



# The Study of Soil Fertility Spatial Variation Feature Based on GIS and Data Mining

Chunan Li, Guifen Chen, Guangwei Zeng, Jiao Ye

► **To cite this version:**

Chunan Li, Guifen Chen, Guangwei Zeng, Jiao Ye. The Study of Soil Fertility Spatial Variation Feature Based on GIS and Data Mining. Daoliang Li; Yingyi Chen. 6th Computer and Computing Technologies in Agriculture (CCTA), Oct 2012, Zhangjiajie, China. Springer, IFIP Advances in Information and Communication Technology, AICT-393 (Part II), pp.211-220, 2013, Computer and Computing Technologies in Agriculture VI. <10.1007/978-3-642-36137-1\_26>. <hal-01348234>

**HAL Id: hal-01348234**

**<https://hal.inria.fr/hal-01348234>**

Submitted on 22 Jul 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# The Study of Soil Fertility Spatial Variation Feature Based on GIS and Data Mining<sup>1</sup>

Chunan Li,Guifen Chen\*,Guangwei Zeng,Jiao Ye

(College of Information and Technology, Jilin Agricultural University, Changchun, 130118, China)

Author: Chunan Li. Tel: 0431-84532775, Email: anchunli123@sina.com

Corresponding author: Guifen Chen. Tel:0431-84532775,Email: guifchen@163.com

**Abstract.** Studying on the spatial distribution and variation characteristics of soil fertility by Application of geographical information systems (GIS) and data mining(DM), with the quantitative analysis of spatial variability of soil fertility in complexity, and provide the effective way for simulations which can be closer to the farmland soil fertility variability. At the same time, interpolation method based on the spatial variability of soil can provide different accuracy data for soil database, researching of space fuzzy clustering based on precise fertilization regional fertility space can provide a decision-making basis for spatial variability of soil fertility evaluation, a precise prediction and evaluation of precision agriculture are the indispensable basic information and theoretical basis of the development and implementation of precision agriculture. In this paper, the authors take the national 863 project which in the Yushu City of Jilin province as the survey region. And then get the best prediction method and discuss the different effect of the soil nutrient under spatial interpolation prediction model on the analysis of the soil nutrient space mutation characteristics of the foundation by the GIS technology and the geography statistical method. Finally, we analysis the soil fertility status by using the weighted space fuzzy clustering as well as the soil nutrient of space

---

<sup>1</sup> Foundation project: Supported by National “863” High-tech Project (2006AA10A309), National Spark Project (2008GA661003) and Changchun Technology Project (09KT14), The research and application of facilities for the safety of vegetables production technology based on Internet of Things(2011-Z20) .

mutation distribution, so as to provide basis for the study on area space mutation characteristics of cultivated land soil fertility which has an important theoretical and practical effect on the precise fertilization of crop and soil fertility evaluation.

**Keywords:** evaluation of soil fertility, GIS, Data Mining, fuzzy clustering

## 1. Introduction

With the development of precision agriculture, soil testing and fertilizer recommendation is often used methods recently. But a long time, people use the mean of the determination of soil fertility to evaluate the soil fertility status of the entire farmland and use single fertilization method to achieve uniform fertilization. But in this way, we cannot reach the required fertilization purposes, and can't effectively improve the utilization rate of fertilizer. At this stage, precision fertilization usually use GPS to make land of an area divided into equally spaced rectangular or square, and then soil testing and sampling in the middle of the rectangle or square of the district. Sample laboratory analysis; re-use GIS technology processing into fertilizer formula grid diagram. After sample analyzed, we use GIS technology to process the fertilization formula grid map. According to the content of specific fertility in the field for each grid operating unit, variable rate fertilizer applicator adjust the corresponding amount of fertilization to achieve low cost, high-income and green environmental protection. But the method according to mesh fertilizing in lack of areas of variable rate fertilizer applicator has been limited. Therefore, according to the actual situation of the current grid fertilization and uniform fertilization, this study attempts to use GIS spatial analysis techniques and weighted spatial fuzzy clustering and decision tree method, which use fuzzy clustering analysis the soil fertility of agricultural cornfield and propose precision fertilization principles and methods, in order to provide a theoretical basis for the lack of variable rate fertilizer applicator region corn-precision fertilizer. Those provide a theoretical basis for precise fertilization of the lack of variable rate fertilizer applicator of the region to promote corn[1].

