Random Walks Partitioning and Network Reliability Assessing in Water Distribution System

Tianwei Mu[1,2] · Yan Lu[3] · Haoqiang Tan[3] · Haowen Zhang[3] · Chengzhi Zheng[4]

Abstract

A novel network partition model is presented within the water distribution system (WDS). Firstly, random walk community detection (RWCD) is employed to divide WDS into different partitions concerning the average pressure of nodes. Then, network reliability is assessed based on hydraulic reliability estimation (HRE), mechanical reliability estimation (MRE), flow entropy function (FEF), and network resilience (NR), via optimizing boundary pipes by the non-dominated sorting genetic algorithm-II (NSGA-II). Finally, pressure-reducing valves (PRVs) are set to pipes for acquiring optimized partitions. The Open Water Analytics (OWA) toolbox and Matlab-2018b is adopted as a hydraulic calculation tool for these extended period simulations (EPS). Seven cases of WDSs were used to verify the practicability of this model. The results demonstrate that network reliability is improved effectively after partitioning and optimizing.

Keywords Random walk community detection (RWCD) · Network reliability · Water distribution system (WDS) · Multi-objective optimization · Non-dominated sorting genetic algorithm-II (NSGA-II)

✉ Haoqiang Tan
    tanhaoqiang126@126.com

[1] College of Environmental Science and Engineering, State Environmental Protection Engineering Center for Pollution Treatment and Control in Textile Industry, Donghua University, Shanghai 201620, China

[2] College of Municipal and Environmental Engineering, Shenyang Jianzhu University, Shenyang 110168, China

[3] School of Civil Engineering & Architecture, Taizhou University, Taizhou, Zhejiang 318000, China

[4] Guangdong Yuegang Water Supply Co.Ltd, Shenzhen, Guangdong 518021, China

1 Introduction

It is essential to reduce the carbon footprint to save non-renewable energy resources and protect our homeland. About 35 % of the water from the water distribution system (WDS) is wasted in most countries (Pasha and Lansey 2014; Zhang et al. 2016). Abnormal condition, such as the pipe burst, water hammer, causes terrible wastefulness in unreliable WDSs. Therefore, water resources management is now faced with a powerful challenge about reducing loss, and improving reliability.

Network partition is an ideal way to improve the quality, efficiency, and safety of the water supply. District Metered Area (DMA) model was presented in 1989, where WDS is divided into different partitions via opening or closing pressure-reducing valves (PRV) (Ciaponi et al. 2016; Giudicianni et al. 2020). Leakage can be reduced only if PRVs are operated through an effective plan (Ulanicki et al. 2003). With the development of sensor equipment, massive data collection becomes easier. Data mining has been applied as a basis of network partition for improving network reliability in WDS (Wu and Rahman 2017). In recent years, some researchers employed the k-means cluster and support vector machine model to partition the WDS, which optimized the locations of pressure monitoring sensors to better detect leakage (Zhang et al. 2016). Fuzzy c-means algorithm was used to zone hydrologic homogenous regions for the flood management (Kumar et al. 2019; Dehghanian et al. 2019). However, these method needs to reposition some sensors' locations. The multi-scale random walk community detection (RWCD) was proposed to process clustering analysis without sensors reposition, which is a heuristic greedy model with modularity optimization (Lambiotte et al. 2014). But it is difficult to select the number of partitions unless optimization algorithms are employed for searching optimized solutions. Genetic algorithm (GA) as a single objective optimization algorithm, has successfully applied in the optimization of energy consumption (Savic et al. 1997). An extended reduced dynamic programming algorithm similar to GA was proposed to optimize energy consumption and maintenance cost in WDS (Zhuan and Xia 2013). Other algorithms such as the cuckoo search (CS) (Yang 2014), CS-GA hybrid algorithm (Ghodrati and Lotfi 2012) can also be used for optimizing WDSs. However, these above algorithms have a fatal weakness that cannot process multiple objective functions' optimization. Non-dominated sorting genetic algorithm (NSGA) is a multiple objective optimization algorithm (Deb et al. 2002), which utilizes a non-dominated sort model to solve multiple objective problems Non-dominated sorting genetic algorithm-II similar to NSGA but an elitist strategy called crowded-comparison is presented, which improves the effect on variety selection (Srinivas and Deb 1994). The network assessment and leakage detection are feasible ways for accessing network reliability which have been successfully implemented by NSGA-II (Gupta and Kulat 2018). Therefore, NSGA-II is a high potential method for the optimization of the network.

Furthermore, a large-scale hydraulic calculation requires a lot of time and computing resource consumption (Zhang and Shen 2007; Mounce et al. 2017). EPANET software is a calculator for hydraulic calculations, which is programmed by the American Environmental Protection Agency (Rossman 2000), which can fast solve the hydraulic status via the extended period simulation (EPS), and also conveniently construct network partitions via opening or closing valves (Ostfeld et al. 2008). EPANET can be re-developed with other platforms such as MATLAB, C program, and JAVA. The Open Water Analytics (OWA) toolbox is a re-developed program using MATLAB platform (McDonnell et al. 2007), which creates a time-saving and programming convenient way for EPS by matrix operations.

For acquiring an ideal solution of a network partition, this study proposes a novel partitioning model. RWCD is first used to divide the WDS into different partitions. Then, WDS reliability is analyzed concerning five functions. Then, the NSGA-II optimization algorithm is applied to access network reliability by controlling boundary pipes via PRVs. In case studies, seven WDSs were selected to verify the practicability of the method, and the results were presented. Five advantages of this model are following as: (1) the number of partitions can be automatically selected by NSGA-II without manual filters; (2) The average pressures of nodes are used as RWCD's variables that are suitable for representing hydraulic conditions; (3) Clear boundaries of partitions are shown by setting PRVs; (4) The results of optimization can cover network reliability in different abnormal situations. (5) The OWA software improves the calculating speed and readability of programs.

