International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 9

ISSN: 2454-4248
59 – 66

A Novel Approach for Management Zone Delineation by Classifying Spatial Multivariate Data and Analyzing Maps of Crop Yield

Naveen Kumar K R
Computer Science and Engineering
Bapuji Institute of Engineering and Technology
Davanagere, INDIA
nvn.dvg@gmail.com

Dr.Nirmala C.R
Computer Science and Engineering
Bapuji Institute of Engineering and Technology
Davanagere, INDIA
crn@bietdvg.edu

***Abstract*—** Precision farming has been playing a distinguished role over last few years. It encompasses the techniques of Data Mining and Information Technology into agricultural process. The acute task in classic agriculture is fertilization, which makes minerals available for crops. Site specific methods result in imbalanced management within fields which affects the crop yield. Treating the whole field as uniform area is merely heedless as it forces the farmers to use costly resources like fertilizers, pesticides etc., at greater expenses. As the field is heterogeneous, the critical task is to identify which part of the field should be considered and the percentage of fertilizer or pesticide required. In order to increase the yield productivity, concept of Management Zone Delineation (MZD) has to be adopted, which divides the agricultural field into homogeneous subfields, or zones based on the soil parameters. Precision Agriculture focuses on the utilization of Management zones (MZs). In this paper, we have collected huge data of Davanagere agricultural jurisdiction during standard farming operations which reflects the heterogeneity of agricultural field. We base our work on a new Data Mining technique called Kriging, which interpolates soil sample values for the specific region, which in turn helps in converting heterogeneous zones to homogeneous subfields.

Keywords- *Management Zone Delineation, Precision Agriculture, Spatial Autocorrelation, Interpolation, Semivariogram, Base Fertilization*

*****

I. INTRODUCTION

Over the past few decades, agricultural field has developed to an extent where larger and faster machines have been utilized to manage farms. But farmers still continued to treat whole fields as uniform elements. The techniques or methods in agriculture are evolving at faster rate, as a part of modern agricultural technique, a new method of farming called Precision Agriculture (PA) has been emerged [4]. In broader sense, Precision Farming can be efficiently defined as the use of Information Technology to improve the decision making process in agricultural production. PA practices contribute in improving the efficiency of yield production and decreasing environmental impact [5]. Since the beginning of PA technology epoch, patterns of crop variability have been considered crucial for variable rate nutrient management. Yield productivity expectations or yield goals influence rate recommendations in much of the cultivated soils in Davangere region. Here we collected huge datasets during standard farming operations, which reflect the heterogeneity of agricultural field. These data can be used for a number of processes there by increasing the efficiency and effectiveness in farming operations. The prime task in classic agriculture is Base Fertilization, which can be described as the process of making the minerals like Nitrogen (N), Phosphorus (P) and Potassium (K) available for the planted crops [6]. PA can also be termed as site specific management. The basic principle in Precision Agriculture or site specific management is that the zone management is implemented in a spatial scale smaller than that of a whole field. Here few simple decision rules and protocols are applied to practically and routinely delineate such Management Zones (MZs) [7]. A number of information sources are used to delineate subfield management zones for site specific crop management. There is lots of data available which contains information about land holdings. Mainly soil properties lead to higher yield which is for farmer's advantage. Based on the biologically valid assumption, a large area of a field contains wide spatial imbalances in soil types and nutrient availability. If these variations are not considered then it result in a loss of productivity. An important role to increase the productivity of agriculture is by using a concept called Management Zone Delineation (MZD) which divides the agricultural field into homogeneous sub field or zones based upon the soil parameters. The concept of management units is proposed in the mid-1990s by Lark and Stafford, 1997 [8] and there are many different delineation methods recorded in the PA as proposed by Taylor et al., 2007 [9]. The majority of statistical methods for MZD, including the original k-means approach by Lark and Stafford (1997), are based on classification algorithms. Classification was performed with the commonly used k-means clustering algorithm using the R Statistical Package [10].

In this paper, we achieve Management Zone Delineation by a method of Interpolation called Kriging, where it interpolates the soil parameters values for a grid location that is not specified in the soil sample database using Nearest Neighbor concept.

II. RELATED WORK

Zhang Xiaohu, et al. (2016) [1] has given a comparison between oval management zones and rectangular management zones, and the author says rectangular management zones improves the delineating of management zones which yields a better performance as well as efficiency when compared to other methods. The method proposed here, is economical and beneficial when applied to large fields. The derived grids have the largest size, and the field in each grid is homogeneous. The experimental results show that the well-designed grids achieved by this method accurately describe the spatial pattern of the soil characteristics and limit the number of instances. Thereby it

 ISSN: 2454-4248
59 – 66

reduces the time required to solve the Binary Integer Linear Programming (BILP) problem and determines the total variance of the delineated management zones.

