Artificial Neural Networks-based methodological approach for Climatic variables prediction

Francklin. Rivas-Echeverría, Edmundo. Recalde, Iván. Bedón, Stalin. Arciniegas, and David. Narváez

Abstract— In this work it's presented a neural networks-based methodological approach for obtaining climatic variables prediction using artificial neural networks. In this methodology it's included statistical data analysis techniques and has been used for temperature, humidity and Pressure prediction using data collected from a weather station in Ibarra, Ecuador.

Keywords—Artificial neural networks, Climate prediction, Environmental pressure prediction, Humidity prediction, Temperature prediction.

I. Introduction

WEATER forecast has been a widely studied area and many international services are focus on this activity [1, 2, 3], because of its impact to critical areas as safety, security, agronomical activities, tourism, building, social development, among others.

For making this climate predictions, there have been used mathematical models, statistical model and more recently intelligent models. Ruano el at [4] have used a multi-objective genetic algorithms for designing off-line radial basis function neural network models. Dombaycıa and Gölcü [5] have developed an artificial neural network model for predicting daily mean ambient temperatures in Denizli, south-western Turkey. For adjusting the model they have used temperature values, measured by The Turkish State Meteorological Service. Concerning other important climatic variables prediction, Shank et al [6] have used artificial neural networks for predicting Dew point temperature from 1 to 12 h ahead using prior weather data as inputs.

One work concerning the relationship between temperature and other areas as tool wear using artificial neural networks, was presented by Choudhurya and Bartarya [7]. Also artificial neural networks have been used for creating model for the greenhouse climate [8] and for surface temperatures prediction in Saudi Arabia [9].

Artificial Neural Networks [10] are a very useful intelligent technique [11, 12], because of it capability for learning behaviors from available data and having generalization capabilities for dealing with data different that the one that was used for generating the models [13].

Some methodological approaches have been made for constructing virtual sensors [14], general applications [12, 7, 15, 14] and classification activities using neural networks [16], in this work it's presented a methodological approach for climatic variables prediction using both statistical data analysis [17, 18, 19, 20, 21, 22] and Artificial Neural Networks including the most typical type of problems that can be found working with climatic measurements.

This work is organized as follows: In section 2 it will be presented the proposed methodological approach for climatic variables forecasting using data analysis and artificial Neural Networks. Section 3 will depict the design of the climatic variables prediction model using data analysis and Artificial Neural Networks for predicting the temperature, humidity and environmental pressure in Ibarra city in Ecuador, including the characteristics and topologies used. In section 4 it will be presented the found results for the diverse forecasting models created and finally in section 5 will be presented the conclusions, recommendations and further works.

II. Methodological approach for climatic variables prediction using data analysis and artificial Neural Networks

In order to be able to create climatic variables prediction models using statistical data analysis and neural networks, it is important to establish a methodological framework, where the diverse aspects, circumstances and important issues to be considered in these tasks are contemplated.

The proposed methodology will be presented in stages and phases in order to make it easier to follow and it's derived from other neural networks and software engineering methodologies [15, 14, 16, 23], but including particularities of the climatic variables prediction.

Francklin. Rivas-Echeverría is with the Laboratorio de Sistemas Inteligentes, Universidad de Los Andes, Mérida, VENEZUELA. Escuela de Ingeniería Pontificia Universidad Católica del Ecuador-sede Ibarra. Ibarra, ECUADOR and Programa Prometeo, Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, Quito, ECUADOR (corresponding autor, phone: +584147427277; e-mail: rivas@ula.ve).

Edmundo. Recalde is with the Escuela de Ciencias Agrícolas y Ambientales Pontificia Universidad Católica del Ecuador-sede Ibarra, Ibarra, ECUADOR (e-mail: erecalde@pucesi.edu.ec).

Iván. Bedón, Stalin. Arciniegas, and David. Narváez are with the Escuela de Ingeniería Pontificia Universidad Católica del Ecuador-sede Ibarra. Ibarra, ECUADOR (e-mail: nibedon@pucesi.edu.ec, smarciniegas@pucesi.edu.ec, ldnarvaez@pucesi.edu.ec).

