Analysis of Soil Fertility Based on FUMF Algorithm

Hang Chen[1,2], Guifen Chen[2(✉)], Yating Hu[2], Liying Cao[2],
Lixia Cai[2], and Sisi Yang[1]

[1] Institute of Scientific and Technical Information of Jilin,
Changchun 130033, China
chenhang0811@163.com, 1046235989@qq.com
[2] College of Information Technology,
Jilin Agricultural University,
Changchun 130118, China
chenhang0811@163.com, guifchen@163.com,
{65447539,45515189,419513823}@qq.com

Abstract. The soil nutrition is an important indicator of soil fertility. The method K-means and FCM are always used to evaluating the soil fertility, but the cluster number need to be set, and the outlier couldn't be eliminated accurately, and there is the deviation between the real result and the soil fertility. So the paper applied the FUMF to analysis the soil nutrient data of Nong An county for eight years, 2005–2012. The result show that the low fertility soils gradually decreased from 2005 to 2012 by precision fertilization, and the moderate and high fertility soil was rising, the overall soil fertility of Nong An had improved significantly. The analysis result was consistent with the actual situation, The FUMF algorithm is proved that was an effective evaluate method of the soil fertility evaluation. It has the practical significance to analyze the large number of soil fertility of high complexity and interactive, it also provided the technical support for precision fertilization decision-making.

Keywords: Fuzzy clustering · Soil fertility · Fertility analysis · Precision fertilization · FUMF

1 Introduction

Soil nutrient content is an important symbol of fertility and productivity of arable land, also it is an important indicator of soil fertility evaluation. With the arrival of precision agriculture era, spatial variability and correlations of wide variety agricultural data which have complex links relationship are more significantly. The attendant massive, diverse and dynamic changes, incomplete, uncertain and a series of characteristics.

Since the 1990s, Data Mining and geographic information systems technology in the agricultural sector has been increasingly widely used. DM and GIS technology can effectively statistics and analysis of massive, complex data. DM Clustering algorithms can dig out the knowledge of soil fertility evaluation from soil nutrient data analysis.

D. Li and Z. Li (Eds.): CCTA 2015, Part I, IFIP AICT 478, pp. 564–573, 2016.
DOI: 10.1007/978-3-319-48357-3_53

Li et al. put forward the application of clustering analysis which is in site classification and soil fertility evaluation [3]. Zheng et al. improved rough K- means algorithm, and put forward the rough K- means clustering algorithm based on density weighted [5]. Chen et al. put forward a weighted spaces fuzzy dynamic clustering algorithm, and proved the validity of method in evaluation of soil fertility [6]. But conventional K-means, FCM and other clustering algorithms have some limitations on soil fertility evaluation. Such as K- means is hard clustering algorithm that can only get a hard divide. Although FCM can get fuzzy clustering divide, both algorithms require artificially set the number of clusters. So it can not eliminate outlier accurately or solve the problem of soil fertility data including complex, dynamic, and interactive fuzzy. Whatever, the clustering results presence of a certain error with the real fertility. For this reason the paper use FUMF algorithm to analyze and evaluate soil fertility.

National measuring territories precise fertilizer projects in Jilin Province for over 10 years. During this period a large number of soil samples were collected and sample of soil nutrients were determined and analyzed. All of this could lay the foundation for soil fertility status by using DM and GIS technology. Thus, this paper use large amounts of data by successive years of soil testing precise fertilizer projects that from Nong An county in Jilin Province. Then, we use GIS and Matlab technical conducted a rapid unsupervised multiscale fuzzy clustering for soil nutrient data from 2005 to 2012. The results show that FUMF algorithm is an effective method for soil fertility evaluation and has practical significance when analyze large amounts of high complexity, strong interaction soil fertility factors. So, it is can provide a technical support for the precise fertilization decision.

2 Experiments and Methods

2.1 The Situation of Research Area

Nong'an is located in Songliao Plain, Changchun, Jilin. specific in northwest of Changchun city away from 60 km, north latitude 43° 54'–44° 56', longitude 124° 32'–125° 45'; The zone is in the temperate semi-humid continental monsoon climate. So, monsoon features obviously, four distinct seasons, abundant sunshine, less rainfall and the annual average temperature of 4.6 °C, annual average sunshine hours 2590 h, the average annual rainfall 507 mm; On the one hand, there are diverse landforms such as high mesa, mesa, two terraces, a terrace, floodplain, sand dunes, depressions, gullies and so on. Thus, most soil is chernozem, meadow soil and black soil; On the other hand, they grow corn sorghum, wheat, millet and soybean and other crops production as the mainstay. It is arguably one of the country's important commodity grain production bases and its total grain production ranked first in the major grain-producing counties.

