

基于 CFSv2 产品和机器学习的东江流域月降水预报*

庄胜杰¹, 王大刚^{1,2}, 林泳恩¹, 林泽群¹, 陈润庭¹

1. 中山大学地理科学与规划学院, 广东 广州 510006
2. 广东省粤北岩溶区碳水耦合野外科学观测研究站, 广东 广州 510006

摘要: 中长期降水预报一直以来是研究水文气象变化的热点, 其精度与可靠性不高的问题亟待解决。以东江流域为研究对象, 通过距平相关系数 ACC、标准化均方根误差 NRMSE、平均绝对误差 MAE 和多模型稳定性指数 MSI 评估 CFSv2 模式产品在月尺度的预测精度与稳定性, 采用 CFSv2 模式降水预报、CFSv2 模式预报因子结合机器学习模型预报 2 种方法预测未来降水。结果表明, 不同预见期下, CFSv2 模式降水预测与实测降水量具有较高的相关性, 对于枯水期的预测效果好于汛期, 但随着起报时间发生改变, 降水预测的差异性较大, 模型稳定性较差; CFSv2 模式预报因子结合机器学习模型提高了预测的稳定性, 相较于 CFSv2 模式降水预测, MSI 从 0.45 降低到 0.25, 在很大程度上减小了由于起报时间改变产生的预报随机性。研究成果可为中长期降水预测提供一种新的思路, 并为中长期水文预报和水资源管理提供决策依据。

关键词: CFSv2; 中长期预报; 机器学习; 产品评估

中图分类号: P457.6 **文献标志码:** A **文章编号:** 2097-0137(2024)04-0009-10

Monthly precipitation forecast in the Dongjiang Basin based on CFSv2 products and machine learning

ZHUANG Shengjie¹, WANG Dagang^{1,2}, LIN Yongen¹, LIN Zequn¹, CHEN Runtong¹

1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China
2. Carbon-Water Observation and Research Station in Karst Regions of Northern Guangdong, Guangzhou 510006, China

Abstract: Mid to long-term precipitation forecasting has always been a hot topic in hydro-meteorological research, with the issue of low accuracy and reliability needing urgent solutions. This study focuses on the Dongjiang Basin and evaluates the prediction accuracy and stability of CFSv2 model products at the monthly scale using the anomaly coefficient of correlation (ACC), normalized root mean square error (NRMSE), mean absolute error (MAE), and the multi-model stability index (MSI). Two methods, namely the CFSv2 model precipitation forecast and the machine learning model forecast combined with CFSv2 model predictors, are employed to predict future precipitation. The results show that under different lead times, the CFSv2 model precipitation forecast exhibits a high correlation with observed precipitation, performing better during the dry season compared to the flood season. However, there is significant variability in precipitation forecasts and poor model stability with

* 收稿日期: 2023-11-15 录用日期: 2024-03-13 网络首发日期: 2024-04-24

基金项目: 国家自然科学基金(52079151, 521111540261)

作者简介: 庄胜杰(2000年生), 男; 研究方向: 水文气象预测;

E-mail: zhuangshj3@mail2.sysu.edu.cn

通信作者: 王大刚(1975年生), 男; 研究方向: 极端气候变化、水文气象预测;

E-mail: wangdag@mail.sysu.edu.cn



ZR20230016

changes in the initial time. Combining CFSv2 model predictors with machine learning models improves the forecast stability, reducing the MSI from 0.45 to 0.25 and effectively reducing the randomness in forecasts caused by changes in the initial time. The findings contribute to providing a new approach for mid to long-term precipitation forecasting and offer decision-making support for mid to long-term hydrological forecasting and water resource management.

Key words: CFSv2; mid to long-term forecast; machine learning; product evaluation

降水作为水文循环的重要组成部分,一直是世界各国科学家们关注的焦点,暴雨带来的灾害对人类生活产生着重大的影响(甄亿位等,2015),降水的准确预报对水资源管理、水安全、旱涝灾害预警预防以及粮食管理安全等十分重要(冯志州,2017)。在全球变暖的背景下,极端气候事件的强度和频率不断增加(Guo et al.,2020),根据中长期水文预报做好水量调度工作以及水资源的调配管理是应对重大干旱洪涝事件的重要手段(刘金凤等,2018),提高中长期水文预报效果的关键在于改善中长期降水预报的质量和稳定性。

然而,中长期降水预报一直是气象预测的重点和难点。目前,降水预报主要参考数值预报模型和基于数据驱动的预报模型(马浩等,2022),但这两种方法均在一定程度上存在不足。由于大气运动的高度非线性和机理认识的限制,数值模式预报存在误差(Wang et al.,2011);数据驱动模型,则无法体现降水影响因子与降水的关系,随着气候条件发生改变,某些高相关影响因子的相关性会显著降低,甚至变得不相关(赵俊虎等,2015;谢舜等,2022)。近年来,全球气候模式(GCM, global climate model)已经成为气候模拟和预测的重要工具,但对降水的量级预测仍然存在不足(Lee et al.,2011),随着气候模式预测产品的日益丰富,CFSv2(coupled forecast system model version 2)、GEFSv12(global ensemble forecast system version 12)等动力模式结果逐渐成为降水预报的重要参考(黄赛男等,2022)。美国国家环境预测中心(NCEP, national centers for environmental prediction)最新的全耦合业务预报系统CFSv2,于2011年进入业务化运行,在我国各级气象部门得到广泛应用(Luo et al.,2013),给出的月、季预测结果具有一定可信度。Guo et al.(2020)通过CFSv2模式降水预报对中国东部夏季降水进行评估,发现其对华南地区短历时强降水预报效果较差,主要偏差来源有海气耦合模式以及预报分辨率过粗的因素;马浩等(2022)通过评估CFSv2对2018年

浙江梅雨降水多尺度预测的性能,发现CFSv2预报产品表现出显著的系统性干偏差,在延伸期尺度上,随着预报时效的缩短,预报效果并非逐步提升,而是存在一个最佳预报时效;Li et al.(2022)通过CFSv2模型输出数据对中国南方地区5月降水进行预测评估,发现由于CFSv2对相关异常环流的捕捉能力较差,导致模型在预测降水及环流因子年际变化方面表现不佳;张东方等(2021)通过CFSv2模式对成都市的延伸期降水预测进行测试,并通过改进的消除偏差一元线性回归方法进行订正,使得预测技巧得到提高。

