

# 基于 GIS 和随机森林算法的宁东土壤 饱和导水率分布与预测

夏子书<sup>1,2</sup>, 白一茹<sup>1,2</sup>, 王幼奇<sup>1,2</sup>, 包维斌<sup>1,2</sup>, 杨帆<sup>1,2</sup>, 钟艳霞<sup>1</sup>, 李鸣骥<sup>1</sup>

(1.宁夏大学资源环境学院, 银川 750021; 2.旱区特色资源与环境治理教育部国际合作联合实验室, 银川 750021)

**摘要:** 为探明宁东土壤饱和导水率( $K_s$ )的空间分布特征,在宁东采集 136 个原状土,采用经典统计和地统计方法分析土壤  $K_s$  的空间结构特征,并以地形因子、土壤属性等作为辅助变量,运用随机森林法(RF)、普通克里格法(OK)和逐步回归克里格法(RK)对区域土壤  $K_s$  进行预测并对 3 种方法的预测结果进行精度评价。结果表明: $K_s$  介于 0.05~7.13 mm/min,平均值为 1.46 mm/min,变异系数为 106.86%; $K_s$  与容重、孔隙度、高程、坡度、坡向、平面曲率和剖面曲率在不同滞后距离下具有自相关关系和交互相关关系;土壤  $K_s$  块金值为 38,表明随机因素引起的土壤  $K_s$  变异性较大,空间异质比为 15.32%,在空间上呈现强变异性;RF 法的预测精度最高,其平均相对误差(MRE)和均方根误差(RMSE)绝对值均为最小,相比 OK 和 RK 方法预测精度分别提高了 5.53%和 2.49%,且对局部细节的描述更准确、模拟效果最佳。RF 法可以较为准确的预测宁东土壤  $K_s$ ,为了解研究区土壤水文过程及林草植被建设提供数据参考。

**关键词:** 饱和导水率; 空间预测; 随机森林模型; 地形因子; 宁东

中图分类号:S152.7

文献标识码:A

文章编号:1009-2242(2021)01-0285-09

DOI:10.13870/j.cnki.stbcxb.2021.01.041

## Distribution and Prediction of Soil Saturated Hydraulic Conductivity in Ningdong Based on GIS and Random Forest Algorithm

XIA Zishu<sup>1,2</sup>, BAI Yiru<sup>1,2</sup>, WANG Youqi<sup>1,2</sup>, BAO Weibin<sup>1,2</sup>,  
YANG Fan<sup>1,2</sup>, ZHONG Yanxia<sup>1</sup>, LI Mingji<sup>1</sup>

(1.College of Resources and Environment, Ningxia University, Yinchuan 750021; 2.Joint Laboratory of International Cooperation for Characteristic Resources and Environmental Governance in Arid Area, Ministry of Education, Yinchuan 750021)

**Abstract:** In order to find out the spatial distribution characteristics of soil saturated hydraulic conductivity ( $K_s$ ), 136 undisturbed soils were collected in Ningdong. The spatial structure characteristics of  $K_s$  were analyzed by classical and geostatistical methods. Taking terrain factors and soil properties as auxiliary variables, the regional soil  $K_s$  were predicted by the random forest method (RF), ordinary Kriging method (OK) and stepwise regression Kriging method (RK), and the accuracy of the prediction results of the three methods were evaluated. The results showed that  $K_s$  ranged from 0.05 to 7.13 mm/min, with an average value of 1.46 mm/min and a coefficient of variation of 106.86%.  $K_s$  had autocorrelation and cross-correlation with bulk density, porosity, elevation, slope, aspect, plane curvature and section curvature under different lag distance. The nugget value of soil  $K_s$  was 38, indicating that the variability of soil  $K_s$  caused by random factor was large, and the spatial heterogeneity ratio was 15.32%, showing strong spatial variability. The prediction accuracy of RF was the highest, and the absolute values of mean relative error (MRE) and root mean square error (RMSE) were both the smallest. Compared with OK and RK, the prediction accuracy of RF was improved by 5.53% and 2.49%, respectively, and the description of local details was more accurate and the simulation effect was the best. RF could accurately predict soil  $K_s$  in Ningdong, and provide data reference for understanding soil hydrological process and forest and grass vegetation construction in the study area.

**Keywords:** soil saturated hydraulic conductivity; spatial prediction; random forest model; terrain factor; Ningdong

收稿日期:2020-07-19

资助项目:国家自然科学基金项目(41867003,41761049);宁夏自然科学基金项目(2018AAC03027);宁夏青年科技人才托举工程项目(2016008);宁夏重点研发计划重大项目(2018BFG02016);宁夏环境保护科学技术研究项目(2018-07)

第一作者:夏子书(1997-),女,硕士研究生,主要从事土壤水分生态、土壤性质空间变异研究。E-mail:xzs0131@163.com

通信作者:王幼奇(1980-),男,博士,副教授,主要从事水土资源与环境研究。E-mail:wyo0563@163.com

土壤饱和导水率(saturated hydraulic conductivity,  $K_s$ )作为不可或缺的土壤水力学参数之一,影响土壤水分、溶质运移及降雨入渗、径流分配<sup>[1]</sup>。一般而言,在土壤容重、孔隙条件和生物过程等众多因素的影响下, $K_s$ 表现出强烈的空间异质性<sup>[2]</sup>,特别在西北旱区  $K_s$  大小直接影响降水入渗、土壤水分运移过程,因此西北地区林草植被恢复、生态环境修复等过程都必须考虑  $K_s$  的空间分布特征<sup>[3]</sup>。宁东是大型煤化工基地,在煤炭开采过程中导致地表植被破坏严重,加之该区域地处西北旱区自然环境脆弱,导致了宁东整体生态系统功能退化严重<sup>[4]</sup>。在我国生态环境建设的大背景下,区域植被恢复和生态重建是目前宁东首要任务之一<sup>[5]</sup>。因此,分析和预测宁东土壤  $K_s$  空间分布特征,有利于充分利用研究区宝贵的水土资源,为该区域林草植被建设提供帮助,能更好地协调宁东经济社会发展与生态建设间的关系。

