Hybrid Soft Computing for Atmospheric Pollution-Climate Change Data Mining

Lazaros Iliadis[1], Vardis-Dimitris Anezakis[2](✉), Konstantinos Demertzis[1], and Stefanos Spartalis[3]

[1] School of Engineering, Department of Civil Engineering, Democritus University of Thrace, University Campus, Kimmeria, 67100 Xanthi, Greece
{liliadis, kdemertz}@fmenr.duth.gr

[2] Department of Forestry and Management of the Environment and Natural Resources, Democritus University of Thrace, 193 Pandazidou st., 68200 N Orestiada, Greece
danezaki@fmenr.duth.gr

[3] School of Engineering, Department of Production and Management Engineering, Democritus University of Thrace, V. Sofias 12, 67100 Xanthi, Greece
sspart@pme.duth.gr

Abstract. Prolonged climate change contributes to an increase in the local concentrations of O_3 and PMx in the atmosphere, influencing the seasonality and duration of air pollution incidents. Air pollution in modern urban centers such as Athens has a significant impact on human activities such as industry and transport. During recent years the economic crisis has led to the burning of timber products for domestic heating, which adds to the burden of the atmosphere with dangerous pollutants. In addition, the topography of an area in conjunction with the recording of meteorological conditions conducive to atmospheric pollution, act as catalytic factors in increasing the concentrations of primary or secondary pollutants. This paper introduces an innovative hybrid system of predicting air pollutant values (IHAP) using Soft computing techniques. Specifically, Self-Organizing Maps are used to extract hidden knowledge in the raw data of atmospheric recordings and Fuzzy Cognitive Maps are employed to study the conditions and to analyze the factors associated with the problem. The system also forecasts future air pollutant values and their risk level for the urban environment, based on the temperature and rainfall variation as derived from sixteen CMIP5 climate models for the period 2020–2099.

Keywords: Self Organizing Maps · Fuzzy Cognitive Maps · Air pollutants
Climate change models of CMIP5 · Soft computing techniques
Athens

N. T. Nguyen and R. Kowalczyk (Eds.): TCCI XXX, LNCS 11120, pp. 152–177, 2018.
https://doi.org/10.1007/978-3-319-99810-7_8

1 Introduction

This is an extended version of the paper "Fuzzy cognitive maps for long-term prognosis of the evolution of atmospheric pollution, based on climate change scenarios: The case of Athens" that was presented in the 8th International Conference on Computational Collective Intelligence (ICCCI 2016).

In our previous research [2] we have used the available dataset in order to forecast atmospheric pollution from the fluctuation of temperature and rainfall till 2100. This was achieved by employing Fuzzy Cognitive Maps and the GFDL_CM 2.0 climate model. The new elements of this research are the following:

The data have been clustered by using Self Organizing Maps in order to estimate the values of the meteorological features that have significant contribution to the atmospheric conditions. The knowledge obtained by the SOM Clustering was used as input in Fuzzy Cognitive Maps in order to produce forecasts on air pollution. Moreover in this research we have used sixteen Climate Models and four Climate Scenarios from the CMIP5 program for the forecast of air pollution based on the fluctuation of temperature and rainfall till 2100. The use of different climate models adds more merit and credibility to our proposed model.

Finally, steps 1, 2 and 6 of the proposed algorithm are new elements of the extended paper.

The increase of human population and production processes over the years, has caused severe negative impacts in the environment, which is often responsible for harmful effects on the health of living organisms and ecosystems. Atmospheric pollution results in alteration of the structure composition and characteristics of the atmosphere. Perhaps it is the most typical example of the environmental burden caused by human activity. Pollutants are either emitted directly into the atmosphere from activities such as energy production by solid or liquid fuel, transport, industry and heating and they are known as primary (e.g. CO, NO, NO_2, SO_2) or they are formed therein when appropriate meteorological or photochemical conditions exist (secondary ones) e.g. O_3.

Under the influence of certain meteorological features, concentrations of primary or secondary pollutants are likely to reach critical alarm values that create inappropriate living conditions, increasing the incidence of cardiovascular and respiratory diseases. It is usually necessary to carry out an in-depth spatiotemporal analysis of the combined effect of meteorological conditions that create and maximize the problem, as well as to search for the interrelationships between the pollutants and the meteorological parameters. More specifically, temperature, humidity and sunshine are determinants in the composition of the smog and photochemical clouds, as opposed to rainfall and wind that contribute to the diffusion and dispersion of pollutants. Smog cloud is formed when there is high concentration of pollutants in the atmosphere, such as CO, SO_2, PM_{10}, $PM_{2.5}$, combined with relatively low temperature high humidity and a few hours of sunshine. The phenomenon is more pronounced during the winter months and especially in the morning hours. In contrast, photochemical cloud occurs under high temperatures, long duration-intensity of sunshine, combined with low relative humidity, high concentration of nitrogen oxides, hydrocarbons and secondary products.

The chemical composition and nature of the pollutants has been shown to cause respiratory problems and can be associated with various forms of cancer. However due to their large heterogeneity and mechanisms by which they enter or leave the atmosphere, it is difficult to calculate their exact contributions to the environment and therefore to calculate their exact effects on human health.

Athens is a characteristic European capital combining Mediterranean climate and low air quality. The increasing air pollution is caused by factors such as poor urban planning, urbanization, topography, climate change and recent economic crisis. A pilot intelligent data mining research of Athens problem, would serve also as a model for other similar cities of the European South.

Technological evolution is offering a solid background for modeling such phenomena, based on sophisticated systems of artificial intelligence and Machine Learning. This research introduces an innovative approach, predicting the atmospheric pollution values with Soft-Computing, based on the most up-to-date specialized and valid scenarios of Climate Change (CC). The system uses Self Organizing Maps (SOM) unsupervised learning, for the grouping of primary environmental data related to the quality of the city's atmosphere. The knowledge extracted by the intelligent cluster analysis is employed to create the conditions for the implementation of individual Fuzzy Cognitive Maps (FCM) one for each cluster obtained by the SOM. This methodology allows the study of the conditions related to atmospheric pollution, the assessment of the problem and its correlation with specialized, realistic situations that eventually lead to emergencies.

The proposed FCM are based on linear correlations rather than the expert's opinion, which allows the modeling of multifactorial systems without any particular constraints. As a result, future air pollutant values and their degree of risk for the urban environment of Athens can be predicted based on the temperature and rainfall variation as derived from sixteen CMIP5 climate models related to the period 2020–2099.

1.1 Literature Review

Significant research efforts have been made towards modeling air quality in medium sized cities or major urban centers like Athens. However, this has been done using statistical methods, without taking into serious consideration the effect of the temperature and precipitation fluctuation, due to the 21st century climate change [29].

Important computational research efforts have been carried out recently towards air pollution modeling. Iliadis et al. [16] have developed Feed Forward Neural Networks in order to model O_3 concentrations in Athens. However, only recently some researchers have used SOM clustering to analyze the problem in major cities. Patterns of air quality have been searched for Mexico City by Neme and Hernández [25], whereas Karatzas and Voukantsis have done the same for the Thessaloniki [18]. Glorennec also applied SOM to forecast O_3 peaks [12]. Li and Chou have investigated spatial variation of air pollution with SOM [21]. Jiang et al. [17] explored the classification and visualization utility of the SOM in the context of New South Wales Australia, using gridded geopotential height reanalysis together with multi-site meteorological and air quality data for Sydney. Pearce et al. [32] applied SOM for Atlanta

Georgia, based on daily air quality data related to ten features, aiming to define the profile of pollutant related day types.

Climate change is the most important environmental risk globally. Our team has already modeled complex systems related to climate change and its direct impacts including increased air pollutants concentrations in the atmosphere [2–9, 16]. In an earlier research [8] we have managed to get a clear and comprehensive view of air quality in the wider urban center of Athens and in the wider Attica basin. This study was based on data selected from nine air pollution measuring stations of the area during the temporal periods (2000–2004, 2005–2008 and 2009–2012). This method was based on the development of 117 partial ANN whose performance was averaged by using ensemble learning approaches. The system used fuzzy logic in order to forecast more efficiently the concentration of each pollutant. A previous research of our team [9], has used SOM performing clustering of the pollutants. The goal, was to find the most isolated outliers' group comprising of extreme values. This specific group has proven to be the most important, as it contains vital information about the hazardous pollutants and the meteorological and temporal conditions under which extreme values occur. The EHF forecasting system [4], allows the prediction of extreme air pollutant values. Its main advantage is that though it does not need to consider raw values as input, it manages to operate quite efficiently. The following unsupervised learning algorithms were employed in this case: SOM, Neural Gas and Fuzzy c-Means clustering. Bougoudis et al. [6] proposed a novel and flexible hybrid machine learning (ML) system that combines Semi-Supervised Classification and Semi-Supervised Clustering, in order to predict air pollutants' outliers and to study the conditions that favor their high concentration. Bougoudis et al. [5] proposed an innovative hybrid system of combined Machine Learning algorithms. They introduced an ensemble capable of forecasting the values of air pollutants. This approach improved the accuracy of existing forecasting models by clustering the data vectors and tracing hidden knowledge.