CFSv2模式降水预报已广泛应用于季节性降水预报,短临降水预报,极端降水预报等多个领域(Lang et al.,2020)。然而,对于月尺度中长期预报的评估仍然较少,且较少的研究关注大气环流因子的预报效果以及不同起报时间对预报因子稳定性的影响,这些都对未来降水的预测起着重要的作用。因此,本研究不仅评估CFSv2模式月尺度降水和预报因子的预报效果,同时聚焦于不同起报时间对各预报因子以及降水的影响,通过建立预报因子与降水之间的非线性关系,从直接预报与间接预报两方面探究降水的中长期预报效果,结合标准化均方根误差(NRMSE, normalized root mean square error)、距平相关系数(ACC, anomaly coefficient of correlation)、平均绝对误差(MAE, mean absolute error)、多模型稳定性指数(MSI, multi-model stability index)等评价指标,客观描述不同预测方式的优劣,得到更具科学性的中长期降水预报结果。

1 研究区域、数据与方法

1.1 研究区域

东江为珠江的一级支流,流域总面积约3.5万 km^2 ,其干流全长约560 km,流域属亚热带季风气候,年均气温约为 $21\text{ }^\circ\text{C}$,年均降雨约为1 800 mm(董满宇等,2013;冯志州,2017)。东江承担香港、广州、深圳、东莞、惠州、河源近5 000万人口的供

水任务,同时兼有防洪、发电、航运、压咸、纳污等多种任务(吴裕珍,2018)。从2008年开始实施的水量调度对发挥东江流域水资源综合效益、保障供水安全、保护生态环境起到了重要作用。然而,受限于降水来水预报的精度,水量调度的效果仍有大幅提升的空间。因此,探索东江流域的中长期降水预报方法,可为开发东江流域降水和来水业务预报模型提供理论基础,有利于进一步提升水量调度效果,更好地保障区域水安全。

1.2 数据

采用东江流域雨量站的降水观测数据,ERA5再分析数据以及CFSv2模式预报数据进行模型搭建与分析。雨量站降水观测数据来源于惠州水文局整编后的东江流域内共73个降水站点,1979年1月至2023年4月逐月降水资料,雨量站分布见图1。

ERA5是由欧洲中期天气预报中心(ECMWF, European centre for medium-range weather forecasts)发布的一套再分析数据集,时间跨度为1950年1月至今,空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$,用于对CFSv2模式与降水相关的预报因子进行评估(陈柯兵等,2020)。预报因子选取1979年1月至2023年3月925、850、700、500和200 hPa共5个高层的经向风(uw, u-component of wind)、纬向风(vw, v-component of wind)、相对湿度(rhu, relative humidity)、位势高(geo, geopotential)以及温度(tem, temperature)和海平面气压(sp, surface pressure)共26个变量,分别用“变量_高度层”(如rhu_

500)形式表示。模式预报数据来自CFSv2实时预报产品,分别为月尺度气象因子预报和降水率预报,资料由CFSv2实时预报网站提供(<https://nomads.ncep.noaa.gov/pub/data/nccf/com/cfs/prod>),时间跨度为2011年4月至今,空间分辨率为 $1^{\circ} \times 1^{\circ}$ 。CFSv2通过将历史观测数据同化到CFS模式后,得到CFSR(coupled forecast system reanalysis)再分析产品,确保了气候记录的连续性,时间跨度为1979年1月至2011年3月,空间分辨率为 $0.5^{\circ} \times 0.5^{\circ}$ (Suranjana et al.,2014)。CFSv2共有24个成员的集合预报系统,预报成员预见期均达到9个月(肖颖等,2023)。

通过泰森多边形方法将雨量站数据处理成东江流域面平均逐月降水序列,其中1979年1月至2021年12月降水序列作为历史降水模拟的评估值,2022年4月至2023年3月降水序列作为未来降水模拟的评估值。由于月尺度时间跨度较大,暂未考虑各气象因子与研究区降水序列之间的遥相关关系以及空间异质性。通过区域裁剪方式,将ERA5气象因子与CFSv2模式预报因子处理为研究区域的面平均序列,共26个特征变量。为评估不同预见期下的模式预测情况,历史评估阶段研究选取月尺度气象因子预报和降水率预报的起报时间均为每月1日。将模式预测的大气环流因子与降水率分别处理成当月预报(M0)和预见期为1~9月的预报序列(M1-M9),共10组;未来降水模拟阶段,从实际业务需要的角度出发,起报时间分别定为每月1、11和21日。其中降水率处理成月降水总量,作为后续研究的基础数据。

1.3 方法

1.3.1 建模方法 由于月尺度的降水时间序列样本量较少,且各特征贡献值无法评定,为减少建模过程中机器学习算法的随机性影响,采用支持向量机和决策树集成两种不同的机器学习方法建立预报因子与降水之间的非线性关系,以此作为和产品预估的结果进行比较。

支持向量回归(SVR, support vector regression)是一种基于支持向量机的回归方法,在处理小样本、高维和非线性问题上具有优势。它通过构造一个或一组超平面来使数据与其距离最小化,从而进行回归预测(刘雅云,2018)。决策树是一种常用的分类和回归算法,但在某些情况下容易出现过拟合的问题(Breiman,2001),随机森林(RF, random forest)回归算法通过采用随机采样和随机特征

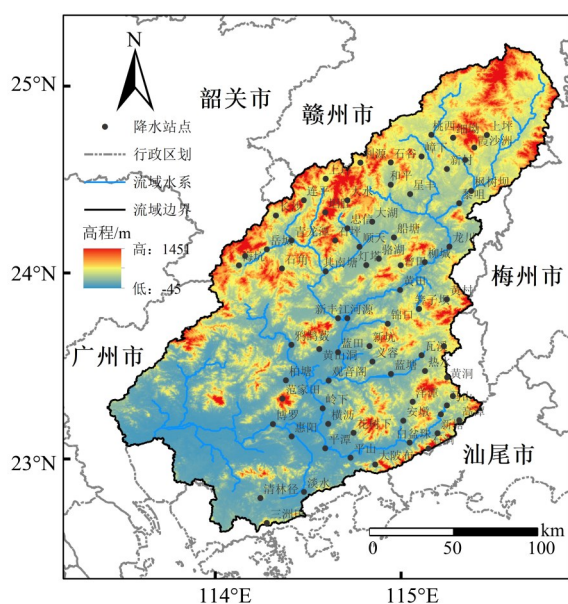


图1 东江流域研究区概况

Fig. 1 Overview of the Dongjiang Basin

选择的决策树集成算法,对训练集进行重复随机采样建立多个决策树,并将多个决策树的预测结果取平均作为最终预测结果,从而达到减少预测误差的目的(Peng et al.,2020)。

