

基于改进 AFSA 算法的河流突发水污染溯源^①



李欣欣^{1,2}, 王 宁², 姜秋隍³, 刘 枢³, 魏建勋⁴, 张 楠⁴

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)

⁴(阜新市生态环境保护服务中心, 阜新 123000)

通讯作者: 王 宁, E-mail: wangning@sict.ac.cn

摘 要: 污染源的排放时间、排放位置以及排放总量的确定是河流突发性水污染溯源问题的关键. 如何快速准确的确定重金属污染源的三个因素, 然后通过 GIS 得到污染源企业的排查清单是本文的研究重点. 本文通过对河流的水文水质的研究, 依据重金属污染物的特性, 确定了一维河流重金属污染物的时空变化的动态解析解. 同时构造了污染物时空溯源模型以及污染物排放总量模型, 并利用改进的 AFSA 算法实现了模型的求解. 研究表明, 该算法使得模型能够更加快速准确地得到三个参数的结果, 然后将该结果通过本文的方法并且借助 GIS 技术更快更准确的为相关工作人员提供污染源企业概率排查清单. 本文提出的方法和模型对于水污染处理以及保护水环境具有一定的指导意义.

关键词: 突发性水污染; 重金属; AFSA 算法; 溯源; GIS

引用格式: 李欣欣, 王宁, 姜秋隍, 刘枢, 魏建勋, 张楠. 基于改进 AFSA 算法的河流突发水污染溯源. 计算机系统应用, 2020, 29(7): 139-144. <http://www.c-s-a.org.cn/1003-3254/7551.html>

Traceability Method of Sudden River Water Pollution Based on Improved AFSA Algorithm

LI Xin-Xin^{1,2}, WANG Ning², JIANG Qiu-Li³, LIU Shu³, WEI Jan-Xun⁴, ZHANG Nan⁴

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)

⁴(Fuxin Ecological Environmental Protection Service Center, Fuxin 123000, China)

Abstract: The key to solve the problem of tracing the source of sudden water pollution in rivers is to determine its discharge time, location, and total amount. This study proposes two models which can quickly and accurately determine these three factors of heavy metal pollution sources and obtain the list of pollution sources through GIS. This study determines the dynamic analytical solution for the spatiotemporal changes of heavy metal pollutants in one-dimensional rivers, through the study of the hydrological and water quality of rivers and the characteristics of heavy metal pollutants. At the same time, a spatial-temporal traceability model of pollutants and a total pollutant discharge model are constructed, and this models are solved using the improved AFSA algorithm. The research results show that the algorithm enables the model to obtain the results of the three parameters more quickly and accurately, and then passes the results through the proposed method and uses GIS technology to provide the relevant staff with a probabilistic checklist of pollution source enterprises more quickly and accurately. The methods and models proposed in this study have certain guiding significance for water pollution treatment and protection of the water environment.

Key words: sudden water pollution; heavy metal; AFSA algorithm; traceability; GIS

① 基金项目: 国家科技重大专项 (2018ZX07601001)

Foundation item: National Science and Technology Major Program (2018ZX07601001)

收稿时间: 2019-12-21; 修改时间: 2020-01-29; 采用时间: 2020-03-11; csa 在线出版时间: 2020-07-03

近年来,国内外突发性重金属水污染事件屡屡发生,由于其发生的突然性和危害的严重性会威胁到广大人民的人身和财产安全,给社会经济的发展带来不可估量的损失,因此国内外对此事件的关注极高,并不断地寻求解决办法^[1].在我国水资源最大的特点是地域分布不均,而我国又是人口大国,缺水状况也是我国一直在解决的问题.我国有跨国界河流40多条,涉及19个主要国家^[2],一旦发生严重的事件,重金属污染物不只对水环境和人体的危害极大,甚至会影响社会治安和国际友好关系.经调查分析,突发性重金属水污染的发生主要是由企业违规排放废水和工厂事故泄漏造成的.因此在自动监测站监测到河水中重金属污染物超标时,能否快速精确的定位污染源企业或工厂,对于后续采取有效措施具有现实意义.本文研究的是由河流附近企业违规排放废水导致的突发性重金属水污染事故的追踪溯源.

