

基于主成分分析和聚类分析的土壤养分特性研究

赵月玲¹, 林玉玲², 曹丽英¹, 马 丽¹, 陈桂芬¹

(1 吉林农业大学 信息技术学院, 吉林 长春 130118; 2 长春嘉和外科医院, 吉林 长春 130051)

摘要:为明确吉林省中北部土壤养分变化特征, 充分了解土壤本身的特性, 达到合理施肥的目的, 本研究对 55 个耕层土壤(0~30 cm)进行采样, 测定分析了土壤 pH 以及土壤中速效氮、速效磷、速效钾、有机质和微量元素的含量, 应用主成分分析和聚类分析方法评价了土壤肥力特性. 结果表明, 研究区的最佳分区数为 6 种养分类型: 第 I 类为缓效钾、有机质、锰和硼极低型; 第 II 类为速效钾、锰、铁和硼偏高, 全磷偏低型; 第 III 类为有机质偏高, 其他元素适中型; 第 IV 类为速效钾偏低, 缓效钾较高, 硼偏低型, 其他元素适中型; 第 V 类为速效钾、铁含量偏低, 其他元素适中型; 第 VI 类为速效钾、全磷、pH 偏高, 锰偏低型. 6 种类型占比分别为 40%、5%、24%、13%、13% 和 5%. 为了简化施肥操作流程, 也可把其分为更少的类型.

关键词:地学统计学; 土壤养分; 空间变异; 主成分分析; 聚类分析
中图分类号: S158 文献标志码: A 文章编号: 1001-411X(2013)04-0484-05

A Study of Soil Nutrients Characteristics Based on Principal Component and Cluster Analysis

ZHAO Yueling¹, LIN Yuling², CAO Liying¹, MA Li¹, CHEN Guifen¹

(1 College of Information and Technology, Jilin Agricultural University, Changchun 130118, China;
2 Jiahe Surgical Hospital, Changchun 130051, China)

Abstract: The objective of the study was to analyze and determine soil characteristics and apply fertilizer rationally by the results of available nitrogen, available phosphorus, available potassium, soil pH, organic matter, trace elements content characteristics in Jilin Province. Principal component analysis and cluster analysis methods were applied to evaluate fertility characteristics in the soil research. Soil characteristics of available nitrogen, available phosphorus and available potassium and organic matter, pH value, soil trace element content index were explored based on 55 soil samples (0–30 cm). The results showed that the optimal partition number was 6 nutrient clusters: I) The content of Slow-release potassium, organic matter, manganese and boron was very low; II) The content of total phosphorus was low, the content of Fe and boron was high; III) The content of organic matter was high; IV) The content of available potassium was high, the content of boron was low; V) The content of available potassium and Fe was low; VI) The content of available potassium, total phosphorus, pH were high, the content of manganese was low. The ratios of these 6 various clusters were 40%, 5%, 24%, 13%, 13% and 5%. In order to simplify the fertilization process, the soil categories could be divided into fewer classes.

Key words: geostatistics; soil nutrients; spatial variability; principal component analysis; cluster analysis

收稿日期: 2012-11-19 网络出版时间: 2013-09-09
网络出版地址: <http://www.cnki.net/kcms/detail/44.1110.S.20130909.1900.008.html>
作者简介: 赵月玲(1975—), 女, 实验师, 博士研究生; 通信作者: 陈桂芬(1956—), 女, 教授, 博士, E-mail: guifchen@163.com
基金项目: 吉林省科技厅青年基金(201101114); 吉林省教育厅研究项目(201247, 201248); 吉林农业大学基金项目(201133)

