Scientific Research

Method to Improve Airborne Pollution Forecasting by Using Ant Colony Optimization and Neuro-Fuzzy Algorithms

Elizabeth Martinez-Zeron[1], Marco A. Aceves-Fernandez[1]*, Efren Gorrostieta-Hurtado[1], Artemio Sotomayor-Olmedo[2], Juan Manuel Ramos-Arreguín[1]

[1]Facultad de Informática, Universidad Autónoma de Querétaro, Querétaro, México
[2]Laboratory of Artificial Intelligence, Faculty of Engineering, Universidad Autónoma de Querétaro, Querétaro, México
Email: *marco.aceves@gmail.com

Abstract

This contribution shows the feasibility of improving the modeling of the non-linear behavior of airborne pollution in large cities. In previous works, models have been constructed using many machine learning algorithms. However, many of them do not work for all the pollutants, or are not consistent or robust for all cities. In this paper, an improved algorithm is proposed using Ant Colony Optimization (ACO) employing models created by a neuro-fuzzy system. This method results in a reduction of prediction error, which results in a more reliable prediction models obtained.

Keywords

1. Introduction

In recent years, the environment has been affected by the presence of particulate pollutants such as the Ozone (O_3), NO_2 nitrogen oxide, carbon monoxide CO, sulfur dioxide (SO_2) and particulate matter less than 10 microns PM_{10} ($\leq$10 microns) and Particles less than 2.5 micrometers $PM_{2.5}$ ($\leq$2.5 microns) [1]. For this reason, pollution monitoring has been necessary for large cities with high concentration of population and industries.

During several years the air quality in Mexico City for prevention of toxicity levels in health and the environment have been observed and evaluated [1].

---

*Corresponding author.

Measures were performed to obtain information efficiency and reliability of the air quality. Some of these contributions consists in predicting pollution levels such as the work of Cortina [2], who forecast levels of ozone pollution in the city of Guanajuato in Mexico by using neural networks, Aceves [3] takes modeling variables between relative humidity and temperature level of contamination. Sotomayor [4] uses support vector machines and kernel functions particulate ozone (O_3), (PM_{10}) and nitrogen dioxide in Mexico City.

Patterns of pollution levels do not show linear behavior [3] [4] whereby a pattern is generated in clusters can be represented by several linear functions, this for ease of interpretation. In this case study, Fuzzy C means is used to generate clusters which have similar characteristics, subsequently establish the cluster centers as membership functions in a fuzzy system [3], Also, neuro-fuzzy inference system (ANFIS) with multiple inputs and one output ("Multiple-Inputs-Single-Output" or MISO) is used to approximate nonlinear functions [5] [6]. Subsequently, three models are generated by the above steps after obtaining improved by using the algorithm of ant colony optimization (ACO) prediction. This methodology shows that it is possible to improve existing algorithms to predict the levels of particulate pollutants, in this case study in Mexico City.

1.1. Fuzzy Inference Methods

A Fuzzy Inference System (FIS) is basically a logic that allows intermediate values defined between conventional evaluations as Yes/No, True/False, etc.

A FIS is the main unit of a fuzzy logic system, the part where the decision is made, by constructing appropriate rules based on the theory of fuzzy sets. The rules are constructed linguistically (IF-THEN-) having the general form of "If A Then B" where A and B are (collections of) propositions containing linguistic variables linked by connectors (AND, OR) to make the correct decision-making [7].

A fuzzy inference system consists of 5 blocks as described in **Figure 1**.

Fuzzy inference system operates as follows: The input data are treated by a method of fuzzification then the Fuzzy rules are formed and the already fuzzyfied input data are analyzed. Defuzzification method is used to convert the value obtained after analysis with fuzzy rules to obtain an output value applicable to the real world.

The steps taken by fuzzy reasoning, *i.e.* inference operations on rules IF - THEN performing the FIS, are:

- Compare the input variables with the membership functions for the linguistic value of each part, this process is called fuzzyfication.
- Combine by an operator ($t-\text{norm}$, usually multiplication or minimum) the membership values to obtain the weight of each rule.
- Add the resulting output to produce a crisp, this step is called defuzzification.

