

基于样点个体代表性的大尺度土壤属性制图方法*

刘京^{1,2†} 朱阿兴^{1,2} 张淑杰¹ 秦承志¹

(1 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)

(2 Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA)

摘要 大空间尺度范围的土壤属性分布信息是陆地表层过程模拟的基础信息。基于野外样点进行空间插值是获得土壤属性空间分布信息的重要手段。现有的空间插值方法通常要求所用样点对研究区土壤属性空间分布规律具有良好的全局代表性。然而,受采样经费和野外采样条件的限制,所采集的样点往往难以全面地反映研究区土壤属性的空间分布规律。基于这样的样点用现有空间插值方法得到的土壤属性分布图通常精度较低,并且由样点全局代表性差带来的推测不确定性也无法得到度量。为了合理利用这些已采集的但全局代表性不好的样点,本文提出了基于样点“个体代表性”推测土壤属性空间分布并度量推测不确定性的方法。该方法在两点环境条件越相似、土壤属性就越相似的假设下,认为每一样点可以代表与其环境条件相似的地区,并且代表程度可以由两点的环境相似度量;通过分析环境相似度计算推测不确定性,并以环境相似度为权重计算样点可代表地区的土壤属性值。将该方法应用于推测新疆伊犁地区土壤表层有机质含量,经验证,本文方法能够有效地利用全局代表性差的样点推测样点能够代表地区的土壤属性空间分布,并且所得的推测不确定性与预测残差呈现正向关系,能够有效地指示推测结果的可靠程度。

关键词 环境相似度;样点代表性;不确定性;土壤属性制图

中图分类号 P934

文献标识码 A

大尺度范围的土壤属性空间分布信息是进行陆地表层地理过程模拟的基础信息^[1]。为了评价模型模拟结果的可靠程度,所用数据的不确定性信息也受到了越来越广泛的重视^[2-4]。基于野外采样点对土壤属性的空间分布进行插值是获得土壤属性空间分布的重要手段之一^[5-7]。为了保证插值结果的精度,现有的插值方法通常要求所用样点对研究区土壤空间变化具有良好的全局代表性^[8]。样点的全局代表性是指:样点所反映的土壤属性空间分布规律能够体现研究区土壤属性的整体分布规律。在数值空间内,具有全局代表性的样点,其土壤属性值应尽可能全面地包含研究区土壤属性的典型值,同时样点属性的值域范围应与研究区土壤属性的值域范围一致;在地理空间内,全局代表性样点是能够全面反映土壤属性空间变异的样点。要获得这样的样点,经典空间统计理论认为需要采集一定数量、并且其分布满足特定规则(如随机或规则分布)的样点^[9]。然而,野外采样是一件耗时耗力耗资的工作,受采样经费和野外采样条件的限

制,按照随机或规则方式采集大量样点通常是不现实的,在国家和区域尺度上尤其如此。目的性采样方法^[10]认为,具有全局代表性的样点可以是少量能够反映土壤空间分布规律的典型样点,然而由于缺乏理论依据,典型样点位置的确定通常较为主观,在大的空间尺度上,土壤空间分布规律通常更为复杂,典型样点的确定显得更为困难。目前,能够用于国家和区域尺度土壤制图的样点通常来源于国家和地区土壤普查^[11],这些样点既无法在数量和分布上满足经典统计理论的要求,也无法确保其典型性^[12]。因此,基于这些样点运用现有的空间插值方法,很难保证所得土壤空间分布信息的精度。此外,现有方法同样无法度量由样点全局代表性差而带来的推测不确定性。

本文针对上述问题,提出了一种基于单个样点代表性推测土壤属性空间变化并度量推测不确定性的方法,旨在突破现有方法对样点全局代表性的严格要求,同时弥补其不能度量由样点代表性差而带来的推测不确定性的局限。本文以新疆伊犁地

* 国家自然科学基金项目“基于样点局域代表性的地理要素空间变化推测方法研究”(40971236)、国家国际科技合作项目“全球数字土壤制图东亚区合作研究”(2010DFB24140)资助

作者简介:刘京(1985—),女,河南开封人,博士研究生,主要从事数字土壤制图及不确定性分析研究。

收稿日期:2012-01-04;收到修改稿日期:2012-06-26

区为例对土壤表层有机质含量空间分布推测进行了实验研究。

1 研究方法

土壤-景观模型^[8]认为土壤与影响其发生发展的环境条件之间存在着协同关系,可以通过环境条件的空间分布来推测土壤的空间分布;土壤景观(环境)单元越相似,其所对应的土壤越相似,并且这些相同或相似的单元在空间上是重复发生的^[13]。基于这样的认识,本文借鉴案例推理的思想(Case-based Reasoning, CBR)^[14],认为每一个样点可以看作包含特定土壤-环境关系的案例,并能够代表与之环境条件相似的地区。本文把单个样点的代表性叫做样点的“个体代表性”,代表程度可以通过环境

