Uncertainty analysis of water supply networks using the fuzzy set theory and NSGA-II

Ali Haghighi *, Arezoo Zahdei Asl

Civil Engineering Department, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

ABSTRACT

This work introduces an approach for taking into account the uncertainty of pipe friction coefficients and nodal demands in the hydraulic analysis of water supply networks. For this purpose, uncertainties are represented by fuzzy numbers and incorporated into the network's governing equations. Input uncertainties are spread out on the network and influence its hydraulic responses, including pipe velocities and nodal pressures. To estimate the responses' uncertainty, input fuzzy numbers are discretized in some levels of membership function. Then, a multiobjective optimization problem is developed for each level to find the extreme values of the node pressures and pipe velocities. The raised problem is solved using the method of Non Dominated Sorting Genetic Algorithm (NSGA-II) coupled to the network hydraulic simulation model. The proposed approach is applied to an example and a real pipe network. It is found that small uncertainties in input variables can significantly influence the network's responses as well as its performance reliability. It is also concluded that NSGA-II has a great role in solving the problem systematically, and improves the computational efficiency of the whole process of network fuzzy analysis.

1. Introduction

For designing a new water distribution network or for condition assessment; calibration and extension of an existing one, it is required to simulate the network hydraulics. In this regard, nodal pressures and pipe velocities are the responses of interest, which may be calculated against several design scenarios or calibration variables. In fact, any decision on the network is taken based on the aforementioned responses. Hence, the reliability of the results obtained by the hydraulic simulation models is quite crucial to every decision making on the network development and rehabilitation. The results' reliability highly depends on accuracy of the mathematical model in one side, and the precision of input variables in another side. The latter is always a major concern with mathematical simulations and can significantly affect the credibility of the results. Some physical parameters used in the hydraulic modeling of water networks cannot be precisely measured and need to be estimated by engineering judgments. This issue results in uncertainty in input variables and consequently, in the network's responses. Major uncertainties in the analysis of pipe networks are resulted from the estimation of pipe friction factors and nodal demands, which are not only imprecise in nature but also significantly change over time. For instance, the network future demands are roughly estimated according to the available experiences and based on statistic data of the population growth and water consumption patterns. Furthermore, pipe friction coefficients as a function of pipe roughness and fluid specifications are continuously under the influences of increasing corrosion and deposition and changes in quantity and quality of water in the system. Depending on the flow characteristics and the network's graph configuration, input uncertainties are spread out over the system and impress the network's responses and performance. For considering these issues, it is required to develop a hydraulic simulation model that can handle such uncertainties in the flow calculations.

Apart from how the uncertainty resources are identified and quantified in a system, there is a challenging issue to apply them as imprecise values to the governing equations. In recent years, this importance has been the subject of interest of many investigators in various fields of engineering. For an engineering system, it is aimed at developing a mathematical model to find out that in what extent the input uncertainties are spread over the output responses. For a long time, the theory of statistics and probability was the predominant approach for handling uncertainties in simulations. In this context, the Monte Carlo method has been one of the most popular approaches for representing and analyzing uncertainties in engineering systems as well as in pipe networks (Bargiela and Hainsworth, 1998; Bao and Mays (1990); Duan et al., 2010; Aghmiuni et al., 2013). Generally, in order to develop
a reliable probability distribution for every parameter having inherent uncertainty, extensive measurements and empirical data are required. Using the Monte Carlo simulation, the system's responses are over and over calculated, each time using a set of random values from the probability functions. It finally produces distributions of possible outcome values. Depending upon the number of uncertainties and the range of variations specified for them, a Monte Carlo simulation could involve more than tens of thousands of recalculations to be completed. In many engineering problems faced with significant uncertainties, e.g., pipe networks analysis and design, the Monte Carlo approach can be computationally very expensive and burdensome.

An alternative approach for the uncertainty analysis is the application of fuzzy set theory and fuzzy logic. Fuzzy set theory was introduced by Zadeh (1965). Principles of the theory in both terms of mathematics and applications were then developed by him and his colleagues as well as by Mamdani and Assilian (1975), Sugeno (1985), Zimmerman (1985), Buckley (1987), Pedrycz (1989), Kandel (1992), Bit et al. (1992), and Bardossy and Duckstein (1995). The fuzzy set theory was initially intended to be an extension of dual logic and/or classical sets theory (Zimmermann, 2010) however, during the last decades, the concept of fuzziness has been highly developed in the direction of a powerful ‘fuzzy’ mathematics. Many scientific and engineering applications can be named which have successfully exploited the fuzzy set theory as a strict mathematical framework. By fuzzy mathematics, vague conceptual phenomena can be more precisely studied. In this view, the fuzzy set theory can be taken into account as a serious rival to the probability theory for the uncertainty analysis in engineering problems. This capability was first pointed out by Zadeh (1973) when he said: “as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.”

At present, the fuzzy set theory is a popular tool for analyzing fuzzy systems, which are equivocal in nature and contain ambiguity in variables and process. Furthermore, this theory has been widely used in different areas of water resources engineering, for example in hydrology (Pesti et al., 1996; Pongracza et al., 1999; Ozelkan and Duckstein, 2001; Bardossy et al., 2002; Srinivas et al., 2008; Srivastava et al., 2010), water quality (Sasikumar and Mujumdar, 1998; Nasiri et al., 2007; Ghosh and Mujumdar, 2010), ground water (Cuan and Aral, 2004; Dixon, 2005; Kurtulus and Razack, 2010), urban flood management (Chang et al., 2008; Fu et al., 2011), reservoirs operation (Saad et al., 1996; Cheng and Chau, 2001; Karaboga et al., 2004), river engineering (Mujumdar and Subbarao, 2004; Ozger, 2009; Kisi, 2010) and pipe hydraulics (Revelli and Ridolfi, 2002; Branas-Svlicic and Ivetic, 2006; Gupta and Bhave, 2005; Haghighi and Keramat, 2012; Spiliotis and Tsakiris, 2012). Focusing on the theme of this study, number of selected investigations that applied the fuzzy set theory to hydraulic simulation of piping systems are briefly reviewed through the next paragraphs.

