

Application of Fuzzy Clustering Analysis in Classification of Soil in Qinghai and Heilongjiang of China

Ping Han^{1,2}, Jihua Wang^{1,2}, Zhihong Ma^{1,2}, Anxiang Lu^{1,2},
Miao Gao^{1,2}, and Ligang Pan^{1,2,*}

¹ Beijing Research Center for Agrifood Testing and Farmland Monitoring,
Beijing 100097, P.R. China

² National Engineering Research Center for Information Technology in Agriculture,
Beijing 100097, P.R. China

Tel.: +86 -10-51503405; Fax: +86-10-51503406

hanp@nercita.org.cn

Abstract. Soil classification systems organize soil variability into useful groupings that can be identified by field investigation and documented in soil survey, and form the basis for the exchange and extension of soil science research and soil resources management. Fuzzy clustering analysis may be used whenever a composite classification of soil incorporates multiple parameters. In this paper, seventy-seven topsoil samples were collected from Qinghai and Heilongjiang of China, and the element contents of topsoil were detected by wavelength dispersive X-ray fluorescence spectroscopy. In fuzzy clustering analysis, all data were standardized, and then a fuzzy similarity matrix was established and the fuzzy relation was stabilized. The results showed that topsoil samples of Qinghai and Heilongjiang were completely grouped into two clusters according to their districts, when given a suitable threshold $\lambda = 0.7580$. This work supplied the quantification classification method of alpine soil (Qinghai) and unsaturation siallitic soil (Heilongjiang).

Keywords: fuzzy clustering analysis, soil classification, elements.

1 Introduction

Soils are the complicated natural bodies and soil system acts as a component for various ecological functions. Soil classification is the process of grouping soil individuals into more or less homogeneous groups with respect to defined objectives [1], thereby highlighting the essential differences in soil properties and functions between classes [2]. Soil classification systems organize soil variability into useful groupings that can be identified by field investigation and documented in soil survey activities to promote effective resource management and technology transfer [3]. Clustering is useful and

* Corresponding author.

plays a key role in searching for structures in data. Each of these structures is called a cluster or class [4]. Cluster analysis is developed in taxonomy and its principal aim is to partition multivariate observations into a number of meaningful multivariate homogeneous groups [5]. Fuzzy clustering analysis method is based on the fuzzy set theory [6] and is one of the most important methods of unsupervised learning and has significant advantages over traditional clustering [7]. Feng Lixiao et al(1992)[8] selected active acid, substitution acid, hydrolysis acid, active aluminum, cation exchange capacity (CEC) and degree of base saturation as the parameters of soil fuzzy clustering analysis to distinguish yellow brown soil and yellow cinnamon soil. Wu Kening et al (1994) [9] studied fuzzy clustering analysis method in soils of transition regions of northern Subtropics in China. The results implicated that fuzzy clustering analysis was in accordance with pedogenesis classification and diagnostic classification. For the development of precision agriculture, fuzzy cluster classification performed to delineate management zones [10, 11].

This article explores the method of soil classification which applied fuzzy clustering analysis. we selected the element contents, especially metal element contents as the soil parameters to do fuzzy clustering analysis and discussed the relationship between element concentration and soil pedogenesis. This would extend soil science and supply classification method for soils which have different soil form processes.

2 Materials and Methods

2.1 Samples

Seventy-seven topsoil samples (0-20cm) were collected from Qinghai and Heilongjiang provinces of China. In these samples, thirty-nine topsoil samples belonged to Qinghai, which were located between longitudes 93.652°E and 95.768°E and latitudes 36.387°N and 36.793°N with an area of about 0.084 km², and, thirty-eight samples belonged to Heilongjiang, which were located between longitudes 131.570°E and 133.304°E and latitudes 46.399°N and 47.605°N with an area of about 0.031 km². At each sampling site, 5 sub-samples were taken from the 4 vertexes and the center of a square block (10m×10m) and mixed thoroughly to select 0.5 kg soil as the representative sample of site. All samples came from farmland.

