

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Air Quality and Bioclimatic Conditions within the Greater Athens Area, Greece - Development and Applications of Artificial Neural Networks

Panagiotis Nastos¹, Konstantinos Moustris²,
Ioanna Larissi³ and Athanasios Paliatsos⁴

¹Laboratory of Climatology and Atmospheric Environment,
Faculty of Geology and Geoenvironment, University of Athens,

²Department of Mechanical Engineering, Technological Educational Institute of Piraeus,

³Department of Electronic-Computer Systems Engineering,
Technological Educational Institute of Piraeus,

⁴General Department of Mathematics, Technological Educational Institute of Piraeus,
Greece

1. Introduction

Over the past few decades the phenomenon of urbanization resulted in severe problems. The quality of human life has been deteriorated in the megacities around the world. This chapter deal with the Artificial Networks (ANNs) forecasting ability in predicting the air quality as well as the bioclimatic conditions in an urban environment.

For this purpose, different ANNs are demonstrated in this chapter. These ANNs have been developed in order to predict the air quality as well as the bioclimatic conditions within the Greater Athens Area (GAA), Greece. The prognosis for both air quality and bioclimatic conditions within GAA concerns the next three days (24 to 72 hours prediction). For the proper ANNs training for both air quality and bioclimatic conditions, hourly values of specific meteorological parameters such as air temperature, relative humidity, wind speed, wind direction, air pressure, sunshine and solar radiation, as well as hourly values of air pollutants concentrations have been used. These hourly data have been recorded in many different sites within GAA from the network of the Greek Ministry of Environment Energy and Climatic Change (GMEECC) during the period 2001-2005. Hourly values of barometric pressure and total solar irradiance for the same time period were acquired from the National Observatory of Athens (NOA).

This chapter is divided into nine sections. The first section is brief introduction concerning ANNs. The second section presents air quality indices that have been used in this work in order to describe the air quality within GAA. The third section presents bioclimatic indices, which describe the human thermal comfort-discomfort due to meteorological conditions. The fourth section presents statistical performance indices that have been used in order to investigate the predictive ability and reliability of the developed ANNs models. The fifth

section demonstrates the examined sites within the GAA and the data/methodology used in this study.

The sixth section presents the ANNs that were developed in order to predict the maximum daily value of the air pollution indices as well as the persistence of the phenomenon, namely the number of consecutive hours within the day with high/strong air pollution. The seventh section presents the ANNs that were developed in order to predict the daily values of the bioclimatic indices as well as the number of consecutive hours within the day with dangerous bioclimatic conditions for humans' health. The eighth section includes the spatial variation for both air quality levels and human comfort/discomfort levels within GAA. The ninth is the last section summarizing briefly the extracted results by the performed analysis and how these results can contribute positively to the economy, energy, environment and quality of human life in general.

Finally the results of this work have shown that the ANNs could give an adequate forecast for both air quality and bioclimatic conditions within the urban environment of the GAA for the next three days at a statistical significant level of $p < 0.01$.

2. Artificial Neural Networks

Artificial Neural Networks (ANNs) are a branch of artificial intelligence developed in the 1950s aiming at imitating the biological brain architecture. They are an approach to the description of functioning of human nervous system through mathematical functions. Typical ANNs use very simple models of neurons. These artificial neurons models retain only very rough characteristics of biological neurons of the human brain (McCulloch & Pitts, 1943). ANNs are parallel-distributed systems made of many interconnected non-linear processing elements (PEs), called neurons (Hecht-Nielsen, 1990). A renewal of scientific interest has grown exponentially since the last decade, mainly due to the availability of appropriate hardware that has made them convenient for fast data analysis and information processing (Viotti et al., 2002).

Figure 2.1 presents the structure of a biological neuron (upper graph) as well as the structure of an artificial neural (lower graph).

ANNs have been applied in time series prediction (Lapedes & Farber, 1987; Werbos, 1988). Although their behaviour has been related to non-linear statistical regression (Bishop, 1995), the big difference is that ANNs seem naturally suited for problems that show a large dimensionality of data, such as the task of identification for systems with great number of state variables. Over the last years, black box approaches have been recognized to constitute a viable alternative to conceptual models for input-output simulation and forecasting and also to allow shortening the time required for the model development. In particular, ANNs concentrated a general consensus in predicting different pollutants time series, as shown by the review of Gardner & Dorling (1998a, 1998b).

Many ANNs were developed for very different environmental purposes. Heymans & Baird (2000) have used network analysis to evaluate the carbon flow model built for the northern Benguela upwelling ecosystem in Namibia. Antonic et al. (2001) have estimated the forest survival after building the hydroelectric power plant on the Drava River, Croatia by means of GIS constructed database and a neural network. Karul et al. (2000) used a three-layer Levenberg-Marquardt feedforward neural network to model the eutrophication process in three water bodies in Turkey. Besides, Moustiris et al. (2011) used ANNs for long term precipitation forecast, using long-term monthly precipitation time series of four meteorological stations in Greece.

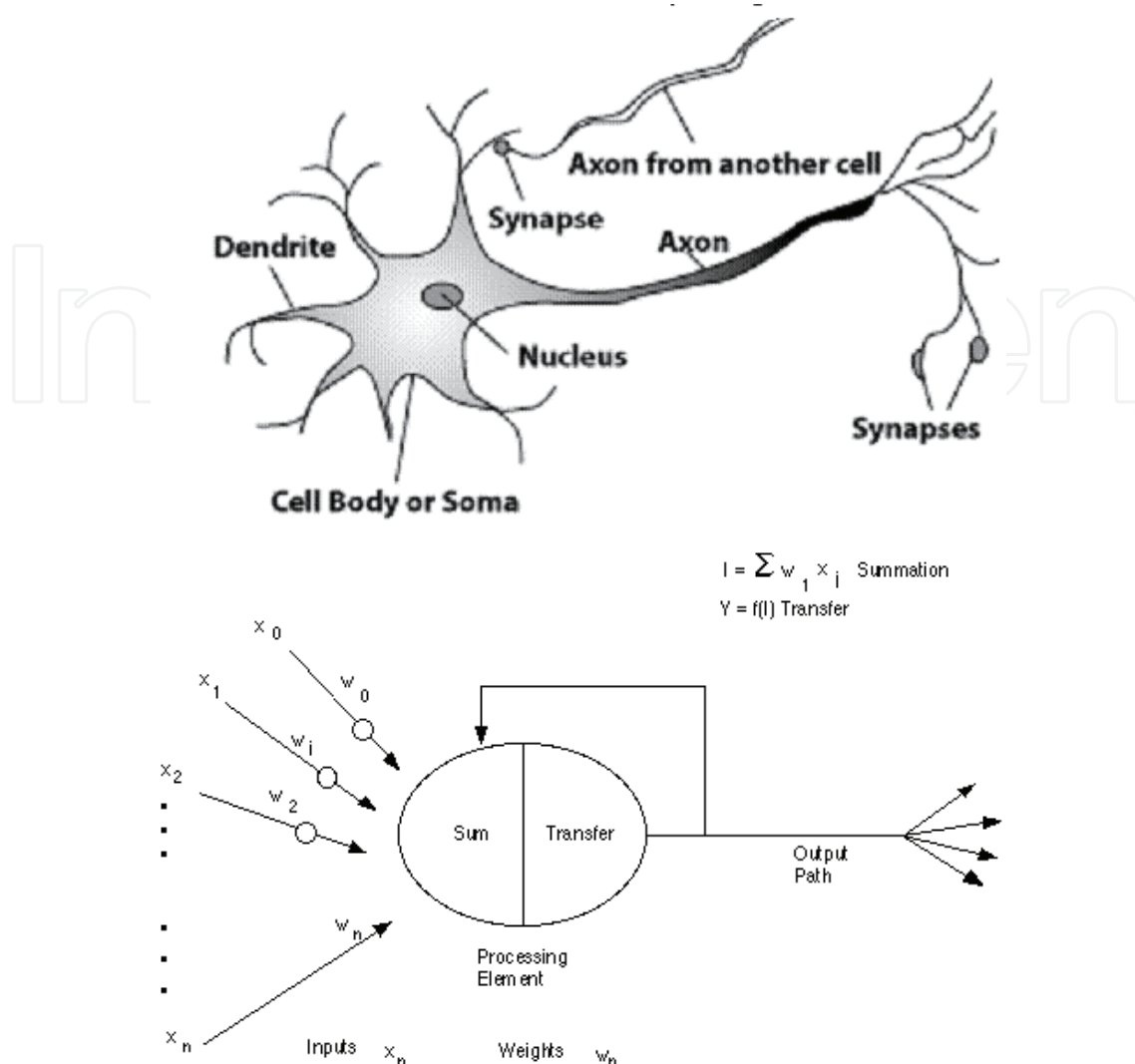


Fig. 2.1. Biological (upper graph) and artificial (lower graph) neuron structure.

Viotti et al. (2002) used ANNs to forecast short and middle long-term concentration levels for some of the well-known pollutants at the urban area of Perugia, Italy. The ANNs approach proved to be viable also for O_3 , PM_{10} , NO_2 , NO_x forecasting, outperforming alternative techniques in different case studies (Nunnari et al., 1998; Prybutok et al., 2000; Kolehmainen et al., 2001; Balaguer Ballester et al., 2002; Schlink et al., 2003; Corani, 2005; Slini et al., 2006; Dutot et al., 2007; Papanastasiou et al., 2007).

2.1 Multi-Layer Perceptron and feed-forward ANNs

The Multi-Layer Perceptron (MLP) is the most commonly used type of ANNs. Its structure consists of Processing Elements (PEs) and connections (Hecht-Nielsen, 1991). PEs, which are called neurons, are arranged in layers. The first layer is the input layer, one or more hidden layers follow and the final layer is the output layer. An input layer serves as buffer that distributes input signals to the next layer, which is a hidden layer. Each neuron of the hidden layer communicates with all the neurons of the next hidden layer, if any, having in each connection a typical weight factor. So, each unit-artificial neuron in the hidden layer sums its input, processes it with a transfer function and distributes the result to the output

layer. It is also possible that there are several hidden layers connected in the same fashion. The units-artificial neurons in the output layer compute their output in a similar manner. Finally, the signal reaches the output layer, where the output value from the ANN is compared to the target value and an error is estimated. Thus, the values of weight factors are amended appropriately and the training cycle repeats until the error is acceptable, depending on the application.

Since data flow within the artificial neural network from a layer to the next one without any return path, such kind of ANNs are defined as feed-forward ANNs. The structure of a feed-forward Multi-Layer Perceptron artificial neural network can be represented as in Figure 2.1.1.

