PROGNOSIS OF MAXIMUM DAILY SURFACE OZONE CONCENTRATION WITHIN THE GREATER ATHENS URBAN AREA, GREECE

MOUSTRIS K.P.\textsuperscript{1}
PROIAS G.T.\textsuperscript{2}
LARISSI I.K.\textsuperscript{3}
NASTOS P.T.\textsuperscript{4}
KOUKOULETSOS K.V.\textsuperscript{5}
PALIATSOS A.G.\textsuperscript{3,}*  
\textsuperscript{1}Department of Mechanical Engineering, Technological Education Institute of Piraeus, Athens, Greece  
\textsuperscript{2}Department of Planning and Regional Development, University of Thessaly, 38334 Volos, Greece  
\textsuperscript{3}Laboratory of Environmental Technology, Electronic Computer Systems Engineering Department, Technological Education Institute of Piraeus, Athens, Greece  
\textsuperscript{4}Laboratory of Climatology and Atmospheric Environment, Department of Geology and Geoenvironment, University of Athens, Athens, Greece

Received: 01/11/2012  
Accepted: 01/10/2014  
Available online: 03/10/2014  
*to whom all correspondence should be addressed: e-mail: agpal@teipir.gr

ABSTRACT

In recent decades, there has been an increasing interest in the prognosis of maximum surface ozone concentrations due to the adverse effects on human health, animal population, agricultural productivity and forestry. The present study deals with the development and application of Artificial Neural Network (ANN) models in predicting the maximum daily surface ozone concentration in several locations within the greater Athens area (GAA), 24-hours in advance. Meteorological and air pollution data during the period 2001 to 2005 were provided by the network of the Hellenic Ministry of the Environment, Energy and Climate Change. Hourly values of barometric pressure and total solar irradiance for the same period have been recorded by the National Observatory of Athens. A training data set for the ANN prognostic model was generated by employing the superposed epoch analysis.

The evaluation of the performance of the developed model, using appropriate statistical indices, clearly indicates that the risk of surface ozone values exceeding the European Union (EU) threshold for human health protection can be successfully predicted. This suggests that the proposed ANN model can be used to issue warnings for the general public and especially certain sensitive groups of the population.

Keywords: ambient air pollution, surface ozone prediction, artificial neural networks.

1. Introduction

Ozone ($O_3$) is an allotropic form of oxygen. It is a relatively unstable gas, colorless, strong oxidizer, highly toxic, with a characteristic odor. It is slightly soluble in water and chemically unstable. The ozone layer in the stratosphere plays a critical role in protecting life on Earth by absorbing most of solar ultraviolet radiation. In the troposphere and mainly at ground level where life exists ozone contributes to poor air quality. Ground level ozone is a secondary pollutant produced from primary pollutants such as nitrogen oxides, hydrocarbons from car exhausts and industry and volatile organic compounds (VOCs). The primary pollutants, with the contribution of solar radiation, react with oxygen (photochemical reaction), especially in warm and sunny weather, to form ozone ($O_3$).
Short duration exposure (1-3 hours) and prolonged duration exposure (6-8 hours) in an ozone polluted environment can result in a number of health effects that are observed in broad segments of the population. Even healthy people will experience effects such as induction of respiratory symptoms, decrements in lung function and inflammation of airways. High daily ozone concentrations are associated with increased asthma attacks, increased hospital admissions and also increased risk of premature death from heart or lung disease. Ozone has also adverse effects on the environment, vegetation and the ecosystems. It causes deterioration of the productive capacity of the agricultural land and destroys the foliage of trees and the aesthetics of forests and parks.

The increasing surface ozone concentration level, in recent decades, has become a major concern worldwide. Numerous studies, over the last decade, refer to the effects of air pollution on public health (Künzli et al., 2000; Katsouyanni et al., 2003; Bartzokas et al., 2004; Paliatsos et al., 2006; Nastos et al., 2008; Nastos, 2008; Bosson et al., 2009; Tonne et al., 2010; Kalantzi et al., 2011). Surface ozone concentrations in the eastern Mediterranean region have been analyzed previously (Ziomas et al., 1989; Álvarez et al., 2000; Kalabokas et al., 2000; Kalabokas et al., 2004; García et al., 2005; Gerasopoulos et al., 2006; Paliatsos et al., 2008). Several studies have shown that the background ozone concentrations in the troposphere have more than doubled (Volz and Kley, 1988; Staehelin and Smith, 1991; Bonasoni et al., 2000). The increased photochemical ozone production observed in the Mediterranean region may be attributed to the high level of solar irradiance in combination with the emissions of anthropogenic ozone precursors. These precursors may be transported over long distances under certain meteorological conditions, resulting in surface ozone formation far from the sources (Bloomfield et al., 1996; Gardner and Dorling, 2000; Dueñas et al., 2002).

