Department of Civil and Environmental Engineering
University of Strathclyde

MAXIMUM ENTROPY BASED EVOLUTIONARY OPTIMIZATION OF WATER DISTRIBUTION NETWORKS UNDER MULTIPLE OPERATING CONDITIONS AND SELF-ADAPTIVE SEARCH SPACE REDUCTION METHOD

Thesis submitted in accordance with the requirements of the
University of Strathclyde
for the degree of

Doctor in Philosophy

by

ANNA MARZENA CZAJKOWSKA

March 2016

Signed:

Date:

MAXIMUM ENTROPY BASED EVOLUTIONARY OPTIMIZATION OF WATER DISTRIBUTION NETWORKS UNDER MULTIPLE OPERATING CONDITIONS AND SELF-ADAPTIVE SEARCH SPACE REDUCTION METHOD

Anna Marzena Czajkowska

ABSTRACT

One of the complexities in designing WDN is evaluation of network performance. The accurate network performance measures such as reliability or failure tolerance are very time consuming calculations, thus surrogate measures are used for water distribution network (WDN) design optimization. Entropy is particularly advantageous since it involves only the flow in the pipe and the demands at the nodes.

This thesis developed efficient new computational methods based on the maximum entropy formalism for the optimization of water distribution systems. Thus the maximum entropy based design approach has been extended here to include multiple operation conditions. Also, the path-related properties of the flow entropy have been exploited to develop a new self-adaptive approach for solution space reduction in multiobjective evolutionary optimization of water distribution systems that resulted in a significant reduction in the number of function evaluations required to find optimal and near optimal solutions.

The novelty and originality of the current research are presented next.

A new penalty-free multi-objective evolutionary optimization approach for the design of WDNs has been developed. It combines genetic algorithm with least cost design and maximum entropy. The approach can handle single operating conditions (SOC) as well as multiple operating conditions (MOC) for any given network. Previously, most of the work has been done for single loading patterns and it was assumed that nodal demands are constant. In reality nodal demand vary over the time so network designed to satisfy one operating condition might not be able to satisfy other loading patterns (i.e. pressure constraints might not be meet). The model has been applied to three well known water distribution networks. The approach has also been implemented on a large real-world network in the literature. Three different methods of designing for multiple loading patterns were investigated. Extensive testing proved that MOC outperform SOC in terms of hydraulic feasibility, pipe size distribution and reliability. The approach is computationally efficient and robust.

The above mentioned penalty-free approach has been extended to form a module that would improve the convergence criteria of the GA by reducing its search space. For large real-world network GA might require extremely large number of function evaluations which could lead to delayed convergence. By reducing the search space, the GA's effectiveness and efficiency will increase as the algorithm will identify the solutions in smaller number of function evaluations. The search space reduction method presented herein is based on entropy and uses the importance of every path through network, which is an inherent property of the entropy function. The developed algorithm is dynamic, self-adaptive and does not require pre-defining the reduced sets of candidate diameters for each pipe. The method has been applied to a large network from the literature. Two cases were studied, one based on full search space and one for reduce search space (RSS) approach. Rapid stabilization was observed for the results obtained using RSS.

ACKNOWLEDGEMENT

I would like to express my special appreciation and thanks to my supervisor, Dr. Tiku Tanyimboh, for his constant guidance and for being a tremendous mentor for me through the course of the PhD program. For his unwavering support, invaluable advice and constructive criticism, I am truly grateful. Without his persistent help, motivation and encouragement this thesis would not have been possible. I also would like to thank Dr. Stella Pytharouli who is acting as my second supervisor for her help and support.

Additionally, I would like to thank the UK Engineering and Physical Sciences Research under Grant Number EP/G055564/1 for funding my PhD programme.

My most sincere and deepest gratitude goes to my beloved mother, Danuta Kucala, father, Edward Kucala, sister, Iwona Kucala-Luniewicz and my parents-in-law, Krystyna Czajkowska and Marek Czajkowski for their help with my daughters and continuous encouragement to complete the PhD. In a special way I would like to thank to my husband, Michal Czajkowski for his moral support, encouragement and faith in me. Also, I would like to thank my daughters, Magdalena and Katarzyna for bringing the happiness and excitement into my life.

Finally, I would like to thank my friends, Calvin Siew, Salah Saleh, Euan Barlow, Upaka Rathnayake and most of all Alemtsehay Seyoum for creating such relaxing and helpful working environment which made my studies even more enjoyable.

DEDICATION

This thesis is dedicated to my mother, Danuta Kucala, father, Edward Kucala, my husband Michal Czajkowski and my daughters, Magdalena and Katarzyna.

TABLE OF CONTENT

LIST OF FIGURES

Chapter Four

Chapter Five

Appendix D

LIST OF TABLES

Appendix B

Appendix C

Appendix D

NOTATION

ΔQ	loop-flow correction
$\Delta Q_l^{(k)}$	loop-flow correction at iteration k to be applied for all pipes belonging to loop l
α_i	parameter to be calibrated in Logit function
β_i	parameter to be calibrated in Logit function
η_{ChM}	dimensionless conversion factor for Chezy-Manning head loss equation
η_{HW}	dimensionless conversion factors for Hazen-Williams head loss equation
ρ	density of water
γ	specific weight of water
a	flow exponent
a_m	probability that link m is in service
b_i	coefficient to be calibrated for node i in Germanopoulos's outflow equation
c_i	coefficient to be calibrated for node i in Germanopoulos's outflow equation
C_{ChM}	Chezy-Manning roughness coefficient
CEND	cost-entropy non-dominated
CFTND	cost-failure tolerance non-dominated
C_{HW}	Hazen-Williams roughness coefficient
C_i	cost of the pipe i
CRND	cost-reliability non-dominated
Cu_i	uniformity coefficient of node i.
D_m	diameter of pipe m
D_{ij}	diameter of pipe ij
DDA	demand driven analysis
DLL	dynamic link library
E	energy dissipated
EA	evolutionary algorithm

EPS	extended period simulation
ESP	entropy stagnation point
f	objective function
f_{ij}	Darcy-Weisbach friction factor in pipe ij
$\underline{F}$	vector of the function values
FEs	function evaluations
FSS	full search space
FT	failure tolerance
g	gravitational acceleration
GA	genetic algorithm
GGM	global gradient method
h_{ij}	head loss in pipe ij
$\underline{H}$	vector of nodal heads
HDA	head dependent analysis
H_i	actual head at node i
H_i^{des}	desired head at node i
H_i^{min}	nodal head at node i below which the outflow is zero
H_i^{req}	required head at node i for fully satisfying the demand
H_k	head of reservoir k
Hn_i	hydraulic gradient level at nodes i
Hn_j	hydraulic gradient level at nodes j
I	set of the source nodes
IJ	set of all links in path p
ILP	Integer-Linear Programming
J_H	Jacobian matrix for the unknown nodal heads
J_x	Jacobian matrix of F
K	arbitrary positive constant from entropy function
K_{ij}	pipe resistance coefficient for pipe ij
l_{ij}	number of loops sharing link ij
L_{ij}	length of pipe ij

M	number of links in the system
ME	maximum entropy
MERSS	maximum entropy reduced search space
MOC	multiple operating conditions
MOEA	multi-objective evolutionary algorithm
MRI	modified resilience index
n_e	exponent parameter which value can vary from 1.5 to 2
nf	flow exponent
n_n,	number of demand nodes
n_{pu}	number of pumps
n_r	number of reservoirs
N	number of nodes in a network
N	number of links in the network
Nj	nodes connected to node j
Nl	number of loops in the network
Nn	number of nodes
NP	non-deterministic polynomial-time
NR	network resilience
NSGA II	elitist non-dominated sorting genetic algorithm
OP	operating condition
$out(N_i)$	set of all pipes from node i
p_i	probability related to the ith event
p_m	probability that only link m is not in service
$p(m,n)$	probability that only links m and n are not in service
$p(0)$	the probability that all links are in service
PAES	Pareto-archive evolution strategy
POF	Pareto optimal front
P_i	fraction of total flow through the system that reach node i
P_j	power introduced to the network by pump j
Pr_i	available pressure at node i

Pr_i^*	pressure at which proportion of the required demand of node *i* is satisfied
q_{ij}	flow rate in link *ij*
Q_i	available outflow that can be delivered by the system at node *i*
Q_{ij}	pipe flow between node *i* and *j*
Q_{ij}^k	corrected flow rate
Q_{ij}^{k-1}	estimated flow rate
Q_{in}	inflow of the pipe
Q_i^{req}	require demand or supply at node *i*
Q_{i0}	demand at node *i*
Q_j	demand or supply at node *j*
Q_k	supply of reservoir *k*
Q_{max}	flow that gives the maximum hydraulic power at the outlet of the pipe
Q_{0i}	demand at node *i*
R	hydraulic reliability
R_i	resistance coefficient
RI	resilience index
RSS	reduced search space
s	surplus power factor
S	network entropy
S_i	entropy of node *i*
SOC	single operating condition
SPEA	strength Pareto evolutionary algorithm
SSR	search space reduction
SSS	steady state simulation
S_0	entropy of the source supplies
T	sum of the nodal demands
T_i	total flow reaching node *i*
$T(m)$	total flows supplied with only link *m* unavailable
$T(m,n)$	total flows supplied with only links *m* and *n* out of service
$T(0)$	total flows supplied with all links are in service

u_m	probability that link *m* is unavailable
VEGA	vector evaluated genetic algorithm
WDN	water distribution network
$\underline{x}$	vector of variables

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

Every human being has the right to have access to clean water as it is one of the basic components necessary for life. For that reason, the water distribution networks (WDNs) are extremely important infrastructure. In order to provide drinking water to each household many different factors needs to be satisfied. Delivered water has to be at adequate pressure, which is demand dependent and vary over time. Drinking water has to have suitable disinfection level, which is also dependent on water pressure and demand. Water storages needs to provide adequate capacity for firefighting purposes etc. All this makes designing WDN as a multi-criterion and highly complex problem, called as NP-hard problem (non-deterministic polynomial-time hard problem).

Evolutionary algorithms (EAs) have been used to solve complex, multi-objective problems for more than two decades. They are capable of finding pareto-optimal solutions in single run and proved to be well suited for such problems. However, the major disadvantage related to multi-objective EA is the inability to directly handle constraints. Widely used method for constraints handling is to penalize infeasible solutions, which could obstruct the search capabilities and direct to suboptimal solutions. The penalty parameters are usually obtained by a trial and error, require tunings and calibration, thus making the whole process extremely time consuming and case sensitive.

Apart from being highly complex to design, the WDNs are also one of the most expensive infrastructure systems. Therefore, the cost minimization is extremely important objective when designing the network. For decades researchers concentrated on minimization of the network cost, which has inevitable influence on network reliability (i.e. cheapest design is not satisfactory in terms of reliability). Nevertheless, the reliability of WDN is equally important as network cost, as the optimal design is the cheapest design that will also satisfy demands under normal and abnormal conditions. Since the estimation of reliability is extremely time-consuming calculation; researchers have been looking for surrogate measures, comparably easier to estimate. Statistical entropy (Tanyimboh and Templeman, 1993), resilience index (Todini, 2000), network resilience (Prasad and Park, 2004), pipe index vector (Vaabel et al. 2006) and modified resilience index (Jayaram and Srinivasan, 2008) are known as reliability indicator. Entropy is particularly advantageous since it involves only the flow in the pipe and the demands at the nodes that are normally given. Over the years, the entropy has been incorporated and tested on many different benchmark networks. Results published in literature suggest that an increase in entropy value corresponds to a better network performance as measured by reliability. Strong positive correlation between entropy and reliability has been demonstrated in many publications.

Water networks are very often designed as systems in which all components are working under normal conditions and are in service. It is also common practice to assume that nodal demands are constant and use steady state modelling when designing WDN. However in reality demands vary with the time of the day and there are many different loading patterns that have to be satisfied by the network. It has been pointed out by few researchers (Alperovits and Shamir, 1977; Prasad 2010) that not one, but few operating conditions should be considered. Moreover, it is already well known, that even if network will satisfy one loading patterns it does not mean that other operating condition will also be satisfied, simply because the pressure constraints may not be satisfied. Most of the work published so far on entropy has been done for single operating condition (SOC), usually maximum daily demand.

Another major problem related to genetic algorithms (GA) is the search space. The size of the search space is highly dependent on parameters such as number of links (i.e. size of the network) and number of commercially available pipe sizes. Therefore, the number of function evaluations required to identify optimal solution can be extremely large. For large, real life networks, with hundreds of pipes and many network components, it can be extremely time consuming process as the algorithm might require even millions of function evaluations. Few researchers made attempts for search space reduction, however none of published so far method would in reality be used for real size system with network components. Vairavamoorthy and Ali (2005) proposed method in which the importance of a pipe is established in relation to the network as a whole. This is done in a few separate steps. Firstly, pipe index is calculated for all pipes in the network. Secondly, lower and upper bound are established in order to reduce the search space before the GA even starts. Therefore, the process requires human analysis, ideally experienced engineering judgement. Moreover, this method requires repeated calculations of pipe index vector, thus making the whole process unnecessary complex. Kadu et al. (2008) used a ‘path concepts’ by converting the loop network into branch and classifying links. The method follows the assumption that the cheapest way of transporting water is along the shortest path. Nevertheless this method is very limited as it is unsuitable for real world networks with multiple loading patterns and networks components like pumps. Therefore, developing robust and efficient methodology for WDN optimization with reduced search space has become an important problem.

1.2 SCOPE OF THE PRESENT RESEARCH

The purpose of this study is to develop a useful and versatile approach able to identify the cost-entropy optimal solutions in one integrated process. The overall aim is to translate the wealth of entropy-based research into large, real world WDNs and therefore demonstrate the advantages of using the entropy-based method to design and optimize WDNs. The research tackles three challenging and complex research aspects related to WDN optimization: 1) identifying the most reliable, cost efficient optimal design of the WDN; 2) incorporating multiple operating conditions into the entropy-based WDN optimization process; 3) to speed up the optimization process by reducing the size of the search space with the use of properties of entropy.

1.3 OBJECTIVE OF THE RESEARCH

The objectives of the present research are as follow:

1) To develop a robust, penalty free EA model capable of identifying diverse solutions in cost-reliability Pareto optimal front (POF). Majority of methods proposed in literature concentrate on looking for cost efficient designs while neglecting hydraulic reliability, which in reality is almost equally important objective.

2) To extend the entropy based EA optimization method of WDN to be able to work under multiple operating conditions. The motivation for this objective is to enable the algorithm to handle more realistic situations when the demands vary with the time of the day, thus to include loading patterns like minimum demand, average demand, fire flow, etc.

3) To test the practical capability, robustness and consistency of the above mentioned model by applying it to hypothetical and real-life networks.

4) To employ the property of entropy function in order to reduce the search space. Doing so will allow the algorithm to look for solutions on feasibility boundary and improve the convergence properties of GA, thus speed up the optimization process.

1.4 A BRIEF DESCRIPTION OF THE METHODOLOGY

This thesis presents innovative penalty-free multi-objective evolutionary algorithm for the optimization of WDNs. The model does not require any constraint handling procedures or penalty functions as the hydraulic feasibility has been incorporated as objective function. The approach is composed of three interactive primary modules: the flow direction handling module was established as part of presented research; hydraulic simulator EPANET 2 and multi-objective evolutionary algorithm NSGA II. All modules are fully integrated and automated, thus no manual intervention is needed. Moreover, apart from identification of initial input data (i.e. population size, crossover and mutation rate) it does not require time consuming calibration and no high-level experience is needed to apply the algorithm to WDN. The algorithm is capable of identifying the set of cost – reliability optimal solutions. Additionally, the algorithm has been modified in order to be able to handle SOC as well as MOC. Furthermore, the search space reduction method has been incorporated into the approach. Entropy maximization is used to reduce the algorithm searching space by limit number available pipe diameters.

1.5 LAYOUT OF THE THESIS

The thesis contains six chapters in total. Following the introduction presented herein, the thesis is organized as follows:

Chapter 2 describes GA optimization technique, its advantages and drawbacks. Various multi-objective algorithms employed in WDNs optimization are also reviewed. Basic principles of hydraulics in WDNs are presented in the literature review chapter for completeness. Finally, the chapter also addresses different measures for reliability assessment of WDN.

Chapter 3 introduce new penalty-free, maximum entropy based multi-objective optimization approach for WDNs that is capable to work under multiple operating conditions (MOC). Three different methods of designing for MOC (i.e. entropy approaches) are investigated. The methodology is explained in detail and applied to three well-known benchmark networks. Comparison of results for single operating condition (SOC) and MOC with identification of the best MOC approach is demonstrated. Moreover, novel technique for identifying POF with diversely spread feasible solutions is presented.

Chapter 4 demonstrates the robustness of reliability based multi-objective GA proposed in previous chapter. A case study based on real life network from Italy is used. Novel technique for identifying POF described in Chapter 3 is also employed and its effectiveness is highlighted. The GA performance analysis is presented.

Chapter 5 proposes novel search space reduction method to improve the convergence of GA based on the entropy function. The methodology is presented in detail and applied to hypothetical network as a case study. Algorithm performance is analysed to demonstrate its robustness, consistency and efficiency.

Chapter 6 presents the conclusions from the present research and provides suggestions for future studies.

CHAPTER TWO

HYDRAULIC MODELLING AND DESIGN OPTIMIZATION OF WATER DISTRIBUTION NETWORKS

2.1 INTRODUCTION

Water distribution networks (WDNs) consist various components such as pipes, pumps, valves, water supply and tanks. All components are necessary for real world WDNs. Their location and operation is dependent on the type of the network, layout topography and distribution of demand nodes. Moreover, the WDNs need to be reliable systems in order to satisfy current and future water demand at sufficient pressure and quantity. There are many various situations that have and influence on networks performance. This includes different loading patterns, pipe burst, network maintenance, valve closures, pump breakage etc. For that reason, hydraulic performance of the system is crucial for WDN design and operation. With the use of formulations of mathematical equations, the hydraulic simulations models are carried out. They replicate the operation of real WDN and provide information of network performance under normal and abnormal operating conditions (i.e. fire flows, network failure or maintenance). Such predictions are highly valuable information that assists the engineers in decision making process.

Two different methods for WDN analysis are identified in the literature. The first one, called demand driven analysis (DDA) is the conventional approach based on nodal flows. It is assumed that nodal demands are fixed and always fully satisfied at all nodes regardless of the nodal pressures. This type of analysis provides accurate results only for situations with sufficient pressure. In case of any pressure shortfall, which may occur due to pump or pipe breakage, the DDA may lead to misleading nodal head demands. In such situations, the head dependent analysis (HDA) that is based on nodal heads would provide realistic results.

There are two types of hydraulics simulations: steady state simulation (SSS) and extended period simulation (EPS). SSS assumes that nodal demands are constant which is extremely unreal as demands vary with the time of the day and there are many different loading patterns that have to be satisfied by the network. It is common practice to use maximum daily demand, often with the fire event when designing WDNs. Nevertheless, designing WDN based on single operating condition, even the one with the highest possible water demand, does not ensure that other operating conditions (i.e. pressure constraints or loading patterns) will be satisfied. EPS is more realistic in evaluating system performance over 24h cycle as this type of simulation takes into consideration various operating conditions and nodal demands, pumps operation, tanks filling and emptying.

Proper evaluation of WDN capability is essential for the network to be able to supply required amount of water despite the circumstances. Spare capacity needs to be included while designing the WDN in order for the network to perform well under both normal and abnormal operating conditions. As a result, the methods for network performance assessment were developed. They can be divided into two groups: accurate measures and surrogate measures. Hydraulic reliability and failure tolerance (Tanyimboh and Templeman, 1998) belongs to accurate measures that determine the ability of the network to satisfy demands. Informational entropy (Tanyimboh, 1993), resilience index (Todini, 2000), network resilience (Prasad and Park, 2004) are examples of surrogate measures proposed in the literature.

Optimization of WDN design is an extremely complex problem, usually multi-criterion called in literature as NP-hard. DeNeufville *et al*. (1971) were one of the first researchers who recognised that water design system involves at least two conflicting objectives (i.e. cost and network performance). For many years, conflicting objectives were often approached by aggregating the objectives into a scalar function and solving them as single-objective optimization problem. Presently, evolutionary algorithms (EAs) are employed for optimization of real world NP-hard problems. EAs are stochastic search methods and well suited for such problems as they search for set of points rather than single point (Goldberg, 1989). Through different EAs, the genetic algorithms (GAs) are well known and widely used. GAs has been extensively employed for solving WDN optimization problem and proved to be robust and efficient.

Due to stochastic nature of GA, the number of evaluation required to identify optimal solution can be enormous. For the large networks, with hundreds of pipes and many network components, it can be extremely time consuming process as the GA might require even millions of evaluation functions. It is therefore desirable to reduce the search space in order to speed up the optimization process.

This chapter provides a literature review for all aspects relevant to research presented herein. Firstly, the GA optimization technique is described in detail as it has been employed for this study. Conventional GA procedure, various processes involved and major advantages are outlined. Several constraint handling techniques and search space reduction methods are presented and their drawbacks highlighted. Different multi-objective algorithms employed by researchers in WDNs optimization are reviewed. Secondly, the fundamentals of hydraulics in WDNs are presented. Demand driven analysis and head dependent analysis are described with shortcomings listed. Two types of network analysis, such as steady state simulation and extended period simulation are discussed alongside with description of different operating condition. Finally, accurate and surrogate network performance measures are presented. Since the entropy is employed in presented study as a network performance method, it has been described with more details. Several studies with correlation between hydraulic reliability and entropy are included and discussed.

2.2 GENETIC ALGORITHMS IN WATER DISTRIBUTION NETWORKS

Evolutionary algorithms are stochastic search techniques well suited for identifying Pareto optimal front in multi-objective, complex optimization problems. General principle of EAs search techniques is to mimic natural selection and genetics in order to survive. In contrast to conventional methods such as linear programming or gradient search EAs possesses the ability of solving non-convex, discontinuous, nonlinear problems with discrete variables. The population of solutions is used in each iteration rather than in single run, thus providing population of solution as outcome. Alike other population based techniques, EAs are capable of searching different regions of solution space simultaneously, hence increasing the possibility of finding a diverse set of solutions. For contradicting, multi-objective problems (i.e. cost and reliability in WDN optimization) identification of multiple optimal solutions in final population is indispensable in WDN planning or management.

For the last decade many population based optimization techniques have been proposed and employed for WDN design optimization. The main categories include genetic algorithm (Holland, 1975; Dandy et al., 1996; Savic and Walters, 1997; Wu and Simpson, 2001; Vairavamoorthy and Ali, 2005; Kadu *et al*., 2008 and many more), differential evolution (Storn, and Price, 1997; Vasan and Simonovic, 2010), particle swarm optimization algorithm (Kennedy and Eberhart, 1995; Montalvo *et al*., 2008), ant colony optimization techniques (Dorigo *et al*., 1999; Maier *et al*., 2003), harmony search (Geem *et al.*, 2001; Mahdavi *et al.* (2007), shuffled frog leaping algorithm (Eusuff and Lansey, 2003).

Among all EAs techniques, the genetic algorithm is the best known and widely employed for WDNs design optimization. Results published in literature prove that GAs are efficient and capable of finding optimal or near optimal solution (Dandy et al., 1996; Savic and Walters, 1997; Wu and Simpson, 2001; Vairavamoorthy and Ali, 2005; Kadu *et al*., 2008). GAs were also applied for network design and rehabilitation (Dandy and Engelhard, 2001), pump operation scheduling (Goldberg and Kuo, 1987), tank sitting and sizing (Prasad, 2010) and water quality optimization (Munavalli and Kumar, 2003). The research carried out for this thesis involved implementation of GA, therefore only this EA technique is described in following subsections.

2.2.1 Conventional Genetic Algorithm Procedure

Genetic Algorithm (GA) is an adaptive heuristic search optimization technique that finds approximate solutions to NP-hard problems. The GA concept was introduced by Holland (1975) and is inspired by Charles Darwin's natural theory of evolution. In nature, the strongest and fittest individuals have higher chance to survive and pass their genes in the reproduction process to the next generation. The weakest individuals extinct in evolution process. It is expected that the population in new generation is better than the old one. Similar rules apply to GAs.

In GA the entire process begins by creating random sets of solutions called population. This is usually done by the user by defining random seed. Solutions, also known as individuals, are encoded as chromosomes. Each chromosome is built from genes, which are design variables (i.e. pipe sizes in WDN optimization). At each generation, the chromosomes from parent population are evaluated in relation to optimization aims (i.e. objective functions). Individuals are assigned fitness values that allow to assess how close the chromosome is from achieving the aim. Then, the solutions are chosen for mating pool where the reproduction phase follows. Individuals with better fitness value have higher chances to be chosen for reproduction. Genetic crossover is applied first to produce offspring, which is based on genetic material from selected parents. Afterwards, small part of offspring individuals undergo genetic mutation to preserve diversity within population. In the next step, the individuals from offspring population are assigned fitness value. Finally, according to their fitness, individuals are selected for the upcoming generation and the whole cycle is repeated. The process continues until termination condition is meet (i.e. usually when specified number of generation is reached). The operation of a simple GA is illustrated on Figure 2.1.

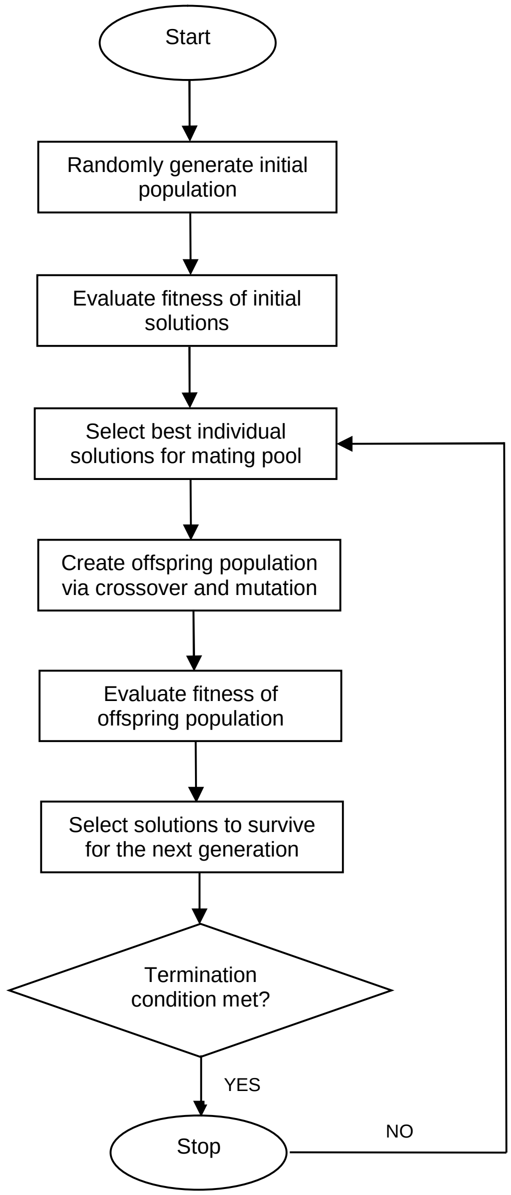


Figure 2.1 Flow chart of a simple genetic algorithm

2.2.2 Advantages of using Genetic Algorithms

Genetic algorithms have been widely used to solve real problems in various disciplines as can solve non-continuous, multi-dimensional and non-differential problems. Any optimization problem, which can be described with chromosome encoding, can be solved by genetic algorithms. As the whole concept is based on natural evolution, it is easy to understand and does not require any information about the structure of the function or advanced mathematical knowledge. The GA works with points of population in parallel. Therefore, it has ability to perform a global search and reduce the chances of being trapped in local optimal solutions. Using population based algorithm has additional advantage as the work required to reach optimal solutions and computational effort are reduced (i.e. optimal solutions is easier to found within the population rather than applying the algorithm many times). Moreover, using reproductive operators (i.e. crossover and mutation) leads to creating offspring population based on successful parent population while the selection operator eliminates reproduction of weak solutions. Finally, as previous studies showed, the GA has adaptability to hybridize with other techniques to solve specific and complex optimization problem.

2.2.3 Solution Representation

Genetic algorithms works on two type of spaces, such as phenotype space (i.e. solution space) and genotype space (i.e. coding space). Encoding is the transformation between phenotype and genotype, while decoding means transformation from genotype to phenotype. Chromosome encoding is the first issue that has to be addressed when the GA is used as optimization tool. It mainly depends on the type of the problem. Originally, the simplest binary encoding was used to represents individuals. Over the time, different encoding methods have been developed to suit various GA optimization problems. Based on type of symbols used to represent the values of genes, the encoding methods can be divided into four groups: binary encoding, real encoding, general data structure encoding and integer permutation encoding (Gen *et al*., 2008).

Binary encoding is the most common, mainly because it is the most straightforward method. Chromosomes are represented by a combination of bits 0 and 1. The binary coding is well suited for optimization problems with discrete values as decision variables (i.e. predefined pipe size in WDN optimization), therefore it has been used for research presented in this thesis. Table 2.1 presents the binary code representation of eight available pipe sizes.

Table 2.1 Available pipe sizes, corresponding binary code and gray code representation

Pipe size (mm)	Binary coding	Gray coding
100	000	000
200	001	001
300	011	010
400	010	011
500	110	100
600	111	101
700	101	110
800	100	111

There are two drawbacks related to binary coding in GA. The main one is the existence of redundant codes (Herrara *et al.* 1998) which occur if the number of decision variables is different than multiplier of number two. For example, for network with 6 available pipe diameter sizes, 3-bit substring would need to be used providing 8 substring (i.e. 2^3). Therefore, 6 substrings would be assigned to decision variable while remaining 2 substring would be redundant. Besley *et al.* (1993) identified three different possibilities for handling the redundant codes. First two options would be to either discard the redundant codes or to assign the low fitness values. Nevertheless, both methods may lead to poor GA behaviour by losing important genetic material (Herrara *et al.* 1998). Another possibility is to link the redundant codes with existing ones. This could be done by random or fixed remapping. At random remapping the redundant codes are randomly assigned to the valid codes. It is bias free (Eshelman *et al.*, 1989) but there is a possibility that less information could be passed from parents to offspring. With fixed remapping the redundant codes are allocated to valid codes based on usually predefined pattern. For example, the codes could be assigned to the decision variable at extreme sides (i.e. lowest and highest decision variable), to the central decision variables or uniformly spread to available decision variables (i.e. to every X^{th} decision variable, depending on number of redundant codes and number of decision variables).

Another disadvantage related to binary encoding is the existence of Hamming cliff. The Hamming cliff describes the situation when completely different genotypes (binary code) represent neighbouring phenotypes (pipe sizes). For example, pipe sizes with 400mm and 500mm are assigned values of 011 and 100 in the binary code (Table 3.1). Therefore in the phenotype domain they are neighbours whilst require three successful bit flips in the genotype domain. Since the mutation probability is rather low, there is small chance that all three bits will be changed at once. An effective way to solve the Hamming cliff problem is to use gray coding, where two neighbouring substrings vary only in one bit (Table 2.1).

In real coding, the genes are represented directly by the real numbers with unique values. Radcliffe (1992) suggested that differentiation between phenotype and genotype is not required for evolution, which is applicable in real coding. Since phenotype space and genotype space have the same topological structures there is no need for encoding on decoding process (Gen *et al*., 2008). Real coding is ideal for optimization problems with continuous search space and problems with large number of decision variables, which in binary coding would be represented by lengthy chromosomes. Real coding GAs have been widely employed into WDN optimization. McCormick *et al.* (1972) and

Eshelman and Schaffer (1993) proved that real coding is more powerful than binary coding for constrained problems optimization. Michalewicz and colleagues (Janikow *et al.*, 1991; Michalewicz, 1992) highlighted advantages in relation to GA efficiency based on comparison between real coded and binary coded GAs. Despite the advantages of employing real coded GA, the decision whether choose real or binary coding should be made by the user as each of the method is appropriate and well suited for different type of fitness functions (Goldberg, 1991 and Eshelman *et al.*, 1993).

2.2.4 Genetic Operators

Genetic operators are key elements of GA convergence and performance as they introduce diversity in searching the solution space. Three components, such as crossover, mutation and selection constitute the genetic operators. Mutation and crossover allow the GA to make changes within chromosomes from one generation to the next one and provide better exploration and exploitation of the searching space. The selection operator leads the search into areas with better solutions.

2.2.4.1 Genetic Crossover

Crossover is the principal operator in reproduction phase. It leads the population to converge on the best solutions found, thus exploiting and concentrating on specific point or direction. Crossover involves the exchange of genetic material of two parents to form two new offspring individuals. Among all chromosomes within population parents with preference towards fitness are chosen. By doing so, it is expected that the new offspring will retain the good genes from their parents. The simplest and basic crossover, called single point crossover takes place when one point is selected and the binary string up to the crossover point is copied from one parent while the rest of the chromosome is taken from second parent. Single point crossover is illustrated in Figure 2.2. Analogously, two point or multiple point crossover could be applied. There are also many different crossover operators. For example, uniform crossover where fixed mixing ration is used between the parents or arithmetic crossover where some arithmetic operations need to be performed in order to make offspring. For the research presented herein single point crossover has been used.

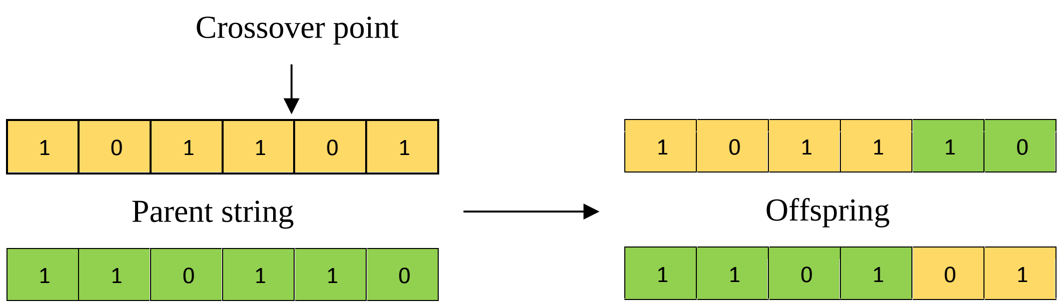


Figure. 2.2 Operation of single point crossover

The number of chromosomes that undergo crossover depend on the ratio of amount of new chromosomes produced (i.e. offspring) to the population size and is called crossover probability. Higher crossover probability allows exchanging more chromosomes thus the GA can explore more of the solution space. Moreover, it reduces the chances of being trapped in local optima.

2.2.4.2 Genetic mutation

Unlike the crossover, the role of mutation is to widen exploring space by introducing random change within chromosome at the gene level. This reintroduces genetic diversity, allows the algorithm to look in different directions and protects it from being trapped at local optima. Number of genes that undergo mutation depend on chromosome length and is usually very low value. Too high mutation probability will lead to random process and hold back the algorithm from quick convergence. Different kind of mutations can be performed, such as inversion mutation, displacement mutation, directional mutation etc. Figure 2.3 illustrates single point mutation where single bit is randomly flipped within the chromosome.

Figure 2.3 Operation of the single point mutation

2.2.4.3 Genetic Selection

The genetic selection is a process to determine which solutions are to be preserved and allowed to reproduce and which should be eliminated. The main objective of selection operator is to emphasise the good solutions and exclude the bad solutions in the population, while keeping the same number of individuals within population. Hence, it directs the GA search towards promising region in search space. The process is dependent on the selection pressure defined as degree to which the better solutions are favoured (Back *et al.*, 2000). The GA convergence is highly dependent by selection pressure as higher selection pressure reflects in quicker convergence. To avoid premature convergence the lower selection pressure is desirable at the early stage of genetic search, in order to widely explore the search space and keep diversity in the population. At the end of the genetic search, the higher selection pressure is preferred to narrow the searching space. Application of appropriate selection pressure allows preserving a good balance between exploration and exploitation necessary for optimization problem (Goldberg and Deb, 1991). There are few different techniques to implement the selection operator into GA however, the roulette wheel selection and tournament selection are the most common and widely used.

Holland (1975) proposed the roulette wheel selection, also known as proportional selection. Each individual in the population has the specified surface on the roulette wheel. The area occupied by solution is proportional to its fitness, so individuals with better fitness have larger area, thus higher chance for selection. Number of rotation of roulette wheel will be equal to the number of individuals in the population. Goldberg and Deb (1991) pointed out that this method increases the chance of losing genetic diversity within population, thus making the algorithm to converge prematurely.

The tournament selection has been developed by Goldberg and Deb (1991). Specified number of solutions are randomly chosen among the population. Their fitness is compared and the best individuals (i.e. highest fitness value) are selected for mating pool. The tournament process continues until the mating pool is sufficiently filled. Major advantage of using tournament selection is its flexibility, as the selection pressure can be easily adjust by changing the tournament size. Smaller tournament size increase the possibility of choosing the individuals with lower fitness value whilst higher tournament size results in a mating pool consisting higher number of fitter individuals.

2.2.5 Constraints Handling Techniques

Evolutionary algorithms are unable to distinguish feasible and unfeasible solutions simply because those optimization techniques were designed to deal with unconstraint problems. For that reason GAs are incapable of handling the constraints directly. Since majority of optimization issues are constraints problems, few different methods has been identified to convert the constraint problem into unconstraint. Michalewicz (1995) grouped existing techniques into four strategies: repairing strategy, modifying strategy, rejecting strategy and penalizing strategy. The most common practice is to penalize infeasible solutions in order to reduce their fitness. Penalizing strategy require incorporating penalty cost applied to the actual cost of WDN for the infeasible solution (Savic and Walters, 1997). The penalty cost is a function of constraint violation and penalty multiplier. Greater constraint violation reflects in higher penalty cost, thus higher chance that solution will be rejected in the next generation. However too high penalty cost restrict the search to the feasible region whilst completely avoiding region with infeasible solutions. It is not desirable as searching through feasible and infeasible regions improves the GA efficiency and identify better solutions than searching through feasible regions only (Glover and Greenberg, 1989). Moreover searching through feasible region only will result in very expensive solutions. On the other hand, low penalty cost will mislead algorithm to rank with similar fitness value the feasible and infeasible solution, thus the search could be confined toward infeasibility region. Adjustment of suitable penalty parameters is complicated task and require extensive fine-tuning before it can be efficiently incorporated into the algorithm.

Researchers endeavour to address the issue related to penalty function. Deb (2000), also Prasad and Park (2004) incorporated method for constraints handling in which the penalty coefficient is not required. This method employs the tournament selection operator wherein feasible solutions are favoured over infeasible. In case of infeasible designs the one with smaller constraint violation is chosen, whilst for two feasible individuals the one with better fitness value is selected. Major drawback of this approach is that the solutions with high cost will be preferred over cost effective, slightly infeasible designs that could be satisfactory in practice.

Farmani *et al.* (2005) developed self-adaptive method that involve application of two-stage penalty but does not require any parameter calibration. Firstly, the highly infeasible solution is assigned penalty cost equal or higher to the cost-efficient feasible design. Then, the penalty cost of that infeasible solution is further increased to the cost of the most expensive design. Penalty costs for remaining infeasible solutions are calculated exponentially in proportion to their infeasibility. This approach can lead to situation when the search will direct to infeasible region, as the low-cost infeasible solutions might be selected over feasible designs with higher cost.

Wu and colleagues (Wu and Simpson, 2002; Wu and Walski, 2005) developed self-adaptive penalty method that lead the GA to search the boundary of feasibility.

Nevertheless, several additional parameters have to be calibrated in advance in order for the approach to be implemented.

Khu and Keedwell (2005) avoid function penalizing by converting the nodal pressure constraint into objective function. Such method imposes unnecessary complications, calculations and computational burden, which will increase with the increasing number of nodes.

2.2.6 Search Space Reduction

The population size, genetic operators, encoding type and other parameters have direct impact on GA performance. However, there is no doubt that number of commercially available pipes and the size of the network have the greatest influence on GA search, its productivity and efficiency. Larger network with more candidate pipe sizes reflect in wider searching space, thus increase computation time required to achieve convergence and reduce the chances for reaching global optimum. For the real world networks, with hundreds of pipes and many network components, it can be extremely time consuming process as the GA might require even millions of evaluation functions. By limiting the number of candidate pipe sizes the search space can be greatly reduced. For example, the hypothetical Six-Loop network (Tanyimboh and Sheahan, 2002) that consist 17 pipes and 12 candidate pipe diameters have a search space of $(12)^{17}$ or 2.218×10^{18}. Eliminating one candidate pipe diameter gives a search space of $(11)^{17}$ or 5.054×10^{17}, thus reduce it by over 77%. Nevertheless, reduction of candidate pipe diameters has to be done in very careful manner as selection of unsuitable candidate pipes may reflect in locating in sub-optimal solutions.

Vairavamoorthy and Ali (2005) presented search space reduction methodology by eliminating candidate pipe diameters based on pipe index (PI). The index is used to measure the impact of the pipe on the hydraulic performance and the importance of the pipe in relation to the whole network. Based on the PI, tighter bound constraints are applied, thus unnecessary sections with infeasible designs can be excluded from the search space. The process is divided into few steps. Firstly, the PI is calculated for each pipe in the network. Secondly, based on PI, the pipes are ranked and divided into groups. Then the initial population is generated. During the GA search, the entire process of calculating PI has to be repeated many times. Since it is highly complex calculation that involve system of linear equations with right hand term it could impose unnecessary computational burden, especially if large network is considered.

Kadu *et al.* (2008) highly reduced GA search space by using critical path concept for the design of WDN (Bhave, 1978). The method is rely on assumption that the cheapest option to deliver water from the source to the demand node is through the shortest path. The looped network is converted into branched network and continuous pipe diameters are obtained for each link of the network. Then, based on the continuous pipe size the closest available commercial pipe size is selected. Subsequently, additional two nearest pipes with higher diameters and two nearest pipes with lower diameters are added thus giving five pipe diameters in total. Employing critical path method allow to reduce the search space substantially, however determining the shortest path may be complex problem itself, especially if the network considered is highly looped with multiple sources and many pipes. More importantly, presented path concept is very limited, as it is unsuitable for networks with components like pumps and tanks, or for network working under multiple operating conditions.

Haghighi *et al*. (2011) proposed hybrid optimization scheme by connecting the GA with Integer-Linear Programming (ILP). The approach requires transforming looped network into quasi branched, in order to define a path from a source to each demand node. Hereby, a single pipe from each loop is excluded and called as ignored pipe. The ignored pipes are optimized by GA whilst ILP optimize all other pipes (i.e. the pipes that are part of branch network). Afterwards, the ILP returns the optimal pipe diameters to the GA and the evolution process in carried on until reaching termination criteria. The approach has been applied to two networks from literature. In both cases great search space reduction has been recognised. Nevertheless, the approach has been applied to the network with single demand operating condition.

2.2.7 Review of Multi-Objective Genetic Algorithms used in Water Distribution Networks

The genetic algorithm belongs to population based search techniques, hence they are capable of searching different regions of solution space simultaneously. This feature makes them are well suited for complex multi-objective optimization problems. Several various multi-objective evolutionary algorithms (MOEA) have been developed in last two decades. Few different classification of MOEA can be found in the literature. Gen *et al.* (2008) grouped algorithms based on fitness assignment while Deb (2001) classified the MOEA as non-elitist and elitist. Elitism is the extremely important feature that favours the best solutions of a population by keeping them intact for next generation. Therefore, the fittest candidates will not be lost in optimization process even if found at early stage and helps in achieving better convergence (Zitzler *et al.*, 2000). Vector Evaluated Genetic Algorithm (Schaffer, 1985), Vector-Optimized Evolution Strategy (Kursawe, 1990) Multi-Objective Genetic Algorithm (Fonesca and Fleming, 1993), Weight-Based Genetic Algorithm (Hajela and Lin, 1992), Non-Dominated Sorting Genetic Algorithm (Srinivas and Deb, 1994) belongs to non-elitist evolutionary algorithm. Examples of elitism MOEAs include Elitist Non-Dominated Sorting Genetic Algorithm (Deb *et al.* 2002), Distance based Pareto Genetic Algorithm (Osyczka and Kundu, 1995), Strength Pareto Evolutionary Algorithm (Zitzler and Thiele, 1998), Pareto-Archive Evolution Strategy (Knowles and Corne, 2000). Few MOEA, widely used in WDN were chosen and described below in more details.

2.2.7.1 Vector Evaluated Genetic Algorithm

Schaffer (1985) proposed the first real implantation of MOEA called Vector Evaluated Genetic Algorithm (VEGA). It is the most straightforward and simplest extension of single objective evolutionary optimization. VEGA is based on vector evaluation, with each element of the vector describing each objective function. In each generation, the population is divided randomly into equal subpopulation. The number of subpopulation is equal to number of objective functions and individuals in particular subpopulation are assigned fitness based on corresponding objective function. The mating of subpopulation is performed by applying crossover and mutation. The major drawback of this MOEA is that each solution is evaluated and tested only for one objective function. Whereby the method is capable of identifying Pareto optimal solutions but does not maintain diversity of solutions in Pareto optimal front.

2.2.7.2 Strength Pareto Evolutionary Algorithm

Strength Pareto Evolutionary Algorithm (SPEA) has been developed by Zitzler and Thiele (1998). The elitism function is introduced by preserving the non-dominated solutions in external population. The fixed number of non-dominated solution is stored from the beginning of the simulation run. After each generation newly chosen non-dominated solutions are compared with stored solutions in order to identify the overall non-dominated designs that later are preserved. Moreover, solutions kept in external population are also engaged in genetic operators alongside with solutions from current population. In order to maintain elitism functionality and direct the search towards global optimum there has to be good balance between number of solutions in current and external populations. If the external population will be too small in relation to current population, the elitism may be lost. Whilst having too large external population could hold back the algorithm from converging to the Pareto optimal front.

2.2.7.4 Elitist Non-Dominated Sorting Genetic Algorithm

Deb *et al.* (2002) proposed Elitist Non-Dominated Sorting Genetic Algorithm (NSGA II). The NSGA II not only includes elitism functionality but also mechanism that preserves diversity among Pareto optimal solutions. Firstly, the random parent population is generated. To create offspring population the mutation and crossover are applied. The offspring and parent populations are combined together and undergo non-dominated sorting. This step ensures elitism by preserving previous and current best individuals. Non-dominated sorting involves dividing results into different fronts and assigning fitness values (i.e. ranks). The first front is non-dominated with the highest fitness value. The second front has assigned second best fitness value and its individuals are dominated by individuals from first front. This goes on until all results have designated fitness values. In addition, the crowding distance is calculated in the objective space for each solution. Crowding distance is a measure of distance between neighbouring individual. High crowding distance means that individual is from less crowded area, thus having such solutions results in better diversity. Based on fitness values and crowding distance of the last front, the best individuals are chosen. The next generation is formed by selecting non-dominated with the highest rank and then subsequent non-dominated fronts in order of their ranking. In case if last accepted front has more results than required, the individuals are chosen based on crowding distance (i.e. results with higher crowding distance will be selected). Therefore, created population include best possible individuals ensuring diversity among solutions. The process is repeated in subsequent generations until reaching termination criteria.

Many researchers in various disciplines employed the NSGA II for solving the optimization problems. Comparisons of NSGA II and other evolutionary techniques can also be found in the literature. Deb *et al.* (2002) has tested NSGA II, PAES and SPEA on nine difficult problems. NSGA II outperforms other methods by achieving better convergence and the most diverse spread of solutions in POF. Farmani *et al.* (2003)

applied SPEA and NSGA II into two hypothetical and one real world WDN. The results indicate that both method are capable of identifying Pareto sets.

In WDN optimization the NSGA II is one of the most popular multi-objective evolutionary algorithm. Majority of the WDN research that employ NSGA II relates to design optimization through pipe sizing. However, different measures for network reliability are sometimes included into optimization as objective function (Prasad and Park, 2004; Farmani *et al.*, 2006; Saleh and Tanyimboh, 2013, 2014). Farmani *et al.* (2006), Prasad (2010), Siew and Tanyimboh (2011) employed NSGA II to optimize the design and upgrade the complex benchmark network called "Anytown". The problem involved pump scheduling, tank sitting and sizing and multi-operating conditions. Farmani *et al.* (2006) additionally included water quality, whilst Siew and Tanyimboh (2011) introduced boundary search techniques to focus the search on feasibility region. Optimal design and rehabilitation of WDN by minimization of network whole life cost and maximization of network performance was presented by Jayarama and Srinivasan (2008). Siew and Tanyimboh (2012) proposed penalty-free multi-objective evolutionary approach with head dependent analysis (HDA). Moreover, Siew *et al.* (2014) extended the approach for the whole life design and rehabilitation. The model took into consideration initial cost, rehabilitation, upgrading cost, repairs and pipe failures as well as deterioration over time. In the area of water drinking safety, Preis and Ostfeld (2008) and Weickgenannt *et al.* (2010) concentrated on contamination detection in WDN by optimizing sensor placements, whilst Jeong and Abraham (2006) focused on minimization strategy of the consequences of intended attack on WDN. Very recently, Salah and Tanyimboh (2013, 2014) proposed multi-objective optimization algorithm for layout, design and reliability optimization of WDN. The approach combines three models: NSGA II as optimization algorithm, hydraulic simulator and algorithm developed for topology confirmation. Barlow and Tanyimboh (2014) presented memetic algorithm that consist NSGA II hybridised with local and cultural improvement operators.

Based on many different and difficult test problems associated with WDN, it can be concluded that NSGA II is robust, efficient and effective multi-objective evolutionary method. For that reason, the NSGA II was chosen as the optimization algorithm for research presented in this thesis.

2.3 HYDRAULIC ANALYSIS OF WATER DISTRIBUTION NETWORKS

2.3.1 Governing Hydraulic Equations

Primarily principles related to hydraulic analysis in WDN is the conservation of mass and energy. Principle of conservation of mass leads to continuity equation. For steady flow process the nodal inflow equals nodal outflow. The equation can be written as

$$\sum_{ij \in in(Nj)} Q_{ij} - \sum_{ij \in out(Nj)} Q_{ij} = Q_j \qquad j = 1, \ldots, N \tag{2.1}$$

where is N the number of nodes on the network, Q_{ij} is the inflow for $ij \in in(Nj)$ and outflow for $ij \in out(Nj)$ at node j; Q_j is demand supply at node j.

Head loss equation arises from conservation of energy principle. For the looped network, the head loss in pipes composing one loop has to be equal to zero. Thus, the head loss equation can be formulated as follows:

$$\sum_{ij \in IJ_{lp}} h_{ij} = 0 \qquad lp = 1, \ldots, Nl \tag{2.2}$$

where IJ_{lp} stand for a loop of closed pipe cycle; h_{ij} the head loss in pipe ij and Nl is the number of loops in the network and can be received from subsequent equation

$$Nl = N - Nn + 1 \tag{2.3}$$

where N represents number of links in the network, while Nn is the number of nodes.

The total head loss for path in WDN must be equal to the difference in heads between starting nodes and ending nodes. The equation can be described as follows:

$$\sum_{ij \in IJ_p} h_{ij} = Hn_i - Hn_j \qquad \forall ij \in IJ \tag{2.4}$$

where H_i and H_j are the hydraulic gradient level at nodes i and j respectively; IJ is the set of all links in path p.