A. Stage 1. Analysis and Description of the climatic prediction Problem

In this stage it should be well studied the climatic conditions in the place where is going to be created the model, including the variables characteristics and behavior, the measurement devices used (meteorological stations, sensors, components, among others). Also is important to know the complete list of available information and the historical databases that will be used for creating the model.

B. Stage 2. Feasibility analysis for climatic variables prediction using Statistical Data Analysis and Neural Networks

According to the information collected in the previous stage it will be studied the feasibility of constructing a neural networks-based model for predicting the climatic variables. It will depend on the quality of the available data, the collected historical data and the diverse measured variables and its relationship with the variables wanted to be predicted.

C. Stage 3. Statistical data analysis of the Variables that will be used for creating the climatic prediction model

Because the model will be created using artificial neural networks, it will highly depend on the quality of the data and the relationship between the variables that will be used for creating the model. It will be used statistical data analysis tools in order to take appropriate decisions concerning the variables and patterns that will be used and to evaluate and complete the data sets for making them useful for making the neural networks-based model. It will be used statistical techniques oriented to the detection of atypical observations, variables transformations if it is required, relationship between variables, data imputation, and data sets selection, among others.

This stage has four important phases:

Phase 3.1 Available Data Description

This phase consists of the description for the data that is going to be used for creating the prediction model. It is necessary to specify the dimensions, ranges, units and environmental meaning of each variable. In this phase the missing values are detected and quantified in order to define if there is any pattern concerning this data loose, because for using imputation techniques it will be necessary to establish that missing data are random and do not present a particular pattern or behavior.

Phase 3.2 Exploratory Data Analysis

Exploratory Data Analysis [24, 18, 21, 22] is used for characterizing the set of climatic variables observations and is formed by four great groups: central tendency measures, dispersion measures, form measures and non-central position measures. It's used for determining the behavior of the data, as well as the possible structures that it present.

Using this technique is important to verify the linear relationship and correlation between variables and detecting possible multicolinearity problems and this will be an important activity for making the variables and patterns selection.

Another important issue for using exploratory data analysis is for Outliers (or atypical observations) detection, that are data that do not correspond or are not coherent with most of the available measurements. There are many techniques for founding this outliers, but the final decision of taking this data off or keep it is based on the researcher criteria and also in the registration of possible problems during the data acquisition.

Phase 3.3 Data processing and imputation

In this phase it should be selected the variables that are going to be used as neural network's inputs (independent or explanatory variables).

The treatment of atypical observations [25] is made and it's decided if this outliers are going to be eliminated, exchanged by other measurements or kept.

Another important activity in this phase is the Data imputation [26, 27, 28, 29] in case that there is some missing values. In this case it should be decided if there is going to be used simple imputation techniques or multiple imputation techniques [30].

Phase 3.4. Training and testing sets Selection

For creating the climatic prediction model using neural models, it is important to split the original data set in two new ones: one of them will be used for adjusting the neural network´s model and the other one will be used for testing the model with non-previously known data in order to verify the generalization capabilities of the model. For this data sets selection some random sampling procedures can be used.

D. Stage 4. Neural Networks-based prediction model construction

In this stage is given the neural networks training, using the arranged data set, the selected inputs and the defined training set. It´s also selected the diverse topologies or neural network´s architecture. It should be decided if there is going to be used some time delay´s models in order to give temporal evolution information to the neural networks. It will be evaluated the results found with diverse neural network´s topologies and which of them presents a better fitting to the original data, and finally the same evaluation process should be done with the testing data set for determining the adjustment of the model to non-previously shown data set and the generalization capabilities. All these activities can be reduced in two phases:

Phase 4.1 Prediction Model Adjustment

It's selected the neural network's architecture or topology (number of layers, neurons in each layer, activation functions, among others). The neural network's inputs were selected in the previous stage and here can also be defined new inputs corresponding to measurements of previous periods of time (time delay). It can also be evaluated combination of neural networks with some other formal or intelligent techniques.

