Study of Water Quality Classifiers

Jorge Camejo[(a)], Osvaldo Pacheco[(a)], Miguel Guevara[(b)]

(a) Institute of Electronics and Telematics Engineering
University of Aveiro, Portugal
jcamejo@ua.pt / orp@ua.pt

(b) Institute of Mechanical Engineering, University of Porto
Rua Dr. Roberto Frias no 400, Porto
mguevara@fe.up.pt

ABSTRACT

Over large parts of the world, humans have inadequate access to drinking water and so, have to use sources contaminated with disease vectors, pathogens or unacceptable levels of toxins or suspended solids that contribute to the spread of diseases. The control and monitoring of the drinking water quality is an open field and has become of prominent concern during the recent decades. This has resulted in the formulation of national and also international regulations for drinking water quality as this is a very important issue both in the first and in the third world. Each member state adjusts the required policing measures to ensure that the legislation is implemented. However, the scientific community has sought to find ways simpler and less costly to classify the water. The most used method of water quality evaluation is to generate a Water Quality Index (WQI) representing a single number based on a reduced number of parameters. This type of water quality assessment is useful not only for its ability to generate understandable classifications, but also for its potential to facilitate behaviour studies over time and also to feed machine learning classifiers for drinking water. In the present work some WQI methods will be studied and compared to finds new classifications approaches using only parameters measured in real time which will provide starting points for machine learning algorithms.

KEY WORDS: *water resources, water quality index, machine learning, hydroinformatic.*

INTRODUCTION

The water quality in lakes, rivers and coastal areas is periodically assessed by scientific and environmental institutions. This classification is based on laboratory analysis of water quality samples collected from stations at select locations and data on water samples are stored in computer systems. From these recorded parameters of water, researchers have been able to take important decisions to safeguard the environment in general and their own consumption.

There are several ways to study and evaluate water quality. One of the most used is the Water Quality Index (WQI), not only for its ability to generate understandable classifications, but also for its potential to facilitate behaviour studies over time. The WQI makes available a single number that expresses the overall water quality based on a reduced number of parameters. The idea of this index is to turn complex water quality data into information that is reasonable and quantifiable by the general public. Usually associated with this numerical rating, a qualitative categorization is attributed. This qualitative classification makes the water quality evaluation process even simpler and may be used in the learning process by artificial intelligence algorithms. Some of the advantages of WQI are: (i) Easy to disseminate through the non-specialized people; (ii) more importance than the individual parameter values; (iii) an average of various variables into a unique number combining different measurement units into one unity is represented. The main WQI disadvantage is that some information regarding individual variables and their interactions may be lost.

The main scope of the present study is to use a WQI classifier as reference to propose a reduced number of parameters measured in real time capable of drinking water classification.

Many researchers and environmental institutions have presented WQI methodologies or have modified the previous methods, Horton & Chase, 1917; Horton, 1965; Brown, 1970; INAG, 2000; Kumar, 2009; CCME, 2001; Rickwood & Carr, 2009; Seilheimer *et al.*, 2009, Said *et al.*, 2004; WEP, 1996; Boyacioglu, 2010; Chaturvedi & Bassin, 2010; Jinturkar *et al.*, 2010; Karbassi *et al.*, 2011; Li *et al.*, 2010; Nikoo *et al.*, 2011; Rajankar *et al.*, 2011; Ramesh *et al.*, 2010; Vasanthavigar *et al.*, 2010; Yadav *et al.*, 2010; Yidana & Yidana, 2010; Semiromi *et al.*, 2011; Soroush *et al.*, 2011.

For the WQI methods the numbers of the parameters to be used are dependent on the water to be classified, i.e. drinking water, environmental protection or estuaries and coastal waters. In general, 6 to 22 parameters are reported to be necessary for the WQI. However, only a few of these parameters can be measured in real time.

A real time monitoring would provides continuous data daily, seasonal, and like this event-driven fluctuations are not missed. This make possible to recognize immediately changes in water-quality conditions. Therefore, the automatic assessment of the drinking water quality is mandatory. The main problems of automatic assessment of drinking water quality are the difficulties of real time measurement of some physical, chemical, and microbiological parameters and the human and financial resources spent to obtain their values.

