Investigation and Forecasting of the Common Air Quality Index in Thessaloniki, Greece

Ioannis Kyriakidis, Kostas Karatzas, George Papadourakis, Andrew Ware, Jaakko Kukkonen

▶ **To cite this version:**

Ioannis Kyriakidis, Kostas Karatzas, George Papadourakis, Andrew Ware, Jaakko Kukkonen. Investigation and Forecasting of the Common Air Quality Index in Thessaloniki, Greece. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.390-400, 10.1007/978-3-642-33412-2_40 . hal-01523061

HAL Id: hal-01523061
https://hal.science/hal-01523061

Submitted on 16 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Investigation and Forecasting of the Common Air Quality Index in Thessaloniki, Greece

Ioannis Kyriakidis[1,2], Kostas Karatzas[3], George Papadourakis[2], Andrew Ware[1] and Jaakko Kukkonen[4]

[1]University of Glamorgan, School of Computing, Pontypridd, Wales,United Kingdom
[2]Department of Applied Informatics & Multimedia, T.E.I. of Crete, Greece
[3]Department of Mechanical Engineering, Aristotle University of Thessaloniki, Greece
[4]Finnish Meteorological Institute, Helsinki, Finland

kyriakidis@teicrete.gr, kkara@eng.auth.gr, papadour@cs.teicrete.gr, jaware@glam.ac.uk, jaakko.kukkonen@fmi.fi

Abstract. Air pollution can affect health and well-being of people and eco-systems. Due to the health risk posed for sensitive population groups, it is important to provide with hourly and daily forecasts of air pollution. One way to assess air pollution is to make use of the Common Air Quality Index (CAQI) of the European Environment Agency (EEA). In this paper we employ a number of Computational Intelligence algorithms to study the forecasting of the hourly and daily CAQI. These algorithms include artificial neural networks, decision trees and regression models combined with different datasets. The results provide with a satisfactory CAQI forecasting performance that may be the basis of an operational forecasting system.

Keywords: air pollution, common air quality index, neural networks, decision trees, linear regression

1 Introduction

High levels of air pollution are associated with a number of health problems and affect quality of life [1]. As an example it may be noted that ambient air pollution, in terms of fine particulate matter (PM2.5), causes about 3% of mortality from cardiopulmonary disease, about 5% of mortality from cancer of the trachea, bronchus, and lung, and about 1% of mortality from acute respiratory infections in children (age<5yr), worldwide [2]. On this basis, the importance of forecasting of poor air quality is evident.

Two forecasting approaches may be employed. The first one makes use of deterministic, numerical models that are capable of "representing" the physical and chemical processes affecting air pollution. They have the advantage of fully covering a geographical area of interest, but they require detailed input in terms of emission data, meteorology and physico-chemical parameters, and they are computationally demanding. The other forecasting approach makes use of data-driven modeling, employing either statistical [3] or Computational Intelligence (CI) methods [4],[5].

We make use of the CAQI of the EEA as the parameter of interest that needs to be forecasted in order to provide health related warnings and information to people. We selected the city of Thessaloniki, Greece as our target city, and we made an effort to construct operational forecasting models, with the aid of CI methods that include Artificial Neural Networks (ANNs) and Decision Trees (DT), as well as statistical methods (Linear Regression Models). In the rest of the paper we are describing the materials and methods that we used (chapter 2), the results of our calculations and the relevant discussion (chapter 3), while we draw our conclusions in chapter 4.

2 Materials and Methods

2.1 Area of interest and datasets used

Thessaloniki is the second largest city of Greece, and is characterized by a pronounced problem of air pollution, especially in terms of Particulate Matter (PM) and more specifically of respirable particles (PM_{10}, i.e. particles of aerodynamic diameter smaller than 10 μm). Air quality monitoring is conducted in various locations of the Greater Thessaloniki Area (www.airthess.gr), and data from these measurements were used in the frame of the current study.

