A DOMAIN AWARE GENETIC ALGORITHM FOR OPTIMUM BOOSTER CHLORINATION IN WATER DISTRIBUTION SYSTEMS

A MASTER’S THESIS

in

Computer Engineering

Atılım University

by

MUSTAFA KEMAL PEKTÜRK

JUNE 2010

A DOMAIN AWARE GENETIC ALGORITHM FOR OPTIMUM BOOSTER CHLORINATION IN WATER DISTRIBUTION SYSTEMS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

MUSTAFA KEMAL PEKTÜRK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF COMPUTER ENGINEERING

JUNE 2010

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

Prof. Dr. İbrahim Akman

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. İbrahim Akman

Head of Department

This is to certify that we have read the thesis "A Domain Aware Genetic Algorithm for Optimum Booster Chlorination in Water Distribution Systems" submitted by "Mustafa Kemal Pektürk" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Selçuk Soyupak	Asst. Prof. Dr. Hürevren Kılıç
Co-Supervisor	Supervisor

Examining Committee Members

Prof. Dr. Selçuk Soyupak

Prof. Dr. Habib Muhammetoğlu

Prof. Dr. Mehmet Önder Efe

Asst. Prof. Dr. Mehmet Cantürk

Asst. Prof. Dr. Hürevren Kılıç

Date: 24.06.2010

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Mustafa Kemal Pektürk

Signature:

ABSTRACT

A DOMAIN AWARE GENETIC ALGORITHM FOR OPTIMUM BOOSTER CHLORINATION IN WATER DISTRIBUTION SYSTEMS

Pektürk, Mustafa Kemal

M.S., Computer Engineering Department

Supervisor: Asst. Prof. Dr. Hürevren Kılıç

Co-Supervisor: Prof. Dr. Selçuk Soyupak

June 2010, 84 pages

Chlorine is generally applied at the exit of pumping stations or service reservoirs before being distributed to consumers. However such while such a single point application may not provide enough free residual chlorine levels for points which are far to the application points, it may end up with undesirably high free residual chlorine levels at consumer taps near to the source It is possible to overcome this problem by applying booster chlorination within networks. Selection of the number and best locations for booster stations, and the chlorine application levels is an important optimization problem. Proper solutions and applications will reduce the water born disease risk at far away locations and also reduce the risk of high levels that may create the formation of cancer causing halomethanes near to the source.

A software has been developed for determining of booster station locations and numbers water distribution networks along with concentration using Genetic Algorithm (GA) approach and Improved Genetic Algorithm (IGA) proposed within the scope of this thesis. Hydraulic and water quality simulations of the selected networks should be realized by EPANET software which is an open-sourced software. Genetic Algorithm and EPANET software should be run to solve the problem interactively. The performance level of the developed software was assessed using synthetic networks and Antalya-Konyaaltı network; and solutions with GA and with Improved GA (IGA) were compared for these networks. As a result, better solution is gained than GA using IGA.

Keywords: Booster Chlorination, EPANET, Genetic Algorithm, Optimization, Water Distribution Network.

ÖZ

İÇME SUYU DAĞITIM ŞEBEKELERİNDE OPTİMUM ARA KLORLAMA İÇİN ALANDAN HABERDAR GENETİK ALGORİTMA

Pektürk, Mustafa Kemal

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Yrd. Prof. Dr. Hürevren Kılıç

Yardımcı Tez Danışmanı: Prof. Dr. Selçuk Soyupak

Haziran 2010, 84 sayfa

İçme suyu şebekelerinde klorlama işlemi genelde tek noktadan olmak üzere ya pompa istasyonu ya da servis rezervuarı çıkışında yapılmaktadır. Bu tür uygulamalar sonucu kaynağa uzak noktalarda yetersiz klor seviyeleri gözlenirken kaynağa yakın bölgelerde ise istenmeyen seviyede klor ölçülebilmektedir. Bu durumun ortaya çıktığı su şebekelerinde, uygun noktalara birkaç ara klorlama istasyonu kurularak problem çözülebilmektedir. Bu işlemin yapılabilmesi için ara (ek) klorlama istasyonlarının sayısı ile yerlerinin seçimi ve klor dozajı büyük önem taşımaktadır. Zira serbest bakiye klorun şebekede çok düşük olduğu yerlerde sudan kaynaklanan bulaşıcı hastalıkların ortaya çıkma riski artarken, çok yüksek olduğu yerlerde ise uzun vadede kanserojen olduğu ileri sürülen bileşikler (halometanlar) oluşabilir.

Bu tez çalışması kapsamında, ara klorlamanın yapılacağı yerlere ve klor miktarına Genetik Algoritmalar (GA) ve bu tez kapsamında önerilen İyileştirilmiş Genetik Algoritmalar (İGA) kullanan bir yazılım geliştirilmiştir. Optimizasyonu yapılacak şebekenin hidrolik çözümünde ve su kalitesi analizinde EPANET adındaki açık kaynaklı yazılım kullanılmaktadır. Genetik Algoritma ile EPANET yazılımı etkileşimli olarak çalıştırılarak problem çözülmektedir. Geliştirilen yazılım, sentetik şebekelerde ve Antalya Konyaaltı bölgesindeki gerçek şebekelerde kullanılarak elde edilen GA ve İGA çözümleri karşılaştırılmıştır. Sonuç olarak İGA kullanılarak klasik GA’ dan daha iyi sonuçlar elde edilmiştir.

Anahtar Kelimeler: Ara Klorlama, EPANET, Genetik Algoritma, İçme Suyu Şebekesi, Optimizasyon.

To Turkey

ACKNOWLEDGMENTS

I express sincere appreciation to my supervisor Asst. Prof. Dr. Hürevren Kılıç and my co-supervisor Prof. Dr. Selçuk Soyupak for their guidance, patience and belief throughout this research. Thanks also go to TÜBİTAK, ASAT, project director Prof. Dr. Habib Muhammetoğlu and staff of the TÜBİTAK–KAMAG project whose project number is 107G088. To my fiancee, Şeyma, I offer sincere thanks for her continuous support and patience during this period.

TABLE OF CONTENT

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENT xi

LIST OF FIGURES xiii

LIST OF TABLES xv

CHAPTER

1. INTRODUCTION 1

2. BACKGROUND INFORMATION 5

  2.1. Problem Overview 5

  2.2. Genetic Algorithms 6

    2.2.1. Individual Representation 8

    2.2.2. Initial Population Construction 9

    2.2.3. Selection 9

    2.2.4. Crossover 9

    2.2.5. Mutation 10

    2.2.6. The Elitist Genetic Algorithm 11

3. EPANET 12

  3.1. What is EPANET? 12

  3.2. Steps in Using EPANET 13

  3.3. The EPANET Programmer's Toolkit 13

    3.3.1. Content 13

    3.3.2. Overview 14

3.3.3. Data Flow Diagram 15

4. APPLICATION OF CLASSICAL GA AND IMPROVED GA 17

4.1. Overview 17

4.2. Classic GA Implementation 19

4.2.1. Individual Representation 20

4.2.2. Initial Population Construction 20

4.2.3. Crossover Operation 20

4.2.4. Selection 21

4.2.5. Fitness Functions 23

4.3. Improved Genetic Algorithm 26

5. IMPLEMENTATION OF THE SOLUTION 31

5.1. Software Components 31

5.2. Tunable Parameters 33

5.2.1. GA Application Parameters 33

5.2.2. WDS Parameters 35

5.3. BetterNet Setup 38

6. RESULTS AND DISCUSSIONS 42

6.1. The Characteristics of the Networks Under Study 42

6.1.1. Test Network 1 (Hypothetical Network) 42

6.1.2. Test Network 2 (Hypothetical Network) 44

6.1.3. Zone 4 (Real Network) 46

6.2. Input Variables and Genetic Algorithm Parameters 49

6.3. Comparison of GA and IGA 50

7. CONCLUSIONS 63

REFERENCES 65

APPENDIX 70

APPENDIX A - Generated Report by BetterNet 70

APPENDIX B - Generation-Fitness Value Graph 78

APPENDIX C – Information about Real Network in Antalya 79

APPENDIX D – Comparison of GA & IGA Solution for Test Network 1 82

LIST OF FIGURES

Figure 2.1 Individual representation [28]. 8
Figure 2.2 Single point crossover [28]... 10
Figure 3.1 Data flow of EPANET [42]. 15
Figure 4.1 Architecture of the solution [28]. 19
Figure 4.2 Subset from individual 1 [29]. 21
Figure 4.3 Construction of child [29].. 21
Figure 4.4 Roulette-wheel selection.. 22
Figure 5.1 Components of BetterNet. 32
Figure 5.2 GA parameters for optimization. 34
Figure 5.3 Progress status screen for tracking optimization. 35
Figure 5.4 WDS parameters for optimization. 36
Figure 5.5 Example EPANET model for working on BetterNet. 38
Figure 6.1 Test Network 1 with 51 nodes, 2 tanks and 1 reservoir........................ 43
Figure 6.2 Demand pattern for Test Network 1. 44
Figure 6.3 Test Network 2 with 39 nodes, 1 tank and 2 reservoirs........................ 45
Figure 6.4 Demand Pattern for Test Network 2. 46
Figure 6.5 Zone 4 with 126 nodes and 1 reservoir.. 47
Figure 6.6 Demand pattern for Zone 4.. 48
Figure 6.7 Chlorination pattern for Zone 4. 49
Figure 6.8 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 4 Nodes. 51
Figure 6.9 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 6 Nodes. 52
Figure 6.10 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 8 Nodes. 53
Figure 6.11 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 4 Nodes. 54

Figure 6.12 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 6 Nodes. 55
Figure 6.13 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 8 Nodes. 56
Figure 6.14 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 4 Nodes. 57
Figure 6.15 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 6 Nodes. 58
Figure 6.16 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 8 Nodes. 59
Figure 6.17 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 4, 4 Nodes. 60
Figure 6.18 Comparison of GA & IGA for CP=0.9, EP=0.5, Objective 4, 6 Nodes. 61
Figure 6.19 Comparison of GA & IGA for CP=0.85, EP=0.4, Objective 4, 8 Nodes. 62
Figure B.1 Generation – fitness value graph for tracking progress. 78
Figure C.1 Overview to Antalya [43] 79
Figure C.2 Konyaaltı with 20 zones [43] 81
Figure D.1 Test Network 1 with 4 booster stations of GA solution...................................... 82
Figure D.2 Test Network 1 with 4 booster stations of IGA solution. 82
Figure D.3 Contour plot of chlorine at 36:00 hrs at GA solution. 83
Figure D.4 Contour plot of chlorine at 36:00 hrs at IGA solution. 83
Figure D.5 Time Series plot of chlorine for node 6, 23, 29 and 47 at GA solution. 84
Figure D.6 Time Series plot of chlorine for node 6, 23, 29 and 47 at IGA solution............. 84

LIST OF TABLES

Table 6.1 The major physical characteristics of the Test Network 1. 43

Table 6.2 The major physical characteristics of the Test Network 2. 45

Table 6.3 The major physical characteristics of the Zone 4. 48

Table 6.4 Parameters for BetterNet runs. 50

Table 6.5 Objective values of Objective 1 for CP=0.9, EP=0.6, 4 Nodes. 50

Table 6.6 Objective values of Objective 1 for CP=0.9, EP=0.6, 6 Nodes. 51

Table 6.7 Objective values of Objective 1 for CP=0.9, EP=0.6,8 Nodes. 52

Table 6.8 Objective values of Objective 2 for CP=0.9, EP=0.6, 4 Nodes. 53

Table 6.9 Objective values of Objective 2 for CP=0.9, EP=0.6, 6 Nodes. 54

Table 6.10 Objective values of Objective 2 for CP=0.9, EP=0.6, 8 Nodes. 55

Table 6.11 Objective values of Objective 3 for CP=0.9, EP=0.6, 4 Nodes. 56

Table 6.12 Objective values of Objective 3 for CP=0.9, EP=0.6, 6 Nodes. 57

Table 6.13 Objective values of Objective 3 for CP=0.9, EP=0.6, 8 Nodes. 58

Table 6.14 Objective values of Objective 4 for CP=0.9, EP=0.6, 4 Nodes. 59

Table 6.15 Objective values of Objective 4 CP=0.9, EP=0.5, 6 Nodes.................................. 60

