

稳定同位素混合模型研究进展与展望

邓 通^{1,2}, 李 哲², 鲁伦慧², 王力维³, 何汶峰³, 张媛媛^{2,*}

(1. 重庆交通大学 河海学院, 重庆 400074; 2. 中国科学院重庆绿色智能技术研究院, 重庆 400714;
3. 中国长江三峡集团有限公司, 长江经济带生态环境国家工程研究中心, 武汉 430000)

摘 要: 精准辨析污染物来源是实现高质量生态环境发展的前提, 稳定同位素混合模型是精准辨析污染物来源的关键技术手段, 但其发展历程尚未得到系统梳理。文章分析了2001—2024年间该模型的研究趋势和热点, 发现模型本身所带来的结果不确定性在近年来得到了更多关注。其次, 详细阐述了稳定同位素混合模型的发展历程及优缺点, 并提供了评价MixSIAR(Bayesian Mixing Models in R)、SIMMR(Stable Isotope Mixing Models in R)模型性能的数学框架, 评估结果表明两种模型在应用中数据正态分布均可降低结果不确定性, 来源个数减少亦会降低精度, 光谱特征指数均不宜单独作示踪剂, 同时建议在实际应用中应根据两种模型实际运行效率需求合理收集和整理先验信息。最后, 评价结果显示MixSIAR更适合复杂、需要精细调整的生态环境系统分析, 而SIMMR则更适用于简单情况下的快速的环境污染物来源评估。

关键词: 文献计量学; 数学原理; 评价分析; 稳定同位素; 混合模型

中图分类号: X142 **文献标识码:** A **文章编号:** 1672-9250(2025)06-0900-13

近年来, 稳定同位素分析技术在环境生态学领域发展迅猛, 为解决复杂污染物来源的环境问题提供了强有力的手段^[1]。由于同位素在生物地球化学过程中存在分馏现象, 因此可通过对不同环境和生物样品中稳定同位素比值的测量来推断它们的起源、历史过程以及环境条件^[2]。稳定同位素分析技术是基于稳定同位素分馏作用, 通过测量样品中特定元素的稳定同位素比值, 以探究生态系统能量流动、气候变化历史、生物地球化学循环等重要问题^[3]。

通过对自然过程的深入研究, 逐渐形成稳定同位素分析技术和数学建模相结合的方法以描述同位素在不同生物地球化学过程中的分布和变化, 并在此基础上逐渐形成稳定同位素混合模型。该模型具有普适性, 能有效且通用于不同环境介质中的物质来源解析。其中, IsoSource、MixSIR(Mixture Stable Isotope in R)及SIMMR(Stable Isotope Mixing Models in R)等模型的应用, 不仅极大地推动了相关学科的研究进展, 也为跨学科研究提供了新的视角和方法。例如, 利用稳定同位素混合模型确定不同营养级食物来源及构建食物网^[4]、

明确植物体内水分的不同来源及确定环境中污染物来源等^[5-8]。稳定同位素混合模型的核心价值在于其能够提供关于物质来源和过程的直接证据, 所以其被广泛用于识别和追踪污染物的来源。同时, 该模型可有效评估不同污染源对环境系统的影响^[9], 进而为污染控制和环境修复提供关键科学依据。

目前, 在生态系统中实现污染物溯源是实现污染治理的前提, 因此利用稳定同位素混合模型对污染物溯源是目前污染治理的热点和难点^[10]。随着稳定同位素混合模型的发展, 其本身的精度已有大幅度提升, 但仍存在模型先验建立不明确、模型数据筛查缺失等问题, 所以亟须对这些已有的模型研究进行系统地梳理和总结。

1 稳定同位素混合模型研究趋势

文献计量方法是指在计算机技术辅助下, 利用清晰明确的知识图谱来可视化文献分析结果, 主要用于描述和分析学科的研究热点及学科动态^[11]。本文利用CiteSpace进行相关分析, 选取Web of Science(WOS)与中

收稿日期: 2024-11-18; 改回日期: 2025-02-27

基金项目: 国家重点研发计划项目(2022YFC3203504); 国家自然科学基金项目(42107273、U2340222); 中国长江三峡集团有限公司科研项目(202403005); 重庆市自然科学基金面上项目(CSTB2024NSCQ-MSX0443)。

第一作者简介: 邓 通(1997—), 男, 硕士研究生, 主要从事河流生态要素循环。email: dengtong@cigit.ac.cn。

*通信作者: 张媛媛(1991—), 女, 博士, 助理研究员, 研究方向为水库地球化学过程。email: zhangyuanyuan@cigit.ac.cn。

© Editorial office of Earth and Environment. This is an open access article under the CC BY-NC-ND license.

国知网(CNKI)作为数据来源,采用索引格式为SCI-EXPANDED、SSCI。选择检索策略为TS=("isoerror" OR "ISOCONC" OR "ISOSOURCE" OR "MIXSIR" OR "SIAR" OR "SIMMR" OR "MIXSIAR"),文献类型限定为"Articles OR Review Articles",对时间跨度不设置要求。但由于稳定同位素混合模型的最早应用时间为2001年,之前均为简单的算术运算,故实际检索时间跨度为2001年1月至2024年12月。WOS检索结果共获得1 478篇文献,经筛查后得1 025篇文献。CNKI检索结果共获得206篇文献,经筛查后得131篇文献。

1.1 文献发表数量及趋势

文献发表数量变化在一定程度上反映其领域的发展速度以及学术界对其的关注程度^[12]。2001年1月至2024年12月,WOS与CNKI中关于稳定同位素混合模型发文1 156篇,呈现逐年稳步增长的态势(图1)。其中,IsoSource模型、SIAR(Stable Isotope Analysis in R)模型及MixSIAR模型的应用广泛,并随着时间变化稳定同位素混合模型的应用出现明显的增长态势。同时,依据发文量变化和模型原理类型,稳定同位素混合模型的发展大致可分为3个阶段:2001年之前,该时期文献数量较少,多以使用简单线性混合模型为主,为第Ⅰ阶段;2001—2008年,年均发文量为7篇,多以使用多元线性混合模型为主,为第Ⅱ阶段;2008年至今,年均发文量为74

篇,多以使用基于贝叶斯算法的混合模型为主,为第Ⅲ阶段。

1.2 关键词分析

关键词是研究型论文的高度精炼及其中心思想的集中体现,通过对热点高频词汇的研究可以快速了解领域的研究热点^[13]。在CiteSpace中以"Keyword"为分析目标,选择时间跨度为2001—2024年,时间切片为1年。中介中心性是可以衡量节点在知识图谱网络中重要程度的关键指标之一,节点的中介中心性值越高,表明其在图谱中的重要性越大。当节点的中介中心性超过0.1时,可认为该节点在整个网络结构中扮演着连接其他节点的枢纽角色。因此,本文对所得的关键词去除检索词及词义相近的,从而筛选得到出现频次及中介中心性前10的关键词,如表1所示。

由表1可知,稳定同位素混合模型的研究主要针对碳氮(C/N)稳定同位素,主要因为其可以提供食物链长度、生态系统能量流动路径和物质循环过程等方面的信息。例如, $\delta^{13}\text{C}$ 值常用来分析消费者食物来源和确定生态系统中不同植被类型的分布^[14-15], $\delta^{15}\text{N}$ 值常用来确定生物在食物网中的营养位置及营养级别^[16]。在众多的关键词中,"River"及"Organic matter"呈现出较高的出现频次和中介中心性,且二者频繁共现,进一步证实了二者之间的紧密关联,即在河流生态系统中有机质的来

