

NONLINEAR SET MEMBERSHIP FORECAST OF URBAN OZONE PEAKS

G. Finzi * M. Milanese ** C. Novara ** M. Volta *

* *DEA, Università di Brescia, Italy.*

Email: finzi@ing.unibs.it, volta@ing.unibs.it

** *DAUIN, Politecnico di Torino, Italy.*

Email: mario.milanese@polito.it, carlo.novara@polito.it

Abstract: In the paper a Nonlinear Set Membership (NSM) prediction method is applied to forecast tropospheric pollution. The NSM method does not require the choice of the functional form of the model used for prediction but only assumes a regularity condition on the regression function defining the model. In this way, the complexity/accuracy problems deriving from the proper choice of a suitable parametrization are circumvented. Here the NSM method is used to forecast the tropospheric ozone concentration in Brescia, a highly populated and industrialized area in Northern Italy. Purpose of this application is to aid local Authorities in decision-making policies for secondary pollution control and prevention. The NSM method is compared to other modelling approaches such as Neural Network, Neuro-Fuzzy, cyclostationary autoregressive and k -nearest neighbor classification. Model performances are assessed through comparison of indices and statistical indicators suggested by the European Environment Agency. *Copyright*© 2005 IFAC

Keywords: Set Membership estimation, nonlinear systems, ozone forecast.

1. INTRODUCTION

Most of forecast methods in the literature are essentially based on the identification of a model of the mechanism generating the data and on the use of it for forecast (see e.g. (Box and Jenkins, 1976; Ljung, 1999; Modha and Masry, 1998)). When possible, the basic laws (e.g. physical, chemical, economical, biological,...) of involved phenomena are used to derive the model structure. This is the case of state space equations, where the functional forms of nonlinear functions are known and depend on some parameters whose values are tuned using measured data.

However, in many practical applications, the required laws are too complex or not well known and input-output (black-box) regression models are used.

The black-box forecast model identification problem can be formalized as follows. Consider a nonlinear dynamic system of the form:

$$y_{t+1} = f_o(\varphi_t) \quad (1)$$

where: $\varphi_t =$

$[y_t \dots y_{t-n_y+1} u_t^1 \dots u_{t-n_1+1}^1 \dots u_t^m \dots u_{t-n_m+1}^m]$,
 $y_t \in \mathbb{R}$ is the variable to forecast, $u_t^1, \dots, u_t^m \in \mathbb{R}$
are exogenous variables, $f_o : \mathbb{R}^n \rightarrow \mathbb{R}$, $n = n_y + \sum_{i=1}^m n_i$.

Consider that a set of noise corrupted measurements \tilde{y}_t and $\tilde{\varphi}_t$ of y_t and φ_t generated by (1) is available. Then, the number m of exogenous variables u and the lag values n_y, n_1, \dots, n_m have to be chosen and an estimate \hat{f} of f_o giving small,

¹ This work was mainly supported by Ministero dell'Università e della Ricerca Scientifica e Tecnologica under the Project "Robustness techniques for control of uncertain systems" and by Università di Brescia under the Project "Photochemical pollution forecast model" (Fondo Giovani Ricercatori).

possibly minimal, identification error $f_o - \hat{f}$ has to be found.

With regard to the choice of m, n_y, n_1, \dots, n_m , several methods have been proposed (Box and Jenkins, 1976; Ljung, 1999; Grassberger *et al.*, 1991; Abarbanel *et al.*, 1993; He and Asada, 1993; Granger, 1980).

With regard to the search of \hat{f} , the usual approach is to consider that f_o belongs to a finitely parametrized set of functions, e.g. $f_o \in \mathcal{F}(\theta) \doteq \{f(\varphi, \theta) = \sum_{i=1}^r \alpha_i \sigma_i(\varphi, \beta_i), \beta_i \in \mathbb{R}^q\}$ where $\theta = [\alpha, \beta]$ and the σ_i 's are given functions. Then, measured data are used to derive an estimate $\hat{\theta}$ of θ . Fundamental to this approach is the proper choice of the parametric family of functions $f(\varphi, \theta)$, typically realized by some search on different functional forms, e.g. linear, polynomial, wavelet, neural networks, neuro-fuzzy networks, etc. ((Sjöberg *et al.*, 1995), for ozone forecast models see also (Nunnari *et al.*, 1998; Finzi *et al.*, 2002; Schlink *et al.*, 2003)). This search may be quite time consuming, and leads to approximate model structures only. The evaluation of the effects of such an approximation on identification errors is still an open problem.

