

基于辅助变量和 GARBF 神经网络的黄河流域土壤镉空间分布预测^①

张成才¹, 郑文豪¹, 闫亚宁¹, 孙雨田¹, 刘 威², 王永辉³

(1 郑州大学水利与交通学院, 郑州 450001; 2 河南水利与环境职业学院, 郑州 450008; 3 河南省地质研究院, 郑州 450001)

摘 要: 为了准确掌握黄河流域土壤镉的空间分布, 以环境因子和土壤理化因子的不同组合作为辅助变量, 利用遗传算法(GA)优化径向基函数(RBF)神经网络对黄河流域土壤镉的空间分布进行了预测, 并与回归克里格、RBF 神经网络预测精度进行了对比, 探究了土壤理化因子和遗传算法对神经网络模型预测精度的影响。结果表明: ①加入土壤理化因子(有机质含量、pH、CEC)可以提高神经网络模型的预测精度, 基于环境因子和土壤理化因子的 GARBF 神经网络模型均方根误差(RMSE)、平均绝对误差(MAE)、平均相对误差(MRE)较仅基于环境因子的 GARBF 神经网络模型分别减小 0.058 mg/kg、0.033 mg/kg、4.4 个百分点; ②遗传算法可以提高神经网络模型的预测精度, 基于环境因子和土壤理化因子的 GARBF 神经网络模型的 RMSE、MAE、MRE 较基于环境因子和土壤理化数据的 RBF 神经网络模型分别减小 0.009 mg/kg、0.005 mg/kg、0.6 个百分点; ③同时加入环境因子和土壤理化因子并使用遗传算法对神经网络模型进行优化得到的预测结果最优, 基于环境因子和土壤理化因子的 GARBF 神经网络模型能用于黄河流域土壤镉的空间分布预测研究。

关键词: 土壤理化因子; 遗传算法; 神经网络; 辅助变量; 空间插值

中图分类号: X53 **文献标志码:** A

Prediction of Spatial Distribution of Soil Cd in Yellow River Basin Based on Auxiliary Variables and GARBF Neural Network

ZHANG Chengcai¹, ZHENG Wenhao¹, YAN Yaning¹, SUN Yutian¹, LIU Wei², WANG Yonghui³

(1 School of Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China; 2 Henan Vocational College of Water Conservancy and Environment, Zhengzhou 450008, China; 3 Henan Geological Research Institute, Zhengzhou 450001, China)

Abstract: In order to accurately grasp the spatial distribution of soil cadmium in the Yellow River Basin, different combinations of environmental factors and soil physicochemical factors were used as auxiliary variables, and the genetic algorithm (GA) was used to optimize the radial basis function (RBF) neural network to predict the spatial distribution of soil cadmium in the Yellow River Basin, and the prediction accuracy of this model was compared with those of the regression Kriging and the RBF neural network, to investigate the effects of soil physicochemical factors and GA on the prediction accuracy of the neural network. The results showed that: 1) The addition of soil physicochemical factors (organic matter content, pH, CEC) could improve the prediction accuracy of the neural network model. The root mean square error (RMSE), mean absolute error (MAE), and mean relative error (MRE) of the GARBF neural network model based on the environmental factors and soil physicochemical factors were reduced by 0.058 mg/kg, 0.033 mg/kg, and 4.4 percentage points, respectively; 2) GA could improve the prediction accuracy of neural network models, and the RMSE, MAE, and MRE of the GARBF neural network model based on environmental factors and soil physicochemical factors were reduced by 0.009 mg/kg, 0.005 mg/kg, and 0.6 percentage points, respectively, compared with the RBF neural network model based on environmental factors and soil physicochemical factors. 3) The prediction results obtained by adding environmental factors and soil physicochemical factors and optimizing the neural network model using GA were optimal, and the GARBF neural network model based on environmental factors and soil physicochemical factors could be

①基金项目: 中央水专项资金项目(20220086A)和河南省自然科学基金项目(222300420539)资助。

作者简介: 张成才(1964—), 男, 河南郸城人, 博士, 教授, 主要从事地理信息技术及水利遥感研究。E-mail: zhangcc@zzu.edu.cn

used to predict the spatial distribution of soil cadmium in the Yellow River Basin.

