**Atmospheric
Chemistry
and Physics
Discussions**

Use of neural networks for tropospheric ozone time series approximation and forecasting – a review

A. A. Argiriou

Laboratory of Atmospheric Physics, Dept. of Physics, University of Patras, Patras, Greece

Correspondence to: A. A. Argiriou (argiriou@physics.upatras.gr)

5739

Abstract

The use of artificial neural networks in atmospheric science expands constantly. During the last years, many papers were published dealing with air pollution modeling. A number of papers deals with the time series approximation and forecasting of tropospheric ozone concentration. Neural networks have been found to outperform other statistical techniques like multiple regression etc. This paper reviews and discusses some practical aspects of the proposed neural network models applied to ozone concentration approximation and forecasting.

1 Introduction

The need of as accurate as possible air quality models has been extensively analyzed in the past and no further arguments on the issue are necessary. Also well known are the impacts of increased tropospheric ozone concentration ($[O_3]$) on the environment and on human health. The complexity however of the mechanisms of ozone formation in the troposphere (Seinfeld and Pandis, 2006), the complexity of meteorological conditions in urban areas and the uncertainty in the measurements of all the parameters involved, make the fast and accurate modeling of $[O_3]$ by combining detailed meteorological and photochemical models very difficult. Therefore short term $[O_3]$ modeling used mainly for operational purposes, i.e. mainly for issuing warnings for high ozone concentrations in urban areas as foreseen by the environmental regulations of many countries, is based mainly on statistical and other techniques. Such models can give acceptable results, although following a "black box" approach (Robeson and Steyn, 1990).

Artificial neural networks (ANNs or briefly NNs) belong to this group of models. Artificial neural network models try to model the operation of biological neural networks and especially those of the human brain that can perform a series of tasks, like pattern recognition, classification etc., with an amazing speed and accuracy, despite the fact

5740

that inputs to the human brain consist most of the time of noisy and erroneous data (Kartalopoulos, 1996). NNs are a relatively new technique; the seminal paper on the issue was published by McCulloch and Pitts in 1943 (Anderson and Rosenfeld, 1988). Since that date NNs due to their potential and flexibility have been used successfully in many applications outperforming previously used methods. This created, sometimes an excessive optimism about their capabilities that together with their easiness (in the sense of user-friendly way) to apply offered by most of the currently used mathematical software packages, led to failures since some essential details that one has to take care of when applying NNs were overlooked and also assumptions that had to be verified prior to their applications, were not.

NNs were also applied in air pollutant time series modeling and air pollutant concentrations forecasting. A systematic flow of published papers on such applications starts in the early 1990s, boosted by the constantly improving performance and the decreasing cost of powerful computers in one hand and by the fact that software packages, commercial or open source, that required a minimum programming effort by non-specialists in NN algorithms were becoming widely available. Gardner and Dorling (1998), almost ten years ago now, published a comprehensive review of NN applications in atmospheric sciences. They focused on the multi-layer perceptron, one of the many existing NN model types (Kartalopoulos, 1996), practically the only that had been widely used by that time. Except of the detailed description of the multi-layer perceptron, the notions of the layers, their architecture and guidelines on their training and use, the paper by Gardner and Dorling (1998) provides important information on NN limitations, problems that may occur and suggestions for application solutions in practical implementations. However, related works published after that review article, do not always clearly state the necessary information about how NNs models they present were developed. Also other important details like i.e. data handling and normalization, use of different data sets for training and validation of the NN, transfer functions used, choice of the hidden layer structure etc., that would allow the replication of the experiment by other researchers, is frequently missing.

NNs, are developing constantly. Until the year 1998 the multi-layer perceptron was practically the only architecture widely used in air quality modeling. Later, papers appeared in which other types of NN models were applied for the same purposes. This paper reviews the models proposed in literature related to the important issue of ozone concentration time series approximation and forecasting using NNs that appeared until this date. The aim is to provide concise information regarding the NN types and architectures used for this purpose, and their results and also to provide a critical evaluation to some crucial points related to NNs and their use for the particular task.

In the present review we designate as forecast models those that using measured (or calculated) input variables at a given time step calculate [O_3] at future time steps. Those that use measured (or calculated) input variables at a given time step to calculate [O_3] at the same time step are designated as time series approximation models.

2   Neural network models

There are many excellent textbooks describing the concept of artificial neural networks. Here we will limit ourselves to a very short description aiming mainly to clarify the the notions and terms used in the remaining of the paper.

The basic module of the artificial neural network is the artificial neuron, shown in Fig. 1.

It has a set of inputs, each one of which arrives to the main processing element after being weighted by a weight factor w. The processing element has a bias term. The neuron produces a signal R when a threshold value Θ is reached or exceeded. The neuron output O is the product of the modulation of a non-linear function f on R. Many neurons together form a network and in the case of a network the neuron is referred to as a node. A network has three types of layers, the input, the hidden and the output layers. The number of layers may vary, as may vary also the number of nodes per layer. The number of layers and nodes per layer define the network topology. Learning is the process by which the network adapts its weight factors and biases in order to converge

to a desired response. There are various learning types and rules developed. The network training depends on the quality and the quantity of information presented to it through the training data sets. The initial performance assessment of the network is performed using data other than the training data, that consist the validation data set.

5 The final performance assessment is often performed by using a third data set. Characteristic artificial neural networks developed are known as paradigms or models. The most commonly used paradigm is the multilayer perceptron already presented. The advantages of neural networks are that they can be trained to approximate virtually any type of relations between input and output data. Their application does not require

10 any a priori knowledge regarding the statistical distribution of data. They can fit highly non-linear functions and can be trained to accurately generalize when fed with new not seen before, data (Gardner and Dorling, 1998). Also, neural network models can be programmed to train themselves periodically and therefore readapt their weights and biases in order to capture eventual new dynamics of the approximated system. This

15 feature is particularly useful in pollutant modeling and forecasting, since the emission characteristics may change with time, due to various reasons. Neural networks have also some disadvantages.They are unstable in the sense that there is no specific procedure allowing the selection of optimal individual models and their topologies; the empirical methods proposed in literature may lead to models that are not optimal. Also,

20 like most empirical models, NN tend to overestimate low ozone concentrations and to underestimate high ozone episodes. Finally neural networks constitute a "black box" approach and this hinders their use.