The spatial variability of soil properties is the variation of soil chemistry, physics, and biological properties in the spatial location. Even in the same soil type, soil characteristics have huge difference. The spatial variability of soil includes two aspects: one is the vertical profile of the soil spatial variability; another is the spatial

variation of soil plane. Many studies have found that the nutrient with great coefficient variation showing a significant negative bias peak distribution and small coefficient of variation of the nutrient state of normal distribution. The research of spatial variability of soil property is to determine the value of spatial variability of soil property, improve the accuracy of soil mapping, and provide important scientific basis for precision agriculture [2].

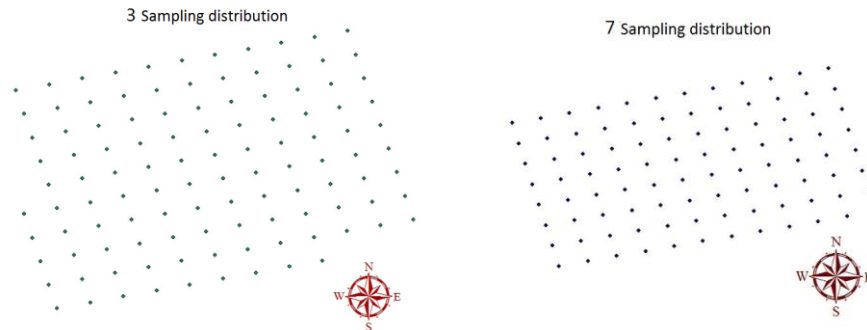
## **2. Materials and methods**

### **2.1 The Overview of research area**

The study area located in the 13th Village, GongPeng Town, Yushu City of Jilin province which is the important commodity grain area of Jilin province, and in which the soil is chernozem and the major crop is corn, soy and so on. This study area is one of the important commodity grain bases in Jilin Province. The terrain is southeast high, northwest low, southeast to northwest tilt. It is at an altitude of 157-220m, longitude 126° 01' 44 " -127° 05' 09 ", latitude 44° 30' 57 " -45° 15' 02 ", belongs to the northeast monsoon in sub-humid temperate climate, average annual precipitation is 500-700mm, rain heat over the same period, longer daylight, conducive to the growth of crops[3].

### **2.2 Sampling methods**

For the implementation of precise fertilization, the first use DGPS to make land divided into grid cell with 40m × 40m, sampling methods see figure 1. We sample in his grid cell, and the sampling depth is 0~20cm. The specific approach: We collect the samples where at the center of the mesh point and the samples that around the center point about 20m away, then mix them for one sample. Using the four division method to process and separate, we collect amount of soil, and put them into the bag with a number mark. Finally, we analyze the samples in laboratory. We use the standard testing method to chemically analysis the nutrient content of pH, available nitrogen, available phosphorus and potassium in the soil samples.



**Fig. 1.** Sampling distribution

### **2.3 Fertility Rating**

According to the soil test results on the 13th Village, Gongpeng Town, Yushu City of Jilin academy of agricultural sciences agriculture information center and using the land statistical interpolation method, we analyzes the soil nutrient (PH, alkali solution nitrogen, phosphorus, all available p, total potassium, rapidly-available potassium) of space mutation characteristics. And then use fuzzy clustering analysis method as well as the decision tree to deal with the problem.

We collect the samples where at the center of the mesh point and the samples that around the center point about 20m away, then mix them for one sample. Using the four division method to process and separate, we collect amount of soil, and put them into the bag with a number mark. Finally, we analyze the samples in laboratory.

