Automatic model based on Artificial Neural Networks to predict the emissions of Carbon Dioxide (CO_2)

ERIK F MÉNDEZ
Systems Department, University of the Andes UNIANDES
Santo Domingo, ECUADOR
us.erikmendez@uniandes.edu.ec

JOSÉ HERRERA
University of the Andes UNIANDES
Santo Domingo, ECUADOR
Joseherrerab79@gmail.com

GABRIELA MAFLA
Indepedent Researcher
Riobamba, ECUADOR
gabriella_1912@hotmail.com

Abstract: This research work is based exclusively on the application of artificial neural networks, aimed at predicting the CO_2 pollution index. For the design of the ANN, a multilayer network of Backpropagation type has been created and the Levenberg-Marquardt method was used for its training. The neural network consists of three layers: input (Input), hidden (Hidden Layer) and output (Output); the architecture was generated with Matlab software. Good quality results were obtained when the actual values and those predicted by the system were checked, demonstrating that it is a highly accepted model for prediction, favoring the planning processes.

Key-Words: carbon dioxide prediction, artificial neural networks, conceptual model, Backpropagation, Levenberg-Marquardt method, pollution

1 Introduction

Air pollution is currently one of the most severe environmental problems worldwide. Economic growth and urbanization, associated with the development of various activities of man, result in an intense consumption of fossil fuels, generating various Greenhouse Gas (GHG) such as carbon dioxide (CO_2) [1].

Artificial Neural Networks are models of information processing, inspired by the functioning of the brain. They have the ability to learn from experience [2].

This document is based exclusively on applying ANN to predict CO_2 emissions in a given region (Santo Domingo de los Colorados, Ecuador), for which the multilayer perceptron Backpropagation artificial neural network model is used with the Levenberg Marquardt learning. The proposed model uses information on the concentrations of the pollutants and the meteorological variables obtained from the proposed city as a case study, linking the need to assist in analysis, planning and decision-making processes for the aforementioned pollutant reduction.

2 Artificial Neural Networks

They are a branch that composes Artificial Intelligence (A.I.), based on the anatomy of the brain of living beings, they receive all the information to be submitted to evaluations by experts [3]. There are many definitions for ANN, using that proposed by McCoulloch and Pitts [4] [5] it is established that an artificial neural network is a mathematical model composed of an input layer given by the vector X = [x_1, x_2, ... x_n], receive stimuli y is modified by a vector w of synaptic weights that represent the incoming pulses, are evaluated, combined with the threshold θ_j, this function is evaluated in the activation process σ (x) that goes to the output layer y, represented by the Eq. (1) [6]:

$$y_i = \sigma\left(\sum_{j=1}^{N} w_{ij} x_j \pm \theta_i\right)$$

For the application of an artificial neural network to a certain problem, the following steps are described [7] [8]:

- Conceptualization of the model, where the inputs and outputs of the information are marked.
- Adequacy of information, learning patterns are ordered and constructed for later validation.
- Learning phase, once appropriate the information is added to the inputs of the neural network, this process is repeated in several stages and the outputs are compared with the desired response.
- Validation phase, once the learning stage is completed, information patterns are presented to validate and calculate the error made by the network, representing the satisfaction of the network.
- Generalization phase, once the appropriate neural network is found, it is used as a prediction model to add new inputs and have desired results.

3 Prediction Model Bas ed On Artificial Neural Networks

The prediction model for predicting the carbon dioxide index consists of the following phases, as shown in Fig. 1.

