Air Qual Atmos Health (2011) 4:263–272
DOI 10.1007/s11869-011-0139-2

Identifying the contribution of physical and chemical stressors to the daily number of hospital admissions implementing an artificial neural network model

Pavlos Kassomenos · Michael Petrakis ·
Dennis Sarigiannis · Alberto Gotti · Spyridon Karakitsios

Received: 19 January 2010 / Accepted: 2 March 2011 / Published online: 2 April 2011

Abstract The relative contribution of chemical (air pollution) and physical (temperature and humidity) health stressors to urban hospitalization rates is the objective of the current study. The data used in the study included the daily number of hospital admissions due to cardiorespiratory diseases, hourly mean concentrations of CO, NO_2, SO_2, O_3, and black smoke in several monitoring stations, as well as meteorological data (temperature, relative humidity, wind speed/direction) in Athens, Greece. The relations among the data above were studied using Generalized Linear Models (GLMs) and Artificial Neural Networks (ANNs). Elevated particulate concentrations are the dominant parameter related to hospital admissions (an increase of 10 μg/m^3 leads to an increase of 8.6% in hospital admissions), followed by O_3 and the other atmospheric pollutants (CO, NO_2, and SO_2). Meteorological parameters also play a decisive role in the formation of air-pollutant levels affecting public health. Both models performed adequately, however the ANN adaptation in complicate environmental issues results in improved modeling outcomes compared to the GLMs. The major finding of the study lies on the flexibility and the adaptation of the methodological approach for assessing non-linear problems and specifically the effect of non-linear parameters, such as the temperature. Moreover, the importance of temperature is established even when the whole dataset is modeled, reflecting the dual mode effect of temperature on cardiorespiratory admissions. Considering the urgent challenge to predict climate change effects on public health, a mathematical tool that successfully captures the direct impact of the affecting meteorological parameters (temperature and humidity) to health outcomes is of high added value.

Keywords Air pollution · Hospital admissions · Temperature · Wind · Humidity · Artificial neural networks · GLMs

P. Kassomenos (✉)
Department of Physics, Lab of Meteorology,
University of Ioannina,
45110 Ioannina, Greece
e-mail: pkassom@uoi.gr

M. Petrakis
National Observatory of Athens, Institute of Environmental Research and Sustainable Development,
12262 Athens, Greece

D. Sarigiannis
Department of Chemical Engineering,
Aristotle University of Thessaloniki,
54124 Thessaloniki, Greece

D. Sarigiannis · A. Gotti · S. Karakitsios
Physical and Chemical Exposure Unit (PCE),
Institute for Health and Consumer Protection (IHCP),
Joint Research Center (JRC), European Commission (EC),
TP-281, Via E. Fermi 1,
Ispra, Varese 21020, Italy

Introduction

Hospital admissions constitute a realistic morbidity metric associable to urban air pollution, since they refer to large parts of the population and constitute a direct way of tackling the cardiorespiratory health burdens imposed by complex air pollution mixtures to the communities exposed (Delfino et al. 1993). To date, a lot of studies have been conducted in order to define the relationship among one or more pollutants and the number of daily hospital admissions in various urban environments (Hauck et al. 2004; Moshammer et al. 2006; Neuberger et al. 2007). Concerns about the effects of particulate matter on public health and the consequential hospital admissions have also driven a number of research groups to study these relations in other cities in Europe (Katsouyanni et al. 2001), the USA (Atkinson et al. 2001), Australia (Morawska et al. 2002), and China (Kan et al. 2004). Apart from particulate matter, ozone plays a very significant role in the impoverishment of public health (McConnell et al. 2002); other research groups dealt with a wider range of pollutants including CO, NO_2, and SO_2 (Katsouyanni et al. 1997). Previous studies concerning the greater Athens area have indicated statistically significant short-term effects of weather and air pollution on mortality (Kassomenos et al. 2001; Kassomenos et al. 2007; Katsouyanni et al. 1997). Shortcomings encountered in air pollution/morbidity association are the potential of several forms of misclassification bias (Lippmann and Lioy 1985), the problem of multi co-linearity between different air pollutants, the harvesting resistance (Katsouyanni et al. 2001), as well as the fact that ambient concentrations recorded by monitors, may be a poor representation of personal exposure (Maddison 2005; Sarigiannis et al. 2004). An additional problem arising in all similar studies is the estimation of the net contribution of meteorological parameters to public health, when their contribution to the formation of air pollution levels is excluded. Higher temperatures appear to be an important factor in increasing the frequency of hospitalization for acute myocardial infarction and congestive heart failure, and they are associated with a decrease in the frequency of visits for coronary atherosclerosis and pulmonary heart disease, while several mechanisms are proposed for linking these interactions (Woodhouse et al. 1994; Gordon 2003).

The determination of the environmental parameters that actually affect hospital admissions still remains open, given that there is a lot of speculation about the suitability of the applied methodologies, or the special conditions of the urban area under study each time. In the present study an attempt has been made to find the short-term relationships among cardiorespiratory morbidity from all causes, air pollution and meteorology in the area of Athens, Greece for a period of 8 years (1992–2000). Analysis of the data was made through a state-of-the-art method (generalized linearized model (GLMs) and an Artificial Neural Network (ANN). The study aims at deriving morbidity functions (given the limitations of data) as well as testing the suitability of these models for assessing the contribution of the meteorological and air pollution factors. Both models where validated as predicting tools of the short-term hospital admissions.

