International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 6097-6101

Ground Water Quality Modelling for Irrigation Using Data Mining Technique and Spatio-Temporal Dates

R.Priya[1] and Dr. R.Mallika[2]

[1]*Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India.*
[2]*Assistant Professor, CBM College, Coimbatore, Tamil Nadu, India.*

[1]*Orcid: 0000-0002-1047-2015*

Abstract

Water availability and quality are the key factors which influence the development of agricultural sector. The quality of water is mainly depends on several parameters of water. Ground water quality may vary with respect to time and space. Ground water quality prediction is an important agricultural problem for farmers and it can be solved with the help of previously available data. Accordingly this paper emphasises prominence of data mining technique using spatio-temporal data to develop useful data models with high accuracy and relevant relationships to predict ground water quality. By and large, several physio-chemical parameters are influencing the quality of ground water; generally the quality of water is assessed based on the Total Dissolved Solids content present in water. For experimental purpose significant parameters such as EC, Na, Mg and Cl were taken as input. Linear regression technique is used in this study to develop data models and TDS is considered as the output of the model.

Keywords: Ground water quality, spatio-temporal data, Prediction models, linear regression, Arc GIS.

INTRODUCTION

Water is one of the most precious natural resource on the earth. In the present era, rapid industrialization in tandem with exponentially growing population generates enormous amount of pollutants and these pollutants are unscrupulously disposed in landfills and water bodies, as a result ground water gets contaminated. This contaminated water becomes unfit for irrigation. Since Ground water quality directly influencing the crop yield it is indispensable to scrutinize or predict the quality of water before going to use it for irrigation purpose [1 2,8]. Owing to lack of time and requisite facilities, scrutinizing the ground water quality before usage is quite difficult in many cases. In order to overcome this difficulty, predicting the ground water quality using the developed data model is a viable solution. Data mining techniques are useful in extracting information from huge amount of data set. Data mining is a renowned tool to develop data models. Data mining is the process of discovering significant patterns, trends and new correlation by analysing through large set of data using statistical and mathematical techniques as well as Patten recognition technology. Especially spatial data mining provides more relevant data for environmental modelling.

Spatial Data Mining (SDM) encompasses objects that are pigeonholed by spatial location or extension by numerous non-spatial attributes.SDM is advantageous in finding inherent regularities, relationships between spatial and non-spatial data [4]. By the same token Geographic Information System (GIS) is expedient for the user can get spatial data and analyse the data analytically. Geographical database establishes a spatio-temporal continuum in that properties relating to a particular place are commonly linked and described based on the properties of its neighbourhood [3, 4]. Temporal aspects related to spatial data are also equally important but are not often taken into account. In case of ground water quality predictions, it is essential to consider temporal aspects. Arc GIS software is used to create thematic maps so that visual interpretation can be made [3].

STUDY AREA

The ground water quality of Noyyal river basin was considered in this study. River Noyyal originates at the Velliangiri hills in Western Ghats and passes through the Coimbatore district, Erode and Karur district in Tamil Nadu state, India and joins Cauvery River at Kodumudy. The total area of the river basin is 3510 sq.km. The climate in the area is hot and humid with an average annual rainfall of about 671 mm. The area is characterized by an undulating topography with gentle slope from west to east. The area is underlined by wide range of high grade metamorphic rocks of the peninsular gneissic complex. The rocks are extensively weathered. Open wells, dug cum bore wells are the groundwater extraction structure in the area. The general gradient of water table in the area is 2.2 m/km towards east. The study area map is shown in fig .1.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 6097-6101

