International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248
Volume: 3 Issue: 10 230 – 233

Terrain Mapping for South Gujarat : A GIS based Solution for Geo-Community

Ms. Jaishree Tailor
Shrimad Rajchandra Institute of Management and Computer Application
Uka Tarsadia University
Bardoli, Surat, India
e-mail: jaishree.tailor@utu.ac.in

Dr. Kalpesh Lad
Shrimad Rajchandra Institute of Management and Computer Application
Uka Tarsadia University
Bardoli, Surat, India
e-mail: kalpesh.lad@utu.ac.in

Abstract— Geographical Information Systems (GIS) are crucial to every field especially for natural resource, land assessment and management. Geo-scientists and geo-community are always concerned about terrain related issues and challenges and continuously put efforts to provide geo-spatial solutions to agriculture and terrain conditions for improving the growth of agricultural yield and resolving the issues of food, fodder and feed.. The authors propose a GIS as a solution to predict geo-spatial variability of terrain attributes for the region of south Gujarat. Thus, this paper discusses the components of the GIS designed and developed along with its data processing issues.

***Keywords**- GIS;Geo-Spatial, Terrain, Soil, Prediction, Gujarat*

*****

I. Introduction

A geographical information system (GIS) is "A system for capturing, storing, checking, integrating, manipulating, analysing and displaying data which are spatially referenced to the Earth". Remote Sensing (RS) & Global Positioning Systems (GPS) facilitates GIS to integrate geo-spatial data with actual variables of interest under study. GISs have been built-up with extensive capabilities that collaborates geo-spatial and statistical measures to symbolize the globe. The ultimate goal of GIS is to cultivate a sense of awareness amongst the communal upon the essentials to conserve our environment [1, 2].

India needs extensive participation and association of industries and academic institution to work and provide in-house geo-spatial solutions because major GIS solutions of India are not native. Thus organizations like NASSCOM, ICAR, ISRO, NIC, AGI etc. have clearly revealed the promotion of home-grown software development. Further it is an irony that about 90% of the software that the country develops targets for the problems outside the country [3]. Hence there is a burning requirement to design and execute bigger roles and responsibilities in terms of propagation, training and uncovering the benefits of geo-spatial technologies [4]. Therefore it is determined that prevalence of G-Revolution in India and that too in Indian Agriculture is still evolving. The initiatives that are taken are not uniformly disseminated between regions due to differentiated nature of demand of the farmers in different areas.

The geo-spatial initiatives are intended to address the needs of the farmers, but they are not being able to take the ultimate benefit meaningfully to meet their needs. Moreover the benefits of such technology is not very popular and regular in terms of availability and usage. And so these efforts have not been able to bring significant impact on the farm productivity and their socio-economic development [5], while designing such native solutions especially with GIS several challenges are faced concerning data and it's pre-processing. This paper therefore reveals the data pre-processing challenges while modelling geo-spatial variability for Southern Gujarat terrain attributes as a part of their research [6].

The proposed GIS that authors have designed will directly benefit the peasant, agronomist, agricultural officers (kheti niyamak), soil testing laboratories and geo-spatial data analyst also. Therefore, the authors have proposed Web-based GIS Model to Predict Geo-Spatial Variability of Southern Gujarat Terrain Attributes to aid the environmental stakeholders to visualize geo-spatial variability of this region and enhance decision-making.

II. Challenges In GIS Terrain Modelling

Terrain input data that enter into the system differ in form and format e.g. soil spreadsheets, and soil web portals etc. This results in to difficulties in data extraction and transform [7]. During calibration of prediction model the authors have come across following challenges:

- Missing values of nutrients for some villages
- Multiple naming conventions for same villages obtained from different sources
- Transliteration issues because some datasets were in Gujarati while others in English
- Erratic values of soil nutrients for some villages (e.g. outlier values)
- Missing taluka wise classification of villages
- No spatial essence .i.e. no coordinates

Realizing the needs of usable format of data so as to conveniently carry out spatial variability of the soil nutrients for the region of interest data pre-processing module was implemented by the authors as a part of GIS for Southern Gujarat Terrain Mapping.

III. Proposed GIS For Terrain Mapping

This segment of the paper describes the components of GIS proposed by the authors that has four major elements namely data source, data pre-processing, predictive modelling and cartographic visualization. The authors have summarized the components in brief and presented the architecture in Figure 1.