Data

Topography and meteorological data

Athens is situated in a small peninsula located in the southeastern edge of the Greek mainland (central Greece). It covers almost 450 km^2 and has about 4,500,000 inhabitants. The built-up area is mainly located in a basin, surrounded by mountains on three sides and open to the sea (the Saronic Gulf) on the fourth side (Fig. 1). The three main mountains are Hymettus (1,050 m) to the E, Pendeli (1,100 m) to the N-NNE, and Parnitha (1,400 m) to the N-NNW. To the west of the basin is the mountain Aegaleo, 450 m high. There are small openings connecting the Metropolitan area of Athens with the rest of Greek mainland to the north, northwest, and northeast of the basin. More than 2.5 million cars are registered in Athens, while industrial activities are mainly located in the west and southwest of the basin. 3-hourly meteorological data (temperature, relative humidity, horizontal components of the wind velocity in the W–E and S–N directions) from the National Observatory of Athens weather station for the period 1992–2000 were used. The weather station is located in the center of the city on a hill at 107 m above sea level.

Hospital admissions data

Daily admissions for the period 1992–2000 were recorded by the Athens Health Authorities and referred to the hospital unit that provides health care to respiratory and cardiovascular events, for the Northern part of the city, inhabited by roughly 1.5 million. The cases considered for this study, in accordance to the ninth revision of the International Classification of Disease ICD-9, were: ischemic heart disease, cerebro-vascular disease, diseases of pulmonary circulation, acute respiratory infections, and other diseases of the respiratory systems or heart. The number of daily hospital admissions, due to respiratory and cardiovascu-

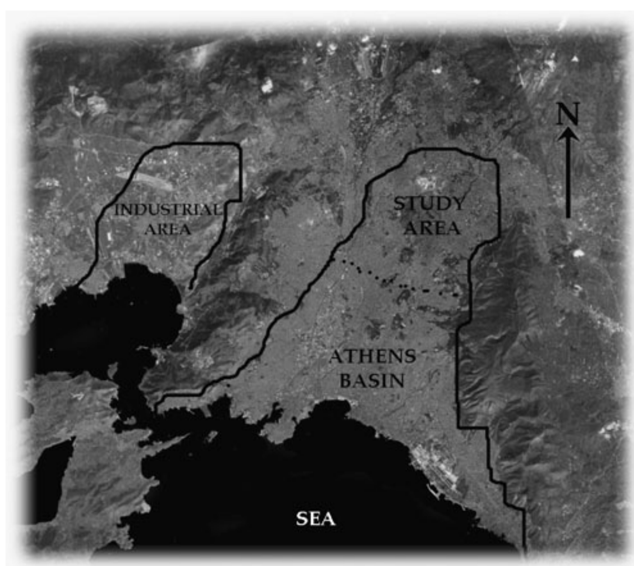

Fig. 1 Topography of the wider area of Athens

lar diseases, ranged between 0 and 10. Totally, 7,436 admissions were recorded (4,285 male/3,151 female).

The number of daily admissions seems to be slightly increased in winter and summer, compared to autumn and spring. In winter the cardiovascular diseases seem to be increased compared to summer, while the opposite occurs for respiratory diseases. Moreover, the seasonal distribution of hospital admissions shows that there are no significant differences among the various age groups.

Air pollution data

A regulatory network of 11 air pollution monitoring stations has been operating in the area of Athens since 1983. The main pollutants measured since that time are SO_2, black smoke (BS), CO, NO_2, NO, and O_3. Details for the monitoring methods can be found in the related authorities document (PERPA 1989). For the needs of this study, data from three air pollution stations, namely Patision, Marousi, and Aristotelous, were considered. These stations are the monitoring sites closest to the hospital, while at the same time measuring all the pollutants used in the study.

The above pollutants were measured hourly, which allows us to consider two possible exposure metrics: the daily maximum, or the daily average. The maximum concentration is more likely to be associated with certain types of acute health effects, but for pollutants primarily emitted from local sources, the daily maximum is likely to pertain to a limited area only. In addition, it is not clear whether sharp peaks in outdoor concentrations are also experienced indoors, where people spend the majority of their time. Since such spatial variations are likely to be smoothened out when the daily average is used, this is the recommended regression metric when the dependent variable is compiled on a daily basis (Lipfert and Hammerstrom 1992).

Although PM_{10} plays a significant role in similar studies worldwide, the authorities did not officially measure this pollutant in Athens until 2001. Up to that date, BS was measured instead. It must be noted that in European urban areas BS is well correlated with PM_{10} with a reported correlation coefficient between 0.53 and 0.74, and higher values during the winter months (Chaloulakou et al. 2005). Data for Athens, covering a 2-year period (1999–2001) revealed a strong linear correlation with R (correlation coefficient) exceeding 0.66 (Chaloulakou et al. 2005; Kyrkilis et al. 2007).

The daily average, maximum, and minimum values of the monitored pollutants are presented in Table 1. The main emission sources of the area are industrial activity (including refineries), traffic emissions and diesel oil-fired heating plants of building during winter. Although a significant improvement on traffic emissions was observed during the years of this study due to the replacement of non-catalytic vehicles with catalytic ones, the overall pollution levels were not reduced significantly. This is attributed to the constantly increasing number of vehicles in the wider area, as well as to the presence of industrial emissions, especially under specific weather conditions (Kassomenos et al. 1999).

Methodology

Generalized linearized models

Generalized linearized models are based on the concepts of the canonical Poisson regression, which is a relative risk model. It is a widely used approach for time-series analyses of air pollution and health that involves daily morbidity counts as the outcome, in linear terms measuring the percentage of increase in morbidity associated with elevations in air pollution levels.

It seems reasonable to suppose that if the population being studied doubled, while keeping its characteristics and all other risk factors constant, the number of increased cases due to air pollution and/or weather (if any) would also double. Since the baseline number of cases would also double under that scenario, this implies a relative risk model. In that model we assume:

$$\mathrm{Log}(E(Y)) = \beta_0 + \beta_1 X_1 + .. + \beta_p X_p \tag{1}$$

where Y is the count of deaths or hospital admissions on a given day, $E(Y)$ is the expected value of Y on that day (corresponding to λ in Eq. 1), $X_1..X_p$ are the predictors of daily counts, and $\beta_1 ... \beta_p$ are the regression coefficients for those predictors.