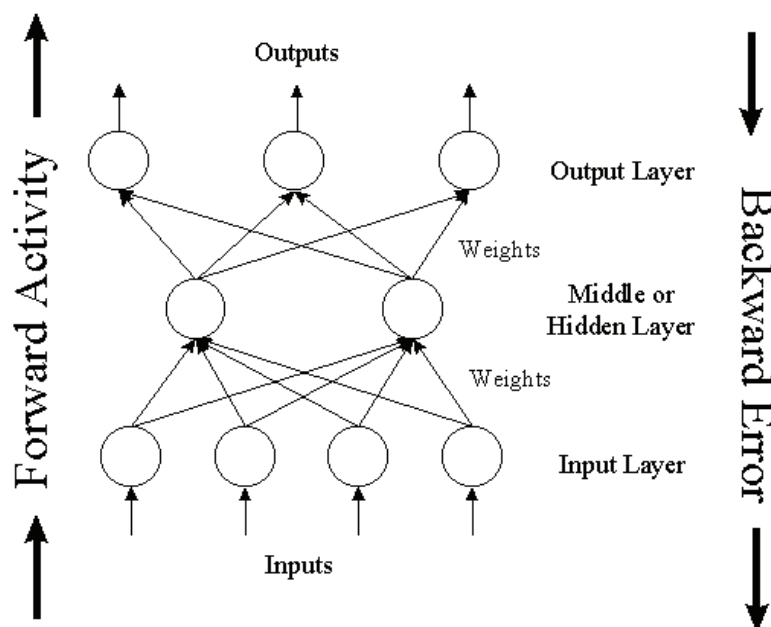


Fig. 2.1.1. Typical MLP feed-forward Artificial Neural Network Structure (Caudill & Butler, 1992).

2.2 Feed-forward ANNs training and the Back-propagation training algorithm

The training-learning process of ANNs can be far from the ensemble optimum in some cases, and the problem can be solved only with a very good database, a best choice of the input configuration for training, or using most powerful learning algorithms (Viotti et al., 2002).

The back-propagation learning algorithm consists of two steps of computation: a forward pass and a backward pass. In the forward pass, an input pattern vector is applied to the sensory nodes of the network, i.e. to the units in the input layer. The signals from the input layer are propagated to the units in the first layer and each unit produces an output. The outputs of these units are propagated to the units in the subsequent layers and this process continues until, finally, the signals reach the output layer, where the actual response of the network to the input vector is obtained (Figure 2.1.1).

During the forward pass, the synaptic weights of the network are fixed. During the backward pass, on the other hand, the synaptic weights are all adjusted in accordance with an error signal, which is propagated backward through the network against the direction of synaptic connections.

The mathematical analysis of the algorithm is as follows (Viotti et al., 2002). In the forward pass, given an input pattern vector $y^{(p)}$, each hidden node-neuron j receives a net input:

$$x_j^{(p)} = \sum_k w_{jk} y_k^{(p)}$$

where w_{jk} represents the weight between the hidden neuron j and the input neuron k . Thus, the hidden neuron j produces an output:

$$y_j^{(p)} = f(x_j^{(p)}) = f\left[\sum_k w_{jk} y_k^{(p)}\right]$$

where $f(x)$ is the activation function of the hidden layer. Different kinds of activation functions are referenced in the literature, such as linear, sigmoid, hyperbolic tangent, logistic, etc. (Norgaard et al., 2000). In the following, we consider a hyperbolic tangent activation function for the neurons in the hidden layer, hence, the value returned by the activation function of neuron j of the hidden layer is:

$$f(x_j) = \frac{e^{2x_j} - 1}{e^{2x_j} + 1}$$

Each output neuron receives the input from the preceding hidden layer by the forecasted value, so that the entry to the output neuron can be written as:

$$x^{(p)} = \sum_j w_j y_j^{(p)} = \sum_j w_j f\left[\sum_k w_{jk} y_k^{(p)}\right]$$

where w_j represents the weight between the output neuron and the hidden neuron j . It therefore produces the final output:

$$y^{(p)} = f(x^{(p)}) = f\left(\sum_j w_j y_j^{(p)}\right) = f\left[\sum_j w_j f\left(\sum_k w_{jk} y_k^{(p)}\right)\right]$$

The presentation of all the patterns is usually called *epoch*. Many epochs are generally needed before the error becomes acceptably small. In the batch neuron the error signal is calculated for each input pattern but the weights are modified only when the input patterns have been presented. The error function is calculated referring to the Mean Square Error (MSE) and the weights are modified accordingly:

$$E = \frac{1}{2}(d - y)^2 = \frac{1}{2}\left\{d - f\left[\sum_j w_j f\left(\sum_k w_{jk} y_k\right)\right]\right\}^2$$

where d is the desired or real output (monitored variable value) and y is the ANN output or the forecasted value. In the batch mode, E is equal to the sum of all MSEs on all the patterns of the training set. E is obviously a differentiable function of all weights (and thresholds) and therefore we can apply the gradient descent method. For the hidden to output connections the gradient descent rule gives:

$$\Delta w_j = -\eta \frac{\partial E}{\partial w_j}$$

where η is a number called *learning rate*. The learning rate is a parameter that determines the size of the weights adjustment each time the weights are changed during the training process. Small values for the learning rate cause small weight changes and large values cause large changes (Attoh-Okine, 1999). The best learning rate is not obvious. If the learning rate is 0.0, the network will never learn.

Refenes et al. (1994) reported that one and two layered network with a learning rate of $\eta=0.2$ and a momentum rate of $0.3 < \alpha < 0.5$ yield the best combination of convergence. The momentum term is a factor used to speed network training. It adds a proportion of the previous weight changes to the current weight changes.

Using the chain rule it can be written as:

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial y^{(p)}} \frac{\partial y^{(p)}}{\partial x^{(p)}} \frac{\partial x^{(p)}}{\partial w_j} = -\sum_p (d^{(p)} - y^{(p)}) f(x^{(p)}) y_j^{(p)}$$

Thus, the hidden to output connections are updated according to the following equation:

$$\Delta w_j = \eta \sum_p (d^{(p)} - y^{(p)}) f(x^{(p)}) y_j^{(p)} = \eta \sum_p \delta^{(p)} y_j^{(p)}$$

where $\delta^{(p)} = (d^{(p)} - y^{(p)}) f(x^{(p)})$

For the input to hidden layer connections the gradient descent rule is:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}}$$

Then using the chain rule, we obtain:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial y_j^{(p)}} \frac{\partial y_j^{(p)}}{\partial x_j^{(p)}} \frac{\partial x_j^{(p)}}{\partial w_{jk}} = \frac{\partial E}{\partial y_j^{(p)}} f(x_j^{(p)}) y_k^{(p)}$$

Particularly, with reference to $\frac{\partial E}{\partial y_j^{(p)}}$, it can be written as:

$$\frac{\partial E}{\partial y_j^{(p)}} = -\sum_p (d^{(p)} - y^{(p)}) \frac{\partial [f(x^{(p)})]}{\partial y_j^{(p)}}$$

and after simple passages, we obtain:

$$\frac{\partial E}{\partial y_j^{(p)}} = -\sum_p (d^{(p)} - y^{(p)}) f(x^{(p)}) w_j$$

Therefore, with reference to $\frac{\partial E}{\partial w_{jk}}$, it can be written as:

$$\frac{\partial E}{\partial w_{jk}} = -\sum_p (d^{(p)} - y^{(p)}) f(x^{(p)}) w_j f(x_j^{(p)}) y_k^{(p)}$$

from which the input to hidden connections updating is obtained as:

$$\Delta w_{jk} = \eta \sum_p (d^{(p)} - y^{(p)}) f(x^{(p)}) w_j f(x_j^{(p)}) y_k^{(p)}$$

and finally we get:

$$\Delta w_{jk} = \eta \sum_p \delta_j^{(p)} w_j f(x^{(p)}) y_k^{(p)} = \eta \sum_p \delta_j^{(p)} y_k^{(p)}$$

with $\delta_j^{(p)} = \delta^{(p)} w_j f(x^{(p)})$

It is worthwhile noting that, a network architecture having just one hidden layer, and activation functions arranged as described above, constitutes a universal predictor and it can theoretically approximate any continuous function to any degree of accuracy. In practice, such degree of flexibility is not achievable because parameters must be estimated from sample data, which are both finite and noisy (Barazzetta & Corani, 2004).

The ANNs work on a matrix containing more patterns. Particularly, the patterns represent the rows while the variables are the columns. This data set is a sample. To be more precise, giving the ANN three different subsets of the available sample we can get the forecasting model; the three subsets concern the training, the validation and the test subsets. These subsets are briefly described:

- *Training subset*, the group of data with which we train-educate the network according to the gradient descent for the error function algorithm, in order to reach the best fitting of the nonlinear function representing the phenomenon.
- *Validation subset*, the group of data, given to the network still in the learning phase, by which the error evaluation is verified, in order to update the best thresholds and weights effectively.
- *Test subset*, one or more sets of new and unknown data for the ANN, which are used to evaluate ANN generalization, i.e. to evaluate whether the model has effectively approximated the general function representative of the phenomenon, instead of learning the parameters uniquely.

3. Air quality indices

Urban air pollution is a growing problem in big cities with large urbanization, where adverse health effects have been established. Bad city design combined with specific topographical and meteorological conditions allowing poor circulation, are associated with frequent episodes of critically high atmospheric pollution, enforcing in some cases extreme actions by the authorities, such as restriction of motor vehicles circulation within large area of the city.

For a better and more effective monitoring and analysis of air quality in big cities, air pollution indices are often used. Most of them have resulted after a series of epidemiological studies, which investigated the impact of air pollution on public health. In this work, two air pollution indices are presented and applied in order to forecast the air quality within GAA using ANNs.

3.1 Description of the European Regional Pollution Index (ERPI)

The European Regional Pollution Index (ERPI) has been proposed and developed by Moustris (2009). This air quality index is based on the air pollution index that is known as Regional Pollution Index (RPI). The New South Wales government in Sydney, Australia used RPI since the mid 1990s (NSW-EPA 1998, 2006).

The calculation of ERPI was performed using the thresholds prescribed by the European Community (EC) based on the framework directive 1996/62/EC and the three affiliated directives 1999/30/EC, 2000/69/EC, and 2002/3/EC (Table 3.1.1). Due to the way of calculation of ERPI, based on EC air pollution thresholds, the Australian RPI was renamed as European Regional Pollution Index (ERPI).

In this work, ERPI was calculated for five main air pollutants. Concretely, the air pollutants concern nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), ozone (O₃) and particulate matter with aerodynamic diameter less than or equal to 10 μm (PM₁₀). For any observed concentration C_i , the value of the sub-index I_i is given by:

$$I_i = \frac{C_i}{Limit_i} \times 50$$

Air Pollutant	Limit values
NO ₂	Hourly value: 200 μg/m ³
SO ₂	Hourly value: 350 μg/m ³
CO	Maximum daily mean value for 8 hours: 10 mg/m ³
O ₃	Maximum daily mean value for 8 hours: 120 μg/m ³
PM ₁₀	Mean daily value: 50 μg/m ³

Table 3.1.1. Limit concentration values of ambient air pollutants according to EC directives.

Once a sub-index I_i is obtained for each air pollutant (Table 3.1.1), the overall ERPI is simply taken as the maximum of all the I_i values according the formula:

$$ERPI = \max\{I_1, I_2, I_3, I_4, I_5\}$$

where $I_1, I_2, I_3, I_4,$ and I_5 are the sub-indices whose values are defined by the NO₂, SO₂, CO, O₃ and PM₁₀, respectively. If ERPI ≥ 50 this means that at least one of the pollutants is over its limit value (Table 3.1.1). Table 3.1.2 presents the classification of air quality according to ERPI values (Moustris, 2009; Moustris et al., 2010).

ERPI	ERPI Class	Classification
0 - 2	1	Very Good
2 - 21	2	Good
21 - 40	3	Satisfactory
40 - 60	4	Sufficient
60 - 79	5	Poor
≥79	6	Very Poor

Table 3.1.2. Air quality classification according to ERPI values.

