

doi:10.3969/j.issn.1000-6362.2021.05.004

施光耀,周宇,桑玉强,等.基于随机森林方法分析环境因子对空气负离子的影响[J].中国农业气象,2021,42(5):390-401

基于随机森林方法分析环境因子对空气负离子的影响*

施光耀^{1,2}, 周宇^{1,2}, 桑玉强³, 张劲松^{1,2**}, 孟平^{1,2}, 蔡露露^{3,4}, 裴松义⁵, 王尧^{1,2}

(1.中国林业科学研究院林业研究所/国家林业局林木培育重点实验室, 北京 100091; 2.南京林业大学南方现代林业协同创新中心, 南京 210037; 3.河南农业大学, 郑州 450002; 4.河南省地球物理空间信息研究院, 郑州 450016; 5.国有建平县黑水机械化林场, 朝阳 122000)

摘要: 空气负离子是衡量一个地区空气清洁度的重要指标, 对人体的心理和生理机能的调节发挥着重要作用。随着森林生态旅游的兴起, 空气负离子的发生过程及影响机制已成为研究热点。本研究基于华北低丘山地森林植被主要生长季的气象数据和栓皮栎人工林空气负离子浓度观测资料, 利用机器学习中随机森林模型从非线性角度全面分析确定影响空气负离子浓度变化的重要环境因子, 通过独立样本对构建的随机森林模型进行模拟和检验, 确定模型的预测精度, 同时筛选出对空气负离子影响程度最大的环境因子。结果表明: 随机森林模型在分析环境因子对空气负离子影响方面具有较高的精度以及较好的拟合效果, 通过对模型的拟合值与实测值进行验证, 均方根误差 (RMSE) 为 59.349, 决定系数 R^2 达到了 0.887。同时利用独立样本数据对随机森林模型进行十折交叉验证, 决定系数 R^2 均达到了 0.904 以上, 且均方误差 (RMSE) 较小, 为 24.851。此外, 模型筛选出影响空气负离子的主要因素, 按重要性排序依次为颗粒物 $PM_{2.5}$ (48.037)、饱和水汽压差 (46.169)、土壤湿度 (43.984)、风速 (43.779)、紫外辐射 (41.130)、土壤温度 (40.107)、总辐射 (36.838)、大气压力 (34.532), 其中对模型重要性贡献相对较高的 3 个变量分别为颗粒物 $PM_{2.5}$ 、饱和水汽压差和风速, 它们对空气负离子的影响起决定性作用。因此, 随机森林模型适合分析环境因子对空气负离子影响, 且拟合效果精度高, 稳定性强。

关键词: 空气负离子; 环境因子; 机器学习; 随机森林; 栓皮栎

Influence of Environmental Factors on Negative Air Ion Using Random Forest Algorithm

SHI Guang-yao^{1,2}, ZHOU Yu^{1,2}, SANG Yu-qiang³, ZHANG Jin-song^{1,2}, MENG Ping^{1,2}, CAI Lu-lu^{3,4}, PEI Song-yi⁵, WANG Yao^{1,2}

(1. Research Institute of Forestry, Chinese Academy of Forestry/Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Beijing 100091, China; 2. Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037; 3. Henan Agricultural University, Zhengzhou 450002; 4. Henan Geophysical Space Information Research Institute, Zhengzhou 450016; 5. State Owned Jianping County Heishui Mechanized Forest Farm, Chaoyang 122000)

Abstract: Negative air ion (NAI) is an essential indicator for measuring the air cleanliness of a given area, which plays an important role in promoting the psychological and physiological functions of the human body. With the development of forest eco-tourism, NAI has attracted substantial attention, while research on NAI has become increasingly active, especially for the topics of the occurrence process and impact mechanism of NAI in related

* 收稿日期: 2020-10-25

基金项目: 中央级公益性科研院所基本科研业务费专项资金项目 (CAFYBB2018ZA002)

** 通讯作者: 张劲松, 研究员, 研究方向为林业气象, E-mail: zhangjs@caf.ac.cn

第一作者联系方式: 施光耀, E-mail: shiguangyao01@163.com

fields. Based on meteorological data and observation data of NAI during the main growing season of *Quercus variabilis* BI. plantation in the hilly area of North China, the random forest model in machine learning was used to analyze the environmental factors that affected NAI concentration changes from a non-linear view, and independent samples were used to simulate the random forest model to determine the prediction accuracy of the model. The estimation model of NAI was established for revealing the response mechanism and predicting the response pattern of NAI to environmental factors for further research. Results showed that the random forest model had higher accuracy and better fitting effect in analyzing the impact of environmental factors on NAI, and by verifying the fitted and measured values of the model, the root mean square error(RMSE) was 59.349, and the coefficient of determination R^2 reached 0.887. While using independent test data to 10-fold cross-validation of the random forest model, the average R^2 was above 0.904 and the root mean square error(RMSE) was small at 24.851. In addition, the model screened out that the main factors affecting NAI were particulate matter $PM_{2.5}$ (48.037), vapor pressure deficit(46.169), soil moisture(43.984), wind speed(43.779), ultraviolet radiation(41.130), soil temperature(40.107), direct radiation(36.838) and atmospheric pressure(34.532) sorted by importance scores. Among them, the three variables contributed relatively high importance to the model were particulate matter $PM_{2.5}$, vapor pressure deficit and wind speed, which prove that they play a decisive role in the variations of NAI. Therefore, the random forest model is better to simulate the NAI with high accuracy and strong stability.

Key words: Negative air ion; Environment factor; Machine learning; Random forest model; *Quercus variabilis* BI.

空气负离子(Negative air ion, NAI)是指获得多余电子的带负电荷的空气离子,因空气中氧分子化学性质活泼优先获得电子,故也称为负氧离子,广泛分布于自然环境中,如森林和湿地^[1-2]。NAI是森林有效清除空气中有机污染物、提升空气质量的重要载体^[1],已成为衡量一个地区空气清洁度的重要指标之一^[2],同时具有多种抗菌作用和生物学效应,对人体的心理和生理机能具有重要的促进作用^[3-4],被誉为“空气维生素”^[5]。

近年来,随着森林生态旅游的兴起,NAI备受关注,对NAI的研究也越来越活跃^[1,6-7],NAI发生过程及影响机制已成为生物气象、森林生态和森林康养等相关领域共同关注的研究热点^[8]。空气负离子复杂多变,受多种环境因子的共同制约,气候条件、森林覆盖率、大气流动、微粒吸附和自然地理条件等,都会影响NAI的浓度^[9]。总的来说,自然界中NAI的来源及影响因素可归纳为物理、生物两大类,物理类指闪电、雷暴、雪暴、风暴、火山爆发的放电现象以及雨水的分解等^[10-11];生物类指森林中树冠、枝叶的尖端放电以及绿色植物光合作用中的光电效应等^[12-16]。环境因子是影响NAI的重要因素,既是物理类,也是生物类的影响因素,前者涉及空气中NAI的扩散和电离以及与颗粒物的作用,后者通过改变小气候间接影响植被的生理过程,进而影

