IJST, Transactions of Civil and Environmental Engineering, Vol. 35, No. C1, pp 105-119
Printed in The Islamic Republic of Iran, 2011

A SELF ORGANIZING MAP BASED HYBRID MULTI-OBJECTIVE OPTIMIZATION OF WATER DISTRIBUTION NETWORKS*

K. NOROUZI, AND G. R. RAKHSHANDEHROO**

Dept. of Civil Eng., Shiraz University, Shiraz, I. R. of Iran
Email: rakhshan@shirazu.ac.ir

Abstract– Water Distribution Networks (WDNs) are an essential infrastructure of every civilization. In the past decades, there has been a lot of work on the optimization of WDNs. This paper presents a hybrid NSGA-II for multi-objective optimization of combinatorial WDN design, utilizing the SOM network as a tool to find the genotypic or phenotypic similarities. SOM is a versatile unsupervised Artificial Neural Network (ANN) that can be used to extract the similarities and find the related vectors with the use of a proper similarity measure. The proposed method, SOM-NSGA-II, derives subpopulations or virtual islands for inbreeding similar individuals to speed up the convergence process of the optimization. The cross-over operation between similar individuals of the subpopulations at the constraint dominated region of the solution space showed a faster convergence and a wider Pareto front for the test problems considered. An added advantage of the method is the application of genotypic sorting of the population by SOM for visual representation of the structure of the Pareto front. The resulted maps showed the extent of variation of the decision variables and their relative importance. This method may be utilized to speed up optimization of large scale WDNs and as an important visual aid for decision makers and designers of WDNs.

Keywords– Multi-objective optimization, water distribution networks, self organizing maps, visualization

1. INTRODUCTION

Water Distribution Networks (WDNs) are an essential infrastructure of every civilization. There have been numerous efforts in the past few decades to minimize costs associated with the operation and construction of WDNs, leading to many single objective optimization algorithms [1]. The most widely used algorithms have been linear programming [2], enumeration techniques [3], nonlinear programming [4], genetic algorithm [5], simulated annealing [6], shuffled frog leaping algorithm [7], and ant colony optimization algorithms [8-11]. The method used by these approaches is to transform the constraints and objectives into one objective by aggregating them with penalty factors. A drawback of this single objective optimization is that the choice of penalty factors may affect the resulting optimal solution [12].

Recent researches deal with WDN optimization as a multi objective optimization, owing to consideration of other objectives such as reliability, quality of service, and head deficit [13]. From a more realistic point of view, multi-objective optimization methods provide the tradeoffs between objectives of interest [14]. For example, benefit and quality of a WDN design were considered as a multi-objective optimization using a fuzzy approach [15]. Minimization of cost and maximization of a surrogate measure of reliability were considered in a Non-dominated Sorting Genetic Algorithm (NSGA) to find the tradeoffs between reliability and cost for benchmark problems in WDNs [16]. Lack of eliticism, need for specification of a sharing parameter, and computational complexity of NSGA were alleviated with the

*Received by the editors April 14, 2009; Accepted July 27, 2010.
**Corresponding author

introduction of NSGA-II by Deb *et al.* [17]. NSGA-II was used to find the tradeoffs between head deficit and cost in the Hanoi bench mark problem [18]. A search space reduction has been proposed for the efficient search of the optimum design regarding only the cost of the network [**19**]. The problem of deploying sensors in a large water distribution network was also considered, in order to detect the malicious introduction of contaminants [20]. A set of realistic objective functions, such as reduction of detection time and the population protected from consuming contaminated water, was used in their approach. With the increasing need for optimal operation of WDNs, Shamir and Salomons developed a method for near-optimal real-time on-line operation of an urban water distribution system with the use of reduced models [21]. They showed that the current optimization methods are not adequate for the demands and should be augmented. Another multi-objective approach was introduced for leak detection and rehabilitation of WDNs using NSGA-II [22]. Their results demonstrated the usefulness of multi-objective approach in leakage detection as a function of pipe age and diameter. Giustolisi *et al.* described a procedure for the robust design of water distribution networks which incorporates the uncertainty of nodal water demands and pipe roughness in a multi-objective optimization scheme aimed at minimizing costs and maximizing hydraulic reliability [23]. Acknowledging the need for efficient optimizers, their method aimed to reduce the number of runs by introducing several density functions, and was tested for real WDNs. Optimal pressure management in water distribution systems through the introduction and regulation of pressure reducing valves were studied by Nicoloini and Zovatto [24]. They argued that reduction in pressure is aimed at controlling water leakages which, being in some cases a high proportion of the total volume supplied, are nowadays one of the major concerns for water utilities. The determination of the number, location, and setting of such valves was formulated as a two criteria optimization problem and was solved with multi-objective genetic algorithms.

Artificial Neural Network (ANN) has been utilized in ANN-based meta-models to address the computational complexity and optimization cost of a WDN [25]. Traditionally, Genetic Algorithms (GAs) have been used to fine-tune topology and weights for the ANN [26, 27]. Self Organizing Map (SOM) is an unsupervised and versatile ANN which has been utilized in many applications involving pattern recognition and clustering [28, 29]. SOM networks may be combined with GA to enhance a single objective optimization process in a real world optimization problem [30]. However, the problem is that global stochastic search methods such as GAs require many iterations to be performed in order to achieve a satisfactory solution, and each iteration may involve running computationally expensive simulations. Recently, this problem has been compounded by the evident need to embrace more than a single measure of performance into the design process, since by nature multi-objective optimization methods require even more iterations [31].

This paper presents a hybrid NSGA-II for multi-objective optimization of WDNs, utilizing SOM network as a tool to find the genotypic or phenotypic similarities at the exploitation phase of the Genetic Algorithm. The proposed method forms virtual islands or subpopulations that evolve within themselves for a specified number of generations to speed up the convergence process of the optimization. The method is coupled with EPANET [32] as the network solver for the analysis of two example networks and the results are compared with NSGA-II.

