PalArch's Journal of Archaeology of Egypt / Egyptology

Ground Water Quality Assessment in Guntur district GIS data Using Data Mining Techniques

Mandadi Vasavi[1], Munugapati Bhavana[2] , S J R K Padminivalli V[3]

[1,3] Assistant Professor, Department of Computer Science & Engineering, R.V.R & J.C College of Engineering, Chowdavaram, Guntur, India.

[2]Assistant Professor, Department of CSE, Geethanjali College of Engineering and Technology, Cheeryal Village, Hyderabad, India.

Mandadi Vasavi , Munugapati Bhavana , S J R K Padminivalli V[3], Ground Water Quality Assessment in Guntur district GIS data Using Data Mining Techniques , Palarch's Journal Of Archaeology Of Egypt/Egyptology 18 (4). ISSN 1567-214x.

Keywords -- Ground Water Quality, prediction, Data Mining, Classification Techniques.

Abstract:

Throughout this article we first attempted to analyse water quality research in the Andhra Pradesh district of Guntur. A thorough study of the consistency of groundwater was undertaken. 31 water samples of various physiochemical parameters, e.g. temperature, pH , electrical conductivity, totally dissolved solids , ammonium nitrates, total hardness, calcium, chloride, magnesium, sulphate, total alkalinity, potassium, total nitrogen, sodium , total phosphorus and dissolved oxygen have been collected and tested. The correlation analysis was also conducted as it is an outstanding method to estimate fair precision of parameter values. This research proposes a new methodological approach in conjunction with an ensemble model for data mining, through the use of the evidence-based confidence function and boosting the BRT regression tree GIS knowledge for groundwater quality visualization in Guntur. Spring areas for training and validation in individual and ensemble methods can be established and subdivided into two groups. Modeling results are drawn up to create potential maps for spring (groundwater). In order to evaluate groundwater content by taking different samples in various towns and cleanly synthesising water parameters that have been applied the diverse Data mining techniques.

Introduction:

Several more researchers employed algorithms for statistics and data mining to map the potential of soil water. Some of them used Spring Sites as markers for groundwater supplies, and others used Spring Sites. According to the literature, one of the most common approaches used by scientists is the frequency ratio, weights of proof and entropy index. Moreover, the ability of groundwaters is commonly evaluated via other methods of data mining like Random Forest, Classifying and Regression Tree and Boosted Retrocess Tree (BRT). While data mining technology has proven reliable in the process of nonlinear and complex data, one of the pitfalls is that it impacts on the accuracy of the predictions and the validity of the models.

Different data mining algorithms, including random-forests, BRT, supply vectors, an artificious neural grid, quadratic discrimination analyses, linear discrimination, flexible discrimination analysis, penalised discrimination analysis, k-nearest neighbours and multivariate adaptive regression splines, w-recent papers by Naghibi and Moradi Diehtpagerdi (2016) and Naghibi et al . (2018), The Proof Belief Function (EBF) approach for mapping groundwater capacity requires additional techniques. The results showed that Nampak et al . performed better than the EBF model. (2014) to chart and compare EBF performance with a logistic regression model, by using the EBF. The success of the EBF model was also investigated and its findings compared to random, random forest, BRT and generalised linear model. In another research project. The findings provided an acceptable score for the EBF model.

The problem of pollution is growing everyday which induces groundwater and other sources of pollution. The consistency of soil water depends on its suitability for drinking , irrigation and other uses. However, it can not use swimming as well as leisure and as a means of raw water supply because of depletion of ground water quality. Contaminated water can affect consumers' health and financial status.

Inadequate sewage storage and recycling (human and animal); agricultural discharges; and overuse of resource water. Pollutants like bacteria , viruses, heavy metals The biggest concern with groundwater is that it is impossible to recover its consistency until it is polluted. Therefore, it is necessary and important that groundwater quality should be safeguarded and managed[1]. A high-use technique for correlating various parameters has been found in mathematical regression analysis.

