

基于机器学习的云南省大雾预报方法研究

孙艳¹ 冯蕾^{2,3} 王传辉¹ 郑春怡⁴ 王涛¹ 钟亦鸣¹

1. 安徽省公共气象服务中心, 合肥 230031

2. 中国气象局公共气象服务中心, 北京 100081

3. 中国气象局交通气象重点开放实验室, 南京 210041

4. 云南省人工影响天气中心, 昆明市 650034

提要: 基于云南省 2016—2025 年 126 个国家气象站观测数据和同期 ERA5 再分析资料, 采用旋转经验正交函数 (REOF) 对云南省区域性大雾空间分布特征进行诊断, 结合地形因子开展大雾分区; 进一步利用随机森林等机器学习方法构建分区大雾预报模型, 将其结果与 SCMOC、ECMWF 数值预报产品进行对比验证。结果表明: 1) 云南大雾发生呈现滇西南、滇东南、滇东北及滇西北四种主导空间模态, 结合地形最终划分为四个大雾区: 1 区(滇西北)、2 区(滇东北)、3 区(滇西南)和 4 区(滇东南); 2) 随机森林模型在各区域建模样本准确率 0.82~0.84, 测试样本准确率 0.80~0.83, 显著优于决策树、K-近邻、逻辑回归和贝叶斯方法; 3) 预报因子在各大雾区的重要度存在明显差异, 高海拔区域(1 区、2 区)对 600—700 hPa 水汽条件依赖较强, 受东部工业活动影响显著的区域(2 区、4 区)初始能见度贡献较大, 相对低海拔区域(3 区、4 区)海平面气压及边界层内温度因子重要性更为突出; 4) 基于随机森林构建的大雾短期预报模型综合性能优于 SCMOC、ECMWF 两种主流预报产品, 尤以秋冬季预报效果提升更为明显, 其中 3 区和 4 区预报精度较高; 针对连续两天的区域性大雾过程, 随机森林模型的 TS 评分分别为 0.37 和 0.40, 显著高于 SCMOC (均 ≤ 0.1) 和 ECMWF (分别为 0.26 和 0.14) 预报评分。

关键词: 大雾; 空间分区; 机器学习; 预报模型; 云南省

Research on Heavy Fog Forecasting Methods in Yunnan Province Based on Machine Learning

SUN Yan¹ FENG Lei^{2,3} WANG Chuanhui¹ Zheng Chunyi⁴ WANG Tao¹

ZHONG Yiming¹

1 Anhui Public Meteorological Service Centre, Hefei 230031

第一作者: 孙艳, 主要从事交通高影响天气监测预报预警服务工作. E-mail: 1834631315@qq.com

通讯作者: 冯蕾, 主要从事交通气象监测预报预警服务技术研究. E-mail: fenglei21cn@163.com

资助项目: 中国气象局创新发展专项(CXFZ2025J071)、云南省气象局“揭榜挂帅”项目(YNJBGS202306)、安徽省气象局自主创新发展专项(AHQXZC202468)

Abstract: Based on the observation data from 126 national meteorological stations in Yunnan Province from 2016 to 2025 and ERA5 reanalysis data in the same period, the spatial distribution characteristics of regional heavy fog in Yunnan are analyzed with Rotated Empirical Orthogonal Function (REOF). Fog zoning is conducted in combination with topographic factors at the same time. Machine learning methods such as random forest are employed furthermore to construct regional heavy fog forecasting models. Then, the results are compared and verified against SCMOC and ECMWF numerical forecast products. The results show that heavy fog in Yunnan Province exhibits four dominant spatial modes: southwestern Yunnan, southeastern Yunnan, northeastern Yunnan, and northwestern Yunnan. Based on these modes and topographic factors, the province is divided into four heavy fog zones: Zone 1 (northwestern Yunnan), Zone 2 (northeastern Yunnan), Zone 3 (southwestern Yunnan), and Zone 4 (southeastern Yunnan). The random forest model achieves an accuracy of 0.82-0.84 for training set samples and 0.80-0.83 for test set samples in each zone, significantly outperforming the decision tree, K-nearest neighbor, logistic regression, and Bayesian methods. There are regional differences in the importance of forecast factors in each heavy fog zone. High-elevation areas (Zone 1 and Zone 2) strongly depend on water vapor conditions of 600-700 hPa. Regions that are significantly affected by eastern industrial activities (Zone 2 and Zone 4) are largely influenced by the initial visibility. In relatively low-elevation areas (Zone 3 and Zone 4), the importance of sea level pressure and temperature within the boundary layer is more prominent. The comprehensive performance of the forecasts by the short-range heavy fog forecast model constructed based on random forest is superior to that of two mainstream forecast products from SCMOC and ECMWF, especially in the improvement of forecast skills in autumn and winter. Among them, the forecasts for Zone 3 and Zone 4 have higher accuracy than that for other zones. Regarding a regional heavy fog event lasting for two consecutive days, the Threat Scores (TS) of the random forest model products are 0.37 and 0.40, which are significantly higher than those of SCMOC (both ≤ 0.1) and ECMWF (0.26 and 0.14, respectively).

Key words: heavy fog, spatial zoning, machine learning, forecasting model, Yunnan Province

引言

大雾通常指悬浮于近地面空气中的微小水滴或冰晶,使水平能见度降至 1 千米以下的天气现象(中华人民共和国国家质量监督检验检疫总局和中国国家标准化管理委员会, 2011)。大雾的发生伴随能见度急剧下降,易引发高速公路封闭、航班延误取消等管制措施,影响高速公路、民航等交通运行安全与运输效率。当出现能见度低于 500 m 的浓雾时,驾驶员对车距和障碍物的判断能力大幅下降,极易诱发交通阻断或事故,甚至造成人员伤亡和重大经济

损失(中国气象局, 2010; Gao et al, 2023; Letsatsi et al, 2025)。雾天通常伴随稳定大气层结与较弱空气垂直交换, 空气扩散条件较差, 雾滴易吸附细颗粒物并滞留在近地面层, 形成雾霾复合污染, 不仅加剧区域空气污染程度, 且细颗粒物通过呼吸进入人体, 对呼吸系统和心肺功能产生不利影响, 危害人体健康(Lyu et al, 2022)。深入认识大雾天气的成因, 提升大雾预报准确性, 对于保障交通安全和公共健康都具有现实意义。

目前, 国内外大雾预报研究已形成丰富的技术体系。传统方法主要以天气学诊断、统计回归预报、数值模式预报及其释用等为主(吴彬贵等, 2017; 田小毅等, 2018; Pithani et al, 2020)。传统预报方法在局地大雾精细化预报中存在预报准确性不足、时空分辨率偏低等短板, 难以满足气象灾害防御与交通保障的应用需求。近年来, 人工智能技术发展迅速, 以机器学习为代表的智能算法凭借强大的非线性拟合与多因子建模能力, 为解决复杂天气预报问题提供了新思路, 已广泛应用在大雾、降水、温度等气象要素预报领域(Sansonnet et al, 2024; 夏杨等, 2026)。Bartok et al (2022)采用支持向量机、决策树、k 近邻等多种机器学习算法对能见度在 300 m 的雾预报研究, 结果表明机器学习方法具备良好的雾预测潜力。张庆等 (2024)利用决策树、随机森林等多种机器学习方法建立京津冀地区低能见度天气预报模型, 引入 ERA5 再分析资料后, 模型的预报性能显著提升。蔡晓杰等 (2025)基于洋山港历史海雾个例库, 建立海雾分型决策树预报模型, 模型雾预报性能优于 ECMWF, 进一步验证了机器学习在局地雾预报中的应用潜力。上述研究多集中于我国东部沿海及平原地区, 针对地形复杂、气候特征独特的低纬高原地区大雾预报的研究仍相对匮乏。

云南地处低纬高原, 山地、高原与盆地相间分布, 地形条件复杂, 气候区域分异显著, 水汽输送与局地环流特征独特, 大雾发生频率高且空间分布不均(陶云等, 2011; 段玮等, 2019)。特殊的地理与气候背景使得大雾形成、维持及消散机制更为复杂(王元等, 2021)。与此同时, 云南省作为西南地区重要的交通枢纽与旅游大省, 高速公路、航空等交通网络密集, 对大雾精细化预报的需求尤为迫切。因此结合机器学习等智能预报技术, 开展云南复杂地形条件下的大雾预报研究, 建立适用于不同区域特征的分区预报模型, 是当前云南大雾精细化预报业务需要解决的关键问题。

