

doi: 10. 11720/wtyht. 2022. 0043

刘凯,戴慧敏,刘国栋,等. 基于主成分聚类法的典型黑土区土壤地球化学分类[J]. 物探与化探, 2022, 46(5): 1132–1140. <http://doi.org/10.11720/wtyht.2022.0043>

Liu K, Dai H M, Liu G D, et al. Geochemical classification of the soil in a typical black soil area using the principal component analysis combined with K-means clustering algorithm[J]. Geophysical and Geochemical Exploration, 2022, 46(5): 1132–1140. <http://doi.org/10.11720/wtyht.2022.0043>

基于主成分聚类法的典型黑土区土壤地球化学分类

刘凯^{1,2,3}, 戴慧敏^{1,2,3}, 刘国栋^{1,2,3}, 宋运红^{1,2,3}, 梁帅^{1,2,3}, 杨泽^{1,2,3}

(1. 中国地质调查局 沈阳地质调查中心, 辽宁 沈阳 110034; 2. 自然资源部 黑土地演化与生态效应重点实验室, 辽宁 沈阳 110034; 3. 辽宁省黑土地演化与生态效应重点实验室, 辽宁 沈阳 110034)

摘要: 土壤地球化学分类对农业区划和生态区划具有指导意义。本文基于多目标区域地球化学调查获取的土壤常量元素数据, 利用主成分分析和 K 均值聚类组合方法——主成分聚类法, 对东北典型黑土区进行地球化学分类研究。结果显示: 成土母质是土壤常量元素特征的主要控制因素, 利用主成分聚类法将典型黑土区土壤样品划分为 5 类最为合理, 各类样品的常量元素含量具有显著性差异 ($P < 0.05$)。土壤地球化学分类结果与第四纪地质单元有一定的对应关系, 而分类结果更能反映成土母质的真实分布情况。同时, 松花江南部黑土区较高的 SiO_2 含量指示了黑土的沙化生态问题, 在黑土地保护中应给予重视。

关键词: 黑土区; 常量元素; 地球化学; 成土母质; 沙化

中图分类号: P632

文献标识码: A

文章编号: 1000-8918(2022)05-1132-09

0 引言

东北黑土地耕地面积占全国面积的 27%, 粮食总产量占全国的 1/4, 是我国十分重要的商品粮基地, 也是保障国家粮食安全的“压舱石”。2003 ~ 2020 年, 中国地质调查局在我国平原区持续开展了多目标区域地球化学调查工作^[1], 其中在东北平原区已完成调查面积约 48 万 km^2 , 采集了 60 余万件土壤样品, 获取了海量的土壤地球化学数据, 为开展黑土区地球化学研究提供了庞大的数据集^[2]。目前, 该数据集在黑土区碳储量计算^[3-5]、土地质量评价^[6]、关键带物质循环^[7-8]等方面得到了广泛应用, 但关于黑土区常量元素的研究尚未见报道。

常量元素是土壤化学成分中最主要的组成部分, 能有效反映土壤的粒度组成、矿物成分等属性^[9]。同时, 在土壤的形成演化过程中, 常量元素

也扮演着重要角色^[10-12]。更重要的是, 由于常量元素性质相对稳定, 通常可以反映成土母质的化学组成, 进而可以利用常量元素对成土母质进行有效预测^[13-14]。

地质过程必然会在土壤地球化学数据中留下痕迹^[15], 多种分析方法的结合更有助于揭示元素地球化学与土壤和质体间的相互关系^[16]。主成分分析和聚类分析都是研究土壤地球化学的常用统计方法, 主成分分析可以通过降维技术把多个变量化为少数几个主成分, 这些主成分仍能够反映原始变量的大部分信息^[17]; 聚类分析可以将性质相似的样品归为一类, 从而达到分类的目的。主成分聚类法则能将两种方法的优势进行融合, 先通过主成分分析获得样品的主成分得分值, 再利用得分值进行聚类分析。目前一些学者将主成分聚类法应用于土地质量评价^[18]、土样识别^[19]以及土壤管理分区^[20]等方面, 但在土壤地球化学分类中的应用仍然较少。

收稿日期: 2022-01-31; **修回日期:** 2022-06-09

基金项目: 中国地质调查局项目“东北黑土地 1: 25 万土地质量地球化学调查”(121201007000161312)、“兴凯湖平原及松辽平原西部土地质量地球化学调查”(DD20190520)

第一作者: 刘凯(1989-), 男, 2014 年毕业于吉林大学, 主要从事生态地球化学研究工作。Email: liu.kai@mail.cgs.gov.cn

通讯作者: 梁帅(1986-), 男, 2015 年毕业于辽宁工程技术大学, 主要从事基础地质和生态地质研究工作。Email: ls476476@163.com

本文基于多目标区域地球化学获得的土壤常量元素数据,尝试利用主成分聚类法,对东北典型黑土区进行地球化学分类研究并探讨其地质意义。

1 研究区概况

目前关于东北黑土存在“广义黑土区”、“黑土区”、“典型黑土区”、“黑土地”等多种说法^[21-22],其定义和分布范围差异较大。本文研究对象主要为中国土壤发生学分类中的黑土区,但由于其空间上分布分散,与草甸土等土壤相互交错,为保证研究区的完整性,根据中国 1 : 100 万土壤图(来源为 <http://westdc.westgis.ac.cn>)绘制了黑土的轮廓作为本次研究区(图 1),总面积约 7 万 km²。

