

土壤模糊隶属度不同数据转换方法及其对空间插值结果的影响

檀满枝¹, 陈杰^{2,1,*}

(1. 土壤与农业可持续发展国家重点实验室 中国科学院南京土壤研究所, 南京 210008;
2. 郑州大学自然资源与生态环境研究所, 郑州 450001)

摘要:应用模糊 c-均值算法对土壤进行连续分类时,其输出的土壤模糊隶属度值具有成分数据的结构特点。直接基于土壤隶属度数据实施普通克里格插值,其空间预测结果缺乏可信度。因此,在进行插值预测之前,必须对土壤模糊隶属度值进行必要的数据转换。研究采用对数正态变换方法、对称对数比转换方法和非对称对数比转换方法对土壤模糊隶属度值进行数据转换,分析了各种数据转换形式对插值结果及其精度的影响。结果表明,对样点土壤模糊隶属度进行简单对数正态转换,其插值结果空间上任意点的土壤对于不同类别的隶属度之和均不为1,因此这样的插值结果理论上缺乏可行性。数据经非对称对数比转换和对称对数比转换后,插值结果均满足各个位置组分之和为1和非负限制,二者相比,后者对区域总体趋势的反映较前者好,且精度较高。因此,在应用对称对数比方法对样点土壤模糊隶属度值进行数据转换的基础上,应用克里格技术实施空间插值可以获得最佳预测结果。

关键词:土壤; 模糊隶属度; 成分数据; 非对称对数比转换; 对称对数比转换; 空间预测

文章编号:1000-0933(2009)06-3147-07 中图分类号:O159,S159,S181 文献标识码:A

Different transforms of fuzzy membership values of sampled soils and their influences on resulted interpolation prediction

TAN Man-Zhi¹, CHEN Jie^{2,1,*}

1 State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China

2 Institute of Natural Resource and Eco-environment, Zhengzhou University, Zhengzhou 450001, China

Acta Ecologica Sinica, 2009, 29(6): 3147 ~ 3153.

Abstract: Soil fuzzy membership values of the sampled soils resulted from fuzzy c-means algorithm, which would be applied in soil predictive mapping, is a kind of compositional data. Owing to the structural characteristics of compositional data, they could not be directly used in prediction of the fuzzy memberships of unknown sites over space by kriging interpolation. To achieve spatial soil prediction, therefore, the membership values of the sampled soils must be transformed before interpolation. In this study, transform of compositional data by several ways were attempted, and influence of different transform approaches on output and precision of prediction compared and analyzed. The results indicated that, membership values of all the spatial predicted sites didn't sum to 1 on condition that known membership values of the sampled soils were simply transformed through logarithm. Obviously, the above predictive result was theoretically unauthentic. Contrarily, membership values of all the spatial predicted sites summed to 1 when the membership values of the known soils were transformed by asymmetry Logratio and symmetry Logratio approaches, demonstrating that two approaches were theoretically accepted. Comparatively, symmetry Logratio transform could lead to a better spatial distribution pattern and higher precision.

基金项目:国家自然科学基金资助项目(40571065, 40701070); 中国科学院南京土壤研究所创新前沿资助项目(ISSASIP0716)

收稿日期:2008-03-04; 修订日期:2008-04-21

致谢:中国科学院南京土壤研究所杜国华研究员在野外调查工作中给予指导和帮助,在此表示感谢。

* 通讯作者 Corresponding author. E-mail: jchen@zzu.edu.cn

Key Words: soil; fuzzy membership; compositional data; asymmetry Logratio transform; symmetry Logratio transform; soil predictive mapping