In all of epidemiology, a basic issue in modeling is to control properly for potential confounding. Time-series studies have some unique features in this regard. Many variables show systematic variation in time (i.e., trend). A second common attribute of many variables that evolve over time is seasonality. Many health, weather, and pollution variables show systematic variation over the course of the year. As for the trend, to focus on possibly causal associations with acute effects, it is necessary to remove these patterns. A final systematic component that may bias time-series regressions involves calendar specific days. Day of week or holiday effects fall in this category. These patterns are not necessarily present in all data, but they occur often enough that they should be checked.

In order to determine the variables that influence the response the most, we ran a preliminary multivariate stepwise linear regression model. On the basis of the results obtained, we selected a list of variables that affected the

Table 1 Daily average, maximum, and minimum values of the monitored pollutants

	CO (mg/m^3)	SO_2 (μg/m^3)	NO_2 (μg/m^3)	O_3 (μg/m^3)	BS (μg/m^3)
Mean	3.4	32.4	72.8	32.6	54.7
Max	12.4	246.5	278.7	103.4	230.3
Min	0.6	8.3	14.8	7.9	13.4

most the response variable, taking into account their statistical significance (p value lower than 0.05). To derive the relative risk (RR) functions, we applied a GLM fitting a Poisson distribution (logarithmic in the response) including only the statistically significant variables in terms on the basis of the above analysis.

Since hospital admissions are considered relatively rare occurrences, the number of occurrences (counts) of admissions per day is generated by a Poisson process where the probability distribution is given by:

$$\mathrm{prob}(y/\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad (2)$$

where the expected number of admssions on any day is λ and *prob(y/λ)* is the probability of y admissions occurring on a given day.

The generalized linear model formula describing the Poisson distribution is hereinafter presented:

$$\mathrm{Log}(E(Y)) = \beta_0 + \beta_1 X_1 + .. + \beta_p X_p \quad (3)$$

Poisson regression is essentially a methodology for the multivariate analysis of counts of relatively rare occurrences as the morbidity data. It assumes that the response variable Y follows a Poisson distribution, and that the logarithm of its expected value can be modeled by a linear combination of unknown parameters. The exponential of the coefficients of each covariate considered in the model allow us to calculate the relative risk.

$$Ln(\mu_t) = \beta_0 + \sum \beta_i x_{it} + \sum \gamma_n Y_{t-n}$$

where:

t	number of a given day (1,…, 2,920)
μ_t	daily expected number (mean) of admissions
i	explanatory variables (humidity, temperature, air pollution)
x_{it}	explanatory variable "i" value at time "t"
Y_{t-n}	number of daily deaths/hospitalizations at time "t" with a lag time "n".
n	lag time of the studied dependent variable (morbidity).
β_0, β_i, and γ_n	are regression coefficients.

The exponential of the coefficients of each covariate considered in the model allows us to calculate the relative risk.

The input parameters of the model were the average daily concentration of the pollutants examined (CO, NO_2, SO_2, PM_{10}, and O_3), the meteorological parameters (wind speed, wind direction, temperature, and humidity), as well as the adjustment for temporal confounding, expressed for 1 year (sin1, cos1), 6 months (sin 2, cos 2), and 4 months (sin 3, cos 3) periods. Of the overall dataset, 5% was left out for validating the model and comparing its predictive capability to the one obtained by the ANN.

ANN development

An artificial neural network was developed as an alternative tool for predicting hospital admissions on the basis of pollution and meteorological measurements. Artificial neural networks form a group of machine learning techniques, inspired by biological neurons. The initial concept of artificial intelligence systems was introduced by Robbins and Monro (Robbins and Monro 1951), who first managed to develop an innovative method for handling statistical challenges. However, it was only after the 1980s that neural nerwork (NN) became widely used, because of their demand for increased computational capacity resources. Since then, neural network models have been adopted in more complex problems, due to their ability to describe highly non-linear relationships. They were first used to model ambient pollution concentrations by Boznar et al. (1993) and since then they have been evolved into a significant tool for air quality modeling and forecasting. Gardner and Dorling (Dorling et al. 2003) gave an overview of various applications of NN in the atmospheric science during the 1990 s.

In general, a neural network is a computer model wherein the architecture of which mimics the knowledge acquisition and organizational skills of the human brain. The basic architecture of neural networks has been widely described by Lippmann (1987). Specifically, NN consist of a number of interconnected processing elements, commonly referred to as neurons. The neurons are logically arranged into two or more layers and interact with each other via weighted connections. These scalar weights determine the nature and strength of the influence between the interconnected neurons. Each neuron is connected to all other neurons in the next layer. There is an input layer where data are presented to the neural network and an output layer that holds the response of the network to the input. It is the intermediate layers, also known as hidden layers, which enable these networks to represent and compute complicated associations between patterns. Neural networks essentially learn through the adaptation of their connection weights.

In our approach, a feed forward, multi-layer network architecture was developed. For training and testing the proposed ANN model the overall dataset was randomly divided in two separate sets, one for training the ANN and one for testing its predictive capacity. The training set comprised 80% of the data, while the remaining 15% of the data formed the test set, while 5% was left out of the training procedure comprising the validation set.

We have evaluated the performance of the proposed methodology using various training algorithms on different network architectures. Specifically, we employed Bayesian regularization (Dan Foresee and Hagan 1997), Resilient Backpropagation (Riedmiller and Braun 1993) Scaled Conjugate Gradient (Møller 1993), as well as the Broyden, Fletcher, Goldfarb, and Shanno (Dennis and Schnabel 1996) and the Levenberg-Marquardt (Hagan and Menhaj 1994) algorithms. With respect to the employed architectures, we modified the number of units used in the hidden layer. In particular, we tested a number of different network architectures using one or two hidden layers with five up to 144 hidden neurons. From the obtained results it was found that Bayesian regularization gives constantly the best results, while the use of one hidden layer with 36 neurons may model successfully the problem under study. Thus, the proposed ANN model uses three layers as shown in Fig. 2. The first (input) layer consists of 18 neurons, one for each input parameter. The second (hidden) layer consists of ten neurons that implement the hyperbolic tangent sigmoid transfer function. Finally, the third (output) layer consists of one linear neuron that corresponds to the predicted hospital admissions. Network training was performed using the Bayesian regularization algorithm. A detailed analysis of the equations constructing the ANN model is given elsewhere (Karakitsios et al. 2007).