Revelli and Ridolfi (2002) published an outstanding paper in which for the first time the fuzzy set theory was used for consideration of uncertainty in analysis of water distribution networks. They coupled a hydraulic simulation model with a gradient-based mathematical optimization method. There, the pipe friction factors and nodal demands were introduced by fuzzy numbers. Through that work, to find the extreme values of each response, the optimization is applied twice to the hydraulic simulation model in each level of uncertainty. The procedure is many times performed until all node pressures and pipe discharges in the network are analyzed in all levels of uncertainty. It was concluded there that the fuzzy responses cannot be explicitly obtained without using optimization. In fact, for every extreme nodal pressure and pipe discharge, there is a critical combination of input uncertainties, which need to be found with the aid of optimization. In spite of the use of a fast classical optimization technique in that investigation, the whole procedure takes a lot of time for computations as well as it may become trapped in local solutions, particularly in large and complex networks.

Branisavljevic and Ivetic (2006) guided a valuable study to investigate the influences of uncertainties on different responses of pipe networks. They found that when the pipe friction factors are considered as the input uncertainties and simulated by fuzzy variables, the network's responses can be classified in two groups. First, there are node pressures that depend on the input variables in a monotonic way and second, there are pipe velocities that depend on the input variables in a non-monotonic way. The node pressures can be analyzed by the hydraulic simulation model in a deterministic iterative way. This is while, for the pipe velocities, the hydraulic model needs to be coupled to an optimization solver. As a consequence, in a general form where both types of responses are of interest, and system may include other uncertainties like in water demands, the problem becomes more complicated, and the optimization is inevitably required.

Gupta and Bhave (2005) paid a special attention to the complexity of the raised optimization in the previous studies and proposed an approach to simplify the problem. They considered the fuzziness of both friction factors and water demands. It was assumed that the maximum impact of the input uncertainties on the responses occurs when the input parameters are at their extreme values. On this basis, a deterministic iterative approach without optimization was proposed to obtain the extreme values of responses. Although the optimization was omitted in that approach, but it still needs heavy combinatorial computations to find the extreme responses by changing the input variables from one extreme to another extreme point for all input variables in all levels of uncertainty.

Spiliotis and Tsakiris (2012) introduced a fuzzy approach for addressing the uncertainty of water demands by the Newton-Raphson method. It was shown that if pipe friction factors are considered to be constant, the principle of monotony of the continuity equation in the network junctions can be exploited. If so, the optimization procedure for handling fuzzy variables would not be required. Thereby, the investigators suggested an explicit algorithm for the fuzzy analysis of pipe networks under fuzzy water demands. However, the method needed to manipulate the governing equations and to develop a particular analysis solver for the network's hydraulics. This can be burdensome for large and complex networks. In addition, the uncertainty of friction factors is still a major concern with that method.

Haghighi and Keramat (2012) introduced a model for studying uncertainty in the transient analysis of pipe networks. It was considered that the water demands; pipe friction factors and wave speeds include uncertainty. These parameters were represented by fuzzy numbers and introduced to the transient analysis solver. In each level of uncertainty, two objective functions were defined at every network junction to minimize the lowest and maximize the highest pressures in the excited transient flow. The simulated annealing method was used there for optimization of the objective functions. The solver was many times applied to the transient simulation model to evaluate the fuzzy responses of the network against the input uncertainties. Burden of computations in both parts of transient analysis and optimization was problematic in that work and made the whole process computationally expensive. It was concluded that in a generalized fuzzy analysis, different input variables such as friction factors, node demands and reservoir levels may include different uncertainties and can be simulated by different shapes of fuzzy numbers. This issue, especially
in complex and large networks, causes changes in the responses in a non-monotonic fashion with the input variables. In fact, for every response, there are two different combinations of input uncertainties that result in its extreme high and low values. Searching for these critical combinations and their corresponding critical responses needs the use of optimization techniques.

This research considers that the iterative and nonsystematic use of single objective optimization for fuzzy analysis of pipe networks highly increases the complexity and computational load of the problem solution. To handle this issue, uncertainties are represented by fuzzy numbers with a given membership function to apply to the governing equations. Fuzzy variables are discretized in some levels, and the network's extreme responses are evaluated in each level separately. To analyze the hydraulic responses' uncertainty, a multiobjective optimization problem is developed. The raised problem consists of a large number of objective functions to evaluate the minimum and maximum values of all node pressures and pipe velocities. For solving the problem, the method of Non Dominated Sorting Genetic Algorithm (NSGA-II) with some modifications is coupled to a hydraulic simulation model. It is applied against an example pipe network and a real case study, and the obtained results are finally discussed. The methodology and main components of the model are in details described in the following.