3.2 Description of Daily Air Quality Index (DAQx)

A new impact-related air quality index obtained on a daily basis and abbreviated as DAQx (Daily Air Quality Index) has been recently developed and tested by the Meteorological Institute of Freiburg, Germany, and the Research and Advisory Institute for Hazardous Substances, Freiburg, Germany (Mayer et al., 2002a, 2002b; Makra et al., 2003). DAQx considers the air Pollutants SO₂, CO, NO₂, O₃ and PM₁₀. To enable a linear interpolation between index classes, DAQx is calculated for each pollutant by:

$$DAQx = \left[\left(\frac{DAQx_{up} - DAQx_{low}}{C_{up} - C_{low}} \right) \times (C_{inst} - C_{low}) \right] + DAQx_{low}$$

with C_{inst}: highest daily 1 hour concentration of SO₂, NO₂, and O₃, highest daily running 8 hours concentration of CO, and mean daily concentration of PM₁₀. C_{up} is the upper threshold of specific air pollutant concentration range; C_{low} is the lower threshold of specific air pollutant concentration range; DAQx_{up} is the value of DAQx according to C_{up}; DAQx_{low} is the value of DAQx according to C_{low} (Table 3.2.1).

The daily value of DAQx is considered the highest value extracted by the calculated values for each pollutant.

SO ₂ (µg/m ³)	CO (mg/m ³)	NO ₂ (µg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	DAQx value	DAQx Class	Classification
0-24	0.0-0.9	0-24	0-32	0.0-9.9	0.5-1.4	1	Very Good
25-49	1.0-1.9	25-49	33-64	10.0-19.9	1.5-2.4	2	Good
50-119	2.0-3.9	50-99	65-119	20.0-34.9	2.5-3.4	3	Satisfactory
120-349	4.0-9.9	100-199	120-179	35.0-49.9	3.5-4.4	4	Sufficient
350-999	10.0-29.9	200-499	180-239	50.0-99.9	4.5-5.4	5	Poor
≥1000	≥30.0	≥500	≥240	≥100	≥5.5	6	Very Poor

Table 3.2.1. Upper and lower limits for air pollutant concentrations and DAQx values, DAQx classes and classification according to Mayer et al. (2002a, 2002b).

4. Bioclimatic indices

The growth of the city of Athens during the last decades and the phenomenon of urbanization (Philandras et al., 1999) have established the well known Urban Heat Island (UHI) at a great areal extent of the city, resulting in explicit effects on human thermal comfort-discomfort. Thermal comfort is defined as the condition of mind, which expresses satisfaction with the thermal environment, absence of thermal discomfort, or conditions in which 80% or 90% of humans do not express dissatisfaction (Givoni, 1998).

Several indices, which describe the human thermal comfort-discomfort, have been developed worldwide. In this chapter three bioclimatic indices will be presented. The Discomfort Index (DI), the Cooling Power index (CP) and the Physiologically Equivalent Temperature (PET). In the process, these indices are briefly described.

4.1 Discomfort Index (DI)

The Discomfort Index (DI) was originally developed by Thom (Thom, 1959) and was supported by later works (Clarke & Bach, 1971; Giles et al., 1990). This index describes the

degree of thermal load under various meteorological conditions, suitable for both outdoor and indoor environments. It is useful to evaluate how current temperature and relative humidity can affect the sultriness or discomfort sensation and cause health danger in the population.

DI (°C)	Classification of human comfort-discomfort sensation
DI < 21	No discomfort feeling
21 ≤ DI < 24	Less than 50% of the total population feels discomfort
24 ≤ DI < 27	More than 50% of the total population feels discomfort
27 ≤ DI < 29	Most of the population feels discomfort
29 ≤ DI < 32	The discomfort is very strong and dangerous
DI ≥ 32	State of medical emergency

Table 4.1.1. Classification of human comfort-discomfort sensation for DI.

Several formulas of the index have been proposed for use along with tables of boundary values that indicate degrees of comfort-discomfort. In the present work we used the following formula of DI, calculated as a combination of air temperature T (°C) and relative humidity RH (Giles et al., 1990):

$$DI = T - (0.55 \times 0.0055 \times RH) \times (T - 14.5), \quad (^\circ\text{C})$$

The classification of the DI values with the equivalent feeling of thermal comfort-discomfort is given in Table 4.1.1 (Giles et al., 1990).

4.2 Cooling Power index (CP)

The Cooling Power Index (CP) was developed by Siple & Passel (1945) and describes the loss of energy, per unit of time and body surface, which a human organism can tolerate. The CP index, in contrast to the DI index, takes into consideration the wind speed instead of relative humidity. It describes the heat flux per surface unit of the human body towards the environment and the vice versa. For the calculation of the CP index hourly values of air temperature (T , °C) and wind speed (V , m/sec) were used. The calculation of CP is based on the following formula (Tzenkova et al., 2003):

$$CP = 1.163 \times (10.45 + 10 \times V^{0.5} - V) \times (33 - T), \quad (\text{W/m}^2)$$

The classification of the CP index values modified by Besancenot et al. (1978), with the equivalent feeling of thermal comfort-discomfort is given in Table 4.2.1.

CP (W/m ²)	Classification of human comfort
CP < 0	Endothermal - very hot discomfort
0 < CP ≤ 174	Atonic - hot discomfort
175 ≤ CP ≤ 349	Hypotonic - hot sub comfort
350 ≤ CP ≤ 699	Neutral - comfort
700 ≤ CP ≤ 1049	Tonic - cold sub comfort
CP ≥ 1050	Cold discomfort

Table 4.2.1. Classification of human comfort-discomfort sensation for CP.

4.3 Physiologically Equivalent Temperature (PET)

The thermal index Physiologically Equivalent Temperature (PET) is based on the total energy balance of the human body. PET values were evaluated (Mayer & Höppe, 1987; Höppe, 1999), in order to interpret the grade of the thermophysiological stress (Table 4.3.1). It describes the effect of the thermal environment as a temperature value (°C) and can be quantified easier for non specialists in this topic. For night time situation, air temperature corresponds very close to the PET value. It has been applied in heat waves and climatic variability studies (Nastos & Matzarakis 2008, Matzarakis & Nastos 2010) and weather impacts on health (Nastos & Matzarakis, 2006).

The PET analysis was performed by the use of the radiation and bioclimate model, RayMan, which is well-suited to calculate radiation fluxes and human biometeorological indices (Matzarakis et al., 1999, 2010) and was chosen for all our calculations of mean radiant temperature and PET. The RayMan model, developed according to the Guideline 3787 of the German Engineering Society (VDI, 1998) calculates the radiation flux in easy and complex environments on the basis of various parameters, such as air temperature, air humidity, degree of cloud cover, time of day and year, albedo of the surrounding surfaces and their solid-angle proportions (Matzarakis et al., 2010).

PET (°C)	Thermal sensation	Physiological stress level
< 4	very cold	extreme cold stress

8	cold	strong cold stress

13	cool	moderate cold stress

18	slightly cool	slight cold stress

23	comfortable	no thermal stress

29	slightly warm	slight heat stress

35	warm	moderate heat stress

41	hot	strong heat stress

> 41	very hot	extreme heat stress

Table 4.3.1. Physiologically Equivalent Temperature (PET) for different grades of thermal sensation and physiological stress on human beings (during standard conditions: heat transfer resistance of clothing: 0.9 clo, internal heat production: 80 W) (Matzarakis & Mayer, 1996)

5. Statistical performance indices

The quality and reliability of the developed ANNs, concerning their ability to forecast both air quality and bioclimatic conditions within GAA, were tested using several statistics

indices that have already been applied in similar studies (Moustris et al., 2010). The statistical performance indices that used in this work are presented and described briefly:

$$\text{Mean Bias Error: } MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)$$

where N is the number of the data points, O_i is the observed data and P_i is the predicted data. The MBE represents the degree of correspondence between the mean forecast (P_i) and the mean observation (O_i). MBE is used to describe how much the model underestimates or overestimates the observed data. Positive/negative values indicate over estimated/under estimated prediction.

$$\text{Root Mean Square Error: } RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}$$

RMSE provides a measure of how well future outcomes are likely to be predicted by the model.

The coefficient of determination (R^2) indicates how much of the observed variability is accounted by the estimated model (Kolehmainen et al., 2001). The coefficient of determination is a number between 0 and +1 and measures the degree of association between two variables. The coefficient of determination is calculated according to the equation (Comrie, 1997):

$$R^2 = \frac{\sum_{i=1}^N (P_i - O_{iave})^2}{\sum_{i=1}^N (O_i - O_{iave})^2}$$

where O_{iave} is the average of the observed data.

A relative measure of error, called the index of agreement (IA), is also discussed in Willmott et al. (1985). Index of agreement is calculated according to the formula:

$$IA = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - O_{iave}| + |O_i - O_{iave}|)^2}$$

where O_{iave} is the average of the observed data. This is a dimensionless measure that is limited to the range of 0-1. If $IA=0$, that means no agreement between prediction and observation and if $IA=1$, that means perfect agreement between prediction and observation.

6. Data and methodology

For the calculation of the bioclimatic indices as well as the air quality indices, appropriate meteorological data in hourly basis were used. More specifically, hourly values of air temperature ($^{\circ}\text{C}$), relative humidity (%) and wind speed (m/s) were used for DI and CP

calculation. In addition to the aforementioned meteorological parameters, total cloudiness cover (octas) was taken into consideration for PET calculation, using the RayMan model (Matzarakis et al., 1999, 2010). The appropriate meteorological parameters used as inputs in the RayMan model were acquired from the National Observatory of Athens, for the period 2001-2004. Besides, hourly values of air pollutants concentrations (NO_2 , SO_2 , CO , O_3 and PM_{10}) were used in order to estimate the two air quality indices ERPI and DAQx. All the above datasets have been recorded by the network of the GMEECC covering the period 2001- 2005 and concern nine (9) different regions within the GAA, namely the regions: Agia Paraskevi, Thracomakedones, Lykovrissi, Maroussi, Liossia, Galatsi, Patission, Aristotelous, and Geoponiki (Fig. 6.1). For a better surveillance, the examined regions-stations listed below with the following abbreviations: Agia Paraskevi (APA), Galatsi (GAL), Liossia (LIO), Maroussi (MAR), Patission (PAT), Aristotelous (ARI), Thracomakedones (THR), Lykovrissi (LYK). The hourly values of air barometric pressure and total solar irradiance for the same time period were obtained from the National Observatory of Athens.

To describe the air quality within the GAA the values of the air quality indices ERPI and DAQx were calculated on an hourly basis in seven different regions-stations (APA, THR, LYK, MAR, LIO, GAL and PAT) with respect to the pollutants NO_2 , SO_2 , CO , and O_3 and in five different regions-stations (APA, THR, LYK, MAR and ARI) with respect to the particulate matter PM_{10} . The maximum value of the 24 hourly values was considered as the daily value for each one of the two air quality indices. Thus, for each one station-region two daily values for each one of the two examined air quality indices were calculated. The first daily value concerns the air pollutants NO_2 , SO_2 , CO , and O_3 and the second concerns the particulate matter PM_{10} . This happened because the daily concentrations of particulate matter PM_{10} as well as the daily concentrations of ozone are both high enough. If only one daily value for each of the two air quality indices was calculated, then, we will not be able to know if that value is due to ozone or PM_{10} . Thereafter, an appropriate number of ANN models were developed and trained in order to predict for the next three days the daily value for each one of the two air quality indices as well as the number of consecutive hours during the day where the value of the index is greater than a threshold value.



Fig. 6.1 Map of Greece (left graph) and spatial distribution of the GMEECC network's stations within GAA (right graph).

