Fuzzy Classification of Cyprus Urban Centers Based on Particulate Matter Concentrations

Nicky Gkaretsa, Lazaros Iliadis, Stephanos Spartalis, Antonios Kalampakas

Fuzzy Classification of Cyprus Urban Centers based on Particulate Matter concentrations

Nicky Gkaretsa[1], Lazaros Iliadis[1], Stephanos Spartalis[2] and Antonios Kalampakas[2]

[1]Democritus University of Thrace, Pandazidou 193, Orestias, Greece
Email: garetsa.nicky@gmail.com, liliadis@fmenr.duth.gr
[2]Democritus University of Thrace, Xanthi, Greece
Email: sspart@pme.duth.gr, akalampakas@gmail.com

Abstract. This research aims in the design and implementation of a flexible Computational Intelligence System (CIS) for the assessment of air pollution risk, caused by PM_{10} particles. The area of interest includes four urban centers of Cyprus, where air pollution is a potential threat to the public health. Available data are related to hourly daily measurements for 2006, 2007 and 2008. This Soft Computing (SC) approach makes use of distinct fuzzy membership functions (FMFs) in order to estimate the extent of air pollution**.** The CIS has been implemented under the MATLAB platform. Some interesting results related to each city are analyzed and useful outcomes concerning the seasonality and spatiotemporal variation of the problem are presented. The effort reveals the severity of air pollution. Risk is estimated in a rather flexible manner that lends itself to the authorities in a linguistic style, enabling the proper design of prevention policies.

Keywords: Z, S, Pi, Gama, Exponential membership functions, Fuzzy Classification, Particulate Matter air pollution

1 Introduction

The presence of any type of pollutants, noise or radiation in the air, can have a potential harmful effect in the health of all living creatures and might make the environment improper for its desired use. Globally, air pollution is considered responsible for a high number of deaths and it also causes several deceases of the breathing system, mainly in urban centers [2]. PM10 are floating particles that have a diameter higher than 0.0002 μm and smaller than 10μm [8], [13]. Immediate actions have to be taken, as studies in the USA have shown that a slight increase of the PM levels only by 10 μg/m3, can increase mortality by 6%-7% [18].

1.1 Literature Review

Numerous papers describing various Soft Computing approaches have been published in the literature lately. Olej, et al., 2010 [16], have developed a FIS (Mamdani) towards air pollution modeling. Garcia, et al., 2010 [7] used a neural system and Artificial Neural Networks (ANNs) for the estimation of air quality related to O3. Iliadis and Papaleonidas, 2009 [12], have developed a distributed multi agent network, employing hybrid fuzzy reasoning for the real-time estimation of Ozone concentration. Aktan and Bayraktar, 2010 [3] used ANNs in order to model the concentration of PM10. Hooyberghs et al., 2005 [9], propose ANN models as tools that forecast average daily PM10 values, in urban centers of Belgium, whereas Iliadis et al., 2007 [11], have done the same forecasting effort for O3 in Athens. A similar research is reported in the literature for Chile by Dı´az-Robles et al., 2009 [5]. Thomas and Jacko, 2007 [20], have conducted a comparison between the application of Soft Computing and typical statistical regression in the case of air pollution. Finally, Mogireddy et al., 2011 [15] and Aceves-Fernandez et al., 2011 [1], have used Support Vector Machines towards air pollution modeling.

1.2 Methodology

As it has already been declared, this paper presents a Soft Computing approach towards the assessment of air pollution levels in Cyprus, by introducing specific fuzzy sets. Soft Computing is an umbrella including neural networks, fuzzy logic, support vector machines and their hybrid approaches [14], [4]. Two types of exponential fuzzy membership functions and also *S, Γ* (Gama), Pi and Z FMFs were applied to determine the linguistics that characterize the severity of the problem in each case. It should be mentioned that this is the first time that such a wide range of FMFs are employed for the case of air pollution with actual field data obtained from urban centers.

