CLUSTERING APPROACH APPLIED ON AN ARTIFICIAL NEURAL NETWORK MODEL TO PREDICT PM_{10} IN MEGA CITIES OF MEXICO

E. MAGAÑA-VILLEGAS[1], J.M. CARRERA-VELUETA[1], S. RAMOS-HERRERA[1], J.R. HERNÁNDEZ-BARAJAS[1], C. GONZÁLEZ-FIGUEREDO[2], J.R. LAINES-CANEPA[1], A. VALDÉS-MANZANILLA[1] & R.G. BAUTISTA-MARGULIS[1]
[1]Professor at the Universidad Juárez Autónoma de Tabasco, México.
[2]Professor at the Instituto Tecnológico y de Estudios Superiores de Occidente, Jalisco, México.

ABSTRACT
A cluster-based artificial neural network model called CLASO (Classification-Assemblage-Association) has been proposed to predict the maximum of the 24-h moving average of PM_{10} concentration on the next day in the three largest metropolitan areas of Mexico. The model is a self-organised, real-time learning neural network, which builds its topology via a process of pattern classification by using an historical database. This process is based on a supervised clustering technique, assigning a class to each centroid of the hidden layer, employing the Euclidean distance as a hierarchical criterion. A set of ARIMA models was compared with CLASO model in the forecast performance of the 24-h average PM_{10} concentration on the next day. In general, CLASO model produced more accurate predictions of the maximum of the 24-h moving average of PM_{10} concentration than the ARIMA models, although the latter showed a minor tendency to underpredict the results. The CLASO model solely requires to be built a historical database of the air quality parameter, an initial radius of classification and the learning factor. CLASO has demonstrated acceptable predictions of 24-h average PM_{10} concentration by using exclusively regressive PM_{10} concentrations. The forecasting capabilities of the model were found to be satisfactory compared to the classical models, demonstrating its potential application to the other major pollutants used in the Mexican air quality index.
Keywords: air quality, artificial neural network, clustering, PM_{10} modelling.

1 INTRODUCTION

In the last decades, the air quality in most populated cities of Mexico has shown a strong tendency towards deterioration. Likewise, restoration capability of the air quality has been diminishing, resulting in a growing probability of appearance of diseases caused by atmospheric pollution. One of the major air pollutants is the particulate matter which either is emitted directly to the atmosphere or is formed as a result of chemical reactions by other pollutants such as SO_x, NO_x and volatile organic compounds. Airborne particulate matter with a diameter larger than 2.5 micrometres and smaller than 10 micrometres is commonly referred as PM_{10} and corresponds to inhalable coarse particles.

Air quality models have been a valuable tool to predict the level of toxicity or injury caused by atmospheric pollution on human population. Modelling simulation results are typically used in design of urban environmental contingency programs and emergency response plans. Because of its complexity in the interaction between several factors affecting pollutants dispersion, the use of stochastic modelling has been increased over the past years. In particular,

This paper is part of the Proceedings of the 24th International Conference on Modelling, Monitoring and Management of Air Pollution (Air Pollution 2016)
www.witconferences.com

ISSN: 1743-7601 (paper format), ISSN: 1743-761X (online), http://www.witpress.com/journals
DOI: 10.2495/SDP-V11-N4-566-577

Artificial Neural Networks (ANNs) have gained interest in mathematical modelling of uncertainties in complex systems [1, 2]. The ANNs are generic frameworks that have shown great capabilities in modelling highly nonlinear systems. Moreover, ANNs are considered to be global nonlinear approximations; this means that provided that the network structure is sufficiently large, any continuous function can be approximated within an arbitrary accuracy by carefully choosing the parameters of the network. Moreover, the ANNs are capable to deal successfully with poor data quality, handling better heterogeneous data, missing data and non-standard noise [3, 4].

In air quality modelling, several ANN topologies have been used to predict concentrations of major air pollutants such as airborne particulate matter, ozone and oxides mainly produced during fuel combustion [5–7]. In particular, several studies of ANN forecasting of the airborne particulate matter concentration have been published in the last years [8–10]. Several studies have proposed hybrid models, assembling ANN models with either stochastic or deterministic approaches. In air pollution predictions, two main hybrid model types can be distinguished; firstly, ANN models using the genetic algorithm or Levenberg-Marquardt schemes as optimisation models [11, 12], and secondly, ANN models combined with a statistical model, either for improvement of the ANN forecast accuracy as in the case of ARIMA models [13], or for selecting of input variables used during the learning phase such as principal component analysis [14, 15].