水动力学反演法,根据河流的水文特征选取合适的水动力学方程,再根据真实的污染物浓度监测数据、污染源的分布数据等,通过构造相关模型并利用相关方法反演求解得到污染源的排放位置、排放时间和排放质量^[2].目前,主要应用了遗传算法、粒子群算法、模拟退火算法等^[3,4],但存在容易得到局部最优和搜索时间较长的问题.基于此类问题,本文选用改进后的AFSA算法,算法具有更强的跳出局部极值的能力和更快的搜索能力.

1 水动力反演

重金属污染源的追踪溯源需要水动力学反演,因此需要选择合适的水动力方程进行模拟,从而得到水动力模拟库,建立相关模型,快速准确的定位重金属污染源.

1.1 重金属污染物的时空变化

在水动力学中,模拟污染物在一维河道中的迁移变化规律可以用以下公式表示:

$$\frac{\partial C}{\partial t} = D_x \frac{\partial^2 C}{\partial x^2} - u_x \frac{\partial C}{\partial t} - KC \quad (1)$$

式(1)中, C 表示河流在 x 断面在 t 时刻污染物的浓度(单位是 mg/L), D_x 表示河流纵向的弥散系数(单位是 m^2/min), u_x 表示河流纵向平均流速(单位 m^2/min), K 表示污染物的衰减系数(单位是 s^{-1})^[5].

本文根据实际情况研究的是瞬时点源排放重金属

污染物,故假设在河流的一个断面 x 处投入质量为 M 的重金属污染物,此时污染物的衰减系数 K 为0,河流的纵向平均流速为 u_x .此时污染物与断面的河水混合,那么方程的初始条件和边界条件是:

$$\begin{cases} C(x,0) = 0, & x \geq 0 \\ C(0,t) = C_0\delta(t), & t > 0 \\ C(+\infty,t) = 0, & t > 0 \\ K = 0 \end{cases} \quad (2)$$

式(2)中,污染物的初始浓度为 $C_0 = M/Q$, M 表示瞬时投放到河流中的污染物的质量(单位是 kg), Q 表示单位时间内河水的流量(单位是 m^3/s)^[5], A 表示河流的断面面积(单位是 m^2).

通过函数 δ 的特性和Laplace的变换得到式(1)在式(2)条件下的解析解,即一维河流中重金属污染物的时空变化方程:

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

式(3)中, $C(x,t)$ 表示污染物在下游的污染物浓度分布(单位是 mg/L), x_0 表示污染源的位置(单位是 m), t_0 表示污染源的排放时刻(单位是 min)^[5],其余同上.

通过式(3)可知,对于突发性重金属水污染的追踪溯源问题,需要确定污染源的排放位置、排放时间和排放量这3个因素.

1.2 水动力情景模拟

如果重金属水污染事故突然发生,自动监测站报警,此时已知自动监测站处的水文信息,包括监测站处河流的断面面积 A 、河流的纵向弥散系数 D_x 、河流的纵向平均流速 u_x 以及监测站监测到的某段离散时间内污染物浓度变化数据等.记自动监测站处得到的随时间变化的污染物浓度数据为 C_i ($i = 1, 2, \dots, n$).