We analyzed data from four monitoring stations located in (a) Agia Sofia (city center), (b) Panorama (a suburb in the northern mountainous area of the city), (c) Sindos (industrial-suburban area in the west of the city influenced by traffic) and (d) Kordelio (a densely populated area close to the industrial area of the city, influenced by traffic). These data included hourly measurements of air pollutants such as Carbon Monoxide (CO), Nitrogen Dioxide (NO_2), Nitrogen Monoxide (NO), Ozone (O_3), Respirable particles (PM_{10}), and Sulphur dioxide (SO_2) and they were complemented by meteorological observations. Data came from the years 2001-2003 (time period with a low percentage of problematic or missing data).

2.2 The Common Air Quality Index

There has been an initiative to combine the impacts of various pollutants and come up with an aggregated measure that may be used for the description of the atmospheric quality. Various air quality indices have been suggested by institutes and authorities in various countries [6], [7]. The European Environment Agency has adopted an index called the Common Air Quality Index (CAQI, www.eyeonearth.org) to identify the atmospheric quality towards public safety and health.

The CAQI is designed to present and compare air quality in near-real time on an hourly or daily basis. The CAQI has 5 categorical levels, corresponding to a range of values starting from 0 (very low) to >100 (very high). Two types of a CAQI are specified, an urban background index and a traffic related index. The urban background index takes into account three main pollutants (NO_2, PM_{10} and O_3) and two auxiliary pollutants (SO_2 and CO) while the traffic index takes to into account two main pollutants (NO_2 and PM_{10}) and also one auxiliary (CO). Those are described in detail in [8], and they are calculated as explained in Table 1. This CAQI was used

to calculate the air quality levels (hourly and daily) for each of the four locations of interest. The traffic and urban background CAQI were calculated per station in accordance to the nature of the monitoring sites: the former was applied for Agia Sofia, Kordelio and Sindos, while the latter for the Panorama station (Table 2)

Table 1. Calculating the two types of CAQI (background and traffic index) by EEA [8]. All concentrations are in μg/m^3. [CO]: 8hours moving average values

$SO_2\_Index([SO_2]) = \begin{cases} [SO_2]/2 & 0<[SO_2]<100 \\ ([SO_2]/8)+37.5 & 100<[SO_2]<500 \\ [SO_2]/5 & [SO_2]>500 \end{cases}$	$NO_2\_Index([NO_2]) = \begin{cases} [NO_2]/2 & 0<[NO_2]<100 \\ ([NO_2]/4)+25 & 100<[NO_2]<200 \\ ([NO_2]/8)+50 & 200<[NO_2]<400 \\ [NO_2]/4 & [NO_2]>400 \end{cases}$
$O_3\_Index([O_3]) = ([O_3]*5)/12$	$PM_{10}\_Index([PM_{10}]) = [PM_{10}]$
$CO\_Index([CO]) = \begin{cases} [CO]/200 & 0<[CO]<5000 \\ ([CO]/100)-25 & 5000<[CO]<10000 \\ ([CO]/400)+50 & 10000<[CO]<20000 \\ [CO]/200 & [CO]>20000 \end{cases}$	
Main Background Index = Max(NO_2_Index, PM_{10}_Index, O_3_Index)	
Auxiliary Background Index = Max(NO_2_Index, PM_{10}_Index, O_3_Index, SO_2_Index, CO_Index)	
Main Traffic Index = Max(NO_2_Index, PM_{10}_Index)	
Auxiliary Traffic Index = Max(NO_2_Index, PM_{10}_Index, CO_Index)	

Table 2. The distribution of the CAQI per level and station. The calculation grid for the CAQI is: Very Low (0-25), Low (26-50), Medium (51-75), High (76-100), Very high (>100).

Index Level	Agia Sofia (2001-2003)		Panorama (2001-2002)		Sindos (2001)		Kordelio (2001-2002)	
	Hourly	Daily	Hourly	Daily	Hourly	Daily	Hourly	Daily
Very high	3510	127	243	1	591	0	2538	95
High	3706	168	613	12	618	0	1958	112
Medium	7414	438	5569	306	1917	3	3866	225
Low	8264	304	9427	377	3558	361	5056	183
Very low	1666	18	885	21	1309	1	1514	17

2.3 Pre-processing of input data and methods for the forecasting of the CAQI

The categorization of the CAQI into various levels reveals that in order to forecast the CAQI for the next hours and days, we need to make use of methods that can effectively classify existing (historical) data, and forecast the classification value of interest. On this basis, and taking into account relevant literature, we employed CI methods, and more specifically Artificial Neural Networks and Decision Trees. ANNs were advanced in the late ’80s, popularizing techniques like Multi-layer Perceptrons (MLP) [9] and Self-Organizing Maps (SOM) [10] while they can be trained to successfully approximate virtually any continuous function [11]. Their advantages

also include greater fault tolerance, robustness, and adaptability, especially compared to expert systems, due to the large number of interconnected processing elements that can be trained to learn from new patterns [12]. These features provide ANNs with the potential to model complex non-linear phenomenon like air pollution [13].

