

Traceability of River Sudden Water Pollution Events based on IBOA Algorithm

Dongqing Li¹ Dongyan Jia^{1,2} Jinling Song^{1,2*} Dichen Song³ Yi Zhao¹

1. School of Mathematics and Information Technology of Hebei Normal University of Science, Qinhuangdao, Hebei, 066004, China

2. Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center, Qinhuangdao, Hebei, 066004, China

3. School of Business Administration of Hebei Normal University of Science, Qinhuangdao, Hebei, 066004, China

Abstract

Sudden water pollution incidents pose a serious threat to water environment security, and rapid and accurate identification of the pollution source is crucial. This paper establishes a contaminant transport and diffusion model as the foundation. It constructs an objective function by combining forward and reverse concentration probability densities, and employs an Improved Butterfly Optimization Algorithm (IBOA) to inversely identify the pollution source's location, release time, and release mass. Experimental results demonstrate that the IBOA method exhibits high identification accuracy and parameter stability, effectively inverting the characteristics of the pollution source. This provides technical support for real-time early warning and emergency decision-making in the water environment.

Keywords

Sudden Water Pollution; Contaminant Transport and Diffusion Model; Reverse Concentration Probability Density; BOA Algorithm (Butterfly Optimization Algorithm)

基于 IBOA 算法的河流突发水污染事件溯源

李冬晴¹ 贾冬艳^{1,2} 宋金玲^{1,2*} 宋帝辰³ 赵怡¹

1. 河北科技师范学院数学与信息科技学院, 中国·河北 秦皇岛 066004

2. 河北省农业数据智能感知与应用技术创新中心, 中国·河北 秦皇岛 066004

3. 河北科技师范学院工商管理学院, 中国·河北 秦皇岛 066004

摘 要

突发水污染事件严重威胁水环境安全, 快速准确识别污染源是关键。本文以污染物迁移扩散模型为基础, 结合正反向浓度概率密度构建目标函数, 并采用改进蝴蝶优化算法 (IBOA) 反演污染源位置、排放时间和排放质量。实验结果表明, IBOA 方法具有高识别精度和参数稳定性, 可有效反演污染源特征, 为水环境实时预警与应急决策提供技术支撑。

关键词

突发水污染; 污染物迁移扩散模型; 逆向浓度概率密度; BOA 算法

【基金项目】河北省省级科技计划资助 (项目编号: 21370103D); 2023 年度河北省高等学校科学研究项目 (项目编号: ZC2023123); 河北省软件工程重点实验室项目 (项目编号: 22567637H); 河北省软件工程重点实验室开放课题 (项目编号: KF2307); 河北省农业数据智能感知与应用技术创新中心开放课题 (项目编号: ADIC2024Y001, ADIC2024Y003, ADIC2025Y005)。

【作者简介】李冬晴 (2002-), 女, 中国河北沧州人, 硕士, 从事水污染溯源研究。

【通讯作者】宋金玲 (1973-), 女, 中国河北滦州人, 博士, 教授, 硕士研究生导师, 从事数据库安全与数据分析处理研究。

1 引言

河流是重要的水资源, 也是动植物生存的基础。但随着快速城镇化, 工业废水、生活污水和农业废水大量排入河流, 使本已紧缺的水资源更加恶化。突发性水污染事件通常因事故短时大量入河, 污染物随水流迁移扩散, 威胁下游生态。因此, 及时识别污染源的排放时间、位置和质量, 即水污染溯源, 是控制和消除污染的关键。

近年来, 水污染溯源技术逐渐向精细化、智能化发展, 主要分为物理模型法和概率模型法。前者基于水动力学和质量守恒, 常用一维、二维甚至三维对流扩散模型, 如辛小康等^[1]将遗传算法与一维模型结合, Li 等^[2]采用混合同调遗

传算法。该方法对边界与初始条件敏感,适用于参数已知的场景。后者则利用贝叶斯推理、MCMC、高斯过程回归等方法量化不确定性,如 Cheng^[3]通过反向定位概率密度函数识别污染源,Wei^[4]结合贝叶斯与 MCMC 实现反演。该方法无需显式物理建模,适合数据丰富但机理复杂的场景。

