

目录

第一章 全球陆地关键水循环要素演变格局及多元驱动因子分析	1
1.1 概述	1
1.2 研究方法	1
参考文献	5

插图清单

图 1.2.1	BFAST 算法流程图	2
图 1.2.2	径流缺失数据插补流程图	3

附表清单

第一章 全球陆地关键水循环要素演变格局及多元驱动因子分析

1.1 概述

1.2 研究方法

1.2.1 时间序列趋势分析方法

1.2.1.1 线性趋势分析方法

研究采用基于最小二乘的一元线性回归方法对时间序列进行拟合, 得到线性变化的斜率和截距, 以此来描述时间序列的变化趋势。同时采用 Mann-Kendall 非参数趋势检验方法对序列线性变化的显著性进行检验。

以时间为自变量, 时间序列的观测值为因变量, 线性回归模型可以表示为:

$$slope = \frac{n \times \sum_{t=1}^n (t \times Y_t) - \sum_{t=1}^n t \times \sum_{t=1}^n Y_t}{n \times \sum_{t=1}^n t^2 - (\sum_{t=1}^n t)^2} \quad (1.2.1)$$

1.2.1.2 线性-非线性趋势识别算法

Jamali 等^[1]

1.2.2 水文气候序列断点检测算法

研究采用 BFAST 算法 (Breaks For Additive Seasonal and Trend) 对水文气候序列进行断点检测。BFAST 是一种基于序列加法分解的断点检测方法, 最初被提出用于处理 NDVI 数据^[2,3], 但是可被扩展用于各种具有周期特征时间序列, 被广泛应用于水文、气象等领域的分析^[4-6]。算法首先基于 STL 算法 (Seasonal and Trend decomposition using Loess) 对时间序列进行分解, 得到趋势项、季节项和剩余项三个部分:

$$Y_t = T_t + S_t + r_t, \quad t = 1, 2, \dots, n \quad (1.2.2)$$

式中 Y_t 为时间序列, T_t 为趋势项, S_t 为季节项, r_t 为剩余项, 即时间序列中的除了趋势和季节之外的部分。BFAST 算法利用基于普通最小二乘 (OLS) 残差的 MOving SUM (MOSUM) 检验来确定分解序列中是否存在断点, 如果存在, 则使用贝叶斯信息准则 (BIC) 来确定断点的位置和数量, 并且分别使用线性函数和三次谐波函数对趋势项和季节项进行分段拟合。假定在趋势项 T_t 中检测出了 m 个断点 t_1, t_2, \dots, t_m , 定义 $t_0=0$, $t_{m+1}=n$, 在趋势项 S_t 中检测出 p 个断点 $\tau_1, \tau_2, \dots, \tau_p$, 定义 $\tau_0=0$, $\tau_{p+1}=n$, 则趋势项和季节项分别被分段拟合为:

$$\widehat{T}_{i,t} = \alpha_i + \beta_i t \quad (t_{i-1} < t \leq t_i, t = 1, 2, \dots, m+1) \quad (1.2.3)$$

$$\widehat{S}_{j,t} = \sum_{h=1}^k [\gamma_{j,h} \sin(\frac{2\pi ht}{f} + \mu_h) + \delta_{j,h} \cos(\frac{2\pi ht}{f} + \nu_h)] \quad (1.2.4)$$

$$(\tau_{j-1} < t \leq \tau_j, j = 1, 2, \dots, p+1)$$

其中 i, j 为趋势项和周期项中突变点所在位置, α_i, β_i 分别为线性模型的拟合斜率和截距, $\gamma_{j,h}, \delta_{j,h}$ 分别为谐波函数中正弦项和余弦项的振幅参数, μ_h, ν_h 为对应项的相位参数, k 为谐波函数的阶数, 在 BFAST 算法中设置为 3。 f 为数据周期频率, 如对于月尺度数据, $f=12$ 。

对分段拟合后的结果进行求和, 形成新的模拟序列, 之后再次进行分解与断点检测步骤, 不断迭代更新序列中断点可能发生的事件和显著性, 直到最终确定分解结果和鲁棒的断点位置。算法的流程如图 1.2.1 所示。