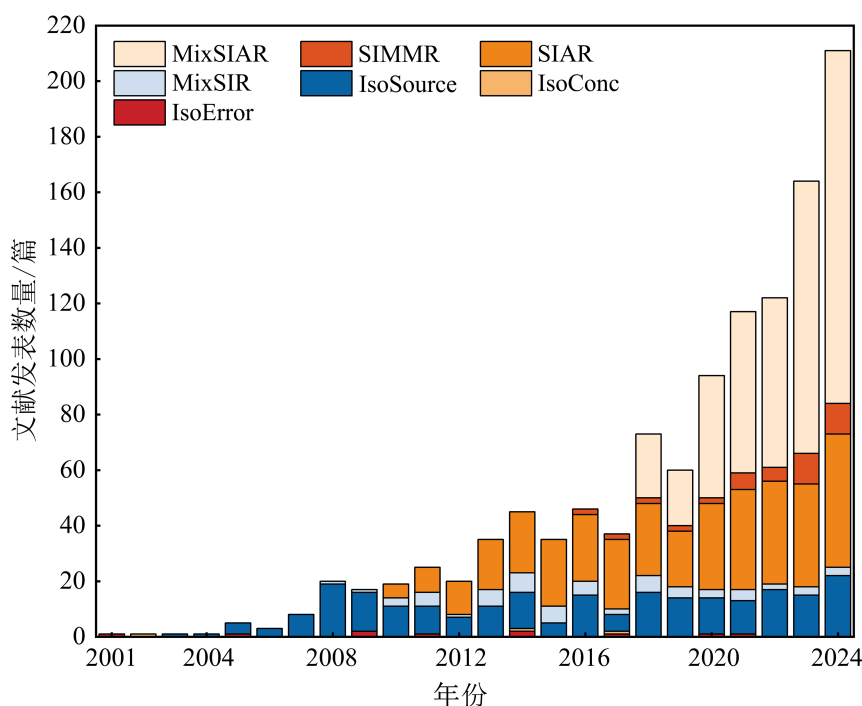


图1 2001—2024年稳定同位素混合模型发文趋势

Fig. 1 Number of publications on stable isotope mixing models from 2001 to 2024

源、分布和转化过程是稳定同位素混合模型应用的重要场景,该模型能够有效揭示河流有机质的动态变化及其与生态系统功能之间的关系。此外,“Fractionation”也呈现出较高的出现频次和中介中心性,表明稳定同位素混合模型的发展正逐步关注到模型输入数据等问题对结果不确定性的影响,尤其是稳定同位素在从源头到测量过程所存在的分馏效应等误差问题。这一问题的探讨不仅涉及模型构建本身,还涵盖样品采集、测量技术以及数据处理等多个环节,需要整合生态、物理、数学等多个学科领域的理论与方法。同时,这也预示着未来稳定同位素混合模型研究将更加注重模型的精细化与精

准化,通过跨学科的融合与创新解决稳定同位素分馏等所导致的模型结果不确定性问题,从而提升模型输出结果的精准性。

2 稳定同位素混合模型的发展历程

根据文献计量所划分的不同发展阶段可以发现稳定同位素混合模型在不同的阶段都具有其阶段性的特点。第Ⅰ阶段为简单线性混合模型阶段(2001年之前),受限于当时的计算能力,研究聚焦于简单线性模型;第Ⅱ阶段为多元线性混合模型阶段(2001—2008年),模型得益于计算机技术飞速发展开始考虑各种不确定性因素,研究内容拓展至复杂生态过程及精细数据处理方法;第Ⅲ阶段为贝叶斯混合模型阶段(2008年至今),基于贝叶斯框架的模型显著提升分析准确性和可靠性,其研究视野不再局限于模型优化,而是逐渐融合多学科前沿理论与方法(图2)。

2.1 简单线性混合模型

20世纪70年代前,物质的源解析基本处于定性分析和半定量分析的阶段,其研究手段主要分为显微镜法和化学法。自20世纪70年代起,化学质量平衡模型(Chemical Mass Balance,CMB)被引入大气气溶胶源解析领域^[17],该方法的早期应用主要基于独特的化学物质作为示踪剂定量解析各类污染源对大气气溶胶的贡献率。然而,由于稳定同位素具有在特定来源中组成确定、化学保守性及安全等特点而逐步成为主要的示踪剂^[18-20],并在CMB基础上形成稳定同位素混合模型。

表1 文献中前十的高频且高中介中心性关键词

Table 1 Top ten high-frequency and high betweenness centrality keywords in the literature

序号	高频关键词	中介中心性
1	Nitrogen	Nitrogen
2	Carbon	Food web
3	Food web	Carbon
4	Diet	Organic matter
5	Organic matter	Fractionation
6	Groundwater	River
7	Fractionation	Community
8	Soil water	Crustacea
9	River	Dynamics
10	Fish	Ecosystem

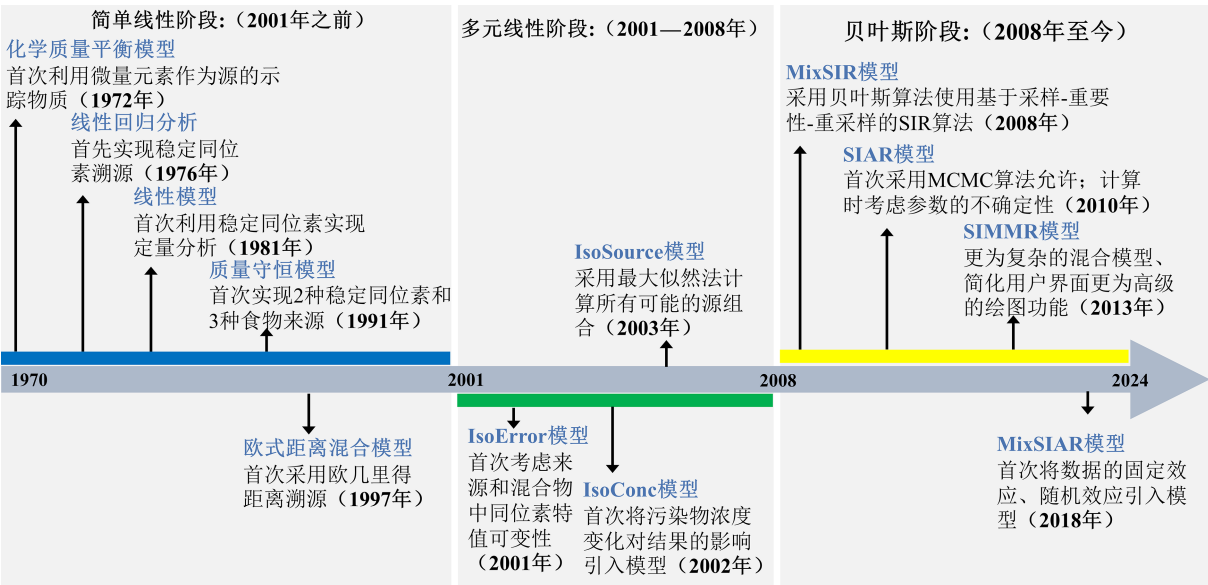


图2 主要稳定同位素混合模型发展历程

Fig. 2 Development history of main stable isotope mixing model

利用稳定同位素混合模型溯源的研究可以追溯到1976年Haines利用C4植物、C3植物、盐沼土与海洋中招潮蟹的碳稳定同位素比值($^{13}\text{C}/^{12}\text{C}$)构建的稳定同位素线性回归模型确定其食物来源^[21]。

最初线性方程受限于稳定同位素的测定技术,仅能对1种稳定同位素和2个来源进行定量计算。随着稳定同位素测量技术的发展,采用多稳定同位素针对多个来源的溯源混合模型逐步出现,主要分为基于CMB进一步得到的质量守恒模型(Mass Conservation Model)和基于欧几里得度量的欧式距离混合模型(Euclidean Distance Mixing Model)^[22-23]。然而,由于距离混合模型忽略质量平衡,导致低估不同来源的贡献及较大的结果不确定性,因此稳定同位素混合模型后期发展均以质量守恒模型为基础。

2.2 多元线性混合模型

简单线性混合模型受限于计算能力仅能够实现点估计,且并未将来源的不确定性和混合物中稳定同位素的可变性、差异性和同位素测试实验误差等问题纳入模型考虑。针对以上简单线性混合模型不能解决的问题,在21世纪初出现了IsoError(2001)、IsoConc(2002)及Iso-Source(2003)3种多元线性混合模型。其中,Phillips等^[24]提出的IsoError模型开创性地解决了来源与混合物中稳定同位素的差异性与不确定性问题。然而,由于不同来源间的同位素组成差异往往导致各元素浓度呈现明显异质性。为此,IsoConc模型创新性地引入源浓度参数作为权重系数,通过加权计算来降低元素浓度差异对源贡献率估算的干扰。但该模型在源浓度难以确定的场景难以实现应用^[25]。最后,Phillips等^[26]在质量平衡方程的基础上进一步开发提出IsoSource模型,其以多元线性混合模型为基础,采用增量叠加的方式运算出所有可能的贡献比例(总和为1),再将每1个组合的加权平均值与实际混合物稳定同位素比值相比较,当差异小于容差设定值时认为是可能的相对贡献,再通过对其频率分析确定可能的贡献比例组,公式如下:

$$N = \left\lceil \frac{(100/i) + (s-1)}{s-1} \right\rceil = \frac{[(100/i) + (s-1)]!}{(100/i)!(s-1)!} \quad (1)$$

式中, i 为源增量参数,一般为1%或2%; s 为来源的个数; N 为组合数量。容差设定值为0.01%~0.1%。虽然Iso-Source模型可以用于计算来源的贡献比例,但是模型采用的是1种最大似然法的分析方法,在模型分析过程中必须假设来源的元素浓度、元素同化率一致^[27],且模型输出结果是1个频率较高的范围而不是1个具体的值,导致模型结果的不确定性较大。