由于两类方法各有局限,单独应用难以解决复杂溯源问题。为融合其优势,本文提出一种新的水污染溯源方法:利用污染物迁移扩散模型描述传输过程,结合正反向浓度概率密度构建目标函数,并通过改进蝴蝶优化算法(İBOA)反演污染源参数。该方法兼具物理模型的机理性和概率模型的适应性,不仅提升识别精度,还增强对不确定性的适应能力,显著提高收敛效率与鲁棒性,为污染防控与应急处置提供支持。

2 污染物迁移扩散模型

河流突发水污染时污染物以瞬时方式排放,排放的污染物在河流中会进行扩散、迁移,污染物在扩散过程满足质量守恒定律,因此污染物排入河流后的浓度变化可表示为:

$$\frac{\partial C}{\partial t} + u_x \frac{\partial C}{\partial x} + u_y \frac{\partial C}{\partial y} + u_z \frac{\partial C}{\partial z} = D_x \frac{\partial^2 C}{\partial x^2} + D_y \frac{\partial^2 C}{\partial y^2} + D_z \frac{\partial^2 C}{\partial z^2} - km \quad (1)$$

其中, C (单位为 kg/L) 表示时刻 t 河道内某一点 (x, y, z) 的污染物浓度, u_x, u_y, u_z (单位为 km/h) 分别指的是河流长度方向、宽度方向以及深度方向的平均水流速度, D_x, D_y, D_z (单位为 km^2/h) 分别是相应方向的弥散系数。

在实际情况中,对于流域较窄、流速较稳定的内陆河流,可以忽略河流的宽度方向 Y 和深度方向 Z 的扩散影响,因此可以将这种河流中的污染物扩散问题从复杂的三维扩散模型简化成一维扩散模型,将式 (1) 简化后得到如下公式:

$$\frac{\partial C}{\partial t} + u_x \frac{\partial C}{\partial x} = D_x \frac{\partial^2 C}{\partial x^2} - kC \quad (2)$$

接着对公式 (2) 进行傅里叶变换处理后,可以得到突发水污染事件中时刻 t 、位置 x 的污染物浓度表达式 (即公式 (2) 的解析解) 为:

$$C(x, t) = \frac{M}{\sqrt{4\pi D_x(t-t_0)}} \exp\left(-\frac{(x-x_0-u_x(t-t_0))^2}{4D_x(t-t_0)} - k(t-t_0)\right) \quad (3)$$

其中 $C(x, t)$ 表示污染物在河流位置为 x , 时间为 t 时刻的污染物浓度; M 为污染物排放的质量 (单位为 kg); x_0 和 t_0 分别表示污染源的排放位置和排放时间。

由公式 (3) 可知, 仅在已知污染源排放质量、位置和时间时, 才能计算下游某位置的污染物浓度; 反之, 若只知浓度, 则无法直接反推出上述三项参数。因此需引入概率模型, 对污染源参数反演进行描述, 并结合污染物迁移扩散模型实现突发水污染溯源。

3 目标函数构建

在污染物传播过程中有两种概率密度函数, 分别是正向位置浓度概率密度函数和逆向位置浓度概率密度函数^[5]。正向位置浓度概率密度指的是污染物传播到某一点的概率, 而逆向位置浓度概率密度指的是从监测点来推测污染物来源的概率。根据文献 [5] 的分析来确定它们之间关系。

