

基于改进 FA 算法的河流突发水污染事件溯源^①



赵栋梁^{1,2}, 周晓磊², 窦志强², 武 曠³

¹(中国科学院大学, 北京 100049)

²(中国科学院 沈阳计算技术研究所, 沈阳 110168)

³(辽宁省生态环境监测中心, 沈阳 110161)

通信作者: 周晓磊, E-mail: zhouxl@sict.ac.cn

摘 要: 针对河流突发水污染事件的溯源问题, 利用正向和逆向质量概率密度之间的耦合关系, 实现污染源的位置、排放时间和污染源排放强度之间解耦, 并结合一维水体扩散模型和美国特拉基河示踪剂实验监测数据, 建立了基于改进萤火虫算法的河流突发水污染事件溯源求解方法. 在求解过程中将监测数据划分为训练集和实验集, 通过训练集数据利用改进的萤火虫算法对河流的水文参数进行调整, 在实验集采用调整后的水文参数, 并通过不同监测断面数据独立求解, 通过方差分析结果, 排除了溯源误差较大的监测断面数据. 研究结果表明, 溯源结果精度较高, 并具有对监测数据的纠错能力, 对实际河流突发水污染事件具有一定的指导意义.

关键词: 河流; 突发水污染; 萤火虫算法 (FA); 耦合概率密度; 溯源

引用格式: 赵栋梁, 周晓磊, 窦志强, 武曠. 基于改进 FA 算法的河流突发水污染事件溯源. 计算机系统应用, 2022, 31(10): 191-198. <http://www.c-s-a.org.cn/1003-3254/8767.html>

Source Tracing of Sudden River Pollution Based on Improved Firefly Algorithm

ZHAO Dong-Liang^{1,2}, ZHOU Xiao-Lei², DOU Zhi-Qiang², WU Jian³

¹(University of Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

³(Liaoning Ecological Environment Monitoring Center, Shenyang 110161, China)

Abstract: This study is conducted to trace the source of sudden river pollution. Specifically, the coupling relationship between forward and reverse mass probability density is used to realize the decoupling of the location, discharge time, and discharge intensity of pollution sources; then, given the one-dimensional water body diffusion model and the monitoring data from the tracer experiment in Truckee River of the United States, a method for tracing the source of sudden river pollution is established on the basis of the improved firefly algorithm (FA). In the method, the monitoring data are divided into a training set and an experimental set, and by the training set data, the improved FA is employed to adjust the hydrological parameters of the river. Then, the adjusted hydrological parameters are used in the experimental set, and data from different monitoring sections are used independently for solutions. Finally, the results are analyzed by variance to exclude the data with large source-tracing errors. The results reveal that the source-tracing results have high accuracy and the ability to correct the monitoring data, which is of certain guiding significance for the actual sudden river pollution.

Key words: rivers; sudden water pollution; firefly algorithm (FA); coupling probability density; source tracing

河流突发水污染事件会对河流下游区域的生态系统造成破坏, 并且随污染排放时间增加, 污染扩散面积

逐渐扩大, 对人民群众以及生态环境的影响逐渐加大, 因此河流突发水污染事件的第一要素是确定污染源位

① 基金项目: 辽宁省中央引导地方科技发展专项 (2021010211-JH6/105); 辽宁省科学技术计划 (2020JH2/10300113); 沈阳市科技计划 (20-203-5-50)

收稿时间: 2021-12-17; 修改时间: 2022-01-29, 2022-02-18; 采用时间: 2022-03-21; csa 在线出版时间: 2022-07-07

置^[1]。本文主要通过研究污染物在河流中的扩散规律,并通过监测站所观测到的污染物浓度变化对污染物排放的位置、时间、质量进行溯源^[2],从而还原污染物排放的变化过程,研究的难点在于污染源溯源求解过程中的求解不唯一、污染源溯源相关的水文参数测量阶段存在误差,污染物回溯模型参数设置存在误差,并且污染物的传播过程不能通过实验来重复^[3]。

水污染溯源的方法最初来自于空气污染的溯源,最开始被应用于地下水污染的研究^[4],Alapati等^[5]通过最小二乘法结合一维地下水扩散模型研究了地下水污染源参数,国内的白玉堃等^[6]通过卡尔曼滤波方法溯源地下水的污染源个数和位置,张双圣等^[7]采用贝叶斯公式结合二维水质的对流模型研究了地下水的污染源瞬时排放模式。在河流污染中也采用了和学者研究地下水污染中类似的方法^[8],国内外学者在对河流污染的污染源排放相关参数的求解做了大量的工作,大致可以把研究分为3大类,分别为数学分析法、模拟优化法、概率分析法,其中,前2个称为确定性方法,最后1个称为不确定性方法。

在数学分析法的研究上,辛小康等^[9]通过遗传算法和数学分析方法相结合,通过一维水质模型研究了单点源和多点源的瞬时污染源参数识别问题;饶清华等^[10]通过有限元法研究了闽江下游不同地点的污染物浓度情况。在模拟优化法上,刘洁等^[11]通过遗传算法在一维对流扩散方程求解目标函数;吴一亚等^[12]通过微分进化算法求解宽浅河道瞬时污染排放迁移问题。确定性方法通过污染物浓度数据反推污染源参数,但是当数据出现一点偏差时通过确定性方法求解结果偏差较大。

