Multi-model based pressure optimization for large-scale water distribution networks

Yuan Zhang, Shaoyuan Li [*], Yi Zheng, Yuanyuan Zou

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, China

ARTICLE INFO

ABSTRACT

In the operation of urban water distribution networks (WDNs), the supply pressure to users should be kept low while meeting the lower bound to satisfy users' water demand, reduce leakages and power consumption. In this paper, a multi-model based pressure optimization method for large-scale WDNs is proposed, where the pressure of the whole WDNs is represented by the pressure of key nodes and key pump station nodes based on the calculation of pressure transfer relationship and a developed network simplification method, the multiple linear models are constructed to approximate the pressure transfer relationships, and the models between the selected key nodes and key pump station nodes are used for the pressure optimization of the whole WDNs. The proposed methods are operation data based, which overcomes the difficulty in constructing accurate and complicated hydraulic models for large-scale WDNs. The network simplification reduces data requirements in multiple model construction and ensures real-time control. In the simulation case, which is part of Shanghai's WDNs, the supply pressure to users is optimized to lower and more homogeneous value without violating the lower bound by applying the proposed approach, which is significant in reducing water leakages and energy consumption.

1. Introduction

1.1. Motivation and incitement

Water distribution networks (WDNs) supply and transport clean water to industrial and domestic users. Raw water is purified in treatment plants through a combination of physical and chemical processes. Clean water is then pumped into a network of pipes and is delivered to users. WDNs also contain valves, pumps, reservoirs, and other components (Brdys & Ulanicki, 1994). In the operation of WDNs, ensuring water supply safety, that is providing sufficient pressure to satisfy users' water demand, is the primary function of WDNs. In addition, the main concerns in WDNs are water leakage and power consumption, because the typical water leakage levels are still approximately 25%–30% (Araujo, Ramos, & Coelho, 2006), and the power consumption accounts for a high proportion of costs. Nowadays, with the expansion of city sizes and populations, the scale of WDNs is growing. It is especially important for the optimization of large-scale WDNs, because even a small percentage of improvements can yield significant savings of water resources and reductions in power consumption.

The pressure in WDNs is one of the key factors affecting water leakage, power consumption, and water supply safety (Lambert, 2001; Pérez, Puig, Pascual, Quevedo, Landeros, & Peralta, 2011). If the pressure is too high, the system will suffer from more serious water leaks,

higher power consumption, and potentially unbearable pipe burst rates. On the other hand, if the pressure is too low, the system will be unable to satisfy users' water demand. Therefore, the ideal pressure optimization result in WDNs is that the pressure of WDNs is kept low and homogeneous while meeting the lower bound. Due to the complexity of the large-scale WDNs and the time-varying characteristics of the water demand, global scheduling is needed to achieve the goal of pressure optimization.

1.2. Literature review

Many studies on the WDNs' pressure optimization have been reported. Some proposed methods rely on the installation and optimization of pressure reducing valves, such as a two-phase method proposed in Araujo et al. (2006). For WDNs already in use, it is impractical to install pressure reducing valves, so more works focus on optimizing the regulation of existing components in WDNs to optimize the pressure.

The non-linear hydraulic model of WDN is the basis of the pressure optimization, and a few strategies are proposed based on the model. For example, in Ulanicki, Bounds, Rance, and Reynolds (2000), an on-line predictive and feedback control scheme is proposed based on a hydraulic model which takes into account a leakage model; a predictive

---

Nomenclature

$\mathcal{N}_\mathcal{G}$, $\mathcal{N}_\mathcal{P}$	General nodes $\mathcal{N}_\mathcal{G} = \{\mathcal{N}_{\mathcal{G},1}, \ldots, \mathcal{N}_{\mathcal{G},n_\mathcal{G}}\}$ and pump station nodes $\mathcal{N}_\mathcal{P} = \{\mathcal{N}_{\mathcal{P},1}, \ldots, \mathcal{N}_{\mathcal{P},n_P}\}$.
$P_{\mathcal{P},i}$, $P_{\mathcal{PK},i}$, $P_{\mathcal{PNK},i}$	The pressure of $\mathcal{N}_{\mathcal{P},i}$, $\mathcal{N}_{\mathcal{PK},i}$ and $\mathcal{N}_{\mathcal{PNK},i}$.
$P_{\mathcal{G},i}$, $P_{\mathcal{GK},i}$, $P_{\mathcal{GNK},i}$	The pressure of $\mathcal{N}_{\mathcal{G},i}$, $\mathcal{N}_{\mathcal{GK},i}$ and $\mathcal{N}_{\mathcal{GNK},i}$.
d_i	Flow of the ith water demand detection node.
$\mathcal{N}_{\mathcal{GK}}$	$\mathcal{N}_{\mathcal{GK}} = \{\mathcal{N}_{\mathcal{GK},1}, \ldots, \mathcal{N}_{\mathcal{GK},n_{\mathcal{GK}}}\}$, general nodes in the simplified WDN, recorded as key nodes.
$\mathcal{N}_{\mathcal{PK}}$	$\mathcal{N}_{\mathcal{PK}} = \{\mathcal{N}_{\mathcal{PK},1}, \ldots, \mathcal{N}_{\mathcal{PK},n_{\mathcal{PK}}}\}$, pump station nodes in the simplified WDN, recorded as key pump station nodes.
$\mathcal{N}_{\mathcal{GNK}}$	$\mathcal{N}_{\mathcal{GNK}} = \{\mathcal{N}_{\mathcal{GNK},1}, \ldots, \mathcal{N}_{\mathcal{GNK},n_{\mathcal{GNK}}}\}$, $\mathcal{N}_{\mathcal{GNK}} = \mathcal{N}_\mathcal{G} \backslash \mathcal{N}_{\mathcal{GK}}$, recorded as non-key nodes.
$\mathcal{N}_{\mathcal{PNK}}$	$\mathcal{N}_{\mathcal{PNK}} = \{\mathcal{N}_{\mathcal{PNK},1}, \ldots, \mathcal{N}_{\mathcal{PNK},n_{\mathcal{PNK}}}\}$, $\mathcal{N}_{\mathcal{PNK}} = \mathcal{N}_\mathcal{P} \backslash \mathcal{N}_{\mathcal{PK}}$, recorded as non-key pump station nodes.
$n_\mathcal{G}$, $n_\mathcal{P}$	The number of general nodes and pump station nodes.
$n_{\mathcal{GK}}$, $n_{\mathcal{PK}}$	The number of key nodes and key pump station nodes.
$n_{\mathcal{GNK}}$, $n_{\mathcal{PNK}}$	The number of non-key nodes and non-key pump station nodes, $n_{\mathcal{GNK}} = n_\mathcal{G} - n_{\mathcal{GK}}$ and $n_{\mathcal{PNK}} = n_\mathcal{P} - n_{\mathcal{PK}}$.
$\mathbf{D}$	$\mathbf{D} = [d_1, \ldots, d_{n_d}]^T$, water demand.
n_d	The number of water demand detection nodes.
$\mathbf{P}_\mathcal{G}$, $\mathbf{P}_{\mathcal{GK}}$, $\mathbf{P}_{\mathcal{GNK}}$	Vectors composed by the pressure of $\mathcal{N}_\mathcal{G}$, $\mathcal{N}_{\mathcal{GK}}$ and $\mathcal{N}_{\mathcal{GNK}}$.
$\mathbf{P}_\mathcal{P}$, $\mathbf{P}_{\mathcal{PK}}$, $\mathbf{P}_{\mathcal{PNK}}$	Vectors composed by the pressure of $\mathcal{N}_\mathcal{P}$, $\mathcal{N}_{\mathcal{PK}}$ and $\mathcal{N}_{\mathcal{PNK}}$.
$\mathbf{L}_\mathcal{G}$	$\mathbf{L}_\mathcal{G} = [l_{\mathcal{G},1}, \ldots, l_{\mathcal{G},n_\mathcal{G}}]^T$, the lower bound of $\mathbf{P}_\mathcal{G}$.
$\mathbf{P}_{\mathcal{GK}}^S$	New lower bound of $\mathbf{P}_{\mathcal{GK}}$.
$\mathbf{P}_{\mathcal{PK}}^*$, $\mathbf{P}_{\mathcal{PNK}}^*$, $\mathbf{P}_\mathcal{P}^*$	Optimal $\mathbf{P}_{\mathcal{PK}}$, $\mathbf{P}_{\mathcal{PNK}}$ and $\mathbf{P}_\mathcal{P}$.
$\boldsymbol{\Theta}_{i,j}$, $\boldsymbol{\Phi}_{i,j}$, $\mathbf{c}_{i,j}$	Coefficient matrices and constant vectors in models.
t	Optimization time.
t_{max}	Acceptable maximum time in every online optimization moment.
a, b	The number of intervals for $P_{\mathcal{P},i}$ and d_i in model construction.
E_{max}	The acceptable maximum error of constructed models.
$\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$	Error vectors of constructed models.

zone control strategy is proposed in Liu and Li (2016) to keep the pressure in a zone, which could reduce the frequency of actuators' change. In these approaches, the accuracy of the model has a large impact on the performance of the pressure optimization. Moreover, the large-scale characteristics of WDNs are not considered, and the non-linearity of the hydraulic model brings challenges to real-time control for large-scale non-linear optimization problems. Hence, methods aiming to solve the issues caused by the non-linearity and large-scale characteristics are studied.

Some researchers propose strategies to handle the non-linearity in the model directly, and schemes to solve the pressure optimization problem are proposed accordingly. For example, in Germanopoulos, Jowitt, Germanopoulos, and Jowitt (1989), a linear theory method is adopted for the iterative linearization of the non-linear hydraulic model, which is one constraint in the formulated excess pressure minimization problem; based on the linear theory method in Germanopoulos et al. (1989), an evolution program encompassing the principles of evolutionary design and genetic algorithms for pressure regulation is proposed in Savic and Walters (1995). For methods in Germanopoulos et al. (1989) and Savic and Walters (1995), the solution of large-scale optimization problem could meet the requirements of real-time control. However, the optimization performance is still subject to the accuracy of the hydraulic model. Some advanced tools and technologies are also used to overcome the issues caused by non-linearity. For example, in Skworcow, Paluszczyszyn, and Ulanicki (2014), the optimization model is automatically generated in the GAMS language from a hydraulic model in the EPANET format and from other constraint information, a module reduction algorithm is used to simplify the optimization model, and a non-linear programming solver CONOPT in GAMS is used to solve the optimization model; in Nazif, Karamouz, Tabesh, and Moridi (2010), an Artificial Neural Network (ANN) model is trained to replace the hydraulic model based on the data generated by a high precision hydraulic simulator model, and the pressure optimization problem is solved by Genetic Algorithm (GA). For methods in Nazif et al. (2010) and Skworcow et al. (2014), the high precision of hydraulic simulator model is the basis for optimization performance improvement.

Some researchers propose network simplification methods to reduce the complexity of the hydraulic model so that the real-time control requirements can be met. In Anderson and Al-Jamal (1995), a simplification method using a parameter-fitting approach is proposed. The layout of the final simplified network is specified a priori; then, pipe conductances and demand distribution are determined by maximizing the fitness between the simplified and the full system performances. In Ulanicki, Zehnpfund, and Martinez (1996), a mathematical model for the hydraulic aggregation of WDNs is presented. The approach computes a network model equivalent to the original system with fewer components by analytical elimination of system components. The reduced nonlinear model preserves the nonlinearity of the original model and approximates the original system within a wide range of operating conditions. In Perelman and Ostfeld (2012), a genetic topology based simplification method, which relies on clustering, is proposed. Strongly and weakly connected sub-graphs are determined using the depth first search and breadth first search graph algorithms.

