

基于用户水质投诉信息的供水管网污染源追踪定位

信昆仑, 项宁银, 陶 涛, 尹兆龙

(同济大学 环境科学与工程学院, 上海 200092)

摘要: 突发性水污染事故发生突然, 污染物扩散迅速, 对供水管网水质安全危害巨大. 在缺乏足够水质监测设备的城市, 用户对管网的水质投诉信息可及时反映管网水质情况. 根据用户水质投诉的位置信息, 利用改进的污染矩阵法查找管网候选污染源节点, 再通过概率理论分析各候选污染源节点中是污染源的可能性, 最后采用波达法则和数据标准化法两种方法评价最具可能性污染源. 该方法的可行性和有效性通过某市管网实例进行了验证.

关键词: 供水管网; 污染源追踪定位; 概率分析; 用户投诉
中图分类号: TU991 **文献标志码:** A

Contamination Source Identification in Water Distribution Network Based on Consumer Complaints

XIN Kunlun, XIANG Ningyin, TAO Tao, YIN Zhaolong
(College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China)

Abstract: Accidental water pollution and rapid diffusion of the pollutant terribly affect water quality in the distribution network; the consumers complaint information to water utilities is an important way for monitoring the condition of the water supply systems timely where there is lack of monitoring station. The research first uses the location information of consumer complaints to find the candidate pollution source nodes through the improved pollution matrix methodology. And then the probability of each candidate pollution source node is calculated on the basis of the total probability formula and evaluated by Borda Count and Normalization Count to identify the most possible pollution source. The feasibility and effectiveness are verified through a source contamination event in a city.

Key words: water distribution system; contamination source identification; probability analysis; consumer complaints

供水管网突发性污染事故的污染源追踪定位技术是一种反演技术. 近年来, 国内外有关污染源定位问题研究较多, Laird 等^[1-2]于 2005 年提出采用非线性化的程序法(non-linear programming), 并于 2006 年对该方法进行了改进, 提出混合整数法(mixed-integer approach), 使该方法从非唯一的方案中选取最可能的注入情景. Cristo 等^[3]于 2008 年提出通过比较最小化模拟值和测量值之间的差值, 利用比例矩阵(water fraction matrix)转化成线性求解最优化问题来确定污染源. Liu 等^[4]于 2011 年探究了一种基于进化算法进行动态优化(evolution algorithm-based dynamic optimization)的方法对污染事件进行实时响应以识别污染源. 国内学者中, 王康乐^[5]利用关系树-线性规划算法对供水管网污染源进行追踪; 吕谋等^[6]通过试验实时监测污染信息, 利用改进的模拟-优化反追踪模型定位污染源.

上述方法均以供水管网中水质在线监测数据充足且准确为前提, 目前国内城市管网中的水质监测设备大多并不完善, 并且监测数据准确度较低. 在一次突发污染事故中没有足够的监测信息可供研究利用时, 用户投诉信息可作为反映管网水质状态的重要信息. 本文提出基于对用户投诉信息进行污染源追踪定位的方法, 将每个用户看作是一个监测点, 通过分析投诉用户的位置信息来确定污染源候选节点, 再根据用户投诉时间信息确定最具可能性污染点. 由于用户投诉信息无法准确反映该位置污染物浓度, 所以本文方法的主要目的是确定污染源实际发生位置, 控制污染继续扩散, 而对污染物投加浓度不予考虑.

1 方法介绍

本文将投诉用户当作是“监测点”, 利用已有的

管网水质模型,通过对某一节点污染后的管网水质变化情况进行动态模拟,通过投诉用户位置信息查找候选污染源节点,根据模拟污染物到达用户投诉节点的时间和用户实际投诉时间等,利用概率分析理论计算该点发生投诉的理论概率,进而根据对所有投诉点针对各候选污染点的投诉概率综合分析,确定其中最具可能性污染源节点。

1.1 候选污染源节点查找

根据 Kessler 等^[7]提出的污染矩阵(pollution matrix)概念来反映在一次污染事件中管网各节点的水质状态,Cristo 等^[3]找到一群候选污染源节点。一般情况下,污染矩阵是由 0 和 1 构成的 $N \times N$ 阶矩阵,其中 N 代表管网节点数,0 和 1 分别代表该节点未污染和已污染。由于一般管网都较大,管网节点数较多,完全采用该方法查找候选污染源节点计算负荷较大。本文对该方法进行改进,只提取用户投诉节点水质信息。由于只有当节点浓度达到一定阈值时才会被人体感知,所以设定当节点污染浓度超过一定感知限时才被认定为 1。污染矩阵法具体计算过程参见文献^[3,7],整个计算过程通过 Matlab 编程实现。

