

DOI: 10.11766/trxb201711250563

不同方法预测苏南农田土壤有机质空间分布对比研究*

谢恩泽^{1, 2} 赵永存^{1, 2†} 陆访仪^{1, 2} 史学正^{1, 2} 于东升^{1, 2}

(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 中国科学院大学, 北京 100049)

摘要 准确把握土壤有机质(SOM)的空间分布规律对于土壤资源的高效持续利用具有重要意义。以江苏南部为研究区, 以辅助因子与SOM的相关性强弱及辅助因子的可获取性为切入点, 运用普通克里格(OK)、回归克里格(RK)和随机森林(RF)方法, 结合地形、气候、土壤类型、土壤理化性质和施肥、碳投入等辅助数据预测了苏南地区农田SOM含量(0~20 cm)的空间分布。结果表明, 三种方法预测的SOM空间分布总体趋势相似, 表现为东高西低, 但局部分异还存在差异; OK预测的精度最低, 100次预测的均方根误差(RMSE)均值为 $6.97 \text{ g} \cdot \text{kg}^{-1}$ 。RK和RF的预测精度则均高于OK方法, 表现为整合与SOM相关性最强的辅助因子全氮(TN)时, RK和RF预测的RMSE分别降低至 $5.25 \text{ g} \cdot \text{kg}^{-1}$ 和 $4.97 \text{ g} \cdot \text{kg}^{-1}$, 而移除相关性最强的辅助因子TN后, RK和RF预测的RMSE亦较OK方法低, 分别为 $6.21 \text{ g} \cdot \text{kg}^{-1}$ 和 $6.29 \text{ g} \cdot \text{kg}^{-1}$; 移除TN后, RK的预测精度稍高于RF, 表明在其他辅助数据与SOM相关性相对较弱的条件下, RK方法有助于提高本研究区SOM预测精度; 同时, 尽管RK和RF的预测精度依然较OK高, 但RK和RF对SOM方差的解释度则分别由51%和55%降低至29%和28%。这表明, 目前容易获取且相对廉价的辅助数据, 对本研究区的SOM空间预测方面, 还面临着数据质量低、预测精度不足等问题。

关键词 空间预测; 克里格插值模型; 随机森林预测模型; 有机质

中图分类号 S159 **文献标识码** A

土壤有机质(Soil organic matter, SOM)是土壤的重要组成部分, 尽管只占土壤总质量的很小一部分, 但其作为土壤肥力形成的基础, 不但影响土壤质量及关键功能, 而且在陆地生态系统物质循环中扮演重要角色^[1-3]。因此, 利用不同空间预测方法对比预测农田SOM, 对于准确把握SOM的空间分布规律和土壤资源管理的科学决策, 进而保障粮食安全、气候安全及土壤安全具有重要意义。

地统计学方法广泛用于土壤属性空间预

测^[4]。而克里格法是其中最准确, 强大的插值方法之一^[5]。徐尚平等^[6]基于半方差函数和普通克里格方法分析了内蒙古地区SOM含量的空间结构特征; Mushtaq等^[7]采用普通克里格方法描绘了克什米尔农业区土壤微量元素的空间分布; Heba等^[8]则利用普通克里格方法对尼罗河三角洲地区的土壤碳库和氮库的空间分布特征进行了深入分析。这些结果表明, 克里格方法在不同区域、不同土壤性质的空间变化中均能得到较好的预测结

* 国家自然科学基金项目(41471177)和国家重点研发计划(2017YFA0603002)项目资助 Supported by the National Natural Science Foundation of China (No. 41471177) and the National Key Research and Development Program of China (No. 2017YFA0603002)

† 通讯作者 Corresponding author, E-mail: yczhao@issas.ac.cn

作者简介: 谢恩泽(1993—), 男, 安徽池州人, 博士研究生, 主要从事土壤有机碳空间预测及不确定性评价研究。E-mail: ezxie@issas.ac.cn

收稿日期: 2017-11-25; 收到修改稿日期: 2018-01-29; 优先数字出版日期(www.cnki.net): 2018-04-11

果。遥感和数字制图技术的快速发展,为土壤空间预测提供了大量廉价且与土壤形成过程直接或间接相关的辅助数据,比如,DEM、植被指数及生物量、光谱反射率、数字化的土壤图、土地利用/覆盖图等。这些辅助数据提供了关于土壤性质变异性 的先验信息,从而使得整合辅助数据先验信息的回归克里格等混合建模方法得到了快速发展。Hengl等^[9]以土壤图和DEM获取的地形因子为辅助数据,采用回归克里格方法预测了克罗地亚SOM的空间分布;Andrew等^[10]则采用回归克里格方法整合红外光谱数据,大大提高了土壤全碳、氮等土壤属性的预测精度。回归克里格方法不但反映了辅助因子对土壤属性空间分异的潜在影响,同时,趋势项的分离和残差的再预测也使得其空间预测精度通常要优于普通克里格方法。近年来,随机森林作为经典的机器学习模型之一,由于具有处理高维度数据、防止过度拟合、提供辅助因子的重要性度量等优势,在土壤空间预测,特别是地形分异明显地区的土壤空间预测中也逐渐得到广泛应用。比如,Mareike等^[11]在厄瓜多尔南部高程起伏1.7~3.2 km的安第斯山脉典型区中,利用随机森林方法成功地预测了土壤质地的空间分布;Dharumaran等^[12]以印度南部安得拉邦阿嫩达布尔区的Bukkarayatasamudrum mandal为研究区,在高程(295~595 m)较复杂多变的情况下利用地形、植被等辅助数据,采用随机森林方法预测了土壤有机碳、pH和电导率的空间分布;而Zhi等^[13]则通过随机森林方法并整合DEM获取的地形因子、遥感影像获取的光谱反射率及植被指数等辅助数据预测了青藏高原地区高山草毡层的空间分布。随机森林方法在地形复杂且人为活动影响相对较弱的区域土壤空间预测中得到了较为成功的应用,而在地形平坦、人为影响强烈的地区应用效果则依然不清楚。

