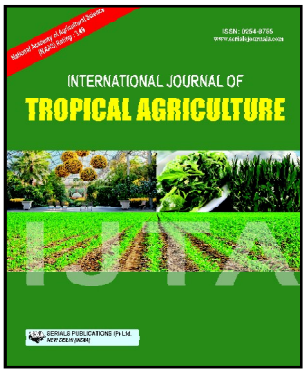

INTERNATIONAL JOURNAL OF TROPICAL AGRICULTURE

ISSN : 0254-8755

available at http: www. serialsjournals. com

Volume 36 • Number 4 • 2018

Spatial Database Management System for Standing Water Resources using Relational Properties

Manibhushan

Corresponding Author: Dr. Manibhushan, Sr.Scientist, ICAR Research Complex for Eastern Region, Patna

Abstract: This research aims to apply an integrated approach of geographic information system (GIS) and relational properties of database to develop a database information system for standing water bodies from extracting required data from digitized maps which is in the form of spatial data in eastern region of country viz. Bihar, Jharkhand and West Bengal. A number of spatial data bases and digitized maps are available related to water resources. But these spatial data do not easily define the available standing water bodies in Bihar, Jharkhand and West Bengal. So it is a challenge to extract required non spatial data in easily understandable form and recreate the spatial data as required by the user. An analysis has been done on available spatial data and maps to get non spatial/ tabular data which are in more easily understandable form. Data extracted district-wise related to area and number of standing water bodies according to their size of Bihar state. Various spatial data, maps, geodatabases have been collected from different organisations and institutes. After that an analysis using some data mining techniques has been done to get non spatial/ tabular data in easily understandable form and some required digitized maps have also been created using GIS technique as required. This type of information is more useful than the spatial data because a common person is able to understand these tabular data and they can use these data for their own purposes. These data can be utilised by a scientific personnel as well as farmers of these area and that will be used in agriculture for better utilization of water resources to enhance agricultural productivity and income of farmers of these states.

Keywords: Spatial data, tabular data, GIS, database, information system

INTRODUCTION

Data extraction sometimes termed as data mining and it is the process to extract important and useful information from a large sets of data (Abello *et al.* 2002). The goal of the data mining process is to extract information from an existing data set and

transform it into human understandable form for its better use. A spatial database is the database which has been specially optimized to store data pertaining to objects in the real world. In other words spatial data is the data which represents objects in geometric space. The objects are stored in database in the form of lines, points and polygons. A Relational Database Management System (RDBMS) with additional features can support spatial databases which are extensively used in environmental studies, Global Positioning System (GPS), and Geographic Information System. Spatial Data Mining (SDM) is a process of discovering trends or patterns from large spatial databases that hold geographical data. Objects in space such as lakes, ponds, rivers, forests, buildings, cities etc., are stored in spatial database. Spatial databases are so complex and make the SDM more difficult when compared with traditional databases. GIS, Microsoft SQL, Oracle, etc., are the used for building spatial and attribute/ tabular/ non spatial databases. Generally a GIS is used to store, retrieve and manipulate spatial data. Spatial data mining has many applications in extracting data in agriculture. Spatial data mining can be used in GIS pertaining to railways. Various techniques were studied that can be used in making a Railway Geographic Information System (RGIS) which can be used for spatial data presentation and statistical analysis besides helping in making well informed decisions (Xu *et al.* 2003). Some of the techniques useful for RGIS include spatial analysis, induction, classification and clustering, trend or spatial characteristic analysis, pattern recognition and digital map image analysis. Other approaches that can also be applied to spatial data mining are visualization approaches, rough set and fuzzy set approaches, genetic algorithms, and artificial neural networks. Clustering area geographical entities were studied by using clustering algorithms (Guang-xue *et al.* 2010). The algorithms work on the concept of geometric space similarity. Artificial neural network was applied to forest resource management in the aspect of insect pest prediction using various layers of geographical data for training models (Peng and Wen 1999). Moreover, the techniques for mining spatial data are to reveal exceptional phenomena implicitly existing inside a given set of geographical data, which are informative in making decisions. Crop yield prediction has important utility towards precision agriculture. Towards this many researchers proposed techniques. Many techniques came into existence using remote sensing data that are linked to crop yield prediction directly or indirectly (Prasad *et al.* 29). The techniques include Temperature Condition Index (TCI), Vegetation Condition Index (VCI) and Normalized Vegetation Index (NDVI). A data warehouse as a storehouse/ big database is a repository of data collected from multiple data sources and is intended to be used as a whole under the same unified schema. A data warehouse gives the option to analyze data from different sources under the same platform. Spatial databases are databases that, in addition to usual data, store geographical information like digitized maps, and global or regional positioning. Such spatial databases present a challenge to extract data using data mining techniques. In this research, a number of digitized maps related to water resources were collected from different sources. After that, data mining techniques were used to extract non spatial data in tabular form which are more understandable to a common person and farmers also. Some digitized maps/ spatial data have also been created as required for analysis purpose. Relational properties of database using SQL, ANN, GIS, etc. have been used for data mining and analysis of data. In the present research work and paper, only standing water bodies data in their area and number of all districts of Bihar have extracted from digitized maps using some data mining techniques. Data mining is the process of discovering previously unknown and potentially interesting patterns in large datasets (Piatetsky-Shapiro and Frawley 1991). The 'mined' information is typically represented as a model of the semantic structure of the dataset, where the model may be