1.2.1. Fuzzy membership functions

Fuzzy Logic (FL) is a universal approximator of real world situations. Several researchers use FL towards systems modeling [18]. The *Z, S, Pi* spline-based FMFs are named after their shape. They are given by the following functions 1 2 and 3 respectively. Function 4 stands for the Gama, denoted after the Greek letter *Γ*, where the exponentials *Ex1* and *Ex2* are given by functions 5 and 6.

$$f(x;a,b)=\begin{cases} 1, & x\le a \\ 1-2\left(\dfrac{x-a}{b-a}\right)^2, & a\le x\le \dfrac{a+b}{2} \\ 2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2}\le x\le b \\ 0, & x\ge b \end{cases} \qquad (1)$$

$$f(x;a,b)=\begin{cases} 0, & x \le a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a \le x \le \dfrac{a+b}{2} \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2} \le x \le b \\ 1, & x \ge b \end{cases} \qquad (2)$$

$$f(x;a,b,c,d)=\begin{cases} 0, & x \le a \\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a \le x \le \dfrac{a+b}{2} \\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2} \le x \le b \\ 1, & b \le x \le c \\ 1-2\left(\dfrac{x-c}{d-c}\right)^2, & c \le x \le \dfrac{c+d}{2} \\ 2\left(\dfrac{x-d}{d-c}\right)^2, & \dfrac{c+d}{2} \le x \le d \\ 0, & x \ge d \end{cases} \qquad (3)$$

$$\mathrm{f(X,a)}=\begin{cases} 0 & \text{if } X \le a \\ \dfrac{k(x-a)^2}{1+k(x-a)^2} & \text{if } X \,\phi\, a \end{cases} \qquad (4)$$

$$\mathrm{f(X,a,b)}=\begin{cases} e^{-\left(\frac{M-X}{a}\right)^2} & \text{if } X \le M \\ e^{-\left(\frac{X-M}{b}\right)^2} & \text{if } X \,\phi\, M \end{cases} \qquad M=\frac{(a+b)}{2} \qquad (5)$$

$$f(X,a,b) = \begin{cases} e^{-\left(\frac{X-C_l}{2Wl}\right)^2} & \text{if } X \pi C_l \\ e^{-\left(\frac{X-C_r}{2Wr}\right)^2} & \text{if } X \phi C_r \\ 1 \text{ in any other case} & \end{cases} \qquad (6)$$

MATLAB has already built in code for the implementation of *Z*, *S* and *Pi* whereas the code for *Γ* and the two exponential functions have been developed in the form of ".m" executable files under the MATLAB platform. It must be clarified that in the case of the *Z, S FMFs*, parameters *a* and *b* locate the extremes of the sloped portion of the curve, whereas for the function *Pi, a* and *d* locate the "feet" of the curve and *b* and *c* locate its "shoulders" [17]. Finally, in function 6 (Ex2) the parameters *wleft* (*wl*), *cleft* (*Cl*), *cright Cr*), *wright* (*wr*), must be positive numbers and their values must be chosen by the user, following the constraint that *cl*<*cr*

Actual MATLAB code for the gamamf.m file:

```
function [ y ] = gamamf (x, params)
%   gamaMF(X, PARAMS) returns an array with the
degrees of membership
%  for an input vector X
%  params=[X0 X1] is a vector with 2 elements
determining the break points of the function
if nargin ~= 2,
error ('Two Parameters are required by
gamaMF.');
elseif length(params) < 2,
error ('gamaMF requires at least two parame-
ters.');
end
x0 = params(1); x1 = params(2);
y = zeros(size(x));  % Creates table Y
index1 = find(x <= x0);   % If X is less than Xo
or equal
if ~isempty(index1),
y(index1) = zeros(size(index1)); % τότε βάζει 0
end
index2 = find(x > x0);    % If X is greater than
31
if ~isempty(index2),
```

```
y(index2) = (x1*(x(index2)-x0).^2)/(1+x1*(x(in-
dex2)-
                     x0).^2);
end
end
```

2 Data and Area of Research

This research effort presents the implementation of a soft computing system (SCS) that is based on fuzzy logic (FL). More specifically *S, Γ, Pi* and *Z FMFs* have been employed in order to determine the proper linguistic that corresponds to the levels of air pollution. The input data include 16,000 data vectors for the four main cities of Cyprus namely: Larnaka, Lemesos, Lefkosia (Nicosia) and Pafos. The field data are related to the period July 2006 to July 2008. Each record contains the following fields: *Date and time*, *Time*, *Day*, *PM10* hourly concentrations (μg/m3), *SEA* (a binary index related to the seasonality of the case), *DoW* (direction of the wind), sin(HoD) and *cos*(*HoD*) (sine and cosine functions related to the effect of the hour of the day) (Ziomas et al, 1995) [21], PM10-24 (average *PM10* related to 24 hours), *T* (Surface Temperature) and *RH* (relative humidity). The PM10 concentrations were measured by using the Tapered Element Oscillating Microbalance device and the Filter Dynamics Measurement system. In order to overcome the variations in the magnitude of the data, they were normalized (standardized to zero) by employing the following function 7:

$$Z = \frac{X - \mu}{\sigma} \tag{7}$$

where μ is the average value and σ is the standard deviation [6].

