Preconditioning Water Distribution Network Optimization with Headloss-Based Design Method

Haixing Liu[1,2] Associate Member ASCE, Christine A. Shoemaker[3] Dist. Member.ASCE Yunzhong Jiang[4], Guangtao Fu[5], Chi Zhang[1]

[1]School of Hydraulic Engineering, Dalian University of Technology, Liaoning, 116023, China

[2]Civil and Environmental Engineering Department, National University of Singapore, Singapore

[3]Industrial Systems Engineering & Management Department and Civil & Environmental Engineering Department, National University of Singapore, Singapore

[4]Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

[5]Center for Water Systems, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, North Park Rd., Exeter, EX4 4QF, UK

Corresponding author: Christine A. Shoemaker

Abstract

This paper develops a new domain knowledge-based initial design method for optimization of water distribution network design. The new initial water distribution network design method, termed as Headloss-based Design Preconditioner (HDP), is based on headloss analysis in the supplying path from source to user. The new HDP-preconditioned search is compared with two algorithms: one preconditioned on a velocity-based initial design method and a simple genetic algorithm without preconditioning. The results show the HDP headloss-based method outperforms the Prescreened Heuristic Sampling Method (PHSM) in terms of the quality of the initial solutions and computational efficiency on all three cases. HDP also outperforms stochastic initialization on two of the three cases. The results obtained imply that the proposed domain knowledge-based design method HDP would be able to also provide effective starting conditions for other optimization algorithms besides genetic algorithm for large water distribution systems since most optimization methods are greatly assisted by a good starting condition.

Keywords:

Water distribution network; preconditioning; optimization; design.

Introduction

The drinking water distribution network (WDN) is critical urban infrastructure that provides the essential service of safe and high-quality drinking water to consumers. Optimal design of the WDN is an important problem because it is a basis for decisions normally involving large investments and hence optimal design can potentially suggest substantial savings. Preconditioning WDN optimization is to provide “good information” for the initiation of the optimization process to improve accuracy and computational efficiency.

Literature Review

In the last several decades, various optimization approaches have been developed and applied to water distribution network optimal design. In particular population-based evolutionary algorithms (EAs) (Fu et al. 2012a, Tolson et al. 2009, Vairavamoorthy and Ali 2000, Wu and Simpson 2001, Zheng et al. 2011, Zheng et al. 2017) have been used. When applying these optimization approaches to real-world large-scale networks however, challenges exist because of the high dimensional decision space (that can include tens of thousands of pipes), potentially high computational demands, and constraints related to engineering practicability (Walski 2015, Wang et al. 2018, Zhang et al. 2018).

Expert knowledge plays a crucial role not only in informed decision making, such as computer-aided decision support systems, but also at various stages of an optimization process, e.g., pre-optimization (Bi et al. 2015, Kang and Lansey 2012), mid-optimization (Johns et al. 2014, Keedwell and Khu 2006, Montalvo et al. 2014) and post-optimization (Andrade et al. 2012). Prior to optimization, expert knowledge can be used to generate an initial population of solutions, and this can improve the efficiency of optimization algorithms (Bi et al. 2015, Kang and Lansey 2012). During the optimization process, domain knowledge can guide the search through establishing the governing rules of search (Keedwell and Khu 2006), strengthen operator behavior of genetic algorithms (Johns et al. 2014), and promote human-computer interactions in the development of computer-aided design by adding good, diverse solutions generated from expert knowledge into the desired search region (Montalvo et al. 2014). After optimization, a domain knowledge-based greedy search algorithm can refine the solutions obtained (Andrade et al. 2012). In general, previous research has proven that domain knowledge can increase the efficiency of optimization algorithms and effectively guide the search direction, thus improving the convergence of optimization algorithms. Thus, there is a critical need to develop efficient, automatic, heuristic approach using domain knowledge to improve the performance of optimization algorithms for complex optimization problems such as WDN design.

Preconditioning, a technique that feeds high performing solutions to the initial population for optimization has been investigated to guide the search towards the global optimum. Preconditioning methods can provide Initial Network Configurations (INC) for the population-based optimization algorithm. Different approaches have been developed to provide promising solutions in the field of WDN optimal design. For example, Fu et al. (2012b) employed a global sensitivity analysis method to decompose the original, complex WDN optimization problem into simple problems with a small set of sensitive design variables, whose solutions are used to precondition the search of the original problem. This method was tested on WDN. Zheng et al. (2011) used a nonlinear programming (NLP) optimization approach to derive solutions for decomposed branched networks, then the solutions are fed to EAs and concurrently the search space is tailored accordingly. Kang and Lansey (2012) and Bi et al. (2015) prescreened heuristic sampling for deriving good solutions based on a network flow velocity analysis. These knowledge-based heuristic methods are proven promising in improving the search efficiency, and thus motivate the current study to develop a more efficient preconditioning method.

Headloss-based Design Preconditioner (HDP) for Finding Initial Network Configurations (INC)

This paper aims to develop a new method (HDP) for WDN that provides high performing solutions to precondition for optimization. The heuristic method is based on physical domain knowledge on headloss. The reasoning behind the method is that, in the optimal network configuration, the minimum pressure head should be as close as possible to the pressure threshold required, and all energy (i.e. pressure head) should be adequately utilized along the supply path without any constraint violations (i.e., minimum pressure requirement). This approach can effectively identify a high performing network configuration for multi-source WDN design problems.

The new method HDP is compared with a flow velocity-based heuristic design approach by

Bi et al. (2015), i.e., Prescreened Heuristic Sampling Method (PHSM), with respect to the computational effort and the quality of solutions.

For both the PHSM method and the HDP preconditioning method proposed here, the best solution can be input to an evolutionary algorithm by making it a member of the initial population in the algorithm (e.g. in a genetic algorithm). We consider three water distribution networks of increasing complexity (Two-reservoir, Modena, and Balerma networks) to demonstrate the efficiency and effectiveness of the proposed HDP method. In this paper we use HDP in conjunction with a genetic algorithm, but it could also be used with other optimization algorithms such as DDS (Tolson and Shoemaker 2007), Differential Evolution (Storn and Price 1997), and Particle Swarm Optimization (Kennedy and Eberhart 1995).

Methods

Headloss-based Design Preconditioner (HDP)

For the optimal design problem of water distribution networks, the ideal optimal solution should, have the least cost solution that satisfies the required minimum pressure at each demand node and all other constraints. The symbols used in this section and the acronyms are listed in Table 1.

The schematic diagram of the HDP method is presented in Figure 1. In this method, three major steps are included: 1) each demand node needs to have a unique supply path, and the supply area from each source is determined first if the network includes multiple sources; 2) Calculating the headloss for each individual pipe based on the supply path; and 3) Calculating the pipe diameter with the headloss equation. The updated pipe diameters will change the flow and headloss of all pipes, and thus the supply zones of water sources and supply paths will change subsequently. An iterative process is comprised by the above three steps, which is stopped when the termination criterion is achieved. The detailed procedures of the HDP method are presented below and summarized in Fig. 1.