Results and discussion

GLM and ANN results

The results obtained from the GLM are presented in Table 2 in terms of regression coefficients, confidence intervals, statistical significance, while in Table 3 is presented the correlation matrix. The results indicated that a 10 μg/m^3 increase of PM_{10} and O_3 correspond to an increase to the daily number of hospital admissions equal to 10 and 11% respectively, while an increase of 1 mg/m^3 of CO corresponds to an increase of 7.2%. With regard to the importance of wind direction, west wind direction seems to be a significant contributor, followed by the south direction. Easterly wind direction seems to have some importance; however, even though the statistical significance criterion is satisfied, it is still higher than the one of west and south

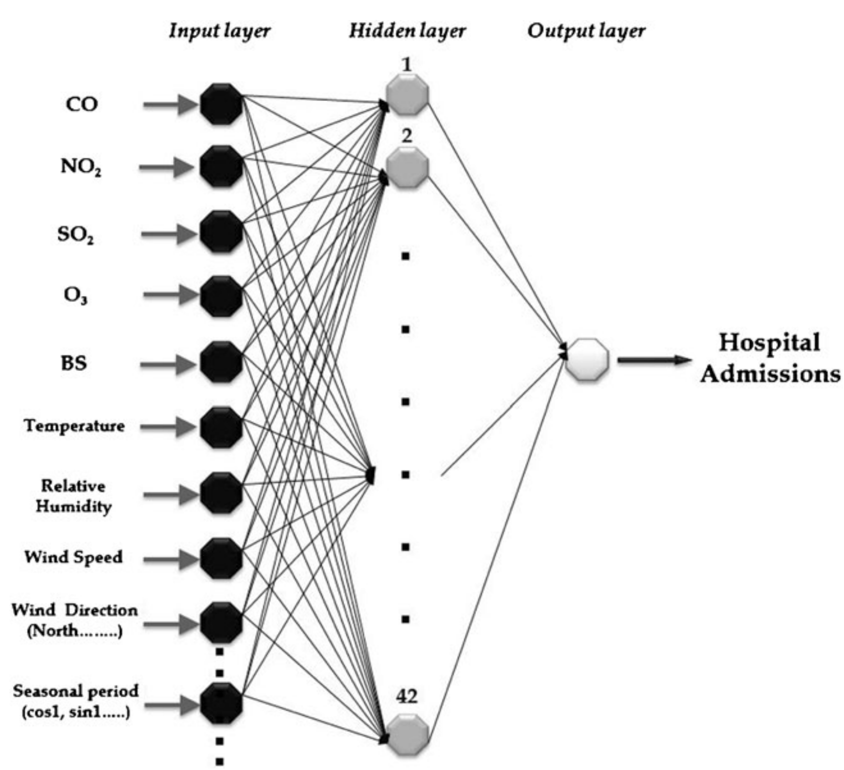

Fig. 2 Construction of the proposed ANN model

Table 2 GLM coefficients

Parameter	B	Standard error	95% Wald confidence interval		Hypothesis test	
			Lower	Upper	Wald chi-square	Significance
Intercept	−1.146	0.1174	−1.377	−0.916	95.328	0.000
PM	0.010	0.0004	0.009	0.011	797.341	0.000
O3	0.011	0.0008	0.009	0.012	184.171	0.000
CO	0.072	0.0077	0.057	0.087	87.601	0.000
Wind west	0.306	0.0331	0.242	0.371	85.668	0.000
COS2	0.065	0.0165	0.033	0.097	15.567	0.000
Wind south	0.237	0.0344	0.169	0.304	47.412	0.000
Temperature	0.014	0.0032	0.008	0.020	18.781	0.000
Relative humidity	0.006	0.0014	0.003	0.009	18.965	0.000
COS1	0.094	0.0339	0.027	0.160	7.633	0.006
Wind east	0.106	0.0443	0.019	0.193	5.729	0.017

direction, as well as the overall contribution weight. The contribution of temperature and humidity seem to be low. In addition, an existing seasonality trend was identified with a period of 6 months, while a secondary seasonal trend was identified regarding the annual cycle. The contribution of the parameters through the ANN network was identified by the sensitivity analysis scores obtained by the model analysis. The results indicated that an increase of 10 $\mu g/m^3$ of PM_{10} and O_3 correspond to an increase to the daily number of hospital admissions equal to 8.6 and 6% respectively, while an increase of 1 mg/m^3 of CO corresponds to an increase of 4.1%.

The performance and validity of the models were evaluated based on the 5% of the data and the results were compared with those obtained by the GLM model (and for the same data set). The evaluation parameters are presented in Table 4, while in Fig. 3 the validation set is presented (predicted vs. observed data). As seen in Table 4 the ANN model performs better in predicting the daily number of hospital admissions under known air pollution and meteorological conditions. Moreover, the Mean Error and the relevant RMSE and RRMSE parameters are smaller for the ANN model. The GLM model is a good predictor for values close to the average value of the dependent parameter - the predictive power of the model grows weaker when trying to predict values close to the upper or lower limits of its validity domain. The ANN model has a more uniform distribution of the prediction error. The only limitation of the ANN is that it can only utilize input data within the range of values included in the training set. All values beyond that range are considered equal to the closest value in the training data set. In a validation method like the one applied here, this situation is not unusual and can lead to excessive maximum errors that are not consistent with the rest of the evaluation parameters.

