

基于随机森林算法对贵阳龙洞堡机场能见度的可预报性研究

邓小光

中国民用航空西南地区空中交通管理局贵州分局, 贵州 贵阳

收稿日期: 2023年3月31日; 录用日期: 2023年5月4日; 发布日期: 2023年5月19日

摘要

利用2017年贵阳龙洞堡国际机场常规观测逐小时数据以及同期贵阳市新华路站点(1446A)环境污染物逐小时数据, 研究在气温、湿度、风速等常规气象要素与环境污染物共同作用下基于随机森林算法对低能见度的变化进行预测研究。研究结果表明: 随机森林模型预测值序列与真实值序列相关系数较高, 表明随机森林算法在能见度变化趋势上预测效果较好。从随机森林算法输出的因素重要性发现环境污染物的贡献较为重要, 进一步研究了各环境污染物的日变化特征和月变化特征。利用HYSPLIT模式确定了机场近地面气团的来源, 这使得机场气象要素以及环境污染物的来源地得以确定。

关键词

随机森林算法, 低能见度, 聚类分析

Research on Visibility Predictability of Longdongbao Airport in Guiyang Based on Random Forest Algorithm

Xiaoguang Deng

Guizhou Branch of Southwest Air Traffic Administration of Civil Aviation of China, Guiyang Guizhou

Received: Mar. 31st, 2023; accepted: May 4th, 2023; published: May 19th, 2023

Abstract

Based on the hourly data of routine observation of Guiyang Longdongbao International Airport in 2017 and the hourly data of environmental pollutants at Xinhua Road station (1446A) in Guiyang

during the same period, the random forest algorithm was used to predict the change of low visibility under the joint action of routine meteorological elements such as temperature, humidity, wind speed and environmental pollutants. The results show that the correlation coefficient between the predicted value sequence and the real value sequence of the stochastic forest model is high, which indicates that the stochastic forest algorithm is effective in predicting the change trend of visibility. According to the importance of factors output by random forest algorithm, it is found that the contribution of environmental pollutants is more important, and the daily and monthly variation characteristics of each environmental pollutant are further studied. The HYSPLIT model is used to determine the source of the air mass near the surface of the airport, which makes the meteorological elements of the airport and the source of environmental pollutants can be determined.

Keywords

Random Forest Algorithm, Low Visibility, Cluster Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

能见度是表征大气透明状况的物理量。在民航飞行中，飞机起飞和着陆对能见度的依赖性极高，只有足够的能见度才能让飞机安全降落。低能见度天气是造成民航事故、影响飞行不正常的主要因素之一。低能见度天气一直是航空气象科研工作和服务的重点[1]。

近年来，国内外学者已经做了不少关于能见度的研究，谢超等[2]使用神经网络模型研究了华南低能见度天气。王楠等[3]使用 EC 数据建立了 SVM 低能见度预测模型，发现模型对于特定天气形势下引发的低能见度天气预报误差较小且预报提前量较大。司林青[4]基于 BP 神经网络方法分类训练了 3 个统计模型并与 WRF 天气模式产品对接，采用分类筛选法研发了龙洞堡机场 24 h 时效的逐时能见度预报产品，结果表明研究效果较好。

从低能见度产生条件可知，气溶胶粒子对低能见度的产生和演变规律也是不可忽略的。首先，气溶胶粒子作为凝结核对大气中的水汽凝聚起着重要作用；其次，不同气溶胶对太阳光的散射和吸收作用会使得环境温度变化，从而间接影响能见度的变化。气溶胶种类多样，可分为矿物气溶胶、碳气溶胶(黑碳和有机碳)、硫酸盐、硝酸盐、铵盐、海盐和生物气溶胶等，不同化学组分下的气溶胶对天气的影响有所不同[5]。而上述气溶胶的直接监测数据缺乏，但可以通过追踪气溶胶前体物 NO₂、CO、SO₂ 以及 O₃、PM_{2.5}、PM₁₀ 等环境污染物的演变来表征空气中气溶胶特性及含量的变化。因此，在本次能见度变化研究中加入了环境污染物的贡献，这对研究效果可能会有一定的提升。

随机森林算法由 Breiman Leo [6]和 Adele Cutler 等[7]提出，该算法结合了 Breiman 的“Bootstrap Aggregating” (自举汇聚法)思想和 Ho [8]的“Random Subspace” (随机子空间)方法。随机森林算法是基于决策树的集成学习算法[9]，决策树是一种广泛应用的树状分类器，在树的节点上，通过选择最优的特征不断对目标进行分类，当达到分类最优时将会停止对树进行分类，并得到分类结果。但是单个决策树对目标预测性能有限，为了改善单个决策树的预测性能，将多个决策树集合起来，通过对每个决策树的分类结果进行组合，就形成了多个决策树组成的随机森林。随机森林的建立能够提升模型的预测精度和泛化能力，避免出现过拟合现象。余星源等[10]使用随机森林算法建立低能见度天气等级判断模型，对比发

