

A novel fuzzy rule based system for assessment of ground water potability: A case study in South India

G. Shobha*, Jayavardhana Gubbi†, Krishna S Raghavan*, Lakshmikanth K Kaushik*, M. Palaniswami†

*Department of Computer Science, R V College of Engineering, India

†ISSNIP, Department of Electrical and Electronic Engineering, University of Melbourne, Australia

Abstract: Groundwater is an important water resource for domestic, irrigation, and industrial needs. The most widely exploited use of this resource is for consumption. Assessment of potability of any ground water samples is a non-trivial task. A new fuzzy rule based system has been proposed to assess the quality of ground-water samples collected from the bore-wells across 24 districts of Karnataka (South India). Eight groundwater quality salts parameters are selected for water quality analysis. A membership function for the fuzzy rule based system for each salt is developed and the weights for each parameter was calculated using Analytic Hierarchy Process (AHP) that relies on pair wise comparison. The system showed that out of 24 districts of Karnataka state, ground water from 51.78% bore-wells was not feasible for consumption.

Keywords: Groundwater quality, Fuzzy rule based system

I. Introduction

Ground water contributes to about eighty percent of the drinking water requirements in the rural areas, fifty percent of the urban water requirements and more than fifty percent of the irrigation requirements of India (Uttam kumar,2009). By 2025, India is projected to be the most populous country in the world with 1/6th of the world's population (Uttam kumar, 2009). The quality and more importantly the basic availability of groundwater is of grave concern, starting from meeting the requirements of basic human needs, to health, agricultural productivity and the associated long term environmental concerns. By 2030, demand for potable water is expected to grow from the current 83 Million liters per day to 189 Million liters per day in India. The ability of water to dissolve all soluble matter with which it comes in contact, leads to ground water containing a higher mineral content than surface water (Ting-Nien, 2010). This is mainly due to very slow percolation and longer periods of contact. Water quality changes with passage of time. In the case of groundwater, the quality varies from stratum to stratum as well as at different depths due to change in chemical composition. Particularly in arid areas of the earth, where evaporation exceeds precipitation, sustained management of groundwater rarely happens. However, there is a need for real time high-density data (both spatially and temporally) ,(Ghisi E 2012). This will aid in management of groundwater at all levels – local, state and national. Recent technological advances in low power integrated circuits and wireless communications have made available efficient, low cost, low power miniature devices for use in sensing applications (Anastassia, 2003). The combination of these factors has improved the viability of utilizing a sensor network consisting of a large number of intelligent sensors, enabling the collection, processing, analysis, and dissemination of valuable information, gathered in a variety of environments. Such networks generate high volumes of data that needs to be converted into usable knowledge. Extracting useful information from a complex sensing environment at different spatial and temporal resolutions is a challenging research problem in artificial intelligence. In the domain of water, this emerging subject is referred to as hydroinformatics. In this paper, we propose a Fuzzy Rule based system to assess the quality of ground water samples based on their salt contents and thus predict its potability.

II. Materials and Methods

The goal of this work is to develop a classification approach that groups data with multiple attributes into relevant categories. It has become the most valuable process for acquiring implicit knowledge among datasets. The overall goal of the project is to develop a fuzzy rule-based system which can assess the water and predict its potability. The techniques used for spatial prediction of groundwater include decision trees, rule-based and nearest-neighbor classifiers, and support vector machines. In this section, we first describe current state-of-the-art followed by the proposed method.

2.1 Existing Systems

In recent years, data mining (Han, J 2000) has emerged as an approach for analyzing natural resources. Specifically, it is a classification approach to grouping data with multiple attributes into relevant categories. It

has become the most valuable process for acquiring implicit knowledge among datasets. Well-developed techniques used for spatial prediction of groundwater include decision trees, rule-based and nearest-neighbor classifiers, and support vector machines (M. C. L. Yu 2013).