设 $C(x_s; x_{ob}; t)$ 表示在时间 t 时, 污染物从排放点 x_s 扩散到观测点 x_{ob} 处的浓度, $m_1(x_s; x_{ob}; t)$ 表示是污染物扩散至该点后的正向位置浓度概率密度, M 为污染源的排放质量, 正向位置浓度概率密度 $m_1(x_s; x_{ob}; t)$ 与 $C(x_s; x_{ob}; t)$ 具有下列关系:

$$m_1(x_s; x_{ob}; t) = \frac{C(x_s; x_{ob}; t)}{M} \quad (4)$$

类似地, 通过对污染源的反向溯源分析, 可以得到逆向位置浓度概率密度 $m_2(x_s; x_{ob}; t)$ 与 $C(x_s; x_{ob}; t)$ 具有下列关系:

$$m_2(x_s; x_{ob}; t_s) = \frac{C(x_s; x_{ob}; t)}{M} \quad (5)$$

从公式 (4) 和 (5) 可以看出, 正向位置浓度概率密度 m_1 和逆向位置浓度概率密度 m_2 在计算中存在高度的耦合性, 即有以下关系:

$$m_1(x_s; x_{ob}; t) = m_2(x_s; x_{ob}; t_s) \quad (6)$$

由公式 (6) 可知, 正向与逆向位置浓度概率密度仅计算方向不同, 二者存在耦合关系。即在已知观测值时, 可利用逆向浓度概率密度函数结合迁移扩散模型反求污染物释放过程, 并得到预测浓度。为使预测浓度序列 Y 与实际观测序列 C 尽可能接近, 采用两者的均方根误差作为目标函数 (式 7)。确定目标函数后, 即可通过优化算法求解污染源参数最优解。

$$\min f(C, Y) = \min \sqrt{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (C_{ij} - Y_{ij})^2} \quad (7)$$

4 改进的蝴蝶优化算法 (İBOA)

蝴蝶优化算法 (BOA) 是一种全局优化算法, 通过模拟蝴蝶嗅觉搜索花蜜或配偶位置, 迭代更新以寻找最优解。算法包含全局和局部搜索两个阶段, 由固定切换概率控制转换。尽管性能优于部分其他算法, 但存在种群多样性低、寻优精度不足和收敛慢等问题^[6], 为了修正上述缺陷, 本文使用改进的 BOA 优化算法进行河流污染溯源。

4.1 反向学习策略

原始的蝴蝶优化算法是在搜索空间中随机初始化种群的位置, 但这种随机方式生成的初始种群, 可能会导致蝴蝶个体的位置分布不均匀。为了提高蝴蝶种群的初始质量、均匀蝴蝶分布, 本实验使用反向学习策略对随机初始化的蝴蝶位置进行优化, 反向学习策略如式 (8) 所示。

$$\hat{x}_{ij} = \lambda(a_j + b_j) - x_{ij} + \epsilon_{ij} \quad (8)$$

其中, x_{ij} 是第 i 只蝴蝶在第 j 维的值; $\hat{x}_{ij}$ 则是 x_{ij} 的反向解; $\lambda \sim U(0.5, 1)$ 为均匀分布的缩放系数; $\epsilon_{ij} = 0.1(a_j - b_j)\xi$, $\xi \sim U(-0.5, 0.5)$ 为基于解空间大小的随机扰动项; a_j 和 b_j 分别为解空间在第 j 维的上界和下界。

为了保证反向解 $\hat{x}_{ij}$ 的可行性, 使用公式 (9) 将其约束在解空间边界内, 其中 a_j 和 b_j 分别为第 j 维的上界和下界。

$$\hat{x}_{ij} \leftarrow \min(\max(\hat{x}_{ij}, a_j), b_j) \quad (9)$$

4.2 柯西变异与概率扰动混合的全局搜索策略

为了增强蝴蝶优化算法的全局搜索性能, 从而能快速接近最优解, 本实验采用柯西变异与概率扰动混合的全局搜索策略。使用公式 (10) 和 (11) 对全局最优解进行柯西变异产生全局候选解并限制在解空间内。通过上述公式产生的当前全局候选解需要和目前的全局最优解进一步比较适应度, 如果全局候选解的适应度值比当前的全局最优解好, 就利用其替换当前全局最优解。

$$x_{\text{candidate}} = x_{\text{best}} + x_{\text{best}} * \text{Cauchy}(0, I_d) \quad (10)$$

$$x_{\text{candidate}} \leftarrow \max(\min(x_{\text{candidate}}, Ub), Lb) \quad (11)$$