Phase 4.2 Prediction Model Evaluation

It's evaluated the diverse obtained neural networks-based prediction models using the data set reserved for testing. It will be evaluated the generalization capabilities of the obtained prediction models and it will be selected the best model according to the data fitness, simpler topology, less error in training and testing phases, among others criteria with the main purpose of choosing the model that better represents the variable wanted to be predicted.

E. Stage 5. Neural networks-based prediction model implantation

In this stage it will be created the operational model that will be used for continuously predicting the climatic variable. It will depend on the characteristics of the systems where the model will be running and the periodicity for having a new on-line prediction.

In Fig. 1, it's depicted the general methodology scheme for climatic variables prediction using statistical data analysis and neural networks.

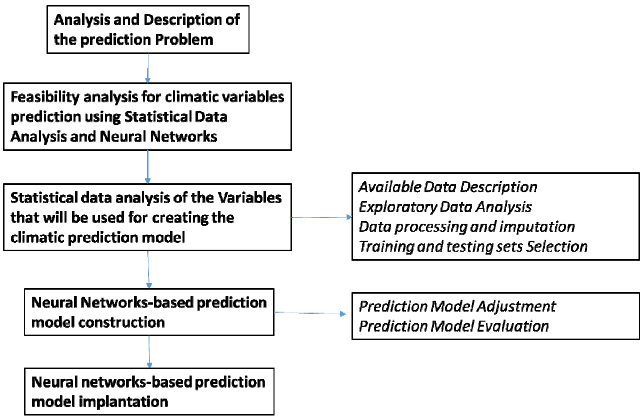


Fig. 1. General Methodology Scheme for climatic variables prediction using statistical data analysis and neural networks

III. Climatic Variables Prediction Model Design Using the Data Analysis and Artificial Neural Networks Proposed Methodology

It was used the proposed methodological approach for predicting the temperature, humidity and pressure in Ibarra city [31], that is located in Provincia de Imbabura in Ecuador. Ibarra is a city with a many climatic changes during the day and is an interesting place for constructing a climatic variables prediction models.

According to the methodology is was studied the climatic station [32] and the variables that have been measured during more than seven year. It was found that there have been using two different meteorological stations with different sample periods, so it was decided to keep the information with the most recently information, that is a station which takes more than sixteen variables every five minutes.

It was studied the relationship between the variables and it was decided to create the climatic variables prediction model using: Temperature, Humidity, Dew Point, Wind speed, Pressure, Rain and Solar Radiation [33].

Because it´s wanted to predict the temperature, humidity and pressure for one day in the future, it was created diverse model using information regarding the time stamp for the climatic variables prediction and it was used information concerning the measurements just from one day before and also with information from previous days up to five former days. The general structure of the neural networks-based models can be seen in Fig.2, Fig.3 and Fig. 4.

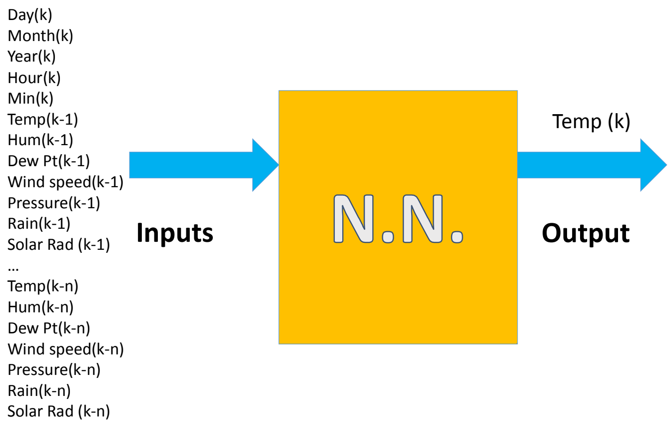


Fig. 2. General structure of the neural networks-based model for temperature prediction in Ibarra, Ecuador