Another critical point is that the estimate $\hat{\theta}$ of θ is usually obtained by a prediction error method, which requires the minimization of a quadratic error function. In several cases, such as neural and neuro-fuzzy networks, the minimization problem is non-convex. This gives rise to possible deteriorations in approximation due to trapping in local minima.

An alternative approach, not requiring the choice of the functional form of f_o , is the Nonlinear Set Membership (NSM) (Milanese and Novara, 2004). NSM method only uses assumptions on f_o regularity, given by bounds on its gradient, whose reliability can be tested by suitable validation analysis. In this way, the complexity/accuracy problems (Haber and Unbehauen, 1990; Sjöberg *et al.*, 1995) posed by the proper choice of the suitable parametrization of f_o are circumvented. Moreover, NSM approach does not require to solve minimization problems, thus avoiding the issue of local minima.

In this paper, NSM method is applied to forecast tropospheric ozone concentrations. NSM method is compared to different modelling approaches such as Neural Network, Neuro-Fuzzy, cyclostationary autoregressive and k -nearest neighbor classification. The models are tested on the one-day ahead forecast of daily maximum ozone concentration in Brescia, a highly populated and industrialized area in the Po Valley (Northern Italy). Model performances are assessed processing both measured and forecasted data, and evaluating indices and statistical indicators suggested by the European Environment Agency (Van Aalst and de Leeuw, 1998).

Forecasts of ozone concentrations are important information for local Authorities in charge of pollution control and prevention. The anthropic emissions (mainly road traffic and combustion processes), the frequent stagnating meteorological conditions and the high solar radiation cause high tropospheric ozone concentrations, especially during summer months. Studies in the literature have described the nonlinear relationships governing the tropospheric ozone levels (Kleinman *et al.*, 1997; Sillman, 1999; Jenkin and Clemitshaw, 2000). They suggest that simple cause-effect functions cannot forecast the link between precursor emissions (NO_x and VOC), meteorological conditions and ozone concentrations.

2. NONLINEAR SET MEMBERSHIP METHOD

In this section the main concepts and results of the Nonlinear Set Membership (NSM) identification method (Milanese and Novara, 2004) are briefly recalled.

Consider that a set of noise corrupted data $\tilde{Y}_T = [\tilde{y}_{t+1}, t = 1, \dots, T]$ and $\tilde{\Phi}_T = [\tilde{\varphi}_t, t = 1, \dots, T]$ generated by (1) is available. Then:

$$\tilde{y}_{t+1} = f_o(\tilde{\varphi}_t) + d_t, t = 1, \dots, T \quad (2)$$

where the term d_t accounts for the fact y_{t+1} and φ_t are not exactly known.

The aim is to derive an estimate \hat{f} of f_o from available measurements $(\tilde{Y}_T, \tilde{\Phi}_T)$, i.e. $\hat{f} = \phi(\tilde{Y}_T, \tilde{\Phi}_T)$. The operator ϕ , called identification algorithm, should be chosen to give small (possibly minimal) L_p error $\|f_o - \hat{f}\|_p$, where $\|f\|_p \doteq [\int_{\Phi} |f(\varphi)|^p dw]^{\frac{1}{p}}$, $p \in [1, \infty]$ and Φ is a given bounded set in \mathbb{R}^n .

Whatever algorithm ϕ is chosen, no information on the identification error can be derived, unless some assumptions are made on the function f_o and the noise d . In the literature, the typical approach is to assume a finitely parametrized functional form for f_o (linear, bilinear, neural network, etc.) and statistical models on the noise (Haber and Unbehauen, 1990; Sjöberg *et al.*, 1995; Narendra and Mukhopadhyay, 1997). In the NSM approach, different and somewhat weaker assumptions are taken, which do not require the selection of a parametric form for f_o but are related to its rate of variation. Moreover, the noise sequence $D_T = [d_1, d_2, \dots, d_T]$ is only supposed to be bounded.