2. MULTI-OBJECTIVE OPTIMIZATION

Optimization is a process of finding a solution or a set of solutions for a problem from potential or feasible solutions that suit a given criteria. When there is only one objective, the preference structure over solutions is well-defined. It is the simple ranking of the objective values and selecting the solution with the extreme value. Therefore, in a single objective problem, there is only one optimum solution which may be identified with certainty [33]. However, in most realistic engineering design problems, there is more than one objective to be considered. The objectives are often conflicting, such that improving one objective may deteriorate the other one(s). Therefore, in a multi-objective optimization, there is more than one solution available and several methods like Non Dominated Sorting Genetic Algorithm (NSGA), weighted min-max, Vector Evaluated Genetic Algorithm (VEGA), Lexicographical ordering, and weighted sum have been presented to find the solution space [33, 34]. NSGA-II [35], which is based on the concept of dominance in order to produce the Pareto front of the objectives, has been evaluated as one of the most successful methods for water distribution network optimizations [20].

By definition, in a minimization problem with M objectives, a feasible solution $x^{(1)}$ "*dominates*" another feasible solution $x^{(2)}$ (stated as $x^{(1)} \leq x^{(2)}$), if both of the following conditions are true [35]:

1) The solution $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives, i.e. $f_j\,(x^{(1)}) \leq f_j\,(x^{(2)})$ for all $j = 1,2,\ldots,M$ objectives:

$$x^{(1)} \leq x^{(2)} \Rightarrow \forall j \in M, f_j(x^{(1)}) \leq f_j(x^{(2)}) \tag{1}$$

2) The design $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective,
or $f_j\,(x^{(1)}) < f_j\,(x^{(2)})$ for at least one $j = 1,2,\ldots,M$ objectives:

$$x^{(1)} \prec x^{(2)} \Rightarrow \forall j \in M, f_j(x^{(1)}) \prec f_j(x^{(2)}) \tag{2}$$

In general, in a multi-objective problem solutions A and B may have the following three possible relations with each other [33]:

A dominates B

B dominates A

Neither A nor B dominate one another.

If two solutions are compared, then the solutions are non-dominated with respect to each other if neither one dominates the other.

A solution $x^{(i)} \in S$ (where S is the set of all feasible solutions) is non-dominated (or a member of Pareto front) with respect to a set $Q \subseteq S$, if there is no $x^{(j)} \in Q$, so that $x^{(j)} < x^{(i)}$.

Such solutions in the objective (or phenotypic) space are called non-dominated solutions [35].

Dominated and Non-dominated solutions for a hypothetical minimization problem with two conflicting objectives are graphically depicted on Figure 1. As shown, the ideal point for the conflicting objectives is considered "unachievable". A search for Pareto front, however, should satisfy two important properties; diversity and convergence. A diverse solution has the largest spread to cover the complete range of objectives and at the same time, it should have well-spaced solutions. Convergence of the search, on the other hand, is a practical guarantee of achieving the Pareto front.

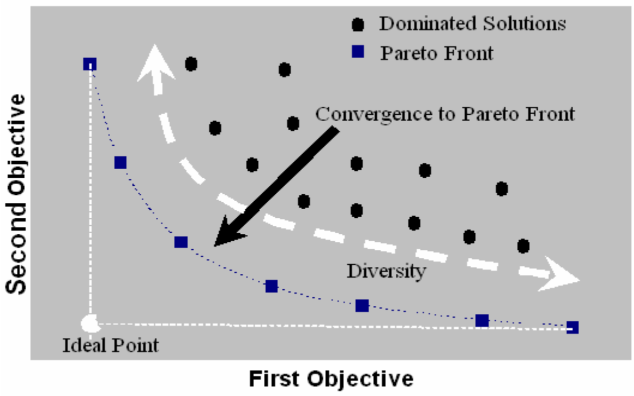

Fig. 1. Pareto front for minimization of two objectives

A problem that all stochastic search methods (such as NSGA-II) face is that they rely on their large population sizes and a large generation number, which can prove to be ineffective for real world situations where objective function evaluations are time consuming and computationally intensive. For these problems, increasing the efficiency of search and reduction of iterations for objective function evaluations can prove to be very efficient. In recent years there have been researches to address the efficiency of the optimization for computationally intensive problems [36, 37]).

3. SELF ORGANIZING MAPS

Self Organizing Map (SOM) was originally developed by Kohonen is as an ANN in order to study the self organization processes in the brain [38]. It is an unsupervised ANN, which consists of an input and an output layer. Figure 2 shows the topology of a SOM. The input vectors are presented to SOM iteratively; SOM extracts their distinguishable features and adjusts the weights of the network to recognize the same features in future. In other words, SOM attempts to map input patterns to nodes such that the nearness or neighborhood relations between the population members (topology) are preserved. Similar patterns are then mapped to nearby neurons [39]. The unique feature of a SOM network is its capability to classify the inputs according to their similarity, without any information regarding the grouping or number of partitioning.

The learning process in a SOM can be summarized as [39]:

1. Assign random values to the network weights, w_{ij}.
2. Present an input pattern or vector, x, to the network.
3. Calculate the distance between pattern x and each weight vector $\mathbf{w_j}$, and identify the winning vector, or

$$d = min \,||\, \mathbf{x} - \mathbf{w_j} \,|| \qquad j=1,2,3,\ldots\, n \tag{3}$$

Where $\|\ \_\ \|$ is the Euclidean norm and w_j is the weight vector of each neuron.

4. Adjust all weights in the neighborhood of the winning neuron, by:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)k(j,t)\{x_i - w_{ij}(t)\} \tag{4}$$

Where $\eta(t)$ is the learning rate at epoch t; and $k(j,t)$ is a suitable neighborhood function.

5. Repeat steps (2) to (4) until a convergence criterion is satisfied.

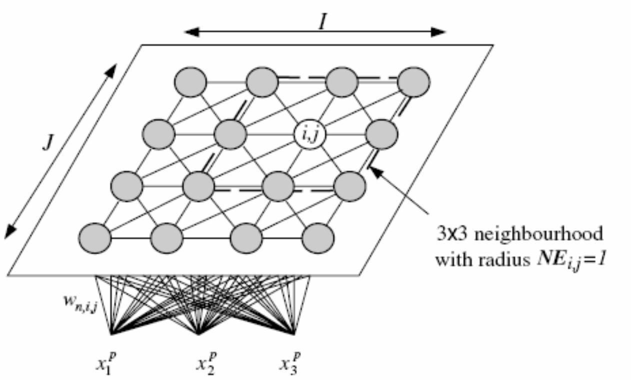

Fig. 2. Schematic representation of an SOM [28]

SOM may be viewed as a technique that clusters input data to similar individuals which, in turn, may be utilized to create virtual islands from a large population size in the exploitation phase of an NSGA-II approach. The clustering technique and evolution of similar individuals are utilized in the proposed method to speed up convergence and increase the performance of multi-objective optimization via NSGA-II.