The pipe head loss can be expressed in various empirical formulae. In reality, three are mainly used to calculate the head loss presented in Eq.2.2 (Bhave and Gupta 2006). The widely known are Darcy-Weisbach equation, Hazen-Williams equation and Chezy-Manning equation are presented in Eq. 2.5, Eq. 2.6 and Eq. 2.7 respectively

$$h_{ij} = \frac{8 f_{ij} L_{ij} Q_{ij}^2}{g D_{ij}^5 \pi^2} \qquad \forall ij \tag{2.5}$$

$$h_{ij} = \frac{\omega_{HW} L_{ij}}{C_{(HW)ij}^{1.852} D_{ij}^{4.87}} Q_{ij}^{1.852} \qquad \forall ij \tag{2.6}$$

$$h_{ij} = \frac{\omega_{ChM} L_{ij} \left(C_{(ChM)ij} Q_{ij}\right)^2}{D_{ij}^{5.33}} \qquad \forall ij \tag{2.7}$$

where f_{ij} is the Darcy-Weisbach friction factor in pipe ij; L_{ij}, h_{ij}, D_{ij} are respectively length in meters, head loss in meters and diameters in meters for pipe ij; Q_{ij} is the flow rate in pipe *ij* in cubic meters per second; *g* is gravitational acceleration; η_{HW} and η_{ChM} stands for Hazen-Williams dimensionless conversion factors (10.67 in SI unit) and Chezy-Manning dimensionless conversion factor (10.29 in SI units), respectively; C_{HW} and C_{ChM} are Hazen-Williams roughness coefficient and Chezy-Manning roughness coefficient respectively.

The head loss h_{ij} could also be expressed using pipe resistance coefficient for pipe *ij* and described as

$$h_{ij} = K_{ij} Q_{ij}^{nf} \qquad \forall ij \qquad (2.8)$$

where K_{ij} is the pipe resistance coefficient for pipe *ij*; *nf* is the flow exponent equal to 1.852 for Hazen-Williams and 2 for Darcy-Weisbach and Chezy-Manning.

2.3.1.1 Formulation of Hydraulic Equations

Basic hydraulic equations (i.e. constitutive equations) can be defined in various ways. In general, there are three unknown variables: pipe flow rates, nodal heads and loop-flow corrections. Hydraulic equations with pipe flow rates as unknown variable are also described as q-equations (Bhave, 1991). Equation representing the flow continuity (Eq.2.1) and the head loss equation (Eq. 2.2) can represents the q-equations.

Hydraulic equations with nodal heads as unknown variable, also known as H-equations (Bhave, 1991) can be presented by rewriting the continuity equation (Eq. 2.1)

$$\sum_{i:H_i>H_j} \left(\frac{H_i - H_j}{K_{ij}} \right)^{\frac{1}{n_f}} - \sum_{i:H_i<H_j} \left(\frac{H_i - H_j}{K_{ij}} \right)^{\frac{1}{n_f}} - Q_j = 0 \qquad \forall ij \in Nj \qquad (2.9)$$

where Nj stands for the nodes connected to node j. In case of H-equations, the loop or path equations are not needed. However, there number of continuity equations will be equal to the number of unknown nodal heads, whilst one nodal head will be known (i.e. usually fixed head at source node).

Hydraulic equations with loop-flow correction as unknown variable (Q_{ij}) are also described as ΔQ-equations (Bhave, 1991). Firstly, it is assumed that initial pipe flow satisfy the flow rate continuity formulation presented in equation 2.1. Then, pipe flow rates are adjusted iteratively by correcting the loop-flow at each loop based on equation expressed as

$$Q_{ij}^{k} = Q_{ij}^{k-1} + \sum_{l \in l_{ij}} \Delta Q_{l}^{k} \qquad \forall ij \in l \tag{2.10}$$

where Q^{k}_{ij} is the corrected flow rate; Q_{ij}^{k-1} is the estimated flow rate; ΔQ_l^k is the loop-flow correction used for all pipes from loop l; $\sum \Delta Q_l^k$ is the total correction of all loops with pipe ij; l_{ij} means number of loops sharing link ij.

Having loop-flow correction (ΔQ_l^k) unknown, with the use of head loss equation (Eq. 2.2), the ΔQ-equations can be presented

$$\sum_{ij \in IJ_l} K_{ij} \left(Q_{ij}^{k-1} + \sum_{l \in l_{ij}} \Delta Q_{l}^{k} \right)^{n_f} = 0 \qquad \forall ij \in IJ_l \, ; I \in Nl \tag{2.11}$$

where IJ_l represents all the pipes in loop l and Nl is the total number of loops in the network. The ΔQ-equations can be solved at the same time using repetitive scheme.

2.3.2 Demand Driven Networks Analysis

Demand driven analysis (DDA) method assumes that nodal demands are fixed and always fully satisfied at all nodes regardless of the nodal pressures. In reality, nodal outflows are actually dependent on nodal pressure. If the nodal pressure gets lower than minimum required level, the flow will be reduced, thus nodal demands may not be satisfied in full. Therefore, the DDA method is capable of analysing WDN only under normal conditions. In case of any pressure shortfall (i.e. pipe or pump breakage, fire flow conditions, network maintenance), the DDA may lead to inaccurate and misleading nodal head results. In such situations, the relationship between pressure and outflow should be taken into consideration and the analysis where nodal demands are dependent on nodal pressure (i.e. head dependent analysis), would be more practical. Nevertheless, due to simplicity of using the DDA method, it is still widely used in water industry. There are four different methods for solving the DDA, such as: Hardy-Cross method (Cross, 1936), Newton-Raphson method (Martin and Peters, 1963), Linear Theory method (Wood and Charles, 1972) and Global Gradient method (Todini and Pilati, 1988). Detailed explanations and equations of those methods can be found in textbooks, however the Global Gradient method in presented herein for completeness.

2.3.2.4 Global Gradient Method

Todini and Pilati (1988) proposed Global Gradient Method (GGM) that simultaneously obtain values of nodal heads and pipe flow rates. Additionally, the gradient method directly obtains amended values of nodal heads and pipe flow rates in the head loss formula, i.e. $H_i^{k+1} = H_i^k + \Delta H_i^k$ and $Q_{ij}^{k+1} = Q_{ij}^k + \Delta Q_{ij}^k$ respectively. Moreover, like in other linear theory method, the GGM does not need satisfying the continuity equations at all nodes in order to start the process. The pipe flow rates and nodal heads are the unknown for the Q-H equations. Taylor's series are used to linearize the non-linear functions. Thus, the non-linear head loss equation for the k^{th} iteration can be presented as

$$\left(H_i^k + \Delta H_i^k\right) - \left(H_j^k + \Delta H_j^k\right) = K_{ij}\left(Q_{ij}^k\right)^{n_f} + n_f K_{ij}\left|Q_{ij}^k\right|^{n_f - 1} \Delta Q_{ij}^k \qquad \forall ij \qquad (2.12)$$

Rewriting and rearranging the above equation (Eq. 2.12) for the corrected nodal heads H_i^{k+1} and H_j^{k+1} leads to

$$H_i^{k+1} - H_j^{k+1} - n_f K_{ij}\left|Q_{ij}^k\right|^{n_f - 1} \Delta Q_{ij}^k = K_{ij}\left(Q_{ij}^k\right)^{n_f} \qquad \forall ij \qquad (2.13)$$

while

$$H_i^{k+1} = H_i^k + \Delta H_i^k \qquad (2.14)$$

$$H_j^{k+1} = H_j^k + \Delta H_j^k \qquad (2.15)$$

Substituting corrected pipe flow rate Q_{ij}^{k+1} to $Q_{ij}^k + \Delta Q_{ij}^k$ in Eq. 2.13 yields to system of linear equations that combine the corrected values of nodal heads and pipe flow rates and could be described as

$$H_i^{k+1} - H_j^{k+1} - n_f K_{ij}\left|Q_{ij}^k\right|^{n_f - 1} \Delta Q_{ij}^k = \left(1 - n_f\right) K_{ij}\left(Q_{ij}^k\right)^{n_f} \qquad \forall ij \qquad (2.16)$$

Linear continuity equation can be presented in terms of corrected pipe flow rates as follows:

$$\sum_{ij \in j} Q_{ij}^k + Q_j = 0 \qquad j = 1, \ldots, N - 1 \qquad (2.17)$$

Corrected nodal heads and pipe flow rates can be obtained by solving Eq. 2.16 and Eq. 2.17 simultaneously. The pipe flow rates Q_{ij}^k can be set to unity or randomly chosen value. With the progression of iteration, the variations in values for nodal heads and pipe flow rates become negligible.

Global gradient method proposed by Todini and Pilati (1988) is incorporated in hydraulic simulation software called EPANET 2 (Rossman, 2000). EPANET 2 is well known and widely used in designing water distribution piping system. It is DDA simulator that can handle any size of the network and carry out steady state analysis as well as extended period simulation. EPANET 2 computes friction head loss along the pipes by using one of the three common formulas (i.e. Hazen-Williams, Chezy-Manning or Darcy-Weisbach). In addition, minor head losses for bends and fittings are included in the software engine. EPANET 2 computes pumping energy and cost. The pumps are modelled using a head-flow curve. It is capable of modelling several types of different pumps, and with a constant or variable speed. It has also capabilities for modelling different types of valves and pressure-dependent flow at sprinkles. Tank sizing functionality (i.e. diameter depending on the tank height) is also included in EPANET 2. Apart from hydraulic modelling, the EPANET 2 provides also wide range of water quality modelling capabilities. It allows user to monitor the age of water through a system and analyse the movement of non-reactive material through the network as well as reactive materials that spreads in the system through the time. Global reaction rate coefficient that can be modified on pipe-by-pipe basis is also included. EPANET 2 engine provide opportunity to track percentage of flow from given node reaching others nodes through time and allows growth or decay reactions to proceed up to limiting concentration. Recently, Siew and Tanyimboh (2012) proposed a pressure dependent analysis extension, called as EPANET-PDX. It is based on EPANET 2 model with all its functionality (i.e. hydraulics, extended period simulation, water quality, demand driven analysis). EPANET-PDX has a logit pressure dependent function (Tanyimboh and Templeman, 2004 and 2010) that provides realistic results for network simulations with pressure shortfall. Tanyimboh and Templeman (2004, 2010) logit function is described in details in subsection 2.3.3.

EPANET 2 is available as an external stand-alone software as well as an open-source Programmer Toolkit. The toolkit is a dynamic link library (DLL) with more than 50 functions that allows researcher to customize EPANET for their own needs. For purposes of research presented in this thesis, the EPANET toolkit has been used.

2.3.3 Head Dependent Analysis

Head dependent analysis (HDA) method takes into account the pressure-dependent nature of flows. Relationship between nodal head and nodal outflow is considered which imply realistic representation of network deficiency. The greatest advantage of using HDA over DDA is its ability to identify the pressure deficient nodes in which the nodal demand is not fully satisfied. Therefore, unlike the DDA, the HDA will provide real nodal heads values in pressure deficiency situations. Over the years, several functions have been proposed for pressure dependency of nodal consumption. All of the functions are based on pressure-outflow relationship concept that nodal demand is satisfied if the nodal head is equal or higher than nodal required head. Few of the HDA functions are presented below.

Wagner *et al.* (1988) and Chandapillai (1991) proposed parabolic function of nodal outflow and pressure that can be expressed as

$$H_i^{des} = H_i^{\min} + R\left(Q_i^{req}\right)^{n_e} \tag{2.18}$$

where R_i represents the resistance coefficient, n_e is an exponent parameter which value can vary from 1.5 to 2 (Gupta and Bhave, 1996). Therefore

$$Q_i = Q_i^{req} \qquad H_i \geq H_i^{req} \tag{2.19}$$

$$Q_i = Q_i^{req}\left(\frac{H_i - H_i^{\min}}{H_i^{req} - H_i^{\min}}\right)^{\frac{1}{n_e}} \qquad H_i^{\min} \leq H_i \leq H_i^{req} \tag{2.20}$$

$$Q_i = 0 \qquad\qquad H_i \le H_i^{\min} \tag{2.21}$$

where Q_i is available outflow that can be delivered by the system at node i; Q_i^{req} is the require demand or supply at node i; H_i is the actual head at node i; H_i^{min} is the nodal head at node i below which the outflow is zero (i.e. in practice H_i^{min} is usually take as nodal elevation); H_i^{req} is the required head at node i for fully satisfying the demand.

Germanopoulos (1985) proposed approximation of nodal flow for a pressure deficient network according to the following equation

$$Q_i = Q_i^{req}\left(1 - b_i e^{-c_i\left(\frac{\Pr_i}{\Pr_i^*}\right)}\right) \tag{2.22}$$

where Pr_i is the available pressure at node i; Pr_i^* is the pressure at which proportion of the required demand of node i is satisfied; b_i and c_i are coefficient which need to be calibrated for node i. For the situation when the field data are not available, b_i and c_i are taken as 10 and 5 respectively, while Pr_i^* is taken as a head to satisfy 93,2% of the required nodal demand. Eq.2.22 has disadvantages as $Q_i \neq 0$ for $H_i = H_i^{min}$ and $Q_i \neq Q_i^{req}$ for $H_i = H_i^{req}$.

Gupta and Bhave (1996) proposed an improved version of Eq.2.22 that satisfies above-mentioned disadvantages. Their equation can be presented as

$$Q_i = Q_i^{req}\left(1 - b_i e^{-c_i\left(\frac{H_i - H_i^{\min}}{H_i^{req} - H_i^{\min}}\right)}\right) \tag{2.23}$$

Fujiwara and Ganesharajah (1993) suggested method where pressure dependent outflow is considered. The function has a complex form but is capable to analyse any WDN and can be presented as follows:

$$Q_i = Q_i^{req} \frac{\int_{H_i^{\min}}^{H_i} \left(H - H_i^{\min}\right)\left(H_i^{req} - H\right) dH}{\int_{H_i^{\min}}^{H_i^{req}} \left(H - H_i^{\min}\right)\left(H_i^{req} - H\right) dH} \qquad H_i^{\min} \le H_i \le H_i^{req} \tag{2.24}$$

Tanyimboh and Templeman (2004, 2010) proposed nodal outflow function, which is based on a Logit function. Tanyimboh function provides smooth transition between zero and partial nodal outflow and between and partial and full demand satisfaction. Therefore, there are no discontinuities in the function or derivatives, which is enormous advantage if compared with other head-outflow functions. Tanyimboh and Templeman equation (2004, 2010) can be described as

$$Q_i = Q_i^{req} \frac{\exp\left(\sim_i + \quad _i H_i\right)}{1 + \exp\left(\sim_i + \quad _i H_i\right)} \tag{2.25}$$

where α_i and β_i are essential parameters in determining the flow and need calibration from field data. Those two parameters determine the shape of the function curve. Therefore, if the field data are not available, Tanyimboh and Templeman (2010) advised substituting the demand satisfaction ratio (DSR) value with 0.01 for situation when $H_i<H_i^{min}$ and 0.999 if $H_i= H_i^{req}$.

When rearranged, Eq. 2.25 can also be presented as

$$\frac{Q_i}{Q_i^{req}} = \frac{\exp\left(\sim_i + \quad _i H_i\right)}{1 + \exp\left(\sim_i + \quad _i H_i\right)} \tag{2.26}$$

where Q_i/Q_i^{req} is the nodal demand satisfaction ratio and can have value between zero and 1.0. The relationship between DSR value and nodal head can be presented as follows:

$$H_i \ge H_i^{req} \qquad \text{if} \qquad DSR = 1 \tag{2.27}$$

$$H_i \le H_i^{\min} \qquad \text{if} \qquad DSR = 0 \tag{2.28}$$

$$H_i^{\min} \le H_i \le H_i^{req} \qquad \text{if} \qquad 0 < DSR < 1 \tag{2.29}$$

In other words, if the DSR equals 1.0 the nodal demand is fully satisfied. When DSR value is between 1 and 0 then nodal demand is partially satisfied and for DSR equals zero there is no nodal outflow.

Tanyimboh and Templeman (2010) developed prototype FORTRAN computer program, which is based on pressure dependent analysis methodology. Program for the Realistic Analysis of the Availability of Water in Distribution Systems (PRAAWDS) generate simulations that are more realistic for pressure deficiency scenarios than DDA simulator. PRAAWDS provides a choice of four different head-outflow relationships functions for running HDA simulation. These include Germanopoulos-Gupta- Bhave (Gupta and Bhave, 1996), Wagner *et al.* (1988), Fujiwara and Ganeshrajah (1993) and Tanyimboh and Templeman (2004). Apart from running HDA, there is also possibility to run conventional DDA simulation. Moreover, there is built-in feasibility evaluation feature that allows to verify the results achieved by using HDA simulator. PRAAWDS has been tested in several studies where the network performance measures (i.e. hydraulic reliability and failure tolerance) that require HDA were employed. Published results (Setiadi et al., 2005, Tanyimboh and Setiadi, 2008a and 2008b) prove that it is very robust and provide accurate results. PRAAWDS has been employed for research presented in this thesis in order to calculate hydraulic reliability and failure tolerance.

2.3.4 Types of Hydraulic Simulations

2.3.4.1 Steady State Simulation

Steady state simulation refers to a condition of the system whose water demands, reservoir and tanks levels remain constant. In other words, it could be compared to snapshot taken over the 24h period. This type of analysis does not reflects the real network operation through the day as demands, water levels in tanks or pump cycling will change over the time. Steady state simulations are widely used for designing WDN as do not require too many data and can identify required network layout and pipes sizes. However there are numerous cases when steady state modelling is rather limited (i.e. checking tank volumes, valves operation, analysing energy consumption), thus simulating WDNs over the time period is necessary.

2.3.4.2 Extended Period Simulation

Extended period simulation (EPS) can be done by linking steady state simulations. It allows keeping track of demand fluctuations, pump operation or tank water level. EPS can be launched for any duration of time, depending of the type of analysis required. In majority of cases, the EPS is run for 24h period, mostly because substantial fluctuations in demands vary on daily pattern. However, for water quality evaluation, longer period, such as few days will probably be used. Relevant expression related to EPS is the hydraulic time step (i.e. time interval). The time step is the length between steady state simulations. It usually varies between 10 minutes to 1 hour. It is assumed, that between the time steps the nodal demands remain constant while water level in tank can change. An EPS is more realistic than steady state simulation and provide better understanding of WDN behaviour (i.e. how water demand vary through the day and how the tank water level changes over the time). Moreover, the optimization of pump scheduling or tank sizing and sitting cannot be done without EPS.

2.3.5 Loading Conditions

2.3.5.1 Single Loading Patterns

Most of the work done so far is based on single operating condition, usually maximum daily demand. For many years it was thought that if network satisfy highest daily demand it will satisfy other conditions as well. Alperovits and Shamir (1977) suggested that when designing the WDN, not only maximum daily demand and fire flow should be included, but also the minimum demand periods have to be considered. Vamvakeridou-Lyroudia *et al*. (2005) stated that any solution, which will work well for peak loading, should also be good enough for normal day operating, thus there is no need to check the minimum pressure constraint for normal loading. Prasad (2010) proved that pressure constraints have to be considered for all loading patterns. He showed that even if the solution satisfy pressure at peak loading and emergency conditions, the tank could remain empty causing pressure deficiency at demand nodes.

2.3.5.2 Multiple Loading Patterns

Predominantly considered loading patterns are as follows:

(a) Maximum daily demand – the highest demand over 24h period. Quite often referred as critical demand, as it is largest demand to be supplied without using storage.

(b) Peak hour demand – very important loading for large WDN. Maximum day demand usually occurs in the evening of the maximum day.

(c) Average daily demand – the average rate of demand for average day.

(d) Minimum daily demand – called also replenishment simulation. Minimum loading pattern is the situation when water consumption is at its lowest level and the tanks are filled.

(e) Fire flows – depend on type of community (i.e. industrial estate would require higher supply than rural area).

(f) Emergency conditions – unplanned situations (i.e. major pipe breakage or power failure) and maintenance activities.

In most cases, either the peak demand or maximum day demand is treated as primary operating condition. Based on this demand and applying demand multiplier or peaking factor, other loading patterns could be calculated.

2.4 PERFORMANCE ASSESSMENT OF WATER DISTRIBUTION NETWORK

The primary objective considered whilst designing the water distribution network (WDN) is the cost minimization alongside with ensuring enough water at required pressure at demand nodes. Chosen solution would have the cheapest possible pipe sizes while satisfying consumers demand. However, it refers to situation when all network components are considered as available, so there is no disruption or failure. In reality, there is high chance that some networks elements will be unavailable due to breakage or deterioration over time (i.e. pipe breakage, pump failure). In such case, the network with smallest and cheapest pipe sizes will most probably not be able to meet required criteria in terms of demand or pressure. Therefore, some spare capacity need to be included while designing the WDN in order for the network to perform well under both, normal and abnormal operating conditions. For that reason, network performance assessment has become equally important as the cost reduction. The methods for performance assessment of WDN can be classified into two groups: accurate performance measures and surrogate performance measures.

2.4.1 Accurate Performance Measures

Hydraulic reliability and failure tolerance are considered to be correct measures of robustness of the design as they determine the ability of network to satisfy demands under normal and abnormal conditions.

2.4.1.1 Hydraulic Reliability

Since there is no general or commonly agreed definition of reliability, the reliability definition used in this thesis is adopted from Tanyimboh and Templeman (2000) and it is defined as the ability of the system to fulfil on average the required nodal demands at adequate pressure whilst considering both normal and abnormal operating. The equation is usually applied to peak or critical loading conditions when demands are constant and can be written as

$$R = \frac{1}{T}\left(p(0)T(0) + \sum_{m=1}^{M} p(m)T(m) + \sum_{m=1}^{M-1}\sum_{n=m+1}^{M} p(m,n)T(m,n) + \ldots \right) + \frac{1}{2}\left(1 - p(0) - \sum_{m=1}^{M} p(m) - \sum_{m=1}^{M-1}\sum_{n=m+1}^{M} p(m,n) - \ldots \right) \qquad (2.30)$$

where R represents the hydraulic reliability; M is the number of links in the system i.e. pipes, pumps and valves; $p(0)=a_1a_2a_3\ldots a_m$ stands for the probability that all links are in service; a_m is the probability that link m is in service at any given time; $p_m=p(0)(u_m/a_m)$ is the probability that only link m is not in service; $u_m=1-a_m$ is the probability that link m is unavailable; $p(m,n)=p(0)(u_m/a_m)(u_n/a_n)$ is the probability that only links m and n are not in service; T is the sum of the nodal demands; $T(0)$ means that total flows supplied with all links are in service; $T(m)$ is for the situation when only link m unavailable and $T(m,n)$ when links m and n out of service.

The pipe availability a_m can be calculated using several formulae from literature. However for the results presented in this thesis formula developed by Cullinane *et al.* (1992) was used.

$$a_m = \frac{0.21218 D_m^{1.462131}}{0.00074 D_m^{0.285} + 0.21218 D_m^{1.462131}} \qquad \forall m \in M \qquad (2.31)$$

where D_m is diameter of pipe m in inches.

Eq. 2.30 contains two main parts (i.e. the two pairs in large parentheses). The first part indicates total demand that is satisfied on average. However, in reality, it is impractical and time consuming to calculate all possible compositions of components failure. Thus, the calculation of first part of Eq. 2.30 underestimates the actual reliability. The second part of Eq. 2.30 corrects the underestimation from first part. Moreover, since the cases with two links out of service are unlikely to happen, thus such scenarios are usually not computed. Likewise, results presented herein are based on situations when only one pipe is out of service at any given moment.

2.4.1.2 Failure Tolerance

The pipe failure tolerance measure (Tanyimboh and Templeman, 1998) provides an estimate of the total demand the WDN is capable of satisfying when some components are out of service. Redundancy represents only spare capacity of the network, as situations when all link are available is not included in the measurement. It has been remarked by Tanyimboh *et al.* (2001) that redundancy could be better measure for supply disruption for a system with failure than the reliability for similar stable system. Tanyimboh and Kalungi (2008, 2009) emphasize the importance of including the redundancy measure alongside with hydraulic reliability for a better WDN performance. The equation for failure tolerance is defined as

$$FT = \frac{R - p(0)T(0)/T}{1 - p(0)} \tag{2.32}$$

where *FT* is the failure tolerance. The failure tolerance calculation is usually carried out once the hydraulic reliability is estimated. Having *R* and *p(0)* makes the redundancy calculation very easy and straightforward.

2.4.2 Surrogate Performance Measures

Due to high computation burden associated with WDN reliability calculation, along with no universal reliability definition, researchers have been looking for alternative measures. Few different surrogate measures of reliability have been proposed over the last decade. The greatest advantage of using such measures is that they are very simple to calculate and do not require repetitive simulations. Moreover, majority of surrogate measures are calculated based on nodal heads or pipe flow rates obtained from WDN modelling. Therefore, they could easily be incorporated in WDN optimization techniques.

2.4.2.1 Informational Entropy

Shannon (1948) used the fact there is some uncertainty related to every probabilistic scheme and introduced informational entropy as a quantitative measure of the quantity of information included in a finite probability distribution. Shannon's informational entropy function can be presented as follows:

$$\frac{S}{K} = -\sum_{i=1}^{n} p_i \ln p_i \tag{2.33}$$

where S represents the entropy; K is the arbitrary positive constant usually taken as 1.0; p_i is the probability related to the ith event. The probabilities represent the finite scheme (i.e. probabilities are exhaustive and independent mutually exclusive), thus

$$p_1 + p_2 + ... + p_n = 1 \tag{2.34}$$

2.4.2.1.1 Entropy in Water Distribution Networks

The idea of using Shannon's informational entropy function for WDNs was first introduced by Awumah *et al.* (1990). Using the flow rates in a probabilistic way to obtain the network entropy values, they showed that the performance of water distribution networks can be measured comparatively.

Tanyimboh and Templeman (1993a, 1993b) were the first to propose the correct definition of the entropy function for WDNs by using the conditional entropy formula of Khinchin (1953) and the multiple probability space model to formulate entropy function for general looped WDNs. Knowing the pipe flow rates, the Tanyimboh equation (Tanyimboh and Templeman 1993a) can be presented as follows:

$$S = S_0 + \sum_{i=1}^{N} P_i S_i \tag{2.35}$$

where S is the WDN entropy; S_i is the entropy of node i; P_i is the fraction of total flow through the system that reach node i and can be described as ratio T_i/T; T_i is the total flow reaching node i; T is the sum of nodal demands; N is the number of nodes in the system; S_0 is the entropy of the source supplies and can be expressed as

$$S_0 = -\sum_{i \in I} \frac{Q_{0i}}{T} \ln\left(\frac{Q_{0i}}{T}\right) \tag{2.36}$$

where Q_{0i} stands for external inflow while I is the set of the source nodes. The entropy of the nodes can be expressed analogously

$$S_0 = -\frac{Q_{i0}}{T_i} \ln\left(\frac{Q_{i0}}{T_i}\right) - \sum_{ij \in out(N_i)} \frac{Q_{ij}}{T_i} \ln\left(\frac{Q_{ij}}{T_i}\right) \qquad i = 1, \ldots, N \tag{2.37}$$

where Q_{i0} is the demand at node *i*; Q_{ij} is the pipe flow between node *i* and *j*; $out(N_i)$ is the set of all pipes from node *i*. Comparison of entropy function proposed by Awumah et al. (1990) and the model developed by Tanyimboh and Templeman (1993a, 1993b, 1993c) is available in Tanyimboh (1993).

Tanyimboh and Templeman (1993a, 1993b) developed non-iterative path-based method for calculating the maximum entropy flows for single source network. Methodology is based on fact that in single source systems all flow paths supplying the nodes start from supply node. A node-weighting technique is used for the path-based method. The single source algorithm was generalised by Yassin-Kassab *et al.* (1999) to the non-iterative method for calculating the maximum entropy flow in multiple source networks. Published results proved that that the method is rigorous and instead involving the solution of a nonlinear optimization problem solves a nonlinear system of equations.

Tanyimboh and Setiadi (2007) incorporate the application of maximum entropy approach to the optimum designs of WDNs with discrete pipe diameters. Genetic algorithm was employed in the optimization process. The approach used in this study was capable of identifying the maximum entropy design with optimum set of flow directions. Prasad and Tanyimboh (2008) also employed maximum entropy approach into optimal designs obtained after solving the well-known, complex "Anytown" benchmark problem.

Czajkowska and Tanyimboh (2012) used entropy maximization method as a surrogate reliability measure and incorporated it as one of the objective functions in the optimization process. The approach is a penalty-free multi objective optimization involving real discrete pipe diameters. Moreover, Czajkowska and Tanyimboh (2012, 2013) extended the entropy based approach into networks capable to handle multiple demand patterns. Designs achieved using multiple operating conditions (MOC) and single operating conditions (SOC) were presented and compared. In both publications (Czajkowska and Tanyimboh; 2012, 2013) was highlighted that designs based on MOC outperform solutions obtained using SOC in terms of feasibility, pipe size distribution and network reliability. Detailed analysis and further relevant results are presented in Chapter 4 of this thesis.

2.4.2.1.2 Correlation between Reliability and Entropy in Water Distribution Networks

Over the years, the entropy has been incorporated into many different benchmark networks and the relationship between this surrogate measure and reliability was tested. The relationship between entropy and reliability of WDNs has been investigated by Tanyimboh and colleagues (Tanyimboh and Templeman 2000, 1993c, Tanyimboh and Setiadi 2008b, Tanyimboh *et al.* 2011, Czajkowska and Tanyimboh 2012). The research proved that the distribution networks that are designed to carry the maximum entropy flows are reliable and that an increase in entropy value corresponds to a better network performance as measured by reliability. It also has been shown that relationship between reliability and entropy is very strong and the reliability of a distribution network improves as the entropy value of the network increases. The evidence suggests that higher entropy values increase the uniformity of the pipe diameters (Awumah *et al*, 1991; Tanyimboh and Templeman, 1993; Czajkowska and Tanyimboh, 2012a) which therefore increases the reliability.

Tanyimboh and Shehan (2002) stated that for different layouts with identical maximum entropy values designs can be expected to be hydraulically similar in general. Research carried out by Setiadi *et al.* (2005) demonstrates that the correlation between entropy and reliability is even stronger when the analysis is done using head-dependent analysis, in comparison to demand-driven analysis.

Tanyimboh *et al.* (2010) analysed the correlation of surrogate reliability measures (i.e. entropy, resilience index, network resilience index, modified resilience index) in relation to the reliability and redundancy. Results proved that entropy outperform other surrogate reliability measures. Another comparison of different reliability surrogate measures was presented in Czajkowska and Tanyimboh (2012b). Entropy correlated well with hydraulic reliability and failure tolerance while plots of other surrogate measures against reliability or redundancy had a lot of scatter. It was also found that designs that have identical maximum entropy values are generally similar with regard to their energy dissipation rates. This supports the hypothesis that entropy can be used as a surrogate measure for reliability in order to evaluate WDNs performance.

2.4.2.2 Resilience Index

Todini (2000) introduced the resilience index as a ratio of the actual power dissipated in the network to the power dissipated in order to meet the required nodal demands and heads of the network. The resilience index can be defined as (Todini, 2000)

$$RI = \frac{\sum_{i=1}^{n_n} Q_i^{req}\left(H_i - H_i^{req}\right)}{\sum_{k=1}^{n_r} Q_k H_k + \sum_{j=1}^{n_{pu}} \frac{P_j}{¥} - \sum_{i=1}^{n_n} Q_i^{req} H_i^{req}} \tag{2.38}$$

where *RI* represents resilience index; Q_i^{req} is the nodal demand; H_i is the head at demand node *i*; H_i^{req} is the demand node head above which the demand is satisfied; Q_k and H_k are the supply and head of reservoir *k*, respectively; P_j is the power introduced to the network by pump *j; γ* is the specific weight of water; n_n, n_r and n_{pu} are the number of demand nodes, number of reservoirs and number of pumps, respectively.

2.4.2.3 Modified Resilience Index

Jayaram and Srinivasan (2008) queried ability of resilience index (Todini 2000) to handle uncertainties for networks with multiple source systems. They proposed modified resilience index, which is altered version of the resilience index, and measure the surplus power as a percentage of the power required at the nodes. Modified resilience index is defined as follow

$$MRI = \frac{\sum_{i=1}^{n_n} Q_i^{req}\left(H_i - H_i^{req}\right)}{\sum_{i=1}^{n_n} Q_i^{req} H_i^{req}} \tag{2.39}$$

where *MRI* stands for modified resilience index.

2.4.2.4 Network Resilience

Prasad and Park (2004) extended the resilience index formulation by including the effect of reliable loops. They introduced the node uniformity coefficient to quantify the uniformity of the diameters of the pipes connected to the node. The node uniformity coefficient was defined as the ratio of average diameter to the maximum pipe diameter. The network resilience equation can be expressed as (Prasad and Park, 2004)

$$NR = \frac{\sum_{i=1}^{n_n} Cu_i Q_i^{req}\left(H_i - H_i^{req}\right)}{\sum_{k=1}^{n_r} Q_k H_k + \sum_{j=1}^{n_{pu}} \frac{P_j}{\text{¥}} - \sum_{i=1}^{n_n} Q_i^{req} H_i^{req}} \tag{2.40}$$

where *NR* is the network resilience and Cu_i is the uniformity coefficient of node *i*.

2.4.2.5 Surplus Power Factor

Vaabel *et al*. (2006) introduced the surplus power factor, which is based on energy transmission of flow and hydraulic power. The surplus power factor has an advantage over other network resilience measures since the pressure heads at the outlet of WDS do not need to be known. The surplus power factor can be defined as (Vaabel *et al*. 2006)

$$s = 1 - \frac{a+1}{a}\left(1 - \frac{1}{a+1}\frac{Q_{in}^{a}}{Q_{\max}^{a}}\right)\frac{Q_{in}}{Q_{\max}} \tag{2.41}$$

where s represents surplus power factor and can vary from 0 to 1 (i.e. $s = 1$ indicates that the system works on its maximum hydraulic capacity), a is the flow exponent dependent on the head loss equation used; Q_{in} is the inflow of the pipe; Q_{max} is the flow that gives the maximum hydraulic power at the outlet of the pipe.

2.4.2.6 Energy Dissipation

Rowell and Barnes (1982) suggested that the efficiency of a pipe can be measured from the rate at which it dissipates energy. For the same rate of flow, the more the energy dissipated by the network the higher the stress levels experienced by the network. Following this approach, Tanyimboh and Templeman (1993) compared alternative designs for water distribution systems. For a given set of flows, the total energy dissipated can be calculated using the following function (Rowell and Barnes, 1982)

$$E = \rho g \sum_{ij \in IJ} q_{ij} h_{ij} \tag{2.42}$$

where E represents energy dissipated, ρ is the density of water; g is the gravitational acceleration; q_{ij} and h_{ij} are the flow rate and head loss in link ij, respectively; and IJ represents the set of links in the network.

2.3 CONCLUSION

Optimization of water distribution network belongs to NP-hard problems that require employing efficient and reliable optimization methods. Classical optimization techniques have many restrictions and are actually unable to solve effectively multi-objective, highly constrained non-linear problems. Genetic algorithm (GA) appears to be efficient in searching discontinues decision spaces and solving complex, multi-objective optimization problems. General GA procedure, alongside with operators involved, have been described in detail. Additional attention was paid to search space reduction methods and constraint handling techniques reported in the literature. The limitation of those procedures has been highlighted. Several commonly employed in WDN multi-objective GA were described.

Second part of the chapter presents fundamentals for WDN modelling. Governing equation and the formulation of a system of non-linear hydraulic equations has been included alongside with several methods for solving them. Restrictions of steady state simulation have been emphasized reference to extended period simulation. Two analysis method were presented. The limitation of conventional demand driven analysis (DDA), unable to simulate system under pressure deficient conditions was highlighted in comparison to head dependent analysis (HDA). Importance of using multiple operating conditions (MOC) over single operating conditions (SOC) has been pointed out.

There is high chance that WDN will be subjected to pressure deficient situation in which some components could unavailable (i.e. pipe burst, pump failure, system maintenance). Therefore, network performance assessment is crucial to estimate the ability of the system to meet required demand under these circumstances. Two accurate performance assessment parameters such as hydraulic reliability and failure tolerance were described in details. Reliability evaluates network performance under normal and abnormal operating condition whilst failure tolerance measure the redundancy available in the network. Due to high computation burden associated with WDN reliability calculation several surrogate measures presented in the literature has also been explained in this chapter. Statistical entropy was described in detail as it is employed in research presented herein. Also correlations between hydraulic reliability versus entropy and reliability vs other surrogate measures has were compared.

CHAPTER THREE

MAXIMUM ENTROPY DESIGN OF WATER DISTRIBUTION NETWORKS UNDER MULTIPLE OPERATING CONDITIONS

3.1 INTRODUCTION

As reviewed earlier in Chapter 2 the entropy formulation was tested on many different networks over the years. Results published in the literature prove that statistical entropy function developed by Tanyimboh and Templeman (1993a) is excellent candidate as a surrogate measure for reliability. Nevertheless, most of the work has been done for single operating condition (SOC), usually referred to as steady-state simulation, which assumes that nodal demands are constant. It is common practice to use maximum daily demand and steady state modelling in designing WDN. However in reality demands vary with the time of the day and there are many different loading patterns that have to be satisfied by the network.

Alperovits and Shamir (1977) suggested that when designing the WDN, not only maximum daily demand and fire flow should be included, but also the minimum demand periods have to be considered. In addition, Prasad (2010) proved that even if a network satisfies peak demands it does not follow that other operating conditions will be satisfied, as the pressure constraints might not be satisfied. Prasad (2010) also demonstrated that designs obtained using multiple operating conditions (MOC) are more reliable than the ones achieved using SOC.

This chapter presents multi-objective genetic algorithm that can work under multiple operating conditions for any given network. The model is applied to three well-known WDNs. The results achieved show that solutions obtained using MOC outperform the solutions based on SOC in terms of hydraulic feasibility, pipe size distribution and reliability. Moreover, the investigation of three different methods for handling the entropy for multiple loading patterns is demonstrated.

3.2 PROBLEM FORMULATION

The optimization of the WDN design is extremely difficult, since it involves multiple objectives that are usually conflicting with each other. For example, minimization of cost and maximization of flows are contradicting. Therefore, there is no single, ideal result since design that is very good for one objective could be bad for others.

Another important problem associated with multi-objective genetic algorithm is the poor ability to handle constraints that are mainly carried out by penalizing infeasible solutions. Therefore, it could obstruct the search capabilities and may direct to suboptimal designs. To ensure that achieved design is optimal or near optimal, the approach does not assign any penalties (i.e. constraint violation penalties are not used in the present approach).

The objectives considered in proposed approach are minimization of the network's initial construction cost, subject to ensuring adequate pressures at all nodes and maximization of entropy. The overall problem formulation can be summarized as follows.

Minimize initial cost:

$$f1 = \sum_{i=1}^{np} C_i(D_i, L_i) \tag{3.1}$$

where $C_i\,(D_i, L_i)$ is the cost of the pipe i with diameter D_i and length L_i; np represents number of pipes in the system. Above formulation is subject to following constraints:

1. Nodal mass balance (Eq. 2.1, Chapter 2) and energy conservation (Eq. 2.2, Chapter 2), that are satisfied externally by EPANET 2 hydraulic solver (Rossman 2000);
2. Discrete pipe sizes selected from a set of commercially available sizes which are included within NSGA II code; and
3. Minimum pressure at critical node must be greater than or equal to the desired pressure at that node in order for the design to be feasible. This hydraulic feasibility constraint was introduced as the objective function and stated below.

Minimize infeasibility: $$f2 = H_i^{des} - H_i\ ; \qquad H_i < H_i^{des} \tag{3.2}$$

where i is the critical node; H_i is the available head at node i; and $_i$; H_i^{des} is the desired head at node i. The desired head is the nodal head above which the demand is satisfied in full and the critical node is the node with the lowest pressure within the network.

Maximize Entropy: $$f3 = S \tag{3.3}$$

where S is the entropy.

3.3 PROBLEM SOLUTION AND METHODOLOGY

The non-dominated sorting genetic algorithm (NSGA II) developed by Deb *et al.* (2002) was chosen for this research as a multi-objective optimization tool. The NSGA II has been widely used by many researchers in various disciplines since it is an efficient evolutionary algorithm based on pareto-rank sorting and elitism approach. The general binary coded NSGA II written in C++ language was modified in this research for the

WDN purposes and coupled with hydraulic simulation software, EPANET 2 (Rossman, 2000) to enable the approach to handle the various system components in the design and optimization process.

In order to achieve cost efficient and reliable solutions, the network performance measure needed to be include within the algorithm. As mentioned earlier, entropy has been chosen as surrogate measure of reliability. Therefore, the external program, that can calculate entropy for any given layout, was developed and tested. Subsequently the subroutine was integrated with NSGA II and EPANET 2. Finally, certain changes within the code were performed allowing the algorithm to handle SOC as well as MOC. Created genetic algorithm can work under any number of operating conditions for any network layout solving multiple objective problems.

The NSGA II is an algorithm that directs the search into objective minimization. Therefore, the objective functions implemented within the code have to satisfy such requirement. Cost minimization is the most straightforward objective, as it only requires implementation of the equation used. In terms of deficit (i.e. hydraulic infeasibility), all values for pressure shortfall are converted into positive values, whilst surplus heads (positive value in reality) are assigned zero value. In such way, the algorithm concentrates on achieving feasible solutions. In case of the entropy, the positive entropy value is converted into negative value. Therefore directing the search to objective minimization leads in reality to entropy maximization.

The procedure of algorithm used in this study is illustrated in Figure 3.1. Initially, a random parent population of size N is generated. To create offspring population of size N, the mutation and crossover are applied. The offspring and parent populations are combined together forming population of size $2N$ which undergoes non-dominated sorting. This step ensures elitism by preserving previous and current best individuals. Non-dominated sorting involves dividing results into different fronts and assigning fitness values (i.e. ranks). The first front is non-dominated with assigned fitness value of 1. The second front has assigned fitness value 2 and its individuals are dominated by individuals from first front. This goes on until all results have designated fitness values. In addition, the crowding distance is calculated for each solution. Crowding distance is a measure of distance between neighbouring individual. High crowding distance means that individual is from less crowded area, thus having such solutions results in better diversity. Based on fitness values and crowding distance of the last front, the best N individuals are chosen. The new generation is filled by each front subsequently until the number of individuals exceeds the population size N. If that happens when results from front i were added, then the individuals from front i are chosen based on crowding distance (i.e. results with higher crowding distance will be selected). Therefore, created population include best possible individuals ensuring diversity among solutions. Then the process is repeated in subsequent generations.

All designs (i.e. initial parent population and child population) go through hydraulic simulations performed by EPANET 2. Depending on the number of loading patterns involved, each operating condition of each design undergoes nodal pressure and pipe flow evaluation (i.e. in case of three operating conditions, one design goes through EPANET 2 three times), as it is necessary to ensure that pressure constraints are satisfied for all loading patterns. In other words, the hydraulic calculation is repeated until the value representing the actual operating condition (OP) is not lower than the value representing maximum operating condition (MaxOP). As a result, each objective function ended up having several values (i.e. depending on the number of operating conditions). Then, to reduce several values (i.e. each representing one loading pattern) into one value for each objective function, different methods have been employed and applied as follows. In case of node pressure deficit, the algorithm chooses the maximum deficit, so if nodal pressure for any loading pattern is lower than desired pressure, the design is treated as infeasible. The cost does not require further analysis, as each operating condition has identical cost for the particular design (i.e. the same set of pipe diameters). For the entropy, three different scenarios were chosen and explained later in paragraph related to MOC entropy approaches (Section 3.5). All three cases were investigated and results presented in the examples (Section 3.6).

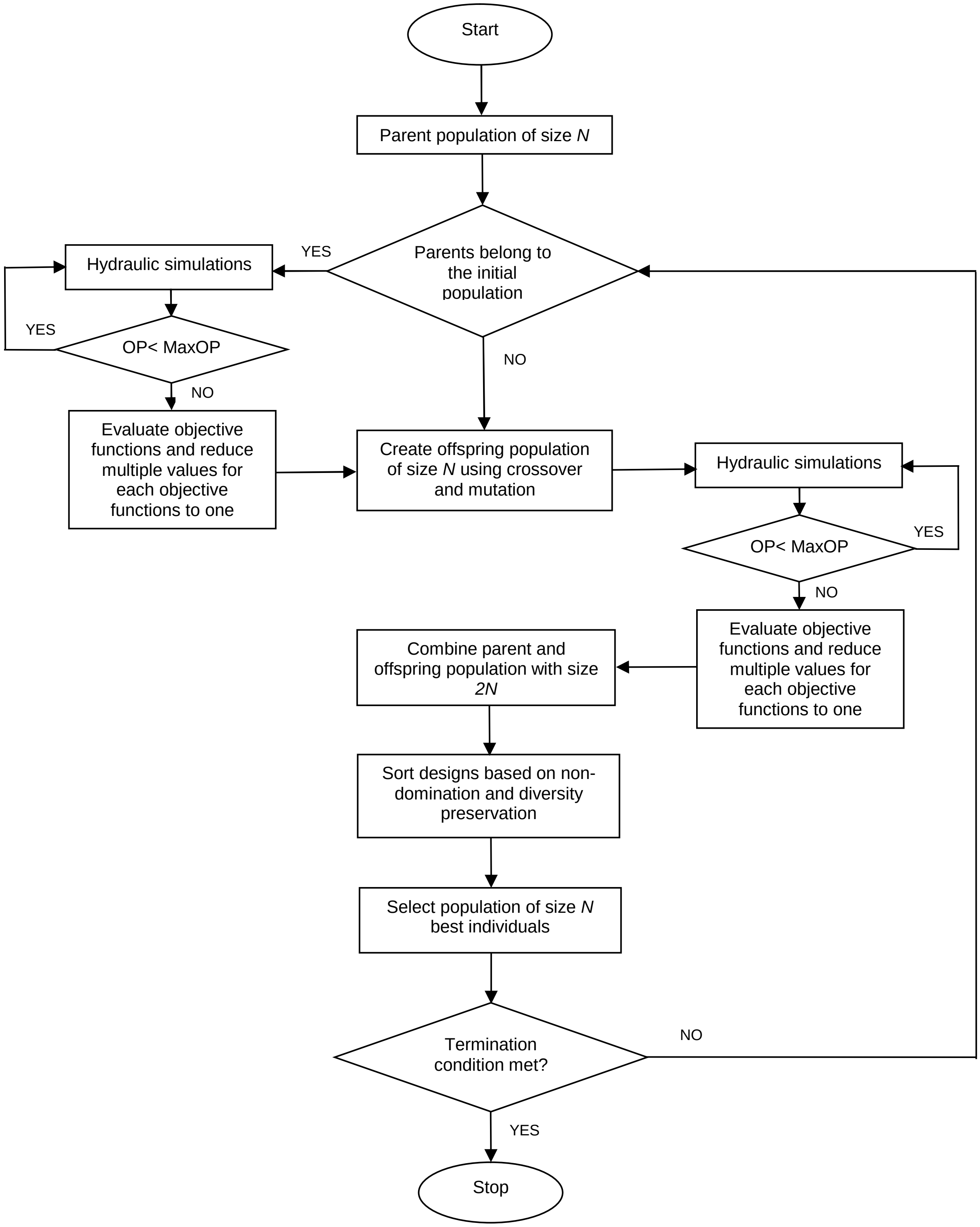


Figure 3.1 Flow chart of presented algorithm

It was observed that the final set of solutions (i.e. last generation) provided by NSGA II in output file is not the same as the feasible set of cost-entropy non-dominated solutions. In other words, there are many infeasible designs in final POF. This may be because the optimization model involves three objectives that are contradicting with each other (i.e. capital cost minimization, statistical entropy maximization and critical node pressure deficit minimization). Moreover, the algorithm does not prioritize any objective function and handle them with equal importance. Thus, even infeasible solutions with low cost and high entropy are treated as valuable; hence they appear in final POF (i.e. last generation). This could be beneficial for the decision maker if infeasible solution with insignificant deficit would be accepted due to low budget level or other cause. Nevertheless, presented research relies mostly on feasible designs, which occur in a small number in the final POF generated by the algorithm. This could be not enough to represent the entire range of possible entropy values for particular network. Thus, the whole analysis process would be more difficult and misleading to drawn proper conclusions.

In order to gather many feasible, non-dominated solutions, the supplementary method to identify such designs was developed. The external software was written in Perl language. Firstly, the feasible designs were selected from entire history of solutions (i.e. feasible designs were picked up from 200,000 results for a GA run with termination criteria set to 200,000 evaluation functions). Then, feasible designs were sorted according to cost and entropy and only non-dominated solutions were chosen. Also the repetitive designs were rejected. This approach provided a good range of candidate solutions for further consideration via keeping the designs that would be rejected by the algorithm.

Entire process of selecting the cost-entropy non-dominated designs over the entire range of solutions is completely automated process and very quick as for a network with 8 pipes simulated for 200,000 function evaluations it requires only about 20 seconds on a PC with the following configuration: Intel Core i3, 2.4GHz, 3GB RAM and Windows XP operating system. It should be highlighted that, unless otherwise stated, all POF presented herein are based on cost-entropy non-dominated screening for the entire history of results.