The decision tree is one of the significant machine learning algorithms. Decision tree involves algorithmic acquisition of structured knowledge in the form of concepts, decisions trees and discrimination nets or production rules. Research has been conducted in Australia to determine an accurate groundwater balance for a region (Expert System for Water Treatment Plant Design,2001). This required estimation of groundwater recharge and discharge rates and also the knowledge of their spatial distribution. The parameters such as climate, soils, near surface geology, landforms, vegetation etc. were identified to build the decision tree. Decision tree used here guides the user to the most appropriate estimate of ground water recharge or discharge given the data availability. The decision tree has parallel pathways directing the user to estimates based on: (i) utilization of statistically significant empirical relationships; and (ii) simple groundwater. These spreadsheets incorporate various parameters, for example, mean annual rainfall, soil type, vegetation type, evapotranspiration and land use change. Other parameters considered include rainfall chloride flux, groundwater chloride concentration, alterable depth, surface soil and regolith type, and aquifer porosity.

Demissie (Yonas, 2012) developed a complementary modeling framework to handle systematic error in physically based groundwater flow model applications that uses data-driven models of the errors during the calibration phase. The effectiveness of four error-correcting data-driven models, namely, artificial neural networks (ANN), support vector machines (SVM), decision trees (DT) and instance based weighting (IBW) is examined for forecasting head prediction errors and subsequently updating the head predictions at existing and proposed observation wells.

Rule based modeling (Manoucher A, 2010) was used for spatial prediction of groundwater quality in Beaufort West, in the Karoo region of South Africa. The groundwater quality data from about 100 bore wells with a 30 years span collected between 1970 and 2007 was used. The variables used in the analyses include chemicals such as chloride, sulphate, magnesium, sodium, phosphates, and calcium. These were used as predictors for groundwater quality and electrical conductivity.

Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. OOBNs (Molina JL, 2013) are applied to the domain of integrated water management and used as a Decision Support System (DSS). The method proposes an integrated analysis of a complex water system supplied by groundwater from four aquifers. After identifying all relevant factors related to water management in the area, key indicators were translated to variables within a Bayesian Network (BN) and the relationships between them investigated. Each network represented one of the four aquifer units. These individual BNs were then linked to form an OOBN, which was used to represent the complex real-world situation. In this way, a DSS to simulate the entire water system was constructed using a group of conventional BNs interlinked to form an OOBN.

2.2 Data Collection

Karnataka is a state in southern India and it is a semi-arid and industrialized. For the purpose of this study, samples from 24 districts have been chosen. Thirteen water quality parameters are collected from several locations in each district by the Government authorities as shown in Table 2. Due to inconsistency in the sensors available, the number of measured

Table 1: Drinking water quality standards as per World Health Organization (WHO)

Sl. No.	Salt	Desirable limit (mg/liter)	Permissible limit (mg/liter)
1	Calcium (Ca)	75	200
2	Chlorine (Cl)	250	1000
3	Electrical Conductivity (EC)	2000	3000
4	Fluoride (F)	1	1.5
5	Magnesium (Mg)	30	100
6	Nitrate (NO ₃)	40	50
7	Thorium (Th)	300	600
8	Sulphate (SO ₄)	200	400
9	Carbonate (CO ₃)	-	-
10	Bi-Carbonate (HCO ₃)	-	-
11	Potassium (K)	-	-
12	Sodium (Na)	-	-
13	Ph	6.5	8.5

parameters varies significantly between locations. Most of the data was collected predominantly between 2006 and 2007. The laboratory tests were conducted for different physico-chemical potable water quality parameters as per the standard procedure (NatarajanVenkat Kumar, 2010),(Rahimi D, 2012). Out of the thirteen measurements, we are concerned with 8 salt compositions (first 8 elements in Table 1). In order to assess the water potability, an automated system is required that converts the 8 salts composition into a human understandable measure of potability. In the presented work, the measure of potability is given by four discrete values: (a) high quality, (b) moderate quality, (c) marginal quality and (d) not potable. Due to the availability of the salt data, the districts in Table 2 were considered for analysis and developing fuzzy rules.