4. THE PROPOSED HYBRID SOM-NSGA-II

In the exploitation phase of GA, combination of extremely different individuals is not likely to produce superior offspring. The competition to excel at the given region or neighborhood of the Pareto front requires specialized genes. The crossover between drastically different individuals, especially at later generations, is likely to produce unfit or even unfeasible individuals or so-called lethal offspring. Therefore, SOM is applied to identify similarities, either genotypic of phenotypic, and use inbreeding to speed up convergence to the Pareto front. The proposed method has two distinct phases. The exploration phase which has the usual selection, cross over, and mutation operators as NSGA-II, and an exploitation phase which utilizes SOM to find similar individuals and form virtual islands or subpopulations. The process of making these subpopulations may be done by using either the similarity in parameter space, also known as the genotype space, or by the use of objective space also known as phenotype space. Phenotypic features are caused by variation in the genotype space. The variable or genotype space usually consists of numerous variables with different structures such that understanding genotypic features via formation of subpopulations would produce very useful information about the role of variables.

The second phase applies a localized evolution or inbreeding of the subpopulations. These islands are handled separately and for a pre-specified number of generations (inbreeding period), and the islands evolve on their own. The subpopulations are merged into one population and the resulting population evolves for one generation. Merging subpopulations in regular intervals provides an opportunity for genetically different offsprings to diffuse or migrate to their more related subpopulations at subsequent generations and provides an overall diversity enforcement mechanism for the population. The method can reduce the production of lethal offering, because the genetic variety inside islands is not large. The offspring have more resemblance to their parents through local selection.

The stepwise algorithm of the method is given below:

1. Initialize map size, map type (genotypic or phenotypic) and chromosomes (population) and evaluate objective functions.
2. Evolve the population for a given number of generations (Exploration Phase):
   chromosomes ← NSGA-II (chromosomes, generations)
3. Make a genotypic or phenotypic SOM of the population.
   sMap ← CreateSOMMap (chromosome, MapSize)
4. Divide the population into N virtual islands or subpopulations according to the resulted SOM.
   SPchromosmes *(k)* ← DividePopulation (sMap, chromosome) $k=1,2,3\ldots N$ subpopulation.
5. Evolve each island separately for inbreeding generations.
   SPchromosmes *(k)* ← NSGA-II (SPchromosomes *(k)*, InbreedingGenerations)
6. Merge islands and evolve the whole population for one generation
   chromosomes ← chromosome $\cup$ SPchromosmes *(j)* $j=1,2,3,\ldots N$
7. Evolve the chromosomes for one generation.
   chromosomes ← NSGA-II (chromosomes, generations = 1)
8. If the stopping criterion is not met, go to step 3, otherwise stop.

According to this algorithm, the updated chromosomes in each iteration eventually converge to the Pareto front. This method has the advantage of explicit parallelism; therefore, every island could be evolved separately with a parallel processor or with a serial implementation. The preferred topology of SOM is the hexagonal lattice, which has the greatest number of neighbors for a given neuron in SOM.

From a general perspective, the proposed method aims at increasing the efficiency of large WDNs optimization through a reduction of function evaluations and a localized search of the Pareto fronts. In this paper, the proposed method and original NSGA-II are applied to two simple combinatorial problems of WDN design optimization and the results are compared. Since NSGA-II optimization problems would involve extended period simulations and evaluation of each network with respect to many objectives, one may expect that the proposed method would converge faster; a factor that would be very helpful, specially in real world (large WDNs) optimization problems. It should be noted that the approach is different from that of Villmann *et al.* [30] in the sense that their method utilized SOM for a single objective optimization, while the present approach utilizes SOM to find subpopulations in multi-objective optimizations in order to converge faster to the Pareto front. The method has an inbreeding generation after each subpopulation is evolved separately, as a controlling step for the overall spread check and suitable spacing of the solutions. Further, the use of genotypic sorting and visualization is a novel approach for understanding the behavior of many similar solutions with the study of representative of the corresponding subpopulation.

In order to compare the performance of multi-objective methods, the Generational Distance (GD) is used [40]. GD is a measure of how far the elements in the set of non-dominated vectors are from those in the Pareto-optimal set and is defined as:

$$GD = \frac{\sqrt{\sum_{i=1}^{n} d_i^2}}{n} \tag{5}$$

Where n is the number of vectors in the set of non-dominated solutions found so far, and d_i is the Euclidean distance (measured in objective space) between each vector and the nearest member of the Pareto-optimal set. Therefore, a value of $GD = 0$ indicates that all generated elements are in the Pareto-optimal set, and any other value indicates how far obtained solutions are from the global Pareto front of the problem.

5. WATER DISTRIBUTION NETWORK OPTIMIZATION

WDN optimization is a nonlinear and complex multi-objective optimization with feasibility constraints involving a number of engineering issues, such as multiple operating conditions (often under uncertainty), reliability, redundancy, resilience, and time scheduling of investments, thus requiring a multi-objective optimization approach [21].

a) WDN modeling and simulation

Mathematical model of a WDN consists of equations of head loss or energy conservation, and flow continuity or conservation of mass. Assuming a network of ***nn*** number of nodes and ***nf*** number of fixed grade nodes (tanks and reservoirs), flow head-loss relation between any node ***i*** and ***j*** connected by a pipe may be written as [33]:

$$\boldsymbol{H_i - H_j = h_{ij} = rQ_{ij}^n + mQ_{ij}^2} \tag{6}$$

Where $\boldsymbol{H}$ is nodal head, $\boldsymbol{h}$ is head loss, $\boldsymbol{r}$ is resistance coefficient, $\boldsymbol{Q}$ is flow rate, $\boldsymbol{n}$ is flow exponent and $\boldsymbol{m}$ is minor loss coefficient. The second set of equations, flow continuity at any node $\boldsymbol{i}$, that has to be satisfied may be expressed as [33]:

$$\boldsymbol{\Sigma Q_{ij} - D_i = 0} \quad \textit{for} \quad \boldsymbol{i = 1,\ldots,nn} \tag{7}$$

