

## 多流域-多预见期下 Stacking 径流预测的适用性评价及影响因素分析\*

陈嘉棋<sup>1</sup>, 江净超<sup>1\*\*</sup>, 江聪<sup>2</sup>, 黄经州<sup>1</sup>, 刘军志<sup>3</sup>

(1: 杭州电子科技大学自动化学院(人工智能学院), 杭州 310018)

(2: 中国地质大学(武汉)环境学院, 武汉 430074)

(3: 兰州大学泛第三极环境中心, 兰州 730000)

**摘要:** Stacking 作为一种先进的集成学习技术, 是提升径流预测性能的重要手段。现有 Stacking 径流预测研究多聚焦于个别流域及少数预见期的准确性评价, 缺乏针对在多流域与多预见期条件下 Stacking 径流预测的适用性评价及影响因素分析。本研究以支持向量回归(SVR)和随机森林(RF)作为个体学习器, 以岭回归作为元学习器, 构建了基于 Stacking 的径流预测模型。利用 CAMELS 数据集的 200 个流域, 以 1~7 天为预见期, 对 Stacking 径流预测的准确性、稳定性及适用性进行了评价, 并分析了其预测效果与流域特征及个体学习器精度的相关性。结果表明: (1) Stacking 在径流预测中的整体准确性和稳定性均高于个体学习器; (2) Stacking 对美国大陆多数流域的径流预测精度具有提升作用, 在降水量和气温较高的流域中提升效果更为显著, 而在降水量和气温较低的流域中效果相对有限; (3) Stacking 倾向于在个体学习器精度较低的流域提高预测精度, 而在个体学习器精度较高的流域则难以显著提高预测精度。研究结果可为集成学习方法在水文预测领域的应用与推广提供技术参考。  
**关键词:** Stacking; 多流域; 多预见期; 径流预测; CAMELS

## Applicability evaluation and influencing factors analysis of Stacking-based runoff prediction under multiple basins and multiple lead times\*

Chen Jiaqi<sup>1</sup>, Jiang Jingchao<sup>1\*\*</sup>, Jiang Cong<sup>2</sup>, Huang Jingzhou<sup>1</sup> & Liu Junzhi<sup>3</sup>

(1: School of Automation (School of Artificial Intelligence), Hangzhou Dianzi University, Hangzhou 310018, P.R.China)

(2: School of Environmental Studies, China University of Geosciences, Wuhan 430074, P.R.China)

(3: Center for the Pan-Third Pole Environment, Lanzhou University, Lanzhou 730000, P.R.China)

**Abstract:** Abstract: As one of the most advanced ensemble learning techniques, Stacking serves as an important approach to enhance the performance of runoff prediction. Existing studies on Stacking-based runoff prediction have mostly focused on accuracy evaluation under a small number of basins and limited lead times, while the applicability evaluation and analysis of influencing factors under multiple basins and multiple lead times remain unexplored. In this study, Support Vector Regression (SVR) and Random Forest (RF) were employed as base learners, and Ridge Regression was used as the meta-learner to construct runoff prediction models based on Stacking. Taking 200 basins from the CAMELS dataset as the study area and 1~7 days as the lead times, the accuracy, stability, and applicability of Stacking-based runoff prediction were systematically evaluated, and the correlations between the effectiveness of Stacking and both basin characteristics and the accuracy of base learners were analyzed.

\* 2025-09-10 收稿; 2026-04-01 收修改稿。

国家自然科学基金项目(42188102, 42171132)联合资助。

\*\* 通信作者; Email: jiangjc@hdu.edu.cn。

The main results are as follows: (1) The overall accuracy and stability of Stacking are higher than those of the individual base learners. (2) Stacking can improve the accuracy of runoff prediction in most basins across the continental United States; the improvement effect is more significant in basins with heavy precipitation and high temperature, but relatively limited in basins with light precipitation and low temperature. (3) Stacking tends to improve prediction accuracy in basins where the accuracy of individual base learners is low, but it is difficult to enhance prediction accuracy in basins where the accuracy of individual base learners is high. This study can provide a reference for the application of Stacking in runoff prediction.

**Keywords:** Stacking; multiple basins; multiple lead times; runoff prediction; CAMELS

径流预测在防洪抗旱与水灾害管理、水资源优化配置、农业生产保障、生态环境保护以及水力发电运营等方面发挥着关键的决策支撑作用, 对社会经济发展具有重大意义<sup>[1-6]</sup>。机器学习凭借其强大的拟合能力, 能够有效挖掘水文气象数据与径流之间的复杂关系, 实现准确的径流预测<sup>[7-9]</sup>。机器学习已成为径流预测的重要工具, 多种机器学习方法在径流预测当中得到广泛应用。例如, 崔东文<sup>[10]</sup>基于反向传播 (Back-Propagation, BP) 神经网络构建径流预测模型, 应用于新疆伊犁河雅马渡站的日尺度径流预测, 结果表明, 多隐层 BP 神经网络模型具有算法稳定、预测精度高的优势。Pham 等<sup>[11]</sup>利用随机森林 (Random Forest, RF), 对太平洋西北水文区 86 个流域进行日尺度径流预测, 发现 RF 在融雪驱动流域表现更佳。郭俊等<sup>[12]</sup>应用改进的支持向量回归 (Support Vector Regression, SVR) 于长江三峡流域日尺度径流预测, 证实其改进方法可提升预测精度。范宏翔等<sup>[13]</sup>利用长短期记忆 (Long Short-Term Memory, LSTM) 网络进行鄱阳湖流域日尺度径流预报, 结果表明, LSTM 能有效捕捉径流过程并反映其短期波动。袁旦等<sup>[14]</sup>利用改进 LSTM 对金沙江流域进行月尺度径流预报, 结果表明, 结合变分模态分解和优化算法帮助 LSTM 能够降低时间序列的非平稳性从而提高预报精度。