Table 6.16 Objective values of Objective 4 CP=0.85, EP=0.4, 8 Nodes................................ 61

Table C.1 Production capacity of resources [43] 80

CHAPTER 1

INTRODUCTION

Disinfection of drinking water is a vital step after all water treatments to prevent deterioration of the microbial quality of water within distribution systems. Chlorine is most-commonly used water disinfectant in drinking water distribution networks [1]. It is not desirable to have low level of chlorine concentrations in water distribution networks: such a practice may be risky to avoid the growth of pathogenic microbes that may cause water-borne diseases. On the other hand high levels of chlorine concentration of water may cause the formation of cancerogenic compounds (trihalomethanes) and spoil taste at the network. Hence, holding the free chlorine residuals (FRC) throughout the water distribution networks within certain limits is very crucial operation for human health. Chlorine reaction and change of flow conditions within distribution networks creates variable chlorine concentrations throughout the water distribution networks. The chlorine reactions involve both the reactions with constituents in the bulk flow and reactions with materials deposited on pipe walls when the water propagates through the water distribution network. Thus, maintaining free chlorine residuals within a specified range is a very hard operation when chlorination of the water distribution network is carried out only at the water treatment plant. [1] highlighted the importance and effectiveness of booster chlorination for maintaining adequate chlorine residuals within distribution networks.

Maintaining free chlorine residuals within a specified range at all nodes in the network or at specified monitoring nodes is possible by adding booster chlorine station(s) to the network. One should realize that the determination of the number and locations along with concentration of chlorine booster stations is a challenging optimization problem considering the complexities of chlorine reactions and the dynamics of the water distribution networks. The objective of this thesis is to develop a software as a decision support tool for determination of number and locations along with concentration of chlorine booster station so that chlorine concentrations at the water distribution system (WDS) are maintained within a specified range while satisfying the problem specific constraints on the WDS.

Although multiple source disinfection schedules can be obtained using direct analysis water quality models, this is a tedious trial and error process in which it is difficult to identify the proper solution and costly.

In earlier studies, [3] proposed an inverse modeling technique for determining the source concentration of chlorine so as to maintain a specified concentration at a single monitoring node under dynamic conditions using a finite difference scheme.

[4] extended the model developed by [3] to the over and underdetermined cases, but used a constant supply source chlorine concentration at all times and failed to address the chlorine concentration at booster stations. It is a well established fact that finite difference schemes are subject to numerical diffusion [3]. Also, Eulerian methods are less memory efficient in simulating chemical transport [5].

The estimation of source dosages at multiple sources was first addressed by [6], and they formulated optimal scheduling of the booster disinfection problem in terms of a linear objective function which minimizes the total chlorine mass injection rate to satisfy residual constraints. The principle of linear superposition, which is applicable only if first-order reaction kinetics (bulk + wall) are assumed for chlorine, is used for determining the dynamic nodal chlorine concentrations. The constraints are written in terms of composite impulse response coefficients and the entire problem constitutes a linear programming (LP) model. However, the introduction of non-first-order bulk and wall reaction kinetics leads to nonlinear behavior of nodal chlorine concentrations with respect to source chlorine inputs. In this case, the constraints cannot be represented by a set of linear equations, and the LP model is inapplicable.

Some studies show that chlorine decay kinetics may be best described by second-order kinetics [7] [8] [9] and a semi-empirical combined first- and second-order model [10] in the bulk phase. Hence, it is essential to include non-first order bulk reaction kinetics for chlorine decay along with either first-order or zero-order wall reaction kinetics [11]. It also appears from objective function formulation and an example of an application by [6] that the water quality sources considered are only of mass injection source type.

Nature inspired computation techniques like Genetic Algorithms (GA) [12] [13], immunocomputing [14] and ant colony optimization [15], [16] are applied in WDS domain, successfully.

Genetic algorithms have been extensively applied to a wide variety of water distribution system optimization problems [17] [18] [19] [12]. [20] applied GA to optimization of pumping in a gas pipeline. Implementation of GAs on parallel computers was illustrated by [21], thereby providing considerable reduction in computing time. A GA is a heuristic search and optimization tool that performs differently from classical search and optimization methods. GA has been increasingly applied to various search and optimization problems because of their broad applicability, ease of use, and global perspective. Genetic algorithms as heuristic procedures are unable to guarantee global optimum solutions. However, experience indicates that they generally provide attainment of near-optimal solutions after a reasonable number of evaluations [18].

Location selection for optimal chlorination for disinfection purpose is a known problem in literature [22] [6]. The booster-based optimum chlorination scheduling research mostly considers linear or nonlinear kinetic decay with a single objective function [6] [23]. An alternative study proposes a pareto-based multi-objective approach for the problem [24]. The use of GA with a single objective function for solving the problem has been proposed in [25]. Two other GA based approaches for finding optimum location and number of booster stations have been introduced in [26] and [27], independently.

In this thesis, determining of booster station locations and concentration is intended using Genetic Algorithm (GA). Hydraulic and water quality simulations of the selected networks should be realized by EPANET software which is an open-sourced software. Genetic Algorithm and EPANET software should be run to solve the problem interactively. The performance level of the developed software was assessed using synthetic networks and Antalya-Konyaaltı network; and solutions with GA and with Improved GA (IGA) were compared for these networks.

At the classic GA, individuals are divided from random points and then crossover takes place. In this thesis, we proposed a new crossover method called domain-aware crossover which uses domain information and so new individuals take the best genes from their parents. Hence IGA is expected to avoid getting stuck in local minimum at the search space.

Rest of the thesis is organized as follows: in Chapter 2, background information about the Genetic Algorithms is given. Chapter 3 describes the EPANET software for hydraulic and water quality simulation of networks. Application techniques of Classical GA and Improved GA for water distribution systems (WDSs) are given at Chapter 4. In Chapter 5, you can find software implementation of the system and Chapter 6 gives experimental results. Chapter 7 summarizes the conclusions.

CHAPTER 2

BACKGROUND INFORMATION

2.1. Problem Overview

Although multiple source disinfection schedules can be obtained using direct analysis of water quality models, this is a tedious and costly trial and error process in which it is difficult to identify the proper solution.

Usually water distribution systems (WDS) are very hard to evaluate because they include many nodal points which are candidates to be booster station. Computing all possibilities and finding the optimum is extremely if not impossible. For example:

- Searching 50 candidate nodes, 5 booster station, 20 chlorination rate, 4 period
- Size of search space:

$$\begin{pmatrix} 50 \\ 5 \end{pmatrix} (20)^{5x4}$$

- Evaluating each point at this search space may take several years. Thus, we cannot find a solution in reasonable time.

Artificial Intelligence (AI) based systematic combinatorial search techniques are needed because of time cost of search of all possibilities and charge of trial and error method. For example:

- Genetic algorithm,
- Gradient Descent,
- Tabu search,
- Simulated annealing, etc. can be optimization method for finding a solution in a reasonable time.

2.2. Genetic Algorithms

In this thesis, Genetic Algorithm (GA) is adopted to find the optimized way for improving chlorination of given water network model. We preferred GA since Genetic Algorithm is a well-established search technique and it has advantage of defining problem easily. Alternative optimization techniques have not been considered because of getting sufficiently satisfactory results from GA.

Basically, two things should be determined in order to be able to use GA on a domain

- A genetic representation of the solution in the given domain,
- A fitness function to evaluate quality of the solution in the given domain.

Genetic Algorithm (GA) is an optimization technique which is based on nature inspired evolutionary mechanism [28]. Genetic Algorithms search the best solution of the multi dimensional search space with respect to principle of surviving of good individuals. In Genetic Algorithms, many candidate solutions are evaluated together in search space and probability of reaching the global best one is increased because of that Genetic Algorithms search the solution over a set that consists of many possible solutions. Each candidate solution on the solutions set is an independent vector on multi dimensional search space. Genetic Algorithms mimics evolutionary mechanisms and its steps. In GA terminology, the set is called “population”. Population consists of set of numbers either called vector, chromosome or individual. Each element of individual is called as gene. Individuals in the population are determined by GA operators such as initializing population, crossover, mutation, etc. Bad individuals are deleted from population with respect to these operators.

Representation of the individual in population is defined specifically to the problem. The representation of the individual, which represents the solution, is the most important factor in success of Genetic Algorithm in solving the problem [28]. A fitness function is used to determine the probability of being solution of each individual in the population. Fitness function is also called objective function. Value, which is returned by fitness function, is called fitness value or objective value. Individuals having high fitness values are considered in crossover operation with higher probability. As a result of crossover operation, new individuals called child come up. These new child individuals carry the properties of parent individuals. Individuals having low fitness values are eliminated from population over time because of not being crossovered. Hence, new population is composed by crossover of the last population’s individuals which have high fitness value so that good properties of the population are propagated from current generation to next generation.

For finding the best solution of the problem:

- Representation of the individual should be defined well,
- Fitness function should be effective and sufficient,
- Genetic Algorithm operators should be selected rightly.

It doesn’t guarantee the global best solution, but it provides acceptable and good solutions at an acceptable time period. Genetic Algorithms are used efficiently only for solving problems which;

- have a very big and complex search space,
- cannot be represented by mathematical models,
- cannot be solved by other optimization methods

2.2.1. Individual Representation

In this subsection, we will define our individual representation specific to the problem. However, different from the problem defined in Section 2.1, each individual (candidate solution) consists of only location (nodal point) and chlorination amount/rate of booster chlorination stations (see Figure 2.1) [28]. Chlorination period is not considered but fixed to a constant value. As a consequence, individual is a vector which consists of nodal points and dosages which are related to these points. Suppose there are m candidate chlorination locations at water distribution system (WDS) and accept that chlorination location count is determined as k by user (k<m). Consider that there are n different concentrations for chlorination at these candidate locations. So, the size of search space, where our solution lies on, is: $\binom{m}{k}(n)^k$.

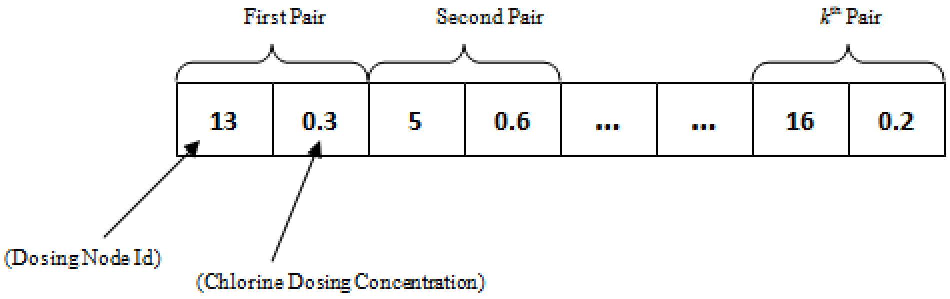


Figure 2.1 Individual representation [28].