The bioclimatic conditions within the GAA are interpreted by the use of the bioclimatic indices DI and CP, which were calculated on hourly basis for eight different regions-stations (APA, THR, LYK, MAR, LIO, GAL, GEO and PAT). In the process, the daily value for each

index in each region-station was calculated. The calculation was carried out only during the warm period of the year (May-September) in order to describe the human discomfort due to heat stress weather conditions. Then, an appropriate number of ANNs were developed and trained in order to predict for the next three days the daily value for each one of the two bioclimatic indices as well as the number of consecutive hours during the day, where the value of the index is greater than a threshold value ($DI \geq 24$ °C) or less than a threshold value ($CP \leq 174$ W/m²). Furthermore, the mean daily values of PET index were estimated only for the National Observatory of Athens, because of the availability of the total parameters needed as inputs in RayMan model. Thereafter, the developed ANN was evaluated in forecasting PET for the next three days.

7. Air quality forecasting using ANNs

7.1 ANNs description

Six different ANNs were developed in order to forecast the air quality levels within the GAA. The first one (ANN#1) was trained in order to forecast the daily value of ERPI (for the pollutants CO, NO₂, SO₂ and O₃) for the next day at seven different areas of GAA (APA, THR, LYK, MAR, LIO, GAL and PAT). The second one (ANN#2) was trained in order to forecast the daily value of DAQx (for the pollutants CO, NO₂, SO₂ and O₃) for the next day at the above seven different areas within the GAA. The third one (ANN#3) was trained in order to forecast the daily number of the consecutive hours for the next day, with at least one of the pollutants concentrations (CO, NO₂, SO₂ and O₃) above a threshold according to directives of European Community, for each one of the seven examined stations within the GAA. The fourth (ANN#4) was trained in order to forecast the daily value of ERPI (with respect to PM₁₀) for the next day, at five different areas of GAA (APA, THR, LYK, MAR, and ARI). The fifth (ANN#5) was trained in order to forecast the daily value of DAQx (with respect to PM₁₀) for the next day, at the mentioned five different areas within the GAA. Finally the sixth (ANN#6) was trained in order to forecast the daily number of the consecutive hours for the next day with the PM₁₀ concentrations above a threshold according to EC directives, for each one of the five examined stations within the GAA.

In each case, the group of data defined as “the training set”, used for ANNs training, concerns the time period 2001-2004. The group of data defined as “the validation set”, given to the network still in the learning phase, accounts 20% of ‘the training set’ for each one of the developed ANN models. Finally “the test set” refers to the year 2005. The year 2005 is absolutely unknown to the models, in order to reveal the models forecasting ability. Table 6.1 presents the input and output data for the six developed ANN models.

The combination of selected data for the appropriate ANN models training was done after a series of several tests (trial and error method). At the end, the combination that gave the best forecasting result in each case was selected (Table 7.1.1).

In this point, we have to mention that for all the constructed ANN models we have used as input data, in addition to other parameters, the maximum and minimum air temperature, the maximum and minimum wind speed for the next day as well as the mean daily air barometric pressure and the mode daily wind direction for the next day. This may produce a limitation in the forecasting attempt, but it is easy to have access to these forecasted values through the network of the Hellenic National Meteorological Service (HNMS).

INPUT DATA (input layer)	ANN#1	ANN#2	ANN#3	ANN#4	ANN#5	ANN#6
Stations' number (1,2,3,4,5,6,7)	√	√	√	√	√	√
Month (1,2,3,...,12)	√	√	√	√	√	√
Mean daily air pressure (mbar) for the six previous days	√	√	√	√	√	√
Daily sum of the global solar irradiance for the six previous days (W/m ²)	√	√	√	√	√	√
Maximum (T _{max}) and minimum (T _{min}) daily temperature (°C) for the six previous days	√	√	√	√	√	√
Maximum (WS _{max}) and minimum (WS _{min}) daily wind speed (m/sec) for the six previous days	√	√	√	√	√	√
Maximum (RH% _{max}) and minimum (RH% _{min}) daily relative humidity for the six previous days		√			√	
Cosine and sine of the mode daily wind direction for the six previous days	√	√	√	√	√	√
ERPI daily value for the six previous days	√		√	√		√
DAQx daily value for the six previous days		√			√	
The number of consecutive hours during the day with ERPI≥50 for the six previous days			√			√
Mean daily air pressure (hPa) one day ahead	√	√	√	√	√	√
Maximum (T _{max}) and minimum (T _{min}) daily temperature (°C) one day ahead	√	√	√	√	√	√
Maximum (RH% _{max}) and minimum (RH% _{min}) daily relative humidity one day ahead		√			√	
Maximum (WS _{max}) and minimum (WS _{min}) daily wind speed (m/sec) one day ahead	√	√	√	√	√	√
Cosine and sine of the mode daily wind direction one day ahead	√	√	√	√	√	√
OUTPUT DATA (output layer)						
ERPI daily value for the next day	√			√		
DAQx daily value for the next day		√			√	
The number of consecutive hours with ERPI≥50 for the next day			√			√

Table 7.1.1. Input and output data for the appropriate training of the six developed ANN models.

7.2 Forecasting of daily ERPI and DAQx values for the next day

The global fit agreement statistical indices as well as the excess statistical indices for the observed and predicted ERPI and DAQx values were calculated and demonstrated for the eight examined stations respectively. More specifically, O_{ave} , P_{ave} , MBE, RMSE, IA and R^2 values for ERPI index are presented in Table 7.2.1.

	ANN#1						ANN#2						ANN#3					
	O_{ave}	P_{ave}	MBE	RMSE	IA	R^2	O_{ave}	P_{ave}	MBE	RMSE	IA	R^2	O_{ave}	P_{ave}	MBE (hours)	RMSE (hours)	IA	R^2
APA	44.0	42.5	-1.460	6.629	0.925	0.752	3.4	3.4	0.013	0.396	0.853	0.557	2.8	3.2	0.374	3.015	0.877	0.605
GAL	35.6	35.7	0.118	0.644	0.903	0.697	3.3	3.3	0.060	0.471	0.751	0.378	0.5	0.7	0.199	1.415	0.657	0.231
LIO	37.7	37.8	0.114	6.081	0.919	0.726	3.3	3.3	0.032	0.463	0.746	0.358	0.7	0.9	0.168	1.761	0.671	0.238
MAR	38.3	37.1	-1.272	5.852	0.920	0.738	3.3	3.3	0.000	0.428	0.795	0.446	0.8	0.8	-0.008	1.825	0.708	0.292
PAT	33.1	32.2	-0.944	9.233	0.717	0.381	3.7	3.7	-0.024	0.343	0.791	0.442	0.3	0.3	0.042	1.365	0.299	0.017
THR	42.4	40.6	-1.810	7.290	0.922	0.760	3.3	3.3	-0.002	0.450	0.876	0.637	3.9	3.7	-0.208	4.924	0.829	0.516
LYK	38.9	38.6	-0.317	7.317	0.937	0.826	3.3	3.3	0.021	0.431	0.889	0.686	2.0	1.4	-0.590	3.092	0.742	0.401
	ANN#4						ANN#5						ANN#6					
	O_{ave}	P_{ave}	MBE	RMSE	IA	R^2	O_{ave}	P_{ave}	MBE	RMSE	IA	R^2	O_{ave}	P_{ave}	MBE (hours)	RMSE (hours)	IA	R^2
APA	40.0	40.0	0.356	18.718	0.674	0.266	3.7	3.6	-0.006	0.639	0.911	0.673	6.0	6.0	-0.310	4.259	0.764	0.377
MAR	46.0	46.0	-0.225	19.441	0.699	0.306	3.9	3.9	0.000	0.646	0.930	0.629	8.0	7.0	-0.141	4.908	0.791	0.425
THR	29.0	31.0	2.026	13.285	0.792	0.487	3.0	3.1	0.076	0.636	0.895	0.689	2.0	3.0	0.485	4.002	0.690	0.361
LYK	53.0	53.0	-0.195	19.927	0.729	0.381	4.2	4.2	-0.066	0.634	0.965	0.611	10.0	10.0	-0.303	5.594	0.816	0.492
ARI	53.0	53.0	0.292	18.352	0.721	0.295	4.3	4.3	0.023	0.528	0.779	0.578	10.0	9.0	-1.084	5.845	0.785	0.430

Table 7.2.1. Global fit agreement indices for ERPI predicted values for the next day.

Concerning the pollutants CO, NO₂, SO₂ and O₃, the R^2 values show a very satisfactory prediction for ERPI-ANN#1 ($0.381 \leq R^2 \leq 0.826$) as well as for the DAQx-ANN#2 ($0.378 \leq R^2 \leq 0.686$) during the test year 2005. Besides, IA values show also a very good prediction for ERPI-ANN#1 ($0.717 \leq IA \leq 0.937$) and the DAQx-ANN#2 ($0.746 \leq IA \leq 0.889$). In all cases, it seems that the prediction for the pollutants CO, NO₂, SO₂ and O₃ is much more successful using the ERPI, which is according to the European Community directives, instead of the DAQx. But using both predictions we can have a better and safe "picture" about air quality one day ahead within the GAA. As far as the air pollution persistence (for the pollutants CO, NO₂, SO₂ and O₃) is concerned, it seems that ANN#3 gives an adequate prediction. The R^2 values range between 0.017 and 0.605 while IA range between 0.299 and 0.877.

Finally, the worst prediction with respect to the air quality index ERPI appears for the region-station PAT (city centre) against the region-station LIO (urban area) concerning the air quality index DAQx. Generally, it seems that the prediction for the stations, which are closer to the GAA's downtown, is not so good compared to the prediction of the peripheral regions-stations. This is likely due to the traffic load and the bad air circulation within the

city's centre, meaning that, more relevant data, associated with the above mentioned factors, are needed for a better ANNs training.

Figure 7.2.1 presents the best prediction (LYK) and the worst prediction (PAT) for ERPI concerning the pollutants CO, NO₂, SO₂ and O₃, while the best prediction (LYK) and the worst prediction (LIO) for DAQx concerning the same pollutants are depicted in Figure 7.2.2. Accordingly, Figure 7.2.3 presents the best prediction (THR) and the worst prediction (APA) for ERPI concerning the pollutant PM₁₀, and Figure 7.2.4 shows the best prediction (LYK) and the worst prediction (ARI) for DAQx with respect to the pollutant PM₁₀. During the warm period of the year (May-September) the values of ERPI (Figure 7.2.1) are greater than 50, meaning that at least one pollutant's concentration is above its threshold according to the EC directives. In most cases (more than 90%) the corresponding pollutant for these high values of ERPI is the ozone. The same results revealed from Figure 7.2.2 regarding DAQx, where during the warm period of the year the daily values of DAQx are greater than 3.5, meaning that a bad air quality exist in most cases. As far as the PM₁₀ concentrations are concerned (Figures 7.2.3 and 7.2.4), it is shown that, for almost half of the days throughout the year are above the threshold concentration value, indicating bad air quality in the most of the examined stations-regions.

