

流域尺度土壤厚度的模糊聚类与预测制图研究^①

王改粉^{1,2}, 赵玉国¹, 杨金玲^{1,2}, 张甘霖^{1*}, 赵其国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 中国科学院研究生院, 北京 100049)

摘要: 基于土壤厚度与景观位置和特征之间的关系, 运用模糊 c 均值聚类(FCM)方法对西苕溪流域的土壤厚度分布进行了空间预测。选取高程、坡度、平面曲率、剖面曲率、径流强度系数和地形湿度指数 6 个地形因子进行模糊聚类, 根据相应的聚类参数将流域地形组合分为 8 类。利用部分调查获得的土壤剖面数据, 结合样点属性和专家经验为典型区赋值, 最后由加权平均得到流域土壤厚度预测图。验证结果表明, FCM 方法可以对地形因子组合进行有效合理的分级, 其预测精度较高, 模型的稳定性较好, 是一种低成本高效率的制图方法。该方法在土壤厚度预测方面具有一定的可靠性。

关键词: 模糊 c 均值聚类; 聚类参数; 土壤厚度; 数字土壤制图

中图分类号: P934

土壤厚度是土壤肥力存在和植物生长的基础, 也是土壤退化及土地生产力评价的一个重要指标。它可以直接反映土壤的发育程度和土壤矿质元素的储存容量, 也是判定土壤侵蚀程度的主要指标^[1-3]。在小流域尺度上, 土壤水分特征与土壤厚度有密切关系, 且与其他土壤性质相比, 土壤厚度可解释更多的土壤水分时空变异性^[4-5]; 土壤厚度也较多地应用到了水分模拟中, 如 Pellenq 等^[6]利用土壤厚度和地形因子模拟土壤水分的空间分布; 徐佩等^[7]认为土壤厚度是影响壤中流产生的重要原因; 土壤水分是研究和模拟地表径流、土壤侵蚀和溶质运移等过程的基础, 也是区域植被和景观格局改善的关键环境因子^[8-10], 因此, 研究土壤厚度的分布又可以这些研究提供必要的的数据支撑。

近年来, 数字土壤制图方法在全世界范围内得到了非常广泛的研究, 覆盖了不同地区、不同环境下各种土壤属性及土壤类型的预测制图^[11-18]。为了克服传统土壤制图中的硬性界限, 更加准确地描述土壤属性和类型的空间变化, 很多学者在数字土壤制图研究中应用了模糊数学方法, 其中模糊 c 均值聚类(fuzzy-c-means, FCM)是应用较为广泛的方法之一。如 Triantafilis 等^[19]对土壤各层次的属性聚类, 并预测土壤层次的空间分布; 赵量等^[11]、孙孝林等^[12]、杨琳等^[13-14]利用 FCM 算法对地形因子聚类, 根据质心属性及隶属度分布进行土壤属性及土壤类型制图; 师华定等^[20]通过模糊聚类提取风蚀和环境关系知识, 对土壤风蚀危险度进行预测。

模糊 c 均值聚类(FCM)以目标函数为基础, 通过优化求解, 获得数据集的模糊划分和聚类^[16, 21], 它是一种非监督分类方法, 它根据分类现象在多属性空间的距离来进行自动分类, 得到各聚类类别与各聚类中心的一组相似度向量, 称之为隶属度(fuzzy membership)^[22]。该方法用连续划分的模糊隶属度代替了传统模型中非此即彼的二值假设, 因此可以更准确地表达连续性地理现象在空间的渐变和过渡, 符合土壤系统的连续性变化本质。

本研究以太湖流域补给区之一的浙江西苕溪流域为研究区, 利用 FCM 方法, 对流域土壤厚度进行了模拟制图与精度评价, 以期为该流域的水文模拟、营养盐输出等研究提供依据。

1 研究区域与方法

1.1 研究区概况

西苕溪流域位于浙江省西北部的天目山北麓, 北临近太湖, 属太湖流域中部丘陵区, 包括安吉县大部和长兴县的一部分, 地理位置在 30°22'58" ~ 30°53'42"N, 119°13'52" ~ 119°57'12"E, 流域周长 340 km, 面积 2071 km²。流域为亚热带湿润季风气候带, 年均气温 12℃ ~ 16℃, 年均降雨量 1300 mm。该地区山丘盆谷相间, 具有山、丘、岗、谷、沟、盆地和平原多种地貌类型。地层复杂, 岩性多变, 除第三系以外, 从震旦系至第四系地层均有发育; 有最古老的震旦系的凝灰质粉砂岩、白云质灰岩, 寒武系硅质岩,