根据自动监测站处的重金属污染物的污染范围信息得重金属污染源的位置范围,即根据情景,假设干流自动监测站 $a(x_a)$ 从未监测到污染物,而在 a 下游的干流自动监测站 $b(x_b)$ 在时刻 T 第一次报警,监测到了污染物,故得到污染源的排放位置范围是 (x_a, x_b) ,而污染源的排放时间与自动监测站 $b(x_b)$ 监测到污染物的时间间隔范围是 $(0, t_{\max})$,其中 t_{\max} 可以由以下公式表示:

$$t_{\max} = \frac{x_b - x_a}{u_x} \quad (4)$$

通过上述过程已知污染源的位置范围和排放时间

范围,由式(3)进行水动力情景模拟.为了便于情景模拟,可以将式(3)进行变形得到以下公式:

$$C(x,t) = \frac{M}{2A\sqrt{D_x\pi t}} \exp\left[-\frac{(x-u_x t)^2}{4D_x t}\right] \quad (5)$$

式(5)中, $C(x,t)$ 表示污染物在自动监测站 $a(x_a)$ 下游的污染物浓度分布,单位是g/L; x 表示污染源距离自动监测站 $b(x_b)$ 的距离长度,单位是m; t 表示污染源在自动监测站 $b(x_b)$ 监测到污染物时刻前 t 分钟排放的污染物,单位是min,其余同上.

在位置和时间范围内进行突发性重金属污染源的情景模拟.假设在重金属污染源的参数是 $[x',t',M']$,代入式(5)得到与自动监测站 $b(x_b)$ 相同的一段离散时间内一系列的污染物的浓度变化数据,记为 C'_i ($i=1,2,\dots,n$).

2 溯源模型建立

重金属污染源的追踪溯源需要位置、时间和质量3个参数,若将重金属污染源追踪溯源问题的3个参数作为一个整体未知参量进行求解,需要构造复杂的模型,模型的复杂会出现溯源时间过长,效率比较低等问题,故不能快速准确地找到污染源.因而本文构建两个模型进行重金属污染物的溯源,将位置和时间这两个参数作为一个整体未知参量建立时空溯源模型,将质量作为一个未知参量建立污染物排放量模型.

2.1 时空溯源模型

假定重金属污染物的排放质量是 M' ,将上述的自动监测站的离散时间的重金属污染物浓度变化序列值 C_i ($i=1,2,\dots,n$)与水文情景模拟得到位置和时间范围内一系列的在相同离散时间内的重金属污染物浓度变化序列值 C'_i ($i=1,2,\dots,n$)进行拟合度分析,利用相关系数 R 来判断真实值与模拟值的相关性.

在时空溯源模型中,未知参数量是位置和时间,质量设为 M' .故若情景模拟的重金属污染源的位置和时间极为接近于真实重金属污染源的位置和排放时间那么相关系数 R 的值也极为接近1^[6].根据相关系数 R 的值从大到小排列, R 越大越大代表是重金属污染源的可能性越大,排序越靠前.相关系数的公式如下:

$$R = \frac{\sum_{i=1}^n (C_i - \bar{C})(C'_i - \bar{C}')}{\sqrt{\sum_{i=1}^n (C_i - \bar{C})^2 \sum_{i=1}^n (C'_i - \bar{C}')^2}} \quad (6)$$

$$\bar{C} = \frac{\sum_{i=1}^n C_i}{n} \quad (7)$$

$$\bar{C}' = \frac{\sum_{i=1}^n C'_i}{n} \quad (8)$$

式中, C_i 表示离散时间内时刻 t 的污染物浓度, \bar{C} 表示离散时间内污染物浓度的平均值, C'_i 表示情景模拟的离散时间内时刻 t 的污染物浓度, \bar{C}' 表示情景模拟的离散时间内污染物浓度的平均值.

根据相关系数 R 存在最大值的性质,为了从一系列的 $[x',t',C'_i]$ ($i=1,2,\dots,n$)中筛选出准确的结果 (x',t') ,故构建以下目标函数:

$$F_1 = \max(R) \quad (9)$$

函数的约束条件为:

$$x_a < x' < x_b; 0 < t' < t_{\max} \quad (10)$$

由式(4)~式(10)构成了时空溯源模型,通过时空溯源模型得到重金属污染源的位置和排放时间,再通过构建污染物排放量模型.