Air quality has emerged as a major factor affecting the quality of living in urban areas, especially in densely populated and industrialized areas. Air pollution control and legislation to regulate various types of pollution are necessary in order to prevent conditions becoming worse in the long run. At the same time, short-term forecasting of air quality is required in order to take preventive and evasive action during episodes of atmospheric air pollution (Lu et al., 2002). It could be possible to avoid excessive medication, reduce the need for hospital treatment and even avoid premature deaths by influencing people’s daily habits or by placing restrictions on traffic and industry. Thus, it is clear that an accurate ozone level prediction system will be a valuable tool. Numerous statistical and ANN models have been developed and tested in order to predict ozone concentrations (Spellman, 1999; Elkamel et al., 2001; Coman et al., 2008; Chattopadhyay and Chattopadhyay-Bandypadhyay, 2008; Ettouney et al., 2009; Zhang et al., 2010; Mahapatra, 2010; Feng et al., 2011).

In the present study, an ANN model was developed and evaluated in order to simultaneously predict the maximum daily 8-hour average values of surface ozone concentration for the next day in seven different regions within the GAA. The innovation of the current work is that the developed predictive model incorporates the key features of previously developed models and, additionally, provides simultaneously real time prediction for all the seven examined sites within the GAA. Therefore, there is no need for a separate model for each site.

2. Data and methodology

2.1 Area and data

The city of Athens is located in an area of complex topography within the Athens basin (∼450 km²) being the southernmost capital on the European mainland. Mountains bound the Athens basin with heights ranging from 400 to 1500 m at the west, north and east sides. Openings exist between these mountains at the northeast and at the west of the basin, while the sea extends southwards (Saronikos Gulf). The Athens basin has a southwest to northeast major axis and is bisected by a cluster of small hills. The prevailing winds blow from N and NE in late summer, fall and winter and from SSW and SW in spring and early summer. The NE and SW directions coincide with the major geographical axis of the basin. The ventilation of the basin is poor during the prevalence of local circulation systems, such as sea/land-
breezes (Larissi et al., 2010). The GAA, like most metropolitan areas in the world, faces severe air pollution problems due to high population density and the accumulation of major economic activities in this region. The contribution of the intense sunshine to the high levels of photochemical air pollution, especially during summer months, is significant. The air pollution problems are often exacerbated by factors that favor the accumulation of air pollutants over the city, such as topography (basin surrounded by mountains), narrow and deep street canyons and adverse meteorological conditions, such as temperature inversions, low wind speed, high air temperature, extensive periods of dryness (Larissi et al., 2010).

The developed ANN model is based on the patterns of ozone data from seven different regions within the GAA. The seven examined monitoring stations are Patission (PAT) in the city center, Galatsi (GAL), Maroussi (MAR), Lykovrissi (LYK), Liossia (LIO), Thrakomakedones (THR) and Agia Paraskevi (APA). Figure 1 shows a map of the GAA with the seven monitoring stations.

![Figure 1](image.jpg)

*Figure 1. The map of GAA with the seven examined monitoring stations*

The meteorological and air pollution data used in this study have been recorded by the network of the Hellenic Ministry of Environment, Energy and Climate Change (HMEECC). A more detailed description of the HMEECC network can be found elsewhere (Larissi et al., 2010). The meteorological data set include hourly values of air temperature, relative humidity, wind speed and wind direction over a five-year period, 2001-2005. The air pollution data set include the 8-hour average values of surface ozone concentration and the hourly NO$_2$ concentrations for the same period. Finally, hourly values of total solar irradiance and barometric pressure recorded by the National Observatory of Athens were used in the analysis.