In Mexico, the air pollution index of Mexico City Metropolitan Area (IMECA, for its acronym in Spanish) has been applied to other urban areas such as Monterrey and Guadalajara as well. The purpose of IMECA is to notify the population about the air pollution levels, the probable health effects and the recommended protection actions. The IMECA was formerly developed on the basis of the USEPA Pollutant Standards Index, and it has been recently updated according to the Air Quality Index developed by USEPA in terms of ambient air quality standards for six criteria pollutants. Nowadays, the IMECA is calculated with the concentrations of the major air pollutants monitored along the metropolitan areas but it has not been estimated via stochastic models for the purpose of prevention of the adverse health effects.

This study aims at proposing an innovative cluster-based ANN model to predict the maximum of the 24-h moving average of PM_{10} concentration on the next day, in the three largest metropolitan areas of Mexico.

2 MATERIALS AND METHODS

2.1 The CLASO model description

The CLASO model is a self-organised, real-time learning neural network, which is constructed initially by a unique input layer, without hidden layers (Fig. 1). The output layer is created with a-priori known classes of each pattern, and it is set once the hidden layers have been defined during network construction (Fig. 1a). The network builds its topology via a process of pattern classification by using an historical database. This process is made based on a supervised clustering technique, assigning a class to each centroid of the hidden layer, using the Euclidean distance as a hierarchical clustering criterion (Fig. 1b). In order to accomplish this, a distance matrix is generated before the neural network model is constructed. The matrix allows determining efficiently the initial radius of centroid to perform the classification. During the centroid creation, the weighting matrix is updated representing the real-time learning. In the pattern classification stage, patterns that were classified by centroids with

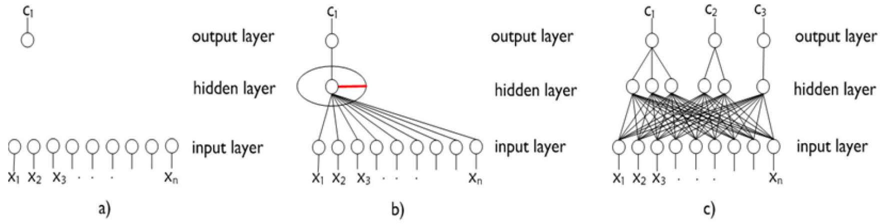

Figure 1: Construction stages of CLASO model, (a) Creation of the output layer, (b) Use of Euclidean distance as a clustering criterion, (c) Constructed ANN.

poor activity can be re-classified via the subsequent presentation to the neural network. This process stops when the number of centroids with poor or null activity remains constant after two consecutive epochs (Fig. 1c). Once the network is constructed, the forecasting is accomplished by using a non-supervised association process of similar patterns.

2.2 Data collection and applicable air quality regulations

The historical database of hourly PM_{10} concentrations during 2009 for the three largest metropolitan areas of Mexico: Monterrey, Guadalajara and Valley of Mexico, was obtained from The National Information System of Air Quality (SINAICA, by its acronym in Spanish). The SINAICA is a federal programme whose purpose is to collect and make public the meteorological and air quality parameters acquired by the automated air monitoring systems of the most populated and industrialised cities in Mexico. The air monitoring stations analysed here were selected on the basis of two criteria: the availability of complete historical data and the spatial representativeness for performing realistic regional assessments. In order to accomplish with temporal completeness and representativeness criteria, the 24-h average PM_{10} concentration must be calculated using a minimum of 75 percent of the hourly PM_{10} concentrations (Mexican regulations: NADF-009-AIRE-2006 and NOM-025-SSA1-1993, in accordance with the U.S. regulation, 40 CFR Pt. 50. App. K. ed. 2006, and the European regulation, Directive 2008/50/EC). Concerning the limit values for the protection of human health, the Mexican maximum value for the daily average PM_{10} concentration is 120 mg/m^3 (NOM-025-SSA1-1993), in contrast to the more tolerant U.S. regulation with a maximum permissible limit of 150 mg/m^3 (40 CFR Pt. 50.6. 2006 until current edition) and the more stringent European limit value, being of 50 mg/m^3 (Directive 2008/50/EC). Similarly, the Mexican maximum value for the annual average PM_{10} concentration is 50 mg/m^3 comparable to the European annual average limit of 40 mg/m^3.