## **3 Results and analysis**

### **3.1 analysis of data statistics**

Analysis of research data statistics using GIS software, to get the data in Table 1 below. Table 1 lists the conventional statistical characteristics values of the study area soil nutrient: minimum, maximum, mean, standard deviation, coefficient of variation and skewed coefficient. Table 2 lists the second national soil survey grading standards of soil level.

**Table 1.** soil nutrient statistic characteristic values

Project	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>	<i>CV%</i>	<i>SK</i>
pH	5.50	6.67	6.03	0.26	4.27	0.49
Alkaline hydrolysis N (mg/kg)	68.67	168.55	121.15	13.00	10.73	0.32
Total phosphorus (g/kg)	0.02	1.38	0.45	0.25	56.87	0.82
Available phosphorus (mg/kg)	6.58	98.21	14.86	11.01	74.14	5.16
Total K (g/kg)	6.04	27.55	18.89	5.24	27.74	-0.35
Potassium (mg/kg)	50.00	180.00	84.68	21.69	25.61	1.22

**Table 2.** Soil nutrient grading standards

Project	One	Two	Three	Four	Five	Six
pH	≤4.5	4.5-5.5	5.5-6.5	6.5-7.5	7.5-8.5	8.5-9.0
Alkaline hydrolysis N (mg/kg)	>150	120-150	90-120	60-90	30-60	≤30
Total phosphorus (g/kg)	>1	0.8-1	0.6-0.8	0.4-0.6	0.2-0.4	≤0.2
Available phosphorus (mg/kg)	>40	20-40	10-20	5-10	3-5	≤3
Total K (g/kg)	>25	20-25	15-20	10-15	5-10	≤5
Potassium (mg/kg)	>200	150-200	100-150	50-100	30-50	≤30

Can be seen from Table 1 and Table 2, the study area, soil pH in the range between 5.5-6.67, soil pH is a three level; soil available nitrogen content between 68.67mg/kg-168.55mg/kg, mean for 121.15mg/kg, in view from the mean, indicating that the soil nitrogen content of the study area in a two level; total phosphorus content of the soil between 0.02g/kg-1.38g/kg, mean for 0.45g/kg. In view from the mean, indicating that the soil phosphorus content in the fourth level; soil available phosphorus content in 6.58mg/kg-98.21mg/kg, the mean for 14.86mg/kg. In view of the mean, indicating that the soil phosphorus content in the three level; total potassium content of the soil between 6.04g/kg-27.55g/kg, averaging 18.89g/kg. In view from the mean, indicating that the study area soil potassium at three level; available potassium 50mg/kg-180mg/kg between, mean 84.68mg/kg, from the mean point of view, indicating that the study of soil available K in the four level. In

accordance with the size of the coefficient of variation describes the degree of variation, coefficient of variation less than 10% belong to the weak variability of the coefficient of variation; greater than or equal to 10% and less than or equal to 100% of moderate variability; coefficient of variation greater than 100% of the strong variability. Regional soil nutrient coefficient of variation between 4.27% -74.14%, in addition to PH other than the other belong to the moderate variability can be seen from Table 1.

The coefficient of variation of soil available phosphorus is max, 74.14%, far greater than the other elements; pH of the coefficient of variation is minimum, 4.27%. The order of the size of the coefficient of variation of nutrient available: P> total P> K> available K> alkaline hydrolysis N> pH.

The distribution of soil nutrients type relates to the correct evaluation of the soil nutrients. The skewed reflect the frequency distribution of the deflection,  $SK = 0$  when the distribution is symmetry, positively skewed  $SK > 0$ , negative bias,  $SK < 0$ . It can be seen from Table 1, total potassium in the skewed coefficient is negative, the other are positive. For the Kriging method, if data distribution is different, require the use of the method is also different. If the data obeys the lognormal distribution, the selection of the lognormal Kriging interpolation method; If it doesn't obey simple distribution, selection of disjunctive Kriging method; When the data in the presence of dominant trends, selected Fankelige method; When you need to know is whether the attribute value exceeds a certain threshold, the choice of indicator Kriging method; When the same thing with two kinds of attribute relationship, and an attribute is not easy to obtain, choose collaborative Kriging way; If the attribute value expectations for a known constant, with simple Kriging; if the attribute value expectation is unknown, the selection of the ordinary Kriging.

