Atmospheric Pollution Research

www.atmospolres.com

Evolutionary procedure based model to predict ground–level ozone concentrations

Jose C.M. Pires, Maria C.M. Alvim–Ferraz, Maria C. Pereira, Fernando G. Martins

LEPAE, Departamento de Engenharia Química, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

ABSTRACT

This study aims to predict the next day hourly average ozone (O_3) concentrations using threshold autoregressive (TAR) models in which the threshold value and the threshold variable are defined by genetic algorithms. The procedure is also able to generate models with statistically significant regression parameters. The performance of TAR models was then compared to the one obtained with autoregressive (AR) and artificial neural network (ANN) models. Different TAR models were generated, corresponding to different threshold variables and values. For the training period, ANN model presented better results than TAR and AR models. However, in the test period, AR and one of the TAR models achieved better predictions of O_3 concentrations than the ANN model. The distinction between the applied models became greater when they were evaluated in the prediction of the extreme values, for which the TAR model presented the best performance. The performance with respect to extreme values is a useful implication for the protection of public health as this model can provide more reliable early warnings about high O_3 concentration episodes.

Keywords:
Time series
Model selection
Genetic algorithms
Ozone concentration forecasting
Threshold autoregressive model
Artificial neural network model

Article History:
Received: 24 March 2010
Revised: 22 May 2010
Accepted: 12 July 2010

Corresponding Author:
Fernando G. Martins
Tel: +351 22 508 1974
Fax: +351 22 508 1449
E-mail: fgm@fe.up.pt

doi: 10.5094/APR.2010.028

1. Introduction

Time series is defined as a sequence of observed points of a variable, usually measured at equally spaced time interval. Considering that Y_1, Y_2, …, Y_n (with n>1) is a time series, the aim of the time series model is to predict the next value Y_{n+1}, based on data already measured (Y_1 to Y_n) (Zou and Yang, 2004). Palit and Popovic (2005) and Gooijer and Hyndman (2006) divided these models into linear and nonlinear models. The linear models try to find a linear relationship between the predicted and the explanatory variables. The most common examples are the auto-regressive (AR), moving average, autoregressive moving average, autoregressive moving average with exogenous inputs (ARMAX) and autoregressive integrated moving average (ARIMA) models. Some examples of nonlinear models are the smooth transition autoregression, autoregressive conditional heteroskedasticity, Markov switching, threshold autoregression (TAR) and bilinear models.

TAR model assumes that the behavior of the series changes for different regimes. The change from one regime to another depends on the past values of the series. Terui and Dijk (2002) presented a TAR model composed by two AR models, each one for a different regime. This model is given by:

$$Y_t = \begin{cases} \hat{\alpha}_0 + \sum_{i=1}^{k} \hat{\alpha}_i Y_{t-i} + \varepsilon_{t,1}, \; if \; Y_{t-d} \le r \\ \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i Y_{t-i} + \varepsilon_{t,2}, \; if \; Y_{t-d} > r \end{cases} \quad (1)$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ (i=0,…,k) are the regression parameters applying the AR model to each regime; $\varepsilon_{\tau,1}$ and $\varepsilon_{\tau,2}$ are the errors associated with the regressions; the values of r and d are the threshold value and delay parameter (the delay parameter defines which input variable should be evaluated and compared with the threshold value to decide the regression equation to use), respectively. In the application of the AR model, the regression parameters were determined by minimizing the sum of squared errors (SSE) (Pires et al., 2008a). Additionally, only the statistically significant regression parameters should be considered. The statistical significance of regression parameters was evaluated through the calculation of confidence intervals for a given significance level. Pires et al. (2008a) assumed that a regression parameter $\hat{v}_i$ (standing for either $\hat{\alpha}_i$ or $\hat{\beta}_i$) was statistically significant if:

$$\left|\hat{v}_i\right| > \frac{t_{n-k-1}^{\alpha/2}\hat{\sigma}}{\sqrt{Sxx_i}} \quad (2)$$

where t is the Student's t distribution, n is the number of data points, k is the number of explanatory variables, α is the significance level, $\hat{\sigma}$ is the standard deviation given by $\sqrt{SSE/(n-k-1)}$, and Sxx_i is the sum of squares related to an explanatory variable x_i given by $\sum_{j=1}^{n}\left(x_{i,j}-\bar{x}_i\right)^2$.

