K-means 聚类方法在黑龙江省低山丘陵区坡耕地类型区划分中的应用

吕志学，孙雪文，刘凤飞

（黑龙江省水土保持科学研究所，黑龙江 哈尔滨 150070）

摘　要：[目的] 探索 K-means 聚类方法在黑龙江省坡耕地类型区划分上应用的可行性，为各市（县）坡耕地水土保持规划和治理提供依据。[方法] 利用 K-means 聚类方法对黑龙江省 70 个低山丘陵区市（县）进行科学区划。[结果] 低山区市（县）共 21 个，丘陵区县市共 49 个。[结论] 所选 16 个分类指标的显著性均小于 0.05，证明分类结果令人满意。

关键词：K-means 聚类方法；低山丘陵区；坡耕地；类型区划分

文献标识码：B　　**文章编号**：1000-288X(2015)01-0124-04　　**中图分类号**：S157

DOI:10.13961/j.cnki.stbctb.2015.01.023

Application of K-means Clustering Method in Regionalization of Slope Farmland in Low Mountain and Hill Area of Heilongjiang Province

LÜ Zhixue, SUN Xuewen, LIU Fengfei

(*Heilongjiang Provincial Soil and Water Conservation Scientific Research Institutes, Harbin, Heilongjiang* 150070, *China*)

Abstract: [Objective] To explore the feasibility of the K-means clustering method in the classification of farmland types in Heilongjiang Province and provide support for the planning and management of soil and water conservation in slope farmland. [Methods] We made a clustering of farmlands in 70 cities(or counties) in low mountain and hilly areas in Heilongjiang Province using K-means clustering method. [Results] The cities or counties in the low mountain is 21 and hill area is 49. [Conclusion] The 16 classification indexes were significant at the level of 0.05, indicating the fitness of classification.

Keywords: **K-means clustering method; low mountain and hill area; slope farmland; regionalization**

中国人口众多，人均资源较少，特别是作为人类赖以生存的耕地资源十分匮乏，人均只有 0.09 hm^2，为世界平均水平的 1/4，美国的 1/9。而且这些耕地资源中大多为低山丘陵区坡耕地，耕地质量差，环境恶劣，直接影响中国的粮食安全[1]。中国东北黑土区是目前世界上仅有的 3 大黑土区之一，该地区是中国的主要商品粮基地，年生产粮食约占全国粮食总产量的 20%，黑龙江省约占其中的 1/2[2]，如何提高作为全国重要商品粮基地—黑龙江省坡耕地的粮食产量，对确保中国粮食安全将起至关重要的作用。黑龙江省幅员面积较大，地貌类型多样，坡耕地水土流失各具特色，治理措施差异显著，特别是低山区坡耕地与丘陵区坡耕地。如低山区耕地作为山地的延伸部分，坡度大、汇流急，雨滴溅蚀地表土壤，而后出现超渗产流，产生水土流失。在治理措施上因该区域主要土壤以棕色针叶林土和山地暗棕壤为主，土层薄而不易于修筑水平梯田，多采用复合地埂[3]等措施进行坡耕地水土流失治理；丘陵区坡耕地坡长坡缓，汇流面积大，地表径流大，降雨出现超渗产流后易形成股状汇流，不仅面蚀严重，在坡面汇水线处更易发生沟蚀。治理措施上因坡耕地黑土层较厚，一般根据不同坡度选择梯田、地埂和改垄措施进行治理[4]。因此，如何科学界定黑龙江省低山区和丘陵区市县，对科学选择、布设水土保持措施将起至关重要的作用。本文在黑龙江省已经确定的 70 个低山丘陵区市县的基础上，选取影响坡耕地质量的 16 项指标，运用成熟的聚类方法，对低山丘陵区市县进行了分类，旨在为不同类型区进行科学治理创造条件。

1　材料与方法

1.1　划分指标的选取

影响坡耕地质量的因素很多，根据已掌握的资

收稿日期：2014-01-17　　**修回日期**：2014-02-17

资助项目：黑龙江省科技攻关项目“黑龙江省山丘区坡耕地综合整治措施体系研究”(GC05 B304)

第一作者：吕志学(1968—)，男(汉族)，黑龙江省哈尔滨市人，硕士，教授级高工，主要从事水土保持基础和应用研究。E-mail：sbslzx@163.com。

料，选取了市县高程＞200 m 耕地加权高程、年均气温、5—9 月平均气温、年降雨量、5—9 月降雨量、年日照时数、≥10 ℃积温、全氮含量、全磷含量、全钾含量、有机质含量、土地人口承载密度、粮食产量、农用化肥施用折纯量、农用机械总动力、耕地侵蚀面积占总耕地面积比共 16 项指标作为类型区划分指标。这些指标分别来自于黑龙江省遥感调查数据、黑龙江省农业地图集[5]及研究数据。