Gordaliza et al. [13] developed coherent storylines about ordinary people living under diverse scenarios of low/high CO_2. Luiz et al. [22] developed FCM aiming to understand the viability of Clean Development Mechanism (CDM) projects in South Africa and how they would influence greenhouse gas (GHG) emissions. Zhang et al. [38] explored the application of FCM on getting stakeholders' perspectives and they employed graph theory indices on quantifying them. Pathinathan and Ponnivalavan [30] analyzed the hazards of plastic pollution using Induced IFCM. IFCMs are a fuzzy-graph modeling approach based on expert's opinion. Amer et al. [1] developed three future scenarios using FCM for the national wind energy sector of a developing country. Mesa-Frias et al. [23] developed a novel method based also on FCM to quantify the framing assumptions in the assessment of health impact. Fons et al. [10] proposed a conceptual model of an eco-industrial park and employed FCM to analyze the impacts of this model in terms of pollution and waste disposal. Papageorgiou and Salmeron [27, 28] surveyed the methods and learning algorithms of FCMs applied to modeling and decision-making tasks.

Mourhir et al. [24] presented the rational and design of a methodology to support Integrated Environmental Assessment using the DPSIR (Driving Forces-Pressures-State-Impact-Response) causal-effect framework and non-monotonic FCM. The methodology was based on key pillars in environmental management, namely connecting the socioeconomic and the natural environment dimensions into a policy oriented context. Paz-Ortiz and García [31] constructed a collective FCM for the qualitative simulation of the earth climate system. The map was developed by considering the subsystems on which the climate equilibrium depends, and by aggregating different expert's opinions over this framework. The linguistic variables were used to fuzzify the edges. Fuzzy aggregation T-Norms were used to produce overall linguistic weights. The obtained outcome was defuzzified using the center of gravity technique and the current state of the earth climate system was simulated and discussed. Finally, a nonlinear Hebbian learning algorithm was used for updating the edges of the map until a desired state was reached. García and Ortiz [11] developed a Soft Computing model for the qualitative analysis of the earth's climate system dynamics throughout the implementation of FCM. For this purpose, they have identified the subsystems in terms of which the dynamics of the whole system can be described. Then, with these concepts they have built a FCM via the study of the documented relations among these concepts. Once the map was built, they used the technique of state vector and the adjacent matrix to find the hidden patterns, i.e. the feedback processes among system's nodes. Later on, they explored the sensitivity of the model to changes in the weights, and to changes in the input data values. Finally, they used fuzzy edges to analyze the causality flux among concepts and to explore possible solutions applied in specific edges.

Tamas et al. [36] have developed air quality forecasting models using ML methods applied to hourly concentrations of O_3, NO_2 and PM_{10} 24 h ahead. Multilayer Perceptrons (MLP) were hybridized with hierarchical clustering and with a combination of SOM and k-means clustering (KMC). Clustering methods were used to subdivide the dataset, and then a MLP was trained on each subset. Khedairia and Khadir [19] proposed clustering performed in two stages based on a combination of SOM and KMC algorithms. The obtained five meteorological clusters were used to better elucidate the dependency of meteorology on air quality in the presence of seven measured pollutants. Olej and Hájek [26] have presented a model for air quality classification of districts into classes according to their pollution, based on Kohonen's SOM and on intuitionistic fuzzy sets. Hájek and Olej [14] designed a model which combines SOM, KMC and classification performed by Learning Vector Quantization neural networks. Therefore, the model generates well-separated clusters and has good generalization ability as well.

1.2 Motivation and Innovative Elements of This Research

Looking at the literature more thoroughly and analytically, we have reached the following conclusions: The SOM and KMC algorithms have been used in many surveys, to classify or cluster raw meteorological and atmospheric data. Fuzzy Cognitive Maps (FCM) and Artificial Neural Networks (ANN) have been employed to construct hybrid forecasting models, based on meteorological parameters. However, there is lack of a holistic approach to address the environmental problem. To the best of our knowledge FCM combined with SOM have not been used in the literature towards air pollutants' clustering and forecasting based on meteorological data and on Climate change models.

Clustering of data in this research, helps to determine the values of the meteorological features that affect the atmospheric conditions in Athens and to isolate and identify potential future extreme values. The features of each cluster are interconnected with synapses in a FCM towards air pollution concentrations' forecasting. This forecasting is based on temperature and rainfall fluctuations, using 16 Climate Models of the CMIP5 program for the period 2020–2099.

This in-depth study, strengthens civil protection mechanisms by serving as a tool for informing public bodies (Ministry of the Environment, Regions, Municipalities) on the potential days of adverse meteorological conditions favoring increased concentrations of pollutants in the atmosphere. An important target of the present study is to inform hospitals promptly to properly manage the potential increased import of patients, by providing the required number of beds and medical supplies plus specialized medical and nursing staff.

The innovation of this research is based on the development of an information system that uses a hybrid Soft Computing model to classify and predict atmospheric data in Athens. Additionally, projection into the future by determining changes in the concentrations of air pollutants based on Climate change models, can lead to preemptive measures-actions towards citizens' protection. It can also contribute to the design of long term improvement of offered health services.

A major innovation introduced herein is FCM modeling which is based solely on measurable factors, arising from the correlation analysis between raw values of the involved parameters and not on the subjective experts' opinion that is the typical practice. This adds credibility and reliability to the overall system.

In addition, innovation appears in the use of unsupervised learning to isolate the representative values of the parameters under consideration that enables forecasting of future pollutants' concentrations.

1.3 Data

The data come from the "Patissia" area in the center of Athens. Relativity FCM analysis was performed for this station, for the period 2000 to 2012, aiming to obtain a symbolic representation of existing complex correlations between the pollutants and other features. The above station is used to store hourly values of CO (in mg/m^3) plus (NO, NO_2, O_3 and SO_2 in $\mu g/m^3$) [5]. It is well known for its consistency and reliability as it has the fewest missing values in the Attica basin [5]. The topography of Athens prevents the diffusion of pollutants. "Patissia" is a heavy traffic area and this is the reason of its high air pollution concentrations. Additionally, records related to the following meteorological factors namely: air temperature (AT), relative humidity (RH), air pressure (PR), solar radiation (SR), wind speed (WS) wind direction (WD) and atmospheric precipitation (AP) were obtained from the station of "Thiseion" near the Acropolis 9 km away from the sea.

During data pre-processing all records with missing values for one or more parameters were removed. Outliers are very important as they are always considered for the activation of the civil protection mechanisms. For this reason, they were not removed from the datasets in order to obtain representative training samples offering potential generalization in future forecasting models. Finally in order to tackle the problem of features with different range, in which the higher values most affect the cost function with respect to the characteristics of the smaller ones without being more important, a normalization of the data vectors was performed in the interval [−1, +1] by employing the following Eq. 1:

$$Y_i = X_i - \frac{\bar{X}}{\sigma} \quad (1)$$

where $\bar{X}$ is the average value of each feature and σ is the standard deviation.

2 Theoretical Background

2.1 Self-organizing Maps

Self-Organizing Maps is an algorithm of competitive learning developed by Kohonen [20]. A SOM includes the Input layer and the layer with the competing neurons which are organized in a 2-D lattice like in Fig. 1. Each one of them is characterized by a weight vector $W_i = (w_{il}, \ldots, w_{id})^T$. When an input vector $X_i = (x_1, \ldots, x_d)^T x \in R$ is introduced, the lattice neurons complete each other and the winning neuron (WINE) m is obtained. Its vector W_m appears to have the highest similarity with vector X_i. Thus SOM depict an input X_i of dimension d, at the coordinates of the grid $R_m = (z_{m1}, z_{m2})^T$ [15, 20].