Where $\boldsymbol{D_i}$ is flow demand at node $\boldsymbol{i}$ and the summation of inflow to node $\boldsymbol{i}$, $\boldsymbol{\Sigma Q_{ij}}$, is over $\boldsymbol{j}$ number of nodes that are connected to node $\boldsymbol{i}$ by a pipe. The method proposed by Todini and Pilati (1987) has been used in *EPANET* to solve the above equations and obtain a static solution for the network [33].

b) Feasibility constraint

Usually, WDN solutions are considered feasible if they satisfy a minimum required head (H^l) at every node. This minimum head requirement is essential to the accuracy of the pressure dependent analysis of the WDN [33]. Utilization of the pressure dependent analysis is not needed here due to the application of the feasibility constraint [41]. From a degree of constraint violation perspective, solutions may be ranked with respect to each other. The constraint violation for any solution can be calculated using a failure index as [42, 43]:

$$I_f = \frac{\sum_{j=1}^{nn} e_j}{\sum_{k=1}^{nr} Q_k H_k + \sum_{i=1}^{npu} P_i / \gamma} \tag{8}$$

Where $e_j = 0$ if $H_j \geq H^l_j$ and $e_j = Q_j\,(H^l - H_j)$ otherwise. A solution is considered feasible when $I_f = 0$ and infeasible otherwise.

In this study, a method is used to compare solutions with different I_f values that does not require penalty coefficients. The method was first introduced by Deb and Agrawal [42]. According to this method, a solution $x(i)$ is constraint-dominating another solution $x(j)$, if any of the following is true:

1. Solution $x(i)$ is feasible and solution $x(j)$ is infeasible,
2. Solutions $x(i)$ and $x(j)$ are both infeasible, but $x(i)$ has a smaller I_f value, or
3. Solutions $x(i)$ and $x(j)$ are both feasible and solution i dominates solution j.

This method has the advantage of being incorporable into a non-domination ranking for any number of constraints, while not requiring any penalty function.

c) Cost objective

In real world situations, the cost of constructing a new WDN may be estimated, with a reasonable accuracy, by considering pertinent parameters like material and labor costs, earth work, equipment, energy consumption, maintenance, and other practical considerations. However, for simplicity, the cost of constructing a new WDN in benchmark problems is usually given by simple equations or tables that relate cost to pipe diameters and length only [30,32] .

d) Reliability objective

Low quality or lack of service in a WDN would cause unpleasant damages to consumers and providers alike. Therefore, reliability of a network has become a major concern in recent optimization problems and neglecting its importance has been a major reason for the limited acceptance of single-objective optimization methods whereby the cost of a WDN is minimized [27]. Reliability is usually defined as the probability that a system will perform its mission within specified limits for a given period of time in a specified environment. For a large system with many interactive subsystems (such as a water distribution system), it is extremely difficult to compute the mathematical reliability analytically. Accurate calculation of a mathematical reliability requires knowledge of the precise reliability of the basic subsystems or components and the impact on accomplishing the mission caused by the set of all possible subsystem (component) failures [1]. Researchers have suggested and used surrogate or indirect methods (indices) to represent the reliability of WDNs.

1- Minimum surplus head index (I_m): The surplus head at a node is equal to the difference between the head H at which the demand Q is supplied, and the minimum required head or design head H^l at that node. This surplus head indicates the available energy for dissipation during failure conditions. The minimum surplus head index I_m is defined as:

$$I_m = \min\{H_j - H_j^l\} \qquad \text{for } j = 1,2,...,nn \tag{9}$$

It is believed that maximization of the available surplus head at the most depressed node improves the reliability of a network [43, 44].

2- Total surplus head index (I_t): Another index that can be used to reflect the reliability of a network is the summation of surplus head at each node. In a mathematical form, the total surplus head index, I_t, may be expressed as:

$$I_t = \Sigma\{H_j - H_j^l\} \qquad \text{for} \qquad j=1,2,...,nn \tag{10}$$

Maximization of I_t improves the system reliability or capability of the network to adjust under stressed conditions [43, 44].

3- Resilience index (I_r): A more practical reliability index often used by researchers [20, 21, 22] was proposed by Todini [43]. It is based on the concept that the power input into a network is equal to the power lost internally (to overcome the friction) plus the available power at demand points:

$$P_{inp} = P_{int} + P_{out} \tag{11}$$

The total input power into a network including power supplied by pumps is given by

$$P_{inp} = \gamma \Sigma Q_k H_k + \Sigma P_i \tag{12}$$

Where Q_k and H_k are discharge and head corresponding to each reservoir node k; P_i is power supplied by pump i, summed over all pumps. Total output power or consumed power in the network is given by:

$$P_{out} = \gamma \Sigma Q_j H_j \tag{13}$$

Where Q_j is demand at node j, and H_j is head at which Q_j is supplied.
The resilience index of a network is then defined as;

$$I_r = \frac{\sum_{j=1}^{nn} Q_j (H_j - H_j^l)}{(\sum_{k=1}^{nf} Q_k H_k + \sum_{i=1}^{npu} P_i / \gamma) - \sum_{j=1}^{nn} Q_j H^l{}_j} \tag{14}$$

Where nn is number of nodes, nf is number of reservoirs, and npu is number of pumps in the network [43, 44].

In a multi-objective WDN optimization problem, networks are sought that satisfy the aforementioned feasibility constraint and are non-dominated (Pareto front) with respect to cost and reliability objectives. In order to prove the effectiveness of the proposed optimization method, two typical WDN example problems and the widely used cost equations and resilience index (as a reliability measure) are considered in this study.

6. EXAMPLE PROBLEMS

a) Example 1: The two looped network

The first example problem solved is the problem presented originally by Alperovits and Shamir [2], whose layout, nodal demands, and pipe lengths are shown in Fig. 3. The network consists of eight 1000 *m* long pipes, a Hazen-Williams *C* value of 130, seven nodes, and a single reservoir. The minimum pressure requirement at every node is 30 *m*. Available pipe diameters and their costs are shown in Table 1.