近年来, 欧洲中期天气预报中心发布的再分析数据集(ERA5)时空分辨率高, 其数据质量与可靠性已得到广泛验证, 目前已被大量应用于各类气象要素预报模型研究中(张庆等, 2024; Murillo-Domínguez et al, 2026)。基于上述背景, 本研究以云南省大雾为研究对象, 结合近年来云南省国家气象站观测资料, 选取区域性大雾个例, 采用 REOF 方法对云南省大雾天气进行分区; 分析归纳出五类影响大雾生消的关键气象要素, 确定机器学习预报因子,

利用再分析数据集建立多种机器学习算法模型，并筛选出各分区最优预报模型，实现云南省大雾精细化客观预报。研究成果可为云南大雾预报业务提供技术参考。

1 数据和方法

1.1 数据

使用的数据主要有：中国气象局气象信息中心提供的 2016 年 1 月—2025 年 7 月云南省 126 个国家气象站观测的能见度、2 m 温度、相对湿度、10 m 风速、10 m 风向和降水数据(站点分布见图 1)，时间分辨率为 1 h。对地面观测数据开展多要素联合质量控制，能见度 ≤ 500 m 时，若相对湿度 $< 70\%$ 、风速 > 4 m/s 或降水量 ≥ 0.1 mm，则判定该组观测为异常数据或雨雾，予以剔除。欧洲中期天气预报中心提供的第五代全球气候和天气再分析数据集(ERA5)，包括 2016 年 1 月—2024 年 6 月海平面气压、不同高度层(1000 hPa、950 hPa、925 hPa、900 hPa、850 hPa、700 hPa、600 hPa 和 500 hPa)气温、相对湿度、风速等气象要素，时间分辨率为 1h，空间分辨率为 $0.25^\circ \times 0.25^\circ$ 。欧洲中期天气预报中心(ECMWF)提供的细网格数值预报数据，包括 2024 年 7 月—2025 年 6 月海平面气压、不同高度层(1000 hPa、950 hPa、925 hPa、900 hPa、850 hPa、700 hPa、600 hPa 和 500 hPa)气温、相对湿度、风速，以及地面能见度等要素，起报时间为北京时 08 时，预报时效为 24 h，时间分辨率为 3 h，空间分辨率为 $0.125^\circ \times 0.125^\circ$ 。中国气象局国家气象中心发布的国家级格点预报指导产品(SCMOC)，该产品基于数值模式输出并结合统计订正、多模式融合等技术生成(胡争光等，2020；潘留杰等，2023)，研究采用 2024 年 7 月—2025 年 6 月能见度预报数据，时间分辨率为 3 h，空间分辨率为 $5\text{ km} \times 5\text{ km}$ 。本文中使用的地图通过中华人民共和国自然资源部批准，审图号为 GS (2024) 0650 号。

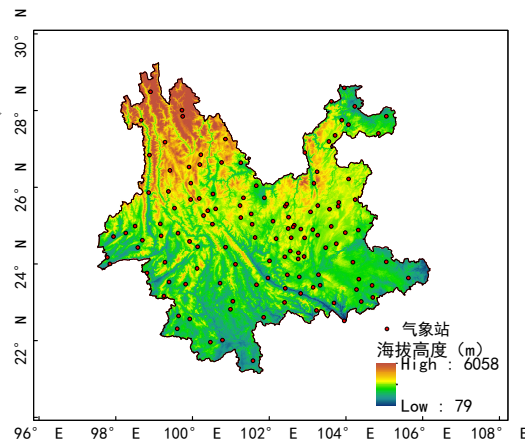


图 1 研究区气象站点及海拔高度空间分布

Fig.1 Spatial distribution of meteorological stations in the study area and the terrain elevation

1.2 方法

1.2.1 旋转经验正交函数

旋转经验正交函数(ReOF)是在经验正交函数(EOF)(张志高等, 2025)基础上, 取累计贡献方差符合要求的前几个主成分作为载荷特征向量, 对其进行方差最大正交旋转, 旋转后的主成分贡献率相比 EOF 更加平均, 简化了特征向量的结构。旋转分解出的空间模态更加清晰, 能够更好的表现地域的差异, 同时反映不同地域的相关分布情况, 实现将气候要素一致的地区划分为同一区域(林颖仪等, 2021; 龚成麒等, 2022)。本研究主要利用 ReOF 方法提取云南省大雾分布的主要空间分布型, 在此基础上分析得到大雾分区。

1.2.2 机器学习方法

研究中构建预报模型时考虑模型性能、可解释性及适用场景等, 选取决策树、逻辑回归、K-近邻、贝叶斯和随机森林五种不同类型的经典机器学习算法(Castillo-Botón et al, 2022; Fridkin and Bendersky, 2026):

决策树是自上而下递归划分的监督学习方法, 通过特征选择准则(如信息增益、基尼系数等)构建树状结构, 模型从根节点开始, 不断将数据集划分为纯度更高的子集, 直至满足停止条件(Safavian and Landgrebe, 1991)。该方法结构直观、易于理解, 能够处理非线性关系。

逻辑回归是一种广义线性分类模型(Hastie et al, 2009), 通过引入 Sigmoid 函数将线性模型的输出映射到 $[0,1]$ 区间, 获得样本属于某一类别的概率估计, 并采用极大似然估计求解参数。该方法计算复杂度低、训练效率高、可解释性强, 可以检验线性可分条件下的分类性能。

K-近邻算法是基于实例的惰性学习方法(Cover and Hart, 1967), 通过计算新样本与训练集中所有样本的距离, 找出最近的 K 个邻居, 以多数投票或距离加权投票的方式确定类别。该方法原理直观, 适用于多分类问题, 能够考察局部样本分布对分类结果的影响。

贝叶斯分类以贝叶斯定理为核心(Hastie et al, 2009), 通过先验概率与条件概率计算样本属于各类别的后验概率, 选择后验概率最大的类别作为结果。该算法所需估计参数少, 计算高效且在小样本与高维特征场景下表现稳定。

随机森林是基于决策树的集成学习算法, 通过自助采样法抽取多个样本子集, 并随机选择特征子集构建多棵决策树, 最终通过投票或平均集成输出结果。该方法能有效降低单棵决策树的过拟合风险, 提升泛化能力, 适用于需提升预测稳定性的分类任务(Stafoggia et al, 2020)。

1.2.3 检验方法

威胁评分(TS)是气象学等领域中用于评估预报精度的一种重要方法,广泛应用于降水、大雾等气象灾害数值预报检验中(李蕊等, 2025; 吴科娟等, 2025)。它综合考量了空报和漏报两种错误,排除了正确否定的干扰,能够直观地反映对关键性天气事件的预报能力。通常与空报率、漏报率等指标结合使用,能够对预报性能进行较为全面的评估(沈文强等, 2023)。研究采用 TS 评分、准确率(EH)、空报率(NH)和漏报率(PO)来检验模型预报效果,具体计算方法见式(1)—(4)。其中 NA 表示有雾预报正确次数、NB 表示空报次数、NC 表示漏报次数、ND 表示无雾预报正确次数。

$$TS = NA / (NA + NB + NC) \quad (1)$$

$$EH = (NA + ND) / (NA + NB + NC + ND) \quad (2)$$

$$NH = NB / (NA + NB) \quad (3)$$

$$PO = NC / (NA + NC) \quad (4)$$

2 结果和分析

2.1 大雾分区

使用 2016—2024 年云南省 126 个国家气象站能见度及相关气象要素数据,统计得到能见度低于 500 m 的雾日。剔除总雾日数不足 5 天的站点,并筛选出雾站数大于总站数 1/10 的日期作为一个个例,共获得 401 个大雾个例。人工剔除站点分布零散个例,最终选取 84 个站点、347 个区域性大雾个例作为分区研究样本。

根据选取的云南省区域性大雾个例,以能见度 ≤ 500 m 的站点取值为 1,能见度 > 500 m 的站点取值为 0,利用 REOF 分析得到其主要空间分布模态(图 2)。第一模态(图 2a)中心区域位于云南省西南部;第二模态(图 2b)中心区域位于云南省东南部,第三模态(图 2c)中心区域位于云南省东北部,第四模态(图 2d)中心区域位于云南省西北部。由于第四和第五模态(图 2e)大部分高相关区重叠,进一步结合云南地形特点(图 1)将云南省大雾区分为四个区域(图 3)。图 3 可以看到 1 区(滇西北)和 2 区(滇东北)海拔较高、地形复杂;3 区(滇西南)位于河谷及低海拔地区,水汽充沛,有利于平流雾与辐射雾并发;4 区(滇东南)地形以丘陵河谷为主,湿度较大,大雾的生成可能与平流过程和地形抬升密切相关。