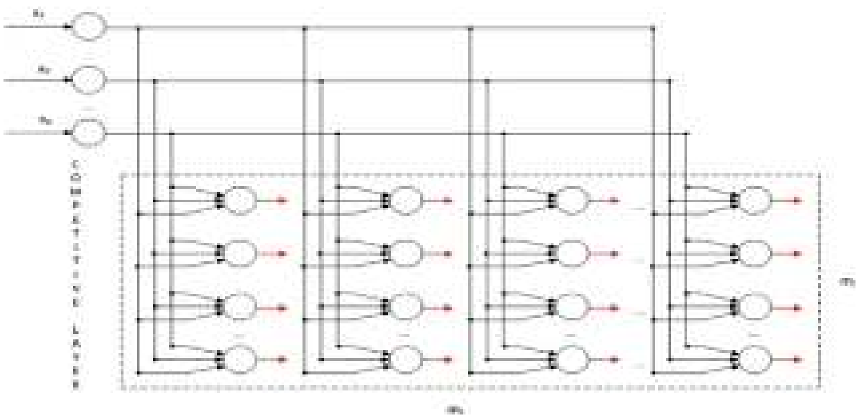

Fig. 1. Self-Organizing Map with d inputs and 2-dimensions lattice m1 × m2

In order to group the data, a self-organization map is formed, initializing the weights $W_i = (w_{il}, \ldots, w_{id})^T$ with small values randomly produced by a random number generator function. The next algorithmic steps follow:

- ***Competition.*** For each training sample X_n the lattice neurons estimate the respective value of the similarity function using the Euclidean distance between the input vector $X_i = (x_1, \ldots, x_d)^T x \in R$ and the weight vector $W_i = (w_{il}, \ldots, w_{id})^T$ of the competing neurons. The neuron with the highest similarity is the winner.

- ***Cooperation.*** The WINE *i* delimits its topological neighborhood $h_{j,i}$ (around the surrounding neurons) which is a function of $d_{j,i}$ which is the distance between the winning neuron i and neuron j. The neighborhood is symmetric to the WINE. The following Gaussian function 2 is used to perform the clustering:

$$h_{j,i}(x) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right) \tag{2}$$

where σ is the effective width of the topological neighborhood which determines the extent to which the neurons in the neighborhood of the winner are involved in the process. This parameter is reduced exponentially in every epoch n, based on the following function 3.

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), \ n = 0, 1, 2, \ldots \tag{3}$$

The parameter σ_0 is the initial value of the active width and the value of τ_1 is given by the following Eq. 4:

$$\tau_1 = \frac{n_0}{\ln(\sigma_0)} \tag{4}$$

Considering a 2-D lattice, we have assigned the value of its radius to the initial value of the active width σ_0 whereas

$$\tau_1 = \frac{1000}{\log \sigma_0} \tag{5}$$

- ***Synaptic Adaption.*** At this last stage of the training process, we have been updating the weights of neurons on the competitive level. The change metric is given by Eq. 6:

$$\Delta w_j = \eta h_{j,i(x)}(x - w_j) \tag{6}$$

The index *i* is used to denote the winner and *j* is a neuron in its neighborhood. Given the weight vector $W_j(n)$ for a certain point in time *n*, we estimated the new vector for the time stamp $n + 1$ from the following function 7:

$$w_j(n+1) = w_j(n) + \eta(n) h_{j,i(x)}(n)(x(n) - w_j(n)) \tag{7}$$

From the above relationship it follows that the learning rate n starts from a value around 0.1 and decreases gradually to the value of 0.01. These values were achieved according to the function 8 below:

$$\eta(\mathrm{n}) = \eta_0 \exp\left(-\frac{\mathrm{n}}{\tau_2}\right)(4)\ \mathrm{n} = 0, 1, 2, \ldots,\ \text{with}\ \eta_0 = 0.1\ \text{and}\ \tau_2 = 1000 \tag{8}$$

It should be stated that τ_2 is given by Eq. 9 accordingly:

$$\tau_2 = \frac{\mathrm{n}_0}{\ln(100 \cdot \eta_0)} \tag{9}$$

where n_0 is the number of iterations of the layout phase, η_0 is the initial learning rate and σ_0 is given by the following Eq. 10.

$$\sigma_0 = \sqrt{w^2 + h^2} \tag{10}$$

It should be clarified that w and h are the length and the height of the 2-D lattice respectively.

2.2 Fuzzy Cognitive Maps

Unlike the majority of complex dynamic systems characterized by nonlinearity and great uncertainty, vague cognitive maps utilize advanced learning techniques to select appropriate weights for causal links between the variables examined. This is done in order to capture the problem with absolute realism. By combining the theoretical background of fuzzy logic, they cover the needs of comparing and characterizing reference sets, solving and modeling complex problems for which there is no exact mathematical model. In the fuzzy cognitive map model, the nodes are interconnected with edges. The direction of the edge implies which node affects which, with the sign of the causality relationship being positive if there is a direct relationship of influence, negative if there is an inverse relation of influence and zero if the two nodes are uncorrelated. Cases of causality that are typically determined by experts are described using fuzzy Linguistic variables and they are fuzzified by using fuzzy membership functions in the real interval [−1, 1] [27, 28, 34, 37].

2.3 Correlation Analysis

In order to test the existence of a linear relationship between meteorological parameters and atmospheric pollutants, the relativity analysis was used with the Pearson (r) parametric correlation coefficient method.

The Pearson linear correlation coefficient of two variables X and Y is defined on the basis of a sample of n pairs of observations (x_i, y_i) $i = 1, 2, \ldots, n$ and it is denoted as $r(X, Y)$ or more simply as r. It is estimated by the following Eqs. 11–15:

$$r = \frac{s_{xy}}{s_x s_y} \tag{11}$$

$$S_{xy} = Cov(X, Y) = \frac{\sum_{i=1}^{v}(x_i - \bar{x})(y_i - \bar{y})}{v - 1} = \frac{\sum_{i=1}^{v} x_i y_i - v\bar{x}\bar{y}}{v - 1} \tag{12}$$

$$s_x = \sqrt{\frac{1}{v - 1}\sum\nolimits_{i=1}^{v} (x_i - \bar{x})^2} \tag{13}$$

$$s_y = \sqrt{\frac{1}{v - 1}\sum\nolimits_{i=1}^{v} (y_i - \bar{y})^2} \tag{14}$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{v}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{v}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{v}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{v} x_i y_i - v\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{v} x_i^2 - v\bar{x}^2}\sqrt{\sum_{i=1}^{v} y_i^2 - v\bar{y}^2}} \tag{15}$$

The correlation coefficient is a pure number and it always takes real values in the interval [−1, 1]. More specifically, when $0 < r \leq 1$, X, Υ are positively linearly correlated whereas when $-1 < r < 0$, then X, Υ are negatively correlated. If $r = 0$, there is no linear correlation [33].

2.4 Climate Change Scenarios

Climate change is the most crucial environmental problem of the century. Our team has already modeled complex systems related to climate change and its direct impacts including increased air pollutants concentrations in the atmosphere [2–9]. The Intergovernmental committee named "*Committee International Panel on Climate Change*" (IPCC) which deals with the assessment of climate change is an international scientific body which until today has published five reports. The aim of the program (CMIP5) that was defined in the 5th Assessment Report on Climate Change (IPCC-AR5, Assessment Report 5) was the design of climate models, aiming to estimate future climatic changes both in the short and in the long range. This objective is achieved by using Earth System Models (ESM) and global climate ocean-atmosphere coupling models "Atmospheric-Ocean General Circulation Models" (AOGCMs).

The latest report (AR5) finds significant improvement in the models to analyze mechanisms of temperature and precipitation, in the study of anthropogenic impact on the environment and in the study of the biochemical cycles. According to the report, four future scenarios of *Greenhouse Gases* (GHGs) concentration in the atmosphere have been developed. These scenarios are known in the literature under the RCPs acronym "*Representative Concentration Pathways*".

Based on the RCP2.6 scenario, a small increase in the emissions of greenhouse gases till the mid of the decade, would result in an increase of the solar radiation (SR) as high as 3 W/m^2 by 2050 and then in a decrease to the level of 2.6 W/m^2 by 2100.

According to the scenarios RCP4.5 and RCP6.0, a moderate increase in the greenhouse gases emissions, would result in the increase of SR. In RCP4.5 the solar radiation values stabilize at about 4.5 W/m^2 before 2100 and respectively in the RCP6.0 they are stabilized at 6.0 W/m^2 after 2100.

Finally, in the most extreme scenario (RCP8.5) with rapid and continuous increase of the GHGs the SR rises as high as 8.5 W/m^2 and continues to rise after 2100. The climate models *bcc_csm1_1, bcc_ csm1_1_m, ccsm4, cesm1_cam5, csiro_mk3_6_0, fio_esm, gfdl_cm3, gfdl_esm2m, giss_e2_h, giss_e2_r, ipsl_cm5a_mr, miroc_esm, miroc_esm_chem, miroc5, mri_cgcm3, noresm1_m* of the CMIP5 project were employed in this research, as the most modern and reliable for finding changes in temperature and precipitation for the time period 2020–2099 [35].