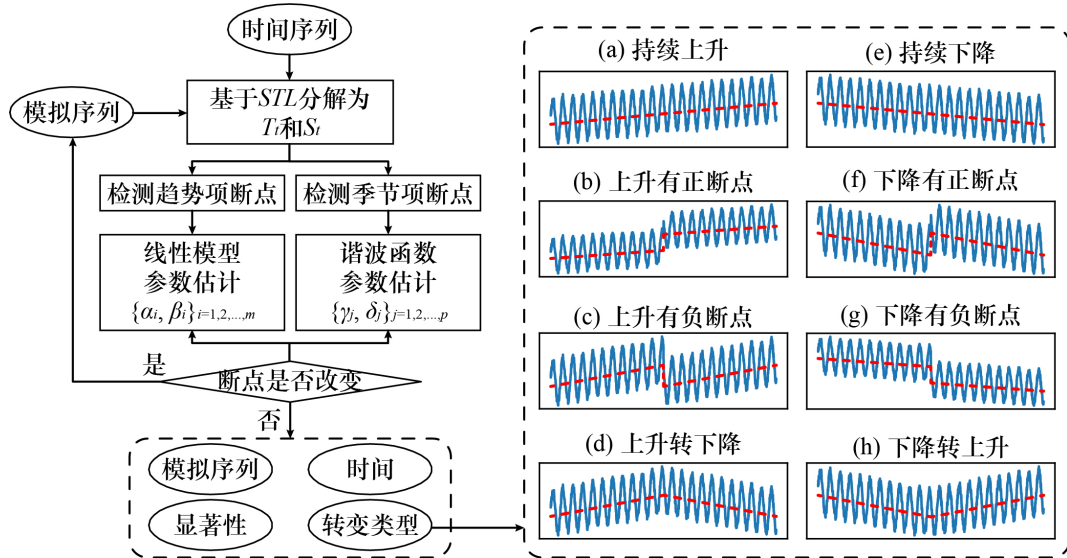


图 1.2.1 BFAST 算法流程图

Fig. 1.2.1 Flowchart of BFAST breakpoint detection algorithm

本研究使用由 R 语言开发的“bfast”程序包, 逐站点对水文气候序列进行断点检测。程序包中, “bfast”和“bfast01”函数分别对应算法的泛化模式和变化检测模式。其中泛化模式能够指定断点数量, 并基于上面介绍的分解拟合算法对原始时间序列进行模拟, 变化检测模式是检测完整时间序列中是否存在一个最显著的断点, 并提取断点的起始时间、变化类型、显著性等详细信息^[7]。根据断点前后序列趋势变化的方向及显著性差异, “bfast01”函数将趋势变化分为 8 种类型, 分别是: 持续上升、持续下降、持续上升并有正向断点、持续上升并有负向断点、持续下降并有正向断点、持续下降并有负向断点、上升转下降、下降转上升, 以提供对时间序列变化断点更详细的

了解。

1.2.3 基于数据驱动的数据缺失值插补方法

由于 BFAST 断点检测算法对时间序列中缺失值敏感，为了保证算法的准确性与可执行性，需要对原始径流序列中的缺失值与无效值进行插补。由于本研究的目的是对实测径流进行分析，因此插补方法需要尽可能保留原始数据的特征，同时尽量减少插补过程中引入的不确定性。在本研究中，采用了基于数据驱动的插补方法，即基于径流序列的历史数据和其他气象要素的数据，通过数据间的相关性来插补缺失值。基于数据驱动的插补方法的基本思想是通过已有数据的特征来推断缺失数据的值，根据非缺失值时间段内的气候数据与径流观测建立联系，训练并评估机器学习模型，利用数据缺失时段的气候数据驱动训练好的模型，对缺失时段的径流进行插补。缺失径流数据插补的基本框架如图1.2.2所示。

