

基于模糊 k-均值算法模型的土壤数值化分类^① ——以河南省境内分布的雏形土为例

孙亚洲¹, 陈杰^{1*}, 吴克宁², 李玲³, 韩杏杏¹, 王海洋¹

(1 郑州大学水利与环境学院, 郑州 450001; 2 中国地质大学(北京)土地科学技术学院, 北京 100083;

3 河南农业大学资源与环境学院, 郑州 450001)

摘要:本研究基于《中国土志·河南卷》中的 97 个雏形土土壤样本点, 根据土壤系统分类标准遴选并确定用于划分雏形土纲中不同层级分类单元的 16 个土壤诊断层和诊断特性, 并依据这些不同指标的属性特征值和不同层级指标的权重建立属性数据集。基于此数据集运用模糊 k-均值算法模型, 确定 6 个聚类中心土壤, 计算样本土壤与中心土壤之间的分类距离, 并对研究区 97 个土壤样本实施数值化分类。研究表明, 基于 97 个样本点与各中心土壤之间的分类距离进行土壤数值化分类, 并通过检验数值化土壤分类结果与谱系式层级土壤分类单元的一致性, 可以评价传统的谱系式层级土壤分类系统的合理性。

关键词: 模糊 k-均值算法模型; 土壤数值化分类; 雏形土

中图分类号: S159.9 **文献标识码:** A

在国内区域尺度数字化土壤制图研究与实践中, 多数学者利用研究区成土因素以及地表人为干预情况作为环境协变量, 构建各种土壤空间预测模型^[1-6]; 部分研究者则基于样点土壤如表层质地、各种物质含量、电导率、pH、土壤颜色和特征土层分布等发生学特性, 以及各土壤样点之间的空间关系, 构建随机模型实施土壤空间预测与数字化制图^[7-9]。无论采用何种技术途径, 数值化土壤分类(numerical soil classification)都是数字化土壤制图的关键环节之一。

数值化土壤分类的重要作用主要表现为两个方面: 一是数值化分类结果可以定量揭示不同土壤之间在性状特征上的相似程度和在发生学上的“亲疏”关系; 二是基于不同土壤类型属性空间上相对差异的数值化表达, 可以直接用于土壤-环境协变量关系的定量拟合^[10-14]。在以往的数值化土壤分类实践中, 常用的算法模型有 3 种, 分别为相似性层级分类、VLADIMIR 半自动分类和模糊连续分类模型^[15-20]。其中, 实现模糊连续的技术途径主要有两种。模糊聚类是一种自动分类方法, 不以现有的常规分类系统为基础。模糊 c-均值算法(FCM: fuzzy c-means algorithm)(也称模糊 k-均值算法)或模糊 c-均值聚类

(也称模糊 k-均值聚类), 是土壤科学中应用最广泛的模糊分类方法之一^[21-25]。

最大限度地发挥土壤诊断层、诊断特性的作用, 让数值化土壤分类结果与谱系式层级土壤分类单元具有最大程度的一致性, 无疑是提高数字化土壤制图与常规土壤图参比水平的必由之路。本文以分布于河南省境内的雏形土为研究对象, 基于诊断层和诊断特性构建雏形土发生特性最小数据集, 确立研究区中心土壤类型, 并应用模糊连续分类算法模型实施数值化土壤分类, 进而完成雏形土不同土族之间分类距离的计算, 为土壤空间预测与数字化土壤制图提供定量信息, 以提高数字化土壤制图与常规土壤图的参比基础、提升数字化土壤图输出结果的应用前景方面提供技术支持^[25]。

1 研究区概况与研究方法

1.1 研究区概况

河南省地处黄河中、下游, 华北平原的南端, 横跨我国第二、三级地貌台阶。地貌类型包括中低山地、丘陵、山间盆地、山前冲积平原和黄泛平原等。由于地处暖温带与北亚热带的过渡地区, 生物气候条

基金项目: 国家科技基础性工作专项(2008FY110600)和国家自然科学基金项目(40971128)资助。

* 通讯作者(jchen@zzu.edu.cn)