2 Network Partition

RWCD is used to divide WDS into different partitions for the first time in this study. This model was first presented by (Lambiotte et al. 2014) and references the community detection (CD) model (Delvenne et al. 2009, 2013). The CD employs modularity (Fortunato 2010) as an indicator to evaluate the clusters of undirected graph nodes. The most common modularity is the Girvan-Newman (GN) quality function. The equation can be simply expressed as follows:

$$Q_{GN} = \sum_{k=1}^{K} \left[\frac{E_k^i}{2m} - \left(\frac{m_k}{2m} \right)^2 \right] \tag{1}$$

Where K is the number of clusters or called partitions for WDS; k is the kth partition; E_k^i is ith links at the kth partition; m_k is the total number of links at kth partition; m is total number of links.

Then, the GN quality function can be expressed by water demands formalism (Girvan and Newman 2002) for programming convenience.

$$Q_{GN} = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(i,j) \tag{2}$$

Where A_{ij} is the adjacent weight between link i and j and A_{ij} is the weight of the link. if link i connects to link j, $A_{ij} = 1$ or $A_{ij} = 0$; d_i and d_j is the number of links connecting to the node i and j, respectively; $\delta(i,j)$ is the Kronecker delta function.

$$\delta(i,j) = \left\{ \begin{array}{ll} 0 & (i \text{ and } j \text{ belong to the same partition}) \\ 1 & (i \text{ and } j \text{ belong to the different partitions}) \end{array} \right. \tag{3}$$

For convenience, the WDS can be abstracted as an undirected graph. The links are regarded as the edges $E=\{e1,e2,\ldots,em\}$ and the nodes are vertexes $V=\{v1,v2,\ldots,vn\}$. And m is the total number of links the same above. Similarly, n is defined as the total number of nodes. The connectivity of WDS can be described as an adjacent matrix A:

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \tag{4}$$

The sum of jth rows is the total number of links attaching to the node j, and can be expressed by d_j. Thus, the total number of links at each node can be defined $D=\{d1,d2,...dn\}^T$. In real WDS, reservoirs and tanks should be regarded as nodes, same as junctions, while valves, pumps are regarded as links like pipes. The combinatorial graph Laplacian can be expressed by $L = D - A$ or $L = D^{-1/2}LD^{1/2}$ (L is equal or greater than zero).

The random walk is a standard dynamic process at a discrete-time extending on the Markov stability formalism (Lambiotte et al. 2014) that can be defined as:

$$p_{t+1} = p_t\left[D^{-1}A\right] = p_t M \tag{5}$$

Where p is a $1 \times n$ vectors of the probability representing the dynamics converge extent of an undirected graph; p will gradually converge to a stationary distribution solution π ($\pi = d^T/2m$); M is a transition matrix at the time t.

The discrete-time form can be as a derivative expression:

$$\frac{dp}{dt} = -p\left[D^{-1}L\right] \tag{6}$$

Then, the WDS will be decomposed into different partitions. Each partition is encoded by the $n \times c$ indicator matrix H. c is the total number partitions; H is an index matrix (for example, H_{ij} is the adjacent value between link i and j and $H_{ij} = 1$ if link i connects to link j, or $H_{ij} = 0$). If a matrix H is set, a clustered auto-covariance matrix of the diffusion process that evaluates WDS at time t can be expressed (Lambiotte et al. 2014):

$$R_t = H^T\left[diag\left(\pi\right)M^t - \pi^T\pi\right]H \tag{7}$$

Where R_t is a $c \times c$ matrix that reflects the probability of the random walk to divide different partitions for each node; M^t is the transition matrix of time step t, which can be expressed as:

$$M = D^{-1}A = -D^{-1}Lt \tag{8}$$

Provide that the minimum of R_t is at the time s, the Markov stability of the results for partitions is (Martelot and Hankin 2012):

$$r_t = \min_{0 \leq s \leq T} Tr[R_s] \tag{9}$$

Where Tr is the trace of matrix; r_t is the minimum R_t; It is notice that the r_1 ($t = 1$) is the same as trace of Eq. 2:

$$r_1 = Tr\left[H^T\left(\frac{A}{2m} - \pi^T\pi\right)H\right] = TrQ_{GN} \tag{10}$$

Thus, the discrete-time $t = 1$ is a particular form of GN modularity.

For converging to the same stationary point as the discrete-time simple random walk, an improved GN modularity can be generated according to the random walk method (Lambiotte et al. 2014):

$$r_t = Tr\left\{H^T\left[diag(\pi)e^{-tD^{-1}L} - \pi^T\pi\right]H\right\} \tag{11}$$

Where t is the time of random walk.

For fitting this modularity to the WDS, each value v_i at the vector V should be redefined as the pressure of the nodes. The matrix A is the average pressure of two nodes:

$$A_{ij} = \frac{\overline{P}_i + \overline{P}_j}{2} \tag{12}$$

Where $-P_i$ and $-P_j$ are average pressures at the node i and j, respectively, which can be solve as:

$$\overline{P}_i = \sum_{t=1}^{T} \frac{P_i}{T} \tag{13}$$

Where T is the total EPS simulation duration.

The total number of nodes d can be revised as the sum of average pressures defined $a = \{a1, a2, \ldots an\}^T$. a_i is the sum of ith column of matrix A.

Thus, the improved GN modularity for WDS is modified as:

$$r_t = Tr\left\{ H^T \left[\phi e^{-t\psi^{-1}Y} - \eta^T \eta \right] H \right\} = Tr\left[H^T f H \right] \tag{14}$$

Where $\eta = a^T/2M$; M is equal to a half of the sum of all rows and columns of matrix A. ψ is the diagonal matrix of E; Y is a matrix and equal to ψ-A; φ is the diagonal matrixη; f is equal to the $diag(\pi)e^{-t\psi^{-1}Y} - \eta^T \eta$.

H is a unit matrix at the start time. The index matrix assumed node 1 is assigned from partition 1 to partition 2. Then, the elements H_{11} changes from 1 to 0 while H_{12} is changing from 0 to 1. Analogizing that, node i belongs to partition j, and partition k is a neighboring partition. Then, the elements H_{ij} changes from 1 to 0 while H_{ik} is changing from 0 to 1 when node i was assigned to partition k.

The variation of modularity can express as follows:

$$\Delta r_i(j \rightarrow k) = r_i(k) - r_i(j) \tag{15}$$

Where $r_i(j)$, $r_i(k)$ are the modularity when the node i belongs to the partition j and partition k, respectively.