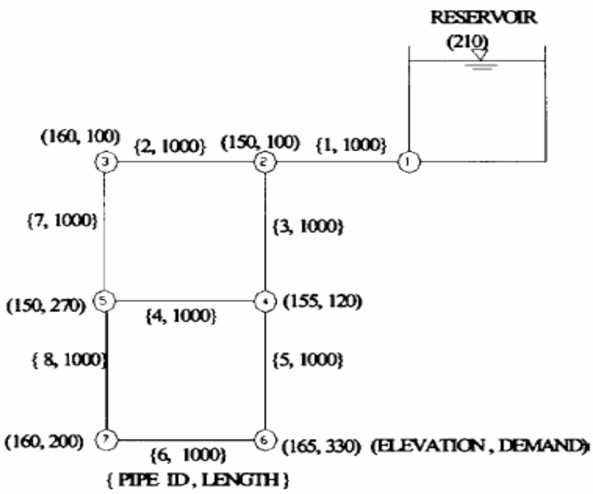

Fig. 3. The two looped network topology [18]

Table 1. Available pipe diameters and their costs

Pipe diameter(in)	1	2	3	4	6	8	10	12	14	16	18	20	22	24
Pipe diameter(mm)	25.4	50.8	76.2	101.6	152.4	203.2	254	304.8	355.6	406.4	457.2	508	558.8	609.6
Cost per unit length($)	2	5	8	11	16	23	32	50	60	90	130	170	300	550

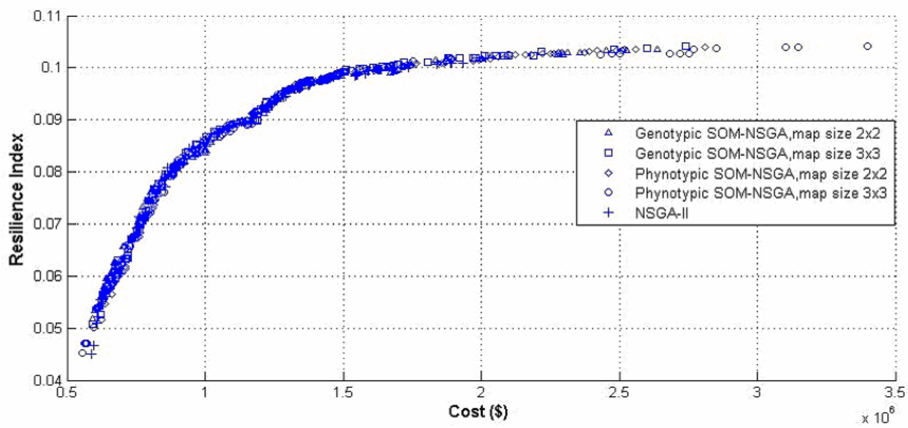

Fig. 4. Pareto front plot for SOM-NSGA-II and NSGA-II for the two looped example

The solution space of two looped network is relatively small and optimization is not particularly challenging. However, as it can be seen from the Pareto front plot (Fig. 4), the range of the Pareto front discovered by the proposed SOM-NSGA-II is larger. Particularly, the Pareto front is more diverse and includes more solutions at the lower cost networks. It was concluded that SOM-NSGA-II has been more successful in finding a diverse range for the Pareto front. For this optimization, a SOM of different map sizes with phenotypic and genotypic inbreeding was used. Every island was evolved for three inbreeding generations and the overall generations for both methods were 100.

Generational Distance, shown in Table (2), is calculated using the distance of individual Pareto front members of each method with the Pareto front of all methods together. The objective values are normalized with the maximum and minimum values from the ensemble of all methods. As reflected in the smaller GD value, the phenotypic method is more successful in finding a better Pareto front with respect to the genotypic method. In general, a larger map size produced better results in both the genotypic and phenotypic approaches, however, the map size in this example was kept in a reasonable range of 2x2 and 3x3. Comparison of GD results for these optimum map sizes in Table 2 shows a smaller GD (i.e. a better performance) for the larger map size.

Table 2. Normalized Generational Distance (GD) for the two looped network

	Genotypic SOM	Phenotypic SOM
SOM-NSGA-map size 2x2	1.86	0.82
SOM-NSGA-map size 3x3	1.25	0.74
NSGA-II	1.52	1.52

b) Example 2: Hanoi WDN

Hanoi water distribution network has 32 nodes and 34 pipes organized in 3 loops [45]. In this benchmark problem there is only a single fixed head source at elevation 100 *m* and no pumping facilities are considered. Minimum head requirement at all nodes is fixed at 30 *m* and commercially available pipe diameters are 12, 16, 20, 24, 30, and 40 inches [14]. Construction cost for a new network is assumed to be a non-linear function of pipe diameter and length,

$$C_{ij} = 1.1\, D_{ij}^{1.5}\, L_{ij} \tag{15}$$

Where C is cost in dollars, D is diameter in inches, and L is length of pipe in meters. A schematic view of Hanoi WDN is shown in Fig. 5. In this example, 3 inbreeding generations for SOM-NSGA-II and 200 generations for both methods were used.

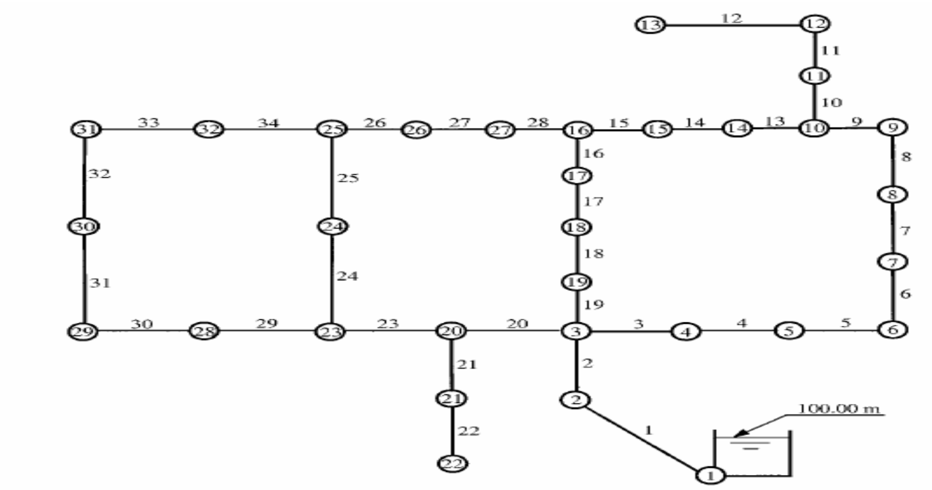

Fig. 5. Hanoi WDN layout [45]

Figures 6 and 7 show the Pareto front discovery by genotypic and phenotypic approaches, respectively. As shown on the figures, both approaches have resulted in very similar satisfactory resilience indexes for reasonable costs at all map sizes. However, the two approaches had different GDs.