集成学习作为一种特殊的机器学习方法, 通过组合若干个体模型 (机器学习模型或机理模型) 的预测结果, 能够有效提升预测的稳定性和精度<sup>[15, 16]</sup>。近年来, 该技术在水文领域的径流预测中得到应用, 常用集成学习算法包括模型平均 (Model Averaging)、堆叠泛化 (Stacking)、装袋法 (Bagging) 和提升法 (Boosting) 等<sup>[17, 18]</sup>。其中, Stacking 采用独特的双层建模架构, 突破了传统集成学习的局限。它通过训练元学习器智能学习不同个体学习器预测结果之间的最优组合关系, 展现出更高的精度潜力, 因此受到水文研究者关注<sup>[19-26]</sup>。例如, 段雅楠等<sup>[19]</sup>利用 Stacking 集成了新安江模型 (Xinanjiang Model, XAJ)、水箱模型 (Tank Model, Tank)、萨克拉门托模型 (Sacramento Model, SAC)、土壤湿度计算及演进模型 (Soil Moisture Accounting and Routing Model, SMAR)、HYMOD 模型 (Hydrological Model) 和 GR4J 模型 (Génie Rural à 4 paramètres Journalier) 共 6 种水文模型, 应用于淮河北庙集流域和东河深渡流域的洪水径流预报, 提高了小时尺度径流预测精度。林泳恩等<sup>[20]</sup>采用 Stacking 方法, 从 SVR、多元线性回归 (Multiple Linear Regression, MLR)、LSTM、前馈神经网络 (Feedforward Neural Network, FFNN)、梯度提升回归树 (Gradient Boosting Regression Tree, GBRT)、自回归积分滑动平均模型 (Auto Regressive Integrated Moving Average Model, ARIMA) 和自适应增强算法 (Adaptive Boosting, AdaBoost) 中选择 2~7 个模型作为个体学习器, 并选择 MLR、SVR 或多层感知机 (Multilayer Perceptron, MLP) 作为元学习器, 构建了 Stacking 模型用于安墩水流域预见期 1~4h 的径流预测, 研究表明, 堆叠集成模型的预测效果与个体学习器的质量呈正相关, 而非其数量。Granata 等<sup>[21]</sup>利用 Stacking 集成 RF 和 MLP 作为个体学习器, 以弹性网络 (Elastic Net, EN) 作为元学习器, 在巴基廖内河 (意大利)、浣熊河 (美国爱荷华州)、威尔逊河 (美国俄勒冈州) 和特伦特河 (英国) 流域构建了预见期 1~3d 的日尺度径流预测模型, 结果显示, Stacking 与双向长短期记忆神经网络 (Bidirectional Long Short-Term Memory Neural Network, BiLSTM) 预测精度相当且计算效率更高。Granata 和 Di Nunno<sup>[22]</sup>以 MLP 和 RF 为个体学习器, EN、保序回归 (Isotonic Regression, IR)、径向基函数 (Radial Basis Function, RBF) 神经网络等算法作为元学习器, 探究其在泰河、泰菲河和泰斯河流域 (英国) 的中短期每日流量预报的有效性, 结果显示, 所有模型均表现出较低的相对误差, 其中 RBF 在预测低流量方面表现出色, 而 EN 在预测洪水流量方面表现出色。Wang 等<sup>[23]</sup>利用 Stacking 集成回声状态网络 (Echo State Network, ESN)、堆叠式自动编码器 (Stacked Autoencoder, SAE)、SVR 和 LSTM 作为个体学习器, 并采用卷积神经网络-双向长

短期记忆神经网络混合模型（Convolutional Neural Network-Bidirectional Long Short-Term Memory Neural Network, CNN-BiLSTM）混合模型作为元学习器，提高了漫湾水电站、峡江水文站和王家汇水文站流域的月尺度径流预测精度。

现有基于 Stacking 的径流预测研究侧重于在个别流域和少数预见期上，从不同个体学习器或元学习器的视角评价 Stacking 径流预测的准确性，而缺乏在多流域和多预见期上对 Stacking 径流预测的适用性和影响因素开展深入研究，仍有一些重要问题亟待解答：（1）Stacking 径流预测的适用性如何，是否在所有流域和预见期上都能提升径流预测精度？（2）Stacking 径流预测的效果是否与流域特征以及个体模型的精度相关？针对上述问题，本研究在 200 个流域和 1~7 天预见期内，选择在模型原理和数据敏感性等方面具有较大差异性和互补性的 SVR 和 RF 作为 Stacking 的个体学习器，选择稳定性高和泛化能力强的岭回归（Ridge Regression）作为 Stacking 的元学习器，分别构建了 1400 个基于 SVR 的径流预测模型、1400 个基于 RF 的径流预测模型以及 1400 个基于 Stacking 的径流预测模型，在此基础上，着重评价了 Stacking 径流预测的适用性以及分析了 Stacking 径流预测效果的影响因素，以期 Stacking 在径流预测中的应用提供参考依据。

## 1 数据资料

### 1.1 研究流域概况

本研究所用的流域数据来源于 Catchment Attributes and Meteorology for Large-sample Studies (CAMELS) 数据集<sup>[27]</sup>，该数据集由美国国家大气研究中心开发，是公开的水文研究数据集，涵盖了美国大陆 671 个受人类活动影响相对较小的盆地流域。本研究从中选取了 200 个流域作为研究区域，所选取流域覆盖了美国大陆的主要气候区和水文分区，具有较强的代表性，所选流域的空间位置分布如图 1 所示。

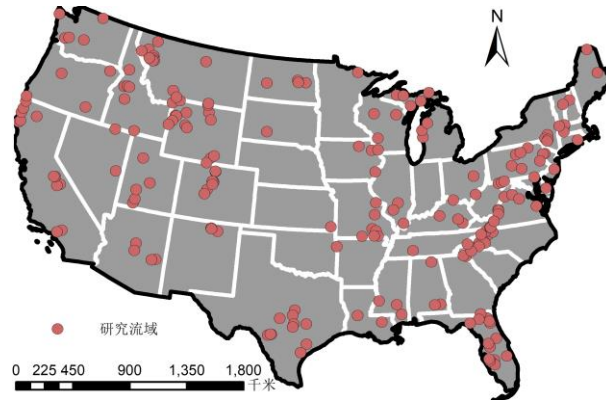


图 1 研究区流域空间位置分布图

Fig1. Spatial distribution map of the basins in the study area

### 1.2 研究流域数据及处理

本研究使用的数据时间跨度为 1995 年 1 月 1 日至 2004 年 12 月 31 日，包括日尺度气象数据和径流数据。其中，流域的日尺度气象数据来自于 Daymet 气象数据集<sup>[28]</sup>，包括降水量、最高气温和最低气温，流域的日尺度径流数来源于 CAMELS 数据集。将 1995 年 1 月 1 日至 2001 年 12 月 31 日作为训练期，相关数据用于模型参数训练和超参数选取；将 2002 年 1 月 1 日至 2004 年 12 月 31 日作为测试期，相关数据用于评价模型的预测性能。

为消除变量间量纲差异并提升模型预测性能，对每个流域的所有气象和径流数据采用最大最小值归一化方法进行标准化处理，其计算公式如下：

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

式中， $x'$  表示归一化后的数据； $x$  表示原始数据； $x_{max}$  和  $x_{min}$  分别表示当前流域数据所在训练期的最小值和最大值。需要说明的是训练集和测试集的气象及径流数据均使用训练期的最小值和最大值进行标准化处理。

## 2 研究方法

### 2.1 基于 SVR 的径流预测模型

支持向量机最初用于分类问题，随后被扩展到回归问题<sup>[29,30]</sup>，此类用于回归的方法被称为 SVR。SVR 的目标是找到一个回归函数  $f(x)$  来逼近目标值。其基本形式可表示为：

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (2)$$

式中， $\mathbf{w}$  为特征空间的权重向量； $\phi(\cdot)$  表示核函数映射，本研究采用 RBF 核函数（见后文公式(6)）作为将低维输入空间映射到高维特征空间的非线性变换函数； $b$  表示偏置项。SVR 表达函数可以通过引入松弛变量  $\xi_i$  和  $\hat{\xi}_i$  形式化为以下问题：

$$\min_{\omega, b, \xi_i, \hat{\xi}_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i)^2 \quad (3)$$

$$\begin{aligned} s.t. \quad & f(\mathbf{x}_i) - y_i \leq \delta + \xi_i \\ & y_i - f(\mathbf{x}_i) \leq \delta + \hat{\xi}_i \\ & \xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, \dots, m \end{aligned} \quad (4)$$

式中， $\mathbf{x}_i$  为第  $i$  个样本的输入特征； $y_i$  为第  $i$  个样本的径流观测值；惩罚参数  $C$  和不敏感损失参数  $\delta$  为 SVR 的超参数； $\xi_i$  和  $\hat{\xi}_i$  为非负松弛变量，用于允许模型存在少量预测误差，解决非线性不可分问题，满足  $\xi_i$ 、 $\hat{\xi}_i$  大于 0。对于优化问题(3)及其约束条件(4)，可通过拉格朗日乘子法可以得到 SVR 的解形如：

$$f(\mathbf{x}) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i)^2 \kappa(\mathbf{x}, \mathbf{x}_i) + b \quad (5)$$

