MONITORING SYSTEM OF WATER AND ITS QUALITY USING GIS & DATA MINING TECHNIQUES

Swapnali Mahadik, [1]
Dhanraj Jadhav[2]

Abstract

Water is an important natural resource that essential to human survive and the environmental health. In India since the past few decades, the increasing activities of population and of industrial area have affects more to water quality. Generally Water quality is depending upon the change of climate and mainly chemical, physical and biological characteristics of water bodies. According to the Maharashtra Pollution Control board (MPCB) there are some predefined water quality criteria's and on the basis of that criteria , collection and integration of data from different resources through Remote Sensing helps to monitor water quality. For collecting and managing a Remote Sensed data there are different techniques are provided by Geographic Information System(GIS) with Data Mining Tools.

This paper examines how the different ways of spatial data utilization are works in combination of geospatial techniques with diverse data mining tools such as Cluster Analysis and Factor Analysis. This data mining tools with remote sensing application will help to enhance decision making of all over the area and environment.

Keywords: Remote Sensing, Spatial data, Data Mining Tools, Cluster Analysis(CA), Factor Analysis(FA), Geographic Information System(GIS).

Introduction

Data mining is a technique used to find new and hidden patterns of data. There are lots of methods in DM which can be use for different applications based on their categories. Such as cluster analysis and Factor analysis. These are the techniques which can be used to evaluate variations of water quality. Now a day's Water quality is a major problem around the world. The quality of water may be described in terms of concentration and some organic and inorganic material present in the water.

In the current research study data mining technique, cluster analysis (CA) and Factor Analysis (FA) is applied to a large environmental data set of chemical and micro-biological indicators of water quality. The main aim of this monitoring program is to collect data through various Remote sensing applications by means of sampling to assess spatial data in water. i.e. it is subjected to multivariate techniques to extract information

[1] Assistant Professor, YMT College of Management, Kharghar- Navi Mumbai
[2] Assistant Professor, Narsee Munje College, Vile Parle, Mumbai.

about similarities and dissimilarities between sampling locations, and influence of a parameters of water from different zone of water resources.

Requirements

Based on the data mining approach first of all following basic requirements can be taken into consideration for water monitoring.

1. Data Requirement

According to the system requirement there has to be a various types of data which can be converted into the information after processing such as environmental data, spatial data, remote sensing data, ground based information, satellite details(name, spectrum range, band no, resolution, data transmission rate)etc.[2]

2. Functional requirement

Monitoring system has to automate the process of remote based data and assessment should be done based on the integration of remote based data and ground based information. It should be the continuous analysis and visualization process.

With this requirements the sequential process has to be followed for monitoring water quality

Process Framework

Here remote sensing based framework used for predicting water quality of different source waters by integrating satellite data and ground-based information through satellites based on provided parameters of water. Figure shows schematic representation of systems framework with different components and their sequential process.

