

基于辅助变量和神经网络模型的土壤有机质空间分布模拟^{*1}

江叶枫^{1, 2} 郭熙^{1, 2*} 叶英聪¹ 孙凯¹ 饶磊¹ 宋青励³

(1. 江西农业大学江西省鄱阳湖流域农业资源与生态重点实验室/国土资源与环境学院, 江西南昌 330045;

2. 南方粮油作物协同创新中心, 湖南长沙 410000;

3. 东华理工大学马克思主义学院, 江西南昌 330013)

【摘要】:为快速准确获取省域尺度下土壤有机质的空间分布状况。以江西省 2012 年测土配方施肥项目采集的 16582 个耕地表层(0~20cm)土壤样点数据,借助四方位搜索法、地统计学和遥感影像分析技术提取环境因子和邻近信息作为辅助变量,构建基于地理坐标与辅助变量的 BP 神经网络模型和普通克里金法结合的方法(BPNN_OK)、基于地理坐标与辅助变量的 RBF 神经网络模型和普通克里金法结合的方法(RBFNN_OK)和普通克里金法(OK 法)3 种方法,模拟省域尺度下耕地表层(0~20cm)土壤有机质的空间分布。对 2416 个验证样点进行独立验证的研究结果显示:基于辅助变量的神经网络模型较普通克里金法有较大提升。BPNN_OK 法对土壤有机质预测结果的均方根误差、平均绝对误差、平均相对误差较 OK 法分别降低了 2.76g/kg、2.34g/kg、9.83%,RBFNN_OK 法较 OK 法分别降低了 2.70g/kg、2.29g/kg、9.61%。研究显示,基于辅助变量的神经网络模型与 OK 法结合的方法明显地提高了土壤有机质空间分布模拟精度,并且存在改进和提高的空间。

【关键词】:土壤有机质;辅助变量;神经网络模型;空间分布模拟

【中图分类号】:S153.6 **【文献标识码】**:A **【文章编号】**:1004-8227(2017)08-1150-09

DOI:10.11870/cjlyzyyhj201708005

土壤有机质(Soil Organic Matter, SOM)含量是表征土壤肥力的重要指标,也是陆地土壤碳库的重要组成部分,在培育土

¹ 收稿日期:2016-12-15; 修回日期:2017-03-13

基金项目:国家自然科学基金项目(41361049) [National Natural Science Foundation of China(41361049)]; 江西省自然科学基金项目(20122BAB204012) [Natural Science Foundation of Jiangxi Province(20122BAB204012)]; 江西省赣鄱英才“555”领军人才项目(201295) [GanPo “555” Talent Research Funds of Jiangxi Province(201295)]

作者简介:江叶枫(1994~),男,硕士研究生,主要从事土壤环境与系统模拟研究. E-mail:jiangyf0308@163.com

***通讯作者** E-mail:xig435@163.com

壤肥力、调节土壤理化性质以及全球碳循环等方面都凸显出极其重要的作用。由于受生物^[1]、气候^[2]、母质^[3]和地形^[4]等因素的影响,土壤有机质含量在空间上呈现非均匀分布。目前基于实地采样获取的土壤有机质含量信息,很难满足精准农业和区域环境保护的实际需求。因此,如何精确地获取区域土壤有机质含量空间分布信息,是有效管理土壤肥力、精确规划土地利用和指导生态环境建设的需要^[5~7]。

土壤有机质空间分布预测方法有很多,其中普通克里金(Ordinary Kriging, OK)法是应用最普遍也最广泛的方法^[8],但在许多情况下因不满足 OK 法的 3 个重要前提条件而导致精度偏低甚至不能应用,而且 OK 法容易造成平滑效应^[9]。尤其是难以表达土壤属性含量突变的区域,导致突变区域信息丢失严重^[10]。为解决这一问题,近年来国内外众多学者将人工神经网络引入土壤有机质的空间分布预测中,人工神经网络模型能够较准确的揭示输入变量与土壤有机质的复杂非线性关系,并取得了显著成就^[11, 12]。目前应用于土壤有机质空间插值的神经网络主要有 BP 神经网络和 RBF 神经网络。

土壤有机质含量的涵养和运移以及分布是复杂的物理、化学和生物过程,受多种地表环境因素共同作用^[13, 14]。研究表明,引入环境因子作为辅助变量的土壤有机质空间分布预测方法,由于量化的环境因子在不同程度上影响土壤有机质的空间分布^[15],其预测精度较仅基于邻近相关采样点权重而不考虑环境因子的空间内插方法有明显的提高^[16, 17]。还有一些学者研究表明,把邻近信息纳入土壤性质的空间分布预测,能够揭示土壤性质空间分布预测的局部变化特征,能在提高精度的同时更加真实的反映土壤性质的空间变异情况^[18, 19]。因此,在预测方法引入环境因子和邻近信息作为辅助变量是进一步提高土壤有机质空间分布预测精度和揭示土壤有机质空间变异信息的有效途径之一。