式中， $\alpha_i$  和  $\hat{\alpha}_i$  是拉格朗日乘子，均为非负实数； $b$  是求解过程确定的偏置常数； $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$  为核函数，其作用是将样本映射到高维特征空间以处理非线性关系。本研究中使用 RBF 核函数，其计算公式如下：

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (6)$$

式中， $\gamma$  为核系数，是 RBF 核函数的重要超参数，控制着函数映射的宽度与复杂度； $\|\mathbf{x}_i - \mathbf{x}_j\|^2$  为样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  的欧氏距离的平方。

本研究对上述 200 个流域的 1~7 天预见期的日尺度径流预测模型分别建模，即为每个流域的每个预见期  $d(d=1,2,\dots,7)$  分别独立训练一个基于 SVR 的径流预测模型，共计 1400 个模型。对于预见期  $d$ ，训练数据定义为  $D_d = \{(\mathbf{X}_i, y_i)\}_{i=d+7}^N$ ，其中， $N$  为训练期天数， $\mathbf{X}_i$  代表第  $i$  天输入特征，其包括从第  $i-6-d$

天到第  $i$  天的降水量、最高温度、最低温度和  $i-6-d$  天到第  $i-d$  天的观测径流量,  $y_i$  代表第  $i$  天观测径流量。

本研究使用 Sklearn 库中 svm.SVR 方法实现基于 SVR 的径流预测模型, 采用网格搜索和 4 折交叉验证联合策略确定模型超参数 (惩罚参数  $C$ 、不敏感损失参数  $\delta$ 、RBF 核参数  $\gamma$ ) 的最优取值。需要说明的是, 在进行交叉验证时, 训练集当中的一部分数据充当验证集, 根据模型在验证集上的表现确定模型超参数, 测试集并未参与交叉验证过程, 因此规避了信息泄露风险。

## 2.2 基于 RF 的径流预测模型

随机森林本质上是一种基于多棵递归回归树的模型<sup>[31, 32]</sup>。本研究为每个流域的每个预见期分别构建一个基于 RF 的径流预测模型, 共计 1400 个模型, 模型的输入特征与基于 SVR 的径流预测模型保持一致。基于 RF 的径流预测模型的建模和预测步骤如下:

步骤 1: 从  $D_d = \{(X_i, y_i)\}_{i=d+7}^N$  中有放回地抽取  $m$  个自助样本子集  $\{S_1, S_2, \dots, S_m\}$ 。

步骤 2: 为每个子集  $S_k$  训练一棵回归决策树  $T_k$ , 共构建  $m$  棵子树。在节点划分时, 随机选择  $z$  个候选特征, 依据均方差最小化 (MSE) 选择最优分裂点。

步骤 3: 最终预测值  $\hat{y}$  为所有决策树输出的均值  $\hat{y} = \frac{1}{m} \sum_{k=1}^m T_k(P_k)$ , 式中  $P_k$  为第  $k$  棵树的输入;  $T_k(\cdot)$  为第  $k$  棵树的输出。

本研究使用 Sklearn 库中 ensemble.RandomForestRegressor 方法实现基于 RF 的径流预测模型, 采用网格搜索和 4 折交叉验证联合策略确定模型超参数 (子集数  $m$ 、候选特征数量  $z$ 、树最大深度、节点可分的最小样本数) 的最优取值。

## 2.3 基于 Stacking 的径流预测模型

Stacking 是一种高阶集成学习框架, 通过双层模型架构整合个体学习器的预测能力, 以提升预测的精度<sup>[33, 34]</sup>。Stacking 通常由不同的个体学习器和一个元学习器组成, 其中, 个体学习器利用原始特征生成多样化的初级预测结果; 元学习器则以这些结果为输入, 学习其与最终输出之间的映射关系。本研究为每个流域的每个预见期分别构建一个基于 Stacking 的径流预测模型, 共计 1400 个模型, 模型的输入特征与基于 SVR 的径流预测模型保持一致。基于 Stacking 的径流预测模型的实现步骤如下:

步骤 1: 首先, 采用 4 折交叉验证将完整数据集  $D_d$  按时间顺序划分为 4 个连续、互斥且长度基本一致的子集  $\{F_1, F_2, F_3, F_4\}$ 。然后, 在第  $j(j=1, 2, 3, 4)$  折上, 将  $D_d$  划分成  $F_j$  和  $D_d \setminus F_j$  分别代表该折的测试集和训练集。接着, 在第  $j$  折上, 采用与 2.1 和 2.2 中的相同超参数, 使用第  $j$  折的训练集  $D_d \setminus F_j$  训练基于 SVR 的径流预测模型  $M_{SVR,j}$  和基于 RF 的径流预测模型  $M_{RF,j}$ 。随后, 使用训练好的个体学习器  $M_{SVR,j}$  和  $M_{RF,j}$

对测试集  $F_j$  进行预测, 得到预测结果  $\hat{y}_{SVR,j,i} = M_{SVR,j}(X_i)$  和  $\hat{y}_{RF,j,i} = M_{RF,j}(X_i)$ , 其中  $\forall X_i \in F_j$ 。最后, 将所有折的预测结果按其原始时间索引顺序整合, 并且合并 SVR 和 RF 个体学习器的交叉验证预测结果,

得到元学习器的输入  $H_{meta} = [h_{d+7}, h_{d+8}, \dots, h_N] = \begin{bmatrix} \hat{y}_{SVR,d+7} & \hat{y}_{SVR,d+8} & \dots & \hat{y}_{SVR,N} \\ \hat{y}_{RF,d+7} & \hat{y}_{RF,d+8} & \dots & \hat{y}_{RF,N} \end{bmatrix}$ 。

步骤 2: 以  $H_{meta}$  作为输入特征, 对应实际径流观测值  $Y = [y_{d+7}, y_{d+8}, \dots, y_N]^T$  作为输出目标, 采用岭回

归作为元学习器  $M_{meta}$  进行训练。岭回归通过最小化目标函数(7)求解参数  $\beta$  :

$$\min_{\beta} \sum_{i=d+7}^N (y_i - \beta^T \mathbf{h}_i)^2 + \lambda \|\beta\|_2^2 \quad (7)$$

式中,  $\beta$  为回归系数向量,  $\lambda$  为正则化超参数。本研究采用网格搜索与 4 折交叉验证相结合的策略确定  $\lambda$  的最优取值。

本研究使用 Sklearn 库中 ensemble.StackingRegressor 和 linear\_model.RidgeCV 方法实现基于 Stacking 的径流预测模型。

#### 2.4 评价指标

本研究使用纳什效率系数 (Nash-Sutcliffe coefficient, NSE) 和均方根误差 (Root Mean Square error, RMSE) 指标评价径流预测精度, 其计算公式如下:

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (9)$$

式中,  $N$  为观测数据的数量;  $y_i$  表示第  $i$  个时刻的观测值;  $\hat{y}_i$  表示第  $i$  个时刻的预测值;  $\bar{y}$  表示观测数据的平均值。

