



论文

我国东北多年冻土区土壤碳储量及其空间分布特征

杨艳^{1,2}, 贾丙瑞^{1,2}, 赫楠^{1,2,3}, 张帆^{1,2,3}, 张典业^{1,2}, 杨元合^{1,2,3*}

1. 中国科学院植物研究所, 植被与环境变化重点实验室, 北京 100093

2. 国家植物园, 北京 100093

3. 中国科学院大学资源与环境学院, 北京 101408

* 联系人, E-mail: yhyang@ibcas.ac.cn

收稿日期: 2025-02-10; 接受日期: 2025-05-06; 网络版发表日期: 2025-07-31

国家重点研发计划(批准号: 2022YFF0801901)资助

摘要 气候变暖导致的多年冻土融化会加速土壤碳释放, 形成冻土碳循环与气候变暖之间的正反馈效应. 针对我国东北多年冻土区土壤碳库特征不明确的问题, 本研究基于800 km样带调查, 结合文献数据, 构建了多种机器学习模型(随机森林、支持向量机、增强回归树和极端梯度提升), 最终采用最优模型(极端梯度提升)系统解析了东北多年冻土区土壤有机碳密度(soil organic carbon density, SOCD)的空间格局. 结果表明, 气候因子(干旱指数)和土壤属性(全氮)共同解释了SOCD空间变异的61.7%; 0~30和0~50 cm土层的SOCD均值分别为(9.8±2.5)和(12.7±2.9) kg C/m², 对应碳储量为(5.5±1.4)和(7.1±1.6) Pg C; SOCD空间分布呈现从西南向东北递增的趋势, 且不同植被类型间存在显著差异(湿地(11.6±2.4) kg C/m²>森林(10.4±2.0) kg C/m²>草地(8.1±2.6) kg C/m²). 上述结果表明东北多年冻土区是个重要碳库, 在持续气候变暖背景下可能成为显著碳源, 未来亟需重点关注该区域碳循环关键过程对气候变化的响应机制.

关键词 土壤有机碳密度, 土壤有机碳储量, 碳循环, 碳-气候反馈, 多年冻土, 机器学习

多年冻土是指连续两年及两年以上处于0℃以下的岩石和土壤, 主要分布于北半球高纬度和高海拔等寒冷地区, 其面积约占全球陆地面积的15%^[1]. 受低温环境的限制, 多年冻土区土壤有机质分解缓慢, 导致土壤碳的积累和储存^[2~4], 其土壤有机碳(soil organic carbon, SOC)储量约占全球3 m深度SOC库总量的1/3以上^[5,6]. 然而, 这一碳库正面临气候变暖的威胁. 值得注意的是, 多年冻土区的气候变暖速率显著高于其他地区^[7,8], 其升温幅度约为全球平均值的2倍^[9], 使其成为对全球气候变化最为敏感的区域之一^[10]. 气候变暖

导致多年冻土融化, 使得长期封存在多年冻土中的有机质被微生物分解, 并以二氧化碳(CO₂)和甲烷(CH₄)等温室气体的形式释放至大气. 这一过程会进一步加剧气候变暖, 进而导致冻土碳循环与气候变暖之间的正反馈效应^[11,12]. 因此, 评估多年冻土区土壤碳库大小及其空间分异特征是准确理解冻土碳-气候反馈关系方向和强度的前提和基础^[11,13].

鉴于其重要性, 近年来多年冻土区土壤碳储量估算受到学术界广泛关注, 并取得了显著进展. 现有研究表明, 北极多年冻土区0~3 m深度土壤碳储量为

引用格式: 杨艳, 贾丙瑞, 赫楠, 等. 我国东北多年冻土区土壤碳储量及其空间分布特征. 中国科学: 生命科学, 2025, 55: 1716–1730

Yang Y, Jia B R, He N, et al. Soil carbon storage and its spatial distribution in the permafrost region of Northeast China (in Chinese). Sci Sin Vitae, 2025, 55: 1716–1730, doi: 10.1360/SSV-2025-0026

1014~1035 Pg C (1 Pg=10¹⁵ g), 而3 m以下的苔原富冰黄土(yedoma deposits)和河流三角洲沉积物(deltaic deposits)分别储存了181~407和91~241 Pg的土壤有机碳^[2,5,6]. 同时, 我国多年冻土区土壤碳储量估算也取得了积极进展. 近期的估算结果显示, 青藏高原多年冻土区0~3 m深度的土壤碳储量为33.0~36.6 Pg C^[14,15]. 然而, 和青藏高原多年冻土区相比, 目前仅有少量研究调查了东北大兴安岭多年冻土区0~20 cm表层土壤的碳储量^[16], 尚不清楚整个东北多年冻土区的碳储量大小. 特别是, 全球尺度多年冻土区土壤碳数据库中并未包括源自东北多年冻土区的观测资料. 东北多年冻土区位于欧亚大陆高纬度多年冻土区的南缘^[17], 是我国冻土生态系统的重要组成部分. 近50年来, 该区域气温上升了0.9~2.2℃^[18,19], 属于气候变化的敏感区域^[20]. 和环北极多年冻土区类似, 该区域也储存着大量碳^[21], 进而为研究高纬度多年冻土区土壤碳的空间分异特征及其影响因素提供了理想平台. 然而, 现有研究对东北多年冻土区土壤碳储量的估算仅限于表层(0~20 cm)^[16,21], 且仅依据林龄预测土壤有机碳密度^[22], 忽略了其他关键因素如植被和土壤理化性质的影响.

为此, 本研究于2023年7~8月在东北多年冻土区开展了区域尺度(30个样点)的土壤剖面调查(0~10, 10~20, 20~30以及30~50 cm深度), 并结合已发表的文献数据构建了迄今为止较为全面的东北多年冻土区土壤碳数据库. 在此基础上, 本研究利用147个样点的土壤碳实测数据, 并以气候变量、植被属性以及土壤属性作为预测变量, 采用机器学习和经验模型相结合的方法构建了土壤碳密度(0~30和0~50 cm)的空间分布预测模型, 并绘制了不同深度(0~10, 0~30和0~50 cm)土壤有机碳密度的空间分布图. 基于机器学习模型, 本研究进一步揭示了影响土壤有机碳库分布的主要环境因子及其相对重要性, 并识别出相对不确定性较高的区域. 此外, 本研究对比了高纬度和高海拔多年冻土区的土壤有机碳密度及储量差异. 研究结果旨在为准确评估东北多年冻土区土壤碳库提供数据基础和方法借鉴, 同时也为进一步认识东北多年冻土区碳循环对气候变暖的响应奠定理论基础.

1 材料与方法

1.1 研究区概况

东北多年冻土区(47°37'~53°26'N, 117°32'~129°

22'E)属于寒温带半湿润气候, 平均海拔为539 m. 研究区近20年的年均气温和年均降水量分别为-4.3~1.7℃和236~680 mm. 该区域大部分活动层厚度范围为1.4~1.8 m, 冻土类型主要为不连续多年冻土和岛状冻土^[23]. 植被类型以典型的落叶针叶林和针阔混交林为主, 部分区域分布着少量草原和湿地. 其中, 森林主要分布在降水量较为丰富的东北地区, 乔木层优势树种为兴安落叶松(*Larix gmelini*), 同时伴生少量白桦(*Betula platyphylla*)和蒙古栎(*Quercus mongolica*), 灌丛则以杜鹃(*Rhododendron simsii*)和杜香(*Ledum palustre*)为主. 草地主要分布于降水量相对较少的西南地区, 其优势种包括羊草(*Leymus chinensis*), 大针茅(*Stipa grandis*)和冷蒿(*Artemisia frigida*). 土壤类型以Cambisol和Chernozems为主^[24].

为揭示东北多年冻土区土壤碳库的空间分布特征, 在国家重点研发计划项目“我国冻土生态系统碳氮磷循环过程、机理及演化趋势”的资助下, 中国科学院植物研究所杨元合课题组联合植物研究所贾丙瑞老师、北京师范大学沈妙根课题组于2023年在该区域开展了800 km的冻土样带调查. 本次调查覆盖了东北地区典型的不连续多年冻土区与岛状冻土区, 旨在系统刻画东北多年冻土区土壤碳储量的空间分布规律和影响因素.

1.2 土壤样品采集与测定

2023年7~8月, 课题组联合相关团队在东北多年冻土区均匀布设30个样点开展调查, 包括23个森林, 4个草地和3个湿地样点(图1)^[25]. 样点的分布比例与各植被类型在该区域的实际面积^[26]比例基本一致. 具体采样方法如下: 在森林区域, 每个样点设置3个20 m×30 m的样方, 并沿每个样方的对角线两端及中心位置采集0~10, 10~20, 20~30和30~50 cm的土壤样品. 在草地和湿地区域, 每个样点内布设1个10 m×10 m的样方, 并沿样方的对角线设置3个1 m×1 m的小样方. 在每个小样方的中心, 采集与森林相同深度的土壤样品, 同时采用环刀法测定各土层的土壤容重. 需要说明的是, 鉴于湿地中积水较多导致常规环刀法难以实施, 本研究采用切割法获取30 cm×30 cm左右的结构完整的土块. 之后, 基于采集的土块使用环刀完成容重样品采集.

