

引用格式: 侯艺璇, 赵华甫, 吴克宁, 等. 基于BP神经网络的作物Cd含量预测及安全种植分区[J]. 资源科学, 2018, 40(12): 2414-2424. [Hou Y X, Zhao H F, Wu K N, et al. Prediction of crop Cd content and zoning of safety planting based on BP neural network[J]. Resources Science, 2018, 40(12): 2414-2424.] DOI: 10.18402/resci.2018.12.09

基于BP神经网络的作物Cd含量 预测及安全种植分区

侯艺璇¹, 赵华甫^{1,2}, 吴克宁^{1,2}, 李 凯¹

(1. 中国地质大学(北京)土地科学技术学院, 北京 100083;

2. 国土资源部土地整治重点实验室, 北京 100035)

摘要: 为科学预测作物重金属含量, 实现重金属污染农用地的安全利用, 本研究利用地理探测器选取对作物Cd含量影响较大的土壤Cd含量、土壤pH值、与交通主干线的距离等10个因素为输入因子, 农作物可食部分Cd含量作为输出因子, 分别建立小麦、水稻、油菜籽、蔬菜可食部分Cd含量的BP神经网络预测模型, 对作物种植污染情况进行预测分析, 划分作物安全利用种植区。研究表明: ① BP神经网络模型预测精度明显优于多元回归预测模型; ② 对小麦、水稻和油菜籽的可食部分Cd含量预测结果进行评价, 得到作物污染可能的空间分布及特征; ③ 依据评价结果, 对3种作物进行配置, 划分得到4种作物适宜种植区, 并提出管控策略。研究可为污染农用地的安全利用及作物种植调整提供思路和依据, 兼具理论和现实意义。

关键词: 作物Cd含量; BP神经网络; 预测模型; 重金属污染; 农用地安全利用; 小麦; 水稻; 油菜籽; 蔬菜; 作物适宜种植区

DOI: 10.18402/resci.2018.12.09

1 引言

随着工业化、城市化快速发展及农药化肥的广泛使用, 中国农用地重金属污染状况持续恶化。“镉大米”、“重金属蔬菜”等一系列食品安全问题的涌现, 使农作物重金属污染受到政府和社会各界的广泛关注。作物重金属含量的影响因素较为复杂, 受作物种类、土壤理化性质、土壤重金属含量、空气污染等诸多因素的影响^[1,2]。如何考虑复杂的自然、社会经济多重因素影响, 实现对作物重金属含量的预测, 并采取有效管控措施, 对保护人体健康和国家粮食安全具有重要的现实意义。

关于作物重金属含量预测模型的构建, 国内外多采用多元回归分析法建立数学模型^[3,4], 部分研究根据作物从土壤或大气中吸收重金属的机理建立

模型^[5,6], 实现对重金属含量的预测。但在回归分析中, 因子表达式选取较为困难, 当因子较多时构建模型较为复杂, 且其相关系数大多在0.4~0.9之间^[7,8], 而机理模型大多仅能预测重金属含量变化的大致趋势^[9], 预测精度有待提升。BP神经网络是利用非线性可微函数进行权值训练的反向传播网络^[9], 能够通过学习逼近任何非线性函数, 达到较高的预测精度^[10]。作为一种智能信息处理系统, BP神经网络不需要确定输入数据与输出数据间的映射关系, 能够通过训练学习的方式表现出事物的因果关系, 在给定输入时得到最接近期望输出值的结果, 适用于本研究预测作物重金属含量这种复杂的问题。

目前, 降低作物重金属含量的方式主要有两种: 一是通过土壤污染修复技术、大气污染治理、限制农

收稿日期: 2017-11-29; 修订日期: 2018-10-19

基金项目: 国家重点研发计划(2017YFD0800305); 国土资源部公益性行业专项经费项目(201511082)。

作者简介: 侯艺璇, 女, 河北邢台市人, 硕士生, 研究方向为土地资源利用与评价。E-mail: yxhou@cugb.edu.cn

通讯作者: 赵华甫, E-mail: huafuzhao@163.com

2018年12月

药化肥使用等途径控制重金属污染,但大多修复技术成本较高,周期较长^[11];二是根据不同作物对重金属富集能力的差异,通过调整作物种植结构,减少重金属向作物可食部分迁移,降低农产品超标风险^[12]。这种方式虽然不能永久地消除污染,但在短期内降低土壤重金属对人体的潜在危害,又能避免农用地修复过程中的大面积闲置。因此,针对较大范围污染区域,可考虑两种方式结合的污染防治措施,在调整种植结构的同时对污染严重区域优先修复治理,既能逐步减轻重金属污染,又能实现农用地的安全生产。

本文通过选取作物可食部分Cd含量的影响因子,构建BP神经网络预测模型,对小麦、水稻、油菜籽和蔬菜可食部分Cd含量进行预测,并根据结果进行作物种植分区,实现Cd污染农用地的安全利用。研究可为重金属污染农用地的安全利用及作物种植调整提供思路和依据,兼具理论和现实意义。

2 研究区概况、数据来源与研究方法

2.1 研究区概况

研究区A镇位于江苏省南部Y市(由于保密需要隐去实际地名),长江三角洲经济开发区,地势南高北低,气候温暖湿润,年均气温15.7℃。农作物一年2~3熟,农业集约化程度较高,是长江三角洲商品粮基地,主要作物为小麦、水稻、油菜籽及各类蔬菜。当地工业企业发达,采矿场、冶炼厂、蓄电池厂等工业活动在生产过程中产生的三废给环境带来了超负荷的压力,造成了周边农用地严重的重金属污染。

2.2 数据来源

本研究收集到的数据及其来源包括:

(1)2015年11月至2016年5月,在Y市布设采集115个土壤样点,并在对应土壤样点上采集作物样本小麦49组、水稻42组、油菜籽39组、蔬菜46组,对土壤理化性质、土壤及作物Cd含量进行测定。其中,土壤与作物中Cd含量采用石墨炉原子吸收光谱法测定。

(2)A镇全域加密布设样点1987个(图1),包含土壤Cd含量及土壤理化性质,数据来源于2011年江苏省Y市耕地质量生态地球化学调查与等级评价数据库,利用ArcGIS反距离权重插值法获得全域信息。

