Geospatial Data Mining to Explore Watershed Development in Rainfed Regions

S. Nallan[1], L. Armstrong[1], Andreas Neuhaus[1,2], B. Croke[3], N. Dunstan[1,4]

[1] School of Computer and Security Science, Edith Cowan University, Perth, Western Australia, Australia

[2]CSBP Limited, Kwinana, Perth, Western Australia, Australia

[3] ANU College of Physical and Mathematical Sciences and ANU College of Medicine, Biology and Environment, Australian National University, Canberra ACT, Australia

[4]School of Science and Technology, University of New England, Armidale, New South Wales, Australia

s.nallan@our.ecu.edu.au

ABSTRACT

Advances in the Information and Communication Technology (ICT) and availability of fine resolution remote sensing spatial data provides an opportunity to identify previously unknown and potentially useful patterns from the huge datasets. Application of geospatial data mining techniques has been used in various agricultural contexts such as drought management, vegetable quality classification, pest management and several other agriculture related areas.

The recent development and implementation of watershed programs in rainfed regions of India has highlighted strategies that can be used to conserve water for irrigation of crops. Due to aberrant rainfall conditions in the rainfed regions and enormous changes in the landuse, the impact of the watershed development has greatly influenced the local hydrology. Any understanding of the impact of watershed development cannot be made through an examination of a individual watershed, but through the examination of all surrounding watersheds. This can be achieved using geospatial datasets and available novel data mining algorithms. For example, these techniques could provide an assessment of the effectiveness of watershed development using different land use patterns, cropping intensity, water availability, aquifer re-charge and positioning of different watersheds structures in a catchment. This paper attempts to explore the application of geospatial data mining techniques to watershed data sets. An evaluation and quantification of hydrological impacts of watershed development under varying climate and management scenarios using advanced techniques of geo-spatial data mining could be one means to improve the understanding of these impacts. This paper reports on the development of a robust data matrix of various parameters that affect the watershed and the application of various data mining algorithms. It is concluded that there is no single technique that can be used to assess the impact of watershed development.

Sreedhar Nallan. C0337 Geospatial data mining to explore watershed development in rainfed regions. EFITA-WCCA-CIGR Conference “Sustainable Agriculture through ICT Innovation”, Turin, Italy, 24-27 June 2013.

Key words: Geospatial, data mining, watershed, hydrology, India

1. INTRODUCTION

Utilization of cutting edge Information and Communication Technology (ICT) advancements such as availability of high resolution remote sensing data, Geographical Information System (GIS) techniques and data mining algorithms helps researchers to identify previously unknown and potentially useful patterns from the huge datasets. This paper reviews research on application of geospatial data mining techniques to different agricultural datasets and how this can be used to describe watershed development impact assessment.

1.1 Watershed Development

In recent years, India has look to watershed development as a way to realize its hopes for agricultural development in rainfed and semi-arid areas (Fig. 1). This intense development of watershed has lead to a number of impact studies by Government and Non-Government research organizations to understand the role inhydrological, social and economic processes.

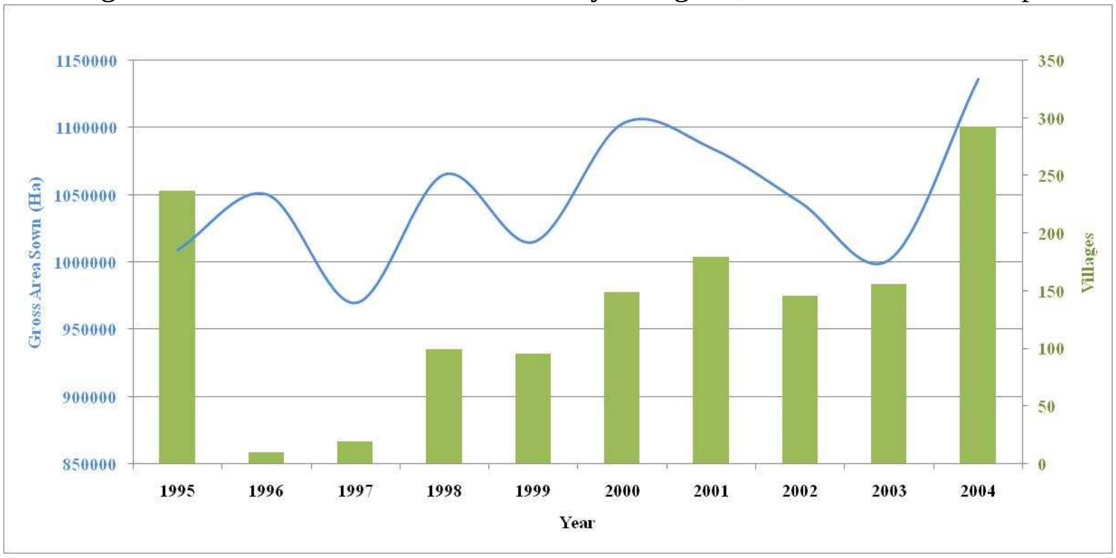

Fig. 1 Gross Area Sown (Ha) and Villages with Watershed Development* in Anantapur District, Andhra Pradesh, India during 1995 to 2004