2.2 Collection and Analysis of Sample Data

On the basis of field research, we are cooperation with cropland capacity survey quality evaluation office and considering soil types, land use, topography, cropping patterns, management measures and production level and other factors according second national soil survey. After that, we determine the sampling point through DGPS and RS systems. Then we can comprehensive analyze the survey plots of soil testing precision fertilizing work from 2005 to 2012. It collected 23,976 samples, sampling map of soil nutrients in Figs. 1 and 2.

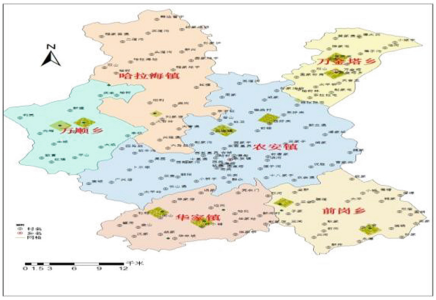


Fig. 1. Nong an soil fertility sampling map

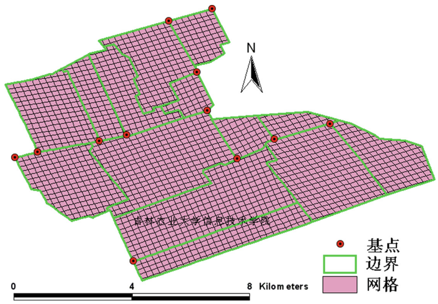


Fig. 2. Part of the grid sampling map from Nong an

Samples were collected depth from 0 to 20 cm. Random multi-point sampling within the same plots. Whatever, after mixing the soil by quartering, we take 1.5 or 2 kg bagging spare. Then, take it back to the laboratory for spare through dry naturally pulverized and sieving. Ultimately, index measuring soil nitrogen, phosphorus and potassium and other nutrients in which a total of 26 kinds of soil types. So, in this paper we analyze the collected 23976 data and calculate maximum, minimum, average value of nitrogen (N), phosphorus (P), potassium (K) according to the different soil types. The data of 2010 is shown in Table 1:

Table 1. Soil fertility data

Town name	Country name	Plot name	Alkalystic N (mg/kg^{-1})	Available P (mg/kg^{-1})	Available K (mg/kg^{-1})	Latitude	Longitude
Nong An	Xi Haolai	Jia Bei	154	28	208	44.58417	125.28982
Nong An	Xi Haolai	Jia Bei	136	31.3	217	44.49895	125.25127
Nong An	Xi Haolai	Jia Bei	132	12.8	217	44.4988	125.2509
Nong An	Xi Haolai	Er Jiedi	132	16.3	198	44.49926	125.25072
Nong An	Xi Haolai	Er Jiedi	121	21.4	208	44.5004	125.24821
Nong An	Xi Haolai	Gong Lubei	118	31.8	227	44.50205	125.24968
Nong An	Xi Haolai	Lu Nan	114	37.9	227	44.50206	125.24738
Nong An	Xi Haolai	Er Jiedi	114	28	237	44.5024	125.24905

(*continued*)

Table 1. (*continued*)

