1

Advances in Airborne Pollution Forecasting Using Soft Computing Techniques

Aceves-Fernandez Marco Antonio, Sotomayor-Olmedo Artemio, Gorrostieta-Hurtado Efren, Pedraza-Ortega Jesus Carlos, Ramos-Arreguín Juan Manuel, Canchola-Magdaleno Sandra and Vargas-Soto Emilio
Facultad de Informática, Universidad Autónoma de Querétaro, México

1. Introduction

There are many investigations reported in the scientific literature about Particulate Matter (PM) 2.5 and PM10 in urban and suburban environments [Vega *et al* 2002, Querol *et al* 2004, Fuller *et al* 2004].
In this contribution, the information acquired from PMx monitoring systems is used to accurately forecast particle concentration using diverse soft computing techniques.
A number of works have been published in the area of airborne particulates forecasting. For example, Chelani[*et al* 2001] trained hidden layer neural networks for CO forecasting at India. Caselli [*et al* 2009] used a feedforward neural network to predict PM10 concentration. Other works such as Kurt's [*et al* 2010] have constructed a neural networks model using many input variables (e.g. wind, temperature, pressure, day of the week, Date, concentration, etc) making the model too complex and inaccurate.
However, not many scientific literature discuss a number of robust forecasting methods using soft computing techniques. These techniques include neuro-fuzzy inference methods, fuzzy clustering techniques and support vector machines. Each one of these algorithms is discussed separately and the results discussed. Furthermore, a comparison of all methods is made to emphasize their advantages as well as their disadvantages.

2. Fuzzy inference methods

Fuzzy inference systems (FIS) are also known as fuzzy rule-based systems. This is a major unit of a fuzzy logic system. The decision-making is an important part in the entire system. The FIS formulates suitable rules and based upon the rules the decision is made. This is mainly based on the concepts of the fuzzy set theory, fuzzy IF–THEN rules, and fuzzy reasoning. FIS uses "IF - THEN" statements, and the connectors present in the rule statement are "OR" or "AND" to make the necessary decision rules.
Fuzzy inference system consists of a fuzzification interface, a rule base, a database, a decision-making unit, and finally a defuzzification interface as described in Chang(et al 2006). A FIS with five functional block described in Fig.1.