### **3.2 Spatial variability of soil fertility**

One starting point of this study is that obviously difference between soil fertility of different sampling points while use the software ArcGIS, and used to guide the production is valuable, and have a clearly concise indication of existing field's fertility[4]. According to the obtained data, using GIS spatial analysis and applying statistics features of ARCMAP 10. With LPI (Local Polynomial Interpolation) Interpolation analysis fertility data of discrete sampled points to get soil available nitrogen, phosphorus and potassium, pH spatial distribution map. Take phosphorus



and potassium spatial distribution map of No.7 and No.3 for comparison.( In Fig.2, Fig.3)

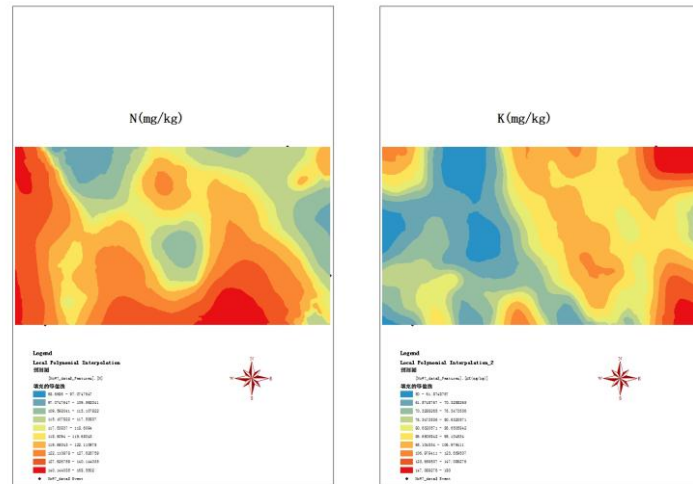


Fig. 2. Spatial distribution of soil fertility in No.7 field

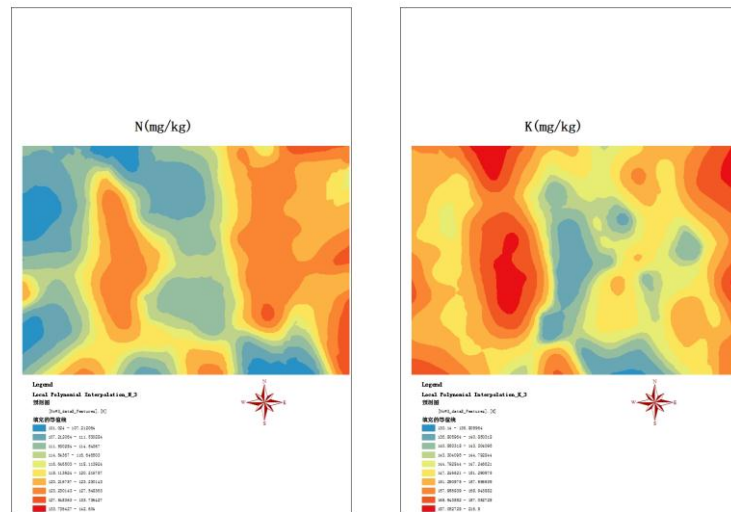


Fig. 3. Spatial distribution of soil fertility in No.3 field

From the figure, there is an obvious difference of the distribution of soil fertility in the third and seventh field. The potassium and nitrogen in the seventh field was

obvious marginalization and unevenly distributed. There is a little peak among the third field and with block feature. Block edges and peaks are mainly caused by the current production mode. Spatial variability of these maps are clearly show the distribution of fertility and easy to apply to the actual production.

### 3.3 Fuzzy clustering analysis

In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster.