The output of the developed FIS comprises of the fuzzy linguistic values Normal, Alert, Alarm which are related to the level of air pollution caused by PM10 and their assigned fuzzy membership values (FMV). European Union (EU) considers 50μg/m3 as the limit for the acceptable maximum daily concentration of PM10 whereas the maximum acceptable average annual boundary is 40μg/m3. Also EU has established the average daily values of 90 μg/m3 and 110 μg/m3as the Alert point and the Alarm limit respectively. The employment of specific numbers as the boundaries between Normal, Alert and Alarm are not rational especially when the measured values might not be very accurate. For example how can we accept the fact that 50μg/m3 are satisfactory for the concentration of PM10 whereas the 50.001 μg/m3 are not. On the other hand the main advantage of the fuzzy linguistic model introduced by this research is the fact that it is flexible and innovative, classifying the conditions of each case based on a real value (in the interval [0,1]) that specifies the degree of belonging to the proper linguistic [10].

The minimum and maximum concentrations of the *PM10* for the period 2006-2008 were input to the MATLAB fuzzy toolbox. Based on the range of these values, the parameters *a, b* for the *Z* and *S* functions and the actual values of the *parameters a, b, c, d* for the Pi FMF were determined automatically.

3 Results

The following table 1 is a small sample of the hourly classification performed for Lemesos Cyprus, on the 10th of July 2006 based on the *PM10* concentrations. It is interesting that there is a continuous Alert situation after eight o'clock in the morning and at 12 o'clock there is an Alarm signal. The situation goes back to Normal after 18:00 whereas there is a continuous Alert situation in the mean time. A classification for Lefkosia on the 12th of October 2006 reveals two Alarm situations, one at 7:00 in the morning and one at 18:00 in the afternoon. Pafos has also been assigned Alert Linguistic from 9:00 in the morning till 17:00 and also two Alarm signals from 18:00 till 19:00 for the 13th of August 2006.

Table 1. Sample of hourly classification for the PM_{10} (Lemesos July 06)

City of Lemesos	**Pollutant**	**Degrees of Membership to the three Linguistics**			
Date-Time	PM_{10}	**Normal Z-FMF**	**Alert Pi-FMF**	**Alarm S- FMF**	**Corre-sponding Linguistic**
10/7/2006 1:00	8.56	0.98	0.00	0.00	Normal
10/7/2006 2:00	47.05	0.34	0.01	0.00	Normal
10/7/2006 3:00	9.50	0.97	0.00	0.00	Normal
10/7/2006 4:00	45.22	0.38	0.00	0.00	Normal
10/7/2006 5:00	79.53	0.00	1.00	0.00	Alert
10/7/2006 6:00	98.71	0.00	0.91	0.04	Alert
10/7/2006 7:00	25.84	0.79	0.00	0.00	Normal
10/7/2006 8:00	90.51	0.00	1.00	0.00	Alert
10/7/2006 9:00	82.28	0.00	1.00	0.00	Alert
10/7/2006 10:00	76.54	0.00	0.98	0.00	Alert
10/7/2006 11:00	68.71	0.04	0.79	0.00	Alert
10/7/2006 12:00	**133.82**	**0.00**	**0.00**	**0.85**	**Alarm**

Table 2. Alarm cases for a period of 12 months in Larnaka

Month 2007	**S**	**Γ**	**Exp1**	**Exp2**	**Month 2007**	**S**	**Γ**	**Exp1**	**Exp2**
Jan	4	7	11	12	July	24	4	36	42
Febr	18	13	24	27	August	19	20	17	19
March	15	30	19	21	Sept	30	12	46	50
April	15	49	23	24	Oct	90	64	116	124
May	27	36	85	89	Nov	59	29	81	90