作者简介: 孙亚洲(1992—), 男, 河南济源人, 硕士研究生, 主要从事数字土壤制图研究。E-mail: sunyazhou00@hotmail.com

件多样,地形地貌类型繁多且地层与岩性复杂,水文及水文地质情况各异,同时河南省具有长期水耕与旱耕的悠久历史,独具特色的自然和人为条件决定了省内复杂多样的土壤形成过程和土壤分布特征。雏形土是指土壤发育程度较弱,并以雏形层为主要诊断层的土壤。在中国土壤系统分类中,除了具有明显其他诊断层、诊断特性的土纲和无诊断特性的新成土之外,其余的发育弱而未成熟的土壤尽归雏形土纲中^[26-28]。雏形土在河南省境内分布广泛,相对集中于省内黄淮海平原地区。根据《中国土系志·河南卷》统计,全省雏形土分布面积 7.7 万 km²,约占河南省土壤总面积的 60%^[29]。

1.2 数据来源

本研究所用的土壤基础数据来源于国家科技基础性专项“我国土系调查与《中国土系志》编制”项目中的“河南省土系调查与《中国土系志·河南卷》编制”课题(编号:2008FY110600)。图 1 是 2009 年 12 月至 2013 年 3 月在河南省境内布设的 97 个雏形土土系调查样点分布图,样点的信息包括土壤分布环境与条件、形态特征及空间位置等^[29]。

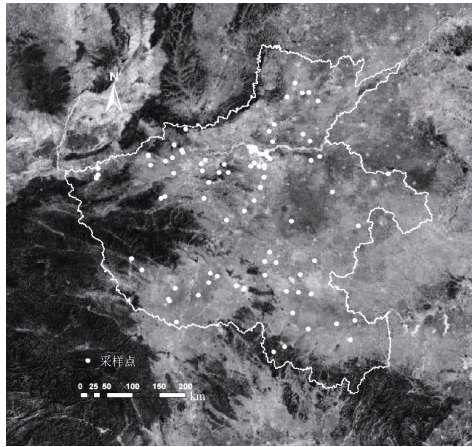


图 1 河南省境内雏形土土系调查样点分布示意图

Fig.1 Spatial distribution of sampled cambisol series in Henan Province

1.3 数值化土壤分类因子遴选与赋值

对《中国土壤系统分类检索(第三版)》中雏形土各级分类单元的设置以及分类标准进行统计分析,排除与雏形土分类无关的土壤发生学属性,然后确定雏形土土纲中用于划分不同层级分类单元的土壤诊断层和诊断特性,共 16 个诊断层和诊断特性被用于河南省境内的雏形土土族以上各分类单元的划分(表 1)^[26-27]。其中,用于划分土纲的属性有 1 个,划分亚纲的有 2 个,划分土类的有 3 个,划分亚类的有 6 个,划分土族的有 4 个。

表 1 用于划分河南省境内雏形土各级分类单元的
诊断层、诊断特性
Table 1 Diagnostic horizons and characteristics of cambisols in
Henan province

分类单元	诊断层、诊断特性
土纲	1、雏形层
亚纲	1、半干润土壤水分状况;2、0~50 cm 土层具有氧化还原特性
土类	1、暗沃表层;2、钙质凝团或结核(砂姜)体积≥10%;3、50~100 cm 土层具有氧化还原特性
亚类	1、水耕现象;2、盐积现象;3、50~100 cm 土层具有氧化还原特性;4、石灰性;5、潜育特征;6、变性现象
土族	1、土壤温度状况;2、控制层段土壤颗粒大小;3、土壤矿物学特性;4、土壤酸碱性