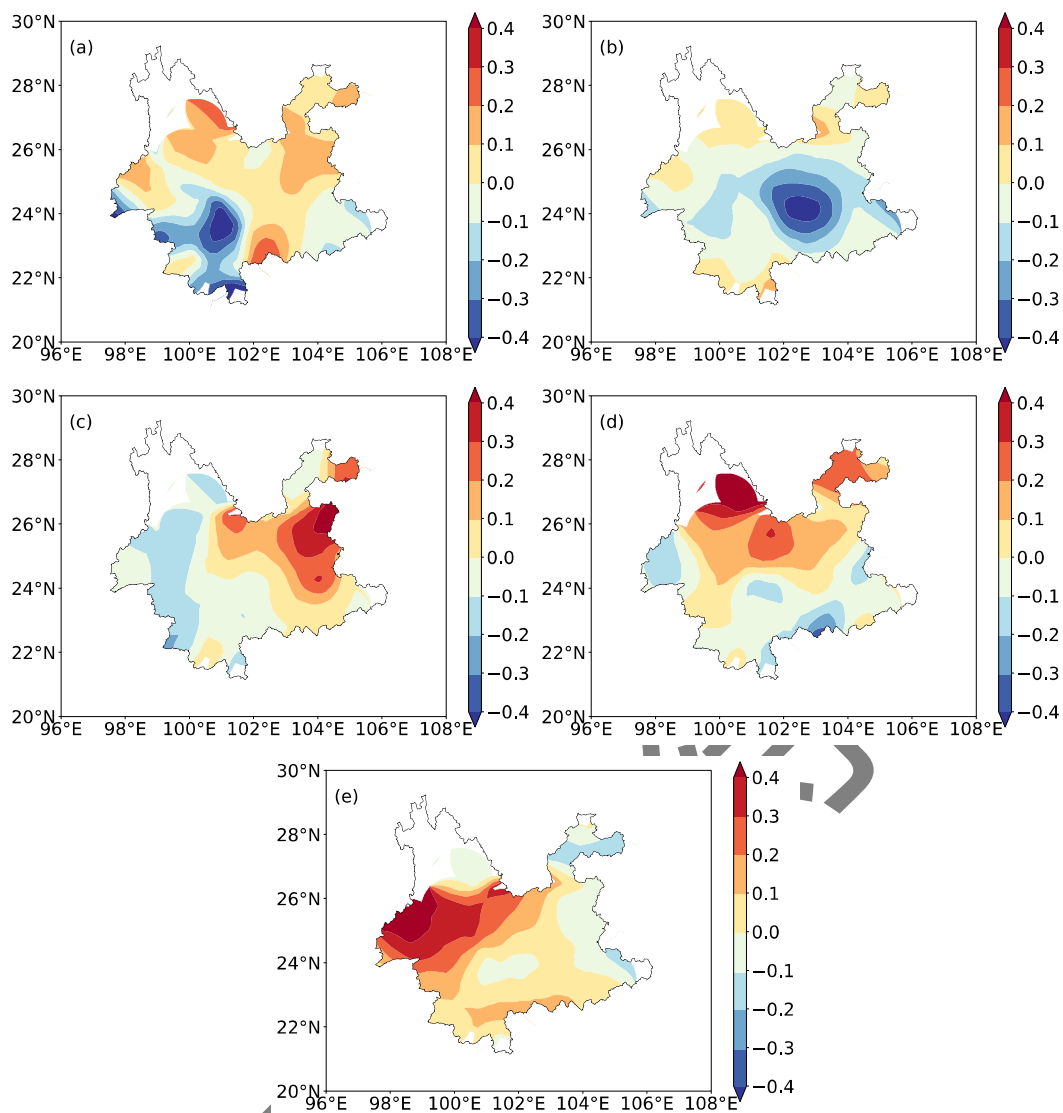


图 2 基于 REOF 分解的 2016—2024 年云南省大雾个例前 5 个模态的空间分布

Fig.2 Spatial distribution of the first five modes derived from REOF decomposition for heavy fog events in Yunnan Province from 2016 to 2024

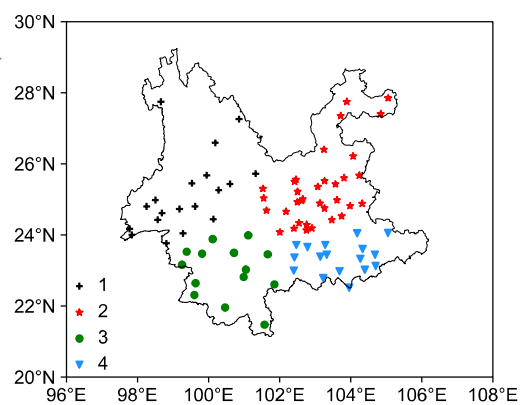


图 3 2016—2024 年利用 REOF 方法分区的云南省大雾分区

Fig.3 Heavy fog zoning by the REOF method in Yunnan Province from 2016 to 2024

2.2 模型样本与预报因子选取

2.2.1 模型样本准备

基于大雾分区结果，对 2016 年 1 月—2024 年 6 月云南省国家气象站大雾观测样本进行分区域统计。由于有雾样本数量偏少、无雾样本偏多，样本分布存在明显不均衡问题，因此通过对不同能见度等级样本赋予相应的样本均衡系数完成样本均衡处理。具体均衡方案为：保留所有能见度范围为 $[0,500)$ m 和 $[500,1000)$ m 的样本，对 $[1000,3000)$ m 及 $(3000,+\infty)$ m 的样本进行随机抽样，随机种子统一设置为 42，最终使能见度范围为 $[0,500)$ m、 $[1000,3000)$ m 和 $(3000,+\infty)$ m 的三类样本数量比例接近 1:1:1。

2.2.2 预报因子选取

从大雾形成机理出发，综合考虑前期降水、初始湿度和能见度实况，以及预报时刻不同高度的风速、湿度、地面气温、逆温、海平面气压等气象因素，同时在逆温层高度选取上考虑了海拔高度，最终选取 35 个变量作为预报因子输入模型(表 1)。预报因子可以分为 5 类。第一类为水汽因子，包括前期降水量、相对湿度等气象要素，直接提供形成大雾所需的水汽条件。在不同区域，如沿海地区由于蒸发较强，相对湿度普遍较高，更易形成平流雾；内陆干旱或高海拔地区则可能因水汽匮乏而使大雾的形成受到抑制(郑征等，2007)。第二类为温度因子，包括地面温度、温度露点差、云量等物理量，温度及其变化直接影响空气的饱和水汽压和冷却速率，进而决定水汽是否凝结成雾滴，温度露点差越小，表明空气越接近饱和状态，更有利于大雾的形成。在山区或高海拔地区，夜间降温幅度大，易形成辐射雾(尤红等，2008)。第三类是稳定度因子，包含海平面气压、风速、逆温相关物理量。逆温抑制垂直扩散，有利于近地面水汽聚集，从而促成大雾的发展；风速适中时利于水汽输送，但过强会破坏雾层，过弱则不利于大雾的均匀形成(田梦等，2025)。第四类为时间因子，考虑能见度时间变化特征，表现为能见度日变化、月变化。在不同纬度与气候区，大雾的日变化和季节分布可能存在差异，这类因子能够捕捉大雾的气候规律和周期性特征，提升模型在特定时段和季节中的预报能力。第五类是能见度初始实况，在不同地理区域，如污染较重的城市区域，可受到气溶胶浓度的影响，此时初始能见度实况对大雾预报具有重要指示作用(崔堃等，2025)。

表 1 大雾预报模型预报因子及其分类

Table 1 Forecast factors and their classification for the fog forecasting model

因子类型	预报因子	简写
水汽因子	过去 12 小时降水量	pre(0-12H)
	起报当天 16 时相对湿度实况	T16_rh
	1000 hPa-500 hPa 各层相对湿度	1000、950、925、900、850、700、600、500hPaRH
温度因子	2 m 气温	T2m
	温度露点差	T2m-DEW
	总云量、低云量	tcc、lcc
稳定度因子	海平面气压	sp
	海平面气压梯度	msldy
	10 m 风速	10mwind
	1000 hPa-500 hPa 各层风速	1000、950、925、900、850、700、600、500hPaWS
	1000 hPa 与地面温差	1000hPa-sufT
	925 hPa 与地面温差	925hPa-sufT
	925 hPa 与 1000 hPa 高度温差	925-1000hPaT
	800 hPa 与 850 hPa 高度温差	800-850hPaT
	750 hPa 与 800 hPa 高度温差	750-800hPaT
	700 hPa 与 750 hPa 高度温差	700-750hPaT
时间因子	所在北京时刻	Hour
	所在月份	Mon
初始条件	起报当天 16 时能见度	T16_vis