Finally, the solutions obtained in this way underwent pipe failure simulations to evaluate hydraulic reliability and pipe failure tolerance. The pressure dependent WDN analysis program PRAAWDS (Program for the Realistic Analysis of the Availability of Water in Distribution Systems) has been used. PRAAWDS is a prototype FORTRAN computer program developed by Tanyimboh and Templeman (2004). Since it is based on head dependent analysis approach, it creates realistic simulations for all pressure regimes. There is possibility to choose between four head-outflow relationships, out of which one is the Tanyimboh and Templeman (2004) function. Moreover, the conventional, demand driven simulation is also included as an option. In addition, it offers a built-in feasibility procedure to evaluate the accuracy of results obtained from pressure dependent analysis. The PRAAWDS program is very easy to use and provides accurate results. Extensive testing has shown that PRAAWDS is efficient and robust (e.g. Setiadi et al, 2005; Tanyimboh and Setiadi, 2008a). Nevertheless, since the PRAAWDS depends on the number of pipes, it requires running the program for numerous times for each design. Single run of PRAAWDS provide the output where only one pipe is closed. For example, if network has 20 pipes, the program needs to be run 20 times producing 20 output files. Then all the values need to be gathered together in order to calculate hydraulic reliability or failure tolerance. It is very time consuming, especially if this needs to be done for many designs with large number of pipes (i.e. hundreds of even thousands). To speed up the process, PRAAWDS extension written in Perl language was developed in this project. The extension makes the process more automated as it replaces the value responsible for pipe closing. In other words, external program forces the PRAAWDS to change the pipe that is closed and to simulate. The process is repeated automatically depending on the number of pipes in network. What is more, all output data are gathered in one external file. Successive steps of the whole procedure (i.e. starting from GA runs and finishing on reliability and failure tolerance calculations) for identification of cost-entropy non-dominated solutions chosen from entire history of results are illustrated on Figure 3.2

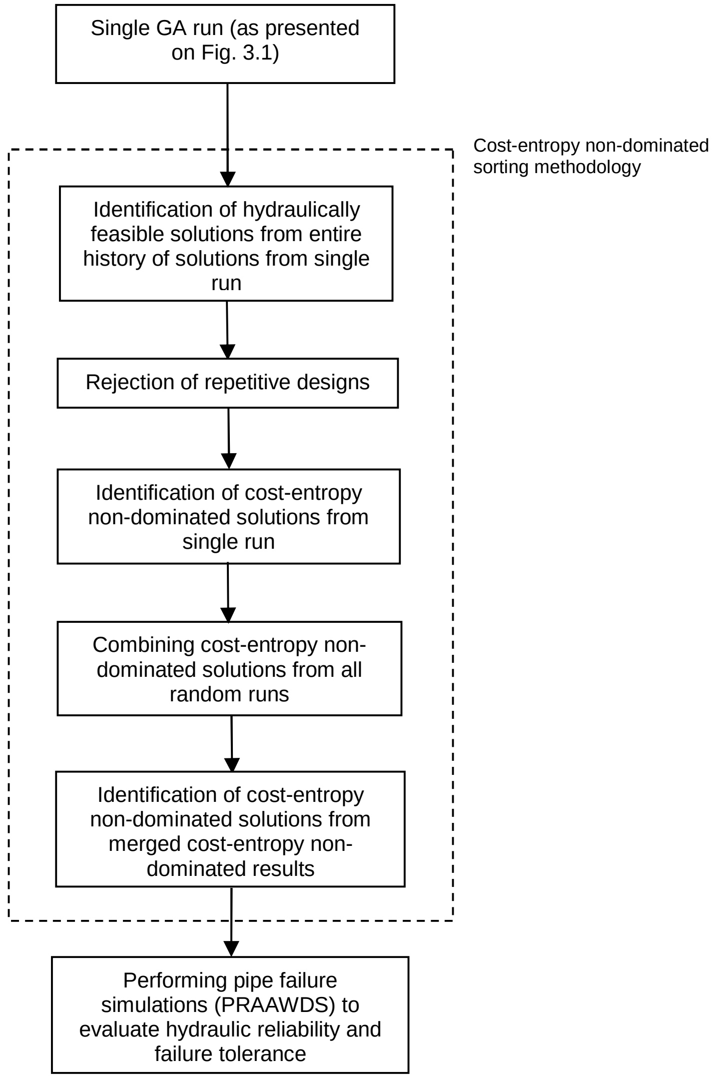


Figure 3.2 Flow chart of successive steps for identification of cost-entropy non-dominated solution chosen from entire history of results

It should be mentioned that for purposes of this study, the Tanyimboh and Templeman (2000) function (Eq. 2.30 and Eq. 2.32, Chapter 2) has been used for all hydraulic reliability and failure tolerance calculations. Therefore, the reliability is defined as the ability of the system to fulfil on average the required nodal demands at adequate pressure whilst considering both normal and abnormal operating conditions. The pipe failure tolerance measure provides an estimate of the total demand the WDN is capable of satisfying when some components are out of service.

3.4 SENSITIVITY ANALYSIS

Implementation of MOC approach required many changes within the original NSGA II code. In order to assess the robustness of created algorithm, the sensitivity analysis was performed before comparing SOC and MOC approaches. Sensitivity analysis was also used to identify the input data, such as population size, mutation rate and crossover point which led to the best results and uniform distribution of solutions in POF. It should be mentioned that sensitivity analysis was carried out for SOC only.

The well-known, hypothetical Two-Loop network (Figure 3.3) was chosen for sensitivity analysis. This single source network was first presented by Alperovits and Shamir (1977) and consists of 8 pipes of length 1000m and 6 demand nodes. The minimum pressure requirement for all nodes is defined as 30m. A Hazen-Williams roughness coefficient of 130 is used for new pipes. The details of 14 discrete pipe size used, as well as cost of these pipes and nodal data can be found in Appendix A. It should be mentioned that for sensitivity analysis only columns with node number, elevation and demand pattern 1 (i.e. data that corresponds to peak demand used in literature for Two-Loop network) were used.

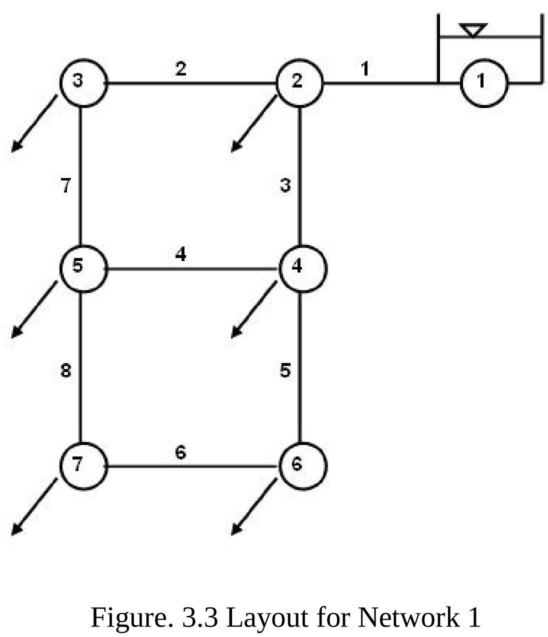


Figure. 3.3 Layout for Network 1

For this network 10 random runs were performed for each case (i.e. each POF presented on graphs is made from 10 random runs). Initial testing showed that for such small network even single run gives quite uniform spread of results in POF. However, for completeness, 10 random runs were initialized. Considering the fact that mutation rate, crossover point and other parameters were investigated; around 180 GA runs were performed. A total of 200 000 function evaluations (i.e. a population size of 200 for 1 000 generations) were allowed for most of the runs. Only the GA runs where the different sizes of population were examined (Figure 3.4), have function evaluations that vary from 100 000 to 500 000 (i.e. number of generations remain the same and was set to 1 000). Unless otherwise stated the probability of crossover and mutation were set to 1.0 and 0.03125 respectively. A 4-bit binary substring was used, thus giving 16 substrings (i.e. 2^4). Having 14 pipe sizes left 2 substrings redundant. Those redundant codes were uniformly allocated to available pipe sizes (i.e. pipes with diameter 152.4mm and 406.4mm were doubled). However, for GA runs where allocation of redundant codes is tested, the 2 redundant substrings were allocated to the lowest pipe size, thus pipe with diameter 25.4mm was tripled. The solution space for this network comprises $14^8=1.48\times10^9$. Average CPU time required for single GA run was about 10 minutes on a PC with the following configuration: Intel Core i3, 2.4GHz, 3GB RAM and Windows XP operating system. Only the GA runs with different population sizes (Figure 3.4) have different CPU times that depend how large the population size was.

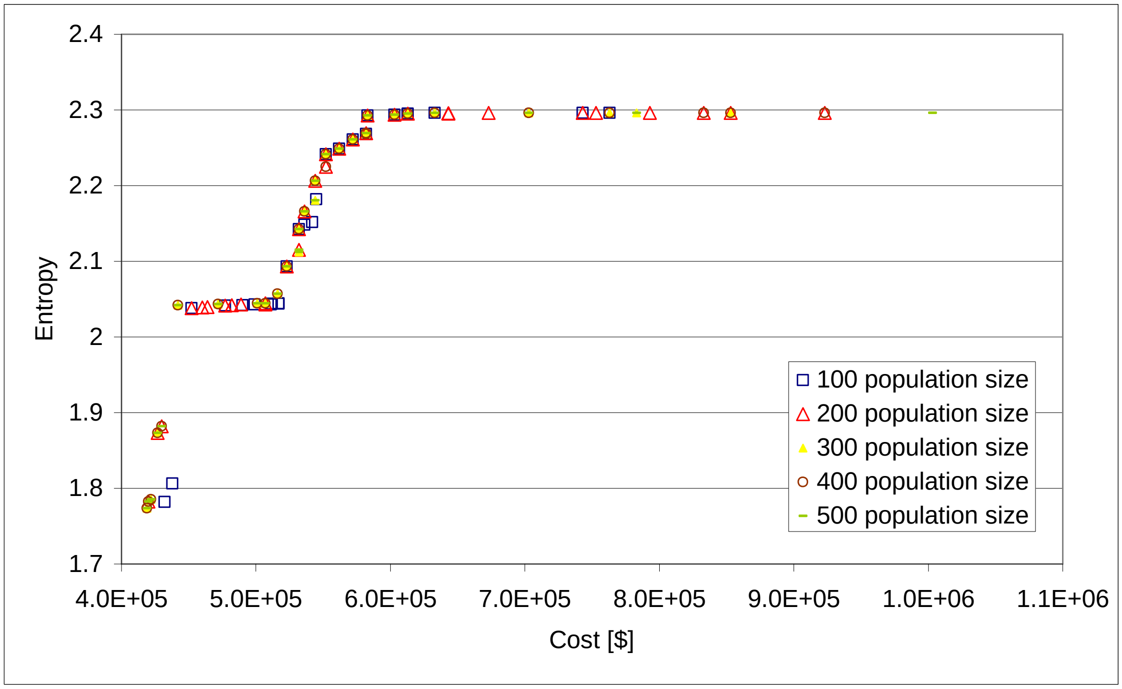


Figure 3.4 Entropy - cost POFs for different population sizes

Comparison of POFs achieved for various population sizes is presented on Figure 3.4. As mentioned earlier, each POF is based on cost-entropy non-dominated solutions chosen from entire history of results for 10 randomly initialized GA runs. It is easy to notice that majority of results overlaps. Only the POF based on population size 100 has small deviation from other POFs. Few individual solutions have higher cost for similar entropy value than the solutions achieved using population size 200 or more. Therefore, if all five POFs would be merged together the results achieved with population size 100, would be dominated. It is also worth to mention that GA with population size 500 generated solution with insignificant increase in entropy value and unnecessary growth in capital cost. The POF achieved using population size of 200 is equally good as POFs obtained with higher population sizes. Moreover, the POF obtained using population size of 200 has more non-dominated solutions than any other POFs. Therefore, it has been decided that there is no point to use value higher than 200, especially that the time length for single GA run increase with the increase of population size (i.e. GA run with population size of 200 takes about twice longer than GA run with population size of 100, etc.).

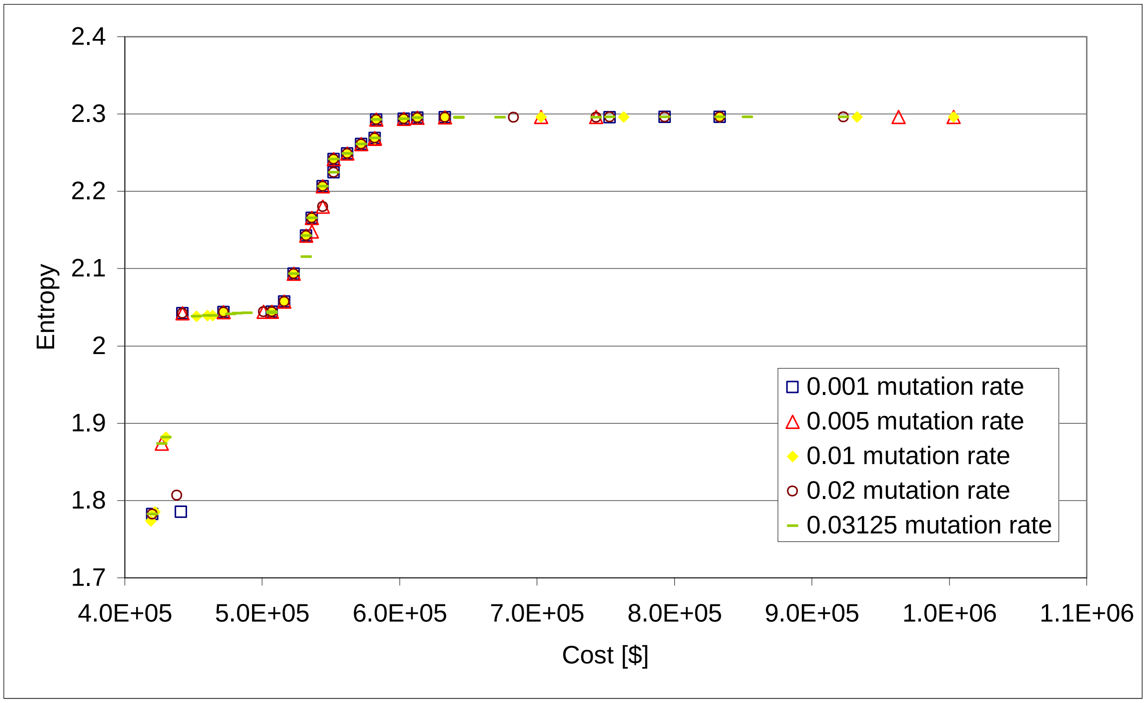


Figure 3.5 Entropy – cost POFs for different mutation rates

For this particular network, the mutation probability is $1/n_g$ = 1/32 (i.e. where n_g is the chromosome length), thus the highest value for mutation is 0.03125. To identify the most efficient mutation rate, different values were tested, starting from the very low such as 0.001 (i.e. 0.1% chance that any single bit would mutate) and finishing on 0.03125 (i.e. 3.125% chance that any single bit would mutate). Achieved POFs were gathered together and presented on Figure 3.5. It is easy to observe that despite the mutation rate, the results produce essentially one front with only two solutions located outside this front. Those outliers belong to GA runs with mutation rates set to 0.001 and 0.02. It has also been noticed that higher mutation rate used, the more non-dominated solutions in POF. Thus, using 0.001 as mutation rate led to 22 solutions; 0.005 - 25 solutions; 0.01 - 27 solutions; 0.02 - 28 solutions and 0.03125 - 35 solutions). Therefore it is worth to use the highest allowed mutation rate in order to achieve more feasible, non-dominated solutions in POF.

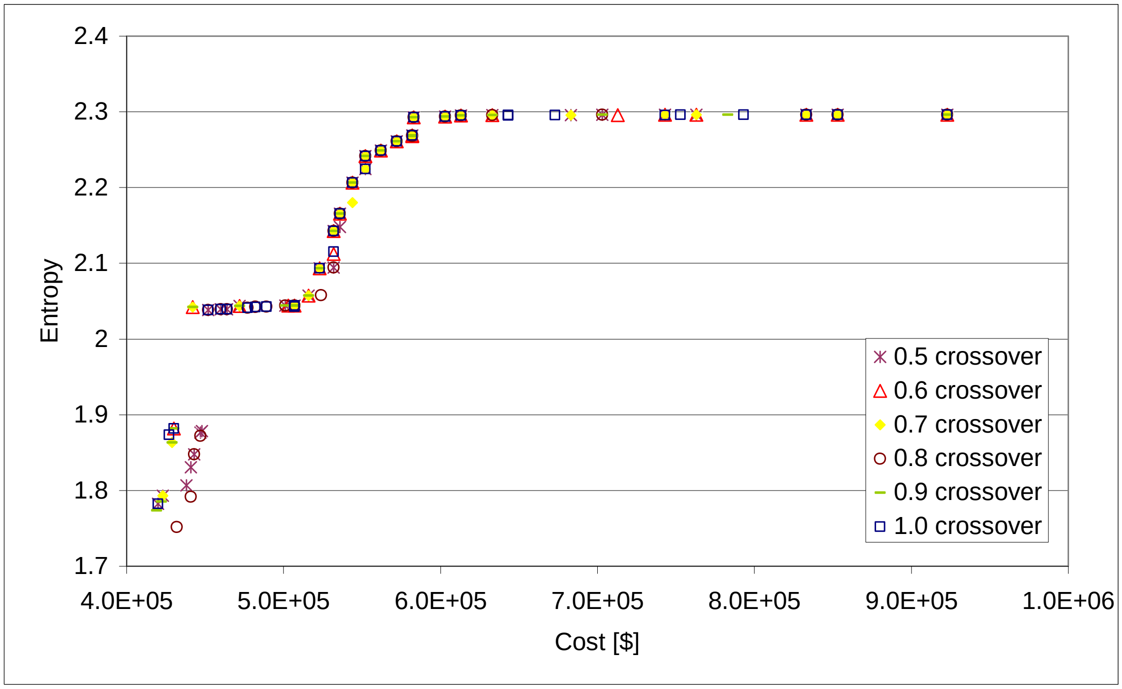


Figure 3.6 Entropy – cost POFs for different crossovers

In order to analyse different crossover rates, single-point crossover operator was used to produce two offspring from two parents. Six different cases were studied with crossover rate varying from 0.5 to 1.0 and the outcome presented on Figure 3.6. Before merging all POFs it has been noticed that the GA runs with crossover rate set to 1.0 produced front with most uniform spread of results. Moreover, using 1.0 as crossover rate led to the largest number of designs in cost-entropy non-dominated POF.

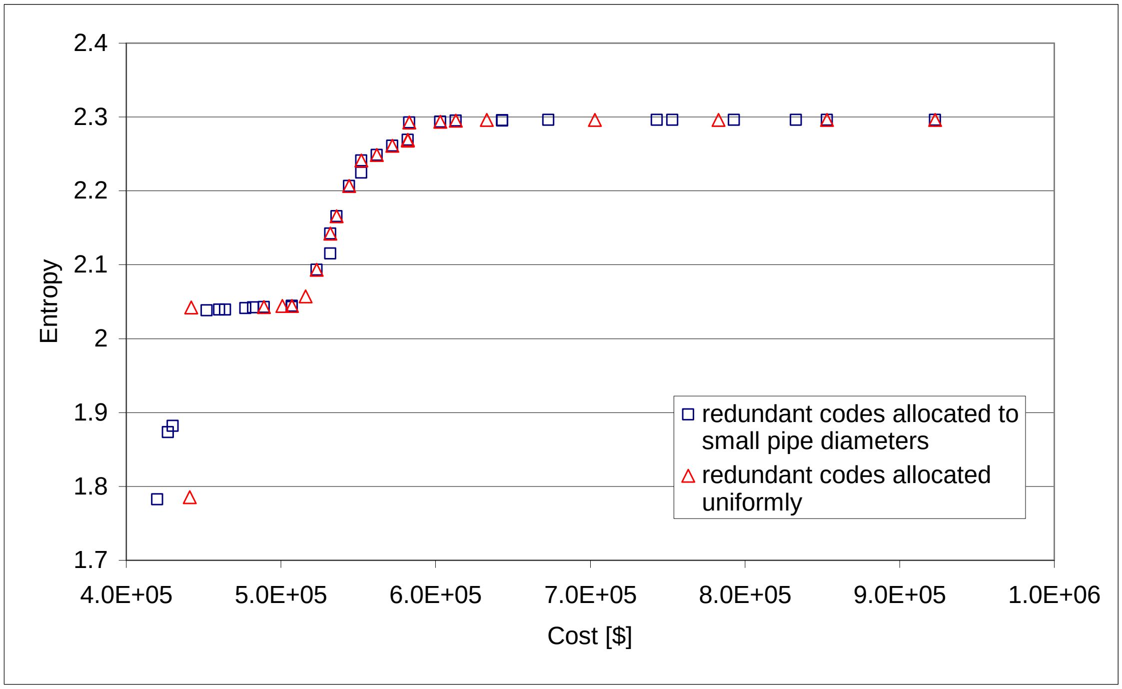


Figure 3.7 Entropy – cost POFs for differently allocated redundant codes

Figure 3.7 present two POFs obtained with the use of the same input parameters (i.e. mutation rate, crossover, random seeds) but different allocation of redundant codes. As mentioned earlier, having 14 pipe sizes left 2 substrings redundant (i.e. 2^4 =16). In first case (i.e. blue squares on the plot), the redundant codes were allocated to lowest pipe size, thus pipe with diameter 25.4mm was tripled. In second case (i.e. red triangles), 2 redundant substrings were uniformly allocated to available pipe sizes (i.e. pipes with diameter 152.4mm and 406.4mm were doubled). It can be noticed on Figure 3.7 that results of both cases mostly overlaps creating uniform front. POF obtained from runs where lowest pipe was tripled has more low cost, low entropy solutions. It seems logical as allocating three codes to lowest pipe caused that algorithm favour solutions with small pipe diameters. However, such situation should be avoided and the redundant codes should be allocated uniformly.

Based on two loop network and presented results it has been decided that for all three network layouts presented in this chapter, the input parameters will be kept the same. It is no doubt that the best values for mutation rate, crossover and population size could slightly vary for different networks. However, all examples of networks presented here are quite small networks, so it is though that the differences in the best input data for particular network would be insignificant.

3.5 FORMULATION OF ENTROPY APPROACH FOR MULTIPLE OPERATING CONDITIONS

As explained earlier, the number of entropy values achieved depends on number of multiple operating conditions. Hence, having three operating conditions leads to three different entropy values for a single design. In order to determine the best design criterion, a number of alternatives may be considered, including:

(a) Maximizing the maximum entropy;

(b) Maximizing the minimum entropy;

(c) Maximizing the total entropy.

3.5.1 Maximizing the Maximum Entropy

To maximize the maximum entropy, the highest entropy value is chosen through all entropies for specific design. Such case seems ideal from logical point of view, as the main aim of using statistical entropy is to maximize its value as a measure of reliability. Nevertheless, GA maximize the best possible entropy value, regardless the values for other operating conditions. In other words, the algorithm tries to find solutions which are very good for particular operating condition while disregarding rest. Thus, the solution with very high entropy value for one operating condition and low entropies for other operating conditions would dominate the solution with slightly lower entropy even if the entropy remain similar for all operating conditions.

3.5.2 Maximizing the Minimum Entropy

In this case, GA maximizes the lowest entropy value among all operating conditions. It eliminates possibility that the best entropy value will be chosen regardless values in other operating conditions. Moreover, the minimum entropy approach ensures that the results for any operating condition will not be lower than the selected one. However, it does not take into consideration operating conditions with entropies higher than the lowest entropy value. Therefore, it underestimates the designs with low entropy value for one operating conditions and high for the rest.

3.5.3 Maximizing the Total Entropy

Maximization of total entropy approach links all entropy values, instead of choosing extremes. In order to achieve the total entropy value, the individual values are added together within the algorithm. This reflects the possibility that different designs can have the same entropy value for some of the operating conditions. It is thought that the more the number of operating conditions involved the lower the chances that all corresponding entropy values will be identical for two different designs. What is more, when using total entropy approach, all entropy values are taken into consideration. Thus, the designs selected should perform reasonably for all operating conditions.

3.6 APPLICATION OF MULTIPLE OPERATING CONDITION APPROACH

To investigate the MOC entropy approaches, three standard WDS layouts documented in the literature were employed. The network presented in sensitivity analysis paragraph, the Two-Loop network (Fig. 3.3), was used as first benchmark. As a second example the Six-Loop network (Fig. 3.8) was chosen. It is no doubt that this network is also quite simple and does not reflect problems of real WDNs. Nevertheless, this benchmark is well known from previous entropy studies and has quite large number of loops as for its size (i.e. the greater the number of loops the more different flow paths possibilities, thus different entropies). The third example is the network with only three loops (Figure 3.17) but two reservoirs and three different demand patterns with associated minimum nodal heads that must be satisfied. Comparison of results achieved by using MOC with different entropy approaches and SOC were carried out and the outcome presented and discussed.

3.6.1 Example 1

The first example, shown in Fig. 3.3 is the simple Two-Loop network used earlier for sensitivity analysis. As this network has only data for one loading pattern, additional operating conditions were calculated based on different networks from literature. Thus, the basic demands for the Two-Loop network have been treated as peak demand and were used to calculate other 2 operating conditions, i.e. average demand and minimum demand. Demand multiplier for average demand was taken as 0.8 and obtained from Surendran *et al.* (2005). To achieve minimum demand, the average demand was multiplied by 0.6 as in well known "Anytown" network (Walski *et al.,* 1987). Both demand multipliers were applied to all nodes, so the demands change with the same ratio. Nodal demands for all operating conditions can be found in Appendix A (Table A-1)

The termination criterion for the GA was taken as 200,000 function evaluations (i.e. 1000 generations for a population size of 200) for all four cases (i.e. SOC and three different entropy approaches for MOC). 10 randomly generated GA runs were performed for each case, giving 40 GA runs in total. The probability of crossover and mutation were set to 1.0 and 0.03125 respectively. The redundant codes were uniformly allocated. Average CPU time required for one operating condition was about 11 minutes and 23 minutes for three operating conditions, on PC with following configuration: Intel Core i3 @ 2,4GHz and RAM 3GB.

It should be mentioned that the entropy value for the results obtained using total entropy approach were 3 times higher than the other cases (i.e. maximum entropy approach, minimum entropy approach and SOC), simply because the entropy values have been added together within the GA in the case of MOC. However, to facilitate the comparison of solutions obtained using SOC and MOC entropy approaches, the entropy values achieved using total entropy approach have been divided by 3 before analysis. This applies to all three examples presented in this chapter.

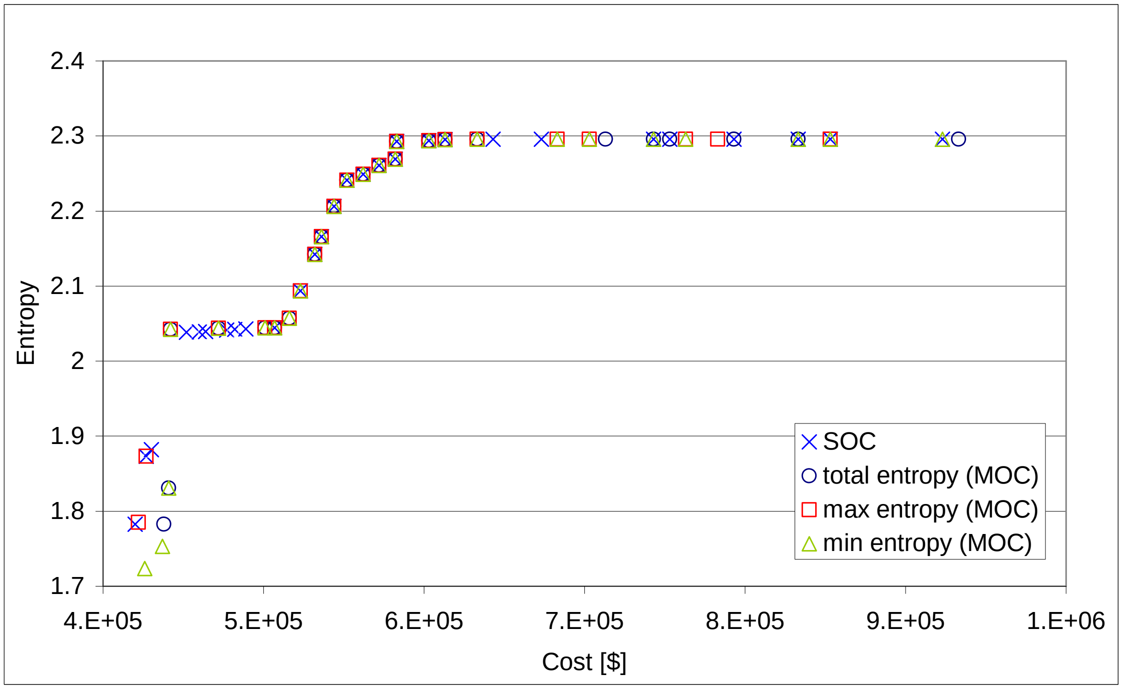


Figure 3.8a Entropy-cost POFs based on entire history of results for whole range of results

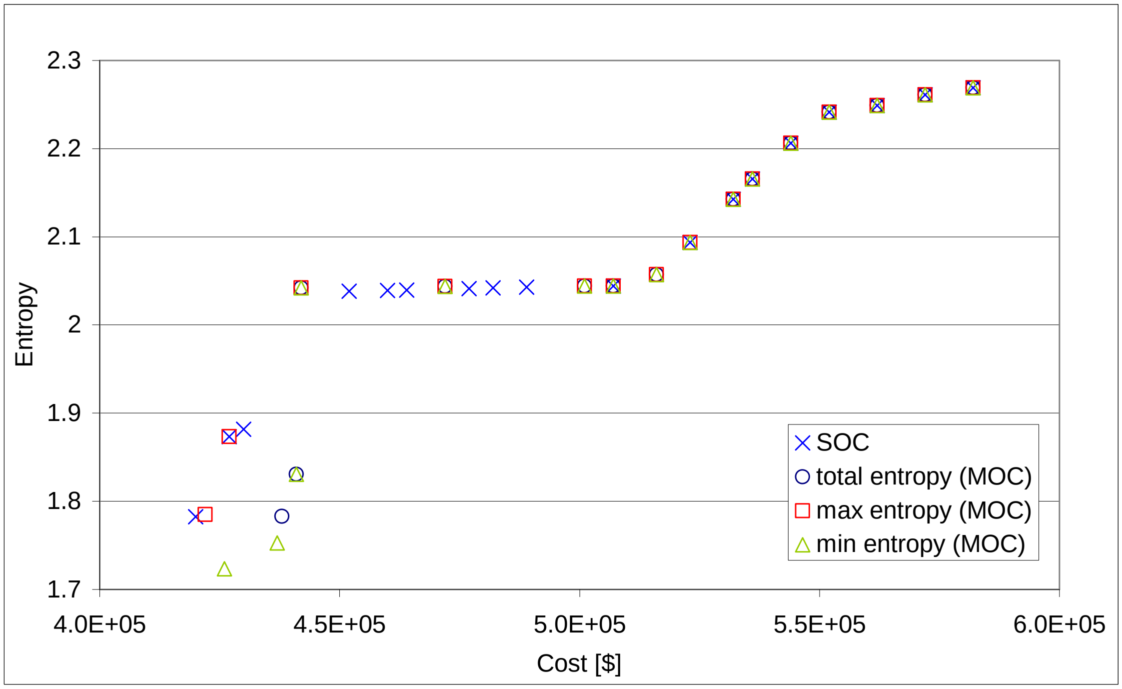


Figure 3.8b Entropy-cost POFs based on entire history of results
for results up to the ESP

Figure 3.8a illustrates POFs of cost versus entropy for SOC and three entropy approaches for MOC, while Fig 3.8b presents the same results but only up to the point beyond which the entropy improvements become very insignificant. There is quite a lot results that have very minor increase in entropy value while high increase in cost. Those designs will always appear in optimization process due to nature of NSGA II that maximizes objective functions. Therefore, the GA tries to reach to the highest entropy value, which will not vary much once near global maximum entropy value is achieved. Since the cost-entropy non-dominated screening is performed on entire history of results and outside the GA, all solutions with increasing entropy value (i.e. as long as are non-dominated) are kept despite the low improvements in entropy. As a result of that, the Entropy Stagnation Point (ESP) was identified as the point beyond which the improvements in entropy value become insignificant. For presented Two-Loop network ESP is the 99% of maximum entropy value achieved.

It can be observed on Fig. 3.8a and Fig. 3.8b that designs based on MOC are very close to SOC solutions in terms of entropy value. Only few solutions achieved using minimum and total entropy approach are located slightly below POFs, thus those designs would be dominated if all POFs would be merged. Nevertheless, the minimum, maximum and total entropy approaches satisfy 3 loading patterns, whilst the designs based on SOC would generally be infeasible for MOC.

Table 3.1 Coefficient of determination of entropy versus other network performance indicators for Example 1

Measure	SOC	MOC		
		Total Entropy	Max Entropy	Min Entropy
Coefficient of determination of entropy vs average pipe diameter				
100%	0.842	0.794	0.835	0.966
99%	0.893	0.899	0.904	0.918
Coefficient of determination of pipe size distribution				
100%	0.663	0.714	0.674	0.773
99%	0.874	0.892	0.875	0.922
Coefficient of determination of reliability vs entropy				
100%	0.813	0.821	0.819	0.826
99%	0.885	0.890	0.892	0.870

Table 3.1 presents coefficient of determination of entropy versus average pipe diameter, pipe size distribution and reliability, illustrated on Fig. 3.9, 3.10 and 3.11, respectively. 100% represent coefficient of determination for whole range of results, while 99% correspond to correlation for solutions up to ESP (i.e. 99% of maximum entropy value). As expected, the coefficients of determination are generally higher for results up to 99% of ME value. This reinforces the hypothesis that results beyond ESP are not necessary and should be removed for analysis purposes. Therefore, remainder of this thesis focuses on results up to the ESP.

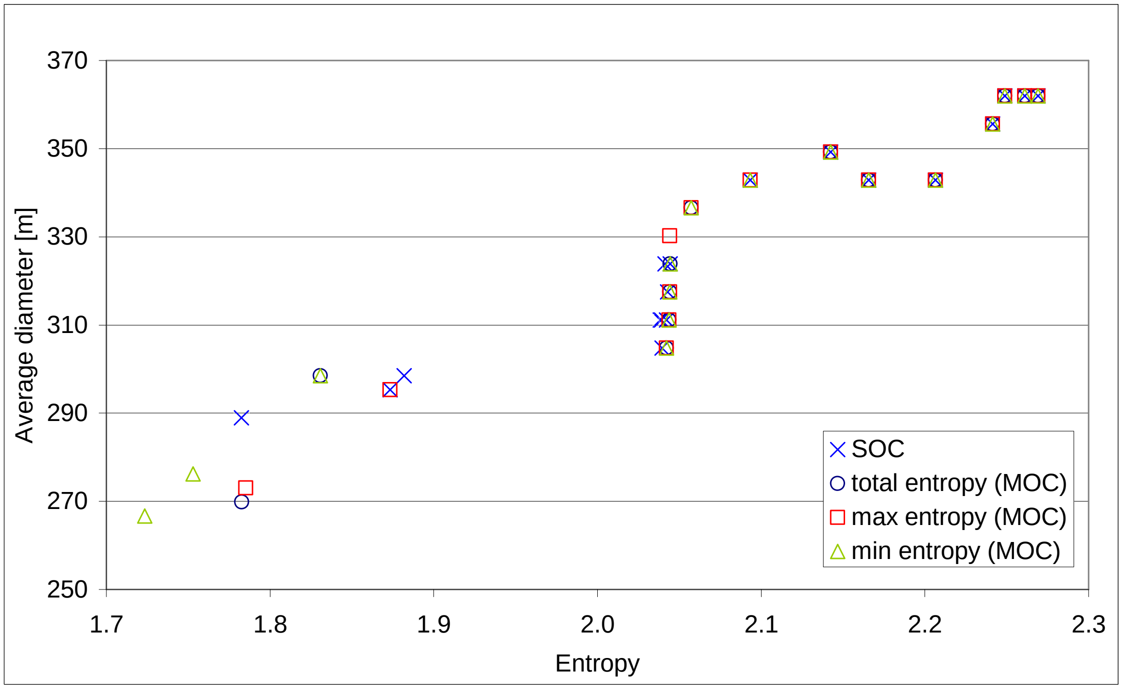


Figure 3.9 Effect of entropy on average pipe size for results up to 99% of the maximum entropy value

Figure 3.9 demonstrate relationship between entropy and average pipe diameter. Entropy increases as the pipes sizes increase, which appears to contribute to the growth of the network reliability, simply because larger pipes have lower breakage rates than the smaller ones (Tanyimboh and Templeman, 1993c; Tanyimboh and Setiadi, 2008b). What is more important, coefficients of determination (Table 3.1) are higher for three operating conditions. It strengthens previous assumption that using multiple operating conditions leads to more uniform pipe sizes. As it is already known, the more uniform pipe sizes result from more uniform flows, which yield higher hydraulic reliability (Awumah *et al,* 1991; Tanyimboh and Templeman, 1993c).

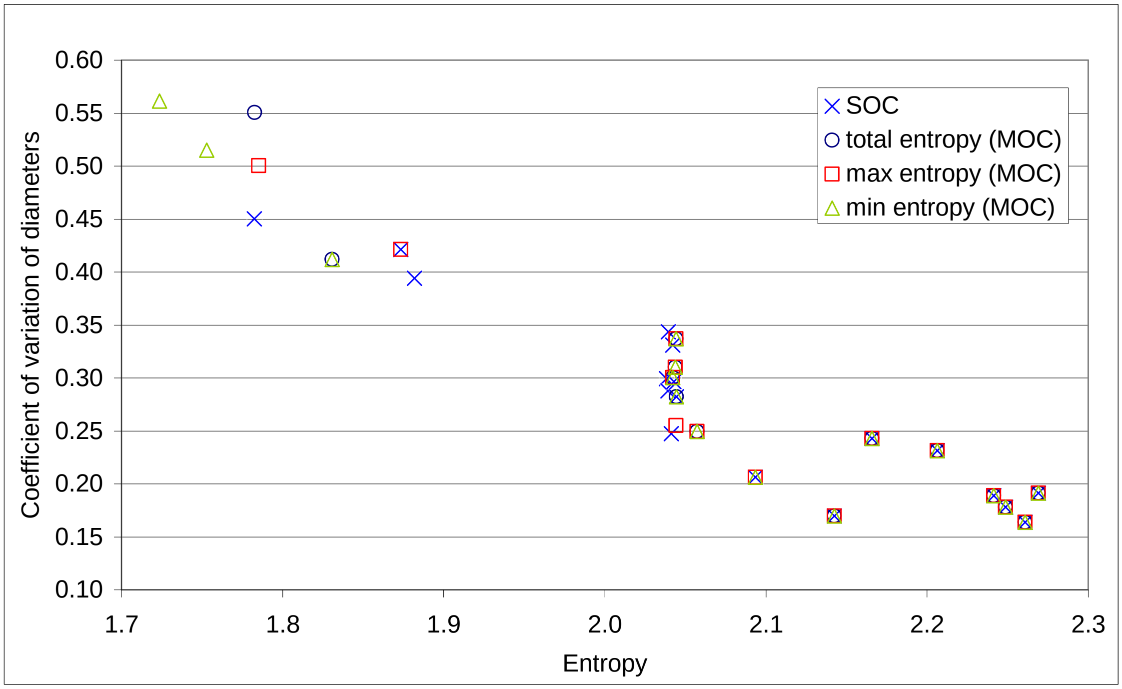


Figure 3.10 Effect of entropy on pipe size distribution for results up to ESP

A comparison of the correlation between pipe size distribution and entropy is shown in Figure 3.10. Pipe size distribution in other words means the coefficient of variation of the diameters - all the pipes have the same length - which measures the uniformity of the pipe diameters. It has been chosen to present the results as previously it has been demonstrated to be simple but very efficient way to show the best solution (Tanyimboh and Setiadi, 2008b). Higher positive correlations for designs achieved using MOC approaches (Table 3.1) suggest that solutions based on MOC outperform the ones obtained by SOC.

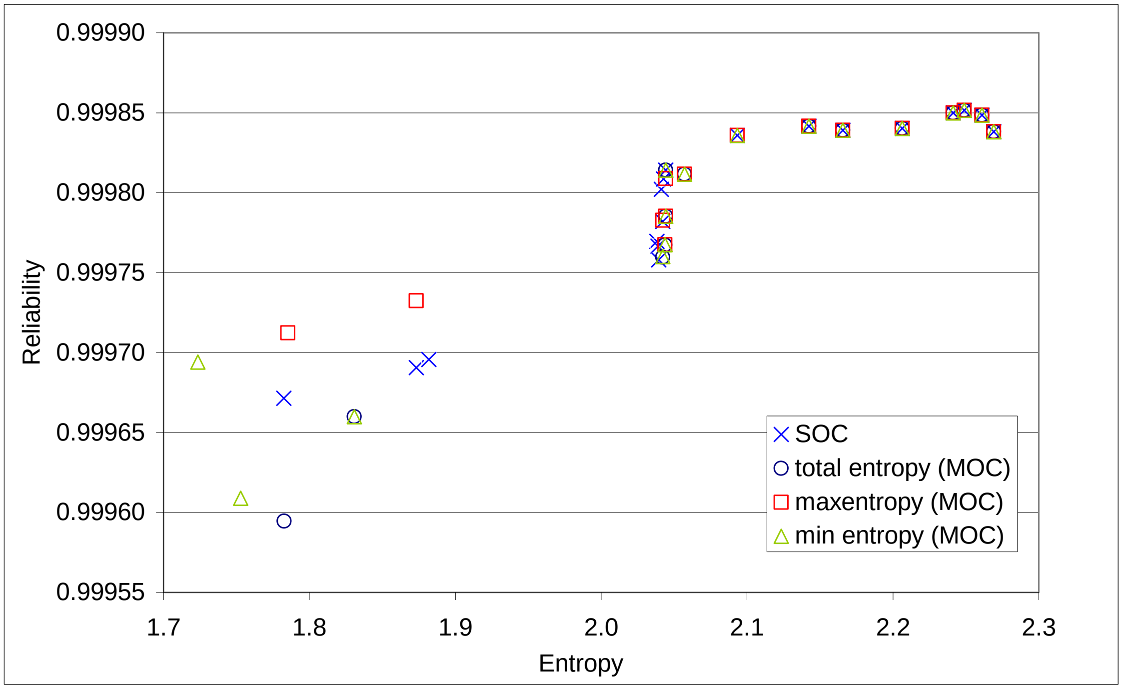


Figure 3.11 Reliability versus entropy for results up to 99% of the maximum entropy value

Figure 3.11 presents relationship between reliability and entropy for SOC and three MOC approaches. PRAAWDS, the head-dependent analysis software, was employed to calculate hydraulic reliability with peak demands used as an input data. Strong positive correlations between entropy and reliability can be observed for all MOC entropy approach as well as SOC (Table 3.1). Nevertheless, designs based on SOC produced less correlation than the ones obtained using MOC entropy approaches (Fig. 3.11 and Table 3.1). This clearly confirms that the approach based on MOC produces designs that are more reliable. Having stated previously that SOC solutions can be infeasible under MOCs, the remainder of this thesis focuses MOC and particularly on selecting the best MOC entropy approach.

As it can be seen from Table 3.1 the maximization of minimum entropy value produce results with stronger relationships between entropy vs. average pipe diameter and coefficients of variation of diameters. However, for the results up to the ESP (i.e. 99%), the maximization of total entropy value has the strongest correlation for reliability vs. entropy. Since there is no clear indication which entropy approach is the best, additional network should be analysed before the conclusion could be made. Moreover, the Two-

Loop network is a small network with very limited possibilities for different flow paths, thus network with higher number of loops should be used.

3.6.2 Example 2

The second example illustrated in Figure 3.12 is a hypothetical Six-Loop network. This network was extensively used in entropy studies (Tanyimboh and Sheahan, 2002; Setiadi *et al*. 2005; Tanyimboh and Setiadi, 2008b; Czajkowska and Tanyimboh, 2012) as despite small number of nodes it has considerable number of loops. Therefore, it has many possibilities for flow paths, thus different entropy values. It was first presented by Awumah *et al.* (1991) and consists of 12 nodes, 17 pipes and single source with total head at 100m. The elevation and required head at all demand nodes are 0m and 30m, respectively. All pipes are 1000m long and have a Hazen-Williams roughness coefficient of 130. A set of 12 commercially available pipe sizes in the range of 100mm to 600mm was used (100, 125, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600).

Having 17 different pipes with 12 possible pipe diameters gives a search space of $(12)^{17}$ or 2.218×10^{18}. A 4-bit binary substring was used, thus giving 16 substrings (i.e. 2^4). Having 12 pipe sizes left 4 substrings redundant. Those redundant codes were uniformly allocated to available pipe sizes (i.e. pipes with diameter 125mm, 250mm, 400mm and 550mm were doubled). The termination criterion for the GA was taken as 200 000 function evaluations (i.e. 1 000 generations for a population size of 200) for all three MOC entropy approaches. 30 randomly generated GA runs were performed for each case, giving 90 GA runs in total. A single-point crossover operator was used to produce two offspring from two parents. 1.0 was used as a crossover probability. A bitwise mutation operator was used to change the bit from 0 to 1 or vice versa. Since the mutation probability was $1/n_g$ = 1/68 (i.e. 68 is the chromosome length), the mutation rate was set to 0.0147 (i.e. there were 1.47% chances that any single bit would mutate). The average CPU time required for single run was about 28 minutes, on PC with following configuration: Intel Core i3 @ 2,4GHz and RAM 3GB.

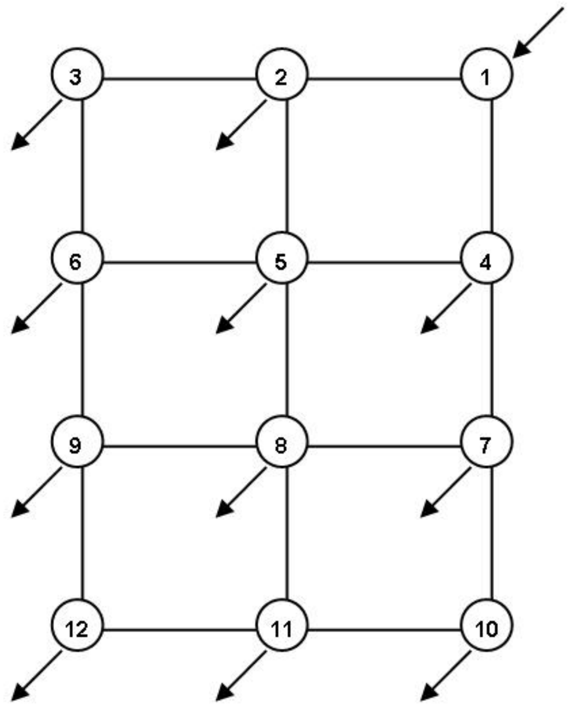


Figure. 3.12 Network layout for Example 2

As in previous network example, also for this layout the nodal demands were treated as peak demands and were used to calculate the demands for other two operating conditions, i.e. the average demand and minimum demand. Since the Six-Loop network is not much bigger than Two-Loop network, the same demands multipliers were used and applied to all nodes. Nodal demands for all operating conditions can be found in Appendix B. Also the entropy values achieved using total entropy approach has been divided by three to facilitate comparison of results.

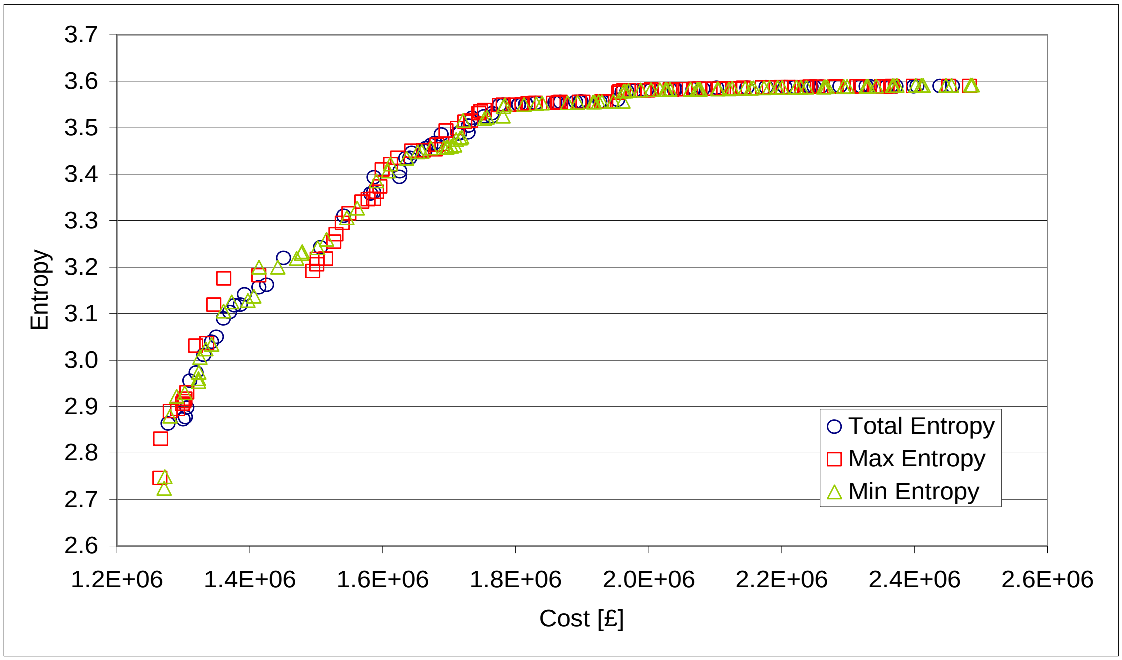


Figure 3.13a Entropy-cost POFs based on entire history of result for whole range of results

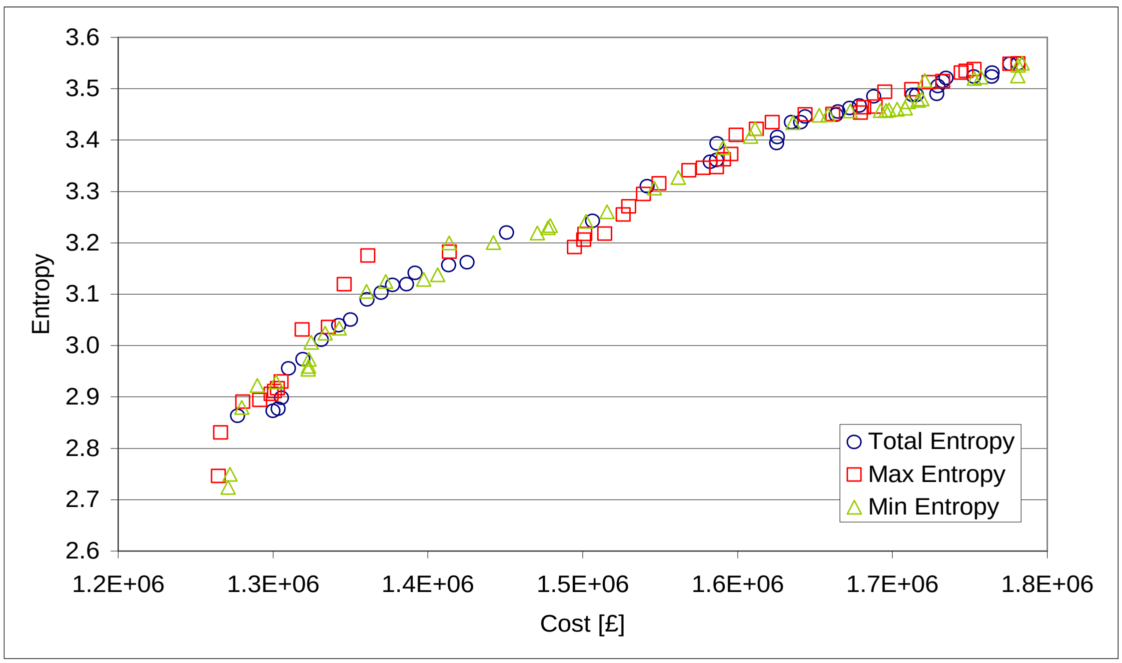


Figure 3.13b Entropy-cost POFs based on entire history of result for results up to ESP

Figure 3.13a presents cost-entropy non-dominated designs for whole range of results generated by GA. It can be noticed that improvement in entropy value become very minor around cost £1.78x10^6. This point correspond to 99% of ME value, as in previous network example. Therefore, 99% of ME value was treated as ESP and solutions up to the ESP are presented in Fig. 3.13b. Once the results beyond EPS were rejected the graph got much clearer and it is easier to notice that few designs obtained using max entropy approach have perceptibly higher entropy value for the same cost than designs achieved using min or total entropy approaches. Nevertheless, there is a gap with no results in max entropy POF between costs £1.4x10^6 and £1.5x10^6. Moreover, above cost £1.5x10^6 there is few results obtained using max entropy approach with lower entropy value for the same cost than designs achieved with remaining entropy approaches. Those fluctuations causes that max entropy POF is not smooth or consistent as expected. It also has some scatter, which is not desired.