其中, $\text{Cauchy}(0, I_d)$ 指的是 d 维标准柯西分布; x_{best} 为当前的全局最优解, $x_{\text{candidate}}$ 为柯西变异后的全局候选解; Ub 和 Lb 分别是解空间的上下界。

然后根据公式 (12) 对每只蝴蝶的位置进行更新, 如果随机数 $U(0, 1)$ 小于 0.3, 蝴蝶更新位置向全局最优解方向移动, 否则使蝴蝶位置在最优解附近小范围随机探索。

$$x_i^{\text{new}} = \begin{cases} x_{\text{best}} + 0.1(Ub - Lb)\xi, & U(0, 1) < 0.3 \\ x_i + U(0, 1) * (x_{\text{best}} - x_i) * FP, & U(0, 1) \geq 0.3 \end{cases} \quad (12)$$

$$FP = s * F^e * (0.9 + 0.2U) \quad (13)$$

其中, $\xi \sim U(-0.5, 0.5)^d$ 表示一个 d 维随机向量, 每个分量 ξ 独立服从区间 $[-0.5, 0.5]$ 上的均匀分布; $U(0, 1)$ 是指均匀随机数; FP 表示参数刺激强度; s 表示参数感官模态; F 为当前解的适应度值; e 表示为参数香味指数。

4.3 基于随机惯性权重的局部搜索策略

在原始的蝴蝶优化算法中, 进入局部搜索阶段后是采用随机的方式更新蝴蝶位置, 这样容易使更新后位置受到前一位置的影响, 导致后期陷入局部最优。为了减小前一位置对更新位置的影响, 在局部搜索阶段中使用随机惯性权重对蝴蝶的位置进行更新, 具体公式如式 (13) 所示。

$$x_i^{t+1} = w^t x_i^t + \epsilon^t (x_i^t - x_k^t) * FP \quad (14)$$

其中, x_i^{t+1} 表示第 i 只蝴蝶的更新后位置; x_i^t 表示第 i 只蝴蝶的前一个位置; $\epsilon^t \sim U(0, 1)$ 随机均匀分布; x_j^t , x_k^t 为随机选择的两个不同个体位置, FP 指的是参数刺激强度; w^t 为随机惯性权重, 计算方式如式 (15):

$$\omega = \mu_{\min} + (\mu_{\max} - \mu_{\min}) * \text{rand}() + \sin\left(\frac{\pi t}{2T_{\max}} + \pi\right) \quad (15)$$

其中, $\mu_{\max}$, $\mu_{\min}$ 分别是惯性权重的上限和下限; $\text{rand}()$ 是在 $[0, 1]$ 范围内随机取值。

4.4 IBOA 算法

IBOA 优化算法的主要流程包括: 首先设置相关参数并初始化蝴蝶种群, 然后计算每个蝴蝶个体的适应度值并找出全局最优位置, 接着根据种群中蝴蝶散发的香气, 通过全局搜索和局部搜索不断迭代每个蝴蝶个体的位置, 当迭代次数达到最大值时则直接输出全局最优解。IBOA 优化算法的流程图如图 1 所示, 算法的具体描述如下。

