Hybrid and Reinforcement Multi Agent Technology for real time air pollution monitoring

Andonis Papaleonidas[1], Lazaros Iliadis[2*],

[1, 2]Democritus University of Thrace , Department of Forestry & Management of the Environment & Natural Resources, 193 Pandazidou st., 68200 N Orestiada, Greece
Email:papaleon@sch.gr [1], liliadis@fmenr.duth.gr* [2]

Abstract. This paper describes the design and implementation of a modular hybrid intelligent model and system, for monitoring and forecasting of air pollution in major urban centers. It is based on Multiagent technologies, Artificial Neural Networks (ANN), Fuzzy Rule Based sub-systems and it uses a Reinforcement learning approach. A multi level architecture with a high number of agent types was employed. Multiagent's System modular and distributed nature, allows it's interconnection with existing systems and it reduces its functional cost, allowing its extension by incorporating decision functions and real time imposing actions capabilities.

1. Introduction

The problem of air quality especially in urban sites is a major and composite task. Nearly all of the air pollutants have been blamed by several epidemic and clinical studies for their direct or indirect involvement in the cause of several serious diseases (Brunekreef and Holgate, 2002) (Nafstad P. et al., 2003)(Lisabeth LD et al., 2008) after a long term exposure. The risk limits and the right of the citizens to have access in vital information related to air quality and to pollutants' concentration levels, have led to the establishment (by the European Commission) of specific safety limits, whereas public awareness services were also created (European Commission, 1990) (European Commission, 1992) (European Commission, 1996) (European Commission, 2000). Even platforms like app store, supply the public with applications that provide real time air quality data in mobile phones in various parts of major cities (diMobile, 2011) (Aratos, 2012). Due to the importance of the problem, various Soft Computing applications have been developed recently for air quality modeling or monitoring, whereas some of them also provide real time proposals when the pollutants' concentration is too high (Triantafilou A.G et. all, 2011), (Iliadis and Papaleonidas, 2009), (Wahab and Alawi, 2002), (Paschalidou et al., 2007), (Iliadis et al., 2007). These efforts use mainly ANN or Multi Agent Systems (MAS). However they only try to face the problem under a monolithic point of view of forecasting or monitoring.
The MAS described in this paper not only records data and presents the actual real time situation (in terms of air pollution) but it also tries to estimate the evolution of

the problem in a short term scale, by employing a hybrid approach that satisfies the needs of a modular wide architecture perceptive design (Estrin et al., 2002).
A main advantage of the proposed system is that it does not work as a stand alone application in a specific place. It rather uses a large number of independent agents of various types that are distributed around several points of an urban center. Each agent is assigned a distinct task depending on the pollutants that characterize the specific area. The agents interact and exchange messages in order to accomplish the task of air quality monitoring. The whole process is supported by the computational power of ANN that have the ability to estimate missing values and the decisions are taken by a fuzzy rule based model. The Reinforcement learning has been applied to enhance the learning ability of the system. It is an integrated holistic approach.

2. General System's Architecture

The system is based on the Jade multiagent platform (Bellifemine F. et al., 2007). Its main advantage is that it can operate under any environment and operation system regardless the processors and the network type. This is due to the fact that it uses Java for the construction of the agents and MySQL in the database mechanism which are both open source environments (under GPL license) and they both allow programming of modular and distributed credible applications of low cost. The system has been divided in five sub-systems with high cohesion level which are assigned distinct functions and as it is shown in the figure1. This architecture enhances the independence and the personal perspective of the comprising agents.

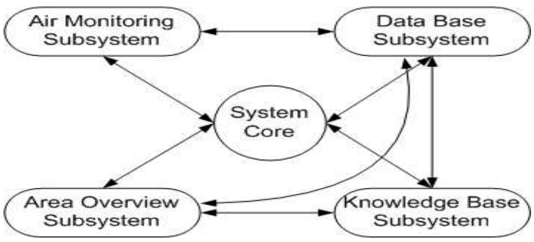

Fig.1. Overall system's architecture

2.1 General description of the subsystems

The Air Monitoring Subsystem (AMS) aims in capturing the instant perception of the environment and in transferring it to the system's core. It comprises of a set of typical air pollution sensors which record the actual measurements and they transfer them to the Sensor Intelligent Agents (SIA) that store them in the Data Base (DB) subsystem acting in a proxy mode. SIA are also responsible for the correct function of the typical sensors, the removal of the improper values, the arrangement of the new measurements' time interval and the update of the administrator for a potential malfunction. The structure of the AMS is seen in figure 2.