Step 0: Initializing parameters

Prior to the first iteration, all pipes in the network are set to the largest diameter among the pipe diameters available, which is gradually reduced through iterations described in Step 3. The parameters IT_{mas} and P_{req} are provided before computing, where the maximum number of iterations IT_{mas} is used as a stop criterion and P_{req} is the minimum pressure threshold that must be maintained at all nodes in the network

Step 1: Grouping nodes in terms of water source supply zone

In multi-source network systems, a demand node may be supplied by several possible water sources, which are primarily determined by the source heads. The traditional way of determining the supplier employs the shortest path method (Bi et al. 2015, Zheng et al. 2011), which only considers the spatial distance between source and node, without involving hydraulic conditions. Since water sources usually have different heads feeding water into the network, the source with the higher pressure head is capable of providing a larger area than that of the lower head. A node i belongs to a specific source k when the shortest paths between the node i and k has the smallest headloss in comparison with the path from node i to any other source. The number of subzones is equal to the number of water sources. Therefore, a new method is proposed here to determine the supply area (i.e. which source should supply water to a given node). The procedure of Step 1 is given as follows:

Step 1.1 Conduct a hydraulic simulation using the current diameter settings (i.e. the largest diameter for each pipe in the Step 0). The simulation results for source heads ($H_k, k = 1, 2, \cdots, N_s$) and pipe flow rates (i.e. including the local demand and conveyance quantity) ($Q_{ij}, i, j = 1, 2, \cdots, N_n, i \neq j$) are recorded where N_s and N_n are the numbers of the sources and the nodes, respectively;

Step 1.2 Calculate the length (L_{ki}) of the shortest paths (P_{ki}) from source k to node i, $P_{ki} = \left(\theta_k, \theta_j \cdots, \theta_i\right)$, using Dijkstra algorithm (Deo 1974) ($i = 1, 2, \cdots, N_n$), and the total number of paths is $N_s \times N_n$;

Step 1.3 Calculate the potentially maximum headloss for each path $HL_{ki} = H_k - P_{req} - E_i$, where E_i represents the elevation at the i th node;

Step 1.4 Calculate unit headloss for each path $UHL_{ki} = HL_{ki} / L_{ki}$, and the total number of unit headlosses is $N_s \times N_n$; and

Step 1.5 Node i has multiple shortest paths corresponding to different sources. Node i is assigned to a specific source if their shortest path poses the smallest unit head loss (UHL_{ki}).

Step 2: Calculate each pipe headloss based on the least headloss path

This heuristic design approach is based on the headloss analysis. Usually, a smaller pipe diameter leads to the greater headloss under the same flow rate condition. This method examines the available largest headloss that could be dissipated in the pipeline. The headloss of a pipe (HL_{ij}) is calculated by the unit headloss (UHL_{ki}) (obtained by Step 1) in the path and the pipe length (L_{ij}). When a pipe is a part of multiple different paths, there are several headloss values for the same pipe. In this situation, the smallest headloss is chosen to represent the headloss of the given pipe, since this guarantees there will be sufficient head preserved for the downstream pipes.

Step 3: Calculating pipe diameters using the headloss of each pipe segment

Step 3.1 The diameter is calculated using the headloss equation (e.g. Hazen-Williams (HW) equation and Darcy-Weisbach (DW) equation (Ormsbee and Walski 2016)) with the known headloss (from Step 2) and flow rates (from Step 1). The specific headloss equation is chosen in terms of the parameters in the cases. The mathematical expression of the HW equation in SI units is given as,

$$D_n = \frac{1.626 L_{ij}^{0.205} Q_{ij}^{0.38}}{C^{0.38} HL_{ij}^{0.205}} \tag{1}$$

where D_n is the diameter updated; C is the Hazen-William coefficient. In addition, the DW equation is also an alternative headloss equation as below,

$$D_n = \left(f \frac{8 Q_{ij}^2 L_{ij}}{g \pi^2 HL_{ij}} \right)^{\frac{1}{5}} \tag{2}$$

where g is the gravity acceleration and f is the Darcy friction factor that can be calculated by the Hagen-Poiseuille formula, the Moody Diagram or the Colebrook-White equation. Because the Darcy-Weisbach equation takes different flow regimes into account, it is thought to be more accurate than the empirical Hazen-Williams equation. The headloss function is chosen in terms of the specific roughness parameter in this study.

Step 3.2 The resulting diameter is rounded up to the closest discrete commercial diameters available. The fitness and constraints are calculated then.

Step 4 The HDP method is an iterative process until the maximum iteration number ($It_{\max}$) is reached or the diameters have no change in a sequence of iterations. The new diameters obtained from Step 3 are used to update the network design and the algorithm goes back to Step 1. Steps 1-3 are repeated until the stopping criteria are met.

As shown in Figure 1, the core idea in the HDP algorithm is to find the least headloss for each pipe in the supply path without violating constraints. We establish the supply zone for grouping nodes. Then, the least headloss per unit length is identified from the headloss paths. If one pipe appears in multiple paths, then the smallest headloss (i.e. the largest diameter) is taken. Third, the headloss of pipes can be translated into pipe diameter by the classical headloss equation (Walski et al. 2003). The HDP method includes an iterative process with three steps. The diameter and flow rate are varied in each iteration and gradually reach an equilibrium (i.e., diameters unchanged through multiple iterations). The headloss along the path is determined based on the optimality principle of the WDN design (i.e. the farthest node heads are at the minimum pressure requirement), which is essential to guarantee a good heuristic solution. The power of HDP is demonstrated by the network cases in the Results section.

The HDP method is demonstrated using an illustrative case study, i.e., the multi-source network with different feeding heads (i.e. water surface elevation of the reservoir). The illustrative network consists of 13 pipes and 8 demand nodes and is symmetrical with pipes of equal length, as shown in Figure 2a. There are two elevated reservoirs feeding the users from east and west sides, respectively. The minimum pressure required is 30 m.

The headloss-based heuristic design (HDP) method is tested in two scenarios of the illustrative network. Figure 2b shows the symmetric diameter configuration for Scenario 1 with the equal heads of the two reservoirs, while Figure 2c shows the Scenario 2 where the source with the higher head could provide a larger supply area. The results are in line with the fact that higher head will reach further demand nodes. This implies the new method is better than the traditional shortest path method (Dijkstra algorithm), which would evenly distribute supply areas in terms of the pipe length. The diameters obtained by HDP in Figure 2 gradually decrease from the source to the lowest pressure node, which is what would be expected based on fluid dynamics. There is no pressure violation in the two configurations of diameters. Therefore, the HDP performs well in the pipe sizing problem and can be applied to a multi-source network.

Comparison to Prescreened heuristic sampling method (PHSM)

PHSM was developed to provide a promising prescreening solution using domain knowledge by Bi et al. (2015). The goal of a prescreening solution is to speed up the identification of an optimal distribution of pipe diameters by providing a good initial guess of what the best solution is, i.e., a set of good values for all the pipe diameters. In Bi et al (2015). PHSM performs well in comparison with Kang and Lansey's heuristic method (KLHM) (Kang and Lansey 2012). Therefore, in this paper PHSM is chosen for comparison with the newly proposed method (HDP). The PHSM method has three primary features : 1) a procedure for estimating a good initial diameter distribution across the network; 2) a velocity-based iterative heuristic design method; and 3) an initializing strategy for an evolutionary optimization algorithm. Following are the PHSM steps:

Step 1 The shortest path method is used to divide the network into sub-zones along the path from source to the farthest nodes. The number of sub-zones is based on the number of the diameter options. In each sub-zone the same initial diameter is assigned to the pipes. The pipes near the source are allocated with a larger diameter, while farther pipes given a smaller diameter.

Step 2 The procedure consists of two loops. The outer loop iterates for the velocity increase until the constraints are violated. The inner loop evolves to achieve an equilibrium between diameter and flows under a certain velocity condition until the diameters convergence (i.e., diameters have no improvement in a sequence of iterations). The velocity is an *a priori* condition in the outer loop, and the flow rate is derived from the hydraulic simulation. The diameter is calculated using $D_n = \sqrt{4Q/\pi v}$, where D_n is the updated diameter (m); Q is the pipe flow rate (m^3/s); and v is the velocity (m/s). The velocity is consistent in an inner loop.

Step 3 The procedure is designed to promote e the diversity of initial solutions, as well as to strengthen the good genes in the population. A probability density function (PDF) is introduced based on the best known solution derived from Step 2. The initial population of an evolutionary algorithm is randomly sampled based on the established PDF. The new PDF replaces the normal distribution, which is widely used in the experimental design (e.g. Latin hypercube sampling). The new PDF is steeper in comparison with the bell shape of the normal distribution in order to significantly concentrate near the best solution. The detailed description of the PHSM method can be found in the study by Bi et al. (2015).