^①基金项目: 中国科学院知识创新工程重要方向项目(KZCX2-YW-Q10-3)、江苏省自然科学基金项目(BK2008058)和国家自然科学基金项目(40771092)资助。

* 通讯作者(glzhang@issas.ac.cn)

作者简介: 王改粉(1986—), 女, 河南洛阳人, 硕士研究生, 主要从事资源演变分异规律与预测制图等方面的研究。E-mail: gfwang@issas.ac.cn

奥陶系的砂页岩, 志留系和泥盆系的砂岩, 二迭系的灰岩, 侏罗系的凝灰岩和白垩系的紫色砂砾岩。就面积而言, 以沉积岩为主, 其次是火成岩, 还有块状分布的燕山期的侵入岩出露。该地区土壤类型主要有: 红壤(湿润富铁土)、黄壤(锥形土)、水稻土(水耕人为土)、粗骨土(新成土)、紫色土(锥形土)和黑色石灰土(淋溶土), 还有小面积的新积土、潮土和石质土(锥形土)。

1.2 数据来源

基础数据包括基于 SRTM 的 $90\text{ m} \times 90\text{ m}$ 数字高程模型 (DEM) 和 44 个实际采样点 (图 1)。在样点布设时, 尽量确保覆盖不同的土壤类型和地形组合, 采样时记录每个样点的地形景观特征、植被、利用方式和母质等, 并根据土壤发生学特征, 对土壤剖面进行发生层划分, 详细记载每个发生层的形态特征, 并对每个发生层采样, 其中水稻土的厚度以出现地下水为止。

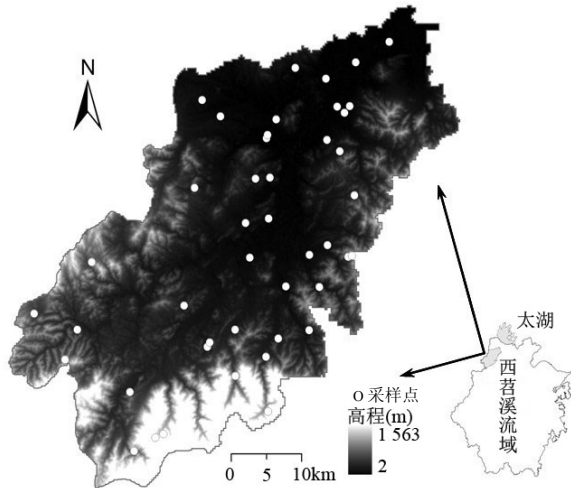


图 1 研究区地理位置和 DEM 及样点分布

Fig. 1 Location, DEM and sampling sites of the study area

1.3 研究方法

1.3.1 数据及预处理 根据影响地表物质循环的主要因素, 将影响该地区土壤厚度的地形因子详细划分成以下因子: 高程、坡度、平面曲率、剖面曲率、径流强度系数和地形湿度指数。其中, 高程表示地面物质流动的趋势和能量, 坡度影响地表物质与能量流动强度, 平面曲率反映物质和能量的相对集中或分散, 剖面曲率反映物质和能量流动的相对加速或减速, 径流强度系数反映土壤中水流侵蚀强度, 地形湿度指数反映土壤中水分状况。

由于 FCM 算法对数据的极端值和异常值比较敏感, 输入的异常值会影响聚类结果, 甚至会导致错误, 所以, 应将这些极值去掉以增强聚类结构。对每层数

据进行基本的统计分析, 观察数据分布情况, 去掉数据两端的离群值以增强聚类结构。

1.3.2 模糊 c 均值聚类 FCM 算法是用隶属度确定每个数据点属于某个聚类的程度的一种聚类算法, 其基本原理是利用统计方法计算多属性空间中的距离使得每个数据点距离自己所属类别中心点最近, 并且以此给每个数据点分配隶属度。