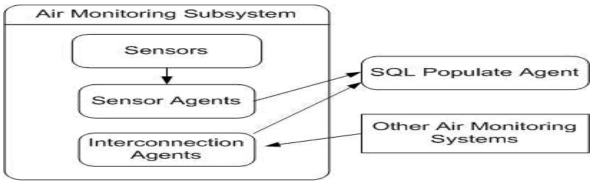

Fig.2. Air Monitoring Subsystem architecture

The next level comprises of the Interconnection Agents (IA) which allow the import of data from other information systems activated in the area under study. Given the high cost of buying and supporting several air pollution sensors, a mechanism was applied that allows the performance with non dedicated sensors. This is done by connecting the system with sensors of other systems aiming in data transfer. Also the IA allow the interconnection between systems located in different areas.
The Data Base Subsystem performs the management, storage and retrieval of data. Its structure can be seen in the following figure 3.

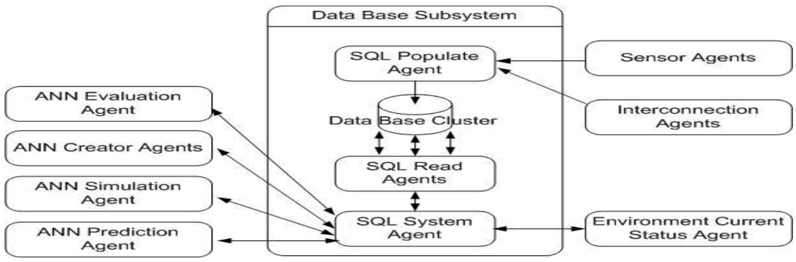

Fig.3. Data Base Subsystem architecture

The SQL Populate Agents (SQLPA) are the simplest ones. Their basic function is to receive the measurements from the SIA and the IIA, to check their validity and to store them after considering the Data Base schema. The SQL System Agent (SQLSA) connects the Data Base Subsystem with the rest, by receiving requests from the agents that require data. It decides to forward the request to the retrieval agents based on the number of queries and the load of the database. No other agent can access the actual data if the SQLSA do not approve. For the same reason, the logic of data base clustering with multiple data bases communicating with a series of SQL Read agents (SQLRA) was selected. In this way the performance of the system was enhanced. The necessity for this approach will be understood further in the next chapter.

3. The role of Reinforcement ANN

ANN employing the Reinforcement approach were developed and incorporated into the multi agent system in order to offer reliable estimation of values when the sensors were malfunctioning and the obtained data were unacceptable. Before referring to the ANN the whole process of data retrieval and storage must be clarified.

3.1. Data manipulation

The SQLRA undertake the task of retrieving data from the data base in order to answer in the request that was diverted by the SQL system agent. Technically, every SQLRA is a compiler which converts the requests from the descriptive language that was developed for the system to SQL query statements, in order to produce the desired data set. The requests are implemented according to the FIPA (Foundation for Intelligent Physical Agents) protocol (FIPA, 2002) and the answers are sent with a message of Inform type (Bellifemine F. et al., 2007). The text in the request message, informs the agent on the data that have to be retrieved and it comprises of a string that describes the data in the following form:

*1001d*dd*01;01h*Kor_SO2*01;*
*24h*mar_pm10*04;01f*pen_tair*03;11p*oin_o3*02;*

Command1. Sample of a retrieval command sent to an SQL Read Agent

The above sample command creates a text file, where each line contains 14 elements and corresponds to a record of the data base. This means that if the DB had only correct hourly measurements for the whole 2005, then the output file would have 365*24=8760 lines with totally 8760*14=122640 values.

3.2. Agents handling ANN storage

The Knowledge Base subsystem manages the development of the ANN which will be potentially used by the system, in order to forecast the evolution of a phenomenon. The structure of the KBS is shown in the following figure 4.