2.2.2. Initial Population Construction

In general, at the beginning, many individual solutions are randomly generated to construct an initial population. Size of the population is dependent to the problem and search space, however traditionally the population consists of several hundreds or thousands of possible solutions. Typically, the randomly generated population covers the entire range of possible solutions.

2.2.3. Selection

Selection is an important GA operator in which individuals are chosen from the population to breed new individuals at the next generation. There are many methods to select the best individuals, for example; roulette wheel selection, tournament selection, rank selection, steady state selection and some others [29].

2.2.4. Crossover

It is the operation of composing new individuals from two individuals which are selected randomly from the population with respect to probability of fitness values (Figure 2.2) [28]. Crossover operation can be defined as an effort of composing new and well-qualified individuals by using parts which compose well-qualified individuals at the population. Crossover is done or not done between two individuals with respect to a probability value which is called crossover probability. After determining that crossover takes place with respect to crossover probability, two individuals are divided into parts at one or two points and new individuals are composed by these parts.

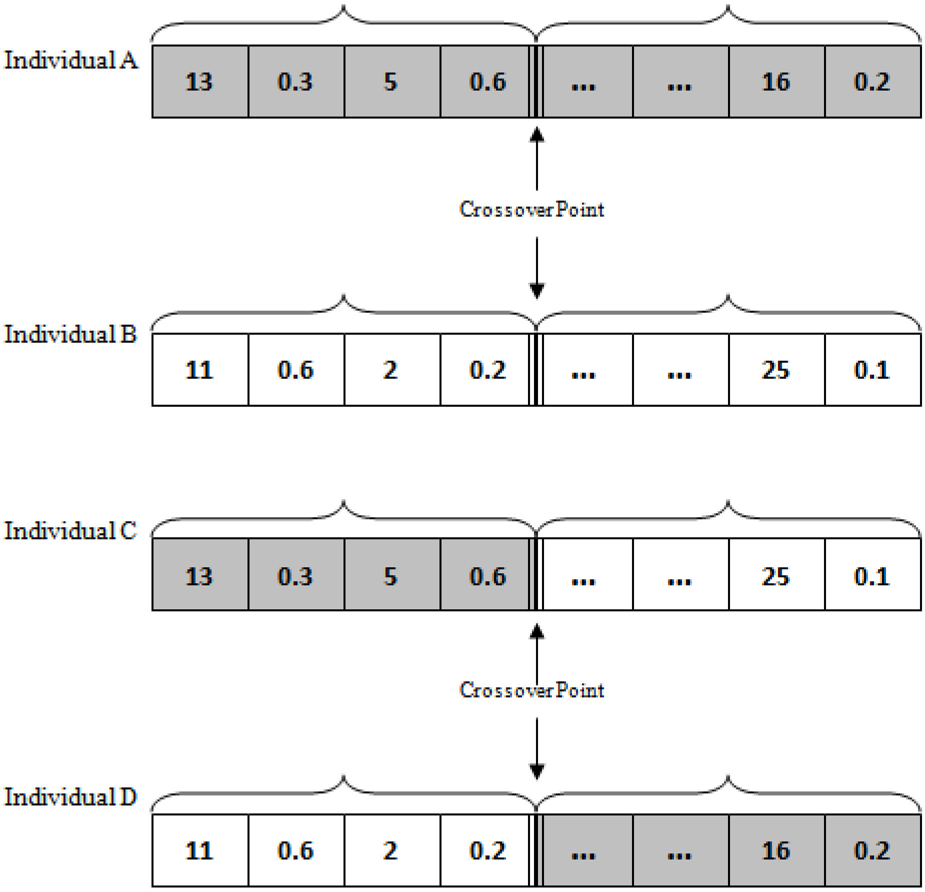


Figure 2.2 Single point crossover [28].

2.2.5. Mutation

Although crossover operation is applied, new individuals may not have better fitness value throughout following generations, because of getting similarity of individuals in the population. Hence, GA is stuck at local minimum/maximum and cannot find better solutions on the search space. To avoid this problem, changing any part of individuals randomly with a small probability is known as a useful operation [30]. In this operation, as a result of user-defined mutation probability, if mutation is needed on individual, any of chlorination location or chlorination amount is selected randomly, and it is changed randomly to assignable value.

An important point which must be noticed here is that if the user defined mutation probability value is high, degeneration probability of good individuals increases because of very high amount of mutation.

2.2.6. The Elitist Genetic Algorithm

The Genetic Algorithm operators may not always lead to better individuals [24]. For example, it is possible that the lowest fitness value of all individuals within the parent population of generation say X is greater than the lowest fitness value of all individuals within the parent population of generation X-1. To prevent the loss of good solutions, an elitist strategy can be used. The Genetic Algorithm elitism process saves a set of the best individuals from current generation (X-1) into next generation (X) directly.

After initial population construction, selection, crossover, mutation operations which are defined above, new population is composed from new child individuals and parent individuals of current population which have good fitness values when all individuals are sorted with respect to fitness values.

CHAPTER 3

EPANET

3.1. What is EPANET?

EPANET 2.0 is a program which solves hydraulic and water quality behavior of WDS [31]. This software is released by the United States Environmental Protection Agency. EPANET is used as a standalone program and as an open-source toolkit. Its engine is used by many software companies which develop more powerful, proprietary packages. EPANET is generally the core engine of these proprietary frameworks. A network consists of pipes, nodes, pumps, valves, storage tanks and reservoirs. EPANET tracks the flow of water in each pipe, the pressure at each node, the height of water in each tank, and the concentration of a chemical species throughout the network during a simulation period which consists of multiple time steps.

EPANET is designed as a research tool for better understanding of the movement and fate of drinking water constituents within WDS. It can be used to analyze different types of applications in WDS. Some examples:

- Sampling program design,
- Hydraulic model calibration,
- Chlorine residual analysis,
- Consumer exposure assessment.

EPANET provides an environment for editing network data, running hydraulic and water quality simulations, and viewing the results in a variety of formats. These include color-coded network maps, data tables, time series graphs, and contour plots.

3.2. Steps in Using EPANET

User typically carries out the following steps when using EPANET to model a water distribution system [31]:

1. Draw a network representation of your distribution system or import a basic description of the network placed in a text file.
2. Edit the properties of the objects that make up the system
3. Describe how the system is operated
4. Select a set of analysis options
5. Run a hydraulic/water quality analysis
6. View the results of the analysis.

3.3. The EPANET Programmer's Toolkit

3.3.1. Content

The EPANET Programmer's Toolkit is a dynamic link library (DLL) of functions that allows developers to use EPANET’s engine and its own function in their specific project [42]. The functions can be called by 32-bit Windows applications written in C/C++, Delphi Pascal, Visual Basic, or any other language that can call functions within a Windows DLL. The Toolkit DLL file is called EPANET2.DLL and is distributed with EPANET 2.0.

EPANET and Programmer's Toolkit were developed by the Water Supply and Water Resources Division of the U.S. Environmental Protection Agency's National Risk Management Research Laboratory.

3.3.2. Overview

The Programmer's Toolkit is an extension for EPANET programmers. It presents a series of functions (see Appendix C) to programmers that want to use EPANET's hydraulic and water quality solution engine in their own applications [42]. Before using it, the programmer should become familiar with how EPANET represents a water distribution network and how it performs a simulation on this network. Information of toolkit and EPANET design can be obtain from reading EPANET's on-line Help file and the EPANET Users Manual.

3.3.3. Data Flow Diagram

The EPANET Toolkit is written in ANSI standard C with separate code modules for input processing, hydraulic analysis, water quality analysis, sparse matrix/linear equation analysis, and report generation [42]. The data flow diagram for analyzing a pipe network is shown below (see Figure 3.1). The processing steps depicted in this diagram can be summarized as follows:

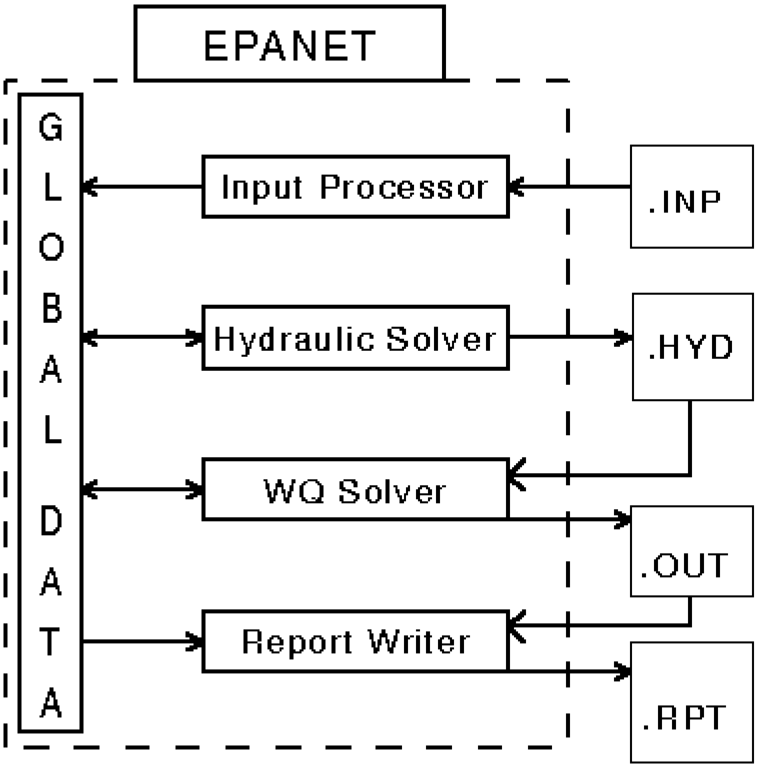


Figure 3.1 Data flow of EPANET [42].

The input processor module receives a network model with .INP extension and it parses, interpretes, and stores the model file's contents in a shared memory area.

The hydraulics solver module applies an extended period hydraulic simulation on given model. The results of every time step can be written to an external, unformatted (binary) hydraulics file (.HYD).

If a water quality simulation is requested, the water quality module accesses the flow data from the hydraulics file which computes and stores substance transport and reaction through the network model for each hydraulic time step. It can write computed hydraulic results and its water quality results to an unformatted (binary) output file (.OUT).

If requested, a report writer module takes the calculated simulation results from the binary output file (.OUT) for each reporting period writes selected values to formatted report file_(.RPT). Error or warning messages taken throughout the simulation are also written to this file.

CHAPTER 4

APPLICATION OF CLASSICAL GA AND IMPROVED GA

4.1. Overview

Abstraction of natural mechanisms and their usage for efficient problem solving is a known systematic approach to tackle with hard optimization problems [32]. Depending on the problem (or problem instances) under study, approaches like evolutionary computing [33]; ant colony optimization [34]; particle swarm optimization [35] and immunocomputing [36] [37] may provide us mathematical tools that are useful in intelligent search, problem solving and optimization. The close relation between problem solving and optimization makes the above approaches natural candidates for efficient problem solving. Still, on the other side, we have the problem of representation. How to represent (or model) real systems under study in order make/take more realistic decisions/actions about their control? Should software solution be an online, automated (even autonomous) one or be an offline decision support level? Answers for these two questions clearly depend on the requirements of problem under study, their criticality and on modeling capability at hand.

Our GA-EPANET based software solution what we call the BetterNet can be positioned as an offline decision support tool for optimum chlorination of water distribution systems. In this study, our answer for better representation is not based on a data-driven automated modeling (as in the use of neural networks [38] for such purpose), but provides an expert-designed, equation-based deterministic models supported by open-source EPANET software. Simply, the proposed solution is nothing but a plug-in level search component that can process different water network optimization criteria.