Genetic algorithms (GAs) have been used to define threshold variable and threshold value for TAR model (Wu and Chang, 2002; Baragona et al., 2004). In this study, besides the optimization of r and d values, GAs were applied to define the explanatory variables that are used in each regression, with the constraint that all regression parameters must be statistically significant. As a case study, the prediction of next day hourly average O_3 concentrations was analyzed due to the importance of this problem for human health protection. Besides its negative effects on human health, O_3 is harmful to vegetation, climate, materials, and atmospheric composition (Leeuw, 2000; Bytnerowicz et al., 2007; Pires et al., 2008b). It is a secondary pollutant, predominantly formed by photochemical reactions involving other air pollutants, under suitable meteorological conditions. Thus, a typical daily concentration profile is generally observed, showing maximum values during the early afternoon and minimum values at night and early morning. Therefore, the use of time series models to predict O_3 concentrations seems promising. Several attempts have been made to predict air pollutant concentrations using statistical models (Nunnari et al., 1998; Salcedo et al., 1999; Prybutok et al., 2000; Kao and Huang, 2000). Nunnari et al. (1998) applied neural techniques for the prediction of concentrations of several air pollutants. Neural network models were compared with time series (AR and ARMAX), obtaining better results. Prybutok et al. (2000) predicted daily maximum O_3 concentrations using ANN models, multiple linear regression and ARIMA models. The ANN model obtained better results. Kao and Huang (2000) developed ANN and time series models to predict SO_2 and O_3 concentrations. ANN model performed slightly better than the time series model. Salcedo et al. (1999) applied a model based on a stepwise approach to time series analysis to predict the daily average concentrations of strong acidity and black smoke. For all analyzed monitoring sites, statistically significant higher frequency (2–4 days) periodic components were observed for both pollutant indicators. As far as it is known, no study has been performed using a TAR model in the air quality modeling field. This study aims to predict the next day hourly average O_3 concentrations through the application of TAR model. Moreover, the performance of TAR model was compared to the ones obtained with AR and ANN models. In this study, the ANN models were formed by three layers. Different numbers of hidden neurons (1 to 8) were tested and for each one, 100 trials were done. Cross–validation was performed to avoid the overtraining using 20% of the training period as the validation data. The selected model corresponded to the least error in the training period.

The remainder of this paper is outlined as follows: in Section 2, GAs are presented and how they are applied to TAR model; Section 3 describes the case study; in Section 4, the results of different TAR models are discussed; and in Section 5, the conclusions are highlighted.

2. Genetic Algorithms

GAs are search and optimization techniques introduced by Holland (Holland, 1975; Lauret et al., 2005; Rothlauf, 2006), based on the Darwin principles of evolution and natural genetics. Three principles are considered important: (i) the existence of a population limited by a maximum number of individuals with different properties and abilities; (ii) the natural creation of new individuals with similar properties of the existing ones; and (iii) the natural selection of fittest individuals.

GAs begin frequently with randomly generated set of individuals (also called chromosomes) that constitute the initial population (Rothlauf, 2006; Bandyopadhyay and Pal, 2007). Each chromosome, a potential solution of a given problem, has genes that represent the model parameters. To evaluate the degree of goodness of the solution represented by each chromosome, a fitness function must be defined. The fittest chromosomes are then submitted to genetic operations (selection, crossover and mutation) to create new individuals. The repetition of this procedure generates a sequence of populations (generations), generally containing better solutions. The termination criteria usually applied to GAs are: (i) stop after a fixed number of generations; or (ii) stop when a chromosome reaches a specific fitness value. In this study, the GA procedure was stopped when the maximum number of generations was achieved. Some examples of GA methodology were presented previously by Holland (1975) and Wu and Chang (2002).

The population size is the number of chromosomes in a population. A large number of chromosomes increases the population diversity, but also increases the computation time due to the fitness evaluation step. Goldberg (1989) reported that the population size selected by many GA researchers usually ranges from 30 to 200. In this study, the population size was fixed at 100 chromosomes. Preliminary simulations showed that for this population size the number of generations should be high to achieve convergence. Thus, the number of generations was 500. Figure 1 shows the codification of each chromosome. Each chromosome, that uses bit string coding representation (with 29 bits), is divided into four sub–strings corresponding to: (i) the value of d (3 bits); (ii) the value of r (8 bits); (iii) the explanatory variables used in the first regression (with the data which $Y_{t-d} \le r-9$ bits); and (iv) the explanatory variables used in the second regression (with the data which $Y_{t-d} > r-9$ bits). The values of d and r are determined converting the binary sub-strings to decimal values. The last two sub–strings are used to decide if a corresponding variable is selected for regression: 1 – the variable is selected; 0 – the variable is not selected. For the prediction of the next day hourly average O_3 concentrations, the TAR model took into account, as the explanatory variables, the O_3 concentrations at the same hour of the previous eight days. Thus, only 3 bits in the chromosome were needed for the delay parameter (d), which represents integer values between 1 (correspondent binary 000) and 8 (correspondent binary 111). The data used in this study presented the minimum and the maximum O_3 concentrations of 0 and 240 μg m^{-3}, respectively. The selection of eight bits ($2^8 = 256$) for the value of r had as objective to obtain a good precision of the threshold value. The selection of eight explanatory variables and the bias needed nine bits coded in the chromosome for each regression in TAR model.