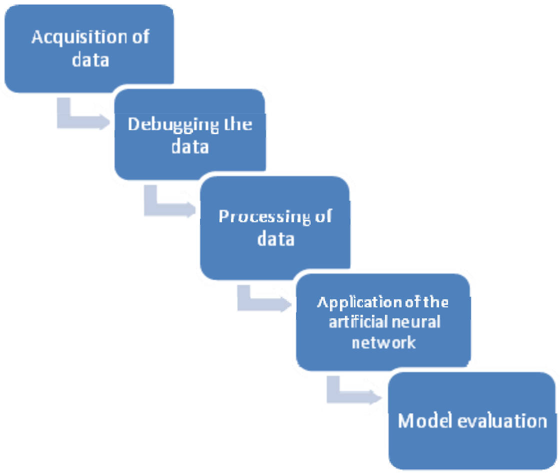

Fig. 1. Prediction model

Phase 1: Acquisition of data. The data is acquired by an electronic CO_2 measuring device, these are reported in parts per million (ppm) with their respective date and time. The designed system stores the information every minute on a micro SD memory card, in a text file (LOG.txt). For the present case study the data acquired from October 1 to December 30 of the year 2018 were considered.

Phase 2: Data analysis. The information acquired may present atypical data that may be out of parameters (300 and 1000 ppm). For this, the average of the data prior to the erroneous data was obtained, in the same way this process was carried out for missing data, with which a reliable database is obtained to train the neural network.

Phase 3: Data processing. The data processing refers to the fact of analyzing and transforming the input and output variables, to speed up the work to the artificial neural network and generate learning patterns. The input data is sectioned by year, month, day and time; while for the output data the values of CO_2 measurements are evidenced.

Phase 4: ANN application. The development of the artificial neural network is done in Matlab software, for which it is necessary to define a series of steps, which are described below [9].

1. Definition of t he training, validation and testing set. The database has 2208 records of carbon dioxide measurements, from which 1824 data are selected for the months of october, november and first 15 days of december of the year 2018; the remaining 15 days will be used to validate the network and determine the prediction reliability.

- Training set, so that the neural network can learn the input patterns and define the weights of the network. The Matlab software automatically selects 1276 data representing 70% of the total set (1824 carbon dioxide measurements).

- Validation set, uses a percentage of information for the final check of the network, Matlab selects 274 data equivalent to 15% of the total.

- Test set, select a certain amount of information to assess the accuracy of the neural network, in this case Matlab chooses 274 data (15%).

2. Selection of the neural network architecture. A multilayer network of Backpropagation type is created, in this process a series of characteristics that an ANN must possess is defined.

- Number of input neurons, corresponds to the data recorded by the prototype measuring carbon dioxide. For the development a time interval of 3 months is considered. In this case the number of input neurons is 4 (year, month, day and hour).

- Number of hidden layers, help the network to generalize and find processing patterns very accurately; In this case the design of the neural network uses a hidden layer.

- Number of hidden neurons, there is no specific rule to determine the number of neurons in the hidden layer, in some cases the network is trained countless times, with different figures and the one that has had the least errors on the data is selected validation; However, Matlab software has a predetermined number of neurons (10), with the possibility of varying the digit to adjust and minimize the error.

- Number of exit neurons, in the case of the present research project, it was determined in one (1), because the neural network will predict the carbon dioxide contamination index in a given time.

- Transfer function, seeks to prevent the neural network from producing outputs with high values, which harms the network in its training and validation; Therefore, the sigmoidal function is widely used for prediction processes.

3. Topology of the neura l network. The neural network consists of three layers; input ("Input") with 4 neurons, hidden ("Hidden Layer") with 10 neurons and output ("Output") with 1 neuron. Fig. 2 shows the structure of an artificial neural network.

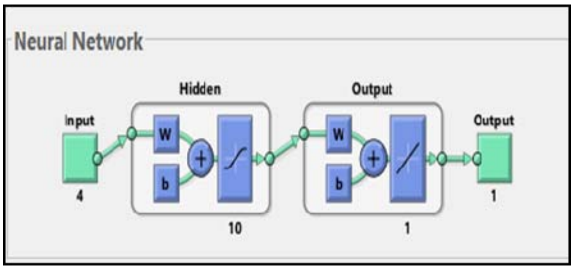

Fig. 2. ANN topology

4. Evaluation criteria. In order to measure the efficiency of the ANN it has been considered to use the "Medium Square Error" (MSE), Eq. (2). The MSE calculates the difference between the network output and the desired response; if the error value is small, the prediction will be more accurate [10] [11].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_p - Y_o)^2 \qquad \text{Eq. (2)}$$

Where "N" is the number of observations considered, "Y_p" is the actual data and "Y_o" is the data estimated by the model.

