基于少量典型样点土壤属性空间分布推测模型中的土壤属性参数敏感性分析

卢岩君[1,2],秦承志[2],邱维理[1],朱阿兴[2]

(1. 北京师范大学地理学与遥感科学学院,北京 100875; 2. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室,北京 100101)

摘要: 针对基于少量典型样点土壤属性空间分布推测模型中的土壤属性参数敏感性问题,以坡位渐变信息结合典型土壤样点的加权平均模型为例,利用东北地形平缓小流域的土壤表层有机质含量样点集,使用阶乘设计、箱线图分析、扰动分析法和本文新设计的 MR 指数评价该模型的参数敏感性。结果表明,该模型中土壤属性参数敏感性较大,其大小与典型样点空间分布有关。敏感性主要由应用该模型时采用的坡位分类体系的不确定性引起。该文的分析方法可用于对基于少量典型样点的土壤属性空间分布推测模型进行参数敏感性综合分析。

关 键 词: 典型样点; 推测模型; 土壤属性参数; 敏感性分析; 模糊坡位

中图分类号: S159+3 **文献标识码**: A **文章编号**: 1000-0690(2011)12-1549-06

精细尺度下土壤属性空间分布信息在自然地理中有广泛应用[1~3]。现有的精细尺度下土壤属性空间分布推测方法主要有: ① 基于大量样点的统计性推测方法[4~9],这类方法需要大量样点建模,野外采样成本高; ② 基于少量典型样点的推测方法,本类方法通常将待推测区域模糊划分成不同类型子区域,然后选取各类型中的少量典型土壤样点建模。基于少量典型样点的推测方法野外采样成本较低,具有广泛的应用前景。现有基于少量典型样点进行推测的方法主要包括: 使用对应于土壤类型的典型样点进行推测[1,10],使用对应于环境因子组合类型的典型样点进行推测[11],使用对应于坡位类型的典型样点进行推测[12]等。这些推测方法采用加权平均模型(公式1)构建典型土壤属性与待推测未知土壤属性的关系:

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}} \quad (1)$$

式中,V_{ij}表示栅格(i,j)上的土壤属性推测值; V^k 代表第 k 个类型上的土壤属性典型值,通过野外采样获得; $S_{ij}{}^{k}$ 表示栅格(i,j)位置与类型 k 的模糊隶属度; n 为区域内划分的类别数。该模型可以描述模糊隶属度向量与土壤属性之间的非线性关系。

参数敏感性分析是评价模型的重要方面,但目前尚未对加权平均模型中的类型模糊隶属度和土壤属性参数进行敏感性分析。其中,土壤属性参数不但野外采样成本较大,且所得的土壤属性典型值存在不可回避的不确定性,其主要来源有: ① 由于土壤属性的空间分布具有不确定性[13],典型类型上的土壤属性值不唯一,更可能为一个属性值范围; ② 实验室测定土壤属性值时的测量结果不确定性,典型土壤属性的测量值不唯一; ③ 由于分类体系或典型样点选择不同,可能产生不同的土壤属性典型值。因此,本文关注于对加权平均模型中的土壤属性参数进行敏感性分析。本文将以东北地形起伏和缓的小流域土壤表层有机质空间分布推测为例,讨论土壤属性典型值的变化对加权平均模型输出的影响,同时探索对以加权平均模型为代表的基于少量典型样点土壤属性空间推测模型

收稿日期: 2010-09-29; 修订日期: 2010-12-24

基金项目: 国家自然科学基金项目(40971235,40671111)、中国科学院知识创新工程项目(KZCX2-YW-Q10-1-5)、国家科技支撑计划(2007BAC15B01)资助。

作者简介: 卢岩君(1987-),男,福建泉州人,硕士研究生,主要从事数字地形分析在数字土壤制图中的应用研究。E-mail: luyj@lreis.ac.cn

通讯作者: 秦承志,副研究员。E-mail: qincz@lreis.ac.cn

的敏感性分析方法。

1 材 料

1.1 研究区及样点集概况

研究区命名为鹤北 8 号小流域(图 1),位于黑龙江省嫩江县境内(49°00′00″~49°00′45″N,125°18′43″~125°20′35″E),流域面积约 4 km^2,主要地貌类型是不同程度切割的山前倾斜平原与河流阶地[14],平均坡度 2.4°,地形起伏和缓。土壤属性数据来自研究区的 48 个土壤样点,样点沿着自山脊至沟谷穿越不同地貌部位的剖面线进行布设(图 1)。

