International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 21 (2018) pp. 15164-15168
© Research India Publications. http://www.ripublication.com

Data Mining as a Tool for the Environmental Quality Analysis in the Basin of the Macheta River (Colombia)

López Sánchez, Wilson Ricardo[1]; Rodríguez Miranda, Juan Pablo[2]; Sánchez Céspedes, Juan Manuel[3]

[1] Electronic Engineer. (Candidate) Magister in Information and Telecommunications Sciences. LASER Research Group. Francisco Jose de Caldas District University

[2] Sanitary and Environmental Engineering. Magister in Environmental Engineering. PhD (Candidate) Associate Professor. Universidad Distrital Francisco Jose de Caldas. Director of the AQUAFORMAT research group. Email: jprodriguezm@udistrital.edu.co . Postal Address: Carrera 5 Este No 15 - 82. Avenida Circunvalar Venado de Oro. Bogotá DC Colombia.

[3] Electronic Engineer. Magister in Administration. GIIRA Research Group. Assistant teacher. Francisco Jose de Caldas District University.

ABSTRACT

This paper presents the assessment of the concurrent environmental quality evaluated in the conditions in the Watershed of the Macheta River, integrating the variables of the water quality (BOD, TSS, N-NO_2 y P_{total}) and precipitation. Using the data mining technique establishes membership functions and describes data and analyzed variables. In the body of surface water, there is a detriment of the environmental quality in the medium and low basin, which is an obvious contamination of the river and therefore the need to do short-term intervention actions in the basin.

Keywords: Watershed, data mining, environmental quality.

INTRODUCTION

The applications of computational systems for decision making and prediction of the behavior of natural phenomena have increased in terms of the techniques that may represent the conditions and abstraction of the phenomenon [1]. The information obtained from different natural phenomena is used in different computational techniques such as machine learning, and databases are widely used to find important information in processes known as data mining [2,3].

Data mining or knowledge discovery in databases is to extract information from the data, to give it meaning and to draw useful conclusions from it, by describing patterns in large data sets provided for To find intelligible models from them [4,5,6,7]. It provides a response according to the linguistic and verbal information of data, considering the assignment of the partial belongings of any object to different subsets of a universal set, instead of belonging to a single set and this membership function is assumed values between zero And one. The purpose of this technique consists in the prediction by means of the classification (associated a discrete value and the objective is to maximize the predictive power of the classification), regression (it has associated a real value and the objective is to learn a real function, with Can minimize the error between the predicted value and the actual value) through a set of input and output attributes, the value of which can be a category or numerical value, ie predict the output value; Another purpose is the description by means of the grouping (to obtain groups of natural form when applying criteria of similarity of data), that appear without labeling or enumerate, which only possess attributes of entrance and the objective is to describe data [7,8,9,10].

Among the different fields, the search for missing parameters and estimation of parameters is considered [11]. This computational technique can cover several areas of knowledge where one has a way of acquiring data or a determined database to which studies of different types can be made [12] with the aim of obtaining a relation or prediction of a Or various variables of the data with which it is counted. Many models describe the behavior of different physical phenomena that require complicated calculations and are not adaptive models [13,14]. However, with data mining, relevant information can be obtained to estimate missing data and, of course, To approximate the knowledge and behavior of the analyzed natural phenomena.

This technique, is a method of approximation where there are no mathematical equations, however the uncertainties and complications of the model are included in the procedure of descriptive diffuse inference [15]. The applications of techniques are usually in the modeling of surface and groundwater quality, estimation of water quality through satellite images, prediction of earthquakes, prediction of basin levels [16,17], Recognition of water quality patterns and sustainable use of water, identification of ecosystem functioning models, improvement of management and control of wastewater treatment plants, urban planning [18,19,20,21].

This paper presents the analysis of the variables of precipitation and water quality (BOD, TSS, N -NO2 and P_{Total}) in order to understand the patterns of behavior, extract attributes, consider membership functions and describe significant data The same in the watershed of the Rio Machetá (Colombia).