1.3.2 评价方法 为评判模型预测的优劣,采用距平相关系数(ACC)、标准化均方根误差(NRMSE)以及平均绝对误差(MAE)作为预报评价指标。通过计算预测值和观测值之间的ACC,可以确定预测模型是否能够准确地反映出观测数据的变化趋势,NRMSE用于评估预测结果的相对误差大小,通过标准化消除不同观测结果之间的量纲差异,MAE用于评估预测结果的绝对误差大小,确定预测模型的平均预测误差。具体公式为

$$\text{ACC} = \frac{\sum_{i=1}^n (\Delta y_i - \bar{\Delta y})(\Delta x_i - \bar{\Delta x})}{\sqrt{\sum_{i=1}^n (\Delta y_i - \bar{\Delta y})^2} \sqrt{\sum_{i=1}^n (\Delta x_i - \bar{\Delta x})^2}},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|,$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}}{y_{\max} - y_{\min}},$$

式中 $\Delta x_i = x_i - \bar{x}$; $\Delta y_i = y_i - \bar{y}$; x_i 为模型预测值; y_i 为实测值; \bar{x} 为预测值均值; \bar{y} 为实测值均值; y_{\max} 为实测最大值; y_{\min} 为实测最小值; n 为总样本数。

对于不同模型预测的稳定性分析,采用多模型稳定性指数(MSI)衡量多个模型对同一预报因子同一时间点预测的一致性程度。多模型定义为不同起报时间对同一时间序列的预测模型。设有 k 个

模型,每个模型对于某个时间点的预测结果为 $x_{i,t}(i=1,\dots,k)$,首先,通过给定的某个时间点 t 的预测结果,得到该时间点预测结果的多模型均值

$$\text{mu}_t = \sum_{i=1}^k x_{i,t}/k.$$

模型在时间 t 的变异系数 CV_t (coefficient of variation)

$$\text{CV}_t = \sqrt{\frac{\sum_{i=1}^k (x_{i,t} - \text{mu}_t)^2}{k}} / \text{mu}_t.$$

在实际计算过程中,由于预报经向风速与纬向风速方向的不稳定性, mu_t 值可能趋近于 0,会导致 CV_t 异常偏大,因此我们计算所有时间点上的 CV 值并对标准差大于 5 倍均值的数据进行剔除;最后,对剩下时间点上 CV 求取平均值,得到多模型稳定性指数

$$\text{MSI} = \frac{1}{m} \sum_{i=1}^m \text{CV}_i,$$

式中 m 为剔除异常值后的样本数。

该指标考虑多模型对时间序列单个时间点预测值的差异,并结合了变异系数的概念,可以评估多个模型的预测结果的一致性和稳定性。

2 研究结果

2.1 历史时期产品预测技巧分析

考虑 CFSv2 产品的发布时间与最长预报时效,选取 2012 年 1 月至 2021 年 12 月作为历史时期产品预测技巧分析的时段,通过对回报值的分析,评估产品的预测技巧。图 2 显示了不同预见期的 CFSv2 模式降水回报结果,对于东江流域全区降水,CFSv2 对整个流域的变化趋势有着较好的预估,不同预见期的降水回报值均在一定程度上反

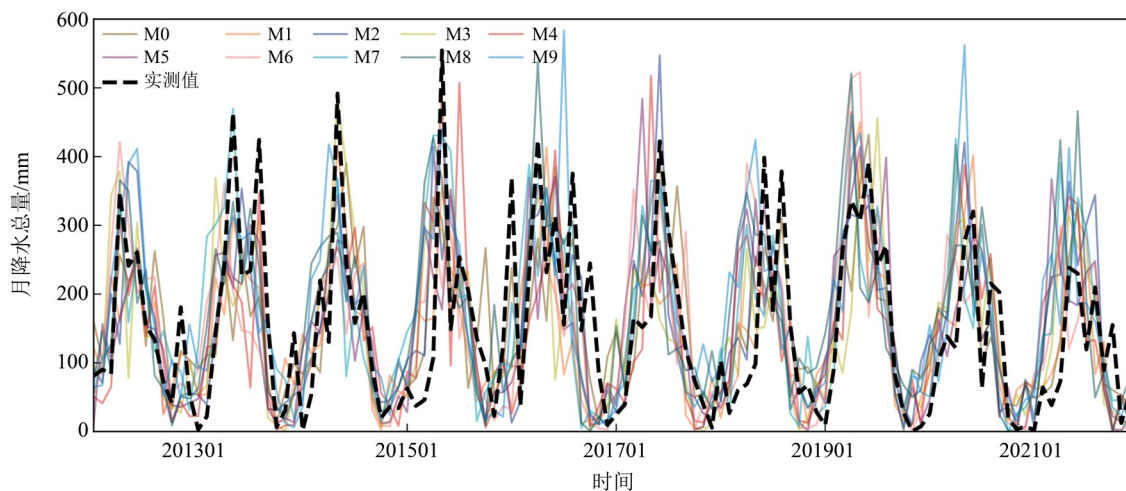


图2 不同预见期CFSv2模式降水回报结果

Fig. 2 CFSv2 model precipitation forecast results with different lead times

映出东江流域的汛枯状况, 但对于同一时间点的预估, 不同预见期的回报值存在较大的差异。整体上, 随着预见期变长, 模式降水预测能力略降, 但NRMSE与ACC的变化不大, 对预见期的变化不敏感。

结合东江水量调度工作的实际需求, 从水文年(4月~次年3月)、汛期(4~9月)、枯水期(10月~次年3月)3个时间尺度进行分析(刘金凤等, 2018)。由CFSv2模式降水不同预见期回报结果在汛期与枯水期的表现可知(图3), 在枯水期,