上述研究可以看出,不同预测方法利用辅助因子在平原地区的应用仍需进一步探讨。所以,选择地势平坦的江苏南部(苏南)为研究区,以影响土壤生态系统服务功能的关键因子SOM为研究对象,以辅助因子与SOM的相关性强弱及辅助因子的可获取性为切入点,通过回归克里格和随机森林方法与传统的普通克里格方法对比分析来研究不同空间预测方法对该区SOM空间分布预测的影响,

以便为准确预测该区SOM空间分布规律,进而为通过科学调控不同区域SOM来实现关键土壤生态系统服务功能提供科学依据。

1 材料与方法

1.1 研究区概况

研究区为江苏省南部地区($30^{\circ} 44' \sim 32^{\circ} 16' N$, $118^{\circ} 29' \sim 121^{\circ} 19' E$)。该区属北亚热带季风气候,四季分明,雨量充沛,年均温 $16^{\circ}C$,年均降雨量1 000 mm。全区地势西高东低,太湖平原地区地势平坦,河流湖泊较多,主要土壤类型为水稻土,沿江地区则以灰潮土为主。研究区西部及西南部多为丘陵、低山岗地,土壤类型主要为黄棕壤。研究区农耕条件优越,工业经济发展迅速,城镇化水平较高,是长三角地区经济最为发达的地区之一。

1.2 数据来源与处理

土壤采样点位置根据该区20世纪80年代第二次土壤普查结果进行布设,共413个(图1),采样时间为2000年,采样深度为0~20 cm。采样同时记录了样点的GPS位置信息以及环境条件和主要农业管理措施信息。土壤有机质(SOM)含量采用重铬酸钾氧化-外加热法测定,全氮采用半微量开氏法测定,有效磷采用碳酸氢钠浸提-钼锑抗比色法测定,速效钾采用乙酸铵浸提-火焰光度法测定,pH采用电位法(土:水=1:2.5)测定,机械组成采用吸管法测定^[14]。其中,土壤全氮、有效磷、速效钾、pH及黏粒含量主要用于SOM空间预测的辅助土壤属性因子。

SOM空间预测的其他辅助因子数据包括:20世纪80年代初始SOM含量(Initial SOM, SOM1980),来源于研究区各县/区的土壤普查报告;江苏省1:20万土壤图;江苏省30 m分辨率DEM计算的主要地形因子,包括高程(Elevation, Elev)、坡度(Slope)、平面曲率(Plane curvature, Planc)、剖面曲率(Profile curvature, Profc)和复合地形指数(Compound terrain index, CTI);1 km×1 km的年均气温(Mean annual temperature, Mat)和降雨量(Mean annual precipitation, Map)数据(1976—2005年平均,来源于碳专项)。除了上述常用于SOM空间辅