2.3 贝叶斯混合模型

贝叶斯混合模型主要包括MixSIR模型、SIAR模型、SIMMR模型及MixSIAR模型^[28-31]。贝叶斯混合模型相对于线性混合模型能够整合数据及来源不确定性,并通过来源贡献比例的先验信息实现不同参数权重的重新分配^[32]。贝叶斯混合模型主要是在贝叶斯统计框架的基础上依据质量守恒模型使用统计分布的方法来表征混合物中不同来源的贡献比例,并通过考虑稳定同位素的变异性、元素浓度变化及来源等因素来降低模型的不确定性^[33],其模型构建及评价流程如图3所示。同时,贝叶斯混合模型的输出结果是以概率分布的形式呈现,输出结果看似与多元线性混合模型的结果类似,但实际上其给出的结果具有相关的概率密度,而多元线性混合模型给出的结果均具有相同的概率^[34]。

贝叶斯混合模型均是通过将来源贡献比例的先验信息和样本信息相结合来构建后验概率分布,再通过后验分布推断各来源对混合物的贡献比例。于贝叶斯抽样的来源贡献比例计算均基于下式^[29]:

$$P(f_q | x) = \frac{L(x|f_q) \times p(f_q)}{\sum L(x|f_q) \times p(f_q)} \quad (2)$$

式中, x 代表混合物中的稳定同位素测量值; $P(f_q|x)$ 为后验概率; $L(x|f_q)$ 为似然函数,它描述了在假设第 q 个来源对稳定同位素值有特定贡献比例 f_q 时所观测到的样本值的概率。 $p(f_q)$ 为模型的先验分布。由于式中的分母理论上是归一化常数,其确保后验分布的总和为1,因此在实际计算中主要关注分子部分,即似然函数 $L(x|f_q)$ 。似然函数的计算涉及到将每个来源的贡献比例与其特定的稳定同位素值分布和分馏系数结合形成1个描述混合物中稳定同位素值整体分布的联合分布函数,并通过这种方法量化不同来源对混合物的贡献及评估模型结果的不确定性。

(1) MixSIR模型

Moore等^[29]于2008年率先开发出基于采样重要性-重采样算法(Sampling-Importance-Resampling, SIR)的贝叶斯混合模型MixSIR。MixSIR模型主要改变模型的后验分布阶段(图3),其具体实现过程主要分为采样、计算权重与重采样,模型会随机生成不同于后验分布 f_q 且易于采样的分布 $g(x)$,并在 $g(x)$ 中随机抽取 N 个样本组成备用样本池 $\{x_1, x_2, \dots, x_N\}$ 。同时,模型会计算每个样本的权重系数 w_i ($w_i = f_q/g(x)$)和权重系数的归一化权重 W_i ($W_i = w_i / \sum_{j=1}^N w_j$)。随后,模型会从 N 个样本中分别以概率 W_i ($i=1, 2, 3, \dots, N$)有放回的抽取

$n(n < N)$ 个样本。再归一化得到每个样本的权重 W'_i ($W'_i = w_{x_i} / \sum_{j=1}^N w_{x_j}$), 根据归一化后的权重, 从原始样本集中进行重新采样产生 1 组新的样本 ($n_1 < n$), 依次往复。当 $N/n_k (k=1, 2, \dots, i) \rightarrow \infty$ 时, 样本 $\{x_1^*, x_2^*, \dots, x_n^*\}$ 将服从后验分布 $f(q)$ [36]。

MixSIR 模型通过考虑多个来源的不确定性, 包括不同来源的变异性、同位素分馏以及测量误差来估计混合物中各个来源的贡献比例。该模型明确指出不确定性与源、分馏和稳定同位素特征的关系, 同时在分析过程中加入先验信息以提高分析的准确性。MixSIR 模型通过贝叶斯框架能够提供关于来源贡献的概率分布, 从而更全面地反映数据的不确定性 [29]。但 MixSIR 模型未考虑到数据误差等不确定性因素, 所以导致其所得到的结果存在置信区间过窄的问题 [37]。

(2) SIAR 模型

Parnell 等 [30] 于 2008 年正式提出了基于马尔科夫链蒙特卡洛算法 (Markov Chain Monte Carlo, MCMC) 的贝叶斯混合模型 SIAR, 并于 2010 年概述出完整的 SIAR 模型的理论框架。SIAR 模型与 MixSIR 模型在实现模型拟合的方式上存在相对较小的差异, 但 SIAR 模型中考虑了 MixSIR 模型中所欠缺的预测值与观测值之间的残差项, SIAR 模型通过残差项将模型拟合中的不确定性, 包括随机测量误差和同位素分馏等因素纳入考虑范围 [30]。

SIAR 模型主要改变模型的后验分布阶段 (图 3), 其在模型运行的采样方法上采用的 MCMC 方法可以分为马尔可夫过程和蒙特卡洛积分两个阶段。其中, 马尔

可夫过程是考虑存在 1 个连续的随机过程且假设每个 X_i 均在空间 θ 中取值 [38]。如果其条件分布函数 [39]:

$$P(X_h | X_s, s \leq t) = P(X_h | X_t), h > t \quad (3)$$

则称为 1 个马尔可夫过程。但如若是 1 个离散的过程, 则需要满足:

$$P(X_h | X_t, X_{t-1}, \dots) = P(X_h | X_t), h > t \quad (4)$$

而处于离散过程的马尔可夫过程即为马尔科夫链, 对于某个时刻, 马尔可夫链中 $X_h (h > t)$ 的值只是关于当前状态的函数, 经过大量的模拟步骤后马尔科夫链会逐渐忘记其初始状态并最终收敛到静态分布, 从而得到理想的目标分布 [40]。通过马尔科夫链从空间 θ 中某一点出发进行随机抽样模拟, 舍弃不稳定的 m 个样本, 利用后 $n-m$ ($n > m$) 个样本进行统计推断, 从而可以得到 1 个稳定的点分布 $\{X_i; i=m+1, \dots, N_n\}$ [38]。再根据大数定律, 样本均值依概率收敛于后验分布 $f(q)$ 的期望 [39]:

$$\frac{1}{n-m} \sum_{i=m+1}^n f(q_i) \xrightarrow{P} E(q | y) \quad (5)$$

SIAR 模型利用 MCMC 算法解决了 SIR 采样方法无法从复杂后验分布中进行有效采样的问题, 并允许其在多维参数空间中探索模型参数的不确定性, 生成非独立但能反映目标分布特性的样本, 并通过对样本的后处理来降低自相关性, 从而使得对混合来源的贡献比例进行准确估计成为可能。虽然 SIAR 加入了元素浓度的校正, 允许各种参数的不确定性, 但其缺乏同时结合浓度依赖和个体水平估计的能力 [30]。

(3) SIMMR 模型

Parnell 等 [41] 于 2013 年提出的 SIMMR 模型是在

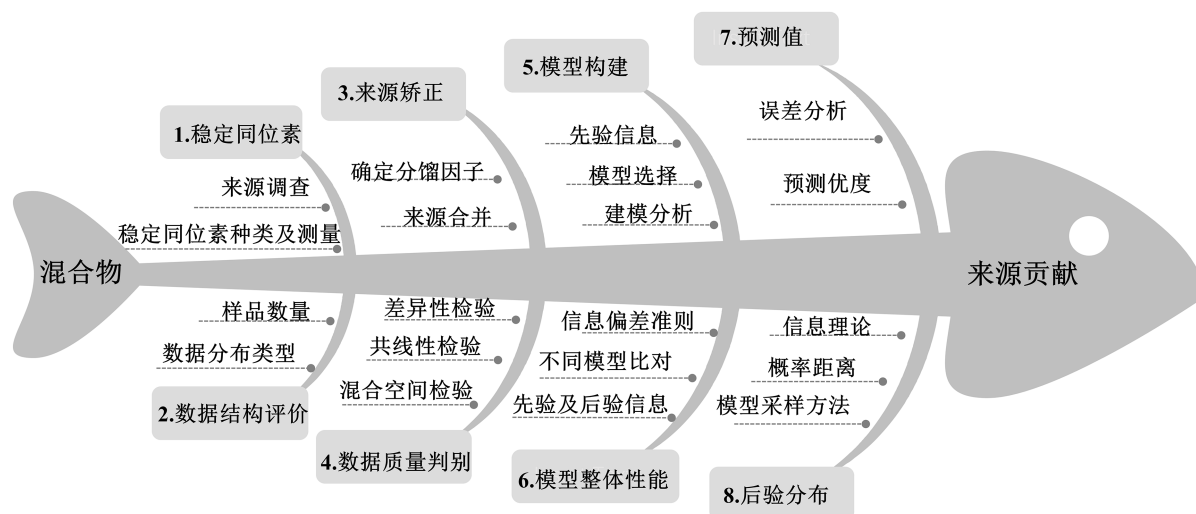


图3 基于贝叶斯算法的稳定同位素混合模型应用及评价流程 [35]