Town name	Country name	Plot name	Alkalystic N (mg/kg^{-1})	Available P (mg/kg^{-1})	Available K (mg/kg^{-1})	Latitude	Longitude
Nong An	Xi Haolai	Yi Jiedi	114	23.3	217	44.5054	125.24378
Nong An	Xi Haolai	Er Jiedi	114	21.3	208	44.50712	125.24472
Nong An	Xi Haolai	San Jiedi	114	19.1	169	44.50855	125.24402
Nong An	Xi Haolai	San Jiedi	121	24.2	160	44.50305	125.23638
Nong An	Xi Haolai	Er Jiedi	118	24.8	140	44.50648	125.24862
Nong An	Xi Haolai	Er Jiedi	118	31.3	179	44.50616	125.24878
Nong An	Xi Haolai	San Jiedi	118	43.6	140	44.50523	125.24845
Nong An	Xi Haolai	San Jiedi	118	32.8	160	44.50502	125.24878
Nong An	Xi Haolai	San Jiedi	110	39.3	169	44.51012	125.2503
Nong An	Xi Haolai	San Jiedi	132	32.8	188	44.50512	125.25013
Nong An	Xi Haolai	San Jiedi	129	37.8	179	44.50522	125.25045
Nong An	Xi Haolai	Er Jiedi	121	23.3	179	44.50755	125.25172
Nong An	Xi Haolai	San Jiedi	114	34.1	227	44.51202	125.2518
Nong An	Xi Haolai	San Jiedi	114	36.8	237	44.5058	125.24893
Nong An	Xi Haolai	Er Jiedi	114	22.7	246	44.50618	125.24475
Nong An	Xi Haolai	Gong Lubei	114	30	217	44.50412	125.23805
Nong An	Xi Haolai	Er Jiedi	118	35.8	208	44.50195	125.2379
Nong An	Xi Haolai	Gong Lubei	129	21.4	237	44.50885	125.24368
Nong An	Xi Haolai	Gong Lubei	140	30.8	227	44.51057	125.24563
Nong An	Xi Haolai	Lu Bei	114	36.3	217	44.51185	125.24018
Nong An	Xi Haolai	Er Jiedi	118	30.8	208	44.51535	125.23695
Nong An	Xi Haolai	San Jiedi	118	36.3	198	44.50502	125.2354
Nong An	Xi Haolai	San Jiedi	118	42.8	227	44.50268	125.23857
Nong An	Xi Haolai	Er Jiedi	103	32.1	198	44.51077	125.25425
Nong An	Xi Haolai	San Jiedi	132	15.8	227	44.50702	125.2542
Nong An	Xi Haolai	Er Jiedi	165	10.8	237	44.51245	125.25438
Nong An	Xi Haolai	Gong Lubei	147	30.8	217	44.51385	125.2544
Nong An	Xi Haolai	Bei Dapian	143	21.3	198	44.5057	125.25578

According to soil grading standards of second soil survey, soil nutrients are divided into six levels, such as shown in Table 2.

Table 2. Soil nutrient grading standards

Project	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Alkalystic N(mg/kg)	>150	120–150	90–120	60–90	30–60	≤30
Available P(mg/kg)	>40	20–40	10–20	5–10	3–5	≤3
Available K(mg/kg)	>200	150–200	100–150	50–100	30–50	≤30

According to preliminary results of the analysis, we begin to accurate classification of soil fertility through data mining.

3 Results and Discussion

3.1 Fast Unsupervised Multiscale Fuzzy Clustering (FUMF) Algorithm

First of all, we clustering the N, P, K three indicators of 23976 data by FUMF algorithm, the purpose is to eliminate the isolated samples point of each index. Then these three indicators data were normalized. Finally, set the parameters of weighted dimensional data for clustering analysis by FUMF.

We can accelerate UMF algorithm through nearest neighbor criterion and get FUMF. Well, FUMF method is divided into two stages:

The first stage: re-expression data by using the nearest neighbor criterion, the data is divided into $\bar{n}$ disjoint subsets S_j, Each subset's data represented by its representative point C_j which is as a whole.

The second stage, implementation of weighted UMF algorithm.

FUMF algorithm is as follows:

Step 1. The re-expression data, initialize the $m-1$, $c_m = \{x_1\}$, $i = 2$ to $N: d(x_i, c_k) = \min_{1 \le j \le m} d(x_i, c_j)$, If $d(x_i, c_k) > \Theta$ and $m < q$ then $m = m + 1$ $c_m = \{x_i\}$ Else $c_k = c_k \cup \{x_i\}$.

Step 2. Clustering UMF, set $j = 1$, set a threshold $\varepsilon > 0$ and $v^{(0)} = c_j$, then using the updated formula 1:

$$v^{(l+1)} = \frac{\sum_{k=1}^{\bar{n}} n_k \cdot c_k \cdot \tilde{d}(v^{(l)}, c_k)}{\sum_{k=1}^{\bar{n}} n_k \cdot \tilde{d}(v^{(l)}, c_k)} \quad (1)$$

Calculate convergence point of c_j, denote as p_j. If $j < \bar{n}$, then $j = 1 + j$, repeat step 2.

Step 3. If $\|p_a - p_b\| \le \varepsilon$, The S_a and S_b of the data points into a class; otherwise, divided into different classes.