### 3 结果分析

#### 3.1 Stacking 径流预测的准确性和稳定性评价

在 200 个流域的测试集上, 采用个体学习器 (SVR、RF) 和 Stacking 集成学习模型对 1~7 天预见期的径流进行了预测,  $NSE$  和  $RMSE$  箱线图如图 2 和图 3 所示。由图 2 和图 3 可知, 在 1~7 天预见期内, Stacking 的  $NSE$  中位数和平均值均最高, SVR 次之, RF 最低; Stacking 的  $RMSE$  中位数和平均值均最低, SVR 次之, RF 最高。采用 Wilcoxon 配对符号秩检验评估了 Stacking 的精度是否显著高于 RF 与 SVR。结果表明, 在 1~7 天预见期内, Stacking 的  $NSE$  显著高于 SVR 和 RF (P 值均小于 0.0001), Stacking 的  $RMSE$  显著低于 SVR 和 RF (P 值均小于 0.0001)。这表明 Stacking 的整体准确性要显著高于个体学习器。

如图 2 和图 3 所示, 随着预见期的延长, SVR、RF 以及 Stacking 的径流预测准确性都呈现出了下降趋势: 在  $NSE$  指标方面, SVR 的  $NSE$  中位数由 0.90 下降到 0.70,  $NSE$  平均值由 0.83 下降到 0.66; RF 的  $NSE$  中位数由 0.87 下降到 0.65,  $NSE$  平均值由 0.81 下降到 0.63; Stacking 的  $NSE$  中位数由 0.91 下降到 0.74,  $NSE$  平均值由 0.85 下降到 0.72。在  $RMSE$  指标方面, SVR 的  $RMSE$  中位数由  $5.36\text{m}^3/\text{s}$  上升到  $8.48\text{m}^3/\text{s}$ ,  $RMSE$  平均值由  $2.32\text{m}^3/\text{s}$  上升到  $4.15\text{m}^3/\text{s}$ ; RF 的  $RMSE$  中位数由  $5.82\text{m}^3/\text{s}$  上升到  $9.18\text{m}^3/\text{s}$ ,  $RMSE$  平均值由  $2.64\text{m}^3/\text{s}$  上升到  $4.87\text{m}^3/\text{s}$ ; Stacking 的  $RMSE$  中位数由  $5.05\text{m}^3/\text{s}$  上升到  $7.62\text{m}^3/\text{s}$ ,  $RMSE$  平均值由  $2.29\text{m}^3/\text{s}$  上升到  $4.00\text{m}^3/\text{s}$ 。从指标变化幅度来看, Stacking 的  $NSE$  下降幅度与  $RMSE$  上升幅度均为最小, SVR 次之, RF 最大, 这表明在径流预测稳定性方面, Stacking 具有比个体学习器更强的稳定性。此外, 三类模型的预测准确性与稳定性呈现出一致性规律: 预测准确性越高的模型, 其稳定性也越强。

#### 3.2 Stacking 径流预测的适用性评价

考虑到  $NSE$  和  $RMSE$  在评价模型径流预测精度方面具有一致性, 本研究基于  $NSE$  评价 Stacking 径流预测的适用性。对于给定的预见期和流域, 若 Stacking 的  $NSE$  同时高于 SVR 和 RF, 则称该流域为 Stacking 有效流域 (即 Stacking 具有适用性); 反之则称该流域为 Stacking 无效流域 (即 Stacking 不具有适用性),

此时 SVR 或 RF 的预测精度最高。

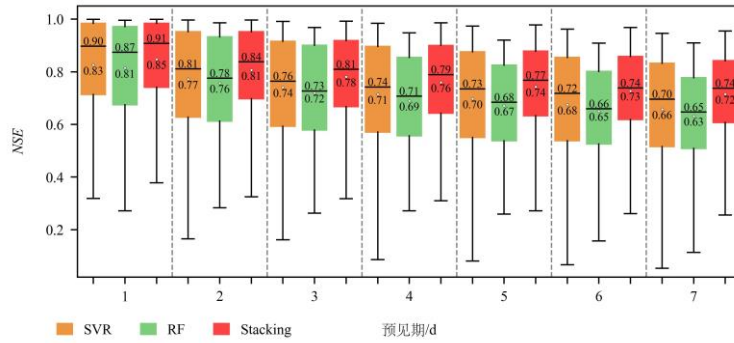


图 2 SVR、RF 以及 Stacking 在测试集上的径流预测 NSE 箱线图

Fig. 2 Box plots of NSE for runoff prediction by SVR, RF and Stacking on the test set

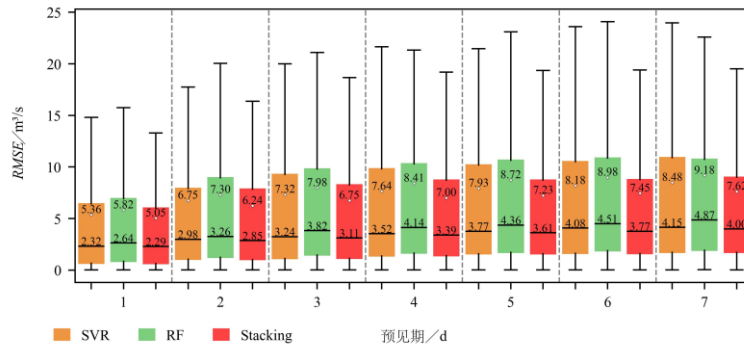


图 3 SVR、RF 以及 Stacking 在测试集上的径流预测 RMSE 箱线图

Fig. 3 Box plots of RMSE for runoff prediction by SVR, RF and Stacking on the test set

在 200 个流域的测试集上，分别确定出了 1~7 天预见期内的 Stacking 有效流域和无效流域，其空间分布和比例统计如图 4 所示。由图 4 可知，在 1~7 天预见期内，Stacking 有效流域的占比为 61%~67.5%，且 4~7 天预见期内的有效流域占比要高于 1~3 天预见期。在 1~7 天预见期内，Stacking 无效流域的占比为 32.5%~39%，其中，SVR 和 RF 精度最高的流域占比分别为 25%~31%、6%~10.5%。这表明，Stacking 虽然无法所有流域上都取得比个体学习器更高的预测精度，但在多数流域上提高了预测精度。Stacking 有效流域和无效流域的空间分布表明，Stacking 有效流域主要集中在美国大陆的东部区域，而 Stacking 无效流域主要集中在美国大陆的西部区域，呈现出较为明显的空间聚集性和分异性。

此外，Stacking 在 75 个流域的 1~7 天所有预见期均具有适用性，流域占比为 37.5%。上述 75 个流域的空间分布如图 5 所示。由图 5 可知，在 1~7 天所有预见期 Stacking 均适用的流域主要集中在美国大陆东部区域。

### 3.3 Stacking 径流预测效果与流域特征的相关性分析

在 1~7 天预见期内，Stacking 有效流域和无效流域的多年平均降水量的箱线图如图 6 所示。由图 6 可知，Stacking 有效流域的多年平均降水量中位数和平均值均在 1200mm 以上，而 Stacking 无效流域的多年平均降水量中位数和平均值均在 1000mm 左右。采用 Mann-Whitney U 检验评估了 Stacking 有效流域的多年平均降水量是否显著高于 Stacking 无效流域。结果表明，在 1~7 天预见期内，Stacking 有效流域的年均降水量显著高于 Stacking 无效流域 (P 值均小于 0.0001)。这表明，Stacking 倾向于在降水量较大的流域提高径流预测精度，而在降水量较小的流域则难以提高径流预测精度，这与图 4 中所体现出的 Stacking 有效流域大多分布在降水量更大的美国大陆东部区域是相一致的。

