DOI: 10.20103/j.stxb.202501130114

郭振敏,汲玉河,周广胜,周梦子,郑凯.基于机器学习的兴安落叶松生态系统在不同时间尺度的碳通量模拟及影响因素分析.生态学报,2025,45 (16); - .

Guo Z M, Ji Y H, Zhou G S, Zhou M Z, Zheng K.Carbon flux simulation and analysis of influencing factors in the *Larix gmelinii* ecosystem at different time scales based on machine learning. Acta Ecologica Sinica, 2025, 45(16): - .

基于机器学习的兴安落叶松生态系统在不同时间尺度 的碳通量模拟及影响因素分析

郭振敏^{1,3},汲玉河^{2,3,*},周广胜^{2,3},周梦子^{2,3},郑 凯^{1,3}

1 郑州大学地球科学与技术学院,郑州 450001

2 中国气象科学研究院灾害天气国家重点实验室,北京 100081

3 中国气象科学研究院郑州大学生态气象联合实验室,郑州 450001

摘要:机器学习已经广泛用于生态系统研究。基于 2014 年 1 月 1 日至 2018 年 12 月 31 日兴安落叶松生态系统碳通量(NEE)观测数据,分析了其动态变化特征,并采用多种机器学习方法进行模拟。结果表明:(1)生长季兴安落叶松生态系统 NEE 日动态 呈"U"变化,整体表现为碳汇,7 月份碳汇能力最强,达 67.57 g C m⁻²月⁻¹,9 月至次年 5 月表现为碳源。(2)结构方程模型分析 表明,兴安落叶松生态系统 NEE 的主要影响因子为潜热通量(*LE*)、净辐射(*Rn*)、叶面积指数(LAI)、空气温度(*Ta*)、相对湿度 (*RH*)、饱和水汽压差(VPD)和土壤含水量(SWC),其中潜热通量和净辐射是影响 NEE 变化的最主导因素。(3)四种机器学习 方法(RF、XGBoost、SVM、ANN)均能较准确地模拟兴安落叶松生态系统 NEE,其中 XGBoost 和 RF 的模拟结果最为相近,但 XGBoost 在模拟精度和计算效率方面优于 RF。研究结果为应用机器学习方法估算生态系统碳通量提供了依据。 关键词:兴安落叶松;净生态系统二氧化碳交换(NEE);碳通量;机器学习;模拟

Carbon flux simulation and analysis of influencing factors in the *Larix gmelinii* ecosystem at different time scales based on machine learning

GUO Zhenmin^{1,3}, JI Yuhe^{2,3,*}, ZHOU Guangsheng^{2,3}, ZHOU Mengzi^{2,3}, ZHENG Kai^{1,3}

1 School of Earth Science and Technology, Zhengzhou University, Zhengzhou 450001, China

2 State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Science, Beijing 100081, China

3 Joint Laboratory of Eco-meteorology, Chinese Academy of Meteorological Sciences, Zhengzhou University, Zhengzhou 450001, China

Abstract: Machine learning has been widely used in ecosystem research. This study analyzed the dynamic variation characteristics of the carbon flux (NEE) data from the *Larix gmelinii* ecosystem, observed from January 1, 2014, to December 31, 2018, and simulated the data using various machine learning methods. The results indicated that: (1) The daily dynamics of the NEE during the growing season of the *Larix gmelinii* ecosystem exhibited a "U" shape, with the ecosystem acting as a carbon sink overall. The carbon sink capacity was strongest in July, with a monthly average of 67.57 g C m⁻². From September to May of the following year, the ecosystem acted as a carbon source. (2) Structural equation modeling revealed that the main influencing factors of NEE in the *Larix gmelinii* ecosystem are latent heat flux (*LE*), net radiation (*Rn*), leaf area index (LAI), air temperature (*Ta*), relative humidity (*RH*), vapor pressure deficit (VPD), and soil water content (SWC), with latent heat flux and net radiation being the most dominant factors affecting NEE

收稿日期:2025-01-13; 网络出版日期:2025-00-00

* 通讯作者 Corresponding author.E-mail: jiyh@ cma.gov.cn

http://www.ecologica.cn

基金项目:中国气象局农业气象重点创新团队(CMA2024ZD02);中国气象科学研究院基本科研业务费(2024Z001);三北工程区陆地生态系统 增汇潜力及风险评估(42141007);生态保护修复增汇潜力与气候变化风险评估(2023Z023)

variations. (3) Four machine learning methods (RF, XGBoost, SVM and ANN) accurately simulated the NEE of the *Larix gmelinii* ecosystem, with XGBoost and RF providing similar results. However, XGBoost outperformed RF in terms of simulation accuracy and computational efficiency. The study provides a basis for using machine learning methods to estimate ecosystem carbon fluxes.

Key Words: Larix gmelinii; net ecosystem carbon exchange; carbon flux; machine learning; simulation

在全球气候变化背景下,陆地生态系统碳循环对环境因子变化的响应已经成为生态学研究的热点问题^[1]。作为陆地生态系统的重要组成部分,森林覆盖了约 26%的陆地面积,贡献了约 60%的陆地植被净初级 生产力,其储存的碳量约为大气的 3 倍,与大气系统之间碳交换量占整个陆地生态系统的 90%以上^[2-3],因此 森林生态系统 CO₂通量的动态变化成为全球变化研究的焦点之一^[4]。明确森林碳循环的调控机制和精确地 估算森林生态系统与大气间的碳通量,不仅是理解陆地碳循环的基础,也是研究全球碳循环的重要组成部分。

目前,观测和模拟是研究陆地生态系统净碳交换特征及其控制机制的主要方法^[5]。在观测方面,碳通量 监测常用方法包括清单法、箱法和微气象法^[6]。其中,微气象法中的涡度相关技术能够在半小时尺度上连续 观测植被群落与大气间的气体交换及其环境主导因子,已成为测定陆地生态系统 CO₂和水热通量的重要工 具^[7]。研究表明,森林与大气之间的碳交换受多种因子的共同影响,例如温度、降水、光合有效辐射及林龄 等^[8–10],且这些因子之间的交互作用对 CO₂通量的影响也不可忽视。例如,东北地区落叶松碳汇主要受光合 有效辐射影响^[11];西双版纳橡胶林的净生态系统碳交换(Net Ecosystem Carbon Exchange, NEE)在干季受物 候因素驱动,而在雨季主要受空气温度和饱和水汽压差共同调控^[12];祁连山青海云杉林的 NEE 主要受植被 指数和净辐射的综合影响,随着植被指数的增加,其碳汇功能显著增强^[13]。上述研究表明,不同森林生态系 统的碳循环机制过程虽存在一定的相似性,但碳储量以及光合速率、呼吸速率因环境异质性的差异而表现出 不同的碳源/汇特征^[14–15]。

在模拟方面,广泛应用的 NEE 模型主要分为两大类:基于生态学过程的静态模型和基于数据驱动的动态 模型^[16-17]。由于 NEE 呈现非线性和非平稳的特征,使得以生态学过程为理论前提的机理模型存在较大的不 确定性^[18]。已有研究表明,基于数据驱动的动态模型在不同时空尺度上的碳通量估算表现优于传统生态学 过程模型^[19-20]。过去 20 年,机器学习在 NEE 研究中的应用日益增多,如使用人工神经网络(Artificial Neural Network, ANN)结合净辐射、温度、湿度和能量通量等变量,模拟中国典型草地、农田和森林生态系统的 NEE^[21];随机森林(Random Forest, RF)、支持向量机(Support Vector Machine, SVM)等模型用于森林涡动相 关缺失数据的插补研究^[22-23];循环神经网络(Recurrent Neural Network, RNN)和长短期记忆网络(Long Short -Term Memory, LSTM)则被用于长时间序列的二氧化碳吸收建模^[24];遗传神经网络(Genetic Algorithm Neural Network, GANN)选择 NEE 预测的主要因素^[25]等。尽管机器学习方法已逐步应用于 NEE 研究中,但现有研 究多集中于单一算法在月尺度或年尺度上的模拟,且由于缺乏对变量重要性的系统筛选,估算结果仍存在一 定不确定性^[5],尤其是在半小时尺度上的精细化模拟及多种机器学习算法的对比分析方面,相关研究仍较为 有限^[26-27]。