Nahuel Raul peralta, et al. (2015) [2] proposed an intuitive method involved in Site-specific management of N (NSSM) by adjusting the fertilizer rates to soil characteristics. The aim to undertake this research is to ensure whether delineating of Management Zone (MZ) within fields improves NUE in wheat. This method of delineation not only affords the opportunity of variable rate application of N fertilizers but, also leads to minimization of pollution risk due to an excessive application of resources.

M. Cordoba, et al. (2013) [3] focused on a procedure which is suitable for large multivariate datasets. The delineation of management areas requires Site-specific crop management. The author used Fuzzy K-means to delineate management classes in the multivariate context for analysis. As it has got some of its drawbacks that led to an extension of the MULTISPATI-PCA algorithm. KM-sPC showed the highest yield differences between delineated classes and the smallest within-class yield variance.

III. MATERIALS AND METHODS

A. Study area description and data collection

The research is conducted at different cultivated locations of Davangere region during the 2016/2017 and 2017/2018 growing seasons. Fields planted with different crops are investigated. The area of study i.e., the coordinate points in each of the experimental sites are marked with a GPS receiver with a satellite-based differentiated signal. An overview of the boundary of the study is shown in Fig. 1.

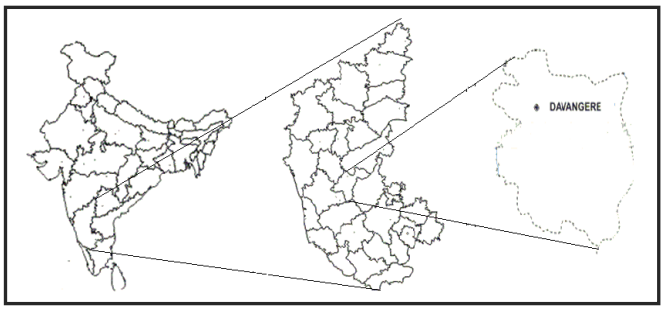

Figure 1. Study Area Boundary.

The soil samples are collected manually and are then taken to the laboratory, air-dried, ground and passed through a 2-mm sieve and then subjected to chemical analysis. The soil properties like pH in suspension, organic matter, available Nitrogen, Phosphorus and Potassium are determined using standard procedures [5]. With the available data three different datasets are made namely Plotted Dataset, Crop Dataset and Map Dataset.

Crop Dataset consists of different crops and minimum soil sample values required to grow them. The different crops that can be grown in Davangere region and the minimum amount of fertilizers or soil sample values like N, P, K and pH required to get the higher productivity are specified in Table I.

TABLE I. CROP DATASET WITH SOIL SAMPLE VALUES

Crop	*N*	*P*	*K*	*pH*
Rice	80	40	40	5.5
Wheat	100	40	0	5.5
Jowar(Sorghum)	80	40	40	5.5
Barley(JAV)	70	40	45	5.5
Bajra(Pearl Millet)	90	30	0	5.5
Maize	80	40	20	5.5
Ragi(naachnnii)	50	40	20	5.5
Chickpeas(Channa)	40	60	80	5.5
French Beans(Farasbi)	90	125	60	5
Fava beans (Papdi - Val)	90	125	60	5
Lima beans(Pavta)	40	60	20	5
Cluster Beans(Gavar)	25	50	25	5
Soyabean	20	60	20	5.5
Peanuts	15	25	0	5.5
Black eyed beans(chawli)	20	60	20	5.5
Kidney beans	20	60	20	5.5

The Map dataset is formed by considering 208 different geographical locations covering the borders of entire Davangere district, where latitude and longitude values are obtained with the aid of Google Maps. TABLE II shows the Map Dataset. Python programming language is used for computation and the code converted these values into a Grid consisting of 125009 points. The grid so obtained is used to predict the percentage of N, P and K values in the soil.

TABLE II. MAP DATASET

Latitude	*Longitude*
13.795882	75.884158
13.794882	75.883158
13.793882	75.882158
13.792882	75.881158
13.791882	75.880158
13.790882	75.879158
13.789882	75.878158
13.788882	75.877158
13.787882	75.876158
13.786882	75.875158
13.785882	75.874158
13.784882	75.873158
13.783882	75.872158
13.782882	75.871158
13.781882	75.870158
13.795882	75.884158

 ISSN: 2454-4248
59 – 66

TABLE III shows the Plotted Dataset for Davangere region. It contains soil sample values like N, P, K and pH and coordinates values where X and Y represent longitude and latitude respectively.