采集的土壤样品运送至中国科学院植物研究所植

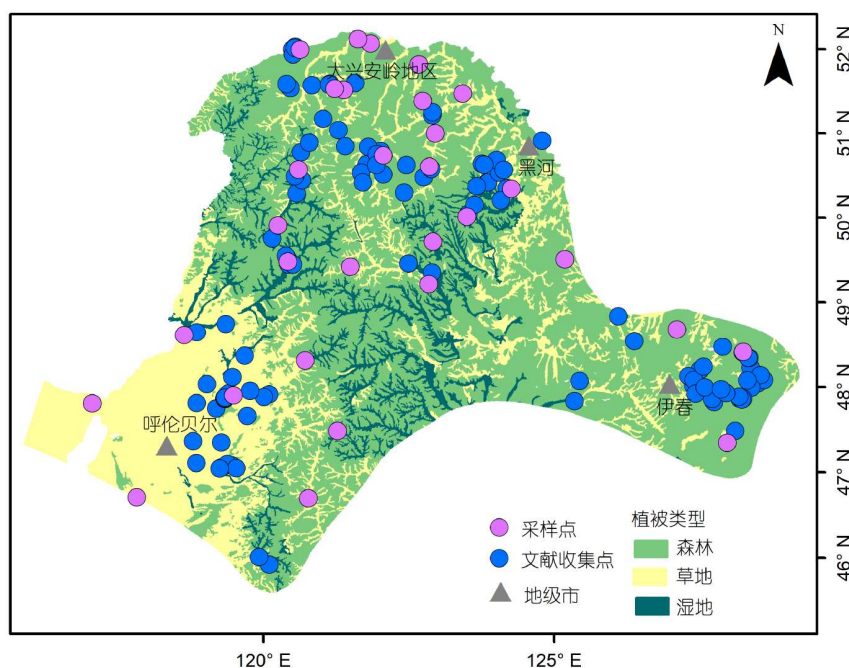


图 1 东北多年冻土区样点分布图. 图中紫色点为样带调查样点, 蓝色点为文献收集样点. 图中东北多年冻土分布范围来源于美国宇航局国家冰雪数据中心^[25]. 植被类型来源于中国科学院植物科学数据中心^[26]. 收集样点的详细文献列表见附录

Figure 1 Distribution of sampling sites in the northeastern permafrost region. The purple dots represent transect survey sites, while the blue dots indicate sites from literature. The map of permafrost distribution in northeastern China is sourced from the National Snow and Ice Data Center (NSIDC)^[25], NASA. Vegetation types are derived from Plant Science Data Center, Chinese Academy of Sciences^[26]. A detailed literature list of collection points was provided in the Appendix

被与环境变化国家重点实验室. 样品经自然风干后, 通过2 mm筛去除石块和根系碎屑, 用于后续理化性质分析. 鉴于该区域土壤基本为酸性^[27], 土壤无机碳含量可忽略不计. 进一步针对无机碳含量的检测结果也证实了这一现象. 随后, 使用Vario EI III元素分析仪 (Elementar, Germany)测定了SOC含量. 土壤容重通过环刀法测定, 即将环刀内样品在105℃条件下烘干至恒重后称重, 结合环刀体积计算容重. 同时, 基于排水法测定土壤样品中的砾石体积. 需要说明的是, 为降低空间异质性对SOC测定结果的影响, 本研究对每个样点的土壤样品进行等质量混匀. 其中, 森林样点的9个重复样品以及草地和湿地样点的3个重复样品分别等量混匀后进行SOC测定.

1.3 数据收集与环境变量来源

为了进一步增加土壤有机碳密度(soil organic carbon density, SOCD)估算的观测数据, 本研究通过Web of Science和Google Scholar以“soil organic matter” or “soil organic carbon” or “soil C storage”为关键词, 检索

截至2024年11月发表的相关文献. 同时, 在中国知网数据库中, 以“土壤有机质”或“土壤有机碳”或“土壤碳储量”为关键词检索了相关文献. 检索到的文献按照以下标准进行筛选: (i) 研究地点须包含经纬度信息, 且位于东北多年冻土区; (ii) 文献中须列出土壤层次和植被类型; (iii) 对于包含实验处理的文献, 仅保留对照组的土壤有机碳或有机质数据, 剔除处理数据. 文献中以图片形式展示的SOC含量数据, 使用Getdata v2.24工具 (<https://getdata.sourceforge.net/download.html>)进行提取. 最终, 筛选出符合标准的文献共37篇(英文17篇, 中文20篇, 详见网络版附录1). 此外, 本文涉及的部分有机碳数据来自开放数据库^[28]和第二次土壤普查^[29]. 总体而言, 本研究共涵盖147个采样点, 包括78个森林样点, 25个草地样点以及44个湿地样点, 共计440条数据. 其中, 约4/5的采样点位于岛状多年冻土区, 与东北多年冻土区冻土类型的面积比例^[25]相近. 基于这些数据, 本研究构建了东北多年冻土区0~10, 10~20, 20~30和30~50 cm深度的SOC含量数据库, 并根据式(1)~(3)^[28]计算了每个采样点的SOCD(kg C/m²)(网络版附图1),

以及整个研究区的碳储量(Pg C).

$$SOC_D = \sum_{i=1}^n SOC_i \times BD_i \times D_i \times (1 - \delta_i) \times 0.01, \tag{1}$$

$$SOC_i = SOM_i \times 0.58, \tag{2}$$

$$SOC_S = A \times SOC_D \times 10^{-6}, \tag{3}$$

式中, SOC_i , BD_i , D_i , δ_i , SOM_i 分别表示第*i*层土壤SOC含量(g/kg)、容重(g/cm³)、土壤深度(cm)、粒径>2.0 mm 砾石的体积百分比(%)和土壤有机质(SOM)含量(g/kg); *n*为土层数. SOC_S 表示土壤碳储量(Pg C), *A*为给定网格的面积(km²). 式(2)中的常数0.58用于将土壤SOM转换为SOC^[30]. 需要说明的是, 对于缺少容重记录的土壤样品, 本研究利用土壤SOC与容重的经验指数方程(网络版附图2)和相应的SOC含量估算了缺失的容重^[31]. 此外, 对于缺少砾石含量记录的样点, 本研究采用相同土壤类型的砾石体积百分比平均值进行替代^[32].

为探究气候因素对东北冻土区土壤碳库的潜在影响, 本研究采用了多源气候数据(表1). 具体而言, 基于各采样点以及文献搜集点的地理坐标, 从国家青藏高原科学数据中心获取2000~2020的年均气温(MAT,

<https://data.tpdac.ac.cn/zh-hans/data/71ab4677-b66c-4fd1-a004-b2a541c4d5bf>)和年均降水量(MAP, <https://data.tpdac.ac.cn/zh-hans/data/faae7605-a0f2-4d18-b28f-5cee413766a2>)数据. 同时, 从全球干旱指数和潜在蒸散气候数据库提取干旱指数(AI, <https://cgiaresci.com-munity/2019/01/24/global-aridity-index-and-potential-evapotranspiration-climate-database-v3/>). AI值越低, 表示干旱程度越高.

为揭示植被属性对东北冻土区土壤碳库的潜在影响, 本研究基于NASA MOD13A3数据集提取2000~2020年的平均归一化植被指数(NDVI, <https://search.earthdata.nasa.gov/search>), 并从地球资源数据云平台获取2000~2020年平均增强型植被指数(EVI, <http://www.gis5g.com/data/zbsj/EVI?id=2126>)用于表征研究区的植被覆盖状况. 同时, 为了表征地上碳输入, 本研究从NASA地球科学数据网获取2000~2020年平均净初级生产力(NPP, <https://lpdaac.usgs.gov/products/mod17a3hgv061/>)数据. 植被类型数据来源于中国科

表 1 本研究使用的土壤有机碳密度预测变量
Table 1 Environmental variables used for SOCD prediction

环境变量	变量	空间分辨率	时间	数据来源
气候变量	年平均气温(MAT)	1 km	2000~2020	国家青藏高原科学数据中心 (https://data.tpdac.ac.cn/zh-hans/data/71ab4677-b66c-4fd1-a004-b2a541c4d5bf)
	年均降水量(MAP)	1 km	2000~2020	国家青藏高原科学数据中心 (https://data.tpdac.ac.cn/zh-hans/data/faae7605-a0f2-4d18-b28f-5cee413766a2)
	干旱指数(AI)	1 km	2022	全球干旱指数和潜在蒸散(ET0)数据库v3 (https://csidotinfo.wordpress.com/2019/01/24/global-aridity-index-and-potential-evapotranspiration-climate-database-v3/)
植被属性	归一化植被指数(NDVI)	1 km	2000~2020	NASA MOD13A3数据集 (https://search.earthdata.nasa.gov/search)
	增强型植被指数(EVI)	1 km	2000~2020	地球资源数据云平台 (http://www.gis5g.com/data/zbsj/EVI?id=2126)
	净初级生产力(NPP)	500 m	2000~2020	NASA地球科学数据网 (https://lpdaac.usgs.gov/products/mod17a3hgv061/)
	植被类型图	10 km	2007	中国科学院植物科学数据中心 (https://www.plantplus.cn/doi/10.12282/plantdata.0155)
土壤属性	黏粉粒含量(Clay+silt)	1 km	2022	国家地球系统科学数据中心 (http://soil.geodata.cn/data/datadetails.html?data-guid=36810085119113)
	土壤全氮(TN)	1 km	2022	国家地球系统科学数据中心 (http://soil.geodata.cn/data/datadetails.html?data-guid=36810085119113)
	土壤pH	1 km	2022	国家地球系统科学数据中心 (http://soil.geodata.cn/data/datadetails.html?data-guid=36810085119113)