3   Time series prediction and forecasting applications

Ruiz-Suárez et al. (1995) apply two NN models, the Bidirectional Associative Memory

25 (BAM) and the Holographic Associative Memory (HAM) in order to approximate $[O_3]$ hourly time series. The data came from pollution monitoring stations of Mexico City, Mexico. The input variables to the BAM model were hourly averages of wind direc-

tion, wind velocity, ambient temperature, relative humidity, CO, SO_2, NO_2, NO_x and the month of the year. The data covered the period from January to May 1992. No information regarding data scaling is provided. Two types of BAM models were tested: One using the value of the above variables measured every day at noon at a given station

5 plus the month and having as output the corresponding $[O_3]$ value for this station and a second, having as inputs hourly values of the above data between 08:00 to 16:00 plus the month and as inputs the corresponding hourly $[O_3]$ values at two locations. The HAM models used as inputs the hour of the day, the month, the meteorological variables and the concentrations of the above mentioned substances in five stations

10 together with $[O_3]$ in four of the above stations, while the output is $[O_3]$ of the fifth station. Five HAM models have been tested calculating the $[O_3]$ at the fifth station 0, 1, 2, 3 and 4 h after the input variables are measured. The authors made several trials in order to optimize their models. The trials consisted to remove some of the input variables with qualitative criteria. The performance criteria used is only the comparison between

15 measured and predicted values in graphical form. No statistical indexes are used for this purpose. There is an acceptable agreement between forecasts and measurements for small forecasting horizons that worsens as the forecasting horizon increases. The paper presents also a short discussion over the advantages of the proposed model with respect to the multilayer perceptron.

20 Yi and Prybutok (1996) first address the problem of predicting the daily maximum 1h-average $[O_3]$ and compare the performance of the NN model to that of a linear regression model and of an autoregressive integrated moving average (ARIMA) model, using data from Dallas, Texas, USA. The input variables were average values between 06:00 to 09:00 in the morning of the forecast day of $[O_3]$, actual maximum daily temperature,

25 carbon dioxide, nitric oxide, nitrogen dioxide, oxide of nitrogen, surface wind speed and direction. Consequently the forecasting window is rather small. Also a dummy variable differentiating work day versus holiday was also used. The training data set covered the period from 1 June to 30 September 1993 because summer presents the worst case for increased $[O_3]$. The data of October 1993 and October 1994 were used to test

the model performance. We should point out here that no October period has been in-
cluded in the training data set and this lack of knowledge of the model may be reflected
to the validation results, since in October $[O_3]$ values are expected to be lower. No
information regarding data scaling is provided. A fully connected feed-forward network
was used, with nine input neurons – one for each input variable, four neurons in a single
hidden layer and one output neuron that produces the estimates for the daily maximum
$[O_3]$, i.e. a 9:4:1 structure. The initial number of hidden layers was taken equal to 3
using an empirical rule. The final number of hidden layers was determined by trial
and error. These tests allowed also the determination of the best combination of the
learning rate, momentum, number of hidden layers, number of hidden layer neurons,
learning rule and transfer function. The back-propagation learning algorithm was used
in the training process. The learning rules tested were the generalized and cumulative
delta rules. The results obtained were analyzed both graphically and statistically. The
NN results were compared to those of the regression and ARIMA models by means
of the Friedman test over the whole data set. A statistically significant superiority of
the NN model over the two other models at the 0.05 significance level was found. A
second exercise was made in order to assess whether the NN model can predict the
high ozone concentration episodes, in this case greater than 0.100 ppm. In this second
case only the NN and regression models participated. The assessment was performed
using the Wilcoxon single-rank test. It was found again that the NN outperforms the re-
gression model at the 0.05 level. However it should be pointed out that the forecasting
horizon of the proposed NN model is rather small, since input data need to be mea-
sured until 09:00 of the forecast day and the maximum $[O_3]$ value is expected to occur
shortly after noon, i.e. within a few hours.

Comrie (1997) compares the results of regression models and NNs under a vari-
ety of meteorological and ozone concentration conditions using ozone data from eight
U.S. cities. It should be noted here that the proposed NN models do not forecast ozone,
but perform a time-series approximation of ozone concentrations since the meteorolog-
ical inputs and the ozone output refer to the same day. The data cover again the period

5745

from May to September but now the time series is longer, from 1991 to 1995. The me-
teorological variables used are only four, namely the daily maximum temperature, the
average daily dew point temperature, the average daily wind speed and daily total sun-
shine duration. These variables were not selected by applying stepwise regression, as
it was the case for Yi and Prybutok (1996), or other statistical technique, but rather on
the theory of tropospheric ozone formation. The input data are scaled between 0.2 and
0.8, a common practice in NN modeling. Model results are reverse – scaled prior to
model comparison. Again the multilayer perceptron with the back-propagation learning
algorithm was used. Two NN models were developed, a 4:6:1 structure, using inputs
only the meteorological variables of the day and a 5:7:1 structure that uses as addi-
tional input the 1-h ozone maximum of the previous day. The number of neurons of the
hidden layer is determined by performing several trials. The output of both structures
is the daily 1-h ozone maximum. NNs and regression models are compared using
the mean absolute error (MAE), the root mean square error (RMSE), the Pearson's
product – moment correlation coefficient (R) and the coefficient of determination (R^2)
or explained variance. The author uses also the Willmott index of agreement d_i, i=1 or
2 (Willmott, 1981; Willmott et al., 1985) that measures the degree to which the predic-
tions of the model are error free. Comparison of the above statistical indexes showed
that the NN models outperform the regression models. The best performing model was
the NN having as an input also the ozone concentration of the previous day. However
both NN and regression models were found to underestimate high ozone episodes and
to overestimate low ozone values.