June	29	22	38	38	Dec	11	11	16	20

Table 3. Alarm cases for 6 months of 2006 in Larnaka

Month 2006	S	Γ	Exp1	Exp2
July	7	12	13	14
August	10	10	22	23
Sept	29	11	38	40
Oct	9	34	13	13
Nov	3	58	5	6
Dec	11	30	14	14

Table 4. Aggregation of the hourly Alarm situations in Larnaka for two years

Year	Number of Alarms (S)	Number of Alarms (Γ)	Number of Alarms (Exp_1)	Number of Alarms (Exp_2)
2007	381	337	512	556
2008	388	480	452	476

It is interesting to compare the classifications for two different cities for two different seasons (summer and winter).

An interesting finding is that in a period of 18 months, the Exponential2 function offers the highest number of Alarms in 10 cases (55.5%.) The exponential1 is second in the frequency of Alarms in 55.5% of the classificationς, whereas the Γ FMF is first in five cases (27.75%).

From table 4 it is concluded that the frequencies of the total hourly worst case situations in Larnaka are more or less the same for 2007 and 2008, regardless the FMF.

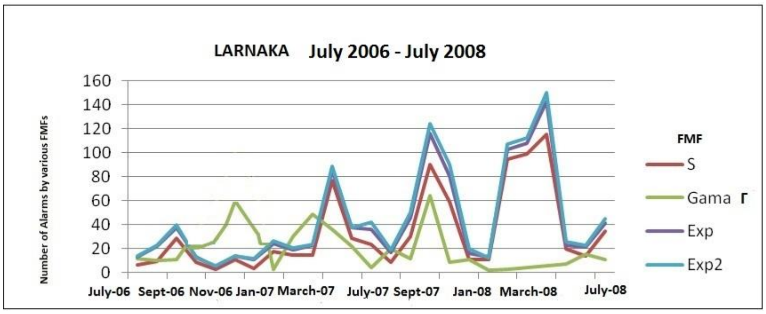

Fig. 1. Evolution of Alarm Signals for Larnaka with various FMFs

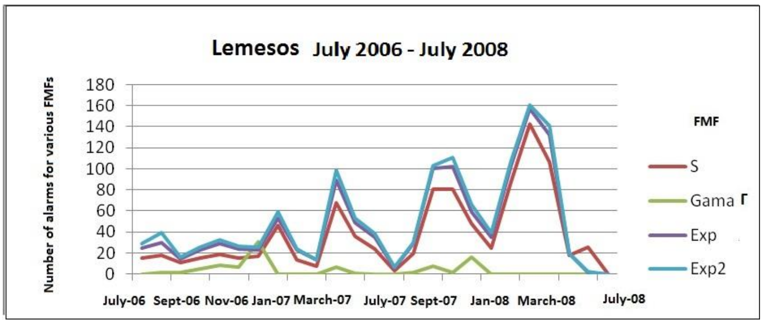

Fig. 2. Evolution of Alarm Signals for Lemesos with various FMFs

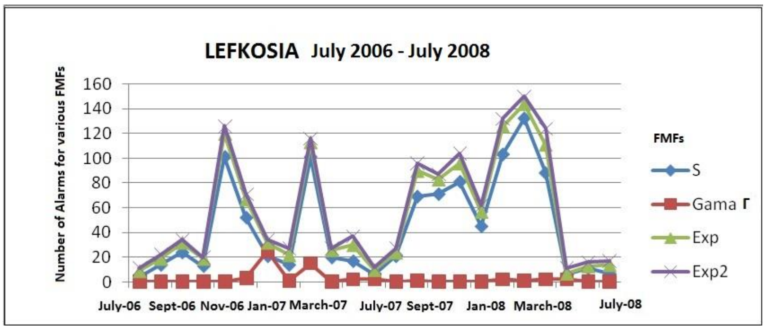

Fig. 3. Evolution of Alarm Signals for Lefkosia with various FMFs

It is obvious from the figures 1,2,3,4, that all of the functions except from the Gama are moving in a parallel mode. The Exp2 FMF always assigns the highest number of Alarms. However (with the exception of the Γ function) the frequency of the Alarms does not have big differences from one membership function to the other. Thus the Exp2 can be employed when the civil protection authorities wish to apply a quite strict policy. Otherwise the S or the Exp1 FMFs can offer a good alternative approach. Another finding is that the Γ function in most of the cases underestimates the risk and it offers rarely only a small number of Alarms. Only for the period October 2006 to January 2007 for Larnaka we have a peak in the number of Alarms introduced by the Γ FMF, which contradicts with the rest FMFs. Thus, the Gama function is not the proper one for this problem.