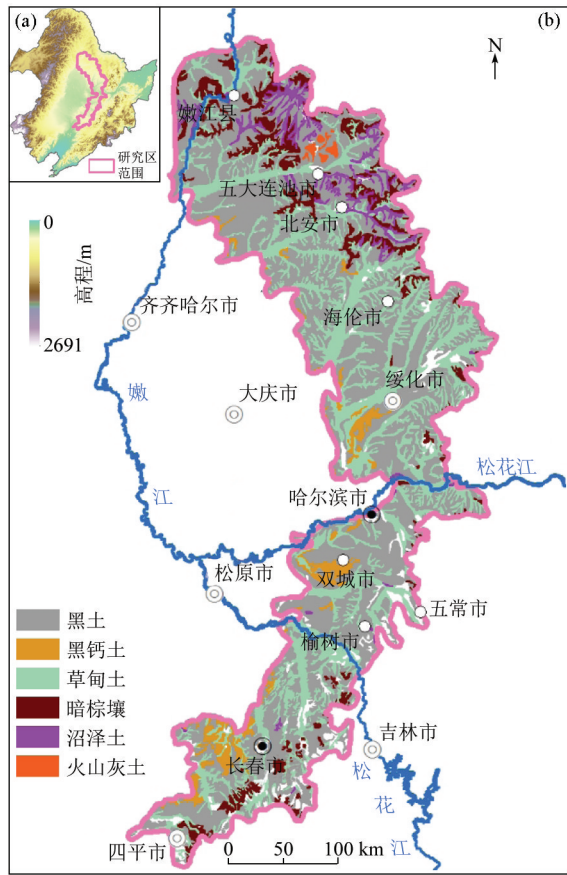


图 1 研究区区域位置(a)及土壤类型分布(b)
Fig.1 Regional location(a) and soil type distribution(b) of the study area

研究区位于松嫩平原东部、北部的山前台地和高平原区,北起黑龙江省嫩江县,南至辽宁省昌图县,沿哈尔滨—北安、哈尔滨—长春铁路沿线形成一条完整的黑土带。该区年降雨量在 500~600 mm,绝大部分集中于 7~9 月,干燥度≤1。年平均气温 0.5~6℃,土壤冻结时间为 120~200 d。研究区土

壤类型以黑土为主,其次为草甸土、暗棕壤,也包含有少量的黑钙土、沼泽土和火山灰土。研究区尚未编制精度较高的成土母质图,不同成土母质的分布尚不清晰,文献中多简单叙述为以第四纪更新世砂砾、黏土层和全新世砂砾、黏土层为主^[23]。

2 材料和方法

2.1 数据来源

本次研究采用的土壤地球化学数据全部来源于多目标区域地球化学调查^[2]。土壤样品获取采用双层网格化土壤测量方法,分别采集表层(0~20 cm)和深层(150 cm 以下)土壤,采样质量为 1 kg,样品在自然风干后过 20 目尼龙筛。表层土壤采样密度 1 点/km²,深层土壤采样密度 1 点/4 km²,每 4 个原始样品组合后测试分析 54 项元素和指标,因此每个表层测试样品代表 4 km² 范围,每个深层测试样品代表 16 km² 范围。测试元素和指标共 54 项,其中常量元素 SiO₂、Al₂O₃、Fe₂O₃、MgO、CaO、Na₂O、K₂O 采用 X 荧光光谱仪测试完成,测试准确度和精密度均达到相关规范^[24] 要求。在剔除个别异常点后,研究区内共选取 22 888 个表层土壤样品和 5 754 个深层土壤样品用于本次研究。

2.2 数据处理方法

2.2.1 描述性统计

元素含量的平均值、中位数、标准差等利用 SPSS 软件进行计算。富集系数(q)代表元素在表层土壤和深层土壤中含量的比值,在 ArcGIS 中按空间对应关系将 1 个深层样品化学属性赋予 4 个表层土壤样品,进而计算每个表层土壤的 q 值。变异系数(C_v)是标准差与平均值的比值,用以代表元素含量的离散程度。

2.2.2 主成分分析

主成分分析是常用的多元统计方法之一,主要用于多元数据的降维,本文使用 SPSS 软件进行主成分分析操作。主成分分析的前提是参与分析的各变量间具有相关性,图 2 显示各常量元素之间多具有显著相关性,KMO 检验系数为 0.6,Bartlett's 检验结果 P<0.001,说明常量元素之间存在共线性,可进行主成分提取。

2.2.3 K 均值聚类

聚类分析是无监督学习的一个重要工具,包括多种具体的聚类方法,如层次聚类分析、基于模型的聚类分析和模糊聚类等^[25]。K 均值聚类是最典型的聚类方法之一,其基本思想是以空间中 k 个点为

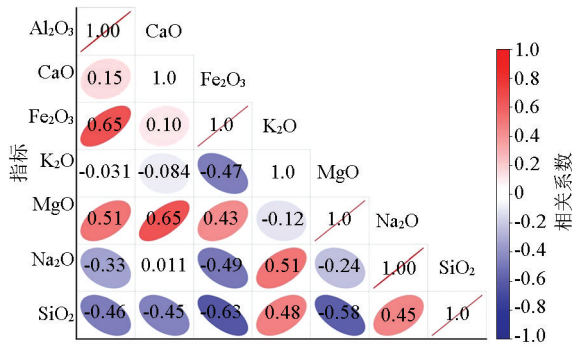


图 2 相关系数矩阵

Fig. 2 Correlation coefficient matrix diagram

中心对样品进行聚类,将最靠近中心的样品归为一类。该算法的最大优势在于简洁和快速,自 Macqueen^[26]提出以来,该方法已成为大量数据分析中最流行的算法之一,对于区域地球化学调查获取的海量样品最为适宜。

K 均值聚类由于要手动输入聚类数目,因此其难点之一就是确定最优分类数(K 值)^[27]。通常,当对数据集的属性非常了解时,可以靠经验来确定分类数,但这对数据处理者的知识储备提出了非常高的要求,往往难以实现。因此,大多数情况下人们通过计算不同分类数的某些特征,进行比较后来决定最优分类数。安光辉等^[28]利用一致性检验和区间差异显著性检验来评价不同分类数目效果,但该方法的主观性较大。Rousseuw^[29]于 1987 年提出利用

轮廓系数来判断分类数,轮廓系数结合了内聚度和分离度两种因素,可以用来在相同原始数据的基础上评价不同算法或者算法不同运行方式对聚类结果所产生的影响,目前得到较广泛的应用。轮廓系数计算公式如下:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)},$$

式中: a_i 为第 i 个对象到其所属簇中所有对象的平均距离; b_i 为该对象到所有非所属簇中对象的平均距离。轮廓系数在 $-1 \sim 1$ 间变化,越趋近于 1 代表内聚度和分离度都相对较优,聚类效果相对较好。因此,本文利用平均轮廓系数对分类结果进行评价。K 均值聚类和轮廓系数计算利用 Matlab 软件实现。