20世纪80年代末期,模糊逻辑开始应用于土壤预测制图领域。在模糊逻辑体系中,土壤样本表现出对于不同土壤类型的多重相似性(multiple similarity)和部分隶属关系(partial membership)。运行模糊算法模型,可获得已知样点的土壤样本属于不同土壤类别的模糊隶属度值^[1~5]。需要指出的是,样点土壤属于不同土壤类别的隶属度之和为1,且属于每一土壤类别的隶属度值均为非负^[6]。从结构上看,样点土壤隶属度值是一种典型的成分数据(compositional data)。基于离散样点土壤模糊隶属度,应用地统计学手段进行空间插值分析是实现土壤连续预测制图的必需程序。但由于成分数据的特殊性,插值分析必须同时满足以下四个条件:每一空间位置上各个组分插值结果为非负,且之和为常数,误差最小化和无偏估计^[7]。在土壤学领域,土壤机械组成同样是一种成分数据,土壤中各粒径颗粒组成之和为100%,且每一组分含量不小于0^[6,8]。国外的相关研究表明,在应用克里格插值手段进行空间预测时,对样点土壤机械组成数据进行对称对数比转换(symmetry Logratio transform)比直接基于原始数据插值预测获得的输出结果更为合理,且精度更高^[7]。而国内一些土壤机械组成空间预测研究中,成分数据的定和与非负限制特性常被忽略,因此预测结果的精度与可信度有待商榷^[9,10]。国外基于样点土壤模糊隶属关系的土壤连续预测制图过程中,除个别案例外^[11],一般都对样点土壤的模糊隶属度值进行必要转换或者对插值条件进行限定^[12~14]。对土壤机械组成、土壤模糊隶属度这类成分数据实施转换的方法有多种,但迄今为止尚没有人系统分析和比较各种方法及其对空间预测结果的影响。本研究以案例地区已获得的土壤模糊连续分类结果,应用不同方法对土壤模糊隶属度值进行数据转换,并对相应的土壤空间预测精度进行评价,以期阐释成分数据不同转换方法的优势和不足,为涉及到成分数据的土壤空间预测研究提供借鉴。

1 材料与方法

1.1 研究区概况与样品采集

研究区位于江苏省南京市东郊麒麟镇东流村附近,面积约为1 km²。北纬32°04'30"~32°05'01",东经118°57'00"~118°58'30"。在对研究区进行系统野外调查的基础上,在不同母质来源、地形部分和土地利用方式下开挖土壤剖面31个,深度为120cm(或至基岩)。对土壤剖面形态特征进行观察、描述与记录,并分层采集土壤样品,同时记载样点的地理位置及其周围的景观信息。结合土壤样品实验室分析数据,在对研究区31个土壤剖面主要发生层进行细分、归整的基础上,划分出9个具有重要土壤发生学意义与分类典型性的特征土层(i水耕表层、ii水耕氧化还原层、iii淡薄表层、iv黄黏层、v黏化层、vi黏盘层、vii弱风化B层、viii砂石表层、ix砂石层)^[15]。随后对研究区进行土钻加密采样,钻取样点共计85个,对照前期确立的特征土层,记录下每个土钻样中相应的特征土层的类型和厚度,同时记录下样点位置及发育的地形部位、母质来源以及利用方式等信息(图1)。

1.2 土壤模糊连续分类

基于研究区116个样点土壤中发育的9类特征土层的厚度数据,建立样点对应特征土层厚度的数据矩阵,对样点土壤缺失的特征土层,其厚度值用“0”表示。确定最佳分类参数后,应用模糊c-均值算法(FCM:fuzzy c-means algorithm)对样点土壤进行模糊聚类,获得4个模糊类别。每一类别的质心值(类别对应每一种特征土层厚度的中心值)和样点土壤属于每一类别的模糊隶属度值分别见表1、2。类别A的代表性土壤,其剖面中主要为砂石表层和砂石层发育,类别B的代表性剖面则为淡薄表层、黏化层和黏盘层组合,类别C剖面为水耕表层和水耕氧化还原层组合,类别D剖面则主要为淡薄表层和弱风化B层组合。参照剖面形态特征及分层样品分析数据,A、B、C、D4个模糊聚类类别分别大致对应于“中国土壤系统分类”中新成土、淋溶土、水耕人为土和雏形土4个土纲下的土壤类型。