For a better understanding of the RWCD (Blondel et al. 2008), a flow chart of this algorithm is illustrated in Fig. S1 in Supplementary Information.

3 Network Reliability Optimization

WDS has been divided into a different number of partitions after the network partition process with respect to the length of iteration times. However, it is unknown which the best partition is, and the number of boundary pipes to be opened or closed. It is necessary to select a partition plan and explore whether the boundary pipes between partitions are opened or closed. Therefore, NSGA-II is re-developed as a multi-objective function optimization method for solving this problem. In this analysis, five objective functions are employed to optimize WDS for balancing pressure and reducing leakage as below.

3.1 Mechanical Reliability Estimation

The mechanical failure has a severe influence on WDS, including the pipe burst, pump damage, and valve out of control, which will consume a huge carbon footprint to offset these losses. The evaluation of mechanical reliability becomes crying needs.

The pipe burst is the most severe accident in WDS. Mechanical reliability estimation (MRE) is presented by (Su et al. 1987; Sirsant and Reddy 2020), which provides a solution to evaluate the stability of pipes. MRE can calculate pipes break based on the different DMA partitions as the various zones have different probabilities of breaks.

The pipe burst is relative to the lengths and diameters of a pipe. The probability of the pipe break is assumed as:

$$p_i = 1 - e^{-\beta_i} \qquad (16)$$

Where p_i is the break probability of pipe i; β_i is the expected number of failures of pipe i which can be solved as:

$$\beta_i = r_i L_i \qquad (17)$$

Where r_i is the per year per unit length expected number of failures of pipe i which can be expressed as Eq. 18; L_i is the length of pipe i.

$$r_i = \frac{0.68}{d_i^{3.26}} + \frac{2.72}{d_i^{1.31}} + \frac{2.77}{d_i^{3.58}} + 0.04 \qquad (18)$$

Where d_i is diameter of pipe i.

Finally, the MRE equation can be solved as:

$$F = \frac{1}{N_p} \sum_{k=1}^{N_p} p_k \frac{D - S_k}{D} \qquad (19)$$

Where F is probability of pipe break of the integrated WDS; N_p is the number of partitions of the WDS; p_k is the break probability of the partition k expressing as Eq. 20; D is the total water demand of WDS; S_k is the total demand of partition k.

$$p_k = \prod_{i=1}^{N_{knode}} p_i \qquad (20)$$

Where N_{knode} is the number of nodes at the partition k.

3.2 Hydraulic Reliability Estimation

The hydraulic reliability estimation (HRE) method is employed as the second objective function, which is proposed by (Gupta and Bhave 1994). In this method, the water demands of nodes are assumed as unknown, independent values. Then, random water demands are distributed to the nodes of WDS in a normal distribution. For improving the stability, some coefficients are appended to the water demands. Five different cases, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0 are considered as factors, which relate to the water requirement of users (Sirsant and Reddy 2020).

The hydraulic calculating is simulated after the water demands are allocated to nodes. Then, node pressures are obtained. The HRE can be solved by:

$$R_h = R_v F_t F_n \tag{21}$$

Where R_h is HRE factor; R_v is volume reliability factor as Eq. 22; F_t is the time factor as Eq. 24; F_n is the node factor as Eq. 25.

$$R_v = \frac{\sum_i \sum_j q_{ij}^r t_i}{\sum_i \sum_j q_{ij}^b t_i} \tag{22}$$

Where q_{ij}^r is the real water demand of node j at the ith time as Eq. 23; q_{ij}^b is the base water demand of node j at the ith time; t_i is a period of EPS during ith time.

The real water demand is calculated as:

$$q_{ij}^r = \begin{cases} q_{ij}^b & H_{ij}^r \geq H^d \\ q_{ij}^b \left(\frac{H_{ij}^r - H^m}{H^d - H^m} \right) & H^m \leq H_{ij}^r \leq H^d \\ 0 & H^r < H^m \end{cases} \tag{23}$$

Where H_{ij}^r is the real pressure of node j at the ith time; H^m is the minimum require pressure; H^d is the desired pressure.

The time factor F_t can be expressed as:

$$F_t = \frac{\sum_i \sum_j \delta_{ij} t_i}{N_{node} T} \tag{24}$$

Where δ_{ij} is a coefficient. If $q_{ij}^r \geq q_{ij}^b$, δ_{ij} is one otherwise is zero; N_{node} is the number of nodes; T is the total time during the EPS.

F_n can be expressed as the function of node reliability:

$$F_n = \left(\prod_{j=1}^{N_{node}} R_{nj} \right)^{\frac{1}{N_{node}}} \tag{25}$$

Where $\prod R_{nj}$ is the product of all the node reliability.

F_n is the node factor, which represents the average of the node reliability R_n. The node reliability of node j can be solved as:

$$R_{nj} = \frac{\sum_i q_{ij}^r t_i}{\sum_i q_{ij}^b t_i} \tag{26}$$

Where R_{nj} is the node j reliability factor during the EPS.

3.3 Flow Entropy Function

Flow Entropy Function (FEF) was firstly presented by (Shannon 1948), and applied in the WDS as an evaluation index for accessing the stability of fluctuating water demands (Tanyimboh and Templeman 1993; Sirsant and Reddy 2020). FEF is employed as the third

objective function for preventing water hammer. The resiliency index can be calculated using the following equation:

$$S = S_s + \sum_j^{N_{node}} P_j S_j \tag{27}$$

Where S_s is the entropy of water supply sources containing reservoirs and tanks, which can be solved by Eq. 28; P_j is the fraction of inflow the node j and the total water demand as Eq. 30; S_j is the entropy of junctions as Eq. 31.

$$S_s = -\sum_i^{N_s} \sum_j^{N_{sp}} p_{ij} \ln^{p_{ij}} \tag{28}$$

Where Ns is the number of water supply sources; N_{sp} is the number of nodes connected with water supply sources; p_{ij} is the fraction of pipe ij outflow and total outflow from node i to node j, expressing as:

$$p_{ij} = \frac{q_{ij}}{Q_{oi}} \tag{29}$$

Where q_{ij} is the external outflow of from node i to node j; Q_{oi} is the total outflow of node i.