Equipped with  $N$  Disaggregated indicators,  $M$  Category object  $X_1, X_2, \dots, X_m$ , each category  $X_i$  have  $N$  objects  $(X_{i1}, X_{i2}, \dots, X_{im})$ . There should be five steps to Fuzzy clustering the  $M$  object.

(1). Selection of the appropriate mathematical formulas to original data  $\{X_{ij}\} (i=1,2,\dots,N, j=1,2,\dots,N)$  for standardization

(2). The selection of the appropriate establishment of mathematical formulas of the  $M$  category object based on fuzzy similarity relation matrix  $R = (r_{ij})_{m \times m}$ .

(3). Using this methods to established relationship matrix  $R$  generally satisfies reflexivity and symmetry, have no passing, not the needed fuzzy equivalence relations, requires the following modifications: squares  $R$  makes to  $R \circ R = R^2$  and then  $R^2 \circ R^2 = R^4$  and so on to make  $R^8, R^{16}, \dots$  until get the equation

$R^{2k} = R^k$ . The  $R^k$  is a fuzzy equivalence relation, sign for  $R^* = (r^*_{ij})_{m \times m}$ .

(4). Gradually reduced from 1 to intercept values  $\lambda$  ( $\lambda > \min(r_{ij})$ ), to make matrix  $R^*$ , generate a dynamic cluster map from fine to coarse.

(5). Select the appropriate threshold values  $\lambda_0$ , to interception matrix  $R^*$  or generate dynamic classification map.

Then multiply this matrix by weights vector  $A = (a_0, a_1, \dots, a_m)$  then make weight fuzzy matrix[5].

$$Y = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ 0 & a_1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & a_m \end{bmatrix}$$

$$= \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}$$

In this paper, we use the weighted clustering algorithm to analysis the data of soil organic matter, available N, P and K nutrition which continuous variable fertility for five years (2008 year's data); Using unweighted algorithm to analysis the data of 2008 year. Clustering results as shown in the table.  $\lambda$  represents the degree of similarity between the data in the table, the class number represent the number of categories when the similarity is  $\lambda$ , , the F value is the value of F distribution that using statistics to calculate. The larger the F value, indicating the distance between classes is larger, and then the classification is better.

**Table 3.** Clustering result table

Weight in 2008			Unweight in 2008		
$\lambda$	Class number	F	$\lambda$	Class number	F
1.000	113	0.339	1.000	117	0.327
0.999	100	0.652	0.999	96	1.065
0.998	86	1.305	0.998	76	1.685
0.997	71	2.304	0.997	65	2.459
0.996	48	3.445	0.996	55	2.836
0.995	40	4.594	0.995	47	2.307
0.994	32	4.875	0.994	38	1.845
0.993	28	4.898	0.993	33	2.957
0.992	23	1.371	0.992	30	2.142
0.991	21	1.469	0.991	23	1.971

0.990	17	1.543	0.990	19	1.915
0.989	15	1.084	0.988	17	1.830
0.988	13	1.069	0.986	16	1.889
0.987	9	1.121	0.985	14	2.063
0.982	7	0.935	0.983	13	2.210
0.981	6	0.920	0.981	12	2.081
0.978	4	0.983	0.978	11	2.092

Comparing the weighted clustering and unweighted clustering results in 2008 year shows that: when  $\lambda$  take 0.993, F can has the largest value, and the classification has the best results.(Table 3) At this point, the value of weighted F is 4.898, the value of unweighted F is 2.957, indicating that the gap between classes are larger when have weighted, so when it has the weighted the classification is better. In addition, the unweighted clustering can assigned the different soil nutrient to the same class level by offsetting the differences between the soil nutrient, but when it has the weighted value it can assign to different classes, and then it can has a better reflect of the real situation of soil nutrients, provide a more reliable and more accurate basis for soil fertility classification.