(Data Source: http://www.watershed.nic.in/ProgressOfSubActivityStatewise.asp and Andhra Pradesh Bureau of Economics and Statistics, Hyderabad)

(* Villages treated under different watershed schemes including completed and ongoing)

Effective water management is dependent on a number of factors such as demography, climate, soils, landuse and topography and these are affected by changing climatic conditions. Rainfed areas in particular, frequently experience water deficits due to irregularity in the rainfall pattern. For example, changes in rainfall intensity and volume in rain fed areas can influence the water table as well its agriculture. The rainfed regions in India often experiences either drought situations or erratic rainfall conditions (Wani, Venkateswarlu et al. 2009) (Fig. 2). Though the annual rainfall may be average, the actual occurrence of the rainfall during the agricultural season is changing. The changes in the intensity of rainfall and prolonged hydrological drought during monsoon season could be some of the reasons which can be attributed to climate change effects, either no runoff or excess runoff (Simvonic 2010). A greater understanding of the dry and wet spells during the monsoon will help in efficient agricultural operations (Singh and Ranade 2009). Watershed programs in the rainfed regions of India has developed many water harvesting structures which are exclusively designed to store the rainwater to raise the ground water table (Garg, Karlberg et al. 2011). It has been highlighted that the criteria used for development of these watersheds should lead to the improvement the agricultural and water productivity (Bhalla, Pelkey et al. 2011).

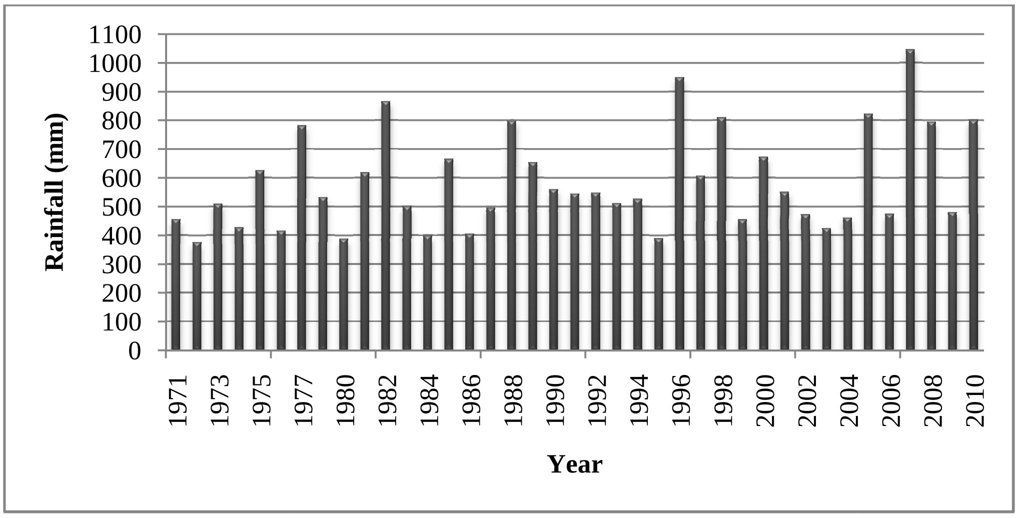

Fig. 1 Annual Rainfall (mm) at Gooty Rain gauge Station, Anantapur District, Andhra Pradesh, India (Data Source: Andhra Pradesh Bureau of Economics and Statistics, Hyderabad, 2011)

Several water harvesting structures has been constructed in the watershed areas in the form of checkdams, percolation tanks, rock fill dams to stop the flow of water and to increase infiltration in that area. However, the investments made by the Government of India watershed programs to construct several water harvesting structures, has failed to give optimum results due to improper location of these structures (Action for Social Advancement 2008). Impact assessments, based on several socio-economic surveys have not examined the impact of water harvesting structures and the hydrological changes after the watershed development. This is crucial and challenging problem which needs further investigation. The hydrological impact cannot be assessed locally as it could be due to change in the rainfall pattern and may not be necessarily from watershed development (Reddy and Soussan 2004). Geospatial analysis utilizing data such as daily rainfall, ground water levels, soil properties and watershed development at different time scales could help in understating climate change impact and hydrological changes in the watershed before and after the development of water harvesting structures.