In addition, the weights assigned to the connections between the layers of the ANN can provide information on the relative importance of each parameter to hospital

Table 3 GLM correlation matrix

	Intercept	PM_{10}	O_3	CO	Wind west	COS^2	Wind south	Temperature	Relative humidity	COS1	Wind east
Intercept	1.000	0.133	−0.334	0.144	−0.230	−0.030	−0.120	−0.603	−0.821	−0.297	−0.077
PM	0.133	1.000	0.107	−0.480	−0.091	−0.077	−0.141	−0.144	−0.193	−0.126	−0.096
O_3	−0.334	0.107	1.000	−0.017	−0.303	0.021	−0.147	−0.023	0.187	0.223	−0.085
CO	0.144	−0.480	−0.017	1.000	−0.026	−0.020	−0.093	−0.182	−0.263	−0.155	0.033
Wind west	−0.230	−0.091	−0.303	−0.026	1.000	−0.026	0.590	0.032	0.176	−0.160	0.453
COS2	−0.030	−0.077	0.021	−0.020	−0.026	1.000	0.035	−0.015	0.059	−0.027	0.067
Wind south	−0.120	−0.141	−0.147	−0.093	0.590	0.035	1.000	0.019	0.048	0.002	0.427
Temperature	−0.603	−0.144	−0.023	−0.182	0.032	−0.015	0.019	1.000	0.207	0.761	−0.041
Relative humidity	−0.821	−0.193	0.187	−0.263	0.176	0.059	0.048	0.207	1.000	−0.094	0.009
COS1	−0.297	−0.126	0.223	−0.155	−0.160	−0.027	0.002	0.761	−0.094	1.000	−0.104
Wind east	−0.077	−0.096	−0.085	0.033	0.453	0.067	0.427	−0.041	0.009	−0.104	1.000

Table 4 ANN vs GLM model evaluation parameters

	ANN	Regression
Mean error	0.6961	0.7763
Maximum error	3.6474	4.0690
Minimum error	0.0022	0.0005
RMSE	0.8950	1.0198
RRMSE	0..1112	0.1225
Correlation coefficient	0.7632	0.6630

admissions. The method for partitioning the connection weights proposed by Garson was used. The technique involves partitioning the hidden–output connection weights of each neuron into components associated with each input neuron. Similarly, using the Wald χ^2 scores to indicate the relative importance of the input parameters, these values were reduced to percentages relative to their sum, and they are presented in parallel to the ANN contributions in Figure 4 as columns representing the relative importance of the various input variables. Clearly the GLM considers PM_{10} (with a relative contribution score of 58.4%) as the parameter dominating the variation in hospital admissions, followed by the contribution of O_3 (relative contribution score of 13.6%) and CO (relative contribution score of 6.4%). The relative contribution scores for the ANN model seem more balanced, being 22.7%, 17%, and 7.7% for PM_{10}, $O_{3,}$ and CO respectively. The temporal (seasonal) confounders of the GLM and the ANN had a relative contribution to physiological response of 1.7% and 0.7% respectively. The results become very interesting with regard to the contribution of the meteorological parameters, where a significantly higher contribution was attributed to all of them by the ANN model. Temperature and humidity were attributed a relative contribution of about 8%, while the respective values for the GLM model were less than 1.5%. Wind speed was not considered as one of the parameters meeting the statistical significance criterion for GLM, thus the attributed contribution was zero, while a similar contribution was found for GLM (10%) and ANN (11.5%) for wind direction.

Discussion

The results of the study cannot be directly compared to existing studies in the wider area, since although there are existing data that come from mortality studies, there is only one relevant morbidity study based on data limited for the year 1988 (Pantazopoulou 1995). The study indicated that the daily number of hospital visits was correlated positively with the levels of air pollution, but this association was not statistically significant for most of the pollutants. However, when the dataset was divided in two seasons, the number of hospital visits was related to a statistically significant degree with all indices during the winter time, providing a first evidence on the association between air pollution and morbidity in the wider area. With regard to the mortality studies conducted in the area, most of the results come from the APHEA project (Katsouyanni et al. 1997; Touloumi et al. 1996), indicating that an increase of 10 μg/m^3 in SO_2 and BS corresponded to 1.2% increase and 0.5% in the daily total numbers of deaths, while an increase of 1 μg/m^3 in CO, the increase was 1%. In addition, a stronger effect of SO_2 on the daily total number of deaths was observed when the levels of BS were >100 μg/m^3. Additional studies conducted in the area were related to the effect of confounding meteorological factors on mortality, (Kassomenos et al. 2001; Kassomenos et al. 2007), identifying the weather types affecting mortality, without to distinguish if these types of weather co-occur with elevated air pollution.

The overall concentration-response functions derived in the study are in good agreement with the results derived by similar but wider-scale studies conducted in Europe (Pelucchi et al. 2009) and worldwide (Ostro 2004). We need to take into account that the data used in the study include the sum of all types of cardio and respiratory (except influenza) causes, thus the higher concentration-response values observed in the current study were to be expected.

In addition, it is important to discuss some issues regarding the clarity of the methodology of the ANN. Although in many studies ANNs have been shown to exhibit superior predictive power compared to the wide variety of regression approaches, they have also been labeled as "black boxes" because they provide little explanatory insight into the relative influence of the independent variables on prediction. This lack of explanatory power is a major concern since the interpretation of statistical models provides knowledge on the causal relationships between environmental stressors and health