2. Fuzzy set theory and the uncertainty representation

For designing a pipe network it is required to compute the nodal pressures and pipe velocities against design scenarios to check whether they satisfy the design criteria or not. In common approaches, a fixed unique value is assigned to every design variable upon some information on system, technical manuals and previous experiences. The network is analyzed against the unique input variables and returns the unique output responses. This is while, with no doubt, every design variable may include uncertainty. Major uncertainties in the hydraulic analysis of water distribution networks are with the estimation of pipe friction coefficients and nodal demands. Depending on the network configuration and the governing flow specifications, input uncertainties are stochastically distributed over the system and result in uncertainty in the responses too. For a reliable design, estimation of the responses' uncertainty is of the central importance. This issue is followed up here by using the fuzzy set theory.

In the fuzzy set theory, the uncertain parameters are represented by fuzzy numbers. A fuzzy number N is, in fact, a set of real numbers so that N ⊂ R. For a normalized fuzzy number N, there is a continuous membership function μ(α)(x) ∈ [0, 1] that describes to which grade a variable x ∈ N belongs to N. When μ(α)(x) = 0 it means that x is “not included” and when μ(α)(x) = 1 it means that x is a “fully included” crisp number in the fuzzy set. In case of 0 < μ(α)(x) < 1, x is a number that is partially belonging to N with the membership of grade μ(α)(x). Fig. 1 illustrates a triangular normalized fuzzy number defined by three special values for x consisting of xα, xc, and xβ such that xα < xc < xβ. As seen in Fig. 1, the membership function μ at xc and xβ is zero while at xα is one. xα is the most likely value of N and the interval [xα, xβ], where the membership function at its extreme points is zero is called the support of the fuzzy number N. One of the useful approaches for handling fuzzy numbers in complex engineering systems is the a−cut decomposition method. An operation so-called a−cut, where a ∈ μ(α), is applied to the fuzzy set to facilitate the interpretation of the fuzziness in engineering systems, which are originally developed on the basis of fixed crisp variables. a = 0 and 1 respectively correspond to the support interval and the most likely value of N. In general, every a−cut on N creates a sub-fuzzy number Nα whose support is interval [xα, xβ] and the most likely value is still xα. In this study, the fuzzy number is used to represent uncertainty in a variable like x which is estimated to be about xα. However, due to the uncertainty at level α, it can also take a value between xα, α and xβ, α. Lower α−cuts introduce more uncertainties from the input variables to the system under consideration. In the fuzzy set theory, this issue is known as the fuzzy number convexity, which mathematically implies that for two arbitrary α1 and α2 if α1 < α2 then Nα1 ⊂ Nα2 and in other words, xα, α1 < xα, α2 and xβ, α1 > xβ, α2.

For uncertainty analysis of pipe networks, it is considered that pipe friction coefficients C as well as nodal demands Q are uncertainty and are treated in simulations by the triangular fuzzy numbers. The capital sign ‘′’ is to emphasize that the variable has uncertainty and is represented by the fuzzy number. The most likely value C′ and Q′ is assigned to every variable on the basis of available information and possible estimations. Then, upon the knowledge one has about the accuracy or inaccuracy of the data and relying on engineering judgments the support of each variable, i.e., {C′, C′} and {Q′, Q′} is introduced. As a consequence, a triangular fuzzy number for every input fuzzy variable is formed. This makes it possible to translate the qualitative terms like “about, close to, probably and approximately” with the variables, in mathematical expressions to take them into account in numerical analyses. When fuzzy variables C′ and Q are introduced to a network, the applied analysis model needs to evaluate the fuzzy responses including, pipe velocities V and nodal pressures P. This task implies that the model must be able to find out that in what extent the input uncertainties are spread over the system and reflected in the responses. This issue cannot be straightforwardly fulfilled by the common approaches and available simulation models like EPANET, which works on the basis of unique input variables and gives unique output responses. For analysis of pipe networks having fuzzy variables, the hydraulic simulation model needs to be equipped with other mathematical techniques as discussed in the following.

3. Fuzzy analysis model

Two conservative rules of flow continuity in nodes and energy in loops govern the steady-state flow distribution in pipe networks. Flow continuity in every node and energy conservation in every loop is respectively met by the following equations.

\[ \Sigma Q_{\text{in}} + \Sigma Q_{\text{out}} = Q \quad \text{for every node} \]  \hspace{1cm} (1)

\[ \Sigma h_{\text{f}} + \Sigma E_{\text{p}} = 0 \quad \text{for every loop} \]  \hspace{1cm} (2)

where Qin and Qout = respectively, the flow discharges toward and out of the node, Q = the node demand, Ep = the energy added to the system by pumps and hf = head losses computed for all pipes in the loop. hf can be estimated by Darcy–Weisbach (DW), Manning and Hazen–Williams (HW) equations however, the latter

![Fig. 1. A triangular normalized fuzzy number.](image)
is more popular in practice. The HW is an empirical equation which, can be written in a general form as follows.

\[ h_L = \frac{\alpha L Q^2}{C D^2} = (Z_u - Z_d) + \left( \frac{P_u - P_d}{\gamma} \right) \]  

(3)

in which, \( C \) = the HW coefficient, \( \omega \) = numerical constant depending on the problem’s units. \( \omega \) is practically ranging from 10.5 to 10.9 in the metric system. \( \alpha \) and \( \beta \) are the HW exponents which are respectively 1.85 and 4.87, \( D \) = pipe diameter, \( L \) = pipe length, \( Z \) = elevation from datum, \( P \) = pressure, \( \gamma \) = specific gravity and subscripts \( u \) and \( d \) respectively denote the upstream and downstream ends of the pipe.

