This article was published in Environmental Science and Pollution Research, 21 (17), 10550–10559, 2014
http://dx.doi.org/10.1007/s11356-014-2977-6

Variation of surface ozone in Campo Grande, Brazil: meteorological effect analysis and prediction

J.C.M. Pires[1] *, A. Souza[2], H.G. Pavão[2], F.G. Martins[1]

[1] LEPABE – Laboratório de Engenharia de Processos, Ambiente, Biotecnologia e Energia, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465, Porto, Portugal.

[2] Instituto de Física, Universidade Federal de Mato Grosso do Sul, Cidade Universitária s/n, caixa postal 549, cep-79090-900- Campo Grande (MS), Brasil.

*Corresponding author

Telephone: +351 22 508 2262

Fax: +351 22 508 1449

E-mail address: jcpires@fe.up.pt

Abstract

The effect of meteorological variables on surface ozone (O_3) concentrations was analysed based on temporal variation of linear correlation and artificial neural network (ANN) models defined by genetic algorithms (GAs). ANN models were also used to predict the daily average concentration of this air pollutant in Campo Grande, Brazil. Three methodologies were applied using GAs, two of them considering threshold models. In these models, the variables selected to define different regimes were daily average O_3 concentration, relative humidity and solar radiation. The threshold model that considers two O_3 regimes was the one that correctly describes the effect of important meteorological variables in O_3 behaviour, presenting also a good predictive performance. Solar radiation, relative humidity and rainfall were considered significant for both O_3 regimes; however, wind speed (dispersion effect) was only significant for high concentrations. According this model, high O_3 concentrations corresponded to high solar radiation, low relative humidity and wind speed. This model showed to be a powerful tool to interpret the O_3 behaviour, being useful to define policy strategies for human health protection regarding air pollution.

Keywords: Air quality; Surface ozone concentrations; Meteorological effect; Statistical models; Artificial neural networks; Genetic algorithms; Evolutionary procedure.

1. Introduction

Surface ozone (O_3) is one of the most important air pollutants due to its negative effects not only on human health, but also on climate, vegetation and materials (Brauer &Brook 1997, Lippmann 1991). It is a secondary pollutant, formed by chemical reactions between its precursors that occur in atmosphere. These chemical reactions are very complex, which attributes to O_3 a characteristic behaviour very difficult to predict. For instance, the O_3 concentrations at rural sites are usually higher than the ones measured at urban sites, where several air pollution sources can be found (Coyle et al. 2002, Jaffe &Ray 2007, Pires et al. 2012a). Aiming the prediction of O_3 concentrations, several research studies were recently performed (Butler et al. 2012, Colette et al. 2012, Huang et al. 2013, Manders et al. 2012, Thunis et al. 2013). The developed models can be used to replace air quality monitoring or to analyse the effect of meteorological and environmental variables on the predicted air pollutant concentration. Additionally, the forecasting of air pollutant concentrations is needed to create preventive and clear actions regarding high concentration episodes.

As mentioned above, the behaviour of O_3 concentrations is very complex, as it is the result of the combination of chemical formation, transport (horizontally and vertically), destruction and deposition. As a photochemical pollutant, it is expected that meteorological variables (mainly temperature and solar radiation) have a strong impact in O_3 concentrations. To develop predictive models, two main approaches can be applied: phenomenological and statistical models. The first approach showed to be accurate. It combines models of emissions with those of meteorological and chemical atmospheric processes. However, the main disadvantages of these approaches are: (i) the scarce emission inventories and meteorological data; (ii) the costly computation; (iii) the difficult operation; and (iv) the requirement of a high level of expertise. On the other hand, statistical models require less data to be developed, being inexpensive and easy to operate. For that reasons, they were commonly used in several recent research studies. Linear and nonlinear models have been applied to predict ozone concentrations. Multiple linear regression, principal component regression, quantile regression, among others, are few examples of linear models (Abdul-Wahab et al. 2005, Baur et al. 2004, Pires et al. 2012b) and, on the other hand, artificial neural networks (ANNs) are the nonlinear models most commonly used (Comrie 1997, Pires &Martins 2011, Yi &Prybutok 1996). As the definition of the model structure have a strong influence of its performance, evolutionary procedures were also used to optimize simultaneously the structure and parameters. Genetic algorithms (GAs) and genetic programming are some of the procedures applied to air quality modelling (Feng et al. 2011, Pires et al. 2010, 2011). GAs are generally associated with ANNs to optimize their structure

(Feng et al. 2011, Pires et al. 2012c). Pires et al. (2012c) presented three different methodologies to create ANNs with GAs, two of which are threshold models, aiming to predict the next-day hourly average O_3 concentrations at an urban site in Porto, Portugal. These methodologies presented good predictive performances and were able to identify different ozone regimes, where the exploratory variables present different influences on the output variable. Taking into account the potential of these methodologies, this paper aims to analyse single and combined effect of meteorological variables on surface ozone behaviour and to predict its daily average concentrations.

2. Data and Models

2.1. Site characterization and data

The study was conducted in the city of Campo Grande. This city is located in central western Brazil, with a population of approximately 724,000 inhabitants and a high degree of economic development, highlighting the importance of agribusiness.

Daily average concentrations of ozone (O_3) were obtained at the Department of Physics of Universidade Federal de Mato Grosso do Sul (UFMS). It is considered an urban site without influence of traffic or industrial sources. The ozone analyser has the working principle of the absorption of ultraviolet radiation by ozone molecule. The analyser is installed near Campo Grande, away from local resources. O_3 is continuously monitored and 15-min average concentration are registered. Daily averages are then calculated based on these values. Meteorological data were collected at Embrapa (Gado de Corte – Campo Grande), with the distance to UFMS of about 5 km. These data are considered representative of all region and include daily average (T_a, ^oC) and maximum (T_h, ^oC) temperature, solar radiation (SR, W m^{-2}), relative humidity (RH, %), wind speed (WS, m/s) and rainfall (RF, mm). The analysed period was from 2004 to 2010.

2.2. Models

The development of ANN models requires the division of the data into three sets: training, validation and test. The training and validation sets corresponded to the period 2004-2008 (1671 data points) and 2009 (365 data points), respectively. The test set corresponded to 2010 (364 data points).

In this study, feedforward ANN with three layers was applied to predict daily average O_3 concentrations using eight input variables: T_a, T_h, SR, RH, 1/RH, WS, RF and O_3 measured in the previous day. A linear function was used as activation function of the output neuron.

Concerning the hidden neurons, four functions were tested: sigmoid, hyperbolic tangent, inverse and radial basis. The early stopping method (training procedure is stopped when an increase of validation error is observed) was applied, to avoid the overfitting. Three methodologies proposed by Pires et al. (2012c) were applied for defining the optimum structure of ANN models using genetic algorithms (GAs) – GA-ANN models (see Table 1). GA-ANN_1 model is the linear combination of three ANN models. In this model, GAs defined the transfer function (for hidden neurons), the number of hidden neurons (up to 8) and the input variables for each ANN model. GA-ANN_2 model considered two O_3 regimes. In this model, GAs defined the threshold variable and value, the transfer function (for hidden neurons), the number of hidden neurons (up to 8) and the input variables for each ANN model (one for each regime). GA-ANN_3 model considered four O_3 regimes. Besides the threshold variables and values, GAs defined the transfer function (for hidden neurons) and the number of hidden neurons (up to 8) that was used in the four ANN models (in this methodology, ANN models have the same structure). All models were coded by the authors in MATLAB® routine.

2.3 Performance indexes

The ANN performances were evaluated through the calculation of five performance indexes: mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), index of agreement of second order (d_2) and Akaike Information Criterion (AIC).

MBE indicates if the experimental values were over or under estimated. MAE and RMSE measures residual errors, which gave a global idea of the difference between the experimental and the predicted values. The values of d_2 show the extent that predicted deviations differ from the observed deviations about the mean observed value, indicating the degree to which model's predictions are error free (Pires 2009) The Akaike Information Criterion (AIC) is a measure of goodness of fit that penalises the model complexity (Pires et al. 2012c). The model with the smallest AIC is the one that most efficiently fits the data.