3 Description of the Proposed Model

The basic methodology of grouping the atmospheric pollution and the meteorological parameters as well as predicting the risk of air quality includes 5 distinct stages (Clustering, Modeling, Grid, Scenarios, and Forecasting) with individual algorithmic steps as shown in the following flowchart. The following chapter presents the analysis of the distinct algorithmic steps.

Step 1 (Clustering): In the first stage, the unsupervised learning algorithm SOM was used to group meteorological data and pollutants into clusters. In this way, similar data with common features is grouped together. The process resulted in 4 clusters with 39% of the data assigned to cluster0, 20% to cluster1, 14% to cluster2 and the remaining 27% to the fourth cluster Cluster3. Statistical analysis and presentation of the threshold values of the involved parameters for each cluster are presented in Table 1 below.

Table 1. Statistical analysis of the clusters

Parameters	Values	Cluster 0	Cluster 1	Cluster 2	Cluster 3
CO	Min	0.1	0.1	0.1	0.1
CO	Max	11.5	7.8	5.3	8.3
CO	Std. dev	0.8059	0.5511	0.4888	0.7152
NO	Min	1	1	1	1
NO	Max	447	393	312	382
NO	Std. dev	39.2074	27.6837	24.6888	31.4461
NO2	Min	1	1	1	1
NO2	Max	173	289	287	353
NO_2	Std. dev	24.0349	33.2698	28.4048	26.3124
O_3	Min	1	1	1	1

(*continued*)

Table 1. (*continued*)

Parameters	Values	Cluster 0	Cluster 1	Cluster 2	Cluster 3
O_3	Max	183	246	284	175
O_3	Std. dev	31.7274	40.2581	39.3866	32.729
SO_2	Min	2	2	2	2
SO_2	Max	213	272	227	220
SO_2	Std. dev	12.6625	20.364	14.1864	12.8126
AT	Min	−2	−2.7	−3.3	−1.21
AT	Max	35.3	41.1	41.33	36.3
AT	Std. dev	6.7545	6.9845	7.5879	6.1947
RH	Min	14	13	8	18
RH	Max	99	96	94	99
RH	Std. dev	14.1309	14.194	14.8345	13.5708
PR	Min	984.71	983.1	986.05	982.1
PR	Max	1025.64	1021.38	1025.51	1021.64
PR	Std. dev	6.1807	5.3403	5.317	5.687
SR	Min	0	0	0	0
SR	Max	551	1040.15	1041.85	565.4
SR	Std. dev	61.5099	252.7237	270.8435	76.3638
SUN	Min	0	0	0	0
SUN	Max	1	1.65	1.1	0.72
SUN	Std. dev	0.1148	0.1921	0.2099	0.0896
WS	Min	0	0.4	0	0.4
WS	Max	17.7	14.6	15.7	16.8
WS	Std. dev	2.5324	1.824	2.7661	1.8201
WD	Min	0	5	0	6
WD	Max	8	16	11	16
WD	Std. dev	1.5179	11.2529	2.8832	11.4951
AP	Min	0	0	0	0
AP	Max	52.3	3	1.5	39.8
AP	Std. dev	0.8784	0.037	0.0225	0.806

In our survey, the inputs to the network are meteorological and pollution data related to the period 2000–2012. This data were divided into four clusters in which the observations within each cluster are homogeneous (they have great similarity and short distance to each other) while the observations of the different clusters differ significantly. Each cluster contains values of different risk level (low, medium, high, and extreme) regarding the involved parameters as shown in Table 1. The above process leads to the choice of the cluster that contains the highest percentage of values forming the proper meteorological and atmospheric conditions that are the most typical for decision making regarding the risk in the center of Athens.

Step 2: Selection of the most representative cluster that will be the most descriptive of the state of meteorological conditions and atmospheric pollution for the period 2000–2012.

From the statistical analysis performed it was observed that clusters 1 and 2 have very low atmospheric precipitation values. For example, the highest atmospheric precipitation for cluster 1 equals to 3 mm. In addition, clusters 1 and 2 do not have a representative sample of CO, NO concentrations and they have many extreme values related to temperature, O_3, solar radiation, and sunshine. The precipitation values contained therein indicate the existence of drought as well as inverse relationship to the extreme values of the above parameters. In addition, the average values of O_3, temperature, sunlight and sunshine, exceeded the average of the 2000–2012 by far. This means that these clusters have accumulated extreme values, making them unable to realistically reflect the typical atmospheric conditions of Athens. Thus, they are unsuitable to be selected as the representative ones for the assessment of future pollutants concentrations evolution due to climate change that will cause a serious temperature increase and atmospheric precipitation decrease.

As far as clusters 0 and 3 are concerned, they account for the largest share of all data (66%) for the period 2000–2012. In order to select the most suitable cluster, it was particularly important that its values should contain a large percentage of the extreme atmospheric conditions. Also, the average values included should not be significantly different from the average of the period 2000–2012.

That is, the values of the selected cluster must not be limited to extreme climatic conditions but they should contain a representative range that will realistically reflect the overall climate and the concentrations of pollutants in the area.

Based on Table 2, the information that helped us select cluster3 as the most representative is that the average temperature is just 1.5 °C different than the average temperature values of the period 2000–2012, as opposed to cluster0 that is differentiated by 3.5 °C. The cluster3 therefore includes a wider range of temperature differences that significantly affect the values of the remaining related features. On the other hand, both clusters contain recordings with mild and moderate precipitation intensity as well as high precipitation values. Therefore, apart from the average, the deviation (range and standard deviation) of temperature and atmospheric precipitation was particularly important for the selection of the appropriate cluster (Tables 1 and 2). Climate simulation by the use of climate models and scenarios is entirely based on changes in temperature and atmospheric precipitation, which are considered as determinants of the influence of atmospheric conditions.

Table 2. Average of each parameter in each cluster

	CO	NO	NO_2	O_3	SO_2	AT	RH	PR	SR	SUN	WS	WD	AP
2000–2012	0.73	16.2	40.7	55.2	12.2	18.8	63.2	1003.5	189.3	0.33	3.3	6.7	464.7
Cluster0	0.85	19.5	39.7	40	10.9	15.33	70.3	1005.6	23.3	0.03	2.9	2.43	699.4
Cluster1	0.62	15	45.1	75.3	17.5	23.28	52.8	1002.1	494.6	0.90	3.5	11.3	1.85
Cluster2	0.51	11	30.7	80.7	11.7	24.8	46.7	1003.8	504.2	0.89	4.7	2.9	0.96
Cluster3	0.77	15.1	44	48.3	10.4	17.38	69.6	1001.4	31.2	0.03	2.6	11.5	711.2

In the second stage **(Modeling)**, all associated parameters of the selected cluster, which were interconnected with synapses, were added and named in order to create the causative negative or positive correlations.

The correlations fuzzification that is the description of any interconnection with Linguistic terms, was made by selecting six fuzzy sets: three positives (low positive, middle positive (+), high positive (+++)) and three negatives, Negative (−), middle negative (−−), high negative (−−−)), representing the fuzzy weights (Table 3).

Table 3. Effect and Value of six Linguistics which correspond to fuzzy weights

Effect	Value
High positive (+++)	1
Middle positive (++)	0.5
Low positive (+)	0.25
Low negative (−)	−0.25
Middle negative (−−)	−0.5
High negative (−−−)	−1

Step 3: Application of the Correlation Analysis method to calculate the potential linear correlation between the variables examined, namely: Carbon Monoxide (CO), Nitrogen Monoxide (NO), Nitrogen Dioxide (NO_2), Ozone (O_3), Sulfur Dioxide (SO_2) Air Temperature (AT), Humidity (RH), Atmospheric Pressure (PR), Solar Radiation (SR), Sunshine (SUN), Wind Speed (WS), Wind Direction (WD) and Atmospheric Precipitation (AP) (see Table 4).