The above equations are developed for all nodes and loops in network and result in a system of nonlinear equations. This system may be arranged with respect to either nodal heads or pipe discharges as the problem unknowns. In any case, the equations must be solved using a numerical method such as the linear theory or gradient method (Todini and Pilati, 1987) which is utilized in EPANET. The governing equations and the solver accept only the unique values for the input parameters like the HW coefficients \( C \) and \( q \), the system’s unknowns are obtained as fuzzy pressures \( P \) and fuzzy discharges \( Q \) and velocities \( V \) too (Fig. 2). In this condition, instead of unique values for input variables, an interval of possible values is introduced to the network. Thereby, for each response, there will be infinite combinations of the input fuzzy variables.

The aim of fuzzy analysis is not to find all these combinations for a response but to obtain the interval that the responses of these combinations belong to. For this purpose, the following algorithm proposed by Revelli and Ridolfi (2002) is adopted here.

The input fuzzy numbers, \( \hat{C} \) and \( \hat{q} \), are discretized with a limited number of \( \alpha \)-cuts including \( \alpha = 0 \) and \( \alpha = 1 \) and a few numbers between them.

1. First, the cut \( \alpha = 1 \) is taken into account at which, all input variables take their most likely values, \( \hat{C} \) and \( \hat{q} \), EPANET is run and the most likely values of network’s responses, \( \hat{P} \) and \( \hat{V} \), are obtained.
2. Another \( \alpha \)-cut is considered, according to the range of variations of input variables, \( [C_{\alpha}, C_{\beta}] \) and \( [q_{\alpha}, q_{\beta}] \), is determined from the input fuzzy numbers (Fig. 1).
3. In every \( \alpha \)-cut, the fuzzy analysis of the pipe network is treated as an optimization problem. For every response, two single-objective optimization problems are introduced such that the response is the objective function that is once maximized and once minimized while, the input fuzzy numbers are the decision variables. The \( \alpha \)-cut under consideration determines the limits of the decision space. This mathematical programming can be written as follows, Maximize\{Minimize

\[ \hat{P}_i = \varphi(C_1, 2, 3, \ldots, m; \hat{q}_{i1,2,3,\ldots,n}) \quad i = 1, 2, 3, \ldots, n \]  

(4)

\[ \hat{V}_j = \varphi(C_1, 2, 3, \ldots, m; \hat{q}_{j1,2,3,\ldots,n}) \quad j = 1, 2, 3, \ldots, m \]  

(5)

subject to,

\[ \hat{C}_{\alpha} \leq \hat{C} \leq \hat{C}_{\beta, \alpha} \quad \text{and} \quad \hat{q}_{\alpha} \leq \hat{q} \leq \hat{q}_{\beta, \alpha} \]  

(6)

where \( n \) and \( m \) respectively, the number of nodes and pipes in network and \( \varphi \) is, in fact, the hydraulic simulation model, EPANET, which acts as a function of input uncertainties and returns the network’s responses. After that the above programming problem was optimized for the whole network, the obtained maximum and minimum objective functions result in the range of variations of the network’s responses corresponding to the \( \alpha \)-cut at hand, i.e., \([P_{b, \alpha}, P_{b, \alpha}]\) and \([V_{b, \alpha}, V_{b, \alpha}]\) for all nodes and pipes. Fig. 2 schematically illustrates the connection between the optimization and EPANET in the described scheme.

4. Next \( \alpha \)-cut is taken into account and the procedure is continued until the support of fuzzy numbers (\( \alpha = 0 \)) is calculated too, and the shape of the membership function for all fuzzy responses is completed.

When the above algorithm is viewed as a close package, it can be mentioned that the fuzzy analysis of pipe networks is a multi-criteria decision making problem with many objective functions. However, the challenges between the objectives do not matter here. One approach to solve this problem is to optimize each objective function separately by a single objective solver. If so, for every \( \alpha \)-cut and for every node pressure or pipe velocity, it is required to carry out the optimization twice, to find the upper and lower extremities of the response (Fig. 2). In this context, Revelli and Ridolfi (2002) used the mathematical method of Quadratic Programming (QP) and Haghighi and Keramat (2012) used the metaheuristic of Simulated Annealing (SA). By using this approach, for fuzzy analysis of the whole network the optimization must be \( 2 \times (m + n) \) times applied to the problem for every \( \alpha \)-cut. This issue is the main concern with the procedure that makes the fuzzy analysis computationally inefficient and burdensome to implement.

In spite of all the aforementioned complexities for the fuzzy analysis of pipe networks, the problem has some particular features that help solve it more efficiently if they are appropriately

\[ \begin{array}{c}
\text{Fuzzy HW coefficients} \\
\alpha \text{-cut operator} \\
\text{Fuzzy nodal demands} \\
\text{Optimization} \\
\text{EPANET} \\
\text{Crisp pressures} \\
\text{Crisp velocities} \\
\text{Network’s configuration and other fixed design parameters}
\end{array} \]

\[ \begin{array}{c}
\hat{C}_1 \\
\hat{C}_2 \\
\vdots \\
\hat{C}_m \\
\hat{q}_1 \\
\hat{q}_2 \\
\vdots \\
\hat{q}_n \\
\hat{P}_1 \\
\hat{P}_2 \\
\vdots \\
\hat{P}_n \\
\hat{V}_1 \\
\hat{V}_2 \\
\vdots \\
\hat{V}_m
\end{array} \]