Prior assumptions on f_o :

$$f_o \in K \doteq \{f \in C^1(\Phi) : \|f'(\varphi)\| \leq \gamma, \forall \varphi \in \Phi\}$$

Prior assumptions on noise:

$$D_T \in \mathcal{D} \doteq \{[d_1, \dots, d_T] : |d_t| \leq \varepsilon_t, t = 1, 2, \dots, T\}$$

■

Here, $f'(\varphi)$ denotes the gradient of $f(\varphi)$ and $\|x\| \doteq \sqrt{\sum_{i=1}^n x_i^2}$ is the Euclidean norm.

A key role in this Set Membership framework is played by the Feasible Systems Set, often called “unfalsified systems set”, i.e. the set of all systems consistent with prior information and measured data.

Definition 2.1. Feasible Systems Set
The Feasible Systems Set FSS_T is:

$$FSS_T \doteq \{f \in K : |\tilde{y}_{t+1} - f(\tilde{\varphi}_t)| \leq \varepsilon_t, \quad t = 1, 2, \dots, T\} \quad (3) \quad \blacksquare$$

The Feasible Systems Set FSS_T summarizes all the information on the mechanism generating the data available up to time T . If prior assumptions are “true”, then $f_o \in FSS_T$, an important property for evaluating identification accuracy.

Typically, in identification theory the problem of checking the validity of prior assumptions arises. The only possibility is to check if prior assumptions are invalidated by the data, i.e. if there exists no system consistent with data and assumptions; in other words, if FSS_T is empty. It is common to introduce the concept of prior assumption validation as follows.

Definition 2.2. Validation of prior assumptions

Prior assumptions are considered validated if:

$$FSS_T \neq \emptyset \quad \blacksquare$$

Necessary and sufficient condition for checking the assumptions validity are given by the following result. Let us define the functions:

$$\begin{aligned} \bar{f}(\varphi) &\doteq \min_{t=1, \dots, T} (\bar{h}_t + \gamma \|\varphi - \tilde{\varphi}_t\|) \\ \underline{f}(\varphi) &\doteq \max_{t=1, \dots, T} (\underline{h}_t - \gamma \|\varphi - \tilde{\varphi}_t\|) \end{aligned} \quad (4)$$

where $\bar{h}_t \doteq \tilde{y}_{t+1} + \varepsilon_t$ and $\underline{h}_t \doteq \tilde{y}_{t+1} - \varepsilon_t$.

Theorem 2.3. (Milanese and Novara, 2004)

- i) $\bar{f}(\tilde{\varphi}_t) \geq \underline{h}_t, t = 1, 2, \dots, T$, is necessary condition for prior assumptions to be validated.
- ii) $\underline{f}(\tilde{\varphi}_t) > \underline{h}_t, t = 1, 2, \dots, T$, is sufficient condition for prior assumptions to be validated. \blacksquare

The validation theorem 2.3 and prior information on the system can be jointly used to assess the values of the constants γ and ε_t appearing in the assumptions on function f_o and on noise d_t , such that sufficient condition holds (Milanese and Novara, 2004).

For a given estimate, $\phi(FSS_T) = \hat{f}$, the related L_p error $\|f_o - \hat{f}\|_p$ cannot be exactly computed, but its tightest bound is given by $\|f_o - \hat{f}\|_p \leq \sup_{f \in FSS_T} \|f - \hat{f}\|_p$. This motivates the following definition of the identification error, often indicated as guaranteed error.

Definition 2.4. Identification error

The identification error of $\hat{f} = \phi(FSS_T)$ is:

$$E[\phi(FSS_T)] = E(\hat{f}) \doteq \sup_{f \in FSS_T} \|f - \hat{f}\|_p \quad \blacksquare$$

The search for algorithms that minimize the identification error leads to the following optimality concepts.

