

水环境多源数据融合方法及应用探讨

马颖丽 汪 涛 李 超
保定学院 河北 保定 071001

摘 要: 水环境承载力评估与预警依赖水文、水质、气象、遥感和社会经济等多源数据,但各类数据在来源、格式、尺度、精度和更新频率上差异显著,传统单一数据源方法难以满足精细化管理需求。本文提出基于概率信息处理的多源数据融合策略,涵盖数据预处理、概率特征提取、不确定性量化与融合推理四个环节,构建适用于水环境管理场景的通用融合框架。研究表明,该方法能有效处理数据的不确定性与异质性,可为承载力评估、风险预警与综合决策提供可靠支撑。

关键词: 水环境承载力; 多源数据融合; 概率信息处理; 异构数据; 数据预处理

引言: 水资源系统具有高度复杂性,单一类型数据难以完整刻画水环境承载力的真实状态。不同监测手段所获数据在时空分辨率、量纲量级和质量可靠性上存在天然差异,导致传统评估方法面临信息不完整与精度不足的双重困境。近年来,大数据技术虽极大丰富了数据来源,但有效整合多源异构数据仍缺乏成熟的方法体系。对此,本文拟从信息融合的角度出发,引入概率信息处理理论,设计一套兼顾普适性与操作性的数据融合策略,为承载力评估与风险预警提供方法支撑。

1 水环境承载力相关数据的特征分析

1.1 数据来源的多样性

水环境承载力评估与预警所依赖的数据来源广泛,根据采集方式和获取渠道的不同,大致可分为以下五类。第一类是水文监测数据,包括水位、流量、流速、蓄水量等指标,通过地面监测站点以固定频率采集,时间连续性强,但空间覆盖有限,主要依赖站点布设密度。第二类是水质监测数据,涵盖溶解氧、COD、氨氮、总磷、总氮等指标,同样来源于地面站点采样与实验室分析,精度较高但监测频次相对较低。第三类是气象观测数据,包括降水量、气温、风速、日照时数等,来源于气象站点,对水环境的水量和水质变化具有重要驱动作用。第四类是遥感数据,通过卫星或无人机平台获取,能够提供大范围的水域面积、叶绿素a浓度、悬浮物浓度等信息,空间连续性好但反演精度受大气条件和算法模型影响。第五类是社会经济数据,包括人口密度、土地利用、工农业产值、污染排放量等,来源于统计年鉴和行政记录,更新周期较长,空间分辨率较粗^[1]。

1.2 数据的异构性表现

多源数据在多个维度上存在异构性,这种异构性是数据融合面临的核心技术难点。格式异构方面,监测站

点数据多为结构化表格数据,遥感数据为栅格影像格式,社会经济数据为统计报表格式,不同格式的数据在存储方式和读取方法上各不相同。尺度异构方面,监测数据以离散点形式存在,遥感数据以连续栅格覆盖,统计数据以行政区划为单元,三者之间的空间尺度匹配困难。精度异构方面,实验室水质分析数据精度高但样本量少,遥感反演数据覆盖广但存在反演误差,在线监测数据实时性强但稳定性不足,不同数据对同一状态变量的描述精度差异显著。时间异构方面,水文水质数据通常为月、旬或日尺度,气象数据可达小时尺度,遥感数据受卫星过境周期和天气条件限制呈不连续分布,社会经济数据通常为年度尺度,多源数据在时间维度上的对齐难度较大。语义异构方面,不同来源数据对同一概念的定义可能存在差异,例如不同部门对“水质达标”的判定标准不尽相同。

1.3 多源数据融合面临的主要挑战

基于上述特征分析,水环境承载力多源数据融合主要面临以下挑战。第一,尺度不匹配。不同数据在空间和时间尺度上的差异导致直接叠加或比较缺乏合理性,需要建立科学的尺度转换和匹配方法。第二,不确定性传播。每种数据源在采集、传输、处理环节均引入不同程度的不确定性,融合过程中如何量化和控制不确定性的累积是一个关键问题。第三,数据冲突处理。当不同数据源对同一状态变量的描述出现矛盾时(如监测站点水质达标而遥感反演显示异常),需要建立合理的冲突消解机制。第四,动态更新能力。水环境数据具有时序动态特征,融合策略需要具备对新数据的在线吸收能力,而非仅适用于静态数据集。第五,普适性要求。不同湖泊水域的数据基础和监测条件差异较大,融合策略应具有一定的通用性和可迁移性。