used on new data for prediction or classification (Cunningham and Holmes 1999). Data mining, also termed as knowledge discovery, is the process of analyzing data from different perspective and summarizing it into valuable or non-trivial information. This information can be used for variety of purposes-research. The data can be analyzed from many different dimensions, categorized & summarized the relationships identified. Technically, data mining simply is the process of finding correlation or patterns among dozens of field in large RDBMS (Ansari 2010). Data mining is a high-level application technique used to present and analyze data for decision-makers. There is an enormous wealth of information embedded in huge databases belonging to enterprises and this has spurred tremendous interest in areas of knowledge discovery and data mining. Agricultural and biological research studies have used various techniques of data analysis including, natural trees, statistical machine learning and analysis methods (Cunningham and Holmes 1999). The aim of this research is to extract data in non spatial/ tabular form of all districts of Bihar using different data mining techniques and their applications to agricultural related areas. Spatial data/ maps have been used for this purpose. Some required digitized maps of available standing water bodies of some districts of Bihar have also been created using GIS. Relation properties of database, SQL, OLAP (On Line Analytical Processing), Artificial Neural Network (ANN), etc. have been used for data mining of district wise standing water bodies in their number and area of Bihar state. Spatial data of water resource have been used for extraction of tabular data of water resources as they are more understandable and useful in better utilization of water resources in agriculture.

MATERIALS AND METHODS

Various spatial data and digitized maps related to available water bodies of all districts of Bihar, Jharkhand and West Bengal have been collected from different organizations and institutes. These data were arranged in required format. After that data mining techniques have been applied to extract data from maps in tabular form which is more understandable. SQL, GIS software, ANN have been used for data mining from digitized maps. Data mining techniques also include Classification, Clustering, Association rules and Regression. The different data mining techniques used for solving different agricultural problem including water resource availability and utilization in agriculture. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. It is a process in which a model learns to predict a class label from a set of training data which can then be used to predict discrete class labels on new samples. To maximize the predictive accuracy obtained by the classification model when classifying examples in the test set unseen during training is one of the major goals of classification algorithm. Data mining classification algorithms can follow two different learning approaches: supervised learning, unsupervised. The different classification techniques for discovering knowledge are Rule Based Classifiers, Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbour(NN), Artificial Neural Network (ANN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms (Beniwal and Arora 2012). In clustering, data are kept in clusters such that the points within each cluster are close to one another. Clustering groups are made in such a manner that similar data are grouped together, while different data belong to different groups. Since the goal of clustering is to discover a new set of categories, the new groups are according to requirement (Rokach and Maimon 2005). There is no prior knowledge about data. The different clustering methods are Hierarchical Methods (HM), Partitioning Methods (PM), Density-based Methods (DBM), Model-based Clustering Methods(MBCM), Grid-based Methods and Soft-computing Methods such as fuzzy and neural network.