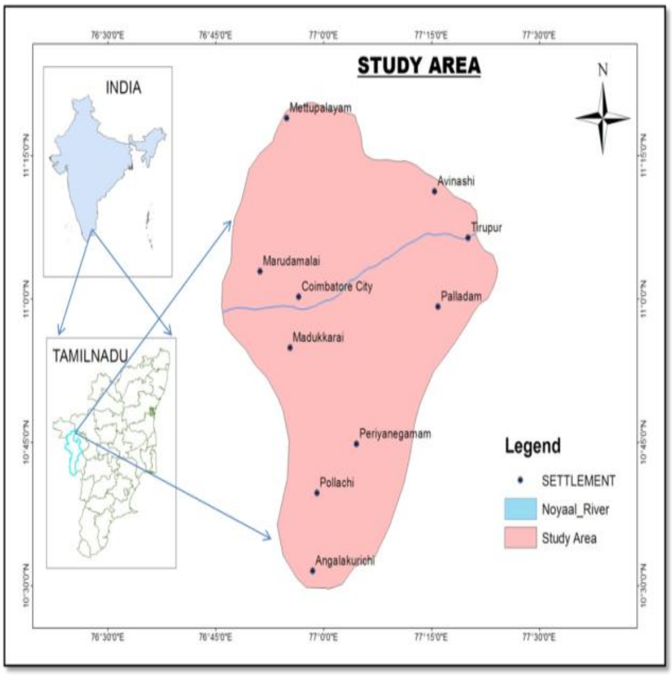

Figure 1: Study Area Map

The study area is located between latitude 11˚.2993' and 10˚.5773', longitude 77˚.0481' and 76˚.9352.12'. Water sample data were collected from 12 different places between the given latitude and longitude. Latitude and longitude of the sampling points selected are given in table.1.

Table 1: Latitude and Longitude

Sample id	Locations	Latitude	Lon Gitude
S1	NEELAMBUR	11˚.2715'	76˚.2239'
S2	KARUMATHAMPATI	11˚.1167'	77˚.1833'
S3	ARASUR	10˚.7857'	77˚.1751'
S4	VEERAPANDI	11˚.1756'	76˚.9622'
S5	METTUPALAYAM	11˚.2993'	76˚.9352'
S6	KODANGIPALAYAM	10˚.7908'	77˚.0356'
S7	SULUR	11˚.0227'	77˚.1257'
S8	THONDAMUTHUR	10˚.5773'	77˚.0481'
S9	MUTHAMPALAYAM	11˚.199'	76˚.9556'
S10	PEELAMEDU	11˚.0303'	77˚.0417'
S11	GANAPATHY	11˚.0393'	76˚.9787'
S12	VELLALUR	10˚.9678'	77˚.0073'

DATA PROCESSING

In order to predict the ground water quality, the data related to physical and chemical parameters of ground water were collected from TWAD (Tamil nadu Water Supply and Drainage department) board for pre-monsoon and post-monsoon (temporal data) from the year 2013 to 2016 for all 12 locations and linear regression data model is developed for each parameter. Linear regression is a learning function that maps a data item to a real-valued prediction variable. Compendium of variable data was taken in to account to train the prediction model to predict the dependant variables.

Spectral analysis is a tedious and time taking process; therefore each and every time it is quite difficult to repeat the same. A mechanism of executing the mapping from nominal to numeric values depends on the transmittance with statistical parameters of water was proposed [5]. For easy predictions, a supporting statistical method was

Developed based on linear regression curve fitting using normalized nominal attributes. Subsequently based on its R^2 value the reliability of the developed empirical equation was assessed.

Development of data model for TDS prediction using spatio-temporal data of Electrical Conductivity (EC)

TDS is considered as the index of general groundwater quality. TDS depends on several chemical parameters present in water. Using regression analysis, data models are developed to ascertain the relationship between the influencing chemical parameters and TDS. The developed data model will help to understand the linear relation between the parameters taken for the analysis. Based on the observed values of the TDS the regression model is developed. The regression model TDS = 0.6643EC + 96.074 is at a dependable correlation level of 99.1% of the sample data. The correlation coefficient shows the reliability of the data model [7].

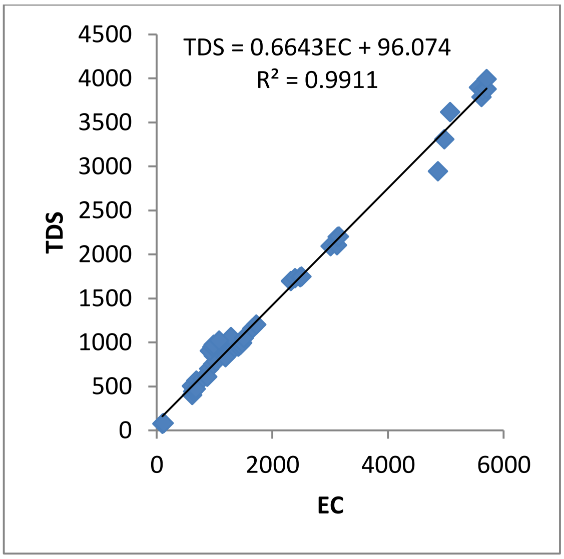

Figure 2: Regression data model for TDS using EC

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 16 (2017) pp. 6097-6101

Spatial distribution of EC

The spatial distribution of high concentration of EC values of samples is presented by pale green colour and it is shown in the Figure 3. From the spatial distribution diagram, it is understood that the groundwater from some scattered areas in North, East, west and south part of the study area which contain groundwater of poor quality with respect to EC during pre-monsoon & post monsoon season. Using IS 2269-1963 the EC ranges were classified in to low, medium, high and very high. The EC ranges were shown in Table.2.