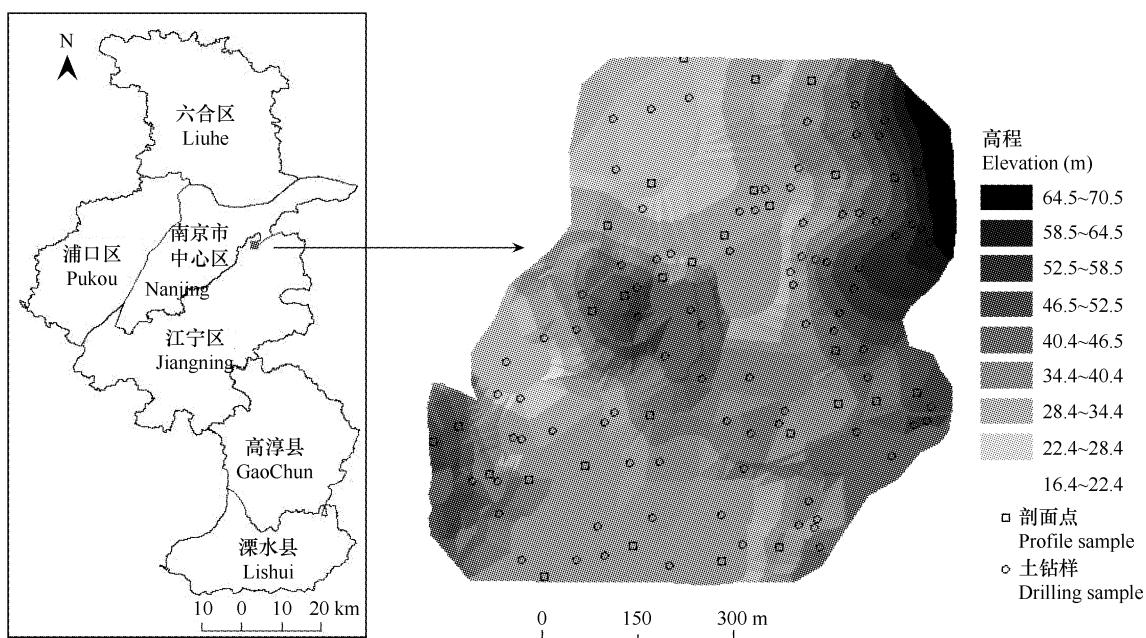


图1 研究区位置及样点分布图

Fig. 1 Location of study area and distribution of sampled sites

表1 4种模糊类别对应的特征土层厚度质心值

Table 1 Centroid values of different characterizing horizons (cm) in the resulted fuzzy clusters

类别 Class	i	ii	iii	iv	v	vi	vii	viii	ix
A	0	0	0.84	0.85	0.07	0.53	0.51	8.25	25.65
B	0	0	18.22	1.57	29.42	70.76	0.68	0.06	0.19
C	24.22	83.30	0.43	0.44	0.55	0.28	0.26	0.04	0.15
D	0	0	20.62	0.14	0.19	0.10	85.18	0.01	0.05

1.3 研究方法

1.3.1 非对称对数比转换

由于成分数据因组分相关性中蕴涵由闭合效应引起的伪相关,而且不服从正态分布,给统计带来困难。Aitchison 于 20 世纪 80 年代提出成分数据的对数比转换方法,同时解决了成分数据统计分析中的闭合效应和统计分布两个方面的问题^[16,17]。Pawlowsky 等将对数比方法与地质统计学方法相结合,提出了区域化成分数据的统计方法^[18]。常用的对数比转换又称为非对称对数比转换(ALR: asymmetry Logratio transform)或者称为加和对数比转换(additive logratio transform)^[3, 7, 12]。具体转换公式如下:

$$\mu'_{ij}(x) = \ln \frac{\mu_{ij}(x)}{\left(\prod_{j=1}^c \mu_{ij}(x) \right)^{1/c}} \quad (1)$$

转回公式为:

$$\mu_{ij}(x) = \frac{\exp \mu'_{ij}(x)}{\sum_{j=1}^c \exp \mu'_{ij}(x)} \quad (2)$$

式中, $\mu_{ij}(x)$ 为第 i 个样点土壤对于第 j 个模糊聚类类别的隶属度, $\mu'_{ij}(x)$ 为第 i 个样点土壤对于第 j 个聚类类别之隶属度的对称对数比转换值。