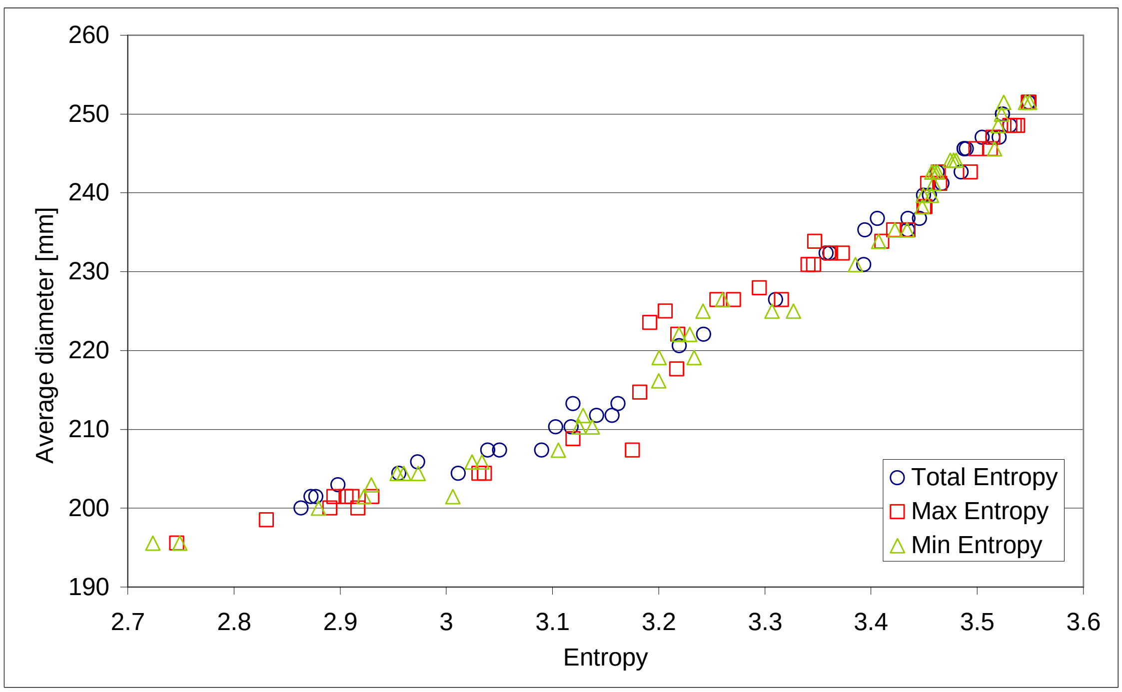


Figure 3.14 Effect of entropy on average pipe size

Figure 3.14 shows the correlation between average pipe diameter and entropy for MOC entropy approaches. There is not much scatter in POFs which reflects in very high coefficient correlations (Table 3.2). Few results obtained using max entropy approach are little bit outside the other two fronts, nonetheless this is not reflected on the relationship between average pipe size and entropy. The coefficient values are very similar, however the relationship achieved when using total entropy approach is slightly stronger than other two, with extremely high value of 0.957 (Table 3.2).

Table. 3.2 Coefficient of determination for network performance indicators for Example 2

Measure	Total Entropy	Max Entropy	Min Entropy
Coefficient of determination of entropy vs average pipe diameter	0.957	0.949	0.942
Coefficient of determination of entropy vs reliability	0.605	0.601	0.554
Coefficient of determination of reliability vs failure tolerance	0.699	0.672	0.622

Finally, using enhanced version of PRAAWDS with additional subroutine, the solutions underwent pipe failure simulations. For this network, not only the reliability was calculated, but also the failure tolerance. The importance of including failure tolerance in addition to reliability for a better representation of the network performance has been demonstrated in Tanyimboh and Kalungi (2009). As in previous example, the peak demands were used.

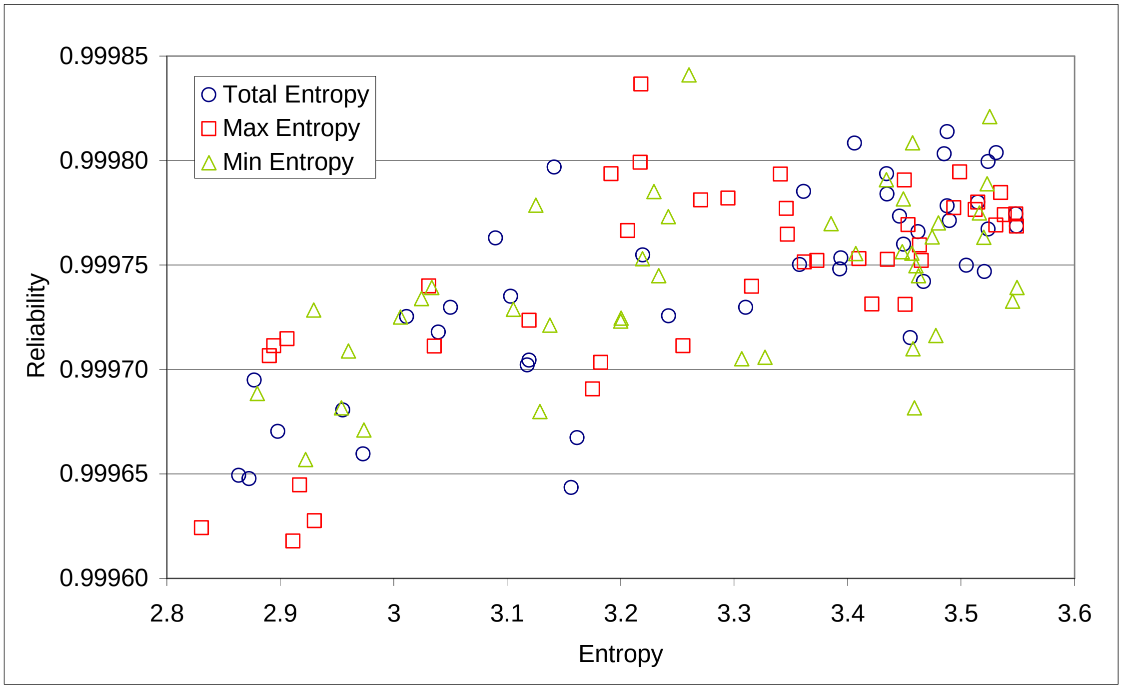


Figure 3.15 Reliability versus entropy

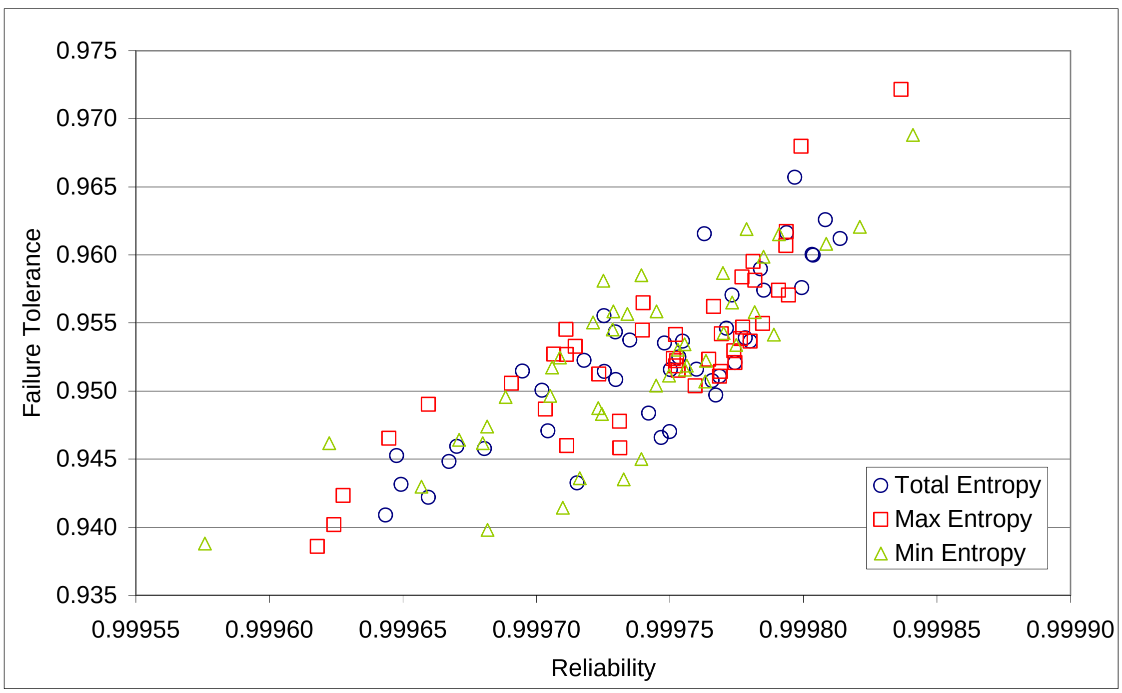


Figure 3.16 Failure tolerance vs reliability

Figure 3.15 and Figure 3.16 present reliability and failure tolerance versus entropy. It can be noticed that the results are very dispersed for all entropy approaches, especially on the graph that represents hydraulic reliability (Figure 3.15). The dispersion of results reflects on lower than expected coefficient of determinations of entropy vs. hydraulic reliability/failure tolerance (Table. 4.2). Nevertheless, for both comparisons (i.e. reliability and failure tolerance) the relationships are higher for results achieved using total entropy approach. Therefore, the total entropy approach is suggested as a possible candidate for handling multiple entropies in MOC approach.

3.6.3 Example 3

The third example is based on Simpson *et al.* (1994). This particular network (Figure 3.17) was chosen as it has three different demand patterns and associated required nodal heads that must be satisfied. The three operating conditions considered represents peak demand pattern and two fire-loading demand patterns. Each fire-loading operating condition has fire flow added to one specified node, while demands at other nodes remain the same as for peak conditions. Therefore, in contrast to previous examples, the nodal demands do not change with the same ratio. Moreover, the minimum nodal head is different for peak demands, fire flows conditions or fire flow node (Appendix B, Table B-4). The minimal nodal head is lower for fire flow conditions, which is similar to the real fire flow requirements. The original design problem was to determine the required pipe sizes for an expansion and upgrading of existing WDN. However, for purposes of simplicity, the network was used for initial design with all pipes to be sized. The network consist 10 demand nodes, 14 pipes and 2 reservoirs. All pipes are assumed to have Hazen-Williams roughness coefficient of 120. Further details of the network such as available pipe diameters, pipe lengths, nodal elevations, demand patterns and required heads can be found in Appendix B.

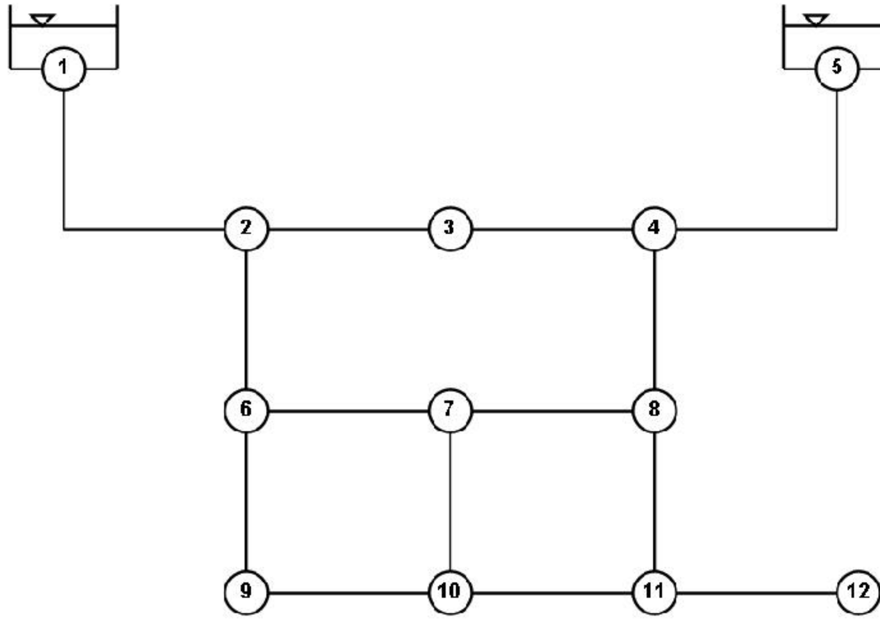


Figure 3.17 Layout for Network 3

There are 8 candidate pipe diameters, so the solution search space for this example comprises $(8)^{14}$ or 4.398×10^{12} infeasible and feasible solutions. A 3-bit binary substring was used, giving 8 substrings (2^3). Therefore, there were no redundant codes for this network. The termination criterion for the GA was taken as 200 000 function evaluations (i.e. 1 000 generations for a population size of 200) for all three MOC entropy approaches. 30 randomly generated GA runs were performed for each case, giving 90 GA runs in total. A single-point crossover operator was used to produce two offspring from two parents. 1.0 was used as a crossover probability. A bitwise mutation operator was used to change the bit from 0 to 1 or vice versa. Since the mutation probability was $1/n_g$ = 1/42 (i.e. 42 is the chromosome length), the mutation rate was set to 0.0238 (i.e. there were 2.38% chances that any single bit would mutate). The average CPU time required for single run was about 17 minutes, on PC with following configuration: Intel Core 2 Duo @ 3,5GHz and RAM 3GB.

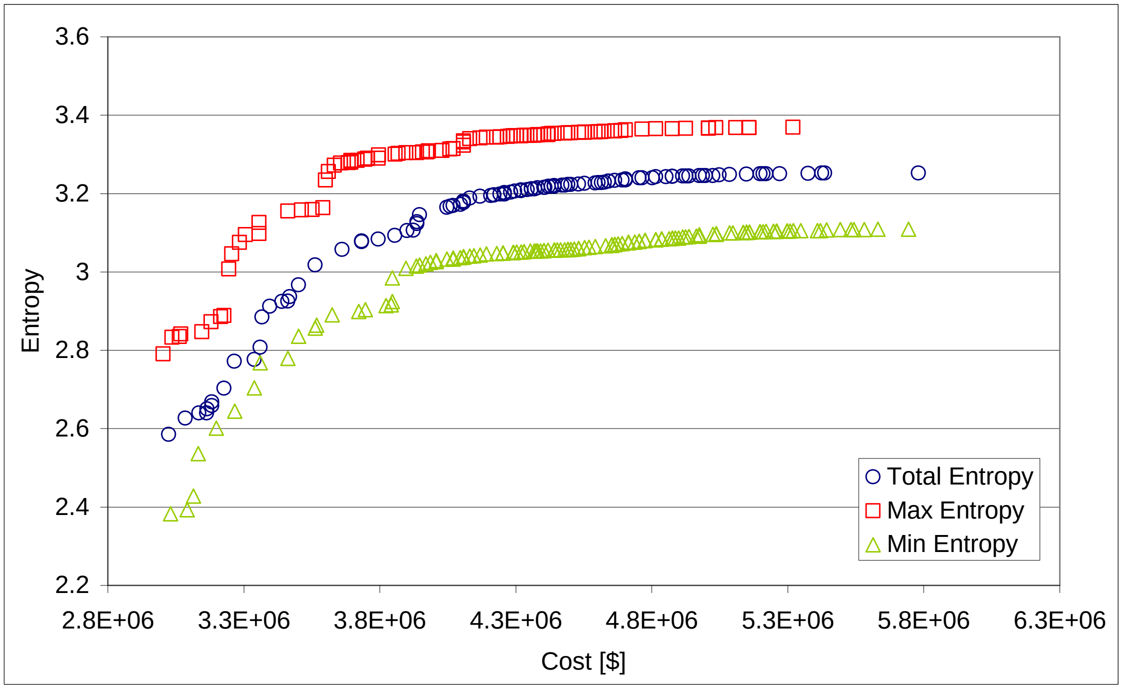


Figure 3.18a Entropy-cost POFs based on entire history of results for whole range of results

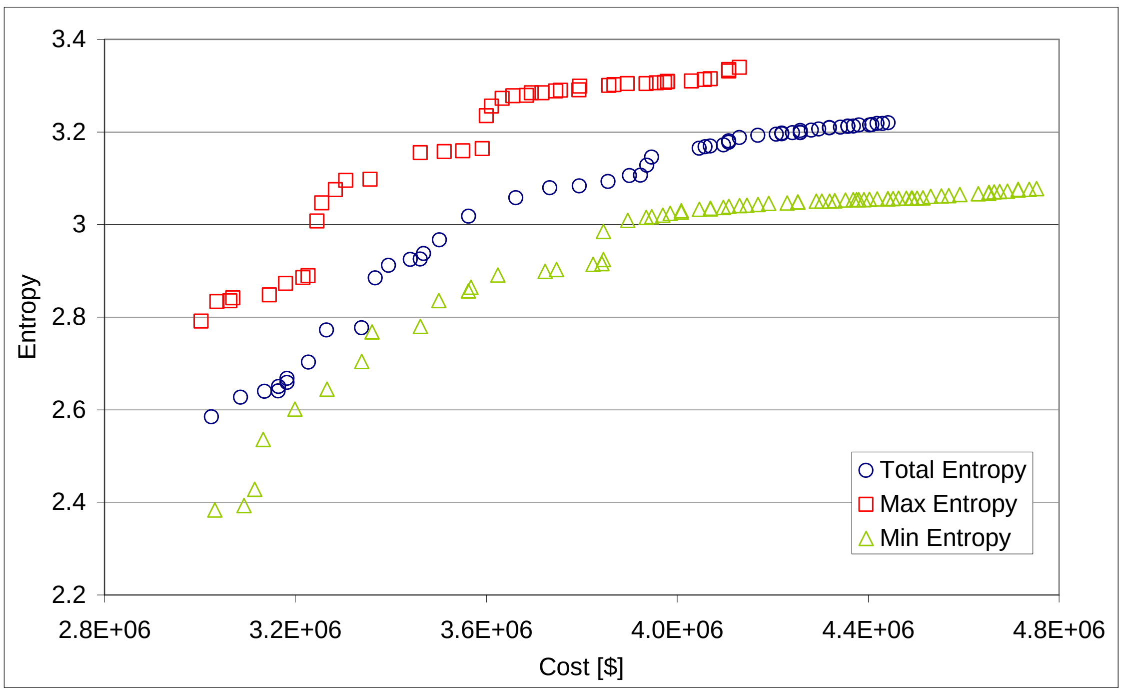


Figure 3.18b Entropy-cost POFs based on entire history of results for results up to ESP

It is noticeable that the entropy approaches have diverse ranges of entropy values in POFs. Also the POFs curves have slightly different shapes and would not overlap if brought to the same entropy range. This is completely expected outcome as a consequence of having three operating conditions with completely different demands (i.e. the demands does not change with the same ratio). It did not happen in the previous two examples. However, similarly to previous networks, the entropy values achieved using total entropy approach has been divided by three. This caused that POF obtained using total entropy approach is located between other two fronts. Once again, it is easy to notice that there is no need to analyse the entire range of results, as a vast number of solutions has high increase in cost with insignificant improvements in entropy. Therefore, the ESP has been set up to 99% of ME and applied to the network. However, the solution with highest entropy, likewise cost, differ much for each POF. Hence, the 99% has been calculated for each entropy approach separately which reflects in shorter (i.e. max entropy approach) or longer (min entropy approach) POFs (Fig. 3.18b)

The next step was to calculate the average pipe diameter and its relationship with entropy. As expected, the correlations are high for all approaches with the highest value for total entropy approach (Table 3.3).

Table 3.3 Coefficient of determination for network performance indicators for Example 3

Measure	Total Entropy	Max Entropy	Min Entropy
Coefficient of determination of entropy vs average pipe diameter	0.952	0.857	0.872
Coefficient of determination of reliability vs entropy for:			
CEND	0.651	0.599	0.007
CRND	0.825	0.660	0.832
CFTND	0.898	0.814	0.832
Number of all feasible designs	1894	1364	2057
Number of designs in POF:			
CEND	53	41	73
CRND	10	5	8
CFTND	5	4	8

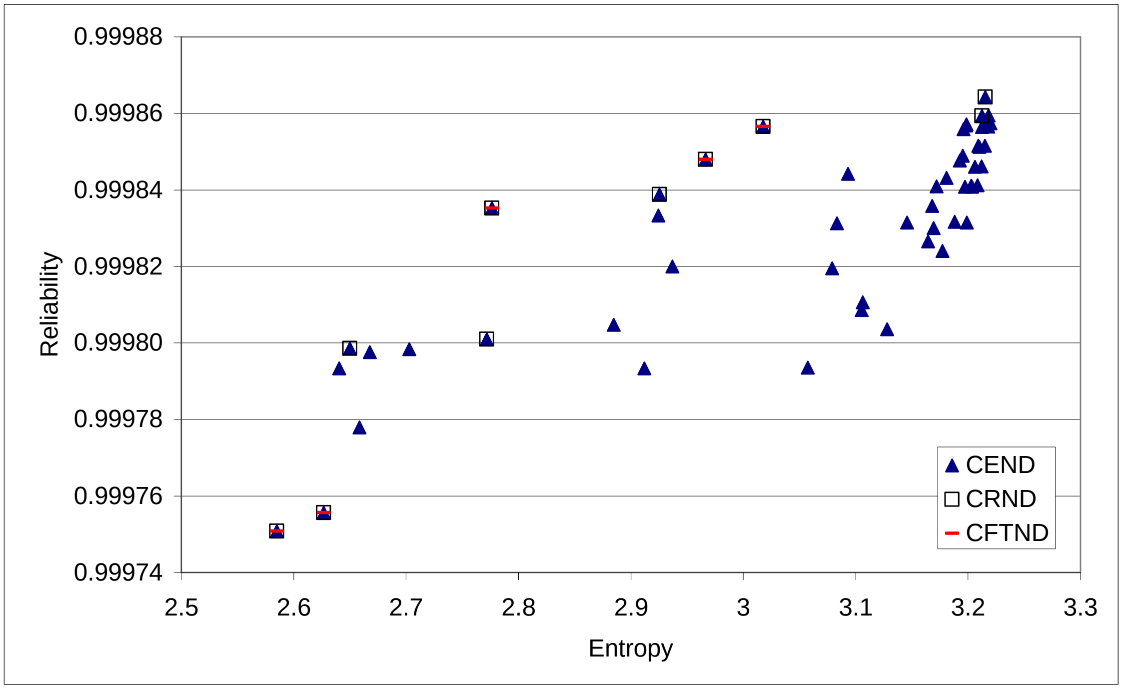


Figure 3.19a Reliability versus entropy for total entropy approach

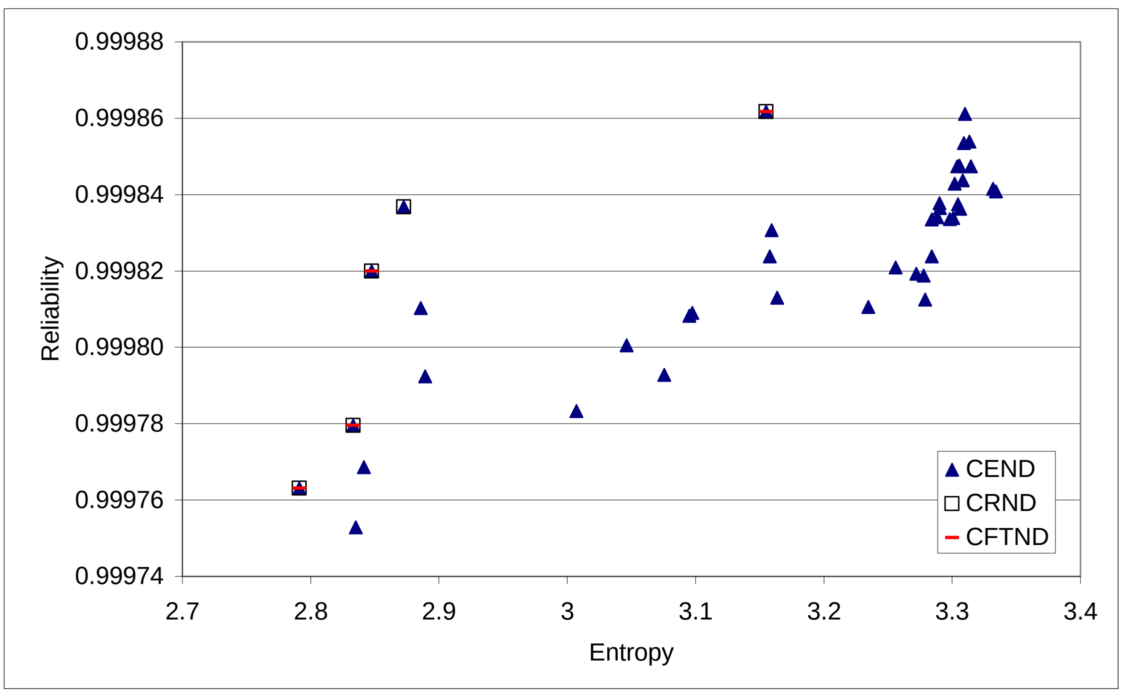


Figure 3.19b Reliability versus entropy for max entropy approach

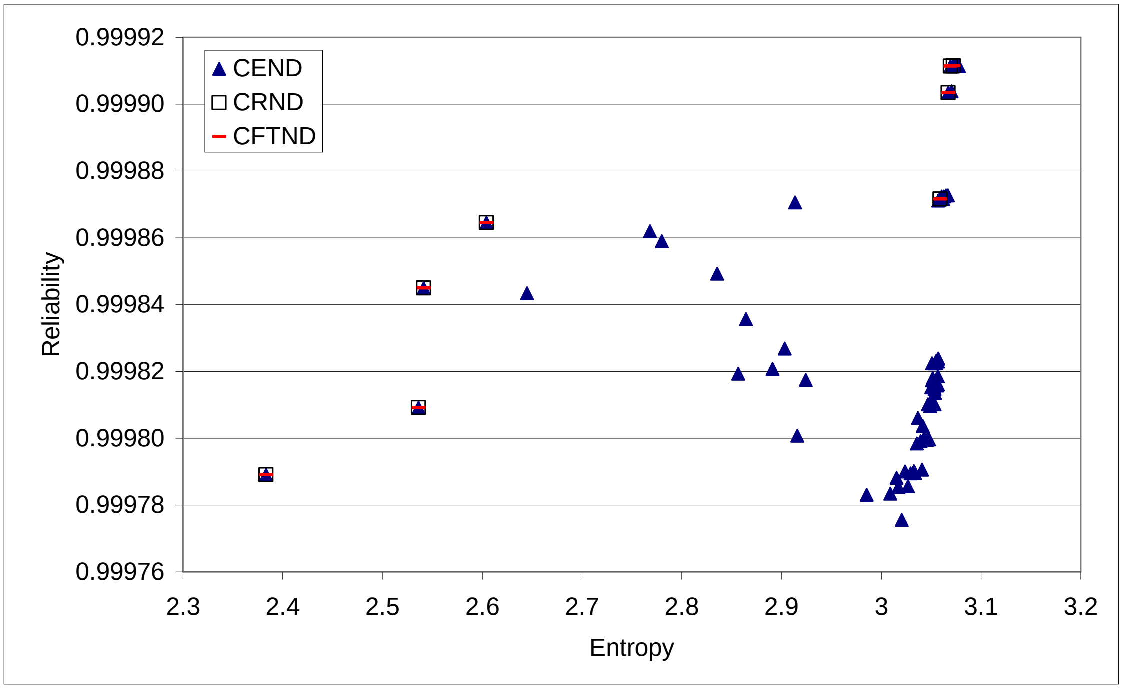


Figure 3.19c Reliability versus entropy for min entropy approach

Figure 3.19 present hydraulic reliability against entropy for cost-entropy non-dominated (CEND) solutions up to the ESP. To take the analysis step further additional cost vs reliability (CRND) and cost-failure tolerance (CFTND) filtering were performed on CEND and CRND, respectively. Related coefficients of determination are gathered in Table 3.3. It should be mentioned, that the peak demands were used to calculate hydraulic reliability and failure tolerance. Unlike the fire flows that are extreme situations, the peak loading is expected to occur on daily basis. Nevertheless, since each designs presented here satisfy all operating conditions, the solutions are also feasible for fire flow loadings.

It is noticeable on Figure 3.19, that all approaches have clusters of results located around highest entropy value. Like it was mentioned before, this happens due to nature of NSGA II that maximizes objective functions and despite using solutions up to ESP, at least half fall into the region of highest entropy (Fig 3.18). This has influence on reliability-entropy relationship by lowering value of coefficient of determination.

Only few solutions in each entropy approach were non-dominated for all aspects (i.e. cost, entropy, reliability and failure tolerance). Those designs remain to choose from.

For example, for the total entropy approach there are 53 CEND solutions (Table 3.3). The 53 CEND solutions represent only 2.8% of all the 1,894 feasible solutions generated by the GA across all generations for all random runs (i.e. 30 random runs for 200 000 function evaluations each). More importantly, the CEND solutions represent only 0.013% of the 6 000,000 candidate designs considered by the GA runs. Furthermore, it can be observed on the total entropy approach that the correlation between reliability and entropy increases from $R^2 = 0.651$ for CEND (Table 3.3) to $R^2 = 0.825$ for CRND and to $R^2 = 0.898$ for CFTND. These results appear to reinforce the hypothesis that, for cost-effective solutions, there is strong positive correlation between entropy and reliability/failure tolerance (Tanyimboh and Templeman, 2000; Tanyimboh and Setiadi, 2008).

In terms of max entropy approach, there are 41 CEND solutions that represent 3.00% of all the 1.364 feasible solutions generated by GA. Then, there are 5 CRND designs and only 4 CFTND solutions. The reliability vs. entropy coefficients of determination increase from $R^2 = 0.599$ for CEND (Table 3.3) to $R^2 = 0.660$ for CRND and to $R^2 = 0.814$ for CFTND. Nevertheless, the values are lower than the entropy-reliability relationships when using total entropy approach.

The min entropy approach generates the largest number of 2,057 feasible designs. This reflects on high value of 73 CEND solutions and 8 solutions for CRND as well as CFTND. Nevertheless the coefficient of determination of entropy vs. reliability for CEND is only 0.007. Therefore there is nearly no correlation at all. It can be noticed that between entropy value of 2.7 and 3.0 hydraulic reliability seems to decrease with increase of entropy. It could be case specific and treated as the outlier if the analysis would depend on single run. However, for this network 30 random runs were performed so there is a possibility that maximization of minimum entropy value does not lead to best (i.e. most reliable) solutions and it is not the ideal entropy approach for presented GA.

Despite which entropy approach was used, it can be seen that the method of identifying CEND designs with additional CRND and CFTND filtering used in this example determines the best (i.e. the most reliable) designs while reducing the computational burden. Nevertheless, the maximization of total entropy seems the most efficient and reliable to use. This reinforces the previous hypothesis that it is important to take entropy values from all operating conditions into consideration.

3.7 CONLCUSIONS

In this chapter, the formulation of a new penalty-free reliability based multi-objective optimization approach for WDNs has been presented. The approach combines multi-objective evolutionary algorithm with least cost design and maximum entropy. The novelty of this research in the context of entropy maximization is that the GA can work under many loading patterns for any given network and can handle discrete pipe sizes. Moreover, three different methods of designing for multiple loading pattern (i.e. entropy approaches) has been investigated.

Sensitivity analysis was carried out in order to assess the robustness of created penalty-free reliability based algorithm and to identify the input data (i.e. population size, mutation rate and crossover point) that increase the chances to achieve the best results and uniform spread of solutions in POF.

The algorithm has been applied to three well-known benchmark networks and the results have been compared in detail. Four cases were studied, one with SOC and three considering different MOC entropy approaches. It is known that designs based on SOC are generally infeasible for MOC, however all four sets of solutions were presented to demonstrate that MOC approach is competitive. In all cases, the proposed model found comparable entropy values for similar cost. However, designs based on any MOC entropy approach clearly outperform solutions obtained by SOC in terms of feasibility, pipe size distribution and reliability. Maximization of total entropy approach produced the best solutions. The significant advantage of total entropy approach over other two approaches is that it takes into consideration entropy values from all operating conditions. Therefore even if any operating condition has extreme values (i.e. minimal or maximal) that is highly dissimilar to other values, it does not have considerably adverse influence on the algorithm process.

Moreover, the total entropy approach seems to be the best choice as it confirms Jaynes' maximum entropy formalism (Jaynes, 1957). Jaynes stated that probability distribution which leaves the largest uncertainty (i.e. the maximum entropy, subject to whatever is known) should be used. In such case, any additional arbitrary assumptions or biases are not included. Therefore, for the different MOC entropy approaches only the total entropy seems to satisfy the maximum entropy principle. Maximization of maximum entropy value for MOC or maximization of minimum entropy value do not include everything or whatever is known (i.e. all provided data).

New method for identifying uniformly spread POF with feasible designs only has been presented. The subroutine for identifying cost-entropy non-dominated solutions over the entire range of results has been tested and proved to be robust and efficient. The biggest advantage of this method it that it allows gathering all best solutions (i.e. cost-entropy non-dominated) in POF, herein producing the front with wide range and nice spread of results. Also, the ESP (i.e. entropy stagnation point) has been proposed. ESP is the point beyond which the improvement in entropy becomes insignificant with high increase in cost. By applying and analysing results up to the ESP, the superfluous results are disregarded, thus leaving the most reliable and cost-effective designs for the decision maker to choose from.

CHAPTER FOUR

MAXIMUM ENTROPY BASED DESIGN OF REAL LIFE NETWORK WITH MULTIPLE OPERATING CONDITIONS

4.1 INTRODUCTION

In previous chapter, the formulation of new penalty-free reliability based multi-objective optimization approach for WDNs has been investigated. Based on three different networks, three entropy approaches for MOC (i.e. maximum, minimum or total entropy) were investigated. The results were analysed and it has been decided that the total entropy approach seems not only most logical but also the best to use since it takes into consideration values from all operating conditions. In addition, the new method for selecting feasible non-dominated solutions from entire range of results (i.e. from all generations up to and including the final Pareto-optimal front) was successfully introduced in previous chapter. The external screening method provides a good range of feasible non-dominated solutions.

In this chapter, the reliability based multi-objective GA developed in this research is applied to a real life WDN. Such network has many more pipes than hypothetical layouts used earlier which makes it harder to design (i.e. more difficult to find feasible, reliable design). Thus, the aim of this chapter is to demonstrate its practicality and capability in integrated optimization process. The total entropy approach is used to confirm that the maximization of total entropy is the most appropriate to handle multiple values of entropy. Furthermore, comparison of POF originally generated by the algorithm and the solutions obtained from entire history was performed and results presented in order to strengthen additional positive aspects of using external screening method of selecting results. Also the GA performance is analysed in order to assess whether the algorithm is efficient, consistent and stable.

4.2 MODEL APPLICATION TO A REAL LIFE NETWORK

4.2.1 Description of the Network and Design Data

The total entropy MOC approach was applied to the WDS serving a zone of the city of Ferrara in Italy. The skeletonized network shown on Figure 4.1 was previously presented by Creaco *et al* (2010, 2012). It consists of 49 nodes, out of which two are the reservoirs (node 1 and node 49), 76 pipes and 29 loops. Under normal conditions (i.e. the basic demands for presented network) reservoirs supply a total demand of 367 l/s. Both reservoirs have the heads set at 30m above sea level. All pipes have Manning roughness coefficient of 0.015 and the total length of the pipes is about 25.2 km. The elevation, minimum head and required head at all demand nodes are set at 0m, 5m and 30m respectively. Further details of the network such as available pipe diameters and the costs, pipe lengths and nodal demands can be found in Appendix C.

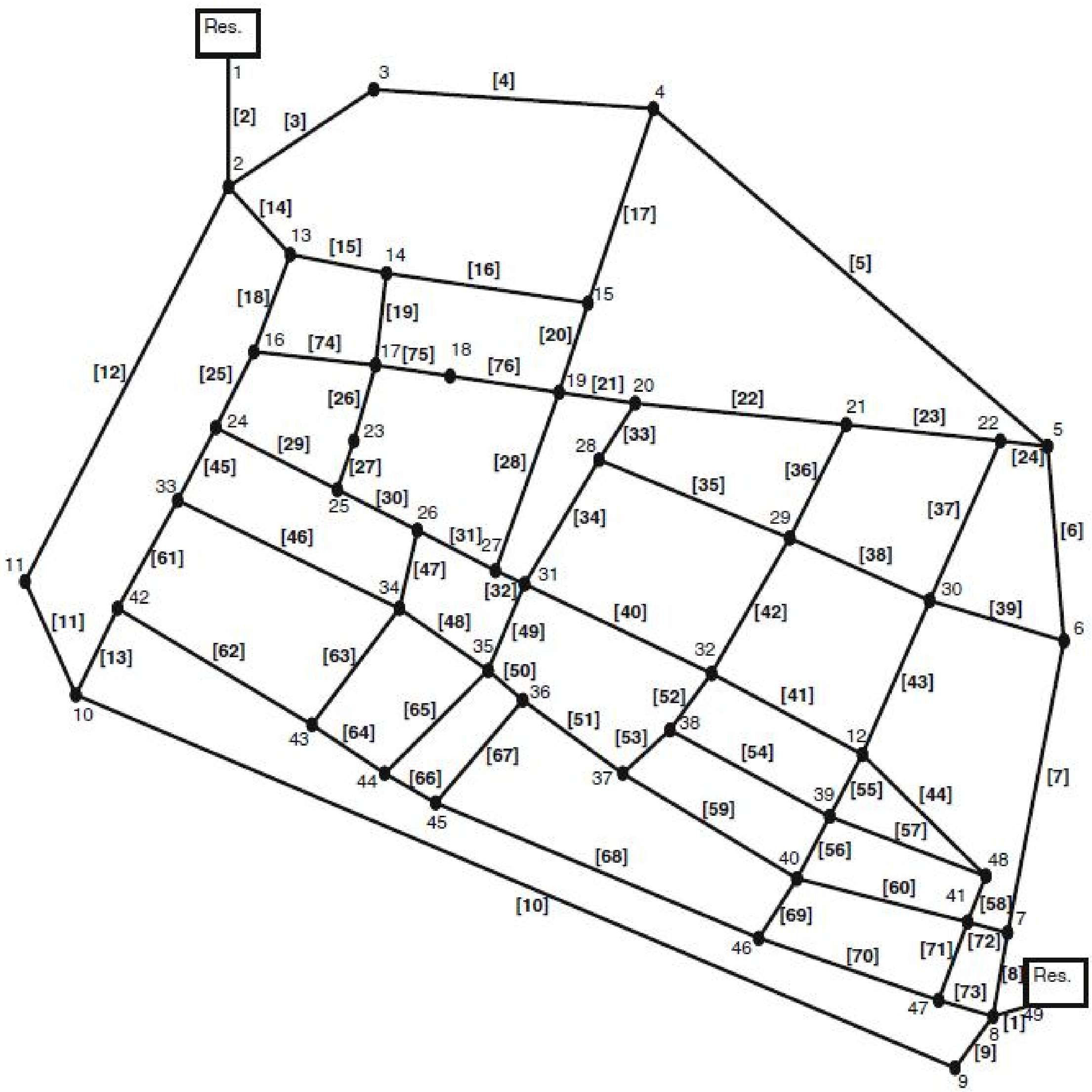


Figure 4.1 Layout of the Ferrara network (numbers in square brackets correspond to pipe numbers)

As this network has only data for one loading pattern, additional operating conditions were calculated based on Example 3 presented in Chapter 3. Example 3 has three operating conditions (peak demand pattern and two fire-loading demand patterns) and associated minimum nodal heads that must be satisfied. Each fire-loading operating condition has fire flow added to one specified node, while demands at other nodes remain the same as for peak conditions. Moreover, the minimum nodal head is different for peak demands, fire flows conditions or fire flow node. Since Example 3 is relatively smaller than network based on city of Ferrara (i.e. presented in this chapter) nodal demands and required heads for additional operating conditions were calculated based on percentage ratio from Example 3. For example, the percentage of increase of required demand for fire flow node was estimated based on Example 3, thus applied to network presented here. All nodal demands for three operating conditions and associated minimum nodal heads can be found in Appendix C.

There are 8 candidate pipe diameters, so the solution search space for this example comprises $(8)^{76}$ or 4.313×10^{68} infeasible and feasible solutions. A 3-bit binary substring was used, giving 8 substrings (2^3). Therefore, there were no redundant codes for this network. The optimization was carried out using NSGA II. The termination criterion for the GA was taken as 500 000 function evaluations (i.e. 1 000 generations for a population size of 500) and 30 randomly generated GA runs were performed. A single-point crossover operator was used to produce two offspring from two parents. 1.0 was used as a crossover probability. A bitwise mutation operator was used to change the bit from 0 to 1 or vice versa. Since the mutation probability was $1/n_g$ = 1/228 (i.e. 228 is the chromosome length), the mutation rate was set to 0.004 (i.e. 0.4% chance that any single bit would mutate). The average CPU time required for single run was about 1 hour and 30 minutes, on PC with following configuration: Intel Core 2 Duo @ 3.5GHz and RAM 3GB.

4.2.2 Results and Discussion

4.2.2.1 Comparison of novel screening method for non-dominated designs with POF originally generated by GA

The new method for selecting feasible non-dominated solutions from entire range of results was successfully introduced in Chapter 3. The external screening method provides a good range of feasible non-dominated solutions. Thus, the same approach was applied to network presented herein. Moreover, to strengthen additional positive aspects of using this novel method of selecting results, comparison of POF originally generated by the algorithm and the solutions obtained from entire history was performed and results presented below.

Similarly like in Chapter 3, majority of results presented herein are obtained by merging non-dominated solutions from numerous runs. POFs based on multiple runs have considerably more solutions than there would be when considering only single run. This happens as different runs have slightly different POFs, therefore merging and choosing non-dominated solutions causes that the gaps between solutions are filled with designs from different runs. Nevertheless, for comparison purposes, POFs for single runs are also presented below.

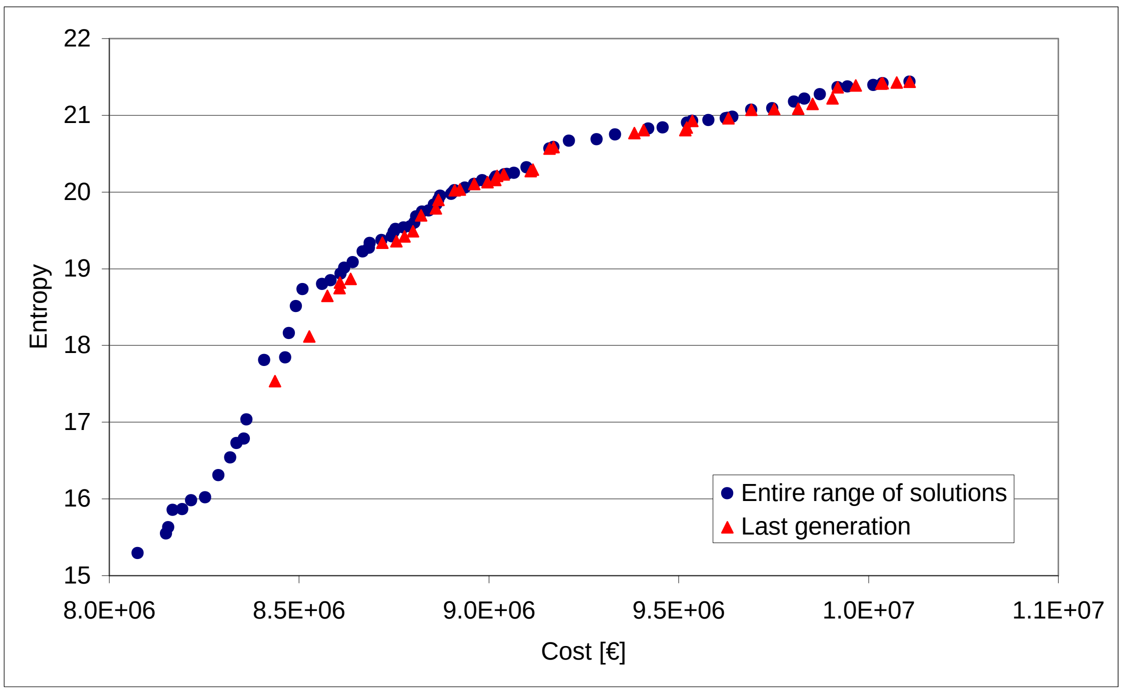


Figure 4.2 Entropy versus cost POFs based on solutions from entire history of results and solutions from the last generation

Figure 4.2 illustrates entropy versus cost POFs based on 30 random runs. Feasible non-dominated solutions chosen from entire history of results are distributed to a span of €2.034 million in cost and 6.148 in entropy, while the solutions from last generation are spread over the span of €1.745 million in cost and 3.935 in entropy. It is also easy to notice from Figure 4.2 that the feasible designs from last generation do not represent the entire range of entropy values that could be achieved for this network. It is completely expected outcome since the role of GA is to maximize entropy as one of the objective functions, thus designs with low entropy value are not kept through the generations. Nevertheless, from practical point of view, the feasible solutions with low entropy value and considerably low cost are valuable and could be chosen by decision maker. More importantly, the majority of solutions in the POF made from last generation are dominated by the results chosen using external screening (i.e. over the entire history of results). In other words, for the same cost, the solutions chosen with the new screening method have higher entropy value, thus reliability, than the ones provided in the last generation. Such case is possible, since despite the elitism used in the NSGA II, there are situations when the valuable solutions are lost. For example, when the population size is smaller than number of good non-dominated solutions, the excess of solutions will not be kept as a consequence of the application of the crowding distance operator.

As mentioned earlier, it has been decided that results for two single runs should be presented. The first run (Figure 4.3a) represents the case with the highest number of feasible solutions among all runs, while the second run (Figure 4.3b) is the opposite (i.e. case with the smallest number of feasible solutions).

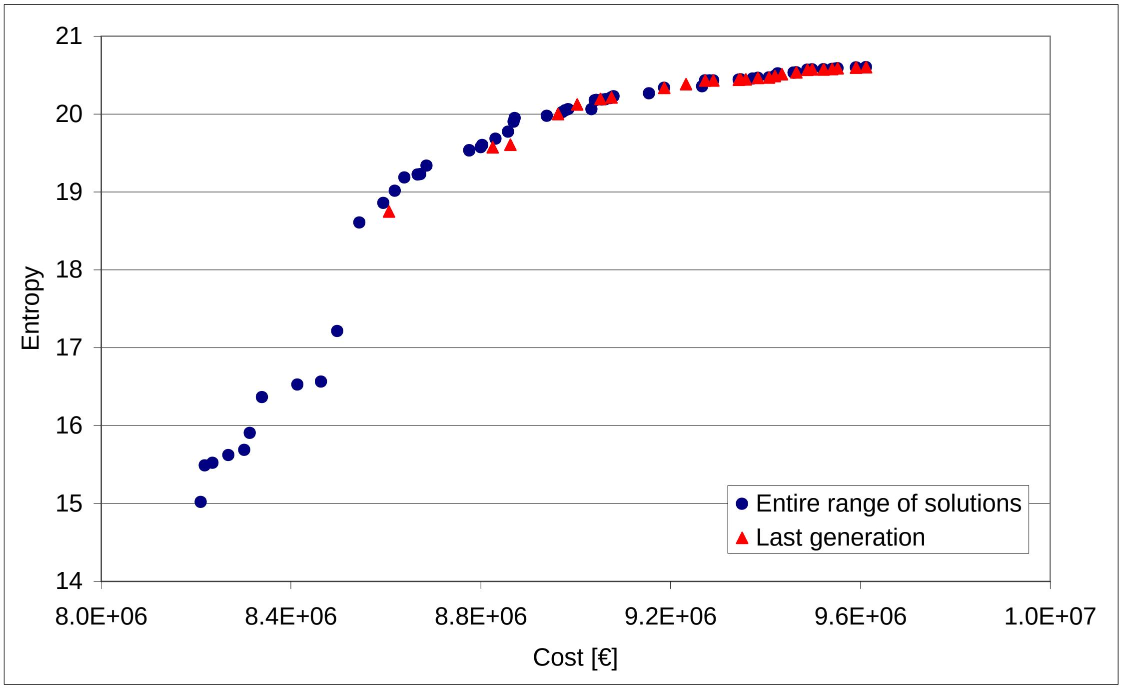


Figure 4.3a Entropy versus cost POFs for one single run (Case 1) based on solutions from entire history of results and solutions from the last generation

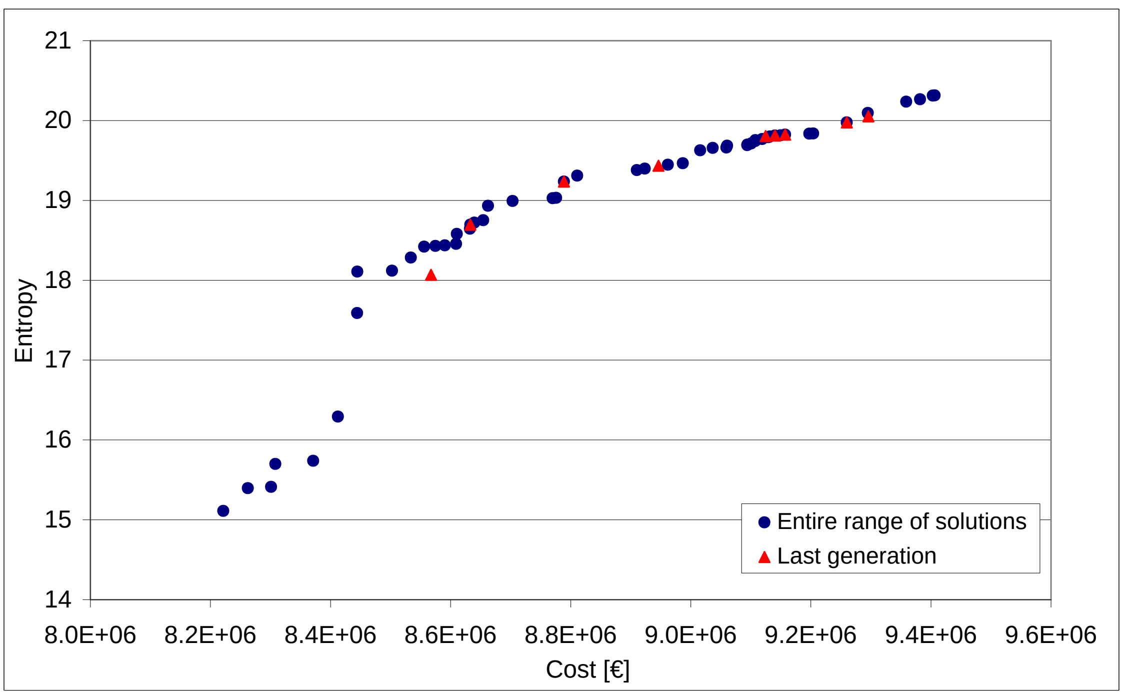


Figure 4.3b Entropy versus cost POFs for one single run (Case 2) based on solutions from entire history of results and solutions from the last generation

As expected, there is less results in POFs when single runs are analysed (Figure 5.3a and Figure 5.3b) for both method of obtaining feasible non-dominated solutions. Nevertheless, the difference in number of designs POF is much greater for solutions provided in last generation. For example, there are 41 feasible non-dominated designs in POF merged from 30 random runs (Figure 5.2). Once the single run is analysed, number of feasible non-dominated solutions drops down to 28 for the first case (Figure 5.3a) and to only 9 feasible results for the second case (Figure 5.3b). It can also be noticed on Figure 5.3a that the solutions obtained from last generation are mostly located between entropy values of 20.0 and 21.0, so around the highest achieved entropy value. Moreover, those solutions, due to insignificant increase in entropy but high increase in cost are not the most valuable solutions and will probably not be chosen by decision maker. As explained in Chapter 3, this happens due to nature of NSGA II that maximizes objective functions. Such situation is not observed on Figure 4.3b mostly due to very low number of feasible solutions obtained in this run in the last generation.