条件的相似程度来度量。基于样点个体代表性推测未采样点土壤属性并计算推测不确定性的整体思路如下:首先,选择环境要素刻画已知样点和待推测点所处的环境条件,通过比较环境条件的相似程度来衡量样点对待推测点的代表程度;其次,根据已有样点对待推测点的代表程度(环境相似度)计算推测不确定性;最后,对于不确定性低的待推测点,以代表程度(环境相似度)为权重综合样点土壤属性值推测其土壤属性值,对于不确定性高的待推测点,由于现有样点集不能很好地代表该点,不做推理并将其属性值赋为空值,最终,在一定的不确定性水平下,得到现有样点能够代表区域的土壤属性空间分布图以及整个研究区推测不确定性分布图。本文方法的整体技术路线如图 1 所示。

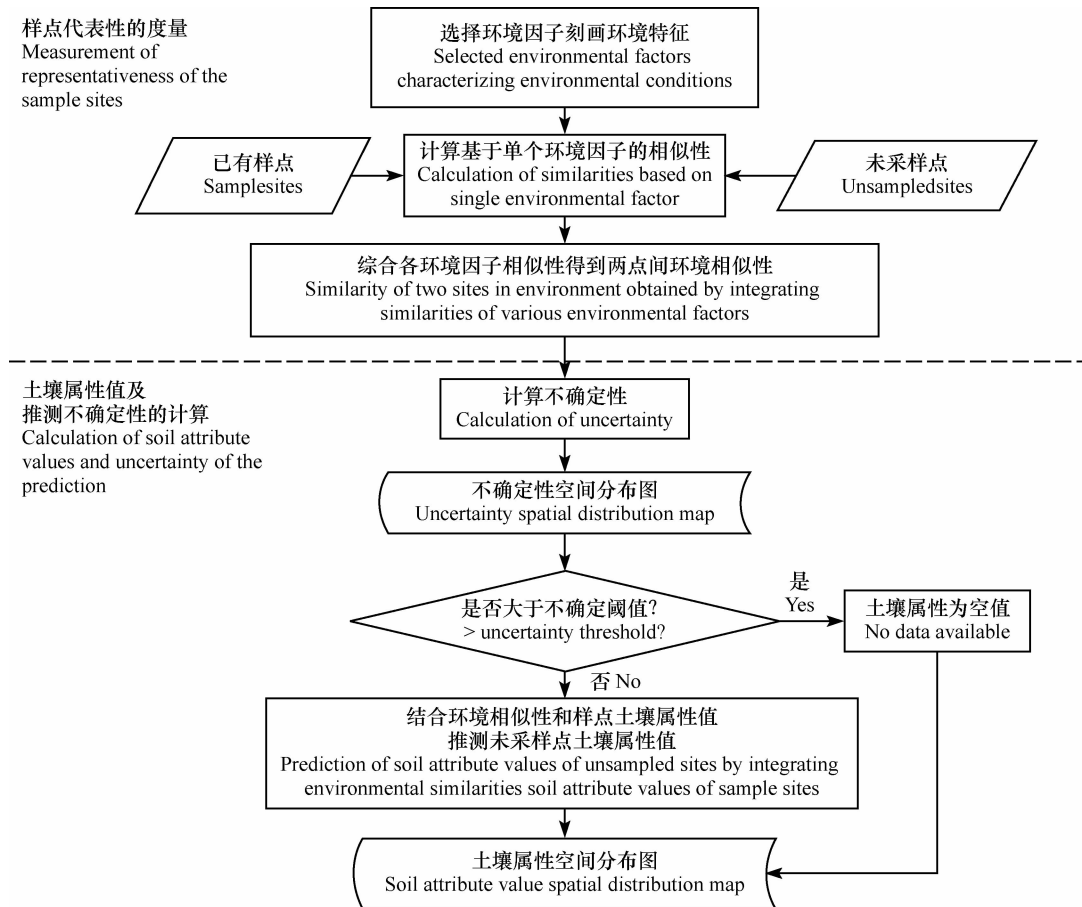


图 1 技术路线图

Fig. 1 Technical flowchart of the proposed method

1.1 样点代表程度的度量

已有样点对待推测点的代表程度可由二者的环境相似度来度量。环境相似度的确定首先需要

选择与土壤在空间上协同分布的环境因子来刻画样点与待推测点的环境特征。根据土壤发生学原理,气候、地形、母质、植被等因子与土壤形成发展

有密切的关系,相关环境变量可以用于指示土壤属性的空间分布。基于所选环境变量,在每一点用环境特征向量(式(1))刻画该点的环境特征,其中, m 是所选环境变量的个数,每一个元素 $e_{v,ij}(v=1,\dots,m)$ 是所选环境变量在待推测点 (i,j) 的特征值。

$$\vec{e}_{ij} = (e_{1,ij}, e_{2,ij}, \dots, e_{m,ij}) \quad (1)$$

基于环境特征向量计算环境相似性需要在两个层面上进行:一是计算基于单个环境变量的相似性;二是综合各环境变量的相似性得到两点的环境相似性(式(2)):