DATA EXTRACTION USING STRUCTURED QUERY LANGUAGE

In principle, data mining is not specific to one type of media or data. Data mining should be applicable to any kind of information repository. However, algorithms and approaches may differ when applied to different types of The most commonly used query language for relational database is structured query language (SQL) which allows retrieval and manipulation of the data stored in the tables, as well as the calculation of aggregate functions such as average, sum, min, max and count. For instance, an SQL query to select the farmers grouped by category (Land Holding group) would be: SELECT count (*) FROM Subsidies WHERE type=small farmer GROUP BY category. As described in this paper, an SQL query to select the number of water bodies having area greater than 10 hectare (ha) of Patna district may be SELECT count (*) FROM Patna WHERE area > '10 ha'. Patna is the name of table having standing water resource data of Patna district. In this way, required data can be retrieved. Data mining algorithms using relational databases can be more versatile than data mining algorithms specifically written for flat files, since they can take advantage of the SQL could provide, such as predicting, comparing, detecting deviations, etc. A relational database consists of a set of tables containing either values of entity attributes, or values of attributes from entity relationships. Tables have columns and rows, where columns represent attributes and rows represent tuples. A tuple in a relational table corresponds to either an object or a relationship between objects and is identified by a set of attribute values representing a unique key.

ON LINE ANALYTICAL PROCESSING

Online analytical processing (OLAP) is an approach is used for multi-dimensional analytical queries and extracting required data. Data mining is a part of OLAP with application such as forecasting or prediction in agriculture. It provides an opportunity of viewing agriculture data from different points of view to better understand what that data means OLAP has been used extensively for analysis of data of water resource as well as crop, soil, etc.. The recent advances in data base technology and data warehouses, the multi dimensional data base, OLAP, SOLAP (Spatial OLAP) and data mining technologies are being successfully applied to the management of natural resources. OLAP technique has been used to mine the available standing water resource data.

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network is an analytic technique formed on the basis of assumed learning process in the human brain. The same way the human brain after learning process is capable of deducing assumptions based on earlier observations, neural networks after learning process are capable to predict change and events in the system. Neural networks are a group of connected input/output units in which every connection has its weight. Learning process is conducted by balancing the network based on connections existing between elements in the examples. Based on importance of cause and effect between certain data stronger or weaker connections are formed between 'neurons'. Network formed in this manner is ready for work on unknown data and reacts based on previous knowledge. Conducted research (Stastny *et al.* 2011) tested the influence of multilayer neural networks on crop yield prediction. Backpropagation algorithm of ANN is used in this research work for data mining of standing water resource data of all districts of Bihar. ANN provides more accurate data than the other methods but data extracted by Relational technique using SQL is also good and relevant. SQL, ANN, GIS, etc. have been used for data mining in this research work and paper. These tabular data may be used for better utilization water resources in agriculture to increase crop productivity in Bihar because the tabular data is more understandable and easy to use.

RESULT AND DISCUSSION

Digital maps/ spatial data were collected from different organisations/ institutes and some water resource maps have also been created according to the requirement. After that non spatial/ tabular data of standing water bodies in their number and area have been extracted from spatial data of all districts of Bihar, Jharkhand and West Bengal. For example, I have shown the spatial and non spatial/ tabular (extracted) data of three districts viz. Patna, Ranchi and Maldah districts of Bihar, Jharkhand and West Bengal respectively (Figure.1-3) and (Table.1-3) but tabular data have been extracted for all districts of Bihar.