CFSv2模式降水回报值较实测值偏大, 整体上存在一定的高估现象; 在汛期, 多数时段的降水回报值较实测值偏低, 且随着预见期的延长, 模式预测的偏差略微增大, M0降水回报值与实测降水量最为相似。图4表明了不同预见期CFSv2模式降水对于不同时间尺度的回报效果, 枯水期平均NRMSE=0.2, ACC=0.68; 汛期NRMSE=0.24, ACC=0.50, 对于整个水文年的预测效果均好于分期预测。

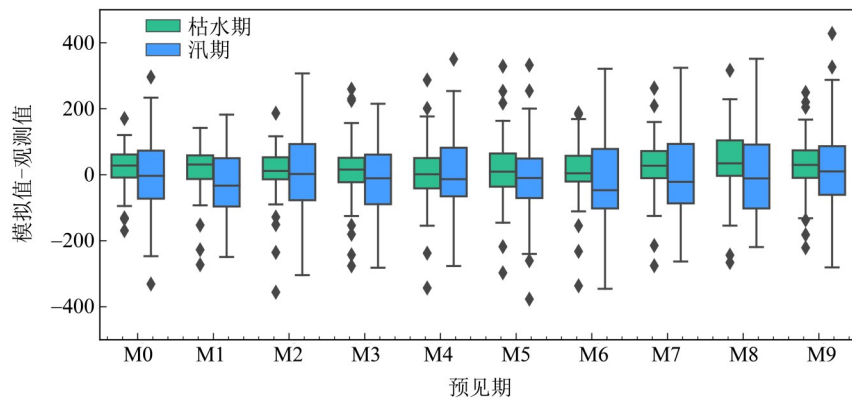


图3 不同预见期枯水期与汛期回报与观测的差值箱形图(黑色菱形点代表超过95%分位数)
Fig. 3 Boxplots of the difference between forecast with different lead times and observed precipitation during dry and flood season (black diamond points represent values exceeding the 95th percentile)

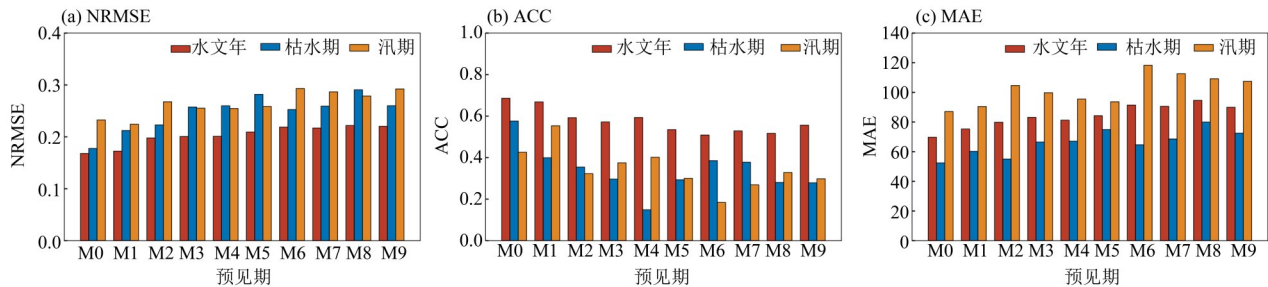


图4 水文年、汛期、枯水期的指标评价
Fig. 4 Evaluation results for hydrological year, flood season, and dry season

上述结果表明, 对于不同的预见期CFSv2模式降水回报值相关性变化不敏感, 从时间尺度上来说, 枯水期的回报效果略好于汛期, 预测技巧略高, 这可能是因为汛期不同月份的降水差异较大, 导致预测的MAE普遍较大。

进一步分析CFSv2模式对气象因子的预测技巧表现, 结果如图5所示, 各预报因子均显出较好的回报水平。M0预报的各气象因子平均NRMSE为0.16, 平均ACC为0.84, 回报效果均好于较长预见期; 对于不同预报因子, 表现较好的有geo、tem、sp与uw, 只有rhu_500与vw_500、vw_700

的预测水平较差。进一步分析可得, 不同预见期的回报效果在时间尺度上拟合趋势并未产生很大的差异, 但不同起报时间对同一时间点的回报值却有很大的差异, 导致在不同时间运行模型产生的不确定性较大。

对于所有变量, 从不同预见期以及不同预报变量两个角度, 分别计算平均ACC、NRMSE以及MSI值, 结果见表1和表2。从预见期来看, M0回报效果最好, 其平均NRMSE、ACC分别为0.16、0.84; 对于不同预报变量, 位势、温度以及海平面气压的回报效果较好, NRMSE、ACC均较其他变