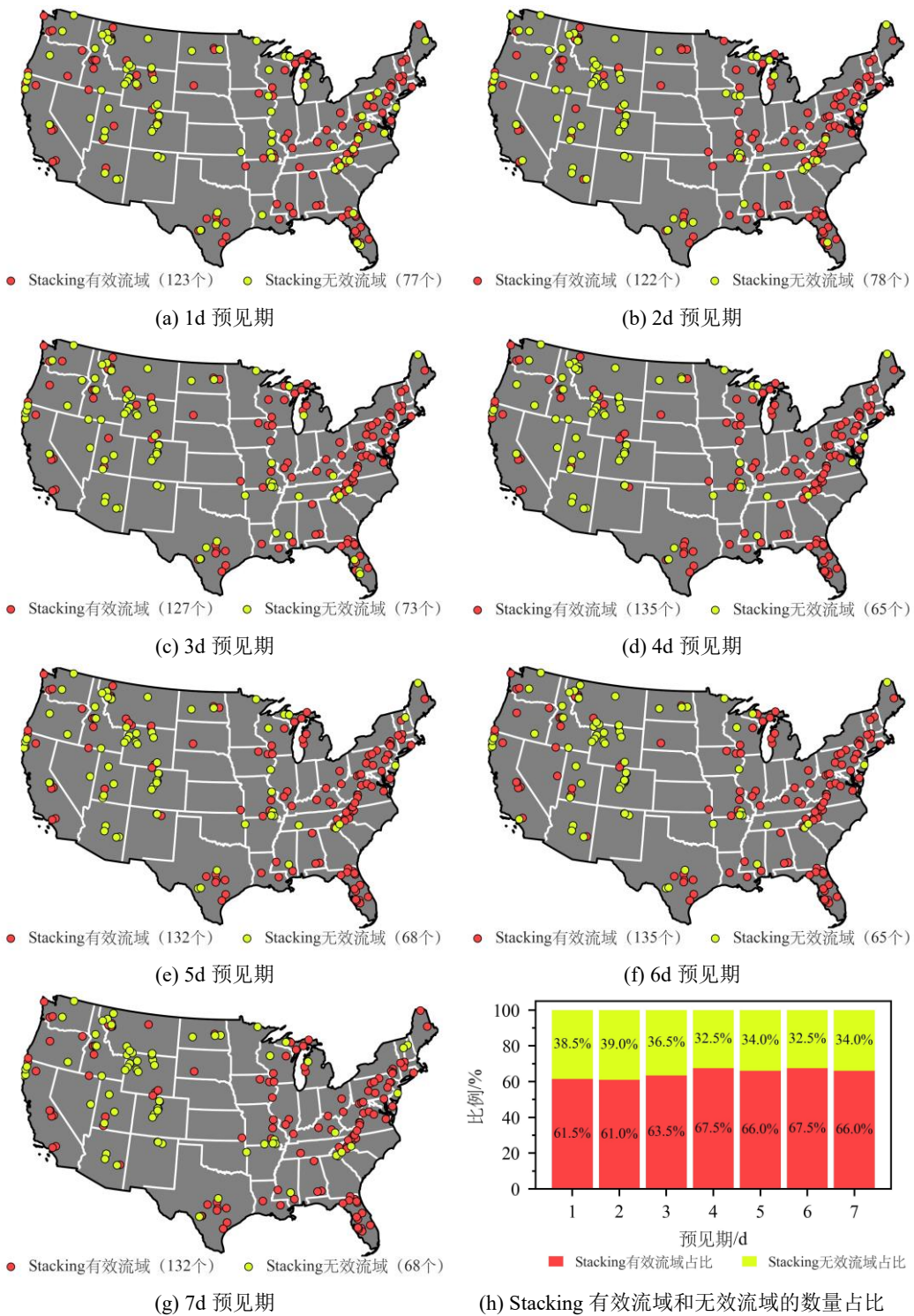


图 4 Stacking 有效流域和无效流域的空间分布及占比

Fig.4 Spatial distribution and percentage composition of Stacking-effective basins and Stacking-ineffective basins

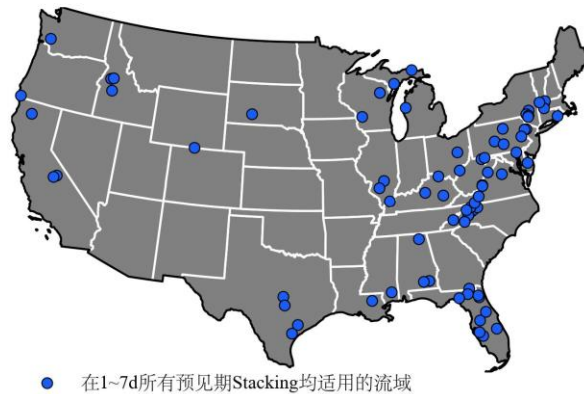


图 5 在 1~7d 所有预见期 Stacking 均适用的流域空间分布

Fig.5 Spatial distribution of basins with improved accuracy during all the lead times of 1~7 days

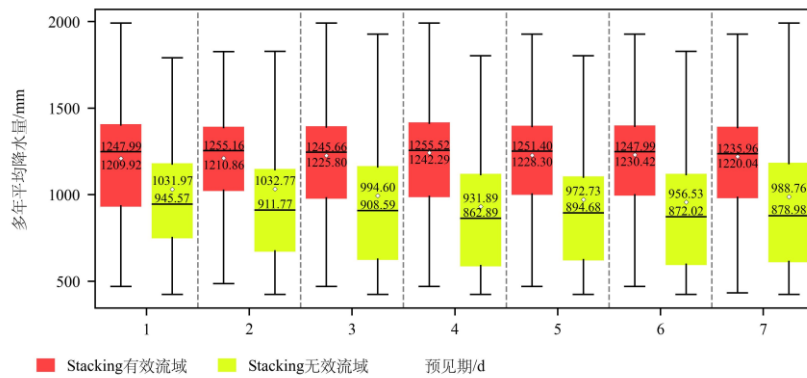


图 6 Stacking 有效流域和无效流域的多年平均降水量箱线图

Fig.6 Box plots of average annual precipitation for Stacking-effective basins and Stacking-ineffective basins

在 1~7 天预见期内，Stacking 有效流域和无效流域的日均气温（由日最高气温和日最低气温计算平均值得到）的箱线图如图 7 所示。由图 7 可知，Stacking 有效流域的日均气温中位数范围为 10.17°C~10.30°C，平均值范围为 10.50°C~11.13°C。Stacking 无效流域的日均气温中位数范围为 4.72°C~6.02°C，平均值范围为 6.04°C~7.63°C。采用 Mann-Whitney U 检验评估了 Stacking 有效流域的日均气温是否显著高于 Stacking 无效流域。结果表明，在 1~7 天预见期内，Stacking 有效流域的日均气温显著高于 Stacking 无效流域（P 值均小于 0.001）。这表明，Stacking 倾向于在气温较高的流域提高径流预测精度，而在气温较低的流域则难以提高径流预测精度。