DISTRICT WISE AVAILABILITY AND AREA OF STANDING WATER BODIES

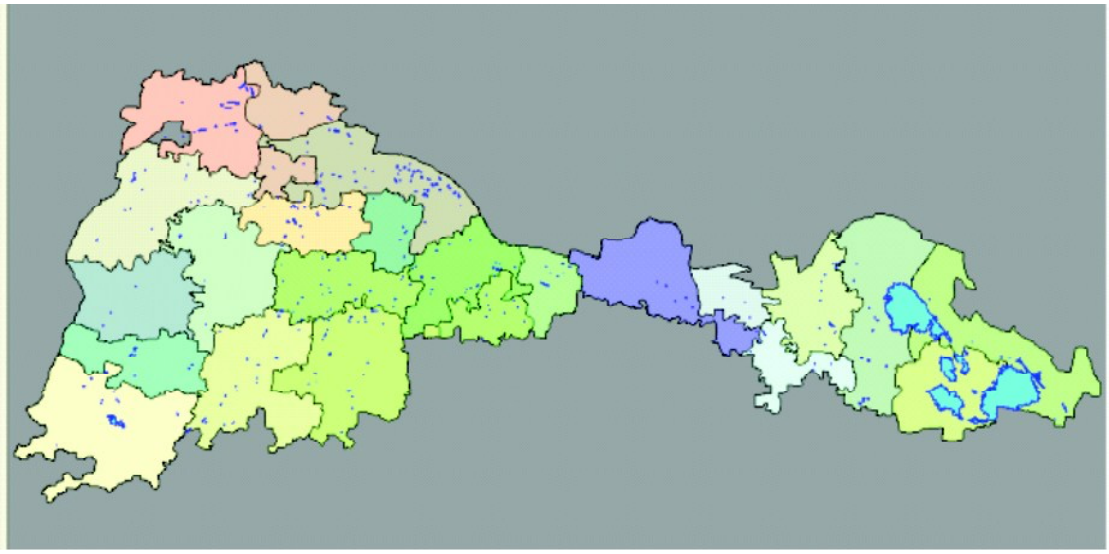

Figure 1: Status of standing water bodies in different blocks of Patna district of Bihar

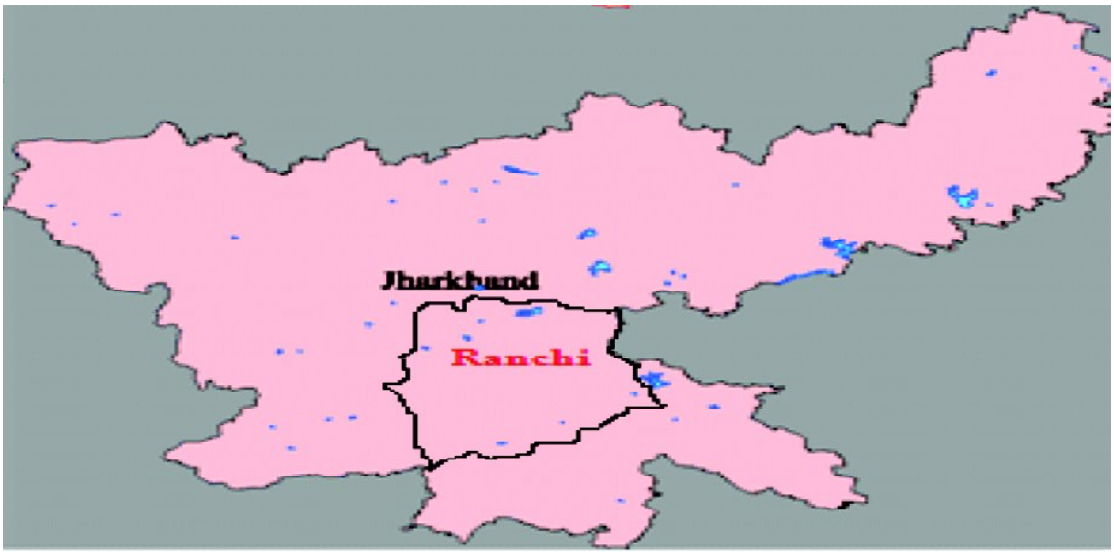

Figure 2: Status of standing water bodies in different blocks of Ranchi district of Jharkhand