1.2 最具可能性的污染源确定

本研究认为节点污染后被投诉是一随机事件 A ,该随机事件建立在 3 种情况(每小时用水量不同、

用水时间不同、污染-投诉时间间隔不同)之上,每种情况影响 A 事件结果的权重程度记作 B ,则 A 事件的概率可用下式表示:

$$P_i(A) = B(1_i)P_{1i} + B(2_i)P_{2i} + B(3_i)P_{3i} \quad (1)$$

式中: P_i 为第 i 号节点发生投诉的概率; P_{1i} 为第 i 号节点基于每小时用水量不同的投诉概率; $B(1_i)$ 为每小时用水量不同影响投诉发生的权重; P_{2i} 为第 i 号节点基于污染发生时间不同的投诉概率; $B(2_i)$ 为污染发生时间不同影响投诉发生的权重; P_{3i} 为第 i 号节点基于时间间隔不同的投诉概率; $B(3_i)$ 为时间间隔不同影响投诉发生的权重。

假定 3 种情况对投诉发生影响相当,即 $B(1_i) = B(2_i) = B(3_i) = 1/3$, $P_i(A)$ 可用下式表示:

$$P_i(A) = \frac{1}{3}P_{1i} + \frac{1}{3}P_{2i} + \frac{1}{3}P_{3i} \quad (2)$$

按 Janke 等^[8]的研究,假设用户发现水质问题只在用水时, P_{1i} 可用下式表示:

$$P_{1i} = \alpha \frac{Q_{\text{node}}}{Q_{\text{all}}} \quad (3)$$

式中: α 为比例系数; Q_{node} 为节点污染时刻节点需水量; Q_{all} 为管网总需水量,以节点污染时刻水厂出水流量计。

根据 Barraj 等^[9]调查的用水模式,假定每个时间段的用水概率 P_2 如表 1 所示。

表 1 每小时用水概率

Tab.1 Probability of water consumption on every hour

时段	6:00— 7:00	7:00— 8:00	8:00— 9:00	9:00— 10:00	10:00— 11:00	11:00— 12:00	12:00— 13:00	13:00— 14:00	14:00— 15:00	15:00— 16:00
用水概率/%	6.2	5.3	5.7	5.5	5.7	5.9	7.2	5.3	5.5	5.4
时段	16:00— 17:00	17:00— 18:00	18:00— 19:00	19:00— 20:00	20:00— 21:00	21:00— 22:00	22:00— 23:00	23:00— 24:00	24:00— 6:00	
用水概率/%	5.0	5.8	6.0	5.4	5.3	5.2	4.0	2.4	3.2	

如上所述,污染物到达节点的时间不同将影响投诉的发生概率,本文采用模拟污染物到达投诉节点时间时的用水概率 P_{2i} 表征,即

$$P_{2i} = \beta P_2 \quad (4)$$

式中: β 为比例系数。

污染到达投诉节点到用户用水发现污染一定会存在时间滞后, T_{ao} 等^[10]假设用户在 2 h 内向供水企业投诉,其研究结果表明该假设具有可行性。基于该假设,本文认为用户发现污染后会在短时间内进行投诉,时间间隔越长投诉的概率越低。因此假设时间间隔 Δt 符合期望为零的正态分布的右半部分,即

$$P_{3i} = \gamma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\Delta t^2}{2\sigma^2}} \quad (5)$$

式中: $\Delta t > 0$; σ 是标准差; γ 是比例系数。为了明显区分不同时间间隔的投诉概率大小,本研究取 σ 为 10,其他取值对结果影响较小,不加详细讨论。方法流程图如图 1 所示。

2 实例研究

以某市一次突发性供水管网污染事件为例。该市拥有城区水厂 2 座,用户总数为 34.56 万户。该市已建立供水管网水力水质模型,模型节点个数为 3 379,此次污染事故中,第 1 天共接到投诉电话 20 起。该市管网拓扑图及投诉用户分布情况如图 2 所示,投诉时间如表 2 所示。

表 2 用户投诉时间

Tab.2 Time of consumer complaints

投诉点号	1	2	3	4	5	6	7	8	9	10
投诉时间	09:51	10:15	10:44	10:57	11:18	11:20	11:50	11:57	12:26	12:42
投诉点号	11	12	13	14	15	16	17	18	19	20
投诉时间	13:20	13:45	14:13	16:46	17:29	18:24	18:45	19:43	19:47	22:20

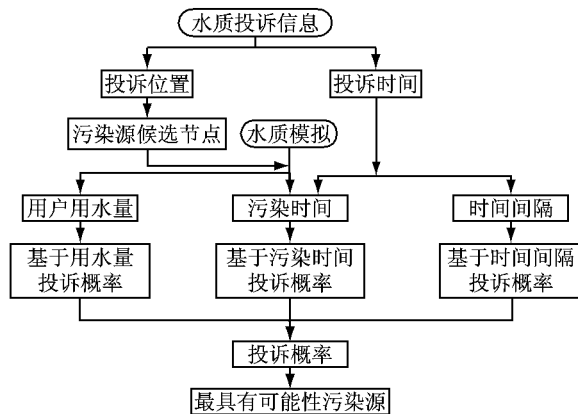


图 1 方法流程图

Fig.1 Flow chart of the method

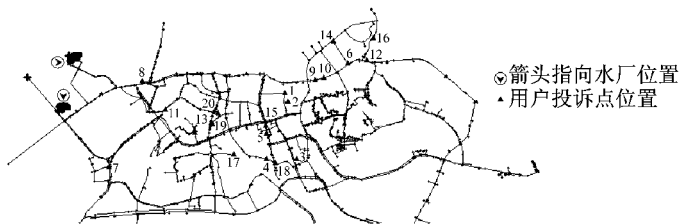


图 2 某市管网拓扑图

Fig.2 The water distribution system of a city

Janke^[8]的研究表明,当污染物总数一定时,污染物释放时间长短对居民危害影响差异性较小; Davis 等^[11]发现注入时间越长,越能减少用水模式不同对污染物发现的影响,所以采用

EPANET 2.0 模拟在管网中进行持续 24 h 等质量浓度注入. 其他参数设置如下:水力步长为 15 min;水质步长为 5 min;报告时间步长为 1 h;模拟起始时间为 0:00;污染物注入质量浓度为 $500 \text{ mg} \cdot \text{L}^{-1}$.

由于本阶段研究主要目的在于提出并验证方法的有效性,取全部 20 个投诉信息进行分析. 根据计算得到的 4 个候选污染源节点号为 99,100,101,3279,各候选污染源节点的位置如图 3 所示.

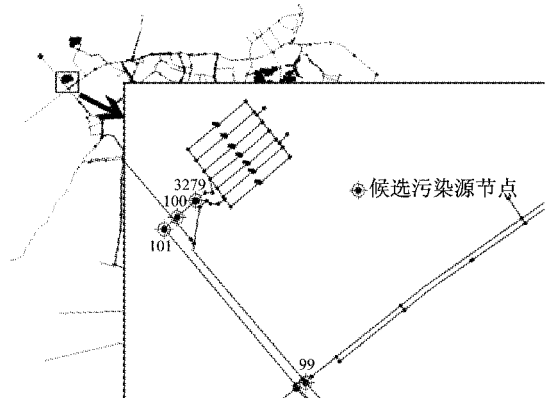


图 3 候选污染源节点位置分布

Fig.3 Location of candidate source nodes

根据模拟结果,4 个候选污染源节点在 00:00 时被注入污染物时,污染物到达投诉点所需时间如表 3 所示.

表 3 污染物到达投诉点所需时间

Tab.3 Time for pollutants arrival to consumer complaint nodes

投诉点号	污染物到达投诉点的模拟时间/h				$\Delta t/\text{h}$	投诉点号	污染物到达投诉点的模拟时间/h				$\Delta t/\text{h}$
	99	100	101	3279			99	100	101	3279	
1	11.26	11.75	11.50	12.00	9.85	11	17.00	17.50	17.25	17.75	13.33
2	19.00	19.75	19.50	20.25	10.25	12	16.50	17.00	16.75	17.25	13.75
3	12.00	12.50	12.25	12.75	10.73	13	4.75	5.25	5.00	5.50	14.22
4	18.75	19.25	19.00	19.50	10.95	14	2.00	2.50	2.25	2.75	16.77
5	23.00	23.75	23.25	24.00	11.30	15	18.50	19.00	18.75	19.25	17.48
6	13.00	13.50	13.25	13.75	11.33	16	15.75	16.25	16.00	16.50	18.40
7	18.50	19.00	18.75	19.25	11.83	17	11.50	12.00	11.75	12.25	18.75
8	12.00	12.50	12.25	12.75	11.95	18	23.75	24.25	24.00	24.50	19.72
9	20.50	21.00	20.75	21.25	12.43	19	20.50	21.00	20.75	21.25	19.78
10	7.75	8.25	8.00	8.50	12.70	20	19.00	19.50	19.25	19.75	22.33

注: Δt 为投诉距污染发生的真实时间.

数据表明,当 0:00 注入污染物时,有 14 个投诉点的投诉时间(表 3 中粗体表示)明显早于污染发生

时间,说明污染发生并非 0:00. 当满足所有投诉点时间要求时,污染注入时间必须在投诉发生前一天上

午 11:00 或之前,以尽快发现污染物为原则,采用以 11:00 为污染物注入时间.由于缺乏实际数据支持,本研究比例系数 α, β, γ 的取值以保证 P_{1i}, P_{2i}, P_{3i} 属于 $[0, 1]$,且接近 1 为原则.

对于 P_{1i} ,计算 $\frac{Q_{\text{node}}}{Q_{\text{all}}} \in [0.01\%, 0.12\%]$,所以取 $\alpha=800$,因此 $P_{1i} \in [0.08, 1]$.

对于 $P_{2i}, P_{2i} \in [2.0\%, 7.2\%]$,取 $\beta=13.8$,因此 $P_{2i} \in [0.27, 1]$.

对于 P_{3i} ,当 σ 为 10, Δt 为 0 时, $\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\Delta t^2}{2\sigma^2}}$ 的值为 0.04,取 $\gamma=25$,使 $P_{3i} \in (0, 1]$.

再根据式(2)计算投诉概率 $P_i(A)$ 如表 4 所示.

表 4 用户投诉概率

Tab.4 Probability of complaints

投诉点号	投诉概率/%				投诉点号	投诉概率/%			
	99	100	101	3 279		99	100	101	3 279
1	38.07	38.98	38.48	39.48	11	60.22	58.55	58.05	59.05
2	69.52	70.44	70.19	66.80	12	66.39	67.39	66.89	70.55
3	56.48	50.04	49.54	50.54	13	32.84	28.67	33.01	28.83
4	54.39	71.60	68.60	71.94	14	74.79	66.55	66.30	66.80
5	72.55	75.39	75.30	75.47	15	41.64	56.35	42.14	56.85
6	59.35	54.55	54.05	55.05	16	56.47	57.30	56.89	57.64
7	52.64	67.27	53.05	67.60	17	30.65	31.15	30.90	24.04
8	62.07	55.71	55.21	56.21	18	47.92	53.06	48.42	53.56
9	67.13	67.88	67.55	70.05	19	53.38	54.30	53.80	59.30
10	52.93	50.76	53.18	51.09	20	53.27	56.69	56.35	57.10

为了直观地比较不同候选污染源节点发生污染情况下的投诉概率,采用波达法则(Borda Count)进行评价.参照波达法将每个投诉点在不同候选污染源节点发生污染情况下的投诉概率进行排序,再将排序的名次转换为分数.将每组投诉点概率进行 Z-score 标准化,对小于 0(即小于均值)的候选污染源节点认定为 0,其他为 1.将 20 组投诉概率分别用两种方法进行打分,最终得分情况如表 5 所示.

表 5 波达计算法和 Z-score 标准化法得分情况

Tab.5 Score based on Borda Count and Normalization Count

候选污染源 节点号	得分	
	波达算法	Z-score 标准化法
99	41	9
100	53	10
101	41	7
3 279	65	12

两种方法均表明 3 279 号候选污染源节点受到污染时,被用户投诉的概率最大,即该点是最有可能的污染源节点.根据该市管网地图可知 3 279 号节点是该市其中之一水厂清水池,所以该污染事故应为水源污染事故,这与实际情况完全吻合.另外,从图 2 可知 100 号候选点与 3 279 号节点位置最为接近,其概率也较其他两个候选点更高.

3 结论

研究在缺乏水质监测数据下利用用户投诉信息

进行污染源定位的方法.首先根据投诉用户的位置信息利用改进污染矩阵法查找候选污染源节点,模拟结果表明该法可从一个节点数大于 3 000 的管网中查找到 4 个候选污染源节点;其次根据污染到达用户的时间确定污染发生的最晚时间.通过概率理论计算各候选污染源节点注入污染物时用户投诉的概率,采用波达法则和数据标准化法两种方法确定最具可能性的污染源,两种方法的结果一致指出了污染源节点.研究表明,利用用户投诉信息可有效地进行污染源定位,该技术有助于在现状管网水质在线监测设备普遍缺乏的情况下应对管网水质突发事件.

为了确保该方法的可行性,研究方法中各拟定的参数数值需要进一步确认;另外由于污染发生到用户投诉的时间间隔存在很大的主观性和随机性,需要进一步研究其发生规律以供后续研究利用.

参考文献:

- [1] Laird C D, Biegler L T, Waanders B, et al. Contamination source determination for water networks[J]. Journal of Water Resources Planning and Management, 2005, 131(2): 125.
- [2] Laird C D, Biegler L T, Waanders B. Mixed-integer approach for obtaining unique solutions in source inversion of drinking water networks[J]. Journal of Water Resources Planning and Management, 2006, 132(4): 242.

(下转第 291 页)