对于样点土壤而言,在某一分类层级上特定诊断层、诊断特性被使用与否是其赋值的主要依据。独立土壤诊断层和诊断特性,例如暗沃表层,如果被用于样点土壤分类就赋值为 1,未使用则为 0;对于土壤水分状况、土壤温度状况、控制层段土壤颗粒大小、土壤矿物学特性等组合土壤属性指标,将其赋予不同数值以最大限度地表现土壤属性差异、提升数值化土壤分类效果^[25]。例如,将河南省境内分布的雏形土的土壤温度状况包括热性土壤温度状况和温性土壤温度状况,分别赋值为 1.2 和 0.8(表 2)。

表 2 数值化土壤分类因子赋值举例
Table 2 Value assignment examples of indexes of soil family in
numerical soil classification

土壤属性	土壤分类因子	特征值
土壤水分状况	潮湿土壤水分状况	1.2
	湿润土壤水分状况	1
	干润土壤水分状况	0.8
土壤温度状况	热性	1.2
	温性	0.8
土壤矿物学特性	蛭石混合型	0.8
	混合型	1.2
控制层段土壤颗粒大小	黏质	1.5
	黏壤质	1.4
	壤质	1.3
	砂质	1.2
	壤质盖粗骨	1.1
	粗骨质	1
	粗骨壤质	0.9
	粗骨盖壤质	0.8
	粗骨盖粗骨壤质	0.7
	粗骨砂质	0.6

1.4 土壤分类因子权重的确定

中国土壤系统分类是谱系式层级分类系统,不同的诊断层、诊断特性在不同的层次分类单元划分时发挥作用。对于参与土壤分类的土壤发生学属性,其作用在划分不同土壤、不同层级的分类单元时是不一样的;对于同一种土壤,不同的发生学属性在不同层级的分类单元划分中的作用也是不一样的。某些诊断层、诊断特性,只用于某些特定土壤类型的划分,譬如锥形土的锥形层,只能用于区分研究区内的锥形土,而不能用于同一地区其他土纲及其下属单元的划分。在同一土纲内,不同发生学属性被用于不同层级的单元划分,但在不同土纲之间,同样的诊断特性可用于划分不同层级的土壤分类单元,譬如土壤水分状况可以在一个土纲中被用于划分亚纲,而在另一个土纲中则被用来划分亚类甚至更低级单元。

由于用于高层级分类单元划分的诊断层、诊断特性在同一土纲中具有更广泛的差异性,可以最先把土壤类型在高级层级区分开来,之后的较低层级的单元续分都是在此基础上进行的。如果将不同层次间的离散分类转换为同一层级上的连续分类,应该赋予用于传统分类中高层级单元划分的发生学属性更高的权重,层级越低,在连续的数值分类中,土壤发生学属性的分类权重就越低(表 3)。

表 3 数值化分类因子在用于不同分类层级时的权重分配
Table 3 Factor weights of different classification hierarchies in Numerical Soil Classification

层级	土纲	亚纲	土类	亚类	土族
权重	1	0.8	0.6	0.4	0.2

毋庸置疑,分类层级越低,分类单元蕴涵的分类信息也就越丰富,参与的诊断层、诊断特性以及其他辅助属性也就越多,这就为定量分析不同分类单元之间的发生学联系提供了坚实的基础。

1.5 土壤数值分类模糊 k-均值算法模型

本研究采用模糊 k-均值算法模型实施土壤数值化分类。对于一组土壤样本而言,模糊 k-均值算法模型的输出结果包括聚类类别数量、中心土族及其样本土壤对于中心土壤的模糊隶属关系(fuzzy memberships)或样本土壤与中心土壤类型之间的分类距离(taxonomic distance)。本研究将之前文中的定量分类因子作为模糊连续分类的输入参数,以基于分类距离的算法输出结果作为讨论对象。具体的模糊 k-均值算法步骤参阅文献[23-24, 30-32]。

本研究计算的分类距离为欧氏距离(Euclidean distance):

$$d_{12} = \sqrt{\sum_{k=1}^n (X_{1k} - X_{2k})^2}$$

式中: d_{12} 为土壤样本与中心土壤之间的分类距离, X_{1k} 、 X_{2k} 分别为两个土壤样本样点的第 k 个属性值。分类距离越大,表明土壤样本与中心土壤类型之间性状相似性越小,发生学关系越远;分类距离越小,表明土壤样本与中心土壤类型之间相似性越大,发生学联系越紧密。