土壤属性图的生成, 是通过典型属性值和各点与各聚类中心的隶属度的加权平均得到, 这主要是基于当地土壤的成土环境与给定土壤类型的成土环境相似, 则当地土壤的性状就与给定的土壤类型性状相似的假设^[23]:

$$V_i = \sum_{k=1}^c u_{ik} \cdot V^k / \sum_{k=1}^c u_{ik}$$

式中, V_i 表示某点 i 的土壤属性值, V^k 代表各聚类中心 k 的土壤属性值, u_{ik} 代表点 i 在类型 k 上的隶属度, c 代表这一地区给定的土壤类型的总数。

把各个聚类中心的土壤属性代入上式, 即可得到该研究区土壤属性连续分布图。

1.3.3 聚类参数 在 FCM 算法中, 模糊度指数 (m) 和聚类类别数 (c) 是两个关键参数。对于模糊度指数 m 和聚类类别数 c 的选择, 很多学者都进行了研究^[24-28], 1976 年 Bezdek^[24]先给出了 FCM 算法在 $m = 2$ 时的物理解释, 1995 年 Pal 和 Bezdek^[27]根据多个聚类有效性函数的表现认为 m 的最佳选取区间为 $[1.5, 2.5]$, 一般情况下可取 2。为了确定最佳分类数 c , 很多学者也进行了研究, 其中 Xie 和 Beni^[25]认为, 用于选择最佳分类数的指标应该能指示分类结果的类内紧密和类间分离程度, 并设计了分类距离 (separate distance, S) 来选择最佳分类数, S 越小, 意味着类内紧密、类间分离的程度越高。因而, 在不同分类下, S 值最小时对应的分类数为最佳分类数。

据此, 在本研究中, 我们选取模糊度指数 m 为 2, 并对 $m = 2$ 时不同聚类类别数的分类距离进行统计, 得到分类距离随着分类类别数的变化曲线, 最终确定本研究中 $m = 2$ 、 $c = 8$ 。

1.3.4 为典型地形组合赋值 从样点中选取训练集, 具体做法为: 分别提取出各个点在 8 类地形组合图上的隶属度, 统计 44 个点在各个类型上的隶属度, 用隶属度较高的点的平均厚度为该类型赋值, 若所有点在某一类的隶属度都很低, 则采用专家经验为其赋值。用于赋值的点做为训练点, 其余的点则为验证点。

1.3.5 精度评价方法 精度评价时采用的指标有: 散点图、相关系数、平均误差 (ME)、平均绝对误差 (MAE)、均方根误差 (RMSE)、变异系数 (CV) 和吻合度 (AC),

并对 31 个验证点的相对误差进行分级统计。

其中,ME 和 MAE 是对模型的总体准确度的度量,其值越接近于 0,说明模型的总体偏差越小;RMSE 和 CV 是对模型精度及稳定性的度量,值越小说明模型的精度和稳定性越高;吻合度(AC)描述模型预测值与实际值的吻合程度,其值在 0 到 1 内变化,越靠近 1,表明预测值与实际值之间吻合程度越高,模型

越好,反之则差。

2 结果与讨论

2.1 模糊聚类

在 $m=2$ 、 $c=8$ 的参数组合下运行 FuzMe^[29]软件,聚类因子为:高程、坡度、平面曲率、剖面曲率、径流强度系数和地形湿度指数,得到 8 个聚类结果(表 1)。

表 1 8 类地形因子组合及其所代表的典型空间单元特征

Table 1 Eight combinations of terrain factors

类别	高程 (m)	坡度 (°)	平面曲率	剖面曲率	径流强度	地形湿度	地形描述
class1	464	20	0.04	-0.04	8.73	9.70	高丘, 陡坡, 沿剖面微凹
class2	203	14	0.02	0.03	8.83	10.20	低丘, 微凸
class3	656	22	0.02	-0.02	8.93	9.81	高丘, 陡坡
class4	320	18	0.03	-0.02	8.69	9.77	中丘, 沿剖面微凹
class5	1 122	21	0.06	-0.10	8.53	9.45	山地, 陡坡, 沿剖面微凹
class6	20	2	0.00	0.02	8.68	12.83	平原, 湿度指数最大
class7	869	23	0.03	-0.04	8.85	9.71	低山, 陡坡
class8	109	8	0.01	0.05	8.99	11.00	低丘, 缓坡, 微凸