2.2 污染物排放量模型

通过时空溯源模型,重金属污染源的排放位置和排放时间已经确定,即可在此模型中作为已知条件,此时重金属污染物溯源的未知参数只有污染源排放污染物的质量.故由式(11)可以大致确定源强度范围.

$$M = \frac{1}{n} \sum_{i=1}^n \frac{C_i}{C'_i} \times M' \quad (11)$$

式中, M 表示所求重金属污染物的质量(单位是g), M' 表示上述过程中假设的重金属污染物的质量(单位是g)^[5],其余同上.

为了确定污染源排放该重金属的总量,通过构建以下函数求最小值:

$$F(M) = \min\left(\sum_{i=1}^n \frac{(C_i - C'_i)^2}{C_i + 1.0}\right) \quad (12)$$

函数的约束条件为:

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

为了方便求解,同时空溯源模型一样将目标函数变为求最大值,目标函数的公式为:

$$F_2 = \max(-F(M)) \quad (14)$$

由式(11)~式(14)构成了污染物排放量模型,通过模型得到重金属污染源的排放量.

3 溯源模型求解

在上述过程中,已经把突发性水污染溯源的两个模型转化成了关于两个求最值的优化模型,本文选用改进后的 AFSA 算法进行模型求解。

改进的 AFSA 求解时空溯源模型时,选取 $F = F_1$ 作为适应度目标函数,人工鱼个体状态是 $X = (x, t)$, 其当前所在位置的食物浓度函数用 $Y = F(X)$ 表示^[7], 求最大值. 人工鱼群算法求解污染物排放量模型时,选取 $F = F_2$ 作为适应度目标函数,人工鱼个体状态是 $X = M$, 其当前所在位置的食物浓度函数用 $Y = F(X)$ 表示^[7], 求最大值。

3.1 改进步长和视野

步长和视野范围是算法的关键参数,步长和视野范围是算法的关键参数,若希望算法的全局搜索能力更强并且收敛迅速,则需要设置较大的视野;若希望算法的局部搜索能力较强,则需要设置较小的视野. 步长大则收敛快,但是步长过大会发生震荡现象,步长小则收敛慢但是求解精度高^[8]. 故本文参考文献^[9]通过式 (15) 动态调整视野 $Visual$ 和步长 $Step$ 增强算法的搜索能力和精确度^[9]:

$$\begin{cases} Visual = Visual_{min} + Visual \times a \\ Step = Step_{min} + Step \times a \\ a = \exp\left(-30 \times \left(\frac{gen}{maxGen}\right)^2\right) \end{cases} \quad (15)$$

式中, $Visual = x_{max}/4$, $Step = visual/8$, $Visual_{min} = 0.001$, $Step_{min} = 0.0002$ 分别为人工鱼的视野、步长的最小值, gen 表示当前的迭代次数, $maxGen$ 表示最大的迭代次数, x_{max} 表示搜索范围的最大值^[9].

3.2 改进觅食行为

在满足前进条件的前提下,只有当随机选择的的状态比当前的状态好,原觅食行为才会选择向该方向移动一步,这样存在搜索速度慢的问题. 故为了加快搜索速度,当前人工鱼 i 的状态为 X_i , 通过式 (16) 随机得到状态 X_j , 如果满足前进条件 $Y_j > Y_i$, 则将当前人工鱼 i 直接移动到 X_j 状态, 如果不满足前进条件则重新按照式 (16) 随机得到状态 X_j . 反之, 当达到 try_number 次后仍然不满足前进条件, 则通过式 (17) 随机生成状态 X_j , 将当前人工鱼 i 直接移动到 X_j 状态。

$$X_j = X_i + (\text{rand}() - 0.5) \times Visual \times X_i \quad (16)$$

$$X_j = X_i + (\text{rand}() - 0.5) \times Step \times X_i \quad (17)$$

式中, $i, j = 1, 2, \dots, fishnum$, $\text{rand}()$ 为 $[0, 1]$ 之间的随机数。

3.3 改进的 AFSA 算法流程

图 1 所示的算法流程中初始化设置算法参数包括: 确定人工鱼群规模 $fishnum$ 、迭代次数 gen 、最大迭代次数 $maxGen$ 、拥挤度因子 δ 、尝试次数 try_number 、个体间的距离 $d_{i,j}$ ^[8]。