Difference in number of feasible solutions between multiple runs and single run is much smaller for proposed new method of screening results. This is mostly due to the fact that solutions are chosen over the entire history so covers whole range of entropy values, thus providing many more feasible solutions than the last generation. For example, there are 74 feasible non-dominated designs in POF merged from 30 random runs (Figure 4.2). Once the single run is analysed, number of feasible non-dominated solutions chosen over entire history of results drops down to 64 for the first case (Figure 4.3a) and to 52 results for the second case (Figure 4.3b). Therefore, for the second case (Figure 4.3b) the solutions obtained from last generation represent only 17% of the 52 results chosen over entire history. This proves that new method of selecting feasible non-dominated solutions is even more important and valuable when only single run or small number of runs can be performed.

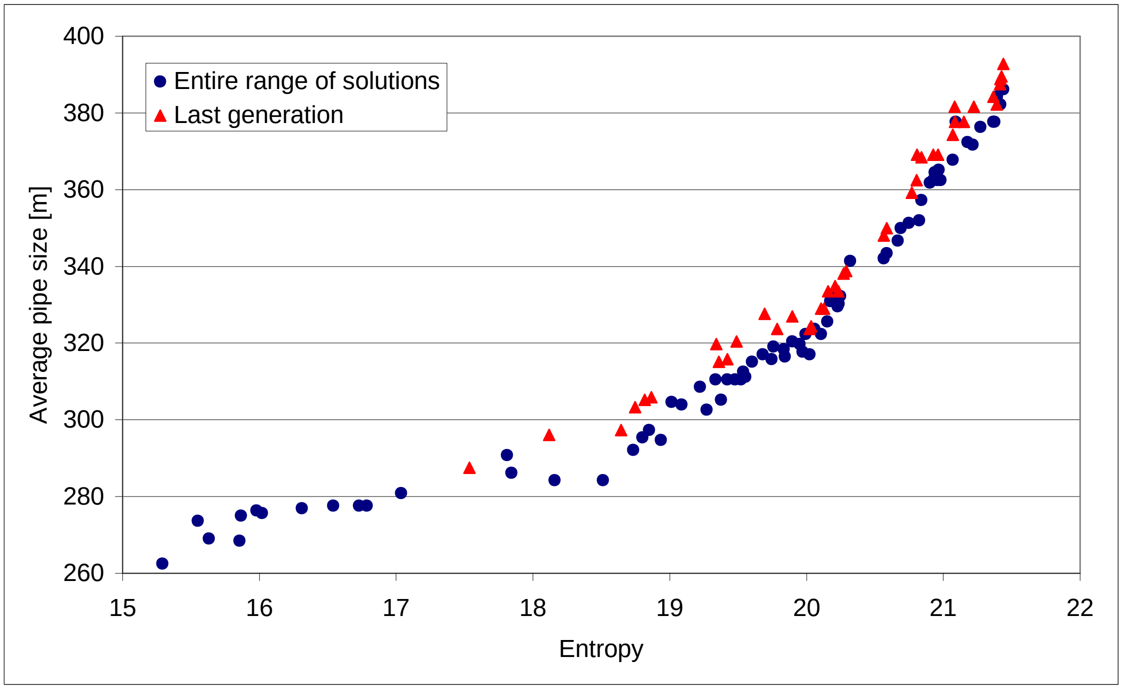


Figure 4.4 Effect of entropy on average pipe size for designs from last generation and designs from entire history of results

Figure 4.4 demonstrates relationship between entropy and average pipe diameter for feasible non-dominated solutions from last generation and feasible non-dominated designs chosen from entire history of results. As expected, the pipe sizes increase as the entropy increases. The coefficients of determination for both cases are extremely high with values of 0.982 for results from last generation and 0.991 for solutions obtained by new screening method.

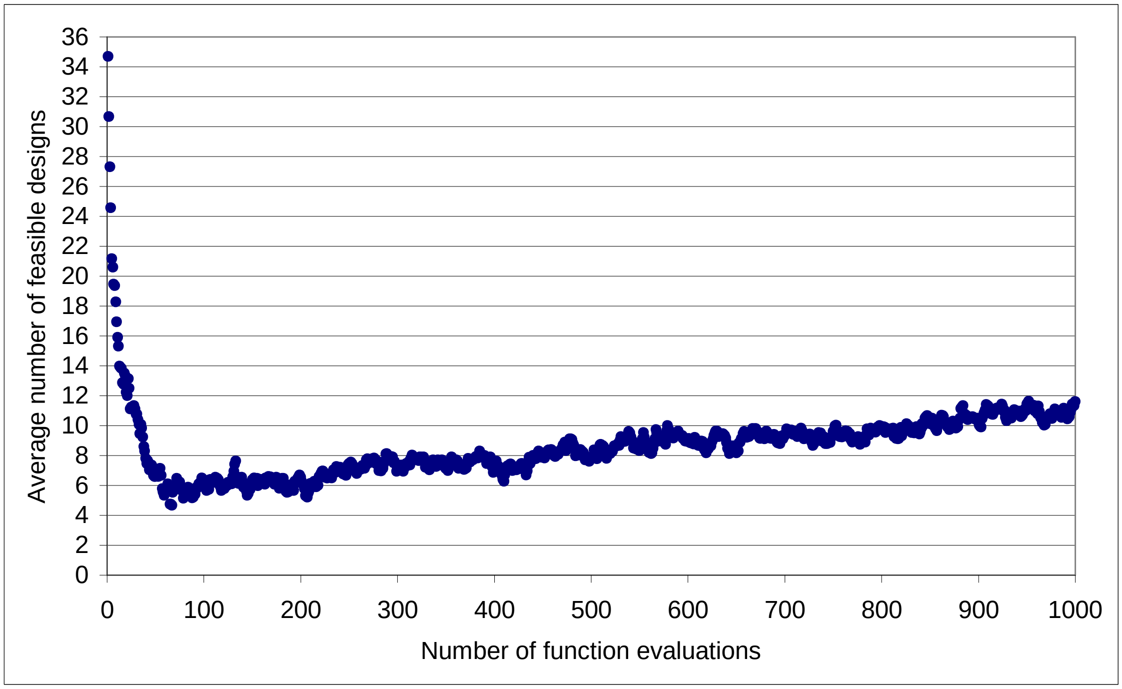


Figure 4.5a Average number of feasible designs from all random runs in each generation

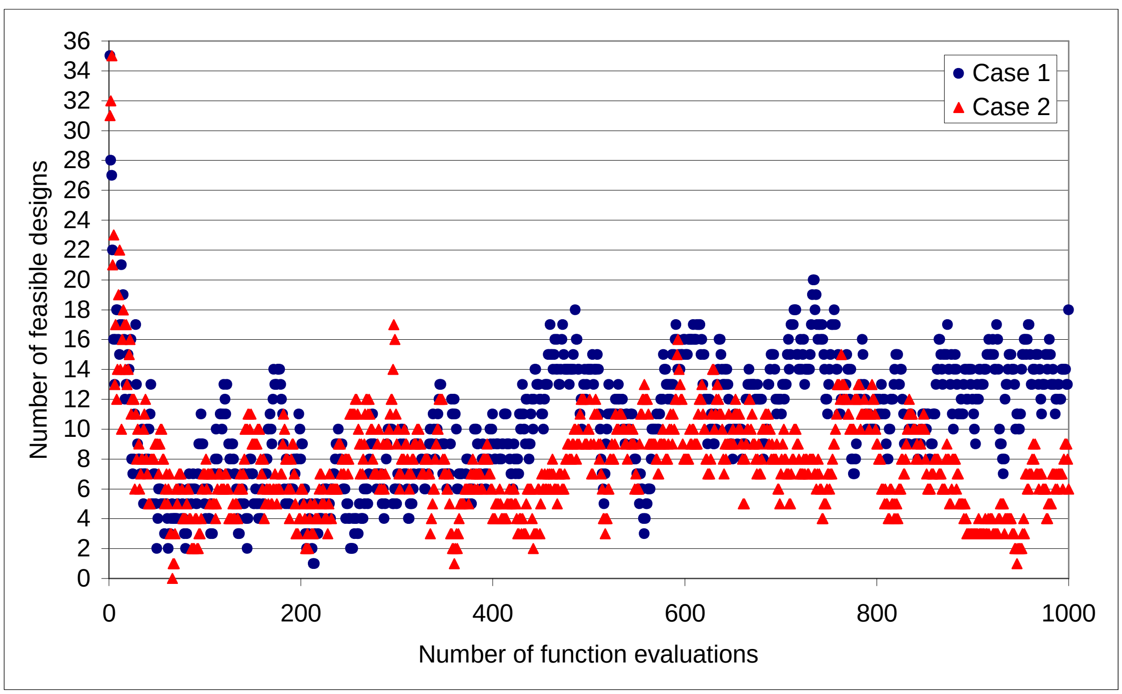


Figure 4.5b Number of feasible designs in each generation for two single runs

Figure 4.5a and 4.5b demonstrate how the number of feasible solutions changes over the GA progression. It can be noticed on both graphs, that there is the highest number of feasible solutions in the first few generations. Nevertheless, those solutions are extremely expensive designs. Therefore in the upcoming generations the GA reduces the cost, thus produce far less feasible designs than earlier. Then, as presented on Figure 4.5a, there is slow increase in number of feasible solutions until the termination criteria are achieved with 11 feasible solutions on average. More fluctuations can be observed on Figure 4.5b, where single runs are presented. Obviously the number of feasible solutions varies for different random runs. Nonetheless in both cases there are quite few generations with higher number of feasible designs than in the last generation. For example, for Case 2, there are high peaks with 17 and 16 feasible designs around function evaluation number 300 and 600 respectively. On the other hand there are only 6 feasible solutions just before the last generation). It leads to conclusion that those feasible solutions identified earlier in the optimization process are not necessarily retained from generation to generation, thus do not appear in final POF. Obviously this could happen because such solutions are too expensive, thus not important or because achieved entropy values are too low. Nevertheless, from the practical point of view it is better to have more solutions that are feasible and choose the ones that are case of interests (i.e. non-dominated) rather than having only few designs. Furthermore, with the use of the new screening procedure presented, there is no possibility that any good and valuable solutions will be lost. As explained in previous chapter, presented method of selecting feasible non-dominated designs over the entire range of solutions is quick and very straightforward. What is more, for such big network as presented in this chapter, simulated for 500 000 function evaluations, the screening procedure takes only about one minute on a PC with following configuration: Intel Core 2 Duo @ 3,5GHz and RAM 3GB.

4.2.2.2 Total entropy based MOC approach

Three entropy approaches (i.e. maximum, minimum or total entropy) were investigated in Chapter 3. Based on results from different network it has been decided that the total entropy approach is the most appropriate to use. The decision was made based on final entropy values provided by GA (i.e. the sum of all entropies for total entropy approach, the highest entropy value for maximum entropy approach etc.). Nevertheless the detailed analyse based on particular entropy values from operating conditions has not been conducted yet. Therefore, for network presented in this chapter, individual entropy values for different loading patterns were analysed and presented below in Figure 4.6 and Table 4.1. It should also be reminded, that only the maximization of total entropy was used in this chapter.

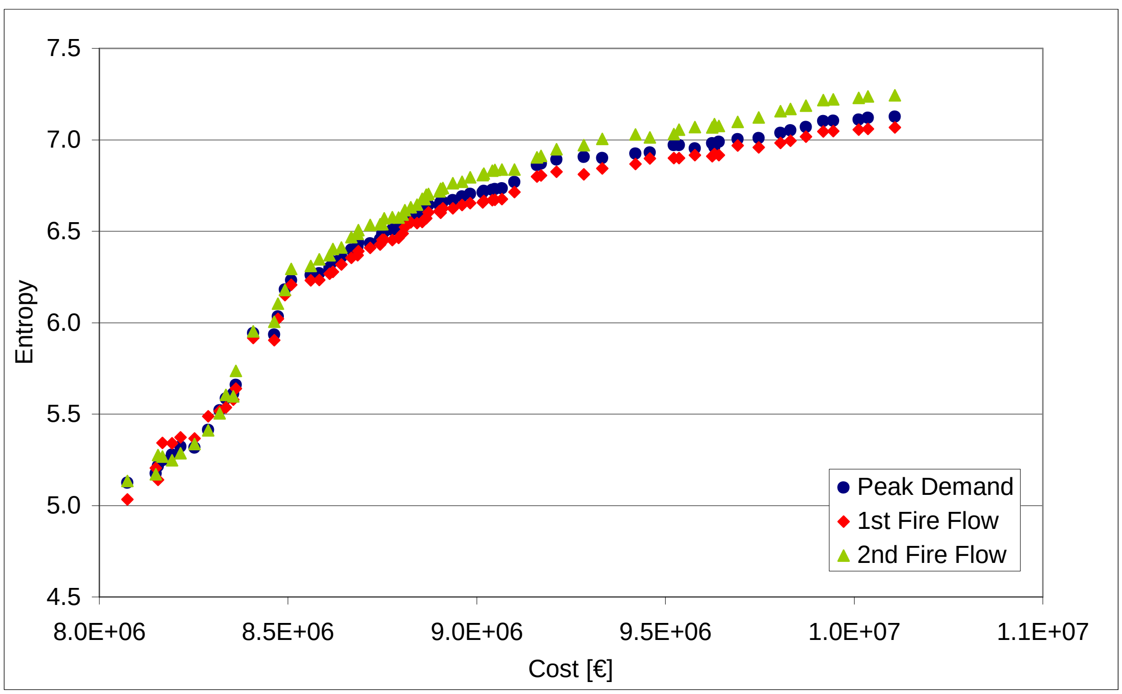


Figure 4.6 POFs of entropy versus cost for individual operating conditions

On both, Figure 4.6 and Table 4.1 can be noticed that there is not one operating condition that would have the highest entropy value for all designs. In case of expensive designs (i.e. starting from €8.5 million in cost) the 2nd fire-loading pattern seems to have the highest entropy value while the 1st fire flows condition the lowest. Nevertheless, when considering the cheaper designs (i.e. up to the €8.5 million in cost) there is no trend suggesting that any operating condition would always have the highest or the lowest entropies. This reinforces conclusion from Chapter 3, that the total entropy approach is the most appropriate to use for MOC optimization, as it takes into consideration all operating conditions. As stated earlier, if maximum or minimum entropy approach would be used, the GA process would depend on only one operating condition. What is more, the operating condition that GA would rely on (i.e. individual entropy value used for objective function evaluation) would probably change from generation to generation. In other words, there would not be any leading operating condition for all loading patterns. That could possibly make the algorithm less stable and less efficient. It should also be highlighted that for this network the entropy values for different operating conditions do not vary much between each other. Nonetheless, it all depends on the nodal demand and other initial data so there could be networks where entropies for individual loading patterns would have significant differences.

Table 4.1 Achieved feasible non-dominated designs for Example 2 with the highest entropy highlighted in bold red and lowest in bold blue

Design	Cost (€10^6)	Critical Surplus Head (m)			Entropy			Total Entropy
		Peak demand	1st Fire Flow	2nd Fire Flow	Peak demand	1st Fire Flow	2nd Fire Flow	
1	8.074	0.030	13.619	13.828	5.125	**5.033**	**5.134**	15.292
2	8.150	0.030	13.768	13.528	5.175	**5.205**	**5.171**	15.551
3	8.155	0.066	13.632	13.422	5.214	**5.140**	**5.277**	15.631
4	8.168	0.087	13.279	13.872	**5.247**	**5.342**	5.267	15.856
5	8.193	0.010	13.479	13.914	5.277	**5.341**	**5.248**	15.866
6	8.215	0.106	13.596	14.007	5.322	**5.372**	**5.286**	15.980
7	8.253	0.173	13.857	13.771	**5.316**	**5.366**	5.337	16.019
8	8.288	0.048	13.484	11.647	5.413	**5.487**	**5.410**	16.310
9	8.319	0.064	13.761	13.636	**5.521**	5.515	**5.504**	16.540
10	8.335	0.104	13.722	13.898	5.586	**5.537**	**5.606**	16.729
11	8.355	0.054	13.788	11.609	**5.612**	**5.579**	5.596	16.787
12	8.362	0.006	13.583	13.423	5.660	**5.639**	**5.737**	17.036

13	8.408	0.123	13.798	13.295	5.942	**5.917**	**5.952**	17.811
14	8.464	0.007	13.637	13.474	5.935	**5.905**	**6.004**	17.844
15	8.473	0.037	13.711	13.555	6.034	**6.021**	**6.104**	18.159
16	8.492	0.050	13.797	12.840	**6.181**	**6.150**	6.180	18.511
17	8.509	0.007	13.697	13.418	6.232	**6.207**	**6.295**	18.734
18	8.561	0.025	13.737	13.042	6.259	**6.231**	**6.311**	18.801
19	8.583	0.044	13.729	13.517	6.270	**6.234**	**6.346**	18.850
20	8.610	0.019	13.701	13.507	6.301	**6.267**	**6.368**	18.936
21	8.619	0.019	13.676	13.555	6.334	**6.278**	**6.402**	19.014
22	8.642	0.002	13.683	13.521	6.358	**6.317**	**6.411**	19.086
23	8.668	0.039	13.720	13.544	6.400	**6.355**	**6.468**	19.223
24	8.684	0.022	13.699	13.506	6.414	**6.370**	**6.486**	19.270
25	8.686	0.032	13.770	13.617	6.438	**6.392**	**6.506**	19.336
26	8.718	0.024	13.714	13.533	6.433	**6.408**	**6.534**	19.375
27	8.744	0.030	13.724	13.492	6.456	**6.428**	**6.537**	19.421
28	8.750	0.010	13.711	13.515	6.485	**6.454**	**6.539**	19.478
29	8.755	0.015	13.666	13.660	6.496	**6.453**	**6.571**	19.520
30	8.776	0.039	13.707	13.510	6.508	**6.452**	**6.577**	19.537
31	8.793	0.022	13.709	13.498	6.513	**6.464**	**6.574**	19.551
32	8.803	0.006	13.663	13.456	6.522	**6.488**	**6.591**	19.601
33	8.809	0.021	13.697	13.486	6.541	**6.522**	**6.616**	19.679
34	8.825	0.005	13.681	13.470	6.566	**6.548**	**6.631**	19.745
35	8.841	0.027	13.711	13.625	6.569	**6.544**	**6.646**	19.759
36	8.856	0.016	13.687	13.613	6.606	**6.550**	**6.679**	19.835
37	8.862	0.028	13.704	13.664	6.598	**6.567**	**6.676**	19.841
38	8.867	0.013	13.701	13.637	6.627	**6.570**	**6.699**	19.896
39	8.872	0.006	13.704	13.383	6.644	**6.602**	**6.703**	19.949
40	8.901	0.014	13.668	13.512	6.643	**6.610**	**6.719**	19.972
41	8.905	0.044	13.744	13.635	6.658	**6.602**	**6.732**	19.992
42	8.910	0.067	13.744	13.555	6.660	**6.626**	**6.736**	20.022
43	8.937	0.001	13.646	13.389	6.669	**6.625**	**6.763**	20.057
44	8.961	0.019	13.669	13.644	6.689	**6.644**	**6.772**	20.105
45	8.982	0.001	13.671	13.278	6.705	**6.653**	**6.795**	20.153
46	9.016	0.021	13.698	13.307	6.711	**6.657**	**6.807**	20.175
47	9.019	0.001	13.676	13.284	6.721	**6.665**	**6.815**	20.201
48	9.041	0.003	13.674	13.408	6.727	**6.669**	**6.832**	20.228
49	9.048	0.020	13.693	13.427	6.730	**6.672**	**6.835**	20.237

50	9.067	0.033	13.708	13.440	6.733	**6.676**	**6.838**	20.247
51	9.100	0.001	13.727	13.519	6.768	**6.715**	**6.837**	20.320
52	9.159	0.000*	13.743	13.058	6.862	**6.799**	**6.903**	20.564
53	9.170	0.011	13.756	13.071	6.870	**6.806**	**6.912**	20.588
54	9.211	0.015	13.721	13.695	6.892	**6.826**	**6.949**	20.667
55	9.284	0.032	13.712	13.695	6.906	**6.812**	**6.970**	20.688
56	9.332	0.003	13.700	13.590	6.901	**6.844**	**7.005**	20.750
57	9.420	0.005	13.705	13.594	6.925	**6.869**	**7.030**	20.824
58	9.458	0.006	13.638	13.453	6.930	**6.897**	**7.013**	20.840
59	9.522	0.003	13.682	13.762	6.970	**6.900**	**7.032**	20.902
60	9.536	0.011	13.692	13.705	6.970	**6.900**	**7.056**	20.926
61	9.578	0.027	13.646	13.607	6.953	**6.916**	**7.069**	20.938
62	9.624	0.009	13.690	13.705	6.982	**6.610**	**7.068**	20.960
63	9.629	0.059	13.672	13.769	6.961	**6.922**	**7.085**	20.968
64	9.641	0.039	13.727	13.740	6.990	**6.916**	**7.076**	20.982
65	9.691	0.026	13.651	13.764	7.003	**6.969**	**7.099**	21.071
66	9.747	0.000**	13.659	13.690	7.010	**6.959**	**7.123**	21.092
67	9.804	0.002	13.656	13.727	7.038	**6.984**	**7.156**	21.178
68	9.832	0.005	13.663	13.716	7.052	**6.995**	**7.168**	21.215
69	9.872	0.008	13.670	13.721	7.070	**7.016**	**7.186**	21.272
70	9.919	0.002	13.665	13.694	7.103	**7.046**	**7.218**	21.367
71	9.945	0.010	13.672	13.724	7.104	**7.048**	**7.221**	21.373
72	10.013	0.006	13.662	13.728	7.111	**7.055**	**7.229**	21.395
73	10.037	0.016	13.678	13.731	7.119	**7.060**	**7.237**	21.416
74	10.108	0.009	13.667	13.724	7.127	**7.069**	**7.244**	21.440

*The actual surplus head at critical node is 0.00013m
**The actual surplus head at critical node is 0.000002m

In addition to investigation of relationship between entropy and cost, also the correlation between actual surplus head and cost or entropy was examined and presented on Figure 4.7a and 4.7b respectively. For both graphs, the surplus heads from peak demand were used since for every single feasible design this operating condition has the lowest value of actual surplus head (Table 4.1). In other words, the peak demand was critical operating condition and once it becomes feasible, the other two patterns were also feasible. It was thought that with the increase of entropy the actual surplus head at critical node will increase as well. Nevertheless, as it can be noticed, there is no correlation between surplus head and entropy (Figure 4.7b) or between surplus head and cost (Figure 4.7a). Therefore, it can be concluded that presented GA with such implemented deficit seems ideal, as once the algorithm achieves feasibility it can concentrate on minimizing the cost and maximizing the entropy (i.e. all feasible solutions, regardless of the actual amount of surplus head are considered equal with respect to feasibility). If the feasible design would have actual value instead of zero the GA would maximize it equally to cost minimization and entropy maximization. Since there is no correlation between increasing entropy or cost and surplus head, the maximization of surplus head would contradict with other two objective functions. Consequently, more time and higher number of function evaluations would be necessary to achieve good trade-off between cost and reliability.

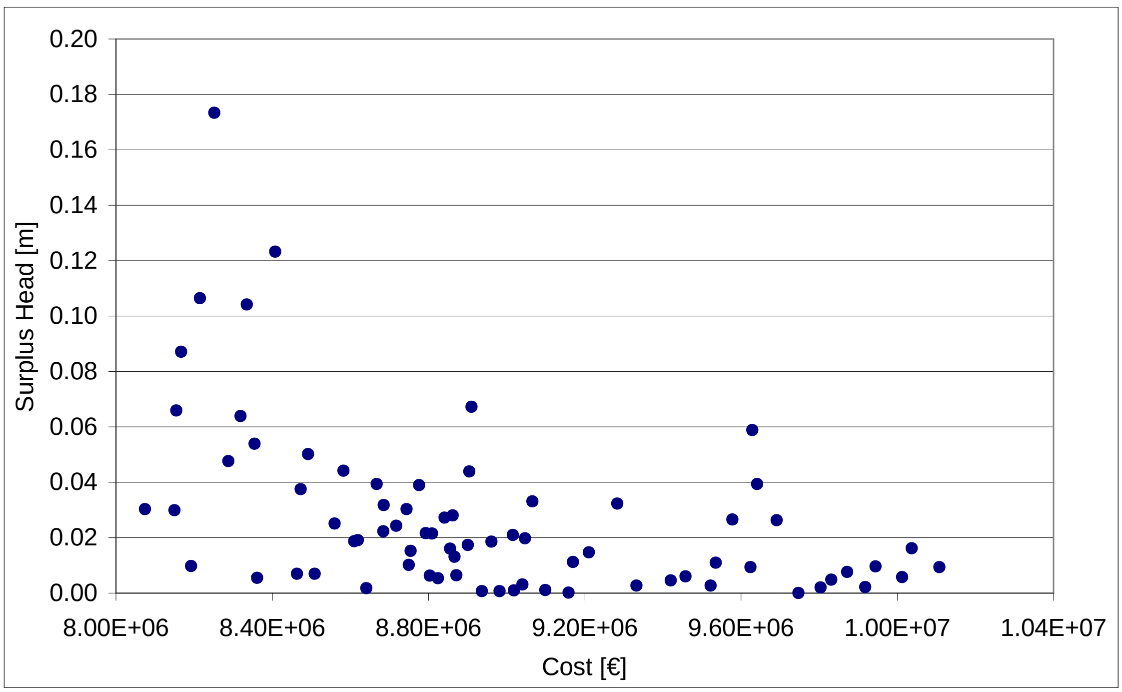


Figure 4.7a Surplus head versus cost for feasible non-dominated solutions from 30 random runs

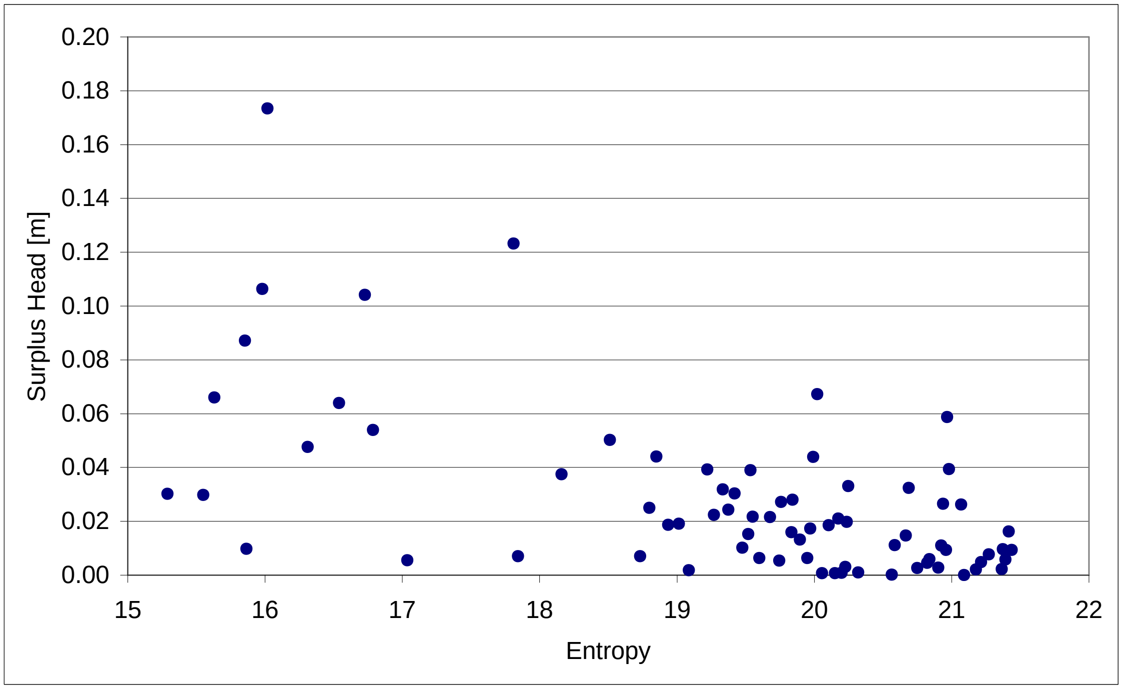


Figure 4.7b Surplus head versus entropy for feasible non-dominated solutions from 30 random runs

4.2.2.3 Algorithm performance

Any changes within GA could reflect in algorithm performance by making it less efficient or unstable. Therefore, it is necessary to ensure that the GA is consistent, converges quickly enough and performs well on all runs despite the initial data (i.e. the randomly generated initial population).

In order to investigate the consistency and performance of GA with such big network simulated for 30 random runs with 500 000 function evaluations in each run, additional software was used. For this purposes, the external software written in Perl language was developed and employed. The software identifies the highest, lowest and average value among 500 population size starting from 1st generation, then for every 20 generations and finishing on the very last one. This could be done for feasible and infeasible solutions or for feasible designs only, depending what data are required for further analysis. Afterwards, data from 30 random runs are gathered together in Excel spreadsheet and maximum, minimum and average values are recognized among all runs. In terms of cost, the external software identified the lowest value, since the aim of the optimization process is to minimize the cost. It should be mentioned that evolution and convergence characteristics for cost are based on feasible and infeasible solutions. For the deficit, the average value among all populations was used as it takes into consideration both infeasible solutions that have the actual deficit value and feasible solutions which have deficit of zero. In terms of entropy, external software identified the highest value, as maximization of the entropy is the purpose of the optimization. More importantly, only the feasible designs were taken into consideration, since only the feasible solutions have realistic or meaningful entropy value. In other words, the entropy value does not represent the proper reliability of the network if the network condition (i.e. feasibility) is not satisfied. Such acquired data were gathered together and presented on Figures 4.8 to 4.10.

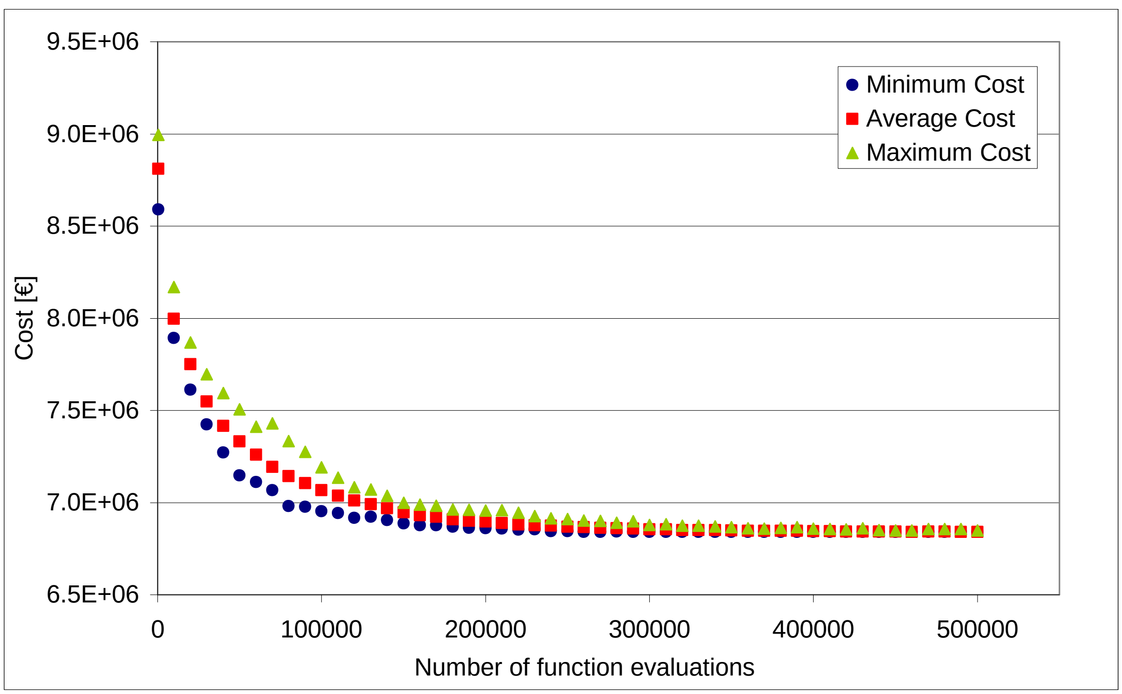


Figure 4.8 Evolution and convergence characteristics for minimum cost based on feasible and infeasible solutions for 30 GA runs

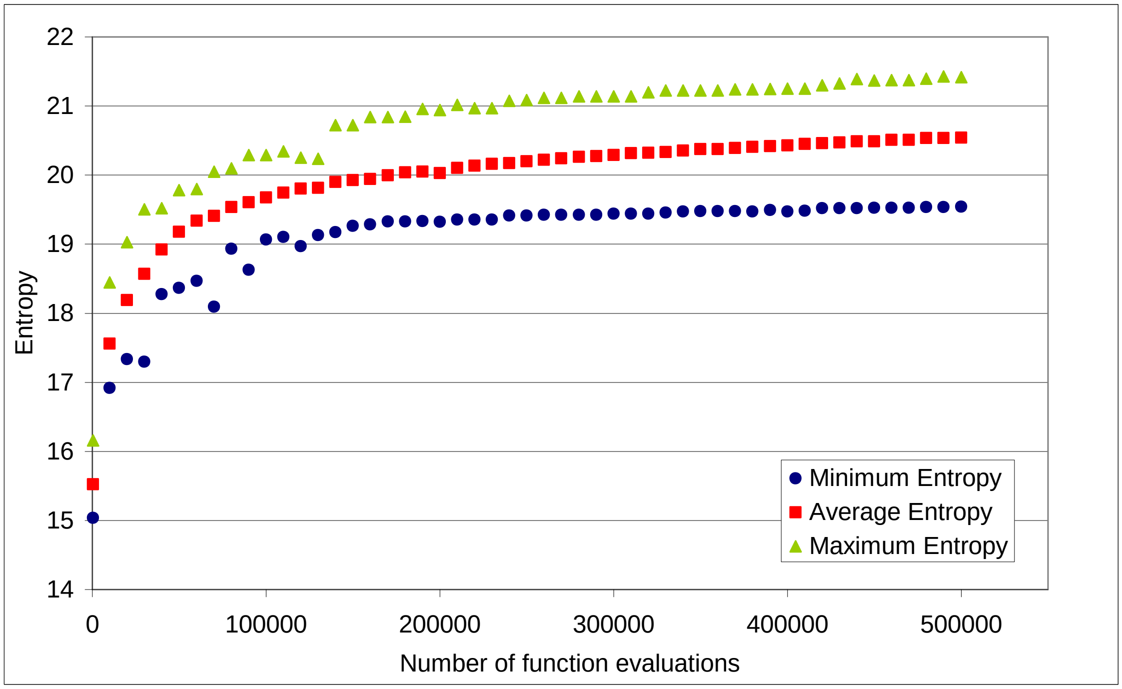


Figure 4.9 Evolution and convergence characteristics for maximum entropy based on feasible solutions for 30 GA runs

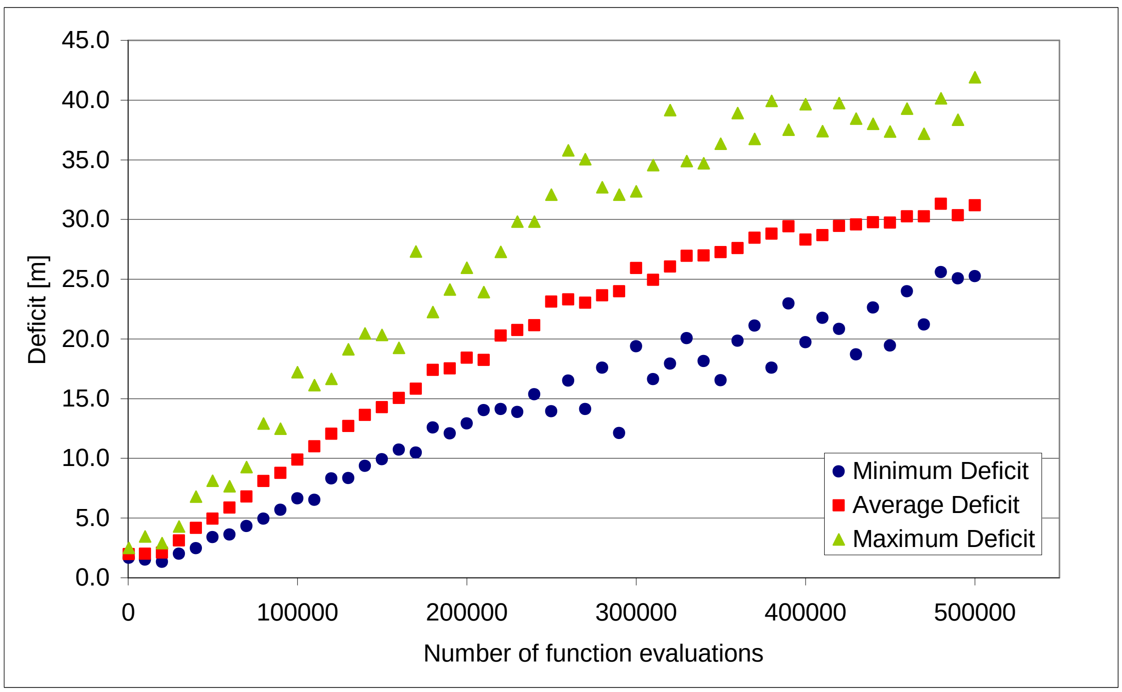


Figure 4.10 Evolution and convergence characteristics for average deficit based on feasible and infeasible solutions for 30 GA runs

The results presented confirm that the GA proposed in this study is stable and efficient. All objective functions behave according to the predictions and expectations. The quickest stabilization can be observed in terms of cost (Figure 4.8) with only 200 000 function evaluations (i.e. 400 generations and population of 500) needed to converge. The evolution and convergence characteristic for entropy is presented on Figure 4.9. The average of maximum entropy increases rapidly within the first generations, then the increase is much slower. There are not many fluctuations on Figure 4.9, although it should be highlighted that some imbalance is likely to appear in evolution and convergence characteristic for entropy. Entropy, in contrast to cost is more sensitive to changes. Once the decision regarding pipe sizes is made, the cost remains the same despite other circumstances with changes like flow paths or nodal heads. Entropy, however, highly depend on flow paths, flow directions and flow rates, so any changes in flow directions or even flow rates will have influence on entropy value. Therefore, for the same pipe sizes but different flows the entropy values can be extremely different. The average deficit (Figure 4.10) increases with the GA progression that is consistent with previously presented results (Figure 4.5). As explained earlier there are many feasible but highly expensive solutions in the first generations causing that the deficit value is fairly low at the early stage. With the cost minimization the number of feasible solutions gets lower, thus giving higher value of average deficit. It should also not be forgotten that all feasible designs have zero deficit in nodal head, while infeasible solutions have positive values. Therefore there are no values that could balance or lower the deficit as feasible solutions are neutral in this context. In addition, since the GA tries to compromise all objective functions, it also generates designs that are very cheap with potentially high but unrealistic entropy values that are highly infeasible, thus increasing the average deficit value.

4.3 CONCLUSIONS

The penalty-free multi-objective genetic algorithm developed has been applied the real life network to illustrate its applicability to water distribution networks in general. Maximization of total entropy value was used for MOC approach. Individual entropy values for different loading patterns were presented and analysed to demonstrate that total entropy approach is the most suitable to employ for MOC optimization. As in previous chapter, the external screening method was used in order to identify the feasible non-dominated solutions from entire history of results. Comparison of POF originally generated by the algorithm and the solutions obtained from entire history was performed and results presented. Overall, it was demonstrated that the screening procedure is extremely efficient and robust with the capability of providing a good improved front with wide range and nice spread of results. Moreover, the analysis of the GA performance strongly suggests that presented algorithm is robust with quick convergence rate.

CHAPTER FIVE

SELF ADAPTIVE SEARCH SPACE REDUCTION METHOD BASED ON MAXIMUM ENTROPY

5.1 INTRODUCTION

It has already been observed in this thesis that evolutionary techniques such as GA have been successfully used in WDNs optimization. Over the last few years many different benchmark networks were used and published results proved that GAs are efficient and capable of finding desired solution (i.e. minimum cost, maximum entropy etc.). However, the size of the search space is highly dependent on parameters such as number of links (i.e. size of the network) and number of commercially available pipe sizes. Therefore, the number of function evaluations required to identify optimal solution can be extremely large. Very simple, theoretical 2 loop network with only 8 pipes and 14 possible pipes diameters has a search space of $(14)^8$ or $1.475\text{x}10^9$. For large, real life networks, with hundreds of pipes and many network components, it can be extremely time consuming process as the GA might require even millions of function evaluations. It is therefore desirable to reduce the search space in order to speed up the optimization process. Hitherto, not much work with valuable results has been done in this aspect and only handful publications present methods for reducing the GA search space.

GA with reduced search space (RSS) approach presented in this chapter is based on entropy. Using the importance of every path through network, which is included in the entropy function, the number of candidate diameters for each pipe was reduced. The method does not involve pre-defining the diameters, which very often happens in different RSS approaches published in literature (Vairavamoorthy and Ali, 2005; Kadu et al. 2008 and Haghighi et al. 2011). The RSS GA developed here is self-adaptive and dynamic. It allows reducing the search space, thus the number of function evaluations required to identify optimal solutions while producing cheaper designs for the same entropy value than conventional GA based on full search space. The method has been tested on network previously used in search space reduction studies.

5.2 PROBLEM FORMULATION

Likewise in previous chapters, the objectives considered herein are minimization of the network's initial construction cost, subject to ensuring adequate pressures at all nodes and maximization of entropy. The overall problem formulation can be summarised as follows.

Minimize initial cost:
$$f1 = \sum_{i=1}^{np} C_i (D_i, L_i) \tag{5.1}$$

where $C_i (D_i, L_i)$ is the cost of the pipe i with diameter D_i and length L_i; np represents number of pipes in the system. Above formulation is subject to constraints such as nodal mass balance (Eq. 2.1, Chapter 2) and energy conservation (Eq. 2.2, Chapter 2) that are satisfied externally by EPANET 2 hydraulic solver (Rossman 2000).

Minimize infeasibility:
$$f2 = H_i^{des} - H_i \; ; \qquad H_i < H_i^{des} \tag{5.2}$$

where i is the critical node; H_i is the available head at node i; and $_i$; H_i^{des} is the desired head at node i. The desired head is the nodal head above which the demand is satisfied in full and the critical node is the node with the lowest pressure within the network.

Maximize Entropy:
$$f3 = S \tag{5.3}$$

where *S* is the entropy.

5.3 PROBLEM SOLUTION AND METHODOLOGY

The binary coded genetic algorithm NSGA II (Deb *et al.* 2002) was modified for WDN purposes and combined with hydraulic simulation software EPANET 2 (Rossman, 2000). Entropy subroutine capable of calculating entropy value for any given network layout was incorporated within the code as a surrogate measure of reliability.

As mentioned earlier, the NSGA II directs the search into objective minimization. Therefore, cost equation remains unchanged, as cost minimization is the objective function. In terms of deficit (i.e. hydraulic infeasibility), all values for pressure shortfall are converted into positive values, whilst surplus heads (positive value in reality) are assigned zero value. In such way, the algorithm concentrates on achieving feasible solutions. In case of the entropy, the positive entropy value is converted into negative value. Therefore directing the search to objective minimization leads in reality to entropy maximization.

The procedure of algorithm used herein can be illustrated as in Figure 5.1. Initially, a random parent population of size *N* is generated. To create offspring population of size *N*, the mutation and crossover are applied. The offspring and parent populations are combined together forming population of size *2N* which undergoes non-dominated sorting. This step ensures elitism by preserving previous and current best individuals. Non-dominated sorting involves dividing results into different fronts and assigning fitness values (i.e. ranks). The first front is non-dominated with assigned fitness value of 1. The second front has assigned fitness value 2 and its individuals are dominated by individuals from first front. This goes on until all results have designated fitness values. In addition, the crowding distance is calculated for each solution. Crowding distance is a measure of distance between neighbouring individual. High crowding distance means that individual is from less crowded area, thus having such solutions results in better diversity. Based on fitness values and crowding distance of the last front, the best *N* individuals are chosen. Among best *N* individuals, the feasible solution with ME value or slightly lower than ME value (i.e. depending on the percentage of ME used) is identified as ‘basic design’. Such solution is used to enforce search space reduction by reducing number of available pipe sizes (i.e. 14 pipe sizes in case of network presented herein) to required number of pipes (i.e. 5 pipes in this study). In order to identify what pipe sizes should be chosen for particular link, (i.e. each link may have different sets of pipe sizes), the algorithm identify the closest pipe sizes to the pipes from ‘basic design’. In other words, the pipe already allocated in ‘basic design’ is treated as ‘middle pipe’. Subsequently, additional two nearest pipes with higher diameters and two nearest pipes with lower diameters are added thus giving 5 pipe diameters in total. If the ‘middle pipe’ appears to be the biggest or smallest available pipe size, it is tripled (i.e. it repeats three times increasing the chances that it will be chosen). Similar situation happens if the ‘middle pipe’ appears to be the next to the biggest or smallest. In this case, the biggest or smallest pipe size is doubled. Therefore, for next generation, the algorithm is searching for solutions through 5 pipe sizes allocated to each link. The process is repeated in subsequent generations until reaches termination condition.

For purposes of this study, that RSS GA was tested on 5 pipe sizes since 5 pipe sizes were used in previous publications for this particular network design considered (Kadu et al. 2008 and Haghighi et al. 2011). Nevertheless, number of available pipe sizes can be easily changed to any odd number required.

It should also be highlighted that the search space reductions are only enforced once the feasible solution is identified. Until then, the RSS GA works exactly the same as FSS GA. Therefore, if the first feasible solution or feasible solutions will appear in, for example, generation 8, the results until generation 8 will be the same for both cases (i.e. RSS and FSS) as long as the simulations were based on the same initial data and random seed.

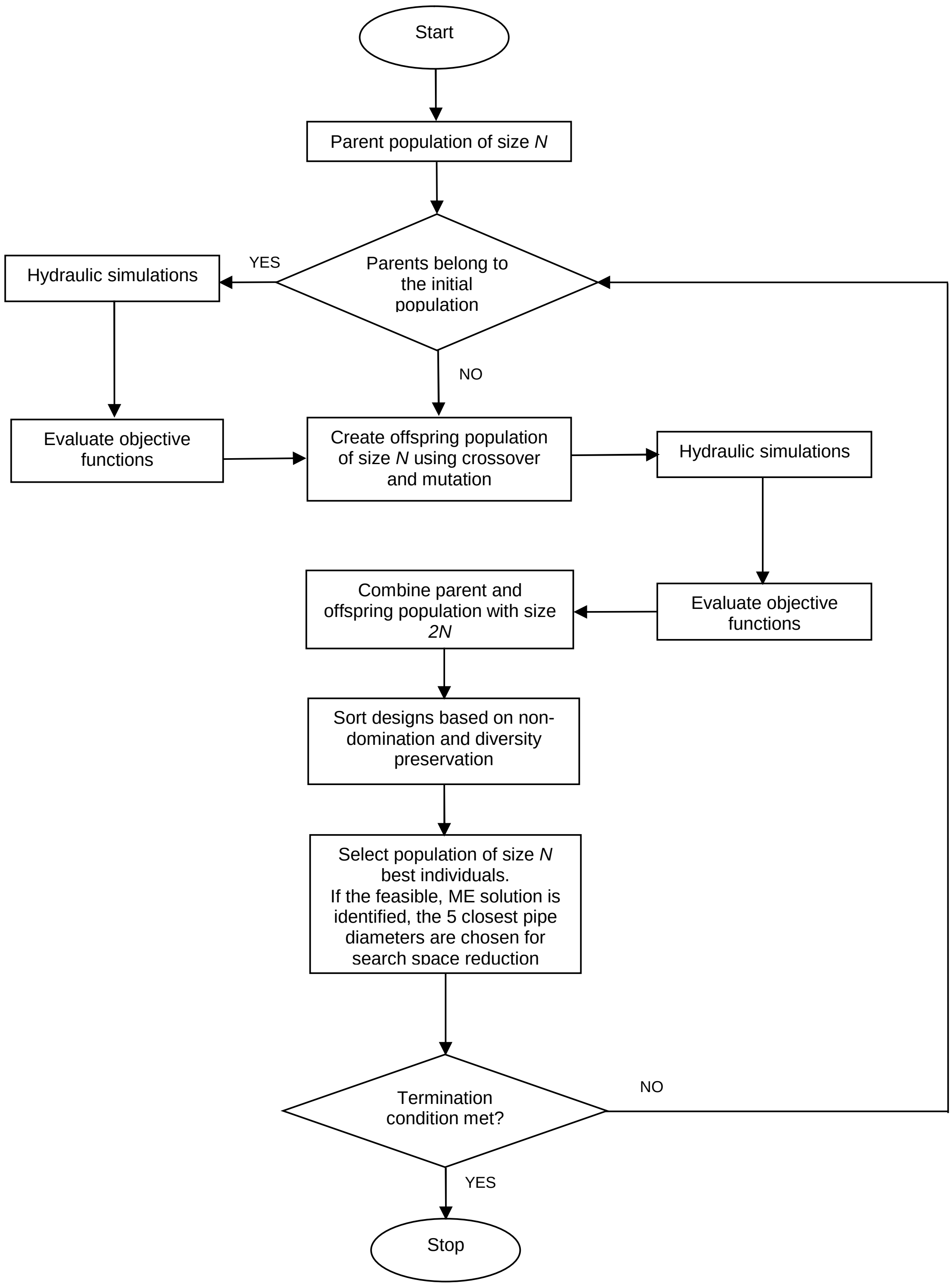


Figure 5.1 Flow chart of proposed approach

5.3.1 Formulation of Search Space Reduction Approaches

Previous studies and results presented in Chapter 3 (Figure 3.7, Figure 3.12 and Figure 3.17) has shown that maximization of entropy leads to numerous solutions with insignificant increase in entropy value but high increase in cost. Since the RSS method is based on entropy maximization it was thought that choosing the solution with ME within each generation, could result in assigning unnecessary big pipe diameters, thus neglecting cheaper solutions with lower entropy. Therefore, it has been decided that not only design with ME value should be used to identify 5 pipe diameters for next generation, but also solutions with slightly lower entropy value than ME for particular generation. Different percentage values of ME were tested and achieved results showed that three percentage approaches proposed herein identify the best solutions, fulfil the gaps between each other and present nearly entire range of possible entropy values for particular network. Single GA run allows using one percentage of ME value. Therefore, depending on required results (i.e. cheap solutions with lower entropy or the most reliable designs with highest entropy) one percentage approach could be employed. Nevertheless to achieve solutions covering the entire range of possible entropy values all three percentage approaches should be used in three separate GA runs. The proposed percentage approaches are as follows.