Key words: Soil physiochemical factors; Genetic algorithm; Neural networks; Auxiliary variables; Spatial interpolation

镉是 I 类致癌物, 可以通过食物链进入人体, 引发肝脏肾脏疾病^[1-2]。2014 年发布的《全国土壤污染状况调查公报》^[3]显示, 重金属镉的点位超标率高达 7%, 是我国土壤的首要无机污染物。工业“三废”排放和农用化学品大量使用是造成土壤镉污染的主要原因^[4]。黄河流域工、农业城市众多, 研究黄河流域土壤镉污染的空间分布状况对确定流域生态安全状况、实现黄河流域高质量发展具有重要意义。

土壤重金属的分布受到人为因素^[5]、环境因子^[6]及土壤理化性质^[7-9]等的影响。引入相关影响因素作为辅助变量的预测模型, 在不同程度上考虑了影响因素对预测变量空间分布的影响, 其预测精度较基于样点数据进行空间内插的方法有明显的提高^[10]。Li 等^[11]以土壤轮作制度、道路交通等人为因素作为预测模型的辅助变量, 预测成都地区的土壤镉分布, 结果表明, 人为因素的引入提高了预测模型的预测精度。江叶枫和郭熙^[12]基于地形数据、植被指数等环境因子预测稻田土壤砷分布, 结果表明, 环境因子的引入提高了预测模型的预测精度。但人为因素作为预测模型的辅助变量难以定量表示, 环境因子中的地形因素难以刻画地形平坦区域土壤重金属的空间分布, 而土壤理化性质能够克服上述缺点, 是土壤重金属空间分布预测模型的理想辅助变量。周伟等^[13]利用地理探测器探究土壤重金属空间分布的影响因子, 提出土壤理化性质可作为土壤重金属空间预测模型的辅助变量。但前人预测土壤重金属的空间分布多以环境因子、人为因素作为辅助变量, 考虑土壤理化性质的研究较少。

在引入辅助变量的预测方法中, 回归克里格法^[14-15]、多元回归模型^[16-17]以及径向基函数(RBF)神经网络模型^[18-19]等被广泛应用于预测土壤重金属等的空间分布。土壤重金属的含量与其空间位置之间往往存在着高度复杂的非线性映射关系^[20], 回归克里格法是基于线性无偏最优估值理论的克里格方法的拓展, 在插值时会产生平滑效应; 多元回归模型无法体现非线性特征^[21]; RBF 神经网络模型具有通过学习逼近任意非线性映射的能力^[22], 但其不准确的隐含层节点个数会影响逼近精度^[23]。由遗传算法(GA)优化的 RBF 神经网络(GARBF 神经网络)可以自适应地确定 RBF 神经网络隐含层节点个数^[24], 解决了 RBF 神经网络难以确定隐含层节点个数的缺陷, 是预测土壤重金属空间分布的理想方法。

基于以上, 本研究以黄河流域为研究区, 以环境因子和土壤理化因子作为辅助变量, 构建基于辅助变量的 GARBF 神经网络模型, 开展土壤重金属空间分布预测, 以期为黄河流域高精度土壤镉空间分布预测提供方法参考。

1 研究区概况与研究方法

1.1 研究区概况

黄河流域位于 32°N~42°N、96°E~119°E(图 1), 自西向东横跨青藏高原、内蒙古高原、黄土高原、黄淮海平原, 流经青海、四川、甘肃、宁夏、内蒙古、陕西、山西、河南、山东 9 个省(区), 最后于山东省东营市垦利区注入渤海, 流域面积约为 $7.95 \times 10^5 \text{ km}^2$ 。流域主要属于南温带、中温带和高原气候区, 年均气温 9℃, 年降水量 200~650 mm。土地利用类型主要有 6 种, 占比从大到小依次为草地 46.2%、耕地 26.6%、林地 13.6%、城镇 3.1%、水体 1.7%, 其余为荒地。流域的土壤类型从上游到下游依次为栗钙土、黑垆土、褐土和潮土。在黄河流经的 9 省区中, 2019 年底常住总人口 42 180.15 万人, 占全国的 30.05%, 地区生产总值约占全国 25%, 在我国社会经济发展中地位十分重要。

1.2 数据获取

1.2.1 土壤重金属镉数据 本研究所采用的土壤重金属镉数据来自布置在黄河流域的 329 个土壤镉金属采样点(图 1)。在采样过程中, 按照五点梅花采样法收集土壤, 同时为减少人为扰动和表层污染积累的影响, 确保土样中重金属镉含量的稳定性和代表性, 采样深度设置为土壤表层 5~15 cm, 并利用手持 GPS 确定采样点经纬度坐标。采样完成后用聚乙烯密封袋避光保存, 后将土样置于朝南风干室(严防阳光直射土样)进行自然风干, 去掉石块、植物根茎等杂物后, 研磨并过 100 目尼龙筛, 装入干净密封袋备用。