Fig. 3 Usage and evaluation processes of stable isotope model based on Bayesian algorithm [35]

SIAR模型的基础上增加复杂的数学框架、对多同位素和多营养级数据的支持、整合不确定性的贝叶斯方法、灵活的数据结构处理能力、自定义先验分布的功能、模型比较与验证工具以及高级绘图功能而得到。其中, SIMMR模型的主要数学创新在于其集中化对数比转换(CLR), 该转换通过处理比例数据来简化模型并解决共线性问题。此外, SIMMR模型能够灵活的处理源数据的不确定性和浓度依赖效应, 允许用户自定义先验分布, 并提供了全面的模型诊断工具, 包括Gelman-Rubin统计量和后验预测检查, 以确保模型的准确性和可靠性。

SIMMR模型主要改变模型的模型构建及后验分布阶段(图3), 其中集中化对数比转换(CLR)主要用于处理先验分布的设置。这是因为成分数据通常受到闭合性约束, 比如混合物来源的所有比例之和为1。这种数据的特点是存在线性依赖, 直接对其进行建模会遇到统计上的挑战。CLR将原始的比例数据转换为可以进行常规统计分析的形式。设有1个成分数据向量 $x=(x_1, x_2, \dots, x_D)$, 其中各成分之和为常数 C (SIMMR模型为来源比例之和)。CLR转换首先计算所有成分的几何平均值 $g(x)$, 然后对每个成分 x_i 计算其相对于几何平均值的比率 $y_i = \frac{x_i}{g(x)}$, 并取自然对数得到 $z_i = \ln(y_i)$ 。这种转换有助于消除成分之间的冗余信息从而改善数据的分布特性, 便于进行比较和分析, 并使数据更接近正态分布, 更适合后续的统计分析和建模工作。

同时, SIMMR模型在模型运行效率上可以选择采用固定形式变分贝叶斯(Fixed Form Variational Bayes Algorithm, FFVB)与MCMC方式。FFVB能够快速直接估计模型后验分布参数, 避免了样本自相关性的问题, 并减少了对马尔可夫链收敛性分析的需求。由于贝叶斯推断通常需要计算模型后验分布的期望, 然而贝叶斯公(式2)因为归一化常数通常是未知的所以难以计算模型后验分布期望值。故SIMMR模型首次采用FFVB方法将推断模型后验分布转化为优化模型后验分布, 通过最大化证据下界(ELBO)来找到最佳近似后验分布。

(4) MixSIAR模型

MixSIAR模型融合了自MixSIR以来贝叶斯混合模型理论几年来的进步^[28], 其主要改进了贝叶斯混合模型构建阶段(图3)。贝叶斯混合模型通过贝叶斯框架整合了多同位素和多营养级的数据, 使其能够同时处理多个营养级和多种同位素。该模型引入连续协变量的线性回归, 允许模型估计协变量对源贡献比例的影响并表

达为线性模型的形式, 其中源比例作为协变量的函数。此外, MixSIAR通过改进误差结构, 考虑残差误差和过程误差的协方差, 更准确地模拟混合物稳定同位素值的方差。该模型还允许贝叶斯模型平均, 通过比较不同模型的留一法交叉验证(LOO)或广泛适用的信息准则(WAIC)来评估模型的预测准确性并计算模型权重, 从而在最终预测中考虑多个模型的贡献。MixSIAR能够有效处理小样本量问题, 通过使用先验分布和后验分布来稳定参数估计, 提高模型的稳健性。

在多数情况下, 将协变量数据(解释变量或自变量)纳入贝叶斯混合模型可以回答有关混合的重要问题, 例如可以提高模型的解释力和预测准确性, 同时控制潜在的复杂变量并降低模型的不确定性等^[42-43]。由于模型均假设所有混合物均为独立同分布(Independent and Identically Distributed, IID)^[44], 所以忽略包括与混合比例相关的协变量将可能导致伪复制。故MixSIAR模型主要创新为模型构建阶段(图3), 其在模型构建中能够将固定效应和随机效应作为协变量来解释混合比例的可变性, 并通过信息标准计算多个模型的相对支持度^[33]。首先MixSIAR模型会根据所提供的来源贡献比例的先验信息修改的狄利克雷先验建立了1组源的贡献比例 p ^[28]。当 p 确定以后, 这些比例将被等距对数比(ILR)转换为ILR空间参数 β_0 ^[41], 通过此变换将 k 维Aitchison-simplex组合等距映射到 $k-1$ 维欧几里得向量空间中。转换后的每1个 β_0 都是正态分布且相互独立的可以通过协变量(β_i)的合并进行修改转换为贡献比例 p_i ^[28], 如下式:

$$p_i = \text{ILR}^{-1}(\beta_0 + \beta_i) \quad (6)$$

MixSIAR模型虽然功能强大, 但仍存在一些限制, 包括对分馏因子选择的敏感性、狄利克雷先验在多源情况下对数据质量和数量的依赖以及模型设定和参数化的挑战。此外, 模型需要较多的计算资源且结果的解释需要一定的统计知识。因此, 在使用MixSIAR模型时需要通过适当的数据预处理和模型验证来提高模型的准确性和可靠性。

整体而言, 稳定同位素混合模型从简单线性混合模型至当前的贝叶斯混合模型经历了显著的发展与改进, 虽然其理论上提供了更精确的解析能力, 但同时也为模型本身引入了新的不确定因素。这些因素可能源自数据质量、模型假设的合理性或计算方法的局限性。因此, 深入探讨这些不确定性来源, 对于确保模型结果的准确性和科学应用的可靠性至关重要。

3 稳定同位素混合模型不确定性来源

在稳定同位素混合模型应用过程中,可以将对稳定同位素混合模型结果具有显著影响的不确定性分为4类:来源不确定性、示踪剂种类不确定性、混合数据结构不确定性以及模型类别不确定性。这些不确定性不仅影响模型的结果准确性,同时也关系到其在污染源应用中的可靠性。因此,深入分析不确定性来源,理解其对模型结果的表现和影响,对于进一步优化模型和提高应用效果具有重要意义。

3.1 来源不确定性

在应用稳定同位素模型时,研究人员通常需要尽可能收集所有的潜在来源,同时模型本身也会对来源缺失十分敏感^[41]。但在模型的分析应用中,来源的增加也会导致模型分析的不确定性显著提升^[33],所以在模型应用过程中需要考虑到来源个数与模型本身不确定性的平衡。故在利用稳定同位素模型时确定合适的来源个数及处理相似来源也是模型获得准确结果的前提。当来源个数超过6种时,模型预测精度通常会出现显著下降,针对该问题目前主要有两类解决方案^[33]:第一类(如MixSIAR、SIMMR)是采用通过先验信息的方法,常通过先验信息对模型中相似来源进行分类组合,从而实现对于模型来源数量的控制,最终达到对于模型不确定性的降低。另一类(如IsoSource)采用先利用稳定同位素混合模型得到所有来源的结果,再通过每个来源的可能的贡献范围,对相似来源进行组合合并,从而实现模型不确定性的降低^[45]。

贝叶斯混合模型具有将样本信息与模型来源贡献比例的先验信息相结合分析的能力,所以当来源贡献比例的先验信息本身误差较小的情况下,模型的来源贡献比例的后验分布结果会先反映先验信息的分布趋势及离散特点,反之先验信息本身具有较大的偏差,则会导致模型后验分布存在更多的不确定性^[46]。

3.2 示踪剂种类不确定性

稳定同位素混合模型的基本假设之一是示踪剂在从源到汇的传输过程中保持含量和性质的相对稳定,所以在利用稳定同位素混合模型对污染物溯源时,示踪剂的选择对结果的准确性有重要影响。然而,由于不同的示踪剂具有不同的环境行为,这将导致模型结果存在不确定性。例如,在利用不同种类示踪剂(^2H 、 ^3H 、 ^{18}O 、 ^{17}O)对植物水分来源解析的不确定性研究中发现示踪剂选择解释了结果的28%的不确定性^[47]。同时,由于考虑经济成本,除稳定同位素以外的示踪剂如光谱特征指数逐渐被应用在混合模型的溯源中,但这些示踪剂的

稳定性、迁移性及生物可利用性等易受到各类环境因子影响的问题并未得到考虑,进而影响稳定同位素混合模型结果的准确性^[48-49]。光谱特征指数在应用于河流DOM溯源时极易受到多种环境因素的干扰^[50-51],例如,水体中pH的变化会影响DOM的光谱特征参数^[52];温度会影响DOM的光化学降解速率和微生物活动;此外,DOM的浓度和组成也会显著影响光谱特征指数^[53],这些因素的综合作用,使得光谱特征指数在有机质溯源的实际应用中存在较大的不确定性。