Based on the previous review, the models used in the optimization are based on hydraulics of WDNs (Germanopoulos et al., 1989; Savic & Walters, 1995; Ulanicki et al., 2000) or data from hydraulic simulator model (Nazif et al., 2010; Skworcow et al., 2014), and the simplification methods use the information of system components (Anderson & Al-Jamal, 1995; Ulanicki et al., 1996) or the topology of WDNs (Perelman & Ostfeld, 2012). In order to construct an accuracy hydraulic model or hydraulic simulator model and apply the network simplification methods, the detailed and accuracy topology information and parameter information for every component need to be obtained or identified (Brdys & Ulanicki, 1994). However, for large-scale WDNs, it is time-consuming or even impossible to meet this requirement in practice because the information recorded by the regulatory agency may be inaccurate, and there are not enough sensors to get data used in parameter identification for all components. Thus, an efficient and practical method for modeling and the corresponding pressure optimization strategy still need to be studied.

1.3. Contribution and paper organization

Along with the development of information technology, the Supervisory Control and Data Acquisition (SCADA) system is applied in the supervisory control system for WDNs (Cembrano, Wells, Quevedo, Pérez, & Argelaguet, 2000). It receives, monitors and records sensor

data about the state of WDNs, such as pressure, flow, water volumes, etc. Researches have been undertaken to develop and apply data-based methods in WDNs. For example, a data-mining-based asset model of WDNs is constructed in Babovic, Drécourt, Keijzer, and Hansen (2002), a dynamic ANN is used to forecast water demand in Ghiassi, Zimbra, and Saidane (2008), and a leakage detection and localization method is developed based on pressure sensor data in Srirangarajan, Allen, Preis, Iqbal, Lim, and Whittle (2013). These ideas inspire the study of practical and efficient pressure optimization algorithms based on operation data.

The main contribution in this paper is that a multi-model based optimization scheme is proposed for the pressure optimization of WDNs, where the pressure delivered to users is optimized by adjusting pump station's outlet pressure. In this method, the pressure of the whole WDNs is represented by the pressure of the selected key nodes and key pump station nodes based on the calculation of the constructed pressure transfer models and the developed network simplification method, where the multiple linear models are used to approximate the pressure transfer relationships as it is an effective method to deal with nonlinearities (Baruch, Lopez, Guzman, & Flores, 2008; Pottmann, Unbehauen, & Seborg, 1993). The pressure transfer models between the key nodes and key pump station nodes are used for the pressure optimization of WDNs. In the proposed network simplification and model construction methods, detailed component and topology information is not needed, only the operation data are used, so the proposed strategy could be easily and quickly implemented in practice for large-scale WDNs.

The remainder of this paper is organized as follows. First, the structure and the control system structure of WDNs are described in Section 2. The issues in pressure optimization of large-scale WDNs is also further clarified in this section. Then the algorithm of the proposed multi-model based pressure optimization approach is presented in Section 3. One part of Shanghai's WDNs, which is one metropolis in China, is taken as the simulation case in Section 4, where the proposed method is tested and compared with the existing experience-based method and particle swarm optimization (PSO) algorithm. The simulation results are discussed in Section 5. Finally, the conclusions are drawn in Section 6.

2. Description of WDNs and issues in pressure optimization

2.1. Structure of WDNs

A simplified schematic diagram of a typical WDN is shown in Fig. 1. Treatment plants provide clean water to the interconnected pipe network. Clean water goes through the pipe network, and then it is delivered to users and consumed by users. The pressure in pipes is boosted by pump stations distributed in the network. The pressure in WDNs ensures that clean water could be transported to all users. Pressure sensors are installed in pump stations' outlets and in some trunk pipes, which are close to users. Let term the locations installed sensors as "nodes". Then the outlets of pump stations are termed as "pump station nodes", $\mathcal{N}_{\mathcal{P}}$. The locations in pipes are termed as "general nodes", $\mathcal{N}_{\mathcal{G}}$. There are also flow sensors to collect the data of users' water demand.

2.2. Control system structure of WDNs

A typical and commonly used WDNs control system structure (Ocampo-Martinez, Barcelli, Puig, & Bemporad, 2012) is shown in Fig. 2, where $\mathbf{L}_{\mathcal{G}}$ is predetermined according to users' pressure requirement. The control system is a hierarchical structure which includes scheduling level and local control level. The scheduling level is responsible for determining $\mathbf{P}_{\mathcal{P}}^{*}$ according to the optimization goal along with the time-varying $\mathbf{D}$. And it sends the scheduled $\mathbf{P}_{\mathcal{P}}^{*}$ to all the local controllers at pump stations. The local controller is responsible for taking the scheduled value, which has been well realized in the operation of WDNs. Therefore, we only focus on the optimization at the scheduling level in this paper without considering the control of pump stations.
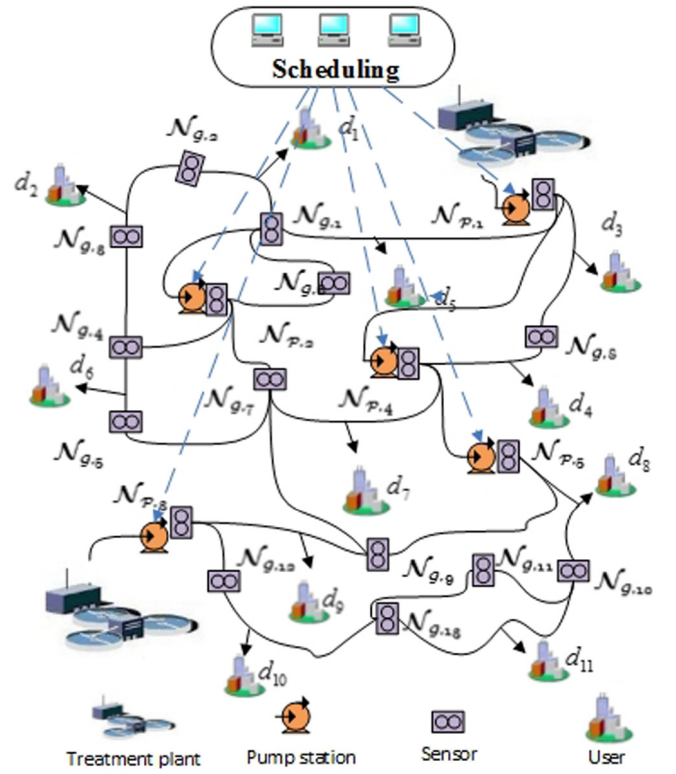

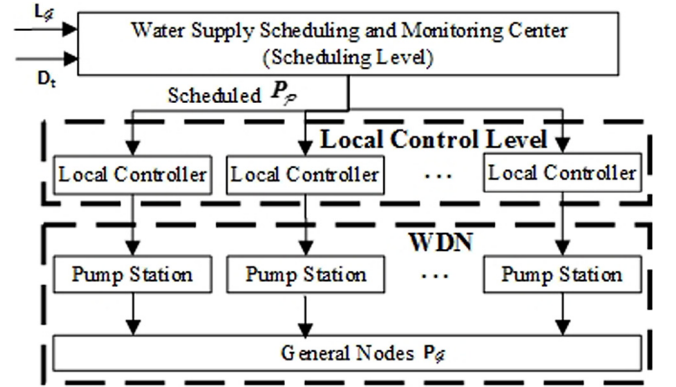**Fig. 1.** Schematic diagram of a typical WDN.

Fig. 2. Control system structure of WDN.

2.3. Issues in pressure optimization of large-scale WDNs

In practical operation, the pressure of general nodes $\mathbf{P}_{\mathcal{G}}$ is used to monitor whether users' water demand is met. Fig. 3 shows one day's $\mathbf{P}_{\mathcal{G}}$ of one part of Shanghai's WDNs. As shown in this figure, the pressure in this area is relevant high, the pressure fluctuates in large range, and there are even situations in which pressure is lower than $\mathbf{L}_{\mathcal{G}}$. This brings high water leakages, high energy consumption and failure to meet users' water demand. Thus, the pressure optimization of WDNs is still a key issue to be considered, and the optimization goal is to keep $\mathbf{P}_{\mathcal{G}}$ homogeneous with lower value. As described in Section 2.1, the controllable variable is $\mathbf{P}_{\mathcal{P}}$. It is necessary to identify the relationship between $\mathbf{P}_{\mathcal{G}}$ and $\mathbf{P}_{\mathcal{P}}$.

As shown in Fig. 1, $\mathcal{N}_{\mathcal{P}}$ and $\mathcal{N}_{\mathcal{G}}$ are connected through pipes. Based on the hydraulics, the nonlinear flow pressure relationship in a pipe can be described by the empirical equation, for instance, the
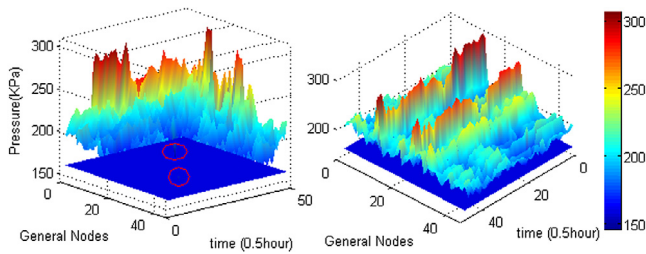
Fig. 3. The pressure distribution of one part of Shanghai's WDN operating with the experience based method.

Hazen–Williams formula (Brdys & Ulanicki, 1994), as follows:

$$p_i - p_j = R_{i,j} q_{i,j} |q_{i,j}|^{0.852}, \tag{1}$$

with $R_{i,j} = (10.67 L_{i,j})/(C_{i,j}^{1.852} D_{i,j}^{4.87})$, where p_i and p_j denote the pressure at both ends of the pipe. $q_{i,j}$ is the flow in the pipe. $L_{i,j}$, $D_{i,j}$ and $C_{i,j}$ denote the pipe length, diameter and roughness coefficient, respectively. In practice, $L_{i,j}$ and $D_{i,j}$ are determined according to the actual information of the pipe, and $C_{i,j}$ can be determined based on experimental method. The flow in pipes is affected by $\mathbf{D}$. Thus, the nonlinear relationship between $\mathbf{P}_P$ and $\mathbf{P}_G$ is time-varying with the change of $\mathbf{D}$ and can be written as

$$\mathbf{P}_G = f(\mathbf{P}_P, \mathbf{D}), \tag{2}$$

where, $f(\cdot)$ represents the nonlinear function. In the case of constant water demand, because pipes are interconnected, the pressure at one general node can be determined by different combinations of several pump stations' outlet pressure. As water demand is time-varying, the variation of flow leads to the change of resistance in pipes, and then the change of the relationship between $\mathbf{P}_G$ and $\mathbf{P}_P$. Therefore, $\mathbf{P}_P$ should be optimized at every optimization moment along with the change of $\mathbf{D}$, where $\mathbf{D}$ can be known by detection.