Definition 2.5. Optimal algorithm
An algorithm ϕ^* is called optimal if:

$$E[\phi^*(FSS_T)] = \inf_{\phi} E[\phi(FSS_T)] \quad \blacksquare$$

The next result shows that the algorithm:

$$\phi_c(FSS_T) = f_c \doteq \frac{1}{2} (\underline{f} + \bar{f}) \quad (5)$$

is optimal for any L_p norm and that the corresponding minimal identification error can be actually computed.

Theorem 2.6. (Milanese and Novara, 2004)

For any $L_p(\Phi)$ norm, with $p \in [1, \infty)$:

i) The identification algorithm $\phi_c(FSS_T) = f_c$ is optimal

ii) $E(f_c) = \frac{1}{2} \|\bar{f} - \underline{f}\|_p = \inf_{\phi} E[\phi(FSS_T)] \quad \blacksquare$

3. CASE STUDY

The city of Brescia is located in the Po Valley in Northern Italy. It is characterized by high industrial, urban and road traffic emissions and continental climate.

The examined data records consist of O_3 , CO , NO and NO_2 hourly concentrations measured by the urban air quality monitoring station. Local temperature monitored and forecasted data are available from the meteorological office. Data were collected during the summer season, from May to September. Prediction models have been identified using the record subset of the period 1995-1998; the best performing models have been selected evaluating their prediction performance on the 1999 data set and then validated on the 2000-2001 data set. The data set corresponding to years 1995-1998 is called *identification set*, the 1999 data set is called *first validation set*, the 2000-01 data set is called *second validation set*.

4. PERFORMANCE INDEXES

In order to test the ability of predictors to foresee if the O_3 concentration will exceed an assigned threshold, European Environment Agency (Van Aalst and de Leeuw, 1998) has defined the following standard contingency table,

Table 1. Contingency table

Alarms	Observed		
	Yes	No	Total
Forecasted	Yes	No	Total
Yes	a	$f - a$	f
No	$m - a$	$N + a - m - f$	$N - f$
Total	m	$N - m$	N

where: N is the total number of data points; f is the total number of forecasted exceedances; m is the total number of observed exceedances; a is the number of correctly forecasted exceedances. Using these definitions, four skill parameters can be defined:

- $SP = \left(\frac{a}{m}\right)100\%$ is the *fraction of correct forecast* smog events (range from 0 to 100 with a best value of 100). The fraction of unexpected events is given by $(100 - SP)\%$;
- $SR = \left(\frac{a}{f}\right)100\%$ is the *fraction of realized forecast* smog events (range from 0 to 100 with a best value of 100); the *fraction of false alarms* is given by $(100 - SR)\%$;
- assuming an equal weight for the correct forecast of smog events and non-smog events, the scoring parameters SP and SR can be combined to give the *success index* $SI = \left(\frac{a}{f} + \frac{N+a-m-f}{N-m} - 1\right)100\%$, ranging from -100 to 100 with a best value of 100;
- $S = 100 \left\{ 1 - \frac{\sum (\hat{y}_{t+1} - y_{t+1})^2}{\sum (y_t - y_{t+1})^2} \right\}$ represents the so-called *skill score*. In this expression \hat{y}_{t+1} and y_{t+1} represents the predicted and measured value at time $(t + 1)$, respectively. The main aim of this parameter is to evaluate the extent to which a given prediction model is globally superior to the *persistent model*. As easily understood, the skill score of a persistent model is zero. Therefore if a given prediction model exhibits $S > 0$, it is globally better than the corresponding persistent model.

5. THE MODELS

The forecast models process chemical and meteorological data collected during day-time until 8 pm, and forecast the maximum expected hourly concentration value of the following day. In general, models are of the form:

$$y_{t+1} = f(\varphi_t) \quad (6)$$

$$\varphi_t = [y_t \ u_t^1 \ u_t^2 \ u_t^3 \ u_t^4]$$

where:

- Time step: one day.
- y_t : maximum ozone concentration $[O_3]$ until 8 pm at day t .
- u_t^1 : mean nitrogen dioxide concentration $[NO_2]$ from 4 pm to 8 pm at day t .
- u_t^2 : mean ozone concentration $[O_3]$ from 4 pm to 8 pm at day t .