Table 2: Ranges of EC

RANGE	LABEL
100-1000	LOW
1000-2000	MEDIUM
2000-3200	HIGH
3200-5700	VERY HIGH

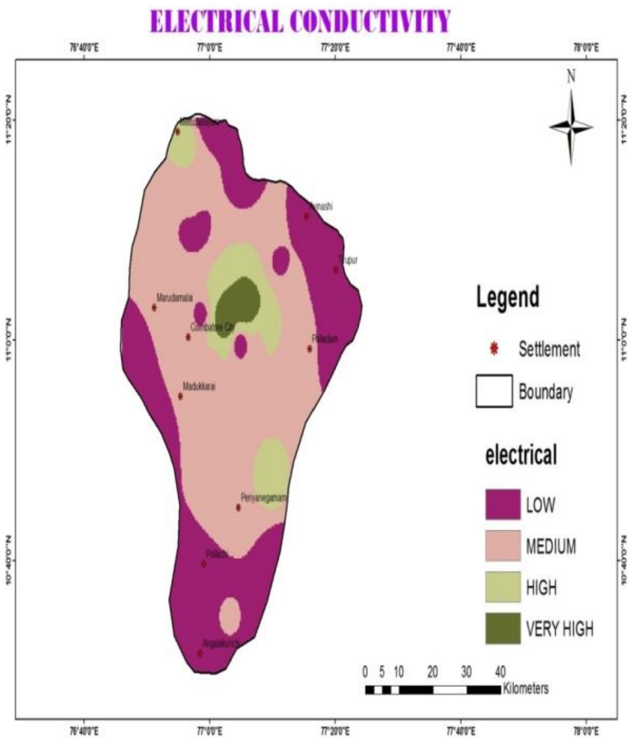

Figure 3: Spatial distribution map for EC

Development of data model for TDS prediction using spatio-temporal data of Magnesium (Mg)

Based on the collected spatio-temporal data for Mg, the data model is developed using regression technique [9].The regression model TDS = 44.68Mg + 311.35 is at a dependable correlation level of 88% of the sample data. The correlation coefficient obtained for the data model was found to be 0.8799 [7] [10].

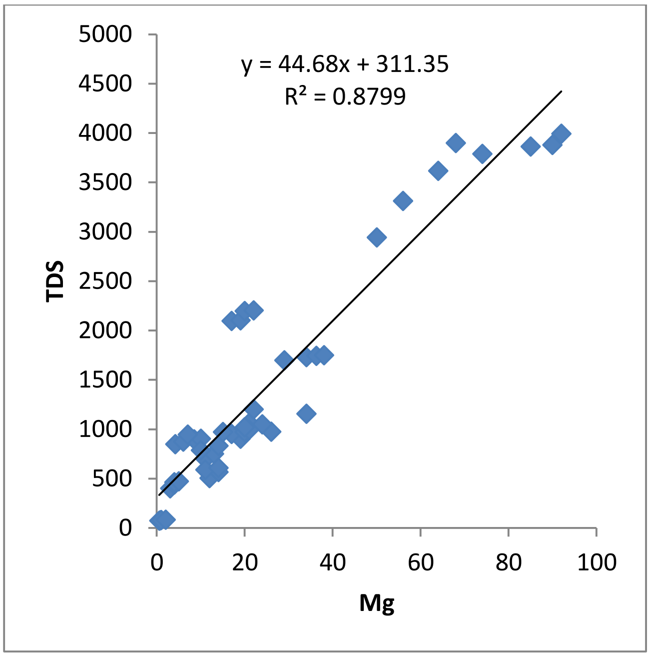

Figure 4: Regression data model for TDS using Mg

Spatial distribution of Mg

The spatial distribution of Mg values of these samples is shown in the Figure.5. From the spatial distribution diagram, it is observed that the groundwater from the south part of the study area & some scattered parts of East, North &Central North part has groundwater with Na contents above the permissible limit during pre-monsoon & post monsoon season [10].The permissible limit value of Mg is taken from IS 2269-1963. Table.3 shows the permissible limit values of Mg.