Fig. 2. A cycle of optimization-hydraulic simulation for the fuzzy analysis of pipe networks.
utilized. First, all objective functions have the same decision variables and decision space. Second, the purpose of the optimization is only finding the extreme values of pressures and velocities in every $\alpha$—cut and the trade-off between them is not important. Third, all objectives are evaluated by a common function named as $\varphi$ in the above mathematical programming. Function $\varphi$ is, in fact, the EPANET hydraulic simulation model which returns all node pressures and pipes velocities once it is run against a combination of input uncertainties (Fig. 2). In other words, every single simulation run of EPANET contains all objective values in a certain set of decision variables. Utilizing this characteristic in the optimization allows to evaluate a candidate solution simultaneously for all objective functions that, leads to exploit the hydraulic solver much more efficiently. In approaches that the problem is treated as an iterative single objective optimization, most computational efforts and results of the hydraulic solver are wasted since at each time only one objective is optimized, and the others are ignored (Fig. 2). Consideration of the above three features implies that the current fuzzy analysis problem can be more efficiently solved if it is treated as a multiobjective optimization. For this purpose, the optimization box in Fig. 2, is equipped with a multiobjective optimization solver which is fed back with all objective functions once EPANET is run. In this study, to optimize functions of (3) and (4) a version of NSGA-II method is utilized.

4. NSGA-II for fuzzy analysis

The Non-dominated Sorting Genetic Algorithm abbreviated as NSGA (Srinivas and Deb, 1995) was one of the first evolutionary algorithms that utilized the principle of Pareto optimality in solving multiobjective problems. Deb et al. (2002) raised some criticisms to the NSGA and developed a powerful approach known as NSGA-II. In comparison with the previous version, the NSGA-II has a less computational complexity, considers elitism, systematically preserves the diversity of Pareto-optimal solutions and adaptively handles the problem constraints. These features have made the NSGA-II very successful and popular in a wide range of engineering problems.

As earlier discussed, for the fuzzy analysis of a pipe network it is required to solve a special multiobjective problem. Herein, a multiobjective genetic algorithm on the basis of the NSGA-II is utilized. Before that, the fundamentals of the method are briefly described and the main steps of the standard NSGA-II is quickly reviewed.

The main concept behind every non-dominated sorting multi-objective optimization is to identify and promote the solutions that dominate other solutions. The basic definition for the notion of dominancy in multiobjective optimization is that; the solution $x$ dominates the solution $y$ if both following conditions are simultaneously met:

1. None of the objectives in $x$ is worse than in $y$.
2. At least one objective in $x$ is better than in $y$.

In solving real-world problems other aspects may be also added to the above definition as Deb et al. (2002) did when developing the NSGA-II. They proposed a new operator that not only fulfills the above criterion but also, in lower levels, preserves the diversity of solutions in the Pareto fronts as well as the feasibility of the search. Through the next paragraph, the main skeleton of the NSGA-II is briefly described. Afterward, the differences between the multiobjective optimization raised in the fuzzy analysis of pipe networks with the common multiobjective problems are addressed and accordingly, the applied NSGA-II in this study is introduced.

**Standard NSGA-II:** The optimization is begun by an initial population that is randomly generated and then evaluated by all objective functions. The population is sorted based on the non-dominance principle. Each chromosome is then assigned a rank number equal to its non-dominance level. For each level, a Pareto front is introduced in which the corresponding chromosomes are stored. Solutions belonging to the best non-dominated fronts are directly transferred to the mating pool to create the next generation. In conventional multiobjective problems, it is also important to maintain a good spread of the solutions in the optimum fronts. To this end, NSGA-II employs the crowded-comparison scheme for preserving diversity among the population members. For every chromosome in a Pareto front, a crowding distance is measured as the distance of the biggest cuboid contacting the two neighboring solutions as shown in Fig. 3. For each objective function, the boundary solutions having the minimum and maximum function values (points A and B in Fig. 3) are assigned a huge distance value, e.g., infinity. These solutions are, in fact, the extremities of the non-dominating front and must be emphasized more than intermediate solutions. As a consequence, every chromosome in the population will have two attributes, including the non-dominance rank (or Pareto front level) and crowding distance. Thereby, a hybrid domination criterion namely the crowded-comparison operator is applied to the selection process so that it is guided at the various stages of the algorithm toward a uniformly spread-out Pareto-optimal front (Deb et al., 2002). Simply speaking, the crowded-comparison operator is so applied to the population that between two solutions with different non-dominance rank, the solution with lower rank is selected. However, if both solutions have a same rank, the one with higher crowded distance is preferred. For the selection process, the binary tournament method is applied to the parent population $P$. Then; the selected parents are mated to generate the new offspring. After mutating new members, the
offspring population $Q$ is formed and merged into the parent population resulting in a combined population $R = P \cup Q$. The population $R$ is sorted based on the non-domination, and the Pareto fronts with different levels are formed. Then, the best half of the combined population $R$ is transferred to the next generation. Since all parents and offspring chromosomes are included in $R$, elitism is automatically provided. The procedure is continued until the desired convergence is achieved.

Deb et al. (2002) also proposed an adaptive approach for constraint handing in NSGA-II for solving constrained multiobjective optimization problems. Similar to the crowded-comparison operator, their constraint handing method also uses the binary

---

**Table 1**

<table>
<thead>
<tr>
<th>Pipe No.</th>
<th>$\hat{\zeta}_{c,a=0}$</th>
<th>$\hat{\zeta}_{\text{crisp}}$</th>
<th>$\hat{\zeta}_{b,a=0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63.19</td>
<td>76</td>
<td>82.87</td>
</tr>
<tr>
<td>2</td>
<td>73.26</td>
<td>80.5</td>
<td>87.74</td>
</tr>
<tr>
<td>3</td>
<td>79.87</td>
<td>88</td>
<td>95.66</td>
</tr>
<tr>
<td>4</td>
<td>67.1</td>
<td>75</td>
<td>80.36</td>
</tr>
<tr>
<td>5</td>
<td>72.58</td>
<td>78</td>
<td>86.94</td>
</tr>
</tbody>
</table>
tournament selection in such a way that two solutions are picked from the population, and the better one is selected. In the presence of constraints, a new definition for domination is defined and replaced for the previous definition. In constrained problems, solution \( x \) dominates \( y \) only if,

1. \( x \) is feasible and \( y \) is not.
2. \( x \) and \( y \) are both infeasible, but \( x \) has a smaller overall constraint violation (with respect to all objective functions and constraints).
3. \( x \) and \( y \) are both feasible, and \( x \) dominates \( y \) by the previously introduced criteria.