(3)Y市和A镇的土壤粘粒含量数据来源于1988年第二次全国土壤普查数据。

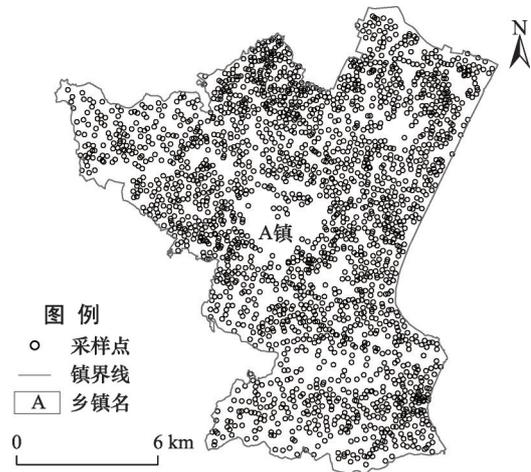


图1 A镇土壤样品采集点分布

Figure 1 Location of soil sampling sites in A town

(4)Y市和A镇的工业、交通、建成区、灌溉等数据来源于2013年江苏省Y市第二次土地利用现状调查成果数据库,利用ArcGIS中Near模块计算得出采样点与各要素距离。

利用Y市数据进行神经网络训练,模型训练完成后,将A镇数据输入模型用于对A镇作物可食部分Cd含量的预测。研究表明pH值、有机质、粘粒等土壤理化性质的年均变化速率较小^[13,14],且在各数据来源年份相近的条件下土地利用变化幅度较小,但数据获取时间不同仍会使预测结果与实际情况之间存在一定的误差。

2.3 模型输入因子的确定

作物中Cd含量往往是多种因素共同作用的结果,其作用机理较为复杂,很难估计它们的独立作用,需综合考虑各影响因子间的交互作用^[15]。地理探测器(GDM)是探测空间分异性,以及揭示其背后驱动力的一组统计学方法,该方法主要包括因子探测、交互探测、风险探测和生态探测四个部分,其中交互探测可以评估两因子共同作用于某变量的解释力^[16]。探测输入因子对作物Cd含量解释力大小主要通过比较各因子在不同类别分区上的总方差与作物Cd含量在整个研究区上的总方差,比率越小则解释力越强。一般表达式为:

$$q = 1 - \frac{\sum_{k=1}^M N_k \sigma_k^2}{N \sigma^2} \quad (1)$$

式中 q 是因子对作物Cd含量的解释力,值域是[0,

1], $q=0$ 表明因子对变量毫无解释力,二者无任何关系, $q=1$ 表明因子完全可以解释变量的空间分布; $k=1, 2, \dots, K$ 为作物 Cd 含量或因子的分类; N_k 和 N 是类别 k 和全区的单元数; σ_k^2 和 σ^2 分别是类别 k 内的方差和全区作物 Cd 含量值的总方差。

利用交互探测选取对作物 Cd 含量解释力较大的因子作为模型输入因子(表1), 因子与土壤 Cd 含量交互后对作物 Cd 含量的解释力较强, 其解释力大小见表2, 所选因子在 A 镇的空间分布状况如图2。

2.4 神经网络模型构建

为了对作物可食部分 Cd 含量进行高精度预测, 建立 BP 神经网络模型。BP 神经网络是由输入层、隐含层和输出层组成的前馈神经网络, 训练过程主要是依据神经网络实际输出与期望输出之间误差的负梯度方向, 从后往前逐层迭代修正各层神经元之间的连接权值^[17], 图3(见第2418页)显示了BP神经网络的架构。

在三层前馈网络中, 输入向量为 $\mathbf{X}=(x_1, x_2, \dots, x_m)^T$, 隐含层输出向量为 $\mathbf{S}=(s_1, s_2, \dots, s_n)^T$, 期望输出向量为 $\mathbf{Y}=(y_1, y_2, \dots, y_j)^T$, 隐含层中第 n 个神经元的输出为^[10]:

$$v_n = f_1 \left(\sum_{m=1}^M \omega_{mn} x_m - \square_n \right) \quad (n=1, 2, \dots, N) \quad (2)$$

式中 v_n 为每个节点的输出值; ω_{mn} 为节点 m 与节点 n 之间的权值; \square_n 为节点 n 的阈值; f_1 为激活函数。输

出层第 j 个神经元的输出为:

$$y_j = f_2 \left(\sum_{n=1}^N w_{nj} v_n - \theta_j \right) \quad (j=1, 2, \dots, J) \quad (3)$$

式中 y_j 为每个节点的输出值; w_{nj} 为节点 n 与节点 j 之间的权值; θ_j 为节点 j 的阈值; f_2 为激活函数。激活函数多采用 sigmoid 函数:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

使用 MATLAB 中的神经网络工具箱, 利用 Y 市采集的样本数据集进行神经网络的学习训练, 分别将 39 组小麦、34 组水稻、31 组油菜籽、37 组蔬菜数据作为训练样本集, 10 组小麦、8 组水稻、8 组油菜籽、9 组蔬菜数据作为检验样本集, 训练并选择预测效果较好的网络^[18]; 再将 A 镇 10 个因子数据集作为输入因子输入到训练好的网络, 利用神经网络模拟预测 A 镇区域内作物可食部分的 Cd 含量。为保证网络快速收敛, 模型采用 Levenberg-Marquardt 反传算法; 由于三层前馈网络可以逼近任何非线性函数, 因此构建单隐含层结构^[10]; 隐含层和输出层中神经元之间采用的激活函数为双曲正切 S 型函数 tansig。

2.5 数据预处理

为了消除纲量不同对结果的影响, 提高网络训练效率, 对样本进行归一化处理, 将所有数据归一化到 $[-1, 1]$ 区间内, 公式如下:

$$X_{\text{norm}} = (a - b) (X - X_{\text{min}}) / (X_{\text{max}} - X_{\text{min}}) + b \quad (5)$$

表1 预测作物可食部分 Cd 含量的 BP 神经网络模型输入因子

Table 1 Input factors of BP neural network for predicting Cd content in edible part of crops

变量	因子	单位	变量	因子	单位
x_1	土壤 Cd 含量	mg/kg	x_6	土壤 Zn 含量	mg/kg
x_2	土壤 pH 值	-	x_7	与交通主干道距离	m
x_3	阳离子交换量(CEC)	cmol/kg	x_8	与建成区距离	m
x_4	土壤有机质含量(SOM)	%	x_9	与采矿用地距离	m
x_5	土壤粘粒含量	%	x_{10}	与灌溉水源距离	m

表2 Y 市模型输入因子与土壤 Cd 含量交互作用对作物可食部分 Cd 含量的解释力大小

Table 2 The explanatory power of interaction between the input factors and the soil Cd content on Cd content of edible parts of crops in Y city

