

Compressed sensing based optimal sensor placement for leak localization in water distribution networks

Xiang Xie, Quan Zhou, Dibo Hou and Hongjian Zhang

ABSTRACT

The performance of model-based leak detection and localization techniques heavily depends on the configuration of a limited number of sensors. This paper presents a sensor placement optimization strategy that guarantees sufficient diagnosability while satisfying the budget constraint. Based on the theory of compressed sensing, the leak localization problem could be transformed into acquiring the sparse leak-induced demands from the available measurements, and the average mutual coherence is devised as a diagnosability criterion for evaluating whether the measurements contain enough information for identifying the potential leaks. The optimal sensor placement problem is then reformulated as a {0, 1} quadratic knapsack problem, seeking an optimal sensor placement scheme by minimizing the average mutual coherence to maximize the degree of diagnosability. To effectively handle the complicated real-life water distribution networks, a validated binary version of artificial bee colony algorithm enhanced by genetic operators, including crossover and swap, is introduced to solve the binary knapsack problem. The proposed strategy is illustrated and validated through a real-life water distribution network with synthetically generated field data.

Key words | artificial bee colony, compressed sensing, diagnosability, leak localization, optimal sensor placement, water distribution network

Xiang Xie
Quan Zhou
Dibo Hou (corresponding author)
Hongjian Zhang
College of Control Science and Engineering,
Zhejiang University,
Hangzhou,
Zhejiang,
China
E-mail: houdb@zju.edu.cn

INTRODUCTION

Water loss management is one of the most significant issues facing urban water utilities around the world. According to the World Bank study (Kingdom *et al.* 2006), a total amount of 45 million cubic meters of water are lost daily in developing countries, enough to serve nearly 200 million people, and causing a conservative estimate of US\$4.7 billion economic losses per year. Among those losses, huge volumes of water are lost through leaks, not being effectively invoiced to customers, as a result of damage or defects in pipelines, lack of maintenance, or increase in pressure. Therefore, detection and localization of water leakages within water distribution networks (WDNs) is extremely urgent for the sustainable development of urban cities.

The past few years have witnessed much progress towards leak detection, localization and pinpointing. In

this field, Hao *et al.* (2012) provides a comprehensive review of the state-of-the-art technologies for condition assessment of underground utilities, especially water supply pipelines. Unfortunately, although effective, these techniques, including widely used acoustic and vibration approaches (Gao *et al.* 2005), are time-consuming and resource-intensive when applied to large-scale WDNs. An alternative is to deploy the WDNs with flow and pressure monitoring devices (Puust *et al.* 2010). Generally, pressure instruments are preferred in practice due to the convenience for installation. Data-driven as well as model-based approaches are adopted, both with their respective advantages. Data-driven approaches barely rely on prior knowledge about the monitored WDNs, since they focus on extracting features of historical data (Bakker *et al.* 2014;

Laucelli *et al.* 2016). Note that applying data-driven approaches may lead to misjudgements when inexperienced events are dealt with. The application of hydraulic transient models to leak localization is another hotspot in the research field. Colombo *et al.* (2009) offer a selective literature review of transient model-based methods. So far, even with long operational range and high positioning accuracy, the transient-based methods are not yet competent for supplanting existing techniques, due to their strong reliance on a number of pressure measuring instruments with high sampling frequency. On the other hand, leak detection and localization techniques on the basis of stationary models has been researched since the first seminal paper of Pudar & Liggett (1992). According to the theory of model-based fault detection and isolation (FDI), residuals online between pressure measurements and estimations using a hydraulic simulator are generated and evaluated. To detect leaks, residuals are checked against thresholds. Further, when some of the residuals violate the thresholds, the residuals are matched against the leak sensitivity matrix to locate potential leaks (Casillas *et al.* 2014; Rosich *et al.* 2015). The stationary model-based methods or mixed model-based/data-driven approaches overall offer more effective and less costly system-wide diagnosis and monitoring schemes (Meseguer *et al.* 2014; Zhang *et al.* 2016).

It is obvious that the location of pressure instrumentations is crucial to provide good quality information for optimizing the performance of different management tasks. Many studies have been conducted for optimal sensor placement, driven by the need for model calibration, network segmentation and leak localization respectively. To put it simply, the optimal sensor placement can be regarded as a multi-objective optimization problem that selectively balances the need for improving the performance with the need to justify the expense of data collection (including metering, installation and maintenance costs). For model calibration, the overall performance is evaluated by the estimation or prediction accuracy, quantified by three most often adopted indexes, the so-called A, D and V-optimality criteria (Savic *et al.* 2009), or combined (Kang & Lansey 2010). The A- and D-optimality are concerned with the estimated parameter uncertainty, aimed at either minimizing the trace or maximizing the determinant of the parameter covariance matrix, while the V-optimality is concerned with

the prediction uncertainty, aimed at minimizing the trace of the prediction covariance matrix. For network segmentation, the sampling-oriented modularity (Simone *et al.* 2016) is employed as a segmentation metric to gain better 'pressure DMAs' configuration. For leak localization, various quantitative indexes are introduced for evaluating the sensor placement schemes on locating possible leaks. Christodoulou *et al.* (2013) define a total entropy describing the ratio of sensing length over pipe length, and optimizes the sensor placement by maximizing the total entropy in the network; Casillas *et al.* (2015) adopt a leak signature space (LSS) method, and employ the criterion of minimizing the number of overlapping signature domain using the integer optimization method; Sarrate *et al.* (2014) propose an isolability index as the number of isolable node pairs under a single operating point scenario, and search for optimal sensor placement by a branch and bound method based on structural analysis; Blesa *et al.* (2016) reformulate the sensor placement problem as a bi-objectives optimization for maximizing both mean and worst leak locatability index, and take into account the dependency of leak localization on sequential operating points; Steffelbauer & Fuchs-Hanusch (2016) comprehensively incorporate uncertainties from different types and sources in the optimal sensor placement problem. To sum up, the key in sensor placement optimization for leak localization is introducing a quantitative index to evaluate the quality of every possible placement scheme, for which a decision-maker would choose a preferred one.