5. ANN training. Consists in discovering the appropriate configuration of weights, so that the ANN can learn from a set of patterns. The objective function is to minimize the sum of the quadratic error (MSE) of the prediction of the training data by pattern. In the Matlab toolbox, the Levenberg-Marquardt method is selected that propagates the error from the output neurons to the input neurons, minimizing prediction errors. The sets represent 70% for training, 15% validation and test the remaining 15% of the data, obtaining training with excellent results at the 59th time of 65 iterations performed, as shown in Figs. 3 and 4 [12].

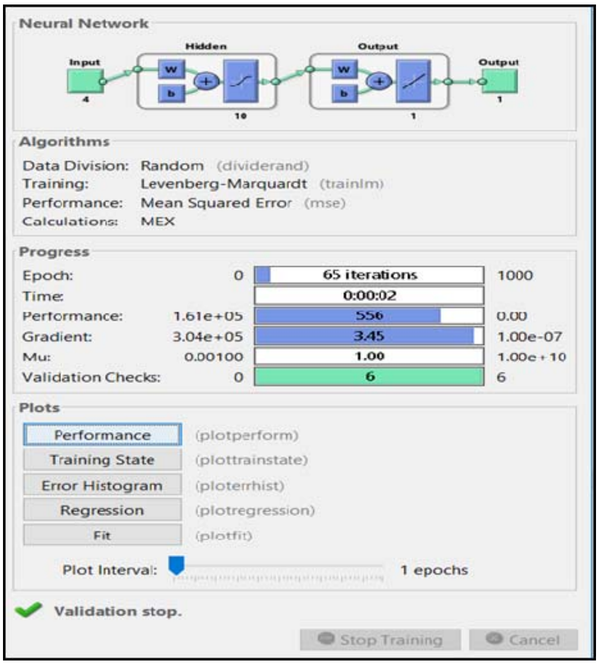

Fig. 3. Training of the artificial neural network for prediction of the CO_2 index

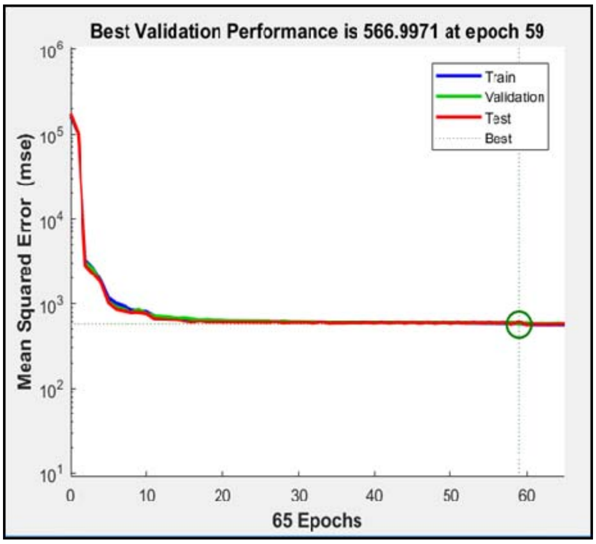

Fig. 4. Training performance

In Fig. 5 it can be seen that the general average correlation coefficient is “R = 0.90719” which tends to approach 1, indicating that there is a high correlation between the real values (“Target”) and the predicted ones ("Output"). The solid gray line indicates the points where the input values are equal to the output values of the artificial neural network.