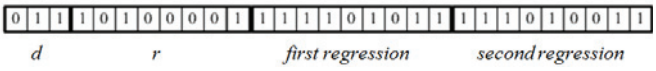

Figure 1. *Codification of chromosomes.*

The selection operation determines which chromosomes are used to generate the next population based on their fitness in the current generation (survival of the fittest). The fitness function measures the performance of the individual with respect to the particular search problem. The fitness function was defined as:

$$\arg\min f = \sqrt{\frac{SSE_1 \times 10^{ip_1} + SSE_2 \times 10^{ip_2}}{n}} \quad (3)$$

where ip is the number of statistically insignificant regression parameters and n is the number of the training points. The indexes *1* and *2* correspond to the first and the second regressions, respectively. Using this fitness function, the best models should present all statistically significant regression parameters to have

the lowest fitness value. If all regression parameters are statistically significant, the fitness value is the root mean squared error (RMSE) of the training data. The selection operation defines the chromosomes that are moved to the next generation. In many selection methods, the best solutions cannot be chosen for the next generation. Therefore, these solutions can be lost after the application of crossover and mutation. To avoid that in this study, the best elements were transferred to the next generation (elitism). However, elitism decreases the population diversity in the next generations. For that reason, after the selection procedure, all chromosomes in the current generation have equal probability of being chosen by the crossover and mutation procedures (Pires et al., 2008c).

The crossover operation consists in exchanging genetic material (binary substrings) between two parents (two chromosomes of the current generation), creating two new individuals. Siriwardene and Perera (2006) reported that high crossover rates increase the population diversity, promoting the mixing of chromosomes. In this study, the selected crossover rate was 0.7.

The mutation operation consists in modifying the chromosomes randomly. In bit string representation, the mutation is done by changing 0 to 1 and vice versa in one or more bits. Siriwardene and Perera (2006) reported that high mutation rates increase the probability of destruction of the best chromosomes. In this study, the selected mutation rate was 0.1.

The GAs procedure applied to TAR models is described by the following steps:

(1) Randomization of the initial population
(2) Calculation of *r* and *d* (threshold and delay parameters of TAR models)
(3) Division of the data into two parts (one which $Y_{t-d} \leq r$ and another with $Y_{t-d} > r$)
(4) Application of AR to the two parts of data and determination of the regression parameters and their statistical significance
(5) Calculation of the fitness values
(6) If the stopping criterion is satisfied (maximum number of generations achieved), the optimal parameters, *d* and *r* values are determined; if not it is necessary to apply the selection, crossover and mutation procedures to obtain new chromosomes. In this case, return to Step 2.

TAR and AR models were estimated using subroutines developed in Microsoft Visual Basic for Applications (MS Excel) created by the authors. ANN models were determined with the Neural Networks Toolbox of Matlab application.

3. Data

The O_3 data was collected in an urban site (Antas) with traffic influences situated in Oporto, Northern Portugal. Pires et al. (2008b) presented a description of this urban site showing the O_3 daily profile and high levels of NO_X at this site. This site belongs to the air quality monitoring network of the Oporto Metropolitan Area and is managed by the Regional Commission of Coordination and Development of Northern Portugal (*Comissão de Coordenação e Desenvolvimento Regional do Norte*), under the responsibility of the Ministry of the Environment. The most important air pollution sources in Oporto Metropolitan Area are vehicle traffic, an oil refinery, a petrochemical complex, a natural gas–fired thermal power plant, an incineration unit and an international shipping port (Pereira et al., 2007).