响NAI。然而,目前环境因子对NAI的影响是促进或抑制以及主导环境因子是什么,已有的研究结果无法给出统一的定论^[17-18],有的甚至出现了相互矛盾的结论^[18-20]。例如,Goldstein等^[21-22]研究表明,潮湿空气中的超氧化物可以产生低浓度过氧化氢(H_2O_2),其分解产物中的 OH^- 是NAI的重要组成部分,从而有助于NAI产生;Fujioka等^[23]研究表明NAI的迁移率随湿度的增加而降低,空气湿度的增加不利于NAI的产生;而黄世成等^[24]研究发现空气湿度与NAI的关系在不同条件下不尽相同;张双全等^[20]研究表明NAI年变化与温度和湿度均无相关性。此外,由于NAI在空气中滞留时间短且产生和消亡过程复杂,受多种环境因子的共同制约,空气中NAI波动大且表现出很强的可变性,致使有关环境因子与NAI关系的研究结论存在着诸多不确定性^[13,18-19]。因此,采用传统统计学方法可能难以全面分析环境因子的共线性以及复杂的非线性关系^[25-27]。此外,由于长时间序列的数据资料堆砌,数据的冗杂、缺失和异常以及计算分析条件的限制,使得环境因素对NAI的影响程度、气象要素对NAI的贡献大小等问题存在争议。

因此,本研究依托黄河小浪底森林生态系统国家定位观测研究站,以华北低丘山地栓皮栎(*Quercus variabilis* BI.)人工林为试验对象,在2018年和2019

年的 6–9 月定位观测获取 NAI 及微气象参数。通过应用机器学习中随机森林 (Random Forest, RF) [28] 的方法筛选出影响 NAI 浓度变化的重要环境因子, 确定随机森林模型在提取影响 NAI 变化的敏感因子中的潜在适用性, 构建随机森林估算模型, 提出并讨论所选变量的重要程度和模型的预测准确性, 并对相关结果进行验证, 旨在为进一步深入研究揭示 NAI 对环境要素变化的响应机制、预测 NAI 变化提供工作基础。

1 材料与方法

1.1 研究区概况

研究区位于黄河小浪底森林生态系统国家定位观测研究站 (35°01'45"N, 112°28'08"E, 海拔高度 410m), 站点面积约 7210hm²。该地区属暖温带大陆季风气候, 年平均气温 12.4℃, 全年日照时数为 2367.7h, 年平均降水量 641.7mm, 年蒸发量为 1400mm, 无霜期 220~230d, 0℃以上年平均有效积温为 5282℃·d, 10℃以上年均积温达 4847℃·d, 植物生长期为 210~220d。受季风气候的影响, 降水季节分配不均匀, 6–9 月平均降水量为 438.0mm, 占全年的 68.3%。土壤成分主要为棕壤和石灰岩风化母质淋溶性褐土, 土壤结构不良, 土壤中石砾含量大, 平均土壤深度为 50cm, pH 值为 7.7~8.5, 春季经常发生季节性干旱, 夏季有暴雨时易受侵蚀。研究区域主要树种为栓皮栎, 平均株高为 8m, 林分密度为 998 棵·hm⁻², 郁闭度为 0.75, 造林时间为 1972 年。

1.2 观测内容及方法

空气负离子的观测采用 RR-9411A 型空气负离子自动监测仪 (中国), NAI 测量范围 0~1.2×10⁷ion·cm⁻³, 迁移率≥0.4cm²·V⁻¹·S⁻¹, 测量精度≤±10%, 采集频率为 1s·次⁻¹, 存储周期为 5min, 仪器安装在研究区域的栓皮栎人工林中, 安装高度垂直地面 1.8m 且无遮挡物, 观测时间为 2018 年和 2019 年的 6–9 月。此外, 利用便携式标准气象站同步观测此区域林冠下方空气温湿度、风速、大气压、太阳辐射、土壤温湿度以及颗粒物浓度等 11 个气象要素, 上述监测指标均通过 CR1000 型采集器 (美国) 进行存储, 通过 DTU900C 型 GPRS 远程传输模块实现数据实时快速传输并自动上传存储至服务器, 具体变量类型详见表 1。

由于饱和水汽压差 (VPD) 是表现空气温湿度的一个综合指标, 相比于气温和相对湿度更具有代表性, 因此, 利用 Smith^[22]经验公式计算 VPD, 用于研究空气温湿度对 NAI 的影响, 其计算式为

$$VPD = 0.61078 \times (1 - RH) \times e^{\frac{17.27 \times Ta}{Ta + 237.3}} \quad (1)$$

式中, Ta 为气温 (℃); RH 为空气相对湿度, 取值 (0, 1), 无量纲。

1.3 数据处理与分析

1.3.1 数据筛选

利用 R 语言对空气负离子数据进行初步筛选, 筛选过程如下, (1) 对时间序列进行筛选, 排除因

表 1 试验观测的环境因子变量类型

Table 1 Types of environmental factor variables observed in the experiment

变量名 Variable name	测量范围 Measuring range	精度 Precision	离地高度 Height from ground	型号 Model	产地 Region
温度 Temperature ambient(Ta)	-80~60℃	±0.12℃	1.8m	HMP155	芬兰 Finland
相对湿度 Relative humidity(Rh)	0~100%	±1PP	1.8m	HMP155	芬兰 Finland
风速 Wind speed(Ws)	0~60m·s ⁻¹	±2%	1.8m	WindSonic	英国 Britain
大气压 Atmospheric pressure(AP)	600~1100hPa	±1hPa	1.8m	AV-410BP	美国 USA
光合有效辐射 Photosynthetically active radiation(PAR)	400~700nm	±5%	2m	LI-190SB	美国 USA
总辐射 Direct radiation(Dr)	400~1100nm	±2%	2m	LI200X	美国 USA
紫外辐射 Ultraviolet radiation(UV)	210~380nm	±2%	2m	TUVR	美国 USA
净辐射 Net radiation(Rn)	300~2800nm	±1%	2m	CNR4	荷兰 Netherlands
土壤湿度 Soil moisture(Ms)	0~100%	±1PP	-0.2m	AV-EC5	美国 USA
土壤温度 Soil temperature(Ts)	-45~65℃	±0.2℃	-0.2m	AV-10T	美国 USA
颗粒物 Particulate matter(PM _{2.5})	1~1000μg·m ⁻³	±1%	1.8m	RR-9421	中国 China

注: 精度一列中 PP 为“百分点”。

Note: PP in the precision means percent point.