在概率分析法的研究上,陈海洋等^[13]通过贝叶斯推理和二维水体扩散规律构建模型,并通过马尔科夫链蒙特卡罗法求解污染源相关参数;程伟平等^[2]通过逆向概率密度法求解了一维水质的污染源参数识别问题;王家彪等^[14]通过结合正向位置概率密度和逆向位置概率密度之间的耦合关系建立求解模型,并通过微分进化算法求解模型,能够实现求解参数之间的解耦,降低时间复杂度,并且结合了确定性和不确定性分析方法,一定程度上降低了参数误差造成的影响。概率分析方法在一定程度上能够避免参数误差造成的误差较大,但是受观测误差影响比较大,具有一定的随机性。

上述研究中,污染物浓度监测数据使用所有监测站点不同时刻的数据,不同监测站间数据没有做区分,

但实际河流中,污染物浓度也存在水流和污染物混合不均匀导致监测浓度不准现象,监测站的数据可以分为不准确、相对准确两种数据类型。基于大部分监测站点的数据都是相对准确的事实,本文提出一种基于多萤火虫种群的求解方法,通过在算法初期隔离不同监测站间的数据来避免不准确数据使相对准确数据失真,在算法后期通过不同种群间相互作用通过多数效应来淘汰不准确数据类型的种群,从而避免某个监测站数据不准确对求解造成的系统误差。

本文采用耦合概率密度方法对所需求解的参数进行建模^[14],并考虑到观测误差和一维水质模型需要污染物混合均匀条件,代入实时的水文特征,采用改进的萤火虫算法(firefly algorithm, FA)算法,引入多种群机制将多个污染物监测站数据同时代入求解,改进后的算法具有更强的全局搜索能力和跳出局部极值能力,对建模后的目标函数进行求解,寻找污染源排放相关的排放位置、排放时间、排放强度。

1 水动力模型构建

河流水污染问题主要是通过已知的水文参数、监测站监测数据信息对污染物的溯源需要根据具体河道的水文特征进行水动力学模型构建,利用污染物在水流中的正向质量概率密度和逆向质量概率密度之间的耦合概率密度关系,使污染物排放强度与排放位置和排放时间之间解耦。

1.1 污染源扩散关系

对于污染物的排放问题,污染物进入河流后的扩散规律满足质量守恒定律^[15],可表示为:

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

其中, m 为河道内点 (x, y, z) 在 t 时刻污染物的浓度,单位 mg/L ; t 为污染物排放后的时间,单位 s ; D_x 、 D_y 、 D_z 分别为河流长度方向、宽度方向、深度方向的弥散系数,单位 m^2/s ; u_x 、 u_y 、 u_z 分别为河流长度方向、宽度方向、深度方向的水流的平均流速,单位 m/s ; k 为污染物的降解系数,单位 s^{-1} 。

对于内陆河流,污染物扩散主要受河流两岸和河流底部的约束,因此污染物在水流中会产生边界反射,水中对流、扩散、吸附等现象。考虑到河流宽度和河

流深度远小于河流长度,在非洪汛期,水流速度基本不变,污染物在被投放到河流内后,能在投放点断面附近迅速扩散,扩散后断面附近污染物浓度基本相同.因此可以把污染源在内河扩散问题简化为一维水体扩散模型,排入河流后的污染物浓度随时间和位置的变化情况可忽略 y, z 方向上的变化^[16],将式(1)简化为:

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

若污染物为瞬时排放模式,则可对式(2)进行傅里叶变换^[17],可以得到 t 时刻在 x 断面处的污染物浓度关系式:

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

其中, $m(x, t)$ 表示污染物在河道 x 位置时间为 t 时刻断面的污染物浓度,单位 g/L ; M 为污染物排放处的强度,单位 g/m^2 ; t_0 表示污染物的排放时间; x_0 表示污染物的排放位置.

由式(3)可知,对于河流突发水污染事件溯源问题,需要确定污染物排放处强度 M ,污染物排放位置 x_0 ,污染物排放时间 t_0 .溯源问题相对于模拟过程具有不能直接求解的特点,但是可以通过监测站的断面信息确定污染源3个参数的约束范围,然后通过迭代试算寻找最优的参数组合.但是3个未知数的组合复杂度较高,寻优过程具有运算量大,求解偏差大,求解时间长的特点,因此需要引入降维方法,降低模型的维度.

本文利用正向位置浓度概率密度(从污染源排放断面判断污染源出现在下游的浓度分布情况)和逆向位置浓度概率密度(从监测断面判断污染源在不同断面处的可能性大小)的关系^[14],将求解污染源的浓度分布问题转化为求解污染源的逆向位置浓度概率问题,从而将 x_0, t_0 的求解和 M 分开.

