

Research on contaminant sources identification of uncertainty water demand using genetic algorithm

Yan Xuesong¹ · Sun Jie¹ · Hu Chengyu¹

Received: 8 November 2016 / Revised: 4 February 2017 / Accepted: 8 February 2017 / Published online: 21 February 2017
© Springer Science+Business Media New York 2017

Abstract Urban water supply network is easily affected by intentional or occasional chemical and biological pollution, which threatens the health of consumers. In recent years, drinking water contamination happens occasionally, which seriously harms social stabilization and safety. Placing sensors in water supply pipes can monitor water quality in real time, which may prevent contamination accidents. However, how to reversely locate pollution sources through the detecting information from water quality sensors is a challengeable issue. Its difficulties lie in that limited sensors, massive pipe network nodes and dynamic water demand of users lead to the uncertainty, large-scale and dynamism of this optimization problem. This paper mainly studies the uncertainty problem in contaminant sources identification (CSI). The previous study of CSI supposes that hydraulic output (e.g., water demand) is known. Whereas, the inherent variability of urban water consumption brings an uncertain problem that water demand presents volatility. In this paper, the water demand of water supply network nodes simulated by Gaussian model is stochastic, and then being used to solve the problem of CSI, simulation–optimization method finds the minimum target of CSI and concentration which meet the simulation value and detected value of sensors. This paper proposes an improved genetic algorithm to solve the CSI problem under uncertainty water demand and comparative experiments are placed on two water distribution networks of different sizes.

Keywords Contaminant sources identification · Uncertainty water demand · Genetic algorithm

✉ Hu Chengyu
huchengyu@cug.edu.cn

¹ School of Computer Science, China University of Geosciences, Wuhan 430074, Hubei, China

1 Introduction

In recent years, unexpected water contamination accidents occasionally happen in China. Some unexpected drinking water contamination accidents and hostile attacks to water supply pipe network cause serious economic losses and baneful social influence. In order to prevent the losses and catastrophes caused by serious water contamination accidents, it is necessary for water supply pipe network of cities and towns to install real-time monitoring systems which maintain drinking water security. In this system, real-time supervision can be achieved through placing water quality sensors in key nodes or water sources [1–6]. However, when contamination accidents happen, how to collect information by water-quality sensors, how to analyze the characteristics of pollution sources and how to forecast pollution locations, injection time, injection duration and injection quality are challengeable.

Particle backward tracing method, reversely deducing the state of every node according to the information detected by sensors, is mainly used in the early study of contaminant sources identification. For example, Shang proposes particle backward method. It regards pollutants as a particle which can reversely deduce pollution sources from the nodes with detected pollution sources according to chronological order within inverse time [7]. Laird uses source-tracking algorithm to locate multiple pollution sources and simulation experiments shows that this method is effective in small-scale pollution sources identification [8]. Particle backward algorithm adopted by De Sanctis can immediately identify possible pollution sources through analyzing the consistency between hydraulic power and water quality. This method is effective in localization problem; however, false and missed locations of sensors are ignored [9]. Costa continuously collects information from sensors and excludes unmatched nodes to sift

possible pollution sources by forward and reverse comparison. Additionally, author discusses the influence of sensor misinformation on algorithm performance [10].

Machine learning algorithm, widely adopted by a majority of experts, can sort the pollution probability of water supply pipe network, which helps to find the node with maximum pollution probability. In reference [11], though pollution sources can be quickly found with the methods of data mining and maximum probability, only approximate range of pollution sources can be found under the circumstance of multiple pollution sources. From the perspective of Perelman, water supply pipe network is divided into different clusters by flow direction and connectivity. The probability calculation is based on Bayesian Rules. The disadvantage of this reference is that only the cluster where pollution source is located can be found and the node in clusters cannot be localized [12]. In 2013, Wang proposed Markov Chain Monte Carlo (MCMC) method to collect samples from stochastic data. Then, Bayesian Rules is adopted to calculate the probability of becoming pollution sources in every node [13]. The combination of MCMC and support vector regression proposed in 2014 greatly shortens computation time [14]. As a forecasting method based on probability, Machine learning can not only dynamically collect samples through MCMC, but also forecast the probability of every node where pollutions are injected. However, its forecast precision should be further improved. Meanwhile, the complexity of calculating node probability will sharply increase because of the growth of pipe network size.

