International Journal of
Science and Engineering Investigations vol. 7, issue 73, February 2018
ISSN: 2251-8843

Artificial Neural Network Intelligent Prediction System for Air Contaminants Monitoring

Jesús Gabriel Martínez Rodríguez[1], José-Isidro Hernández-Vega[2], Sergio Nesmachnow[3], Jonam L. Sánchez Cuevas[4]
[1,2,4]División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Nuevo León, Nuevo León, México
[3]Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay
([1]jesusgabrielmr@gmail.com, [2]jose.isidro.hernandez@itnl.edu.mx, [3]sergion@fing.edu.uy, [4]jonam.leonel.sanchez@itnl.edu.mx)

Abstract- This paper presents the application of an Artificial Neural Network (ANN) to predict air contaminants behavior. Input data were obtained from the Atmospheric Deposit Network (REDDA) in Mexico City. The objective of the ANN was to predict Ozone (O3) concentration, Network training was based on environment conditions and represented as a function of time. For the experimental analysis, real data obtained from REDDA, in the period from January 1 to October 31, 2017 were used. The design of the Network was under a three-layer feed-forward architecture using "nnet" package of the R software. The results obtained showed that the ANN can infer the concentration of the pollutant studied based on its training, it is graphically represented and compared with real data taken from the REDDA, reaching the conclusion that the ANN designed predicts similar behavior of the contaminants to a real sampling in a period of time.

Keywords- *Artificial Neural Network, ANN, Pollution Prediction*

I. INTRODUCTION

Air pollution is one of the main environmental and public health problems in the world due to the effects that pollutants have on human health and ecosystems. It is a problem complex to evaluate and control due the amount and variety of emitting sources, the transformation that pollutants suffer in the atmosphere and the harm they cause to living beings [1].

The most widely used tools for the estimation of atmospheric pollutants concentration are the mathematical dispersion models, to evaluate the dispersion of pollutants there is a variety of software such as ADAM (Air Force Dispersion Assessment Model) y AERMOD Modeling System (AERMOD_Model) [2].

ANN have been used for prediction as intelligent systems in medical diagnoses [3], control systems [4]. It is important to make pollutant prediction to generate and provide needed information to scientists, legislators and planners so that they make the right decisions in favor of the management and improvement of the environment [5] In Nuevo León, in the past two years, the established limits of pollution levels have been exceeded during 60% of the days of the year according to the Sustainable Development Secretariat [6]. One of the problems arising from the monitoring of pollutants is that fixed sources are not able to identify on time when permitted limits are being exceeded. In section II, the prediction of pollutant problem using ANN is presented. The ANN paradigm is described in section III of the article. The details of the design and implementation of the neural network for the prediction of pollutant concentration are shown in section IV. Section V reports the experimental analysis carried out. The details of design and implementation of the neural network for prediction of pollutant concentration are shown in section VI. Future work is presented in section VII.

II. POLLUTANT PREDICTION PROBLEM USING ANN

Different models of ANN have been used over the years for prediction, pattern classification and function approximation purposes. The proposed models like Multilayer Perceptron with a hidden layer predict a concentration between 1 and 48 hours in advance. Some models use historical pollution data plus meteorological information, traffic information, day of the week [7]. The air quality forecasting system of Mexico City was developed by the Secretariat of the Environment of Mexico City collaborating in the project the Barcelona Supercomputing Center (CBS). The objective of the air quality forecasting system of Mexico City, is to prepare an operational forecast for the next 24 hours of the air quality status, using current knowledge on atmospheric chemistry and dynamics of the region [1]. The model that was developed raises the use of meteorological variables as factors that influence the concentration of pollutants.

A. Description of the problem

With the 8 data: Sampling Time(TM), Ozone (O3) concentration, Solar Radiation (UVA), Atmospheric Pressure (PRES), Relative Humidity (HR), Temperature (TMP), Wind Speed (WSP), Wind Direction (WDR), obtained from the Atmospheric Deposit Network of Mexico City, it is intended to predict the behavior of the Ozone 12 hours and 24 hours, after the sample, thus achieving to anticipate the pollutant concentration value close to the real.

III. ARTIFICIAL NEURAL NETWORKS

Artificial Intelligence (AI) is a field of knowledge composed of a set of techniques based on imitating different human skills and other functions related to the intelligence of a human being in a computational way.