$$S_{ij,k} = P \left(E_v(e_{v,ij}, e_{v,k}) \right) \quad (2)$$

式中, $S_{ij,k}$ 为待推测点 (i,j) 与第 k 个样点之间的环境相似度, $e_{v,ij}$ 、 $e_{v,k}$ 分别为第 v 个环境变量在点 (i,j) 和样点 k 处的特征, m 为所选环境变量的个数。 E_v 为用于计算单个环境变量相似度的函数,其形式取决于环境变量 v 的类型:若环境变量特征为名称量或序数量,可选用布尔函数;若数据类型为间隔量或比率量,则可选用距离函数(如高斯距离、马氏距离等)。 P 为用于综合各变量相似度得到两点环境相似度的函数,其形式的确定应考虑环境变量之间的关系^[15],可选用的方法有加权平均法^[8]、最小限制因子法^[16]、基于规则的方法^[17]等。最后在每一个待推测点可以得到一个环境相似度向量(式(3)),该向量中的每一个元素 $S_{ij,k}(k=1,2,\dots,n)$ 是待推测点 (i,j) 与第 k 个样点的环境相似度, n 为样点数量。环境相似度向量表达了已有样点集对待推测点的代表程度:

$$\vec{S}_{ij} = (S_{ij,1}, S_{ij,2}, \dots, S_{ij,n}) \quad (3)$$

1.2 土壤属性值和推测不确定性的计算

基于环境相似度向量,本文采用式(4)计算由样点代表性带来的推测不确定性:

$$Uncertainty_{ij} = 1 - \max(S_{ij,1}, S_{ij,2}, \dots, S_{ij,n}) \quad (4)$$

式中,取最大值的操作可以得到待推测点与其代表性最好(即环境最相似)的样点之间的环境相似度,若该相似度仍然较低,说明现有样点集不能很好地代表该点,那么用现有样点集推测待推测点的土壤属性值将会有较高的不确定性。

对于推测不确定性过高的点,由于现有样点不

能很好地代表该点,不宜用现有样点推测该点的土壤属性值。本文通过设置不确定性阈值来检测现有样点是否能够很好地代表待推测点:若待推测点的不确定性高于阈值,则不做推理而将其属性赋为空值;若待推测点的不确定性低于阈值,则以代表程度(即环境相似度)为权重,通过线性加权的方式综合各样点土壤属性值,进而计算待推测点土壤属性值(式(5)):

$$V_{ij} = S_{\max} * V_{s_{\max}} + (1 - S_{\max}) * \frac{\sum (S_{others} * V_{others})}{\sum S_{others}} \quad (5)$$

式中, V_{ij} 为待推测点 (i,j) 的土壤属性值, S_{\max} 和 $V_{s_{\max}}$ 分别为代表性最好的样点与待推测点的环境相似度和土壤属性值, S_{others} 和 V_{others} 分别为其他样点与待推测点的环境相似度和土壤属性值。

2 研究区及样点分布

研究区位于新疆伊犁地区(42°14'16"~44°50'30"N,80°09'42"~84°56'50"E),东西长约350 km,南北宽约280 km,总面积约为56 400 km²,该区东、南、北三面被天山山脉环绕,西面敞开,整体地势西低东高,包含“四谷一盆”五大地貌单元,其中伊犁河谷海拔最低,约为500~900 m,最高为昭苏盆地和特克斯谷地,约为900~2 000 m。本区位于温带大陆性生物气候带,各个地形区气候植被类型差异显著,土壤也因此呈现明显的地带性分布。伊犁地区各河谷平原的成土母质母岩主要有以下几种类型:淡黄色沙及少量粉砂的石英砂、灰色石英砂云母砂、第三纪红色泥岩、洪积冲积黄土状母质、冲积黄土状母质等。

本文从《中国土种志》^[18]、《新疆伊犁地区土壤》^[19]中共收集了75个样点作为建模点,其空间分布如图2中圆点所示。可以看出,这些样点的空间分布不符合随机或规则的分布模式,且多集中于河谷地区,不能代表整个研究区土壤空间分布特征。

为了验证本文方法的有效性,本文收集了32个独立验证点¹⁾(图2中三角点)。虽然这些样点主要

1) 独立验证点来自于:

- ① 由新疆农业科学院土壤肥料研究所马兴旺研究员提供。样点详细资料参见硕士学位论文:于秀秀. 新疆伊犁河谷新垦区土壤质量评价. 中国农业大学, 2008
- ② 由中国科学院地理科学与资源研究所自然地理与全球变化研究部土地覆被变化与土地资源研究室张红旗副研究员提供
- ③ 由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张百平研究员提供