3 结果与讨论

3.1 常量元素含量特征

表 1 为研究区表层和深层土壤的常量元素统计数据,其元素含量高低趋势一致,均为 $\text{SiO}_2 > \text{Al}_2\text{O}_3 > \text{Fe}_2\text{O}_3 > \text{K}_2\text{O} > \text{Na}_2\text{O} > \text{CaO} > \text{MgO}$,其中 SiO_2 、 Al_2O_3 和 Fe_2O_3 含量占 80% 以上,具有明显的硅铝土特点。从元素含量分布形态看,除 CaO 以外的其他元素偏度在 1 附近,基本符合正态分布,而 CaO 表层和深层偏度值分别为 6.25 和 4.30,呈右偏形态,说明存在一定数量的极大值。

表 1 黑土区常量元素参数统计
Table 1 Statistical table of major elements in black soil area

类型	元素	最小值/%	最大值/%	平均值/%	中位数/%	标准偏差/%	偏度	峰度	变异系数
表层土壤	SiO ₂	41.87	80.46	64.55	64.80	2.95	-0.67	2.87	0.05
	Al ₂ O ₃	7.88	16.62	13.95	13.98	0.78	-0.59	2.95	0.06
	Fe ₂ O ₃	0.95	10.49	4.77	4.76	0.62	0.08	4.67	0.13
	K ₂ O	1.72	3.98	2.50	2.50	0.15	0.63	4.31	0.06
	Na ₂ O	0.72	4.66	1.62	1.56	0.29	1.82	6.44	0.18
	CaO	0.55	15.03	1.56	1.39	0.82	6.25	57.76	0.53
	MgO	0.21	3.50	1.29	1.30	0.21	0.24	3.21	0.17
深层土壤	SiO ₂	44.67	75.41	64.34	64.66	2.51	-0.36	3.10	0.04
	Al ₂ O ₃	9.01	18.50	14.79	14.89	0.80	-0.80	3.61	0.05
	Fe ₂ O ₃	1.38	9.08	5.12	5.16	0.78	-0.27	2.58	0.15
	K ₂ O	1.61	3.68	2.62	2.63	0.17	0.30	4.69	0.06
	Na ₂ O	0.53	3.66	1.67	1.64	0.28	1.38	5.60	0.17
	CaO	0.35	15.41	1.41	1.25	0.66	4.30	28.73	0.47
	MgO	0.36	3.09	1.37	1.40	0.24	-0.37	2.92	0.17

富集系数 q 代表了元素在土壤表层和深层的变化程度,当 q 在 $0.9 \sim 1.1$,说明元素含量在表层和深层变化不明显, $q > 1.1$ 时说明元素在表层富集, $q < 0.9$ 时说明元素在表层贫化而在深部富集。图 3 所示, SiO_2 的 q 值全部在 $0.9 \sim 1.1$ 之间,说明在垂向

上最为稳定, Al_2O_3 、 K_2O 、 MgO 、 Fe_2O_3 和 Na_2O 的表深土壤含量比分布范围略大,但仍以 $0.9 \sim 1.1$ 为主,CaO 则表现为在表层更富集。变异系数 C_v 代表了元素含量的离散程度,CaO 表层和深层 C_v 值分别为 0.53 和 0.47,为中等强度变异^[30],说明 CaO

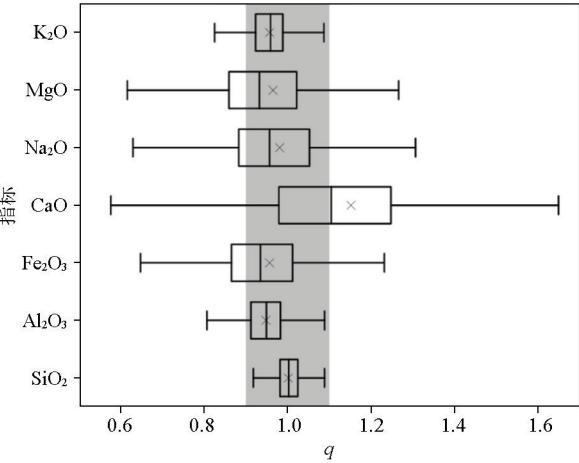


图 3 常量元素富集系数箱线图

Fig.3 Box diagram of enrichment coefficient of major elements

含量分布较不均匀。其他常量元素 C_v 值均小于 0.3,说明其空间变异程度较低。
以上结果表明,表层和深层土壤的常量元素整

体特征相似,体现了表层土壤对成土母质的继承性^[31],而表层土壤常量元素变异系数略大,则可能是成土过程中元素的迁移和再分配造成的^[32]。由于地球化学分类更突出元素含量高低和空间分布的差异性^[16],因此本次选用表层土壤用于地球化学分类研究。

图 4 为研究区表层土壤常量元素的空间分布图,插值方法为反距离权重法。 SiO_2 的高值区主要分布于黑土区南部和东北部, Al_2O_3 和 Fe_2O_3 的空间分布趋势整体上与 SiO_2 相反。 K_2O 和 Na_2O 分布特征类似,在黑土区南部和西北部为高值区。 CaO 和 MgO 分布特征基本一致,表现为西高东低的分布格局。不同元素的空间分布特征有一定的相似性,但差异性也很明显,因此难以直接利用单个常量元素进行地球化学区的划分,必须借助于主成分分析、聚类分析等方法来挖掘控制元素分布特征的内在因素^[33]。

