Check for updates

An intelligent traceability method of water pollution based on dynamic multi-mode optimization

Qinghua Wu[1] · Bin Wu[2] · Xuesong Yan[3]

Abstract

Drinking water safety is a safety issue that the whole society attaches great importance to currently. For sudden water pollution accidents, it is necessary to trace the water pollution source in real time to determine the pollution source's characteristic information and provide technical support to emergency management departments for decision making. The problems of water pollution's real-time traceability are as follows: non-uniqueness and dynamic real time of pollution sources. Aiming at these two difficulties, an intelligent traceability algorithm based on dynamic multi-mode optimization was designed and proposed in the work. As a multi-mode optimization problem, pollution traceability could have multiple similar optimal solutions. Firstly, the new algorithm divided the population reasonably through the optimal subpopulation division strategy, which made the nodes' distribution in a single subpopulation more similar and conducive to local optimization. Then, a similar peak penalty strategy was used to eliminate similar solutions and reduce the non-unique solutions' number, since real-time traceability required higher algorithm convergence than traditional offline traceability and dynamic problems with parameter changes, historical information preservation, and adaptive initialization strategies could make reasonable use of the algorithm's historical knowledge to improve the population space and increase the population convergence rate when the problem changed. The experimental results showed the proposed new algorithm's effectiveness in solving problems—accurately tracing the source of pollution, and obtain corresponding characteristic information in a short time.

1 Introduction

Drinking water safety is a public safety issue that is highly concerned by urban residents, and it is closely related to people's lives. With the improvement in people's living standards and the enhancement of environmental

protection awareness, more and more people pay attention to the safety of drinking water quality. As the basic industry and public facility of the national economy and socioeconomic development, the drinking water supply system is prone to be damaged due to its wide coverage, long-term operation, and strong relevance, which will cause panic and huge economic losses to the society. Sudden drinking water pollution incidents mainly include biological pollution, chemical pollution, and mixed pollution. In addition to human factors and natural disasters, sudden public health events also significantly affect the safety of drinking water quality (medical sewage and domestic sewage). Since the outbreak of the Novel Coronavirus, the urban drinking water supply system has faced more severe challenges: How to improve the water

✉ Xuesong Yan
  yanxs@cug.edu.cn

1 Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China

2 School of Economics and Management, Nanjing Tech University, Najing 211816, China

3 School of Computer Science, China University of Geosciences, Wuhan 430074, China

supply system's real-time fault tolerance capacity and shorten the response time after water pollution's occurrence has become key issues. Therefore, the study on drinking water quality pollution traceability has great significance to the guarantee of urban water resources environment and water system safety in the new environment [1].

The urban drinking water supply system is a complex system composed of several spatially interrelated components, including water head, water source, and water pump connected by water pipes. Its topology can be represented as an undirected graph. In the drinking water supply system, water quality sensors can be arranged at key nodes, which can send out early warning messages and continuously record real-time water quality information data once water pollution is detected. With the passage of pollution time, the information collected by the water quality sensor increases. It is a huge challenge to trace or locate water pollution in real time through its characteristics and quickly infer the injection location, injection time, and the concentration and quality of pollutants. The real-time traceability of water pollution depends on the water quality and hydraulic models for parameter inversion or prediction through the node pollutant concentration data provided by water quality sensors. The main information of the inverted pollution source contains four dimensions: the node index represented by the pollution source, the start time of pollution, the duration of injected pollutants, and the injection quality per unit time step. Real-time traceability starts the simulation solution when the pollutants are detected by the water quality sensor for the first time, which is convenient for the urban water quality supervision department to carry out relevant treatment measures in time, such as the release of water quality indicators or the issuance of emergency notifications. However, the cost of high-precision water quality sensors and the time-consuming of hydraulic and water quality model simulations will increase the difficulty of water pollution's traceability, which gives the problem the following complex characteristics:

(1) *Dynamic real time* First, the model established by the inversion of the intelligent optimization algorithm can be defined as a dynamic optimization problem, since it contains the time-varying characteristics of the dynamic optimization problem; that is, the problem parameters, variables, scale, and constraints will change with time. Second, due to the speed requirement of the problem, which is hoped that the pollution source can be accurately located when the pollution event is monitored, the water pollution's traceability is a process of obtaining real-time solutions, that is, obtaining a satisfactory solution to the problem within an acceptable time. Based on

the two factors, the high requirements for the convergence speed and solution accuracy of the intelligent optimization algorithm make it extremely difficult to solve the problem of water pollution's real-time traceability.

(2) *Non-uniqueness of solutions* Simulation data generated by simulation software needed to be compared with the real monitoring data in the solution process due to the inversion problems' characteristics. Since the data size in a fixed period was limited and the water quality sensors' arrangement was not global, the monitoring data size was not large. The data's finiteness led to multiple solutions satisfying the characteristics of the existing data during solving the problem, which satisfied the characteristics of multimode optimization problems. The basic requirement of intelligent water pollution traceability was to accurately locate the water pollution source, which posed a higher challenge to the solution algorithm: the intelligent optimization algorithm could obtain the only accurate solution on the premise of a small amount of data.

Hydraulic models and water quality models were firstly established in most studies on water quality pollution traceability. Combined with the forward simulation of water quality, the solutions of water quality pollution traceability were mainly divided into three categories: simulation–optimization method, probability selection method, and other methods. The simulation–optimization method was classic to solve the pollution traceability problem. It was good in solution accuracy and robustness but had the disadvantages of high calculation cost and slow solving speed, which could only be applied to small or medium-scale problems. The probability selection method mainly used the Bayesian network model to discriminate the pollution source information and select the most likely location in a probabilistic manner. Others include analytically based methods such as particle backtracking, hydraulic and water quality models, neural networks, or algorithms that combined multiple strategies. The work mainly discussed the simulation–optimization method.

In 1998, Zierolf et al. [2] proposed the I/O model, which is a backward tracking model of reverse thinking, that is, tracking the particles' transportation process in the pipe network according to the reverse time. According to this model, Shang et al. [3] used the particle backtracking algorithm for pollution source location in 2002, which decomposes the water quality model's basic elements, calculates the impact factors used to describe the reservoir and multiple pollution sources output functions, and re-establishes the water quality model's pipeline constraints. The experimental results show that the optimal solution is

valid. In 2006, Guan et al. [4] proposed a new simulation–optimization backtracking model, which is simulated by EPANET software and solved by the gradient descent method. The model results' accuracy and computational efficiency are good and have good scalability to accommodate various optimization algorithms.

Liu et al. [5] proposed a dynamic simulation–optimization framework in 2011, which connects the inversion problem with the dynamic optimization problem. It studies the dynamics of the problem and used some adaptive operators to track the optimal solutions in real time. The experimental results are significantly better than previous ones. In 2015, Hu et al. [6] quantified the non-uniqueness of the pollution source location problem's solution and proposed a parallel niche genetic algorithm based on MapReduce to quickly and accurately obtain the location, quality, and time vector of the pollution sources. In 2016, Arpan et al. [7] compared three algorithms for solving pollution source location problems, which were the Bayesian probability-based method, pollution source state-based method, and simulation optimization method. Finally, the simulation optimization method can solve most pollution intrusion events. In 2017, Wang et al. [8] proposed a temporal Bayesian method. First, the sensor observations' probability distribution is obtained when the node is a polluted one. Then, the polluted node's posterior probability is calculated in real time and decomposed into a hierarchical tree structure. Finally, the node with the highest posterior probability is selected as the polluted node in real time. In 2017, Yan et al. [9] simulated uncertain water demand through a Gaussian model and proposed a real-time positioning algorithm. Experiments show that the method can find real pollution events in a short time with fewer sensor data. In 2019, Yan et al. [10] continued to reduce the positioning time on this basis to simulate the real-time location and solved the problems according to the designed simulation optimization model and algorithm. The real pollution source can be located by a small amount of sensor data.

More and more people pay attention to the influence of the water pipe network's uncertain factors on the positioning problems. In 2009, Torres et al. [11] pointed out that understanding the water pipe network's uncertainties can establish a robust model, perform sensitivity analysis, and effectively uncertainty levels. On this basis, in 2011, Preis and Ostfeld et al. [12] proposed a method to add hydraulic ambiguity on the premise of knowing a small amount of hydraulic data and the single sensor's effective data. Statistical knowledge is used to process the node demand's upper and lower bounds, and the experiments are carried out in three different situations. In 2009, Vankayala et al. [13] considered the node demand as the biggest driving force for the water pipe network's uncertainties

under the condition that the hydraulic water quality model remained unchanged. Therefore, they used the noisyGA (noisy Genetic Algorithm) algorithm to solve the positioning problem when the water demand was uncertain. The uncertain water demand is closer to the real pipeline network situation. Yan et al. [14, 15] further simulated the water demand model to solve the water demand's uncertainties and designed the traceability algorithms based on the Poisson model and the autoregressive model. The experimental results prove the algorithm's effectiveness for water pollution's traceability. Yan et al. [16–18] proposed an intelligent optimization method based on expensive optimization methods to solve water pollution's intelligent traceability in the large-scale pipe network nodes. The results prove the effectiveness of the large-scale pipe networks. Hu et al. used optimization algorithms for sensor placement in water distribution system [19–21].

The work's remaining chapters are organized as follows: Chapter 2 mainly introduces the problem model of water-pollution traceability and the process of real-time traceability. Chapter 3 is the algorithm's design of water pollution's intelligent traceability based on multi-mode optimization, including the algorithm ideas, multi-strategy improvements, and the algorithm's overall steps. Chapter 4 shows the simulation experiment and analysis, where the proposed algorithm is verified through different pollution scenarios.

2 Traceability of water pollution

The general process of solving the traceability problem of water pollution by the simulation–optimization method was as follows: Firstly, take the pollution source information as the input event of the water quality and hydraulic model simulation software EPANET for the corresponding simulation. Then, the offline data generated by the simulation software were used to calculate the function's fitness and trace the source. The solution to this problem mainly included four dimensions of the information: the node index represented by the pollution source, the pollution's start time, the duration of injected contaminant, and the injection quality per time step. The work adopted real-time simulation to trace the water pollution's source. When the fitness function was calculated, the water pollution traceability's optimization target was changed from a static problem to a dynamic problem by inputting dynamic simulation-time parameters to simulation software EPANET. The real-time positioning's fitness function is shown in Eq. 1.

$$\text{Minimize} f = \frac{\sqrt[2]{\sum_{t=t_0}^{T} \sum_{i=1}^{N_s} \left(C_{i,t}^{\text{obs}} - C_{i,t}^{*} \right)^2}}{N_s}$$
$$\text{S.T.} C = (idx, t_s, t_d, M) \tag{1}$$
$$M = (m_1, m_2, \ldots m_n), \; m_i > 0$$
$$t \in T$$

where C^{obs} and C^{*} separately represent the sensor concentration under the simulated and the real pollution events, mainly determined by the input pollutant information and the simulation duration. It contains four related variables of $\{idx, t_s, t_d, M\}$ with idx as the source point location, t_s as the start time, $t_d a$ s the duration, and M as the injection quality per time step, whose dimension changes with t_d. Besides, T represents the current simulation duration and the current actual time and N_s the sensors' number. The optimization target minimizes the concentration difference between the real pollution source represented by f and the simulated pollution source event, which is a dynamic optimization function that minimized the target.

Once the pollution event occurs, the polluted water body will be monitored by the sensor when it passes through the water quality monitoring sensor pre-arranged in the network, which is the same as the scene when the actual pollution occurs. Then, the time detected for the first time is taken as the algorithm's starting time t_0 and be simulated with a present time step of Δt. Similar to the dynamic optimization's target, the optimal solution of the current period is obtained in each time step Δt. Meanwhile, the most possible pollution source node and pollutant injection information are found in a relatively short time, and relevant departments are guided to make decisions on public events, thus minimizing the public impact caused by pollution. In the previous non-real-time positioning studies, the sensor's concentration 24 or 48 h after the first detection of the pollutant is often used as the offline simulation data for optimization.

However, in the case of real-time positioning, the sensors' readings may fluctuate greatly in different events due to the difference in pipe network characteristics and the sensor's layout scheme. In Fig. 1, the sensor detects the pollution for the first time around 9 h. Assuming that the simulation time step is 1 h, in the first hour of the simulation, the real pollution source's concentration is too small that any simulated pollution source injected a small amount of pollution in this period leads to an extremely small concentration difference, which cannot guide the population evolution correctly. During the 10 to 12 h of the simulation, the simulated observation concentration fluctuates sharply compared with the previous one, resulting from a large number of previous high-quality solutions' quality reduction. The real-time positioning requires a

reasonable response to the dynamic changes caused by the simulation duration's change to ensure the algorithm's solving ability. Figure 2 shows the real-time traceability flow of the work.

In the optimization process, discrete or continuous traditional optimization algorithms can be selected as the basis for improvement according to different coding methods. The main improvement directions can be roughly divided into two types: 1. The exploration and inquiry capabilities of improving the algorithm through optimization strategies, representing the algorithm's evolutionary optimization and refinement capabilities, respectively. 2. Based on the pollution sources' dynamic characteristics, the algorithm is dynamically adjusted during the evolution process to better adapt to the false influence caused by the dynamic environment and the parameters. Finally, the most possible pollution source events are obtained by outputting the optimization algorithm.

3 Intelligent real-time traceability method based on dynamic multi-mode optimization

3.1 Algorithm idea

When the intelligent real-time water pollution's traceability is combined with the dynamic optimization, the actual problems' peculiarities should be considered, such as the objective function's variation characteristics, the pipe network parameters' characteristics, and the simulation time. Experimental analysis is needed to improve the strategy. In dynamic optimization cases, changes in environmental parameters cause the continuous change of the optimal solution's position. The unchanged optimal solution should be found before each change, with the algorithm's evolution direction adjusted. With swarm intelligence algorithms, if the population cannot be dynamically adjusted over time, the intelligence optimization algorithm's convergence characteristics will cause the individuals in the population to lose diversity, resulting in the algorithm's stagnation and the inability to continue moving toward the new optimal solution. However, blindly pursuing the population's diversity would cause the randomness and dispersion of the entire population, and sometimes the optimal value cannot be obtained at all. Therefore, a dynamic optimization algorithm is needed to balance the population's diversity while the historical optimal solution is rationally used. Simultaneously, the non-uniqueness problem caused by the incompleteness of the pipeline network data should be considered when the water pollution's source is traced through inversion based on intelligent optimization methods.
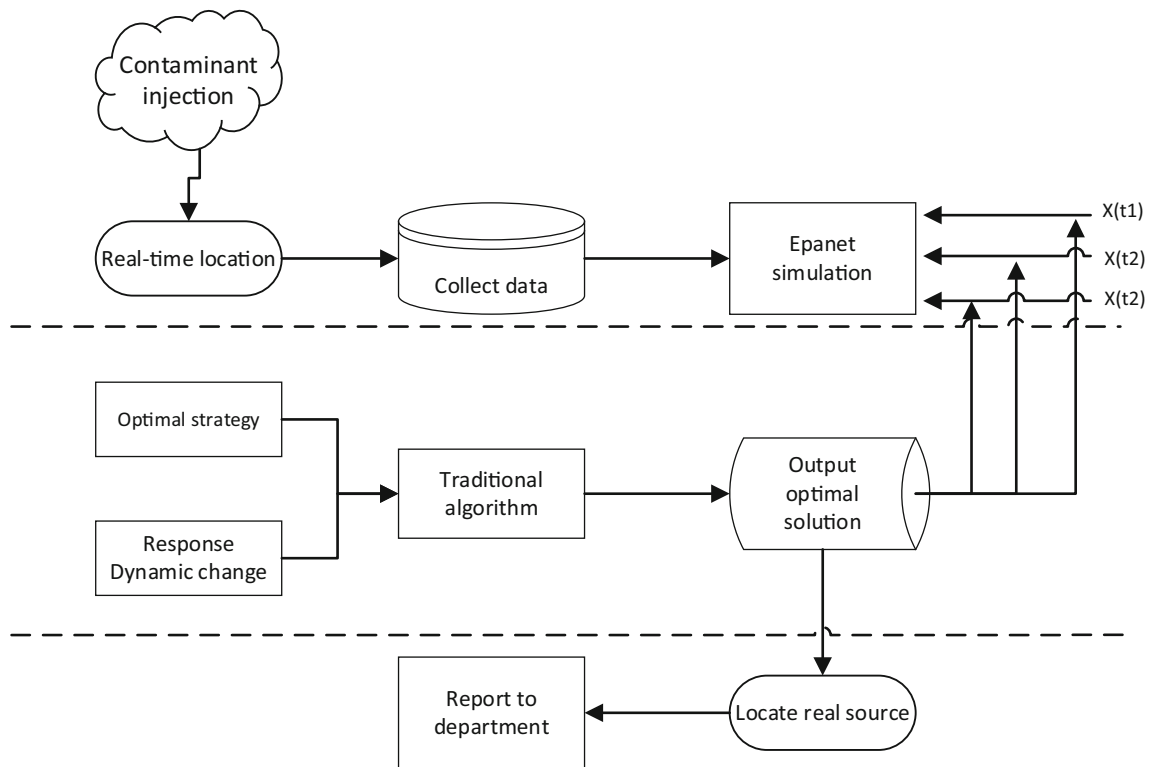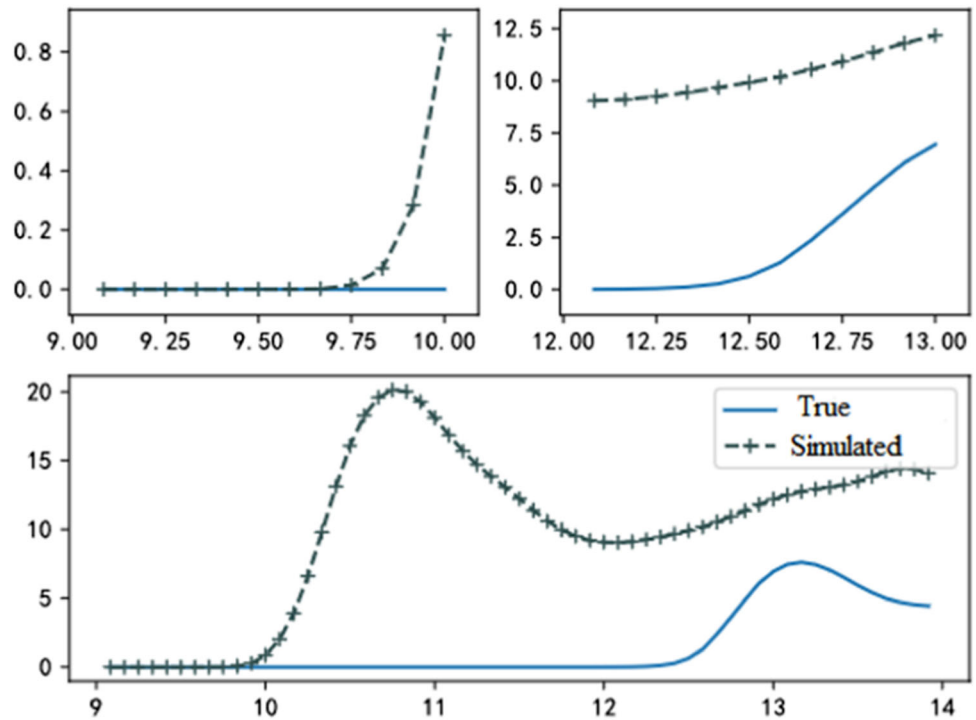
Fig. 1 Concentration contrast

Fig. 2 Real-time positioning framework

Based on the above ideas, the work designed an intelligent real-time traceability algorithm for water pollution based on dynamic multi-mode optimization. Figure 3 shows the algorithm's overall flow, which mainly includes:

1. Optimal subpopulation division strategy. 2. Historical information retention strategy. 3. Similar peak punishment strategy. 4. Quality local search strategy. 5. Adaptive initialization strategy. 6. Redundant individual elimination
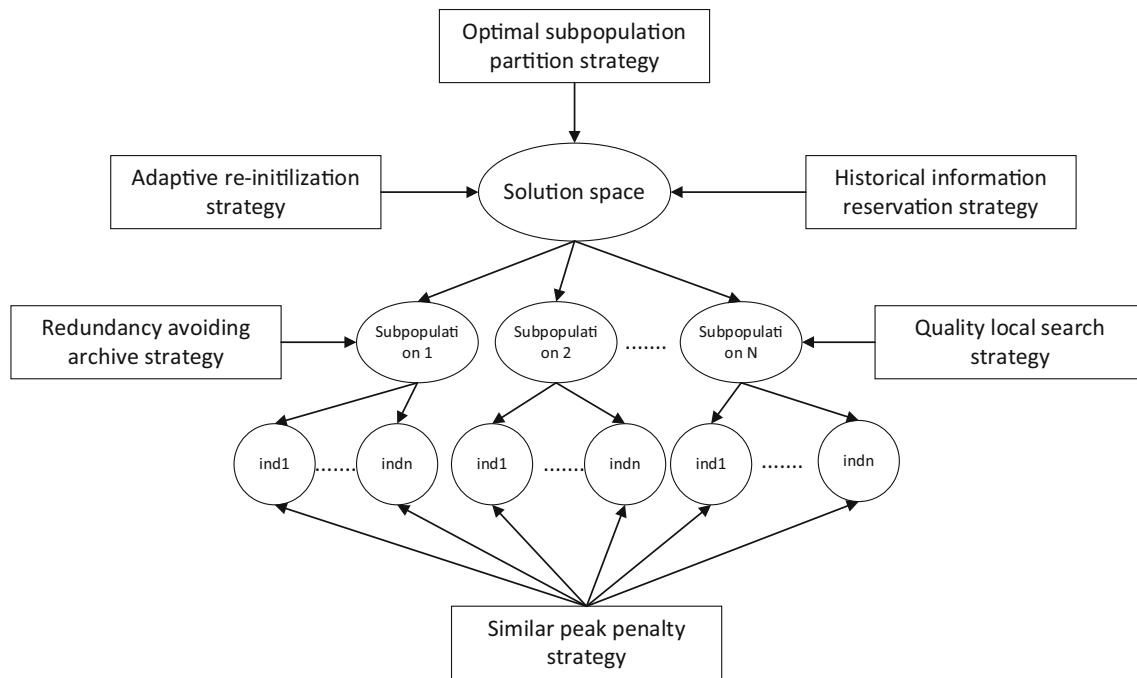
Fig. 3 Intelligent water pollution traceability algorithm framework based on dynamic multi-mode optimization

strategy. The multi-modal algorithm adopted a multi-population approach to solve the non-uniqueness problem. The optimal subpopulation division strategy was used to cluster the pollution source feature information randomly generated, with the performance indicators proposed for evaluation. Finally, iteratively select the clustering division with the optimal classification. After the single population converged, the convergent local optimal solution would be saved. At the same time, the quality partial search strategy should be introduced to improve the algorithm's accuracy. The reason was that in the pollution source's characteristic information, the dimension of the injection quality changed with the injection duration. The dimension of the real pollution source's injection quality to be solved was uncertain, and the most direct influence was that the characteristic information of which part injected quality was not precise enough. After the algorithm's solving ability improved, strategies should be set up to cope with the environmental changes as they occurred. When the environmental changes' degree was uncertain, the adaptive initialization strategy could be adopted, which made reasonable use of the historical optimal solution reserved for each period to improve the algorithm's convergence speed. Finally, when multiple populations evolved to the same peak, the populations with poor quality should be punished to save algorithm resources.

3.2 Improved multiple strategies

3.2.1 Optimal subpopulation division strategy

The problem is considered as multi-mode optimization due to the non-uniqueness of water pollution traceability [22]. Therefore, the whole population is generally divided into multiple subpopulations to better search the entire solution space so that each subpopulation can have an optimal solution and perform independent optimization iterations, thus preserving the entire population's diversity. However, it is difficult to determine the way to divide the population and the subpopulations' number for the solution space with unknown local peaks. If there are too many subpopulations in the entire solution space, the single subpopulation is prone to stagnation due to the small number of individuals. On the contrary, if there are too few subpopulations, then the single population can easily fall into the local optimum, and the algorithm as a whole can hardly find the global optimum.

The work adopted a clustering method to self-adaptively divide the sub-populations, thus overcoming the above shortcomings of multi-group division. Different from the clustering methods [23, 24] in the previous multi-population optimization algorithms, the optimal subpopulation partition strategy (OSPS) proposed in the work could adaptively adjust the population quantity. The strategy's execution process was as follows: After randomly

initializing the entire population, OSPS began to search for the most reliable population division strategy:

(1) The k-means clustering algorithm is used to generate a series of division schemes, where the k value changes with each iteration until reaching the preset maximum population number MAX_CLUSTER_NUM. (2) During each iteration, judge the division schemes with different k values, calculate the population division metric of the current division scheme, and keep the population division scheme represented by the minimum value of *pdm*. Until it reaches the maximum number of iterations MAX_GEN, the process terminates and the optimal division scheme is adopted. *pdm* is calculated by Eq. (2).

$$pdm = \min_{1 \le g \le \text{MAX\_GEN}} \frac{\sum_{i=1}^{subN} \sum_{j=1}^{size(i)} d\left(indi_j, CS_i^g\right)}{\min_{1 \le i \le subN}\left\{\min_{j \ne i} d_{\min}\left(C_i^g, C_j^g\right)\right\}},$$
(2)

where

$$d_{\min}\left(C_i^g, C_j^g\right) = \min_{indi_i \in C_i^g, indi_j \in C_j^g} d_{\min}\left(indi_i, indi_j\right)$$

where $d\left(indi_i, indi_j\right)$ is the Euclidean distance between individuals i and j; C_i^g and CS_i^g are the ith cluster with g generations' evolution and its cluster center, respectively; subN is the subpopulations' number after using k-means clustering; $d_{\min}$ is the minimum distance between two different clusters, determined by the minimum distance between two populations of individuals. The smaller the pdm, the stronger the cohesion of a single subpopulation (the smaller the molecule in Eq. 2), the larger the distance between each population, and the smaller the possibility of mutual coverage. After dividing the optimal subpopulation, it should consider the situation that the single population is too discrete and the individuals' number is too small. As shown in Fig. 4a, the fewer individuals are merged if the

right half has fewer individuals than NP/MAX_CLUSTER_NUM (NP is the subpopulations' number). Figure 4b shows the division effect after merging. Meanwhile, it is necessary to keep the single population's number not too large (more than $2 \times$ NP/MAX_CLUSTER_NUM) or select the population with too far cluster center distance in terms of the two populations selected to be merged.

3.2.2 Historical information preservation and redundant individual elimination strategies

As mentioned above, when the environment changes in the dynamic optimization problems, it is necessary to reasonably balance the population diversity according to the changes. One of the methods is to preserve the historical optimal solution before the environmental changes. Using the historical archive-based method, the effective information in the predecessor period can be transferred to the current environment, leading the populations' cross-mutation in a favorable direction. In some studies, the historical archive method has been proved to be effective in accelerating population convergence. However, on the other hand, it is also necessary to consider other influences caused by the excessively large historical archive solution set and the same or similar individuals' excessive redundancy: Waste of computing resources, the diversity's loss caused by too intensively reusing individuals in the population. Therefore, the historical information reservation strategy (HIRS) was adopted in the work to preserve the historical optimal individuals; meanwhile, the redundant avoiding archive strategy (RAAS) was used to remove similar individuals. The first is the judgment of similarity. Three variables of the pollution source characteristic information except for the injection quality are taken as the benchmark, and individuals with the same source location,
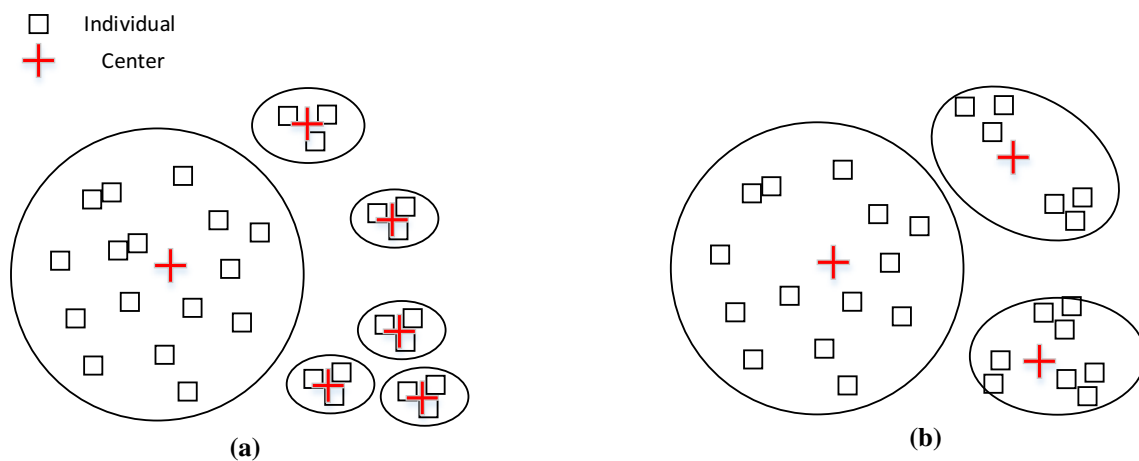

Fig. 4 Subpopulation division: **a** before the subpopulations merge; **b** after the subpopulations merge

starting, and duration time are similar. In the pollution source characteristic information, these three variables having a relatively large impact on the objective function's value are discrete variables with a relatively small range. For contaminant injection quality, it is difficult to take equal values as the basis for the similarity of two individuals because of the real value.

The second is the time to preserve the population's optimal solution. The breadth search was the intelligent optimization algorithm's main focus in the early stages, searching for potential peaks in the solution space, and then taking fitness function and evolution operator as the direction to specifically explore the optimal value node on a peak. Therefore, preserving the populations' optimal solution should not be in the algorithm evolution's exploration stage, where the adaptability is often poor and cannot afford to represent a local optimum. Instead, the local optimal value should be retained gradually after the population entered the convergence stage. In the study [25], the population converges are judged mainly by the optimal individual fitness or the variations in the population radius, but the premise is established based on the general test function.

It is inaccurate to use Euclidean distance to calculate the population radius in water pollution traceability, because the source location information between individuals does not represent their spatial distribution in the water supply network, but only serves as a number. Not only that, in the early stage of real-time pollution traceability, some subpopulations have poor fitness due to the lack of pollutant concentration data and remain unchanged after multiple iterations. The misjudgment of the population convergence leads to the preservation of inferior individuals in the populations' historical archives. After the population converges, the population will move toward the optimal individual in the current population affected by the genetic operators, and there will be a large number of similar individuals in the population. Therefore, when the optimal individual's repetition rate in a population exceeds certain degree rr, it is considered as the population converges.

$$\text{pBest}_i : \left(idx_b, t_{s,b}, t_{d,b}, M_b \right), \text{ind}_i : \left(idx, t_s, t_d, M \right)$$
$$if : idx_b = idx \text{then} : \text{pBest}_i = \text{ind}_i, N_{rr} = N_{rr} + 1 \quad (3)$$

$$\frac{N_{rr}}{N_i} > rr \quad (4)$$

$$f_{\text{cur}} = \frac{\sqrt[2]{\sum_{t=t_0}^{T} \sum_{i=1}^{N_s} \left(C_{i,t}^* \right)^2}}{N_s} \quad (5)$$

As shown in Eq. (3), pBest_i is the best individual in population i, and ind_i is any individual in population i. When judging two individuals locating at the same

pollution source node, the current population's repeat number N_{rr} increases by 1. N_i is the individual number of the population i, and the current population is considered to converge when the proportion of N_{rr} exceeds repetition threshold rr. Judge the fitness of the subpopulation's optimal individual before the historical archive. If the current individual fitness value satisfies Eq. (5), it will not be added to the historical archive. The main purpose is to eliminate individuals with all zero sensor concentration values simulated in the early stage of real-time positioning and eliminate some effects caused by insufficient data. Besides, the one with the smaller fitness for similar individuals is retained in the historical archive.

3.2.3 Adaptive initialization policy

After setting up the historical archive, it is necessary to adjust the use of these historically optimal individuals. An Adaptive Re-initialization Strategy (ARS) was formulated according to the dynamic environment's changing degree in the work, aiming at dynamically adjusting the individuals' number in the HIRS archive to balance the populations' diversity. After the environment changed, the change of parameter T of the pollution traceability caused varying degrees of changes in the individual fitness of the population, which directly affected the average fitness of the population. Moreover, the populations needed to reorient their evolution direction toward the new peak. In the real-time evolution process, the set evolution time step and the iterations' maximum number on each step could be considered as the changing moment in the dynamic optimization problem.

Then, before and after each change, the impact ecs(environmental change severity) caused by the current change is calculated according to Eq. (6), which mainly expresses the fitness changing degree of the same individual in different environments. N^{es} represents the sentries' number set by the current algorithm, and its main function is to calculate the environmental changes' degree. After that, the environmental changes' degree should be compared with the fitness changing degree in the stable evolution state to be defined. Equation (7) shows the specific calculation.

$$ecs = \frac{\sum_{i=1}^{N^{es}} \left| \text{fit}^{\text{new}}(\text{es}_i) - \text{fit}^{\text{old}}(\text{es}_i) \right|}{N^{es}} \quad (6)$$

where er (evolution rate) is the average degree of the change of all the populations' fitness, which is calculated at the algorithm's each iteration. The adaptive initialization strategy retains the calculated maximum and minimum, and forms the upper and lower bounds ecr, as shown in Eq. (8).
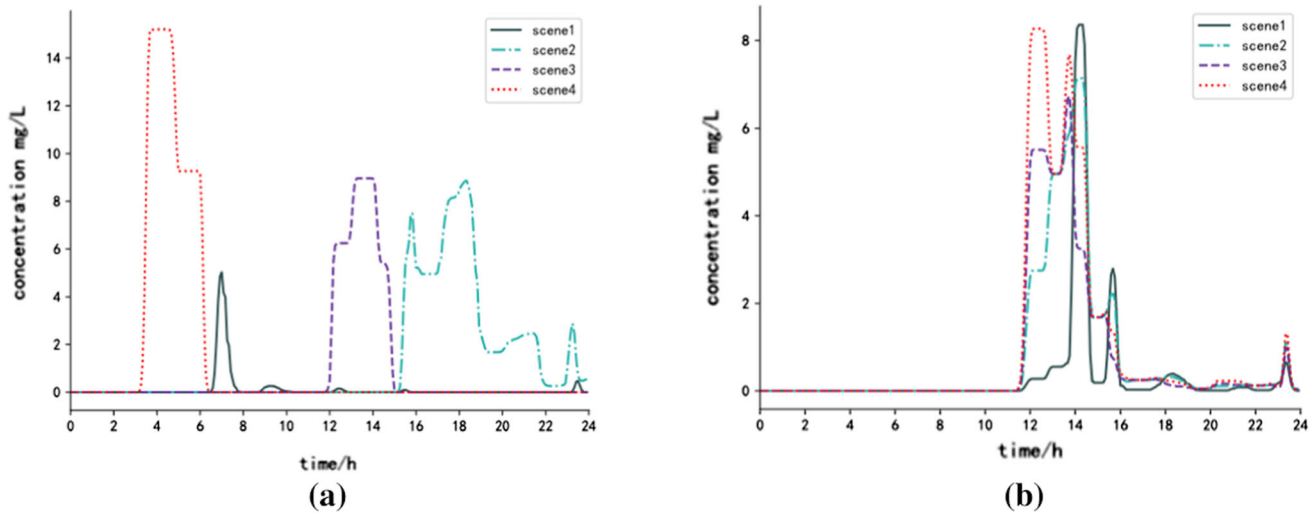
Fig. 5 Comparison of concentration of pollution source events

$$\text{er}_g^{\text{avg}} = \frac{1}{NP} \sum_{i=1}^{\text{sub}N} \sum_{j=1}^{\text{size}(i)} \left| \text{fit}_g\left(\text{ind}_{i,j}\right) - \text{fit}_{g-1}\left(\text{ind}_{i,j}\right) \right| \qquad (7)$$

$$\text{ecr} = \left[\min\left(\text{er}_g^{\text{avg}}\right), \ \max\left(\text{er}_g^{\text{avg}}\right) \right] \qquad (8)$$

When ecs is relatively large and exceeded the upper bound of ecr, the current environment changes dramatically. The high-quality individuals in the previous environment may not be suitable for the next environment; the population can initialize almost all individuals at random. Conversely, when the change degree of ecs is at the lower bound of ecr, the environmental change has almost no effect on the evolution. Its effect is less than a variation in the stable evolution process, so a large number of historical solutions can be used to continue the evolution. Equation (9) shows the specific execution. In general, the calculated *ecs* and *ecr* are used to adaptively adjust the number of the population's individual applying history solution N^{reserved}, and the individuals out of the proportions are still randomly initialized. When ecs exceeds the upper bound of ecr, number 1 in Eq. (9) means only the optimal solution of the previous environment is retained.

$$N^{\text{reserved}} = \begin{cases} NP, \text{ecs} < \min\left(\text{er}_g^{\text{avg}}\right) \\ \dfrac{\max\left(er_g^{\text{avg}}\right) - \text{ecs}}{\max(\text{er}_g^{\text{avg}}) - \min(\text{er}_g^{avg})} \cdot NP, \text{ others} \\ 1, \text{ecs} > \max\left(\text{er}_g^{\text{avg}}\right) \end{cases} \qquad (9)$$

3.2.4 Quality local search strategy

In actual water pollution events, the information such as the node where the pollution occurs, start, and duration of

the pollution is often more critical, which can guide relevant agencies to carry out response measures. At the same time, when the simulation software simulates the events, pollution events with different node positions, start, and duration often differ greatly in the function fitness. Conversely, when these three parameters are the same, changing the injection quality in each period only affects a relatively small part of the fitness. When the node location, start time, and duration are different, and the injection quality is the same, and the pollution concentration of the pollution source event shown in Fig. 5a fluctuates differently. However, in Fig. 5b, the first three characteristic information of the four pollution events is all {44,5,3}, and the injection quality is different, but their concentrations fluctuate in a very similar way. It also proved that the work is reasonable to select three characteristic information as the benchmark data for judging the similarity of nodes in the previous section. Therefore, from the perspective of judging whether the individual is converged, the algorithm should focus on searching the first three main pieces of information in the first stage, and then the refinement search, such as for quality. For finding more accurate information such as quality, the quality local search strategy (QLSS) in this section further searched for the quality of high-quality individuals in the population to improve the algorithm's convergence accuracy, based on judging the population's convergence.

After the algorithm reaches the maximum iterations, it performs the quality local search on the convergent population that has been judged in the historical information retention strategy. Local search adopts particle swarm optimization (PSO). In the algorithm's initialization stage, the quality of the converged individuals in the converged population is initialized, keeping the first three main

variables unchanged. Different main variables only generate one population. A new multi-population with different population numbers is generated in this way. Each population refines the quality of its population and outputs the optimal solution at the end of the algorithm. After the end of the local search algorithm, a new optimal pollution source consistent with the convergent populations' number is generated.

3.2.5 Similar peak penalty strategy

Although the population division scheme generated by the optimal subpopulation division strategy has reasonably balanced the distance within the population and the distance between the populations, multiple subpopulations evolving in the same direction are inevitable in the actual algorithm evolution process and finally located on the same peak. Corresponding to the pollution source problem that the optimal individuals of multiple populations are similar, the three main characteristics of the pollution source are the same. Therefore, the situation needed to be optimized, and the optimal solution is to merge similar populations into one population size, leaving some high-quality individuals in two other populations to solve the algorithm resources. For the population individual's reduction caused by fusion, new individuals are randomly generated again to supplement, and the optimal subpopulation division strategy is used again. Besides, the similar peak penalty strategy (SPPS) in the work is based on population convergence. There is no need to judge whether it is covered with other populations for populations without convergence, and the algorithm resources can be used more rationally simultaneously.

3.3 Steps of the algorithm

After analyzing the real-time water pollution traceability's characteristics, the traditional intelligent traceability algorithm is strengthened by a multi-strategy approach, so it can better cope with the environmental changes of intelligent real-time water pollution traceability. The process of the intelligent real-time traceability algorithm based on the dynamic multi-mode optimization is as follows.

Step 1 Initialize the individuals of the population randomly and implement the optimal subpopulation division strategy.

Step 1.1 Set SP_MAX_GEN as the maximum iterations of the optimal subpopulation division, 3 as the initial value of k, and start the iteration.

Step 1.2 Divide the population according to the current k value, and calculate *pdm* according to Eq. (3.5). Save the

minimal plan of pdm as $P_{\min(\text{pdm})}$, and k increase by 1 until reaching $k_{\max}$.

Step 1.3 Iterate to the maximum iterations SP_MAX_GEN, and let the initial population $P = P_{\min(\text{pdm})}$.

Step 2 t_0 is the moment that the sensor detected the pollutant for the first time, and the current time is $t_c = t_0 + \Delta t$. MAX_GEN is the maximum iterations for each time step Δt, and g is the current iterations.

Step 2.1 Let $g = g + 1$, and enter the genetic algorithm selection stage. Use the roulette method to select the parent individual according to the individual's fitness.

Step 2.2 In the cross-phase, generate the random number rand. When rand $< P_c$, cross the parental individual using the selection phase. The source position, start, and duration of the pollution source individual are coded by integers, and two-point crossover in integer coding is adopted. The quality injection part use SBX to simulate binary crossover to produce offspring individual ind_c.

Step 2.3 In the variation stage, when random number rand $< P_m$, the source position, start, and duration of the pollution source use single-point variation in integer coding, and inject the quality real number single-point variation to produce new offspring individual ind_m.

Step 2.4 In the update phase, add the new offspring individuals to the current subpopulation P_i, and discard the individuals in the current population with the worst fitness.

Step 2.5 Determine whether the current population converges according to Eq. (4). If it converges, add the current population P_i to the convergent population set and add the optimal solution to the historical archive δ_i. The redundancy elimination mechanism is used to filter the repeated archived solutions. For the set φ_c, judge whether there is a similar peak evolution, and retain only one population size of the similar peak individual according to the similar peak penalty strategy.

Step 2.6 Calculate er_g^{avg} according to Eq. (7), and update the upper and lower bounds of ecr. Determine whether the current iteration g exceeds MAX_GEN, and enter Step3 if it exceeds.

Step 3 Use the quality local search strategy to perform a quality search on convergent population φ_c. Execute Algorithm 3.1 to get set LS_Best to update the historical archive δ.

Step 4 The multi-mode algorithm generates the multiple subpopulation's optimal solution Gbest$_i$, and output the individual with the least fitness as the real pollution source located at the current time step. $t_c = t_c + \Delta t$ is the current time and judge whether t_c is greater than the maximum positioning time T; if not, enter Step5. The algorithm terminates when it exceeds.

Step 5 Adaptive initial strategy restarts the population. Use Eq. (9) to calculate N^{reserved}. The parameters can determine how many individuals in the historical archive δ_i are used, and the remaining individuals are initialized randomly.

4 Experimental results and analysis

The multi-strategy improved algorithms were compared to explore the impact of different improved strategies on the algorithm's performance to verify the effect of the new algorithm proposed in the work and analyze the algorithm's performance in different pipe networks. Then, the new algorithm was compared with the traditional dynamic optimal algorithm to verify the accuracy and real-time performance of the new algorithm.

4.1 Experimental parameter settings

This section mainly experiments and compares the pipe network Net3_Rossman2000, BWSN1_Ostfeld2008, and

ky5_Jolly2013 [26]. Three pipe networks separately include different node numbers and scales and have the same hydraulic water quality steps in the parameter settings to ignore the parameters' influence. Besides, the settings of the sensors' number change as the nodes increases due to the limitation of the pipe network's size. Table 1 shows the specific parameters related to the pipe network.

Table 2 shows the new algorithm's basic parameters and the basic parameters needed to improve the multi-strategy. In the optimal subpopulation division strategy, the maximum divided populations per generation $k_{\max}$ is set to + 10 in the three pipe networks in the work, which need further adjustments after the expansion of the pipe network. The crossover and variation probability of the genetic algorithm is adjusted to 80 and 90% to improve the population's convergence speed. The sentinel individuals' number N^{es} is generally set to be close to the population size, and pop_size represents the individual number generated when randomly initializing the population. In addition, repeated individuals' threshold rr is 0.6 during judging whether the population converges. Under the current population size, the individual's number in a single
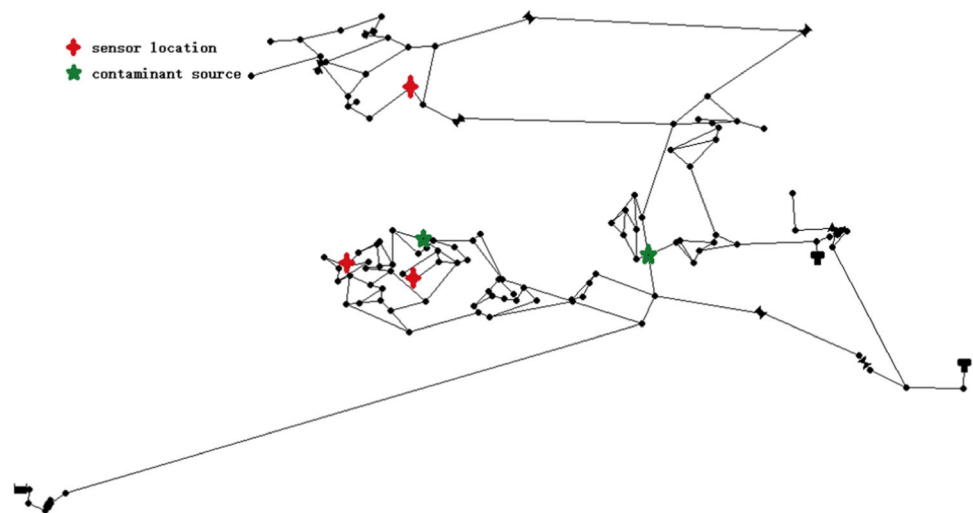
Table 1 Network parameters

Network parameters	Net3_Rossman2000	BWSN1_Ostfeld2008	ky5_Jolly2013
Node number	97	129	430
Pipe network number	119	178	507
Reservoir	2	4	4
Pool	3	3	3
Hydraulic step	1 h	1 h	1 h
Water quality step	5 min	5 min	5 min
Sensor position distribution	37, 61	10, 83, 100	6, 22, 30, 34, 40, 42, 43, 76, 80, 87

Table 2 Parameters of the new algorithm proposed in the work

Parameter	Parameter description	Parameter size
$k_{\max}$	Maximum clustering k-value	10
MAX_CLUSTER_GEN	Clustering iterations	3
pop_size	Population size	100
subpop_size	Individual boundaries of subpopulations	[5, 20]
P_c	Crossover probability	80%
P_m	Variation probability	90%
T^s	Simulation time	5 h
Δt	Timestep	10 min
rr	Population individual repetition threshold	0.6
N^{es}	Sentry individual detecting the environmental changes	80
w	Search weights of quality PSO	0.8
c_1 and c_2	Search factors of quality PSO	1.8

Fig. 6 BWSN1_Ostfeld2008
Pipe network topology

Table 3 pollution scenario of BWSN1_Ostfeld2008 pipe network

Parameter Scenario	Node number	Start time of the pollution injection	Duration (h)	Injection contamination concentration(mg/L)
Pollution scenario 1	44	2	4	300, 180, 240, and 180
Pollution scenario 2	92	2	4	300, 180, 240, and 180

Table 4 Error data of BWSN1_Ostfeld2008 pipe network

Algorithm	Pollution scenario 1				Pollution scenario 2			
	Start time error	Duration error	Injection quality error	acc	Start time error	Duration error	Injection quality error	acc
NGA	0.2853	1.0124	561.682	0.5548	0.2153	0.7281	346.4867	0.5629
GA-SBX	0.2523	1.0528	592.9434	0.7967	0.1847	0.727	330.2962	0.5967
GA-OSPS	0.2227	0.9992	564.7464	0.795	0.2204	0.6709	352.2975	0.5790
GA-HIRS	0.0164	1.0796	469.9973	0.725	0.0574	0.6389	295.9846	0.633
GA-ARS	0.0885	1.0818	486.1	0.81	0.072	0.7068	288.6287	0.725
GA-QLSS	0.2698	1.0741	510.3191	0.6758	0.1947	0.7414	316.7342	0.6112
New algorithm	0.0283	0.999	443.1	0.8612	0.0844	0.5622	282.8936	0.6887

subpopulation is small, and it increases when the number of individuals in the subpopulation increase.

Experimental environment: Personal PC, Memory 16G, CPU corei7-8750H, and Windows10 operating system.

4.2 Algorithm's performance index

In terms of the evaluation criteria for the algorithm performance, different references have designed different evaluation criteria for pollution traceability, most of which are based on absolute or relative error values. The work carried out an optimized design on this basis to reflect the performance characteristics of the water pollution traceability algorithm, which mainly includes the following points:

The first is the accuracy of positioning. The pollution source feature information in the work mainly includes the source location, start and duration, and pollution injection quality array. For the problem of pipe network positioning, the most decisive information is which node the pollution

Table 5 Real-time data of BWSN1_Ostfeld2008 pipe network

Algorithm	Pollution scenario 1			Pollution scenario 2		
	Average earliest location time	Earliest location time interval	Average earliest location algebra	Average earliest location time	Earliest location time interval	Average earliest location algebra
NGA	26	[10, 90]	8.15	21.5	[10, 50]	10.87
GA-SBX	14.5	[10, 30]	9.984	17	[10, 40]	9.33
GA-OSPS	19	[10, 70]	8.1	25	[10, 60]	13.141
GA-HIRS	15	[10, 50]	2.72	56.5	[10, 180]	2.55
GA-ARS	20	[10, 50]	2.33	28	[10, 120]	3.25
GA-QLSS	25.5	[10, 90]	9.32	26.5	[10, 50]	9.46
New algorithm	14	[10, 30]	4.13	26.5	[10, 130]	4.04

source is generated on. Then, the work sets whether the algorithm locates the real pollution source node as an indicator, and establishes the accuracy rate acc (accuracy) as the evaluation standard. In the entire simulation duration, for the optimal result obtained by the algorithm within each Δt, if the node of the optimal individual is the real pollution source node, then the node is hit once at this step. Equation (10) shows the specific calculation.

$$\text{acc} = \frac{T^s \sum_{i=1}^{N^{\text{run}}} N^{\text{hit}}}{\Delta t \cdot N^{\text{run}}} \tag{10}$$

where N^{hit} represents the total number of times that the real pollution source node is located in an independent operation. Dividing T^s and Δt can get the number of evolution steps in a simulation time, and it also corresponds to the number of environmental changes. N^{run} is the number of independent runs of the algorithm experiment, which is generally set to 20 times in the work. Besides, after locating the true source point, the remaining three pieces of information of the pollution source characteristics are evaluated. The average absolute error value under all simulation steps is used to represent the accuracy of the algorithm for solving different information. In the case of different injection quality dimensions caused by different durations, the unequal part is calculated as 0 mg injection.

As to reflecting the real-time nature of the algorithm for traceability of pollution sources, two indicators are set. The first is the earliest time for the algorithm to locate the real pollution source. After the evolution of each time step is over, if the optimal solution obtained by the algorithm in this evolution is the real pollution source node, then the difference between the current time and the time when the pollution is first discovered is the time when the real pollution source is located for the first time. After the evolution of each time step, if the optimal solution obtained by the algorithm in this evolution is the real pollution source

node, then the difference between the current time and the time when the pollution first discovered is the time when the real pollution source is located for the first time. It represents how long the algorithm can find the true pollution source location after the pollution occurs. Meanwhile, the smaller the simulation time step, the fewer simulation data that can be used for evolution, which reflects the algorithm's optimization ability and the ability to cope with the non-uniqueness of the real-time location of pollution sources. Within each step, the average algebra of the algorithm to locate the true pollution source for the first time is also used as an indicator. The fewer the iterations, the better the algorithm's ability to converge.

4.3 Impact of different strategies on the algorithm

Different improvement strategies in the dynamic multimode optimization algorithm have different influences on the algorithm. In this section, different experimental scenarios and control groups are designed to test the algorithm's effect through combining different strategies. The experimental pipe network is the BWSN1_Ostfeld2008 water supply pipe network. Table 1 shows the characteristic parameters, and Fig. 6 shows the topology. There are two groups of pollution events in the experiment, and Table 3 presents the characteristic information.

Experiments were conducted according to different strategies and genetic algorithm combinations to explore different effects on the positioning results. The control group mainly included: (1) the normal genetic algorithm (NGA), (2) the improved genetic algorithm (GA-SBX), (3) the improved genetic algorithm and optimal subpopulation strategy (GA-OSPS); (4) the improved genetic algorithm and historical information retention strategy (GA-HIRS), (5) the improved genetic algorithm and adaptive
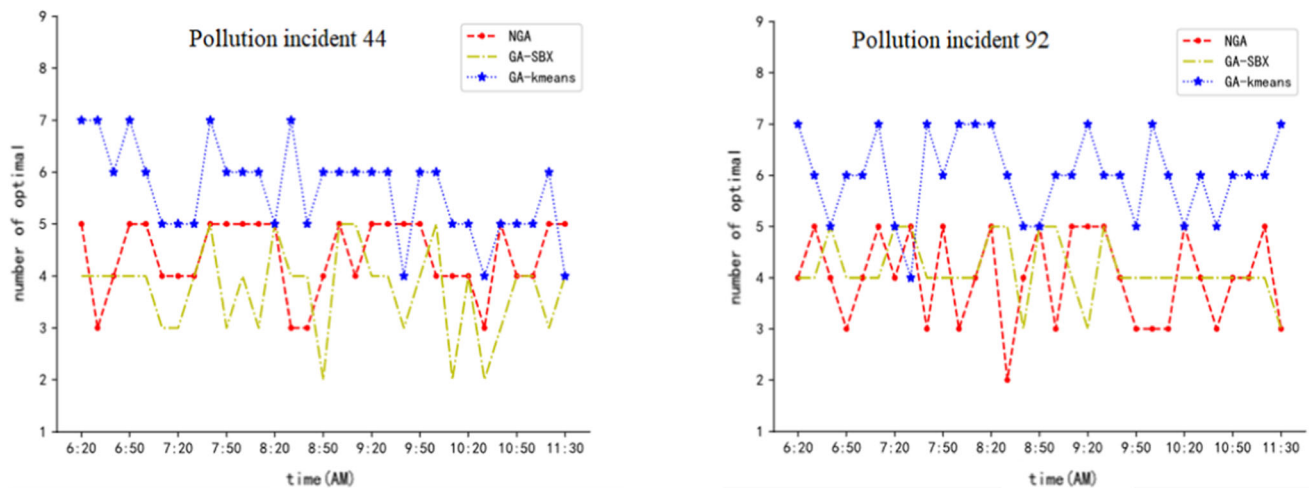
Fig. 7 Individuals' optimal number of BWSN1_Ostfeld2008

initialization strategy (GA-ARS), (6) the improved genetic algorithm and quality local search strategy (GA-QLSS), and (7) the multi-strategy dynamic multi-mode optimization algorithm (MDMMOA). Tables 4 and 5 show the experimental results.

Table 4 shows the error value of the pollution source characteristic information and the positioning accuracy of the genetic algorithm with different strategy combinations, which mainly reflect the positioning accuracy of different strategy combinations. GA-SBX is an improved algorithm based on the traditional genetic algorithm combined with simulated binary crossover and hybrid coding in the work. Compared with the experimental data of traditional NGA, the improved genetic algorithm can reduce the error and improve the location accuracy in pollution scenario 1. After adding the optimal subpopulation division strategy, the algorithm can select the parent individuals more reasonably when the algorithm crosses due to a more reasonable population division. The algorithm's positioning accuracy increases, but the error was not different when compared with NGA and GA-SBX. Both GA-ARS and GA-HIRS respond to the dynamic environment's changes by reusing the historical optimal individuals. The difference is that GA-HIRS adopts a simpler method of reusing all the optimal historical information individuals without adapting to the extent where dynamic changes occur. In

Table 4, GA-HIRS performs well in three characteristic information's error values. The excellent individuals in each different period are completely used when the algorithm is optimized. As a result, the algorithm has converged when the population is initialized and tends to use a single optimization in the overall process. Therefore, in the pollution scenario with a low degree of uniqueness and non-uniqueness of pollution sources, GA-HIRS can obtain more iterations to refine the characteristic information due to its fast convergence characteristics and finally reduce the error. However, note that if the non-uniqueness degree of the pollution scenario is too high, and the algorithm locates the error source with less fitness at the initial stage, GA-HIRS will not find the real source again as the simulation time increases. Finally, it is completely impossible to locate the real pollution source, which has been confirmed in part of the experimental data. Besides, the algorithm's positioning accuracy is not high. When GA-QLSS is used, the error value of the pollution-source characteristic information is reduced compared with NGA because the particle swarm algorithm is used to further evolve the results of the algorithm. The error value of GA-ARS is smaller than that of GA-QLSS, due to the algorithm's rapid convergence after dynamic changes and relatively more iterations for evolution refinement. Finally, the new

Table 6 Pollution scenarios of Net3_Rossman2000 pipe network

Parameter Scenario	Node number	Start time of the pollution injection	Duration (h)	Injection contamination concentration (mg/L)
Pollution scenario 1	16	2	4	300, 180, 240, 180
Pollution scenario 2	86	2	4	300, 180, 240, 180

Fig. 8 Net3_Rossman2000 Pipe Network
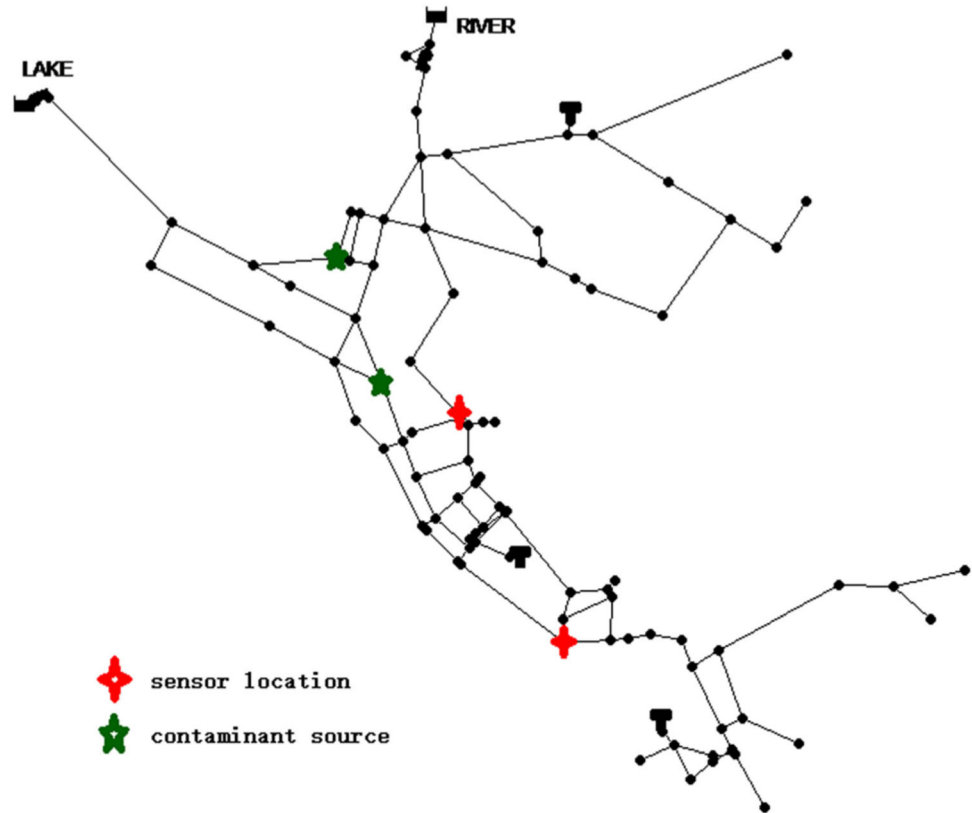
Table 7 Error data of the Net3_Rossman2000 pipe network

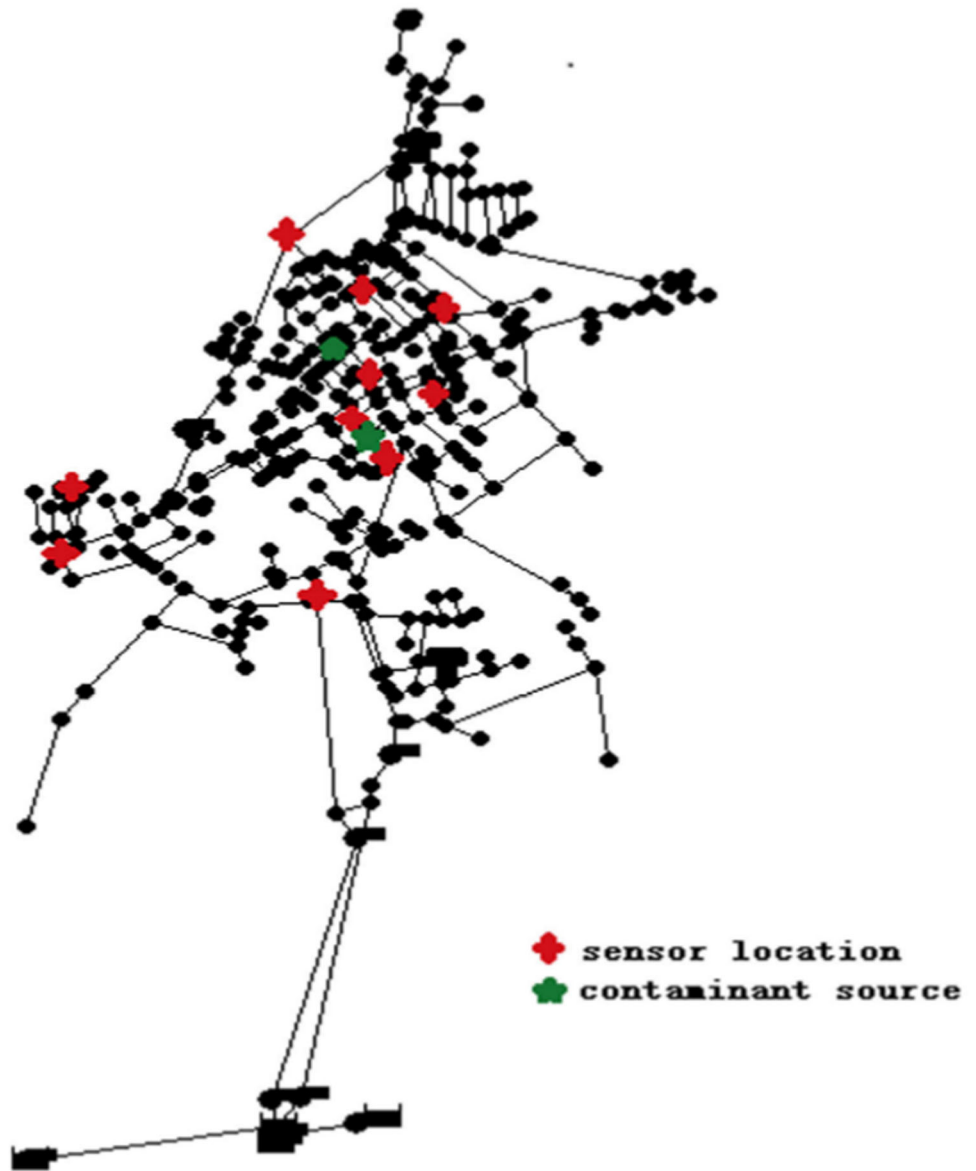| Algorithm	Pollution scenario 1				Pollution scenario 2			
	Start-time error	Duration error	Injection quality error	acc	Start-time error	Duration error	Injection quality error	acc
DOA	0.27	0.69	374.5	0.459	0.21	0.78	365.8	0.556
New algorithm	0.04	0.54	400.4	0.96	0.01	0.74	300.8	0.745

algorithm proposed in the work performs well in both pollution scenarios.

The data in Table 5 are whether the algorithm can use less time to locate the real solution in real-time evolution, representing the real-time optimization and improvement in different strategies. In the two scenarios, GA-SBX has more excellent optimization capabilities due to its more reasonable crossover method. Limited data are used to find the real source of pollution in the case of a small simulation step size and act as the algorithm's optimal individual. In terms of the average number of earliest positioning iterations, two algorithms using the historical solution set strategy have achieved significant results. It is because the algorithm can converge due to the historical optimal individual during restart to locate quickly. After MDMMOA

adds a redundant elimination strategy, a large number of locally optimal and similar individuals due to non-uniqueness are eliminated. Its algorithm's positioning speed is balanced compared with GA-ARS and GA-HIRS, and the convergence speed is relatively reduced; however, a better positioning effect is received.

GA-OSPS has no significant effect on positioning accuracy and real-time performance. However, multi-population can provide multiple optimal solutions with comparable quality during solving the pollution sources' non-uniqueness, because the sensor concentration produced by non-homogeneous individuals is the same. Meanwhile, the algorithm always retains a large number of optimal solutions because of the similar peak penalty strategy. This ability to locate multi-mode solutions is very important for

Fig. 9 ky5_Jolly2013 pipe network

Table 8 Pollution scenarios of the ky5_Jolly2013 pipe network

Parameter Scenario	Node number	Start time of the pollution injection	Duration (h)	Injection contamination concentration (mg/L)
Pollution scenario 1	31	2	4	300, 180, 240, and 180
Pollution scenario 2	155	2	4	300, 180, 240, 180

solving the non-uniqueness problem. Only when the number of multiple solutions is sufficient can the optimal solution individual be judged based on the fitness comparison, thus optimally locating the real pollution source in multiple parts.

Figure 7 reflects the algorithm's optimal solutions number of different combination strategies. Using an optimal subpopulation division strategy can guarantee multiple locally optimal solutions and improve the positioning accuracy.

Table 9 Error data of ky5_Jolly2013 pipe network

Algorithm	Pollution scenario 1				Pollution scenario 2			
	Start time error	Duration error	Injection quality error	acc	Start time error	Duration error	Injection quality error	acc
DOA	0.361	0.734	539.5	0.67	0.249	1.175	556.85	0.37
New algorithm	**0.07**	**0.914**	**400.4**	**0.98**	**0.017**	**1.01**	**533.95**	**0.63**

4.4 Analysis of algorithms performance under different pipe networks

Two other small and medium-scale pipe networks were used for experimental comparison to further compare the performance of the dynamic multi-mode optimization algorithm and verify the algorithm's real-time traceability.

(1) Net3_Rossman2000 pipe networks

This group of experiments adopted Net3_Rossman2000 water supply pipe network, with relevant parameters shown in Table 6, and the topology shown in Fig. 8. The work conducted experiments on two pollution scenarios on the Net3_Rossman2000 water supply pipe network, where the injected pollution concentration was continuously injected in the units of 1 h.

For the two pollution scenarios shown in Fig. 8, the traditional dynamic optimal algorithm (DOA) was adopted for experimental comparison with the new algorithm. The pipe network adopted the experiment's average data of 5 h simulation time and 20 independently running times (see Table 7).

Comparing the experimental data, the algorithm's accuracy of real-time positioning has greatly improved after using the new algorithm, and the algorithm has a high probability of obtaining the optimal solution node in different simulation time steps. In addition, the start time error was zero, and the duration error and the injection quality error were not much different from the DOA in terms of the solution error. One of the reasons was that the real pollution source nodes' number located by DOA was relatively small. In this case, the optimal solution's quality was relatively good, so the average error obtained would be small. Other unshown data were that the injection quality error of the optimal solution could reduce to zero after the new algorithm's simulation time increased. However, DOA could not converge to this accuracy, showing that the new algorithm had a relatively good ability to explore the optimal solution.

(2) ky5_Jolly2013 Pipe Network

This group of experiments adopted a relatively enlarged ky5_Jolly2013 pipe network, with relevant parameters shown in Table 1 and the topology shown in Fig. 9. The work still conducted experiments on two pollution scenarios on this pipe network to explore the new algorithm's ability of real-time pollution traceability in the medium-sized pipe network. Table 8 shows the pollution scenarios' parameters.

Table 9 shows that two pollution scenarios are simulated experiments. As the pipe network scale has increased compared with the pipe network used above, the sensor layout has also increased to 10, which directly increases the time of calling the simulation software EPANET for a single simulation. Therefore, the simulation duration of this experiment is set to 3 h to reduce the experiment's overall time. Figure 9 shows the specific experimental data.

The new algorithm's positioning accuracy has been greatly improved compared with the common dynamic optimal algorithm in the medium-sized pipe network, and it has excellent positioning and optimization capabilities. There is also a significant improvement in terms of error; the error is negligible for the starting time. Among the unshown experimental data, the new algorithm can accurately locate nodes using 10 to 20 min of the sensor concentration data on average, and no more than 60 min at the latest, which is very important for the ability to cope with real-time missing positioning data.

5 Conclusions

Drinking water safety, related to people's vital interests, is a key concern of the whole society. Real-time water pollution traceability is a powerful guarantee for urban drinking water safety. Considering the challenges of the real-time water pollution traceability, the work elaborated and formulated combining the problem's dynamic optimal characteristics, and proposed a dynamic multi-mode optimal algorithm based on it, which improved a variety of

strategies to enhance the algorithm's real-time traceability. In terms of the non-unique of water pollution traceability, the optimal subpopulation division strategy was adopted to improve the algorithm's optimization ability, thus dividing and conquering for solving and increasing the optimal solutions including the real pollution source. In addition, an adaptive initialization strategy was introduced to reuse historical information to cope with the environmental change caused by the simulation duration parameters' variation in real-time simulation. Besides, the strategy can guide the algorithm to restart according to the environmental change degree. Finally, the quality of the pollution-source characteristic information was also refined locally to improve the algorithm's optimal solution quality. Through the new algorithm's combination experiment, the effectiveness of different improvement strategies for improving the algorithm's performance was proved. Meanwhile, another part of the experiment showed that the new algorithm had achieved good improvement effects for small and medium-sized pipe networks.

Declarations

Conflicts of interest The authors declare there are no conflicts of interest.

References

1. Rasekh A (2012) Risk analysis and adaptive response planning for water distribution systems contamination emergency management. Texas A&M University, Texas
2. Zierolf ML, Polycarpou MM, Uber JG (1998) Development and autocalibration of an input-output model of chlorine transport in drinking water distribution systems. IEEE Trans Control Syst Technol 6(4):543–553
3. Shang F, Uber JG, Polycarpou MM (2002) Particle backtracking algorithm for water distribution system analysis. J Environ Eng 128(5):441–450
4. Guan J, Aral MM, Maslia ML et al (2006) Identification of contaminant sources in water distribution systems using simulation–optimization method: case study. J Water Resour Plan Manag 132(4):252–262
5. Liu L, Ranjithan SR, Mahinthakumar G (2011) Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure. J Water Resour Plan Manag 137(2):183–192
6. Hu C, Zhao J, Yan X et al (2015) A MapReduce based parallel niche genetic algorithm for contaminant source identification in water distribution network. Ad Hoc Netw 35:116–126
7. Seth A, Klise KA, Siirola JD et al (2016) Testing contamination source identification methods for water distribution networks. J Water Resour Plan Manag 142(4):04016001
8. Wang C, Zhou S (2017) Contamination source identification based on sequential Bayesian approach for water distribution network with stochastic demands. IISE Trans 49(9):899–910
9. Yan X, Sun J, Hu C (2017) Research on contaminant sources identification of uncertainty water demand using genetic algorithm. Clust Comput 20(2):1007–1016
10. Yan X, Li T, Hu C et al (2019) Real-time localization of pollution source for urban water supply network in emergencies. Clust Comput 22(3):5941–5954
11. Torres JM, Brumbelow K, Guikema SD (2009) Risk classification and uncertainty propagation for virtual water distribution systems. Reliab Eng Syst Saf 94(8):1259–1273
12. Preis AO (2011) Hydraulic uncertainty inclusion in water distribution systems contamination source identification. Urban Water J 8:267–277
13. Vankayala P, Sankarasubramanian A, Ranjithan SR, Mahinthakumar G (2009) Contaminant source identification in water distribution networks under conditions of demand uncertainty. Environ. Forensics 10:253–263
14. Yan X, Hu C, Sheng VS (2020) Data-driven pollution source location algorithm in water quality monitoring sensor networks. Int J Bio-Inspired Comput 15(3):171–180
15. Yan X, Gong J, Wu Q (2021) Pollution source intelligent location algorithm in water quality sensor networks. Neural Comput Appl 33(1):209–222
16. Yan X, Yang K, Hu C et al (2018) Pollution source positioning in a water supply network based on expensive optimization. Desalin Water Treat 110:308–318
17. Gong J, Yan X, Hu C et al (2019) Collaborative based pollution sources identification algorithm in water supply sensor networks. Desalin Water Treat 168:123–135
18. Hu Y, Yan X (2021) Neural network-assisted expensive optimization algorithm for pollution source rapid positioning of drinking-water. Int J Bio-Inspired Comput 17(4):227–235
19. Hu C, Dai L, Yan X et al (2020) Modified NSGA-III for sensor placement in water distribution system. Inf Sci 509:488–500
20. Hu C, Shu X, Yan X et al (2020) Inline wireless mobile sensors and fog nodes placement for leakage detection in water distribution systems. Softw-Pract Exp 50(7):1152–1167
21. Hu C, Yan X, Gong W et al (2020) Multi-objective based scheduling algorithm for sudden drinking water contamination incident. Swarm Evol Comput 55:1–11
22. Yan X, Zhao J, Hu C et al (2019) Multimodal optimization problem in contamination source determination of water supply networks. Swarm Evol Comput 47:66–71
23. Li C, Yang S, Yang M (2014) An adaptive multi-swarm optimizer for dynamic optimization problems. Evol Comput 22(4):559–594
24. Li C, Nguyen TT, Yang M et al (2016) An adaptive multipopulation framework for locating and tracking multiple optimal. IEEE Trans Evol Comput 20(4):590–605
25. Das S, Mandal A, Mukherjee R (2014) An adaptive differential evolution algorithm for global optimization in dynamic environments. IEEE Trans Cybern 44(6):966–978
26. Ostfeld A, Uber JG, Salomons E et al (2008) The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms. J Water Resour Plan Manag 134(6):556–568

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.