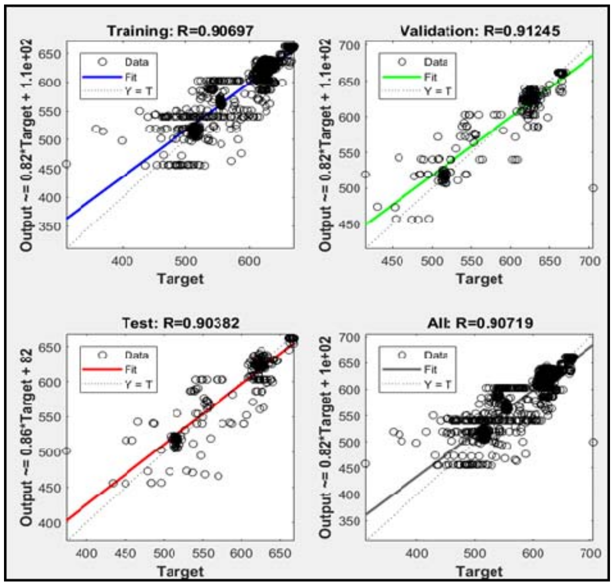

Fig. 5. Correlation between real values and CO_2 forecast

6. System implementation based on the artificial neural network.

For the implementation of the carbon dioxide prediction system, it was considered to create the algorithm focused on forecasting in order to support the decision making of mitigation of the environmental pollutant. From the definition of the variables, input (year, month, day, hour) and output (CO2 measurements) described in previous stages, the prediction periods for the delivery of results by the system are specified.

Taking into account the types of data and the information to be delivered in certain periods, the flow charts were developed with the online tool “Lucidchart” for the general process of the system, which will be encoded in the “M” language of the software Matlab computing. The algorithms developed to predict the impact of CO2 per day and week are shown in Figs. 6 and 7.