5.3.1.1 100% of Maximum Entropy

GA with 100% of maximum entropy reduce search space approach (100% MERSS) identify the solution with highest entropy value among all feasible designs for particular generation. Based on that solution, the algorithm allocates pipe sizes for the next generation. Overall, the aim of 100% MERSS is to identify the most reliable solutions.

5.3.1.2 99% of Maximum Entropy

Algorithm with 99% of maximum entropy reduce search space approach (99% MERSS) also identify the highest entropy value among all feasible designs for particular generation. Then, the 99% of this ME is calculated and the solution with closest entropy value (i.e. either lower or higher) is used to allocate pipe sizes for next generation. The 99% MERSS will most probably not be able to find the designs with highest entropy value. Nevertheless, extensive testing has shown that this percentage approach can fill possible gap between results achieved using 100% MERSS and 98% MERSS.

5.3.1.3 98% of Maximum Entropy

GA with 98% of maximum entropy reduce search space approach (98% MERSS) works on similar basis as the 99% MERSS approach. The only difference is that the solution closest to 98% of the highest entropy value is used to allocate pipe sizes for the next generation. Because of that, the GA is able to identify the solutions from the other end of entropy range (i.e. designs with lowest entropies and lowest cost). It should be highlighted that those low cost, lower entropy were not identified by GA based on FSS. Moreover, those solutions are important from engineering point of view as they could be chosen by the decision maker (i.e. depending on the budget level).

5.4 APPLICATION OF SEARCH SPACE REDUCTION APPROACH

5.4.1 Description of the Network and Design Data

To present the potential of proposed approach the network previously introduced by Kadu et al (2008) was used. This network represents a substantial challenge as despite its medium size (i.e. number of pipes) it has large number of decision variables, thus resulting in large solutions space. It has not been widely studied, however all previous publications relate to search space reduction. Therefore, this network is ideal to present novel self-adaptive search space reduction method. The skeletonised network shown on Figure 5.2 consist 26 nodes, 34 pipes and 9 loops. Two reservoirs are labelled 1 and 2 with the heads set at 100 and 95m above sea level, respectively. All pipes have Hazen-Williams roughness coefficient of 130. Nodal demands in m^3/min, as well as nodes and pipe numbers are shown in the Figure 5.2. Further details of the network such as available pipe diameters and its cost, pipe lengths and minimum HGL can be found in Appendix D.

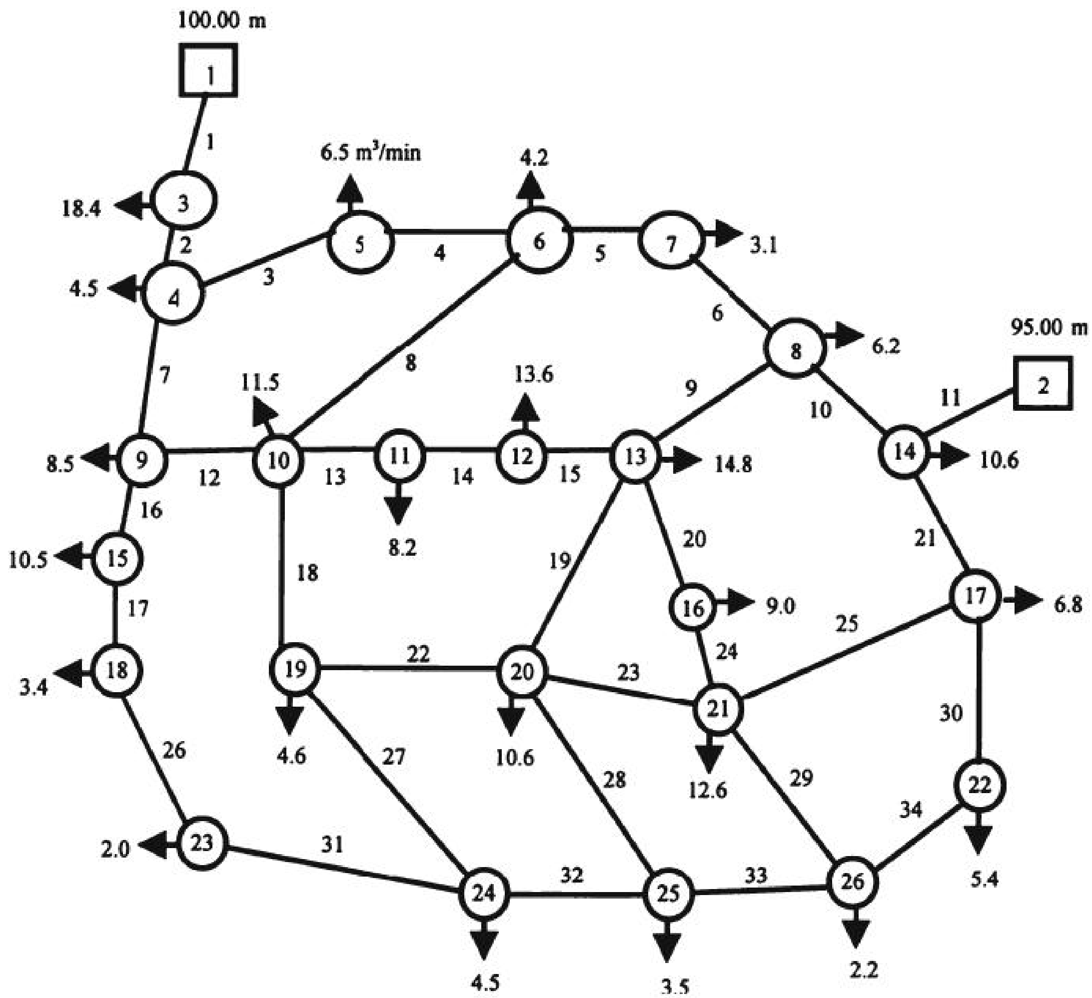


Figure 5.2 Network layout

Presented network has 14 candidate pipe diameters, so the solution search space for this example comprises $(14)^{34}$ or $9.3\text{x}10^{38}$ infeasible and feasible solutions. A 4-bit binary substring was used, giving 16 substrings (i.e. 2^4). Having 14 decision variables left 2 substrings redundant. Those redundant codes were uniformly allocated to available pipe sizes. Uniform allocation means that the redundant codes are assigned to every i^{th} number of pipe preserving the same space at the beginning, end and between pipes that are doubled. Therefore, in case of 14 pipes and 2 redundant codes, pipe fifth and tenth are doubled. Once the reduction in available pipe sizes takes place (i.e. the most appropriate feasible solution is found and number of decision variables from 14 for each pipe is reduced to 5), the number of redundant codes increases and is automatically uniformly allocated to available pipe sizes. In other words, when design that is the basis for another generation is identified, the nearest pipe sizes (i.e. two smaller and two larger sizes) for this particular link are identified and redundant codes are allocated among them. Such allocation of redundant code is dependent what pipe sizes for particular links had chosen design. Therefore, each link may have different pipe sizes. The whole process is fully dynamic and does not require any manual interference.

The termination criterion for the GA was taken as 100 000 function evaluations (i.e. 1 000 generations for a population size of 100) and 50 randomly generated GA runs were performed. A bitwise mutation operator was used to change the bit from 0 to 1 or vice versa. Since the mutation probability was $1/n_g = 1/136$ (i.e. 136 is the chromosome length), the mutation rate was set to 0.007 (i.e. there was a 0.7% chance that any single bit would mutate). A single-point crossover operator was used to produce two offspring from two parents and 1.0 was used as a crossover probability. The average CPU time required for single run is almost the same for all four cases (i.e. FSS, 100%MERSS, 99%MERSS, 98%MERSS) and it takes between 217 and 220 seconds for the full GA run of 100,000 function evaluations (Table 5.1), on PC with following configuration: Intel Core 2 Duo @ 3.5GHz and RAM 3GB.

5.4.2 Results and Discussion

The methodology for selecting feasible non-dominated solutions from entire range of results was successfully introduced in Chapter 3 and Chapter 4. Presented results proved that the external screening method provides a good range of feasible non-dominated solutions and ensures that all valuable solutions are kept. Therefore, the same approach was applied to network presented herein. Firstly, the feasible, cost-entropy non-dominated designs for each run were identified. Later, cost-entropy non-dominated solutions from 50 runs were merged together in order to identify final POF. The process was repeated four times, for each POF separately (i.e. FSS, 100%MERSS, 99%MERSS and 98%MERSS). Comparison of all POFs is presented in Figure 5.3a. It can be noticed that RSS POFs supplement each other and even overlap. Therefore, the non-dominated solutions for all three RSS POFs were merged and screened for final cost-entropy non-dominated designs. Such achieved POF is presented alongside with FSS POF on Figure 5.3b. It can be observed on Figure 5.3a and 5.3b that FSS POF is dominated by RSS POFs almost on the entire length. In other words, the solutions obtained using RSS have higher entropy, thus reliability for similar cost and vice versa, for the same entropy value, the RSS solutions are cheaper. What is more, the RSS method identifies cheap solutions with lower entropy (i.e. between 3.1 and 3.3 entropy value) that are not recognised by FSS. FSS on the other hand produce solutions with the highest entropy value, which are not identified by RSS. This small limitation occurs because the SSR presented herein is based on 5 pipe diameters. So for this particular network, 14 available pipe sizes are automatically reduced to 5 once the feasible solution is achieved. Having 5 pipe diameters instead the full range makes the algorithm progressing faster but it may get harder to achieve the near global maximum entropy value. Nevertheless, it should be highlighted that there are very low chances that design with near global maximum entropy value would be chosen by decision maker due to very high cost. Solutions with slightly lower entropy value and lower cost are in practice more valuable since the solutions near global entropy value are usually highly expensive. For example, the solution with highest entropy, identified by FSS approach, has an entropy value of

4.373 and a cost of 4.66x10^8 Rupees. Whereas the design with highest entropy produced by 100%MERSS has an entropy value of 4.330 and cost of 4.5x10^8 Rupees. Therefore, the increase in entropy between 100%MERSS and FSS is marginal and smaller than 1%, whilst the increase in cost is nearly 3.4%. Hence, it is not worth to pay this extra cost with so low increase in entropy, thus reliability. In order to identify where exactly the FSS solutions are not dominated by RSS designs, the FSS POF was merged with RSS POF and cost-entropy non-dominated solutions were identified and presented on Figure 5.3c. It can be noticed that only very small fraction of the plot is covered by solutions obtained using FSS. In much larger part of the plot, solutions identified by RSS approaches has higher entropy for similar cost, thus dominate FSS designs. Moreover, there are as little as 23 solutions identified by FSS that if considered without RSS solutions, would not be enough for the decision maker to choose from. Whereas the RSS approaches produced 124 cost-entropy non-dominated solutions (47 designs identified by 100%MERSS, 62 by 99%MERSS and 15 by 98%MERSS).

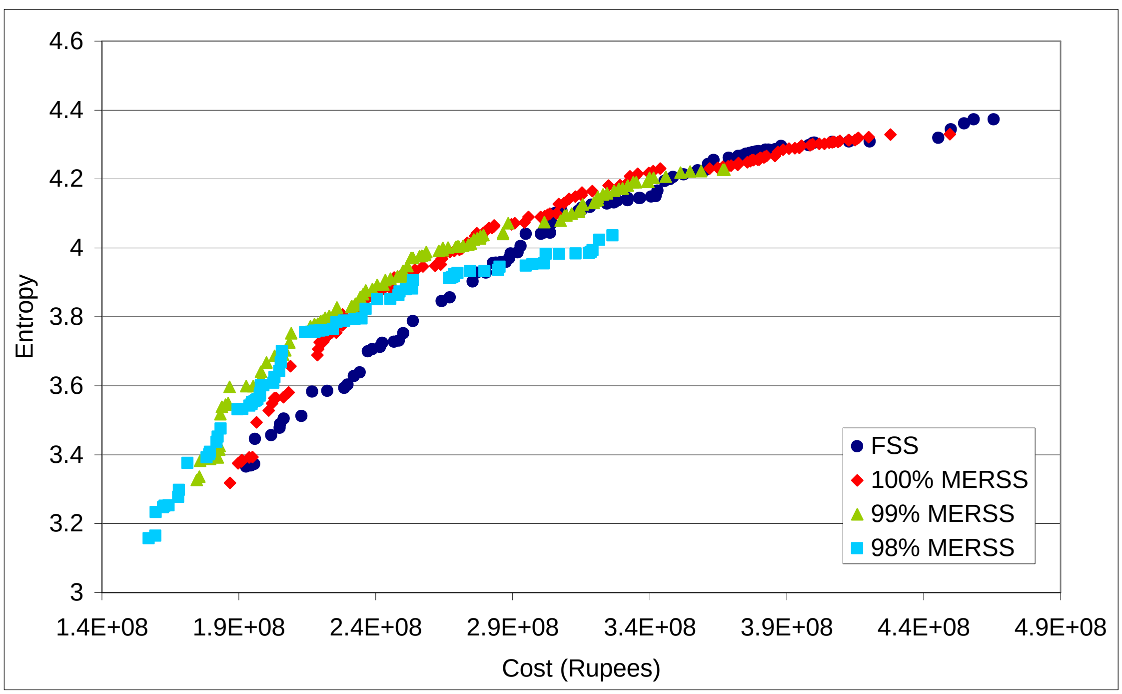


Figure 5.3a Entropy versus cost POFs for FSS, 100%MERSS, 99%MERSS, 98%MERSS based on entire history of results from 50 runs

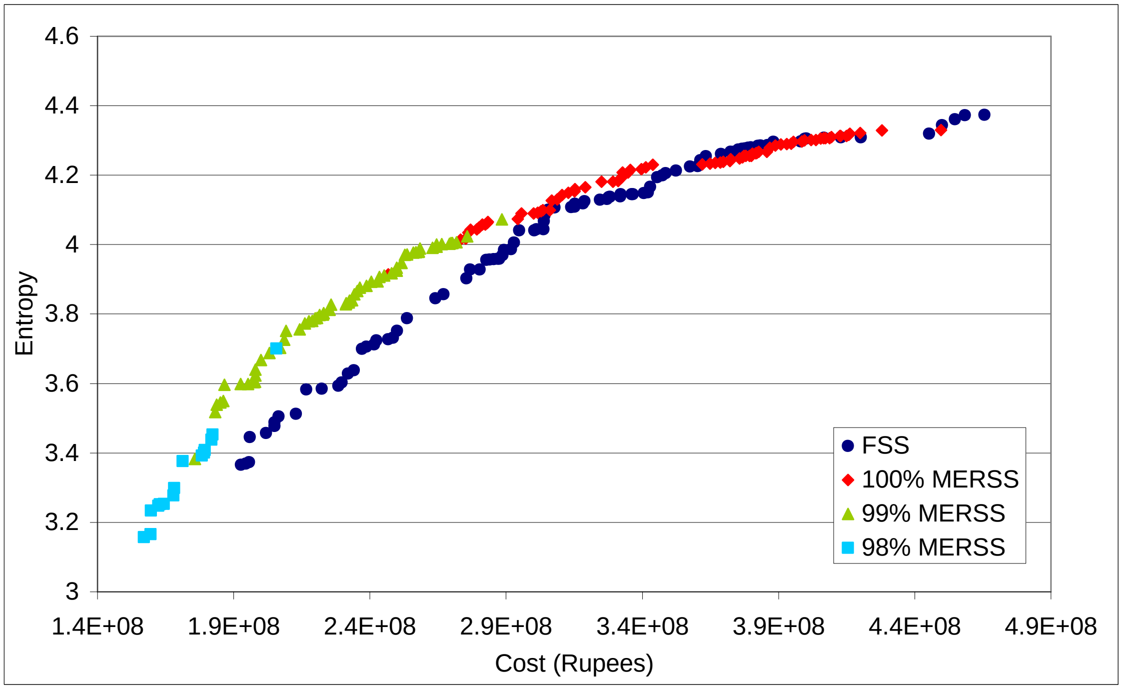


Figure 5.3b Entropy versus cost POFs for FSS and merged MERSS based on entire history of results from 50 runs

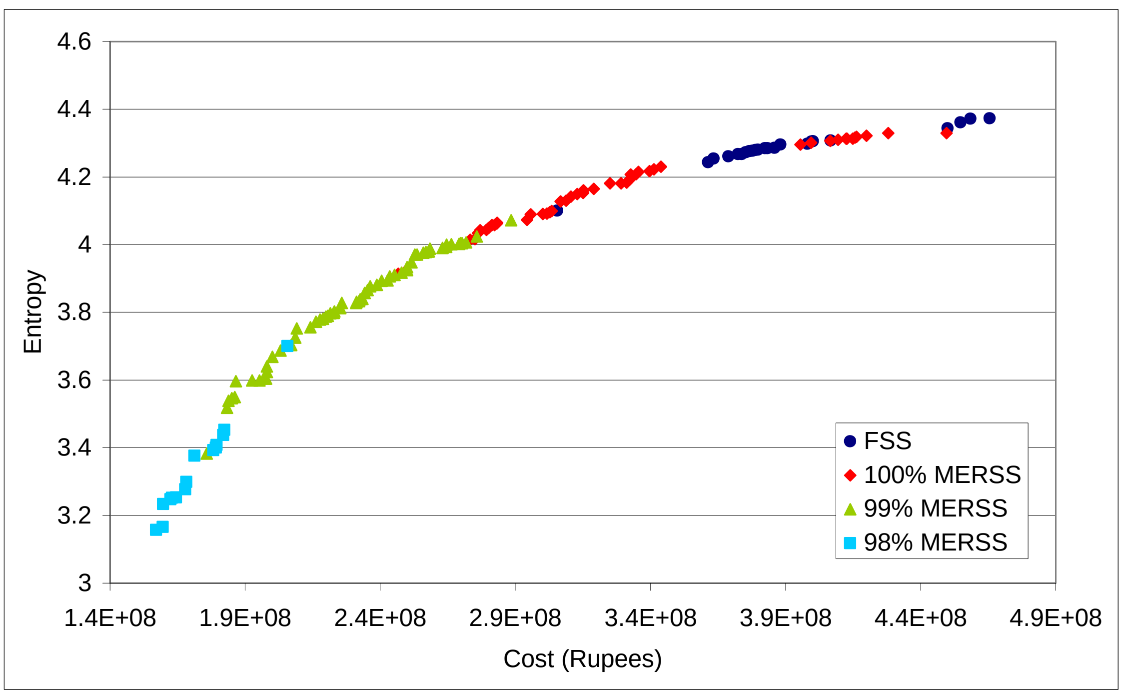


Figure 5.3c Entropy versus cost POFs for merged FSS and MERSS based on entire history of results from 50 runs

In addition to final merged POF presented on Figure 5.3, further analysis was also performed on individual cost-entropy non-dominated POFs for 50 randomly generated runs. However, due to space limitation all 50 GA runs were gathered and attached in Appendix D. The entropy and cost scales were kept the same for all 50 graphs for ease of comparison. Despite the fact, whether it is FSS or RSS POF, the graphs can vary much between each other (i.e. plot from one run could be completely dominated by plot from other run). This is absolutely normal since the initial population depend on initial data thus random seed. Therefore it is important to have multiple runs instead single run in order to assess not only the GA performance but also in case of designing actual network since the single run may produce the best or worst POF and be misleading.

Even a quick look at individual plots is enough to identify that majority of RSS POFs outperforms FSS POFs in terms cost and entropy. Nevertheless, all 50 runs were divided into four subcategories in order to identify number of runs with the best and worst RSS POFs in comparison to FSS POFs. Such data are presented in Figure 5.4

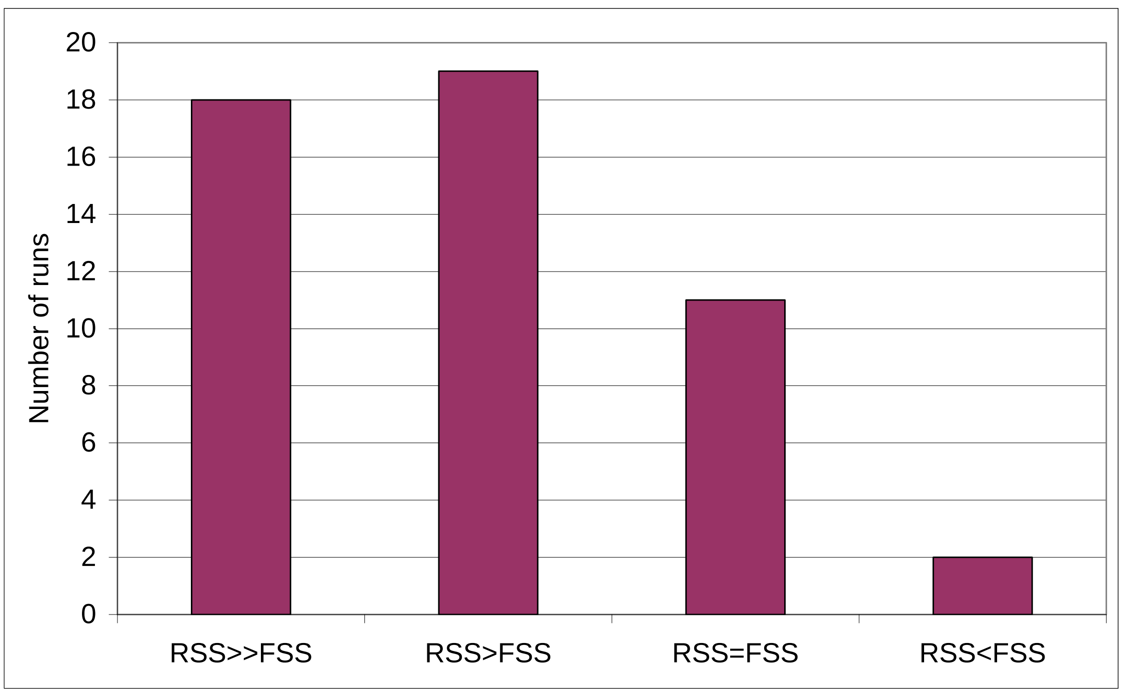


Figure 5.4 Comparison of performance of RSS POFs and FSS POFs for individual runs

First column (i.e. RSS>>FSS) represents runs where RSS approaches produced POFs with not only cheaper solutions for similar entropy value, but also higher or at least the same maximum entropy value. Good example of this category would be run 1, 6, 13, 19, 20, 21, etc. (Appendix D). Second column (i.e. RSS>FSS) correspond to runs where POFs based on RSS have cheaper design for similar entropy value, however the FSS approach produce designs with highest entropy value. This group of results is the biggest among all four. Third column (i.e. RSS=FSS) stands for runs where RSS solutions dominate the FSS designs, but the maximum entropy values for designs achieved using RSS are considerably lower than highest entropies obtained using FSS. For example run 2, 5, 7, 18, etc. (Appendix D) would belong to this category. As mentioned earlier, the limitation in entropy range (i.e. actually in high entropy values) is the results of pipe sizes reduction (i.e. FSS is based on 14 pipe size diameters while RSS approach on 5 pipe size diameters). The last column (i.e. RSS<FSS) corresponds to runs when FSS POFs would dominate the RSS POFs. It has to be highlighted that only 2 runs (run 10 and 43) belongs to this category, which in comparison to total number of all 50 runs is relatively small number.

Even though it has been proved that the novel method of selecting non-dominated feasible designs from entire range of results is superior to originally generated POF (i.e. last generation) there are high chances that in some cases only this POF will be used. Therefore, additional comparison of POFs originally generated by GA, were presented on Figure 5.5a and Figure 5.5b. Once again, Figure 5.5a presents four separate POFs (i.e. FSS, 100%MERSS, 99%MERSS and 98%MERSS) obtained after merging 50 random runs. Whilst Figure 5.5b shows two POFs; one for FSS and one for RSS, which represents final cost-entropy non-dominated designs for three RSS approaches. It can be noticed, the only difference between Figure 5.3a and 5.5a, also 5.3b and 5.5b is that the POF obtained from last generation (Figure 5.5a and Figure 5.5b) has less designs than POF made with the use of entire history of solutions (Figure 5.3a and Figure 5.3b). Overall conclusion that can be drawn is that the results achieved using SSR outperform

FSS solutions with higher entropy for the same cost and vice versa, lower cost for the same entropy value.

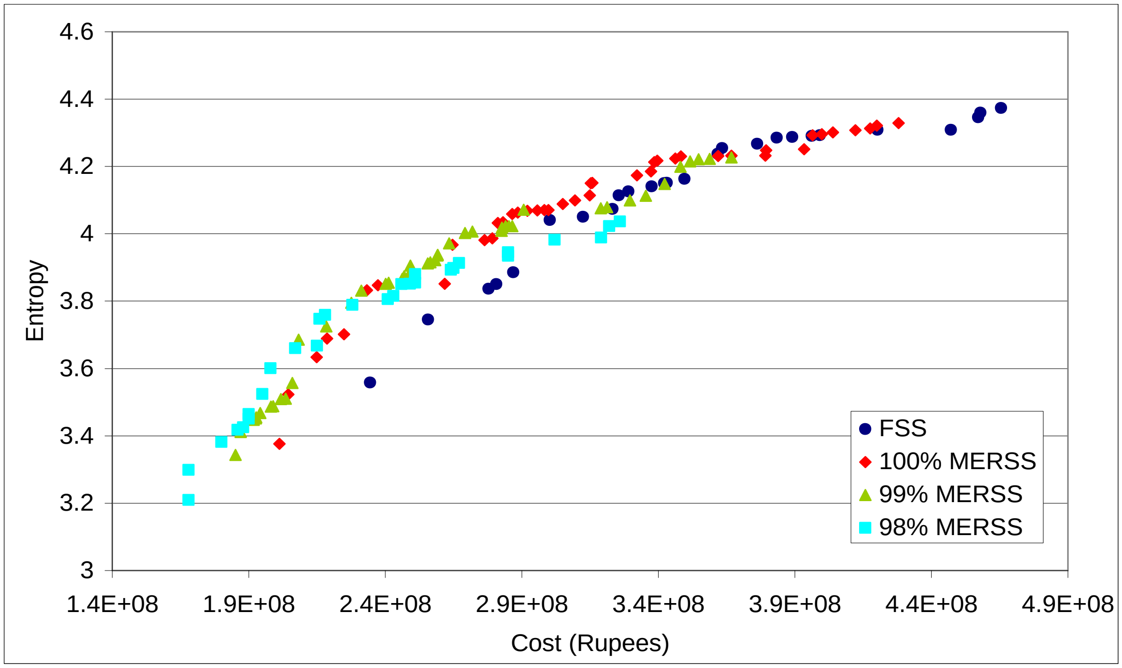


Figure 5.5a Entropy versus cost POFs for FSS, 100%MERSS, 99%MERSS, 98%MERSS based on results from last generation from 50 GA runs

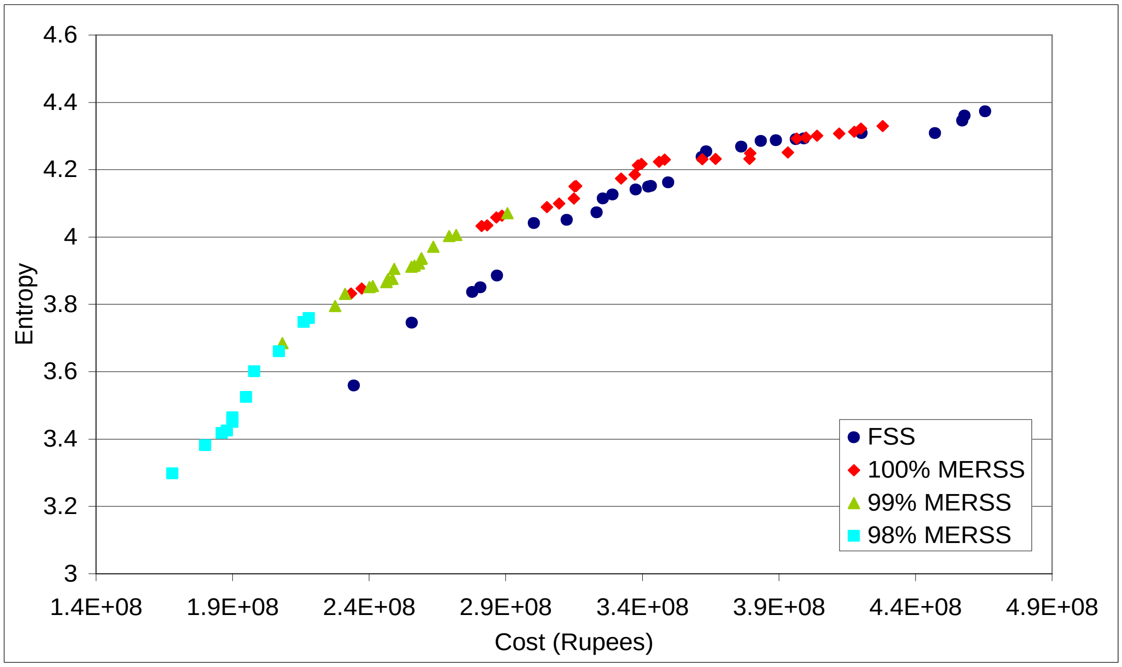


Figure 5.5b Entropy versus cost POFs for FSS and merged MERSSs based on results from last generation from 50 GA runs

In order to investigate the consistency and performance of presented GA with such big network simulated for 50 random runs with 100 000 function evaluations in each run, the additional software had to be used. For these purposes, the external software written in Perl language that has already been described in Chapter 4 was modified and employed. As mentioned before, the software identifies the highest, lowest and average value among population of size 100 starting from 1st generation, then for every 20 generations and finishing on the very last one. This was done for feasible and infeasible solutions as well as for feasible designs only. Afterwards, data from 50 random runs were gathered together in excel spreadsheet and maximum, minimum and average values were recognized among all runs. Such acquired data were gathered together and presented on Figures 5.6 to 5.9. In terms of cost, the external software identified the lowest value, since the aim of the optimization process is to minimize the cost. Two cases were analysed; one for feasible and infeasible solutions (Figure 5.6) and another one for feasible designs only (Figure 5.7). For the deficit (Figure 5.9), the average value among all population was used as it takes into consideration both infeasible solutions that have the actual value and feasible solutions which are equal to zero. In terms of entropy (Figure 5.8), external software identified the highest value, as maximization of the entropy is the purpose of the optimization. Similarly as in previous chapter, only the feasible designs were taken into consideration, since only the feasible solutions have realistic entropy value.

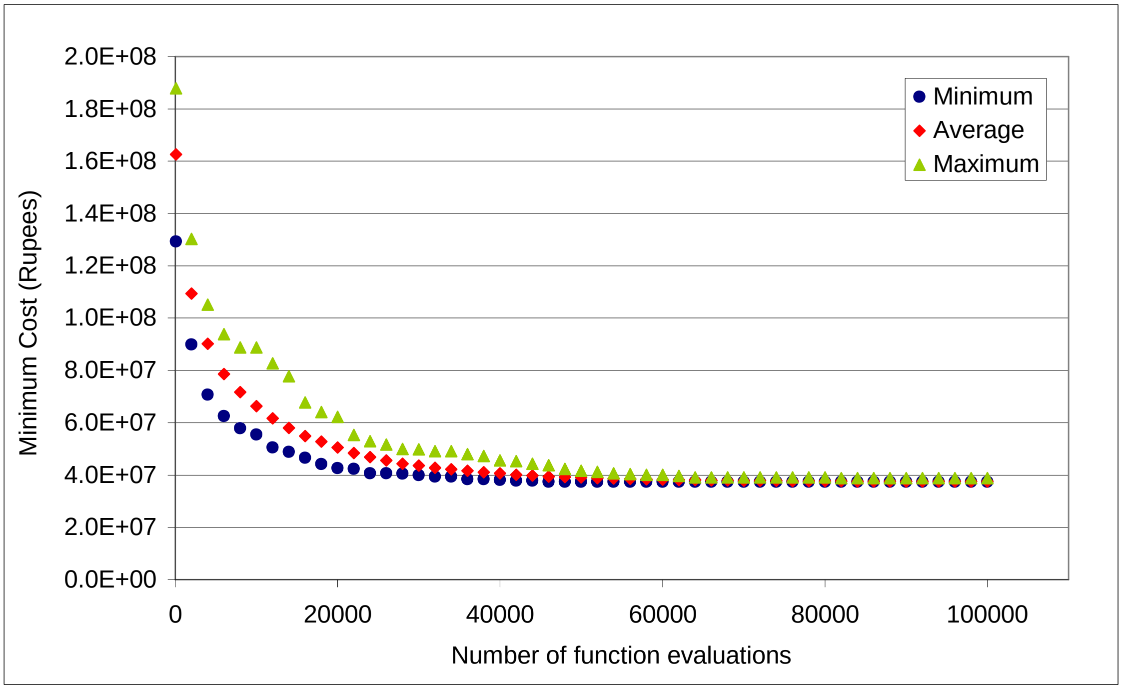


Figure 5.6a Search performance of FSS GA based on feasible and infeasible designs from 50 GA runs

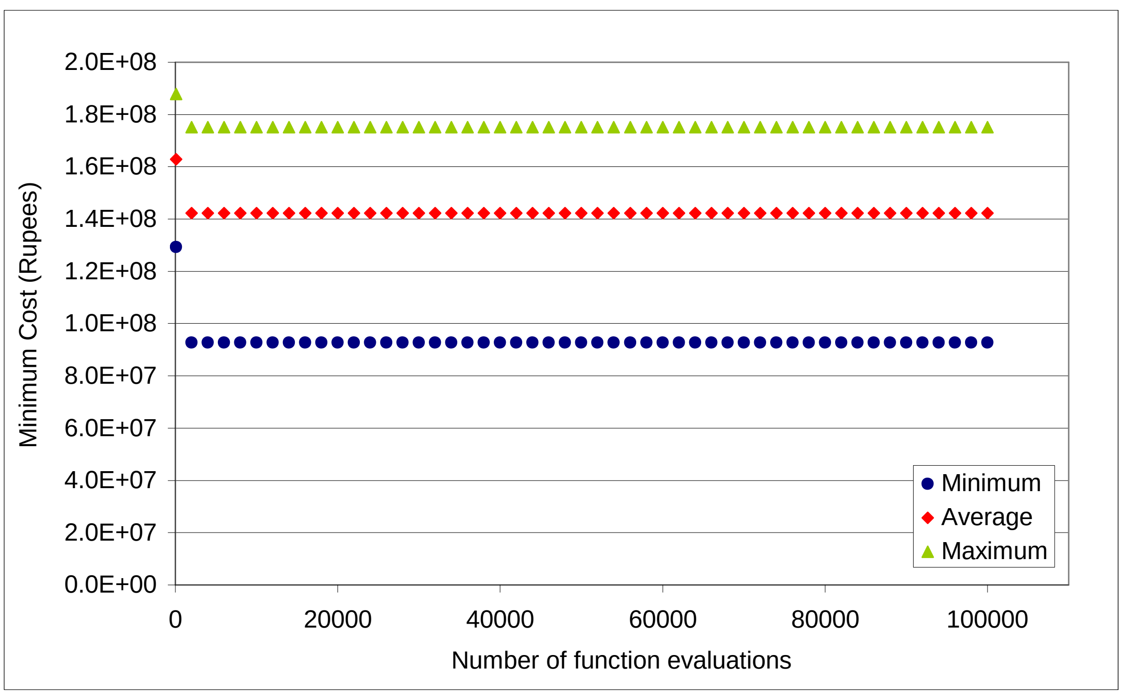


Figure 5.6b Search performance of 100%MERSS GA based on feasible and infeasible designs from 50 GA runs

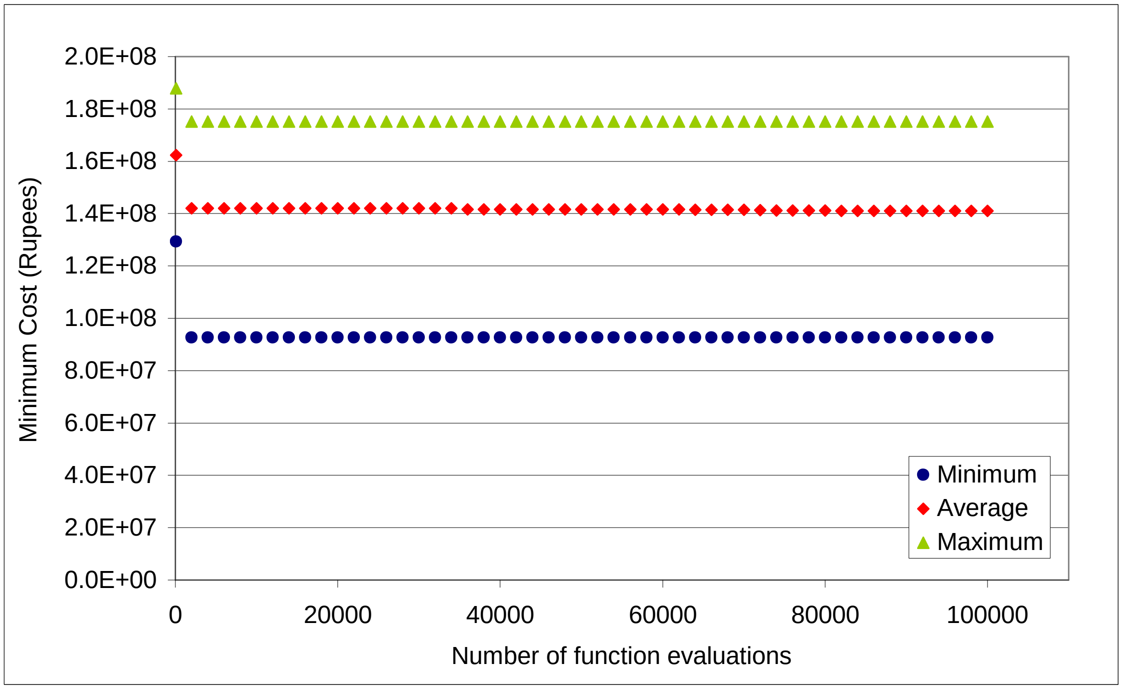


Figure 5.6c Search performance of 99%MERSS GA based on feasible and infeasible designs from 50 GA runs

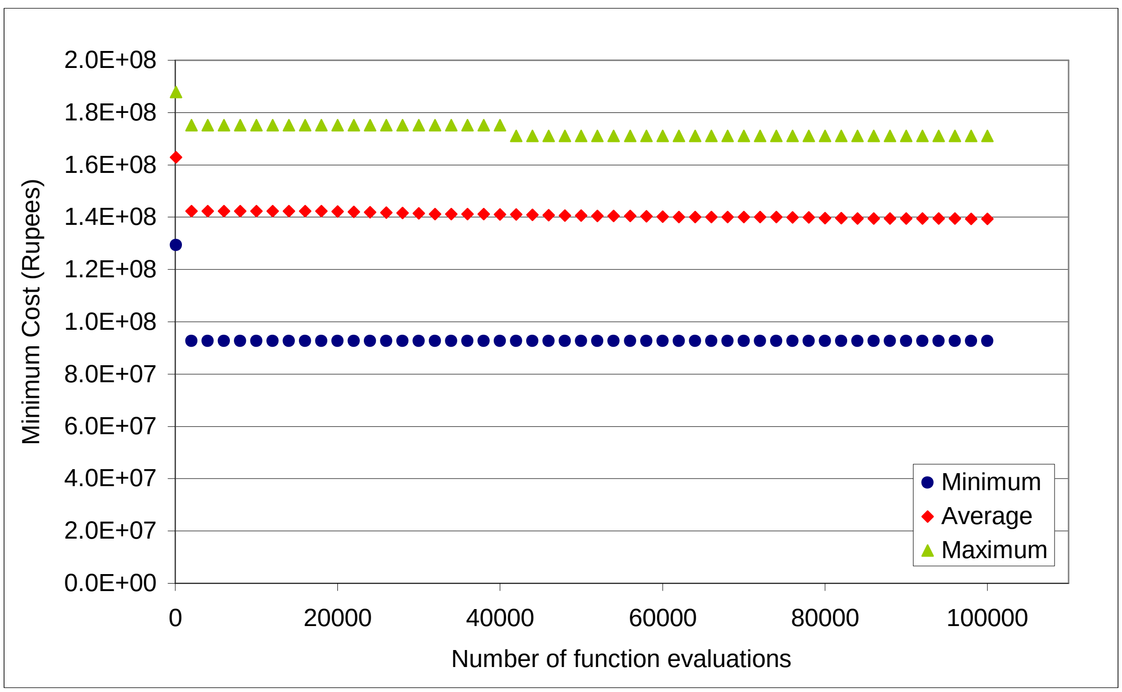


Figure 5.6d Search performance of 98%MERSS GA based on feasible and infeasible designs from 50 GA runs

Figure 5.6 demonstrates GA performance in terms of cost for FSS and three cases of RSS. It should be mentioned, that the scale for cost was kept the same for all Figures (5.6a, 5.6b, 5.6c and 5.6d), so it is easy to identify differences between the costs for FSS and RSS, which will be explained later in this paragraph.

It can be noticed that RSS converged extremely quickly in comparison to FSS approach. For all three cases of RSS stabilization can be observed with only 2 000 function evaluations (i.e. 20 generations and population of 100) needed to converge. This rapid stabilization is the result of reducing the number of candidate pipe sizes once the GA with RSS reaches feasible solution. This leads to a reduction in the size of the solution space that corresponds to $(5/14)^{34} = 6.3\times10^{-16}$. In other words the solution space is reduced to a fraction of 6.3×10^{-16} or $1/1.60\times10^{15}$ of the full solution space. Therefore, the full range of 14 pipes diameters is reduced to 5 pipe diameters which makes the algorithm progress much faster. In the case of FSS about 40 000 function evaluations (i.e. 400 generations) are needed to converge. This leads to a conclusion that all three RSS cases (i.e. 100%MERSS, 99%MERSS and 98%MERSS) could be performed within the same or even lower number of function evaluations that is required for the FSS to converge. Moreover, simulating three RSS cases would not take more time than simulating FSS. This way, not only wider range of available entropy values could be achieved (Figure 5.3) but also cheaper solutions with the same entropy would be identified, as presented in Figure 5.3.

As pointed out earlier, there is some difference in values of costs between FSS and RSS. Algorithm based on FSS produce cheaper designs than solutions generated with RSS method. Nevertheless, it should be remembered that those very cheap designs obtained using FSS are also highly infeasible, which makes them useless. Whereas, the RSS method produce solutions located mostly at the boundary between feasible and infeasible region, thus making the cost higher. For that reason, it has been decided that GA performance in terms of cost should also be performed on feasible designs only. This should clarify and prove that the actual cost of feasible solutions is lower for designs achieved using RSS approaches. Those results were presented in Figure 5.7.

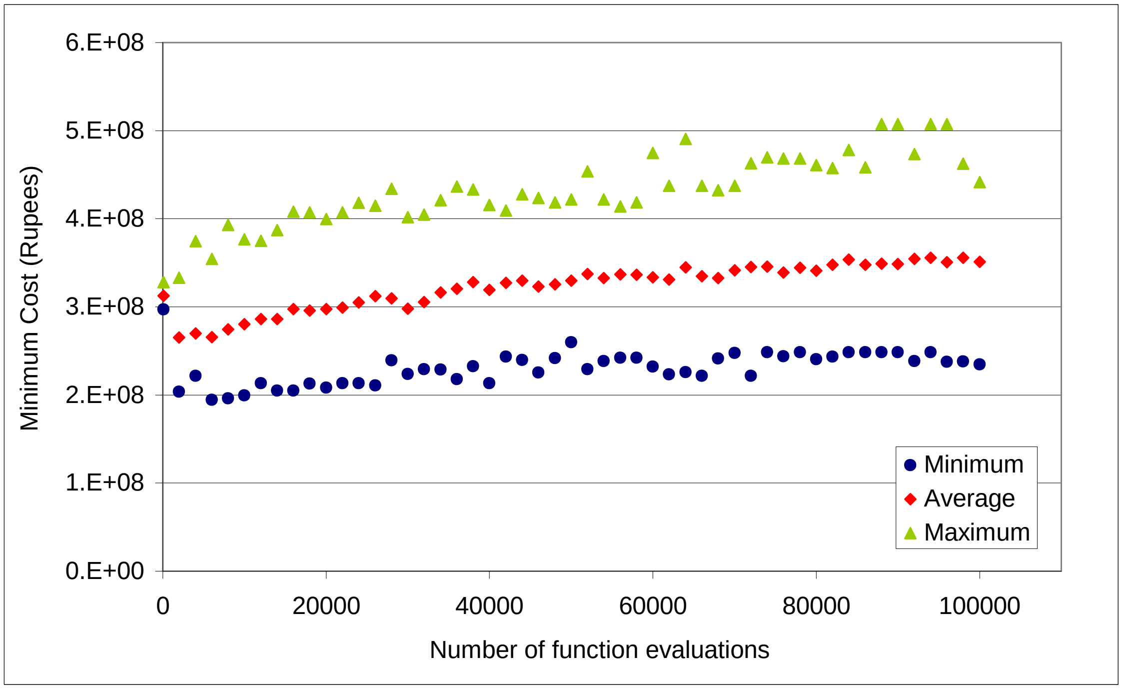


Figure 5.7a Search performance of FSS GA based on feasible designs from 50 GA runs

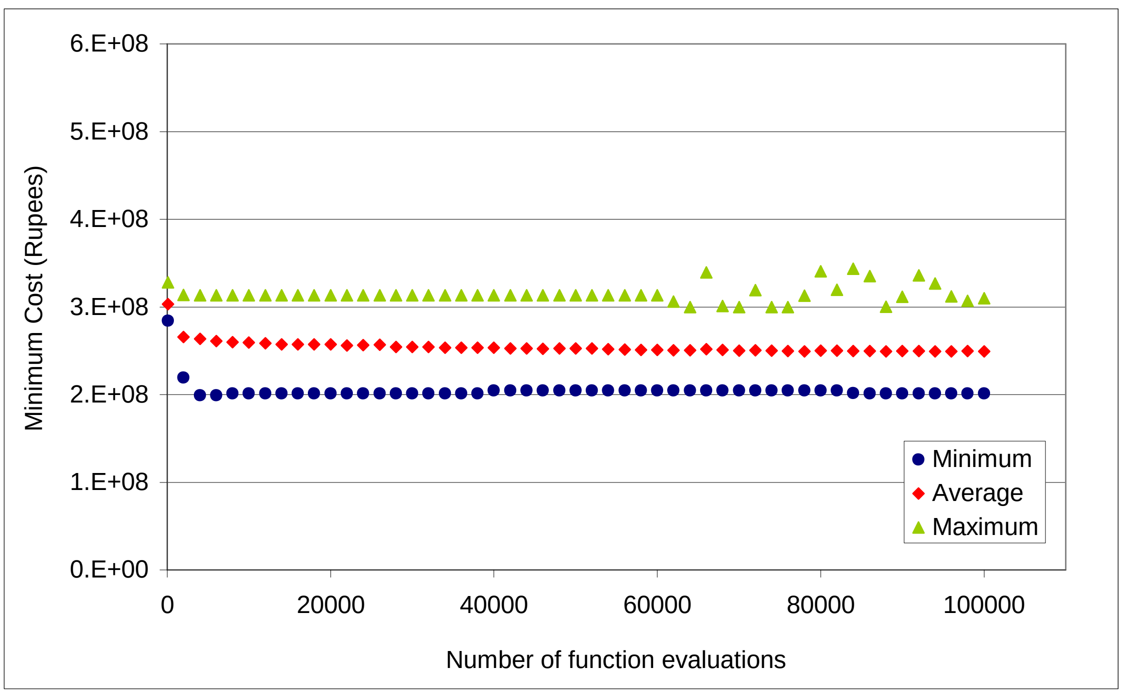


Figure 5.7b Search performance of 100%MERSS GA based on feasible designs from 50 GA runs

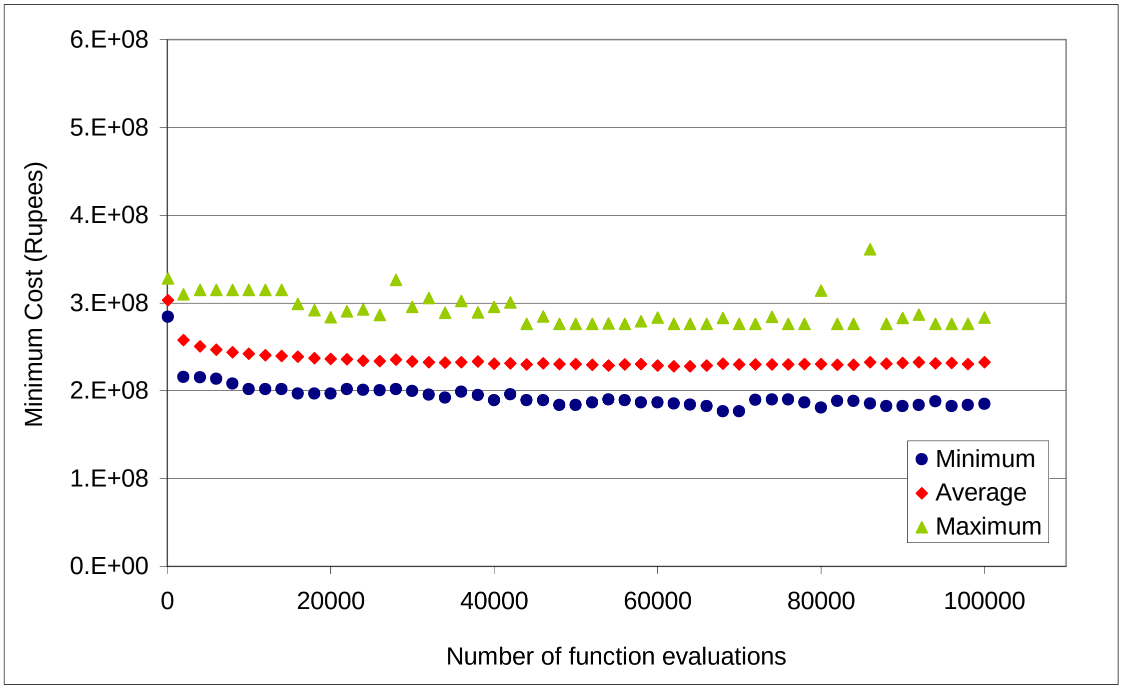


Figure 5.7c Search performance of 99%MERSS GA based on designs from 50 GA runs

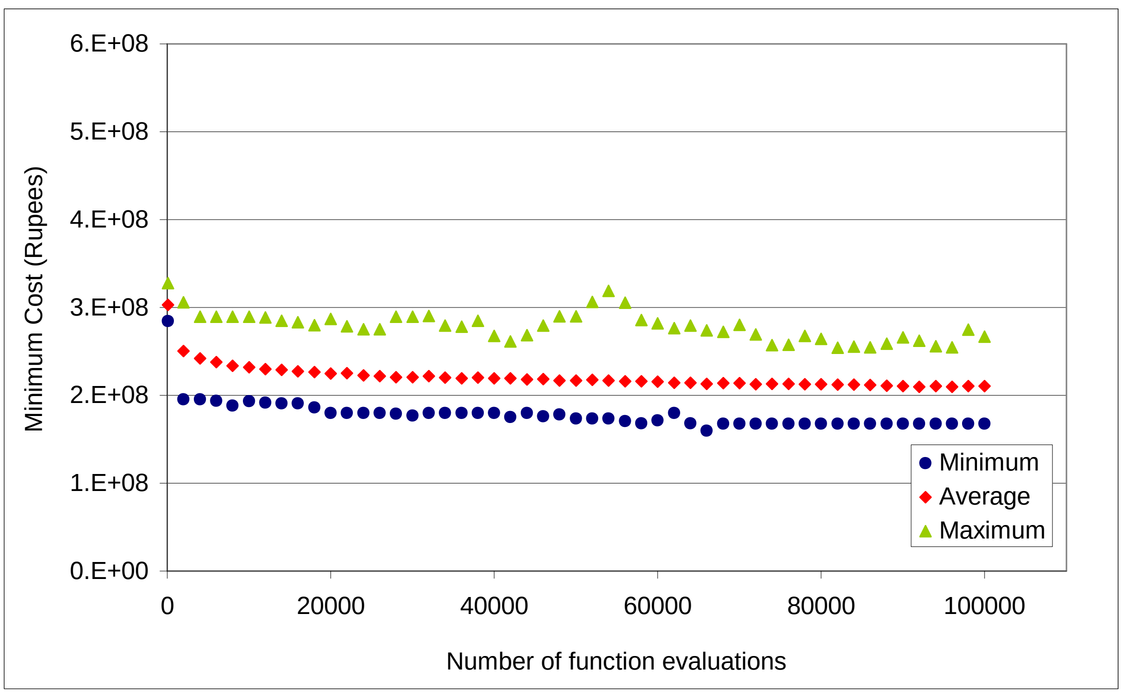


Figure 5.7d Search performance of 98%MERSS GA based on feasible designs from 50 GA runs

It is easy to notice on Figure 5.7 that for all RSS cases the cost decrease with the increase of function evaluations, so exactly the same way as when the feasible and infeasible solutions were taken into consideration (Figure 5.6). Nevertheless, it does not happen for FSS approach, as it is clear from the Figure 5.7a that the minimum cost decrease at the very beginning and then slowly increases with increasing number of function evaluation. This is expected outcome as apart from cost minimization the FSS GA also try to reach the global maximum entropy value, that as explained earlier in this chapter, raises the cost significantly and disproportionately to the increase of entropy. Therefore, when considering the feasible designs only, the cost and entropy get higher with the increase of number of function evaluations. Obviously the aim of RSS GA is the same as FSS GA (i.e. reaching for the highest possible entropy value, while minimizing the cost and achieving feasibility), but since the RSS method is based on reduced search space (i.e. lower number of pipe diameters), there is less possibilities for different pipe diameters configurations. Therefore the RSS GA search mostly the space at feasibility boundary, thus there is more chances to find feasible designs with high entropy and considerably low cost. Moreover, it is noticeable on Figure 5.6 that all three RSS cases achieve lower minimum cost for feasible solutions than the FSS. This goes along with previous results presented on Figure 5.3 that RSS method allows to identify cheaper feasible deigns than FSS method. Some fluctuations can be observed on all four graphs. Such imbalance is an expected outcome as presented results are based on feasible solutions only, thus it is not like proper GA performance regarding stabilisation and convergence. Nevertheless, the RSS has fewer fluctuations and seems to converge far quicker.