A Decision Tree on the other hand, is a hierarchical data structure implementing the divide-and-conquer strategy [14]. It is an efficient nonparametric method, which can be used for both classification and regression. DTs are essentially a map of the reasoning process in which a tree-like graph is constructed to explore options and investigate the possible outcomes of choosing the options. The reasoning process starts from a root node, transverses along the branches tagged with decision nodes, and terminates in a leaf node [15].

The above two CI methods were used together with Linear Regression (LR) models, which are based on a linear combination of the input parameters. Such models may take into account the "memory" and inertia interwoven to the air quality system of any area and have been proven successful especially if they are "fed" with lagged values, i.e. past values of the parameter of interest.

From the results of Table 2 it was made evident that Agia Sofia station demonstrates the highest values in terms of "High" and "Very High" CAQI classes. For this reason, it was selected as the station for the development of CAQI forecasting models. Air quality and meteorological data used for LR, ANNs and DT model construction, originate from this location. In order to forecast hourly and daily CAQI, three datasets were used, depending on the forecasting target.

1. Dataset 1: Dataset with only lagged (i.e. previous) CAQI values (from 1 to 10), either hourly or daily, according to the type of index being calculated. Dataset 1 consists of 1042 daily records and 23671 hourly records.
2. Dataset 2: this included air quality (CO, SO_2, O_3) and meteorological values (Temperature, Dew Point, Humidity, Sea Level Pressure, Wind Direction). Dataset 2 consists of 1043 daily records and 23672 hourly records.
3. Dataset 3: this included lagged (previous) CAQI values, (from 1 to 5), hourly or daily, according to the type of index being calculated. In addition, this dataset also included air quality and meteorological values. Dataset 3 consists of 1042 daily records and 23671 hourly records.

Concentration data were preprocessed before used as input to the forecasting algorithms (ANNs and DTs). The first step of the preprocessing procedure was to remove all records that did not include values for the target characteristic(s). Specifically, for Datasets 1 and 3, all records that do did include CAQI values, were removed. In the case of Dataset 2 all records that did not include all target values (CO, SO_2 and O_3) were also removed. The next step was to use an interpolation method to calculate all missing values [16]. The wind direction values were encoded by using the formula $WD=1+\sin(\theta+\pi/4)$, in order to replace the cyclic nature of this variable with a linear one. In addition, as ANNs require numerical values as input, nominal values were converted to numerical values for the CAQI, as described in Table 3. The specific numerical values were selected, because of the hyperbolic tangent sigmoid transfer function that was used to normalize the input data in the case of ANNs and DTs. Table 4 presents the way that the continuous numerical values (forecasts) were mapped back to nominal CAQI values. For this mapping the numerical values were divided in five equal in range groups.

Table 3. Conversion table from CAQI Levels (nominal values) to numerical values.

CAQI Level Nominal Value	Numerical Value
Very Low	-1
Low	-0.5
Medium	0
High	0.5
Very High	1

Table 4. Conversion table from continuous numerical values to CAQI Levels (nominal values).

Numerical Value (n)	CAQI Level Nominal Value
n<=-0.6	Very Low
n>-0.6 AND n<=-0.2	Low
n>-0.2 AND n<0.2	Medium
n>=0.2 AND n<0.6	High
n>=0.6	Very High

The performance of the models was estimated with the aid of statistical indicators like the Index of Agreement and the Cohen's Kappa, which are defined as follows:

$$IA = 1 - \frac{\sum_{i=1}^{n} |P_i - O_i|^2}{\sum_{i=1}^{n} \left(|P_i - \overline{O}| + |O_i - \overline{O}| \right)^2} \quad (1)$$

where P_i and O_i are the forecasted and observed concentration values respectively. The index of agreement ranges between 0 and 1, with 0 stating no agreement, and best agreement given by the limiting case of 1.