2.3 Classical regression models and statistical goodness-of-fit parameters

As classical regression models, the Box-Jenkins time series analysis in the type of autoregressive, integrated, and moving average (ARIMA) models were selected to be compared with CLASO ANN model in the forecast performance of the maximum value of 24-h average PM_{10} concentration on the next day. The Box-Ljung test was applied to the residual series from ARIMA (p, d, q) models to determine the randomness of residuals as a criterion for model selection. In case of CLASO model, a unique hidden layer was considered, and the patterns were constituted by values of 24-h moving average PM_{10} concentration at 0, 6, 12

Table 1: Statistical goodness-of-fit parameters used for model comparison.

PARAMETER	EXPRESSION*
Pearson's Product-Moment Correlation Coefficient	$r=\dfrac{\sum_{i=1}^{N} O_i P_i - N\bar{O}\bar{P}}{N\sigma_O\sigma_P}$
Sum of Squared Errors	$SSE=\sum_{i=1}^{N}\left(O_i-P_i\right)^2$
Normalized Mean Square Error	$NMSE=\dfrac{\overline{\left(O-P\right)^2}}{\bar{O}\bar{P}}$
Fractional Bias	$FB=2\dfrac{\bar{O}-\bar{P}}{\bar{O}\bar{P}}$
Willmott's Index of Agreement	$IA=1-\dfrac{\sum_{i=1}^{N}\left(P_i-O_i\right)^2}{\sum_{i=1}^{N}\left[\lvert P_i-\bar{O}\rvert+\lvert O_i-\bar{O}\rvert\right]^2}$

* O_i and P_i denote observed and predicted values, respectively, and refer to the mean of observed and predicted values, N is the number of observed values, s_O and s_P symbolise the standard deviation of observed and predicted values.

and 18 hours, and the minimum and maximum values of PM_{10} concentration as well. On the other hand, the ARIMA models used values of the 24-h average PM_{10} concentration. Both ARIMA and CLASO models for each meteorological station in the three cities were calibrated by using data of 24-h moving average PM_{10} concentration from January to November 2009. Afterwards, these models were tested employing data from December 2009 and were compared using several statistical goodness-of-fit parameters (Table 1) such as Pearson's product-moment correlation coefficient (r), Sum of Squared Errors (SSE), Normalized Mean Square Error ($NMSE$), Fractional Bias (FB) and Willmott's Index of Agreement (IA).

3 RESULTS AND DISCUSSION

3.1 Temporal variations of 24-h average PM_{10} concentrations

Temporal variations of 24-h moving average PM_{10} concentration in the three largest metropolitan areas of Mexico during 2009 were analysed. A characteristic temporal variation was created using PM_{10} concentration values collected from four selected monitoring stations of Monterrey metropolitan area (Fig. 2). In this urban area, daily average limit value exceedances of PM_{10} concentration are extremely frequent, for example, the limit value was exceeded 62 times in the northwest region, 28 times in the centre, 94 times in the southwest and 33 times in the northeast, summarising 217 times in 2009 (60% of the year). Furthermore, the southwest region exhibited the maximum daily average PM_{10} concentration values greater than 250 mg/m^3. Similarly, the Mexican limit value of the annual average PM_{10} concentration was exceeded at the four monitoring stations, that is, annual average values of 96 mg/m^3 in

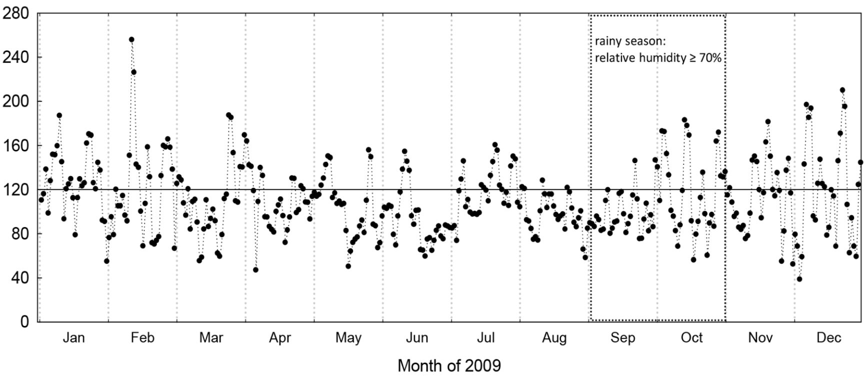

Figure 2: Characteristic temporal variation of PM_{10} concentration in Monterrey during 2009. Rainy season ranged from September 2, to November 1.

the northwest region, 83 mg/m^3 in the centre, 101 mg/m^3 in the southwest and 82 mg/m^3 in the northeast. It should be noticed that the rainy season, arbitrarily defined here as the time interval exhibiting a daily average moisture equal or greater than 70%, showed a negligible effect on the PM_{10} concentration values.