变量	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
小麦	0.810	0.818	0.839	0.879	0.853	0.754	0.897	0.932	0.758
水稻	0.649	0.849	0.912	0.669	0.896	0.593	0.852	0.890	0.663
油菜籽	0.591	0.783	0.409	0.496	0.824	0.786	0.810	0.716	0.684
蔬菜	0.966	0.962	0.531	0.964	0.958	0.365	0.369	0.535	0.523

2018年12月

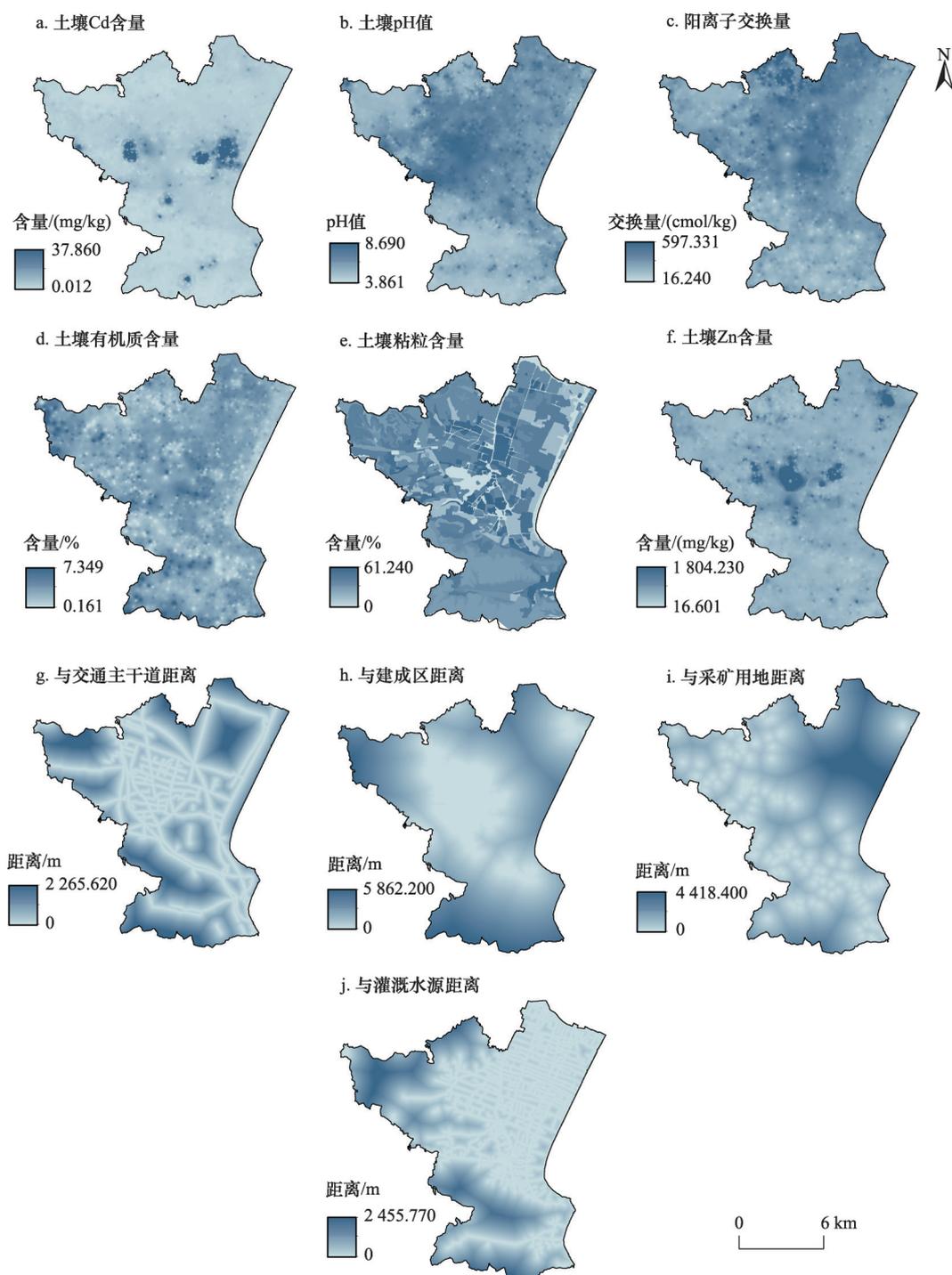


图2 A镇预测模型输入因子空间分布

Figure 2 Spatial distribution of input factors of prediction model in A town

式中 X 、 X_{norm} 、 X_{max} 和 X_{min} 分别为样本的实测值、归一化值、最大值和最小值; a 、 b 分别为归一化区间的最大值和最小值。对于输出结果还需进行反归一化处理。

2.6 模型精度分析

为了衡量模型的性能,选取决定系数(R^2)、均方

根误差(RMSE)、平均绝对误差(MAE)、平均相对误差(MRE)为模型精度评价指标。

R^2 越大, RMSE、MAE和MRE值越小,模型精度越高。

2.7 作物可食部分污染等级评价方法及标准

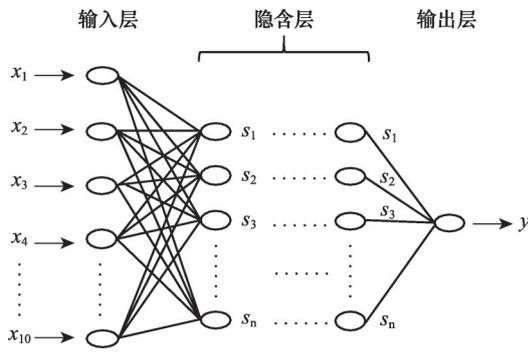


图3 BP神经网络结构示意图

Figure 3 Schematic of the structure of BP neural network

目前,中国常用的农作物可食部分重金属限量超标评价方法大多采用单因子指数法,其公式为:

$$E = \frac{CC}{CS} \quad (6)$$

式中 E 为作物可食部分中 Cd 含量的超标指数; CC 为作物可食部分中 Cd 含量; CS 为作物可食部分中 Cd 的评价标准值,采用国家标准《食品中污染物限量》(GB 2762-2017)^[19](表3);参考环保部《全国土壤污染状况评价技术规定》(环发[2008]39号)^[20],根据 E 的大小,将作物重金属超标程度划分为五级(表4)。

3 结果及分析

3.1 BP神经网络的训练及检验

经过不断调试,选择训练及预测效果较好的模型。小麦、水稻、油菜籽、蔬菜的训练样本集和检验样本集的相关评价指标见表5、图4。其中,小麦、水稻、油菜籽三种模型的训练效果较好, R^2 较高,皆在 0.01 水平上显著; RMSE、MAE、MRE 较小,其中 MRE 不超过 30%; 模型预测精度较高,说明 BP 神经网络对于 3 种作物 Cd 含量的预测具有很好的适用性。但蔬菜模型的预测效果相对其他 3 种作物模型较差,训练集和检验集的误差较大。主要原因可能是蔬菜样