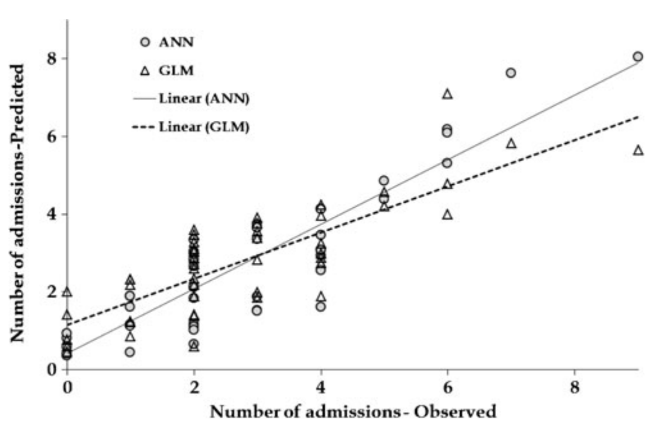

Fig. 3 Evaluation of ANN and GLM model in comparison to the recorded hospital admission data

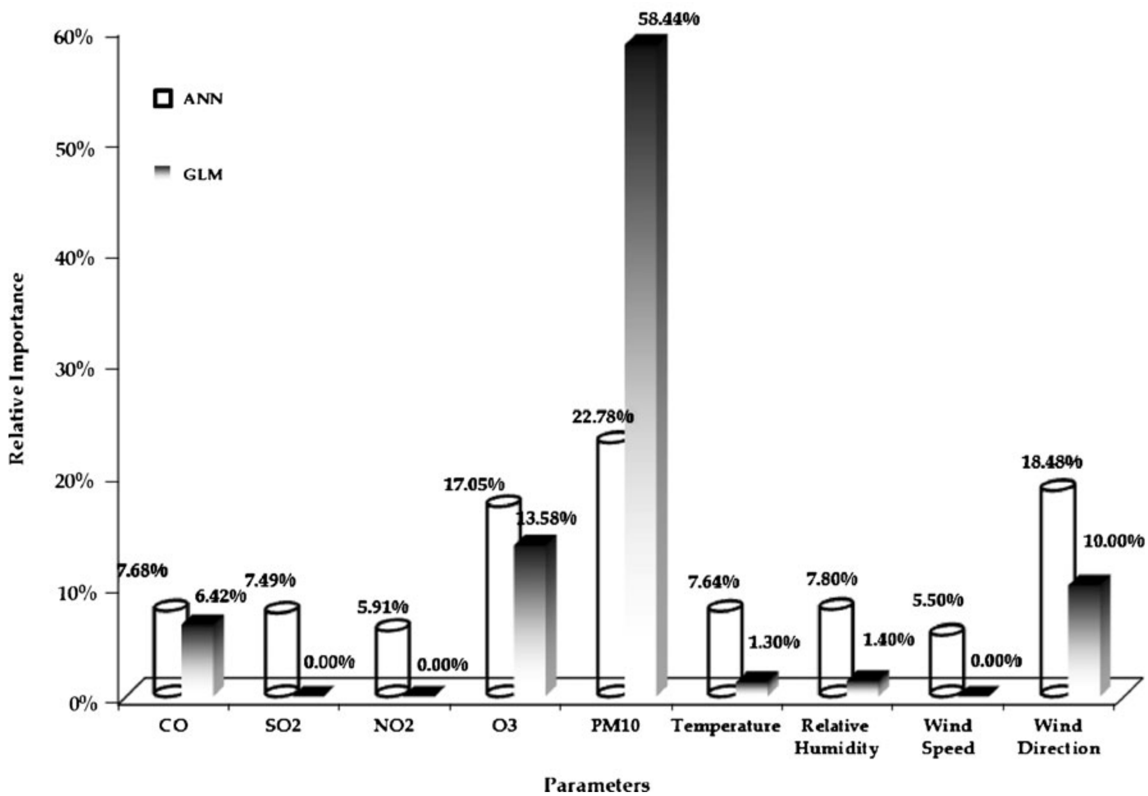

Fig. 4 Relative importance of the parameters affecting hospital admissions obtained by ANN and GLM models

outcomes. However, when the proper tools are used for the interpretation of the results, as it is the case in the current study, (Garson's algorithm, sensitivity analysis), we can assess the parameters governing the phenomenon with a smaller degree of uncertainty compared to regression methods. From the comparison of the results of the two models against the validation dataset, it can be deduced that the ANN model is more capable than GLMs to model. The results regarding the importance of the parameters as obtained by the ANN model seem to be more reliable. The ANN model attributed a higher significance to the physical stressors (temperature and relative humidity) than the GLM model. Especially regarding temperature, significant contributions to morbidity are expected within small ranges of values, a fact supported in the study conducted in Madrid (García-Herrera et al. 2005), where a triggering effect on mortality was identified when maximum daily temperature exceeds a given threshold (34°C in Lisbon and 36°C in Madrid). Thus, the need for an adaptive model capturing these changes is emerging. The capability of the ANN model to capture to a higher extent the effects of physical stressors is of great value in the context of assessing the impacts of the interaction between air pollution and climate change on public health.

From the methodological point of view, it should be noted that although the link between air pollution and hospital admissions is well established, the statistically significant quantification of this causal association is subject to many uncertainties (Sahsuvaroglu and Jerrett 2007). One major problem is the relation between the measured air pollution levels and the corresponding exposure levels of the population in the area under study. Unfortunately, most of the time, the concentrations measured in the monitoring sites are hardly representative of the exposure of the general population (Sarigiannis and Saisana 2008). This is due to several reasons, such as:

- The location of the monitoring sites, which does not always reflect the actual population exposure levels. The monitoring sites considered in epidemiological studies must be located in areas where people live and work and population density has to be homogeneously distributed.
- The temporal distribution of the maximum levels of the recorded air pollution levels. Ozone for example is a secondary pollutant, the ambient air level of which presents its maximum value usually at noon, i.e. delayed compared to other pollutants like PM_{10}. In such a case, it is possible that the levels of both pollutants are elevated the same day, but that due to the different time when the respective maxima are reached, the population is exposed to a different degree to each of them.