本研究在 R 语言中运用 k-均值算法模型计算土壤样本与中心土壤之间的分类距离,基于分类距离判断土壤样本类别的归属^[33]。

2 结果与讨论

2.1 研究区锥形土聚类类别与中心土族

运用模糊 k-均值模型将研究区分属于 79 个土系的 97 个土壤样点聚类为 6 个类别,各类别的发生学属性值心值见表 4。

基于表 4 中各聚类质心的诊断层、诊断特性数据,依照《中国土壤系统分类检索(第三版)》^[27],将各聚类中心土族分别确认为壤质混合型温性非酸性-石灰底锈干润锥形土(中心土族 1)、黏壤质混合型热性非酸性-普通淡色潮湿锥形土(中心土族 2)、壤质混合型热性非酸性-普通淡色潮湿锥形土(中心土族 3)、黏壤质混合型热性-普通砂姜潮湿锥形土(中心土族 4)、壤质混合型温性非酸性-石灰筒育干润锥形土(中心土族 5)、壤质混合型温性非酸性-石灰淡色潮湿锥形土(中心土族 6)。

2.2 研究区锥形土分类距离

研究区锥形土 6 个中心土族之间分类距离的计算结果见表 5,样点土壤与各中心土族之间的分类距离见表 6。

表 5 数据显示,中心土族 2、3(黏壤质混合型热性非酸性-普通淡色潮湿锥形土、壤质混合型热性非酸性-普通淡色潮湿锥形土)之间的分类距离最小,表明二者在土壤诊断特征上具有最大的相似性,在成土过程中具有最密切的发生学联系;而中心土族 4、5(黏壤质混合型热性-普通砂姜潮湿锥形土、壤质混合型温性非酸性-石灰筒育干润锥形土)之间的分类距离最大,二者的土壤诊断特征的相似性最低、成土过程中的发生学联系最弱。

据表 6 数据统计,与第 1 个中心土壤剖面距离最小的土壤剖面样本有 36 个,与第 2 个中心土壤距离最小的土壤剖面样本有 4 个,与第 3 个中心土壤距离最小的土壤剖面样本有 3 个,与第 4 个中心土壤距离最小的土壤剖面样本有 4 个,与第 5 个中心土壤距离

表 4 研究区雏形土聚类别发生学属性质心值

Table 4 Centroid values of soil genetic properties of different cambisol clustering categories in study region

土壤属性	类别 1	类别 2	类别 3	类别 4	类别 5	类别 6
暗沃表层	0.033	0.150	0	0	0	0
半湿润土壤水分状况	0.444	0	0	0	0.640	0
湿润土壤水分状况	0.244	0	0	0	0	0
潮湿土壤水分状况	0	0.960	0.960	0.960	0	0.960
温性土壤状况	0.098	0	0	0.040	0.160	0.160
热性土壤状况	0.093	0.240	0.240	0.180	0	0
变性特征	0	0	0	0.100	0	0
0 ~ 50 cm 土层具氧化还原特性	0	0.800	0.800	0.800	0	0.800
50 ~ 100 cm 土层具氧化还原特性	0.483	0	0	0	0	0
石灰性	0.200	0.100	0	0	0.391	0.400
钙质凝团或结核(砂姜)体积 ≥ 10%	0	0	0	0.600	0	0
水耕现象	0	0.100	0	0	0	0
盐积现象	0	0	0	0	0	0.267
混合型	0.227	0.240	0.240	0.240	0.240	0.240
蛭石混合型	0.009	0	0	0	0	0
黏壤质	0.109	0.280	0	0.140	0.048	0
壤质	0.123	0	0.260	0.130	0.205	0.260
砂质	0.013	0	0	0	0.010	0
黏质	0	0	0	0	0.001	0
粗骨质	0.006	0	0	0	0	0
粗骨壤质	0.020	0	0	0	0.008	0
粗骨盖壤质	0	0	0	0	0.003	0
粗骨盖粗骨壤质	0	0	0	0	0.003	0
粗骨砂质	0.003	0	0	0	0	0
壤质盖粗骨	0	0	0	0	0.005	0
非酸性	0.122	0.150	0.200	0.050	0.170	0.133