The O_3 measurements were performed through UV–absorption photometry using the 41 M UV Photometric Ozone Analyser from Environment S.A. (Pires et al., 2008b). This equipment was submitted to a rigid maintenance programme, being calibrated for every four weeks. Measurements were continuously made and hourly average concentrations (in $\mu g\, m^{-3}$) were recorded.

The data was organized in such a way that the predicted hourly average O_3 concentration was the function of the O_3 concentrations at the same hour of the eight days before. The analysed period was from May to August 2006. Table 1 presents some important statistics of the used dataset. It was divided in the training (May, June and July, corresponding to 2 199 data points) and test (August, corresponding to 733 data points) periods. The explanatory variables were Z standardized to zero mean and unit standard deviation.

Table 1. *Statistics of the used dataset*

Average	Percentile				
	0	**25**	**50**	**75**	**100**
49	0	22	47	70	200

4. Results and Discussion

Different TAR models were obtained, having different threshold variables and values. These models were compared with AR and feed–forward ANN models. The statistical significance of the regression parameters in TAR and AR models was evaluated by a *t*–test with a significance level of 0.05. Table 2 shows the

Table 2. *Statistically significant regression parameters of threshold autoregressive models (M1 to M6) and autoregressive model (AR) models and the root squared mean error (RMSE) value of the training data for all models (including artificial neural network – ANN-model)*

Model	$\hat{\nu}_0$	$\hat{\nu}_1$	$\hat{\nu}_2$	$\hat{\nu}_3$	$\hat{\nu}_4$	$\hat{\nu}_5$	$\hat{\nu}_6$	$\hat{\nu}_7$	$\hat{\nu}_8$		RMSE
M1	48.2	13.7	2.7	2.7	3.0	2.0		2.4	-2.1	*If* $Y_{t-1} \leq 88.4$	23.17
	45.4	20.5			17.8	-12.4	-22.8	26.3		*If* $Y_{t-1} > 88.4$	
M2	48.5	13.6	3.2	2.9	3.5	1.7		2.9	-2.2	*If* $Y_{t-2} \leq 125.1$	23.17
	134.7	16.7	-19.2		13.9		-41.9		31.1	*If* $Y_{t-2} > 125.1$	
M3	48.6	14.2	3.4	3.0	3.4	1.4		2.9	-2.1	*If* $Y_{t-2} \leq 159.6$	23.21
	-94.3	-14.5		79.4	21.9		24.8	-102.9	54.3	*If* $Y_{t-2} > 159.6$	
M4	48.5	13.6	3.3	2.9	3.5	1.7		3.0	-2.2	*If* $Y_{t-2} \leq 118.7$	23.25
	84.5	14.5					-33.3		42.5	*If* $Y_{t-2} > 118.7$	
M5	49.2	14.3	3.6	3.0	4.1	2.5		2.7	-1.9	*If* $Y_{t-6} \leq 96.1$	23.26
	48.8	11.5	-4.7					5.3		*If* $Y_{t-6} > 96.1$	
M6	48.8	14.2	3.1	2.8	4.0			2.9		*If* $Y_{t-7} \leq 116.6$	23.44
	24.2	-10.2						13.1		*If* $Y_{t-7} > 116.6$	
AR	48.3	14.4	2.5	3.1	3.7			2.7	-1.5		23.84
ANN											22.11

regression parameters of the six TAR models (M1 to M6) determined with fitness value higher than that obtained with AR model and the RMSE of the training data for all models (including the ANN model). Concerning the training period, the ANN model (which presented 8 hidden neurons) overcame the TAR and AR models. For the TAR and AR models, all regression parameters were considered statistically significant. Therefore, the fitness value corresponded to the RMSE of the training data. Additionally, it was also observed that the two regimes in TAR models did not have an equal number of data. The first regime (with $Y_{t-d} \leq r$) had always more data compared to the second regime, due to high threshold values achieved.

In the test period, the O_3 concentrations given by TAR, AR and ANN models were determined by the application of the regression equations obtained in the training period. The models were compared through the calculation of the following statistical parameters: mean bias error (MBE), mean absolute error (MAE), RMSE, Pearson correlation coefficient (R) and index of agreement of the second order (d_2) (Gardner and Dorling, 2000; Chaloulakou et al., 2003). Table 3 presents the performance indexes presented by the TAR, AR and ANN models.