据《安吉土壤》^①记载,自第三纪末开始的喜马拉雅造山运动,承袭了历史上地壳运动的替演,对天目山脉产生了较为深刻的影响,其升降运动建造了境内多级地貌,依次可分为 5 级古剥离平面:1 000~1 300 m、800~900 m、600 m、350 m 和 200 m。可以看出,由 FCM 算法得到的地形因子组合可以较好地反映这种分级,并加入平原地形,从而使地形组合更加全面。

2.2 聚类类型的典型厚度赋值

统计 44 个点在 8 个类型上的隶属度,其中在 class1 和 class5 两个类型上的隶属度都较低,所以采用专家经验为这两个类型赋值,其他 6 个类型均用隶属度较高的点的平均厚度赋值,从而得到 8 个类型对应的典型厚度值(表 2)。用于赋值的点为训练点,剩余的点则为验证点(图 2)。

表 2 各聚类类型对应的典型土壤厚度值

Table 2 Typical value of soil thickness of each soil class by clustering

类别	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8
厚度 (cm)	70	42	67	90	10	100	12	85

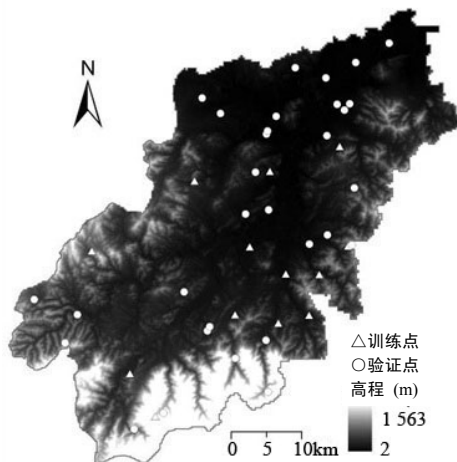


图 2 训练点及验证点分布

Fig. 2 Training sites and validation sites of the study area

2.3 土壤厚度预测制图

由上述 8 类典型地形组合及其聚类中心的属性值,根据土壤景观模型原理,在 ArcGIS 中进行栅格计算,即可得到整个流域内的土壤厚度预测图(图 3)。随着地形变化,土壤厚度也有明显的变化趋势:南部高山沟谷区,山岭平均高度都在海拔 800 m 以上,在其自然地上发育的土壤土层很薄,多在 20 cm 左右;中部及北部河谷平原区,河床开阔,比降减少,河流流速大大减缓,沉积物越积越深厚,该地区预测厚度多在 100 cm 左右;分布于两者之间的低山及丘陵区,

海拔差异较大,其土层厚度差异也较大,一般海拔较高且坡陡地区土层薄,缓坡地带土层较深厚。根据本研究所得到的土壤厚度结果在空间中呈现连续的

①安吉县土壤普查办公室.安吉土壤.1985

分布,具有较高的分辨率(90 m),比较好地体现了土壤随地形因子变化的特点,并且任意位置都能有模拟的结果,十分有利于相关的土壤水文特征模拟和计算。

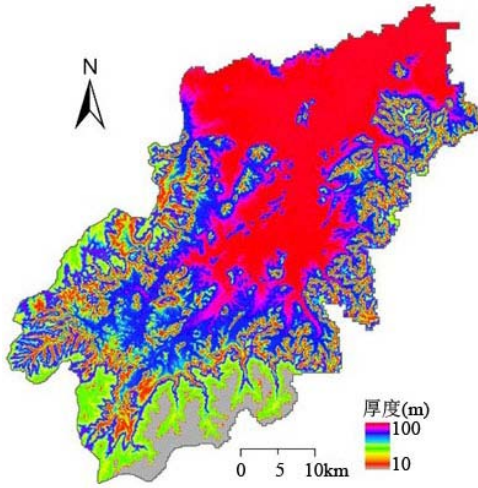


图 3 研究区土壤厚度分布

Fig. 3 Predicted soil depth of the study area

2.4 精度评价

本研究中,对预测精度进行评价,其变异系数(CV)为0.15,变异系数较小,吻合度(AC)为0.92,接近于1,说明模型的精度与稳定性较高。观察验证点散点图(图4)也可看出模型的预测精度较高。