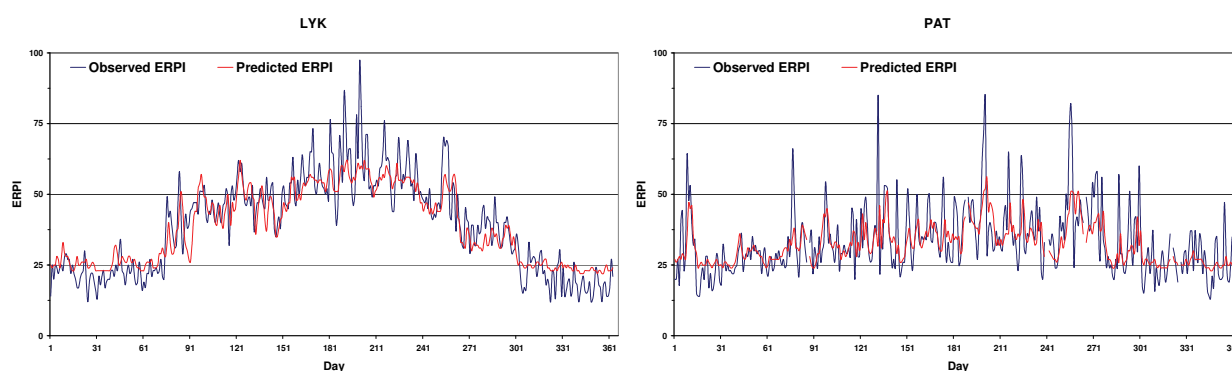


Fig. 7.2.1. Predicted vs. observed ERPI values for the CO, NO₂, SO₂ and O₃, pollutants during the test year 2005.

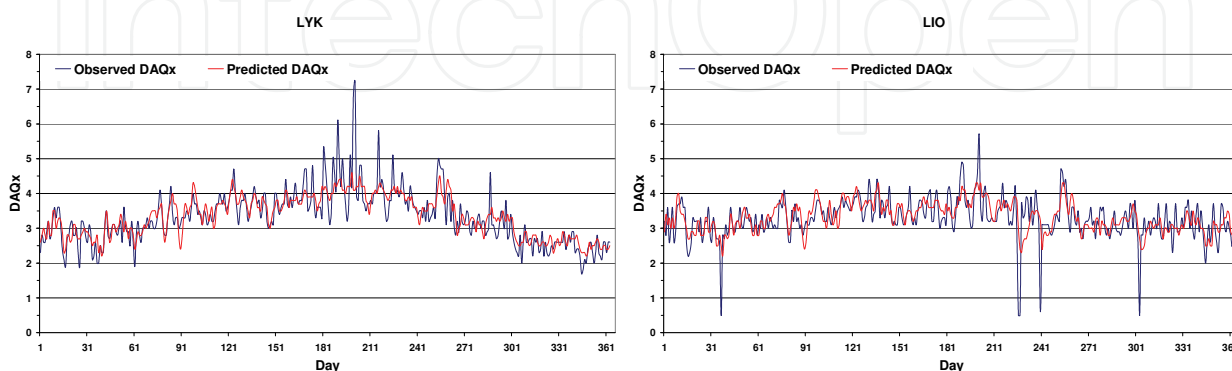


Fig. 7.2.2. Predicted vs. observed DAQx values for the CO, NO₂, SO₂ and O₃, pollutants during the test year 2005.

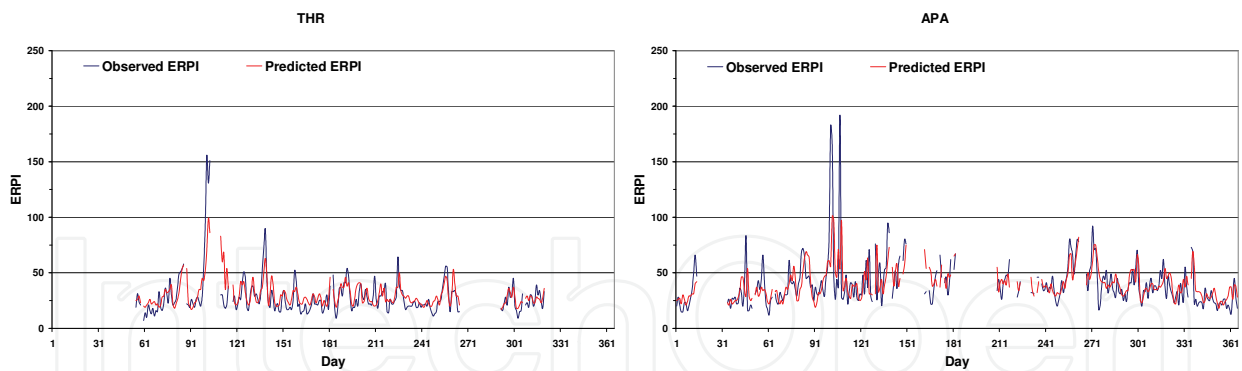


Fig. 7.2.3. Predicted vs. observed ERPI values for the PM₁₀ pollutant during the test year 2005.

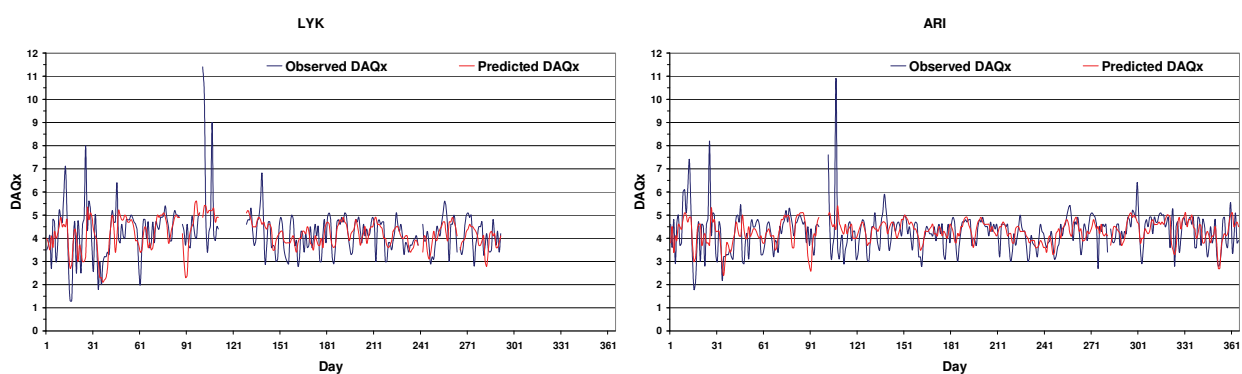


Fig. 7.2.4. Predicted vs. observed DAQ_x values for the PM₁₀ pollutant during the test year 2005

8. Bioclimatic conditions forecasting using ANNs

8.1 ANNs description for DI and CP forecasting

Four different ANN models were developed in order to forecast the bioclimatic conditions within the GAA during the warm period of the year (May-September). The first one (ANN#7) was trained in order to forecast the daily value of Thom's DI index for the next day at eight different areas of GAA (APA, THR, LYK, MAR, LIO, GAL, GEO and PAT). The second one (ANN#8) was trained in order to forecast the daily value of CP index for the next day at the above mentioned eight different areas within the GAA. The third one (ANN#9) was trained in order to forecast the daily number of the consecutive hours with DI ≥ 24 °C for the next day at each one of the eight examined stations within the GAA. Finally, the fourth (ANN#10) was trained in order to forecast the daily number of the consecutive hours with CP ≤ 174 W/m² for the next day at each one of the eight examined stations within the GAA.

In each case the group of data named as "the training set" used for ANNs training concerns the time period 2001-2004. The group of data named as "the validation set" given to the network still in the learning phase accounts 20% of the training set for each one of the above ANNs. Finally "the test set" refers to the year 2005, which is absolutely unknown to the models in order to reveal the models forecasting ability. Table 8.1.1 presents the input and output data for the four developed ANNs. The combination of selected data for the appropriate ANN models training was done after a series of several tests (trial and error method). At the end, the combination that gave the best forecasting result in each case was selected (Table 8.1.1).

INPUT DATA (input layer)	ANN#7	ANN#8	ANN#9	ANN#10
Stations' number (1,2,3,4,5,6,7)	√	√	√	√
Month (5,6,7,8,9)	√	√	√	√
The maximum (T_{max}) daily temperature for the six previous days.	√	√	√	
The maximum (RH_{max}) daily relative humidity for the six previous days.	√		√	
The maximum (DI_{max}) daily value of DI for the six previous days.	√		√	
The daily number of consecutive hours with $DI \geq 24$ °C for the six previous days.	√		√	
The maximum (V_{max}) daily wind speed for the six previous days.		√		
The minimum (CP_{min}) daily value of CP for the six previous days.		√		
The daily number of consecutive hours with $CP \leq 174$ W/m ² for the six previous days.		√		√
The maximum (T_{max}) and minimum (T_{min}) daily temperature for the six previous days.				√
The maximum (V_{max}) and minimum (V_{min}) daily wind speed for the six previous days.				√
The maximum (CP_{max}) and minimum (CP_{min}) daily value of CP for the six previous days.				√
OUTPUT DATA (output layer)				
The maximum (DI_{max}) daily value of DI for the next day.	√			
The minimum (CP_{min}) daily value of CP for the next day.		√		
The daily number of consecutive hours with $DI \geq 24$ °C for the next day.			√	
The daily number of consecutive hours with $CP \leq 174$ W/m ² for the next day.				√

Table 8.1.1. Input and output data for the appropriate training of the four developed ANNs.

8.2 DI and CP daily value forecasting for the next day

The global fit agreement statistical indices as well as the excess statistical indices for the observed and predicted values were calculated and demonstrated for the eight examined stations respectively. More specifically, O_{ave} , P_{ave} , MBE, RMSE, IA and R^2 values for DI are presented in Table 8.2.1.

The R^2 values show a very satisfactory prediction for DI-ANN#7 ($0.676 \leq R^2 \leq 0.841$) during the test year 2005 as well as for the CP-ANN#8 ($0.591 \leq R^2 \leq 0.814$). Concerning the IA values, a very satisfactory prediction for DI-ANN#7 ($0.849 \leq IA \leq 0.956$) as well as for the CP-ANN#8 ($0.813 \leq IA \leq 0.948$) appears. Taking into consideration the persistence of the phenomenon with respect to the daily number of consecutive hours with high discomfort conditions, due to strong heat stress, it seems that ANN#9 and ANN#10 give an adequate prediction. Additionally, the R^2 values show a very satisfactory prediction for ANN#9 (0.140

	ANN#7						ANN#8					
	O _{ave}	P _{ave}	MBE (°C)	RMSE (°C)	IA	R ²	O _{ave}	P _{ave}	MBE (W/m ²)	RMSE (W/m ²)	IA	R ²
APA	23.2	23.3	0.082	0.972	0.942	0.794	141	135	-6.196	49.754	0.920	0.731
GAL	24.1	24.0	-0.057	0.889	0.952	0.824	98	97	-1.137	39.863	0.948	0.814
GEO	23.2	23.3	0.111	0.831	0.956	0.841	143	130	-12.739	44.964	0.942	0.806
LIO	23.5	23.4	-0.069	0.929	0.944	0.801	126	126	0.145	48.369	0.930	0.764
MAR	24.0	23.9	-0.145	1.046	0.934	0.773	115	113	-2.154	48.893	0.926	0.743
PAT	25.0	25.1	0.090	0.889	0.953	0.832	82	78	-4.338	46.999	0.936	0.792
THR	19.9	20.7	0.779	1.584	0.849	0.676	292	246	-45.137	80.978	0.813	0.591
LYK	22.5	22.7	0.247	1.051	0.929	0.771	167	160	-6.791	55.554	0.878	0.664
	ANN#9						ANN#10					
	O _{ave}	P _{ave}	MBE (hours)	RMSE (hours)	IA	R ²	O _{ave}	P _{ave}	MBE (hours)	RMSE (hours)	IA	R ²
APA	3.6	4.2	0.654	2.610	0.930	0.760	7.0	8.0	1.333	4.372	0.907	0.702
GAL	6.6	6.4	-0.170	3.135	0.951	0.832	12.0	12.0	0.255	4.158	0.946	0.812
GEO	4.0	4.3	0.301	2.526	0.946	0.810	8.0	9.0	0.667	4.292	0.931	0.762
LIO	4.1	4.3	0.204	2.475	0.943	0.795	8.0	9.0	0.849	3.959	0.932	0.764
MAR	6.1	6.0	-0.112	3.254	0.943	0.797	10.0	10.0	0.430	4.436	0.934	0.766
PAT	10.9	11.0	0.103	3.601	0.901	0.861	13.0	14.0	0.524	4.923	0.928	0.751
THR	0.1	0.6	0.541	1.126	0.368	0.104	1.0	2.0	1.281	2.716	0.750	0.443
LYK	2.0	2.6	0.669	2.083	0.897	0.680	5.0	6.0	1.118	3.836	0.903	0.689

Table 8.2.1. Global fit agreement indices for DI predicted values for the next day.