Table 2: Summary of data collection including the time of collection, number of locations and the number of parameters collected

District Name	Data Collection Period (predominant)	Number of Locations	Number of parameters measured
Bagalkote	2006-2007	29	5
Bangalore Rural	2006-2007	34	5
Belagavi	2006-2007	66	5
Bellary	2004-2007	13	13
Bidar	2004-2007	25	13
Chamarajanagar	2005-2007	18	13
Chikkamagalur	2005-2007	12	13
Chitradurga	2005-2007	21	5
Coorg	2006-2007	23	5
Dakshina Kannada	2006-2007	30	5
Davanagere	2006-2007	11	13
Dharwada	2006-2007	23	13
Gulbarga	2006-2007	19	5
Hassan	2006-2007	40	5
Haveri	2006-2007	17	13
Kolar	2006-2007	51	13
Koppala	2006-2007	26	13
Mandya	2006-2007	76	13
Mysore	2006-2007, 2011	20, 14	13
Raichur	2005-2007	29	13
Shimoga	2006-2007	27	13
Tumkur	2006-2007	19	5
Udupi	2006-2007	21	5
Uttara Kannada	2006-2007	32	13

III. Development of Fuzzy Rule Based System

A Fuzzy Rule based system is developed to assess the water quality and predict its potability. A fuzzy rule-based system developed consists of two parts (a) knowledge base and (b) Inference Engine. A knowledge base is an information repository that provides a means for information to be collected, organized, shared, searched and utilized. It can be either machine-readable or intended for human use. Data in a knowledge base may be of two types. They are a) Facts: consisting of measurable quantities and quantities that can be derived from these measurable quantities and b) Heuristics: that are data obtained from human expertise on that subject. The knowledge base of our project contains two types of data. One is the concentrations of salts of various water samples collected from the bore-wells. This data is stored in a multidimensional format. The weka tool was used to perform various forms of analytics on this data (Rossen, 2007). The other is the accepted limits of all these salts prescribed by World Health Organization (WHO) standard that is given in Table 1. This data is stored in the form of a relational database. The rules are generated from the facts using the proposed system. An inference engine as shown in Figure 1 tries to derive answers for a real world problem using data stored in a knowledge base. It is the *brain* that the expert system uses to reason about the information in the knowledge base for the ultimate purpose of formulating conclusions (Yonas, 2012). Basically this unit consists of all the rules, which will be used to assess the quality of water. A Fuzzy Logic approach is proposed that uses the principle of *fuzzification* to convert crisp data into fuzzy values to predict the potability of the water quality. Membership

function was generated in order to fuzzify the concentrations of each salt in the water samples. Using the fuzzy values for each salt for a particular region, an overall membership value was calculated for a particular bore-well. Thus using these fuzzy values, the rule-based system makes a decision as to whether the water sample is potable or not. The output is the quality of the drinking water sample in four discrete values.

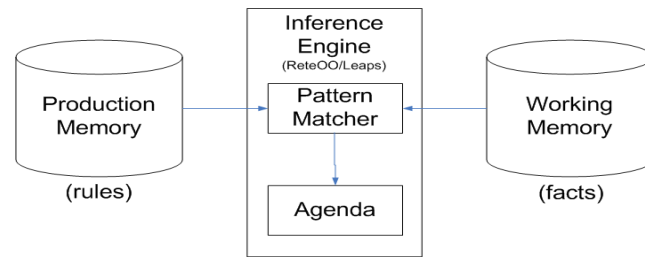


Figure 1: Inference Engine

1.1 Fuzzy Logic and its Role

Fuzzy logic was initially developed by Zadeh *et. al.* in 1965 as a generalization of classic logic (L.A. Zadeh 1968). The Fuzzy set is defined as “a class of objects with a continuum of grades of memberships”; the value 0 means that x is not a member of the fuzzy set; the value 1 means that x is fully a member of the fuzzy set. The values between 0 and 1 characterize fuzzy members that belong to the fuzzy set only partially. Traditionally, thematic maps are represented with discrete attributes based on Boolean memberships, such as polygons, lines and points. These types of entities either have a value or do not have it; an intermediate option is not available. With the Fuzzy theory, the spatial entities are associated with membership grades that indicate to what extent the entities belong to a class (Thomas L, 2009). Mathematically, a fuzzy set can be defined as

$$A = \left\{ \frac{\mu(x)}{x} \right\} \text{ For each } x \in X \quad (1)$$

where, $\mu(x)$ is the function membership function (MF) that defines the grade of membership of x in A . The MF takes values between and including 1 and 0 for all A . If $X = \{x_1, x_2, \dots, x_n\}$, equation (1) can be written as

$$A = \left\{ \frac{\mu(x_1)}{x_1} + \frac{\mu(x_2)}{x_2} + \dots + \frac{\mu(x_n)}{x_n} \right\} \quad (2)$$