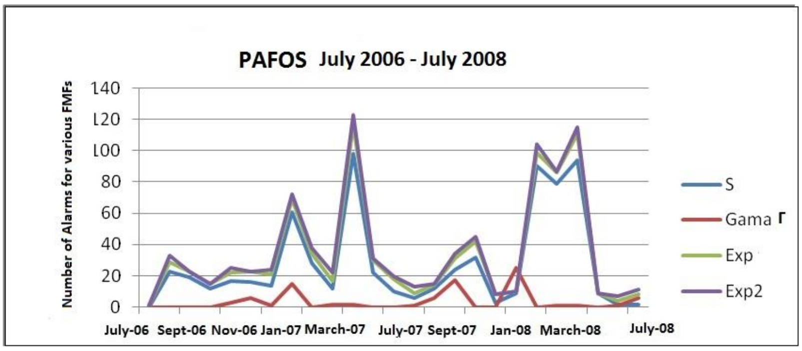

Fig. 4. Evolution of Alarm Signals for Pafos with various FMFs

3.1 Comparative analysis

The following figures 5, 6 represent a graphical display of the classes Normal, Alert, Alarm for Larnaka and for Lemesos for the same winter period. Number 1 stands for Normal, Number 2 for Alert and number 3 for Alarm.

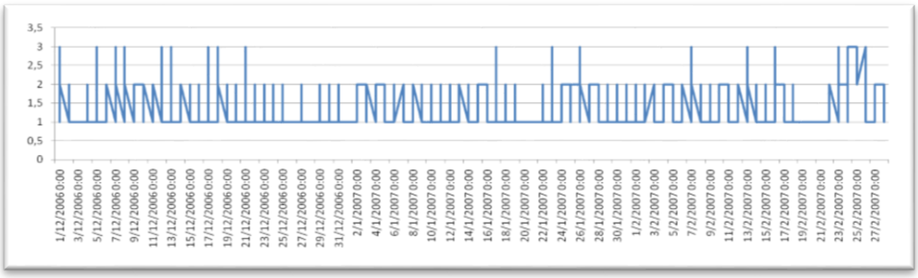

Fig. 5. Classification in 3 classes for Larnaka (December 06-February 07)

It is obvious from figures 5 and 6, that Lemesos has by far the most Alarm and the fewest Normal signals, whereas Larnaka is in an Alert state most of the times and more often than Lemesos.

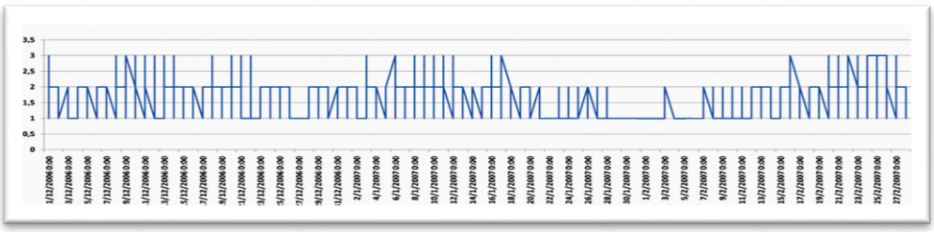

Fig. 6. Classification in 3 classes for Lemesos (December 06-February 07)

The following table 5 presents the Alerts frequency distribution on a daily basis, for *Larnaka*, *Lemesos*, *Lefkosia* and *Pafos* for the period (July 2006 – July 2008). It is clearly shown that in Larnaka the highest total number of Alerts has been obtained for Monday followed by Friday and Tuesday. In Lemesos the corresponding ranking was Tuesday, followed by Monday and Friday. In Lefkosia, again the same days are taking the lead. Friday has the most total number of Alerts, followed by Tuesday and Monday. Finally in Pafos Monday is first, followed by Sunday and Friday whereas Tuesday is very close only with a difference of two cases. It is clearly shown that the three most risky days for all cities (with a small differentiation in the order) are Monday, Friday and Tuesday.