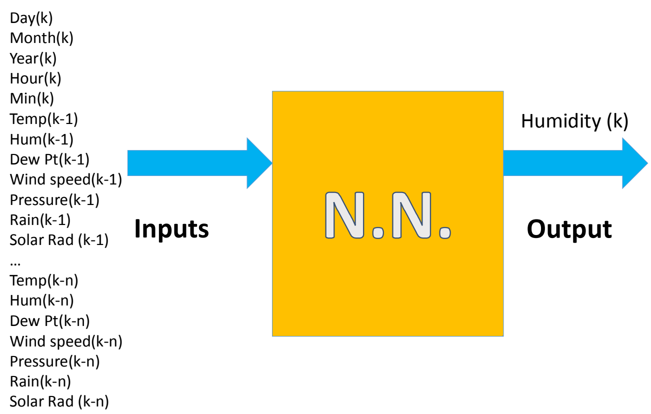


Fig. 3. General structure of the neural networks-based model for humidity prediction in Ibarra, Ecuador

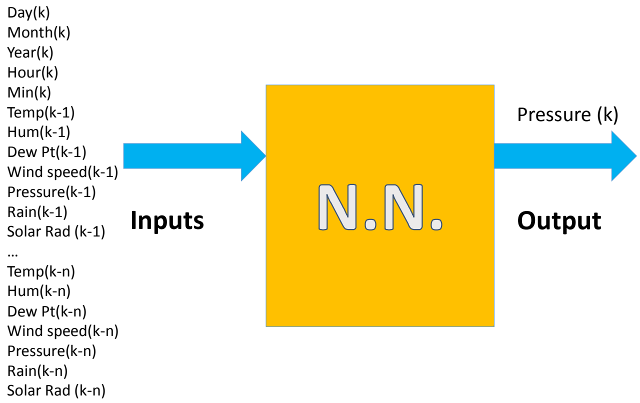


Fig. 4. General structure of the neural networks-based model for pressure prediction in Ibarra, Ecuador

Concerning the statistical data analysis, it was analyzed the data taken every five minutes from January 1st 2012 until may 15th 2015. It was detected the outliers and took decisions about this data. Also it was found the incomplete data and it was made the data imputations in order to fill this spaces in the data sets.

It was selected the data set that was going to be used for training the artificial neural networks (240.000 patterns) and the data set used for testing the model (116.640 patterns).

The model that gave better result for the temperature prediction (according to the general model depicted in figure 2) was the one created using information regarding all the previous days until the fourth day. The results can be seen in Fig. 5, Fig. 6 and Fig. 7.