TABLE III. PLOTTED DATASET OF DAVANGERE REGION

Location	*Latitude*	*Longitude*	*N*	*P*	*K*	*pH*
Khudapura	14.4212044	76.5545525	104	23	106	6.98
Manamainah atti	14.4251505	76.5226052	125	28	125	6.77
Turuvanur	14.4003794	76.4305865	103	13	159	5.98
Ganjigatte	14.2341279	76.1157275	63	9	159	8.14
Gyarahalli	14.2285746	76.1075307	54	10	159	8.16
Muthugaduru	14.2197345	76.1154271	103	20	170	7.54
Saasalu	14.1993182	76.1131633	103	15	169	6.43
Daginakatte	14.2005352	75.781889	103	17	166	7.26
Hunesehalli	14.2148461	75.7170703	74	10	179	8.14
Basavapura	14.2251706	75.7335002	85	21	184	7.27
Holemadapura	14.2192406	75.6751802	79	10	154	7.62
Halivana	14.3233453	75.7592645	95	9	198	5.26
Chikkatham manahalli	14.3079906	75.7423603	113	14	190	6.29
Dibbadahalli	14.3363521	75.7415808	76	9	151	5.83
Khudapura	14.4212044	76.5545525	104	23	106	6.98
Manamainah atti	14.4251505	76.5226052	125	28	125	6.77

B. *Dataset Construction*

Finally we get the following datasets,

1. Plotted dataset of soil sample values like N, P, K and pH and coordinates values (latitude and longitude) which are taken from different cultivated locations of Davangere region.
2. Crop dataset consisting of crops name and minimum soil sample values required to grow them.
3. Whole Davangere region Map coordinates which includes 2 fields' latitude and longitude.

C. *Mathematical Model*

Management Zone Delineation is achieved by an interpolation technique called Kriging [15][16] where an overview about Kriging is presented in this part. Kriging is a part of Geostatistics, which is used in mapping of surfaces from minimum sample data and the estimation of values at unsampled locations efficiently. Here Kriging uses a weighted average of neighboring sample to estimate the unknown values at a given known location. The weight is optimized using a semivariogram model as shown in Fig. 3 Spatial Variation of soil can be quantified using the semivariance, which is the expected squared difference between the values of the variable at two different locations.

$$\gamma(h) = \frac{1}{2}E\{[Z(x) - Z(x+h)]^2\} \qquad (1)$$

Where Z(x) is value of the variable of interest (soil properties like N, P, K and pH) at location (x, y) and E { } is the statistical expectation operator.

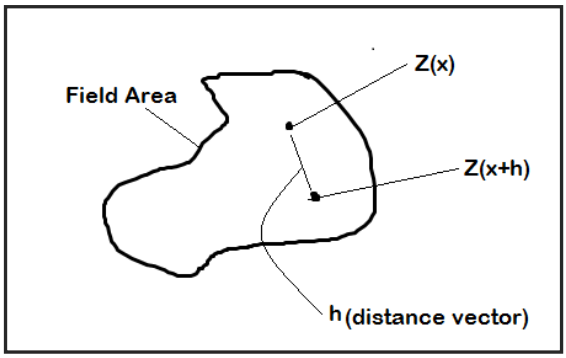

Figure 2. Description

From Fig. 2, it can be stated that the semivariance is about how large the spatial variation is as a function of the distance between two locations. 'E' in equation (1) is the expected value. The equation concludes that larger the distance between two locations, greater will be the difference between soil properties. Also as the distance increases the correlation decreases and becomes zero at some point. The point at which correlation becomes zero is said to be the Range, which is the range of semivariogram, which signifies the distance up to which there is a spatial correlation.

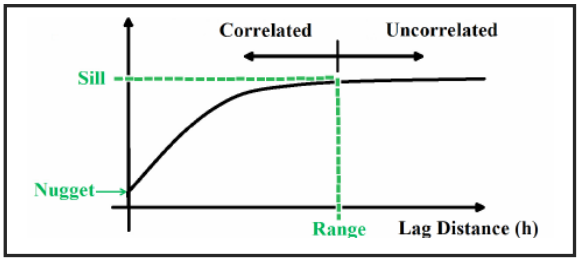

Figure 3. Semivariogram

Sill is called as the maximum value of semivariogram, which is the variance of dataset. Nugget is the measurement error. The main problem is to determine the best predicted value at each uniformed location when only data related to spatial location is available. To make this simple, we modify the statistical expectation operator in equation (1) by replacing x by u and thus we obtain the new estimator Z*(u) as shown below.

$$E\{[Z^*(u) - Z(u)]\} = 0 \qquad (2)$$

Consider the situation when estimation is required at only one location. The simple linear sum can be defined as,

$$Z^*(u) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(u_\alpha) \qquad (3)$$

Equation (2) becomes,

$$\sum_{\alpha=1}^{n} \lambda_\alpha E[Z(u_\alpha)] = \sum_{\alpha=1}^{n} \lambda_\alpha E[Z(u)] \qquad (4)$$

If the expected value is stationary, i.e., constant over the domain, then the unbiased condition becomes,