The correlation analysis measures the close relationship between the chosen independent and dependent variables. The likelihood that the linear equation between x and y is nearer to +1 or +1 shows. This analyses the essence of the relationship between the factors and, therefore, the nature of the relationship between them and thus provides a forecast or forecast process.
Within the scope of the present study the status of quality of water was emphasised and the correlation analysis of different parameters was provided in groundwater samples obtained from different areas of the Guntur district[5].

For this function, the regression tree was chosen to be used as a data mining strategy as well as to incorporate stochastic gradient boosts in the collection of features (Naghibi et al. 2016) in order to decrease variance and bias. The BRT model also determines the significance for the simulation phase of the effect variables. In consideration of the strong characteristics of the BRT model above, this model has been selected in combination with the EBF model to maximise its predictive accuracy.

In this review, the proposed BRT ensemble method reinforces the shortcomings of each method to evaluate groundwater connexions between a single layer and with each class of individual layers and integrates the benefits. Since in soil water capacity evaluation this combined method is almost new, its efficacy and capability can be studied with this study. This study aims to enhance the efficiency of statistical science by applying an ensemble of data mining models to map future groundwater. The goals of the thesis are therefore:

(1) Assessment of the BRT model success in assessment of the capacity of groundwater.

(2) Rating the importance and the relationship between groundwater capacity and GCFs of groundwater conditioning variables.

(3) Provide geographic knowledge and guidelines in the Guntur groundwater management decision-making processes.

Even though groundwater does have a geographical background, analysis of its quality, quantity and variation of both across a wide area is significant, especially in the context of rural water supply programmes. If interpreted spatially, any natural resource with multi-dimensional implications is well known. In particular, the natural resources, water – surface water and groundwater, as well as related landscape elements, which regulate the distribution of a particular resources, if they are represented in the right manner, give a better perspective to help formulate and manage strategies.

Each household has a specific past, as groundwater is accessible in the landscape and used for its sustenance by the society. The traditional mapping technique, maintaining the chemicals research documents and attempt to apply to the analytical approaches such as multi-layer integration, tabular / characteristic data relation required for planning is complicated and time consuming. At around the same time, if correctly connected and with a synoptic perspective, the multi-dimensional information provides a knowledge foundation that is critical for development-oriented planning. The village wise database was created and analytical data were divided into two groups, i.e. allowable limit and limit categories. For average calculations, the maximum number of samples falling in a segment was considered and tabled in XLS format.

Related Work

In the year 2018 S.V.S Ganga Devi[1] has proposed ground water quality data analysis using classification techniques . Bayesian methods have been investigated. Artificial Neural Networks (ANN)s were used to develop model for prediction and monthly values of water quality parameters. The study was carried out in Kadapa located in A.P, There were 58 water samples collected and analysed in laboratory for physico-chemical properties, conditional inference tree technique was used for analysis. R language was used to perform the water data analysis for accuracy assessment.

In the year 2018 JyotiBansaland A.K. Dwivedi [2] proposed the study of water quality using Water quality index (WQI) measurement for Indore groundwater and physico-chemical parameters. The water quality indices used included: Electric conductivity (EC), pH , total hardness, ODC, total allalinity, TDS, sulphate, chloride, turbidity. Once WQI is measured, the results were compared with IS: 10500-2012.The results of water quality assessment show that the parameters were high in wet season than summer season.

In the year 2017 Javadi.S and Hashemy.M. [3] proposed ground water vulnerability assessment using vulnerability maps. They have used DRASTIC index. K-means clustering technique was applied in four features of depth to ground water. They have discussed about advantages and disadvantages of DRASTIC method. The study area of the project was Hashtgerd plain located in central part of Alborz Province. There were 4936 samples points data were used in this study.

In the year 2017 Ajinkya P. Chatur, Prof. R. V. Mante, Amol R. Dhakne[4] has proposed the review of various procedures for forecast of ground water level and ground water quality. The methods discussed for ground water level prediction were SVM, KNN, CBA (Classification Based on Association rule), Naïve Bayes. The techniques discussed for ground water quality and ground water levels were ANN, NB, Decision Tree, FNN(Fuzzy Neural Networks) and Back Propagation Neural Network. They concluded that CBA out performed when compared to other methods in terms of their performances.