METHODS

This study is divided in two parts. First, the most reported WQI methods will be analysed. In the second part machine learning algorithms will be explored to study the correlation between parameters measured in real time. In order to estimate the performance of this task, two datasets, located in different regions, were collected from 2000 to 2011: a) The Provincial Water Quality Monitoring Network from Ontario, Canada and b) National Hydrologic Information System from the Central Region of Portugal (Figure 1).

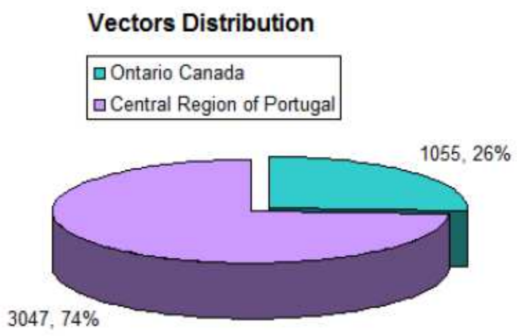

Figure 1. Vectors Distribution by dataset.

Water Qualities Classifiers

In general, water quality indices are developed in three major steps: (i) selection of the variables out of a large number of possible variables; (ii) developing the subindices function and weightages; and (iii) aggregation of the candidate variables selected, (Kumar, 2009). As referred above there are different methods to classify the water quality.

In 1974, Ralph D. Harkins proposed an objective, nonparametric statistical procedure for combining a number of water quality parameters to arrive at a water quality index. These indexes are directly comparable and may be plotted to show time trends or station differences or analyzed statistically to obtain probability levels. The index is suitable for summarizing water quality and determining overall trends, but it does not replace careful analysis of individual parameters to determine potential sources of pollution problems. Parameters included in the index are given equal weights (Harkins, 1974).

The Canadian Council of Ministers of the Environment (CMECC) had been working in an efficient WQI (CCME, 2001; Hurley *et al.*, 2012). The CCME WQI incorporates three elements: Scope - the number of water quality parameters (variables) not meeting water quality objectives (F(1)); Frequency - the number of times the objectives are not met (F(2)); and Amplitude. the extent to which the objectives are not met (F(3)). The index produces a number between 0 (worst) to 100 (best) to reflect the water quality. This method, similar with Ralph method, is suitable to classify the water only for a period of time but is not able to classify the water from independent samples.

The Portuguese water management service (INAG, Agência Portuguesa do Ambiente) proposed a support to evaluate the quality of the superficial waters. In this method, the water was classified in non linear scale A, B, C, D, or E where A denotes no pollution and E denotes extreme pollution, which represents serious risks in terms of public and environmental health (Table 1). (INAG, 2000). The drawback of this method is that only one parameter is sufficient to not be in the certain range of values and the water will be wrong classify in inferior category. No correlation between the parameters is established in this method.

Brown *et al.*, developed a water quality index (Brown, 1970) comparable in structure to Horton's index (Horton, 1965) but with much greater rigour in selecting parameters, developing a common scale, and assigning weights for which elaborate Delphi technique were performed. This effort was supported by the National Sanitation Foundation (NSF) from United Sates. For formulating NSF-WQI, Brown *et al*. assembled a panel of 142 experts from various professions throughout the United States with expertise in various aspects of water quality management. The members of the panel were mailed a series of three questionnaires.

Although this method is based on expert rates which can destroyed comparability and objectivity, it is very useful to classify the water from independent samples. This method permits valuable comparison between waters from different geographical regions and also trends studies for regions of interest.

Many researches and science institutions in various countries have been applied this method to water classification (Chaturvedi & Bassin, 2010) in India and other researcher has been apply this methods to check other methods (Bhatt & Pandit, 2010) in India, Perez-Castillo & Rodriguez, 2008 in Costa Rica.

In 2009 Kumar and Alappat applied the same method of Brown for the formulation of NSF-WQI but with some modifications as follows: reversal of significance values; and proper incorporation of expert opinion.