1.2 逆向位置关系

设 $m(x_s, x_{ob}, t)$ 表示的监测浓度对应的正向位置浓度概率密度为 $m_1(x_s, x_{ob}, t)$,逆向位置浓度概率密度为 $m_2(x_s, x_{ob}, t_s)$,根据文献[2]可知,它们之间存在以下关系:

$$m_1(x_s, x_{ob}, t) = m(x_s, x_{ob}, t)/M \quad (4)$$

$$m_2(x_s, x_{ob}, t_s) = m(x_s, x_{ob}, t)/M \quad (5)$$

其中, x_s 为污染源位置, x_{ob} 为观测断面位置, t 为污染物正向移动的时间点, t_s 为通过观测断面溯源计算的污染源排放时间点; $m(x_s, x_{ob}, t)$ 表示时间为 t 时,污染源从排

放断面 x_s 扩散到 x_{ob} 观测断面的浓度; $m_1(x_s, x_{ob}, t)$ 为污染源从排放断面 x_s 扩散到 x_{ob} 观测断面 t 时刻观测断面的正向位置浓度概率密度,量纲为 m^{-1} ; $m_2(x_s, x_{ob}, t_s)$ 为从观测断面 x_{ob} 推测出污染源在 t_s 时刻处在 x_s 处的逆向位置浓度概率密度,量纲为 m^{-1} .

由式(4)、式(5)可知,正向位置浓度概率密度和逆向位置浓度概率密度两者在计算中具有高度耦合性,在同一位置,计算方向不同,计算结果概率密度相同^[18].因此可以得出:

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

由于 $m_2(x_s, x_{ob}, t_s)$ 具有 m^{-1} 的量纲,因此也满足式(2),对其也进行傅里叶变换得到逆向位置浓度概率密度为^[19]:

$$m_2(x_s, x_{ob}, t_s) = \frac{1}{\sqrt{4\pi D_x (t_{ob} - t_s)}} \exp\left(-\frac{(x_{ob} - x_s - u(t_{ob} - t_s))^2}{4D_x (t_{ob} - t_s)} - k(t_{ob} - t_s)\right) \quad (7)$$

由式(7)可知,污染源的逆向位置浓度概率与污染物排放强度无关,可以通过求解 $m_2(x_s, x_{ob}, t_s)$ 来计算排放位置 x_0 ,排放时间 t_0 .

2 溯源优化模型构建

2.1 位置时间模型

由式(4)可知, $m_2(x_s, x_{ob}, t_s)$ 和 $m(x_s, x_{ob}, t)$ 之间具有线性相关性,相关系数为污染物排放浓度的倒数.设 m^i ($i=1, 2, 3, \dots, n$)为一系列观测断面浓度数据, m_2^i ($i=1, 2, 3, \dots, n$)为一系列与 m^i 相对于的通过观测断面 x_{ob} 确定的逆向位置浓度概率密度^[20].两者相关系数为 R ($R \leq 1$), R 值越接近1代表此时逆向位置浓度概率密度越接近真实值,可知其有以下相关性:

$$R = \frac{\sum_{i=1}^n (m^i - \bar{m})(m_2^i - \bar{m}_2)}{\sqrt{\sum_{i=1}^n (m^i - \bar{m})^2} \sqrt{\sum_{i=1}^n (m_2^i - \bar{m}_2)^2}} \quad (8)$$

其中, n 为观测断面浓度数据和对应的逆向位置浓度概率密度数据的个数, \bar{m} 表示观测断面浓度数据的平均值, \bar{m}_2 表示逆向位置浓度概率密度数据的平均值.

设 x_0, t_0 为真实的污染源排放位置和排放时间,优化求解的目标是寻找最接近真实值的组合.因此将溯源问题转化求解最优的 x', t' 组合使当前的相关系数 R 最接近1,可通过求解以下目标函数完成求解^[20].

$$\min(1 - R) \quad (9)$$

目标函数的约束条件是通过已有的监测站断面数确定的排放位置的范围和排放时间的范围:

$$x_{\text{ob}} < x' < x_{\text{max}}; t_{\text{ob}} < t' < t_{\text{max}} \quad (10)$$

通过求解式(9),寻找其最小值即可得到最近接近 x_0, t_0 的 x', t' 组合.

2.2 污染物排放强度模型

通过位置时间模型求得的 x', t' 组合,可以作为污染源排放量模型的已知条件,因此污染源排放量模型只需要求解污染物排放处的强度 M 这一个未知参数,根据文献[20]可知,污染物的正向位置浓度概率密度和逆向位置浓度概率密度具有线性相关性,可得式(11):

$$M = \frac{1}{n} \sum_{i=1}^n \frac{m^i}{m_2^i} \quad (11)$$

可以通过式(11)估算污染物排放处的强度 M 的大小范围作为约束范围.

为了进一步确定 M 的大小,可以通过正向位置浓度概率密度构建目标函数:

$$\min \left[\sum_{i=1}^n (Mm_2^i - m^i)^2 \right] \quad (12)$$

目标函数中 M 的约束范围由式(11)确定的范围确定:

$$M_{\min} \leq M \leq M_{\max} \quad (13)$$

通过求解式(12),寻找其最小值即可得到最接近真实值的污染物排放处的强度 M .