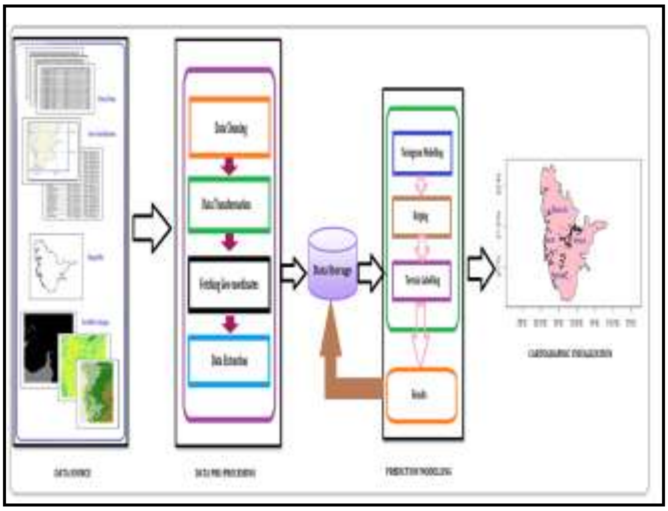

Figure 1. Architecture for Terrain Mapping GIS

A. Data Source

The inputs to the proposed GIS for Terrain Mapping consists of several forms. The purpose of this Data Source component of this GIS is to access and gather data from varying sources and in varying formats. For terrain mapping as well as for terrain attribute prediction the basic inputs are point data that consist of attribute and non-geo-spatial information like taluka, village, and soil nutrients tested values etc. The attribute data is made geo-spatial by attaching geo-coordinates to each village, therefore, geo-coordinates are the second input as Longitude and Latitude.

Further for representing an area using a map, relevant administrative boundary of a region is required, to fulfil this Shape files a globally accepted vector file formats have been used. Considering impacts of other topographic features like Elevation, NDVI, TPI, TRI etc., on terrain mapping, provision for harnessing remote sensing and satellite images is also included in this module. These input data is then pre-processed by the Data Pre-processing module of the GIS the representation of which is shown in Figure 2 and Figure 3.

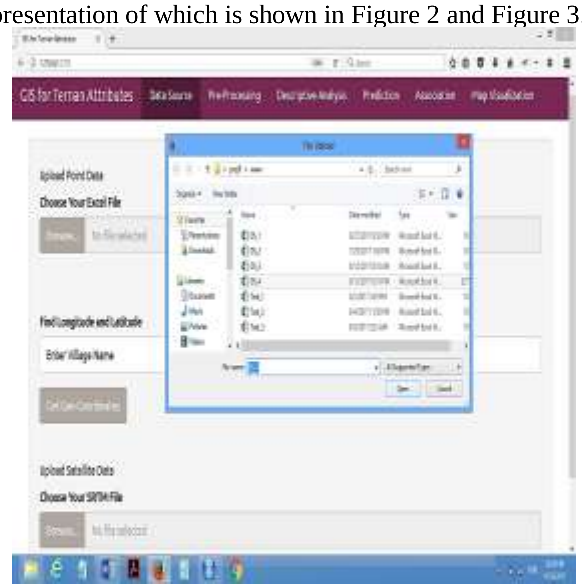

Figure 2. Screen Layout to upload Data

B. Data Pre-processing

Before data is used for predictions it necessary to clean data. Original data that are downloaded from soil portals or other data sources like sugar factories have unnecessary information that remains unused and consumes unnecessary memory and increase processing time. So only those information in the form of fields that are required are retrieved using sub-setting. The fields that are discarded are Survey number, farmers name, area, sample number, laboratory numbers, and field surface structure etc. While the fields that participate for further processing are District, Taluka, Village, pH, EC and OC.

Further it also deals with missing and NA values of pH, EC and OC for a village for taluka. For those villages without any measurements or missing values or NA, average of entire taluka has been calculated and assigned as the terrain value to such a village. Outliers have been identified while implementing QQ-Plot and Histogram and thus have been discarded from the data. The raw data with inconsistencies in the form of redundancies, NA values and outliers are depicted in Figure 3.