In practice, due to the difficulty in obtaining accurate parameter and topology information of all the pipes, using hydraulic model in pressure optimization problem is challenging. Therefore, multiple linear models are constructed using operation data to approximate the nonlinear model (2) in this paper. In the modeling for a model with n inputs, the range of each input is divided into m segments, then there are m^n linear models to be constructed. Furthermore, in order to obtain the optimal scheduling results, the optimization problem is solved m^n times. For large-scale WDNs with multiple inputs, the number of linear models could be very large, then a huge amount of data is needed to make identification, and the real-time control cannot be guaranteed. Therefore, it is necessary to reduce the dimension of large-scale WDNs before applying the multi-model method, and the connection between the original system and the simplified system needs to be constructed.

3. Multi-model based pressure optimization method

In order to overcome the issues in pressure optimization of WDNs, a multi-model based pressure optimization method is proposed and shown in Fig. 4. The approach can be divided into offline stage and online stage. As shown in Fig. 4(a), in the offline stage, the general nodes and pump station nodes are simplified respectively to reduce the number of constraints and reduce the number of variables to be optimized in the optimization problem. In addition, three models are constructed offline: (i) multiple linear models between $\mathbf{P}_{GK}$ and $\mathbf{P}_{PK}$ are used in the smaller-scale pressure optimization problem; (ii) the model between $\mathbf{P}_{GNK}$ and $\mathbf{P}_{GK}$ is used to calculate the new lower bound $\mathbf{P}_{GK}^S$ for $\mathbf{P}_{GK}$ to guarantee the satisfaction of $\mathbf{L}_G$; (iii) multiple linear models between $\mathbf{P}_{PNK}$ and $\mathbf{P}_{PK}$ is used to calculate $\mathbf{P}_{PNK}^*$ when $\mathbf{P}_{PK}^*$ has been obtained by solving the smaller-scale pressure optimization problem. As shown in Fig. 4(b), at every optimization moment t in the
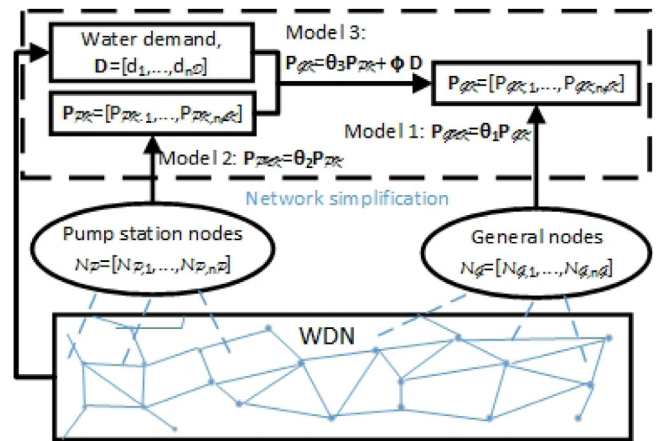
(a) Network simplification and models construction
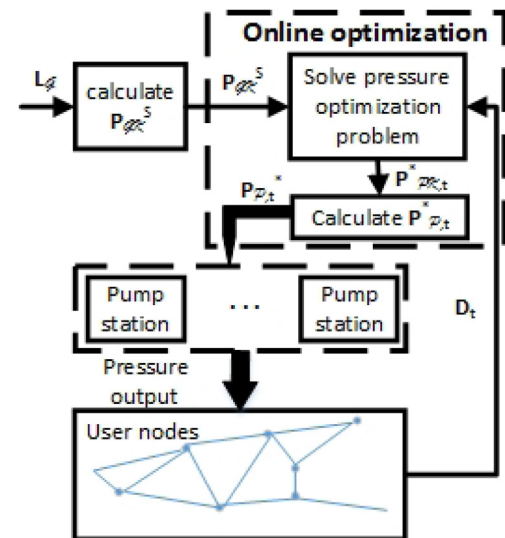
(b) Pressure optimization

Fig. 4. Diagram of the proposed multi-model based pressure optimization approach.

online stage, the smaller-scale pressure optimization problem is solved iteratively for every linear model with the new lower bound $\mathbf{P}_{GK}^S$ to obtain $\mathbf{P}_{PK}^*$. Then $\mathbf{P}_{PNK}^*$ is calculated, and $\mathbf{P}_P^*$ are applied to the original WDN at last. In subsequent part of this section, the proposed network simplification algorithm is described first. Next, the pressure transfer models are constructed. Then the optimization problems used in the proposed pressure optimization scheme are illustrated. The algorithm of the proposed multi-model based pressure optimization method is summarized at last.

3.1. Network simplification algorithm

The goal of network simplification is to determine $\mathcal{N}_{KG}$ and $\mathcal{N}_{KP}$. In order to obtain good optimization results, $\mathbf{P}_{GK}$ and $\mathbf{P}_{PK}$ should reflect the pressure of the original WDN as much as possible, then the accuracy of constructed models could be guaranteed. Considering that the historical pressure data for all nodes are available, we propose a practical network simplification method based on the widely used singular value decomposition (SVD) (Kong, Goodwin, & Seron, 2015; Signoretto, Cevher, & Suykens, 2013; Tian, Ma, Lu, & Wang, 2015) because the characteristics of the historical pressure data are uncertain, and the use of SVD is not limited by the type of matrix.

In the pressure data collection, set the same sampling interval for every node and continuously sample m times. Then, the pressure data of n nodes constitute an $m \times n$ matrix $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_n]$, where $\mathbf{m}_i$ is the vector composed with the pressure data of the ith node. Perform SVD on $\mathbf{M}$, and obtain

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T, \tag{3}$$

with $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_k]_{m \times k}$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]_{n \times k}$ and singular value matrix $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$, where $\sigma_1 > \cdots > \sigma_k > 0$. $\mathbf{V}^T$ is the transpose of $\mathbf{V}$. In SVD, the larger the singular value σ_i, the more the information on $\mathbf{M}$ contained in $\mathbf{u}_i$. To ensure that no key information is missing, all the left vectors $\mathbf{u}_i$, $i \in \mathbb{I}_{1:k_{\mathcal{K}}}$ whose corresponding singular values are significantly greater than the other singular values should be taken as key vectors, and the minimum number of key vectors is obtained and recorded as $k_{\mathcal{K}}$. The number of key vectors will be adjusted subsequently based on real-time control requirements and model's accuracy. Considering that the nodes corresponding to key vectors do not exist in the actual system, a procedure to determine the key nodes which are most similar to the key vectors is illustrated next.

Here, the pressure data of each node compose a vector, and the angle between two vectors is used to measure the similarity between them (Jones & Furnas, 1987). For each key vector $\mathbf{u}_i$, the angle between $\mathbf{u}_i$ and the actual node with vector $\mathbf{m}_j$ can be calculated as

$$\alpha_{\mathbf{u}_i, \mathbf{m}_j} = \arccos\left(\frac{\mathbf{u}_i \mathbf{m}_j}{|\mathbf{u}_i||\mathbf{m}_j|}\right). \tag{4}$$

Then the actual vectors which are most similar to $\mathbf{u}_i$, $i \in \mathbb{I}_{1:k_{\mathcal{K}}}$ are obtained by solving

$$\mathbf{n}_i = \arg\min_{\mathbf{m}_j} \arccos\left(\frac{\mathbf{u}_i \mathbf{m}_j}{|\mathbf{u}_i||\mathbf{m}_j|}\right), i \in \mathbb{I}_{1:k_{\mathcal{K}}}, \forall j \in \mathbb{I}_{1:n}, \tag{5}$$

which means that the selected actual vectors have the minimum angles with the key vectors. Then the nodes whose pressure vectors are $\mathbf{n}_i$, $i \in \mathbb{I}_{1:k_{\mathcal{K}}}$, are the selected key nodes. The whole algorithm to select key nodes is summarized in Algorithm 1.

Algorithm 1 Network simplification algorithm

Input: Matrix $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_n]$, where the column vector $\mathbf{m}_i$, $i \in \mathbb{I}_{1:n}$, is composed by pressure data of the i-th node.

Output: The selected key nodes.

1: $\mathbf{U}, \Sigma$ and $\mathbf{V} \leftarrow$ apply SVD to $\mathbf{M}$;
2: Determine the number of key nodes $k_{\mathcal{K}}$ based on the singular values;
3: **for** $i = 1$ to $k_{\mathcal{K}}$ **do**
4:     Record $\mathbf{u}_i$;
5: **end for**
6: $\boldsymbol{\alpha} \leftarrow (5)_{1 \times k_{\mathcal{K}}}$;
7: **for** $i = 1$ to $k_{\mathcal{K}}$ **do**
8:     **for** $j = 1$ to n **do**
9:         $\alpha_{\mathbf{u}_i, \mathbf{m}_j} \leftarrow \arccos\left(\frac{\mathbf{u}_i \mathbf{m}_j}{|\mathbf{u}_i||\mathbf{m}_j|}\right)$;
10:         **if** $\alpha_{\mathbf{u}_i, \mathbf{m}_j} < \boldsymbol{\alpha}(i)$ **then**
11:             $\boldsymbol{\alpha}(i) \leftarrow \alpha_{\mathbf{u}_i, \mathbf{m}_j}$;
12:             $\mathbf{n}_i \leftarrow \mathbf{m}_j$;
13:         **end if**
14:     **end for**
15: **end for**
16: Key nodes $\leftarrow$ the nodes whose pressure vectors are $\mathbf{n}_i$, $\forall i \in \mathbb{I}_{1:k_{\mathcal{K}}}$.

This algorithm can be applied to select both $\mathcal{N}_{\mathcal{KG}}$ and $\mathcal{N}_{\mathcal{KP}}$ after obtaining $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$. The appropriate value of $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$ can be obtained by an optimization problem introduced in Section 3.3.

3.2. Models construction

There are three types of models that need to be built. The method based on parameter identification using historical operation data is adopted to construct all of these models. These models could reflect the numerical relationship between variables. The specific form of each kind of model is described next.

As described in Section 2.3, the pressure transfer relationship between two nodes is nonlinear and affected by the change of water demand. Therefore, in all three types of model, water demand $\mathbf{D}$ should be one part of the independent variable. However, since $\mathbf{D}$ is time-varying, if $\mathbf{D}$ is used as a part of model between $\mathbf{P}_{\mathcal{GNK}}$ and $\mathbf{P}_{\mathcal{GK}}$, $\mathbf{P}_{\mathcal{GK}}^S$ needs to be calculated online as long as $\mathbf{D}$ changes. In order to reduce the burden of online computing, in the construction of model between $\mathbf{P}_{\mathcal{GNK}}$ and $\mathbf{P}_{\mathcal{GK}}$, the variable $\mathbf{D}$ is neglected. As this model is used to determine $\mathbf{P}_{\mathcal{GK}}^S$ based on $\mathbf{L}_{\mathcal{G}}$, only one model around $\mathbf{L}_{\mathcal{G}}$ needs to be constructed. The form of this model is selected as

$$\mathbf{P}_{\mathcal{GNK}} = \boldsymbol{\Theta}_1 \mathbf{P}_{\mathcal{GK}} + \mathbf{c}_1, \tag{6}$$

with $\mathbf{P}_{\mathcal{GNK}} = [P_{\mathcal{GNK},1}, \ldots, P_{\mathcal{GNK},n_{\mathcal{GNK}}}]^T$, $\boldsymbol{\Theta}_1 = [\theta_{1,1}, \ldots, \theta_{1,n_{\mathcal{GNK}}}]^T$, $\mathbf{P}_{\mathcal{GK}} = [P_{\mathcal{GK},1}, \ldots, P_{\mathcal{GK},n_{\mathcal{GNK}}}]^T$, $\mathbf{c}_1 = [c_{1,1}, \ldots, c_{1,n_{\mathcal{GNK}}}]^T$. Where $\mathbf{c}_1$ reflects the effect of other variables on $\mathbf{P}_{\mathcal{GNK}}$. In the parameter identification procedure, if $P_{\mathcal{GNK},i}, i \in \mathbb{I}_{1:n_{\mathcal{GNK}}}$ is around $l_{\mathcal{GNK},i}, i \in \mathbb{I}_{1:n_{\mathcal{GNK}}}$, then the data sets are selected to identify $\theta_{1,i}$ and $c_{1,i}, i \in \mathbb{I}_{1:n_{\mathcal{GNK}}}$.