表3 农作物可食部分中 Cd 限量标准值^[19]

农作物种类	小麦	水稻	油菜籽	蔬菜
限量标准值	0.10	0.20	0.50	0.05

表4 农作物可食部分 Cd 超标指数对应的超标程度^[20]

超标等级	E	超标程度
I	$E \leq 1$	未超标
II	$1 < E \leq 2$	轻微超标
III	$2 < E \leq 3$	轻度超标
IV	$3 < E \leq 5$	中度超标
V	$E > 5$	重度超标

本中包含白菜类、甘蓝类等多个品种,对 Cd 的吸收能力差异较大,无法达到较好的学习与预测效果。

4 种作物模型预测结果如图 5,可以看出,模型在对作物 Cd 含量较高区间的预测误差较大而在低区间预测较准确,可能与数据代表性有关,在采集数据中作物 Cd 含量较低的样本所占比例较大,造成样本数据分布不均匀,网络训练结果不稳定,因而模型预测结果在高 Cd 含量区间较差^[21]。

3.2 与多元回归法对比分析

以往作物预测模型多采用多元回归法构建,为对比两种方法的预测精度,使用多元回归法将作物可食部分 Cd 含量与上述因子进行拟合。为压缩变量尺度、消除异方差,采用对数模型,以相同比例划分得到建模集和检验集,得到 4 种作物的多元回归模型(表 6,见第 2420 页),并计算精度评价指标(表 7,见第 2420 页)。由表 7 可得,与 BP 神经网络模型相比,多元回归模型的 R^2 较低, RMSE、MAR 和 MRE 较大,多元回归模型预测精度远不如 BP 神经网络模型。

3.3 预测污染分布特征

将 A 镇因子数据输入已训练好的 BP 神经网络

表5 BP神经网络训练样本和检验样本精度评价指标

Table 5 Results of accuracy evaluation indexes of train and test samples of BP neural network

	R^2		RMSE		MAE		MRE	
	训练集	检验集	训练集	检验集	训练集	检验集	训练集/%	检验集/%
小麦	0.998**	0.979**	0.010	0.132	0.008	0.087	12.231	27.084
水稻	0.993**	0.969**	0.021	0.129	0.008	0.082	8.995	22.998
油菜籽	0.969**	0.916**	0.009	0.038	0.007	0.029	19.413	27.517
蔬菜	0.961**	0.546*	0.331	0.429	0.240	0.258	134.060	83.530

注: *、** 分别表示在 0.05、0.01 水平上显著。

2018年12月

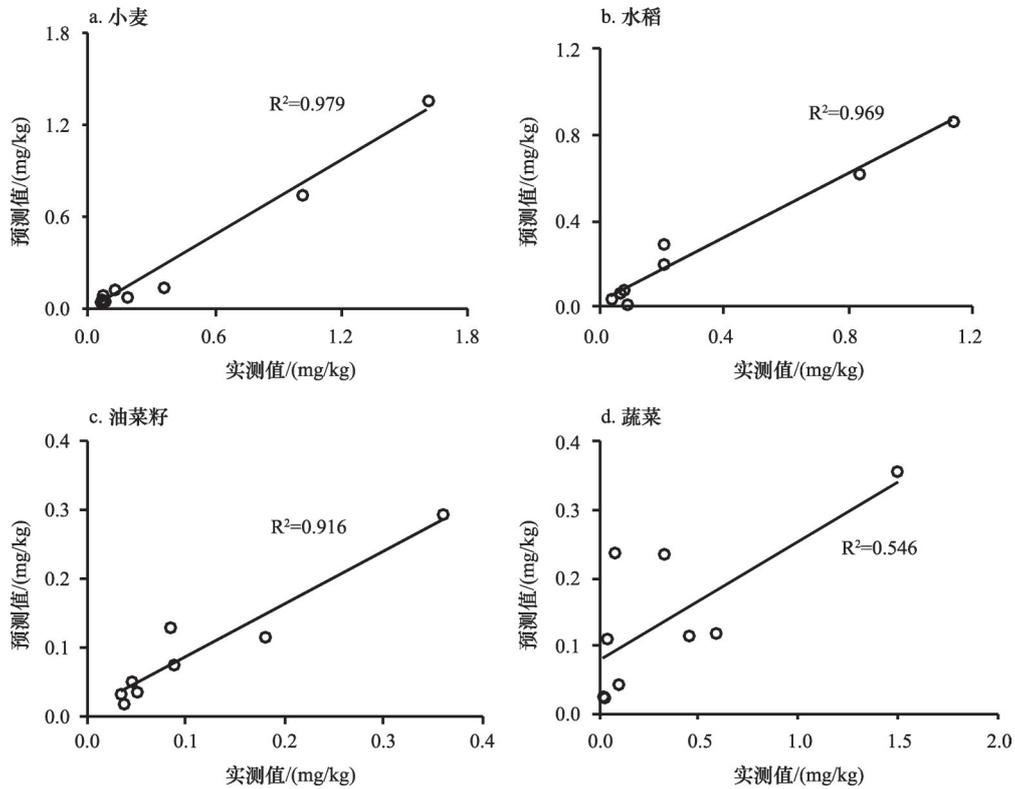


图4 BP神经网络模型检验集预测结果相关性分析

Figure 4 Correlation analysis of prediction results of test set of BP neural network