MATERIALS AND METHODS

The method used is a combination of real and exact observation and knowledge of an empirical, complex situation and inductive reasoning, which would consist in deriving a new knowledge from particular phenomena and knowledge already obtained, and establishing propositions analyzed from their Causes and real effects, that is, from the particular to the general [22,23]. It is worth mentioning that according to the analysis and scope of the results, the type of research was analytical - quasi experimental, since it analyzes an event, understands it and in terms of its obvious aspects and discovers the elements that make up the totality and connections Which explain their integration, that is, it facilitates the study and deeper understanding of the event under study [24,25,26].

Precipitation information was obtained from the climatological stations of the Autonomous Regional Corporation of Cundinamarca (CAR) located in each of the municipalities that belongs to the hydrographic basin of the Macheta River; information water quality parameters $_{BOD5,}$ TSS, N and P $_{total\ -NO2,}$ both surface water quality as plants wastewater treatment (including treatment flow) located in towns in To the basin in question, were taken from the Environmental Laboratory of the Regional Autonomous Corporation of Cundinamarca (CAR).

The analysis period for precipitation and water quality information is from year 2012 to 2014. A database was developed with the estimation or replacement of missing data, thus a database of 123 is constructed with 5 different variables and for this the mean and variance for each analyzed variable is determined, then it is ordered upwards with respect to the calculated variance.

RESULTS

The following are the results of applying data mining in the Macheta River Basin.

In Figure 1 for the parameter $_{BOD5,}$ a correlational pattern structure indicator of environmental quality high variability in the data reported in the period of analysis 2012 – 2014 is observed, where especially in 2011 we notice a reduction of environmental quality in the basin.

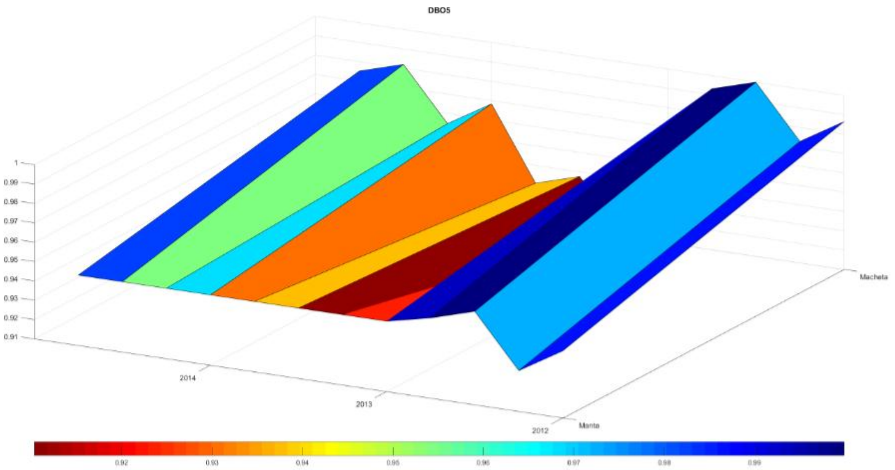

Figure 1. Three- dimensional Diagram for la BOD.

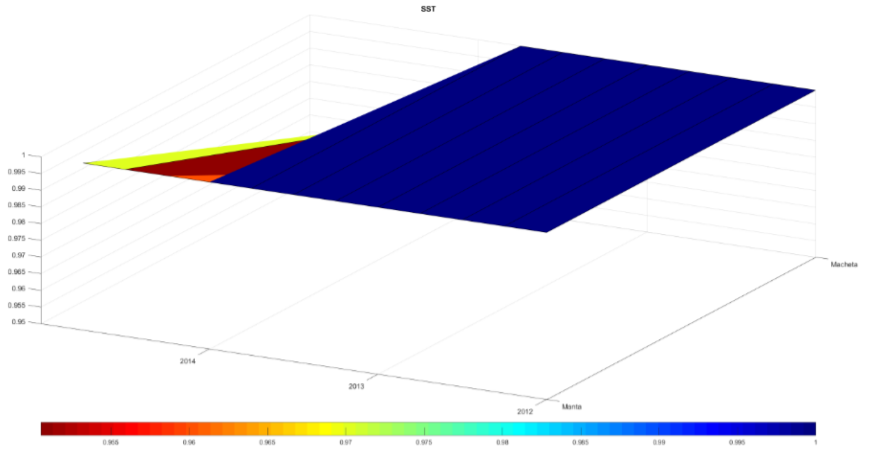

Figure 2. Three-dimensional diagram for TSS.