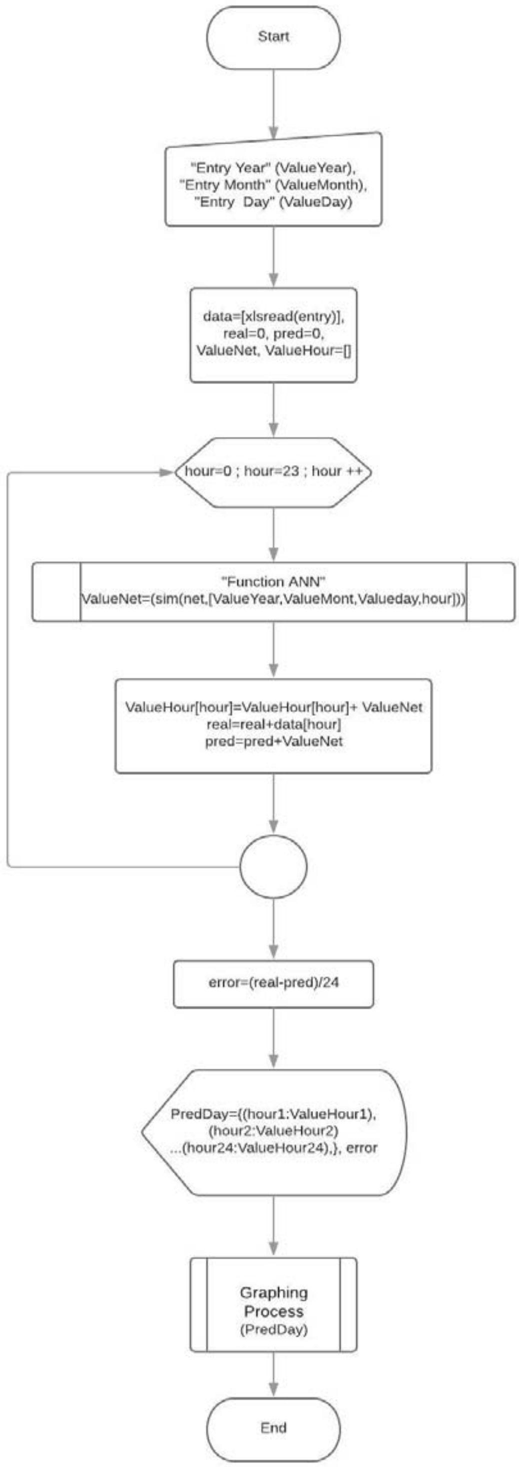

Fig. 6. Prediction flowchart per day

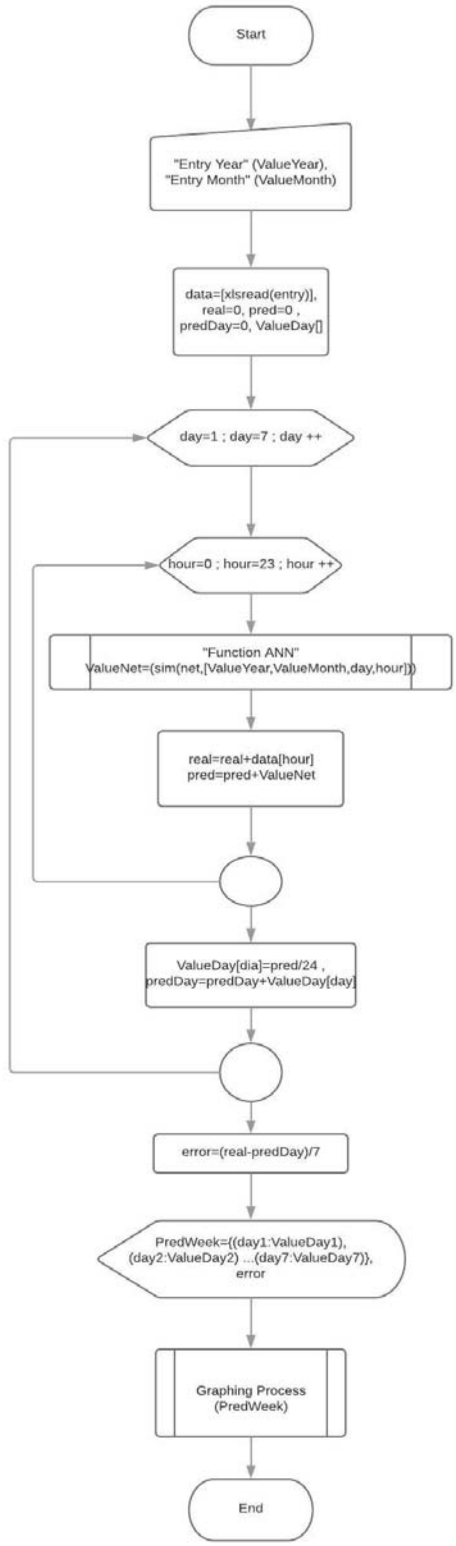

Fig. 7. Prediction flow chart per week

After the development of the algorithms, the coding is carried out. In Figs. 8 and 9, a part of the prediction algorithm code is shown.

```
load('redCO2.mat');                    ...Carga la red neuronal artificial
disp('Predice Día');
data=xlsread('D:\RNA_TESIS\datos_red\datos.xlsx','salida');
Y_real=data((2161:2184),1);            ... Vector de datos reales

X_hora =[] ;                           ... Vector de almacenamiento de hora
Y_pred =[] ;                           ... Vector de almacenamiento de datos de predicción

for hora= 0:23
    ValorRed = sim (net,[2018 12 16 hora]');     ... Calculo del valor de predicción
    X_hora= vertcat (x_hora, hora);              ... Vector para grafico de eje X
    Y_pred= vertcat (y_pred,ValorRed);           ... Vector para grafico de eje y

end
real = round(mean(Y_real),2);                    ... Promedio de los datos reales
prediccion = round(mean(Y_pred),2);              ... Promedio de los valores predicción
error =round(abs(real-prediccion),2);            ... Error del pronostico
```