The Cohen's Kappa or critical success index (CSI) verifies how well the events of interest (CAQI values) were forecasted and it is unaffected by the number of correct forecasts. Its calculation is based on the formulae: $CSI = \frac{a}{a+b+c}$ where *a*,*b*, *c* and *d* are calculated as follows: a= true positives, b= false negatives, c= false positives, d=true negatives. Positives or negatives refer to the occurrence or not of the event of interest (i.e correct classification of the CAQI), respectively. The CSI used in this study does not use the "*d*" events (true negatives). As a consequence, the CSI is tending to give poorer scores for rare events, but it takes into account both false negative events and false positive events. Therefore, the CSI can be characterized as a more balanced score [17].

3 Results and discussion

3.1 CAQI Forecasting with the aid of Linear Regression Models

Forecasting of the numerical hourly and daily CAQI values

LR models were used as the reference method for the development of the CAQI forecasting models. For this reason lagged values of the CAQI (a total of up to 10 values, Dataset 1) were tested. Results indicate that the hourly CAQI numerical values for Agia Sofia can be forecasted by a LR model by employing only one lagged value (IA = 0.92). Contrary to the hourly, the daily CAQI numerical values were not as successfully forecasted (IA = 0.76).

Forecasting of the nominal hourly and daily CAQI levels

The calculations were repeated for the CAQI levels (nominal values), in addition to their numerical values. Overall results are presented in Tables 5 and 6. On this basis it

is made evident that both hourly and daily CAQI levels are not forecasted in a satisfactory manner with the aid of LR models. This can be verified via their low Cohen's Kappa Index (Table 7).

Table 5. Observed vs forecasted hourly CAQI Levels for Agia Sofia via LR. Rows sum 100%.

Hourly CAQI Levels		forecasted CAQI Levels				
		Very low	Low	Medium	High	Very High
Observed CAQI Levels	Very low	65.31	31.88	1.10	0.72	0.99
	Low	7.14	74.74	16.49	1.18	0.44
	Medium	0.23	23.22	61.78	12.43	2.34
	High	0.17	4.11	33.55	46.95	15.21
	Very High	0.03	1.30	6.74	23.00	68.93

Table 6. Observed vs forecasted daily CAQI Levels for Agia Sofia via LR. Rows sum 100%.

Daily CAQI Levels		Forecasted CAQI Levels				
		Very low	Low	Medium	High	Very High
Observed CAQI Levels	Very low	23.53	35.29	41.18	0.00	0.00
	Low	5.75	62.94	23.64	6.07	1.60
	Medium	0.64	33.26	52.36	9.66	4.08
	High	0.00	12.87	47.95	25.73	13.45
	Very High	0.00	3.94	24.41	29.92	41.73

As a next step, the performance of the LR models for the next day's CAQI levels forecast was further investigated (nominal values). For this purpose, the equal as well as the weighted participation of hourly CAQI values in the forecasting model was employed, as described below.

Forecasting of the numerical 24-hourly values to calculate the daily CAQI levels

The fact that the forecast of hourly CAQI values reached a 0.92 IA led us to investigate the forecasting of the daily CAQI levels via forecasting first the 24 hourly values of the next day. To do so, the hourly CAQI values of the previous day were used in the following manner: for the forecasting of the first hourly CAQI value for the next day, the 24 hourly values of the previous day (observed ones) were used. For the forecasting of the second hourly CAQI value of the next day, the previous (forecasted) value and the 23 hourly values of the previous day (observed ones) were used, and so forth. Via this procedure, the hourly CAQI values were forecasted and then we calculated the daily average CAQI value.

Table 8 presents the comparison of daily CAQI forecasting performance via LR models at the station of Agia Sofia, by using lagged observed numerical values and by (equally) using 24 hourly forecasted values. The hourly CAQI numerical values for the Agia Sogia station achieved an IA = 0.70, on the basis of LR models. The performance was decreased in comparison to the performance of the model that uses ten observed lagged values, which reached IA = 0.92.On the other hand, the daily CAQI Levels performance was increased when using 24 forecasted lagged values with an IA=0.5078. This difference in the performance may be attributed to the use of

a higher percentage of observations for the forecasting of the hourly CAQI values for the first hours of the next day, thus allowing information of the previous day to "penetrate" the target day of the forecast.