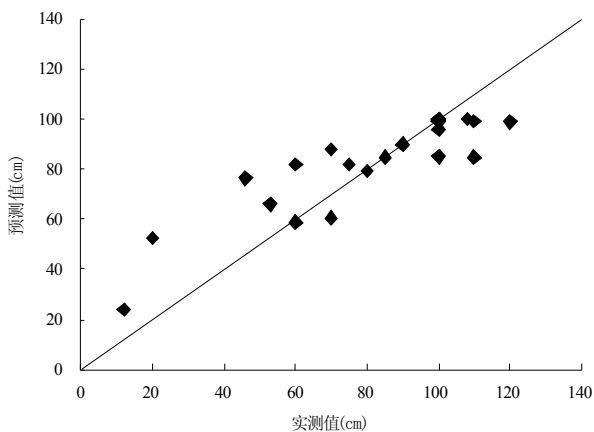


图 4 验证点散点图

Fig. 4 Scatter diagram of validation points

与验证点上的实测值相比,土体厚度预测值的平均误差(ME)为0.91 cm,平均绝对误差(MAE)为7.00 cm,均方根误差(RMSE)为11.81 cm。在Florinsky

等^[30]的类似研究中,不同地形部位上(分散区、过渡区、汇集区),土体厚度预测的平均误差分别为6、10和34 cm,整体平均绝对误差为17 cm。可以看出,本研究的预测达到了一定的精度。在Zhu和Band^[22]的研究中,土壤属性预测值的均方根误差是观测平均值的26%,低于传统土壤图的38%,而本研究中,均方根误差是观测平均值的15%。由此可以看出,本研究中预测误差的变异较小,预测值的可信度较高。

对实测值和预测值进行统计,来评价本研究的可靠性。两者相关性达到极显著,其相关系数为0.90($p < 0.01$, $n = 31$)。对相对误差进行统计可知:除去个别样点的误差较大外,大部分验证点的相对误差在25%以下,其中相对误差小于20%的点有24个,占总验证点的77.4%;相对误差小于15%的点有22个,占总验证点的71.0%;相对误差小于10%的点有21个,占总验证点的67.7%;相对误差小于5%的点有18个,占总验证点的58.1%。

按相对误差排序后发现(表3),水田预测精度较高,而其他土地利用类型预测结果误差分异较大。该流域内水田多位于沉积层深厚的河谷平原,其地形平坦,成土环境比较稳定,因而由地形预测的结果可以较准确地反映其实际情况。林地和其他类型在丘陵上的分布较广,而丘陵地土层厚度差异较大,在坡度较大的地区水土流失严重,土层较薄,在缓坡地带,土层较深厚,且多为经济林地和旱田,茶园面积也较大。可以明显看出,在土层较厚的地区其预测精度均较好,而相对误差在25%以上的点均为土壤厚度观测值较小的点,其土地利用类型均为林地,并且其土壤厚度预测值比实际观测值偏大,这可能有两个原因,一是与当地的林地利用有关,由于西苕溪流域内主要以经济林为主,虽然当地的表观植被覆盖度较高,但存在一定的林下流现象,导致土壤厚度变薄;二是与本研究的模型适应性有关,模糊聚类算法对数据有一定集中作用,而野外土壤形成因素复杂,土壤厚度的离散程度较高,所以会导致观测值较小的点的预测结果偏大。赵量等^[11]的研究也证明了模糊聚类算法在土壤发育较好、土层较厚的地区预测效果较好,适用性较强。

必须指出的是,本研究只选择了地形参数,尚未考虑母质、母岩、土地利用和植被参数,当局部地区因为母质、母岩、土地利用方式及地表植被覆盖等的不同对土壤厚度产生较大影响时,仅使用地形参数进行预测会影响预测的精度。因此,我们将在下一步研究中加入更多因子,以期得到更好的预测结果。

