

基于多尺度地理环境变量的随机森林空间推测建模

吴 慧, 吴田军

长安大学理学院, 陕西 西安

收稿日期: 2023年10月30日; 录用日期: 2023年12月23日; 发布日期: 2023年12月30日

摘 要

在数据科技快速发展的时代下, 空间推测(Spatial Prediction)是指利用空间理论技术对空间对象进行建模、描述和表示, 并以此为基础对空间对象间的空间关系进行定性或定量分析和处理的过程。机器学习方法作为空间智能领域的常用推测方法之一, 它能够融合多维辅助变量的信息, 提高模型的推测精度, 但是不能有效结合空间位置关系信息来进一步提高推测精度。本文基于随机森林空间推测框架, 将协变量的多尺度特征图层与随机森林空间推测算法结合, 提出了基于多尺度地理环境变量的随机森林空间推测建模方法。应用Meuse数据集对所提出的方法进行对比实验, 并与随机森林方法、随机森林的空间推测方法相对比, 检验所提方法的推测精度。实验表明, 本文提出的基于多尺度地理环境变量的随机森林空间推测建模方法相较于RF、RFsp-MS ($k = 0$)、RFsp-MS ($k = 1$)、RFsp-MS ($k = 2$), 可获得更加精准的空间推测结果, 土壤制图结果也具有更加合理的空间分布和丰富的细节信息。因此, 多尺度地理环境变量的随机森林空间推测方法能够有效结合自变量信息与空间位置关系信息, 有效提高土壤变量推测的精度, 同时能够输出有效的不确定性表达信息。

关键词

空间推测, 随机森林, 多尺度, 土壤锌浓度, 不确定性, Meuse数据集

Random Forest Spatial Prediction Modeling Based on Multi-Scale Geographic Environment Variables

Hui Wu, Tianjun Wu

School of Sciences, Chang'an University, Xi'an Shaanxi

Received: Oct. 30th, 2023; accepted: Dec. 23rd, 2023; published: Dec. 30th, 2023

Abstract

In the era of rapid development of data science and technology, spatial prediction refers to the process of using spatial theory technology to model, describe, and represent spatial objects, and based on this, qualitatively or quantitatively analyze and process the spatial relationships between spatial objects. Machine learning method, as one of the commonly used inference methods in the field of spatial intelligence, can fuse information from multi-dimensional auxiliary variables to improve the inference accuracy of the model. However, it cannot effectively combine spatial position relationship information to further improve the inference accuracy. Based on the framework of random forest spatial inference, this article combines the multi-scale feature layer of covariates with the random forest spatial inference algorithm, and proposes a modeling method for random forest spatial inference based on multi-scale geographic environment variables. Meuse dataset was used to conduct comparative experiments on the proposed method and compare it with the random forest method and the spatial inference method of the random forest to test the inference accuracy of the proposed method. The experiment shows that the random forest spatial inference modeling method based on multi-scale geographic environment variables proposed in this article can obtain more accurate spatial inference results compared to RF, RFsp-MS ($k = 0$), RFsp-MS ($k = 1$), and RFsp-MS ($k = 2$), and soil mapping results also have more reasonable spatial distribution and rich detailed information. Therefore, the random forest spatial inference method for multi-scale geographic environment variables can effectively combine independent variable information and spatial position relationship information, effectively improve the accuracy of soil variable inference, and output effective uncertainty expression information.

Keywords

Spatial Prediction, Random Forest, Multi-Scale, Soil Zinc Concentration, Uncertainty, Meuse Dataset

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

卫星定位系统(GPS)、地理信息系统(GIS)和遥感(RS)等理论与技术主要包括在现代空间信息技术中,结合先进的计算机技术,与此同时对空间数据进行采集、处理与应用[1] [2],是推动社会发展的重要力量。空间数据因为其有时空相关性的特点,与其他类型的数据相比,对独立性变量假设的经典统计学方法难以使用,这也使得几何空间中的牛顿推测等方法并不适用。1970年,地理学第一定律由Tobler [3]提出,也被称作空间相关性定律,其具体指的是“All things are related, but nearby things are more related than distant things”。空间相关性,指的是“地物之间的相关性与距离有关,通常来说,任何东西与别的东西之间都是相关的,但近处的东西比远处的东西相关性更强”。空间数据的分析和应用是由地理学第一定律提供理论基础,在地学领域中,最为重要的研究问题之一是空间推测中将离散点的测量数据转化为连续数据。通过几十年的不断研究与发展,空间推测算法趋于完善,并逐渐被应用到地质勘探、土壤环境、海洋水质、空气质量等诸多领域。但随着社会生产力水平的进步与提高,进一步凸显出对地理空间信息的多样性、时效性需求,这也进一步促使空间推测算法对地理空间智能领域提出了更高的要求。