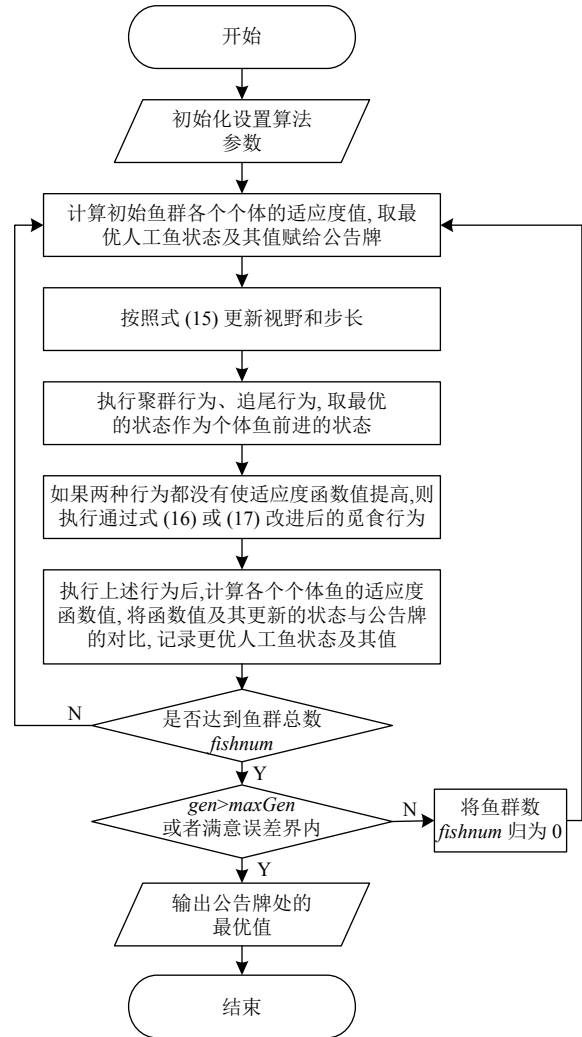


图 1 改进的 AFSA 算法流程图

4 污染源的排查

通过调查分析, 根据实际情况, 通过上述模型得到污染源在一维河道的位置, 排放时间以及排放量后筛选得到的涉铜企业名单, 当企业距离河流越近并且企业的信誉越低, 则是污染源企业的可能性越高. 故用 L 表示企业到纳污河流的距离, 用 E 表示企业的信誉度, 其等级分为优/良 (用 0 表示), 差 (用 1 表示). 通加入权值后得到公式:

$$F_i = w_1 \frac{1}{L_i} + w_2 E_i \quad (18)$$

式中, $w_1 = 0.9975$ 、 $w_2 = 0.0025$, L_i 表示企业*i*到河流的距离, E_i 表示企业*i*的信誉度, F_i 表示加入权值后企业*i*为污染源的函数值。

当上述企业名单有*k*个企业时, 通过概率计算, 将企业名单根据概率从大到小排列, 从而得到相关工作人员的排查名单. 概率计算公式为:

$$P_i = \frac{F_i}{\sum_{i=1}^k F_i} \quad (19)$$

5 算法验证

本文研究的是瞬时点源突然超标排放重金属铜导致的河流重金属水污染, 并利用已有的 GIS 进行结果展示^[10].