本文以江西省耕地表层土壤有机质为研究对象,基于 2012 年测土配方项目的数据,以地理坐标、环境因子与邻近信息作为辅助变量,运用 BP 神经网络模型与 OK 法结合(Back Propagation Neural Network Combined with OK, BPNN_OK)、RBF 神经网络模型与结合 OK 法(Radial Basis Function Neural Network with OK, RBFNN_OK)和 OK 法,模拟江西省耕地表层土壤有机质的空间分布。结果可为省域尺度下土壤有机质的空间插值模型选取和精度优化提供参考。

1 材料和方法

1.1 研究区域概况

江西省地理坐标介于 $N24^{\circ} 29' 14'' \sim 30^{\circ} 04' 41''$, $E113^{\circ} 34' 36'' \sim 118^{\circ} 28' 58''$ 之间,总面积为 $1.67 \times 105 \text{ km}^2$,人口 4566 万,辖 11 个地级行政区、100 个县级行政区、1 个国家级新区。东邻浙江省、福建省,南连广东省,西接湖南省,北毗连湖北省、安徽省,且共接长江。年平均气温 $20 \sim 30^{\circ}\text{C}$,雨量充沛,年均降水量 $1340 \sim 2000 \text{ mm}$;全年无霜期约 $240 \sim 300 \text{ d}$,为亚热带湿润气候。省内东、西、南三面环山,中部丘陵和河谷平原交错分布,北部则为鄱阳湖平原。土地利用类型以耕地和林地为主(图 1(a)),其中耕地面积为 $3.09 \times 106 \text{ hm}^2$,约占全省土地总面积的 23.21%。土壤类型主要有水稻土、红壤、黄褐土、石灰土和潮土等。粮食作物以水稻为主,水稻种植面积约占全省耕地总面积的 80%。植被以常绿阔叶林为主,具有典型的亚热带森林植物群落。

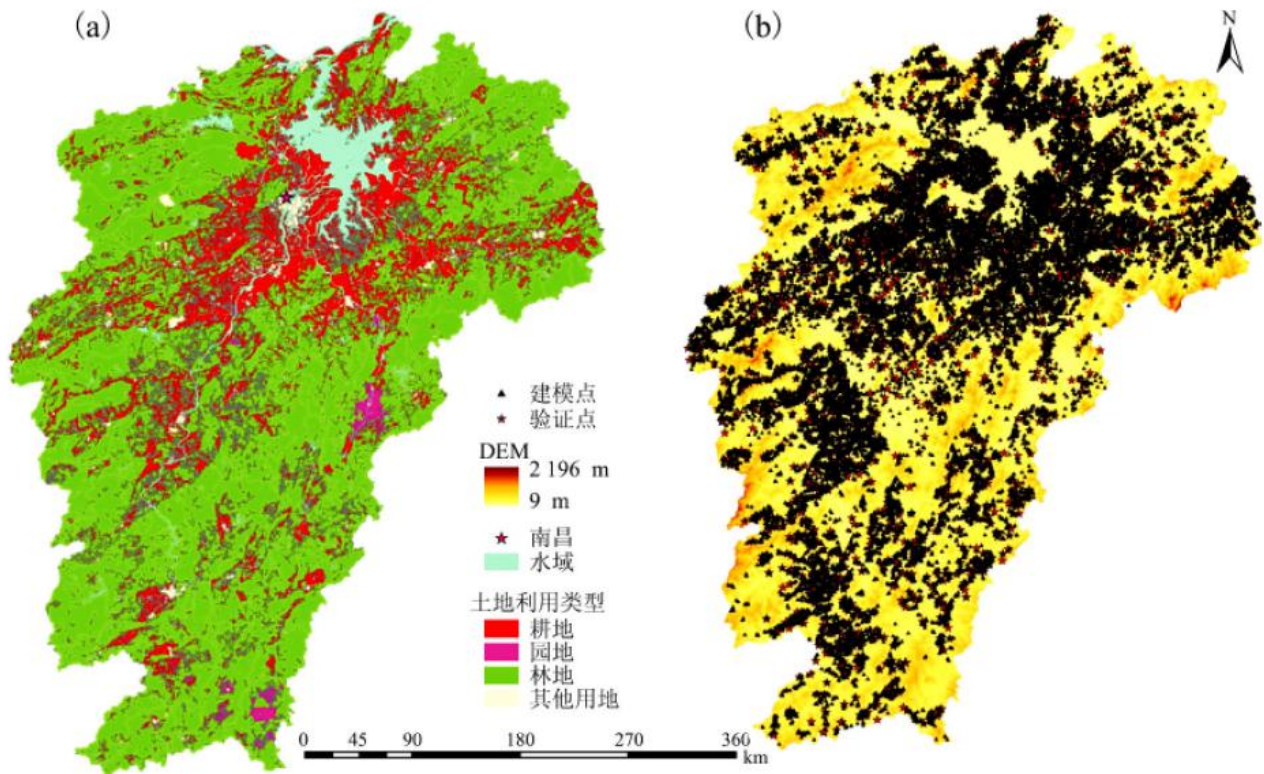


图1 江西省土地利用类型图(a)、土壤样点分布图和DEM数据(b)

Fig. 1 Land-use Types (a), Distribution of Soil Sampling Sites and DEM (b) in Jiangxi Province