Nunnari et al. (1998) in their paper compare an ARMAX, a multilayer perceptron NN
and a fuzzy neural network. Fuzzy neural networks are a special category of neural
networks the weights of which represent a set of fuzzy rules (Kartalopoulos, 1996). The
developed schemes forecast ozone concentrations 1-h and 24-h ahead. The inputs are
measured concentrations of the following substances O_3, NMHK, NO_2, NO_2 and NO_x,
up to 3 time steps before the time step of the output value. A rather unusual feature of
the Nunnari et al. approach is that they do not use any meteorological parameter as

5746

input. The data comes from an industrial area in Sicily and cover a limited time period of three months which is not specified. No information regarding eventual scaling of the data and about the structure and other features of the NN is provided except that the back-propagation learning algorithm is used. The model performance is assessed by comparing the following indexes: the mean value and also the standard deviation of the residuals (between forecasted and measured values for each model), the ratio between the variance of the residual and the variance of the actual time series of a given pollutant, the correlation coefficient between the forecasted and the measured time series, the percentage of correctly forecasted values that exceed a given threshold and finally the percentage of wrongly forecasted values, i.e. values that are forecasted to exceed a given threshold while the corresponding measurement does not. The comparison of the above mentioned indexes revealed that the multilayer perceptrons outperformed the ARMAX and fuzzy neural network models.

Guardani et al. (1999) apply the multilayer perceptron NN and the back propagation learning algorithm both for time series fitting and forecasting using data from four air quality monitoring stations located at the metropolitan area of São Paolo, Brazil. The data used for training and validation of the NN models cover only two summer months of the southern hemisphere, namely October and November 1996. The authors state that in order to avoid extrapolations during the validation step, the test variables were kept within the range of the validation variables, but no more details on this are given. A number of 8 hourly averaged variables, both meteorological and chemical, were selected as inputs, namely air temperature and humidity, wind speed and direction and concentrations of CO, NO, NO_2 and NMHC. The authors proceeded to several exercises in order to develop NN models. The first exercise aimed at developing a model for fitting the ozone concentration time series. A 6:6:1 structure was adopted, but no information is given which are the two out of the eight input variables that were excluded. By applying the model to three stations they obtain a coefficient of determination R^2 between 0.91 and 0.93. No other statistical indicator is given. The second is a forecasting exercise. Using morning measurements of NO and NO_2 together with meteorological

5747

variables as inputs the authors developed models that forecast the $[O_3]$ in the afternoon. The resulted R^2 values were lower, between 0.6 and 0.7. Unfortunately no other information regarding the structure of these models is provided. The third exercise aimed at correlating ozone precursors generated in the city center where traffic is dense with meteorological variables and ozone concentrations measured in surrounding stations. The input variables were the morning averages of concentrations of NO, NO_2, NMHC, CO, wind velocity and wind direction measured at the city center station together with the afternoon averages of solar radiation level and temperature at each of the surrounding stations where the $[O_3]$ needs to be forecasted. Several intervals for calculating the morning and afternoon averages were tested and the final selection was from 8 to 11 h for the morning averages and from 12 to 17 h for the afternoon averages. A 8:8:1 model structure was adopted, the output node providing the forecasted $[O_3]$ value at the surrounding station the model refers to. The obtained R^2 is of about 0.85. An interesting feature of this paper is that the relative importance of the input variables is assessed from the sum of the absolute values of the weights between the input and hidden layers. In this particular case the variables in decreasing order of importance were found to be: temperature, wind direction, radiation level NMHC, NO, NO_2, wind velocity and CO. A final exercise was to use the NNs described above with the weights and biases calculated using the October and November 1996 data, to fit the time series of $[O_3]$ measured during October and November 1997. This resulted to R^2 values ranging between 0.73 and 0.80, showing a poorer performance with respect to that of the first exercise, but this is only to be expected since the number of training data (one value per input parameter per day for two months) is very limited to train a NN model of this structure, in order to approximate time series.

Spellman (1999) proceeds to an exercise similar to that of Comrie (1997) that is to compare regression and NN models, this time only to forecast daily 1-h maximum ozone concentrations and not for time series fitting. The data refer to five UK stations selected in order to cover a variety of location types (urban, rural, proximity to the sea, etc.), and climatic regions and cover the period from May to September for the years

5748

1993 to 1996. Among the available meteorological parameters the maximum daily temperature and the hours of sunshine were selected as input variables on the basis of stepwise regression. An additional input variable is the previous day's 1-h maximum ozone concentrations. No information regarding data scaling is provided. The multilayer perceptron with the back propagation algorithm was used. The models used sigmoid transfer functions and the delta rule gradient descent technique. The number of hidden layers was determined by trial and error procedure for each of the five stations. The final structures were a 3:3:3:1 (i.e. three nodes in the input layer, three in the first hidden layer, three in the second hidden layer and one node in the output layer) for three of the stations, 3:10:8:1 for the fourth and 3:6:9:1 for the fifth. The regression models used the same input variables. Many statistical indicators were used in order to assess the performance of the various models. It was found that NNs perform better than the regression models, but the improvement in the performance is not dramatic. Again the low ozone concentration values are overestimated and the high concentration values are underestimated. The R^2 ranges from 0.596 to 0.283 for the various stations while Comrie (1997) reports slightly better values between 0.37 to 0.69 for his stations. It should be reminded here that Comrie uses two additional meteorological variables as inputs, namely the average daily dew point temperature and the average daily wind speed, but only one hidden layer. Soja and Soja (1999) studied the possibility of building very simple linear and non-linear regression models for direct prediction of daily ozone exposure indexes like the 7-h mean (09:00–16:00 h), the daily integrated ozone dose AOT40 (accumulated over a threshold of 40 ppb during daylight hours) and the maximum half hour mean. Their aim was to use a limited number of meteorological variables as inputs to the model, namely the daily maximum temperature and sunshine duration. The data comes from a rural area in Austria located at a distance 30–50 km downwind of important industrial areas and covers the period from 1993 to 1995. Two NN architectures were used: the multilayer perceptron and the Bayesian networks. The optimal structure was found to be 2:2:1 in both cases. Three groups of models were developed, one for each input type. The models were assessed by comparing the sum

5749

of the absolute values of the residuals for each one of the calculated ozone indexes. No other statistical indicators were used therefore we can not proceed to any comparison with similar cases found in literature. The authors found that the performance of the NN models is comparable to and frequently worst than that of the individual functions, but state that the limitations of the neural network model is probably due to its simple structure. Unfortunately they did not test more complex structures.

