Application of Ordinary Kriging on Educational Data using FOSS4G: R Platform

Jaishree K. Tailor[1], Dr. Ravi M. Gulati[2] and Rashvin Tailor[3]
[1, 3] Shrimad Rajchandra Institute of Management and Computer Application-UTU
[1, 3] Bardoli, India, [1]jaishree.tailor@utu.ac.in ,
[2]Department of Computer Science,Veer Narmad South Gujarat University,Surat, India, rmgulati@gmail.com
[3] rashvin.tailor@utu.ac.in

***Abstract*— Geographical Information Systems (GIS) along with Remote Sensing plays an important role for almost all the areas to facilitate real time location based decision making. Public and private sector as well as academics are benefited with these systems. Geostatistics is an important component of GIS featured to predict or estimate the variables of interest at unknown locations through varying kriging algorithms. The renaissance fetched by Free and Open Source Software for Geo-spatial (FOSS4G) have opened up abundant options for geo-spatial processing. This paper reviews FOSS4G libraries, plugins, GISs as well as Exploratory Spatial Data Analysis software (ESDAS) available for kriging and its types along with testing the feasibility of applying Ordinary Kriging on educational datasets using geoR and Gstat packages of R. Secondary data in the form of overall results of GTU (Gujarat Technological University) for July'2010 examination of thirty two MCA institutes of Gujarat have been selected for this study. Summary statistics, semi-variogram of these institutes have been calculated and variograms using spherical and exponential models have been fitted. Further Ordinary Kriging was applied to predict the results of sample institutes. The simulated results obtained from Kriging are then compared with actual results. RMSE and MSE is used to check the validity of these models. Predicted results prove that this research opens up a platform for applying Kriging on educational datasets using R one of the FOSS4G tool.**

Keywords- Data mining; Educational; FOSS; GIS; Interpolations; Kriging; Remote Sensing; Semivariogram; Variogram

I. INTRODUCTION

Interpolations like Splines, Inverse Distance Weighting, and Kriging being very important components of GIS are widely applied in the field of geosciences. Kriging is a geostatistical method of spatial interpolation used to assess the quality of precision with estimated prediction errors. A Kriging technique which uses an average of a subset of neighboring points to produce a particular interpolation point and the summation of weights tends to unity is known as Ordinary Kriging [1]. The application areas of kriging ranges from Mining, Geosciences, Hydrology, Fisheries, and Environmental to Economical, Socio-Economical, Health, Computer Science. Predictions are also performed in the field of educational mining. Data mining on educational datasets is usually done using proprietary software SPSS or using open source mining tools such as Weka, Rapid Miner, KNIME, Orange to predict student performance, effective teaching methodology, prospective admissions, efficient educational policy etc. [2]. Since, the general task of Kriging is to predict or estimate the variables of interest at unknown locations, therefore there exist a probability to predict unknown educational parameters at unknown locations using Kriging. Thus this study tests the feasibility of applying Kriging on educational datasets. The basic aim of this paper is to highlight the extent of application of Kriging especially kriging on educational datasets using FOSS4G tools.

II. REVIEW OF KRIGING APPLICATIONS & FOSS4G

A. Kriging Application Areas

In [3] the authors applied kriging in financial sector to predict the value of stock exchange index and its comovement. While the authors in [4] predicted prices of real estate using cokriging on the basis of location. Kriging have been applied for estimation of Census attributes like distribution of house income on the basis of socio-economic parameters [5]. The effects of air pollution on student performance was determined in [6] using kriging whereas estimation of risk and portfolio loss was carried out with kriging in [7]. Kriging is widely used for prediction of soil fertility and soil mapping as presented in [8]. [9] Highlights the use of punctual kriging for image restoration and [10] discusses about application of kriging in health sector to estimate health related hazard due to concentration of heavy metals in blood and urine.

B. FOSS4G Tools for Kriging

Free Open Source Software (FOSS), also well known as Open Source or Free Software, are characterized by their freedom to use, modify, and distribute. FOSS provides more quality solutions with capability, security, and sustainability at much less cost, and are increasingly the first software option for individuals and organizations [11]. Since Geographical Information Systems play an important role for almost all the sectors for effective management of environment, the need for remotely sensed data and geospatial analysis on such data for effective policy making is very essential for the public and the private sector along with academic community. FOSS4G have turned out to be a boon in this context by providing ample

solutions in the form of Remote Sensing software, Libraries, Mapping Tools, Exploratory Spatial Data Analysis Software, Desktop and Web-based Geographical Information Systems (GIS) as well as Mobile GIS [12]. This section of the paper exhibits review of only those FOSS4G that provides the functionality of Kriging and those that integrate with other FOSS4G to accomplish this geostatistical interpolation. Thus in all a short review of thirty eight free and open source geospatial tools in attended in this paper subject to access of websites FOSS4G, AI-Geostat, OSGeo and Spatial analysis online [13][14][15][16]. Table I shows rich set of free and open source geospatial libraries that are available for Kriging. Whereas Table II and III presents free and open source GISs and ESDAS that fulfills geostatistical functionality.