学院植物科学数据中心^[26](<https://www.plantplus.cn/doi/10.12282/plantdata.0155>). 在此基础上, 结合实际调查的植被类型, 本研究将研究区域划分为森林、草地和湿地三种类型. 此外, 为解析土壤属性对东北多年冻土区土壤碳库的潜在影响, 本研究从国家地球系统科学数据中心获取东北多年冻土区表层土壤的黏粉粒含量、土壤全氮(TN)和pH栅格数据(<http://soil.geodata.cn/data/datadetails.html?dataguid=36810085119113>). 需要指出的是, 实测点上的TN含量通过Vario EI III元素分析仪(Elementar)测定, 文献收集点中缺失的TN数据源自国家地球系统科学数据中心(<http://soil.geodata.cn/data/datadetails.html?dataguid=36810085119113>).

1.4 模型预测过程与精度检验

本研究利用四种机器学习方法: 随机森林(random forest, RF)^[33]、支持向量机(support vector machine, SVM)^[34]、增强回归树(boosted regression tree, BRT)^[35]以及极端梯度提升(extreme gradient boosting, XGBoost)^[36], 建立了表层土壤(0~10 cm)有机碳密度的预测模型. 这些方法能够捕捉复杂的非线性关系, 具有较高的灵活性和适应性. 其中, 随机森林是一种基于多棵决策树集成的预测方法, 通过随机选择辅助变量进行模型训练并预测目标变量. 增强回归树采用随机梯度增强提高建模性能. 支持向量机通过将数据映射到高维或无限维空间中寻找最佳分割超平面, 从而简化复杂的非线性关系^[37,38]. XGBoost是一种基于梯度增强决策树的高效方法, 通过引入正则化项控制模型复杂性并防止过拟合^[39,40].

各模型均采用网格搜索法进行调参^[34]. 特别是, 为优化XGBoost模型性能, 本研究对主要超参数(如学习率、树的最大深度、每棵树的样本比例、最大提升迭代次数)进行了调优. 随后, 为评估各模型的预测精度, 采用10倍交叉验证方法^[41], 并以100次迭代结果的平均值作为最终预测结果(空间分辨率为1 km). 模型性能通过决定系数(R-Square, R^2), 均方根误差(root mean square error, RMSE), 以及平均绝对误差(mean absolute error, MAE)进行评价, 计算公式^[42,43]如下:

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}}, \quad (5)$$

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n}, \quad (6)$$

式中, n 为样本量, i 为样本编号, O_i , P_i 以及 $\bar{O}$ 分别表示实测值, 模型预测值以及实测值的平均值. 最终, 通过比较各模型的 R^2 , RMSE以及MAE, 本研究选择最优预测模型(XGBoost)对土壤表层SOC(0~10 cm)进行空间格局模拟.

为了评估上述模型的不确定性, 采用100次bootstrap抽样作为非参数重复抽样方法^[44,45]. 通过随机分离建模数据集100次, 每次构建预测模型, 并以100次建模结果的平均值作为最终预测结果. 同时, 计算预测结果在95%置信区间(CIs)内的可能性, 以量化预测结果的不确定性, 其计算公式^[46]如下:

$$CIs = \bar{Y} \pm a \times \frac{SD}{\sqrt{n}}, \quad (7)$$

$$Uncertainty = \frac{CI_{upper} - CI_{lower}}{\bar{Y}}, \quad (8)$$

式中, $\bar{Y}$ 表示100次估算中土壤SOC的平均值, SD表示标准差, n 表示重复次数($n=100$). a 为正态分布在95%置信水平下的临界值, 取值为1.96, CI_{upper} 和 CI_{lower} 分别表示置信区间的上限和下限. 需要说明的是, 深层土壤中有限的样本量会在很大程度上制约机器学习方法的交叉验证效果. 因此, 为了获取东北多年冻土区不同深度土壤(0~30和0~50 cm)SOC的空间分布, 本研究参考Ding等人^[38]的研究, 基于实测数据和文献收集的0~10, 10~20, 20~30以及30~50 cm深度数据, 计算了0~30和0~50 cm深度的SOC. 随后, 本研究构建了表层(0~10 cm)和不同深度层(0~30和0~50 cm)的线性回归模型, 并利用该模型将表层SOC外推至更深的土层, 从而揭示其空间分布格局.

1.5 统计分析

本研究中涉及的数据分析主要包括以下三个方面. 首先, 为避免各环境变量对土壤有机碳密度的过度拟合, 本研究通过Spearman相关分析筛选与土壤有机碳密度显著相关的解释因子^[47], 并剔除年均温和黏粉粒含量等不相关变量. 随后, 对各筛选后的变量进行共线性检验, 各变量的方差膨胀因子(VIF)均 <10 ^[48]. 将筛选后的变量纳入模型预测, 并使用“varImp”函数评估最佳预测模型(XGBoost)中各解释因子的相对重要性.

评估结果以百分比形式表示, 所有变量的重要性之和为100%. 为进一步揭示SOC D对各解释因子变化的响应模式, 使用“iml”包中“pdb”函数对模型响应的异质性进行了可视化分析. 通过生成部分依赖图(partial dependence plot, PDP), 明确展示了主要预测因子对SOC D的影响, 同时考虑其他因子的平均效应^[45].

其次, 基于线性回归分析构建了10倍交叉验证结果与实测值之间的相关关系, 以评估模型的预测性能. 最后, 为比较不同植被类型下土壤SOC D的空间分布差异, 采用非参数Kruskal-Wallis检验, 并结合事后Dunn检验对组间差异进行显著性检验. RF, SVM, BRT以及XGBoost模型分别基于的R语言中的“randomForest”和“caret”包、“gbm”包、“e1071”包、“xgboost”包实现. 所有统计分析均在R 4.3.2^[49]中进行.

2 结果与讨论

2.1 东北多年冻土区土壤有机碳库的空间格局

为了揭示SOC D的空间分布特征, 本研究使用与表层(0~10 cm)SOC D显著相关的环境变量进行建模. 从各模型的交叉验证结果来看, 与其他机器学习模型(RF, BRT和SVM)相比, XGBoost在预测性能上表现最佳, 其 R^2 更高, 且RMSE和MAE相对较低(表2, 图2A). 因此, 本研究基于XGBoost模型预测了东北冻土区0~10 cm SOC D的空间分布, 并进一步通过建立0~10 cm SOC D与0~30和0~50 cm SOC D的线性回归模型推算出0~30和0~50 cm的SOC D的分布(图2, 网络版附图3, 均 $P<0.001$). 结果显示, 0~10 cm土壤SOC D范围为0.1~8.0 kg C/m², 平均值为(4.2±1.1) kg C/m²(平均值±标准差, 下同), 其中60.2%分布在4~6 kg C/m²之间(图3A); 0~30和0~50 cm的SOC D在0.3~18.2和1.7~22.5 kg C/m²之间(图3B, 网络版附图3).

SOC D呈现从西南向东北逐渐增加的趋势(图3), 这与Li等人^[16]和Tang等人^[50]报道的结果一致. 并且, 不同植被类型间的SOC D存在显著差异($P<0.05$), 其中表层土壤碳密度的分布趋势为: 湿地((5.0±1.1) kg C/m²)>森林((4.5±0.8) kg C/m²)>草地((3.5±1.1) kg C/m²)(图3C). 与表层类似, 0~30 cm SOC D湿地最高((11.6±2.4) kg C/m²), 其次为森林((10.4±2.0) kg C/m²)和草地((8.1±2.6) kg C/m²)(图3E). 进一步研究显示, SOC D相对不确定性较高的区域主要集中在湿地以及草地和森林的过渡带, 而森林土壤的不确定性相对较低(图4). 这可能与湿地季节性水淹和非水淹状态的变化^[51], 以及林草过渡带较高人为干扰的影响有关^[52]. 研究区0~10, 0~30和0~50 cm的土壤有机碳储量分别为(2.4±0.6), (5.5±1.4) Pg C(图3D和F)和(7.1±1.6) Pg C. 对于0~30 cm土壤碳库而言, 森林贡献最大((3.6±0.7) Pg C), 其次是草地((1.3±0.4) Pg C)和湿地((0.6±0.1) Pg C)(图3F). 森林碳储量占比高的主要原因可能与其分布面积较大有关. 尽管湿地的SOC D显著高于森林土壤, 但在东北冻土区森林面积约为湿地面积的7倍^[26]. 因此, 森林生态系统的广泛分布使其在东北多年冻土区SOC储量中占据主导地位.