1.2　K-means 聚类方法原理

K-means 聚类是聚类方法中的一种，是常见 4 类聚类方法（划分方法、层次方法、基于密度的方法和基于网络的方法）中最典型的“划分方法”，又称作 K 中心聚类，属硬聚类算法。通常我们要预先确定分类数，然后才能进行这个聚类分析。基本思想是初始随机给定 K 个簇中心，按照最邻近原则把待分类样本点分到各个簇。然后按平均法重新计算各个簇的质心，从而确定新的簇心。一直迭代，直到簇心的移动距离小于某个给定的值或达到给定的迭代次数[6]。

K-Means 聚类算法主要分为 3 个步骤：

第 1 步是为待聚类的点寻找聚类中心。

第 2 步是计算每个点到聚类中心的距离，将每个点聚类到离该点最近的聚类中去。

第 3 步是计算每个聚类中所有点的坐标平均值，并将这个平均值作为新的聚类中心。

反复执行第 2，3 步，直到聚类中心不再进行大范围移动或者聚类次数达到要求为止。

1.3　具体操作步骤

具体操作步骤：（1）先准备好数据，在打开软件的菜单栏上选定“K-均值聚类”分类方法；（2）输入聚类用到的指标变量和市县名称；（3）设置聚类的类别数目。这个数目的确定主要依据工作经验和前人的研究证明。本研究欲将黑龙江省低山丘陵区市县划分成 2 个类型区，所以设置的类别数为 2；（4）在主对话框中，设置最大迭代次数，默认值一般为 10，但是数据量越大，迭代次数就应该越多，所以我们设置为 100；（5）点击“确定”按钮，开始运行数据，并输出数据结果。具体操作步骤详见图 1。

2　结果与讨论

黑龙江省低山丘陵区坡耕地类型划分初始聚类中心见表 1。它列出每一个类别初始的中心点，如低山区这个类别中，16 项指标中的“市（县）高程＞200 m 耕地加权高程”这个指标的初始中心点为－0.365 91，即该类别中所有市县的该项指标都是距离这个中心点相对较近，而距离丘陵区 0.347 78 这个中心点相对较远，其他 15 项指标也是如此，最终根据 16 项指标的“综合距离”确定某个市县应该划归为那个类别中，即确定初步分类结果。这些中心点都是 SPSS 自动生成的。

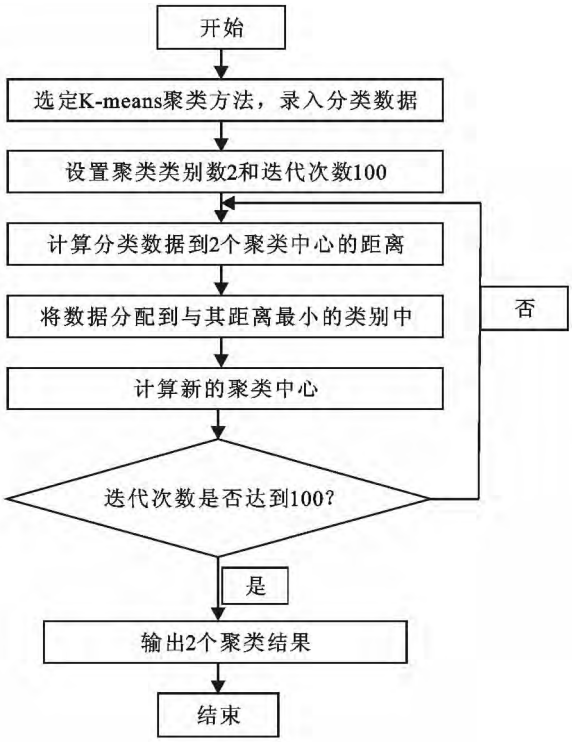

图 1　K-means 聚类具体操作步骤

表 1　黑龙江省低山丘陵区坡耕地类型划分初始聚类中心

指 标	分 区	
	低山区	丘陵区
高程＞200 m 耕地加权高程	－0.365 91	0.347 78
年均气温	0.833 99	－4.559 45
5—9 月平均气温	1.097 38	－4.151 59
年降雨量	1.783 29	－1.509 81
5—9 月降雨量	1.854 64	－1.761 65
年日照时数	－0.180 52	－1.444 14
≥10 ℃积温	0.953 42	－2.763 96
全氮含量	0.635 37	0.635 37
全磷含量	0.414 73	0.973 03
全钾含量	－0.330 47	0.633 39
有机质含量	－0.719 87	0.945 95
土地人口承载密度	0.672 11	－0.476 10
粮食产量	3.353 96	－1.146 72
农用化肥施用折纯量	3.544 72	－1.249 17
农用机械总动力	1.197 18	－1.403 73
耕地侵蚀面积占总耕地面积比	－0.281 58	0.267 83