2.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a branch of artificial intelligence developed in the 1950s aiming at imitating the biological brain architecture. They are an approach to the description of functioning of human nervous system through mathematical functions. Typical ANNs use very simple models of neurons. These artificial neurons models retain only the very rough characteristics of biological neurons of the human brain (McCulloch and Pitts, 1943). ANNs are parallel-distributed systems made of many interconnected non-linear processing elements (PEs), called neurons (Hecht-Nielsen, 1991). A renewal of scientific interest has grown exponentially since the last decade, mainly due to the availability of appropriate hardware that has made them convenient for fast data analysis and information processing (Viotti et al., 2002). Many ANN models have been developed in the last fifteen years for very different
environmental purposes (Nunnari et al., 1998; Prybutok et al., 2000; Heymans and Baird, 2000; Karul et al., 2000; Antonic et al., 2001; Kolehmainen et al., 2001; Balaguier Ballester et al., 2002; Schlink et al., 2003; Corani, 2005; Slini et al., 2006; Dutot et al., 2007; Papanastasiou et al., 2007; Moustris et al., 2010a; b; 2011).

The Multi-Layer Perceptron (MLP) is the most commonly used type of ANNs. Its structure consists of PEs and connections (Hecht-Nielsen, 1991; Caudill and Butler, 1992). The PEs are arranged in layers. The first layer is the input layer followed by one or more hidden layers and after that the final layer, which is the output layer. An input layer serves as buffer that distributes input signals to the next layer, which is a hidden layer. Each neuron of the hidden layer communicates with all the neurons of the next, if any, hidden layer, having in each communicating connection a typical weight factor. Each unit-artificial neuron in the hidden layer sums its input, processes it with a transfer function and distributes the result to the output layer. The number of hidden layers, all connected in the same fashion, may vary. The units-artificial neurons in the output layer compute the final output in a similar manner. The produced output value from the ANN is compared against a target value and an error is estimated. The previously described procedure, called a training cycle, is repeated. In each training cycle the values of the weight factors are modified in an orderly way until the estimated error is within acceptable limits, depending on the application. Since data flow, within the artificial neural network, from one layer to the next without any return path, such an ANN is defined as feed-forward ANN.

2.3 Methodology

The long-term objective of an environmental management system, with regard to surface ozone level in ambient air, is to limit the number of days with average ozone concentrations above 120 µg m⁻³. This is a target value for the protection of human health set out by EU directive (EU, 2002).

Data from days before the episode day, when violations of the target value occurred, were used to train the ANN predictive model. In order to estimate best the number of days that should be taken into account, daily maximum 8-hour averages of surface ozone concentration, during days with violations of the surface ozone threshold, were organized in superposed epoch analysis illustrations (Panofsky and Brier, 1968; Singh and Badruddin, 2006). These are depicted in Figure 2. The “zero” day represents the mean value of the maximum daily 8-hour surface ozone concentrations during days with violations at any of the measuring sites of HMECC’s network. The other days (named as -1, -2, etc) represent the mean values of the maximum daily 8-hour ozone for the same sites. Figure 2 indicates that when an air pollution episode has occurred there is a significant increasing trend of the maximum daily 8-hour mean values of surface ozone concentrations five days before the episode day. Furthermore, it appears that the phenomenon is smoothed out within four to five days after the episode day.

The best ANN structure was selected based on a set of exploratory experiments. For the chosen ANN architecture, the trial-and-error method was applied (Spellman, 1999; Elkamel et al., 2001; Ettouney et al., 2009; Mahapatra, 2010) that is, after a training period, the number of PEs in the hidden layer were increased or reduced until the smallest prediction error was obtained. The selected ANN model consists of one input layer with 32 PEs, one hidden layer with 4 PEs and one output layer with 1 PE. According to the above analysis, the input data necessary for training the ANN model are: the station number (1, 2, 3, 4, 5, 6, 7) and month number (1, 2, ..., 12), the maximum daily value of the 8-hour moving average of surface ozone concentration for the five previous days, the maximum daily value of NO₂ hourly concentrations for the five previous days, the mode daily value of the wind direction for the five previous days, the mean daily value of the wind speed for the five previous days, the maximum daily value of air temperature for the five previous days, the daily total irradiance for the five previous days, the mean daily value of barometric pressure for the five previous days, the mean daily value of relative humidity for the five previous days and finally, the maximum daily value of the 8-hour moving average of surface ozone concentration for the next day. The final result (target value) produced by the model is the maximum daily value of the 8-hour moving average of surface ozone concentration for the next day.