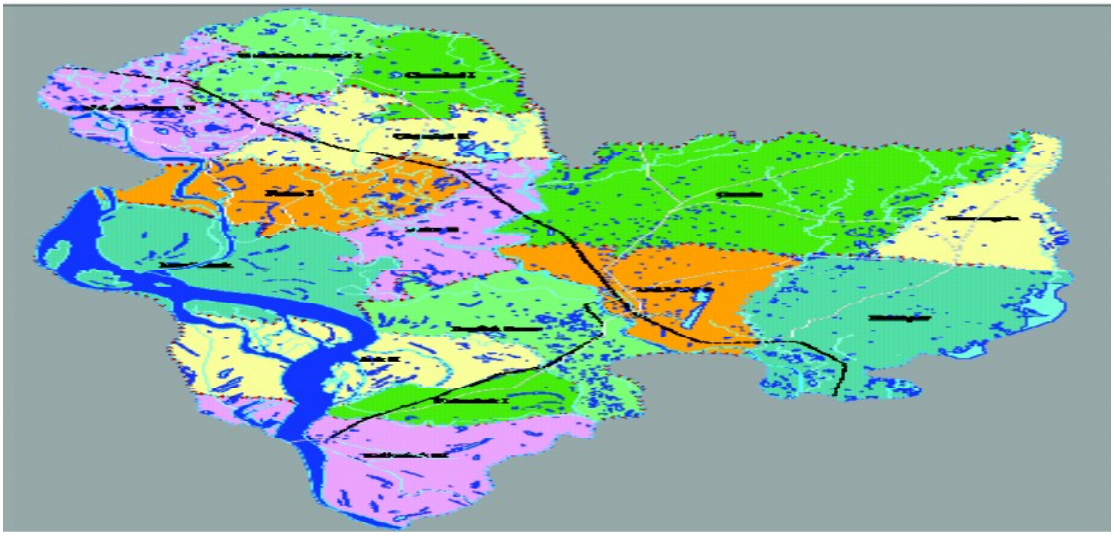

Figure 3: Status of standing water bodies in different blocks of Maldah district of West Bengal

Table 1
Avaialbility of standing water bodies in number and area according to the size of water bodies of Patna district of Bihar

Name of District	*Category of water body (Area-wise) (ha)*	*Number of water bodies*	*Total Area of water bodies (ha)*
Patna	0.5-10	335	520.18
	10-50	09	136.08
	50-100	0	0
	100-500	03	681.78
	500-1000	01	936.03
	More than 1000	02	6155.61
	Total no. of water bodies	350	8429.68

Table 2
Avaialbility of standing water bodies in number and area according to the size of water bodies of Ranchi district of Jharkhand

Name of District	*Category of water body (Area-wise) (ha)*	*Number of water bodies*	*Total Area of water bodies (ha)*	*Number of Perennial water bodies*
Ranchi	> 0.5 and < 2	893	796.58	64
	> 2 and < 5	77	224.13	06
	> 5 and < 10	21	139.01	01
	> 10 and < 50	15	279.76	01
	50 and above	04	3024.83	0
	Total number of water bodies	1010	4463.83	72

Table 3
Avaialbility of standing water bodies in number and area according to the size of water bodies of Maldah district of West Bengal

Name of District	*Category of water body (Area-wise) (ha)*	*Number of water bodies*	*Total Area of water bodies (ha)*
Maldah	> 0.5 and < 10	2983	5651.62
	> 10 and < 50	253	5108.45
	> 50 and < 500	55	5479.95
	> 500 and < 1000	01	776.56
	1000 and above	02	5478.76
	Total	3294	22495.33

From the above data, it is clear that among three districts, the maximum number of water bodies are available in Maldah district there after Ranchi and the least number of water bodies are in Patna district among mentioned three districts of three states. Maximum area of water bodies are present in Maldah district, there after Patna and Ranchi district has the least area of standing water bodies among three above mentioned districts (Table 4 & Figure 4). In Patna district, Mokama tal having less number of water bodies but the area of water bodies is very large because after monsoon season, water becomes stagnant for few months in mokama tal area because these area is low land area. After 3 to 4 months, pulse crops such as lentil, chickpea, pea and tisi, etc. have been cultivated. Many standing water bodies in West Bengal may be used for fish farming because 3-4 months, wet land remains idle. So there may be land use planning for this tal area such fish production, chest nut (Singhara) cultivation, etc. can be done in water logged period that will give farmers more income. So that the livelihood of farmers may also become better. So these spatial and tabular data especially tabular data which is more understandable to farmers and a common person can be used for increasing the agricultural productivity and ultimately the income of farming community using these available excess standing water resources in these areas. So in this way, standing water resource data may be used to increase crop productivity for all