#### 4 Conclusions

In this paper, we use the Yushu City demonstration field data,ArcGIS geostatistical modules and weighted fuzzy clustering algorithm, to research the distribution, and variability of the 576 soil samples' nutrients (including the pH, nitrogen, total phosphorus, available phosphorus, potassium, potassium) in 3 and 7 field in-depth and meticulous. And the result was compared with the second national soil survey. Using ordinary Kriging, inverse distance weighted, local polynomial and global polynomial interpolation method to predict the spatial variation of soil, we compared the accuracy of the four forecasting methods and draw the following main conclusions:

(1) The demonstration field soils is acidic, its pH value is mainly distributed between 5.5-6.5; available nitrogen content in a two level (121.15mg); content of total P(0.45g) and K (84.68mg)are low, in a four level; content of total potassium (18.89g) and available phosphorus (14.86mg) in a three level; from the coefficient of variation

(CV%), the coefficient of variation of soil nutrients in demonstration field in the 4.27% -74.14%, in addition to the PH the other belong to the moderate variation. The coefficient of variation of soil available phosphorus is 74.14 percent, far greater than other elements, the spatial distribution is uneven, appear high or low state; PH variation coefficient is the smallest, only 4.27%. The CV% order of nutrient is available P> total P> K> available K> alkaline hydrolysis N> pH. There is a big difference of the spatial variability of different plots of soil nutrients in the soil. The mobility of available nutrient directly affects their spatial distribution in the soil. Clear the different variation of available nutrient in the soil can be variable fertilization on soil to achieve the purpose of both saving fertilizer and to meet the demand for soil nutrients.

(2) The test results show that the weighted fuzzy clustering algorithm is an effective method of soil fertility evaluation; It can provide more reliable and more effective science basis for variable rate fertilization and soil fertility evaluation through the application of weighted fuzzy clustering algorithm, and thus in the same block, by variables fertilization to significantly reduce the regional differences of soil nutrients, and to increase production.

(3)In the ordinary Kriging, inverse distance weighted, global polynomial and local polynomial prediction method, the accuracy is the highest by using ordinary Kriging Interpolation method to predict the spatial distribution of pH and potassium. But using the local polynomial method to create the spatial variability of N, total P, available P and K can have the highest prediction accuracy. Generally speaking, the Kriging method and local polynomial interpolation method have a high precision to predict soil nutrient spatial distribution, while the global polynomial interpolation in predicting the spatial variation of soil has the least accurate, but has the minimum value.

(4)Fertility evaluation work is an important work of precision agriculture, usually, the fertility evaluation, involved less space factors; it has most experts in the field to participate. They use the experimental data to analyze, it is subjectivity, and also more difficult to express the nonlinear relationship between soil properties and environmental variables and uncertainty information. In this study, we use GIS technology and statistical techniques, clustering algorithms to evaluate of soil fertility which provides a new way and method for spatial data mining and knowledge discovery [6].

## Reference:

1. Liying CAO, Soil Nutrient Evaluation Based on GIS and Fuzzy Clustering Analysis[J] , Journal of Anhui Agri Sci, 2010, 38( 16) : 8595- 8596, 8661
2. Fan Mingfeng, Spatial Variability of Soil Nutrients and Comparison of Predictive Methods Using GIS[M],Southwest University
3. Guifen Chen, Yue Wang,Guowei Wang. Research and application of the maize precise fertilization system. *Journal of Jilin Agricultural University*, 2006, 28(5): 586-590.
4. Yueling Zhao, Guifen Chen, Yue Wang, A Study on Spatial Variability of Soil Nutrients Based on GIS,*Acta Agricultural Boreali-occidentalis Sinica*,2005,14(6):195~198.
5. Guifen Chen, Liying CAO, Guowei WANG, Application of Weighted Spatially Fuzzy Dynamic Clustering Algorithm in Evaluation of Soil Fertility, *Scientia Agricultura Sinica*, 2009,42(10):3559-3563
6. Guifen Chen, Research and Application of Spatial Data Mining Technology for Precision Agriculture[M], Jilin University.