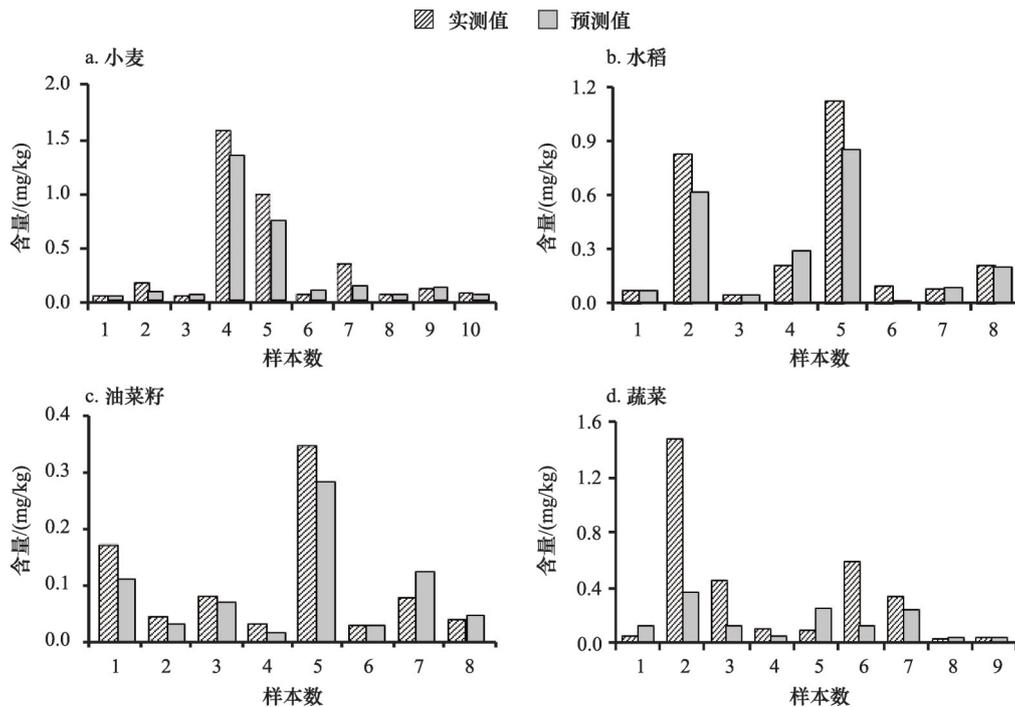


图5 研究区BP神经网络模型检验集作物可食部分Cd含量预测值与实测值

Figure 5 Predicted values and measured values of Cd content of crops edible part in test set of BP neural network model in study area

表6 4种作物可食部分Cd含量与各因子的多元回归方程

Table 6 Multivariate regression equations of Cd content in edible parts of 4 crops and factors

预测方程	
小麦	$\lg Cd_{(小麦)} = 1.0001g x_1 - 2.2671g x_2 - 0.8751g x_3 - 0.4271g x_4 - 0.1691g x_5 + 0.1231g x_6 + 0.0471g x_7 + 0.1691g x_8 - 0.2231g x_9 + 0.0411g x_{10} + 3.491$
水稻	$\lg Cd_{(水稻)} = 1.0591g x_1 - 2.5391g x_2 - 0.6631g x_3 - 1.8161g x_4 - 0.7301g x_5 - 0.8151g x_6 + 0.2001g x_7 - 0.1741g x_8 - 0.1531g x_9 + 0.3121g x_{10} + 7.699$
油菜籽	$\lg Cd_{(油菜籽)} = 0.4101g x_1 - 0.1741g x_2 - 1.3741g x_3 + 0.7611g x_4 + 0.6291g x_5 - 0.6041g x_6 - 0.1541g x_7 + 0.0501g x_8 - 0.1371g x_9 + 0.0131g x_{10} + 2.304$
蔬菜	$\lg Cd_{(蔬菜)} = 0.5411g x_1 + 2.2821g x_2 - 1.5981g x_3 - 0.0941g x_4 - 1.1991g x_5 - 0.5631g x_6 - 0.1651g x_7 - 0.1211g x_8 - 0.0801g x_9 - 0.5351g x_{10} + 6.429$

表7 多元回归建模样本和检验样本精度评价指标

Table 7 Results of accuracy evaluation indexes of train and test samples of multiple regression

	R ²		RMSE		MAE		MRE	
	建模集	检验集	建模集	检验集	建模集	检验集	建模集/%	检验集/%
小麦	0.756**	0.854**	0.200	0.177	0.081	0.106	43.774	59.649
水稻	0.519**	0.783**	0.304	0.229	0.151	0.144	116.162	52.261
油菜籽	0.507**	0.915**	0.035	0.114	0.023	0.062	45.636	28.354
蔬菜	0.128*	0.552*	1.498	0.450	0.448	0.262	132.524	120.077

注: *、**分别表示在0.05、0.01水平上显著。

模型,得到作物Cd含量的预测结果,并对其进行评价,利用ArcGIS10.2将A镇预测及评价结果可视化(见图6)。由于蔬菜样本包含多个品种,模型预测效果较差,对于预测和管理分区支撑作用有限,故仅考虑小麦、水稻和油菜籽这三类大田作物的分布特征。

(1)小麦Cd污染超标地区主要分布在A镇的东部和南部地区,同时在西北部和东北部也有少量分布。主要原因可能是在东部和南部地区,土壤Cd含量较高,交通线路较为密集,此外西北部和南部地

区的土壤pH、CEC含量较低,可以在一定程度上提高土壤Cd的活性,促进小麦对土壤中Cd的吸收^[22]。

(2)水稻Cd污染超标地区与小麦分布较为接近,主要分布在A镇西北部、南部和东部地区,此外在建成区附近也有分布。其分布原因与小麦相似,可能由于土壤Cd含量较高或土壤pH、CEC较低而促进了土壤Cd向水稻的迁移。

(3)油菜籽的分布表现为全域不超标。主要原因是油菜籽对Cd具有较低的富集能力^[23],在对A镇种植油菜籽可食部分Cd含量的预测结果中,其平均

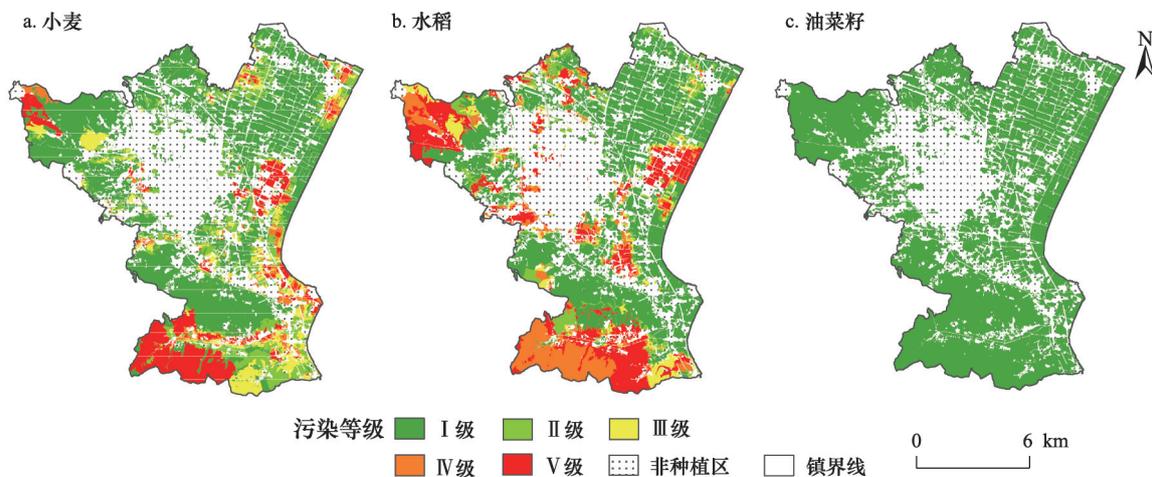


图6 A镇作物可食部分Cd污染预测等级评价结果

Figure 6 Evaluation results of predicted Cd pollution for edible parts of crops in A town