参照 HJ1315—2023《土壤和沉积物 19 种金属元素总量的测定 电感耦合等离子体质谱法》^[25], 土壤镉含量的测定步骤大致为: 称取 0.2 g 土壤放入微波消解罐中, 分别加入 9 mL $\text{HNO}_3(\rho=1.42 \text{ g/mL})$ 、3 mL $\text{HCl}(\rho=1.19 \text{ g/mL})$, 充分混匀、反应平稳后, 加盖拧紧, 7 min 内升温至 120℃并保持 3 min, 消解完成后, 冷却至室温; 之后消解罐中内容物全部转移至坩

竭, 在坩埚中加入 2 mL HF($\rho=1.16$ g/mL), 并将坩埚置于电热板上, $120 \sim 140$ °C 加热至内容物呈不流动的黏稠状, 取下坩埚, 冷却至室温, 加入 1 mL HClO₄ ($\rho=1.67$ g/mL), $160 \sim 180$ °C 继续加热至白烟几乎冒尽, 内容物呈黏稠状; 取下坩埚、稍冷, 滴加少量稀硝酸溶液(1.41%)冲洗坩埚内壁, 温热溶解内容物,

冷却至室温后, 转移至 50 mL 容量瓶中, 待测。同时做空白试验。最后, 利用电感耦合等离子体质谱仪(ICP-MS, Thermo Fisher, USA)进行测定。试验同时, 制定标准溶液, 建立标准曲线。试验分析时均使用符合国家标准的优级纯试剂, 试验用水为不含目标物的超纯水。

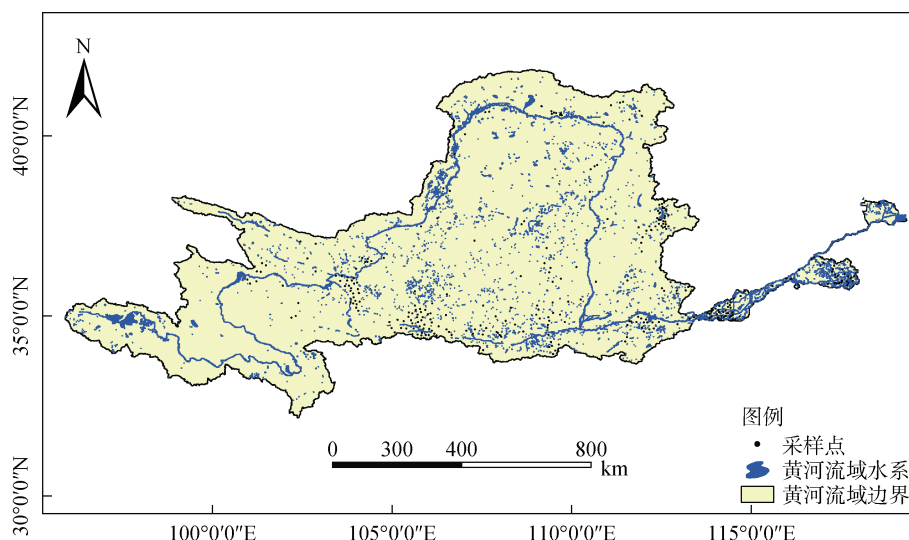


图1 黄河流域土壤采样点分布

Fig. 1 Distribution of soil sampling sites in the Yellow River Basin

1.2.2 土壤理化因子数据 本研究所采用的土壤理化因子数据来源于中国土壤信息格网基本属性数据集(国家地球系统科学数据中心土壤分中心: <http://www.soil.geodata.cn>), 其空间分辨率为 90 m, 包含我国范围内基本土壤物理和化学属性。为确保获得的研究区土壤理化因子数据与实际土样采样点对应, 核对采样点的经纬度、时间信息, 选取数据集 6 个土层深度中的 5 ~ 15 cm 深度数据, 从栅格数据中提取对应采样点的土壤 pH、有机质含量(OM)、阳离子交换量(CEC)、黏粒含量数据集为土壤理化数据。

1.2.3 环境因子数据 基于环境因子预测土壤属性是当前的主流技术, 环境因子以地形因素和植被指数为主。黄河流域存在大面积的平原, 地势平坦, 地形因子对该区域的土壤重金属空间分布影响较小。因此, 本研究仅选取归一化植被指数(NDVI)作为预测土壤镉空间分布的环境因子。NDVI 由 Landsat 8 影像(影像日期为 2022 年 01 月 01 日)的第四、五波段在 ArcGIS 中进行栅格计算获取, Landsat 8 数据下载自地理空间数据云(<http://www.gscloud.cn>)。为了与土壤理化因子数据集的空间分辨率保持一致, 将 NDVI 数据的分辨率重采样至 90 m。