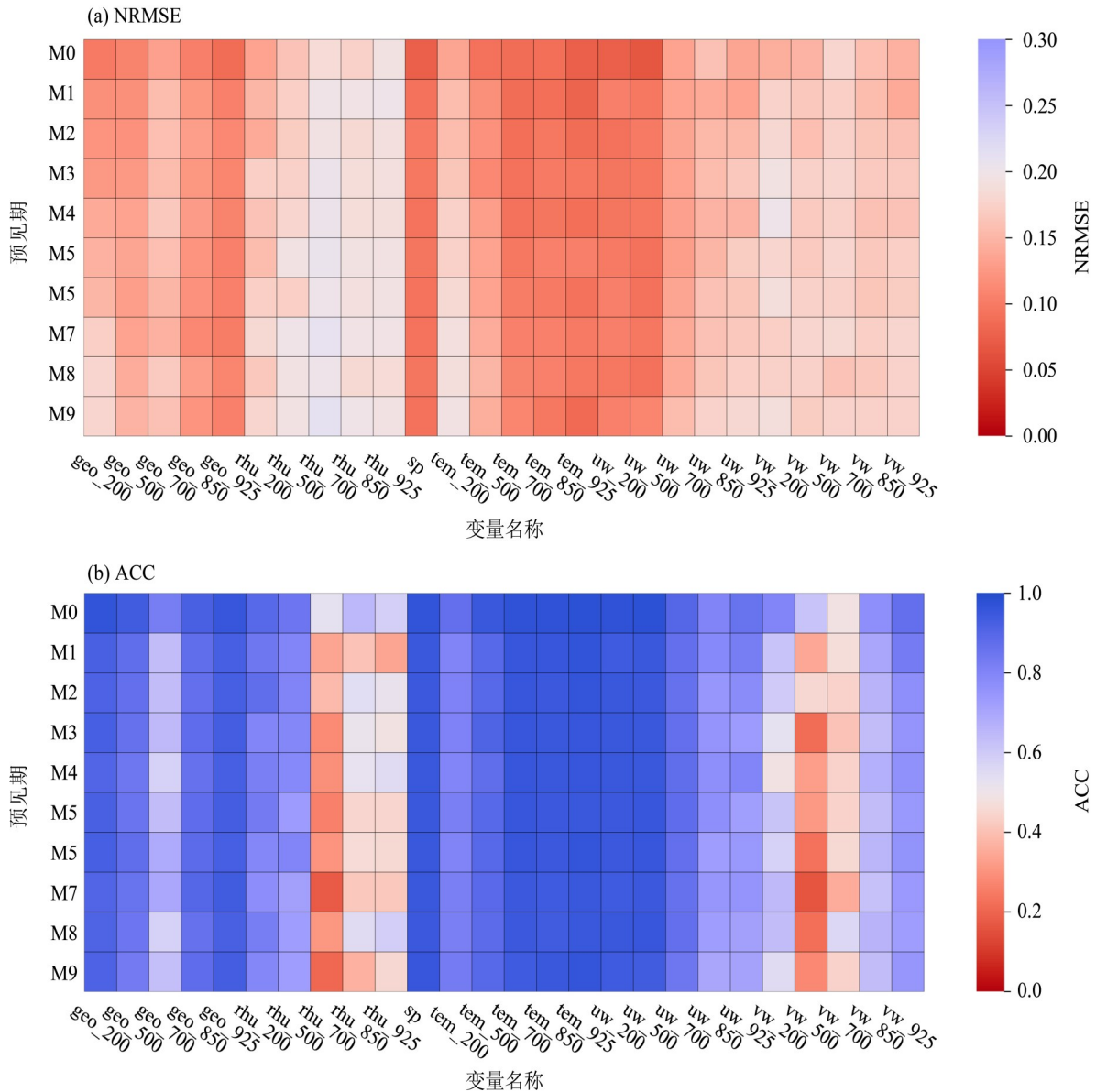


图 5 CFSv2模式对26个气象因子的预测技巧评价

Fig. 5 Evaluation of the forecasting skill for 26 meteorological variables by the CFSv2 model

表 1 不同预见期所有变量平均值

Table 1 Averaged values of statistics NRMSE and ACC for all variables at different lead time

| 预见期 | NRMSE | ACC | 预见期 | NRMSE | ACC |
|-----|-------|------|-----|-------|------|
| M0 | 0.16 | 0.84 | M5 | 0.22 | 0.65 |
| M1 | 0.18 | 0.80 | M6 | 0.19 | 0.79 |
| M2 | 0.20 | 0.75 | M7 | 0.20 | 0.76 |
| M3 | 0.19 | 0.78 | M8 | 0.19 | 0.79 |
| M4 | 0.21 | 0.75 | M9 | 0.22 | 0.72 |

量较好，MSI趋近于0，不同起报时间模型的稳定性也较为出色；从模型预测精度指标来看，各气象因子的NRMSE、ACC大多好于降水回报结果，

仅 rhu_700、vw_500、vw_700 三个指标相对较差；从模型预测稳定性指标来看，相对湿度、经向风速表现略差，MSI略高但整体上小于降水预报MSI，纬向风速稳定性效果最差，其MSI值均大于0.5。图6为部分变量对于不同起报时间回报结果的敏感性分析，对于纬向风速，大多时间点的CV值超过0.5，受不同起报时间影响最大，降水影响略好于纬向风速，但低于经向风速以及相对湿度，其受不同起报时间的影响也较大。

综上所述，通过历史时期产品预测技巧分析，CFSv2模型在预测东江流域降水趋势方面表现较好。M0、M1的回报效果最佳，随着预见期的延长，降水回报能力有所下降，枯水期的拟合效果

表 2 不同变量所有预见期平均值

Table 2 Averaged values of statistics NRMSE and ACC for different variables across all lead time

| 变量 | MSI | NRMSE | ACC | 变量 | MSI | NRMSE | ACC |
|---------|------|-------|------|---------|------|-------|------|
| geo_200 | 0.00 | 0.22 | 0.92 | tem_850 | 0.00 | 0.10 | 0.95 |
| geo_500 | 0.00 | 0.17 | 0.87 | tem_925 | 0.00 | 0.13 | 0.96 |
| geo_700 | 0.00 | 0.24 | 0.65 | uw_200 | 0.33 | 0.12 | 0.95 |
| geo_850 | 0.01 | 0.13 | 0.88 | uw_500 | 0.30 | 0.22 | 0.95 |
| geo_925 | 0.01 | 0.10 | 0.93 | uw_700 | 0.40 | 0.17 | 0.87 |
| rhu_200 | 0.22 | 0.15 | 0.83 | uw_850 | 0.48 | 0.22 | 0.75 |
| rhu_500 | 0.27 | 0.20 | 0.77 | uw_925 | 0.35 | 0.26 | 0.76 |
| rhu_700 | 0.12 | 0.29 | 0.31 | vw_200 | 0.63 | 0.28 | 0.61 |
| rhu_850 | 0.07 | 0.25 | 0.50 | vw_500 | 0.69 | 0.30 | 0.32 |
| rhu_925 | 0.07 | 0.26 | 0.50 | vw_700 | 0.60 | 0.24 | 0.45 |
| sp | 0.00 | 0.20 | 0.95 | vw_850 | 0.51 | 0.24 | 0.66 |
| tem_200 | 0.00 | 0.21 | 0.82 | vw_925 | 0.52 | 0.22 | 0.77 |
| tem_500 | 0.00 | 0.10 | 0.90 | pr | 0.47 | 0.20 | 0.57 |
| tem_700 | 0.00 | 0.10 | 0.95 | | | | |