3. Results and discussion

3.1. Profiles of O_3 concentrations and meteorological variables

Table 2 presents the annual averages of the meteorological variables and O_3 concentrations. The annual average concentrations ranged from 15.1 to 20.1 ppb, being the highest value obtained for 2007. In this period, slightly higher temperatures and solar radiation were observed as well as low RH and WS. This observation is in agreement with the research studies performed by other authors (Camalier et al. 2007, Jacob &Winner 2009). Camalier et al. (2007)

developed a generalised linear model developed to predict O_3 concentrations that explained 80% of the variance. This model showed a positive correlation with temperature and negative with RH, considering these variables as the most important ones. Regarding WS, low values in polluted regions cause O_3 increase due to the longer reaction time and high aerodynamic resistance to dry deposition (Baertsch-Ritter et al. 2004, Dawson et al. 2007). In addition, high WS promotes the dispersion of O_3 precursors and thus the decrease of its concentration.

Figure 1 presents the annual average profiles of O_3 concentrations during the analysed period. For each year, the highest monthly average concentrations occurred mainly in September (in 2005, it occurred in August), period between dry and wet seasons. Figure 2 shows the monthly average values of meteorological variables. The highest monthly average temperatures were observed in the period from November to January, while the lowest values occurred between May and July (values between 18.6 and 27.5 ^{o}C). Similar profile was observed for SR. These meteorological variables are often associated to high O_3 concentrations, but in this study the highest values of these variables did not occur at the same period. However, in 2007 (when the highest monthly average O_3 concentration occurred – 39.2 ppb), the highest temperatures were also measured in September, which may contribute to the increase of photochemical production of this secondary pollutant. Comparing to the average values measured in September, it was observed an increase of about 40% in O_3 levels in 2007 due to the different annual profile of temperatures observed in this year. Regarding other meteorological variables, September 2007 was the month with the lowest RH value (36.9%) from all period and it was the month with lowest RF comparing with same period of other years.

3.2. Linear correlation analysis

Figure 3 shows the temporal variation of linear correlation (monthly basis) between O_3 concentrations and meteorological variables. With respect to temperature, positive correlations were usually observed which is in agreement with what was expected. The highest correlation values were observed in September 2005 (R=0.864) and 2008 (R=0.839), and April 2010 (R=0.804). In September 2007, the correlation value was also high (R=0.737). SR also presented positive correlation with O_3. On the other hand, RH was almost always negatively correlated with O_3 concentrations. The influence of this variable took more relevance in 2008 with six absolute correlation values greater than 0.75. In September 2007, this variables presented a low and positive correlation value (R=0.129). Concerning WS and RF, no strong correlations (>0.75) with O_3 concentrations were observed.

3.3. Prediction of O_3 concentrations

Several GA-ANN models were developed to predict daily average O_3 concentrations using meteorological variables as inputs. Two of these models are threshold models that considered two or more regimes where relationship between the output and input variables are different. The change from one regime to another depends on the value (threshold value) of a specific input variable (threshold variable). GAs were used to optimise the ANN structure, input variables, transfer function and threshold variable and value (for threshold models GA-ANN$_2$ and GA-ANN$_3$). Table 3 shows the best achieved GA-ANN models. GA-ANN$_1$ considered the average of the outputs of three ANN models with radial basis function as activation function in the 8 hidden neurons, having 211 parameters. All the variables were selected. GA-ANN$_2$ selected the daily average O_3 concentrations of previous day as threshold variable. For O_3 concentrations less than 34.6 ppb, the daily average concentration of the next day was dependent on T_h, T_a, SR, RH, RF and O_3 and for concentrations higher than threshold value, WS was also considered significant in the prediction and T_h was not selected. GA-ANN$_2$ also considered radial basis function as activation function in 8 and 5 hidden neurons in the two ANN models, having 114 parameters. GA-ANN$_3$ considered two threshold variables: RH and SR. All ANN models considered hyperbolic tangent as activation function in the 8 hidden neurons. GA-ANN$_3$ selected all input variables, having 243 parameters. These models presented the best performance in the fitting of O_3 concentrations using the training and validation set (see Table 4). In this period, GA-ANN$_2$ presented better performances than GA-ANN$_1$ and GA-ANN$_3$ models (lower MAE, RMSE and AIC, and higher d_2). These models were then applied to test set (not used in their development) to evaluate their predictive performance. GA-ANN$_1$ presented better performance indexes (lower MAE and RMSE, and higher d_2); however, taking into account the complexity of the achieved model (using AIC parameter), GA-ANN$_2$ model was the one that most efficiently predicted O_3 concentrations. GA-ANN$_3$ presented the worst performance in both training and test periods. Figure 4 shows the model predictions of GA-ANN$_1$ and GA-ANN$_2$ during test period. GA-ANN$_1$ prediction values were almost always closer to the measured data.

3.4. Influence of meteorological variables in different O_3 regimes

The analysis of the influence of meteorological variables on O_3 concentrations were performed through the GA-ANN$_1$ and GA-ANN$_2$ models, which presented good performances in the O_3 prediction. The combination effect of two meteorological variables were tested for T_h, SR, RH and WS (the most selected variables by the models). The tested values belonged to the range defined by the data used for the models' development: $10.9<T_h<39.5$, $9.3<SR<359.8$, $19.2<RH<98.0$ and $0<WS<12.1$. Figure 5 shows the influence of the combination of two meteorological variables on O_3 concentrations according GA-ANN$_1$ model. The effect of T_h on O_3 concentrations was clear in all tested combinations: with SR, RH and WS. O_3 concentrations increase with T_h and presented the maximum values for $48<SR<243$ (maximum O_3 concentration of 24.0 ppb), $19<RH<81$ (maximum O_3 concentration of 24.3 ppb) and $4<WS<5$ (maximum O_3 concentration of 23.0 ppb). Regarding SR, all combinations showed that O_3 concentrations corresponded to middle or lower values. This observation is contrary to what is expected in terms of O_3 chemistry (Camalier et al. 2007, Jacob &Winner 2009). As a photochemical pollutant, its concentration should increase with solar radiation. With respect to RH, the GA-ANN$_1$ model identified two distinct O_3 behaviours. The combination effect T_h-RH showed that low O_3 concentrations corresponded to low RH, while in SR-RH the opposite relationship was observed. Chen et al. (2011) have demonstrated that RH favours O_3 decomposition, showing a negative correlation. Concerning WS, the combination of T_h-WS showed that high O_3 concentrations corresponded to high values. This observation is also at odds with what was expected. High WS values promote the dispersion of pollutants and thus their concentration tends to be low. In addition, no significant variation on O_3 concentration was observed with the tested combinations of WS with SR and RH.

Figure 6 shows the combined effect of the selected meteorological variables according GA-ANN$_2$ model. For low O_3 concentrations ($O_3 \leq 34.6$ ppb), WS was not selected by the model, i.e. WS did not present any significant influence on O_3. Thus, only the binary effects T_h-SR, T_h-RH and SR-RH were analysed in this O_3 regime. High O_3 concentrations was observed for high T_h and SR and low RH. For $O_3 > 34.6$ ppb, T_h was not considered significant in this O_3 regime. High O_3 concentrations (near 50 ppb) were observed for high SR and low RH and WS, being in agreement with other research studies (Baertsch-Ritter et al. 2004, Camalier et al. 2007, Dawson et al. 2007, Jacob &Winner 2009, Ordonez et al. 2005). The importance of WS in high O_3 concentrations was also identified by other authors. For instance, Baertsch-Ritter et al. (2004) reported that O_3 peak concentration lowers 15% when WS was doubled.

This analysis of the effect of meteorological variables in O_3 concentration should be performed for the development of predictive models. In this study, GA-ANN$_1$ model achieved the best predictive performance, but did not describe the real effect of meteorological variables on O_3 concentrations. On the other hand, GA-ANN$_2$ also obtained a good predictive performance with less complexity (low AIC value). Additionally, this model presented the accepted relationship between the studied variables, in special for high O_3 concentrations (which is important for the definition of policy measures for human health protection). Accordingly, GA-ANN$_2$ methodology should be applied to predict the O_3 concentrations.