Artificial Neural Networks are mathematical-computational models that have a high capacity for generalization and treatment of both linear and non-linear problems, and do not require an exhaustive knowledge of the distribution of the variables under study [8].

ANN is a Parallel Distributed Processing model (PDP) composed of a large number of processors. It can learn complex non-linear cartographies between an input and output space. It has a set of processing units (neurons), a connectivity pattern (topology), a representation of the environment, a neuronal state (or activation state) [9].

A. Feed-forward Neural Network(FNN):

Feed-forward neural networks, belongs to the class of most studied by the scientific field networks and the most commonly used in various fields of application.

The different types of ANN are distinguished from each other by the number of neurons or nodes that constitute the basic processing element, the network architecture described by the weighted connections between the nodes, and by the training algorithm used to find the parameters of the network [8]. FNN has three types of neuronal layer: input, hidden layer and output.

B. Training of the FNN

The training is based on the back-propagation algorithm by the supervised descendant gradient method, in which the weights are updated using a set of desired inputs and outputs, and the comparison with the real output of the neural network. The back-propagation algorithm consists of two phases: first, an input pattern is applied, which propagates through all the layers that make up the network until producing the output of the network. This output is compared with the desired output and the error committed by each output neuron is calculated. These errors are transmitted backwards, starting from the output layer, to all the neurons in the intermediate layers. Each neuron receives an error that is proportional to its contribution on the total error of the network. Based on the error received, the errors of the synaptic weights of each neuron [10] are adjusted.

IV. DESIGN AND IMPLEMENTATION OF THE PROPOSED NEURAL NETWORK

This section presents the details of the design and implementation of the proposed Neural Network.

A. R Language

R is an environment and a statistical analysis programming language. It is an implementation of the S open source language. It is one of the most used languages used in research by the statistical community, being also popular in other fields such data mining, biomedical research, bioinformatics, financial mathematics. To this contributes the possibility of loading different libraries or packages with the functionalities of calculation and graphics [9].

Package nnet is a feed-forward neural networks software with a simple hidden layer, and for polynomial linear log models [11].

B. Methodology

The first step was to collect the data from the Atmospheric Deposit Network of Mexico City, then data were uploaded to the R Studio software with the "nnet" package.75% of data were normalized to train a one hidden layer network with 10 neurons, 8 input parameters and one output. 670 iterations were performed for the convergence of the network. Due to the output data are used as a reference, is a supervised learning. Subsequently, 25% of the normalized data were used for testing. Figure 1 shows the neural network to be trained for the forecast of pollutants concentration with the collected data. The first layer consists of input variables: Sampling Time(TM), Ozone (O3) concentration, Solar Radiation (UVA), Atmospheric Pressure (PRES), Relative Humidity (HR), Temperature (TMP), Wind Speed (WSP), Wind Direction (WDR), the hidden layer contain the neurons H1,H2…H10, B1 is the hidden layer Bias, O1 is the output layer and B2 is the output layer Bias.

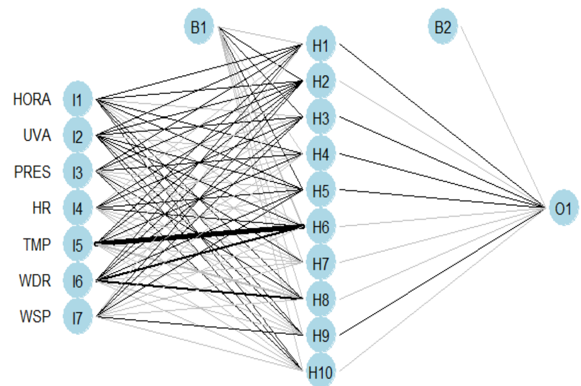

Figure 1. Neural Network generated to predict ozone concentration.

V. EXPERIMENTAL RESULTS

This section presents the details of design and implementation of the proposed neural network.

A. Platform for development and execution

For the neural network training RStudio version 1.0.136 – © 2009-2016 software were used, with the 'nnet' package. The equipment used for the experimental analysis has an Intel Pentium CPU N3710 processor at 1.60 GHz, 8 GB of RAM and Windows 10 64-bit operating system.

B. Instance of the problem

To train the neural network data obtained from Atmospheric Deposit Network of Mexico City were used corresponding to the period from January 1 to October 31, 2017. The input variables used to generate the networks are:

- Hour (0 – 23)
- Ozone concentration (parts per billion)
- Solar Radiation (kWh/m2)
- Atmospheric Pressure (mmHg)
- Relative Humidity (percentage)
- Temperature (Celsius degrees)
- Wind Velocity (km/h)
- Wind Direction (degrees)

The ozone concentration parameter was used as a predictor in the neural network.