1.3.2 对称对数比转换

对称对数比转换(SLR: symmetry Logratio transform)或称为改进的加和对数比转换(modified additive

logratio transform)^[3,7,12]。转换公式为:

$$\mu'_{ij}(x) = \ln \frac{\mu_{ij}(x) + \eta}{\left(\prod_{j=1}^c (\mu_{ij}(x) + \eta) \right)^{1/c}} \quad (3)$$

转回公式为:

$$\mu_{ij}(x) = \left(\frac{\exp \mu'_{ij}(x)}{\sum_{j=1}^c \exp \mu'_{ij}} - \frac{\eta}{1 + \sum_{j=1}^c \eta} \right) \left(1 + \sum_{j=1}^c \eta \right) \quad (4)$$

其中, $\mu_{ij}(x)$ 为第 i 个样点土壤对于第 j 个模糊聚类类别的隶属度, $\mu'_{ij}(x)$ 为第 i 个样点土壤对于第 j 个聚类类别之隶属度的对称对数比转换值, η 为常数, 取研究区除 0 外最小隶属度值的一半。

本研究半变异函数分析和普通克里格插值以及各种公式的栅格计算均在 ArcGIS 环境中完成。

2 结果与讨论

2.1 研究区土壤类型模糊隶属度值及数据转换结果

应用模糊 c-均值算法(FCM), 得到样点土壤对于各聚类类别的模糊隶属度值。具体数据及转换结果见表 2。

表 2 样点土壤对于各聚类类别的隶属度值及其对称对数比、非对称对数比转换值

Table 2 The memberships of sampled soils to the four fuzzy resulted clusters and their transformed data by different ways

ID	A	B	C	D	SLR-A	SLR-B	SLR-C	SLR-D	ALR-A	ALR-B	ALR-C	ALR-D
p1	0.018	0.034	0.018	0.930	-1.110	-0.525	-1.157	2.792	0.330	0.592	0.314	16.321
p2	0.006	0.009	0.005	0.980	-1.419	-0.883	-1.463	3.765	0.242	0.414	0.231	43.244
p3	0.001	0.001	0.001	0.997	-1.849	-1.411	-1.889	5.150	0.157	0.244	0.150	173.881
p4	0.021	0.038	0.020	0.921	-1.086	-0.494	-1.127	2.707	0.338	0.610	0.324	14.984
p5	0.001	0.002	0.001	0.996	-1.758	-1.304	-1.792	4.853	0.172	0.272	0.166	128.867
p6	0.046	0.059	0.043	0.852	-0.781	-0.524	-0.834	2.140	0.458	0.592	0.434	8.497
p7	0.031	0.041	0.029	0.899	-0.906	-0.608	-0.958	2.472	0.404	0.545	0.383	11.848
...

SLR: 对称对数比转换 Symmetry logratio transform; ALR: 非对称对数比转换 Asymmetry logratio transform

2.2 原始数据和经转换后数据分布形态比较

原始土壤隶属度数据不符合正态分布, 但除类别 D 外, 其它 3 个类别的隶属度值符合对数正态分布。图 2 显示的是经过非对称对数比、对称对数比转换后的数据形态分布, 可以看出对称对数比转换比非对称对数比转换更近似正态分布, 采用对称对数比转换后的数据进行克里格插值可望获得较为理想的空间预测结果。

2.3 各种数据转换方法空间预测精度比较

首先基于原始土壤模糊隶属度值及其经对数正态转换的数据, 应用普通克里格插值方法对研究区土壤进行了空间预测。将获得的 4 个单一类别隶属度栅格图进行栅格加和, 结果发现栅格隶属度之和均不为 1(图 3-I), 不能满足属于每一类别的隶属度之和为 1 的要求。采用对数正态转换(图 3-II), 4 种类别的隶属度插值栅格图加和结果显示, 13% 的栅格隶属度之和大于 1、87% 的栅格隶属度之和小于 1, 没有一个栅格隶属度之和等于 1 的情况出现。最大之和达到 7.25, 最小之和仅为 0.02, 栅格隶属度之和平均值为 0.59。且土壤对于单个类别 C 和 D 的隶属度值都有大于 1 的情况出现, 最大值达 7.24。显然, 这样的插值结果没有可信基础。说明不考虑成分数据的特殊性, 对原始数据直接或经简单对数正态转换后进行插值, 其预测结果将蕴涵必然的不确定性和不可靠性。尽管基于原始数据和简单对数正态变换进行克里格插值的平均预测误差和均方根误差都较小(表 3), 按理来说这应该是理想的预测结果, 但实际证明对原始数据和简单对数正态变换进行插值不可靠, 这点同时突出说明了依靠一种方法进行预测精度验证是不可靠的问题。