In Figure 2 for the TSS parameter, an identifiable behavior of patterns of reduced variability, apparent homogeneity and line entities that generate the condition of a good environmental quality related to this parameter for this body of surface water in the period 2012 - 2013, however in the period of 2014, there is a reduction in the environmental quality in the River, due to a dumping of the municipalities of Macheta and Manta.

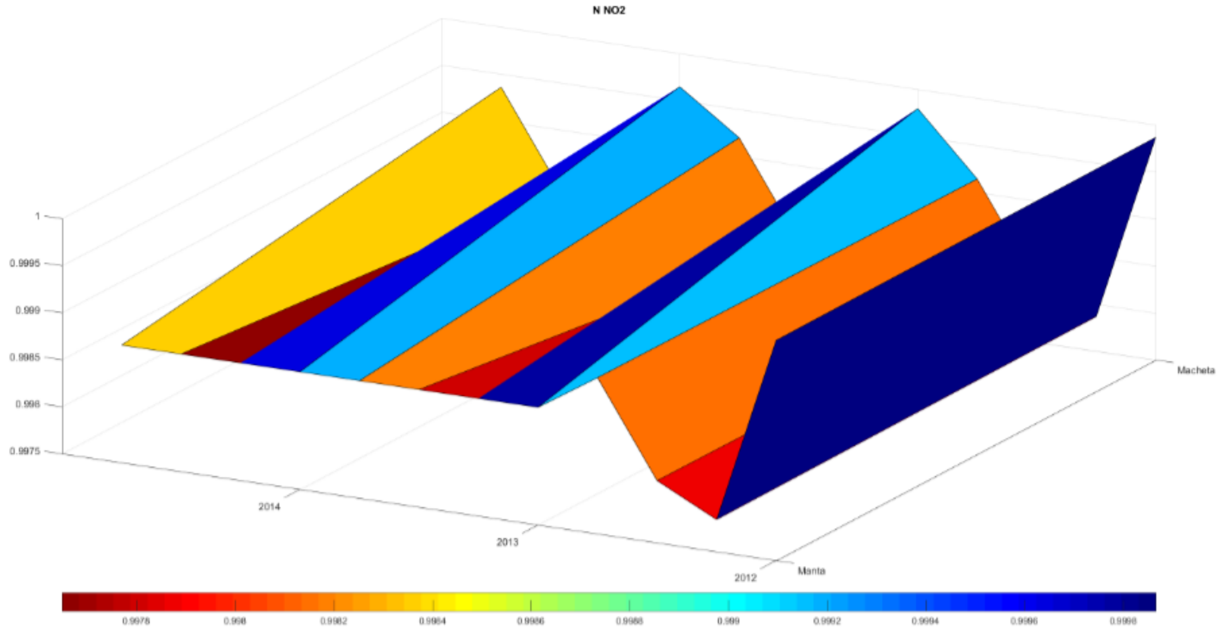

Figure 3. Three - dimensional diagram for N-NO $_2$.

In Figure 3, the behavior of the parameter N- NO $_2$ is observed, with a correlation pattern variability and heterogeneity in environmental quality between good and fair, confluent in a topology environmental quality in the basin in the period 2012 to 2014. Although in the period 2014, a significant contribution of nitrogen by spills in the municipalities of Macheta and Manta is evidenced, generating a detriment in the environmental quality.