总之, 进一步发展空间推测模型, 对空间制图水平提出了更高的要求, 对机器学习领域具有重要的现实意义。

目前, 空间推测方法主要包括以下几类: 地统计方法[4]、回归模型方法[5]、机器学习方法。目前空间推测方法中广泛应用的方法是地统计方法。地统计方法的基础是描述空间自相关性的地理学第一定律和描述空间异质性的第二定律。基于地理学第一定律的最近邻插值、反距离加权、径向基函数和局部多项式等确定性推测方法, 以及普通克里金(OK)、简单克里金和回归克里金(RK)等都包括在地统计方法中。与确定性推测方法相比, 地统计方法能在给出无偏最优推测结果的同时, 得到推测结果的不确定性估计, 但是难以解决空间异质性问题。例如: 2012年, 姚旭峰等[6]考虑空间中点和点之间的相关性, 将空间统计学中的克里金方法引入到机械加工的平面度误差推测中, 得出了克里金方法能够显著提高对平面度误差推测的准确程度的结论。

回归模型方法以地理学第三定律为基础, 可以分为以下几类: 多元线性回归模型、逐步回归模型、广义线性模型、广义加性模型。地理学第三定律认为: 地理环境越相似, 地理目标特征越相近。多元线性回归模型具有解释性强等优点, 遥感、地形、植被、土壤电导率等数据均可作为环境变量与此模型相结合。然而, 传统的MLR受限于数据分布: 例如, 若原始数据不服从正态分布, 推测结果误差可能会比较大。逐步回归模型(Stepwise Regression Model, SRM)可在结合环境变量上对模型进行优化, 利用AIC信息指标, 通过统计指标和异常值测试进行识别和剔除冗余变量, 将多重共线性降至最低。回归模型方法可以充分利用自变量数据, 改善模型的推测精度。但是偶尔会出现的情况是部分数据难以全部精确获得, 为了防止模型过拟合, 与此同时, 模型会产生残差。因为残差往往具有空间自相关性, 所以在考虑属性相似性的基础上, 综合利用空间位置关系信息能够在一定程度上提高空间推测精度。比如: 2022年, 关莹莹等[7]利用地理加权回归模型计算了分级人居环境适宜性指数, 并探讨了其与人口、GDP密度的空间一致性, 验证了适宜性较好的土地的利用效率。

机器学习方法在处理多维协变量数据方面具有极其重要的作用, 2001年, Breiman [8]提出随机森林(RF), 并且说明了RF适用于回归问题; 2018年, Tomislav 等[9]提出了随机森林空间推测框架(Random Forest for spatial predictions framework, RFsp), 其将预估点到样本点的缓冲距离作为推测因子加入到模型中, 通过实验证明其相较于地统计方法推测精度得到提高; 2020年, Sekulić 等[10]提出了随机森林空间插值法(Random Forest Spatial Interpolation, RFSI)。该方法将邻近点的观测值与其到预估点的距离作为推测因子加入到随机森林中, 以改进空间插值结果, 并验证RFSI的推测结果优于克里金。目前, 空间制图领域最前沿的研究方向之一是如何将空间位置关系信息融入到机器学习方法中[11], 提高空间推测精度。鉴于此, 本文提出了基于多尺度地理环境变量的随机森林的空间推测模型(Spatial Prediction Model for Random Forest Based on Multi-scale Geographical Environment Variables), 主要是在随机森林空间推测框架的基础上, 将协变量的多尺度特征图层引入到模型中, 多尺度策略较单尺度参数更能满足建模的需要, 并以荷兰默兹河的土壤锌浓度数据作为案例, 与RF、RFsp对比分析, 验证了该方法的空间推测效果。

2. 研究区与数据来源

受地区、人为活动等各种复杂因素影响, 在空间推测研究方面土壤锌浓度[12]广泛应用, 因为土壤锌浓度往往呈现出比较复杂的空间分布趋势, 本文通过Meuse数据集所对应的研究区以及包含的具体数值加以介绍, 来更好地解释多尺度地理环境变量的随机森林空间推测模型[13]。