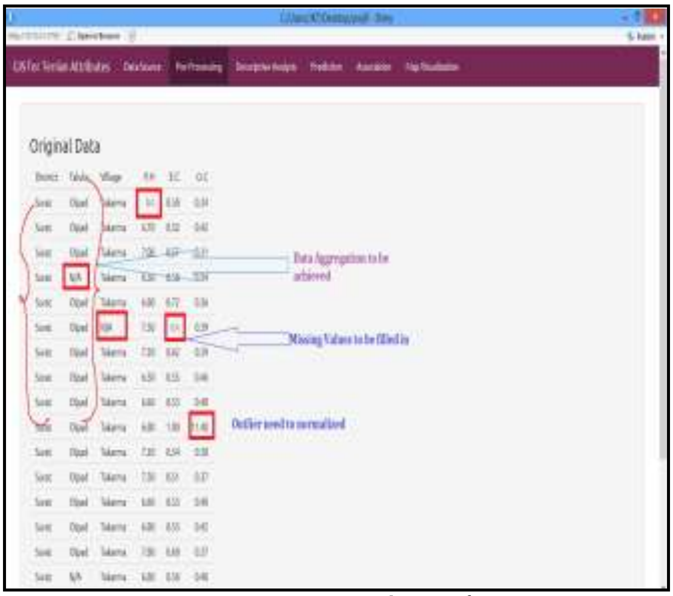

Figure 3. Pre-processing Web-page of GIS

- **Data Transformation:** For each village approximately there were more than 20 samples for a taluka. There are no disclosed alternatives that are online or offline available that maps Survey Number and Area of the field to the coordinates. Moreover a village can have multiple fields without any co-ordinates, therefore it was essential to summarize these multiple records to a single record. Such task was accomplished by calculating the average value for each village. In this manner the researchers applied **data aggregation** which saves execution time as well as memory. Attributes like Latitude, Longitude and Distance are very much important for prediction and modelling geo-spatial variability. Therefore these have been implemented by authors and thus have deployed **attribute construction**. Further it is essential that before geo-spatial prediction is carried out appropriate Co-ordinate Referencing System and Datum has to be assigned followed by relevant projections. To accomplish this WGS84 Datum has been used for same.

- **Fetching Geo-Coordinates:** To make the soil data spatial it is essential to have coordinates well known as Longitudes (x-axis) and Latitudes (y-axis). The proposed GIS Prototype encompasses to fulfil this crucial task whereby after the data have been pre-processed, the system fetches coordinates and store them into the file. While performing attribute construction by fetching geo-codes, it was found that in some cases the result of geocodes returned were NA or NULL. While analysing the issue it was found that this was due to the mismatch between the spellings of village names that appeared on Open Street Map or Google Maps and that with the spelling names that existed in the datasets. The example of same is given below.

  e.g. village name in Gujarati appears : જોલવા

  English transliterated village name on Google Map appears: Jolva. Spelling variations found in datasets: Jolva, Jorva, Jolvaa, Jolwa, Jorwa, Jolwaa. Moreover fetching of co-ordinates online all the time for the same villages is time consuming as well as not a feasible approach. Therefore, to overcome this issue a concept of string similarity has been implemented. Village Master file is created. The most correct spelling of the village name availed from the maps have been replaced with the spellings that appear in the master file. A taluka master is created with Taluka ID and Taluka Name. The village master file that was prepared is now being updated by adding three more fields Taluka ID, Longitude and Latitude that are to be retrieved only once from the Google Map. List of the villages (taluka wise) is stored in a master text file which is then converted to a data frame in R for smooth retrieval process.
- **Data Reduction:** A geo-scientist or other geo-stakeholders upon having adequate terrain attributes and topographic features might also be interested to study the impact of each attribute individually or in combination on the terrain fitness. Thus, to accommodate the impact of these attributes and features like Elevation, Roughness Index, Flow Direction etc., Principal Component Analysis and Step-wise Regression have been facilitated for reducing the multiple independent variables
- **Data Integration and Storage:** The terrain attribute data is derived from various sources, therefore some files may have only taluka and village names but no district specified, the other have taluka but no district. Some sources mention about year whereas the others did may not. Manually doing the task is very time consuming. Therefore the solution was achieved through data integration wherein we designed attribute tables to act as a standard for new inputs. If the input attributes deviated from those of the standards, the proposed algorithm would be executed to resolve these issues.

C. Predictive Modelling

The next module is predictive modelling which is designed to analyze the geo-spatial variation as well as input to the terrain attribute prediction model. The agronomist will enter the village of his interest. The output of pre-processing module acts as an input the variogram sub-module.

- **Variogram Modelling:** The module will select neighboring villages based on the distance from the village of prediction. This number by default we take as five considering the results derived from calibration and validation phase of authors' research on modelling geo-spatial variability. Here in this component, first the sample variogram will be computed and then fitted with Spherical Model along with four Nugget, Sill and Range combinations as proposed by the authors, for the calculation of model weights [14, 15].
- **Terrain Attribute Prediction:** Once the Variogram model has been fitted along with four NSR combinations proposed by us, this module will then predict the terrain attribute value using Ordinary Kriging. Since four NSR combinations have been entered with Spherical model, predictions will be calculated four times for each value of pH, EC and OC. Calculation of prediction accuracy and intermediate results preparation is also done. The maximum out of the four values will be then selected for classification of terrain attributes [10-13]. The entire flow is presented in Figure 4 and predictions appear in Figure 5.