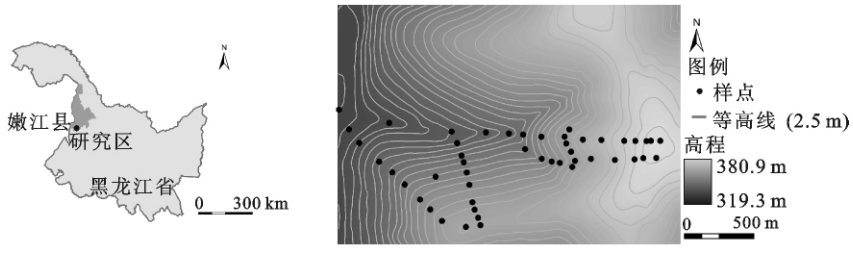


图 1 研究区及样点

Fig. 1 Sketch of study area and soil samples

土壤属性与坡位之间存在协同变化关系。秦承志等[12]将该模型应用于东北地形起伏和缓的小流域,结果表明该模型可以使用较少的野外土壤样点,获得较好的推测效果。

以该模型为例进行敏感性分析的原因是:一方面,地形信息较土壤类型信息更容易获取,而且当前在小流域内精细尺度下的土壤信息推测方法常使用地形信息作为辅助信息,而坡位是地形信息全面的、综合的反映,所以该模型可以作为典型方法,代表在小流域精细尺度下基于少量样点的土壤属性空间分布推测方法;另一方面,坡位类型较土壤类型更易识别与获取,该模型有较好的易用性,以其为例进行敏感性分析也具有实践意义。

为获取建模所需的坡位空间渐变信息,首先根据研究区地形起伏和缓的特点,选取由山脊、坡肩、背坡、坡脚和沟谷等 5 类坡位组成的坡位分类体系[16];然后使用文献[15]提出的基于坡位典型位置相似度的算法,利用数字地形分析软件 SimDTA[14]定量提取研究区的坡位渐变信息。该算法根据原型理论,充分利用典型坡位所包含的丰富的地学概念和地形信息,可克服现有其他算法[17,18]的不足之处,较准确地提取定量化坡位渐变信息[15]。

1.2 基于坡位渐变信息的土壤属性空间分布推测模型

本文以基于定量化坡位渐变信息的土壤属性空间分布推测模型为例,进行土壤属性参数敏感性分析。定量化坡位渐变信息可以全面、有效地描述使用常用地形属性难以描述的空间相对位置关系,而且与土壤空间分布具有关联关系[15]。该模型通过对各类典型坡位上土壤样点属性值的进行加权平均推测土壤属性(公式 1)。模型的基本假设为:① 在小流域内,地形是影响土壤属性空间分布的主要因素;② 典型坡位上出现典型土壤属性,

2 敏感性分析实验设计

由于各类坡位的土壤属性典型值都存在不确定性,所以需要分析各类坡位的土壤属性典型值的不唯一性对模型输出的影响。本文采用文献[19]所介绍的阶乘设计(factorial design)方法设计敏感性分析实验。阶乘设计为每个需要进行敏感性分析的参数选取一个取值集合,然后将所有集合进行组合得到建模方案集,使用建模方案集中的每个方案进行建模和应用,得到一系列的输出结果用以进行敏感性分析,该方法适用于多参数多取值时的敏感性分析。

本文的实验流程如图 2 所示:首先选定对应于各坡位类型模糊隶属度大于一定的阈值(本文取 0.9)的区域为各类坡位的典型区域,处于典型区域内的土壤样点构成各类坡位的典型样点集(图 3),进行野外采样和实验室测量,得到各类坡位土壤属性典型值集,然后将各类坡位的土壤属性典型值进行组合得到建模方案集,建模方案集中的每一个方案都是一套由 n 个土壤属性典型值组成的建模参数集(n 为坡位类别数,本文为 5)。进行推测实验时,每次从建模方案集中取出一个方案进行应用,