3 溯源优化模型求解

在第2节将污染物溯源需要求解的污染物排放位置 x_0 ,污染物排放时间 t_0 ,污染物排放强度 M 三个参数转化为对式(9)、式(12)两个目标函数求最小值问题,本文选用改进的萤火虫算法^[21]求解目标函数,为了使算法更适合河流污染溯源的求解,做了以下改进.

3.1 改进参数控制策略和约束条件

在原始的萤火虫算法中,控制随机扰动的步长 α 是一个固定值,主要目的是作为随机扰动项增加算法的搜索能力和一定程度上保持种群多样性. α 越大,全局搜索能力越强,但是算法后期可能存在跳过最优解,发生在最优解附近震荡的现象; α 越小,局部搜索能力越强.

对应到本问题,在求解目标函数式(7)时,污染物排放位置 x 的范围相对于污染物排放时间范围较大,因此需要提供一个缩放系数 k ,来调整不同维度的步长

相对值,可表示为:

$$\alpha_0 = k_i \alpha_0 \quad (14)$$

其中, $k_i (i=1, \dots, d)$,分别表示在1到 d 维的缩放因子,大小由不同维度之间的搜索范围比例确定, α_0 为控制随机扰动的步长初始值,大小为 $[0, 1]$.

由于污染物溯源问题在初期搜索范围较大需要更大的步长来加快搜索速度和全局搜索能力,在后期需要较小的步长来提升搜索精度,因此针对本问题参考文献[21]提出一种自适应步长改进策略,可表示为:

$$\alpha_{t+1} = \left(1 - \frac{t}{T_{\max}} \right) \cdot \alpha_t \quad (15)$$

其中, $t (t=1, \dots, n)$ 表示当前的迭代次数, α_t 表示第 t 次迭代时随机扰动的步长大小, T_{\max} 表示迭代次数的最大值.

原始的萤火虫算法在移动位置时没有考虑在该维度的搜索范围,对应到污染物溯源问题上, x_0, t_0, M 三个参数都可以通过监测站的参数和河道信息确定其大致范围,因此需要对萤火虫的移动范围做出限制^[16],可表示为:

$$x_i = \begin{cases} x_{\min}, & x_i < x_{\min} \\ x_{\max}, & x_i > x_{\max} \end{cases} \quad (16)$$

其中, x_{\min} 表示在该维度搜索范围的最小值, x_{\max} 表示在该维度搜索范围的最大值, x_i 表示第 i 只萤火虫的位置.

3.2 改进种群策略

原始的萤火虫算法中,萤火虫亮度较高的个体只会对其附近的亮度相对较低的个体有吸引力,但是对距离较远的萤火虫个体没有吸引力,因此原始算法容易陷入局部极值过早收敛导致求解误差较大.针对上述问题本文参考文献[22]提出的多萤火虫种群的优化策略改进算法,具体内容见第3.3节算法流程中步骤3-11.

3.3 算法流程

对于污染物溯源问题,早期的学者采用耦合概率密度分析方法建模求解时,没有区分不同监测站的污染物浓度数据,某个监测站的监测值可能因为污染物和河流混合不均匀或者某个设备本身存在系统误差,从而导致通过该监测站数据求解误差较大.

针对本问题提出如下改进:首先监测站间数据相互独立,使用单个监测站数据依次使用多种群萤火虫算法求解,最后分析各个监测站的求解结果,采用标准差分析法,若通过某一个监测站求解的结果相对于其他使用其他监测站数据求解结果偏差较大,证明该监

监测站数据为不准确数据, 淘汰该结果^[23]. 整体算法结构如图1所示, 具体算法流程如下.

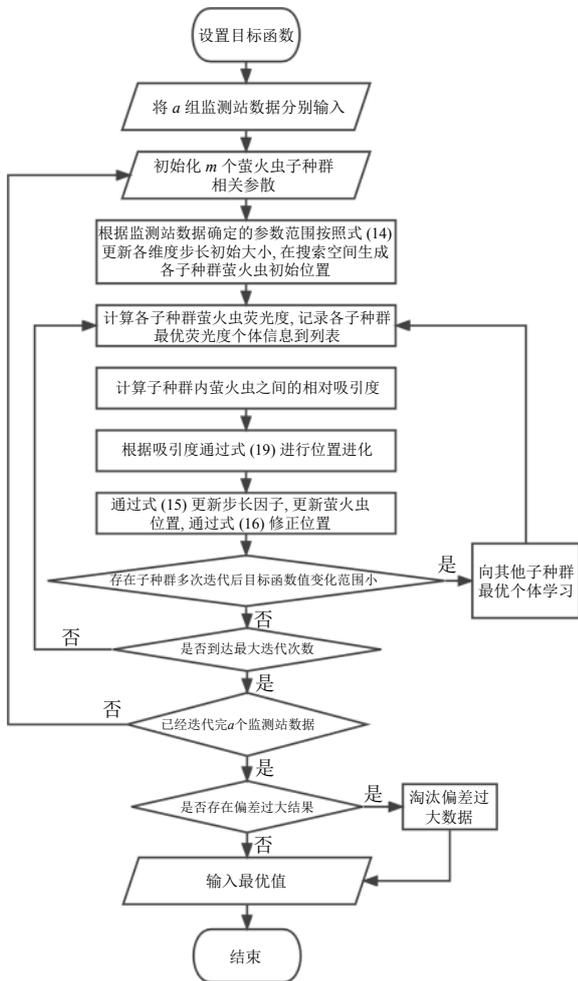


图1 改进的FA算法流程图

步骤1. 确定溯源优化模型求解需要求解的内容, 式(9)、式(12).