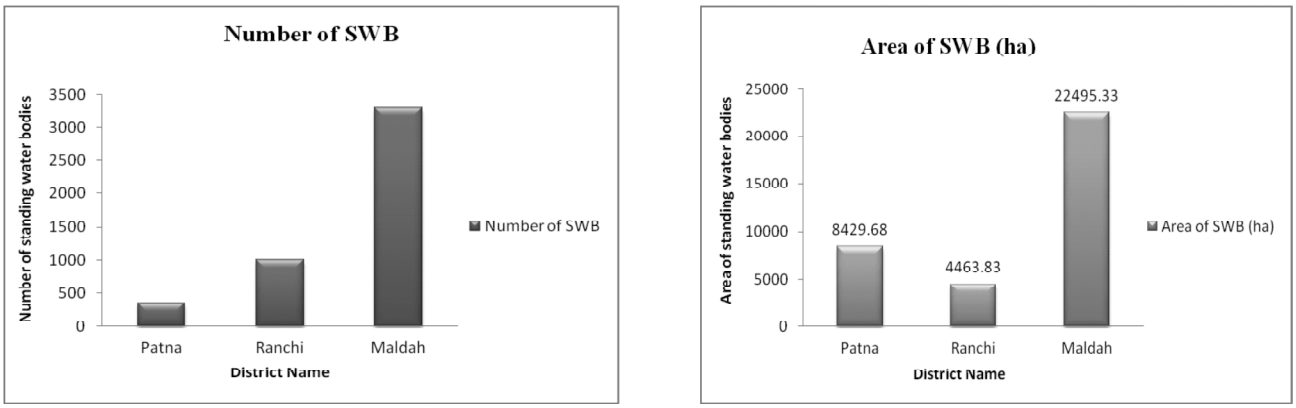

Figure 4: Number and area of standing water bodies of Patna, Ranchi and Maldah districts of Bihar, Jharkhand and West Bengal states

districts of Bihar, Jharkhand and West Bengal. In this way these tabular data will be very useful for farmers to do better utilization of standing water bodies for increasing the crop productivity.

Table 4
Number and area of standing water bodies Patna, Ranchi and Maldah districts of Bihar, Jharkhand and West Bengal

District Name	*Number of SWB*	*Area of SWB (ha)*
Patna	350	8429.68
Ranchi	1010	4463.83
Maldah	3294	22495.33

SWB: Standing Water Body

CONCLUSION

There are many applications of data mining techniques in agriculture on amount of data that are currently available from many resources in spatial and non spatial form. Agricultural organizations and their management try every day to find information in large databases in decision making to the best utilization of natural resources for increasing the crop productivity. Data mining, through better management and data analysis, can assist agricultural organizations to achieve higher crop productivity. However, available raw agricultural data are widely distributed at different places. These data should be collected and stored in organized forms and integrated to form an information system in agricultural organization. Data mining technology provides user oriented access to new and hidden patterns in data from which knowledge is generated which can help with decision making in agricultural organizations for better planning of agricultural activities and use of soil, water and other natural resources. So in this research paper, data mining techniques have been applied to extract useful non spatial data in tabular form from spatial data/ maps. Relational properties of database using SQL statement, artificial neural network, GIS, etc. have been applied in data mining for getting useful district wise data of standing water resources in their area and number of Bihar state that can be used for better utilization of water resources in agriculture to increase crop productivity and finally the income of farmers of Bihar, Jharkhand and West Bengal.

REFERENCES

Abello J, Pardalos PM, Resende M. (2002). Handbook of massive data sets. Kluwer, New York.

Beniwal, S. and Arora, J. (2012). Classification and feature selection techniques in data mining. International Journal of Engineering Research & Technology (IJERT), 1(6).

Lior Rokach, Oded Maimon. (2005). The Data Mining Knowledge Discovery Handbook. Clustering Methods, Chapter-15, pp 321-352.

Stastny J, V Konecny and O Trenz. (2011. Agricultural prediction by means of neural network. *Agric. Econ.* – Czech. 57(7): 356–361.

Ansari A and Ansari S. (2010). The concept of data mining, its applications & Issues. pp.67-78.

Wei XU, Yong QIN, Houkuan Huang. (2003). The Application of Spatial Data Mining in Railway Geographic Information Systems. *IEEE*.0 (0), pp1467-1471.

CHEN Guang-xue, LI Xiao-zhou, CHEN Qi-feng and LI Xiao-zhou. (2010). Clustering Algorithms for Area Geographical Entities in Spatial Data Mining. *IEEE*.0 (0), pp1630-1633.

Prasad Anup K, Chai Lim, Singh Ramesh P and Kafatos Menas. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. International Journal of Applied earth observations and Geo information, Volume 8, Issue 1, pp 26-33.

Peng C.H. and Wen, X. (1999). Recent applications of artificial neural networks in forest resource management: an overview. In: U. Corté and M.Sànche-Marrè (Eds.), *Environmental Decision Support Systems and Artificial Intelligence.* (pp. 15-22).

AAAI Technical Reports WS-99-07, AAAI Press, Menlo Park, CA, USA.

Piatetsky-Shapiro, G and Frawley W.J. (1991). *Knowledge Discovery in Databases*. Menlo Park, CA, AAAI Press.

Cunningham S. J and Holmes G. (1999). Developing innovative applications in agriculture using data mining. In the *Proceedings of the Southeast Asia Regional Computer Confederation Conference*.