This paper formulates the leak localization problem as a FDI problem with underdetermined linear equations, and adopts a diagnosability criterion to guide the sensor placement optimization so that leaks can be located. By maximizing the diagnosability using an enhanced binary artificial bee colony (ABC) algorithm, a quasi-optimal sensor placement can be guaranteed. The proposed sensor placement optimization approach is illustrated through a real-life network with synthetically generated field data.

DIAGNOSABILITY CRITERION AND OPTIMAL SENSOR PLACEMENT

In this section, a compressed sensing based diagnosability criterion is adopted to guide the sensor placement

optimization and enhance the overall performance for leak detection and localization. The following subsection proposes a summary of the model-based leak detection and localization procedure. Particularly, an analytical expression for the leak sensitivity matrix is provided. Next, diagnosability analysis is presented for a leak localization procedure with the theory of compressed sensing. Average mutual coherence, a typical diagnosability criterion, is adopted for evaluating the quality of sensor placement. This is followed by a subsection describing the implementation of an enhanced binary ABC algorithm to support the decision making process, aimed at maximizing the degree of diagnosability (minimizing the average mutual coherence in this paper) and thus optimizing the sensor placement.

Leak detection and localization in WDNs

Model-based leak detection and localization techniques are adopted in this paper, aimed at providing a more effective and less costly system-wide diagnosis and monitoring scheme for WDNs. According to the theory of FDI, a set of online residuals between available pressure measurements and corresponding estimations using hydraulic simulators are generated and evaluated. In the absence of leaks, all of the generated residuals remain below given thresholds. Otherwise, when some of the residuals violate the thresholds, the residuals are matched against the leak sensitivity matrix, searching for the most favorable combination of leak candidates that leads to the consistency of the observed and expected residuals.

The model-based diagnosis methodology relies on the sensitivity-to-leak analysis, describing the theoretical pressure disturbances induced by leak candidates. An analytical expression of the leak sensitivity matrix is provided in this paper. For pressurized water steadily flowing through a network, the governing equations can be expressed as:

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{R}|\mathbf{q}|^{e-1} & \mathbf{A}^T \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \mathbf{H} \end{bmatrix} + \begin{bmatrix} -\mathbf{Q} \\ \mathbf{A}_{10}\mathbf{H}_0 \end{bmatrix} = \mathbf{0} \quad (1)$$

where \mathbf{A} = the $n \times p$ incidence matrix relating pipes to junction nodes (with unknown heads) in the WDN, in which the

elements of \mathbf{A} take values 0, +1 (assumed flow enters the node) and -1 (assumed flow exists the node); similarly, \mathbf{A}_{10} relates pipes to fixed head nodes; n and p = the number of junction nodes and pipes; \mathbf{R} = the pipe friction parameters vector, depending on the diameter, roughness and length of the pipe; e = flow exponent, the value is 1.852 when using the Hazen-Williams equation; \mathbf{q} and \mathbf{Q} = the vector of pipe flows and user required demands; \mathbf{H} and \mathbf{H}_0 = the vector of unknown and known heads. Here, it is assumed that minimum service head required for supplying demand \mathbf{Q} is guaranteed among the whole network. In other words, the nodal demands \mathbf{Q} is independent of available heads \mathbf{H} . Since exact modelling of any potential leaks is intractable in practice, assumption has been made that leaks only appear at existing nodes. Thus, for the sake of simplicity, leaks are modelled as extra nodal demands. The total nodal demands \mathbf{Q}' can be written as:

$$\mathbf{Q}' = \mathbf{Q} + \mathbf{I}(\mathbf{H}) \quad (2)$$

where \mathbf{I} = the $n \times 1$ vector representing extra nodal demands induced by potential leaks in WDNs. From a systematic point of view, the analytical solution of Jacobian matrix J_{H-Q} regarding nodal heads to nodal demand can be derived (Piller et al. 2017):

$$\mathbf{J}_{H-Q} \triangleq \frac{\partial \mathbf{H}}{\partial \mathbf{Q}'} = -(\mathbf{A}\mathbf{B}\mathbf{A}^T + \nabla_{\mathbf{H}}\mathbf{I})^{-1} \quad (3)$$

where $\mathbf{B} = (e\mathbf{R}|\mathbf{q}|^{e-1})^{-1}$ and ∇ is the nabla operator. In this paper, the focus is placed on locating medium magnitude unreported leaks with reasonable \mathbf{I} , since large leaks would be visible to end users and need no further study. In this case, the term $\nabla_{\mathbf{H}}\mathbf{I}$ is relatively small and can be neglected in the calculation. Then, an approximate expression for residual vector \mathbf{r} , as the difference between the nominal pressure in the model without leaks and the observed pressure in the network with specific leaks \mathbf{I} , can be established:

$$\mathbf{r} = (\mathbf{A}\mathbf{B}\mathbf{A}^T)^{-1} \mathbf{I} + \delta \quad (4)$$

where δ represents errors resulting from systematic errors induced by nodal demand uncertainties or inaccurate node