基于非对称对数比转换后的数据进行克里格插值预测, 生成的单一类别土壤隶属度图中每一栅格单元的

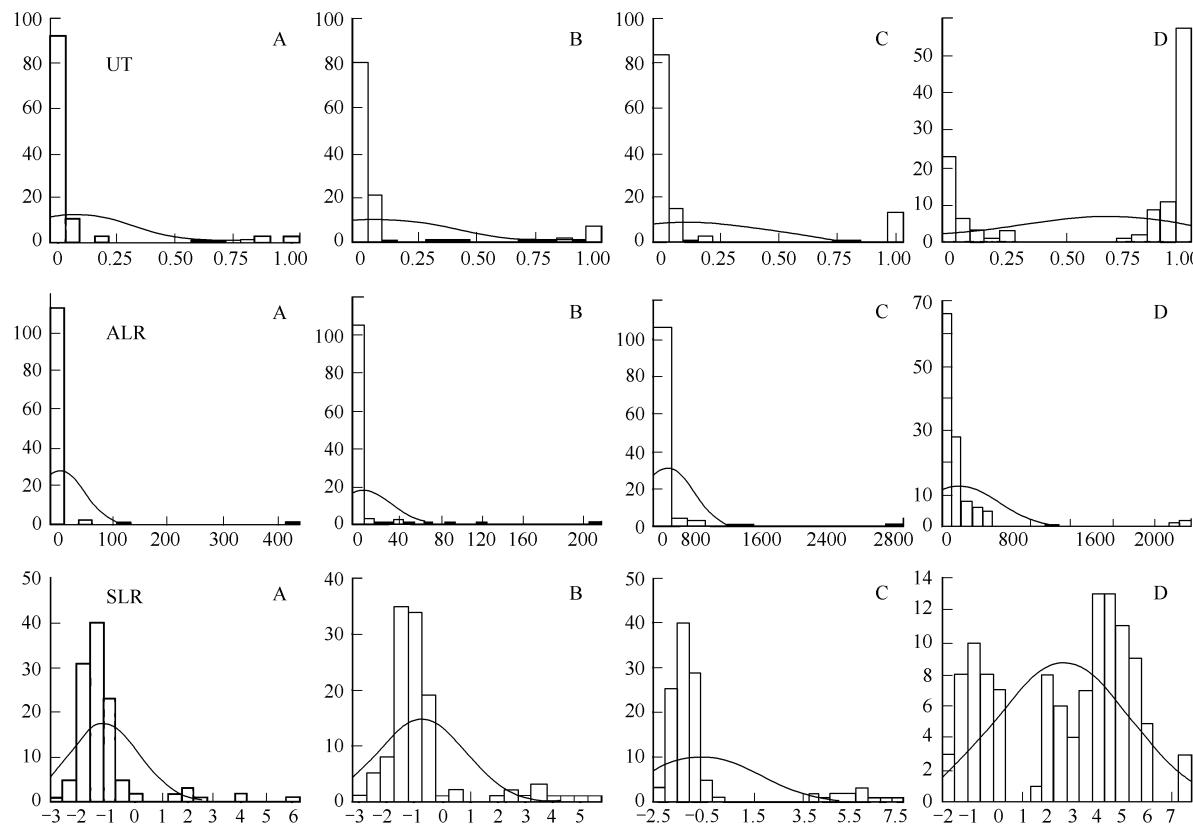


图2 原始土壤隶属度(UT)及其转换数据分布形态

Fig. 2 Distributions of soil memberships and their transformed data

ALR:非对称对数比转换 Asymmetry logratio transform; SLR:对称对数比转换 Symmetry logratio transform

