

DOI: 10.13758/j.cnki.tr.2024.02.021

杨清, 范晓梅, 王林林, 等. 基于贝叶斯最大熵的黄河三角洲土壤含盐量空间分布预测. 土壤, 2024, 56(2): 406–414.

# 基于贝叶斯最大熵的黄河三角洲土壤含盐量空间分布预测<sup>①</sup>

杨清, 范晓梅\*, 王林林, 唐影

(南京信息工程大学地理科学学院, 南京 210044)

**摘要:** 黄河三角洲土壤盐渍化问题是制约当地农业生产和生态稳定的关键因素。为了准确掌握盐渍土的空间分布, 提高土壤含盐量的空间预测精度, 本研究根据 2022 年 5 月黄河三角洲的 193 个采样点两个深度土壤含盐量分析数据, 结合高程和 Landsat 9 遥感影像等数据, 采用地理加权回归(Geographically weighted regression, GWR)模型构造区间型软数据, 进而建立贝叶斯最大熵(Bayesian maximum entropy, BME)模型对研究区土壤含盐量的分布进行了预测, 并同传统的地统计模型普通克里格(Ordinary kriging, OK)及 GWR 模型的预测结果进行了比较。结果表明: BME 模型对土壤含盐量的预测精度高于另外两种模型。与 OK 模型相比, BME 模型的 RMSE 在土壤表层(0~30 cm)和底层(90~100 cm)分别降低 25% 和 21%,  $R^2$  分别提高 0.543 2 和 0.352 7。BME 模型作为本研究最佳土壤含盐量空间预测模型, 展现了多源数据整合及非线性估计的优势。黄河三角洲表层土壤盐渍化率(88%)高于底层(68%), 土壤含盐量大体呈现由西南向东北递增的趋势, 沿海地区大于内陆地区, 黄河三角洲北部是整个区域盐渍化最为严重的地区。

**关键词:** 贝叶斯最大熵; 黄河三角洲; 土壤含盐量; 地理加权回归

**中图分类号:** S151.9; S156.4+1 **文献标志码:** A

## Prediction of Spatial Distribution of Soil Salinity Content in Yellow River Delta Based on Bayesian Maximum Entropy Model

YANG Qing, FAN Xiaomei\*, WANG Linlin, TANG Ying

(College of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** Soil salinization in the Yellow River Delta (YRD) is a key factor restricting local agricultural productivity and ecological stability. In order to accurately grasp the spatial distribution of saline soils and improve the spatial prediction accuracy of soil salinity, based on measured soil salinity at 193 sampling points and two soil layers in the YRD in May, 2022, and data such as digital elevation model and Landsat 9 remote sensing imagery were combined, the Geographically Weighted Regression (GWR) model was used to construct interval-type soft data, and then the Bayesian Maximum Entropy (BME) model was established to predict the distribution of soil salinity in the study area, and finally the prediction results were compared with the traditional geographic statistical model Ordinary Kriging (OK) and the GWR model. The results showed that the prediction accuracy of the BME model was higher than those of the other two models. Compared with OK, the prediction errors (RMSE) of BME in the soil surface (0–30 cm) and bottom (90–100 cm) layers were decreased by 25% and 21%, respectively, and the  $R^2$  were improved by 0.543 2 and 0.352 7, respectively. As the best spatial prediction model in this study, BME showed the advantages of multi-source data integration and nonlinear estimation. The salinization rate (88%) of the surface layer in the YRD was higher than that of the bottom layer (68%). Generally, soil salinity was increased from southwest to northeast, the coastal areas were more serious than the inland areas, and soil salinization was most prominent in the northern part of the YRD.

**Key words:** Bayesian maximum entropy; The Yellow River Delta; Soil salinity content; Geographically weighted regression

土壤盐渍化是指可溶性盐在土壤表层的集聚现象, 是农业发展、生态稳定和土壤资源可持续利用的主要限制因素<sup>[1]</sup>。全球广泛且不断扩张的盐渍土使盐渍化成为

当前备受关注的环境和资源问题<sup>[2-4]</sup>。黄河三角洲作为典型的盐渍化滨海区, 由于长期的海水渗透以及相对干旱的气候条件, 盐渍化问题日益显著。此外, 人类活动如

①基金项目: 国家自然科学基金项目(41971107)资助。

\* 通讯作者(fanxm@lreis.ac.cn)

作者简介: 杨清(1997—), 男, 湖南衡阳人, 硕士研究生, 主要从事土地退化与空间预测研究。E-mail: 1503118401@qq.com

不合理的土地利用和重灌轻排的耕作方式造成严重的次生盐渍化<sup>[5]</sup>,使盐渍化问题更加复杂,造成农业生产下降、生物多样性锐减、生态环境恶化等问题<sup>[3]</sup>。

掌握土壤含盐量(Soil salinity content, SSC)的空间分布特征可为盐渍土改良治理工作提供重要依据。传统地统计模型中的普通克里格(Ordinary kriging, OK)是最常用的 SSC 插值方法。OK 通过各样本点之间 SSC 的空间依赖关系,在未知点提供最优线性无偏估计<sup>[6]</sup>。然而,SSC 由于受多种环境因素的影响而具有高空间变异特征,难以满足传统地统计学要求的数据平稳性假设,因此,SSC 样本的规模和特征较大地影响着地统计模型的应用<sup>[6-7]</sup>。受样本数据影响较小的线性回归模型可以有效地利用环境影响因素来反演 SSC 的空间分布,是近年来颇受欢迎的 SSC 插值方法。地理加权回归(Geographically weighted regression, GWR)模型利用空间变化的环境协变量系数捕捉样本数据的空间非平稳性<sup>[8-9]</sup>,在最优带宽范围内建立局部线性方程对 SSC 进行预测。然而,线性回归建模受环境协变量与 SSC 之间的相关性的影响较大,导致模型的不稳定性。

贝叶斯最大熵(Bayesian maximum entropy, BME)是 Christakos 等<sup>[10]</sup>和 He 等<sup>[11-12]</sup>于 20 世纪 90 年代提出的一套灵活而严谨的非线性时空估计理论框架。与 OK 插值相比,其可以充分利用硬数据和软数据进行空间分析与估计,且不需要传统地统计学高斯分布和二阶平稳假设等前提条件。近年来,BME 模型被广泛地应用于土壤属性的空间预测。高风杰等<sup>[13]</sup>将我国东北黑土丘陵区海沟河小流域表层有机质作为硬数据,不同土地类型作为软数据,通过 BME 模型预测有机质的空间分布,并与协同克里格方法进行对比,精度评价表明 BME 模型的整体预测精度高于协同克里格方法。刘欢瑶等<sup>[14]</sup>将亚热带红壤丘陵区有机碳密度作为硬数据,借助 GWR 模型结合地形、土地利用等多源数据获取软数据,应用 BME 结合 GWR 的耦合模型对土壤有机碳密度进行空间预测,得出耦合模型的预测精度高于 GWR 模型。Zhang 等<sup>[15]</sup>将土壤有机质作为硬数据,将多元线性回归和 GWR 模型构造的 95% 置信区间作为软数据,通过 BME 模型对湖北省沙洋县的有机质进行空间预测,得出相较于地统计学方法和线性回归模型,BME 模型具有更低的预测误差和预测不确定性。Douaik 等<sup>[16]</sup>利用 BME 模型结合区间软数据对霍尔多巴吉地区的 SSC 进行预测制图,得出 BME 模型的制图精度高于地统计模型。总之,国内外研究结果均表明,BME 模型

对土壤成分的空间预测精度要明显高于经典的地统计学方法和线性回归模型,但目前针对 SSC 方面的研究并不多见。

此前已有众多学者针对黄河三角洲盐渍土的空间分布展开了深入细致的研究,研究方法以经典的地统计方法和 GWR 方法为主,基于 BME 模型的 SSC 空间预测分析在当地尚未见报道。本研究以黄河三角洲 SSC 为研究对象,以高程、Landsat 9 波段指数及距海距离等作为协变量构造区间型软数据,进而建立 BME 模型对 SSC 的空间分布进行预测,最后对不同方法 SSC 空间分布预测结果进行对比,以验证 BME 方法的有效性。

## 1 材料与方 法

### 1.1 研究区概况

研究区位于山东省东营市(图 1)黄河入海口地区,东临渤海和莱州湾,地处 37°25'N ~ 38°10'N, 118°30'E ~ 119°20'E,面积约 3 309 km<sup>2</sup>。该区域地形平坦,地势由西南向东北倾斜,高程在 -13.5 ~ 10.2 m,区域内地下水埋深浅,易受海水的侵蚀,地下水矿化度高<sup>[17-18]</sup>。气候类型为暖温带半湿润大陆性季风气候,年降雨的 74% 集中在 6—9 月<sup>[1]</sup>,多年平均降水量约为 560 mm<sup>[18]</sup>,多年平均蒸发量约为 1 962 mm,蒸降比约为 3.5 : 1。土壤类型以轻沙壤为主,毛细管作用强烈。在海水侵蚀和季节性干旱的气候条件下,区域内原生盐渍土广泛分布。此外,由于当地对土地资源的不合理开发利用,以及重灌轻排的耕作方式,导致土地次生盐渍化和生态恶化等问题频发。

### 1.2 土样采集与测定

本研究采用的土壤调查数据来自均匀布置在黄河三角洲的 193 个土壤采样点(图 1),每个采样点包括表层(0 ~ 30 cm)与底层(90 ~ 100 cm)两个深度的 SSC 数据。采样点的间距约为 3 km,采样时间为 2022 年 5 月 23 日至 6 月 6 日,在雨季来临之前。本研究对两个土层深度的 SSC 进行研究,分析区域内土壤盐渍化的垂直剖面特征,可为准确掌握黄河三角洲土壤盐渍化特征及其发展趋势提供可靠的数据支撑。在采样过程中使用取土器采集土样并进行封装标记,利用 GPS 记录采样点的经纬度,后将采样点投影到横轴墨卡托投影的坐标系下。将土样带回实验室后进行自然风干,剔除植被根系与砂石,然后研磨过 0.5 mm 孔筛,按照 5 : 1(V : m)的水土比浸提土壤,SSC 采用烘干称重法测定,pH 采用 pH 计测定。根据 3 $\sigma$  原则将平均值  $\pm 3\sigma$  之外的 SSC 值视为异常值,并将异常值修改为正常最大或最小值,以避免异常值对插值结果的影响<sup>[19]</sup>。

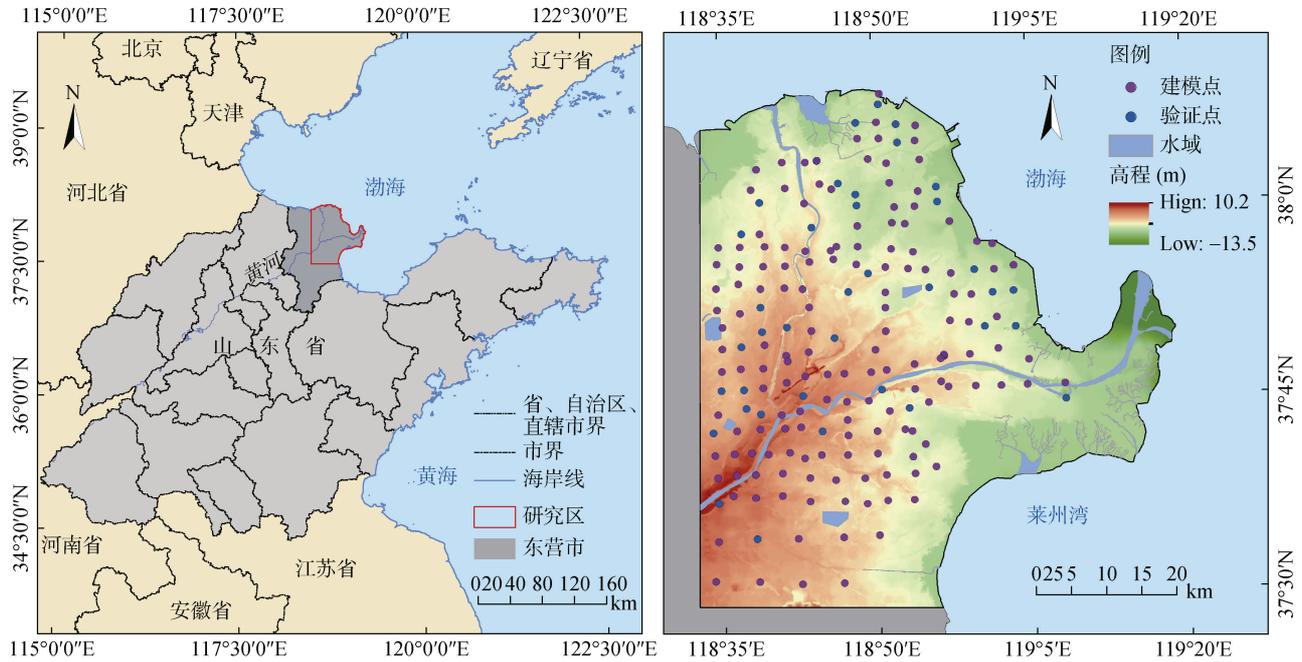


图 1 研究区位置及采样点分布图

Fig. 1 Location of the study area and distribution of sampling points

1.3 环境协变量数据获取及预处理

1.3.1 遥感数据及其他数据的获取 本研究使用的 2022 年 5 月 15 日的 Landsat 9 OLI2 遥感影像来自美国地质勘探局(https://earthexplorer.usgs.gov/), 其成像时间与土壤采样时间接近, 空间分辨率为 30 m。利用 ENVI 软件对影像进行了辐射定标、大气矫正、几何矫正和裁剪等处理。DEM(Digital elevation model) 数据为根据 1 : 10 000 地形图等高线数字化后, 在 ArcGIS 中生成, 分辨率为 10 m。

1.3.2 环境协变量的提取 大量基于遥感影像的 SSC 反演研究表明, 遥感影像的红光和近红外波段对土壤中的盐分较为敏感<sup>[20]</sup>, 且根据这些敏感波段计算出的植被指数、盐度指数和组合波段指数能有效地反

映 SSC 的空间分布<sup>[21-22]</sup>。此外, SSC 的空间格局还受外界环境和人类活动的直接影响。因此, 本研究在前人研究成果的基础上选取了 13 个环境协变量, 包括高程、距海距离(Distance to the sea, DTS)、pH、单波段(近红外(Near infrared, NIR)和短波红外(Shortwave infrared, SWIR1))、组合波段(Combined band, C2)、植被指数(归一化植被指数(Normalized difference vegetation index, NDVI)、差值植被指数(Difference vegetation index, DVI)、绿度归一化植被指数(Green normalized difference vegetation index, GNDVI)和改进土壤调节植被指数(Modified soil adjusted vegetation index, MSAVI))和盐度指数(Salinity Index, SI3、SI9、SI-T)。环境协变量的定义如表 1 所示。

表 1 环境协变量及其定义

Table 1 Environmental covariates and their definitions

协变量	定义
pH	由采样点土样经化验分析得到 pH, 后通过普通克里格方法插值到栅格表面, 分辨率为 30 m
高程	距海平面距离, 根据 1 : 10 000 地形图等高线数字化后提取, 分辨率为 10 m
NIR、SWIR1	近红外和短波红外波段, 波段范围分别为 0.85 ~ 0.88 μm 和 1.57 ~ 1.65 μm
NDVI	$NDVI = (NIR - Red) / (NIR + Red)$ , 反映植被覆盖程度
DVI	$NDVI = NIR - Red$ , 对土壤背景的变化较为敏感
GNDVI	$GNDVI = (NIR - Green) / (NIR + Green)$ , 常用于评估植被的光合活性
MSAVI	$MSAVI = 2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}$ , 消除土壤的影响, 反映植被状况
SI3、SI9	$SI3 = \sqrt{Green^2 + Red^2 + NIR^2}$ , $SI9 = (NIR - Red) / Green$ , 可用于快速盐渍化动态评估
SI-T	$SI-T = (Red / NIR) \times 100$ , 可用于快速盐渍化动态评估
DTS	距海距离, 根据海岸线通过欧式距离分析提取, 分辨率为 30 m
C2	$C2 = Green + Red + NIR$ , 绿光、红光和近红外波段的组合, 用以增强对 SSC 的敏感性

1.4 预测模型构建

1.4.1 BME 模型 1)贝叶斯理论基础。时空随机场模型是 BME 的理论基础，该模型将自然过程抽象成具有空间和时间两个参数的随机变量的集合。 $p$  是空间  $s$  和时间  $t$  定义的点，即  $p=(s, t)$ 。时空随机场  $X(p)$  的期望值  $m_x=E(X(p))$  表示时空随机场的整体趋势和时空结构性，协方差  $c_x(p, p')$  表示任意两点之间的时空相关性<sup>[10]</sup>。

BME 将所有的信息和数据分为两类，即广义知识和特定知识。广义知识主要包括协方差函数、物理法则、经验关系等；特定知识包括硬数据和软数据。BME 方法有两个重要的步骤：①使用广义知识，基于最大熵原理，计算研究区内未测点变量分布的先验概率密度函数(probability density function, pdf)，最大熵原理即在约束条件下，使一个系统的信息熵最大，引入拉格朗日乘数  $\mu_\alpha$ ，得到先验 pdf 公式如下：

$$f_G(Z_{\text{map}}) = \frac{1}{A} \exp\left(\sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(Z_{\text{map}})\right) \quad (1)$$

式中： $f_G(Z_{\text{map}})$ 表示先验 pdf； $Z_{\text{map}}=(Z_{\text{hard}}, Z_{\text{soft}}, Z_0)$ ， $Z_{\text{hard}}$ 、 $Z_{\text{soft}}$ 、 $Z_0$  分别表示硬数据点、软数据点和估计点； $g_\alpha$  是关于  $Z_{\text{map}}$  的已知函数； $N_c$  为条件总数； $A$  为归一化约束。

$$A = \int \exp\left(\sum_{\alpha=1}^{N_c} \mu_\alpha g_\alpha(Z_{\text{map}})\right) dZ_{\text{map}} \quad (2)$$

②使用特定知识，基于贝叶斯条件概率公式，更新上一步得到的先验 pdf，得到研究区内预测点变量分布的后验 pdf：

$$f_k(Z_0) = f_G(Z_0 | Z_{\text{hard}}, Z_{\text{soft}}) = \frac{\int_l^u f_G(Z_0, Z_{\text{hard}}, Z_{\text{soft}}) dZ_{\text{soft}}}{\int_l^u f_G(Z_{\text{hard}}, Z_{\text{soft}}) dZ_{\text{soft}}} \quad (3)$$

式中： $f_k(Z_0)$ 表示后验 pdf； $l$  和  $u$  分别为置信区间的下界和上界。根据后验 pdf，利用预测均值进行 SSC 制图，预测均值定义如下：

$$\hat{X}_{\text{mean}}(Z_0) = \int Z_0 f_k(Z_0) dZ_0 \quad (4)$$

式中： $\hat{X}_{\text{mean}}(Z_0)$ 表示估计点  $Z_0$  处 BME 的预测值。本研究基于 BME 的协方差分析及 SSC 估计均在 MATLAB 中的 SEKS-GUI v0.69 软件包中完成。

2) 软数据构造。本研究将 BME 协方差分析得到的协方差函数和正规化约束作为广义知识，硬数据和构造的软数据作为特定知识。本文的硬数据是指采样点获取的土样经化验分析得到的 SSC 数据，软数据

则是根据 GWR 模型的结果在研究区内 1 200 m 分辨率下 2 271 个空间格网点上计算得到的 95% 置信区间。以表层 SSC 软数据构造为例，公式如下<sup>[15]</sup>：

$$GI_{\text{GWR}}^{1-\alpha}(S_k) = \left[ \begin{array}{l} \bar{X}_{\text{GWR}} - \frac{S_{\text{GWR}}}{\sqrt{\text{NL}}} t_{(\alpha/2)}(\text{NL}-1), \\ \bar{X}_{\text{GWR}} + \frac{S_{\text{GWR}}}{\sqrt{\text{NL}}} t_{\alpha/2}(\text{NL}-1) \end{array} \right] \quad (5)$$

式中： $GI_{\text{GWR}}^{1-\alpha}(S_k)$ 表示某空间网格点  $S_k$  上的 95%置信区间； $\bar{X}_{\text{GWR}}$  和  $S_{\text{GWR}}$  分别表示在  $S_k$  上 12 个 GWR 模型的预测均值和预测方差；NL 为 GWR 模型的个数，这里为 12； $t_{(\alpha/2)}(\text{NL}-1)$ 表示自由度为 NL-1，置信水平为  $1-\alpha$  的 t 分布的临界值， $\alpha$  设置为 0.05。底层 SSC 软数据构造方式与表层相似。

1.4.2 其他插值方法 1) OK 法。OK 法是目前应用最广泛的空间预测方法，其根据半变异或半方差函数确定各样本点对预测点的影响权重，在满足平稳假设、高斯假设等前提条件下，对研究对象进行最优线性无偏估计<sup>[23]</sup>。

2) GWR 模型。GWR 模型利用环境协变量对带宽范围内的样本点进行局部线性拟合，并且利用空间变化的协变量系数捕捉 SSC 的空间非平稳性。在 GWR 中，用高斯核函数表征 SSC 随样本点距离增加而衰减的形式，本文选用自适应带宽函数作为核函数，利用校正的 Akaike 信息标准法(AICc)确定带宽参数。GWR 估计公式如下：

$$y_{\text{GWR}(u)} = \beta_{0(u)} + \sum_{i=1}^n \beta_{i(u)} X_{i(u)} + \varepsilon_{(u)} \quad (6)$$

式中： $y_{\text{GWR}(u)}$ 为模型在位置  $u$  处的预测值； $\beta_{0(u)}$ 为截距； $\beta_{i(u)}$ 为  $i$  协变量在位置  $u$  的回归系数； $X_{i(u)}$ 为位置  $u$  处  $i$  协变量的值； $n$  为协变量的个数； $\varepsilon_{(u)}$ 为预测误差。本文利用安装在 ArcGIS 10.2 中的 GTWR 插件进行 GWR 建模。

1.4.3 模型精度评价 本研究将土壤采样点 SSC 数据集划分为建模集(约占 80%)和验证集(约占 20%)，选用平均绝对误差(Mean absolute error, MAE)、均方根误差(Root mean square error, RMSE)和决定系数( $R^2$ )来验证各模型对 SSC 的估算精度，并对比模型的效果。MAE 和 RMSE 越低、 $R^2$  越高表示模型预测精度越高，反之越低。具体公式如下：

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n (|o_i - p_i|) \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - p_i)^2} \quad (8)$$

$$R^2 = \left[ \frac{\sum_{i=1}^n (o_i - \bar{o})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^n (o_i - \bar{o})^2 \sum_{i=1}^n (p_i - \bar{p})^2}} \right]^2 \quad (9)$$

式中： $o_i$ 为实测值； $p_i$ 为预测值； $\bar{o}$ 和 $\bar{p}$ 分别为实测值和预测值的平均值； $n$ 为验证集样本数。

## 2 结果与分析

### 2.1 土壤含盐量描述性统计

研究区采样点 SSC 统计特征见表 2 所示。表层 SSC 在 0.458 g/kg ~ 54.976 g/kg, 平均值为 8.628 g/kg, 变异系数为 152%, 属强空间变异(<10% 为弱变异, 10% ~ 100% 为中等变异, >100% 为强变异)。底层 SSC 在 0.459 g/kg ~ 26.599 g/kg, 平均值为 4.516 g/kg, 变异系数为 120%, 属强空间变异, 平均值和变异系数较表层分别降低 48% 和 21%。可见, 研究区 SSC 普遍较高, 具有高空间变异特征, 底层土盐渍化相对较轻, 土壤盐分表聚现象明显。

表 2 采样点土壤含盐量的描述性统计特征

Table 2 Descriptive statistics of soil salinity contents in surface and bottom layers at sampling points

深度(cm)	数据集	数量	最小值(g/kg)	最大值(g/kg)	均值(g/kg)	标准差(g/kg)	变异系数(%)	分布类型
0 ~ 30	全部样点	193	0.458	54.976	8.628	13.132	152	对数正态
	建模集	154	0.458	54.976	8.636	13.485	156	对数正态
	验证集	39	0.459	54.976	8.598	11.633	135	对数正态
90 ~ 100	全部样点	193	0.459	26.599	4.516	5.405	120	对数正态
	建模集	154	0.459	26.599	4.592	5.396	117	对数正态
	验证集	39	0.460	26.599	4.214	5.429	129	对数正态

表 3 不同土层土壤含盐量与环境协变量相关性

Table 3 Correlations between soil salinity content and covariates at different soil layers

协变量	0 ~ 30 cm		90 ~ 100 cm		
	0 ~ 30 cm	90 ~ 100 cm	协变量	0 ~ 30 cm	90 ~ 100 cm
高程	-0.41**	-0.38**	GNDVI	-0.42**	-0.31**
pH	-0.53**	-0.67**	MSAVI	-0.39**	-0.30**
DTS	-0.35**	-0.31**	SI3	-0.27**	-0.26**
NIR	-0.37**	-0.31**	SI9	-0.26**	-0.26**
SWIR1	-0.12	-0.17*	SI-T	0.39**	0.30**
NDVI	-0.43**	-0.29**	C2	-0.17*	-0.38**
DVI	-0.41**	-0.25**	表层土含盐量		0.60**

注：\*、\*\*分别表示在  $P < 0.05$ 、 $P < 0.01$  水平上显著相关。

### 2.3 不同方法的半变异函数模型

为了方便对比, 本研究借助安装在 MATLAB 中的 SEKS-GUI v0.69 软件包完成 OK 和 BME 模型的协方差分析, 相关协方差模型参数见图 2。块基比( $C_0/(C_0+C)$ )表示随机变异与系统总变异的比值, 该比值

### 2.2 环境协变量筛选和软数据构建

表土盐渍化主要由底土和地下水中盐分的迁移和积累造成<sup>[3]</sup>, 因此, 底层和表层 SSC 的空间格局存在一定的联系。本文将表层 SSC 预测结果作为一个底层 SSC 的预测因子, 以提高底层 SSC 的空间预测精度<sup>[3,24]</sup>。由表 3 中 SSC 与环境协变量的相关性可知, 底层 SSC 和表层 SSC 空间分布具有较高的正相关关系, 此外, 除 SI-T 与 SSC 呈现正相关外, 其他协变量与 SSC 均呈现显著负相关。大部分环境协变量与 SSC 在  $P < 0.01$  水平上显著相关, 但相关系数偏低(相关系数绝对值大部分在 0.3 ~ 0.7)。基于遥感光谱的植被和盐度指数(如 NDVI、DVI、GNDVI、MSAVI、SI-T)的相关性在土壤底层有所降低, 这与遥感影像对深层 SSC 的敏感性降低有关<sup>[2-3]</sup>。

本研究将显著性水平  $P < 0.05$  和方差膨胀因子  $VIF < 7.5$  作为筛选一组环境协变量的标准, 最终在土壤表层筛选 12 组协变量建立 12 个 GWR 模型, 在土壤底层筛选 4 组协变量建立 4 个 GWR 模型, 基于表层 12 个和底层 4 个 GWR 模型的结果按式(5)构造出 95% 置信区间, 分别作为表层和底层 SSC 软数据。

<0.25 表示高空间相关性, 0.25 ~ 0.75 表示中等强度空间相关性, >0.75 表示弱空间相关性。不同深度 OK 模型的块基比在 0.61 ~ 0.66, 属中等强度空间相关, 而 BME 模型的块基比均 <0.25, 属高空间相关, 说明软数据的加入增强了 SSC 样本之间的空间相关性。

图中圆圈表示经验模型，曲线表示协方差模型，两者拟合度越高表明协方差模型越好。图 2C、2D 中 BME 的协方差模型与经验模型的拟合度明显高于 OK 模型(图 2A、2B)，说明 BME 模型的可靠性更高。

2.4 模型精度评价

基于验证样本计算的各模型精度如表 4 所示。综合各精度评价指标来看，不同深度 SSC 的 GWR 和 BME 模型的预测精度较 OK 模型均有显著提高，其中 BME 模型的预测精度最高。对于表层土壤，BME

和 GWR 模型对 SSC 估算的 RMSE 与 OK 模型相比分别降低了 25% 和 14%， $R^2$  从 0.069 7 分别提高到 0.612 9 和 0.305 1。对于底层土壤，BME 和 GWR 模型对 SSC 估算的 RMSE 与 OK 模型相比分别降低了 21% 和 18%， $R^2$  从 0.343 2 分别提高到 0.695 9 和 0.537 4。表明 BME 模型预测值与实测值的拟合程度显著高于其他方法。证明利用 BME 方法，结合与 SSC 显著相关的辅助变量，能有效地提高 SSC 的空间预测精度，较好地反映 SSC 的空间变异特征。

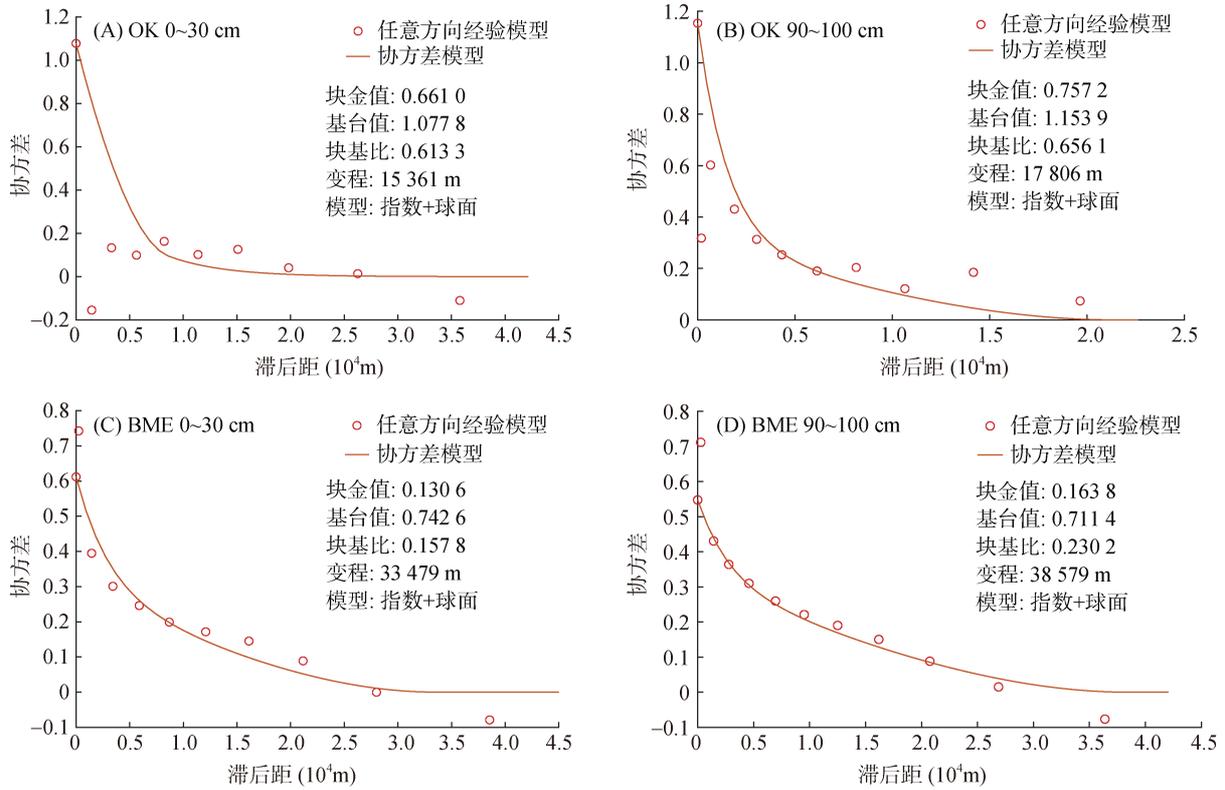


图 2 不同深度 OK 和 BME 的协方差模型参数  
Fig. 2 Covariance model parameters of OK and BME models for different soil layers

表 4 模型精度评价  
Table 4 Accuracy assessment of different models

深度 (cm)	模型	MAE (g/kg)	RMSE (g/kg)	$R^2$
0 ~ 30	OK	6.867	11.896	0.069 7
	GWR	5.558	10.276	0.305 1
	BME	4.625	8.919	0.612 9
90 ~ 100	OK	2.641	4.684	0.343 2
	GWR	2.101	3.864	0.537 4
	BME	1.922	3.703	0.695 9

2.5 土壤含盐量空间分布特征

根据中国半湿润环境气候区 SSC 分级标准将盐渍化分为 5 个等级<sup>[25]</sup>：非盐渍化(0 ~ 2.0 g/kg)、轻度盐渍化(2.0 ~ 4.0 g/kg)、中度盐渍化(4.0 ~ 6.0 g/kg)、

重度盐渍化(6.0 ~ 10.0 g/kg)和盐土(>10.0 g/kg)。对 3 种模型得到的 SSC 空间分布进行分级，得到 SSC 空间分布如图 3 所示。不同方法预测的 SSC 空间分布特征基本一致，即沿海地区大于内陆地区，东北地区

大于西南地区,黄河沿岸 SSC 较低但存在多个局部土壤盐渍化区域,北部的老黄河入海口 SSC 明显大于现黄河入海口,大体呈现由西南向东北递增的趋势。基于 OK 模型预测的 SSC 平滑连续(图 3A、3D),盐渍土沿带状连续分布,但反映不出 SSC 分布差异的细节;基于 GWR 模型的预测图中(图 3B、3E),存在界限分明且 SSC 较高的条块状区域,这是由于各

辅助变量来自遥感图像数据,SSC 分布图受遥感光谱特征的影响,与自然条件下土壤盐分连续渐变的特征不符。基于 BME 方法的预测图(图 3C、3F)克服了 OK 插值的平滑效应,同时又结合了基于 GWR 模型构造的区间型软数据,得到的 SSC 空间分布特点更为合理,能够较为准确地反映黄河三角洲 SSC 的空间分布。

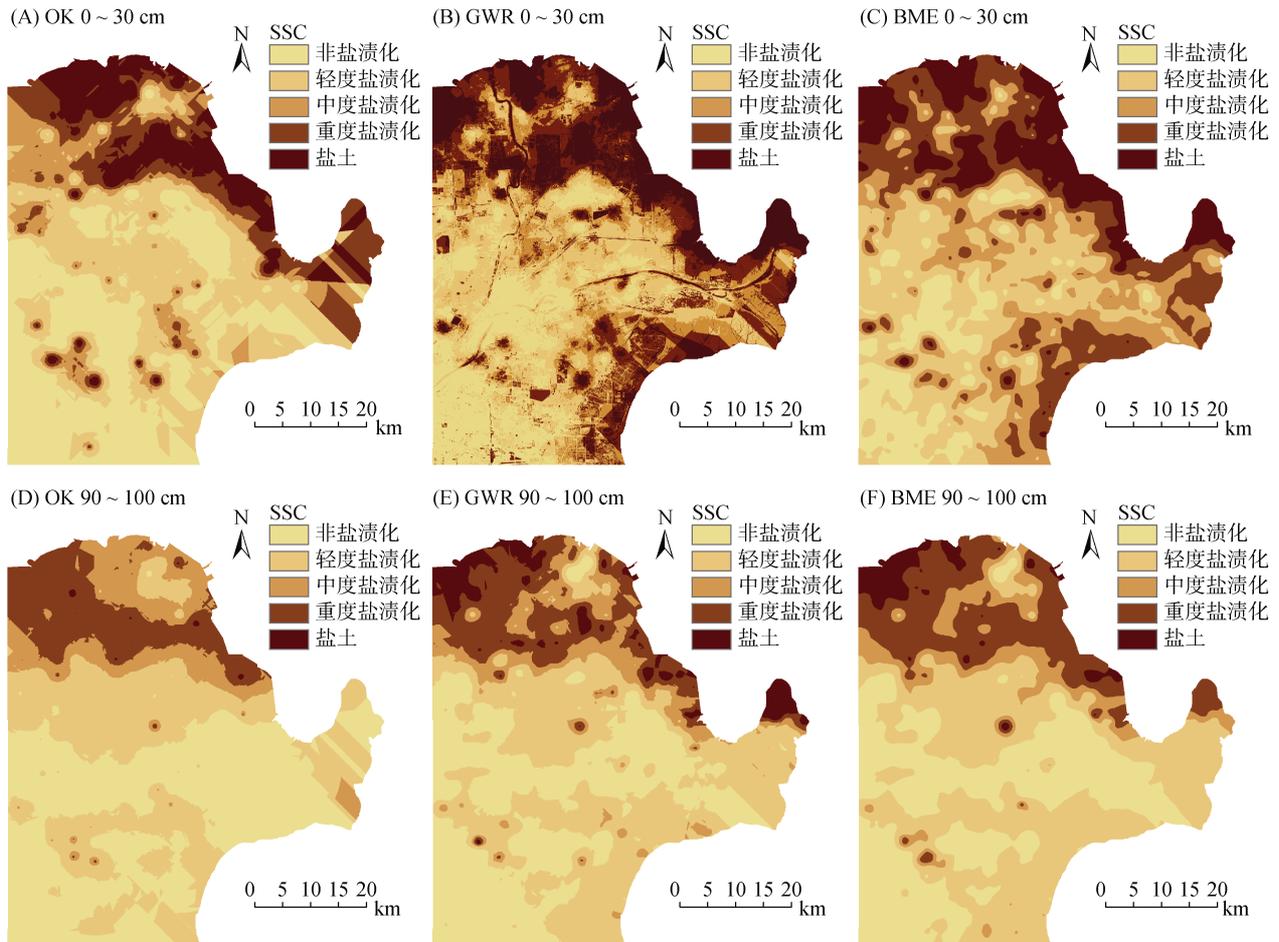


图 3 不同深度不同模型土壤含盐量的空间分布

Fig. 3 Spatial distribution of soil salinity content of different models at different soil layers

根据 BME 插值结果对研究区不同等级盐渍土面积进行统计,结果见图 4。结合图 3 和 4 可知,黄河三角洲表层土壤约有 88% 为不同程度的盐渍化土壤,其中重度盐渍化及以上的盐渍土约占 37%,主要分布在北部黄河故道和孤北水库一带,以及东北沿渤海湾和东南沿莱州湾的滨海区域,在内陆也有一些零散分布。黄河沿岸两边的灌区由于排灌条件好,加上多年的人工改造和耕作历史,表层 SSC 相对较低。底层土壤中盐渍化土壤的面积占比约为 68%,并且以轻度盐渍化土壤为主(约占 38.5%)。底层土壤中的重度盐渍化及以上的土壤主要分布在黄河三角洲北部,

而东南部莱州湾的盐渍化程度较低。可见,黄河三角洲北部的盐渍土属于上下均匀型,而南部的盐渍土属于盐分表聚型。

### 3 讨论

#### 3.1 环境协变量选用

本研究选用了 13 个环境协变量包括基于遥感光谱的指数、高程、距海距离和 pH 等指标进行 GWR 建模及软数据构造。相关研究表明,遥感影像的 SSC 敏感波段(红外和近红外波段)以及基于这些敏感波段计算的植被和盐度指数可用于 SSC 的快速空间反

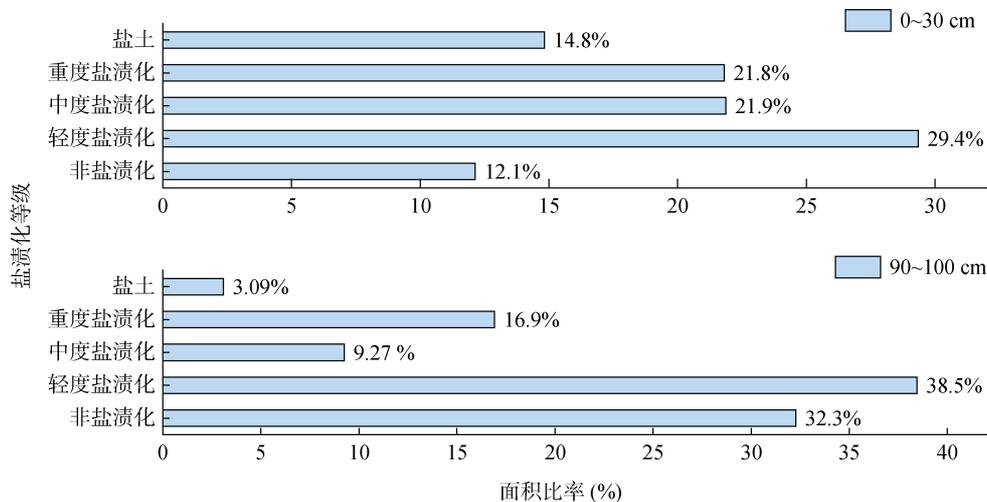


图 4 不同土层不同类型盐渍土面积统计

Fig. 4 Area statistics of soils under different degrees of salinization at different soil layers

演<sup>[20-22]</sup>,但影像中不同地物之间的光谱连续特征导致较大的预测偏差,造成 GWR 模型得出了非真实的 SSC 空间分布的问题。此外,遥感影像只记录了表土信息,导致遥感指数与底层 SSC 的相关性较表层有所降低,以预测精度较高的表层 SSC 空间分布作为中间变量可以提高其他土壤深度 SSC 的空间分布预测精度<sup>[3,24]</sup>。Li 等<sup>[3]</sup>以表层土壤盐度为中间变量,提高了多深度土壤盐度的预测精度,弥补了遥感光谱对深层盐度不敏感的缺陷; Taghizadeh-Mehrjardi 等<sup>[24]</sup>以两个深度实测盐度为中间变量,结合环境协变量对其他深度盐度进行预测,与不加盐度中间变量相比,预测精度有明显提高。基于地形的环境协变量如高程、坡向、坡度和地表粗糙度等是反映土壤盐分运移规律的有效指标<sup>[2,26]</sup>,由于黄河三角洲地势平缓,地形比降在 1/8 000 ~ 1/12 000,因此,除高程以外,地形因子中的其他指标与 SSC 的相关关系并不显著。水的运动控制着盐分的运移、集聚和分布,距海距离是表征盐分来源的重要指标,与 SSC 存在显著的相关关系,在滨海低地区域,地下水埋深浅,矿化度高,盐分表聚现象明显,盐渍化问题突出<sup>[27]</sup>。

### 3.2 贝叶斯最大熵模型预测效果

根据本研究中各模型的精度评价结果, BME 模型对 SSC 的预测精度有显著的提升,但与相关研究相比其  $R^2$  仍然较低。主要原因如下: ①提取的环境协变量与 SSC 的相关性偏低。环境协变量与 SSC 的相关性影响软数据的质量,进而对 BME 的预测精度产生影响<sup>[15]</sup>,导致 BME 模型的拟合优度偏低。相关 SSC 预测研究表明,协变量与 SSC 的相关系数一般不超过 0.7,这在一定程度上限制了模型预测精度的提高。Li 等<sup>[3]</sup>利用 Landsat 遥感波段数据提取的植被

和盐度指数与 SSC 的相关性在 0.5 左右;吴春生等<sup>[5]</sup>提取了与 SSC 相关系数绝对值分别为 0.379、0.499 和 0.045 的高程、NDVI 和河流距离作为协变量,导致建立的 GWR 模型的拟合优度偏低; Peng 等<sup>[21]</sup>提取了大量遥感指数和地形指数作为协变量,这些指数与盐度的相关系数绝对值均小于 0.65。此外, BME 可以接受多种形式的软数据,软数据的获取不应局限于单一数据源,利用历史实测盐度数据、土壤电导率和地下水电导率等构造软数据也可能使 BME 达到理想的预测效果。②研究区和研究对象的特殊性。由于受河流、海洋、气候和人类活动等的影响,黄河三角洲土壤盐渍化具有明显的空间和季节变化特征,此外,受地表景观破碎化等因素的影响,其 SSC 具高度变异性和非连续性的特点。吴春生等<sup>[5]</sup>将高程、NDVI 和河流距离作为协变量,应用 GWR 模型对黄河三角洲 SSC 进行空间预测,得到 GWR 模型的  $R^2$  为 0.421,在本研究 GWR 模型的  $R^2$ (0.3 ~ 0.54)范围内,说明环境协变量的线性回归对 SSC 的解释度较低,黄河三角洲 SSC 具有特殊性。

## 4 结论

1) SSC 和环境协变量的相关性分析结果表明,大部分提取的环境协变量与 SSC 在  $P < 0.01$  水平上显著相关,但相关系数普遍偏低。除盐度指数 SI-T 与 SSC 呈显著正相关外,其他环境协变量与 SSC 均呈显著负相关。基于遥感光谱的植被和盐度指数与 SSC 的相关性随着土层深度的增加而有所降低。

2) 综合各精度评价指标, GWR 和 BME 模型的预测精度较 OK 模型均有显著提高,其中 BME 模型是最佳 SSC 预测模型。对于土壤表层, BME 模型对

SSC 估算的 RMSE 与 OK 和 GWR 模型相比分别降低了 25% 和 13%,  $R^2$  分别从 0.069 7 和 0.305 1 提高到 0.612 9。对于底层土壤, BME 模型对 SSC 估算的 RMSE 与 OK 和 GWR 模型相比分别降低了 21% 和 4.2%,  $R^2$  分别从 0.343 2 和 0.537 4 提高到 0.695 9。从 SSC 空间分布看, BME 模型充分展现了多源数据整合以及非线性估计的优势, 能较好地反映 SSC 的空间变异特征, 提高空间预测精度, 较准确地反映了黄河三角洲 SSC 的空间分布特点。

3) 黄河三角洲土壤普遍存在盐渍化现象, SSC 总体呈现由西南向东北递增的趋势。土壤表层重度盐渍化及以上的盐渍土约占 37%, 主要分布在东北沿渤海湾和东南沿莱州湾区域, 以及一些内陆零散区域, 空间变异特征明显。土壤底层重度盐渍化及以上的盐渍土约占 20%, 集中分布在研究区北部。

#### 参考文献:

- [1] Zhu W, Yang J S, Yao R J, et al. Soil water-salt control and yield improvement under the effect of compound control in saline soil of the Yellow River Delta, China[J]. *Agricultural Water Management*, 2022, 263: 107455.
- [2] Taghizadeh-Mehrjardi R, Schmidt K, Toomanian N, et al. Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models[J]. *Geoderma*, 2021, 383: 114793.
- [3] Li Y S, Chang C Y, Wang Z R, et al. Remote sensing prediction and characteristic analysis of cultivated land salinization in different seasons and multiple soil layers in the coastal area[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 111: 102838.
- [4] 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.
- [5] 吴春生, 黄翀, 刘高焕, 等. 黄河三角洲土壤含盐量空间预测方法研究[J]. *资源科学*, 2016, 38(4): 704-713.
- [6] 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.
- [7] 孙越琦, 孙笑梅, 巫振富, 等. 样点规模与采样方法对表层土壤 pH 空间预测精度的影响[J/OL]. *土壤学报*, 2022: 1-17. (2022-10-11). <https://kns.cnki.net/kcms/detail/32.1119.P.20221009.1741.002.html>.
- [8] Zhang S Y, Chen Y Y, Zhang Z Y, et al. VNIR estimation of heavy metals concentrations in suburban soil with multi-scale geographically weighted regression[J]. *CATENA*, 2022, 219: 106585.
- [9] Wang D, Li X X, Zou D F, et al. Modeling soil organic carbon spatial distribution for a complex terrain based on geographically weighted regression in the eastern Qinghai-Tibetan Plateau[J]. *CATENA*, 2020, 187: 104399.
- [10] Christakos G, Serre M L. BME analysis of spatiotemporal particulate matter distributions in North Carolina[J]. *Atmospheric Environment*, 2000, 34(20): 3393-3406.
- [11] He J Y, Christakos G, Cazelles B, et al. Spatiotemporal variation of the association between sea surface temperature and chlorophyll in global ocean during 2002-2019 based on a novel WCA-BME approach[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 105: 102620.
- [12] He J Y, Christakos G, Wu J P, et al. Spatiotemporal BME characterization and mapping of sea surface chlorophyll in Chesapeake Bay (USA) using auxiliary sea surface temperature data[J]. *Science of the Total Environment*, 2021, 794: 148670.
- [13] 高凤杰, 吴啸, 师华定, 等. 基于贝叶斯最大熵的黑土区小流域土壤有机质空间分布预测[J]. *环境科学研究*, 2019, 32(8): 1365-1373.
- [14] 刘欢瑶, 孟岑, 邹冬生, 等. 基于 BME-GWR 法的景观单元土壤有机碳密度空间预测[J]. *广东农业科学*, 2021, 48(2): 75-83.
- [15] Zhang C T, Yang Y. Can the spatial prediction of soil organic matter be improved by incorporating multiple regression confidence intervals as soft data into BME method?[J]. *CATENA*, 2019, 178: 322-334.
- [16] Douaik A, Van Meirvenne M, Tóth T. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data[J]. *Geoderma*, 2005, 128(3/4): 234-248.
- [17] Xia J B, Ren J Y, Zhang S Y, et al. Forest and grass composite patterns improve the soil quality in the coastal saline-alkali land of the Yellow River Delta, China[J]. *Geoderma*, 2019, 349: 25-35.
- [18] 胡鑫隆, 齐继, 于千钧, 等. 黄河三角洲不同长势柽柳对土壤水盐异质性的影响[J]. *土壤*, 2022, 54(6): 1291-1299.
- [19] 王幼奇, 张兴, 赵云鹏, 等. 基于 GIS 和地理加权回归的砂田土壤阳离子交换量空间预测[J]. *土壤*, 2020, 52(2): 421-426.
- [20] Allbed A, Kumar L, Aldakheel Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region[J]. *Geoderma*, 2014, 230/231: 1-8.
- [21] Peng J, Biswas A, Jiang Q S, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. *Geoderma*, 2019, 337: 1309-1319.
- [22] Masoud A A, Koike K, Atwia M G, et al. Mapping soil salinity using spectral mixture analysis of landsat 8 OLI images to identify factors influencing salinization in an arid region[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2019, 83: 101944.
- [23] 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.
- [24] Taghizadeh-Mehrjardi R, Minasny B, Sarmadian F, et al. Digital mapping of soil salinity in Ardakan region, central Iran[J]. *Geoderma*, 2014, 213: 15-28.
- [25] 鲍士旦. 土壤农化分析[M]. 3 版. 北京: 中国农业出版社, 2000.
- [26] Wei Y, Ding J L, Yang S T, et al. Soil salinity prediction based on scale-dependent relationships with environmental variables by discrete wavelet transform in the Tarim Basin[J]. *CATENA*, 2021, 196: 104939.
- [27] Song C Y, Ren H X, Huang C. Estimating soil salinity in the Yellow River Delta, eastern China—An integrated approach using spectral and terrain indices with the generalized additive model[J]. *Pedosphere*, 2016, 26(5): 626-635.