4. Conclusions

Correlation analysis between O_3 concentrations and meteorological variables showed the positive impact of temperature and solar radiation and negative influence of relative humidity. The highest O_3 concentrations were observed in a period, when temperature presented high correlation value and low values and low impact were observed for relative humidity. Wind speed and rainfall did not show strong influence on O_3 concentrations.

Three different methodologies were applied to define ANN models through GAs to predict daily average O_3 concentrations. Two of them are threshold models and, despite not presenting the best predictive performance, the one that assumes two regimes was selected. This model presented less complexity (fitted the data most efficiently) and it describes the real relationship between the O_3 concentrations and the meteorological variables. In addition, it assumes that the meteorological effect on O_3 concentrations changed, when O_3 concentrations surpassed 34.6 ppb. Solar radiation, relative humidity and rainfall were considered significant for both O_3 regimes; however, wind speed (dispersion effect) was only significant for high concentrations. The analysis of meteorological effect on O_3 concentration through the model showed that high O_3 concentrations are associated to high solar radiations, low relative humidity and wind speed. The good predictive performance of the GA-ANN models showed that it can be useful to minimize the population exposure to high O_3 concentration episodes and to improve the political policies regarding environmental health planning.

Acknowledgements

J.C.M. Pires thanks the Foundation for Science and Technology, POPH-QREN and FSE for the Post-Doctoral fellowship SFRH/BPD/66721/2009.

References

Abdul-Wahab SA, Bakheit CS, Al-Alawi SM (2005): Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environ Modell Softw 20, 1263-1271

Baertsch-Ritter N, Keller J, Dommen J, Prevot ASH (2004): Effects of various meteorological conditions and spatial emission resolutions on the ozone concentration and ROG/NOx limitation in the Milan area (I). Atmos Chem Phys 4, 423-438

Baur D, Saisana M, Schulze N (2004): Modelling the effects of meteorological variables on ozone concentration - a quantile regression approach. Atmospheric Environment 38, 4689-4699

Brauer M, Brook JR (1997): Ozone personal exposures and health effects for selected groups residing in the Fraser Valley. Atmospheric Environment 31, 2113-2121

Butler TM, Stock ZS, Russo MR, van der Gon HACD, Lawrence MG (2012): Megacity ozone air quality under four alternative future scenarios. Atmos Chem Phys 12, 4413-4428

Camalier L, Cox W, Dolwick P (2007): The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. Atmospheric Environment 41, 7127-7137

Chen HH, Stanier CO, Young MA, Grassian VH (2011): A Kinetic Study of Ozone Decomposition on Illuminated Oxide Surfaces. J Phys Chem A 115, 11979-11987

Colette A et al. (2012): Future air quality in Europe: a multi-model assessment of projected exposure to ozone. Atmos Chem Phys 12, 10613-10630

Comrie AC (1997): Comparing neural networks and regression models for ozone forecasting. J Air Waste Manage 47, 653-663

Coyle M, Smith RI, Stedman JR, Weston KJ, Fowler D (2002): Quantifying the spatial distribution of surface ozone concentration in the UK. Atmospheric Environment 36, 1013-1024

Dawson JP, Adams PJ, Pandis SN (2007): Sensitivity of ozone to summertime climate in the eastern USA: A modeling case study. Atmospheric Environment 41, 1494-1511

Feng Y, Zhang WF, Sun DZ, Zhang LQ (2011): Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. Atmospheric Environment 45, 1979-1985

Huang M, Carmichael GR, Chai T, Pierce RB, Oltmans SJ, Jaffe DA, Bowman KW, Kaduwela A, Cai C, Spak SN, Weinheimer AJ, Huey LG, Diskin GS (2013): Impacts of transported background pollutants on summertime western US air quality: model evaluation, sensitivity analysis and data assimilation. Atmos Chem Phys 13, 359-391

Jacob DJ, Winner DA (2009): Effect of climate change on air quality. Atmospheric Environment 43, 51-63

Jaffe D, Ray J (2007): Increase in surface ozone at rural sites in the western US. Atmospheric Environment 41, 5452-5463

Lippmann M (1991): Health-Effects of Tropospheric Ozone. Environ Sci Technol 25, 1954-1962

Manders AMM, van Meijgaard E, Mues AC, Kranenburg R, van Ulft LH, Schaap M (2012): The impact of differences in large-scale circulation output from climate models on the regional modeling of ozone and PM. Atmos Chem Phys 12, 9441-9458

Ordonez C, Mathis H, Furger M, Henne S, Huglin C, Staehelin J, Prevot ASH (2005): Changes of daily surface ozone maxima in Switzerland in all seasons from 1992 to 2002 and discussion of summer 2003. Atmos Chem Phys 5, 1187-1203

Pires JCM 2009: Development and Application of Statistical Methods to Support Air Quality Policy Decisions, Faculty of Engineering, University of Porto, Porto, 144 pp

Pires JCM, Alvim-Ferraz MCM, Pereira MC, Martins FG (2010): Evolutionary procedure based model to predict ground-level ozone concentrations. Atmos Pollut Res 1, 215-219

Pires JCM, Alvim-Ferraz MCM, Pereira MC, Martins FG (2011): Prediction of tropospheric ozone concentrations: Application of a methodology based on the Darwin's Theory of Evolution. Expert Syst Appl 38, 1903-1908

Pires JCM, Martins FG (2011): Correction methods for statistical models in tropospheric ozone forecasting. Atmospheric Environment 45, 2413-2417

Pires JCM, Alvim-Ferraz MCM, Martins FG (2012a): Surface ozone behaviour at rural sites in Portugal. Atmos Res 104, 164-171

Pires JCM, Alvim-Ferraz MCM, Pereira MC, Martins FG (2012b): Comparison of several linear statistical models to predict tropospheric ozone concentrations. J Stat Comput Sim 82, 183-192

Pires JCM, Goncalves B, Azevedo FG, Carneiro AP, Rego N, Assembleia AJB, Lima JFB, Silva PA, Alves C, Martins FG (2012c): Optimization of artificial neural network models through genetic algorithms for surface ozone concentration forecasting. Environ Sci Pollut R 19, 3228-3234

Thunis P, Pernigotti D, Gerboles M (2013): Model quality objectives based on measurement uncertainty. Part I: Ozone. Atmospheric Environment 79, 861-868

Yi JS, Prybutok VR (1996): A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environ Pollut 92, 349-357

Table 1. Structures of artificial neural network models (from Pires et al. (2012c))

Model	Structure
GA-ANN$_1$	$y = \sum_{i=1}^{3} b_i \times net_i(x_j)$
GA-ANN$_2$	$y = \begin{cases} net_1(x_j), & if\ x_d \leq r \\ net_2(x_j), & if\ x_d > r \end{cases}$
GA-ANN$_3$	$y = \begin{cases} \begin{cases} net_1(x_j), & if\ x_e \leq r_2 \\ net_2(x_j), & if\ x_e > r_2 \end{cases}, & if\ x_d \leq r_1 \\ \begin{cases} net_3(x_j), & if\ x_f \leq r_3 \\ net_4(x_j), & if\ x_f > r_3 \end{cases}, & if\ x_d > r_1 \end{cases}$

Table 2. Annual average of meteorological variables and O_3 concentrations

	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
T_h	23.2	23.4	23.5	23.7	23.4	23.5	23.8
T_a	29.9	29.8	30.2	30.7	29.9	29.7	30.2
SR	209.0	199.0	205.0	208.9	206.9	197.5	189.1
RH	66.6	68.1	65.5	62.5	65.7	70.0	64.1
WS	3.0	3.1	3.1	3.2	5.9	5.8	6.0
RF	3.3	4.5	3.2	3.4	3.6	4.7	3.4
O_3	15.1	18.0	18.3	20.1	17.3	16.1	18.1

T_h – maximum temperature (°C), T_a - average temperature (°C), SR – solar radiation (W m^{-2}), RH - relative humidity (%), WS - wind speed (m/s), RF – rainfall (mm), O_3 – ozone concentrations (ppb).

Table 3. GA-ANN models: their input variables, activation functions (AF), number of hidden neurons (HN) and number of model parameters (MP).