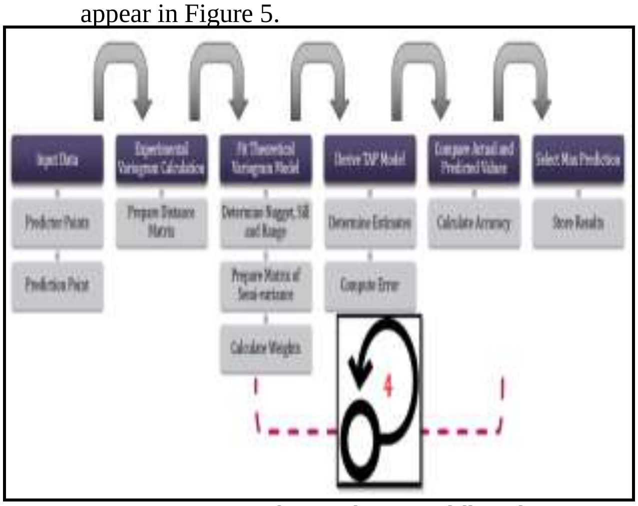

Figure 4. Terrain Attribute Prediction Modelling Flow

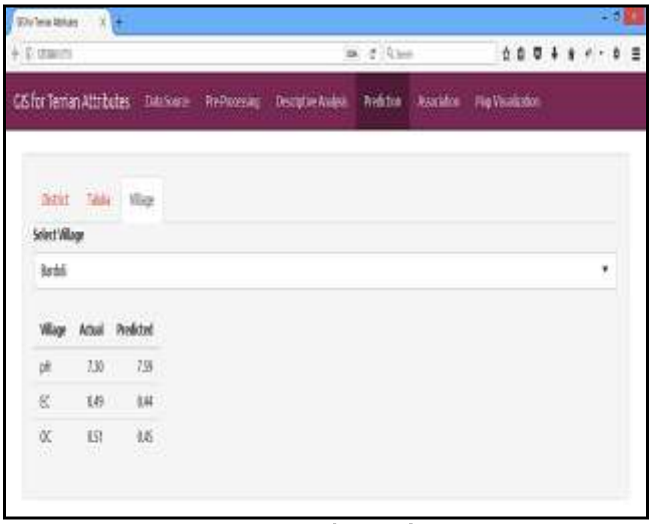

Figure 5. Terrain Attribute Prediction Page

ISSN: 2454-4248
230 – 233

TABLE 1 TERRAIN LABELS

Labels			
pH	***EC***	***OC***	***Combined***
PH1	EC1	OC1	PH1EC1OC1
PH1	EC1	OC2	PH1EC1OC2
PH1	EC1	OC3	PH1EC1OC3
PH1	EC2	OC1	PH1EC2OC1
PH1	EC2	OC2	PH1EC2OC2
PH1	EC2	OC3	PH1EC2OC3

- **Terrain Attribute Labelling:** After performing Variogram analysis and prediction of the chemical properties under study i.e. pH, EC and OC the next step is to categorize the prediction of a particular village to its relevant quality label. The quality labels have been generated using standard values of pH, EC and OC that divides pH into 7 classes, EC into 4 and OC into three classes as show in Table 1.

D. Cartographic Visualization

Every GIS is incomplete without Cartographic visualization or map. The proposed GIS for Terrain Mapping facilitates the geo-community to view maps of their region of interest. Therefore this module has been designed by the researcher to achieve this functionality. District-wise, taluka-wise village maps with terrain attributes can be viewed by the beneficiaries for further decision making. Sample map of district bearing taluka and villages for which data has been collected in represented in Figure 6, and Figure 7 respectively.