 ISSN: 2454-4248
59 – 66

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \quad (5)$$

Equation (4) becomes,

$$E[Z(u_\alpha)] = E[Z(u)] \quad (6)$$

The above assumption of a stationary expected value is occasionally useful in practical modeling and hence the phenomenon under study can be decomposed into mathematical terms as,

$$Z(u) = M + R(u) \quad (7)$$

M is some unknown constant which is nothing but expected value of mean (m) and R(u) often represents the term residual, which is spatially varying. Therefore the equation becomes,

$$Z(u) = m + R(u) \quad (8)$$

The above formula can be expanded as,

$$Z(u) = m(u) + R(u) \; with \; E[R(u)] = 0 \; \forall \; u \quad (9)$$

That means,

$$E[Z(u)] = E[m(u)] \quad (10)$$

Subtracting the mean value on both the sides in equation (2), we get

$$E\{[Z^*(u)] - m(u)\} = E[Z(u) - m(u)] \quad (11)$$

Further modifying the above equation, we get

$$E\{[Z^*(u)] - m(u)\} = \sum_{\alpha=1}^{n} \lambda_\alpha E[Z(u_\alpha) - m(u_\alpha)] \quad (12)$$

Here the estimator at only one location is required and hence the above equation becomes,

$$Z^*(u) - m(u) = \sum_{\alpha=1}^{n} \lambda_\alpha [Z(u_\alpha) - m(u_\alpha)] \quad (13)$$

The equation (13) so obtained is the Kriging equation.
Where,

- u, uα are the location vectors for estimation point, the neighboring data point is indexed by α.
- n(u) is the total number of data points in the local neighborhood used for estimation of Z*(u).
- m(u), m(uα) are the expected values of means for Z(u) and Z(uα).
- $\boldsymbol{\lambda_\alpha(u)}$ refers to the estimation location.
- Z(uα) is the Kriging weight assigned to node.
- Z(u) is treated as a random field with the mean, m(u) and the residual component, R(u) = Z(u) - m(u).

Model used:
Ste - Matern, M. Stein's parameterization

$$\gamma(h) = c_0\left[1 - \frac{2}{\Gamma(v)}\left(\frac{h\sqrt{v}}{a}\right)^{v} K_v\left(2\frac{h\sqrt{v}}{a}\right) for\, h > a\right] \quad (14)$$

where $v > 0$ is the smoothness factor parameter. Matern model is a special case of Exponential and Gaussian models. The Matern semivariance γ(h) is a class of semivariance models that provide the flexibility to describe the spatial autocorrelation.

D. Algorithm for Management Zone Delineation

The algorithm for Management Zone Delineation is based on the Interpolation technique called Kriging and K-means clustering algorithms. The former has been described by Hans Wackernagel [13]. The latter has been fully described by Fridgen [7].

1. ALGORITHM Management Zone Delineation
2. // Performs automatic kriging on the given dataset.
3. // Input: formula, input_data, new_data.
4. // Output: An interpolated object containing the results of predictions and the sample variogram.
5. {
6. convert spatial coordinates to create a Spatial object.
7. place Spatial object on the Earth's surface using Coordinate Reference System (CRS).
8. convert the coordinates of input_data into the widely used WGS84 CRS.
9. perform kriging on the given dataset to delineate zones.
10. }

1) Algorithm Description

This algorithm performs Management Zone Delineation. The function takes formula, input_data, new_data as input parameters and outputs an interpolated object containing the results of the interpolation (prediction, variance and standard deviation), the sample variogram, the variogram model and the sums of squares between the sample variogram and the fitted variogram model. The algorithm first converts spatial coordinates into a spatial object, the geometry by which spatial coordinates (attributes) are aggregated. The spatial object so obtained is then placed on the earth's surface using CRS. The delineated zones are obtained by performing Kriging on the given dataset. Here the Kriging predictions (results of Kriging) are calculated as weighted averages of different soil parameters (N, P, & K).

 ISSN: 2454-4248
59 – 66

Arguments

Formula	formula that defines the dependent variable as a linear model of independent variables, for ordinary kriging: 'N~1'
input_data	An object of the SpatialPointsDataFrame-class containing the data to be interpolated.
new_data	A sp object containing the prediction locations. new_data can be a point set, a grid or a polygon.
sample_variogram	An optional way to provide a different dataset for the building of the variogram then for the spatial interpolation.