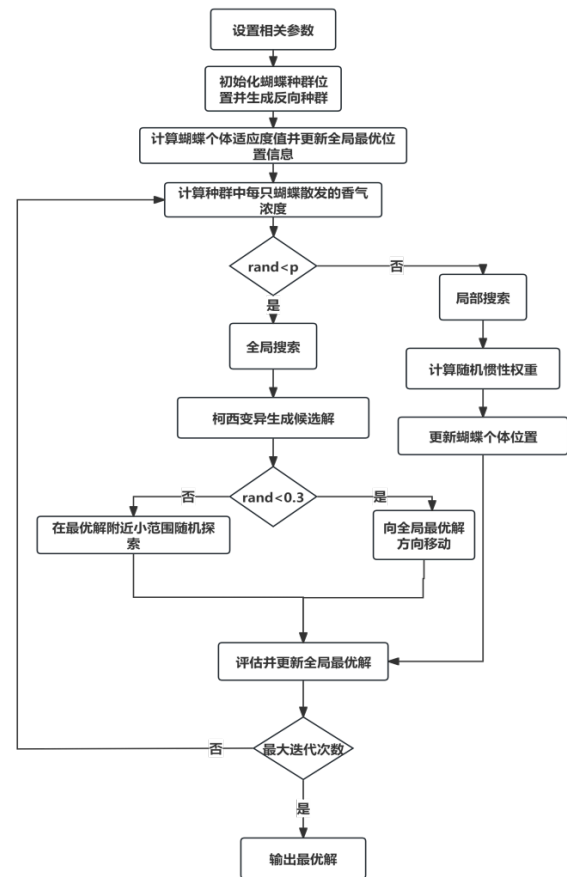


图 1 改进蝴蝶优化算法 (IBOA) 的流程图

步骤 1: 设置蝴蝶种群规模 S , 搜索空间维度 dim , 迭代次数 Max , 香味指数 e , 感官模态 b , 柯西变异尺度 c , 惯性权重的上限 $\mu_{\max}$, 惯性权重下限 $\mu_{\min}$, 切换概率 p 等参数。

步骤 2: 在搜索空间范围内随机生成每只蝴蝶个体的位置 (即蝴蝶的初始化种群), 初始化种群根据公式 (8)、(9) 进行反向学习, 获得蝴蝶精英种群。

步骤 3: 依据公式 (7) 的目标函数对蝴蝶个体进行评价, 选出当前最优解, 并用最优解更新全局最优位置。

步骤 4: 根据蝴蝶种群的适应度值, 利用香味指数 e 计算每只蝴蝶的香气浓度。

步骤 5：在 $(0,1)$ 范围内随机取值和 p 进行对比，如果小于 p 的话（代表蝴蝶感知到更高香味个体），那么进入步骤 6 的全局搜索阶段，否则（代表无更高香味个体时），进入步骤 7 的局部搜索阶段。

步骤 6：假如进入的是全局搜索阶段的话，根据公式（10）、（11）利用柯西变异生成候选解优化全局最优解，再根据公式（12）以概率扰动混合的方式对蝴蝶位置进行更新。

步骤 7：假如进入的是局部搜索阶段的话，根据公式（15）计算随机惯性权重，将得到的随机惯性权重带入公式（14）更新蝴蝶位置。

步骤 8：根据蝴蝶的当前位置计算适应度值并与当前全局最优解进行比较，当出现适应度小于全局最优解的情况，那么用当前的蝴蝶位置更新全局最优解。

步骤 9：判断是否达到最大迭代次数，达到则直接输出全局最优解，否则返回到步骤 4 继续迭代。

5 基于 IBOA 算法的水污染溯源实验

为了验证 IBOA 对河流突发水污染溯源问题上的可行性，利用 IBOA 对待求的污染源参数（污染源质量 M ，污染源位置 x_0 ，污染源排放时间 t_0 ）进行反演，并将反演结果与真实的污染源参数进行对比检验。

5.1 实验数据

在较窄河流的某处设置瞬时点源污染，污染源排放质量 M 为 17.88kg，污染源排放时间 T 为 0.5h，在下游 2 千米和 2.5 千米处分别设有监测断面，每个断面以 6 分钟为间隔进行 10 次采样，共获取 20 组污染物浓度的监测数据。该河流扩散系数 E 为 $2\text{km}^2/\text{h}$ 和河流平均流速 U 为 $5\text{km}/\text{h}$ ，根据上述河流情况本实验选用污染物迁移扩散模型模拟污染物扩散情况。需要反演的污染源参数如表 1 所示：

表 1 待反演参数

参数	污染物排放质量 $M(\text{kg})$	污染物排放时间 $t_0(\text{h})$	污染物排放位置 $x_0(\text{km})$
真实值	17.88	0.5	2