elevation, and noises in measurements. Let \mathbf{A}_1 be a $n \times n$ non-singular submatrix of \mathbf{A} , corresponding to a spanning tree that uses n pipes to connect unknown head nodes and the known head node (root), and \mathbf{A}_2 denotes a $n \times (p-n)$ submatrix of \mathbf{A} corresponding to the co-tree with $\mathbf{N} = \mathbf{A}_1^{-1}\mathbf{A}_2$. \mathbf{B}_1 is the partitioned matrix of \mathbf{B} containing the spanning tree pipes, while \mathbf{B}_2 is the partitioned matrix containing the co-tree. Then, according to the graph theory, the leak sensitivity matrix \mathbf{S} can be further simplified as:

$$\mathbf{S} \triangleq (\mathbf{A}\mathbf{B}\mathbf{A}^T)^{-1} = \mathbf{P}^T (\mathbf{B}_1 + \mathbf{N}\mathbf{B}_2\mathbf{N}^T)^{-1} \mathbf{P} \quad (5)$$

where \mathbf{P} = the path matrix of the chosen spanning tree, only correlative with the network topology. According to the theory of graph (Bapat 2014), $\mathbf{P} = \mathbf{A}_1^{-1}$. In Equation (5), $\mathbf{B}_1 + \mathbf{N}\mathbf{B}_2\mathbf{N}^T$ is a sparse matrix, depending on the nondeterministic network demands, or the operating point scenario. When dealing with large-scale WDNs, the running speed would be significantly improved while the precision maintains almost the same compared to Equation (3).

Based on Equation (4), the target of inferring the location of potential leaks can be transformed into acquiring the extra nodal demand vector \mathbf{l} . Considering the industry practice, the number of deployed sensors m is generally much less than the number of leak candidates n , that is m should be far less than n . Therefore, the inversion problem of acquiring the extra nodal demands \mathbf{l} from the $m \times 1$ dimensionality-reduced residual vector \mathbf{r}_m is ill-posed due to the column rank deficiency of $\Lambda\mathbf{S}$, in which Λ is the $m \times n$ binary measurement matrix determined by sensor placement. Assuming the occurrence of leaks is independent from each other, it is obvious that the probability of having fewer leaks is higher than that of having more. Therefore, to handle the above problem, a sparse solution with minimal number of nonzero elements in \mathbf{l} is desired. With the theory of compressed sensing, the sparse solution can be obtained by solving the following optimization problem:

$$\mathbf{l} = \arg \min \left\{ \|\mathbf{r}_m - \Lambda\mathbf{S}\mathbf{l}\|_2^2 + \lambda \|\mathbf{l}\|_0 \right\} \quad (6)$$

where λ is a formal variable in the compressed sensing problem for regulating the relative contribution between the two objectives. Without loss of generality, Equation (6) is also appropriate for describing leak localization along a

given time horizon, since it is simply a row augmentation in matrix \mathbf{S} . Note that the leak-induced demand through a constant area hole changes with the nodal pressure (Giustolisi et al. 2008). Therefore, the column of the sensitivity matrix from each time instant should be normalized with respect to the square root of the corresponding nodal pressure. The optimization problem is apparently nonlinear and NP-hard. While under certain conditions several approximate recovery algorithms are available, this is beyond the scope of this paper and will be studied further in the future.

Compressed sensing based diagnosability analysis

In compressed sensing, diagnosability is a reasonable measure for evaluating whether the measurements contain enough information for exact recovery of every possible sparse vector \mathbf{l} , regardless of the practicality of the reconstruction scheme. So far, the diagnosability criterion has been under extensive research and is manifested as some requirements on $\Lambda\mathbf{S}$ (Bastani et al. 2016). As the leak sensitivity matrix \mathbf{S} is specific according to the network topology and operating point scenario, a better measurement matrix Λ is thus desired, corresponding to a better sensor placement scheme that leads to better compressed-sensing reconstruction performance.

The restricted isometry property (RIP) and mutual coherence are two common criteria for evaluating the diagnosability in compressed sensing theory. Compared with RIP, the coherence is a more computationally effective metric. The mutual coherence μ of matrix $\tilde{\mathbf{S}} = \Lambda\mathbf{S}$ is the largest absolute and normalized inner product between different l_2 -normalized columns, formulated as follows:

$$\mu(\tilde{\mathbf{S}}) \triangleq \max_{1 \leq i \neq j \leq n} \frac{|\langle \tilde{\mathbf{S}}_i, \tilde{\mathbf{S}}_j \rangle|}{\|\tilde{\mathbf{S}}_i\| \cdot \|\tilde{\mathbf{S}}_j\|} \quad (7)$$

in which $\tilde{\mathbf{S}}_i$ is the i th column of matrix $\tilde{\mathbf{S}}$. In general, the mutual coherence measures the worst cosine similarity between the columns, a value that exposes the solution's vulnerability (Elad 2007). As such, a large value of μ would potentially confuse any recovery algorithms. Therefore, matrix Λ should be optimized, in order to minimizing the mutual coherence of $\Lambda\mathbf{S}$.