Recently WQI methods based on parameters measured in real time have been proposed. The main purpose of these methods is the urgency to protect water supplies from incidental contaminations, poisoning and even from bioterrorist acts. Some of them detect drinking water contaminants in Real-Time as warning system using conventional sensor, ENDETEC, 2012; Murray *et al.*, 2012; Yang *et al.,* 2009. These works act directly on the drinking water pipes and are very useful to identify the presence of specific polluters on the water pipes that affect human health. The main disadvantage of these systems is that they are not giving a complete indication on the water quality and is only taking into account those specific parameters that are directly affecting the life human safety without any correlation between the parameters. Is not possible any kind of trend studies neither any comparison between qualities of different regions waters.

In addition to the feasibility of the classification mentioned above, there are other important factors in the application of this framework (Table 2).

Table 1. Values ranges proposed by INAG (INAG, 2000)

Parameters	***Categories/Classes***				
	A	***B***	***C***	***D***	***E***
pH values (Sorense)	6.5 -8.5		6.0 - 9.0		5.5 - 9.5
Temperature / ^oC	≤20	>20 ≤25	>25 ≤28	>28 ≤30	>30
Conductivity / µS cm^{-2}	≤750	>750 ≤1000	>1000 ≤1500	>1500 ≤3000	>3000
Total Suspended Solids /mg dm^{-3}	≤25	>25 ≤30	>30 ≤40	>40 ≤80	>80
Dissolved Oxygen (% sat)	≥90	<90 ≥70	<70 ≥50	<50 ≥70	>50
Oxidability /mg(O_2) dm^{-3}	≤3	>3 ≤5	>5 ≤10	>10 ≤25	>25
5-day BOD (20^{o}C) / mg(O_2) dm^{-3}	≤3	>3 ≤5	>5 ≤8.	>8 ≤20	>20
COD / mg (O_2) dm^{-3}	≤10	>10 ≤20	>20 ≤40	>40 ≤80	>80
Ammonia nitrogen mg NH_4^+ dm^{-3}	≤0.1	>0.1 ≤1	>1 ≤2	>2 ≤5	>5
Nitrate	≤5	>5 ≤25	>25 ≤50	>50 ≤80	>80
Kjeidahl nitrogen mg$_N$dm^{-3}	≤0.5	>0.5 ≤1	>1 ≤2	>2 ≤3	>3
Total reactive phosphurus mg P_2O_5 dm^{-3}	≤0.54		>0.54 ≤0.94		>0.94
Total Coliforms / n^o / 100 cm^3	≤50	>50 ≤5000	>5000 ≤50000	>50000	
Fecal Coliforms/ n^o / 100 cm^3	≤20	>20 ≤2000	>2000 ≤20000	>20000	

In the present study a real time prediction of the overall water quality of natural source as the best prevention of human health and environment will be proposed. This prediction is possible because it is based on strong correlation (using artificial intelligence algorithms) between parameters measured in real time and others that are not.

ORIGINAL RESULTS

In order to propose a drinking water quality classification in real time four parameters that can be measured automatically will be used. The selected parameters are: pH, Dissolved Oxygen, Nitrates and Temperature.

The next step was to apply a water quality index method to associate a class for each instance of the dataset, for using as the classification criterion. After this step, three different machine learning (ML) methods were applied: K-Nearest Neighbour, Partial Decision Tree and Artificial Neural Net. Details of the proposed method are illustrated in Figure 2.

In all ML algorithm applied, Ten folds cross validations was programmed to evaluate all iterations, which means that the average and standard deviations were evaluated (Table 3.).

By each iteration about 80 % of these data were applied to train a model, 10 % for validation and 10 % to test for the three models used.

Similar tendencies are observed comparing the three algorithms by iterations. The use and application of these algorithms in the field of water resource are thus well justified. However, kNN algorithm gave slightly better results. The Bregman Divergences and Mahalanobis distance were used to support nearest neighbor method. In the case of PART algorithm, the confidence was 95 % and the third method was an Artificial Neural Network backpropagation with three hidden layers (3 and 1 neurons respectively).