2.3 预报模型训练及评估

基于选取的预报因子对 4 个分区建立预报模型。利用选取的样本，根据其时间计算相应的预报因子，并将样本按照 7:3 的比例分为训练集和测试集，利用训练集分别采用随机森林、逻辑回归、K-近邻、贝叶斯、决策树 5 种机器学习方法进行训练。图 4 给出各预测方法的训练样本准确率和测试样本准确率。从训练样本准确率来看，随机森林在四个区域中均表现最优(1 区 0.84，2 区 0.83，3 区 0.82，4 区 0.83)，说明其对训练数据具有较强的拟合能力；决策树与 K-近邻方法也显示出较好的拟合效果，而贝叶斯方法拟合能力明显偏弱，在各区训练准确率均低于 0.64。测试样本准确率方面，随机森林仍然保持最高，平均达 0.81，且与

训练准确率较为接近,表明模型泛化能力良好;决策树方法准确率次之,平均准确率为 0.77;逻辑回归方法在测试集上表现稳定,训练与测试准确率基本一致,说明模型结构简单但稳健(平均准确率为 0.75);K-近邻方法在测试集上准确率普遍明显低于建模结果,可能存在过拟合;贝叶斯方法准确率最低,平均不足 0.60。综合比较,随机森林在各区域训练与测试阶段均表现最优,选取为最终模型预测方法。

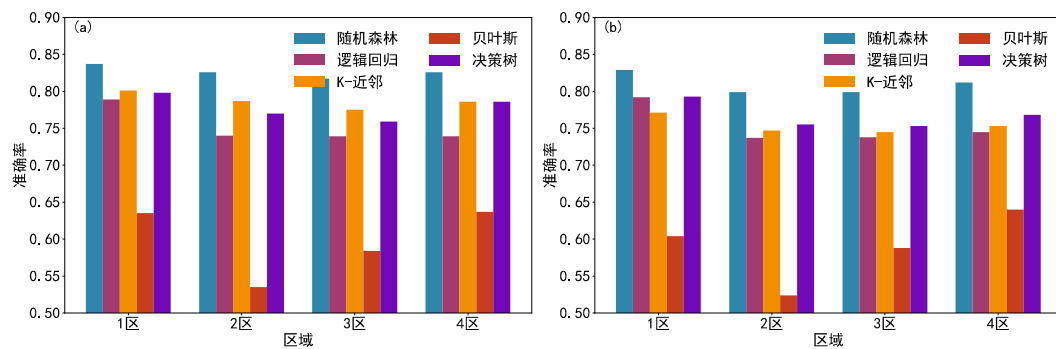


图 4 2016—2024 年云南省 5 种机器学习方法大雾分区域建模的(a)训练集和(b)测试集样本的准确率

Fig.4 The accuracy of (a) training set and (b) test set samples of regional heavy fog modeling by five machine learning methods in Yunnan Province from 2016 to 2024

2.4 预报因子重要度分析

随机森林算法可以通过计算基尼系数来评估各因子重要度(刘帅等, 2024), 将重要度最高的因子定义为 100, 计算其他各因子与其相对数值。计算各区域随机森林算法模型中各预报因子特征重要度, 并提取前十位因子(图 5)。可以看到, 1 区和 2 区中 600—700 hPa 相对湿度作用突出, 其中 1 区 600 hPa 相对湿度重要度达 100, 这可能与该区域海拔多高于 2000 m 有关, 高海拔地区边界层顶较高, 中层水汽平流输送对辐射雾的形成更为关键。2 区和 4 区 16 时初始能见度的重要度分别位列第一、二位, 或与东部工业城市的污染物输送有关, 初始能见度偏低往往表明气溶胶浓度较高或近地面已具备一定水汽条件, 为夜间辐射降温成雾提供有利前置环境。3 区和 4 区海平面气压重要度均居首位, 反映出大尺度气压系统对该区域大雾生消的重要调控作用。海平面气压的变化不仅影响大气稳定性和垂直运动条件, 还与水汽输送和近地面风场特征密切相关。另外 4 区海拔相对较低, 800 hPa-850 hPa 和 925 hPa-1000 hPa 高度层的温差影响也十分重要, 表明边界层内的热力结构对低海拔地区大雾的形成具有直接的控制作用。边界层内温差反映逆温强度与低层稳定度, 温差越大, 表明层结越稳定, 越有利于水汽在近地面累积成大雾。总体而言, 水汽、温度、稳定度及初始能见度在各区域均为关键因子, 但受地形与人类活动等影响, 不同区域因子重要度存在差异。

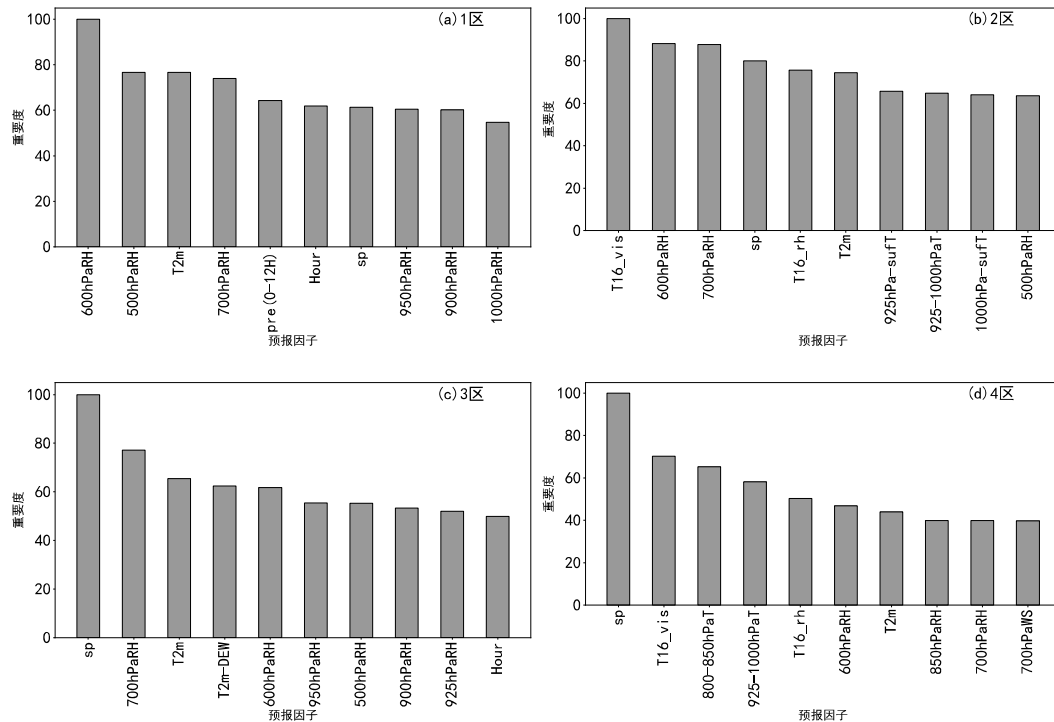


图 5 2016—2024 年云南省基于随机森林方法的大雾分区域重要度排名前十的预报因子

Fig.5 The top ten forecast factors of importance ranking by Random Forest method for regional heavy fog forecasting in Yunnan Province from 2016 to 2024

2.5 预报检验

2.5.1 模型效果预报检验

利用 2024 年 7 月—2025 年 6 月的模型客观预报数据对模型的实际预报能力进行独立检验。由于云南大雾的形成、发展主要出现在夜间至次日日出前后，将前一日 20 时至当日 08 时作为模型检验时段，该时段内某站点实况出现能见度低于 500 m 的情况记为该站点出雾，同期预报结果只要出现为 1 的情况，则判断预报正确。在此基础上计算 TS、EH、NH 和 PO。另外获取同期 SCMOC(客观原因缺少 2025 年 1—2 月预报数据)和 ECMWF 能见度预报产品，进行检验对比。

筛选 2024 年 7 月—2025 年 6 月选出出雾站数大于总站数 1/10 的大雾过程，得到 87 个大雾个例进行检验。从全省平均来看(图 6)，短期预报的 TS 评分(0.24)最高，其次是 ECMWF 预报(0.13)，SCMOC 预报最低(0.06)；短期预报的空报率低于 SCMOC 和 ECMW；漏报率明显低于 ECMWF 和 SCMOC。

