

基于灰色关联度分析与随机森林算法的龙岩上杭梅花始花期预报模型研究

邱子键 江帆 赵娜 何清芳*

福建省龙岩市气象局

DOI:10.32629/as.v9i7.4174

[摘要] 为提升龙岩上杭梅花花期预报精度、支撑文旅服务需求,本文以2016—2024年梅花始花期观测数据及同期冬季气象资料为基础,采用灰色关联分析法筛选关键气象因子,构建随机森林(RF)与台湾需冷量两类花期预报模型并进行对比验证。结果表明:研究区梅花始花期集中于12月中下旬,年际差异明显;日均气温、5cm地温、最低气温及相对湿度为影响花期的主导因子;随机森林模型预报精度显著优于需冷量模型,始花期预报均方根误差仅1.2天。本研究可为当地梅花观赏旅游气象服务、花期管理及景观规划提供科学支撑。

[关键词] 梅花始花期;灰色关联分析;随机森林;需冷量模型

中图分类号: S685.17 文献标识码: A

Study on the First Bloom Period Forecast Model of Plum Blossoms in Shanghang, Longyan Based on Grey Relational Analysis and Random Forest Algorithm

Zijian Qiu Fan Jiang Na Zhao Qingfang He*

Longyan Municipal Meteorological Bureau, Fujian Province

[Abstract] To improve the forecast accuracy of plum blossom bloom period in Shanghang, Longyan and meet the demands of cultural and tourism services, this study adopted the grey relational analysis method to screen key meteorological factors based on the observed data of the first bloom period of plum blossoms and synchronous winter meteorological data from 2016 to 2024. Two bloom period forecast models, namely the Random Forest (RF) model and Taiwan chill unit model, were constructed and comparatively verified. The results showed that the first bloom period of plum blossoms in the study area was mainly concentrated in the middle and late December with significant inter-annual variations. Daily average temperature, 5 cm ground temperature, minimum temperature and relative humidity were the dominant meteorological factors affecting the bloom period. The Random Forest model exhibited significantly higher forecast accuracy than the chill unit model, with the root mean square error of the first bloom period forecast as low as 1.2 days. This study can provide scientific support for meteorological services for plum blossom ornamental tourism, bloom period management and landscape planning in the local area.

[Key words] first bloom period of plum blossoms; grey relational analysis; random forest; chill unit model

1 引言

“赏花游”是春日旅游的热门,近年来,多地旅游部门把赏花与旅游文化相结合开展各种富有特色的旅游活动以助推旅游产业的发展和提升财政收入。花期预测成了公众旅游规划、部门组织赏花庆典活动的重要依据。上杭县梅花山景区是龙岩市已成规模的赏梅胜地,具有很高的观赏、文化、经济、价值。针对发展梅花旅游的旺盛需求,建立适合本地梅花始花期预报模

型,及时准确的预报海洋镇观音井景区的梅花始花期,为景区宣传及游客科学安排赏花旅游路线等均有重要意义。

梅花喜温耐寒,需要满足阳光充足、排水良好和空气清洁等基本条件,其开花都与气象息息相关,尤其始花期与12月至翌年2月的气象因子关系紧密^[1-5]。探讨气候因子对花期的影响并建立二者的相关性、精确定梅花始花期及花期长短,可为观光旅游业的发展提供有价值的借鉴。

目前国内普遍采用的花期预报方法有不同下限温度积温分析法建立花期预报模型,并将预报结果与检验年份实况进行对比,检验不同积温指标对花期预报的效果,从中筛选出最优预报方案。例如:赵桂琼^[6]等基于桂花物候观测与气象资料开展成都市秋桂花前、花期与气象因子的关系研究,采用逐步回归方法建立秋桂始花期预测模型,并能较常年花期提前20天进行预测。目前国内外对梅花始花期的预报方法相对较少,检索到较为成熟的梅花花期预报方法有廖碧婷^[3]等参考多种梅花需冷量模型:7.2℃模型、0~7.2℃模型、尤他模型、台湾模型,并引用对广州梅花最优的台湾模型,利用自动站逐时气温数据和梅花花期数据标定界限温度,建立了广州地区梅花花期的预报模型。这类方法想要预测花期,必须有足够的梅花花期和气象观测数据,且花期预报主要采用的方法主要参考温度等热量条件,花期预报准确率有待提高。