Table 3. *Performance indexes of the threshold autoregressive models (M1 to M6), autoregressive model (AR) and artificial neural network (ANN) models for the test period*

	M1	M2	M3	M4	M5	M6	AR	ANN
MBE	-1.42	-1.84	-1.66	-1.85	-3.25	-3.87	-1.50	-1.76
MAE	18.99	19.99	19.86	20.29	19.33	20.04	18.94	21.97
RMSE	25.26	27.76	27.47	28.39	26.36	27.96	25.19	28.44
R	0.78	0.73	0.73	0.71	0.76	0.72	0.78	0.72
d_2	0.86	0.83	0.84	0.82	0.83	0.81	0.85	0.79

MBE was negative in all models, meaning that, on average, the predicted O_3 concentrations were underestimated. MAE and RMSE measure residual errors, which give a global idea of the difference between the observed and modelled values. Thus, M1 and AR were the models that presented the best performance indexes. Table 4 shows the performance indexes of M1, AR and ANN models in the test period when hourly average O_3 concentrations above 88.4 μg m^{-3} (threshold value of M1 model) were recorded.

Table 4. *Performance indexes of the best threshold autoregressive model (M1), autoregressive model (AR) and artificial neural network model (ANN) obtained for the test period when hourly average O_3 concentrations above 88.4 μg m^{-3} (threshold value of M1 model) were recorded*

	M1	AR	ANN
MBE	-33.67	-36.83	-47.87
MAE	38.47	39.77	48.24
RMSE	45.42	46.16	54.41
R	0.53	0.53	0.47
d_2	0.59	0.57	0.49

In test set, 96 data points with O_3 concentrations (percentile 90 of the all dataset; percentile 87 of the test set) above this value were recorded. The prediction of high O_3 concentrations is an important issue due to the negative effects that this air pollutant causes at these levels. In this range of O_3 concentrations, the difference between these models was more significant, with M1 being the model that presents the best performance. On the other hand, ANN model presented the worst predictions. Figure 2 (a and b) show, as an example, the predictions with M1 and AR models in August 6, 7, 20 and 21, 2006. It was shown that the M1 model led to better predictions for high O_3 concentrations.

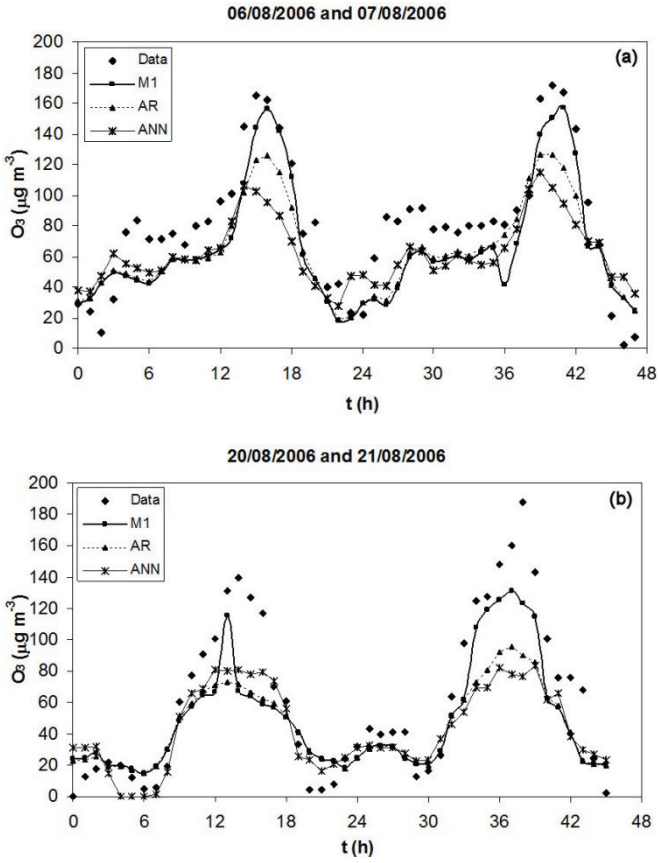

Figure 2. *Prediction of hourly average O_3 concentrations using M1, AR and ANN models:* ***(a)*** *August 6 and 7, 2006; and* ***(b)*** *August 20 and 21, 2006.*

5. Conclusions

GAs were applied to define TAR models for prediction of the next day hourly average O_3 concentrations. Different models were obtained with different threshold variables and values. For the training period, the ANN model presented better results than TAR and AR models. However, in the test period, AR and one of the TAR models showed the best predictions of O_3 concentrations (better predictions than that obtained with the ANN model). The distinction between the applied models became greater when they were evaluated on their ability to predict extreme values (> 88.4 μg m^{-3}). TAR model allowed more efficient predictions of extreme O_3 concentrations, which are very important to develop efforts to reduce the negative effects of O_3.