3.3 混合数据结构不确定性

稳定同位素模型结果的准确性同样也会受到样品数量及代表性的影响。在稳定同位素的各类模型的研究及建立过程中,均采用均值及方差来表征每种来源的同位素特征的不确定性^[1],需要指出稳定同位素模型基于中心极限定理,只有当样本量达到30及以上时,中心极限定理才会成立^[54]。但是由于在实际的研究中,受制于采样时间、采样费用及稳定同位素测试费用等影响,使得样本量不能够满足模型数学分布的假设,导致数据格式出现均匀分布或者离散分布等数据分布形式。同时,小样本量及不同的数据分布结构容易引发均值与方差的估计偏差增大,导致统计量难以保证对总体参数的无偏估计,从而导致稳定同位素模型结果的不确定性增加^[54]。

3.4 模型类别不确定性

从稳定同位素混合模型发展历程可以发现每个稳定同位素混合模型都有各自的优缺点。例如,IsoSource采用最大似然法估计,未充分考虑数据的不确定性,且必须假定食物来源同位素的浓度、同化效率一致,不同组织间无同位素判别值的差异;SIAR模型加入了元素浓度的校正,允许参数的不确定性,但缺乏同时结合浓度依赖和个体水平估计的能力;SIMMR模型只能对同一生态系统的食物来源贡献进行分析,无法处理跨生态系统的情况;MixSIAR模型存在包括对分馏因子选择的敏感性、在多源情况下对数据质量和数量的依赖性,同时其需要较多的计算资源。尽管这些模型为确定污染物来源提供了重要的方法,但不同的方法针对同一问题可能会得到不一样的结果,这是由混合模型的算法和输入数据的不确定性引起的^[55]。例如,在利用不同稳定同位素模型(IsoSource、SIAR、MixSIR、MixSIAR)对植物水分来源解析的不确定性研究中,稳定同位素模型的选择解释了结果37%的不确定性^[47]。所以如何针对合适的环境条件及实验目标选取合适的稳定同位素模型成为能否完成较为精准溯源的第一步。

4 稳定同位素混合模型不确定性评价

针对以上4种稳定同位素混合模型结果不确定性来源,本文提出1个环境污染来源识别的模型评估框架旨在针对混合物数据的数学统计分布,以及在混合物数据结构(如正态分布)确定的前提下,对混合物数据的数量、来源及其先验信息、示踪剂的种类和数量进行系统性评价(图4)。这种方法将有助于探究主要稳定同位素模型(MixSIAR模型及SIMMR模型)在不同场景下的适用性,并为模型选择和结果解释提供更加坚实的科学基础。通过这种评价体系可以更好地界定不同来源的不确定性对模型结果的影响,从而提高模型应用的准确性和可靠性。

以水环境有机质(OM)污染来源识别为例,本文从文献中收集了各种环境来源OM的稳定同位素指标、光谱指标及化学计量比指标作为溯源示踪剂(表2)。同时按照表2所示混合比例生成单个混合物示踪剂值,再利用单个混合物示踪剂值为期望或中间值随机生成各种数学分布。

目前用于衡量模型的拟合程度的评价方法并没有固定的标准。可以利用均方根误差(RMSE)来评估预测值的准确程度,其可表示稳定同位素实测值和预测值之间的平均差异;其值较低表明模型预测结果具有较小的误差和更准确的预测。也可以采用平均绝对百分比

误差(MAPE)来衡量模型拟合程度偏离的相对大小(即百分率)。相对来说,MAPE不容易受极端值的影响,其使用百分率来衡量模型拟合程度偏离的大小,而RMSE采用误差的平方,会放大预测误差,所以对于离群数据更敏感。总体而言,较小的RMSE、MAPE表明模型的性能更好,公式如下^[47]:

$$\text{RMSE} = \left[\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \right]^{1/2} \quad (7)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - O_i}{O_i} \right| \quad (8)$$

其中, n 是样本的数量; O_i 和 P_i 分别代表实际与预测的贡献比率;RMSE的单位为‰。

对于混合数据分布结构,2个模型在正态分布和离散分布数据结构下并未表现出显著差异。然而,当数据遵循均匀分布时,无论是均方根误差(RMSE)还是平均绝对百分比误差(MAPE)均有所增加,这表明在数据分布接近正态分布时,模型的性能往往更为稳定和可靠(图5a)。这说明虽然稳定同位素本身的自然属性符合高斯分布定理,但在不同的数据结构下,稳定同位素模型性能会出现较为显著的波动,所以在实际应用模型时需要较高的样本数量来满足数据分布结构的假设,同时模型实际应用前建议利用数据探索和预处理技术等统计方法来验证数据分布假设的合理性。根据这些分析

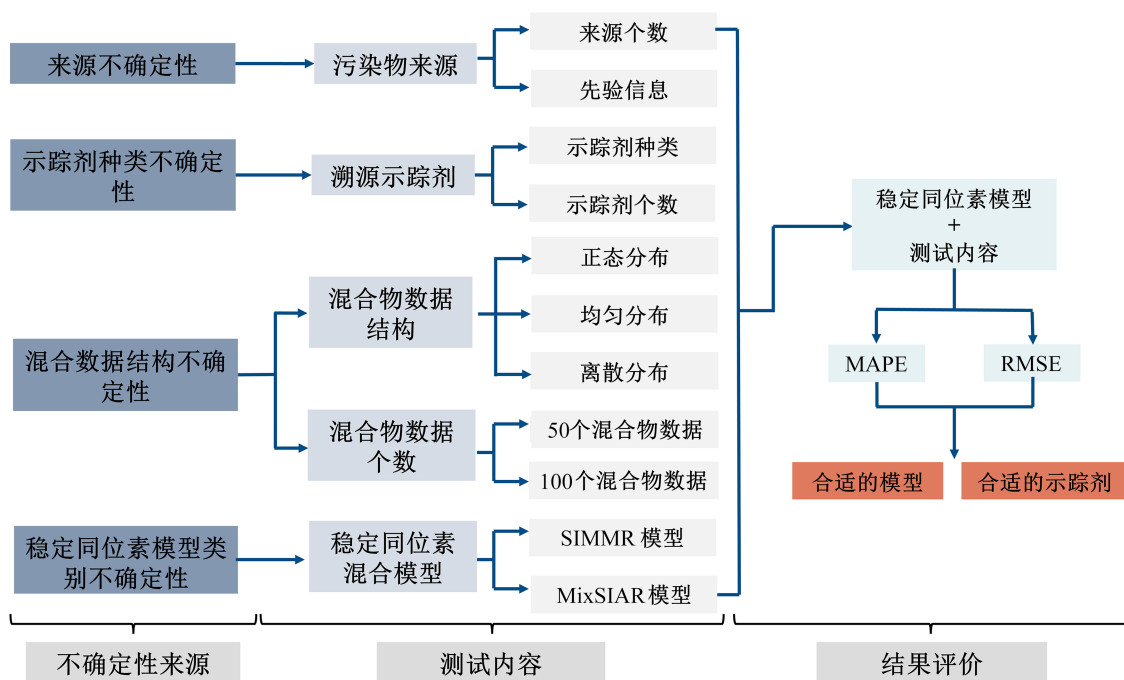


图4 环境污染来源识别的模型评价框架示意图

Fig. 4 Schematic diagram of model evaluation in identification of environmental pollution sources

结果,如有必要,可对数据进行适当的转换,以增强模型的适用性和预测准确性^[64]。对于不同混合物数据个数,2个模型的RMSE与MAPE均相差不大,表明混合物数据个数的变化并不能影响到模型结果(图5b)。

在减少来源个数的情况下,2个模型的RMSE与MAPE均出现了增加(图6a),说明来源个数的减少会降低模型运行的精度。针对先验信息,MixSIAR模型与SIMMR模型的RMSE与MAPE均未出现较大的差异,说明先验信息并不能提高模型运行的精度(图6a);对于不同的示踪剂种类及个数,当使用光谱特征指数作为示踪剂时,MixSIAR模型与SIMMR模型的RMSE与MAPE均呈现出显著的增加。因此,评价结果可以证实光谱特征指数在环境中易受生物地球化学过程的影响,单独将其作为有机质溯源的示踪剂是不适宜的。然而,在利用光谱特征指数、稳定同位素及化学计量比等联合作为示踪剂时,模型的RMSE与MAPE均呈现出降低,这表明多种

示踪剂的联合使用可以提高模型的预测准确性。因此,建议结合稳定同位素和光谱特征指数,以提高有机质溯源的准确性和可靠性。同时,MixSIAR模型的RMSE与MAPE均显著高于SIMMR模型,说明除采用稳定同位素作为示踪剂以外,采用其他指标时SIMMR模型输出结果均优于MixSIAR模型(图6b),这可能归因于SIMMR模型采用的CLR通过处理比例数据、简化模型、解决共线性问题、改善数据分布特性、消除冗余信息以及解决闭合效应等方式,其有助于提升模型的预测准确性和稳定性。