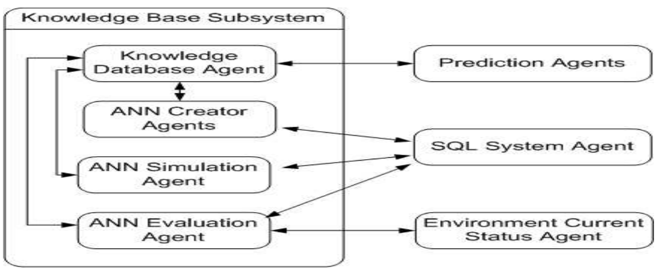

Fig.4. Knowledge Base Subsystem architecture

The Knowledge Base Agent stores all ANN used by the system, together with their characteristics and the *overall score* of each and decides which one of them will be used every time that a forecast is requested. The *characteristics* are obtained by rules determining in which cases each network can be used (e.g. other ANN is reliable for

the winter period, other for the summer). The choice of the proper ANN is done by the application of filters on the data. These filters consider the criteria like: a) The number of cases in which the network had the best estimation b) The average error of each c) The overall score of each one which is estimated during the network validation process by the Evaluation Agent (EVA). The network choice mechanism checks based on the characteristics of the available ANNs, which ones can be used in each specific case and from the emerging candidates it nominates the ANN with the highest score. The validation and assessment of the neural networks is done with the cooperation of the ANN Simulation Agent (ANNSA) and the EVA.
The score for every network is obtained based on a Reinforcement approach in combination with fuzzy logic. According to the Reinforcement technique every ANN receives merit or it is punished by raising or reducing its score, every time that the ANNSA is executed. The amount of penalty or benefit is determined by the employment of a fuzzy algebra model.
It is really important that based on the system's philosophy the ANN which are available in the knowledge base are not executed only when there is a request for data but they are re-adjusted every time that the actual perception of the environment is refreshed. The ANNSA runs all available networks after each new measurement obtained, regardless if this was done after a request or not.
In this point the Knowledge Base is informed on which network had the best performance per estimated value, so that the table with the *characteristics* is updated after the EVA estimates the new scores.

4. Fuzzy modeling

For each parameter under study four boundary values A, B, C, D were stored in the database. These values were used by two semi trapezoidal fuzzy membership functions (Iliadis, 2007) to define the linguistics (fuzzy sets) *high error, small error*, between the estimated by the ANN and the actual values of each feature (Iliadis et al., 2008). The numbers A and B correspond to error quite close to zero, whereas C and D are related to higher errors and their range depends on the range of the values of the actual feature under study. Even if someone picks a little higher numbers for A,B,C,D the slight change in the fuzzy sets does have any significant effect in the membership values. The design of the fuzzy sets has been done is a way that the system is flexible and adaptable. These membership functions can be seen in the following figure 5.

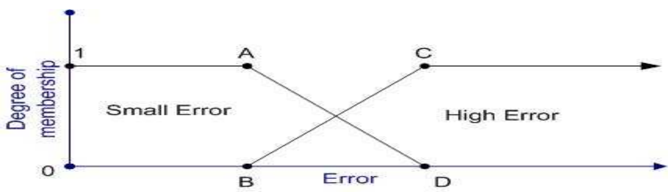

Fig.5. Semi Trapezoidal definition of small and high error

The degree of membership (DOM) of each ANN to every error fuzzy set is the score based on which it will be rewarded or punished. The DOM to the *small error* linguistic is added to the overall score of each ANN, whereas its DOM to the *high error* fuzzy set is subtracted. By using this flexible approach we avoid to change the overall score of an ANN in a rare case of an extremely bad or extremely good performance and thus the final score reflects the whole performance of the network.

5. Agents for the development and assessment of the ANN

The ANN Creator Agents (ANNCA) offer the capability of an automatic creation and initial assessment of an ANN. The system accepts as input by the user, the ever recorded width of the values of each considered parameter and potential restrictions which are also incorporated in the data base, the number of hidden neurons and the number of hidden sub layers, potential transfer functions and the minimum score under which the networks are not considered credible.
The ANNCA follow a trial and error approach by offering a range of values to the above parameters. In this way they automatically construct multiple ANN which cover the set of all potential combinations. This approach leads to the creation of a vast number of ANN in a combinatorial explosion mode and thus computational power is required. As it will be clarified in the results section where the system was executed for a specific case, a huge number of 2200 networks were automatically created. A threshold value of R^2 was used as the minimum criterion of acceptance for a network to be recorded in the knowledge base and it will not be subjected to continuous assessment as it was mentioned before. The Air Overview sub system is responsible for the visualization of the data and of the forecasts and it comprises of two agent types, namely the Area Monitor Agents and the Prediction Agents Its structure is shown in the following figure 6.