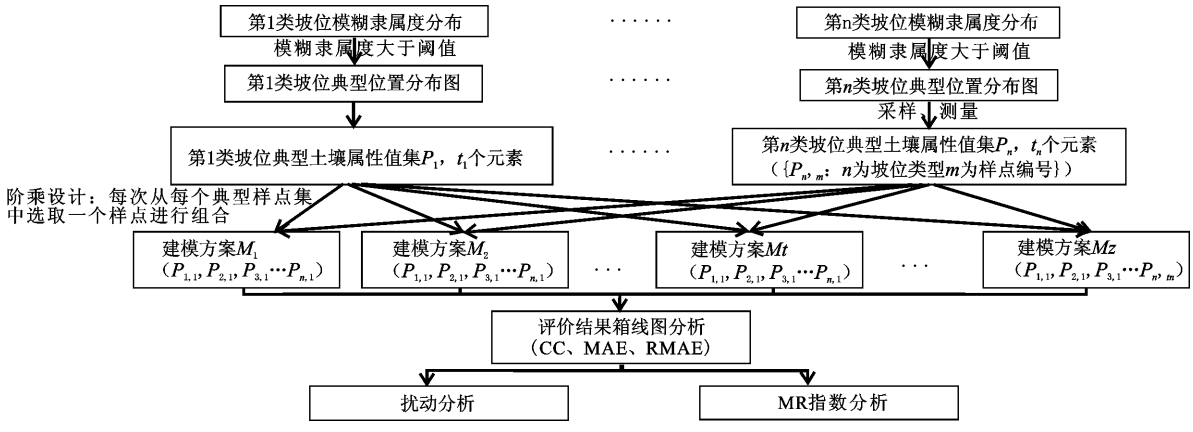


图 2　敏感性分析流程

Fig. 2　Flow chart of sensitivity analysis

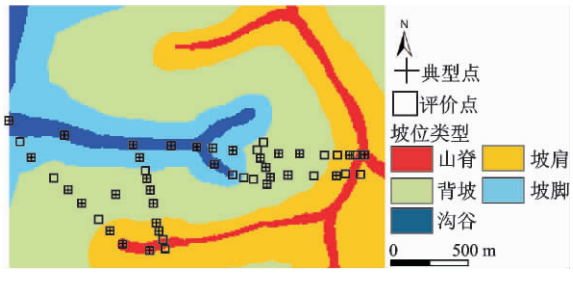


图 3　典型点与评价点分布

Fig. 3　Distribution of typical samples and validation samples

直到所有方案都被应用为止。

限于样点数，本文主要针对单一坡位上土壤属性典型值参数进行敏感性分析。尽管有所简化，但这种处理方式同样可反映多个参数变化时敏感性。其原因是: 首先，下述的分析方法可以反映多个参数的不确定性对推测效果的综合影响; 其次，加权平均模型中的所有土壤属性参数对模型的影响作用是相似的，敏感性分析结果适用于其它坡位。

2.1　评价结果集的箱线图分析

为了保证评价点有足够数目，且覆盖各类坡位的典型位置以及坡位间的过渡，本文以研究区内所有 48 个样点作为评价点(图 3) ，计算预测值与观测值之间的 Pearson 相关系数(Correlation Coefficient，CC，下式值为 C_{CC}) 、平均绝对误差(Mean Absolute Error，MAE，下式值为 M_{MAE}) 和均方根误差(Root Mean Square Error，RMSE，下式值为 R_{RMSE}) 等 3 个评价指标(公式 2 ~ 4) ，用以描述不同建模方案推测效果的差异。

$$C_{CC} = \frac{N\sum OP - \sum O\sum P}{\sqrt{N\sum O^2 - (\sum O)^2}\sqrt{N\sum P^2 - (\sum P)^2}} \quad (2)$$