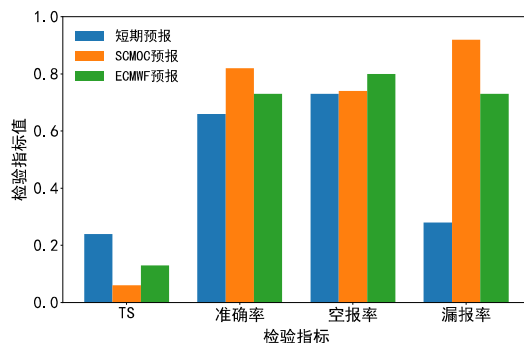


图 6 2024 年 7 月至 2025 年 6 月云南省大雾过程短期预报模型、SCMOC 和 ECMWF 预报检验结果

Fig.6 Forecast verification results of short-range forecast model, SCMOC and ECMWF for heavy fog events in Yunnan Province from July 2024 to June 2025

分月检验结果如图 7 所示,短期预报逐月 TS 评分在 0.08~0.33 之间,准确率在 0.55~0.89 之间,空报率在 0.47~0.76 之间,漏报率在 0.18~0.90 之间。总的来看短期预报在全省平均及多数月份综合预报效果最好,TS 评分较高、空报率和漏报率较低;其中秋冬季节预报效果最好,在春季和夏秋季节转换时期漏报情况有所增多。

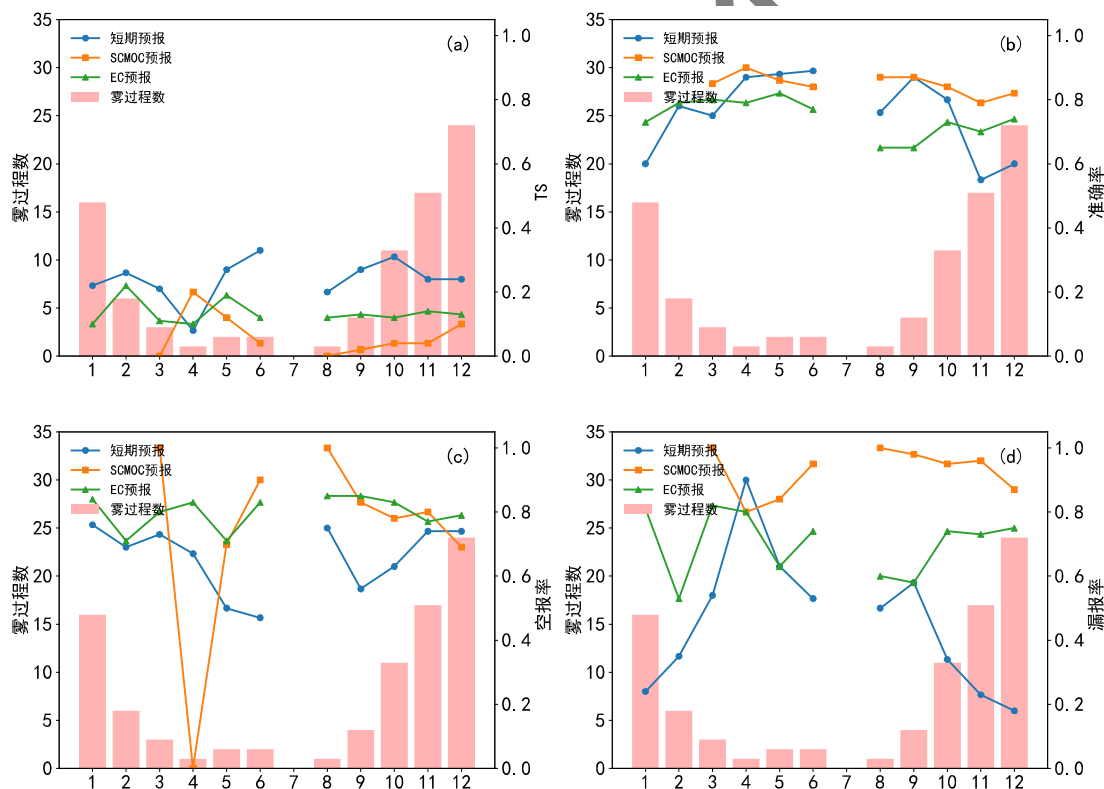


图 7 2024 年 7 月至 2025 年 6 月云南省大雾过程短期预报模型、SCMOC 和 ECMWF 逐月预报检验结果

Fig.7 Monthly forecast verification results of short-range forecast model, SCMOC and ECMWF for heavy fog events in Yunnan Province from July 2024 to June 2025

进一步根据云南省大雾客观分区结果，针对各大雾区选取出雾站点大于该区域总站数 1/10 的大雾过程，进行分区检验(图 8)。从 TS 评分来看，短期预报在 1-4 区域 TS 评分均明显高于 SCMOC 和 ECMWF 预报，其中 3 区 TS 评分最高(0.44)。短期预报在 3 区和 4 区的准确率最高(分别为 0.76 和 0.85)，SCMOC 预报在 1 区和 2 区的准确率最高(分别为 0.83 和 0.80)。从空报率来看，短期预报在 2-4 区的空报率最低，ECMWF 预报在 1 区的空报率最低(0.74)，其次是短期预报(0.76)、SCMOC 预报(0.78)。从漏报率来看，短期预报在各区域的漏报率均低于 ECMWF 和 SCMOC，尤其在 2 区和 3 区漏报率最低(均为 0.30)。总之，短期预报在各分区预报效果最好，其中 3 区和 4 区更优。

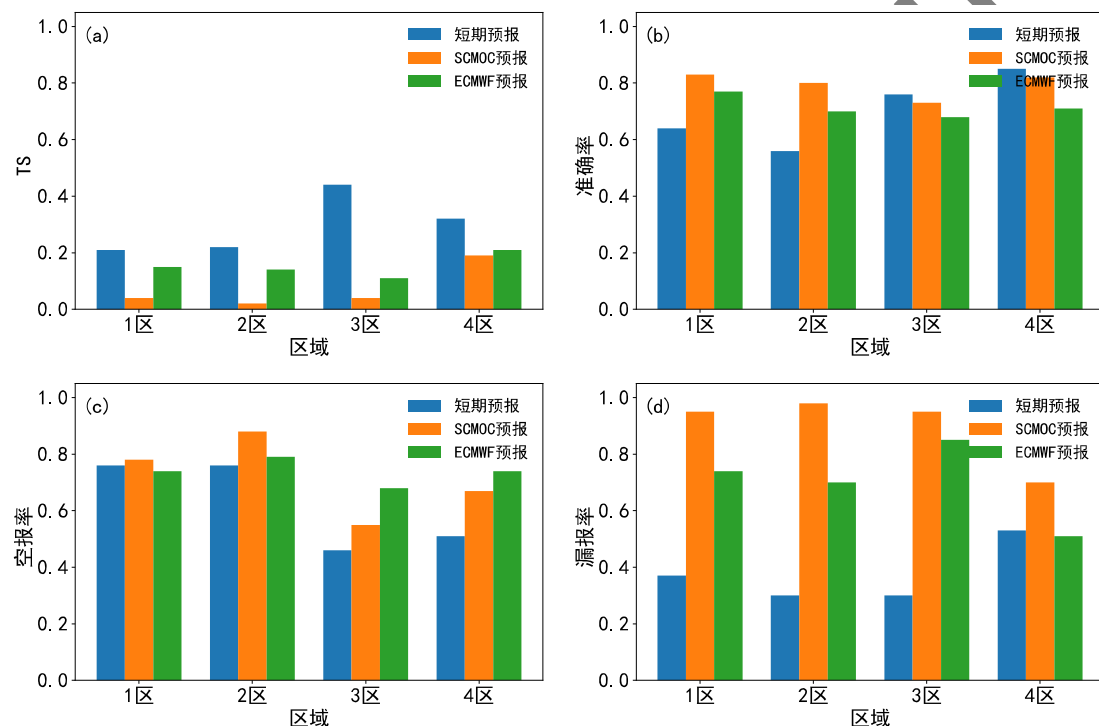
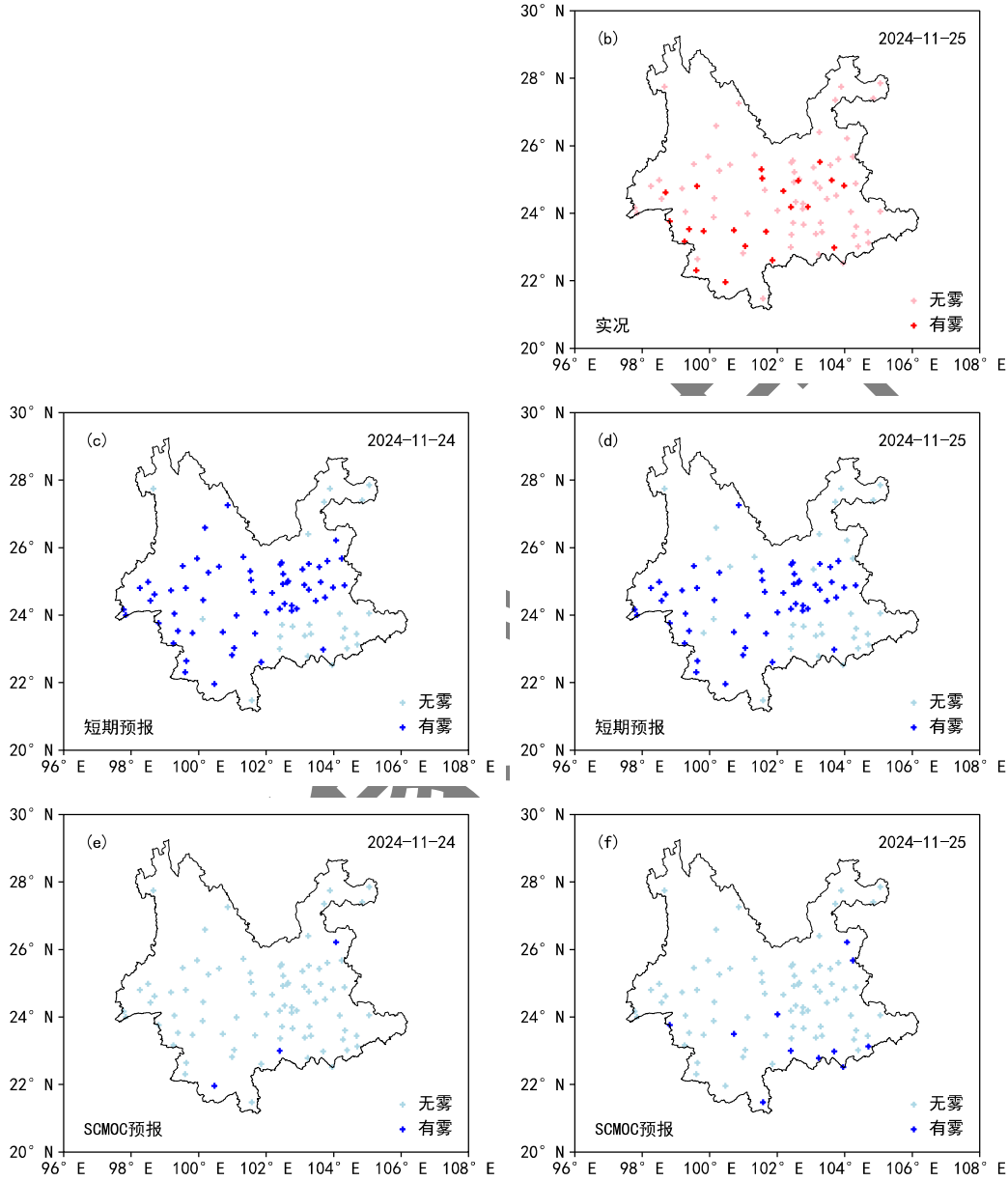