2.1. 研究区概况

本文选取收集在荷兰Stein村附近的默兹河洪泛区作为研究区域, 位于北纬 $51^{\circ}42'54''$ 、东经 $4^{\circ}40'4''$

之间。

2.2. 数据来源

本研究数据集包括了 155 个站点观测的土壤锌浓度数据、DEM 模型数据[14]、地表水数据等，具体说明如下：

1) 土壤锌浓度数据。研究采用的是 sp 软件包中包含的 Meuse 数据集。该数据集提供了位置以及观察地点的许多土壤和景观变量，其中土壤锌浓度来自面积约为 $15\text{ m} \times 15\text{ m}$ 的复合样品。该数据集包括了 155 个站点的土壤锌浓度，单位为 mg/kg。具体信息如表 1 所示。

Table 1. Statistical information of soil zinc concentration data at 155 stations of Meuse

表 1. 默兹河 155 个站点土壤锌浓度数据的统计信息

名称	数值
观察点的数目	155
最大值	1839
最小值	113
均值	326
方差	134743.2
标准差	367.0738

2) DEM 模型数据。数字高程模型(Digital Elevation Model, 简称 DEM)是 DTM 中最基本的部分，它是对地球表面地形地貌的一种离散的数学表达。<http://ahn.nl/> [15]是本研究采用的 DEM 模型数据的获取途径，其空间分辨率为 1 km。

3. 研究方法

3.1. 空间推测方法

目前，在空间推测方法的发展过程，可以分为两个阶段：第一个阶段是确定性和地统计等传统科技方法的发展[16]，其理论性高，有大量假设条件，并且由于数据集的复杂性，很难满足。第二个阶段是机器学习方法的发展，机器学习作为一种新方法既属于统计学模型范畴，也包含半经验混合模型特性，可基于大数据获取对象特征，也可采取结合多种模型的方法来发挥各自优点，具有耗时短、准确度高、分辨率高和推测时间长等优点。

3.1.1. 随机森林空间推测方法与模型

随机森林(RF)是袋装树木的延伸，是一种基于分类回归的集成学习方法[17]。一些基准研究已经证明它是目前可用的最好的机器学习技术之一。

RF 本质上是一种非空间的空间推测方法[18]-[23]，因为在 MLA 模型参数的估计过程中忽略了采样位置和一般的采样模式。这可能会导致次优推测和可能的系统性高推测或低推测，特别是在目标变量的空间自相关性高和点模式显示出明显的抽样偏差。为了克服这个问题，我们提出了 RFsp (Random Forest for spatial prediction)模型，模型的表达式为：

$$Z(s) = f(x_G, x_R, x_P)$$

其中， $Z(s)$ 为空间样本点 s 上的目标变量取值； x_R 是表面反射率的协变量，即通常是遥感图像的光谱信息；

x_p 是基于过程的协变量, 即辅助变量; x_G 是考虑观测值之间地理邻近性和观测值之间空间关系的协变量 (模拟克里格法中使用的空间相关性)。

3.1.2. 多尺度地理环境变量的随机森林空间推测方法与模型

在 RFsp 中使用观测和未知位置的解释变量数据来研究空间关联性[24]。然而, 在样本位置外提取不同地理信息的模型仍然有限, 而且仍然越来越需要更有效地提取整个空间的地理信息。为了解决上述问题, 本研究在 RFsp 的基础上将协变量的多尺度特征图层引入到模型中, 根据模型精度, 选择最适宜的建模尺度。其大致可以分为改进协变量和构建模型两部分, 核心内容是针对 RFsp 中的地理环境变量进行改进, 根据训练点和预估点的坐标, 计算出预估点到训练点的欧式距离:

$$d_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

式中: (x_i, y_i) 为预估点的坐标, (x_0, y_0) 为训练点的坐标, 尺度参数的表达式为:

$$r_k = r_0 + \Delta r \cdot k (k = 0, 1, k)$$

其中, 尺度间隔为 $\Delta r = 100$ m, 选取 0 m、100 m、200 m 作为定量的尺度标准, 将预估点到训练点的欧氏距离 d_i 与尺度参数 r_k 进行比较, 当 $d_i \leq r_k$ 时, 将范围内所有的点对应的协变量的值取平均值 x_i , 计算其平均值 x_i 作为协变量的新的取值, 从而得到一组不同尺度的协变量的值, 将其作为建模变量输入, 与原有的环境推测变量构建形成多尺度地理环境变量的随机森林空间推测(RFsp-MS)方法与模型。模型的表达式为:

$$Z(s) = f(x'_G, x_R, x_p)$$

其中, x_G 是经过计算之后得到的新的地理环境变量, 具体展开如下:

$$\begin{aligned} & x_G x'_G \\ x_1 & \rightarrow \{x_{1r_0}, x_{1r_1}, \dots, x_{1r_k}\} \cup \\ x_2 & \rightarrow \{x_{2r_0}, x_{2r_1}, \dots, x_{2r_k}\} \cup \\ & \dots \\ x_i & \rightarrow \{x_{ir_0}, x_{ir_1}, \dots, x_{ir_k}\} \end{aligned}$$

多尺度地理环境变量的随机森林空间推测算法流程图可以如下表示:

协变量大致可以分为三类: 地表水、到河边的距离、DEM 模型数据, 数据处理之后, 得到了地理环境变量的多尺度特征图层以及参与推测的协变量有土地利用、铅浓度、土壤类型、海拔等, 再加入已知训练点的土壤锌浓度建立模型。

建模的步骤可以如下:

步骤 1: 训练 RFsp-MS 模型

- 1) 删除 155 个样本点中的缺失值, 并将剩余站点的观测值与推测变量值组合为 Data;
- 2) Data 按 7:3 的比例, 划分为训练集与测试集;
- 3) 对训练集中的每个点, 根据欧式距离公式计算其到预估点的距离 Dist, 将其与不同的尺度参数进行比较, 计算出新的地理环境变量, 并将其作为建模变量输入;
- 4) 训练 RFsp-MS 模型, 推测测试集的土壤锌浓度并计算测试集的 RMSE;
- 5) 比较不同的尺度参数对应的 RMSE, 选取最小的 RMSE, 并输出其对应的随机森林的参数。

步骤 2: 预估点推测

- 1) 对步骤 1 输出的最优随机森林参数以及新的地理环境变量构建 RFsp-MS 模型;
- 2) 利用构建的 RFsp-MS 模型推测预估点的土壤锌浓度。

步骤 3: 交叉验证

- 1) 运用五折交叉验证法来进行验证;
- 2) 初始化 $k = 1$;
- 3) k 对应的数据集为测试集, 除 k 之外的 4 份数据集为训练集, 对步骤 1 输出的随机森林的最优参数以及新的地理环境变量构建 RFsp-MS 模型, 利用 RFsp-MS 模型推测测试集的土壤锌浓度;
- 4) 执行 $k = k + 1$, 返回第三步, 直到 $k = 5$;
- 5) 根据公式计算模型评价指标 MAE、RMSE、CCC、 R^2 。

3.2. 模型评价标准

为验证实验推测结果的有效性及其准确性, 选取 RF、RFsp-MS ($k = 0$)及不同尺度的随机森林的空间推测模型进行对比, 并利用平均绝对误差(Mean Absolute Error, MAE), 其指的是模型推测值与真实值的误差绝对值的平均值; 均方根误差(Root Mean Square Error, RMSE), 反映的是模型的精度, 误差越大, 表示模型的精度越低; 决定系数(Coefficient of Determination, R^2), 在回归模型中该标准发挥了评价模型好坏的作用。具体而言, 回归模型对因变量 y 变化的解释程度是由其评估的。 R^2 越大, 表示模型的拟合度越好, 可解释程度越高; 反之, 该值越小表示模型拟合有问题; 一致相关系数(Concordance Correlation Coefficient, CCC)这四个评价标准加以比较, 公式如下:

$$MAE = \frac{1}{N} \sum_{j=1}^N |\hat{z}(s_j) - z(s_j)|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{z}(s_j) - z(s_j))^2}$$

$$R^2 = \left[1 - \frac{SSE}{SST} \right] \%$$

$$CCC = \frac{2\rho\sigma_{\hat{z}}\sigma_z}{\sigma_{\hat{z}}^2 + \sigma_z^2 + (\mu_{\hat{z}} - \mu_z)^2}$$

式中: $\hat{z}(s_j)$ 是交叉验证点 s_j 的预估值; $z(s_j)$ 是交叉验证点 s_j 的实际观测值; N 是交叉验证点的总数; SSE 是交叉验证点的误差平方和; SST 是总平方和; $\mu_{\hat{z}}$ 与 μ_z 是预估值与观测值的均值; $\sigma_{\hat{z}}$ 与 σ_z 是预估值与观测值的标准差; ρ 为预估值与观测值之间的相关系数。