中国北方针叶林生态系统是典型的寒温带针叶林,其深厚的腐殖质层使得土壤碳库与植被碳汇极为丰富。然而,由于观测站点稀缺和数据不足,该生态系统与大气间通量交换的研究仍不充分,现有文献中关于其碳源/汇功能及基于机器学习的 NEE 模拟研究较为稀缺。为此,本研究基于中国北方林森林生态系统定位研究站的涡度相关长期观测数据,分析兴安落叶松 NEE 的变化特征及其环境影响因子,并采用随机森林(RF)、极端梯度提升(eXtreme Gradient Boosting, XGBoost)、支持向量机(SVM)和人工神经网络(ANN)4种不同机器学习方法对 NEE 进行模拟,验证机器学习理论及技术在生态模型模拟中的有效性。

1 材料与方法

1.1 研究区概况

中国北方林森林生态系统定位研究站(呼中站)地处我国东北大兴安岭北部——黑龙江省呼中国家级自然保护区内,通量观测场地理坐标为51°46′52″N,121°01′04″E,海拔773 m。总面积为16.7 万 hm²,其中有林地面积约占14.7 万 hm²,森林覆盖度为90%以上。该站点地处寒温带大陆性季风气候区,年均气温-4.4 ℃,最暖月7月均温17.2 ℃,最冷月12月均温-23.54 ℃。年均降雨量为481.6 mm,约70%的降水发生在夏季,具有明显的季节性变化。土壤类型为棕色针叶林土,植被类型为地带性寒温性针叶林,以兴安落叶松(Larix gmelinii)为单优势种,白桦(Betula platyphylla)为主要伴生树种,林下植被主要包括杜香(Ledum palustre)、越桔(Vaccinium vitisidaea)、兴安杜鹃(Rhododendron dauricum)等植物。

1.2 森林净生态系统交换观测试验系统

观测数据来源于中国北方林森林生态系统定位研究站的开路式涡度相关通量监测系统和微气象梯度观测系统。开路式涡度相关观测系统相关设备安装在通量塔高度约 35 m 处主风方向伸出的支臂上,数据采集频率为 10 Hz,并提供 30 min 在线通量数据。微气象梯度观测系统数据采集频率为 2 s,在线计算半小时平均值/累积值,并相应输出各种观测变量的日平均值/累积值。具体碳水通量及微气象要素观测设备信息如表 1 所示。

表1 中国北方林森林生态系统定位研究站(呼中站)碳水通量及微气象要素观测设备信息

 Table 1
 Information on observation equipment for carbon water flux and micro meteorological elements at the North China Forest Ecosystem

 Positioning Research Station (Huzhong Station)

rositioning Research Station (Tuziong Station)							
	测定要素 Measurement elements	传感器型号 Sensor model	制造商 Manufacturer	观测高度/深度 Observation height/depth			
开路式涡度相关通量监测系统	CO2、H20通量	LI-7500	LI-COR	35 m			
Open-path eddy covariance flux	三维超声风速仪	CSAT3	Campbell	35 m			
measurement system	感热通量	CSAT3	Campbell	35 m			
微气象梯度观测系统	光合有效辐射	LI-190SB	LI-COR	35 m			
Micrometeorological gradient	净辐射	CNR1	Kipp&Zonen	35 m			
observation system	三维超声风速传感器	034B	Campbell	30 m			
	空气温度/湿度	HMP45C	Campbell	2,15,25,34 m			
	土壤温度	107L	Campbell	-5,-10,-20,-50,-100 cm			
	土壤水分	CS616	Campbell	-10,-20,-30,-50,-100 cm			
数据采集器	通量采集	CR5000	Campbell				
Data acquisition system	气象要素采集	CR23XTD	Campbell				

1.3 数据处理

由于天气、电力及仪器故障等因素,野外观测数据通常存在缺失或异常。为确保数据质量,需要对通量观测数据进行处理和质量控制。数据处理流程主要包括高频原始数据的预处理、数据校正、质量控制、缺失数据插补、CO₂通量组分拆分等^[28]。首先剔除传感器测量的野点资料,随后采用 Eddypro7.0.6 软件进行采用二次坐标旋转对数据资料进行校正,确保三维风速仪坐标与自然风坐标匹配^[6],最后对湍流通量进行频率订正和WPL 校正以消除密度效应影响^[29]。但经过上述步骤处理得到半小时尺度的 CO₂通量数据仍存在不符合理论假设的异常值,需要进一步进行质量控制,具体步骤为^[30]:(1)剔除降雨同期数据;(2)利用差分法剔除通量时间序列中明显异常的数据;(3)剔除夜间摩擦风速低于临界摩擦风速 u*的通量数据,以确保观测数据的质量。

本研究通量数据选取 CO₂通量、潜热通量(Latent heat flux, LE) 和感热通量(Sensible heat flux, H), 微气

象数据选择空气温度(Air temperature, Ta)、相对湿度(Relative humidity, RH)、净辐射(Net radiation, Rn)、土 壤温度(Soil temperature, Ts)、土壤含水量(Soil water content, SWC)、饱和水汽压差(Vapor pressure deficit, VPD)。数据时间跨度为2014年1月1日至2018年12月31日,时间尺度为30min(共1826d,87646条),经 过质量控制后 CO₂通量有效数据共28456条,约占每年比例31.4%—34.1%。为了分析 NEE 昼夜、季节动态 特征,需要对质量控制后的数据进行插补。其中,通量缺失数据采用边际分布采样法、平均日变化法和线性内 插法,对数据质量控制后缺失的半小时尺度通量数据进行插补。气象缺失数据对于小于2h的采用线性内插 法进行插补,而对于超过2h的则采用平均日变化进行插补。

遥感数据使用 MOD15A2H 数据中空间分辨率 500 m、时间分辨率 8 d、时间范围为 2014—2018 年的叶面 积指数(LAI)产品,该数据产品在谷歌地球引擎(Google earth engine, GEE)云平台进行下载,以中国北方林森 林生态系统定位研究站为中心,提取半径1公里范围内多个像元 LAI 平均值。可根据 MODIS 产品提供的相 应质量标志,采用 S-G 滤波对其进行平滑,剔除因重云、传感器故障等原因损坏的不良质量数据^[31]。随后, 使用线性插值法将 LAI 数据转换为逐日数据,并对日重复值进行填补,以实现与半小时数据及其他数据格式 的匹配。