1.3 研究方法

1.3.1 基于辅助变量的 GARBF 神经网络 基于辅助变量的 GARBF 神经网络以辅助变量作为输入, 以土壤镉作为网络输出, 利用遗传算法(GA)对 RBF 神经网络的结构和权值进行优化。遗传算法具有全局逼近的能力, 可避免神经网络陷入局部最优的问题。具体步骤如下: ①二进制编码。隐含层与输出层之间的权值向量构成个体, 对权值向量中的权值进行二进制编码, 构成染色体。②初始化种群。进化以种群为单位, 按照权值向量中各个权值范围, 随机生成一定数量的权值向量, 构成种群。③确定适应度函数。将训练样本点的数据代入由权值矩阵组成的神经网络, 计算个体误差的绝对值(公式(1)), 根据个体绝对误差确定适应度函数(公式(2), 式中, y_i' 表示样本点预测值, y_i 表示样本点实际值, T 表示遗传代数, E 表示绝对误差, $f(x)$ 表示适应度函数)。④选择。用轮盘赌的方法进行个体选择, 避免单纯以适应度高低作为选择依据可能导致的局部收敛问题。⑤交叉和变异。通过交叉和变异使子代获得与上一代不同的基因, 通过步骤④的选择对基因进行筛选。⑥迭代停止。当迭代达到要求的误差范围, 迭代停止, 此时, 对应于染色

体的参数即为神经网络的最优参数。⑦实现空间插值。将研究区划分为 90 m 大小的网格,以网格中心点对应辅助变量作为网络的输入,利用遗传算法优化的神经网络具有的预测功能,得到插值点处土壤镉含量的预测结果。将预测结果连同位置信息保存为 Excel 文件,导入 ArcGIS10.8 中获取插值图。

$$E=\frac{1}{T}\sum_{i=0}^T|y_i-y_i'|$$

(1)

$$f(x)=\frac{1}{E}$$

(2)

基于辅助变量的 GARBF 神经网络计算和建模均以 Matlab 为平台,运用其神经网络工具箱和遗传工具箱编程实现。

1.3.2 对照方法 选用回归克里格和基于辅助变量的 RBF 神经网络作为研究的对照方法,分别探究相同辅助变量下,GARBF 神经网络与前二者的预测精度差异;相同插值方法下,NDVI 和不同土壤理化因子组合作为辅助变量对 RBF 和 GARBF 神经网络精度的影响。回归克里格预测土壤镉时,首先,利用多元逐步回归构建采样点辅助变量与土壤镉含量的回归方程,预测流域内土壤镉的空间分布;其次,对

采样点预测残差进行普通克里格插值;最后,将土壤镉含量预测值和残差值叠加即获得研究区土壤镉含量。基于辅助变量的 RBF 神经网络以采样点辅助变量作为输入,以采样点镉含量作为网络输出,进行神经网络结构的训练,利用训练好的神经网络预测研究区土壤镉分布。

1.3.3 预测精度评价 随机从 329 个土壤样点提取 80%(263 个)作为训练样本,其余 20%(66 个)作为检验样本,通过样本镉含量实测值与预测值之间的均方根误差(RMSE)、平均绝对误差(MAE)和平均相对误差(MRE)3 个指标进行模型精度评价。

2 结果与分析

2.1 土壤镉含量描述性统计特征

研究区 329 个采样点土壤中镉含量的平均值为 0.75 mg/kg,范围为 0.02 ~ 32.69 mg/kg,标准差为 2.77 mg/kg (表 1)。土壤重金属镉的变异系数大于 100%,说明土壤重金属镉的分布受外界因素的影响较大。从统计数据看,训练样本点与检验样本点大体上保留了全部采样点的结构特征。

表 1 土壤镉含量描述性统计
Table 1 Descriptive statistics of soil Cd content

项目	样本数(个)	最大值(mg/kg)	最小值(mg/kg)	平均值(mg/kg)	标准差(mg/kg)	变异系数(%)
全部采样点	329	32.69	0.02	0.75	2.77	369
训练样本点	263	32.69	0.02	0.74	2.75	372
检验样本点	66	28.34	0.01	0.79	2.82	357

2.2 土壤镉含量与其他理化因子的相关性

使用 SPSS19.0 进行土壤镉含量与其他理化因子的 Pearson 相关性分析(表 2),筛选与镉含量在 $P<0.01$ 水平上极显著相关的指标。从表 2 可以看出,镉含量与黏粒含量相关性较弱,与 OM、CEC 呈显著正相关,与 pH 呈显著负相关,其中镉含量与 CEC 的相关性最高,相关系数为 0.56,与 OM、pH 的相关系数分

表 2 土壤镉含量与其他土壤理化性质的相关系数
Table 2 Correlation coefficients between soil Cd contents and other physio-chemical properties