A fuzzy rule has a general structure in a fuzzy model as shown on Equation (1):

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

where A, B fuzzy sets in the previous data sets are, $z = f(x, y)$ is a consequent function. Usually $f(x, y)$ is polynomial in the input variables x and y, the output of the system with a fuzzy region specified by the predecessors of the rule.
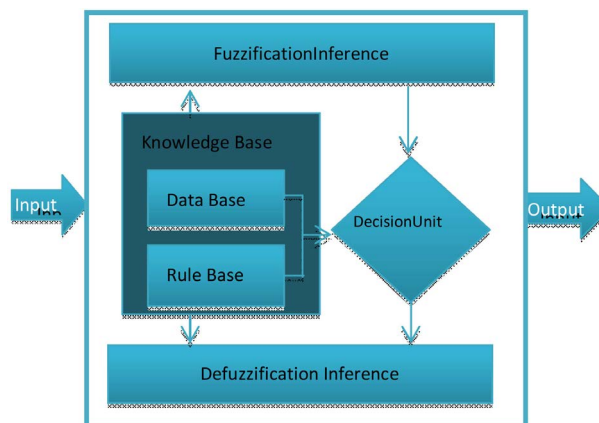

Figure 1. Block Diagram of a Generic Fuzzy inference system.

A typical rule in FIS model has the form as shown in Equation (2) [8]:

$$\text{IF Input } 1 = x \text{ AND Input } 2 = y, \text{THEN Output is } z = ax + b\,y + c \tag{2}$$

The final output of the FIS system is the weighted average of all the outputs of the rules:

$$\text{Final Ouput} = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \tag{3}$$

The Takagi-Sugeno model or TSK model proposed in 1985 [8] [9], is widely used in the theoretical analysis, application control and Fuzzy Modeling.

A Fuzzy system needs a precedent and consequent to express a logical connection between the input and output data used as a basis to produce the desired behavior of the system.

1.2. Clustering

Clustering is the grouping of data based on certain criteria that clusters have similar properties and characteristics, clusters are based mainly on the distance interpreted as similarity [10].

1.2.1. Fuzzy C-Means

Fuzzy C-Means (FCM) is a clustering method which allows a data to belong in one or more clusters, this method developed by Dunn [11] and improved by Bezdek [10], is often used for pattern recognition, it is an iterative optimization algorithm that minimizes the cost function given by:

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^{m} \left\| x_k - v_i \right\|^2 \tag{4}$$

where n is the number of data, c is the number of clusters, x_k is the kth data point, v_i is the center of the ith cluster, μ_{ik} is the degree of membership of the kth data in the ith cluster, m is any real number greater than 1, (typically $m = 2$) [10]. The degree of membership μ_{ik} is defined by:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\dfrac{\left\| x_k - v_i \right\|}{\left\| x_k - v_j \right\|} \right)^{2/(m-1)}} \tag{5}$$

Starting with the desired number of cluster c and an initial randomly chosen center for each cluster v_i, $v_i = 1, 2, 3, L, c$, FCM will converge in a solution to v_i which represents both a local minimum or a point function cost [10]. This clustering method uses fuzzy partition such that each point can belong to several clusters with different membership values between 0 and 1. Fuzzy c-means, has predefined parameters such the weight of the exponent m and the number of clusters.

To measure the effect of clustering algorithm, validity and accuracy are required. Validity: If the algorithm can find all internal type in the data collected. Truth: If the algorithm can set the same kind of data to the same group, different types of data at different clusters. We define the rate of decline of Clustering (C_r) to measure the effect of clustering [7]:

$$C_r = \frac{m_{\text{best}}}{m_{\text{result}}} \times 100 \tag{6}$$

m_{best} is best clustering number and m_{result} is the actual number of cluster after clustering.

$$v_i = \frac{\sum_{x \in X} \left(\mu_C \left(x \right) \right)^m}{\sum_{x \in X}^{n} \left(\mu_C \left(x \right) \right)^m}, \qquad 1 \le i \le k \tag{7}$$