Table 2. Comparison between classifiers methodologies

Classical Method	***Pipes Method Real-Time***	***Source Method Real-Time***
• Time - consuming	• Real Time warning	• Real Time warning
• Expensive analysis	• Expensive installation	• Reasonable cost
• No remote control	• Remote control	• Remote control
• Good accuracy	• Only contaminants warning	• Reasonable accuracy

Table 3. Accuracy and Sdv values per Algorithm

ML Algorithm	***Accuracy***	***Sdv***
KNN	98.02	1.50
PART	97.31	2.03
ANN	93.79	2.67

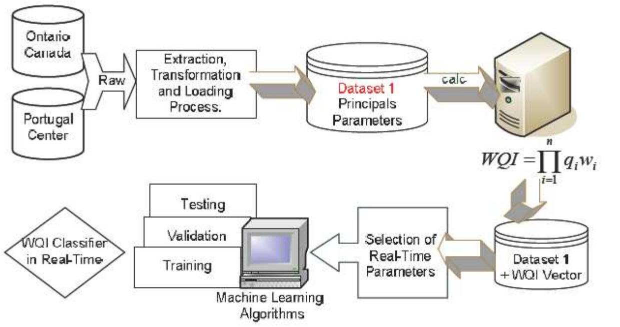

Figure 2. The proposed method.

CONCLUSIONS

In this paper the most used WQI methods have been reviewed. It has been shown that the calculation of a WQI is a challenging task due to the various variables that must be taking into account as: type of water sources and different locations with specific environments which influence the parameters share into the WQI value.

All the methods have their advantages none of them are complete. Some can classify the water quality for a certain period of time being able to make accurate correlation between parameters over time. Other methods can give indication on the WQI from independent samples which permit better comparison between samples points independently on the period of time. This allow to feed machine learning algorithms from each individual cases.

The original results of this study demonstrate the machine learning algorithm can be applied for real time classification of water quality.

ACKNOWLEDGEMENT

The authors acknowledge the Institute of Electronics and Telematics Engineering of the University of Aveiro (IEETA/UA) and Águas DA Região de Aveiro (AdrA) for the collaboration in this work. Jorge Camejo is thankful to the Foundation for Science and Technology (FCT) for the fellowship SFRH/BD/61253/2009, to European Social Fund (FSE) and to the Operational Human Potential Program (POPH). Prof. Guevara acknowledges POPH - QREN-Tipologia 4.2– Promotion of scientific employment funded by the ESF and MCTES, Portugal.

LITERATURE CITED

Bhatt, J. P., & Pandit, M. K., 2010. A macro-invertebrate based new biotic index to monitor river water quality. Current Science, 99(2), 196-203.

Boyacioglu, H., 2010. Utilization of the water quality index method as a classification tool. Environmental Monitoring and Assessment, 167(1-4), 115-124. doi: DOI 10.1007/s10661-009-1035-1

Brown, R. M. et. al., 1970. A water quality index - Do we dare? Water Sewage Works 11, 339-343.

CCME, 2001. Canadian Water Quality Guideline. Canadian Council of Ministers of the Enviroment.

Chaturvedi, M. K., & Bassin, J. K., 2010. Assessing the water quality index of water treatment plant and bore wells, in Delhi, India. Environmental Monitoring and Assessment, 163(1-4), 449-453. doi: DOI 10.1007/s10661-009-0848-2

ENDETEC, 2012. Global Sensor Platform of Veolia Water Solutions & Technologies, 2013, from http://www.endetec.com/en/

Harkins, R. D., 1974. Objective Water-Quality Index. Journal Water Pollution Control Federation, 46(3), 588-591.

Horton, R. K., 1965. An index number system for rating water quality. J. Water Pollut. Control Fed., 300-306.

Horton, T., & Chase, E. S.,1917. A Study of the Application of the Score. System to the Sanitary Quality of Public Water Supplies in New York State. American Journal of Public Health, 7(4), 380-390.