- u_t^3 : maximum temperature T until 8 pm at day t .
- u_t^4 : estimate of maximum temperature until 8 pm at day $t + 1$.

The input u_t^4 , i.e. the estimate of maximum temperature until 8 pm at day $t + 1$, has been obtained using a linear auto-regression model.

The model inputs and the lag values have been chosen and handled by means of statistical techniques (analysis of ozone and precursor typical day and causality issue between ozone and chemical-meteorological parameters (Granger, 1980)).

Nonlinear Set Membership model NSM

The NSM model is of the form (6), with:

$$f(\varphi) = f_c(\nu\varphi)$$

where f_c is defined by (5). The regressor φ has been scaled by the vector ν in order to adapt to data. The used scaling vector is $\nu = [0.49 \ 0.12 \ 0.2 \ 1.04 \ 0.097]$. The values $\gamma = 10.6$ and $\varepsilon_t = 40, \forall t$ have been chosen for f_o gradient norm bound and noise bound on the base of the validation analysis of (Milanese and Novara, 2004).

Neural Network model NN

The structure considered in this work is the one hidden layer neural network (see e.g. (Hertz *et al.*, 1991) and (Vapnik, 1995)), i.e. a function f_{NN} of the form:

$$f_{NN}(\varphi) = \sum_{l=1}^r \alpha_l \sigma(\beta_l \varphi - \lambda_l) + \zeta \quad (7)$$

where r is the number of neurons, $\alpha_l, \lambda_l, \zeta \in \mathbb{R}$, $\beta_l \in \mathbb{R}^n$, are parameters and $\sigma(x) = 2/(1 + e^{-2x}) - 1$ is a sigmoidal function. Neural networks learn on a training data set, tuning parameters $\alpha, \lambda, \zeta, \beta$ by means of a back-propagation algorithm.

The NN model is of the form (6), with $f = f_{NN}$. Several neural networks of the form (7), with different values of r , have been trained on the identification set. A neural network with $r = 8$, showing the best prediction performances on the first validation set, has been chosen for the model NN.

Neuro-Fuzzy model NF

In neuro-fuzzy systems ((Shing and Jang, 1993), (Babuška and Verbruggen, 2003)), neural networks are used to tune the membership functions of the fuzzy system and to automatically extract fuzzy rules from numerical data. In this work, a four-layer neuro-fuzzy network has been assumed.

The nodes of the first layer (layer A) represent the regressor φ . The activation functions of the second layer nodes act as membership functions. Each neuron of the third layer acts as a rule node so that this layer provides the fuzzy rule base. The output of this layer determines the activation level at the output memberships. When φ_i is the i -th node in layer A , o_j^l is the j -th output of generic

layer L and ϕ_{ij}^L is the weight of the link between $j - th$ neuron at layer $L + 1$ and $i - th$ neuron at layer L , each layer output can be described as follows:

$$\begin{aligned} \text{Layer B:} & \quad o_j^B = o^B(\varphi_i, \phi_{ij}^{Ap}, \phi_{ij}^{Al}) \\ \text{Layer C:} & \quad o_j^C = \min_i(\phi_{ij}^B \cdot o_i^B) \\ \text{Layer D:} & \quad o_j^D = \frac{\sum_i(\phi_{ij}^C \cdot o_i^C)}{\sum_i o_i^C} \end{aligned}$$

The output of the network is a scalar variable, i.e. $o_j^D = o^D \in \mathbb{R}$. The four-layer neuro-fuzzy network is then a function f_{NF} of the form:

$$f_{NF}(\varphi) = o^D(o^C(o^B(\varphi))) \quad (8)$$

As ordinary neural nets, the neuro-fuzzy net learns on a training data set, tuning membership functions and rules by means of a back-propagation algorithm.

The Neuro-Fuzzy model has been identified assuming Gaussian membership functions and sumprod inference mechanism.

The NF model is of the form (6), with $f = f_{NF}$.

Auto Regressive Ciclostationary with exogenous inputs model ARCX

ARCX *grey box* stochastic models are extended ARX models which take into account the non-stationary and nonlinear behavior of the process through parameters depending on time varying classes.