2.2 东北多年冻土区土壤有机碳库的驱动因素

在构建机器学习模型之前, 本研究基于Spearman相关分析筛选了与表层土壤(0~10 cm) SOC D显著相关的环境变量. 结果表明, SOC D与部分气候属性(MAP和AI)、植被属性(NDVI, EVI和NPP)以及全氮含量呈显著正相关, 与土壤pH呈显著负相关, 而与MAT及黏粉粒含量无显著相关性(网络版附图4). 基于筛选的相关变量, 本研究进一步采用XGBoost模型评估了各变量的相对重要性. 模型结果显示, 全氮含量是影响东北多年冻土区SOC D的最主要的协变量, 可解释41.4%的变异(图5A). 部分依赖分析进一步验证

表2 不同机器学习方法预测的0~10 cm土壤有机碳密度交叉验证结果比较

Table 2 Comparison of cross-validation results of SOC D at 0~10 cm predicted by different machine learning methods

模型	决定系数	均方根误差	平均绝对误差
RF	0.45	1.52	1.15
BRT	0.41	1.54	1.20
SVM	0.40	1.59	1.20
XGBoost	0.47	1.52	1.20

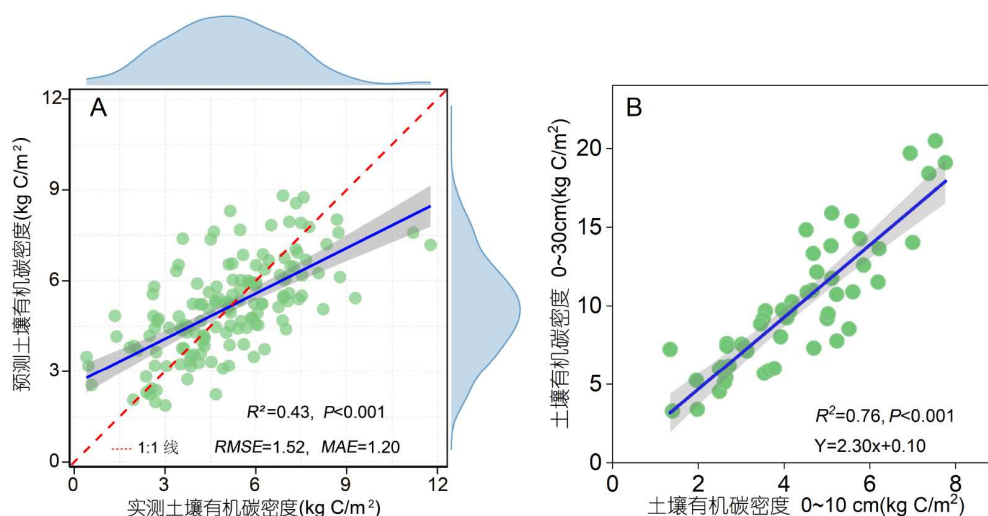


图2 基于机器学习极端梯度提升(XGBoost)模型的SOCd与实测SOCd值的关系。A: 十折交叉验证结果; B: 0~30与0~10 cm 实测SOCd值之间的关系。蓝色实线表示显著的线性回归, 灰色阴影区域为95%置信区间。图A中蓝色区域表示SOCd密度分布
Figure 2 Relationship between SOCd based on the XGBoost machine learning model and measured SOCd values. A: 10-fold cross-validation results; B: relationship between measured SOCd at depths of 0~30 and 0~10 cm. The blue solid line denotes the significant linear regression, and the grey shaded area is 95% confidence interval. The blue area in panel (A) shows the density distribution of SOCd

了这一点, 表明随着全氮含量的增加, SOCd显著上升(图5B)。这一现象主要与土壤碳氮之间的化学计量关系有关^[53]。

除TN之外, AI解释了20.3%的SOCd变异(图5A), 且对SOCd的影响呈现非线性的趋势(图5C)。AI的显著作用表明, 水分条件是影响土壤SOCd的重要因素。由于多年冻土区水分条件的空间异质性较大, 区域内不同水热条件可能导致SOCd的显著空间分异^[6,16]。根据联合国环境规划署(UNEP)^[54]的分类标准, AI处于0.3~0.5时为半干旱系统, 0.5~0.65为半湿润系统, 而AI>0.65为湿润系统。本研究的研究发现, 当AI>0.5时, SOCd显著增加(图5C)。这可能由于较高的水分可利用性, 使得湿润期更长, 为植物根系生长提供有利条件, 并促进底物输入的垂直和横向扩散, 进而增强与矿质结合态有机质的相互作用, 最终促进土壤有机碳的积累^[55]。

总体来看, 土壤属性对SOCd变异的解释率最高(47.6%), 气候属性和植被属性的贡献相当, 分别为26.6%和25.8%。这一结果与以往研究认为, SOCd在大尺度空间格局上主要受气候调节的结论存在差异^[56,57], 可能是由于东北多年冻土区气候梯度较小, 且特有的低温条件使得土壤有机质分解具有对低温的长期适应性, 土壤碳储量受土壤属性的影响更为显著^[58]。

2.3 东北多年冻土区土壤有机碳储量与其他冻土区的比较

为了揭示不同冻土区SOCd的差异, 本研究比较了东北多年冻土区、青藏高原多年冻土区以及北极多年冻土区的0~30 cm SOCd。结果表明, 环北极多年冻土区的SOCd最高(13.0 kg C/m², 95%CI: 10~25^[5]和17.3 kg C/m², 95%CI: 15.3~19.3 kg C/m²^[6]), 东北多年冻土区次之(本研究为9.8 kg C/m², 95%CI: 9.4~10.2 kg C/m²), 均高于青藏高原多年冻土区(6.0 kg C/m², 95%CI: 3.5~9和3.9 kg C/m², 95%CI: 3.4~4.4 kg C/m²)^[6,38] (图6A)。上述差异可能与以下三个方面有关。首先, 植被类型分布是导致上述差异的关键因素。环北极多年冻土区主要以森林、苔原和湿地生态系统为主^[59], 东北冻土区作为环北极多年冻土区的南部延伸, 其植被也以森林为主^[27], 而青藏高原多年冻土区则主要以草地为主^[60,61]。因此, 环北极和东北多年冻土区的植被碳输入(分别为315.7和425.5 g/m²)^[62,63]显著高于青藏高原多年冻土区(68.8 g/m²)^[57]。

其次, 泥炭地分布是导致不同多年冻土区土壤碳库差异的另外一个重要因素。泥炭地约占环北极多年冻土区面积的13%^[5], 在东北多年冻土区中占比为10%^[26], 而青藏高原泥炭地仅占其总面积的0.2%^[64]。

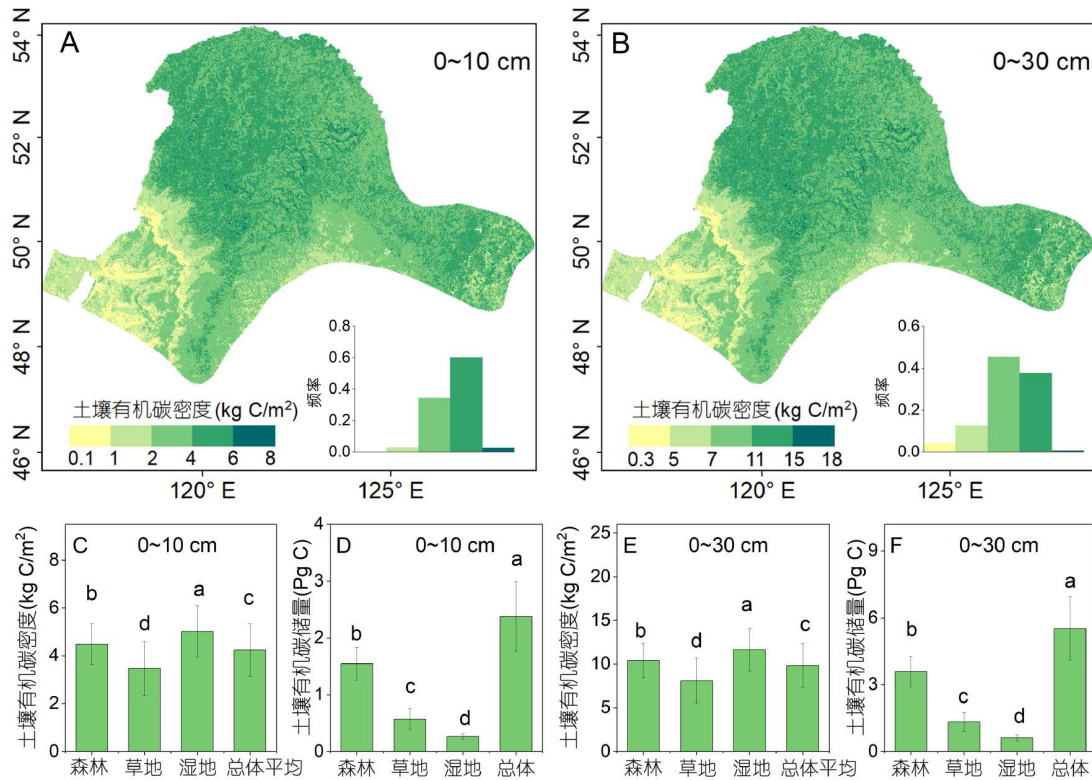


图 3 基于极端梯度提升(XGBoost)预测的东北多年冻土区SOCD的空间分布, 以及不同植被类型下的SOCD均值和碳储量. 基于线性回归外推得到的0~10 cm (A)和0~30 cm (B) SOCD空间分布. 空间分辨率为1 km. 直方图展示了各层次土壤有机碳密度的频率分布. 各植被类型下的碳密度和储量0~10 cm (C, D), 0~30 cm (E, F). 柱状图以平均值±标准差表示, 不同的小写字母表示不同植被类型间存在显著差异($P < 0.05$)