表3 验证点相关信息

Table 3 Related informations of validation points

编号	土壤类型	土地利用类型	实测值 (cm)	预测值 (cm)	相对误差
AJ04	红壤	山地旱田	100	99.96	0%
AJ19	紫色土	平原水田	100	99.95	0%
AJ22	红壤	丘陵旱田	100	99.93	0%
AJ02	粗骨土	丘陵水田	100	99.90	0%
AJ42	红壤	林地	85	84.78	0%
AJ08	红壤	林地	90	89.76	0%
AJ28	黑色石灰土	林地	100	99.62	0%
AJ25	红壤	林地	100	99.46	1%
AJ26	潴育型水稻土	平原水田	100	99.24	1%
AJ16	渗育型水稻土	平原水田	100	99.23	1%
AJ29	潴育型水稻土	平原水田	100	99.23	1%
AJ35	潴育型水稻土	平原水田	100	99.21	1%
AJ15	渗育型水稻土	平原水田	100	99.19	1%
AJ43	潴育型水稻土	平原水田	100	99.12	1%
AJ01	红壤	林地	80	79.18	1%
AJ34	红壤	平原水田	100	98.95	1%
AJ11	红壤	林地	60	58.80	2%
AJ31	红壤	桑茶果园	100	95.91	4%
AJ17	红壤	疏林地	108	99.91	7%
AJ06	黑色石灰土	林地	75	81.85	9%
AJ14	潴育型水稻土	平原水田	110	99.20	10%
AJ30	黑色石灰土	丘陵旱田	70	60.39	14%
AJ44	潴育型水稻土	林地	100	84.97	15%
AJ07	紫色土	疏林地	120	98.83	18%
AJ10	红壤	林地	110	84.78	23%
AJ39	黄壤	林地	53	65.86	24%
AJ05	黑色石灰土	林地	70	87.93	26%
AJ13	红壤	林地	60	81.90	37%
AJ36	黑色石灰土	林地	46	76.36	66%
AJ40	粗骨土	林地	12	24.02	100%
AJ37	石质土	林地	20	52.46	162%

3 结论

本研究利用少量样点, 进行一个中型规模的流域土壤厚度模拟制图, 结果表明, FCM 方法可以对地形因子组合进行有效合理的分级, 预测精度较高, 模型的稳定性较好, 是一种低成本高效率的制图方法。本研究只选择了地形参数, 尚未考虑母质母岩、土地利用及植被覆盖等其他参数, 对于更大区域的土壤厚度分布研究还应该考虑母质风化特点等基础上进一步深入。

致谢: 中国科学院南京地理与湖泊研究所李恒鹏博士为本研究提供了研究区土壤类型分布图, 特此致谢。

参考文献:

- [1] 刘创民, 李昌哲, 史敏华, 梁海英. 多元统计分析在森林土壤肥力类型分辨中的应用. 生态学报, 1996, 16(4): 444-447
- [2] 张本家, 高岚. 辽宁土壤之土层厚度与抗蚀年限. 水土保持研究, 1997, 4(4): 57-59
- [3] Yanda PZ. Use of soil horizons for assessing soil degradation and reconstructing chronology of degradation processes: The case of Mwisanga Catchment, Kondoa, central Tanzania. *Geomorphology*, 2000, 34: 209-225
- [4] Rhoton FE, Lindbo DL. A soil depth approach to soil quality assessment. *Journal of Soil and Water Conservation*, 1997, 52(1): 66-72