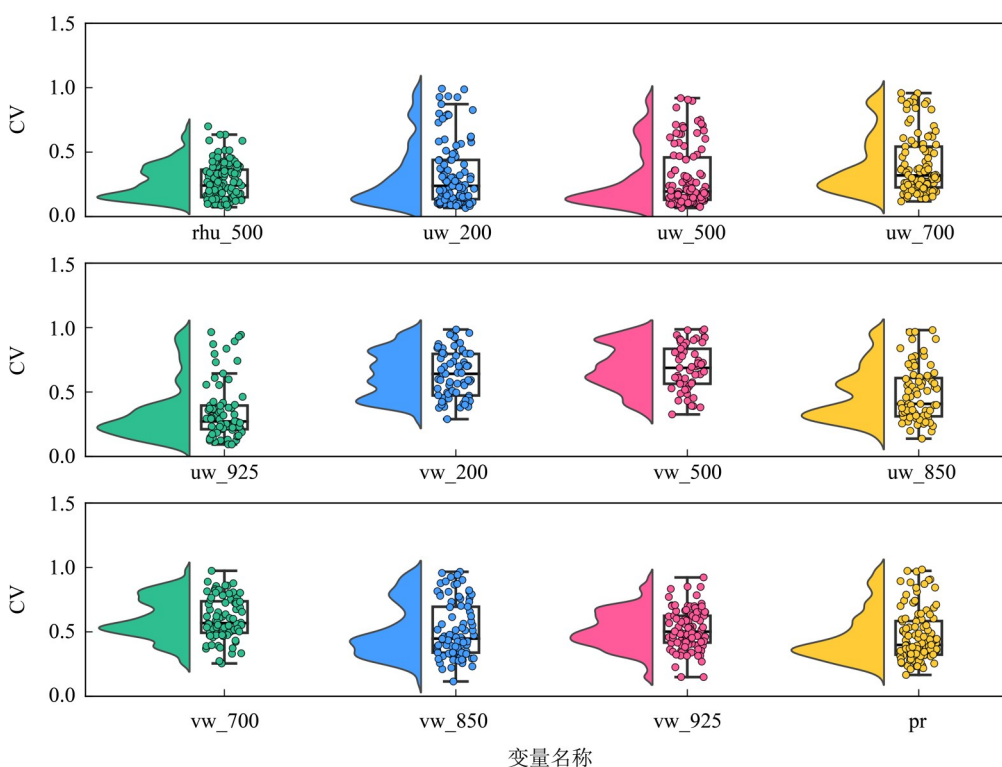


图 6 部分变量对于不同起报时间预测结果的敏感性分析

Fig. 6 Sensitivity analysis of selected variables with respect to forecast results at different initial times

稍好于汛期。此外, 不同起报时间对于同一时间点的回报值有着显著差异, 这带来了一定的不确定性。总体而言, CFSv2模式的预报结果具有一定的可信度, 但在具体应用中仍需考虑其局限性和不确定性。

2.2 结合机器学习方法的降水预测分析

结合历史阶段CFSv2对各气象因子以及降水的预测技巧评估结果, 对于各气压层温度、位势高和低气压层的相对湿度、经向风速、纬向风速以及海平面气压的回报效果好于直接的降水预报,

且这些气象因子的预测值随起报时间不同而产生的波动范围小于降水预报的波动范围。因此,在CFSv2在气象因子预报的基础上,引入机器学习模型,把预报的气象因子作为模型输入进行降水预报,是降水预报一个新思路。根据对气象因子的预测效果评价,所挑选的26个气象因子在不同预见期下的预测能力有所差异,其中 rhu_{700} 、 vw_{500} 以及 vw_{700} 的ACC分别为0.31、0.32和0.45,距平相关系数较低,因此,将剩下23个气象变量作为输入进行模型搭建。鉴于月尺度降水数据时间序列长度较短,暂不考虑时间滞后相关性与空间异质性,且预报因子的挑选较少,采用随机森林模型(RF)与支持向量回归模型(SVR)作为降水预测模型,模型训练期为1979年1月至2011年12月,验证期为2012年1月至2021年12月,模型参数设置见表3。

图7揭示了结合预报因子的机器学习模型与CFSv2模式降水预报(以M0为例)在历史时期上的模拟结果,RF模型在训练期NRMSE的平均值为0.06,验证期为0.13,SVR模型在训练期NRMSE

的平均值为0.06,验证期为0.15,模型具有较好的泛化性。对于距平较大的月份,M0模拟效果略优于RF模型与SVR模型,对极端降水值的预测较好,但在验证期其NRMSE的平均值为0.17,模拟效果略逊于机器学习模型。

为深入研究CFSv2模式降水预报(CFS模型)以及预报因子结合RF模型(RF-CFS模型)与SVR模型(SVR-CFS模型)对汛期以及枯水期预报效果之间的差异与优劣,选取2022年4月至2023年3月作为模型评价阶段,其中汛期起报月份为2022年2和3月,枯水期起报月份为2022年8和9月。图8为不同起报时间不同模型对汛期以及枯水期的预报效果。

对于2022年4~9月汛期预测,起报时间为2022年2月21日的CFS降水模式预报效果最好,其NRMSE与MAE分别为0.20、72 mm,均表现为最低水平,与实际降水偏差较少,但2月1日起报模式预测效果最差,NRMSE达到0.4,MAE高达164 mm;对于RF-CFS模型预报效果来说,其NRMSE为0.25~0.36,MAE为91~140 mm,ACC

表3 模型参数设置

Tabel 3 Model parameter setting

| 模型 | 参数范围 | 参数取值 |
|----------|---|------|
| 随机森林模型 | 决策树个数: {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} | 90 |
| | 决策树深度: {3, 4, 5, 6, 7, 8, 9, 10} | 9 |
| | 叶子节点的最小样本数量: {1, 3, 5, 7, 9} | 9 |
| | 分支节点的最小样本数量: {2, 4, 6, 8, 10} | 2 |
| 支持向量回归模型 | 核函数: {Linear, Rbf} | Rbf |
| | 惩罚系数: {0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256} | 128 |
| | 核函数系数: {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8} | 0.1 |
| | 容忍系数: {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8} | 0.8 |