Figure 3 Spatial distribution of SOCD in the permafrost region of Northeast China predicted by XGBoost machine learning model, and carbon density and carbon storage under different vegetation types. Spatial distribution of SOCD at 0~10 cm (A) and 0~30 cm (B) that was extrapolated using linear regression. The spatial resolution is 1 km. Histograms show the frequency distribution of SOCD values. Carbon density and storage across vegetation types are shown for 0~10 cm (C, D) and 0~30 cm (E, F). Bar plots display mean±standard deviation, with different lowercase letters indicating significant differences among vegetation types ($P < 0.05$)

泥炭地因其更强的持水能力和厌氧环境特性, 加之低温条件下的反复沉积作用^[65], 显著抑制了微生物活性, 从而抑制了有机质分解过程, 进一步促进了SOC的长期累积^[65,66].

再次, 气候条件的差异可能进一步放大了不同冻土区间的SOCD的差异. 环北极多年冻土区和东北多年冻土区较高的土壤水分($1.2 \text{ m}^3/\text{m}^3$, CMIP6, <https://aims2.llnl.gov/search>)也有助于抑制微生物分解活动, 降低有机质的分解速率^[2,5]. 相比之下, 青藏高原冻土区受较高海拔和强烈太阳辐射的影响, 土壤水分含量较低 ($0.3 \text{ m}^3/\text{m}^3$, 0~100 cm)^[67], 碳输入有限, 导致SOC的累积效率相对较低^[46]. 同时, 环北极冻土区的年均气温相较于青藏高原与东北多年冻土区更低(-6.9°C

vs. -1.9 和 -1.7°C)^[68,69]. 这些植被和气候条件的协同作用共同加剧了不同冻土区间SOCD的差异. 此外, 本研究结果显示, 东北多年冻土区的SOCD ($(9.8 \pm 2.5) \text{ kg C/m}^2$)高于从全国SOCD数据中提取的东北多年冻土区对应的SOCD($(7.7 \pm 2.1) \text{ kg C/m}^2$)^[70], 这一差异可能与本研究中涉及的样本量更大、代表性更强有关.

在上述研究的基础上, 本研究进一步比较了东北多年冻土区、青藏高原多年冻土区以及北极多年冻土区的0~30 cm土壤碳库大小(图6B). 结果发现, 环北极多年冻土区的储量最高($217 \sim 232 \text{ Pg C}$)^[5,6], 远超东北多年冻土区(5.5 Pg C , 本研究结果)和青藏高原多年冻土区($4.4 \sim 7.0 \text{ Pg C}$)^[6,38](图6B). 这种差异不仅与环北极

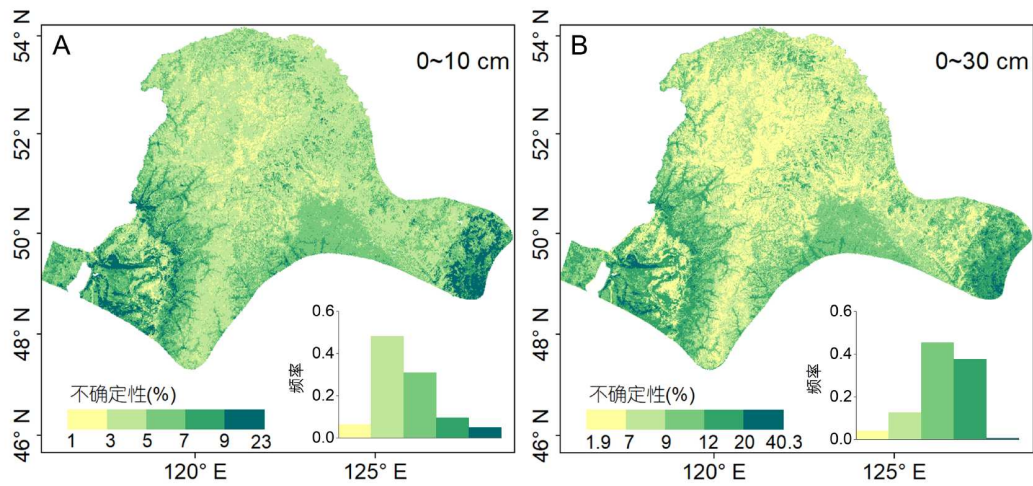


图 4 东北多年冻土区不同层次土壤有机碳密度的不确定性分布. A: 0~10 cm; B: 0~30 cm. 直方图展示了不确定性的频率分布
Figure 4 Uncertainty distribution of surface SOC density in the permafrost region of Northeast China. A:0~10 cm; B: 0~30 cm. The inserted histograms illustrate the frequency distribution of the corresponding uncertainty values

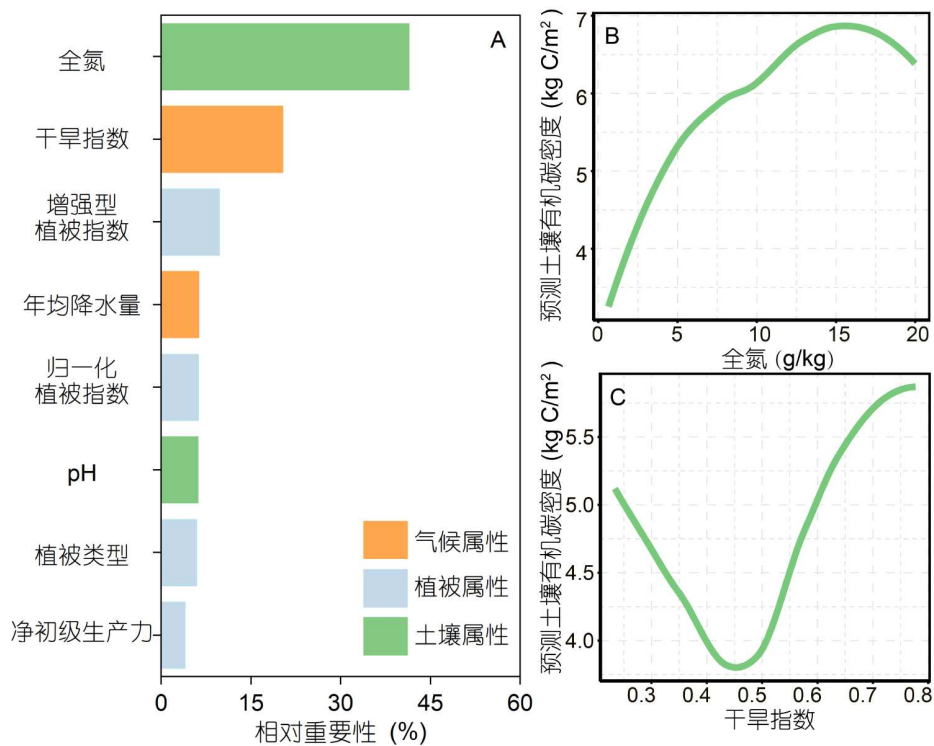


图 5 极端梯度提升(XGBoost)模型中各变量的相对重要性以及预测的SOCD与环境变量关系. A: 各变量的相对重要性, 包括气候变量(干旱指数和年均降水量); 植被属性(归一化植被属性, 增强型植被指数, 净初级生产力, 以及植被类型)和土壤属性(全氮, pH). B和C: SOCD与土壤全氮含量(B)以及干旱指数(C)之间部分依赖图

Figure 5 Relative importance of variables in the XGBoost model and the relationship between SOCD and environmental variables. A: Relative importance of variables, including climate variables (aridity index and mean annual precipitation), vegetation attributes (normalized difference vegetation Index, enhanced vegetation index, net primary productivity, and vegetation type), and soil properties (total nitrogen and pH); B and C: partial dependence plots showing relationships of SOCD with total nitrogen (B) and aridity index (C)