This was the main structure of the NSGA-II which has been so far applied to many complex multiobjective optimization problems successfully. However, the current problem in some senses is different from the common multiobjective problems. The main differences were already pointed out. In this problem, the number of objective functions is very large but the trade-off between them does not matter. The main goal of the current multiobjective optimization is only to find the optimum boundary solutions, and the diversity preservation among the solutions is not important. On the contrary, it is desired here to guide the solutions in Pareto fronts toward the boundary solutions, i.e., the maximum and minimum values of each objective function. This issue requires an absolutely different definition for the crowded-comparison operator in the selection process. With respect to this issue and the fact that the problem is unconstrained, a new version of NSGA-II for the fuzzy analysis of pipe networks is introduced as follows.

1. An initial random population of size \( N \) is generated.
2. The flow simulation model is called to analyze the network against the population. All \( M \) objective function values, pressure and flow responses, are evaluated for every individual by a single one run of EPANET.
3. The population is ranked based on the non-domination, and the Pareto fronts from level 1 to \( l \) are formed.
4. Instead of the crowding distance in the standard NSGA-II, a new metric is defined here to measure the distance between every solution and the extreme points in the Pareto front. This metric named as the closeness-distance is computed for every individual \( i \) (\( i = 1 \) to \( N \)) with respect to every the objective function \( j \) (\( j = 1 \) to \( M \)) by the following equation,

\[
d_{ij} = \frac{f_{ij} - f_{jmin}}{f_{jmax} - f_{jmin}}
\]

where \( f_{jmin} \) are \( f_{jmax} \) are respectively the minimum and maximum values of the objective function \( j \) in the population, and \( f_{ij} \) is the value of the objective function \( j \) in the solution \( i \). The closeness-distance schematically shown in Fig. 4. \( d_{ij} \) is a normalized dimensionless value which differs from zero in the boundary solution A to 1 in the most distant solution, i.e., solution B in the Pareto front (Fig. 4). After that for all individuals the closeness-distances with respect to all objective functions are computed.
objective functions were computed, the minimum of $d_{i1}$ to $M$ is assigned to $i$ as its selected closeness-distance. Thereby, the boundary solutions, points A and B in Fig. 4, as well as the solutions close to them are emphasized through the selection process.

5 The selection of parents is done according to the non-dominination and closeness-distance criteria. Using the tournament method, two individuals $x$ and $y$ are randomly picked up from the population. With respect to the non-domination rank number and the selected closeness-distance, between $x$ and $y$ the one wins the tournament that (a) belongs to a better level of Pareto front and (b) has a smaller closeness-distance if they both belong to a same front.

6 An appropriate crossover technique is applied to the selected parents to create $N$ offspring. In this work, the blend crossover method (BLX-$\alpha$) proposed by Eshelman and Shaffer (1993) is adopted.

7 A few genes of the new chromosomes are randomly mutated.

8 The children and parents’ populations are combined together. The resulted population is sorted based on the Pareto non-domination criterion.

9 Old individuals are replaced by better offspring, and the new generation is formed.

10 If the stopping criteria are met the optimization is terminated otherwise; it goes to step 2.

At the end of optimization, the boundary solutions in the first Pareto-optimal front represent the extreme fuzzy responses of the network according to the $\alpha$-cut introduced to the input variables. The described procedure has been also presented by a flowchart in Fig. 5.
5. Examples

In this part, the proposed algorithm is applied against two pipe networks. The first example is a small network taken from the literature, and the second one is a real case study.

5.1. Example 1

To evaluate the proposed fuzzy approach, a small pipe network originally introduced by Revelli and Ridolfi (2002) is taken into account. The network as shown in Fig. 6, is consisting of two loops, five pipes and four nodes. The pipe diameters and lengths as well as nodal demands and elevations are given in Fig. 6. In this example, it is supposed that only pipe friction coefficients (Hazen–Williams) include uncertainty and need to be treated by fuzzy numbers. The maximum uncertainty of Hazen–Williams coefficients, i.e., the support of input fuzzy numbers in $\alpha = 0$ is presented in Table 1. In the present analysis, fuzzy responses of nodal pressures at nodes 2, 3 and 4 and pipe discharges in all pipes are computed in three levels of $\alpha = 0, 0.5$ and 1. Revelli and Ridolfi (2002) utilized a single-objective optimization model to solve this problem. They applied the Quadratic Programming (QP) method as a classical optimization technique, to maximize and minimize each hydraulic response in each $\alpha$-cut. On the other hand, in that work, for finding the maximum and minimum values of three nodal pressures and five pipe discharges in two $\alpha$-cuts (0 and 0.5), overall, the QP solver was carried out 32 times independently. The QP results are shown in Figs. 7 and 8 by dash lines.