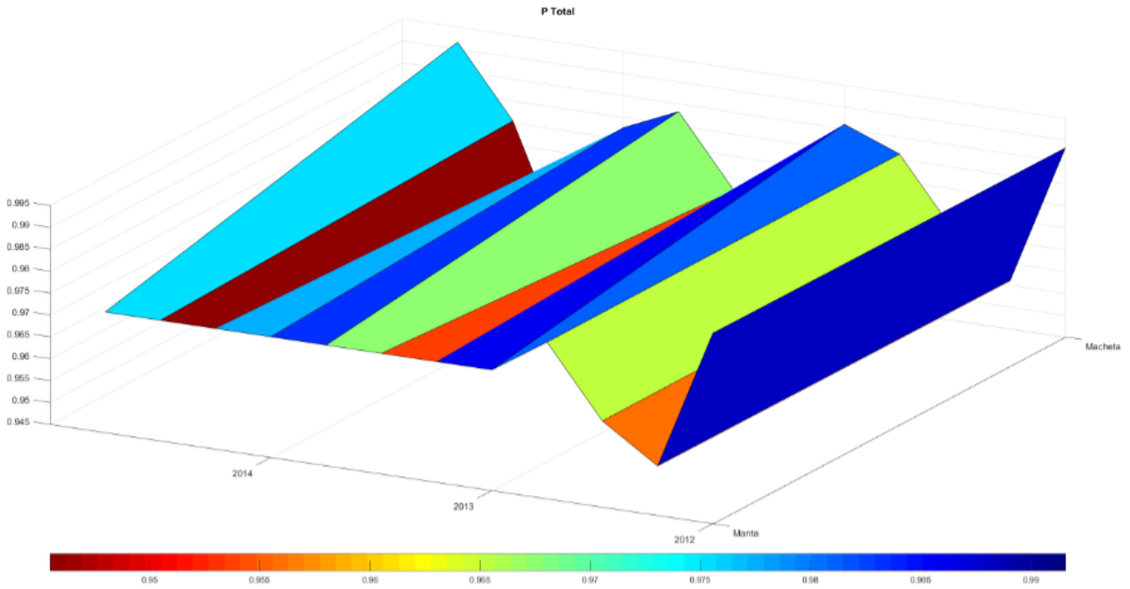

Figure 4. Three-dimensional Diagram for P$_{total}$.

In Figure 4 for the parameter $P_{Total,}$ entities connecting line between stations and observe period-analyzed 2012-2014, with high variability that presents a spatially high dispersion relationship in the data compiled and therefore volatile fluctuation with recurrent frequency that causes a variable structure of the environmental quality in the body of surface water.

CONCLUSIONS

Data mining applied to the analysis of the variables of water quality and precipitation in the basin for the Macheta River, considers understand and comprehend patterns of behavior, extract attributes, consider membership functions and describe significant data of BOD, TST, N -NO_2, P_{total} and precipitation, in order to identify suitable actions in the short term intervention in the basin in terms of spatial identification of scenarios in sections or sectors of the surface water bodies especially in the middle and lower basin, due to the high anthropic pressure, the predominant environmental effect and the evident alteration of the environmental quality in the river. With the membership functions, it is possible to establish a marginal approach to the installation or optimization of wastewater treatment systems (WWTS /WWTP), related to aspects of the sensitivity, adaptability of the technology to be implemented for the increase of environmental quality in the middle and lower basin of the river.

ACKNOWLEDGEMENTS

The authors are grateful for the Doctor of Engineering program of the Francisco José de Caldas District (Bogotá, Colombia), the Autonomous Regional Corporation of Cundinamarca (CAR) and the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) for the information provided by the Basin analyzed.

REFERENCES

[1] Refonaa J, "“Analysis and prediction of natural disaster using spatial data mining technique” ," in International Conference on Circuit, Power and Computing Technologies, USA, 2015, pp. 1-10.

[2] L Pulvirenti, ""La discriminación de las superficies de agua, las lluvias abundantes, y Wet nieve usando las observaciones COSMO-SkyMed de eventos de tiempo severo"," IEEE Transactions on ciencias de la tierra y la teledetección, pp. Vol 52, Numero 2, 152 - 189., 2014.