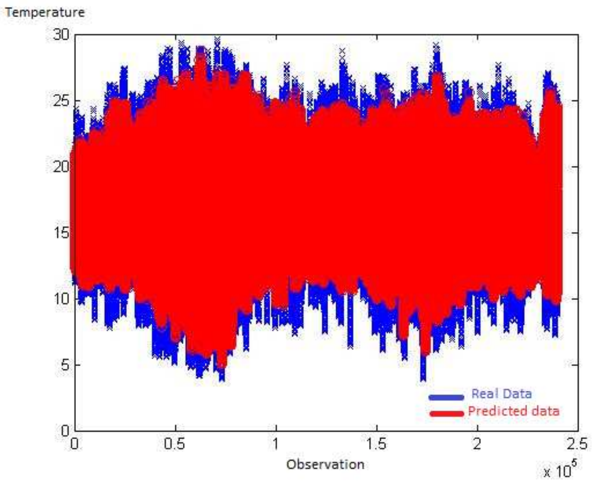


Fig. 5. Result for the training data set using temperature prediction model

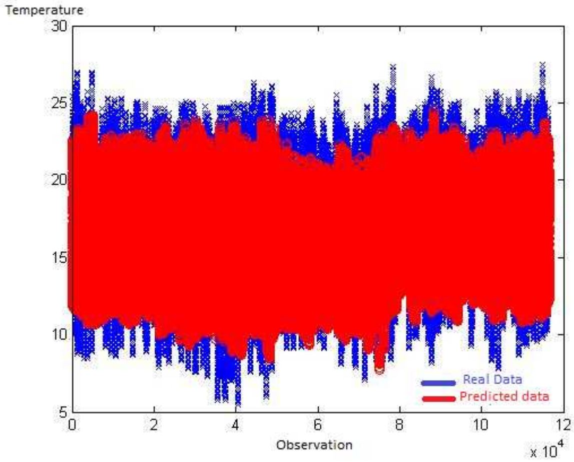


Fig. 6. Result for the testing data set using temperature prediction model

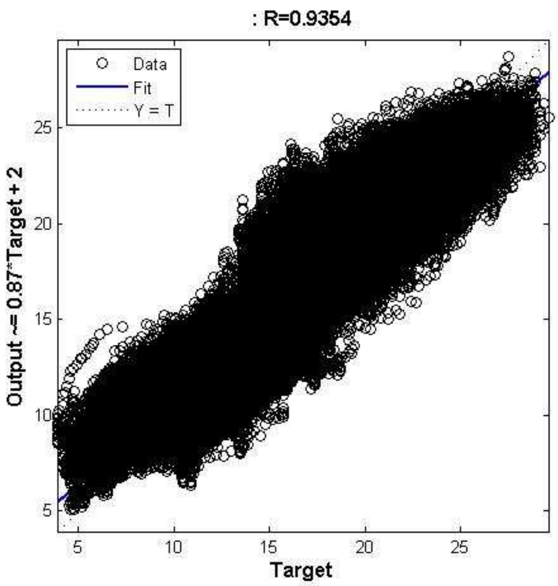


Fig. 7. Regression plot for temperature prediction model

Concerning the humidity prediction, the model that gave better result (according to the general model depicted in figure 3) also was the one created using information regarding all the previous days until the fourth day. The results can be seen in Fig. 8, Fig. 9 and Fig. 10.

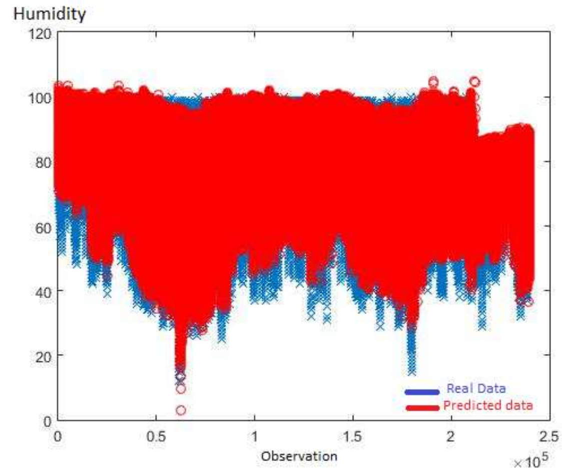


Fig. 8. Result for the training data set using humidity prediction model

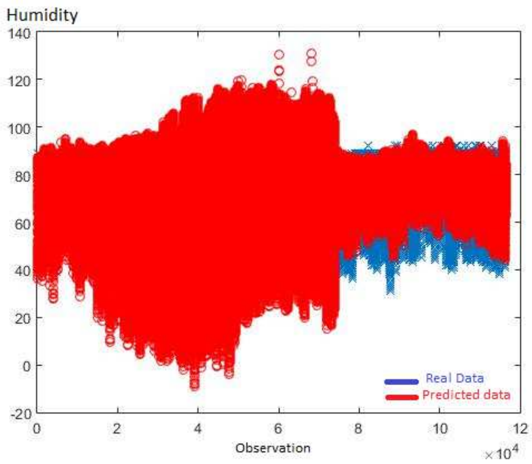


Figure 9. Result for the testing data set using humidity prediction model

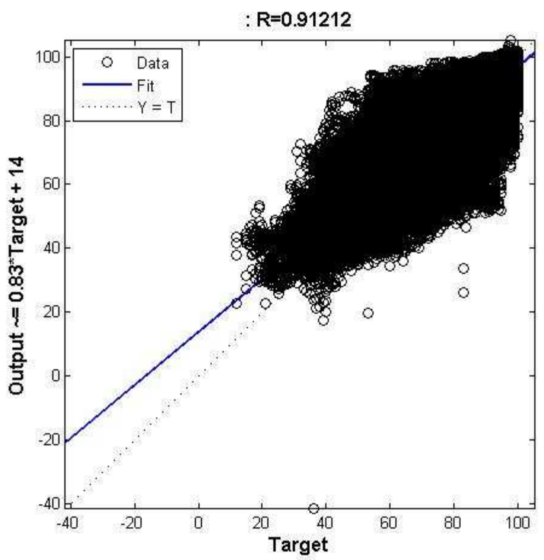


Fig. 10. Regression plot for humidity prediction model

The model that gave better result for the environmental pressure prediction (according to the general model depicted in figure 4) was the one created using information regarding all the previous days until the fifth day. The results can be seen in Fig. 11, Fig. 12 and Fig. 13.