在模型运行时间方面,无论在什么情况下,SIMMR模型运行时间均比MixSIAR模型短,前者以秒计,后者以小时计。在增加混合物数据个数的情况下,两种模型运行时间均呈现倍数增长趋势,其中MixSIAR模型增长更多,可达10倍左右;而在不同数据分布情况下,MixSIAR模型运行时间差异较大,尤其是当数据采

表2 不同来源示踪剂值^[50, 56-63]
Table 2 Tracer values of different sources^[50, 56-63]

来源	$\delta^{13}\text{C}_{\text{OM}}/\text{‰}$	$\delta^{15}\text{N}_{\text{OM}}/\text{‰}$	SUVA ₂₅₄	HIX	BIX	FI	C/N	混合比例	信息先验
植物	-27.69	8.8	0.87	0.71	0.62	1.71	18.31	0.05	0.04
土壤	-28.29	5	3.85	5.45	0.67	1.34	11.38	0.20	0.18
地下水	-23.5	9.12	0.7	3.85	0.87	1.88	30.74	0.05	0.06
猪畜废水	-22.58	11.29	1.58	5.71	0.79	1.65	21.89	0.10	0.12
农业废水	-23.7	5.84	1.71	8.05	0.7	1.49	9.66	0.20	0.16
工业废水	-17.84	10.6	1.9	9.28	0.93	1.68	24.59	0.04	0.06
城市污水	-23.26	14.2	1.66	9.26	0.78	1.64	5.81	0.18	0.18
污水处理厂出水	-25.07	3.9	2.02	0.95	1.21	1.58	0.92	0.18	0.20

注:SUVA₂₅₄—芳香性指数;HIX—腐化指数;BIX—微生物指数;FI—荧光指数;C/N—碳氮化学计量比。

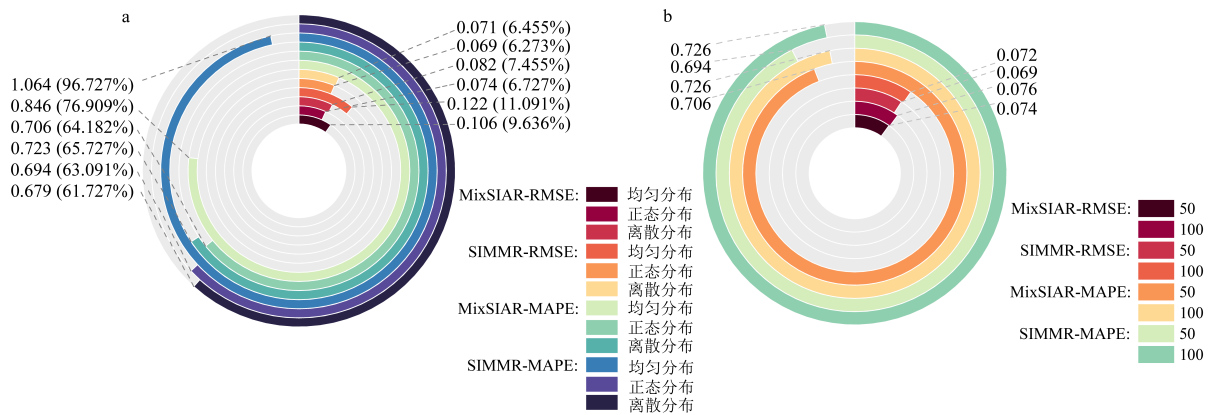


图5 数据结构及混合物数据个数评价结果
Fig. 5 Evaluation results of the data structure and number of mix data

用均匀分布时,其运行时间可达100 h左右(图7a),但SIMMR模型运行的时间变化不大(图7b);针对不同来源数量,两种模型的运行时间均无明显变化;同时,SIMMR模型无论是否具有先验信息,其运行时间均保持稳定,而MixSIAR模型在无先验信息的情况下,运行时间显著增加,最高可达10倍;而不同示踪剂的情况下,示踪剂个数越多,模型运行时间越长,光谱指标能明显

增加两种模型的运行时间,尤其是MixSIAR模型。由于提高样本数量会使得数学分布结构更加满足稳定同位素模型的高斯分布的假设,但样本数量的增加将导致模型运行效率显著降低,所以在实际的应用过程中需要考虑样本数量导致的模型数学分布及模型效率之间的矛盾。同时,先验信息对于模型的运行效率有着极为显著的影响,所以为提高模型的运行效率,在实际的应用中