With respect to the PM_{10} concentration in Guadalajara metropolitan area during 2009, the characteristic temporal variation based on four selected monitoring stations is shown in Fig. 3. In contrast to temporal variation in Monterrey, 24-h average limit value exceedances could be associated with other specific events pointed out in Fig. 3.

The temporal behaviour of PM_{10} concentration in the Valley of Mexico urban area is illustrated in Fig. 4. Again, the typical variation was constructed using the daily average values of four representative monitoring stations, for instance, north and northeast regions recorded maximum daily average values of 150 mg/m^3, and they exhibited 18 limit value exceedances in the year. Concerning the annual average PM_{10} concentration value, the north, northeast and centre regions exceeded the Mexican limit value recording 61, 60 and 56 mg/m^3, respectively. The exceedances in Valley of Mexico area were frequently associated with thermal inversion phenomena that occur commonly in winter.

3.2 Classical regression models *vs.* CLASO model

In Fig. 5, forecasted values of 24-h average PM_{10} concentration for ARIMA and CLASO models were compared using data from the Monterrey area recorded during December 2009. In general, ARIMA models exhibited a predominant overprediction of those daily average PM_{10} concentration values lower than Mexican daily average limit value (120 mg/m^3), and a predominant underprediction with those daily average PM_{10} concentration values greater than Mexican daily average limit value. This trend can be clearly observed in three regions of Monterrey area: Northwest (Fig. 5a), Centre (Fig. 5b) and Northeast (Fig. 5d). In order to confirm this trend, the Fractional Bias, a statistical parameter measuring deviation of the expected values was computed. From Table 2, this parameter presented negative values for the three regions of Monterrey mentioned above, indicating a global overprediction of expected values by using the ARIMA models. In contrast, the PM_{10} concentration forecast using the CLASO model showed positive values of *FB* for the four monitoring stations of

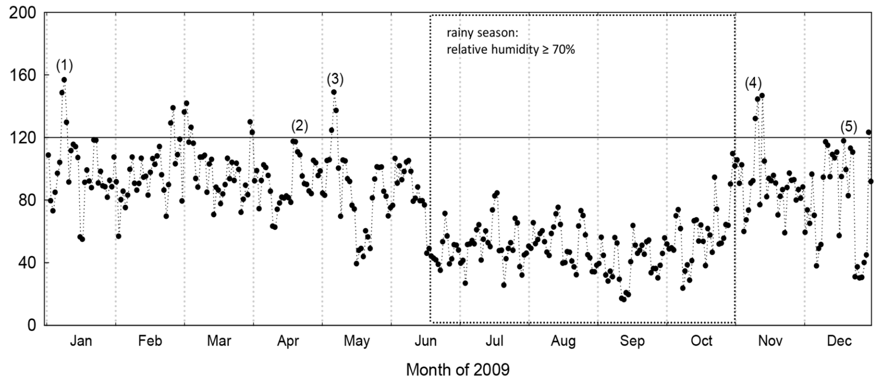

Figure 3: Characteristic temporal variation of PM_{10} concentration in Guadalajara during 2009. Rainy season ranged from June 17 to November 1. Numbers on the graph indicate specific events generating airborne particulate matter pollution: (1) Thermal inversion and low temperatures, (2) Four successive wildfires in La Primavera Forest, (3) Photochemical smog, high solar radiation, highest temperatures of the year and calm wind, (4) Clandestine operation of traditional brick factories, (5) Fireworks and bonfires celebrating the New Year.

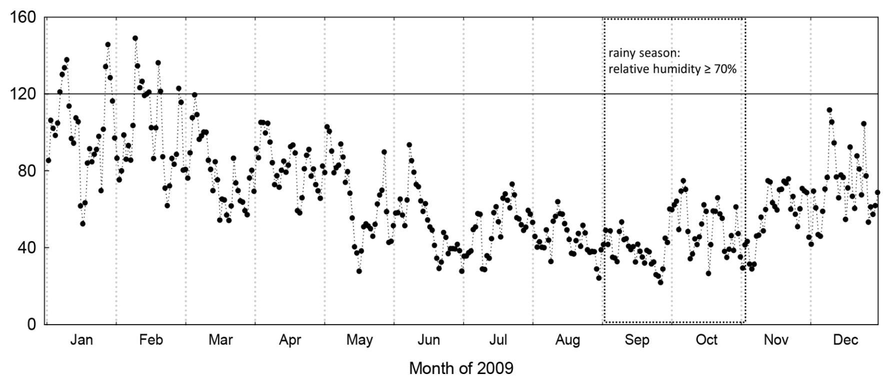

Figure 4: Characteristic temporal variation of PM_{10} concentration in Valley of Mexico during 2009. Rainy season ranged from September 3 to November 4.