大数据时代背景下,机器学习作为人工智能领域的重要分支,国内外越来越多的学者将机器学习技术应用于各个领域,在预测模型方面初显优势。这些分析方法能够从多水平、多因素着手,综合分析各指标的整体效应,使筛选出的结果更具科学性^[7-12]。吉莉等基于主成分分析法(PCA),通过BP神经网络算法及逐步回归算法,构建了重庆蜡梅花初花期预测模型,均能很好地预测花期^[7-12]。

本研究以上杭湖洋观音井地区梅花为研究对象,基于2016—2023年梅花始花期观测资料,以及同期冬季(12月—翌年2月)气象要素数据,利用机器学习方法开展花期预报模型构建与对比研究。采用灰色关联分析法量化各因子与始花期的关联程度,筛选出关联度较高的关键气象因子作为模型输入变量,明确影响该地区梅花始花期的主导气象要素。以筛选得到的关键气象因子为输入,采用随机森林(RandomForest, RF)算法构建机器学习花期预报模型,此外引入花期预报中广泛应用的需冷量模型作为对比,结合龙岩本地气候特征修正模型界限温度,构建适配区域特点的梅花始花期需冷量预报模型。以均方根误差、预报准确率为核心指标,对两类模型的预报精度进行对比分析与评估。可为当地梅花观赏旅游气象服务、花期管理及景观规划提供科学支撑。

2 数据与方法

2.1 数据及其来源

梅花花期资料为课题组对上杭观音井景区实地走访调查所得。本文采用降水观测数据为福建省2016—2024年上杭国家气象站的观测资料。包括冬季12月至翌年2月逐日最高气温、最低气温、日均气温、日降水量、日照时数、日均5cm地温、日均相对湿度等气象观测资料。

2.2 灰色关联法

灰色关联分析方法是因素之间发展趋势的相似或相异程度作为衡量因素间关联程度的一种方法。以2016—2024年始花期作为母序列,以同期气象观测资料为系统影响因素子序列,计算母序列与各子序列之间的关联度,选取关联程度相对较高的因

子作为模型的输入因子从而建立预测模型^[13]。

2.3 随机森林

随机森林(RandomForest, RF)算法是以决策树为基分类器的一个集成学习模型。通过自助法随机抽样构建独立决策树,以平均输出回归结果。它能有效避免单棵决策树的过拟合问题,泛化能力强,还可输出特征重要性排序,广泛应用于物候预测等回归与特征筛选任务^[14]。样本数据集由影响始花期的气象因子、始花期日序构成。

2.4 需冷量模型

研究中选用已证实可行或已经投入业务运行的花期预报模型与基于机器学习建立的预报模型进行分析比较。台湾模型规定7.2℃为打破休眠的最有效低温范围,此温度下1h计为1CU,而7.3~15.0℃的低温范围只有半效作用,1h计为0.5CU,15.1~26.6℃为无效温度,1h计为0CU,26.7~27.8℃低温效应部分被解除,该温度范围内1h计为-0.5CU,27.8℃以上低温效应完全被解除,该温度范围内1h计为-1CU。统计过去2016—2024年梅花进入始花期的需冷量,并以此为指标,计算达到此需冷量的日期即为预报的梅花始花期。

3 主要结论

3.1 龙岩上杭观音井梅花始花期和花期的年统计特征

2016—2024年上杭观音井景区梅花始花期与花期长度均呈现显著年际波动与非线性演变特征(图1)。始花期主要集中在12月中下旬,年际前移与推迟变幅剧烈:2018年最早,显著提早至12月10日;2016年、2021年及2023年最晚,均推迟至12月25日前后。这种波动与暖冬、冷冬交替的年际气候背景高度契合,表明龙岩冬季气温变化对梅花休眠解除与花芽催醒具有强烈的调控作用。