E. Workflow

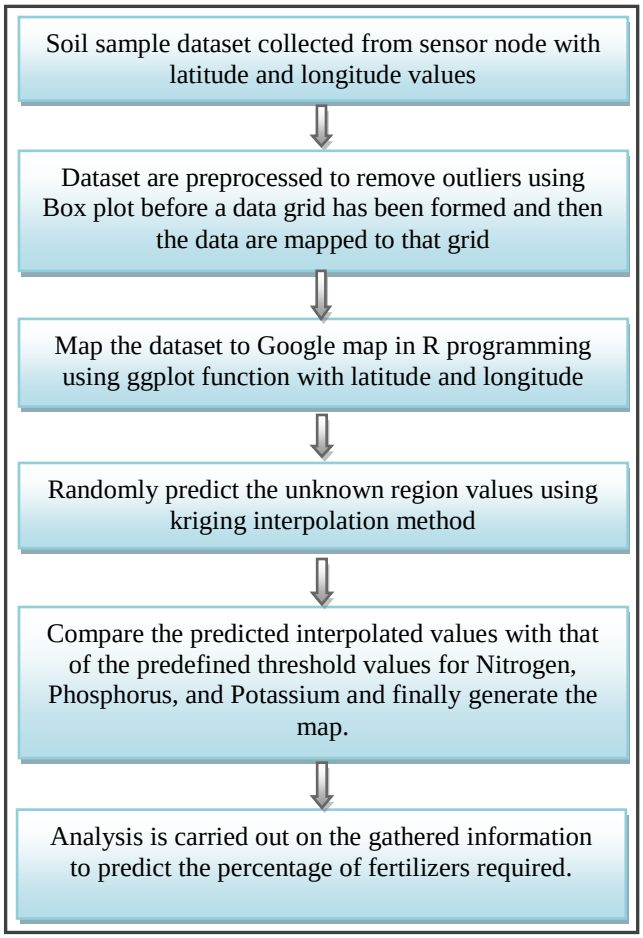

Figure 4. Workflow procedure

Fig. 4 shows the overview of the six major steps. The first step is to collect the soil sample values from different cultivated fields of Davangere region and also longitude and latitude of agriculture fields is recorded. Requisite data, such as physical and chemical properties of soil like N, P, K and pH are obtained by simulation and prediction. With the available data three different datasets namely Crop Dataset, Map dataset and Plotted Dataset are constructed as described earlier. In the second step the data are preprocessed to remove outliers using Box plot previous to that a data grid has been formed and then the data are mapped to that grid and the log transformation techniques are applied to get the data in the desired structure. The third step is to map the datasets into Google map in R programming using ggplot( ) function with longitude and latitude values. The fourth and fifth steps are of at most importance where an interpolation technique called Kriging, which is the best method to achieve Management Zone Delineation, is used to randomly predict the unknown region values.

To achieve this two tasks are essential. The first task is to exhibit the dependency rules. The second is to build the predictions. To accomplish these two tasks, Kriging follows a two-step process. At first it creates a variogram and some covariance functions in order to estimate the spatial autocorrelation which depends on the model that is used for autocorrelation. Then it predicts the unknown values. The major soil properties can be quantified and the respective surface maps are prepared. The final step is to analyze the information gathered from the map which leads to a decision of applying fertilizers or pesticides based on the amount of N, P, K and pH in the soil.

F. Data Analysis

1) Data Preprocessing

The raw data may contain few points that are much larger or smaller than the rest. Such points are called outliers. Sometimes outliers result from data entry errors or measurement error; the resulting variables are not necessarily geographically co-referenced with each other. Outliers should always be scrutinized, and any outlier that is found to result from an error should be corrected or deleted.

The datasets are preprocessed to remove outliers. As mentioned before a data grid has been formed and then the data are mapped to that grid. The geographic coordinates are converted to Cartesian coordinates. Upon performing data preprocessing, boxplots of various soil parameters like N, P and K are obtained, which can be placed side-by-side. This facilitated an easy visual comparison of the features of several soil samples.

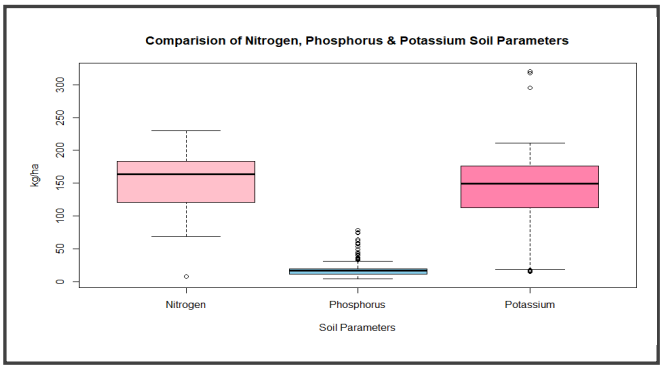

Figure 5. Comparative Boxplot for Nitrogen, Phosphorus and Potassium Soil Parameters

 ISSN: 2454-4248
59 – 66

Fig. 5 presents a side-by-side comparison of the Nitrogen, Phosphorus, Potassium soil parameters.