Monterrey, resulting in a global underprediction of the expected values; however, the CLASO model underprediction is significantly smaller than the overprediction obtained with ARIMA models.

The above requires an additional confirmation, therefore, a new criterion called the percent of acceptable predictions (Table 2) would be appropriate here to define the proportion of days in a month whose absolute value of the difference between observed and predicted values is lower than 5 mg/m^3. According to this criterion, the ARIMA models exhibited a poor proportion of acceptable predictions, with 6% (southwest region) being the minimum and 26% (NE region) being the maximum percent of acceptable predictions. Overall, the percent of

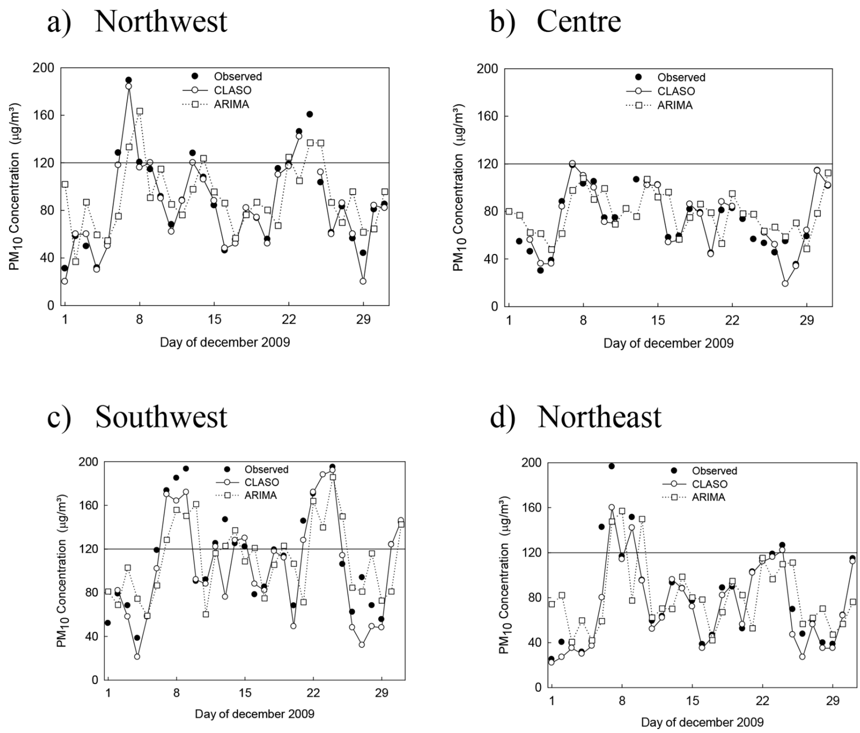

Figure 5: Comparison of ARIMA and CLASO models in the prediction of observed temporal variations of 24-h average PM_{10} concentration in four meteorological stations of Monterrey during December 2009.

acceptable predictions for Monterrey metropolitan area using ARIMA models was of 15.6%. On the other hand, the CLASO model obtained a percent range of acceptable predictions between 47% (southwest region) and 80% (northwest region), with an overall proportion of acceptable predictions for Monterrey metropolitan area of 68.1%. It should be noticed that, both ARIMA and CLASO models, showed a poor forecast of the observed daily average PM_{10} concentration values recorded in Monterrey SW region.

Forecast capabilities for ARIMA and CLASO models by using data of Guadalajara recorded during December 2009 is depicted in Fig. 6. The Fractional Biases showed that both ARIMA and CLASO models tend to underpredict the PM_{10} concentration values, although the set of ARIMA models exhibited, in each meteorological station, a minor degree of under-prediction. In relation to acceptable predictions, ARIMA models obtained 31.7% of acceptable predictions in Guadalajara area, varying between 14% and 55% for southeast and north regions, respectively. On the other hand, using the CLASO model, 74.5% of acceptable predictions in Guadalajara were obtained, varying between 62% and 87% for south and southeast regions, respectively. It is worth mentioning that the Mexican limit value of 24-h average PM_{10} concentration was not exceeded during December 2009 in Guadalajara metropolitan area.

Table 2: Results of statistical goodness-of-fit parameters for ARIMA and CLASO models.