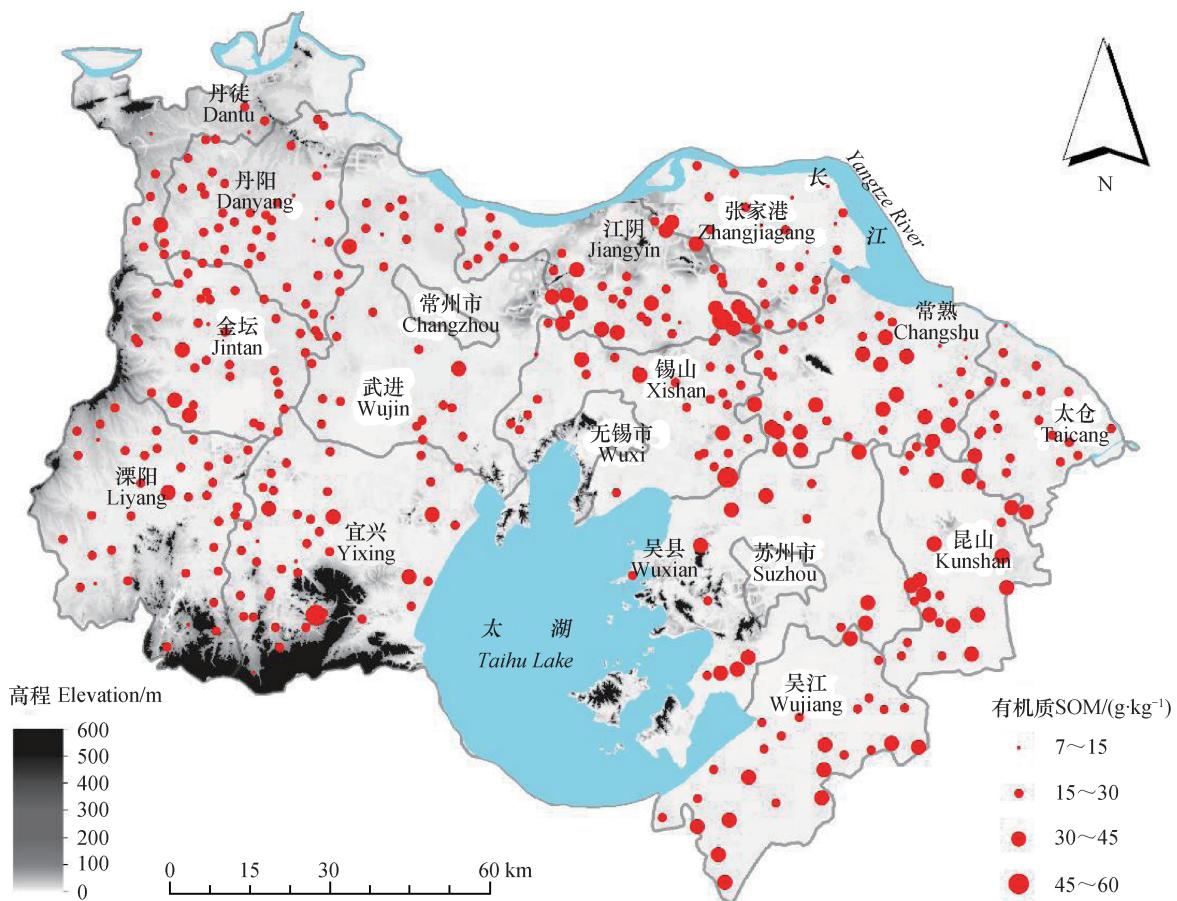


图1 苏南地区地形及土壤采样点分布图

Fig. 1 Topographic map of South Jiangsu and distribution of soil sampling sites

助预测的因子外,本研究也考虑了一些容易获取且免费的数据产品,以检验其在提高SOM空间预测精度中的可行性,这些因子包括:土壤温度(Soil temperature, ST)和土壤湿度(Soil moisture, SM)1 km×1 km栅格数据(2000年月平均,分别来源于欧洲中期天气预报中心(ECMWF)和Earthdata);遥感反演的5 km×5 km净初级生产力(Net Primary Production, NPP)数据(1980—2000年平均,来源于北京师范大学);5 km×5 km年均氮肥用量(Nitrogen fertilizer, N-fert)栅格数据(1994—2001年平均,来源于美国国家航空航天局(National Aeronautics and Space Administration, NASA)社会经济数据应用中心)。其中遥感NPP主要用于估算农田作物还田碳输入。农作物碳投入主要包括:秸秆还田、作物根茬残留和作物地下生物量。根据周睿等^[15]区域农田有机物质的估算方法利用NPP数据进行农田碳投入的碳估算:

$$B_{\text{total}} = \alpha \times \text{NPP}$$

式中, B_{total} 为作物整株生物量(包括地上和地下),单位为 $\text{g} \cdot \text{m}^{-2}$ (生物量按C计算);NPP为净初级生产力; α 为碳-干物质系数。由作物的根冠比、收获指数、秸秆还田比例计算农田碳投入。

对于土壤SOM、TN、Ap、Ak、pH和Clay等点位测定数据,采用普通克里格方法预测提取300 m×300 m栅格数据,插值后的栅格数据作为SOM空间预测的辅助因子。对于30 m分辨率DEM计算得到的Elev、Slope、Planc、ProfC和CTI等辅助预测因子,采用邻域平均法将尺度降至300 m×300 m栅格,以消减土壤样点可能存在的GPS定位误差对空间预测的影响。而对于空间分辨率相对较低的ST、SM、和N-fert等辅助预测数据则采用普通克里格方法插值将尺度升至300 m×300 m栅格作为SOM空间预测的辅助因子。

1.3 SOM空间预测辅助因子的筛选

首先,采用皮尔森相关分析对SOM空间预测的辅助因子进行初步筛选。表1相关分析的结果表

明, Clay、SOM1980、TN、Ak、pH、SD、CT、Elev、ProfC、Mat、Map、InputC、SM、ST、

N-fert共15个辅助预测因子与SOM含量显著相关, 可潜在地用于SOM含量空间分布预测。

表1 土壤有机质含量与辅助预测因子的相关系数

Table 1 Pearson correlation coefficients between soil organic matter content and auxiliary factors

| 初始有机质SOM1980 | 全氮 | 有效磷 | 速效钾 | pH | 黏粒 | 水稻土 |
|--------------|-----------|--------|---------|--------|---------|-------|
| | TN | Ap | Ak | | Clay | SD |
| 0.54** | 0.75** | -0.10 | 0.15* | -0.18* | 0.28** | 0.17* |
| 潮土 | 土壤温度 | 土壤湿度 | 年均温 | 年均降雨量 | 高程 | 坡度 |
| CT | ST | SM | Mat | Map | Elev | Slope |
| -0.22** | 0.44** | 0.41** | 0.28** | 0.18** | -0.24** | 0.07 |
| 平面曲率 | 剖面曲率ProfC | 地形复合指数 | 根茬碳投 | 氮肥施用量 | | |
| Planc | | CTI | InputC | N-fert | | |
| -0.13 | 0.16* | 0.10 | -0.30** | 0.35** | | |

注: *表示相关性显著 ($P<0.05$); **表示相关性极显著 ($P<0.01$) Note: * denotes significant correlation ($P<0.05$); **denotes extremely significant correlation ($P<0.01$)

其次, 对于皮尔森相关分析初步筛选的15个辅助预测因子, 进一步采用逐步回归分析进行筛选, 以便避免因子间的共线性效应。同时, 为探讨辅助预测因子与SOM相关性强弱对不同空间预测方法预测精度的潜在影响, 在逐步回归分析过程中, 考虑了保留相关性最强的辅助预测因子TN和移除TN两种情况以便进行对比分析。逐步回归的筛选主要是以最小信息准则 (AIC) 为衡量标准, 通过选择最小的AIC信息统计量, 来达到

删除变量的目的, 直到当去掉某一因子时AIC均增加, 逐步回归分析终止, 达到最优模型。由此得到, 保留相关性最强的辅助预测因子TN的条件下, 逐步回归筛选得到的SOM辅助预测因子分别为: TN、SOM1980、SD、Mat、SM、Clay; 移除与SOM相关性最强的TN预测因子的条件下, 逐步回归保留的辅助预测因子包括SD、Elev、Mat、InputC、Map、ST和SOM1980共七个因子 (表2)。