3.2 Soil Nutrient Content Analysis

Through statistical analyze 23,976 samples of soil nutrient content, we summarizes the changes of soil nutrients from early, metaphase and anaphase data. As shown in Table 3:

Table 3. Nong an Part of the township of soil nutrient content in different years descriptive statistics

Year	Index	Alkalystic N (mg kg^{-1})	Available P (mg kg^{-1})	Available K (mg kg^{-1})
2005	Range	9.0–193.0	2.0–145.0	8.0–960.0
	Mean	105.79	18.7	138.1
	CV (%)	21.83	67.15	51.15
	Number of samples	2297	2297	2297

(*continued*)

Table 3. (*continued*)

Year	Index	Alkalystic N (mg kg^{-1})	Available P (mg kg^{-1})	Available K (mg kg^{-1})
2009	Range	19.0–211.0	2.1–96.9	40.0–382.0
	Mean	131.84	22.25	184.22
	CV (%)	19.35	54.74	25.87
	Number of samples	5115	5115	5115
2012	Range	88.0–213.0	5.5–82.6	15.0–413.0
	Mean	135.98	24.68	168.47
	CV (%)	13.77	41.97	28.19
	Number of samples	6329	6329	6329

3.3 FUMF Analysis

Taking into account the soil sampling N, P, K three indicators' observed values are different. Data will inevitably be contaminated during sampling that resulting in some isolated points Therefore before cluster analysis of soil nutrient, we need to pre-processing the data set. Pretreatment divided into the following steps:

(1) Executing clustering algorithm for N, P, K three indicators respectively. If it contains a small number of data points when clustering, indicating this category may be constituted by isolated point. In the experiment, we analyze categories which data points lower than 20 and delete those isolated points which beyond the normal range of values.
(2) Because of three indicators of N, P, K have differences in dimension as raw data. Therefore, each of these three indicators were normalized so that the mean of each index is 0 and variance is 1.
After process the raw data, each sample as a data point for clustering. Due to Evaluation of soil fertility mainly depends on the content of P indexes, and P indexes are generally lower than the value of N, K. Thus, we should weighted N, P, K as 1:10:1 before performing clustering algorithm.
(3) Parameter settings: the convergence of the scale parameter is 0.15; convergence precision is 10^{-5}; maximum number of iterations is 100; fuzzy factor is m = 2; data reduction parameters is 0.8; convergence scale parameter 0.14 multiplied mean value; After performing clustering algorithm to pretreatment and weighted data, using inverse transform to get clustering results.

3.3.1 The Initial Precision Fertilization Clustering Results

In this paper, we collected 2297 samples from 27 towns in 2005 to establish the experimental data set(remove isolated points of 38 when prepossessing), all of the data come from Bajilei, Bangchai, Binghe, Fuquanlong, Gaojiadian, Halahai and so on. Then we clustering by FUMF. The clustering results shown in Table 4 and Fig. 3(a 2005).

Table 4. Clustering results in 2005

Category	Available N (mg kg^{-1})	Available P (mg kg^{-1})	Available K (mg kg^{-1})	The amount of data	Fertility
1	105.30	11.30	119.05	824	Low
2	102.83	13.87	122.90	623	Medium
3	105.93	19.50	134.85	427	Medium
4	112.61	64.58	140.52	21	High
5	105.14	54.18	448.18	2	High
6	102.83	32.72	135.60	95	High
7	126.20	47.54	399.26	6	High
8	108.14	38.36	158.07	26	High
9	111.16	40.29	144.33	38	High
10	107.81	44.89	138.15	51	High
11	132.25	60.23	125.78	4	High
12	101.39	29.55	131.55	132	Medium
13	103.13	27.59	129.29	2	Medium
14	107.73	54.61	107.13	5	High
15	132.78	60.18	264.45	3	High

3.3.2 The Middle Precision Fertilization Clustering Results

Experimental data sets with 5115 samples from 23 towns in 2009 (remove isolated points of 24 when prepossessing). The data come from Bajielei, Dehui, Gaojiadian, Halahai, Helong and so on. And clustering results Table 5 and Fig. 3(b 2009).