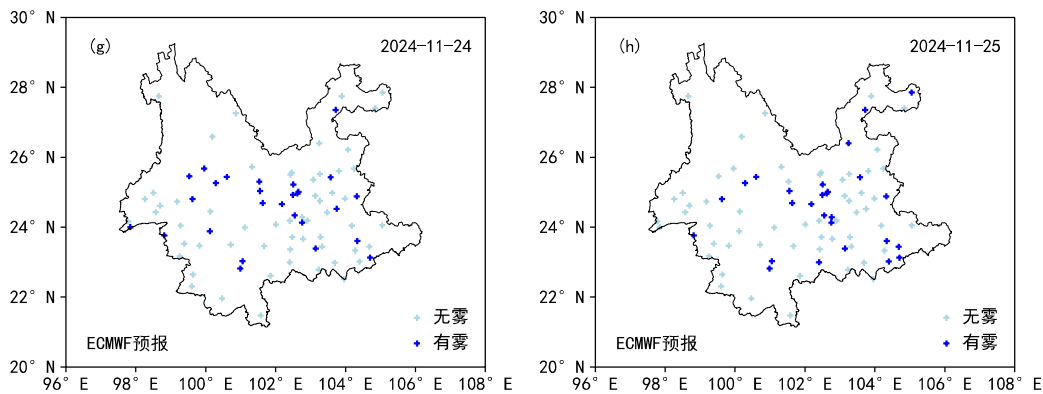
图 8 2024 年 7 月—2025 年 6 月云南省各分区大雾过程短期预报模型、SCMOC 和 ECMWF 预报检验结果
Fig.8 Forecast verification results of short-range forecast model, SCMOC and ECMWF for heavy fog events in the four zones of Yunnan Province from July 2024 to June 2025

2.5.2 典型大雾天气过程检验

选取出雾范围较大的连续两天的区域性大雾过程(2024 年 11 月 24 日-25 日)进行预报检验。图 9 给出了大雾过程能见度实况和三种预报方法的预报结果。11 月 24 日(指北京时间 11 月 23 日 20 时-24 日 08 时)和 11 月 25 日(指北京时间 11 月 24 日 20 时-25 日 08 时)，主要于云南西南部和中南部地区发生区域性大雾。图 10 为连续两天大雾过程检验情况。可以看到：短期预报模型预报的大雾区基本上包括了实况大雾区，两天大雾过程 TS 评分分别为 0.37 和 0.40,漏报率很低，分别为 0.12 和 0.05，空报率分别为 0.62 和 0.60；SCMOC 预报在两天大

雾过程中均只预报出云南南部局部大雾区，TS 评分均不超过 0.1，漏报严重；ECMWF 预报在两天大雾过程中主要预报出云南中部和南部大雾区，TS 评分分别为 0.26 和 0.14，空报率分别为 0.59 和 0.78，漏报率分别为 0.58 和 0.73。由此可见短期预报模型在连续两天的区域性大雾过程中综合预报能力最佳，相比于主流数值预报方法，大雾预报效果具有明显提升。





(注：有雾表示检验时段内出现能见度 ≤ 500 m；无雾表示检验时段内能见度均 > 500 m)

图9 2024年11月(a,c,e,g)24日和(b,d,f,h)25日区域性大雾过程能见度(a,b)实况和(c-h)三种方法
预报结果

Fig.9 Comparison between (a, b) the observed visibility and (c-h) the forecasted visibility by three methods for regional heavy fog events on (a, c, e, g) 24 and (b, d, f, h) 25 November 2024

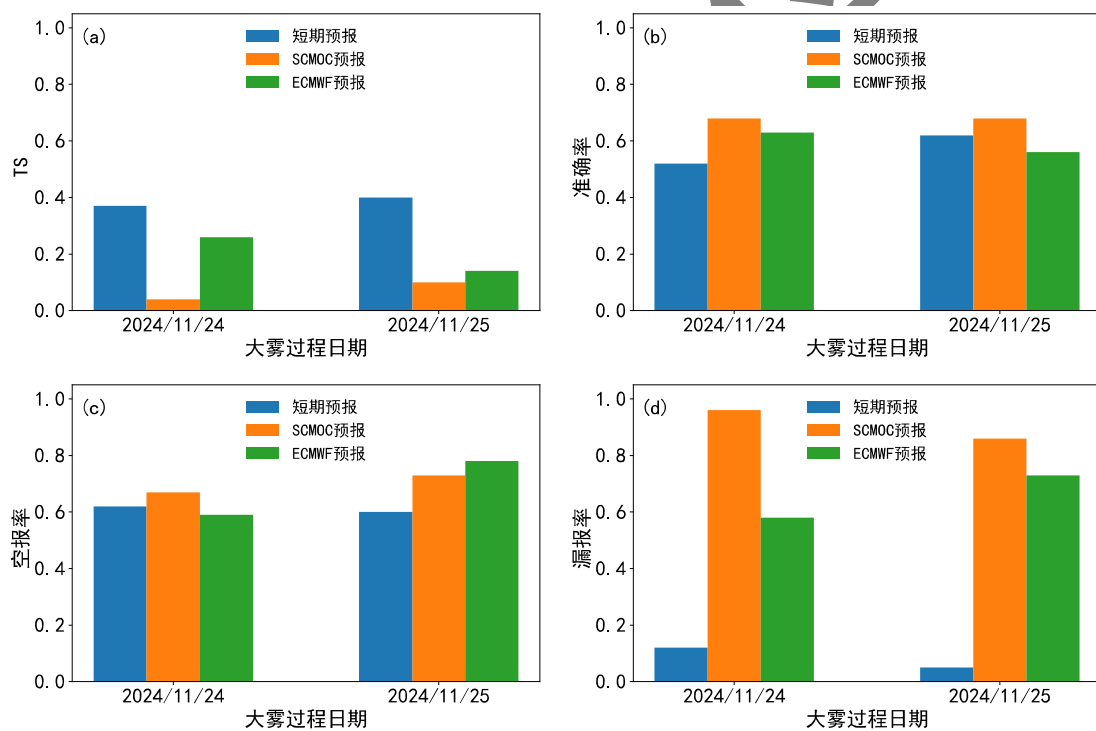


图10 2024年11月24—25日区域性大雾过程短期预报模型、SCMOC和ECMWF预报检验结果

Fig.10 Forecast verification results of short-range forecast model, SCMOC and ECMWF for the two-day consecutive regional heavy fog event from 24 to 25 November 2024