Table 4. Correlation values of the involved features

	CO	NO	NO_2	O_3	SO_2	AT	RH	PR	SR	SUN	WS	WD	AP
CO	1	0.87	0.63	−0.59	0.36	−0.19	0.22	0.2	0	−0.01	−0.21	−0.16	−0.02
NO	0.87	1	0.48	−0.48	0.44	−0.21	0.21	0.17	0.08	0.04	−0.15	−0.09	−0.02
NO_2	0.63	0.48	1	−0.62	0.34	−0.03	0.16	0.23	0.03	0.02	−0.31	−0.22	−0.03
O_3	−0.59	−0.48	−0.62	1	−0.24	0.4	−0.47	−0.2	0.01	0.02	0.21	0.09	−0.02
SO_2	0.36	0.44	0.34	−0.24	1	−0.26	−0.07	0.19	0.14	0.08	−0.06	0.04	−0.04
AT	−0.19	−0.21	−0.03	0.4	−0.26	1	−0.39	−0.26	−0.02	0.01	−0.1	−0.17	−0.05
RH	0.22	0.21	0.16	−0.47	−0.07	−0.39	1	0.06	−0.03	−0.02	−0.09	−0.11	0.17
PR	0.2	0.17	0.23	−0.2	0.19	−0.26	0.06	1	−0.01	0.02	−0.26	0.08	−0.07
SR	0	0.08	0.03	0.01	0.14	−0.02	−0.03	−0.01	1	0.32	0.12	−0.01	−0.01
SUN	−0.01	0.04	0.02	0.02	0.08	0.01	−0.02	0.02	0.32	1	0.05	−0.01	−0.02
WS	−0.21	−0.15	−0.31	0.21	−0.06	−0.1	−0.09	−0.26	0.12	0.05	1	−0.16	0.04
WD	−0.16	−0.09	−0.22	0.09	0.04	−0.17	−0.11	0.08	−0.01	−0.01	−0.16	1	−0.01
AP	−0.02	−0.02	−0.03	−0.02	−0.04	−0.05	0.17	−0.07	−0.01	−0.02	0.04	−0.01	1

Step 4: Partitioning of the selected cluster variables into negative and positive correlations, by fuzzifying the crisp numerical correlation values with three triangular membership functions. In this way Linguistics are assigned to the correlations.

Table 5 analytically presents the fuzzification of the correlations of the examined variables that emerged after the Correlation Analysis process.

Table 5. Fuzzification of the correlation analysis with proper Linguistics

	CO	NO	NO_2	O_3	SO_2	AT	RH	PR	SR	SUN	WS	WD	AP
CO	1	+++	++	−−−	++	−−	++	+	−	−	−−	−−	−
NO	+++	1	++	−−−	++	−−	+	+	+	+	−	−	−
NO_2	++	++	1	−−−	++	−	+	++	+	+	−−	−−	−
O_3	−−−	−−−	−−−	1	−−	++	−−−	−−	+	+	+	+	−
SO_2	++	++	++	−−	1	−−	−	+	+	+	−	+	−
AT	−−	−−	−	++	−−	1	−−	−−	−	+	−	−−	−
RH	++	+	+	−−−	−	−−	1	+	−	−	−	−	+
PR	+	+	++	−−	+	−−	+	1	−	+	−−	+	−
SR	−	+	+	+	+	−	−	−	1	++	+	−	−
SUN	−	+	+	+	+	+	−	+	++	1	+	−	−
WS	−−	−	−−	+	−	−	−	−−	+	+	1	−−	+
WD	−−	−	−−	+	+	−−	−	+	−	−	−−	1	−
AP	−	−	−	−	−	−	+	−	−	−	+	−	1

Step 5 (Grid): The third stage involves the design of FCM by introducing and interconnecting all associated variables of the selected cluster, based on the Linguistic correlations that occurred after the fuzzification of the numerical values. The FCM node interaction simulation algorithm was implemented by performing iterative estimations of a new value for each node in every step. The new value depends on the values of the nodes that start edges and point to it, but also on the weight of the acne that joins them. Transfer functions were used to transform the result of the sum of the products of node activation levels. In this way, the new value of each node and the corresponding weight of the edge connecting were calculated. This is practically the approach used to translate the actual value of each variable of the model to the interval [−1, 1]. The negative type of influence is depicted with an orange color and the positive with a blue color. The degree of influence depends on the thickness of each line. The higher the influence the thicker is the line [2]. The FCM of Fig. 2 above presents the degree of influence between Air Temperature (AT), Atmospheric Precipitation (AP) and Air Pollutants (Table 6).

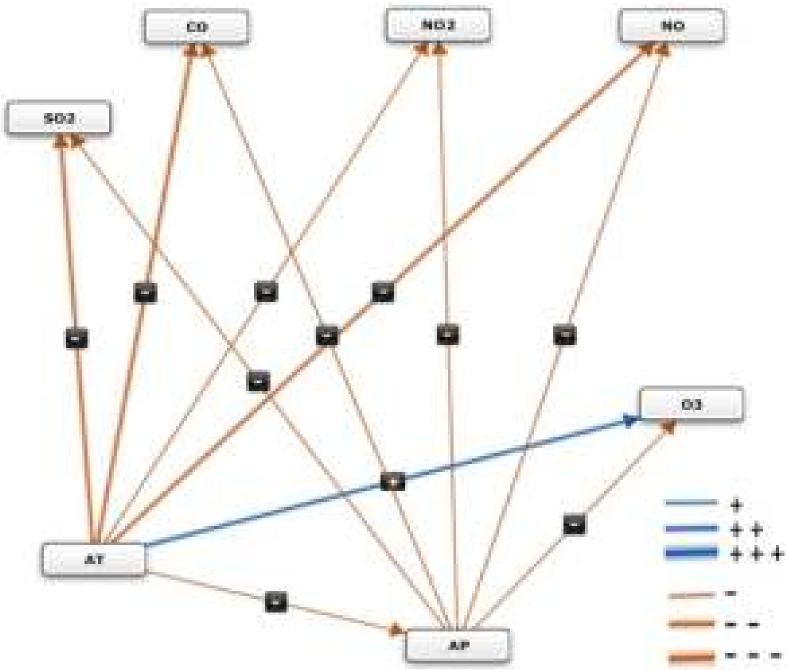

Fig. 2. FCM between air temperature, atmospheric precipitation and air pollutants (Color figure online)

Table 6. The degree of influence between air temperature, atmospheric precipitation and air pollutants

	CO	NO	NO_2	O_3	SO_2	AT	AP
AT	−−	−−	−	++	−−	1	−
AP	−	−	−	−	−	−	1

Step 6 (Scenarios): The fourth stage includes the fuzzification of the temperature and atmospheric precipitation change for the period 2020 to 2099 emerging from the sixteen CMIP5 climatic models, as well as extensive testing of the four climate change scenarios (Fig. 3).

Step 7: Partitioning of scenario variables based on changes in temperature and atmospheric precipitation of sixteen climatic models, as well as fuzzification of the crisp numerical values, with the creation of 12 triangular fuzzy membership functions (FMF). The FMF constructed were related to temperature variations and cover the interval from 0 °C to +7.6 °C, which is also the highest temperature increase for the area under consideration for that period. Numerical values were fuzzified using four triangular FMF and eight semi-triangular ones (S-FMF). Two FMF and four S-FMF were related to temperature changes in the closed interval [0, + 7.64] °C.