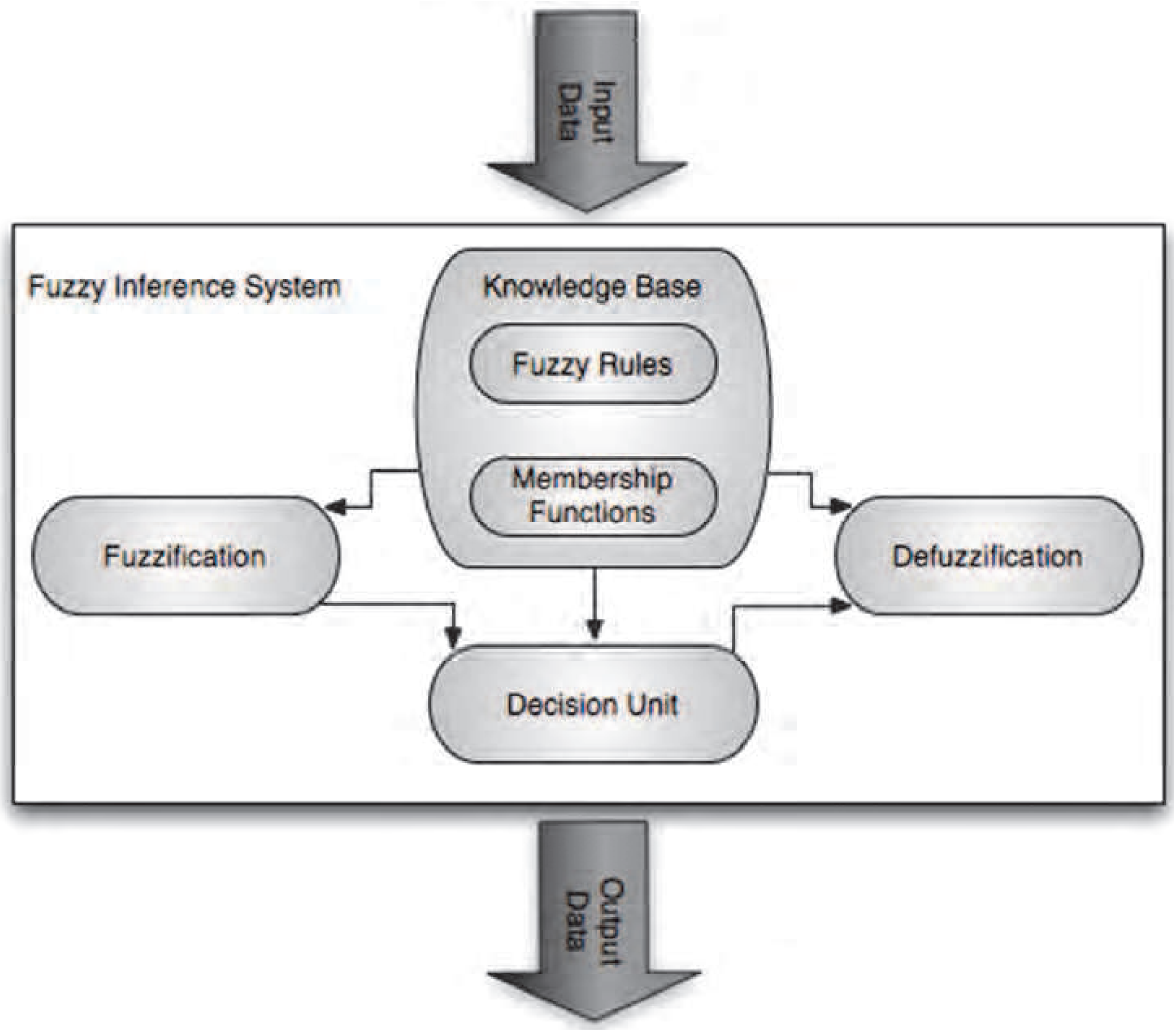


Fig. 1. Fuzzy Inference System

The function of each block is as follows:

- A rule base containing a number of fuzzy IF–THEN rules;
- A database which defines the membership functions of the fuzzy sets used in the fuzzy rules;
- A decision-making unit which performs the inference operations on the rules;
- A fuzzification interface which transforms the crisp inputs into degrees of match with linguistic values; and
- A defuzzification interface which transforms the fuzzy results of the inference into a crisp output.

The working of FIS is as follows. The inputs are converted in to fuzzy by using fuzzification method. After fuzzification the rule base is formed. The rule base and the database are jointly referred to as the knowledge base.

Defuzzification is used to convert fuzzy value to the real world value which is the output.

The steps of fuzzy reasoning (inference operations upon fuzzy IF–THEN rules) performed by FISs are:

- Compare the input variables with the membership functions on the antecedent part to obtain the membership values of each linguistic label. (this step is often called fuzzification.)
- Combine (through a specific t-norm operator, usually multiplication or min) the membership values on the premise part to get firing strength (weight) of each rule.
- Generate the qualified consequents (either fuzzy or crisp) or each rule depending on the firing strength.
- Aggregate the qualified consequents to produce a crisp output. (This step is called defuzzification.)

A typical fuzzy rule in a fuzzy model has the format shown in equation 1

$$\text{IF } x \text{ is } A \text{ and } y \text{ is } B \text{ THEN } z = f(x, y) \tag{1}$$

where AB are fuzzy sets in the antecedent; Z = f(x, y) is a function in the consequent. Usually f(x, y) is a polynomial in the input variables x and y, of the output of the system within the fuzzy region specified by the antecedent of the rule.
A typical rule in a FIS model has the form (Sugeno et al1988): IF Input 1 = x AND Input 2 = y, THEN Output is z = ax + by + c.
Furthermore, the final output of the system is the weighted average of all rule outputs, computed as

$$FinalOutput = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \tag{2}$$

3. Fuzzy clustering techniques

There are a number of fuzzy clustering techniques available. In this work, two fuzzy clustering methods have been chosen: fuzzy c-means clustering and fuzzy clustering subtractive algorithms. These methods are proven to be the most reliable fuzzy clustering methods as well as better forecasters in terms of absolute error according to some authors[Sin, Gomez, Chiu].
Since 1985 when the fuzzy model methodology suggested by Takagi-Sugeno [Takagi *et al* 1985, Sugeno *et al* 1988], as well known as the TSK model, has been widely applied on theoretical analysis, control applications and fuzzy modelling.
Fuzzy system needs the precedent and consequence to express the logical connection between the input output datasets that are used as a basis to produce the desired system behavior [Sin *et al* 1993].

3.1 Fuzzy clustering means (FCM)

Fuzzy C-Means clustering (FCM) is an iterative optimization algorithm that minimizes the cost function given by:

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{m} \|x_k - v_i\|^2 \tag{3}$$

Where n is the number of data points, c is the number of clusters, xk is the kth data point, vi is the ith cluster center μik is the degree of membership of the kth data in the ith cluster, and m is a constant greater than 1 (typically m=2)[Aceves *et al* 2011]. The degree of membership μik is defined by:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{2/(m-1)}} \tag{4}$$

Starting with a desired number of clusters c and an initial guess for each cluster center vi, i = 1,2,3… c, FCM will converge to a solution for vi that represents either a local minimum or a saddle point cost function [Bezdek *et al* 1985]. The FCM method utilizes fuzzy partitioning such that each point can belong to several clusters with membership values between 0 and 1. FCM include predefined parameters such as the weighting exponent m and the number of clusters c.

3.2 Fuzzy clustering subtractive

The subtractive clustering method assumes each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. Consider m dimensions of n data point (x1,x2, …, xn) and each data point is potential cluster center, the density function Di of data point at xi is given by:

$$D_i = \sum_{i=1}^{n} e^{\left(\frac{\|x_i - x_j\|^2}{\left(\frac{r_a}{2}\right)^2}\right)} \tag{5}$$

where r_a is a positive number. The data point with the highest potential is surrounded by more data points. A radius defines a neighbour area, then the data points, which exceed r_a, have no influence on the density of data point.