In order to solve the location and chlorination amount selection problem for Free Residual Chlorine (FRC) level control in WDS, we combined a nature inspired computing solution namely Genetic Algorithms [39] with EPANET based deterministic modeling solution. Intelligent search mechanism provided by GA and realistic modeling via the adopted deterministic approach were two basic components of the proposed solution. Components execute together in the form of a wrapper depicted in Figure 4.1. During BetterNet execution, the GA component searches for the best chlorination location and amount that tries to keep FRC levels as much as possible in acceptable ranges in terms of public health concerns. The model responses are obtained by extended period simulations (EPS) that are executed on the deterministic network model. The responses are considered as the fitness results for the current chlorination strategy.

The search component is implemented as a GA introduced in [40]. We represent the genotypes (i.e. individuals or chromosomes) as chlorination strategy patterns. Patterns correspond to candidate solutions to the problem. Each candidate solution is constituted by a vector of node-id and chlorine concentration pairs. Fitness score of an individual is related to its chlorination success that achieves the objective function under study. The score is measured based on the feedback obtained from the EPS result. BetterNet implements Roulette Wheel selection algorithm [41] for individual selection. Crossover operator is implemented as a single line column-wise cut. Variation is provided by a single location or a single chlorination amount mutation.

Note that, the solution uses the open-source EPANET software for equation-based deterministic model description and its EPS engine for modeling and simulation purposes, respectively. By this way, deterministic solutions can be obtained. However, it requires an expertise in water distribution network design and analysis. The user is supposed to be able to do hydraulics modeling, calibration and water quality modeling using EPANET.

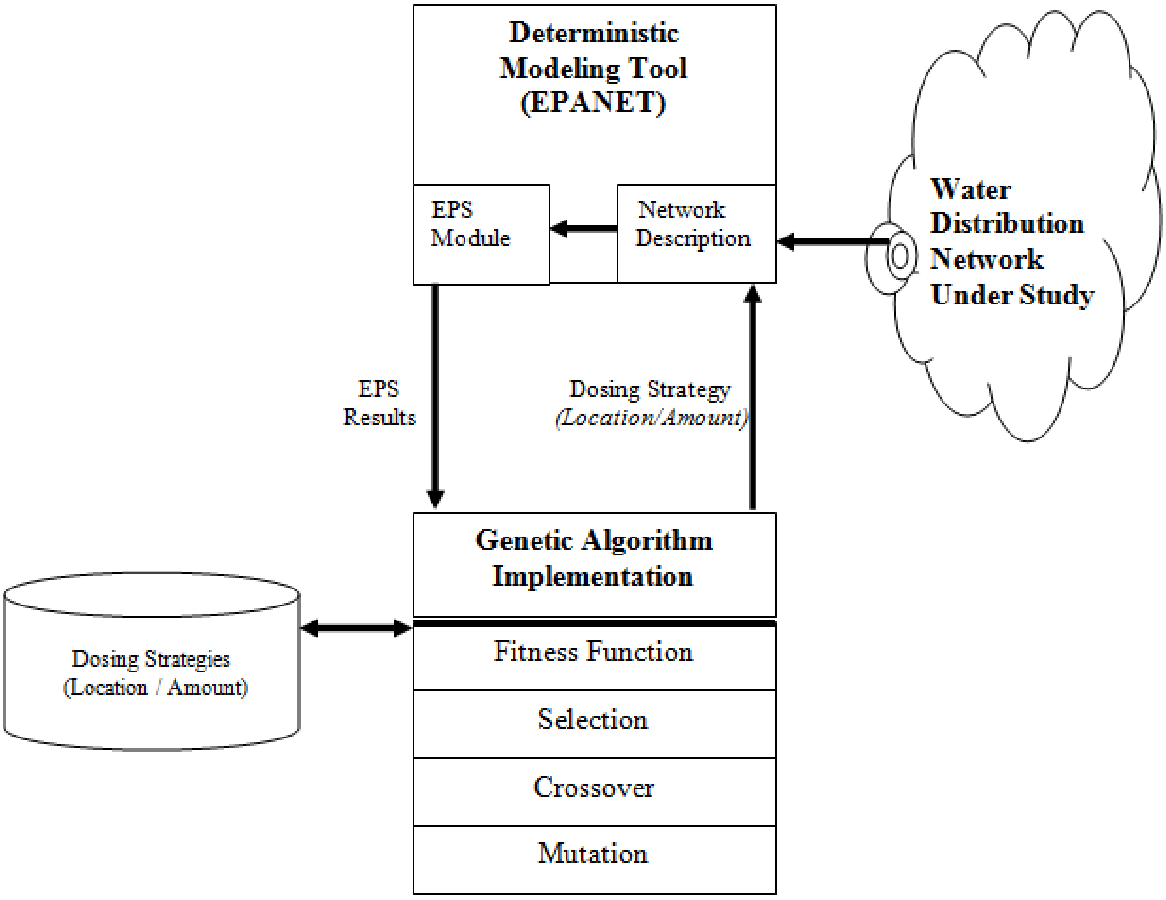


Figure 4.1 Architecture of the solution [28].

4.2. Classic GA Implementation

In BetterNet, we implemented the Classical GA description. This subsection explains the details of the GA implementation realized in BetterNet [28].

4.2.1. Individual Representation

Each individual (candidate solution) consists of location (nodal point) and chlorination amount of booster chlorination stations. In other words, individual is a vector which consists of nodal points and concentrations which are related to these points. Refer back to subsection 2.2.1 for details.

4.2.2. Initial Population Construction

Initial population can be constructed with two different ways. Firstly, user enters population size value to the software. User may want that individuals of initial population are composed randomly by the software. Uniformly distributed random number generator is used to generate random number. Other alternative distribution function can be Gaussian, Exponential, Binomial and Bernoulli but they are not considered and compared as an effort at this thesis. Alternatively, user may want to enter some individuals, which consist of chlorination locations and amounts of chlorination, to population on the software and if there is any empty slot on population, the software completes it randomly. With this alternative way, experience of user on determining chlorination locations and amounts is used to provide utility on finding the solution efficiently.

4.2.3. Crossover Operation

It is the operation of composing new individuals from two individuals which are selected randomly from the population with respect to probability of fitness values. Crossover operation can be defined as effort of composing new and well-qualified individuals by using parts which compose well-qualified individuals at the population. Crossover is done or not done between two individuals with respect to a probability value which is called crossover probability. After determining that crossover takes place with respect to crossover probability, two individuals are divided to parts by one or two points and new individuals are composed by these parts. In this way, there may be duplication in child individuals with respect to Node ID. Duplication of Node ID is not allowed because any node must be shown one time at an individual.

An algorithm to solve the duplication problem is that [29]:

- Choose an arbitrary part from the first parent (Figure 4.2.)
- Copy this part to the first child
- Copy the numbers that are not in the first part, to the first child (Figure 4.3):
    - starting right from cut point of the copied part,
    - using the order of the second parent
    - and wrapping around at the end
- Analogous for the second child, with parent roles reversed

Copy randomly selected set from the first parent

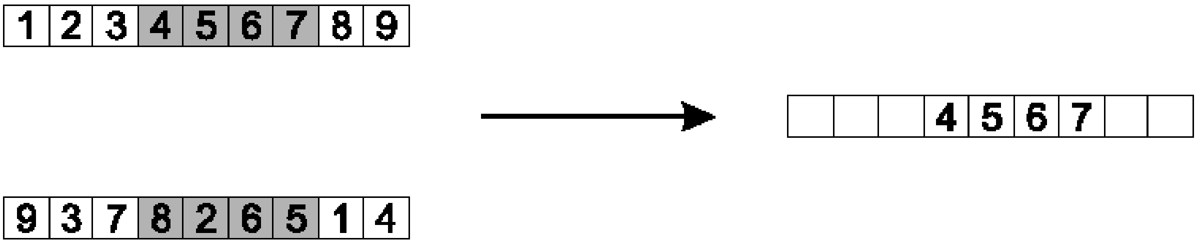


Figure 4.2 Subset from individual 1 [29].

Copy rest from second parent in order 1, 9, 3, 8, 2.

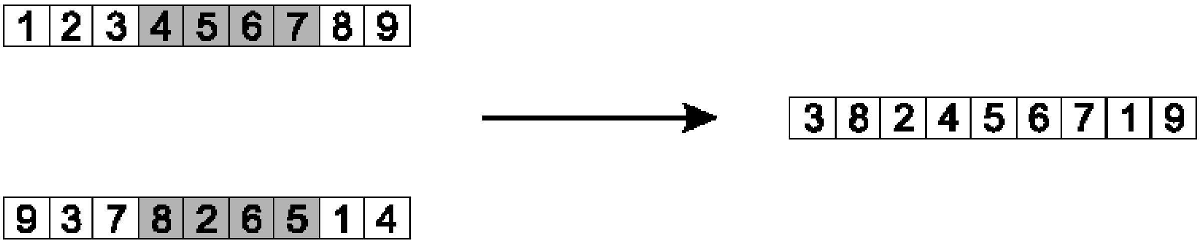


Figure 4.3 Construction of child [29].

4.2.4. Selection

There are many methods to select the best individuals, for example; roulette wheel selection, tournament selection, rank selection, steady state selection and some others

(Eiben and Smith 2003) [29]. In BetterNet solution, roulette wheel selection was implemented as a selection operator (Figure 4.4) because fitness values of individuals in the population are very close to each other [40]. In the roulette wheel selection, parents are selected with respected to their fitness value. The better individuals have more chance to be selected. Imagine a roulette wheel where all individuals are placed in the population, each one has its place big with respect to its fitness value, such as the following picture.

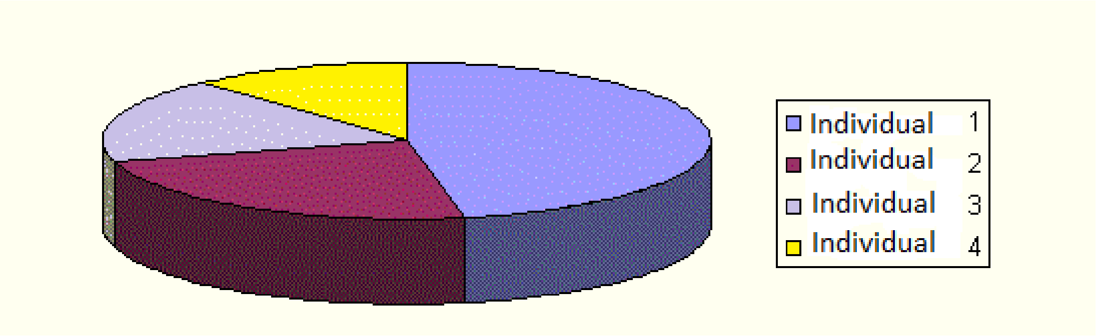


Figure 4.4 Roulette-wheel selection.

Then a marble is thrown there and selects the individual. Individuals with bigger fitness will be selected more times.

Roulette wheel selection procedure in the proposed solution:

- The fitness function is computed for each individual and fitness value is assigned to each one, which is mapped later.
- The population is sorted by ascending fitness values. Fitness values are mapped to 0 and 1. Mapped means that each fitness value is transformed to

value between 0 and 1. The best individual fitness value is 1 and the worst one fitness value is 0.

- Individuals have percentage on the wheel with respect to their mapped fitness value.
- A random value is used to determine the selected individuals according to mapped fitness value.

For mutation and elitism refer to subsections 2.2.5 and 2.2.6, respectively.