1.2 土壤采样与数据处理

土壤样品按照《全国耕地地力调查与质量评价技术规程》采集于2012年11月至2013年2月,采用多点混合的方法采集0~20cm的耕地表层土壤样品16582个。每个采样点均以GPS记录其海拔和坐标,每个样点采集样品1kg,土壤样品经过自然风干后,在实验室磨碎过筛,采用重铬酸钾(K₂Cr₂O₇)油浴加热测定土壤有机质含量。本文采用拉依达准则法(3倍的标准差)对采样点数据的土壤有机质含量进行处理^[20],剔除后总耕地样点为16109个,本文相关内容的研究均采用剔除异常值后的数据。

1.3 辅助变量获取

1.3.1 环境因子的获取

选取高程(H)、坡度(S)、坡向(AS)、曲率(C)、坡度变率(SOS)、坡向变率(SOA)、地形起伏度(QFD)、河流动能指数(Ω)、地形湿度指数(TI)和植被覆盖指数(NDVI)10个因素作为影响土壤有机质空间分布的环境因子。其中植被覆盖指数由Landsat8OLI影像(拍摄日期为2016年12月26日,由各县数据合成,空间分辨率30m)的第四波段和第五波段在ArcGIS10.2中进行栅格计算获取;其他地形因子可通过GIS空间分析从DEM数据获得。各环境因子的计算公式见参考文献^[12~14]。Landsat8OLI影像和DEM数据来源于地理空间数据云(<http://www.gscloud.cn/>),借助ArcGIS10.2软件处理提取。

1.3.2 邻近样点信息的获取

利用四方位搜索法(the four direction search method)获取插值点附近土壤有机质含量值。四方位搜索法是基于反距离

加权算法思想^[21]和空间自相关理论^[22]，根据采样点之间的距离越近影响越大，距离越远影响越小的思想，以采样点为坐标原点，变程 a 为半径，在每个象限内选择一个邻近点土壤属性值 P_1 、 P_2 、 P_3 、 P_4 。四方位搜索算法步骤：

1) 在 MATLAB 中新建一个 Point Record 类型的数组 point [] 文件，把 16109 个采样点地理坐标(x, y)和土壤有机质含量值(z)导入；

2) 从文件中随机抽取一个点，赋值给 Q_i ；

3) 在 point [] 文件剩余点中随机抽取一个点，赋值给 W_i ；

4) 计算 Q_i 、 W_i 两点之间的欧氏距离，设为 d，令 Temp1=Temp2=Temp3=Temp4=a，若 $d > a$ ，则回到步骤 3；

5) 若 $W_i(x) > Q_i(x)$ ， $W_i(y) > Q_i(y)$ 且 $d \leq \text{Temp1}$ ，则 Temp1=d， $P1=W_i(z)$ ；

6) 若 $W_i(x) < Q_i(x)$ ， $W_i(y) > Q_i(y)$ 且 $d \leq \text{Temp2}$ ，则 Temp2=d， $P2=W_i(z)$ ；

7) 若 $W_i(x) < Q_i(x)$ ， $W_i(y) < Q_i(y)$ 且 $d \leq \text{Temp3}$ ，则 Temp3=d， $P3=W_i(z)$ ；

8) 若 $W_i(x) > Q_i(x)$ ， $W_i(y) < Q_i(y)$ 且 $d \leq \text{Temp4}$ ，则 Temp4=d， $P4=W_i(z)$ ；

9) 重复步骤 2)~8)，直至遍历 mat 文件中所存放在文本中。

四方位搜索法在 MATLABR2014a 中通过编程实现，出现的空值由其他象限的平均值进行计算。其中 P_1 、 P_2 、 P_3 、 P_4 分别代表以采样点为坐标原点，4 个象限内离采样点最近点的土壤有机质含量值； Q_i 和 W_i 表示采样点的地理坐标， $Q_i(x)$ 和 $W_i(x)$ 表示经度， $Q_i(y)$ 和 $W_i(y)$ 表示纬度。

1.4 研究方法

1.4.1BP 神经网络模型与 OK 法结合的方法(BPNN_OK)

BPNN_OK:通过 BP 神经网络建立土壤有机质与地理坐标、环境因子以及邻近样点信息的最佳非线性映射关系，得到代表确定性部分的趋势项和代表随机性部分的残差项，然后把趋势项的预测结果导入 ArcGIS 中转成栅格文件，同时利用 OK 法对残差项进行插值，最后将两者的结果进行叠加，其过程可表示为：

$$f_1(x) = g_1(x) + m_1(x)$$

式中： $f_1(x)$ 为 BPNN_OK 在 x 处的插值结果； $g_1(x)$ 为 BPNN 在 x 处的趋势项； $m_1(x)$ 为 BPNN 残差项在 x 处的 OK 插值结果。在 matlab 中调用函数为 newcf 建立一个三层网络，由于 BPNN 对输入样本进行了归一化处理所以隐含层和输出层分别采用 tansig 和 purelin 函数，训练函数使用 trainlm 函数，网络的拓扑结构为 16-22-1、步数为 50、学习率为 0.05，修改权值采用 LMS(Least Mean Square Algorithm)算法。

1.4.2RBF 神经网络模型与 OK 法结合的方法(RBFNN_OK)