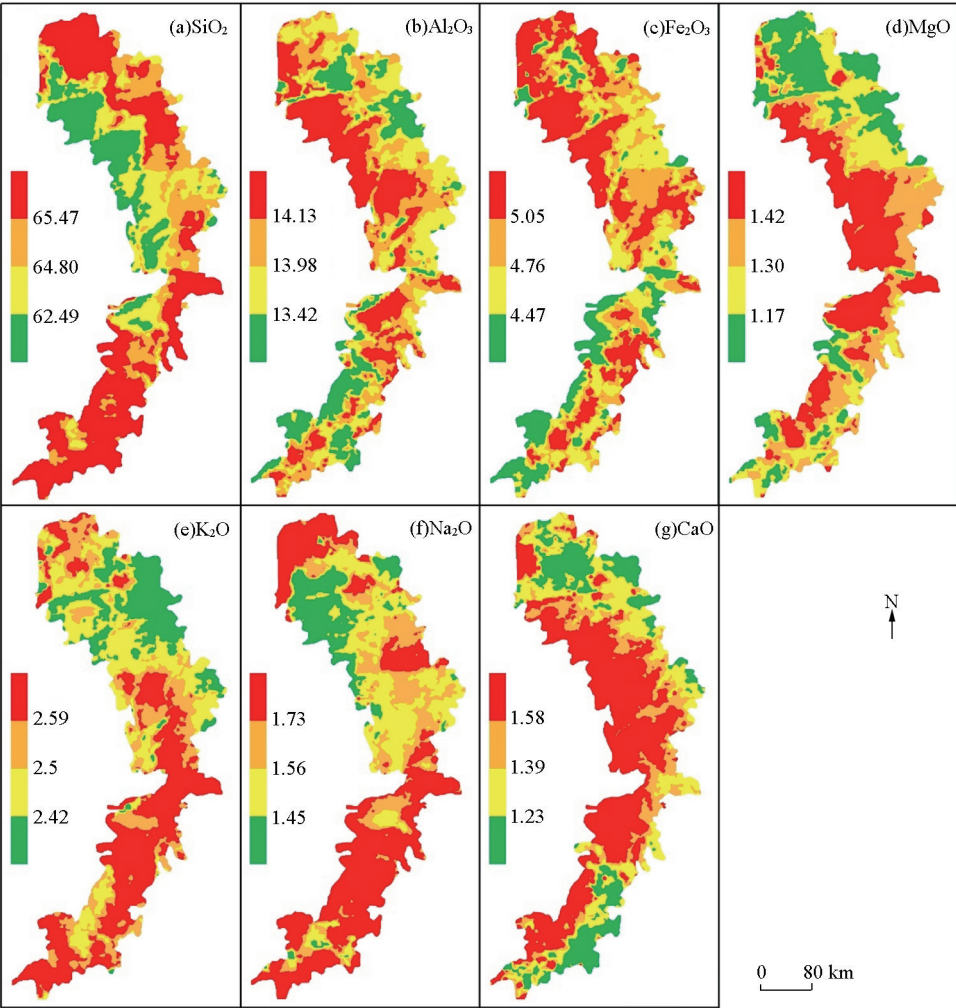


图 4 东北黑土区常量元素含量空间分布(各元素含量区间取值为四分位数,数据单位为%)

Fig.4 Spatial distribution of major elements in black soil area of Northeast China
(interval value of element content is quartile and data unit is %)

3.2 主成分分析结果

本文采用累积贡献率来判断主成分的最优数量。表 2 显示前 3 位主成分的特征值大于 1, 分别解释 46.708%、21.056% 和 15.289% 的数据变异, 因此本研究最终选取 3 个主成分 (PC1、PC2、PC3), 提取后的主成分共可解释 83.05% 的数据变异。

图 5 显示 PC1 与 Al_2O_3 、 Fe_2O_3 、 MgO 和 CaO 为正相关, 指示 PC1 代表黏土矿物和铁铝氧化物。PC2 与 CaO 及 MgO 呈正相关, 表明 PC2 代表了碳酸盐矿物。PC3 与 K_2O 、 Na_2O 相关性最强, 说明

PC3 代表长石类矿物。

表 2 主成分特征值及方差贡献率
Table 2 Principal component characteristics and variance contributions

主成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积贡献率/%	总计	方差百分比	累积贡献率/%
PC1	3.270	46.708	46.708	3.270	46.708	46.708
PC2	1.474	21.056	67.764	1.474	21.056	67.764
PC3	1.070	15.289	83.053	1.070	15.289	83.053

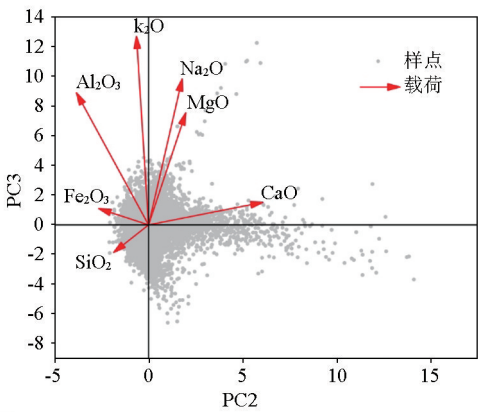
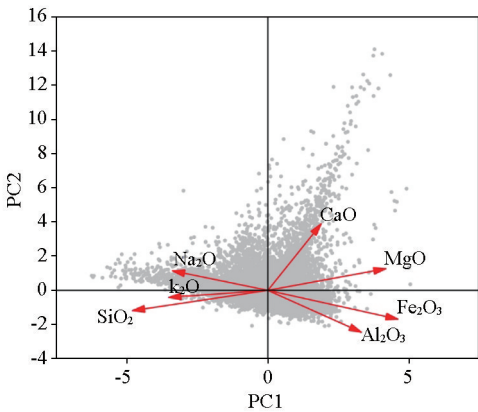


图 5 主成分分析得分

Fig. 5 Principal component analysis score diagram

3.3 K 均值聚类结果

利用样品的主成分得分值进行 K 均值聚类。首先将分类数 K 值分别设置为 2~8, 并利用 Matlab 中 evalclusters 函数计算对应的平均轮廓系数。结果显示, 当分类数为 5 时, 平均轮廓系数达到最大值 0.57 (图 6), 因此, 将研究区内数据分为 5 类较为合理, 使用 Turkey 多重比较法检验各类别间差异的显著性, 结果显示各类别常量元素含量具有显著性差异 ($P < 0.05$), 各类别样品数及元素平均含量见表 3。

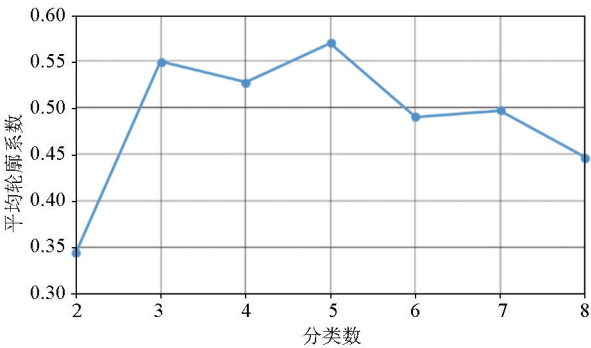


图 6 不同分类数及对应的平均轮廓系数

Fig. 6 Mean silhouette coefficients of different classification numbers

表 3 各类别样品常量元素平均值统计

Table 3 Statistical table of mean values of major elements in various classifications %