GD was calculated using the same method of normalization of values with the maximum and minimum of all methods together (Table 3). As in the previous example, the phenotypic approach performed better than the genotypic approach and standard NSGA-II in all map sizes as evidenced by smaller GD values in Table 3. Although in this example increasing the map size decreased the SOM clustering error, it did not show a consistent effect in minimizing GD. Further increase of the map size (beyond 5x5) did not improve SOM clustering performance as more cells remained empty. In general, the enhanced performance of optimization in SOM-NSGA-II is mainly due to the clustering effect of SOM applied to the population.

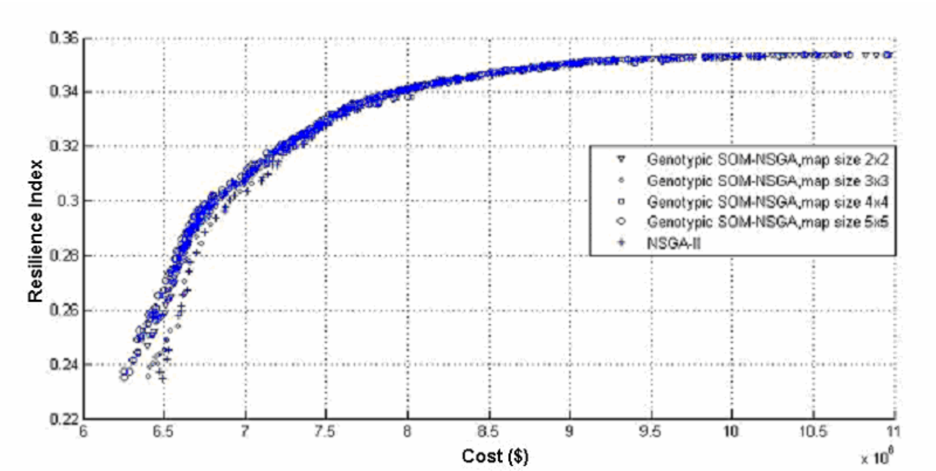

Fig. 6. Pareto front plot for genotypic SOM-NSGA-II and NSGA-II for Hanoi WDN

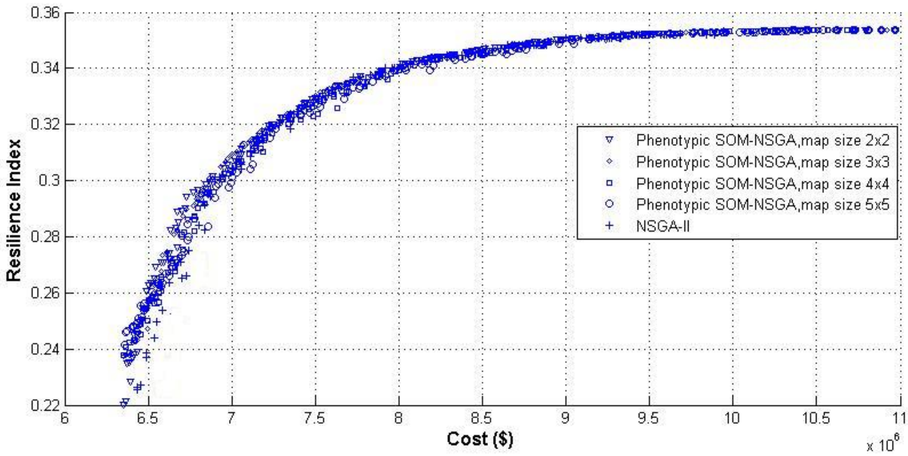

Fig. 7. Pareto front plot for phenotypic SOM-NSGA-II and NSGA-II for Hanoi WDN

Table 3. Normalized Generational Distance for Hanoi WDN

	Genotypic SOM	Phenotypic SOM
SOM-NSGA-map size 2x2	1.90	0.75
SOM-NSGA-map size 3x3	2.88	2.18
SOM-NSGA-map size 4x4	2.14	0.46
SOM-NSGA-map size 5x5	2.08	0.65
NSGA-II	2.40	2.40

Figure 8 shows two visual representations of the entire population. The unsorted genetic structure of the population from NSGA-II is shown on the left hand side of the figure and the genotypic sorted population shown on the right hand side. As clearly depicted on this figure, the unsorted population is not coherent, nor does it have any structure or pattern. On the contrary, the sorted population is much better structured and similar individuals are recognizable in the four groups of subpopulations.

The structures of subpopulations are depicted in Fig. 9. In this figure, all 200 solutions in the Pareto front are categorized in 4 similar subpopulations with 70, 30, 70, and 30 solutions in each subpopulation, respectively. The color coding corresponds to the pipe sizes (*mm*) that are allocated for every link (out of 34). For example, some lighter colored pipes (pipes 1 to 5) do not change color in all Pareto front solutions, which reflects their importance in the network. A feasible Pareto front network requires high diameters for such pipes, and hence, a limited pressure drop there. In other words, pressure drop in such pipes has a significant importance to the pressure distribution in the entire network. On the other hand, there are pipes (pipes 30 to 34) that take up different colors in different Pareto front solutions. Pressure drop in such pipes is not crucial to the feasibility of the network.

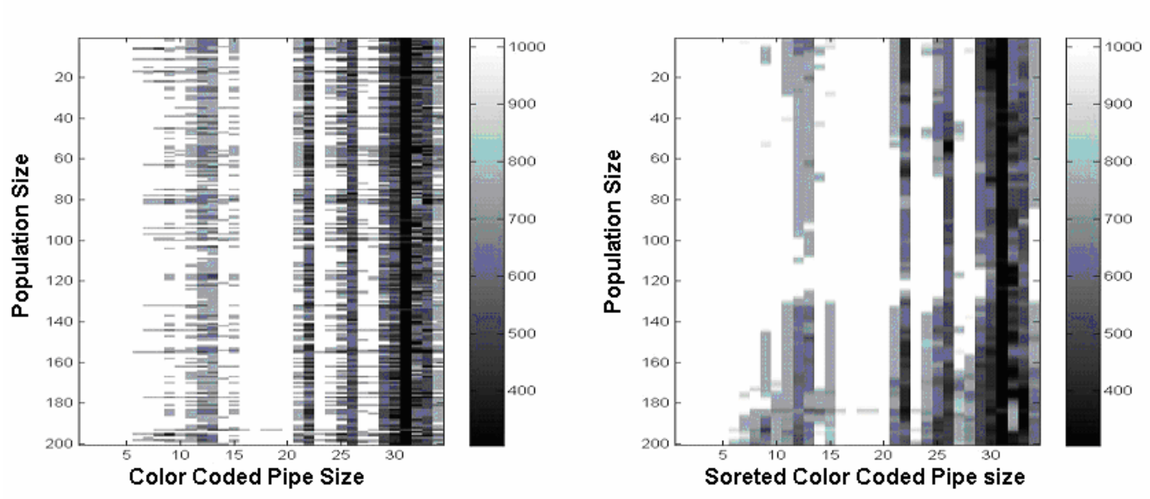

Fig. 8. Unsorted (left) and genotypic sorted (right) color coded population of Hanoi example problem (pipe diameters in *mm*)