C. RESULTS

In the experimental analysis, from the neural network, the results obtained in the first case were graphs of ozone concentration versus ozone concentration simulated by the network at 12 hours after the sample, as shown in figure 2. Figure 3 shows the concentration of ozone vs actual ozone concentration 12 hours after the sample. Figure 4 represents the concentration of ozone vs concentration of ozone simulated by the network 24 hours after the sample. Figure 5 represents the concentration of ozone vs actual ozone concentration 24 hours after the sample.

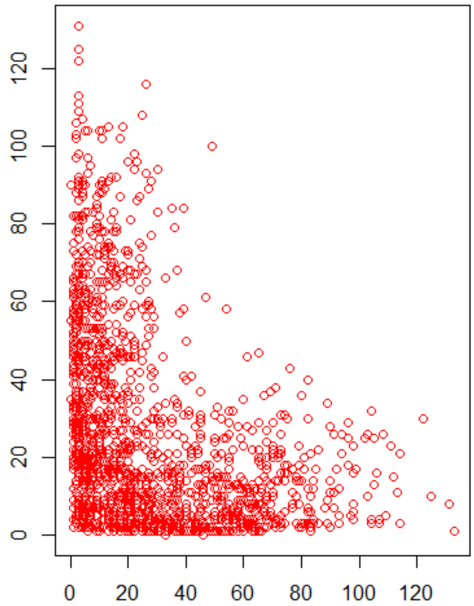

Figure 3. Real data graphic, data corresponding to the ozone concentration in x axis versus ozone concentration 12 hours after the sample in y axis.

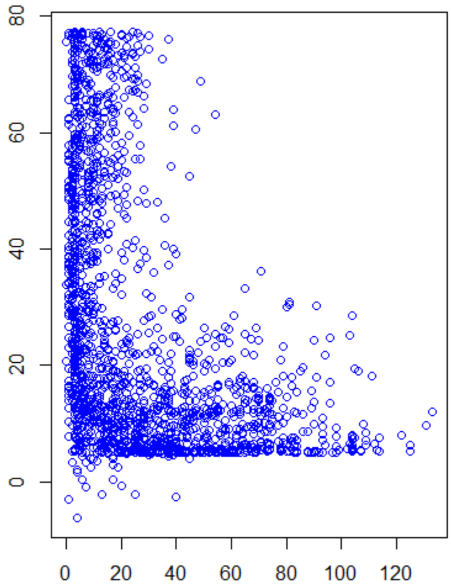

Figure 2. Simulated data graphic, data corresponding to the ozone concentration in x axis versus ozone concentration 12 hours after the sample in y axis.

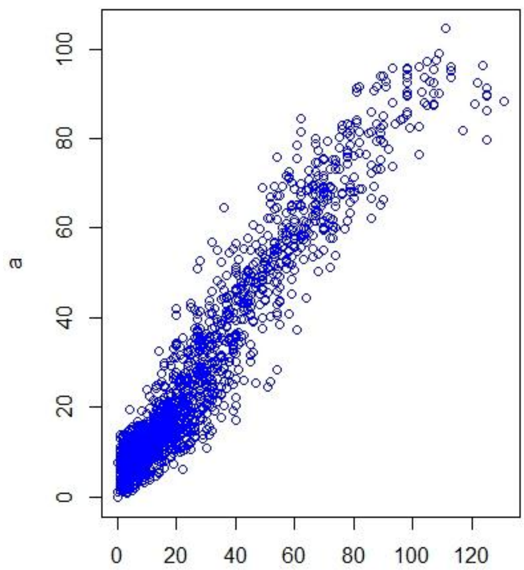

Figure 4. Simulated data graphic, data corresponding to the ozone concentration in x axis versus ozone concentration 24 hours after the sample in y axis.