$\leq R^2 \leq 0.832$) as well as for the CP-ANN#10 ($0.443 \leq R^2 \leq 0.812$) during the test year 2005. Besides, the IA values, show a very satisfactory prediction regarding ANN#9 ($0.368 \leq IA \leq 0.951$) and ANN#10 ($0.750 \leq IA \leq 0.946$). The worst prediction for the daily number of consecutive hours with high discomfort conditions, due to strong heat stress, refers to the region-station of THR (suburban region-station). This may be attributed to the fact that, in this suburban region (Thrakomakedones) the bioclimatic conditions are better than all the other examined regions within the GAA due to lower temperature values. Both discomfort indices, DI and CP, present daily values over their thresholds for a short period of time during the examined period. Thus, there is not a "memory-experience" of the persistence in THR, so the developed ANN models cannot have the appropriate training in order to forecast the number of consecutive hours with strong discomfort.

Figure 8.2.1 reveals that within the city's centre (PAT), the strong discomfort conditions ($DI \geq 24$ °C) appear from the end of June to the first half of September. At the suburban station (THR) there is not a significant discomfort, according to DI values. Just a few days during the warm period of the year appear to be over the threshold of $DI \geq 24$ °C; namely at least 50% of the population feels discomfort due to heat stress.

Figure 8.2.2 illustrates that close to the city's center (urban area of Galatsi), the hot sub comfort conditions according to CP values ($CP \leq 174$ W/m²) appear from the middle of June until the first half of September. At the suburban station (THR), the discomfort due to heat stress conditions starts at the beginning of July until the middle of August. In all the above cases it seems that the prediction of bioclimatic conditions one day ahead with the use of ANN models is very satisfactory and realizable.

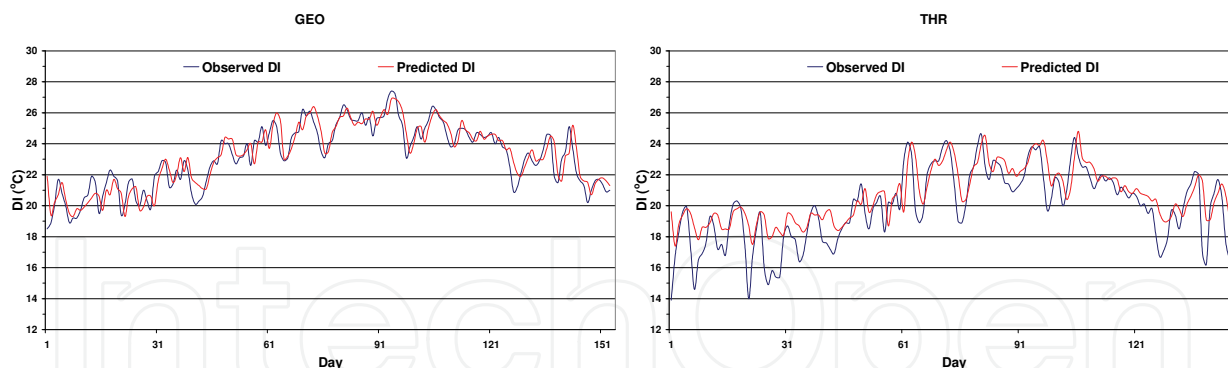


Fig. 8.2.1. Predicted vs. observed values of DI for the next day, concerning the best (GEO) and the worst (THR) prediction for DI daily maximum value, during the warm period of the test year 2005.

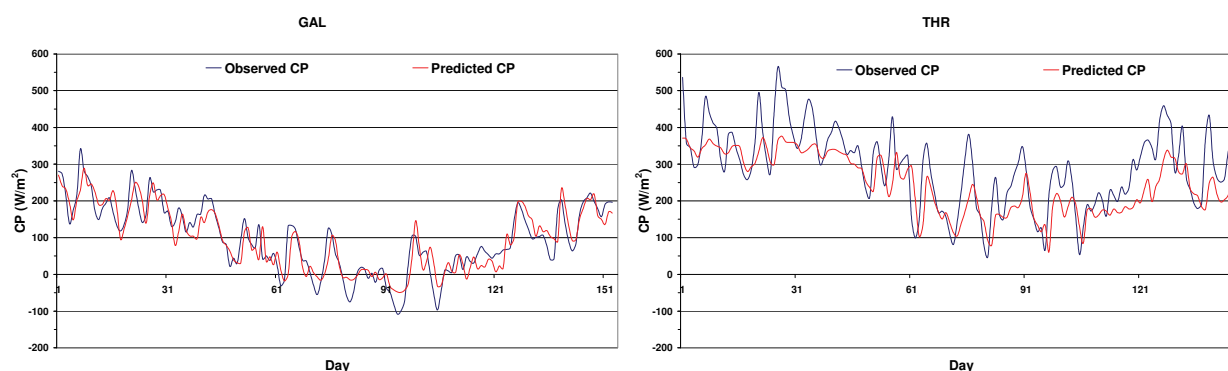


Fig. 8.2.2. Predicted vs. Observed values of CP for the next day, concerning the best (GAL) and the worst (THR) prediction for CP daily minimum value, during the warm period of the test year 2005.

8.3 ANNs description for PET forecasting

Three developed ANNs were trained using back-propagation algorithm to forecast the mean daily PET value for the next day (ANN#11), two next days (ANN#12) and three next days (ANN#13). The training dataset concern the period 2001-2003, while the validation dataset concern the year 2004, which was absolutely unknown to the constructed model, in order to test the predictive ability of the model. Superposed epoch analysis on the training datasets indicated that three days before the incidence of strong heat/cold stress are adequate to forecast PET value for the next days. Thus, the input data (Table 8.3.1) which were taken for ANNs training concern the mean daily air temperature, relative humidity, wind speed and sunshine for the previous three days from the National Observatory of Athens.

Table 8.3.2 presents the fit agreement indices between the observed and the predicted PET values, for the validation year 2004. It is remarkable the high values of IA and R², which indicate that the constructed ANNs have an excellent forecasting ability of PET for the next three days. This gives evidence that the developed ANNs, taking into account simple meteorological parameters recorded in the previous three days, are capable to predict a bioclimatic index, which is not easily calculated (PET was estimated using the RayMan

model), while the most remarkable finding is that of pronounced agreement between observed and predicted PET values. Figure 8.3.1 depicts the predicted and observed mean daily PET time series for the next day (a), the next two days (b) and the next three days (c), along with the respective scatter plots.

INPUT DATA (input layer)	ANN#11	ANN#12	ANN#13
Mean daily air temperature (°C) for the three previous days	√	√	√
Mean daily wind speed (m/s) for the three previous days	√	√	√
Mean daily relative humidity (RH%) for the three previous days	√		
The sunshine duration (hours) for the three previous days	√	√	√
OUTPUT DATA (output layer)			
Mean daily PET value for the next day	√		
4Mean daily PET value for the next two days		√	
Mean daily PET value for the next three days			√

Table 8.3.1. Input and output data for the appropriate training of the developed ANN#11.

	MBE	RMSE	IA	R ²
Mean daily PET value for the next day (ANN#11)	+0.5	2.8	0.982	0.933
Mean daily PET value for the next two days (ANN#12)	+0.5	3.8	0.966	0.874
Mean daily PET value for the next three days (ANN#13)	+0.4	4.3	0.956	0.839

Table 8.3.2. Global fit agreement indices for PET predicted values for the next one, two and three days.

9. Spatial distribution of air quality and bioclimatic conditions in the GAA

9.1 Spatial variation of air quality within GAA

The mean annual value for both air quality indices ERPI and DQAx was calculated at all the examined regions within GAA, during the time period 2001-2005. Figure 9.1.1 shows the spatial variation of air quality levels within GAA. As far as the air quality index ERPI is concerned, only the station THR appears a satisfactory air quality level in annual basis (ERPI < 40). The stations MAR, APA and GAL appear a tolerable air quality level (ERPI < 50). Moreover, the air quality levels at LYK, LIO and PAT stations are very close to the limit value of ERPI ≥ 50. Finally, the air quality level appears to be poor in the city centre station ARI. This may be attributed to the high PM₁₀ concentration levels almost during the whole year. In this point, we have to mention that the station PAT is also in the centre of the city and very close to the ARI station, but unfortunately for this station we don't have any PM₁₀ observations. Similar conclusions are extracted with respect to the air quality index DAQx. The only exception is the LIO station in which the air quality levels seems to be much closer to the stations GAL, MAR and APA.