5.1 背景

以自动监测站 *b* 处铜突发污染超标报警为例, 2018 年的 12 月 16 日上午 11 点, 自动监测站 *b* 第一次铜超标报警, 此时监测值为 7.5 mg/L. 而在同一时刻图 2 (黑点代表企业, 三角形代表涉铜企业) 标注的正常运行的自动监测站 *a*、自动监测站 *c* 以及其他在图 2 中未标注的正常运行的自动监测站并没有发生铜突发污染超标报警, 故污染源存在于自动监测站 *a* 和自动监测站 *b* 之间. 将该时间作为初始监测时间, 然后每间隔 1 小时监测一次, 监测到下午 14 点结束. 自动监测站 *b* 处的水文信息如表 1 所示.



图 2 河流地图

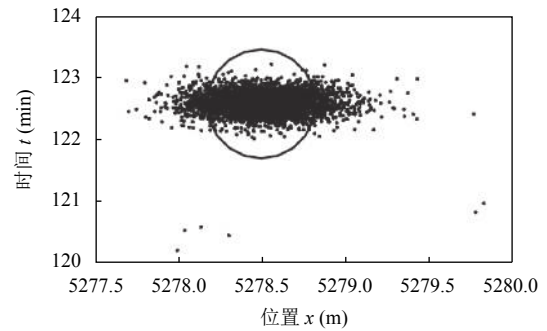
表 1 自动监测站 *b* 的水文信息

参数	纵向弥散系 D_x (m^2/min)	河流断面面积 A (m^2)	河流纵向流速 u_x (m/min)
数值	1350	24	36

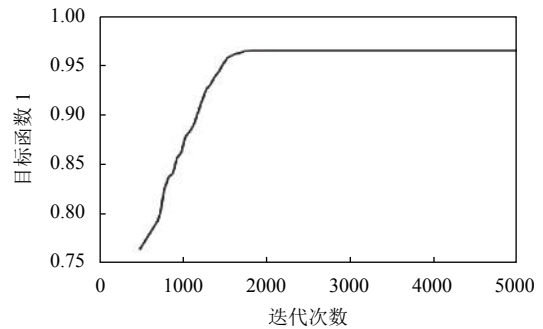
5.2 溯源结果

通过上述两个模型得到运行结果图 3 和图 4. 图 3

中的寻优过程图中圆的圆心坐标数据为表 2 中溯源结果的污染源位置和污染源排放时间, 图 4 中的寻优过程图中圆的圆心坐标数据为表 2 中溯源结果的污染源排放量. 表 2 中, 污染源位置即河流中污染源的位置距离自动监测站 *b* 的距离, 污染源排放时间即污染源排放时刻与自动监测站 *b* 第一次报警的时刻的时间差, 污染源排放量即污染源的排放质量.



(a) 寻优过程图



(b) 迭代过程图

图 3 目标函数 1

5.3 溯源企业分析

通过上述结果可以看出上述两个模型能够快速准确地得到污染源的 3 个参数, 之后根据这 3 个参数并利用 GIS 地图得到图 5. 其中标注圆的半径通过式 (20) 确定.

$$r = \frac{\sum_{i=1}^n L_i}{n} \quad (20)$$

其中, L_i 代表第*i*个涉铜企业到其纳污河流的距离 ($i = 1, 2, \dots, n$).

图 5 中红色圆内涉铜污染源企业有 6 个, 表明这 6 个企业可能为污染源企业, 再由式 (18) 和式 (19) 计算这 6 个企业为污染源的的概率并按照概率从小到大排列得到概率排查清单, 即表 3.