近年来,随着国内外精准作业技术体系的不断提高和应用,根据土壤养分自身的变异特性划分农田养分管理分区,进而进行变量施肥管理决策,在国内外农业研究中已成为一种常用的方法之一^[1-4].对土壤营养状况的了解,直接影响到农业生产潜力,有必要对土壤养分含量水平进行综合分析评价,它是有效利用农业资源和技术的一种途径.采用管理分区划分技术可以代替网格采样并在变量施肥中得到广泛应用^[5].这样既能提高土壤养分利用效率,又能节省资源,获得较好的经济效益,达到保护有限的农业生态资源的目的^[6].众多科技研究者运用与农业生产相关的一种或几种因素结合来进行管理分区划分,使这种划分技术得到了进一步的提升.管理分区技术的主要关键点在于根据各种限制产量的不同因子相似性,将目标地块划分成子区域进行精细管理.因此,了解土壤养分的空间变异情况和自相关性是农田精准养分管理的重中之重^[7].如 Kravchenko 等^[8]根据地形特征和土壤属性来定义管理分区,通过利用多年作物产量数据的综合分析,进行精准农业管理分区的提取研究.檀满枝等^[9-10]在研究土壤重金属的分类时采用 c -多次组合法,与此同时确定了适宜的模糊类别数和模糊加权指数,取得了令人满意的结果.

采用主成分分析法和聚类分析方法,以量化形式表现土壤理化性状的综合指标,可以大大减少单个指标所反映出的物理特性带来的差异性^[11-12].本研究根据不同的土壤养分指标,进行统计学分类,探讨多因素分析技术在土壤养分管理分区方面研究的可行性,应用主成分分析和聚类分析相结合的分析方法,对 55 个采样点的多指标相关矩阵进行研究,分析耕层土壤养分的空间变异性和自相关性,了解土壤养分含量及其空间变异特征,确定田间管理分区级别,为农田土壤养分指标评价提供新的途径和方法.

1 材料与方法

1.1 研究区域概况

研究区位于吉林省中北部,北纬 $44^{\circ}30'57'' \sim 45^{\circ}15'02''$ 、东经 $126^{\circ}01'44'' \sim 127^{\circ}05'09''$.该地属半湿润温带大陆性季风气候,年平均气温 $4.6 \sim 5.6^{\circ}\text{C}$,年降水量在 $500 \sim 600 \text{ mm}$,绝大部分集中于暖季,占年降水量的 90% 左右,土壤类型主要是典型的黑土,主要种植玉米和大豆等作物.

1.2 指标测定与方法

本研究选用吉林省中北部玉米主产区土壤,由于研究土壤的范围较大,研究人员间隔一定距离随

机选取每块农田的中央部分,施肥前对 $0 \sim 30 \text{ cm}$ 耕层土壤进行多点采集、混合取样,共采集 55 个样点(图 1).应用 DGPS 装置(全球卫星差分定位系统)进行精准定位确定土壤采样地点,每个采样点取样 0.8 kg 左右,装袋密封后带回实验室,经自然风干后磨细,过 1 mm 筛,用于土壤养分的测定.另取一部分样品,过 0.25 mm 筛,用于有机质和全量养分元素的测定.土壤有机质用重铬酸钾滴定法测定;碱解氮用碱解扩散法测定;速效磷用钼锑抗比色法测定;速效钾用醋酸铵浸提-火焰光度法测定;pH 用雷磁-25 型酸度计法(水土质量比为 $1:5$)测定;土壤中微量元素如有效铁、锰,用 1.0 mol 醋酸铵浸提-原子吸收分光光度法测定;有效铜、锌用 0.1 mol 盐酸浸提-原子吸收分光光度法测定;钼用光度法(比色法)测定.