花期长度介于37~52天,最大年际差值达15天,离散度较高且极端年份特征明显。2022年花期最长达52天,与当年较早始花期及持续平稳的低温条件密切相关;2018年、2020年及2024年花期最短,均为37天,多因后期气温快速回升导致花期提前结束。



图1 2016—2024年梅花始花期及花期变化图

3.2 梅花始花期气象因子筛选

根据灰色系统理论,通过对2016—2024年上杭观音井梅花始花期前各气象因子的变化序列与始花期日序(DOY)母序列进行灰色关联度计算与排序,得到了各气象驱动因子的关联序列

表(表1)。结果显示,不同气象因子对梅花始花期的调控效应存在显著的层级差异,整体呈现出“温度与地温主导、湿度及光照协同、日最高气温弱相关”的驱动特征。

在所有参选的气象因子中,日均气温与始花期的灰色关联系数最高,达到了0.7245,这表明日均气温的年际波动与梅花始花期的早晚演变趋势最为同步。其次是日均5cm地温,其关联系数高达0.6918,与日均气温表现出极强的协同驱动力。日最低气温的关联系数也达到了0.6582。日均相对湿度和日照时数的关联系数分别为0.5834、0.5421,属于较强的协同调节因子。因此,在后续构建梅花始花期机器学习预报模型时,选择日均气温、日均5cm地温、日均最低气温和日均相对湿度等核心高关联的4个因子作为模型的输入变量。

表1 始花期前气象因子与梅花始花期的灰色关联序

气象因子	关联度系数
日均气温/℃	0.7245
日均5cm地温/℃	0.6918
日最低气温/℃	0.6582
日均相对湿度/%	0.5834
日照时数/h	0.5421
日降水量/mm	0.4816
日最高气温/℃	0.3475

3.3 预测模型结果分析

模型检验与对比：为实现上杭观音井梅花花期精准预报,本研究基于灰色关联法筛选的核心气象因子构建特征集,采用时间序列划分法,将2016—2021年历史观测数据作为训练集,2022—2024年独立数据作为测试集,分别构建随机森林(RF)机器学习模型与台湾需冷量(CU)物候模型开展横向对比试验,两类模型的需冷量统计及精度检验结果分别见表2、表3。

表2 基于台湾模型计算的2016-2024年上杭观音井梅花各物候期需冷量(CU)

年份	初花期	凋谢期
2016 年	78.5	290.5
2017 年	85.0	310.0
2018 年	42.5	225.5
2019 年	90.0	320.0
2020 年	93.5	368.0
2021 年	137.5	470.5
2022 年	112.0	415.0
2023 年	124.0	432.5
2024 年	70.5	265.0
平均值	92.6	344.1

结果表明,随机森林模型在非线性拟合与泛化能力上优势

显著。始花期预测中,RF模型均方根误差(RMSE)仅1.2天、平均绝对误差(MAE)0.9天,较需冷量模型(RMSE=2.8天)精度提升57.1%;在波动剧烈的花期长度预测中,RF模型RMSE仍稳定在1.5天,同样优于需冷量模型。

精度差异源于两类模型的作用机制不同。需冷量模型虽能清晰解释梅花低温休眠解除的生理机理,但仅依赖逐时气温单一变量,无法兼顾地温、湿度等因子对花蕾发育的协同调控,在气候波动年份易产生系统性偏差。而随机森林通过多决策树集成,可有效解析多源气象因子的复杂交互作用,即便面对2022年52天的异常长花期极端样本,仍保持较强抗干扰能力,展现了机器学习算法在精细化旅游气象服务中的应用潜力。

表3 随机森林模型与台湾需冷量模型在测试集(2022—2024年)上的精度指标对比

预报目标	评估模型	均方根误差 (RMSE)/d	平均绝对误差 (MAE)/d
始花期	台湾需冷量模型 (CU)	2.8	2.2
	随机森林模型 (RF)	1.2	0.9
花期长度	台湾需冷量模型 (CU)	3.4	2.7
	随机森林模型 (RF)	0.5	1.1