4.2.5. Fitness Functions

In BetterNet, for all objective function applications the algorithm works in such a way that we try to keep ($c_{min} \leq c \leq c_{max}$) for any node and for any time [28]. The objective functions 1 and 2 are alternative objective functions to reduce variability of FRC levels within any WDS.

1. Centralization of FRC levels:

$$Min(\ SSD = \sum_{i=1}^{m} \sum_{j=1}^{T} \left| c(i,j) - c_{median} \right|^2 \ ) \qquad \text{where}$$

SSD(m, T) : The sum of square of differences between FRC concentration levels and desirable average FRC concentration value (median of maximum and minimum allowable concentrations) throughout the EPS process,

m : Total number of nodes,

T : EPS duration excluding the starting transition period,

i, j : Indices for the nodes and time steps,

c(i,j) : Calculated FRC concentration level at i^th^ node for the j^th^ time step,

$$c_{median} = \frac{c_{\max} - c_{\min}}{2}$$

c_{min} : Minimum allowable FRC concentration level within WDS,

$c_{\max}$: Maximum allowable FRC concentration level within WDS,

Allowable range: ($c_{\min} \leq c_{i,j} \leq c_{\max}$)

2. Minimization of variance of FRC levels within WDS:

$$Min\ \sigma^2 = \frac{\sum_{j=1}^{T} \sum_{i=1}^{K} (c_{i,j} - \bar{c}\ )^2}{N-1} \qquad \text{where}$$

$c_{i,j}$: FRC concentration level of node i in time j ,

σ^2 : Variance of concentrations ,

: The index of a node ,

T : EPS duration excluding the starting transition period,

$\bar{c}$: Average of the calculated concentrations ,

N : T*K

3. Minimization of risk of occurrence probability of FRC concentration values outside of the allowable range:

The purpose of this objective function is to minimize the risk level of consumption of water with FRC concentrations outside the allowable range. The minimum and maximum risk levels that can be calculated are 0 and 1 respectively.

$$Min\ (R = risk\ of\ consumption\ of\ water\ with\ FRC\ levels\ outside\ of\ allowable\ range = (1 - \frac{\sum_{j=1}^{T} \sum_{i=1}^{KP} Q_{i,j}\ c_{i,j}}{\sum_{j=1}^{T} \sum_{i=1}^{NN} Q_{i,j}\ c_{i,j}})$$

where

$c_{i,j}$: Chlorine concentration level of node i in time j,

$Q_{i,j}$: The amount of flow demand at node i in time j,

KP : Twhere $c_{\min} \le c_{i,j} \le c_{\max}$

NN : Total number of nodes,

T : EPS duration excluding the starting transition period,

4. Minimization of trihalomethane formation risk index:

This objective function has been selected to minimize system specific trihalomethane formation risk index (TFRI) by keeping FRC level as low as possible. There is no universal limit for upper TFRI levels. The values might vary and the relative magnitudes should be compared under different proposed solutions for the same WDS.

$$Min\ (SSD = TFRI = \sum_{i=1}^{NN}\sum_{j=1}^{T}\left|c(i,j) - c_{\min}\right|^{2}) \quad \text{where}$$

SSD : The sum of square of differences between FRC concentration levels and permitted lowest concentration values throughout the EPS process,

NN : Total number of nodes,

T : EPS duration excluding the starting transition period,

i, j : Indices for the nodes and time steps, c(i, j) : FRC concentration level at node i in time step j,

It must be noted here that some of these objectives might be conflicting with others. The BetterNet solution was not intended to have multi-objective optimization capabilities.

4.3. Improved Genetic Algorithm

For classic GA, individuals are divided from random points and then crossover takes place. In this thesis, we proposed a new crossover method called domain-aware crossover which uses domain information and so new individuals take the best genes from their parents. For improved method, third fitness function (Risk of occurrence probability of FRC concentration values outside of the allowable range) is used to be domain aware in the proposed solution. All genes are considered together and contribution of each gene to individual is computed with respect to risk function. Hence, good genes are passed to first child and bad genes are passed to second one. Bad child is lost at the next generation later because of low fitness value. This crossover method (IGA) can be applied to problem which has an appropriate fitness value when one gene of individual is turned-off. IGA can be used for any one of combinatorial problem. It is not applicable for non-combinatorial problems. For example, this method cannot be applied to 8-queens problem because when we turn-off one of genes, we lose one column and one row of board, thus fitness value cannot be computed rightly after turning off one of genes and the method cannot be applicable more.

In the domain aware crossover, first and second parents, which are chosen by selection operator, are evaluated together in the given WDS model. In this scenario, firstly chlorination of each node in these two parents is done with respect to given chlorine concentration of them and risk value of the WDS model is computed accordingly. Secondly contribution of each node to minimizing the computed risk value is computed by closing chlorination of each node one by one.

Procedure to perform domain aware crossover:

- Chlorination is applied by adding booster stations of selected two parents together with given concentrations. Then, risk value of WDS is computed considering these two parents as one individual.
- Chlorination of each node is turned-off and then risk value is computed for each turning-off operation. One node turned-off risk value is divided to overall risk value which is calculated at first step; hence contribution of each node (or gene) to minimizing the overall risk value is calculated.
- All genes from first and second parents are sorted in descending order with respect to contribution of them, which is calculated at second step. Then, first half of sorted genes composes new child with better genes, second half composes second new child with worse genes.

Parent 1:

7	0,3	16	0,35	35	0,4

Parent 2:

23	0,4	10	0,3	39	0,4

<u>Step 1:</u>

Combine two parents to calculate overall risk value:

7	0,3	16	0,35	35	0,4	23	0,4	10	0,3	39	0,4

Risk value (R) calculated with respect to this vector

Minimization of risk of occurrence probability of FRC concentration values outside of the allowable range:

$$\text{R=risk of consumption of water with FRC levels outside of allowable range} = (1-\frac{\sum_{j=1}^{T}\sum_{i=1}^{KP}Q_{i,j}c_{i,j}}{\sum_{j=1}^{T}\sum_{i=1}^{NN}Q_{i,j}c_{i,j}})$$

Where

$c_{i,j}$: Chlorine concentration level of node i in time j,

$Q_{i,j}$: The amount of flow demand at node i in time j,

KP : Twhere $c_{\min} \leq c_{i,j} \leq c_{\max}$

NN : Total number of nodes,

T : EPS duration excluding the starting transition period,

Step 2:

Calculate contribution of each gene k : $C_k = \dfrac{1-\dfrac{\sum_{j=1}^{T}\sum_{i=1}^{KP}Q_{i,j,k}c_{i,j,k}}{\sum_{j=1}^{T}\sum_{i=1}^{NN}Q_{i,j,k}c_{i,j,k}}}{1-\dfrac{\sum_{j=1}^{T}\sum_{i=1}^{KP}Q_{i,j}c_{i,j}}{\sum_{j=1}^{T}\sum_{i=1}^{NN}Q_{i,j}c_{i,j}}}$

Where

$c_{i,j}$: Chlorine concentration level of node i in time j,

$c_{i,j,k}$: Chlorine concentration level of node i in time j without chlorination of node k,

$Q_{i,j}$: The amount of flow demand at node i in time j,

$Q_{i,j,k}$: The amount of flow demand at node i in time j without chlorination of node k,

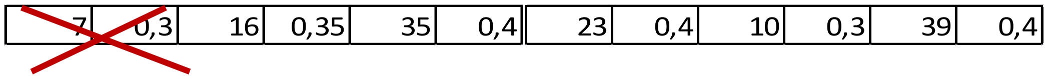
k : The chlorination node,

KP : Twhere $c_{min} \leq c_{i,j} \leq c_{max}$

NN : Total number of nodes,

T : EPS duration excluding the starting transition period,

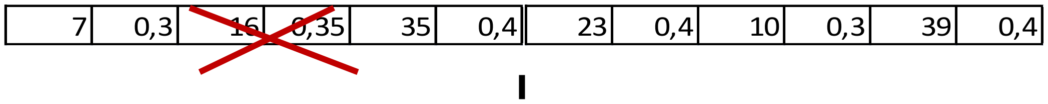
Contribution of gene 1:



~~7~~	~~0,3~~	16	0,35	35	0,4	23	0,4	10	0,3	39	0,4

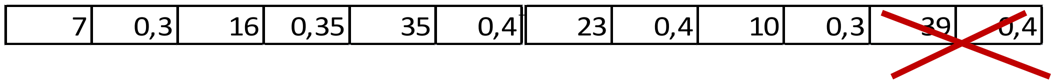
Risk value (R1) of that individual without gene 1 is calculated and that represents the contribution of gene 1 to overall risk value (Cm).

Contribution of gene 2:

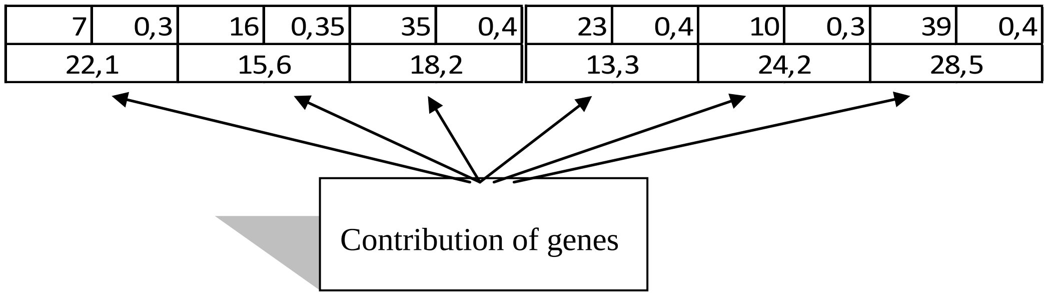


7	0,3	~~16~~	~~0,35~~	35	0,4	23	0,4	10	0,3	39	0,4

Contribution of gene 6:



7	0,3	16	0,35	35	0,4	23	0,4	10	0,3	~~39~~	~~0,4~~



7	0,3	16	0,35	35	0,4	23	0,4	10	0,3	39	0,4
22,1		15,6		18,2		13,3		24,2		28,5	

Step 3:

After sorting in descending order:

39	0,4	10	0,3	7	0,3	35	0,4	16	0,35	23	0,4
28,5		24,2		22,1		18,2		15,6		13,3	

Child 1(with better genes):

39	0,4	10	0,3	7	0,3

Child 2(with worse genes):

35	0,4	16	0,35	23	0,4

CHAPTER 5

IMPLEMENTATION OF THE SOLUTION

5.1. Software Components

BetterNet is developed as a stand-alone application written in C/C++ programming language. The program calls EPANET Dynamic Link Library (DLL) of functions that allows developers to customize EPANET's computational engine for their own specific needs. The functions can be incorporated into 32-bit Windows applications in any language that can call functions within the DLL. The program evaluates models by performing extended period simulation (EPS) of hydraulic and water quality behavior within networks. BetterNet also uses available Interop.dll to give some graphical view to user. The software is implemented as a collection of five components (Figure 5.1). Graphical user interface provides user an interface to enter model parameter, GA parameter and the model itself to the system. It also shows a screen for watching progress status of the program. Briefly, this component is a bridge between user and GA component. GA component includes genetic operators: Selection, crossover, mutation, fitness functions, etc. It calls EPANET DLL to computing fitness values of individuals within network model. Firstly, it solves hydraulic behavior of water distribution systems, and then it analyzes water quality behavior of the network with each individual one by one. EPANET performs extended period simulation of hydraulic and water quality behavior within pressurized pipe networks. A network can consist of pipes, nodes (pipe junctions), pumps, valves and storage tanks or reservoirs. EPANET tracks the flow of water in each pipe, the pressure at each node, the height of water in each tank, and the concentration of a chemical species throughout the network during a multi-time period simulation. In addition to chemical species, water age and source tracing can also be simulated.