步骤2. 设总共有 a 个监测站的监测数据, 每个监测站的污染物浓度系列为 m^i ($i=1, 2, \dots, n$), 分别将这 a 个监测站的数据作为初始条件各执行一次步骤3-12的算法内容.

步骤3. 假设所有子种群的萤火虫数目之和为 m , 子种群数目为 n , 分别为子种群初始化不同的光强吸收系数 γ 、最大吸引度因子 β_0 、步长因子 α , 使各个子种群具有不同的迭代过程.

步骤4. 根据监测站数据和河道信息通过式(16)确定需要求解的参数 x_0 , t_0 , M 的上下限.

步骤5. 根据参数 x_0 , t_0 , M 的上下限间的大小比例通过式(14)更新各子种群不同维度的步长因子 α 的初始值.

步骤6. 根据步骤4结果设定不同维度的搜索空间范围, 根据该范围在不同维度随机生成萤火虫的初始位置:

$$X_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijd}) \quad (17)$$

其中, $j=1, 2, \dots, n$; $i=1, 2, \dots, m/n$, X_{ij} 表示在子种群 j 中第 i 只萤火虫的位置, d 表示参数的维度.

步骤7. 将生成的 X_{ij} 代入式(9), 将式(9)计算结果设置为子种群 j 中第 i 只萤火虫的荧光度 I_{0ij} , 分别记录各个子种群中萤火虫个体荧光度的最大值 I_{0j}^{best} , 将各个子种群 I_{0j}^{best} 和位置 X_{ij} 记录在全局信息 $List$ 列表内.

步骤8. 计算子种群内萤火虫之间的相对吸引度:

$$\beta = \beta_0 e^{-\gamma r_{ii}^2} \quad (18)$$

其中, r_{ii} 表示子种群 j 中第 i 和第 i' 只萤火虫之间的空间距离, 定义为 $r_{ii'} = \|X_{ij} - X_{i'j}\|$, $X_{i'j}$ 表示子种群 j 中第 i' 只萤火虫的位置.

各子种群中萤火虫个体开始位置进化, 根据式(19)更新空间位置:

$$X_{ij}(t+1) = X_{ij}(t) + \beta(X_{i'j}(t) - X_{ij}(t)) + \alpha(rand - 1/2) \quad (19)$$

其中, α 通过式(15)在每一次迭代前更新大小, $rand$ 为在 $[0, 1]$ 内服从均匀分布的随机因子, t 表示当前为第几次迭代.

根据式(16)修改 $X_{ij}(t+1)$ 的值, 使其不超过步骤4结果中的上下限.

步骤9. 查询 $List$ 列表数据, 若存在某个子种群 I_{0j}^{best} 值在最近10次迭代变化范围小于 10^{-3} , 在 $List$ 列表中寻找其他子种群中最优秀的萤火虫个体.

步骤10. 如果步骤9结果存在则向其学习, 如果不存在则通过遗传算法中的变异操作来调整该种群最优萤火虫个体的位置, 从而增加该子种群的种群多样性, 如式(20)、式(21)所示:

$$\eta = 1 - \frac{t}{T_{\max}} \quad (20)$$

$$X_{ij} = X_{ij} + \eta \times X_{ij} \times N(0, 1) \quad (21)$$

其中, η 为根据迭代次数 t 不断调整的值; $N(0, 1)$ 为满足高斯分布均值为0, 方差为1的随机值.

步骤11. 判断是否达到最大迭代次数, 若达到转步骤12, 没有达到转步骤7继续迭代过程.

步骤12. 收集步骤2中的 a 个监测站分别迭代后获取的 a 组污染物位置 x , 污染物排放时间 t 污染物排放浓度 M 的数据^[24]. 由式(2)可知, 在一维模型情况下, 给定 x 时, t 也唯一确定, 并且 M 是由确定的 x_0 , t_0 作为

前提条件求得,因此污染物位置 x 对其他参数有直接影响,求 a 组数据中污染物位置 x 的标准差,若标准差过大,代表某个监测站存在不准确浓度数据,将其溯源结果排除,输出排除后的结果。

4 案例分析

为了验证提出的改进 FA 算法的可行性,本文采用文献 [25] 美国地质勘探局公布的 2006 年在特拉基河

做的染料示踪实验数据.为了研究混合不均匀情况下监测值对溯源结果的影响^[24],采用文献中低流量 (4.05–18.04 m³/s) 情景,在监测点 1 处投放 rhodamine WT 染料 0.82 kg、在监测点 2–6 处设置监测断面,监测点 1–12 位置如图 2 所示,在监测断面 2 处染料与河流混合不均匀,不同断面不同时刻的染料浓度系列为文献 [25] 中数据,监测点 2–6 距染料投放点 1 处的位置关系如表 1 所示。