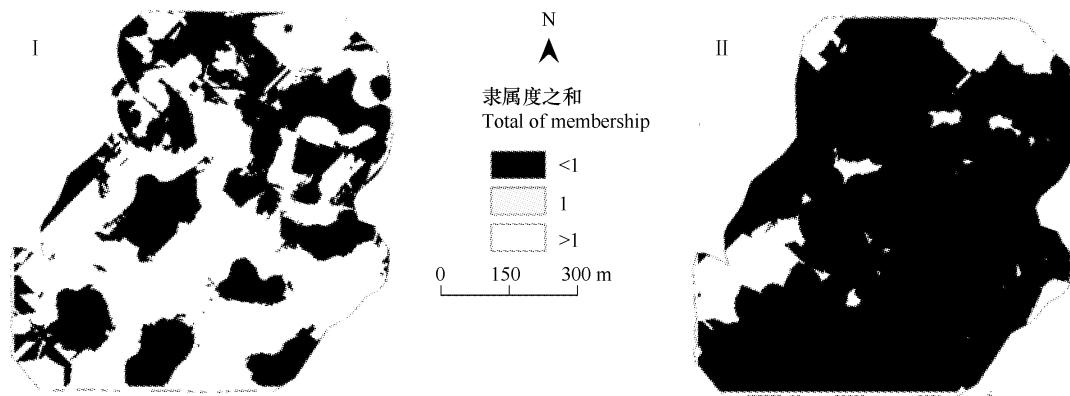


图3 4种模糊类别土壤隶属度加和栅格图

Fig. 3 Grid map of addition of four soil memberships based

I 基于原始数据 Original data, II 基于对数转换数据 Logarithm transformed data

隶属度之和虽然均为1,但预测结果精度低,其预测精度评估参数:均方根预测误差平均值为180.22,最高的达到359.4(表3)。另外,4个单一类别土壤隶属度预测结果图中,空间上土壤类别隶属度连续渐变过渡特征不明显,4个类别的土壤隶属度栅格图显示隶属度基本在0.1以下和0.9以上两个区间范围取值,这种分布格局明显不合理(图4-I)。

对基于对称对数比转换后的数据插值生成的单一类别土壤隶属度图进行栅格加和,其结果均为1,且栅格样点对于每一类别的隶属度值都在0~1之间,满足成分数据的定和和非负条件。同时空间预测精度较高,

平均预测误差接近0,均方根预测误差平均值为1.66。总之,采用对称对数比方法对原始土壤隶属度值进行转换,不仅可以满足成分数据克里格插值的定和和非负限制,同时预测结果也较为精确。此外,其输出结果单一类别隶属度图体现了土壤连续分类的空间渐变过渡特征,对区域总体趋势的反映较非对称对数比转换后插值结果好(图4-II)。

表3 4个模糊类别的土壤隶属度值未转换和经转换后预测结果的精度评价

Table 3 Precision evaluation of Kriging interpolation using untransformed and transformed data from soil membership values to four fuzzy clusters

类别 Class	均方根误差 RMSE				平均预测误差 ME			
	原始 UT	自然对数 ln	对称对数比 SLR	非对称对数比 ALR	原始 UT	自然对数 ln	对称对数比 SLR	非对称对数比 ALR
A	0.168	1.952	1.017	42.77	-0.001	0.024	0.001	-0.105
B	0.261	2.292	1.484	27.31	0.001	0.026	0.002	0.041
C	0.254	2.335	1.863	291.40	-0.010	-0.053	-0.068	-3.656
D	0.373	2.196	2.268	359.40	0.010	0.080	0.050	2.615
平均 Mean	0.264	2.194	1.658	180.22	0	0.019	-0.004	-0.276

UT: Untransformed; ln: logarithm transformed; SLR: Membership values transformed by symmetry log-ratio; ALR: Membership values transformed by asymmetry log-ratio