After calculating the density function of each data point is possible to select the data point with the highest potential and find the first cluster center. Assuming that X_{c1} is selected and D_{c1} is its density, the density of each data point can be amended by:

$$D_i = D_i - D_{c1} e\left(-\frac{\|x_i - x_{c1}\|^2}{\left(\frac{r_b}{2}\right)^2}\right) \tag{6}$$

The density function of data point which is close to the first cluster center is reduced. Therefore, these data points cannot become the next cluster center. r_b defines an neighbour area where the density function of data point is reduced. Usually constant $r_b > r_a$. In order to avoid the overlapping of cluster centers near to other(s) is given by [Yager *et al* 1994]:

$$r_b = \eta \cdot r_a \tag{7}$$

4. Support vector machines

The support vector machines (SVM) theory, was developed by Vapnik in 1995, and is applied in many machine-learning applications such as object classification, time series prediction, regression analysis and pattern recognition. Support vector machines (SVM) are based on the principle of structured risk minimization (SRM) [Vapnik *et al* 1995, 1997].

In the analysis using SVM, the main idea is to map the original data x into a feature space F with higher dimensionality via non-linear mapping function ϕ, which is generally unknown, and then carry on linear regression in the feature space [Vapnik 1995]. Thus, the regression approximation addresses a problem of estimating function based on a given data set, which is produced from the ϕ function. SVM method approximates the function by:

$$y = \sum_{i=1}^{m} w_i \phi_i(x) + b = w\phi(x) + b \tag{8}$$

where $w = [w_1, \ldots, w_m]$ represent the weights vector, b is defined as the bias coefficients and $\phi(x)=[\phi 1(x), \ldots, \phi m(x)]$ the basis function vector.
The learning task is transformed to the weights of the network at minimum. The error function is defined through the ε-insensitive loss function, $L\varepsilon(d,y(x))$ and is given by:

$$L_\varepsilon(d, y(x)) = \begin{cases} |d - y(x)| - \varepsilon & |d - y(x)| \geq \varepsilon \\ 0 & others \end{cases} \tag{9}$$

The solution of the so defined optimization problem is solved by the introduction of the Lagrange multipliers αi, α_i^* (where i=1,2,…,k) responsible for the functional constraints defined in Eq. 9. The minimization of the Lagrange function has been changed to the dual problem [Vapnik *et al 1997*]:

$$\phi(\alpha, \alpha^*) = \left[\sum_{i=1}^{k} d_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{k} (\alpha_i - \alpha_i^*) \right.$$

$$\left. -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} (\alpha_i, \alpha_i^*)(\alpha_j, \alpha_j^*) K(x_i, x_j) \right] \tag{10}$$

With constraints:

$$\begin{gathered} \sum_{i=1}^{k} (\alpha_i, \alpha^*{}_i) = 0, \\ 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C \end{gathered} \tag{11}$$

Where C is a regularized constant that determines the trade-off between the training risk and the model uniformity.
According to the nature of quadratic programming, only those data corresponding to non-zero $\alpha_i - \alpha_i^*$ pairs can be referred to support vectors (nsv). In Eq. 10 $K(x_i, x_j)=\phi(x_i)*\phi(x_j)$ is the inner product kernel which satisfy Mercer's condition [Osuna *et al 1997*] that is required for the generation of kernel functions given by:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \tag{12}$$

Thus, the support vectors associates with the desired outputs $y(x)$ and with the input training data x can be defined by:

$$y(x) = \sum_{i=1}^{Nsv} (\alpha_i, \alpha_i^*) K(x, x_i) + b \tag{13}$$

Where x_i are learning vectors. This leads to a SVM architecture (Fig. 2) [Vapnik 1997, Cristianini *et al* 2000].

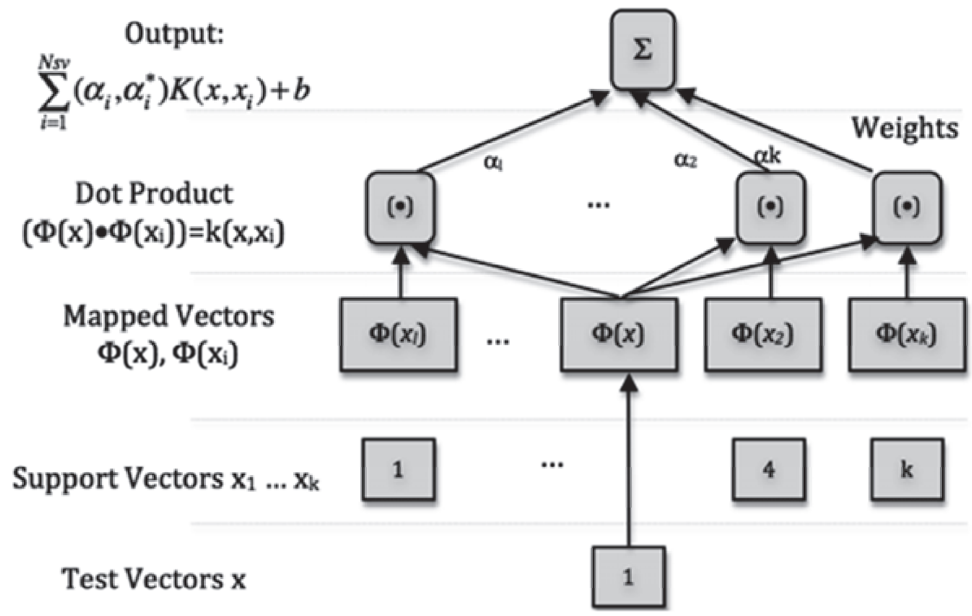


Fig. 2. Support Vector Machine Architecture.