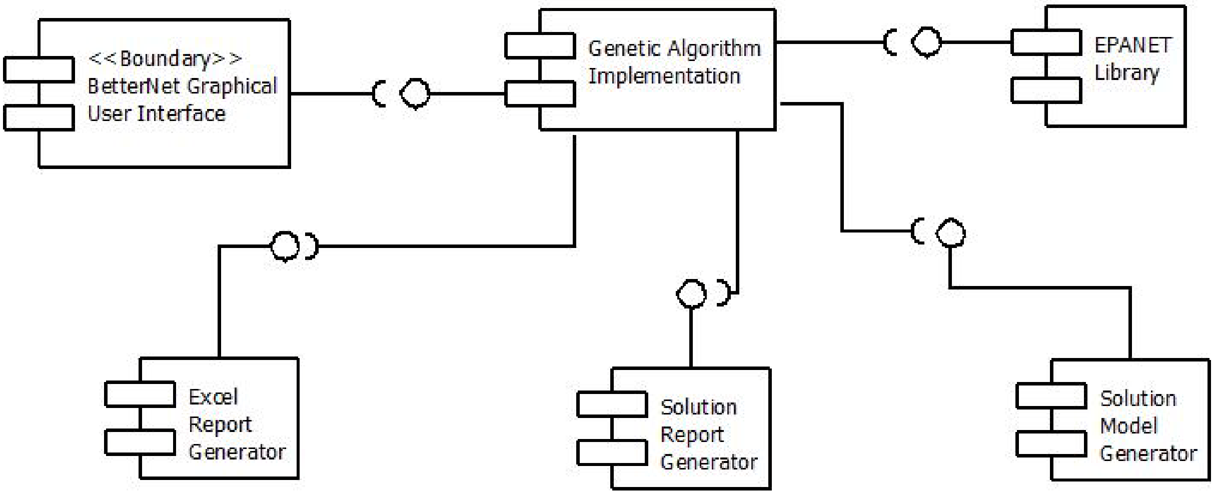


Figure 5.1 Components of BetterNet.

BetterNet can generate a solution report within solution report generator component. This report includes model path, solution date, all parameters and fitness values of all fitness functions for each generation and the solution. The program gives a table for following each generation and corresponding fitness value via generation tracking component. It provides a chart with generation number and fitness value for understanding solution graphically. Finally, after the program finds the solution, solution model generator component prepares a model with booster stations which are determined by GA component. User can open this model within EPANET, directly.

Analysis of water quality of model is very difficult because WDS is a dynamic system. Pressure, demand, flow, etc. of the system can change simultaneously and fast with time. Chlorine concentrations at nodes can change rapidly with respect to these hydraulic properties. So, WDS cannot be analyzed only topologically and it must be analyzed with its dynamic properties. Thus analysis of water quality of model requires many computations with dynamic properties. BetterNet user doesn’t need to consider these hard computations more.

BetterNet is developed as a decision support system for WDS. It should be used by domain experts who understand water network design and some genetic algorithm. BetterNet Manual has required information about GA for educating domain experts. BetterNet’s aim is to improve the planning of the WDS and it doesn’t tend to be a control system.

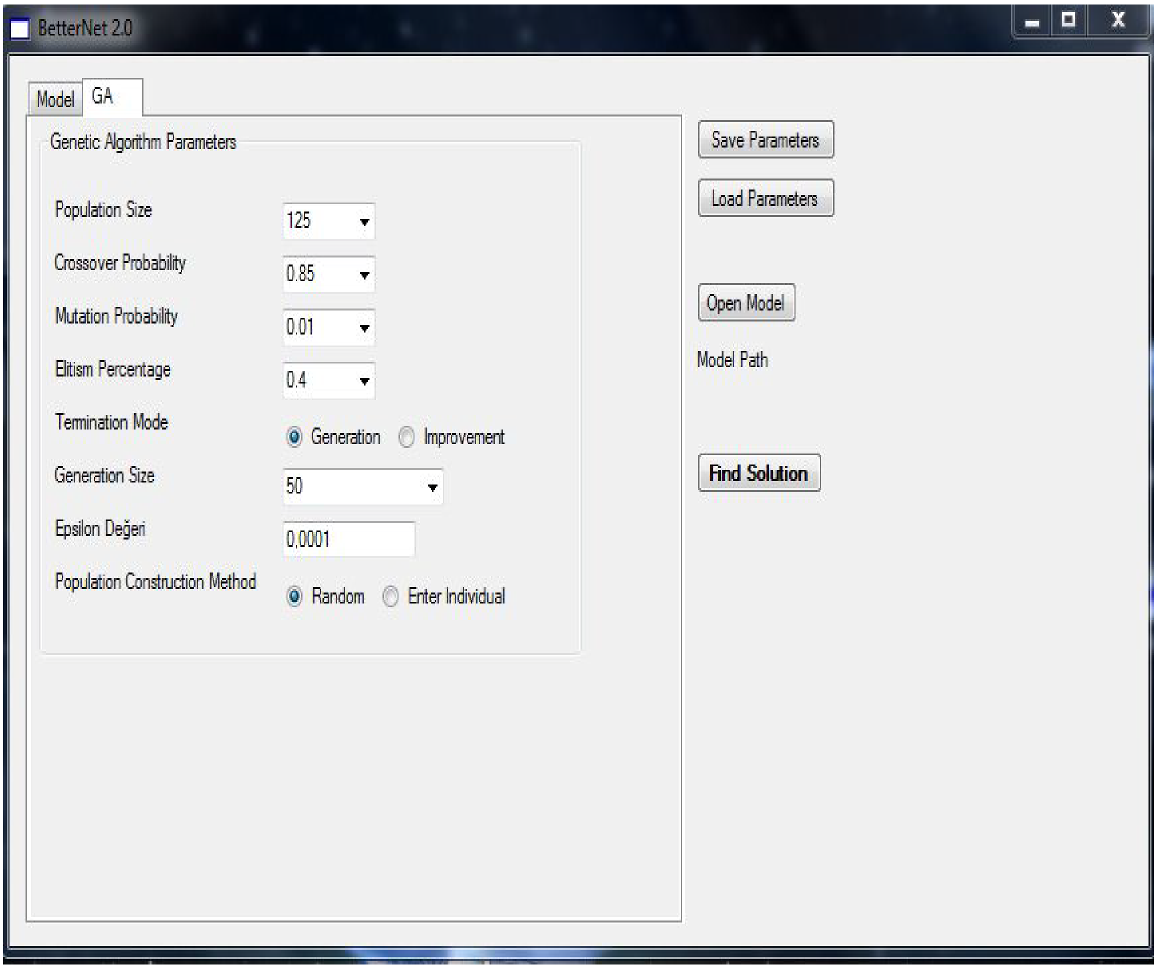
5.2. Tunable Parameters

BetterNet supports two types of tunable system parameters: GA parameters and WDS parameters. User can execute different water distribution network chlorination candidate applications on the model under study by trying alternative scenarios, automatically. All system parameters can be saved for later experimentation and can be reloaded from a file. User can observe/report/pause/continue or stop the search process at any time during execution. The following two subsections introduce capabilities of the BetterNet provided via the tunable system parameters.

5.2.1. GA Application Parameters

There are eight GA parameters/switches that can be set by the user (Figure 5.2) [28]. Population Size (PS) defines the current scope of the search process. Higher PS value increases the probability of better quality individuals to be encountered at the cost of overall processing time. Initial Population (IP) can be set either “randomly” via program or by entering them “manually”. The advantage of manual entry is to allow good candidate solutions decided by the domain expert to appear in the initial search set. The crossover is the basic mechanism for new individuals to be constructed however the risk is to ignore nearby better solutions residing at the search space.

Therefore, BetterNet provides its user a control over the Crossover Probability (CP). Mutation on the other hand is nothing but to permit random modifications over an individual’s description that may lead to possible backward movements in the search space. So, like CP, user can define the Mutation Probability (MP) that facilitates the control over amount of individual’s randomization. Higher MP values turn the so claimed intelligent search process into a random walk over the landscape.

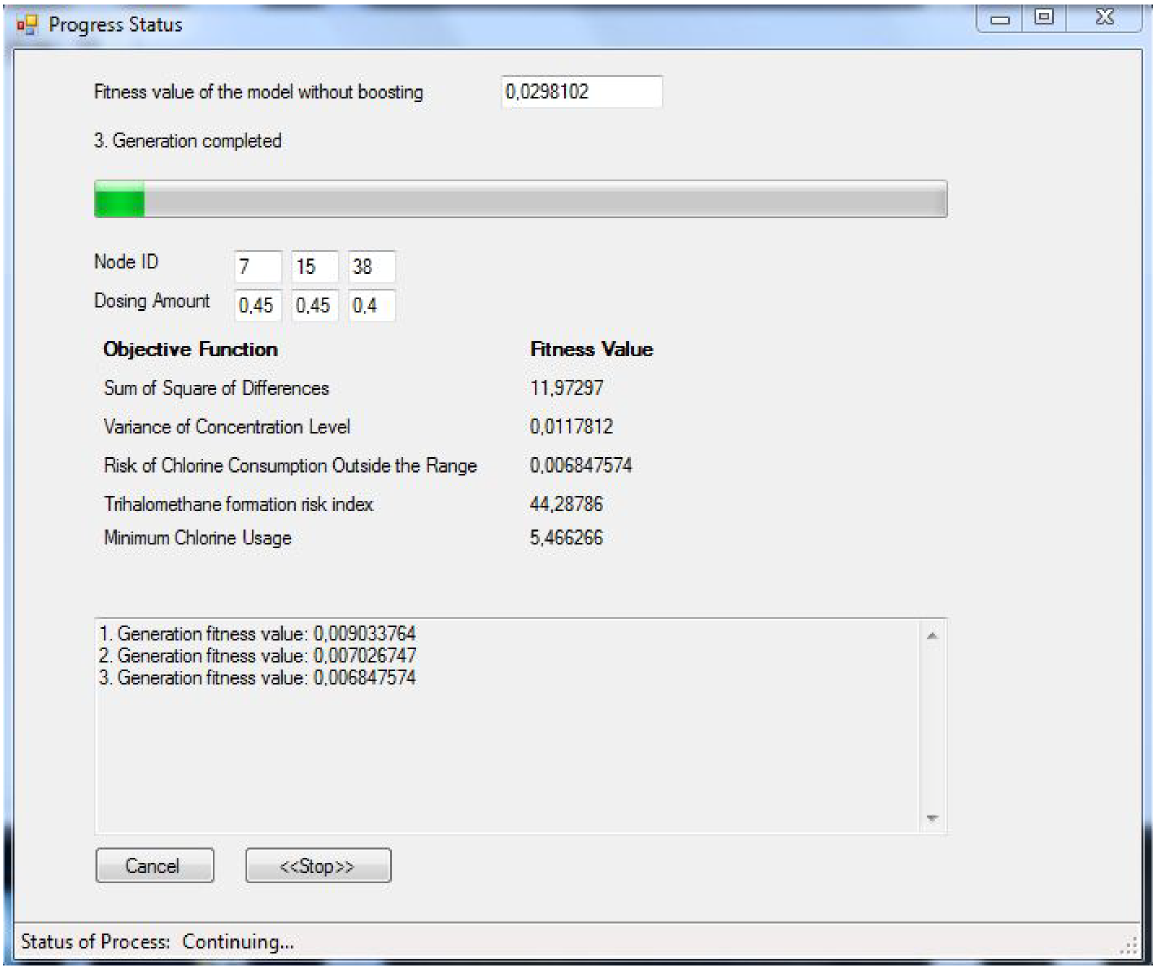


Genetic Algorithm Parameters

Population Size 125
Crossover Probability 0.85
Mutation Probability 0.01
Elitism Percentage 0.4
Temination Mode ● Generation ○ Improvement
Generation Size 50
Epsilon Değeri 0,0001
Population Construction Method ● Random ○ Enter Individual

Figure 5.2 GA parameters for optimization.