设备存储中断、故障引起的时间序列不连续以及异常数据；(2) 将每个数值与其前后的值进行对比，若该值小于前数值的3倍或1/3，舍弃并记录为NA；(3) 将连续6个或以上的相同数据值判定为异常值，记录为NA；(4) 对小于10的数值进行差值计算，以前后2个数据取均值并取整，记录为该时刻插值后的数值；(5) 赋值后再次筛选，将仍小于10的数值剔除，记录为NA，输出所有有效数据；而对气象因子的筛选只进行前两部分的筛选，剔除异常值。试验期间共收集34000组完整数据，剔除异常值后选取有效数据约30000组进行分析。

1.3.2 随机森林模型的建立与分析

1.3.2.1 随机森林模型简介

随机森林模型是一种通过聚集大量决策树来提高预测准确性的组合模型，该模型可用于处理非线性关系、分类、回归、高阶相关、评估变量的重要性以及插值丢失的数据等^[29]。已有研究表明随机森林在筛选一些复杂多变的因素干扰方面具有准确的预测能力^[30-32]，该方法已应用于生态学研究，显示了高精度和预测建模变量间复杂相互作用的能力。首先，对训练数据利用自助抽样形成N个样本；其次，对每个样本分别建立回归树模型，每一回归树都是由部分独立样本生成；最后，以N个回归树模型结果的平均值来确定因变量的预测值。此外，该模型要求每个二叉分裂点仅考虑预测变量的一个子集，从而使得弱相关性变量获得更多参与回归树模型构建的机会，因此所得到的回归树具有更高的可信度^[28,33]。为了获得优化的随机森林模型，本研究采用替换取样法将原始数据分为6个随机子样本集（每个样本5000组数据），选取其中任意1个样本集作为独立样本，其余5个样本作为模型样本，每个模型样本集随机分为训练样本（70%）和验证样本（30%），对随机森林模型进行训练和验证，而独立样本对通过测试的模型进行模拟，再次验证模拟的精度。随机森林模型构建利用R语言RandomForest包来实现。

1.3.2.2 随机森林模型精度评价

采用均方根误差（RMSE）和决定系数（ R^2 ）对模拟结果进行精度检验，其计算式为

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

$$R^2 = 1 - \sum_{i=1}^n (p_i - o_i)^2 / \sum_{i=1}^n (o_i - \bar{o}_i)^2 \quad (3)$$

式中， o_i 和 p_i 分别为空气负离子的观测值和模型拟合值； \bar{o}_i 为观测值的均值。

为了进一步检验模型泛化能力，基于独立样本数据进行十折交叉验证法（10-fold cross-validation）评价已构建好的随机森林模型的预测误差。十折交叉验证法是将所有数据随机分成大小相近的10组，先以第1组数据作为验证数据，剩下9组数据作为拟合模型；再以第2组数据作为验证数据，剩下9组用作拟合模型。依次类推，重复10次，以10次测试误差的平均值作为最后精度评价。

1.3.2.3 随机森林模型中重要性评价

随机森林模型中，在拟合或分类时会用到每个通过自助抽样形成的大约2/3样本数据，剩余样本数据称为袋外样本数据（OOB），用于评估模型整体误差以及变量的重要性。若 X_j 为输入变量之一，则计算变量 X_j 在第n株树的重要性 VI_n 时，首先利用自助抽样的样本数据建立回归树模型 T_n ，计算OOB的预测误差，然后随机置换变量 X_j 的观测值，再次建立回归树模型 T'_n ，计算OOB'（随机置换变量观测后袋外数据）的预测误差，将2次袋外数据预测误差的差值进行标准化处理后，将所有回归树结果取均值即为变量 X_j 在第n株树的重要性 $VI_n(X_j)$ ^[34]，其计算式为

$$VI_n(X_j) = \left\{ \sum_{i=1}^{N_{OOB}} I[f(X_i) = f_n(X_i)] - \sum_{i=1}^{N_{OOB}} I[f(X_i) = f_n(X'_i)] \right\} / N_{OOB} \quad (4)$$

变量 X_j 在整个随机森林中的重要性得分 $VI(X_j)$ ，这个指标代表准确度的平均下降程度，数值越大，自变量对因变量的影响越大，负值表示模型预测受到因变量的负面影响^[35]，其计算式为

$$VI(X_j) = \sum_{n=1}^{N_t} VI_n(X_j) / N_t \quad (5)$$

式中， N_{OOB} 为袋外样本数； $f(X_i)$ 为袋外数据中第i个观测值； $f_n(X_i)$ 为在随机置换变量 X_j 的观测值前第n株树上袋外数据第i个观测值所对应的预测值； $f_n(X'_i)$ 为在随机置换变量 X_j 的观测值后第n株树上袋外数据第i个观测值所对应的预测值； $I[f(X_i)=f_n(X_i)]$ 和 $I[f(X_i)=f_n(X'_i)]$ 为判别函数，当 $f(X_i)=f_n(X_i)$ 或 $f(X_i)=f_n(X'_i)$ 时，取值为1，否则为0；

N_t 代表随机森林模型中树的数量。

1.3.2.4 随机森林模型输入环境变量

以空气负离子浓度为因变量, 随机森林模型输入的环境因子包括 VPD、Ws、AP、PAR、Dr、UV、Rn、Ms、Ts 和 $PM_{2.5}$ 为自变量, 其中 VPD 由式 (1) 计算得到, 代表空气温湿度的综合指标, 其余 9 个环境因子均由便携式移动气象站观测获取。皮尔逊相关分析表明, 这 10 个环境因子与空气负离子 (NAI) 均极显著相关 ($P < 0.001$)。

以输入参数对模型的重要性得分及其影响显著性为依据, 评价环境因子对空气负离子的影响。环境因子重要性得分及显著性检验通过 R 语言 rfPermute 包实现。

2 结果与分析

2.1 随机森林模型参数优化结果

随机森林模型中有 2 个基本参数, 一是模型中每次拆分时尝试的变量个数 (M); 二是模型中树的数量 (N_t)。变量数 (M) 是通过内部随机森林函数确定的, 该函数从默认值变量总数的 1/3 开始计算最佳变量数。目前, 随机森林模型中 M 的取值大多基于经验公式, 本研究采用逐次筛选的方法确定最优参数 M , 即尝试所有 M 的取值 (输入变量数 1~10), 取模型误差最小时对应的 M 。图 1a 显示, 当 M 取值为 6 时, 模型误差达到最小。随机森林模型中树的数量 (N_t) 是为了避免潜在的过度拟合, 首先使用训练样本数据对模型进行预处理, 运行时将树的数量 (N_t) 设置为 1000, 运行结果见图 1b, 在树的数量 (N_t) 达到 300 以后模型误差趋于平稳, 而变量数 (M) 达到 6 以后模型误差最低, 因此最终变量数 (M) 和树的数量 (N_t) 的取值分别为 6 和 300。