2.2 Soil Samples Preparation and Measurements

The samples were air-dried, ground, passed 250mm nylon sieve. Four grams of soil powder were pressed into pellet at 10 tones using a manual hydraulic press (pellet diameter = 32 mm). The concentrations of Cu, Pb, Zn, Cr, Ni, Fe, Mn, Rb, Sr, V, MgO, CaO, Na₂O, F, S, Cl, Ce, Ba, Co, Ga, Zr, La, Al₂O₃, TiO₂ and As in soil pellets were analyzed by wavelength dispersive X-ray fluorescence spectroscopy (TW2404, PHILIPS Company).

2.3 Fuzzy Clustering Analysis

All data except content value of Mn, x_{ij} were standardized by formula (a). i and j represented soil sample ID (1, 2, ..., 77) and soil parameters ID (1, 2, ..., 24).

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (a)$$

Note: In formula (a), x'_{ij} was standardization value.

$$\bar{x}_j = \frac{1}{77} \sum_1^{77} x_{ij} \quad (h) \quad s_j = \sqrt{\frac{\sum_1^{77} (x_{ij} - \bar{x}_j)^2}{77 - 1}} \quad (i)$$

According to the standardization data, the relation r'_{ik} between i^{th} and k^{th} objects of classification was calculated by formula (b). Then, the similar matrix R' was established and described as equation (c).

The fuzzy similar matrix R was established via the r_{ik} which was set by the relation r'_{ik} on the interval $[0, 1]$ by ordering $r_{ik} = 0.5 + 0.5 \times r'_{ik}$ and was described as equation (d). The fuzzy similar matrix R was not a stabilized one. That is, it met the reflexivity [12] and symmetry [13], but not transitivity [13, 14]. The fuzzy similar matrix R will have to be changed into the fuzzy equivalent matrix via self-squared method when clustering [15]. We transformed R into R^* using formula (e): $r_{ij} = \bigvee (r_{ik} \wedge r_{jk}) = (R \times R)_{ij}$ (e). If $R^* = R^k = R^{2k}$ (f), then the R^* had become a stabilized fuzzy relation. In this case the R^* was stable at $R^4 = R^8$. And R^* was described as equation (g).

$$r'_{ik} = \frac{\sum_j^n x_{ij} x_{kj}}{\sqrt{\sum_{j=1}^n x^2_{ij} \sum_{j=1}^n x^2_{kj}}} \quad (b) \quad R' = \begin{pmatrix} r'_{1,1} & r'_{1,2} & \dots & \dots & r'_{1,77} \\ r'_{2,1} & r'_{2,2} & \dots & \dots & r'_{2,77} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r'_{77,1} & r'_{77,2} & \dots & \dots & r'_{77,77} \end{pmatrix} \quad (c)$$

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & \dots & \dots & r_{1,77} \\ r_{2,1} & r_{2,2} & \dots & \dots & r_{2,77} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{77,1} & r_{77,2} & \dots & \dots & r_{77,77} \end{pmatrix} \quad (d) \quad R^* = \begin{pmatrix} r^*_{1,1} & r^*_{1,2} & \dots & \dots & r^*_{1,77} \\ r^*_{2,1} & r^*_{2,2} & \dots & \dots & r^*_{2,77} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r^*_{77,1} & r^*_{77,2} & \dots & \dots & r^*_{77,77} \end{pmatrix} \quad (g)$$

The R^* value embodied the similarity and classification probability of all soil samples. If the value was 0, then the two soil samples did not have similarity as different classifications. If the value was 1, then the two soil samples had very similarity as the same classification. If the value was on the interval $[0, 1]$, then the value indicated the probability that the two soil samples could be classified into the same clusters. λ was set as probability threshold. If λ was greater than r^*_{ij} , then r^*_{ij} was on the order of 1. If λ was less than r^*_{ij} , then r^*_{ij} was on the order of 0. We assigned a value from 1 to 0 to λ , reduced the value gradually, and according to the same λ value, divided some soil samples into same class as r^*_{ij} which was 1 or 0. The algorithms repeated the above steps at different λ value until no soil samples were divided into the same class.