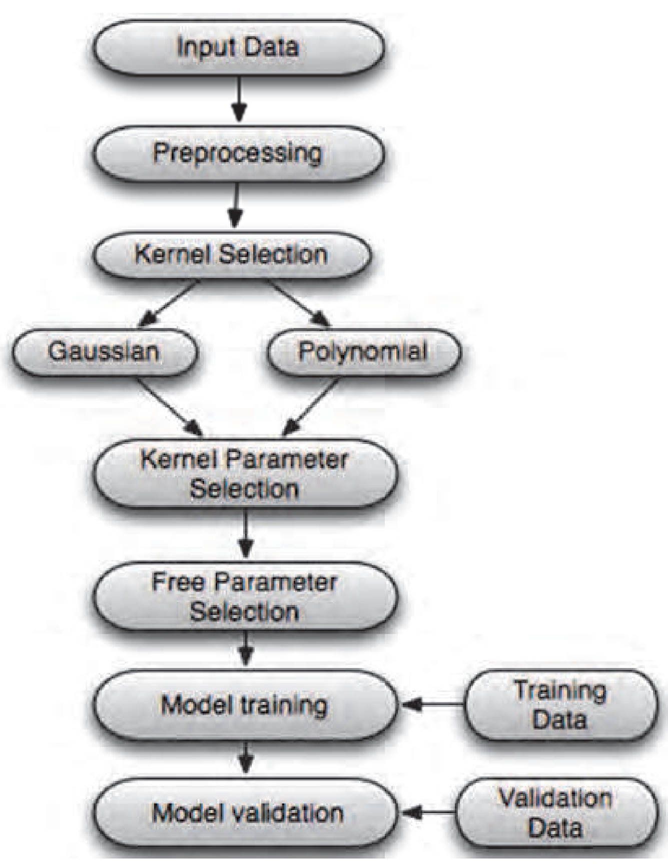


Fig. 3. Support Vector Machine Methodology.

The methodology used for the design, training and testing of SVM is proposed as follows based in a review of Vapnik, Osowski [*et al* 2007] and Sapankevych[*et al* 2009]

a. Preprocess the input data and select the most relevant features, scale the data in the range [−1, 1], and check for possible outliers.
b. Select an appropriate kernel function that determines the hypothesis space of the decision and regression function.
c. Select the parameters of the kernel function the variances of the Gaussian kernels.
d. Choose the penalty factor C and the desired accuracy by defining the ε-insensitive loss function.
e. Validate the model obtained on some previously, during the training, unseen test data, and if not pleased iterate between steps (c) (or, eventually b) and (e).

5. Discussion of results

Simulations were performed using fuzzy clustering algorithms using the equations [3-7], in this case study, the datasets at Mexico City in 2007 were chosen to construct the fuzzy model. Likewise, the data of 2008 and 2009 from the same geographic zone in each case were used to training and validating the data, respectively. The result of the fuzzy clustering model was compared then to the real data of Northwest Mexico in 2010.
The results obtained show an average least mean square error of 11.636 using Fuzzy Clustering Means, whilst FCS shows an average least mean square error of 10.59. Table 1 shows a list of the experiments carried out. An example of these results is shown in figure 4 for FCM and figure 5 shows the estimation made using FCS at Northwest Mexico City.