An alternative way to understand the mutual coherence is through the adoption of the Gram matrix $\mathbf{G} = \bar{\mathbf{S}}^T \bar{\mathbf{S}}$, where $\bar{\mathbf{S}}$ is the column-normalized matrix. Apparently, the mutual coherence is equal to the off-diagonal entry g_{ij} with the largest magnitude, that is $\mu = \max(g_{ij})$ ($i \neq j$). Notably, the pure mutual coherence only reflects the worst case, and ignores the average behavior of all other columns. Therefore, an average measure of mutual coherence instead is introduced, as a more reasonable objective function (Elad 2007). Predictably, this is more appropriate to improve the actual performance of the leak localization procedure. The average mutual coherence is defined as follows:

$$\mu_{avg}(\tilde{\mathbf{S}}) = \frac{\sum_{1 \leq i \neq j \leq n} |g_{ij}|}{n(n-1)} \quad (8)$$

To put it simply, the target is to minimize the criterion $\mu_{avg}(\Lambda \mathbf{S})$ with respect to the binary measurement matrix Λ , assuming the matrix \mathbf{S} is known and fixed.

Optimal sensor placement using enhanced binary ABC algorithm

The optimal sensor placement problem can be roughly stated as the choice of m most informative rows of \mathbf{S} out of n available ones such that $\mu_{avg}(\Lambda \mathbf{S})$ is minimized. For the sake of simplicity, an n -dimensional binary vector \mathbf{x} is adopted as a replacement for Λ , and each Λ corresponds to a unique \mathbf{x} :

$$\mathbf{x} = [x_1, \dots, x_i, \dots, x_n] \quad (9)$$

where $x_i = 1$ if a sensor is installed in node i , otherwise $x_i = 0$. Therefore, the optimal sensor placement problem is comes down to a particular $\{0, 1\}$ quadratic knapsack problem, with 'one' and 'zero' respectively denoting the presence and absence of a sensor located at corresponding node:

$$\begin{aligned} \mathbf{x} &= \arg \min \mu_{avg}(\mathbf{x}, \mathbf{S}) \\ \text{s.t. } \mathbf{x}: x_i &\in \{0, 1\}, \sum_{i=1}^n x_i = m \end{aligned} \quad (10)$$

Ozturk et al. (2015) demonstrate that the enhanced binary ABC algorithm outperforms other analogous algorithms in solving the binary knapsack problem. The ABC

algorithm is a population-based metaheuristic technique, motivated by the intelligent foraging behavior of a honey bee swarm, for effectively solving non-convex and non-smooth optimization problem. The standard ABC algorithm can effectively optimize continuous problems, and for taking advantage of the strong find-best ability of the ABC algorithm, some modifications on ABC are required in order to suit the binary optimization for the knapsack problem. Eventually, it should be successfully applied to the sensor placement optimization problem and maximize the overall diagnosability. The kernel ideas of the algorithm are described as follows, and the implementation detail of the algorithm can be found in Ozturk et al. (2015).

In the ABC algorithm, the feasible solution to the optimization problem, or a sensor placement \mathbf{x} in this paper, is denoted as a food source, with the nectar amount indicating the quality or fitness of the solution. Artificial bees are divided into three categories, that is employed, onlooker and scout bees. Each employed bee is corresponding to a food source, responsible for detecting a new food source in the neighborhood and then memorizing the detected food source greedily when the nectar amount of the detected food source is higher than that of the current one. The genetic operators, including two-point crossover and swap (mutation) operators, are integrated into the binary neighborhood searching procedure. First, two-point crossover operators between x_i, x_j, x_k, x_{gbest} are applied to generate eight children food sources, where x_j and x_k are two food sources randomly selected from the population of employed bees and x_{gbest} is the best food source among the population. Then, a swap operator is applied to eight children food sources to produce eight grandchildren food sources. The survival food source among the 16 children and grandchildren food sources is selected as the neighborhood food source. After that, employed bees are then sharing information about the food sources with onlookers. Each onlooker selects the food source probabilistically in a process of 'roulette wheel selection', based on the nectar amount of each food source. Therefore, quality food sources are selected with higher probability, and thus can be fully exploited. Similarly, each onlooker searches for a new food source in the binary neighborhood of its selected food source and moves to it if the nectar amount of the new food source is higher than that of the selected. If the

food source does not improve for a predetermined number of iterations, this food source is deemed to be completely exploited and should be abandoned. These employed bees become scout bees, randomly searching for new food sources in the solution space. In short, exploration and exploitation procedures are carried out together for a robust search for the optimal solution. The flowchart of the enhanced binary ABC algorithm is presented in Figure 1.

RESULTS AND DISCUSSION

This section illustrates the proposed optimal sensor placement strategy with a real-life case study. The network is located in a major city in central China, consisting of two water plants, with 171 nodes and 239 pipes allocated on its distribution mains, shown in Figure 2. In order to acquire an optimal sensor placement under the most general