Fig. 8. Programming code of the CO2 prediction algorithm per day

```
load('redCO2.mat');                ...Carga la red neuronal artificial
disp('Predice Semana');
data=xlsread('D:\RNA_TESIS\datos_red\semana.xlsx','semanal');
Y_real=data(:,1);                  ... Vector de datos reales
Xmes_dia=[] ;                      ... Vector de almacenamiento de dia
Ymes_pred= [] ;                    ... Vector de almacenamiento de datos de predicción del mes

 for dia= 16:23

    Valorhora =[] ;                ... Vector de almacenamiento de hora
    for hora= 0:23
        ValorRed = sim (net,[2018 12 dia hora]');     ... Calculo del valor de predicción
        Valorhora= vertcat (valorhora, ValorRed);     ... Vector de almacenamiento de valo de p
     end
     MediaDia= mean(Valorhora);                       ...Promedio
     Xmes_dia= vertcat (xmes_dia, dia);               ... Vector para grafico de eje X
     Ymes_pred= vertcat (ymes_pred, MediaDia);        ... Vector para grafico de eje y
end
real = round(mean(Y_real),1);                        ... Promedio de los datos reales
prediccion = round(mean(Ymes_pred),1);               ... Promedio de los valores predicción
error =round(abs(real-prediccion),1);                ... Error del pronostico
```

Fig. 9. Programming code of the CO2 prediction algorithm per week

Phase 5: Evaluation of the ANN model. The system will be evaluated by implementing the algorithm in the Matlab software tool, with the purpose of creating the "Script" of compilation, exposing comparative graphs and determining the percentage of prediction error (MSE - "Mean Square Error"), for which is the data acquired in the phases described above. The model will be valued under criteria described in Table 1 [13].

Table 1. Model Rating

Error rate	Assessment
0 – 5	High mind accepted
6 – 10	Acceptable
11 – 15	Accepted with uncertainty
16 – 20	Unpredictable
21 onwards	Not accepted

4 Tests and Results

To validate the ANN model, the information acquired by the CO_2 measurement prototype is considered, which was obtained in the months of october, november and december of the year 2018, selecting a total of 1824 records for the development of the tests.

4.1 Validation of the prediction model with data acquired per day

This test was performed based on the day variable, for which a random record of the database obtained by the carbon dioxide measurement prototype was taken; this information corresponds to 12/16/2018, it was compared with the data predicted by the model, resulting in an average error of 1.27. Therefore, it follows that the prediction is highly accepted, based on the valuation model established in Table 1. Table 2 shows the computational results of the training algorithm.

Table 2. Prediction of the model for the day 12/16/2018.

Hour	Real	Prediction	Error
0:00	517,30	523,73	-6,43
1:00	513,42	523,73	-10,31
2:00	512,15	523,73	-11,58
3:00	515,87	523,73	-7,87
4:00	514,32	523,81	-9,49
5:00	515,82	528,62	-12,80
6:00	622,28	595,80	26,48
7:00	618,92	613,01	5,91
8:00	622,40	615,22	7,18
9:00	627,50	619,44	8,06
10:00	618,83	623,35	-4,52
11:00	641,52	624,97	16,55
12:00	618,77	625,41	-6,64
13:00	638,93	625,51	13,42
14:00	644,87	625,54	19,33
15:00	618,08	625,54	-7,46
16:00	614,23	625,54	-11,31
17:00	626,97	625,54	1,42
18:00	624,97	625,54	-0,58
19:00	630,35	625,54	4,81
20:00	628,95	625,54	3,41
21:00	630,25	625,54	4,71

22:00	630,05	625,54	4,51
23:00	629,15	625,54	3,61
Average Error			**1,27**

In Fig. 10, the actual measurements with a daily average of 599 ppm and the prediction with 597.73 ppm in the environment are shown, it can also be seen that the forecast curve maintains its trend between 9 and 23 hours, due to training patterns on the days of each month, indicating equality of the CO_2 index in the period of time to future.