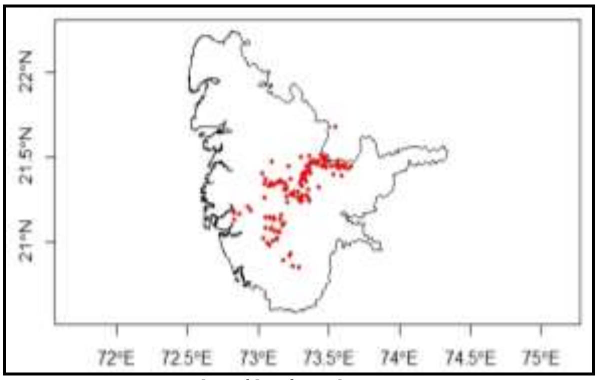

Figure 6 Shapefile of South Gujarat Region

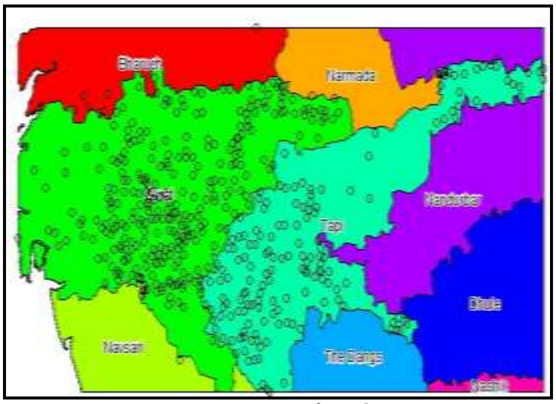

Figure 7 District Map of South Gujarat

IV. Conclusion

The proposed system prototype was validated with extensive experiments which resulted to overall accuracy more than 70%. The data related issues though can be resolved through readily available solutions, still it is necessary to have higher end of customization for better results. Therefore, native GIS solutions like that proposed in this paper by authors are reasonably reliable if data and logical inconsistencies are considered keenly and overcome with customize solutions.

References

[1] Buckey, D. (2009, August). Raster Data Formats. The GIS Primer : An Introduction to Geographic Information Systems. Retrieved October 13, 2015, from http://planet.botany.uwc.ac.za/nisl/gis/gis_primer/page_17.htm

[2] Burrough, P. (1990). Principles of Geographical Information Systems for Land Resource Assessment (6 ed.). Oxford Clarendon Press

[3] A. Dasgupta, "Make it in India," Geospatial World, 7 March 2016.

[4] ICAR, "Vision 2050," Indian Institute of Soil Science, Bhopal, 2015.

[5] Indian Society of Soil Science, "Soil Health: Concept, Status and Monitoring," Bulletin of the Indian Society of Soil Science, pp. 1-98, May 2016.

[6] C. Eick , "Extracting Regional Knowledge from Spatial Datasets," Department of Computer Science, University of Houston, 2010.

[7] Indian Institute of Soil Science, "Vision 2030," Director, Indian Institute of Soil Science, Bhopal, 2011.

[8] Zhang, "Data preprocessing," 11 May 2011. [Online]. Available: http://www.cs.wustl.edu/~zhang/teaching/cs514/Spring11/Data-prep.pdf. [Accessed 21 December 2016].

[9] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, 3 ed., Waltham: Morgan Kaufmann Publishers Elsevier, 2012, pp. 83-111.

[10] K. Mulatani, A. Mohammad, M. Jadav and J. Tailor, "Analyzing Spatial Autocorrelation in Distribution of Soil Chemical Properties using Moran'I and Variogram for Talukas of Surat District," National Journal of System and Information Technology, vol. 9, no. 2, pp. 97-107, December 2016.

[11] N. Kumar, "Investigation the Potentiality of Regression Kriging in the Estimation of Soil Organic Carbon Versus the Extracted Result from the Existing Soil Map," Indian Institute of Remote Sensing, Dehradun, 2009.

[12] ESRI, "ESRI Shapefile Technical Description," ESRI White Papers, pp. 1-34, July 1998.

[13] H. Patel, F. Patel , J. Prajapati and J. Tailor, "Regression Kriging : Disguised and Unveiled," Nationa Journal of System and Information Technology, vol. 6, no. 2, pp. 167-177, December 2012.

[14] R. Rathod, G. Kanet, N. Sardhara, H. Solanki, B. Vaghani and J. Tailor , "Application of Ordinary Kriging on Educational Datasets," National Journal of System and Information Technology, vol. 7, no. 2, pp. 91-98, December 2013.

[15] J. Tailor, R. Gulati and R. Tailor, "Application of Ordinary Kriging on Educational Data using FOSS4G: R Platform," in OSGEO-FOSS4G Proceedings, Dehradun, 2015.

[16] J. Tailor and R. Gulati, "Comparing Prediction Accuracy of OK and RK for the Soils of Surat Talukas," in IEEE TIAR-2015, Chennai, 2015.

[17] J. Tailor, K. Lad and R. Gulati, "Measuring Spatial Correlation of Soil pH and Fe using Theoretical Variograms," International Journal of Computer Science and Networks, vol. 6, no. 5, pp. 533-538, October 2017.