$$M_{MAE} = \frac{\sum |P - O|}{N} \quad (3)$$

$$R_{RMSE} = \sqrt{\frac{\sum |P - O|^2}{N - 1}} \quad (4)$$

式中，P 为预测值; O 为观测值; N 为评价点个数。

将每个评价指标的所有评价结果按照选取的单一坡位上的土壤属性典型值进行分类统计，绘制箱线图。箱线图中的单个箱线可以直观地反映当某一类坡位土壤属性参数值得到确定时的两方面信息: ① 模型的平均推测效果，即评价指标的中位数; ② 其他各类坡位土壤属性参数值的变化对推测效果的综合影响，即评价指标的变异程度。对箱线图中的不同箱线进行比较，可以反映单一坡位的土壤属性参数值变化时对模型推测效果的影响。本文进一步采用以下两种方法挖掘箱线图中的敏感性信息: ① 水文模型中的扰动分析法; ② 本文提出的综合考虑推测效果及其变异性的定量化评价指标——MR 指数分析法。

2.2　扰动分析法

如果存在使模型推测效果达到最优的土壤属性典型值，可借鉴水文模型敏感性分析中的扰动分析法[20] 定量分析参数敏感性。扰动分析法是一种最简单的参数敏感性分析方法，即在某个参数最佳估计值附近给定一个人工扰动，计算参数在小范围内产生波动所导致模型输出变化的敏感性指数(式 5) ，并根据敏感性指数的范围(表 1) 判断参数敏感性大小[21] 。

$$I = [y' - y_0) / y_0] / [x' - x_0) / x_0] \quad (5)$$

式中，I 为敏感性指数; y'为经过“扰动”后的输出，

y_0 为“扰动”前的输出，本文选取的输出为各评价指标箱线的中位数；x'为经过“扰动”后的参数值，x_0 为“扰动”前的参数值。

表 1 敏感性指数与敏感性程度对应关系表

Table 1 Sensitivity levels defined by sensitivity index

敏感性指数绝对值区间	[0, 0.05]	(0.05, 0.2]	(0.2, 1]	(1, +∞)
敏感性程度	低敏感性	中敏感性	高敏感型	极高敏感性

2.3 综合考虑推测效果及其变异性的定量化评价指标——MR 指数分析法

由于单独考虑模型的推测效果或者推测效果的变异性较为片面，本文设计一个将推测效果及其变异性相结合的定量化评价指标，称为 MR 指数（M 代表推测效果统计指标的中值，R 代表推测效果统计指标的极差，式 6 中值为 M_{MR}），用于综合评价模型的推测效果与变异性，可以作为考虑模型推测效果情况下的敏感性定量化描述指标。

$$M_{MR}(I) = \mathrm{Sgn}(I) \times \mathrm{Median}(I) + \mathrm{Range}(I) + \mathrm{Offset}(I) \quad (6)$$

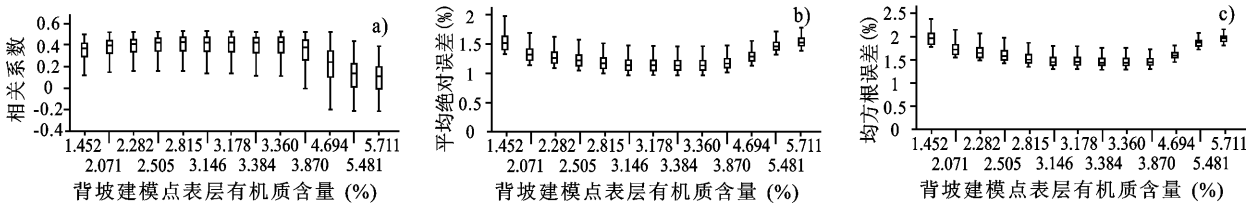
式中，I 为推测效果的统计评价指标，如 CC、MAE 或 RMSE；Sgn(·) 为符号函数，Sgn(CC) = −1，Sgn(MAE) = Sgn(RMSE) = 1；Median(I) 为评价指标中值的标准化数据；Range(I) 为评价指标极差的标准化数据；Offset(·) 为偏移函数，Offset(CC) = 1，Offset(MAE) = Offset(RMSE) = 0。MR 指数值越小，模型在考虑推测效果时敏感性越低。由于不同推测效果评价指标对推测效果的响应不同，推测效果越好，CC 越大、MAE 和 RMSE 越小，因此在式 6 中设置符号函数 Sgn(·) 和偏移函数 Offset(·)，维护各指标 MR 指数数值意义的统一性，使 MR 指数的值域保持为 [0, 2]。MR 指数与敏感性指数在反映敏感性方面的区别在于后者反映被选取的单一参数变化引起的模型输出变化，而前者反映其它未被选取的参数共同变化引起的模型输出变化。