In Fig. 5, the median is nearer to the third quartile in boxplot for Nitrogen, Phosphorus and Potassium, which showcases that the sample values are comparatively closely packed between the median and the third quartile and more scattered between the median and the first quartile. The upper whisker of Phosphorus is slightly longer than the lower one, which indicates that the data are skewed towards the right. Whereas in Potassium, the lower whisker is a little longer than the upper whisker, which suggests that the data are skewed to the left.

The upper and lower whiskers of Nitrogen lie at 60 kg/ha and 240 kg/ha respectively. Beyond the lower whisker there is an outlier. Similarly the upper and lower whiskers of Phosphorus lie at 6 kg/ha and 85 kg/ha respectively. Here we can see few outliers beyond the upper whisker. The upper and lower whiskers of Potassium lie at 8 kg/ha and 323 kg/ha respectively. Outliers can be seen beyond both the whiskers.

The box for Nitrogen and Potassium are bit taller compared to Phosphorus. Also the upper and the lower whiskers of Nitrogen and Potassium are longer than the Phosphorus and hence we can conclude that the spread in values is slightly larger for Nitrogen and Potassium. In addition, there are several outliers among the data for all three soil parameters.

By applying Boxplot on the dataset the outliers are identified and removed. The next step is to obtain the data in the required form.

2) Data Transformation

Upon removing the outliers, the data are still in undesirable format. Hence one of the log transformation techniques has to be applied to get the data in the desired format.

The main reason behind using the Log Transformation technique is skewness in data, i.e., one or a few points are much larger than the bulk of the data. When Data Preprocessing is performed, it is observed that the data are skewed towards right and left for Phosphorous and Potassium dataset respectively. Hence Log Transformation technique is used.

By the end of this stage, the outliers from data have been removed and Log Transformation has made highly skewed distributions less skewed thus responding to skewness towards large values. This can be valuable both for making patterns in the data more interpretable and for helping to meet the assumptions of inferential statistics.

G. Results and Discussion

Soil and terrain variables vary gradually rather than abruptly over space. Hence, delimiting the spatial boundaries of the subfield areas unambiguously is quite difficult. Here we exploited the spatial autocorrelation to improve the performance of k-means clustering for subfield Management Zone Delineation. We observed that identifying the MZs and their size is strongly related not only to existing within field variation, but also to the different classification procedures and clustering algorithms used. The necessity of the spatial structure of variables included in subfield partitioning has been recognized. Our results show that combining Kriging with k-means clustering technique provides a useful method for subfield classification into Management zones under autocorrelation of input variables. Here we used a variogram to identify the neighborhood of each data point. A graph of semi variances versus distances between ordered data is plotted and as a result a semivariogram is generated.

The availability domain of Nitrogen lies between 22.7 kg/ha to 294 kg/ha. When Kriging is applied on the Nitrogen value dataset of 653 observations, we obtained Kriging predictions (left) and associated Kriging variances or Kriging standard error (right) (for the Nitrogen content). Kriging classified the availability domain into various ranges like [22.7,61.46], (61.46,100.2], (100.2,139], (139,177.7], (177.7,216.5], (216.5,255.2], (255.2,294], which are indicated by different colors in the map shown below.

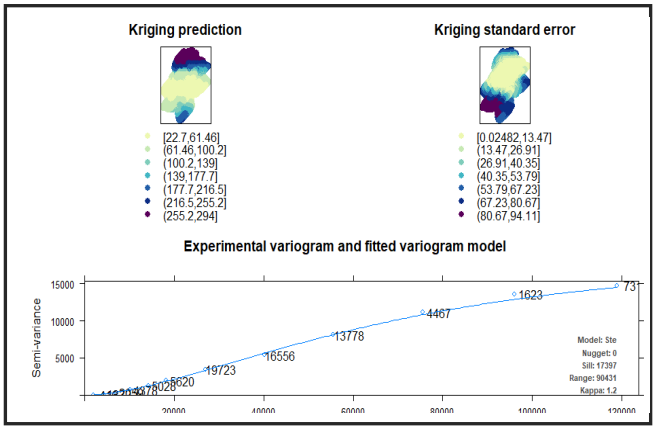

Figure 6. Kriging on Nitrogen

Fig. 6 shows the variogram that is generated for Nitrogen value dataset. The Nugget value is 0, which indicates that there is no measurement error. The Sill value is 17,397, beyond which there is no correlation, the variable is a purely random variable and the variogram is flat. The range value is 90,431 at which the model first flattens out. The three semivariogram parameters namely, nugget effect, range and sill significantly contribute to prediction errors.

The availability domain of Phosphorus lies between 6 kg/ha to 85 kg/ha. When Kriging is applied on the Phosphorus value dataset of 653 observations, we obtained Kriging predictions (left) and associated Kriging variances or Kriging standard error (right) for the Phosphorus content. Kriging classified the availability domain into various ranges like [22.7,61.46], (61.46,100.2], (100.2,139], (139,3.177.7], (177.7,216.5], (216.5,255.2], (255.2,294], which are indicated by different colors in the map shown below.