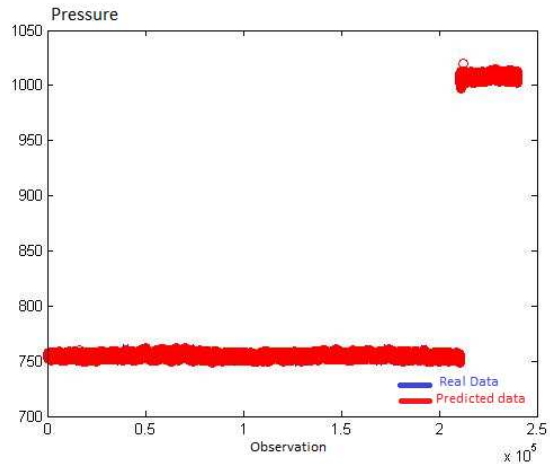


Fig. 11. Result for the training data set using pressure prediction model

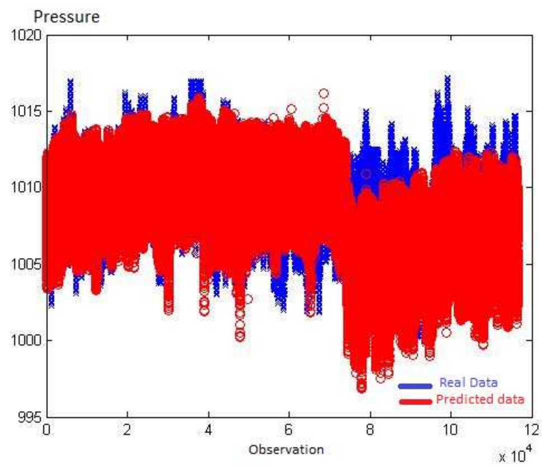
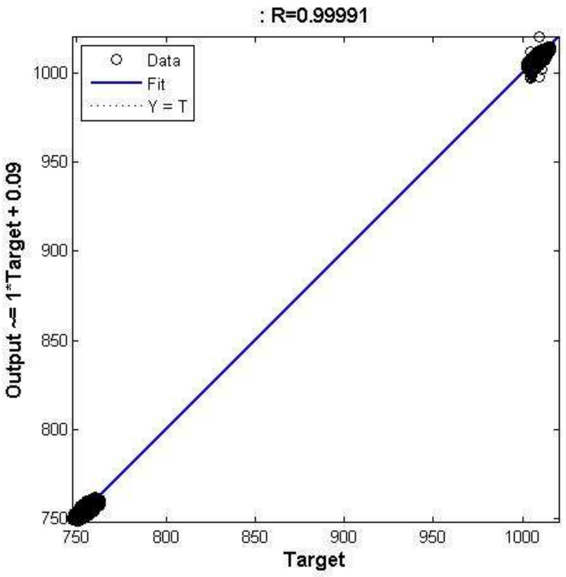


Fig. 12. Result for the testing data set using pressure prediction model



Fig, 13. Regression plot for pressure prediction model

IV. Conclusion

In this paper it was presented a methodological framework proposed for creating climatic variables prediction models using statistical data analysis and artificial neural networks.

This proposed methodological approach considers the typical kind of problems found in climatic prediction problems and the best statistical and analytical ways for handling this problems.

It was used this methodological approach for creating particular models for temperature, humidity and pressure predictions for the city of Ibarra in Ecuador and it was found good results with diverse artificial neural networks structures, founding that for each variable the best model contains different information requirements concerning the measurements from previous days

It will be continued this research, creating new models and it can be studied the feasibility of using other intelligent or hybrid techniques for having better results.