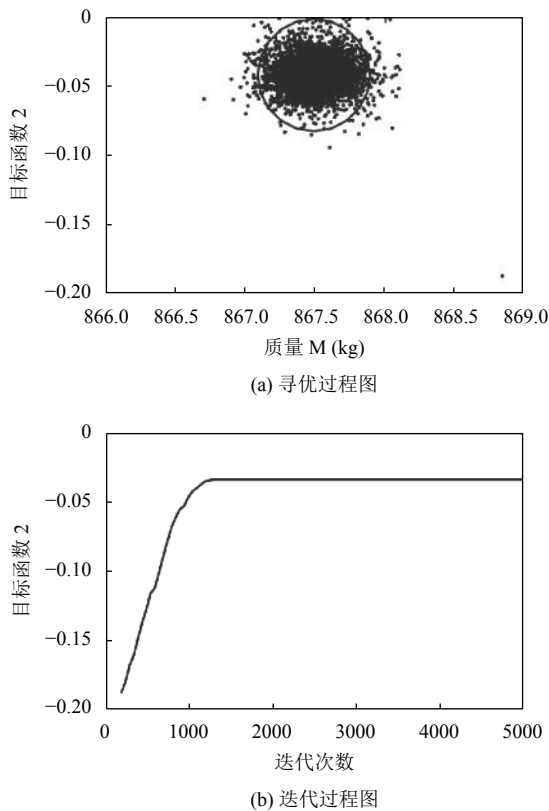


图4 目标函数2

表2 溯源结果分析

参数	真实值	溯源结果	相对误差 (%)
污染源位置 x	5216 m	5278.40 m	1.20
污染源排放时间 t	120 min	122.40 min	2.0
污染源排放量 M	860 kg	867.31 kg	0.85



图5 溯源后的河流地图

表3 概率排查清单 (单位: %)

企业	概率
企业6	26.1
企业3	22.2
企业5	10.4
企业4	6.5
企业1	5.2
企业2	3.5

当时发生重金属铜污染后, 相关工作人员通过走访排查自动监测站 b 的上游相关涉铜企业进行污染源企业的追踪溯源, 最终确定为企业 6 导致此次河流重金属铜污染事故的发生, 与表 3 中概率最大的企业一致.

6 结论

通过实验发现概率最高的企业 6 就是人工排查得到的污染源企业. 目前已有的研究工作主要是求得污染源的排放位置, 排放时间以及污染物排放量这 3 个参数后再通过人工排查确定污染源企业, 没有明确的人工排查顺序, 故排查过程耗时耗力, 不能及时的找到污染源企业从而采取相应的措施处理问题. 实际上这 3 个参数值总因为各种因素导致存在相应的误差, 本文提出的方法能在水文信息较少的情况下能更加快速准确地求得污染源的 3 个参数, 并且考虑到由于各种因素导致这 3 个参数的误差无法完全排除, 构造相关公式并考虑到企业信誉度因素, 最终为相关工作人员提供排查的概率清单, 进而快速准确地找到污染源企业, 及时采取措施解决突发性重金属铜水污染问题, 保护好我们的水环境.

参考文献

- 田源. 突发性水污染事件应急法律制度研究[硕士学位论文]. 太原: 山西财经大学, 2016.
- 郑军, 张立, 杨常青, 等. 跨国界流域重金属污染溯源体系框架初步构建. 水资源保护, 2015, 31(6): 57-61. [doi: 10.3880/j.issn.1004-6933.2015.06.009]
- 王浩. 基于优先节点定位和人工鱼群优化的 DV-Hop 算法研究[硕士学位论文]. 徐州: 中国矿业大学, 2016.
- Jha M, Datta B. Three-dimensional groundwater contamination source identification using adaptive simulated annealing. Journal of Hydrologic Engineering, 2012, 18(3): 307-317.
- 杨海东. 河渠突发水污染追踪溯源理论与方法[博士学位论文]. 武汉: 武汉大学, 2014.
- 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.
- 刘刚. 一种人工鱼群算法及其应用研究[硕士学位论文]. 上海: 华东理工大学, 2015.
- 姚凌波. 人工鱼群智能优化算法的研究及应用[硕士学位论文]. 无锡: 江南大学, 2017.
- 王联国, 洪毅, 赵付青, 等. 一种改进的人工鱼群算法. 计算机工程, 2008, 34(19): 192-194. [doi: 10.3969/j.issn.1000-3428.2008.19.065]
- 王永桂, 张潇, 张万顺. 流域突发性水污染事故快速模拟与预警系统. 环境科学与技术, 2018, 41(7): 164-171.