Daily CAQI Levels based on 24-h Forecasted Values with Weighting Factors

In order to further investigate the influence of input values to the performance of the forecasting models, additional daily CAQIs were calculated by using weighted factors. These weighted factors dictated that the first forecasted values of each day (values at midnight, 1 a.m., 2 a.m. etc.) will contribute more than the last forecasted values. Table 9 presents one of the weighting methods tested, named "Factor4", which provides the best forecasting performance, in comparison to others. When using the hourly forecasted CAQI values to calculate the daily CAQI, the results are slightly better (IA= 0.5105, i.e. an improvement of 1,51%) than those when forecasting directly the daily CAQI.

Table 7. Results of hourly and daily CAQI level forecasts via LR models (Ag. Sofia).

	IA	Cohen's Kappa Index
Hourly CAQI Levels	0.6510	0.53
Daily CAQI Levels	0.4954	0.28

Table 8. Comparison of daily CAQI forecasting performance via LR models (Ag. Sofia).

Forecast by using:	Forecasting performance (IA)	
	Daily CAQI Values	Daily CAQI Levels
10 observed lagged values-Dataset 1	0.92	0.4954
24 forecasted lagged values	0,70	0.5078

Table 9. "Factor4" weighting of hourly values at Agia Sofia.

Hourly Values	Contribution to daily value
1-6	46.25%
7-12	28.75%
13-18	18.75%
19-24	6.25%

3.2 CAQI Forecasting with the aid of ANNs and Decision Trees

In order to develop ANN and DT models, data were firstly normalized by applying the hyperbolic tangent sigmoid transfer function, which was also applied on the hidden layers. On the output layer the linear transfer function was used. This is a common structure for function approximation (or regression) problems. The first 50% of the data is used for training, the next 25% for validation and finally the last 25% for testing. Validation data are used in order to early terminate training (i.e. if the network performance on the validation data fails to improve), thus avoiding over-fitting. Test data are used for the estimation of the model performance and the calculation of the relevant indices. Concerning the ANN architecture, one hidden layer was used in all models. The number of neurons for each model was equal to the number of the input parameters. For the training phase we implemented the Levenberg-Marquardt Back propagation algorithm [18]. Principal Component Analysis (PCA) [19] was performed to the data in order to change their dimensionality, as this helps the ANN to find patterns into the data. It should also be noted that none of the input parameters were excluded. Several experiments were performed in order to forecast the CAQI levels by using ANNs and Decision Trees:

1. Hourly and daily CAQI (numerical) values were forecasted with ANNs, in order to calculate the hourly and daily CAQI (nominal) levels. Table 10 presents the forecasting results of different datasets for hourly and daily CAQI levels, indicated as Model 3 (when Dataset 1 was used), Model 4 (when Dataset 2 was used) and Model 5 (when Dataset 3 was used).
2. Hourly and daily NO_2 and PM_{10} values were forecasted with ANNs, in order to calculate the hourly and daily CAQI Levels by using the formulas of Table 1. As input to the ANN, Dataset 2 was used (Model 6 in Table 10).
3. With the aid of ANNs the hourly and daily CAQI Levels were forecasted. In order to convert these numerical values to CAQI Levels (nominal values), the logical expressions of Table 4 were used. Table 10 present the forecasting results of different datasets for hourly and daily CAQI levels, indicated as Model 7 (when Dataset 1 was used), Model 8 (when Dataset 2 was used) and Model 9 (when Dataset 3 was used).
4. With the aid of DTs the hourly and daily CAQI Levels were forecasted. In Table 10 these are indicated as Model 10 (when Dataset 1 was used), Model 11 when Dataset 2 was used) and Model 12 (when Dataset 3 was used).