2.4 Data Analysis

All data were analyzed by Office Excel 2003, SPSS 18.0, and MATLAB 7.0.

3 Results and Discussion

3.1 T-Test of the Data

Table 1 showed the statistical results of the 25 metal concentrations in the topsoil of Qinghai and Heilongjiang. All data were analyzed by independent samples t-test analysis in SPSS 18.0 software. T-values of metal contents as soil parameters showed significant differences or very significant differences except Mn between Qinghai and Heilongjiang.

Table 1. Data analysis results of elements in topsoil of Qinghai and Heilongjiang ^a

Parameters ^b	Sampling sites ^c	N	Mean	Std. Deviation	Std. Error Mean	t
Cu	HLJ	38	22.31	3.95	0.64	4.332**
	QH	39	17.80	5.09	0.81	
Pb	HLJ	38	22.09	3.52	0.57	7.318**
	QH	39	13.59	6.31	1.01	
Zn	HLJ	38	61.81	15.77	2.56	2.316*
	QH	39	54.82	9.97	1.60	
Cr	HLJ	38	63.61	9.42	1.53	5.594**
	QH	39	52.17	8.51	1.36	
Ni	HLJ	38	47.19	7.61	1.24	2.476*
	QH	39	43.77	3.85	0.62	
Fe	HLJ	38	3.32	0.78	0.13	2.163*
	QH	39	3.04	0.19	0.03	
Rb	HLJ	38	102.10	10.13	1.64	14.980**
	QH	39	60.77	13.84	2.22	

Table 1. (continued)

Sr	HLJ	38	177.58	34.55	5.61	-4.772**
	QH	39	216.10	36.23	5.80	
V	HLJ	38	86.11	10.97	1.78	10.135**
	QH	39	62.05	9.84	1.58	
MgO	HLJ	38	1.21	0.43	0.07	-17.268**
	QH	39	3.05	0.50	0.08	
CaO	HLJ	38	1.89	1.30	0.21	-30.037**
	QH	39	9.44	0.86	0.14	
Na ₂ O	HLJ	38	1.27	0.20	0.03	-9.243**
	QH	39	1.65	0.15	0.02	
F	HLJ	38	435.39	134.00	21.74	-11.441**
	QH	39	742.31	98.18	15.72	
S	HLJ	38	402.11	141.35	22.93	-5.076**
	QH	39	811.10	482.37	77.24	
Mn	HLJ	38	580.23	276.38	44.84	-0.719
	QH	39	612.82	41.74	6.68	
Cl	HLJ	38	65.49	19.64	3.19	-4.946**
	QH	39	923.28	1082.80	173.39	
Ce	HLJ	38	82.10	16.99	2.76	8.194**
	QH	39	52.89	14.20	2.27	
Ba	HLJ	38	631.13	36.12	5.86	27.985**
	QH	39	445.21	19.56	3.13	
Co	HLJ	38	12.23	2.63	0.43	4.332**
	QH	39	9.97	1.86	0.30	
Ga	HLJ	38	17.95	2.75	0.45	6.751**
	QH	39	13.67	2.81	0.45	
Zr	HLJ	38	260.29	39.24	6.37	12.713**
	QH	39	166.28	23.50	3.76	
La	HLJ	38	39.82	18.01	2.92	4.005**
	QH	39	25.86	11.87	1.90	
Al ₂ O ₃	HLJ	38	13.54	1.24	0.20	9.195**
	QH	39	11.36	0.78	0.12	
TiO ₂	HLJ	38	0.84	0.027	0.004	42.026**
	QH	39	0.59	0.025	0.004	
As	HLJ	38	9.16	2.59	0.42	-4.369**
	QH	39	11.65	2.41	0.39	

^a Independent samples T-test analysis were used. * $p < 0.05$, * * $p < 0.01$;

^b The concentration units of Fe, MgO, CaO, Na₂O, Al₂O₃, TiO₂ were percentage, others were mg/kg;

^c HLJ and QH represented Heilongjiang and Qinghai respectively.