Acknowledgment

Authors want to thanks the support given to this project by the Secretaría de Educación Superior, Ciencia, Tecnología e Innovación of Ecuador and Prometeo Program.

References

[1] National Weather Service. http://www.cpc.ncep.noaa.gov/

[2] Hong Kong Observatory. http://www.hko.gov.hk/wxinfo/currwx/fnd.htm

[3] Australian Weather. http://www.weatherzone.com.au/

[4] Ruano, A.E., Crispim, E.M., Conceição, E.Z.E., Lúcio, M.M.J.R. Prediction of building's temperature using neural networks models. *Energy and Buildings*. Volume 38, Issue 6, June 2006, pp. 682–694.

[5] Dombaycıa, Ö. A., Gölcü, M. Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. *Renewable Energy*. Volume 34, Issue 4, April 2009, pp. 1158–1161.

[6] Shank, D. B., Hoogenboom and McClendon, R. W. Dewpoint Temperature Prediction Using Artificial Neural Networks. *Journal of Applied Meteorology and Climatology*. Volume 47, Issue 6, June 2008, pp. 1757–1769.

[7] Choudhurya, S.K., Bartarya, G. Role of temperature and surface finish in predicting tool wear using neural network and design of experiments. *International Journal of Machine Tools and Manufacture*. Volume 43, Issue 7, May 2003, pp. 747–753.

[8] Seginer, I., Boulard, T., Bailey, B.J. Neural Network Models of the Greenhouse Climate. *Journal of Agricultural Engineering Research*. Volume 59, Issue 3, November 1994, pp. 203–216.

[9] Tasadduqa, I., Shafiqur Rehmanb, S., Bubshaita, K. Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. *Renewable Energy* Volume 25, Issue 4, April 2002, pp. 545–554

[10] Hagan, M., Demuth, H., and Beale, M. *Neural Network Design*. Hagan Publishing. 2002.

[11] Aguilar, J., Rivas F. *Introducción a las técnicas de computación inteligente*. Mérida, Venezuela: Editorial Meritec. 2001.

[12] Bravo, C.E., Saputelli, L., Rivas, F. Pérez, A.G., Nickolaou, M., Zangl, G., De Guzmán, N., Mohaghegh S.D., Nunez, G. State of the Art of Artificial Intelligence and Predictive Analytics in the E&P Industry: A Technology Survey. *SPE Journal*. Vol 19. Nro 4. 2014, pp. 547-563.

[13] Honik, K., Stinchcombe M. and White, H. Multilayer feedforward networks are universal aproximators. *Neural Networks*, vol 2, Issue 5, 1989, pp. 359-366.

[14] Pérez, A. Nava, L. Rivas, F. Colina, E. Neural Networks-based Virtual Sensors Methodology. *Proceedings of WSEAS International Conference on Neural Networks*. Interlaken, Switzerland. 2002.

[15] López, T, Pérez, A., Rivas, F. Data Analysis Techniques for Neural Networks- based virtual sensors. *Proceedings of 8th WSEAS Int. Conf. On Neural Networks*. Vancouver, Canada. 2007.

[16] Pérez-Méndez, A. Torres, E. Rivas, F. Maldonado–Rodríguez, R. A Methodological Approach for Pattern Recognition System using discriminate analysis and artificial neural networks. *Proceedings of 6th WSEAS International Conference on Neural Networks*. Lisbon, Portugal. 2005.

[17] Hawkins, D. (1974). The Detection of Errors in Multivariate Data Using Principal Components. *JASA*, 69, 346. 1974.

[18] Hintze, J. On the Use of "Elemental Analysis" in Multivariate Variable Selection. *Technometrics*, 22, 4. 1980.

[19] Johnson, R. Wichern, D. *Applied Multivariate Statistical Analysis*. Prentice Hall. United States. 2001.

[20] Lebart, L. Morineau, A. Warwick, K. *Multivariate Descriptive Statistical Analysis*. John Wiley & Sons Inc., New York. 1984.

[21] Morrison, D. *Multivariate Statistical Methods*. Third Edition. McGraw Hill Publishing Company. 1979.

[22] Tukey, J. W. *Exploratory Data Analysis, Vol 1*. Addison-Wesley, California. 1977.