表2 土壤有机质 (SOM) 预测辅助因子的逐步回归分析

Table 2 Stepwise regression analysis of auxiliary factors for prediction of soil organic matter (SOM)

| 去除相关性强辅助因子 | | | 保留相关性强辅助因子 | | |
|---|--------------------|--------|---|--------------------|--------|
| Removal of highly related auxiliary factors | | | Retention of highly related auxiliary factors | | |
| 模型保留因子 | 因子系数 | AIC | 模型保留因子 | 因子系数 | AIC |
| Model reserved factors | Factor coefficient | | Model reserved factors | Factor coefficient | AIC |
| 年均温Mat | -6.08 | 766.46 | 年均温Mat | -3.02 | 666.33 |
| 高程Elev | -0.14 | 767.05 | 水稻土SD | 1.74 | 666.47 |
| 土壤温度ST | 5.69 | 767.25 | 黏粒Clay | 0.10 | 667.17 |
| 年均降雨量Map | 0.04 | 767.64 | 初始有机质SOM1980 | 0.11 | 667.35 |
| 水稻土SD | 3.12 | 769.45 | 土壤湿度SM | 2.00 | 668.18 |
| 根茬碳投入InputC | -0.02 | 770.12 | 全氮TN | 12.78 | 771.02 |
| 初始有机质SOM1980 | 0.39 | 797.60 | - | - | - |

1.4 SOM的空间预测方法

普通克里格 (Ordinary kriging, OK) : 基于土壤属性的空间自相关性, 通过邻近的相关观测点的权重均值来预测采样点位置的土壤属性^[16]。本文通过SOM含量半方差函数及拟合参数, 采用OK方法预测SOM空间分布。

回归克里格 (Regression kriging, RK) : 首先基于逐步回归分析保留的辅助预测因子, 采用多元线性回归 (MLR) 对SOM进行预测, 然后对多元回归预测的SOM残差进行OK预测, 最后将回归预测值与残差的OK预测值相加得到SOM的预测值^[8]。

随机森林 (Random forest, RF) : 首先基于逐步回归筛选的预测因子和SOM观测数据, 对随机森林模型中的关键参数每次分裂向量数 (mtry) 进行参数优化, 然后采用经参数优化后的RF模型对SOM空间分布进行预测。参数优化过程中, 分裂节点随机向量数量ntree设定为1 000, 以预测误差最小化为目标函数, 通过遍历比较确定mtry的最优值^[17] (根据预测因子的个数设置mtry参数为1~7, 并进行建模), 以运行100次的平均RMSE结果寻找最优参数。结果表明, 当保留与SOM相关性最强的预测因子TN时, mtry最优值为4, 而移除相关性最强的辅助因子TN时, mtry最优值为1。

1.5 空间预测精度分析

SOM的空间预测精度采用独立验证进行评价。为了使模型每次验证的结果覆盖到整个样点数据集, 将样点数据集分成预测集和验证集进行三折独立验证 (即将样点数据随机分成三份, 分三次取其中两份为预测数据集, 一份为验证数据集), 循环执行三折独立验证100次, 通过比较100次得到的验证数据点位置上SOM均方根误差进行空间预测精度分析。其中, 均方根误差 (RMSE) 的计算公式为:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (z_i^* - z_i)^2}$$

式中, z_i^* 为第*i*个样点的预测值; z_i 为第*i*个样点的实际值; *n*为样点个数。RMSE的结果越小表明预测精度越高。

本研究中辅助数据的筛选、尺度转换等数据处理均采用R软件进行分析, 其中OK、RK和RF及空间预测精度分析采用R语言编程计算, SOM空间分

布采用Arcmap10.2软件进行制图。

2 结 果

2.1 SOM含量的描述性统计特征

研究区413个样点的描述统计分析 (图2) 结果表明, 农田土壤耕层SOM平均含量为29.29 g·kg⁻¹, 中值与均值较为接近, 为28.79 g·kg⁻¹, 高于江苏北部的黄淮平原和东部滨海平原, 略高于江淮平原地区^[18]; 标准差为7.18 g·kg⁻¹, 变异系数 (CV) 为25%, 变异程度中等; SOM含量介于7.75~58.31 g·kg⁻¹之间, 极差为50 g·kg⁻¹。SOM含量的频率分布直方图表现为稍向均值右侧偏离, 偏度系数为0.35, 峰度值为0.75, 表现为近似正态分布。

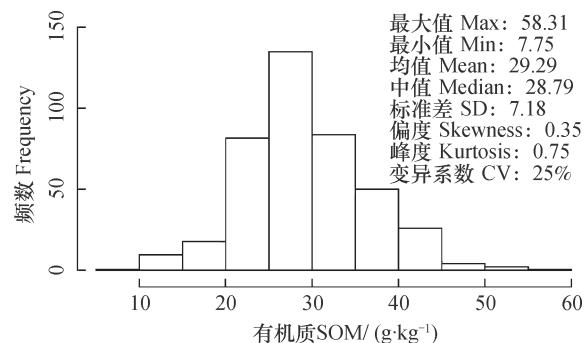


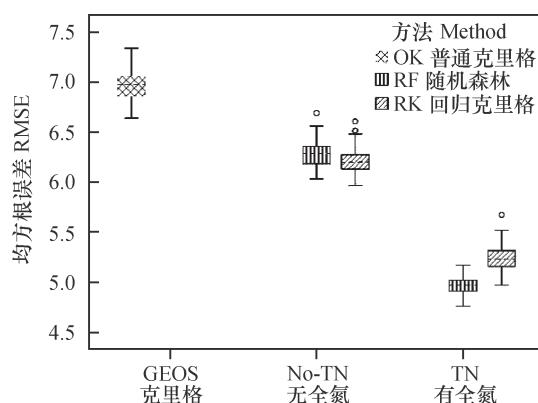
图2 土壤有机质含量的描述统计及频数分布直方图

Fig. 2 Descriptive statistics and histogram of soil organic matter (SOM) contents