3 结论与讨论

基于云南省2016—2025年的国家气象站观测数据和ERA5再分析资料，采用REOF方

法对云南省大雾进行分区,利用随机森林等机器学习方法构建分区预报模型,并与 SCMOC、ECMWF 数值预报产品进行对比检验。得到以下结论:

(1) 云南省大雾空间分布具有显著的区域性差异。根据 REOF 分析以及地形因子将云南划分为四个大雾区:1 区(滇西北)和 2 区(滇东北)以高海拔为主、3 区(滇西南)多为河谷及低海拔区域和 4 区(滇东南)以丘陵河谷为主

(2) 随机森林方法在四个分区均表现最优,建模准确率为 0.82~0.84,测试准确率为 0.80~0.83,兼具高拟合能力与良好泛化性能。决策树预测准确率次之(平均 0.77),K-近邻虽拟合效果较好但泛化能力欠佳,逻辑回归稳健但预测精度有限(平均 0.75),贝叶斯方法预测准确率明显偏弱(平均<0.60)。

(3) 不同区域预报模型中各气象因子重要度具有明显差异:高海拔区(1、2 区)高度依赖 600—700 hPa 中高层水汽,尤其 1 区中 600 hPa 相对湿度重要度达 100;受工业影响区(2、4 区)侧重初始能见度;低海拔区(3、4 区)以海平面气压为首要因子,3 区地面温度和 4 区边界层温度差也起到重要作用。

(4) 随机森林模型在全省及各分区综合预报能力均优于 SCMOC 和 ECMWF 数值预报产品。时间上看,秋冬季预报效果最佳;空间上,3 区、4 区 TS 评分和准确率较高。在连续两天的区域性大雾过程 TS 分别为 0.37 和 0.40,明显优于 SCMOC(TS 均 $\leq$ 0.1)和 ECMWF(分别为 0.26 和 0.14)。

研究在客观分区基础上,结合随机森林等机器学习算法构建了分区大雾预报模型,整体预报效果优于主流数值预报产品,但仍存在以下局限:1)受复杂地形和站点分布不均的影响,模型预报能力有限,部分区域空报和漏报偏多。朱毓颖等(2024)采用 PhyDNet-ATT 深度学习方法建立江苏省能见度预报模型,也指出在观测站点密集区域的预报误差显著低于观测站点稀疏区域。本研究的分区建模方法在一定程度上减轻了复杂下垫面对大雾预报的影响,但在站点稀疏地区预报能力仍然不足。2)当前模型主要依赖常规气象要素,未考虑气溶胶浓度等人为影响因子,已有研究表明在高污染区域气溶胶消光对 1 公里以下低能见度过程具有显著贡献(Xie et al, 2024),本模型通过初始能见度因子间接表明了城市化及污染传输对低能见度的影响,但对其刻画能力有限。3)模型在春季或春秋转换季节预报能力不足,可能由于单一机器学习模型存在固有不确定性,不同算法在特定天气形势下或不同类型雾事件中的表现差异明显(Xiang et al, 2025)。

未来可从以下方面进行改进:一是融合卫星遥感、无人机探测及高分辨率自动站等多源观测数据,改善复杂地形区域站点稀疏问题;二是引入气溶胶浓度、下垫面类型、人为热源

等因子,构建包括人为影响的综合预报因子集,加深对局地大雾生消机制的理解;三是开展多模型集成研究,如将随机森林与 XGBoost、LightGBM 及深度学习模型进行集成,针对不同分区、不同季节或不同天气型建立自适应权重分配模型,实现分区、分型、分时的精细化预报。上述改进可进一步提升云南大雾精细化预报水平,为当地交通气象保障和灾害风险防控提供更扎实的技术支撑。

参考文献

- 蔡晓杰, 朱智慧, 岳彩军, 等, 2025. 洋山港海雾特征分析及分型决策树模型预报评估[J]. 气象, 51(8): 954-963. Cai X J, Zhu Z H, Yue C J, et al, 2025. Sea fog characteristics and forecast evaluation of classification decision tree models at Yangshan Port[J]. Meteor Mon, 51(8): 954-963 (in Chinese).
- 崔堃, 王红磊, 严殊祺, 等, (2025-06-05)[2026-03-09]. 雾的观测及其与气溶胶和边界层相互作用研究进展[J/OL]. 三峡生态环境监测, <https://link.cnki.net/urlid/50.1214.x.20250604.1819.002>. Cui K, Wang H L, Yan S Q, et al, (2025-06-05)[2026-03-09]. Advances in fog observations and their interaction with aerosols and the boundary layer[J/OL]. Ecol Environ Monit Three Gorges, <https://link.cnki.net/urlid/50.1214.x.20250604.1819.002> (in Chinese). (查阅网上资料, 未找到本条文献引用日期, 请确认)
- 段玮, 段旭, 邢冬, 等, 2019. 影响云南地区适航条件的视程障碍天气气候特征[J]. 气象科技, 47(2): 329-336. Duan W, Duan X, Xing D, et al, 2019. Influence of visual obstruction weather on airworthiness conditions and its climatic characteristics in Yunnan[J]. Meteor Sci Technol, 47(2): 329-336 (in Chinese).
- 龚成麒, 董晓华, 董立俊, 等, 2022. 雅砻江流域 REOF 分区的降水特征及其未来趋势分析[J]. 水土保持研究, 29(3): 78-87. Gong C Q, Dong X H, Dong L J, et al, 2022. Analysis of precipitation characteristics and future trends in the REOF subarea of the Yalong River Basin[J]. Res Soil Water Conserv, 29(3): 78-87 (in Chinese).
- 胡争光, 薛峰, 金荣花, 等, 2020. 智能网格预报应用分析平台设计与实现[J]. 气象, 46(10): 1340-1350. Hu Z G, Xue F, Jin R H, et al, 2020. Design and implementation of gridded forecast application analysis platform[J]. Meteor Mon, 46(10): 1340-1350 (in Chinese).
- 刘帅, 王涛, 曹佳文, 等, 2024. 基于优化随机森林模型的降雨群发滑坡易发性评价——以西秦岭极端降雨事件为例[J]. 地质通报, 43(6): 958-970. Liu S, Wang T, Cao J W, et al, 2024. Susceptibility assessment of precipitation-induced mass landslides based on optimal random forest model: taking the extreme precipitation event in western Qinling mountains as an example[J]. Geol Bull China, 43(6): 958-970 (in Chinese).
- 李蕊, 张萍萍, 刘奕辰, 等, 2025. 四种数值模式对山东半岛两类暴雨天气预报的检验分析[J]. 暴雨灾害, 44(5): 632-641. Li R, Zhang P P, Liu Y C, et al, 2025. Verification and analysis of four numerical models for forecasting rainstorm associated with two weather patterns in Shandong Peninsula[J]. Torrential Rain Disasters, 44(5): 632-641 (in Chinese).