在 1~7 天预见期内，Stacking 有效流域和无效流域的流域面积的箱线图如图 8 所示。由图 8 可知，Stacking 有效流域的流域面积中位数范围为 323.61km<sup>2</sup>~385.74km<sup>2</sup>，平均值范围为 766.59km<sup>2</sup>~870.59km<sup>2</sup>。Stacking 无效流域的流域面积中位数范围为 233.78km<sup>2</sup>~437.52km<sup>2</sup>，平均值范围为 740.23km<sup>2</sup>~950.22km<sup>2</sup>。采用 Mann-Whitney U 检验评估了 Stacking 有效流域的流域面积是否与 Stacking 无效流域存在显著差异。结果表明，在 1~7 天预见期内，Stacking 有效流域的流域面积与 Stacking 无效流域的流域面积无显著差异（P 值均大于 0.1）。这表明，Stacking 的效果与流域面积无显著相关性。

### 3.4 Stacking 径流预测效果与个体学习器精度的相关性分析

考虑到 NSE 和 RMSE 在评价模型径流预测精度方面具有一致性，本研究基于 NSE 分析 Stacking 径流预测效果与个体学习器精度的相关性。在 1~7 天预见期内，Stacking 有效流域和无效流域的个体学习器的平均 NSE（即 SVR 模型 NSE 与 RF 模型 NSE 的平均值）箱线图如图 9 所示。由图 9 可知，Stacking 有效

流域中的个体学习器平均 *NSE* 的中位数和平均值均低于 Stacking 无效流域。采用 Mann-Whitney U 检验评估了 Stacking 有效流域的个体学习器平均 *NSE* 是否显著低于 Stacking 无效流域。结果表明，在 1~7 天预见期内，Stacking 有效流域的个体学习器平均 *NSE* 显著低于无效流域 (P 值均小于 0.01)。这表明，Stacking 更倾向于在个体学习器预测精度较低的流域提高预测精度，而在个体学习器预测精度较高的流域难以提高径流预测精度。

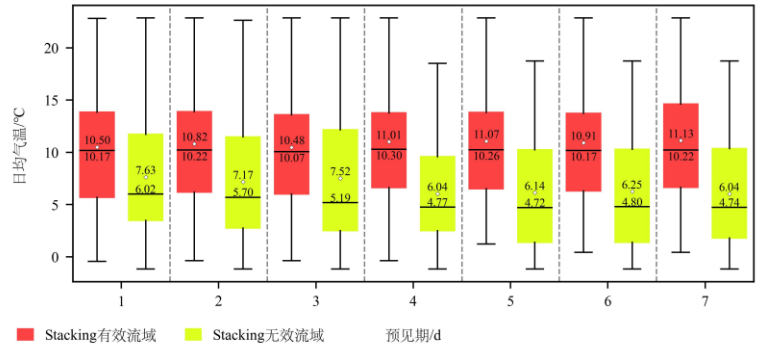


图 7 Stacking 有效流域和无效流域的日均气温箱线图

Fig.7 Box plots of average daily temperature for Stacking-effective basins and Stacking-ineffective basins

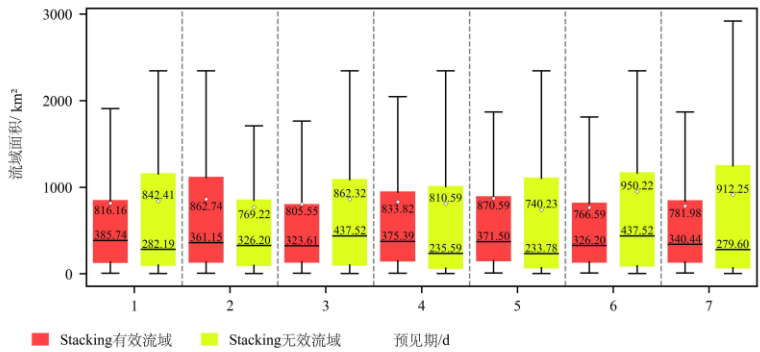


图 8 Stacking 有效流域和无效流域的流域面积箱线图

Fig.8 Box plots of basin areas for Stacking-effective basins and Stacking-ineffective basins

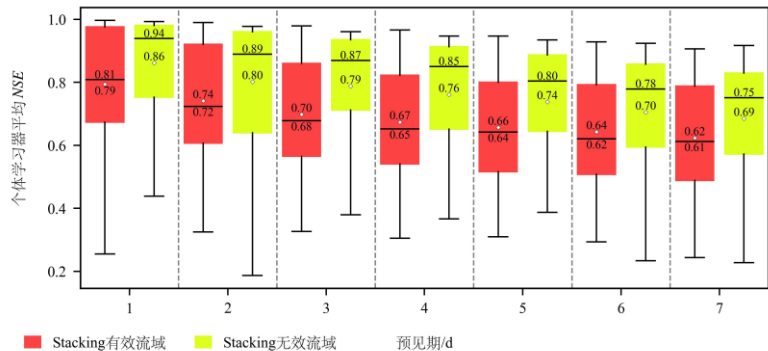


图 9 Stacking 有效流域和无效流域的个体学习器平均 *NSE* 箱线图

Fig.9 Box plots of the average *NSE* of individual learners for Stacking-effective basins and Stacking-ineffective basins

## 4 讨论

### 4.1 Stacking 径流预测效果的影响因素分析

尽管 Stacking 在大多数流域中提高了径流预测精度，但其径流预测效果并不具备普遍适用性，Stacking 径流预测效果与流域特征、个体学习器精度以及个体学习器-元学习器组合的适配性等因素相关。

首先，Stacking 径流预测效果与流域气候特征高度相关，不同气候类型流域的径流产生机制不同。就本研究而言，美国大陆东部流域的气候类型主要为湿润/半湿润季风气候，降水量较大、气温较高，其径流机制复杂或者存在多种径流机制，不同类型的个体学习器可能擅长捕捉复杂径流机制的不同方面或不同的径流机制，不同个体学习器对上述流域的径流预测结果具有较高的互补性，元学习器能够对个体学习器的预测结果进行有效融合，从而获得比个体学习器更高的径流预测精度。而对于降水量较小的美国大陆西部干旱/半干旱流域，往往以单一超渗径流机制为主，使用单一的个体学习器已能有效捕捉核心径流机制，不同个体学习器预测结果趋同，互补性较差，Stacking 难以提升径流预测精度。此外，Stacking 径流预测效果还与流域数据特征相关，不同类型流域的数据特征往往不同。美国大陆东部湿润/半湿润流域的降水-径流事件丰富，包含了多样的径流模式，径流信息密度高，有助于不同的个体学习器产生互补的预测结果，同时有助于元学习器充分地学习出集成策略，从而提高径流预测精度。而美国大陆西部干旱/半干旱流域降水-径流事件稀少，径流模式的多样性不足，极端性较强，导致不同个体学习器的预测结果趋同，互补性差，元学习器无法对不同径流预测结果进行有效集成，甚至产生过拟合问题，难以提高径流预测精度。

其次，Stacking 径流预测精度的提升依赖于个体学习器预测结果的互补性和误差独立性。当所有的个体学习器预测精度均较高时，其预测结果趋于相同，互补性较弱，而误差结构高度相关。在此情况下，元学习器无法从个体学习器的输出中学习额外的互补信息，而是可能过度学习噪声，容易产生过拟合，从而难以实现精度提升，甚至出现径流预测精度下降的现象。此外，从另外一个视角来看，当所有个体模型的径流预测精度已经很高时，引入元学习器，相当于增加了模型复杂度，容易造成泛化能力和径流预测能力的下降。而当个体学习器精度适中时，不同个体学习器存在较为明显的误差互补空间，有利于元学习器充分学习到互补信息，从而提高径流预测精度。在本研究中，Stacking 有效流域的个体学习器 *NSE* 中位数范围为 0.62~0.81，预测精度适中；Stacking 无效流域的个体学习器 *NSE* 中位数为 0.75~0.94，预测精度较高。研究结果印证了个体学习器精度对 Stacking 适用性的影响。