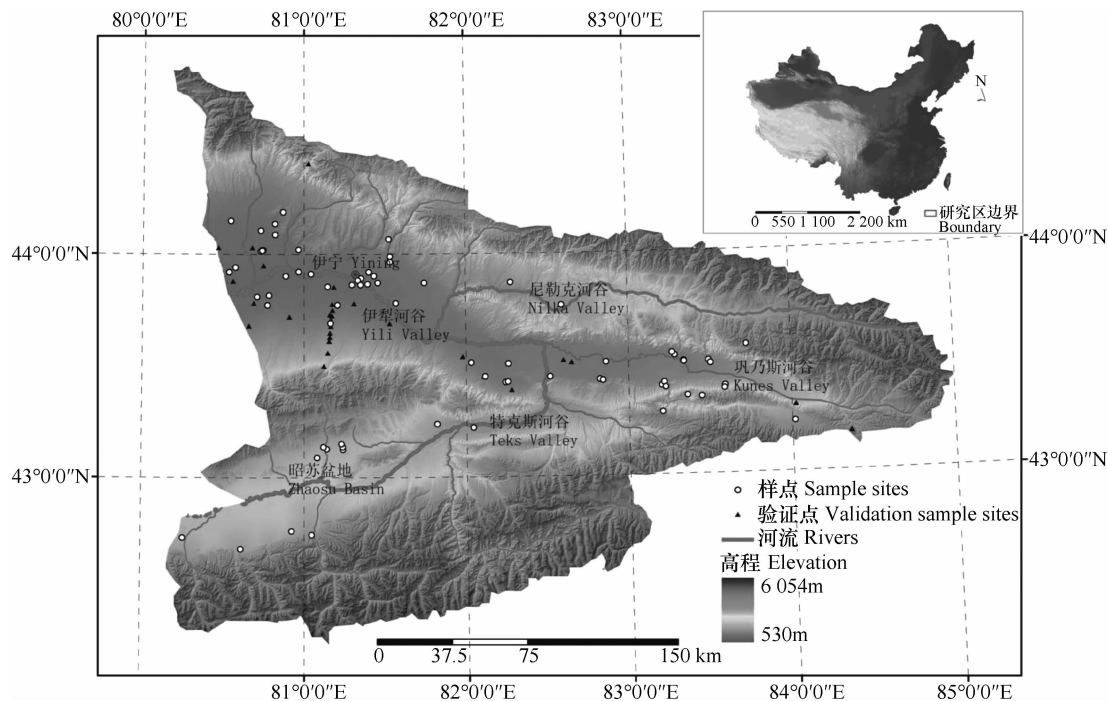


图2 研究区及样点分布图

Fig. 2 Study area and distribution of sampling sites

集中于伊犁河谷区,但是由于最终推理结果中不确定性较低的区域也主要集中于伊犁河谷,这些验证点仍能够用来评价推理结果的精度。

3 方法应用与结果检验

3.1 样点代表程度的度量

根据对本区土壤与环境关系的认识^[19],选择了气候、母质、地形三种环境要素来刻画研究区与土壤空间分布相关的环境条件特征。没有选择植被要素的原因是由于本区植被条件的空间分布直接受气候和地形条件的影响,气候和地形条件的组合特征可以较好地代表植被特征。气候数据由中国科学院地理科学与资源研究所生态系统研究网络观测与模拟重点实验室提供^[20],包括年均降水量、年均温度和年均相对湿度3个变量,空间分辨率为1 km;母质数据来源于中国1:250万地质图²⁾,在ArcGIS软件中将其进行栅格化,空间分辨率为1 km;地形数据包括高程、坡度、坡向、曲率、表面积比率和地形湿度指数,这六种地形环境变量均派生于数字高程模型(原始数字高程模型数据来自于

SRTM 90 m 分辨率的地形数据产品)³⁾。坡度、坡向、沿剖面曲率、沿等高线曲率根据 ArcGIS 软件所采用的算法生成;表面积比率根据 Jenness^[21]提出的算法生成;在计算地形湿度指数时采用了 Qin 等^[22]提出的多流向算法。地形数据的分辨率均为 90 m。

由于环境变量数据分辨率不同,选择最终成图栅格大小为 1 km。对于气候和母质变量,由于其分辨率与成图分辨率相同,且这些变量通常在 1 km × 1 km 栅格范围内空间变异较小,本文采用单一值来刻画这些环境变量在一个成图栅格的特征;对于地形变量,由于其分辨率小于成图分辨率,在一个成图栅格内部有多个值,且地形特征在 1 km × 1 km 栅格范围内可能具有较大的空间变异性,单一的平均值可能无法正确表达一个栅格内的地形特征。针对这一问题,本文通过核密度估计的方法(kernel density estimation, KDE)^[23]估计地形变量在每一栅格的概率密度分布函数(probability distribution function, PDF),用以刻画地形在每一栅格内部的变化特征。核密度估计是一种非参数估计方法,其优点在于它只从数据本身出发而不需引入对数据分布的

2) 由国土资源数据共享网下载 <http://gsd.cgs.cn/download.asp>

3) 来源于中国科学院计算机网络信息中心国际科学数据镜像网站 <http://datamirror.csdb.cn>

先验假设,并且能够得到平滑连续的概率密度分布函数。

基于上述环境因子,可以计算样点与待推测点的环境相似性。在计算基于单个环境变量相似性时,对于气候变量选择了 GOWER 相似度(式(6))^[24],该方法基于标准化的变量值计算相似度,能够消除不同数量单位对计算结果的影响。其中 $SC_{ij,k}$ 为待推测点 (i,j) 与第 k 个案例点的气候相似度, $e_{v,ij}$ 、 $e_{v,k}$ 分别为第 v 个气候变量在待推测点 (i,j) 和案例点 k 处的特征值, $Range(e_v)$ 为第 v 个气候变量的值域, m 为气候变量的个数。

$$SC_{ij,k} = \frac{1}{m} \sum_{v=1}^m \left(1 - \frac{|e_{v,ij} - e_{v,k}|}{Range(e_v)} \right) \quad (6)$$

对于母质变量采用了布尔函数,即:若待推测点 (i,j) 处母质类型与案例点 k 处类型相同,则二者的母质相似度 $SP_{ij,k} = 1$, 否则 $SP_{ij,k} = 0$ 。对于地形变量,由于采用了概率密度分布曲线表达每一地形变量在成图栅格的特征,本文采用了如式(7)所示的计算方法^[25]计算地形相似度,其中 $ST_{ij,k}$ 为待推测点 (i,j) 与案例点 k 的地形相似度, $A_{v,ij}$ 、 $A_{v,k}$ 分别为点