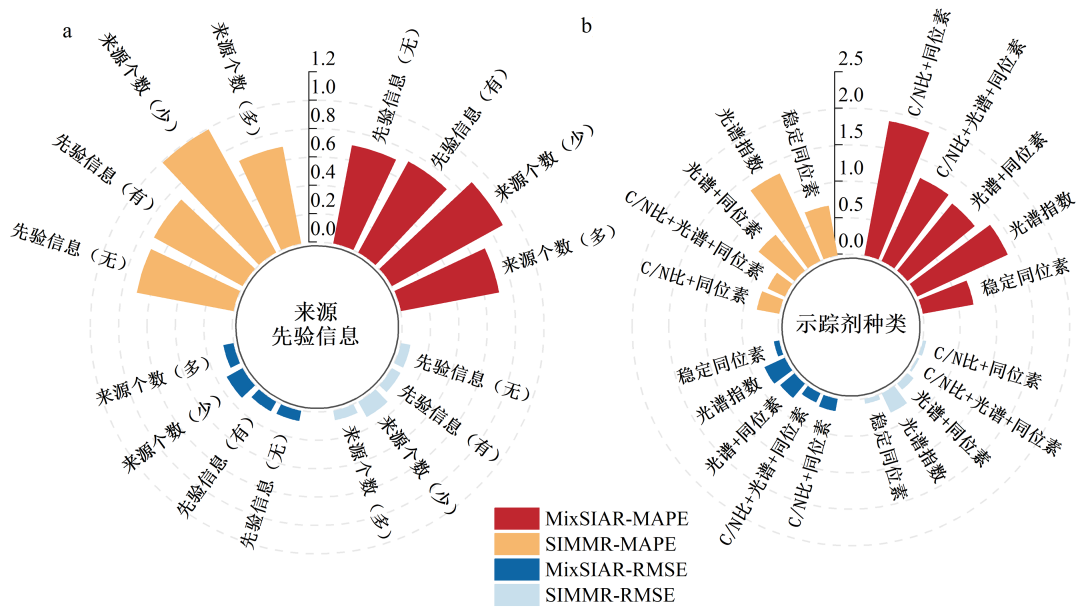


图6 来源、先验信息及示踪剂种类评价结果

Fig. 6 Evaluation results of source, prior information, and tracer types

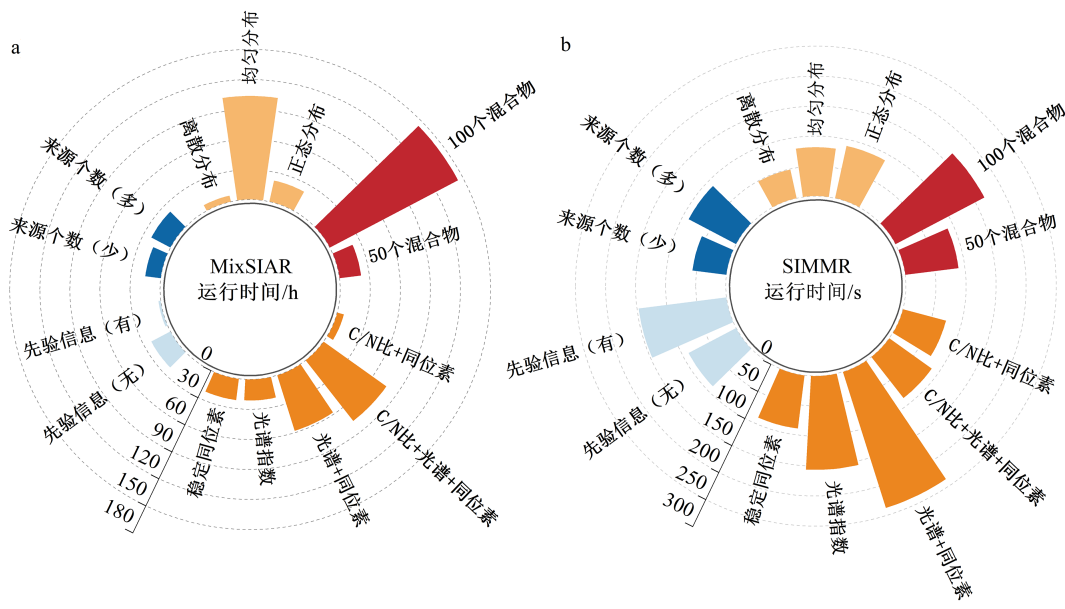


图7 MixSIAR模型及SIMMR模型运行时间

Fig. 7 Running time of MixSIAR and SIMMR

应根据实际需求采用绝对主成分分析、流域模型等方式或共轭先验的方式获取先验信息。

最后,在不同的研究中研究者应结合不同环境特点及模型评价结果,因地制宜的选取模型进行污染物溯源。例如,在多污染源环境中,MixSIAR模型允许用户在模型中包含多个层次(如个体、群体、环境因素等),并支持贝叶斯框架的细致定制。这使得MixSIAR在需要考虑多个潜在影响因素时更为适用,尤其适用于污染源空间分布复杂的情况^[28]。SIMMR模型虽能较好估算多源混合问题中各污染源的贡献比例,但污染源数量过多时性能会受影响,其适用于结构较简单的生态系统,不需要太多复杂的因素建模^[31]。而在样本稀缺场景下,SIMMR模型适用性较高,其CLR变换有助于提升预测准确性和稳定性,而MixSIAR模型此时运行时间和计算复杂度可能增加,效率降低。因此,MixSIAR更适合复杂、需要精细调整的生态系统分析,而SIMMR更适合在较简单的情况下进行快速评估。

5 结论与展望

本文首先利用Citespace软件对相关文献的发表数量及关键词进行了分析,进而系统梳理了稳定同位素混合模型的发展历程,并对其不确定性来源进行了深入探讨,最后提出1个模型适用场景检验的思路,通过不同的来源、示踪剂、混合物数据结构等方式评估对比SIMMR模型与MixSIAR模型的性能。由于稳定同位素混合模型具有局限性,研究者在选取模型进行环境问题分析时更应该注意到模型应用的相关条件,尽可能提升模型的精准性。

1)随着技术的发展,目前多以基于贝叶斯算法的稳定同位素混合模型的应用为主,尤其是SIMMR、Mix-SIAR。在稳定同位素混合模型运行前,由于给定模型的先验信息会影响混合模型结果的分布特征并显著影响到模型的运行效率,所以在应用模型前应根据实际运行效率的需求采用绝对主成分分析、流域模型等方式或共轭先验的方式获取先验信息;同时来源的个数也会影响模型运行的精度,故在确定所有来源后应根据来源的相似性进行来源的组合与合并。

2)稳定同位素混合模型运行时不同种类及个数的示踪剂会导致运行结果呈现出显著差异,由于光谱特征指数在环境中容易受到生物地球化学作用的影响而不适合单独作为稳定同位素模型溯源的示踪剂。同时,增加示踪剂种类及个数可以提高模型精度,所以建议在实际应用中采用多种示踪剂的联合使用。

3)稳定同位素混合模型运行时应对混合物数据结构进行检查,理想情况下,数据应遵循正态分布,若非正态分布则需进行参数检验或数据转换。

4)MixSIAR更适合复杂、需要精细调整的生态环境系统分析,而SIMMR更适合在较简单的情况下进行快速的环境污染物来源评估。

5)稳定同位素混合模型在未来可以考虑融合能够挖掘数据中复杂模式的机器学习等算法来降低模型的不确定性。同时,也可以利用动态建模等技术来考虑时空变化,使模型更加贴合实际场景的动态特性,进而增强模型的场景适用性和预测能力。最后,未来在模型改进时应整合多同位素体系以提供更全面的信息,提高模型分辨率和可靠性。

参 考 文 献

- [1] 赵永松,单秀娟,杨涛,等. 稳定同位素模型在水生生态系统食物网中的研究进展[J]. 渔业科学进展, 2024, 45(1): 1-13.
- [2] 张昊,李建平. 稳定碳同位素在草地生态系统碳循环中的应用与展望[J]. 水土保持研究, 2021, 28(1): 394-400.
- [3] 陈世革,白永飞,韩兴国. 稳定性碳同位素技术在生态学研究中的应用[J]. 植物生态学报, 2002, 26(5): 549-560.
- [4] Vander Zanden M J, Rasmussen J B. Variation in $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ trophic fractionation: implications for aquatic food web studies[J]. Limnology and Oceanography, 2001, 46(8): 2061-2066.
- [5] 杨开丽,刘淑亮,万惠芸,等. 基于氮氧同位素和MixSIAR模型的烟台北部地表水硝酸盐来源分析[J]. 中国地质, 2024, 51(6): 2066-2076.
- [6] Saka D, Antwi E O, Skrzypek G, et al. Tracing sulfate sources in a tropical agricultural catchment with a stable isotope Bayesian mixing model[J]. Science of the Total Environment, 2024, 951: 175502.
- [7] 王刚,程全国,雷坤,等. 飞云江水系氮磷污染及硝酸盐氮来源解析[J]. 环境污染与防治, 2023, 45(4): 486-491.
- [8] 雍亮,冯民权,范锐谦. 长河流域枯水期水体硝酸盐氮解析[J]. 地球与环境, 2023, 51(3): 274-286.
- [9] 文浩龙,杨平恒,华茂松,等. 岩溶地下河硝酸盐转化与来源对比研究[J]. 西南大学学报(自然科学版), 2023, 45(5): 172-184.
- [10] 马文娟,刘丹妮,杨芳,等. 水环境中污染物同位素溯源的研究进展[J]. 环境工程技术学报, 2020, 10(2): 242-250.
- [11] Mou J, Cui Y, Kurcz K. Bibliometric and visualized analysis of research on major e-commerce journals using Citespace [J]. Journal of Electronic Commerce Research, 2019, 20: 219-237.
- [12] 江卉卿,汪洁琼. 基于文献计量分析的河流生境质量评价研究[J]. 城市建筑, 2021, 18(22): 79-83.