2018年12月

值仅为0.09mg/kg,含量最高值为0.36mg/kg,远小于其评价标准的安全阈值0.50mg/kg。

3.4 种植区划调整

将4种作物未污染地区叠加(图7a),以农用地安全利用为目标,为便于分区管理,对细碎斑块进行整合,得到4种作物种植安全区(图7b),并根据污染状况提出以下管控策略:

(1)小麦、水稻、油菜籽适宜种植区。该区面积4197.47 hm²,占农用地总面积的33.94%。该区的土壤Cd污染程度低,应实施优先保护策略,防止污染输入,维护现有安全状态,同时通过优化农艺生产措施如施肥、间作等降低作物对土壤Cd的富集,确保农用地安全利用。

(2)水稻、油菜籽适宜种植区。该区面积2203.58hm²,占农用地总面积17.82%。该区土壤Cd污染程度中等,多处于交通用地和采矿用地周边,应实施综合监控策略,对土壤、灌溉水、大气环境实施动态监测,预防人类活动对作物的污染。

(3)小麦、油菜籽适宜种植区。该区面积2791.15hm²,占农用地总面积22.57%。该区土壤Cd污染程度中等,作物污染主要由于土壤性质等因素,应通过优化农艺措施,抑制作物对Cd的吸收。

(4)油菜籽适宜种植区。该区面积3175.12hm²,占农用地总面积25.67%。该区多为土壤Cd污染超标严重的地区,由于国家标准《食品中污染物限

量》^[19]中油菜籽的限量值较高,种植油菜籽的预测值远低于该限量值,因此在该区域种植油菜籽不足以对消费者膳食安全产生较大威胁,但针对该区应严格管控治理污染源,并进行一定程度的修复治理,对农作物、土壤、大气等污染状况进行监测。

4 讨论

由两种方法的精度评价结果来看,BP神经网络模型的预测效果优于多元回归模型。由于BP神经网络模型是一种黑箱模型,可根据多个相互作用复杂的因子对某一变量进行预测,能够处理复杂模糊的映射关系且不需知道数据的分布形式和与变量间的关系,克服了多元回归模型对复杂因子关系模拟较差的缺点^[24],因此相对于多元回归法来说,BP神经网络可以更好地模拟作物富集Cd这一过程的非线性规律,具有一定的适用性和优越性。

本文利用GDM,在综合考虑各影响因子的交互作用下,选取了土壤Cd含量、土壤pH、CEC、与交通主干线的距离等10个因子,克服了以往神经网络在输入因子选择中仅考虑线性相关性而忽略了某些因子,进而限制了模型精度提升^[21]。选取因子与冯金飞的研究较为相似,该研究采用BP神经网络预测高速公路周边作物重金属含量,选择了与公路的距离、车流量、土壤pH、SOM、土壤重金属含量等作为输入因子,达到了较高的预测精度^[25]。大量实

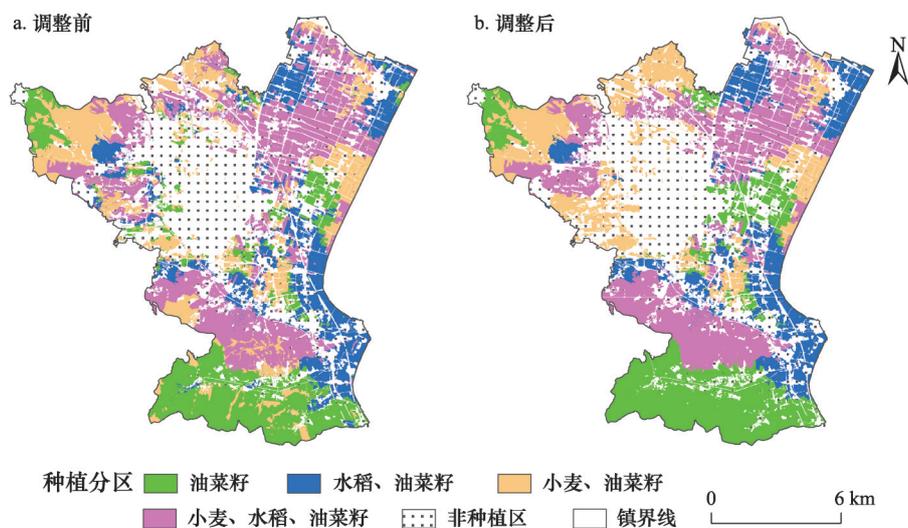


图7 A镇作物适宜种植分区

Figure 7 Zoning of suitable planting for crops in A town

验发现,相关土壤性质如土壤pH、CEC、SOM、粘粒含量及其他重金属含量能够影响土壤Cd的活性,进而影响Cd从土壤到作物的迁移^[26]。因此,在以往多元回归模型中,大多选取上述土壤性质因素^[8]。但同时,研究表明作物Cd含量还会受到空气污染、灌溉等人类活动的影响,道路、工业企业排放的废气、废水将在一定程度上增加作物Cd含量^[27,28];Novotná M等的研究表明,在建模过程中,空气污染途径不应该被忽略^[8],因此本文在因子选取上还加入了与交通主干道、建成区、采矿用地等与潜在污染源距离因素。

与冯金飞构建的BP神经网络模型相比^[25],本文的相对误差较大。在其研究中,小麦和水稻可食部分Cd含量的检验集相对误差范围分别为1.200%~9.800%和1.030%~16.940%,而本文小麦和水稻检验集的MRE分别为27.084%和22.998%。原因可能是在高速公路沿线区受车流量影响较大,研究区较小,主导影响因素相对简单,而相对于线状研究区,面域研究区的影响因素更为复杂,因此本研究预测难度更大。同时,不同作物品种或同一品种不同基因型对Cd的富集能力不同^[9],也会造成一定的误差。

本研究中BP神经网络预测精度相对较高,模型对于作物Cd含量预测及种植分区研究有很好的适用性。结合因子分析、遗传算法等方法对BP神经网络加以改进,是否会进一步提高网络精度及训练效率,是今后进一步的研究方向。此外,增加网络训练样本数量以及对蔬菜样本的具体类别进行划分,会在一定程度上提高模型的预测精度。由于模型预测存在误差,对于分区的准确性会产生一定影响,因此适宜种植区边界的划定并不是绝对的而是一个相对的范围,建议当地尽量选择种植安全阈值较高或非食用的植物,如油菜籽、林木、芦苇、鲜花等,并对农用地进行监测预警。另外,种植作物的分区还应考虑当地经济投入成本及农户的可接受度等问题。上述问题也为今后的补充完善提供了研究方向。

5 结论

本文通过GDM选取对作物Cd含量影响较大

的10个因子,利用BP神经网络分别构建小麦、水稻、油菜籽、蔬菜可食部分Cd含量预测模型,结果表明:

(1)BP神经网络模型预测效果较好,该方法对于作物Cd含量预测具有很好的适用性。在检验集中,小麦、水稻、油菜籽、蔬菜的 R^2 分别为0.979、0.969、0.916和0.546;RMSE分别为0.132、0.129、0.038和0.429;MAE分别为0.087、0.082、0.029和0.258;MRE分别为27.084%、22.998%、27.517%和83.530%,模型精度明显优于多元回归预测模型。而蔬菜预测效果相对较差的原因可能是蔬菜样本包含多个品种,不同品种对Cd的吸收差异较大。

(2)由于蔬菜预测结果误差较大,仅对小麦、水稻和油菜籽三种作物的可食部分Cd含量预测结果进行评价,得到三种作物污染预测分布图。其中,小麦和水稻的污染超标地区主要位于东部、南部和西北部,油菜籽全域不污染。

(3)依据评价结果,以农用地安全利用为目标,对研究区小麦、水稻和油菜籽三种作物进行重新配置,划分得到4种作物种植安全区,并根据污染状况和当地实际提出污染农用地安全利用管控策略。研究可为实现重金属污染农用地安全利用提供思路 and 依据,兼具理论和现实意义。

参考文献(References):

- [1] Legind C N, Trapp S. Comparison of prediction methods for the uptake of As, Cd and Pb in carrot and lettuce[J]. *Sar&Qsar in Environmental Research*, 2010, 21(5-6):513-525.
- [2] Pandey J, Pandey U. Atmospheric deposition and heavy metal contamination in an organic farming system in a seasonally dry tropical region of India[J]. *Journal of Sustainable Agriculture*, 2009, 33(4):361-378.
- [3] 王梦梦, 何梦媛, 苏德纯. 稻田土壤性质与稻米镉含量的定量关系[J]. *环境科学*, 2018, (4):1-11. [Wang M M, He M Y, Su D C. Quantitative relationship between paddy soil properties and cadmium content in rice grains[J]. *Environmental Science*, 2018, (4): 1-11.]
- [4] Bešter P K, Lobnik F, Eržen I, et al. Prediction of cadmium concentration in selected home-produced vegetables [J]. *Ecotoxicology & Environmental Safety*, 2013, 96(6):182-190.
- [5] Hough R L, Young S D, Crout N M J. Modelling of Cd, Cu, Ni, Pb and Zn uptake, by winter wheat and forage maize, from a sewage disposal farm[J]. *Soil Use & Management*, 2010, 19(1):19-27.