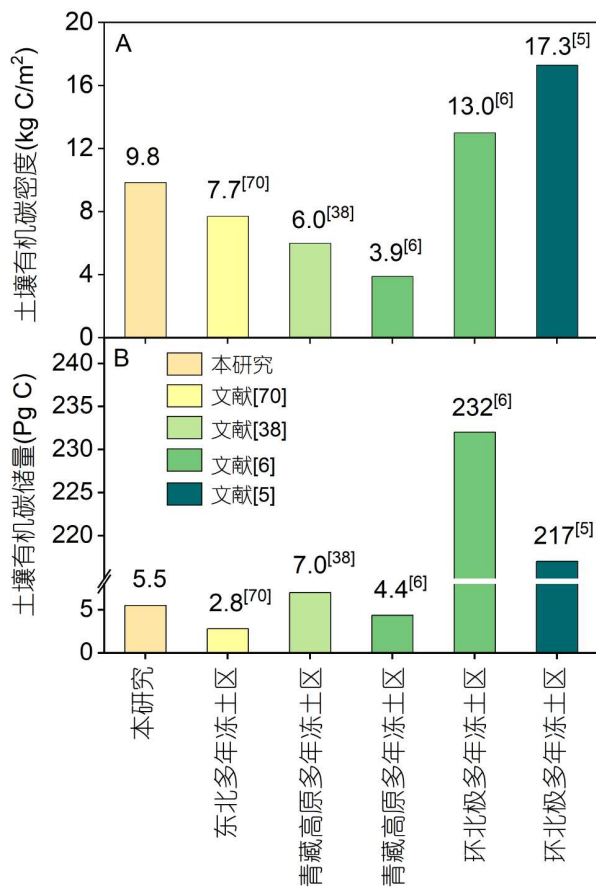


图 6 不同研究估算的高纬度和高海拔多年冻土区0~30 cm 深度SOCD和有机碳储量的比较. A: SOCD; B: 有机碳储量
Figure 6 Comparison of SOCD and soil organic carbon storage at 0~30 cm depth in high-latitude and high-altitude permafrost regions derived from various studies. A: SOCD; B: soil organic carbon storage

多年冻土区较高的 SOCD有关, 还与其更大的面积密切相关. 环北极多年冻土区的面积($3.6 \times 10^7 \text{ km}^2$)^[5,71]约为青藏高原($1.1 \times 10^6 \text{ km}^2$)^[38]和东北多年冻土区面积(本研究 $5.6 \times 10^5 \text{ km}^2$)总和的22倍, 导致其碳储量的主导地位. 总体而言, 植被类型、泥炭地分布、气候条件以及区域面积共同决定了不同多年冻土区之间SOCD和碳储量的差异.

2.4 不确定性分析

尽管XGBoost模型及其外推方法在东北多年冻土区土壤SOCD估算中表现良好, 但模型预测仍存在不确定性. 模型不确定性主要来源于以下三个方面. 首先, 实测SOC的不确定性. 如前所述, 本研究使用的部分数据源自第二次全国土壤普查数据, 其中剖面采样

基于土壤发生层次, 导致部分数据因层次与本研究的层次不一致而被剔除. 同时, 第二次全国土壤普查距今已近30年, 期间SOC的动态变化也可能会引入数据源的不确定性^[72]. 更为突出的是, 现有的文献研究大多集中于表层土壤, 东北冻土区深层土壤的样点显著不足^[20,28], 尤其是冻土层土壤的观测数据极为缺乏^[73,74], 这极大地限制了冻土层土壤碳库的估算. 考虑到深层土壤碳储量对区域土壤碳库的重要贡献, 其数据缺失加剧了模型预测的不确定性. 因此, 未来研究应统一采样标准, 合理覆盖不同土壤层次(包括活动层和冻土层)和冻土类型(包括岛状冻土和非连续冻土), 并增加深层土壤的采样密度, 以提高土壤碳库估算的准确性.

其次, 预测协变量的不确定性. 区域土壤碳库的估算精度一定程度上受限于输入数据的空间分辨率以及模型在不同尺度上的适用性. 然而, 目前所使用的遥感和环境数据(如气候因子、植被覆盖等)难以充分捕捉生态系统的空间异质性, 例如, 大尺度环境数据无法充分反映湿地与周边地区在地表覆盖度(如NDVI值)和气候条件(MAT, MAP)的差异^[75]. 这可能导致湿地生态系统的预测值略低于实测值, 进而影响区域SOC储量的准确性. 此外, 多年冻土生态系统的特殊性(如冻土类型、活动层深度和冻融交替频率)对土壤碳储存具有显著影响^[76]. 然而, 现有模型在预测协变量中未充分纳入这些关键因素, 可能导致预测结果的不确定性增加. 为此, 未来研究应注重多尺度数据的整合, 结合高分辨率遥感数据与地面实测数据, 并加强对活动层深度和冻融节律的监测, 将其作为关键协变量纳入模型, 以发展适用于多年冻土区土壤碳库的估算模型. 同时, 还需加强基于高分辨率数据的区域校准, 例如, 引入高分辨率土壤湿度数据^[77]优化模型输入, 从而显著提高区域土壤碳库估算结果的精度.

再次, 估算方法的不确定性. 机器学习方法(如XGBoost)在建模过程中受到方法本身的局限性影响, 例如, 模型选择、参数优化及输入变量依赖性问题, 可能导致预测结果出现差异^[78]. 为降低单一模型的不确定性, 未来可以通过融合多模型方法(如深度神经网络DNN^[42]、卷积神经网络CNN^[79,80])来综合不同模型的优点, 从而提升模型的稳定性和预测的准确性. 此外, 区域尺度土壤碳库估算还面临模型外推能力的限制. 现有模型在高异质性区域(如湿地、冻土区或土地

利用变化剧烈的地区)中的预测准确性较低^[6,81], 这主要源于模型训练数据与外推区域数据的不一致性。例如, 模型在样本量丰富的区域表现良好, 但在样本量稀少的区域(如冻土区深层土壤或特定湿地区域)因缺乏足够的训练数据而产生较大的预测偏差^[82]。这种外推能力的局限性进一步加剧了预测的不确定性。总体来看, 未来研究应从提升数据质量(增加深层采样点)、统一取样标准、整合多尺度数据、降低估算方法不确定性, 以及改进模型外推能力等方面进行优化。

3 结论与展望

基于大尺度样带调查, 结合文献数据整合, 本研究构建了全新的东北多年冻土区土壤有机碳数据库, 该数据库有望为全面评估全球多年冻土区土壤碳储量提供重要的数据补充。基于该数据库, 同时结合机器学习模型, 本研究预测了东北多年冻土区土壤有机碳密度的空间分布特征。研究结果显示, 0~10, 0~30和

0~50 cm深度的土壤碳密度分别为 (4.2 ± 1.1) , (9.8 ± 2.5) 和 (12.7 ± 2.9) kg C/m², 对应的土壤碳储量分别为 (2.4 ± 0.6) , (5.5 ± 1.4) 和 (7.1 ± 1.6) Pg C。研究区土壤有机碳密度呈现由西南向东北逐渐增加的分布趋势, 其中森林土壤的碳储量最高, 草地和湿地次之。此外, 基于统一深度(0~30 cm)的比较结果显示, 环北极多年冻土区和东北多年冻土区的土壤碳密度均高于青藏高原多年冻土区。尽管东北多年冻土区的面积远小于青藏高原多年冻土区, 但两者的碳储量相当。上述结果表明, 东北多年冻土区是个不可忽视的碳库, 可能在区域尺度乃至国家尺度碳循环中发挥着重要作用。为进一步提高东北多年冻土区土壤碳储量估算的精度, 未来研究应加强对东北多年冻土区更多样点、更深土壤层次的碳储量调查, 并结合高分辨率卫星遥感资料和深度神经网络等方法实现尺度转换。此外, 为了深入理解该区域潜在的冻土碳-气候反馈关系, 未来研究应重点关注东北多年冻土区碳循环关键过程及其对气候变暖的响应机制。

致谢 感谢北京师范大学沈妙根教授课题组在野外采样中给予的帮助。

参考文献

- 1 Zhang T, Barry R G, Knowles K, et al. Statistics and characteristics of permafrost and ground-ice distribution in the Northern Hemisphere. *Polar Geogr*, 1999, 23: 132–154
- 2 Tarnocai C, Canadell J G, Schuur E A G, et al. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob Biogeochem Cycle*, 2009, 23: 2008GB003327
- 3 Knoblauch C, Beer C, Liebner S, et al. Methane production as key to the greenhouse gas budget of thawing permafrost. *Nat Clim Change*, 2018, 8: 309–312
- 4 Schuur E A G, Abbott B W, Commane R, et al. Permafrost and climate change: carbon cycle feedbacks from the warming Arctic. *Annu Rev Environ Resour*, 2022, 47: 343–371
- 5 Hugelius G, Strauss J, Zubrzycki S, et al. Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences*, 2014, 11: 6573–6593
- 6 Mishra U, Hugelius G, Shelef E, et al. Spatial heterogeneity and environmental predictors of permafrost region soil organic carbon stocks. *Sci Adv*, 2021, 7: eaaz5236
- 7 Mountain research initiative EDW working group. Elevation-dependent warming in mountain regions of the world. *Nat Clim Change*, 2015, 5: 424–430
- 8 Biskaborn B K, Smith S L, Noetzel J, et al. Permafrost is warming at a global scale. *Nat Commun*, 2019, 10: 264
- 9 Intergovernmental Panel on Climate Change. Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report [Masson-Delmotte V, Zhai P, Pirani A, et al. (eds.)]. Cambridge: Cambridge University Press, 2021. In press
- 10 Serreze M C, Walsh J E, Chapin III F S, et al. Observational evidence of recent change in the northern high-latitude environment. *Clim Change*, 2000, 46: 159–207
- 11 Schuur E A G, McGuire A D, Schädel C, et al. Climate change and the permafrost carbon feedback. *Nature*, 2015, 520: 171–179