2.2 不同空间预测方法的预测精度

从不同空间预测方法100次三折独立验证的结果 (图3) 来看。OK方法的预测误差最大, 空间预测的均方根误差 (RMSE) 均值为6.97 g·kg⁻¹, 仅能解释SOM方差的9.7%; 当整合与SOM相关性最强的辅助预测因子TN后, RK和RF预测方法的RMSE均值分别降低至5.25 g·kg⁻¹和4.97 g·kg⁻¹, 分别降低了25%和29%, 同时, RK和RF方法预测结果对SOM方差的解释程度分别提高到了51%和55%, RF方法均方根误差的四分位距亦较RK方法小, 表明整合与SOM相关性最强的辅助预测因子TN时, RF方法预测的稳健性更好。然而, 当不整合与SOM相关性最强的辅助预测因子TN时, 进入空间预测过程的其余辅助预测因子与SOM的相关性减弱 (表1、表2), RK和RF方法预测的RMSE均值也分别增加至6.21 g·kg⁻¹和6.29 g·kg⁻¹, 相应

的SOM方差解释程度亦分别降低至29%和28%，同时，RK方法均方根误差的四分位距也较RF方法稍小，表明在不整合与SOM相关性较强的辅助因子TN时，RK方法预测的精度及稳健性较RF方法稍好。此外，尽管不整合与SOM相关性最强的辅助预测因子TN时，SOM的预测精度要较整合辅助预测因子TN时低，但其预测精度依然较OK方法高。



注：GEOS表示克里格方法；TN代表考虑全氮因子；No-TN代表不考虑全氮因子 Note: GEOS stands for kriging method; TN for the factor of total TN; No-TN for exclusion of the factor of TN

图3 不同空间预测方法的预测误差对比

Fig. 3 Comparison between the spatial prediction methods in deviation

2.3 SOM的空间分布特征

从不同方法预测的SOM含量空间分布图来看（图5），SOM含量空间分布的总体趋势大体一致，主要表现为研究区西部的丹阳-金坛-溧阳一线含量相对较低，SOM含量变化范围集中在 $15 \sim 35 \text{ g} \cdot \text{kg}^{-1}$ ，而研究区东部到东南部的常熟-昆山-吴江一线SOM含量相对较高，大多数为 $25 \sim 45 \text{ g} \cdot \text{kg}^{-1}$ 。OK方法预测的SOM含量空间分布模式表现为具有较多的、局部分布的岛状高值区和低值区，且岛状区域分布较为零散。与OK方法相比，RK和RF两种方法预测的SOM空间分布的低值区连续性稍好，同时，昆山南部和吴江南部的SOM高值区也得到了更好的表达。然而，考虑到是否整合与SOM相关性最强的辅助预测因子时，RK和RF两种方法预测的SOM空间分布也存在一定的差异性。当在空间预测过程中整合了与SOM相关性最强的辅助预测因子（比如TN）时，除了RF方法能够进一步表达太仓东部地区的SOM低值区外，RK和RF两种方法预测的SOM空间分布基本完全一致；当

移除与SOM相关性最强的辅助预测因子TN后，其余辅助预测因子与SOM的相关性较弱，此时，与RK方法预测的SOM空间分布模式相比，RF预测结果中，研究区西部丹阳-金坛-溧阳一线的SOM含量低值区范围变大，同时研究区东部的SOM高值区的范围亦明显增加。

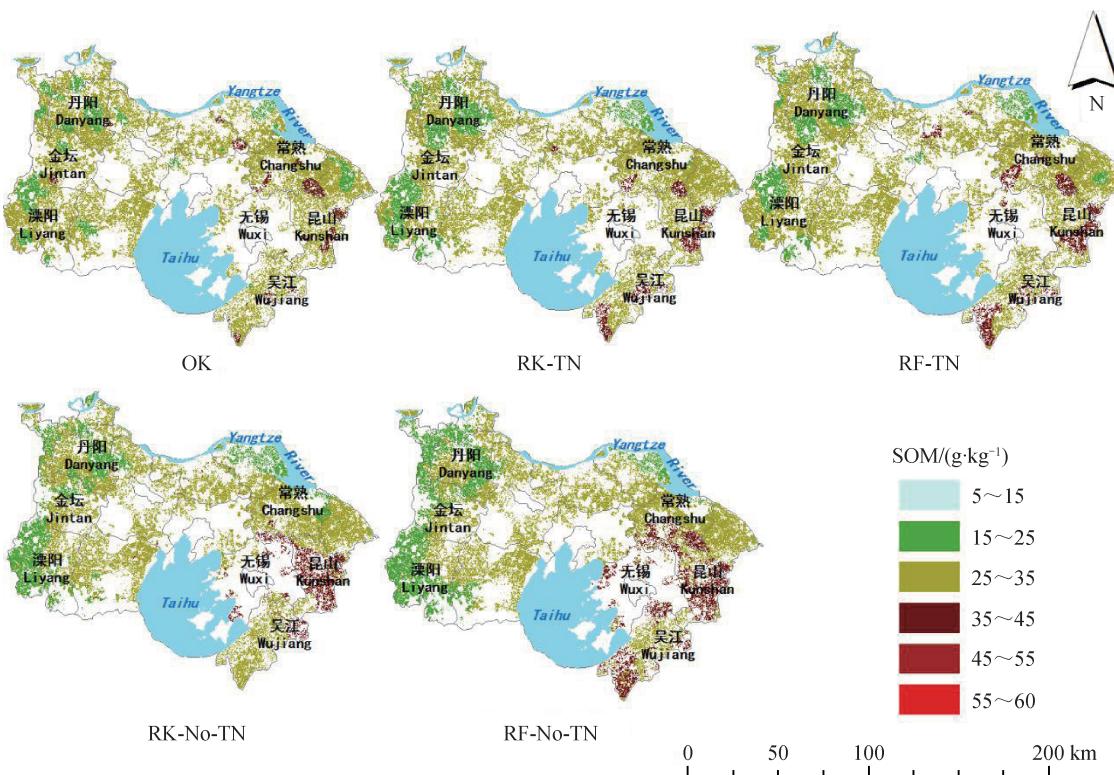
3 讨 论

3.1 不同方法的预测精度

在样点数量及样点空间位置不变的条件下，OK方法的空间预测精度在很大程度上取决于待预测目标土壤属性的空间变异结构^[19]。从SOM含量半方差函数的拟合结果来看（表3），研究区SOM含量的空间相关性较弱，其中块金值与基台值的比值高达83%，表明耕作、施肥等随机因素的强烈影响削弱了SOM的空间自相关性^[20]，可能是导致OK方法预测精度低的主要原因之一。