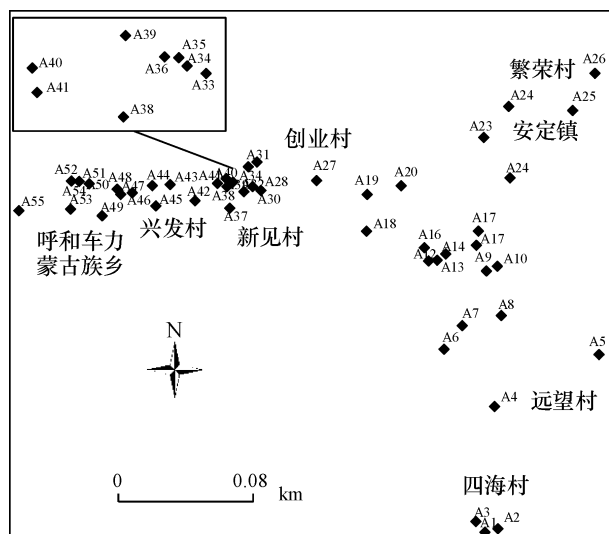


图 1 55 个土壤采样点分布图

Fig. 1 Distribution map of 55 soil samples

1.3 数据分析方法

所有试验数据采用 SPSS 18.0 软件进行分析.先对 55 个样本的 16 个指标进行标准化处理,求出相对应的相关矩阵,然后计算相关矩阵的特征根、方差贡献率及累计贡献率,确定因子个数,根据因子得分进行聚类分析,最后将土壤肥力的各项指标及分类结果输入到土壤数据库中.

2 结果与分析

2.1 土壤养分特性

在土壤学研究中,通常用变异系数(CV)的大小表示土壤特性空间变异性的强弱, $CV \leq 0.1$ 时为弱变异性, $0.1 < CV < 1.0$ 时为中等变异性, $CV \geq 1.0$ 时为强变异性^[13-15].对 55 个采样点的主要养分和微

量元素进行统计分析,结果见表 1. 由表 1 可见,吉林省典型土壤养分的变异性状各不相同,土壤中不同养分性质具有较大的差异,其中,缓效钾、速效磷、全氮、有机质、锌 5 个养分指标 CV 较大,分别为 0.77、1.10、1.14、0.71 和 1.40,表明速效磷、全氮和锌具有强变异特性;土壤 pH 具有弱变异的特性;其他养分则具有中等变异的特性.

表 1 55 个采样点土壤养分的主要特征

Tab.1 The main nutrient properties of 55 soil samples

养分 或 pH	$w(\text{养分})^{1)}/$ ($\text{mg} \cdot \text{kg}^{-1}$)	偏度	峰度	变异系数 (CV)
缓效钾	310.27 ± 238.25	1.71	3.31	0.77
速效氮	73.75 ± 33.15	1.82	5.98	0.45
速效磷	9.26 ± 10.23	1.98	4.75	1.10
速效钾	130.71 ± 51.52	0.92	0.99	0.39
全氮	390.00 ± 453.00	1.70	2.25	1.14
全磷	540.00 ± 278.00	0.52	0.46	0.51
全钾	43 360 ± 20 110	0.11	0.33	0.46
有机质	8 080 ± 5 770	1.87	4.56	0.71
有效硫	27.56 ± 18.46	0.75	0.85	0.67
有效锰	6.17 ± 4.14	0.72	-0.69	0.67
有效铁	4.58 ± 2.79	1.18	1.65	0.61
有效锌	0.65 ± 0.91	2.96	9.61	1.40
有效铜	1.03 ± 0.44	1.93	3.72	0.43
有效硼	0.21 ± 0.13	0.58	-0.56	0.60
有效钼	56.26 ± 20.41	0.19	-0.70	0.36
pH	8.33 ± 0.19	-0.507	0.57	0.02

1) 表中数据为平均值 ± 标准差.

2.2 主成分分析

为明确各个养分指标的相对重要性,应用 SPSS 18.0 统计软件对土壤中 16 个养分指标状态进行了主成分分析,首先进行 Bartlett 检验,其中 KMO 检验值为 0.743, $P < 0.05$,因此可以对原始数据进行主成分分析. 标准化处理各样品养分的测定结果,经主成分分析计算得到各主成分的特征值和贡献率(表 2). 由表 2 可以看出,第 1 个主成分的方差贡献率约 19%,第 2 个主成分的方差贡献率约 18%,第 3 个主成分的方差贡献率约 11%,第 4 个主成分的方差贡献率约 10%,第 5 个主成分的方差贡献率约 8%,第 6 个主成分的方差贡献率约 7%,第 7 个主成分的方差贡献率约 7%,第 8 个主成分的方差贡献率约 7%. 即一个 8 因子模型解释了试验数据的 87%,而主成分 1、2、3、4、5 这 5 个因子已包含样本的大部分信息(累计方差贡献率为 66%).