Model	AF	HN	MP
GA-ANN1			
$O_{3\|t+24} = \frac{1}{3} n_1(all\ va1a.) + \frac{1}{3} n_2\left(T_h, T_a, SR, RH, \frac{1}{RH}, RF, O_3\right) + \frac{1}{3} n_3(T_h, T_a, SR, WS, O_3)$	*net1* – radial basis *net2* – radial basis *net3* – radial basis	*net1* – 8 *net2* – 8 *net3* – 8	211
GA-ANN2			
$O_{3\|t+24} = \begin{cases} n_1\left(T_h, T_a, SR, RH, \frac{1}{RH}, RF, O_3\right), if\ O_3 \leq 34.6 \\ n_2\left(T_a, SR, \frac{1}{RH}, WS, RF, O_3\right), if\ O_3 > 34.6 \end{cases}$	*net1* – radial basis *net2* – radial basis	*net1* – 8 *net2* – 5	114
GA-ANN3			
$O_{3\|t+24} = \begin{cases} \begin{cases} n_1(all\ va1a.), if\ RH \leq 41.6 \\ n_2(all\ va1a.), if\ RH > 41.6 \end{cases}, if\ RH \leq 65.1 \\ \begin{cases} n_3(all\ va1a.), if\ SR \leq 197.4 \\ n_4(all\ va1a.), if\ SR > 197.4 \end{cases}, if\ RH > 65.1 \end{cases}$	*neti* – hyperbolic tangent	*neti* – 8	243

Table 4. Performance of achieved models in both training and test periods

Model	Training and validation					Test				
	MBE	MAE	RMSE	d_2	AIC	MBE	MAE	RMSE	d_2	AIC
$GA\text{-}ANN_1$	-0.02	2.57	3.62	0.94	5661	-0.02	1.05	1.49	0.96	714
$GA\text{-}ANN_2$	-0.03	2.52	3.52	0.95	5348	0.12	1.22	1.73	0.95	626
$GA\text{-}ANN_3$	-0.09	2.68	3.89	0.94	6014	-0.93	2.57	4.08	0.78	1509

MBE - mean bias error, MAE - mean absolute error, RMSE - root mean squared errors, d_2 - index of agreement, AIC - Akaike Information Criterion.

Figure Captions

Figure 1. Annual average profiles of O_3 concentrations (in ppb) during the analysed period.

Figure 2. Monthly average values of: (a) average temperature; (b) solar radiation; (c) relative humidity; (d) wind speed; (e) and rainfall.

Figure 3. Temporal variation of linear correlation between O_3 concentrations and: (a) average temperature; (b) solar radiation; (c) relative humidity; (d) wind speed; (e) and rainfall.

Figure 4. Model predictions of GA-ANN$_1$ and GA-ANN$_2$ during test period.

Figure 5. Combined effect of meteorological variables on daily average O_3 concentrations (ppb) according GA-ANN$_1$ model: (a) maximum temperature and solar radiation; (b) maximum temperature and relative humidity; (c) maximum temperature and wind speed; (d) solar radiation and relative humidity; (e) solar radiation and wind speed; and (f) relative humidity and wind speed.

Figure 6. Combined effect of meteorological variables on daily average O_3 concentrations (ppb) according GA-ANN$_2$ model for $O_3 \leq 34.6$ ppb (a, c and e) and $O_3 > 34.6$ ppb (b, d and f): (a) maximum temperature and solar radiation; (b) solar radiation and relative humidity; (c) maximum temperature and relative humidity; (d) solar radiation and wind speed; (e) solar radiation and relative humidity; and (f) relative humidity and wind speed.

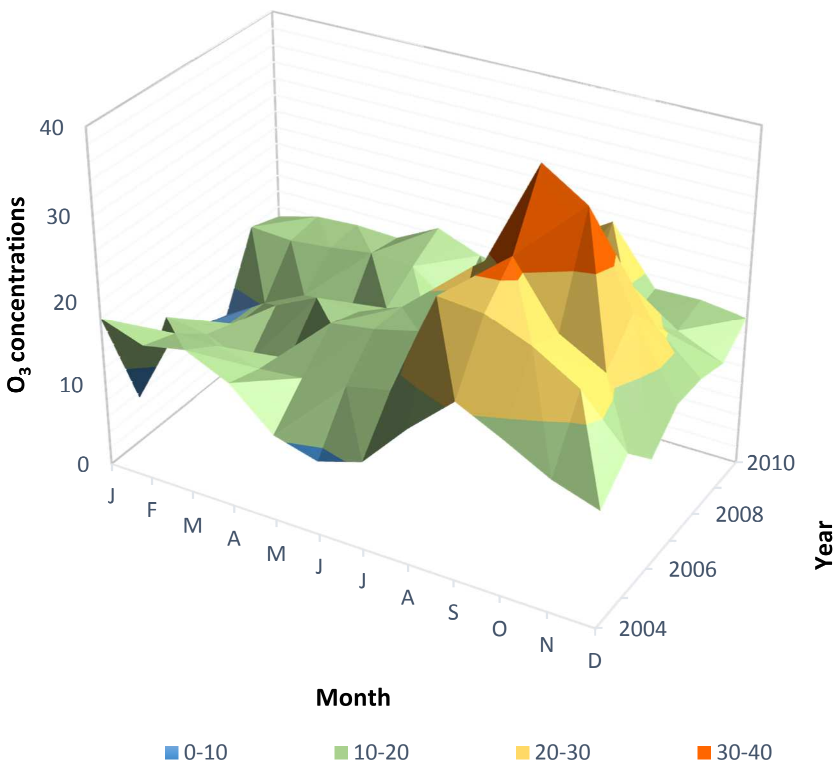
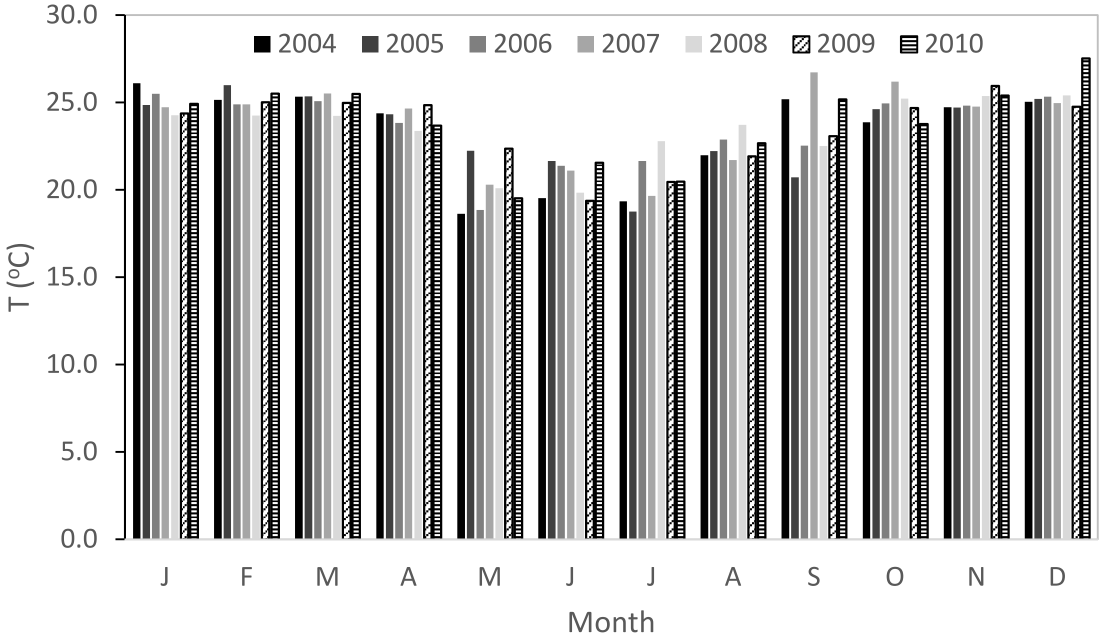
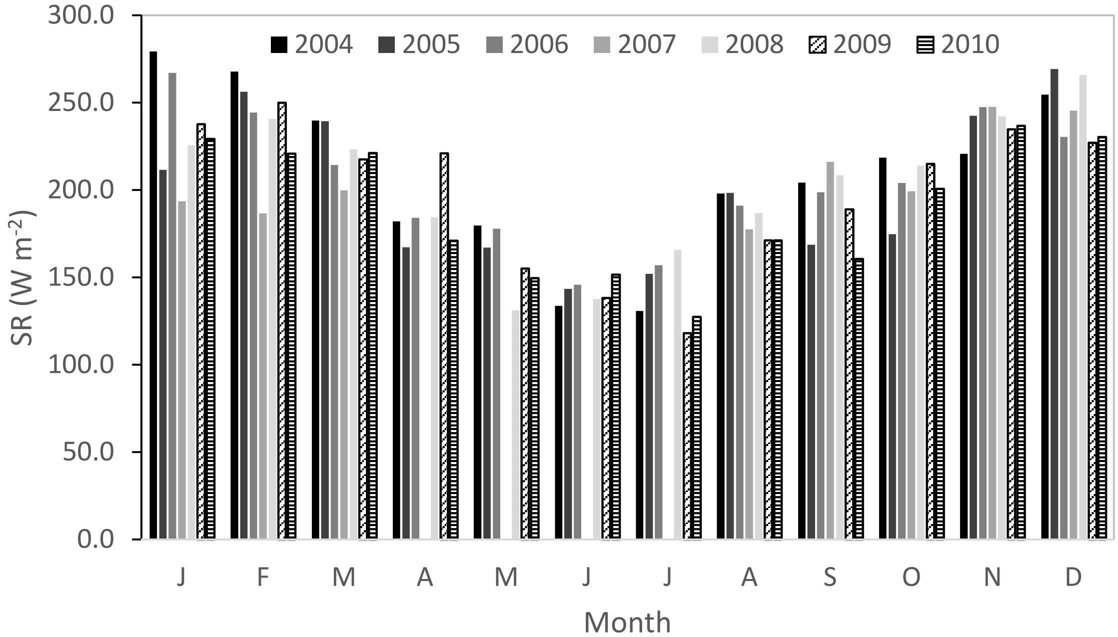