Hurley, T. et al., 2012. Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality. Water Research, 46(11), 3544-3552. doi: DOI 10.1016/j.watres.2012.03.061

INAG, 2000. The Portuguese Water Management Service. Retrieved from http://www.inag.pt/.

Jinturkar, A. M.et al., 2010. Determination of water quality index by fuzzy logic approach: a case of ground water in an Indian town. Water Science and Technology, 61(8), 1987-1994. doi: Doi 10.2166/Wst.2010.095

Karbassi, A. R. et al., 2011. Development of Water Quality Index (WQI) for Gorganrood River. International Journal of Environmental Research, 5(4), 1041-1046.

Kumar, D. & Alappat, B., 2009. NSF-Water Quality Index : Does it represent the experts´ opinion? [Technical Notes]. Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management.

Li, P. Y., et al., 2010. Groundwater Quality Assessment Based on Improved Water Quality Index in Pengyang County, Ningxia, Northwest China. E-Journal of Chemistry, 7, S209-S216.

Murray, S.et al., 2012. Real-Time Water Quality Monitoring: Assessment of Multisensor Data Using Bayesian Belief Networks. Journal of Water Resources Planning and Management-Asce, 138(1), 63-70. doi: Doi 10.1061/(Asce)Wr.1943-5452.0000163

Nikoo, M. R., et al., 2011. A probabilistic water quality index for river water quality assessment: a case study. Environmental Monitoring and Assessment, 181(1-4), 465-478. doi: DOI 10.1007/s10661-010-1842-4

Perez-Castillo, A. G. & Rodriguez, A., 2008. Physicochemical water quality index, a management tool for tropical-flooding lagoons. Revista De Biologia Tropical, 56(4), 1905-1918.

Rajankar, P. N. et al., 2011. Groundwater quality and water quality index at Bhandara District. Environmental Monitoring and Assessment, 179(1-4), 619-625. doi: DOI 10.1007/s10661-010-1767-y

Ramesh, S., et al., 2010. An innovative approach of Drinking Water Quality Index-A case study from Southern Tamil Nadu, India. Ecological Indicators, 10(4), 857-868. doi: DOI 10.1016/j.ecolind.2010.01.007

Rickwood, C. J., & Carr, G. M., 2009. Development and sensitivity analysis of a global drinking water quality index. Environmental Monitoring and Assessment, 156(1-4), 73-90. doi: DOI 10.1007/s10661-008-0464-6

Said, A., et al., 2004. An innovative index for evaluating water quality in streams. Environmental Management, 34(3), 406-414. doi: DOI 10.1007/s00267-004-0210-y

Seilheimer, T. S., et al., 2009. Comparative study of ecological indices for assessing human-induced disturbance in coastal wetlands of the Laurentian Great Lakes. Ecological Indicators, 9(1), 81-91. doi: 10.1016/j.ecolind.2008.02.001

Semiromi, F. B., et al., 2011. Evolution of a new surface water quality index for Karoon catchment in Iran. Water Science and Technology, 64(12), 2483-2491. doi: Doi 10.2166/Wst.2011.780

Soroush, F., et al., 2011. A Fuzzy Industrial Water Quality Index: Case Study of Zayandehrud River System. Iranian Journal of Science and Technology Transaction B-Engineering, 35(C1), 131-136.
Vasanthavigar, M., et al., 2010. Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. Environmental Monitoring and Assessment, 171(1-4), 595-609. doi: DOI 10.1007/s10661-009-1302-1
WEP, 1996. Lower Great Miami Watershed Enhancement Program (WEP).
Yadav, A. K., et al., 2010. Water Quality Index Assessment of Groundwater in Todaraisingh Tehsil of Rajasthan State, India-A Greener Approach. E-Journal of Chemistry, 7, S428-S432.
Yang, Y. J., et al., 2009. Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results. Journal of Environmental Management, 90(8), 2494-2506. doi: DOI 10.1016/j.jenvman.2009.01.021
Yidana, S. M., & Yidana, A., 2010. Assessing water quality using water quality index and multivariate analysis. Environmental Earth Sciences, 59(7), 1461-1473. doi: DOI 10.1007/s12665-009-0132-3