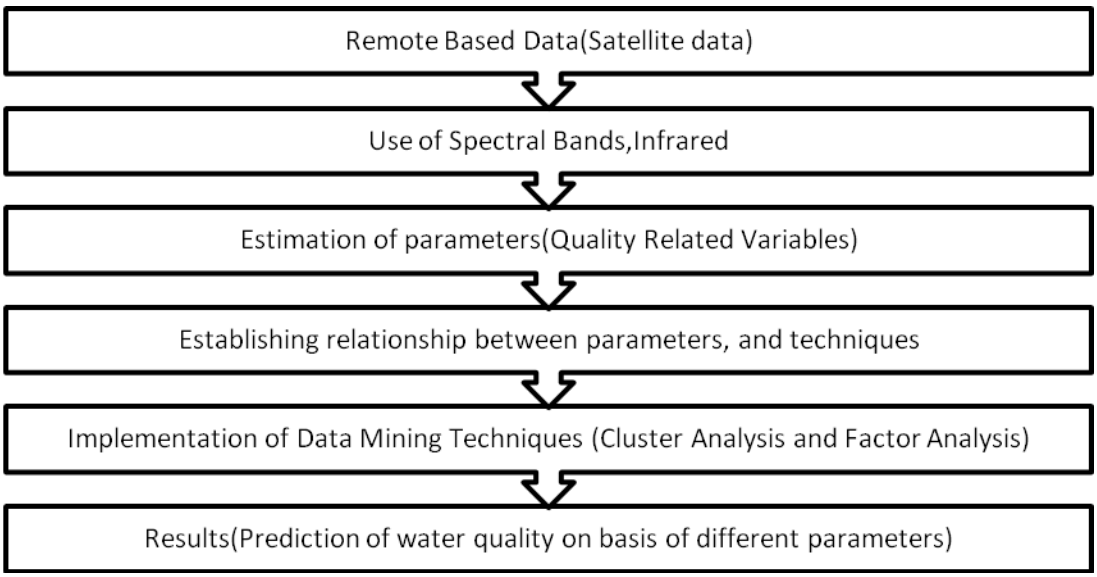

Fig 1. A remote sensing based process framework for monitoring water quality

Traditionally, water related data had been collected using water level stations and occasional measurements using flow meters. Typically system gathers data for years

without assessing each variable or parameter for any change like seasonal or random. this may affect the results. These parameters can include pH value, water temperature, air temperature, dissolved oxygen, turbidity, precipitation, water level, etc. But in the current era with the help of new satellites and sensors researchers can overcome with this problem.

Applications of remote sensing for data extraction

Remote sensing is the technique of making measurements of the earth using sensors on satellites. These sensors, the information gathering device ,register an electromagnetic energy and provides photographs, electrical signals and based on this it gathers data and provide specialized capabilities for manipulating, analyzing, and visualizing those data through which GIS allows to link databases and maps to integrate with the system.

Different Remote sensing methods may be applied to measure different components of the water. Various applications generally used combinations of this techniques. Remote sensing is the process of inferring surface parameters from measurements of the upwelling electromagnetic radiation from the land surface. This radiation is both reflected and emitted by the land. The former is usually the reflected solar while the latter is in both the thermal infrared (TIR) and microwave portions of the spectrum. There is also reflected microwave radiation as in imaging radars. The reflected solar is used in hydrology for snow mapping vegetation/land cover and water quality studies.[8]. This same technique with minor difference can continue with the water bodies such as main primary set of state variables which include water quality, water temperature, water equivalent ,surface temperature , moisture and many more different criteria's.

Water quality monitoring is the process to determine the chemical, physical and biological characteristics of water bodies and identifying the source of any possible pollution or contamination which might cause degradation of the water quality. These water quality indicators can be categorized as: (i) Biological: bacteria, algae, (ii) Physical: temperature, turbidity and clarity, color, salinity, suspended solids, dissolved solids, (iii) Chemical: pH, Dissolved oxygen, biological oxygen demand, nutrients (including nitrogen and phosphorus), organic and inorganic compounds (including toxicants) and (iv) Aesthetic: odors, taints, color, and floating matter.[1]

Such as suspended sediments are the most common pollutant both in weight and volume in surface waters of freshwater systems (Robinson, 1971; Lal, 1994). Suspended sediments increase the radian emergent from surface waters in the visible and near infrared proportion of the electromagnetic spectrum (Ritchie et al.,1976). In situ , controlled laboratory, aircraft and satellite measurements have shown that surface water radiance is affected by sediment type, texture, and color (Novo et al., 1989), sensor view, and sun angles (Ritchie et al., 1975), and water depth(Blanchard and Leamer, 1972).Spectral sensors on boat, aircraft, and satellite platforms have all been used to study suspended sediment patterns.

While extracting and monitoring all this parameters through RS, data consists of spatial, spectral and radiometric resolutions which provides atmospheric, radiometric and terrain corrections.