- 12 Koven C D, Ringeval B, Friedlingstein P, et al. Permafrost carbon-climate feedbacks accelerate global warming. *Proc Natl Acad Sci USA*, 2011, 108: 14769–14774
- 13 Jobbágy E G, Jackson R B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol Appl*, 2000, 10: 423
- 14 Mu C, Zhang T, Wu Q, et al. Editorial: organic carbon pools in permafrost regions on the Qinghai-Xizang (Tibetan) Plateau. *Cryosphere*, 2015, 9: 479–486
- 15 Ding J, Wang T, Piao S, et al. The paleoclimatic footprint in the soil carbon stock of the Tibetan permafrost region. *Nat Commun*, 2019, 10: 4195
- 16 Li F Z. Dynamics of top soil organic carbon density in the greater Khingan mountains (in Chinese). Dissertation for Doctoral Degree. Inner Mongolia: Inner Mongolia Agricultural Univ, 2022 [李凤滋. 大兴安岭林区表层土壤有机碳密度动态研究. 博士学位论文. 内蒙古: 内蒙古农业大学, 2022]
- 17 Liu R, Zang S Y, Zhao L, et al. Climate and environmental changes in the Mohe Basin permafrost region of the Greater Hinggan Mountains since 30 ka BP (in Chinese). *Acta Geogr Sin*, 2024, 79: 2280–2296 [刘锐, 臧淑英, 赵林, 等. 30 ka BP以来大兴安岭漠河盆地多年冻土区气候与环境变化. *地理学报*, 2024, 79: 2280–2296]
- 18 Peng W H, Mu C C, Chang Y H, et al. Carbon storage of forested wetland ecosystems in the cold temperate permafrost region, northeast China (in Chinese). *Acta Pedol Sin*, 2020, 57: 1526–1538 [彭文宏, 牟长城, 常怡慧, 等. 东北寒温带永久冻土区森林沼泽湿地生态系统碳储量. *土壤学报*, 2020, 57: 1526–1538]
- 19 Jin H J, Yu S P, Lu L Z, et al. Degradation of permafrost in the Da and Xiao Hinggan Mountains, Northeast China, and preliminary assessment of its trends (in Chinese). *J Glaciol Geocryol*, 2006, 28: 467–474 [金会军, 于少鹏, 吕兰芝, 等. 大小兴安岭多年冻土退化及其趋势初步评估. *冰川冻土*, 2006, 28: 467–474]
- 20 Wei Z, Jin H J, Zhang J M, et al. Prediction of permafrost changes in Northeastern China under a changing climate. *Sci China Earth Sci*, 2011, 54: 924–935
- 21 Li H, Wu Y, Liu S, et al. Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Glob Change Biol*, 2022, 28: 3394–3410
- 22 Wang H, Jin H, Li X, et al. Changes in carbon stock in the Xing'an permafrost regions in Northeast China from the late 1980s to 2020. *GISci Remote Sens*, 2023, 60: 2217578
- 23 Huang S, Ding Q, Chen K, et al. Changes in near-surface permafrost temperature and active layer thickness in Northeast China in 1961–2020 based on GIPL model. *Cold Regions Sci Tech*, 2023, 206: 103709
- 24 Mantel S, Dondeyne S, Deckers S. World reference base for soil resources (WRB). *Encyclopedia of Soils in the Environment*, 2nd ed. Netherlands: Elsevier, 2023. 206–217
- 25 Brown J, Ferrians O, Heginbottom J, et al. Circum-arctic map of permafrost and ground-ice conditions, version 2. NASA National Snow and Ice Data Center Distributed Active Archive Center, 2002
- 26 Zhang X S. Vegetation map of the People's Republic of China (1:1000000) (in Chinese). Beijing: Geology Press, 2007 [张新时. 中华人民共和国植被图(1:1000000). 北京: 地质出版社, 2007]
- 27 Yang Y, Cheng S, Fang H, et al. Interactions between soil organic matter chemical structure and microbial communities determine the spatial variation of soil basal respiration in boreal forests. *Appl Soil Ecol*, 2023, 183: 104743
- 28 Xu L, Yu G, He N, et al. Carbon storage in China's terrestrial ecosystems: a synthesis. *Sci Rep*, 2018, 8: 2806
- 29 Shi J, Song G. China soil native species database: a nationwide soil dataset based on the second national soil survey. Science Data Bank, 2016 [施建平, 宋歌. 中国土种数据库—基于第二次土壤普查的全国性土壤数据集. 中国科学数据, 2016]
- 30 Perie C, Ouimet R. Organic carbon, organic matter and bulk density relationships in boreal forest soils. *Can J Soil Sci*, 2008, 88: 315–325
- 31 Hossain M, Chen W, Zhang Y. Bulk density of mineral and organic soils in the Canada's arctic and sub-arctic. *Inf Process Agric*, 2015, 2: 183–190
- 32 Xu L, Yu G, He N. Increased soil organic carbon storage in Chinese terrestrial ecosystems from the 1980s to the 2010s. *J Geogr Sci*, 2019, 29: 49–66
- 33 Breiman L. Random forests. *Mach Learn*, 2001, 45: 5–32
- 34 Kavzoglu T, Colkesen I. A kernel functions analysis for support vector machines for land cover classification. *Int J Appl Earth Observation Geoinf*, 2009, 11: 352–359
- 35 Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors).

- [Ann Statist](#), 2000, 28: 337–407
- 36 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, Assoc Comput Mach, 2016, 785–794
- 37 Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification. Technical report, Department of Computer Science, Taiwan University, 2003. Available from: URL: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- 38 Ding J, Li F, Yang G, et al. The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. [Global Change Biol](#), 2016, 22: 2688–2701
- 39 Nguyen T T, Pham T D, Nguyen C T, et al. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. [Sci Total Environ](#), 2022, 804: 150187
- 40 Emadi M, Taghizadeh-Mehrjardi R, Cherati A, et al. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. [Remote Sens](#), 2020, 12: 2234
- 41 Szatmári G, Pásztor L, Heuvelink G B M. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. [Geoderma](#), 2021, 403: 115356
- 42 Odebiri O, Mutanga O, Odindi J, et al. Modelling soil organic carbon stock distribution across different land-uses in South Africa: a remote sensing and deep learning approach. [ISPRS J Photogramm Remote Sens](#), 2022, 188: 351–362
- 43 Geng J, Tan Q, Lv J, et al. Assessing spatial variations in soil organic carbon and C:N ratio in Northeast China's black soil region: insights from Landsat-9 satellite and crop growth information. [Soil Tillage Res](#), 2024, 235: 105897
- 44 Viscarra Rossel R A, Webster R, Bui E N, et al. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. [Glob Change Biol](#), 2014, 20: 2953–2970
- 45 Zhou Z, Ren C, Wang C, et al. Global turnover of soil mineral-associated and particulate organic carbon. [Nat Commun](#), 2024, 15: 5329
- 46 Wu T, Wang D, Mu C, et al. Storage, patterns, and environmental controls of soil organic carbon stocks in the permafrost regions of the Northern Hemisphere. [Sci Total Environ](#), 2022, 828: 154464
- 47 Devine S M, O' Geen A T, Liu H, et al. Terrain attributes and forage productivity predict catchment-scale soil organic carbon stocks. [Geoderma](#), 2020, 368: 114286
- 48 Sáez-Sandino T, García-Palacios P, Maestre F T, et al. The soil microbiome governs the response of microbial respiration to warming across the globe. [Nat Clim Chang](#), 2023, 13: 1382–1387
- 49 R Core Team R. A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2022
- 50 Tang X, Zhao X, Bai Y, et al. Carbon pools in China's terrestrial ecosystems: new estimates based on an intensive field survey. [Proc Natl Acad Sci USA](#), 2018, 115: 4021–4026
- 51 Song C C, Song Y Y, Wang X W, et al. Advanced in researches on carbon and nitrogen cycles in wetland ecosystems under climate change (in Chinese). [Wetland Sci](#), 2018, 16: 424–431 [宋长春, 宋艳宇, 王宪伟, 等. 气候变化下湿地生态系统碳、氮循环研究进展. [湿地科学](#), 2018, 16: 424–431]
- 52 Hu J, Hartemink A E, Desai A R, et al. A continental-scale estimate of soil organic carbon change at NEON sites and their environmental and edaphic controls. [J Geophys Res Biogeosci](#), 2023, 128: e2022JG006981
- 53 Rocci K S, Cleveland C C, Eastman B A, et al. Aligning theoretical and empirical representations of soil carbon-to-nitrogen stoichiometry with process-based terrestrial biogeochemistry models. [Soil Biol Biochem](#), 2024, 189: 109272
- 54 Franklin W, Cardy G. The united nations data bases on desertification. In: Arnalds O, Archer S, eds. Rangeland desertification. Advances in Vegetation Science. Dordrecht: Springer, 2000. 131–141
- 55 Heckman K A, Possinger A R, Badgley B D, et al. Moisture-driven divergence in mineral-associated soil carbon persistence. [Proc Natl Acad Sci USA](#), 2023, 120: e2210044120
- 56 Post W M, Emanuel W R, Zinke P J, et al. Soil carbon pools and world life zones. [Nature](#), 1982, 298: 156–159
- 57 Yang Y, Mohammad A, Feng J, et al. Storage, patterns and environmental controls of soil organic carbon in China. [Biogeochemistry](#), 2007, 84: 131–141
- 58 Diáková K, Čapek P, Kohoutová I, et al. Heterogeneity of carbon loss and its temperature sensitivity in East-European subarctic tundra soils. [FEMS Microbiol Ecol](#), 2016, 92: fiw140
- 59 Wang J, Liu D. Models overestimate ecosystem water use efficiency for northern permafrost regions. [Agric For Meteorol](#), 2023, 339: 109594