In the smaller-scale pressure optimization problem, multiple linear models between $\mathbf{P}_{\mathcal{GK}}$ and $\mathbf{P}_{\mathcal{PK}}$ are constructed to approximate the nonlinear model between $\mathbf{P}_{\mathcal{GK}}$ and $\mathbf{P}_{\mathcal{PK}}$. The form of the ith model is selected as

$$\mathbf{P}_{\mathcal{GK}} = \boldsymbol{\Theta}_{i,2} \mathbf{P}_{\mathcal{PK}} + \boldsymbol{\Phi}_{i,1} \mathbf{D}^{1.852} + \mathbf{c}_{i,2}, \tag{7}$$

with $\mathbf{P}_{\mathcal{GK}} = [P_{\mathcal{GK},1}, \ldots, P_{\mathcal{GK},n_{\mathcal{GK}}}]^T$, $\boldsymbol{\Theta}_{i,2} = [\theta_{i,2,1}, \ldots, \theta_{i,2,n_{\mathcal{GK}}}]^T$, $\mathbf{P}_{\mathcal{PK}} = [P_{\mathcal{PK},1}, \ldots, P_{\mathcal{PK},n_{\mathcal{PK}}}]^T$, $\boldsymbol{\Phi}_{i,1} = [\phi_{i,1,1}, \ldots, \phi_{i,1,n_{\mathcal{GK}}}]^T$, $\mathbf{D} = [d_1, \ldots, d_{n_d}]^T$, $\mathbf{c}_{i,2} = [c_{i,2,1}, \ldots, c_{i,2,n_{\mathcal{GK}}}]^T$. Where $\mathbf{c}_{i,2}$ reflects the effect of other variables on $\mathbf{P}_{\mathcal{GK}}$. $\boldsymbol{\Theta}_{i,2}$, $\boldsymbol{\Phi}_{i,1}$ and $\mathbf{c}_{i,2}$ are identified by using the data of $\mathbf{P}_{\mathcal{GK}}$, $\mathbf{P}_{\mathcal{PK}}$ and $\mathbf{D}$. The data of $\mathbf{P}_{\mathcal{GK}}$, $\mathbf{P}_{\mathcal{PK}}$ and $\mathbf{D}$ at the same sampling moment compose a data set. At the beginning of the identification, each $P_{\mathcal{PK},j}$, $j \in \mathbb{I}_{1:n_{\mathcal{PK}}}$ is divided into a equal intervals, and each d_l, $l \in \mathbb{I}_{1:n_d}$ is divided into b equal intervals. Therefore, there are a total of $a^{n_{\mathcal{PK}}} b^{n_d}$ combinations of value intervals, that is, a total of $a^{n_{\mathcal{PK}}} b^{n_d}$ models need to be identified. Then all the historical data sets are categorized according to the interval of $P_{\mathcal{PK},j}, j \in \mathbb{I}_{1:n_{\mathcal{PK}}}$ and $d_l, l \in \mathbb{I}_{1:n_d}$. Finally, $\boldsymbol{\Theta}_{i,2}$, $\boldsymbol{\Phi}_{i,1}$ and $\mathbf{c}_{i,2}$, $i \in \mathbb{I}_{1:a^{n_{\mathcal{PK}}} b^{n_d}}$ are identified by using the categorized historical data sets.

The construction of model between $\mathbf{P}_{\mathcal{PNK}}$ and $\mathbf{P}_{\mathcal{PK}}$ is similar with the construction of model between $\mathbf{P}_{\mathcal{GK}}$ and $\mathbf{P}_{\mathcal{PK}}$. The form of the ith model is selected as

$$\mathbf{P}_{\mathcal{PNK}} = \boldsymbol{\Theta}_{i,3} \mathbf{P}_{\mathcal{PK}} + \boldsymbol{\Phi}_{i,2} \mathbf{D}^{1.852} + \mathbf{c}_{i,3}, \tag{8}$$

with $\mathbf{P}_{\mathcal{PNK}} = [P_{\mathcal{PNK},1}, \ldots, P_{\mathcal{PNK},n_{\mathcal{PNK}}}]^T$, $\boldsymbol{\Theta}_{i,3} = [\theta_{i,3,1}, \ldots, \theta_{i,3,n_{\mathcal{PNK}}}]^T$, $\mathbf{P}_{\mathcal{PK}} = [P_{\mathcal{PK},1}, \ldots, P_{\mathcal{PK},n_{\mathcal{PK}}}]^T$, $\boldsymbol{\Phi}_{i,2} = [\phi_{i,2,1}, \ldots, \phi_{i,2,n_{\mathcal{PNK}}}]^T$, $\mathbf{D} = [d_1, \ldots, d_{n_d}]^T$, $\mathbf{c}_{i,3} = [c_{i,3,1}, \ldots, c_{i,3,n_{\mathcal{PNK}}}]^T$. Where $\mathbf{c}_{i,3}$ reflects the effect of other variables on $\mathbf{P}_{\mathcal{PNK}}$. The division of the interval is the same as before, so there are also a total of $a^{n_{\mathcal{PK}}} b^{n_d}$ models. The identification of $\boldsymbol{\Theta}_{i,3}$, $\boldsymbol{\Phi}_{i,2}$ and $\mathbf{c}_{i,3}$ is also carried out as before.

So far, all the models used in optimization have been built. These models are constructed using historical data offline. Since the operation of WDNs is full of uncertainty, in general, as long as the historical data is sufficient, it is possible to find data sets for every model. Therefore, these models can be successfully obtained.

Remark 1. For the determination of the intervals' number a, it is necessary to comprehensively consider the adequacy of operation data, the model accuracy and the requirements of real-time control. For the adequacy of operation data, if a is too large, for some intervals, it may not be possible to obtain enough data used in the model identification. In order to guarantee real-time control, $(n_{\mathcal{GK}}, n_{\mathcal{PK}})$ is calculated using

a subsequent method to be introduced with constant a. In general, the number of intervals a and $(n_{\mathcal{GK}}, n_{\mathcal{PK}})$ are negatively correlated. So if a is too large, the calculated $(n_{\mathcal{GK}}, n_{\mathcal{PK}})$ is relatively small, and the accuracy of constructed models may be affected. However, if a is smaller, the calculated $(n_{\mathcal{GK}}, n_{\mathcal{PK}})$ is larger, but the accuracy of constructed models is still not easy to forecast. As the models' accuracy is related to the pressure optimization performance, so the accuracy should be tested. Moreover, when the models are not accurate enough, the number of intervals a should be adjusted, and the network simplification procedure is repeated with adjusted a. The details of the test and adjustment procedure are described in the algorithm of the proposed method.

3.3. Optimization problems

There are 3 optimization problems in the multi-model based pressure optimization approach. The first is used to determine $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$, the second is used in calculation of $\mathbf{P}_{\mathcal{GK}}^S$, and the third is the smaller-scale pressure optimization problem to determine $\mathbf{P}_{\mathcal{PK}}^*$. The first two are used in the offline stage, and the third is used in the online stage at every optimization moment.

As $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$ are related to the total calculation time and the accuracy of models, so $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$ should be set as large as possible with constant a while the total calculation time is not bigger than t_{max}. As the calculation time for an optimization problem is related to $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$, define $g(n_{\mathcal{GK}}, n_{\mathcal{PK}})$ as the function to calculate the maximum calculation time for optimization problems with different $n_{\mathcal{KG}}$ and $n_{\mathcal{PK}}$. Then the total number of smaller-scale pressure optimization problems solved in every optimization moment is $a^{n_{\mathcal{PK}}}$. The optimization problem used to determine $n_{\mathcal{GK}}$ and $n_{\mathcal{PK}}$ is defined as $\mathbb{P}_1$ and as follows: $\mathbb{P}_1$:

$$\min_{n_{\mathcal{GK}}, n_{\mathcal{PK}}} \|a^{n_{\mathcal{PK}}} g(n_{\mathcal{GK}}, n_{\mathcal{PK}}) - t_{max}\|^2, \tag{9a}$$

subjects to

$$a^{n_{\mathcal{PK}}} g(n_{\mathcal{GK}}, n_{\mathcal{PK}}) \leq t_{max}, \tag{9b}$$

$$n_{\mathcal{GK}}, n_{\mathcal{PK}} \in \mathbb{I}_{>1}. \tag{9c}$$

The value of t_{max} is determined based on the optimization interval. The maximum calculation time function $g(\cdot)$ can be approximated by experimental method. $g(\cdot)$ varies according to the computing power and algorithm used. Generally, as long as a is not too big, $n_{\mathcal{GK}}$ and $n_{\mathcal{PK}}$ should be larger than the minimum value obtained based on analysis of singular values because the calculation for problem with linear models is fast. With obtained $n_{\mathcal{GK}}$ and $n_{\mathcal{PK}}$, the corresponding key nodes and key pump station nodes can be determined using the proposed network simplification algorithm.

In the operation of WDNs, an official fixed and unified lower bound $\mathbf{L}_{\mathcal{G}}$ ensures the satisfaction of users' water demand, which should be taken as one constraint in the pressure optimization problem. For the smaller-scale pressure optimization problem, the constraints on $\mathbf{P}_{\mathcal{GNK}}$ should be represented by the constraints on $\mathbf{P}_{\mathcal{GK}}$. Therefore, the optimization problem $\mathbb{P}_2$ is defined as follower to calculate $\mathbf{P}_{\mathcal{GK}}^S$, which will be used in the smaller-scale pressure optimization problem. $\mathbb{P}_2$:

$$\min_{\mathbf{P}_{\mathcal{GK}}} \|\mathbf{P}_{\mathcal{GNK}} - \mathbf{L}_{\mathcal{GNK}}\|^2, \tag{10a}$$

subjects to (6) and

$$\mathbf{P}_{\mathcal{G}} \geq \mathbf{L}_{\mathcal{G}}, \tag{10b}$$

where the optimization objective (10a) is to make the $\mathbf{P}_{\mathcal{GNK}}$ as close as possible to $\mathbf{L}_{\mathcal{GNK}}$. Constraint (10b) guarantees that $\mathbf{P}_{\mathcal{G}}$ is higher than $\mathbf{L}_{\mathcal{G}}$, so as to satisfy water demand. The optimal solution of $\mathbb{P}_2$ is $\mathbf{P}_{\mathcal{GK}}^S$.

In the online stage, the data of $\mathbf{D}$ is obtained at optimization moment t, then the following smaller-scale pressure optimization problems $\mathbb{P}_{3,i}, i \in \mathbb{I}_{1:a^{n_{\mathcal{PK}}}}$ are solved iteratively to find $\mathbf{P}_{\mathcal{PK}}^*$. $\mathbb{P}_{3,i}, i \in \mathbb{I}_{1:a^{n_{\mathcal{PK}}}}$:

$$\min_{\mathbf{P}_{\mathcal{PK},t}} (1/n_{\mathcal{GK}})\omega_1\|\mathbf{P}_{\mathcal{GK},t} - \mathbf{P}_{\mathcal{GK}}^S\|^2 + (1/n_{\mathcal{PK}})\omega_2\|\Delta\mathbf{P}_{\mathcal{PK},t}\|^2, \tag{11a}$$

subjects to (7),

$$\mathbf{P}_{\mathcal{PK},t} \in C_i, \tag{11b}$$

$$\mathbf{P}_{\mathcal{GK},t} \geq \mathbf{P}_{\mathcal{GK},t}^S, \tag{11c}$$

where ω_1 and ω_2 are weighting terms, $\Delta\mathbf{P}_{\mathcal{PK},t} = \mathbf{P}_{\mathcal{PK},t} - \mathbf{P}_{\mathcal{PK},t-1}$, C_i is the set that consists the upper and lower bound of $\mathbf{P}_{\mathcal{PK}}$ corresponding to model (7). There are two parts in the objective (11a). The first part makes $\mathbf{P}_{\mathcal{GK}}$ as close as possible to $\mathbf{P}_{\mathcal{GK}}^S$, which corresponds with the goal of reducing leakages and energy consumption. The second part aims to reduce the fluctuation of $\mathbf{P}_{\mathcal{PK},t}$ between two adjacent optimization times, which reduces the impact on the pipes, and then reduces leakage. Constraint (11b) specifies the range of $\mathbf{P}_{\mathcal{PK}}$ and constraint (11c) makes sure that water demand could be satisfied.

3.4. Algorithm of multi-model based pressure optimization method

In the offline stage of the proposed method, the number of intervals a is firstly determined as the maximum value which could guarantee the models' identification with the collected operation data. Next, $n_{\mathcal{GK}}$ and $n_{\mathcal{PK}}$ are calculated with $\mathbb{P}_1$, then $\mathcal{N}_{\mathcal{KG}}$ and $\mathcal{N}_{\mathcal{KP}}$ are determined using proposed Algorithm 1. With the simplification results, the models introduced in Section 3.2 are identified, and the models' errors are tested. If the tested maximum error is larger than a predetermined error E_{max}, then let $a = a - 1$, and repeat the network simplification procedure and model identification procedure accordingly. Otherwise, the simplification results and models are considered as acceptable, and $\mathbb{P}_2$ is solved to determine $\mathbf{P}_{\mathcal{GK}}^S$.

In the online stage, all the pressure optimization problems $\mathbb{P}_{3,i}, i \in \mathbb{I}_{1:a^{n_{\mathcal{PK}}}}$ are solved to determine $\mathbf{P}_{\mathcal{PK}}^*$. Then the value of $\mathbf{P}_{\mathcal{PNK}}^*$ is calculated based on appropriate model in (8). The complete algorithm is summarized in Algorithm 2 and the algorithm flow chart is shown as Fig. 5.

Algorithm 2 Algorithm of multi-model based pressure optimization method

---

Input: Historical operation data of $\mathbf{P}_{\mathcal{G}}$, $\mathbf{P}_{\mathcal{P}}$ and $\mathbf{D}$. Lower bound $\mathbf{L}_{\mathcal{G}}$. Initial value of $\mathbf{P}_{\mathcal{P}}$, a and b. Water demand data $\mathbf{D}_t$. The maximum acceptable calculation time t_{max}. The maximum acceptable error E_{max}.

Output: Optimal pressure distribution of $\mathbf{P}_{\mathcal{G},t}, t \in \mathbb{I}_{>0}$.

1: **Offline stage**
2: $\mathbf{E}_1 \leftarrow (1)_{n_{\mathcal{GNK}} \times 1}, \mathbf{E}_2 \leftarrow (1)_{n_{\mathcal{GK}} \times 1}, \mathbf{E}_3 \leftarrow (1)_{n_{\mathcal{PNK}} \times 1}$;　▷ Vectors used to record the errors of models;
3: **while** $\mathbf{E}_1 > (E_{max})_{n_{\mathcal{GNK}} \times 1}$ or $\mathbf{E}_2 > (E_{max})_{n_{\mathcal{GK}} \times 1}$ or $\mathbf{E}_3 > (E_{max})_{n_{\mathcal{PNK}} \times 1}$ **do**
4:　　$n_{\mathcal{GK}}, n_{\mathcal{PK}} \leftarrow$ Solve $\mathbb{P}_1$;
5:　　$\mathcal{N}_{\mathcal{GK}}, \mathcal{N}_{\mathcal{PK}} \leftarrow$ Apply Algorithm 1 to $\mathcal{N}_{\mathcal{G}}$ and $\mathcal{N}_{\mathcal{P}}$;
6:　　Construct Model (6), (7) and (8) based on Section 3.2;
7:　　Calculate $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$;
8:　　**if** $\mathbf{E}_1 > (E_{max})_{n_{\mathcal{GNK}} \times 1}$ or $\mathbf{E}_2 < (E_{max})_{n_{\mathcal{GK}} \times 1}$ or $\mathbf{E}_3 > (E_{max})_{n_{\mathcal{PNK}} \times 1}$ **then**
9:　　　　$a \leftarrow a - 1$;
10:　　**end if**
11: **end while**
12: $\mathbf{P}_{\mathcal{GK}}^S \leftarrow$ Solve $\mathbb{P}_2$;
13: **Online stage**
14: $t \leftarrow 1$
15: **while** No emergencies and changes in WDNs **do**
16:　　$J_t^* \leftarrow 1000$;　　　　　　▷ The minimum objective value of $\mathbb{P}_3$
17:　　$\mathbf{P}_{\mathcal{P},t}^* \leftarrow \mathbf{0}$;
18:　　Obtain $\mathbf{D}_t$;
19:　　Select models in (7) used in optimization based on $\mathbf{D}_t$;
20:　　**for** $i = 1$ to $a^{n_{\mathcal{PK}}}$ **do**
21:　　　　$J_{i,t}^*, \mathbf{P}_{\mathcal{PK},i,t}^* \leftarrow$ Solve $\mathbb{P}_{3,i}$;
22:　　　　**if** $J_{i,t}^* < J_t^*$ **then**
23:　　　　　　$\mathbf{P}_{\mathcal{PK},t}^* = \mathbf{P}_{\mathcal{PK},i,t}^*$

24:          **end if**
25:      **end for**
26:      Calculate $\mathbf{P}^*_{\mathcal{PNK}}$ using appropriate model in (8) with $\mathbf{D}_t$ and
        $\mathbf{P}^*_{\mathcal{PK},t}$;
27:      Send $\mathbf{P}^*_{\mathcal{P},t}$ to the local controllers in pump stations;
28:      $t \leftarrow t+1$ and waiting for the next optimization moment;
29: **end while**

So far, the algorithm of multi-model based pressure optimization approach has been described. Compared to the approach of using multi-model directly, in the proposed strategy, the number of models that need to be constructed has been reduced significantly, the number of optimization problems that need to be solved online at each optimization moment has been reduced from a^{n_P} to $a^{n_{PK}}$, and the calculation time for one optimization problem decreases because of smaller-scale, which makes the real-time control possible. To further demonstrate the effect of the proposed method for pressure optimization, the simulation results based on a real case, which is a part of Shanghai's WDNs, are presented in the next section.

4. Simulation

4.1. Description of simulation case

A real case in Shanghai, the biggest city of China, is used to perform the simulation. The WDN used in the simulation covers an area of 250 km^2 and supplies clean water for about 2 million people. The trunk pipeline (diameter ≥ 800 mm) structure and the sensor network are shown in Fig. 6. Fig. 6 is drawn using geographic information system (GIS) data provided by relevant authority for research. There are 63 pressure sensors, including 45 general detection sensors and 18 pump station sensors. These nodes are numbered $n1$-$n45$ and $n46$-$n63$, respectively. The number of sensors to detect the water demand of users is 3 in this network. The rationality of the pressure detection sensors' distribution is not considered in this paper.

4.2. Network simplification result

Pressure data of all the general nodes and pump station nodes are collected every 10 minutes for one month (31 days). In total, there are 4464 data points for every node. The raw data are cleaned before use. For the data collected at every node, the missing data are very limited relative to the total amount of data, and generally, no data are missing for more than two consecutive sampling times. So the missing data are added by linear interpolation. As this area is connected with other parts of the WDNs, and all the nodes are collected based on GIS information. There may be some outliers that are not relevant to most nodes in this case. For general nodes, correlation results between each two nodes are analyzed in combination with geographic location. If one node is less relevant to all other nodes and is at the edge of the case, then it is considered as an outlier. For pump station nodes, correlation results between pump station nodes and general nodes are also analyzed in combination with geographic location. If one pump station node is less relevant to all general nodes and is located at the edge of the case, then it is considered as an outlier. Then the outliers in this case are $n14$, $n15$, $n21$, $n32$, $n34$, $n39$, $n41$, $n46$, $n49$, and $n53$. Finally, $n_G = 38$ general nodes, $n_P = 15$ pump station nodes and $n_d = 3$ water demand nodes are retained in this WDN.

Considering the range of collected pressure data for pump station nodes, the initial number of interval a is set to 3. Based on the characteristics of water demand changes, the optimization interval is set to 1 h (Grosso, Ocampo-Martínez, Puig, & Joseph, 2014; Pascual, Romera, Puig, Cembrano, Creus, & Minoves, 2013; Sankar, Kumar, Narasimhan, Narasimhan, & Bhallamudi, 2015), and the corresponding t_{max} is set to 5 min. Then the calculated n_{GK} and n_{PK} are both 9 by solving $\mathbb{P}_1$. Based on Algorithm 1, it obtains that $\mathcal{N}_{GK}$ includes $n6$, $n9$, $n10$, $n13$, $n18$, $n25$, $n27$, $n30$, and $n33$, $\mathcal{N}_{PK}$ includes $n47$, $n51$, $n54$, $n57$, $n58$, $n59$, $n60$, $n61$, and $n63$. These key nodes and key pump station nodes are renumbered as $\mathcal{N}_{GK} = \{\mathcal{N}_{GK,1}, \ldots, \mathcal{N}_{GK,9}\}$ and $\mathcal{N}_{PK} = \{\mathcal{N}_{PK,1}, \ldots, \mathcal{N}_{PK,9}\}$. The distribution of $\mathcal{N}_G$ and $\mathcal{N}_P$ is shown in Fig. 7.
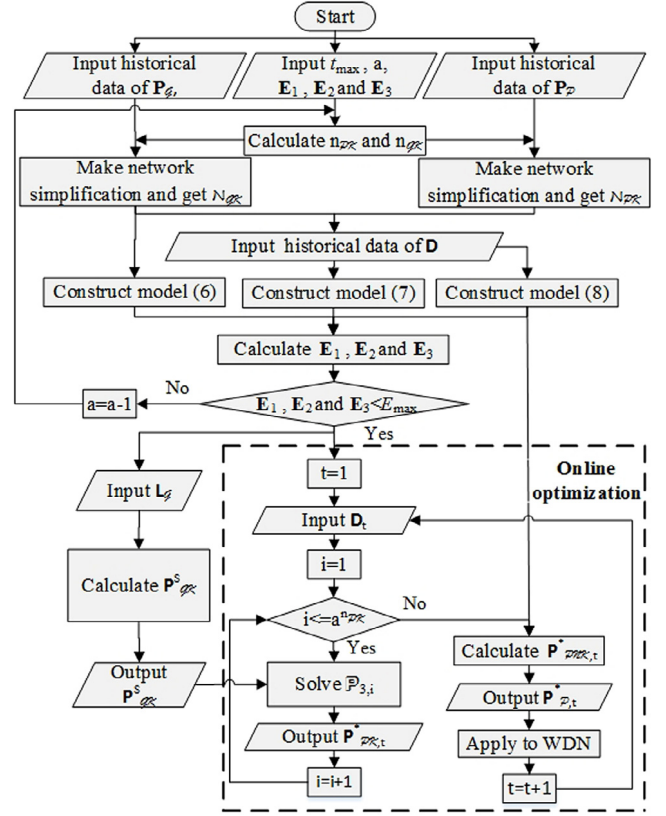

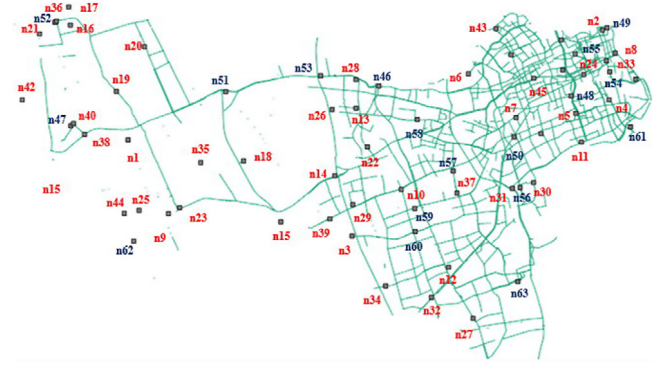**Fig. 5.** Flow chart of the proposed multi-model based pressure optimization approach.

Fig. 6. Trunk pipe line structure ($diameter \geq 800$ mm) and sensor network in the simulation case.
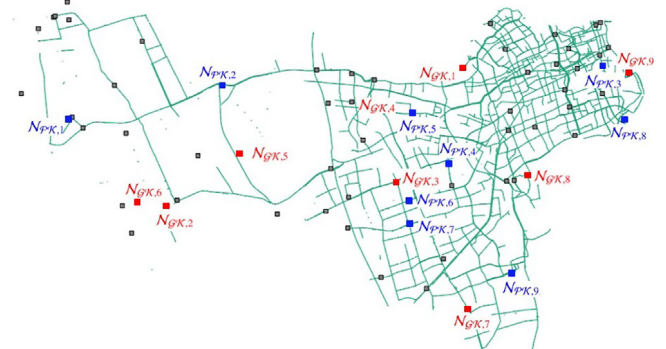

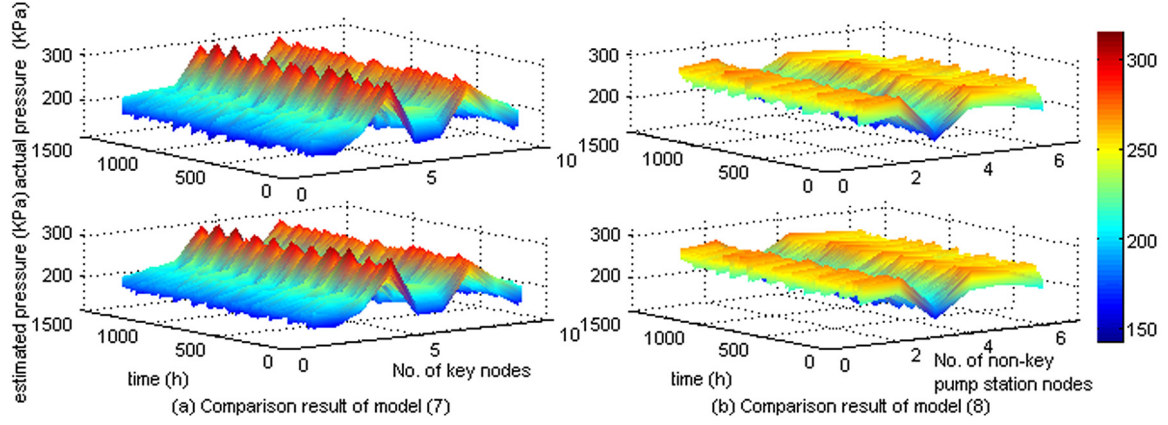**Fig. 7.** Distribution of key nodes and key pump station nodes.

Fig. 8. Comparison results of model (7) and (8).

Table 1
$\overline{\text{MAPE}}$, MAPE_{max}, and MAPE_{min} of model (7) and (8).

Model	(7)	(8)
$\overline{\text{MAPE}}$	1.21%	1.47%
MAPE_{max}	1.91%	2.71%
MAPE_{min}	0.49%	0.74%

Table 2
New lower bound of $\mathbf{P}_{\mathcal{GK}}$.

No.	value (KPa)	No.	value (KPa)	No.	value (KPa)
$P^S_{\mathcal{GK},1}$	165	$P^S_{\mathcal{GK},4}$	211	$P^S_{\mathcal{GK},7}$	227
$P^S_{\mathcal{GK},2}$	171	$P^S_{\mathcal{GK},5}$	166	$P^S_{\mathcal{GK},8}$	210
$P^S_{\mathcal{GK},3}$	165	$P^S_{\mathcal{GK},6}$	165	$P^S_{\mathcal{GK},9}$	165

4.3. Models construction and verification

In the data selection for model identification, the sampling interval is set to 1 s, and the time span is 1 year (365 days). The data of $\mathbf{P}_{\mathcal{G}}$, $\mathbf{P}_{\mathcal{P}}$ and $\mathbf{D}$ are collected. The lower bound $\mathbf{L}_{\mathcal{G}}$ is 160 KPa in Shanghai according to the requirements of the regulatory authorities. For the ith non-key node, if $P_{\mathcal{GNK},i} \leq 180$ KPa, then the data set including $P_{\mathcal{GNK},i}$ and $\mathbf{P}_{\mathcal{GK}}$ is retained for identification of $\theta_{1,i}$ and $c_{1,i}$ in model (6). All the coefficients in model (6) are obtained. In the identification of models in (7) and (8), a and b are set to 3. Then there are $3^{n_{\mathcal{PK}}+n_d}$ models in model (7) and (8). For most of the models, the collected data are sufficient to identify the coefficients. As the constructed models are static models, the data used in identification do not need to be continuous. Then for some interval combinations which could get enough data during this time period, more historical data are used for identification.

The accuracy of model (7) and (8) is verified using 10 day's data. The sampling interval is set as 10 min. Although it is not practical to verify all the constructed models, the validity of this modeling method can be shown from the comparison results. The estimated data by models are compared with the actual data. The comparison results are shown in Fig. 8. We use Mean Absolute Percentage Error (MAPE) to quantify the error of these models. MAPE is calculated as MAPE = $(1/n) \sum_{i=1}^{n} |(\widehat{m_i} - m_i)/m_i|$, where n is the number of sampling times, $\widehat{m_i}$ is the estimated data and m_i is the actual data. The maximum acceptable E_{max} is set as 3%. The average $\overline{\text{MAPE}}$, the maximum MAPE, MAPE_{max}, and the minimum MAPE, MAPE_{min} of the models are shown in Table 1.

As shown in Fig. 8, the estimated data from models are almost the same as the actual operation data. These results can also be illustrated by the MPAE of the models. As shown in Table 1, the MAPE_{max} of these two kinds of models are not more than $E_{max} = 3\%$ and $\overline{\text{MAPE}}$ is a little more than 1%. Therefore, the models are considered as accurate and can be used in the optimization. The high accuracy of models also implies that no key information is missing after network simplification.

4.4. Pressure optimization

The new lower bound of $\mathbf{P}_{\mathcal{G}}$ is obtained by solving $\mathbb{P}_2$, where $\mathbf{L}_{\mathcal{G}}$ is set to 165 KPa, which is slightly higher than the minimum value (160 KPa) required by regulatory authorities, in order to maintain a margin
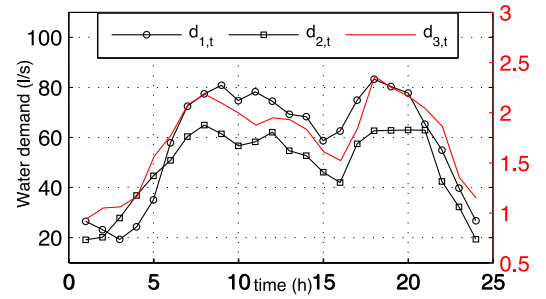

Fig. 9. Water demand in one day.

for model error. The solved $\mathbf{P}^S_{\mathcal{KG}}$ is shown in Table 2 and is used in the online optimization problem $\mathbb{P}_3$.
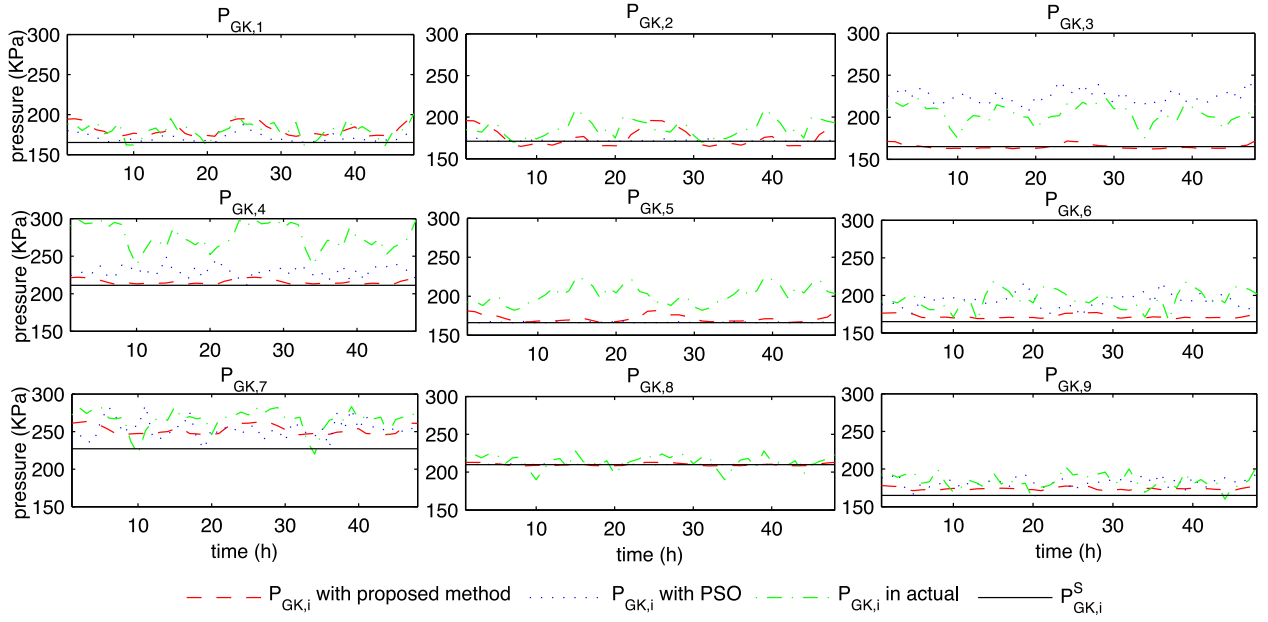
The online simulation is performed in two scenarios: (i) one day's actual water demand data on Oct. 15th 2018 are used; (ii) water data with high fluctuations are used.

4.4.1. Scenario with actual water demand data

In this scenario, the simulation lasts for 2 days (48 optimization moments). The weighting terms in $\mathbb{P}_{3,i}$, $i \in \mathbb{I}_{1:3^9}$, ω_1 and ω_2 are set to 0.7 and 0.3 respectively. It is assumed that the outlet pressures of pump stations can be kept at $\mathbf{P}^*_{\mathcal{P}}$ by local controllers in two adjacent optimization moments. One day's actual water demand data on Oct. 15th 2018, as shown in Fig. 9, are used as feed forward signal $\mathbf{D}_t$ in the simulation, that is, the water demand data repeat twice in the simulation. The simulation is performed as described in Algorithm 2. The calculated $\mathbf{P}^*_{\mathcal{P},t}$ and $\mathbf{D}_t$ are applied to an overall model which contains all the pump station nodes and general nodes. The simulation is performed in Matlab software and the optimization problem is solved using interior point convex algorithm.

In order to illustrate the effect of the proposed scheme, results obtained by applying the PSO algorithm and the experience-based heuristic method in actual operation are used for comparison. The actual pressure data on Oct. 15th 2018 are used for comparison.

The comparison results of $\mathbf{P}_{\mathcal{GK}}$ are shown in Fig. 10. As shown in this figure, the optimized $\mathbf{P}_{\mathcal{GK}}$ with proposed method is more homogeneous than that with other methods.

Fig. 10. $\mathbf{P}_{GK}$ with 3 methods.

(a) Dimensional color map of general nodes with 3 methods

(b) Plane color map of general nodes with 3 methods
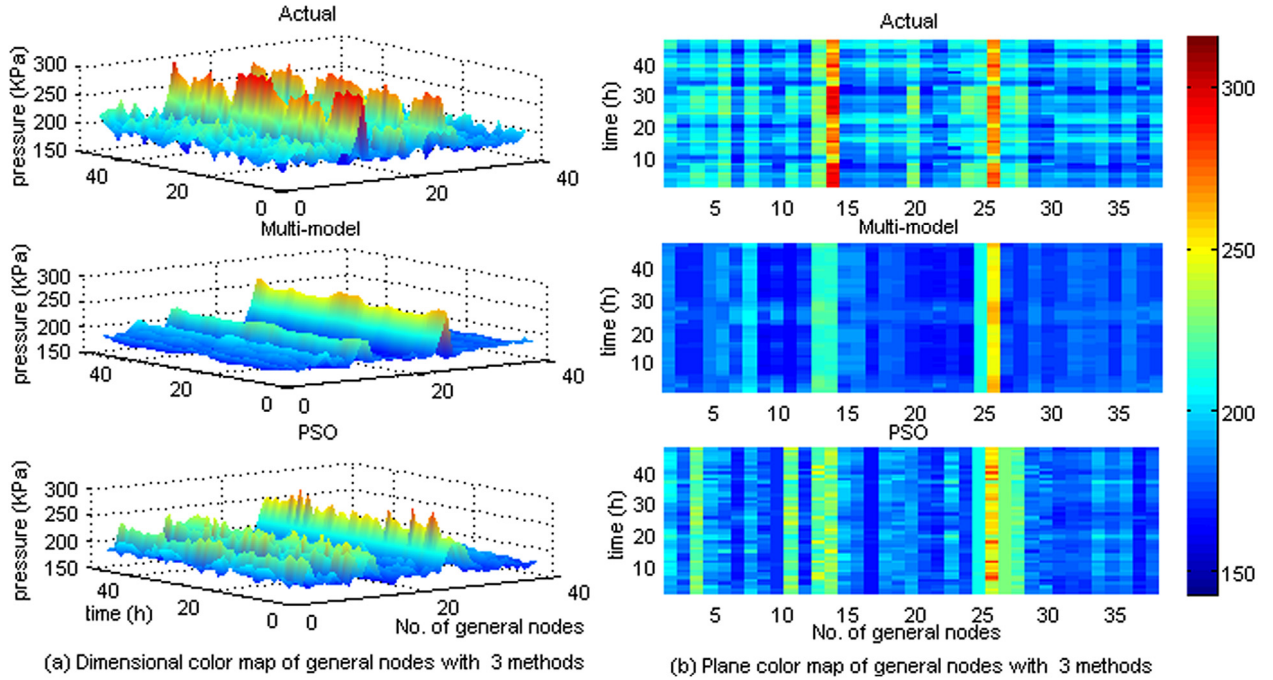
Fig. 11. $\mathbf{P}_{G}$ with 3 methods.

More generally, the comparison results of $\mathbf{P}_{G}$ are shown in Fig. 11. The average, maximum and minimum values of each $P_{G,i}$, $i \in \mathbb{I}_{1:n_G}$ are collected and shown in Fig. 12, the statistical data for all the general nodes are gathered in Table 3, where $\sigma_{\mathbf{P}_G}$ is the standard deviation of all the general nodes' pressure data. As shown in Fig. 11, $\mathbf{P}_{G}$ is the most homogeneous when applying the proposed method ($\sigma_{\mathbf{P}_G} = 3.8$ is the smallest). Moreover, as shown in Fig. 12, with applying the proposed method and PSO, the lower bound could be guaranteed (the minimum value is 160 KPa and 165 KPa), which is better than the actual operation (the minimum value is 155 KPa). Furthermore, when applying the proposed method, the average pressures are lower for most of the nodes as shown in Fig. 12, and the average pressure is reduced in large range (9%, from 200 KPa to 182 KPa). Overall, with applying the proposed

method, users' pressure can be kept lower and more homogeneous, which could reduce energy consumption and improve water supply security.

The pressure of pump station nodes $\mathbf{P}_P$ is shown in Fig. 14, and the average, maximum and minimum values of each $P_{P,i}$, $i \in \mathbb{I}_{1:n_p}$ are collected in Fig. 13. The statistical data are collected in Table 4, where $\sigma_{\mathbf{P}_p}$ is the standard deviation of all the pump station nodes' pressure data.

As shown in Fig. 14 and Table 4, the fluctuate of $\mathbf{P}_P$ is reduced when applying the proposed method, while it increases when applying PSO. The average $P_{P,i}$, $i \in \mathbb{I}_{1:n_p}$ is reduced when applying the proposed method (from $227 KPa$ to $205 KPa$) and PSO (from $227 KPa$ to $196 KPa$). These results are consistent with the results obtained from $\mathbf{P}_G$. As the
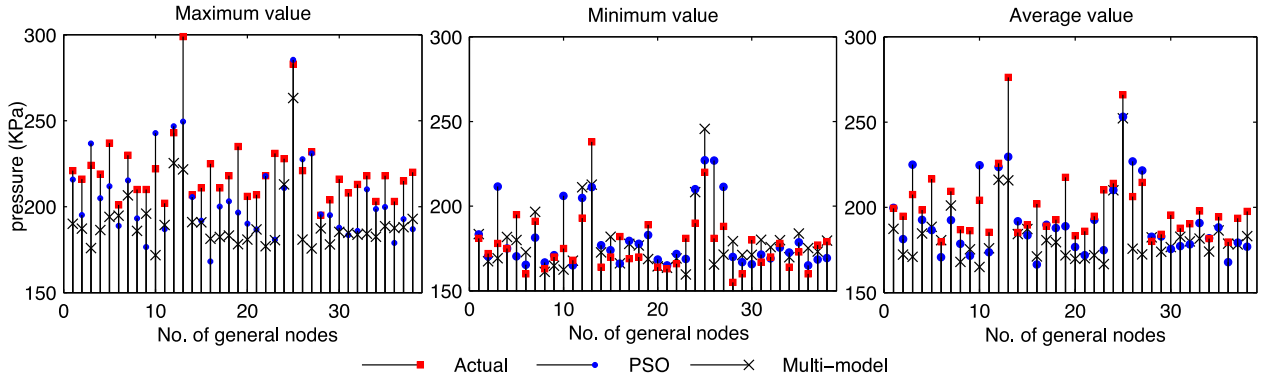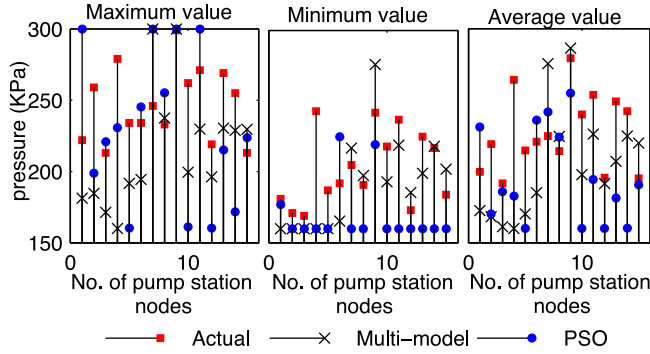
Fig. 12. Average, maximum and minimum values of $\mathbf{P}_{\mathcal{G}}$ with 3 methods.

Fig. 13. Average, maximum and minimum values of $\mathbf{P}_{\mathcal{P}}$ with 3 methods.

Table 3
Statistical data of $\mathbf{P}_{\mathcal{G}}$.

Method	Max (KPa)	Min (KPa)	Ave (KPa)	$\sigma_{\mathbf{P}_{\mathcal{G}}}$
Actual	299	155	200	10.8
Multi-model	263	160	182	3.8
PSO	285	165	179	5.4

Table 4
Statistical data of $\mathbf{P}_{\mathcal{P}}$.

Method	Max (KPa)	Min (KPa)	Ave(KPa)	$\sigma_{\mathbf{P}_{\mathcal{P}}}$
Actual	299	169	227	11.4
Multi-model	300	160	205	7.3
PSO	300	160	196	14.2

Fig. 14. $\mathbf{P}_{\mathcal{P}}$ with 3 methods.

Fig. 15. Water demand with high fluctuation.

cost of power represents a large proportion of the total operation cost of WDNs, the cost could be reduced when applying the proposed method and PSO.

4.4.2. Scenario with high fluctuations in water demand

The water demand data used in this scenario are shown in Fig. 15. All the other settings are the same as in Section 4.4.1. The results obtained by applying the proposed method and PSO are compared. The comparison results of $\mathbf{P}_{\mathcal{G}}$ and $\mathbf{P}_{\mathcal{P}}$ are shown in Figs. 16 and 17 respectively.

As shown in Figs. 16 and 17, when there is a large change in water demand, the solved $\mathbf{P}_{\mathcal{P}}^*$ change immediately, and the corresponding $\mathbf{P}_{\mathcal{G}}$ reach the optimal value when the proposed method is applying. In the case with PSO, the solved $\mathbf{P}_{\mathcal{P}}^*$ is changing when the water demand is constant because some nodes fall into local optimum in the searching, then the corresponding $\mathbf{P}_{\mathcal{G}}$ is also changing.

Moreover, in both scenarios, the time used to solve optimization problems is tested. We solve problem $\mathbb{P}_{3,i}$ 100 times, and get the longest
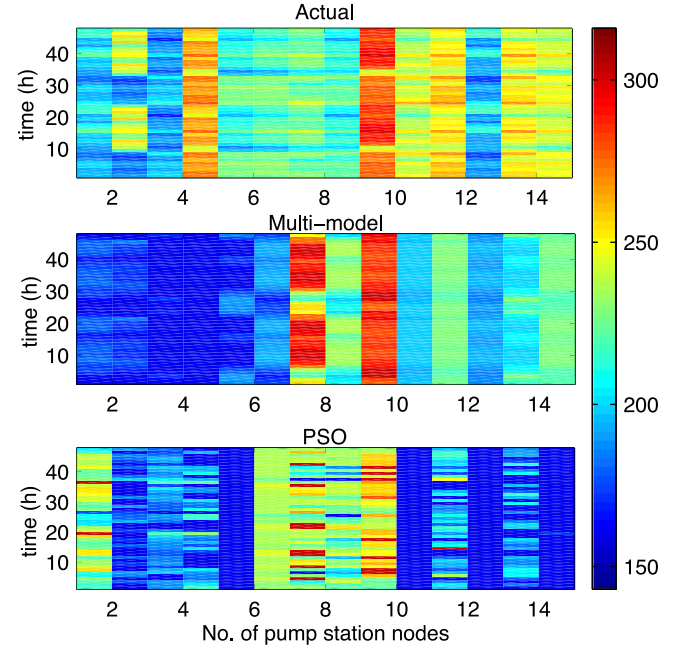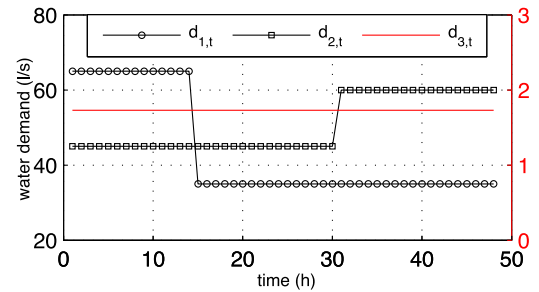
and the average time are 0.03142 s and 0.014426 s, respectively. There are 3^9 problems to be solved at each optimization moment, so the average total time required is 4.73 min and the maximum time required is no more than 10.31 min, then the real time optimization can be achieved.
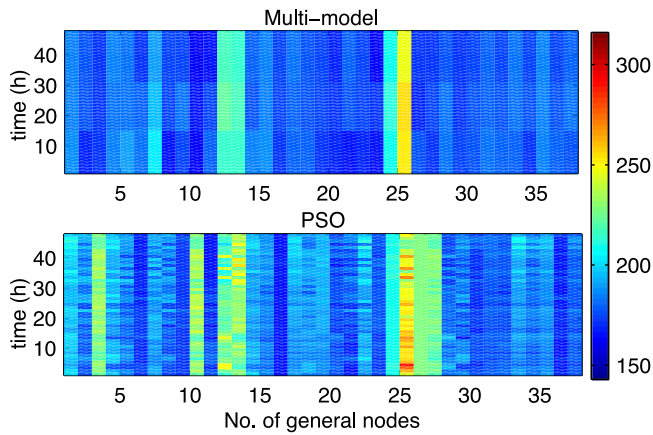
Fig. 16. $\mathbf{P}_{\mathcal{G}}$ under 2 methods with high fluctuation water demand.
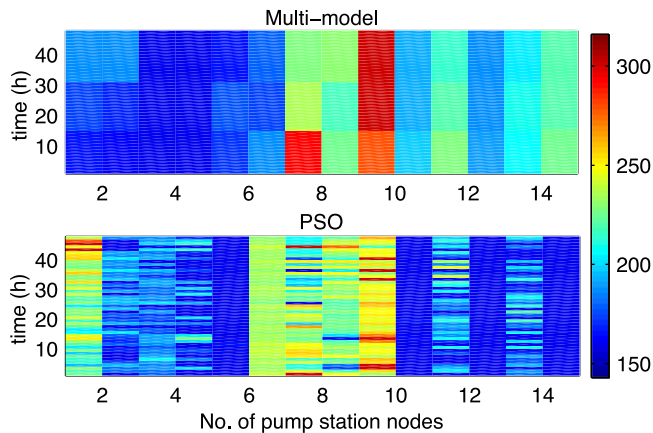

Fig. 17. $\mathbf{P}_{\mathcal{P}}$ under 2 methods with high fluctuation water demand.

5. Discussion

In the simulation, $\mathcal{N}_{\mathcal{GK}}$ and $\mathcal{N}_{\mathcal{PK}}$ are selected based on our proposed data based method. As shown in Fig. 7, $\mathcal{N}_{\mathcal{GK}}$ and $\mathcal{N}_{\mathcal{PK}}$ are distributed over the whole area and not clustered. As the pressure in different sub-areas has their own characteristics from the point of mechanism, this result is reasonable and could prove the effectiveness of the proposed network simplification method from another aspect.

For models construction, with applying the network simplification, the number of models which need to be identified is significantly reduced, so the data required are reduced accordingly, which makes the multi-model based modeling method easier to be applied in practice.

For the calculation time of PSO, it depends on the number of iterations and particles, and it is usually time-consuming for large-scale nonlinear models. For the proposed method, as the constructed models are linear, the calculation time could be significantly reduced, which makes the real-time control is achieved.

6. Conclusions

Pressure optimization is one of the issues in modern urban water distribution networks (WDNs) as it relates to meeting users' water demand, reducing leakages and energy consumption. This paper proposes a method to optimize the pressure of large-scale WDNs by simplifying the pressure transfer models. The proposed method makes use of the operation data in the monitoring system, then it could be efficiently applied to modern large-scale WDNs. The simulation results, which are based on actual operation data collected from part of Shanghai's

WDN, are compared with the experience based heuristic method and PSO algorithm. With applying the proposed strategy, the constructed multiple linear models have acceptable accuracy, the optimized pressure is more homogeneous (i.e the standard deviation is reduced from 10.8 to 3.8) with meeting users' water demand (i.e the minimum pressure is more than the lower bound 160 KPa), the average pressure is reduced by 9% (from 200 KPa to 182 KPa), and the corresponding pump station's outlet pressure is reduced by 9.7% (from 227 to 205) with smaller fluctuations (i.e the standard deviation is reduced from 11.4 to 7.3), which is important in reducing water leakages and energy consumption. Moreover, the ease of application of this method also has great practical significance.

Declaration of competing interest

In the manuscript entitled "Multi-model based Pressure Optimization for Large-scale Water Distribution Networks", we propose a multi-model based scheme used in the pressure optimization of large-scale water distribution networks (WDNs). The work is based on the structure information and operation data of Shanghai's WDNs. We clear that our work has no conflicts of interest with other related work.

References

Anderson, E. J., & Al-Jamal, K. H. (1995). Hydraulic-network simplification. *Journal of Water Resources Planning and Management*, *121*(3), 235–240.

Araujo, L., Ramos, H., & Coelho, S. (2006). Pressure control for leakage minimisation in water distribution systems management. *Water Resources Management*, *20*(1), 133–149.

Babovic, V., Drécourt, J.-P., Keijzer, M., & Hansen, P. F. (2002). A data mining approach to modelling of water supply assets. *Urban Water*, *4*(4), 401–414.

Baruch, I. S., Lopez, R. B., Guzman, J.-L. O., & Flores, J. M. (2008). A fuzzy-neural multi-model for nonlinear systems identification and control. *Fuzzy Sets and Systems*, *159*(20), 2650–2667.

Brdys, M., & Ulanicki, B. (1994). *Operational control of water systems: structures, algorithms and applications*. Prentice Hall.

Cembrano, G., Wells, G., Quevedo, J., Pérez, R., & Argelaguet, R. (2000). Optimal control of a water distribution network in a supervisory control system. *Control Engineering Practice*, *8*(10), 1177–1188.

Germanopoulos, G., Jowitt, P. W., Germanopoulos, G., & Jowitt, P. W. (1989). Leakage reduction by excess pressure minimization in a water supply network. *Proceedings of the Institution of Civil Engineers Part Research & Theory*, *87*(2), 195–214.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2008). Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning & Management*, *134*(2), 138–146.

Grosso, J. M., Ocampo-Martłnez, C., Puig, V., & Joseph, B. (2014). Chance-constrained model predictive control for drinking water networks. *Journal of Process Control*, *24*(5), 504–516.

Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science*, *38*(6), 420–442.

Kong, H., Goodwin, G. C., & Seron, M. M. (2015). A cost-effective sparse communication strategy for networked linear control systems: an SVD-based approach. *International Journal of Robust and Nonlinear Control*, *25*(14), 2223–2240.

Lambert, A. (2001). What do we know about pressure: Leakage relationships in distribution systems? In *Proc. IWA system approach to leakage control and water distribution systems management*.

Liu, D. M., & Li, S. Y. (2016). Predictive zone control of pressure management for water supply network systems. *International Journal of Automation & Computing*, 1–8.

Nazif, S., Karamouz, M., Tabesh, M., & Moridi, A. (2010). Pressure management model for urban water distribution networks. *Water Resources Management*, *24*(3), 437–458.

Ocampo-Martinez, C., Barcelli, D., Puig, V., & Bemporad, A. (2012). Hierarchical and decentralised model predictive control of drinking water networks: Application to barcelona case study. *IET Control Theory & Applications*, *6*(1), 62–71.

Pascual, J., Romera, J., Puig, V., Cembrano, G., Creus, R., & Minoves, M. (2013). Operational predictive optimal control of Barcelona water transport network. *Control Engineering Practice*, *21*(8), 1020–1034.

Perelman, L., & Ostfeld, A. (2012). Water-distribution systems simplifications through clustering. *Journal of Water Resources Planning & Management*, *138*(3), 218–229.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., & Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, *19*(10), 1157–1167.

Pottmann, M., Unbehauen, H., & Seborg, D. (1993). Application of a general multi-model approach for identification of highly nonlinear processes-a case study. *International Journal of Control*, *57*(1), 97–120.

Sankar, G. S., Kumar, S. M., Narasimhan, S., Narasimhan, S., & Bhallamudi, S. M. (2015). Optimal control of water distribution networks with storage facilities. *Journal of Process Control*, *32*, 127–137.

Savic, D. A., & Walters, G. (1995). An evolution program for optimal pressure regulation in water distribution networks. *Engineering Optimization*, *24*(3), 197–219.

Signoretto, M., Cevher, V., & Suykens, J. (2013). An SVD-free approach to a class of structured low rank matrix optimization problems with application to system identification. In *IEEE conf. on decision and control*.

Skworcow, P., Paluszczyszyn, D., & Ulanicki, B. (2014). Pump schedules optimisation with pressure aspects in complex large-scale water distribution systems. *Drinking Water Engineering and Science*, *7*(1), 121–149.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B., & Whittle, A. J. (2013). Wavelet-based burst event detection and localization in water distribution systems. *Journal of Signal Processing Systems*, *72*(1), 1–16.

Tian, Y., Ma, J., Lu, C., & Wang, Z. (2015). Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mechanism and Machine Theory*, *90*, 175–186.

Ulanicki, B., Bounds, P. L. M., Rance, J. P., & Reynolds, L. (2000). Open and closed loop pressure control for leakage reduction. *Urban Water*, *2*(2), 105–114.

Ulanicki, B., Zehnpfund, A., & Martinez, F. (1996). Simplification of water distribution network models. In *Proc., 2nd Int. Conf. on Hydroinformatics* (pp. 493–500). Netherlands: Balkema Rotterdam.