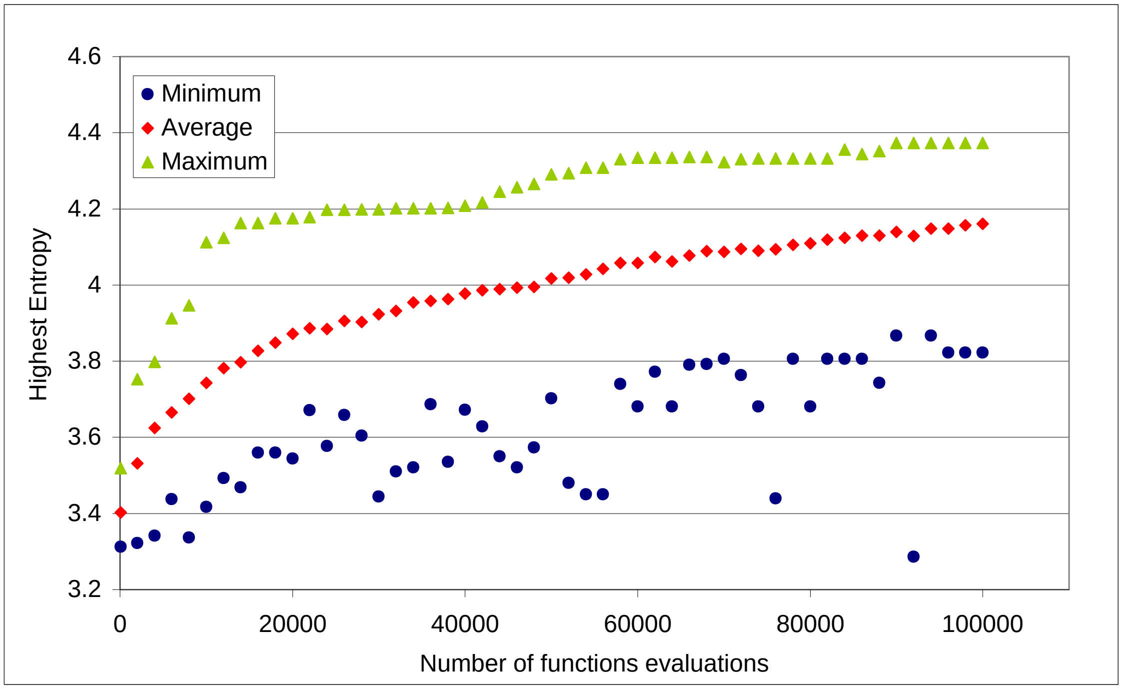


Figure 5.8a Search performance of FSS GA based on feasible designs from 50 GA runs

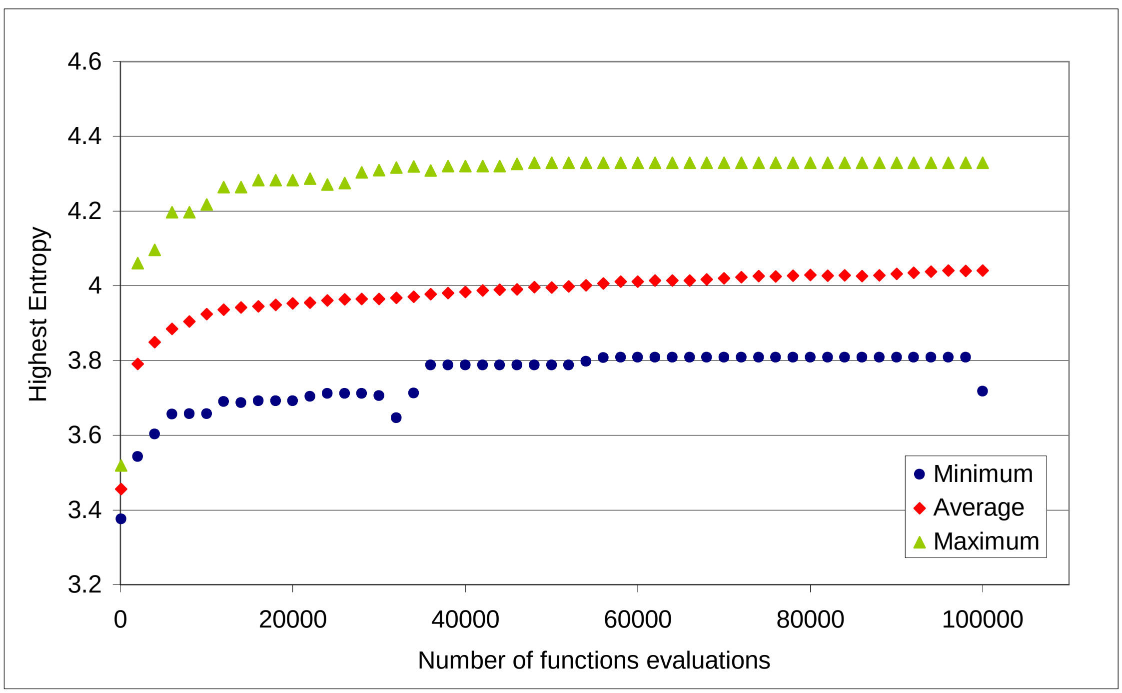


Figure 5.8b Search performance of 100%MERSS GA based on feasible designs from 50 GA runs

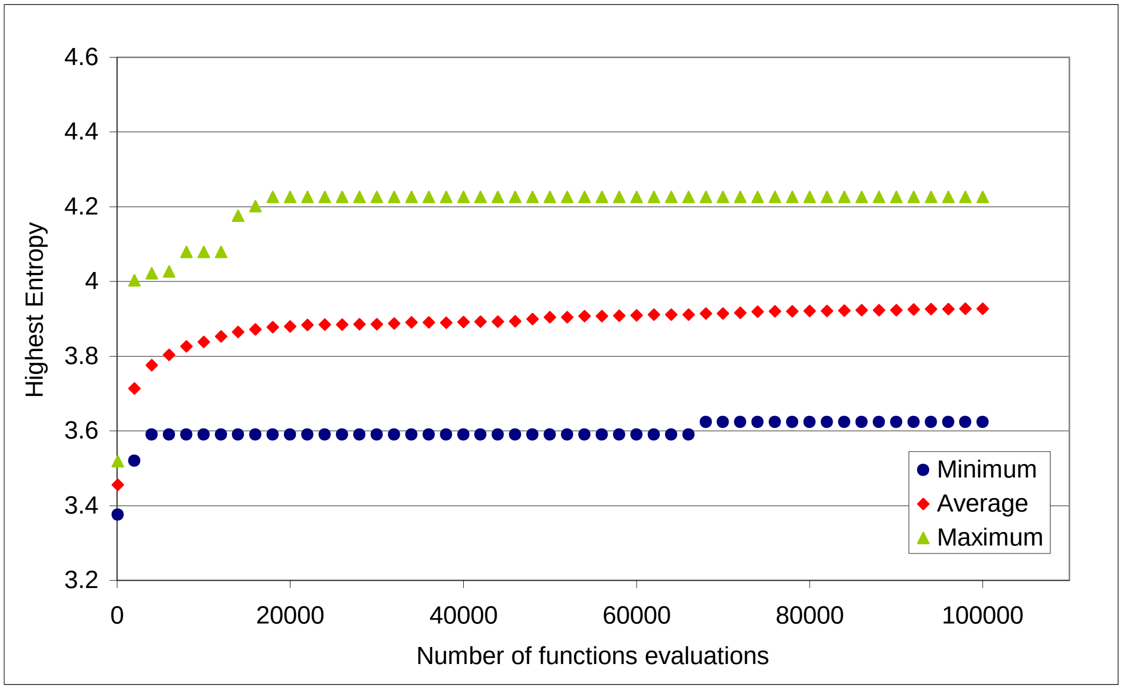


Figure 5.8c Search performance of 99%MERSS GA based on designs from 50 GA runs

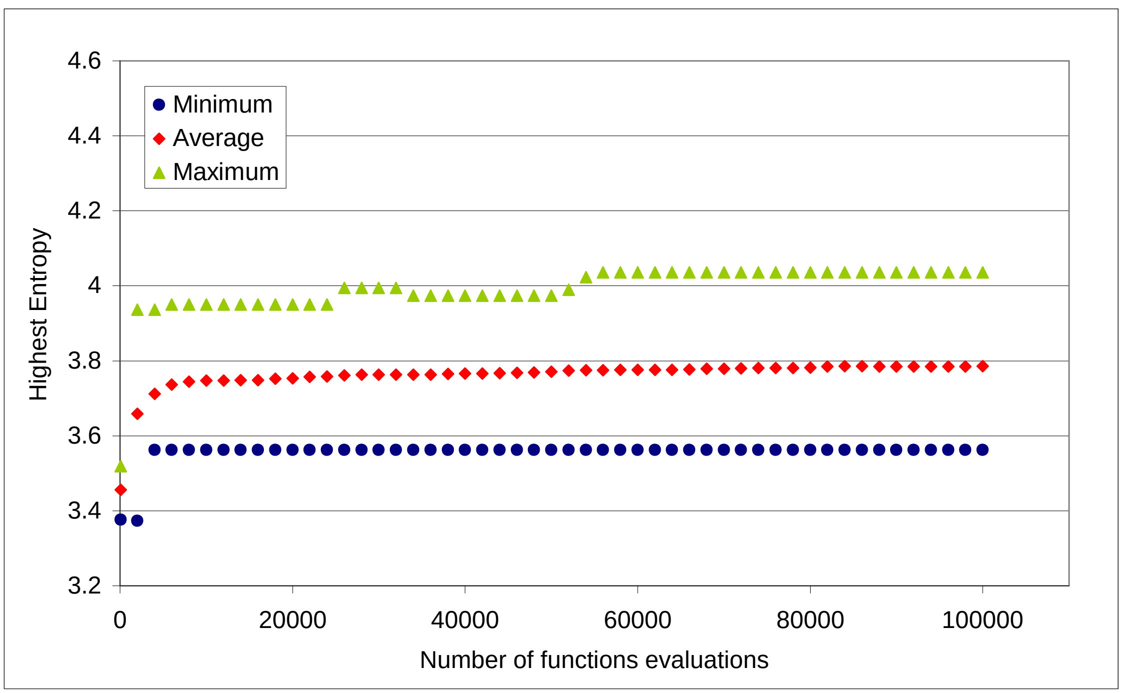


Figure 5.8d Search performance of 98%MERSS GA based on feasible designs from 50 GA runs

The evolution and convergence characteristic for entropy is presented on Figure 5.8. Only feasible solutions were considered. Similarly to results showed in Figure 5.7, analysed objective (i.e. entropy in this case) increase and stabilise far quicker for RSS approaches than FSS method. For the RSS approaches (Figure 5.8b, Figure 5.8c and Figure 5.8d) the entropy value seems to be boosted rapidly within first few generations, and then the increase is much slower, nearly insignificant. This rapid stabilization is the result of reducing the number of decision variables once the GA with RSS reaches feasible solution. Having 5 pipe diameters instead the full range of 14, makes the algorithm progress faster and reach to the maximum possible entropy value (i.e. highest for particular set of reduced pipe diameters) within fewer generations. For the FSS approach (Figure 5.8a), the entropy has also the quickest growth in the first generations, however it does not seems to stabilise but slowly increases until the GA termination. Moreover, quite a few fluctuations can be observed for results based on FSS approach, especially when the minimum of highest entropy is considered.

It is also easy to notice from Figure 5.8 that the GA based on FSS produces results with slightly higher entropy value than the RSS approaches. Obviously the percentage of ME value used is of great importance. It is not even expected that GA will identify highest entropy value when 98% MERSS or 99%MERSS are used since the main role of those percentage approaches is to identify cheaper solutions, thus lower entropy values. Nevertheless, the maximum of highest entropy value for 100% MERSS is 4.329 while the maximum of highest entropy value for FSS is 4.370. The small difference equals to 0.041 between those entropy values is the limitation that occurs because the RSS is based on restricted number of pipe sizes. Detailed explanations have been presented earlier in this chapter when Figure 5.3 was analysed.

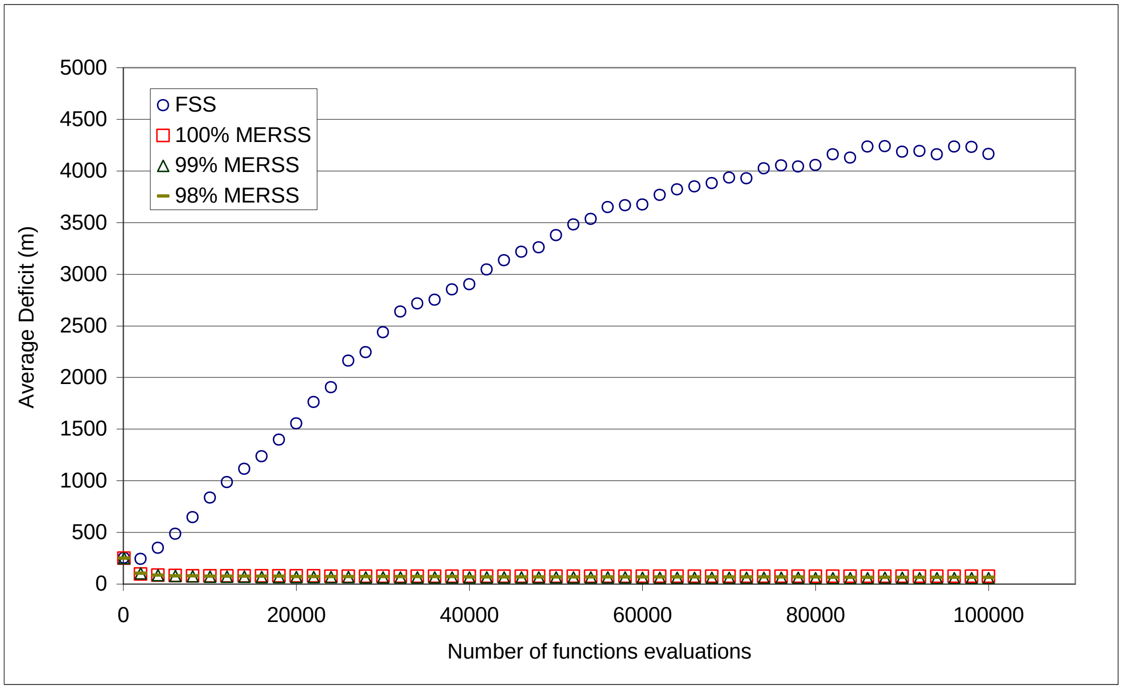


Figure 5.9 Search performance of GA based on feasible and infeasible designs from 50 GA runs

Figure 5.9 demonstrates GA performance in terms of deficit for FSS and three cases of RSS. For the FSS, the average deficit increases with the GA progression which is consistent with results presented in previous chapter (Figure 4.10). As explained in Chapter 4, there are many feasible but highly expensive solutions in the first generations causing that the deficit value is fairly low at the early stage. With the cost minimization the number of feasible solutions gets lower, thus giving higher value of average deficit. The situation is completely different for all RSS approaches as the average deficit decreases to the level close to 0m (i.e. lower than 100m of deficit) and remains stable until the GA termination. The lowest average deficit occurs for 99%MERSS and is around 58m (65m for 98%MERSS and 73m for 100%MERSS). Obviously, the value of average deficit will never be equal to zero since some infeasible solutions will always occur. It should also not be forgotten that all feasible designs are treated as neutral (i.e. zero value is assigned to feasible solutions despite the actual surplus head, while infeasible solutions have positive values). Therefore there are no values that could balance or lower the deficit. Nevertheless, the FSS approach has the average deficit value around 4 162m just before the GA termination. Hence, the average deficit below

100m for the RSS approaches is extremely small. This highly desirable attribute of identifying the solutions at the boundary of feasibility is achieved due to reduction of all set of possible pipe. Using ME based RSS approaches allows narrowing the entire search space to boundary between feasible and infeasible regions. As Kadu et al. (2008) highlighted, the optimal (i.e. cost effective) designs are located at the boundary of feasible and infeasible solutions. Siew and Tanyimboh (2012) proved that algorithm with boundary search techniques is able to locate optimal solution in considerably lower number of function evaluations required. Presented GA with RSS approaches reinforce above statements.

Table 5.1 Convergence and consistency statistics based on 50 random GA runs

Measure	**Minimum**	**Mean**	**Median**	**Maximum**	**Standard deviation**
CPU time (for entire 100,000 FEs) [s]:					
FSS	210	217	216	227	4.693
100% MERSS	210	219	220	227	2.968
99% MERSS	214	220	221	229	3.557
98% MERSS	212	219	218	227	3.443
Number of feasible designs per run:					
FSS	2336	3985	3984	5750	945.831
100% MERSS	6990	11831	11362	18389	2540.812
99% MERSS	4737	10363	10172	16831	2613.134
98% MERSS	4939	10183	9571	16597	3099.759
Function evaluations for convergence in terms of entropy[a] (3% tolerance used as convergence criterion)					
FSS	$4.3*10^{-35}$	$39.8*10^{-35}$	$32.3*10^{-35}$	$107.5*10^{-35}$ [b]	31.269
100% MERSS	$4.3*10^{-35}$	$11.2*10^{-35}$	$4.3*10^{-35}$	$107.5*10^{-35}$ [b]	22.396
99% MERSS	$4.3*10^{-35}$	$9.3*10^{-35}$	$4.3*10^{-35}$	$81.7*10^{-35}$	13.780
98% MERSS	$4.3*10^{-35}$	$7.8*10^{-35}$	$4.3*10^{-35}$	$25.8*10^{-35}$	6.683

[a] Point beyond which the improvements become very insignificant.
[b] Evaluation function at which the GA run was terminated (i.e. last evaluation function).

Table 5.1 presents convergence and consistency statistics based on 50 random GA runs. It can be noticed that the CPU time is nearly the same for all three approaches based on 100,000 FEs. Therefore, the RSS approach (i.e. identification of design with ME and recognition of set of 5 pipe sizes that could change from generation to generation) does not slow down the GA process significantly. Furthermore, considering that RSS requires far fewer function evaluations to converge, the 3 different RSS approaches could be performed within the same time as one full FSS simulation. It is also clear from Table 5.1 that RSS approach produces more feasible solutions than FSS. On average, there are approximately three times more feasible solutions per single run when using any of the RSS approaches. Therefore, there are more alternative solutions to consider for any additional in depth analyses, for example reliability and failure tolerance. Consequently, more solutions will remain for the decision maker to choose from. Similar conclusions were drawn based on results presented on Figures 5.3 and Figure 5.5.

Function evaluations (FEs) for convergence in terms of entropy (Table 5.1) means the ratio of average FEs to achieve convergence to total size of solution space. Since the search space for this particular network is extremely large (i.e. $14^{\wedge 34}$) rates for convergence are coming out as low values. Based on initial analysis for all 50 random runs it has been decided that improvements in entropy becomes insignificant if is lower than 3% between neighbouring generations. Therefore, the last generation that has increase in entropy equal or lower than 3% in comparison to previous generation is treated as final generation required to achieve convergence. Obviously all generations afterwards must also have increase in entropy equal or lower than 3%.

It can be noticed that minimum rate for convergence is the same for all approaches (i.e. FSS and three RSS approaches). Nevertheless, it should be mentioned that in case of FSS it is only one such case, while for all RSS approaches it happens more often. The values of median prove that for the RSS approaches the minimum rate for convergence occurs very often as for those cases median is equal to minimum value. The same applies to maximum rate for convergence for FSS and 100%MERSS. Therefore, the minimum and maximum rate for convergence should be treated more like guidance rather than proper measures. Mean and median are more appropriate measures for this case. In terms of mean, the values for RSS approaches are significantly lower than for

FSS method. This confirms the fact that GA with RSS approach converges far quicker than GA based on FSS method. It is also noticeable that the lower percentage of ME for RSS, the smaller number of FEs required achieving convergence.

5.5 CONCLUSIONS

In this chapter, the formulation of efficient reliability-based search space reduction for WDNs has been presented. The approach is based on entropy and uses the importance of every path through network, which is included in the entropy function. The novelty of this research in the context of GAs is that the developed algorithm is dynamic, self-adaptive and does not require any initial testing or pre-defining the diameters. Moreover, three different percentage values of ME have been integrated in the procedure.

The algorithm has been applied to network previously used in search space reduction studies. Full search space method and reduced search space approach with three near ME values were studied. For the FSS the GA used all 14 pipe sizes, while for RSS only 5 pipe sizes were employed once the first feasible solution was identified. Sufficient number i.e. 50 randomly generated GA runs was performed for each case (i.e. FSS and RSS approaches). Detailed comparison for merged POFs based on FSS and merged POFs obtained using RSS has been presented alongside with analysis for single POFs of all 50 GA runs. It has been showed that different percentage approaches of ME proposed fill the gaps between each other and present nearly entire range of possible entropy values for particular network. Moreover, solutions obtained using RSS dominate and clearly outperform designs based on FSS in terms of cheaper cost for similar entropy value. Another significant advantage of any RSS approach over other FSS method is that it allows narrowing the entire search space to boundary between feasible and infeasible regions.

The consistency and performance of presented GA has been investigated for all objective functions (i.e. cost, entropy and deficit). It is evident that in case of RSS approach much smaller number of function evaluations is required for the GA to converge. This leads to conclusion that three near ME RSS simulations (i.e. three percentages of ME) could be performed within time required for single FSS simulation. That would provide not only better solutions (i.e. cheaper cost for similar entropy) but also wider range of entropy values and more solutions to choose from.

CHAPTER SIX

CONCLUSION

6.1 INTRODUCTION

Water distribution networks (WDNs) are one of the most expensive infrastructure systems. That makes the cost minimization as extremely important objective when designing the network. For many years researchers concentrated on minimization of the network cost. It has inevitable influence on network reliability, as the cheapest design is not satisfactory in terms of reliability. Then, it was recognized that the reliability of WDN is equally important as network cost, as the optimal design is the cheapest one that will satisfy demands under normal and abnormal conditions. Nevertheless, the estimation of reliability is extremely time-consuming calculation as requires simulating the WDN a very large number of times. This encouraged researchers to look for surrogate measures, comparably easier to estimate than reliability.

Statistical entropy is one of few existing reliability indicator. Entropy is particularly advantageous since it involves only the flow in the pipe and the demands at the nodes that are normally given. Over the years, the entropy has been incorporated and tested on many different benchmark networks. Results presented in literature suggest that an increase in entropy value corresponds to a better network performance as measured by reliability. Strong positive correlation between entropy and reliability has been demonstrated in many publications.

For more than decade, different evolutionary algorithms (EA) have been used to solve complex, multiobjective problems, mostly because they are capable of finding multiple pareto-optimal solutions in single optimization run. Nevertheless, the immense problem related to multi-objective EA is the poor ability to handle constraints. The constraints are mainly carried out by penalizing infeasible solutions, thus it could obstruct the search capabilities and may direct to suboptimal designs. Moreover, the penalty parameters are usually obtained by a trial and error making the whole process unnecessary time consuming and dependent on the parameters chosen.

The goal of this research was to develop a useful and versatile tool able to identify the cost-entropy optimal solutions in one integrated process and thereby to demonstrate the advantages of using the entropy-based method in WDN optimization. The foremost challenge solved herein was employing the entropy as an indicator for reducing the search space without having negative impact on network performance. The research investigated the problem with proper manner and developed a consistent, straightforward and robust optimization tool. The approach does not require any initial testing or time consuming calibration and no experience is needed to apply the algorithm to WDN.

Both approaches (i.e. MOC and RSS) presented in this thesis are referred as penalty-free multi-objective optimization algorithms and have been developed by incorporating few separate models. Two main tools employed for the research were borrowed from literature. The fast, elitist multi-objective genetic algorithm NSGA II was employed to solve optimization problem and coupled with hydraulic simulation software EPANET 2. Moreover, the model capable of identifying pipe flow directions and calculating entropy for any given network has been developed and coupled with above-mentioned algorithm. Both approaches have been successfully applied. The MOC model was used to optimize problems such as design, reliability and multiple loading patterns, as detailed in Chapter 4 and Chapter 5. The RSS model has been applied to optimize issues involving design, reliability and algorithm search space, as detailed in Chapter 6. This chapter concentrates on overall summary and conclusion for presented research with several suggestions for future work.

6.2 SUMMARY AND CONCLUSION

6.2.1 Multiple Operating Conditions and Maximum Entropy Based Design

The MOC ME approach has been formulated to demonstrate the advantages of using the entropy-based method to design and optimize WDNs and extend the optimization method to multiple loading patterns (i.e. minimum demand, average demand, fire flow etc.) as detailed in Chapter 3. The approach is based on three primary models: a new algorithm developed to calculate entropy for any given network, hydraulic simulation software, EPANET 2 and the fast non-dominated sorting elitist multiobjective genetic algorithm, NSGA II.

The approach involves three objectives. The first objective is to minimize network initial cost, which is dependent on pipe diameter and length. The second objective is to ensure that all nodal demands are fully satisfied. In order for the design to be feasible, the minimum pressure at critical node must be greater or equal to the desired pressure at that node. Third objective is to maximise entropy function as a measure of reliability.

In Chapter 3, the entropy based MOC approach has been applied to three well-known benchmark networks, i.e. the Two-loop network, the Six-loop network and network with three different demand patterns (Simpson et al. 2004). Four cases were studied, one with SOC and three considering MOC entropy approaches. All three MOC cases succeeded in obtaining competitive designs to the SOC method in terms of cost and entropy. Moreover, solutions obtained using MOC outperform designs obtained by SOC in terms of feasibility, pipe size distribution and reliability. Maximization of total entropy approach produced the best solutions. The significant advantage of total entropy approach over other two approaches (i.e. maximization of maximum entropy and maximization of minimum entropy) is that it takes into consideration entropy values from all operating conditions.

In Chapter 4, the maximum entropy MOC approach has been applied the real life network to illustrate its practicality and capability in integrated optimization process. Presented network has many more pipes than hypothetical layouts used earlier which makes it harder to design (i.e. more difficult to find feasible, reliable design). Maximization of total entropy value was used for MOC approach. Individual entropy values for different loading patterns were analysed to demonstrate that total entropy approach is the most suitable to employ for MOC optimization.

The MOC approach was tested on significant number of network layouts (i.e. three hypothetical and one real world) for two reasons: to identify the best entropy approach for MOC and to verify the algorithm for its consistency, accuracy and robustness. Multiple random runs were performed for each network, with the number of runs depending on network size. In case of six-loop network (Example 2 from Chapter 3) and Example 3 from Chapter 3, 30 random runs were performed for each ME approach, thus giving 90 random runs for each network (i.e. 30 runs for maximum entropy, 30 runs for minimum entropy and 30 runs for total entropy). For the real world network presented in Chapter 4, 30 random runs were performed. All those random runs generated result that are consistent and prove that the algorithm is robust and stable. It should also be highlighted that for publication purposes, the researchers usually use no more than 30 random runs for large networks, as it seem to be sufficient to verify the algorithm robustness. Nevertheless, sensitivity analysis was carried out in order to assess the robustness of created algorithm. Moreover, additional algorithm performance analysis was performed based on real world network. Presented results prove that algorithm is consistent with quick convergence rate and reliable solutions.

6.2.2 Search Space Reduction Method

Search space reduction approach has been formulated to improve convergence criteria of the GA. The method is based on entropy and uses the importance of every path through network, which is included in the entropy function. Developed algorithm is dynamic, self-adaptive and does not require any initial testing or pre-defining the diameters. Three different percentage values of ME have been integrated in the procedure ensuring much better reliability and robustness of the algorithm. The search space reductions are only enforced once the feasible design is identified, then the process of identifying reduced number of pipe sizes is repeated from generation to generation. Available pipe diameters can be reduced to any odd number.

The approach is based on three primary models: a new algorithm developed to calculate entropy for any given network, hydraulic simulation software, EPANET 2 and the fast non-dominated sorting elitist multiobjective genetic algorithm, NSGA II. The objectives considered are minimization of the network's initial cost, while ensuring network feasibility and maximization of entropy. In order to avoid penalizing infeasible solution, hydraulic feasibility has been incorporated into the approach as objective function. As such, three entities are evaluated for each proposed design: cost, entropy and network feasibility.

The algorithm has been applied to WDN previously introduced by Kadu et al (2008). The network represents substantial challenge due to large number of decision variables, thus resulting in large solution space. It is known in literature from search space reduction studies. Two cases were studied: one based on FSS and one RSS approach. RSS run include three near-ME values (i.e. three different percentage of ME value). Using three near-ME values allows identifying nearly entire range of entropy values for particular network. Detailed comparison for merged POFs based on FSS and merged POFs obtained using RSS has proved that solutions obtained using RSS dominate and clearly outperform designs based on FSS in terms of cheaper cost for similar entropy value. Another significant advantage of any RSS approach over other FSS method is that it allows narrowing the entire search space to boundary between feasible and infeasible regions. Therefore, there are more chances to find feasible design with high entropy and considerably low cost.

Search space reduction methods presented in the literature (Vairavamoorthy and Ali, 2005; Kadu et *al.* 2008) concentrate only on cost minimization (i.e. reliability optimization in not included) and are not self-adaptive methods. Therefore, the work presented in this thesis is completely novel and nothing as such exists in the literature. For that reason, the algorithm was verified for its consistency and robustness. Firstly, network previously used for search space reduction (Kadu et *al.* 2008) has been implemented in this study. Great number of 50 random runs was performed for each approach (i.e. one FSS and three different ME approaches), thus giving 200 random runs in total. Results for all approaches from all random runs were presented herein in Appendix D. Those results are consistent and demonstrate the algorithm's robustness. Moreover, the consistency and performance of presented algorithm has also been investigated for each of the objective functions. Feasible and infeasible solutions were used for the cost examination. Rapid stabilization was observed for the results obtained using RSS in comparison to FSS approach. In terms of entropy, only the feasible solutions were considered. Once again, the entropy value seems to be boosted quickly within the first few generations. Such rapid stabilization is the result of reducing the number of decision variables. It is evident that in case of RSS approach much smaller number of function evaluations is required for the GA to converge.

6.2.3 Cost-Entropy Non-Dominated Sorting Module and Genetic Algorithm Performance Module

During the work on research it has been found that additional software is required for analysis purposes. Two external modules have been developed. The aim of first one is to identify the cost-entropy non-dominated solutions over the entire history of results in order to provide good range of feasible candidate solutions. Second module was developed for reassurance of GA consistency and performance as the algorithm changes that were performed could reflect in making the GA less efficient and stable.

The external screening method was used in all research chapters. Comparison of POF originally generated by the algorithm and the solutions obtained from entire history was performed and results presented. Overall, it was demonstrated that the screening procedure is extremely efficient and robust with the capability of providing a good improved front with wide range and nice spread of results.

The GA performance module was employed in Chapter 4 and Chapter 5. Presented results ensure that the GA is consistent, converges quickly enough and performs well on all runs despite the initial data (i.e. the randomly generated initial population).

6.3 SUGGESTIONS FOR FUTURE WORKS

The research presented herein has addressed a broad scope and challenging issues related to WDN optimization. Nevertheless, there is more room for additional problems that have been revealed while solving issues presented in this thesis. Moreover, some ideas, that could possibly be carried out to improve and complete present research even further, has arisen.

Implementation of MOC in penalty-free reliability based multi-objective approach for WDNs has been an immense step as most of the work has been done for SOC, which assumes that nodal demands are constant. Nevertheless the networks used for this phase of research were fairly simple layouts with only pipes and reservoirs included. Therefore, good suggestion would be to extend the existing MOC approach to include other network components, such as valves, storage tanks and pumps and to cover longer period of time using extended period simulation. Additionally, water quality could be incorporated within presented study. Hydraulic model EPANET 2, employed in developed approach, has capability to model water age, thus predict chlorine residuals within water distribution systems. Therefore, it would be beneficial and straightforward to implement water quality within present study.

A very good suggestion would be to employ more realistic hydraulic analysis model. The present study is based on demand driven analysis (DDA). DDA assumes that all nodal demands are satisfied at all times so it works accurately for a network performing under normal operating conditions. In case of any pressure deficient scenarios, the DDA may leads to unrealistic nodal pressures. To avoid such misleading situations the head dependent analysis (HDA) could be employed. The HDA considers the relationship between pressure and outflow. Thus, replacing the DDA model with HDA model incorporated in Epanet-PDX (Siew and Tanyimboh, 2011) is worth considering.

A good suggestion would be to investigate new method for handling redundant codes in binary coded genetic algorithm. Performance of presented approach should be compared with other remapping and bias free methods proposed in literature.
Successfully used novel method for identifying uniformly spread POF with feasible designs strongly suggest that the technique should be employed inside the algorithm (i.e. as a part of GA single run) in order to avoid running any additional programmes. There are few different techniques to incorporate such subroutine. The easiest way would be to implement the method after the last generation but before closing the program. Nevertheless, it would not be much different from existing method. Therefore, it would not interfere with the GA process and results. More complicated with possible impact on GA process would be to identify feasible non-dominated solutions after specified number of generations. Then, with the use of elitism, the best results would have to be kept for another generations with the whole process repeated until the GA stopping criteria.

Another alternative, that could possibly lead to required results (i.e. feasible, cost-entropy non-dominated), would be to prioritise objective functions like cost and entropy while taking into account feasible solutions only. However, such strategy should be implemented carefully and wisely as penalizing infeasible solutions could obstruct the search capabilities and may lead to suboptimal solutions.

Implementation of ME based RSS method had enormous effect in designing WDN. The success of RSS approach applied to SOC strongly suggests extending the method to cover MOC. The next step would be to widen the method to cover longer period of time using extended period simulation and to include other network components, such as pumps, valves and storage tanks. Additionally, achieved approach would need to be tested for calibrating on real world distribution networks.

It would also be interesting to validate the developed algorithm by comparing the results with solutions obtained from other search space reduction approaches. Nevertheless, to do so, the additional approaches involved would also have to have the reliability performance indicator included. So far, it is impossible to make such comparison, as alternative solutions where cost and network reliability are optimized are not available elsewhere.

RSS approach used in this study has been tested on reduced 5 pipe diameters; however any number of pipe sizes could be used. It would be interesting to analyse results achieved if different number of pipe sizes would be employed. It is thought that the smaller number of reduced pipe sizes the smaller number of function evaluations required for GA to converge, but harder to achieve near global maximum entropy value.

A good suggestion would be to increase number of reduced pipe sizes (i.e. from 5 to 7, from 7 to 9 etc.) when the highest achieved entropy reach stagnation point. It could prevent the GA of being trapped in local optimum. Moreover, it would allow the algorithm to widen explore regions of search spaces with good potential solutions. Even if the GA would finally reach the highest number of available pipe diameters (i.e. full search space), the optimal design should still be identified in lower number of function evaluations than it would happen if FSS would be applied for entire GA run.

REFERENCES:

Alperovits, E., and Shamir, U. (1977). Design of optimal water distribution systems. *Water Resources Research*, 13(6), 885-900.

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

Awumah, K., Goulter, I., and Bhatt, S.K. (1991). Entropy-based redundancy measures in water distribution network design. *Journal of Hydraulic Engineering*, 117(5), 595-614.

Back, T., Fogel, D.B., and Michalewicz, Z. (2000). *Evolutionary computation 1: Basic algorithms and operators*. Vol. 1. CRC Press.

Barlow, E. and Tanyimboh, T. T. (2014) Multi-objective memetic algorithm applied to the optimization of water distribution systems. *Water Resources Management*, 28(8), 2229-2242.

Besley, D., Bull, D.R., and Martin, R.R. (1993). An overview of Genetic Algorithms: Part 2, Research Topics, University Computing, 15(4), 170-181.

Bhave, P.R. (1978). Noncomputer optimization of single-source network. *Journal of Environmental Engineering*, 104(4), 799-814.

Bhave, P.R. (1991) *Analysis of Flow in Water Distribution Networks*. Lancaster, PA: Technomic Publishing.

Caruana, R.A., and Schaffer, J.D. (1988). Representation and hidden bias: Gray vs. binary coding for genetic algorithms. *In Proceedings of the 5th International Conference on Machine Learning*, 153-161, Los Altos: CA, Morgan Kaufmann.

Chandapillai, J. (1991). Realistic simulation of water distribution system. *Journal of Transportation Engineering*, 117(2), 258-263.

Creaco, E., Franchini, E., and Alvisi, S. (2010). Optimal placement of isolation valves in water distribution systems based of valve cost and weighted average demand shortfall. *Water Resources Management*, 24(15), 4317-4338.

Creaco, E., Franchini, E., and Alvisi, S. (2012). Evaluating water demand shortfalls in segment analysis. *Water Resources Management*, 26(8), 2301-2321.

Cross, H. (1936). Analysis of flow in networks of conduits or conductors. Bulletin No. 286, University of Illinois Engineering, Experimental Station, Urbana, 111.

Czajkowska A.M., and Tanyimboh T.T. (2012a). Maximum entropy design of water distribution systems under multiple operating conditions, 10th International Conference on Hydroinformatics, Hamburg, Germany, 14-18 July, 2012**.**

Czajkowska, A.M., and Tanyimboh, T.T., (2012b). Reliability assessment of water distribution networks using surrogate measures, *14th Water Distribution Systems Analysis Conference*, Adelaide, South Australia, 24-27 September 2012.

Czajkowska, A.M., and Tanyimboh, T.T. (2013). Water distribution network optimization using maximum entropy under multiple loading patterns. *Water Science & Technology: Water Supply*, 13(5), 1265-1271.

Dandy, G.C., and Engelhard, M.O. (2001). Optimal scheduling of water pipe replacement using genetic algorithm. *Journal of Water Resources Planning and Management*, 127(4), 214-223.

Dandy, G.C., Simpson, A.R., and Murphy, L.J. (1996). An improved genetic algorithm for pipe network optimization. Water Resources Research, 32(2), 449-458.

Deb, K. (2000). An efficient constraint handling method for genetic algorithms. *Computer Methods in applied Mechanics and Engineering*, 186(2), 311-338.

Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms* (Vol. 16). John Wiley & Sons.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197.

DeNeufville, R., Schaake, J., and Stafford, J. (1971). Systems analysis of water distribution networks. *Journal Sanitary Engineering Division ASCE*, 97(SA6), 825–842.

Dorigo, M., Caro, G.D., and Gambardella, L.M. (1999). Ant algorithms for discrete optimization. Artificial Life, 5(2), 137-172.

Eshleman, L.J., Caruana, A., and Schaffer, J.D. (1989). Biases in the crossover landscapes. *Proceeding of the Third International Conference on Genetic Algorithm*, 86-91, San Mateo, Morgan Kaufmann.

Eshleman, L.J., and Schaffer, J.D. (1993). Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms 2*, San Francisco, Morgan Kaufmann Publishers, Vol.2, 187-202.

Eusuff, M.M., and Lansey, K.E. (2003). Optimization of water distribution network design using the shuffled frog leaping algorithm. *Journal of Water Resources Planning and Management*, 129(3), 210-225.

Fonseca, C.M., and Fleming, P.J. (1993). Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In *Proceedings of the International Conference on Genetic Algorithms*, vol. 5, 416–423, Morgan Kaufmann, Los Altos, CA.

Farmani, R., Savic, D.A., and Walters, G.A. (2003). Multi-objective optimization of water system: a comparative study. *Cabrera et al. (eds.) Pumps, Electromechanical Devices and Systems Applied to Urban Water Management*, vol.1, 247-256.

Farmani, R., Wright, J.A., Savic, D.A., and Walters, G.A. (2005) Self-adaptive fitness formulation for evolutionary constrained optimization of water systems. *Journal of Computing in Civil Engineering*, 19(2), 212-216.

Farmani, R., Walters, G., and Savic, D. (2006). Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. reliability vs. water quality. *Journal of Hydroinformatics*, 8(3), 165-179.

Fujiwara, O., and Ganesharajah, T. (1993). Reliability assessment of water supply systems with storage and distribution networks. *Water Resources Research*, 29(8), 2917-2924.

Geem, J.Z., Kim, W., and Loganathan, G.V. (2001). A new heuristic optimization algorithm: harmony search. *Simulation*, 76(2), 60–68.

Gen, M., Cheng, R., and Lin, L. (2008). *Network models and optimization: Multiobjective genetic algorithm approach.* Springer Science & Business Media.

Germanopoulos, G. (1985). A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models. *Civil Engineering Systems*, 2(3), 171-179.

Glover, F., and Greenberg, H. (1989). New approaches for heuristic search: a bilateral linkage with artificial intelligence. *European Journal of Operational Research*, 39, 19-130.

Goldberg, D.E. (1989). Genetic algorithms in search, optimization and machine learning. 1st edition. USA: Addison-Wesley Longman Publishing Co.

Goldberg, D.E. (1991). Real coded genetic algorithms, virtual alphabets and blocking. *Complex Systems*, 5, 139-167.

Goldberg, D.E., and Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. *Foundations of Genetic Algorithms*, 1, 69-93.

Goldberg, D.E., and Kuo, C.H. (1987). Genetic algorithms in pipeline optimization. *Journal of Computing in Civil Engineering*, ASCE, 128-141.

Gupta, R., and Bhave, P. R. (1996). Comparison of methods for predicting deficient-network performance. *Journal of Water Resources Planning and Management*, 122(3), 214-217.

Hajela, P., and Lin, C.Y., (1992). Genetic search strategies in multi-criterion optimal design. *Structural Optimization*, 4(2), 99-107.

Haghighi, A., Samani, H. M., and Samani, Z. M. (2011). GA-ILP method for optimization of water distribution networks. *Water Resources Management*, 25(7), 1791-1808.

Herrera, F., Lozano, M., and Verdegay, J.L. (1998). Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, 12(4), 265-319.

Holland, J.H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* Oxford, England: U Michigan Press.

Janikow, C.Z., and Michalewicz, Z. (1991). An experimental comparison of binary and floating point representation in genetic algorithms. *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Mateo, Morgan Kaufmann Publishers, 31-36.

Jayaram, N., and Srinivasan, K. (2008). Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing. *Water Resources Research*, 44(1).

Jaynes, E.T. (1957). Information theory and statistical mechanics. *Physical Review*. 106, 620-630 and 108. 171-190.

Jeong, H.S., and Abraham, D.M. (2006). Operational response model for physically attacked water networks using NSGA-II. *Journal of Computing in Civil Engineering*, 20(5), 328-338.

Kadu, M.S., Gupta. R., and Bhave. P.R. (2008). Optimal design of water networks using a modified genetic algorithm with reduction in search space. *Journal of Water Resources Planning and Management*, 134(2), 147-160

Kennedy, J., and Eberhart, R.C. (1995). Particle swarm optimization. *Proceedings of IEEE International Conference on Neural Networks*, Piscataway, 1942-1948.

Khu, S.T., and Keedwell, E. (2005) Introducing more choices (flexibility) in the upgrading of water distribution networks: the New York City tunnel network example. *Engineering optimization*, 37(3), 291-305.

Khinchin, A.I., (1953). The entropy concept in probability theory. *Uspekhi Mathematicheskikh Nauk.* 8(3), 3-20. Translation in A.I. Khinchin (1975). *Mathematical Foundations of Information Theory*, Dover, New York, 1-28.

Knowles, J.D., and Corne, D.W. (2000). Approximating the non-dominated front using the Pareto archived evolution strategy. *Evolutionary computation*, 8(2), 149-172.

Kursave, F. (1990). A variant of evolution strategies for vector optimization. *Parallel Problem Solving from Nature 1*, 183-197.

Mahdavi, M., Fesanghary, M., and Damangir, E. (2007). An improved harmony search algorithm for solving optimization problem. *Applied Mathematics and Computation*, 188(2), 1567-1579.

Maier, H.R., Simpson, A.R., Zecchin, A.C., Foong, W.K., Phang, K.Y., Seah, H.Y., and Tan, C.L. (2003). Ant colony optimization for design of water distribution systems. Journal of Water Resources Planning and Management, 129(3), 200-209.

Martin, D.W., and Peters, G. (1963). The application of Newton's method to network analysis by digital computer. *Journal of Institute of Water Engineers*, 17, 115-129.

McCormick, W.T., Scheitzer, P.J., and White, T.W. (1972). Problem decomposition and data reorganization by cluster technique. *Operations Research*, 20(5), 993-1009.

Michalewicz, Z. (1992). Genetic algorithms + data structures = evolution programs. Springer-Verlag, New York.

Michalewicz, Z. (1995). A survey of constraint handling techniques in evolutionary computation methods. *In:* McDonnell *et al.* eds. *Evolutionary Programming IV*, Massachusetts: MIT Press, 135-158.

Montalvo, I., Izquierdo, J., Perez, R., and Tung, M.M. (2008). Particle swarm optimization applied to the design of water supply systems. *Computers and Mathematics with Applications*, 56(3), 769-776.

Munavalli, G.R., and Kumar, M.S. (2003). Optimal scheduling of multiple chlorine sources in water distribution systems. *Journal of Water Resources Planning and Management*, ASCE, 129(6), 493-504.

Osyczka, A., and Kundu, S. (1995). A new method to solve generalized multicriteria optimization problems using the simple genetic algorithms. *Structural Optimization*, 10(2), 94-99.

Prasad, T.D. (2010). Design of pumped water distribution networks with storage. *Journal of Water Resources Planning and Management*, 136(1), 129-132.

Prasad, T.D., and Park, N.S., (2004). Multiobjective genetic algorithms for design of water distribution networks. *Journal of Water Resources Planning and Management*, 130(1), 73-82.

Prasad, T.D., and Tanyimboh, T.T., (2008). Entropy based design of "Anytown" water distribution network. *Proceedings of the 10th Annual Water Distribution Systems Analysis Conference*, Kruger National Park, South Africa, August 17-20.

Preis, A., and Ostfeld, A. (2008). Multiobjective contaminant sensor network design for water distribution systems. *Journal of Water Resources Planning and Management*, 134(4), 366-377.

Radcliffe, N.J., (1992). Non-linear genetic representations. *Parallel Problem Solving from Nature 2*, Elsevier Science Publisher, Amsterdam, 259-268.

Rossman, L.A. (2000). EPANET 2. User's Manual. Water Supply and Water Resources Division, National Risk Management Research Laboratory, US EPA, Cincinnati, OH, USA.

Rowell, W.F., and Barnes, J. (1982). Obtaining layout of water distribution systems. *Journal of Hydraulics Division*, 108(1), 137-148.

Saleh, S.H.A., and Tanyimboh T.T. 2011. Global maximum entropy minimum cost design of water distribution systems. *13th International Water Distribution System Analysis Symposium*, 22-26 May 2011, Palm Springs, California. USA: ASCE, 206-213.

Saleh, S.H., and Tanyimboh, T.T. (2013). Coupled topology and pipe size optimization of water distribution systems. *Water Resources Management*, 27(14), 4795-4814.

Saleh, S.H., and Tanyimboh, T.T. (2014). Optimal Design of Water Distribution Systems Based on Entropy and Topology. *Water Resources Management*, 28(11), 3555-3575.

Savic, D.A., and Walters, G.A. (1997). Genetic algorithms for least cost design of water distribution networks. *Journal of Water Resources Planning and Management*, 123(2), 66-67.

Schaffer, J. D. (1985) Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st International Conference on Genetic Algorithms, Pittsburgh, PA, USA*, 93-100.

Setiadi, Y., Tanyimboh, T.T., and Templeman, A.B. (2005). Modelling errors, entropy and the hydraulic reliability of water distribution systems. *Advances in Engineering Software*, *36*(11), 780-788.

Siew, C., and Tanyimboh, T.T. (2011). Design of the "Anytown" network using the penalty-free multi-objective evolutionary optimization approach. In Proceedings of the 13th Annual Water Distribution Systems Analysis Conference, WDSA, 22-26.