In 2017 Mr.Sudhir M. Gorade, Prof. AnkitDeo, Prof. PreeteshPurohit [5] proposed investigation of a few information mining arrangement strategies. There were two stages in Classification. The initial step was a model dependent on informational index, and in second step the model was utilized to order an obscure tuple into a class mark. They examined about different order models, K-Nearest Neighbor, Decision Trees, Support Vector Machines, Neural Networks, Naive Bayesian Classifiers. They examined points of interest and burdens of these methods. They reasoned that the vast majority of the students incline toward Decision tree, Bayesian classifiers, back spread, bolster vector machines, in light of the fact that these methods use preparing tuples to build a speculation display. Some languid student like closest neighbor classifiers and case-based thinking. These procedures store preparing tuples in example space and hold up until gave a test tuple.

In the year 2016 Priya, Dr. R. Mallika[6] has used linear regression to estimate crop yield water quality. Noyyal River Basin was the research area of this initiative. 12 separate sites obtained water samples. The TDS data model using Electrical Conductivity (EC) spatio-temporal data was developed. Data sources have been obtained by the TWAD board of directors for the years 2013 to 2016. Linear Regression data model for TDS using Mg, Na, Cl, EC also developed in this project.

In the year 2013 Kamakshaiah Kolli, Dr. R. Seshadri [7] has specified the assessment of the content of surface water has been proposed using data mining techniques which have been studied with a reference to F- concentration in rural areas of Guntur, Andhra Pradesh , India. It has been suggested that ground water is alkaline, mild to very hard and mostly blackish.

Study Area and Preparation of the Conditioning Factors

Study Area

Tadepalli has been on Krishna's southern bank. It is located at a latitude of 16° 24' N to 16° 24' N and a longitude of 80° 34' E to 80° 34' E in the district of Guntur. This is 22 metres (72 ft) in average height. Villages in Tadepalle mandal include: Ippatam, Chirravur, Kunchanapalli, Mellampudi, Prathuru, Tadepalli, Undavalli and Vaddeswaram. Village of Tadepalle mandal is a part of Tadepalle mandal. The entire surface area is 19618.87 acres, the highest humidity in the summer and the lowest humidity in the early winter. The mean temperature is 38 ° C, while the minimum is 18 ° C and the population is 80. 887 (Census 2001). Each year, the region of the sample is 1040 mm.

Ground water quality data generation

The analysis has obtained a sufficient number of ground water samples and physicochemical analysis analyses the selected water quality parameters. The metrics of water quality chosen include pH, electrical conductivity, chlorides, sulphates, nitrate, overall durability, total solids, fluorides, Alkalinity, sodium , potassium and phosphates. There are a total of 45 samples of ground water provided at default sites.

Water Quality Index

A large number of water quality data is measured to reduce WQI to a single numerical value that represents the cumulative water quality at a given location and time on the basis of many criteria for the water quality. It is also known as a ranking that represents the cumulative effect of various parameters on the overall water quality. Horton's first suggestion (1965) was according to the definition of indices that reflect water quality gradation. The key aim of the Water Quality Index is to convert nuanced data on water quality into publicly comprehensible and accessible information. The Water Quality Index can provide a simple indicator of water quality based on certain very important parameters.

Materials and methods

A spring could be emphasized by the fact that allows groundwater to flow from either a tank to the surface. This analysis indicates, based on physiographical and hydrological features of the study area, that events of natural spring and their release rate may be correlated with the possible soil supply within the studied Basin. A Ground water Potential Map is suggested in order to measure this relationship as an instrument for calculating the relationship between the spring occurrences and successful factors known as conditioning factors. Two data sets, including a spring position inventory and GCF, were developed for the modelling of groundwater capacity. The above data sets used the EBF model and the arcGIS 10.4 for the corresponding GPM. The next step was the extraction and use of EBF values as an input to the BRT model and the training of the set BRT model. Ultimately , the effectiveness of the BRT approaches was confirmed by the application of the ROC map.

Data preparation

Based on field surveys, a spring stock dataset was prepared containing 94 springs (2014). The data set was then divided into two testing subsets (70% of the data set: 66 fountains) and the model validated (30% of the dataset: 28 fountains). It is important to remember that the spring dataset was split into two sub-sets.