RBF 神经网络对非线性连续函数具有一致逼近性和学习速度快等优点，是目前预测研究的热点之一^[19, 20]。RBFNN_OK:通过 RBF 神经网络建立土壤有机质与地理坐标、环境因子以及邻近样点信息的最佳非线性映射关系，得到代表确定性部分的趋势项和代表随机性部分的残差项，然后把趋势项的预测结果导入 ArcGIS 中转成栅格文件，同时利用 OK 法对残差项进行插值，最后将两者的结果进行叠加，其过程可表示为：

$$f_2(x) = g_2(x) + m_2(x)$$

式中： $f_2(x)$ 为 RBFNN_OK 在 x 处的插值结果； $g_2(x)$ 为 RBFNN 在 x 处的趋势项； $m_2(x)$ 为 RBFNN 残差项在 x 处的 OK 插值结果。以误差最小为约束条件，调用 MATLAB 中的 newrb 函数对网络进行训练，设置误差容为 $1e-8$ 、扩散因子为 18、最大神经元个数为 95。

1.4.3 参照方法

本次研究将 OK 法作为参照方法，OK 法基于土壤有机质的区域化特征，通过邻近相关采样点的权重来对未采样点位置的土壤有机质含量进行预测，被广泛应用于土壤有机质的空间插值中^[23, 24]。

1.5 插值精度分析

通过 ArcGIS10.2 中运用地统计模块生成样本数据子集：其中随机均匀选取 85%样点进入建模子集；剩下 15%样点进入验证子集用于验证插值精度。得到建模子集和验证子集。以均方根误差 (RMSE)、平均绝对误差 (MAE)、平均相对误差 (MRE) 对建模子集和验证子集预测值与实际采样值进行对比分析，得出精度评价结果，其公式分别为：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\bar{Z}_i - Z_i)^2}{n}} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\bar{Z}_i - Z_i| \quad (2)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|\bar{Z}_i - Z_i|}{Z_i} \times 100\% \quad (3)$$

$$RI = \frac{R_{RBFNN_OK/ BPNN_OK} - R_{OK}}{R_{OK}} \times 100\% \quad (4)$$

式中： $\bar{Z}_i$ 为样点的预测值； Z_i 为样点的实际观测值； n 为样点数，RMSE，MAE 和 MRE 值越小，说明误差越小、模拟精度越高。 R_{RBFNN_OK} 、 R_{BPNN_OK} 、 R_{OK} 分别是 RBFNN_OK、BPNN_OK、OK 的预测值与实际观测值的相关系数。RI 用来两两对比不同插值方法精度的提高程度。

2 结果与分析

2.1 一般统计分析

表 1 为研究区 16109 个土壤采样数据的统计结果。从表 1 可以看出, 土壤有机质含量的平均值为 30.22g/kg, 值域范围为 12.00~50.00g/kg。从 K-S 检验的 P 值大于 0.05 可以得出土壤有机质含量服从正态分布。根据变异系数的大小江西省土壤有机质的变异系数在 10%~100%之间, 为中等程度的变异性。对建模子集和验证子集进行常规统计分析, 从值域范围、均值与变异系数可以看出, 建模子集与验证子集大体上保留了全部样点的结构特征。同时对 RBFNN 残差与 BPNN 残差进行 K-S 检验, 从 P 值可以看出二者均符合正态分布, 可以进行半变异函数分析和 OK 法插值。

表 1 描述性统计

Tab. 1 Descriptive Statistics Characteristics

项目	样本数	最大值 (g/kg)	最小值 (g/kg)	平均值 (g/kg)	标准差 (g/kg)	偏度系数	K-S 检验	变异系数 (%)
全部样点	16 109	50.00	12.00	30.22	8.88	0.12	0.56	29.38
建模子集	13 693	50.00	12.00	30.25	8.88	0.10	0.42	29.36
验证子集	2 416	50.00	12.20	30.06	8.89	0.16	0.36	29.57
RBFNN 残差	13 693	24.73	-29.11	0.05	5.45	-0.05	0.86	/
BPNN 残差	13 693	25.90	-28.25	-0.02	5.38	-0.09	0.95	/

2.2 土壤有机质与辅助变量相关性分析

表 2 为研究区土壤有机质与环境因子及邻近信息等辅助变量的相关性。从表 2 可以看出, 土壤有机质含量与高程呈显著的负相关, 相关系数为-0.016($P<0.05$), 表明地势越高, 土壤有机质含量越低。土壤有机质含量与植被覆盖指数与地形湿度指数的相关系数分别为 0.002($P<0.01$)和 0.005($P<0.05$), 表明湿度较大、植被覆盖较大的地方有机质含量也相对较高。这与以往研究结果较为一致^[13, 14]。与其他环境因子相关性均不显著, 这可能与本文的研究尺度有关。以往研究都集中在县域尺度且地形较为单一, 环境变化明显, 而省域尺度下环境变化更为复杂, 环境因子与县域尺度相比较模糊。此外, 邻近信息 P1、P2、P3、P4 与土壤有机质均有极显著的相关性($P<0.01$), 相关系数均达到 0.2 以上, 表明在省域尺度土壤有机质存在较强的自相关性, 邻近信息对土壤属性有着非常重要的影响。因此, 在进行土壤有机质空间分布模拟时有必要引入邻近信息。

表 2 土壤有机质与辅助变量的相关系数
Tab. 2 Correlations Between Soil Properties and
Environmental Factors, Neighbor Information