After determining the number of clusters, the value of m, and the stopping criterion the FCM algorithm performs two steps, first calculates the membership functions using Equation (5), as a second step updates the proto types using Equation (7) the two steps are repeated iteratively to achieve the stability criterion, in which the change of the output values by the previous equations is minimal. The **Figure 2** shows an example FCM function.
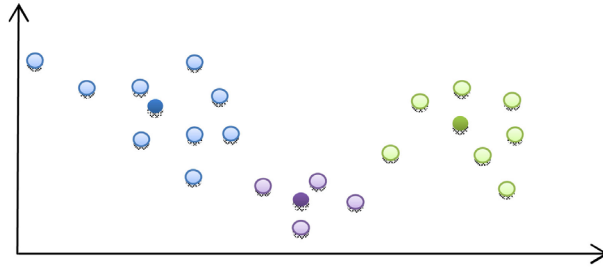
Figure 2. Example of Fuzzy *C*-Means showing three different clusters with a set of data points each, and their centroids.

1.3. The Adaptive Neuro-Fuzzy Inference System (ANFIS)

This method is a type of neural network that is based on a method for the process of fuzzy modeling to learn information about a data set by generating fuzzy rules, this is called neuro-fuzzy, this architecture was proposed by Jang in 1993 [12]. Has a network structure with directional links connecting a number of nodes. Each node has a function with adjustable or fixed parameters.

The classical ANFIS consists of 5 layers with specific tasks as show in **Figure 3** [13].

First Layer (Layer 1). Layer one is performed the input fuzzification. That is, each entry will set a value of belonging that only covers the total of the input variables to be treated [13]. This is expressed in mathematical terms as follows:

$$\text{Out}_{ij}^{(1)} = \mu_j \left(\text{In}_i^{(1)} \right) \tag{8}$$

While:

$\text{Out}_{ij}^{(1)}$ is the layer 1 node's output which corresponds to the j-th linguistic term of the i-th input variable. The membership is:

$$\mu_j = \frac{1}{1 + \left| \dfrac{x_i - c_{ij}}{a_{ij}} \right|^{bo}}, \quad i = 1, \cdots, \text{NumInVars}, \quad j = 1, \cdots, \text{NumIntTer} \tag{9}$$

While the parameters $\left(a_{ij}, b_{ij}, c_{ij} \right)$ are referred to as premise parameters or non-linear parameters and they adjust the shape and the location of the membership function.

Second Layer (Layer 2 from **Figure 3**). Executes the fuzzy AND of the antecedent part of the fuzzy rules. This results to each node's output being the product of all of its inputs (every input term node that is connected to that rule node) [14]:

$$\text{Out}_k^{(2)} = w_k = \prod_{i=1}^{N_{\text{Inputs}}} \text{Out}_{ij}^{(1)}, \tag{10}$$

For all the terms nodes j connected to the k-th the rule node, $k = 1, \cdots, \text{NumRules}$

Third layer (Layer 3 from **Figure 3**). Normalizes the membership functions (MFs).

The output of the k-th node is the firing strength of each rule divided by the total sum of the activation values of all the fuzzy rules. This results in the normalization of the activation value for each fuzzy rule. This operation is simply written as:

$$\text{Out}_k^{(3)} = \overline{w_k} = \frac{\text{Out}_k^{(2)}}{\sum_{m=1}^{N_{\text{Rules}}} \text{Out}_m^{(2)}}, \qquad k = 1, \cdots, \text{NumRules} \tag{11}$$

Fourth layer (Layer 4 from **Figure 3**) executes the consequent part of the fuzzy rules. Each node k in this layer is accompanied by a set of adjustable parameters a_{1k}, a_{2k}, $\cdots$, $a_{N_{\text{Inputs}}k}$, a_{0k} and implements the linear function:

$$\text{Out}_k^{(4)} = \overline{w_k} f_k = \overline{w_k} \left(a_{1k} \text{In}_1^{(1)} + a_{2k} \text{In}_2^{(1)} + \cdots + a_{N_{\text{Inputs}}k} \text{In}_{N_{\text{Inputs}}k}^{(1)} + a_{0k} \right), \qquad k = 1, \cdots, \text{NumRules} \tag{12}$$
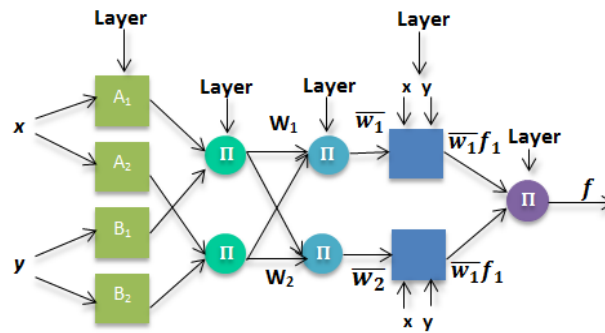
Figure 3. ANFIS architecture.

The weight w_k is the normalized activation value of the k-th rule, calculated by from Equation (7). Those parameters are called *consequent parameters* or *linear parameters* of the ANFIS system and are adjusted by the RLS algorithm.

Layer 5 computes the output of fuzzy system by summing up the outputs of layer fourth.

$$\text{Out}^{(5)} = \sum_{k=1}^{N_{\text{Rules}}} \text{Out}_k^{(4)} = \sum_{k=1}^{N_{\text{Rules}}} \overline{w_k} f_k = \frac{\sum_{k=1}^{N_{\text{Rules}}} w_k f_k}{\sum_{k=1}^{N_{\text{Rules}}} w_k} \tag{13}$$

1.3.1. Optimization

The optimization algorithm is a numerical method which finds a value $xi \in Rn$, where Rn is an n-dimensional search space, which minimizes or maximizes the function optimization $J(x)$; through the systematic selection of values of the variable xi using some restrictions. Here, $J(x)$ is called the objective function. A feasible solution that minimizes or maximizes the objective function is called an optimal solution [15].

As mentioned in [15] the techniques used for providing solution to complex combinatorial optimization problems have evolved from constructive methods to local search methods and finally to population-based algorithms. These techniques are classified as: Extensive and not exhaustive. These non-analytical computational techniques are based, and the classification is performed based on the solution space.

Exhaustive optimization techniques are those that guarantee always find the optimal (maximum or minimum) walking in the worst case the whole solution space (which can be considerably large) [16].

Non-exhaustive techniques rely on getting good enough solutions without exceeding the time constraints established or memory. Solutions are based on steps taken to find new solutions that approximate the optimal solution more, increasing quality.

Some of the best known are:
- Genetic Algorithms
- Taboo search
- Ant Colony Optimization
- Greedy Randomized Adaptive Search Procedure (GRASP).
- Scatter Search
- Simulated Annealing

1.3.2. Ants Colony Optimization

The ant colony optimization (ACO) is a bio-inspired technique of classification based on the external behavior of a real ant colony when foraging. This algorithm was created in 1992 by Dorigo [17].

The ACO algorithm solves complex problems of combinatorial optimization in various fields of engineering, commerce, industry, among others. When a problem does not have a polynomial equation that describes it belongs to the NP-hard problems. The results provided by ACO can be approximated to address the combinatorial optimization problemsas shown in the **Figure 4**. [18].

The probability of an ant k to choose the next node, where x is the current state and y is the next state we have:
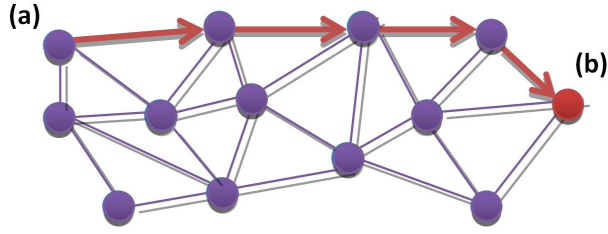
Figure 4. Example of ACO where ants choose the shortest path from node 1 to node 2.

$$p_{xy}^{k} = \frac{\left(\tau_{xy}^{\alpha}\right)\left(\eta_{xy}^{\beta}\right)}{\sum\left(\tau_{xy}^{\alpha}\right)\left(\eta_{xy}^{\beta}\right)} \tag{14}$$

where:

η_{xy} is the level of convenience that has a transition state, $\frac{1}{d_{xy}}d_{xy}$ is the distance.