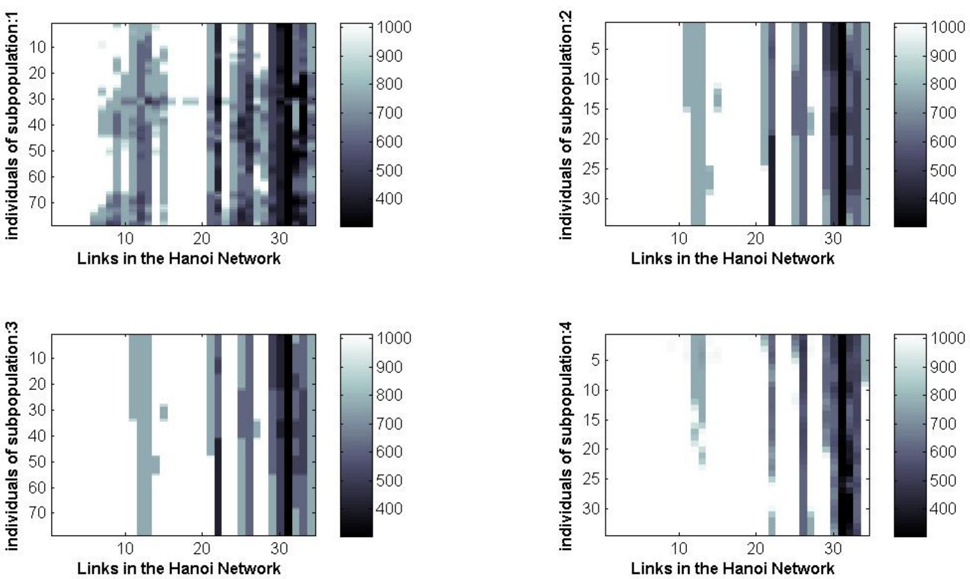

Fig. 9. Color coded genotypic diversity of Pareto front for Hanoi WDN (pipe diameters in *mm*)

7. CONCLUSION

A hybrid method, SOM-NSGA-II, was presented for multi-objective optimization of combinatorial WDNs. In this method, SOM was incorporated into NSGA-II and applied to two WDN example problems. SOM successfully discovered similar individuals in the population and formed virtual islands or subpopulations. The crossover operation within subpopulations at the constraint dominated region of the solution space resulted in a faster convergence (in Hanoi WDN) and a wider Pareto front (in the two looped WDN). One would expect the method to be more computationally effective than standard NSGA-II for realistic optimization problems, where the computations are more intense than simple example problems. From a fundamental perspective, SOM incorporation pushed the method closer to the natural evolution as crossover predominantly happens within similar individuals or niches in the nature. An added advantage of the method is the usage of genotypic sorting of the population by SOM for visual representation of the structure of the Pareto front. The resulted genotypic maps showed the relative importance of pipe diameters and their variation extent in the Pareto front. This can prove to be a very important visual aid for decision makers and designers of WDNs.

REFERENCES

1. Mays, L. W. (2000). *Water distribution systems handbook*. McGraw-Hill. New York.
2. Shamir, A. E. (1977). Design of optimal water distribution systems. *Water Resources Research*. Vol. 13, No. 6, pp. 885-900.
3. Gessler, J. (1985). Pipe network optimization by enumeration. *Proc. Computer Applications in Water Resources*, ASCE, Buffalo, New York, pp. 572-581.
4. Chiplunkar, A. V., Mehndiratta, S. L. & Khanna, P. (1986). Looped water distribution system optimization for a single loading. *J. Environ. Eng*., Vol. 112, No. 2, pp. 264-279.
5. Savic, D. A. & Walters, G. A. (1997). Genetic algorithms for least-cost design of water distribution networks. *J. Water Resour. Plan. Manage*., Vol. 123, No. 2. pp. 67-77.
6. Cunha, M. D. C. & Sousa, J. (1999). Water distribution network design optimization: Simulated annealing approach. *J. Water Resour. Plan. Manage*., Vol. 125, No. 4, pp. 215-221.
7. Liong, S. Y. & Atiquzzaman, M. (2004). Optimal design of water distribution network using shuffled complex evolution. *J. Inst. Eng*. Singapore, Vol. 44, No. 1, pp. 93-107.
8. Maier, H. (2003). Ant colony optimization for design of water distribution systems. *J. Water Resour. Plan. Manage.*, Vol. 129, No. 3, pp. 200-209.
9. Afshar, M. H. (2007). Application of Ant Algorithm to Pipe Network Optimization. *Iranian Journal of Science & Technology, Transaction B: Engineering*, Vol. 31, No. B5, pp. 487-500.
10. López-Ibáñez, M., Devi Prasad, T. & Paechter. (2008). Ant colony optimization for optimal control of pumps in water distribution networks. *J. Water Resour. Plan. Manage.* Vol. 134, No. 4, pp. 337-346.
11. Zecchin, A. C., Maier, H. R., Simpson, A. R., Nixon, L & Nixon, J. B. (2007). Ant Colony optimization applied to water distribution system design: Comparative study of five algorithms. *J. Water Resour. Plan. Manage.*, Vol. 133, No. 1, pp. 87-92.
12. Wu, Z. Y. & Walski, T. M. (2004). Self-adaptive penalty cost for optimal design of water distribution systems. *ASCE EWRI Conf*., Salt lake city.
13. Walaski, T. M. (2001). The wrong paradigm-why water distribution optimization does not work. *J. Water Resour. Plan. Manage*., Vol. 129, No. 2, pp. 203-295.
14. Farmani, R., Walters, G. A. & Savic, D. A. (2005). Trade off between total cost and reliablity for anytown water distribution network. *J. Water Resour. Plan. Manage*. Vol. 131, No. 3, pp. 161-171.
15. Vamvakeridou-Lyroudia, L. S., Walters, G. A. & Savic, D. A. (2005). Fuzzy multiobjective optimization of water distribution networks. *J. Water Resour. Plan. Manage*. Vol. 131, No. 6, pp. 467-475.
16. Prasad, T. D. & Park, N. S. (2004). Multiobjective genetic algorithms for design of water distribution networks. *J. Water Resour. Plan. Manage*., Vol. 130, No. 1, pp. 73-82.
17. Deb, K., Agrawal, S., Pratap, A. & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Proc., Parallel Problem Solving from Nature*, *VI Conf*., pp. 849-858.
18. Atiquzzaman, M., Liong, S. Y. & Xinying, Yu. (2006). Alternative decision making in water distribution network with NSGA-II. *J. Water Resour. Plan. Manage*., Vol. 132, No. 2, pp. 122-128.
19. Kadu, M. S., Gupta, R. & Bhave, P. R. (2008). Optimal design of water networks using a modified genetic algorithm with reduction in search space. *J. Water Resour. Plan. Manage.* Vol. 134, No. 2, pp. 147-160.

20. Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. & Faloutsos, C. (2008), Efficient Sensor Placement Optimization for Securing Large Water Distribution Networks*. J. Water Resour. Plan. Manage.* Vol. 134. No. 6, pp. 516-526
21. Shamir, U. & Salomons, E. (2008). Optimal real-time operation of urban water distribution systems using reduced models. *J. Water Resour. Plan. Manage.*Vol. 134, No. 3, pp. 181-185
22. Alvisiand, S. & Franchini, M. (2009) Multiobjective optimization of rehabilitation and leakage detection scheduling in water distribution systems. *J. Water Resour. Plan. Manage*., Vol. 135, No. 6, pp. 426-439
23. Giustolisi, O., Laucelli, D. & Colombo, A. F. (2009), Deterministic versus stochastic design of water distribution networks. *J. Water Resour. Plan. Manage.*Vol. 135, No. 2, pp. 117-127.
24. Nicolini, M. & Zovatto, L. (2009). Optimal location and control of pressure reducing valves in water networks*. J. Water Resour. Plan. Manage.,* Vol. 135, No. 3, pp. 178-187
25. Broad, D. R., Dandy, G. C. & Maier, H. R. (2005). Water Distribution system optimization using metamodels. *J. Water Resour. Plan. Manage*, Vol. 131, No. 3, pp.172-180.
26. Whitley, D., Starkweather, T. & Bogart, C. (1990). Genetic algorithms and neural networks: optimizing connections and connectivity*. Parallel Computing*. Vol. 14, No.3, pp. 347-361.
27. Yao, X. (1993). A review of evolutionary artificial neural networks. *Int. J. Intell. Systems,* Vol. 8, No.4, pp. 539-567.
28. Poogaard, E. P. & Mynett, A. E. (1998). Self organizing feature maps for the analysis of hydrological and ecological data. *Proc. Hydroinformatics*. Vol. 1, No. 1, pp.733-740.
29. Nouruzi, K. & Rakhshandehroo, G. R. (2007). Most probable pollution source identification in rivers by neural networks. *Iranian Journal of Science & Technology, Transaction B: Engineering*, Vol. 31, No. B5, pp. 501-508.
30. Villmann, T., Villmann, B. & Slowikb, V. (2004). Evolutionary algorithms with neighborhood cooperativeness according to neural maps. *New Aspects in Neurocomputing: 10th European Symposium on Artificial Neural Networks*, Vol. 57, pp. 151-169.
31. Pierro, F., Khu, S. T., Savic, D. A. & Berardi, L. (2009). Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms. *Environmental Modeling & Software.,* Vol. 24, No.2, pp. 202-213.
32. Rossman, L. A. (2000). *EPANET 2 user's manual*. EPA. National Risk Management Lab. Cincinnati. OH. 45267.
33. Gen, M. & Cheng, R. (2000). *Genetic algorithms and engineering optimization*. John Wiley & Sons, Inc., New York.
34. Coello Coello, C. A. & Lamont, G. B. (2007). *Applications of multi-objective evolutionary algorithms*. Vol. 1, World Scientific Publishing, Singapore.
35. Deb, K., Agrawal, S. & Pratap, A. (2002). A fast and elitist multio-bjective genetic algorithm for multi-objective optimization: NSGAII. *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp. 182-197.
36. Jourdan, L., Corne, D. W., Savic, D. A. & Walters, G. (2006). LEMMO: hybridizing rule induction and NSGA II for multi-objective water systems design. *Proc. of the Eighth International Conference on Computing and Control for the Water Industry*. Vol. 2, pp. 45-50.
37. Knowles, J. (2006). ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Transactions on Evolutionary Computation*. Vol. 10, No. 1, pp. 50-66.
38. Kohonen, T. (1990). The self organizing map. *Proceeding of IEEE*, Vol. 78, No. 9, pp. 1464-1480.
39. Bowden, G. J., Dandy, G. C. & Maier, H. R. (2004). Input determination for neural network models in water resources applications. Part 1 - background and methodology. *Journal of Hydrology*, Vol. 301, No. 1 ,pp. 75-92.

40. Abraham, A, Jain, L. & Goldberg, R. (2005). Evolutionary multiobjective optimization. Springer-Verlog London Limited.
41. Tabesh, M. & Dolatkhahi, A. (2006). Effects of pressure dependent analysis on quality performance assessment of water distribution networks. *Iranian Journal of Science & Technology, Transaction B: Engineering*, Vol. 30, No. B1, pp.119-128.
42. Deb, K. & Agrawal, S. (1999). A niched-penalty approach for constraint handling in genetic algorithms. *Proc., ICANNGA-99*, *Portoroz, Slovenia*, pp. 123-135.
43. Todini, E. (2000). Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water*, Vol. 2, No. 2, pp. 115-122.
44. Parasad, T. D. & Park, N. S. (2004). Multiobjective genetic algorithms for design of water distribution networks. *J. Water Resour. Plan. Manage*., Vol. 130, No. 1, pp. 73-81.
45. http://centres.exeter.ac.uk/cws/benchmarks/layout/37-hanoi-water-distribution-network (accessed 2010, 24 January).