Among the various NN models published the one that uses a maximum number of concentrations of various chemical substances is that of Hadjiiski and Hopke (2000). The dataset used included hourly averages of fifty-three hydrocarbon (C_2–C_{10}) compounds, O_3, nitrogen oxides (NO, NO_2, NO_x) from June through November 1993. The monitored meteorological parameters included air temperature, uv solar radiation, wind speed and direction. Again the multilayer perceptron architecture was used. Several NN models were developed. The first model performed a time series approximation of the [O_3] recordings. The structure of the model is 57:3:1. The input data were scaled within the interval [0.1,0.9]. This model was used in order to estimate the model response to uniform changes in the input variables. This sensitivity exercise revealed that the most important inputs were: cyclopentene and c-2-hexene, isopropylbenzene, n-propylbenzene, methylcyclohexane, n-nonane, NO, NO_2, temperature and solar radiation. Hadjiiski and Hopke proposed another NN model for forecasting ozone one-hour ahead [O_3] values. The inputs to this model were the ten variables (determined as important from a sensitivity analysis) measured one time step before the forecast horizon plus ozone concentration values two time steps before the forecast horizon. The optimal model structure was found to be 12:5:1. Model performance was assessed by means of the RMSE and R^2 which for the forecasting model was found to be 0.96. This value was by far the better reported in literature until this paper was published, it should be noted however that the Hadjiiski and Hopke forecasting model performs a 1-h ahead forecast using measurements of the current hour, while the other forecasting models already presented perform daily maximum 1-h average forecasts. The authors studied also the impact of the number of prior values of [O_3] on the accuracy of the model and

5750

found out that the addition of more than two prior values reduces the accuracy.

Cannon and Lord (2000) presented a detailed application of NN modeling to forecast daily maximum 1h-average ozone concentrations for 10 stations located in British Columbia, Canada. The authors tried to address the problem of model instability by means of bootstrap aggregation (bagging). This is an ensemble method that generates different training data sets by re sampling with replacement from the available cases. NN models are trained on the re sampled data sets and results from the individual members are averaged to provide the ensemble model outputs. Results of bagged NN models were compared to those of individual NN and regression models. The data sets used for training and validation covered the period from 1991 to 1996. Input variables were restricted to parameters commonly measured by surface observing stations. These included daily maximum temperature (for day 0, −1, −2 i.e. relative to the forecast day) in three locations, daily maximum temperature change (forecast day value minus previous day value for day 0, −1, −2) in three locations, surface wind speed at 18:00 UTC in three locations, cloud opacity at 18:00 UTC in three locations, sea-level pressure differences between three preselected locations at 12:00 and 00:00 UTC, the maximum ozone concentration of the previous day, the Julian day (in form of a trigonometric function) and a precipitation index. This gives 12 input variables in total. All of the parameters except the maximum ozone concentration of the previous day and the Julian day come from weather forecast models. The models were based on the multilayer perceptron architecture. The training algorithm was the resilient back propagation. Models were developed for ten locations where $[O_3]$ is monitored. For the performance assessment the slope and intercept of measured versus forecasted values are used, together with the MAE, RMSE and R^2. It was found that R^2 ranges from 0.27 to 0.74 depending on the model type and the location. The authors conclude that bootstrap aggregation increased both stability and accuracy of the multilayer perceptron NN models since bagged models outperformed the simple NN models in all cases.

Cobourn et al. (2000) compare the performance of a hybrid nonlinear regression

5751

model and a neural network, in forecasting daily maximum 1h-average ozone concentrations. Cobourn et al. make a clear distinction between O_3 forecasts where the input variables of the model are meteorological forecasts and O_3 hind casts where the input variables are measured. The NN models they propose use only meteorological variables as inputs, namely maximum daily temperature, dew point temperature (09:00–13:00 average) , wind speed (09:00–15:00 average), cloud cover (09:00–14:00 average), nighttime calms and precipitation. The validation data set covered the period 1993–1997, the validation and test datasets were measurements of the years 1998 and 1999 respectively. The NN model was developed with the 6:1:10:1 structure; the back propagation algorithm is used. The statistics used in order to assess the model performance were the absolute error (bias), MAE and RMSE and for the firs time in the literature under review, the detection rate (DR), defined as the percentage of the observed threshold exceedances detected by the alarms and the false alarm rate (FAR), defined as the percentage of alarms in which observed concentrations were below the alarm level. It should be noted that the error statistics were calculated separately for the overall results and for high $[O_3]$ predictions. The authors conclude that when the input variables to the regression and to the NN models were not measurements but forecasted values, both models showed practically a similar performance. When measured values were used as inputs the two models performed very close but the regression model was slightly superior: the DR was 92% for the regression model and 75% for the NN model, while the FAR was 35% for the regression model and 51% for the neural network. Prybutok et al. (2000) present a work quite similar to that of Yi and Prybutok (1996). They state that the models they proposed in 1996 were developed using data from Dallas-Forth Worth which did not contain any measurements higher than the EPA non-attainment level (0.120 ppm). Therefore in their new contribution they develop a NN, an ARIMA and a regression model for Houston where an important number of days above the threshold are observed. The NN architecture and structure and the type of input data is essentially the same as the one in Yi and Prybutok (1996). The validation data set covers the period from 1 June to 30 September 1994

5752

and testing was performed from 1 October to 10 October 1994. Despite the authors' statement that the motive of this new paper was to test their models for ozone levels higher than 0.120 ppm, the graph where they compare the performance of their model shows that the daily maximum [O_3] did not exceed 0.100 ppm during the test period. The MAE of the NN model for Houston and Dallas-Fort Worth is 0.012945 and 0.0064 i.e. almost double. The regression and the ARIMA model were found to perform also less good with the Houston data. Unfortunately no other parameters that would allow a more thorough comparison between the two sites are given. Statistical test showed again that the NN model outperforms the regression and ARIMA model.