The regression function is a linear function of regressor φ , whose coefficients are variable in time:

$$f(\varphi) = \theta_t \varphi + b_t \quad (9)$$

In general, coefficients θ_t and b_t depend on regressor φ , exogenous variables $v_t^1, \dots, v_t^l \in \mathbb{R}$, and time t : $\theta_t = \theta(\varphi, v_t^1, \dots, v_t^l, t)$, $b_t = b(\varphi, v_t^1, \dots, v_t^l, t)$. The coefficients θ_t and b_t are estimated in such a way that $\theta_t \varphi_t + b_t \simeq f_o(\varphi_t)$ in a neighborhood of the working point φ_t .

A particular kind of non-stationarity is cyclostationarity, occurring whenever the process has an underlying periodic component (for instance: daily, weekly, yearly). In this case, the parameters vary with time according to a periodic function.

The ARCX model is of the form (6) where f is given by (9) and the coefficient vector θ_t are time-varying with a weekly periodic component.

Nearest Neighbor Classification model NNC

The performance indices SP, SR and SI defined in Section 4 measure the capability of a predictor to foresee if the O_3 concentration exceeds a given threshold. Since the problem of forecasting if a variable exceeds an assigned threshold can be treated as a classification problem, a k -nearest neighbor decision rule has been implemented. The k -nearest neighbor rule classifies a vector φ to the

class most heavily represented among its k nearest neighbors in the Φ space using Euclidean metrics.

Let $\Phi_t^k \subseteq \tilde{\Phi}_T$ be the set of the k points of $\tilde{\Phi}_T$ that are closest to φ_t . Let \bar{k}_t be the number of points of Φ_t^k corresponding to an exceedance. Then, the k -nearest neighbor decision rule is defined by:

$$y_{t+1} = \begin{cases} 1 & \text{if } \bar{k}_t \leq k/2 \\ 2 & \text{if } \bar{k}_t > k/2 \end{cases}$$

where class 1 corresponds to values not exceeding the threshold, and class 2 corresponds to values exceeding the threshold.

Several k -nearest neighbor rules with k ranging from 1 to 20 have been considered. The 13-nearest neighbor rule, leading to the best performances on the first validation set, has been chosen for the NNC model.

Persistent model PER

The persistent model is often used in the forecast literature for comparison. In particular, its forecast performances can be considered as a lower bound on the performances that a "good" model should display. The persistent model PER is defined by: $y_{t+1} = y_t$.

6. PREDICTION RESULTS

Prediction performances of the identified models have been evaluated on the basis of the performance indices defined in section 4 for $130\mu\text{g}/\text{m}^3$ threshold. In Tables 2 and 3 the performance indices obtained by the models on the first and second validation set are shown.

Table 2. Prediction performances on the first validation set (1999).

Model	SP [%]	SR [%]	SI [%]	S
NSM	72.4	73	52.3	27.4
NN	69.8	72.1	55.7	32.6
NF	63.5	74.1	51.8	27.8
ARCX	61.9	75	51.1	23.7
NNC	61.4	73.6	44.8	-
PER	65.1	66.1	47.6	0

Table 3. Prediction performances on the second validation set (2000-01).

Model	SP [%]	SR [%]	SI [%]	S
NSM	71	54	62.3	5.8
NN	53.8	60.0	49.6	20.4
NF	66.7	55.3	60.2	22.1
ARCX	35.9	48.3	31.3	22.7
NNC	50	51.4	43.1	-
PER	41.5	42.5	34.4	0

7. CONCLUSIONS

In this paper a Nonlinear Set Membership (NSM) prediction method is applied to a problem of troposphere pollution forecast. The NSM method does not require assumptions on the functional form of involved nonlinearities, thus reducing the effects of modelling errors. The issue of local minima is avoided since no minimization problem needs to be solved. Moreover, the method does not use statistical assumptions such as stationarity, uncorrelation, etc., whose validity is difficult to check and would be lost in the presence of approximate modelling. On the basis of these theoretical features, it is expected that forecasts obtained by means of NSM method may show good robustness versus imprecise knowledge of involved nonlinearities and noise properties. These theoretical expectations appear to be confirmed by the numerical results, where the NSM model exhibits interesting performances with respect to other well-established methods.