2.2 随机森林模型输出的环境因子重要性排序

根据随机森林模型的参数优化结果对 5 个训练样本数据进行分析, 模拟得到了 5 个训练样本集的重要性得分 $VI(X_j)$, 根据重要性得分的高低对所选变量进行排序, 结果见图 2。由图中可见, 5 个训练样本得到的结果较为一致, 所选的 10 个变量重要性得分均超过了 20, 其中 $PM_{2.5}$ 在影响空气负离子 (NAI) 变化方面发挥了决定性作用。利用对应的 5 个验证数据集对结果进一步验证表明, 使用随机森林模型

创建的 5 个中间模型显示了 88% 以上的方差解释比例, 空气负离子观测值与预测值的相关性均在 0.93 以上 (表 2)。为了使用较少的变量也能较好地反映模型的预测精度, 剔除部分重要性得分较低的变量, 如图 2 所示, 5 个样本的重要性得分显示在重要性得分约 30 时出现了阈值, 因此将低于该阈值的变量进行剔除, 选取高于该阈值的变量作为随机森林模型的最终变量。因此, 对 5 个样本的重要性得分综合考虑, 最终选取以下 8 个变量作为最终模型变量, 根据重要性从高到低排序依次为颗粒物 ($PM_{2.5}$)、饱和水汽压差 (VPD)、土壤湿度 (Ms)、风速 (Ws)、大气压 (AP)、土壤温度 (Ts)、总辐射 (Dr) 和紫外辐射 (UV), 而光合有效辐射 (PAR) 和净辐射 (Rn) 被排除在模型之外。

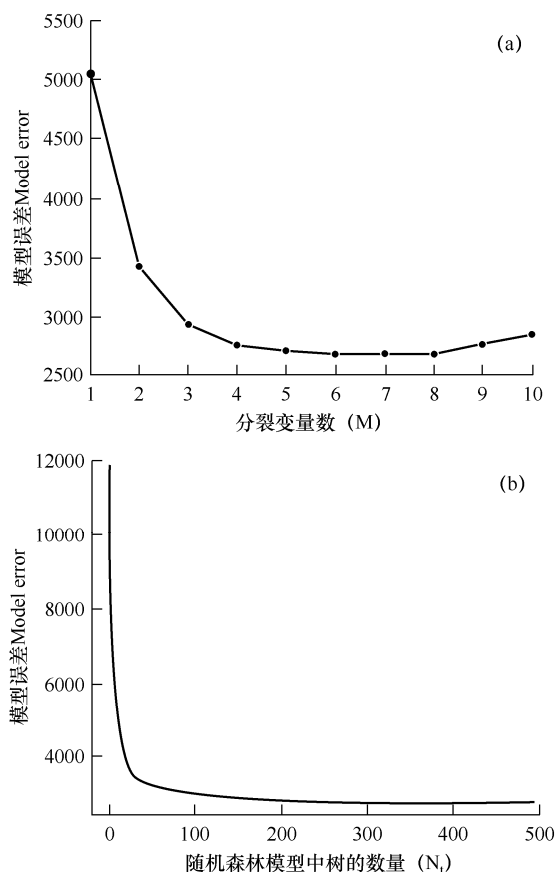


图 1 模型参数 M 和 N_t 对模型误差的影响

Fig. 1 Influence of number of splitting variables (M) and number of trees in random forest model (N_t) to the model error

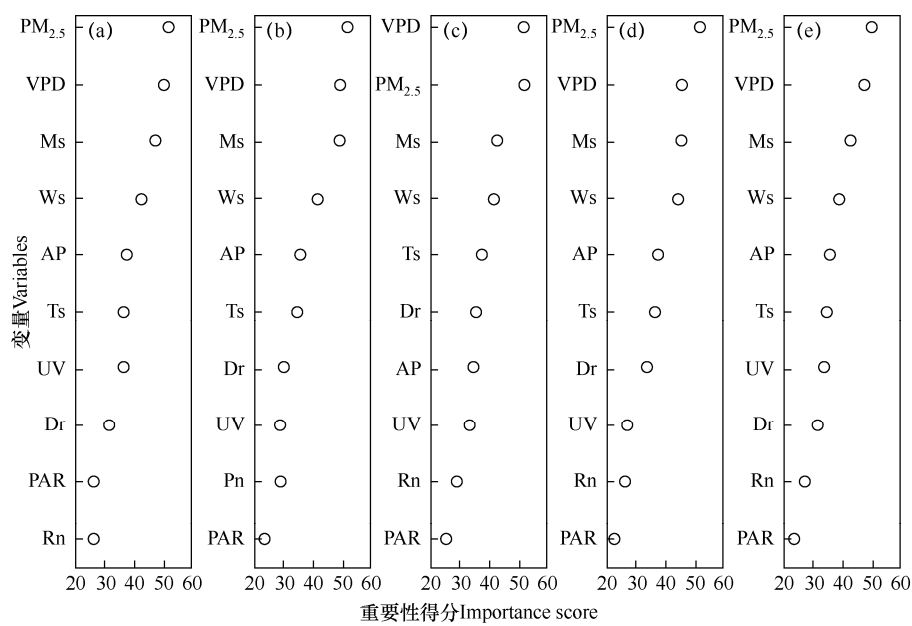


图 2 基于随机森林模型输出的 5 个训练样本集的环境因子重要性排序

Fig. 2 The variable importance of random forest model in descending order based on 5 training sample sets

表 2 随机森林模型的结果及相关值

Table 2 The results of random forest intermediate model and correlation values

样本编号 Sample No.	方差解释率 Variance explained(%)	相关系数 R
1	88.1	0.938
2	88.6	0.941
3	88.7	0.942
4	88.4	0.940
5	88.8	0.942

2.3 环境因子重要性综合排序

基于重要性得分较高的 8 个变量建立最终的随机森林模型，利用修正后的 5 个训练样本建立随机森林模型，并基于平均值的方法对预测结果进行综合评价，最终得到的随机森林模型由表 3 所示的 8 个变量建立，该模型的方差解释率为 88.9%。从随机森林模型可以看出，环境因子对 NAI 影响的重要性得分表现为 PM_{2.5} (48.037) > VPD (46.169) > Ms (43.984) > Ws (43.779) > UV (41.130) > Ts (40.107) > Dr (36.838) > AP (34.532)，PM_{2.5} 和 VPD 在影响 NAI 变化方面发挥了决定性作用，说明这两个因子是影响 NAI 变化的重要因素。