It is sometimes necessary to identify long-term changes in the concentrations of pollutant time series that are independent of the meteorological fluctuations. Referring to tropospheric ozone, these long-term changes will reflect changes in ozone due to climate, policy or economic change. Gardner and Dorling (2001) apply NN models in order to derive these long-term changes and compare their method with another technique, based on time series decomposition using a filter. Gardner and Dorling train a multi-layer perceptron model to learn the short-term and seasonal components of ozone time series. Such a model does not have the means to take into account long term changes in the emissions. If such long term changes occur, then the model performance becomes invalid and the long-term changes will be reflected in the model residuals. The NN models developed had a structure 8:10:10:1. They were trained using the scaled conjugate gradient algorithm. The authors investigated the problem of how to split the available data set into a training and validation set. They tested two methods. The first consisted in training one network for each year in the time series, using the data of that year for the training and using the data of the subsequent year for validation. Then the network was tested over the whole time series and calculated the average residual from all networks. The second method consisted of training only one network over the whole time series except for one year which was used for validation. The model was tested over the whole time series to obtain the residuals. Although both methods were found to perform equally well, the second was chosen because

it maximized the use of the available data and could be more robust when dealing with missing data. The data used were daily maximum ozone concentrations from various US sites for the period 1984–1995. The input variables were the daily maxima of surface temperature, specific humidity, wind speed, opaque cloud cover and ceiling height. Additional inputs were the daily totals of global solar irradiance and the sine and cosine of the day of the year. It was found than NN models can be successfully used to identify long-term trends in ozone time series when compared with another statistical method.

Elkamel et al. (2001) propose a 13:25:1 multilayer perceptron model to fit a 5-min average ozone concentrations time series measured in an industrial area close to Kuwait City, Kuwait, during March and April 1995. The input variables consisted of simultaneous averages of the following variables: NMCH, CO_2, CH_4, CO, SO_2, NO, NO_2, suspended dust, surface air temperature, relative humidity, global solar irradiance, wind speed and wind direction. The 90% of the data series was used for training and the remaining 10% for the validation of the data set, but no information is provided regarding how and with which criteria this separation was made. Calculated values are compared to measurements by means of a cross plot graph and by calculating minimum, maximum and average absolute errors. No other statistical indicators are provided. Based on these indicators and graphs, the authors claim that there is a good agreement between measurements and predictions. It shoud be noted however that [O_3] measurements in the time series used did not exceed 60 ppb. In order to check the forecasting capabilities of their model, Elkamel et al. (2001) apply to it data measured two months later, in May and June 1995. An interesting point in this paper is the application of a method allowing to determine the relative importance of the various input variables to the output value. The major three influences in decreasing order were found to be those of CO, NO and ambient temperature. Solar irradiance is ranked almost in the end. This could be explained by the fact that under the Kuwait climate, temperature and irradiance levels are highly correlated. A year later was published another paper referring to ozone time series fitting in Kuwait City, this time the

data coming from the city center (Abdul-Wahab and Al-Alawi, 2002). The authors developed three NN models that had the same input variables as those of Elkamel et al. (2001), but the length of the measurements was even shorter this time: only June 1997 was used. The dataset was separated randomly in two parts. The first containing 85% of the data was used for model training and the remaining 15% for testing. The first model approximated the ozone concentration time series during a 24-h period. The second model took into account only those measurements recorded between sunrise and sunset, when maximum ozone concentration values occur and the third was developed to fit daily maximum ozone concentration values. These models were based on the back propagation scheme, but no further information regarding their structure is provided. The model performance was assessed by means of the mean square error, mean absolute error, correlation coefficient and coefficient of determination. Of course, an good agreement was found, as it was the case in all ozone time series fitting exercises already reviewed.

An interesting feature in Abdul-Wahab and Al-Alawi (2002) is the application of a method that uses the network weights in order to assess the importance of each of the input variables to the modeled output. Comparing the ranking of the input variables of the three NN models presented here, but also the similar exercise based on the same variables presented by Elkamel et al. (2001) coming from a nearby cite, no direct conclusion can be drawn. Viotti et al. (2002) develop NN models to predict various atmospheric pollutants measured in the city of Perugia, Italy. Their data cover the period 1997–1999. They developed modes for short-term (i.e. 1-h ahead), middle (next day) and long-term (next week) forecasts. The NN models were applied to ozone only for middle and long term forecasts. The input variables are meteorological parameters, namely surface air temperature, relative humidity, pressure, wind speed, rainfall amount and solar irradiance and also traffic density data. No ozone precursors were used. Two categories of models were tested: one using as inputs hourly values of the above mentioned variables and a second, using daily averages. The multilayer perceptron architecture with the back propagation learning rule was used. The model has

a structure of 7:5:1. The only statistical indicator provided is the mean standard error. Unfortunately this is expressed in relative concentration units and therefore is not comparable with similar exercises. For a 48-h test data set, ozone measurements ranging approximatively between 0.1–0.45 (relative values) are compared with NN forecasts. The MSE given is 0.126. For a 500-h data set [O_3] ranged between 0.1–0.7 and the MSE was 0.196, which is not negligible. Furthermore from a graphical comparison provided in the paper of Viotti et al. (2002), it can be seen that NN models do not manage to forecast most of the highest [O_3] values.

An innovative approach dealing with a 24-h ahead forecasting of hourly averages of [O_3] is presented by Balaguer-Ballester et al. (2002). Their data comes from three sites in Spain, covering thus both rural and urban environments. Two NN architectures are used, the multilayer perceptron and the finite impulse response. The models were developed with the aim to predict 1 and 2-days ahead hourly [O_3], using an iterative procedure that takes into account the predicted values at time t-1 in order to predict the output value at time t. This procedure is repeated until [O_3] at time t+24 is forecasted. The model performance is assessed by comparing the RMSE, MAE, coefficient of determination and also the d_1 and d_2 indexes of agreement (Willmott, 1981; Willmott et al., 1985). The R^2 values for the MLP network range between 0.75–0.81, the lowest value corresponding to data from an urban site and the highest to the rural one. The same parameter for the finite impulse response network ranges between 0.73–0.71, the highest and the lowest value occurring both at the two urban sites. Wang et al. (2003) employ neural networks in order to forecast maximum [O_3] in Hong Kong. They apply an improved concept of the radial basis function (RBF) which is the adaptive radial basis function (ARBF) network. RBF networks are considered as effective for fast learning in feed-forward networks, requiring less computing time compared to the back-propagation algorithm. Since an RBF disposes only one hidden layer, this architecture may not be adequate for accurate modeling in all purposes. In this case the adaptive RBF network may be used, that can determine the number of hidden nodes dynamically. Wang et al. (2003) provide in their paper a detailed discussion regarding