Fig. 7 shows the variogram that is generated for Phosphorous value dataset. The Nugget value is 0, which indicates that there is no measurement error. The Sill value is 1541, beyond which there is no correlation, the variable is a purely random variable and the variogram is flat. The range value is 87963 at which the model first flattens out.

 ISSN: 2454-4248
59 – 66

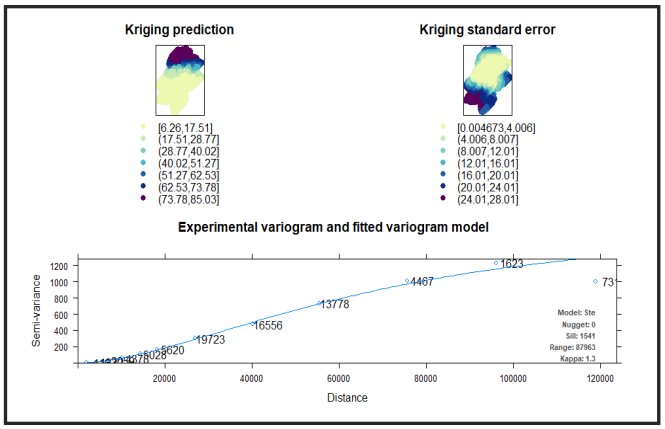

Figure 7. Kriging on Phosphorus

The availability domain of Potassium lies between 88 kg/ha to 324 kg/ha. When Kriging is applied on the Potassium value dataset of 653 observations, we obtained Kriging predictions (left) and associated Kriging variances or Kriging standard error (right) for the Potassium content. Kriging classified the availability domain into various ranges like [8.457,53.52], (53.52,98.58], (98.58,143.6], (143.6,188.7], (188.7,233.8], (233.8,278.8], (278.8,323.9], which are indicated by different colors in the map shown below.

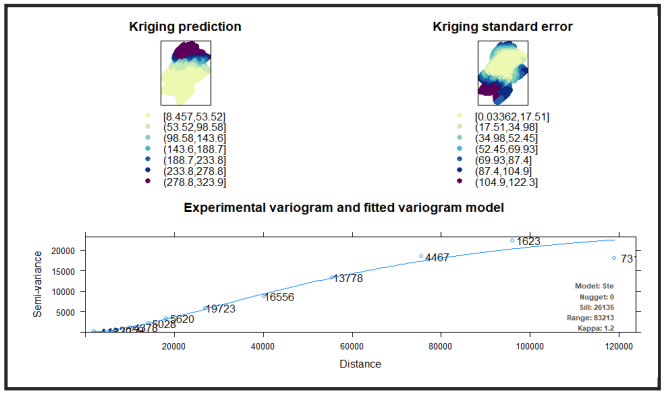

Figure 8. Kriging on Potassium

Fig. 8 shows the variogram that is generated for Potassium value dataset. The Nugget value is 0, which indicates that there is no measurement error. The Sill value is 26135, beyond which there is no correlation, the variable is a purely random variable and the variogram is flat. The range value is 83213 at which the model first flattens out.

Our results showed that the Kriging relies on semivariogram. In simple terms, it can be concluded that semivariogram quantify autocorrelation, because it graphs out the variance of all pairs of data according to distance. The prediction is always strong in Kriging. Input data are used to build a mathematical function with a semivariogram, and create a prediction surface, then validate it. The prediction surface thus facilitated us in describing how well our model predicts. Kriging classified the availability domain of different soil parameters N, P, K into various ranges which are indicated by different colors in the map shown. With this the prediction becomes easy, so that the percentage of N, P, K at different locations can be easily surmised.

H. Conclusion

The research work focuses on a simple approach towards Management Zone Delineation (MZD) by using Kriging. It gave an analysis of how Kriging results in better MZD. The implemented procedure is suitable for large multivariate datasets and requires previous fitting of a variogram model. The algorithm (function) defined performs automatic kriging on the given dataset. We have also investigated the spatial relationship between the data records based on the degree of correlation by calculating nuggets to sill ratio. The generated variogram is used to fit a model of the spatial correlation of the observed phenomenon. The heterogeneity of soil parameters is indicated by variance analysis. Semivariogram model helps in defining the size of variable zones in fields. It provides an easy approach to define zones in fields.

The analysis helps the farmer to identify the region where he can grow the crops with the optimal use of fertilizer and also helps to improve the yield prediction.

Thus instead of using the methods of classic agriculture, as a substitute, we can adopt the concept of MZD to increase yield productivity and to minimize the environmental pollution risks by overuse of fertilizers and chemicals.