表 3 随机森林模型变量重要性得分

Table 3 The variable importance score of random forest model

变量名 Variables	重要性得分 Importance score
颗粒物 PM _{2.5}	48.037
饱和水汽压差 Vapor pressure deficit	46.169
土壤湿度 Soil moisture	43.984
风速 Wind speed	43.779
紫外辐射 Ultraviolet radiation	41.130
土壤温度 Soil temperature	40.107
总辐射 Direct radiation	36.838
大气压力 Atmospheric pressure	34.532

2.4 不同环境条件下随机森林模型模拟效果评价

利用独立样本数据对随机森林模型的拟合值和实测值进行拟合检验，结果见图 3。由图中可见，实测值与拟合值之间均方根误差 (RMSE) 为 59.349，线性相关系数 R² 为 0.887，表明模型具有较高的预测精度，显示出良好的拟合效果。为了进一步验证随机森林模型的稳定性，再次利用独立的测试数据对随机森林模型进行十次交叉验证，结果如表 4 所示，其 R² 平均值为 0.904，均方误差平均值为 24.851，说明随机森林模型模拟精度高，且具有很好的稳定性。

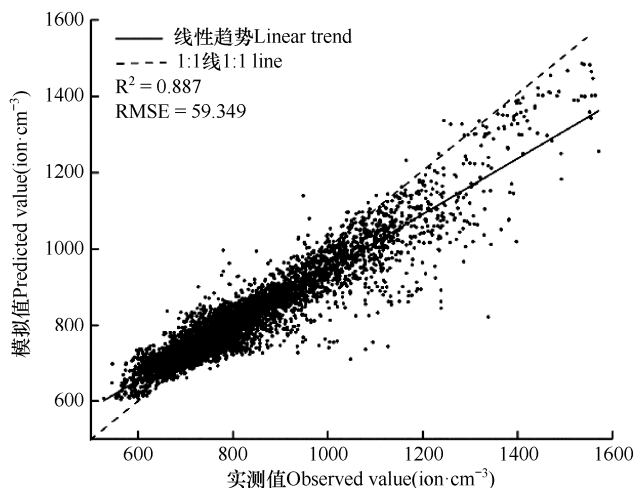


图 3 随机森林模型对空气负离子浓度的模拟精度评价

Fig. 3 Evaluation of simulation accuracy of negative air ions by random forest model

表 4 模型十折交叉验证结果

Table 4 Result of model 10-fold cross-validation

次数 Time	均方误差 RMSE	决定系数 R^2
1	44.663	0.868
2	28.135	0.887
3	23.362	0.906
4	22.476	0.909
5	20.627	0.917
6	21.037	0.915
7	21.507	0.913
8	21.388	0.914
9	22.603	0.908
10	22.713	0.908
平均值 Mean	24.851	0.904

3 讨论与结论

3.1 讨论

空气负离子受环境条件的影响是高度可变的，是多维的，而且变量之间存在高度复杂的非线性关系。加之测量过程中的不可控因素诸多，如山区供电与信号传输、仪器损耗与损坏、极端异常天气等，都可能导致数值丢失^[36]。因此，传统统计分析方法可能无法提供令人满意的结果，不足以揭示复杂过程中的模式和关系^[37-38]。因此，通过 R 语言应用随机森林的方法，研究 NAI 与环境因子之间的关系，从非线性角度全面分析确定影响 NAI 浓度变化的重要环境因子。模型中变量的筛选应具有较强的代表性，从而使模型更加简洁稳定而易于推广^[39-40]，本

研究中随机森林模型从原始的 10 个因子中选择了 8 个变量，其中重要性较高的 3 个变量分别为，颗粒物 ($PM_{2.5}$)、饱和水汽压差 (VPD) 和风速 (Ws)，表明它们对 NAI 的影响具有重要意义。

颗粒物 ($PM_{2.5}$) 是随机森林模型筛选出来的最重要变量，主要是因为自由离子不容易在空气中长期存在，容易附着在粒子上，通过颗粒污染物与负离子结合，形成大分子沉淀物，降低了环境中 NAI 浓度^[41-42]。相关文献^[43]表明 $PM_{2.5}$ 与空气离子呈负相关关系，空气离子参与了颗粒形成过程。例如，在研究城市热岛问题时有关学者发现^[44-45]，城市的空气污染程度越大，NAI 的浓度越低，同时 NAI 在城乡梯度上的差异主要是由 $PM_{2.5}$ 等大气污染物和空气温湿度等气候因素决定的，因此 NAI 是衡量一个地区空气清洁度的重要指标之一。

饱和水汽压差 (VPD) 是模型中筛选出来的重要变量，在随机森林中处于第二位，饱和水汽压差是空气温湿度的综合表现形式，由二者共同作用，其对 NAI 的影响不能忽略。温度的升高主要增加了分子运动速度和分子间碰撞的可能性，同时也提高了氧分子被电离的水平，从而有助于 NAI 的形成^[46]。空气湿度对 NAI 的影响主要是水分子存在会形成大量的 OH^- ，是 NAI 的重要组成成分，其大小取决于相对湿度^[22]。例如，当 NAI 在湿润空气中，其稳定状态的主要团簇离子为 $O_2^-(H_2O)_n$ 和 $CO_3^-(H_2O)_n$ 水合物，二者共同影响 NAI 浓度的高低^[47-48]；Reiter 等^[49]研究表明，空气湿度的增加可以改变颗粒碰撞和凝结的主要作用力，增强凝结效果，使小颗粒凝结并沉降成大颗粒，从而减少 NAI 的损失，维持 NAI 浓度；Bowers 等^[1]研究发现瀑布附近的负离子浓度增加主要与水汽有关，这些中间离子在很大程度上可以由一个 OH^- 和一个水团簇离子组成。

土壤湿度也是最重要的变量之一，在随机森林中处于第三位。土壤湿度的高低直接影响植被水分的吸收和传递，影响植被的生理状态，而森林植被是 NAI 产生的主要稳定来源^[50]，对 NAI 的影响具有重要意义。相关文献表明^[51]，植物能够通过气孔释放挥发性有机化合物，其中一部分在植物释放过程中电离，另一部分由于太阳辐射或其它辐射而电离，这个因素增加了正、负空气离子的浓度。此外植物释放挥发性有机化合物的速度还取决于植物体内的代谢过程，这在很大程度上取决于植物的种类、光

照以及土壤的湿度。本研究在植被主要生长季的6–9月开展试验，此时期为当地降水的主要季节，雨量充沛，土壤湿度不是影响植物正常生理活动的限制因子，因此土壤湿度对NAI的影响需要在具有明显干湿季的地区进一步加以分析。