现随机森林在准确度和计算效率方面较其他模型均有一定优势。付旭东等[11]通过构建随机森林 1~6 h 风场预报模型,表明随机森林算法在风场预报中有较好的泛化能力,对地面 10 米风场有较好的预报水平。朱国栋等[12]结合不同数量的决策树进行模型训练后,研究建立基于随机森林方法的乌鲁木齐机场逐时温度回归预报模型,预测效果较好。徐敏辉等[13]通过建立随机森林回归模型,对低 RVR 的大雾过程进行拟合,发现背景光亮度、相对湿度、低云高、温度对 RVR 的影响较大,随机森林回归模型的测试机相关系数大 0.886,说明其具备较好的学习能力和较优的拟合效果。

贵阳龙洞堡机场位于中国贵州省贵阳市东郊,坐标为(106.7988°E, 26.5438°N),海拔高度 1139 米,地处云贵高原东麓,其复杂的地形条件导致能见度预报变得尤为困难。本文利用机器学习随机森林算法对贵阳机场气象要素变化和环境污染变化共同作用下的能见度变化特征进行预测研究,并利用 HYSPLIT 后向轨迹模式计算贵阳机场近地面气团的来源问题,这一工作为贵阳龙洞堡机场能见度的预报、预警工作提供一定参考和指导。

2. 数据与方法

2.1. 数据

贵阳龙洞堡国际机场 2017 年 1 月至 12 月地面常规观测逐小时数据,包括风速(Sp)、温度(T)、露点(Td)、相对湿度(RH)、场面气压(QFE)等。

贵阳龙洞堡机场常规观测数据中缺乏环境污染物数据,选用距离贵阳机场最近的新华路站点(1446A)数据代表贵阳机场环境污染物,本次研究选取 2017 年 1 月至 12 月近地面环境污染物逐小时数据,变量包括 PM10、PM2.5、SO₂、NO₂、CO、O₃等。环境污染物每小时数据可在如下网址查看:

<https://www.zq12369.com/?city=%E8%B4%B5%E9%98%B3&tab=city>。

HYSPLIT 模式由美国国家海洋和大气管理局(NOAA)的空气资源实验室以及澳大利亚气象局联合研发的一种用于计算和分析大气污染物输送、扩散轨迹的专业模型。模式模拟主要有后向传输模型和前向扩散模型,本文中使用了其后向传输模型,对贵阳龙洞堡机场近地面层上气团的来源进行确定。拉格朗日后向轨迹模式(HYSPLIT)采用的气象资料来源于美国国家环境预报中心(National Centers for Environmental Prediction, NCEP)提供的同时段 GDAS (global data assimilation system)数据,该数据是利用全球资料同化系统,把全球 1° × 1°大气状态数据插值到正形投影的地图上,时间分辨率为 6 h (记录时间分别为 00:00, 06:00, 12:00 和 18:00 (UTC)),垂直方向分为 23 层。其中气象要素场包括水平和垂直风速、温度、气压、相对湿度、降水等。本文选取位于贵阳龙洞堡机场所在点位(26.54°N, 106.80°E)为受点位置。

本文所涉及的时间均为北京时。

2.2. 方法

2.2.1. 随机森林算法

随机森林算法是一种非线性统计集成算法,由多科分类与回归决策树组成,是一种基于决策树算法改进的高级算法,它的本质属于机器学习中的集成学习。该算法采用随机的方式建立一个森林,里面有很多的决策树,每一棵决策树之间是没有关联的。当有一个新的输入样本进入的时候,就让森林中的每一棵决策树分别进行预测,得到该样本的类别预测结果。最终,森林中所有树的预测结果是哪一类的最多,就作为最终预测结果。

本文在构建随机森林模型时,首先从观测数据集中能见度低于 3000 的样本中随机抽取 270 个样本生成训练集,在此基础上生成决策树,多个决策树共同构成随机森林。在模型中,假设共有 M 个训练指标,系统随机选取 m 个(m ≤ M)作为节点指标,选取基尼最小值当作分支标准,依照决策树的预测结果,以投