Siew, C., and Tanyimboh, T.T. (2012). Pressure-dependent EPANET extension. Water Resources Management 26(6), 1477–1498.

Siew, C., Tanyimboh, T.T., and Seyoum, A.G. (2014). Assessment of penalty-free multi-objective evolutionary optimization approach for the design and rehabilitation of water distribution systems. *Water Resources Management,* 28(2), 373-389.

Simpson, A.R., Dandy, G.C., and Murphy, L.J. (1994). Genetic algorithms compared to other techniques for pipe optimization. *Journal of Water Resources Planning and Management*, 120(4), 423-443.

Shannon, C., (1948). A mathematical theory of communication. *Bell Systems Technical Journal.* 27(3), 379-428.

Srinivas, N., and Deb, K. (1994). Multiobjective optimization using non-dominated sorting in genetic algorithms. *Evolutionary computation*, 2(3), 221-248.

Storn, R., and Price, K. (1997). Differential Evolution - a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, *Kluwer Academic Publishers*, Vol. 11, 341 - 359.

Surendran, S., Tanyimboh, T.T., and Tabesh, M. (2005). Peaking demand factor-based reliability analysis of water distribution system. *Advances in Engineering Software*, 36(11-12), 789-796.

Tanyimboh, T.T. (1993). *An entropy based approach to the optimum design of reliable water distribution networks*. Thesis (PhD), University of Liverpool.

Tanyimboh, T.T., and Kalungi, P. (2008). Optimal long term design, rehabilitation and upgrading of water distribution systems. *Water Science and Technology*, IWA, 39(4), 249-255.

Tanyimboh, T.T., and Kalungi, P. (2009). Multi-criteria assessment of optimal design, rehabilitation and upgrading schemes for water distribution networks. *Civil Engineering and Environmental Systems*, 26(2), 117–140.

Tanyimboh, T.T., and Setiadi, Y. (2007). Maximum-entropy design of water distribution networks using discrete pipe diameters. *Civil-Comp Press*, Paper 238.

Tanyimboh, T.T., and Setiadi, Y. (2008a). Joint layout, pipe size and hydraulic reliability optimization of water distribution systems. *Engineering Optimization*, 40(8), 729-747.

Tanyimboh, T.T., and Setiadi, Y. (2008b). Sensitivity analysis of entropy constrained designs of water distribution systems. *Engineering Optimization*, 40(5), 439-457.

Tanyimboh, T.T., and Sheahan, C. (2002). A maximum entropy based approach to the layout optimization of water distribution systems. *Civil Engineering and Environmental Systems*, 19(3), 223-253.

Tanyimboh, T.T., Tabesh, M., and Burrows, R. (2001). An appraisal of source head methods for calculating the reliability of water distribution networks. *Journal of Water Resources Planning and Management*, 127 (4), 206-213.

Tanyimboh, T.T., and Templeman, A.B. (1993a). Maximum entropy flows for single-source networks. *Engineering Optimization*, 22(1), 49-63.

Tanyimboh, T.T., and Templeman, A.B. (1993b). Calculating maximum entropy flows in networks. *Journal of the Operational Research Society*, 44(4), 383-396.

Tanyimboh, T.T., and Templeman, A.B. (1993c). Optimum design of flexible water distribution networks. *Civil Engineering Systems*, 10(3), 243-258

Tanyimboh, T.T., and Templeman, A.B. (1995). A new method for calculating the reliability of single-source networks. *Developments in Computational Techniques for Civil Engineering* (Topping, B.H.V. (Ed.)). Civil-Comp Press.

Tanyimboh, T.T., and Templeman, A.B. (1998). Calculating the reliability of single source networks by source head method. *Advances in Engineering Software*, 29(7-9), 449-505.

Tanyimboh, T.T. and Templeman, A.B. (2000). A quantified assessment of the relationship between the reliability and entropy of water distribution systems. *Engineering Optimization*, 33(2), 179-199.

Tanyimboh, T.T., and Templeman, A.B. (2004). A new nodal outflow function for water distribution networks. *Proceedings of the 4th International Conference on Engineering Computational Technology, BHV Topping and CA Mota Soares (Eds.),* Civil-Comp Press, Stirling, UK, Paper 64, page 12.

Tanyimboh, T.T., and Templeman, A.B. (2010). Seamless pressure-deficient water distribution system model, *Journal of Water Management, ICE*, 163(8), 389-396.

Tanyimboh T.T., Tietavainen, M.T., and Saleh, S.A.E. (2011). Reliability assessment of water distribution systems with statistical entropy and other surrogate measures. *Water Science and Technology – Water Supply*, 11(4), 437-443.

Todini, E. (2000). Looped water distribution networks using resilience index based on heuristic approach. *Urban Water*, 2(3), 115-122.

Todini, E., and Pilati, S. (1988). A gradient algorithm for the analysis of pipe networks. *Computer applications in water supply: Systems analysis and simulation*, vol. 1 (Coulbeck, B., and Chun-Hou, O. (Eds.)). Research Studies Press, Taunton, UK, 1–20.

Vaabel, J., Ainola, L., and Koppel, T. (2006). Hydraulic power analysis for determination of characteristics of a water distribution system. *8th Annual Water Distribution Systems Analysis Symposium*, ASCE, Reston, VA.

Vairavamoorthy, K., and Ali, M., (2005). Pipe index vector; A method to improve genetic algorithm based pipe optimization. *Journal of Hydraulic Engineering*, 131(12), 1117-1125.

Vamvakeridou-Lyroudia, L.S., Walters, G.A., Savic, D.A. (2005). Fuzzy multiobjective optimization of water distribution networks. *Journal of Water Resources Planning and Management*, 131(6), 467-476.

Vasan, A., and Simonovic, S.P. (2010) Optimization of water distribution network design using differential evolution. *Journal of Water Resources Planning and Management*, 136(2), 279-287.

Wagner, J.M., Shamir, U., and Marks, D.H. (1988). Water distribution reliability: simulation methods. *Journal of Water Resources Planning and Management*, 114(3), 276-294.

Walski T. M., Brill, E.D., Gessler, J., Goulter, I.C., Jeppson, R.M., Lansey, K., Lee, H.-L., Liebman, J.C., Mays, L., Morgan, D.R., and Ormsbee, L. (1987). Battle of the network models: Epilogue. *Journal of Water Resources Planning and Management*, 113(2), 191-203.

Weickgenannt, M., Kapelan, Z., Blokker, M., and Savic, D.A. (2010) Risk-based sensor placement for contaminant detection in water distribution systems. *Journal of Water Resources Planning and Management*, 136(6), 629-636.

Wood, D.J., and Charles, C.O. (1972). Hydraulic network analysis using linear theory. *Journal of the Hydraulics Division*, 98(7), 1157-1170.

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

Wu, Z.Y., and Simpson, A.R. (2002) A self-adaptive boundary search genetic algorithm and its application to water distribution systems. *Journal of Hydraulic Research*, 40(2), 191-203.

Wu, Z.Y., and Walski, T. (2005) Self-adaptive penalty approach compared with other constraint-handling techniques for pipeline optimization. *Journal of Water Resources Planning and Management*, 131(3), 181-192.

Yassin-Kassab, A., Templeman, A.B., and Tanyimboh, T.T. (1999). Calculating maximum entropy flows in multi-source, multi-demand networks. *Engineering Optimization*, 31(6), 695-729.

Ziztler, E., and Thiele, L. (1998). An evolutionary algorithm for multiobjective optimization. The strength Pareto approach. *Technical Report 43, Zurich, Switzerland: Computer Engineering and Networks Laboratory (TIK)*, Swiss Federal Institute of Technology (ETH).

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2), 173-195.

APPENDIX A

INPUT DATA AND ADDITIONAL RESULTS FOR NETWORK USED FOR SENSITIVITY ANALYSIS PRESENTED IN CHAPTER THREE

Table A-1. Node data for network used for sensitivity analysis

Node No.	Elevation (m)	Demand (l/s)	H_{min} (m)
2	150	27.78	30
3	160	27.78	30
4	155	33.33	30
5	150	75.00	30
6	165	91.66	30
7	160	55.56	30

Table A-2. Reservoir data for network used for sensitivity analysis

Reservoir No.	Total Head (m)
1	210.00

Table A-3. Available pipe diameters and unit costs for network used for sensitivity analysis

Diameter (mm)	Cost (units)
25.4	2
50.8	5

76.2	8
101.6	11
152.4	16
203.2	23
254.0	32
304.8	50
355.6	60
406.4	90
457.2	130
508.0	170
558.8	300
609.6	500

Table: A-4. CPU time for different populations

Population Size	Average CPU time [s]
100	289
200	604
300	932
400	1252
500	1460

Table A-5. Number of CEND solutions depending on mutation rate

Mutation Rate	Number of CEND designs (for all 10 random runs)
0.0010	22
0.0050	25
0.0100	27
0.0200	28
0.03125	35

APPENDIX B

INPUT DATA FOR EXAMPLES IN CHAPTER THREE

Table B-1: Node data for network in Example 1

Node No.	Elevation (m)	Demand Pattern 1		Demand Pattern 2		Demand Pattern 3	
		Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)
2	150	27.78	30	22.22	30	13.33	30
3	160	27.78	30	22.22	30	13.33	30
4	155	33.33	30	26.66	30	16.00	30
5	150	75.00	30	60.00	30	36.00	30
6	165	91.66	30	73.33	30	44.00	30
7	160	55.56	30	44.45	30	26.67	30

Table B-2. Node data for network in Example 2

Node No.	Demand Pattern 1		Demand Pattern 2		Demand Pattern 3	
	Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)
2	27.8	30	22.24	30	13.34	30
3	41.7	30	33.36	30	20.02	30
4	41.7	30	33.36	30	20.02	30
5	41.7	30	33.36	30	20.02	30
6	27.8	30	22.24	30	13.34	30
7	55.5	30	44.40	30	26.64	30
8	55.5	30	44.40	30	26.64	30
9	55.5	30	44.40	30	26.64	30
10	27.8	30	22.24	30	13.34	30
11	41.7	30	33.26	30	20.02	30
12	27.8	30	22.24	30	13.34	30

Table B-3. Reservoir data for network in Example 2

Reservoir No.	Total Head (m)
1	100.00

Table B-4. Node data for network in Example 3

Node No.	Elevation (m)	Demand Pattern 1		Demand Pattern 2		Demand Pattern 3	
		Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)
2	320.04	12.62	28.18	12.62	14.09	12.62	14.09
3	326.14	12.62	17.61	12.62	14.09	12.62	14.09
4	332.23	0	17.61	0	14.09	0	14.09
6	298.70	18.93	35.22	18.93	14.09	18.93	14.09
7	295.66	18.93	35.22	82.03	10.57	18.93	14.09
8	292.61	18.93	35.22	18.93	14.09	18.93	14.09
9	289.56	12.62	35.22	12.62	14.09	12.62	14.09
10	289.56	18.93	35.22	18.93	14.09	18.93	14.09
11	292.61	18.93	35.22	18.93	14.09	18.93	14.09
12	289.56	12.62	35.22	12.62	14.09	50.48	10.57

Table B-5. Reservoir data for network in Example 3

Reservoir No.	Total Head (m)
1	365.76
5	371.86

Table B-6. Pipe data for network in Example 3

Pipe No.	Start Node	End Node	Length (m)
1	1	2	4828
2	2	3	1609
3	3	4	1609
4	4	5	6437

5	2	6	1609
6	4	8	1609
7	6	7	1609
8	7	8	1609
9	6	9	1609
10	7	10	1609
11	8	11	1609
12	9	10	1609
13	10	11	1609
14	11	12	1609

Table B-7. Available pipe diameters and unit costs for network in Example 3

Diameter (mm)	Pipe Cost ($/m)
152	49.5
203	63.3
254	94.8
305	132.9
356	170.9
406	194.9
457	231.3
508	262.5

APPENDIX C

INPUT DATA FOR EXAMPLE IN CHAPTER FOUR

Table C-1: Nodal demand for network in chapter four

Node No.	Elevation (m)	Demand Pattern 1		Demand Pattern 2		Demand Pattern 3	
		Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)	Demand (l/s)	H_{min} (m)
2	0	5	28	5	14	5	14
3	0	3	28	3	14	3	14
4	0	6.5	28	6.5	14	6.5	14
5	0	10.5	28	10.5	14	10.5	14
6	0	9.5	28	9.5	14	9.5	14
7	0	6.5	28	6.5	14	6.5	14
8	0	8.5	28	8.5	14	8.5	14
9	0	9.5	28	9.5	14	9.5	14
10	0	10	28	10	14	10	14
11	0	4.5	28	4.5	14	18	8.4
12	0	5.5	28	5.5	14	5.5	14
13	0	3.5	28	3.5	14	3.5	14
14	0	4	28	4	14	4	14
15	0	5	28	5	14	5	14
16	0	4	28	4	14	4	14
17	0	8	28	8	14	8	14
18	0	6.5	28	6.5	14	6.5	14
19	0	7.5	28	7.5	14	7.5	14
20	0	6	28	6	14	6	14
21	0	8	28	8	14	8	14
22	0	8.5	28	8.5	14	8.5	14
23	0	2.5	28	2.5	14	2.5	14
24	0	4.5	28	4.5	14	4.5	14
25	0	4.5	28	4.5	14	4.5	14

26	0	7.5	28	7.5	14	7.5	14
27	0	6	28	6	14	6	14
28	0	6	28	6	14	6	14
29	0	8	28	8	14	8	14
30	0	7.125	28	7.125	14	7.125	14
31	0	9	28	39	8.4	9	14
32	0	8.625	28	8.625	14	8.625	14
33	0	5	28	5	14	5	14
34	0	11.5	28	11.5	14	11.5	14

Table C-2. Pipe data for network in chapter four

Pipe No.	Length (m)	Pipe No.	Length (m)
1	1	39	239.08
2	1	40	404.28
3	260	41	367.51
4	600	42	391.18
5	1200	43	229.78
6	600	44	404.76
7	800	45	215.2
8	200	46	575.37
9	500	47	210.89
10	2100	48	257.55
11	500	49	277.95
12	1000	50	48.1
13	304.16	51	309.11
14	341.16	52	150.25
15	210.5	53	174.42
16	422.28	54	396.55
17	529.66	55	169.47
18	254	56	178.62
19	252.9	57	339.77
20	230.65	58	98.51
21	159.46	59	458.41
22	456.91	60	386.4

23	313.99	61	135.01
24	73.8	62	559.91
25	192.73	63	319.1
26	176.53	64	193.48
27	124.94	65	393.04
28	478.71	66	60
29	282.03	67	243.45
30	262.9	68	646.8
31	202.43	69	236.22
32	52.91	70	422.08
33	154.54	71	244.93
34	352.16	72	5
35	165.08	73	31.84
36	325.05	74	258.71
37	377.94	75	154.7
38	275.39	76	229.87

Table C-3. Available pipe diameters and unit costs for network in chapter four

Diameter (mm)	Pipe Cost (€/m)
150	271.94
200	299.43
250	328.01
300	359.54
350	399.03
400	438.63
450	461.34
500	502.78

APPENDIX D

INPUT DATA AND ADDITIONAL RESULTS FOR NETWORK PRESENTED IN CHAPTER FIVE

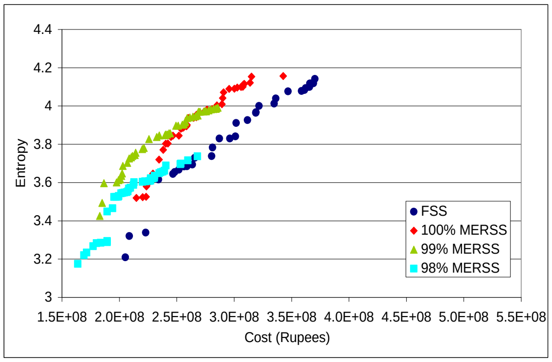
Table D-1. Pipe and node details for example from chapter five

Pipe details		Node details		
Pipe No.	Length (m)	Node No.	Minimum HGL (m)	Demand (m3/min)
1	300	1	100	-
2	820	2	95	-
3	940	3	85	18.4
4	730	4	85	4.5
5	1 620	5	85	6.5
6	600	6	85	4.2
7	800	7	82	3.1
8	1 400	8	82	6.2
9	1 175	9	85	8.5
10	750	10	85	11.5
11	210	11	85	8.2
12	700	12	85	13.6
13	310	13	82	14.8
14	500	14	82	10.6
15	1 960	15	85	10.5
16	900	16	82	9.0
17	850	17	82	6.8
18	650	18	85	3.4
19	760	19	82	4.6
20	1 100	20	82	10.6
21	660	21	82	12.6
22	1 170	22	80	5.4
23	980	23	82	2.0

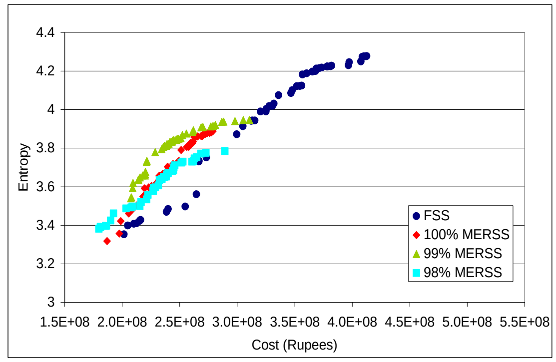
24	670	24	80	4.5
25	1 080	25	80	3.5
26	750	26	80	2.2
27	900			
28	650			
29	1 540			
30	730			
31	1 170			
32	1 650			
33	1 320			
34	3 250			

Table D-2. Available pipe diameters and unit costs for example from chapter five

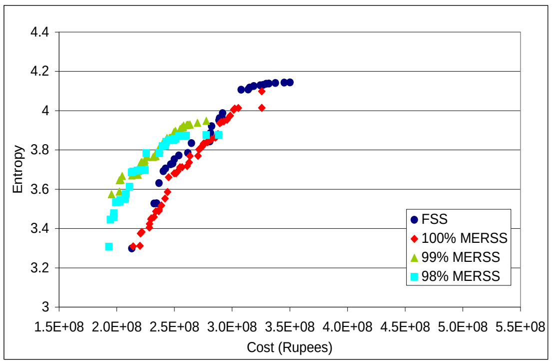
Diameter (mm)	Unit cost (rupees)
150	1 115
200	1 600
250	2 154
300	2 780
350	3 475
400	4 255
450	5 172
500	6 092
600	8 189
700	10 670
750	11 874
800	13 261
900	16 151
1 000	19 395



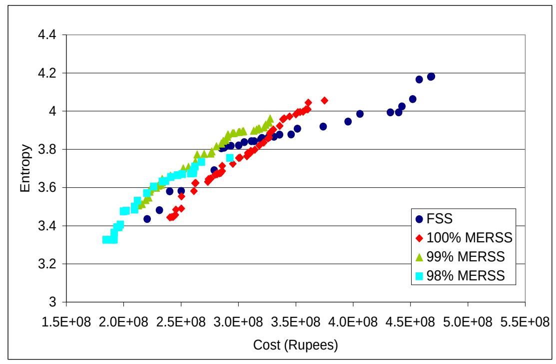
Run 1



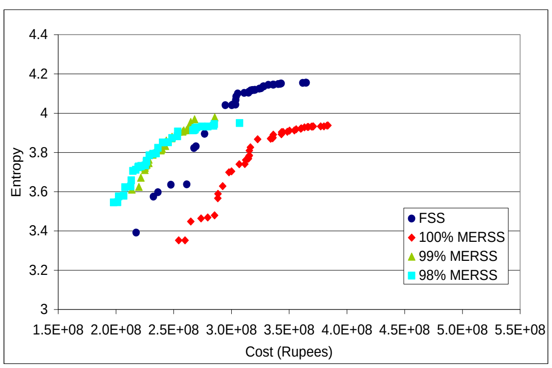
Run 2



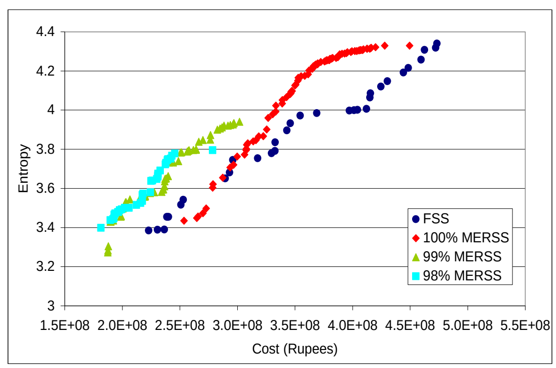
Run 3



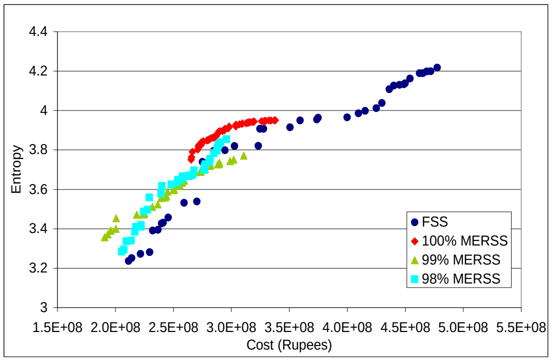
Run 4



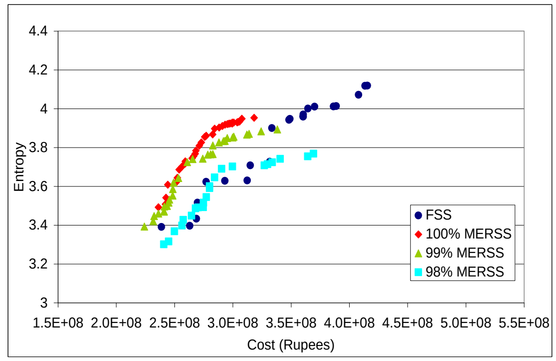
Run 5



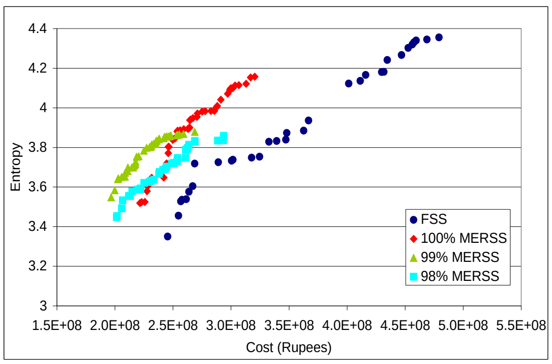
Run 6



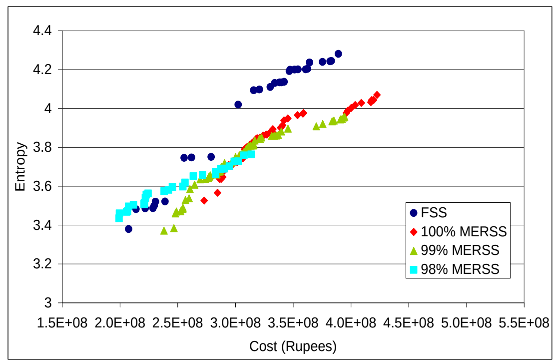
Run 7



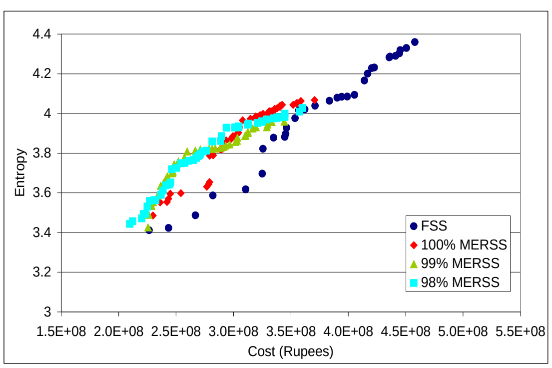
Run 8



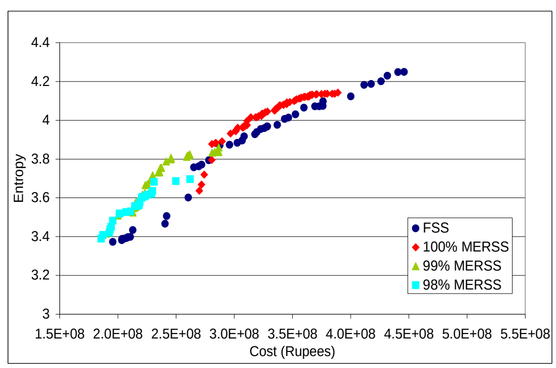
Run 9



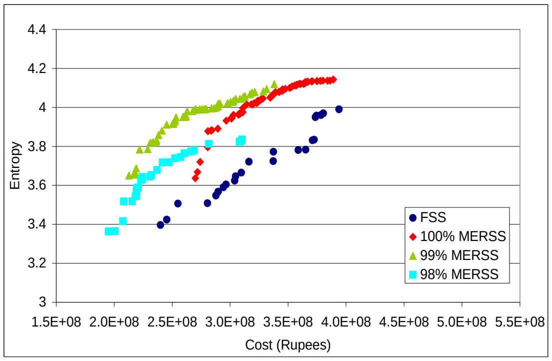
Run 10



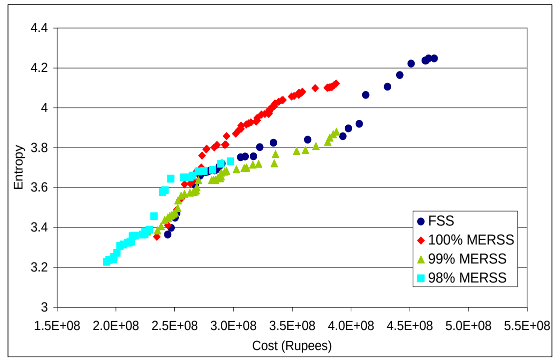
Run 11



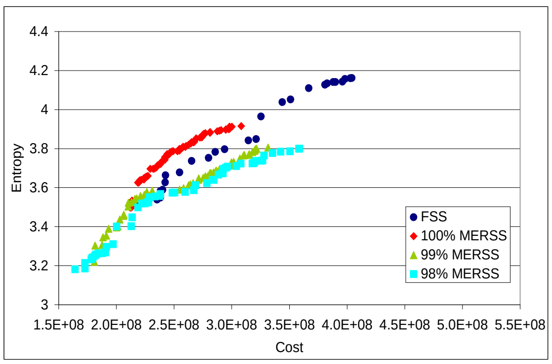
Run 12



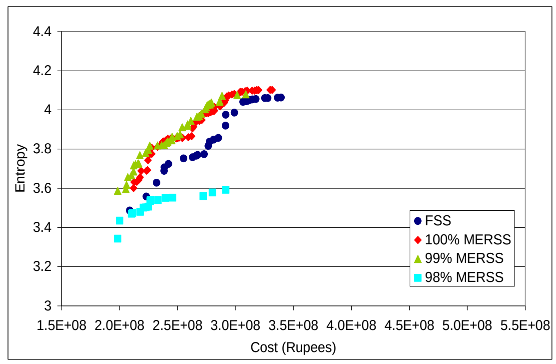
Run 13



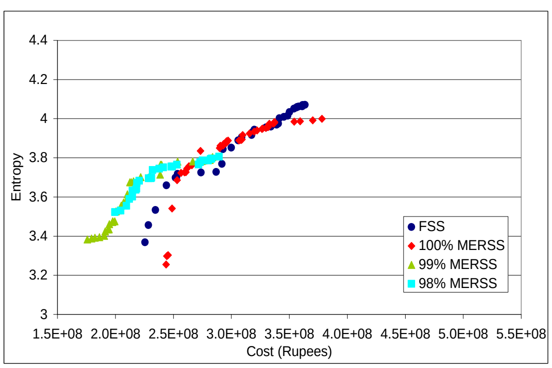
Run 14



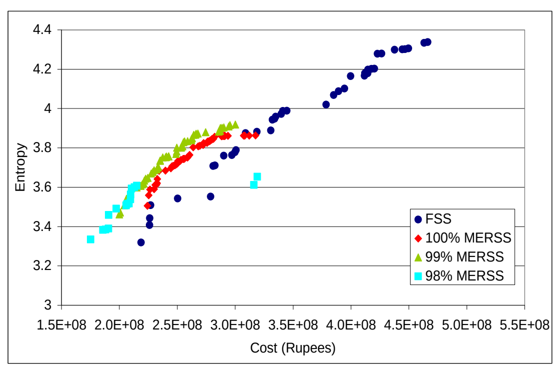
Run 15



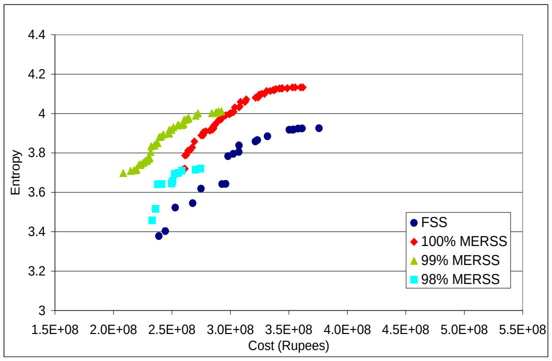
Run 16



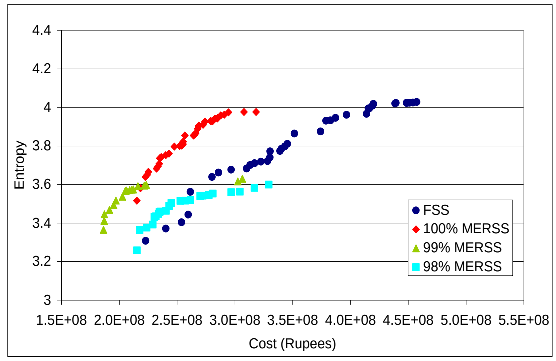
Run 17



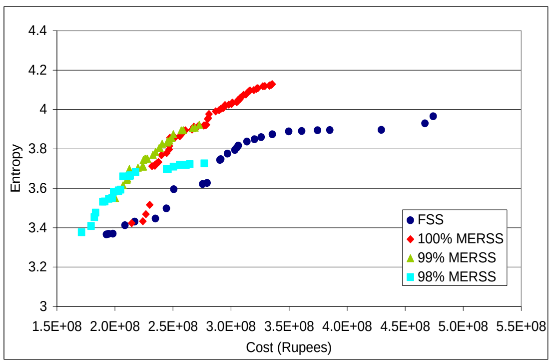
Run 18



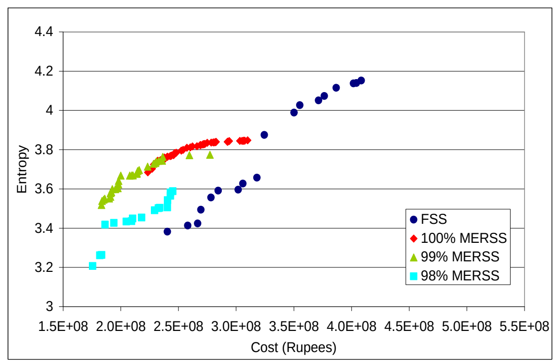
Run 19



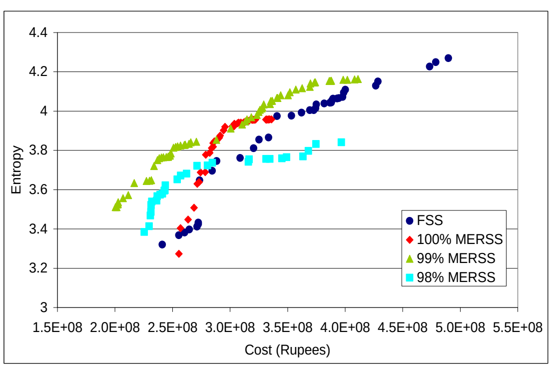
Run 20



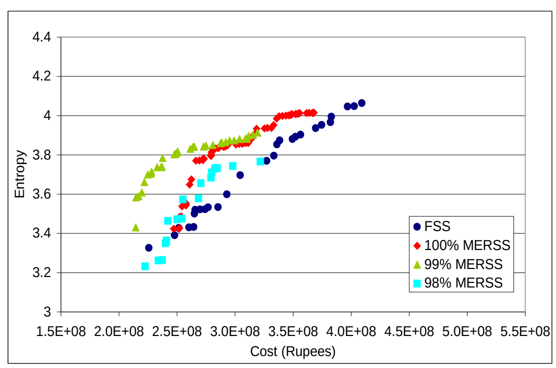
Run 21



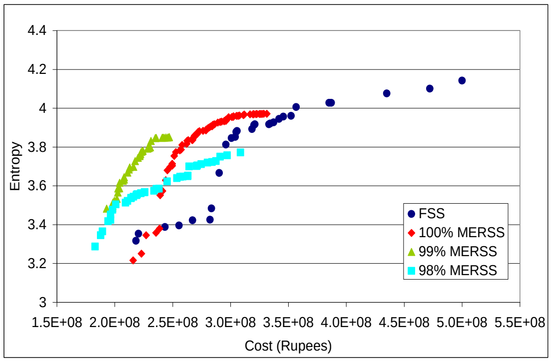
Run 22



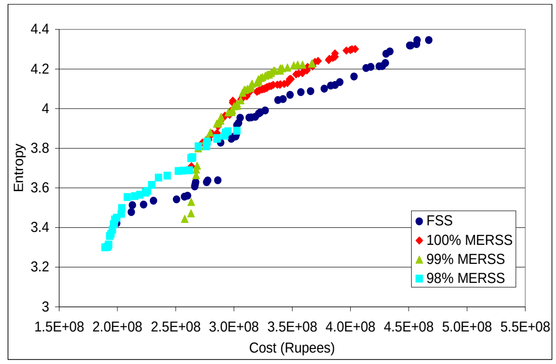
Run 23



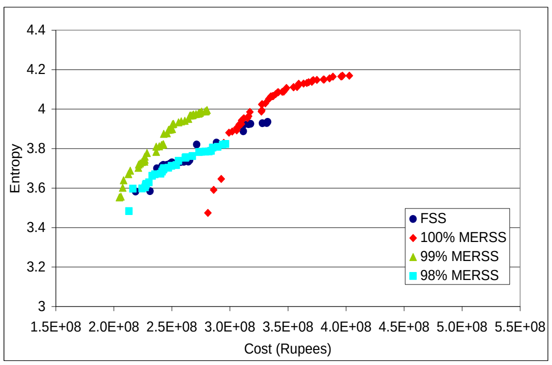
Run 24



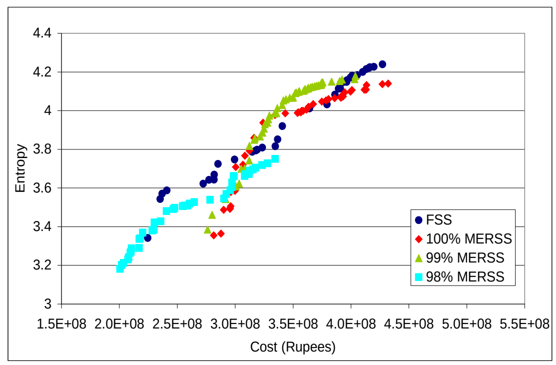
Run 25



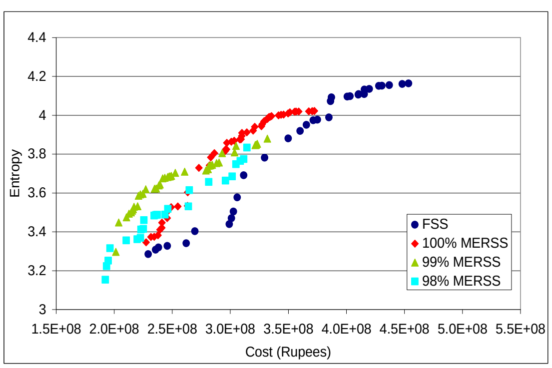
Run 26



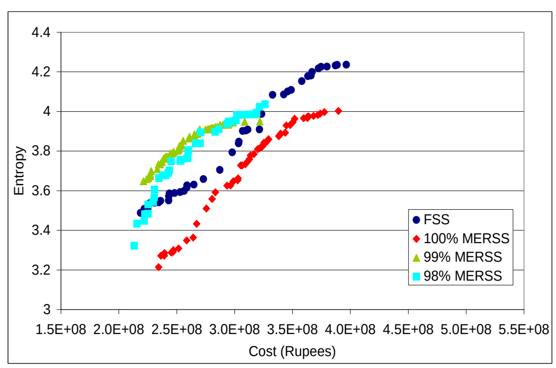
Run 27



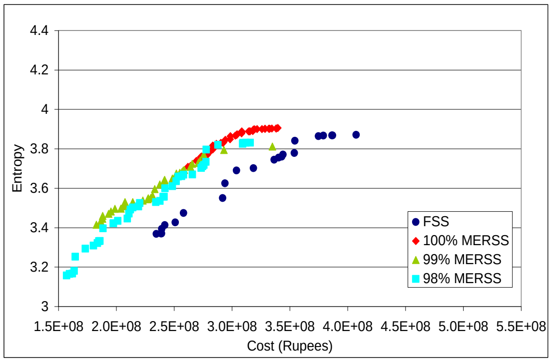
Run 28



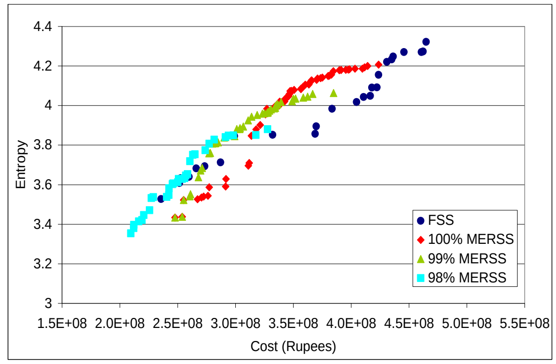
Run 29



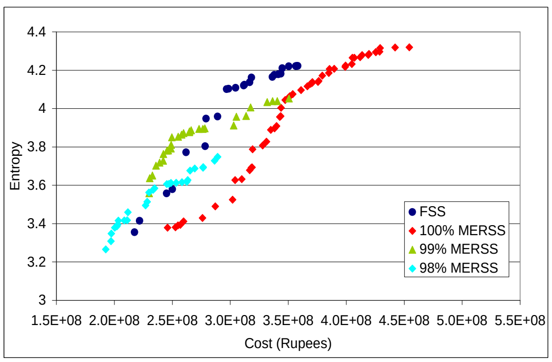
Run 30



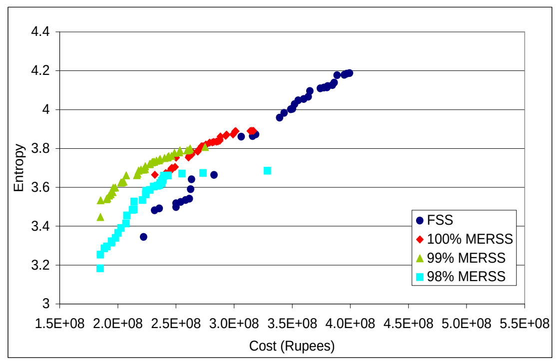
Run 31



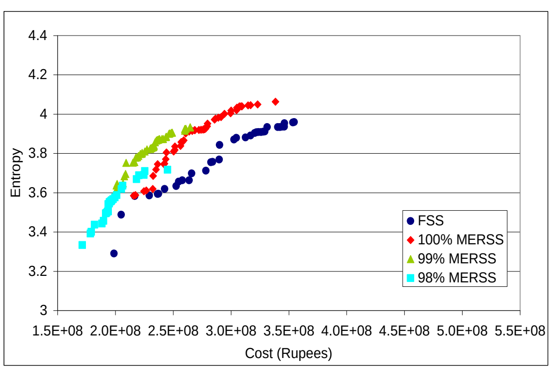
Run 32



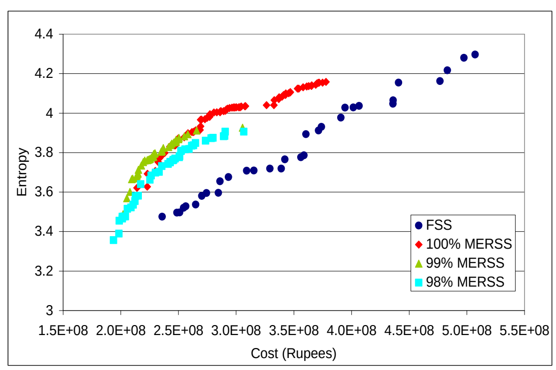
Run 33



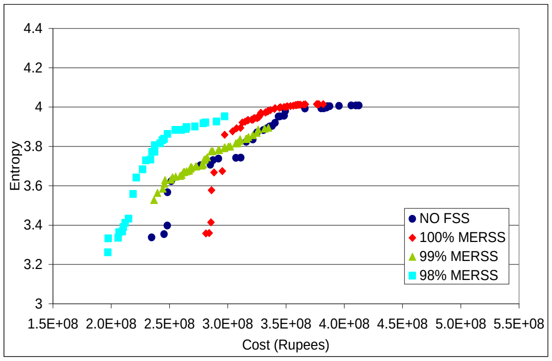
Run 34



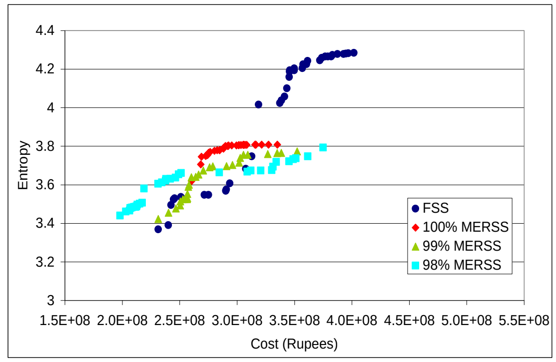
Run 35



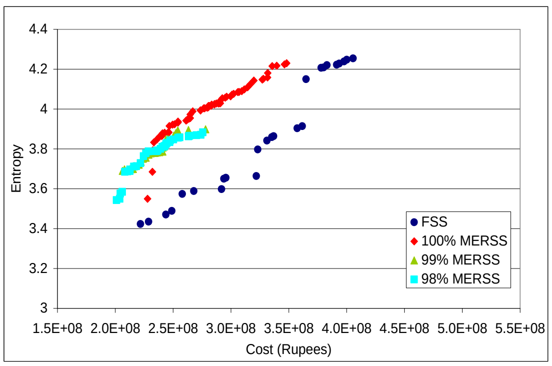
Run 36



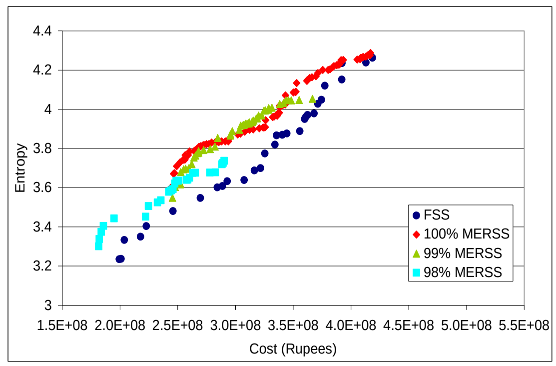
Run 37



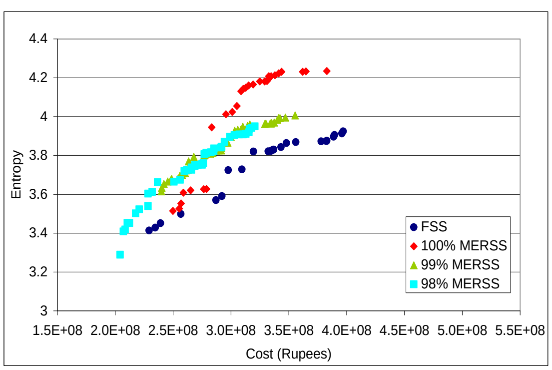
Run 38



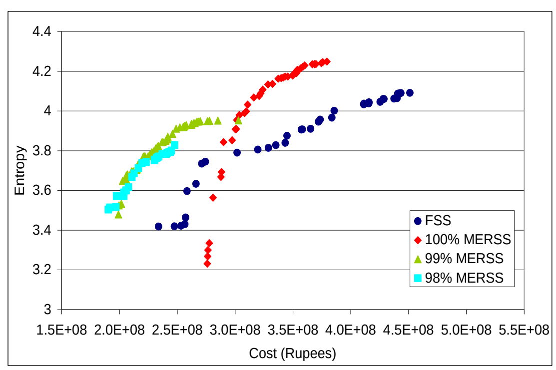
Run 39



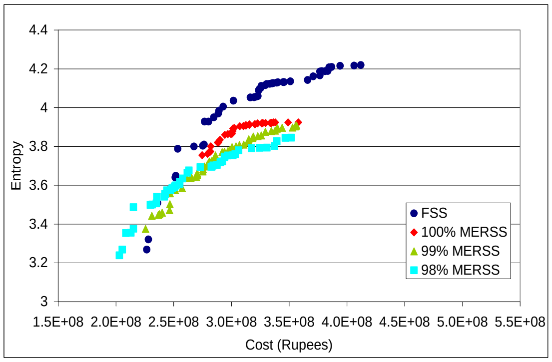
Run 40



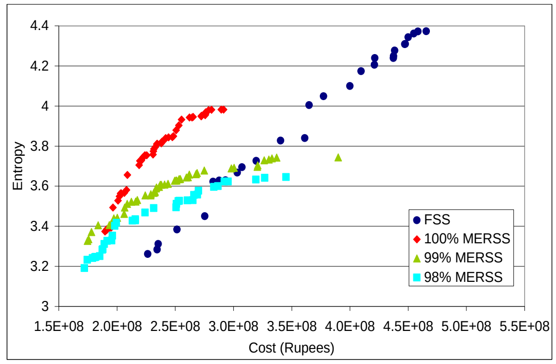
Run 41



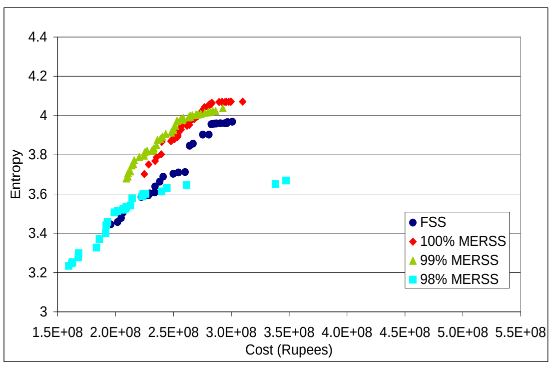
Run 42



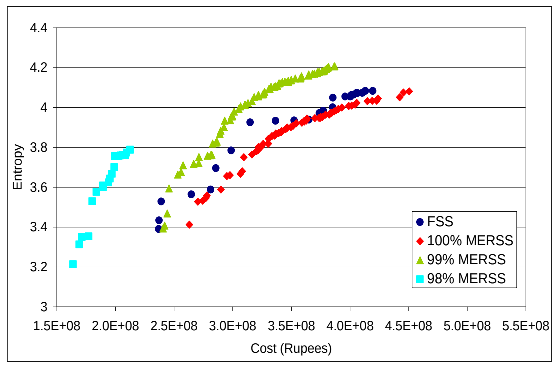
Run 43



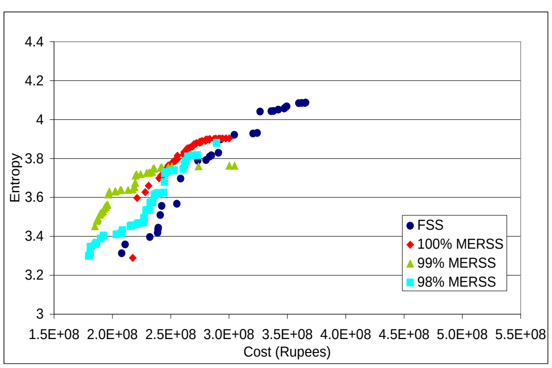
Run 44



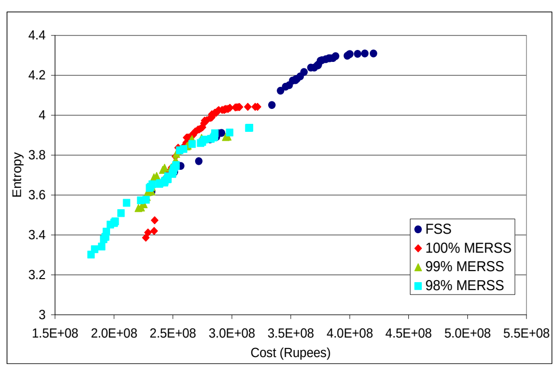
Run 45



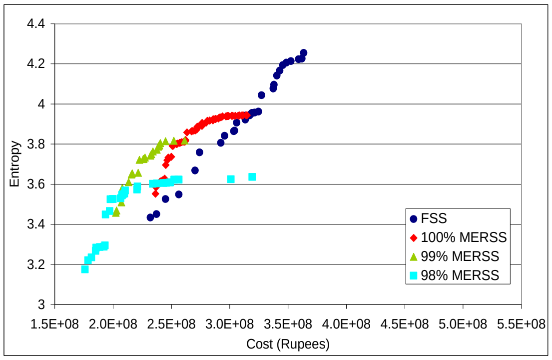
Run 46



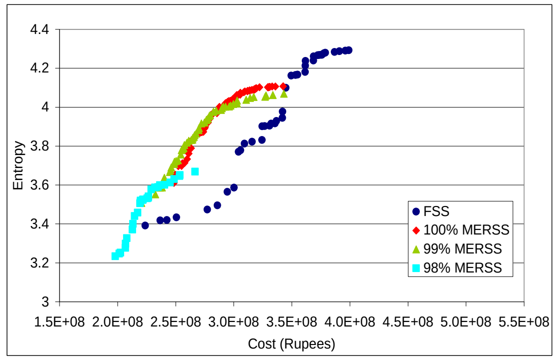
Run 47



Run 48



Run 49



Run 50

Figure D-1. Individual cost-entropy non-dominated POFs for 50 randomly generated runs

APPENDIX E

LIST OF PUBLICATIONS

JOURNALS PAPERS

Czajkowska, A.M., and Tanyimboh, T.T. (2013). Water distribution network optimization using maximum entropy under multiple loading patterns. *Water Science & Technology: Water Supply*, 13(5), 1265-1271.

Tanyimboh, T.T., Siew, C., Saleh, S. and **Czajkowska A.M.** (2016) Comparison of surrogate measures for the reliability and redundancy of water distribution systems, *Water Resources Management*, DOI:10.1007/s11269-016-1353-3, 2016.

CONFERENCE PAPERS

Czajkowska A.M., and Tanyimboh T.T. (2012). Maximum entropy design of water distribution systems under multiple operating conditions, *10th International Conference on Hydroinformatics*, Hamburg, Germany, 14-18 July, 2012.

Czajkowska, A.M., and Tanyimboh, T.T. (2012). Water distribution network optimization using maximum entropy under multiple loading patterns, *World Water Congress and Exhibition*, Busan, South Korea, 16-21 September 2012.

Czajkowska, A.M., and Tanyimboh, T.T., (2012). Reliability assessment of water distribution networks using surrogate measures, 14th Water Distribution Systems Analysis Conference, Adelaide, South Australia, 24-27 September 2012.