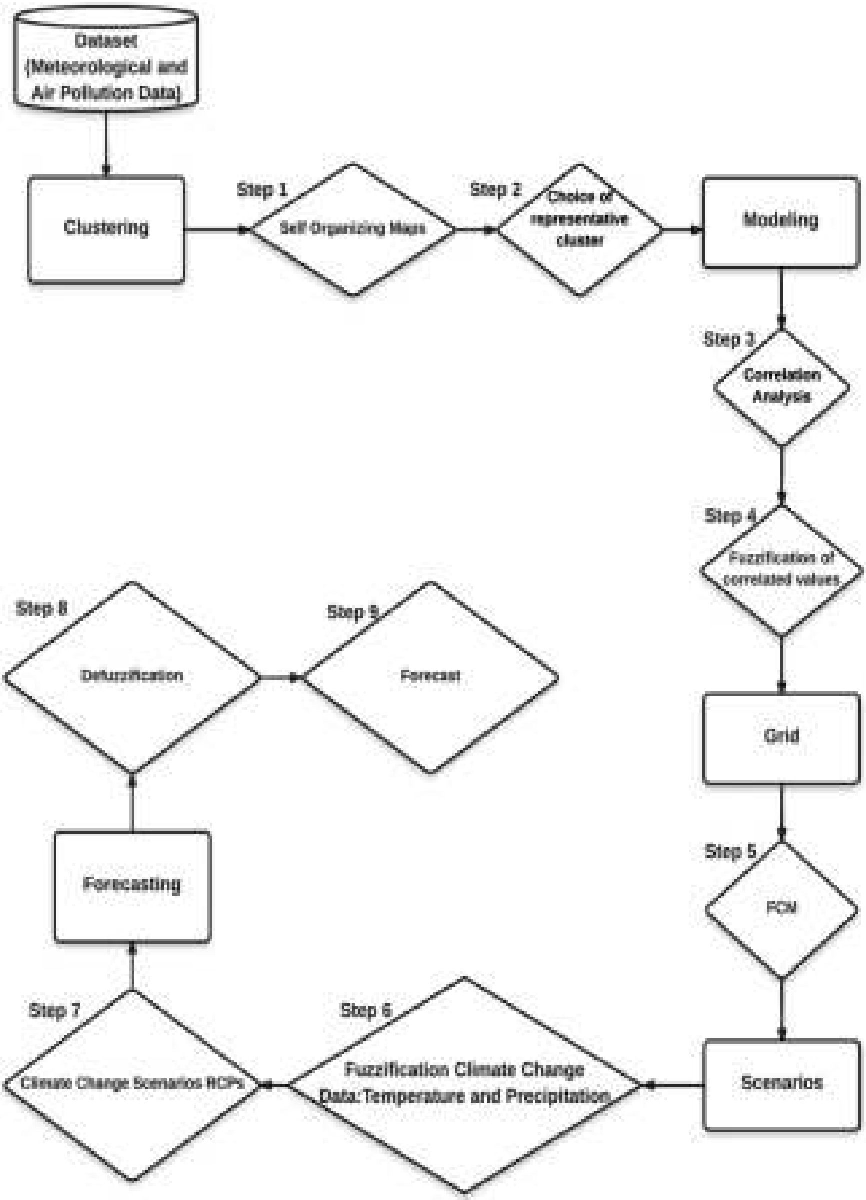

Fig. 3. Flowchart of the proposed model

The first S-FMF, the FMF and the next S-FMF refer to the interval [0, 3.82 °C (median)]. These FMF correspond to the fuzzy sets: low negative (−), middle negative (−−), high negative (−−−), whereas the linguistic high negative (−−−) contains values close to the smallest estimated change. The next S-FMF, the FMF and the last S-FMF refer to the interval [3.82 (median), 7.64 (the highest °C fluctuation based on the most extreme scenario)]. These FMF correspond to the low positive (+), middle positive (++), high positive (+++), with the high positive (+++) being close to the maximum temperature change.

In the same way, two FMF and four S-FMF were developed for the atmospheric precipitation with crisp values in the interval [0, −544.64 mm].

Due to the fact that atmospheric precipitation appears reduced in the future values, the first S-FMF, the FMF and the last S-FMF cover the atmospheric precipitation reduction in the interval [−272.32, −544.64 mm] which corresponds to high atmospheric precipitation reduction. The next S-FMF, the FMF and the last S-FMF were used for the smaller changes [0, −272.32 (median)] mm which correspond to the fuzzy sets low positive (+), middle positive (++), high positive (+++), with high positive (+++) declaring the zero decrease of atmospheric precipitation (0 mm) (Table 7).

Table 7. FMF and S-FMF boundaries (Air Temperature, Atmospheric Precipitation)

Fuzzy Sets corresponding to temperature and Atmospheric precipitation changes	FMF and S-FMF boundaries in the closed interval [0, + 7.64] °C	FMF and S-FMF boundaries in the closed interval [0, −544.64] mm
−−− (S-FMF)	0 1.528	−544.64 −435.7
−− (FMF)	0.382 1.91 3.438	−517.4 −408.5 −299.6
− (S-FMF)	2.292 3.82	−381.2 −272.32
+ (S-FMF)	3.82 5.348	−272.32 −163.4
++ (FMF)	4.202 5.73 7.258	−245.1 −136.2 −27.28
+++ (S-FMF)	6.112 7.64	−108.9 0

Extensive tests of scenarios have been carried out on the basis of changes in temperature and atmospheric precipitation and on their impact on the quality of the atmosphere in Athens. The fuzzy Linguistics produced by the use of climate change scenarios, were defuzzified in order to forecast the air pollutants' potential future crisp values. In this way, we have performed a projection in the distant future for the problem of environmental degradation due to air pollution.

Step 8 (Forecasting): In the fifth stage, the defuzzification procedure is performed to produce the crisp arithmetic values from the fuzzy sets that have been generated previously. The centroid defuzzification method was used, according to which the center of gravity of the distribution of the fuzzy set was calculated by the following Eq. 16:

$$x = \frac{\int x \cdot \mu(x)dx}{\int \mu(x)dx} \tag{16}$$

Step 9: Predicting future pollutant values is based on the amount of relative change of each parameter as a result of the following Eq. 17:

$$Relative\,Change = \frac{FutureValue - InitialValue}{InitialValue} \tag{17}$$

4 Results and Discussion

After extensive testing has been carried out on multiple climate models and scenarios and all possible cases of change in temperature and atmospheric precipitation have been examined, the relevant changes to the pollutants under consideration are shown in Table 8. Finally, 10 different scenarios result from the combination of temperature and atmospheric precipitation change by 2099. Table 8 presents the historical values of the parameters for the period 2000–2012.

4.1 Description of the Ten Scenarios

Several scenarios were used to reduce the uncertainties related to the different configurations used by the distinct climate models. The results produced by a set of simulations of various climate models and scenarios of greenhouse gas emissions were analyzed aiming to lead to a reliable approach of precipitation and temperature changes estimation. Based on these scenarios, we cannot precisely verify long-term climate changes, but it is necessary to study the models that have been widely accepted by the scientific community and oversee the evolution of the impact of climate change on the quality of life and the environment.

The ten scenarios presented represent the time-period 2080–2099 where the most rapid changes in the climate are expected, applying the most extreme climate change scenario RCP8.5 to all climate models. Specifically, Table 8 shows a detailed presentation of the temperature and atmospheric precipitation fluctuations in each scenario as well as the relative changes in the estimated concentrations of the pollutants.

Looking at a detailed presentation of the most significant changes observed, it is evident that a rise in temperature up to the value of (+3.82 °C), coupled with a reduction in millimeters of atmospheric precipitation to a value of (−272.32 mm), contributes significantly towards the increase in all primary pollutants, while reducing secondary pollutants such as O_3 (ID1).

Also, a temperature increase, up to the median value of (+3.82 °C) combined with a great reduction in millimeters of atmospheric precipitation (−415.94 mm) contributes to the increase of all pollutants, while O_3 remains unchanged (ID2).

A slight increase in temperature above the value of median (from 4.15 °C to 4.47 °C), combined with a large reduction in millimeters of atmospheric precipitation (from −370.95 mm to −544.64 mm), contributes to the increase in all pollutants (ID4, ID10).

Table 8. Relative changes of air pollutants based of the climate change scenarios

ID	AT	AP	CO	NO	NO_2	O_3	SO_2
2000-2012 2080-2099	17.38(°C)	711.18mm	48.28	10.44	0.77	15.13	44.01
1	low negative (+3.23)(°C)	low negative (-315.94 mm)	0.05	0.03	0.03	-0.01	0.05
2	low negative (+2.96)(°C)	medium negative (-415.94) mm	0.05	0.04	0.04	0	0.06
3	medium positive (+5.25)(°C)	medium negative (-392.55) mm	-0.02	0	0.01	0.02	0
4	low positive (+4.15)(°C)	high negative (-544.64) mm	0.03	0.03	0.04	0.01	0.05
5	medium negative (+1.93)(°C)	medium negative (-371.57) mm	0.07	0.05	0.05	-0.01	0.07
6	high positive (+7.64)(°C)	medium negative (-432.79) mm	-0.09	-0.06	-0.03	0.05	-0.05
7	medium positive (+5.85)(°C)	medium positive (-77.49) mm	-0.07	-0.05	-0.04	0.01	-0.06
8	medium negative (+2.11)(°C)	low negative (-325.15) mm	0.06	0.04	0.04	-0.01	0.07
9	low positive (+4.54)(°C)	low positive (-227.45) mm	-0.02	-0.01	-0.01	0	-0.02
10	low positive (+4.36)(°C)	medium negative (-370.95) mm	0.01	0.01	0.02	0.01	0.02

Temperature changes of approximately (+2 °C) much lower than the median, combined with a moderate decrease of atmospheric precipitation millimeters (from −325.15 mm to −371.57 mm) contribute to the increase of all pollutants except O_3 (ID5, ID8).

The highest positive relative change in the values of pollutants (CO, NO, NO_2, SO_2) is observed in the ID5 scenario. According to ID5, all pollutants will increase significantly whereas O_3 will show a slight decrease. On the other hand, according to scenarios ID6, ID7 a moderate and rapid temperature increase, irrespective of millimeters of atmospheric precipitation, will contribute to a significant reduction in concentrations of all pollutants as opposed to significantly increased O_3.

The ID6 scenario shows the largest positive relative change in O_3 values and is considered the most extreme scenario as it predicts a particularly significant increase in temperature and O_3 affecting the morbidity and mortality of residents of an urban center such as Athens.