III. MATERIALS AND METHODS

This section describes about the sample data and experiments done in R 3.2.0 for Ordinary Kriging on educational data sets:

A. Study Area, Sample Datasets and Data processing

The spatial variability measurement for the performance of GTU affiliated institutions is carried by obtaining institute wise results of MCA programmed of GTU for the month of July 2010. In all thirty one institutes' results were taken into consideration which did not have co-ordinates. Co-ordinates were extracted from the Open DEM and EPSG's website [17] [18].The study area falls under EPSG: 32644 WGS 84 / UTM zone 44N with minimum latitude value as 20.290827 and maximum as 26.829077 while minimum value for longitude is 70.745747 and maximum value is 80.055567. Five institutes were randomly selected as samples. For this study LibreOffice Calc was used to enter data with namely five fields: Institute Id, Institute Name, Latitude, Longitude and Result. The file was then saved as Comma Separated Value (.csv). This file was then imported into R and further converted into spatial data frame which is shown in Figure 1. Descriptive statistics as shown in Table-IV was calculated for the twenty six institutes under study in R 3.2.0.

TABLE I. FREE & OPENSOURCE GEOSPATIAL LIBRARIES FOR KRIGING

Sr.No.	Library	In-built	Support	Description
1.	Dot Spatial		✓	A GIS library written for .Net4 used for spatial data modelling, analysis, mapping with an aid for GIS extension.
2.	GDAL/OGR	✓		A combined library for raster and vector
3.	GeoPack	✓		Geostatistical Software for Waste Management supporting Point or Block kriging, Cokriging, Disjunctive kriging and Cokriging
4.	geoR	✓		Provides geostatistical functionality using R platform. Avails Simple, Ordinary, Universal Kriging and KED.
5.	GMS		✓	With support of GSLIB provides Simple, Ordinary, Universal, Indicator kriging.
6.	GSLIB		✓	Geostatistical package for 2D and 3D with Simple and Ordinary Kriging, Kriging with a trend model, Kriging with an External Drift (KED) and Cokriging
7.	Gstat	✓		Program, and R package, for uni- or multivariable geostatistical modelling, prediction and simulation. Support for Ordinary, Simple, Multivariable Universal kriging, KED.
8.	GSTL	✓		A C++ library with comprehensive tools and algorithms for geostatistics. Supports Simple, Ordinary Kriging, Co-kriging and Kriging with trend.
9.	Kriging (R)	✓		A library supporting Ordinary kriging
10.	Kriging (IBM)	✓		Specific module for Ordinary Kriging only
11.	OpenEV		✓	OpenEV is a software library and application for viewing and analysing raster and vector geospatial data. Takes support of GDAL for Kriging.
12.	Sextante		✓	A spatial data analysis library written in Java with the purpose to provide a platform for implementation, deployment and usage of geoprocessing functionality..
13.	TerraLib with Terra Explorer		✓	TerraLib is an open source library of GIS classes and functions library, allowing a collaborative environment for the development of multiple GIS tools. Ordinary and Indicator Kriging.
14.	UNCERT	✓		Package with 3D geostatistics. Simple and Ordinary kriging