票方式决定新样本的类别[14]。每次抽样未被选中的数据构成了袋外数据(OOB)，利用袋外数据估计内部误差，称为袋外误差(E_{OOB})，公式如下：

$$E_{OOB} = \frac{1}{n} \sum_{i=1}^n [\hat{Y}(X_i) - Y_i]^2 \quad (1)$$

式(1)中： n 为OOB的样本个数； $\hat{Y}(X_i)$ 为根据给定样本 X_i 基于模型的输出数据； Y_i 为观测数据。

此外，该模型通过对算法中OOB误差的估计评价特性变量的重要性。先计算每个决策树的袋外误差，然后在训练指标的数据随机中加入噪声并计算袋外误差，预测指标的重要性公式如下：

$$V(i) = \frac{1}{N} \sum (E_{OOB2} - E_{OOB1}) \quad (2)$$

式(2)中改变指标 i 造成的袋外误差 E_{OOB2} 越大，表明变量 i 越重要[15]。

为了检验随机森林模型的回归预测效果是否可行，将贵阳机场的观测能见度数据作为真实能见度数据，选取皮尔逊相关系数 R 作为检验指标， R 值越高，表示模型预测能见度变化效果越好。

随机森林算法是一种很灵活实用的方法，它有如下优点：

- 1) 在当前所有算法中，具有较好的准确率；
- 2) 能够有效运行在大数据集上，训练速度快，容易做成并行方法；
- 3) 能够处理具有高维特征的输入样本，并且不需要降维；
- 4) 能够评估各个特征在分类问题上的重要性；
- 5) 在生成过程中，能够获取到内部生成误差的一种无偏估计；
- 6) 对于缺省值问题也能够获得很好的结果；
- 7) 既能处理离散型数据，也能处理连续型数据，数据集无需标准化。

2.2.2. HYSPLIT 模式

HYSPLIT 模型属于 Eulerian-Lagrangian 混合型的扩散模式，其平流和扩散计算采用了 Lagrangian 法。该模式采用地形 σ 坐标，气象数据在水平坐标上保持其原来格式，而垂直方向内插到地形 σ 坐标系统：

$$\sigma = \frac{Z_{top} - Z_{mst}}{Z_{top} - Z_{gl}} \quad (3)$$

式(3)中： Z_{top} 为轨迹模式坐标的顶部， Z_{gl} 为地形高度， Z_{mst} 为坐标下边界高度。计算气团所携带粒子运动轨迹时，其最终位置由初始位置(Q)和后推位置(Q')的平均速率计算得到：

$$\begin{aligned} Q'(t + \Delta t) &= Q(t) + V(Q, t) \Delta t \\ Q(t + \Delta t) &= Q(t) + 0.5 \times [V(Q, t) + V(Q', t + \Delta t)] \Delta t \end{aligned} \quad (4)$$

式(4)中： Δt 为后推的时间步长，本研究中 Δt 取12小时。

2.2.3. 聚类分析

利用 HYSPLIT 模式对 2017 年 1 月至 12 月贵阳龙洞堡机场近地面气团 12 小时后向轨迹进行大量计算，得到共 1336 个后向轨迹结果，采用聚类分析方法(略)对上述后向轨迹结果进行分类得到不同类别并进行分析。

3. 能见度变化的客观预报

考虑贵阳机场能见度低于 3000 米后会对航班存在一定的影响，因此，针对 2017 年 1 月至 12 月的数据，筛选出能见度小于 3000 米的小时能见度数据和同期的气象观测数据以及环境污染物数据，共得到

294 个样本数据, 每个样本数据包括风速、温度、露点、相对湿度、场面气压、SO₂、NO₂、CO、O₃、PM10、PM2.5 以及能见度数据。随机选择 270 个样本作为随机森林算法模型的训练集, 并对随机森林模型进行训练, 剩余的 24 个样本数据作为测试集。

研究中, 决策树数目设置为 100, 叶子树设为 5。图 1 为随机森林预测误差曲线图, 横坐标为决策树数目, 纵坐标为误差值, 决策树数目较少时, 计算误差较大, 最大误差可达 0.052, 波动也较大。随着决策树数目的增多, 误差迅速减小, 并趋于稳定, 决策树达到 100 时, 其误差分布在 0.035 左右。