τ_{ij} is the trace levels, or that the amount of pheromone which will be deposited in a state of transition.

β is a parameter that controls the influence of η_{xy}.

α is a parameter that controls the influence of τ_{xy}.

Updated Trail

Pheromone must be updated because this changes which followed the iterations by offering increased likelihood based on the accumulated experience [19].

In this case the level of pheromone raises based on the best solution path according to Equation (14)

$$\tau_{ij} = \left(1-\rho\right)\tau_{ij} + \Delta\tau_{ij} \tag{15}$$

where: $\rho \in (0,1]$ It is the persistence factor trail

$$\Delta\tau_{ij} = \sum_{k=1}^{\text{NrosAnts}} \Delta\tau_{ij} \tag{16}$$

Accumulating trace is proportional to the quality of solutions.

$$\Delta\tau_{ij} = \begin{cases} \dfrac{1}{L_k}, & \text{if ant } k \text{ used edge } p_i, p_j; \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

L_k Represents the objective value of the solution k (in the denominator for minimization problems).

Every time an ant travels a node applies the rule:

$$t_{ij} \leftarrow \left(1-\varphi\right)\times t_{ij} + t_0 \tag{18}$$

$\varphi \in (0,1]$ Decrement pheromone parameter.

The stop condition

This process is iterative until it stops for compliance with established criteria

2. Methodology

In **Figure 5**, the environmental monitoring map shown in Mexico City, in this study we only evidence in the northwest area of the spring season.

The list of the sites used in this contribution is shown on **Figure 5**, whilst the proposed methodology for this contribution is shown in the **Figure 6**.

The monitoring stations used for this case study were chosen due to data availability and specific industrial or polluted areas.

Since obtaining data from such stations is filtered data, *i.e.*, data values are complete and validated readings
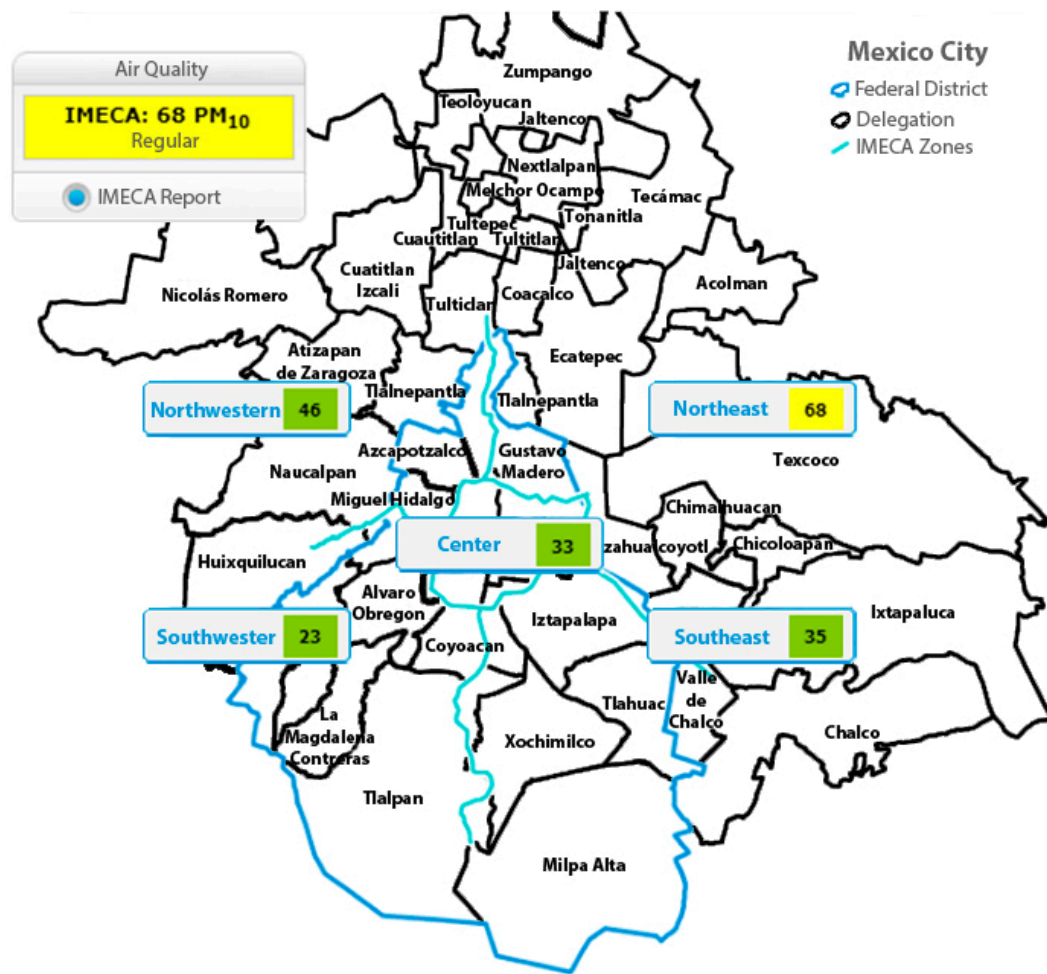
Figure 5. Monitored areas.