2.4 隶属度空间预测结果

表4为基于对称对数比转换后的隶属度半变异函数分析结果,每个模糊土壤类别经最佳拟合模型和最佳预测参数,进行普通克里格插值,得到4种类别的土壤模糊连续分布图(图4-II)。对比结果显示,原始土壤隶属度数据经对称对数比方法转换后,应用普通克里格插值获得的单一类别土壤模糊连续分类图具有最好的表达效果。

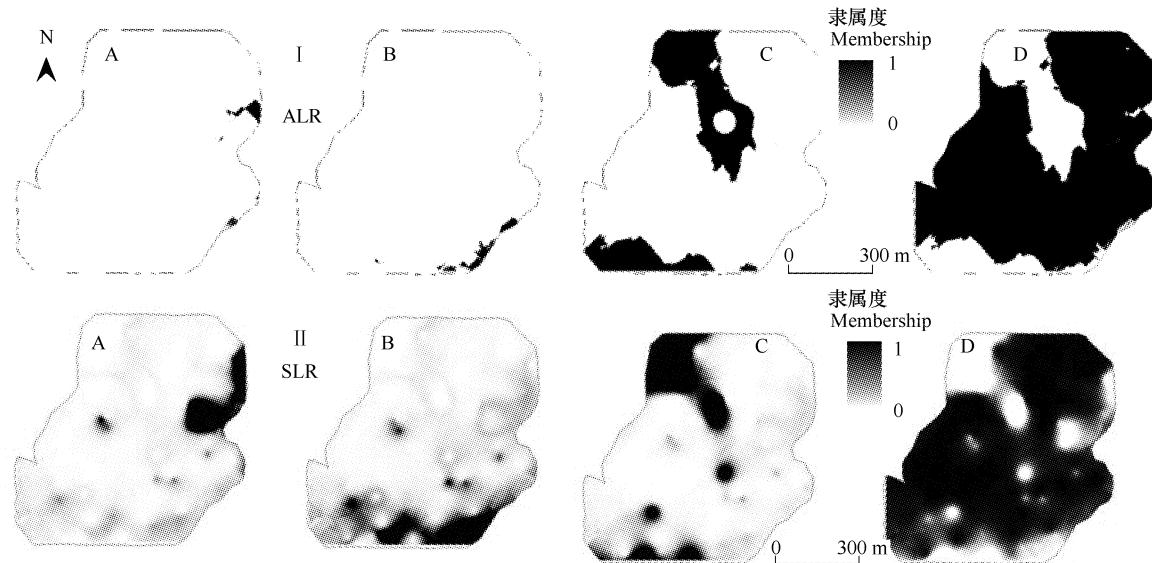


图4 单一类别土壤类型连续分布图

Fig. 4 Predictive mapping of membership of soils to different fuzzy clusters

ALR: 数据经非对称对数比转换; SLR: 数据经对称对数比转换 ALR: Membership values transformed by asymmetry log-ratio; SLR: Membership values transformed by symmetry log-ratio

3 结论

(1) 应用模糊逻辑及其相关算法获得的样点土壤模糊隶属度值是一组典型的成分数据。基于原始土壤隶属度数据或经简单对数正态转换进行克里格插值,其空间预测精度虽然较高。但空间上任意点的土壤对于不同类别的隶属度之和不为1,因此这样的插值结果理论上没有可信基础。

(2) 原始土壤隶属度数据经非对称对数比和对称对数比转换后, 插值结果均满足空间上各点土壤隶属度之和为1和非负的必要条件。二者相比, 前者对区域总体趋势的反映较前者好, 且精度较高。因此, 所有土壤模糊隶属度数据转换方式中, 对称对数比转换最理想, 是一种能较好解决成分数据空间插值问题的数据转换技术。

表4 对称对数比转换数据半变异函数分析结果

Table 4 Theoretical semivariogram models of the transformed membership values by symmetry log-ratio and their corresponding parameters

类别 Type	模型类型 Model	块金值 Nugget	偏基台值 Partial sill	基台值 sill	块金效应 Nugget effect	有效变程 Effect range
A	球状模型 Spherical model	0.34	1.40	1.75	19.56	108
B	球状模型 Spherical model	1.47	0.60	2.07	70.92	804.9
C	指数模型 Index model	0	4.53	4.53	0	174.16
D	指数模型 Index model	0	6.84	6.84	0	123.04