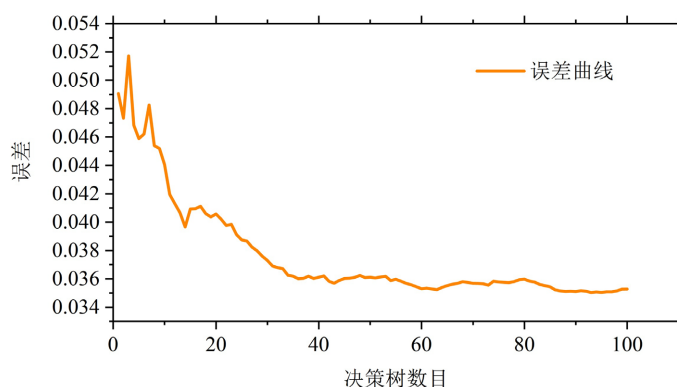


Figure 1. Random forest prediction error curve

图 1. 随机森林预测误差曲线图

经过 270 个样本数据进行训练后, 将剩余的 24 个样本数据作为测试集输入随机森林模型, 得到 24 个能见度的预测值, 将预测值与真实值进行对比, 绘制如图 2。通过分析预测值与真实值差异可知, 随机森林预测值序列与真实值序列相关系数达 0.80, 通过 99% 的显著性检验, 说明随机森林算法在能见度变化趋势上预测效果较好。从具体数值上分析, 预测值相比真实值较为保守, 例如第 5 个时次真实值为 300 米, 但预测值为 707 米; 第 20 个时次真实值为 400 米, 但预测值为 1002 米; 第 12 个时次真实值为 2800 米, 但预测值为 2058 米。说明在精度上还存在些许不足。但总体上可以认为随机森林算法对贵阳机场能见度的预测结果较为可信。

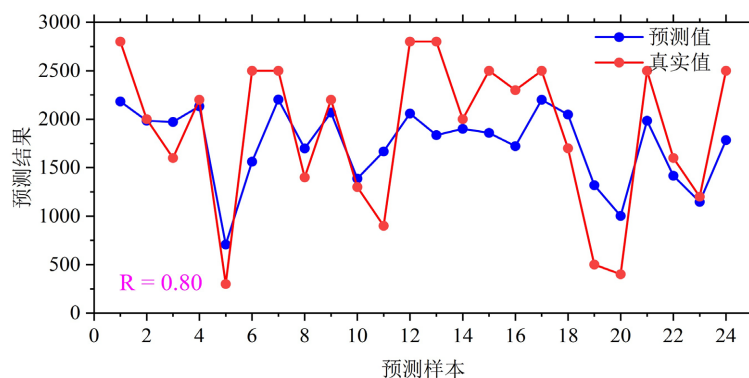


Figure 2. Prediction results of visibility change of Guiyang Airport by Random Forest algorithm. The abscissa represents 24 predicted sample values, and the ordinate represents the corresponding predicted results. The red line is the true value and the blue line is the predicted value

图 2. 随机森林算法对贵阳机场能见度变化的预测结果。横坐标表示 24 个预测样本值, 纵坐标表示相对应的预测结果。红线表示真实值, 蓝线表示预测值

通过随机森林算法输出的因素重要性数据绘图，见图 3。从因素重要性上分析可知，随机森林算法对贵阳机场能见度的预测算法中，所有因素对随机森林预测的准确度影响占比有所不同，其重要性从大到小可列为： $O_3 > QFE > T > Td > Sp > CO > RH > NO_2 > PM_{2.5} > PM_{10} > SO_2$ 。其中近地面臭氧重要性最高，系数可达约 1.4，而硫酸盐气溶胶前体物 SO_2 的重要性最低，仅约为 0.2。这也说明了随机森林算法预测的能见度数值变化对 SO_2 浓度敏感性较低，而对 O_3 浓度的变化敏感性较高。经研究发现，大气中 $PM_{2.5}$ 浓度显著下降时，会导致光辐射增强，有利于臭氧生成，因此臭氧的浓度变化在一定程度上可以反应大气中气溶胶的浓度变化。

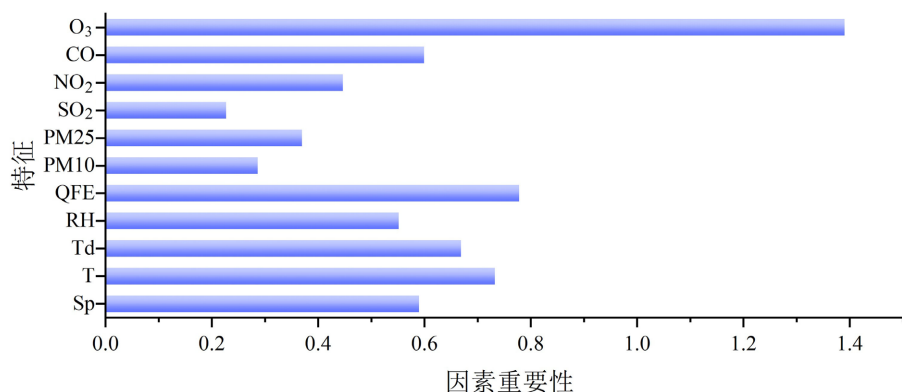


Figure 3. Importance of factors. Ordinate lists the influence characteristic values
图 3. 因素重要性，纵坐标列出了各影响特征值