Table 5. Clustering results in 2009

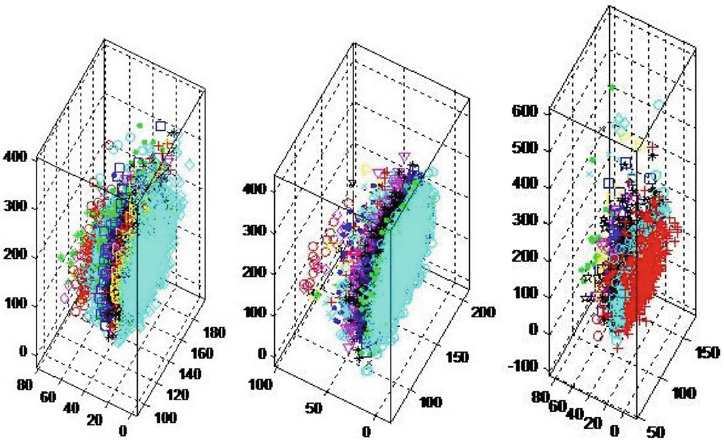
Category	Available N (mg kg^{-1})	Available P (mg kg^{-1})	Available K (mg kg^{-1})	The amount of data	Fertility
1	123.68	43.40	179.97	102	High
2	131.79	13.81	187.19	1499	Low
3	133.58	22.55	185.53	1156	Medium
4	135.92	17.73	189.62	957	Medium
5	132.67	15.66	187.31	147	Medium
6	134.86	19.69	186.40	441	Medium
7	133.14	36.41	186.40	272	High
8	134.98	30.84	185.22	398	High
9	132.40	62.92	178.31	6	High
10	138.88	76.07	152.41	6	High
11	148.39	67.57	224.91	10	High
12	111.89	81.76	163.05	16	High
13	109.95	72.23	136.23	3	High
14	116.30	53.27	144.66	50	High
15	112.51	57.82	140.77	14	High
16	123.68	43.40	179.97	11	High
17	131.79	13.81	187.19	3	High

3.3.3 The Late Precise Fertilization Clustering Results

Experimental data sets with 6329 samples from 17 towns in 2012 (remove isolated points of 17 when prepossessing). All of the data come from Helong, Qiangang, Bajilei, Fuquanlong, Gaojiadian, Halahai, Huajia and so on. Then we clustering by FUMF. The clustering results shown in Table 6 and Fig. 3(c 2012).

Table 6. Clustering results in 2012

Category	Available N (mg kg^{-1})	Available P (mg kg^{-1})	Available K (mg kg^{-1})	The amount of data	Fertility
1	131.79	20.12	170.56	742	Medium
2	134.85	37.22	160.72	561	High
3	133.05	21.75	166.18	399	Medium
4	137.10	31.71	162.83	666	High
5	133.89	25.76	166.14	699	Medium
6	133.15	28.58	166.18	593	Medium
7	134.18	14.20	162.77	1312	Low
8	133.13	23.58	164.96	260	Medium
9	133.80	18.54	169.35	460	Medium
10	135.91	15.77	165.41	316	Medium
11	135.36	30.17	163.07	58	Medium
12	134.90	47.52	161.72	204	High
13	140.12	63.81	133.31	25	High
14	134.71	45.89	170.48	4	High
15	136.68	35.46	160.15	4	High
16	131.79	20.12	170.56	9	High



(a 2005) (b 2009) (c 2012)

Fig. 3. Soil nutrient spatial clustering map

3.4 Clustering Analysis

When compared the clustering results from 2005, 2009 to 2012, we can derive trend of soil fertility that soil fertility tend towards equilibrium and rise after precision fertilization.

Table 7. Clustering analysis table and soil fertility

Year	The number of samples	Classification number			The percentage of each category in the total sample		
		High	Medium	Low	High	Medium	Low
2005	2259	251	1184	824	11.11 %	52.42 %	36.48 %
2009	5091	891	2701	1499	17.50 %	53.05 %	29.44 %
2012	6312	1473	3527	1312	23.34 %	55.88 %	20.79 %

4 Conclusions

Through the clustering results we analyzed, soil fertility tend towards equilibrium and rise after precision fertilization, so it can reflect the trend of soil fertility better. The results can be seen from Table 7:

(1) We clustering according to the parameters which is set by clustering algorithm. The data is from 2005, 2009 and 2012 and the number of samples are 2259, 5091 and 6312. Then we can derived its cluster classification results are consistent with the actual high fertility, the fertility and low fertility referring soil grading standards. So, it could prove that FUMF algorithm is an effective method to soil fertility evaluation.