1.4 数据分析

为探究兴安落叶松 NEE 的昼夜及季节变化特征,首先使用 Excel 2016 和 Origin 2022 对数据进行整理及 可视化处理。其次,利用 Pearson 相关分析方法来分析 CO₂通量与环境变量之间的相关关系^[32],通过 Python 3.9 中的 SciPy 库实现。其次通过计算方差膨胀因子(Variance infation factor, VIF)检验环境变量之间的共线 性^[33]:当 VIF<5,表示自变量之间没有共线性问题,可作为独立变量保留;VIF > 10,表示存在较强的共线性, 需要删除变量或正则化;VIF 介于 5—10 之间,表示存在中等程度的共线性,可根据实际情况进行对自变量选 择^[34],该部分在 Spss 26.0 软件中完成。最后基于结构方程模型量化环境因子对 NEE 的直接、间接和总影响, 采用 X^2/df 检验(1 < X^2/df < 3)、拟合优度指数(GFI > 0.9、AGFI > 0.9)和近似均方根误差(RMSEA<0.05)来 评估模型的适用性和充分性^[35],该部分通过 Amos 26.0 软件实现。以上数据分析流程主要完成兴安落叶松 NEE 与主导环境因子之间的调控机制研究。

本研究对比了4种不同机器学习模型:随机森林(RF)、极端梯度提升(XGBoost)、支持向量机(SVM)和 人工神经网络(ANN),以评估对兴安落叶松 NEE 的模拟能力,为基于大数据的森林生态系统固碳研究提供方 法参考。

(1)随机森林(RF)

随机森林是由多棵分类与回归树构成的非参数机器学习算法^[36]。该算法可有效地运行大量数据集,并 处理数千个输入变量,在分类问题上具有构建速度快,分类精度高,可以在一定程度上避免过拟合等优点^[37]。 随机森林在抽样过程中采用了有放回抽样,随机选择一部分样本分割分类树的每个节点,剩下未被抽取的样 本被称为袋外数据(Out of bag)。利用大约 37%的袋外数据进行分类结果的精度评价以及计算不同特征变量 的重要性^[38]。具体过程为:(1)将所有特征变量均参与到随机森林的袋外精度中,并计算每个特征变量的重 要性;(2)根据特征重要性排序,生成不同的特征组合,并为每个组合构建随机森林模型,计算其袋外样本精 度;(3)对每个组合进行 N 次循环计算,取袋外精度的平均值作为最终精度,选择精度最高的特征组合作为最 优变量集。

(2)极端梯度提升(XGBoost)

极端梯度提升是一种基于 Boosting 集成学习的机器学习方法。该算法以 CART 决策树作为基分类器,逐步加入新的 CART 树以拟合前一轮预测的残差,最终将所有树的预测结果相加以生成最终预测值^[39]。 XGBoost 因其高效性和良好的预测精度受到广泛关注,在表格型数据处理方面常优于传统机器学习方法及部分深度神经网络模型^[40]。

(3)支持向量机(SVM)

支持向量机是一种基于监督学习的广义线性分类器,已被广泛应用于水质评价、水华预测等领域^[41]。 SVM 通过预设的核函数(如线性核、多项式核、高斯径向基核)将有限维度空间映射到高维特征空间,从而可 以将非线性回归变为线性回归^[42]。其优点是在样本数据有限的情况下,可以在复杂的模型与学习能力之间 寻求最佳,以求获得最优的泛化能力。

(4)人工神经网络(ANN)

人工神经网络通过自主学习构建输入变量与目标值之间的非线性映射关系,通常包括输入层、隐藏层和 输出层^[43]。输入层负责接收输入数据,输出层负责输出整个神经网络的计算结果,隐藏层则通过连接权重和 偏置反映变量间的复杂层次关系^[44]。神经网络上每个节点称为神经元,各层之间的神经元通过一定的权重 相互连接,其基本思想是通过调整连接权重和偏置来最小化损失函数,从而提高模型的预测能力^[45]。神经网 络拥有并行化分布式计算的特点,能够在短时间内寻找到优化解,充分发挥了计算机高速运算的能力,在解决 高维度,非线性的问题上有极大的优势^[46]。

为进一步优化模型参数,本研究使用 Python 3.9 中 Scikit-learn 的 Randomized search CV 进行超参数随机 搜索。该方法在计算成本和搜索效率方面表现出显著优势,能够在相对较短的时间内找到性能较优的超参数 配置^[47]。各模型重要参数设置如下:在 RF 模型中,max_depth 设置为 33,max_features 选择 sqrt,min_samples_ split 为 23,n_estimators 为 408;在 XGBoost 模型中,学习率设置为 0.029,max_depth 为 8,min_child_weight 为 4, n_estimators 为 303;在 SVM 模型中,惩罚因子为 3.437,损失函数为 0.024,核函数选择径向基函数(RBF);在 ANN 模型中,优化器选择 Adam 方法进行优化,隐藏层的大小设置为 1,隐藏层节点数为 100,dropout 率为 0.2, 激活函数为 ReLU,训练轮次设置为 280。

1.5 精度验证与评价指标

本研究将数据集按照 8:2 的比例随机划分为训练集和测试集。在训练集上进一步采用十折交叉验证,以 优化模型参数并评估模型的泛化性能。研究采用决定系数(*R*²)、均方根误差(RMSE)、平均绝对误差(MAE) 三个评价指标对不同机器学习模型模拟的结果进行精度检验。

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}$$
(1)

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

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$
(3)

式中, y_i 是第*i*个观测值, \hat{y}_i 是第*i*个预测值, \bar{y} 是观测值的均值,*n*是样本数量。

2 结果与分析

2.1 兴安落叶松 NEE 时间变化特征

2.1.1 NEE 昼夜变化特征

如图 1 所示,兴安落叶松生态系统 NEE 在生长季(5 月至 9 月)表现出明显的昼夜变化特征,尤其在 6—8 月变化幅度较大,范围为-0.59—0.27 mg CO₂ m⁻² s⁻¹。白天,随着光照强度和气温的升高,植物光合作用逐渐 增强,NEE 转为负值,生态系统作为碳汇吸收大气中的 CO₂,日平均碳汇强度为-0.22 mg CO₂ m⁻² s⁻¹。7 月碳 吸收能力最强,NEE 月净碳吸收量高达-67.57 g C m⁻²。夜间,由于光合作用停止,呼吸作用占主导地位,NEE 转为正值,表现为碳源,夜间碳源强度为 0.13 mg CO₂ m⁻² s⁻¹。凌晨 0:00 至 4:00 期间,NEE 保持相对稳定,随 着日出,光合速率逐步增加,在 5:30 至 6:30 NEE 从正值转为负值,生态系统从碳源转变为碳汇;8:00 至 9:00 碳吸收达到峰值,随后逐渐下降;17:00 至 18:00 NEE 再次转为正值,生态系统向大气释放 CO₂,兴安落叶松 从碳汇转变为碳源。夜间 20:00 至次日凌晨 4:00, NEE 趋于稳定状态。

在非生长季(10月至次年4月),NEE 昼夜变化相对平缓,整体表现为碳排放(图2)。由于低温条件下光 合作用受到抑制,兴安落叶松逐渐进入休眠状态,生态系统主要通过呼吸作用释放 CO₂。10月为生长季末期 的过渡阶段,温度和辐射逐渐降低,但总初级生产力(GPP)尚未完全归零,在 10:00 至 11:00 期间,NEE 短暂 呈负值,系统表现出碳汇特征。





2.1.2 NEE 季节变化特征

兴安落叶松 NEE 表现出明显的季节性动态变化, 整体呈现典型的 U 型单峰趋势(图 3)。每年 5 月至 6 月是兴安落叶松的萌芽、展叶和开花阶段,随着叶片的 生长发育,生态系统的碳吸收能力逐步增强,NEE 逐渐 转为负值。6 月底至 8 月是生长季的旺盛期,树木叶片 完全舒展,新陈代谢活动高度活跃,NEE 达到碳吸收峰 值,生态系统碳汇功能最强。从 8 月下旬至 9 月中旬, 树木逐渐形成冬芽并停止生长,碳吸收逐渐过渡为碳排 放,且碳排放峰值集中在这一阶段。自 9 月下旬起,植 被逐渐进入休眠,系统表现为碳源状态。在 10 月至次 年 4 月期间,植物完全进入休眠,生态系统代谢活动趋 于微弱,NEE 日总值波动范围在 0.12—1.79 g C m⁻² d⁻¹ 之间。