总结上述研究可知，通过随机森林算法预测能见度值的研究发现，该算法通过引入机场风速、气温、水汽条件以及环境污染物 O_3 、 CO 、 NO_2 、 $PM_{2.5}$ 、 PM_{10} 以及 SO_2 等气象要素(见表 1)可以较好得对机场能见度变化特征进行预测，同时也发现上述环境污染物均对贵阳龙洞堡机场能见度的预测结果产生一定程度的影响。

Table 1. Variable factors affecting visibility
表 1. 影响能见度的变量因子

变量因子	单位	变量因子	单位
10 米风速(Sp)	m/s	PM ₁₀	$\mu\text{g}/\text{m}^3$
2 米气温(T)	$^{\circ}\text{C}$	PM _{2.5}	$\mu\text{g}/\text{m}^3$
2 m 露点温度(Td)	$^{\circ}\text{C}$	SO_2	$\mu\text{g}/\text{m}^3$
相对湿度(RH)	%	CO	mg/m^3
场面气压(QFE)	hPa	NO_2	$\mu\text{g}/\text{m}^3$
		O_3	$\mu\text{g}/\text{m}^3$

4. 近地面污染物浓度特征分析

研究贵阳机场环境污染物变化特征对预测能见度的变化研究存在一定意义。图 4 截取了贵阳市环境污染物站点分布图，新华路站点位于贵阳机场西北侧，其观测数据特征可近似代表贵阳机场环境特征。 O_3 、 NO_2 、 $PM_{2.5}$ 、 SO_2 、 PM_{10} 、 CO 等环境污染物的日变化特征见图 5。分析可知， O_3 在一天中的午后

至傍晚时分浓度最高，最高值达 $81 \mu\text{g}/\text{m}^3$ ，其余时间浓度较低，上午 8 点左右浓度最低，约为 $37 \mu\text{g}/\text{m}^3$ ； NO_2 在一天中的上午 6 点和下午 15 点出现浓度最低极值，最低约为 $21 \mu\text{g}/\text{m}^3$ ，夜间的浓度较高，约为 $37.6 \mu\text{g}/\text{m}^3$ ； $\text{PM}_{2.5}$ 粒子浓度在夜间分布较高，极大值为 $41.4 \mu\text{g}/\text{m}^3$ ，白天分布较低，极小值为 $30.5 \mu\text{g}/\text{m}^3$ ； SO_2 浓度分布特征与 NO_2 较为相似，夜间和中午的浓度较高，而上午和下午的浓度较低，极大值约为 $14.8 \mu\text{g}/\text{m}^3$ ，极小值约为 $9.4 \mu\text{g}/\text{m}^3$ ； PM_{10} 粒子浓度分布特征与 $\text{PM}_{2.5}$ 较为相似，极大值 $68 \mu\text{g}/\text{m}^3$ ，分布在 1 点，极小值 $50 \mu\text{g}/\text{m}^3$ ，分布在 15 点； CO 的浓度变化波动相对较小，5 点~7 点以及 14 点~17 点浓度较低，极小值约为 $0.62 \text{ mg}/\text{m}^3$ ，极大值 $0.77 \text{ mg}/\text{m}^3$ ，出现在 22 点。



Figure 4. Distribution of environmental pollutant sites in Guiyang
图 4. 贵阳市环境污染物站点分布