Figure. 1. Processing datasets from.csv to R package

TABLE II. FREE & OPENSOURCE GIS WITH KRIGING

SR.No.	GIS	Description	Kriging Types supported
1.	GRASS	Open Source GIS with support for raster and vector and is used for geospatial data management, analysis, image processing, graphics, spatial modeling and visualization.	Ordinary and Block Kriging
2.	gvSIG	Desktop GIS supporting raster and vector formats.Works with Sexante and GRASS for kriging.	Ordinary kriging
3.	ILWIS OPEN	GIS and RS software with image processing, spatial analysis and digital mapping.	Ordinary Kriging, Kriging with anistrophy, with trend and Co-Kriging
4.	QGIS	Open Source Geographic Information System (GIS) licensed under the GNU General Public License.	Ordinary and Universal Kriging
5.	SADA	GIS with 2D geostatistical module	Ordinary kriging, Indicator kriging,
6.	SAGA	Open Source GIS with geostatistical modeling	Ordinary and Universal kriging
7.	S-GeMS	Software for 3D geostatistical modeling	Simple Kriging, Multivariate kriging (co-kriging
8.	SGS	Software for geostatistical simulations	Simple and Ordinary Kriging
9.	SpatDesign	Sampling optimization	Universal kriging, Bayesian linear kriging, Gaussian kriging
10.	Spherekit	Spatial interpolation programs using spherical geometry	Ordinary point kriging
11.	Surface III	Spatial interpolation and contouring program	Universal kriging
12.	uDig	uDig provides a complete Java solution for desktop GIS data access, editing, and viewing	Ordinary Kriging
13.	Vesper	Program for automatic Variogram Estimation and Spatial Prediction with Error.	Ordinary Kriging, Simple, Block Kriging
14.	WhiteboxGAT	An open-source and cross-platform Geographic information system (GIS) and RS software package.	Kriging

TABLE III FREE & OPENSOURCE ESDAS, PLUGINS, PROGRAMS WITH KRIGING

Sr.No.	ESDAS	Description	Kriging Types supported
1.	Agromet	Program for multivariable geostatistical modelling (2D).	Ordinary point kriging , Cokriging
2.	ExploStat	GIS with geostatistical functions for 2D analyses	Ordinary point kriging, Block kriging
3.	E{Z}-Kriging	Teaching Kriging	Block Kriging
4.	Fragstats	A toolkit for ecological raster data.	Kriging
5.	Geo-EAS	Geostatistics in 2D	Ordinary Kriging, Simple Kriging, Block Kriging, Co-Kriging, Universal Kriging
6.	Geostatistical Tool	A pluging using DotSpatial framework	Ordinary Kriging, Simple Kriging, Co-Kriging (to be included next)
7.	Geostatistical Toolbox	ESDA with geostatistics for 1D,2D and 3D	Ordinary Kriging, Simple Kriging, Block Kriging, Co-Kriging
8.	GRADGRID4	An advanced spatial interpolation tool combining GRASS and R functions	Co-kriging
9.	R-Spatial	Project based on R providing mathematical and statistical libraries	Ordinary Kriging, Universal Kriging
10.	SurGe	A computer program enabling generation of surfaces with interpolations	Kriging with Minimum Curvature

TABLE IV SUMMARY STATISTICS

Measures	Size	Min. Value	Median	Mean	Max. Value26
Values	26	11.11	31.74	36.67	62.07

TABLE V OBSERVED AND PREDICTED RESULTS

Sr. No. of the institute	Latitude	Longitude	Results		
			Observed	Spherical	Exponential
26	20.2908	72.8789	28.57	49.95	48.8
8	21.7257	73.04320	63.27	41.74	40.75
30	22.2161	70.5823	45.65	39.84	39.94
2	23.0163	72.5050	44.44	37.1	37.13
3	23.0257	72.5590	37.93	38.23	36.91
		Average	43.972	41.372	40.706

B. Semivariogram and Ordinary Kriging of Institute Results

- Kriging first requires semi-variogram analysis which was done using Gstat and geoR as shown in Figure 2.
- Then both Exponential and Spherical variogram models were fitted as shown in Figure 3 and 4.