此外，鉴于流域径流过程的复杂性以及不同流域间的差异性，不仅不存在一个适用于所有流域的机器学习模型，同样也难以找到一个能够普遍提升所有流域径流预测精度的个体学习器-元学习器组合。具体而言，适用于不同流域的个体学习器-元学习器组合可能存在差异。本研究以 SVR 和 RF 作为个体学习器，以岭回归作为元学习器，在 60% 以上的流域提高了径流预测精度。倘若考虑不同流域的特性，为每个流域适配性地选择个体学习器-元学习器组合，有望进一步提高 Stacking 的适用性。此外，在某些流域中，由于水文响应机制独特或数据特征复杂，难以寻找到有效的个体学习器-元学习器组合，导致径流预测精度提升有限，甚至出现预测精度下降；而在另一些流域中，预测性能对个体学习器-元学习器组合相对不敏感，多数常用的组合方式均能取得一定的改进效果。这种由流域特性所导致的模型组合适配性差异，是基于集成学习的径流预测研究中必须正视的问题。

### 4.2 与现有 Stacking 径流预测研究的对比分析

现有 Stacking 径流预测研究通常基于数量稀少的流域样本，所构建的模型规模也相对较小，这种数据基础与模型体量的不足，导致难以提炼出具有一定普适性的规律与结论。其次，现有研究大多呈现的是 Stacking 径流预测的正面案例，鲜少探讨其失效的负面情形。最后，相关研究侧重于评价不同个体学习器或元学习器对 Stacking 准确性的影响，而缺少对 Stacking 适用性和影响因素的分析。

本研究利用了覆盖不同水文气象条件的 200 个流域的数据，分别构建了 1400 个 SVR 模型、1400 个 RF 模型及与之对应的 1400 个 Stacking 集成模型，形成了规模可观的径流预测模型集合。在此基础上，本研究不仅评价了 Stacking 的预测精度和稳定性，还初步揭示了 Stacking 的适用性及其效果与流域特征、

个体学习器精度等因素相关性。本研究与现有研究的切入点不同，两者在研究内容上相互补充，共同深化了对 Stacking 径流预测问题的理解。

#### 4.3 本研究的不足与展望

受计算资源与模型训练成本的限制，本研究选择以 SVR 与 RF 作为个体学习器，以岭回归作为元学习器，所涵盖的模型种类相对有限。本研究选用的个体学习器（SVR 与 RF）虽在径流预测领域具有广泛的应用，但尚未引入如 LSTM 等深度学习模型作为个体学习器。不同的个体学习器-元学习器组合可能为 Stacking 径流预测带来一定的不确定性。本研究将气温、降水、气温和历史径流作为模型输入特征，未考虑潜在蒸散和土壤湿度等要素。此外，本研究基于 200 个流域和 7 个预见期开展了 Stacking 径流预测应用研究，在流域数量和预见期长度方面仍有进一步扩展的空间。

考虑到不同的个体学习器-元学习器组合可能为 Stacking 径流预测带来一定的不确定性，后续研究拟引入更多类型的机器学习模型（特别是深度学习模型）作为个体学习器，选用岭回归之外的其他机器学习模型作为元学习器。将潜在蒸散和土壤湿度等要素作为模型输入特征，以增强结果解释维度。在更大规模的流域样本和更长的预见期上进一步探索 Stacking 径流预测中的适用性与影响因素。

## 5 结论

本研究以 SVR 和 RF 作为个体学习器，以岭回归作为元学习器，构建了基于 Stacking 的径流预测模型。在 CAMELS 数据集的 200 个流域及 1~7 天预见期内，系统评价了 Stacking 径流预测的准确性和稳定性，着重评价了 Stacking 径流预测的适用性和分析了 Stacking 径流预测效果的影响因素。主要结论如下：

- (1) 在 1~7 天预见期内，Stacking 的整体预测准确性要显著高于个体学习器 SVR 和 RF。
- (2) Stacking、SVR 和 RF 的径流预测精度随着预见期的延长，均呈现出下降趋势，Stacking 的预测精度下降幅度低于 SVR 和 RF，呈现出更强的稳定性。
- (3) Stacking 无法保证在所有流域上均能提高预测精度，Stacking 在 61%~67.5% 的流域上提高了预测精度，预测精度提高的流域多集中在美国大陆东部湿润地区。
- (4) Stacking 的效果与流域的降水量和气温相关，Stacking 倾向于在降水量较大和气温较高的流域提高预测精度，而在降水量较小和气温较低的流域难以提高精度。
- (5) Stacking 的效果与个体学习器的预测精度相关，Stacking 倾向于在个体学习器精度较低的流域提高预测精度，在个体学习器精度较高的流域难以提高预测精度。

## 6 参考文献

- [1] Deng C, Sun PY, Yin X *et al.* Runoff simulation in the upper Han River Basin using physics-informed machine learning model. *J Lake Sci*, 2025, **37**(1): 279-292. DOI: 10.18307/2025.0143. [邓超, 孙培源, 尹鑫等. 基于概念性水文模型与长短期记忆模型耦合的汉江上游流域径流模拟. *湖泊科学*, 2025, **37**(1): 279-292.]
- [2] Hu LY, Jiang XL, Zhou JH *et al.* Application of LSTM considering time steps in runoff prediction of Ganjiang River Basin. *J Lake Sci*, 2024, **36**(4): 1241-1251. DOI: 10.18307/2024.0454. [胡乐怡, 蒋晓蕾, 周嘉慧等. 考虑记忆时间的 LSTM 模型在赣江流域径流预报中的应用. *湖泊科学*, 2024, **36**(4): 1241-1251.]
- [3] Kreibich H, Van Loon AF, Schröter K *et al.* The challenge of unprecedented floods and droughts in risk management. *Nature*, 2022, **608**(7921): 80-86. DOI: 10.1038/s41586-022-04917-5.
- [4] Wang N, Li MW, Chen SZ *et al.* A watershed runoff reconstruction method based on stacked ensemble machine learning under incomplete data: A case study of Juzhanghe River. *Water Resources and Hydropower Engineering*, 2024, **55**(10): 85-97. DOI: 10.13928/j.cnki.wrahe.2024.10.007. [王楠, 李明蔚, 陈首志等. 基于堆叠机器学习的非充分数据条件下流域径流还原方法研究: 以沮漳河为例. *水利水电技术(中英文)*, 2024, **55**(10): 85-97.]
- [5] Li GY. Further comprehensively deepen water conservancy reform to promote high-quality development of water conservancy and safeguard national water security: a speech at the 2025 national water conservancy work conference. *Water Resources*