表 5 研究区雏形土中心土族之间的分类距离

Table 5 Taxonomic distance matrix between centroid soil families of cambisol in study region

	中心土族 1	中心土族 2	中心土族 3	中心土族 4	中心土族 5	中心土族 6
中心土族 1	0	1.471	1.470	1.575	0.629	1.486
中心土族 2	1.471	0	0.437	0.681	1.506	0.650
中心土族 3	1.470	0.437	0	0.659	1.488	0.565
中心土族 4	1.575	0.681	0.659	0	1.603	0.831
中心土族 5	0.629	1.506	1.488	1.603	0	1.432
中心土族 6	1.486	0.650	0.565	0.831	1.432	0

表 6 研究区雏形土样点土壤与各中心土族之间的分类距离

Table 6 Taxonomic distance matrix between sampled and centroid soil families of cambisol in study region

样点	中心土族 1	中心土族 2	中心土族 3	中心土族 4	中心土族 5	中心土族 6
样点 1	1.475	0.212	0.382	0.659	1.518	0.682
样点 2	1.470	0.437	0	0.659	1.488	0.565
样点 3	1.470	0.437	0	0.659	1.488	0.565
样点 4	1.470	0.437	0	0.659	1.488	0.565
样点 5	1.642	0.752	0.816	0.393	1.681	0.992
... ..						
样点 93	0.661	1.476	1.536	1.615	0.312	1.481
样点 94	0.661	1.476	1.536	1.615	0.312	1.481
样点 95	0.686	1.522	1.530	1.632	0.313	1.474
样点 96	0.686	1.522	1.530	1.632	0.313	1.474
样点 97	0.661	1.449	1.483	1.564	0.500	1.534

最小的土壤剖面样本有 47 个，与第 6 个中心土族距离最小的土壤剖面样本有 3 个。

2.3 研究区雏形土数值化分类结果参比与检验

通过对照 97 个样点土壤在《中国土系志·河南卷》中的土族划分及其亚类、土类级别的归属，检验本研究中与特定中心土族分类距离最小的一组样点土壤在《中国土系志·河南卷》中是否同属一个亚类或者土类。结果表明，其中与第 5 个中心土壤(壤质混合型温性非酸性-石灰简育干润雏形土)分类距离最小的 47 个土壤剖面样本包含且只包含所有的简育干润雏形土(表 7)，也就是说以这个中心土族为代表的聚类类别

中，样点土壤均来自于简育干润雏形土，并且所有的简育干润雏形土都属于这个类别，表明其常规土壤分类在这种土族划分上较为合理，现有的划分简育干润雏形土的诊断层和诊断特性可以区分该土类与其他土类，并且土类内部的相似性很高；而《河南省土系志》中的同属淡色潮湿雏形土的土族样本在数值化分类中被分别划分为第 2、3、6 三个不同的聚类类别(表 7)，表明传统的谱系式层级分类系统中的淡色潮湿雏形土的划分标准不合理，基于现有的诊断层和诊断特性划分的淡色潮湿雏形土的内部差异很大，可以把淡色潮湿雏形土拆分为多个土类，或者增加限定条件。

表 7 研究区雏形土样本土壤数值化分类结果与中心土族参比
Table 7 Reference between results of numerical soil classification and centroid soil families of cambisol in study region

中心土族		分类距离最近的土壤样点土类	数量
中心土族 1	壤质混合型温性非酸性-石灰底锈干润雏形土	简育湿润雏形土	11
		底锈干润雏形土	25
中心土族 2	黏壤质混合型热性非酸性-普通淡色潮湿雏形土	暗色潮湿雏形土	1
		淡色潮湿雏形土	
中心土族 3	壤质混合型热性非酸性-普通淡色潮湿雏形土	淡色潮湿雏形土	9
中心土族 6	壤质混合型温性非酸性-石灰淡色潮湿雏形土	淡色潮湿雏形土	
中心土族 4	黏壤质混合型热性-普通砂姜潮湿雏形土	砂姜潮湿雏形土	4
中心土族 5	壤质混合型温性非酸性-石灰简育干润雏形土	简育干润雏形土	47