According to scenarios ID1, ID5 and ID8 the temperature will be lower than the mean value and it will not increase significantly compared to the historical data for the period 2000–2012. However, there will be a significant increase in CO, NO, NO_2, SO_2 whereas O_3 will remain unchanged.

It is observed that there is an inverse relationship to the variation of CO, NO, NO_2, SO_2 and O_3 which is determined by the temperature. When the temperature increases the values of CO, NO, NO_2, SO_2 decrease whereas O_3 increases and vice versa.

The last twenty years of the 21st century, are of great interest (2080–2099) in which two extreme scenarios are observed. In the ID6 scenario the temperature will show an increase of 44% (7.64 °C, high positive (+++)) and the rainfall will decrease by 61% (−432.79 mm, medium negative (−−))

Based on the ID4 scenario, the temperature will rise, up to 24% (4.15 °C, low positive (+)) and the rainfall will be reduced by 76% (−544.64 mm, high negative (−−−)). The combination of changes in the two most important climatic parameters of the two extreme scenarios resulted in an increase of O_3 when ID6 scenario has been applied whereas all pollutants have been increased according to the ID4 scenario (Table 9).

Table 9. Air pollutants forecasted values based on the two extreme scenarios

Period	CO	NO	NO_2	O_3	SO_2
2000–2012	0.77	15.13	44.01	48.28	10.44
2080–2099 (ID4)	0.80	15.59	45.77	48.76	10.96
Absolute increase value	**+0.03**	**+0.46**	**+1.76**	**+0.48**	**+0.52**
2080–2099 (ID6)					
Absolute increase value	**0.71** **−0.06**	**14.22** **−0.91**	**42.69** **−1.32**	**50.69** **+2.41**	**9.91** **−0.53**

In our previous work [2] related to the extreme scenario of the 2080–2099 period that includes the application of the climatic model GFDL_CM2.0 the temperature increased by 15% and the rainfall decreased by 49%. If we assume that this scenario will prevail at some time by 2100, temperature and atmospheric precipitation will significantly affect pollutants as it is included in Table 8 ID5.

By looking at an assessment of the air quality risk based on the extreme scenario (ID6), the increase in O_3 entails an increase in temperature and solar radiation by affecting cardiovascular and respiratory conditions while equating to the frequent presence of photochemical cloud and temperature reversal in the summer months (Table 10).

Table 10. Air pollutants forecasted values based on the ID5 scenario

Period	CO	NO	NO_2	O_3	SO_2
2000–2012	0.77	15.13	44.01	48.28	10.44
2080–2099 (ID5)	0.83	15.89	46.21	47.8	11.17
Absolute increase value	**+0.06**	**+0.76**	**+2.2**	**−0.48**	**+0.73**

Also, the reduction of SO_2 and CO values suggest a decrease in the frequency of the smog cloud in the winter months and in the decrease of diseases related to these pollutants. Additionally, the ID4 scenario indicates the presence of adverse meteorological conditions in both winter and summer due to increased concentrations of all pollutants. The graph below shows the values of atmospheric pollutant concentrations as these will be formed by the application of the three climate scenarios (ID4, ID5, ID6) over the period 2080–2099 Fig. 4.

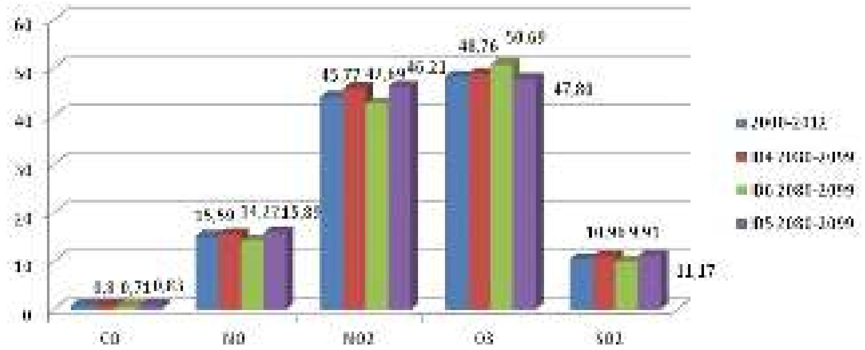

Fig. 4. Graphical Display of air pollutants based on the three scenarios of climate change

According to the most extreme scenario, the rise in temperature by +7.64 °C, with a mean decrease in millimeters of rainfall entails a significant increase in Ozone alone and therefore an increase in photochemical cloud and respiratory diseases. However, if mild meteorological conditions prevail in terms of temperature change by increasing by 2.86 °C to 4.15 °C while reducing millimeters of rainfall then atmospheric conditions are differentiated. Under these conditions, concentrations of CO, NO, SO_2 and even higher in NO_2 will be observed. According to this scenario O_3 will be more volatile as its concentrations will decrease with a lower temperature rise. Conversely, its concentrations will increase as the temperature rises, while the combination of temperature increase with a significant decrease in millimeters of precipitation will increase its concentrations even further.

While in few scenarios the temperature is dramatically rising above 5 °C the decrease of precipitation appears to be rapidly evolving, as a 50% rainfall reduction has been observed in most of the scenarios. If the decline in millimeters of rainfall is verified, the absence of precipitation for a long time will pose risks such as drought or water scarcity, increased forest fires, disease outbreaks, low food production and increased hospital admissions.

The results of this research are significant but also equally alarming as they indicate that over the years climate change will change the concentrations of pollutants by affecting the health of the residents of Athens. Therefore, continuous monitoring, climate forecasting as well as constant update and alertness of state mechanisms are required to minimize the impact on health and the environment from future increases in concentrations of pollutants in the atmosphere.

Summarizing, we conclude that the CMIP5 climate models used, can provide a specific set of useful simulations of the potential short and long-term climate changes and they offer a clear and rational identification of the factors responsible for the differences in models' results. The models used, perform long-term simulations. Their input data are reliable and safe output of the coupled ocean-atmosphere models (AOGCMs).

5 Conclusions

This paper proposes the use of an innovative method of air quality study, as well as a hybrid intelligent system for the clustering of the involved parameters and forecasting of atmospheric pollution. This whole effort is based on a combination of Soft computing sophisticated methods. In particular, clustering of data is performed using SOM in order to isolate and study the values of the parameters that determine and represent the problem of air pollution. In addition, this model uses FCM to analyze and model the causal interconnections of the parameters determining the atmospheric pollution, while fuzzifying the correlated values of the meteorological parameters and pollutants of Athens center for the period 2000–2012. Concentrations of atmospheric pollutants are evaluated in relation to the variation in temperature and atmospheric precipitation values as these are modeled on the projections of sixteen climatic models for the period up to 2100. The application of climate models testing various scenarios over the period 2080–2099 presented important data on the air quality risk and on the effects of air pollution on the environment and on public health.

Future research could involve its model, which will combine semi supervised methods and online learning. Also, could be improved towards a better online learning with self-modified the number of hidden nodes. Moreover, additional computational intelligence methods could be explored, tested and compared on the same task in an ensemble approach. Finally, the ultimate challenge would be the scalability of model with bio-inspired optimization algorithms in parallel and distributed computing in a real-time system.

References

1. Amer, M., Jetter, A.J., Daim, T.U.: Scenario planning for the national wind energy sector through fuzzy cognitive maps. In: Proceedings of PICMET 2013: Technology Management in the IT-Driven Services, pp. 2153–2162. IEEE, San Jose (2013)
2. Anezakis, V.-D., Dermetzis, K., Iliadis, L., Spartalis, S.: Fuzzy cognitive maps for long-term prognosis of the evolution of atmospheric pollution, based on climate change scenarios: the case of Athens. In: Nguyen, N.-T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds.) ICCCI 2016. LNCS (LNAI), vol. 9875, pp. 175–186. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45243-2_16
3. Anezakis, V.-D., Iliadis, L., Demertzis, K., Mallinis, G.: Hybrid soft computing analytics of cardiorespiratory morbidity and mortality risk due to air pollution. In: Dokas, I., Bellamine-Ben Saoud, N., Dugdale, J., Díaz, P. (eds.) Proceedings of Information Systems for Crisis Response and Management in Mediterranean Countries, ISCRAM-med 2017, Lecture Notes in Business Information Processing, (LNCS), vol. 301, pp. 87–105. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67633-3_8
4. Bougoudis, I., Demertzis, K., Iliadis, L.: Fast and low cost prediction of extreme air pollution values with hybrid unsupervised learning. Integr. Comput. Aided Eng. **23**(2), 115–127 (2016). https://doi.org/10.3233/ica-150505