- 林颖仪, 王式功, 马盼, 等, 2021. 基于 REOF 方法的海南岛人体舒适度特征分析[J]. 干旱气象, 39(5): 838-846. Lin Y Y, Wang S G, Ma P, et al, 2021. Characteristics of human body comfortable degree in Hainan Island based on REOF method[J]. J Arid Meteor, 39(5): 838-846 (in Chinese).
- 潘留杰, 张宏芳, 刘静, 等, 2023. 智能网格 SCMOC 及多模式降水预报对比[J]. 大气科学学报, 46(2): 217-229. Pan L J, Zhang H F, Liu J, et al, 2023. Comparative analysis of SCMOC and various numerical models for pre-cipitation forecasting[J]. Trans Atmos Sci, 46(02): 217-229 (in Chinese).
- 沈文强, 钱浩, 马昊, 等, 2023. 浙江梅汛期暴雨预报的客观订正方案对比分析[J]. 气象, 49(6): 697-707. Shen W Q, Qian H, Ma H, et al, 2023. Objective correction schemes on the forecast of torrential rain during the Meiyu Period in Zhejiang[J]. Meteorol Mon, 49(6): 697-707 (in Chinese).
- 陶云, 段旭, 段长春, 等, 2011. 云南近 50a 雾的变化特征[J]. 云南大学学报(自然科学版), 33(3): 308-316. Tao Y, Duan X, Duan C C, et al, 2011. The change characteristics of fog in Yunnan during the nearly 50 years[J]. J Yunnan Univ, 33(3): 308-316 (in Chinese).
- 田梦, 吴彬贵, 廖云琛, 等, 2025. 雾顶辐射冷却驱动的湍流混合对华北平原雾的影响作用[J]. 大气科学, 49(2): 460-476. Tian M, Wu B G, Liao Y C, et al, 2025. Influence of turbulent diffusion driven by fog-top radiative cooling on heavy fog in the North China Plain[J]. Chin J Atmos Sci, 49(2): 460-476 (in Chinese).
- 田小毅, 朱承瑛, 张振东, 等, 2018. 长江江苏段江面雾的特征和预报着眼点[J]. 气象, 44(3): 408-415. Tian X Y, Zhu C Y, Zhang Z D, et al, 2018. Characteristics and forecasting focus for river fog in Jiangsu section of Yangtze River[J]. Meteor Mon, 44(3): 408-415 (in Chinese).
- 王元, 牛生杰, 陆春松, 等, 2021. 西双版纳热带雨林地区冬季辐射雾理化特征的观测研究[J]. 中国科学: 地球科学, 51(12): 2098-2111. Wang Y, Niu S J, Lu C S, et al, 2021. Observational study of the physical and chemical characteristics of the winter radiation fog in the tropical rainforest in Xishuangbanna, China[J]. Sci Sin Terr, 51(12): 2098-2111 (in Chinese).
- 吴彬贵, 张建春, 李英华, 等, 2017. 天津港秋冬季低能见度数值释用预报研究[J]. 气象, 43(7): 863-871. Wu B G, Zhang J C, Li Y H, et al, 2017. Research on numerical interpretative forecast for low-visibility at Tianjin Port in autumn and winter[J]. Meteor Mon, 43(7): 863-871 (in Chinese).
- 吴科娟, 张苏平, 李昕蓓, 等, 2025. 基于机器学习的西北太平洋海雾预报模型研究[J]. 海洋气象学报, 45(3): 18-29. Wu K J, Zhang S P, Li X B, et al, 2025. Research on sea fog prediction model over Northwest Pacific based on machine learning[J]. J Mar Meteor, 45(3): 18-29 (in Chinese).
- 夏杨, 谢英, 王晓峰, 等, 2026. 基于机器学习的上海地区站点低能见度预报改进和检验[J]. 热带气象学报, 42(1): 153-164. Xia Y, Xie Y, Wang X F, et al, 2026. Optimized forecasting and verification of low visibility for Shanghai stations based on machine learning[J]. J Trop Meteor, 42(1): 153-164 (in Chinese).
- 尤红, 杨明, 郭荣芬, 等, 2008. 云南昆洛高速峨山段典型山地雾的诊断分析[J]. 气象, 34(8): 87-94. You H, Yang M, Guo R F, et al,

2008. Diagnostic analysis of typical mountainous fog process in Eshan along Kunming-Daluo high way in Yunnan[J]. Meteorol Mon, 34(8): 87-94 (in Chinese).
- 张庆, 张楠, 陈子煊, 等, 2024. 多种机器学习方法在京津冀地区低能见度天气预报中的应用[J]. 气象科学, 44(2): 362-374. Zhang Q, Zhang N, Chen Z X, et al, 2024. Application of multiple machine learning methods in low-visibility weather forecasting in Beijing-Tianjin-Hebei region[J]. J Meteor Sci, 44(2): 362-374 (in Chinese).
- 张志高, 胡柱钰, 贾梦薇, 等, 2025. 基于 EOF 和 REOF 的河南省人体舒适度时空分布特征[J]. 西南大学学报(自然科学版), 47(9): 189-199. Zhang Z G, Hu Z Y, Jia M W, et al, 2025. Spatial and temporal distribution characteristics of human comfort in Henan Province based on EOF and REOF[J]. J Southwest Univ (Nat Sci Ed), 47(9): 189-199 (in Chinese).
- 郑征, 李佑荣, 张树斌, 等, 2007. 西双版纳海拔变化对水湿状况的影响[J]. 山地学报, 25(1): 33-38. Zheng Z, Li Y R, Zhang S B, et al, 2007. Influence of the altitudinal increase on water and humidity conditions, Xishuangbanna[J]. J Mt Sci, 25(1): 33-38 (in Chinese).
- 中国气象局, 2010. X/T 111-2010 高速公路交通气象条件等级[S]. 北京: 气象出版社. China Meteorological Administration, 2010. QX/T 111-2010 Grade of weather conditions for freeway transportation[S]. Beijing: China Meteorological Press (in Chinese).
- 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会, 2011. GB/T 27964-2011 雾的预报等级[S]. 北京: 中国标准出版社. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, 2011. GB/T 27964-2011 Grade of fog forecast[S]. Beijing: China Standards Press (in Chinese).
- 朱毓颖, 郑玉, 张备, 2024. 基于 PhyDNet-ATT 的能见度预报方法[J]. 应用气象学报, 35(6): 667-679. Zhu Y Y, Zheng Y, Zhang B, 2024. Visibility forecast based on PhyDNet-ATT deep learning algorithm[J]. J Appl Meteor Sci, 35(6): 667-679 (in Chinese).
- Bartok J, Šišan P, Ivica L, et al, 2022. Machine learning-based fog nowcasting for aviation with the aid of camera observations[J]. Atmosphere, 13(10): 1684.
- Castillo-Botón C, Casillas-Pérez D, Casanova-Mateo C, et al, 2022. Machine learning regression and classification methods for fog events prediction[J]. Atmos Res, 272: 106157.
- Cover T, Hart P, 1967. Nearest neighbor pattern classification[J]. IEEE Trans on Inf Theory, 13 (1): 21-27.
- Fridkin S, Bendersky M, 2026. Interpretable machine learning: a comprehensive review of foundations, methods, and the path forward[J]. Wires Data Min Knowl, 16(1): e70075.
- Gao J J, Tian H, Li A X, et al, 2023. Analysis of agglomerate fog meteorological characteristics in Anhui province based on traffic accident data[J]. Pure Appl Geophys, 180(1): 313-333.
- Hastie T, Tibshirani, Friedman J, 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction[M]. 2nd ed. New

York: Springer, 119-134.

Letsatsi N, Dube K, Mearns K, 2025. Fog and global aviation: the state of knowledge evolution[M]//Mukandavire Z, Al Ali H, Dube K.

Aviation Industry Trends-Technology Advancements, Regulations, and Sustainability. Cham: Springer: 271-283.

Lyu R, Gao W, Peng Y R, et al, 2022. Fog-haze transition and drivers in the coastal region of the Yangtze River Delta[J]. Int J Environ Res Public Health, 19(15): 9608.

Murillo-Domínguez J, Molina-Almaraz M, García-Sánchez E, et al, 2026. Lightweight deep learning surrogates for ERA5-based solar forecasting: an accuracy-efficiency benchmark in complex terrain[J]. Technologies, 14(2): 97.

Pithani, P, Ghude, S D, Jenamani, R K, et al, 2020. Real-time forecast of dense fog events over Delhi: the performance of the WRF model during the WiFEX field campaign[J]. Wea Forecasting, 35(2): 739-756.

Safavian S R and Landgrebe D, 1991. A survey of decision tree classifier methodology[J]. IEEE Transactions on Systems, Man, and Cybernetics, 21(3):660-674.

Sansonnet C, Schaeffer M, Baarsch F, 2024. Seasonal forecast of two-meter temperature and precipitation in Tanzania: a hybrid cluster and point-by-point machine learning approach[J]. J Appl Meteor Climatol, 63(10): 1255-1271.

Stafoggia M, Johansson C, Glantz P, et al, 2020. A random forest approach to estimate daily particulate matter, nitrogen dioxide, and ozone at fine spatial resolution in Sweden[J]. Atmosphere, 11(3): 239.

Xiang Y Q, Zhang Q H, Wang M Q, et al, 2025. A hybrid forecasting approach for sea fog based on machine learning[J]. Artif Intell Earth Syst, 5(1): e250013.

Xie Y, Wang X F, Gao Y Q, et al, 2024. Improving visibility forecasting during haze-fog processes in shanghai and eastern China: the significance of aerosol and hydrometeor extinction[J]. Atmos Environ, 337: 120756.