3.3 Comparison of results

On the basis of the overall results presented in Table 10 we can evaluate the forecasting performance for each daily and hourly model that was used. For this purpose, we used the "Percentage Agreement" (PA), in order to calculate the percentage of the times that the forecasted CAQI Levels were identical with the observed CAQI Levels, in order to evaluate the models. The best forecasting performance of daily CAQI Levels (PA=61%), was achieved when Model 9 (*Forecast Directly the CAQI Levels with Dataset 3*) was used with the aid of ANNs, with an increase of forecasting performance reaching 10% compared to the best forecasting performance when LR models were used, and with an increase of 4% compared to the best forecasting performance when DTs were used. The best forecasting performance of hourly CAQI Levels (PA=65%), was achieved when LR models were used, with an increase of 5% compared to the best forecasting performance when ANNs were used, and with an increase of 2% compared to the best forecasting performance when DTs were used.

Table 10. Comparison of the best forecasting performance of all Forecasting Models for Daily and Hourly CAQI Levels.

Model	Sub Model	Daily CAQI Levels		Hourly CAQI Levels	
		Percentage Agreement	Cohen's Kappa	Percentage Agreement	Cohen's Kappa
Model 1	-	49.54%	0.28	**65.10**%	**0.53**
Model 2	Forecasted CAQI - Factor4	51.05%	0.31	-	-
Model 3	2 lagged values	45.00%	0.14	57.00%	0.41
	8 lagged values	49.00%	0.22	55.00%	0.39
Model 4	-	53.00%	0.30	46.00%	0.25
Model 5	1 lagged value	58.00%	0.37	57.00%	0.41
	4 lagged values	55.00%	0.32	60.00%	0.45

Model 6	-	53.00%	0.30	44.00%	0.23
.Model 7	1 lagged value	41.00%	0.11	54.00%	0.37
	10 lagged values	52.00%	0.28	53.00%	0.34
Model 8	-	54.00%	0.32	46.00%	0.25
Model 9	1 lagged value	60.00%	0.41	55.00%	0.37
	2 lagged values	**61.00%**	**0.43**	54.00%	0.37
Model 10	1 lagged value	50.00%	0.23	63.00%	0.51
Model 11	-	54,00%	0,32	46,00%	0,26
Model 12	4 lagged values	57,00%	0,36	62,00%	0,47
	5 lagged values	57,00%	0,36	62,00%	0,48

Model Details:
1. Calculate the CAQI Levels from forecasted Daily/Hourly CAQI values by using regression models
2. Calculate the CAQI Levels from forecasted Hourly CAQI values by using regression models
3. Calculate the CAQI Levels by forecasting the CAQI values via ANNs, with Dataset 1
4. Calculate the CAQI Levels by forecasting the CAQI values via ANNs, with Dataset 2
5. Calculate the CAQI Levels by forecasting the CAQI values via ANNs, with Dataset 3
6. Calculate the CAQI Levels by forecasting NO_2 and PM_{10} via ANNs, with Dataset 2.
7. Forecast Directly the CAQI Levels via ANNs, with Dataset 1
8. Forecast Directly the CAQI Levels via ANNs, with Dataset 2
9. Forecast Directly the CAQI Levels via ANNs, with Dataset 3
10. Forecast Directly the CAQI Levels via Decision Trees, with Dataset 1
11. Forecast Directly the CAQI Levels via Decision Trees, with Dataset 2
12. Forecast Directly the CAQI Levels by via Decision Trees, with Dataset 3

4. Conclusions and further research directions

In the present paper the CAQI was used to calculate the pollution levels (hourly and daily) for four locations in Thessaloniki, Greece. On the basis of the results, the Agia Sofia station was chosen as the one for which forecasting models were developed and tested. LR models, ANNS and DTs were used in order to perform the forecasting of hourly and daily CAQI levels (i.e. nominal) and values (i.e. numerical). Different input datasets were used for that purpose, with a different number of lagged values. We observed that when the forecasted daily CAQI value was calculated on the basis of the forecasted hourly CAQI values, the performance was better in comparison to the one achieved via the direct forecast of the daily CAQI values. Moreover, if weighting factors were used in order to calculate the daily CAQI values, the forecasting performance improved. The best performance in terms of the Cohen's kappa was 0.53 for the hourly CAQI and 0.44 for the daily index levels, thus acceptable for operational purposes [19].