Table 5. Alerts on a daily basis (2006-2008) for four Cyprus cities

	Daily distribution of Alarms for Lemesos						
Year	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
2006	21	20	13	18	13	8	2
2007	64	71	55	67	74	57	60
2008	62	80	49	47	56	61	53
TOTAL	147	**171**	117	132	143	126	**115**
	Daily distribution of Alarms for Lefkosia						
Year	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
2006	4	2	9	7	20	9	4
2007	87	97	60	93	130	59	48
2008	78	110	68	75	74	72	69
TOTAL	169	209	137	175	**224**	140	**121**
	Daily distribution of Alarms for Pafos						
Year	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
2006	10	14	16	24	13	6	4
2007	53	37	30	48	65	25	64
2008	54	53	25	24	28	52	46
TOTAL	**117**	104	**71**	96	106	83	114
	Daily distribution of Alarms for Larnaka						
Year	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
2006	14	12	12	10	18	2	1
2007	69	45	43	43	78	45	58
2008	82	76	43	35	49	44	59
TOTAL	**165**	133	98	88	145	91	**118**

The following figure 7 provides clearly, additional arguments to support the conclusions discussed above. Again it is shown that Monday, Friday and Tuesday are the most risky days.

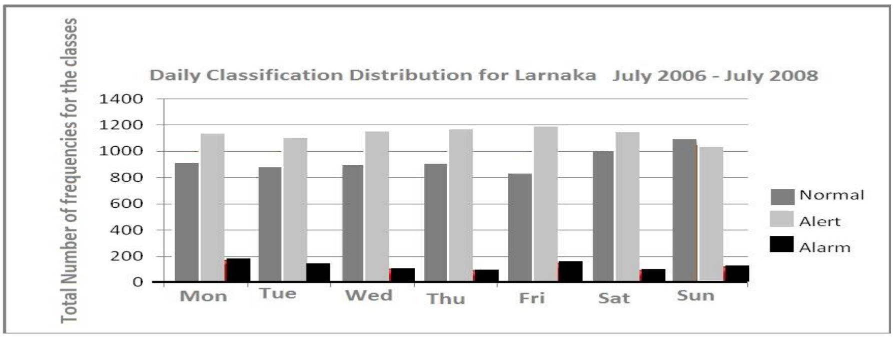

Fig. 7. Classification distribution in 3 classes for Larnaka

Also in figure 7 it is shown that in most of the cases, there is an "Alert situation" regardless the day of the week. Sunday and Saturday have the most Normal days, followed by Thursday.

4 Discussion - Conclusions

A first certain conclusion is that the situation in Cyprus regarding the PM10 concentration is becoming worse every year. A first glance at the graphs 1,2,3,4 reveals that the year 2008 is by far the worst compared to 2006 and 2007. Also it is obvious that there is a continuous and rapid increase of the problem which is detected by the huge differentiations between the peaks. The worst year with the highest number of Alarms is 2008 for Larnaka (150), Lemesos (160), Lefkosia (155). Only for Pafos May 2007 is the most risky with 120 Alarms but the peak of February 2008 is very close with 106. The most dangerous seasons are either October to December or February to April.

Concluding it can be suggested that the problem of the particulate matter concentration in the urban centers of Cyprus is quite serious and the Alert cases are very common, whereas the Alarms are quite frequent. Also it has been shown that the problem is seasonal and the day or the month, play a very serious role. This of course is due to human activities or weather conditions. The S function is a little bit more optimistic, giving the least number of Alarms, whereas the Exp2 FMF is the most strict one.

The Gama function has a very low estimation of the risk (the problem is underestimated significantly) and it is differentiated from all other FMFs that have very similar behavior.

Future research will be the conversion of the CIS to a real time one. Working in a real time mode and offering real time classifications, it can enable the effective handling by the local authorities.