黑龙江省低山丘陵区坡耕地类型划分最终聚类中心见表 2。它列出每一个类别最终的中心点，如低山区这个类别中，16 项指标中的“市县高程＞200 m 耕地加权高程”这个指标的最终中心点为－0.188 21，即该

类别中所有市县的该项指标都是距离这个中心点相对较近，而距离丘陵区0.690 09这个中心点相对较远的，其他15项指标也是如此，最终根据16项指标的“综合距离”确定某个市县应该划归为那个类别中，即确定最终分类结果。依据这些最终中心点，即可自动生成最终聚类结果表。

表3是我们根据最终聚类结果表整理后的聚类结果。从表3中我们可以看出，低山区这个类别中主要包含尚志市、鸡西市、鸡东县等21个市县，主要位于黑龙江省北部和东南脚的大小兴安岭及张广才岭、老爷岭低山区。该区耕地为山地的延伸部分，坡度大、汇流急，耕地总面积27 329.55 km^2，耕地水土流失面积14 842.50 km^2，占耕地总面积的54.31%，主要土壤以棕色针叶林土和山地暗棕壤为主；丘陵区主要包含哈尔滨市、方正县、木兰县等49个市县，主要位于松嫩平原和三江平原内。该区耕地总面积104 726.52 km^2，耕地水土流失总面积36 232.48 km^2，占耕地总面积的34.60%。该区地势波状起伏，坡长坡缓，汇流面积大，地表径流大，主要土壤为黑土、黑钙土。

表2 黑龙江省低山丘陵区坡耕地类型最终聚类中心

指 标	分 区	
	低山区	丘陵区
高程＞200 m耕地加权高程	−0.188 21	0.690 09
年均气温	0.359 85	−1.319 43
5—9平均气温	0.359 59	−1.318 49
年降雨量	−0.061 24	0.224 56
5—9月降雨量	−0.074 05	0.271 51
年日照时数	0.096 94	−0.355 43
≥10 ℃积温	0.408 35	−1.497 30
全氮含量	−0.245 02	0.898 40
全磷含量	−0.174 01	0.638 05
全钾含量	0.011 27	−0.041 31
有机质含量	−0.242 84	0.890 42
土地人口承载密度	0.156 29	−0.573 05
粮食产量	0.190 62	−0.698 94
农用化肥施用折纯量	0.192 70	−0.706 56
农用机械总动力	0.046 14	−0.169 16
耕地侵蚀面积占总耕地面积比	−0.102 58	0.376 12

表3 黑龙江省低山丘陵区坡耕地类型聚类结果

类 别	聚类成员	成员数量/个	分区
1	尚志市、鸡西市、鸡东县、伊春市、嘉荫县、呼玛县、塔河县、漠河县、牡丹江市、海林市、宁安市、穆棱市、东宁县、林口县、绥芬河市、黑河市、嫩江县、逊克县、孙吴县、北安市、五大连池市	21	低山区
2	哈尔滨市、方正县、木兰县、通河县、延寿县、宾县、阿城市、五常市、巴彦县、依兰县、鹤岗市、铁力市、勃利县、齐齐哈尔市、依安县、克山县、克东县、拜泉县、讷河市、富裕县、甘南县、龙江县、密山市、虎林市、萝北县、双鸭山市、集贤县、友谊县、宝清县、饶河县、佳木斯市、汤原县、桦川县、富锦市、同江市、抚远县、桦南县、七台河市、肇州县、海伦市、明水县、望奎县、青冈县、安达市、兰西县、肇东市、北林区、庆安县、绥棱县	49	丘陵区
合 计		70	

表4给出了计算机的原始方差分析结果。从表4中看出，自由度为一组数据中能独立变化数据的多少。本文欲将70个市县划分成2类，因此总的自由度为70−1=69，即最多有69个市县可自由划分为1类，其余自然确定为另一类。并且，因为只划分为2类，所以类间自由度为1，即最多只有1类可自由确定，剩下的自然为另一类；类内自由度为68，即每类中最多有68个能独立变化的数据[7]。F检验值即为类间平均方差与类内平均方差的比值，该F值越大，即类间均方大于类内均方，也就是类间变异量大于类内变异量，说明各类间的差异远超出总期望值离差，代表各类的平均数存在明显的差异，同时，F值越大说明其在分类中所起的作用越大[8]。如表中≥10 ℃积温、年均气温等。判断文中所选影响坡耕地质量的16项指标是否具有代表性取决于其显著性检验结论，显著性水平一般确定为0.05，即当显著性计算结果小于0.05时，所选指标对整体具有很好的代表性，否则代表性不强[9]。从表4中看，所选分类指标的显著性均小于0.05，说明该分类指标在统计学有意义，即所选分类指标具有很好的代表性，分类结果有效。