Future work should extend the proposed method for TAR models with more than two regressions. Genetic algorithms could also be conjugated with ANN models to: (i) determine the best combination of the input variables; and (ii) define threshold regressions using these models.

Acknowledgements

Authors are grateful to *Comissão de Coordenação da Direcção Regional–Norte* and to *Instituto Geofísico da Universidade do Porto*, for kindly providing the air quality and meteorological data. This work was supported by *Fundação para a Ciência e Tecnologia* (FCT). J.C.M. Pires also thanks the FCT for the fellowship SFRH/BD/23302/2005.

References

Bandyopadhyay, S., Pal, S.K., 2007. *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*. Springer–Verlag, Berlin Heidelberg, pp. 19-51.

Baragona, R., Battaglia, F., Cucina, D., 2004. Estimating threshold subset autoregressive moving–average models by genetic algorithms. *Metron* 62, 39-61.

Bytnerowicz, A., Omasa, K., Paoletti, E., 2007. Integrated effects of air pollution and climate change on forests: a northern hemisphere perspective. *Environmental Pollution* 147, 438-445.

Chaloulakou, A., Saisana, M., Spyrellis, N., 2003. Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. *Science of the Total Environment* 313, 1-13.

De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. *International Journal of Forecasting* 22, 443-473.

De Leeuw, F.A.A.M., 2000. Trends in ground level ozone concentrations in the European Union. *Environmental Science and Policy* 3, 189-199.

Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improved methodology to account for non–linear behaviour. *Atmospheric Environment* 34, 21-34.

Goldberg, D.E, 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Kluwer Academic Publishers, Boston, MA, pp. 1-412.

Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, pp. 1-183.

Kao, J.J., Huang, S.S., 2000. Forecasts using neural network versus Box–Jenkins methodology for ambient air quality monitoring data. *Journal of the Air and Waste Management Association* 50, 219-226.

Lauret, P., Boyer, H., Riviere, C., Bastide, A., 2005. A genetic algorithm applied to the validation of building thermal models. *Energy and Buildings* 37, 858-866.

Nunnari, G., Nucifora, A.F.M., Randieri, C., 1998. The application of neural techniques to the modelling of time-series of atmospheric pollution data. *Ecological Modelling* 111, 187-205.

Palit, A.K., Popovic, D., 2005. *Computational intelligence in time series forecasting: theory and engineering applications,* Springer, London, pp. 1-388.

Pereira, M.C., Santos, R.C., Alvim-Ferraz, M.C.M., 2007. Air quality improvements using European environment policies: a case study of SO_2 in a coastal region in Portugal. *Journal of Toxicology and Environmental Health*, Part A 70, 347-351.

Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C., 2008a. Selection and validation of parameters in multiple linear and principal component regressions. *Environmental Modelling and Software* 23, 50-55.

Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim-Ferraz, M.C.M., Martins, F.G., 2008b. Management of air quality monitoring using principal component and cluster analysis - Part II: CO, NO_2 and O_3. *Atmospheric Environment* 42, 1261-1274.

Pires, J.C.M., Martins, F.G., Alvim-Ferraz, M.C.M., Pereira, M.C., 2008c. Genetic algorithm based technique for defining threshold regression models. *iEMSs 2008: International Congress on Environmental Modelling and Software*, 303-310, 7-10 July 2008, Barcelona.

Prybutok, V.R., Yi, J., Mitchell, D., 2000. Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research* 122, 31-40.

Rothlauf, F., 2006. *Representations for Genetic and Evolutionary Algorithms,* Springer–Verlag, Berlin Heidelberg, pp. 1-180.

Salcedo, R.L.R., Alvim Ferraz, M.C.M., Alves, C.A., Martins, F.G., 1999. Time–series analysis of air pollution data. *Atmospheric Environment* 33, 2361-2372.

Siriwardene, N.R., Perera, B.J.C., 2006. Selection of genetic algorithm operators for urban drainage model parameter optimisation. *Mathematical and Computer Modelling* 44, 415-429.

Terui, N., van Dijk, H.K., 2002. Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting* 18, 421-438.

Wu, B., Chang, C.L., 2002. Using genetic algorithms to parameters (d, r) estimation for threshold autoregressive models. *Computational Statistics and Data Analysis* 38, 315-330.

Zou, H., Yang, Y., 2004. Combining time series models for forecasting. *International Journal of Forecasting* 20, 69-84.