	黏粒含量	pH	OM	CEC	Cd
黏粒含量	1				
pH	-0.31**	1			
OM	-0.11	0.36**	1		
CEC	0.27**	0.15	0.51**	1	
Cd	0.19	-0.46**	0.48**	0.56**	1

注:黏粒是指粒径 <0.002 mm 土壤颗粒,** 表示在 $P<0.01$ 水平极显著相关。

别为 0.46、-0.48。因此,本研究选择 pH、OM、CEC 作为土壤镉含量的解释变量。

2.3 模型变量筛选

多个解释变量的共同作用可以增强或减弱对因变量的解释能力^[26],将解释变量 pH、OM、CEC 进行不同组合,分别和环境因子 NDVI 作为神经网络模型的输入变量,比较不同输入变量下训练样本的逼近误差,筛选最优的土壤理化指标组合,结果如表 3 所示。从表 3 可以看出,仅基于 NDVI 而无土壤理化数据输入时的回归克里格、RBF 神经网络和 GARBF 神经网络模型训练样本逼近误差最大,pH、OM、CEC 的组合与 NDVI 作为回归克里格、RBF 神经网络和 GARBF 神经网络模型输入的训练样本逼近误差最小。因此,选择 pH、OM、CEC 的组合和 NDVI 一起作为神经网络模型的辅助变量。

2.4 模型预测精度

将检测样本点带入回归克里格、RBF 神经网络、

表 3 基于不同输入变量的训练样本误差
Table 3 Errors of the training samples based on different input variables

输入变量	回归克里格			RBF 神经网络			GARBF 神经网络		
	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)
NDVI	0.239	0.169	25.1	0.240	0.171	25.3	0.238	0.167	24.5
pH+NDVI	0.236	0.158	24.4	0.237	0.160	24.7	0.235	0.155	23.9
OM+NDVI	0.234	0.156	24.5	0.235	0.158	24.6	0.232	0.154	23.8
CEC+NDVI	0.232	0.152	24.1	0.234	0.155	24.2	0.229	0.150	23.5
pH+OM+NDVI	0.220	0.149	22.5	0.222	0.150	22.7	0.215	0.148	22.3
pH+CEC+NDVI	0.198	0.144	22.2	0.200	0.148	22.4	0.197	0.144	22.1
OM+CEC+NDVI	0.194	0.141	21.8	0.197	0.142	22.1	0.190	0.140	21.5
pH+OM+CEC+NDVI	0.182	0.135	20.2	0.184	0.138	20.8	0.179	0.132	19.8

GARBF 神经网络模型，通过比较检测样本的误差，检验土壤理化因子和 NDVI 对神经网络模型预测精度的影响。表 4 是不同辅助变量下 3 种预测方法检验样本的预测误差。检验样本误差反映方法的预测精度和可靠性，由检验样本的插值结果可以看出，基于 pH、OM、CEC 和 NDVI 的 GARBF 神经网络模型预测土壤镉的 RMSE 为 0.173 mg/kg、MAE 为 0.131 mg/kg、MRE 为 18.9%，较基于相同辅助变量的 RBF 神经网络模型的 RMSE、MAE、MRE 分别降低 0.009 mg/kg、0.005 mg/kg、0.6 个百分点，说明遗传算法可以提高神经网络模型的预测精度；较基于相同辅助

变量的回归克里格模型的 RMSE、MAE、MRE 分别降低 0.003 mg/kg、0.002 mg/kg、0.2 个百分点，说明基于辅助变量的 GARBF 神经网络模型预测精度高于回归克里格模型。由表 4 还可知，在预测方法相同的情况下，以 NDVI 和 pH、OM、CEC 组合作为辅助变量的插值方法精度较高，以 NDVI 作为辅助变量的插值方法精度较低；在辅助变量相同的情况下，基于辅助变量的 GARBF 神经网络插值方法精度最高。上述结果表明，在 NDVI 的基础上，添加土壤理化因子作为辅助变量，可以提高神经网络模型的预测精度，遗传算法改进的神经网络方法(GARBF)具有更好的插值结果。

表 4 基于最佳输入变量的检验样本误差
Table 4 Errors of the testing samples based on best input variables

输入变量	回归克里格			RBF 神经网络			GARBF 神经网络		
	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)	RMSE (mg/kg)	MAE (mg/kg)	MRE (%)
NDVI	0.234	0.169	23.5	0.238	0.175	23.8	0.231	0.164	23.3
pH+OM+CEC+NDVI	0.176	0.133	19.1	0.182	0.136	19.5	0.173	0.131	18.9