- [5] 陈佳, 史志华, 李璐, 罗璇. 小流域土层厚度对土壤水分时空格局的影响. *应用生态学报*, 2009, 20(7): 1565-1570
- [6] Pellenq J, Kalma J, Boulet G, Saulnier GM, Wooldridge S, Kerr Y, Chehbouni A. A disaggregation scheme for soil moisture based on topography and soil depth. *Journal of Hydrology*, 2003, 276(1/4): 112-127
- [7] 徐佩, 王玉宽, 傅斌, 王道杰, 王先拓, 王勇强. 紫色土坡耕地壤中产流特征及分析. *水土保持通报*, 2006, 26(6): 14-18
- [8] Meyles E, Williams A, Ternan L, Dowd J. Runoff generation in relation to soil moisture patterns in a small Dartmoor catchment, Southwest England. *Hydrological Processes*, 2003, 17: 251-264
- [9] Entekhabi D, Rodriguez-Iturbe I, Castelli F. Mutual interaction of soil moisture state and atmospheric processes. *Journal of Hydrology*, 1996, 184(1/2): 3-17
- [10] Chen LD, Huang ZL, Gong J, Fu BJ, Huang YL. The effect of land cover/vegetation on soil water dynamic in the hilly area of the loess plateau, China. *Catena*, 2007, 70: 200-208
- [11] 赵量, 赵玉国, 李德成, 孙孝林, 张甘霖. 基于模糊集理论提取土壤-地形定量关系及制图应用. *土壤学报*, 2007, 44(6): 961-967
- [12] 孙孝林, 赵玉国, 赵量, 李德成, 张甘霖. 应用土壤-景观定量模型预测土壤属性空间分布及制图. *土壤*, 2008, 40(5): 837-842
- [13] 杨琳, 朱阿兴, 李宝林, 秦承志, 裴韬, 刘宝元, 李润奎, 蔡强国. 应用模糊c均值聚类获取土壤制图所需土壤-环境关系知识的方法研究. *土壤学报*, 2007, 44(5): 784-791
- [14] 杨琳, 朱阿兴, 秦承志, 李宝林, 裴韬, 刘宝元. 运用模糊隶属度进行土壤属性制图的研究—以黑龙江鹤山农场研究区为例. *土壤学报*, 2009, 46(1): 9-15
- [15] Zhu AX, Yang L, Li BL, Qin CZ, Pei T, Liu BY. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 2010, 155: 164-174
- [16] McBratney AB, Odeh IOA. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, 1997, 77(2/4): 85-113
- [17] 檀满枝, 陈杰, 徐方明, 郑海龙, 张学雷. 基于模糊集理论的土壤重金属污染空间预测. *土壤学报*, 2006, 43(3): 389-396
- [18] Amini M, Afyuni M, Fathianpour N, Khademi H, Flühler H. Continuous soil pollution mapping using fuzzy logic and spatial interpolation. *Geoderma*, 2005, 124(3/4): 223-233
- [19] Triantafyllis J, Ward WT, Odeh IOA, McBratney AB. Creation and interpolation of continuous soil layers classes in the Lower Namoi Valley. *Soil Science Society of America Journal*, 2001, 65: 403-413
- [20] 师华定, 高庆先, 齐永清, 刘纪远, 胡云峰. 蒙古高原土壤风蚀危险度的 FCM 模糊聚类研究. *自然资源学报*, 2009, 24(5): 881-889
- [21] 高新波. 模糊聚类分析及其应用. 西安: 西安电子科技大学出版社, 2004
- [22] Zhu AX, Band LE. A knowledge-based approach to data integration for soil mapping. *Canadian Journal of Remote Sensing*, 1994, 20(4): 408-418
- [23] Zhu AX, Band LE, Vertessy R, Dutton B. Derivation of soil properties using a soil land inference model (SoLIM). *Soil Science Society of America Journal*, 1997, 61: 523-533
- [24] Bezdek JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981
- [25] Xie XL, Beni G. A validity measure for fuzzy clustering. *IEEE transactions on pattern analysis and machine intelligence*, 1991, 13(8): 841-847
- [26] Odeh IOA, Chittleborough DJ, McBratney AB. Soil pattern recognition with fuzzy-c-means: Application to classification and soil-landform interrelationships. *Soil Science Society of America Journal*, 1992, 56: 505-516
- [27] Pal NR, Bezdek JC. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, 1995, 3(3): 370-379
- [28] 孙孝林, 赵玉国, 张甘霖, 李德成. 预测性土壤有机质制图中模糊聚类参数的优选. *农业工程学报*, 2008, 24(9): 31-37
- [29] Minasny B, McBratney AB. FuzME version 3.0, Australian Centre for Precision Agriculture, The University of Sydney, Australia. 2002. <http://www.usyd.edu.au/agriculture/acpa/software/fuzme>
- [30] Florinsky IV, Eilers RG, Manning GR, Fuller LG. Prediction of soil properties by digital terrain modeling. *Environmental Modelling & Software*, 2002, 17: 295-311

Prediction and Mapping of Soil Depth at A Watershed Scale with Fuzzy-c-means Clustering Method

WANG Gai-fen^{1,2}, ZHAO Yu-guo¹, YANG Jin-ling^{1,2}, ZHANG Gan-lin¹, ZHAO Qi-guo¹

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

2 *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: Soil depth of the west Tiaoxi catchment was predicted using fuzzy c-means clustering(FCM) based on the relationships between soil depth and landscape parameters. Six terrain factors, i.e., elevation, slope, planform curvature, profile curvature, runoff intensity and topographic wetness index were clustered, then the whole catchment was classified into eight combinations of these factors. Typical soil depths from the training soil dataset, combined with attribute of samples and expert knowledge, were assigned to each cluster center. Soil depth map was predicted with weighted average model. Results showed that, FCM method could rationally and effectively classify the combination of terrain factors, and it is a low cost and high efficiency mapping method with satisfactory prediction precision and model stability and could be possible applied to areas with the similar landscape conditions.

Key words: Fuzzy c-means clustering, Clustering parameters, Soil depth, Digital soil mapping