A hypothetical CSI problem that pollution can be injected in any pollution sources is constructed through pollution concentration accumulated by the node of analog sensors. If the information of pollution sources in a forward simulation model is known, the accumulated pollution concentration of sensor can be concluded through simulation. CSI, a backward problem of this process, is defined as an optimization problem whose goal is to find a pollution source which minimize the difference between accumulated simulated concentration and practical accumulated detected concentration of sensor nodes. Simulation and optimization method solves problems by optimization algorithm which drives simulators. In 2006, Guan proposed simulation–optimization method to solve non-linear CSI problem. Forecasts are optimized and pollution sources are revised by the continuous reading of sensor data. Finally, pollution sources and the history of pollutants release are identified [15]. Liu proposes a self-adapting dynamic-optimization method based on evolution algorithm to search the characteristics of pollution sources (starting time, location and release history). The characteristics of pollution sources gradually converge until the only optimization solution is obtained when new workable sensors are constantly added in [16]. Hu et al. think CSI problem is characterized by its extremely high computation com-

plexity, uncertainty and non-uniqueness of the solution in a large-scale water distribution network with dynamic water demands. To tackle this issue, they develop a Map-Reduce based Parallel Niche Genetic Algorithm (MR-PNGA) that is not only able to achieve high identification accuracy but also to explore the cloud resources for performance improvement [17]. Yan et. al., formulated the CSI problem into an optimization problem, and used hybrid encoding method to code the problem according to the properties of a variable, so as to improve the convergence speed and accuracy. They also used different size of pipe network data in experiment, which results finally verified the validity and robustness of the proposed method [18].

At present, most of the study about this problem is based on the presupposition that the water demand of users is known. It is also the input of model. However, in real world, the uncertainty water demand of users leads to the uncertain input of model. In order to improve the precision of model, the water demand as unknown input is should be firstly defined when the model is established. In this paper, the water demand of users simulated by Gaussian model is treated as the input of simulation–optimization model and an improved genetic algorithm is used as optimization algorithm to solve CSI problem under uncertain water demand. The effectiveness of algorithm is verified by simulation experiments in the end.

2 CSI problem

2.1 The model of CSI

CSI problem is to find the location of pollution sources and pollution injection history of the time. Pollution injection history includes starting time, duration and the quality of injected pollution. In the view of optimization, when the minimum variance between accumulated simulated-concentration of contamination accidents and practical accumulated detected-concentration is zero or less than a certain threshold value ϵ , this contamination accident is a real one. In previous references, if the data of water demand is known, the model of CSI is formulated as formula (1). However, the data of water demand in this paper are unknown and we suppose the starting time and duration of pollutants injection are known. This means the CSI problem is simplified as finding the location of pollution sources and the accumulated concentration of injection pollution. So the model of CSI is formulated as formula (2). In formula (2) uses the max-min principle and it is proved that max–min principle is better than least square error principle used in formula (1).

$$\begin{aligned} & \text{Find}\{L, M_{t_c}, T_0\} \\ & \text{minimize } F = \sqrt{\frac{\sum_{t=t_0}^{t_c} \sum_{i=1}^{N_S} [C_{it}^{obs} - C_{it}(L, M_{t_c}, T_0)]^2}{N_S * t_c}} \end{aligned} \quad (1)$$

In formula (1), F indicates for forecasting error; L indicates for the nodes of pollution source position; T_0 indicates for the starting time of pollution injecting; t_0 indicates for the time when sensor firstly detects pollution; t_c indicates for the time step of the time; M_{t_c} is the vector of pollution injecting quality, indicating for the situation of pollution injecting quality from time T_0 to t_c ; i indicates for the position node placed by analog sensor; t indicates for observing time step; N_s indicates for the total number of sensors, $M_{t_c} = (m_{T_0}; m_{T_0+1}; \dots; m_{t_c})$; C_{it}^{obs} indicates for the concentration detected at sensor i and time step t ; $C_{it}(L, M_{t_c}, T_0)$ indicates for the concentration simulated by analog sensor at sensor i and time step t .

$$\text{Find}\{L, C_0\}$$

$$\text{minimize } F = \max_{k=1, \dots, N_s} \left\{ \sum_{t=1}^T (C_k^{obs}(t) - C_k^*(L, C_0, t))^2 \right\} \quad (2)$$

In formula (2), F indicates for forecast error; L indicates for the position node of pollution source; C_0 indicates for the concentration of pollutants source within known pollution-injecting time; t indicates for the time step of the time; T the total number of sample time steps; k indicates for the position of a certain sensor; N_s indicates for the total number of sensors; C_k^{obs} indicates for the practical accumulated detected concentration at sensor k ; $C_k^*(L, C_0)$ indicates for the simulated accumulated concentration at sensor k ; Notice: the uncertainty of water demand is in the calculation of C_k^* not C_k^{obs} , the threshold value is empirical value.

2.2 Simulation–optimization model

In the solving of CSI problem by simulation–optimization model, EPANET simulator is used to simulate the pollution data in sensors. The main job of EPANET simulator is to forward input the information of water quality which is compared with the information of water quality practically detected by sensors. If the error value is smaller than the given value ε , the solution conducted from optimization algorithm is an effective value, which means the pollution sources location and concentration have been found. Or else, optimization algorithm will continue to work until a specified condition is found, and then the program will come to an end. The framework of simulation–optimization model is shown in Fig. 1, EPANET simulator has three input conditions: (1) the known input composed by pipe network information and hydraulic condition; (2) unknown water demand data which can be collected through water demand model, (3) unknown pollution source characteristics. Optimization algorithm is used to reduce the errors in accumulated detected data and accumulated simulated data in every iteration process.