3.2 Fuzzy Clustering

The fuzzy clustering graphic was made by MATLAB7.0 soft (Fig.1).

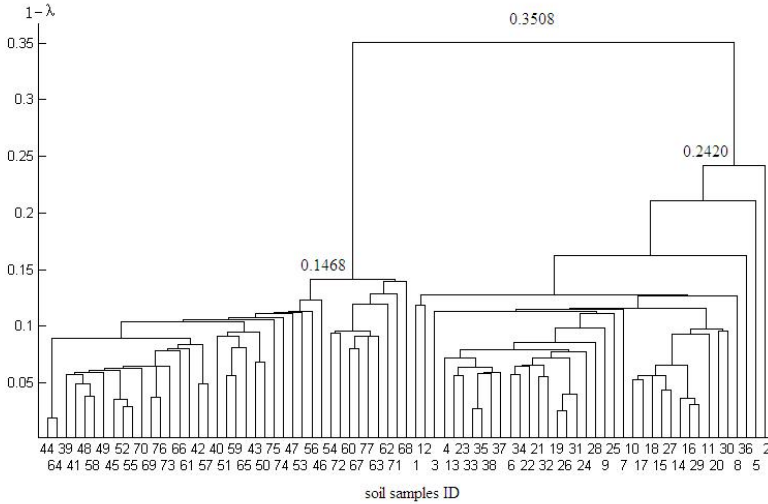


Fig. 1. The fuzzy clustering graphic

Note: Soil samples ID from 1-38 represent the soil of Heilongjiang, and soil sample ID from 39-77 represent the soil of Qinghai.

The fuzzy clustering graphic showed that the topsoil samples of Qinghai and Heilongjiang were completely grouped into two clusters according to their districts when given probability threshold $\lambda = 0.7580$. According to the soil order of China, the soil of Qinghai and Heilongjiang belonged to alpine soil and un-saturation sillic soil respectively [16]. So alpine soil and un-saturation sillic soil could be divided into two classes based on their metal content. The topsoil samples of Qinghai and Heilongjiang were regarded as the same class when given probability threshold $\lambda = 0.6492$. It implied that topsoil of Qinghai and Heilongjiang had 64.9% similarity based on their 24 species of metal concentration. When given probability threshold $\lambda = 0.7580$ and $\lambda = 0.8532$, the topsoil soil samples of Heilongjiang and the topsoil soil samples of Qinghai were divided into the same class respectively. The results showed that similarity of the 39 topsoil samples of Qinghai was greater than the 38 topsoil samples of Heilongjiang, that is, with respect to total homogeneity based on 24 species of metal concentration, Qinghai was better than Heilongjiang. Maybe this result was related to the topsoil sample sites. All topsoil samples of Qinghai were collected from the Golmud region, as well as, the topsoil samples of Heilongjiang were collected from the Tongjiang region and the Shuangyashan region. On the other hand, this difference could come from soil pedogenesis of the two districts.

4 Conclusions

The metal element content of soil can be used for qualitative classification. For this purpose, the 77 soil samples from Qinghai and Heilongjiang of China have been analyzed to determine their Cu, Pb, Zn, Cr, Ni, Fe, Mn, Rb, Sr, V, MgO, CaO, Na₂O, F, S, Cl, Ce, Ba, Co, Ga, Zr, La, Al₂O₃, TiO₂ and As contents. T-test analysis was conducted on all soil metal element concentration. By way of T-test results, all metal element contents except Mn were used as parameters in the fuzzy clustering analysis. Fuzzy classification algorithms of the soil samples based on the 24 element concentrations allowed for an objective interpretation of their similarities and differences. The results show that this fuzzy clustering method can be applied in qualitative classification of the soil of Qinghai and Heilongjiang. This method helps develop the qualitative classification of soil, extend soil science, and is propitious to farmland management.