2.2 兴安落叶松 NEE 的环境影响因素

NEE 是生态系统碳输入和排放之后的碳净交换量,反映了植物光合与呼吸作用的综合结果,受温度、光照、水分等多种环境因子的共同影响。为揭示兴安落叶松生态系统 NEE 的环境影响机制,本研究选取空气温





Fig.2 Monthly average diurnal variation of *Larix gmelini* NEE during the non growing season







度(Ta)、相对湿度(RH)、净辐射(Rn)、饱和水汽压差(VPD)、土壤温度(Ts)、土壤含水量(SWC)、叶面积指数(LAI)、潜热通量(LE)和感热通量(H)共9个关键环境因子,采用多种方法分析其对 NEE 的影响。

2.2.1 Pearson 相关性分析

基于半小时尺度的数据对 NEE 与各环境因子进行了 Pearson 相关性分析(图4)。结果表明:NEE 与所有 环境因子均存在显著相关性(P<0.001)。其中,NEE 与 RH 呈显著正相关关系(P<0.001);与 Ta、Rn、Ts、 SWC、VPD、LAI、LE、H 为极其显著的负相关关系(P<0.001),在负相关因子中,LE 与 NEE 的负相关系数最大 (-0.66),其次为 Rn(-0.52),表明潜热通量和净辐射是影响兴安落叶松生态系统碳吸收的关键因子。



图 4 兴安落叶松生态系统 NEE 与环境因子的相关性分析

Fig.4 Correlation analysis between NEE and environmental factors in the Larix gmelinii ecosystem

NEE:净生态系统碳交换 Net ecosystem carbon exchange; *Ta*:空气温度 Air temperature; *RH*:相对湿度 Relative humidity; *Rn*:净辐射 Net radiation; *Ts*:土壤温度 Soil temperature; SWC:土壤含水量 Soil water content; VPD:饱和水气压差 Vapor pressure deficit; LAI:叶面积指数 Leaf area index; *LE*: 潜热通量 Latent heat flux; *H*:感热通量 Sensible heat flux; *表示 *P*<0.05; **表示 *P*<0.01; ***表示 *P*<0.001

2.2.2 共线性检验

基于相关性分析结果发现,部分环境变量之间相关性较强。在模型中同时引入多个高度相关的变量可能导致信息冗余,进而影响模型的解释力和稳定性。为此,本研究采用方差膨胀因子(VIF)对环境变量进行共线性检验。结果如表2所示,*RH*、SWC、*LE*和LAI的VIF值均低于5,变量之间共线性较低,可作为后续建模中独立特征变量。*Rn*、*H*和VPD的VIF值介于5.115—7.500之间,存在一定程度的共线性,但通常不会对回归模型的稳定性和解释力产生显著影响。鉴于生态系统通量交换的复杂性,单个因子难以全面解释 NEE 的

Table 2 Variance inflation factor between environmental factors and NEE					
环境因子	方差膨胀因子	环境因子	方差膨胀因子		
Environmental factor	Variance infation factor	Environmental factor	Variance infation factor		
空气温度 Ta	12.834	VPD	5.115		
相对湿度 RH	4.528	LAI	3.110		
净辐射 Rn	7.500	LE	2.578		
土壤温度 Ts	12.556	Н	5.175		
土壤含水量 SWC	1.674				

Ta:空气温度 Air temperature; RH:相对湿度 Relative humidity; Rn:净辐射 Net radiation; Ts:土壤温度 Soil temperature; SWC:土壤含水量 Soil water content; VPD:饱和水气压差 Vapor pressure deficit; LAI: 叶面积指数 Leaf area index; LE: 潜热通量 Latent heat flux; H: 感热通量 Sensible heat flux

表 2 环境因子与 NEE 之间的方差膨胀因子

变化,多变量的交互作用对 CO₂通量交换具有重要影响, Rn、H 和 VPD 也被纳入后续建模中。空气温度和土 壤温度的 VIF 值均大于 10,存在显著的多重共线性,结合两者与 NEE 的相关系数(NEE 与 Ta 的相关系数为 -0.38,与 Ts 的相关系数为-0.35),选择空气温度作为后续分析中的代表温度变量。综上所述,最终选择 RH、 SWC、LE、LAI、Rn、H、VPD 和 Ta 构建结构方程模型,用于分析兴安落叶松生态系统 NEE 的环境影响机制。 2.2.3 结构方程模型

在构建结构方程模型时,由于感热通量(H) 难以满 足模型的构建需求,本研究基于相关性分析将其剔除重 新构建模型,并以 NEE 为因变量,选择 Ta、RH、Rn、VPD 和 SWC 作为自变量,LAI 和 LE 作为中介变量,重新构 建兴安落叶松生态系统的结构方程模型,以揭示环境因 子对 NEE 的直接和间接影响路径。研究结果表明,调 整后的模型拟合效果良好(X²/df=2.818,GFI=1.000, AGFI=0.999, RMSEA=0.009),能够解释 NEE 变化的 59%(图 5),该模型能够较好地反映兴安落叶松生态系 统中 NEE 的环境驱动机制。

结合图 5 和表 3 所示,尽管 Ta 和 RH 能够对 NEE 产生直接效应,但其路径系数未通过显著性检验,其对 NEE 的主要影响是通过间接途径实现的,路径系数分 别为-0.25 和-0.30。温度升高可能会加速兴安落叶松 的呼吸作用,从而增加碳排放,但由于其耐寒特性,高温 效应可能通过促进 LAI 和蒸散来缓解其对 NEE 的负面 影响。Rn 对 NEE 的间接效应大于直接效应,且二者均 表现为负向效应。SWC 和 VPD 对 NEE 产生了显著正 向直接效应,VPD 又通过其他途径还表现出负向的间 接效应。潜热通量和 LAI 对 NEE 均产生了显著的负向



 $\chi^2/df = 2.818$ GFI = 1.000 AGFI = 0.999 RMSEA = 0.009

 图 5 兴安落叶松生态系统 NEE 与个环境因子的结构方程模型
 Fig.5 Structural equation model of NEE and various environmental factors in the *Larix Gmelinii* ecosystem
 箭头数值表示路径系数大小,实线通过显著性检验(P<0.001),

虚线表示未通过显著性检验,红色箭头为正向作用,蓝色箭头为 负向作用;间接效应仅显示显著结果。 χ^2/df :卡方/自由度 Chi square/degree of freedom; GEI: 拟合优度指数 Goodness of fit index;AGFI:调整拟合优度指数 Adjusted goodness of fit index; RMSEA:近似均方根误差 Root mean square error of approximation

直接效应,其路径系数分别为-0.60 和-0.13。各环境因子对 NEE 的总效应顺序为(表3):LE > Rn > RH > Ta > LAI > VPD > SWC。其中,LE、Rn、LAI、RH 和 Ta 对 NEE 的总影响为显著的负向作用,而 VPD 和 SWC 对 NEE 的总影响为显著的正向作用,潜热通量和净辐射是影响 NEE 变化的关键因子。