目前,一些学者使用不同方法对不同区域  $K_s$  的空间变化规律进行了分析和预测。在石羊河流域,李涛等<sup>[6]</sup>运用线性回归方法较好地分析和预测了研究区表层土壤  $K_s$  的空间分布特征;在黄土高原北部典型坡面,赵春雷等<sup>[7]</sup>应用状态空间方程对土壤  $K_s$  的空间分布进行了高精度地模拟;在新疆地区,李祥东等<sup>[8]</sup>发现,传递函数方程可以较好地分析和预测土壤水力学参数(土壤  $K_s$ 、田间持水量和永久萎蔫系数)的空间分布特征;在桂西北喀斯特地区,张川等<sup>[9]</sup>使用克里格插值法研究了典型灌丛坡地表层  $K_s$  的分异特征。可以看出在不同区域学者们利用不同方法和手段对土壤  $K_s$  的空间结构和特征进行分析和预测,并取得了较好的效果。随着计算机技术和机器学习算法的成熟,随机森林(random forest, RF)等方法成为近年来的研究热点,由于其强大的自学习能力等优势被迅速应用到土壤属性预测和制图领域,并取得良好效果<sup>[10-11]</sup>。与大多数统计建模方法相比 RF 模型具有较大优势,它对噪声及异常值的容忍性较强、不易出现过度拟合的情况,能在评价指标重要性的基础上准确刻画变量间的非线性关系<sup>[12]</sup>。如相关学者利用 RF 模型精确预测了区域土壤有机质<sup>[10]</sup>、全氮<sup>[13]</sup>、容重<sup>[14]</sup>等,其效果明显优于线性回归和传递函数方程等传统方法。目前在宁东关于土壤  $K_s$  研究较少,特别是利用 RF 方法模拟其空间分布特征尤为缺乏。因此在分析宁东土壤  $K_s$  空间分异特征的基础上,以土壤容重、孔隙度、高程、坡度和曲率等作为辅助变量,运用 RF 模型预测研究区土壤  $K_s$  空间分布特征,对了解区域土壤水文过程和林草植被恢复具有重要作用,并为宁东工业园区今后转型升级提供参考和帮助。

## 1 材料与方法

### 1.1 研究区概况

宁东地处宁、陕、甘、蒙毗邻地区(106°20′—106°52′E, 37°50′—38°23′N),位于毛乌素沙地西南缘,地形平缓开阔,平均海拔 1 424 m;气候属于中温带干旱气候,年均降水量为 200 mm 左右,年均蒸发量 1 800 mm 左右,日照时间长,年均气温为 7.8 °C,年均风速为 2.6 m/s;土壤以风沙土为主伴有少许盐碱土,土层较薄;植被以荒漠草原为主,植被盖度为 30%左右,天然植被分布稀少且不均匀,主要以一年生或多年生、早生或超早生灌木、半灌木或草本植物为主,是宁夏主要的沙生植物柠条、沙蒿、牛枝子、酸枣、沙冬青、花棒、怪柳、白刺、甘草、猫头刺、蒺藜、黄蒿、骆驼蒿等集中分布区<sup>[15]</sup>。随着“一带一路”战略的提出,宁东依托其内部丰富的矿产资源逐渐发展成为宁夏的经济动脉<sup>[16]</sup>,而近年来生态环境的破坏使“三北”防护林建设、退耕还林还草等建设工程被纳入宁东生态建设规划并给予政策扶持<sup>[5]</sup>。

### 1.2 样点布设及土样采集

2019 年 6 月,经过前期调研并结合宁东地形分布和自然地理条件,以野外采样、实地考察为研究手段,依据典型性和代表性原则随机选择代表性样点进行取样,利用手持 GPS 记录每个样点的经纬度。每个采样点利用环刀(体积 100 cm<sup>3</sup>)采集表层(0—5 cm)原状土壤样品,在环刀顶面和底面放置滤纸,防止土壤颗粒流失,原状土样品带回实验室分析处理。采集土壤样品共 136 个,其中 103 个用于方程拟合,33 个用于方程验证。取样点位置分布见图 1。

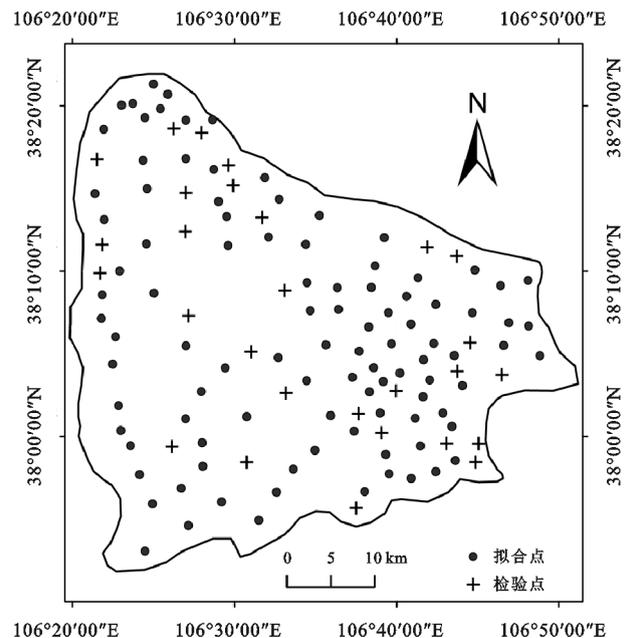


图 1 研究区土壤样点分布

### 1.3 测定项目及方法

土壤  $K_s$  用定水头法测定,并按照质量守恒定律和达西定律得出  $K_s$  的计算公式<sup>[17]</sup>为:

$$K = 10Q \times L / A \times \Delta H \times t \quad (1)$$

式中: $K$  为不同温度下的饱和导水率(mm/min); $Q$  为渗透量(mL); $L$  为土层厚度(cm); $\Delta H$  为渗流路径的总水头差(cm); $t$  为渗透时间(min); $A$  为水流经过的横截面积(cm<sup>2</sup>)。

为了便于比较不同温度下的  $K$  值,将其转换为 10 °C 时的饱和导水率:

$$K_s = K / (0.7 + 0.03t) \quad (2)$$

式中: $K_s$  为温度 10 °C 时的饱和导水率(mm/min); $t$  为测定时水的温度(°C)。

用环刀法测定原状土壤容重( $Bd$ )、总孔隙度、毛管持水量、毛管孔隙度和通气孔隙度<sup>[18]</sup>。

### 1.4 地形因子的提取

数字高程模型(digital elevation model, DEM)采用空间分辨率为 30 m 的 ASTER GDEM 数据,从地理空间数据云平台下载,地形因子主要包括高程(elevation, Ele)、坡向(aspect, Asp)、坡度(slope, Slo)、剖面曲率(profile curvature, Pc)、平面曲率(horizontal curvature, Hc),由 DEM 数据在 ArcGIS 10.3 空间分析模块提取。其中高程由 Spatial Analysis Tools 直接提取,坡度、坡向、平面曲率及剖面曲率等由 3D Analysis 工具计算后提取。高程介于 982~3 205 m,平均约为 1 424 m;坡度、坡向通过重分类命令分为平缓坡(-1)、阴坡(0°~45°,315°~360°)、半阳坡(45°~135°)、阳坡(135°~275°)、半阴坡(275°~315°)。

### 1.5 分析方法

#### 1.5.1 地统计及普通克里格(ordinary kriging, OK)

半变异函数是地统计学研究局部变量空间变化的重要工具,也是克里格空间插值的输入参数。其表达式为:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i)Z(x_i+h)]^2 \quad (3)$$

式中: $\gamma(h)$  为半变异函数; $h$ ,  $N(h)$  为间隔距离和配对数; $Z(x_i)$  和  $Z(x_i+h)$  分别是在空间位置  $x_i$  和  $x_i+h$  上的观测值( $i=1,2,\dots,N(h)$ )。

OK 法是基于结构分析及半变异函数理论的地统计学内容之一,根据各采样点数值间的线性组合,对已知采样点数据进行最优、连续的区域化变量评价<sup>[19]</sup>。

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (4)$$

式中: $Z^*(x_0)$  表示待估点  $x_0$  的值; $Z(x_i)$  表示第  $i$  个有效观测值( $i=1,2,3,\dots,n$ ); $\lambda_i$  是通过半变异函数生成的权重且  $\sum \lambda_i = 1$ 。

1.5.2 回归克里格(regression kriging, RK) RK 法中回归分析采用多元线性逐步回归,是以普通最小二乘法为基础,利用给定的多个解释变量数据研究因变量与多个自变量之间的线性关系,表示为:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_n x_{ni} \quad (5)$$

式中: $x_i$  ( $i=1,2,3,\dots,n$ ) 为自变量; $y_i$  ( $i=0,1,2,\dots,n$ ) 和  $\beta_j$  ( $j=0,1,2,\dots,n$ ) 分别为因变量和自变量的回归系数。

然后分离空间采样点上的趋势项,用实际值减去趋势项来确定其残差,对残差进行线性无偏估计处理,最后通过将回归分析的趋势项与残差插值结果相加,得到目标变量的空间估计值<sup>[20]</sup>,可表示为:

$$z(s) = m(s) + \epsilon(s) \quad (6)$$

式中: $z(s)$  为目标变量在  $s$  处的预测值; $m(s)$  为使用线性逐步回归得到的趋势项; $\epsilon(s)$  为使用 OK 插值得到的残差项。

1.5.3 RF 模型 RF 模型是一种集成学习方法,属于机器学习的一部分<sup>[21]</sup>。它集成了多个 CART 决策树,从原始样本中运用 bootstrap 重抽样法随机选择多个样本,建立决策树模型。RF 模型建立的过程中涉及到 2 个关键的参数:mtry 和 ntree。mtry 为每次重建决策树模型时的节点分裂数;ntree 为决策树数目,即使用 bootstrap 重抽样法选择样本的次數。这 2 个关键参数的确定基于 RF 模型操作中产生的实际误差大小。RF 的分类结果是由每棵分类决策树的输出结果进行投票来决定的<sup>[22]</sup>,而在回归预测中,预测值为所有回归树输出结果的平均值,表达式为:

$$\bar{h}(x) = \left( \frac{1}{k} \right) \sum_{i=1}^k h(X; \theta_i) \quad (7)$$

式中: $\bar{h}(x)$  为预测值; $\theta_i$  是一个独立分布的随机向量,它能够决定决策树的生长; $X$  为输入矩阵; $h(X; \theta_i)$  为第  $i$  棵回归树的输出结果; $k$  是回归树的数量。

### 1.6 精度评价

随机生成的 103 个用于拟合的点参与所有分析过程,其余 33 个验证点用于验证空间预测结果。用验证点估计值和实测值的平均相对误差(MRE)和均方根误差(RMSE)来评价不同方法对研究区土壤  $K_s$  预测的准确性<sup>[14]</sup>。

$$MRE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

式中: $\hat{y}_i$  是在  $i$  处土壤  $K_s$  的估计值; $y_i$  是在  $i$  处土壤  $K_s$  的实测值; $n$  为验证点个数,在本文中  $n=33$ 。当 MRE 和 RMSE 接近于 0 时,插值精度较高。

以 RMSE 为指标,比较不同插值方法与参考方法的相对精度提高值( $RI$ ),计算公式为:

$$RI = \frac{RMSE_R - RMSE_K}{RMSE_R} \times 100\% \quad (10)$$

式中:  $RMSE_R$  为参照方法的 RMSE,  $RMSE_K$  为其余方法的 RMSE。

### 1.7 数据处理

土壤  $K_s$  的计算采用 Microsoft Excel 2010 软件,各变量描述性统计、多元线性逐步回归及自相关、交互相关分析由 SPSS 22.0 软件进行,变异函数模型采用 GS+ 9.0 软件进行拟合,RF 模型由 R 3.5.3 软件的 RandomForest 包进行,采用 ArcGIS 10.3 软件对土壤  $K_s$  进行克里格插值。

## 2 结果与分析

### 2.1 土壤各变量描述性统计

宁东土壤  $K_s$ 、 $Bd$ 、孔隙度、毛管持水量、毛管孔隙度和通气孔隙度的基本统计特征见表 1。土壤  $K_s$  介于

0.05~7.13 mm/min,变化范围较大,均值为 1.46 mm/min,高于黄土高原地区均值 0.53 mm/min<sup>[23]</sup>,这与本地区较高的砂粒含量和较低的黏粒含量有关<sup>[24]</sup>。研究区内土壤  $K_s$  的变异系数为 106.86%,按照 Warrick 等<sup>[25]</sup>对土壤性质变异水平的分类,属于强变异性,表明其空间分布极不均匀。 $K-S$  正态检验的结果表明,土壤  $K_s$  数据不符合正态分布,采用自然对数转换后符合正态分布( $p=0.97>0.05$ )。土壤  $Bd$  变化范围介于 0.89~1.59 g/cm<sup>3</sup>,均值为 1.31 g/cm<sup>3</sup>,略低于黄土高原地区的均值 1.35 g/cm<sup>3</sup><sup>[26]</sup>,变异程度较弱,变异系数为 9.16%。土壤孔隙度介于 32.45%~66.51%,变异系数为 10.38%,属于正态分布。土壤毛管持水量变化范围为 0.01%~13.74%,变异系数为 169.09%,属于对数正态分布。毛管孔隙度介于 0.01%~18.33%,变异系数为 170.78%,属于对数正态分布。通气孔隙度变化范围为 12.79%~65.17%,变异系数为 15.26%,属于正态分布。

表 1 各变量基本统计信息

指标	最大值	最小值	平均值	标准差	变异系数/%	峰度	偏度	$K-S$	分布类型
$K_s/(mm \cdot min^{-1})$	7.13	0.05	1.46	1.56	106.86	2.56	1.74	0.97	对数正态
$Bd/(g \cdot cm^{-3})$	1.59	0.89	1.31	0.12	9.16	3.98	-0.99	0.25	正态
土壤孔隙度/%	66.51	32.45	44.68	4.64	10.38	3.98	0.99	0.25	正态
毛管持水量/%	13.74	0.01	1.65	2.79	169.09	15.90	3.73	0.35	对数正态
毛管孔隙度/%	18.33	0.01	2.43	4.15	170.78	16.93	3.73	0.23	对数正态
通气孔隙度/%	65.17	12.79	42.26	6.45	15.26	4.11	-0.66	0.17	正态

### 2.2 土壤 $K_s$ 与其他变量的自相关和交互相关分析

对  $K_s$  进行空间分布预测前需要了解不同滞后距离下各变量及变量之间的相互关系。与传统的皮尔逊相关分析相比,交互相关分析的优势在于能够反映不同滞后距离下 2 个变量之间的相关性<sup>[7]</sup>。图 2 和图 3 分别显示了各变量的空间自相关关系及  $K_s$  与其他变量的交互相关关系。自相关分析结果表明,以下 6 个变量在空间上存在自相关关系, $K_s$  在 3 个滞后距离(即 3 个采样间距)下具有显著的空间自相关性, $Bd$ 、孔隙度、 $Ele$ 、 $Asp$ 、 $Pc$  分别在滞后距离 5 和 9,5 和 9,1—16,3 和 6,4 和 9 上自相关系数大于 95%置信水平下的临界值为 0.17。因此,本研究的取样密度足以确定变量具有自相关关系的距离,从而揭示与  $K_s$  相关的各因素空间自相关结构。

交互相关分析结果(图 3)表明, $K_s$  在不同滞后距离和不同方向上与以下 8 个影响变量具有不同的空间相关性。当滞后距离为 4 时  $Bd$  与  $K_s$  存在交互负相关关系,交互相关系数为 -0.184,在其他滞后距离下无交互相关关系。孔隙度、通气孔隙度、 $Ele$ 、 $Slo$ 、 $Asp$ 、 $Pc$  和  $Hc$  交互相关的滞后距离分别为 4,4,-6,0,1,6,-7。除  $Slo$  与  $K_s$  存在交互负相关关

系,其余变量均与  $K_s$  存在交互正相关关系。

### 2.3 土壤 $K_s$ 地统计分析与时空预测模型构建

2.3.1 地统计分析 利用半方差函数分析土壤  $K_s$  的空间变异性(表 2)。指数模型为  $K_s$  最佳拟合模型,块金值代表随机因素影响的大小,主要是在最小采样间隔内由自然过程引起的变化<sup>[27-28]</sup>,研究区土壤  $K_s$  具有较大的块金值( $C_0=38$ ),表明随机因素引起的土壤  $K_s$  变异性较大<sup>[29]</sup>。变程反映空间变量在多远距离范围内具有空间相关性,研究区  $K_s$  在 14 820 m 范围内具有空间依赖性。邵明安等<sup>[30]</sup>指出,最佳采样点间距应小于变程的 1/2,采样间距较合适。空间异质比表示数据的空间相关性,25%以下为强空间相关性,75%以上为弱空间相关性,25%~75%表示中等空间相关性<sup>[31]</sup>。 $K_s$  的空间异质比为 15.32%,说明研究区土壤  $K_s$  具有强的空间依赖性。

2.3.2 土壤  $K_s$  的 RK 预测 逐步回归不仅确保了显著影响土壤  $K_s$  的自变量完全进入回归模型,而且消除了各自变量间的共线性<sup>[32]</sup>。本研究以与土壤  $K_s$  具有空间相关性的 8 个影响因素( $Bd$ 、孔隙度、通气孔隙度、 $Ele$ 、 $Slo$ 、 $Asp$ 、 $Pc$  和  $Hc$ )为自变量,对其进行多元线性逐步回归分析,模型为:

$$\ln K_s = -0.011 * Slo + 0.12 * Bd + 1.929 \quad (11)$$

式中:  $Slo$  为坡度( $^{\circ}$ );  $Bd$  为容重( $g/cm^3$ )。由模型可知,  $Bd$  和  $Slo$  是研究区土壤  $K_s$  进行多元线性逐步回归分析的最佳辅助变量。  $Bd$  是进行土壤  $K_s$  模拟时常被考

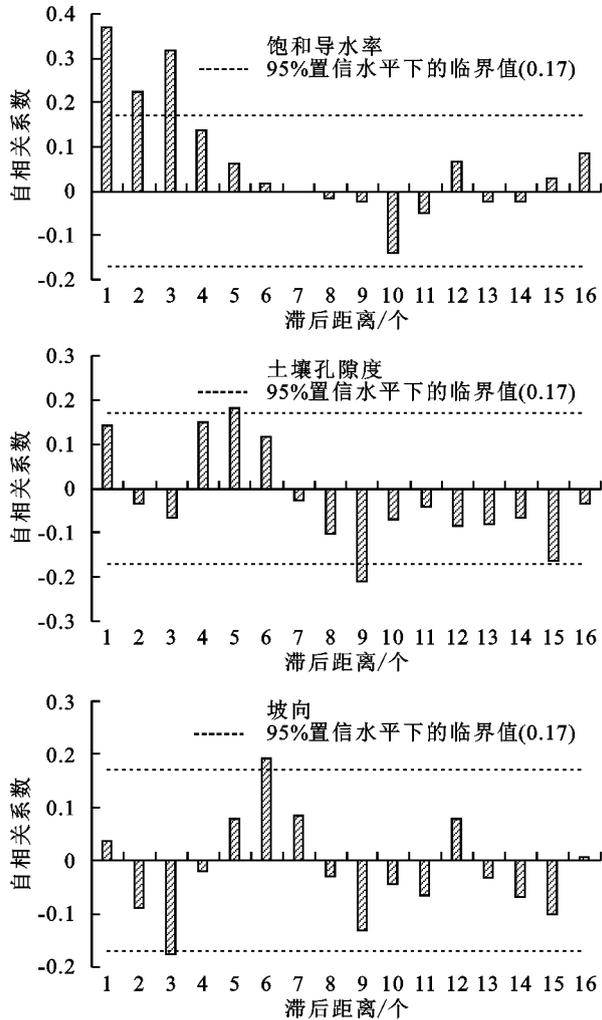
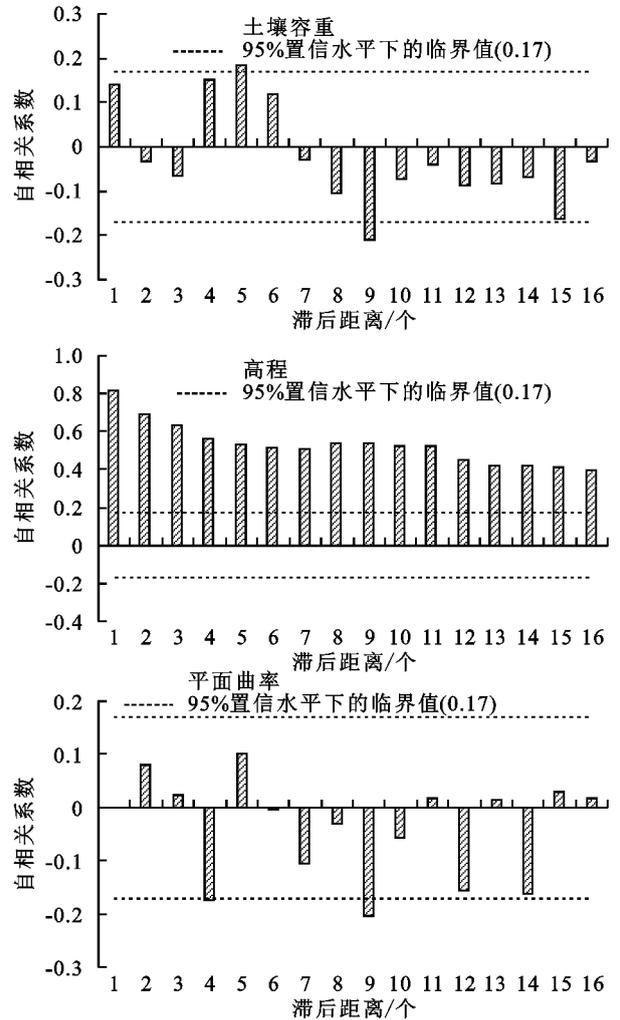


图 2  $K_s$  及其影响因素的自相关分析

通过逐步线性回归预测土壤  $K_s$  的决定系数较低, 仅为 0.234, 这种现象出现是由于研究区土壤  $K_s$  受采样点分布和人为活动的影响, 与环境具有较强的非线性关系。逐步线性回归只能解释其空间分异特征中很小一部分, 为提高预测结果精度对其残差进行 OK 插值, 对逐步线性回归无法解释部分进行重新分析, 最后将残差插值结果和逐步线性回归结果进行叠加, 得到 RK 预测图。

**2.3.3 土壤  $K_s$  的 RF 预测** 土壤  $K_s$  的空间分异反映在土壤形成过程中, 地理位置的变化会导致其自然环境的不同, 环境因素对土壤  $K_s$  的影响也存在明显差异, 使得土壤  $K_s$  空间分布具有空间变异性特点<sup>[10]</sup>。为考虑地理位置对地面特征的影响, 选取经纬度坐标作为建模指标; 同时利用 RF 模型对环境变量进行重要性排序, 最终确定  $Bd$ 、孔隙度、通气孔隙度、 $Ele$ 、 $Slo$  和  $P_c$  6 个变量作为自变量; 用 R 3.5.3 软件中的 RandomForest 软件包预测土壤  $K_s$ 。通过逐次试验,

虑到的因素,  $K_s$  的大小与土壤孔隙数量及大小有着密切联系,  $Bd$  的差异将改变土壤的孔隙状况, 从而导致土壤  $K_s$  的变化<sup>[33]</sup>。因此,  $Bd$  是影响土壤  $K_s$  分布的主要因素, 这与 Hu 等<sup>[34]</sup> 在黄土高原地区的研究结论相同。



确定 RF 模型中  $mtry$  和  $ntree$  参数的最优值<sup>[35]</sup>, 结果表明, 当  $mtry$  值为 3、 $ntree$  值为 1 300 时, RF 拟合效果最好, 模型稳定性最高。

## 2.4 土壤 $K_s$ 空间分布预测结果

为直观反映研究区  $K_s$  的空间变化特征, 基于半方差函数, 通过克里格插值将不同方法预测的土壤  $K_s$  绘制成图(图 4)。从土壤  $K_s$  的空间分布预测图中(图 4)可以看出, 不同方法预测的宁东土壤  $K_s$  的空间分布趋势大致相同, 高值区和低值区的位置和范围基本一致, RF 预测结果与 OK、RK 预测结果具有相似性, 即研究区土壤  $K_s$  分布由北向南逐渐增多。从制图效果看, OK 法得到的图较为粗略, 斑块非常零碎, 不符合现实中土壤  $K_s$  的空间渐变特征, 仅能反映区域土壤  $K_s$  空间变异总体趋势, 对局部变异情况难以反映, 预测精度明显低于其余 2 种方法。RF 预测结果和 RK 的区别在于 RF 对土壤  $K_s$  预测结果

更精细,能体现更多地形信息,斑块边界更为光滑,感

官上更符合土壤  $K_s$  含量的实际分布状况。

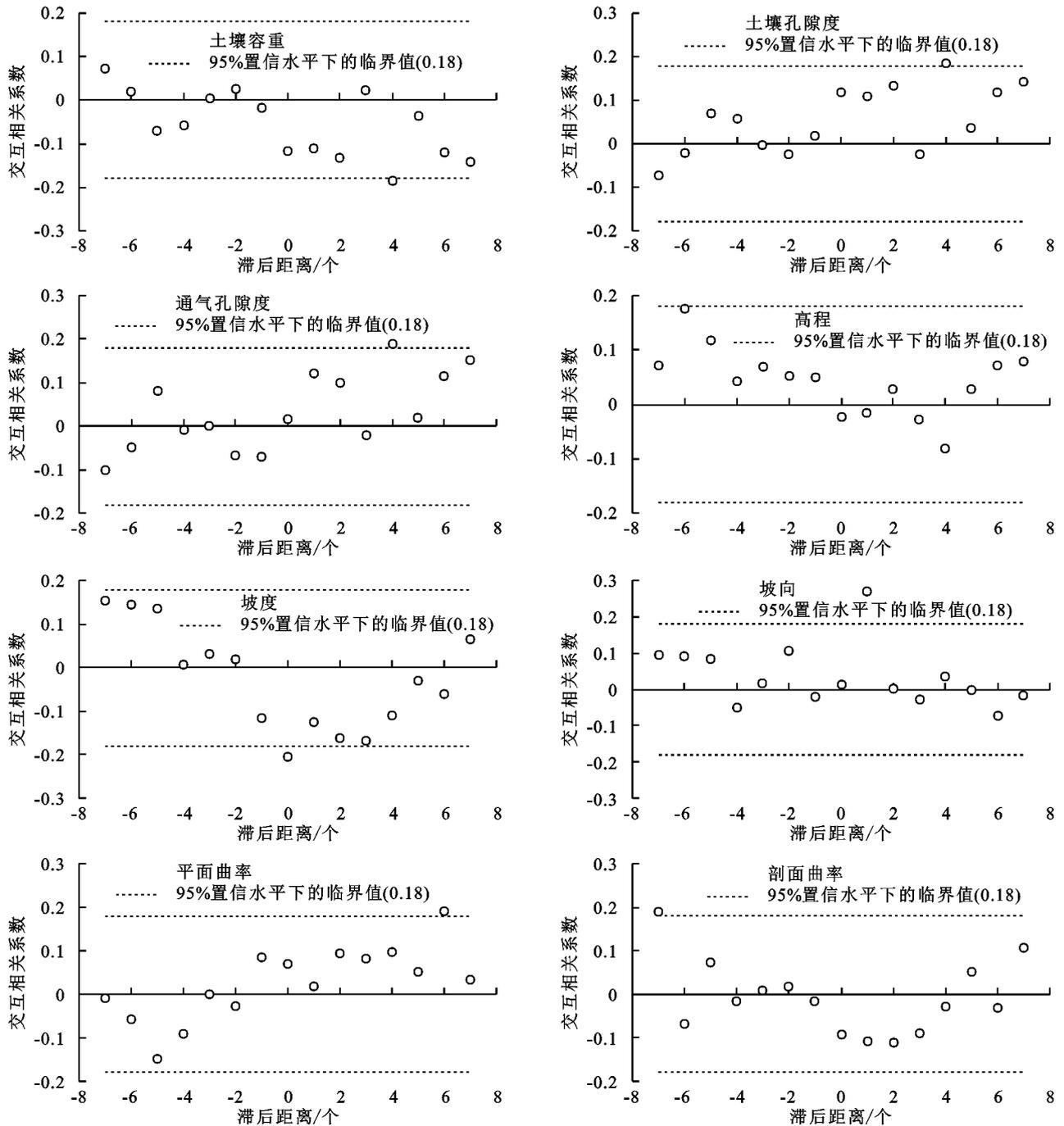


图 3 土壤  $K_s$  及其影响因素的交互相关分析

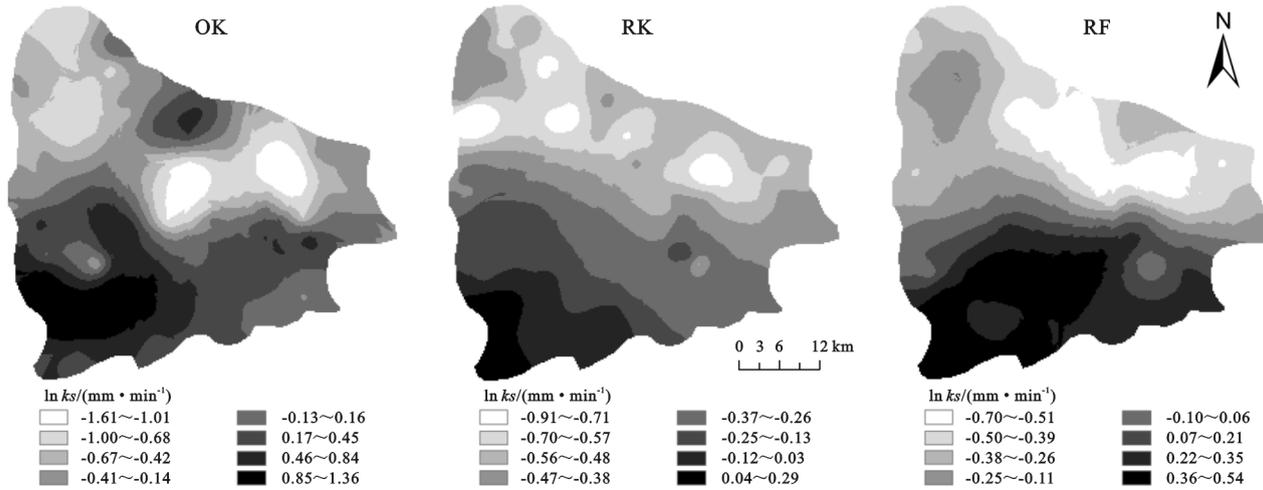
表 2 土壤  $K_s$  半方差函数模型及其结构参数

模型	块金值 $C_0$	基台值 $C_0+C$	变程 A/m	空间异 质比/%	决定 系数 $R^2$
指数模型	38.00	248.00	14820.00	15.32	0.73
线性模型	131.81	283.13	25258.60	46.55	0.67
球状模型	43.90	234.30	9390.00	18.74	0.67
高斯模型	82.20	235.50	8712.22	34.90	0.66

## 2.5 预测精度评价

通过计算 33 个验证点的 MRE、RMSE 和  $RI$ , 检验不同方法的预测精度。由表 3 可知, OK 法预测土壤  $K_s$  的 MRE 和 RMSE 绝对值均为最大, 表明其预

测精度最差; RK 法预测精度也较差, 这主要是实际情况中土壤  $K_s$  与环境因子间并不是简单的线性关系, 更多是由于地形、气候和人类活动等因素影响下所存在的复杂非线性关系<sup>[36]</sup>。相对而言, RF 法的预测精度最好, 这与郭澎涛等<sup>[13]</sup>采用不同预测方法对土壤属性预测结果一致。通过公式(10)计算可知, 以 OK 法作为对照, RK 法和 RF 法在预测土壤  $K_s$  的精度上分别相对提高了 3.11%, 5.53%; 以 RK 法作为参照, RF 法在预测土壤  $K_s$  的精度上提高了 2.49%。表明与 OK 和 RK 法相比, RF 法能更好地捕捉土壤  $K_s$  与环境因子间的复杂关系, 预测精度高。

图 4  $K_s$  空间分布预测图表 3  $K_s$  的不同预测方法精度比较

预测方法	MRE	RMSE
OK	-0.66	1.29
RK	-0.50	1.25
RF	-0.42	1.21

### 3 讨论

描述性统计分析发现,研究区表层土壤  $K_s$  均值为 1.46 mm/min, 变异系数为 106.86%, 属于强变异性。傅子洵等<sup>[37]</sup> 研究神木县土壤  $K_s$  平均值为 1.28 mm/min, 变异系数为 97.33%, 与本研究的结果相近; 于冬雪等<sup>[23]</sup> 对整个黄土区进行了调查发现, 典型黄土区表层土壤  $K_s$  平均值为 0.53 mm/min, 变异系数为 82%, 土壤  $K_s$  均值低于本研究结果; 白一茹等<sup>[38]</sup> 对宁夏压砂地土壤  $K_s$  进行分析得出, 平均值为 0.05 mm/min, 变异系数为 80%, 土壤  $K_s$  均值远低于本研究。原因是宁东和神木县以风沙土为主, 土壤质地较粗, 而黄土高原其他区域土壤质地较细, 所以宁东土壤  $K_s$  值与神木接近, 但是明显高于黄土高原土壤  $K_s$  均值<sup>[9]</sup>。同时宁东地形和土地利用类型差异较大所以在一定程度上也增加了土壤  $K_s$  的变异性<sup>[37]</sup>。研究区  $K_s$  的空间异质比为 15.32%, 在空间上表现出较强的自相关性, 说明在目前研究尺度上引起空间变异的主要因素为气候、地形和成土母质等<sup>[39]</sup>。确定变程可为研究取样尺度提供数据参考<sup>[40]</sup>, 研究区土壤  $K_s$  变程为 14 820 m, 大于取样间距。在统计抽样独立的原则下, 可以适当增加采样距离、降低采样密度, 既能反映研究区土壤  $K_s$  的空间结构, 又可节约采样时间和测试成本<sup>[41]</sup>。

在土壤属性空间预测中没有绝对的最佳方法, 因此有必要在不同研究区中找出最佳预测手段和方法<sup>[42]</sup>。土壤性质的空间自相关程度决定了 OK 法预测的准确性<sup>[43]</sup>, 研究区地形相对复杂, 导致其对研究区土壤  $K_s$  预测精度相对较差, 不能很好地反映土壤

$K_s$  在不同地形情况下的空间差异; RK 法虽引入地形因子, 考虑了土壤  $K_s$  与环境因子间交互相关特征, 但研究区土壤  $K_s$  与相关环境因素间并非简单的线性关系<sup>[8]</sup>, 其受采样点分布和人为活动等的影响, 与环境的关系较复杂; RF 法既考虑了土壤  $K_s$  的空间依赖性, 而且能更准确地揭示土壤  $K_s$  与各因素间的复杂关系<sup>[11]</sup>, 提高了对土壤  $K_s$  空间分布的预测精度, 预测结果有助于较好地揭示研究区不同地形条件下土壤  $K_s$  随各辅助因素变化的细节。因此对于宁东土壤  $K_s$  空间分布预测精度 RF 法较明显高于 OK 和 RK 法。为了取得预期的预测效果, 在采用 RF 法时需要将采样点的实测数据进行充分分析并多次比较, 同时要考虑采样地点的实际情况, 以选择最佳的  $mtry$  和  $ntree$  参数<sup>[44]</sup>。

此外, 在今后的研究中还需要进一步将土壤  $K_s$  与植被覆盖度、土壤水分、土壤侵蚀、土壤孔隙结构等因素和过程结合起来, 以确定影响土壤  $K_s$  的主要因素。

### 4 结论

(1) 宁东表层土壤  $K_s$  均值为 1.46 mm/min, 变化范围为 0.05~7.13 mm/min, 变异系数为 106.86%, 属于强变异性, 符合对数正态分布。

(2)  $K_s$ 、 $Bd$ 、孔隙度、 $Ele$ 、 $Asp$ 、 $Pc$  在不同滞后距离上均具有空间自相关结构。  $K_s$  在不同的滞后距离下与  $Bd$  和  $Slo$  具有显著交互负相关关系, 与孔隙度、通气孔隙度、 $Ele$ 、 $Asp$ 、 $Pc$  和  $Hc$  具有显著交互正相关关系。

(3) 3 种方法预测的  $K_s$  空间分布总体趋势均表现为南高北低, 但局部细节存在一定的差异性。 RF 法能较好捕捉和模拟一个地区环境因子与  $K_s$  之间的关系, 其 MRE 和 RMSE 分别为 -0.42 和 1.21, 绝对值均为最小, 预测精度相较于 OK、RK 法分别提高了 5.53%, 2.49%, 具有最佳的模拟效果。

- 参考文献:
- [1] Wang Y Q, Shao M A, Liu Z P, et al. Regional-scale variation and distribution patterns of soil saturated hydraulic conductivities in surface and subsurface layers in the loessial soils of China [J]. *Journal of Hydrology*, 2013, 487(487): 13-23.
  - [2] 李永宁, 王忠禹, 王兵, 等. 黄土丘陵区典型植被土壤物理性质差异及其对导水特性影响[J]. *水土保持学报*, 2019, 33(6): 176-189.
  - [3] Machiwal D, Jha M K, Mal B C. Modelling infiltration and quantifying spatial soil variability in a wasteland of Kharagpur, India [J]. *Biosystems Engineering*, 2006, 95(4): 569-582.
  - [4] 韩永伟, 高吉喜, 李咏红, 等. 宁东能源化工基地规划生态环境影响评价研究[J]. *环境与可持续发展*, 2008, 9(1): 58-60.
  - [5] 李风军, 冯晓秀, 陆桂琴. 宁东能源化工基地生态环境脆弱性评价研究[J]. *生态科学*, 2014, 33(5): 1017-1022.
  - [6] 李涛, 李晓琳, 郝新梅, 等. 应用土壤质地预测干旱区葡萄酒土壤饱和导水率空间[J]. *农业工程学报*, 2014, 30(9): 84-92.
  - [7] 赵春雷, 邵明安, 贾小旭. 黄土高原北部坡面尺度土壤饱和导水率分布与模拟[J]. *水科学进展*, 2014, 25(6): 806-815.
  - [8] 李祥东, 邵明安, 赵春雷. 西北干旱区土壤水力参数空间变异与模拟[J]. *干旱区研究*, 2019, 36(6): 1325-1332.
  - [9] 张川, 陈洪松, 张伟, 等. 喀斯特坡面表层土壤含水量、容重和饱和导水率的空间变异特征[J]. *应用生态学报*, 2014, 25(6): 1585-1591.
  - [10] 任丽, 杨联安, 王辉, 等. 基于随机森林的苹果区土壤有机质空间预测[J]. *干旱区资源与环境*, 2018, 32(8): 141-146.
  - [11] 齐雁冰, 王茵茵, 陈洋, 等. 基于遥感与随机森林算法的陕西省土壤有机质空间预测[J]. *自然资源学报*, 2017, 32(6): 1074-1086.
  - [12] 杨其坡, 王伟, 刘洪斌. 基于地形因子和随机森林的丘陵区农田土壤有效铁空间分布预测[J]. *中国生态农业学报*, 2018, 26(3): 422-431.
  - [13] 郭澎涛, 李茂芬, 罗微, 等. 基于多源环境变量和随机森林的橡胶园土壤全氮含量预测[J]. *农业工程学报*, 2015, 31(5): 194-201.
  - [14] 卢宏亮, 赵明松, 刘斌寅, 等. 基于随机森林模型的安徽省土壤属性空间分布预测[J]. *土壤*, 2019, 51(3): 602-608.
  - [15] 李占文, 骆有庆, 王东菊, 等. 宁东能源化工基地大环境绿化有害生物危害现状及防控对策[J]. *宁夏工程技术*, 2010, 9(4): 376-379.
  - [16] 智静, 乔琦, 傅泽强. 干旱地区工业系统资源代谢水平分析: 以宁东能源煤化工基地为例[J]. *干旱区地理*, 2015, 38(1): 155-162.
  - [17] 白一茹, 王幼奇, 展秀丽. 陕北农牧交错带土地利用方式对土壤物理性质及分布特征的影响[J]. *中国农业科学*, 2013, 46(8): 103-111.
  - [18] 中国科学院南京土壤研究所. 土壤理化分析[M]. 上海: 上海科学技术出版社, 1983.
  - [19] Zimmerman D L, Zimmerman M B. A Comparison of spatial semivariogram estimators and corresponding ordinary Kriging predictors [J]. *Technometrics*, 1991, 33(1): 77-91.
  - [20] 姚雪玲, 傅伯杰, 吕一河, 等. 基于 GIS 和统计模型的黄土丘陵沟壑区土壤水分插值方法[J]. *水土保持学报*, 2013, 27(6): 93-96.
  - [21] 包青岭, 丁建丽, 王敬哲, 等. 基于随机森林算法的土壤有机质含量高光谱检测[J]. *干旱区地理*, 2019, 42(6): 1404-1414.
  - [22] Lindner C, Bromiley P A, Ionita M C, et al. Robust and accurate shape model matching using random forest regression-voting [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1862-1874.
  - [23] 于冬雪, 贾小旭, 黄来明. 黄土区不同土层饱和导水率空间变异与影响因素[J]. *土壤通报*, 2018, 49(5): 1073-1079.
  - [24] Dai Y J, Shan G W, Duan Q Y, et al. Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling [J]. *Journal of Hydrometeorology*, 2015, 14(3): 869-887.
  - [25] Warrick A W, Nielsen D R. Spatial variability of soil physical properties [J]. *Applications of Soil Physics*, 1980, 42(3): 319-344.
  - [26] 连纲, 郭旭东, 傅伯杰, 等. 黄土高原小流域土壤容重及水分空间变异特征[J]. *生态学报*, 2006, 26(3): 647-654.
  - [27] 赵越, 罗志军, 廖牧鑫, 等. 泰和县耕地土壤养分空间分布及影响因素[J]. *水土保持学报*, 2018, 32(5): 299-306.
  - [28] 张珍明, 周运超, 黄先飞, 等. 喀斯特小流域土壤有机碳密度及碳储量空间分布异质性[J]. *水土保持学报*, 2017, 31(2): 184-190.
  - [29] 王军, 傅伯杰, 邱扬, 等. 黄土丘陵小流域土壤水分的时空变异特征 - 半变异函数[J]. *地理学报*, 2000, 55(4): 428-438.
  - [30] 邵明安, 王全九, 黄明斌. 土壤物理学[M]. 北京: 高等教育出版社, 2006: 228-261.
  - [31] Cambardella C A, Moorman T B, Parkin T B, et al. Field-scale variability of soil properties in central iowa soils [J]. *Soil Science Society of America Journal*, 1994, 58(5): 1501-1511.
  - [32] 刘志鹏, 邵明安, 王云强. 区域尺度下黄土高原土壤全钾含量的空间模拟[J]. *农业工程学报*, 2012, 28(22): 132-140.
  - [33] Chapuis, Robert P. Predicting the saturated hydraulic conductivity of soils: A review [J]. *Bulletin of Engineering Geology & the Environment*, 2012, 71(3): 401-434.
  - [34] Hu W, Shao M A, Si B C. Seasonal changes in surface bulk density and saturated hydraulic conductivity of

- natural landscapes [J]. *European Journal of Soil Science*, 2012, 63(6): 820-830.
- [35] Sonobe R, Tani H, Wang X, et al. Parameter tuning in the support vector machine and random forest and their performances in cross- and same-year crop classification using TerraSAR-X [J]. *International Journal of Remote Sensing*, 2014, 35(23): 7898-7909.
- [36] Wagner B, Tarnawski V R, Hennings V, et al. Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set [J]. *Soil Science Society of America Journal*, 2001, 108(1/2): 145-147.
- [37] 傅子洹, 王云强, 安芷生. 黄土区小流域土壤容重和饱和导水率的时空动态特征[J]. *农业工程学报*, 2015, 31(13): 128-134.
- [38] 白一茹, 王幼奇, 王菲, 等. 压砂地土壤导水特性空间格局及影响因子[J]. *干旱地区农业研究*, 2016, 34(4): 55-61.
- [39] 张晗, 欧阳真程, 赵小敏, 等. 江西省耕地土壤氮素空间变异特征及其主控因素[J]. *水土保持学报*, 2018, 32(5): 307-315.
- [40] 刘春利, 胡伟, 贾宏福, 等. 黄土高原水蚀风蚀交错区坡地土壤剖面饱和导水率空间异质性[J]. *生态学报*, 2012, 32(4): 1211-1219.
- [41] 王幼奇, 赵云鹏, 白一茹, 等. 中宁枸杞土壤碳组分分布特征及其空间异质性[J]. *环境科学*, 2018, 39(3): 1396-1403.
- [42] 谢恩泽, 赵永存, 陆访仪, 等. 不同方法预测苏南农田土壤有机质空间分布对比研究[J]. *土壤学报*, 2018, 55(5): 14-24.
- [43] Zhu Q, Lin H S. Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes [J]. *Pedosphere*, 2010, 20(5): 594-606.
- [44] 王茵茵, 齐雁冰, 陈洋, 等. 基于多分辨率遥感数据与随机森林算法的土壤有机质预测研究[J]. *土壤学报*, 2016, 53(2): 342-354.
- (上接第 284 页)
- [12] 肖克飏, 吴普特, 雷金银, 等. 不同类型耐盐植物对盐碱土生物改良研究[J]. *农业环境科学学报*, 2013, 31(12): 2433-2440.
- [13] Garland G, Bunemann E K, Oberson A, et al. Phosphorus cycling within soil aggregate fractions of a highly weathered tropical soil: A conceptual model [J]. *Soil Biology and Biochemistry*, 2018, 116: 91-98.
- [14] 仇建飞, 窦森, 邵晨, 等. 添加玉米秸秆培养对土壤团聚体胡敏酸数量和结构特征的影响[J]. *土壤学报*, 2011, 48(4): 781-787.
- [15] Joergensen R G, Emmerling C. Methods for evaluating human impact on soil microorganisms based on their activity, biomass, and diversity in agricultural soils [J]. *Journal of Plant Nutrition and Soil Science*, 2006, 169: 295-309.
- [16] Pérez C S, Elsa E C, Robert W, et al. Soil microbial responses to drought and exotic plants shift carbon metabolism [J]. *The ISME Journal*, 2019, 13: 1776-1787.
- [17] Mohammad B, Falk H, Sofia K F, et al. Structure and function of the global topsoil microbiome [J]. *Nature*, 2018, 560: 233-237.
- [18] Liu S L, Baidengsha M, Joergensen R G, et al. Response of soil microorganisms after converting a saline desert to arable land in central Asia [J]. *Applied Soil Ecology*, 2016, 98: 1-7.
- [19] Kruse J, Abraham M, Amelung W, et al. Innovative methods in soil phosphorus research: A review [J]. *Journal of Plant Nutrition and Soil Science*, 2015, 178(1): 43-48.
- [20] Balemi T, Negisho K. Management of soil phosphorus and plants adaptation mechanisms to phosphorus stress for sustainable crop production: A review [J]. *Journal of Soil Science Plant Nutrition*, 2012, 12: 547-562.
- [21] Sato S, Solomon D, Hyland C, et al. Phosphorus speciation in manure and manure-amended soils using XANES spectroscopy [J]. *Environmental Science and Technology*, 2005, 39(19): 7485-7491.
- [22] Robles-Aguilar A A, Pang J J, Postma J A, et al. The effect of pH on morphological and physiological root traits of *Lupinus angustifolius* treated with struvite as a recycled phosphorus source [J]. *Plant and Soil*, 2019, 434: 65-78.
- [23] Schachtman D P, Shin R. Nutrient sensing and signaling: NPKS [J]. *Annual Review of Plant Biology*, 2007, 58: 47-69.
- [24] Jalali M, Jalali M. Relation between various soil phosphorus extraction methods and sorption parameters in calcareous soils with different texture [J]. *Science of the Total Environment*, 2016, 566: 1080-1093.
- [25] Zhou Y, Staver A C. Enhanced activity of soil nutrient-releasing enzymes after plant invasion: A meta-analysis [J]. *Ecology*, 2019, 100(11): e02830.