2.5 模型预测结果

将最优变量组合的预测结果以相同分辨率的栅格格式在 ArcGIS10.8 中进行显示，如图 2 所示。由图 2 可以看出，各预测方法的土壤镉含量空间分布格局大体相同，土壤高含镉区域主要位于甘肃、陕西、内蒙古、宁夏、山东。其中，甘肃省白银市镉污染最为严重，这可能与该市支柱产业为有色重金属加工、冶炼^[27]有关。李有文等^[28]的研究也表明，白银市地表土壤中重金属镉的含量远远超过土壤背景值。与回归克里格、RBF 神经网络模型预测结果相比，GARBF 模型预测结果多出陕西榆林、山西省太原西部、内蒙古包头市西南部 3 处土壤高含镉区域，这 3 处区域的土壤镉含量在 1 ~ 5 mg/kg。经资料分析发现，榆林

地区采矿业以及太原钢铁冶炼工业均为当地经济支柱产业；包头市钢铁冶炼工业和稀土产业极为著名，有着“草原钢城”“稀土之都”称号，且都集中分布在包头市西南部，这 3 种产业的生产过程均可能造成土壤重金属镉含量的升高。综上，基于土壤理化因子和 NDVI 的 GARBF 神经网络模型预测结果图更接近实际土壤重金属含量分布。