3 结论

1) 本研究将《中国土系志·河南卷》中的 97 个雏形土土壤样点进行 k-均值聚类，依照《中国土壤系统分类检索(第三版)》将 6 个聚类中心土族分别定为壤质混合型温性非酸性-石灰底锈干润雏形土、黏壤质混合型热性非酸性-普通淡色潮湿雏形土、壤质混合型热性非酸性-普通淡色潮湿雏形土、黏壤质混合型热性-普通砂姜潮湿雏形土、壤质混合型温性非酸性-石灰简育干润雏形土和壤质混合型温性非酸性-石灰淡色潮湿雏形土。

2) 本研究基于河南省境内的雏形土土壤样本的诊断层、诊断特性信息，应用模糊连续分类算法模型实施数值化土壤分类，通过检验数值化分类结果与谱系式层级土壤分类单元的一致性，可以检验传统谱系式层级分类系统的合理性。结果表明，壤质混合型温性非酸性-石灰简育干润雏形土的土壤分类标准较为合理；淡色潮湿雏形土的分类标准不合理。

参考文献：

[1] 赵量. 基于模糊集理论的土壤——景观定量关系提取及制图应用[D]. 南京: 南京农业大学, 2007

[2] 周银. 基于决策树方法的县级土壤数字制图研究[D]. 杭州: 浙江大学, 2011

[3] 郭澎涛. 丘陵山地预测性土壤制图研究[D]. 重庆: 西南大学, 2009

[4] 张春弟. 模糊土壤制图的研究和应用[D]. 武汉: 华中农业大学, 2014

[5] 王良杰. 基于 GIS 的中比例尺数字土壤制图研究[D]. 南京: 南京农业大学, 2009

[6] 王琴. 基于土壤图的土壤与环境关系的获取及推理制图[D]. 武汉: 华中农业大学, 2013

[7] 刘鹏飞. 基于分类距离的数字土壤制图技术的研究[D]. 郑州: 郑州大学, 2011

[8] 李立东, 陈杰, 宋轩, 等. 空间回归模型在区域数字化土壤制图中的应用——以河南封丘县为例[J]. 土壤学报, 2013, 50(1): 21-29

[9] 章明奎, 夏建强, 符娟林. 基于土层空间变异性的比例尺土壤调查制图研究[J]. 浙江大学学报(农业与生命科学版), 2006, 32(3): 346-349

[10] Carré F. and McBratney A. B. Digital terrain mapping[J]. Geoderma, 2005, 128: 340-353

[11] Carré F. and Jacobson M. Numerical classification of soil profile data using distance metrics[J]. Geoderma, 2009, 148: 336-345