For each of the seven examined sites the available data set, from 2001 to 2005, was divided into two subsets. The first subset is the four-year period, from 2001 to 2004, and the second the one-year period.
for 2005. The data from the first subset was used for training the ANN model while the second data subset was only used for evaluating the model’s prediction accuracy.

![Graph showing 8-hour average surface ozone concentrations from HMEECC’s monitoring network seven days before and after the zero-day where an air pollution episode occurs during 2001-2005.](image)

**Figure 2.** 8-hour average surface ozone concentrations from HMEECC’s monitoring network seven days before and after the zero-day where an air pollution episode occurs during 2001-2005.

In order to evaluate the results and the predicting performance of the developed model, statistical indices such as the coefficient of determination ($R^2$), the mean bias error (MBE), the root mean square error (RMSE) and the index of agreement (IA) were used (Kolehmainen et al., 2001; Moustris et al., 2010a). The accuracy of the proposed prognostic model, to predict the days that surface ozone concentration exceed the EU’s threshold value of 120 μg m$^{-3}$, was assessed by using appropriate statistical indices such as the true predicted rate (TPR), the false positive rate (FPR), the false alarm rate (FAR) and the success index (SI) were applied (Moustris et al., 2010a).

### 3. Results

Table 1 presents the validation statistical indices between the observed and the predicted ozone concentrations for the seven examined stations and for the next 24-hours, respectively.

**Table 1.** Statistical indices for the evaluation of the ANN model forecasting accuracy, for 1-day ahead prediction (year 2005).

<table>
<thead>
<tr>
<th>Station</th>
<th>$R^2$</th>
<th>MBE (μg m$^{-3}$)</th>
<th>RMSE (μg m$^{-3}$)</th>
<th>IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>APA</td>
<td>0.622</td>
<td>+0.920</td>
<td>17.715</td>
<td>0.883</td>
</tr>
<tr>
<td>GAL</td>
<td>0.639</td>
<td>+2.204</td>
<td>18.649</td>
<td>0.893</td>
</tr>
<tr>
<td>LIO</td>
<td>0.727</td>
<td>+1.403</td>
<td>16.578</td>
<td>0.923</td>
</tr>
<tr>
<td>MAR</td>
<td>0.710</td>
<td>+2.112</td>
<td>16.946</td>
<td>0.912</td>
</tr>
<tr>
<td>PAT</td>
<td>0.226</td>
<td>+2.418</td>
<td>20.250</td>
<td>0.641</td>
</tr>
<tr>
<td>THR</td>
<td>0.728</td>
<td>+2.466</td>
<td>18.665</td>
<td>0.921</td>
</tr>
<tr>
<td>LYK</td>
<td>0.777</td>
<td>+0.928</td>
<td>20.076</td>
<td>0.929</td>
</tr>
</tbody>
</table>
Figure 3. Observed (blue line) and predicted (red line) maximum daily 8-hour surface ozone concentrations for 1-day ahead prediction for LYK (a) and PAT (b) stations (year 2005).

The coefficient of determination values indicates that there was a close agreement between the recorded and the predicted 8-hour surface ozone concentration values for all forecasting cases (p<0.01). More specifically, the coefficient of determination takes values between 0.226 (PAT) and 0.777 (LYK), while the index of agreement ranges between 0.641 (PAT) and 0.929 (LYK). This shows that the predicted values are close enough to the corresponding observed values. The best results appear to be from LYK’s input dataset while the worst from PAT’s input dataset.

The MBE values varied between 0.920 μg m⁻³ and 2.466 μg m⁻³, while the RMSE ones varied between 16.946 μg m⁻³ and 20.250 μg m⁻³. These values indicate a good agreement between the predicted and the recorded values in the majority of the examined cases.

The predicted and the observed time series of the daily maximum 8-hour average surface ozone concentration for the next day for the station of LYK (best prediction) and the station of PAT (worst prediction) are depicted in Figures 3a and 3b respectively.

LYK station is a suburban station, for which the ANN model had the best performance (Figure 3a, right panel) in predicting the maximum daily 8-hour average surface ozone concentration value for the next day. On the contrary, the city center PAT station is the one with the worst model performance (Figure 3b, right panel). However this might be attributed to the heavy traffic load of the region.