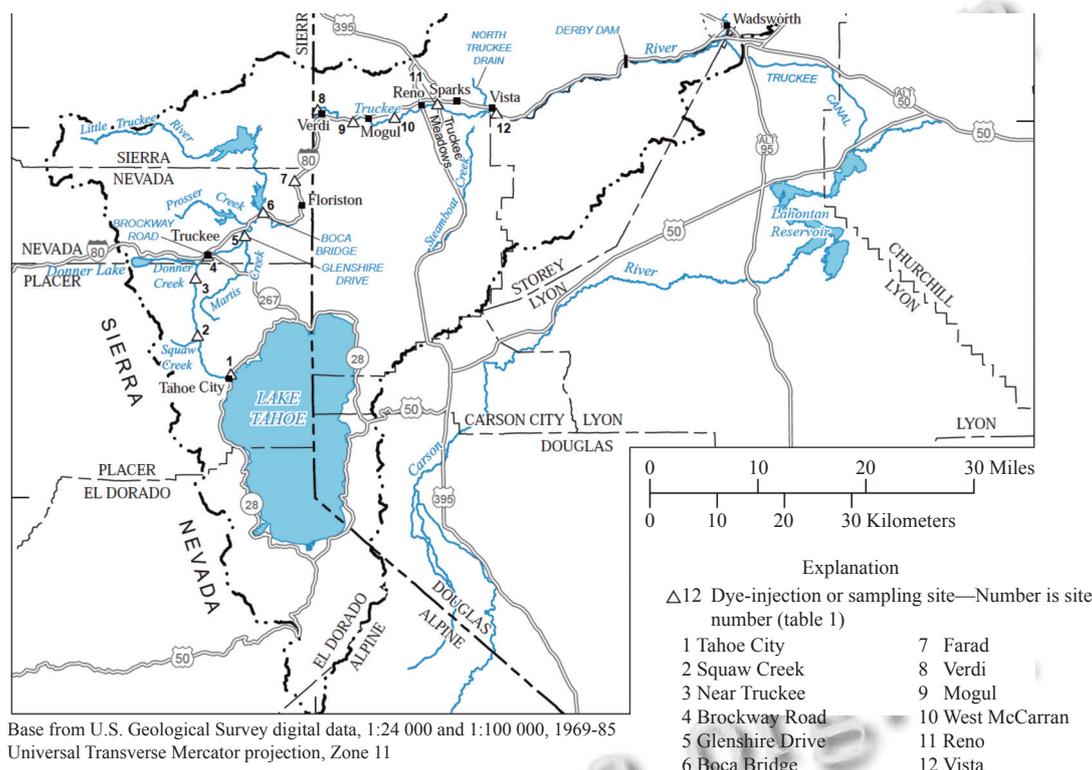


图2 特拉基河监测站点图

4.1 溯源参数优化调整

本文使用特拉基河监测断面的数据,采用改进的 FA 的算法对污染物的排放相关的 x_0, t_0, M 进行求解.由于算法设置萤火虫子种群个数 n ,每个子种群萤火虫的个数 m/n ,光强吸收系数 γ 、最大吸引度因子 β_0 ,步长因子 α 和最大迭代次数 T_{max} 等参数对目标函数,求解时间等有较大影响.因此本文通过多次运行实验程序,代入不同的萤火虫算法相关参数,进行对比实验,最终选取的相关参数如表 2 所示。

表 2 中 FA 算法为单种群算法,只需要输入一组参数;而本文采用的改进 FA 算法在案例分析中子种群数量设置为 3,3 个子种群各代入一组参数,由于每组参数

中 γ, β_0, α 值不相同,因此各个子种群具有不同的迭代过程,从而避免各子种群同时陷入局部极值的情况,当某个子种群陷入局部极值时可向其他子种群的较优个体学习,跳出局部极值,增加了整个算法的种群多样性,增强了改进的 FA 算法跳出局部极值能力,提高了通过改进 FA 算法溯源的准确性。

表 1 示踪剂投放监测断面分布情况

监测点序号	流量 (m ³ /s)	距投放点长度 (m)
1	1.87	0
2	4.05	9.9
3	5.49	20.33
4	7.53	24.77
5	7.56	32.32
6	18.04	40.25

表2 算法参数表

算法	β_0	α	γ	T_{max}	m	n
FA	1	0.97	1	500	10	1
	1	0.97	0.99	500		
改进的FA	0.99	0.96	0.99	500	30	3
	0.98	0.98	1	500		

针对河流的水文参数,由文献[22]可大致估计,污染物的降解系数 k 为 $1.7 \times 10^{-10} \text{ s}^{-1}$,监测点2-6号处河流平均流速 u 为15 m/s,河流长度方向扩散系数 D_x 为 $40 \text{ m}^2/\text{s}$. D_x 为估计值,实际河流中存在 D_x 、 D_y 、 D_z 三种扩散方向,因此需要对参数进行调整.为保证结论正确性,将监测数据分为两组:监测断面2-监测断面5号数据设置为实验集,用来验证改进FA算法溯源的准确性;监测断面6号数据设置为训练集[25],用来调整河流的水文参数.计算方式采用式(9),将 D_x 设为待求参数,污染源相关参数设为已知,通过改进的FA算法通过Matlab软件进行100次迭代试算,选取可以使目标函数相对最优的 D_x 数据为 $15.63 \text{ m}^2/\text{s}$.