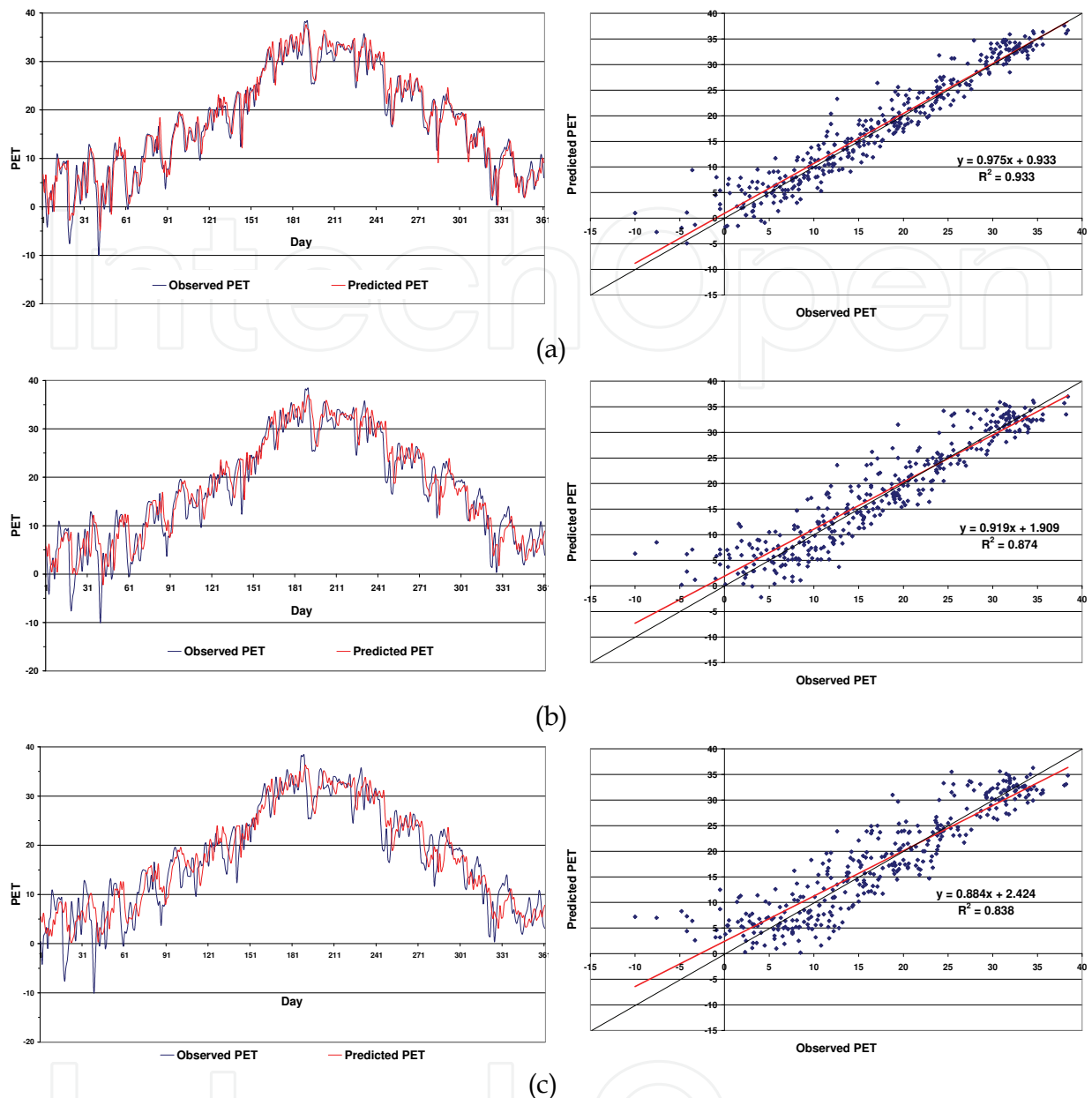


Fig. 8.3.1. Mean daily PET predicted values vs. observed values, for the next day (a), next two days (b) and next three days (c) three days, for the test year 2004.

9.2 Spatial variation of bioclimatic conditions within GAA

During the period 2001-2005, the mean annual value for both bioclimatic indices DI and CP was calculated at all the examined regions within the GAA. Figure 9.2.1 depicts the spatial variation of the bioclimatic conditions within the GAA during the warm period of the year (May-September), where three different bioclimatic zones appear. The first zone is the north suburban zone (THR), which can be characterized as a comfortable zone. The second zone extends peripherally the city's center (LIO, LYK, MAR and APA) and can be marginally characterized as a comfortable zone or warm zone. Finally, the third zone concerns the city's center (GAL, PAT and GEO), which can be characterized as an uncomfortable zone or a strong heat stress zone.

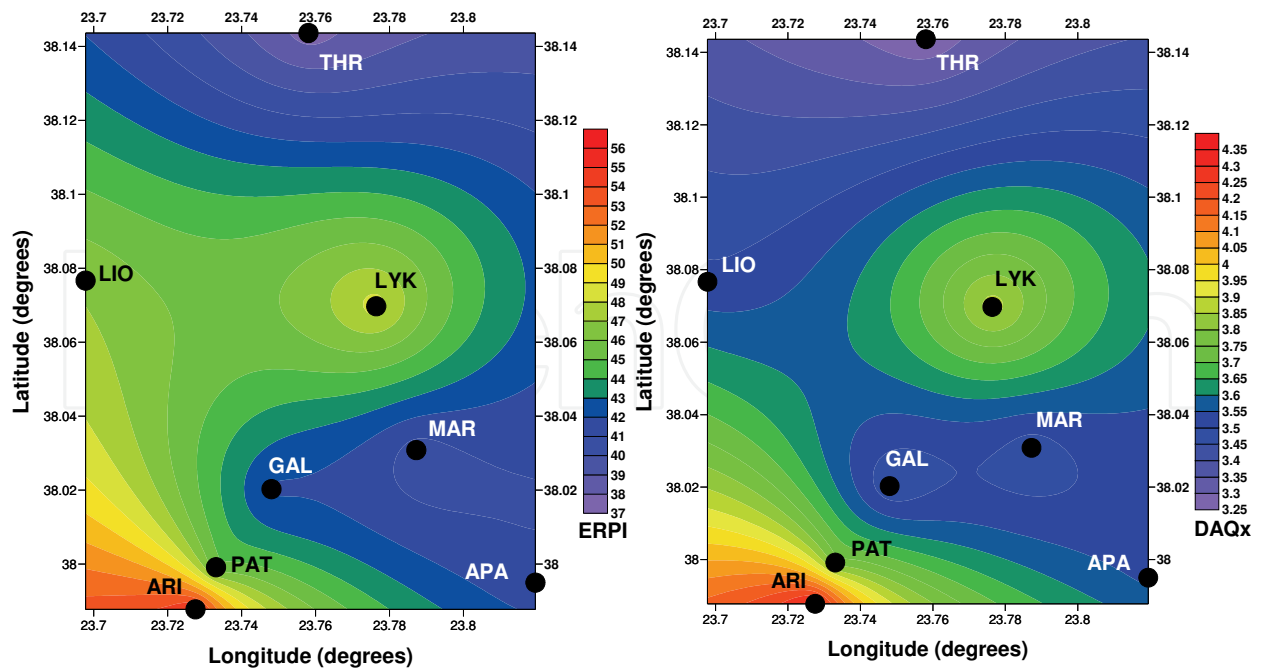


Fig. 9.1.1. Spatial variation of air quality levels within GAA for the pollutants NO_2 , SO_2 , CO , O_3 and PM_{10} . Mean annual values of ERPI (left graph) and DAQx (right graph), during 2001-2005.

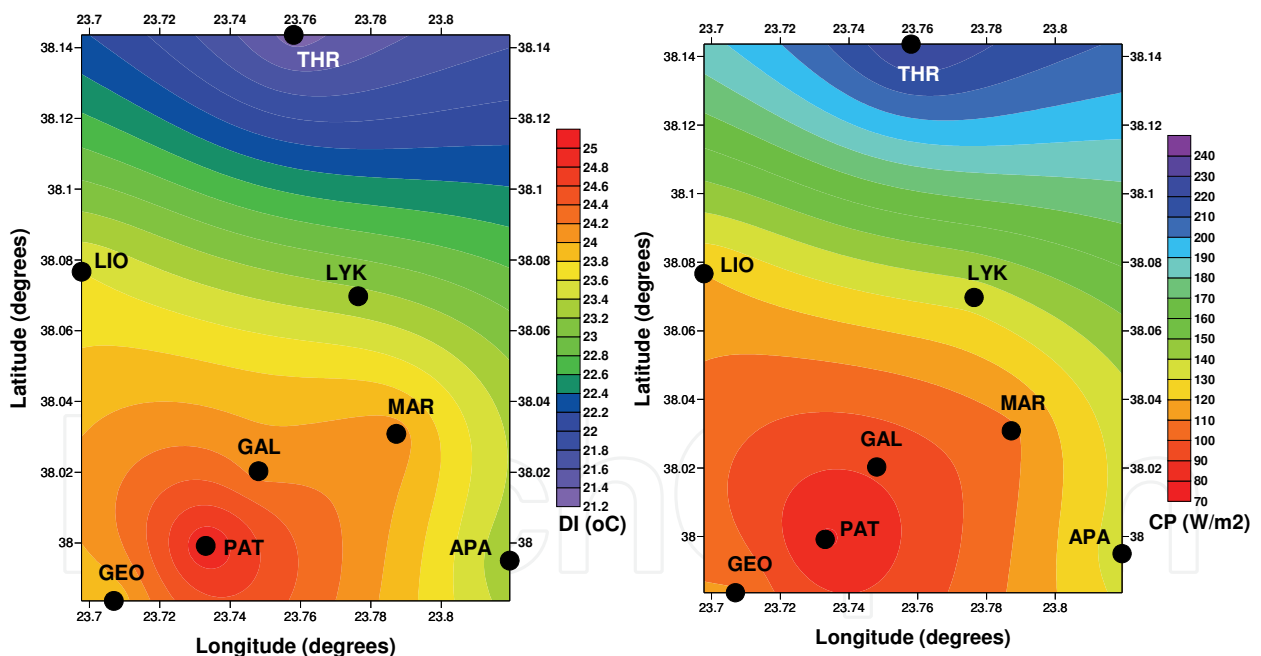


Fig. 9.2.1. Spatial variation of bioclimatic conditions within the GAA. Mean annual values of DI (left graph) and CP (right panel), for the warm period of the year during 2001-2005.

As far as the persistence of discomfort during the examined period 2001-2005 is concerned, the mean seasonal number of consecutive hours during the day with high levels of human discomfort appears in the station PAT; 11.3 and 13.6 consecutive hours with respect to DI and CP, respectively, against 1.0 and 2.7 consecutive hours at the station THR, respectively. All the other examined regions-stations within the GAA present a bioclimatic behavior

between PAT and THR. This means that for a given building within the city's center region (PAT), we need 5 to 11 times more energy for cooling during the warm period of the year than the energy for cooling at the north suburban area (THR).

10. Conclusions

In this study an application, which concerns the development and the use of ANN models on environmental issues and generally in environmental management, is presented. A number of ANN models have been developed and trained in order to forecast the air quality levels, as well as the bioclimatic conditions in different regions within the GAA. The findings of this work appoint the ANN models forecasting capacity.

The Results showed that the use of ANN models as forecasting tool is realizable and satisfactory at a statistically significant level of $p < 0.01$. In particular for the air quality forecasting for the next day, the R^2 values ranged between 0.381 and 0.826 (ERPI) and between 0.378 and 0.686 (DAQx). Besides, the IA index between the predicted and observed values ranged between 0.717 and 0.937 for ERPI forecasting, while it ranged between 0.746 and 0.889 for DAQx forecasting. It seems that in all cases, the air quality forecasting is more sufficient using the ERPI air quality index than the DAQx. In this point we have to mention that the ERPI is according to the European Community directives for the air quality levels. The same results are extracted regarding the forecasting of the persistence of air pollution episodes and especially the number of consecutive hours during the day with poor air quality.

Concerning the forecasting of bioclimatic conditions for the next day, the R^2 values ranged between 0.676 and 0.841 for DI and between 0.591 and 0.814 for CP. The IA values ranged between 0.849 and 0.956 for DI and between 0.813 and 0.948 for CP. Taking into account the persistence of the phenomenon (the number of consecutive hours during the day with high discomfort conditions due to strong heat stress), it seems that ANN#9 (consecutive discomfort hours according to DI values) and ANN#10 (consecutive discomfort hours according to CP values) give an adequate prediction.

A remarkable finding of this research is that the high values of IA (0.956 – 0.982) and R^2 (0.839 – 0.933) with respect to PET forecasting for the next three days indicate that the constructed ANNs have an excellent forecasting ability of PET, a more complex bioclimatic index based on the human energy balance. This gives evidence that the developed ANNs, taking into account simple meteorological parameters recorded in the previous three days, are capable to predict a bioclimatic index, which is not easily calculated (PET was estimated using the RayMan model).