5.2 污染源参数反演实验与分析

针对以上实验数据，本节利用 IBOA 算法对污染源参数（污染源质量 M 、污染源排放时间 t_0 以及污染源排放位置 x_0 ）进行反演，并与原始蝴蝶优化算法 BOA 的反演结果进行对比。IBOA 算法的参数设置如下：蝴蝶种群 S 为 50，搜索空间维度 dim 为 3，迭代次数 Max 为 150，香味指数为 e 为 0.08，感官模态 b 为 0.02，柯西变异尺度 c 为 0.15，惯性权重的上限 $\mu_{\max}$ 为 0.9，下限 $\mu_{\min}$ 为 0.4，初始切换概率 p 为 0.85。IBOA 与 BOA 算法都独立运行 50 次，每次运行的迭代次数为 50 次，根据运行中的反演参数变化来判断哪个算法更适合水污染溯源。

图 1 展示了 BOA 算法在 50 次独立运行中的 3 个参

数反演结果。子图 1 表明质量参数反演值波动剧烈，最小 10、最大约 28，与真实值偏差大，显示其全局寻优不稳定，易受初始种群和搜索路径影响。子图 2 显示排放时间反演值多低于真实值 0.5h，部分明显偏高，反映出收敛速度不稳定。子图 3 中位置参数多数低于真实值 2km，也存在偏大值，说明搜索精度不足。

综上，BOA 算法在质量控制、运行效率和收敛精度上均表现出波动性，为后续改进提供了方向。

图 2 展示了 IBOA 算法在 50 次运行中的反演结果。相比 BOA，IBOA 在稳定性和整体性上显著提升。子图 1 显示质量参数紧密分布在真实值 17.88kg 附近，波动极小，明显减少了性能抖动，体现出更强的稳定性与一致性。子图 2 表明排放时间始终接近真实值 0.5h，收敛速度稳定，时间控制良好。子图 3 中位置参数均集中在 2km 附近，几乎无偏移，验证了 IBOA 在搜索精度上的优越性。

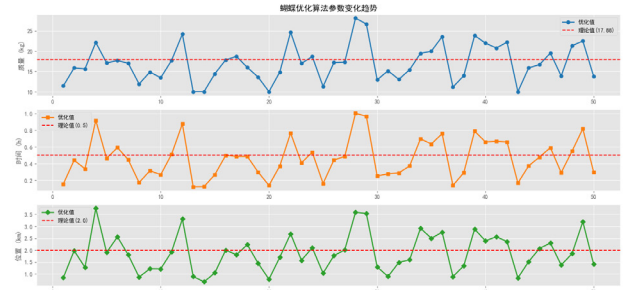


图 2 蝴蝶优化算法反演参数的变化趋势

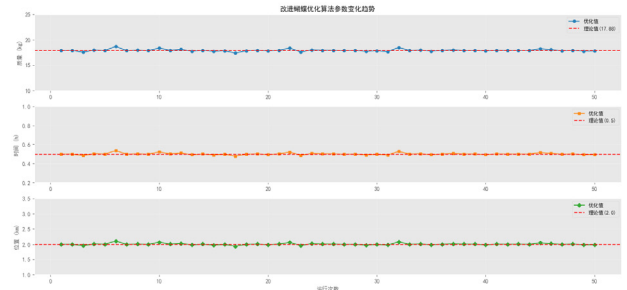


图 3 改进蝴蝶优化算法反演参数的变化趋势

为进一步比较 IBOA 与 BOA 在 50 次运行中的性能，对三个反演参数进行了统计分析，结果如表 2、表 3 所示。IBOA 的最优值与真实值几乎一致，三项参数的绝对误差均在 0.003 以内、相对误差不超过 0.04%，定位精度极高；而 BOA 虽接近真实值，但仍存在偏差，特别是质量参数相对误差达 0.1607%。在最差值方面，BOA 的相对误差高达 44.0716%、74.9314%、55.2836%，表现出明显不稳定；而 IBOA 的最差值误差均低于 8%，仍能保证溯源准确性。在平均值方面，IBOA 结果与真实值高度一致，质量和时间几乎重合，位置参数虽略偏低（2.0310%），但仍优于 BOA。标准差对比进一步显示，BOA 波动较大（4.5285、0.2361、