Table 2 presents the values of validation statistical indices of exceedances, a day when the maximum daily 8-hour moving average ozone’s concentration exceeds the threshold value of 120 μg m⁻³ (EU, 2002).
Table 2. Statistical indices for the evaluation of the ANN model forecasting accuracy to predict the exceedance days (1-day ahead prediction, year 2005).

<table>
<thead>
<tr>
<th>Station</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>W</th>
<th>TPR (%)</th>
<th>FPR (%)</th>
<th>FAR (%)</th>
<th>SI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>APA</td>
<td>112</td>
<td>13</td>
<td>40</td>
<td>186</td>
<td>89.6</td>
<td>17.7</td>
<td>26.3</td>
<td>84.9</td>
</tr>
<tr>
<td>GAL</td>
<td>12</td>
<td>28</td>
<td>11</td>
<td>305</td>
<td>30.0</td>
<td>3.5</td>
<td>47.8</td>
<td>89.0</td>
</tr>
<tr>
<td>LIO</td>
<td>21</td>
<td>30</td>
<td>45</td>
<td>255</td>
<td>41.2</td>
<td>15.0</td>
<td>68.2</td>
<td>78.6</td>
</tr>
<tr>
<td>LYK</td>
<td>80</td>
<td>21</td>
<td>27</td>
<td>223</td>
<td>79.2</td>
<td>10.8</td>
<td>25.2</td>
<td>86.3</td>
</tr>
<tr>
<td>MAR</td>
<td>25</td>
<td>32</td>
<td>31</td>
<td>268</td>
<td>43.9</td>
<td>10.4</td>
<td>55.4</td>
<td>82.3</td>
</tr>
<tr>
<td>PAT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>362</td>
<td>-</td>
<td>0.0</td>
<td>-</td>
<td>100.0</td>
</tr>
<tr>
<td>THR</td>
<td>96</td>
<td>8</td>
<td>30</td>
<td>205</td>
<td>92.3</td>
<td>12.8</td>
<td>23.8</td>
<td>88.8</td>
</tr>
</tbody>
</table>

According to Table 2, it appears that the developed ANN model shows a very good accuracy in predicting exceedance days. The TPR index shows that the model predicts correctly from 30.0% (GAL) up to 92.3% (THR) the observed exceedance days. The SI index, which measures the accuracy of the model to predict correctly if the next day will be or not an exceedance day, varies from 78.6% (LIO) up to 100.0% (PAT). The overall results show that predictive accuracy of exceedance days of the developed prognostic ANN model is very good.

4. Conclusions

Ozone is one of the main air pollutants that degrade the quality of atmospheric environment within the GAA. The objective of this study was to develop and assess an ANN model for 1-day prediction for the maximum daily 8-hour surface ozone concentration values simultaneously in seven different regions inside the GAA. The performed analysis shows that the coefficient of determination between the real and the predicted 8-hour surface ozone concentration values (0.226–0.777), for the year 2005, 24-hours in advance, are statistically significant (p<0.01). The index of agreement between the real and the predicted 8-hour surface ozone concentration values (0.641–0.929), for the year 2005, 24-hours in advance, indicate a very good agreement between prediction and observation. In addition, the accuracy of the developed model is confirmed by the fact that the maximum value of MBE is up to 2.466 μg m$^{-3}$ while the maximum RMSE value is 20.250 μg m$^{-3}$. Finally, the proposed prognostic ANN model shows a very satisfactory 1-day ahead prediction accuracy with respect to the the next day to be or not an exceedance day.

The performance of the ANN should be further evaluated. The model could give more reliable predictions for the maximum daily 8-hour surface ozone concentrations by increasing the available input data necessary for training the model and improving data quality (e.g. eliminate blank days). The model also did not take into account other environmental factors, such as heavy traffic, that certainly affects the results. One further direction that should be investigated could be to train the model using the extreme values only (e.g. those>100 μg m$^{-3}$). It is estimated that the success of the model would have been higher since the goal is to predict the episodes and the low 8-hour ozone concentrations.

REFERENCES


Caudill M. and Butler C. (1992), Understanding Neural Networks, Computer Explorations, MIT.


