

# 基于负梯度法的供水管网污染源识别

信昆仑, 盛希夫, 陶 涛, 项宁银

(同济大学 环境科学与工程学院, 上海 200092)

**摘要:** 利用负梯度法研究供水管网突发污染事故时污染源识别问题。首先针对供水管网的污染源识别问题构建最优化问题数学模型, 然后用 EPANET 作为内嵌的水质模拟计算引擎, 沿着目标函数的负梯度方向进行问题的最优解搜索, 并针对管网算例, 对污染源侵入时间、地点和侵入过程进行计算求解。分析差分步长、梯度求解方法等因素对问题求解效率的影响, 结果表明采用基于水质模拟的差分法求解梯度向量并以水质公差作为差分步长, 具有最好的求解效果。

**关键词:** 供水管网; 污染源识别; 梯度法; 最优化

**中图分类号:** TU991

**文献标志码:** A

## Contamination Source Identification in Water Distribution Networks Based on Negative Gradient Method

XIN Kunlun, SHENG Xifu, TAO Tao, XIANG Ningyin

(College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China)

**Abstract:** This paper discusses the problem of pollution source identification in water distribution systems. First, a mathematical optimization model on the identification problem is built. Then with the EPANET Toolkit as the embedded water quality simulator, the optimization problem is solved with a negative gradient method. The methodology is finally applied to a sample network model published within the EPANET software, and the result shows that it is able to identify the time, location and concentration of the contamination injection sources effectively. To evaluate the performance of the algorithm, impacts of factors such as difference step and gradient method were further discussed, and the results show that, with water quality tolerances as the differential step, the water quality simulation-based finite difference method is an effective algorithm to solve the pollution source identification problem in water distribution system.

**Key words:** water distribution networks; contamination source identification; gradient method; optimization

近几年,城市供水管网的突发污染事故屡有发生,除水源地污染引起管网饮用水问题之外,如管道错接等供水管网内部的原因也会造成水质的局部乃至全局恶化。因此,如何快速准确地定位污染源是饮用水安全保障急需解决的课题。关于污染源的定位研究, Laird 等<sup>[1]</sup>提出了用非线性规划方法来求解这一问题。首先用源头追踪算法替代 EPANET 水质模拟计算引擎从而产生必要的数据,然后通过带正则项的非线性规划方法求出污染源侵入的时间和地点。针对这一方法解的非唯一性问题, Laird 等<sup>[2]</sup>又提出了先用混合整数二次规划的方法确定一个候选污染源节点集,然后在此污染源节点集的基础上利用前一方法筛选出最有可能是污染源的节点。但这个方法对于大中型管网的适用性和此模型与实际管网相符程度仍然有待研究。Guan 等<sup>[3]</sup>利用 EPANET 作为内嵌的水质模拟器再结合负梯度法提出了模拟优化算法。这一方法在已知管网模型的运行模式和水质监测数据充足可靠的前提下能够较好地解出污染源侵入的时间和地点。但只考虑了水质反应为线性的情况,也没有给出确定候选污染源节点的方法。Pries 等<sup>[4]</sup>针对此问题提出了关系树—线性规划算法。通过构造上下游节点的污染物浓度关系树并形成线性规则,再利用关系树的线性规则反向求解线性规划问题,得到污染源节点位置及污染注入属性。Cristo 等<sup>[5]</sup>提出了离散最优化的方法来求解污染源的识别问题。主要通过构造水质污染矩阵来确定可能的污染源注入点,再通过最小化模拟值和监测值的差来从可能的污染源点中确定污染源的位置。另外, Liu 等<sup>[6]</sup>同样利用 EPANET 作为内嵌的水质模拟计算引擎用遗传算法随机搜索最优

解. 由于以上方法均需要管网的水质监测点可以报告实时的污染物浓度, 对于水质监测点仅能判断是否出现污染物的情况, Yang 等<sup>[7]</sup>提出用贝叶斯方法计算已检出污染物的水质监测节点其上游节点为污染源的概, 然后根据概率的大小判断出可能的污染源.