40
30
20
10
0
O_3 concentrations
J
F
M
A
M
J
J
A
S
O
N
D
Month
2004
2006
2008
2010
Year
0-10
10-20
20-30
30-40

Figure 1.



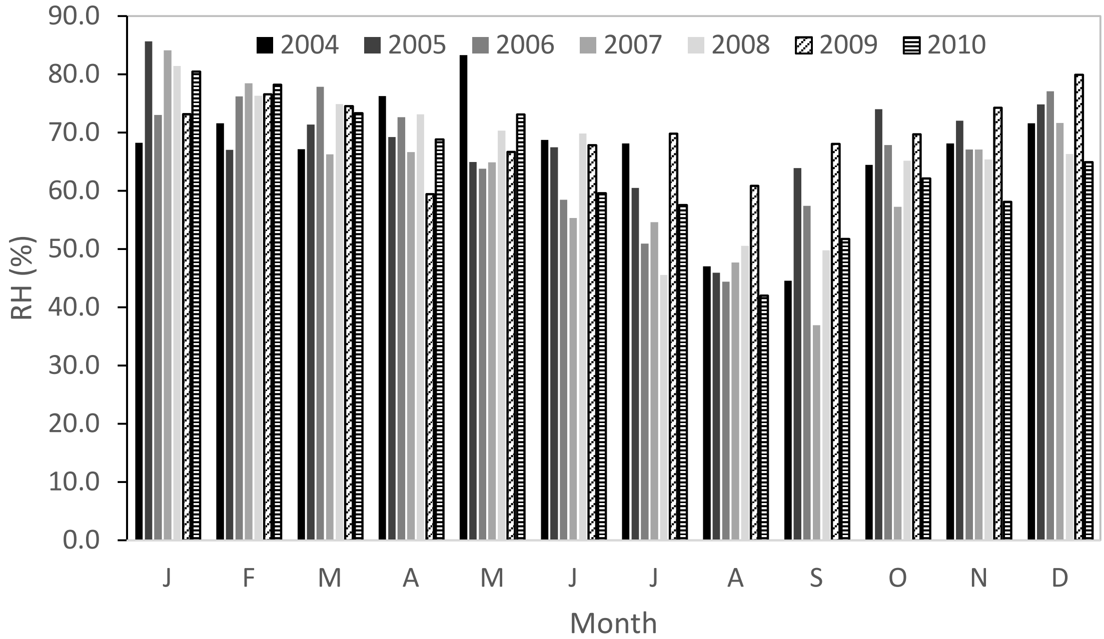
2004
2005
2006
2007
2008
2009
2010
T (ºC)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
J
F
M
A
M
J
J
A
S
O
N
D
Month

(a)



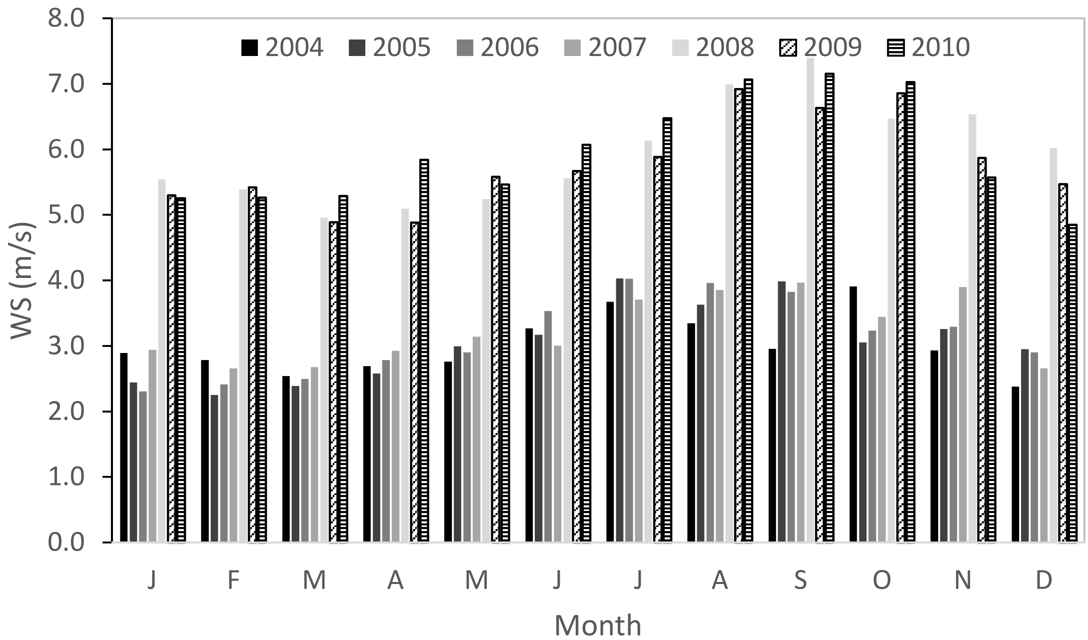
2004
2005
2006
2007
2008
2009
2010
SR (W m-2)
300.0
250.0
200.0
150.0
100.0
50.0
0.0
J
F
M
A
M
J
J
A
S
O
N
D
Month

(b)