these two issues. The inputs to the ARBF model were chosen among the available parameters by performing a correlation analysis in order to assess the major factors affecting [O_3] in the area of interest. These were found to be three pollutants ([O_3], [NO_2] and [NO_x]) and three meteorological variables, namely wind speed, solar irradiance and ambient temperature. The effectiveness of the ARBF model is also assessed by comparing its performance to that of a general RBF model. The later uses the same inputs plus [CO] and wind direction, according to the primary principle of ozone formation. Both models forecast the next day's maximum 1-h average [O_3]. Hourly data from 1999 were used as test samples and those from 2002 for validation. The ARBF approach was found to perform better compared to the general RBF model. Unfortunately, the only statistical indicators provided by the authors are the minimum, maximum and average errors and therefore not directly comparable to other models reported in literature. However some judgment can be made by the following numbers: the maximum [O_3] levels in the Tsuen Wan area during 1999 and 2000 was 201 $\mu g\,m^{-3}$, while the annual averages were 56.8 $\mu g\,m^{-3}$ and 49.3 $\mu g\,m^{-3}$ respectively, while The maximum absolute error of the ARBF predictions was 113.9 $\mu g\,m^{-3}$, the minimum 0 $\mu g\,m^{-3}$ and the average absolute error 23.2 $\mu g\,m^{-3}$.

Chaloulakou et al. (2003) compared neural networks and regression models for forecasting the next day's maximum hourly ozone concentration in Athens, Greece. They adopted the multi-layer perceptron architecture with the feed forward algorithm. Training was performed using the Levenberg-Marquard algorithm. Their data base includes data sets only for the high season period, namely from April to October, for the years 1992 to 1999. Four data sets were used, each corresponding to a different air pollution measurement station of the Athens area. Data of each station were split into a test set, comprising the two thirds of randomly selected data of the 1992 to 1998 period, a validation set comprising the remaining one third of this data period and a training data set comprising measurements of the year 1999. The adopted neural network has 11 nodes in the input layer, and one output node (the maximum hourly [O_3]. The input data include the following parameters: the morning wind speed defined as

5757

the average value between 07:00–10:00 h, the nocturnal wind speed defined as the value at 14:00 h, incoming solar energy per unit area between 10:00–14:00, ambient relative humidity between 10:00–13:00, temperature at the 850 hPa isobaric level (obtained from radiosonde data at 14:00), temperature change from the previous day at the 850 hPa isobaric level, ambient temperature range, wind direction between 13:00 to 14:00 in form of a wind direction index and the hourly maximum [O_3] of the previous three days. The authors claim that these were selected as the most important parameters accounting for the variability in the summertime hourly maximum [O_3], through statistical and graphical analysis, without naming the methods used. The optimum number of the hidden layer nodes was found to be equal to five for one of the stations and equal to six for the other three stations. The performance of the derived models was tested using wide range of statistical indicators. The authors present an interesting table that compares the values of the various statistical indicators derived in their study with those that could be found in common in other papers aiming also to forecast the next day's maximum hourly [O_3]. The performance of the proposed NNs is compared to that of multiple linear regression models developed in the same study. Overall, NNs were found to outperform multiple linear regression models and present a similar performance to those NN models developed for areas having different topography and emission characteristics.

Heo and Kim (2004) use in their work several NN models in order to forecast daily maximum ozone concentrations at four monitoring sites in Seoul, Korea (Heo and Kim, 2004). Here the NN models are not used autonomously, as it was the case in all papers reviewed until now, but as part of a method that includes also a fuzzy expert part. It is out of the scope of this review to consider also fuzzy logic based forecasting systems, therefore we will limit our presentation to the neural part of this method. The multilayer perceptron has a 36:36:1 structure. The inputs are given at various time intervals and at several times. The 1-h interval data comprise [SO_2], [CO], [NO_2], [O_3], wind speed and direction and solar energy per unit area. Ambient temperature and relative humidity are given at a 3-h interval. Temperature, wind speed and direction at the 500 hPa isobaric

5758

level are used at a 12-h interval. The NN models are not used directly to forecast the [O_3] values, but as intermediate estimators. Therefore the comparison of their performance to that of other models appearing in literature is meaningless. Pastor-Barcénas et al. (2005) use the multilayer perceptron architecture combined with the back propagation training algorithm, in order to approximate surface [O_3]. They use a 20 days long hourly data set (1–20 April 2002) for training and the following 10 days hourly data set for validation purposes. The inputs include [NO], [NO_2], wind speed, temperature, pressure, solar irradiance and relative humidity. The NN performance is assessed by calculating the correlation coefficient, coefficient of determination, mean absolute error, root mean square error and the index of agreement. The optimum model was found to be that having 12 hidden neurons. Despite the acceptable values of these statistical indicators, the model during the 10-days validation period, failed to simulate values higher than 60 ppb and in a particular day those higher than 50 ppb.

Gómez-Sanchis et al. (2006) use NN models in order to assess the relevance of various input parameters used in order to forecast 1-h ahead [O_3] values. Three data sets comes from the area of Valencia, Spain are used. Data cover the month of April for the years 1997, 1999 and 2000. The first 20 days of each month are used as a training data set and the remaining ten days as a validation data set. From the graphs provided in the paper, it can be seen that the maximum [O_3] value does not exceed 70 ppb. Inputs include [O_3], [NO], [NO_2], ambient temperature, wind speed, relative humidity, solar radiation and barometric pressure. Their selection is based on literature survey. The multilayer perceptron is used and trained with the back propagation algorithm. The optimal network architecture was found to be 8:7:1, for 1997, 8:14:1, for 1999 and 8:16:1, for 2000. Since the number of nodes in the hidden layer varies with time, the proposed model is suitable only for time series approximation only and lacks of any generalization possibilities. The NN performance is assessed using statistical indicators. Although the authors claim that the optimal network matches actual observations very appropriately, from graphs it can be seen that although the proposed models are able to follow the diurnal [O_3] variation in general, often they cannot repro-