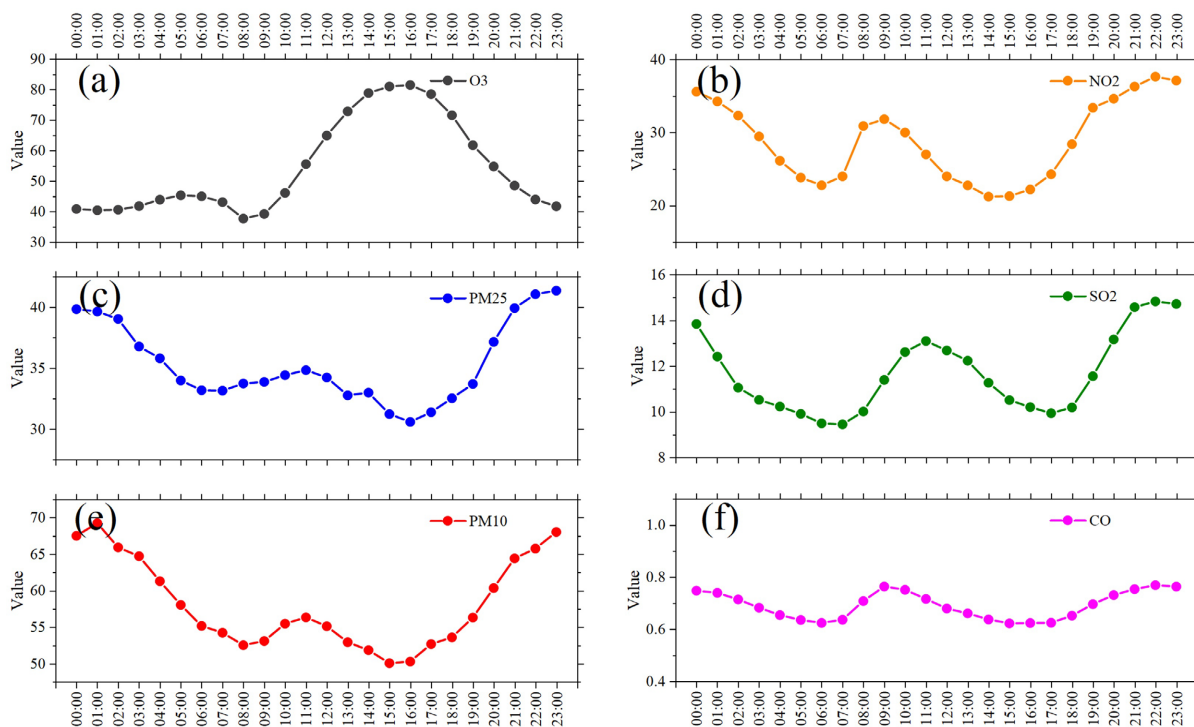


Figure 5. Diurnal variation characteristics of environmental pollutants, and the abscissa represents Beijing time. (a) O_3 ; (b) NO_2 ; (c) $\text{PM}_{2.5}$; (d) SO_2 ; (e) PM_{10} ; (f) CO

图 5. 环境污染物日变化特征，横坐标表示北京时间。(a) O_3 ; (b) NO_2 ; (c) $\text{PM}_{2.5}$; (d) SO_2 ; (e) PM_{10} ; (f) CO

O₃、NO₂、PM_{2.5}、SO₂、PM₁₀、CO 等环境污染物的月变化特征见图 6。分析可知，O₃ 浓度大值区出现在 4 月~5 月，极大值约为 86 μg/m³，浓度小值区出现在 11 月、12 月以及 1 月，极小值约为 40 μg/m³；NO₂ 浓度大值区出现在 3 月、11 月和 12 月，极大值约为 39.2 μg/m³，浓度小值区分布在 5 月~10 月，极小值约为 22.3 μg/m³；PM_{2.5} 粒子浓度大值区分布在 1~3 月以及 11~12 月，极大值约为 54.3 μg/m³，浓度小值区分布在 5 月~10 月，极小值约为 20.5 μg/m³；SO₂ 浓度大值区分布在 1~2 月和 11~12 月，极大值约为 27.3 μg/m³，浓度小值区分布在 3~10 月，极小值约为 5.4 μg/m³；PM₁₀ 粒子浓度大值区分布在 1~3 月以及 11~12 月，极大值约为 95.5 μg/m³，浓度小值区分布在 6~10 月，极小值约为 33.8 μg/m³；CO 浓度大值区分布在 1~3 月，极大值约为 1.05 mg/m³，浓度小值区分布在 4~10 月，极小值约为 0.52 mg/m³。