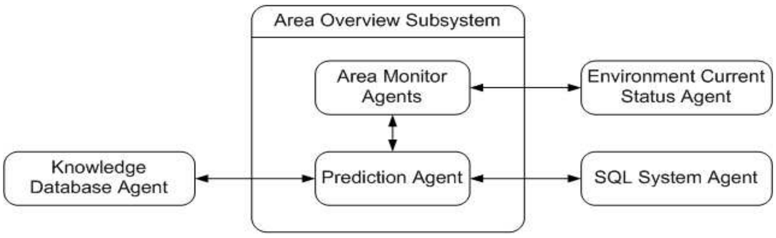

Fig.6. Structure of the Air Overview subsystem

The Area Monitor Agents (AMA) constitute an awareness area of the system. By the term area the system understands any group of data and forecasts that it can manage and it can be an air pollution measurement station, a hospital service that monitors the evolution of an air pollutant concentration or a wider area represented by dozens measurements stations, or a simple user that needs to monitor specific data in certain locations. It other words it can combine any number of features to construct a logical area of study. The visualization of the situation in a specific area can be done by using fuzzy sets (Linguistics) of the type "*Low*", "*High*" "*Critical*". This is shown in figure 7.

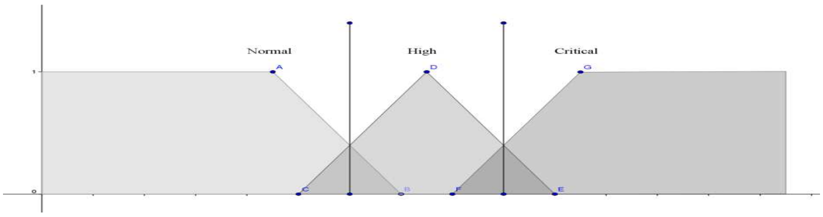

Fig.7. Fuzzy determination of the situation in a specific area

The Prediction Agent (PREA) is responsible to provide forecasting when an area asks for one. When it receives a request for a forecast it communicates with the Knowledge Database Agent (KDA) and asks the nominated ANN to perform estimation for a specific parameter. Then, the PREA asks the SQL System Agent to provide the data required by the network in order to perform forecast. In the rare case of an ANN that cannot function due to the fact that the SQL system Agent did not find the necessary data, the Prediction agent will contact the KDA again and it will require the next available network. This process repeats in an iterative manner until a proper solution to be found.

6. System Core Agents

The last subsystem is the System Core (SYSCO). The SYSCO handles the general attendance of the system. It provides access rights to all agents that try to connect to the platform and it offers an overall image for the situation of the system and for the area under study. It also informs the administrator for potential malfunctions and connection failures of an agent.

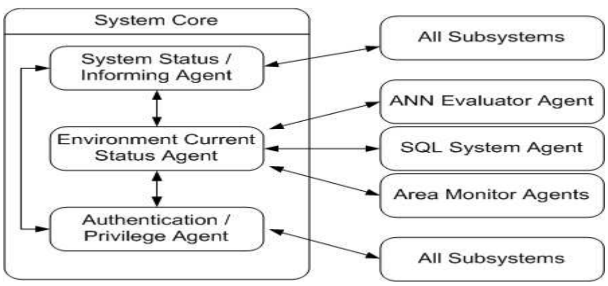

Fig.8. System Core architecture

The System status Agent monitors if the system performs functional tests and informs the administrator accordingly. It communicates with all of the system's agents to

check their functional status. It also keeps the list with the minimum available agents required for the proper operation of each subsystem.
The Environment Current Status Agent is responsible for the temporary storage of the current status of all factors monitored by the system. Its usefulness is in the fact that it can update all agents on the current status of the environment without using the database.
Finally the Authentication/Privilege Agent defines the level of information for which each agent will have access. It does not allow the access in the database of Sensor agents which are not declared in the system and in the database schema and it defines the credentials of the Area Agents when they are not public.

7. Results and Discussion

In order to check the proper function and the performance of the system, a preliminary testing was performed by using actual historic data. The target was the short term forecast and also the current, plus 1,2,3 and 6 hours ahead estimation of Ozone (O_3) in the measurements station "*Athinas*" located in the center of Athens. Actual hourly data records were used coming from 11 surrounding stations (9 of them measuring air pollutants and 2 meteorological data) located in the center of Athens. The data were measured on a 24 hours basis for the whole annual period and only when the measuring stations functioned properly. When a station was malfunctioning the obtained value was defined to be as high as -9999.00 and it was never considered in any analysis. The data come from the website of the Greek ministry of environment and Climate change (Minenv, 2012) and they are historical data from the establishment of each station until year 2010. The stations used and the date of their first operation plus the parameters employed can be seen in the following table 1.