The fraction of inflow can be expressed as:

$$P_i = \frac{Q_i}{Q} \tag{30}$$

Where Q_i is the total internal inflow of node i; Q is the total water demand.

The entropy of junctions can be expressed as:

$$S_j = -\sum_i^{N_{node}} \sum_{j(j \neq i)}^{N_{node}} p_{ij} \ln^{p_{ij}} \tag{31}$$

3.4 Network Resilience

Network resilience (NR) is proposed by (Todini 2000; Prasad and Park 2004), which reflects the internal energy dissipation. Huge power dissipation may be caused when pipe diameters change. NR is employed as the fourth objective function to avoid power loss of WDS. NR can be solved as:

$$I_n = \frac{\sum_{i=1}^{N_{node}} U_i q_{min,i} \left(H_i - H_{min,i}\right)}{\sum_{i=1}^{N_s} q_{R,k} H_{R,k} - \sum_{i=1}^{N_{node}} q_{min,i} H_{min,i}} \tag{32}$$

Where U_i is the uniformity factor of node I, expressing as Eq. 33; $q_{min,i}$ is the minimum water demand at node i during the EPS period; $H_{min,i}$ is the minimum pressure at node i during the EPS period; $q_{R,k}$ is the outflow of reservoir k; $H_{R,k}$ is the pressure of reservoir k;

$$U_i = \frac{\sum_{j=1}^{N_i} d_j}{N_i \max\{d_1, \ldots, d_{N_i}\}} \tag{33}$$

Where d_j is the diameter of link j; N_i is the number of pipes connected with node i.

3.5 Connectivity Optimization

For better water resources management, the number of boundary pipes between two partitions needs to be minimized.

Thus, the number of boundary pipes between two partitions is selected as the fifth objective function.

$$f = \sum_{i=1}^{N_p} \sum_{j=1}^{N_{jnode}} s_{ij} \tag{34}$$

Where s_{ij} is the boundary pipe j at ith partition. When the valve is open, s_{ij} is one, otherwise zero; N_{jnode} is the sum of boundary pipes at the partition i; f is the sum of opened boundary pipes.

s_{ij} is divided into l groups for analysis. For example, with ten boundary pipes and four partitions and each partition has three boundary pipes in WDS, if these pipes status are opened or closed, it can be written in a binary style: [1,1,1,0,1,1,0,1,1,1,0,1]. The first three elements are the first partition boundary pipes which valves in this pipe are all opened. The next three elements are second partition boundary pipes, in which the first valve is closed. The rest can be explained in the same manner. Also, boundary pipes are sort of by their index.

3.6 Non-dominated Sorting Genetic Algorithm-II

NSGA-II is a multiple objective function optimization model utilizing non-dominated sort search method.

In the WDS, the equation of continuity and the conservation of energy should be obeyed. Two constraints need to be considered:

$$H_{min}^{i} \leq H^{i} \leq H_{max}^{i} \tag{35}$$

$$L_{min}^{i} \leq L^{i} \leq L_{max}^{i} \tag{36}$$

Where H_{max}^{i} is the maximum of ith node pressures during the EPS periods; H_{min}^{i} is the minimum of ith node pressures; L_{min}^{i} and L_{max}^{i} are the minimum and maximum of ith tank water levels, respectively.

NSGA-II is employed to optimize five objectives above, and some revised for this algorithm is as follows:

All nodes belong to same partition are imported into a matrix to facilitate management. Secondly, all pipes is expressed as both ends nodes index. For example, L_{ij} is the pipe that ends nodes are i and j. Find the boundary pipes between two partitions by identification of the pipe's ends nodes if belong to different partitions. Next, initial variables are set randomly as the connected states of pipes (opened or closed). For instance, if the number of partitions is 3, and boundary pipes are 5, the connected states matrix is:

$$X = \begin{bmatrix} s_{11} & \cdots & s_{15} \\ \vdots & \ddots & \vdots \\ s_{31} & \cdots & s_{35} \end{bmatrix} \tag{37}$$

Where s_{ij} is the status of pipe j for population (partition) i.

The objective function can be expressed by:

$$\min \{f_1, f_2, f_3, f_4, f_5\} = f(\boldsymbol{x}) \tag{38}$$

Where f_1 is the reciprocal of MRE as Eq. 19; f_2 is the reciprocal of HRE as Eq. 21; f_3 is FEF as Eq. 27; f_4 is the reciprocal of NR as Eq. 32; f_5 is the connectivity function as Eq. 34. $\boldsymbol{x}$ is an independent vector containing water demands of nodes, partition numbers, nodes pressures, flows of pipes, and boundary pipes status.

Using Eq. 37 to solve **Eq. 38**, if the results meet the requirement of Eqs. 35–36, this results will be retained, otherwise function values are set to infinity.

Then, non-dominated sort was processed (Srinivas and Deb 1994) after all populations have been solved. The five objective functions generates a matrix F, which can be defined as:

$$F = \begin{bmatrix} f_1^1 & \cdots & f_5^1 \\ \vdots & \ddots & \vdots \\ f_1^3 & \cdots & f_5^3 \end{bmatrix} \tag{39}$$

Where f_i^j is the value of the objective function f_i in the population of j.

Then, the domination count initial as $n_p=0$ and search one of the best population and set it to F_1. The rank of F_1 is set to one. Next, traverse F_1 and find the domination solutions S_i for each i. Traverse each S_i all the individual n_l and then execute $n_p=n_p$ -1. If $n_l = 0$, save it into a set of Q. Then, $F_2 = Q$, and repeat this process until F_n is a null set.

For instance, assign values for Eq. 39:

$$F = \begin{bmatrix} 3 & 4 & 6 & 6 & 5 \\ 4 & 5 & 7 & 8 & 9 \\ 5 & 5 & 8 & 9 & 10 \end{bmatrix} \begin{matrix} \{1\} \\ \{2\} \\ \{3\} \end{matrix} \tag{40}$$

All objective function values in {1} are less than {2}, {3}. The n_1 is equal to 0 and (2), (3) are combined to the $S_1=[\{2\},\{3\}]$ and $F_1=\{1\}$.

Then, mark each element of {1} to rank (1) Similarly, {2} is less than {3} but larger than {1}, so n_2 is 1 and $S_2 = \{3\}$. The {3} is larger than {1}, {2}, so n_3 is equal to (2) F_1 is not a null set, so traverse F_1. $S_1 = [\{2\},\{3\}]$ and n_2-1 = 1–1 = 0, so Q={2}. Mark element of {2} to rank 2, but the n_3-1 = 2 − 1 = 1, reject it. $F_2 = Q$ and traverse F_2. $S_2 = \{3\}$, $n_3 - 1 - 1 = 2-1-1 = 0$, so Q={3} and mark element of {2} to rank (3) $F_3 = Q$ and traverse F_3. Then, S_3 is a null set, so Q is a null set. $F_4 = Q$ is null. {1}, {2} and {3} are ranked as 1, 2, and 3.

After populations ranking, the crowding distance should be assigned to the whole population. For each population, the values should be sorted in rank, allowing calculation of the crowding distance for the population at each rank and in descending order with distance. The crowding distance equation is as follows:

$$n_d(i) = n_d(i) + [f_m(i+1) - f_m(i-1)] \tag{41}$$

Where n_d is the crowding distance; f_m is the objection function m.

Next, the new rank and distance should be calculated, for each population position to be updated, using copy, cross, and mutation method like simple genetic algorithm.

Then, combine half of the old population with half of the new population, which are all half of higher sort elements.

For instance: assuming the old population is:

$$F = \begin{bmatrix} 3 & 4 & 6 & 6 & 5 \\ 4 & 5 & 7 & 8 & 9 \end{bmatrix}$$

Where the first of row is rank 1 and second is rank 2.

Assuming the new population is:

$$F = \begin{bmatrix} 3 & 5 & 5 & 6 & 5 \\ 6 & 8 & 9 & 8 & 9 \end{bmatrix}$$

Where the first of row is rank 1 and second is rank 2.

Then, combine first of row of old population and first of row of new population and reject others. A new combination can be produced:

$$F = \begin{bmatrix} 3 & 5 & 5 & 6 & 5 \\ 3 & 4 & 6 & 6 & 5 \end{bmatrix}$$

Finally, non-dominated sort will be continued processing with new population until reaching the maximum iteration times.

The process of optimization is illustrated in Fig. S2. In the RWCD process, the number of partitions is equal to the number of iterations. Thus, a huge number of partitions will be obtained. Then, some partition plans that exist nodes of disconnected regions or repeated plans should be removed. The depth-first search algorithm is employed to judge the connectivity of all nodes at each partition (Rodrigues et al. 2019). Then, all selected partitions plans are inputted to the NSGA-II model, using five objective functions to optimize. Finally, the best partition plan can be selected manually with network reliability.

4 Example of the Network Partition With Net1

An example is employed called Net1 (Hwang and Lansey 2017), which has 9 junctions, 1 reservoir, 1 tank, 12 pipes, and 1 pump. The average water demand is 249.81 m^3/h. The network topology is shown in Fig. S3.

Eleven nodes in Net1 are firstly distributed to 11 partitions shown in Fig. S4 (a). The average pressures of nodes are solved by EPS, illustrated in Table S1. Then, each weight A_{ij} is solved by Eq. 12 if two nodes are connected, or is zero, illustrated in Fig. S4 (b). There are 13 weights of non-zeros including A_{12}, $A_{1,10}$, A_{23}, A_{25}, A_{34}, A_{36}, $A_{3,11}$, A_{47}, A_{56}, A_{58}, A_{67}, A_{69}, A_{89} and the matrix is a symmetric matrix ($A_{ij}=A_{ji}$). Thus, 26 elements are non-zeros. The sum of all rows and columns of matrix A is 2 M = 2066.96. M is equal to 1033.48. The vector $\boldsymbol{a}$ is equal to the sum of each row of A matrix.

$$\mathbf{a} = [128.66, 253.38, 316.52, 171.53, 252.27, 338.62, 172.51, 64.69, 163.64, 43.45, 61.92]^{T}$$

The η solves by $a^T/2M$, η=[0.062, 0.123, 0.153, 0.083, 0.122, 0.164, 0.083, 0.080, 0.079, 0.021, 0.030]^T. ψ and φ are diagonal matrix $\boldsymbol{a}$ andη, which are obtained via replacing the elements to the diagonal lines of matrix.

The matrix H is initialized and generates an 11*11 unit matrix. The GN modularity matrix is solved by Eq. 14 and results are illustrated in Table S2. The range of step time is from 0.5 to 100 and increases at 0.5 per step. The trace of this matrix $r_1(1)$ is -0.6240, when the time is 0.5.

Then, node 1 is transferred from partition 1 to partition 2. The elements H_{11} changes from 1 to 0 while H_{12} is changing from 0 to 1 when node 1 was assigned to partition 2.

Then, GN modularity is recalculated at the new partition by Eq. 14. The trace of this matrix $r_1(1)$ is -0.6536. Thus, variations of modularity as node 1 is assigned from partition 1 to partition 2 can be solved by Eq. 15. $\Delta r_1(1 \rightarrow 2) = r_1(2) - r_1(1)$=-0.0296. Similarly, the variation as node 1 from partition 1 to partition 3 to 11 are − 0.027, -0.140, -0.077, -0.010, -0.139, -0.145, -0.146, -0.215, and − 0.225. The maximum of the GN modularity is $\Delta r_1(1 \rightarrow 6)$ = -0.010<0. The partition of node 1 does not transfer to the other partition for the next iteration. Others nodes' variations of GN modularity matrix are shown in Table S3, when the step time is 0.5. No one variation is larger than zero with nodes 10, 11, in which the first iteration will not change the partition of any nodes. Nodes (index of 2–9) have some elements greater than 0. For instance, the maximum of the variation of node 2 is $\Delta r_2(2 \rightarrow 3)$>0. Node 2 is transferred from partition 2 to partition 3. Finally, a new partition solution is obtained, which has five partitions [{1}, {2,4,6}, {3,5,7,8,9}, {10}, {11}], shown in Fig. S5 (a).

The second step is processed when the time is 1.0 (0.5 + 0.5). The final partition solution is obtained at time 29.5, after iterations of 59 times, shown in Fig. S5 (b). Two partitions are obtained: [{1,2,10}, {2,3,4,5,6,7,8,9,11}]. Boundary pipes are {2–3}, {2–5} between two partitions, where we can set valves to control the connectivity.

After the network partition by RWCD, 2000 partition plans would be obtained. Four different conditions including six, five, three, two partition states are retained, which are shown in Fig. S6.

NSGA-II is employed to optimize MRE, HRE, FEF, NR, and the number of connectivity. The highest and lowest pressures are 140 m and 110 m, respectively. The maximum and minimum levels of the tank are 45.72 m and 30.48 m, respectively. The results are shown in Table 1 that the reliability of WDS is improved after the partition process. The maximum opened boundary pipes are 8, 7, 4, and 2, respectively at six, five, three, two partition conditions of Fig. S6 (a)-(d). The best result can be selected in 2 or 3 partition conditions. HRE at 3 numbers of partitions is greater than 2, and MRE is nearly equal to 2 numbers of partitions, and FEF also greater than 2 numbers of partitions, which indicated that the leakage and water hammer broken will occur in a higher probability in 2 partitions condition. Moreover, only one boundary pipe will cause an unsafe status and cut off contact between two partitions if the pipe breaks. Therefore, the best number of open valves was 2 ($L_{2,3}$ and $L_{3,6}$, where 2, 3, and 6 are the index of ends nodes), where MRE, HRE, FEF, NR have the best-optimized values at the 3 partitions condition. Two PRVs should be installed in the

Table 1 The results of the best solution at different partition states

Partitions	MRE	HRE	FEF	NR	Boundary pipes	Open pipes	Binary expressed
no	0.8400	0.0967	1.1059	0.0005	13	13	-
6	0.6393	0.0902	0.8946	0.0007	8	5	11,100,101
5	0.7805	0.1040	0.8828	0.0003	7	5	1,110,011
3	0.9784	0.1877	1.1110	0.0007	4	2	1001
2	0.9721	0.1268	0.9153	0.0003	2	1	10

second and third boundary pipes ($L_{2,5}$ and $L_{6,7}$ where 2, 3 and 6, 7 are the index of ends nodes) and closed.

5 Case studies

Seven cases of WDS were employed to verify the practicability of this method. Cases of frequently-used WDSs are employed, including Net3 (Hwang and Lansey 2017), EXA7 (Bentley Systems 2013), three Kentucky (KY) systems (Jolly et al. 2014) from the Kentucky statewide database, and BWSN developed by (Ostfeld et al. 2008; Amin et al. 2013). In addition, a real case, City F in the south of China is applied to reveal real effectiveness. The parameters of these WDSs are listed in Table S4. The topologies of WDS are shown in Fig. S7–S13. The time interval of hydraulic calculation is set to one hour in the EPS process and the maximum time is set to 24th hour. The OWA toolbox is employed for all hydraulic calculations. The range of partition step time is from 0.5 to 100 and increasing 0.5 per step. The computer devices are: CPU is i7-9700K; RAM is DDR4-96G; ROM is intel-SSD-1T; The operating system is Windows 10, and Matlab-2018b is used.

For the case of Net3, each step of GN modularity is first calculated, then network partition is processed. The process of iterations of random walk is illustrated in Fig. S14. The highest and lowest pressures are 200 m and 100 m. The maximum and minimum levels of tanks (name IDs are 1, 2, and 3) are (0 m, 9.78 m), (1.98 m, 12.28 m), and (1.22 m, 10.82 m), respectively. The partition numbers will stop variation when the step is 356. After 2000 steps, Net3 is divided into 14, 10, 9, 8, 7, 6, 5, 4, and 3 partitions at the step of 1, 2, 3, 5, 7, 8, 15, 24, 49, and 356. Then, NSGA-II is employed to optimize pipes and select the best partitions. The populations and maximum iteration time are 20 and 50. The solutions of optimization are manually filtered. Six high potential plans are selected, illustrated in Table S3. The partition is shown in Fig. S15. The four partition plan in Table S5 has the largest MRE, HRE, and NR while FEF is small enough (MRE, HRE, FEF, NR are 1.0000, 0.0908, 1.5183, 0.0051), which indicated that the leakage and water hammer will occur in a lower probability. Therefore, the best solution is four partitions and the two of six boundary pipes (name IDs are 120 and 240) should be opened and PRVs should be set to others, shown in Fig. 1.

The other topological graph can be solved in the same way. For the cases of EXA7, KY3, KY5, KY11, city F, and BSWN Numbers of 15, 5, 5, 12, and 9 partitions plans are obtained by random walk, shown in Fig. S16-S21. The highest and lowest pressures are 80 m and 10 m with EXA7, respectively. The maximum and minimum levels of the tank (name ID 882) are 12.19 m and 9.78 m, respectively. The solutions of NSGA-II optimization are manually filtered, illustrated in Table S6. The best solution of EXA7 has eight partitions and nine boundary pipes should be opened, shown in Fig. S22. The name IDs of these boundary pipes are 421, 504, 614, 617, 708, 733, 768, 769, 821. The other boundary pipes should be closed. The highest and lowest pressure of KY3, KY5, and KY11 are the same as 100 m and 40 m. Three tanks including T-1, T-2, T-3, maximum and minimum levels are (42.88 m, 48.98 m), (44.88 m, 52.50 m), (28.69 m, 39.36 m) in KY3. After NSGA-II optimization, the solutions of NSGA-II optimization are manually filtered and high potential plans are selected, illustrated in Table S7. The best result of KY3 has three partitions with 4 opened boundary pipes, shown in Fig. S23. The name IDs of pipes are P-151, P-341, P-342, and P-346. For the case of KY5, tanks including T-1, T-2, T-3, maximum and minimum levels are (17.56 m, 25.18 m), (7.18 m, 17.85 m), and (8.19 m, 18.85 m), respectively. The results of NSGA-II optimization are
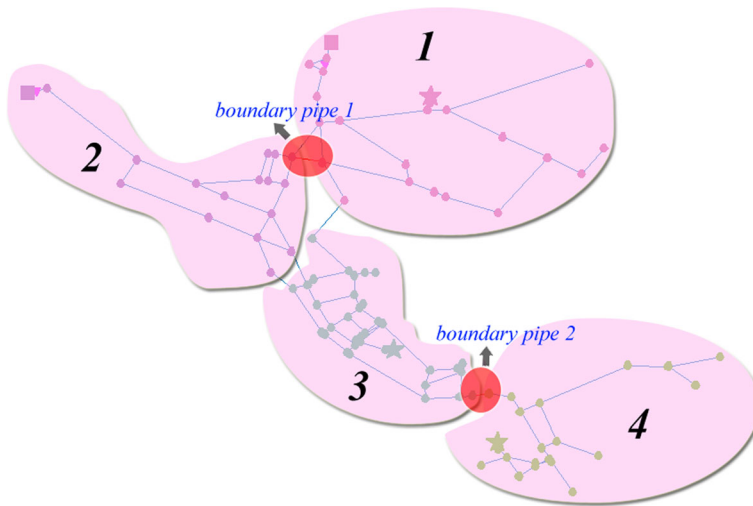
Fig. 1 The best network partition plans of Net3

illustrated in Table S8. The best result of KY5 has six partitions with 7 opened boundary pipes, shown in Fig. S24. The name IDs of pipes are P-13, P-151, P-226, P-400, P-447, P-458, P-63. For case of KY11, 25 tanks maximum and minimum levels are (0.00, 4.67), (0.00, 3.96), (30.07, 39.21), (12.22, 20.15), (1.87, 11.92), (3.80, 10.50), (13.83, 20.54), (28.02,35.64), (18.22, 21.27), (22.38, 30.00), (0.00, 9.34), (33.89, 38.46), (30.74, 39.89), (25.99, 33.61), (24.47, 30.56), (0.13, 8.66), (41.03, 51.69), (0.00, 7.54), (18.65, 26.57), (34.33, 43.17), (21.19, 25.76), (0.00, 8.00), (25.01, 31.10), (32.94, 39.65), (15.70, 21.80), (12.18, 19.19), (14.47, 27.28), (6.35, 10.92) meters. The optimized results are illustrated in Table S9. The final result of KY11 has nine partitions with 7 opened boundary pipes that name IDs of pipes are P-101, P-111, P-127, P-128, P-13, P-152, P-205, shown in Fig. S25. For the case of city F, no tank exists in the topological graph. Thus the restriction of tank level can be ignored. The optimized results are illustrated in Table S10. The final result of city F has ten partitions with 22 opened boundary pipes that name ID are 89,825, 107,851, 145,471, 154,060, 156,025, 156,761, 156,801, 157,093, 157,778, 157,956, 158,096, 158,097, 158,618, 158,730, 158,997, 159,144, 159,154, 159,163, 159,187, 159,192, 159,194, 159,236, shown in Fig. S26. The results of BSWN for optimization are illustrated in Table S11. Two tanks' maximum and minimum levels are (0 m, 9.78 m) and (0 m, 12.77 m). The final result of BSWN has three partitions with 2 opened boundary pipes that name IDs are LINK-41 and LINK-85, shown in Fig. S27.

MRE of cases of Net3, EXA7, KY3, KY5, KY11, city F, and BSWN all increase by 0.051, 0.021, 0.047, 0.045, 0.131, 0.004, and 0.030, respectively. HRE all increase by 0.090, 0.286, 0.3696, 0.373, 0.131, 0.235, and 0.003. NR are all increased by 0.004, 0.001, 0.014, 0.019, 0.074, 0.001, and 0.008. Moreover, FEF reduce by 0.419, 2.246, 2.155, 0.897, 2.121, 39.428, and 0.896. The results demonstrate that the FEF values at each partition for each specific case is decreased and MRE, HRE, and NR are improved, thus this method has a great effect on network partition. Meanwhile, each partition can be effective controlled via PRVs.

Above all, comparing with the conditions of no partition, these seven cases all show the improvement of network reliability. Energy and economic losses are decreased due to the reduced leakage and water hammer phenomenon.

6 Summary and Conclusion

An automatic network partition is presented. RWCD model is employed to divide WDS into various partitions according to the node pressures at first. Five functions including MRE, HRE, FEF, NR, and the number of boundary pipes are used as objective functions to access the reliability of a network. Then, the NSGA-II algorithm is applied to optimize WDSs conditions, with respect to boundary pipes and partitions. A high potential strategy of partitioning can be obtained via setting PRVs to the boundary pipes. An example of the Net1 topological graph is employed to briefly explain this process. This method is also verified by seven cases of WDS. The OWA toolbox was used as a hydraulic calculation tool for these EPS processing.

Base on the results obtained the following observations can be made:

(1) Network reliability of all of WDS can be improved after network partition comparing with no partition condition, which reduces leakages and water hammers.

(2) The best partition plan can be selected by NSGA-II without manual filters.

(3) Average pressure is an effective variable for partitioning.

(4) A PRV setting to close pipe can get a clear boundary of partitions for convenient management.

Therefore, this method is a practicability method for water resources management.

However, some weaknesses might be found in this method:

(1) The more iterations and the fewer steps, the more partition plans will be generated, which may spend much time searching eligible plans in the RWCD model. Meanwhile, the more partition plans, the more optimization processes should be measured. Thus, a time-saving modification should be developed.

(2) The maximum of the MRE accessing function is 1.0 (100 %). Some cases have optimized it to up to maximum value and cannot improve further. Thus, a better evaluation function should be invented.

(3) In addition, the hydraulic condition of WDS is only considered as independent variables in the random walk partitioning model. Water quality is also an important indicator for WDS that might also be taken into account.

Thus, researching and improving this method should continue in the future.

Data Availability Supplementary data to this article can be found online.

Declarations

Ethical Approval Compliance with Ethical Standards.

Consent to Publish All authors agree to publish.

Consent to Participate Not applicable.

Conflict of Interest The authors declare no conflicts of interest.

References

Amin S, Litrico X, Sastry SS, Bayen AM (2013) Cyber security of water SCADA Systems—Part II: attack detection using enhanced hydrodynamic models. IEEE Trans Control Syst Technol 21:1679–1693. https://doi.org/10.1109/TCST.2012.2211874

Bentley Systems (2013) WaterGEMS V8i users'manual. Watertown, CT

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008:P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Ciaponi C, Murari E, Todeschini S (2016) Modularity-Based Procedure for Partitioning Water Distribution Systems into Independent Districts. Water Resour Manag 30:2021–2036. https://doi.org/10.1007/s11269-016-1266-1

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6:182–197. https://doi.org/10.1109/4235.996017

Dehghanian N, Nadoushani SSM, Saghafian B, Akhtari R (2019) Performance evaluation of a fuzzy hybrid clustering technique to identify flood source areas. Water Resour Manag 33:4621–4636. https://doi.org/10.1007/s11269-019-02385-7

Delvenne J-C, Yaliraki SN, Barahona M (2009) Stability of graph communities across time scales. http://arxiv.org/abs/0812.1811

Delvenne J-C, Schaub MT, Yaliraki SN, Barahona M (2013) The stability of a graph partition: a dynamics-based framework for community detection. pp 221–242. https://doi.org/10.1007/978-1-4614-6729-8_11

Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Ghodrati A, Lotfi S (2012) A Hybrid CS/GA algorithm for global optimization. pp 397–404. https://doi.org/10.1007/978-81-322-0487-9_38

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99:7821–7826. https://doi.org/10.1073/pnas.122653799

Giudicianni C, Herrera M, di Nardo A, Adeyeye K (2020) Automatic multiscale approach for water networks partitioning into dynamic district metered areas. Water Resour Manag 34:835–848. https://doi.org/10.1007/s11269-019-02471-w

Gupta R, Bhave PR (1994) Reliability analysis of water-distribution systems. J Environ Eng 120:447–461. https://doi.org/10.1061/(ASCE)0733-9372(1994)120:2(447)

Gupta A, Kulat KD (2018) A selective literature review on leak management techniques for water distribution system. Water Resour Manag 32:3247–3269. https://doi.org/10.1007/s11269-018-1985-6

Hwang H, Lansey K (2017) Water distribution system classification using system characteristics and graph-theory metrics. J Water Resour Plan Manag 143:04017071. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000850

Jolly MD, Lothes AD, Sebastian Bryson L, Ormsbee L (2014) Research database of water distribution system models. J Water Resour Plan Manag 140:410–416. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000352

Kumar A, Kumar P, Singh VK (2019) Evaluating different machine learning models for runoff and suspended sediment simulation. Water Resour Manag 33:1217–1231. https://doi.org/10.1007/s11269-018-2178-z

Lambiotte R, Delvenne JC, Barahona M (2014) Random walks, Markov processes and the multiscale modular organization of complex networks. IEEE Trans Netw Sci Eng 1:76–90. https://doi.org/10.1109/TNSE.2015.2391998

Martelot EL, Hankin C (2012) Multi-scale community detection using stability optimisation within greedy algorithms. http://arxiv.org/abs/1201.3307

McDonnell BE, Salomons E, Uber J, Klise K (2007) Open Water Analytics (OWA). KIOS Research Center for Intelligent Systems and Networks of the University of Cyprus. http://wateranalytics.org/

Mounce SR, Ellis K, Edwards JM et al (2017) Ensemble decision tree models using RUSBoost for Estimating risk of iron failure in drinking water distribution systems. Water Resour Manag 31:1575–1589. https://doi.org/10.1007/s11269-017-1595-8

Ostfeld A, Uber JG, Salomons E et al (2008) The Battle of the Water Sensor Networks (BWSN): A design challenge for engineers and algorithms. J Water Resour Plan Manag 134:556–568. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:6(556)

Pasha MFK, Lansey K (2014) Strategies to develop warm solutions for real-time pump scheduling for water distribution systems. Water Resour Manag 28:3975–3987. https://doi.org/10.1007/s11269-014-0721-0

Prasad TD, Park N-S (2004) Multiobjective genetic algorithms for design of water distribution networks. J Water Resour Plan Manag 130:73–82. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:1(73)

Rodrigues GPW, Costa LHM, Farias GM, de Castro MAH (2019) A depth-first search algorithm for optimizing the gravity pipe networks layout. Water Resour Manag 33:4583–4598. https://doi.org/10.1007/s11269-019-02373-x

Rossman L (2000) EPANET's user manual. US EPA, Washington, DC. https://nepis.epa.gov/Adobe/PDF/P1007WWU.pdf

Savic DA, Walters GA, Schwab M (1997) Multiobjective genetic algorithms for pump scheduling in water supply. pp 227–235. https://doi.org/10.1007/BFb0027177

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sirsant S, Reddy MJ (2020) Assessing the performance of surrogate measures for water distribution network reliability. J Water Resour Plan Manag 146:1–15. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001244

Srinivas N, Deb K (1994) Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evol Comput 2:221–248. https://doi.org/10.1162/evco.1994.2.3.221

Su Y, Mays LW, Duan N, Lansey KE (1987) Reliability-based optimization model for water distribution systems. J Hydraul Eng 113:1539–1556. https://doi.org/10.1061/(ASCE)0733-9429(1987)113:12(1539)

Tanyimboh TT, Templeman AB (1993) Maximum entropy flows for single-source networks. Eng Optim 22:49–63. https://doi.org/10.1080/03052159308941325

Todini E (2000) Looped water distribution networks design using a resilience index based heuristic approach. Urban Water 2:115–122. https://doi.org/10.1016/S1462-0758(00)00049-2

Ulanicki B, Shipley N, Prescott S (2003) Analysis of district metered area (DMA) performance. In: Advances in Water Supply Management. Taylor & Francis, Abingdon

Wu ZY, Rahman A (2017) Optimized deep learning framework for water distribution data-driven modeling. Procedia Eng 186:261–268. https://doi.org/10.1016/j.proeng.2017.03.240

Yang X-S (2014) Nature-inspired optimization algorithms. Elsevier. https://doi.org/10.1016/C2013-0-01368-0

Zhang M, Shen Y (2007) Study and application of steady flow and unsteady flow mathematical model for channel networks. J Hydrodyn 19:572–578. https://doi.org/10.1016/S1001-6058(07)60155-3

Zhang Q, Wu ZY, Zhao M et al (2016) Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. J Water Resour Plan Manag 142:04016042. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000661

Zhuan X, Xia X (2013) Optimal operation scheduling of a pumping station with multiple pumps. Appl Energy 104:250–257. https://doi.org/10.1016/j.apenergy.2012.10.028