Acknowledgments

The financial support of National High Technology Research and Development Program 863 (2010AA10Z403), (2007AA10Z202), and Beijing Municipal Science and Technology Commission Program (Z09090501040901). Thanks are due to Heilongjiang Academy of Agricultural Science and Golmud Bureau of Agriculture and Animal Husbandry of Qinghai Province for collecting topsoil samples.

References

- [1] Cline, M.G.: Basic principles of soils classification. *Soil Science* 67(2), 82–91 (1949)
- [2] Rossiter, D.G.: Classification of urban and industrial soils in the word reference base for soil resources. *J. Soils Sediments* 7(2), 96–100 (2007)
- [3] Effland, W.R., Pouyat, R.V.: The genesis, classification, and mapping of soils in urban areas. *Urban Ecosystems* 1, 217–228 (1997)
- [4] Sarbu, C., Einax, J.W.: Study of traffic-emitted lead pollution of soil and plants using different fuzzy clustering algorithms. *Anal. Bioanal. Chem.* 390, 1293–1301 (2008)
- [5] Templ, M., Filzmoser, P., Reimann, C.: Cluster analysis applied to regional geochemical data: problems and possibilities. *Applied Geochemistry* 23, 2198–2213 (2008)
- [6] Zadeh, L.A.: Fuzzy sets. *Information and Control* 8, 338–353 (1965)
- [7] Liu, L., Zhou, J.Z., An, X.L., Li, Y.H., Liu, Q.: Improved fuzzy clustering method based on entropy coefficient and its application. In: Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W. (eds.) *ISNN 2008*, Part II. LNCS, vol. 5264, pp. 11–20. Springer, Heidelberg (2008)
- [8] Feng, L.X., Xiao, J.Z., Guo, W.X.: Application of fuzzy clustering method on the classification of yellow brown soil and yellow cinnamon soil in southern Shaanxi. *Chinese Journal of Soil Science* 23(3), 108–110 (1992)
- [9] Wu, K.N., Kang, C.: Fuzzy cluster analysis of soils in transition regions of northern Subtropics in China. *Tropical and Subtropical Soil Science* 3(3), 163–168 (1994)
- [10] Moral, F.J., Terrá, J.M., Marques da Silva, J.R.: Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research* 106, 335–343 (2010)

- [11] Li, Y., Shi, Z., Wu, C.F., Li, H.Y., Li, F.: Determination of potential management zones from soil electrical conductivity, yield and crop data. *Journal of Zhejiang University: Science B* 9(1), 68–76 (2008)
- [12] Zadeh, L.A.: Similarity relations and fuzzy orderings. *Information science* 3, 177–206 (1971)
- [13] Zimmerman, H.J.: *Fuzzy set theory and its applications*. Kluwer Nijhoff Publishing, Norwell (1985)
- [14] Kung, H.T., Ying, L.G., Liu, Y.C.: Fuzzy clustering analysis in environmental impact assessment- a complement tool to environmental quality index. *Environment Monitoring and Assessment* 28, 1–14 (1993)
- [15] Wang, S.L., Wang, X.Z.: A fuzzy comprehensive clustering method. In: Alhajj, R., Gao, H., Li, X., Li, J., Zaïfane, O.R. (eds.) *ADMA 2007*. LNCS (LNAI), vol. 4632, pp. 488–499. Springer, Heidelberg (2007)
- [16] Wei, F.S., Chen, J.S., Wu, Y.Y., Zheng, C.J.: Study on soil environmental background values of China. *Environmental Science* 12(4), 12–19 (1991)