The BRT is indeed a data mining learning approach that includes both a correlational method and techniques of boosting, and can be used for problems of both regression and classification. It attempts to improve the effectiveness and predictability of single approaches by the combination of multiple fitted versions. Boosting is used to integrate the outcomes of decision-making trees close to model averages. Some of the parameters in this model need to be optimised, for example, a number of trees, a reduced shrinkage and a depth of interaction. The value of trees in a construct model is defined by shrinking or learning rates (Naghibi et al . , 2016). The number of nodes in trees is calculated by connexion depth or complexity.

You should describe the BRT model as follows

Weights starting to fi = 1 / n. Classifier for m = 1 to Cm):

1. Run the weighted data classifier Cm
2. Calculate error rate rm RM
3. Consider the αmlog (1-rm)/rm classification weight
4. Wi= wi exp[αmI(yi = Cm)] Measure weights wi

The major vote will ultimately be achieved by:

sign =ΣM, m−1 αm Cm (X)

The best parameter set in BRT is chosen using the exactness index and the Water Quality index, which can be determined as follows:

Accuracy = TP + TN / TP + TN + FP + FN

WQI = Pobs−Pexp / 1−Pobs

Pobs = TP + TN /n

Pexp = (TP + FN)(TP + FP) + (FP + TN) (FN + TN) / √N

While n seems to be the correctly classified ratio of cells and N indicates a total number of cells of preparation, TP, FP, TN and FN, respectively, represent true positive, incorrect positive and true negative.

The EBF model was used first in implementing the new data mining ensemble model, and the values for the various GCF groups were allocated. The search feature in ArcGIS 10.4 then generated new maps of each factor. For training the data-mining model (i.e., BRT), a new dataset was given. In this dataset, 1 was allocated to non-spring positions in the spring and 0. The random description of the nonspring positions using ArcGIS 10.4 is noted. The BRT model has been implemented using R open-source software using the gbm kit, using the current training dataset and current GCF layers with Bel values.

The BRT approach was developed by a 10-fold cross-validation, which was considered to be sufficient runs for the parameter optimisation. The GPMs developed with the EBF and BBF-BRT methods must be explained as being classified in four categories — small , medium, moderate and high by the system of natural break classification (Naghibi et al. 2018).

Sampling and sample analysis

This research gathered groundwater samples from 20 wells of 10 to 90 m depth and discharges from various land use patterns of 1 to 6 L / s, including from open-tube and agribusiness areas for the purposes of assessing heavy metal emissions during spring and summer 2019. The sampling stations in the sample area are shown in Figure 1. The samples were gathered in 200 mL of polyethylene acid-washed bottles to resist predictable characteristic shifts according to normal procedures. The samples obtained were purified and frozen at 6N HNO3 and deposited for further study at a temperature of 4 ° C. A water sample with ICP-OES (Variant, 710-ES, Australia) has been defined as concentration of heavy metals (As, Zn , Pb and Cu).

Index of Heavy Metal Pollution

The HPI contains the best water content of heavy metals based on a qualitatively average weighted arithmetic process. In two phases, the HPI is generated. First of all, a grading scale is defined for each pa-rameter that gives weight, and second, the pollution parameter that should be the basis for the index is chosen. The rater system is randomly between 0 and 1 and is chosen according to the significance of each consistency criterion, or it can be tested by translating values to the recommended parameter norm.

Heavy Metal Evaluation Index

In water samples HEI emphasizes on heavy metals to boost the consistency of water (33). The water quality indice is graded as HEI < 400 (low heavy metals), 400 (moderately to heavier metals) and HEI>800 (high heavy metals). The Water Quality Index is classificationled into three groups.

Results and Discussions

The effects of the concentrations of heavy metals for spring and summer in freshwater collections. Even the matrix for spring-summer correlations between el-ements.

The findings suggested that the concentrations of As, Zn , Pb, and Cu from a given city level obtained from Guntur during springtime ranged from 0,08 to 7,48, from 0,12 to 15,64, from 0,09 to 5,50 and from 0,89 to 13,58 μg / L.