风速是随机森林模型筛选出来的对NAI影响较大的第4位重要变量。首先，风速大小是决定空气洁净度的一个关键因素，而空气洁净度与NAI的存活密切相关；其次，从微观上分析，在NAI形成过程中，需要从分子态获得能量才能到达离子态，强风条件下分子获得能量的机会增加，分子态向离子态转变的过程会加快^[52]。相关研究表明^[13,53–54]，风速与NAI呈正相关，风速能加速林冠产生的NAI的扩散速度以及空气的流动和污染物的扩散，风速的增加有利于NAI的形成。例如，王薇等^[55]研究显示，NAI浓度与风速有较为密切的关系，其中最主要的影响因素是水，其次是风；然而，相关研究^[56]发现，在多雨和非多雨条件下，热带雨林地区NAI浓度与风速的关系不显著。这可能是由于研究区处的气候带不同，风速不是影响NAI的主要限制因子。

在随机森林模型中，土壤温度的重要性得分处于第六位，说明土壤温度对NAI的形成具有重要作用，这是以往研究未涉及或者忽略的，土壤温度对NAI的影响可能与植被根系的生理活动有关。土壤温度可以通过影响植被根系的呼吸作用以及其它生理活动，包括水分以及营养元素的吸收，从而影响整个植株的生理状态，促进或抑制植被气孔的开放，影响负离子的释放。此外，考虑到空气离子浓度与土壤中氢的析出直接相关，而影响氢析出的动力学的主要因素是湿度和土壤热梯度^[57]。因此，土壤温度的变化可能与氢的析出有关^[58]，从而影响负离子浓度。

模型中辐射对NAI的影响处于相对较低的位置，但其作用不能忽视。例如，光合有效辐射的增强有利于提升森林植被光合作用，植物可以通过光合作用向大气释放更多的氧气和负离子，而氧分子具有较强的吸收空气中自由电子的能力，因此空气中负离子的浓度随着光合有效辐射的增加而升高^[59]。相关研究结果表明，NAI浓度与光合有效辐射强度呈正比^[60]。例如，徐兆辉等^[61]在分析天堂寨风景区NAI浓度日观测资料的基础上，选取晴天和阴天天气类型的NAI浓度及相应的气象因子进行单

因素相关分析，结果表明，NAI浓度与光合有效辐射量呈极显著的线性正相关。

模型中大气压的重要性最低，但大气压对NAI的影响是存在的，这在以往的研究中也有涉及。例如，冯建涛等^[62]研究选取大气压作为与NAI相关的气象因子，建立了模型方程，用来预测空气负氧离子浓度。

综上所述，森林作为NAI形成的主要稳定来源，其协同作用机制较为复杂，不是简单的线性关系。基于非线性关系的随机森林能较好地预测环境因素对NAI的影响^[59]，这是以往研究常忽略的。当使用随机森林模型对环境因子的重要性进行排序时，由于在回归过程中不可能超出训练集的数据范围进行预测，也不可能引入所有指标^[63]，这可能是本研究使用模型的限制。本研究结果与以往研究有所不同，主要是环境条件对NAI影响的是复杂多变的，研究必须明确地区环境和社会条件是否一致，排除人为因素的干扰。引入机器学习^[64]中随机森林的分析方法，可以检验变量之间的非线性关系，确定了随机森林方法在NAI建模中的潜力。通过对模型变量进行相对重要性排序，从而强调了它们在解释环境因子对NAI影响关键因子的重要性，为更好地了解NAI对环境因子的响应机制提供重要的理论基础。模型应用表明，随机森林模型显示出较高的预测精度，反映了非线性趋势的存在。未来研究可以考虑不同时空尺度下预测因子的变化，以评估自变量在不同气候区的解释力；可以应用于其它不同类型的气候区，以了解气候区差异对所选变量重要性水平和模型整体预测能力；还可以利用支持向量机、神经网络和结构方程等深层次的数据挖掘方法，为研究森林以及城市中NAI的影响因素提供新的手段。

3.2 结论

本研究基于机器学习中随机森林模型研究环境因子对空气负离子影响的关键因子，结果表明，随机森林模型从原始的10个因子中选取8个因子作为自变量，利用所选变量建立的随机森林模型具有较高的预测精度，在拟合NAI变化方面具有较好的适用性，显示出较好的拟合效果，决定系数 R^2 达到了0.887，均方根误差（RMSE）为59.349。对模型重要性贡献相对较高的3个变量分别为颗粒物（ $PM_{2.5}$ ）、饱和水汽压差（VPD）和风速（Ws），表明了它们对

空气负离子的影响起决定性作用。颗粒物 ($PM_{2.5}$) 主要通过吸附 NAI 至表面, 形成大分子沉淀物, 从而降低了环境中的 NAI; 饱和水汽压差 (VPD) 是空气温湿度的综合表现, 通过空气温湿度升高增强水分子间运动速度和分子间碰撞的可能, 促进了分子向离子态转化的过程; 风速 (W_s) 可以加快林冠产生的 NAI 的扩散速度以及空气的流动和污染物的扩散, 从而有利于 NAI 的形成。