4. 空间推测实验与结果分析

4.1. 空间推测制图结果分析

基于 RF、RFsp-MS ($k = 0$)以及不同尺度的随机森林的空间推测模型的空间推测结果以及不确定性图(如图 1、图 2 所示), 可知土壤锌浓度含量推测结果大体上相似, 但局部特征差异比较明显: 由图 1(d)推测结果可知, 从颜色的深浅上来区分, 默兹河东部地区土壤锌浓度含量较低, 西部地区土壤锌浓度含量较高。由图 2(c)的不确定性结果可知, 推测标准差保持在较低的水平集中在绝大部分地区, 较高的地区分布主要在默兹河的南部区域, 由实际地理位置可以得出, 默兹河的东南部地区海拔比其他地区高, 导致站点数据稀缺, 导致空间推测方法难以在这一地区捕捉到有效信息, 如需更准确地掌握该地区的土壤锌浓度情况, 还需要进一步获取样本点的信息。相比于其他方法, 多尺度地理环境变量的随机森林空间

推测方法由于使用了辅助数据, 从而具有更丰富的细节信息, 空间分布变化也更加合理。

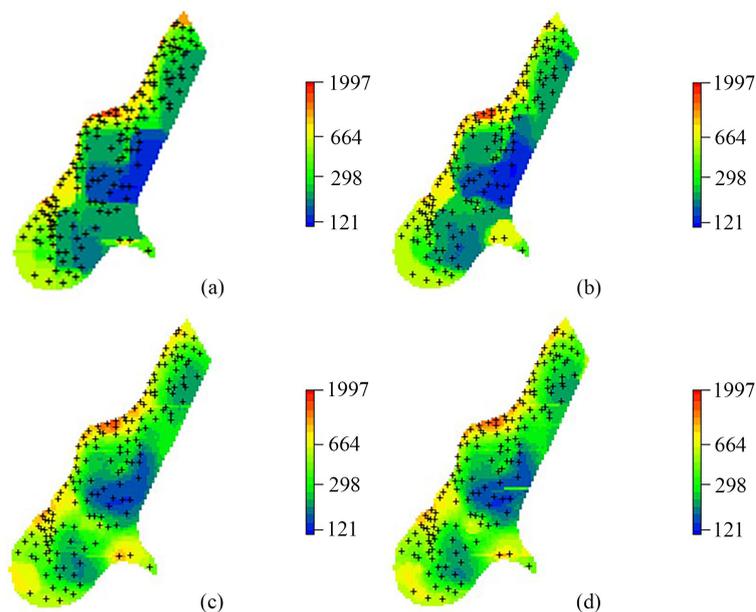


Figure 1. RF (a), RFsp-MS ($k = 0$) (b), RFsp-MS ($k = 1$) (c), and RFsp-MS ($k = 2$) (d) soil zinc concentration prediction maps

图 1. RF (a)、RFsp-MS ($k = 0$) (b)、RFsp-MS ($k = 1$) (c)、RFsp-MS ($k = 2$) (d)土壤锌浓度推测图

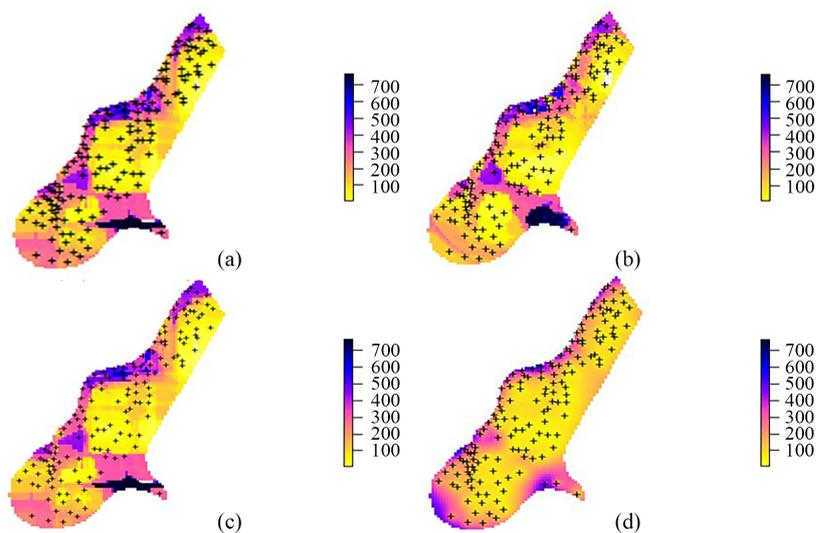


Figure 2. RF (a), RFsp-MS ($k = 0$) (b), RFsp-MS ($k = 1$) (c), and RFsp-MS ($k = 2$) (d) soil zinc concentration prediction uncertainty maps