2 概率信息处理方法的基本原理

2.1 概率信息处理的理论基础

概率信息处理是一套以概率论为基础的信息处理框架,其核心思想是将所有可用信息转化为概率分布的形式,通过贝叶斯推断等工具实现信息的组合与更新。与水环境多源数据融合的需求相契合,概率信息处理方法具有几个关键的理论优势。(1) 概率框架能够统一表示不同类型的不确定性。在水环境数据中,不确定性既来源于观测误差(偶然不确定性),也来源于模型简化和数据缺失(认知不确定性),概率分布能够同时表达这两类不确定性。(2) 贝叶斯推断提供了增量式信息更新的自然机制。当新的监测数据到达时,可以将当前的概率分布作为先验,通过贝叶斯公式吸收新信息得到后验分布,实现数据融合的动态更新。(3) 概率框架具备冲突信息的合理处理能力。当不同数据源给出不同结论时,贝叶斯方法通过似然函数的设定来权衡各数据源的可靠性,不依赖简单的加权平均或投票机制^[2]。(4) 概率信息处理方法的输出是一个完整的概率分布,而非单一数值,这为后续的承载力评估和风险预警提供了更为丰富的信息基础。

2.2 常用方法

在水环境多源数据融合中,常用的概率信息处理方法包括贝叶斯网络、证据理论和概率主成分分析等。贝叶斯网络是一种基于有向无环图的概率图模型,能够表示变量之间的条件依赖关系,适合处理水环境中水文要素、水质指标与外部驱动因子之间的复杂关联。通过贝叶斯网络,可以将不同来源的观测数据作为证据节点输入,推演目标变量的后验概率分布。证据理论(D-S理论)是概率论的推广形式,能够显式表示对命题的信任区间(置信度与似真度),在处理信息不完全或证据冲突时具有独特优势。概率主成分分析将传统主成分分析拓展至概率框架,在高维水环境数据的降维和特征提取中具有更好的鲁棒性。

3 水环境承载力多源数据融合策略设计

3.1 融合策略的整体架构

基于上述数据特征分析和概率信息处理方法的理论优势,本文提出一套面向水环境承载力评估与预警的多源数据融合策略。该策略包含四个核心环节:数据预处理与标准化、概率特征提取、不确定性量化、概率融合推理。数据预处理与标准化环节负责将原始多源数据进行格式统一、异常检测、缺失填补和尺度转换,形成可进入概率框架的基础数据集。概率特征提取环节将标准化后的数据转化为概率分布形式的特征表达。不确定性

量化环节对每一数据源的不确定性来源进行分类、量化和标定,为融合过程中的权重分配提供依据。概率融合推理环节以贝叶斯推断为核心引擎,将各数据源的概率特征和不确定性估计进行综合推理,输出融合后的状态估计及其置信区间。

3.2 数据预处理与标准化

数据预处理是融合策略的基础层,直接决定后续环节的输入质量。针对水环境多源数据的异构特征,预处理包括以下核心步骤。异常值检测与处理:采用统计方法与规则方法相结合的方式识别异常值,前者基于数据分布的离群程度判定,后者基于水环境指标的物理合理阈值筛选。缺失值填补:针对不同类型缺失数据采用差异化策略——时间序列监测数据采用线性插补与季节性分解相结合的方法;空间覆盖不足的遥感数据采用邻近区域插值或协同克里金方法;偶发性缺失可利用同一周期内相关指标进行回归估计^[3]。尺度统一与转换:空间尺度上,建立监测站点、遥感栅格和统计单元数据间的转换关系,采用面积加权法实现统计数据空间降尺度,采用最近邻或双线性插值实现遥感数据的站点尺度提取。时间尺度上,将日、旬、月、年等不同周期的数据统一插值到标准时间网格。

3.3 概率特征提取

在数据标准化的基础上,将各数据源的观测值转化为概率分布形式的特征表达。对于监测站点数据,观测值及其测量误差直接构成高斯分布。对于遥感反演数据,反演算法通常能够提供误差估计,结合历史验证数据对误差分布进行校准。对于社会经济统计数据,受统计口径和调查方法影响,原始数值可能无法完全反映真实状态,需要引入专家判断或比例因子法将统计数值转化为概率区间。

3.4 不确定性量化

不确定性量化是概率融合的关键环节。水环境数据的不确定性来源可以分为观测不确定性、采样不确定性、模型不确定性和人为不确定性四类。量化方法依据数据类型和不确定性来源而定。对于有明确物理约束或历史比对数据的不确定性来源,采用统计标定方法确定误差分布;对于依赖专家经验的判断型不确定性,采用概率区间赋值方法。最终将每一数据源的不确定性以概率分布参数的形式输出,传递至融合推理环节。在融合过程中,不确定性越小的数据源在贝叶斯推断中获得越高的权重。