Table 1. Description of measurement stations

Name	**CODE**	**Type**	**Established**	**Data**
Ag Paraskevis	AGP	Air Pollution	2000	O_3,NO, NO_2
Amarousion	MAR	Air Pollution	1987	O_3,NO,NO_2, CO
Peristeriou	PER	Air Pollution	1990	O_3,NO,NO_2, CO, SO_2
Athinas	ATH	Air Pollution	1984	O_3,NO,NO_2, CO, SO_2
Pathsion	PAT	Air Pollution	1990	O_3,NO,NO_2, CO, SO_2
Aristotelous	ARI	Air Pollution	1994	NO, NO_2
Geoponikis	GEO	Air Pollution	1984	O_3,NO,NO_2, CO, SO_2
Peiraias	PIR	Air Pollution	1984	O_3,NO,NO_2, CO, SO_2
N Smyrnis	SMY	Air Pollution	1984	O_3,NO,NO_2, CO, SO_2
Pendelis	PEN	Meteorological	1999	Temp, Wspeed, SunTime, Wdirection, Radiation, RH
Thiseiou	THI	Meteorological	1985	Temp, Wspeed, SunTime, Wdirection, Illumin, RH

The Ozone for the "*Athinas*" station was used as output whereas the data of the rest 10 stations except of their Ozone values were used as input. The assessment process of the system included the following steps:
A) Creation of the Database schema according to the data management system's needs. The following table 2 presents the structure of the database and the number of records.

Table 2. Description of the Database

Table	Records	Type	Size
array01h	236,754	InnoDB	82.6 MB
day_names	7	InnoDB	16.0 KB
factors01h	51	InnoDB	32.0 KB
measurements01h	237,168	InnoDB	15.0 MB
values01h	12,196,690	InnoDB	1.2 GB

InnoDB engine was selected not only because it is the default storage engine for MySQL, as of MySQL 5.5, providing a transaction safe environment with rollback and crash-recovery capabilities in order to protect data but as at the same time provides high speed data access (MySQL, 2012).
B) Creation of an Interconnection Agent which aimed in reading the text files in order to send the data in the proper form to the Populate agent that would store them in the database.
C) Creation of an ANN Creator Agent which automatically implemented the required networks. It was asked from the ANNCA to construct all potential networks' combinations, using data at least from one meteorological station and two air pollution ones, until all combinations between the stations described in table 1 were done. The above combinations produced 741 ANN. For the best control of the system it was asked by each network to train and test itself using a variable number of hidden neurons, providing the system with three values (0.5, 0.75, 1) which would be multiplied with the number of the proposed hidden neurons in order to produce their actual number. The ANN Creator Agent produced automatically the data files used for training and testing. These files contained values related to each factor from the first year of the station's establishment until 2009. Each of the 741 ANN asked the SQL System Agent the creation of a measurements' file using the right format. To show the computational complexity we must add that totally there were 10,132,450,485 SQL queries performed and the whole process used a cluster with 8 dataset servers (based on average pc configuration). For all of the process it took less than 7 days. This was actually the reason for choosing the solution of clustering in the data base. After the implementation of the ANN the training and their registration in the Knowledge Database Agent followed.
D) The last step in the system's testing was the use of the 2010 data for the total simulation of its function based on the knowledge base created in the previous step.
A Simulation Agent was created which ran the system with hourly values from the 01/01/2010 until 31st of December 2010. Also the ANNSA and the ANN Evaluation

Agent were used. Totally the system was executed 8,760 times (365*24) and managed to obtain results for the 8,738 of them. For the 22 cases there was a lack of data due to simultaneous interrupt of function of both meteorological stations. In 167 cases it was not possible to use the ANN Evaluation Agent due to the fact that for these hours the "Athinas" station was not functioning and thus we could not estimate the ANN error. As it was mentioned before in this chapter the ANN had 5 output values whereas the error during simulation and testing of the system was estimated only for the current value. When the system will be fully executed the error will be estimated separately for each output.
The following table 3 presents the results for the simulation of the year 2010.

Table 3. Results for the simulation of the year 2010

Output	**R**	**R^2**	**Mean Square Error (MSE) (Root MSE)**
Current estimation	0,9146	0,8365	139,712 (11.820)
1 hour forecast	0,9057	0,8203	154,735 (12.439)
2 hours forecast	0,8909	0,7937	173,413 (13.169)
3 hours forecast	0,8754	0,7663	193,302 (13.903)
6 hours forecast	0,8441	0,7125	230,007 (15,166)

The system in this phase covers a wide range of requirements and potentials, whereas in its direct extensions it will have Agents that will materialize mechanisms of retraining the ANN when they appear to have a reduced performance and a continuous reduction in their overall score. In this way the ANN will be automatically adjustable without the interference of the user in systemic changes (e.g. close a specific area to automobiles) which change the weight of the used parameters in an ANN. Additionally a complementary decision support subsystem could be implemented that will use rules and it will process the available data to impose actions that will improve the situation in the environment. This of course (in an iterative manner) will readjust the system under the new improved environment.