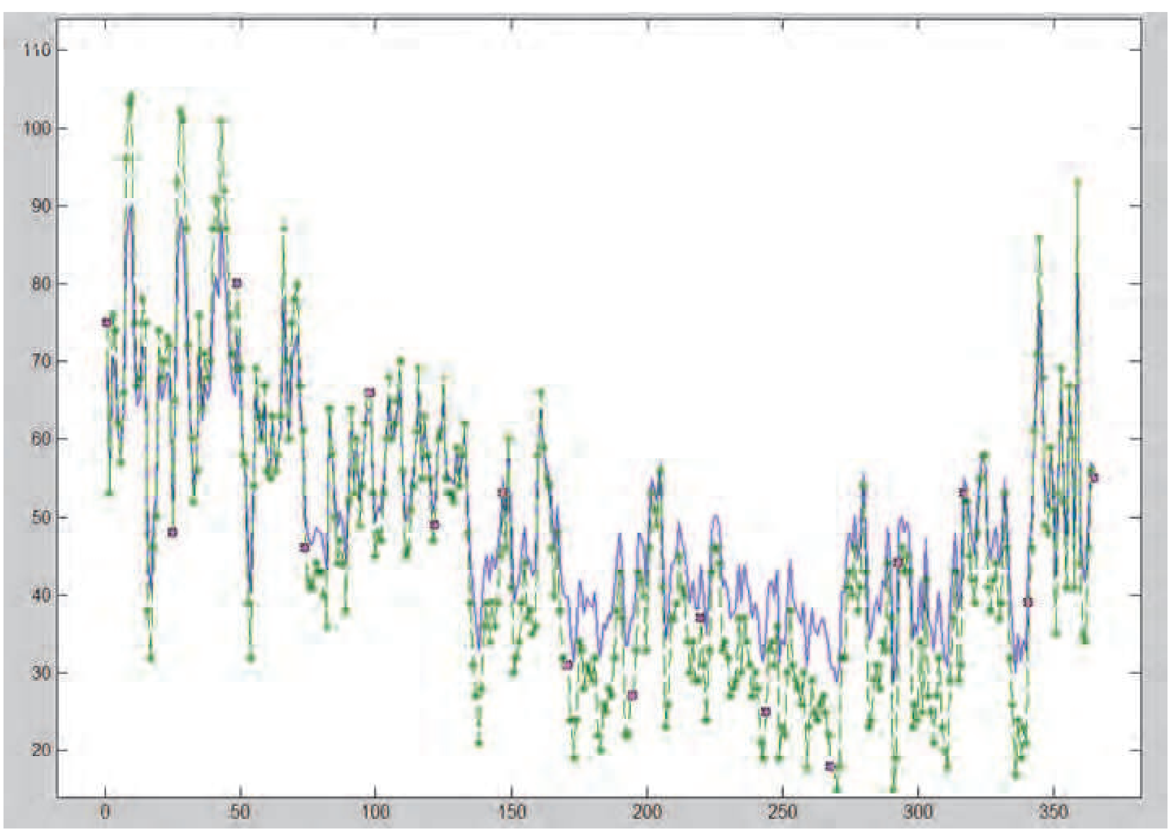


Fig. 4. Fuzzy Clustering Means (FCM) Results at Northwest Mexico City. Raw Data VS. Fuzzy Model

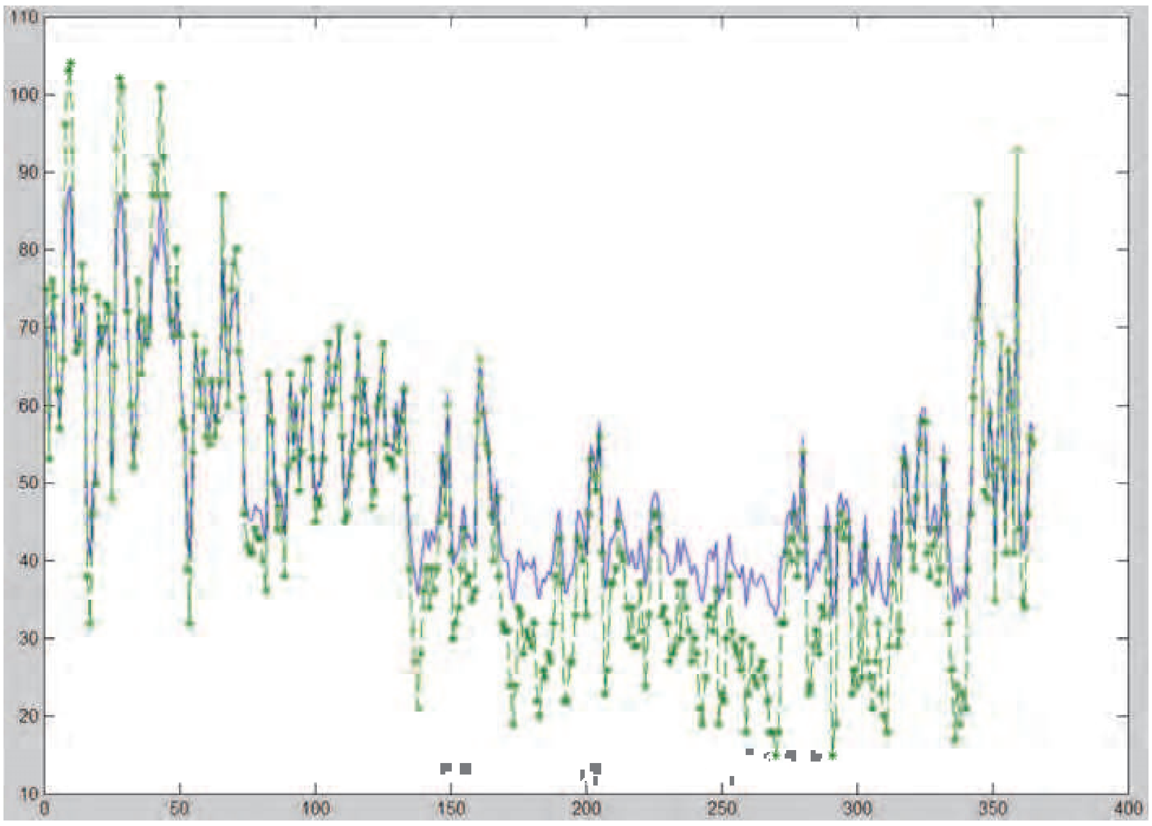


Fig. 5. Fuzzy Clustering Subtractive (FCS) Results at Northwest Mexico City. Raw Data VS. Fuzzy Model