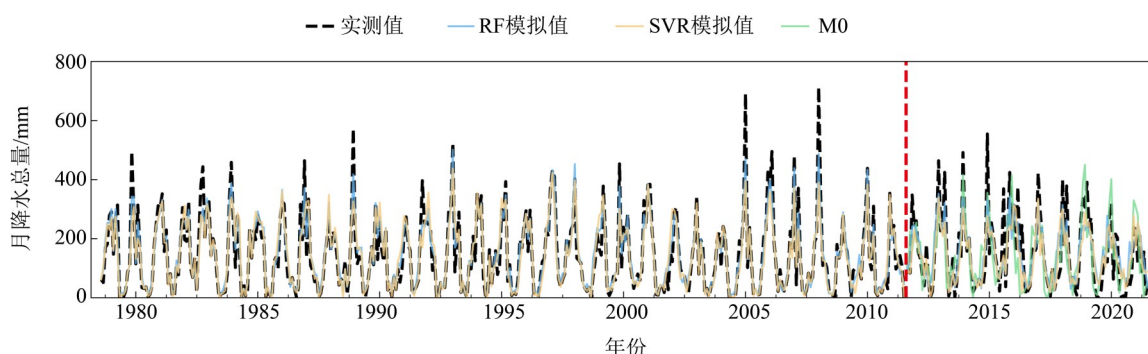


图7 机器学习模型历史降水模拟情况(红色竖虚线之前为训练期,之后为验证期)

Fig. 7 Historical precipitation simulation of the machine learning model

(The red dashed line represents the training period, and the subsequent period represents the test period.)

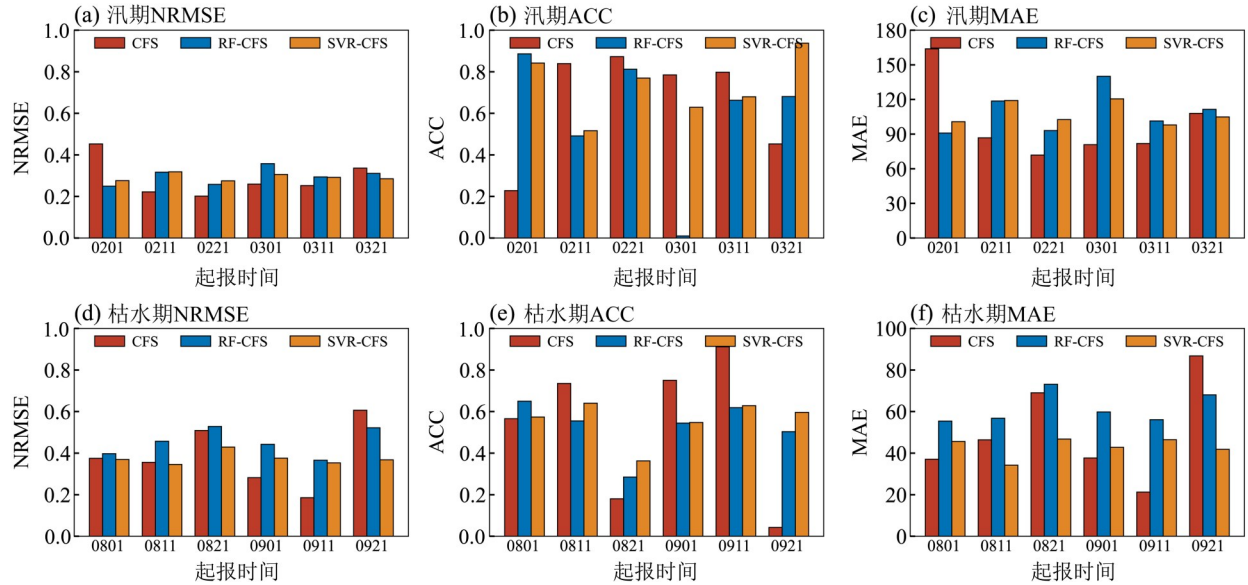


图 8 不同模型对 2022 年汛期和枯水期的预测指标评价

Fig. 8 Evaluation of indicators for the flood and dry season of the 2022 by different models

最大为 0.89, 最小值仅为 0.01, 除 3 月 1 日起报时间预报效果较差, 其余起报时间预报效果趋于平稳; SVR-CFS 模型在 3 种预报方式里表现最为平稳。对于枯水期的预报情况与汛期情况表现形势差别不大, CFS 降水模式预报起伏最大, 起报时间为 9 月 1 日的预报效果在所有模型里表现最好, NRMSE 仅为 0.18, MAE 为 21 mm, ACC 达到 0.91, 但起报时间为 9 月 11 日的预报效果与实际情况相差最大, NRMSE 与 MAE 为所有模型中的最高值, ACC 仅 0.18; 在预报稳定性方面, 预报因子结合机器学习模型预测的效果均较 CFS 降水模式预报效果稳定。

由不同预报方式对于 2022 年水文年的整体预报情况可知(图 9), CFS 降水模式预报 MSI 为 0.45, 而 RF-CFS 与 SVR-CFS 模型降水预报 MSI 分别为 0.31 与 0.25, 对于不同起报时间各模型的平均值在

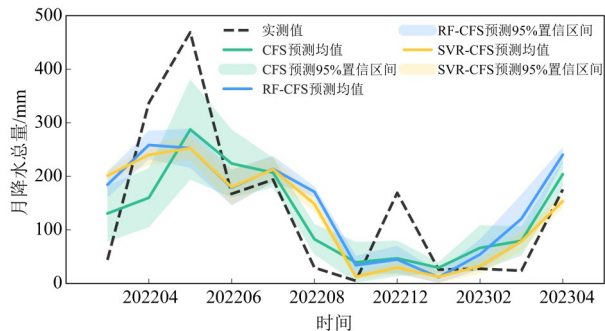


图 9 不同预报方式对于 2022 年水文年的整体拟合情况
Fig. 9 Overall fitting of different forecasting methods for the hydrological year of 2022

一定程度上能够预测实测降水情况, 但 CFS 降水模式预报的 95% 置信区间较 RF-CFS 与 SVR-CFS 有着明显的偏大, 其预测波动范围较大, 而 RF-CFS 与 SVR-CFS 的置信区间较小, 起报时间不同带来的影响较小。

综上所述, 仅通过 CFS 降水模式预报结果具有很大的不确定性, 各项评价指标的波动范围较大, 预报效果随起报时间改变的规律难以发掘; 通过构建大气环流因子与降水之间的非线性关系在一定程度上削减了由起报时间不同带来的预报结果的不稳定性, 但模型的预报精度还有待提高。