- Development Research*, 2025, **25**(01): 1-12. DOI: 10.13928/j.cnki.wrdr.2025.01.001. [李国英. 进一步全面深化水利改革为推动水利高质量发展、保障我国水安全作出新的贡献——在 2025 年全国水利工作会议上的讲话. 水利发展研究, 2025, **25**(01): 1-12.]
- [6] Miao C, Gou J, Fu B *et al.* High-quality reconstruction of China's natural streamflow. *Science Bulletin*, 2022, **67**(5): 547-556. DOI: 10.1016/j.scib.2021.09.022.
- [7] Li B, Tian FQ, Li YK *et al.* Development of a spatiotemporal deep-learning-based hydrological model. *Advances in Water Science*, 2022, **33**(06): 904-913. DOI: 10.14042/j.cnki.32.1309.2022.06.005. [李步, 田富强, 李钰坤等. 融合气象要素时空特征的深度学习水文模型. 水科学进展, 2022, **33**(06): 904-913.]
- [8] Kratzert F, Klotz D, Brenner C *et al.* Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 2018, **22**(11): 6005-6022. DOI: 10.5194/hess-22-6005-2018.
- [9] Rasouli K, Hsieh WW, Cannon AJ. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 2012, **414**: 284-293. DOI: 10.1016/j.jhydrol.2011.10.039.
- [10] Cui DW. Application of Hidden Multilayer BP Neural Network Model in Runoff Prediction. *Journal of China Hydrology*, 2013, **33**(01): 68-73. DOI: 10.3969/j.issn.1000-0852.2013.01.013. [崔东文. 多隐层 BP 神经网络模型在径流预测中的应用. 水文, 2013, **33**(01): 68-73.]
- [11] Pham LT, Luo L, Finley AO. Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. *Hydrology and Earth System Sciences Discussions*, 2021, **25**(6): 2997-3015. DOI: 10.5194/hess-25-2997-2021.
- [12] Guo J, Zhou JZ, Zhang YC *et al.* Daily Runoff Forecast Based on Improved Support Vector Machine Regression Model. *Water Power*, 2010, **36**(03): 12-15. DOI: 10.3969/j.issn.0559-9342.2010.03.004. [郭俊, 周建中, 张勇传等. 基于改进支持向量机回归的日径流预测模型. 水力发电, 2010, **36**(03): 12-15.]
- [13] Fan HX, He HD, Xu LG *et al.* Simulation and attribution analysis based on the long-short-term-memory network for detecting the dominant cause of runoff variation in the Lake Poyang Basin. *J Lake Sci*, 2021, **33**(3): 866-878. DOI: 10.18307/2021.0319. [范宏翔, 何菡丹, 徐力刚等. 基于长短记忆模型的鄱阳湖流域径流模拟及其演变的归因分析. 湖泊科学, 2021, **33**(3): 866-878.]
- [14] Yuan D, Tan YG, Zhu YX *et al.* Research of monthly runoff forecast of Jinsha River Basin based on VMD-PSO-LSTM. *Water Resources and Hydropower Engineering*, 2024, **55**(S1): 28-38. DOI: 10.13928/j.cnki.wrahe.2024.S1.004. [袁旦, 谭尧耕, 朱艳霞等. 改进 LSTM 在金沙江流域中长期径流预报研究. 水利水电技术(中英文), 2024, **55**(S1): 28-38.]
- [15] Dong X, Yu Z, Cao W *et al.* A survey on ensemble learning. *Frontiers of Computer Science*, 2020, **14**(2): 241-258. DOI: 10.1007/s11704-019-8208-z.
- [16] Sagi O, Rokach L. Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 2018, **8**(4): e1249. DOI: 10.1002/widm.1249.
- [17] Zounemat-Kermani M, Batelaan O, Fadaee M *et al.* Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 2021, **598**: 126266. DOI: 10.1016/j.jhydrol.2021.126266.
- [18] Tosan M, Nourani V, Kisi O *et al.* Evolution of ensemble machine learning approaches in water resources management: a review. *Earth Science Informatics*, 2025, **18**(2): 416. DOI: 10.1007/s12145-025-01911-z.
- [19] Duan YN, Liang ZM, Zhao JF *et al.* Combined Forecasting of Hydrological Model Based on Stacking Integrated Framework. *Water Resources and Power*, 2022, **40**(09): 27-30+39. DOI: 10.20040/j.cnki.1000-7709.2022.20212422. [段雅楠, 梁忠民, 赵建飞等. 基于 Stacking 集成框架的水文模型组合预报研究. 水电能源科学, 2022, **40**(09): 27-30+39.]
- [20] Lin YE, Meng Y, Du Y *et al.* Study on Influence Factors about Runoff Forecasting Performance of Stacking Integrated Model. *Journal of China Hydrology*, 2023, **43**(01): 57-61. DOI: 10.19797/j.cnki.1000-0852.20210399. [林泳恩, 孟越, 杜懿等. 堆叠集成模型径流预报效果的影响因素研究. 水文, 2023, **43**(01): 57-61.]
- [21] Granata F, Di Nunno F, de Marinis G. Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: A comparative study. *Journal of Hydrology*, 2022, **613**: 128431. DOI: 10.1016/j.jhydrol.2022.128431.
- [22] Granata F, Di Nunno F. Forecasting short- and medium-term streamflow using stacked ensemble models and different meta-

- learners. *Stochastic Environmental Research and Risk Assessment*, 2024, **38**(9): 3481-3499. DOI: 10.1007/s00477-024-02760-w.
- [23] Wang W-C, Gu M, Li Z *et al.* A stacking ensemble machine learning model for improving monthly runoff prediction. *Earth Science Informatics*, 2024, **18**(1): 120. DOI: 10.1007/s12145-024-01544-8.
- [24] Qian X, Wang B, Chen J *et al.* An explainable ensemble deep learning model for long-term streamflow forecasting under multiple uncertainties. *Journal of Hydrology*, 2025, **662**: 133968. DOI: 10.1016/j.jhydrol.2025.133968.
- [25] Deb D, Arunachalam V, Raju KS. Daily reservoir inflow prediction using stacking ensemble of machine learning algorithms. *Journal of Hydroinformatics*, 2024, **26**(5): 972-997. DOI: 10.2166/hydro.2024.210.
- [26] Lu M, Hou Q, Qin S *et al.* A Stacking Ensemble Model of Various Machine Learning Models for Daily Runoff Forecasting. *Water*, 2023, **15**(7): 1265. DOI: 10.3390/w15071265.
- [27] Addor N, Newman AJ, Mizukami N *et al.* The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 2017, **21**(10): 5293-5313. DOI: 10.5194/hess-21-5293-2017.
- [28] Thornton PE, Thornton MM, Mayer BW *et al.* Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Management of Daily Surface Weather Data Products. Oak Ridge, Tennessee: Oak Ridge National Laboratory, 2014. DOI: 10.3334/ORNLDAAC/1219
- [29] Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 1998, **2**(2): 121-167. DOI: 10.1023/A:1009715923555.
- [30] Basak D, Pal S, Patranabis DC. Support vector regression. *Neural Information Processing-Letters and Reviews*, 2007, **11**(10): 203-224. DOI: 10.1007/978-1-4302-5990-9\_4.
- [31] Breiman L. Random forests. *Machine learning*, 2001, **45**: 5-32. DOI: 10.1023/a:1010933404324.
- [32] Loh WY. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 2011, **1**(1): 14-23. DOI: 10.1002/widm.8.
- [33] Wolpert DH. Stacked generalization. *Neural networks*, 1992, **5**(2): 241-259. DOI: 10.1016/S0893-6080(05)80023-1.
- [34] Pavlyshenko B. Using Stacking Approaches for Machine Learning Models. In: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). Lviv: IEEE, 2018: 255-258. DOI: 10.1109/DSMP.2018.8478522.