[3] M. Karim, "A comprehensive study on the effects of using data mining techniques to predict tie strength, ," Computers in Human Behavior, pp. Vol 60, Julio. 534 - 541, 2016.

[4] R. Benítez, Inteligencia artificial avanzada. España: Fundación Universidad Oberta de Cataluña, 2013.

[5] Ferley Medina, "Funcionalidades de la minería de datos.," Revista Ingeniería y región, vol. 12, pp. 31-40, November 2014.

[6] H. Escobar, "Aplicaciones de minería de datos en marketing.," Revista Publicado, vol. 3, no. 8, pp. 503-512, 2016.

[7] Miguel García, Aplicación de técnicas metaheurísticas en minería de datos. España: Universidad de Laguna. Servicio de publicaciones, 2007.

[8] Santos Riquelme, Roberto Ruíz, and Karina Gilbert, "Minería de datos: conceptor y tendencias. ," Revista Iberoamericana de Inteligencia Artificial, vol. 10, no. 29, pp. 11-18, 2006.

[9] R. Ruiz, "Presentación: minería de datos. ," Revista Iberoamericana de Inteligencia Artificial, vol. 10, no. 29, pp. 7-9, 2006.

[10] M. Itati, "Revisión de algoritmos de Redes Neuronales en dos herramientas de Minería de Datos. ," Técnica administrattiva, vol. 11, no. 33, pp. 10-15, 2012.

[11] G Ssali, "Computational intelligence and decision trees for missing data estimation," in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), USA, 2008, pp. 20-25.

[12] J Zhun, "Advances and challenges in building engineering and data mining applications for energy-efficient communities," Sustainable Cities and Society, , pp. Vol 25, August, 33 - 38, 2016.

[13] S Chapra, "Surface Water Quality Modelling.," in The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS, EPA/600/3-87- 007. USA: Mc Graw Hill. Brown, L.C., and Barnwell, T.O. Environmental Protection Agency, 1987, p. 189.

[14] Chapra S, "QUAL 2K: A Modeling Framework for Simulating River and Stream Water Quality.," in QUAL 2K: A Modeling Framework for Simulating River and Stream Water Quality. USA: EPA. Mc Graw Hill, 2008, p. 89.

[15] M. Erkan, "River flow estimation from upstream flow records by artificial intelligence methods.," Journal of Hydrology, vol. 369, no. 1-2, pp. 71-77, May 2009.

[16] M. Bonansea, "Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina)," Remote Sensing of Environment, pp. Vol 158, No 1, March, 28 -41. , 2015.

[17] T. Harvey, "Satellite-based water quality monitoring for improved spatial and temporal retrieval of chlorophyll-a in coastal waters," Remote Sensing of Environment, pp. Vol 158, No 1, March, 417-430, 2015.

[18] M. Ay, "Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques," Journal of Hydrology, no. 511, pp. 279-289, 2014.

[19] H. Sari, "Fuzzy-logic modeling of Fenton’s strong chemical oxidation process treating three types of landfill leachates.," Environmental Science and Pollution Research, vol. 20, no. 6, pp. 4235-4253, June 2013.

[20] T. Pai, "Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality," Applied Mathematical Modelling, vol. 35, no. 8, pp. 3674 -3684, august 2011.

[21] Timoty Ross, Fuzzy Logic, with engineerin applications. , Third edition. ed. New Mexico, USA: WILEY, 2010.

[22] Vergel G., Metodología. Un manual para la elaboración de diseños y proyectos de investigación. Compilación y ampliación temática. Barranquilla: Publicaciones Corporación UNICOSTA, 2010.

[23] M Balestrini, Cómo se elabora el proyecto de investigación. Caracas, Venezuela: BL Consultores asociados, 2001.

[24] Hurtado J., Metodología de la investigación holística.. Caracas: Fundación SYPAL, 2000.

[25] G. Vergel, Metodología: un manual para la elaboración de proyectos de investigación. Barrranquilla.: Unicosta, 2010.

[26] R. Hernández, Metodología de la investigación.. México: Mc Graw Hill, 2010.