4.2 改进FA算法结果分析

将修正后的水文参数应用到实验集监测断面2-监测断面6号数据中,分别采用改进的FA算法和FA算法对目标函数式(9)、式(12)进行求解,从而对污染源参数排放位置 x_0 、排放时间 t_0 、排放强度 M 进行求解,改进的FA算法和FA算法的迭代过程图如图3、图4所示,其中图3为求解目标函数式(9)即求解参数排放位置 x_0 、排放时间 t_0 的迭代过程图,图4为目标函数式(12)即求解参数排放强度 M 的迭代过程图.表3为通过Matlab软件运行100次,去掉偏差过大数据后对剩余数据求平均值得到的数据.

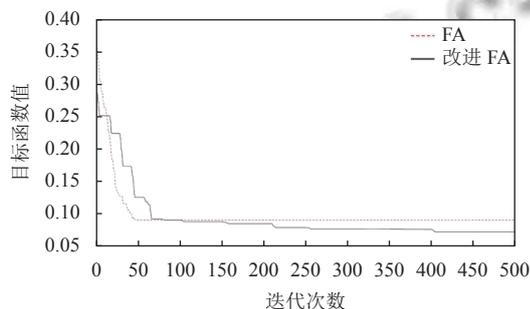


图3 目标函数1迭代过程图

通过图3、图4和表3可知,FA算法在迭代到50代左右时,开始快速收敛,但与真实值相比求解结果偏差较大,问题为陷入局部极值,种群多样性较低,没办法跳出局部极值.改进的FA算法在300代左右开始

收敛,并且图像呈现阶梯下降趋势,虽然迭代速度相对于FA慢,但求解精度高,具有以下优点:由于引入了自适应步长策略,使算法前期全局搜索能力较强,后期局部搜索精度更高所以具有更好的种群多样性,全局搜索能力更强;引入了多种群互相学习策略,当某个种群陷入局部极值时会向其他种群学习或者自身进行高斯扰动,所以具有更好的跳出局部极值能力.

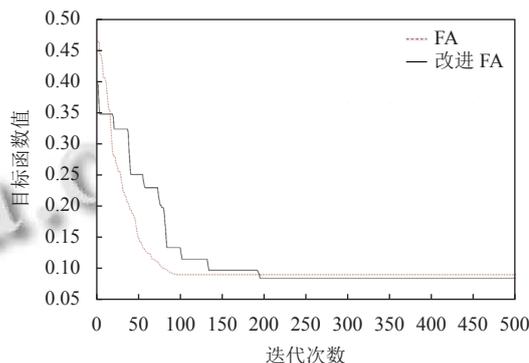


图4 目标函数2迭代过程图

表3 多断面溯源结果

算法	参数	真实值	计算值			
			断面1	断面2	断面3	断面4
FA	x_0 (m)	0	533.64	332.73	288.73	-227.13
	t_0	22:30	22:01	22:52	22:49	22:12
	M (g)	820	1026.73	930.21	897.54	730.42
改进FA	x_0 (m)	0	323.49	127.42	165.24	-156.36
	t_0	22:30	22:08	22:22	22:17	22:35
	M (g)	820	895.36	835.43	841.47	801.45

由表3可知,断面2计算值明显偏大,根据本文第3.2节算法步骤,对断面2-5溯源得到的污染源位置 x_0 的标准差和去掉断面2数据溯源结果对比如表4所示,排除断面2的数据后,标准差明显降低.

表4 多断面溯源位置标准差

类型	断面2-5	断面3-5
标准差	173.04	143.52
平均值	114.95	45.43

通过上述溯源结果分析,改进的FA算法溯源结果污染源位置范围 $[-156.36, 165.24] \text{ m}$ 、污染源排放时间范围 $[22:17, 22:35]$ 、污染源排放强度范围 $[801.45, 895.36] \text{ g}$ 相对于污染源真实值排放位置 0 m 、排放时间 $22:30$ 、排放强度 820 g 偏差不大.本文方法在实际的河流污染物溯源分析中具有相对的准确性,并且在通过标注差排除异常监测断面溯源结果后,不同断面溯源结果相对真实值偏差范围在可接受范围内.

