Input Space". *IEEE Trans. on Neural Networks* 5, no. 1:15–23.

Whitley, D. and T. Hanson. 1989. "Optimizing Neural Networks Using Faster, More Accurate Genetic Search". In *Proc. 3rd Int. Conf. on Genetic Algorithms.*

Yao L. and W. A. Sethares. 1994. "Nonlinear Parameter Estimation via the Genetic Algorithm" *IEEE on Signal Processing* 42, no. 4: 927–935.

Zhang, Q. and A. Benveniste. 1992. "Wavelet Networks". *IEEE Trans. on Neural Networks* 3, no. 7: 889–899.

BIOGRAPHY

Vladimír Sýs received the M.Sc. degree from the Czech Technical University at Prague in Control Engineering. He is currently a candidate for the Ph.D. degree in the Computing and Control Engineering Department of the Faculty of Chemical Engineering of the Prague University of Chemical Technology. His research interests include methods of time-series analysis, neural networks, genetic algorithms and system modelling.

Aleš Procházka received the M.Sc. degree from the Czech Technical University at Prague and Ph.D. degree from the Prague University of Chemical Technology both in Control Engineering. Since 1990 he is the Assoc. Professor in the Computing and Control Engineering Department of the Faculty of Chemical Engineering of the Prague University of Chemical Technology. His research interests include digital signal processing methods, system identification and modelling and neural networks use for signal classification and prediction. He is a member of IEEE and EURASIP.