(i,j) 和案例点 k 所在成图栅格内部第 v 个地形变量的概率密度分布曲线下的面积, $A_{v,ijk}$ 为两条概率密度分布曲线相交下的面积, m 为地形变量的个数。

$$ST_{ij,k} = \frac{1}{m} \sum_{v=1}^m \frac{2 \times (A_{v,ijk})}{A_{v,ij} + A_{v,k}} \quad (7)$$

在综合气候、母质、地形相似度时,考虑到气候和母质主要影响土壤在宏观尺度上的分布规律,而地形主要控制土壤在微观尺度上的分异规律,采用了如图3所示的综合方法^[26]。其中 $S_{ij,k}$ 为待推测点 (i,j) 与案例点 k 的环境相似度。为了得到 $S_{ij,k}$, 先判断气候相似度 $SC_{ij,k}$, 如果气候足够相似(此处选择 0.5 为阈值,在其他实际应用中可根据具体情况进行调整),则再进一步判断母质相似度 $SP_{ij,k}$ 是否为 1(即母质是否相同),若 $SP_{ij,k} = 1$, 则最后对气候相似度 $SC_{ij,k}$ 和地形相似度 $ST_{ij,k}$ 进行加权平均得到 $S_{ij,k}$ 。在进行加权时的权重选择,由于缺乏在该区域微观尺度上气候和地形在决定土壤属性方面的相对重要性,本文在此选择赋予二者相同的权重。在其他实际应用中,可以根据具体情况对二者的权重进行调整。

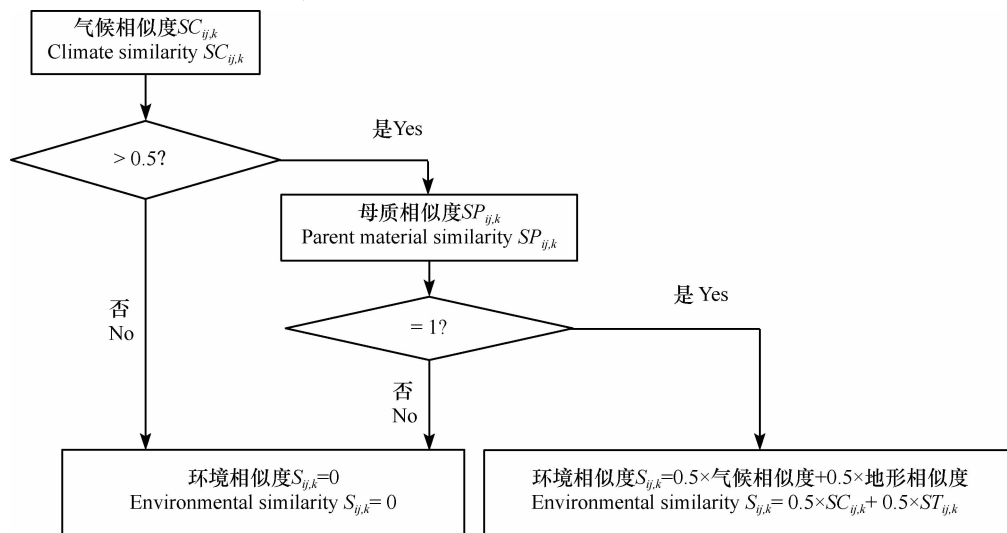


图3 综合气候、母质、地形相似度的方法

Fig. 3 An approach to integrating climate, parent material and topographic similarities

3.2 推测不确定性与土壤表层有机质含量的计算

将待推测点与各案例点的环境相似度代入式(4)计算推测不确定性,得到如图4推测不确定性空间分布图。设定一定的不确定性阈值,如 0.2,认为只有当环境相似度大于 0.8 时,案例点能够很好地代表待推测点;而其他待推测点,因为与案例点环境相似度过低,被认为不能由案例点代表,用案例点进行

推测的不确定性过高。对于能够被案例点代表的待推测点,运用式(5)以环境相似度为权重综合各案例点计算得到待推测点的表层有机质含量;对于不能由现有案例点代表的待推测点,不宜用现有样点进行推理,将其属性值赋为空值“*No data*”。最终得到的土壤表层有机质含量分布图如图5所示。本文在 3.3 节就不确定性阈值的选择进行了探讨。