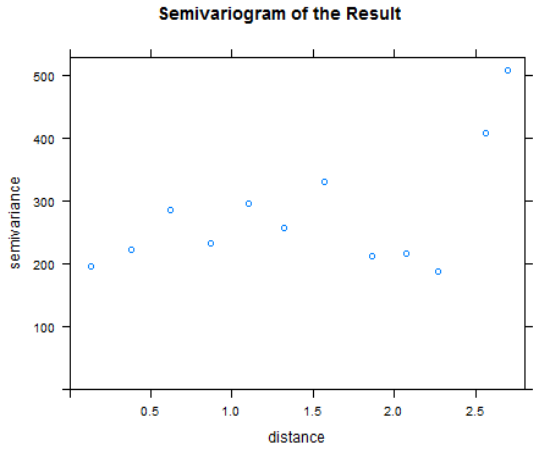

Figure 2. Semivariogram for Results in R

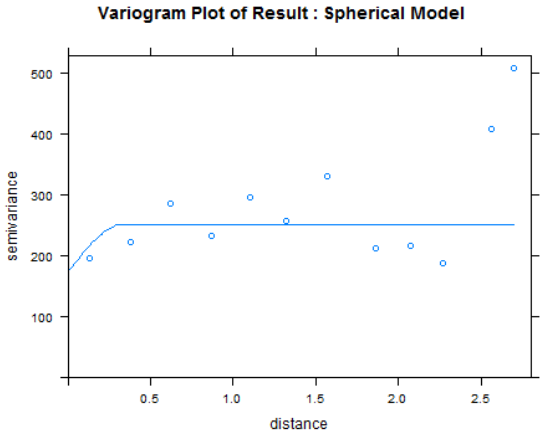

Figure 3 Variogram Plot for Results : Spherical in R

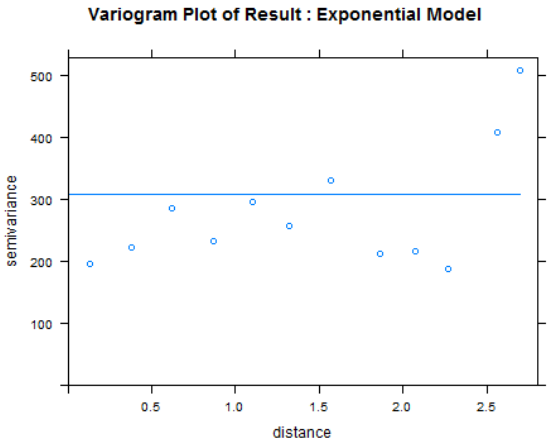

Figure 4 Variogram Plot of Result: Exponential in R

- This was followed by prediction of results for sample five institutes using kriging.
- Semivariogram was plotted using **plot variogram()** function of **gstat** package whereas to estimate the values of the five institutes **ksline()** function of **geoR** package was used.
- Finally to retrieve the estimated values of the sample institutes **predict** was used both for spherical and exponential models and then were compared with original result values.

IV. RESULTS AND DISCUSSION

The average percentage result accounted to 36.67 for the remaining 26 institutes with a minimum result of 26% and a maximum of 62.07%. Semivariogram plot yielded nugget as 200, sill as 400 and range as 2.5. The lag distance was 0.5. Variogram with Spherical model had partial sill as 78.91 and range as 0.34 without nugget. While Exponential model turned out to be a singular model with partial sill as zero and range as 2.5 which means the range values are too small. The observed and the simulated results are shown in Table V. The krigged value of institute number 3 that is 38.23 and 36.91 is very much closer to its observed value that is 37.93. Whereas there exist a difference of almost 7 for institute number 30 and 2 with observed values of 45.65 and simulated values of 39.84, 39.94 and 37.1, 37.13 respectively for spherical and exponential models. Whereas there is a huge difference in the observed and estimated values of institute number 8 and 26. The root mean square error (RMSE) and mean absolute error (MAE) for exponential model was 14.17 and 11.35 respectively. While the values of RMSE and MAE for spherical model were 14.20 and 11.27 respectively which proves that Spherical model is more appropriate in this study.

V. CONCLUSION

The conclusion of this study is categorized into three subsection firstly from the review of kriging applications it can be concluded that kriging has been applied in financial, asset management, econometrics, pollution, geology, image processing and health sectors. Secondly the availability of free and open source tools available for kriging presented in this study it can be concluded that out of fourteen geospatial libraries, plugins eight support Ordinary kriging with other kriging techniques inbuilt whereas the rest use either support of other libraries or frameworks to execute the same. In case of GIS software there are in all ten GISs that support Ordinary kriging along other kriging methods and only 4 GISs do not have support for Ordinary kriging in direct form. However contribution towards support of Ordinary kriging by ESDAS software is also significant with statistics of six out of ten. Thus there are ample of options availed by Foss community for Kriging but the selection of a software depends upon several parameters based on Objectives, Evaluation Criteria and its importance and License type as suggested by Steiniger & Hunter,2013. The later part of the conclusion emphasizes upon applicability of kriging on predicting results of five institutes of Gujarat state affiliated to Gujarat Technological University. The predictions through ordinary kriging algorithm indicates that out of the five institutes the results of three institutes are closure to the observed results which means that Ordinary kriging can definitely be applied on these types of datasets using R platform. Thus this study establishes a platform to apply kriging on educational data sets. Furthermore study and research is required for future enhancement by deploying appropriate auxiliary variables like overall internal results of the institution, merit score for admission etc. in addition to percentage result to improve the values of prophecy and thereby accuracy.