5 结论与展望

本研究采用耦合概率密度法和一维水质扩散模型进行建模,通过改进的FA算法对模型进行求解,为了检验算法可靠性采用特拉基河的示踪剂实验真实场景数据.在实验求解过程中通过多次试算的形式对改进FA算法的最大迭代次数、子种群个数等参数进行合理选取,在确定河流水文参数时,结合资料给定的水文参数(水流速度、降解系数、扩散系数)和通过算法分析对河流长度方向扩散系数进行修正,进一步提高在对河流突发水污染事件溯源的可靠性.最后通过多监测断面分别进行溯源求解,采用标准差分析法,排除了因为污染物混合不均匀导致的监测数据偏差.通过特拉基河染料示踪实验表明,该方法在单污染源识别问题上,其精度高于原始的FA算法,具有一定的可靠性.

参考文献

- 张芳,王炜亮,成杰民.流域突发性水污染事故应急体系构思.环境科技,2010,23(1):57-60.[doi:10.3969/j.issn.1674-4829.2010.01.016]
- 程伟平,廖锡健.基于逆向概率密度函数的一维污染源排放重构.水动力学研究与进展A辑,2011,26(4):460-469.[doi:10.3969/j.issn1000-4874.2011.04.010]
- 刘晓东,姚琪,薛红琴,等.环境水力学反问题研究进展.水科学进展,2009,20(6):885-893.
- Sidauruk P, Cheng AHD, Ouazar D. Ground water contaminant source and transport parameter identification by correlation coefficient optimization. Groundwater, 1998, 36(2): 208-214. [doi: 10.1111/j.1745-6584.1998.tb01085.x]
- Alapati S, Kabala ZJ. Recovering the release history of a groundwater contaminant using a non-linear least-squares method. Hydrological Processes, 2000, 14(6): 1003-1016.
- 白玉莹,卢文喜,李久辉.卡尔曼滤波方法在地下水污染源反演中的应用.中国环境科学,2019,39(8):3450-3456.[doi:10.3969/j.issn.1000-6923.2019.08.039]
- 张双圣,强静,刘汉湖,等.基于贝叶斯公式的地下水污染源识别.中国环境科学,2019,39(4):1568-1578.[doi:10.3969/j.issn.1000-6923.2019.04.027]
- 刘晓东,王珏.地表水污染源识别方法研究进展.水科学进展,2020,31(2):302-311.
- 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.水电能源科学,2014,32(7):52-55,136.
- 饶清华,曾雨,张江山,等.闽江下游突发性水污染事故时空模拟.环境科学学报,2011,31(3):554-559.
- 刘洁,陈昊辉,张丰帆,等.基于改进遗传算法的河流水质模型多参数识别.东北农业大学学报,2020,51(1):73-82.[doi:10.3969/j.issn.1005-9369.2020.01.009]
- 吴一亚,金文龙,吴云波,等.宽浅河道瞬时源游项反问题及反演精度主要影响因子分析.水资源保护,2015,31(5):58-61.[doi:10.3880/j.issn.1004-6933.2015.05.011]
- 陈海洋,滕彦国,王金生,等.基于Bayesian-MCMC方法的水体污染识别反问题.湖南大学学报(自然科学版),2012,39(6):74-78.
- 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.水利学报,2015,46(11):1280-1289.
- 韩龙喜,陈丽娜.可降解的地表水污染物随机游走改进模式.水科学进展,2009,20(5):689-694.[doi:10.3321/j.issn:1001-6791.2009.05.014]
- 杨海东,肖宜,王卓民,等.突发性水污染事件溯源方法.水科学进展,2014,25(1):122-129.
- 范玉科,侯为根,徐浩.河道污染的一维对流-扩散方程源项识别反问题.安徽工业大学学报(自然科学版),2014,31(3):323-327.
- 王忠慧,贡力,康春涛,等.基于BAS算法的河渠突发水污染溯源.水资源保护,2020,36(5):87-92.[doi:10.3880/j.issn.1004-6933.2020.05.013]
- Neupauer RM, Wilson JL. Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resources Research, 1999, 35(11): 3389-3398. [doi: 10.1029/1999WR900190]
- Cheng WP, Jia YF. Identification of contaminant point source in surface waters based on backward location probability density function method. Advances in Water Resources, 2010, 33(4): 397-410. [doi: 10.1016/j.advwatres.2010.01.004]
- 符强,童楠,钟才明,等.基于改进型进化机制的萤火虫优化算法.计算机科学,2014,41(3):228-231,248.[doi:10.3969/j.issn.1002-137X.2014.03.049]
- 符强,童楠,赵一鸣.一种基于多种群学习机制的萤火虫优化算法.计算机应用研究,2013,30(12):3600-3602,3617.[doi:10.3969/j.issn.1001-3695.2013.12.021]
- 陈小雪,尉永清,任敏,等.基于萤火虫优化的加权K-means算法.计算机应用研究,2018,35(2):466-470.[doi:10.3969/j.issn.1001-3695.2018.02.031]
- 陶宇夏,蒋晶,魏永长,等.基于差分进化算法的河流突发性污染事故溯源.工业工程与管理,2021,26(2):9-14.
- Crompton EJ. Traveltime for the Truckee River between Tahoe City, California, and Vista, Nevada, 2006 and 2007. U.S. Geological Survey, 2008.

(校对责编:牛欣悦)