Herein, the example is also solved using the NSGA-II method. For this purpose, the network is analyzed in two steps, including all hydraulic responses in cuts $\alpha = 0$ and 0.5. As a result, in each $\alpha$-cut, the NSGA-II must optimize 16 objective functions simultaneously. As earlier discussed, a new feature namely the closeness–distance was added to the NSGA-II to improve its efficiency in solving fuzzy problems. This feature would lead the Pareto solutions toward the end points (A and B in Fig. 3) which are the responses of interest in this fuzzy analysis. To evaluate the performance of this feature in solving the problem at hand, both standard and new NSGA-II schemes were applied to the network analysis. For both algorithms, the population size was considered to be 200 and the mutation ratio was set 0.02. Both schemes ultimately resulted in the same responses. However, the new version was faster than the standard version. The closeness–distance operator in the new version significantly speeds up the algorithm convergence in finding the extreme points on the optimal Pareto front presented in Figs. 7 and 8. Several primary runs were done for this example, and it was concluded that, averagely, the new version of NSGA-II converges to the results shown in Figs. 7 and 8 in less than 100 generations while, the standard NSGA-II gives the same results after about 250 generations. This turns out that the new feature works well in analyzing fuzzy systems in which the diversity of solutions is not important.

Furthermore, as seen in Figs. 7 and 8, the recently obtained results by NSGA-II are very close to the previously obtained by the QP. Not only that, but in some cases like Q4 and Q5, NSGA-II has given slightly better results. This confirms that the proposed fuzzy

![Fig. 10. Maximum uncertainty in the nodal pressures.](image1)

![Fig. 11. Maximum uncertainty in the pipe velocities.](image2)
approach is capable of uncertainty analysis in the network with an efficient and reliable algorithm.

5.2. Example 2 (case study)

A real pipe network is now analyzed by the described fuzzy approach. The network has been constructed in a 120 ha area to supply the water demands of an industrial park near the Persian Gulf in southwest of Iran. The network has 46 nodes and 65 pipes and a reservoir with 27 m piezometric head that feeds the demands by gravity. The network’s layout configuration, including the pipe and node names as well as the most likely demands are depicted in Fig. 9. The network has been constructed in a quite flat area which is about 2 m higher than the mean sea level. The pipes are made of polyethylene (PE80) and the most likely HW coefficient for all of them was considered to be 130 when designing the network. The pipe lengths and inner diameters are also presented in Table 2. As the design criteria for sizing the pipe diameters and reservoir elevation, the minimum required nodal pressure and pipe velocity were respectively considered to be 18 m and 0.3 m/s. The network was designed according to the future estimated demands, shown in Fig. 9, and the aforementioned friction coefficient. As discussed in the manuscript, both water demand and friction coefficient are significantly under the influence of uncertainties during the lifetime of the system. The operators are curious to know how the system reacts to the input uncertainties, and its performance reliability is affected. The fuzzy analysis method is used to respond this question. For this purpose, the pipe friction coefficients and the nodal demands are represented by triangular fuzzy numbers (Fig. 1) while, their most likely values are considered to be those used in the original design. The HW coefficients and the nodal demands are respectively assumed to have ±10% and ±15% uncertainty with respect to their most likely quantities. On this basis, the support of every fuzzy number is obtained. For example, for a pipe HW coefficient with the most likely value \( C_c = 130 \), an isosceles triangular fuzzy number is created whose the support interval is \([117, 143]\). Now, we are going to find out that what will happen to these uncertainties when they are introduced to the network, how they are spread out over the network and in what extent; they influence the network’s hydraulic performance. In this case study, except the reservoir, there are 45 nodes and 65 pipes. If all nodal pressures and pipe velocities are the responses of interest, there will be 110 objective functions in each \( \alpha \)-cut. Since every objective function needs to be once maximized and once minimized, the number of required optimizations becomes 220 for every \( \alpha \)-cut. For the current pipe network, the fuzzy numbers are discretized in three \( \alpha \)-cuts including \( \alpha = 0, 0.5 \) and 1. Except for \( \alpha = 1 \) that is corresponding to the most likely values, for fuzzy analysis of the whole network in \( \alpha = 0 \) and 0.5, the problem is totally consisting of 440 single objective functions, which indeed introduce a very large nonlinear optimization problem. To make the problem more manageable during the optimization, it is divided into eight independent sub-problems in every \( \alpha \)-cut. Before that, to calculate the most likely values of the pipe velocities \( V_c \) and nodal pressures \( P_c \)

Fig. 12. Membership function of the fuzzy pressures.
corresponding to \( a = 1 \), the hydraulic solver, EPANET, is run against the most likely values, \( c_i \) and \( q_i \). Afterward, in each of the remaining \( a - \) cuts \( (a = 0.5 \text{ and } 0) \) the extreme values of \( \{ \bar{P}_{a,a}, \tilde{P}_{b,a} \} \) and \( \{ V_{a,a},V_{b,a} \} \) for the whole network are calculated through solving the following eight multiobjective optimization sub-problems. Each sub-problem has a different set of objective functions but the same decision variables and search space.

1. Minimization of \( P_1 \) to \( P_{23} \) to find \( \bar{P}_{a,a} \) values.
2. Minimization of \( P_{24} \) to \( P_{45} \) to find \( \tilde{P}_{b,a} \) values.
3. Maximization of \( P_1 \) to \( P_{23} \) to find \( \bar{P}_{a,a} \) values.
4. Maximization of \( P_{24} \) to \( P_{45} \) to find \( \tilde{P}_{b,a} \) values.
5. Minimization of \( V_1 \) to \( V_{33} \) to find \( V_{a,a} \) values.
6. Minimization of \( V_{34} \) to \( V_{65} \) to find \( V_{a,a} \) values.
7. Maximization of \( V_1 \) to \( V_{33} \) to find \( V_{b,a} \) values.
8. Maximization of \( V_{34} \) to \( V_{65} \) to find \( V_{b,a} \) values.