表 2 8 个主要因子特征值及贡献率

Tab.2 The eigen values and contribution rates of 8 main factors

因子	转化前			转化后		
	特征值	贡献率 /%	累积贡 献率/%	特征值	贡献率 /%	累积 贡献率/%
1	5.378	33.613	33.613	2.983	18.642	18.642
2	2.311	14.444	48.057	2.903	18.145	36.787
3	2.072	12.952	61.009	1.739	10.866	47.654
4	1.257	7.856	68.865	1.637	10.232	57.885
5	0.843	5.269	74.134	1.279	7.996	65.882
6	0.795	4.967	79.101	1.152	7.202	73.083
7	0.632	3.950	83.051	1.103	6.894	79.977
8	0.577	3.608	86.659	1.069	6.681	86.659

从表 3 可以看出,第 1 主因素和锰高度正相关,第 2 主因素和硼高度正相关,第 3 主因素与铁高度正相关,第 4 主因素和全磷高度负相关,第 5 主因素和缓效钾高度正相关,第 6 主因素和速效钾高度正相关,第 7 主因素和有机质高度正相关,第 8 主因素与 pH 高度负相关. 因为总方差近 87% 的贡献来自这 8 个主因素,所以可以认为锰、硼、铁、全磷、缓效钾、速效钾、有机质和 pH 这 8 个主要因素是土壤养分的特征元素.

表 3 8 个主要因子对应的载荷矩阵

Tab.3 Loading matrix of 8 principal components reserved

养分或 pH	1	2	3	4	5	6	7	8
缓效钾	0.646	-0.138	-0.130	0.311	0.506	0.166	0.025	0.019
速效氮	0.529	0.306	0.342	0.187	-0.464	-0.316	0.043	0.006
速效磷	0.750	0.236	0.345	-0.141	0.192	-0.008	-0.056	-0.188
速效钾	-0.256	-0.064	0.471	0.422	-0.210	0.648	0.053	-0.039
全氮	0.868	-0.176	-0.019	-0.014	-0.097	-0.187	0.108	0.040
全磷	0.049	0.440	0.058	-0.772	0.051	0.190	0.146	0.205
全钾	-0.845	0.105	0.261	-0.041	-0.020	0.008	0.143	0.038
有机质	0.383	0.396	-0.464	0.139	-0.086	0.062	0.591	-0.283
有效硫	-0.435	0.037	0.551	0.200	0.386	-0.274	0.316	0.224
有效锰	0.892	-0.119	-0.075	0.076	0.234	0.037	-0.078	-0.027
有效铁	0.573	-0.086	0.717	0.003	-0.060	-0.017	-0.048	0.041
有效锌	0.115	0.772	0.300	-0.156	0.147	0.186	-0.037	-0.065
有效铜	0.782	-0.060	0.397	-0.023	-0.108	0.056	0.087	0.103
有效硼	0.074	0.816	-0.124	0.178	0.014	-0.056	-0.320	-0.140
pH	-0.700	0.236	0.349	0.156	0.172	-0.235	-0.010	-0.311

2.3 聚类分析

聚类分析作为一种探索性的分类方法,是一种建立分类的多元统计分析方法,它能够将一批样本(或变量)数据根据其诸多特征,按照性质上的亲疏程度在没有先验知识的情况下进行自动分类,产生

多个分类结果^[16]. 为减少土壤养分分类的指标数量, 保证样品分类结果的可信度与科学性, 本文将主成分分析的结果(即 55 个样品的 8 个指标)进行聚类分析研究. 采用 SPSS 18.0 聚类分析程序对 55 个样本进行聚类分组. 在方法上采用欧氏距离测量, 使用平均联接(组间)的树状图重新调整距离聚类合并, 每 2 个样本间用 Average linkage 法连结, 8 个指标作为聚类变量, 可将 55 个样品分成 6 大类, 聚类结果如图 2.