I. Future Work

The future development focuses on the strengthening of the proposed work. For spatial interpolation, the performance of the distance-based Kriging can be improved using Linear Regression model. Special attention should be paid to spatial structure modeling, because this stage largely determines the shape of interpolated space as well as prediction errors. The form of the semivariogram is very difficult to identify when limited data is available.

ACKNOWLEDGEMENT

This research work is funded by VGST, Bengaluru, Karnataka and supports to meet the Scientific and Technological needs. I extend my sincere thanks to my research supervisor for her continuous support and encouragement throughout this work. I also take this opportunity to express my sincere gratitude towards my Institution for providing all the necessary research facilities which helped in carrying out the work.

REFERENCES

[1] Zhang, Xiaohu, Jiang, Li, Qiu, Xiaolei, Qiu, Jianxiu, Wang, Juan, Zhu, and Yan, “An improved method of delineating rectangular management zones using a semivariogram-based technique,” Volume 121, pp. 74-83 ,February 2016.

[2] Nahuel RaúlPeralta, Jose Luis Costa, Mónica Balzarini, Mauricio Castro Franco, Mariano Córdoba, and Donald Bullock, “Delineation of management zones to improve nitrogen management of wheat,” Volume 110, January 2015, Pages 103-113.

[3] M.Córdoba, C. Bruno, J. Costa, and M. Balzarini, “Subfield

management class delineation using cluster analysis from spatial principal components of soil variables," Volume 97, September 2013, Pages 6-14

[4] Georg Ruß and Rudolf Kruse, "Exploratory hierarchical clustering for management zone delineation in precision agriculture" Otto-von-Guericke-Universität Magdeburg, Germany, Springer-Verlag Berlin Heidelberg 2011, P. Perner (Ed.): ICDM 2011, LNAI 6870, pp. 161–173, 2011

[5] Rodrigo A.Ortega and Oscar A.Santibáñez, "Determination of management zones in corn (Zea mays L.) based on soil fertility," Volume 58, Issue 1, August 2007, pp. 49-59.

[6] Georg Ruß and Rudolf Kruse, "Exploratory hierarchical clustering for management zone delineation in precision agriculture," ICDM 2011: Advances in Data Mining. Applications and Theoretical Aspects pp 161-173.

[7] Jon J. Fridgen, Newell R. Kitchen, Kenneth A. Sudduth, Clyde, and Fraisse, "Management zone analyst (MZA): software for subfield management zone delineation," Volume 121, pp. 74-83 ,February 2016.

[8] R M Lark and J V Stafford, "Classification as a first step in the interpretation of temporal and spatial variation of crop yield," Ann. Biol. (1997), 130:111-121

[9] James Arnold Taylor, Alex B Mcbratney and B. M. Whelan, "Establishing management classes for broadacre agricultural production," Agron. J. 99, 1366–1376.9

[10] R Development Core Team, "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing", 2008 Vienna, Austria.

[11] Matt A Wood., "Python and Matplotlib Essentials for Scientists and Engineers". June 2015

[12] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From data mining to knowledge discovery in databases," Copyright © 1996, American Association for Artificial Intelligence. All rights reserved. 0738-4602-1996 / $2.00.

[13] Hans Wackernagel., "Ordinary Kriging" 1998

[14] M.Córdoba, C.Bruno, J.Costa and M.Balzarini, "Subfield management class delineation using cluster analysis from spatial principal components of soil variables," Computers and Electronics in Agriculture Volume 97, September 2013, Pages 6-14.

[15] Celal Güvendik, A. Esat Genç and Özgür Tamer, "Improving the performance of Kriging based interpolation application with parallel processors," Signal Processing and Communications Applications Conference (SIU), 2012 20th. IEEE, 2012.

[16] Horvath, Janina, and Kristina Novak Zelenika. "Identifcation of palaeoenvironments using clustering methods and indicator kriging, case study from Late Miocene sandstones, the Sava depression." Nafta 62.11-12 (2011): 364-376.

[17] F. Guastaferro, A. Castrignanò, D. De Benedetto, D. Sollitto, A. Troccoli and B. Cafarelli, "A comparison of different algorithms for the delineation of management zones," December 2010, Volume 11, Issue 6, pp 600–620.

[18] R. M. Lark, "Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield," International Journal of Geographical Information Science, 1998, vol. 12, Pages 83-98.

[19] M. A. Oliver and R. Webster, "A geostatistical basis for spatial weighting in multivariate classification," Mathematical Geology, January 1989, Volume 21, Issue 1, pp 15–35.

[20] Georg Ruß and Alexander Brenning, "Data mining in precision agriculture: management of spatial information," © Springer-Verlag Berlin Heidelberg 2010, pp 350-359.

[21] Andreas Lichtenstern, "Kriging methods in spatial statistics," Bachelor's Thesis, Department of Mathematics, Technische Universit¨at Mu¨nchen, 13.08.2013.