Keeping relatively good solutions at hand (or simply in memory) may facilitate the search process and may provide a kind of momentum throughout the search process. The control over its amount is provided by Elitism Percentage (EP) parameter indicating the ratio of the best solutions coming from the previous generation and kept in the current generation. Execution of BetterNet can be visualized and stopped at any time by its user (Figure 5.3). The Stopping Criterion (SC) can either be a Fixed Number of Generations (FNG) set by the user or a small constant threshold value Epsilon (E) showing the amount of improvement in the objective function value. The search process continues as long as the change between current and the previous objective function value is greater than a desirable satisfactory minimum value pre-set by the user.



Fitness value of the model without boosting 0,0298102

3. Generation completed

Node ID	7	15	38
Dosing Amount	0,45	0,45	0,4

Objective Function	Fitness Value
Sum of Square of Differences	11,97297
Variance of Concentration Level	0,0117812
Risk of Chlorine Consumption Outside the Range	0,006847574
Trihalomethane formation risk index	44,28786
Minimum Chlorine Usage	5,466266

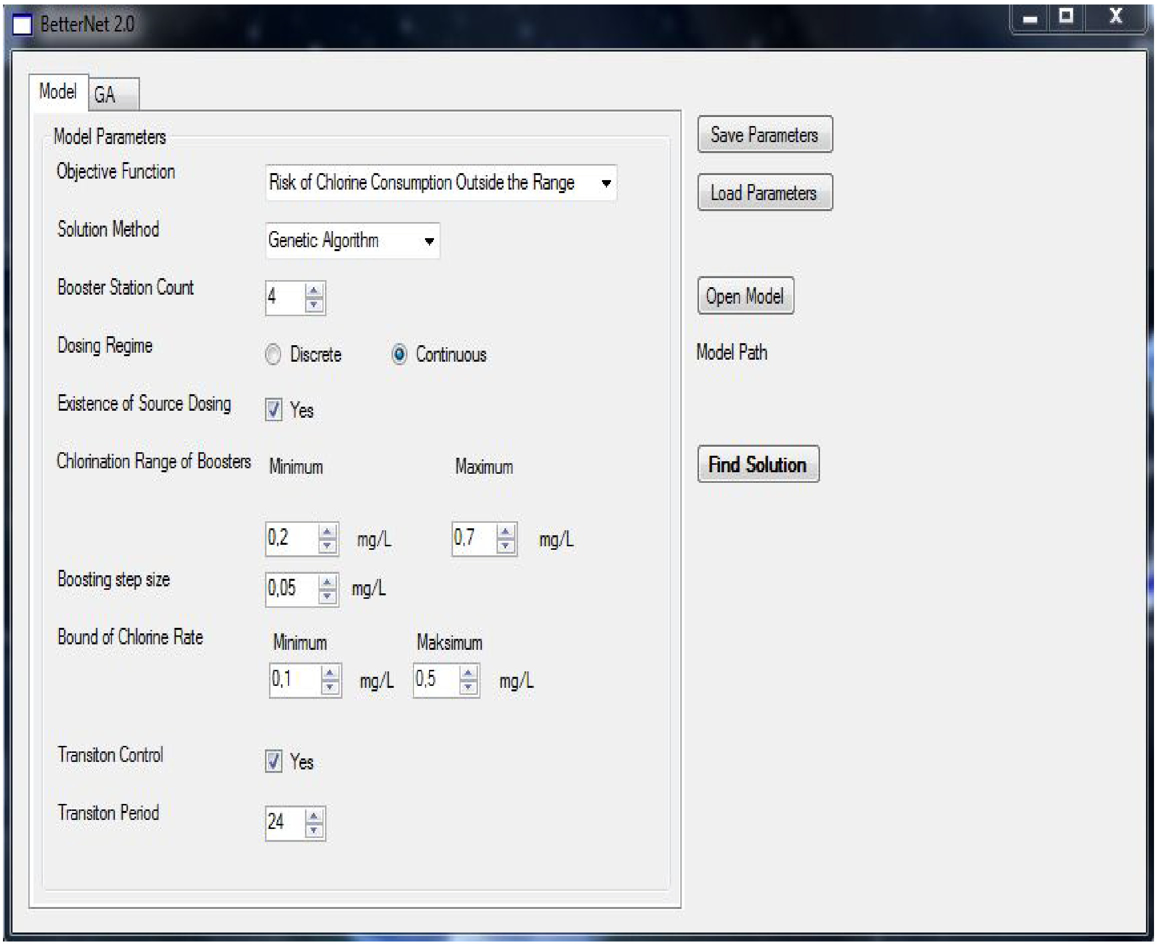
```
1. Generation fitness value: 0,009033764
2. Generation fitness value: 0,007026747
3. Generation fitness value: 0,006847574
```

Status of Process: Continuing...

Figure 5.3 Progress status screen for tracking optimization.

5.2.2. WDS Parameters

There are twelve WDS parameters/switches that can be set by the user (Figure 5.4) [28]. Different from GA parameters, WDS parameters provide domain specific preferences to be specified. User can set any one of five different domain specific objective (or fitness) functions decided by domain expert. Their calculations are mainly based on the amount of feedback concentrations taken from the junction points (i.e. nodes) that are obtained by the end of EPANET software extended period simulation. The upper and lower threshold values for concentrations are allowed to change between 0.20 (c_{min}) and 0.50 (c_{max}) ppm. Note that current version of the BetterNet does not support multi-objective decision making. However, in addition to a selected objective function, user can observe changes of other non-selected objective function values attained throughout the search process. Below, we list the supported fitness/objective functions.



Model Parameters

Objective Function: Risk of Chlorine Consumption Outside the Range

Solution Method: Genetic Algorithm

Booster Station Count: 4

Dosing Regime: ○ Discrete ● Continuous

Existence of Source Dosing: ☑ Yes

Chlorination Range of Boosters: Minimum 0,2 mg/L Maximum 0,7 mg/L

Boosting step size: 0,05 mg/L

Bound of Chlorine Rate: Minimum 0,1 mg/L Maksimum 0,5 mg/L

Transiton Control: ☑ Yes

Transiton Period: 24

Figure 5.4 WDS parameters for optimization.

Chlorination done at common reservoir (or source) site is mostly a natural choice for effective and efficient WDS chlorination. However, existence/nonexistence of “source dosing” throughout the search process is provided as a choice to the user (Existence of Source Dosing - ESD variable). In addition to the source, the number of other chlorination locations can be set via variable Number of Additional Dosing Location (NADL). For an acceptable time performance, the upper limit for NADL is supposed to be 35 locations.

Dosing Regime (DR) can either be applied continuous or discrete time steps. The Range of Additional Chlorine Concentration (RACC) amount can be set by user. The unit of concentration is mg/L. The number of alternative chlorination amount depends on the Step Size (SS) variable decided by user. The target range of FRC in nodes can be determined by entering related minimum and maximum values. As an initial stage of EPS one can specify a transition period (in hours) by the end of which the network is supposed to enter a cyclic time series and becomes stable. The FRC results obtained during transition period should be ignored in objective function calculation.

Getting good results is dependent to selected parameters directly. When population size and number of generation are increased, search space increases, response time of program increases, however probability of finding better results increases also. When they are decreased, probability of finding better results decreases too. When defined mutation probability value is high, then degeneration probability of good individuals increases because of very high amount of mutation. So mutation probability must be selected very small such as 0.01-0.05. When crossover probability is high, breeding a new child probability increases and big part of next generation consists of new individuals. When it decreases, big part of next generation consists of selected parents. It must be selected between 0.8 and 0.95 for better results. If elitism percentage is high, big part of next generation consists of last generation. Otherwise big part of next generation consists of new individuals. It must be selected between 0.4 and 0.6 for better results.

5.3. BetterNet Setup

Software installation is started double clicking the file which is called "Water_Optimization_Setup" [28]. The program will be installed to "C: \ \ Program Files \ \BetterNet 2.0" directory. User can choose another directory for installation if user wants. Your computer must have “.Net Framework 3.5” and Microsoft Office 2007. Otherwise, errors are taken while running the program.

To remove the program from the computer, enter the directory and click "uninstall”.

Some virus program stops the operation of the software. When faced with such a problem, temporarily stop your virus program.

Before starting the optimization work, network files are needed, which have been calibrated in terms of hydraulics and water quality, and whose chlorine concentrations can be accurately predicted [28]. One of the synthetic networks which are used in this study is given below as an example (Figure 5.5). Network .inp file extension is required at the software.

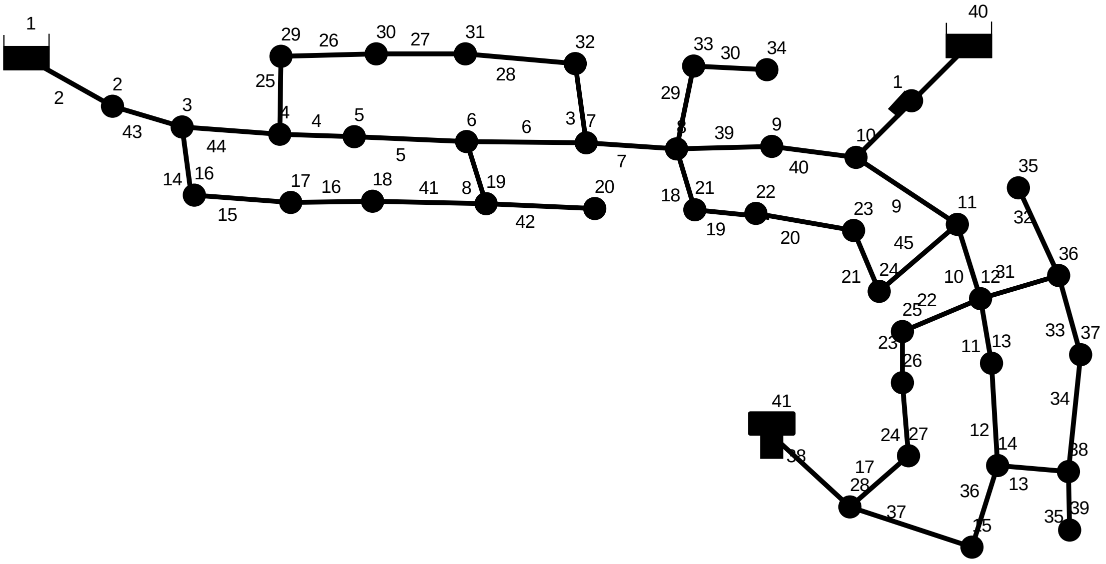


Figure 5.5 Example EPANET model for working on BetterNet.

“Fitness Function”: Using the top left corner, user must choose one of four optimization functions:

- Sum of Square of Differences
- Variance of Concentration Level
- Risk of Chlorine Consumption outside the Range
- Trihalomethane Risk Index

“Optimization Method”: User must enter an optimization method for a solution. Optimization methods can be selected as follows:

- "Exhaustive Search”: This method tries to find all possible outcomes. Choosing this method, up to three additional stations to find the chlorination is recommended. Additional chlorine number of stations increases, the problem is growing search space will take a long time to reach a solution. A very simple network can also be solved with exhaustive search but user must select chlorine boundary simply.