Table 3 Effects and standa	ardized coefficien	ts of various e	nvironmental	factors on NE	E in the Larix	Gmelinu ecosy	ystem	
影响类型		环境因子 Environmental factor						
Impact type	Ta	Rn	RH	SWC	VPD	LE	LAI	
直接影响 Direct Impact	—	-0.23	—	0.03	0.29	-0.60	-0.13	
间接影响 Indirect effect	-0.25	-0.32	-0.30	—	-0.22	—	-0.09	
总影响 Total impact	-0.25	-0.55	-0.30	0.03	0.07	-0.60	-0.22	

表 3 环境因子对兴安落叶松生态系统 NEE 影响效应及标准化系数

2.3 基于机器学习方法的 NEE 模拟

通过结构方程模型筛选出的因子作为特征变量输入随机森林模型,计算特征重要性并依据排序选择不同 数量的前 n 个特征组合。随后,基于每组特征组合计算袋外精度,重复循环 100 次以获得平均值。最终选择 袋外精度最高的特征组合作为模型输入变量^[48]。结果表明,影响 NEE 的最优变量组合为:*LE*、*Rn*、LAI、*Ta*、 *RH*、VPD 和 SWC,袋外精度为 0.66。为确保模型的泛化能力和稳健性,进一步采用十折交叉验证,并通过 *R*²、 RMSE 和 MAE 指标对模型的拟合优度进行综合评估。 4 种模型对半小时尺度 NEE 测试值与模拟值的验证散点图如图 6 所示,模拟值与实际值的变化趋势拟合 结果如图 7 所示。通过对比可以看出,各模型均能在不同程度上较好地拟合 NEE,预测值与实际值的变化趋 势较为一致,但不同模型的预测效果仍存在一定差异。半小时尺度各机器学习对 NEE 模拟的评价指标值表 明(表 4),除 ANN 外,RF、XGBoost 和 SVM 模型在测试集上的 R^2 均大于 0.70,RMSE 均小于 0.14 mg CO₂ m⁻² s⁻¹,MAE 均小于 0.08 mg CO₂ m⁻² s⁻¹。其中,XGBoost 模型在各指标上表现最优,测试集 R^2 为 0.754,RMSE 为 0.128 mg CO₂ m⁻² s⁻¹,MAE 为 0.066 mg CO₂ m⁻² s⁻¹。其次是 RF 模型, R^2 为 0.745,RMSE 为 0.130 mg CO₂ m⁻² s⁻¹,MAE 为 0.067 mg CO₂ m⁻² s⁻¹。相比之下,SVM 模型预测性略逊于 XGBoost 和 RF,其测试集 R^2 分别低 0.047和 0.038,RMSE 分别高 0.012 和 0.010 mg CO₂ m⁻² s⁻¹,MAE 分别低 0.008 和 0.007 mg CO₂ m⁻² s⁻¹。ANN 模型表现相对较弱, R^2 为 0.685,RMSE 为 0.142 mg CO₂ m⁻² s⁻¹,MAE 为 0.076 mg CO₂ m⁻² s⁻¹。



图 6 2014—2018 年各机器学习半小时尺度 NEE 测试值与模拟值的验证散点图

Fig.6 Scatter plot of validation between NEE test values and simulated values for 30 minutes of machine learning from 2014 to 2018 RF:随机森林 Random forest;XGBoost:极端梯度提升 eXtreme gradient boosting;SVM:支持向量机 Support vector machine;ANN:人工神经网络 Artificial neural network

- .	
表 4	各机器学习下半小时尺度 NEE 训练集和测试集的模拟精度指标

Table 4 Simulation accuracy	metrics for NEE	training and te	esting sets under va	arious machine	learning approach	ies
	决定系	系数	均方根	误差	平均绝为	对误差
模型	R^2		$\text{RMSE/(mg CO_2 m^{-2} s^{-1})}$		MAE/(mg CO ₂ m ⁻² s ⁻¹)	
Model	训练集	测试集	训练集	测试集	训练集	测试集
	Training set	Test set	Training set	Test set	Training set	Test set
随机森林 Random forest	0.841	0.745	0.101	0.130	0.053	0.067
极端梯度提升 eXtreme gradient boosting	0.878	0.754	0.088	0.128	0.051	0.066
支持向量机 Support vector machine	0.709	0.707	0.137	0.140	0.060	0.074
人工神经网络 Artificial neural network	0.707	0.685	0.139	0.142	0.075	0.076

R²:决定系数 Coefficient of determination; RMSE:均方根误差 Root mean square error; MAE:平均绝对误差 Mean absolute error



图 7 2014—2018 年各机器学习半小时尺度 NEE 模拟值与实测值拟合图 Fig.7 Fitting chart of simulated and measured NEE values for 30 minutes of machine learning from 2014 to 2018

基于半小时尺度的完整特征变量数据,利用4种机器学习模型对兴安落叶松生态系统的 NEE 进行模拟, 生成连续半小时尺度 NEE 模拟值。将模拟结果累加至月尺度,生成月尺度 NEE 模拟值,与涡度相关法观测 得到的日尺度 NEE 进行对比分析。4 种模型在月尺度上的模拟结果总体能够反映兴安落叶松 NEE 的季节性 变化特征,这一趋势与生态系统的碳交换规律相符。然而,在捕捉月尺度变化的幅度和细节上,不同模型的模 拟效果存在显著差异(图 8)。XGBoost 和 RF 模型在趋势方向上总体与观测值较为一致,能够较好地反映兴 安落叶松生态系统 NEE 的季节性变化特征,但也存在部分月份的高估或低估现象。相比之下,SVM 模型在趋 势方向上与前两个模型一致,但在变化幅度的还原上不如 XGBoost 和 RF。ANN 模型整体表现较弱,其预测值 波动较小,不仅在趋势方向上偏离观测值,对季节性变化的捕捉能力也明显不足。从模拟精度来看(表 5), XGBoost、RF 和 SVM 模型在月尺度上指标结果较为接近,其中 XGBoost 模型表现最佳,*R*²为 0.945,RMSE 为 8.056 g C m⁻²月⁻¹,MAE 为 5.668 g C m⁻²月⁻¹。ANN 性能最差,*R*² 仅为 0.709,RMSE 和 MAE 分别为 18.509 g C m⁻²月⁻¹和 14.764 g C m⁻²月⁻¹,在月尺度上的模拟精度明显低于其他模型。

Table 5 Simulation accuracy indicators of NEE training and testing sets for each machine learning month							
模型	决定系数	均方根误差	平均绝对误差				
Model	R^2	RMSE/(g C m ⁻² 月 ⁻¹)	MAE/(g C m ⁻² 月 ⁻¹)				
随机森林 Random forest	0.939	8.502	5.820				
极端梯度提升 eXtreme gradient boosting	0.945	8.056	5.668				
支持向量机 Support vector machine	0.928	9.210	6.698				
人工神经网络 Artificial neural network	0.709	18.509	14.764				



图 8 2014—2018 年各机器学习月尺度模拟值与实测值累积拟合图

Fig.8 Cumulative fitting chart of monthly scale simulated values and measured values for machine learning from 2014 to 2018