PARAMETER	MONITORING STATIONS BY METROPOLITAN AREA											
	Monterrey				Guadalajara				Valley of Mexico			
	NW*	CE	SW	NE	N	CE	S	SE	N	CE	S	NE
ARIMA Models												
r	0.61	0.61	0.67	0.57	0.43	0.30	0.30	0.36	0.36	0.70	0.61	0.53
SSE	30.8	20.7	34.7	34.0	7.7	13.4	20.7	19.9	18.2	12.8	14.1	16.8
NMSE	0.12	0.07	0.10	0.18	0.05	0.05	0.05	0.03	0.09	0.05	0.07	0.06
FB	−0.04	−0.05	0.02	−0.01	0.06	0.01	0.01	0.02	0.04	0.03	0.03	0.03
IA	0.75	0.87	0.80	0.74	0.64	0.94	0.95	0.98	0.84	0.82	0.78	0.82
Days predicted	31	29	31	31	31	25	24	21	28	31	31	30
% acceptable pred.	13	21	6	26	55	28	21	14	21	26	39	27
CLASO Neural Network Model												
r	0.98	0.95	0.93	0.95	0.49	0.78	0.96	0.92	0.97	0.83	0.83	0.94
SSE	6.7	8.5	20.4	14.8	9.8	12.3	6.8	10.5	4.5	9.0	11.6	5.7
NMSE	0.01	0.01	0.03	0.04	0.08	0.04	0.00	0.01	0.00	0.02	0.04	0.00
FB	0.02	0.01	0.09	0.10	0.14	0.07	0.01	0.07	0.02	0.04	0.08	0.00
IA	0.99	0.99	0.96	0.96	0.69	0.96	0.99	0.99	0.99	0.99	0.96	0.99
Days predicted	30	25	30	31	31	23	23	21	22	19	26	23
% acceptable pred.	80	72	47	74	71	78	87	62	86	74	88	78

*NW: Northwest, CE: Centre, SW: Southwest, NE: Northeast, N: North, S: South, SE: Southeast

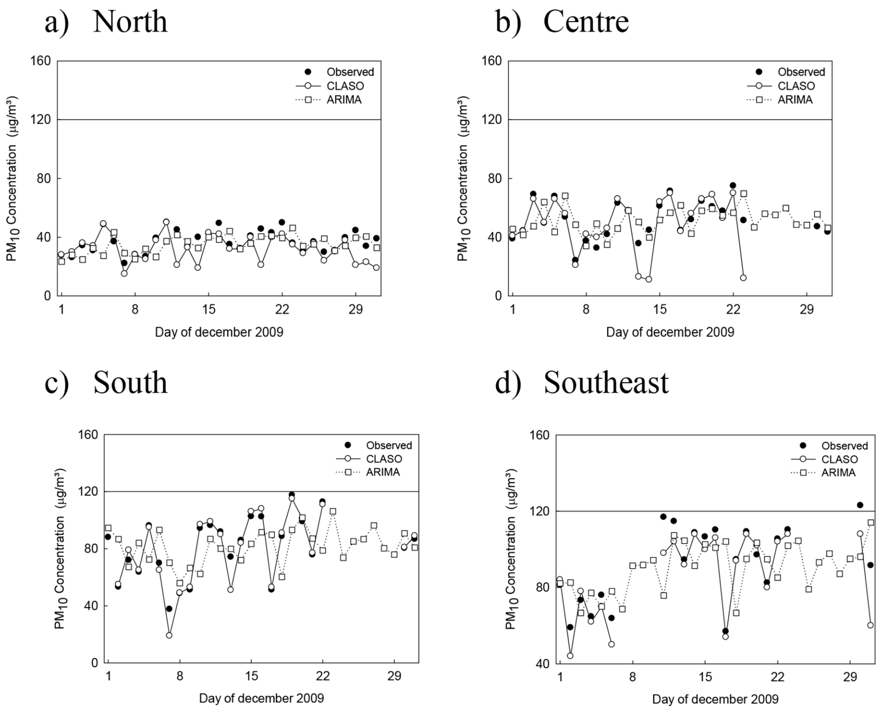

Figure 6: Comparison of ARIMA and CLASO models in the prediction of observed temporal variations of 24-h average PM_{10} concentration in four meteorological stations of Guadalajara during December 2009.