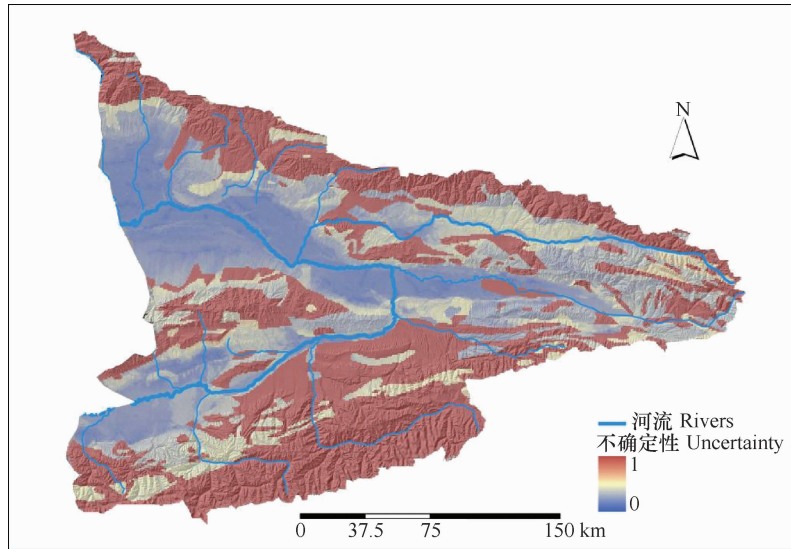


图 4 不确定性分布图

Fig. 4 Uncertainty distribution map

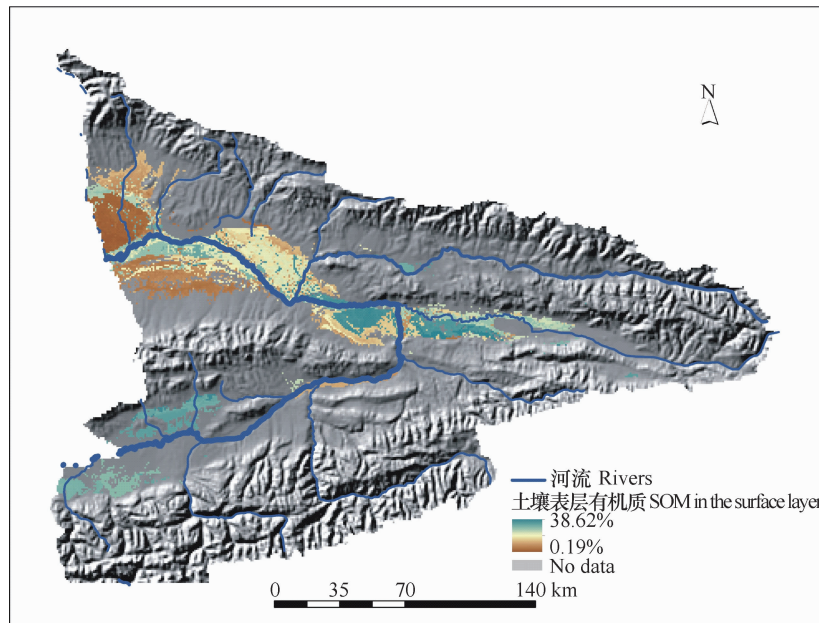


图 5 土壤表层有机质含量空间分布图

Fig. 5 Spatial distribution map of A-horizon soil organic matter (SOM) content

3.3 结果检验

基于 32 个独立验证点对所得结果进行以下检验和讨论:

(1) 在精度检验方面:基于验证点的预测残差(推测值与实际观测值的差异)计算均方根误差(Root Mean Square Error, RMSE),并与实际观测值的标准差(Standard Deviation, SD)相比较:基于 32 个独立验证点所得的 RMSE 为 1.827,SD 为 3.317。由于标准差(SD)度量了观测值的平均值相对实际

观测值的偏差,而 RMSE 度量了推测值相对实际观测值的偏差,因此若 RMSE 小于 SD,则说明方法所得推测值相对于实际观测值的平均值更接近实际观测值。

验证点实测值与预测值 1:1 校验散点图如图 6 所示。可以看到,多数验证点位于 1:1 线附近,但普遍存在预测值略高于实测值的现象,此外,实测值为 13.358% 的验证点存在较大偏差。预测值略高于实测值的原因可能在于式(5)计算土壤属性值的

方法。该公式为了强调最相似案例点对决定待推测点土壤属性值的影响,将其单独列为一项,并未参与其他案例点土壤属性值的归一化。而在实测值为 13.358% 的验证点处的较大偏差则是由于案例点中并未包含土壤有机质含量如此之高的案例,而该点位于巩乃斯县附近的河谷平原区,当前所选环境因子未能将该点与现有案例点进行有效区分。

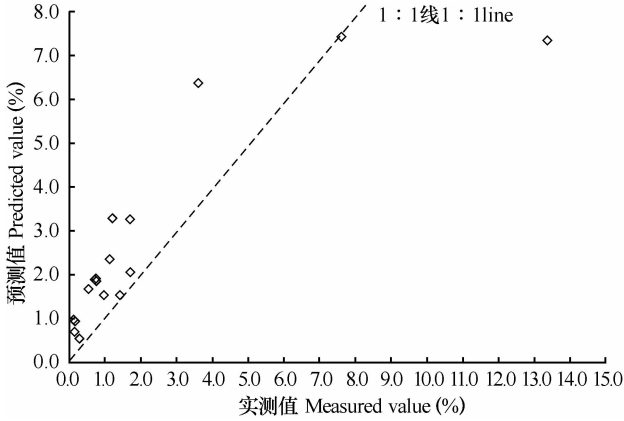


图 6 实测值与预测值 1:1 校验图

Fig. 6 1:1 Validation of measured values and predicted values

(2) 在不确定性信息有效性的检验方面:对每一验证点处的预测残差和不确定性值进行散点分析,如图 7 所示,可以看出,随着不确定性的增加,预测残差呈增大的趋势。说明所得到的不确定性信息能够指示推测结果的可靠程度。需要说明的是,图 7 中只包含了 19 个不确定性低于 0.2 的验证点,这是由于在 0.2 的不确定性阈值下,不确定性高于 0.2 的区域被赋予空值“*No data*”。若有验证点位于这些空值区,没有相应的预测值,因此无法计算残差并参与验证。