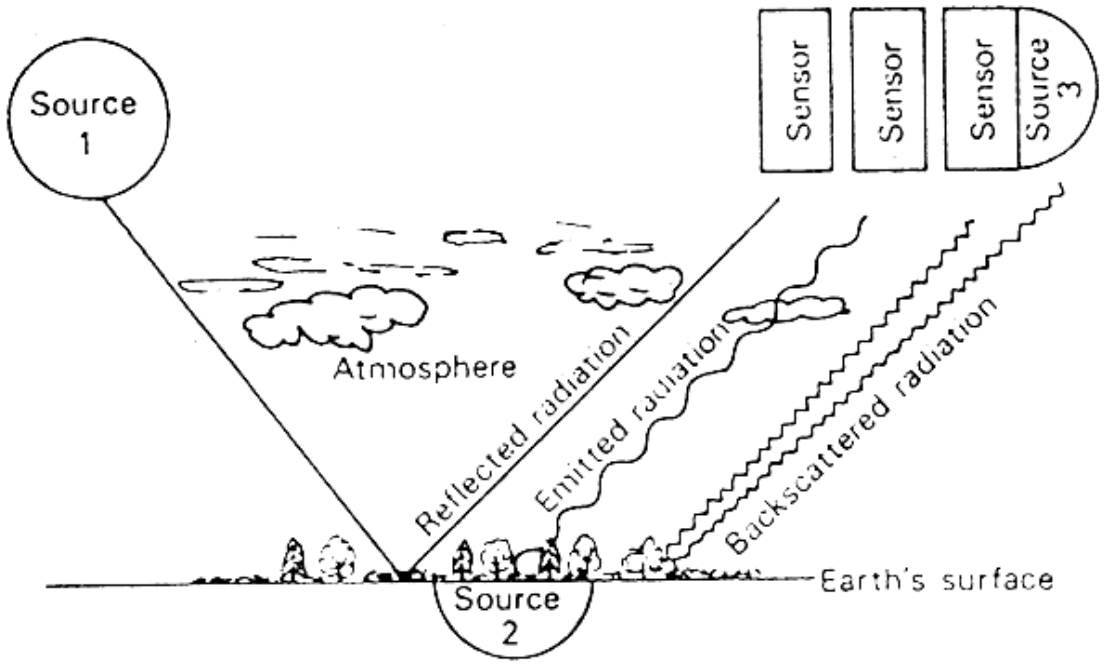

Fig2:The Key Features of the Remote Sensing Data Collection Process [10]

There are Several image processing techniques also which have been introduced in the extraction of water features from satellite data such as pixel classification , thresholding , spectral transformation PCA(Principal Compound Analysis), etc. With this Monitoring , various changes of water surface using different sensors used in various applications, such as disaster monitoring , forest and vegetation change, etc.

Next RS application, Simultaneous multi-spectral platforms such as Landsat (thematic mappers) can take images in multiple wavelengths of electro-magnetic radiation (multi-spectral) including Landsat program or the IKONOS satellite. This Landsat images are used by regulatory agencies such as KYDOW to indicate water quality parameters including Secchi depth(water clarity), chlorophyll a density and total phosphorus content.[9]

Like this data from several recently launched satellites sensors i.e. ,SeaWiFS, Modular Optical Scanner (MOS), Ocean Color and Temperature Scanner (OCTS), IKONOS are available which supports to extraction and integration of data for hydrological systems.

Statistical analysis based on the available data

Here theses two data mining algorithms are analyzed in prediction of water quality at different levels. The main factors of water quality monitoring are the collection of water samples, measurements at particular time and particular location with final outcome.

Cluster Analysis (CA)

Here CA can be used to detect similarities between different sampling sites for different seasons and it generates dendogram which indicates pollution status such as sulfate, chloride, Ammonium , Nitrates , Dissolved oxygen etc. as well as the effect of contamination of the samples such as grouping the sampling sources and the time on the basis of percentage of similarity and dissimilarity of different water quality parameters. A dendogram is a tree-structured graph used in maps to show the result of a hierarchical clustering.

Cluster Analysis (CA) provides groups of objects (cases) into classes (clusters) on the basis of similarities within a class and dissimilarities between different classes[3].Here the sampling strategy can be designed in such way to cover a broad range of determinants at key sites that accurately represents the water environment quality. This provides a way of segregation of data with cluster analysis. This results into interpreting data and indicate the spatial and temporal patterns. In this case of cluster analysis, the similarities-dissimilarities are quantified through Euclidean distance measurements, the distance between two objects, i and j, is given as;

$$d_{ij}^2 = \sum_{k=1}^{m} \left(Z_{ik} - Z_{jk} \right)^2$$

Where d^2_{ij} donates the Euclidean distance, Z_{ik} and Z_{jk} are the values of variable k for object i and j, respectively, and m is the number of variables. [4]

It also shows , CA organizes different entities into discrete group and dendogram shows grouping of clusters which varies with significant level.