整合与SOM具有一定相关性的辅助预测因子后，RK和RF方法均能在一定程度上提高SOM的空间预测精度。Knotters^[21]和Ahmed^[22]研究结果表明，在插值方法中使用辅助因子有两个重要特性需要考虑，首先是辅助因子和目标变量之间的相关性强弱，其次是预测的残差是否具有一定的空间结构性。当整合与SOM相关性较强的辅助预测因子（比如TN）后，RF的预测精度显著提高，表明RF方法预测精度主要受目标变量与辅助预测因子间的相关性强弱所控制。而对于RK方法而言，整合高相关性的辅助预测因子的预测精度也较不整合时要高，但由于整合高相关性的辅助预测因子后分离出的残差缺乏空间相关性而导致RK方法的预测精度低于RF方法。

在不整合高度相关的辅助预测因子TN的条件下，RK方法预测精度仍然较RF方法稍高，表明RK方法的预测精度除了与目标变量（比如本研究中的SOM）和辅助因子间的相关性强弱有关外^[23]，还在很大程度上受目标变量多元回归（MLR）预测残差的空间变异结构影响^[24]。比如，从整合与不整合辅助预测因子TN的MLR残差的半方差函数对比来看（表3），在辅助数据与SOM相关性较强的条件下（整合TN后），由于SOM的大部分趋势组分被MLR所分离，导致SOM的残差半方差函数基



注: OK: 普通克里格; RK-TN: 考虑TN因子的回归克里格; RF-TN: 考虑TN因子的随机森林; RK-No-TN: 移除TN因子的回归克里格; RF-No-TN: 移除TN因子的随机森林 Note: OK: Ordinary kriging; RK-TN: Regression kriging including TN factor; RF-TN: Random forest including TN factor; RK-No-TN: Regression kriging excluding TN factor; RF-No-TN: Random forest excluding TN factor

图4 不同方法预测SOM空间分布

Fig. 4 SOM spatial distribution maps based on prediction using different prediction methods

表3 土壤有机质含量的半方差函数模型及拟合参数

Table 3 Semi-variogram model and fitting parameters of soil organic matter content

| 变量 Variables | 模型 Models | 块金值 Nuggets | 基台值 Sill | 块金值/基台值 Nugget/Sill/% | 变程 Ranges | 决定系数 R^2 | 残差平方和 RSS |
|------------------------|-----------------|----------------|-------------|--------------------------|--------------|---------------|--------------|
| 普通克里格 OK | 球状 Spherical | 35 | 42 | 83 | 22 000 | 0.17 | 148 |
| 保留全氮的回归克里格 RK-TN | 线性 Linear | 27 | 27 | 100 | - | 0.09 | 59 |
| 去除全氮的回归克里格 RK-No-TN | 球状 Spherical | 29 | 37 | 78 | 12 000 | 0.34 | 50 |

本为纯块金效应, 即趋势性组分预测精度高而残差预测精度降低; 而在辅助数据与SOM的相关性较弱的条件下(不整合TN), 尽管辅助数据MLR分离的SOM趋势组分相对较少, 但SOM残差的半方差函数相比于整合TN条件下依然具有一定的结构性, 从而残差的预测精度能够有所提升。从数据来源分析, 由于土壤全氮含量获取成本较高, 而移除全氮因子后回归模型中保留的气象、DEM、土壤

图及初始SOM含量等辅助预测数据(表1)则相对较为容易获取, 因此, 从这一点来看, 在辅助预测数据与SOM有相关性, 但相关性相对较弱的条件下, 传统的RK方法在本研究区SOM空间预测上仍然具有较高的应用价值。

3.2 不同方法预测结果局限性

本研究区地势平坦, DEM获取的地形因子在SOM预测应用中还有一定的局限性, 同时, 本研

究区土壤受人为活动影响强烈, 目前容易获取、相对廉价且可能用于辅助SOM空间预测的数据在提高本研究区SOM空间预测精度方面还面临诸多挑战。比如, 相对于本研究区的面积(2.7万km²)而言, 尽管土壤湿度、温度、氮肥施用量等辅助数据比较容易获取, 但依然存在空间分辨率相对较低, 难以和SOM观测点数据进行尺度匹配等问题。

Li^[25]利用RK方法预测SOM空间分布的研究结果表明, 目标变量和辅助因子的空间尺度匹配性直接影响SOM的空间预测精度。因此, 如何对现有容易获取且廉价但相对粗糙的辅助数据进行尺度转换以便更好地识别辅助数据与SOM间潜在的相关性是未来需要进一步研究的问题。此外, 对本研究区而言, SOM除了受土壤类型等环境因子影响外, SOM含量特别是表层SOM含量还受施肥、耕作、秸秆管理等农业管理措施的强烈影响^[26]。作物残茬的C输入是影响农田土壤SOM含量的重要因素之一^[27], 整合土壤C输入来预测SOM的空间分布不但能在一定程度上间接地反映农业管理措施等人为活动对SOM空间分布区域差异性的潜在影响, 也可能在一定程度上提高SOM空间预测精度。然而, 从本研究区C输入与SOM间的相关性分析来看, 其直接的相关性还相对较弱。这不但与NOAA AVHRR遥感数据估算的NPP空间分辨率相对较低且存在混合像元等问题有关, 也与研究区的农田土壤C输入来源相对复杂(比如, 除了根茬C输入外还有有机肥等C输入等)有关。因此, 在平原地区人为活动因子影响强烈的条件下, 寻找与SOM相关性强的、容易获取且廉价的其他辅助预测替代因子, 或者开发适宜的辅助数据降尺度方以实现辅助数据与SOM实测点数据的尺度匹配, 将是实现平原地区强烈人为作用影响下SOM空间有效预测的重要途径之一。