5759

duce the extreme values, either maxima or minima, of the time series. This is only to be expected since the number of data used for training is rather limited and consequently, extreme concentration values is even lower for the NN to acquire the necessary knowledge. The assessment of the impact of the various input values to the model output is assessed via a sensitivity analysis, which is to remove one input value each time and compare the result of the new model with that having the complete set of input values. The importance of the removed value is then considered to proportional to the observed difference. Sousa et al. (2007) developed a NN model based on the multilayer perceptron in order to predict the next day hourly [O_3]. The inputs to the model are [O_3], [NO], [NO_2], ambient temperature, wind velocity and relative humidity. The NN model is compared to a multiple linear regression model. Data comes from an urban air quality measurement station of Oporto, Portugal. The dataset covers only one month, July 2003, the 26 first days of which are used for training and the remaining 5 days for validation purposes. An innovative approach is also used: a second NN model has been developed the inputs of which are not the environmental parameters described above, but their principal components. This is done in order to avoid the problem that may occur in NN performance when some of the input variables are highly cross correlated. It was found that NN perform better compared to the linear models and that the principal component NN model performs better compared to the common one. However, as it can be seen from the graphs presented in this paper, no model can follow the observed [O_3] time series in a satisfactory manner and the improvement from introducing the reduced inputs is rather not dramatic. Finally, all models fail to predict values higher than 100 $\mu g\,m^{-3}$. Here again the problem seems to be the size of the training data set and it would be of interest if the principal component NN model development concept could be applied using a significantly longer data set.

5760

4 Conclusions

In this paper we reviewed a series of papers presenting artificial neural network models for tropospheric ozone concentration modeling. The works reviewed can be classified in two categories. The first comprises models used for [O_3] time series approximation. The second comprises models that can be used to forecast [O_3] values. The models in this second category have not all the same forecasting horizons; this may range from 1 h to 1 day. Many papers refer to "ozone prediction" but we consider the word "prediction" here to be somehow misleading since it refers more to forecasting, but in most of the cases it was used to describe time series approximation models.

Models have been developed based on data from several locations all over the world from various sites, urban, suburban, rural and industrial, where ozone concentrations vary significantly. A major problem is related to the size of the data bases used for validating but mainly for training the NN models. By reviewing the literature it was found that this may vary from some weeks to several years. Others refer to year-round measurements and others to the so-called high ozone season, i.e. from April to September or October for the northern hemisphere. It is well known from the NN theory that the more input variables are used, assuming that the desired model output is only ozone concentration, the more complex the network architecture has to be (Cybenko's demonstration that a neural network with only one hidden layer is able to approximate any function provided that there are enough nodes in this one hidden layer (Cybenko, 1989), is often not valid in practical implementations as is shown by various NN modeling optimization exercises) and therefore a longer training time series is needed for the network to learn from it. Some models fail to perform not because of errors in their concept, but because they were not presented with the required number of training cases to learn from. Another important point for selecting the time series: the model developer should take care, depending on what the proposed model is intended to do, that the selected time series for training and validation, include a fair number of cases that the model will be confronted to when applied. If, for example, the model is

to be used for public warning, the time series used should include enough high ozone episodes. The majority of the proposed models is based on the multilayer perceptron. Among the 23 papers reviewed here the non multilayer perceptron models found were one bidirectional associative memory network, one holographic associative memory network, one bayesian network, one radial basis function network and one adaptive radial basis function network. However, from the information provided, one cannot arrive to a general conclusion regarding the relative performance of these different types of models.

Most of the models use both meteorological and pollutant measurements as inputs. Only one model was identified that relied only on pollutant measurements. Also only one model used as inputs only meteorological variables. The meteorological variables used in most models are commonly measured in meteorological and air pollution stations. In one case upper atmosphere radiosounding data were also used. The concern of most authors however is to develop models having as inputs easily available variables. An important issue is the selection of input parameters. In the relatively early stages of NN model development, the selection was based on the theory of tropospheric ozone formation, i.e. a somehow empirical approach, since not all variables governing such a complex process are available. Later, statistical techniques like stepwise regression started to be used for this selection, the main aim being to avoid the use of strongly cross-correlated input parameters and therefore to limit the number of inputs, that leads to a more simplified model structure. When the number of parameters measured at a specific site is small, then the modeler has no other choice but to select them all as inputs. An important issue when using NN models is the normalization of the input parameters prior to their use. This is done in order to avoid discontinuities, which could affect the overall model performance. when using some nonlinearity functions. These nonlinearity functions range usually between [−1,1] or [1,0] or ±1/2. It is therefore recommended that the inputs should be normalized within these ranges prior to their use. However the information on whether this normalization was performed or not and the normalization range could be found in few papers only.

A complex issue is the selection of the model structure. Each model has at least three layers, the input, the hidden and the output layer, each one having a specific number of nodes. The selection of the number of nodes for the input and output layer is obvious. The number of the input nodes equals the number of the input parameters and the number of the output nodes is usually one, that of the $[O_3]$ to be simulated or forecasted. The number of nodes in the hidden layer and also the number of hidden layers has to be determined by the modeler. A novice neural network model developer when studying the literature he will find contradictory information regarding the same topic. As an example Guardani et al. (1999) in their paper explain why they used a single hidden layer and not a multiple one, by provide a reference stating Cybenko (1989). On the other hand, Spellman (1999) provides another reference according to which at least two layers are needed as it is the second layer that increases network power and can permit the modeling of more complex non-linear functions. A third layer is needed when the function is extremely complex, noisy or discontinuous. The conclusion is that the number of nodes either in single or multiple hidden layer networks and also the number of hidden layers should be a result of optimization, determined by trial and error. Practice showed that networks having a number of nodes split into more than one hidden layer, may sometimes outperform the single hidden layer one. The initial number of hidden layers and nodes per layer is defined by the modeler intuitively. The model performance in some papers was assessed only graphically without providing any statistical estimators. The latest papers however provide quite a number of statistical estimators that allow the reader a more objective judgment.

Overall, the various papers published up to now demonstrate that the artificial neural networks for tropospheric ozone time series approximation and forecasting is promising. The relative success of the models depends on how strict is the methodology followed by the developer, regarding the selection of the appropriate input data, concerning the type of variables, length of the time series, range, data normalization and preconditioning etc, and the extensive experimentation on the various parameters (i.e. selection of non-linear functions, number of layers and nodes) that determine the per-

formance of NN models in order to arrive to an optimal model. All, but one, papers show that neural networks, when properly applied, outperform other statistical methods. However, when developing them, the modeler should be fully aware about the specific points of this technique and treat them carefully.