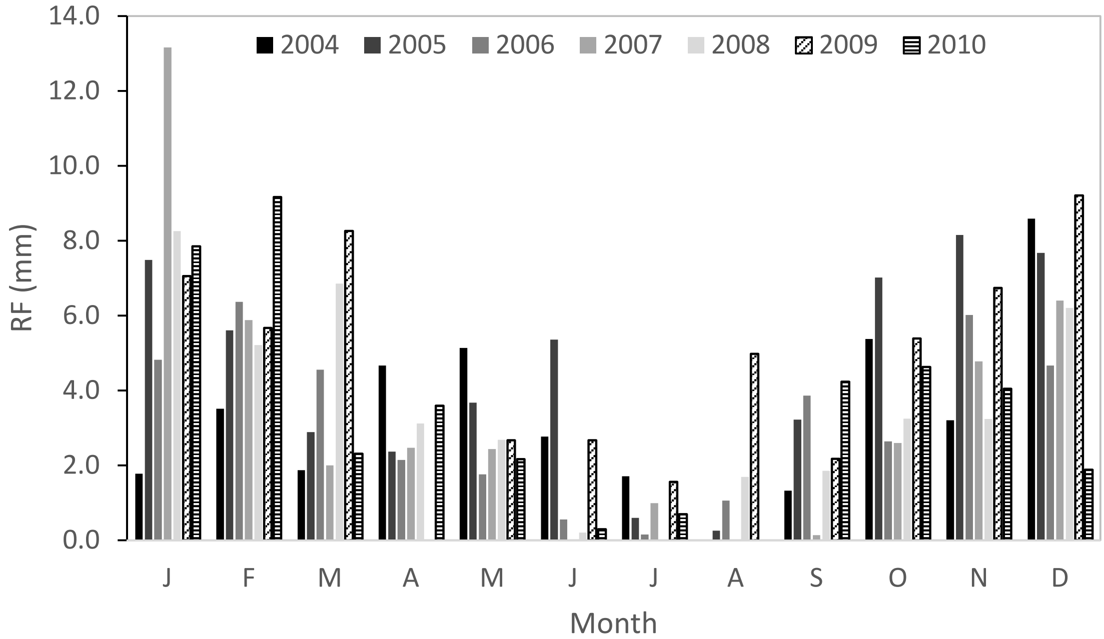
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
RH (%)
2004
2005
2006
2007
2008
2009
2010
J
F
M
A
M
J
J
A
S
O
N
D
Month

(c)



8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
WS (m/s)
2004
2005
2006
2007
2008
2009
2010
J
F
M
A
M
J
J
A
S
O
N
D
Month

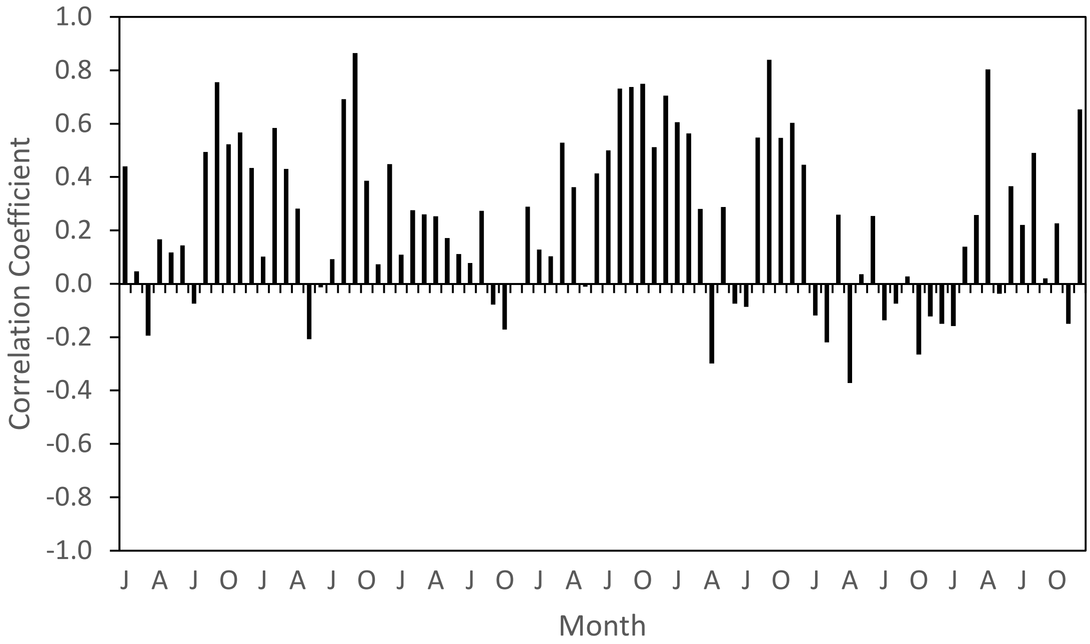
(d)



14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
2004 2005 2006 2007 2008 2009 2010
RF (mm)
J F M A M J J A S O N D
Month

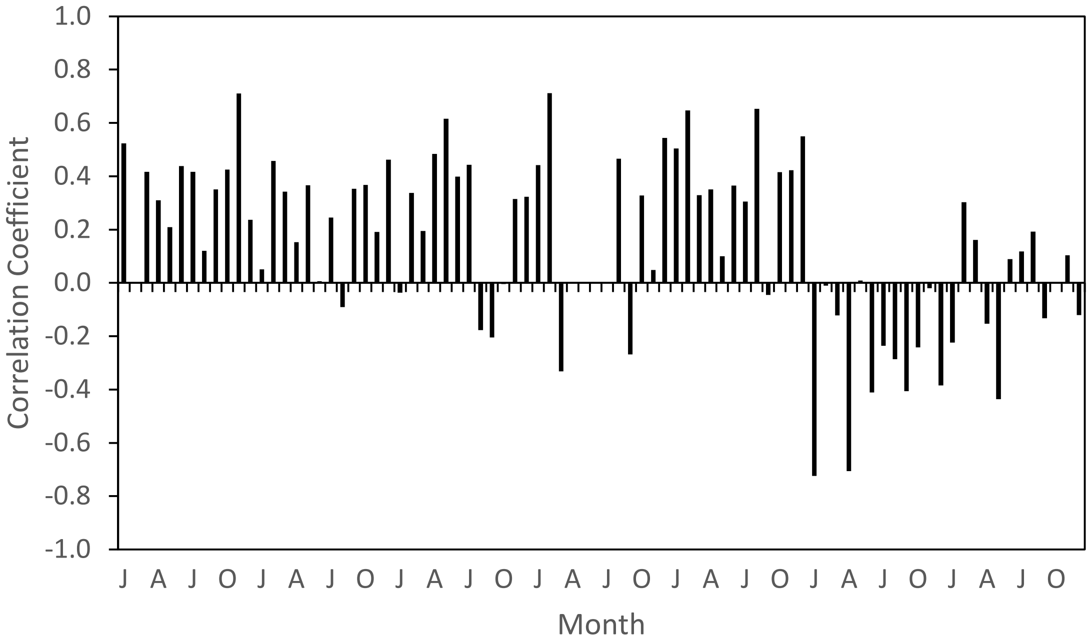
(e)

Figure 2.



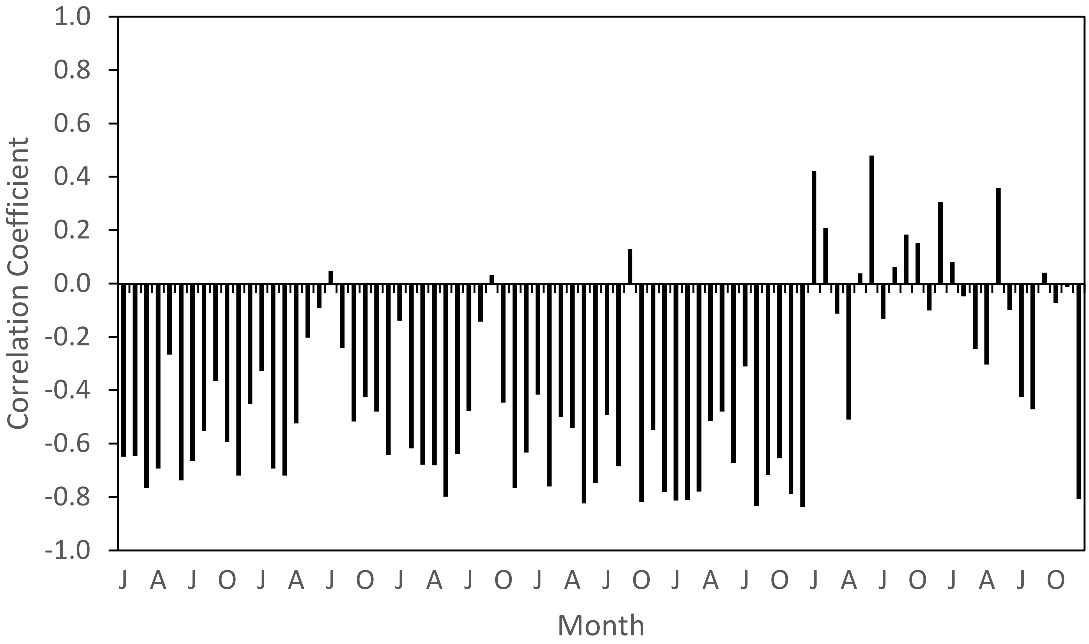
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Correlation Coefficient
J A J O J A J O J A J O J A J O J A J O J A J O J A J O
Month