4 结 论

OK、RK和RF三种方法对苏南地区农田表层SOM含量空间预测结果的对比分析表明: (1) 三种方法预测的SOM空间分布总体趋势大体一致, 总体表现为东高西低, 但SOM空间分布的局部细节存在一定的差异性。从预测精度来看, OK方法预测的平均RMSE为6.97 g·kg⁻¹, RK和RF在保留

相关性最强的辅助因子TN时平均RMSE分别为5.25 g·kg⁻¹和4.97 g·kg⁻¹, 而在移除相关性最强的辅助因子TN后, 平均RMSE分别为增加至6.21 g·kg⁻¹和6.29 g·kg⁻¹, 因此, RK和RF预测精度均较OK要高, 且RK和RF均随辅助因子与SOM的相关性变强而预测精度增加。(2) RK和RF的预测精度存在一定的差异, 表现为保留相关性最强的TN时, RF的预测精度较RK要高, 但移除TN后, RK较RF精度稍高。因此, 考虑到TN的获取成本及移除TN后进行预测模型的辅助因子可获取性问题, RK方法在辅助因子与SOM相关性较弱的条件下对本研究区依然有较高的应用价值。(3) RK和RF的预测精度均与辅助因子和SOM间的相关性强弱有关, 但目前能够通过公开途径免费获取的相关辅助数据对于本研究区而言还存在分辨率相对较低等问题, 因此, 如何寻找与SOM相关性强的, 容易获取且廉价的其他辅助因子预测替代因子, 或者开发适宜的辅助因子降尺度方法以实现辅助因子与SOM实测数据的尺度匹配, 将是实现平原地区强烈人为作用影响下SOM空间预测的重要挑战。

参 考 文 献

- [1] 黄昌勇, 徐建明. 土壤学. 第3版, 北京: 中国农业出版社, 2000: 1—380
Huang C Y, Xu J M. Pedology (In Chinese). 3rd ed. Beijing: China Agriculture Press, 2000: 1—380
- [2] 张法升, 曲威, 尹光华, 等. 基于多光谱遥感影像的表层土壤有机质空间格局反演. 应用生态学报, 2010, 21 (4): 883—888
Zhang F S, Qu W, Yin G H, et al. Spatial pattern of surface soil organic matter based on remotely sensed multispectral image (In Chinese). Chinese Journal of Applied Ecology, 2010, 21 (4): 883—888
- [3] Jenny H. Factors of soil formation. Soil Science, 1941, 42 (5): 415
- [4] Sauer T J, Cambardella C A, Meek D W. Spatial variation of soil properties relating to vegetation changes. Plant and Soil, 2006, 280: 1—5
- [5] Rama M P, Jiro K, Shinya T. A kriging method of interpolation used to map liquefaction potential over alluvial ground. Engineering Geology, 2013, 152: 26—37
- [6] 徐尚平, 陶澍, 曹军, 等. 内蒙古土壤pH值、黏粒和有机质含量的空间结构特征. 土壤通报, 2001, 32 (4): 145—148

- Xu S P, Tao S, Cao J, et al. Spatial structure pattern of soil pH, clay and organic matter contents in the Inner Mongolia area (In Chinese). *Chinese Journal of Soil Science*, 2001, 32 (4): 145—148
- [7] Mushtaq A W, Wani J A, Bhat M A, et al. Mapping of soil micronutrients in Kashmir agricultural landscape using ordinary kriging and indicator approach. *Journal of the Indian Society of Remote Sensing*, 2013, 41 (2): 319—329
- [8] Heba E, Mohamed A, Adel A A, et al. Spatial variation of soil carbon and nitrogen pools by using ordinary kriging method in an area of north Nile Delta, Egypt. *Catena*, 2014, 113: 70—78
- [9] Hengl T, Heuvelink G B M, Stein A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 2004, 120: 75—93
- [10] Andrew S, Ganesh P, Keith S. Evaluating regression-kriging for mid-infrared spectroscopy prediction of soil properties in western Kenya. *Geoderma Regional*, 2017, 10: 39—47
- [11] Mareike L, Bruno G, Bernd H. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma*, 2012, 170: 70—79
- [12] Dharumaran S, Hegde R, Singh S K. Spatial prediction of major soil properties using random forest techniques- A case study in semi-arid tropics of South India. *Geoderma Regional*, 2017, 10: 154—162
- [13] Zhi J J, Zhang G L, Yang F, et al. Predicting matic epipedons in the northeastern Qinghai-Tibetan Plateau using random forest. *Geoderma Regional*, 2017, 10: 1—10
- [14] 张甘霖, 龚子同. 土壤调查实验室分析方法. 北京: 科学出版社, 2012: 1—254
- Zhang G L, Gong Z T. Soil survey laboratory methods (In Chinese). Beijing: Science Press, 2012: 1—254
- [15] 周睿, 潘贤章, 王昌坤, 等. 区域农田有机物质投入量的估算方法——以陕西长武县为例. *土壤*, 2013, 45 (5): 862—867
- Zhou R, Pan X Z, Wang C K, et al. Estimation method for inputs of organic matter to regional farmland-Taking Changwu County as a case study (In Chinese). *Soils*, 2013, 45 (5): 862—867
- [16] 苏晓燕, 赵永存, 杨浩, 等. 不同采样点数量下土壤有机质含量空间预测方法对比. *地学前缘*, 2011, 18 (6): 34—40
- Su X Y, Zhao Y C, Yang H, et al. A comparison of predictive methods for mapping the spatial distribution of soil organic matter content with different sampling densities (In Chinese). *Earth Science Frontiers*, 2011, 18 (6): 34—40
- [17] 赵北庚. 基于R语言randomForest包的随机森林建模研究. *计算机光盘软件与应用*, 2015 (2): 152—153
- Zhao B G. The research of random forest model based on random Forest package by using R language (In Chinese). *Computer CD Software and Applications*, 2015 (2): 152—153
- [18] 沈雨. 基于模型和GIS的江苏省农田土壤有机碳动态研究. 南京: 南京农业大学, 2003
- Shen Y. Simulation and prediction of soil organic carbon dynamics in Jiangsu Province based on model and GIS techniques (In Chinese). Nanjing: Nanjing Agricultural University, 2003
- [19] 秦耀东. 土壤空间变异研究中的半方差问题. *农业工程学报*, 1998, 14 (4): 42—47
- Qin Y D. Semi-variogram problems in soil spatial variability (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 1998, 14 (4): 42—47
- [20] Hu K L, Li H, Li B G, et al. Spatial and temporal patterns of soil organic matter in the urban-rural transition zone of Beijing. *Geoderma*, 2007, 141: 302—310
- [21] Knotters M, Brus D J, Oude Voshaar J H. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 1995, 67: 227—246
- [22] Ahmed S, De Marsily G. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*, 1987, 23: 1717—1737
- [23] Watt M S, Palmer D J. Use of regression kriging to develop a carbon: Nitrogen ratio surface for New Zealand. *Geoderma*, 2012, 183: 49—57
- [24] 赵永存, 史学正, 于东升, 等. 不同方法预测河北省土壤有机碳密度空间分布特征的研究. *土壤学报*, 2005, 42 (3): 379—385
- Zhao Y C, Shi X Z, Yu D S, et al. Different methods for prediction of spatial patterns of soil organic carbon density in Hebei Province, China (In Chinese). *Acta Pedologica Sinica*, 2005, 42 (3): 379—385
- [25] Li Y. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information. *Geoderma*, 2010, 159: 63—75
- [26] 杨帆, 徐洋, 崔勇, 等. 近30年中国农田耕层土壤有