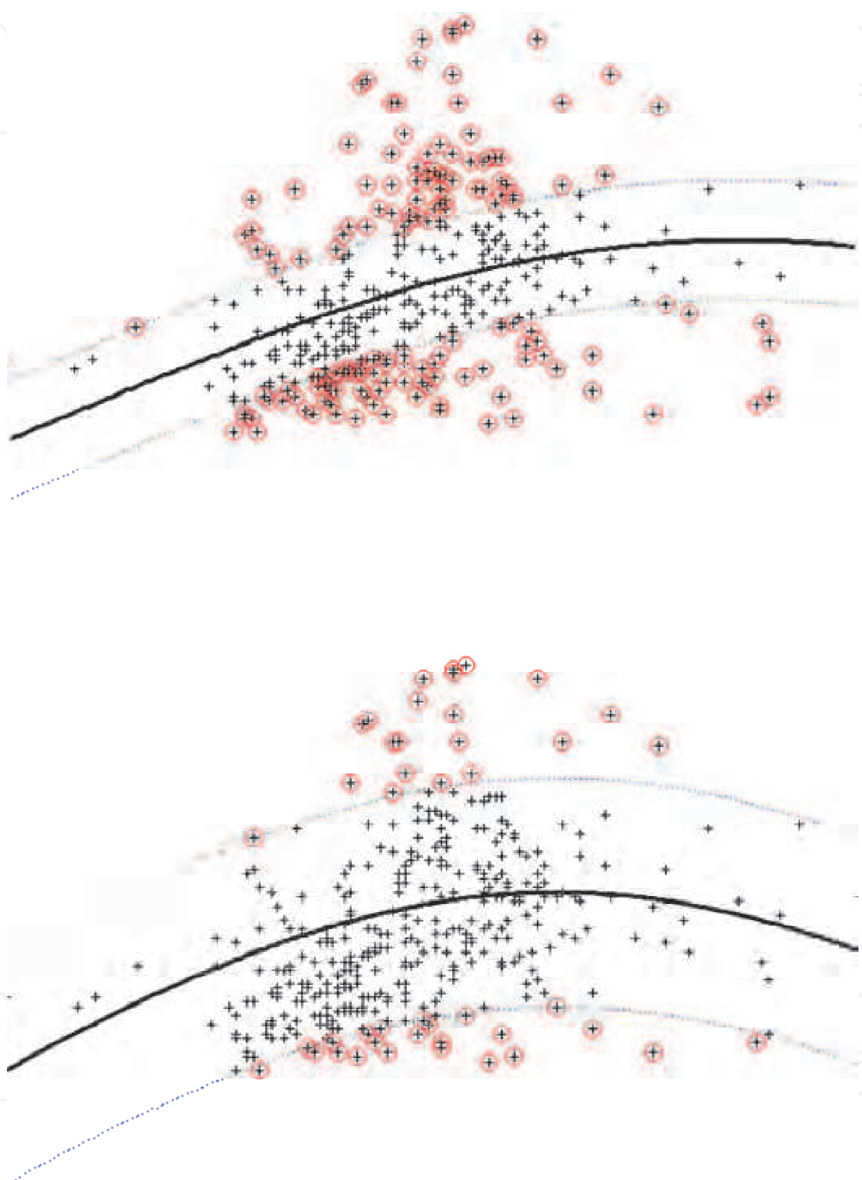
In figures 4 and 5, the raw data (shown in blue solid line) and the constructed fuzzy model (in dashed-starred green line) shown that the trained model is approximated to the raw data with an average least mean square error of 8.7%, implying that a fuzzy model can be accurately constructed using this technique.

Site	LMSE using FCM	LMSE using FCS
Northwest	10.1917	7.4807
Northeast	13.6282	13.7374
Center	18.5757	15.1409
Southwest	5.0411	7.4953
Southeast	10.7428	9.1188

Table 1**.** List of the experiments carried out using FCM and FCS.

In table 1 is shown that the best prediction in terms of error percentage is given at southwest for both fuzzy clustering means and fuzzy clustering subtractive, whilst the lessen estimation is given at the city center. This may be due to the high variations in terms of PM10 particles making it more difficult to predict. However, more research is needed to confirm this.
Furthermore, detailed simulations were carried out using Support Vector Machines following the proposed methodology shown in figure 3. These simulations were carried out

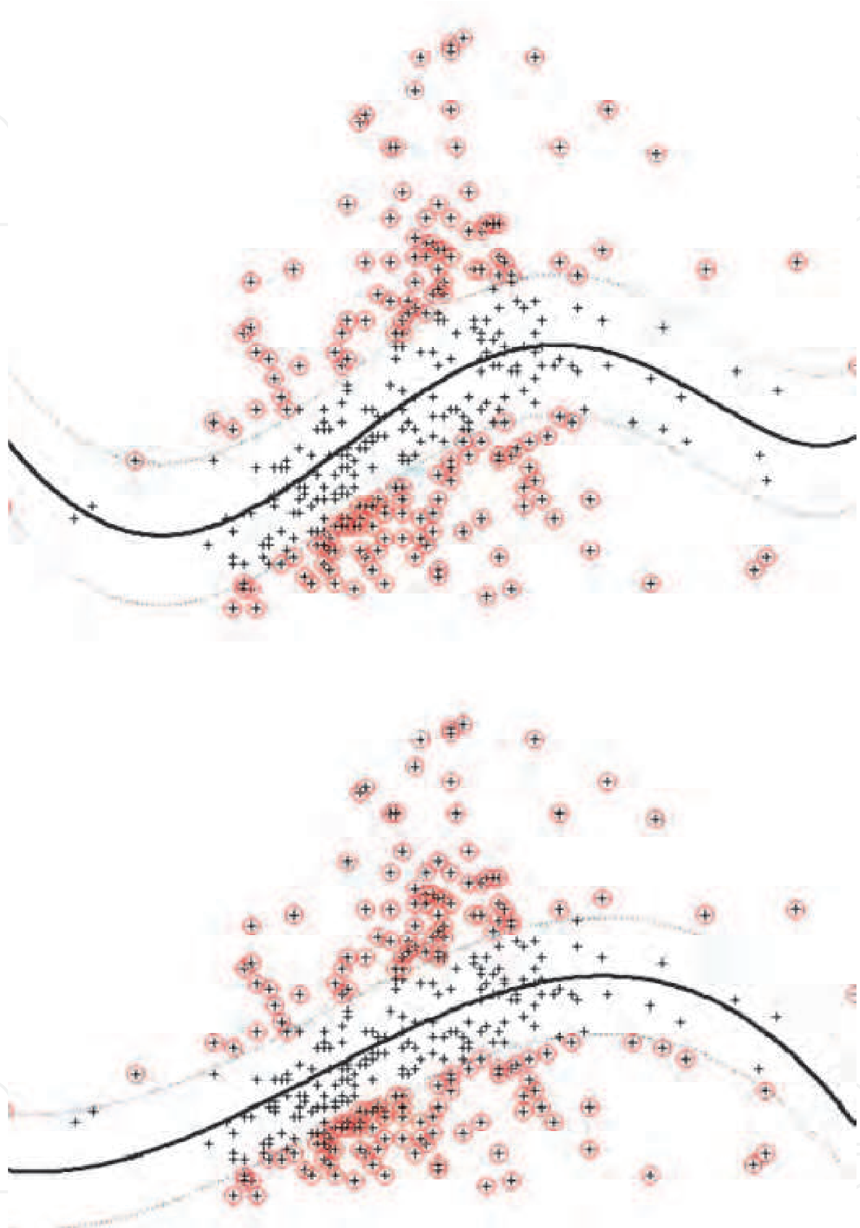
using the same dataset as the fuzzy clustering technique. In this case, values 2 σ was chosen, and an ε of 11 and 13 were chosen since it was demonstrated to give better results in previous contributions (Sotomayor *et al* 2010, Sotomayor *et al* 2011). Figure 6 shows the results of the model using support vector machines with a Gaussian kernel, whilst figure 7 shows the results using the same datasets, with polynomial kernel