类别	样品数	SiO_2	Al_2O_3	Fe_2O_3	CaO	K_2O	MgO	Na_2O
I 类	18	56.15	14.37	6.16	3.06	2.90	3.63	2.19
II 类	6512	64.66	13.59	4.87	1.26	2.39	1.21	1.45
III 类	10023	63.07	14.46	5.19	1.57	2.50	1.45	1.54
IV 类	778	59.24	12.86	4.39	5.04	2.35	1.70	1.66
V 类	5557	66.75	13.51	4.24	1.38	2.65	1.18	1.94

为研究分类结果的实际内涵及应用价值, 绘制了样品类别的空间分布 (图 7)。

第 I 类样品数最少, 仅为 18 个, 但其元素含量特征明显, SiO_2 含量最低, 平均值仅 56.15%, 而 Fe_2O_3 、 MgO 、 CaO 含量最高 (表 3), 常量元素含量特征与基性岩较一致。第 I 类样品空间上位于五大连池玄武岩区, 土壤类型为火山灰土。

第 II 类样品与第 III 类样品元素含量特征较为类似, 其样品数量也占绝对优势, 分别占全区的 28.5% 和 43.8%。第 II 类样品的 SiO_2 含量 (平均值 64.66%) 略高于第 III 类样品 (平均值 63.07%), 反映其相对含有更高的砂质成分。在空间上, 第 II 类样品主要分布在小兴安岭西麓丘陵地区, 海拔普遍大于 200m (图 1), 该地区在第四纪时期以隆升为

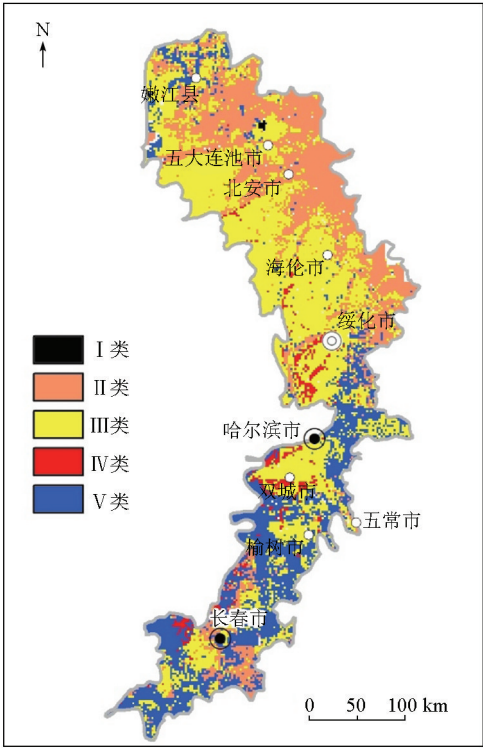


图 7 样品类别空间分布

Fig. 7 Spatial distribution map of sample categories

主,地形切割较剧烈,下部的砂层和砂砾石层多出露近地表,土壤类型以暗棕壤为主。

第Ⅲ类样品数量最多,占全区的 43.8%,其常量元素特征与典型黑土的成土母质——黄土状亚黏土^[34]基本一致,广泛分布在松辽平原东北部的高平原区,土壤类型多为黑土和草甸土。

第Ⅳ类样品仅 778 件,常量元素表现为含有较丰富的 CaO 和 MgO,其平均含量分别达 5.04% 和 1.7%,代表土壤中碳酸盐矿物含量较高。空间上沿个别河流分布,并均分布在隆起抬升的陡坡一侧,与黑钙土的分布吻合度较高。

第Ⅴ类样品 SiO₂、K₂O、Na₂O 含量较高,说明土壤中含有较多的石英和长石类矿物。在空间上,第Ⅴ类样品主要分布在松花江南部各河流两岸及一级阶地。

3.4 地球化学分类结果与地质单元的关系

土壤中元素地球化学特征受成土母质、气候、地形、生物、时间等因素综合控制,而以往研究表明,土壤常量元素特征与成土母质的关系最为密切^[32,35],土壤常量元素的空间分布与成土母质和地质单元往往具有较高的吻合度^[36]。

本次聚类结果中,除第Ⅰ类样品与玄武岩区完全对应外,其他 4 类样品在空间上与地质单元的对应性并不明显(图 7、图 8),其统计学结果也得出同

样结论(图 9)。原因可能包括两个方面:一方面,地质图的比例尺与土壤地球化学调查尺度不匹配,在以往开展的基础地质调查工作中,第四系广泛覆盖的平原区并不是调查重点,东北平原区目前编制的第四系地质图仅为 1:100 万比例尺,精度较低;另一方面,同一地质单元的地表岩性差异显著,第四纪地质单元通常由多个沉积地层构成,如本区冲湖积物通常上部为亚黏土,下部则为砂层、砂砾层,受地表侵蚀、沉积、风化等作用影响,同一地质单元在不同地区出露岩性有较大差异,因此,地质单元与地球化学的对应关系并不完全一致^[37-38],也是造成二者对应性低的一个原因。