0.8009), 而 IBOA 数值显著更低 (0.2140、0.0098、0.0247), 说明其稳定性和鲁棒性远优于 BOA。

表 2 原始蝴蝶优化算法的反演结果与真实值对比

	质量 (kg)	时间 (h)	位置 (km)
真实值	17.88	0.5	2
最优值	17.8513	0.4993	2.0014
绝对误差	0.0287	0.0007	0.0014
相对误差	0.1607%	0.1351%	0.0688%
最差值	10	0.125343173	0.894327708
绝对误差	7.88	0.374656827	1.10567229
相对误差	44.0716%	74.9314%	55.2836%
平均值	17.0494	0.4544	1.8779
绝对误差	0.8307	0.0456	0.1221
相对误差	4.6452%	9.1235%	6.1044%
标准差	4.5285	0.2361	0.8009

表 3 改进蝴蝶优化算法反演结果与真实值对比

	质量 (kg)	时间 (h)	位置 (km)
真实值	17.88	0.5	2
最优值	17.8779	0.4998	1.9994
绝对误差	0.0021	0.0002	0.0006
相对误差	0.0115%	0.0352%	0.0310%
最差值	18.6651	0.5355	2.1017
绝对误差	0.7851	0.0355	0.1017
相对误差	4.3911%	7.0912%	5.0858%
平均值	17.8924	0.5005	1.9594
绝对误差	0.0124	0.0005	0.0406
相对误差	0.0692%	0.1060%	2.0310%
标准差	0.2140	0.0098	0.0247

通过可视化与误差对比可见, IBOA 在整体性能上明显优于 BOA。BOA 虽偶尔接近真实值, 但整体波动大、稳定性不足; 而 IBOA 在参数曲线上收敛平稳, 最优值与平均值更贴近真实值, 最差值偏差减小, 标准差显著降低, 表现出更强的稳定性和鲁棒性。综上, IBOA 在精度与收敛速度上均优于 BOA, 适用于水污染溯源问题。

6 结论

针对突发性水污染事件的污染源识别, 本文提出一种融合物理与概率模型的反演方法。以污染物迁移扩散模型为物理基础, 利用正向与逆向浓度概率密度关系构建目标函数, 并采用改进蝴蝶优化算法 (IBOA) 反演污染源质量、排放时间和位置。为验证 IBOA 的适用性, 将其与原始算法分别运行 50 次, 对比反演参数趋势及最优值、最差值、平均值的误差和标准差, 结果表明 IBOA 在精度和稳定性上均优于原始算法, 具有可靠性。

然而, 该方法仍有限制, 一维迁移扩散模型仅适用于河道较窄、流速稳定的场景, 难以刻画复杂断面及非均匀流速下的扩散过程, 影响岸边污染及复杂地形中的精度。但所提框架独立于模型维度, 具备推广至二维 / 三维模型的潜力。未来可结合高维模型提升适应性, 并进一步优化算法性能, 拓展其在多源污染识别、实时预警与应急决策中的应用, 为水环境治理提供更高效智能的支撑。

参考文献

[1] 辛小康,韩小波,李建,等.基于遗传算法的水污染事故污染源识别模型[J].水电能源科学,2014,32(07):52-55+136.

[2] Li J , Lu W , Fan Y .Groundwater Pollution Sources Identification Based on Hybrid Homotopy-Genetic Algorithm and Simulation Optimization[J].Environmental Engineering Science, 2021.

[3] Cheng, W. P., & Jia, Y. (2010). Identification of contaminant point source in surface waters based on backward location probability density function method. Advances in Water Resources, 33(4), 397-410.

[4] WEI G, CHI Z, YU L, et al. Source identification of sudden contamination based on the parameter uncertainty analysis[J].Journal of Hydroinformatics, 2016, 18 (6): 919-927.

[5] 王家彪,雷晓辉,廖卫红,等.基于耦合概率密度方法的河渠突发水污染溯源[J].水利学报,2015,46(11):1280-1289.

[6] Arora S ,Singh S .Butterfly optimization algorithm: a novel approach for global optimization[J].Soft Computing,2019,23(3):715-734.