a) SVM Estimated with free parameters of ε = 13 and σ = 2
b) SVM Estimated with free parameters of ε = 11 and σ = 2

Fig. 6. SVM Results at Northwest Mexico City using Gaussian Kernel.

Figure 6 indicates a summary of the results with the Support vector machine (in red circles), the raw data (black cross) and the behavior of the data (solid black line). These results show that for Gaussian Kernel (fig 6) gives 11.8 error using the same LMSE Algorithm than the fuzzy model with an epsilon of 13 giving a total number of support vector machines of 157. In the case of figure 5b, using the Gaussian kernel, it was also used the same σ and an epsilon of 11. For this figure, the support vector shows an improvement by having an LMSE of 8.7.



a) SVM Estimated with free parameters of ε = 13 and σ = 2
b) SVM Estimated with free parameters of ε = 11 and σ = 2

Fig. 7. SVM Results at Northwest Mexico City using Polynomial Kernel.

For figure 7a, the estimation gives an error of 9.8 using an σ of 2 and an epsilon of 11 using 177 support vector machines. Likewise, figure 7b also shows the estimation using a third degree polynomial kernel with an ε of 13. In this case, a 10.1 LMSE is shown by having 183 support vector machines.

6. Conclusions and further work

An assessment in the performance of both fuzzy systems generated using Fuzzy Clustering Subtractive and Fuzzy C-Means was made taking in account the number or membership functions, rules, and Least Mean Square Error for PM10 particles. As a case study, Estimations were made at Northwest Mexico City in 2010, giving consistent results.

In case of SVMs, it can be concluded that for this case study an ε of 11 gives a better estimation than an ε of 13 for the Gaussian kernel. In general, the Gaussian kernel gives better results in terms of estimation than its corresponding polynomial kernel. In general terms, fuzzy clustering gives a better estimation than Gaussian and polynomial kernels, although in-depth studies are needed to corroborate these results for other scenarios.

For future work, more SVM kernels can be implemented and comparison can be made to find out which kernels give better estimation. Also, SVMs can be implemented along with other techniques such as wavelet transform to improve the performance of these algorithms

7. References

Aceves-Fernández M.A., Sotomayor-Olmedo A., Gorrostieta-Hurtado E., Pedraza-Ortega J.C., Tovar-Arriaga S., Ramos-Arreguin J.M., Performance Assessment of Fuzzy Clustering Models Applied to Urban Airborne Pollution, *CONIELECOMP 2011, 21th International Conference on Electrical Communications*, pp. 212-216 (2011).

Bezdek, J. C., “Pattern Recognition with Fuzzy Objective Function Algorithms”, *Plenum Press, NY*, 1981.

Caselli M. & Trizio L. & de Gennaro G. & Ielpo P., “A Simple Feedforward Neural Network for the PM10 Forecasting: Comparison with a Radial Basis Function Network and a Multivariate Linear Regression Model”, *Water Air Soil Pollut* (2009) *201*:365–377

Chang Wook A., “Advances in Evolutionary Algorithms: Theory, Design and Practice”, Springer, ISSN: 1860-949X, 2006.

Chelani A.B.; Hasan M. Z., “Forecasting nitrogen dioxide concentration in ambient air using artificial Neural networks”, *International Journal of Environmental Studies*, 2001, Vol. *58*, pp. 487-499

Chiu S, “Fuzzy model identification based on cluster estimation”, *Journal of Intelligent and Fuzzy Systems*; September 1994, *2*, pp. 267–78.

Cristianini, N., Shawe-Taylor, J., An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, Cambridge, UK (2000)

Fuller G W and Green D., “The impact of local fugitive PM10 from building works land road works on the assessment of the European Union Limit Value”, *Atmospheric Environment* 2004**,** *38*, pp. 4493–5002.