The NSGA-II is separately applied to either of the above sub-problems. In all optimization runs the population size is set to be 250, and the mutation ratio is considered to be 0.025. Following the algorithm shown in Fig. 5, the fuzzy analysis in every \( a - \) cut is started by a randomly generated population and is terminated by checking the convergence criteria. Averagely in all runs, the best Pareto-optimal front was obtained after a bout 600–700 generations.

After optimization of all sub-problems, the uncertainty distribution over the network is obtained. One of the most important results from this analysis is the support of the fuzzy responses. The support values represent the maximum uncertainty in the responses corresponding to the maximum uncertainty in the input variables. Figs. 10 and 11 respectively present the extreme values within the support intervals for the fuzzy pressures and velocities along with their most likely values already used to design the network. These figures clearly show that how the input uncertainties are spread out on the network and influence the results. As it is rationally expected, the nodes more distant from the reservoir are more affected by uncertainty, while the closer nodes are more resistant to the input uncertainties. The network nodes averagely include \(-10.73\%\) (ranging from \(-0.16\%\) to \(-24.43\%\)) and \(+6.13\%\) (ranging from \(+0.08\%\) to \(+14.60\%\)) uncertainty with respect to the original design values. Furthermore, to better see how the input triangular fuzzy numbers are distributed to the system and appear in the responses; some nodal pressures, including large uncertainty have been selected to show in Fig. 12. Some nodal pressures have about \(-24\%\) uncertainty resulting in that, the reliability in these nodes with respect to the minimum required pressure may fail. For example, in nodes 37, 41, 42, 43, 45 and 46, the most likely pressure (used in the original design) is respectively 19.11, 18.71, 18.45, 18.42, 19.12 and 18.5 m. However, due to the introduced uncertainty their pressures may become 14.84, 14.25, 13.96, 13.92, 14.95 and 14.03 m, all of which are lower than the required minimum value (18 m). Also, by inspecting the entire network it is found that about 33.3% of nodes may experience lower than 18 m pressure values.

A similar argument can be provided on the pipe velocities. As seen in Fig. 11, the input uncertainties have made more

![Fig. 13. Membership function of the fuzzy velocities.](image-url)
significant changes to the flow distribution and consequently, to the pipe velocities rather than the nodal pressures. The network pipes generally include a range from about 39.71% (ranging from 12.72% to about 100%) and +47.88% (ranging from +14.29% to +127.27%) uncertainty with respect to the original design values. In Fig. 13, the fuzzy numbers of some selected pipes having the most uncertainty in the whole network are presented. As seen, the velocity in all of these pipes may become zero due to the input uncertainty while, the recommended minimum velocity is 0.3 m/s. The most likely values in Figs. 11 and 13 clearly show that the constraint of minimum velocity has been strictly met in the original design so that all pipes in the network (in the absence of uncertainty) have velocity greater than 0.3 m/s. However, because of uncertainty, it is found that about 31% of pipes may experience less than 0.3 m/s velocity.

6. Conclusion

Estimation of water demands and pipe friction coefficients introduces significant uncertainties to analysis and design of water networks. These uncertainties are spread out over the system and make the network's responses uncertain too. This study introduced a mathematical model for taking into account the input uncertainties in the network hydraulic simulation. For this purpose, an analysis package utilizing the fuzzy set theory, NSGA-II method and the hydraulic simulation model of EPANET was developed. EPANET accepts only unique (crisp) values for nodal demands and pipe friction coefficients and gives only unique values for nodal pressures and pipe velocities. When the input uncertainties are represented by the fuzzy numbers and introduced to the network's governing equations by using the α-cut method, a multi-criteria decision making problem is formed, which needs optimization techniques to be solved. The raised problem has many objective functions, which do not challenge with each other and therefore, the trade-off between them does not matter. The problem is neither single nor multi-objective optimization such that, in the previous investigations, it was treated as a multi-single-objective optimization. The conventional approach to solve the problem is to apply a single-objective solver, every time to only one of the objective functions. In this study, a multiobjective optimization method on the basis of the NSGA-II operator in the standard NSGA-II. This new metric guides the network's responses uncertain too. This study introduced significant uncertainties to analysis and design of water networks. These uncertainties are spread out over the system and make the network's responses uncertain too. This study introduced a mathematical model for taking into account the input uncertainties in the network hydraulic simulation. For this purpose, an analysis package utilizing the fuzzy set theory, NSGA-II method and the hydraulic simulation model of EPANET was developed. EPANET accepts only unique (crisp) values for nodal demands and pipe friction coefficients and gives only unique values for nodal pressures and pipe velocities. When the input uncertainties are represented by the fuzzy numbers and introduced to the network's governing equations by using the α-cut method, a multi-criteria decision making problem is formed, which needs optimization techniques to be solved. The raised problem has many objective functions, which do not challenge with each other and therefore, the trade-off between them does not matter. The problem is neither single nor multi-objective optimization such that, in the previous investigations, it was treated as a multi-single-objective optimization. The conventional approach to solve the problem is to apply a single-objective solver, every time to only one of the objective functions. In this study, a multiobjective optimization method on the basis of the NSGA-II operator in the standard NSGA-II. This new metric guides the network's responses uncertain too.

References