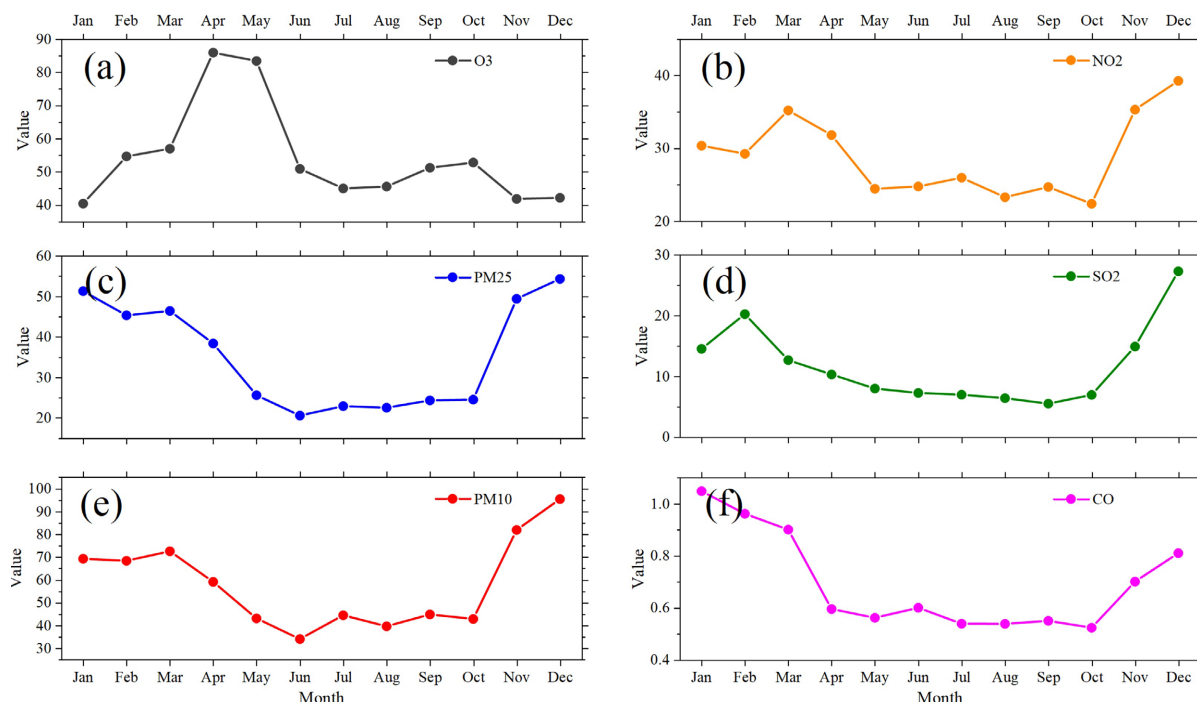


Figure 6. Monthly variation characteristics of environmental pollutants, with the abscissa representing January to December. (a) O₃; (b) NO₂; (c) PM_{2.5}; (d) SO₂; (e) PM₁₀; (f) CO

图 6. 环境污染物月变化特征，横坐标表示 1~12 月。(a) O₃；(b) NO₂；(c) PM_{2.5}；(d) SO₂；(e) PM₁₀；(f) CO

5. 贵阳机场近地面气团来源分析

利用 HYSPLIT 后向轨迹模式对贵阳机场近地面气团进行来源确定，高度设置为 50 米，计算 12 小时后向轨迹，每 6 小时计算一次后向轨迹，得到 1336 条轨迹，利用聚类分析方法将轨迹集合按照来向最佳条件可分为 6 类。图 7 即为 2017 年贵阳机场 HYSPLIT 后向轨迹聚类分析结果，由图可知贵阳机场近地面的气团有 2%来自机场西北方向、25%来自机场偏北方向、11%来自机场东北方向、24%来自机场偏东方向、21%来自机场东南方向以及 17%来自机场偏南方向。进一步分析可知，来自偏北、偏东、东南以及偏南方向的气流占比较多，因此在进行预测研究时考虑上述方向上游的气象条件以及污染物浓度分布情况对预测效果存在一定的作用。

6. 结论

本文利用 2017 年贵阳龙洞堡国际机场常规观测数据以及同期贵阳市新华路站点(1446A)环境污染物

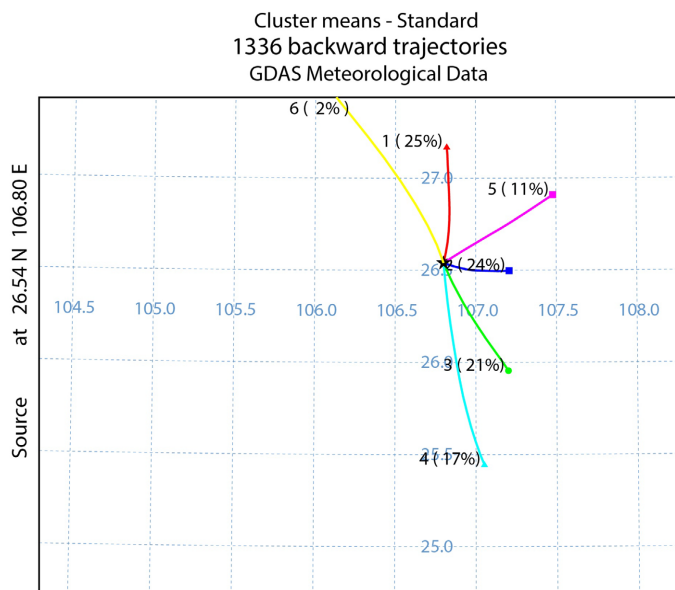


Figure 7. Cluster diagram of HYSPLIT model 12-hour backward trajectory of Guiyang Airport from January to December 2017. The black five-pointed star represents the longitude and latitude coordinate position of Guiyang, and different lines represent different source tracks