Results suggest a range of 0,57 to 7,21, 0,41 to 16,42, 0,19 to 4,46, and 6,54 to 15,76 μg / L in the summertime concentrations of As, Zn , Pb and Cu, collected from Guntur from the stated city plain.

The 5 percent value association (P>0.05) for the demon was not important for the following pairs: As and Zn , Pb, Cu; Zn and Pb; Cu and Pb and Cu in spring and summer season water tests.

Cd, HPI and HEI calculation values for each site, correlation of index values with metal concentration, and correlation of spring and summer indices.

The seasons of Spring and the summer were -2.80 and -2.67 and show low pollution respectively. The computed HPI has demonstrated that the average value for drinking water at all points for spring and summer was 9.74 and 9.51, respectively. In comparison, the HEI determined indicates that the mean amounts is 1.20 and 1.32 in both spring and summer and suggest low heavy metal emissions.

For spring and summer samples a comparison between indices and the concentration of heavy metal has been shown to react dramatically with Pb. This reveals that Pb is the major contributors. In comparison, Cd, HPI and HEI are strongly associated. Consequently, there are the three existing methods; the Contamina FI, the HPI and the HEI.

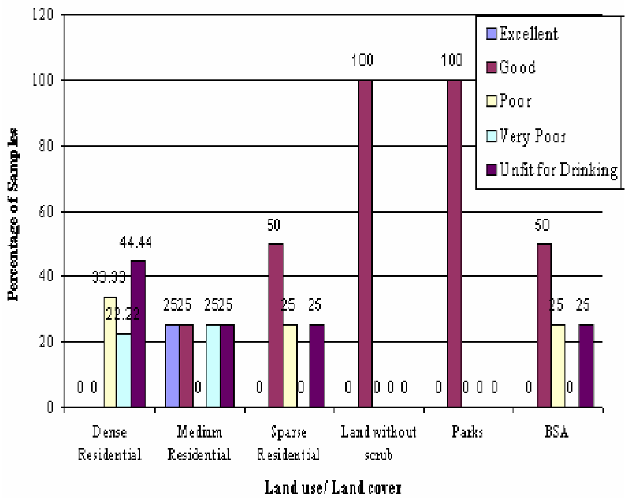

Figure1: Correlation with water quality of land use / coverage

Conclusions:

Groundwater potential modelling, which has attracted numerous scientists worldwide, is considered a key aspect in groundwater studies. The study introduced a new BRT model ensemble and assessed its performance in the field of groundwater mapping. A training set of the belief values taken from the results of the EBF model was used in the BRT model. The BRT models have been evaluated using the ROC curve. The results showed that the BRT model worked better than the basic EBF model. Thus the implementation of the BRT model can be inferred to increase the prediction power of the EBF model but the two models have an acceptable performance in this analysis. The BRT model's improved performance may be attributed to the stronger aspects of the BRT model such as its ability to handle nonlinear phenomena.

A review of the finding from the different stages of the work shows that the Remote Sensing and GIS integration is useful tools for planning different digital thematic layers and maps that demonstrate that different water quality parameters are spread spatially. The superposition on the satellite images of spatial distribution of water quality maps is a very authentic idea for defining and compare water quality issues with land use to explain why the quality of water deteriorates.

To ensure sustainable management of soil water, monitoring pollution patterns and trends in urbanisation is an important task. A Remote Sensing and GIS combined analysis reveals that the effect of land-use / land use on groundwater quality can be measured and evaluated. Relevant geographic distribution diagrams with different contamination criteria are used to comprehensively demarcate the local water distribution to help propose a systematic approach to groundwater contamination protection to remediation steps.

References :

[1] F. Aburub and W. Hadi, ―Predicting Groundwater Areas Using Data Mining Techniques : Groundwater in Jordan as Case Study,‖ vol. 10, no. 9, pp. 1621–1624, 2016.

[2] J. Bansal and A. K. Dwivedi, ―Assessment Of Ground Water Quality By Using Water Quality Index And Physico Chemical Parameters : Review Paper,‖ vol. 7, no. 2, pp. 170–174, 2018.

[3] A. P. Chatur, P. R. V Mante, and A. R. Dhakne, ―Brief Survey of Different Techniques for Prediction of Groundwater Level and Groundwater Quality,‖ vol. 4, no. 11, pp. 168–172, 2017.