Factor Analysis and Principal Component Analysis (PCA)

Factor analysis is working on the variables that are linked to find out specific factors that are associated with grouping and chemical characteristics of water sample. This technique is applied to reduce the dimensionality of a data set consisting of large number of interrelated variables, and this reduction is achieved by transforming the data set into a new set of variables-the principal components (PCs).The PCA is a data reduction technique and suggests how many varieties of water parameters are important to explain the observed variance in the data.[5]Also it gives the correlation between the observations in terms of the underlying factors, which are not directly observable in water sample. There are three stages in factor analysis for all the variables a correlation matrix

is generated, factors are extracted from the correlation matrix based on the correlation coefficients of the variables, to maximize the relationship between some of the factors and variables, the factors are rotated. [7]

Based on obtained information, it is possible to design an optimal sampling strategy, which could reduce the number of sampling stations, the frequency of sampling, the number of samples collected and associate costs and will also help to understand complex nature of water quality issues and determine priorities to improve water quality.[3]

Conclusion

This paper represents water quality predictions using data mining techniques at different geographic locations. In which cluster analysis technique provides facility to grouped different monitoring stations into few cluster of similar water quality characteristics for processing and the data reduction technique factor analysis. With this it focuses on various Remote sensing applications with integration of GIS to improve techniques of water analysis for monitoring water quality through different parameters. Also this research focus on understanding the effect of infrared and electromagnetic radiation on water surface so that new models can be developed related to water quality parameter measurements made by remote sensing and data mining tools and techniques.

Basically overall analysis demonstrates the use of statistical techniques to study the source of different parameters in water system resources to improve the quality which is helpful in planning the things like Supply of water for agriculture, Supply of drinking water as per the quality, Development of possible sources for recharging of groundwater, etc.

This research is mainly represented for the main task of study to find out how implementation of GIS could be integrated as useful, and even an essential tool to water monitoring system with Remote Sensing and data mining techniques.

References

[1] Use of Remote Sensing and GIS in Monitoring Water Quality Norsaliza Usali & Mohd Hasmadi Ismail (Corresponding author)

[2] Design of the water quality monitoring system for inland lakes based on remote sensing data.
zheng zhoua*, liangming liub, yuanling zhaoc

[3] Use of Cluster Analysis-A Data mining tool for improved water quality monitoring of River Satluj
Neetu Arora, Amarpreet Singh Arora, Siddhartha Sharma, Dr. Akepati S. Reddy
International Journal of Advanced Networking Applications (IJANA) ISSN No. : 0975-0290

[4] Investigation of Cluster Analysis in Surface Water in Yesilirmak River
Nurgul Ozbay, Suheyla Yerel, Huseyin Ankara
1st International S yposium on Sustainable Development, June 9-10 2009,

[5] Statistical Assessment of Water Quality Parameters for Pollution Source Identification in Sukhnag Stream: An Inflow Stream of Lake Wular (Ramsar Site), Kashmir Himalaya
Salim Aijaz Bhat,1 Gowhar Meraj,2 Sayar Yaseen,1 and Ashok K. Pandit1

[6] water quality prediction using data mining techniques: a survey
prof & hod, dept of computer science &engg., shoba g, dr. Shobha g.

[7] GUPTA AK, GUPTA SK and PATIL RS (2005) Statistical analyses of coastal water quality for a port and harbour region in India. Environ. Monit. Asses. 102 179-200.
[8] Remote sensing in hydrology
Thomas J. Schmugge a,*, William P. Kustas a, Jerry C. Ritchie a, Thomas J. Jackson a,Al Rango b
[9] Guo, Huadong; Huang, Qingni; Li, Xinwu; Sun, Zhongchang; Zhang, Ying (2013). "Spatiotemporal analysis of urban environment based on the vegetation–impervious surface–soil model" (Full text article available). Journal of Applied Remote Sensing 8: 084597. Bibcode:2014JARS....8.4597G. doi:10.1117/1.JRS.8.084597
[10] Geographical information systems and remote sensing in inland fisheries and Food And Agriculture Organization of united nation: corporate documentary Repository , Produced by: Fisheries and Aquaculture Department
[11] Sensor networks for monitoring and control of water distributed systems.
A. J. Whittle 1, m. Allen2, a. Preis3 and m. Iqbal3