图 2. RF (a)、RFsp-MS ($k = 0$) (b)、RFsp-MS ($k = 1$) (c)、RFsp-MS ($k = 2$) (d)土壤锌浓度推测不确定性图

与 RF、RFsp-MS ($k = 0$)、RFsp-MS ($k = 1$) 三种模型的不确定性图相比, RFsp-MS ($k = 1$) 模型的标准差相对较小, 但是其标准差趋势呈现均匀分布, RFsp-MS ($k = 0$) 的不确定性信息包含的信息相对较少, 即无法更加标准地度量特殊点(预估点周围的样本点相对较少)的标准差。基于 RF 的三种不同的空间推测方法, RF、RFsp-MS ($k = 0$)、RFsp-MS ($k = 1$) 与 RFsp-MS ($k = 2$) 在推测标准差方面不同之处主要集中在

默兹河偏东南的区域, RF、RFsp-MS (k = 0)、RFsp-MS (k = 1)的标准差基本保持在相对较高的水平, RFsp-MS (k = 2)虽然减少部分地区的标准差, 但仍有小部分区域保持在较高水平。结合四种不同的空间推测方法的推测结果与不确定性信息来看, RFsp-MS (k = 2)推测结果保持着较高的精度水平, 并且在不确定性方面, RFsp-MS (k = 2)与 RF、RFsp-MS (k = 0)相比标准差更小, 与 RFsp-MS (k = 1)相比, 不确定性更加具有时效性与信息性[25]-[28]。因此, RFsp-MS (k = 2)模型的空间推测结果的分布更加合理。

4.2. 交叉验证结果分析

本文提出的多尺度的随机森林的空间推测模型[29]与 RF、RFsp-MS (k = 0)、RFsp-MS (k = 1)交叉验证结果如表 2 所示。通过表 2 可以得出, 在推测精度方面, RFsp-MS (k = 2)最高, RF 最低, 从模型的拟合程度来看, RFsp-MS (k = 1)、RFsp-MS (k = 2)都高于 0.5, 说明模型的拟合度较好。

Table 2. Accuracy of four speculative methods based on fifty fold cross validation
表 2. 基于五折交叉验证的四种推测方法的精准度

方法	MAE	RMSE	CCC	R ²
RF	19.68986	29.49024	0.5036752	0.3466746
RFsp-MS (k = 0)	18.22457	26.97146	0.55	0.4535104
RFsp-MS (k = 1)	17.08588	25.09717	0.5354774	0.5711109
RFsp-MS (k = 2)	16.99222	24.14737	0.5481469	0.5868849