In simple terms, the equations mean that for every x that belongs to the set X , there is a membership function that describes the degree of belonging of x in A . A new membership function as shown in equation 1 is generated to fuzzify the data.

$$\mu(x) = f(x) = \left\{ \begin{array}{l} 0, x \leq a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, x \geq b \end{array} \right\} \quad (3)$$

This rule is used to fuzzify the data that consists of the salt concentrations of various salts in the ground water samples that were collected. Then weights are assigned to each of the fuzzy values. An overall membership value is calculated using equation 4 based on weights assigned.

$$\sum w_i X\mu_i = \mu(A). \text{ Where } \sum_{i=1}^{i=n-1} w_i = 1 \quad (4)$$

The weights are calculated using Analytical Hierarchical Processing (AHP) (Oliver Meixner, 2009),(Raffaelelannone, 2011). AHP is a structured method for organising complex decisions such as water potability. This technique uses a pair wise comparison matrix which has an entry corresponding to each salt pair in the data. By using AHP, parameters that are considered for potability of water is decomposed into hierarchy of more easily comprehended sub-problems, each of which is analysed independently. Once the hierarchy is built, its elements are systematically evaluated by pairwise comparison with respect to their impact on an element above them in the hierarchy. The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight is derived for each element of the hierarchy that allows complex decision making.

IV. Result

As explained earlier, for every salt, the membership value is calculated for that salt of a particular water sample. In addition, an overall membership value is calculated considering the membership values of all the salts present in the water sample (as shown in equation 4). Based on this overall membership value, the water sample is categorized into one of the four categories (high, low marginal and moderate) to describe the quality of water as shown in Table 3.

Table 3: Classification of Groundwater based on calculated Fuzzy values

Fuzzy values	Groundwater quality	Class
0-0.25	Highly quality	1
0.25-0.5	Moderate quality	2
0.5-0.75	Marginal quality	3
0.75-1	Non suitable	4

AHP weights are assigned to the salts that were used to assess the portability of water. The AHP weights thus calculated for each salt is shown in

Table 4.

Table 4 shows the permissible and desirable limits for various salts according to WHO standards with their AHP weights. It is clear that the concentration of F, NO₃ and Mg are the major causes for non-potability of water in Karnataka state (southern state in India) based on the weights allocated to them.

Table 4: AHP Weights assigned to each salt based on our input dataset

Salt	Desirable (mg/liter)	Permissible (mg/liter)	Weight
Ca	75	200	0.10507
Cl	250	1000	0.02217
EC	2000	3000	0.01662
F	1	1.5	0.33744
Mg	30	100	0.17919
NO ₃	40	50	0.21178
SO ₄	200	400	0.07185
Th	300	600	0.05584

In order to validate our method, one of the major districts (Mysore) is considered. Table 5 summarizes the results with original raw data collected for Mysore district. As it can be seen, excess values of NO₃ in 7 out of 14 bore-wells of Mysore renders them not fit for drinking. The results of the proposed method match these values with full agreement. As per Table 5, our results indicate water from 50% bore-wells was not suitable for drinking exactly matching the data provided by the authorities (100 percent accuracy for Mysore). Moreover, by close observation of NO₃- values, it becomes clear that the non-potability was due to higher concentration of NO₃-.

Table 5: Results for Mysore District (in 2011) along with input data (non-permissible NO₃ values are highlighted)

LOCATION	pH	EC	CO ₃ ⁻⁻	HC O ₃ ⁻	Cl ⁻	NO ₃	SO ₄	F ⁻	Ca ⁺⁺	Mg ⁺⁺	TH	Na ⁺	K ⁺	Results
Udatyagiri	7.9	1750	0	555	177	80	90	0.7	40	128	630	111	3	Not fit for drinking
Indiranagar	7.8	890	0	372	57	40	24	0.8	48	51	330	50	3.1	Fit for drinking
Sachidananda Ashram	7.6	1240	0	433	121	24	64	0.7	72	48	380	108	2.7	Fit for drinking
Srirampura	7.4	2600	0	610	390	118	142	0.5	84	189	990	130	3	Not fit for drinking
Nachanallypalya	7.7	970	0	390	57	51	30	1.5	24	53	280	91	4.5	Not fit for drinking
Railway Colony	7.9	1090	0	366	106	27	66	0.67	76	46	380	69	4	Fit for drinking