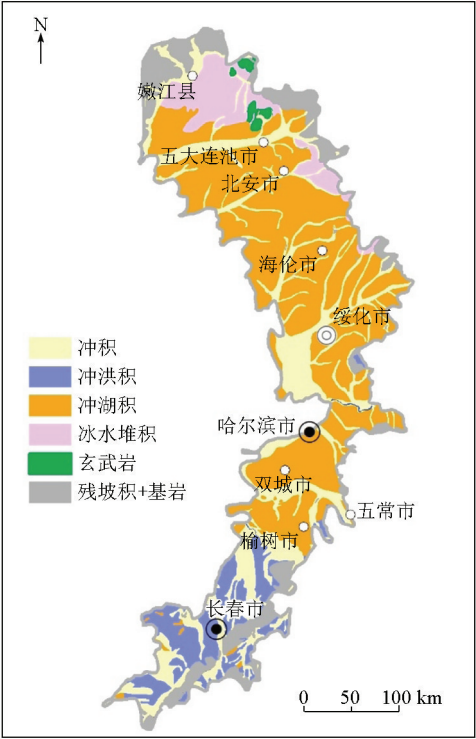


图 8 典型黑土区第四纪地质

Fig. 8 Quaternary geological map of typical black soil area

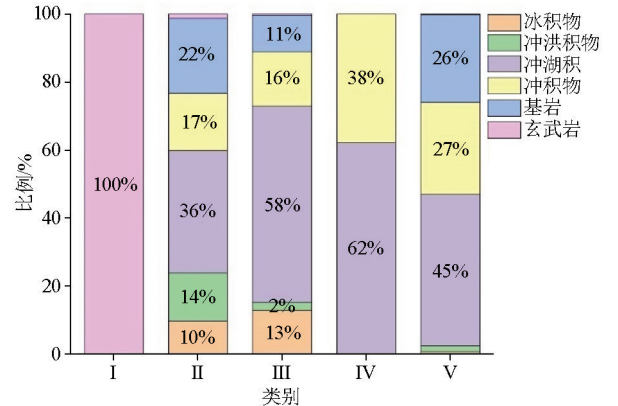


图 9 不同类别所包含地质单元的累积频率

Fig. 9 Cumulative frequency map of geological units in different classifications

由于区域多目标地球化学调查为 1 : 25 万比例尺,采样密度大,因此地球化学分类结果比地质图能更清晰地反映成土母质的空间分布情况。综合地理、地质和土壤特征,可将 5 个类别分别命名为:Ⅰ—五大连池玄武岩母质发育的火山灰土;Ⅱ—小兴安岭山前丘陵冲洪积物母质发育的暗棕壤;Ⅲ—松辽平原东部高平原低钙黄土状土母质发育的黑土;Ⅳ—松辽平原东部高平原边缘富钙冲湖积母质上发育的黑钙土;Ⅴ—松辽平原东南部冲积物母质上发育的黑土。该结果较客观地反映了典型黑土区的成土母质分布情况。

3.5 地球化学分类结果反映的黑土生态环境问题

地球化学分类结果表明,松花江南部土壤中砂质含量较北部明显偏高(图 7),原因一方面是由于南部成土母质以河流冲积物、冲洪积物和残坡积物为主(图 8),普遍含有较高的砂质成分,但另一方面可能也反映出局部地区的黑土存在沙化现象。近期研究表明,中国东北部地区已成为沙尘暴的多发区,尤其是松花江、嫩江流域的局部草地退化和沙化地区^[39]。受季风和沙尘暴作用影响,每年都有大量细粒沙尘由西向东吹至黑土区。以哈尔滨地区为例,现代尘暴中干沉降和湿沉降粉尘的 REE 模式和 Sr-Nd 同位素组成都表明,沙尘来源于科尔沁沙地和浑善达克沙地^[40-41],这些沙尘的加入会增加黑土中砂质含量,从而造成土壤沙化。在本次研究结果中,松花江南部冲湖积母质上发育的黑土中目前也具有较高的 SiO₂ 含量,指示这些地区土壤可能已发生沙化,未来应作为黑土区的生态问题加以关注。

4 结论

1) 典型黑土区常量元素 SiO₂、Al₂O₃ 和 Fe₂O₃ 含量占 80% 以上,具有明显的硅铝土特点,表层土壤和深层土壤的常量元素比以 0.9~1.1 为主,说明土壤常量元素主要受控于成土母质。

2) 利用主成分聚类法将黑土区土壤样品划分为 5 类最为合理,各类别地球化学特征差异显著,其空间分布能较清晰地反映出成土母质的空间分布情况。根据地理、地质及土壤特征分别命名为:Ⅰ—五大连池玄武岩母质发育的火山灰土;Ⅱ—小兴安岭山前丘陵冲洪积物母质发育的暗棕壤;Ⅲ—松辽平原东部高平原低钙黄土状土母质发育的黑土;Ⅳ—松辽平原东部高平原边缘富钙冲湖积母质上发育的黑钙土;Ⅴ—松辽平原东南部冲积物母质上发育的黑土。

3) 松花江南部冲湖积母质发育的黑土局部具有较高的 SiO₂ 含量,指示该地区黑土已出现沙化生态问题,在黑土地保护中应给予关注。

参考文献 (References):

[1] 李括,彭敏,赵传冬,等. 全国土地质量地球化学调查二十年[J]. 地学前缘,2019,26(6):128-158.
Li K, Peng M, Zhao C D, et al. Vicennial implementation of geochemical survey of land quality in China[J]. Earth Science Frontiers, 2019, 26(6): 128-158.

[2] 戴慧敏,赵君,刘国栋,等. 东北黑土地质量调查成果[J]. 地质与资源,2020,29(3):299.
Dai H M, Zhao J, Liu G D, et al. Progress in the quality survey of black soil in northeast China[J]. Geology and Resources, 2020, 29(3): 299.

[3] 戴慧敏,刘驰,宫传东,等. 东北平原土壤碳库构成及其与土壤性质的关系[J]. 第四纪研究,2013,33(5):986-994.
Dai H M, Liu C, Gong C D, et al. Soil carbon pool in northeast plain of China and its relations between the soil properties[J]. Quaternary Sciences, 2013, 33(5): 986-994.

[4] 刘国栋,戴慧敏,杨泽,等. 三江平原土壤碳库时空变化和影响因素研究[J]. 现代地质,2021,35(2):443-454.
Liu G D, Dai H M, Yang Z, et al. Temporal and spatial changes of soil carbon pool and its influencing factors in the Sanjiang Plain[J]. Geoscience, 2021, 35(2): 443-454.

[5] 刘国栋,李禄军,戴慧敏,等. 松辽平原土壤碳库变化及其原因分析[J]. 物探与化探,2021,45(5):1109-1120.
Liu G D, Li L J, Dai H M, et al. Change in soil carbon pool in Songliao Plain and its cause analysis[J]. Geophysical and Geochemical Exploration, 2021, 45(5): 1109-1120.

[6] 刘国栋,杨泽,戴慧敏,等. 黑龙江省海伦市长发镇土地质量地球化学评价及开发建议[J]. 地质与资源,2020,29(6):533-542.
Liu G D, Yang Z, Dai H M, et al. Geochemical evaluation of land quality and development suggestion of land in Hailun city, Heilongjiang Province[J]. Geology and Resources, 2020, 29(6): 533-542.