3.5 概率融合推理

概率融合推理是整个融合策略的核心引擎。本研究

采用贝叶斯推断作为主要的融合推理工具。其基本框架是：对于目标变量（如水环境承载力综合指数或特定水质指标浓度），预先设定先验分布；当获得来自不同数据源的观测证据时，通过贝叶斯公式依次更新后验分布，最终输出综合所有数据源信息的后验估计及其置信区间。为降低计算复杂度，可采用贝叶斯网络进行变量间条件独立结构的建模，将大规模融合问题分解为局部计算。在数据源之间出现明显冲突时，引入冲突系数检测机制，对冲突程度超过阈值的数据源进行降权或标记待核验处理。

4 融合策略在水环境管理中的应用探讨

4.1 在水环境承载力评估中的应用

在水环境承载力评估中，融合策略的价值体现在两个方面。一方面是特征指标的协同构建。承载力评估需要同时获取水资源量、水环境质量和生态完整性等多维度信息，任何单一数据源都无法完整覆盖。通过概率融合方法，可将监测站点提供的精确水质数据、遥感影像提供的水域面积和植被覆盖信息、气象数据提供的水量补给信息以及统计年鉴提供的人类活动压力信息进行综合，形成对承载力状态的全面描述。另一方面是评估结果的置信度表征。传统评估方法输出单一数值或等级，无法反映因数据缺失或不准确而导致的评估结果不确定性。融合策略输出的概率分布形式评估结果，使管理者能够了解评估结论的可靠程度。

4.2 在水环境风险预警中的应用

在风险预警场景中，融合策略的作用主要体现在三个方面。首先，提高预警信号的可靠性。先期预警模型依赖对未来气象、水文条件的预测数据，这些预测本身存在不确定性。概率融合方法可将预测数据的不确定性纳入预警概率的计算，避免因忽略不确定性而产生的虚假预警或漏报。其次，多指标联合判断。水质超标风险往往由多个因素共同触发，融合策略可综合水量、水质、气象等多源信息进行联合概率推断，识别单一指标分析难以发现的风险信号^[4]。最后，跟踪预警中的实时更新。在风险事件发生后的跟踪预测阶段，融合策略能够将不断接入的新监测数据通过贝叶斯更新机制实时修正预测结果，使预测曲线随着信息积累逐步收敛至真实状态。

4.3 策略的普适性与实施要点

无论针对哪个湖泊水域，水环境承载力评估与预警面临的数据多源异构问题具有共性，该策略的四环节架构具有良好的可迁移性。但在具体实施中，需注意以下几个要点。第一，先验信息的合理设定。贝叶斯融合结果在一定程度上依赖先验分布的设定，应尽量利用历史数据和专家知识构建信息性先验，避免使用无信息先验导致融合结果过度依赖数据。第二，不确定性标定的持续校准。各数据源的不确定性估计需要通过反复验证进行标定，建议定期利用新增实测数据对误差分布进行重新估计，使不确定性标定保持与实际数据质量一致。第三，融合结果的局部验证。建议选取监测条件较好的子区域对融合结果进行独立验证，确认融合策略有效后再向全水域推广应用。

结束语

水环境承载力评估与预警所需数据具有多源异构特征，传统单一数据源方法难以满足精细化管理需求。本文系统梳理了水环境承载力相关数据的基本特征与融合面临的主要挑战，在此基础上设计一套基于概率信息处理的多源数据融合策略，以数据预处理、概率特征提取、不确定性量化和概率融合推理为核心环节，以贝叶斯推断为主要工具。该策略在承载力评估与风险预警场景中展现了良好应用前景，可为水环境大数据管理和综合决策提供方法参考，后续需进一步提升自动化程度与实时计算能力。

课题：2023 年保定市科学技术研究与发展指导计划项目“水资源承载力监测预警关键技术研究”（2340ZZ031）

参考文献

- [1]程丹明.水污染源追踪的多维数据融合方法研究——以宁国市为例[J].生态与资源,2025(5):139-141.
- [2]张念辰,邹士倩,柴泉,等.基于多传感器数据融合的区域水环境污染程度监测方法研究[J].环境科学与管理,2025,50(1):110-113.
- [3]李学科,张文锦,后小龙,等.基于多源监测数据融合的水质评价预警方法及交换共享机制[J].环境保护与循环经济,2024,44(11):68-73,82.
- [4]司源,董飞,廉秋月,等.基于多源监测与数据融合的水质动态评价方法[J].人民黄河,2021,43(2):88-94.