Sunnada Keri	7.6	840	0	281	78	39	42	0.61	52	41	300	50	2.8	Fit for drinking
Chamraja Mohalla	7.4	670	0	238	57	39	24	0.7	44	29	230	44	3.4	Fit for drinking
Kukkarahally	7.6	900	0	360	78	18	26	0.6	32	60	330	53	2.8	Fit for drinking
Ontikoppal	7.9	1450	0	378	199	63	78	0.4	128	14	380	152	2.1	Not fit for drinking
Hinkal	7.8	940	0	366	71	41	32	0.9	56	53	360	46	2.8	Fit for drinking
Kavery Nagar	7.3	1360	0	335	142	110	98	0.63	128	34	460	97	4.4	Not fit for drinking
Veeranagere	7.4	1320	0	409	135	80	60	0.1	64	41	330	97	90	Not fit for drinking
Dhanaganahally	7.4	1750	0	366	241	140	112	0.6	152	70	670	84	4.7	Not fit for drinking

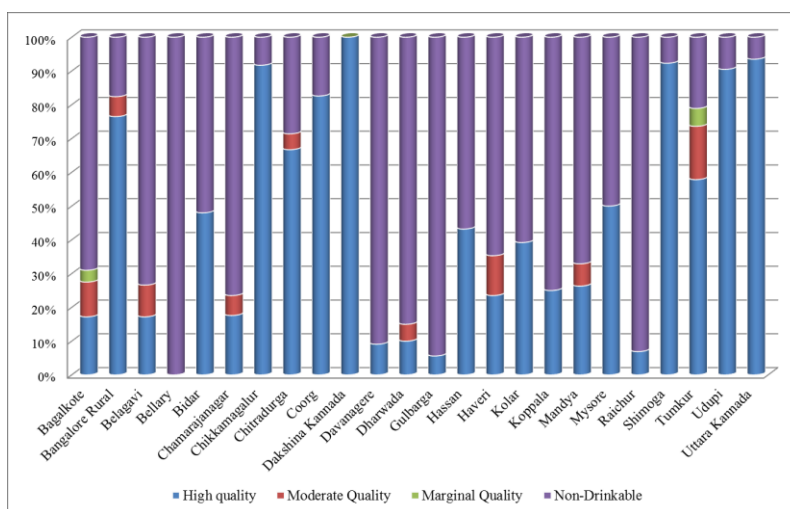


Figure 2: Results of water quality level for all the districts

Figure 2 summarizes the results of water potability obtained for all the districts using the proposed method. Interestingly, 15 out of 24 districts have non-potable ground water in 50% locations tested with Bellary (a mining district) having no potable ground water. On the other hand, Dakshina Kannada has high quality ground water in all the locations tested. For improving the quality of potable ground water, it is critical to know the reasons behind contamination.

Table 6 summarizes the contribution of different properties measured in ground water contamination. As it can be seen, a total of 48.22% of the locations are potable in Karnataka. 38.31% of the locations are affected by high NO₃ contamination with minor contributions from high electric conductance and presence of Fluoride. Thorium and Magnesium is present in three locations tested; Ca, Cl and SO₄ contamination is not an issue in Karnataka.

Table 6: Percentage of affected locations due to different salts

Salts	Percentage
EC	6.95%
Cl	0.00%
NO ₃	38.31%
SO ₄	0.00%
F	5.92%
Ca	0.00%

Mg	0.44%
Th	0.15%
Unaffected	48.22%

V. Conclusions

Potability of ground water is critical in India due to high population and repeated failure of monsoon rains. Many contaminations affect potability of ground water. An automated system that can take salt concentrations as input and generates level of potability of ground water as output is essential. In this paper, a Fuzzy rule-based system is reported to assess the quality of water for potability based upon the concentration of salts. It was found that twenty districts in Karnataka have a large percentage of ground water that is not suitable for consumption and only few districts like Chikmagalur and Dakshina Kannada have good drinkable ground water.