5. Bougoudis, I., Demertzis, K., Iliadis, L.: HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens. Neural Comput. Appl. **27**(5), 1191–1206 (2015). https://doi.org/10.1007/s00521-015-1927-7
6. Bougoudis, I., Demertzis, K., Iliadis, L., Anezakis, V.-D., Papaleonidas, A.: Semi-supervised hybrid modeling of atmospheric pollution in urban centers. In: Jayne, C., Iliadis, L. (eds.) EANN 2016. CCIS, vol. 629, pp. 51–63. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44188-7_4
7. Bougoudis, I., Demertzis, K., Iliadis, L., Anezakis, V.D., Papaleonidas, A.: FuSSFFra, a fuzzy semi-supervised forecasting framework: the case of the air pollution in Athens. Neural Comput. Appl. **29**, 375–388 (2017). https://doi.org/10.1007/s00521-017-3125-2
8. Bougoudis, I., Iliadis, L., Papaleonidas, A.: Fuzzy inference ANN ensembles for air pollutants modeling in a major urban area: the case of Athens. In: Mladenov, V., Jayne, C., Iliadis, L. (eds.) EANN 2014, CCIS, vol. 459, pp. 1–14. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-319-11071-4_1
9. Bougoudis, I., Iliadis, L., Spartalis, S.: Comparison of self organizing maps clustering with supervised classification for air pollution data sets. In: Iliadis, L., Maglogiannis, L., Papadopoulos, H. (eds.) AIAI 2014, IFIP AICT, vol. 436, pp. 424–435. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44654-6_42
10. Fons, S., Achari, G., Ross, T.: A fuzzy cognitive mapping analysis of the impacts of an eco-industrial park. J. Intell. Fuzzy Syst. **15**(2), 75–88 (2004)
11. García, C.G., Ortiz, I.P.: Stability analysis of climate system using fuzzy cognitive maps. In: Obaidat, M.S., Filipe, J., Kacprzyk, J., Pina, N. (eds.) Simulation and Modeling Methodologies, Technologies and Applications. AISC, vol. 256, pp. 211–222. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-03581-9_15
12. Glorennec, P.Y.: Forecasting ozone peaks using self-organizing maps and fuzzy logic. In: Sportisse, B. (ed.) APMS 2001, pp. 544–550. Springer, Heidelberg (2002). https://doi.org/10.1007/978-3-662-04956-3_52
13. Gordaliza, J.A., Flórez, R.E.V.: Using fuzzy cognitive maps to support complex environmental issues learning. In: Proceedings of New Perspectives in Science Education Conference, 2nd edn (2013)
14. Hájek, P., Olej, V.: Air quality modeling by Kohonen's self-organizing feature maps and LVQ neural networks. WSEAS Trans. Environ. Dev. **4**(1), 45–55 (2008)
15. Haykin, S.: Neural Networks and Learning Machines, 3rd edn. Pearson Education, New York (2009)
16. Iliadis, L., Spartalis, S., Paschalidou, A., Kassomenos, P.: Artificial neural network modeling of the surface ozone concentration. Int. J. Comput. Appl. Math. **2**(2), 125–138 (2007)
17. Jiang, N., Betts, A., Riley, M.: Summarising climate and air quality (ozone) data on self-organising maps: a Sydney case study. Environ. Monit. Assess. **188**(2), 103 (2016). https://doi.org/10.1007/s10661-016-5113-x
18. Karatzas, K.D., Voukantsis, D.: Studying and predicting quality of life atmospheric parameters with the aid of computational intelligence methods. In: Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A., Guariso, G. (eds.) International Environmental Modeling and Software Society (iEMSs 2008), vol. 2. pp. 1133–1139. iEMSs (2008)
19. Khedairia, S., Khadir, M.T.: Impact of clustered meteorological parameters on air pollutants concentrations in the region of Annaba, Algeria. Atmos. Res. **113**, 89–101 (2012). https://doi.org/10.1016/j.atmosres.2012.05.002
20. Kohonen, T.: Self-Organization and Associative Memory, 3rd edn. Springer, Berlin (1989). https://doi.org/10.1007/978-3-642-88163-3

21. Li, S.T., Chou, S.W., Pan, J.J.: Multi-resolution spatio-temporal data mining for the study of air pollutant regionalization. In: Proceedings of the 33rd Hawaii International Conference on System Sciences, USA, p. 33. IEEE (2000)
22. Luiz, J., Muller, E.: Greenhouse gas emission reduction under the Kyoto protocol: the South African example. Int. Bus. Econ. Res. J. **7**, 75–92 (2008)
23. Mesa-Frias, M., Chalabi, Z., Foss, A.M.: Assessing framing assumptions in quantitative health impact assessments: a housing intervention example. Environ. Int. **59**, 133–140 (2013). https://doi.org/10.1016/j.envint.2013.06.002
24. Mourhir, A., Rachidi, T., Papageorgiou, E.I., Karim, M., Alaoui, F.S.: A cognitive map framework to support integrated environmental assessment. Environ. Model. Softw. **77**, 81–94 (2016). https://doi.org/10.1016/j.envsoft.2015.11.018
25. Neme, A., Hernández, L.: Visualizing patterns in the air quality in Mexico City with self-organizing maps. In: Laaksonen, J., Honkela, T. (eds.) WSOM 2011. LNCS, vol. 6731, pp. 318–327. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21566-7_32
26. Olej, V., Hájek, P.: Air quality modelling by Kohonen's self-organizing feature maps and intuitionistic fuzzy sets. In: Proceedings of the 12th IASTED International Conference on Artificial Intelligence and Soft Computing, ASC 2008, Spain, pp. 22–27. Elsevier B.V. (2008)
27. Papageorgiou, E.I., Salmeron, J.L.: A review of fuzzy cognitive maps research during the last decade. IEEE Trans. Fuzzy Syst. **21**(1), 66–79 (2013). https://doi.org/10.1109/TFUZZ.2012.2201727
28. Papageorgiou, E.I., Salmeron, J.L.: Methods and algorithms for fuzzy cognitive map-based modeling. Intell. Syst. Ref. Libr. **54**, 1–28 (2014). https://doi.org/10.1007/978-3-642-39739-4_1
29. Paschalidou, A.: University of Ioannina, Ph.d. thesis development of box model for the air pollution forecasting in medium size cities (2007). (in Greek)
30. Pathinathan, T., Ponnivalavan, K.: The study of hazards of plastic pollution using induced fuzzy cognitive maps (IFCMS). J. Comput. Algorithm **3**, 671–674 (2014)
31. Paz-Ortiz, I., Gay-García, C.: Fuzzy cognitive mapping and nonlinear Hebbian learning for the qualitative simulation of the climate system, from a planetary boundaries perspective. In: Obaidat, M.S., Ören, T., Kacprzyk, J., Filipe, J. (eds.) Simulation and Modeling Methodologies, Technologies and Applications. AISC, vol. 402, pp. 295–312. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26470-7_15
32. Pearce, J.L., et al.: Exploring associations between multipollutant day types and asthma morbidity: epidemiologic applications of self-organizing map ambient air quality classifications. Environ. Health Glob. Access Sci. Sour. **14**(1), 1–12 (2015). https://doi.org/10.1186/s12940-015-0041-8
33. Rodgers, J.L., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. Am. Stat. **42**(1), 59–66 (1988). https://doi.org/10.1080/00031305.1988.10475524
34. Salmeron, J.L., Froelich, W.: Dynamic optimization of fuzzy cognitive maps for time series forecasting. Knowl. Based Syst. **105**, 29–37 (2016). https://doi.org/10.1016/j.knosys.2016.04.023
35. Scafetta, N., Willson, R.C.: ACRIM total solar irradiance satellite composite validation versus TSI proxy models. Astrophys. Space Sci. **350**(2), 421–442 (2014). https://doi.org/10.1007/s10509-013-1775-9
36. Tamas, W., Notton, G., Paoli, C., Nivet, M.L., Voyant, C.: Hybridization of air quality forecasting models using machine learning and clustering: an original approach to detect pollutant peaks. Aerosol Air Qual. Res. **16**(2), 405–416 (2016). https://doi.org/10.4209/aaqr.2015.03.0193

37. Vidal, R., Salmeron, J.L., Mena, A., Chulvi, V.: Fuzzy cognitive map-based selection of TRIZ trends for eco-innovation of ceramic industry products. J. Clean. Prod. **107**, 202–214 (2015). https://doi.org/10.1016/j.jclepro.2015.04.131
38. Zhang, H., Song, J., Su, C., He, M.: Human attitudes in environmental management: fuzzy cognitive maps and policy option simulations analysis for a coal-mine ecosystem in China. J. Environ. Manag. **115**, 227–234 (2013). https://doi.org/10.1016/j.jenvman.2012.09.032