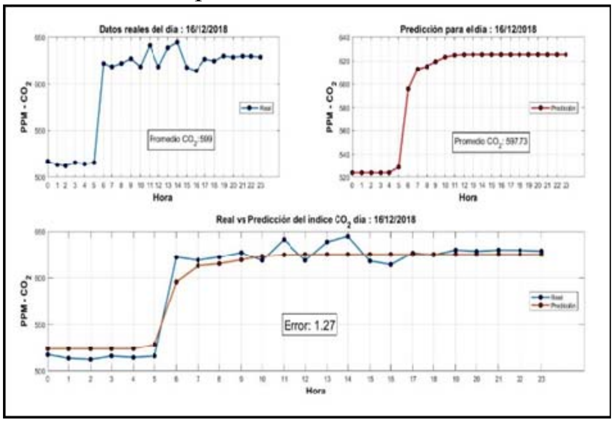

Fig. 10. Prediction of the model for the day 12/16/2018

4.2 Validation of the p rediction model with data acquired per week

This test was carried out based on the week variable, as in the previous case, a random record was taken of the database obtained by the carbon dioxide measurement prototype, corresponding to the third week of december (days 16 to 23) of the year 2018, were compared with the data predicted by the system for the corresponding dates; resulting in an error of 1.6. Therefore, it indicates that it is a highly accepted model based on the valuation model established in Table 1. Table 3 shows the computational results of the training algorithm.

Table 3. Results prediction system for the third week of December 2018 year

Day	**Real**	**Prediction**	**Error**
16	599,71	597,73	1,98
17	598,71	597,49	1,22
18	599,67	597,25	2,42
19	598,50	597,01	1,49
20	598,38	596,77	1,61
21	598,21	596,53	1,68
22	597,29	596,28	1,01
23	597,46	596,04	1,42
Average Error			**1,60**

In summary, it can be affirmed that the prediction model through its forecasting algorithms, demonstrated that it is a highly accepted model because it has shown excellent results in the different validation cases proposed above.

5 Conclusions

The development of the flow chart was coded in the Matlab computer tool, due to the high performance of its components to create artificial neural networks, allowing the prediction system to perform ANN training and validation, in order to obtain effective forecasts for guarantee support for decisions to mitigate CO_2 pollutants.

The comparisons between the real data and those predicted by the model in the different validation cases, demonstrated that it is a highly accepted model.

References:

[1] E. Andrade, «Study of the main types of neural networks and the tools for their application,» Salesian Polytechnic University, Cuenca, 2013.

[2] D. Cabrera, «Design of an artificial neural network for the demand for electrical predictio,» National University of Loja, Loja, (2014).

[3] E. Medina, «Artificial neural networks for predicting the concentration of PM10 in the Ate district. UNIVERSIDAD NACIONAL AGRARIA DE LA SELVA,» Tingo Maria, (2016).

[4] P. Ponce, « Artificial intelligence with engineering applications,» Mexico D.F.: Alfaomega, 2010.

[5] C. Paoli, G. Notton, M.-L. Nivet, M. Padovani y J.-l. Savelli, «A Neural Network Model Forecasting for Prediction of Hourly Ozone

Concentration in Corsica,» *International Conference on Environment and Electrical Engineering, EEEIC.EU 2011,* vol. 10, pp. 1-4, 2011.

[6] M. G. Cortina, « Application of artificial intelligence techniques to the prediction of air pollutants,» UPM Digital File, Madrid, 2012.

[7] C. Gallo, F. Contó y M. Fiore, «A Neural Network Model for Forecasting CO2 Emission,» *Agris on-line Papers in Economics and Informatics.,* vol. 6, nº 2, pp. 31-36, 2014.

[8] M. Huang, T. Zhang, J. Wang y L. Zhu, «A new air quality forecasting model using data mining and artificial neural network,» *IEEE International Conference on Software Engineering and Service Science (ICSESS),* pp. 259-262, 2015.

[9] M. Fernández y N. Lazzo, «Estimation of the CO2 emissions of the students of the UCB (Tupuraya Campus), for the use of transportation and mitigation proposals,» *RevActaNova,* pp. 433-450, 2018.

[10] IEA, « International Energy Agency,» 03 2018. [On line]. Available: https://www.iea.org/publications/freepublications/publication/GECO2017.pdf.

[11] E. Caicedo y J. López, « A practical approach to Artificial Neural Networks,» Cali: University of the Valley, 2017.

[12] M. Rubiolo, « Development of new models and algorithms based on neural networks for data mining tasks,» Santa Fe de la Vera Cruz, 2014.

[13] MathWorks, «The MathWorks, Inc.,» 12 12 2018. [On line]. Available: https://la.mathworks.com.