(3) 本文研究表明, 不考虑成分数据的特殊要求, 模糊隶属度空间插值结果将带来必然的不确定性或不可靠性。成分数据的空间结构分析和区域化变量预测为相关土壤学研究提出了新课题。开发具有坚实理论基础且实用、便捷、可操作性强的模型算法与数据转换技术具有重要意义。

References:

- [1] De Gruijter J J, McBratney A B. A modified fuzzy k-means method for predictive classification. In: Bock, H. H. ed. Classification and Related Methods of Data Analysis. Elsevier, Amsterdam, 1988. 97—104.
- [2] Odeh I O A, McBratney A B, Chittleborough D J. Soil pattern recognition with fuzzy-c-means: applications to classification and soil-landform interrelationships. *Soil Sci. Soc. Am. J.*, 1992, 56: 505—516.
- [3] McBratney A B, De Gruijter J J, Brns D J. Spatial prediction and mapping of continuous soil classes. *Geoderma*, 1992, 54(12), 39—64.
- [4] McBratney A B, De Gruijter J J. A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science*, 1992, 43:159—175.
- [5] Burrough P A, van Gaans P F M, Hootsmans R. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma*, 1997, 77: 115—135.
- [6] Walvoort D J J, De Gruijter J J. Compositional kriging: a spatial interpolation method for compositional data. *Math. Geol.*, 2001, 33:951—966.
- [7] Odeh I O A, Todd A J, Triantafilis J. Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, 2003, 168(7):501—514.
- [8] Zhou D. Geological compositional data analysis: difficulties and solutions. *Journal of China University of Geosciences Earth Science*, 1998, 23:147—152.
- [9] Feng N N, Li T X, Zhang X Z, et al. Spatial Variability of soil particle composition in hilly tea plantation soils with different sampling scales. *Journal of Soil and Water Conservation*, 2006, 20(3):123—128.
- [10] Liu F C, Shi X Z, Pan X Z. Characteristics of spatial variability of soil granules in a typical area of Southern Jiang Province. *Chinese Journal of Soil Science*, 2003, 34(4):247—249.
- [11] Odeh I O A, McBratney A B, Chittleborough D J. Fuzzy-c-means and Kriging for mapping soil as a continuous system. *Soil Sci. Soc. Am. J.*, 1992, 56:1848—1854.
- [12] Triantafilis J, Ward W T, Odeh I O, et al. Creation and interpolation of continuous soil layer classes in the Lower Namoi Valley. *Soil Sci. Soc. Am. J.*, 2001, 65:403—413.
- [13] De Gruijter J J, Walvoort D J J, Van Gaans P F M. Continuous soil maps—a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma*, 1997, 77: 169—95.
- [14] Bragato G. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. *Geoderma*, 2004, 118: 1—16.
- [15] Du G H, Zhang G L, Gong Z T. Discussion of soil characteristic horizons and soil series classification, *Soil*, 2001, 1:1—6.
- [16] Aitchison, J. The statistical Analysis of Compositional Data. Chapman & Hall, London, 1986.
- [17] Aitchison, J. The statistical analysis of compositional data. *J. Royal Stat. Soc. B.*, 1982, 44:139—177.
- [18] Pawlowsky V, Olea R A, Davis J C. Estimation of regionalized compositions: a comparison of three methods. *Mathematical Geology*, 1995, 27(1): 105—127.

参考文献:

- [8] 周蒂. 地质成分数据统计分析——困难和探索. *中国地质大学学报(地球科学)*, 1998, 23:147~152.
- [9] 冯娜娜, 李延轩, 张锡洲, 等. 不同尺度下低山茶园土壤颗粒组成空间变异性特征. *水土保持学报*, 2006, 20(3):123~128.
- [10] 刘付程, 史学正, 潘贤章, 等. 苏南典型地区土壤颗粒的空间变异特征. *土壤通报*, 2003, 34(4):247~249.
- [15] 杜国华, 张甘霖, 龚子同. 论特征土层与土系划分. *土壤*, 2001, 1:1~6.