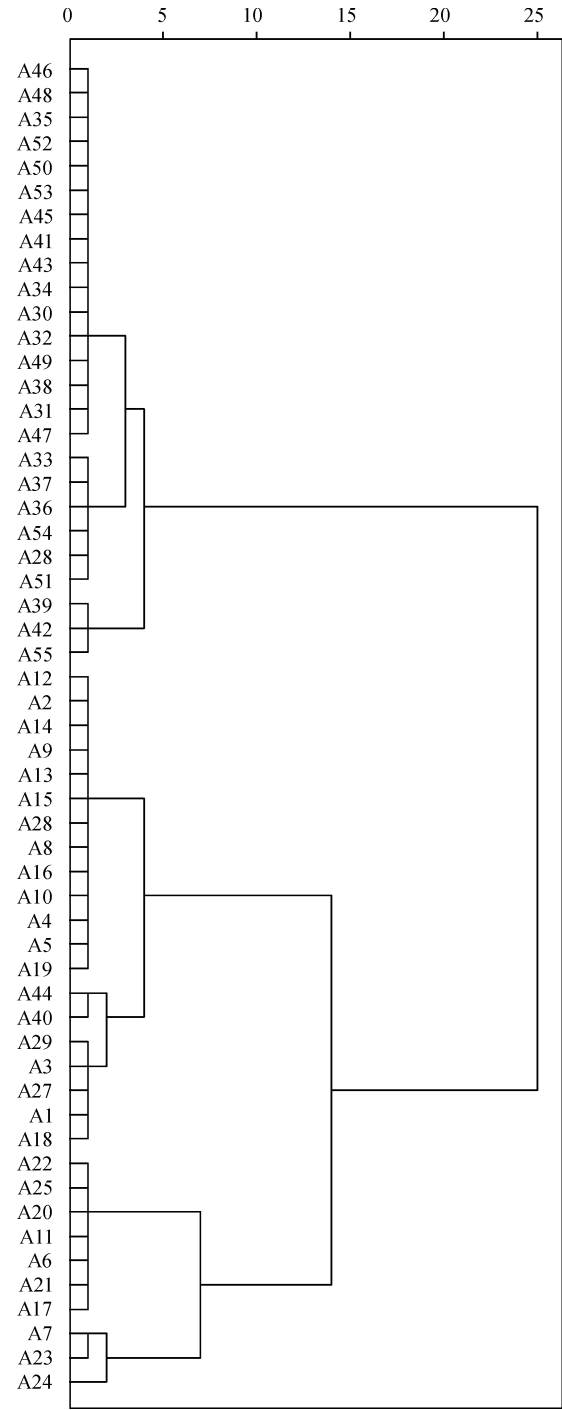


图 2 55 个土壤样品聚类图

Fig. 2 Clustering figure of 55 soil samples

6 类土壤的养分特性见表 4. 由图 2 和表 4 可见, 6 类养分特性具有明显差异: 第 I 类, 主要包括采样点 A46、A48、A35、A52、A50、A53、A41、A43、A34、A30、A32、A49、A38、A31、A47、A33、A37、A36、A54、A28 和 A51, 占总样本的 40%, 这类土壤主要特征是缓效钾、有机质、锰和硼极低; 第 II 类, 主要包括采样点 A39、A42 和 A55, 占总样本的近 5%, 这类土壤的主要特征是铁和硼偏高, 全磷偏低; 第 III 类, 主要包括采样点 A12、A2、A14、A9、A13、A15、A26、A8、A16、A10、A4、A5 和 A19, 占总样本的 24%, 这类土壤的主要特征是有机质偏高, 其他元素适中; 第 IV 类, 包括 A44、A40、A29、A3、A27、A1 和 A18, 占总样本的 13%, 这类土壤的主要特征是速效钾偏低, 缓效钾较高, 硼偏低, 其他元素适中; 第 V 类, 包括 A22、A25、A20、A11、A6、A21 和 A17, 占总样本的 13%, 这类土壤主要特征是速效钾、铁含量偏低, 其他元素适中; 第 VI 类, 包括 A7、A23 和 A24, 占总样本的 5%, 这类土壤的主要特征是速效钾、全磷、pH 偏高, 锰偏低.