3 应用结果及讨论

本文以研究区内表层有机质含量典型值数量最多的背坡为例进行分析。从背坡上不同表层有机质含量典型值对应推测效果箱线图可见（图 4，每个箱线所用的评价结果数目为 360），在本研究区，随着背坡表层有机质含量典型值的增大，3 个评价指标中位数的变化都反映出推测效果先升高而后降低，取得最佳推测效果的表层有机质含量典型值为 3.360%，可以认为该值为该参数的最优值（以下简称为最优参数）；3 个指标的极差变化则有所不同，CC 的极差总体上呈现增大趋势，而 MAE 与 RMSE 的极差总体上呈现减小趋势。

本文以最优参数作为参数的最佳估计值，非最优推测结果对应的建模点表层有机质含量作为受到扰动后的参数值，使用扰动分析法计算敏感性指数，所得结果显示（图 5）：最优参数附近的参数值对应各评价指标的敏感性指数较小，大都处于中低敏感性区，而距离最优参数较远的参数值对应各评价指标的敏感性指数较大，大都处于高敏感性区。如果假设最优参数为土壤有机质含量的真值，在实验室有机质含量测量误差为 ±5% 的情况下，各评价指标的敏感性指数中有一半处于中低敏感性区，处于高敏感性区的敏感性指数值也接近中低敏感性区，因此，由实验室有机质含量测量误差引起的敏感性不高。进一步计算各参数对应的 MR 指数（图 6）可以发现，对 MAE 与 RMSE 两个指标，在最优参数附近的参数值对应模型的 MR 指数相对较小，这说明当参数值越接近最优参数时，模型敏感性越低。

通过扰动分析和 MR 指数分析可以发现一个共同现象：无论是否考虑模型推测效果，最优参数



a. 相关系数(CC)；b. 平均绝对误差(MAE)；c. 均方根误差(RMSE)

图 4 研究区背坡建模点有机质含量 – 评价指标箱线

Fig. 4 Box plots of evaluation indices for typical values of organic matter content on back slope

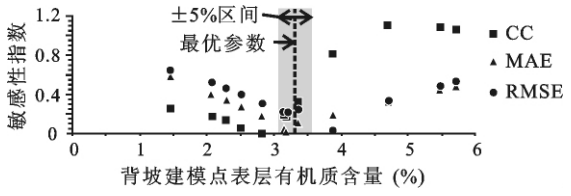


图 5　背坡表层有机质典型值变化的敏感性指数散点

Fig. 5　Scatter plot of sensitivity index

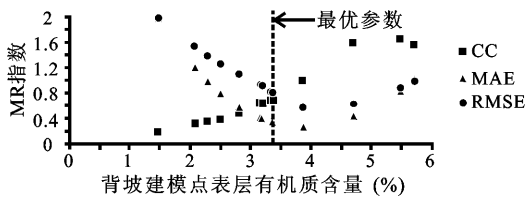


图 6　背坡表层有机质典型值变化的 MR 指数散点

Fig. 6　Scatter plot of MR index

附近的参数取值的敏感性较低。进一步考查不同敏感性区域对应的建模点在空间上的分布时可以发现,敏感性较低时建模点所处的空间位置水平曲率绝对值较小,说明这些建模点所处位置的地表形态在水平方向上较为平直,即典型的直坡,而敏感性较高时建模点所在区域的水平曲率绝对值较大,说明其所处位置的地表形态在水平方向上弯曲程度大。而大多数背坡上的评价点都位于直坡。所以,当建模点与评价点所处区域地形条件相似时,模型的参数敏感性较小,反之则较大。一般而言,同属一类坡位的区域地形条件是相似的,但相似程度与坡位分类的详细程度有关,坡位分类越详细,相似程度越大。因此,加权平均模型在本研究区中总体敏感性高的原因可能是坡位分类体系不够详细,导致某些表层有机质含量典型值代表性不强。