参考文献 References

- [1] Bowers B, Flory R, Ametepe J, et al. Controlled trial evaluation of exposure duration to negative air ions for the treatment of seasonal affective disorder[J]. *Psychiatry Research*, 2017, 259: 7-14.
- [2] 邵海荣, 贺庆棠. 森林与空气负离子[J]. *世界林业研究*, 2000(5): 20-24.
Shao H R, He Q T. Forest and air anion[J]. *World Forestry Research*, 2000(5): 20-24. (in Chinese)
- [3] 吴楚材, 郑群明, 钟林生. 森林游憩区空气负离子水平的研究[J]. *林业科学*, 2001, 37(5): 75-81.
Wu C C, Zheng Q M, Zhong L S. A study of the aero-anion concentration in forest recreation area[J]. *Scientia Silvae Sinicae*, 2001, 37(5): 75-81. (in Chinese)
- [4] Jovanić B R, Jovanić S B. The effect of high concentration of negative ions in the air on the chlorophyll content in plant leaves[J]. *Water Air and Soil Pollution*, 2001, 129(1-4): 259-265.
- [5] 王薇, 余庄. 中国城市环境中空气负离子研究进展[J]. *生态环境学报*, 2013, 22(4): 705-711.
Wang W, Yu Z. Research progress on negative air ions in urban environment in China[J]. *Ecology and Environmental Sciences*, 2013, 22(4): 705-711. (in Chinese)
- [6] Virkkula A, Hirsikko A, Vana M, et al. Charged particle size distributions and analysis of particle formation events at the Finnish Antarctic research station Aboa[J]. *Boreal Environment Research*, 2007, 12(3): 397-408.
- [7] Miao S, Zhang X, Han Y, et al. Random forest algorithm for the relationship between negative air ions and environmental factors in an urban park[J]. *Atmosphere*, 2018(9): 463-476.
- [8] Ling X, Jayaratne R, Morawska L. Air ion concentrations in various urban outdoor environments[J]. *Atmospheric Environment*, 2010, 44(18): 2186-2193.
- [9] Tikhonov V P, Tsvetkov V D, Litvinova E G, et al. Generation of negative air ions by plants upon pulsed electrical stimulation applied to soil[J]. *Russian Journal of Plant Physiology*, 2004, 51(3): 414-419.
- [10] Zhang J, Yu Z. Experimental and simulative analysis of relationship between ultraviolet irradiations and concentration of negative air ions in small chambers[J]. *Journal of Aerosol Science*, 2006, 37(10): 1347-1355.
- [11] Luts A, Parts T. Evolution of negative small air ions at two different temperatures[J]. *Journal of Atmospheric and Solar Terrestrial Physics*, 2002, 64(7): 763-774.
- [12] 单晟烨, 齐超, 张冬有. 漠河针叶林空气负离子浓度日变化特征研究[J]. *中国农学通报*, 2015, 31(25): 13-18.
Shan S Y, Qi C, Zhang D Y. Daily variation of negative air ions concentration in the coniferous forest in Mohe county[J]. *Chinese Agricultural Science Bulletin*, 2015, 31(25): 13-18. (in Chinese)
- [13] 齐冰, 邵碧嘉. 杭州市空气负离子变化特征分析[J]. *气象与减灾研究*, 2011, 34(4): 68-71.
Qi B, Shao B J. Characteristics of anion variation in Hangzhou[J]. *Meteorology and Disaster Reduction Research*, 2011, 34(4): 68-71. (in Chinese)
- [14] Shao H R, He Q T, Yan H P, et al. Spatio-temporal changes of negative air ion concentrations in Beijing[J]. *Journal of Beijing Forestry University*, 2005, 27(3): 35-39.
- [15] 柏智勇, 吴楚材. 空气负离子与植物精气相互作用的初步研究[J]. *中国城市林业*, 2008, 6(1): 58-60.
Bai Z Y, Wu C C. A preliminary study on interaction of negative air ion with plant aromatic substance[J]. *Journal of Chinese Urban Forestry*, 2008, 6(1): 58-60. (in Chinese)
- [16] Liang H, Chen X, Yin J, et al. The spatial-temporal pattern and influencing factors of negative air ions in urban forests, Shanghai, China[J]. *Journal of Forestry Research*, 2014, 25(4): 847-856.
- [17] Iwama H. Negative air ions created by water shearing improve erythrocyte deformability and aerobic metabolism[J]. *Indoor Air*, 2004, (14): 293-297.
- [18] 陈欢, 章家恩. 空气负离子浓度分布的影响因素研究综述[J]. *生态科学*, 2010, 29(2): 181-185.
Chen H, Zhang J E. Review on factors influencing the concentration distribution of negative air ions[J]. *Ecological*

- Science,2010,29(2):181-185.(in Chinese)
- [19] 季玉凯.棋盘山风景区空气负离子分布与变化规律的研究[D].沈阳:沈阳农业大学,2007:1-46.
- Ji Y K.Study on the concentration of negative air ions in Qipanshan scenic spot[D].Shenyang:Shenyang Agricultural University,2007:1-46.(in Chinese)
- [20] 张双全,谭益民,吴章文.空气负离子浓度与空气温湿度的关系研究[J].中南林业科技大学学报,2011,31(4):114-118.
- Zhang S Q,Tan Y M,Wu Z W.The relationship between air anion concentration and air temperature and air relative humidity[J].Journal of Central South University Of Forestry And Technology,2011,31(4):114-118.(in Chinese)
- [21] Goldstein N I,Goldstein R N,Merzlyak M N.Negative air ions as a source of superoxide[J].International Journal of Biometeorology,1992,36(2):118-122.
- [22] Smith W L.Note on the relationship between precipitable water and surface dew point[J].Journal of Applied Meteorology,1966,(5):726-727.
- [23] Fujioka N,Tsunoda Y,Sugimura A,et al.Influence of humidity on variation of ion mobility with life time in atmospheric air[J].IEEE Trans Power Syst,1983,102(4):911-917.
- [24] 黄世成,徐春阳,周嘉陵.城市和森林空气负离子浓度与气象环境关系的通径分析[J].气象,2012,(11):1417-1422.
- Huang S C,Xu C Y,Zhou J L.Path analysis on negative air ion concentration and the meteorological environment in urban and forest zones[J].Meteorological Monthly,2012,(11):1417-1422.(in Chinese)
- [25] 吴甫成,姚成胜.岳麓山空气负离子及空气质量变化研究[J].环境科学学报,2006,26(10):1737-1744.
- Wu F C,Yao C S.Distribution of negative air ions and its relation to air quality of the Yuelu Mountain[J].Acta Scientiae Circumstantiae,2006,26(10):1737-1744.(in Chinese)
- [26] Lin H F,Lin J M.Generation and determination of negative air ions[J].China Pulp & Paper Industry,2017,1(1):1-6.
- [27] 李爱博,周本智,李春友,等.基于控制实验的6个典型亚热带树种空气负离子效应[J].林业科学研究,2019,(4):120-128.
- Li A B,Zhou B Z,Li C Y,et al.Negative air ion effect of six typical subtropical tree species based on control experiment[J].Forestry Research,2019(4):120-128.(in Chinese)
- [28] Cutler D R,Edwards T C,Beard K H,et al.Random forests for classification in Ecology[J].Ecology,2007, 88(11):2783-2792.
- [29] Lam K L,Cheng W Y,Su Y,et al.Use of random forest analysis to quantify the importance of the structural characteristics of beta-glucans for prebiotic development[J].Food Hydrocolloids,2020,108:1-13.
- [30] Lozano F J,Suárez-Seoane S,Kelly M,et al.A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees:a case study in a mountainous Mediterranean region[J].Remote Sensing of Environment,2008,112(3):708-719.
- [31] Amatulli G,Rodrigues M J,Trombetti M,et al.Assessing long-term fire risk at local scale by means of decision tree technique[J].Journal of Geophysical Research Biogeosciences, 2006,111(4):15.
- [32] 秦鹏程,刘志雄,万素琴.基于决策树和随机森林模型的湖北油菜产量限制因子分析[J].中国农业气象,2016,37(6):691-699.
- Qing P C,Liu Z X,Wan S Q.Identification of yield limiting factors for oilseed rape in Hubei province based on CART and random forest model[J].Chinese Journal of Agrometeorology, 2016,37(6):691-699.(in Chinese)
- [33] Archer K J,Kimes R V.Empirical characterization of random forest variable importance measures[J].Computational Statistics and Data Analysis,2008, 52(4):2249-2260.
- [34] Delgado B M,Maestre F T,Reich P B,et al. Microbial diversity drives multifunctionality in terrestrial ecosystems[J].Nature Communications,2016(7):1-8.
- [35] Genuer R,Poggi J M,Tuleau-Malot C.Variable selection using random forests[J].Pattern Recognition Letters,2010, 31(14):2225-2236.
- [36] Prasad A M,Iverson L R,Liaw A.Newer classification and regression tree techniques:bagging and random forests for ecological prediction[J].Ecosystems,2006,9(2):181-199.
- [37] Peters J,Baets B D,Verhoest N E C,et al.Random forests as a tool for ecohydrological distribution modelling[J].Ecological Modelling,2007,207:304-318.
- [38] De A G,Fabricius K E.Classification and regression trees:a powerful yet simple technique for ecological data analysis[J].Ecology,2000,81:3178-3192.
- [39] Miao S,Zhang X,Han Y,et al.Random forest algorithm for