- [13] Chen C M. Visualizing scientific paradigms: an introduction[J]. *Journal of the American Society for Information Science and Technology*, 2003, 54(5): 392–393.
- [14] 易现峰. 海北高寒草甸土壤有机碳同位素组成及 C_3/C_4 碳源的变化[J]. *西北植物学报*, 2005, 25(2): 336–342.
- [15] 李伟业, 丁虎, 韩晓昆, 等. 渤海湾西岸入海河流中有机碳来源及其影响因素分析[J]. *地球与环境*, 2023, 51(2): 185–195.
- [16] 王玉玉, 于秀波, 张亮, 等. 应用碳、氮稳定同位素研究鄱阳湖枯水末期水生食物网结构[J]. *生态学报*, 2009, 29(3): 1181–1188.
- [17] Miller M S, Friedlander S K, Hidy G M. A chemical element balance for the Pasadena aerosol[J]. *Journal of Colloid and Interface Science*, 1972, 39(1): 165–176.
- [18] 吴健, 沙晨燕, 李大雁, 等. 稳定碳同位素技术在多环芳烃源解析中的应用进展[J]. *环境科学与技术*, 2021, 44(3): 170–179.
- [19] 孙学丽. 基于稳定同位素技术的食品检验应用研究进展[J]. *食品安全导刊*, 2024(1): 154–157.
- [20] 李哲, 谢宝瑜, 季荣, 等. 稳定同位素分析在昆虫生态研究中的应用[J]. *昆虫学报*, 2004, 47(3): 394–399.
- [21] Haines E B. Relation between the stable carbon isotope composition of fiddler crabs, plants, and soils in a salt marsh[J]. *Limnology and Oceanography*, 1976, 21(6): 880–883.
- [22] Ben-David M, Flynn R W, Schell D M. Annual and seasonal changes in diets of Martens: evidence from stable isotope analysis[J]. *Oecologia*, 1997, 111(2): 280–291.
- [23] Schwarcz H P. Some theoretical aspects of isotope paleodiet studies[J]. *Journal of Archaeological Science*, 1991, 18(3): 261–275.
- [24] Phillips D L, Gregg J W. Uncertainty in source partitioning using stable isotopes[J]. *Oecologia*, 2001, 127(2): 171–179.
- [25] Koch P L, Phillips D L. Incorporating concentration dependence in stable isotope mixing models: a reply to Robbins, Hilderbrand and Farley (2002)[J]. *Oecologia*, 2002, 133(1): 14–18.
- [26] Phillips D L, Gregg J W. Source partitioning using stable isotopes: coping with too many sources[J]. *Oecologia*, 2003, 136(2): 261–269.
- [27] 高春霞, 戴小杰, 田思泉, 等. 基于稳定同位素技术的浙江南部近海主要渔业生物营养级[J]. *中国水产科学*, 2020, 27(4): 438–453.
- [28] Stock B C, Jackson A L, Ward E J, et al. Analyzing mixing systems using a new generation of Bayesian tracer mixing models[J]. *PeerJ*, 2018, 6: e5096.
- [29] Moore J W, Semmens B X. Incorporating uncertainty and prior information into stable isotope mixing models[J]. *Ecology Letters*, 2008, 11(5): 470–480.
- [30] Parnell A C, Inger R, Bearhop S, et al. Source partitioning using stable isotopes: coping with too much variation[J]. *PLoS One*, 2010, 5(3): e9672.
- [31] Govan E, Jackson A L, Inger R, et al. Simmr: a package for fitting Stable Isotope Mixing Models in R[EB/OL]. 2023: 2306.07817. <https://arxiv.org/abs/2306.07817v1>.
- [32] Ellison A M. Bayesian inference in ecology[J]. *Ecology Letters*, 2004, 7(6): 509–520.
- [33] Phillips D L, Inger R, Bearhop S, et al. Best practices for use of stable isotope mixing models in food–web studies[J]. *Canadian Journal of Zoology*, 2014, 92(10): 823–835.
- [34] Newsome S D, Yeakel J D, Wheatley P V, et al. Tools for quantifying isotopic niche space and dietary variation at the individual and population level[J]. *Journal of Mammalogy*, 2012, 93(2): 329–341.
- [35] 祝孔豪, 李斌, 王康, 等. 稳定同位素质量平衡混合模型的性能评估[J]. *水生生物学报*, 2022, 46(3): 427–438.
- [36] Smith A F M, Gelfand A E. Bayesian statistics without tears: a sampling-resampling perspective[J]. *The American Statistician*, 1992, 46(2): 84.
- [37] Hopkins J B 3rd, Ferguson J M. Estimating the diets of animals using stable isotopes and a comprehensive Bayesian mixing model[J]. *PLoS One*, 2012, 7(1): e28478.
- [38] Wu H Y, Chen Y, Cai Z N, et al. Gradient pyramid mechanism and Nelder–Mead simplex enhanced Colony Predation Algorithm for optimal power flow problems[J]. *Energy Reports*, 2024, 11: 2901–2920.
- [39] 汪嘉冈. 现代概率论基础[M]. 2版. 上海: 复旦大学出版社, 2005.
- [40] Gilks W R, Richardson S, Spiegelhalter D. Introducing Markov chain Monte Carlo[M]//Markov Chain Monte Carlo in Practice. London: Chapman and Hall/CRC, 1995: 19–38.
- [41] Parnell A C, Phillips D L, Bearhop S, et al. Bayesian stable isotope mixing models[J]. *Environmetrics*, 2013, 24(6): 387–399.
- [42] Francis T B, Schindler D E, Holtgrieve G W, et al. Habitat structure determines resource use by zooplankton in temperate lakes[J]. *Ecology Letters*, 2011, 14(4): 364–372.
- [43] Ogle K, Tucker C, Cable J M. Beyond simple linear mixing models: process–based isotope partitioning of ecological processes [J]. *Ecological Applications*, 2014, 24(1): 181–195.
- [44] Hurlbert S H. Pseudoreplication and the design of ecological field experiments[J]. *Ecological Monographs*, 1984, 54(2): 187–211.
- [45] Ward E J, Semmens B X, Phillips D L, et al. A quantitative approach to combine sources in stable isotope mixing models[J]. *Ecosphere*, 2011, 2(2): art19.
- [46] Choy S L, O’Leary R, Mengersen K. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models[J]. *Ecology*, 2009, 90(1): 265–277.
- [47] Gai H Q, Shi P J, Li Z. Untangling the uncertainties in plant water source partitioning with isotopes[J]. *Water Resources Research*, 2023, 59(12): e2022WR033849.
- [48] 刘靳, 涂耀仁, 段艳平, 等. Cu同位素示踪技术应用于环境领域的研究进展[J]. *环境保护科学*, 2020, 46(2): 85–92.
- [49] 冯虎元, 安黎哲, 王勋陵. 环境条件对植物稳定碳同位素组成的影响[J]. *植物学通报*, 2000, 17(4): 312.

- [50] Kim M S, Lim B R, Jeon P, et al. Innovative approach to reveal source contribution of dissolved organic matter in a complex river watershed using end-member mixing analysis based on spectroscopic proxies and multi-isotopes[J]. *Water Research*, 2023, 230: 119470.
- [51] 徐爱喆, 韩晓昆, 刘铭轩, 等. 渤海湾两条河流及其近岸海域水体中DOM的光谱特征及影响因素[J]. *地球与环境*, 2022, 50(4):526–536.
- [52] 于莉莉, 钟晔, 孙福红, 等. pH值对滇池水体溶解性有机质(DOM)光降解作用的影响[J]. *光谱学与光谱分析*, 2019, 39(8): 2533–2539.
- [53] 傅芳敏, 姚爱林, 明玥, 等. 2022年6月珠江流域洪水对下游溶解有机物浓度和组成的影响[J]. *海洋与湖沼*, 2024, 55(2): 397–408.
- [54] Kwak S G, Kim J H. Central limit theorem: the cornerstone of modern statistics[J]. *Korean Journal of Anesthesiology*, 2017, 70(2): 144–156.
- [55] Jackson A L, Inger R, Bearhop S, et al. Erroneous behaviour of MixSIR, a recently published Bayesian isotope mixing model: a discussion of Moore & Semmens (2008)[J]. *Ecology Letters*, 2009, 12(3): E1-E5.
- [56] 孙太虎. 基于稳定同位素的河流硝酸盐来源定量解析[D]. 上海: 华东师范大学, 2023.
- [57] 黄建伟. 植物叶片C、N、P化学计量特征及影响因素: 以中国22个城市为例[D]. 开封: 河南大学, 2023.
- [58] 张俊萍. 利用N、O同位素技术研究太湖运河中氮的分布特征及来源[D]. 淮南: 安徽理工大学, 2015.
- [59] 王晓琳, 王寅, 何奕成, 等. 黑河下游植物碳氮化学计量特征及其影响因素[J]. *北京林业大学学报*, 2023, 45(4): 50–59.
- [60] 李娟, 马忠明, 王文丽, 等. 小麦秸秆调节猪粪碳氮比对产沼气的影响[J]. *中国沼气*, 2012, 30(4): 30–32.
- [61] 熊杏, 熊清华, 郭熙, 等. 南方典型丘陵区耕地土壤全氮、有机碳和碳氮比空间变异特征及其影响因素[J]. *植物营养与肥料学报*, 2020, 26(9): 1656–1668.
- [62] 陈强, 王伟, 汪传新, 等. 改良分段进水工艺处理低C/N城市污水的体积优化控制[J]. *水处理技术*, 2015, 41(4): 71–76.
- [63] 杨明雪, 高娜, 李凌云, 等. 食品工业废水委托水质净化厂处理的实践与探究[J]. *中国给水排水*, 2023, 39(9): 122–126.
- [64] 王康, 祝孔豪, 郭钰伦, 等. 基于稳定同位素的消费者营养来源溯源的方法和过程[J]. *水生生物学报*, 2022, 46(5): 767–777.

Advances and Prospects in Stable Isotope Mixing Model Researches

DENG Tong^{1,2}, LI Zhe², LU Lunhui², WANG Liwei³, HE Wenfeng³, ZHANG Yuanyuan^{2,*}

(1. The College of River and Ocean Engineering, Chongqing Jiaotong University, Chongqing 400074, China;

2. Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China;

3. National Engineering Research Center of Eco-Environment in the Yangtze River Economic Belt, Wuhan 430000, China)

Abstract: Accurately identifying the sources of pollutants is a prerequisite for achieving high-quality ecological and environmental development. Stable isotope mixing models are a key technological means for precisely identifying pollutant sources, but their development has not yet been systematically reviewed. This paper analyzes the research trends and hotspots of these models from 2001 to 2024 and finds that the uncertainty of the results generated by the models themselves has received more attention in recent years. It also provides a detailed elaboration of the development history of stable isotope mixing models, their advantages and disadvantages, and offers a mathematical framework for evaluating the performance of MixSIAR (Bayesian Mixing Models in R) and SIMMR (Stable Isotope Mixing Models in R) models. The evaluation results show that in the application of both models, normally distributed data can reduce result uncertainty, while a decrease in the number of sources will lower the precision. Spectral feature indices should not be used alone as tracers. It is also suggested that in practical applications, prior information should be collected and organized according to the actual operational efficiency needs of the two models. Finally, the evaluation results indicate that MixSIAR is more suitable for analyzing complex ecological and environmental systems that require fine-tuning, while SIMMR is better suited for rapid assessment of pollutant sources in simpler situations.

Keywords: bibliometrics; mathematical principles; evaluation and analysis; stable isotopes; mixing models