综上,虽然土壤属性典型值是加权平均模型的一个敏感性较大参数,但可考虑通过以下两种方法降低模型对该参数敏感性: ① 采用较为详细的坡位分类体系(尤其是对地表形态较为复杂区域),本研究区中可考虑根据水平曲率进一步细分坡位类型,例如将背坡进一步分为凸背坡、直背坡和凹背坡等[15]; ② 如果改变坡位分类体系带来的实际工作量较大,可以考虑预先调查待推测区域的地形特征,若某类坡位某种子类型在待推测区域中的范围较大,则可将该类坡位的建模点布设在范围较大子类型处,有可能降低参数敏感性对模型推测结果的影响。这将同时涉及到对加权平均模型中的模糊隶属度向量参数的敏感性分析。

4　结论与展望

为了对加权平均模型中土壤属性参数进行敏感性分析,同时探索对基于少量典型样点的土壤属性空间推测模型的参数敏感性分析方法,本文在东北地形起伏和缓的小流域内,以基于坡位渐变信息的土壤表层有机质含量空间分布推测模型为例,使用阶乘设计生成的箱线图对加权平均模型进行土壤属性参数敏感性分析,并从两个方面做深入分析: ① 使用扰动分析结合敏感性指数分析法对由单一坡位土壤属性典型值变化体现的参数敏感性分析; ② 提出 MR 指数,并使用该指数对由多个坡位土壤属性典型值共同变化体现的参数敏感性进行分析,同时结合模型推测效果进行讨论。

结果表明,总体上加权平均模型的土壤属性参数敏感性较大,但敏感性与土壤属性典型点的空间分布有关,若分布位置与大多数受其影响较大的推测区具有相近的地形条件,则土壤属性典型值导致的敏感性较小,敏感性的主要来源可能是分类体系的不确定性。因此,通过选择更详细的分类体系,或保障土壤属性典型点在对应坡位中能代表的面积,将可能降低敏感性。下一步研究还将关注对加权平均模型中的另一个重要参数——模糊隶属度向量的敏感性分析,以及研究如何根据区域特征确定最合适的坡位分类体系,以降低参数敏感性。

本文采用基于箱线图的扰动分析法和 MR 指数分析法对加权平均模型的敏感性进行定量刻画。其中,MR 指数可综合反映模型敏感性和推测效果等两个重要方面,用于对加权平均模型进行综合敏感性分析。这些分析方法只应用于模型输出结果所形成的箱线图,并不依赖于特定的模型结构,因此本文的分析方法适用于对其它类型模型的敏感性分析。

参考文献:

[1] Zhu A X, Band LVertessy R, et al. Derivation of soil properties using a Soil Land Inference Model (SoLIM) [J]. Soil Science Society of America Journal, 1997, **61**(2): 523 – 533.

[2] 阎伍玖. 芜湖市城市边缘区土壤重金属污染空间特征研究[J]. 地理科学, 2008, **28**(2): 282 ~ 285.

[3] 白军红, 丁秋祎, 高海峰, 等. 向海湿地不同植被群落下土壤氮素的分布特征[J]. 地理科学, 2009, **29**(3): 381 ~ 384.

[4] Bishop T F A, McBratney A B. A comparison of prediction methods for the creation of field – extent soil property maps [J]. Geoderma, 2001, **103**(1 – 2): 149 – 160.

[5] 连　纲, 郭旭东, 傅伯杰, 等. 基于环境相关法的土壤属性空

间分布特征研究——以黄土丘陵沟壑区小流域为例 [J]. 地理科学,2008,28(4): 554 ~ 558.

[6] 孙孝林,赵玉国,秦承志,等. DEM 水平分辨率对多元线性土壤 - 景观模型及其制图应用的影响 [J]. 土壤学报,2008,45(5): 971 ~ 977.