References

1. Aceves-Fernandez M. A., O. A. Sotomayor, E. Gorrostieta-Hurtado, J. C. Pedraza-Ortega, J. M. Ramos-Arreguín, S., Canchola-Magdaleno, E. Vargas-Soto (2001). "Advances in Airborne Pollution Forecasting Using Soft Computing Techniques" pp. 3-14 Book Chapter, Air Quality - Models and Applications INTECH Publications ISBN 978-953-307-307-1
2. Air Quality in Cyprus, (2013) The Department of Labour Inspection (DLI), Ministry of Labour and Social Insurance http://www.airquality.dli.mlsi.gov.cy
3. Aktan M, H. Bayraktar (2010). The Neural Network Modeling of Suspended Particulate Matter with. Autoregressive Structure. Ekoloji 19, 74, 32-37
4. Chaturvedi D. K., (2008). Soft Computing Techniques and its Applications in Electrical Engineering Springer, Berlin
5. Díaz-Robles L.A., J.S. Fu and G.D Reed (2009). "Seasonal Distribution and Modeling of Diesel Particulate Matter in the Southeast US." Environment International, 35 (6), 956–964,
6. Dogra K, (2010.) "Autoscaling" QSARword-A Stand Life Sciences Web Resource http://www.qsarworld.com/qsar-statistics-autoscaling.php
7. European Commission (1999) Instructions 1999/30/EC. (22/04/1999), related to the limited values of SO2, Particulate Matter in the air http://www.modus.gr/site1/gr/000F4240/Data/1999-30-EK.pdf
8. Garcia I., J Rodriguez, Y. Tenorio (2010). "Artificial Neural Network Models for Prediction of Ozone Concentrations in Guadalajara, Mexico" Book Chapter "Earth and Planetary Sciences - Oceanography and Atmospheric Sciences Chapter 3 "Air Quality-Models and Applications", book edited by Nicolas Mazzeo INTECH Publications
9. Gentekakis I., B., (1999). Air Pollution Consequences, Control and Alternative Technologies (in Greek) Tziolas Edition, Thessaloniki, Greece
10. Hooyberghs J., C. Mensink, G. Dumont, F. Fierens, O. Brasseur (2005). "A neural network forecast for daily average PM10 concentrations in Belgium" Atmospheric Environment 39/18 3279-3289
11. Iliadis L., (2007) Intelligent Information Systems and Applications in Risk estimation Stamoulis Publishers, Thessaloniki, Greece
12. Iliadis L., S. Spartalis, A. Paschalidou, P. Kassomenos, (2007), "Artificial Neural Network Modelling of the surface Ozone Concentration" International Journal of Computational and Applied Mathematics. ISSN 1819-4966 Volume 2 Number 2 (2007), pp. 125–138
13. Iliadis L., A. Papaleonidas, (2009). "Intelligent Agents Networks Employing Hybrid Reasoning: Application in Air Quality Monitoring and Improvement" Communications in Computer and Information Science CCIS Vol. 43, p.p. 1-16 SPRINGER Proceedings of the 11th EANN (Engineering Applications of Neural Networks) London, UK
14. Kamataki A. E., (2009). "High scale transfer of Particulate Matter pollution in Thessaloniki" (in Greek) Graduate thesis, Aristotle University of Thessaloniki, Greece

15. Kecman V., (2001). “Learning and Soft Computing”, The MIT Press, Cambridge, MA., USA
16. Mogireddy K., K. Reddy, (2011). “Physical Characterization of Particulate Matter Employing Support Vector Machine Aided Image Processing” Master of Science Thesis in Electrical Engineering, University of Toledo, Spain, Electrical Engineering http://etd.ohiolink.edu/view.cgi?acc_num=toledo1297358759
17. Olej V., Obrsalova I., Krupka J., (2009) “Modelling of Selected Areas of Sustainable Development by Artificial Intelligence and Soft Computing” University of Parduvice, Czech Republic
18. Sumathi S. and Surekha P., (2010) Computational Intelligence Paradigms Theory and Applications using Matlab CRC Press, Taylor and Francis Group USA
19. Technical chamber of Greece, (2013) http://portal.tee.gr
20. Theocharis I., (2006). Fuzzy Systems Aristotle University of Thessaloniki, Greece
21. Thomas S., R.B. Jacko (2007). A stochastic model for estimating the impact of highway incidents on air pollution and traffic delay. Journal of Transportation Research Record A Journal of the Transportation Research Board, USA
22. Ziomas I, P., Suppan, B. Rappengluck, D. Balis, P. Tzoumaka, D. Melas, A. Papayiannis, P. Fabian, C. Zerefos, (1995). A contribution to the study of photochemical smog in the Greater Athens area Beitr. Phys. Atmosph. 68, 191-203