(2) The data were compared from 2005, 2009 and 2012, the clustering results show that high fertility soils were increased from 8.16 % to 13.99 % and 15.30 %; The second soil fertility were increased from 53.64 % to 62.80 % and 65.80 %; low soil fertility dropped 22.60 % from 38.10 % to 18.93 % respectively. It is shown that from 2005 to 2012, the low soil fertility decreases and other soil fertility increase after precise fertilization. So, soil fertility has improved significantly.

(3) The analysis results are consistent with the actual situation, it is not only shows FUMF algorithm is an effective method for soil fertility evaluation, but also proved that after precise fertilization soil fertility has improved significantly in general. Therefore, we believe that the method is meaningful by using data mining to analyze fertility factors of large the high complexity and, strong interaction data. So it can provide technical support for precision fertilization decisions.

Acknowledgment. This work was supported by the National "863" High-tech Project (2006AA10A309), Jilin province science and technology development projects(key science and technology research project): "The development and demonstration of corn production's monitoring and traceability system based on Internet of things technology" (20140204045NY), Jilin province agricultural committee projects: "The demonstration and generalization of Corn precision operation system based on Internet of things".

References

1. Zeitouni, K.: A survey of spatial data mining methods databases and statistics point of views. In: Becker, S. (ed.) Data Warehousing and Web Engineering, pp. 229–242. IRM Press, London (2002)
2. Sharma, L.K., Vyas, O.P., Tiwary, U.S., Vyas, R.: A novel approach of multilevel positive and negative association rule mining for spatial databases. In: Perner, P., Imiya, A. (eds.) MLDM 2005. LNCS (LNAI), vol. 3587, pp. 620–629. Springer, Heidelberg (2005). doi:10.1007/11510888_61
3. Li, L., Li, L.: Application of clustering analysis in site classification and soil fertility evaluation. In: Proceedings of 2010 Third International Conference on Education Technology and Training (2010)
4. Guo, X., Liu, X., Li, X.: Improvement and analysis of hierarchical clustering algorithm. Comput. Appl. Softw. **25**(6), 243–244 (2005)
5. Zheng, C., Miao, D., Wang, R.: Improved K-means clustering algorithm based on rough density weighted. Comput. Sci. (2009)
6. Cheng, G., Cao, L., Wang, G.: Application of weighted spatial fuzzy dynamic clustering algorithm in soil fertility evaluation of the black soil zone. Chin. Agric. Sci. **42**(10), 3559–3563 (2009)
7. Zhang, J., Liu, X.: Research and application of K-means algorithm based on clustering Analysis. Appl. Comput. **24**(5), 166–168 (2007)
8. Li, Y., Yin, P.: Speculative multithreading partitioning algorithm based on fuzzy clustering. J. Comput. **37**(3), 580–592 (2014)
9. Yang, Y., Guo, S.: FCM image segmentation algorithm based on kernel function and spatial information. Jilin Univ. (Eng. Sci.) **41**(2), 283–287 (2011)
10. Liu, F., Shi, X., Yu, D.: Research of mapping of soil properties in Taihu based on geostatistics and GIS–In total soil nitrogen in cartography example. J. Soil **41**(1), 20–27 (2004)
11. Zhao, Y., Shi, X., Yu, D.: Discussion of mall-scale spatial variability of soil nutrients and its influencing factors– In urban areas in Wuxi city, Jiangsu province **37**(2), 214–219 (2006)
12. Huang, S., Jin, J., Yang, L.: Research Grain Crop spatial variability of soil nutrients and the partition management technology in Region of County. J. Soil **40**(1), 79–88 (2003)
13. Yang, Y., Shi, X., Yu, D.: Research of region scale soil nutrients spatial variability and influencing factors. Geogr. Sci. **28**(6), 788–792 (2008)
14. Guo, X., Fu, B., Ma, K.: Research on spatial variability of soil fertility based on GIS and geostatistics—case by Hebei Zunhua. J. Appl. Ecol. **11**(4), 557–563 (2000)
15. Sun, Y., Lu, Y.: High-dimensional data flow subspace clustering algorithm based on grid. Comput. Sci., 199–203 (2007)
16. Shan, S., Yan, Y., Zhang, X.: Clustering algorithm based on K most similar clustering subspace. Comput. Eng., 4–6 (2009)
17. Qiu, B., Zheng, Z.: Clustering algorithm based on the local density and dynamically generated mesh. Comput. Eng. **31**(2), 385–387 (2010)