表 4 各类别土壤主要养分指标的平均值

Tab. 4 The average values of main nutrient indexes of different soil categories

类别	$w/(mg \cdot kg^{-1})$					$w/(g \cdot kg^{-1})$		pH
	缓效钾	速效钾	有效锰	有效铁	有效硼	全磷	有机质	
I	180.13	137.16	4.30	4.47	0.09	0.53	5.56	8.38
II	571.93	121.10	13.40	9.02	0.40	0.33	6.89	8.09
III	321.61	135.48	6.00	3.75	0.28	0.56	12.48	8.29
IV	525.14	115.28	10.12	4.70	0.21	0.48	7.89	8.22
V	256.32	116.82	6.17	3.87	0.38	0.50	8.13	8.43
VI	225.51	140.86	4.20	5.88	0.50	0.93	9.07	8.43

3 讨论与结论

土壤中各种养分因子的选定与分级标准的划分目前尚无一个统一的标准^[17]. 本研究结果表明, 速效磷、全氮和锌具有强变异特性, 土壤 pH 具有弱变异的特性, 其他养分则具有中等变异的特性. 这与陈朝阳等^[14]对植烟土壤的研究结果相同, 但有的元素研究结果不同, 这可能与土壤种植的作物有关.

在土壤养分特性研究中, 为减少聚类分析的样本矩阵, 通过应用主成分分析的结果来减少数据的冗余; 这种方式计算简单, 所得的结论比较客观^[18]. 聚类分析方法是根据不同养分性状间所具有的相关性或者是样品间的相似性或差异性指标, 定量地确定样本间的亲疏关系. 这种分析目标数据的方法, 在不同的研究领域正在逐步展开并得以应用^[11-12]. 于江等^[19]应用主成分分析方法, 在复杂的土壤肥力指

标体系中筛选出若干个彼此不相关,但反映出原来全部指标所提供的大部分信息综合性指标,并依此主要信息聚类分析,进一步划分出最佳处理和评判修复效果.通过主成分分析和聚类分析把土壤的养分指标进一步细化分类,在植烟土壤中已有应用,如陈朝阳等^[14]针对土壤的主要理化性状指标把土壤分为5大类,然后针对不同类之间不同的理化性状指标,进行分类施肥指导种烟生产,具有一定的实际意义.赵瑞蕊等^[20]应用主成分和聚类分析方法对烤烟指标也进行了综合评价,将样本分为好、较好、中等、稍差、差5类指标.本研究按照这种研究方法将吉林省中北部土壤分为6类:第Ⅰ类缓效钾、有机质、锰和硼极低;第Ⅱ类速效钾、锰、铁和硼偏高,全磷偏低;第Ⅲ类有机质偏高,其他元素适中;第Ⅳ类速效钾偏低,缓效钾较高,硼偏低,其他元素适中型;第Ⅴ类速效钾、铁含量偏低型,其他元素适中型;第Ⅵ类速效钾、全磷、pH偏高,锰偏低.

田间土壤的养分特性在很大程度上受到成土母质、气候、地形、耕作制度等因素影响.一些因子在空间分布上呈一定的连续性,所以,土壤养分在相邻近点之间养分分布具有一定的相关性.用主成分聚类的定量分析方法对农田土壤养分进行综合评价并分类,所得结论客观、可信、较有说服力,为农田土壤施肥提供重要参考和依据;利用得到的主成分指标组成新矩阵作为聚类分析的样本矩阵,原理清晰,计算简单,大大减少了计算工作量.