3 讨论

3.1 环境因子对 NEE 的影响分析

在生长季内,兴安落叶松的 NEE 呈现典型的昼夜"U"型变化曲线,这种变化是由植物的光合作用和呼吸 作用共同驱动所形成的,研究发现,其碳吸收峰值通常出现在上午9点,与其他生态系统在中午或下午达到峰 值的情况显著不同^[49-50]。这一现象与兴安落叶松所处高纬度地区的地理位置和特殊气候条件密切相关。由于 日出较早,清晨光照充足,温度和湿度适宜,光合作用快速启动,碳吸收量迅速上升,并在上午9点左右达到峰值。 然而,在中午强光下,气孔逐渐关闭以减少水分流失,抑制光合作用,形成"光合午休"现象。非生长季阶段,尽管 光合作用几乎停滞,土壤微生物和植物呼吸作用仍在进行,导致生态系统表现为二氧化碳的净排放状态。 基于相关系数、共线性检验及结构方程模型的结果进行分析,兴安落叶松 NEE 受到多重环境因素的协同 影响。其中潜热通量和净辐射是影响兴安落叶松生态系统 NEE 的主导因素。He 等^[21]的研究利用 ANN 模型 模拟多种生态系统的 NEE,发现潜热通量不仅通过能量分配影响碳通量,还通过调节气孔行为显著增强碳吸 收能力。徐亚彬等^[51]通过涡动相关数据观测长白山森林生态系统,发现潜热通量与 CO₂通量之间存在紧密 联系,验证了潜热通量在碳水耦合过程中的核心地位。Zhang 等^[52]研究指出,潜热通量作为净辐射的主要消 耗项,通过调节 VPD 间接影响气孔行为,在能量平衡与碳循环中发挥了重要作用。在本研究中,结构方程模 型分析发现 Rn 和 VPD 对 LE 的路径系数分别为 0.52 和 0.30,表明 Rn 通过促进光合作用和蒸腾作用增强,显 著推动了 LE 的增加。VPD 则通过调节气孔行为,影响蒸腾作用强度,进一步促进 LE 的增加^[53]。在动态平 衡机制下,当潜热通量消耗过多水分时,气孔导度降低,限制 CO₂的吸收;而在水分充足时,潜热通量的增加促 进气孔开放,增强了 CO₂的吸收^[54],这些发现进一步揭示了 LE \Rn 和 VPD 在兴安落叶松生态系统碳水耦合 过程中的协同作用。

太阳辐射是碳水循环的主要能量来源^[55],牛晓栋等^[50]在天目山衰老林中指出,NEE 随着辐射升高显著 增加。朱苑等^[56]的研究进一步表明,辐射对东北地区翅片叶桦林碳通量最为显著。在本研究中,*Rn* 不仅直 接影响光合作用效率,还能通过影响 LAI 和 *LE* 间接影响碳吸收,这种双重负向效应突出表明,辐射驱动的水 热动态是影响兴安落叶松生态系统 NEE 的重要机制,其增加能够为植物提供更多光合作用能量,驱动 CO₂的 固定,从而显著提高碳吸收效率。特别是在兴安落叶松生态系统中,辐射驱动的水热动态机制发挥着核心作 用,进一步推动了 NEE 的提升,这一点与周丽艳等^[30]的研究一致。

LAI 在该生态系统中受到温度、湿度等环境因素的显著影响,通过对光合作用和水分消耗的调节作用影响 NEE^[57]。祁连山青海云杉林中, NEE 主要受植被指数和净辐射的综合影响,随着植被指数的增加,其碳汇 功能显著增强。与吕富成等^[58]的研究不同,本研究的结构方程模型量化了各环境因子对 NEE 的复杂交互效 应,揭示 *LE*、LAI 在 NEE 动态变化中具有重要的中介作用。此外, VPD 和 *RH* 通过调节蒸腾作用影响植物的 水分利用效率,与温度变化密切相关^[59]。当空气干燥时,植物蒸腾作用加剧,水分利用效率下降,从而抑制碳 吸收;而水分亏缺则是高温下限制光合作用和固碳能力的主要原因^[60]。尽管在本研究中 SWC 对 NEE 的直 接效应较小,但作为光合作用的必要条件,其在维持生长季内碳吸收中不可或缺。

3.2 不同机器学习对 NEE 模拟的对比分析

本研究基于随机森林模型的袋外精度对环境因子进行了优化筛选,确定的最优变量组为 LE、Rn、LAI、 Ta、RH、VPD 和 SWC,这与结构方程模型中路径分析得出的主导因子高度一致,为 NEE 模型的构建奠定了坚 实的变量基础。

对 4 种机器学习模型的对比分析表明,各模型虽然在算法基础和计算路径上存在差异,但都能够捕捉到 NEE 与环境变量之间的非线性依赖关系,从而实现较为精准的估算。本研究中 XGBoost 表现最佳,其次为 RF,SVM 和 ANN 表现稍逊。XGBoost 和 RF 都是基于树结构的集成学习方法,XGBoost 是基于提升树方法通 过逐步学习残差的方式,能够在处理高维复杂数据时有效降低误差^[27]。RF 主要通过独立训练多棵决策树并 集成这些结果来降低模型的方差。相关研究表明,集成学习模型在复杂生态系统中能够准确捕捉 NEE 的动 态变化趋势,在短时间尺度下对 NEE 波动具有较强的还原能力。Reichstein 等^[41]在极端气候条件下比较了 这两种模型对不同生态系统 NEE 的预测能力,结果表明 NEE 预测效果有所下降,但 XGBoost 相比 RF 表现出 更高的鲁棒性,且计算效率更高,这与本研究一致。Zeng 等^[62]使用随机森林方法将涡度协方差数据扩展至全 球尺度,在分析 1999—2019 年间的 NEE 动态变化时表现出较高的准确性,其所生成的数据产品被广泛应用 于全球碳循环模型的验证和改进。相比之下,SVM 模型适合小样本数据,在处理大样本时,由于计算复杂度 高且对超参数敏感,其性能有所下降。Yao 等^[63]在利用机器学习方法估算中国 NEP(-NEE)时,SVM 在月尺 度上能够按近 RF 的趋势捕捉能力,但在高峰期和低谷期的表现仍显不足,尤其对极值的还原能力较弱。在 对三峡水库 CO₂通量进行预测时,SVM 模型在性能方面不如 RF 模型,这可能与对超参数和核函数的调整优 化不足有关^[40]。ANN 模型在本研究中的表现受限,尽管神经网络在处理非线性关系方面具有潜在优势^[21,64-65],但其预测能力依赖于网络结构的合理性和超参数的优化。相关研究表明,ANN 在数据复杂度和数据量大的情况下的性能会有下降趋势,在数据量少时则存在饱和效应^[66]。在本研究中,由于兴安落叶松生态系统位于寒温带气候区,地理环境要素之间交互关系复杂,ANN 在训练过程中未能充分捕捉这些复杂变量间的非线性依赖性。此外,受数据高维性和训练过程中参数优化不足的影响,ANN 模型可能出现收敛至局部

13

总体而言,本研究中 XGBoost 在处理复杂非线性关系和大规模数据集方面展现了卓越性能。机器学习方法在生态系统通量预测中显示出显著潜力,尤其是在高维数据和复杂变量相互作用下,其自动化特征选择和误差优化极大地提升了预测的可靠性和精确性。

最优解的情况,最终导致其预测效果明显低于 RF 和 XGBoost 等模型。

4 结论

(1)受气象因子调控,2014—2018年兴安落叶松生态系统整体表现为碳汇,具有显著的季节变化。生长季阶段 NEE 表现出单峰形变化特征,碳吸收高峰出现在 7月(NEE 为-67.57 g C m⁻²月⁻¹)。其中,生长季阶段 5月和 9月针叶林生态系统为碳源,6月、7月和 8月表现为碳汇,非生长阶段则整体表现为碳源。

(2) 潜热通量和净辐射是影响兴安落叶松生态系统 NEE 变化的主要驱动因子, LAI 是碳通量变化的重要 中介变量。空气温度、相对湿度、土壤含水量通过直接与间接的方式对 NEE 进行共同调控。