1.2 Geospatial Data Mining

Data mining is a new inter-disciplinary concept involving data analysis and knowledge discovery from the databases (Rok, Matjaz et al. 2007). It is considered as a multi-facet approach which includes statistical analysis, data visualization, neural networks, knowledge discovery, pattern recognition and data base management (Feelders, Daniels et al. 2000).

Geospatial data mining is an emerging active research field which uses the massive spatial databases for knowledge discovery and extraction of unknown and useful patterns. Spatial data mining involves various tasks such as spatial classification, spatial clustering, spatial association and spatial trend analysis. Spatial data mining efficiently handles huge datasets for process based computational algorithms. This makes an effective visualization to explore complex patterns (Guo and Mennis 2009). The data mining techniques have be applied to many areas in the agriculture such as soil classification, water quality monitoring, weed detection and atmospheric pollution forecast (Mucherino, Papajorgji et al. 2009). Data mining techniques are often used to study soil characteristics (Eklund 1998).

2. METHODOLOGY

The present study investigated spatial association data mining algorithm on watershed data to understand the spatial and temporal association of watershed development. The methodology framework is described in the Fig. 3. Data collected from different parameters with spatial and temporal attributes has been used for the analysis and this has been tabulated into a mega dataset with all available parameters contributing to watershed development and agriculture development. Spatial data has been collected from the freely available MODIS data. Other data on land use / land cover mapping, soil properties data, slope of the stream, height and width of the checkdam, agriculture bore wells in the vicinity of the water harvesting structure, rainfall data, ground water levels has been collected and tabulated.

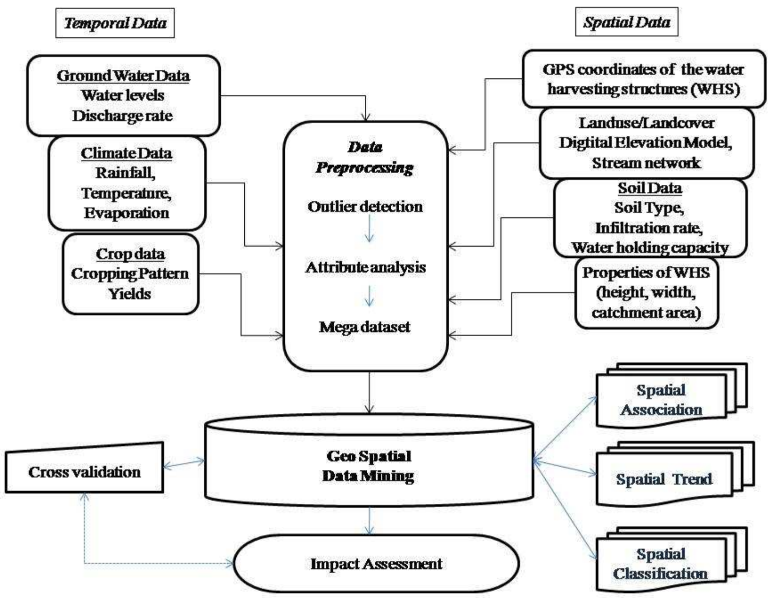

Fig. 3 Geospatial Data Mining approach framework

After analysing the spatial characteristics of watersheds, the mega watersheds has been further divided into small watersheds with various categories. The key variables identified for this analysis are tabulated in Table 1. Water harvesting techniques are based on the local needs of the agriculture. In some cases, checkdams are constructed to divert the stream flow into a natural farm pond and in other cases its used for ground water recharge. This study explored the spatial association of checkdams and its impact in the surrounding hydrology. Data on Geomorphology, Geology and Lithology along with land use and soil properties helps in understanding the location specific hydrological characters of the watershed.

Table 1. Identified spatial and temporal variables for watershed data mining

Spatial	**Temporal**
Stream network	Rainfall
Water harvesting structures	Landuse changes
Natural recharge structures (farm pond, tank)	Ground water levels
Slope	Bore well Pump Yield
Geomorphology	Demography
Geology	
Slope	
Soil	

The association of each variable with the watershed is made with different combinations using association algorithms. For example, :IS_A(X,”CHECKDAM”) ^ CLOSE_TO(X, “TANK”) = RECHARGE_INDEX(X, “HIGH”) [0,5%, 80%] (80% of tanks recharge based on close proximity of check dam in 0,5% of dataset). In this case, the water harvesting structure is very close to a natural farm pond (tank). The association rule mining can be used to identify the location specific issues in the water harvesting.