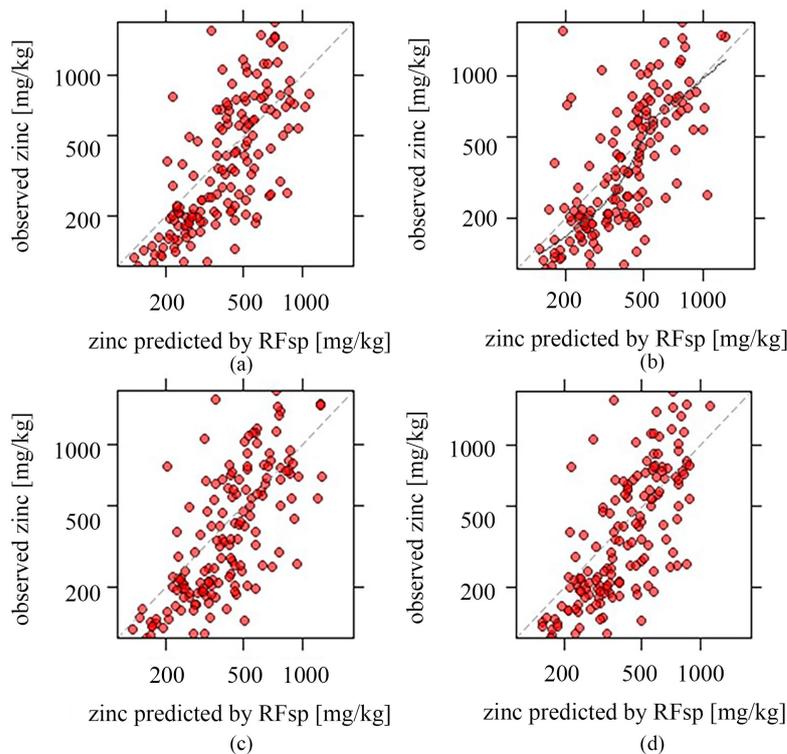


Figure 3. Correlation diagrams of RF (a), RFsp-MS (k = 0) (b), RFsp-MS (k = 1) (c), and RFsp-MS (k = 2) (d) based on observation values and predicted values

图 3. RF (a)、RFsp-MS (k = 0) (b)、RFsp-MS (k = 1) (c)、RFsp-MS (k = 2) (d)基于观测值与推测值的相关图

进一步分析, 由于 RF [30]是一种基于决策树的集成学习的方法, 其没有考虑空间自相关性, 可能会导致次优推测, 在 Meuse 数据集的推测结果及图 3 的观测值与推测之的相关图中, 没有表现出很高的推测精度; RFsp-MS ($k = 0$)将空间位置关系考虑在内, 在本次实验中, 推测精度处于一个中等的水平, 但是模型训练过程中距离的应用可能导致推测结果的偏差; RFsp-MS ($k = 2$)是随机森林的空间推测与协变量的多尺度特征图层的结合, 其在 Meuse 数据集的推测结果中, 表现出了更高的推测精度, 因此, 相较于 RF、RFsp-MS ($k = 0$), RFsp-MS ($k = 2$)在空间推测方面不失为一种好的选择。

5. 总结与展望

总览全部内容, 本文主要的创新内容体现在基于随机森林空间推测的框架, 结合地理环境变量的多尺度特征图层, 提出了多尺度地理环境变量的随机森林的空间推测的方法[31], 该方法在空间推测中能够有效结合反映土壤锌浓度及影响因素的多维辅助变量信息与空间位置关系信息, 可以有效地提高模型的推测精度, 得到更为准确的空间推测结果。在 RFsp 模型中, 我们使用了 155 个训练点来建模, 随着数据大小的增加和更多的协变量层的增加, RFsp 通常会在验证点导致令人满意的 RMSE 和 MAE, 而在交叉验证标准差中没有显示出空间自相关[13]。本文所提出的空间推测方法相较于其他方法, 推测精度提高了 10% 以上。同时, 其推测结果具有更丰富的细节变化信息和合理的空间变化。本文所提出的方法较适用于具有复杂空间异质性的土壤变量下的空间推测, 但是其推测效果还有待于在其他数据集中检验。

参考文献

- [1] 李海涛, 邵泽东. 空间插值分析算法综述[J]. 计算机系统应用, 2019, 28(7): 1-8.
- [2] 王福涛, 李景文, 李占元. GIS 空间数据表达与存储研究综述[J]. 中国水运(学术版), 2006, 6(11): 139-141.
- [3] Tobler, R.W. (2016) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240. <https://doi.org/10.2307/143141>
- [4] 尹镇南. 地质统计学(空间信息统计学)基本理论与方法应用[M]. 北京: 地质出版社, 2012: 83-85.
- [5] Mohanasundaram, S., Udmale, P., Shrestha, S., et al. (2020) A New Trend Function-Based Regression Kriging for Spatial Modeling of Groundwater Hydraulic Heads under the Sparse Distribution of Measurement Sites. *Acta Geophysica*, **68**, 751-772. <https://doi.org/10.1007/s11600-020-00427-y>
- [6] 姚旭峰, 杜世昌, 费兰, 等. 基于克里金方法的平面建模和误差预测[C]//中国质量协会, 卓越国际质量科学研究院. 使命与责任——以质量方法促转型升级——第五届中国质量学术与创新论坛论文集(上). 北京: 中国质量杂志出版社, 2012: 6.
- [7] 关莹莹, 李雪铭, 杨俊, 等. 基于地理加权回归模型的辽宁省城市人居环境综合适宜性评价[J]. 地理科学, 2022, 42(12): 2097-2108.
- [8] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [9] Tomislav, H., Madlene, N., Wright, M.N., Heuvelink, G.B.M. and Gräler, B. (2018) Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ*, **6**, e5518. <https://doi.org/10.7717/peerj.5518>
- [10] Sekulić, A., Kilibarda, M., Heuvelink, B.G., et al. (2020) Random Forest Spatial Interpolation. *Remote Sensing*, **12**, Article 1687. <https://doi.org/10.3390/rs12101687>
- [11] 黄玉婧, 徐智, 单添敏, 等. 基于多尺度空间的直升机滚动轴承故障诊断[J]. 测控技术, 2022, 41(10): 52-57, 65.
- [12] 李召艺. 基于随机森林的北京市城市热岛效应与三维景观格局多尺度关系研究[D]: [硕士学位论文]. 长春: 吉林大学, 2022.
- [13] 刘亚彬, 刘大有. 空间推理与地理信息系统综述[J]. 软件学报, 2000, 11(12): 1598-1606.
- [14] 钱江波, 陈涤, 王夏晖, 等. 基于机器学习的区域土壤重金属污染风险诊断研究[J/OL]. 环境工程, 2023: 1-12. <https://kns.cnki.net/kcms/detail/11.2097.x.20230609.1146.006.html>
- [15] 余浪, 胡德勇, 汪小勇. 赫章县铅锌矿典型区域土壤及水体重金属污染现状研究[J]. 中国资源综合利用, 2022, 40(8): 148-151.