(3)基于随机森林和结构方程模型筛选出的兴安落叶松生态系统 NEE 模拟的最优变量组合为 LE 、Rn、 LAI、Ta、RH、VPD 和 SWC。四种机器学习方法(RF、XGBoost、SVM 和 ANN)均能较好地模拟兴安落叶松 NEE, 其中以 XGBoost 表现最佳,RF 次之。

参考文献(References):

- Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, Rödenbeck C, Altaf Arain M, Baldocchi D, Bonan G B, Bondeau A, Cescatti A, Lasslop G, Lindroth A, Lomas M, Luyssaert S, Margolis H, Oleson K W, Roupsard O, Veenendaal E, Viovy N, Williams C, Ian Woodward F, Papale D. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science, 2010, 329(5993): 834-838.
- [2] 王春林,于贵瑞,周国逸,闫俊华,张雷明,王旭,孙晓敏.鼎湖山常绿针阔叶混交林 CO₂通量估算.中国科学(D辑:地球科学),2006, 36(S1):119-129.
- [3] Rollinger J L, Strong T F, Grigal D F. Forested soil carbon storage in landscapes of the northern Great Lakes region. Management of carbon sequestration in soil: CRC Press, 2019. 335-350.
- [4] 于贵瑞,牛栋,王秋凤.《联合国气候变化框架公约》谈判中的焦点问题.资源科学,2001,23(6):10-16.
- [5] 唐欢,叶剑,高振翔,孟韩春,刘隆,李成.我国典型生态系统净 CO₂交换量与蒸散量的多机器学习模型对比分析.干旱区资源与环境, 2022, 36(8): 92-100.
- [6] 王兴昌,王传宽.森林生态系统碳循环的基本概念和野外测定方法评述.生态学报,2015,35(13):4241-4256.
- [7] Zhang K D, Gong Y, Escobedo F J, Bracho R, Zhang X Z, Zhao M. Measuring multi-scale urban forest carbon flux dynamics using an integrated eddy covariance technique. Sustainability, 2019, 11(16): 4335.
- [8] Grossiord C, Granier A, Gessler A, Pollastrini M, Bonal D. The influence of tree species mixture on ecosystem-level carbon accumulation and water use in a mixed boreal plantation. Forest Ecology and Management, 2013, 298: 82-92.
- [9] Xie J, Jia X, He G M, Zhou C X, Yu H Q, Wu Y J, Bourque C P, Liu H, Zha T S. Environmental control over seasonal variation in carbon fluxes of an urban temperate forest ecosystem. Landscape and Urban Planning, 2015, 142: 63-70.
- [10] Wharton S, Falk M. Climate indices strongly influence old-growth forest carbon exchange. Environmental Research Letters, 2016, 11(4): 044016.
- [11] 李小梅, 张秋良. 兴安落叶松林生长季碳通量特征及其影响因素. 西北农林科技大学学报(自然科学版), 2015, 43(6): 121-128.
- [12] 刘雅琪,郑惠旭,张晶,焦麟杰,张一平,陈星,陈耀亮,宋清海.西双版纳热带雨林与橡胶林净生态系统碳交换特征及其影响因子比较.地球与环境,2024,1-12[2025-04-18].https://doi.org/10.3724/EE.1672-9250.2024.52.032
- [13] 裴薇薇,杨喆,王云英,王新,杜岩功.祁连山区青海云杉林碳汇特征及调控因子.中国农业科技导报,2024,26(1):226-233.
- [14] 李润东,范雅倩,冯沛,宋泽,李鑫豪,闫惠娟,马莉,查天山.北京松山天然落叶阔叶林生态系统净碳交换特征及其影响因子.应用生

态学报, 2020, 31(11): 3621-3630.

- [15] Makkonen M, Huttunen S, Primmer E, Repo A, Hildén M. Policy coherence in climate change mitigation: an ecosystem service approach to forests as carbon sinks and bioenergy sources. Forest Policy and Economics, 2015, 50: 153-162.
- [16] 齐建东,黄俊尧. 基于深度学习的草地生态系统净碳交换模拟. 农业机械学报, 2020, 51(6): 152-161.
- [17] 窦佳慧,梁宇,怀保娟,吴苗苗,刘波,马天啸,王耀.不同地形条件下青藏高原森林生产力和碳收支动态. 生态学杂志, 2024, 43(6): 1521-1530.
- [18] 康满春,朱丽平,许行,查同刚,张志强. 基于 Biome-BGC 模型的北方杨树人工林碳水通量对气候变化的响应研究. 生态学报, 2019, 39 (7): 2378-2390.
- [19] Zhao J F, Liu D S, Zhu Y J, Peng H W, Xie H F. A review of forest carbon cycle models on spatiotemporal scales. Journal of Cleaner Production, 2022, 339: 130692.
- [20] Zhang L L, Zhang J Y, Lv H, Sun B W. Analysis of carbon flux in terrestrial ecosystems from GOSAT data in China. E3S Web of Conferences, 2018, 53: 03012.
- [21] He H L, Yu G R, Zhang L M, Sun X M, Su W. Simulating CO₂ flux of three different ecosystems in ChinaFLUX based on artificial neural networks. Science in China Series D: Earth Sciences, 2006, 49(2): 252-261.
- [22] 冯新妍, 贾昕, 黄金泽, 高圣杰, 袁敏, 刘甜甜, 靳川. ANN-BiLSTM 模型在温带荒漠灌丛碳通量数据缺失值插补中的应用. 北京林业大学学报, 2023, 45(9): 62-72.
- [23] 王少影,张宇,孟宪红,宋敏红,尚伦宇,苏有琦,李照国.机器学习算法对涡动相关缺失通量数据的插补研究.高原气象,2020,39 (6):1348-1360.
- [24] Renchon A A, Griebel A, Metzen D, Williams C A, Medlyn B, Duursma R A, Barton C V M, Maier C, Boer M M, Isaac P, Tissue D, Resco de Dios V, Pendall E. Upside-down fluxes down under: CO₂ net sink in winter and net source in summer in a temperate evergreen broadleaf forest. Biogeosciences, 2018, 15(12): 3703-3716.
- [25] 温旭丁,赵仲辉,邓湘雯.基于神经网络的杉木人工林碳通量影响因素的选择.广东林业科技, 2014, 30(2): 23-28.
- [26] 陈亮,周国模,杜华强,刘玉莉,毛方杰,徐小军,李雪建,崔璐,李阳光,朱迪恩.基于随机森林模型的毛竹林 CO₂通量模拟及其影响因子.林业科学,2018,54(8):1-12.
- [27] Liu J Z, Zuo Y J, Wang N N, Yuan F H, Zhu X H, Zhang L H, Zhang J W, Sun Y, Guo Z Y, Guo Y D, Song X, Song C C, Xu X F. Comparative analysis of two machine learning algorithms in predicting site-level net ecosystem exchange in major biomes. Remote Sensing, 2021, 13 (12): 2242.
- [28] 燕宇杰,周莉,周广胜,贾丙瑞,宋家欣,张森. 2014—2018 年呼中北方林森林生态系统碳水通量观测数据集.中国科学数据(中英文网络版), 2023, 8(2): 7-16.
- [29] Kondo F, Tsukamoto O. Experimental validation of WPL correction for CO₂ flux by eddy covariance technique over the asphalt surface. Journal of Agricultural Meteorology, 2012, 68(4): 183-194.
- [30] 周丽艳,贾丙瑞,周广胜,曾伟,王宇.中国北方针叶林生长季碳交换及其调控机制.应用生态学报,2010,21(10):2449-2456.
- [31] Wang Y, Zhou L, Ping X Y, Jia Q Y, Li R P. Ten-year variability and environmental controls of ecosystem water use efficiency in a rainfed maize cropland in NorthEast China. Field Crops Research, 2018, 226: 48-55.
- [32] 匡苗苗,周广胜,周梦子.西藏高寒草地物种多样性和生产力的环境驱动机制.生态学报, 2024, 44(14): 6254-6264.
- [33] 张雷雨,杨毅,梁霄.地理加权回归模型的多重共线性诊断方法.测绘与空间地理信息,2017,40(10):28-31.
- [34] Senaviratna N A M R, Cooray T M J A. Diagnosing multicollinearity of logistic regression model. Asian Journal of Probability and Statistics, 2019, 5 (2): 1-9.
- [35] 席振华,王玉阳,马耀明,马伟强.珠峰北坡高寒灌丛草原生长季碳、水通量特征分析.高原气象, 2023, 42(4): 887-898.
- [36] 侯蒙京,殷建鹏,葛静,李元春,冯琦胜,梁天刚.基于随机森林的高寒湿地地区土地覆盖遥感分类方法.农业机械学报,2020,51(7): 220-227.
- [37] 关晓蔷,王文剑,庞继芳,孟银凤.基于空间变换的随机森林算法.计算机研究与发展, 2021, 58(11): 2485-2499.
- [38] Belgiu M, Drăguţ L. Random forest in remote sensing: a review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31.
- [39] Chen T Q, Guestrin C. XGBoost: A Scalable Tree Boosting System////Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. ACM, 2016. 785-794.
- [40] 赵敬皓, 王娜娜, 蒋嘉铭, 田亚峻. 基于 BayesianOpt-XGBoost 的煤电机组碳排放因子预测. 中国环境科学, 2024, 44(1): 417-426.
- [41] 欧阳常悦,秦宇,刘臻,梁越. 基于机器学习的水-气界面 CO₂、CH₄扩散通量预测及影响因素分析——以三峡水库为例. 湖泊科学, 2023, 35(2): 449-459.