辅助变量	土壤有机质	辅助变量	土壤有机质
植被覆盖指数	0.002 **	地形起伏度	-0.015
高程	-0.016 *	河流动能指数	-0.009
坡度	0.001	地形湿度指数	0.005 *
坡向	0.012	P_1	0.354 **
曲率	-0.014	P_2	0.232 **
坡度变率	-0.010	P_3	0.346 **
坡向变率	0.013	P_4	0.264 **

注：* 表示 $P < 0.05$ ；** 表示 $P < 0.01$ 。

2.3 半变异函数分析

在 GS 中对土壤有机质、RBFNN 残差和 BPNN 残差进行半变异函数的拟合，用半变异函数描述土壤有机质的空间变异性；从表 3 可以看出，土壤有机质、RBFNN 残差和 BPNN 残差的块金效应值(随机性部分引起的空间变异占系统总变异的比例) [25, 26] 均大于 75%，呈弱空间相关性。土壤有机质与 BPNN 残差的最优模型为球状模型；RBFNN 残差的最优模型为指数模型。从模型的参数可以看出，各模型的拟合度较高(大于 0.85)；变程相对较小，这与江西省复杂地形地貌空间结构特征相符合。

表 3 土壤有机质、RBFNN 预测残差和 BPNN 预测残差的半变异函数参数
Tab. 3 Semi-variogram Parameters of Log-transformed SOM and Prediction Residuals by BPNN and RBFNN

土壤养分	数据项	模型	块金值	基台值	块金效应 (%)	变程 (m)	决定系数
有机质	SOM 含量值	球状	55.91	69.99	80.36	18 734.25	0.88
	RBFNN 残差	指数	25.46	33.51	76.25	8 125.63	0.95
	BPNN 残差	球状	24.86	31.87	78.59	8 423.19	0.96

2.4 预测能力对比分析

将 BPNN、RBFNN 和 OK 法对土壤有机质的预测结果与实测值进行相关性分析，结果表明两种神经网络对土壤有机质的预测值与真实值的相关系数要大于 OK 法的预测结果(如表 4)。其中，BPNN 对土壤有机质的建模、验证和全部样点的相对提高高度较 OK 法分别为 32.31%、52.86%、34.11%。RBFNN 较 OK 法相对提高高度分别为 31.98%、51.53%、34.28%。这表明神经网络模型能有效的揭示土壤有机质和辅助变量的非线性映射关系。

表 4 神经网络模型和 OK 法预测结果与实测值的相关性

Tab. 4 Correlations Between Actual Values and Prediction Values by Neural Network Model and OK Method

模型	建模集			验证集			全部样点		
	OK	RBFNN	BPNN	OK	RBFNN	BPNN	OK	RBFNN	BPNN
<i>R</i>	0.616 ^{**}	0.813 ^{**}	0.815 ^{**}	0.524 ^{**}	0.794 ^{**}	0.801 ^{**}	0.598 ^{**}	0.803 ^{**}	0.802 ^{**}
<i>RI</i>	—	31.98%	32.31%	—	51.53%	52.86%	—	34.28%	34.11%

注：* 表示 $P < 0.05$ ；** 表示 $P < 0.01$ 。

2.5 模拟精度评价

模拟精度分析结果表明(见表 5)，引入环境因子与邻近信息作为辅助变量进行预测的神经网络方法(RBFNN_OK 与 BPNN_OK)要明显优于仅基于采样点进行空间内插的 OK 法。从建模集预测效果看，与 OK 法进行比较，RBFNN_OK 法对土壤有机质预测结果的 RMSE、MAE、MRE 分别降低了 2.08g/kg、1.78g/kg、7.33%；BPNN_OK 法分别降低了 2.14g/kg、1.83g/kg、7.51%。从验证集预测效果看，与 OK 法进行比较，RBFNN_OK 法对土壤有机质预测结果的 RMSE、MAE、MRE 分别降低了 2.70g/kg、2.29g/kg、9.61%；BPNN_OK 法分别降低了 2.76g/kg、2.34g/kg、9.83%。从全部采样点预测效果看，与 OK 法进行比较，RBFNN_OK 法对土壤有机质预测结果的 RMSE、MAE、MRE 分别降低了 2.21g/kg、1.88g/kg、7.79%；BPNN_OK 法分别降低了 2.23g/kg、1.89g/kg、8.14%。

表 5 土壤有机质预测方法精度对比

Tab. 5 Precision Compared Indexes of Different Methods for Predicting SOM

	方法	建模集			验证集			全部采样点		
		RMSE	MAE	MRE	RMSE	MAE	MRE	RMSE	MAE	MRE
有机质	OK	6.99	5.61	21.04	7.71	6.22	23.82	7.14	5.73	21.60
	RBFNN_OK	4.91	3.83	13.71	5.01	3.93	14.21	4.93	3.85	13.81
	BPNN_OK	4.85	3.78	13.53	4.95	3.88	13.99	4.91	3.84	13.46

