

---

# 城市环境所等构建水环境动态数据挖掘框架

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/18751.html>

**本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！**

水环境动态性是环境科学和生态学在水域生态系统的主要交叉研究主题之一。根据水域生态系统边界特征，水环境动态性关联机制包括外部和内部两方面，一是从环境科学角度看水环境动态性受到系统外部因素的驱动，例如流域内气候水文和人类活动叠加作用下水体环境变化；二是从生态学角度看水环境动态性与系统内部水生生物之间具有相互作用，如通过有机物生产和分解等过程改变水体的环境。水体野外原位观测数据和科学文献与日俱增，为环境科学和生态学在水环境动态性外部和内部关联机制多尺度定量剖析带来了新机遇。

水环境动态性的研究随着水体野外原位观测数据快速递增带来“喜悦”的同时，增长中的“烦恼”也随之而来。观测数据和科学文献的快速递增亟待破解数据共享中遇到的一系列难题。例如，不同研究中在观测期限和频率上的差异，明显成为数据共享的主要壁垒之一。中国科学院城市环境研究所联合美国新墨西哥州立大学，以水库总氮这一关键营养物季节变化为案例，构建了针对科学文献水环境动态数据挖掘的系统性框架。

水环境动态数据挖掘框架包括三个逐级关联的模块：第一级是数据提取模块，其模块功能是文献下载、筛选和分组以及文献图表数据的提取与汇总；第二级是数据分流模块，其模块功能是根据观测时长、观测频率和原始数据表达方式等限制条件对第一级模块产生的数据进行多次分流，即分出“混浊”数据，洗出“清洁”数据产生时间序列数据矩阵；第三级是数理统计模块，其模块功能是对时间序列数据矩阵进行归一化处理 and 凝聚层次聚类分析及分析结果可视化。

在案例分析中，该框架识别了水库总氮动态的三种模式：一是夏低谷模式，即水库总氮动态由浮游植物的生长和降雨等自然过程驱动；二是夏高峰模式，即湿沉降和农业径流导致水库总氮在夏季处于峰值；三是春高峰模式，即水库总氮季节动态可能受到人为排放的干扰。这个案例分析并不意味着全球总氮动态仅有三种模式。随着水库总氮动态原位观测数据和科学文献的继续递增，迭代应用该框架可能会识别出更多的模式，或有更充足的证据表明仅存在三种模式。

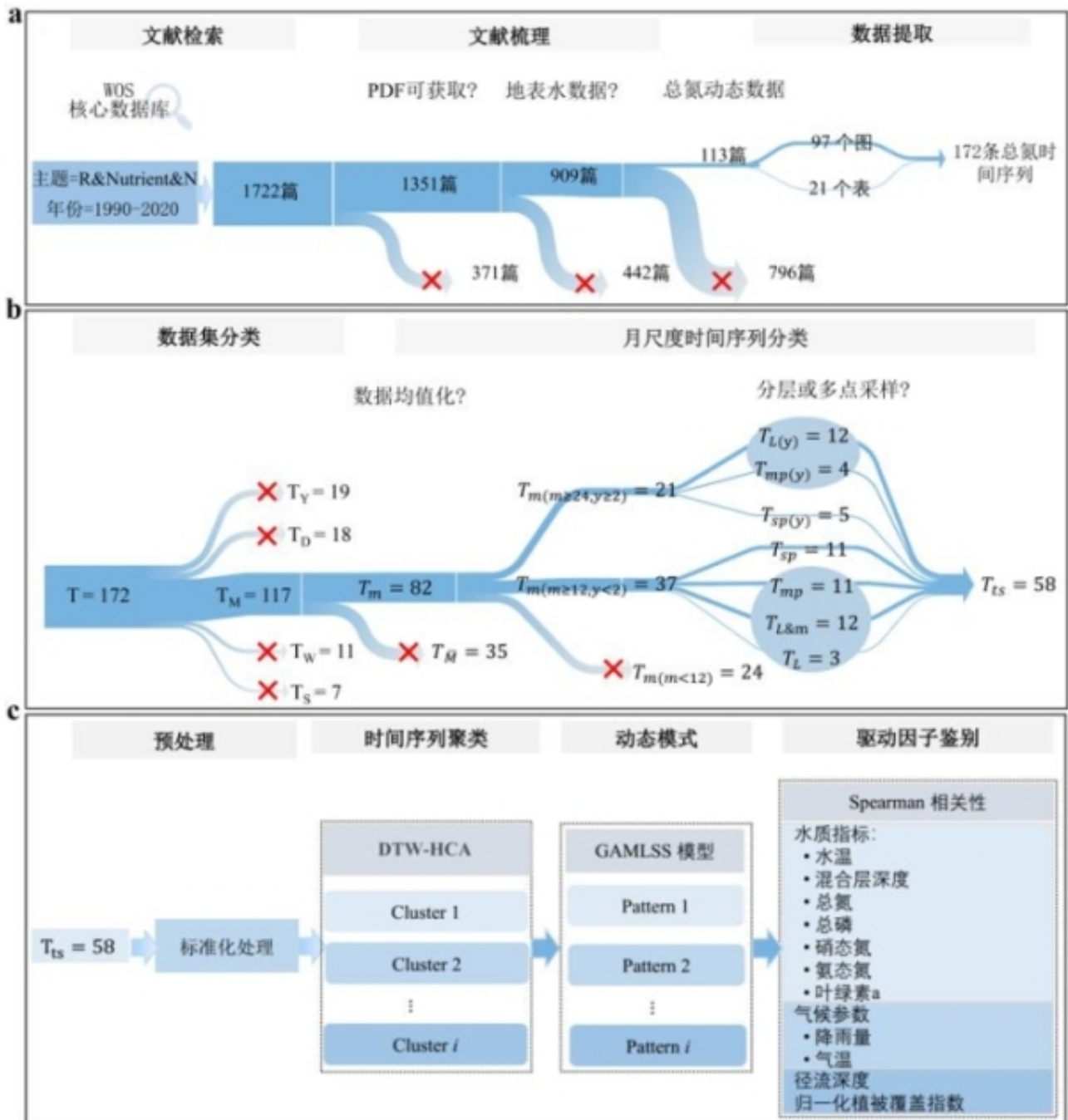
水环境动态数据挖掘框架进一步整合和夯实了科研人员多年前两项研究中的数据分析路径。该框架的第一级和第二级模块，对其中一项针对产油藻类种群动态科学文献数据提取路径进行了分级和优化；第三级模块对另一项关于水库浮游植物动态同步性检验统计学方法进行了相应改进。这些路径分级和优化及方法改进，不意味着该框架已能够完全适用于水环境动态数据挖掘。事实上，因水环境参数多种多样及数据观测过程和表达方式差异化，其科学文献数据共享存在不同程度的挑战。这预示着水环境动态数据挖掘框架的优化和升级研究尚待发展。

水环境动态数据挖掘框架和案例分析结果以Revisiting seasonal dynamics of total nitrogen in reservoirs with a systematic framework for mining data from existing publications为题，发表在Water

Research上。研究团队在同期Water Research上以Bayesian change point quantile regression approach to enhance the understanding of shifting phytoplankton-dimethyl sulfide relationships in aquatic ecosystems

为题，发表了关于水域生态系统冷室气体物质对浮游植物生物量响应的边界变点模型的合作研究成果。

[论文链接](#)



数据挖掘框架中提取-分流-统计逐级关联模块

研究团队单位：城市环境研究所

---

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发