- [42] 药静宇. 气候变化背景下干旱区碳通量的特征分析[D]. 兰州:兰州大学, 2021.
- [43] 田秋燕,赵正勇,杨旗,朱航勇.基于全球数据库的森林土壤呼吸模型研究.林业与环境科学,2019,35(6):1-6.
- [44] 齐建东,黄金泽,贾昕. 基于 XGBoost-ANN 的城市绿地净碳交换模拟与特征响应. 农业机械学报, 2019, 50(5): 269-278.
- [45] 周宇,黄辉,张劲松,孟平,孙守家.森林生态系统涡度相关法碳通量长时间连续性缺失数据插补方法的比较.中国农业气象,2021,42 (4):330-343.
- [46] 黄金泽. 神经网络模型在碳通量数据模拟的研究与应用[D]. 北京:北京林业大学, 2019.
- [47] Shanthi D L, Chethan N. Genetic algorithm based hyper-parameter tuning to improve the performance of machine learning models. SN Computer Science, 2022, 4(2): 119.
- [48] Janitza S, Hornung R. On the overestimation of random forest's out-of-bag error. PLoS One, 2018, 13(8): e0201904.
- [49] 张悦,冯会丽,王维枫,薛建辉,吴永波,于水强.洪泽湖地区杨树人工林碳水通量昼夜和季节变化特征.南京林业大学学报(自然科学版),2019,43(5):113-120.
- [50] 牛晓栋, 江洪, 张金梦, 方成圆, 陈晓峰, 孙恒. 浙江天目山老龄森林生态系统 CO2 通量特征. 应用生态学报, 2016, 27(1): 1-8.
- [51] 徐亚彬,宋博,任妙春,刘福利.长白山森林生态系统二氧化碳通量与涡动相关研究.北京农业,2012,(27):100-101.
- [52] Zhang J H, Chen P S, Liu D Q, Liang R H, Cui G S. Vapor pressure deficit plays a pivotal role in the carbon dioxide sink of the Jingxin Wetland. EGUsphere(preprint), 2024, 1-24[2025-04-15]. https://doi.org/10.5194/egusphere-2024-2787.
- [53] Baldocchi D D, Vogel C A. Energy and CO₂ flux densities above and below a temperate broad-leaved forest and a boreal pine forest. Tree Physiology, 1996, 16(1/2): 5-16.
- [54] Jin Y, Liu J H, Zhang X. Energy balance closure problem over a tropical seasonal rainforest in Xishuangbanna, southwest China: role of latent heat flux. Water, 2022, 14(3): 395.
- [55] Tong X J, Mu Y M, Zhang J S, Meng P, Li J. Water stress controls on carbon flux and water use efficiency in a warm-temperate mixed plantation. Journal of Hydrology, 2019, 571: 669-678.
- [56] 朱苑, 刘帆, 王传宽, 王兴昌. 帽儿山温带落叶阔叶林净生态系统碳交换的日变化及光响应特征. 应用生态学报, 2020, 31(1): 72-82.
- [57] Xu Z P, Man X L, Cai T J, Shang Y X. How potential evapotranspiration regulates the response of canopy transpiration to soil moisture and leaf area index of the boreal larch forest in China. Forests, 2022, 13(4): 571.
- [58] 吕富成, 酒名扬, 韩立钦, 陈晓虹. 中国北方针叶林碳通量动态变化及其影响因素分析. 气候与环境研究, 2024, 29(3): 243-252.
- [59] 唐祥,陈文婧,李春义,查天山,吴斌,王小平,贾昕.北京八达岭林场人工林净碳交换及其环境影响因子.应用生态学报,2013,24 (11):3057-3064.
- [60] Carrara A, Janssens I A, Curiel Yuste J, Ceulemans R. Seasonal changes in photosynthesis, respiration and NEE of a mixed temperate forest. Agricultural and Forest Meteorology, 2004, 126(1/2): 15-31.
- [61] Reichstein M, Bahn M, Ciais P, Frank D, Mahecha M D, Seneviratne S I, Zscheischler J, Beer C, Buchmann N, Frank D C, Papale D, Rammig A, Smith P, Thonicke K, van der Velde M, Vicca S, Walz A, Wattenbach M. Climate extremes and the carbon cycle. Nature, 2013, 500 (7462): 287-295.
- [62] Zeng J Y, Matsunaga T, Tan Z H, Saigusa N, Shirai T, Tang Y H, Peng S S, Fukuda Y. Global terrestrial carbon fluxes of 1999—2019 estimated by upscaling eddy covariance data with a random forest. Scientific Data, 2020, 7(1): 313.
- [63] Yao Y T, Li Z J, Wang T, Chen A P, Wang X H, Du M Y, Jia G S, Li Y N, Li H Q, Luo W J, Ma Y M, Tang Y H, Wang H M, Wu Z X, Yan J H, Zhang X Z, Zhang Y P, Zhang Y, Zhou G S, Piao S L. A new estimation of China's net ecosystem productivity based on eddy covariance measurements and a model tree ensemble approach. Agricultural and Forest Meteorology, 2018, 253: 84-93.
- [64] Papale D, Valentini R. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Global Change Biology, 2003, 9(4): 525-535.
- [65] van Wijk M T, Bouten W. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. Ecological Modelling, 1999, 120(2/3): 181-197.
- [66] Abbasian H, Solgi E, Mohsen Hosseini S, Hossein Kia S. Modeling terrestrial net ecosystem exchange using machine learning techniques based on flux tower measurements. Ecological Modelling, 2022, 466: 109901.