- the relationship between negative air ions and environmental factors in an urban park[J]. *Atmosphere*, 2018,9:463-476.
- [40] Vilar L, Woolford D G, Martell D L, et al. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain[J]. *International Journal of Wildland Fire*, 2010, 19(3): 325-337.
- [41] Tammet H, Hrrak U, Laakso L, et al. Factors of air ion balance in a coniferous forest according to measurements in Hyytiälä, Finland[J]. *Atmospheric Chemistry and Physics*, 2006, 6(11): 3377-3390.
- [42] Vana M, Ehn M, Petäejae T, et al. Characteristic features of air ions at mace head on the west coast of Ireland[J]. *Atmospheric Research*, 2008, 90(2): 278-286.
- [43] Kolarž P, Filipović D. Measurements and correlations between several atmospheric parameters[J]. *Facta universitatis-series: Physics, Chemistry and Technology*, 2008, 6(1): 99-104.
- [44] Zhang K, Wang R, Shen C, et al. Temporal and spatial characteristics of the urban heat island during rapid urbanization in Shanghai, China[J]. *Environmental Monitoring and Assessment*, 2010, 169: 101-112.
- [45] Zhao S, Da L, Tang Z, et al. Ecological consequences of rapid urban expansion: Shanghai, China[J]. *Frontiers in Ecology and the Environment*, 2006(4): 341-346.
- [46] 叶彩华, 王晓云. 空气中负离子浓度与气象条件关系初探[J]. *气象科技*, 2000, 28: 51-52.
- Ye C H, Wang X Y. Preliminary study on the relationship between the concentration of negative ions in the air and meteorological conditions [J]. *Meteorological Science and Technology*, 2000, 28: 51-52. (in Chinese)
- [47] Reiter R. Frequency distribution of positive and negative small ion concentrations, based on many years recordings at two mountain stations located at 740 and 1780m ASL[J]. *International Journal of Biometeorology*, 1985, 29(3): 223-231.
- [48] Luts A. Evolution of negative small ions at enhanced ionization[J]. *Journal of Geophysical Research Atmospheres*, 1995, 100(1): 1487-1496.
- [49] Skalny J D, Mikoviny T, Matejcek S. An analysis of mass spectrometric study of negative ions extracted from negative corona discharge in air[J]. *International Journal of Mass Spectrometry*, 2004, 233: 317-324.
- [50] Wang J, Li S H. Changes in negative air ions concentration under different light intensities and development of a model to relate light intensity to directional change[J]. *Journal of Environmental Management*, 2009, 90(8): 2746-2754.
- [51] Sinicina N, Skromulis A, Martinovs A. Impact of microclimate and indoor plants on air ion concentration[J]. *Reformation and Strategy*, 2015, 1(1): 66-73.
- [52] Liu X, Guanter L, Liu L. Downscaling of solar-induced chlorophyll fluorescence from canopy level to photosystem level using a random forest model[J]. *Remote Sensing of Environment*, 2019, 231: 110772.
- [53] Zhang Z P, Yu Y W, Meng M H, et al. Progresses of aeroanion in tourism environments[J]. *Journal of Zhejiang Forestry College*, 2006, 28(11): 1417-1422.
- [54] Kalivitis N. Night-time enhanced atmospheric ion concentrations in the MBL[J]. *Atmospheric Chemistry and Physics*, 2012, (12): 3627-3638.
- [55] 王薇, 余庄, 郑非艺. 不同环境场所夏季空气负离子浓度分布特征及其与环境因子的关系[J]. *城市环境与城市生态*, 2012(2): 38-40.
- Wang W, Yu Z, Zheng F Y. Characteristics of negative air ion concentration and its relationships with environmental factors in various environments in summer[J]. *Urban Environment and Urban Ecology*, 2012(2): 38-40. (in Chinese)
- [56] Si T, Luo Y, Zhao Z. Relationship between negative air ion concentration and meteorological elements in Diaoluoshan tropical rainforest[J]. *Resources Science*, 2014, 36(4): 788-792.
- [57] Duenas C, Fernandez M C, Carretero J, et al. Release of ^{222}Rn from some soils[J]. *Ann Geophysicae*, 1997, 15: 124-133.
- [58] Kolar P, Vaupoti J, Kobal I, et al. Thoron, radon and air ions spatial distribution in indoor air[J]. *Journal of Environmental Radioactivity*, 2017, 173: 70-74.
- [59] Deng L. Review on research of the negative air ion concentration distribution and its correlation with meteorological elements in mountain tourist area[J]. *Earth Sciences*, 2019, 8(1): 60-68.
- [60] Laakso L, Hussein T, Aarnio P, et al. Diurnal and annual characteristics of particle mass and number concentrations in urban, rural and Arctic environments in Finland[J].

- Atmospheric Environment,2003,37(19):2629-2641.
- [61] 徐昭晖.安徽省主要森林旅游区空气负离子资源研究[D].合肥:安徽农业大学,2004:1-45.
- Xu Z H.Study on air anion resources in main forest tourism areas of Anhui province[D].Hefei:Anhui Agricultural University,2004:1-45.(in Chinese)
- [62] 冯建涛.城市带状绿地空气负离子分布规律研究[D].吉林:北华大学,2017:1-69.
- Feng J T.Study on the distribution of negative air ions in urban belt green space[D].Jilin:Beihua University,2017:1-69.(in Chinese)
- [63] 陈纪波,胡慧,陈克垚,等.基于非线性 PLSR 模型的气候变化对粮食产量的影响分析[J].中国农业气象,2016,37(6):674-681.
- Chen J B,Hu H,Chen K Y,et al.Effects of climate change on the grain yield based on nonlinear PLSR model[J].Chinese Journal of Agrometeorology,2016,37(6):674-681.(in Chinese)
- [64] 冯禹,崔宁博,龚道枝.机器学习算法和 Hargreaves 模型在四川盆地 ET_0 计算中的比较[J].中国农业气象,2016,37(4):415-421.
- Feng Y,Cui N B,Gong D Z.Comparison of machine learning algorithms and Hargreaves model for reference evapotranspiration estimation in Sichuan basin[J].Chinese Journal of Agrometeorology,2016,37(4):415-421.(in Chinese)