图 7. 2017 年 1 月~12 月贵阳机场的 HYSPLIT 模式 12 小时后向轨迹聚类图。黑色五角星表示贵阳经纬度坐标位置，不同线条表示不同的来源轨迹

数据，挑选影响能见度变化的变量因子作为输入数据集，采用机器学习随机森林算法对低能见度的变化进行预测分析。由于环境污染物对低能见度变化起到重要作用，进一步研究了各环境污染物的日变化特征和月变化特征。最后，利用 HYSPLIT 模式确定了机场近地面 50 米高度上气团的来源，这一研究有助于了解机场气象要素以及环境污染物的来源地。结论如下：

1) 随机森林预测值序列与真实值序列相关系数达 0.80，通过 99% 的显著性检验，说明随机森林算法在能见度变化趋势上预测效果较好，但预测值相对较为保守，在精度上还存在些许不足。总体上可以认为随机森林算法对贵阳机场能见度的变化预测效果较好。

2) 从随机森林算法输出变量重要性中可以看出，环境污染物的影响占比较大，其中近地面臭氧浓度变化对能见度变化的敏感性最高，说明低能见度数值的变化不仅仅是与温度、水汽等有关，还与近地面臭氧，PM_{2.5} 等密切相关。

3) 各环境污染物日变化和月变化特征均有所不同，浓度大值区和小值区的分布时段存在差异，因此造成的影响也有所不同。

贵阳机场 HYSPLIT 后向轨迹聚类分析结果表明，贵阳机场近地面的气团有 2% 来自机场西北方向、25% 来自机场偏北方向、11% 来自机场东北方向、24% 来自机场偏东方向、21% 来自机场东南方向以及 17% 来自机场偏南方向。其中，来自偏北、偏东、东南以及偏南方向的气流占比较多，因此在进行预测研究时考虑上述方向上游的气象条件以及污染物浓度分布情况对预测结果准确度存在一定的作用。

参考文献

- [1] 李惠彬. 从飞行事故看低云低能见度对飞行的影响[J]. 空中交通管理, 1997(4): 40-41.
- [2] 谢超, 马学款, 张恒德. 华南低能见度天气特征及客观预报研究[J]. 气象科学, 2019, 39(4): 556-561.

-
- [3] 王楠, 朱蕾, 周建军, 刘倩, 杨洪儒. 基于 EC 细网格产品在乌鲁木齐机场低能见度预测中的释用[J]. 沙漠与绿洲气象, 2020, 14(2): 81-89.
- [4] 司林青. 贵阳机场辐射雾天气特征及客观预报研究[J]. 气候变化研究快报, 2022, 11(3): 232-239. <https://doi.org/10.12677/CCRL.2022.113022>
- [5] 谭敏. 污染条件下气溶胶和水汽时空分布探测和数据分析研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2018.
- [6] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [7] Adele, C., David, R.C. and John, R.S. (2012) Ensemble Machine Learning. Springer, Boston, 157-176.
- [8] Ho, T.K. (1998) The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 832-844. <https://doi.org/10.1109/34.709601>
- [9] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术, 2018, 12(1): 49-55.
- [10] 余星源, 聂颖, 刘红年. 南京禄口机场低能见度天气统计分析判断模型[J]. 气象科学, 2022, 42(5): 661-667.
- [11] 付旭东, 王金艳, 李龙燕, 陈金车, 苏士翔, 常伟, 王明. 基于随机森林算法的风场预报[J]. 兰州大学学报(自然科学版), 2021, 57(4): 503-509. <https://doi.org/10.13885/j.issn.0455-2059.2021.04.010>
- [12] 朱国栋, 孙建杰, 陈阳权, 王照刚. 随机森林方法在机场温度预测中的应用[J]. 气象科技进展, 2021, 11(4): 34-36.
- [13] 徐敏辉, 徐颖, 梁秋枫. 2010-2021 年厦门高崎机场大雾天气 RVR 特征及预测[J]. 海峡科学, 2022(10): 26-31.
- [14] 任梅芳, 庞博, 徐宗学, 赵彦军. 基于随机森林模型的雅鲁藏布江流域气温降尺度研究[J]. 高原气象, 2018, 37(5): 1241-1253.
- [15] 赵丽. 基于随机森林模型的黑龙江省地表温度降尺度的研究[J]. 农业灾害研究, 2023, 13(1): 97-99.