(3) 在不确定性阈值的选择方面:该阈值控制着能够被案例点代表的待推测点的空间范围——不确定性阈值越小,与案例点环境更相似的待推测点才能被案例点代表,其土壤属性值才能够用现有案例点进行推测,反之,则不能用现有案例点进行推测,将其属性赋为空值。图 8 显示了不同不确定性阈值下土壤属性图的均方根误差 RMSE。可见,不确定性阈值越高,所得土壤属性图将包含更多的误差。这是由于在高不确定性阈值下,与现有案例点环境条件不够相似的点(即推测不确定性较高的点)也用现有案例点进行了推测,其结果包含较高的误差。因此,可根据所能承受的误差水平来选择合适的不确定性阈值。

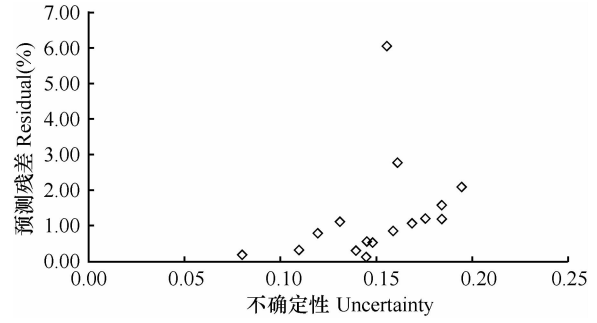


图 7 不确定性阈值为 0.2 时不确定性
与预测残差散点图

Fig. 7 Scatter plot of uncertainty and predicted residual with uncertainty threshold being 0.2

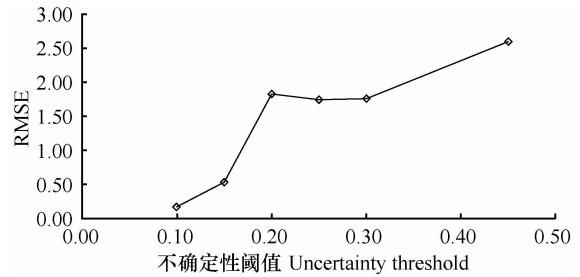


图 8 不同不确定性阈值下土壤属性图的
均方根误差 RMSE

Fig. 8 RMSEs of soil attribute maps with different uncertainty thresholds

4 结论与展望

本文从样点的个体代表性出发,以环境相似度作为对待推测点代表程度的度量,基于环境相似度推测已有样点能够代表地区的土壤属性值,突破了现有方法对样点集全局代表性的严格要求,为利用大尺度空间范围内数量有限、分布任意的样点推测土壤属性空间分布提供了有效途径。此外,本文在样点个体代表性的基础上,通过分析环境特征的相似性,度量由样点代表性引入的推测不确定性。在新疆伊犁地区的应用研究表明,该方法所得不确定性度量与预测残差之间存在正向关系,对推测结果的可靠程度有指示作用,能够为评价推测结果提供有价值的信息。

本文方法所推测的土壤属性空间分布范围并没有覆盖整个研究区,原因是现有样点不具有良好的全局代表性,需要进一步采样来完备样点集对研究区土壤-环境关系的代表性,扩增可推测面积。考虑到本方法所度量的不确定性信息来源于样点代

表性程度,并对推测结果的可靠程度有很好的指示作用,因此在后续工作中可以以不确定性信息分布图为指导设计补充采样方案,达到有效扩增可推测面积并降低推测不确定性的目的。