土壤综合肥力是由水、肥、气、热4大指标体系构成,在以后的研究中,应该加大采样数量和研究范围,考虑更多的影响土壤肥力的指标并加以分析,同时在研究方法上要加强技术的交叉和融合.在土壤养分状况研究中,这里主要考虑的是由土壤主要养分和微量养分所构成的指标体系得出的结果,并未把土壤容重、垂直剖面等物理性状指标作为影响因子进行研究.所以这将是以后的研究重点方向之一.

参考文献:

- [1] 自由路,金继运,杨俐苹,等.基于GIS的土壤养分分区管理模型研究[J].中国农业科学,2001,34(1):1-4.
- [2] SCHEPERS A R, SHANAHAN J F, LIEBIG M K, et al. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years[J]. Agron J, 2004, 96: 195-203.
- [3] 李翔,潘瑜春,赵春江,等.基于空间连续性聚类算法的精准农业管理分区研究[J].农业工程学报,2005, 21(8):78-82.
- [4] 朱静,黄标,孙维侠,等.长江三角洲典型地区农田土壤有机质的时空变异特征及其影响因素[J].土壤,2006, 38(2):158-165.
- [5] KHOSLA R, FLEMING K, DELGADO J A, et al. Use of site specific management zones to improve nitrogen management for precision agriculture[J]. J Soil Water Conserv, 2002, 57: 513-518.
- [6] 赵其国.我国现代农业发展中的若干问题[J].土壤学报,1997, 34(1):1-9.
- [7] 陈防,刘冬碧,万开元,等.精准农业与农田精准养分管理现状及展望[J].湖北农业科学,2006, 45(4): 515-518.
- [8] KRAVCHENKO A N, BULLOCK D G. Correlation of corn and soybean grain yield with topography and soil properties [J]. Agron J, 2000, 92:75-83.
- [9] 檀满枝,陈杰,郑海龙,等.模糊c-均值聚类法在土壤重金属污染空间预测中的应用[J].环境科学学报,2006,26(12):2086-2092.
- [10] 檀满枝,陈杰,徐方明,等.基于模糊集理论的土壤重金属污染空间预测[J].土壤学报,2006,43(3):389-396.
- [11] 张建逵,康廷国,窦德强.林下山参与园参无机元素的聚类分析和主成分分析[J].中草药,2012,43(9): 1835-1840.
- [12] 曾洁,孙俊良,李光磊,等.基于主成分分析和Q型聚类分析的玉米品种特性研究[J].沈阳农业大学学报,2009,40(1)53-57.
- [13] 自由路,金继运,杨俐苹,等.农田土壤养分变异与施肥推荐[J].植物营养与肥料学报,2001,7(2):129-133.
- [14] 陈朝阳,陈星峰.南平烟区植烟土壤理化性状聚类分析与施肥对策[J].中国烟草科学,2012,33(3),17-21.
- [15] 刘继明,宋启亮,李芝茹,等.大兴安岭白桦低质林生态功能评价指标的灰色关联聚类分析[J].东北林业大学学报,2012,40(8):112-115.
- [16] 薛薇.基于SPSS的数据分析[M].北京:中国人民大学出版社,2006:327-377.
- [17] 张步翀.调亏灌溉春小麦农田土壤养分综合评价[J].灌溉排水学报,2007,28(4):1-5.
- [18] 鲍艳,胡振琪,柏玉,等.主成分聚类分析在土地利用生态安全评价中的应用[J].农业工程学报,2006,22(8):87-90.
- [19] 于江,朱昌雄,郭萍,等.生物腐植酸对新疆甘草产地沙退化土壤修复效果评价:以土壤养分指标为例[J].中国农业气象,2010, 31(3): 369-373.
- [20] 赵瑞蕊,何结望,王海明,等.基于主成分和聚类分析的湖北烤烟质量指标综合评价[J].中国烟草科学,2012.33(4):90-94.