Table 3: Range of Magnesium

RANGE	LABEL
2-30	LOW
30-50	MEDIUM
50-100	HIGH
100-200	VERY HIGH

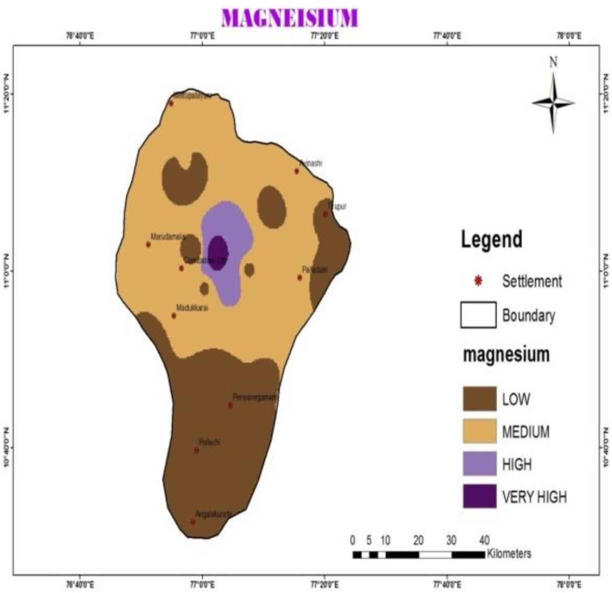

Figure 5: Spatial distribution map for Mg

Development of data model for TDS prediction using spatio-temporal data of Na

The regression model developed to predict TDS was reckoned as TDS = 6.4217Mg + 214.89 [8].The maximum correlation level was obtained at 91.5% of the data scattered getting distributed evenly on both sides of the model equation curve. The correlation co-efficient (R^2=0.9152) shows the reliability of the developed model.

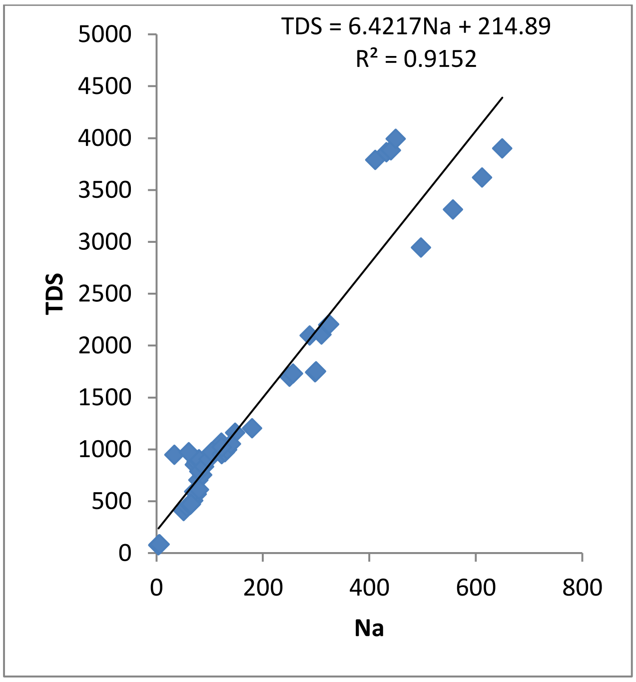

Figure 6: Regression data model for TDS using Na

Spatial distribution of Sodium (Na)

The spatial distribution of Na values of these samples is shown in the Figure.5. From the spatial distribution diagram, it is understood that the groundwater from the south part of the study area & some scattered parts of East, North, Central North & southern part has groundwater with Sodium content above the permissible limit during pre-monsoon & post monsoon season. The permissible limit values of Na were taken from IS 2269-1963. Table.3 shows the permissible limit values of Na.

Table 4: Range of Sodium

RANGE	LABEL
2-30	LOW
30-50	MEDIUM
50-100	HIGH
100-200	VERY HIGH

Development of data model for TDS prediction using spatio-temporal data of chloride (Cl)

The spatio-temporal data for chloride content was obtained for groundwater samples and the spatial distribution diagram was plotted for interpretations. Based on the obtained data, the chloride values were distributed as desirable, permissible and above permissible limits and illustrated. The regression model TDS= 3.6271Cl + 133.24 is used to envisage the TDS value based on Cl value. The maximum correlation level was obtained at 97.1% (R^2=0.9719).