– The activity patterns of the exposed population. This is related also to the above parameter. It is important to know the activities of the population that are exposed, especially during the hours of the peak concentrations. If for example the majority of the population the peak times is commuting to workplace, the exposure values raise significantly. On the other hand, there are times when even though pollution levels are elevated people avoid going out of their home due to possible high ambient temperature. In these cases, the model might underestimate the importance of air pollution on public health.

In addition, there are factors that alter the physiological response of the population to air pollution:

– The susceptibility of the population. Not all the people respond in the same way to xenobiotics, due to differences in metabolism, health status, gender, and age.
– The harvesting resistance of the population to air pollution. In this case, the effects of air pollution are not only manifested in a short-term context; they may also appear a few days later (Dilaveris et al. 2006).
– The chemical composition of urban air pollution. Ambient air is a complex mixture of xenobiotics. Consequently, we ignore possible mixture effects, which may be even more complicated to assess when taking into account the effect of physical stressors, as indicated by a small number of related studies reviewed by Gordon (2003). In addition, the chemical composition of PM may alter significantly its systemic toxicity and consequently its contribution to morbidity and mortality.

Without any doubt, a mechanistic full chain assessment that takes into account emission processes, environmental fate, actual personal exposure, biokinetics, co-exposure to several chemical and physical stressors, toxicity pathways, early biological responses, and finally possible health outcomes would be ideal, although the complexity and the multidisciplinary needed could make it cumbersome. For the moment, integrated approaches taking into account that “full chain assessment” are still limited to long term exposure effects related to carcinogenic health endpoints (Karakitsios et al. 2008; Sarigiannis et al. 2011; Sarigiannis et al. 2009; Georgopoulos et al. 2008).

Conlusions

In the present study, the problem of hospital admissions and air pollution/meteorology in the Athens urban basin has been addressed through GLM and ANN modeling. The GLM indicated that O_3 has a slightly higher impact on public health than PM_{10}; however, emission reduction of these primary (for PM_{10}) and secondary (for PM_{10} and O_3) pollutants have to be a priority for environmental and public health managers. Westerly and southern winds lead to pollutant accumulation in the area, and thus have an adverse effect on public health especially when they are accompanied by elevated temperatures and relative humidity. Between the two models, ANN gave a higher contribution to the meteorological parameters, capturing more efficiently the combined effect of chemical (air pollution) and physical (temperature and humidity) stressors on health outcome indicators.

In addition, the conclusions drawn by the study are interesting from the environmental policy point of view, since the contribution of the special air pollution and climatic characteristics of the area to the daily number of hospital admissions was successfully modeled. The ANN methodology suggested herein is a more flexible and adaptive mathematical framework than GLMs and other simpler forms of regression analysis as shown by the prediction results. Thus, besides the assessment of the significance of the environmental stressors affecting hospital admissions, the model is useful for forecasting the days of higher morbidity risk. This predictive capacity of the ANN model makes it a valid tool supporting decisions regarding preventive measures. Our future work will extend this methodology to other potential application sites (e.g., Rome, Madrid) capturing the effects of harvesting on the predictive capacity of the model.

References

Atkinson RW, Anderson HR, Sunyer J, Ayres J, Baccini M, Vonk JM, Boumghar A, Forastiere F, Forsberg B, Touloumi G, Schwartz J, Katsouyanni K (2001) Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2 project. Am J Respir Crit Care Med 164(10 I):1860–1866

Boznar M, Lesjak M, Mlakar P (1993) A neural network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain. Atmos Environ B 27(2):221–230. doi:10.1016/0957-1272(93)90007-s

Chaloulakou A, Kassomenos P, Grivas G, Spyrellis N (2005) Particulate matter and black smoke concentration levels in central Athens, Greece. Environ Int 31(5):651–659

Dan Foresee F, Hagan MT (1997) Gauss-Newton approximation to Bayesian learning. In: Neural Networks. International Conference on 1997, 9–12 Jun, vol. 1933, pp 1930–1935

Delfino RJ, Becklake MR, Hanley JA (1993) Reliability of hospital data for population-based studies of air pollution. Arch Environ Health 48(3):140–146

Dennis JEJ, Schnabel RB (1996) Numerical methods for unconstrained optimization and nonlinear equations. Soc Ind Appl Math

Dilaveris P, Synetos A, Giannopoulos G, Gialafos E, Pantazis A, Stefanadis C (2006) CLimate impacts on Myocardial infarction deaths in the Athens TErritory: the CLIMATE study. Heart 92(12):1747–1751

Dorling SR, Foxall RJ, Mandic DP, Cawley GC (2003) Maximum likelihood cost functions for neural network models of air quality data. Atmos Environ 37(24):3435–3443

García-Herrera R, Díaz J, Trigo RM, Hernández E (2005) Extreme summer temperatures in Iberia: health impacts and associated synoptic conditions. Ann Geophys 23(2):239–251

Georgopoulos PG, Wang SW, Yang YC, Xue J, Zartarian VG, McCurdy T, Ozkaynak H (2008) Biologically based modeling of multimedia, multipathway, multiroute population exposures to arsenic. J Expo Sci Environ Epidemiol 18(5):462–476

Gordon CJ (2003) Role of environmental stress in the physiological response to chemical toxicants. Environ Res 92(1):1–7

Hagan MT, Menhaj MB (1994) Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Hauck H, Berner A, Frischer T, Gomiscek B, Kundi M, Neuberger M, Puxbaum H, Preining O, Amoako-Mensah T, Bauer H, Broer S, Ctyroky P, Danninger E, Eiwegger T, Frühauf P, Galambos Z, Gartner C, Hann W, Horak F Jr, Horvath H, Iro A, Kalina M, Klocker J, Kreiner P, Krejci W, Kromp-Kolb H, Krüger B, Lavric T, Limbeck A, Matzke W, Moshammer H, Piegler B, Pouresmaeil P, Pühringer W, Putschögl B, Raber W, Riess P, Salam A, Schimek MG, Schmid H, Schuster B, Semmelrock G, Syeda B, Stopper S, Studnicka M, Tarmann V, Wartlik E, Zarkada A (2004) AUPHEP - Austrian Project on Health Effects of Particulates - General overview. Atmos Environ 38 (24):3905–3915