are sought.

The data are arranged in time series along the day; hence the clustering is performed to generate the characteristics of groups depending on pollution levels. These are grouped according through the fuzzy c means algorithm, which generates fuzzy clusters, based on a center that iteratively changes to accommodate more data groups. This provides greater robustness for modeling. Data modeling is performed using a fuzzy logic system, wherein each generated center is interpreted as fuzzy rules language.

The fuzzy system is improved with training data, which in this case, the training data are data from a previous year validated dataset. The training is through the use of neuro-fuzzy system AN-FIS which generates a more robust data modeling system.

The training is carried out using of neuro-fuzzy system ANFIS which generates a more robust data modeling system.

Three models for the prediction of air quality are generated; these models are subject to the ant colony optimization algorithm to improve the prediction closer to an outcome more accurately. Such three models along with the real measured data and its corresponding prediction are shown on **Figure 7**.

The mean square error for each model generated with neuro-fuzzy and neuro-fuzzy with ACO to measure the difference between the estimated and the model.

3. Results and Discussion

The models are created using Fuzzy Logic where the rules are formed by clustering, the ANFIS to improve each model and make each one more accurate and the ACO algorithm use these three models to improve the forecast
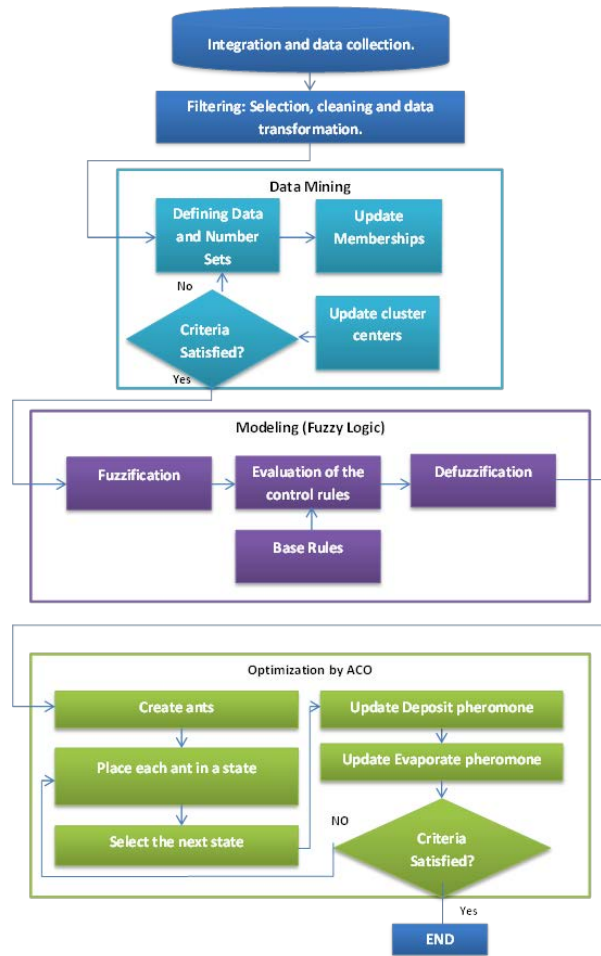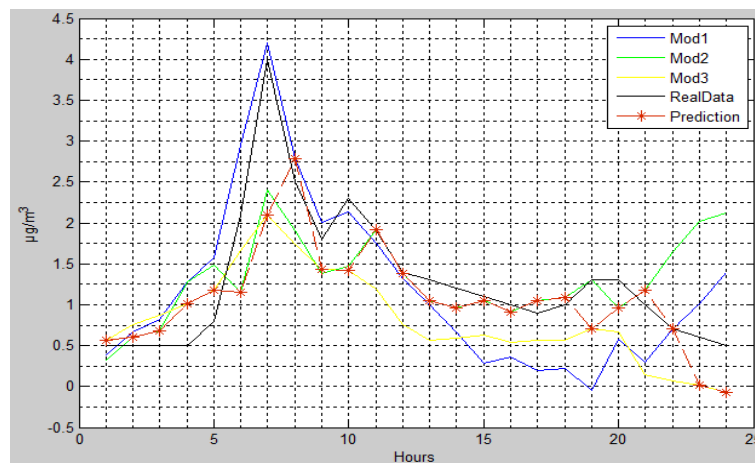
Figure 6. Proposed methodology.