In optimization algorithm, every individual in species stands for a contamination accident which can be simulated by EPANET simulator. EPANET simulator can also input the pollutants accumulated concentration of sensor nodes and the adaption degree of every individual can be calculated through the comparison between the pollutants accumulated concentration of sensor nodes and practical accumulated detected concentration of sensors. It is concluded that the smaller the adaptive value is, the higher occurring possibility the contamination accident has.

3 Water demand model

The water demand of urban residents follows a certain rule. As it is mentioned in the articles of Praveen et al. [19], in order to find a residential water-demand model which is time-variable and stochastic, Buchberger and Wells are applied to analyze the water demand information from about 21 families in Milford within 30 days in 1996. A rule of residential water demand is concluded from the observed data. In this paper, through the experimental measurement of the water-consumption data from Milford residents, the weighting factors of water consumption of every hour can be calculated from the total water consumption of every day. The total water consumption of every hour can be calculated with the help of weighting factors and given water-demand data of pipe network of every day. Then, the total water demand of every hour will be distributed to every node by the weighting factors of every node. The average water consumption of every hour is shown in Fig. 2. We can conclude that though the residential water demand is uncertain, it still follows a certain rule. Therefore, we adopt Gauss model to simulate the rule of water demand in this paper.

Gauss model, a complete dissociation model added with simple white noise interference, is applied in the situation without prior probability. In Gauss model, the water demand of every time step in every node is added with a Gauss disturbance, which is shown in formula (3):

$$d_i^j \rightarrow d_i^j [1 + \sigma \cdot N(0, 1)] \quad (3)$$

In formula (3), d_i^j indicates for the water demand in pipe node i at time step j ; σ , a constant, indicates for the percentage of changing requirements; in this essay, we suppose that the value of σ is 0.25, $N(0, 1)$ indicates for a standard normal distribution whose mean value is 0 and variance is 1; the basic water-demand in every node produces a random water-demand set through the calculation of formula (3).

If Gauss model is used to simulate random water demand as shown in Fig. 3, an uncertain water demand needed by experimental pipe network is gained. As the complete dissociation of Gauss model, a certain difference and randomness