REFERENCES

- Abarbanel, H. D. I., R. Brown, J. J. Sidorowich and L. S.H. Tsimring (1993). The analysis of observed chaotic data in physical system. *Rev. Mod. Phys.* **65**(4), 1331–1392.
- Babuška, R and H. Verbruggen (2003). Neuro-fuzzy methods for nonlinear system identification. *Annual Reviews in Control* **27**, 73–85.
- Box, G.E.P. and G.M. Jenkins (1976). *Time series analysis: Forecasting and control*. Holden Day. San Francisco, CA.
- Finzi, G., G. Nunnari and M. Volta (2002). Modelling and short term forecasting of photochemical pollution by soft computing techniques. In: *proc. IEMSS*. pp. 120–123.
- Granger, C.W.J. (1980). Testing for causality, a particular viewpoint. *Journal of Economic Dynamic Control* **2**, 329–352.
- Grassberger, P., T. Sreiber and Schaffrath (1991). Nonlinear time sequence analysis. *J. Bifurcation and Chaos* **1**, 482–489.
- Haber, R. and H. Unbehauen (1990). Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica* **26**, 651–677.
- He, X. and H. Asada (1993). A new method for identifying orders of input-output models for nonlinear dynamical systems. In: *Proc. of the American Control Conference*. San Francisco, CA. pp. 2520–2523.
- Hertz, J., A. Krogh and G. Palmer (1991). *Introduction to the theory of neural computation*. Addison - Wesley, Santa Fe Institute studies in the sciences of complexity. Reading (Mass.).
- Jenkin, M.E. and K.C. Clemitshaw (2000). Ozone and Other Secondary Photochemical Pollutants: Chemical Processes Governing their Formation in the Planetary Boundary Layer. *Atmos. Environ.* **34**, 2499–2527.
- Kleinman, L.I., P.H. Daum, J.H. Lee, Y. Lee, L.J. Nunnermacker, S.R. Springston and L. Newman (1997). Dependence of Ozone Production on *NO* and Hydrocarbons in the Troposphere. *Geophys. Res. Letters* **24**(18), 2299–2302.
- Ljung, L. (1999). *System identification: theory for the user*. Prentice Hall. Upper Saddle River, N.J.
- Milanese, M. and C. Novara (2004). Set membership identification of nonlinear systems. *Automatica* **40**/6, 957–975.
- Modha, D. S. and E. Masry (1998). Memory-universal prediction of stationary random processes. *IEEE Transaction on Information Theory* **44**, 117–133.
- Narendra, K. S. and S. Mukhopadhyay (1997). Neural networks for system identification. In: *Sysid 97*. Vol. 2. pp. 763–770.
- Nunnari, G., A.F.M. Nucifora and C. Randieri (1998). The application of neural techniques to the modelling of time-series of atmospheric pollution data. *Ecological Modelling* **111**, 187–205.
- Schlink, U., S. Dorling, E. Pelikan, G. Nunnari, G. Cawley, H. Junninen, A. Greig, R. Foxall, K. Eben, T. Chatterton, J. Vondracek, M. Richter, M. Dostal, L. Bertucco, M. Kolehmainen and M. Doyle (2003). A rigorous inter-comparison of ground-level ozone predictions. *Atmospheric Environment* **37**, 3237–3253.
- Shing, J. and R. Jang (1993). Anfis: Adaptive-network-based fuzzy inference system. *IEEE trans. on System, Man and Cybernetics* **23**, 665–683.
- Sillman, S. (1999). The Relation Between Ozone, *NO_x* and Hydrocarbons in Urban and Polluted Rural Environments. *Atmos. Environ.* **33**, 1821–1845.
- Sjöberg, J., Q. Zhang, L. Ljung, A. Benveniste, B.Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica* **31**, 1691–1723.
- Van Aalst, R.M. and F.A.A.M. de Leeuw (1998). National ozone forecasting system and international data exchange in northwest europe. Technical Report 9. European Environmental Agency.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag.