引用格式：刘明杰,徐卓揆,郜允兵,等.基于机器学习的稀疏样本条件下的土壤有机质估算方法[J].地球信息科学学报,2020,22(9):1799-1813. [Liu M J, Xu Z K, Gao Y B, et al. Estimating soil organic matter based on machine learning under sparse sample[J]. Journal of Geo-information Science, 2020,22(9):1799-1813.] **DOI:**10.12082/dqxxkx.2020.190441

基于机器学习的稀疏样本下的土壤有机质估算方法

刘明杰[1,2],徐卓揆[1,3],郜允兵[2,4*],杨　晶[2,4],潘瑜春[2,4],高秉博[5],周艳兵[2,4],周万鹏[2,6],王　凌[7]

1. 长沙理工大学交通运输学院,长沙 410114; 2. 国家农业信息化工程技术研究中心,北京 100097; 3. 长沙理工大学公路地质灾变预警空间信息技术湖南省工程实验室,长沙 410114; 4. 北京农业信息技术研究中心,北京 100097; 5. 中国农业大学,北京 100083;6. 河南理工大学,焦作 454003; 7. 河北省农林科学院农业资源环境研究所,石家庄 050051

Estimating Soil Organic Matter based on Machine Learning Under Sparse Sample

LIU Mingjie[1,2], XU Zhuokui[1,3], GAO Yunbing[2,4*], YANG Jing[2,4], PAN Yuchun[2,4] , GAO Bingbo[5], ZHOU Yanbing[2,4], ZHOU Wanpeng[2,6], WANG Ling[7]

1. School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410114, China; 2. Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China; 3. Engineering Laboratory of Spatial Information Technology of Highway Geological Disaster Early Warning in Hunan Province (Changsha University of Science & Technology),Changsha 410114, China; 4.National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; 5. China Agricultural University, Beijing 100083, China; 6. Henan Polytechnic University, Jiaozuo 454003, China; 7. Institute of Agricultural Resources and Environment, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050051, China

Abstract: To improve the accuracy of soil organic estimation in the case of sparse samples and to construct the soil organic predictive models applying the machine learning methods, GRNN (Generalized Regression Neural Network) and RF(Random Forest). The soil was diluted into 8 samples with different sampling density (2703, 1352, 676, 339, 169, 85, 43, 22 samples) according to the soil organic matter sampling data of Daxing agricultural land in 2007 applying the MMSD (Minimization of the Mean of the Shortest Distances) criterion. GRNN (Generalized Regression Neural Network), RF (random forest) and Ordinary Kriging are applied to predict each sampling density espectively. Cross Validation is used to verify the prediction accuracy of unknown samples at each sampling density. With the decrease of sampling point density, the spatial correlation between sampling points decreases gradually, thus the semivariogram's fitting precision deteriorates, the errorofprediction point result increases, and the confidence of the prediction decreases. The spatial correlation between sampling points is close to disappear when the sample is diluted under 43 and 22 samples, and the coefficient of

收稿日期：2019-08-13；**修回日期：**2019-12-14.

基金项目：国家重点研发计划课题(2017YFD0801205)；北京市农林科学院科技创新能力建设专项(KJCX20170407、KJCX20200414)；湖南省教育厅资助科研项目(13B129)；湖南省工程实验室开放基金资助项目(KFJ180602)。[**Foundation items:** The National Key Research and Development Program of China, No.2017YFD0801205; The Science and Technology Innovation Capacity Building Project of Beijing Academy of Agriculture and Forestry Sciences No.KJCX20170407, KJCX20200414; Scientific Research Project Funded by The Education Department of Hunan Province, No.13B129; Project Supported by Open Fund of Hunan Engineering Laboratory, No. KFJ180602.]

作者简介：刘明杰(1995—)，男，贵州贵阳人，硕士生，研究方向为地理信息系统。E-mail: 2210478688@qq.com

***通讯作者：**郜允兵(1976—)，男，河南焦作人，博士，主要从事农业环境监测与时空分析技术研究。E-mail: gybgis@163.com

determination of the semivariogram function is low and the residual is large. The impacts the Ordinary Kriging receives, which are from the changes in the number of the sampling points, sampling density and spatial structures of samples is obvious. The prediction accuracy of the method decreases with the decrease of the number of sampling points. There is no significant correlation between the predicted values and the observed values at or below 85 sampling points. The prediction accuracy of GRNN and RF is almost independent of the sampling density. The predicted values fluctuate within a certain threshold space around the observed values, and has good correlation. At sampling points of 85 and below, the prediction accuracy is greatly improved compared with Ordinary Kriging. Ordinary Kriging is not suitable for spatial interpolating calculation in the case of sparse samples, especially in the case of weak spatial correlation. The machine learning models can fully learn the environmental information and spatial proximity information of soil sampling points. They combine attribute similarity and spatial correlation and have better stability and adaptability, not being easy to be affected by the number of sampling points, configuration and sampling density, and can make stable and accurate predictions even when the spatial autocorrelation between sampling points is very weak.

Key words: soil organic matter; spatial interpolation; machine learning; attribute similarity; spatial correlation; Daxing County; sparse sample; sampling density

***Corresponding author:** GAO Yunbing, E-mail: gybgis@163.com

摘要：采用GRNN（Generalized Regression Neural Network）和RF（Random Forest）2种机器学习方法构建土壤有机质预测模型，以提高稀疏样本情况下的土壤有机质估算精度。依据北京市大兴区农用地2007年的土壤有机质采样数据，按MMSD准则（Minimization of the Mean of the Shortest Distances）抽稀为8种不同采样密度的样本（分别为2703、1352、676、339、169、85、43、22个样本），分别采用GRNN、RF和Ordinary kriging对各采样密度下的未知采样点进行预测，采用交叉检验的方式验证各采样密度下未知样点的预测精度。随着采样点密度的下降，样点间的空间自相关性逐渐减弱，半变异函数的拟和精度变差，预测点结果误差增大，预测的置信度降低。当抽稀到43个和22个采样点时，样点间的空间自相关性接近歼灭，半变异函数的决定系数较低且残差较大。普通克里格受到采样点数量和采样密度、样点的空间结构的影响比较明显，其预测精度随采样点数量的下降而下降。在85个采样点及以下时，其预测值与观测值之间没有显著的相关性。GRNN和RF的预测精度受采样密度的影响不大，其预测精度在一个较小的范围内波动，其预测值围绕观测值在一定阈值空间内震荡波动，具有较好的相关性，在85个及以下的采样密度时，预测精度相对普通克里格有较大的提升。普通克里格法不适合在稀疏样本条件下空间插值计算，尤其是在空间自相关性比较弱的情况下。机器学习模型能充分学习土壤间环境信息、样点空间邻近效应信息，兼顾属性相似性和空间自相关，具有更好的稳定性和适应性，不容易受到采样点数量、构型和采样密度等因素的影响，即使在采样点空间自相关性很弱的情况下也能做出稳定预测精度。

关键词：土壤有机质；空间插值；机器学习；属性相似性；空间自相关；大兴区；稀疏样本；采样密度

1 引言

土壤有机质（Soil Organic Matter，SOM）是土壤的重要组成部分，对土壤物理、化学、生物过程以及土壤肥力积累和植物的生长具有重要影响，掌握土壤有机质的赋存情况对于农田土壤环境保护、土壤肥力的合理调节以及农业可持续发展等方面都有重要的意义[1-3]。土壤是一个时空连续的自然变异体。土壤有机质与区域尺度有一定的关联性，尺度不同影响因素也往往不同[4-5]，从土壤有机质的形成机理出发，微生物是土壤有机物质分解和周转的主要驱动力，凡能影响微生物活动及其生理作用的因素都会影响有机物质的分解和转化，同时也就是影响土壤有机质含量的因素[6-8]。县域尺度下地形、土地利用方式、施肥条件、畜牧粪便等因素对微生物生理及活动产生重要影响，研究表明县域尺度下上述因素均对土壤有机质含量有较大的影响[9-14]。孔祥斌等[15-16]通过对北京大兴区20余年的土壤养分变化数据分析，表明农田土壤长期受到施肥、耕作措施、种植制度等各种人为活动的影响，有机质分布的空间自相关性减弱，朝均一化方向发展。胡克林等[17]的研究发现大兴区土壤有机质含量主要受土壤质地、土地利用方式、人为耕作管理措施等因素影响。赵汝东等[18]的研究表明，地形、土壤质地、

耕地利用类型及施肥管理等因素对有机质含量的分布具有较大影响。国内外许多学者利用土壤环境变量运用机器学习的方法进行土壤有机质估算。李启权等[19]以平均气温、降水、相对湿度、日照时数、太阳辐射、坡度和归一化植被指数等量化指标作为RBF神经网络的输入建立模型，对全国的土壤有机质进行估算；其后又以高程、坡度、地形湿度指数和增强型植被指数等指标输入RBF模型对县域有机质含量进行插值[20]。江叶枫等[21]构建了集成BP神经网络模型、多元逐步回归的RBF-NN模型[22]、与普通克里格相结合的BPNN模型和RBFNN模型[23]，用于对县域和省域的有机质进行插值。另外，Pouladi N[24]、杨联安[25]等有代表性的学者采用极限学习机和随机森林等方法开展土壤有机质预测的研究。前人关于估算土壤属性变量的相关研究主要集中在采样密度相对较大的情形，对于空间自相关性较弱，且稀疏的样本点下研究工作开展较少，缺乏一种利用小样本进行精准土壤有机质插值的技术手段。土壤有机质的形成受到环境因素的影响，相同环境或相似环境下的有机质含量往往差异较小。同时，采样点之间还存在空间相关性，在一定尺度范围内，距离越近的点则有机质的含量更趋于接近。本文以北京大兴区农用地为例，基于土壤采样点的空间自相关性与土壤影响因素的环境属性相似性，结合多维环境变量，利用机器学习方法预测该区域土壤有机质的含量，评价稀疏样点下基于机器学习的方法下土壤有机质估算能力的稳定性。

2　研究方法、研究区概况与数据来源

2.1　研究方法

2.1.1　研究路线

本文选取土壤有机质影响因素构建特征空间，利用GRNN、RF和普通克里格法分别在不同采样密度下建立预测模型进行估计，并对估计结果进行分析和评价，结合不同采样密度下采样点的空间分析结果得出结论，研究路线如图1所示。

为了使样本点在整个抽样空间里最大可能地分布均匀，本文采用MMSD准则（Minimization of the Mean of the Shortest Distances）进行样本抽稀。具体描述为在采样区域任意一点与最临近采样点的距离平均值最小，本实验过程通过R语言编程实现样本数据的抽稀。根据MMSD准则对采样点进行逐步抽稀，尽量均匀的分布采样点，建立采样点个数分别为2703、1352、676、339、169、85、43、22的8个实验组（为避免随机现象的影响，43个和22个采样点分别各5组进行实验）。然后在每个实验组下分别用GRNN、RF和普通克里格法对土壤有机质进行预测。机器学习模型输入变量为坐标（x，y）和3个环境因子（土壤质地、土地利用类型、畜禽粪便影响强度），输出为采样点的有机质含量值，该过程通过MATLAB_R2017b实现。普通克里格法利用GS+9.0构建半变异函数模型然后在ArcGIS 10.2中插值。

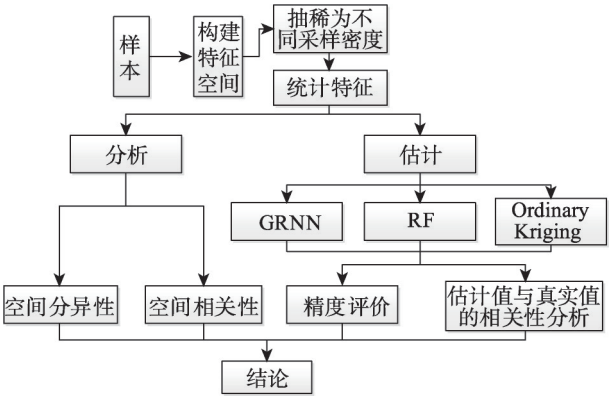


图1　研究路线

Fig. 1　Research Roadmap

在每一组实验中，以10折交叉验证所得的最终精度作为该组实验的预测精度（由于样本量太小，43个采样点为7折交叉验证，22个采样点为5折交叉验证）。具体做法是把采样点被分为10份，第一份保留作预测的检验集，其余9份作训练集，下一次把第二份保留作预测的检验集，其余9份作训练集，重复10次，遍历整个数据集。

实验结果通过误差均方根（RMSE）、平均绝对误差（MAE）、平均相对误差（MRE）3个指标进行评价预测精度，其中RMSE和MAE可以反映预测结果与样本之间整体的偏移情况，MRE可以反映预测结果相对样本本身的偏移情况。

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

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|(\widehat{y_i}-y_i)\right| \tag{2}$$

$$MRE=\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{y_i}-y_i}{y_i}\times 100\% \tag{3}$$

式中：$\hat{y}_i$为预测值；y_i为观测值；n为检验样本数。

2.1.2 机器学习模型

近年来，机器学习方法因其优异的非线性映射能力，已经被成功的应用于土壤科学[26-27]、农学[28-29]、水文学[30-31]、气象学[32]以及环境科学等[33-35]诸多领域。其中，广义回归神经网络和随机森林在预测方面具有较强的稳定性，适合于小样本下的函数逼近。因此，本文选择GRNN和RF 2种模型进行土壤有机质的估算。

广义回归神经网络（Generalized Regression Neural Network，GRNN）最早由美国学者Sprecht提出，它是径向基函数神经网络的一种，是由输入层、模式层、求和层和输出层组成的4层网络，本文中对应的输入（样本特征）X=[*x*, *y*, *texture*, *landuse*, *dung*]^T（*x*=坐标*x*值、*y*=坐标y值、*texture*=土壤质地、*landuse*=土地利用类型、*dung*=畜禽粪便利用强度），输出（标签）*Y*=[*SOM*]^T（*SOM*=土壤有机质值），样本的个数为*m*，GRNN模型结构如图2所示。

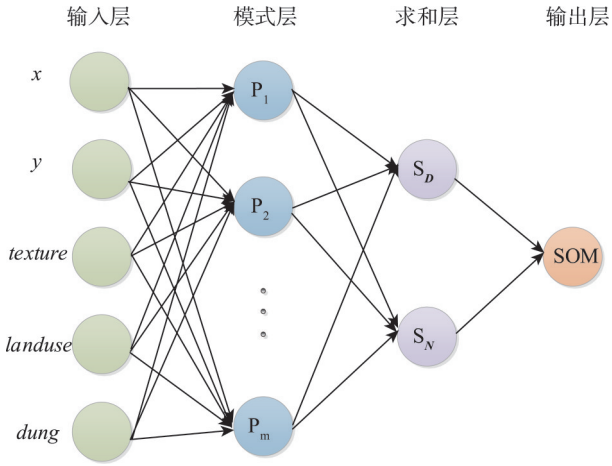


图2 广义回归神经网络结构

Fig. 2 The structure chart of General Regression Neural Network

输入层节点个数等于样本的特征维度，各神经元是简单的分布单元，直接将输入变量传递给模式层。

模式层节点个数等于训练样本的个数*m*，各神经元对应不同的样本，传递函数为

$$P_i = \exp\left(-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right) (i=1, 2, \cdots, m) \quad (4)$$

式中：*X*为网络的输入变量；X_i为神经元*i*对应的学习样本（*i*=1, 2, ⋯, *m*）；σ为平滑因子，是模型的超参数。

求和层节点个数等于输出样本维度加1，分为两部分，第一个节点输出模式层输出的算术和，即

$$S_D = \sum_{i=1}^{m} P_i \ (i=1, 2, \cdots, m) \quad (5)$$

其余节点输出模式层输出的加权和，即

$$S_{Nj} = \sum_{i=m}^{m} Y_{ij} P_i \ (i=1, 2, \cdots, m) \quad (6)$$

式中：*j*代表输出的维度；Y_{ij}代表样本X_i对应的标签，在本文中输出只有SOM，因此式（3）又可以写成

$$S_N = \sum_{i=m}^{m} SOM_i P_i \ (i=1, 2, \cdots, m) \quad (7)$$

输出层节点个数等于输出集的维度，每个节点的输出等于对应的求和层输出与求和层第一个节点输出相除，即

$$\widehat{SOM} = \frac{S_N}{S_D} \quad (8)$$

$\widehat{SOM}$为输入变量*X*对应的土壤有机质估计值。

随机森林（RF，Random Forest）是以决策树或回归树为基学习器的集成算法，在本文中选择回归树作为基学习器构建模型，其结构如图3，具体步骤为：

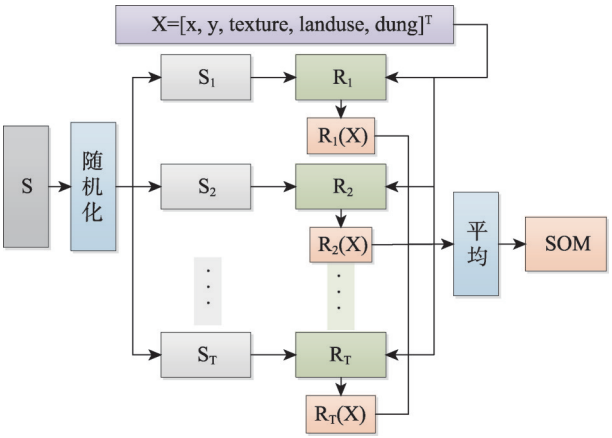


图3 随机森林结构

Fig. 3 The structurechart of Random Forest

（1）用Bootstrap方法对训练样本*S*重采样，随机产生*T*个训练集$S_1, S_2, \cdots, S_T$。

（2）利用每个训练集，生成对应的回归树$R_1, R_2, \cdots, R_T$；在每个非叶子节点分裂属性前，从M个属性中随机抽取mtry个属性作为当前节点的分裂属性集，并以这*mtry*个属性中最好的分裂方式对该节点进行分裂，在整个森林的生长过程中，*mtry*的值维持不变。在本文中*M*就是输入样本*X*的维度，*mtry*为模型的超参数。

（3）重复第（1）和第（2）步，直到生成 *ntree* 棵回归树，*ntree* 为超参数，需要足够大。

（4）对于测试集样本 X，利用每个决策树进行测试，得到对应的回归 $R_1(X), R_2(X), \cdots, R_T(X)$。

（5）对全部决策树的回归进行加和求平均作为输出。

在训练机器学习模型的过程中，随机森林生成决策树的个数（*ntree*）选择在500以上，逐渐增加至精度无明显变化为止；决策树的属性拆分（*mtry*）选择在 $\sqrt{M}$ 以下，调试选择为使精度最高的参数，调试的方法是使用袋外数据检验。GRNN的平滑因子（σ）取值选择在0~1之间，同样通过预留的数据集调试最优。

2.2 研究区概况

研究区为北京市大兴区南部耕作农田，基本情况如图4所示，其位于北京市城南116°13′E—116°34′ E，39°26′ N—39°51′ N之间。全境属永定河冲积平原，土地总面积约1036 km^2，农用地面积约563.69 km^2，地势总体自西北向东南略倾斜，海拔高程在13.4~52 m之间，坡度在0.8%~1.0%左右，全区成土母质为永定河冲积物，土壤类型以砂质潮土、壤质潮土为主，其面积占研究区总面积的90.72%，且由西南向东北，土壤质地由砂变粘。大兴区属暖温带半湿润大陆季风气候，年平均气温为11.6 ℃，年平均降水量556.4 mm，雨热同季，但季节分配不均，76.2%雨水集中在夏季。总人口53.5万人，其中农业人口占37万人，是北京市主要的粮、菜、瓜、果生产基地，是重要的京郊农业大区。

样品采集于2007年4—9月，采用空间系统布样方式，样点均匀布设在北京市大兴区南部农用地面积比重大的地区，包括庞各庄镇、北臧村镇、礼贤镇、长子营镇、魏善庄镇、安定镇、采育镇、榆垡镇。共有2703个采样点，利用GPS定位每个采样点的地理坐标，并对采样点编号且记录采样时间、采样地点和土地利用方式。每一个采样点采集耕层土壤（0~25 cm）样品，每个样品均采用混合样，由采样点周围50 m范围内的5~10个土壤样品均匀混合，测定方法按土壤有机质测定方法执行（GB 9834-88）[36]。

2.3 土壤环境因子

土壤有机质具有明显的尺度效应，不同尺度上的有机质其主导影响因素不尽相同[37-39]。在全国尺度下，气候类型在一定程度上对有机质的分布起到

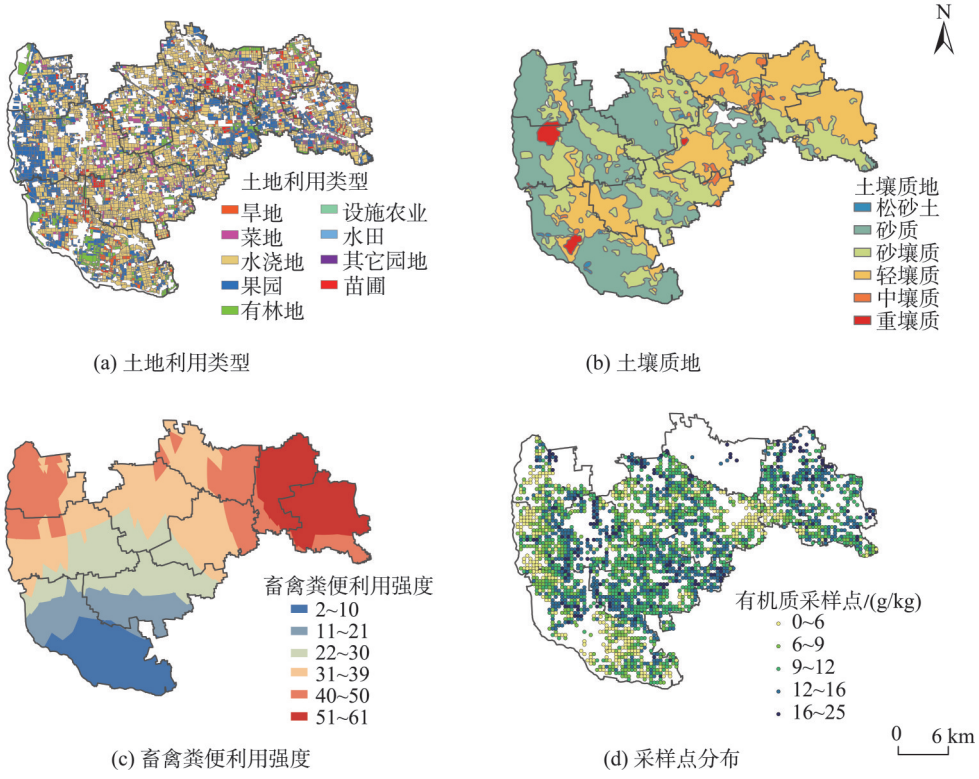


图4　北京市大兴区基本概况

Fig. 4　The overview of Daxing district of Bejing

了控制和主导作用。在省域尺度下,气候的主导性降低,植被、地貌地形、土壤类型、土壤质地及土地利用类型等因素对有机质的分布影响更强烈。县域及以下尺度仍然受到国家和省域尺度的环境因素制约,但施肥条件、耕作制度以及灌溉方式等人为管理措施的影响更为深刻,表现为局部异质性和分区异质性。

本研究区,研究区地势平坦,以平原地形为主,不存在主要的山脉或丘陵,海拔西高东低,最高处海拔为44.25 m,最低处为15.10 m,研究区地形的分区异质性不明显,对有机质空间变异影响不大。研究区土壤类型绝大多数为潮土,约占整个研究区面积的86.41%,其余土壤类型中面积较多的为潮褐土和盐化潮土,分别占研究区总面积的7.19%和4.31%,3种土壤类型的性质比较接近,土壤类型的分区异质性同样不明显。本文在开展实验设计时,采用地理探测器对上述影响因素对土壤有机质的贡献测算,均无显著性,不作为本研究的影响因子。从调查和统计研究发现,由于技术外溢和扩散作用,研究区内农户的化肥施用常年稳定,呈现村组、各户异质性不是十分突出。但当地农民普遍有就近取粪沤肥的耕作习惯,因此可以用畜禽粪便的利用强度代表研究区内有机肥的施用情况,作为有机质的影响因素进行计算。当地的作物轮作情况与土地利用方式密切相关,因此土地利用数据基本可以表达作物轮作的特征。

利用方差分析研究在低采样密度下土地利用类型、土壤质地、畜禽粪便利用强度、土壤类型及植被指数(Normalized Difference Vegetation Index,NDVI)对土壤有机质的影响力,在85个点时分析结果见表1。在用地类型、土壤质地和畜禽粪便利用强度这3个因素的影响下,土壤有机质的组间均方均大于组内均方(F>1),F检验显示在0.01的水平下显著(P<0.01),说明土地利用类型、土壤质地和畜禽粪便利用强度在该研究区内对有机质的含量具有重要的影响。而土壤类型、植被指数这两个因素影响下的有机质组间均方均小于组内均方(F<1),且没有通过F的显著性检验(P>0.01),说明土壤类型、植被指数在此情形下对土壤有机质的分布没有起到显著的影响。

因此,选取土壤质地、土地利用类型以及畜禽粪便的影响强度这些环境因子作为研究区土壤有机质的环境控制因素。其中,对类型变量数据参考高凤杰等[40]研究先进行概率化处理,具体为:

表1 土壤有机质含量方差分析

Tab. 1 Variance analysis of soil organic matter

	方差来源	偏差平方和	自由度Df	均方	F	P
用地类型	组间	351.750	4	87.938	6.871	0.000
	组内	1023.868	80	12.796		
	总体	1375.618	84			
土壤质地	组间	356.405	3	118.802	9.442	0.000
	组内	1019.213	81	12.583		
	总体	1375.618	84			
畜禽粪便利用强度	组间	241.351	5	48.270	3.362	0.008
	组内	1134.267	79	14.358		
	总体	1375.618	84			
土壤类型	组间	0.945	2	0.472	0.028	0.972
	组内	1374.674	82	16.764		
	总体	1375.618	84			
植被指数	组间	17.592	2	8.796	0.531	0.590
	组内	1358.027	82	16.561		
	总体	1375.618	84			

（1）将研究区的有机质值的范围记为 $[S_{\min}, S_{\max}]$，将其等值分为 m 组（m 通常取6~12之间），则第 K 组的取值范围可表示为：

$$S_K = \left[S_{\min} + \frac{S_{\max} - S_{\min}}{m} \times (K-1),\ S_{\min} + \frac{S_{\max} - S_{\min}}{m} \times K \right] \tag{9}$$

（2）以土壤质地（记为 T）为例，则有机质与土壤质地的定量关系可以表示为：

$$R_{(S_K, T)} = \left[\left(T_1, \frac{C(1)_K}{C_K} \right), \cdots, \left(T_X, \frac{C(X)_K}{C_K} \right) \right] \tag{10}$$

式中：$R_{(S_K, T)}$ 为 S_K 范围内有机质和土壤质地的概率关系；X 为土壤质地的类别数；$C(1)_K$ 为 S_K 范围内 T_1 类别的采样点个数；C_K 为 S_K 范围内的采样点个数。

（2）基于有机质与土壤质地的定量关系求取有机质的模糊分布表达式

$$P^K = R_{(S_K, T_i)} \quad (i = 1, 2, \cdots, X) \tag{11}$$

$$f_i = [(S_1, P^1), (S_2, P^2), \cdots, (S_m, P^m)] \tag{12}$$

P^K 表示在 T_i 情况下，预测值与 S_K 类别的相似度；f_i 表示在 T_i 情况下，有机质的模糊分布表达式。

畜禽粪便影响强度通过下列过程得到。收集整理2007年以前存在于大兴区及周边的规模化畜禽养殖场数据，Ⅱ级规模赋予权重1，Ⅰ级规模赋予权重2，通过缓冲区分析与叠置分析，得到每个养殖场覆盖区的规模分级权重加和，即大兴区的畜禽粪便影响强度（无量纲）。养殖规模分级参考现行的畜禽养殖业污染物排放标准（GB18596-2001）[41]，缓冲区半径设置为10 km，参考走访调研的结果与李艳霞等的研究[42]。

3　结果及分析

3.1　土壤有机质含量的统计特征

将研究区各实验组在SPSS20.0中对土壤有机质含量进行描述性统计分析（表2），研究区表层（0~25 cm）土壤有机质含量范围在1.2032~24.7293 g/kg之间，按第二次全国土壤普查养分分级标准（表3）比

表2　研究区所有实验组土壤有机质含量描述性统计

Tab. 2　Descriptive statistics of soil organic matter in all experimental groups in the study area

实验组	极大值/（g/kg）	极小值/（g/kg）	平均值/（g/kg）	标准差/（g/kg）	变异系数	偏度	峰度	K-S双侧显著性
D2703	24.73	1.20	10.49	3.91	37.28	0.147	−0.098	0.302
D1352	24.73	1.65	10.53	3.92	37.23	0.159	−0.003	0.632
D676	24.73	1.65	10.58	4.11	38.87	0.318	0.036	0.640
D339	24.73	1.81	10.55	3.94	37.32	0.223	0.292	0.946
D169	22.98	1.98	10.20	3.99	39.08	0.290	0.112	0.964
D85	18.05	2.03	10.93	4.05	37.01	−0.181	−0.804	0.934
D43_1	17.97	4.35	11.07	3.46	31.24	−0.187	−0.616	0.972
D43_2	17.73	2.98	10.07	4.17	41.39	0.122	−0.893	0.951
D43_3	17.84	3.53	11.10	3.53	31.80	−0.091	−0.900	0.855
D43_4	19.67	2.02	10.33	4.04	39.13	0.128	−0.064	0.906
D43_5	17.68	2.19	10.39	4.46	42.90	−0.329	−0.998	0.445
D22_1	18.28	3.53	10.70	4.30	40.13	0.293	−0.988	0.611
D22_2	17.61	3.91	11.28	3.84	34.03	−0.037	−0.634	0.940
D22_3	16.40	3.72	10.49	3.60	34.29	−0.380	−0.692	0.783
D22_4	15.34	4.29	10.56	3.24	30.72	−0.320	−0.825	0.912
D22_5	17.86	2.36	9.35	4.36	46.60	0.260	−0.540	0.999

注：为排除随机现象的干扰，43、22个采样点时各抽5组进行实验。

表3　土壤有机质含量分级标准

Tab. 3　Soil organic matter content grading standard

土壤属性	丰富	较丰富	中等	较缺乏	缺乏	极缺乏
有机质/（g/kg）	>40	30~40	20~30	10~20	6~10	< 6

较处于偏低的水平。从样点的正态分布特征看，随样点数的减少，偏度有从正偏态转变为负偏态的趋势，峰度也有降低的趋势。变异系数通常被认为能较好地揭示随机变量的离散程度，按照变异系数的划分等级：变异系数＜10%为弱变异性、变异系数在10%~100%之间为中等变异性、变异系数＞100%为强变异性。各组的变异系数在30.72%~46.60%之间，属于中等变异程度，随采样点的减少，变异系数变得不稳定。通过K-S检验，各组均符合正态分布，至少在较低采样密度时满足普通克里格插值的需求，极少采样点时很难拟合出稳态的半变异函数，在此仅用来说明稀疏样本下的问题。

大兴区土壤有机质的采样点分布情况如图5所示，土壤有机质呈现出局部聚集的空间特征。在大兴区东北部与中西部地区出现土壤有机质的高值聚集区，中部偏东与西北部地区为低值的聚集区。大兴区的土壤有机质的含量整体偏低，主要集中分布在6~16 g/kg之间。从整体观察，土壤有机质的分布与土壤质地的分布具有一定的吻合。

利用GS+软件为每个实验组构建半变异函数模型，然后选择拟合最优的模型进行普通克里格插值，模型的选择标准为决定系数 R^2 尽量大，残差RSS尽量小。最终选择的模型如图5所示，图中 C_0 为块金值，C_0+C 为基台值，R 为变程，单位是m，R^2 表示决定系数，RSS表示模型拟合的残差。从图中可以看出，随采样点数量的减少，模型的决定系数逐渐减小，残差逐渐增大，这表明模型的可信度随着采样密度降低逐渐减小。

3.2 预测结果

抽选D2703、D676、D167、D43共4个实验组的预测结果（图6），在2703个采样点的情况下，3种模型都能基本反映研究区土壤有机质的分布情况。随样本数量的降低，3种预测模型对有机质值刻画的精细程度都有所下降，但整体拟合程度GRNN和RF的结果要优于普通克里金。

随采样点数量的减少，普通克里格法的3个误差指标（RMSE、MRE和MAE）都在增大，而GRNN和RF的误差指标则在一个较小的区间范围内波动，各组实验具体的预测精度如表4所示。从图7中进一步可以发现，3种预测方法RMSE随采样点减少的变化趋势，克里格预测结果的RMSE随样本数减少而增大，在85个采样点及以下时RMSE达到最大并且趋于稳定，而GRNN和RF预测结果的RMSE随样本数减少没有特别明显的变化规律。这个现象表明普通克里格方法容易受到采样点的间距和数量的影响，而机器学习方法受到采样密度和采样点构型的影响较小。

在D43和D22的5次抽样中各预测方法的精度均发生了一些变化，但变化趋势并不完全相同，且克里格法的变化幅度相对较大。这说明3种方法的预测均受到了采样点构型的影响，但影响的原理并不相同，且克里格法的表现更为剧烈。从普通克里格插值分析的结果来看，在低采样密度的情况下，各精度评价指标显示其插值结果误差偏大，R^2 较小，说明此时样点数据集所获得的空间变异分析结果不能满足普通克里格插值的需求。

4 讨论

4.1 样本的空间分析

采用全局Moran's I 指数对各实验组的样本进行空间自相关分析，随着采样点数量的下降，样点之间的平均最短距离不断增加，使得土壤有机质的空间自相关性逐渐减弱，Moran's I 指数检验结果的置信度逐渐降低（当P>0.10，且−1.65<Z< 1.65时，置信度不足90%），直到最后空间自相关性的表现为不显著，具体结果如表5所示。

利用地理探测器对各实验组数据进行探测，从q值的变化可以看出，随着采样点数量的减少，样本之间的空间分异性有逐渐增强的趋势，与此同时样本之间的空间自相关性正在逐步减弱。结合表5可看到，在样本间平均最短距离超过2000 m之后，空间分异性基本占据了主导地位。机器学习模型通过学习综合利用了采样点的空间相关性和采样点的背景环境信息，因此这可以解释在此情形下机器学习模型的预测结果比克里格法的预测结果更好的原因，分异属性及因子探测的结果如表6所示。

4.2 GRNN、RF和Ordinary Kriging预测方法的比较

将各实验组的预测值与观测值进行相关性分析（表7），在85个采样点及以下的各组实验中，普通克里格法的预测结果与观测值之间并没有显著的线性相关关系，这表明当前的预测结果不能反映土壤有机质分布的结构特征，是不可靠的预测结果，这也与之前的空间自相关分析及半变异分析的

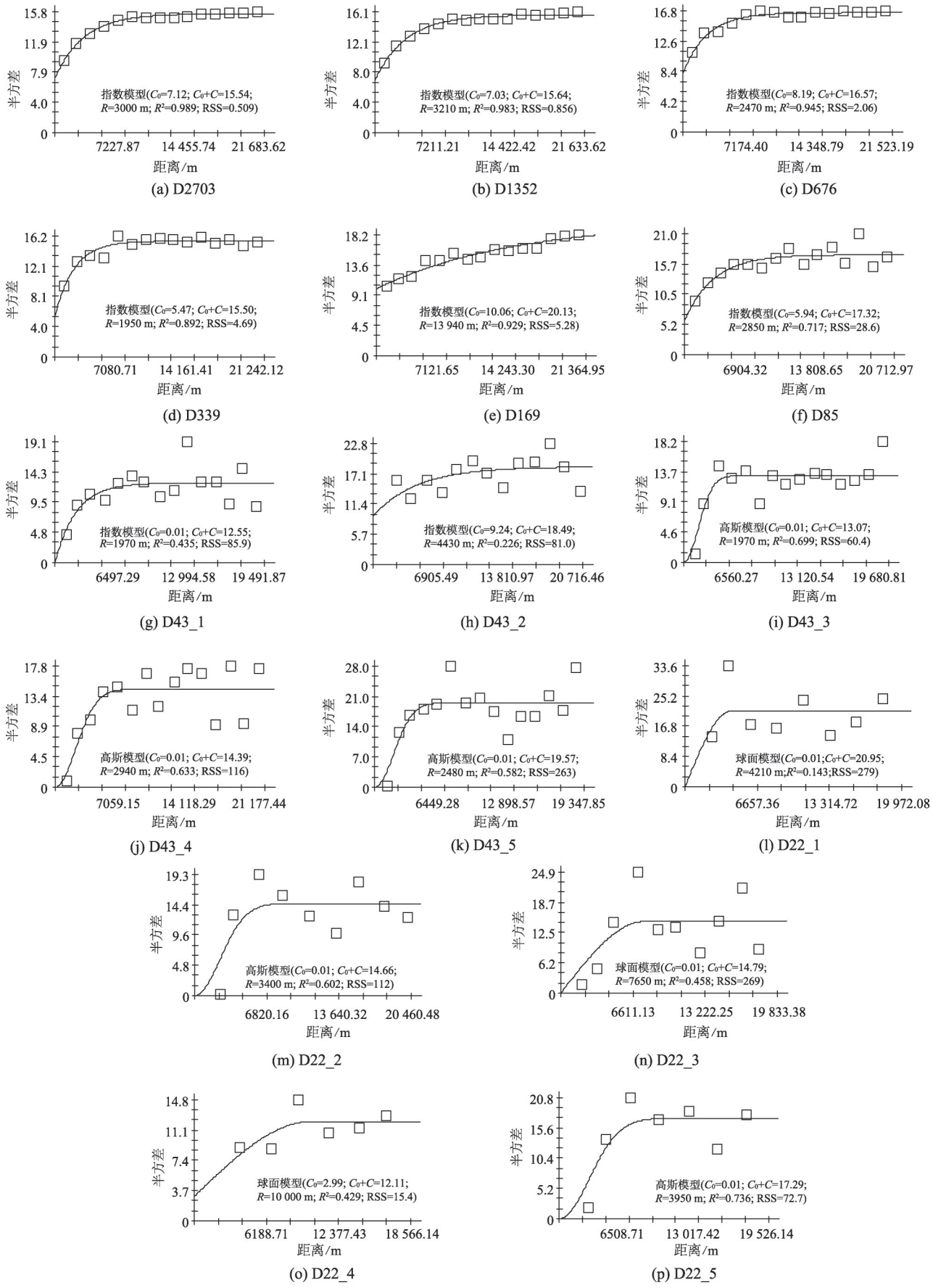


图5 所有实验组半变异函数模型及相关参数

Fig. 5 Variograms founction and related parameters of all experimental groups

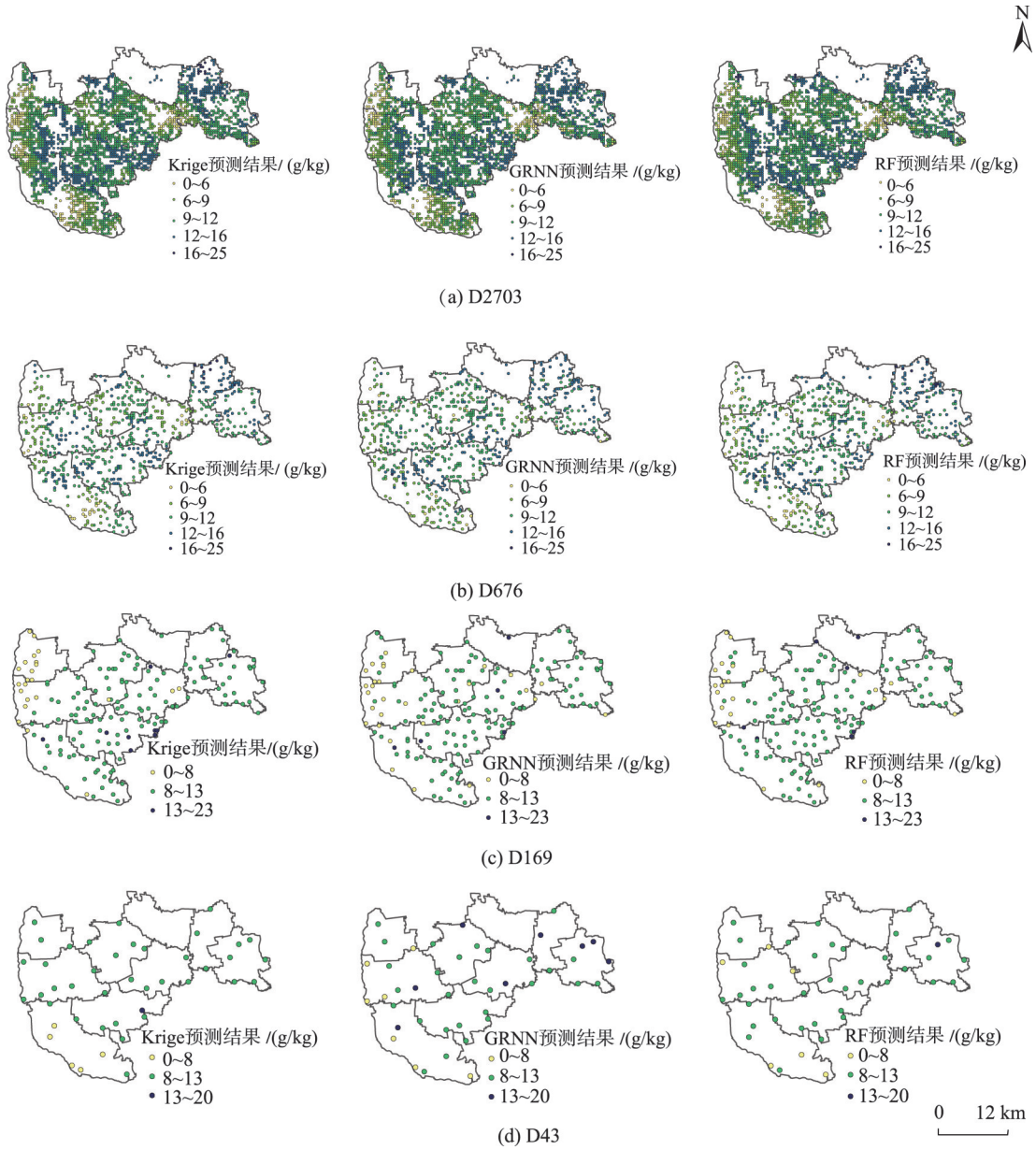


图6 D2703、D676、D169、D43实验组3种方法预测结果对比

Fig. 6 Comparison of the results predicted by the three methods in D2703, D676, D169 and D43

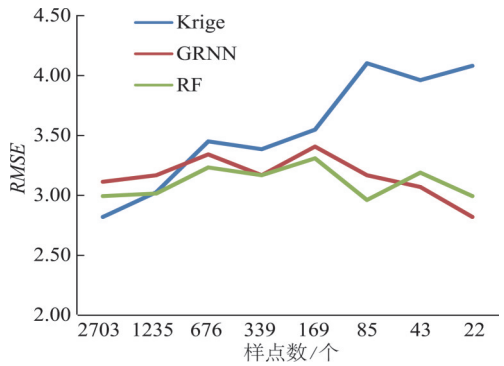
结果相互印证。随采样点的抽稀，采样点的数量下降，样点间的距离增加，有机质的空间自相关性减弱，甚至消失，难以构建可靠的半变异模型，导致最终的预测结果不可信。而机器学习方法的预测值与观测值之间始终具有一定的相关性，表明其预测的结果仍然具备一定的可信度，其方法可以作为低采样密度下的一种预测手段，具体结果如表6所示。

对土壤有机质的预测结果表明，在低采样密度下机器学习方法对有机质样本预测的准确性优于普通克里格法，即使在采样点空间自相关性消失的情况下，也能得到具有一定置信度的预测结果。杨联安等[43]利用随机森林结合辅助变量准确预测土壤有机质空间分布的研究也表明，机器学习方法能够有效刻画土壤属性与环境因子间存在的复杂非线性关系，对比普通克里格、回归克里格，机器学习的预测精度较高，尤其是在成土环境复杂地区，这和本研究的结果基本相同。沈掌泉等[44]利用GRNN和Krige在不同采样密度下对土壤pH、有效P、有效K、有效Mg和有效S进行插值估计，结果表明GRNN的插值精度超过了Krige的插值精度，尤其是当样点数仅为原始布局的不到15 %的时候，与Krige相比，即使在适当减少采样数量和加大采样

表4　所有实验组GRNN、RF和普通克里格法的预测精度

Tab. 4　Prediction accuracy of GRNN, RF and Ordinary Krigingin all experimental groups

实验组	RMSE			MRE/%			MAE		
	Krige	GRNN	RF	Krige	GRNN	RF	Krige	GRNN	RF
D2703	2.82	3.11	2.99	26.89	29.59	28.92	2.21	2.43	2.35
D1352	3.02	3.17	3.01	29.54	30.28	29.50	3.02	2.46	2.36
D676	3.45	3.34	3.23	33.00	31.94	30.50	2.74	2.65	2.57
D339	3.38	3.17	3.17	35.60	30.64	31.19	2.82	2.47	2.53
D169	3.55	3.41	3.31	35.92	36.64	34.39	2.83	2.75	2.61
D85	4.10	3.17	2.96	43.76	33.37	30.45	3.47	2.72	2.46
D43_1	3.36	2.84	2.72	30.40	25.63	23.73	2.74	2.31	2.18
D43_2	4.13	3.19	3.11	44.58	34.52	33.70	3.34	2.70	2.65
D43_3	3.68	3.14	3.31	35.53	30.26	31.93	3.17	2.67	2.91
D43_4	3.95	2.76	3.18	44.86	33.59	37.36	3.21	2.36	2.69
D43_5	4.69	3.42	3.61	67.44	38.46	47.36	4.06	2.70	3.01
D22_1	4.34	3.22	3.42	43.60	35.06	33.76	3.77	2.86	2.91
D22_2	4.31	2.43	2.62	65.64	24.60	26.55	3.83	2.13	2.28
D22_3	3.90	2.89	3.07	77.11	27.63	33.04	3.54	2.27	2.64
D22_4	3.53	2.56	2.86	58.38	23.23	27.42	3.03	2.18	2.45
D22_5	4.31	2.98	2.98	146.35	35.83	41.13	3.79	2.44	2.55



注：43、22个采样点为5组预测的平均值。

图7　GRNN、RF及Ordinary Kriging预测方法RMSE随采样点数量减少的变化

Fig. 7　The change of RMSE of GRNN, RF and Ordinary Kriging when the number of sampling points decreases

间隔时，也能获得较好的结果，这也与本文的研究结果相吻合。

在实验的过程里尝试过调整机器学习模型中输入的影响因素，发现减少土壤质地、土地利用类型、畜禽粪便影响强度这3个因素中的任意一个都会造成预测精度的下降；增加研究区的土壤类型或植被指数作为输入，发现仅在D2703和D1352的预测精度保持基本不变，其余各组的预测精度都有不同程度的下降。这可以认为在大兴区土壤类型和植被对有机质的空间分布没有起到主导作用，与孔祥斌等[16]、胡克林等[17]的研究结果相吻合。

关于克里格法适用范围的土壤采样密度研究中，有人认为应控制在0.28个/km^{2}以上，才能满足克里格空间插值分析的需求[45]，也有人认为合理的采样

表5　所有实验组样本的空间自相关分析

Tab. 5　Spatial correlation analysis of all experimental groupsamples

实验组	平均最短距离/m	Moran's *I*值	*Z*得分	*P*值	实验组	平均最短距离/m	Moran's *I*值	*Z*得分	*P*值
D2703	371.45	0.35	56.78	0.00	D43_3*	2615.74	0.15	1.15	0.25
D1352	425.63	0.33	30.83	0.00	D43_4*	3109.01	0.15	1.30	0.19
D676	567.68	0.33	14.99	0.00	D43_5*	2909.11	0.08	0.83	0.41
D339	810.79	0.20	8.24	0.00	D22_1*	4237.90	−0.32	−1.39	0.17
D169	1285.91	0.17	5.68	0.00	D22_2*	4384.49	−0.15	−0.53	0.60
D85	1914.45	0.19	3.06	0.00	D22_3*	3662.95	−0.10	−0.19	0.85
D43_1	2792.95	0.22	2.41	0.02	D22_4*	4342.87	0.10	0.74	0.46
D43_2	3110.80	0.14	1.52	0.13	D22_5*	4723.02	−0.11	−0.48	0.63

注：*表示检验结果为空间自相关性不显著。

表6 3种土壤有机质影响因素探测结果

Tab. 6 Detected result of three influence factor of soil organic matter

实验组	土地利用类型		土壤质地		畜禽粪便影响强度	
	*q*值	*p*值	*q*值	*p*值	*q*值	*p*值
D2703	0.182	0.000	0.199	0.000	0.148	0.000
D1352	0.209	0.000	0.190	0.000	0.139	0.000
D676	0.198	0.000	0.233	0.000	0.171	0.000
D339	0.212	0.000	0.189	0.000	0.138	0.000
D169	0.210	0.000	0.229	0.000	0.202	0.000
D85	0.256	0.067	0.259	0.016	0.175	0.035
D43	0.500	0.004	0.177	0.071	0.345	0.050
D22	0.496	0.016	0.477	0.049	0.742	0.034

表7 GRNN、RF和普通克里格法的预测值与观测值的相关性分析

Tab. 7 The correlation analysis between the observed values and predicted values of GRNN, RF and Ordinary Kriging

实验组	Krige	GRNN	RF	实验组	Krige	GRNN	RF
D2703	0.686**	0.608**	0.642**	D43_3	0.155	0.388*	0.392**
D1352	0.635**	0.590**	0.637**	D43_4	0.153	0.669**	0.576**
D676	0.543**	0.581**	0.616**	D43_5	−0.067	0.591**	0.577**
D339	0.450**	0.580**	0.578**	D22_1	−0.137	0.572**	0.590**
D169	0.433**	0.500**	0.545**	D22_2	−0.343	0.709**	0.630**
D85	0.175	0.599**	0.660**	D22_3	−0.094	0.584**	0.440*
D43_1	0.210	0.532**	0.594**	D22_4	−0.102	0.547**	0.398*
D43_2	0.129	0.559**	0.627**	D22_5	0.054	0.635**	0.762**

注：检查手段为Pearson相关系数。*表示在0.05水平（双侧）上显著相关；**表示在0.01水平（双侧）上显著相关。

密度应该在0.59个/km^{2}以上[46]。本文通过预测值与观测值的相关性分析发现，在169个采样点以下，各组的普通克里格插值结果与观测值之间没有显著的相关关系，且3种精度指标显示其误差偏大，综合分析认为在县级研究区适合利用普通克里格插值分析的采样密度至少应该在0.31个/km^2（169个点）及以上。

5 结论

本文在县域尺度下从土壤有机质的形成与影响因素出发，采用随机森林、广义神经网络模型，从训练数据中学习样点的环境属性相似特征和空间临近的自相关特征，稀疏样本下未知的采样点进行土壤有机质估算，并对不同采样密度下的样本预测值精度进行交叉验证。实验结果表明：

（1）随着采样点密度、强度的降低，样本之间的空间自相关性逐渐减弱，土壤有机质的空间分异性随着采样尺度增加而逐渐增强，适合利用普通克里格插值分析的采样密度至少应该在0.31个/km^{2}及以上。

（2）GRNN、RF等机器学习方法在小样本的情况下容易受到训练样本中极值的干扰，在使用前需要结合研究区域的实际情况对样本进行适当的数据预处理工作。样本应尽量分散，容纳整个研究区域的数据特征，尽量保证数据的多样性，如果样本集中分布在某一小区域内（这个区域既指空间范围，也指属性范围），那么机器学习的预测结果基本只会接近这一区域的数据，从而失去对研究区域整体的反映。样本数量的减少可能会导致样本整体蕴含的信息量下降，这不利于机器学习方法的使用，因此采用多年的时空数据在理论上可以提高机器学习模型的预测精度，具体情况还有待进一步的研究。

（3）随机森林、广义神经网络模型等机器学习方法受采样数量和采样点间距的影响不大，在样本量稀疏的条件下，相对于普通克里格法预测精度提高了23%~37%。机器学习方法能充分学习土壤采样点的环境信息和空间邻近信息，能兼顾土壤采样点的属性相似性与空间自相关，有利于充分挖掘环境信息的价值进行未知点的属性估算。

参考文献(References)：

[1] Ondrasek G, Bakić Begić H, Zovko M, et al. Biogeochemistry of soil organic matter in agroecosystems & environmental implications[J]. Science of The Total Environment, 2019,658:1559-1573.

[2] 徐云鹤,方斌.江浙典型茶园土壤有机质空间异质性分析[J].地球信息科学学报,2015,17(5):622-630. [Xu Y H, Fang B. Study on spatial heterogeneity of the soil organic matter in typical tea gardens of Jiangsu Province and Zhejiang Province[J]. Journal of Geo- information Science, 2015,17(5):622-630.]

[3] 陈桂香,高灯州,曾从盛,等.福州市农田土壤养分空间变异特征[J].地球信息科学学报,2017,19(2):216-224. [Chen G X, Gao D Z. Characteristics of the spatial variation of soil nutrients in farmland of Fuzhou City[J]. Journal of Geo-information Science, 2017,19(2):216-224.]

[4] Wiesmeier M, Urbanski L, Hobley E, et al. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales[J]. Geoderma, 2019,333:149-162.

[5] Guo Z, Adhikari K, Chellasamy M, et al. Selection of terrain attributes and its scale dependency on soil organic carbon prediction[J]. Geoderma, 2019,340:303-312.

[6] 黄昌勇,徐建明.土壤学(第三版)[M].北京:中国农业出版社,2010:29-43 [Huang C Y, Xu J M. Soil Science (3rd edition)[M]. Beijing: China Agricultural Press, 2010:29-43.]

[7] Paul E A. The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization[J]. Soil Biology and Biochemistry, 2016, 98:109-126.

[8] Khan K S, Mack R, Castillo X, et al. Microbial biomass, fungal and bacterial residues, and their relationships to the soil organic matter C/N/P/S ratios[J]. Geoderma, 2016,271:115-123.

[9] Tu C, He T, Lu X, et al. Extent to which pH and topographic factors control soil organic carbon level in dry farming cropland soils of the mountainous region of Southwest China[J]. CATENA, 2018,163:204-209.

[10] Sun W, Zhu H, Guo S. Soil organic carbon as a function of land use and topography on the Loess Plateau of China [J]. Ecological Engineering, 2015,83:249-257.

[11] 杨晓.县域土壤有机质空间分布研究——以山东省诸城市为例[D].济南:山东农业大学,2018. [Yang X, Study on spatial distribution of soil organic matter in the county area: A case study of Zhucheng City, Shandong Province [D]. Jinan: Shandong Agricultural University, 2018.]

[12] 高凤杰,马泉来,韩文文,等.黑土丘陵区小流域土壤有机质空间变异及分布格局[J].环境科学,2016,37(5):1915-1922. [Gao F J, Ma Q L, Han W W, et al. Spatial variability and distribution pattern of soil organic matter in a mollisol watershed of China[J]. Environmental Science, 2016, 37(5):1915-1922.]

[13] 宋莎,李廷轩,王永东,等.县域农田土壤有机质空间变异及其影响因素分析[J].土壤,2011,43(1):44-49. [Song S, Li T X, Wang Y D, et al. Spatial variability of soil organic matter and its influencing factors at county scales[J]. Soils, 2011,43(1):44-49.]

[14] 王宗明,张柏,宋开山,等.东北平原典型农业县农田土壤养分空间分布影响因素分析[J].水土保持学报,2007(2):73-77. [Wang Z M, Zhang B, Song K S, et al. Analysis of related factors for soil nutrients in croplands of typical agricultural county, Northeast Plain, China[J]. Journal of Soil and Water Conservation, 2007(2):73-77.]

[15] 孔祥斌,张凤荣,王茹.近20年城乡交错带土壤养分时间空间变异特征分析——以北京市大兴区为例[J].土壤,2004(6):636-643. [Kong X B, Zhang F R, Wang R. Spatial and temporal variation of soil nutrients in periurban region: A case study of Daxing County in Beijing City[J]. Soils, 2004(6):636-643.]

[16] 秦静,孔祥斌,姜广辉,等.北京典型边缘区25年来土壤有机质的时空变异特征[J].农业工程学报,2008,24(3):124-129. [Qin J, Kong X B, Jiang G H, et al. Characteristics of spatio-temporal changes of soil organic matter in typical fringe in Beijing for 25 years[J]. Transactions of the Chinese Society of Agricultural Engineering, 2008,24(3): 124-129.]

[17] 胡克林,余艳,张凤荣,等.北京郊区土壤有机质含量的时空变异及其影响因素[J].中国农业科学,2006,38(4):764-771. [Hu K L, Yu Y, Zhang F R, et al. The Spatial-Temporal variability of soil organic matter and its influencing factors in suburban area of Beijing[J]. Scientia Agricultura Sinica, 2006,38(4):764-771.]

[18] 赵汝东,孙焱鑫,王殿武,等.北京地区耕地土壤有机质空间变异分析[J].土壤通报,2010,41(3):552-557. [Zhao R D, Sun Y X, Wang D Y, et al. Research on spatial variability of soil organic matter in Beijng field[J]. Chinese Journal of Soil Science, 2006(4):764-771.]

[19] 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.]

[20] 李启权,王昌全,张文江,等.基于神经网络模型和地统计学方法的土壤养分空间分布预测[J].应用生态学报,2013,24(2):459-466. [Li Q Q, Wang C Q, Zhang W J, et al. Prediction of soil nutrients spatial distribution based on neural network model combined with geostatistics[J]. Chinese Journal of Applied Ecology, 2013,24(2):459-466.]

[21] 江叶枫,郭熙,叶英聪,等.应用集成BP神经网络模型预测土壤有机质空间分布[J].江苏农业学报,2017,33(5):1044-1050. [Jiang Y F, Guo Z, Ye Y C, et al. Spatial distribution of soil organic matter predicted by BP neural network ensemble model[J]. Jiangsu Journal of Agricultural Sciences, 2017,33(5):1044-1050.]

[22] 江叶枫,郭熙.基于辅助变量和回归径向基函数神经网络(R-R BFNN)的土壤有机质空间分布模拟[J].浙江农业学报,2018,30(4):640-648. [Jiang Y F, Guo Z. Prediction of soil organic matter distribution based on auxiliary variables and regression-radial basis function neural network (R-RBFNN) model[J]. Acta Agriculturae Zhejiangensis, 2018,30(4):640-648.]

[23] 江叶枫,郭熙,叶英聪,等.省域尺度土壤有机质空间分布的神经网络法预测[J].江苏农业学报,2017,33(4):828-835. [Jiang Y F, Guo Z, Ye Y C, et al. Prediction of spatial distribution of soil organic matter at provincial scale with neural network[J]. Jiangsu Journal of Agricultural Sciences, 2017,33(4):828-835.]

[24] Pouladi N, Møller A B, Tabatabai S, et al. Mapping soil organic matter contents at field level with Cubist, Random Forest and Krige[J]. Geoderma, 2019,342:85-92.

[25] 宋英强,杨联安,冯武焕,等.基于多源辅助变量和极限学习机的蔬菜地土壤有机质预测研究[J].土壤通报,2017,48(1):118-126. [Song Y Q, Yang L A, Feng W H, et al. Prediction for soil organic matter in vegetable fields based on cooperative variables and extreme learning machine algorithm[J]. Chinese Journal of Soil Science, 2017,48(1):118-126.]

[26] Zeraatpisheh M, Ayoubi S, Jafari A, et al. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran[J]. Geoderma, 2019,338:445-452.

[27] Feng Y, Cui N, Hao W, et al. Estimation of soil temperature from meteorological data using different machine learning models[J]. Geoderma, 2019,338:67-77.

[28] 何勇,张淑娟,方慧.基于人工神经网络的田间信息插值方法研究[J].农业工程学报,2004,20(3):120-123. [He Y, Zhang S J, Fang H. Interpolation method of field information based on the artificial neural network[J]. Transactions of The Chinese Society of Agricultural Engineering, 2004,20(3):120-123.]

[29] Chen D, Chang N, Xiao J, et al. Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms[J]. Science of The Total Environment, 2019,669:844-855.

[30] Ransom K M, Nolan B T, A. Traum J, et al. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA[J]. Science of The Total Environment, 2017,601-602:1160-1172.

[31] Knoll L, Breuer L, Bach M. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning[J]. Science of The Total Environment, 2019,668:1317-1327.

[32] Saurabh Das, Rohit Chakraborty, Animesh Maitra. A random forest algorithm for nowcasting of intense precipitation events[J]. Advances in Space Research, 2017,60(6):1271-1282

[33] Bonelli M G, Ferrini M, Manni A. Artificial neural networks to evaluate organic and inorganic contamination in agricultural soils[J]. Chemosphere, 2017,186:124-131.

[34] 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 N, 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.]

[35] Khosravi V, Doulati Ardejani F, Yousefi S, et al. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods[J]. Geoderma, 2018,318:29-41.

[36] GB 9834-88.土壤有机质测定法[S]. [GB 9834-88. Method for determination of soil organic matter[S].]

[37] 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.

[38] 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.

[39] 许信旺.不同尺度区域农田土壤有机碳分布与变化[D].南京:南京农业大学,2008. [Xu X W, Regional distribu-

tion and variation of SOC storage in agricultural soils at different scales[D]. Nanjing Agricultural University, 2008.]

[40] 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预测[J].环境科学研究,2019,32(8):1365-1373. [Gao F J, Wu X, Shi H D, et al. Prediction of spatial distribution of soil organic matter in a mollisol watershed of China based on BME method[J/OL]. Research of Environmental Sciences, 2019,32(8):1365-1373.]

[41] GB 18596-2001.畜禽养殖业污染物排放标准[S]. [GB 18596- 2001. Discharge standard of pollutants for livestock and poultry breeding[S].]

[42] 李帷,李艳霞,杨明,等.北京市畜禽养殖的空间分布特征及其粪便耕地施用的可达性[J].自然资源学报,2010,25(5):746-755. [Li W, Li Y X, Yang M, et al. Spatial distribution of livestock and poultry production and land application accessibility of animal manure in Beijing[J]. Journal of Natural Resources, 2010,25(5):746-755.]

[43] 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. [Ren L, Yang L A, Wang H, et al. Spatial prediction of soil organic matter in apple region based on random forest [J]. Journal of Arid Land Resources and Environment, 2018,32(8):141-146.]

[44] 沈掌泉,周斌,孔繁胜,等.应用广义回归神经网络进行土壤空间变异研究[J].土壤学报,2004,41(3):471-475. [Shen Z Q, Zhou B, Kong F S, et al. Study on spatial variety of soil properties by means of generalized regression neural network[J]. Acta Pedologica Sinica, 2004,41(3):471-475.]

[45] 范曼曼,吴鹏豹,张欢,等.采样密度对土壤有机质空间变异解析的影响[J].农业现代化研究,2016,37(3):594-600. [Fan M M, Wu P B, Zhang H, et al. Effect of sampling density on spatial variability analysis of soil organic matter[J]. Research of Agricultural Modernization, 2016,37(3):594-600.]

[46] 姜怀龙,李贻学,赵倩倩.县域土壤有机质空间变异特征及合理采样数的确定[J].水土保持通报,2012,32(4):143-146. [Jiang H L, Li Y X, Zhao Q Q. County-scale spatial variability of soil organic matter distribution and determination of reasonable sampling density[J]. Bulletin of Soil and Water Conservation, 2012,32(4):143-146.]