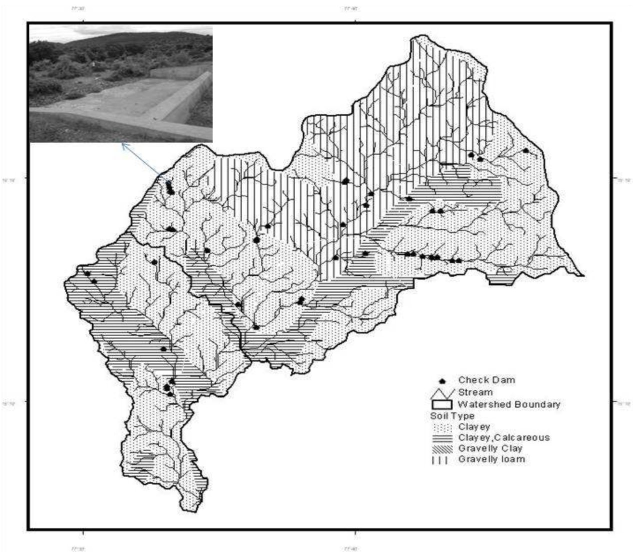

Fig. 3 Check dams in a watershed having different soil types (Inset: Check dam structure)

3. SUMMARY

Application of novel algorithms enhances the quality of the output as compared to existing indigenous techniques. Geospatial data mining in the field of agriculture related domains and in particular to watershed impact assessment is a current challenge. The data mining requires crucial data to understand the impacts at a micro scale. Geospatial data mining algorithms enhances the understanding of spread, location and its impact on the local hydrology-within watershed and within the contiguous watersheds. The effect of climate change at micro level watershed scale is difficult to understand but the impact can be experienced indirectly.

The rainfall data at an hourly rate and the infiltration rate at different check dam structures and the levels of ground water at a less interval during rainy season can provide greater accuracy. Unavailability of year wise water harvesting structures data with its location specific coordinates, also may hinder this type of spatial analysis. The techniques applied to the real-time data will enhance the skills of decision making in a better way for the planners by simulating with the location specific conditions with a simple decision support system. However, with the revolution in the software and hardware availability and the ease of optimal storages can give more options for collection of micro level data sets for a crucial study using geospatial data mining techniques.

4. REFERENCES

Action for Social Advancement (2008). Report on Study on Workings of Check Dams in Madhya Pradesh. Action for Social Advancement, State Planning Commission , Madhya Pradesh, India.

Bhalla, R., et al. (2011). "Application of GIS for evaluation and design of watershed guidelines." Water Resources Management **25**(1): 113-140.

Eklund, P. W. (1998). "Data mining and soil salinity analysis." International Journal of Geographical Information Science **12**(3): 247-268.

Feelders, A., et al. (2000). "Methodological and practical aspects of data mining." Information and Management **37**: 11.

Garg, K. K., et al. (2011). "Assessing impacts of agricultural water interventions in the Kothapally watershed, Southern India." Hydrological Processes **25**.

Guo, D. and J. Mennis (2009). "Spatial data mining and geographic knowledge discovery - An introduction." Computers, Environment and Urban Systems **33**: 6.

Mucherino, A., et al. (2009). "A survey of data mining techniques applied to agriculture." Operational Research International Journal **9**: 20.

Reddy, R. V. and J. Soussan (2004). "Assessing the Impacts of Watershed Development Programmes: A Sustainable Rural Livelihoods Framework." Indian Journal of Agricultural Economics **59**(3): 331.

Rok, R., et al. (2007). "Integrating data mining and decision support through data mining based decision support system." The Journal of Computer Information Systems **47**(3): 89.

Simvonic, S. P. (2010). "A new methodology for the assessment of climate change impacts on a watershed scale." Current Science **98**(8): 9.

Singh, N. and A. A. Ranade (2009). Climatic and Hydroclimatic Features of Wet and Dry Spells and thier Extremes across India, Research Report No. RR 122. Indian Institute of Tropical Meteorology, Pune, India.

Wani, S. P., et al., Eds. (2009). Best-bet Options for Integrated Watershed Management. Comprehensive Assessment of Watershed Programs in India (SBN 978-92-9066-526-7). ICRISAT, Patancheru 502 324, Andhra Pradesh, India, International Crops Research Institute for the Semi-Arid Tropics. pp. 312.