[4] S. V. S. G. Devi, ―Ground Water Quality Data Analysis Using Classification Techniques.‖

[5] S. M. Gorade and P. A. Deo, ―A Study of Some Data Mining Classification Techniques,‖ pp. 3112–3115, 2017.

[6] Kamakshaiah Kolli, Dr. R. Seshadri – Ground Water Quality Assessment Using Data Mining techniques,‖ IJCA, vol. 76, no. 15, pp. 0975–8887, 2013.

[7] Bhaskar CV, Kumar K, Nagendrappa G. Assessment of heavy metals in water samples of certain locations situated around Tumkur, Karnataka, India. E-J Chem 2010; 7(2): 349-52.

[8] Rajankar PN, Gulhane SR, Tambekar DH, Ramteke DS, Wate SR. Water quality assessment of groundwater resources in Nagpur Region (India) Based on WQI. E-J Chem 2009; 6(3): 905-08.

[9] Rizwan R, Gurdeep S, Manish Kumar J. Application of heavy metal pollution index for ground water quality assessment in Angul District of Orassia, India. Int J Res Chem Environ 2011; 1(2): 118-22.

[10] Sobhanardakani S, Tayebi L, Farmany A, Cheraghi M. Analysis of trace elements (Cu, Cd and Zn) in muscle, gill and liver tissues of some fish species using anodic stripping voltammetry. Environ Monit Assess 2012; 184(11): 6607- 11.

[11] Hosseini SV, Aflaki F, Sobhanardakani S, Tayebi L, Babakhani Lashkan A, Regenstein JM. Analysis of mercury, selenium and tin concentrations in canned fish marketed in Iran. Environ 12. Monit Assess 2013; 185(8): 6407-12.

[12] Carvalho ML, Santiago S, Nunes ML. Assessment of the essential element and heavy metal content of edible fish muscle. Anal Bioanal Chem 2005; 382(2): 426-32.

[13] Järup L. Hazards of heavy metal contamination. Br Med Bull 2003; 68(1): 167-82.

[14] Tahsin N, Yankov B. Research on accumulation of zinc (Zn) and cadmium (Cd) in sunflower oil. Journal of Tekirdag Agricultural Faculty 2007; 4(1): 109-12.

[15] Sobhanardakani S, Jamshidi K. Assessment of metals (Co, Ni, and Zn) content in the sediments of Mighan Wetland using geo-accumulation index. Iranian Journal of Toxicology 2015; 9(30): 1386-90.

[16] Abou-Arab AAK, Ayesh AM, Amra HA, Naguib K. Characteristic levels of some pesticides and heavy metals in imported fish. Food Chem 1996; 57(4): 487-92.

[17] Dahiya S, Karpe R, Hegde AG, Sharma RM. Lead, cadmium and nickel in chocolate and candies from suburban areas of Mumbai. India. J Food Comp Anal 2005; 18 (6): 517-22.

[18] Hosseini SV, Sobhanardakani S, Tahergorabi R, Delfieh P. Selected heavy metals analysis of Persian sturgeon's (Acipenser persicus) caviar from Southern Caspian Sea. Biol Trace Elem Res 2013; 154(3): 357-62.

[19] Adekunle IM, Akinyemi MF. Lead levels of certain consumer products in Nigeria: a case study of smoked fish foods from Abeokuta. Food Chem Toxicol 2004; 42(9): 1463-8.

[20] Saracoglu S, Tuzen M, Soylak M. Evaluation of trace element contents of dried apricot samples from Turkey. J Hazard Mater 2009; 167(1-3): 647-52.

[21] Ackah M, Anim AK, Zakaria N, Osei J, Saah-Nyarko E, Gyamfi ET, et al. Determination of some heavy metal levels in soft drinks on the Ghanaian market using atomic absorption spectrometry method. Environ Monit Assess 2014; 186(12): 8499-507.

[22] Prasad B, Sangita K. Hevay metal pollution index of ground water of an abandoned open cast mine filled with fly Ash: a case study. Mine Water and the Environment 2008; 27(4): 265-7.