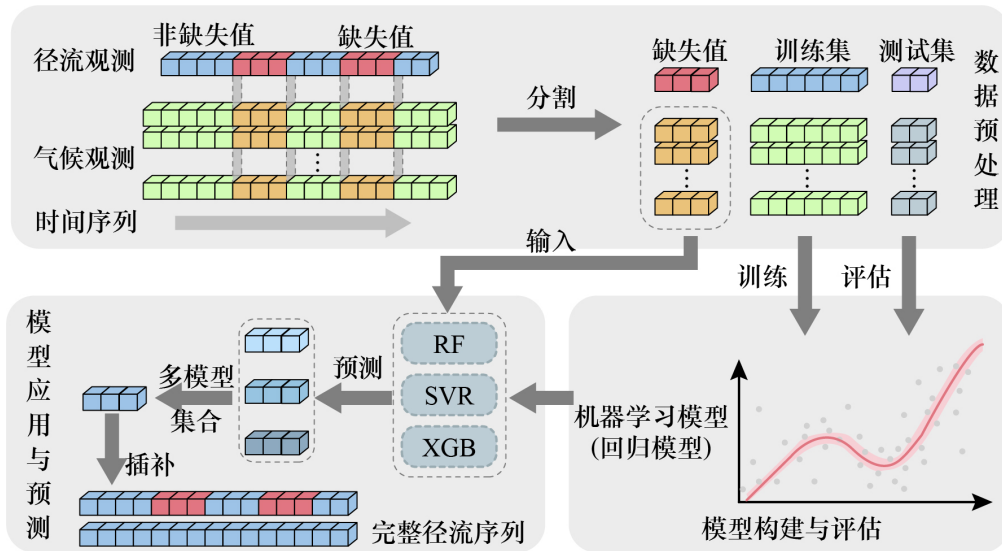


图 1.2.2 径流缺失数据插补流程图

Fig. 1.2.2 Interpolation flowchart of missing runoff data

本研究采用了三种常用的回归类型机器学习算法构建流域气候条件与实测径流之间的关系，分别是支持向量机回归（Support Vector Machine Regression, SVR）、随机森林（Random Forest, RF）和极端梯度提升（eXtreme Gradient Boosting, XGBoost）。这三种算法在机器学习领域中被广泛应用于回归问题，具有较好的拟合能力和泛化能力。在本研究中，对于每个流域，将径流数据作为目标变量，其他气象要素（降水、潜在蒸散发、最高气温、最低气温和平均气温）的数据作为特征变量，构建回归模型。考虑到气候条件对径流的滞后影响，每种气候变量的滞后阶数设置在 0-3 之间，因此有 5（气象要素）×4（滞后时段）共 20 种特征变量。在模型训练阶段，采用了 5 折交叉验证的方法，以减少模型过拟合的风险。采用预测结果与实测结果的纳什效率系数（Nash-Sutcliffe Efficiency Coefficient, NSE）、相对误差（Relative Error, RE）和相关

系数 (Correlation Coefficient, CC), 均方根误差 (Root Mean Square Error, RMSE) 作为评价指标, 以评估模型的模拟结果与泛化能力。

$$NSE = 1 - \frac{\sum_{t=1}^n (Q_{o_i} - Q_{s_i})^2}{\sum_{t=1}^n (Q_{o_i} - \overline{Q_o})^2} \quad (1.2.5)$$

$$RE = \frac{\overline{Q_s} - \overline{Q_o}}{\overline{Q_o}} \quad (1.2.6)$$

$$CC = \frac{\sum_{t=1}^n ((Q_{o_i} - \overline{Q_o})(Q_{s_i} - \overline{Q_s}))}{\sqrt{\sum_{t=1}^n (Q_{o_i} - \overline{Q_o})^2 \sum_{t=1}^n (Q_{s_i} - \overline{Q_s})^2}} \quad (1.2.7)$$

$$RMSE = \sqrt{\sum_{t=1}^n \frac{(Q_{s_i} - Q_{o_i})^2}{n}} \quad (1.2.8)$$

式中 Q_o 和 Q_s 分别代表了观测值和模拟值, $\overline{Q_o}$ 和 $\overline{Q_s}$ 分别代表了观测值和模拟值的平均值, n 为样本数量。

在模型应用阶段, 利用数据缺失时段的气候数据驱动训练好的模型, 预测缺失时段的径流数据。为了减少模型预测结果的不确定性, 对于每个流域, 采用三种机器学习算法分别进行插补, 最终插补结果取三种算法的平均值作为最终插补结果。

参考文献

- [1] Jamali S, Seaquist J, Eklundh L, et al. Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel [J]. *Remote Sensing of Environment*, 2014, 141: 79-89. <https://doi.org/10.1016/j.rse.2013.10.019>.
- [2] Verbesselt J, Hyndman R, Newnham G, et al. Detecting trend and seasonal changes in satellite image time series [J]. *Remote Sensing of Environment*, 2010, 114(1): 106-115. <https://doi.org/10.1016/j.rse.2009.08.014>.
- [3] Verbesselt J, Hyndman R, Zeileis A, et al. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series [J]. *Remote Sensing of Environment*, 2010, 114(12): 2970-2980. <https://doi.org/10.1016/j.rse.2010.08.003>.
- [4] Li J, Li Z L, Wu H, et al. Trend, seasonality, and abrupt change detection method for land surface temperature time-series analysis: Evaluation and improvement [J]. *Remote Sensing of Environment*, 2022, 280: 113222. <https://doi.org/10.1016/j.rse.2022.113222>.
- [5] Jiang P, Ding W g, Yuan Y, et al. Identifying trend shifts in vegetation greenness in China from 1982 to 2015 [J]. *Land Degradation & Development*, 2022, 33(9): 1434-1445. <https://doi.org/10.1002/ldr.4240>.
- [6] Bernardino P N, De Keersmaecker W, Fensholt R, et al. Global-scale characterization of turning points in arid and semi-arid ecosystem functioning [J]. *Global Ecology and Biogeography*, 2020, 29(7): 1230-1245. <https://doi.org/10.1111/geb.13099>.
- [7] 姜萍. 1982-2015 年中国植被覆盖变化及其对气候变化的敏感性分析 [D]. 兰州大学, 2022.