Figure 7. Model selection by ACO.

having different options base on the three previous models. The models are named Mod 2 Mod 1 and Mod 3, respectively. These models are generated based on the characteristics of the stations, *i.e.* stations show greater similarity between their measurements generate a model.

The results are shown on **Table 1** and **Table 2**. CO is sensed to the northwest, with three models generated

Table 1. Results of spring in northwestern federal district with CO pollutant.

	Station 1	Station 2	Station 3	Station 4
Mod 1	0.845	0.431	0.678	0.652
Mod 2	1.732	0.591	0.572	0.890
Mod 3	0.97	2.540	1.041	0.860
Mod ACO	0.882	0.373	0.559	0.594

Table 2. Results of spring in northwestern federal district with O_3 pollutant.

	Station 1	Station 2	Station 3	Station 4
Mod 1	0.0067	0.0143	0.0116	0.0103
Mod 2	0.0193	0.0124	0.0283	0.0179
Mod 3	0.0078	0.0213	0.0407	0.0718
Mod ACO	0.0074	0.0100	0.0103	0.0117

based on the area in 2009, and data validation are the stations belonging to the same area but in 2010. Likewise the **Table 2** is showing the models of Ozone (O_3). In the following tables the mean square error, it can be seen that is closer to the real value, the prediction of stations Station 1, Station 2, Station 3 and Station 4 based Mod 1, Mod 2 and Mod 3 models. The station in which there was no good approximation was at Station 1.

Figure 7 shows how the algorithm is based on the nearest location from the current node to the next node following the ACO algorithm as shown on Equations (14)-(16) and the methodology shown on **Figure 6**. The influence of the pheromone that can choose the model that is closest to the actual data, likewise considering also the shortest distance.

The **Figure 7** shows the comparison between the models created by the trained logic system with ANFIS and the ACO model. This to appreciate the difference between the accuracy for each model generated with neuro-fuzzy and neuro-fuzzy improved with ACO.

4. Conclusions

The algorithms based on swarm intelligence are feasible for solving problems different than the typical machine learning methods. It has shown that a combination of modeling and optimization methods may be used to improve the prediction for this type of non-linear problem. Such combination of methods showed that systems working cooperatively is well complemented and integrated to generate a task system.

For these reasons ACO is an algorithm that provides an improvement in the prediction of contamination if used properly. ACO prediction has a point of view on the combinatorial type issue, and thus, the use of this algorithm improves the contaminants prediction by making the shortest search and the use of pheromone provides a best approximation to real data.