A comparison between ARIMA and CLASO models in the prediction of temporal variations for a 24-h average PM_{10} concentration in the Valley of Mexico is shown in Fig. 7. Similarly as in Guadalajara area, *FB* values indicated that both models generally underpredict the maximum of the 24-h moving average PM_{10} concentration in Valley of Mexico during December 2009. Regarding the proportion of acceptable predictions, the ARIMA models produced the 28.3% of acceptable predictions in Valley of Mexico area, varying between 21% and 39% for north and south regions, respectively. In contrast, the CLASO model showed a significant number of accurate predictions, resulting in 82.2% in Valley of Mexico area, varying between 74% and 88% for centre and south regions, respectively.

Besides the fractional bias and the percent of acceptable predictions, other statistical parameters were calculated for the purpose of comparing ARIMA and CLASO models (Table 2). According to the computed Pearson's *r* coefficients, ARIMA models were incapable of properly predicting 24-h average PM_{10} concentration values obtaining coefficient values as low as 0.30 for the southeast region in Guadalajara, and the highest coefficient value was calculated for the data correlation of centre region in Valley of Mexico. On the other hand, the *r* coefficients estimated using the CLASO model showed a better prediction; that is, the computed minimum value estimated for the north region of Guadalajara was 0.49, and the maximum value estimated for the northwest region in Monterrey area was 0.98. It should be noticed that, according to the *r* coefficient, the data collected in the Monterrey metropolitan area exhibited the best correlation by using the CLASO model achieving values between 0.92

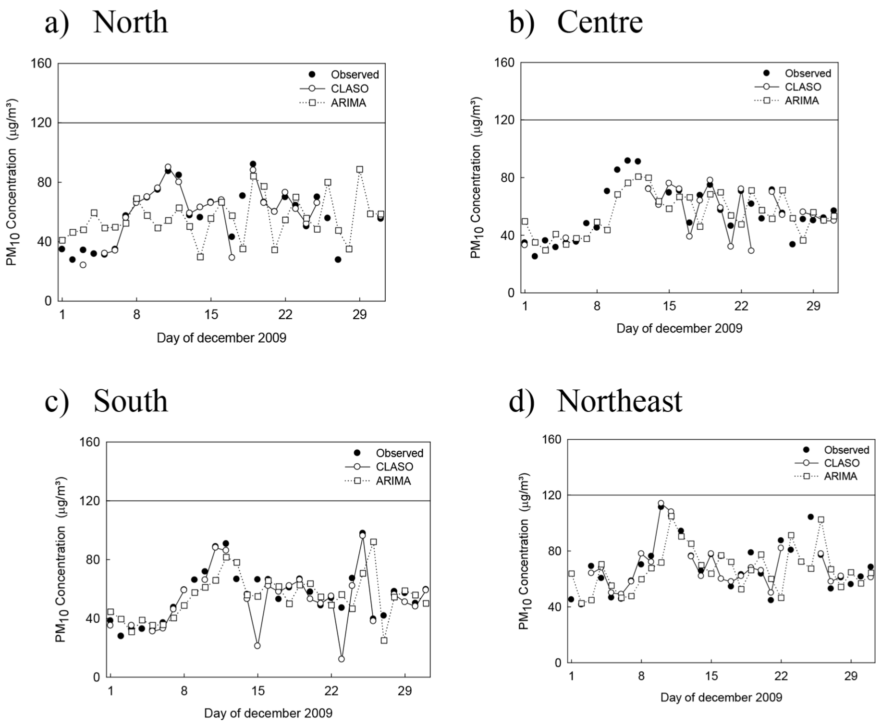

Figure 7: Comparison of ARIMA and CLASO models in the prediction of observed temporal variations of 24-h average PM_{10} concentration in four meteorological stations of the Valley of Mexico during December 2009.

and 0.98. In contrast, the data obtained from Guadalajara area showed a mixed correlation, being a poor correlation for the data of north region ($r = 0.49$) and a suitable correlation for data of the south region ($r = 0.96$). For the data fitting of the three major urban areas of Mexico, CLASO model exhibited a remarkable fitness.

4 CONCLUSIONS

A novel cluster-based artificial neural network model, referred to as CLASO, has been proposed. CLASO model is a dynamic neural network, which is subjected to a learning phase and the number of neurons in each hidden layer is also dynamically defined.

CLASO model produced more accurate predictions for the maximum of 24-h moving average of PM_{10} concentration than the ARIMA models, although the latter exhibited a minor tendency to underprediction. CLASO has demonstrated acceptable predictions for the maximum of 24-h moving average of PM_{10} concentration in the three Mexican largest urban areas by using exclusively regressive PM_{10} concentrations. This feature is especially useful in historical databases from air quality monitoring systems with lack of information either incomplete or unrepresentative meteorological records. In the future, CLASO model will be tested using other metrics as clustering criterion and an optimisation scheme focused to improve the clustering creation may be developed. As the ultimate goal, CLASO could be applied to other major pollutants used in the Mexican air quality index.