将 3 种方法以相同分辨率在 ArcGIS10.2 中显示(见图 2)，可以看出，3 种方法在土壤有机质空间分布预测中非常相似，说明 RBFNN_OK 和 BPNN_OK 预测土壤有机质的空间分布可信度较高。与 OK 法相比，RBFNN_OK 和 BPNN_OK 法较 OK 法在预测范围、预测误差、相对提高度都有较大的改善。从预测范围看，BPNN_OK 预测范围最大(14.4183~45.4952g/kg)，OK 预测值域范围最小，为 15.8341~43.5949g/kg。从预测误差来看(表 4)，两种神经网络模型的预测误差较 OK 法都有不同程度的降低。从相对提高度来看(表 4)，两种神经网络模型拟合的相关系数较 OK 法有很大的提升。RBFNN_OK 和 BPNN_OK 预测有机质的空间分布克服了普通克里金法仅基于样点自身数据而难以捕捉复杂环境条件下的土壤有机质空间分布的缺点，同时也说明神经网络模型与克里金法结合的方法可以比较准确的预测全省土壤有机质空间分布。

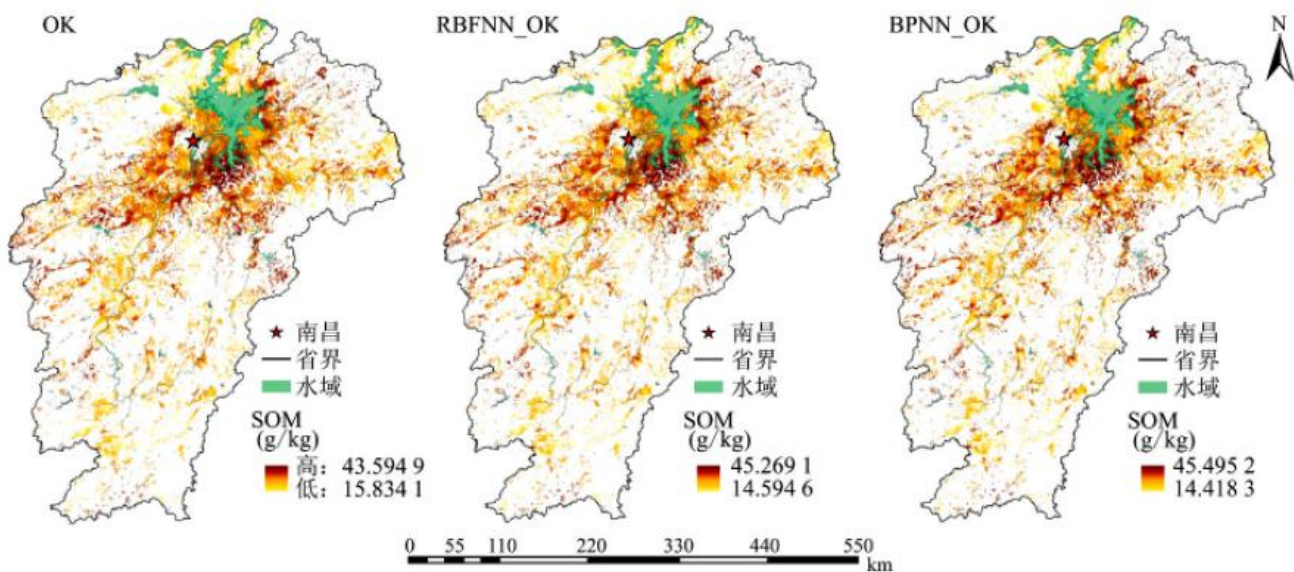


图2 江西省耕层土壤有机质空间分布预测结果

Fig. 2 Soil Organic Matter Map from Farmland Surface in Jiangxi Province Based on Different Method

3 讨论

神经网络作为一种新的空间插值技术，用于土壤有机质的空间异质性研究相对较少。李启权等^[27]根据影响土壤有机质的环境因子和地理坐标，建立环境因子、地理坐标与土壤有机质的神经网络模型，从而获取了研究区土壤有机质的空间分布。徐剑波等^[18]以地理坐标与邻近信息作为网络的输入，探讨神经网络对土壤有效磷的空间分布模拟能力，并与OK法作比较，研究表明神经网络的模拟误差较OK法有明显的降低，这与本研究结果基本一致，这是因为神经网络模型具有极强的非线性映射能力，能以任意精度逼近任意连续函数，因此对研究区的预测精度较为理想。为了更加真实的反映土壤有机质的空间分布情况，本研究通过把地理坐标、环境因子和邻近信息共同作为网络的输入，运用BPNN_OK法和RBFNN_OK法来模拟江西省土壤有机质的空间分布。神经网络模型与OK法结合的方法不仅能够准确地捕捉土壤有机质与辅助变量的复杂关系，而且在残差存在空间自相关的情况下充分利用神经网络残差进行普通克里金法插值，考虑了采样点的结构性和随机性的空间模拟。

在土壤有机质空间分布模拟方法中，OK法是应用最普遍也最具代表性的，能够通过邻近相关采样点土壤属性值权重来预测未知点、对预测结果给出误差等优点，但该方法没有考虑环境因子的影响，同时其应用条件较多、会受到新的样点数据构型的影响^[28]，导致其制图效果和预测精度上还不是理想。虽然有些学者尝试运用环境因子来预测土壤性质的空间分布，但却忽略了邻近信息的重要性，导致其预测精度和对土壤性质的空间异质性描述能力较低。在复杂环境条件下，如果模拟方法不能反映有环境因子与邻近信息引起的土壤有机质空间异质性，必然会导致土壤有机质过高或过低的预测^[27]。因此，引入环境因子与邻近信息作为辅助变量预测的神经网络方法的模拟结果和精度评价结果要优于OK法。