3 结 论

本文对 CFSv2 模式降水预报在东江流域预报效果进行评估, 通过统计水文年、汛期以及枯水期的 ACC、NRMSE 和 MAE 评估不同预见期模式预报效果的好坏。对于不同预见期, 模式预报与实测降水之间的相关性并无太大差异, 但不同起报时间对于同一时间点的降水量预测有着较大的差异, 因此提出 MSI 指标衡量不同起报时间预测降水量的差异性。同时, 发现 CFSv2 模式对于大气环流因子的预报效果好于降水预报, 因此建立大气环流因子与降水之间的非线性关系从而进行间接预报, 以此减少对于不同起报时间预测降水量的差异性。本文得出以下主要结论:

1) 随着预见期的变化, CFSv2 模式降水预测与实测降水之间的相关性变化不敏感, ACC 最小为 0.17, 最大为 0.22。从分期来看, 枯水期的预报

效果较好,相应的 NRMSE 为 0.2, ACC 为 0.68; 汛期预报效果相对较差,预报值的 NRMSE 为 0.24, ACC 为 0.50。

2) CFSv2 模式对多数大气环流因子预报效果好于降水预报,且对于不同起报时间,多数大气环流因子的稳定性更高。位势、温度、经向风速以及海平面气压的预测效果均优于降水预测效果,不仅体现在时间上高度相关,而且不同起报时间

产生的波动也较小。模式降水预报虽然有着较高的时间相关性,但对于起报时间的变化较为敏感,不同起报时间对同一时间的预报值相差较大。

3) 构建大气环流因子与降水之间的非线性关系在一定程度上降低了由起报时间不同带来的预报结果的不稳定性。2022 年 4 月至 2023 年 3 月,预报因子结合机器学习模型的降水预报置信区间远远小于 CFSv2 模式的降水预报置信区间。

参考文献:

- 陈柯兵,郭生练,王俊,等,2020. 长江上游 ECMWF 降水和径流预报产品评估[J]. 人民长江,51(3): 73-80.
- 冯志州,2017. 基于统计方法的东江流域季节降水预报研究[D]. 广州: 中山大学.
- 董满宇,王炳钦,廖剑宇,等,2013. 近 50 年东江流域极端降水事件变化特征[J]. 资源科学,35(3): 521-529.
- 黄超,李巧萍,谢益军,等,2022. 机器学习方法在湖南夏季降水预测中的应用[J]. 大气科学学报,45(2): 191-202.
- 黄赛男,李文韬,段青云,2022. GEFSv12 降水再预报数据在淮河流域的适用性评估[J]. 南水北调与水利科技(中英文),20(5): 925-934.
- 刘金凤,田兆伟,沈雪娇,2018. 近 60 年东江流域降雨径流特性分析[J]. 广东水利水电,(12): 31-36+74.
- 刘雅云,2018. 基于机器学习的降水数据分析算法的研究[D]. 南京: 南京信息工程大学.
- 马浩,于翡,葛敬文,等,2022. CFSv2 对 2018 年浙江梅雨降水多尺度预测的性能评估[J]. 气象科学,42(5): 690-702.
- 唐兆康,2021. 地基垂直观测网数据对数值预报的影响评估研究[D]. 南京: 南京信息工程大学.
- 吴裕珍,2018. 基于 CFSv2 的东江流域中长期降雨预报[D]. 广州: 中山大学.
- 肖颖,庞轶舒,马振峰,等,2023. NCEP CFSv2 模式对川渝夏季降水次季节预测技巧评估及预报偏差分析[J]. 高原气象,42(6): 1576-1588.
- 谢舜,孙效功,张苏平,等,2022. 基于 SVD 与机器学习的华南降水预报订正方法[J]. 应用气象学报,33(3): 293-304.
- 张东方,成青燕,何慧根,等,2021. CFSv2 模式资料在成都市延伸期降水预测中的应用评估[J]. 气象研究与应用,42(2): 1-6.
- 赵俊虎,封国林,2015. 全球变暖背景下中国东部夏季三类雨型预测概念模型新建[J]. 中国科学: 地球科学,45(4): 414-426.
- 甄亿位,郝敏,陆宝宏,等,2015. 基于随机森林的中长期降水量预测模型研究[J]. 水电能源科学,33(6): 6-10.
- BREIMAN L, 2001. Random forests[J]. Mach Learn, 45(1): 5-32.
- GUO Y, NIE H W, 2020. Summertime daily precipitation statistics over East China in CFSv2[J]. Phys Chem Earth, 115: 89-95.
- LANG Y, LUO L F, YE A Z, et al, 2020. Do CFSv2 seasonal forecasts help improve the forecast of meteorological drought over mainland China[J/OL]. Water, 12(7): 2010. <https://doi.org/10.3390/w12072010>.
- LEE S S, LEE J Y, HA K J, et al, 2011. Deficiencies and possibilities for long-lead coupled climate prediction of the western North Pacific-East Asian summer monsoon[J]. Clim Dyn, 36(5): 1173-1188.
- LI C Z, YANG S, MO W Q, et al, 2022. Seasonal prediction for may rainfall over Southern China based on the NCEP CFSv2[J]. J Trop Meteorol, 28(1): 29-44.
- LUO L F, TANG W, LIN Z H, et al, 2013. Evaluation of summer temperature and precipitation predictions from NCEP CFSv2 retrospective forecast over China[J]. Clim Dyn, 41: 2213-2230.
- PENG T, ZHI X, JI Y, et al, 2020. Prediction skill of extended range 2-m maximum air temperature probabilistic forecasts using machine learning post-processing methods[J]. Atmosphere, 11(8): 823.
- SAHA S, MOORTHY S, WU X, et al, 2014. The NCEP climate forecast system version 2[J]. J Clim, 27: 2185-2207.
- VALVERDE RAMÍREZ M C, de CAMPOS VELHO H F, FERREIRA N J, 2005. Artificial neural network technique for rainfall forecasting applied to the São Paulo region[J]. J Hydrol, 301(1/2/3/4): 146-162.
- WANG W Q, XIE P P, YOO S H, et al, 2011. An assessment of the surface climate in the NCEP climate forecast system reanalysis[J]. Clim Dyn, 37: 1601-1620.