Problem Formulation and Optimization Algorithms

The water distribution network (WDN) optimal design is formulated as a single objective least cost problem. The objective is penalized when design constraints are violated. The mathematical expression of the objective (F) is given below and is used as a fitness function in the genetic algorithm:

$$F = \begin{cases} \sum_{i=1}^{pn} L_i f\left(D_i\right),\ if\ \ V\left(D\right) \leq 0 \\ C_{\max} + penalty,\ if\ \ V\left(D\right) > 0 \end{cases} \tag{3}$$

where L_i and D_i are the length and the diameter of pipe i, respectively. $f(D_i)$ is the unit cost length function of pipe i with diameter D_i. pn is the total number of pipes. Each pipe has a maximum cost (typically for the largest diameter). $C_{\max}$ is the sum of these maximum costs over all the pipes. If the design constraints are violated, the objective is greater than the maximum cost. The higher the violation, the larger the penalty. $V(D)$ is the violation. The term $penalty$ is the sum of the constraint violations (i.e. the possible maximum cost). The constraint violation is calculated as the sum of pressure deficits,

$$V(D) = -\sum_{i=1}^{nn}\left[\min\left(H_i - H_{req}, 0\right)\right]. \tag{4}$$

where H_i is the pressure head at node i, H_{req} is the pressure head requirement. nn is the total number of nodes. The hydraulic equations that compute pressures and flows in each pipe are solved by EPANET 2 (Rossman 2000).

We are comparing alternative methods for generating initial trial solutions (preconditioner) to help speed up the optimization search. For optimization we use a Genetic algorithm (GA) as a typical evolutionary algorithm. Genetic algorithms can find (near) optimal solutions in various complex nonlinear problems with integer variables (Fu et al. 2008, Meng et al. 2016, Reca and Martínez 2006, Sweetapple et al. 2014, Wu and Walski 2005). Nicklow et al. (2010) summarized four components of GA including: 1) generation of the initial population; 2) computation of the fitness function; 3) select parents and reproduce offspring solutions; 4) mutation of each offspring solution to maintain the population diversity. This paper focuses on the first point so the initial population contains a good solution that can enhance the gene pool and potentially lead to a better solution.

Since we are comparing our HDP to an alternative initial starting solution computation method PHSM, we use both methods to start a simple GA (SGA) optimization. We choose GA as the optimization method since this was the method used in the PHSM (Bi et al. 2015). Integer coding is used for the discrete pipe sizing problems. The operators chosen are roulette wheel selection, uniform probability distribution for crossover point, and Gaussian mutation. The case study and the specific parameters for each case are provided in the next section.

To integrate the HDP or PHSM with the GA, the initial population of GA is produced by Latin Hypercube sampling (LHS) method, and the best solution derived from either the PHSM or the HDP is added to the initial population before starting the optimization iterations.

Case Study Networks

Three distribution networks (called "Two reservoir", "Modena", and "Balerma" Networks) with an increasing number of variables and network complexity are used to test the newly proposed HDP approach. Three network cases and their optimization parameters are summarized in Table S1 in the supplemental materials. Each network is explained in more detail below.

Case #1

The "Two-Reservoir" network (TRN) from Gessler (1985) has 14 pipes and 10 junctions, as shown in Figure S1 in the supplemental materials. The design problem has been modified in this study, and all pipes are considered as variables. Two reservoir heads are fixed at 365.76 m (left) and 371.86 m (right). All pipes have the same Hazen-Williams roughness coefficient of 120. The available diameters are [152, 203, 254 ,305 ,356 ,407 ,458 ,509 mm]. The minimum pressure required at all nodes is 30 m. Only the normal demand scenario is used here.

Case #2

Modena network (MOD) (Bragalli et al. 2008), as shown in Figure S2 in the supplemental materials, is a real-world network in Italy, which includes 317 pipes, 268 nodes, and 4 reservoirs with a fixed head in the range from 72.0 m to 74.5 m. A Hazen-Williams roughness coefficient of 130 is applied to all pipes. The available pipe diameters are [100, 125, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800 mm]. The minimum pressure head requirement of all the demand nodes is 20 m. The maximum pressure thresholds are also considered as given in the studies (Bragalli et al. 2008, Wang et al. 2015). The flow velocity in pipes should be less than 2.0 m/s.

Case #3

Balerma Network (BN) in Italy (Reca and Martínez 2006) includes 454 relatively short length pipes, 443 nodes, and 4 reservoirs with fixed heads within 112 m to 127 m, as shown in Figure S3 in the supplemental materials. The material of pipes is polyvinyl chloride (PVC). The available diameters are [113, 126.6, 144.6, 162.8, 180.8, 226.2, 285, 361.8, 452.2, 581.8 mm]. The Darcy-Weisbach fraction factor of 0.0025 mm is applied to all the pipes. The minimum pressure head above ground elevation is 20 m for all the demand nodes.

For all three networks, the options of commercially available diameters, the corresponding unit pipe costs, and EPANET input files (.inp) can be found at the website of Exeter CWS (https://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/design-resiliance-pareto-fronts/summary-of-benchmark-problems/) and also refer to Wang et al. (2015).

Results and Discussion

Comparison of Heuristic Design Methods

Tables 2-4 show the optimal design solutions for TRN, MOD, and BN, respectively, using the two initial approaches HDP and PHSM. With the HDP, each row in tables represents the results of the network cost for each iteration. The PHSM has two iterative processes: 1) the outer loop of velocity increase and 2) the inner loop, which is associated with the iterative process for balancing flow and diameter in every velocity increment.

The design solutions from HDP are better than those from PHSM for all three case studies, with a HDP cost saving of 20.7%, 16.6%, and 29.9% in Tables 2, 3 and 4, respectively. The simulation is executed once in each optimization iteration. So, the number of iterations (each doing one hydraulic simulation) represents the computational burden of the preconditioning methods since the computational time in each hydraulic simulation is consistent for a network. The PHSM needs a larger number of iterations to achieve convergence than the HDP method. The number of hydraulic simulations in the HDP method is only 16.7%, 12.7%, and 6.1%, respectively, of the simulations required by PHSM in the three cases. In summary, the results show the HDP method greatly outperforms PSHM in terms of computational efficiency and solution quality.

Recall that the PHSM method is based on velocity analysis of design networks, with a fundamental assumption of uniform velocity across the network. The method starts with the possible minimum velocity and increases the velocity until the constraints are violated. This assumption results in a deviation from the optimal solutions. The assumption that uniform velocity is distributed in the entire network represents a general understanding of network characteristics, which shows the lower capital cost network may have a higher average velocity. Thus, it is difficult to achieve a lower cost with uniform velocity as required by the PHSM method.

In contrast, the new method (HDP) uses the headloss based analysis in which the possible maximum headloss (in a path) is the criterion to determine the designed diameters. All available heads are effectively distributed along the pipe in order to obtain the cost-effective diameters. On the basis of the given headloss for each pipe, the equilibrium relationship between flow and diameter is achieved within a sequence of iterations. Through the three case studies, the resulting diameters are able to converge quickly on the basis of the criterion of the improvement of the solution quality.

PHSM attempts a wide range of velocities and starts from a relatively low one for ensuring the solution is feasible. For each velocity trial, PHSM requires an inner loop to balance flow and diameter. The computational effort for PHSM is therefore significantly increased compared to HDP. The MOD and BN cases are large-scale, real-world WDNs with hundreds of variables. However, the HDP can achieve the convergence in only a few iterations, which demonstrates that the HDP design method based on domain knowledge is very efficient.

According to Tables 2 and 3, the costs derived from HDP exhibit a decreasing trend in the iterative process. However two infeasible solutions are obtained in iterations 2 and 3 in Table 3, where the values in the brackets are the constraint violation values. The solutions after the third iteration are feasible ones including the final solution. Infeasible solutions are likely to appear at the beginning stage when diameters vary widely. That is because the flows would oscillate around the loop in the network when the diameters are being adjusted. In some cases, the flow direction also varies at times.

Figure 3 shows that all the designed network satisfies the fundamental principle of network design, i.e., the allocated pipe diameters decrease from source to end users along the supply paths. This is an advantage over randomly selecting an initial design because it is using known information about the WDS. . In Figure 3, the pipes that link to sources with a higher head have larger diameters in all three cases. This implies that the source with a higher head can supply more water and cover a larger supply area. Further, the HDP method identifies smaller trunk mains compared with PHSM without violating the head constraints. . These design solutions represent good solutions that can be used in practice if no optimization is going to be undertaken to determine pipe sizes. . However, the solutions could be improved further, with the aid of optimization technologies. This is demonstrated by adding the HDP solutions into the initial population of for the genetic algorithm search.

Optimal Solutions

We want to compare the impact of the different initial network configurations (INC) on the final solutions obtained from the optimization (here using GA as the optimization tool). We compare the INC solutions from HDP and PHSM to the simple GA method, in which all the solutions in the initial population are produced randomly using Latin Hypercube Sampling (LHS).

For each of the three network case studies, ten trials were conducted considering the random nature of GAs. The search convergences of the GAs during the entire evolutionary process are shown in Figure 4. The shaded areas represent the maximum and minimum ranges from 10 random trials. The bold black lines are the average values of the multiple trials. The statistics of the optimization results are shown in Table 5.

The results in Table 5 show the best solution, average value of solutions and worst solution found in multiple trial runs for each of the algorithms. From Column 3, all three initialization methods are able to achieve an optimal solution of US$3.521 million in the TRN case, since the TRN is a small case and easy to converge. However, the variation of the solutions of HDP is smaller than that of PHSM and LHS (see Figures 4a, 4b and 4c), which indicates HDP is more reliable at reaching a good solution.

In the MOD problem, the LHS obtained the best final and average solutions but is the worst on the other two problems. The final solution variation range of LHS is larger than the range of HDP in MOD. It indicates that LHS shows an unstable performance. It should be noted that the percentage difference between LHS and HDP in MOD is much smaller than the difference for TRN and BN cases. Moreover, the convergence of PHSM and HDP at the beginning is significantly better than LHS. If the computational resource is limited for the optimization problems, the preconditioning method is quite effective for accelerating the convergence within the limited computational budget. The reason why LHS can obtain better results occasionally will be explained in the discussion on Figure 5.

More importantly, this study is to explore and compare the quality of the solutions and the speed of the convergence between the two preconditioning methods that are based on the domain knowledge. For all cases HDP outperforms PHSM methods regardless of whether the comparisons are based on the best solution, on the average solution or on the worst solution in Table 5. The variation range of the HDP solutions is smaller than PHSM as shown in Figure 4. It shows that HDP can save massive computational cost compared to PHSM when achieving the same quality solution.

The results of TRN and BN cases demonstrate that the optimization started with a better initial solution may lead to the better optimized results. The results imply that the good solution that is fed into the initial population could effectively guide the search process and accelerate the search convergence. This preconditioning is effective for large networks with a high dimensional search space.

As shown in Figure 5b and 5c, the sources 1 and 4 are the main sources (i.e. the sources are linked by the larger pipes) in the obtained optimization solutions. Consistently, the main sources are the same as the initialization solutions derived from the PHSM and HDP respectively (see Figures

3c and 3d). Compared to the optimal solution derived from LHS, the only main source is Source 3 (Figure 5a). Therefore, the preconditioning methods of PHSM and HDP provide the starting points that guide the inferior search direction (i.e. not towards the optimal area), then result in the local optimum in the optimization. It may thus be inferred that the crossover operator of GAs can spread the superior genes of the solution in the population quickly, however the mutation operator is hard to assist the search to escape the local trap (preconditioning solutions). A potential reason is that the mutation probability in the large case is relatively low, for example the mutation probability in MOD is 0.003 in comparison with the TRN case of 0.07. Hence the initial network configuration (INC) design methods contribute more to the exploitation of EAs, while less to the exploration in the large WDNs.

Comparing with the Literature Solutions

Table 6 shows the comparison of optimal solutions from the literature using GA-related algorithms or mathematical programming for the MOD case. The best solution of the MOD network obtained in this paper is $2,531,934. SGA has a great potential for applying to the water network problems. The optimization solution combined with the HDP method performs well in comparison with other results from the literature (Table 6) Although the GA is used in this study, but the HDP method is easy to combine with other sophisticated optimization algorithms (e.g., Particle Swarm Optimization, Harmony Search, Differential Evolution, Dynamically Dimensioned Search).

Similarly,, for the BN case, the best solutions that are associated with GA in the literature are summarized in Table 7. Table 7 shows that the HDP initial design (added to the initial SGA population) combined with SGA derives the best solution (€ 1,941,349) among all the algorithms when executing 10 million evaluations. The preconditioning method (HDP+SGA) is better than the hybrid algorithms (Sadollah et al. 2015, Sheikholeslami et al. 2015, Tolson et al. 2009) in the BN (Balerma) case study.

Here we show HDP enables the optimization to substantially reduce the number of objective function evaluations necessary to obtain accurate solutions for search with a simple GA. Future research can explore how to combine this technique with more sophisticated optimization methods. The improved solutions are highly expected.

Conclusions

In this paper, a new heuristic design method (HDP) based on domain knowledge is proposed to provide an initial network configuration (INC) that is used in the population in the first generation of the simple GA search for water distribution network design problems. The domain knowledge includes an understanding of the physical factors affecting the relationship between pipe sizes and heads in a network. The method employs headloss analysis to determine the pipe diameters in the network through an iterative process to get a good initial guess of what an efficient allocation of pipe sizes might be. Its performance is compared with another (INC) design method (PHSM), developed by Bi et al. (2015). The results from three networks show that the HDP method is significantly superior to PHSM in terms of the quality solution and the computational burden. When the solutions from the heuristic methods are fed to the population-based GA, the performance of the algorithm has been improved substantially. More importantly, the HDP based genetic algorithm search is more efficient and effective compared to those based on PHSM.

Combining a deterministic heuristic network design method with the evolutionary algorithm is promising as it brings domain knowledge to the optimization algorithms in order to strengthen the algorithm’s search capability and convergence. This can effectively balance the global search algorithm (exploration) and guided search (exploitation) by engineering judgment to solve the optimal design problems in real-world large-scale WDNs. The performance of HDP could be tested on more case studies and their use with other optimization algorithms.

Data Availability Statement

All data, models, and code that support the findings of this study are available from the first or corresponding author upon reasonable request.

Acknowledgements

This work was initiated when Dr. Liu Haixing visited Civil and Environmental Engineering Department, National University of Singapore (NUS) with Professor Shoemaker for one year. This study was funded by the National Natural Science Foundation of China (91647201, 91747102, 51579027, 51708086). This study was also supported by the Liaoning Natural Science Foundation (2019-MS-043) and the Fundamental Research Funds for the Central Universities (DUT18RC(3)072). Prof. Shoemaker's start up grant from NUS provided partial support. We would like to thank the editors and the anonymous reviewers for their insightful comments that have helped improve the quality of the paper.

Supplemental Materials

Figures S1-S3 and Tables S1 are available in the ASCE Library (ascelibrary.org).

References

Andrade, M., Kang, D., Choi, C. and Lansey, K. (2012) Heuristic Postoptimization Approaches for Design of Water Distribution Systems. Journal of Water Resources Planning and Management 139(4), 387-395.

Bi, W., Dandy, G.C. and Maier, H.R. (2015) Improved genetic algorithm optimization of water distribution system design by incorporating domain knowledge. Environmental Modelling & Software 69, 370-381.

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A. and Toth, P. (2008) Water network design by MINLP. Rep. No. RC24495, IBM Research, Yorktown Heights, NY.

Deo, N. (1974) Graph Theory with Applications to Engineering and Computer Science (Prentice Hall

Series in Automatic Computation), Prentice-Hall, Inc.

Fu, G., Butler, D. and Khu, S.-T. (2008) Multiple objective optimal control of integrated urban wastewater systems. Environmental Modelling & Software 23(2), 225-234.

Fu, G., Kapelan, Z., Kasprzyk, J. and Reed, P. (2012a) Optimal design of water distribution systems using many-objective visual analytics. Journal of Water Resources Planning and Management 139(6), 624-633.

Fu, G., Kapelan, Z. and Reed, P. (2012b) Reducing the complexity of multiobjective water distribution system optimization through global sensitivity analysis. Journal of Water Resources Planning and Management 138(3), 196-207.

Gessler, J. (1985) Pipe network optimization by enumeration. Computer applications in water resources, 572-581.

Johns, M.B., Keedwell, E. and Savic, D. (2014) Adaptive locally constrained genetic algorithm for least-cost water distribution network design. Journal of Hydroinformatics 16(2), 288-301.

Kang, D. and Lansey, K. (2012) Revisiting Optimal Water-Distribution System Design: Issues and a Heuristic Hierarchical Approach. Journal of Water Resources Planning and Management 138(3), 208-217.

Keedwell, E. and Khu, S. (2006) Novel Cellular Automata Approach to Optimal Water Distribution Network Design. Journal of Computing in Civil Engineering 20(1), 49-56.

Kennedy, J. and Eberhart, R. (1995) Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks 4, 1942-1948 vol.1944.

Meng, F., Fu, G. and Butler, D. (2016) Water quality permitting: From end-of-pipe to operational strategies. Water Research 101, 114-126.

Montalvo, I., Izquierdo, J., Pérez‐García, R. and Herrera, M. (2014) Water Distribution System Computer‐Aided Design by Agent Swarm Optimization. Computer‐Aided Civil and Infrastructure Engineering 29(6), 433-448.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. and Zechman, E. (2010) State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management. Journal of Water Resources Planning and Management 136(4), 412-432.

Ormsbee, L. and Walski, T. (2016) World Environmental and Water Resources Congress 2016, pp. 455-464, American Society of Civil Engineers.

Reca, J. and Martínez, J. (2006) Genetic algorithms for the design of looped irrigation water distribution networks. Water Resources Research 42(5), W05416.

Rossman, L.A. (2000) EPANET 2: users manual, US EPA.

Sadollah, A., Yoo, D.G. and Kim, J.H. (2015) Improved mine blast algorithm for optimal cost design of water distribution systems. Engineering Optimization 47(12), 1602-1618.

Sheikholeslami, R., Zecchin, A.C., Zheng, F. and Talatahari, S. (2015) A hybrid cuckoo–harmony search algorithm for optimal design of water distribution systems. Journal of Hydroinformatics 18(3), 544-563.

Storn, R. and Price, K. (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11(4), 341-359.

Sweetapple, C., Fu, G. and Butler, D. (2014) Multi-objective optimisation of wastewater treatment plant control to reduce greenhouse gas emissions. Water Research 55, 52-62.

Tolson, B.A., Asadzadeh, M., Maier, H.R. and Zecchin, A. (2009) Hybrid discrete dynamically dimensioned search (HD-DDS) algorithm for water distribution system design optimization. Water Resources Research 45(12), W12416.

Tolson, B.A. and Shoemaker, C.A. (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research 43(1), W01413.

Vairavamoorthy, K. and Ali, M. (2000) Optimal design of water distribution systems using genetic algorithms. Computer ‐ Aided Civil and Infrastructure Engineering 15(5), 374-382.

Walski, T. (2015) Real-World Considerations in Water Distribution System Design. Journal of Water Resources Planning and Management 141(9), 02515002.

Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W., Beckwith, S. and Koelle, E. (2003) Advanced water distribution modeling and management, Haestad Press, Waterbury, CT.

Wang, Q., Guidolin, M., Savic, D. and Kapelan, Z. (2015) Two-Objective Design of Benchmark Problems of a Water Distribution System via MOEAs: Towards the Best-Known Approximation of the True Pareto Front. Journal of Water Resources Planning and Management 141(3), 04014060.

Wang, Y., Chen, A.S., Fu, G., Djordjević, S., Zhang, C. and Savić, D.A. (2018) An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features. Environmental Modelling & Software 107, 85-95.

Wu, Z. and Walski, T. (2005) Self-Adaptive Penalty Approach Compared with Other Constraint-Handling Techniques for Pipeline Optimization. Journal of Water Resources Planning and Management 131(3), 181-192.

Wu, Z.Y. and Simpson, A.R. (2001) Competent genetic-evolutionary optimization of water distribution systems. Journal of Computing in Civil Engineering 15(2), 89-101.

Zhang, C., Chen, X., Li, Y., Ding, W. and Fu, G. (2018) Water-energy-food nexus: Concepts, questions and methodologies. Journal of Cleaner Production 195, 625-639.

Zheng, F., Simpson, A.R. and Zecchin, A.C. (2011) A combined NLP-differential evolution algorithm approach for the optimization of looped water distribution systems. Water Resources Research 47(8), W08531.

Zheng, F., Zecchin, A., Newman, J., Maier, H. and Dandy, G. (2017) An Adaptive Convergence-Trajectory Controlled Ant Colony Optimization Algorithm with Application to Water Distribution System Design Problems. IEEE Transactions on Evolutionary Computation.

Table 1 Notation for the symbols and acronym

Symbol	Definition	Where discussed
$D_{\max}$	Maximum pipe diameter	Step 0
IT	Current iteration number and	Step 0
$IT_{\max}$	maximum number of iterations	Step 0
P_{req}	Minimum pressure threshold	Step 0 and Step 1.3
H_k	Water head of source k ($k = 1, 2, \cdots, N_s$)	Step 1.1
N_s	Number of sources	Step 1.1
N_n	Number of nodes	Step 1.1
Q_{ij}	Pipe flow rate (the pipe from node i to node j, and $i \neq j$)	Step 1.1
L_{ki}	Length of the shortest path from source k to node i	Step 1.2
P_{ki}	Shortest path ($\theta_k, \cdots, \theta_i$) from source k to node i	Step 1.2
E_i	Elevation at node i	Step 1.3
HL_{ki}	Maximum headloss for path P_{ki}	Step 1.3
UHL_{ki}	Unit headloss for path P_{ki}	Step 1.4
HL_{ij}	Headloss of pipe ij	Step 2
D_n	Updated Diameter	Step 3
D_o	Diameter obtained in the last iteration	Step 3
L_{ij}	Length of pipe ij	Equations 1 and 2
g	Gravity acceleration	Equation 2
C	Roughness coefficient	Equation 1
f	Darcy friction factor	Equation 2
HDP	Headloss-based Design Preconditioner	
PHSM	Prescreened Heuristic Sampling Method	
INC	Initial Network Configurations	
HW	Hazen-Williams equation	
DW	Darcy-Weisbach equation	
SGA	Simple genetic algorithm	
LHS	Latin hypercube sampling	

Table 2 TRN network initial solution results of heuristic design methods

	HDP		PHSM		
	Iteration	Cost (m$)	Velocity (m/s) for outer loop	Iterations for inner loop	Cost (m$)
	1	3.979	0.1	5	6.577
	2	3.918	0.2	4	5.637
			0.3	3	4.941
Total	**2**	**3.918**	**Total**	**12**	**4.941**

Table 3 MOD network initial solution results of heuristic design methods

HDP						PHSM		
Iteration	**Cost (m$)**	**Iteration**	**Cost (m$)**	**Iteration**	**Cost (m$)**	**Velocity (m/s) for outer loop**	**Iterations for inner loop**	**Cost (m$)**
1	7.556	8	2.872	15	2.839	0.1	24	21.87
2	3.041 (1.98)[a]	9	2.881	16	2.829	0.2	39	14.63
3	2.945 (0.34)[a]	10	2.882	17	2.829	0.3	28	9.530
4	2.869	11	2.862	18	2.831	0.4	12	4.311
5	2.873	12	2.855	19	2.823	0.5	39	3.779
6	2.844	13	2.846			0.6	8	3.385
7	2.864	14	2.844					
Total				**19**	**2.823**	**Total**	**150**	**3.385**

[a] The sum of pressure violations (unit: m).

Note: The velocity constraint is not violated.

Table 4 BN network results of heuristic design methods

	HDP		PHSM		
	Iteration	Cost (m€)	Velocity (m/s) for outer loop	Iterations for inner loop	Cost (m€)
	1	4.004	0.1	9	13.38
	2	2.687	0.2	19	10.83
	3	2.604	0.3	16	8.982
	4	2.504	0.4	20	8.160
	5	2.503	0.5	27	7.295
	6	2.461	0.6	20	6.255
	7	2.447	0.7	14	5.094
	8	2.426	0.8	19	4.476
	9	2.428	0.9	16	3.958
	10	2.427	1	11	3.716
	11	2.429	1.1	9	3.466
Total	**11**	**2.429**	**Total**	**180**	**3.466**

Table 5 Optimization results for three network cases

Problem	Initialization method	Number of different trial runs	Best solution found ($m)	Average cost solution ($m)	Worst solution found ($m)	Average number of evaluations used by HDP to find equivalent best solution in PHSM
	1	2	3	4	5	6
TRN	LHS	50	**3.521**	3.612	3.995	**-**
	PHSM	50	**3.521**	3.599	3.717	21,667
	HDP	50	**3.521**	**3.571**	**3.642**	**7,500**
MOD	LHS	10	**2.532**	**2.557**	2.576	**-**
	PHSM	10	2.617	2.672	2.723	9,550,000
	HDP	10	2.549	2.564	**2.574**	**730,000**
BN	LHS	10	2.151	2.189	2.2.13	**-**
	PHSM	10	1.986	1.999	2.014	9,500,000
	HDP	10	**1.941**	**1.954**	**1.980**	**1,300,000**

Table 6 The best solutions of the MOD network in the literature

MOD references	Algorithm	Best solution (m$)	Computational budget (Evaluations)
This study	SGA	**2.532**	10,000,000
	HDP+SGA	2.549	10,000,000
Bragalli et al. (2008)	MINLP	2.565	7,200(s)[a]
Bragalli et al. (2012)	BONMIN	2.577	7,200(s) [a]

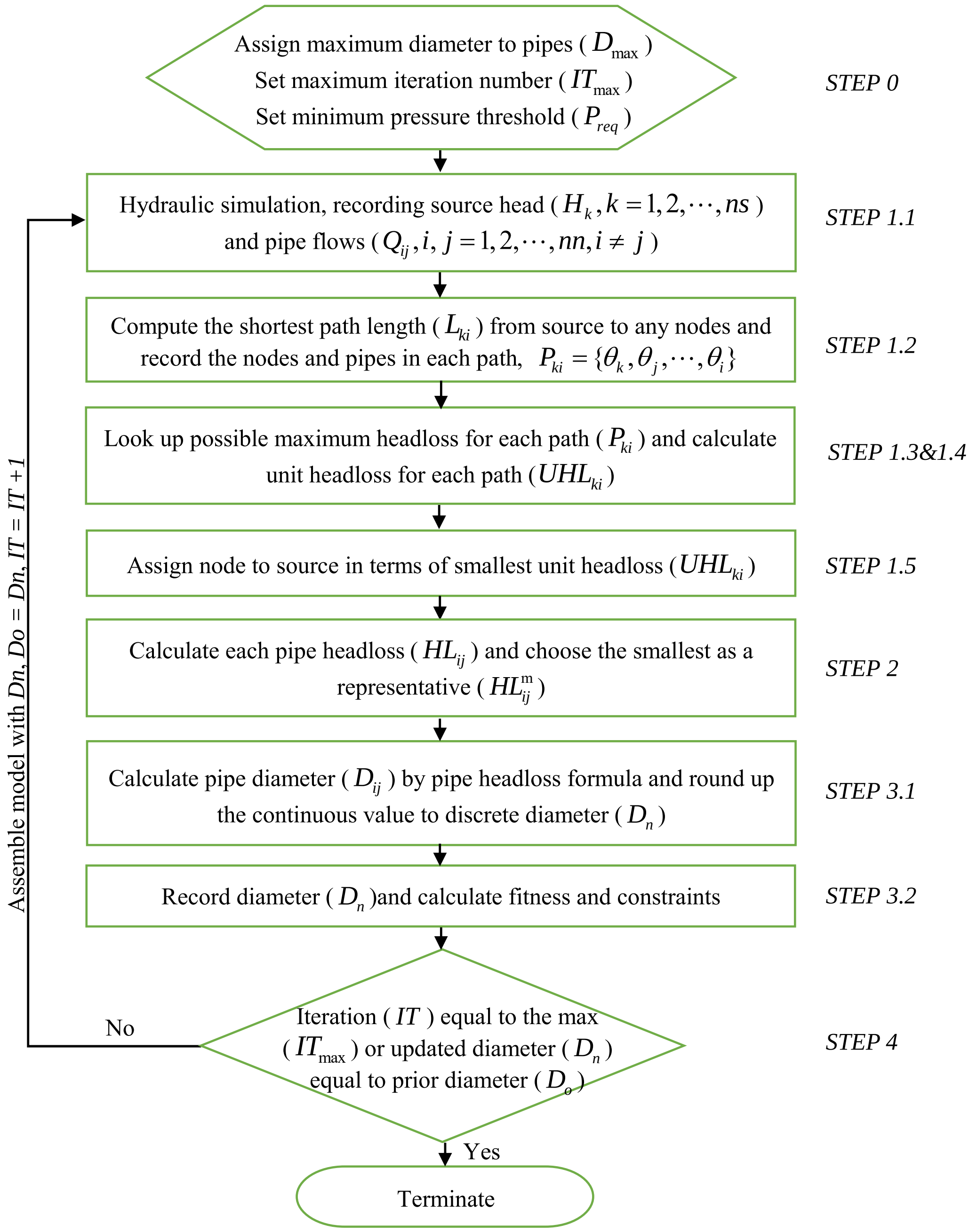
[a] The time is used to obtain the solution.

"+" represent two algorithms are conducted subsequently.

Table 7 The best solutions of the BN networks in the literature

BN references	Algorithm	Best solution (m€)	Computational budget (Evaluations)
This study	SGA	2.151	10,000,000
	HDP+SGA	**1.941**	10,000,000
Bi et al. (2015)	PHSM+SGA	2.061	1,000,000
Reca and Martínez (2006)	GENOME	2.302	10,000,000
Tolson et al.(2009)	HDDDS	1.956	10,000,000
	HDDDS+Local	1.941	30,000,000
Sadollah et al. (2015)	IMBA	2.014	250,000
Sheikholeslami et al. (2015)	CSHS	1.988	3,000,000
Sheikholeslami and Talatahari (2016)	BB-BC-PSO	1.987	3,000,000

"+" represent two algorithms are conducted subsequently.


Assign maximum diameter to pipes ($D_{\max}$)
Set maximum iteration number ($IT_{\max}$)
Set minimum pressure threshold (P_{req})
STEP 0
Hydraulic simulation, recording source head ($H_k, k=1,2,\cdots,ns$) and pipe flows ($Q_{ij}, i, j=1,2,\cdots,nn, i\neq j$)
STEP 1.1
Compute the shortest path length (L_{ki}) from source to any nodes and record the nodes and pipes in each path, $P_{ki}=\{\theta_k,\theta_j,\cdots,\theta_i\}$
STEP 1.2
Look up possible maximum headloss for each path (P_{ki}) and calculate unit headloss for each path (UHL_{ki})
STEP 1.3&1.4
Assign node to source in terms of smallest unit headloss (UHL_{ki})
STEP 1.5
Calculate each pipe headloss (HL_{ij}) and choose the smallest as a representative (HL^{m}_{ij})
STEP 2
Calculate pipe diameter (D_{ij}) by pipe headloss formula and round up the continuous value to discrete diameter (D_n)
STEP 3.1
Record diameter (D_n)and calculate fitness and constraints
STEP 3.2
Iteration (IT) equal to the max ($IT_{\max}$) or updated diameter (D_n) equal to prior diameter (D_o)
STEP 4
No
Assemble model with Dn, $Do = Dn$, $IT = IT+1$
Yes
Terminate

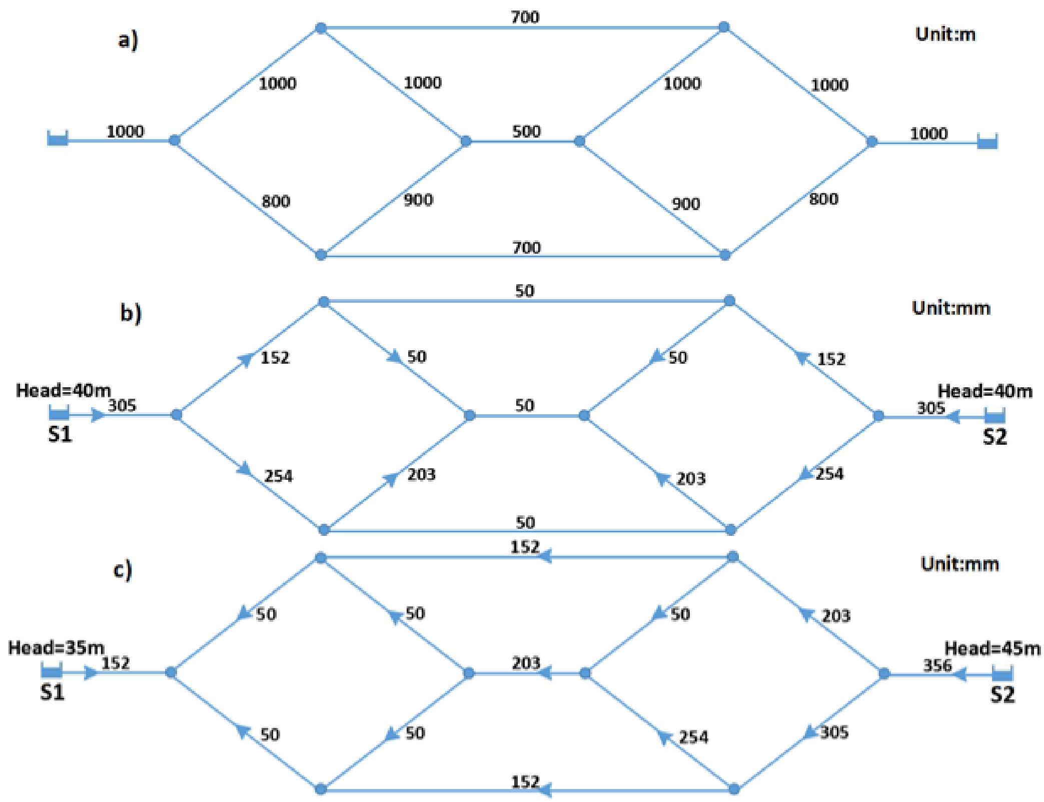

a)
Unit:m
700
1000
1000
1000
1000
1000
500
1000
800
900
900
800
700
b)
Unit:mm
50
152
50
50
152
Head=40m
305
S1
50
Head=40m
305
S2
254
203
203
254
50
c)
152
Unit:mm
50
50
50
203
Head=35m
152
S1
203
Head=45m
356
S2
50
50
254
305
152

Figure3

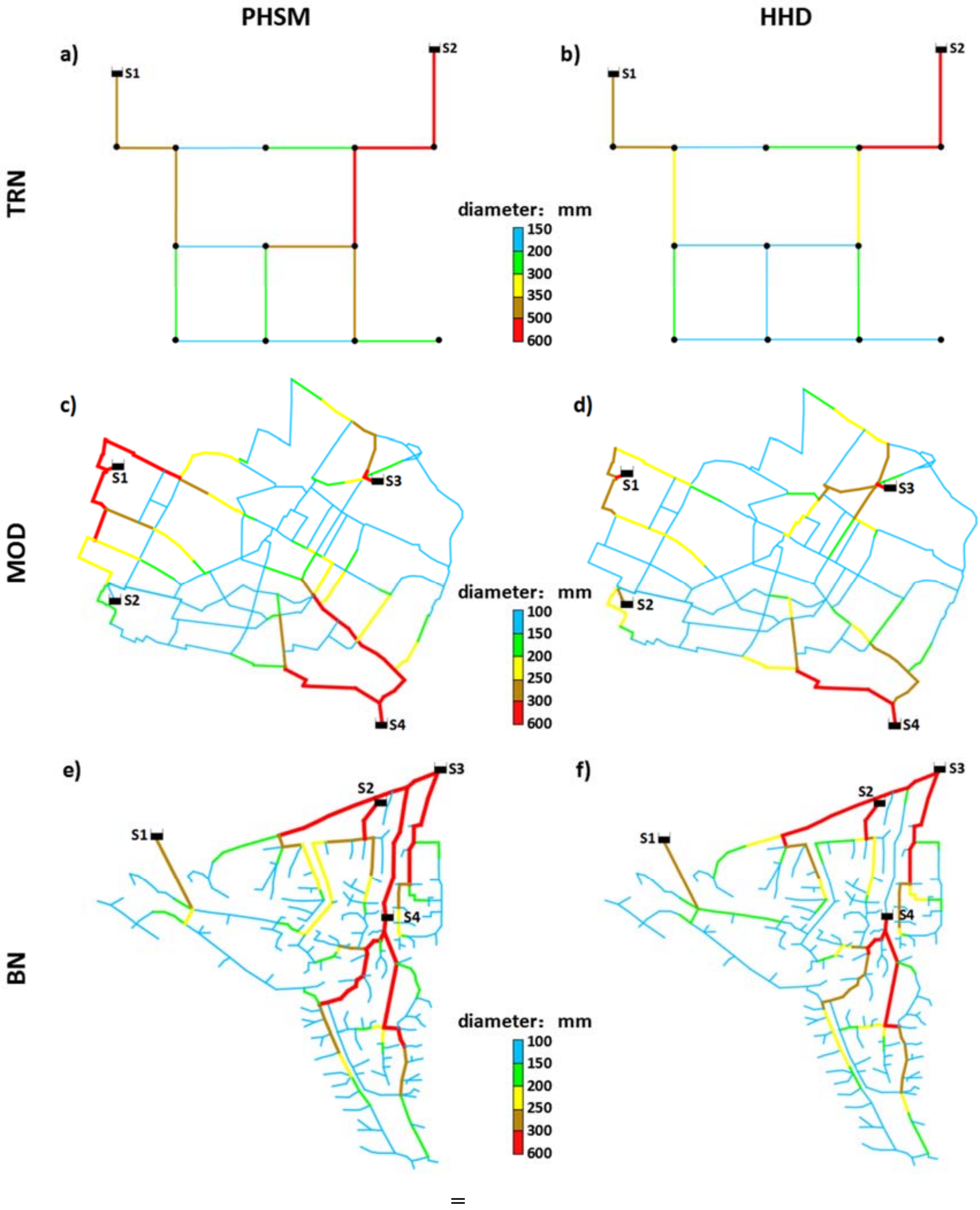

PHSM
HHD
TRN
MOD
BN
a)
b)
c)
d)
e)
f)
S1
S2
S3
S4
diameter: mm
150
200
300
350
500
600
diameter: mm
100
150
200
250
300
600
diameter: mm
100
150
200
250
300
600

Figure4

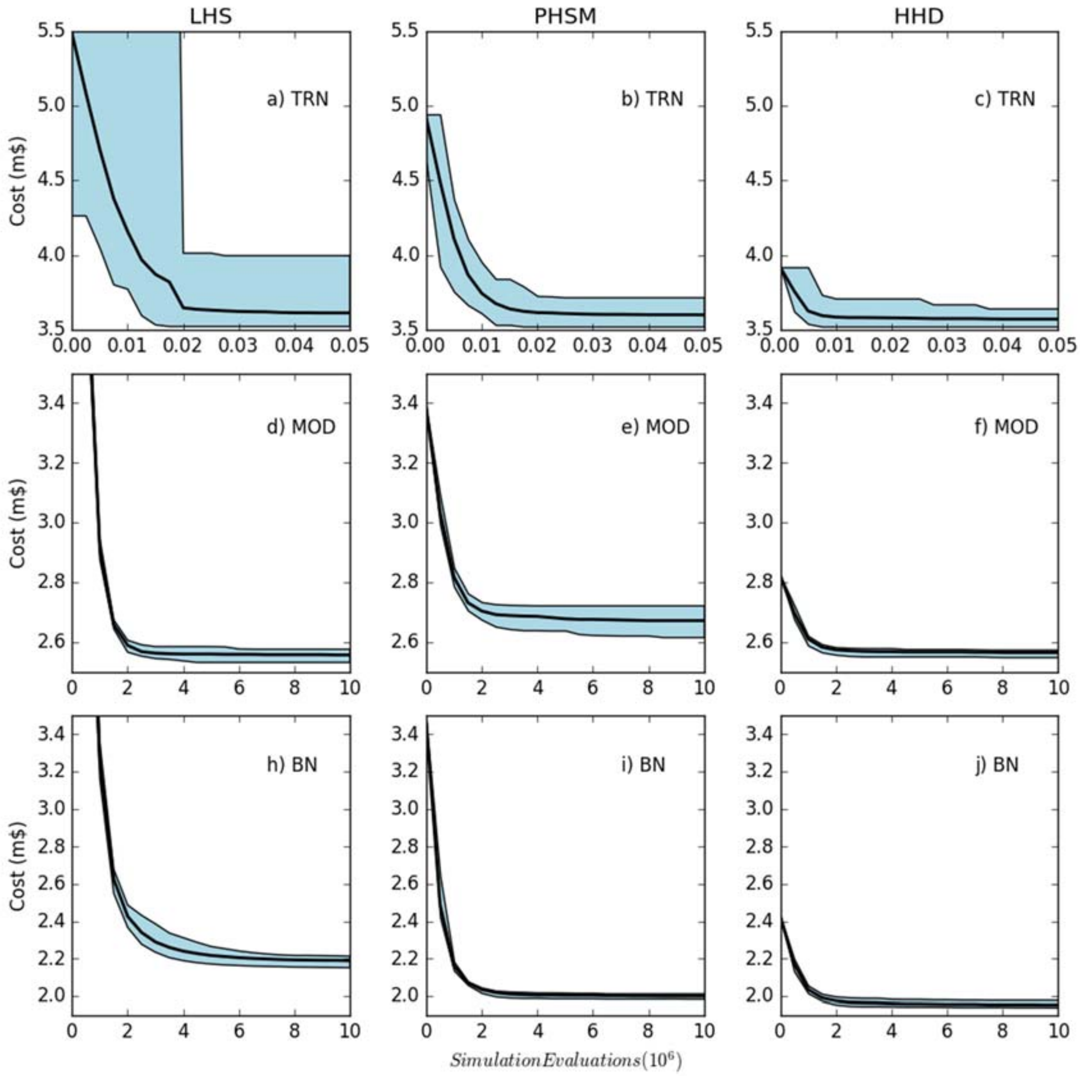

LHS
PHSM
HHD
a) TRN
b) TRN
c) TRN
d) MOD
e) MOD
f) MOD
h) BN
i) BN
j) BN
Cost (m$)
Cost (m$)
Cost (m$)
5.5
5.0
4.5
4.0
3.5
0.00
0.01
0.02
0.03
0.04
0.05
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
0
2
4
6
8
10
$Simulation Evaluations (10^6)$

Figure1

Click here to access/download;Figure;Fig 5.pdf

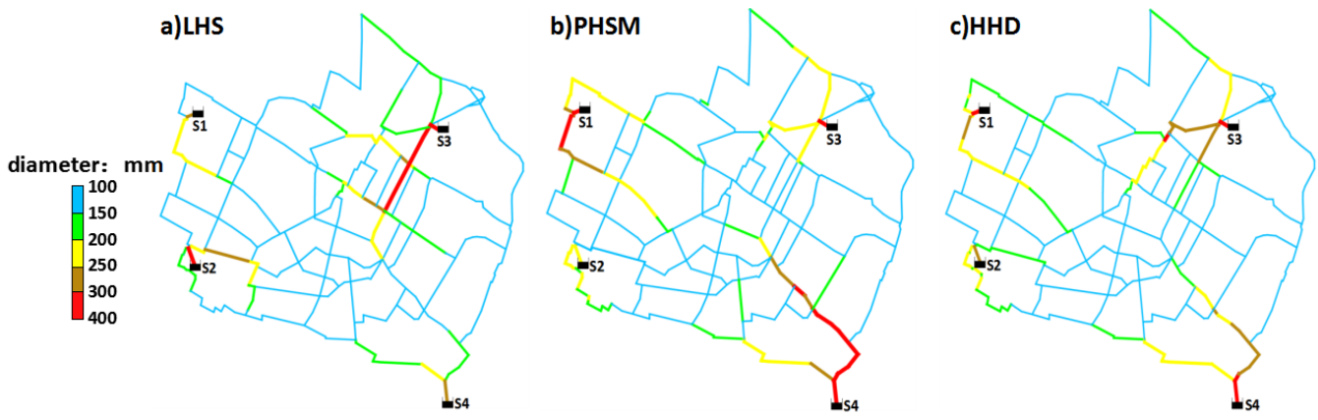


Figure 1 Flowchart of Headloss-based Design Preconditioner (HDP) algorithm

Figure 2 The illustrative case network: Pane a) is the topology of the network where the numbers represent pipe lengths (m). Panes b) and c) are the results (pipe diameters and flow directions) of HDP from equal (Scenario 1) and unequal (Scenario 2) heads of sources (m), respectively, where the numbers represent pipe diameters (mm).

Figure 3 Comparison of the initial network configurations derived from PHSM and HDP methods. Panels a) and b) show the results of the initial solutions in Two Reservoir problem derived from PHSM and HDP, respectively. Panels c) and d) are for the MOD problem; Panels e) and f) are for the BN problem.

Figure 4 Convergence of the GAs preconditioned on the solutions from PHSM and HDP for three WDNs. “LHS” refers to a GA with an initial population selected by Latin Hypercube sampling. The shaded areas represent the ranges (minimum to maximum) from 10 random trials.

Figure 5 Network configurations of the optimal solutions in MOD