[23] Rivas F., Colina E., Rivas C. Expert Systems methodology for Management. *Proceeding of IASTED International Conference on Software Engineering*. Las Vegas, USA. 1998.

[24] Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Third Edition. John Wiley & Son. United Sates. 2003.

[25] Mardia, K., Kent, J., y Bibby, J. *Multivariate Analysis*. Academic Press, London. 1979.

[26] Afifi, A. and Elashoff, R. Missing observations in multivariate statistics III: Large sample analysis of simple linear regression. *Journal of the American Statistical Association*. 64, 1969, pp. 337-358

[27] Little, R. and Rubin, D. *Statistical Analysis with Missing Data*. John Wiley & Sons Inc., New York. 1987.

[28] Madow, W., Nisselson, H. y Olkin, I. *Incomplete Data in Sample Surveys, Volume I, Report and case Studies*. Academic Press, New York. 1983.

[29] Madow, W., Olkin, I. y Rubin, D. *Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies*. Academic Press, New York. 1983.

[30] Schafer, J., Olsen, M. Multiple imputation for multivariate missing-data problems: a data analyst's persepective. *Multivariate Behavioral Research*. Volume 33, Issue 4, 1998, pp. 545-571.
[31] Ibarra city. https://en.wikipedia.org/wiki/Ibarra,_Ecuador
[32] Weather Station Davis Vantage Pro 2. http://www.davisnet.com/weather/products/weather_product.asp?pnum=06162
[33] Rivas, F., Recalde, E., Bedon, I., Arciniegas, S., Narvaez, D. Environmental temperature prediction using a Data analysis and Neural Networks methodological approach. *Proceedings of 16th International Conference on Neural Networks (NN '15)*. Rome, Italy. 2015.

Francklin Rivas-Echeverría. Systems Engineer (1993), Magister Scientiae in Control Engineering (1996), Doctor in Applied Sciences (2000). Full time professor at Universidad de Los Andes (ULA), Venezuela, Invited Research at Pontificia Universidad Católica del Ecuador-sede Ibarra with Prometeo Program, Ecuador. President of the Board of Directors of diverse national and international consortiums. Founding member and Coordinator for the ULA's Intelligent Systems Laboratory. More than 230 scientific articles in journals, books, and international conference proceedings. Author of four books. Awarded by Halliburton for contributions and dedication to the development of petroleum technology. Recognition awarded by Revista Gerente as one of the 100 most successful Managers in Venezuela.

Edmundo Recalde. Full-time professor at the Pontificia Universidad Católica del Ecuador Sede Ibarra, Agroindustrial Engineer with a Higher Diploma in University Teaching and a MBA in Quality and Productivity. He is a PhD candidate in Agricultural and Forestry Research at Universidad Santiago de Compostela. Professor fort more than 15 years of Statistics and Design of Experiments in undergraduate and master level. His research area is alternative energy, currently is working with unicellular algae. He has participated in national and international research projects.

Iván Bedón. Director of the Engineering School at the Pontificia Universidad Católica del Ecuador Sede Ibarra, he has obtained his engineering and Master degree from the Universidad San Francisco de Quito, in Geographic Information systems presenting the thesis "Geo-portal as decision support tool for indigenous community development projects. Corporation Imbabura". Internal Reviewer of "Axioma" indexed journal, Professor of the master's programs in education sciences and Technologies for the Management and Teaching Practices.

Stalin Arciniegas. Mechatronic Engineer (2011), Magister in Technologies for the Management and Teaching Practices (2015). Full time professor at Universidad Católica del Ecuador-sede Ibarra (PUCE-SI), Ecuador. Author of the patent Eco-colector Automático de Apitoxina, IEPI-2015-36251.

David Narvaez. Electronic and field netwoks engineer. (2012), Magister in Technologies for the Management and Teaching Practices (2015). Full time professor at Universidad Catóica del Ecuador-sede Ibarra (PUCE-SI), Ecuador. Research coordinator on Systems Engineering School, Author of the patent Eco-colector Automático de Apitoxina, IEPI-2015-36251. Electronic and mathematics project developer.