Gomez, A. F., M. Delgado, and M. A. Vila, “About the Use of Fuzzy Clustering Techniques for Fuzzy Model Identification”, *Fuzzy Set and System*,. 1999, pp. 179-188.

Kurt Atakan, Oktay Ayse Betül, “Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks”, *Expert Systems with Applications*, *37* (2010) 7986–7992.

Osowski S. and Garanty K., "Forecasting of the daily meteorological pollution using wavelets and support vector machine," *Engineering Applications of Artificial Intelligence, vol. 20*, no. 6, pp. 745-755, September 2007.

Osuna, E., R. Freund, F. Girosi.:Support vector machines: Training and applications. AI Memo 1602, Massachusetts Institute of Technology, Cambridge, MA 44. (1997).

Querol X, Alastvey A, Ruiz C.R., Avtinano B, Hausson H.C., Harrison R.M, Buringh E, Ten Brink H.M, Lutz M, Bruckmann P, Straehl P and Schnerflev J., "Speciation and origin of PM10 and PM 2.5 in selected European cities", *Atmospheric Environment.* 2004, *38*, pp. 6547 – 6555.

Sapankevych I. and Sankar R., "Time series prediction using support vector machines: A survey," *Computational Intelligence Magazine*, IEEE, *vol. 4, no. 2*, pp. 24-38, 2009.

Schölkfopf B.: Smola A. J.: and Burges C.: Advances in Kernel Methods –Support Vector Learning. *Cambridge, M.A.: MIT Press*. 1999.

Sin, S. K., and De Figueiredo, "Fuzzy System Designing Through Fuzzy Clustering and Optimal preDefuzzification", *Proc. IEEE International Conference on Fuzzy Systems*. 1993 *2*, 190-195.

Sotomayor-Olmedo A., Aceves-Fernandez M.A., Gorrostieta-Hurtado E., Pedraza-Ortega J.C., Ramos-Arreguin J.M., Vargas-Soto J.E., Tovar-Arriaga S., "Modeling Trends of Airborne Particulate Matter by using Support Vector Machines", *7th International Conference on Electrical and Electronics Engineering Research (CIIIEE 2010)*, November 10-12 2010, Aguascalientes, Ags. Mexico, ISBN: 978-607-95060-3-2

Sotomayor-Olmedo A., Aceves-Fernandez M.A., Gorrostieta-Hurtado E., Pedraza-Ortega J.C., Vargas-Soto J.E., Ramos-Arreguin J.M., Villaseñor-Carillo U., "Evaluating Trends of Airborne Contaminants by using Support Vector Regression Techniques", *CONIELECOMP 2011, 21th International Conference on Electrical Communications*, pp. 137-141 (2011).

Sugeno, M., and G. T. Kang. "Structure Identification of Fuzzy Model", *Fuzzy Sets and Systems*. 1988, *28*, pp. 15-33.

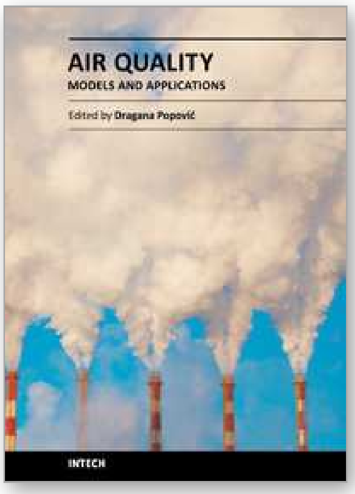
Takagi, T., and M. Sugeno, "Fuzzy Identification of Systems and its Application to Modeling and Control", *IEEE Trans. Systems Man and Cybernetics*. 1985 -*15*, pp. 116-132.

Vapnik, V.: The Nature of Statical Learning Theory. *Springer-Verlang, New York*. 1995.

Vapnik, V., Golowich, S., Smola A.: Support method for function approximation regression estimation, and signal processing. Advance *in Neural Information Processing System 9*. MIT Press, Cambridge, MA. 1997.

Vega E, Reyes E Sanchez G, Ortiz E, Ruiz M, Chow J, Watson J and Edgerton S, "Basic Statistics of PM2.5 and PM10 in the atmosphere of Mexico City", *The science of the total environment* 2002, *287*,pp. 167-176.

Yager, R. and D. Filev, "Generation of Fuzzy Rules by Mountain Clustering", *Journal of Intelligent & Fuzzy Systems*, 1994, 2, pp. 209- 219.



Air Quality-Models and Applications
Edited by Prof. Dragana Popovic

ISBN 978-953-307-307-1
Hard cover, 364 pages
Publisher InTech
Published online 09, June, 2011
Published in print edition June, 2011

Air pollution has been a major transboundary problem and a matter of global concern for decades. High concentrations of different air pollutants are particularly harmful to large cities residents, where numerous anthropogenic activities strongly influence the quality of air. Although there are many books on the subject, the one in front of you will hopefully fulfill some of the gaps in the area of air quality monitoring and modeling, and be of help to graduate students, professionals and researchers. The book is divided in five sections, dealing with mathematical models and computing techniques used in air pollution monitoring and forecasting; air pollution models and application; measuring methodologies in air pollution monitoring and control; experimental data on urban air pollution in China, Egypt, Northeastern U.S, Brazil and Romania; and finally, the health effects due to exposure to benzene, and on the influence of air pollutants on the acute respiratory diseases in children in Mexico.

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

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