- [16] 盛美玲. 基于环境相似性的作物生长模型参数空间推测方法研究[D]: [博士学位论文]. 南京: 南京师范大学, 2019.
- [17] 付杰, 宋伦, 雷利元, 等. 最优尺度和随机森林算法下盘锦海岸带遥感监测[J]. 遥感信息, 2022, 37(2): 7-15.
- [18] 王铭鑫, 范超, 高秉博, 等. 融合半变异函数的空间随机森林插值方法[J]. 中国生态农业学报(中英文), 2022, 30(3): 451-457.
- [19] 黄明. 基于地统计方法的成都市住宅价格空间格局分析[J]. 经纬天地, 2019(5): 8-13.
- [20] Pekel, J.F., Cottam, A., Gorelick, N. and Belward, A.S. (2016) High-resolution Mapping of Global Surface Water and Its Long-Term Changes. *Nature*, **504**, 418-422. <https://doi.org/10.1038/nature20584>
- [21] 焦思佳, 吴田军, 董世英, 等. 基于反距离加权随机森林的空间推测方法研究[J]. 昆明理工大学学报(自然科学版), 2022, 47(4): 46-54.
- [22] 曾芸芸, 张红英, 袁明东. 多尺度融合的双分支特征提取人群计数算法[J/OL]. 计算机工程与应用, 2023: 1-10. <https://kns.cnki.net/kcms/detail/11.2127.TP.20230628.1602.013.html>
- [23] 杨鹏杰, 徐宇, 郑晨一. 基于多尺度自适应残差卷积神经网络的新能源配电网故障定位技术[J]. 水利水电技术(中英文), 2023, 54(S2): 439-446.
- [24] 李玲玲, 刘劲松, 李智, 等. 人口密度随机森林模型优化实验研究[J]. 地理学报, 2023, 78(5): 1304-1320.
- [25] 张志斌, 蔡来良, 杜庄, 等. 多尺度点云特征随机森林滤波算法[J]. 激光杂志, 2023, 44(2): 76-82.
- [26] 张亦然, 刘廷玺, 董新, 等. 基于多源遥感和机器学习方法的科尔沁沙地植被覆盖度反演[J]. 中国沙漠, 2022, 42(3): 187-195.
- [27] 丁学辉, 许海林, 罗颖婷, 等. 基于随机森林特征优选与 MAEPSO-ELM 算法的变压器 DGA 故障诊断[J]. 电力科学与技术学报, 2022, 37(2): 181-187.
- [28] 陈飞, 王斌, 刘婷, 等. 基于时移多尺度注意熵和随机森林的水电机组故障诊断[J]. 水利学报, 2022, 53(3): 358-368, 378.
- [29] 王雪洁, 施国萍, 周子钦, 等. 基于随机森林算法对 ERA5 太阳辐射产品的订正[J]. 自然资源遥感, 2022, 34(2): 105-111.
- [30] 余新华, 赵维清, 朱再春, 等. 基于遥感和作物生长模型的多尺度冬小麦估产研究[J]. 光谱学与光谱分析, 2021, 41(7): 2205-2211.
- [31] 董红瑶, 王弈丹, 李丽红. 随机森林优化算法综述[J]. 信息与电脑(理论版), 2021, 33(17): 34-37.