由于本文研究尺度较大，土壤有机质与辅助变量的关系还随着空间区位的变化进而产生差异；而且区域内土壤类型、土属类型、秸秆还田、施肥量和人为因素均会对江西省土壤有机质空间分布产生影响。因此，在预测方法中充分考虑各种因素和它们之间的关系以及各因素对土壤有机质分布的作用机理，才能进一步提高省域尺度下土壤有机质的空间分布模拟效果，使其更加接近实际空间分布。

4 结论

本文以江西省为研究区,在神经网络模型与OK法结合的方法理论框架指导下,构建基于地理坐标与辅助变量的BPNN_OK模型和RBFNN_OK模型对江西省耕地表层土壤有机质空间分布进行模拟。对13693个采样点进行建模,2416个采样点进行独立验证的结果表明:BPNN_OK法对土壤有机质预测结果的均方根误差、平均绝对误差、平均相对误差较OK法分别降低了2.76g/kg、2.34g/kg、9.83%,RBFNN_OK法较OK法分别降低2.70g/kg、2.29g/kg、9.61%。可以得出,神经网络模型结合OK法预测精度明显高于OK法,该方法可为省域尺度下土壤有机质的空间插值模型选取和精度优化提供参考。在今后的研究中,将选取更多的定性因素与定量因素甚至是时间因素来提高对土壤有机质空间插值精度。

参考文献:

- [1] MCBRATNEY A B, PRINGLE M J. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture [J]. Precision Agriculture, 1999, 1(2) : 125—152.
- [2] DAI W, HUANG Y. Relation of soil organic matter concentration to climate and altitude in zonal soils of China [J]. Catena, 2006, 65(1) : 87—94.
- [3] SARKAR S, ROY A K, MARTHA T R. Soil depth estimation through soil-landscape modelling using regression kriging in a Himalayan terrain [J]. International Journal of Geographical Information Science, 2013, 27(12) : 2436—2454.
- [4] 连纲,郭旭东,傅伯杰,等. 黄土高原县域土壤养分空间变异特征及预测——以陕西省横山县为例 [J]. 土壤学报, 2008, 45(4) : 577—584.
- 【LIAN G, GUO X D, FU B J, et al. Spatial variability and prediction of soil nutrients on county scale on the loess plateau:A case study of Hengshan county, Shanxi province [J]. Acta Pedologica Sinica, 2008, 45(4) : 577—584. 】
- [5] MCBRATNEY A B, ODEH I O A, BISHOP T F A, et al. An overview of pedometric techniques for use in soil survey [J]. Geoderma, 2000, 97(3—4) : 293—327.
- [6] SUMFLETH K, DUTTMANN R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators [J]. Ecological Indicators, 2008, 8(5) : 485—501.
- [7] ELBERLING B, GREGORICH E G, HOPKINS D W, et al. Distribution and dynamics of soil organic matter in an Antarctic dry valley [J]. Soil Biology & Biochemistry, 2006, 38 (10) :3095—3106.
- [8] 李晓军,张振远. 基于指示和普通克里金的不连续地层厚度估计方法 [J]. 岩土力学, 2014, 35(10) : 2881—2887.
- 【LI X J, ZHANG Z Y. A combined indicator-ordinary Kriging method for estimating thicknesses of discontinuous geological strata [J]. Rock and Soil Mechanics, 2014, 35(10) : 2881 —2887. 】
- [9] 沈掌泉,施洁斌,王珂,等. 应用集成BP神经网络进行田间土壤空间变异研究 [J]. 农业工程学报, 2004, 20(3) : 35—39.

【SHEN Z Q, SHI J B, WANG K, et al. Spatial variety of soil properties by BP neural network ensemble[J]. Transactions of the CSAE, 2004, 20(3) : 35—39. 】

[10] 董敏, 王昌全, 李冰, 等. 基于 GA RBF 神经网络的土壤有效锌空间插值方法研究[J]. 土壤学报, 2010, 47(1) : 42—50.

【DONG M, WANG C Q, LI B, et al. Study on soil available zinc with GA-RBF-Neural-Network-Based spatial interpolation method [J]. ActaPedologicaSinica, 2010, 47(1) : 42—50. 】

[11] 沈润平, 丁国香, 魏国栓, 等. 基于人工神经网络的土壤有机质含量高光谱反演[J]. 土壤学报, 2009, 46(3) : 391—397.

【SHEN R P, DING G X, WEI G S, et al. Retrieval of soil organic matter content from Hyper-Spectrum based on ANN [J]. ActaPedologicaSinica, 2009, 46(3) : 391—397. 】

[12] 李启权, 王昌全, 岳天祥, 等. 基于 RBF 神经网络的土壤有机质空间变异研究方法 [J]. 农业工程学报, 2010, 26(1) : 87—93.