REFERENCES

[1] Jordan, M.I. & Bishop, C.M., Neural networks. *ACM Computing Surveys*, **28**(1), pp. 73–75, 1996.
http://dx.doi.org/10.1145/234313.234348

[2] Ordieres, J.B., Vergara, E.P., Capuz, R.S. & Salazar, R.E., Neural network prediction model for fine particulate matter ($PM_{2.5}$) on the US-Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Modelling & Software*, **20**, pp. 547–559, 2005.
http://dx.doi.org/10.1016/j.envsoft.2004.03.010

[3] Alexandridis, A., Patrinos, P., Sarimveis, H. & Tsekouras, G., A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models. *Chemometrics and Intelligent Laboratory Systems,* **75**, pp. 149–162, 2005.
http://dx.doi.org/10.1016/j.chemolab.2004.06.004

[4] Cherkassky, V., Krasnopolsky, V., Solomatine, D.P. & Valdes, J., Computational intelligence in earth sciences and environmental applications: issues and challenges. *Neural Networks*, **19**, pp. 113–121, 2006.
http://dx.doi.org/10.1016/j.neunet.2006.01.001

[5] Gómez-Sanchís, J., Martín-Guerrero, J.D., Soria-Olivas S., Vila-Francés, J., Carrasco, J.L. & Del Valle-Tascón, S., Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration. *Atmospheric Environment*, **40**, pp. 6173–6180, 2006.
http://dx.doi.org/10.1016/j.atmosenv.2006.04.067

[6] Kurt, A. & Oktay, A.B., Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, **37**, pp. 7986–7992, 2010.
http://dx.doi.org/10.1016/j.eswa.2010.05.093

[7] Pfeiffer, H., Baumbach, G., Sarachaga-Ruiz, L., Kleanthous, S., Poulida, O. & Beyaz, E., Neural modelling of the spatial distribution of air pollutants. *Atmospheric Environment*, **43**, pp. 3289–3297, 2009.
http://dx.doi.org/10.1016/j.atmosenv.2008.05.073

[8] Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F. & Brasseur, O., A neural network forecast for daily average PM_{10} concentrations in Belgium. *Atmospheric Environment,* **39**, pp. 3279–3289, 2005.
http://dx.doi.org/10.1016/j.atmosenv.2005.01.050

[9] Slini, T., Kaprara, A., Karatzas, K. & Moussiopoulos, N., PM_{10} forecasting for Thessaloniki, Greece. *Environmental Modelling & Software* **21**, pp. 559–565, 2006.
http://dx.doi.org/10.1016/j.envsoft.2004.06.011

[10] Caselli, M., Trizio, L., De Gennaro, G. & Lelpo, P., A simple feedforward neural network for the PM_{10} forecasting: comparison with a radial basis function network and a multivariate linear regression model. *Water Air & Soil Pollution*, **201**, pp. 365–377, 2009.
http://dx.doi.org/10.1007/s11270-008-9950-2

[11] Grivas, G. & Chaloulakou, A., Artificial neural network models for prediction of PM_{10} hourly concentrations, in the Greater area of Athens, Greece. *Atmospheric Environment*, **40**, pp. 1216–1229, 2006.
http://dx.doi.org/10.1016/j.atmosenv.2005.10.036

[12] Nejadkoorki, F. & Baroutian, S., Forecasting extreme PM_{10} concentrations using artificial neural networks. *International Journal of Environmental Research*, **6**(1), pp. 277–284, 2012.

[13] Díaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G. & Moncada-Herrera, J.A., A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of temuco, Chile. *Atmospheric Environment,* **42**, pp. 8331–8340, 2008.
http://dx.doi.org/10.1016/j.atmosenv.2008.07.020

[14] Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A. & Kolehmainen, M., Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Science of the Total Environment*, **409**, pp. 1266–1276, 2011.
http://dx.doi.org/10.1016/j.scitotenv.2010.12.039

[15] Paschalidou, A.K., Karakitsios, S., Kleanthous, S. & Kassomenos, P.A., Forecasting hourly PM_{10} concentration in cyprus through artificial neural networks and multiple regression models: implications to local environmental management. *Environmental Science & Pollution Research,* **18**, pp. 316–327, 2011.
http://dx.doi.org/10.1007/s11356-010-0375-2