- 60 Fang K, Chen L, Qin S, et al. Mineral and climatic controls over soil organic matter stability across the Tibetan alpine permafrost region. *Global Biogeochem Cycle*, 2021, 35: e2021GB007118
- 61 Chen L, Yang G, Bai Y, et al. Permafrost carbon cycle and its dynamics on the Tibetan Plateau. *Sci China Life Sci*, 2024, 67: 1833–1848
- 62 Epstein H E, Reynolds M K, Walker D A, et al. Dynamics of aboveground phytomass of the circumpolar Arctic tundra during the past three decades. *Environ Res Lett*, 2012, 7: 015506
- 63 Running S, Zhao M. MOD17A3HGF MODIS/terra net primary production gap-filled yearly L4 global 500 m SIN grid V006 (NASA EOSDIS Land Processes DAAC), 2019
- 64 Yang G, Peng C, Chen H, et al. Qinghai-Tibetan Plateau peatland sustainable utilization under anthropogenic disturbances and climate change. *Ecosyst Health Sustain*, 2017, 3: e01263
- 65 Craft C, Mangroves I. C. Creating and Restoring Wetlands. Boston: Elsevier. 2016
- 66 Hugelius G, Loisel J, Chadburn S, et al. Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proc Natl Acad Sci USA*, 2020, 117: 20438–20446
- 67 Zhao L, Hu G J, Zou D F, et al. Permafrost changes and its effects on hydrological processes on Qinghai-Tibet Plateau (in Chinese). *Bull Chin Acad Sci*, 2019, 34: 1233–1246 [赵林, 胡国杰, 邹德富, 等. 青藏高原多年冻土变化对水文过程的影响. 中国科学院院刊, 2019, 34: 1233–1246]
- 68 Fan Y, van den Dool H. A global monthly land surface air temperature analysis for 1948–present. *J Geophys Res Atmos*, 2008, 113: 2007JD008470
- 69 Peng S, Ding Y, Liu W, et al. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth Syst Sci Data*, 2019, 11: 1931–1946
- 70 Liu F, Wu H, Zhao Y, et al. Mapping high resolution national soil information grids of China. *Sci Bull*, 2022, 67: 328–340
- 71 Zhu D, Peng S, Ciais P, et al. Simulating soil organic carbon in yedoma deposits during the Last Glacial Maximum in a land surface model. *Geophys Res Lett*, 2016, 43: 5133–5142
- 72 Shi Y, Yang C, Zhu J L, et al. Estimation of national and provincial carbon emissions, terrestrial carbon sinks and their relative contribution to emission reductions during 1980–2020 (in Chinese). *Sci Sin-Vitae*, 2024, 54: 2459–2478 [石岳, 杨晨, 朱江玲, 等. 2024. 中国及省域碳排放、陆地碳汇及其相对减排贡献, 1980–2020. 中国科学: 生命科学, 2024, 54: 2459–2478]
- 73 Yang Y H, Zhang D Y, Bai Y X, et al. Research progress about processes, mechanisms and trajectories of carbon, nitrogen and phosphorus cycling in China's permafrost ecosystems (in Chinese). *China Basic Sci*, 2024, 26: 53–59 [杨元合, 张典业, 白宇轩, 等. 2024. 我国冻土生态系统碳氮磷循环过程、机制及演化趋势研究进展. 中国基础科学, 2024, 26: 53–59]
- 74 He M, Qin S Q, Feng X H, et al. Spatial distributions and key drivers of microbial necromass carbon in permafrost deposits across the Tibetan Plateau (in Chinese). *Chin Sci Bull*, 2023, 69: 925–938 [贺美, 秦书琪, 冯雪徽, 等. 青藏高原多年冻土微生物残体碳的分布特征及其驱动因素. 科学通报, 2023, 69: 925–938]
- 75 Adam E, Mutanga O, Rugege D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecol Manage*, 2010, 18: 281–296
- 76 Matzner E, Borken W. Do freeze-thaw events enhance C and N losses from soils of different ecosystems? A review *Eur J Soil Sci*, 2008, 59: 274–284
- 77 Minasny B, Berglund Ö, Connolly J, et al. Digital mapping of peatlands—a critical review. *Earth-Sci Rev*, 2019, 196: 102870
- 78 Reddy N N, Das B S. Digital soil mapping of key secondary soil properties using pedotransfer functions and Indian legacy soil data. *Geoderma*, 2023, 429: 116265
- 79 Meng X, Bao Y, Luo C, et al. SOC content of global Mollisols at a 30 m spatial resolution from 1984 to 2021 generated by the novel ML-CNN prediction model. *Remote Sens Environ*, 2024, 300: 113911
- 80 Bao Y, Meng X, Liu H, et al. Assessing the improvement potentials of climate model partitioning and time-variant feature extraction for soil organic carbon prediction. *Catena*, 2024, 241: 108014
- 81 Siewert M B. High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. *Biogeosciences*, 2018, 15: 1663–1682
- 82 Ren Y, Li X, Mao D, et al. Northeast China holds huge wetland soil organic carbon storage: an estimation from 819 soil profiles and random forest algorithm. *Plant Soil*, 2023, 490: 469–483

Soil carbon storage and its spatial distribution in the permafrost region of Northeast China

YANG Yan^{1,2}, JIA BingRui^{1,2}, HE Nan^{1,2,3}, ZHANG Fan^{1,2,3}, ZHANG DianYe^{1,2} & YANG YuanHe^{1,2,3*}

¹ Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

² China National Botanical Garden, Beijing 100093, China

³ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China

* Corresponding author, E-mail: yhyang@ibcas.ac.cn

The thawing of permafrost due to climate warming is expected to accelerate soil carbon release, potentially triggering a positive feedback between permafrost carbon cycle and climate warming. To address the unclear characteristics of soil carbon pools in the Northeast permafrost region of China, this study integrated soil organic carbon data from an 800 km transect survey, combined with relevant published literature. Using these site-level observations, we adopted several machine learning models, including random forest (RF), support vector machine (SVM), boosted regression trees (BRT), and extreme gradient boosting (XGBoost), to generate spatial patterns of soil organic carbon pool. Among these, XGBoost demonstrated the highest accuracy and was employed to assess soil organic carbon storage and its spatial distribution across the region. The results revealed that climate (aridity index) and soil properties (total nitrogen, TN) together explained 61.7% of the variability in soil organic carbon density (SOCD). The average SOCDs at depths of 0~30 and 0~50 cm were (9.8 ± 2.5) and (12.7 ± 2.9) kg C/m², respectively, with the corresponding soil carbon storage being (5.5 ± 1.4) and (7.1 ± 1.6) Pg C, respectively. The SOCD generally increased from southwest to northeast, with notable differences across vegetation types (wetland: (11.6 ± 2.4) kg C/m² > forest: (10.4 ± 2.0) kg C/m² > grassland: (8.1 ± 2.6) kg C/m²). These findings underscore the Northeast China permafrost region as a significant carbon reservoir, which could become a substantial carbon source in the context of ongoing climate warming. Therefore, it is crucial to explore the key processes that govern the carbon cycle and their responses to climate warming in Northeast China's permafrost region.

soil organic carbon density (SOCD), soil organic carbon storage, carbon cycle, carbon-climate feedback, permafrost, machine learning

doi: [10.1360/SSV-2025-0026](https://doi.org/10.1360/SSV-2025-0026)