3 讨论

3.1 土壤理化因子对神经网络模型预测精度的影响

Mcbratney 等^[29]的研究证明，辅助变量数据可以提高神经网络模型对土壤重金属含量预测的准确性。为了准确预测土壤重金属含量的空间分布，应充分考

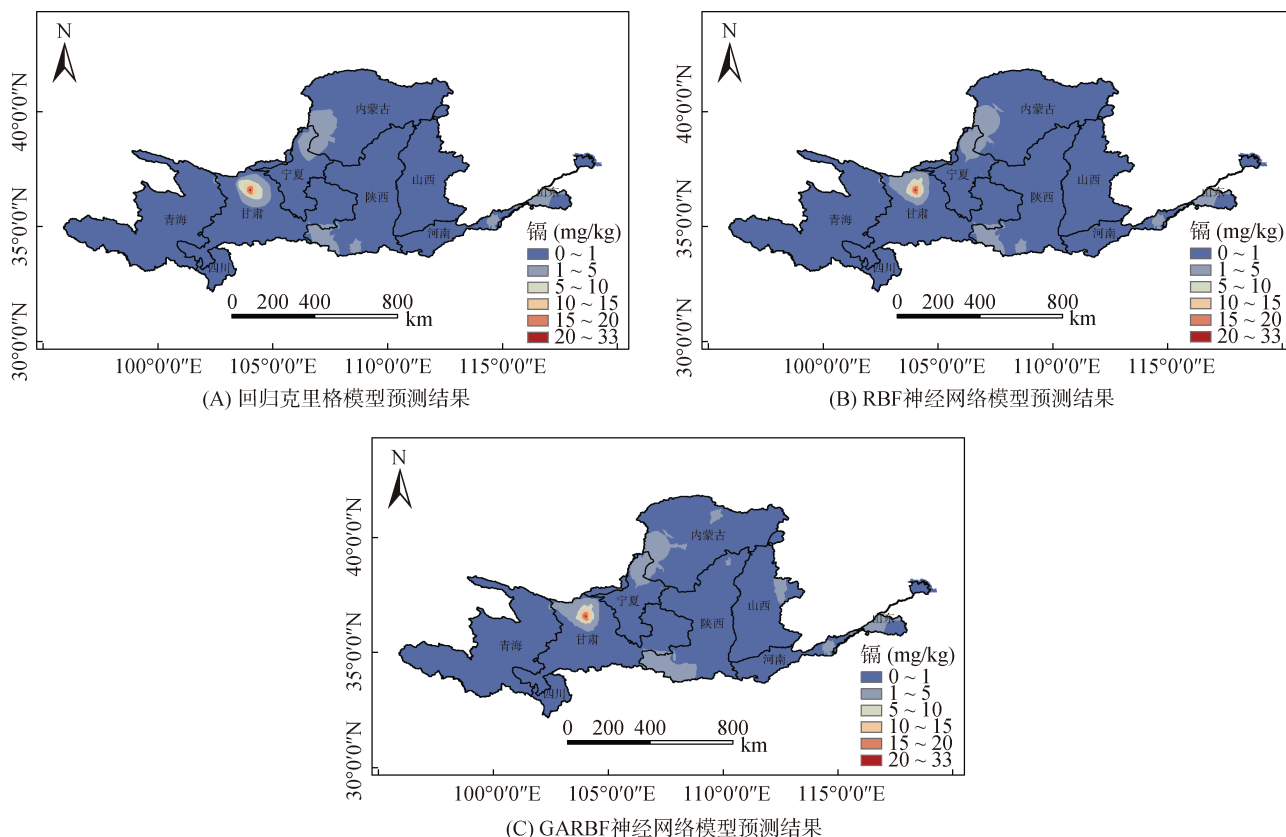


图 2 不同预测方法镉的空间分布

Fig. 2 Spatial distribution of cadmium in different prediction methods

考虑影响土壤重金属分布的因素,筛选出获取简单、相关性强的辅助数据。除环境因子外,土壤理化因子中 pH、OM、CEC 与土壤中镉含量相关性较强、可以量化且容易获取。pH 通过影响土壤中矿物质、水合氧化物等的表面负电荷来影响土壤对镉的吸附;OM 具有大量的功能团,吸附镉的能力高于其他矿质胶体,且 OM 分解形成的小分子有机酸、腐殖酸等可以与镉形成稳定的络合物;而 CEC 反映胶体的负电荷量,CEC 越高,负电荷量越高,吸附镉的能力越强。本研究显示,仅基于 NDVI 的神经网络模型预测精度最低,pH、OM、CEC 分别单独与 NDVI 作为辅助变量均能提高神经网络模型的预测精度,pH、OM、CEC、NDVI 共同作辅助变量时,神经网络模型的预测精度最高。

3.2 预测模型对比

在辅助变量相同的情况下,GARBF 神经网络模型的预测精度高于回归克里格模型和 RBF 神经网络模型。GARBF 神经网络模型通过遗传算法优化了 RBF 神经网络模型隐层到输出层权值,从而降低了 RBF 神经网络模型的插值误差,使 RBF 神经网络具

有更好的曲面逼近能力和全局寻优能力,因此遗传算法可以提高 RBF 神经网络模型的预测精度。回归克里格模型中的逐步多元回归计算过于简单,难以准确描述多个自变量与土壤镉之间的复杂非线性映射关系。此外,回归克里格模型中预测残差的克里格方法存在平滑效应。上述原因导致回归克里格模型预测精度较低。江叶枫等^[30]也认为应用神经网络模型来代替回归克里格的预测方法和误差插值方法。

4 结论

1) 基于土壤理化因子 OM、pH、CEC 和环境因子 NDVI 的 RBF 和 GARBF 神经网络模型的预测精度高于仅基于 NDVI 的 RBF 和 GARBF 神经网络方法;基于土壤理化因子和 NDVI 的回归克里格方法的预测精度高于仅基于 NDVI 的回归克里格模型。表明加入土壤理化因子(OM、pH、CEC)作为辅助变量可以提升神经网络模型的预测精度。

2) GARBF 神经网络模型通过遗传算法优化,在利用的辅助变量与 RBF 神经网络相同的情况下,GARBF 神经网络模型的预测精度高于 RBF 神经网络

模型; 基于土壤理化因子 OM、pH、CEC 和环境因子 NDVI 的 GARBF 神经网络模型, 通过遗传算法优化, 其预测精度高于基于土壤理化因子(OM、pH、CEC)和 NDVI 的回归克里格模型。

3) 根据 GARBF 神经网络模型预测结果, 黄河流域的土壤高含镉区域主要分布在甘肃白银市、陕西榆林市、内蒙古包头市西南部、山西太原市西部、内蒙古、宁夏、山东等地, 与实际相符, 表明基于土壤理化因子(OM、pH、CEC)和 NDVI 的 GARBF 神经网络模型预测黄河流域土壤镉空间分布具有较高的精度。

参考文献:

- [1] 马骄阳, 保欣晨, 王坤, 等. 土壤镉污染的人体健康风险评估研究: 生物有效性及毒性效应[J]. 生态毒理学报, 2021, 16(6): 120–132.
- [2] Turgut C, Katie Pepe M, Cutright T J. The effect of EDTA and citric acid on phytoremediation of Cd, Cr, and Ni from soil using *Helianthus annuus*[J]. Environmental Pollution, 2004, 131(1): 147–154.
- [3] 环境保护部, 国土资源部. 全国土壤污染状况调查公报[OL]. 2014-04-07. <https://www.gov.cn/foot/site1/20140417/782bcb88840814ba158d01.pdf>
- [4] 骆永明. 中国土壤环境污染态势及预防、控制和修复策略[J]. 环境污染与防治, 2009, 31(12): 27–31.
- [5] 张云菲, 孜比布拉·司马义, 杨胜天, 等. 农田土壤重金属污染特征、生态风险评估与来源分析[J]. 江苏农业科学, 2020, 48(4): 266–272.
- [6] 戴倩倩, 徐梦洁, 庄舜尧, 等. 基于地理探测器的封丘县农田土壤重金属分布影响因素研究[J]. 土壤, 2022, 54(3): 564–571.
- [7] 苗秀荣, 来雪慧, 李梦茜, 等. 不同钝化剂对土壤有效态重金属含量及其在小白菜中累积的影响[J]. 河南农业科学, 2020, 49(8): 63–71.
- [8] 梅雪, 刘鸿雁, 吴龙华, 等. 基于 HDXRF 和 ICP-MS 的黔西北土壤重金属空间分布及影响因素研究[J]. 土壤, 2023, 55(2): 399–408.
- [9] 汪洁, 朱有为, 杨肖娥, 等. 农用地土壤可持续安全利用的研究与实践——以浙江省镉污染农田为例[J]. 土壤学报, 2023, 60(6): 1662–1674.
- [10] 李启权, 王昌全, 岳天祥, 等. 基于定性和定量辅助变量的土壤有机质空间分布预测——以四川三台县为例[J]. 地理科学进展, 2014, 33(2): 259–269.
- [11] Li Q Q, Wang C Q, Dai T F, et al. Prediction of soil cadmium distribution across a typical area of Chengdu Plain, China[J]. Scientific Reports, 2017, 7(1): 7115.
- [12] 江叶枫, 郭熙. 基于多源辅助数据和神经网络模型的稻田土壤镉空间分布预测[J]. 环境科学学报, 2019, 39(3): 928–938.
- [13] 周伟, 李丽丽, 周旭, 等. 基于地理探测器的土壤重金属影响因子分析及其污染风险评估[J]. 生态环境学报, 2021, 30(1): 173–180.
- [14] 高中原, 肖荣波, 王鹏, 等. 融合自然-人为因子改进回归克里格对土壤镉空间分布预测[J]. 环境科学, 2021, 42(1): 343–352.
- [15] Ye H C, Huang W J, Huang S Y, et al. Effects of different sampling densities on geographically weighted regression Kriging for predicting soil organic carbon[J]. Spatial Statistics, 2017, 20: 76–91.
- [16] 于灏, 苏智杰, 祝培甜, 等. 水稻、小麦与土壤中重金属 Cd 含量的关系模拟研究[J]. 地学前缘, 2021, 28(1): 438–445.
- [17] Micó C, Recatalá L, Peris M, et al. Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis[J]. Chemosphere, 2006, 65(5): 863–872.
- [18] 江叶枫, 郭熙, 叶英聪, 等. 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[J]. 长江流域资源与环境, 2017, 26(8): 1150–1158.
- [19] 尹群, 郭纪敏, 张世文, 等. 基于辅助变量的县域土壤有机质预测[J]. 江苏农业科学, 2020, 48(24): 267–273.
- [20] 李保杰, 顾和和, 于法展, 等. 徐州市区土壤重金属空间分布研究[J]. 测绘科学, 2011, 36(5): 82–84, 89.
- [21] 解雪峰, 濮励杰, 沈洪运, 等. 滨海重度盐碱地改良土壤盐渍化动态特征及预测[J]. 土壤学报, 2022, 59(6): 1504–1516.
- [22] 柴杰, 江青茵, 曹志凯. RBF 神经网络的函数逼近能力及其算法[J]. 模式识别与人工智能, 2002, 15(3): 310–316.
- [23] Wang G X, Gertner G, Parysow P, et al. Spatial prediction and uncertainty analysis of topographic factors for the Revised Universal Soil Loss Equation (RUSLE)[J]. Journal of Soil and Water Conservation, 2000, 55: 374–384.
- [24] 董敏, 王昌全, 李冰, 等. 基于 GARBF 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.
- [25] 中华人民共和国生态环境部. 土壤和沉积物 19 种金属元素总量的测定 电感耦合等离子体质谱法: HJ 1315—2023[S]. 北京: 中国标准出版社, 2023.
- [26] Wang J F, Li X H, Christakos G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China[J]. International Journal of Geographical Information Science, 2010, 24(1): 107–127.
- [27] 南忠仁, 李吉均, 张建明, 等. 白银市区土壤作物系统重金属污染分析与防治对策研究[J]. 环境污染与防治, 2002, 24(3): 170–173.
- [28] 李有文, 曹春, 巨天珍, 等. 白银市不同区域蔬菜地土壤重金属污染特征及生态风险评估[J]. 生态学杂志, 2015, 34(11): 3205–3213.
- [29] McBratney A, Field D J, Koch A. The dimensions of soil security[J]. Geoderma, 2014, 213: 203–213.
- [30] 江叶枫, 孙凯, 郭熙, 等. 基于环境因子和邻近信息的土壤属性空间分布预测[J]. 环境科学研究, 2017, 30(7): 1059–1068.