Acknowledgements

References

[1] http://www.calidadaire.df.gob.mx/calidadaire/index.php?opcion=4&opcionrecursostecnicos=3

[2] Cortina. M. (2012) Aplicación de inteligencia artificial a la predicción de contaminantes atmosféricos. Tesis Doctoral, E.T.S.I. Telecomunicación (UPM).

[3] Aceves-Fernández, M.A., Sotomayor-Olmedo, A., Gorrostieta-Hurtado, E., Pedraza-Ortega, J.C., Tovar-Arriaga, S. and Ramos-Arreguin, J.M. (2011) Performance Assessment of Fuzzy Clustering Models Applied to Urban Airborne Pollution. 2011 21*st International Conference on Electrical Communications and Computers* (*CONIELECOMP*), San Andres Cholula, 28 February-2 March 2011, 212-216.

[4] Sotomayor-Olmedo, A., Aceves-Fernández, M.A., Gorrostieta-Hurtado, E., Pedraza-Ortega, C., Ramos-Arreguín, J.M. and Vargas-Soto, J.E. (2013) Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A

Kernel Performance Approach. *International Journal of Intelligence Science*, **3**, 126-135. http://dx.doi.org/10.4236/ijis.2013.33014

[5] (2005) Adaptive Neuro-Fuzzy Inference Systems (ANFIS). In: Nedjah, N. and de Macedo Mourelle, L., Eds., *Fuzzy Systems Engineering: Theory and Practice*. *Vol.* 181, Springer, Rio de Janeiro, 53-83.

[6] Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Library for Simulink Ilias S. Konsoulas M.Sc., CEng, MIET.

[7] Aceves-Fernandez, M.A., Sotomayor-Olmedo, A., Gorrostieta-Hurtado, E., Pedraza-Ortega, J.C., Ramos-Arreguín, J.M, Canchola-Magdaleno, S. and Vargas-Soto, E. (2011) Advances in Airborne Pollution Forecasting Using Soft Computing Techniques. In: Popovic, D., Ed., *Air Quality*: *Models and Applications*, INTECH Publisher, Querétaro, 1-14. http://dx.doi.org/10.5772/16273

[8] Sugeno, M. and Kang, G.T. (1988) Structure Identification of Fuzzy Model. *Fuzzy Sets and Systems*, **28**, 15-33. http://dx.doi.org/10.1016/0165-0114(88)90113-3

[9] Takagi, T. and Sugeno, M. (1985) Fuzzy Identification of Systems and Its Application to Modeling and Control. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-15**, 116-132.

[10] Bezdek, J.C. (1981) Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York. http://dx.doi.org/10.1007/978-1-4757-0450-1

[11] Dunn, J.C. (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, **3**, 32-57. http://dx.doi.org/10.1080/01969727308546046

[12] Jang, J.S.R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems*, *Man and Cybernetics*, **23**, 665-685.

[13] Ghanbari, A., Abbasian-Naghneh, S. and Hadavandi, E. (2011) An Intelligent Load Forecasting Expert System by Integration of Ant Colony Optimization, Genetic Algorithms and Fuzzy Logic. 2011 *IEEE Symposium on Computational Intelligence and Data Mining* (*CIDM*), Paris, 11-15 April 2011, 246-251.

[14] Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Library for Simulink

[15] Muñoz, M., López, J. and Caicedos, E. (2008) Inteligencia de enjambres: Sociedades para la solución de problemas (una revisión). *Ingeniería e Investigación*, **28**, 119-130.

[16] Carretero, J.C.O., Cánovas, D.G. and Pereira, F.D.Q. (2008) Técnicas Heurísticas para Problemas de Diseño en Telecomunicaciones.

[17] Dorigo, M. (1992) Optimization, Learning and Natural Algorithms. Ph.D. Thesis, Politecnico di Milano, Italian.

[18] Calixto, C.A.C. (2005) Implementacion En Hidroinformática De Un Método De Optimización Matemática Basado En La Colonia De Hormigas. Tesis no publicada, Pontificia Universidad Javeriana, Bogota, Colombia.

[19] Kole, A., Chakrabarti, P. and Bhattacharyya, S. (2013) An Ant Colony Optimization Algorithm for Uncapacitated Facility Location Problem.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.