参 考 文 献

- [1] 毛嘉富,王斌,戴永久. 陆地生态系统模型及其与气候模式耦合的回顾. 气候与环境研究,2006,11(6):763—771. Mao J F, Wang B, Dai Y J. Perspective on terrestrial ecosystem models and their coupling with climate system models (In Chinese). Climatic and Environmental Research, 2006, 11(6): 763—771
- [2] Goovaerts P. Geostatistical modeling of uncertainty in soil science. Geoderma, 2001, 103: 3—26
- [3] Keur P, Iversen BV. Uncertainty in soil physical data at river basin scale. Hydrology and Earth System Science Discussions, 2006, 3: 1 281—1 313
- [4] Refsgaard J C, van der Sluijs J P, Højberg A L, et al. Uncertainty in the environmental modelling process-A framework and guidance. Environmental Modelling and Software, 2007, 22: 1 543—1 556
- [5] 李启全,岳天祥,范泽孟,等. 中国表层土壤全氮的空间模拟分析. 地理研究, 2010, 29(11): 1 981—1 992. Li Q Q, Yue T X, Fan Z M, et al. Spatial simulation of topsoil TN at the national scale in China (In Chinese). Geographical Research, 2010, 29(11): 1 981—1 992
- [6] Henderson B L, Bui E N, Moran C J, et al. Australia-wide predictions of soil properties using decision trees. Geoderma, 2005, 124: 383—398
- [7] Minasny B, McBratney A B. Methodologies for global soil mapping//Boettinger J L, Howell D W, Moore A C, et al. Digital soil mapping: Bridging research, environmental application, and operation. Netherlands: Springer, 2010: 429—436
- [8] McBratney A B, Mendonça S M, Minasny B. On digital soil mapping. Geoderma, 2003, 117: 3—52
- [9] Goovaerts P. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma, 1999, 89: 1—45
- [10] 杨琳,朱阿兴,秦承志,等. 基于典型点的目的性采样设计方法及其在土壤制图中的应用. 地理科学进展, 2010, 29(3): 279—286. Yang L, Zhu A X, Qin C Z, et al. A purposive sampling design method based on typical points and its application in soil mapping (In Chinese). Progress in Geography, 2010, 29(3): 279—286
- [11] FAO/IIASA/ISRIC/ISS-CAS/JRC. Harmonized World Soil Database. Version 1.1. Rome: FAO and IIASA, 2009: 1—43.
- [12] Carre F, McBratney A B, Minasny B. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma, 2007, 141: 1—14
- [13] Hudson B D. The soil survey as paradigm-based science. Soil Science Society of America Journal, 1992, 56: 836—841
- [14] Aamodt A, Plaza E. Case-based Reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, 1994, 7(1): 39—59
- [15] Shi X, Zhu A X, Wang R. Fuzzy representations special terrain feature using a similarity-based approach//Petry F, Robinson V B. Fuzzy modeling with spatial information for geographic problems. New York: Springer, 2005: 233—251
- [16] Zhu A X, Band L E. A knowledge-based approach to data integration for soil mapping. Canada Journal of Remote Sensing, 1994, 20: 408—418
- [17] Mallavan B P, Minasny B, McBratney A B. Homosoil, A procedure for identifying areas with similar soil forming factors. 3rd Global Workshop on Digital Soil Mapping, 2008, Logan, Utah
- [18] 第二次全国土壤普查办公室. 中国土种志. 北京: 中国农业出版社, 1995. The Office for the Second National Soil Survey of China. Soil species of China (In Chinese). Beijing: China Agriculture Press, 1995
- [19] 吴荣镇,权志诚. 新疆伊犁地区土壤. 新疆: 伊犁地区土壤调查办公室, 1985. Wu R Z, Quan Z C. Soils in Yili Region Xinjiang (In Chinese). Xinjiang: The Office for Soil Survey of Ily Region, 1985
- [20] Feng X, Liu G, Chen J M, et al. Net primary productivity of China's terrestrial ecosystems from a process model drive by remote sensing. Journal of Environment Management, 2007, 85: 563—573
- [21] Jenness J S. Calculating landscape surface area from digital elevation models. Wildlife Society Bulletin, 2004, 32: 829—839
- [22] Qin C Z, Zhu A X, Pei T, et al. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. International Journal of Geographical Information Science, 2007, 21(4): 443—458
- [23] Silverman B W. Density estimation for statistics and data analysis. London: Chapman and Hall, 1986: 1—175
- [24] Gower J C. A general coefficient of similarity and some of its properties. Biometrics, 1971, 27: 857—871
- [25] Zhu A X. A personal construct-based knowledge acquisition process for natural resource mapping using GIS. International Journal of Geographic Information Science, 1999, 13(2): 119—141
- [26] Mallavan B P, Minasny B, McBratney A B. Homosoil, a procedure for identifying areas with similar soil forming factors. 3rd Global Workshop on Digital Soil Mapping, 2008, Logan, Utah

LARGE-SCALED SOIL ATTRIBUTE MAPPING METHOD BASED ON INDIVIDUAL REPRESENTATIVENESS OF SAMPLE SITES

Liu Jing^{1,2†} Zhu A-xing^{1,2} Zhang Shujie¹ Qin Chenzhi¹

(1 *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*)

(2 *Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA*)

Abstract The information about spatial distribution of soil attributes over a large area is the basic one required for land surface process modeling. Spatial interpolation based on sampling sites in the field is an important way to acquire such information. The existing spatial interpolation methods usually call for sound representativeness of all the sampling sites for the whole study area in soil attribute spatial distribution. However, limited by budget and/or field accessibility for sampling, in most cases, the sampling sites chosen can hardly reflect comprehensively the spatial distribution of soil attributes of the study area. A soil attribute soil map worked out with the existing interpolation methods based on such sampling is often low in precision and the poor representativeness of the sampling makes it hard to measure the uncertainty it brings about in prediction. In order to make reasonable use of such samples and measure prediction uncertainty effectively, this paper presents a new method that uses the individual representativeness of each sample to predict soil attribute of unvisited locations and to quantify prediction uncertainty. Based on the assumption that the more similar the environment conditions of two sites, the more similar the soil attributes of the two, it is held that every sampling site can be used to represent an area similar to the site in environment, and the representativeness can be measured by similarity degree of the two. By analyzing environmental similarity, uncertainty in prediction can be measured, and with environmental similarity as weight, soil attribute values of the area a sampling site represents can be worked out. This method was once used to predict soil organic matter (SOM) content in the surface soil layer of the Yili District, Xinjiang Uyghur Autonomous Region for validation. As demonstrated, the method is an effective approach to using a sampling site that is poor in global representativeness to predict soil attribute spatial distribution of the area it represents, and the uncertainty of the prediction is positively related to the predicated residual, which can be used effectively to indicate credibility of the prediction.

Key words Environmental similarity; Sample representativeness; Uncertainty; Soil attribute mapping

(责任编辑:陈德明)