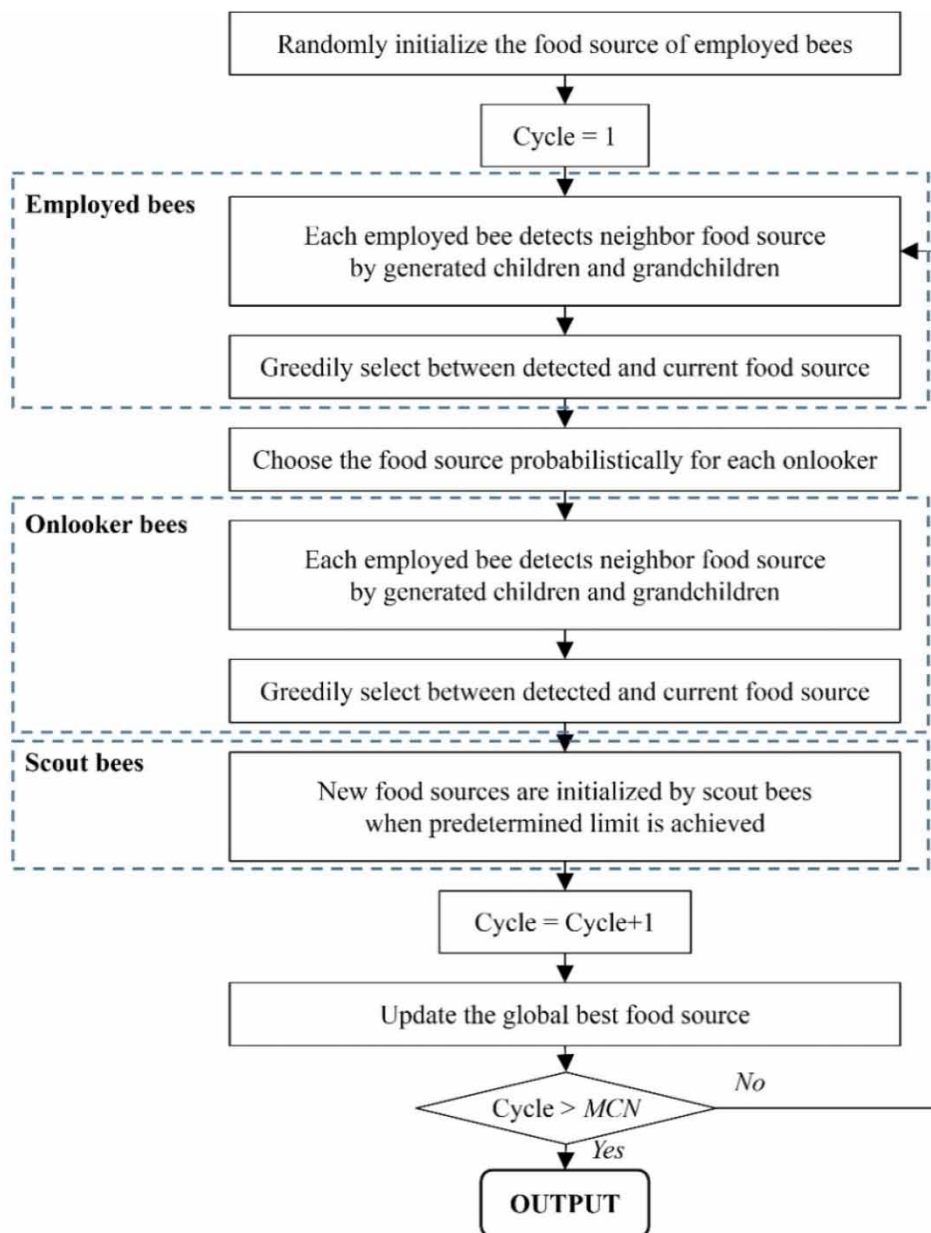


Figure 1 | The flowchart of the enhanced binary ABC algorithm.

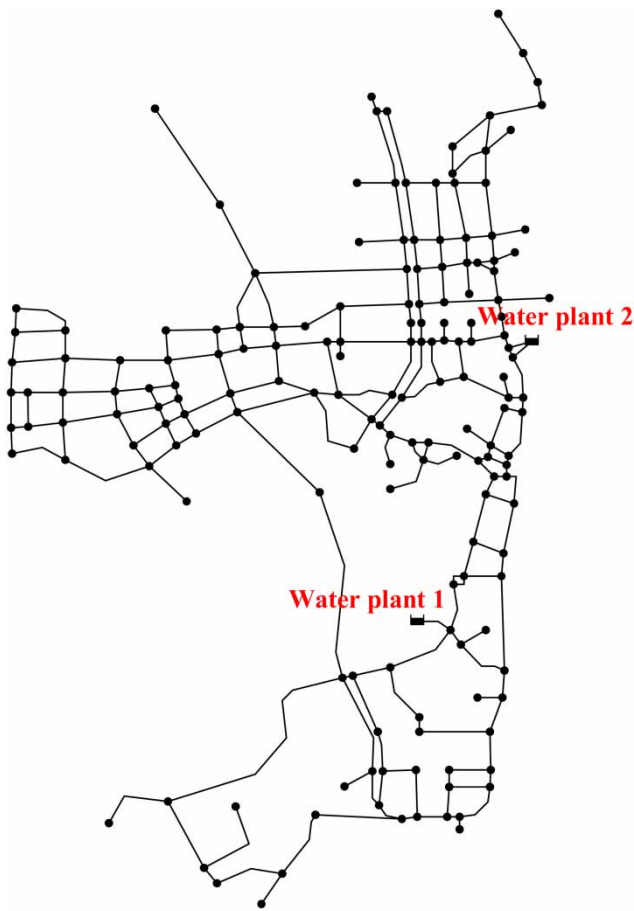


Figure 2 | The layout of the benchmark water distribution network.

situation, the typical nodal base demands and corresponding diurnal curves are previously estimated and provided. In the remainder of this section, the number of pressure sensors required in the network is roughly estimated by a greedy algorithm, and the enhanced binary ABC algorithm is then adopted aimed at intelligently searching for an optimal sensor placement scheme.

In general, the exhaustive search method, which necessarily provides the exact result, is typically computationally prohibitive. In this paper, the use of the greedy worst-out algorithm is explored to determine the proper number of sensors required for the network shown here. The row of \mathbf{S} that would minimize the average mutual coherence maximally is sequentially removed from \mathbf{S} at each iteration. Hence, a set of locations that seems not suitable for sensing potential leaks can be acquired in order. The result is shown in Figure 3. As the figure shows, the redundant information

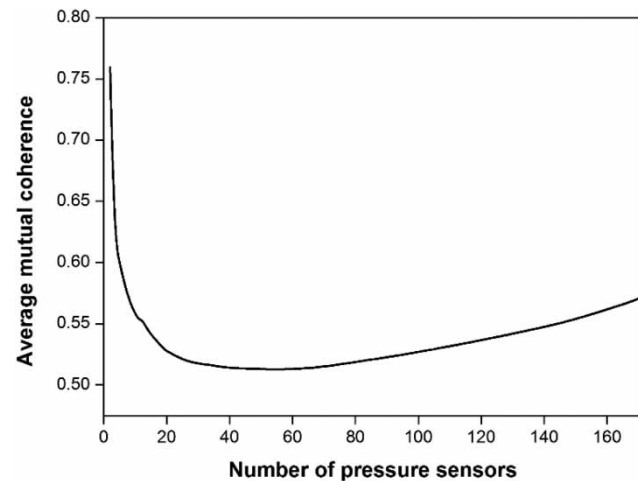


Figure 3 | Average mutual coherence versus the number of deployed sensors.

on potential leaks is gradually eliminated when the number of sensors decreases from 171 to 52. Mathematically speaking, with the decrease of the number of sensors, the rows in \mathbf{S} that are nearly the linear combination of other rows are removed one by one. With a further decrease of the number of sensors from 52, effective information is lost as well. In this case, due to the constraint of sensor costs, the number of deployed sensors is decided to be 26, considering the diagnosability criterion deteriorates only by 1.5% while the costs halve compared to the estimated best result.

The clustering techniques are frequently adopted in the optimal sensor placement searching process to reduce the initial set of candidate sensors, so that the computational complexity can be limited (Sarrate *et al.* 2014; Blesa *et al.* 2016). However, ignoring the differences between sensors in one cluster does not meet the fundamental purpose of the optimal sensor placement problem. Alternatively, heuristic methods are introduced to reduce the computational burden. Compared with the genetic algorithm or particle swarm optimization, the enhanced binary ABC algorithm is inherently more suitable in solving binary optimization problems, such as the optimal sensor placement, which is a particular $\{0, 1\}$ knapsack problem (Ozturk *et al.* 2015). By using the enhanced binary ABC algorithm, the computation requirement should be affordable unless the scale of WDN is too large.

In this case, the enhanced binary ABC algorithm is proven to be a valid option to reduce the unacceptable computing costs of the exhaustive search (the diagnosability

criterion must be evaluated for $\binom{171}{26} \approx 3.8 \times 10^{30}$ times). The initial population for the algorithm is generated randomly. As shown in Figure 4, the global minimum average mutual coherence tends to a constant quickly along with the increasing number of generations. A fairly good characteristic of convergence is provided, which needs only about 150 generations to reach a quasi-optimal average mutual coherence of 0.520218. Some fluctuations appear in the evolution of average and current minimum objective function due to the appearance of scout bees. This indicates that the unchanged food sources are discarded in order to avoid being trapped into the local optimal.

The optimal placement of 26 sensors is shown in Figure 5. For branching pipe network, sensors should be deployed near the end of the branch, furthest away from the inlet, while for a loop pipe network, there exists no clear regularity for the optimal sensor placement, and the result heavily depends on both the topology and normal nodal demands.

A reduction in the number of candidate sensor locations does contribute to speeding up the optimization. According to the above analysis result, one-sidedness considering the degree of diagnosability, all candidate sensors that are located at the intermediate section of branching pipes can be directly eliminated, without much influence on the final sensor placement result. In this case, a total number of 29 candidate sensor locations can be eliminated, leaving only 142 candidate locations to be considered.

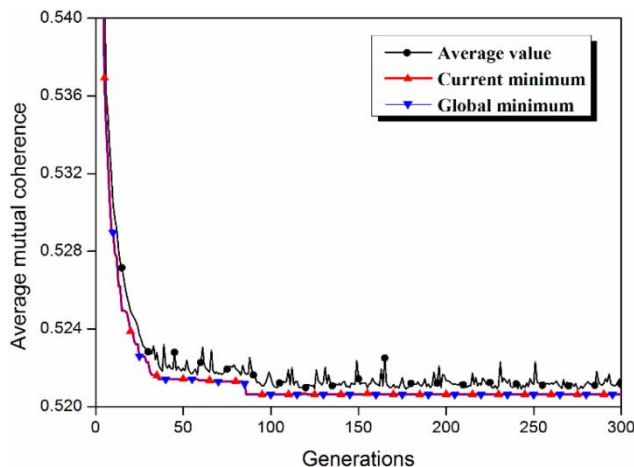


Figure 4 | Evolution progress of the average mutual coherence.

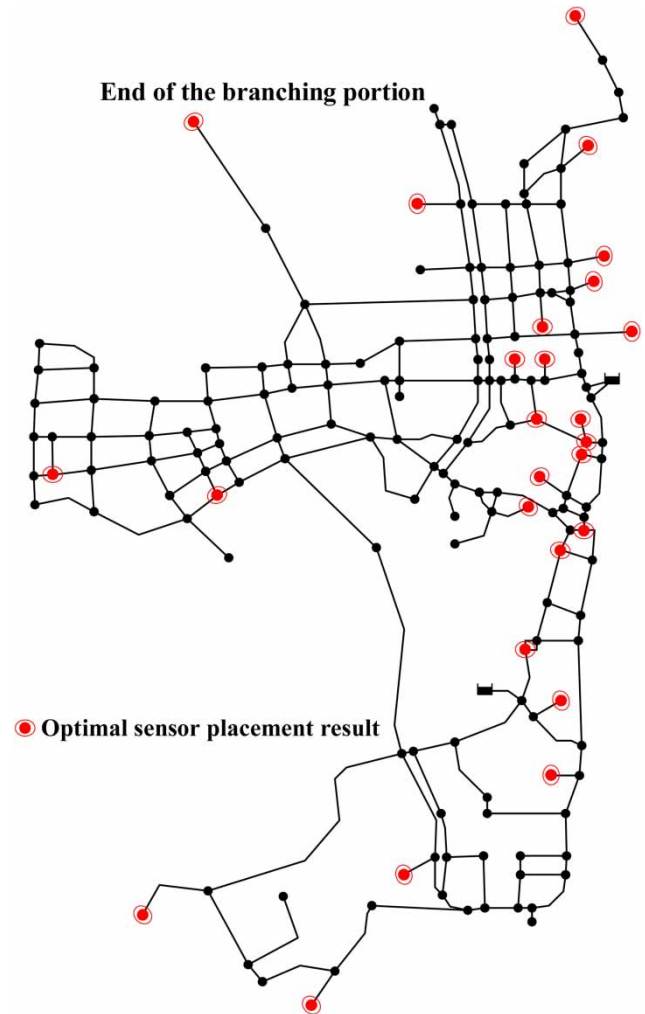


Figure 5 | Sensor placement result for the network.

To date, multiple quantitative indexes have been offered to evaluate the quality of every possible placement scheme. These indexes mainly fall into two major categories. The first category takes the number of overlapping signature domains in the LSS as the sensor placement performance measure (Casillas et al. 2015). That is to say, with proper sensor placement, the Euclidean distances between each pair of leak signatures are expected to exceed the corresponding model output uncertainties. However, due to the introduction of the logical function, the performances of similar sensor placement schemes are almost identical and thus hard to distinguish. The second category lays their basis on the cosine similarities between each two column vectors in leak sensitivity matrix. One typical among them is the leak

isolability index (Blesa et al. 2016):

$$I = \sum_{1 \leq i < j \leq n} 1 - \frac{\tilde{s}_i^T \cdot \tilde{s}_j}{\|\tilde{s}_i\| \cdot \|\tilde{s}_j\|} \quad (11)$$

It is interesting to note that maximizing the leak isolability index is almost equivalent to minimizing the average mutual coherence, and even these two indexes are derived from different perspectives. The value of average mutual coherence and leak isolability index for two sensor placement schemes are provided in Table 1. The sub-optimal sensor placement is obtained by a greedy worst-out algorithm, while the quasi-optimal placement is obtained by an enhanced binary ABC algorithm. These two sensor placement schemes are the same except for one sensor. The result indicates that both these two indexes show favorable distinguishing ability between similar placement schemes.

The proposed method is of guiding significance to deploying pressure instruments in newly built or reformed WDNs. However, in addition to maximizing the degree of diagnosability, the engineering factors are almost equally decisive. Putting the degree of diagnosability aside, in practical applications, the pressure instruments are also preferred to be installed at the points of higher elevation. Because pressure instruments located at higher elevations reduce their requirements for the range of measurements, instruments are most accurate when measuring pressures within 50–75% of the maximum value on the scale. Moreover, the installation and maintenance costs depend on the sensor placement to a large extent as well. In order to provide a good trade-off between the diagnosability, the overall costs of metering, installation and maintenance combined, a node-specific weight can be easily integrated into the sensor placement optimization.

CONCLUSIONS

In this paper, an optimal sensor placement strategy is proposed to enhance the overall performance for model-based

leak detection and localization in WDNs. The model-based leak detection and localization technique is based on a relatively small number of pressure measurements and the associated sensitivity-to-leak analysis. By recovering the sparse representation of a leak-induced nodal demand vector, potential leaks can be detected and located. This paper provides an analytical expression for the leak sensitivity matrix, which contributes to calculation efficiency. According to the theory of compressed sensing, the average mutual coherence is adopted to evaluate whether the measurements contain enough information for identifying potential leaks. An optimization method using the enhanced binary ABC algorithm is proposed for optimal sensor placement by minimizing the average mutual coherence, so that the degree of diagnosability can be maximized.

The proposed method is successfully applied to a real-life network located in a major city in central China. By minimizing the average mutual coherence using an enhanced binary ABC algorithm, a quasi-optimal sensor placement can be guaranteed. Based on the resulting optimal sensor placement obtained using the proposed method, the conclusion is reached that, for branching pipes, when candidate sensors located at the intermediate section own similar or lower elevation and similar or higher costs compared with candidate near the end of the branch, it is recommended that those candidates should be eliminated to reduce the number of candidate sensor locations. Otherwise, a node-specific weight should be integrated into the objective function for a balance between maximizing the degree of diagnosability and minimizing the overall costs.

ACKNOWLEDGEMENTS

This work has been funded by the National Natural Science Foundation of China (No. U1509208; 61573313) and by the Key Technology Research and Development Program of Zhejiang Province (No. 2015C03G2010034).

REFERENCES

- Bakker, M., Vreeburg, J. H. G., Van de Roer, M. & Rietveld, L. C. 2014 *Heuristic burst detection method using flow and pressure measurements*. *J. Hydroinform.* **16** (5), 1194–1209.

Table 1 | The comparison of average mutual coherence and leak isolability index

Index	Average mutual coherence	Leak isolability index
Sub-optimal	0.5206	6967.8
Quasi-optimal	0.5202	6973.6

- Bapat, R. B. 2014 *Graphs and Matrices*. Springer, London, pp. 17–18.
- Bastani, K., Kong, Z., Huang, W. Z. & Zhou, Y. Q. 2016 Compressive sensing-based optimal sensor placement and fault diagnosis for multi-station assembly processes. *IIE Trans.* **48** (5), 462–474.
- Blesa, J., Nejjari, F. & Sarrate, R. 2016 Robust sensor placement for leak location: analysis and design. *J. Hydroinform.* **18** (1), 136–148.
- Casillas, M. V., Garza-Castanon, L. E. & Puig, V. 2014 Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. *J. Hydroinform.* **16** (3), 649–670.
- Casillas, M. V., Garza-Castanon, L. E. & Puig, V. 2015 Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. *Water* **7** (11), 6496–6515.
- Christodoulou, S. E., Gagatsis, A., Xanthos, S., Kranioti, S., Agathokleous, A. & Fragiadakis, M. 2013 Entropy-based sensor placement optimization for waterloss detection in water distribution networks. *Water Resour. Manage.* **27** (13), 4443–4468.
- Colombo, A., Lee, P. & Karney, B. W. 2009 A selective literature review of transient-based leak detection methods. *J. Hydroenviron. Res.* **2** (4), 212–227.
- Elad, M. 2007 Optimized projections for compressed sensing. *IEEE Trans. Signal Process.* **55** (12), 5695–5702.
- Gao, Y., Brennan, M. J., Joseph, P. F., Muggleton, J. M. & Hunaidi, O. 2005 On the selection of acoustic/vibration sensors for leak detection in plastic water pipes. *J. Sound Vib.* **283** (3–5), 927–941.
- Giustolisi, O., Savic, D. & Kapelan, Z. 2008 Pressure-driven demand and leakage simulation for water distribution networks. *J. Hydr. Eng.* **134** (5), 626–635.
- Hao, T., Rogers, C. D. F., Metje, N., Chapman, D. N., Muggleton, J. M., Foo, K. Y., Wang, P., Pennock, S. R., Atkins, S. G., Swingler, S. G., Parker, J., Costello, S. B., Burrow, M. P. N., Anspach, J. H., Armitage, R. J., Cohn, A. G., Goddard, K., Lewin, P. L., Orlando, G., Redfern, M. A., Royal, A. C. D. & Saul, A. J. 2012 Condition assessment of the buried utility service infrastructure. *Tunnell. Undergr. Space Technol.* **28**, 331–344.
- Kang, D. & Lansey, K. 2010 Optimal meter placement for water distribution system state estimation. *J. Water Resour. Plann. Manage.* **136** (3), 1–11.
- Kingdom, B., Liemberger, R. & Marin, P. 2006 *The World Bank: The Challenge of Reducing Non-Revenue Water in Developing Countries—How the Private Sector Can Help: A Look at Performance-Based Service Contracting*. Available from: <https://openknowledge.worldbank.org/handle/10986/17238> (accessed 20 January 2017).
- Laucelli, D., Romano, M., Savic, D. & Giustolisi, O. 2016 Detecting anomalies in water distribution networks using EPR modelling paradigm. *J. Hydroinform.* **18** (3), 409–427.
- Meseguer, J., Mirats-Tur, J. M., Cembrano, G., Puig, V., Quevedo, J., Perez, R., Sanz, G. & Ibarra, D. 2014 A decision support system for on-line leakage localization. *Environ. Model. Softw.* **60**, 331–345.
- Ozturk, C., Hancer, E. & Karaboga, D. 2015 A novel binary artificial bee colony algorithm based on genetic operators. *Inform. Sci.* **297**, 154–170.
- Piller, O., Elhay, S., Deuerlein, J. & Simpson, A. 2017 Local sensitivity of pressure-driven modeling and demand-driven modeling steady-state solutions to variations in parameters. *J. Water Resour. Plann. Manage.* **143** (2), 1–12.
- Pudar, R. S. & Liggett, J. A. 1992 Leaks in pipe networks. *J. Hydr. Eng.* **118** (7), 1031–1046.
- Puust, R., Kapelan, Z., Savic, D. A. & Koppel, T. 2010 A review of methods for leakage management in pipe networks. *Urban Water J.* **7** (1), 25–45.
- Rosich, A., Puig, V. & Cassillas, M. V. 2015 Leak localization in drinking water distribution networks using structured residuals. *Int. J. Adapt. Control Signal Process.* **29** (8), 991–1007.
- Sarrate, R., Blesa, J., Nejjari, F. & Quevedo, J. 2014 Sensor placement for leak detection and location in water distribution networks. *Water Sci. Technol. Water Supply* **14** (5), 795–803.
- Savic, D. A., Kapelan, Z. S. & Jonkergouw, P. M. R. 2009 Quo vadis water distribution model calibration. *Urban Water J.* **6** (1), 3–22.
- Simone, A., Giustolisi, O. & Laucelli, D. B. 2016 A proposal of optimal sampling design using a modularity strategy. *Water Resour. Res.* **52** (8), 6171–6185.
- Steffelbauer, D. B. & Fuchs-Hanusch, D. 2016 Efficient sensor placement for leak localization considering uncertainties. *Water Resour. Manage.* **30** (14), 5517–5533.
- Zhang, Q. Z., Wu, Z. Y., Zhao, M., Qi, J. Y., Huang, Y. & Zhao, H. B. 2016 Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. *J. Water Resour. Plann. Manage.* **142** (11), 1–15.

First received 22 March 2017; accepted in revised form 29 June 2017. Available online 24 August 2017