[7] 刘凯,杜守营,戴慧敏,等. 黑龙江省五常市东部土壤中硒分布及影响因素[J]. 地质与资源,2020,29(6):597-602.
Liu K, Du S Y, Dai H M, et al. Selenium distribution and influencing factors of soil in eastern Wuchang City, Heilongjiang Province[J]. Geology and Resources, 2020, 29(6): 597-602.

[8] 宋运红,张哲寰,杨凤超,等. 黑龙江海伦地区垦殖前后典型黑土剖面主要养分元素垂直分布特征[J]. 地质与资源,2020,29(6):543-549.
Song Y H, Zhang Z H, Yang F C, et al. Vertical distribution of major nutrient elements in typical black soil sections in Hailun, Heilongjiang Province: Before and after reclamation[J]. Geology and Resources, 2020, 29(6): 543-549.

[9] 李国刚,胡邦琦,李军,等. 山东半岛沿岸海域表层沉积物的常量元素及其地质意义[J]. 海洋地质与第四纪地质,2012,32

(3):45-54.

Li G G, Hu B Q, Li J, et al. Geochemistry of major elements in the surface sediments of the offshore area of Shandong Peninsula and its geological implications[J]. *Marine Geology and Quaternary Geology*, 2012, 32(3):45-54.

[10] Kumaravel V, Sangode S J, Siddaiah N S, et al. Major element geochemical variations in a Miocene-Pliocene Siwalik paleosol sequence: Implications to soil forming processes in the Himalayan foreland basin [J]. *Journal of the Geological Society of India*, 2009, 73(6):759-772.

[11] Du J, Luo Y, Zhang W, et al. Major element geochemistry of purple soils/rocks in the red Sichuan Basin, China: Implications of their diagenesis and pedogenesis [J]. *Environmental Earth Sciences*, 2013, 69(6):1831-1844.

[12] Veronica M N J, Georgesivc E E. A comparative analyses of granulometry, mineral composition and major and trace element concentrations in soils commonly ingested by humans [J]. *International Journal of Environmental Research and Public Health*, 2015, 12(8):8933-8955.

[13] Aldis M, Aherne J. Exploratory analysis of geochemical data and inference of soil minerals at sites across Canada [J]. *Mathematical Geosciences*, 2021, 53:1201-1221.

[14] Hossein T, Farhad K, Arash A, et al. Geochemistry of soils derived from selected sedimentary parent rocks in Kopet Dag, North East Iran [J]. *Journal of Geochemical Exploration*, 2018, 194:52-70.

[15] Hassan T, Ute M, Raimon T D, et al. Surficial and deep earth material prediction from geochemical compositions [J]. *Natural Resources Research*, 2018, 28:869-892.

[16] Wang J, Zuo R G, Caers J. Discovering geochemical patterns by factor-based cluster analysis [J]. *Journal of Geochemical Exploration*, 2017, 81:106-115.

[17] Jan S, Radim V, Jarmilac, et al. Regional geochemical zonation of cultivated floodplains-Application of multi-element associations for soil quality evaluation along the Ohře (Eger) River, Czech Republic [J]. *Journal of Geochemical Exploration*, 2020, 212:106491.

[18] 陈加兵, 曾从盛. 主成分分析、聚类分析在土地评价中的应用——以福建沙县夏茂镇水稻土为主要评价对象 [J]. *土壤*, 2001, 33(5):243-246, 256.

Chen J B, Zeng C S. Application of principle component analysis and hierarchical cluster analysis in land evaluation [J]. *Soil*, 2001, 13(5):243-246, 256.

[19] 王同兴, 郭骏杰, 王强. 基于 K 均值动态聚类分析的土样识别 [J]. *建筑科学*, 2010, 26(7):52-56, 71.

Wang T X, Guo J J, Wang Q. The recognition of soil sample based on the K-means dynamic clustering analysis [J]. *Building Science*, 2010, 26(7):52-56, 71.

[20] 郭燕, 田延峰, 吴宏海, 等. 基于多源数据和模糊 k-均值方法的农田土壤管理分区研究 [J]. *土壤学报*, 2013, 50(3):441-447.

Guo Y, Tian Y F, Wu H H, et al. Zoning of soil management based on multi-sources data and fuzzy-K means [J]. *Acta Pedologica Sinica*, 2013, 50(3):441-447.

[21] 赵玉明, 程立平, 梁亚红, 等. 东北黑土区演化历程及范围界定研究 [J]. *土壤通报*, 2019, 50(4):765-775.

Zhao Y M, Cheng L P, Liang Y H, et al. Developing history and defining boundary of the black soil regions in Northeast China [J]. *Chinese Journal of Soil Science*, 2019, 50(4):765-775.

[22] 刘宝元, 张甘霖, 谢云, 等. 东北黑土区和东北典型黑土区的范围与划界 [J]. *科学通报*, 2021, 66(1):96-106.

Liu B Y, Zhang G L, Xie Y, et al. Delineating the black soil region and typical black soil region of northeastern China [J]. *Chinese Science Bulletin*, 2021, 66(1):96-106.

[23] 中国科学院林业土壤研究所. 中国东北土壤 [M]. 北京: 科学出版社, 1980.

Institute of forestry and soil, Chinese Academy of Sciences. *Soil in Northeast China* [M]. Beijing: Science Press, 1980.

[24] 中华人民共和国国土资源部. DZ/T 0258—2014 多目标区域地球化学调查规范 (1:250 000) [S]. 北京: 中国标准出版社, 2015.

Ministry of Land and Resources of the People's Republic of China. DZ/T 0258—2014 Specification of multi-purpose regional geochemical survey (1:250 000) [S]. Beijing: China Standard Press, 2015.

[25] Clemens R, Peter F, Karl F, et al. The concept of compositional data analysis in practice: Total major element concentrations in agricultural and grazing land soils of Europe [J]. *Science of the Total Environment*, 2012, 426:196-210.

[26] Macqueen J. Some methods for classification and analysis of multivariate observations [C]//Proc of Berkeley Symposium on Mathematical Statistics & Probability, 1965.

[27] 冯波, 郝文宁, 陈刚, 等. K-means 算法初始聚类中心选择的优化 [J]. *计算机工程与应用*, 2013, 49(14):182-185, 192.