Kan H, Chen B, Chen C, Fu Q, Chen M (2004) An evaluation of public health impact of ambient air pollution under various energy scenarios in Shanghai, China. Atmos Environ 38 (1):95–102

Karakitsios S, Sarigiannis D, Gotti A, Kassomenos P, Pilidis G (2008) An integrated exposure and risk model for benzene in the ambient air. Epidemiology 19(6):S356. doi:310.1097/1001.ede.0000340432.0000391307.0000340433e

Karakitsios SP, Papaloukas CL, Kassomenos PA, Pilidis GA (2007) Assessment and prediction of exposure to benzene of filling station employees. Atmos Environ 41(40):9555–9569

Kassomenos P, Gryparis A, Samoli E, Katsouyanni K, Lykoudis S, Flocas HA (2001) Atmospheric circulation types and daily mortality in Athens, Greece. Environ Health Perspect 109 (6):591–596

Kassomenos P, Skouloudis AN, Lykoudis S, Flocas HA (1999) 'Air-quality indicators' for uniform indexing of atmospheric pollution over large metropolitan areas. Atmos Environ 33(12):1861–1879

Kassomenos PA, Gryparis A, Katsouyanni K (2007) On the association between daily mortality and air mass types in Athens, Greece during winter and summer. Int J Biometeorol 51(4):315–322

Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, Monopolis Y, Rossi G, Zmirou D, Ballester F, Boumghar A, Anderson HR, Wojtyniak B, Paldy A, Braunstein R, Pekkanen J, Schindler C, Schwartz J (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology 12(5):521–531

Katsouyanni K, Zmirou D, Spix C, Sunyer J, Schouten JP, Ponka A, Anderson HR, Le Moullec Y, Wojtyniak B, Vigotti MA, Bacharova L, Schwartz J (1997) Short-term effects of air pollution on health: a European approach using epidemiologic time series data. The APHEA project. Public Health Rev 25(1):7–18

Kyrkilis G, Chaloulakou A, Kassomenos PA (2007) Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: relation to potential health effects. Environ Int 33 (5):670–676

Lipfert FW, Hammerstrom T (1992) Temporal patterns in air pollution and hospital admissions. Environ Res 59(2):374–399

Lippmann M, Lioy PJ (1985) Critical issues in air pollution epidemiology. Environ Health Perspect 62:243–258

Lippmann R (1987) An introduction to computing with neural nets. IEEE ASSP Mag 4(2):4–22

Maddison D (2005) Air pollution and hospital admissions: an ARMAX modelling approach. J Environ Econ Manage 49 (1):116–131

McConnell R, Berhane K, Gilliland F, London SJ, Islam T, Gauderman WJ, Avol E, Margolis HG, Peters JM (2002) Asthma in exercising children exposed to ozone: a cohort study. Lancet 359 (9304):386–391

Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6(4):525–533

Morawska L, Vishvakarman D, Mengersen K, Thomas S (2002) Spatial variation of airborne pollutant concentrations in Brisbane, Australia and its potential impact on population exposure assessment. Atmos Environ 36(21):3545–3555

Moshammer H, Hutter HP, Neuberger M (2006) Gas cooking and reduced lung function in school children. Atmos Environ 40 (18):3349–3354

Neuberger M, Rabczenko D, Moshammer H (2007) Extended effects of air pollution on cardiopulmonary mortality in Vienna. Atmos Environ 41(38):8549–8556

Ostro B (2004) Outdoor air pollution: assessing the environmental burden of disease at national and local levels. WHO Environmental Burden of Disease Series. Geneva

Pantazopoulou A (1995) Short-term effects of air pollution on hospital emergency outpatient visits and admissions in the greater Athens, Greece area. Environ Res 69(1):31–36

Pelucchi C, Negri E, Gallus S, Boffetta P, Tramacere I, La Vecchia C (2009) Long-term particulate matter exposure and mortality: a review of European epidemiological studies. BMC Public Health 9(1):453

PERPA (1989) Database information, vol. 1. Air Pollution in Athens area. Technical Note. Helenic Ministry of the Environment - Directorate of Air Pollution and Noise control, Athens

Riedmiller M, Braun HA (1993) Direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Neural Networks. IEEE International Conference on 1993, vol. 581, pp 586–591

Robbins H, Monro S (1951) A stochastic approximation method. Ann Math Stat 22:400–407

Sahsuvaroglu T, Jerrett M (2007) Sources of uncertainty in calculating mortality and morbidity attributable to air pollution. J Toxicol Environ Health A 70(3–4):243–260

Sarigiannis D, Gotti A, Karakitsios S (2011) A computational framework for aggregate and cumulative exposure assessment. Epidemiology 22(1):S96–S97. doi:10.1097/1001.ede.0000391962.0000303834.0000391966

Sarigiannis D, Karakitsios S, Gotti A (2009) Mechanistic full chain approach for ETS carcinogenicity impact assessment in the EU. Epidemiology 20(6):S88. doi:10.1097/1001.ede.0000362974.0000301246.0000362996

Sarigiannis DA, Saisana M (2008) Multi-objective optimization of air quality monitoring. Environ Monit Assess 136(1–3):87–99

Sarigiannis DA, Soulakellis NA, Sifakis NI (2004) Information fusion for computational assessment of air quality and health effects. Photogramm Eng Remote Sens 70(2):235–245

Touloumi G, Samoli E, Katsouyanni K (1996) Daily mortality and "winter type" air pollution in Athens, Greece - A time series analysis within the APHEA project. J Epidemiol Community Health 50(Suppl. 1)

Woodhouse PR, Khaw KT, Plummer M, Foley A, Meade TW (1994) Seasonal variations of plasma fibrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease. Lancet 343(8895):435–439