3 结 论

最初我们选择了21个分类指标进行上机运算，经多次反复，筛选确定其中的16个指标为最终分类指标，指标涵盖了地形地貌、自然条件、经济条件、社会条件和水土流失等方面。其中“市县高程＞200 m耕地加权高程”来自于课题研究结论，“耕地侵蚀面积占总耕地面积比”来自于遥感调查数据，其他来自于黑龙江省农业地图集。该图集资料比较完整、系统，虽然年代较早，但仍是目前最新的图集，其中的指标能很好的代表黑龙江省70个低山丘陵区市县耕地整体质量。

表 4　黑龙江省低山丘陵区坡耕地类型原始方差分析

指 标	类 间		类 内		F 检验	显著性
	平均方差	自由度	平均方差	自由度		
市县高程＞200 m 耕地加权高程	9.091	1	0.881	68	10.319	0.002
年均气温	33.235	1	0.526	68	63.191	0.000
5—9 月平均气温	33.188	1	0.527	68	63.018	0.000
年降雨量	3.961	1	0.801	68	4.945	0.035
5—9 月降雨量	4.406	1	0.894	68	4.928	0.039
年日照时数	4.410	1	0.779	68	5.661	0.030
≥10 ℃积温	42.800	1	0.385	68	111.084	0.000
全氮含量	15.409	1	0.788	68	19.552	0.000
全磷含量	7.772	1	0.900	68	8.632	0.005
全钾含量	4.031	1	0.814	68	4.952	0.034
有机质含量	15.136	1	0.792	68	19.109	0.000
土地人口承载密度	6.269	1	0.923	68	6.796	0.011
粮食产量	9.326	1	0.878	68	10.628	0.002
农用化肥施用折纯量	9.531	1	0.875	68	10.898	0.002
农用机械总动力	4.545	1	0.807	68	5.632	0.032
耕地侵蚀面积占总耕地面积比	3.703	1	0.775	68	4.778	0.042

K-means 聚类算法是一种比较成熟的聚类方法，本文将该方法首次应用在黑龙江省低山丘陵区市县划分上，采用计算机程序进行运算，获得了较为理想满意的分类结果，通过对划分的 2 个类型区进一步统计分析。

可以看出低山区与丘陵区的平均数差异显著，如山区平均土地人口承载密度为 68 人，市县高程＞200 m耕地加权高程为 341 m，耕地侵蚀面积占总耕地面积比为 54.31%。

丘陵区平均土地人口承载密度为 156 人，市县高程＞200 m 耕地加权高程为 256 m，耕地侵蚀面积占总耕地面积比为 34.60%，也进一步验证了分类结果的可信性，这对今后不同类型区坡耕地采取分类治理具有十分现实的意义。

同时由于影响坡耕地质量的因素很多，很多因素还在不断变化，因此今后应根据掌握的最新数据进一步对分类结果进行核试验证。

[参 考 文 献]

[1] 崔秀珍，吴国梁. 新时期我国耕地资源总量动态平衡的维持与可持续利用的途径[J]. 安徽农业科学，2006，34(2)：298-299.

[2] 刘兴土，阎百兴. 东北黑土区水土流失与粮食安全[J]. 中国水土保持，2009(1)：17-19.

[3] 屈远强，勇丽波. 穆棱市水保生态建设措施体系构建与实践[J]. 水土保持应用技术，2010(3)：24-26.

[4] 王树清. 拜泉县生态农业发展战略与实践[J]. 中国生态农业学报，1995(4)：77-78，76.

[5] 黑龙江省地图集编纂委员会. 黑龙江省农业地图集[M]. 哈尔滨：哈尔滨地图出版社，1999.

[6] 张建辉. K-means 聚类算法研究及应用[D]. 武汉：武汉理工大学，2007.

[7] 卢淑华. 社会统计学[M]. 北京：北京大学出版社，2005.

[8] 靳庭良，张宝青. 回归分析中 t 检验与 F 检验关系的进一步探讨[J]. 统计与决策，2009(21)：7-9.

[9] 孙逸敏. 利用 SPSS 软件分析变量间的相关性[J]. 新疆教育学院学报，2007，23(2)：120-123.