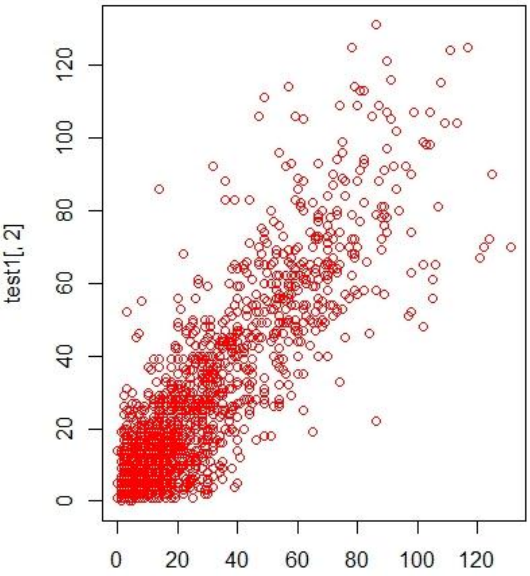

Figure 5. Real data graphic, data corresponding to the ozone concentration in x axis versus ozone concentration 24 hours after the sample in y axis.

Observing the graphs of the ozone concentration simulated by the artificial network against the real measure, a similarity between them is noted, being bigger in the 12 hours after the sample in comparission to the 24 hours after the sample case.

VI. FUTURE WORK

Apply ANN to predict the concentration of several criteria pollutants from the data provided by sensors assembled in an air monitoring unit to generate a collection of contaminants data gathered in real time, so that the network can be trained and make forecasts in a specific primary source of pollutants such as an industrial chimney.

REFERENCES

[1] Secretaría del medio Ambiente, "Red de depósito atmosférico," 2017. [Online]. Available: http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBk%27.

[2] N. A. and S. Administration, "NASA's Global Change Master Directory," 2018. [Online]. Available: https://gcmd.nasa.gov/index.html.

[3] J. J. Sprockel, J. J. Diaztagle, W. Alzate, and E. González, "Redes neuronales en el diagnóstico del infarto agudo de miocardio," *Rev. Colomb. Cardiol.*, vol. 21, no. 4, pp. 215–223, 2014.

[4] K. J. Hunt, D. Sbarbaro, R. Żbikowski, and P. J. Gawthrop, "Neural networks for control systems-A survey," *Automatica*, vol. 28, no. 6, pp. 1083–1112, 1992.

[5] Instituto nacional de ecología, "Manual 1 Principios de Medición de la Calidad del Aire" 2010

[6] Gobierno de Nuevo León, "Boletin Gobierno Ciudadano," 2018. [Online]. Available: http://www.nuevoleon.gob.mx/noticias/suspende-desarrollo-sustentable-20-pedreras.

[7] M. G. Januch, "Aplicación de Técnica de Inteligencia Artificial a la Predicción de Contaminantes Atmosféricos," 2012.

[8] W. N. Venables and B. D. Ripley. "Modern Applied Statistics with S." Springer, New York, fourth edition, 2002.

[9] R Studio. Rstudio: integrated development environment for r. RStudio Inc, Boston, Massachusetts, 2012

[10] L. F. Bertona, "Entrenamiento de redes neuronales basado en algoritmos evolutivos," *Grado, Tesis Ing. Informática, Univ. BUENOS AIRES*, p. 253, 2005.

[11] B. Ripley and W. Venables, "R Package 'nnet': Feed-Forward Neural Networks and Multinomial Log-Linear Models," CRAN R Core Team (2016). R A Lang. Environ. Stat. Comput. R Found. Stat. Comput. Vienna, Austria, no. © The R Foundation., "R Core Team (2016).M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

Jesús Gabriel Martínez Rodríguez: Computer Systems Engineer with specialty in Databases, student of master's Degree in Mechatronics, currently working on the Intelligent Mechatronic Systems Research line at the Instituto Tecnológico de Nuevo León.

José Isidro Hernández Vega: Computer Systems Engineer with master's Degree in computer science. Professor with recognition of Desirable Profile by PRODEP since 2012. Member of the Academic of Sustainable Technologies and Intelligent Mechatronics Systems Research Group**.**

Sergio Nesmachnow: Has a Ph.D. degree in Computer Science from Universidad de la República, Uruguay. He currently holds an Aggregate Professor position in the Numerical Center (CeCal) at Computer Science Institute, Engineering Faculty. His main research interests are parallel and distributed computing, scientific computing, evolutionary algorithms and metaheuristics, and numerical methods.

Jonam Leonel Sánchez Cuevas: Electrical Engineer with master degree in Electronic Science by Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). Currently working in Intelligent Mechatronics Systems research line and member of the Academic of Sustainable Technologies and Intelligent Mechatronics Systems Research Group.