(a)

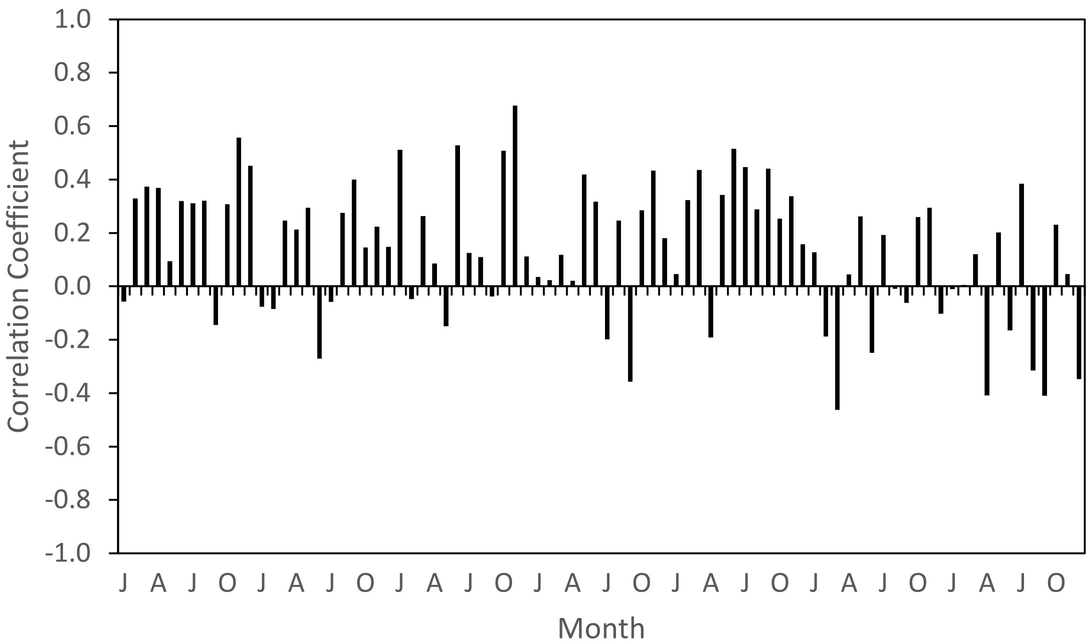


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Correlation Coefficient
J A J O J A J O J A J O J A J O J A J O J A J O J A J O
Month

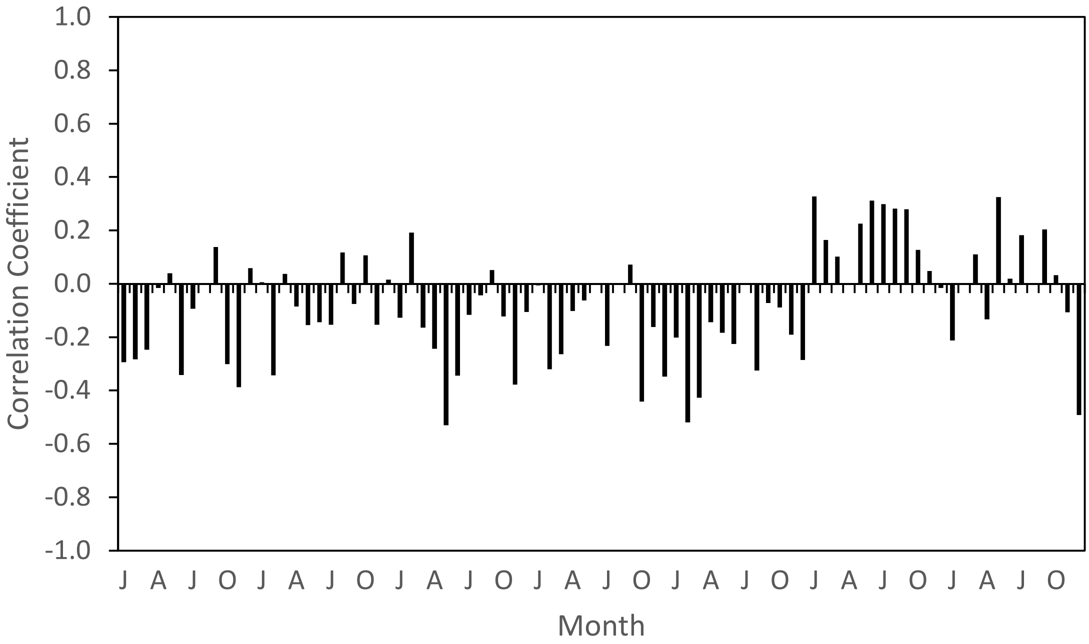
(b)


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Correlation Coefficient
J A J O J A J O J A J O J A J O J A J O J A J O J A J O
Month

(c)


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Correlation Coefficient
J A J O J A J O J A J O J A J O J A J O J A J O J A J O
Month

(d)


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Correlation Coefficient
J A J O J A J O J A J O J A J O J A J O J A J O J A J O
Month

(e)

Figure 3.

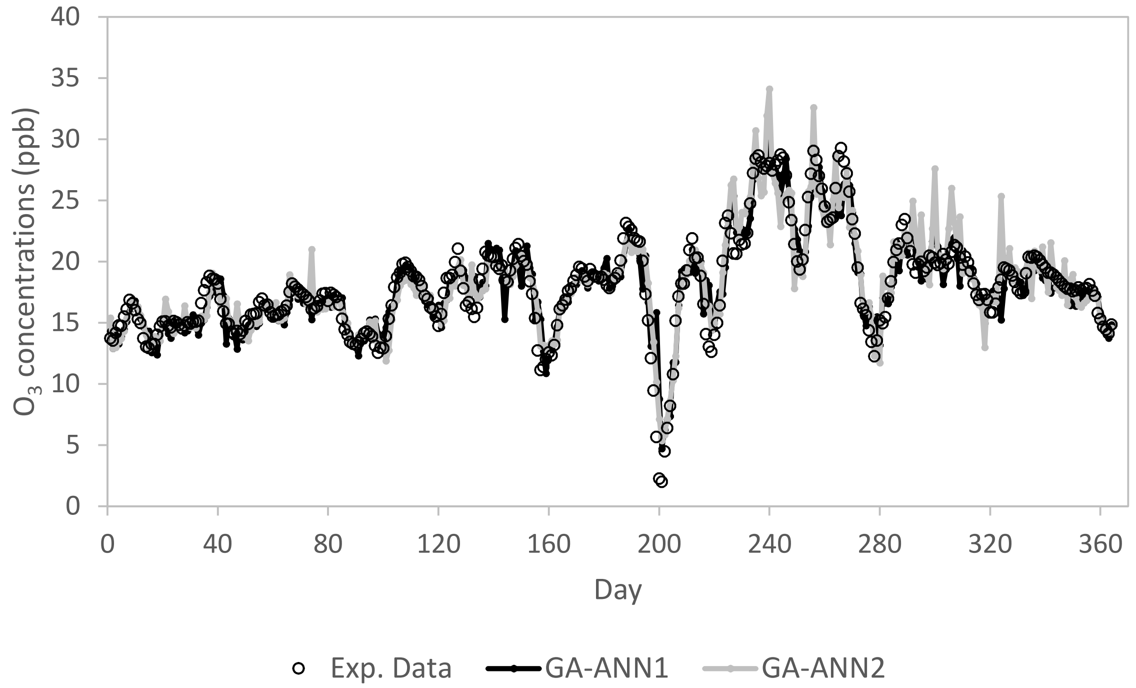


Figure 4.