本质上来说, 供水管网污染源的识别问题是管网水质模拟的反演问题, 从数学的角度来看也是一个优化问题. 本文基于 Guan 等<sup>[3]</sup>利用负梯度法求解供水管网污染源识别问题, 结合 Cristo 等<sup>[5]</sup>提出的候选污染源节点集合方法, 进一步考虑供水管网水质反应的复杂性, 用差分步长法代替权重系数法求解梯度方向, 采用 EPANET 作为内嵌的水质模拟计算引擎结合负梯度法来求解污染源侵入的时间、地点和侵入过程, 并结合实例管网进行算法求解性能的分析.

## 1 模型构建

在城市供水管网拥有相当精度的水力水质模型且管网上安装了若各个水质监测点的情况下, 污染源的识别问题可以转化为一个数学上的最优化问题. 本文以模拟时段内各个水质监测点的实际污染物浓度值与模拟值差的平方和为目标函数, 通过分析水质监测点不同时刻模拟浓度值与实测浓度值差来修正计算候选污染源在特定时刻的节点浓度. 待模拟时段内所有的候选污染源节点浓度更新后, 再次进行水质模拟. 当模拟结果达到预设的条件后, 输出候选污染源节点浓度矩阵. 由于求解所得的候选污染源污染物累计注入量与其作为真实污染源的可能性成正比, 因此可以此作为确定污染源节点的依据.

首先定义

$$\begin{aligned} \mathbf{X} &= [x_1(t), x_2(t), \dots, x_n(t)]^T \\ \mathbf{Y} &= [y_1(t), y_2(t), \dots, y_m(t)]^T \end{aligned} \quad (1)$$

式中:  $x_i(t)$  表示候选污染源节点  $i$  在  $t$  时刻的污染物浓度值, 对于  $k$  个时间段, 则  $\mathbf{x}_i$  是一个行向量, 即  $\mathbf{x}_i = [x_i(t_1), x_i(t_2), \dots, x_i(t_k)]$ ;  $y_j(t)$  表示水质监测点  $j$  在  $t$  时刻的模拟污染物浓度值;  $n$  表示候选污染源节点总数;  $m$  表示布置的水质监测点总数;  $t$  表示模拟时刻.

监测点模拟浓度可以表示为

$$\mathbf{Y} = G(\mathbf{X}, t) \quad (2)$$

结合监测点的实际监测结果, 污染源识别问题

的目标函数可以表示为<sup>[3]</sup>

$$\begin{aligned} f(\mathbf{X}^*) &= \min_{\mathbf{X}} \left\{ \sum_{i=1}^m \sum_{t=0}^T [y_i(t) - y'_i(t)]^2 \right\} \\ \text{s. t. } & y_i(t) = G(\mathbf{X}, t) \end{aligned} \quad (3)$$

式中:  $\mathbf{X}^*$  为污染源识别问题的最优解,  $y'_i(t)$  为水质监测点  $i$  在  $t$  时刻的实际浓度值. 为了便于求解, 这里需作几点假设: ①污染物的流量相对于管网水量而言可忽略不计, 不影响管网原来的水力工况; ②污染物在管网节点处投加, 投加点可以是多个, 投加量可以随时间变化, 但在每一水质时间步长之内是恒定不变的; ③污染物侵入管网中每个节点的概率是相同的, 也就是说污染物是随机侵入的; ④管网中的水质监测点能够监测到任意浓度的污染物.

## 2 模型求解

由于式(3)表示的最优化问题目标函数和约束条件均为非线性函数, 且约束条件是由供水管网恒定流方程、连续性方程、管道水质反应动力学方程共同决定的复杂非线性等式约束, 因此该问题为高度复杂的非线性规划问题. 考虑到约束条件可采用成熟的 EPANET 水力水质分析模块进行隐式求解, 故基于负梯度法, 通过差分法求目标函数梯度向量对此非线性规划问题进行求解. 求解过程如下:

**步骤1** 给定初始近似点  $\mathbf{X}^{(0)}$  及精度  $\epsilon > 0$ , 若  $\|\nabla f(\mathbf{X}^{(0)})\|^2 \leq \epsilon$ , 则  $\mathbf{X}^{(0)}$  即为近似极小点. 需要说明的是梯度向量的每一个元素均为目标函数对  $\mathbf{X} = [x_1, x_2, \dots, x_n]$  相应元素的偏导数. 该偏导数通过差分法近似求解

$$\frac{\partial f(\mathbf{X}^{(k)})}{\partial x_i} = \frac{f(\mathbf{X}^{(k)} + \Delta \mathbf{X}) - f(\mathbf{X}^{(k)})}{\delta} \quad (4)$$

式中:  $\Delta \mathbf{X}$  是第  $i$  个元素为  $\delta$  且其余元素为 0 的列向量;  $\delta$  为差分步长, 其取值大小应该合理确定;  $f$  即目标函数(见式(3)). 当水质反应为线性关系时, 水质监测点所测得的污染物浓度值可以表达成污染物侵入节点的污染物浓度加权值<sup>[3]</sup>, 即

$$y_j(t) = \sum_i a_{j,i} x_i \quad (5)$$

偏导数的求法可以简化为

$$\frac{\partial f(\mathbf{X}^{(k)})}{\partial x_i} = 2 \sum_j \sum_t a_{j,i} [y_j(t) - y'_j(t)]^k \quad (6)$$

式中:  $a_{j,i}$  为水质监测点  $j$  对候选污染源节点  $i$  的权重系数, 即当候选污染源节点  $i$  有一个单位的增量时, 在水质监测点  $j$  有相应的污染物浓度增量. 权重系数可以通过 EPANET 水质分析中的源头追踪功

能得到。

**步骤 2** 若  $\|\nabla f(X^{(0)})\|^2 > \epsilon$ , 求步长  $\lambda_0$ , 并计算

$$X^{(1)} = X^{(0)} - \lambda_0 \nabla f(X^{(0)}) \quad (7)$$

求最佳步长本文采用黄金分割法, 最佳步长即为使得目标函数最小的步长。若  $\|\nabla f(X^{(1)})\|^2 \leq \epsilon$ , 则输出  $X^{(1)}$ , 否则转入步骤 3。

**步骤 3** 一般地说, 若  $\|\nabla f(X^{(k)})\|^2 \leq \epsilon$ , 则  $X^{(k)}$  即为所求的近似解; 若  $\|\nabla f(X^{(k)})\|^2 > \epsilon$ , 则求步长  $\lambda_k$ , 并确定下一个近似点

$$X^{(k+1)} = X^{(k)} - \lambda_k \nabla f(X^{(k)}) \quad (8)$$

如此迭代下去, 直至达到所设定精度要求为止。

### 3 算例研究

以供水管网实例 Net3 为例进行污染源的识别问题求解。Net3 管网模型含节点数 92 个, 水库 2 座, 网中水塔 3 个, 管段数 117 个, 水泵 2 台, 如图 1 所示。图中虚线范围内的节点为候选污染源节点。

假定在节点 101 和 119 同时注入污染物, 注入方式为设置点注入<sup>[8]</sup>, 注入时间段为 3:00 am~

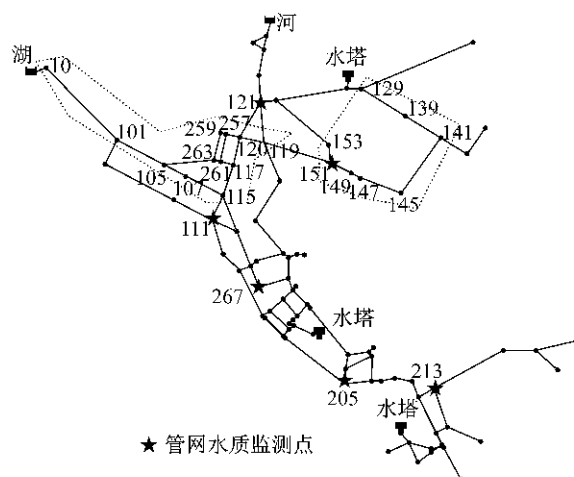


图 1 管网拓扑结构图

Fig.1 Topology layout of the case study network

10:00 am, 污染物注入质量浓度变化模式见图 2。管网水质模拟时间总长 24 h, 水力时间步长 1 h, 水质时间步长 5 min。在管网中随机选取 6 个节点作为水质监测点(分别为连接节点 121, 151, 111, 267, 205, 213), 如图 1 所示。通过水质模拟, 以这些节点模拟污染物质量浓度数据作为监测结果数据(见表 1)。