4 总结与讨论

本研究聚焦福建省龙岩市上杭观音井景区梅花花期演变规律,采用灰色关联分析法筛选关键气象因子,构建随机森林(RF)预报模型,并与台湾需冷量模型开展对比验证。结果表明,2016—2024年该区域梅花始花期与花期长度均呈显著年际非线性波动,始花期集中于12月中下旬,花期长度平均42.7天。热量与地温是影响花期的主导因素,灰色关联分析显示,日均气温、日均5cm地温、日均最低气温和日均相对湿度与始花期的关联度均大于0.5。

模型对比试验表明,随机森林模型的预报精度与稳定性显著优于传统需冷量模型。其在独立测试集上的始花期预测均方根误差(RMSE)仅1.2天,花期长度预测RMSE为1.5天,较需冷量模型分别提升57.1%和55.9%。需冷量模型虽能清晰解释低温解除休眠的生理机理,但输入特征单一,无法捕捉地温、湿度等因子对花蕾发育的交叉调控;而随机森林通过决策树集群可有效解析多源气象特征,即便面对2022年52天的异常长花期极端样本,仍保持较强抗干扰泛化能力。

本研究仍存在改进空间。随机森林模型存在“黑箱”问题,难以定量还原植物生理发育过程。未来可将需冷量累积值作为物理约束嵌入机器学习损失函数,构建物理驱动的融合模型,实现物候机理与数据智能的深度结合,进一步提升预报效果。

[基金项目]

龙岩市科技计划重点项目：2024LYF9066(龙财教[2024]57号); 龙岩市科技计划重点项目 2024LYF9065(龙财教[2024]57号)。

[参考文献]

[1]贾坤,张黎霞,赵天禄,等.北京地区梅花的花期预报模式初建[J].北京林业大学学报,2010,32(S2):97-100.

[2]杨亚会,李庆卫.部分梅花品种的需冷量研究[J].北京林业大学学报,2013,35(S1):47-51.

[3]廖碧婷,王四化,杜尧东,等.广州地区梅花花期的预报方法[J].广东气象,2016,38(4):55-58.

[4]易延勇,杨平会,陈波,等.贵州荔波野生梅花花期预报研究[J].安徽农业科学,2013(32):12656-12657.

[5]陈俊愉.梅花研究六十年[J].北京林业大学学报,2002,24(5):224-229.

[6]赵桂琼,邓锐捷,陈志龙,等.成都市秋桂开花期气象指标及预测[J].中国农学通报,2022,38(29):135-139.

[7]孙家清,张志薇,艾文文.BP神经网络在油菜花期预报中的应用[J].气象与环境科学,2019,42(4):22-26.

[8]赵洪杰,闫景东,陈广瑞,等.积温对比法在鸭梨始花期预报中的应用[J].气象科技,2019,47(5):872-878.

[9]冯敏玉,孔萍,胡萍,等.基于花前物候利用灰色关联分析法建立油菜花期预报模型[J].中国农业气象,2021,42(11):929-938.

[10]徐敏,赵艳霞,张顾,等.基于机器学习算法的冬小麦始花期预报方法[J].农业工程学报,2021,37(11):162-171.

[11]李德,陈文涛,乐章燕,等.基于随机森林算法和气象因子的砀山酥梨始花期预报[J].农业工程学报,2020,36(12):143-151.

[12]姚日升,涂小萍,丁焯毅,等.宁波桃树花期预报方法[J].气象科技,2014,42(1):180-186.

[13]吉莉,李强.重庆市蜡梅开花期气候适宜度模型构建及潜在种植区研究[J].河南农业科学,2025,54(3):147-156.

[14]李德,乐章燕,陈文涛,等.基于等距拆分和随机森林算法的皖北小麦始花期气象预报[J].麦类作物学报,2023,43(6):784-797.

作者简介:

邱子键(1995--),汉族,福建长汀人,硕士研究生,助理工程师,灾害性天气的预报预警技术。

江帆(1982--),女,汉族,福建上杭人,职称:高级工程师,学士,研究方向:水库、旅游等专业气象服务 龙岩市气象局。

赵娜(1987--),女,汉族,辽宁铁岭人,职称:工程师,大学,研究方向:气象科普。