机质含量变化.土壤学报, 2017, 54 (5) : 1047—1056

Yang F, Xu Y, Cui Y, et al. Variation of soil organic matter content in cropland of China over the last three decades (In Chinese). *Acta Pedologica Sinica*, 2017, 54 (5) : 1047—1056

[27] Anyanzwa H, Okalebo J R, Othieno C O, et al. Effects of conservation tillage, crop residue and cropping systems on changes in soil organic matter and maize-legume production: A case study in Teso District. *Nutrient Cycling in Agroecosystems*, 2010, 88: 39—47

Comparison Analysis of Methods for Prediction of Spatial Distribution of Soil Organic Matter Contents in Farmlands South Jiangsu, China

XIE Enze^{1,2} ZHAO Yongcun^{1,2†} LU Fangyi^{1,2} SHI Xuezheng^{1,2} YU Dongsheng^{1,2}

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

(2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract 【Objective】Soil organic matter (SOM) plays a key role in maintaining soil quality and functions and crop productively and a critical role in global C recycling. Therefore, knowledge and understandings of spatial distribution patterns of SOM are important to sustainable utilization of soil resources, guarantee of food security, and mitigation of the momentum of global climate change. 【Method】A total of 413 cropland topsoil samples (0~20 cm) were collected from the southern parts of Jiangsu Province for analysis of SOM contents, and ordinary kriging (OK), regression kriging (RK), and random forest (RF) methods were employed for mapping spatial distribution of SOM contents. Auxiliary data such as soil type, topographic factors derived from DEM, climate, soil temperature and moisture, soil properties, and agricultural management practices (N fertilization rate and C input by crop residues) varying in spatial resolution were first scaled into grids 300 m in resolution using either the kriging interpolation or neighborhood averaging method, and then the auxiliary factors screened by a stepwise-regression process were used in RK and RF predictions of SOM. Moreover, to identify impacts of the correlations between auxiliary factors and SOM content on spatial prediction of SOM contents in accuracy, the root mean square errors (RMSE) derived by RK and RF methods were also compared between the situations of removing and retaining the auxiliary factor with highest correlation coefficient. 【Result】The SOM spatial distribution patterns derived with the OK, RK and RF methods were quite similar, that is to say, SOM contents in the eastern parts of the study area are relatively high, whereas those in the western parts are low. But local differences did exist in detail of SOM distribution prediction between the methods can be intuitively observed. The OK method was the lowest in prediction accuracy, with mean RMSE being 6.97 g·kg⁻¹ and lower than the RK and RF methods, of which the mean RMSE of the RK and RF methods was lowered down to 5.25 and 4.97 g·kg⁻¹, respectively, when total nitrogen (TN) of the auxiliary factors that were most closely related to SOM was integrated. However, when TN of these auxiliary factors was removed, the RMSE predicted with RK and RF was lower than that with OK, being 6.21 and 6.29 g·kg⁻¹, respectively, while the average explained variance was decreased to 29% and 28%, respectively. However, the RK and RF methods are still better than the OK method, as the RMSE derived by OK was as high as 6.97 g·kg⁻¹, and the explained variance of OK was only 9.7%. 【Conclusion】RK and RF are both higher than OK in prediction accuracy, however, the difference in prediction accuracy between RK and RF depends

on degree of the correlation between the auxiliary factors and SOM. When auxiliary factors most closely related to SOM, such as TN was included in the prediction, RF was better than RK; while those were excluded, RK was slightly better than RF, indicating that RK is still promising due to the relatively high-cost of TN measurement. In addition, prediction accuracy of RF largely depends on degree of the correlation between the auxiliary data and SOM, when TN was removed from the RF prediction, the predicted RMSE increased significantly, indicating that the current easily attainable and available low-cost auxiliary are still facing many challenges in improving SOM prediction accuracy in plain regions with strong anthropogenic influences. Consequently, developments of new scaling methodology for raw auxiliary data or new higher resolution auxiliary data for quantifying relationships between auxiliary data and SOM are critical for improving accuracy of the prediction of SOM in plain areas with intensifying anthropogenic influences.

Key words Spatial prediction; Kriging; Random forest models; Soil organic matter

(责任编辑: 檀满枝)