11. References

- Antonic, O., Hatic, D., Krian, J. & Bukocev, D. (2001). Modeling groundwater regime acceptable for the forest survival after the building of the hydro-electric power plant. *Ecol. Model.*, Vol. 138, pp. 277-288.
- Attoh-Okine, N.O. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Adv. Eng. Softw.*, Vol. 30, No. 4, pp. 291-302.
- Balaguer Ballester, E., Valls, G., Carrasco-Rodriguez, J., Soria Oliva, E., Valle-Tascon, S. (2002). Effective 1-day ahead prediction of hourly surface ozone concentrations in

- eastern Spain using linear models and neural networks. *Ecol. Model.*, Vol. 156, pp. 27-41.
- Barazzetta, S., Corani, G. (2004). First results in the prediction of PM10 in Milan: the Air Sentinel project. *Proceedings of the 9th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Garmisch, May 2004.
- Besancenot, J.P. (1978). Le bioclimat humain de Rio. In: *Recherches de Climatologie en Milieu Tropical et Mediterranean*, Suchel, J.B., Altes, E., Besancenot, J.P. & Maheras, P. (Editors), Cahier No. 6 du Centre de Recherches de Climatologie, Universite de Dijon, Dijon, France.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, ISBN 019853864, Oxford, U.K.
- Caudill, M. & Butler, C. (1992). *Understanding Neural Networks; Computer Explorations*, MIT Press-Cambridge, ISBN 0262530996, MA, USA.
- Clarke, J.F., Bach, W. (1971). Comparison of the comfort conditions in different urban and suburban microenvironments. *Int. J. Biometeorol.*, Vol. 15, pp. 41-54.
- Comrie, A.C. (1997). Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manage.*, Vol. 47, pp. 653-663.
- Corani, G. (2005). Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. *Ecol. Model.*, Vol. 185, pp. 513-529.
- Council Directive 96/92/EC (1996). *On ambient air quality assessment and management*. Official Journal of the European Communities, L296, 21.11.1996, pp. 55-63.
- Council Directive 1999/30/EC (1999). *Limit values of sulphur dioxide, nitrogen dioxide and oxides of nitrogen, particulate matter and lead in ambient air*. Official Journal of the European Communities, L163, 29.6.1999, pp. 41-60.
- Directive 2000/69/EC of the European Parliament and the Council (2000). *Limit values for benzene and carbon monoxide in ambient air*. Official Journal of the European Communities, L313, 13.12.2000, pp. 12-22.
- Directive 2002/3/EC of the European Parliament and the Council (2002). *Ozone in ambient air*. Official Journal of the European Communities, L67, 9.3.2002, pp. 14-31.
- Dutot, A.L., Rynkiewicz, J., Steiner, F.E. & Rude, J. (2007). A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environ. Modell. & Softw.*, Vol. 22, No. 9, pp. 1261-1269.
- Gardner, M.W., Dorling, S.R. (1998a). Artificial Neural Networks, the multilayer perceptron. A review of applications in the atmospheric sciences. *Atmos. Environ.*, Vol. 32, pp. 2627-2636.
- Gardner, M.W., Dorling, S.R. (1998b). Neural network modeling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmos. Environ.*, Vol. 33, pp. 709-719.
- Giles, B.D., Balafoutis, C.H., Maheras, P. (1990). Too hot for comfort: the heatwaves in Greece in 1987 and 1988. *International Journal of Biometeorology*, Vol. 34, pp. 98-104.
- Givoni, B. (1998). *Climatic considerations in building and urban design*, John Wiley & Sons, ISBN 0471291773, New York.
- Hecht-Nielsen, R. (1990). *Neurocomputing*, Addison-Wesley, ISBN 0201093553, Reading, M.A.
- Heymans, J.J., Baird, A. (2000). A carbon flow model and network analysis of the northern Benguela upwelling system, Namibia. *Ecol. Model.*, Vol. 126, pp. 9-32.

- Höppe, P.R. (1999). The physiological equivalent temperature—a universal index for the biometeorological assessment of the thermal environment. *Int. J. Biometeorol.*, Vol. 43, pp. 71–75.
- Karul, C., Soyupak, S., Cilesiz, A.F., Akbay, N. & Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecol. Model.*, Vol. 164, pp. 145-152.
- Kolehmainen, M., Martikainen, H., Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. *Atmos. Environ.*, Vol. 35, pp. 815-825.
- Lapedes, A. & Farber, R. (1987). *Non-linear signal processing using neural networks*. Technical Report no. LA-UR-2662 Los Alamos National Laboratory.
- Makra, L., Mayer, H., Beczi, R. & Borsos, E. (2003). Evaluation of the air quality of Szeged with some assessment methods. *Acta Climatologica et Chorologica, Universitatis Szegediensis*, Tom. 36-37, pp. 85-93.
- Matzarakis, A., Mayer, H. (1996). *Another kind of environmental stress: Thermal stress*. WHO Collaborating Centre for Air Quality Management and Air Pollution Control, NEWSLETTERS No. 18, pp. 7-10.
- Matzarakis, A., Mayer, H., Iziomon, M. (1999). Applications of a universal thermal index: physiological equivalent temperature. *Int. J. Biometeorol.*, Vol. 43, pp. 76-84.
- Matzarakis, A., Rutz, F., Mayer, H. (2010). Modelling Radiation fluxes in simple and complex environments – Basics of the RayMan model. *Int. J. Biometeorol.*, Vol. 54, pp. 131-139.
- Mayer, H., Höppe, P. (1987). Thermal comfort of man in different urban environments. *Theor. Appl. Climatol.*, Vol. 38, pp. 3-49.
- Mayer, H., Kalberlah, F. & Ahrens, D. (2002a). TLQ-Am impact related air quality index obtained on a daily basis. *Proceedings of the fourth Symposium on the Urban Environment*, Norfolk, Virginia (USA), 20-24 May 2002, pp. 80-81.
- Mayer, H., Kalberlah, F., Ahrens, D. & Reuter, U. (2002b). Analysis of indices for the assessment of the air (in German). *Gefahrstoffe-Reinhaltung der Luft*, Vol. 62, pp. 177-183.
- McCulloch, W.S., Pitts W. (1943). A logical calculus of ideas immanent in Nervous activity. *Bulletin of Mathematical Biophysics*, Vol. 5, (1943), pp.115-133.
- Moustris, K.P. (2009). *Air quality forecasting with the use of Artificial Neural Networks in the greater Athens area*. PhD Thesis, School of Chemical Engineering, National Technical University of Athens, Athens, Greece.
- Moustris, K.P., Ziomas, I.C., Paliatsos, A.G. (2010). 3-Day-Ahead Forecasting of Regional Pollution Index for the Pollutants NO₂, CO, SO₂, and O₃ Using Artificial Neural Networks in Athens, Greece. *Water Air Soil Poll.*, Vol. 200, pp. 29–43.
- Moustris, K.P., Larissi, I.K., Nastos, P.T., Paliatsos, A.G. (2011). Precipitation forecast using artificial neural networks in specific regions of Greece. *Water Resour. Manag.*, DOI 10.1007/s11269-011-9790.
- Nastos, P.T., Matzarakis, A. (2006). Weather impacts on respiratory infections in Athens, Greece. *Int. J. Biometeorol.*, Vol. 50, pp. 358-369.
- Nastos, P.T., Matzarakis, A. (2008). Variability of tropical days over Greece within the second half of the twentieth century. *Theor. Appl. Climatol.*, Vol. 93, pp. 75-89.
- Matzarakis, A., Nastos, P.T. (2010). Human-Biometeorological assessment of heat waves in Athens. *Theor. Appl. Climatol.*, DOI 10.1007/s00704-010-0379-3.

- Norgaard, M., Ravn, O., Poulsen, N.K., Hansen, L.K. (2000). *Neural Networks for Modelling and Control of Dynamic Systems*, Springer-Verlag, ISBN 1852332271, London.
- Nunnari, G., Nucifora, M., Randieri, C. (1998). The application of neural techniques to the modeling of time-series of atmospheric pollution data. *Ecol. Model.*, Vol. 111, pp. 187-205.
- Papanastasiou, D.K., Melas, D., Kioutsioukis, I. (2007). Development and Assessment of Neural Network and Multiple Regression Models in Order to Predict PM10 Levels in a Medium-sized Mediterranean City. *Water Air Soil Poll.*, Vol. 182, pp. 325-334.
- Philandras, C.M., Metaxas, D.A., Nastos, P.T. (1999). Climate variability and urbanization in Athens. *Theor. Appl. Climatol.*, Vol. 63, No. 1-2, pp. 65-72.
- Prybutok, R., Junsub, Y., Mitchell, D. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Huston's daily maximum ozone concentrations. *Eur. J. Oper. Res.*, Vol. 122, pp. 31-40.
- Refenes, A.N., Zapranis, A. & Francis, G. (1994). Stock performance modeling using neural networks: A comparative study with regression models. *Neural Networks*, Vol. 7, No. 2, pp. 375-388.
- Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., Doyle, M. (2003). A rigorous inter-comparison of ground-level ozone predictions. *Atmos. Environ.*, Vol. 37, pp. 3237-3253.
- Siple, P.A., Passel, C.F. (1945). Measurements of dry atmospheric cooling in subfreezing temperatures. *Proceedings of the American Philosophical Society*, Vol. 89, No. 1, pp. 177-199.
- Slini, T., Kaprara, A., Karatzas, K. & Moussiopoulos, N. (2006). PM10 Forecasting for Thessaloniki, Greece. *Environ. Model. Softw.*, Vol. 21, No. 4, pp. 559-565.
- Thom, E.C. (1959). The discomfort index. *Weatherwise*, Vol. 12, pp. 57-60.
- VDI (1998). VDI 3787, Part I: Environmental Meteorology, Methods for the human biometeorological evaluation of climate and air quality for the urban and regional planning at regional level. Part I: Climate, Beuth, Berlin.
- Viotti, P., Liuti, G. & Di Genova, P. (2002). Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. *Ecol. Model.*, Vol. 148, No. 1, pp. 27-46.
- Werbos, P. (1988). Generalization of Backpropagation with application to a recurrent gas market model. *Neural Networks*, Vol. 1, pp. 339-356.
- Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J. & Rowe, C. (1985). Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, Vol. 90, pp. 8995-9005.



Advanced Air Pollution

Edited by Dr. Farhad Nejadkoorki

ISBN 978-953-307-511-2

Hard cover, 584 pages

Publisher InTech

Published online 17, August, 2011

Published in print edition August, 2011

Leading air quality professionals describe different aspects of air pollution. The book presents information on four broad areas of interest in the air pollution field; the air pollution monitoring; air quality modeling; the GIS techniques to manage air quality; the new approaches to manage air quality. This book fulfills the need on the latest concepts of air pollution science and provides comprehensive information on all relevant components relating to air pollution issues in urban areas and industries. The book is suitable for a variety of scientists who wish to follow application of the theory in practice in air pollution. Known for its broad case studies, the book emphasizes an insightful of the connection between sources and control of air pollution, rather than being a simple manual on the subject.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Panagiotis Nastos, Konstantinos Moustris, Ioanna Larissi and Athanasios Paliatsos (2011). Air Quality and Bioclimatic Conditions within the Greater Athens Area, Greece - Development and Applications of Artificial Neural Networks, Advanced Air Pollution, Dr. Farhad Nejadkoorki (Ed.), ISBN: 978-953-307-511-2, InTech, Available from: <http://www.intechopen.com/books/advanced-air-pollution/air-quality-and-bioclimatic-conditions-within-the-greater-athens-area-greece-development-and-applica>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen