1   # Reliability surrogate measures for water distribution system

2   # design: a comparative analysis

3   Haixing Liu[1], Dragan A. Savić[2], Zoran Kapelan[3], Enrico Creaco[4], Yixing Yuan[5]

4   [1] Associate Research Fellow, School of Hydraulic Engineering, Dalian University of

5   Technology, Dalian 116024, China; Formerly, State Key Laboratory of Urban Water

6   Resource and Environment, School of Municipal and Environmental Engineering, Harbin

7   Institute of Technology, Harbin, 150090, China. E-mail: lhx_526@163.com

8   [2] Professor, College of Engineering, Mathematics and Physical Sciences, University of

9   Exeter, Exeter, EX4 4QF, UK. E-mail: d.savic@exeter.ac.uk

10   [3] Professor, College of Engineering, Mathematics and Physical Sciences, University of

11   Exeter, Exeter, EX4 4QF, UK. E-mail: z.kapelan@exeter.ac.uk

12   [4] Assistant Professor, Dipartimento di Ingegneria Civile ed Architettura, University of Pavia,

13   Via Ferrata 3, 27100 Pavia, Italy. E-mail: creaco@unipv.it

14   [5] Professor, State Key Laboratory of Urban Water Resource and Environment, School of

15   Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin, 150090,

16   China. E-mail: yyx1957@163.com

17   Corresponding author: Haixing Liu

18   Email:lhx_526@163.com

Abstract:

Due to the large computational burden associated with the direct assessment of reliability, the indirect indices of reliability have recently received more attention in the framework of water distribution system design. Two new energy related indices for reliability evaluation (i.e., available power index and pipe hydraulic resilience index) are developed in this paper. The performance of these new indices is evaluated and compared with that of four existing indices (other three energy related indices, i.e., resilience index, network resilience index, modified resilience index, and the entropy-based method, i.e., diameter-sensitive flow entropy) according to the following two-step methodology. In the first step, the application of the multi-objective optimization makes it possible to determine optimal network configurations that trade-off the installation cost (to be minimized) against the generic indirect reliability index (to be maximized). In the second step, the performance of the optimal solutions in terms of explicit reliability assessment is examined under conditions in which the original network is perturbed by applying demand variations and random pipe failures to account for future operating uncertainties. The Hanoi and the Fossolo benchmark networks are used as case studies. The results obtained show that energy based indices yield an overall superior estimate of reliability in comparison with the diameter-sensitive flow entropy. Furthermore, the new indices show some advantages in the evaluations performed under demand and pipe failure uncertainties.

Introduction

41 Traditionally, the optimal design of water distribution systems (WDSs) in the single objective

42 optimisation framework is to obtain the least cost solution in which optional pipe diameters

43 are usually the decision variables and the water sources and layout are known a priori [Savic

44 and Walters, 1997]. In the drive to improve cost effectiveness and ensure water security, the

45 WDS design should incorporate many more aspects [Walski, 2001]. A multiobjective

46 optimisation model generally takes cost as a common objective and combines one or more

47 benefit objectives, such as reliability, robustness, water quality, into a trade-off relationship

48 [Farmani et al., 2005; Kapelan et al., 2005; Creaco and Franchini, 2012; Ostfeld, 2012;

49 Marchi et al., 2014]. Reliability can be classified into three categories: mechanical reliability,

50 hydraulic reliability, and water quality reliability [Ostfeld, 2004; Atkinson et al., 2014;

51 Shafiqul Islam et al., 2014; Gheisi and Naser, 2015]. Reliability is defined as the ability of a

52 WDS to provide an adequate level of service to customers under both normal and abnormal

53 operating conditions within a prescribed time interval [Xu and Goulter, 1999]. Although Jung

54 et al. [2014] argued that robustness definition covers additional information that assesses not

55 only the probability of pressure meeting requirement but also the variation in pressure,

56 reliability undoubtedly remains the most widely applied method for the assessment of WDS

57 performance and is substantially investigated by the scientific community [Dandy and

58   Engelhardt, 2006; Martinez, 2010; Ostfeld, 2012; Piratla and Ariaratnam, 2012; Torii and

59   Lopez, 2012; Shafiqul Islam et al., 2014].

60   Disruptions in supply could potentially result in a financial loss to a WDS operator. The

61   concern of how to improve reliability of supply during the design phase is vital. Fujiwara and

62   Tung [1991] stated that reliability could be improved by increasing pipe size. Additionally,

63   tanks play an important role in reliability improvement studies as they help balance water

64   supply and demand in the system, as well as supply water during an emergency situation

65   [Farmani et al., 2005; van Zyl et al., 2008]. Walski [1993] highlighted the importance of

66   adequate isolation valve systems for providing reliability for WDS. Other methods, e.g.,

67   adaptive pump operation, multiple sources, are also important for reliability assessment

68   [Zhuang et al., 2013].

69   Reliability is usually assessed using a probabilistic approach, such as Monte Carlo Simulation,

70   First Order Reliability Method and/or the Cut-set method [Park et al., 1998]. These

71   methodologies are computationally expensive because the assessment involves sampling

72   methods or a traversing process across all combinations of pipe failure. Thus they are

73   difficult to implement within the multiobjective optimisation design methods for a real WDS

74   [Bao and Mays, 1990; Maier et al., 2001; Kapelan et al., 2005; Yannopoulos and Spiliotis,

75   2013]. As WDS is a complex system and changes over time, the network design that is

76   assessed to be reliable at the design stage is likely to become unreliable in meeting the

77   performance requirements in the future. Another related concept – resilience – can be used to

78   assess reliability indirectly in both conventional [Farmani et al., 2005] and adaptive WDS

79   design [Basupi and Kapelan 2013]. Resilience is defined as the ability of system to 'bounce

80   back' following some failure and hence can be estimated as the capacity of a system to

81   absorb disturbance while the system is undergoing changes so as to still retain essentially the

82   same function, structure, identity and feedbacks [Folke, 2006]. Todini [2000] proposed the

83   surplus power at nodes to characterise the resilience of looped water networks. Many

84   improvements and applications associated with Todini's resilience index appeared thereafter

85   [Prasad and Park, 2004; Farmani et al., 2005; Jayaram and Srinivasan, 2008; Basupi et al.,

86   2013]. Creaco et al. [2014] showed that a good indication of the network reliability under

87   such critical operation scenarios as those related to pipe isolation and hydrant activation can

88   be obtained when the multi-objective optimization is performed considering three objective

89   functions: 1) cost to be minimized; 2) Todini's resilience to be maximized; 3) loop diameter

90   uniformity to be maximized. Despite the advantages of the methodology proposed by Creaco

91   et al. [2014], the use of three objectives may be too cumbersome in the optimization context

92   for a large network problem.

93   Other studies have developed indices to assess reliability of water supply. Flow entropy

94   characterising uniformity of pipe flows can be employed to assess reliability of a WDS

95   [Awumah et al., 1990; Tanyimboh and Templeman, 2000; Liu et al., 2014]. Park et al. [1998]

96   introduced an indicator of hydraulic power capacity, which combines flow and pressure in a

97   single dimension for measuring capacity (or reliability) of a WDS. It is, therefore, of interest

98   to researchers and practitioners to determine which reliability assessment method is the most

99   appropriate measure for optimal design studies.

100   Raad et al. [2010] conducted a comparative analysis of four reliability surrogate measures

101   based on the assumptions of demand and pipe failure uncertainties. They only considered one

102   assessment indicator, i.e., the average demand satisfaction as a result of the demand

103   variability and pipe failure consequence analyses. Baños et al. [2011] compared three existing

104   reliability indices by using the indicators of average percentage of unfeasible scenarios and

105   average minimum over-demand in the scheme of demand uncertainty. Most recently,

106   Atkinson et al. [2014] investigated the correlations amongst resilience index, entropy and

107   minimum surplus head, and examined the hydraulic performance of the solutions for the

108   Anytown benchmark network.

109   Despite the many existing contributions to the field, the assessment of the best reliability

110   surrogate measure(s) is still an open issue. In this paper, two new measures are proposed and

111   their performance compared to four existing measures in the context of the multiobjective

112   design of water distribution systems. In particular, a novel *a posteriori* performance

113   assessment is used to analyse and compare the surrogate measures in two benchmark

114   networks.

115   The rest of the paper is organized as follows. Firstly, the surrogate measures (including two

116   newly developed) are presented. Secondly, the optimization process and an a posteriori

117   performance assessment method are described. Lastly, the applications are demonstrated and

118   the conclusions drawn. Note that several abbreviations are used in this paper. A list of these

119   abbreviations can be found in Table 1.

Reliability surrogate measures

In this study, three existing reliability surrogate measures from literature are analysed, including the resilience index (RI) [Todini, 2000], network resilience index (NRI) [Prasad and Park, 2004] and modified resilience index (MRI) [Jayaram and Srinivasan, 2008]. The above three measures are all derived from the power analysis of water systems. The diameter-sensitive flow entropy (DSFE), newly introduced reliability index based on an extension of the flow entropy formulation [Liu et al., 2014], is also considered. Furthermore, the analysis is also extended to two new measures, the available power index (API) and the pipe hydraulic resilience index (PHRI), which belong to the classes of the power and pipe hydraulic gradient measures, respectively.

The following subsections report the formulations of the two new measures (API and PHRI). The other four surrogate measures (RI, NRI, MRI and DSFE) are reviewed in the supplementary material.

Available power index

A WDS is a dynamic, constantly changing and balanced system, which obeys the energy conservation law. The input power (energy per unit time) is derived from sources, pumps and tanks located in high elevations, while the output power could be divided into the available and unavailable power. The available power represents the output power at demand nodes; the unavailable power includes pipe friction losses, minor losses from pumps and bends, and leakage. The output power function (P_{out}) is given by

140
$$P_{out} = P_{ava} + P_{unava} \tag{1}$$

141   where P_{ava} is the available portion of output power; P_{unava} is the unavailable portion.

142   A WDS with higher reliability has the additional capacity to resist the perturbation caused by

143   external uncertainty sources. An obvious way to improve reliability would be to provide a

144   high level of available power while reducing the unavailable power as much as possible. In

145   this paper, a new method, the available power index, is introduced to measure reliability,

146   which is defined as a ratio of the available power (P_{ava}) to the total output power (P_{out}). The

147   available power index is formulated as,

148
$$API = \frac{P_{ava}}{P_{out}} \tag{2}$$

149   The available power is delivered to users, which is calculated in terms of flow (Q_i) and head

150   ($H_{ava,i}$) at each demand node, shown as follows,

151
$$P_{ava} = \gamma \sum_{i=1}^{nnode} Q_i H_{ava,i} \tag{3}$$

152   The input power (P_{in}) includes the power introduced into the network by pumps (P_k) and

153   provided by reservoirs and tanks.

154
$$P_{in} = \gamma \sum_{j=1}^{nreservoir} Q_j H_j + \sum_{k=1}^{npump} P_k + \gamma \sum_{l=1}^{ntank} Q_l H_l \tag{4}$$

155   where Q_l is the tank flow into or out of the network and H_l is the free surface elevation at

156   tank l; $ntank$ is the number of tanks. The flow of tank is positive when the water comes

157 out of the tank, and vice versa. Due to the input power equal to the output power ($P_{in} = P_{out}$),

158 Equation 2 could be substituted with Equation 3 and 4. The available power index is written

159 as,

160
$$API = \frac{\sum_{i=1}^{nnode} Q_i H_{ava,i}}{\sum_{j=1}^{nreservoir} Q_j H_j + \sum_{k=1}^{npump} \frac{P_k}{\gamma} + \sum_{l=1}^{ntank} Q_l H_l}$$
(5)

161 A tank can be viewed as a feeding or demanding node. The tank in Equation 5 is put in the

162 supply term, since the tank power avoids the denominator becoming zero when the pump

163 does not supply water to the pressurized system. The available power index is the limiting

164 form of Todini's resilience index, i.e., RI's required head equal to zero. However, they have

165 entirely different physical meanings. API is based on the analysis of the energy conservation

166 principle and is expressed as the available power in a WDS, while the RI concept emphasizes

167 the surplus power in the system. In addition, due to requiring one parameter less (i.e., the

168 required head at demand nodes), the available power index is more straightforward and

169 readily computed in comparison with the resilience index. Furthermore, the available power

170 index explicitly expresses the significance of tanks. This is essential, since tanks not only

171 balance the flow difference between on-peak and off-peak periods, but also store energy

172 during normal operation and emergency situations (e.g., fire fighting, pipe breaks). Hence

173 tanks play a key role of energy transfer within a WDS are necessary in the analysis of the

174 available power.

Pipe hydraulic resilience index

176 The aforementioned methods associated with resilience of WDSs belong to a same class in

177 which power at nodes is taken into account. A new method in terms of the hydraulic gradient

178 along pipes is developed and presented here. Hydraulic gradient is related to the head

179 difference at both ends of a pipe and the pipe length. When water flows through a pipe, the

180 hydraulic grade line (HGL) will decline due to friction loss, as shown in Figure 1. If upstream

181 pipes dissipate less energy, the downstream pipes will obtain more head. The magnitude of

182 available head could reflect the capacity of a WDS to handle the perturbation caused by

183 uncertainty. However, available heads are insufficient to quantify reliability on their own, as

184 one metre head loss in a pipe of 100-metre length differs from that of a one-kilometre long

185 pipe. Hence, a new measure for WDS resilience which combines pressure head and pipe

186 length is developed.

187 The product of pressure head and pipe length is represented by an area, as shown in Figure 1.

188 The blue triangle area (A_i) is characterized by the head loss and the projection length of the

189 pipe, while the red triangle area (S_i) is determined by the available head and the projection

190 length of the pipe. Pipe hydraulic resilience index is defined as

191
$$PHRI = \frac{\sum_{i=1}^{n} S_i}{\sum_{i=1}^{n} (A_i + S_i)} \qquad (6)$$

192 where n is the number of pipes in the WDS. It is worth noting that the areas are summed up

193 prior to being divided in order to avoid simultaneous reduction of pipe length. The functions

194 of area S_i, $S_i + A_i$ and the projection length of pipe i $\left(L_{pro,i}\right)$ are expressed as follows,

$$S_i = \frac{1}{2}\left(H_{ds,i} - H_{req}\right)L_{pro,i} \tag{7}$$

$$S_i + A_i = \frac{1}{2}\left(H_{us,i} - H_{req}\right)L_{pro,i} \tag{8}$$

$$L_{pro,i} = \sqrt{L_i^2 - \left(Z_{us,i} - Z_{ds,i}\right)^2} \tag{9}$$

198 where $H_{ds,i}$ and $H_{us,i}$ are the heads at downstream and upstream nodes, respectively, for

199 pipe i; H_{req} is the required head of the system; L_i and $L_{pro,i}$ are the length and the

200 projection length, respectively, for pipe i. $Z_{ds,i}$ and $Z_{us,i}$ are the elevations at downstream

201 and upstream nodes for pipe i.

202 ## Assessment and comparison methodology

203 A comprehensive framework is established here to evaluate and compare the surrogate

204 reliability measures. The assessment procedure is based on a set of solutions derived from the

205 optimal design of WDSs associated with each reliability measure. The design samples (i.e.,

206 the optimal solutions of network design) are tested in scenarios perturbed by demand spatial

207 fluctuation and random pipe failure uncertainties. The evaluation indicators are proposed to

208 assess the performances of the design samples, and then determine the performance of

209 reliability surrogate measures.

210   The sample of WDS designs is obtained by applying a multiobjective optimisation process.

211   Each of reliability surrogate measures introduced above is respectively used as a conflicting

212   objective against the WDS design cost. The optional diameters of new pipes are the only

213   decision variables. A well-known multiobjective optimisation algorithm, the elitist

214   Non-dominated Sorting Genetic Algorithm II (NSGA-II), is used to solve this problem [Deb

215   et al., 2002]. A modified version of EPANET2 [Rossman, 2000] is used as the hydraulic

216   engine. In order to compare the performance of the six different surrogate measures, the

217   optimal solutions for each pair of surrogate reliability and cost (i.e. six different formulations)

218   are respectively derived by the optimisation process. The same optimization process is

219   conducted ten times for each of the six formulations to avoid the influence of the randomness

220   of generating initial solutions. Ten sets of non-dominated solutions are aggregated to yield an

221   optimal solution set by using Pareto dominance concept.

222   *WDS performance in perturbed scenarios*

223   The uncertainty of WDSs stems from a variety of aspects, while in this paper the two

224   common uncertainties of demand spatial fluctuation and pipe random failure are considered.

225   Probabilistic methods are typically used for handling uncertainty analysis in WDSs [Kapelan

226   et al., 2005; Filion et al., 2007; Sumer and Lansey, 2009; Fu and Kapelan, 2011]. Moreover, a

227   marginal quantity method could be used to quantify the ability of systems to withstand

228   uncertain incidents. The marginal quantity could be represented by a number or a polyhedron

229   [Housh et al., 2011; Sun et al., 2011]. In order to establish the accessible and versatile

230   evaluation indicators for reliability, the probabilistic method is still applied herein.

231   The demand uncertainty is often characterised by an appropriate probability density function

232   (PDF) through a considerable amount of sampling [Fu and Kapelan, 2011]. Many researchers

233   conducted the analysis for demand fluctuation in the assumption of uncertainty following

234   various PDFs including normal, log normal, uniform, β, Pearson type III, to name a few [Bao

235   and Mays, 1990; Giustolisi et al., 2009; Kang et al., 2009]. The normal distribution is used

236   most frequently out of those mentioned above, and thus is adopted in this paper. The mean of

237   normal distribution is equal to the design demand. The coefficient of variation is assumed to

238   be 0.5, which implies a wide range of demand variation, e.g., when the change of usage type

239   occurs due to area development. The Latin Hypercube Sampling (LHS) is used to sample

240   varying demand in terms of the PDF for each node. The Latin Hypercube Sampling (LHS)

241   technique is selected here to reduce the computational effort substantially in comparison with

242   the Monte Carlo sampling procedure [Kang et al., 2009]. LHS is implemented based on 1,000

243   samples in this study.

244   The second part of the assessment framework for reliability surrogate measures is established

245   using perturbation related to pipe failure. Under the assumption that a failed pipe can be

246   specified and isolated, the procedure involves the assessment of the system performance. To

247   this end a single pipe isolation is considered, since the probability of simultaneous occurrence

248   of two pipes failure is extremely low, and the procedure then traverses each pipe one by one

249   across the network. Raad et al. [2010] conducted the comparison of reliability measures

250   associated with a single pipe failure under the assumption of the equal probability of failure

251   for all pipes. However, the failure probability of pipes with large diameters is much lower

252　than in the case of small-diameter ones, since former have thicker walls and are usually better

253　maintained (due to higher importance in the system). The new method combines the

254　probability of pipe failure with the reliability evaluation indicators. Some common

255　assumptions with respect to pipe failure are stated as follows:

256　　1) N-type isolation valves: There are isolation valves at both ends of every pipe, and thus

257　　the isolation segment is just limited to a pipe;

258　　2) Simulation of a failed pipe: A pipe outage is simulated by closing the corresponding

259　　　link in the model; and

260　　3) Nodal demand: Water is distributed along pipes. However, it is common modelling

261　　practice to aggregate demand at nodes. When a given pipe is closed, the reduced quantity

262　　supplied by that pipe is ignored.

263　The testing for demand fluctuation and pipe failure intervention are both based on the same

264　set of design solutions for each reliability surrogate measure, but they are independent and

265　executed separately.

Reliability evaluation indicators

267　The best performance associated with reliability is achieved when the solution sample is able

268　to sustain in a normal operation/level of service under various kinds of perturbation or have

269　the least impact on the performance under the failed scenarios. A failed scenario refers to the

270　case when there is at least one failed node in the system, i.e., such that it cannot achieve the

271　prescribed level of service (supplied water cannot meet the demand). Farmani et al. [2005]

272   indicated that a reliability measure should reflect the way in which water users are affected,

273   i.e., the number of users with restricted or no service and the degree (magnitude) of the effect

274   on the service level. Improved assessment indicators are introduced in this study, consisting

275   of:

276   ● The sum of the number of scenarios with at least one failed node;

277   ● The spatial extent of failed nodes in the network, which is represented by a failed

278   node count under a particular failed scenario;

279   ● The failure magnitude associated with the level of service at nodes.

280   In order to evaluate these indicators, pressure driven demand (PDD) needs to be integrated

281   into the WDS hydraulic solver. The modified version of EPANET2 [Morley and Tricarico,

282   2008] is implemented in this study. Three global indicators of system performance are

283   proposed as follows:

284   ● *Percentage of Failure Scenarios (PFS).* For each solution (i.e., network design), the

285   PFS function returns the percentage of the total evaluated scenarios in which the

286   network fails to achieve the prescribed demand target for at least one node.

287   ● *Failure Node Count (FNC).* For each failed scenario, the FNC function returns the

288   total number of failed nodes where the demand at nodes does not meet the required

289   level of service. For all failed scenarios of a given optimal solution, the average of

290   FNC $\left(FNC_{mean} \right)$, the median of FNC $\left(FNC_{median} \right)$, the 25th percentile of FNC

291   $\left(FNC_{25th} \right)$ and the 75th percentile of FNC $\left(FNC_{75th} \right)$ could be calculated.

292   ● *Fraction of Failure Degree (FFD).* For failed nodes, the FFD function returns a ratio

293   of the inadequate amount of water (i.e. the difference between the demand and the

294   actual supplied water at a node) to the sum of the water required. Similarly, some

295   statistics with respect to the FFD indicator are proposed: the average of FFD

296    $\left(FFD_{mean} \right)$, the median of FFD $\left(FFD_{median} \right)$, the 25th percentile of FFD $\left(FFD_{25th} \right)$

297   and the 75th percentile of FFD $\left(FFD_{75th} \right)$.

298 The greater the values of PFS, FNC and FFD, the greater the impact of perturbation on the

299 network is, and thus the worse the performance of reliability surrogate measure. A method for

300 evaluating overall performance of all solutions of a surrogate measure is proposed, which is

301 defined as an integral of an indicator value at a fixed cost step. In detail, the integral method

302 is implemented by the sum of the indicator values for each solution multiplied by the cost

303 step between two consecutive solutions. The smaller the integral value, the better the overall

304 performance of the reliability surrogate measure is. The integral method is mathematically

305 expressed as follows:

306
$$OP = \int_{\omega} f\left(P_{indicator} \right) \, d\mathbb{P}(\omega) \tag{10}$$

307 where OP is the overall performance of an indicator at a range of costs; $\mathbb{P}(\omega)$ is the

308 vector space of costs; ω is the cost variable; $f(P_{indicator})$ represents the statistical method

309 of an indicator; $P_{indicator}$ represents the value of a reliability evaluation indicator

Case studies

Benchmark networks

We use two different case studies, the Hanoi and Fossolo benchmark networks, to demonstrate and test our methodology. The comparative analysis of reliability surrogate measures in the first case study is based on a modified benchmark network [Fujiwara and Khang, 1990]. The original Hanoi network is supplemented by a tank with an initial head of 80 m which is connected to the network by a 100m-long pipe, as shown in Figure 2(a). The network is flat as the elevation of all nodes (apart from the elevated tank) are the same. The head at all nodes of the network is required to be larger than or equal to 30 m. The search space for the Hanoi network design problem is equal to 2.87×10^{26}.

The second case study is a benchmark network adopted from Wang et al. [2014]. The Fossolo network, which includes 58 pipes, 36 demand nodes, and one reservoir with a fixed head of 121 m, is a much more complex design problem to solve as its search space is equal to 7.25×10^{77}. The Fossolo network shows a variation of node elevation, ranging from 61.24 m to 67.90 m. The minimum pressure head of all the demand nodes is to be maintained at or above 40 m, and a maximum pressure head is specified for each node, respectively. In addition, the flow velocity in each pipe should be no greater than 1 m/s. Figure 2(b) depicts the layout of the network. Both networks use the same cost objective, given as:

$$Cost = \sum_{i=1}^{N} U_c(D_i)L_i \tag{11}$$

329 where U_c is the unit cost of pipe i which is a function of diameter D_i; L_i is the length

330 of pipe i; N is the total number of pipes.

Sample of Solutions

332 Optimizations are carried out considering, as objective functions, the network cost and each

333 of the six reliability measures described in the methodology section. In particular, for each

334 pair of objective functions (i.e., one of the reliability surrogate measures and cost), a set of 10

335 optimization runs is carried out in order to explore the influence of randomly generated initial

336 populations. For the Hanoi network, a preliminary sensitivity analysis led to the choice of a

337 population of 200 individuals and of a total of 500 generations for the optimization sets with

338 RI, NRI, MRI, API and PHRI. Additional analyses determined that optimisation runs with

339 these NSGA-II parameters and using DSFE were not converging. Therefore, a population of

340 200 individuals and a total of 5,000 generations were allowed for the optimization set with

341 DSFE. For the Fossolo network, the parameters used for the optimization are the population

342 of 200 individuals and 5,000 generations per run, while 10,000 generations were allowed for

343 the runs using the DSFE measure.

344 The optimal solutions are generated using the Pareto dominance approach by aggregating

345 best solutions from all ten optimization runs. This was done to ensure that as optimal as

346 possible Pareto fronts are obtained and compared. The solution samples for Hanoi are

347 selected from the overlapping area of costs, as shown in Figure 3, in order to implement the

348 integral method and provide a fair comparison. The values of RI, NRI, API and PHRI range

349 from 0 to 1, while MRI and DSFE only have a lower bound (i.e., 0). The Pareto front of

350   DSFE is less smooth when compared to other fronts, as a consequence of the more

351   complicated search for solutions in this case. The Pareto optimal solutions for Fossolo are

352   shown in Figure 4(a), and the shared cost solutions in the range from 0.03 to 0.7 million

353   dollars are selected as samples to be evaluated. The Figure 4(b) represents the enlarged area

354   from Figure 4(a).

355   The comparative analysis of the surrogate measures is fair on the premise of the same

356   convergence of the optimization process. For both networks used in this paper, the near

357   optimal solutions were validated in comparison with the results obtained by Wang et al [2014]

358   where they used massive computational resources in order to achieve the best-know

359   approximation of the true Pareto front.

360   ## *Comparison of results: Case #1*

361   **Perturbation of demand**

362   The samples of the solutions obtained in the previous section are evaluated first using the

363   indicators of PFS, FNC and FFD under demand perturbation. In Figure 5, FNC_{mean} ,

364   FNC_{median} , FNC_{25th} and FNC_{75th} are calculated for each solution. For all surrogate

365   measures, the extent of FNC variation tends to be small and FNC_{75th} decreases with the

366   increase in cost in the testing of demand perturbation. The reason is that the system with

367   higher cost has consistently greater redundancy to cope with demand changes for all

368   surrogate measures. Obviously, the values of FNC (i.e., the number of the failed nodes) of RI,

369   NRI, MRI and API decline steeply for lower costs. As a result, when RI, NRI, MRI and API

370   measures are used to investigate reliability, increasing the capital investment at the lower

371 costs will have a significant effect on mitigating the affected (failure) area under demand

372 perturbation. It is interesting to observe that FNC_{mean} of DSFE remains at an almost

373 constant value of about nine failure nodes across all the costs in Figure 5(f). The increased

374 cost is not necessarily beneficial for reducing the impact of demand perturbation on the

375 failure extent captured by the DSFE measure. The vast majority of FNC_{median} values of

376 DSFE are greater than FNC_{mean}. More values are clustered toward the upper end of the range

377 or few extreme values at the lower end are causing this skewness. For the network, a few

378 cases of demand fluctuations have a less significant impact on the spatial extent of the

379 deficient nodes.

380 The indicator of FFD is a widely used measure, and in this assessment framework, FFD

381 represents the inadequate degree of water delivered at failed nodes. FFD statistics for all

382 solutions are shown in Figure 6. The overall trend is that FFD values decline with the cost

383 increase except for DSFE. Unlike the FNC, the trends of FFD_{25th} and FFD_{75th} are

384 relatively consistent. The FFD values (i.e., the fraction of failure degree) of RI, NRI, MRI

385 and API experience no significant change for the lower cost values, while the FNC values of

386 these solutions decrease dramatically. This implies that, among these solutions, the spatial

387 extent of failed nodes has not expanded, but the failure degree of these failed nodes becomes

388 more severe.   The failure degree of PHRI at failed nodes declines throughout all the costs.

389 The FFD statistics of DSFE are approximately invariable.

390 Reliability surrogate measures are analysed using the integral method (i.e., Equation 10)

391 based on two kinds of scale: 1) one-segment cost step for which the integral procedure is

392   calculated once throughout the cost values; and 2) five-segment cost step for which the

393   integral calculation will be conducted at each fifth of the cost interval. The computation at

394   two scales is based on Equation 10, and the difference is that the one fifth range of cost is

395   respectively calculated in the five-segment integral. The integral results under demand

396   uncertainty perturbation are shown in Figure 7. The one-segment integral is plotted as light

397   grey columns, and that of five-segment results are represented by a line and symbol plot.

398   Figure 7(a) shows that the PFS values of PHRI are the lowest in the low cost region, but they

399   grow with the increase in cost. The figure also shows that the PHRI solutions with the smaller

400   diameter redundancy have the smallest failure scenario count and the strongest resistance

401   capacity with respect to the perturbation caused by demand variation. Overall, the PHRI

402   solutions have the smallest failure scenario count in terms of the 1-segment integral. DSFE

403   performs poorly compared with other surrogate measures throughout the entire cost spectrum.

404   RI, NRI, MRI and API have the similar PFS values for the five-segment cost steps. In the

405   comparison for 1-segment cost step, the NRI is slightly better than RI, MRI and API (see

406   Figure 7(d).

407   It is worth noting that the FNC statistics are only considered based on the failed scenarios.

408   For PHRI solutions, the probability of occurrence of failed scenarios associated with demand

409   perturbation is lower (see Figure 7d), but there will be a greater extent of impact on system

410   when a failure occurs (see Figure 7e).

411   The values of FNC_{mean} are used to calculate the integral of FNC, whose results are shown as

412   Figure 7(b) and (e). The smallest value of 1-segment integral of FNC is derived from RI.

413 However, the RI measure shows the lowest integral values at the higher cost steps for

414 5-segement integral of FNC in comparison with NRI, MRI and API. NRI has the greatest

415 mean value of failure nodes among the power-related family of surrogate measures. The

416 measures of PHRI and DSFE show the larger extent of impact under demand perturbation.

417 The mean FFD related to demand uncertainty is used to calculate the integral of FFD, as

418 shown in Figure 7(c) and (f). PHRI has the smallest integral value of FFD for 1-segment. The

419 analysis of the 5-segment integral shows that the PHRI plays a dominant role for the lower

420 cost region in comparison to other surrogate measures. NRI is the second best for failure

421 degree at failed nodes. Again, DSFE does not perform well and has an increasing difference

422 from other indices towards the high cost.

423 **Perturbation of pipe failure**

424 Pipe failure uncertainty is considered in this paper by using the concept of network

425 availability [Cullinane et al., 1992], which is defined as the proportion of the time that the

426 system will satisfactorily fulfil its function. Only the FFD indicator involves the network

427 availability, since PFS and FNC indicate the actual number of failed scenarios and nodes,

428 respectively. All failure scenarios considered are based on the fully connected network

429 configurations.

430 The failure node count values for each solution are shown in Figure 8. There are similar,

431 relatively flat patterns of variation for FNC statistics amongst RI, NRI, MRI, API and PHRI.

432 The figure also shows that most of solutions have a relatively small number of failed nodes

433 and some critical nodes are significantly influenced by pipe failure. The FNC_{median} values

434 are lower than FNC_{mean} (Figure 8a-8f). This skewness is inverse in comparison with the

435 results of FNC in demand perturbation scenarios (Figure 5). It demonstrates that several

436 critical pipe failures could result in a great number of nodes affected (failed). It can be

437 observed that FNC_{75th} of PHRI declines slightly as the cost increases in Figure 8(e).

438 In Figure 9, the failure degree at failed nodes declines with the increase in cost for all

439 reliability surrogate measures. However, the failed node count does not decrease in the pipe

440 failure testing in Figure 8. The downward trend of FFD can be attributed to the cost increase,

441 either larger diameter pipe leading to a lower probability of pipe failure or conveying more

442 water to failed nodes through other pathways. Although it is difficult to identify which is the

443 dominant reason, this trend reflects better the practical situation. The declining trends of

444 FFD_{25th} and FFD_{75th} are consistent throughout the cost range for RI, NRI, MRI, API,

445 PHRI (see Figure 9a-9e). The redundancy of pipes increases consistently with increases in

446 cost. However, with reference to DSFE, some of pipes taken out of service could lead to the

447 severely decreased system performance (see Figure 9f).

448 The integral of PFS is shown in Figure 10(a) and (d). The trends of all surrogate measures

449 show a decline with the increase in cost, which illustrates that by increasing capital cost one

450 can reduce the impact of pipe failure on WDSs. The comparison of surrogate measures

451 demonstrates that API has the lowest value of 1-segment integral and the lowest failure

452 scenario count under pipe failure perturbation. RI has the second lower value for the PFS

453 1-segment integral. There is a similar situation with DSFE for the case of demand uncertainty.

454   The failure scenario count of DSFE shows the similar results with other measures at the

455   lowest cost step, but the DSFE has great differences at other cost steps.

456   FNC_{mean} is used as a representative value to calculate the integral of FNC, as shown in

457   Figure 10(b) and (e). PHRI has the best performance with respect to failure node count for

458   1-segment integral and outperforms other indices almost at all cost steps for 5-segment

459   integral. In the power family of surrogate measures, MRI obtains a better performance than

460   other power-related indices. Once again, DSFE has the greatest number of failed node count.

461   The integral of FFD_{mean} is shown in Figure 10 (c) and (f). The integral values of NRI and

462   DSFE for 1-segment are relatively small. However, they are not ideal options since DSFE

463   values vary significantly and the NRI measure outperforms other indices only at the highest

464   cost step in the 5-segment integral. By the comparison of Figure 7 and 10, DSFE seems to

465   perform better in resisting the perturbation of pipe failure than the disturbance of demand

466   changes due to taking into account the impact of diameters on reliability.

467   *Comparison of results: Case #2*

468   **Perturbation of demand**

469   The Pareto optimal solutions for the Fossolo network are derived via the previously described

470   optimization process. The assessment associated with demand and pipe failure perturbations

471   is performed post optimization based on the obtained optimal solutions. The assessment

472   results are represented by a set of indicator statistics. The indicator based on the failure node

473   count (FNC) is shown in Figure S1 in the supplementary material. For RI, NRI, MRI, API

474   and PHRI, the FNC values declines sharply at the lower cost, while the changes of the FNC

475 indicator become relatively flat at the high cost. Meanwhile, the 25th percentile and 75th

476 percentile and median merge together with the increase in cost. This indicates that the

477 solutions at the high cost end have only one or no failed scenarios in 1,000 demand

478 fluctuation evaluations, which can also been seen in Figure 11(a). It should be emphasized

479 that the indicators can achieve zero at the highest cost when using RI, NRI, MRI and API

480 surrogate measures, which shows these networks exhibit high reliability with respect to

481 demand fluctuations.

482 The DSFE surrogate measure has the worst performance with respect to the FNC indicator. In

483 the case of increasing cost, the networks derived using DSFE still have a large number of

484 failed nodes. However, the range of variation of FNC is small for DSFE. For the NRI

485 measure, the networks with the higher cost exhibit a wide range of changes of failed nodes. It

486 is worth noting that the Latin Hypercube sampling method uses the same random seed for all

487 surrogate measures, but the NRI solutions at higher costs are not as good as other surrogate

488 measures with respect to the demand perturbation. In other words, the solutions obtained by

489 using NRI show a larger spatial extent of failures under the demand perturbation condition.

490 The indicator of fraction of failure degree (FFD) is calculated and the results of its statistics

491 are shown in Figure S2 in the supplementary material. All the surrogate measures except

492 DSFE exhibit a reduction in failure degree with the increase in cost. Furthermore, at the

493 lower costs, the FFD decreases sharply and then became relatively flat at the higher costs. RI,

494 MRI, API and PHRI have very narrow ranges of variation for the demand perturbation. In

495 contrast, the NRI has a wider range of variation, which demonstrates that although network

496   configurations derived from NRI can resist a proportion of the demand perturbation very well,

497   they are less effective for other perturbations. This is likely due to NRI being based on a

498   uniformity coefficient for pipe diameters, which means that diameters in the network tend to

499   increase evenly from the source to the extremities. If the increased demands appear near the

500   source, the network can easily handle this perturbation. However, if the increased demands

501   are applied at the nodes far away from the source, it is most likely to result in a failed

502   scenarios. Different from source-bias diameter layout of the NRI, it is common engineering

503   practice that larger diameters are associated with trunk mains, and the relatively small

504   diameters are deployed as offshoots from the trunks.

505   The integral method is used to compare the surrogate measures. The integral results of the

506   Fossolo network is given in Figure 11. The DSFE surrogate has the largest values for all three

507   indicators. It should, however, be noted that DSFE solutions with smaller costs (basically in

508   the first cost segment) are competitive with other surrogate measures, but when the cost

509   increases, it becomes inferior.

510   NRI is the second worse measure when demand perturbations are applied to the Fossolo

511   network. This is probably because NRI tends to favour diameter redundancy (i.e., larger

512   diameters) near the source of the network. However, the demand perturbations are applied

513   uniformly across the network. As a results, NRI's bias for diameter redundancy fails to

514   perform well for this kind of global demand perturbations.

515   The surrogate measures of RI, MRI and API show similar behaviour to each other under

516   demand perturbations. The newly developed PHRI outperforms other measures with respect

517   to other indicators in the number of failed scenarios. It could be seen in Figure 11(d) that the

518   tiny advantage is stemmed from the second cost interval in the five-segment integral. But

519   PHRI becomes slightly inferior for other two indicators.

520   **Perturbation of pipe failure**

521   Failure node count is calculated next with respect to the pipe failure perturbation, as shown in

522   Figure S3 in the supplementary material. The RI, MRI, and API measures show similar

523   behaviour in terms of FNC values such that with the increase in cost, the number of failed

524   nodes decreases generally. Here, the variation of FNC is different from its behaviour shown

525   in the demand perturbation graphs. However, the non-monotonic decreasing trend is the same

526   as in the pipe failure perturbation case in the Hanoi network. The pipe faliure perturbation

527   ususlly causes the node failure at the downstream of the incident pipe, and these failed nodes

528   tend to locate at the extremities. Hence, it indicates that the pipe failure often fails the critical

529   nodes in a certain extent.

530   The values of the FNC indicator exhibit significant changes at the lower cost values for PHRI

531   and NRI. And, at most higher costs, the FNC values of PHRI and NRI are eqaul to zero. This

532   indicates that there is no node that fails to meet the requirement in the network, no matter

533   which pipe is taken out of service. On the contrary, DSFE does not show an obvious trend

534   regarding the FNC indicator.

535   The fraction of failure degree indicator (FFD) is shown in Figure S4 in the supplementary

536   material. All surrogate measures, except DSFE, show a similar trend in which FFD decreases

537   with the increase in cost. Particularly, NRI and PHRI achieve a good performance for the

538   pipe failure perturbation. However, these two surrogate measures are worse than RI, MRI and

539   API in the assessment of demand perturbation in the Fossolo network. This demonstrates that

540   the network designs derived from MRI and PHRI measures are good at coping with local

541   perturbations (i.e., the pipe failure), rather than global perturbations (i.e., demand fluctuation)

542   for this particular network. The range of variation of DSFE is relatively small at the lower

543   costs, but its statistics takes a much wider range at the higher cost end.

544   The integral values are calculated by Equation 10, and the results are shown in Figure 12.

545   With respect to the pipe failure perturbation in the Fossolo network, NRI achieves the best

546   performance for all three indicators, as it obviously has a much lower value than other

547   surrogate measures. PHRI is the second best in the assessment, while RI, MRI and API have

548   similar values in both one segment and five-segment integrals.

549   DSFE has the worst results overall by using the one segment integral analysis, but it can be

550   seen that it facilitates searching for reliable solutions, especially in the lower cost range.

551   Another point which is worth noting is that DSFE's solutions have a better performance in

552   handling the pipe failure perturbation than that for the demand uncertainty perturbation. This

553   can be reflected in both case studies. With a sufficient budget, DSFE tends to deploy multiple,

554   and long, trunk mains in order to achieve uniformly distributed flows based on the flow

555   entropy measure. The pipe flows take a long route to the users, and thus more energy is

556 dissipated, which leads to a more severe impact once the demand fluctuation and pipe

557 isolation occur. Besides, it may be that the higher cost solutions derived by using DSFE are

558 sub-optimal for this network.

Conclusions

560 The paper presented two new reliability surrogate measures (i.e., available power index and

561 pipe hydraulic resilience index) and compared these to the four existing measures from the

562 literature (i.e., resilience index, network resilience index, modified resilience index and

563 diameter-sensitive flow entropy). The comparison was done under the conditions of demand

564 and pipe failure uncertainties on the problem of optimal WDS design. The demand

565 uncertainty, characterised using normal distribution, is implemented by a stochastic sampling

566 procedure which simulates different demand spatial fluctuations. The pipe failure uncertainty

567 is described by the existing probability of a pipe failure model. The performance of obtained

568 optimal WDS designs, derived by minimising total cost of interventions and maximising

569 system reliability of supply (achieved by using above reliability surrogate measures, one at a

570 time), is examined by using three indicators: (i) percentage of failure scenarios, which

571 represents the overall performance of a system under uncertain conditions; (ii) failure node

572 count, which indicates the spatial extent of the level of service reduction in failed scenarios;

573 and (iii) the fraction of failure degree, which denotes the magnitude of service deficiency at

574 failed nodes.

575   The above methodology was applied to the case studies of Hanoi and Fossolo WDSs. The

576   results obtained show that the pipe hydraulic resilience index and available power index

577   proposed performed well under demand and pipe failure perturbations. Opposite of this, the

578   diameter-sensitive flow entropy performed relatively poorly. It is worth noting that the

579   low-cost DSFE solutions performed much better than that at the larger cost. The NRI

580   surrogate measure has a better performance in the pipe failure perturbation than that for the

581   demand fluctuation perturbation. Generally speaking, it seems that the power family of

582   reliability surrogates, including RI, MRI and API, achieved similar results despite small

583   differences in implementation.

584   Further work would be necessary to investigate the reliability surrogate measures in a wide

585   range of large-scale real-world WDSs. Note that not only should surrogate measures assess

586   reliability accurately, but also it is a concern that, in the case of large-scale real-world WDSs,

587   the measure can be implemented in the search of optimal solutions with high computational

588   efficiency.

589 Acknowledgements

References

Atkinson, S., R. Farmani, F. Memon and D. Butler (2014). Reliability Indicators for Water Distribution System Design: Comparison, *Journal of Water Resources Planning and Management*, 140(2), 160-168, doi:10.1061/(ASCE)WR.1943-5452.0000304.

Awumah, K., I. Goulter and S. Bhatt (1990). Assessment of reliability in water distribution networks using entropy based measures, *Stochastic Hydrology and Hydraulics*, 4(4), 309-320, doi:10.1007/bf01544084.

Baños, R., J. Reca, J. Martínez, C. Gil and A. Márquez (2011). Resilience Indexes for Water Distribution Network Design: A Performance Analysis Under Demand Uncertainty, *Water Resources Management*, 25(10), 2351-2366, doi: 10.1007/s11269-011-9812-3.

Bao, Y. and L. Mays (1990). Model for water distribution system reliability, *Journal of Hydraulic Engineering*, 116(9), 1119-1137, doi:10.1061/(ASCE)0733-9429(1990)116:9(1119).

Basupi, I., Z. Kapelan and D. Butler (2013). Reducing life-cycle carbon footprints in the redesign of water distribution systems, *Journal of Water and Climate Change*, 4(3), 176-192, doi:10.2166/wcc.2013.004.

Creaco, E. and M. Franchini (2012). Fast network multi-objective design algorithm combined with an a posteriori procedure for reliability evaluation under various operational scenarios, *Urban Water Journal*, 9(6), 385-399, doi: 10.1080/1573062x.2012.690432.

Creaco, E., M. Franchini and E. Todini (2014). The combined use of resilience and loop diameter uniformity as a good indirect measure of network reliability, *Urban Water Journal*, 1-15, doi: 10.1080/1573062X.2014.949799.

615   Cullinane, M., K. Lansey and L. Mays (1992). Optimization-availability-based design of

616   water-distribution networks, *Journal of Hydraulic Engineering*, 118(3), 420-441,

617   doi:10.1061/(ASCE)0733-9429(1992)118:3(420).

618   Dandy, G. and M. Engelhardt (2006). Multi-Objective Trade-Offs between Cost and Reliability in the

619   Replacement of Water Mains, *Journal of Water Resources Planning and Management*, 132(2), 79-88,

620   doi:10.1061/(ASCE)0733-9496(2006)132:2(79).

621   Deb, K., A. Pratap, S. Agarwal and T. Meyarivan (2002). A fast and elitist multiobjective genetic

622   algorithm: NSGA-II, *Evolutionary Computation, IEEE Transactions on*, 6(2), 182-197, doi:

623   10.1109/4235.996017.

624   Farmani, R., G. Walters and D. Savic (2005). Trade-off between total cost and reliability for Anytown

625   water distribution network, *Journal of Water Resources Planning and Management*, 131(3), 161-171,

626   doi:10.1061/(ASCE)0733-9496(2005)131:3(161).

627   Filion, Y., B. Adams and B. Karney (2007). Stochastic Design of Water Distribution Systems with

628   Expected Annual Damages, *Journal of Water Resources Planning and Management*, 133(3), 244-252,

629   doi:10.1061/(ASCE)0733-9496(2007)133:3(244).

630   Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses,

631   *Global Environmental Change*, 16(3), 253-267, doi: 10.1016/j.gloenvcha.2006.04.002.

632   Fu, G. and Z. Kapelan (2011). Fuzzy probabilistic design of water distribution networks, *Water*

633   *Resources Research*, 47(5), W05538, doi: 10.1029/2010WR009739.

634   Fujiwara, O. and D. B. Khang (1990). A two-phase decomposition method for optimal design of

635   looped water distribution networks, *Water Resources Research*, 26(4), 539-549,

636   doi:10.1029/WR026i004p00539.

637   Fujiwara, O. and H. D. Tung (1991). Reliability improvement for water distribution networks through

638   increasing pipe size, *Water Resources Research*, 27(7), 1395-1402, doi:10.1029/91wr00882.

639   Gheisi, A. and G. Naser (2015). Multistate Reliability of Water-Distribution Systems: Comparison of

640   Surrogate Measures, *Journal of Water Resources Planning and Management*, 141(10), 04015018,

641   10.1061/(ASCE)WR.1943-5452.0000529.

642   Giustolisi, O., D. Laucelli and A. Colombo (2009). Deterministic versus Stochastic Design of Water

643   Distribution Networks, *Journal of Water Resources Planning and Management*, 135(2), 117-127,

644   doi:10.1061/(ASCE)0733-9496(2009)135:2(117).

645   Housh, M., A. Ostfeld and U. Shamir (2011). Optimal multiyear management of a water supply

646   system under uncertainty: Robust counterpart approach, *Water Resour. Res.*, 47(10), W10515, doi:

647   10.1029/2011wr010596.

648   Jayaram, N. and K. Srinivasan (2008). Performance-based optimal design and rehabilitation of water

649   distribution networks using life cycle costing, *Water Resources Research*, 44(1), W01417, doi:

650   10.1029/2006WR005316.

651   Kang, D. S., M. F. K. Pasha and K. Lansey (2009). Approximate methods for uncertainty analysis of

652   water distribution systems, *Urban Water Journal*, 6(3), 233-249, 10.1080/15730620802566844.

653   Kapelan, Z. S., D. A. Savic and G. A. Walters (2005). Multiobjective design of water distribution

654   systems under uncertainty, *Water Resources Research*, 41(11), W11407,

655   doi:10.1029/2004WR003787.

656   Liu, H., D. Savić, Z. Kapelan, M. Zhao, Y. Yuan and H. Zhao (2014). A diameter-sensitive flow

657   entropy method for reliability consideration in water distribution system design, *Water Resources*

658   *Research*, 50(7), 5597-5610, doi: 10.1002/2013WR014882.

659   Maier, H. R., B. J. Lence, B. A. Tolson and R. O. Foschi (2001). First-order reliability method for

660   estimating reliability, vulnerability, and resilience, *Water Resources Research*, 37(3), 779-790, doi:

661   10.1029/2000WR900329.

662   Marchi, A., E. Salomons, A. Ostfeld, Z. Kapelan, A. Simpson, A. Zecchin, H. Maier, Z. Wu, S.

663   Elsayed, Y. Song, T. Walski, C. Stokes, W. Wu, G. Dandy, S. Alvisi, E. Creaco, M. Franchini, J.

664   Saldarriaga, D. Páez, D. Hernández, J. Bohórquez, R. Bent, C. Coffrin, D. Judi, T. McPherson, P. van

665   Hentenryck, J. Matos, A. Monteiro, N. Matias, D. Yoo, H. Lee, J. Kim, P. Iglesias-Rey, F.

666   Martínez-Solano, D. Mora-Meliá, J. Ribelles-Aguilar, M. Guidolin, G. Fu, P. Reed, Q. Wang, H. Liu,

667   K. McClymont, M. Johns, E. Keedwell, V. Kandiah, M. Jasper, K. Drake, E. Shafiee, M. Barandouzi,

668   A. Berglund, D. Brill, G. Mahinthakumar, R. Ranjithan, E. Zechman, M. Morley, C. Tricarico, G. de

669   Marinis, B. Tolson, A. Khedr and M. Asadzadeh (2014). Battle of the Water Networks II, *Journal of*

670   *Water Resources Planning and Management*, 140(7), 04014009,

671   doi:10.1061/(ASCE)WR.1943-5452.0000378.

672   Martinez, J. (2010). Cost and reliability comparison between branched and looped water supply

673   networks, *Journal of Hydroinformatics*, 12(2), 150-160.

674   Morley, M. S. and C. Tricarico (2008). Pressure driven demand extension for EPANET

675   (EPANETpdd). Technical Report 2008-02. Exeter, UK, Centre for Water System, University of

676   Exeter.

677   Ostfeld, A. (2004). Reliability analysis of water distribution systems, *Journal of Hydroinformatics*,

678   6(2004), 281-294.

679   Ostfeld, A. (2012). Optimal reliable design and operation of water distribution systems through

680   decomposition, *Water Resources Research*, 48(10), W10521, doi:10.1029/2011wr011651.

681   Park, J. I., J. H. Lambert and Y. Y. Haimes (1998). Hydraulic power capacity of water distribution

682   networks in uncertain conditions of deterioration, *Water Resources Research*, 34(12), 3605-3614, doi:

683   10.1029/98wr01377.

684   Piratla, K. R. and S. T. Ariaratnam (2012). Reliability based optimal design of water distribution

685   networks considering life cycle components, *Urban Water Journal*, 9(5), 305-316,

686   10.1080/1573062x.2012.660961.

687   Prasad, T. and N. Park (2004). Multiobjective genetic algorithms for design of water distribution

688   networks, *Journal of Water Resources Planning and Management*, 130(1), 73-82,

689   doi:10.1061/(ASCE)0733-9496(2004)130:1(73).

690   Raad, D. N., A. N. Sinske and J. H. van Vuuren (2010). Comparison of four reliability surrogate

691   measures for water distribution systems design, *Water Resour. Res.*, 46(5), W05524, doi:

692   10.1029/2009wr007785.

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

694   Savic, D. A. and G. A. Walters (1997). Genetic algorithms for least-cost design of water distribution

695   networks, *Journal of Water Resources Planning and Management*, 123(2), 67-77, doi:

696   10.1061/(ASCE)0733-9496(1997)123:2(67).

697   Shafiqul Islam, M., R. Sadiq, M. Rodriguez, H. Najjaran and M. Hoorfar (2014). Reliability

698   Assessment for Water Supply Systems under Uncertainties, *Journal of Water Resources Planning and*

699   *Management*, 140(4), 468-479, doi:10.1061/(ASCE)WR.1943-5452.0000349.

700   Sumer, D. and K. Lansey (2009). Effect of uncertainty on water distribution system model design

701   decisions, *Journal of Water Resources Planning and Management*, 135(1), 38-47, doi:

702   10.1061/(ASCE)0733-9496(2009)135:1(38).

703   Sun, S., S. T. Khu, Z. Kapelan and S. DjordjeviC (2011). A fast approach for multiobjective design of

704   water distribution networks under demand uncertainty, *Journal of Hydroinformatics*, 13(2), 143-152.

705   Tanyimboh, T. T. and A. B. Templeman (2000). A quantified assessment of the relationship

706   between the reliability and entropy of water ditribution systems, *Engineering Optimization*, 33(2),

707   179-199, doi:10.1080/03052150008940916.

708   Todini, E. (2000). Looped water distribution networks design using a resilience index based heuristic

709   approach, *Urban Water*, 2(2), 115-122, doi: 10.1016/S1462-0758(00)00049-2.

710   Torii, A. and R. Lopez (2012). Reliability Analysis of Water Distribution Networks Using the

711   Adaptive Response Surface Approach, *Journal of hydraulic engineering*, 138(3), 227-236,

712   doi:10.1061/(ASCE)HY.1943-7900.0000504.

713   van Zyl, J., O. Piller and Y. le Gat (2008). Sizing Municipal Storage Tanks Based on Reliability

714   Criteria, *Journal of Water Resources Planning and Management*, 134(6), 548-555,

715   doi:10.1061/(ASCE)0733-9496(2008)134:6(548).

716   Walski, T. (2001). The Wrong Paradigm—Why Water Distribution Optimization Doesn't Work,

717   *Journal of Water Resources Planning and Management*, 127(4), 203-205,

718   doi:10.1061/(ASCE)0733-9496(2001)127:4(203).

719   Walski, T. M. (1993). Water distribution valve topology for reliability analysis, *Reliability*

720   *Engineering & System Safety*, 42(1), 21-27, doi:org/10.1016/0951-8320(93)90051-Y.

721   Wang, Q., M. Guidolin, D. Savic and Z. Kapelan (2014). Two-Objective Design of Benchmark

722   Problems of a Water Distribution System via MOEAs: Towards the Best-Known Approximation of

723   the True Pareto Front, *Journal of Water Resources Planning and Management*, 141(3), 04014060,

724   10.1061/(ASCE)WR.1943-5452.0000460.

725   Xu, C. and I. C. Goulter (1999). Reliability-based optimal design of water distribution networks,

726   *Journal of Water Resources Planning and Management*, 125(6), 352-362,

727   doi:10.1061/(ASCE)0733-9496(1999)125:6(352)

728   Yannopoulos, S. and M. Spiliotis (2013). Water distribution system reliability based on minimum

729   cut–set approach and the hydraulic availability, *Water Resources Management*, 27(6), 1821-1836,

730   doi:10.1007/s11269-012-0163-5.

731   Zhuang, B., K. Lansey and D. Kang (2013). Resilience/Availability Analysis of Municipal Water

732   Distribution System Incorporating Adaptive Pump Operation, *Journal of hydraulic engineering*,

733   139(5), 527-537, doi:10.1061/(ASCE)HY.1943-7900.0000676.

734

Table 1. List of Abbreviations

API	Available power index
DSFE	Diameter-sensitive flow entropy
FFD	Fraction of failure degree
FNC	Failure node count
RI	Resilience index
HGL	Hydraulic grade line
LHS	Latin hypercube sampling
MRI	Modified resilience index
NRI	Network resilience index
PDD	Pressure driven demand
PDF	Probability density function
PFS	Percentage of failure scenario
PHRI	Pipe hydraulic resilience index
WDS	Water distribution system

a)

Demand

320.00
580.00
840.00
1100.00
CMH

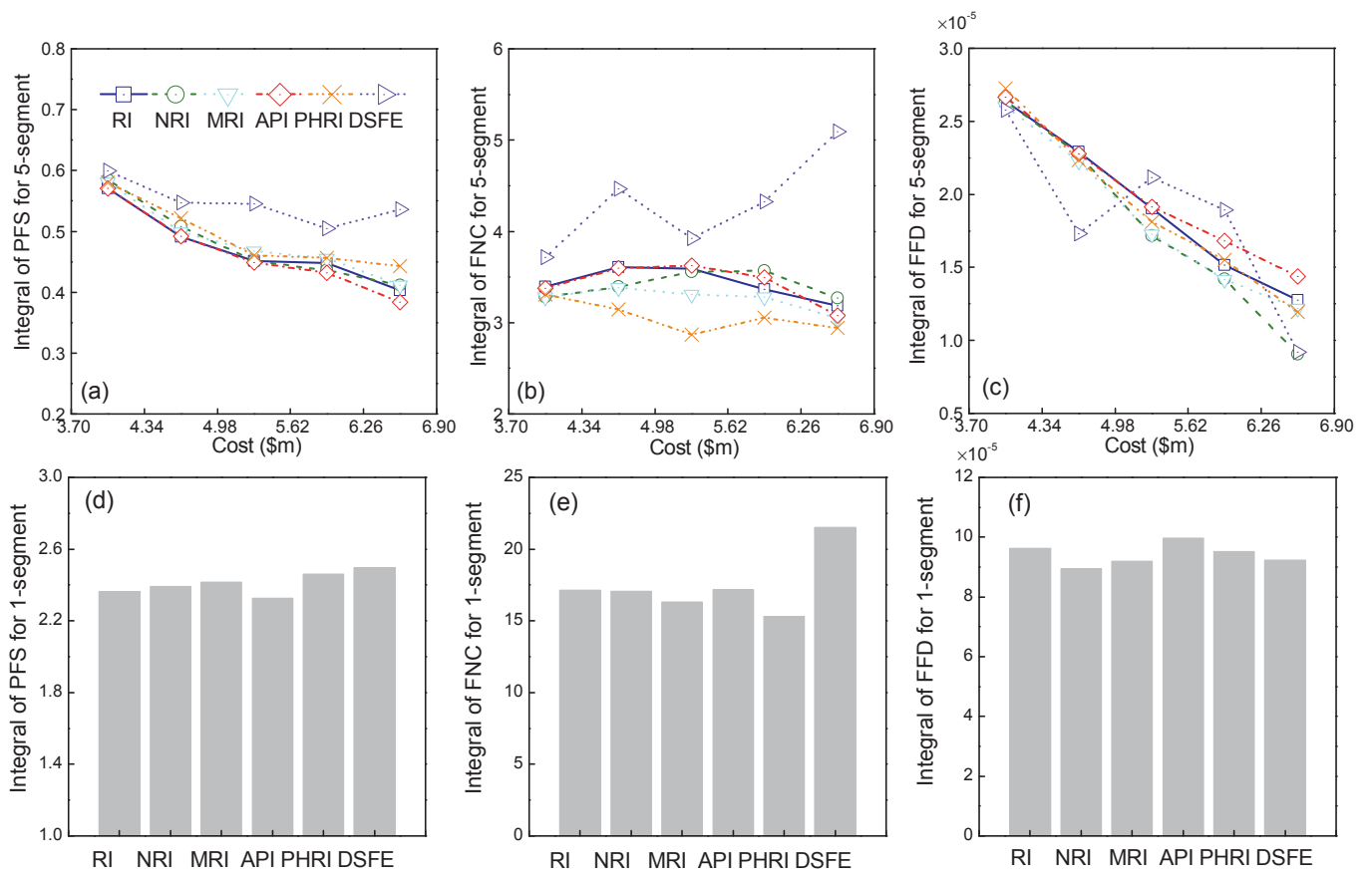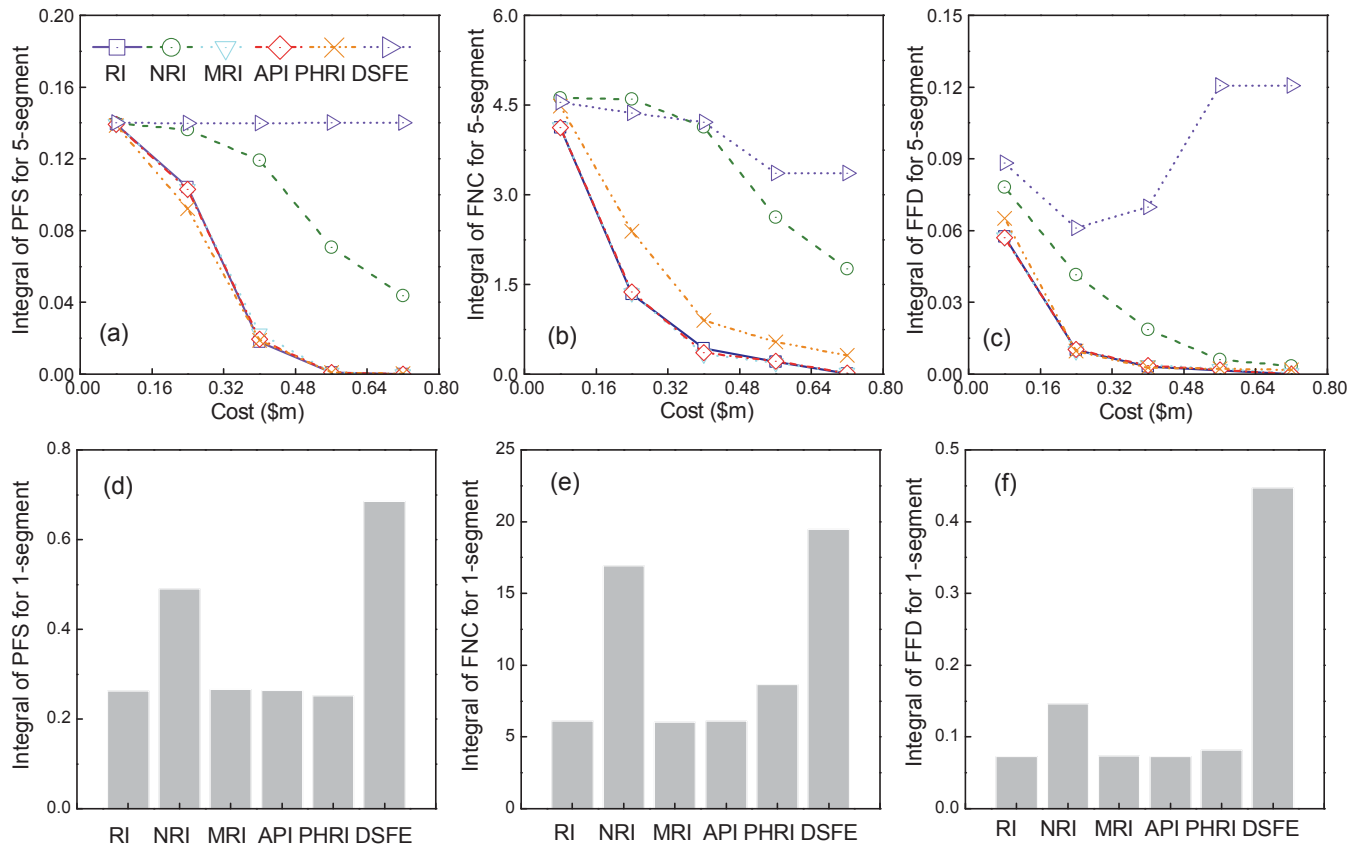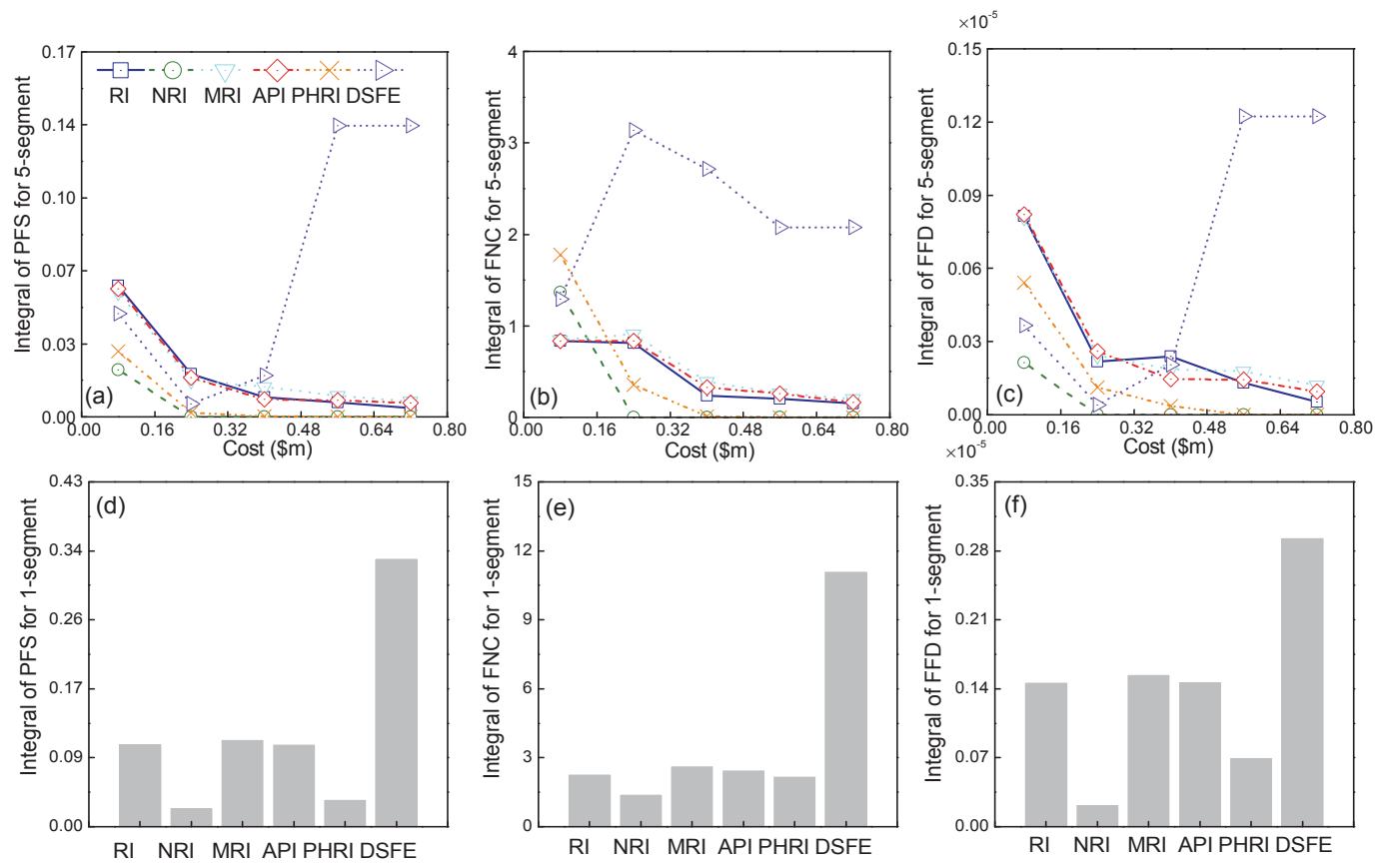b)

Demand

0.60
1.20
1.80
2.40
LPS
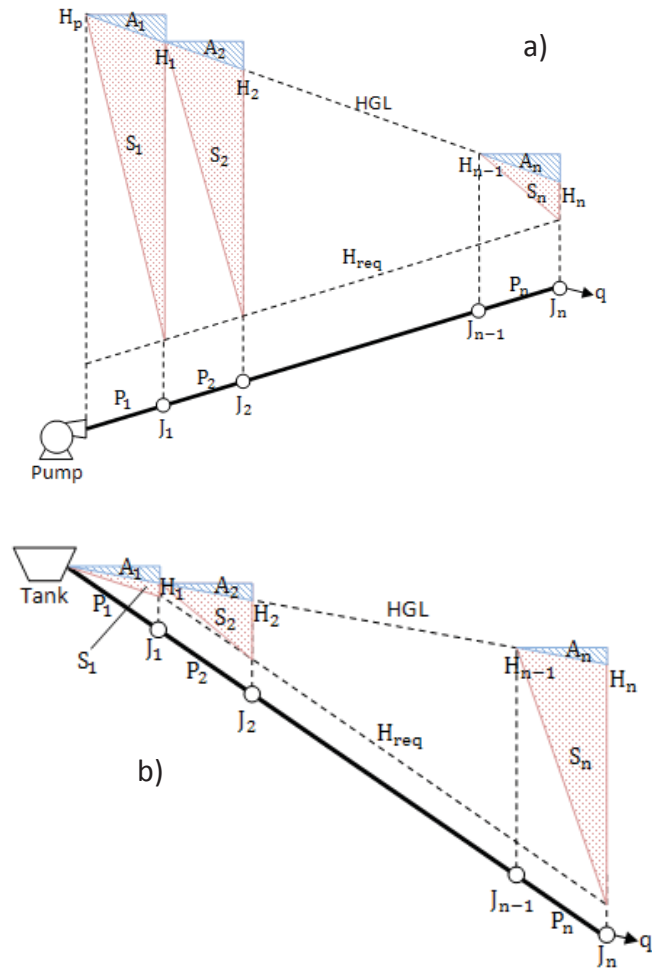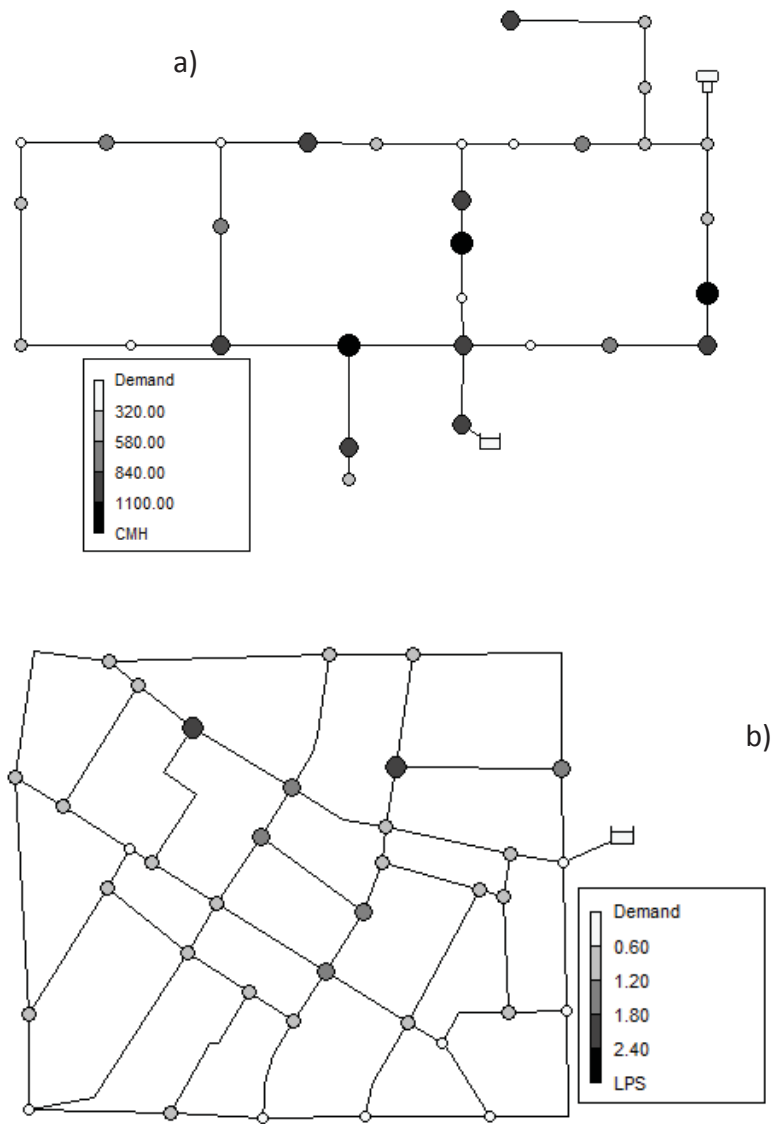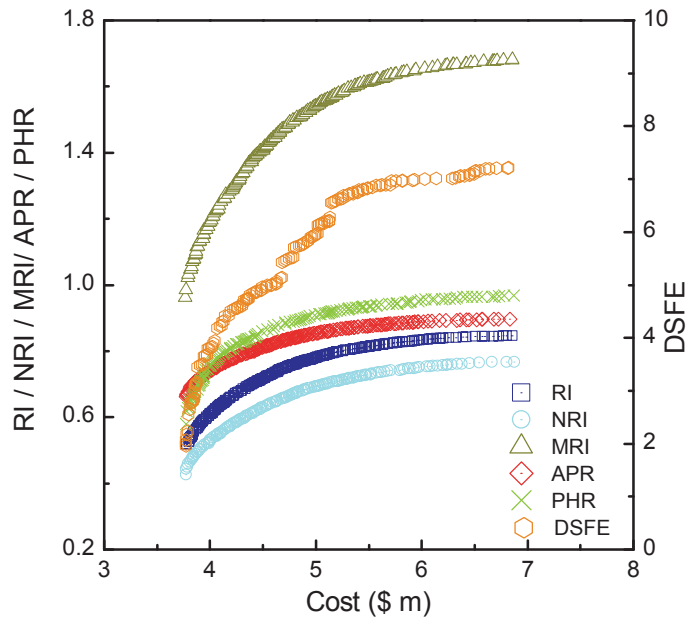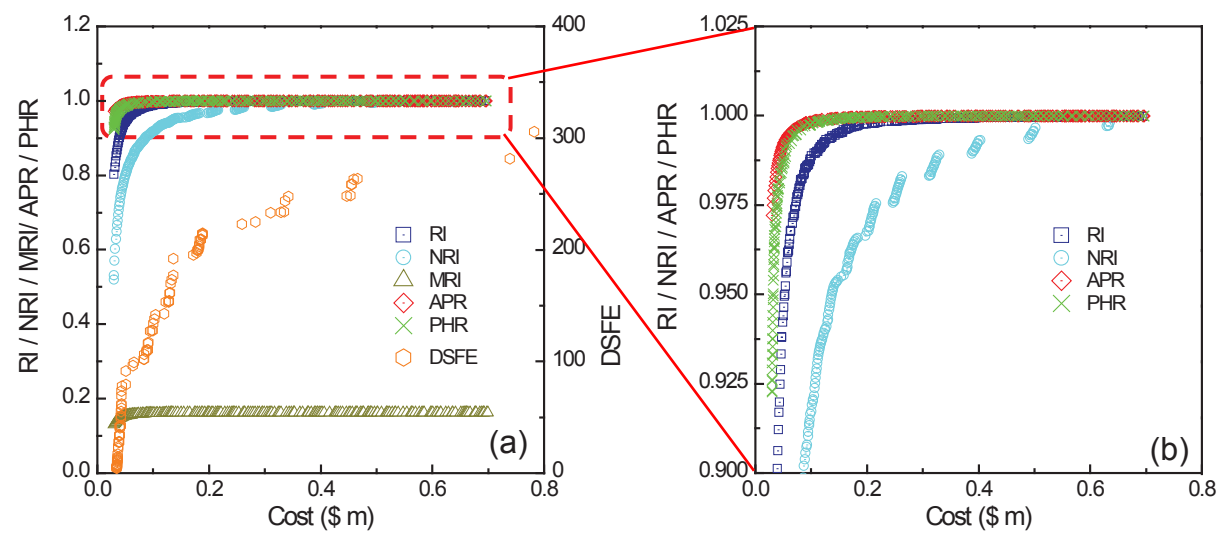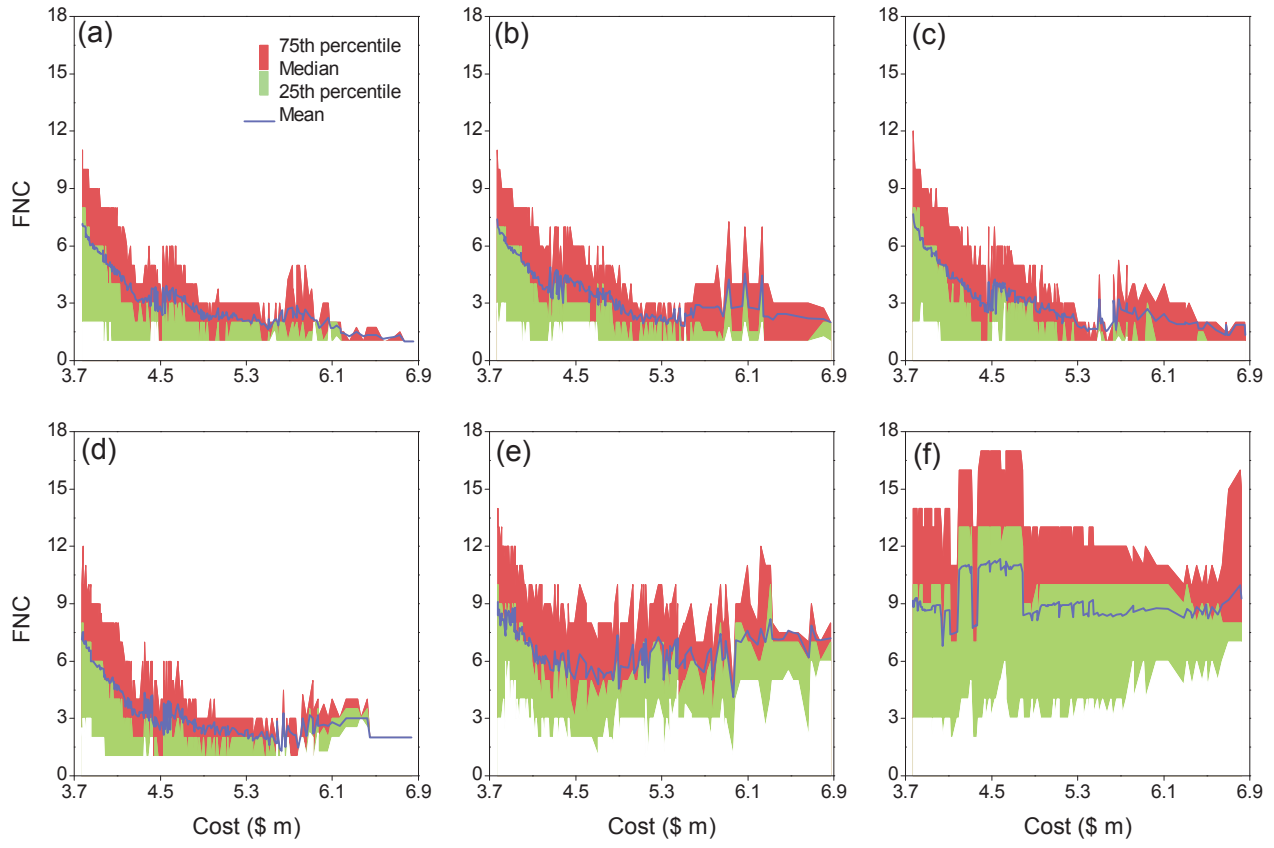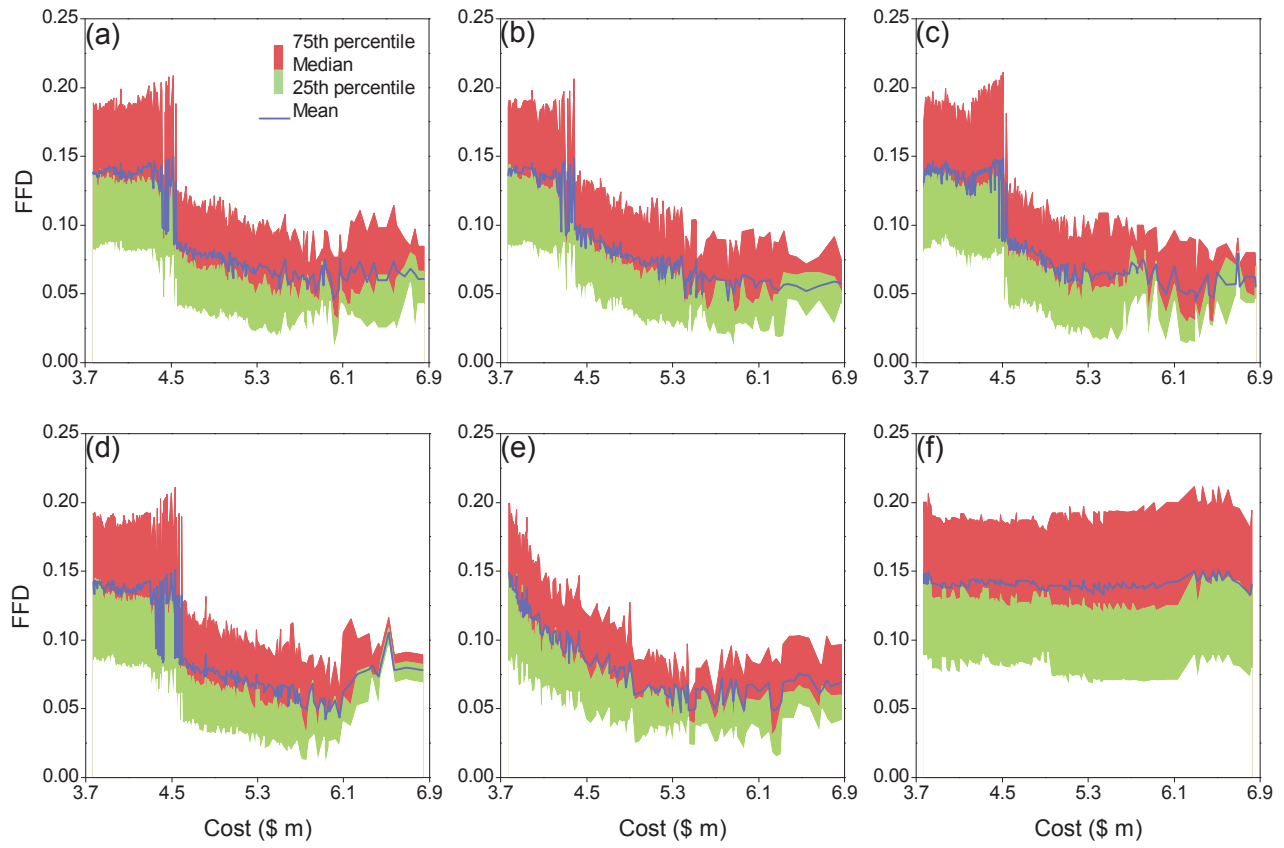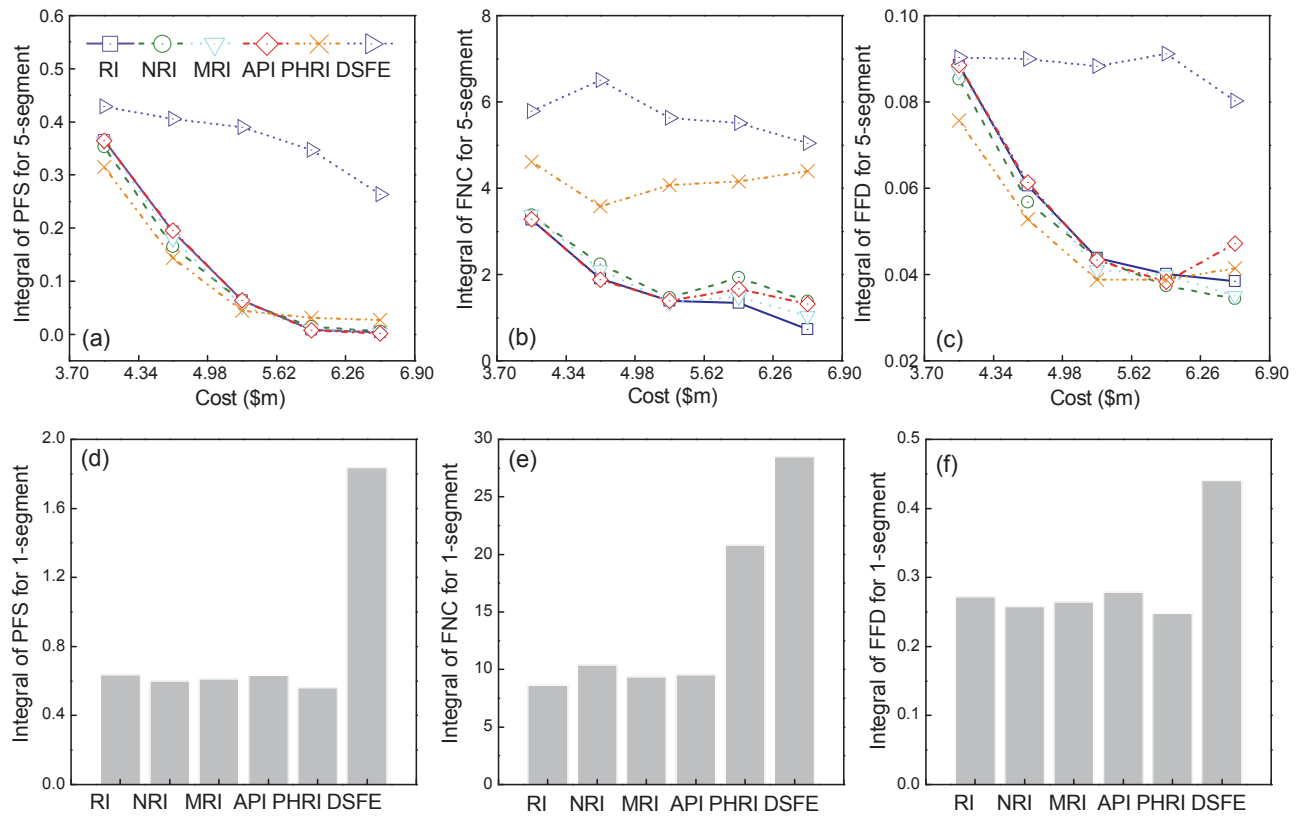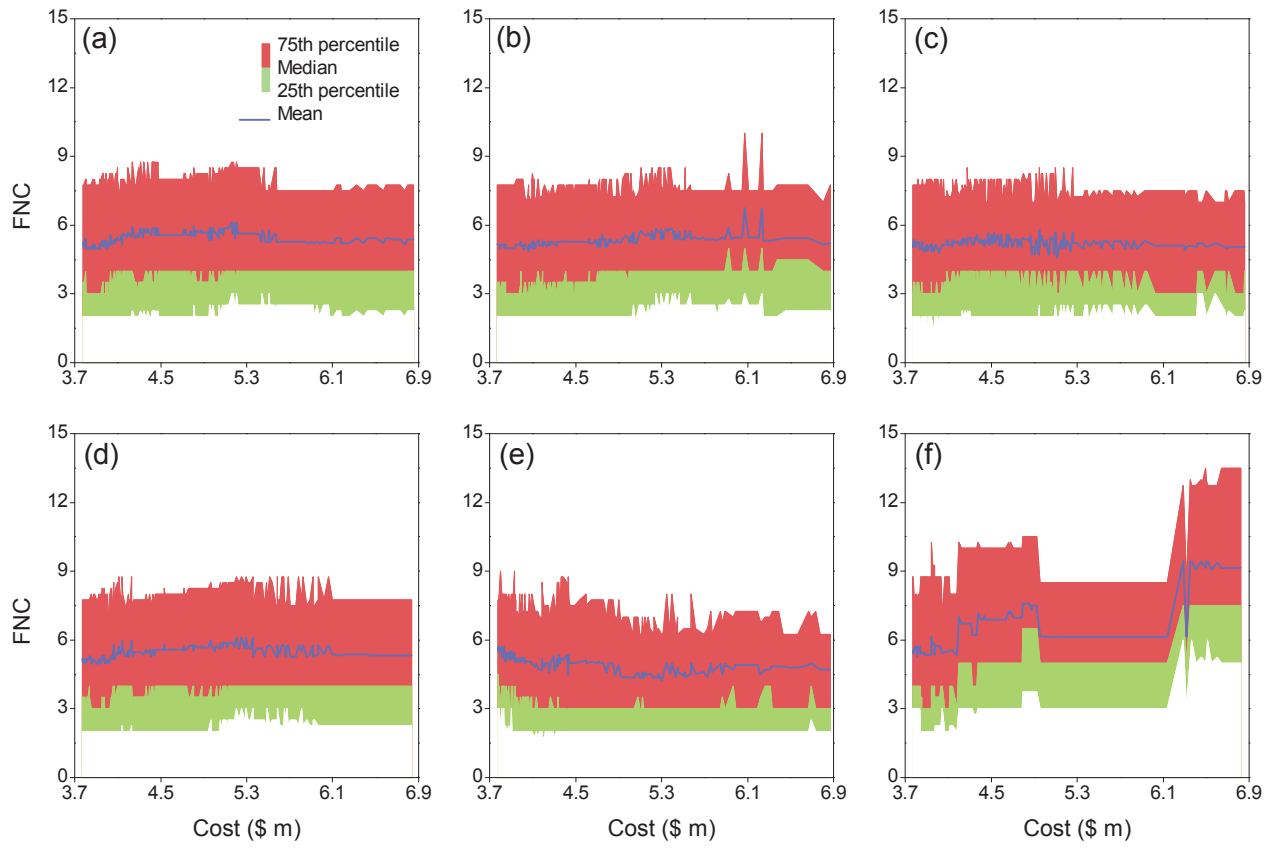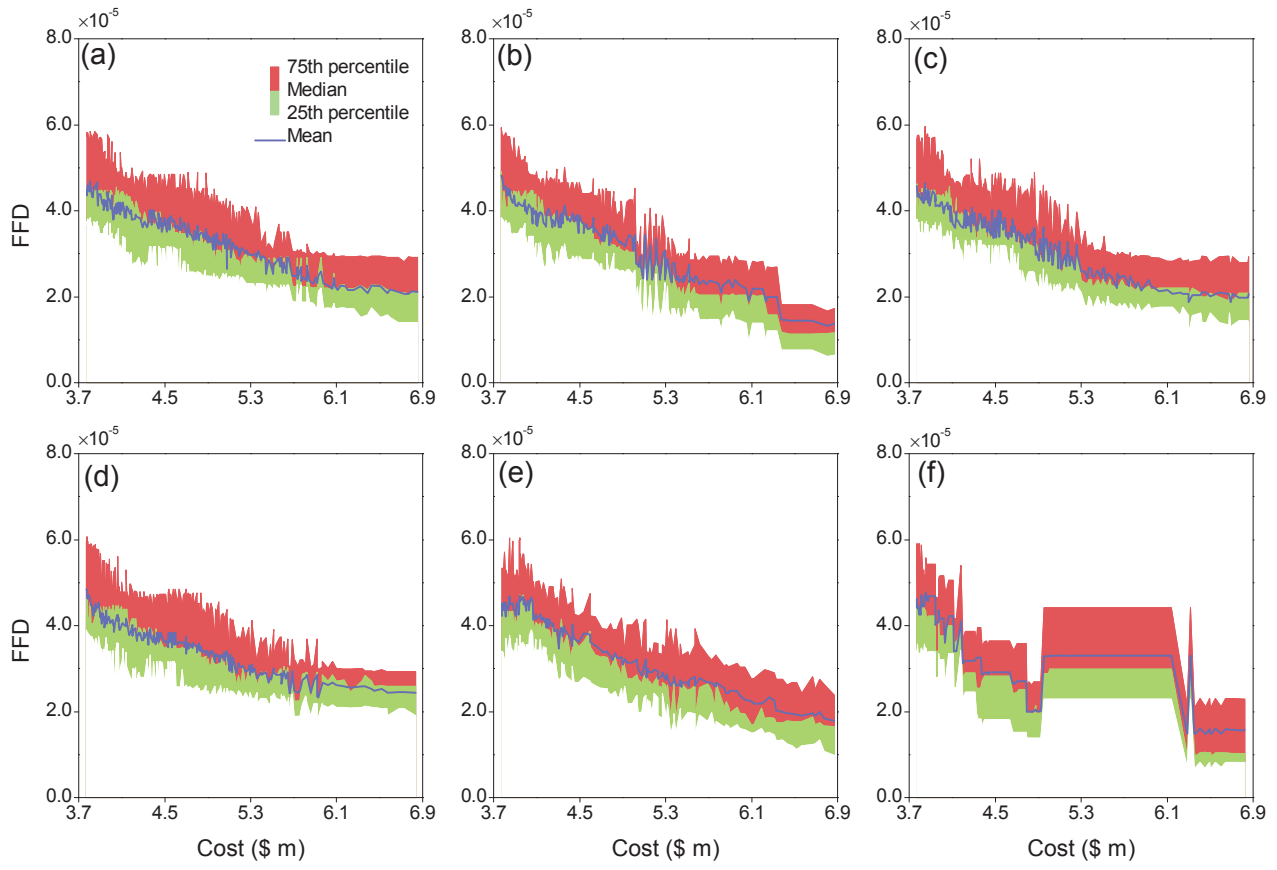
Figure captions

Figure 1. A diagram of water distribution systems with HGL (a) a pressurized supply system, (b) a gravity supply system. The two systems are assumed with only a fixed demand q at node J_n

Figure 2. The schematic diagram with nodal demand represented by the bubble size: (a) Hanoi network; (b) Fossolo network

Figure 3. Pareto fronts obtained using each reliability surrogate measure and cost objectives for the Hanoi network

Figure 4. Pareto fronts obtained using each reliability surrogate measure and cost objectives for the Fossolo network: (a) Pareto fronts derived by six surrogate reliability measures; (b) enlarged part of Figure 4a

Figure 5. Failure node count under perturbation of demand uncertainty in the Hanoi network: (a) RI, (b) NRI, (c) MRI, (d) API, (e) PHRI, and (f) DSFE

Figure 6. Fraction of failure degree under perturbation of demand uncertainty in the Hanoi network: (a) RI, (b) NRI, (c) MRI, (d) API, (e) PHRI, and (f) DSFE

Figure 7. Integral of evaluation indicators under perturbation of demand uncertainty in the Hanoi network. PFS is the percentage of failure scenarios with both integral approaches (Figures 7a and 7d); FNC is the failure node count with both integral approaches (Figures 7b and 7e); and FFD is the fraction of failure degree with both integral approaches (Figures 7c and 7f).

Figure 8. Failure node count under perturbation of pipe failure in the Hanoi network: (a) RI, (b) NRI, (c) MRI, (d) API, (e) PHRI, and (f) DSFE

Figure 9. Fraction of failure degree under disturbance of pipe failure uncertainty in the Hanoi network: (a) RI, (b) NRI, (c) MRI, (d) API, (e) PHRI, and (f) DSFE

Figure 10. Integral of evaluation indicators under perturbation of pipe failure uncertainty in the Hanoi network. PFS is the percentage of failure scenarios with both integral approaches (Figures 10a and 10d); FNC is the failure node count with both integral approaches (Figures 10b and 10e); and FFD is the fraction of failure degree with both integral approaches(Figures 10c and 10f)

Figure 11. Integral of evaluation indicators under perturbation of demand uncertainty in the Fossolo network. PFS is the percentage of failure scenarios with both integral approaches (Figures 11a and 11d); FNC is the failure node count with both integral approaches (Figures 11b and 11e); and FFD is the fraction of failure degree with both integral approaches (Figures 11c and 11f)

Fiugre 12. Integral of evaluation indicators under perturbation of pipe failure uncertainty in the Fossolo network. PFS is the percentage of failure scenarios with both integral approaches (Figures 12a and 12d); FNC is the failure node count with both integral approaches (Figures 12b and 12e); and FFD is the fraction of failure degree with both integral approaches (Figures 12c and 12f)