REFERENCES

[1] ESRI,http://support.esri.com/en/knowledgebase/GISDictionary/term/ordinary%20kriging, Accessed online: May 2, 2015.

[2] E. Osmanbegovic, M. Suljic, "Data Mining Approach for Predicting Student Performance", Journal of Economics and Business, vol. 10, no.1, May 2012

[3] G. Aviles, J. Montero, A. Orlove, "Spatial Modeling of Stock Market Comovements", Finance Research Letters Elsevier, vol. 9, no. 4, pp. 202-212, December 2012

[4] J. Olmo, "Prediction of Housing Location Price by a Multivariate Spatial Method: Cokriging", Journal of Real Estate, vol. 29, no. 1, pp.91-114 , April 2007

[5] G. Meng, "Social and Spatial Determinants of Adverse Birth Outcome Inequalities in Socially Advanced Societies", University of Waterloo Canada, Thesis, October 2010, http://hdl.handle.net/10012/5607

[6] M. Sebastian, V. Mauricio, "The Effects of Air Pollution on Educational Outcomes: Evidence from Chile", IDB Working Paper Series, vol. 468, pp. 2-42, December 2013

[7] M. Broadie, Y. Diu, C. Moallemi, "Risk estimation via Regression", Published in Proceeding of 2011 Winter, Simulation Conference(WSC), Phoenix, AZ, pp. 3854-3865, December 2011.

[8] E. Sayed and E. Omran, "Improving the Prediction Accuracy of Soil Mapping through Geostatistics," International Journal of Geosciences, vol. 3, pp. 574-590, May 2012

[9] Chaudhary, J. Kim, T. Tuan, "Impact of Directional Variograms in Fuzzy Type-II Punctual Kriging based Image Restoration", Life Science Journal, vol. 10, no. 1, pp.-1010-1021, 2013, http://www.lifesciencesite.com.

[10] P. Chiang, D. Chan, "A GIS-Aided Assessment of the Health Hazards of Cadmium in Farm Soils in Central Taiwan", International Journal of Environmental Res Public Health, vol. 8, no.9, pp.-3759-3763, September 2011, doi: 10.3390/ijerph8093759

[11] M.Akbari, M. Rajabi, "Evaluation of Desktop Free/Open Source GIS Software based on Functional and Non-functional capabilities", Technique Vjesnik, vol.20, no. 5, pp. 755-764, May 2013.

[12] W. Mohammed, "Free and Open Source GIS: An Overview on the Recent Evolution of Projects, Standards and Communities", Published in Proceeding of The 9th National GIS Symposium, Saudi Arabia, KSA, pp.1-13, March 2015, doi: 10.13140/2.1.3938.2408.

[13] http://wiki.osgeo.org/wiki/FOSS4G [Accessed on 12th May 2015]

[14] http://www.ai-geostats.org/bin/view/AI_GEOSTATS/SoftwarePoints [Accessed on 15th May 2015]

[15] http://www.aigeostats.org/bin/view/AI_GEOSTATS/SoftwareGeostatistical [Accessed on 15th May 2015]

[16] http://en.wikipedia.org/wiki/Open_Source_Geospatial_Foundation [Accessed on 20th May 2015]

[17] http://www.opendem.info/opendem_client.html [Accessed on 9th April 2015]

[18] http://epsg.io/32644/map [Accessed on 9th April 2015]

[19] http://www.grss-ieee.org/open-source-software-related-to-geoscience-and-remote-sensing/ [Accessed on 23rd May 2015]

AUTHORS PROFILE

Ms. Jaishree Tailor is working as an Assistant Professor at Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University(UTU)-Bardoli Gujarat . She has more than 11 years of experience in management and computer science field. Her area of specialization includes GIS and Open Source Technologies. She is pursuing her Ph.d. in Geographical Information Systems from UTU.

Dr. Ravi Gulati is working as an Associate Professor at Department of Computer Science at Veer Narmad South Gujarat University (VNSGU) Surat, Gujarat. He has more than 25 years of experience in academics. His area of interest includes Data Structures, Client Server, and DBMS. He has guided many Ph.D. and P.Phil students.

Mr. Rashvin Tailor is working as an Assistant Professor at Shrimad Rajchandra Institute of Management and Computer Application affiliated to Uka Tarsadia University-Bardoli Gujarat. He has more than 11 years of experience in management field. His area of specialization includes innovation management and open source software development process. He is pursuing his Ph.D. from VNSGU Surat-Gujarat in the area of new product development.