2018年12月

- [6] 施亚星, 吴绍华, 周生路, 等. 土壤-作物系统中重金属元素吸收、迁移和积累过程模拟[J]. 环境科学, 2016, 37(10):3996-4003. [Shi Y X, Wu S H, Zhou S L, et al. Simulation of the absorption, migration and accumulation process of heavy metal elements in soil-crop system[J]. *Environmental Science*, 2016, 37(10):3996-4003.]
- [7] Chen H, Yuan X, Li T, et al. Characteristics of heavy metal transfer and their influencing factors in different soil-crop systems of the industrialization region, China[J]. *Ecotoxicology & Environmental Safety*, 2016, 126(2):193-201.
- [8] Novotná M, Mikeš O, Komprdová K. Development and comparison of regression models for the uptake of metals into various field crops[J]. *Environmental Pollution*, 2015, 207(3):357-364.
- [9] 周开利, 康耀红. 神经网络模型及其MATLAB仿真程序设计[M]. 北京:清华大学出版社, 2005. [Zhou K L, Kang Y H. *Neural Network Model and Its MATLAB Simulation Program Design* [M]. Beijing: Tsinghua University Press, 2005.]
- [10] 周政. BP神经网络的发展现状综述[J]. 山西电子技术, 2008, (2):90-92. [Zhou Z. Survey of current progress in BP neural network[J]. *Shanxi Electronic Technology*, 2008, (2):90-92.]
- [11] 樊霆, 叶文玲, 陈海燕, 等. 农田土壤重金属污染状况及修复技术研究[J]. 生态环境学报, 2013, (10):1727-1736. [Fan T, Ye W L, Chen H Y, et al. Review on contamination and remediation technology of heavy metal in agricultural soil[J]. *Ecology and Environmental Sciences*, 2013, (10):1727-1736.]
- [12] 张倩, 赵华甫, 吴克宁, 等. 重金属超标农用地安全利用的地类转换模式研究[J]. 土壤通报, 2016, 47(5):1071-1076. [Zhang Q, Zhao H F, Wu K N, et al. Land type transfer mode for the safety utilization of heavy metals over standard agriculture land[J]. *Chinese Journal of Soil Science*, 2016, 47(5):1071-1076.]
- [13] 朱静, 黄标, 孙维侠, 等. 农田土壤有效态微量元素的时空变化及其影响因素研究[J]. 南京大学学报(自然科学), 2007, 43(1):1-12. [Zhu J, Huang B, Sun W X, et al. Temporal-Spatial Variability and Its Influencing Factors of Available Trace Metals in Soils[J]. *Journal of Nanjing University (Natural Sciences)*, 2007, 43(1):1-12.]
- [14] 王振宇. 保护性耕作对冀西北高原土壤物理性质和养分的影响[D]. 保定:河北农业大学, 2010. [Wang Z Y. Effects of Conservation Tillage on Soil Physical Properties and Nutrient in Northwest Plateau of Hebei Province[D]. Baoding: Hebei Agricultural University, 2010.]
- [15] Dayton E A, Basta N T, Payton M E, et al. Evaluating the contribution of soil properties to modifying lead phytoavailability and phytotoxicity[J]. *Environmental Toxicology & Chemistry*, 2006, 25(3):719-725.
- [16] 王劲峰, 徐成东. 地理探测器:原理与展望[J]. 地理学报, 2017, 72(1):116-134. [Wang J F, Xu C D. Geodetector: principle and prospective[J]. *Acta Geographica Sinica*, 2017, 72(1):116-134.]
- [17] 张雨浓, 杨逸文, 李巍. 神经网络权值直接确定法[M]. 广州:中山大学出版社, 2010. [Zhang Y N, Yang Y W, Li W. *Neural Network Weight Direct Determination Method*[M]. Guangzhou: Sun Yat-sen University Press, 2010.]
- [18] 邹志红, 王学良. BP模型在河流水质预测中的误差分析[J]. 环境科学学报, 2007, 27(6):1038-1042. [Zou Z H, Wang X L. The errors analysis for river water quality prediction based on BP-modelling[J]. *Acta Scientiae Circumstantiae*, 2007, 27(6):1038-1042.]
- [19] 中华人民共和国国家卫生和计划生育委员会, 国家食品药品监督管理总局. 食品安全国家标准食品污染物限量(GB 2762-2017)[EB/OL](2017-04-14)[2018-09-08]. <http://www.nhfp.gov.cn/sps/s7891/201704/b83ad058ff544ee39dea811264878981.shtml>. [National Health and Family Planning Commission of People's Republic of China, China Food and Drug Administration. National Food Safety Standard for Limit of Contaminants in Food(GB 2762-2017)[EB/OL] (2017-04-14)[2018-09-08]. <http://www.nhfp.gov.cn/sps/s7891/201704/b83ad058ff544ee39dea811264878981.shtml>.]
- [20] 中华人民共和国环境保护部. 关于印发《全国土壤污染状况评价技术规定》的通知(环发[2008]39号)[EB/OL]. (2012-01-12)[2018-09-08]. <http://www.er-china.com/PowerLeader/html/2012/01/20120112005153.shtml>. [Ministry of Environmental Protection of People's Republic of China. Notice on Printing and Distributing the Technical Regulations on the Evaluation of Soil Pollution Status in the Whole Country (Huanfa [2008]No. 39) [EB/OL] (2012-01-12)[2018-09-08]. <http://www.er-china.com/PowerLeader/html/2012/01/20120112005153.shtml>.]
- [21] 陈博, 欧阳竹. 基于BP神经网络的冬小麦耗水预测[J]. 农业工程学报, 2010, 26(4):81-86. [Chen B, Ouyang Z. Prediction of winter wheat evapotranspiration based on BP neural networks[J]. *Transactions of the CSAE*, 2010, 26(4): 81-86.]
- [22] 夏运生, 王凯荣, 张格丽. 土壤镉生物毒性的影响因素研究进展[J]. 农业环境科学学报, 2002, 21(3):272-275. [Xia Y S, Wang K R, Zhang G L. Research advances in influence factors of phytotoxicity of cadmium in soil[J]. *Journal of Agro-Environment Science*, 2002, 21(3):272-275.]
- [23] 牛雅典. 不同油菜品种对Cd耐性差异及其初步机理研究[D]. 武汉:华中农业大学, 2012. [Niu Y D. Differences of Cd Tolerance among Different Rapeseed Species and Their Preliminary Mechanism[D]. Wuhan: Huazhong Agricultural University, 2012.]
- [24] 张景阳, 潘光友. 多元线性回归与BP神经网络预测模型对比与运用研究[J]. 昆明理工大学学报(自然科学版), 2013, (6):61-67. [Zhang J Y, Pan G Y. Comparison and application of multiple regression and BP neural network prediction model [J]. *Journal of Kunming University of Science and Technology (Natural Science Edition)*, 2013, (6):61-67.]
- [25] 冯金飞. 高速公路沿线农田土壤和作物的重金属污染特征及规律[D]. 南京:南京农业大学, 2010. [Feng J F. Heavy Metal Pollution Characteristics and Laws of Farmland Soils and Crops along Expressway[D]. Nanjing: Nanjing Agricultural University, 2010.]
- [26] 汤丽玲. 作物吸收Cd的影响因素分析及籽实Cd含量的预测

- [J]. 农业环境科学学报, 2007, 26(2):699-703.[Tang L L. Effects of soil properties on crop Cd uptake and prediction of Cd concentration in grains[J]. *Journal of Agro-Environment Science*, 2007, 26(2):699-703.]
- [27] 严连香, 黄标, 邵学新, 等. 不同工业企业周围土壤-作物系统重金属Pb、Cd的空间变异及其迁移规律[J]. 土壤学报, 2009, 46(1):52-62.[Yan L X, Huang B, Shao X X, *et al.* Spatial variability and transfer of Pb and Cd in soil-crop system around different types of factories[J]. *Acta Pedologica Sinica*, 2009, 46(1):52-62.]
- [28] Sharma R K, Agrawal M, Marshall F M. Atmospheric deposition of heavy metals (Cu, Zn, Cd and Pb) in Varanasi City, India[J]. *Environmental Monitoring & Assessment*, 2008, 142(1-3):269-278.

Prediction of crop Cd content and zoning of safety planting based on BP neural network

HOU Yixuan¹, ZHAO Huafu^{1,2}, WU Kening^{1,2}, LI Kai¹

(1. School of Land Science and Technology, China University of Geosciences (Beijing), Beijing 100083, China;

2. Key Laboratory of Land Regulation Ministry of Land and Resources, Beijing 100035, China)

Abstract: In order to predict the heavy metal content of crops and realize the safe utilization of heavy metal contaminated agricultural land, 10 factors including soil Cd content, soil pH value, soil organic matter, cation exchange capacity, distance to mining land, and distance to traffic lines etc. which affect Cd content in crops were selected by GeoDetector model, being used as input factors. The Cd content in edible parts of crops was used as output factor. The BP neural network prediction models of Cd content in edible parts of wheat, rice, rapeseed, and vegetables were established respectively. The pollution of crop planting was predicted and analyzed. The safety zones of planting was categorized. The results illustrate that: (1) The BP neural network exhibit a good predictive effect, which has a nice applicability to the prediction of Cd content of edible parts of crops. The prediction accuracy of BP neural network models was better than that of the multiple regression prediction models according to the accuracy evaluation indexes. (2) Due to the low prediction accuracy of vegetables prediction model, the support for prediction and management partitions was limited. Only the prediction results of Cd content in edible parts of wheat, rice, and rapeseed were evaluated. The crop pollution prediction distribution map and distribution characteristics were obtained. (3) With the goal of safe use of agricultural land, the three crops in the study area were re-allocated based on the evaluation results, which was divided into four crop suitable planting safety zones. The findings of this study can provide ideas and basis for safe utilization of contaminated agricultural land and crop planting adjustment, which is of both theoretical and practical significance.

Key words: Cd content of crops; BP neural network; prediction models; heavy metal pollution; safety use of agricultural land; wheat; rice; rapeseed; vegetables; zoning of suitable planting for crops