[7] Mueller T G,Pierce F J. Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales [J]. Soil Science Society of America Journal,2003,67(1): 258 - 267.

[8] 李 翔,潘瑜春,赵春江,等. 利用不同方法估测土壤有机质及其对采样数的敏感性分析 [J]. 地理科学,2007,27(5): 689 ~ 694.

[9] Pei T,Qin C,Zhu A,et al. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow - direction algorithms and kriging methods [J]. Ecological Indicators,2010,10(3): 610 - 619.

[10] Zhu A X,Qi F,Moore A,et al. Prediction of soil properties using fuzzy membership values [J]. Geoderma,2010,158(3-4): 199 - 206.

[11] 杨 琳,朱阿兴,秦承志,等. 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 [J]. 土壤学报,2009,46(1): 9 ~ 15.

[12] 秦承志,卢岩君,邱维理,等. 模糊坡位信息在精细土壤属性空间推测中的应用 [J]. 地理研究,2010,29(9): 1706 ~ 1714.

[13] Wetterlind J,Stenberg B,Soderstrom M. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models [J]. Geoderma,2010,156(3 - 4): 152 - 160.

[14] 秦承志,卢岩君,包黎莉. 简化数字地形分析软件 SimDTA(V1.0) 及其在坡位信息提取中的应用 [J]. 地球信息科学学报,2009,11(6): 737 ~ 743.

[15] Qin C,Zhu A X,Shi X,et al. Quantification of spatial gradation of slope positions [J]. Geomorphology,2009,110(3 - 4): 152 - 161.

[16] 秦承志,朱阿兴,李宝林,等. 坡位的分类及其空间分布信息的定量化 [J]. 武汉大学学报(信息科学版),2009,34(3): 374 ~ 377.

[17] Burrough P A,van Gaans P F M,MacMillan R A. High - resolution landform classification using fuzzy k - means [J]. Fuzzy Sets and Systems,2000,113(1): 37 - 52.

[18] MacMillan R A,Pettapiece W W,Nolan S C,et al. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic [J]. Fuzzy Sets and Systems,2000,113(1): 81 - 109.

[19] Hamby D M. A review of techniques for parameter sensitivity analysis of environmental models [J]. Environmental Monitoring and Assessment,1994,32(2): 135 - 154.

[20] 王中根,夏 军,刘昌明. 分布式水文模型的参数率定及敏感性分析探讨 [J]. 自然资源学报,2007,22(4): 649 ~ 655.

[21] Lenhart T,Eckhardt K,Fohrer N,et al. Comparison of two different approaches of sensitivity analysis [J]. Physics and Chemistry of the Earth,2002,27(9 - 10): 645 - 654.

Sensitivity Analysis of Soil Property Parameter in Typical-sample-based Prediction Model of Digital Soil Mapping

LU Yan-jun[1,2], QIN Cheng-zhi[2], QIU Wei-li[1], ZHU A-xing[2]

(1. *School of Geography, Beijing Normal University, Beijing* 100875, *China*; 2. *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Science and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101, *China*)

Abstract: Although the typical-sample-based prediction models of the spatial distribution of soil property are promising in digital soil mapping, the sensitivity analysis of the soil property parameter in these models is still lack. This paper presents an attempt to establish weighed average model on a fuzzy slope position based on typical samples for digital soil mapping. The soil samples of A-horizon soil organic matter were collected in a low-relief small watershed in Northeast China. A "factorial design" method was applied to simulating scenarios of modeling point set. The performances of the model under different scenarios were analyzed with box plot. Further, we designed an MR index, which was calculated from the median and the range of the prediction error, for parameter sensitivity analysis based on the prediction performance of model. Both MR index and existing sensitivity index were used in this study. The result shows that the weighed average model is sensitive to the soil property parameter. The sensitivity degree is related to the locations of typical soil sample. The method proposed in this paper is also available for the parameter sensitivity analysis of other typical-sample-based prediction models of digital soil mapping.

Key words: typical sample; digital soil mapping; soil property; sensitivity analysis; fuzzy slope position