References

1. Nemours Organization, KidsHealth website: Ozone, Air Quality, and Asthma. (2010) http://kidshealth.org/parent/medical/allergies/ozone_asthma.html
2. Cohen, A.J., Anderson, R.H., Ostro, B., Pandey, K.D., Krzyzanowski, M., Kunzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet J.M. and Smith K.: The Global Burden of

Disease Due to Outdoor Air Pollution, J. of Toxicology and Env. Health, vol. 68(13-14), pp. 1301--307 (2005)
3. Vlachogianni A., Kassomenos P., Karppinen A., Karakitsios S., Kukkonen J.: Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki, Science of The Total Environment, 409(8), pp. 1559--1571 (2011)
4. Atakan K., Ayşe Betül O.: Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks, Expert Systems with Applications, 37 (12), pp. 7986--7992 (2010)
5. Tzima F., Mitkas P., Voukantsis D. and Karatzas K.: Sparse episode identification in environmental datasets: the case of air quality assessment, Expert Systems with Applications, 38(5), pp. 5019--5027 (2011).
6. Poupkou A., Nastos P., Melas D. and Zerefos Ch.: Climatology of Discomfort Index and Air Quality Index in a Large Urban Mediterranean Agglomeration, Water, Air, & Soil Pollution, 222(1), pp. 163--183 (2011)
7. Kassomenos P., Kelessis A., Petrakakis M., Zoumakis N. and Christidis T.: Characterizing the quality of the atmosphere over an urban complex, 11th Int. Conf. on Env.l Science and Technology, Chania, Crete, Greece, 3-5 Sept. , vol. 2, pp. 424--431 (2009)
8. Elshout, S. and Leger K.: Comparing Urban Air Quality Across Borders.(2007) http://www.airqualitynow.eu/download/CITEAIR-Comparing_Urban_Air_Quality_across_Borders.pdf
9. Gallant, S. I.: Perceptron-based learning algorithms, IEEE Transactions on Neural Networks, 1(2), pp. 179--191 (1990)
10. Kohonen, T.: Self-Organizing Maps, 2nd edn. Series in Information Sciences, Springer, Heidelberg (1997)
11. Hornik, K., Stinchcombe, M. and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, vol. 2, pp. 359--366 (1989)
12. Lippman, R.P.: An introduction to computing with neural nets, IEEE ASSP Mag. (1987)
13. Chelani, A.B., Gajghate, D.G. and Hasan, M.Z.: Prediction of ambient PM_{10} and toxic, metals using artificial neural networks, J. of Air and Waste Management Ass, vol. 52, pp. 805--810 (2002)
14. Ethem, A.: Introduction to machine learning. 2nd edn. Massachusetts Inst. of Tech. (2010)
15. Lakhmi, C.J. and Chee, P.L.: Advances in Intelligent Methodologies and Techniques, Horia-Nicolai Teodorescu, JunzoWatada, and Lakhmi C. Jain (eds.): Intelligent Systems and Technologies, SCI 217, Springer-Verlag Berlin Heidelberg (2009)
16. Kyriakidis, I., Karatzas, K. and Papadourakis, G.: Using Preprocessing Techniques in Air Quality forecasting with Artificial Neural Networks, Proceedings of the Fourth Int. ICSC Symposium on Information Technologies in Environmental Engineering, Thessaloniki, Greece, pp. 28--29 May, Springer Series: Environmental Science and Engineering. (2009)
17. Eun-Young Ji, Moon Y.-J., Kim K.-H., Lee G.-H.: Statistical comparison of interplanetary conditions causing intense geomagnetic storms (Dst<=-100 nT), Journal of Geophysical Research, 115, pp. 1--7 (2010)
18. Saini, L.M. and Soni, M.K.: Artificial neural network based peak load forecasting using Levenberg-Marquardt and quasi-Newton methods, Generation, Transmission and Distribution, IEE Proceedings, vol. 149(5), pp. 578 -- 584 (2006)
19. Voukantsis D., Karatzas K., Kukkonen J., Räsänen T. Karppinen A. and Kolehmainen M.: Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki, Science of the Total Environment, vol. 409, pp. 1266--1276 (2011)