表 1 监测点模拟污染物质量浓度

Tab.1 Simulated concentrations at monitoring sites

mg · L<sup>-1</sup>

时间	监测点 121	监测点 151	监测点 111	监测点 267	监测点 205	监测点 213
04:00 am	0.00	1.01	98.73	97.73	72.82	0.00
05:00 am	0.00	26.26	100.00	100.00	100.00	98.53
06:00 am	0.00	51.40	100.00	100.00	100.00	95.10
07:00 am	0.00	68.37	100.00	100.00	100.00	98.77
08:00 am	0.00	38.76	99.03	98.29	98.81	99.12
09:00 am	0.00	27.08	99.63	99.34	99.57	98.97
10:00 am	0.00	33.71	100.00	100.00	100.00	98.75
11:00 am	0.00	16.41	100.00	100.00	100.00	94.80
12:00 am	0.00	0.00	0.00	0.00	10.24	84.44
13:00 pm	0.00	0.00	0.00	0.00	0.00	0.02

#### 3.1 确定候选节点集

候选节点集是指根据水质监测点对污染物的监测情况所确定的可能是管网污染事故污染源发生位置的节点集合。目前关于管网突发污染事故污染源追踪的研究, 仍然以通过经验法确定候选节点集合为主<sup>[3]</sup>, 候选节点集合确定的范围和准确程度都会影响到污染源识别问题最优化算法的性能。为此, 本

$$S_{\text{candidate}} = \{N_j \mid j = 10, 20, 60, 61, 101, 103, 105, 107, 109, 111, 115, 117, 119, 120, 121, 123, 125, 127, 129, 139, 141, 145, 147, 149, 151, 153, 257, 259, 261, 263\}$$

#### 3.2 结果分析

考虑到按照负梯度方向搜索得到的是局部最优解, 且绝大多数候选污染源节点均不是污染源注入点, 可将初始解向量的每个元素均设置为 0。求解过

研究根据文献[5,9]提出的最小覆盖集方法进行候选节点集合的确定。首先求出节点的监测时间矩阵, 以 24 h 为它们的服务水平, 可以确定出每个监测点所覆盖的范围。通过表 1 可知, 本次污染事故假定模式下, 实际上仅水质监测点 151, 111, 267, 205, 213 能监测到污染物。因此可以把 5 个监测点共同覆盖的节点作为候选节点集(如图 1 所示), 即

程中, 差分步长  $\delta$  取 0.010 0, 最高迭代次数取 360。图 2 中显示采用差分法计算梯度向量求解得到的结果, 由于求解的多数候选点的质量浓度值均很小, 图中仅显示 30 个候选污染源节点中所求解的注入质

量浓度较大的4个节点的注入质量浓度随时间变化情况。候选污染源节点101是各时刻累计投加质量浓度值最大的节点,初步选定为污染源节点。同时,候选污染源节点10从3:00 am到10:00 am基本上以 $100 \text{ mg} \cdot \text{L}^{-1}$ 的质量浓度持续注入。图1表明,节点10和节点101是同一个管段的起止节点。该现象表明有一个污染源在以节点10和节点101为起止节点的管道中。事实上,候选污染源节点101的注入质量浓度变化曲线基本与实际污染源节点101的质量浓度变化曲线相吻合。此外,候选污染源节点119的投加质量浓度呈抛物线型变化也与实际的污染源119投加质量浓度变化相吻合,可以确定这里也是一个污染源注入点。另外,候选污染源节点151作为污染源节点119的邻近节点,也有一定质量浓度的污染物注入,这就更表明了另一个污染源就是节点119或者在其附近。

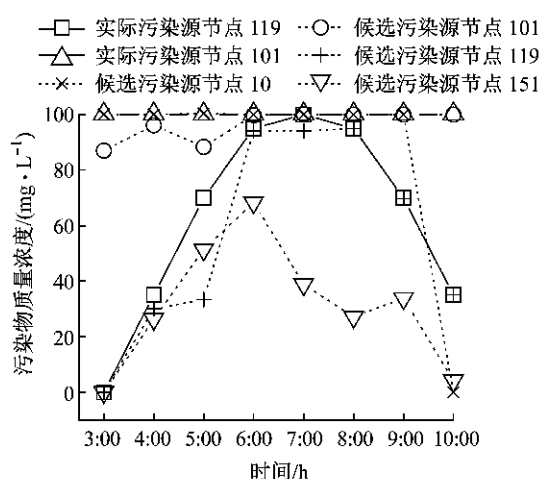


图2 污染物注入质量浓度变化曲线(差分法)

Fig.2 Injection concentrations of contamination source (finite difference method)

### 3.3 影响求解结果的因素分析

#### (1) 差分步长

对于图1所显示的小型算例管网,将迭代次数均设定为360次,差分步长 $\delta$ 依次取值为0.100 0, 0.010 0, 0.001 0和0.000 1时,计算结果的梯度向量模长值依次为16.52, 0.24, 0.15和0.33。为了定量地衡量计算结果与实际结果的拟合程度,计算污染源节点101和119各个时段假定注入浓度和解出注入浓度差的平方和总和,依次为:1 564.09, 1 914.74, 7 703.17, 64 201.58,如表2所示。

由表2可知,差分步长取0.010 0时对应的计算结果(梯度向量模长为0.24,目标函数值为1 914.74,而梯度向量模长和目标函数值是反映梯

度法求解质量的两个主要指标)最好,此时差分步长取值正好与EPANET水质选项中设定的水质公差值相同。其原因在于水质公差表示管网水质的最小变化值,也即一部分水体与另一部分水体水质的最小差值<sup>[9]</sup>,这个值也表征了管网中水质监测点的污染物最低监测限度,因此当差分步长取水质公差值时,计算结果最好。若差分步长继续减小,由于梯度计算将逐步超出管网模型的计算精度范围,计算结果反而变差。

表2 差分步长求解结果对比

Tab.2 Comparison of results with different  $\delta$ s

序号	迭代次数	差分步长 $\delta$	梯度向量模长	目标函数值
1	360	0.100 0	16.52	1 564.09
2	360	0.010 0	0.24	1 914.74
3	360	0.001 0	0.15	7 703.17
4	360	0.000 1	0.33	64 201.58

#### (2) 梯度计算方法

针对本文研究的算例管网,分别分析了通过差分法(见式4)以及线性比例法(见式6)进行梯度计算的求解效果。线性比例法计算后得到的结果如图3所示,其中迭代次数为400次,收敛时间为1 min,梯度向量的模长为3 230.83;通过差分法编码计算后得到的结果如图2所示,其中迭代次数为400次,收敛时间为5 min,梯度向量的模长为0.25。

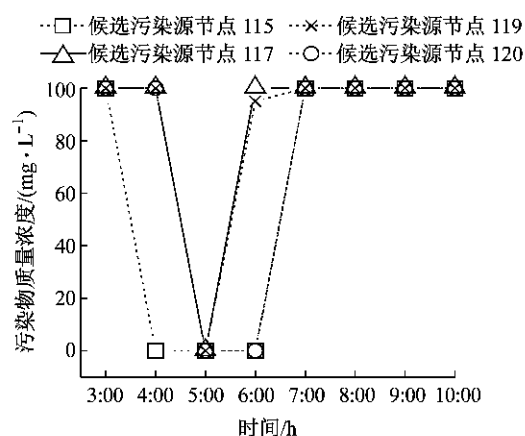


图3 污染物注入质量浓度变化曲线(线性比例法)

Fig.3 Injection concentrations of contamination sources with linearly proportional method

图3表明在候选污染源节点115,117,119和120处有污染物注入,所定位出的污染源位置在假定污染源注入点附近区域,但对于污染源节点污染物浓度的变化识别精度较差;而通过差分法计算(见图2)则能够直接确定污染源注入的时间、地点和侵入过程。此外,迭代次数相同时,采用线性比例法计算

结果的梯度向量模长远大于采用差分法计算的梯度向量模长. 因而采用线性比例法计算的结果要比用差分步长编码计算的结果差, 但用线性方式编码计算的收敛速度快于用差分法计算的速度. 在采用线性比例法进行计算时, 还需要为迭代变量设置一个可取值区间, 以保证解的非负性和所能取的最大值, 否则将会出现无法收敛的情况. 而用差分法计算时, 仅需要保证解的非负性, 无需设置解的最大限值. 分析其原因在于, 线性比例法计算时, 由于水质监测点对于某一候选污染源节点的权重系数, 是取相应时间步长时段内的权重系数平均值, 因此所求得的梯度向量并不能精确地给出目标函数的收敛方向, 导致计算结果较差. 因此, 建议在需要快速确定污染源点位时可采用线性比例法, 当需要更精确地分析污染物注入时间、浓度及变化模式时, 采用基于管网水质模拟的差分法进行计算.

## 4 结论

负梯度法可有效解决供水管网单点污染物注入模式和多点污染物注入模式下的污染源定位问题. 相对于遗传算法等随机搜索算法, 沿负梯度方向搜索具有更高的搜索效率, 通过引入差分算法可有效降低目标函数对于决策变量求导的难度. 而通过将利用最小覆盖集确定候选污染源节点集合的方法与梯度法相结合, 可进一步提高对污染源识别问题的求解能力. 此外, 通过不同差分步长下的计算结果表明, 当其值取水质模拟所采用的水质公差值时, 算法求解结果最好. 而相对于用线性比例法求解梯度向量, 差分法可获得更高的求解质量, 不但可以定位污染源, 并能以较高精度识别污染物注入的时间和浓度变化模式.

由于污染源识别问题是利用有限个水质监测点获得的数据反演污染源的侵入过程, 因此, 水质监测点的数量及空间布置对于问题的求解有直接的影

响. 因此, 在将来的研究中, 本方法应进一步与水质监测点的优化布置结合, 以更进一步提高算法的求解效率. 此外, 对于无法有效获取管网中水质的精确浓度数据的情况, 如何对突发污染事件进行污染源的辨识, 是未来需要进一步研究的问题.

## 参考文献:

- [1] Laird C D, Biegler L T, van Bloemen Waanders B G, et al. Contamination source determination for water networks[J]. *Water Resources Planning and Management*, 2005, 131(2): 125.
- [2] Laird C D, Biegler L T, van Bloemen Waanders B G. Mixed-integer approach for obtaining unique solutions in source inversion of water networks[J]. *Water Resources Planning and Management*, 2006, 132(4): 242.
- [3] GUAN Jiabao, Aral M M, Maslia M L, et al. Identification of contaminant sources in water distribution systems using simulation optimization method: case study [J]. *Water Resources Planning and Management*, 2006, 132(4): 252.
- [4] Pries A, Ostfeld A. Contamination source identification in water systems: a hybrid model trees-linear programming scheme[J]. *Water Resources Planning and Management*, 2006, 132(4): 263.
- [5] Cristo D C, Leopardi A. Pollution source identification of accidental contamination in water distribution networks[J]. *Water Resources Planning and Management*, 2008, 134(2): 197.
- [6] LIU Li, Ranji Ranjithan S, Mahinthakumar G. Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure[J]. *Water Resources Planning and Management*, 2011, 137(2): 183.
- [7] YANG Xueyao, Boccelli D L, De Sanctis A E. A Bayesian approach for probabilistic contamination source identification [C]//World Environmental and Water Resource Congress. [S. l.]: ASCE, 2011.
- [8] Rossman L A. EPANET 2 users' manual[M]. Cincinnati: U. S. Environmental Protection Agency, 2000.
- [9] 王凤仙. 供水管网水质管理软件应用和预警技术研究[D]. 上海: 同济大学环境科学与工程学院, 2010.  
WANG Fengxian. Study on application of water quality management software and early-warning technology for water distribution systems[D]. Shanghai: College of Environmental Science and Engineering of Tongji University, 2010.