References

Abdul-Wahab, S. A. and Al-Alawi, S. M.: Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks, Environmental Modelling and Software, 17, 219–228, 2002. 5755

Anderson, J. A. and Rosenfeld, E.: Neurocomputing, Foundations of Research, MIT Press, Cambridge, MA, 1988. 5741

Balaguer-Ballester, E., i Valls, J. L., Carrasco-Rodriguez, G. C., Soria-Olivas, E., and del Valle-Tascon, S.: Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks, Ecological Modelling, 156, 27–41, 2002. 5756

Cannon, A. J. and Lord, E. R.: Forecasting summertime surface-level ozone concentrations in the Lower Fraser valley of British Columbia: An ensemble neural network approach, J. Air Waste Manage. Assoc., 50, 322–339, 2000. 5751

Chaloulakou, A., Saisana, M., and Spyrellis, N.: Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, Sci. Tot. Environ., 29, 555–562, 2003. 5757

Cobourn, W. G., Dolcine, L., French, M., and Hubbard, M. C.: A comparison of nonlinear regression and neural network models for ground-level ozone forecasting, J. Air Waste Manage. Assoc., 50, 1999–2009, 2000. 5751, 5752

Comrie, A. C.: Comparing neural networks and regression models for ozone forecasting, J. Air Waste Manage. Assoc., 47, 653–663, 1997. 5745, 5748, 5749

Cybenko, G.: Approximations by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems, 2, 303–314, 1989. 5761, 5763

Elkamel, A., Abdul-Wahab, S., Bouhamra, W., and Alper, E.: Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach, Adv. Environ. Res., 5, 47–59, 2001. 5754, 5755

Gardner, M. and Dorling, S.: Artificial neural network-derived trends in daily maximum surface ozone concentrations, J. Air Waste Manage. Assoc., 51, 1202–1210, 2001. 5753

Gardner, M. W. and Dorling, S. R.: Artificial neural networks (The multilayer perceptron) - A review of applications in the atmospheric sciences, Atmos. Environ., 32, 2627–2636, 1998. 5741, 5743

Gómez-Sanchis, J., Martín-Guerro, J. D., Soria-Olivas, E., Vila-Francés, J., Carrasco, J. L., and del Valle-Tascón, S.: Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration, Atmos. Environ., 40, 6173–6180, 2006. 5759

Guardani, R., Nascimento, C. A. O., Guardani, M. L. G., Martins, M. H. R. B., and Romano, J.: Study of atmospheric ozone formation by means of a neural network-based model, J. Air Waste Manage. Assoc., 49, 316–323, 1999. 5747, 5763

Hadjiiski, L. and Hopke, P.: Application of artificial neural networks to modeling and prediction of ambient ozone concentrations, J. Air Waste Manage. Assoc., 50, 894–901, 2000. 5750

Heo, J. S. and Kim, D. S.: A new method of ozone forecasting using fuzzy expert and neural network systems, Sci. Tot. Environ., 325, 221–237, 2004. 5758

Kartalopoulos, S. V.: Understanding Neural Networks and Fuzzy Logic – Basic Concepts and Application, IEEE Press, New York, 1996. 5741, 5746, 5767

Nunnari, G., Nucifora, A. F. M., and Randieri, C.: The application of neural techniques to the modelling of time-series of atmospheric pollution data, Ecological Modelling, 111, 187–205, 1998. 5746

Pastor-Barcénas, O., Soria-Olivas, E., Martin-Guerro, J. D., Camps-Valls, G., Carrasco-Rodríguez, J. L., and del Valle-Tascón, S.: Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling, Ecological Modelling, 182, 149–158, 2005. 5759

Prybutok, V. R., Yi, J., and Mitchell, D.: 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, 2000. 5752

Robeson, S. M. and Steyn, D.: Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations, Atmos. Environ., 24B, 303–312, 1990. 5740

Ruiz-Suárez, J. C., Mayora-Ibarra, O., Torres-Jiménez, J., and Ruiz-Suárez, L. G.: Short-term ozone forecasting by artificial neural networks, Advances in Engineering Software, 23, 143–149, 1995. 5743

5765

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics, Wiley, New York, 2006. 5740

Soja, G. and Soja, A.-M.: Ozone indices based on simple meteorological parameters: potentials and limitations of regression and neural network models, Atmos. Environ., 33, 4299–4307, 1999. 5749

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., and Pereira, M. C.: Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environmental Modelling Software, 22, 97–103, 2007. 5760

Spellman, G.: An application of artificial neural networks to the prediction of surface ozone concentrations in the United Kingdom, Applied Geography, 19, 123–136, 1999. 5748, 5763

Viotti, P., Liuti, G., and Di-Genova, P.: Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia, Ecological Modelling, 148, 47–59, 2002. 5755, 5756

Wang, W., Lu, W., Wang, X., and Leung, A. Y. T.: Prediction of maximum daily ozone level using combined neural network and statistical characteristics, Environment International, 29, 555–562, 2003. 5756

Willmott, C. J.: On the validation of models, Physical Geography, 2, 184–194, 1981. 5746, 5756

Willmott, C. J., Ackelson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., and Rowe, C. M.: Statistics for the evaluation and comparison of models, J. Geophys. Res., 90(C5), 8995–9005, 1985. 5746, 5756

Yi, J. and Prybutok, V. R.: A neural network forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, Environ. Pollution, 92, 349–357, 1996. 5744, 5746, 5752

5766
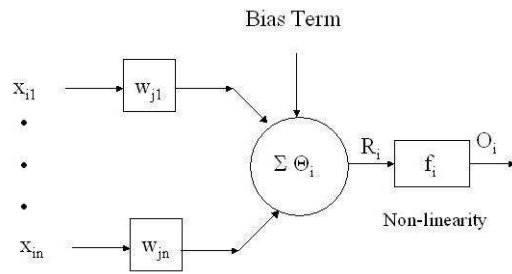
Fig. 1. Basic model of a neuron (Kartalopoulos, 1996).