Feng B, Hao W N, Chen G, et al. Optimization to K-means initial cluster centers [J]. *Computer Engineering and Applications*, 2013, 49(14):182-185, 192.

[28] 安光辉, 马蓉, 陈伟, 等. 基于 K-均值聚类的绿洲农田管理分区提取的研究 [J]. *石河子大学学报: 自然科学版*, 2011, 29(6):757-761.

An G H, Ma R, Chen W, et al. Delineation of precision agriculture management zones in oasis field based on K-means algorithm [J]. *Journal of Shihezi University: Natural Science*, 2011, 29(6):757-761.

[29] Rousseeuw P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis [J]. *Journal of Computational and Applied Mathematics*, 1987, 20:53-65.

[30] 张明, 陈国光, 高超, 等. 华东多目标区域地球化学调查区土壤常量元素地球化学特征 [J]. *吉林大学学报: 地球科学版*, 2014, 44(3):995-1002.

Zhang M, Chen G G, Gao C, et al. Geochemical characteristics of macro elements in soils in the region covered by multi-purpose geochemical survey in Eastern China [J]. *Journal of Jilin University: Earth Science Edition*, 2014, 44(3):995-1002.

[31] Michael D, Paul D, Andreas S, et al. Principal component analysis of the geochemistry of soil developed on till in Northern Ireland [J]. *Journal of Maps*, 2013, 9(3):373-389.

[32] Sandra B, Efren G O, Francisco J G N, et al. Geochemical distribution of major and trace elements in agricultural soils of Castilla-La Mancha (central Spain): Finding criteria for baselines and delimiting regional anomalies[J]. *Environmental Science and Pollution Research*, 2017, 26(4): 3100–3114.

[33] Drew L J, Grunsky E C, Sutphin D M, et al. Multivariate analysis of the geochemistry and mineralogy of soils along two continental-scale transects in North America[J]. *Science of the Total Environment*, 2010, 409(1): 218–227.

[34] 韩晓萌, 戴慧敏, 梁帅, 等. 黑龙江省拜泉地区典型黑土剖面元素地球化学特征及其环境指示意义[J]. *地质与资源*, 2020, 29(6): 556–563.

Han X M, Dai H M, Liang S, et al. Element geochemistry of the typical black soil sections in Baiquan area, Heilongjiang Province: Environmental implication[J]. *Geology and Resources*, 2020, 29(6): 556–563.

[35] Chipres J A, Calleja A D L, Tellez J I, et al. Geochemistry of soils along a transect from Central Mexico to the Pacific Coast: A pilot study for continental-scale geochemical mapping[J]. *Applied Geochemistry*, 2009, 24(8): 1416–1428.

[36] Wilson M J. The importance of parent material in soil classification: A review in a historical context[J]. *Catena*, 2019, 182: 104–131.

[37] Fabian E G, Jasmin B, Volkmar M, et al. From geological to soil parent material maps: A random forest-supported analysis of geological map units and topography to support soil survey in South Tyrol[J]. *Geoderma*, 2019, 354: 113884.

[38] Rudolph S M, Justin B R. Investigating surficial geologic controls on soil properties, inorganic nutrient uptake, and northern hardwood growth in Western Massachusetts, USA[J]. *Journal of Soil Science and Plant Nutrition*, 2020, 20: 19–20.

[39] 杨雪艳, 张丽, 袁祝香, 等. 东北地区春季沙尘天气变化特征及其与大气环流变化的关系[J]. *气象与环境学报*, 2018, 34(4): 75–83.

Yang X Y, Zhang L, Qiu Z X, et al. Characteristics of sand-dust events and their relationships with atmospheric circulation in spring in Northeast China[J]. *Journal of Meteorology and Environment*, 2018, 34(4): 75–83.

[40] 袁方, 谢远云, 迟云平. 哈尔滨沙尘暴天气沉降物的物质组成及其对物源的限制[J]. *中国地质*, 2018, 45(6): 1177–1187.

Yuan F, Xie Y Y, Chi Y P. Material characteristics of dust fallouts during the dust-storm weather in Harbin: Constraint on the provenance[J]. *Geology in China*, 2018, 45(6): 1177–1187.

[41] 谢远云, 孟杰, 郭令芬, 等. 哈尔滨沙尘沉降物稀土元素地球化学特征及其物源分析[J]. *地球科学: 中国地质大学学报*, 2013, 38(5): 923–933.

Xie Y Y, Meng J, Guo L F, et al. REE Geochemistry for sand-dust fallouts in Harbin, Heilongjiang Province and provenance analysis[J]. *Earth Science: Journal of China University of Geosciences*, 2013, 38(5): 923–933.

Geochemical classification of the soil in a typical black soil area using the principal component analysis combined with K-means clustering algorithm

LIU Kai^{1,2,3}, DAI Hui-Min^{1,2,3}, LIU Guo-Dong^{1,2,3}, SONG Yun-Hong^{1,2,3}, LIANG Shuai^{1,2,3}, YANG Ze^{1,2,3}
(1. *Shenyang Center of China Geological Survey, Shenyang 110034, China*; 2. *Key Laboratory of Black Soil Evolution and Ecological Effect, Ministry of Natural Resources, Shenyang 110034, China*; 3. *Key Laboratory of Black Soil Evolution and Ecological Effect, Liaoning Province, Shenyang 110034, China*)

Abstract: The geochemical classification of soils is significant for agricultural and ecological regionalization. Based on the data on major elements in soil obtained from the multi-purpose regional geochemical survey, this study conducted the geochemical classification for a typical black soil area in northeast China using the principal component analysis combined with the K-means clustering algorithm (also referred to as the principal component clustering method). The results are as follows. The soil parent materials are the main factor controlling the characteristics of major elements in the soil. It is the most appropriate to divide the soil samples from the typical black soil area into five categories using the principal component clustering method. Various samples had significantly different major element contents ($P<0.05$). The geochemical classification results corresponded to the Quaternary geological units to a certain degree and can better reflect the actual distribution of soil parent materials. Moreover, the high SiO_2 content in the black soil area in the southern Songhua River indicates desertification, to which much attention should be paid in the protection of the black soil.

Key words: black soil area; major elements; geochemistry; parent material; desertification

(本文编辑:蒋实)