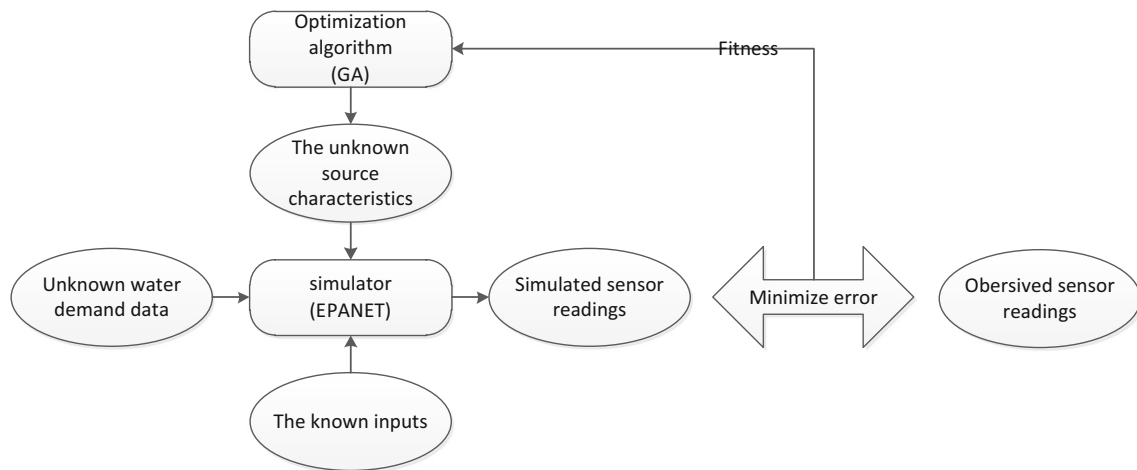


Fig. 1 Framework of simulation–optimization

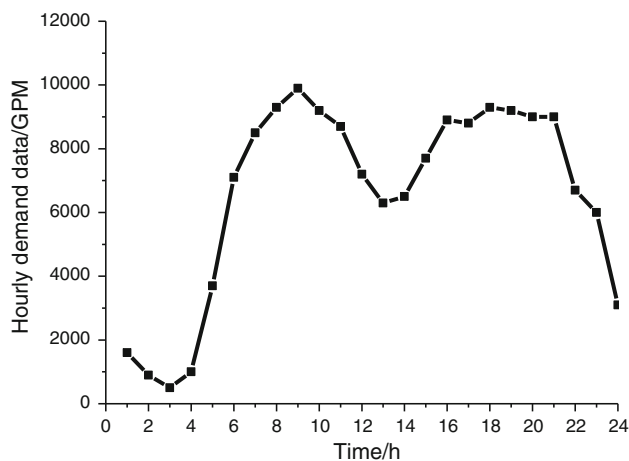


Fig. 2 Mean hourly demand of Milford

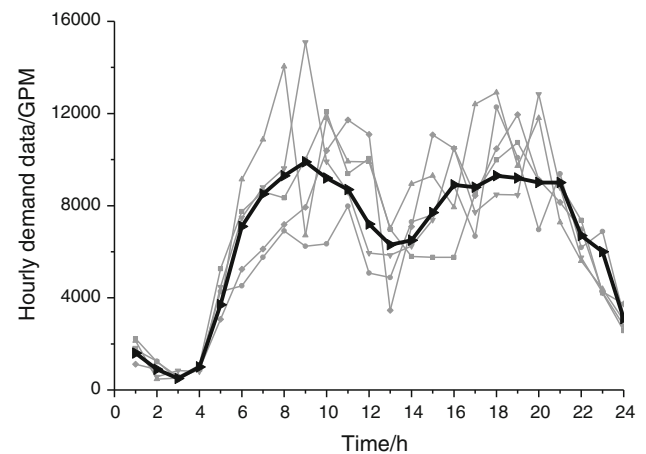


Fig. 3 Demand of Gaussian model

exist between the water demand curve produced by Gauss model and basic water demand curve produced by statistic data.

4 Solving CSI problem based on improved genetic algorithm

Genetic algorithm is considered to be an intelligence heuristic search algorithm which originates from nature. This intelligent optimization algorithm was introduced by Holland [20] in 1975 and next his student Goldberg [21] made a complete description about simple genetic algorithm in the book “Genetic algorithms in search, optimization, and machine learning”. In the book, it involves all important parameters discussion of genetic algorithm which includes crossover operator, mutation operator and fitness calculation, etc. And the genetic algorithm has been used in the

real life. The main characteristic of genetic algorithm is that swarm search strategy and information exchange and search among individuals in swarm don’t depend on gradient information. As a representative of heuristic probability search, the genetic algorithm is concise, adaptable, and easy to parallel programming design. It is widely applied in much areas such as bioscience, data mining, mechanical control, intelligent optimization, etc. And the genetic algorithm is an extremely important keyword about intelligent computing in this century. In the genetic algorithm, it is a practical matter about the components of genetic chromosome which need encoding firstly and replace the original mathematical expression. Therefore, the algorithm can be easily settled no matter it is discrete variable or continuous variable. The genetic algorithm dominates the algorithm searching process without depending on other auxiliary external information so that it doesn’t require to design function or understand spatial smoothness or continuity in order to easily handle many practical mathematical models; on the other hand,

evolution genetic algorithm has a lot of different research directions in scientific research [22–29]. The evolution genetic algorithm can effectively improve algorithm parameter and searching strategy and fuse other heuristic algorithms which then compose of hybrid genetic algorithm.

4.1 Chromosome coding

In improved genetic algorithm, every individual, standing for a contamination accident, covers two variables: node position and the quality of injected pollutant (the above paper supposes that the starting time and duration of injected pollutants are known). The node position, an integer variable, adopts integer coding; the quality of injected pollutants, a real number, adopts real coding. For example, a polluted scene is illustrated: 300mg pollutants are continuously injected into position 374, the chromosome of polluted scene is shown as {374, 300.0f}. (Note: continuous injection and instantaneous injection have no differential influence on the code of chromosome, only affecting the pollution sources injection method in pipe network); time multiplier corresponds to the method of pollutants injection. If the injection time is one hour, time multiplier is required: {1.0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}; if the injection is instantaneous, time multiplier is zero.

4.2 Crossover operator

Hybrid cross method is adopted in this paper. Decimalize uniform crossover is for position, and arithmetic crossover method is for concentration. The crossover probability is 0.8. New individuals produced by crossover are compared with parent individuals. The individuals with low adaptive degree will be saved.

If the individual 1 is: {259, 672.7663}, the individual 2 is: {126, 550.8366}. It is assumed that the mask work of stochastic equal length of decimal cross is 011 (0, 1 are character strings, crossover probability is 0.8); if the arithmetic divisors of arithmetic crossover are 0.2 and 0.6, the two new individuals after crossover are: individual 3: {226, $672.7663 \times 0.2 + 550.8366 \times 0.6$ } and individual 4: {159, $672.7663 \times 0.6 + 550.8366 \times 0.2$ }.

4.3 Optimization strategy

Optimizing population structure and keeping population diversity: population should be sorted according to its location. Only one individual is left in very node and other individuals are reinitialized, which refers to stochastic pollution sources position and concentration. The goal is that the convergent rate of pollution sources position almost presents an exponential convergent rate without optimization process.

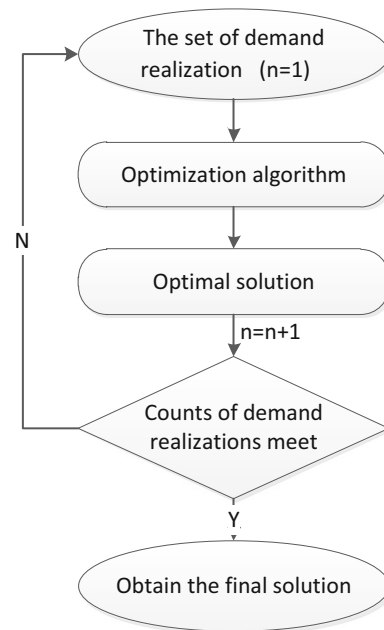


Fig. 4 Uncertainty water demand optimization algorithm

Jumping out of the local optimal solution: when decision algorithm is caught into the local optimal solution, we will compare the adaptive degree of local optimal solution with that of global optimal solution. If the adaptive degree of local optimal solution is smaller than that of global optimal solution, the global optimal solution will be updated and this individual and its adaptive degree will be saved; if not, the global optimal solution will not be updated.

4.4 Algorithm flowchart

A group of stochastic water demand produced by Gaussian model is used as algorithm input when optimization algorithm is used to solve the problem of pollution sources localization of uncertainty water demand. Figure 4 is the block diagram of uncertainty water demand solving by optimization algorithm. In order to calculate the precision of optimization algorithm when pollution source localization problem of uncertainty water demand is solved, n stochastic groups of uncertainty water demand produced by Gaussian model are indispensable. The times of correct pollution source information through statistic algorithm stands for the precision rate of solved algorithm.

In this paper we proposed the improved genetic algorithm, in the new algorithm we added with population structure optimization algorithm and jump out of local optimal solution method, as optimization algorithm to solve the pollution sources localization problem of uncertainty water demand. Population structure optimization means that the individuals in population with same pollution sources location are

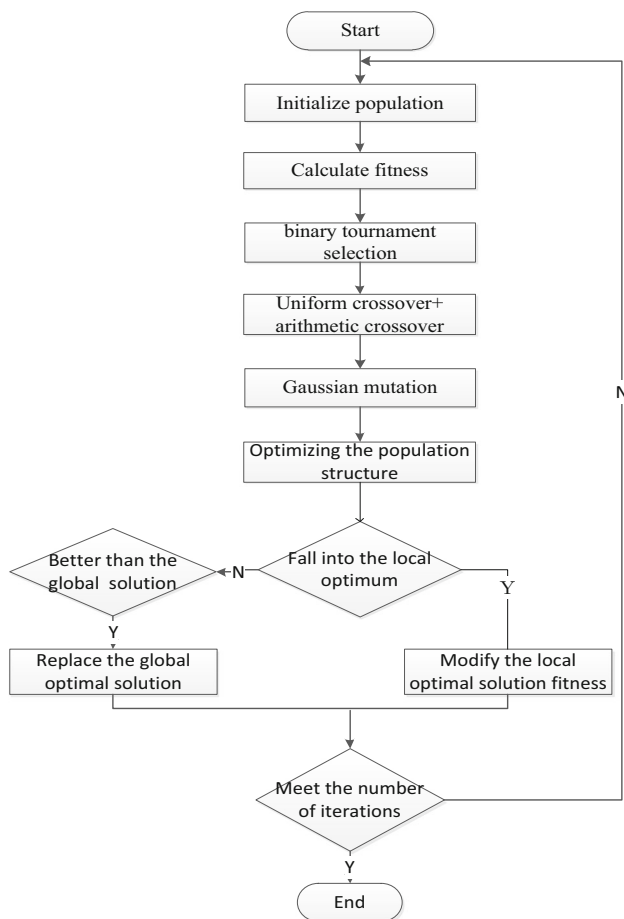


Fig. 5 Improved genetic algorithm flowchart

sorted according to adjective degree after a crossover and variation and before algorithm enters into a next iterative. Among individuals with the same pollution sources location, the one whose adjective degree is the best will be saved and others will be reinitialized. If identical local optimization solutions are produced by multiple-iteration based on genetic algorithm, the optimal solutions fall into local optimization solution. As falling into local optimization solution, optimal solutions should jump out. Or else, optimal solution will be eliminated because algorithm converges too early. The improved genetic algorithm flowchart shown in Fig. 5.

5 Experimental simulation and analysis

5.1 Parameter setting

This paper does experiments on two different water supply pipe networks: standard testing pipe network 1 includes 92 nodes, 3 reservoirs, 2 pools and 2 water pumps; testing pipe network 3 includes 430 nodes, 4 reservoirs, 3 pools and 11 water pumps. 4 sensors are randomly placed in pipe network

in every experiment. We suppose that the threshold value of sensor detected concentration is 0.0001 mg/l. All the nodes in pipe network are supposed to supply water to residential water-supply area. With the incorporation with EPANET simulator, the specific parameters of these two pipe networks are shown in Table 1 and the parameter setting genetic algorithm in experiments is shown in Table 2.

What we need to emphasize is that as white noise algorithm requires 8 times of recurrent water demand for every individual, the water demand parameter of white-noise genetic algorithm is 100 on condition that total water demand realizes certain times.

In this paper, we suppose the real pollution source is a single one. There are two groups of experiments done in this essay: experiment 1, placed on standard pipe network, is to verify the accuracy of algorithm. The reasonability and precision of the algorithm in experiment 1 is checked by the basic algorithm and improved white-noise algorithm mentioned in the article of Praveen et al. for verifying the robustness of algorithm, experiment 2 is placed on big pipe network with basic genetic algorithm and improved genetic algorithm. Statistic algorithm is correspondent to the rate of finding real pollution source.

5.2 Experiment 1: verifying the accuracy of algorithm

The goal of pollution-source-localization problem is to forecast the position of pollution sources and the concentration of injecting pollution. Four sensors are randomly placed in detected pipe network 1. Standard genetic algorithm, white-noise genetic algorithm and genetic algorithm are used in the experiment based on Gaussian model, which can verify the accuracy of the proposed algorithm. Three experimental scenes are shown below (scene 1 and scene 2 are mentioned in the experiments involved in the paper [19]):

- Scene 1: Gaussian model and standard genetic algorithm optimization method;
- Scene 2: Gaussian model and white-noise genetic algorithm optimization method;
- Scene 3: Gaussian model and improved genetic algorithm optimization method.

What we need to notice is that the pollution-source concentration in scene 1 and scene 2 is processed with maximum integer hypothesis, for example, the injection concentration is 2674.21 which is set into 3000 in practical experiments. However, this paper is not so.

800 groups of water demand data randomly produced by Gaussian model are adopted to do experiments in pipe network 1 in every scene. This paper uses hit probability to describe algorithm precision which means water demand times in practical pollution scene optimally simulated by sev-

Table 1 Parameters of network

Parameters	Detected pipe network 1	Detected pipe network 2
The size of pipe network	97	430
Style of pollutions	Single source	Single source
Water demand	Uncertainty water demand but it follows a certain rule	
Time step of hydraulic power	1 h	1 h
Time of water quality	5 min	30 min
Total time of simulation	24 h	48 h
Node of pollution injecting	15	374
Time of injecting pollution	/	0:00
Duration of injecting pollution	/	1 h
Number of sensors	12/12/4	4
Quality of pollution injecting	2674.21	300

Table 2 Algorithm parameter setting

Parameter	Standard GA	White-noise GA	Improved GA
Size of population	80	80	80
Iterations	40	40	40
Realized times of water demand	800	100	800
Probability rate	0.8	0.8	0.8
Mutation probability	Gaussian mutation	Gaussian mutation	Gaussian mutation

Table 3 Algorithm hit probability

Pollution scene	Standard GA (%)	White-noise GA (%)	Improved GA (%)
Hit probability	68	80	100

eral water demand times are divided into the total times of water demand. It is shown in formula (4):

$$P = \frac{N_f}{N_s} \tag{4}$$

In formula (4), P indicates for hit probability, N_f indicates for the times when real pollution sources are found, N_s indicates for the total times of simulated water demand.

As it is shown in Table 3, the rates of finding pollution source position and concentration with three different algorithms are 68% (Standard GA), 80% (White-noise GA) and 100% (Improved GA). Figure 6 shows that: in scene 1, the possible pollution source nodes found by standard genetic algorithm are presented by circles; in scene 2, the possible pollution sources nodes found by white-noise genetic algorithm are expressed by triangles; in scene 3, the experimental results is that the node position of pollution sources is only marked by triangles. Triangle stands for the nodes position of real pollution sources and stars and circles stands for the position of jamming nodes which are introduced by algorithm uncertainty. This is because genetic algorithm added with improved strategy helps algorithm to effectively jump out of local optimal solution. Meanwhile, the size of pipe network, containing not too many nodes, is relatively small,

therefore, disturbance items can be excluded and real pollution source can be found.

5.3 Experiment 2: the robustness of verification algorithm

In experiment 1, it is proved that the accuracy rate of improved algorithm is the highest one among three testing algorithms. Now, we need to verify whether this algorithm is still effective in big pipe network or not. Four sensors are placed in experiment 2 which runs in pipe network 2. Two scenes are set:

- Scene 4: Gaussian model and standard genetic algorithm;
- Scene 5: Gaussian model and improved genetic algorithm.

As it is shown in Table 4, the rates of finding pollution source through algorithm are 33% (SGA) and 65% (Improved GA). Figure 7 presents the position nodes of real pollution sources marked by triangles. Figure 8 is an amplification result of real pollution sources location, coming from Fig. 7. Circles stand for the node location of possible pollution sources found by improved algorithm; stars stand for the node location of possible pollution sources found by standard genetic

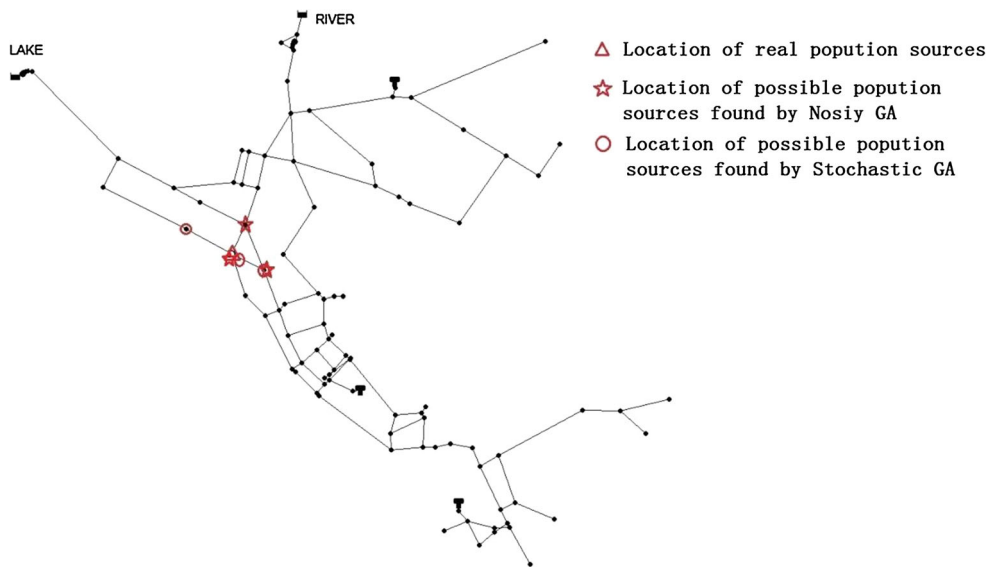


Fig. 6 Experiment 1 results

Table 4 Algorithm hit probability

Pollution scene	Standard GA (%)	Improved GA (%)
Hit probability	33	65

△ Location of real pollution sources



Fig. 7 Experiment 2 results

△ Location of real pollution sources
○ Location found by Improved GA
☆ Location found by Stochastic GA

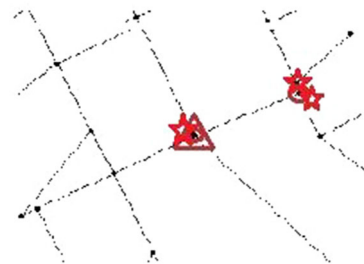


Fig. 8 Experiment 2 amplification results

algorithm. Compared with the experiment 1 results, basic genetic algorithm is not suitable to localize pollution sources in big pipe network. While, though improved genetic algorithm cannot find the node location of real pollution sources with a 100%, it has a very high percentage of finding the node location of real pollution sources between two false locations. Thus, improved genetic algorithm can solve the problem of pollution sources localization in relatively big pipe network.

6 Conclusion

Pollution sources localization problem is an interdisciplinary problem, covering environmental science and computing science. This essay studies the pollution sources localization problem under uncertainty water demand. The problem of pollution sources identification are defined as optimization problem by known hypothesis of water supply-

ing pipe network and simulation–optimization model is used in proof. In order to solve this problem, the water demand randomly produced by Gaussian model is used to simulate uncertainty water demand and improve genetic algorithm is proposed as optimization algorithm. Through simulation experiments, we check the accuracy of algorithm which is applied in two different pipe networks and the robustness of algorithm is checked as well. In the study of pollution sources localization, if the pipe network nodes in cities and towns are more than 1000 and residential water demand constantly changes, this problem can be regarded as an abstract optimization problem which are dynamic, large-scale and multimodal function. Therefore, further proposing a method to solve a dynamic, large scale and multi-model problem is the following research of this paper.

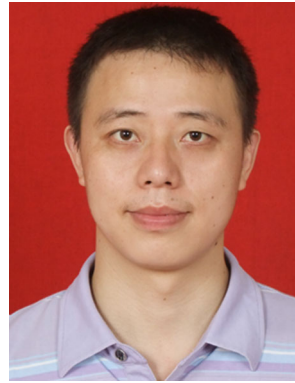
Acknowledgements This paper is supported by Natural Science Foundation of China (Nos. 61402425, 41404076, 61501412, 61673354, 61672474).

References

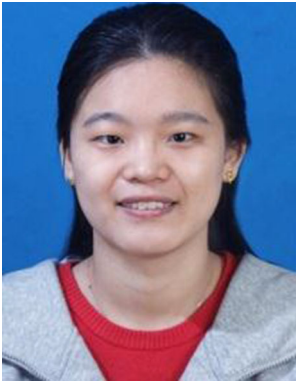
- Zhangjie, F., Xingming, S., Qi, L., Lu, Z., Jiangang, S.: Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing. *IEICE Trans. Commun.* **98**(1), 190–200 (2015)
- Gu, B., Sheng, V.S., Wang, Z., Ho, D., Osman, S., Li, S.: Incremental learning for ν -support vector regression. *Neural Netw.* **67**, 140–150 (2015)
- Guo, P., Wang, J., Geng, X.H., Kim, C.S., Kim, J.-U.: A variable threshold value authentication architecture for wireless mesh networks. *J. Internet Technol.* **15**(6), 929–935 (2014)
- Xie, S., Wang, Y.: Construction of tree network with limited delivery latency in homogeneous wireless sensor networks. *Wirel. Personal Commun.* **78**(1), 231–246 (2014)
- Hu, C., Li, M., Zeng, D., Guo, S.: A survey on sensor placement for contamination detection in water distribution systems. *Wirel. Netw.* (2016) 1–15. doi:[10.1007/s11276-016-1358-0](https://doi.org/10.1007/s11276-016-1358-0)
- Zeng, D., Teng, C., Gu, L., Yao, H., Liang, Q.: Flow setup time aware minimum cost switch-controller association in software defined networks. 11th International Conference on IEEE. 259–264 (2015)
- Shang, F., Ulber, J.G., Polycarpou, M.M.L.: Particle back-tracking algorithm for water distribution systems analysis. *J. Environ. Eng.* **128**, 441–450 (2002)
- Laird, C.D., Biegler, L.T., van BloemenWaanders, B.G., Bartlett, R.A.: Contamination source determination for water networks. *J. Water. Resour. Plan. Manag.* **131**, 125–134 (2005)
- De Sanctis, A.E., Shang, F., Uber, J.G.: Real-time identification of possible contamination sources using network backtracking methods. *J. Water. Resour. Plan. Manag.* **136**, 444–453 (2010)
- Costa, D.M., Melo, L.F., Martins, F.G.: Localization of contamination sources in drinking water distribution systems: a method based on successive positive readings of sensors. *Water Resour. Manag.* **27**, 4623–4635 (2013)
- Huang, J.J., McBean, E.A.: Data mining to identify contaminant event locations in water distribution systems. *J. Water Resour. Plan. Manag.* **135**, 466–474 (2009)
- Perelman, L., Ostfeld, A.: Bayesian networks for source intrusion detection. *J. Water Resour. Plan. Manag.* **139**, 426–432 (2013)
- Wang, H., Jin, X.: Characterization of groundwater contaminant source using Bayesian method. *Stoch. Environ. Res. Risk Assess.* **27**, 867–876 (2013)
- Wang, H., Harrison, K.W.: Improving efficiency of the Bayesian approach to water distribution contaminant source characterization with support vector regression. *J. Water Resour. Plan. Manag.* **140**, 3–11 (2014)
- Guan, J., Aral, M.M., Maslia, M.L., Grayman, W.M.: Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. *J. Water Resour. Plan. Manag.* **132**, 252–262 (2006)
- Liu, L., Ranjithan, S.R., Mahinthakumar, G.: Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure. *J. Water Resour. Plan. Manag.* **137**, 183–192 (2010)
- Hu, C., et al.: A MapReduce based parallel niche genetic algorithm for contaminant source identification in water distribution network. *Ad Hoc Netw.* **35**, 116–126 (2015)
- Yan, X., et al.: Contaminant source identification in water distribution network based on hybrid encoding. *J. Comput. Methods Sci. Eng.* **16**, 379–390 (2016)
- Vankayala, P., et al.: Contaminant source identification in water distribution networks under conditions of demand uncertainty. *Environ. Forensics* **10**, 253–263 (2009)
- Holland, J.H.: *Adaptation in Natural and Artificial Systems*, 1st edn. MIT Press, Cambridge (1975)
- Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley Reading, Menlo Park (1989)
- Dogan, A., Ozguner, F.: Genetic algorithm based scheduling of meta-tasks with stochastic execution times in heterogeneous computing systems. *Cluster Comput.* **7**(2), 177–190 (2004)
- Ramirez, A.J., et al.: Plato: a genetic algorithm approach to runtime reconfiguration in autonomic computing systems. *Cluster Comput.* **14**(3), 229–244 (2011)
- Moghaddam, F.F., et al.: Carbon-aware distributed cloud: multi-level grouping genetic algorithm. *Cluster Comput.* **18**(1), 477–491 (2015)
- Shojafar, M., et al.: FUGE: a joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method. *Cluster Comput.* **18**(2), 829–844 (2015)
- Wu, Q., et al.: Multi-label classification algorithm research based on swarm intelligence. *Cluster Comput.* **19**(4), 2075–2085 (2016)
- Sheng, X., Kefeng, W.: Motif identification method based on Gibbs sampling and genetic algorithm. *Cluster Comput.* 1–9 (2016). doi:[10.1007/s10586-016-0699-x](https://doi.org/10.1007/s10586-016-0699-x)
- Yan, X., et al.: Hybrid genetic algorithm for engineering design problems. *Cluster Computing.* 1–13 (2016). doi:[10.1007/s10586-016-0680-8](https://doi.org/10.1007/s10586-016-0680-8)
- Meng, X., et al.: A genetic algorithm using K-path initialization for community detection in complex networks. *Cluster Computing.* 1–10 (2016). doi:[10.1007/s10586-016-0698-y](https://doi.org/10.1007/s10586-016-0698-y)



Yan Xuesong received the M.S. degree in Computer Application Technology from China University of Geosciences in 2003, and the Ph.D. degree in Computer Software and Theory from the School of Computer Science, Wuhan University in 2006. He is currently an associate professor in China University of Geosciences. His research interests are in the areas of evolutionary computation, data mining, machine learning and computer applications.



Hu Chengyu received the M.S. degree in Automation from Wuhan University of Technology in 2003, and the Ph.D. degree in Automation from School of Mechanical Science and Engineering, Huazhong University of Science and Technology in 2010. He is currently an associate professor in China University of Geosciences. His research interests are in the areas of intelligence computation and automation.



Sun Jie received the B.S. degree in Computer science and technology from China University of Geosciences in 2014. She is currently a Master candidate in China University of Geosciences. Her research interests are in the areas of evolutionary computation and machine learning.