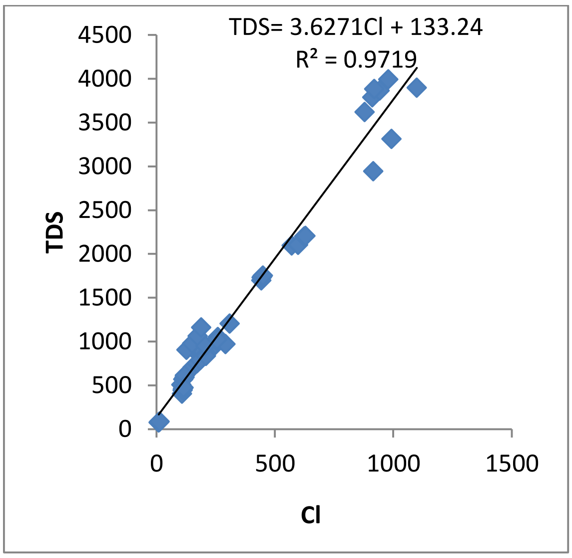

Figure 6: Regression data model for TDS using Cl

Spatial distribution of Cl

Spatial distribution diagram was plotted for the collected data. From the spatial distribution diagram, it was observed that the groundwater from the central part & southern part of the study area has groundwater with chloride contents within desirable to permissible limit and the rest part has chloride content below the desirable limit. The permissible limit values of Cl were taken from IS 2269-1963. Table.5 shows the permissible limit values of Cl.

Table 5: Range of Chloride

RANGE	LABEL
14-230	LOW
230-360	MEDIUM
360-590	HIGH
590-1100	VERY HIGH

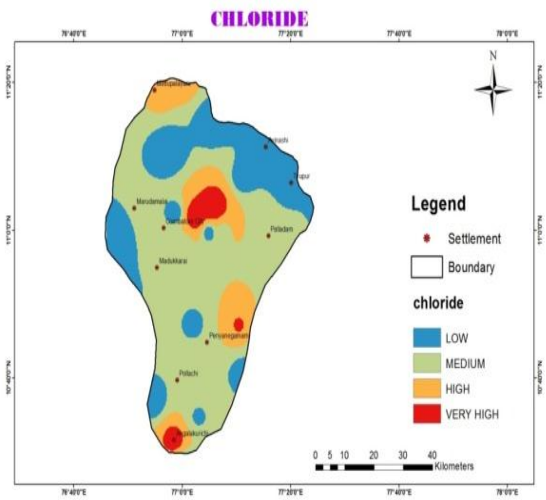

Figure 7: Spatial distribution map for Cl

CONCLUSION

Data mining is nove land eminent research area in agriculture. Agriculture is a water-based industry and the required crop yield for major crops can be obtained only adequate and quality supply of water is provided. In this paper linear regression, an effective data mining technique was used to predict water quality and thereby forecast the crop yield. The developed model analyses the ground water quality in terms of TDS and attempts to approximate future values of data as a linear combination of parameters. It is decisively concluded that this may help the decision maker to predict the crop yield with respect to water quality before harvesting the crop.

REFERENCES

[1] Niketa Gandhi, lesia Armstrong “Applying Data Mining Techniques to Predict Yield Of Rice in Humid Subtropical climatic zone of India” IEEE Paper 2016.

[2] Rajshekhar Borate, Rahul Ombale, Sagar Ahire, Manoj Dhawade, Mrs. Prof. R. P. Karande “Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India” IEEE paper Issue 1, April 2016.

[3] Aakunuri Manjula, Narsimha“XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction” IEEE 2015.

[4] Feng Gao, Martha C. Anderson, Donghui Xie “Spatial and temporal information fusion for Crop condition Monitoring” IEEE 2016.

[5] Shoba G, Dr. Shobha G.” Water Quality Prediction Using Data Mining techniques: A Survey” IJECS 2014.

[6] Darcy Miller, Jaki McCarthy, Audra Zakzeski “A Fresh Approach to Agricultural Statistics: Data Mining and Remote Sensing” JSM2009.

[7] Samson S. and Elangovan K “Hydro chemical analysis and estimation of ground water quality in namakkal District “, tamil nadu,,International journal of Ecotoxicology and environmental pollution 2011.

[8] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir,Kallal Das, Faridur Rahman, Rashedur M Rahman “Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh” IEEE paper Issue 1, June 2015.

[9] Mohammad Motiur Rahman, Naheena Haq and Rashedur M Rahman “Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield”, 17th. IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, 2014.

[10] Samson S. and Elangovan K”Assessment of ground water quality for drinking purposes during per monsoon Season in namakkal district tamil nadu,india.Pollution Research 2011.