References

A.G. Triantafyllou, S. Skordas, Ch.Diamantopoulos, E. Topalis (2011) «The New Dynamic Air Quality Information System and its Application in West Macedonia, Hellas», 4th Environmental Conference of Macedonia, 18 – 20 March 2011

Abdul-Wahab, S.A., Al-Alawi, S. (2002) "Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks". Environmental Modelling & Software 17 (3), 219–228.

Aratos, (2012) Ozone Index Saves Earth By Aratos Technologies S.A. http://itunes.apple.com/us/app/ozone-index-saves-earth/id468073482?mt=8 (Valid at 15/04/2012).

B.Brunekreef, S.Holgate (2002) “Air pollution and health” The Lancet, Volume 360, Issue 9341

Council Directive 90/313/EEC of 7 June 1990 on the freedom of access to information on the environment, Official Journal L 158 , 23/06/1990 P. 0056 - 0058

Council Directive 92/72/EEC of 21 September 1992 on air pollution by ozone, Official Journal L 297 , 13/10/1992 P. 0001 - 0007

Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management, Official Journal L 296 , 21/11/1996 P. 0055 - 0063

Council Directive COM(2000) 613 final, directive of the European parliament and of the council relating to ozone in ambient air

diMobile, (2011) Ozone Position By diMobile http://itunes.apple.com/ke/app/ozone-position/id439508448?mt=8 (Valid at 15/04/2012).

Estrin D., Culler D., Pister K. and Sukhatme G. (2002) “Instrumenting the Physical World with Pervasive Networks”, IEEE Pervasive Computing, 1 (1) P. 56-69.

F. Bellifemine, G. Caire, and D. Greenwood (2007). Developing Multi-Agent Systems with JADE. John Wiley & Sons. New York, USA.

FIPA, (2002), FIPA ACL Message Structure Specification, http://www.fipa.org/specs/fipa00061/index.html (Valid at 15/04/2012).

Iliadis L, S. Spartalis, S. Tachos, (2008) "Application of fuzzy T-norms towards Artificial Neural Networks’ evaluation: A case from wood industry” Journal Information Sciences (Informatics and Computer Science Intelligent Systems Applications) Volume 178, Issue 20, P. 3828-3839

Iliadis L., Spartalis S., Paschalidou A., Kassomenos (2007). Artificial Neural Network Modeling of the surface Ozone concentration International Journal of Computational and Applied Mathematics vol 2, no 2, P.125-138

Iliadis, L. (2007) “Intelligent Information systems and Applications in risk estimation”. Stamoulis A., publishing, Thessaloniki, Greece.

Iliadis L. and Papaleonidas A., (2009) “Intelligent Agents Networks Employing Hybrid Reasoning: Application in Air Quality Monitoring and Improvement”, EANN 2009, CCIS 43, P. 1–16.

Lisabeth LD, Escobar JD, Dvonch JT, Sánchez BN, Majersik JJ, Brown DL, Smith MA, Morgenstern LB. (2008) “Ambient air pollution and risk for ischemic stroke and transient ischemic attack” , Annals of Neurology, DOI:10.1002/ana.21403

Ministry of Environment, Energy & Climate Change, (2012) Air Pollution Measurements. http://www.ypeka.gr/Default.aspx?tabid=495&language=el-GR (Valid at 15/04/2012).

MySQL, (2012) Chapter 14. Storage Engines. http://dev.mysql.com/doc/refman/5.0/en/storage-engines.html (Valid at 15/04/2012).

Nafstad P, Håheim LL, Oftedal B, Gram F, Holme I, Hjermann I, Leren P. (2003) “Lung cancer and air pollution: a 27 year follow up of 16 209 Norwegian men”, Thorax 2003;58:1071-1076.

Paschalidou A., Iliadis L., Kassomenos P., Bezirtzoglou C. (2007). Neural Modeling of the Tropospheric Ozone concentrations in an Urban Site Proceedings of the 10th International Conference Engineering Applications of Neural Networks P. 436-445