【LI Q Q, WANG C Q, YUE T X, et al. Method for spatial variety of soil organic matter based on radial basis function neural network [J]. Transactions of the CSAE, 2010, 26(1) : 87—93. 】

[13] 张素梅, 王宗明, 张柏, 等. 利用地形和遥感数据预测土壤养分空间分布 [J]. 农业工程学报, 2010, 26(5) : 188—194.

【ZHANG S M, WANG Z M, ZHANG B, et al. Prediction of spatial distribution of soil nutrients using terrain attributes and remote sensing data [J]. Transactions of the CSAE, 2010, 26(5) : 189—194. 】

[14] 李启权, 王昌全, 岳天祥, 等. 基于神经网络模型的中国表层土壤有机质空间分布模拟方法 [J]. 地球科学进展, 2012, 27(2) : 175—184.

【LI Q Q, WANG C Q, YUE T X, et al. Method for spatial Simulation of topsoil organic matter in China based on a neural network model [J]. Advances in Earth Science, 2012, 27(2) : 175—184. 】

[15] THOMPSON J A, KOLKA R K. Soil Carbon Storage Estimation in a Forested Watershed Using Quantitative Soil-Landscape Modeling [J]. Soil Science Society of America Journal, 2005, 69(4) : 1086—1093.

[16] WANG B W, ZHOU W J, SUM, et al. Regression-kriging of soil organic matter using the environmental variables derived from MODIS and DEM [J]. Agricultural Science & Technology, 2012: 838—842.

[17] LV J, LIU Y, ZHANG Z, et al. Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils [J]. Journal of Hazardous Materials, 2013, 261(13) : 387—397.

[18] 徐剑波, 宋立生, 夏振, 等. 基于 GARBF 神经网络的耕地土壤有效磷空间变异分析 [J]. 农业工程学报, 2012, 28(16) : 158—165.

【XU J B, SONG L S, XIA Z, et al. Spatial variability of available phosphorus for cultivated soil based on GARBF neural network [J]. Transactions of the CSAE, 2012, 28 (16) : 158 —165. 】

[19] 李启权, 王昌全, 岳天祥, 等. 不同输入方式下 RBF 神经网络对土壤性质空间插值的误差分析 [J]. 土壤学报, 2008, 45(2) : 360—365.

【LI Q Q, WANG C Q, YUE T X, et al. Error analysis of soil property spatial interpolation with RBF artificial neural network with different input methods [J]. Acta Pedologica Sinca, 2008, 45(2) : 360—365. 】

[20] 王丹丹, 史学正, 于东升, 等. 东北地区旱地土壤有机碳密度的主控自然因素研究 [J]. 生态环境学报, 2009, 18(3) : 1049—1053.

【WANG D D, SHI X Z, YU D S, et al. Main natural factors controlling soil organic carbon density in the uplands of Northeast China [J]. Ecology and Environmental Sciences, 2009, 18(3) : 1049—1053. 】

[21] CHEN C F, ZHAO N, YUE T X, et al. A generalization of inverse distance weighting method via kernel regression and its application to surface modeling [J]. Arabian Journal of Geosciences, 2014, 8(9) : 1—11.

[22] LIU Z G, LI M C, SUN Y, et al. Study on spatial autocorrelation of urban land price distribution in changzhou city of jiangsuprovince [J]. Chinese Geographical Science, 2006, 16 (2) :160—164.

[23] 黄魏, 韩宗伟, 罗云, 等. 基于地形单元的土壤有机质空间变异研究 [J]. 农业机械学报, 2015, 46 (4) : 162—167.

【HUANG W, HAN Z W, LUO Y, at al. Spatial Distribution of Soil Organic Matter Based on Topographic Unit [J]. Transactions of The Chinese Society of Agricultural Machinery, 2015, 46(4) : 162—167. 】

[24] 李志斌. 基于地统计学方法和 Scorpan 模型的土壤有机质空间模拟研究 [D]. 北京: 中国农业科学院, 2010.

[25] 阎波杰, 潘瑜春, 赵春江. 区域土壤重金属空间变异及合理采样数确定 [J]. 农业工程学报, 2008, 24(S2) : 260—264.

【YAN B J, PAN Y C, ZHAO C J. Spatial variability and reasonable sampling number of regional soil heavy metals [J]. Transactions of the CSAE, 2008, 24(S2) : 260—264. 】

[26] 王晋民, 王俊鹏, 胡月明, 等. 栗钙土农田土壤养分空间变异特性及采样方法研究 [J]. 干旱地区农业研究, 2006, 24(5) : 59—63.

【WANG J M, WANG J P, HU Y M, et al. Studies on spatial variability of soil nutrients and sampling methods in a field of chestnut soil [J]. Agricultural Research in the Arid Areas, 2006, 24(5) : 59—63. 】

[27] 李启权, 王昌全, 张文江, 等. 丘陵区土壤有机质空间分布预测的神经网络方法 [J]. 农业环境科学学报, 2012, 31(12) : 2451–2458.

【LI Q Q, WANG C Q, ZHANG W J, et al. 2012. Predict the spatial distribution of soil organic matter for a hilly region with radial basis function neural network [J]. Journal of Agro-Environment Science, 31(12) : 2451–2458. 】

[28] 王政权. 地统计学及在生态学中的应用 [M]. 北京: 科学出版社, 1999.