[12] 王恒钦, 潘剑君, 卢浩东, 等. 诊断层和诊断特性与景观关系探讨——以江苏省句容某样区为例[J]. 土壤, 2015, 47(4): 758-764

- [13] 邱霞霞, 李德成, 赵玉国, 等. 基于不同土壤分类系统的土壤景观格局分析——以我国西北黑河流域中游地区为例[J]. 土壤, 2016, 48(5): 1022–1031
- [14] 黄魏, 罗云, 汪善勤, 等. 基于传统土壤图的土壤—环境关系获取及推理制图研究[J]. 土壤学报, 2016, 53(1): 72–80
- [15] Rayner J H. Classification of soils by numerical methods[J]. Journal of Soil Science, 1966, 17: 79–92
- [16] Carré F. and Girard M.C. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes[J]. Geoderma, 2002, 110: 241–263
- [17] 刘鹏飞, 宋轩, 刘晓冰, 等. 基于 k-均值算法模型的区域土壤数值化分类及预测制图[J]. 生态学报, 2012, 32(6): 1846–1853
- [18] 刘晓冰. 区域数字土壤制图方法对比研究[D]. 郑州: 郑州大学, 2013
- [19] 刘超, 卢玲, 胡晓利. 数字土壤质地制图方法比较——以黑河张掖地区为例[J]. 遥感技术与应用, 2011, 26(2): 177–185
- [20] 韩卫东. 耕地土壤肥力数字化制图方法对比研究[D]. 杭州: 浙江大学, 2014
- [21] De Gruijter J J, McBratney A B. A modified fuzzy k-means method for predictive classification[C] // Bock H H. Classification and related methods of data analysis. Amsterdam: Elsevier, 1988: 97–104
- [22] Verheyen K, Dries A, Hermy M, et al. High-resolution continuous soil classification using morphological soil profile descriptions[J]. Geoderma, 2001, 101: 1–48
- [23] 檀满枝, 陈杰. 模糊逻辑在土壤连续分类和制图表达中的应用及展望[J]. 土壤学报, 2009, 46(1): 137–143
- [24] 檀满枝, 陈杰. 模糊 c-均值算法在区域土壤预测制图中的应用[J]. 土壤学报, 2009, 46(4): 571–577
- [25] 孙亚洲. 基于分类距离的土壤系统分类土族单元划分研究[D]. 郑州: 郑州大学, 2017
- [26] 龚子同, 张甘霖, 陈志诚, 等. 土壤发生与系统分类[M]. 北京: 科学出版社, 2007
- [27] 中国科学院南京土壤研究所土壤系统分类课题组, 中国土壤系统分类研究协作组, 中国土壤系统分类检索(第三版)[M]. 合肥: 中国科学出版社, 2001: 161–180
- [28] 李德成, 张甘霖. 中国土壤系统分类土系描述的难点与对策[J]. 土壤学报, 2016, 53(6): 1–8
- [29] 吴克宁, 李玲, 鞠兵, 等. 中国土系志·河南卷[M]. 北京: 科学出版社, 2017
- [30] McBratney A B, De Gruijter J J. A continuum approach to soil classification by modified fuzzy k-means with extragrades[J]. Journal of Soil Science, 1992, 43: 159–175
- [31] De Gruijter J J, Walvoort D J J, van Gamsb P F M. Continuous soil maps-A fuzzy set approach to bridge the gap between aggregation levels of process and distribution models[J]. Geoderma, 1997, 77: 169–95
- [32] McBratney A B, Odeh, et al. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurement, and fuzzy decisions[J]. Geoderma, 1997, 77: 85–113
- [33] 徐建华. 现代地理学中的数学方法[M]. 北京: 高等教育出版社, 2002: 71–72

Numerical Soil Classification Based on Fuzzy k-means Algorithm Model——A Case Study of Cambisols in Henan Province

SUN Yazhou¹, CHEN Jie^{1*}, WU Kening², LI Ling³, HAN Xingxing¹, WANG Haiyang¹

(1 College of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China; 2 School of Land Science and Technology, China University of Geosciences, Beijing 100083, China; 3 College of Resource and Environment, Henan Agricultural University, Zhengzhou 450001, China)

Abstract: According to Chinese Soil Taxonomy(CST), we selected 16 soil diagnostic horizons and diagnostic characteristics, which can be used to classify cambisols at different levels based on 97 cambisol soil profiles in *Soil Series of China (Henan)*, and we built a data matrix with different characteristics and hierarchy weights of these indexes. Then we extracted 6 centroid soil families by fuzzy k-means algorithm model, and calculated the taxonomic distance between sampled and centroid soil families, meanwhile numerically classified the 97 cambisol soil profiles. We find that the rationality of the traditional hierarchic CST can be evaluated according to the consistence of traditional hierarchic classifying units and the numerical soil classification derived from the taxonomic distance between sampled and centroid soil families.

Key words: Fuzzy k-means algorithm model; Numerical soil classification; Cambisols