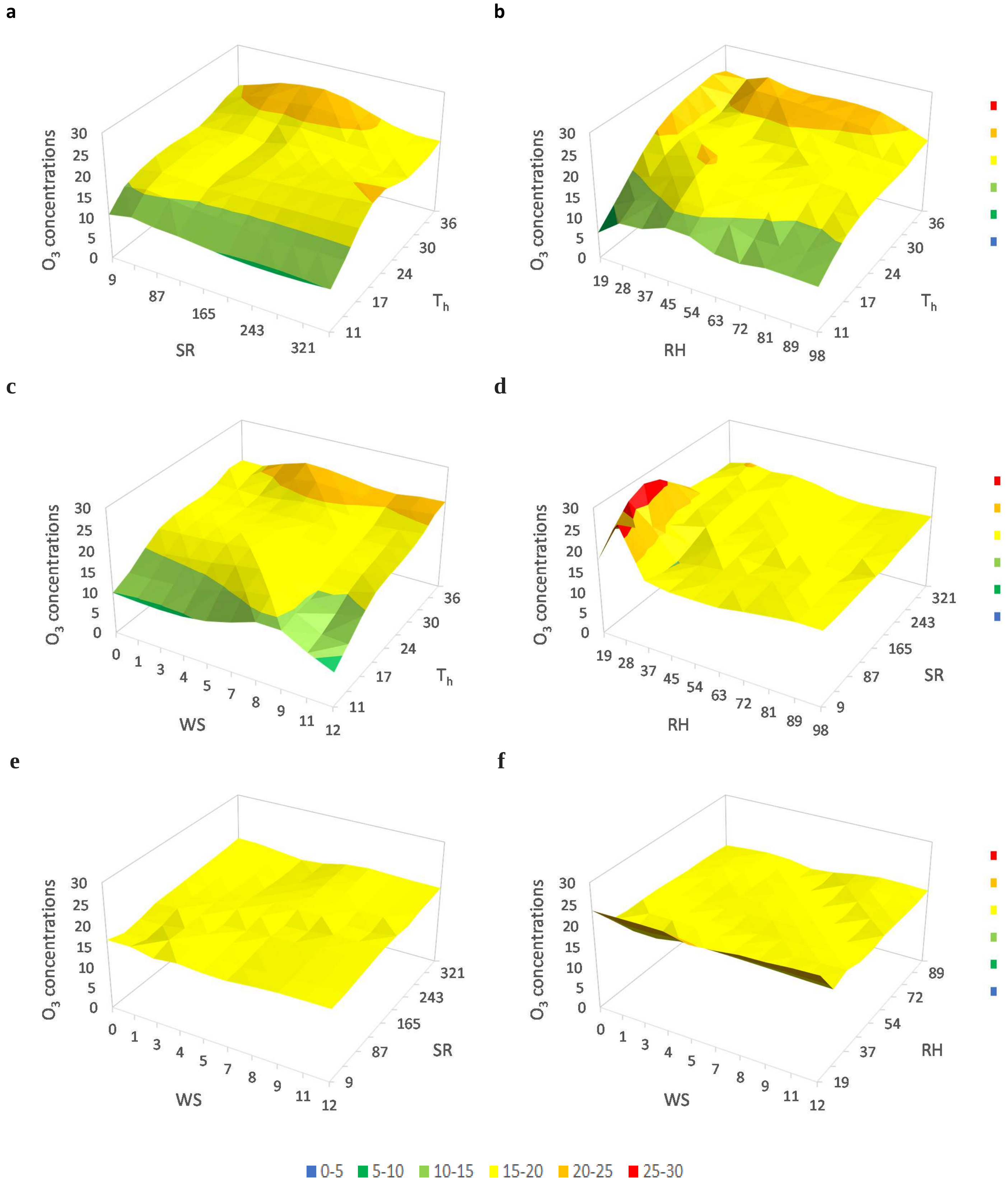

a
b
c
d
e
f
O3 concentrations
30
25
20
15
10
5
0
SR
9
87
165
243
321
Th
11
17
24
30
36
RH
19
28
37
45
54
63
72
81
89
98
WS
0
1
3
4
5
7
8
9
11
12
0-5
5-10
10-15
15-20
20-25
25-30

Figure 5.

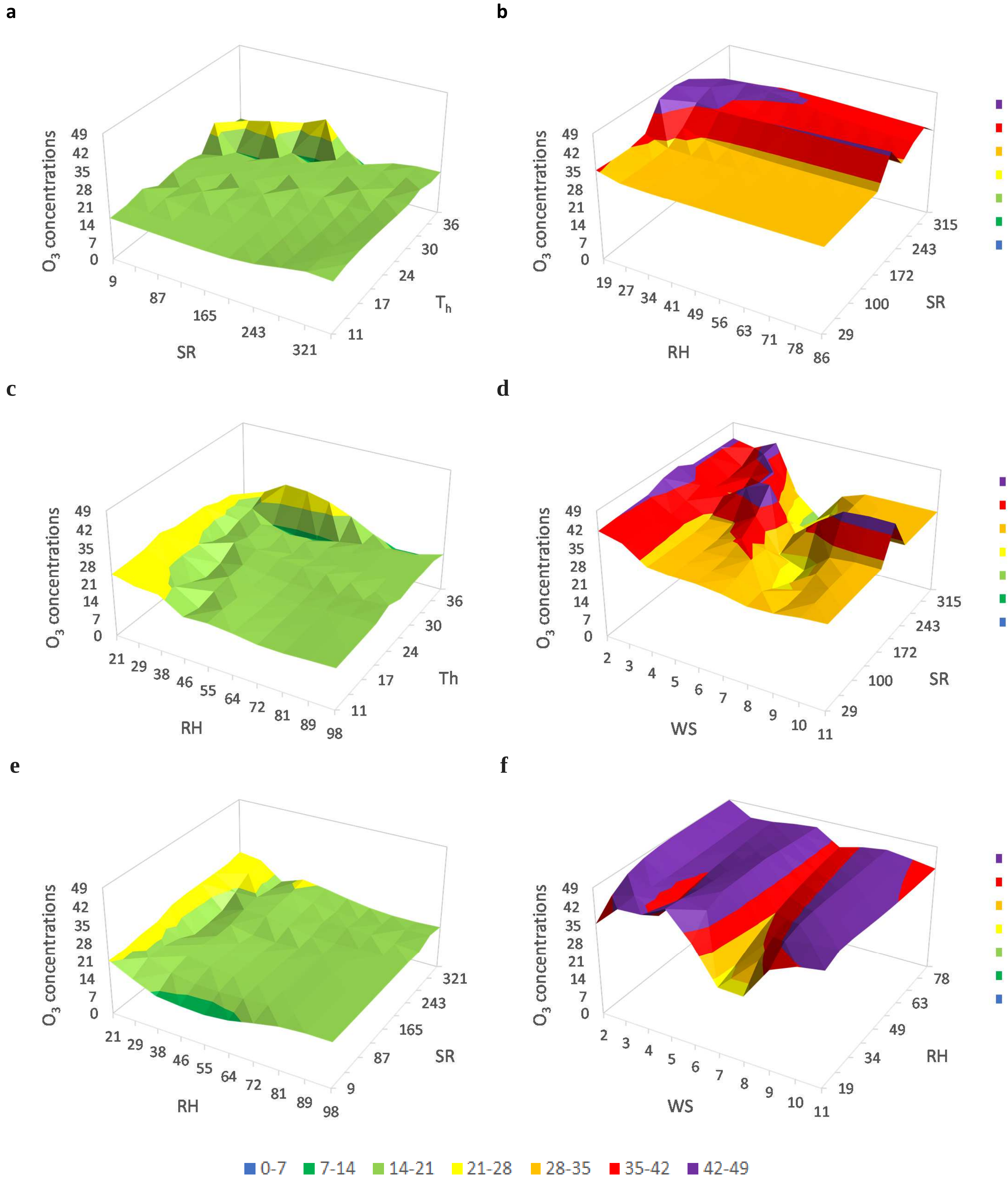

a
b
c
d
e
f
O3 concentrations
SR
Th
RH
WS
0-7
7-14
14-21
21-28
28-35
35-42
42-49

Figure 6.