Acknowledgement

We would like to thank Central Ground Water Board, Bangalore, Karnataka, for sharing the data required for our experimental analysis. I also thank Mr. Naveen Kumar H N for helping me out in fine tuning my paper.

References

- [1] Anastassia Ailamaki, Christos Faloutsos, Paul S. Fischbeck, Mitchell J. Small, Jeanne VanBriesen, An environmental sensor network to determine drinking water quality and security, SIGMOD Record, Vol. 32, No. 4, December 2003, pp. 47-52.
- [2] Expert System for Water Treatment Plant Design - A dissertation, Rand Affrikans University, Sept 2001.
- [3] Ghisi E, Cardoso KA, Rupp RF, Short-term versus long-term rainfall time series in the assessment of potable water savings by using rainwater in houses, Journal of Environment management, 2012 Jun 15;100:109-19.D.O.I. : 10.1016
- [4] Han, J.; Kamber, M. Data Mining: Concepts and Techniques; Morgan Kaufmann: 2000.
- [5] L.A. Zadeh, "Fuzzy algorithms," Info. & Ctl, Vol. 12, 1968, pp. 94-102.
- [6] M. C. L. Yu, I. Cartwright, J. L. Braden, and S. T. de Bree, Examining the spatial and temporal variation of groundwater inflows to a valley-to-flood plain river using 222Rn, geochemistry and river discharge: the Ovens River, southeast Australia, Hydrol. Earth Syst. Sci. Discuss, 10, 5225-5267, 2013.D.O.I. :10.5194/hessd-10-5225-2013.
- [7] Manoucher A, Mini karim C, Annette John, A comparison on a different rule based statistical models for modeling geogenic groundwater contamination, Journal of Environmental modeling & Software, Vol 25, Issue 12, December 2010, pp. 1650-1657.
- [8] Molina JL, Pulido-Velazquez M, LiopisAlbert C, Periahero S, Stochastic hydro-economic model for groundwater quality management using Bayesian networks., Journal of Water Science technology, 2013;67(3):579-86. doi: 10.2166/wst.2012.598
- [9] NatarajanVenkat Kumar, Samson Mathew, Ganapathiram Swaminathan, A Hybrid Approach towards the Assessment of Groundwater Quality for Potability: A Fuzzy Logic and GIS Based Case Study of Tiruchirappalli City, India. Journal of Geographic Information System, 2010, 2, 152-162.
- [10] Oliver Meixner, Fuzzy AHP Group Decision Analysis and its Application for the Evaluation of Energy Sources, Proceedings of the 10 th International Symposium on the Analytic Hierarchy/Network Process Multi-criteria Decision Making, July 29 - August 1 2009, University of Pittsburgh
- [11] RaffaeleIannone, Salvatore Miranda, Veronica Napoli and Stefano Riemma - A structured approach based on AHP and Fuzzy Logic to estimate value of industrial assets. 2011 3rd International Conference on Information and Financial Engineering IPEDR vol.12.
- [12] Rahimi D, Mokarram M. - Assessing the groundwater quality by applying fuzzy logic in GIS environment- A case study in Southwest Iran. International Journal Of Environmental Sciences Volume 2, No 3, 2012
- [13] Rossen Dimov, Weka: Practical machine learning tools and techniques with Java implementations, AI Tools Seminar University of Saarland, WS 06/07, 2007
- [14] Ting-Nien Wu, Jan-Yee Lee, Chen-Hsiang Huang, Application of Time Series Analysis on Temporal Variation of Fluoride in Groundwater around Southern Taiwan Science Park. 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010).
- [15] Thomas L. Saaty - Decision making with the analytic hierarchy process. Int. J. Services Sciences, Vol. 1, No. 1, 2008.13-19.
- [16] Uttam kumar, Mukhopodhya C, Ramachandra T V, Spatial Data mining and modelling for visualization and urbanization, SCIT Journal, ISSN-0974-5076, Volume 9, Aug 2009, pp. 1-9.
- [17] Yonas Kassa Demissie, Data-driven models to enhance physically-based groundwater model predictions, Doctoral thesis, University of Illinois at Urban- Champaign Champaign, IL, USA, ISBN: 978-0-549-63954-1, 2012.