A deterministic approach for optimization of booster disinfection placement and operation for a water distribution system in Beijing

Fanlin Meng, Shuming Liu, Avi Ostfeld, Chao Chen and Alejandra Burchard-Levine

ABSTRACT

Previous studies on booster disinfection optimization were commonly based on 'blank networks', neglecting the impact of existing disinfection facilities, which could result in misleading solutions. To overcome this limitation, a method, which incorporates the existing disinfection facilities, is developed and demonstrated in this study. A particle backtracking algorithm, which traces the upstream pathways of the disinfection insufficiency nodes, is employed to narrow down the potential positions for booster stations. Deterministic optimization results are then efficiently yielded by the introduction of a 'coverage matrix'. The proposed method is applied to a real life water distribution system in Beijing, China. Results show the methodology effectiveness in optimizing booster disinfection placement and operation for real life water distribution systems. For the explored case study, results suggest that adding a booster disinfection station at 0.1% of the nodes of the system can satisfy chlorine residual at about 97.5% of all nodes.

Key words | booster disinfection, deterministic approach, optimization, particle backtracking algorithm, water distribution systems

Fanlin Meng
Shuming Liu (corresponding author)
Chao Chen
School of Environment,
Tsinghua University,
Beijing, 100084,
China
E-mail: *shumingliu@tsinghua.edu.cn*

Avi Ostfeld
Faculty of Civil and Environmental Engineering,
Technion—Israel Institute of Technology,
Haifa 32000,
Israel

Alejandra Burchard-Levine
Tsinghua-Veolia Joint Research Center,
Tsinghua University,
Beijing, 100084,
China

INTRODUCTION

A large percentage of water resources in China have been polluted under the pressure of economic and population growth. The pollution not only increases the amount of hazardous chemicals in water, but also promotes the growth of microorganisms in water distribution systems (Rossman *et al.* 1994; Courtis *et al.* 2009). Disinfection is used at water plants to ensure water safety of the network. Due to the large size of the water distribution systems, and physical, chemical and biological reactions in pipes, disinfectants usually run out before reaching the far ends of the network (Vasconcelos *et al.* 1997; Powell *et al.* 2000; Hallam *et al.* 2002, 2003; Boccelli *et al.* 2003). However, it is impractical to add too much disinfectant at the water plant, because its concentration would exceed the maximum tolerant limit at the network entrance, along with bad taste and excess disinfection by-products (DBPs) at the far ends of the system. Booster disinfection could to some extent alleviate the problem by injecting chlorine at some intermediate locations. This can provide consumers with safe water with enough disinfectant residuals (Tryby *et al.* 1999).

Research has been conducted about the optimization of booster disinfection station locations and related injection rates. Boccelli *et al.* (1998) was the first to apply the linear superposition theory and linear programming models for efficient waste load allocation in river basins to scheduling of booster disinfection doses in water distribution systems. Given a number of booster stations at specified locations, the booster injection scheduling was optimized to minimize the total injection mass of chlorine subject to maintaining sufficient residual concentrations at monitored nodes. By assuming periodicity of mass injection and network hydraulics, the infinite time-varying problem was converted to a

doi: 10.2166/hydro.2013.149

solvable finite-time optimal scheduling problem with boundaries. Tryby *et al.* (2002) incorporated booster station locations as decision variables, employing a mixed-integer linear programming (MILP) and branch and bound technique for both the optimization of booster station locations and injection scheduling. To satisfy the concentration constraints, Tryby *et al.* (2002) removed some demand nodes from the monitoring constraint set using a pruning strategy. This is because some nodes, especially those at the far ends of the network, could not be sufficiently disinfected even with a large number of booster stations. Prasad *et al.* (2004) utilized a multi-objective optimization model to avoid node pruning. By switching the concentration constraint to an optimization objective, which is to maximize the percentage of safe water with disinfectant concentration above a minimum concentration level, all nodes could be included in the evaluation. Besides supplying more biologically safe water, the booster disinfection can also be optimized to reduce the production of DBPs (Cozzolino *et al.* 2005) or to respond to contamination in the water distribution systems (Isovitsch & VanBriesen 2009). The operation of pumps and valves could be conjunctively optimized as well for better disinfection (Ostfeld & Salomons 2006; Kang & Lansey 2010).

Although the methodology of optimizing booster disinfection stations has been improving, previous research was commonly based on 'blank networks', which assumed that the systems were not disinfected by any facilities prior to optimization (Boccelli *et al.* 1998; Tryby *et al.* 2002; Prasad *et al.* 2004). However, in reality, an important task for any water utility, and in China, is to improve the existing disinfection plan to ensure that the water quality at end users will meet the national regulations. Therefore, a model based on existing disinfection facilities that could optimize extra booster disinfection stations is essential for water quality management. Meanwhile, for most of the previous research, the locations of the booster stations have been optimized from a small sized 'candidate pool' of potential locations selected according to engineering judgment. Although it has greatly eased the computational burden, the selected set by experience might miss the optimal solution. Besides, other simplifying approaches, e.g. node pruning and stochastic algorithm, further reduce the chance of obtaining the optimal result.

The method proposed in this study is based on the current disinfection situation of the network. With the aim of minimizing the disinfectant insufficiency situations, the candidate set of booster stations is, instead of subjectively selected, narrowed down to upstream nodes of the disinfection weak points, which are yielded through the hydraulic calculation by the Particle Backtracking Algorithm (PBA). The optimization of the location and injection scheduling of the booster disinfection stations based on the reduced set of 'candidate pool', is further accelerated by the introduction of a 'coverage matrix'. The matrix improves optimization efficiency in four ways: (a) constraints of the optimization problem are incorporated in the calculation of the matrix, thus the optimization process of booster station locations is simplified to mathematical operations of the matrix; (b) optimal injection rates for each locating option could be attained in the computation process of the 'coverage matrix', and thus require no further optimization; (c) the matrix presents clear hydraulic connections between potential booster stations and disinfectant insufficient nodes, and among the booster stations, which could further decrease the complexity of the problem; and (d) water quality values, which are key elements in determining component values in the matrix, are calculated with linear superposition theory, in place of time-consuming water quality simulations in software (e.g. EPANET). The following sections first briefly introduce the PBA (detailed information could be referred to in Shang *et al.* (2002)). After that the proposed methodology is presented, and its effectiveness is demonstrated through an application on a real life water distribution system in Beijing, China.

Particle backtracking algorithm

Generally, traditional models for the simulation of the dynamic transport and decay of disinfectant in water distribution systems can be divided into two categories: Eulerian models and Lagrangian models. The two types of models could provide output information for all nodes over time, but are not able to disclose the internal details of the links between input and output. Zierolf *et al.* (1998) developed the first Input/Output (I/O) model which relates chlorine injected at treatment stations to concentrations at nodes. The algorithm tracks the travel of water parcels in networks

in reverse time and direction. It could find all the paths from output to input and their corresponding time, and has been used to estimate pipe chlorine reaction rates off-line. However, such an algorithm could only model chlorine transport in networks without storage tanks. Shang *et al.* (2002) extended the work of Zierolf *et al.* (1998) in the PBA, by considering the existence of tanks and allowing multiple water sources and water quality inputs.

Shang *et al.* (2002) also developed a library *UpstreamNode* for upstream tracing analysis using functions in the PBA. The analysis could yield water quality impact coefficients of all nodes in the network to the observed output node. The parameter reflects water dilutions and reactions on the way from any designated node to the observed output node. If the water quality impact coefficient $(s_j^k)_i$ is not zero, it indicates part of the water flow to output node j at time t $(t \in (k-1)\delta t_s, k\delta t_s]$, δt_s is the time step for impact coefficients integration, which is 1 h) passed through node i at time $t - t_{i\to j}^k$ (where $t_{i\to j}^k$ is the flow travel time from node i to node j). Thus continuous disinfectant injection at node i could improve water quality condition at node j at least at time interval k. However, if $(s_j^k)_i$ is zero, it implies node i is not an upstream node of node j at time step k, or it is far too upstream to have obvious water quality impact on node j, in which case it is not labeled as an upstream node either in this study. This method could find all upstream nodes in different flow paths, and is used in the presented study to trace the upstream nodes of disinfection insufficient nodes in the designated time interval.

METHODOLOGY

Problem statement

Through upstream backtracking analysis, the potential location set for booster stations could be reduced by PBA. Even so, the computational work could still be intensive. This can be illustrated from a simplest case: establishing only one booster disinfection station. As can be anticipated, the station should be situated upstream enough, so that the disinfectant injected could reach as many nodes in need as possible. But if the position is too upstream, some nodes which could be offered sufficient disinfectant by a less upstream node might not receive enough disinfectant anymore due to the increased distance for disinfectant decay. The injection rate is in a similar situation, where too low injection might not satisfy enough nodes to its best extent, while too high a value may breach the upper concentration limit at some nodes. If the examined scope extends from one monitoring interval to several intervals, it can be observed that the disinfectant insufficient nodes would be changing at different intervals, along with the upstream nodes. Thus to set a booster station that could solve the most disinfection inadequacy problems at all times is a complex issue. The problem becomes further more intricate if more than one booster station is to be built up. As their influence ranges may overlap with each other, coordination among each station is required to achieve the best outcome.

Therefore, although PBA could reduce the location solution space, massive calculations are still needed when location(s), injection rate(s) and constraints are considered together. Stochastic optimization strategies (such as Genetic Algorithm) could be adopted to achieve a satisfactory result in a limited time, yet it cannot guarantee a best result. Therefore, a method is proposed in this paper that could yield a deterministic optimal result within reasonable time. Before the methodology is presented, some key concepts used in this paper are first explained as follows.

Coverage, target cases and candidate pool

The term coverage has been commonly used in the field of water quality monitoring and sensor placement in water distribution systems (Lee & Deininger 1992; Kumar *et al.* 1997), where it denotes the possibility to infer water quality at upstream nodes based on water quality measurements at downstream locations. The coverage defined in this study is quite similar, where the monitoring of water quality at a downstream node is replaced by disinfectant injection at an upstream node, and the inference of water quality at upstream nodes by the level of meeting a disinfectant residual demand at a downstream node. As amelioration of water quality at one node should not be achieved by violation of the upper concentration limit at other nodes, disinfectant residual levels at other nodes are also considered in the definition of coverage. The determination of 'coverage', along with two other concepts ('target cases'

and 'candidate pool'), is illustrated by the example in Table 1.

In this example, there are nine nodes in the network, the disinfectant concentration values of which are monitored at each hour (although the water quality simulation time step δt_q is 5 minutes, the monitoring time step δt_m is 1 h, which is the same as the case study in the paper). The triangle (Δ) in the cell indicates that the residual concentration at the related node and monitoring time is below the lower limit required. In this paper, we refer to this situation as a 'target case'. Hence in this example, there are five 'target cases' in total.

The aim of booster disinfection is to reduce the number of 'target cases'. PBA is employed to first find the upstream nodes of the 'target node' at 'target time interval' of each 'target case'. Results are shown in Table 1 as the dark shading cells. Note that the target nodes are also included as the 'upstream nodes', as injection at their own places could certainly improve the disinfection situation. The upstream nodes of each target case are combined into a set, which is named as the 'candidate pool' for booster station locations. In this case, there are six components in the pool, which are node 1, node 2, node 3, node 5, node 6, and node 9. The calculation of the 'candidate pool' excludes all irrelevant nodes (i.e. node 4, node 7 and node 8) from the following optimization and thus could improve the computational efficiency.

Each component in the pool, or combination of the components if more than one booster station is to be established, together with their respective injection rate(s), is a potential solution for booster disinfection. If a target case is solved by one potential solution, while concentration values in the whole network at all monitoring intervals do not break the upper limit, the target case is regarded as 'covered' by the potential solution, i.e. the coverage is 1. Otherwise, the coverage would be 0. The concept of coverage is to ensure that any final solution is 'safe', without risking breach of the upper limit at other nodes in the network.

Table 1 | Example for terminology explanation

	Monitoring interval								
Node ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1			Δ						
2					Δ	Δ			Δ
3						Δ			

Optimization variables and objectives

To achieve the best booster disinfection plan, three variables need to be optimized, which are the number and locations of the booster disinfection stations, as well as the injection rate at each station. No doubt that the percentage of 'covered' target cases would increase with the increase of the number of booster stations. However for each number of booster stations, there is only one optimal solution, or several equally optimal solutions, relying on the optimization objective(s). Thus to make the problem simpler and results clearer, the booster station number is known as given for each optimization of the locations and injection rates. The number would start from one and step up by one after each optimization. After several optimizations, a trade-off would be observed between the number of booster stations and improvement of the water quality condition in the network. The comparison of the optimization results would suggest a reasonable solution, which achieves a satisfactory water quality improvement requiring a moderate investment, although the final decision is made by local authorities.

As for the optimization objectives, there are mainly two concerns: maximization of coverage of the 'target cases' and minimization of the disinfection injection rate. If the two objectives are equally important, there would be several non-dominated optimal solutions on a Pareto front. However, in reality, these two objectives have different priorities. For the sake of supplying safe drinking water, the first objective (maximization of coverage of nodes having a disinfection deficiency problem) has higher priority than the second one (minimization of the disinfection injection rate). In some situations, for example the case study in this research, the decision makers only take the second objective into account when two or more solutions have the same value of the first objective. Therefore, the optimal booster disinfection problem is constructed as a two-step single optimization problem.

At step 1, the booster disinfection optimization model can be written as the maximization of coverage of the 'target cases'. The mathematical formulation of

this model is:

$$\text{Maximize} \sum_{i=1}^{n_b} \sum_{m=1}^{n_t} \sum_{j=1}^{n_m} (\text{Cov}_j^m)_i \quad (1)$$

Subject to:

$$0 \le v_i;\ i = 1, \cdots, n_b \quad (2)$$

where n_b = number of booster disinfection stations for optimization (their locations are chosen from the 'candidate pool'); n_t = number of monitoring time intervals (the assumption of periodicity of mass injection and network hydraulic in Boccelli's research is adopted in this study, thus a long-term simulation is run until the disinfectant concentration dynamics at nodes become periodic, and then the last hydraulic cycle is monitored for booster disinfection optimization. In the case study, a hydraulic cycle lasts for 24 h and the monitoring time step is 1 h, so $n_t = 24$); n_m = number of disinfectant insufficiency nodes at the m^{th} monitoring time interval ($\sum_{m=1}^{24} n_m$ = number of all 'target cases'); $(\text{Cov}_j^m)_i$ = coverage of the i^{th} booster station to the j^{th} target node at the m^{th} monitoring time step; v_i = injection dosage rate (mg/s) at the i^{th} booster disinfection station (the dosage in this study varies with inflow rate to generate a constant concentration increment at the injection node, i.e. flow paced injection). Maximization of coverage will eventually increase the percentage of safe drinking water which satisfies bounds on residual disinfectant concentration according to the required regulations on drinking water quality.

If more than one optimal result is generated which maximizes coverage, a second objective of minimizing the disinfectant injection rate is applied. This is step 2 of the methodology. As the flow rate at each node changes every hour, the injection rate alters accordingly at the booster station at all times. The sum of the injection rate at all monitoring time intervals in the last hydraulic cycle, or the total injection mass during that period could be used as the optimization evaluation parameter; however in this study the average injection rate during the last cycle is used so that the results are more comparable with outcomes from other research:

$$\text{Minimize} \sum_{i=1}^{n_b} C_i \cdot \left[\left(\sum_{s=1}^{n_h} q_i^s \right) / n_h \right] \quad (3)$$

where C_i = constant concentration increase after the i^{th} booster station; n_h = number of hydraulic steps in a hydraulic cycle (the hydraulic time step δt_s is 1 h in this study, which is the same as the monitoring time step); q_i^s = flow rate moving through the i^{th} booster station at the s^{th} hydraulic step; $\sum_{s=1}^{n_h} q_i^s / n_h$ represents the average flow rate through the i^{th} booster station during the last hydraulic cycle. Minimization of the average injection rate not only reduces the total disinfectant injection mass, but also inhibits the formation of the DBPs.

The above two-step optimization could be continued for all solutions with different coverage, which would generate a Pareto front in a deterministic way. Although this can be achieved by the presented method in a straightforward way, the trade-off between coverage and injection rate is not the focus of this research, and thus is not calculated and presented in the paper.

Solution procedure

A computer program based on PBA was developed. It consists of two phases: the setup phase and the solution phase. In the setup phase, a whole water quality simulation is executed in EPANET2.0 (Rossman 2000) for the network with existing disinfection facilities, and the 'target cases' information in the last hydraulic cycle is recorded using EPANET Programmer's Toolkit. The library *UpstreamNode* (Shang *et al.* 2002) is employed to generate the 'candidate pool' of booster station locations. Parallel to that, composite response coefficients are calculated by perturbation as in Boccelli *et al.* (1998). The coefficients quantify the response of the monitoring nodes to a unit periodic dose at a booster disinfection station. As disinfectant injection in the case study is flow paced, a unit water source quality (1 mg/L) is set up at one node while all previous disinfection inputs in the network are removed. After a water quality simulation in EPANET2.0, concentration values of all nodes in the network, i.e. the composite coefficients, during the last

hydraulic cycle are saved in a matrix. The same process is carried on with other nodes being the water quality source node. The composite coefficients are basic elements in the solution phase for the calculation of coverage and concentration of the network using linear superposition theory. Linearity means that the concentrations at nodes in the network are proportional to input dosage rate and superposition implies two separate responses from two inputs could be added at an observed node. The adoption of the linear superposition principle avoids running the time-consuming water quality simulation in EPANET2.0, which greatly improves the optimization efficiency.

In the solution phase, the coverage value is calculated for each upstream node in the 'candidate pool' to each 'target case'. With the gap between lower concentration limit and the current value of the target case, the minimum injection concentration at the upstream node could be determined using the composite coefficient at the target node to the upstream node in the target time interval. If the injection concentration does not exceed the upper constraint, the concentration profile of all nodes over all monitoring periods is calculated using the linear superposition theory. If no violation of the upper concentration limit is observed across the whole network, the coverage is set to 1. Otherwise, if the injection concentration is above the upper limit, or any concentration value at the network under the minimum injection concentration exceeds the upper constraint, the coverage is set to be 0. The minimum injection concentration, which is the value calculated when coverage is 1, or 0 if coverage is 0, is recorded for the following injection optimization. Note that the injection concentration is used in the coverage calculation, because it is the setting parameter in the flow paced injection mode; however the injection rate is employed as an optimization objective for its physical meaning.

The results of the coverage calculation are summarized in a 0–1 matrix Cov_mat shown in Figure 1, and each minimum injection concentration result, multiplied by average flow rate through the respective upstream node in the last hydraulic cycle, is saved in another matrix Min_inj (Figure 2). The values in each row of Cov_mat display the coverage scope of the related upstream node. For instance, node 1 in Figure 1 could maximally cover four 'target cases': node 18 at time interval h_1, node 17 at both interval h_2 and h_3, and node 55

		monitoring time interval h_1			monitoring time interval h_2			monitoring time interval h_3	
		node 17	node 18	node 37	node 17	node 37	node 15	node 17	node 55
upstream nodes	node 1	0	1	0	1	0	0	1	1
	node 7	1	1	1	1	0	0	1	0
	node 9	0	1	0	0	1	1	0	1
	node 11	1	1	1	0	1	0	1	0
	node 12	1	0	1	0	1	1	0	1
	node 35	1	1	0	0	1	1	0	1
	node 40	1	0	0	1	0	0	1	0

Figure 1 | Coverage matrix Cov_mat.

		monitoring time interval h_1			monitoring time interval h_2			monitoring time interval h_3	
		node 17	node 18	node 37	node 17	node 37	node 15	node 17	node 55
upstream nodes	node 1	0	0.042	0	0.030	0	0	0.100	0.051
	node 7	0.031	0.230	0.019	0.045	0	0	0.700	0
	node 9	0	0.420	0	0	0.640	0.330	0	0.250
	node 11	0.370	0.021	0.340	0	0.550	0	0.310	0
	node 12	0.230	0	0.120	0	0.024	0.056	0	0.510
	node 35	0.770	0.240	0	0	0.560	0.078	0	0.210
	node 40	0.570	0	0	0.650	0	0	0.045	0

Figure 2 | Minimum injection rate matrix Min_inj.

in interval h_3. The values in each row of Min_inj reveal the minimum average injection rates to achieve different coverage results. As shown in Figure 2, the maximum coverage of node 1 is only achieved when the average injection rate is 0.10 mg/s; while injection value of 0.042 mg/s could only cover two 'target cases'. As the calculation of each injection rate already takes into account the upper concentration constraints at all nodes over all monitoring intervals, the selection of the maximum average injection rate would not violate the concentration constraints at other 'target cases' covered by a lower injection rate.

The optimization in this study is based on the two matrices yielded. If only one booster station is to be established, the sum of each row in Cov_mat would be first calculated. The solution with the maximum sum is the optimal result, and the maximum value of the related row in Min_inj is the respective average injection rate of the solution. If more than one solution has the optimal coverage, their minimum average injection rates will be compared and the one with the smallest value is the final solution.

If the booster station number n_b is more than one, all combinations of n_b rows of Cov_mat are listed as alternative solutions. For each solution, if the values in column c are not

all zero, it means the 'target case' which column c represents could be covered by the solution. Summing up the coverage results of each column yields the coverage result of the solution. As for the injection rates, the sum of the maximum value of each row is the total minimum average injection rate of the solution. However, the optimality of this solution is only guaranteed when the coverage ranges of the upstream nodes in the solution do not overlap with each other. For other cases, the optimal solution(s) needs to be checked to be conclusive.

To start with, the compliance of the upper concentration limits under this optimal solution should be investigated. As stated before, the upper limits have only been checked for a single booster station. So for multiple booster stations optimization, the minimum injection rate of each station calculated before should be used carefully because: (a) there might be a risk of violating the upper limits when all the stations reach their maximum coverage abilities; and (b) if the upper constraints are not violated, there might be a lower total injection rate that could achieve the same maximum coverage, i.e. the optimal result obtained before might need further optimization. So if the upper constraints are not violated, a matrix Jud_nonviro is established in this study to find a better solution. Figure 3 shows the matrix (n rows and m columns) for solution of two booster stations, where d_{xy} denotes the total average injection rate of the solution when the injection rate at the first booster station reduces to the $y+1^{\text{th}}$ maximum value, and the injection rate at the second booster station reduces to the $x+1^{\text{th}}$ maximum value. When only adjusting the injection rate at the first booster station, the maximum coverage of the solution would not be maintained if the injection rate reduces beyond the $m+1^{\text{th}}$ maximum value. Similarly, d_{n0} represents the total injection rate when injection at the second booster station goes to its extreme low value while keeping the maximum coverage result. All possibilities of injection reductions between the two stations are calculated and stored in the matrix. If d_{pq} has the minimum value in the matrix, it reveals that the solution could be best improved compared with the original one (i.e. d_{00}), if the injection rate at the first station reduces to the $q+1^{\text{th}}$ maximum value and the second station to the $p+1^{\text{th}}$ value.

$$\begin{bmatrix} d_{00} & d_{01} & \cdots & d_{0m} \\ d_{10} & d_{11} & & d_{1m} \\ \vdots & & \ddots & \vdots \\ d_{n0} & d_{n1} & \cdots & d_{nm} \end{bmatrix}$$

Figure 3 | Matrix Jud_nonviro.

Likewise, if the upper concentration constraints are violated, a similar matrix Jud_viro like the one in Figure 3 is created. The violation indicates that the current injection rates need to be decreased in order to satisfy the constraints. Thus the decrease could start from the first station until the violation is relieved, and the total coverage results of the adjusted solution are recorded along the process. In a similar way, the total coverage values of injection reductions at the second station and reduction combinations between the two stations are also recorded in the matrix. If the value on the x^{th} row and the y^{th} column is the maximum in the matrix, the solution it represents is the improved optimal result. If several elements in the matrix have the same maximum coverage, the one with minimum total injection rate would be the final solution. As the coverage result of the optimal solution might change after the check, the rank of the coverage of all possible solutions needs to be adjusted. The optimal solution in the new rank would be investigated with the upper limits compliance again, and the check processes above would be executed repeatedly until the confidently optimal solution is found. The scheme of the whole procedure is illustrated in Figure 4. The procedure could yield deterministic optimal booster

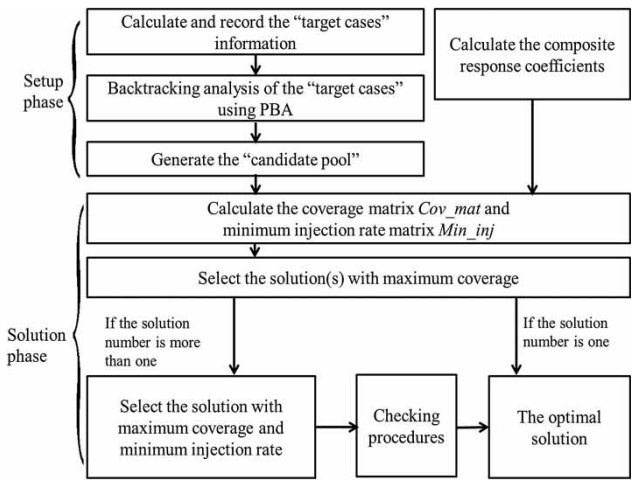

Figure 4 | Flowchart of the procedure.

disinfection solutions for any existing network. In real-life networks, as in the case study of this paper, the information revealed in the coverage matrix could be utilized to simplify the calculation and further improve the computational efficiency.

CASE STUDY

Network description

The network used in this case study is located in the suburban area of Beijing, China. It has 3,339 nodes, 3,363 links and 170.12 km of pipes in total (Figure 5). Almost 97% of the pipes are ductile iron pipes lined with cement, while only some small pipes are polyethylene. The network contains 33 underground water wells, 37 pumps and 3 tanks, which supply about 30,000 m^3/day of water. The three cubical tanks are modeled as continuous flow stirred tank reactors. For simplification, the quantity and variation of the water demand at the consumers are assumed to be deterministic. The water consumers are grouped into 11 categories according to their demand patterns. For each category, the water consumption patterns at some selected typical consumers are measured, the average of which is assigned to all consumers in the same category. Water usage variation at each node is modeled as basic water demand multiplied by the water demand pattern. The variation of water demand and pump operation are assumed to be varying hourly and periodic on a 24-h cycle, and follow the same pattern over the whole simulation period.

Pretreatment of the network

As disinfectant is usually added at the tanks in the water plants, nodes located before the tanks (including the wells)

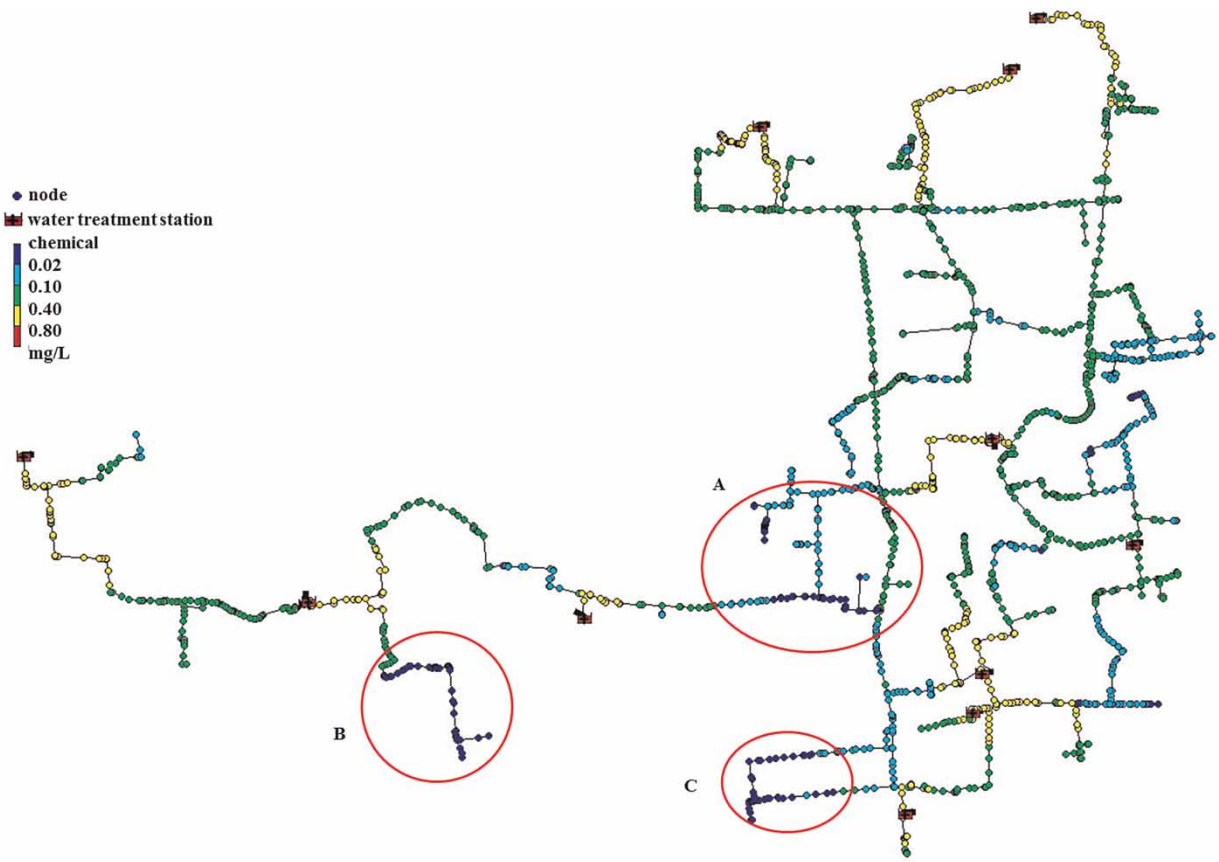

Figure 5 | Network schematic and concentration profile at hour-190 before optimization.

would have no disinfectant residuals. Thus these nodes would be unreasonably considered as 'target nodes'. Therefore, the combination of 'wells + nodes + a tank + pumps' is simplified to 'a single well + a single pump', which constantly supplies water to the network and keeps a stable water level. In addition, there are some end nodes that have no water consumption, or some abandoned water consumption points, which would also give 'false' signal of disinfection insufficiency. These end nodes are pruned from the network.

In the backtracking analysis, the algorithm can only recognize nodes with non-zero nodal demand on the upstream pathways. Nodes with zero nodal demand will not be recognized and thus be excluded from the candidate pool, although they are potential locations for installing booster disinfection stations. This will result in misleading results. To avoid this, each node in the network is assigned with a very small and extra water demand (0.000001 L/s) before the backtracking analysis initiates.

Disinfectant injection and residual measurement

The existing disinfectant applied to this network was chlorine dioxide, which was added into the network at 11 water treatment stations (shown as red rectangles in Figure 5). At each injection location, chlorine dioxide is injected with a flow-paced approach at a concentration of 0.8 mg/L, which is the upper concentration limit according to the national regulation for drinking water quality in China. As the injections at the water sources have already reached the limits, they are not optimized in this study.

For the network examined in this research, the pressures and water flow rates are measured online and recorded in a 'supervisory control and data acquisition' (SCADA) system. The concentration of chlorine dioxide residual is measured on site by engineers regularly. The chlorine dioxide concentration at the water plants is measured by the titration method four times a day. In the network, the chlorine dioxide concentration is measured on site at 16 sampling sites using a HACH Pocket Colorimeter II Test Kit twice a month. These concentration values are used to calibrate the chlorine decay model.

Chlorine dioxide decay kinetics

A key assumption in the linear superposition theory is the first order disinfectant decay kinetics, thus surveys on decay kinetics were carried out to check whether the linear superposition principle could be adopted. Besides, parameters of disinfectant decay kinetics are key elements in the water quality model of the network. To investigate the bulk decay of chlorine dioxide, all water treatment stations are classified into four groups according to their water characteristics. Water samples are taken from one water treatment station from each group to conduct the jar test. Results for this water treatment station are used to represent its group. In the jar test, the concentration of chlorine dioxide is measured every 2 h in the first 2 days and every 24 h in the following 10 days. The concentration values from each water treatment station are then fitted with four different decay kinetics formulations: zero-order, first-order, parallel first-order, and limited first-order. The R^2 of the four groups for each decay kinetics formulation are listed in Table 2. As shown in the table, there are no significant differences between the performances of the four decay kinetics formulations. Each formulation shows slight advantage and disadvantage in some cases. Considering first order kinetics formulation can be easily modeled in most water distribution network modeling software and could significantly simplify the optimization of booster disinfection, it is selected to simulate the decay process of chlorine dioxide in this study. The bulk decay coefficient of the model, which is calculated as the average of the bulk decay coefficients of the four groups, is $-0.076\ \text{day}^{-1}$.

For the wall decay kinetics, the decay order of the examined system is set to be 1 considering the average age (8 years) and the corrosion on the inner surface of the pipes. As nearly all pipes in the network are ductile iron and

Table 2 | Bulk decay kinetics orders and the corresponding R^2

	Water plant			
Decay order (R^2)	Group A	Group B	Group C	Group D
Zero-order	0.9843	0.8867	0.9566	0.9735
First-order	0.9783	0.9323	0.9794	0.9727
Parallel first-order	0.9485	0.9808	0.9904	0.9624
Limited first-order	0.9558	0.9655	0.9921	0.9656

have the same pipe age, the wall decay coefficient is assumed to be the same across the whole network. In view of the high cost of measurement on site, the wall decay coefficient is calculated using a heuristic method. By measuring the residual concentration at five sites along a trunk main pipe (600 mm) during one day, the value of wall decay coefficient is optimized using the Genetic Algorithm (Goldberg 1989; Liu *et al.* 2007). The wall decay coefficient is calculated to be -0.32 day^{-1}.

Setup phase calculation

Besides calculating the composite response coefficients, a 192-h water quality simulation is executed in the setup phase. The disinfectant residual concentration profile in the last 24 h at various nodes of the network is recorded in the matrix Conc, through which 7,558 'target cases' are detected during this period. As shown in the snapshot of residual profile at hour-190 in Figure 5, the disinfectant weak points are mainly gathered in three regions (A, B and C). The information (node ID and monitoring time period) of each 'target case' is recorded in a text file, which is used by *UpstreamNode* for the backtracking analysis. After the analysis, 1,713 nodes are selected into the 'candidate pool'.

Solution phase calculation and results

In the solution phase, coverage matrix Cov_mat (1,713 rows and 7,558 columns) and the same sized minimum injection rate matrix Min_inj are calculated.

In the case of one booster station, the optimal location is node J0772 (Figure 6(a)) and the related average injection rate is 13.94 mg/s. This booster station could cover 1,831 'target cases' and the concentration profile after the booster disinfection is also shown in Figure 6(a). Most disinfection insufficiency problems in region A could be solved. Besides, as it is on a trunk through which water from sources is transported and distributed to nodes in the western part of the network, it could supplement disinfectant to nodes along the trunk pipe due to its favorable hydraulic location.

For the case of two booster stations, there are two optimal solutions in terms of coverage: (a) node J0772 and node S0876; and (b) node J0772 and node S0934. The minimum injection rates at node S0876 and node S0934 are 5.25 and 5.35 mg/s, respectively. Thus the optimal locations are node J0772 and node S0876 (Figure 6(b)) with injection rates of 13.94 and 5.25 mg/s, respectively. The total coverage of the two booster stations is 3,120 'target cases'. As node S0876 is also on a trunk pipe of the network, it could enhance disinfection at the eastern part of the network.

When further increasing the booster disinfection station number to three, the number of solved 'target cases' rises to 4402, and the optimal locations for booster disinfection stations are node J0772, node S0876 and node S446 (Figure 6c), with average injection rates of 13.94, 5.25 and 0.43 mg/s. Node S446 could solve most of the disinfection problems in area B.

As the combination number increases about 1,000 times each time the booster station number grows, the numeration of all possible combinations becomes computationally expensive for more than three booster stations, limited by the available computation facilities in this study. Thus a simplification strategy is employed to obtain some rough optimal results for reference. The optimization of n booster stations is based on the optimal results of $n-1$ booster stations. That is to say, the 'target cases' covered by the first $n-1$ booster stations would be deleted from further optimization. The n^{th} station only needs to find the best solution for the rest of the 'target cases'. Using this strategy, the fourth, fifth, sixth and seventh booster stations are found at node J1266, node J2020, node B057 and node S484-1, and the increased coverage of the 'target cases' by each station is 1,147, 689, 303 and 208, respectively. Although these results (referred to as 'preliminary results' in this paper) are not the final ones, they could be utilized to reduce the computational efforts of further accurate optimizations.

The coverage range of node J1266, the fourth preliminarily yielded station, is compared with the first three optimized stations, and overlap of coverage scope is only observed between node J0772 and node J1266. The result could be directly verified through Figure 7(a), in view of the geographic locations of the four nodes and distances between them. As the coverage of node J1266 is quite high compared with the coverage of the first three stations and the preliminarily calculated fifth station, the fourth station should be located in this area. Considering that node S0876 and node S446 are not hydraulically connected with this area, 'target cases' covered by these two stations are removed from this

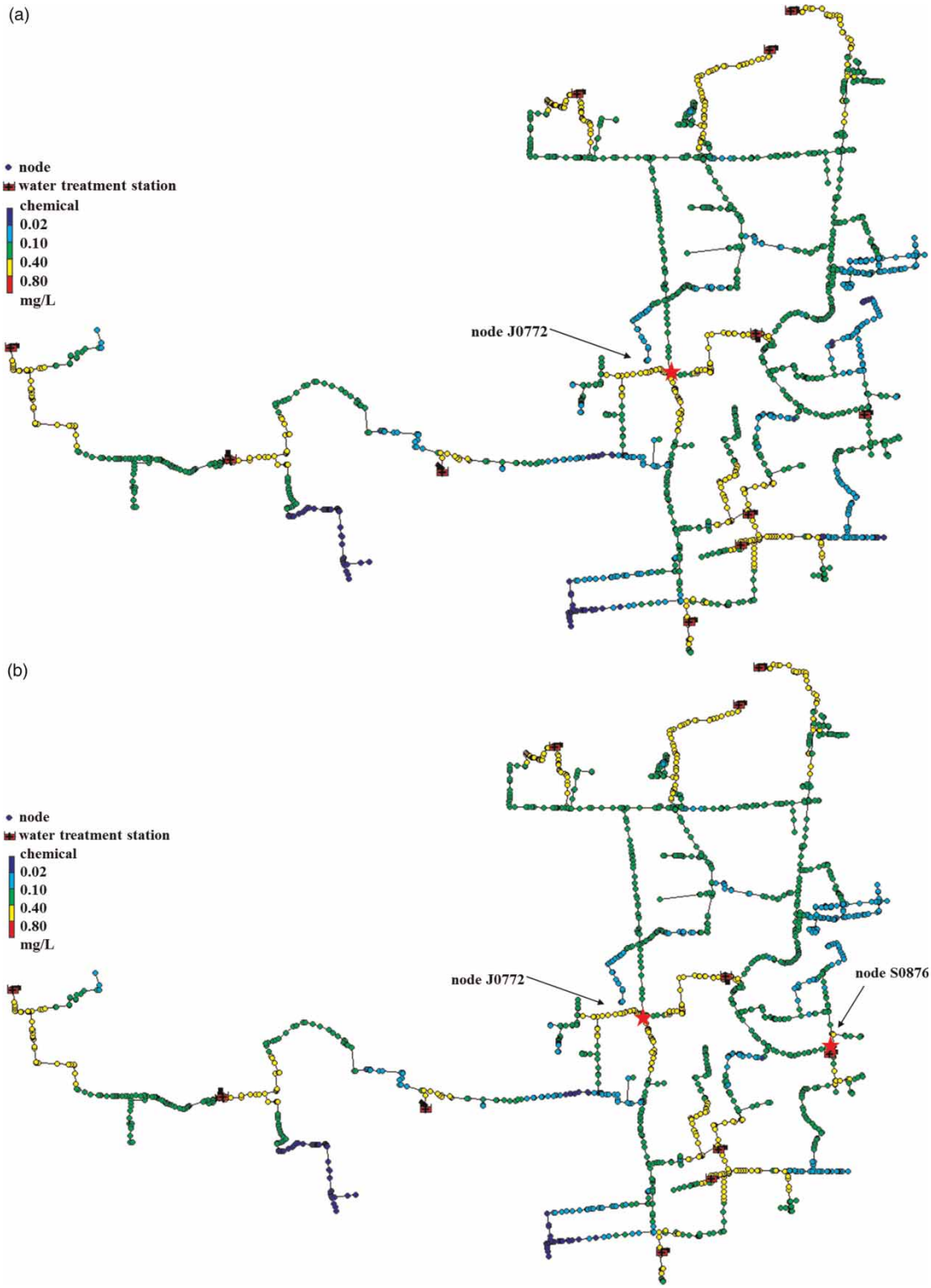

Figure 6 | Concentration profile with one, two and three booster disinfection stations. (a) One booster disinfection station. (b) Two booster disinfection stations. (c) Three booster disinfection stations.

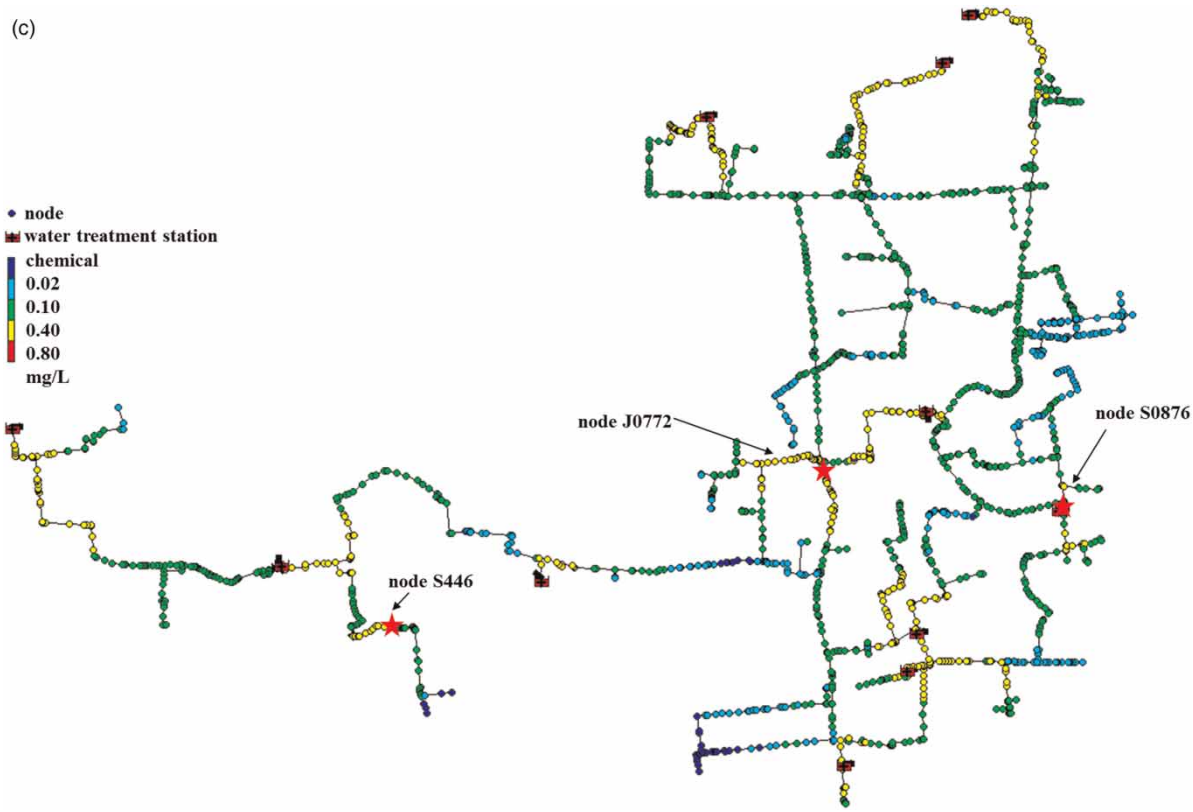

Figure 6 | continued.

optimization. Thus two booster stations are optimized for the rest of the 'target cases', and results with best coverage (5,543 covered cases) are found at node J0772 and node J1266 – the same as the preliminary result. The combined disinfection effects of the two stations are added on the existing network with their single optimized injection rates. No violation of the upper concentration limits is observed and any attempt to reduce the injection rate would reduce the total coverage. Thus the fourth station should be located at node J1266 with average injection rate of 0.15 mg/L, which supplements disinfectant to many of the weak points in region C.

As the preliminary result of the fifth booster station (node J2020) is also in area C (Figure 5 and Figure 7(b)), the coverage scope of node J2020 is compared with that of node J0772 and node J1266. Overlaps are found with both two stations. Thus a global optimization is needed to find the three optimal stations besides node S0876 and node S446. The optimal solution is the same as the preliminary result once again, which is node J0772, node J1266 and node J2020, covering 6,232 target cases in total. The concentration profile of the network is added with the three single optimal injection rates, and violations of upper constraints are detected. It indicates that the current injection rates need to be reduced. Of all the three nodes, node J2020 is the only one able to remove the violation with only one step reduction in injection rate, from 0.50 to 0.49 mg/s; moreover, the total coverage still remains the same. As the violation has been released, the next step is to optimize the injection rates in the solution while still keeping the coverage. Node J0772 can reduce its injection rate to 13.91 mg/s by one step of reduction. Injection at node J1266 can be reduced to 0.13 mg/L by two steps. While injection at node J2020 could be considerably reduced from 0.49 to 0.09 mg/s by 23 steps. A $2 \times 3 \times 24$ sized matrix could be set up for the optimization of injection rates among the three stations. However, this computation could be avoided in this case, because all the maximum reductions could be achieved at the same time without sacrificing the total coverage. Most of the disinfection problems in area C are solved by the addition of a booster station at node J2020. Thus

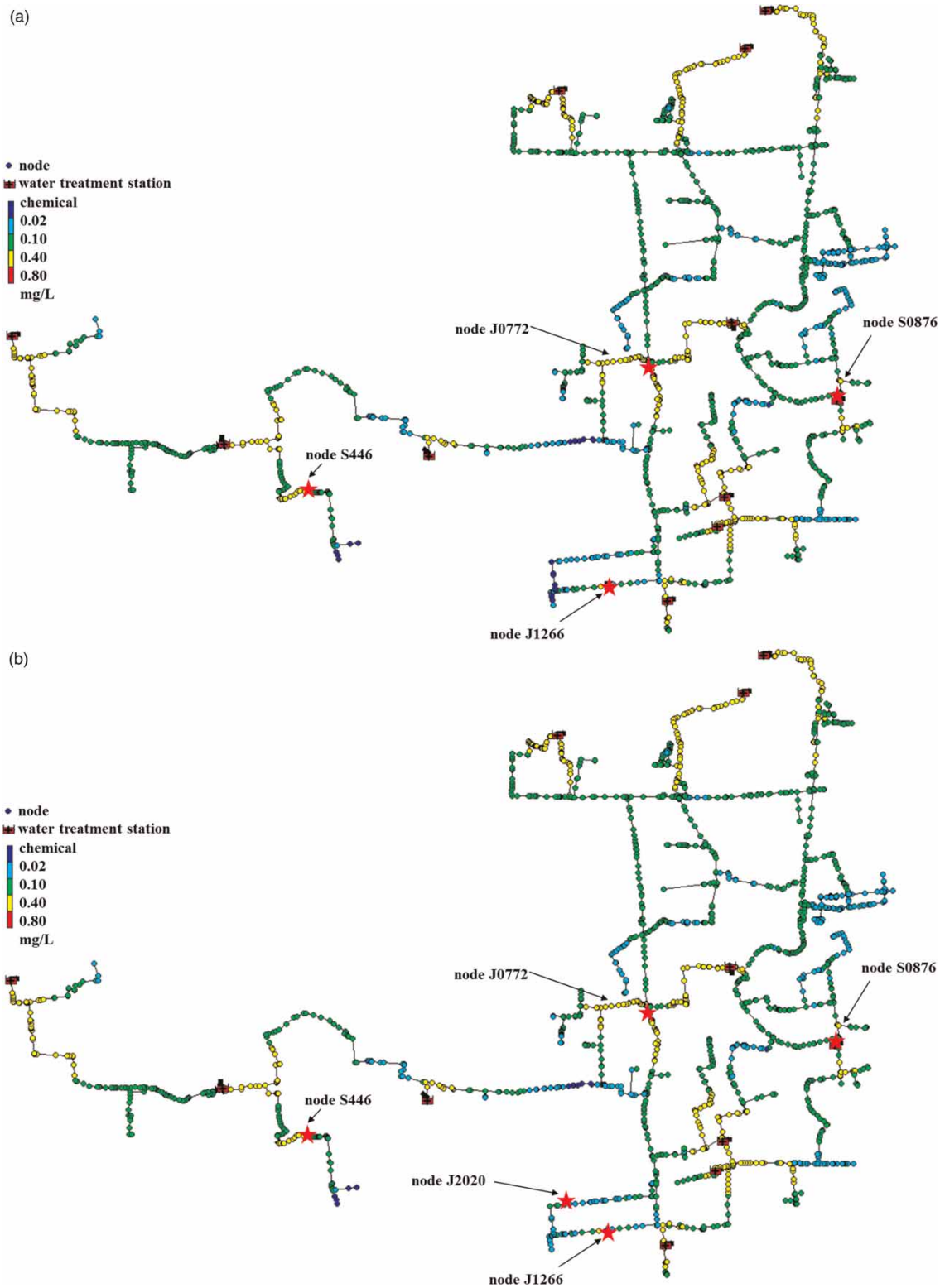

Figure 7 | Concentration profile with four, five and seven booster disinfection stations. (a) Four booster disinfection stations. (b) Five booster disinfection stations. (c) Seven booster disinfection stations.

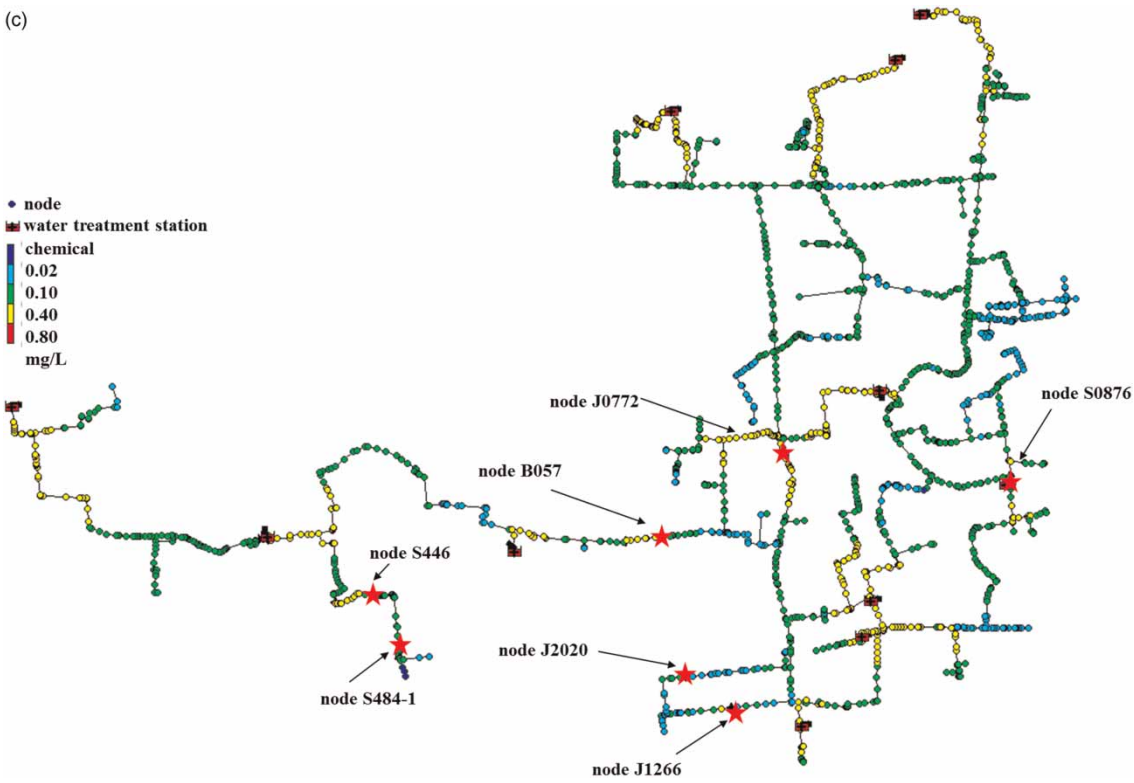

Figure 7 | continued.

the optimal locations for five booster stations are at node J0772, node S0876, node S446, node J1266 and node J2020, with respective minimum average injection rates of 13.91 mg/s, 5.25mg/s, 0.43 mg/s, 0.13 mg/L and 0.09 mg/s.

As for the sixth booster station, the preliminary result shows it to be at node B057. Its coverage range only overlaps with that of node J0772. Thus the target cases covered by previously optimized stations except J0772 could be eliminated from the solution space, and two stations need to be optimized for the rest target cases. The location results are still node J0772 and node B057 as the preliminary result, with a total of 6,535 covered cases. The upper limits are not violated and the injection rates, which are 13.94 and 1.30 mg/s respectively, could not be further reduced.

The seventh booster station in the preliminary result is at node S484-1, near the end of area B (Figure 5 and Figure 7(c)). Further optimization validated this result. The upper limits are not violated, but the injection rate at node S446 could be reduced to 0.37 mg/s while the total coverage of the two nodes is not affected. The optimal injection rate at node S484-1 is 0.02 mg/s. With this solution, there are a total of 6,743 target cases covered.

DISCUSSION

Although the simplification strategy for the rough preliminary results is not rigorous mathematically, their results are the same as the ones from the accurate optimizations in the case study. This might indicate that, for an existing network, there are a few key locations that have strong hydraulic influence to improve the disinfection situation. For instance, the first two booster stations in the case study are optimized both on the trunk pipe, suggesting that the upstream node(s) with large flow rate(s) is very efficient in transporting disinfectant, and also very effective in spreading the disinfectant to more nodes in the network.

After the majority of the disinfection insufficiency problems are satisfied by the first few stations, the booster stations are prone to 'clog together' if the optimization continues. Figure 8 shows the accumulated percentage of coverage and total average injection rate as the booster station number grows. The coverage percentage is ever increasing, although the increasing rate slows down after the fourth booster station. In contrast, the total injection rate is not always rising. The two pits in the figure show the reduced total injection rate if two booster stations are close together (i.e. node J2020 & node J1266, and node S446 & node S484-1). Although the coverage is still increasing, the cost-effectiveness of setting a new booster station near another station needs to be considered. Figure 9

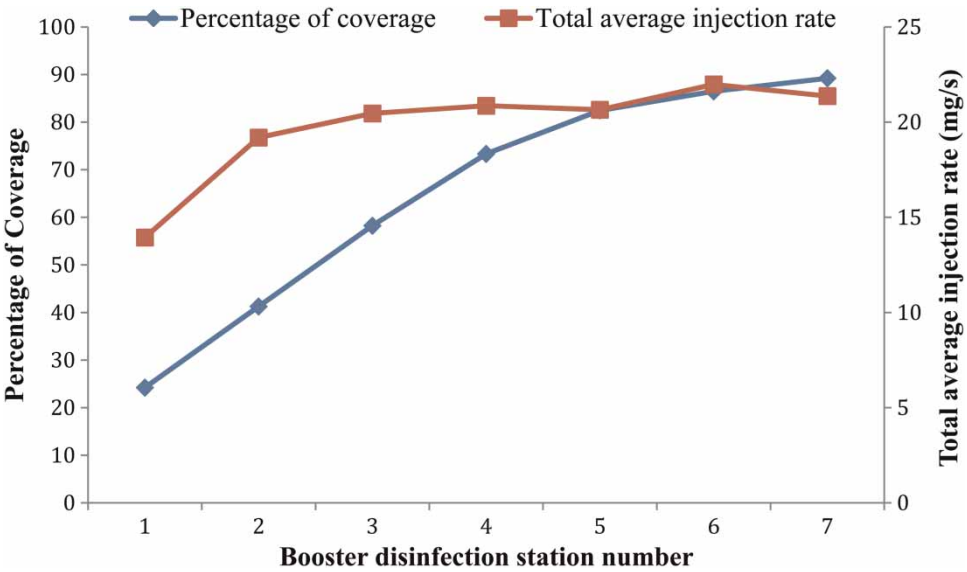

Figure 8 | Relationship of percent of coverage and total average injection rate with number of booster disinfection stations.

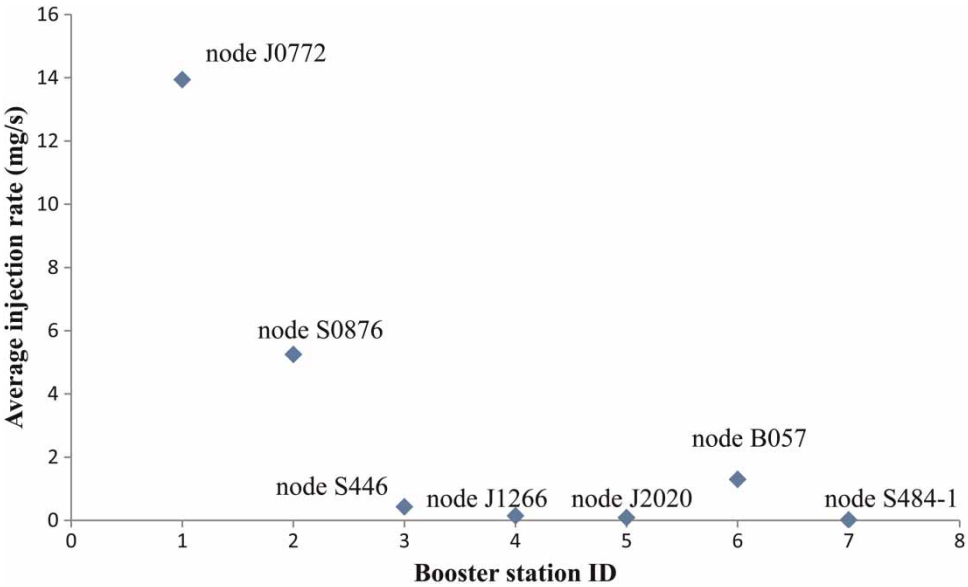

Figure 9 | Relationship of average injection rate at each booster station.

presents the average injection rate of each booster station. The injection rate at a certain optimized booster station changes under different numbers of total booster stations, however considering the minor difference, the larger value is used for this figure. As shown in Figure 9, although the injection rate at the sixth station bounces back a bit, the general trend of trivial injection rates after the fourth station suggests that, investing more booster stations to achieve a better coverage might not be a wise solution.

For the case of four booster stations, although there are still 2015 'target cases', the number of average uncovered cases in each monitoring time interval is about 83, which only accounts for 2.5% of the total nodes. In other words, for the case study network, installing new booster disinfection stations at only 0.1% of total nodes (i.e. $4/3{,}339 = 0.1\%$) could ensure the chlorine residual at about 97.5% of the total nodes satisfying the national regulation. Further investigation on the flow rates in the uncovered areas reveals that the uncovered target nodes are mostly located in the 'dead zone', where water demand is small and water age is high. It might be better to solve such a problem using other approaches, for example, opening valves regularly to speed up water circulation, rather than adding more booster disinfection stations. Thus four booster stations are recommended for this case study area.

CONCLUSIONS

A simple matrix-based method to solve an optimal booster disinfection problem, which employs the PBA, is developed in this study and applied to a real life network in Beijing, China.

The following conclusions are drawn.

1. The proposed method is based on the current disinfection state of the network, and could provide deterministic optimal solutions for booster disinfection, which maximally cover the disinfection deficiency problems with minimum disinfectant injection. The procedures introduced in this paper are applicable to any existing network in real-life.
2. The research avoids huge computations through a few approaches:
   (a) The hydraulic calculation of the PBA scales down the potential booster disinfection locations to the upstream nodes. In the case study, 1,626 irrelevant nodes out of all 3,339 nodes in the network are removed from the optimization.
   (b) A constraints-embedded matrix is introduced, thus only simple mathematical operations of the matrix are needed for the optimization of the locations and injection rates. The matrix is calculated under the linear superposition theory, which avoids any time-consuming water quality simulation in software. Furthermore, the hydraulic information the coverage matrix contains could be used to simplify the computation process.
3. Results from the case study show that although the percentage of covered 'target cases' increases with the increment of the number of booster disinfection stations, the increasing rate slows down after a certain number, with also the decrease of injection rate at each station. Four booster stations are recommended to be cost-effective for this case study, which by adding booster disinfection station at 0.1% nodes can satisfy the chlorine residual at about 97.5% of the total nodes.

ACKNOWLEDGEMENTS

The authors would like to express thanks to three anonymous reviewers for their valuable comments. The authors also thank Tsinghua Independent Research Project (2011080993) and National Nature and Science Foundation of China (51290284) for their financial support to this research.

REFERENCES

Boccelli, D. L., Tryby, M. E., Uber, J. G., Rossman, L. A., Zierolf, M. L. & Polycarpou, M. M. 1998 Optimal scheduling of booster disinfection in water distribution systems. *J. Water Res. PL-ASCE* **124** (2), 99–111.

Boccelli, D. L., Tryby, M. E., Uber, J. G. & Summers, R. S. 2003 A reactive species model for chlorine decay and THM formation under rechlorination conditions. *Water Res.* **37** (11), 2654–2666.

Courtis, B. J., West, J. R. & Bridgeman, J. 2009 Temporal and spatial variations in bulk chlorine decay within a water supply system. *J. Environ. Eng.-ASCE* **135** (3), 147–152.

Cozzolino, L., Pianese, D. & Pirozzi, F. 2005 Control of DBPs in water distribution systems through optimal chlorine dosage and disinfection station allocation. *Desalination* **176**, 113–125.

Goldberg, D. E. 1989 *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison-Wesley Longman, Boston, MA.

Hallam, N. B., Hua, F., West, J. R., Forster, C. F. & Simms, J. 2003 Bulk decay of chlorine in water distribution systems. *J. Water Res. PL-ASCE* **129** (1), 78–81.

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C. & Spencer, I. 2002 The decay of chlorine associated with the pipe wall in water distribution systems. *Water Res.* **36** (14), 3479–3488.

Isovitsch, P. S. L. & VanBriesen, J. M. 2009 Booster disinfection for response to contamination in a drinking water distribution system. *J. Water Res. PL-ASCE* **135** (6), 502–511.

Kang, D. & Lansey, K. 2010 Real-time optimal valve operation and booster disinfection for water quality in water distribution systems. *J. Water Res. PL-ASCE* **136** (4), 463–473.

Kumar, A., Kansal, M. L. & Arora, G. 1997 Identification of monitoring stations in water distribution system. *J. Environ. Eng.-ASCE* **123** (8), 746–752.

Lee, B. H. & Deininger, R. A. 1992 Optimal locations of monitoring stations in water distribution system. *J. Environ. Eng.-ASCE* **118** (1), 4–16.

Liu, S. M., Butler, D., Brazier, R., Healthwaite, L. & Khu, S. T. 2007 Using genetic algorithms to calibrate a water quality model. *Sci. Total Environ.* **374** (2–3), 260–272.

Ostfeld, A. & Salomons, E. 2006 Conjunctive optimal scheduling of pumping and booster chlorine injections in water distribution systems. *Eng. Optimiz.* **38** (3), 337–352.

Powell, J. C., West, J. R., Hallam, N. B., Forster, C. F. & Simms, J. 2000 Performance of various kinetic models for chlorine decay. *J. Water Res. PL-ASCE* **126** (1), 13–20.

Prasad, T. D., Walters, G. A. & Savic, D. A. 2004 Booster disinfection of water supply networks: multiobjective approach. *J. Water Res. PL-ASCE* **130** (5), 367–376.

Rossman, L. A. 2000 *EPANET 2 User's Manual*. National Risk Management Research Laboratory, Cincinnati, OH.

Rossman, L. A., Clark, R. M. & Grayman, W. M. 1994 Modeling chlorine residuals in drinking-water distribution-systems. *J. Environ. Eng.-ASCE* **120** (4), 803–820.

Shang, F., Uber, J. G. & Polycarpou, M. M. 2002 Particle backtracking algorithm for water distribution system analysis. *J. Environ. Eng.-ASCE* **128** (5), 441–450.

Tryby, M. E., Boccelli, D. L., Koechling, M. T., Uber, J. G., Summers, R. S. & Rossman, L. A. 1999 Booster chlorination for managing disinfectant residuals. *J. Am. Water Works Assoc.* **91** (1), 95–108.

Tryby, M. E., Boccelli, D. L., Uber, J. G. & Rossman, L. A. 2002 Facility location model for booster disinfection of water supply networks. *J. Water Res. PL-ASCE* **128** (5), 322–333.

Vasconcelos, J. J., Rossman, L. A., Grayman, W. M., Boulos, P. F. & Clark, R. M. 1997 Kinetics of chlorine decay. *J. Am. Water Works Assoc.* **89** (7), 54–65.

Zierolf, M. L., Polycarpou, M. M. & Uber, J. G. 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.

First received 28 September 2012; accepted in revised form 7 January 2013. Available online 11 February 2013