- "Genetic Algorithms": With this selection method, genetic algorithm optimization process is carried out according to the user-defined parameters.

- "Domain Aware Genetic Algorithms ": If choosing this method, genetic algorithm operates according to domain aware crossover.

"Number of Additional Dosing Location": The user enters how many additional chlorination stations are wanted in the network to here. Additional chlorination station number is fixed at the upper limit of 35.

"The range of additional chlorine concentration": Lowest (minimum) and highest (maximum) concentration of chlorine must be entered, which can be applied in the additional chlorination stations.

"Step size": The difference between (step size) concentrations must be specified during the investigation of the most appropriate chlorine levels will be applied.

"Network stability control": EPANET software operates on a network solution for the EPS and the conversion of the results of repeated time series is gained after a certain period of transition is beginning. Transition time (warm-up time) must be specified by user. If the transition period is defined, in that case any assessment for that period is not done.

"GA" tab is clicked for entering Genetic Algorithm parameters.

"Population Size": It represents the number of individuals that will be used in Genetic Algorithm. Search space is increased by increasing this parameter for getting better solution.

"Crossover Probability": It is the probability of crossover of selected two individuals.

"Mutation Probability": It is the probability of that an individual will be mutated.

“Elitism Percentage": When the new generation is determined, user specified percentage of the value is taken into account between parents and children selected. For example, if elitism percentage is 0.4, 40% of the best individuals of previous generation are taken and 60% of the new generation’s the best individuals are selected when new generation is created.

“Termination Mode ": Optimization work can be finished in two different ways: after a fixed number of repeats or when difference between two consecutive best solutions is less than a particular value.

- "Fixed number of generations": The user should enter a generation number for the Genetic Algorithm. Optimization continues until the entered generation number and genetic algorithm is terminated.

- "Improvement based program termination" (ε): User enters an expected improvement rate (ε) between two consecutive generations and the genetic algorithm terminates when rate between consecutive generations is smaller than entered rate.

"Population Creation Options": Initial population can be constructed with two different ways. Firstly, user enters population size value to the software. User may want that individuals of initial population are composed randomly by the software. Alternatively, user may want to enter some individuals, which consist of chlorination locations and amounts of chlorination, to population on the software and if there is any empty slot on population, the software completes it randomly.

After entering the parameters "Find Solution" button is clicked to start the optimization work and the results are reported in "solution.txt" file for the user.

"Saving parameters": All parameters entered by the user can be saved by "Save Parameters" option in the program. So that the user doesn’t need to enter a large number of parameters again when saved parameters will be used again. Parameters are saved using ".PRM" extension. "Parameter Load" option should be used to download saved information from an existing parameter file.

After optimization parameters have been entered or loaded, network model is introduced to software by "Model Select" command. Network model with ".INP" file extension should be introduced to the software. "Find Solution" button is used to start the optimization process.

"Progress Screen” and "Progress Tracking Chart" has been developed for users to follow the results during GA process. Also, options of being able stop the ongoing process and cancelling the process are developed. Process output report is produced, which called "Solution.txt" (see Appendix A) and contains information on the results. The Program composes the solution model with “.INP” extension which can be imported to EPANET, thus user can see the solution as an EPANET model. The program represents the generation-fitness value graph to user also (Appendix B).

CHAPTER 6

RESULTS AND DISCUSSIONS

In order to compare GA and IGA, we developed independent optimization studies based on supported objective functions.

6.1. The Characteristics of the Networks Under Study

6.1.1. Test Network 1 (Hypothetical Network)

A hypothetical network (Figure 6.1) has been designed to exploit the optimization capabilities of BetterNet.

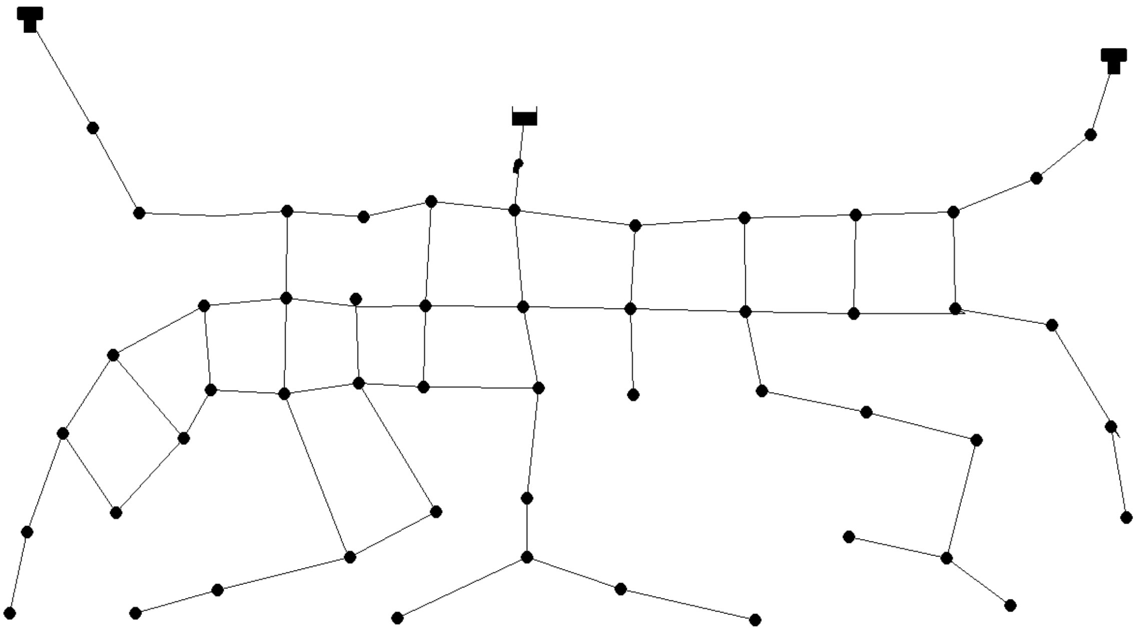


Figure 6.1 Test Network 1 with 51 nodes, 2 tanks and 1 reservoir.

Table 6.1 summarizes the major physical characteristics of the network under study. The designed network is considerably complex with 1 supply reservoirs with source chlorination and there is a pump in the system. It contains 2 tanks. There is a balancing reservoir. Total of pipe lengths is 39. 1 km.

Number of Nodes	Number of Source Reservoirs	Number of Tanks	Number of Pumps	Number of Pipes	Total of Pipe Lengths (km)	Total Base Nodal Demand (cmh)	Diameter Range (mm)
51	1	2	1	64	39. 1	759	200-500

Table 6.1 The major physical characteristics of the Test Network 1.

The assumed diurnal demand pattern is periodic with 24 hour intervals and summarized by Figure 6 2. The hydraulic and decay kinetics are assumed to be pre-calibrated in order to simplify the presentation. The supply reservoirs feed chlorine with a constant concentration level of 0.5 ppm for the four objective functions.

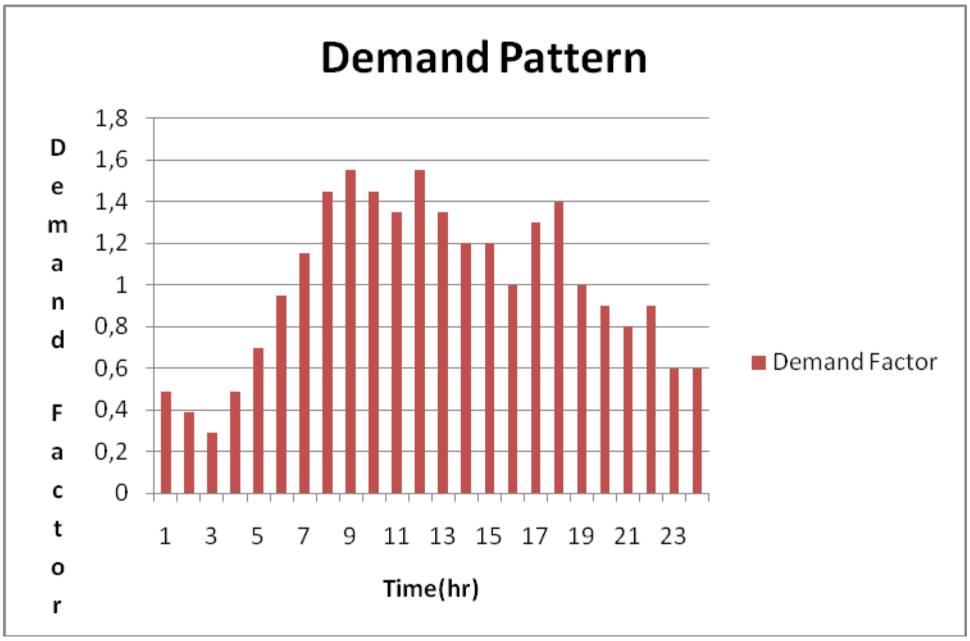


Figure 6.2 Demand pattern for Test Network 1.

6.1.2. Test Network 2 (Hypothetical Network)

A hypothetical network (Figure 6.3) has been designed to exploit the optimization capabilities of BetterNet.

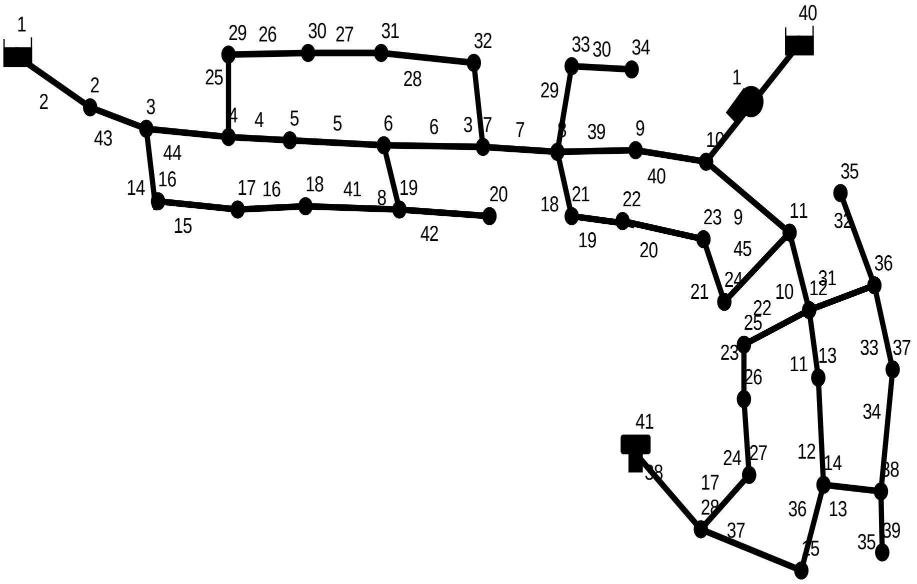


Figure 6.3 Test Network 2 with 39 nodes, 1 tank and 2 reservoirs.

Table 6.2 summarizes the major physical characteristics of the network under study. The designed network is considerably complex with 2 supply reservoirs with source chlorination and there is a pump in the system. There is a balancing reservoir. Total of pipe lengths is 43.2 km.

Number of Nodes	Number of Source Reservoirs	Number of Tanks	Number of Pumps	Number of Pipes	Total of Pipe Lengths (km)	Total Base Nodal Demand (cmh)	Diameter Range (mm)
39	2	1	1	45	43. 2	754	200-400

Table 6.2 The major physical characteristics of the Test Network 2.

The assumed diurnal demand pattern is periodic with 24 hour intervals and summarized by Figure 6.4. The hydraulic and decay kinetics are assumed to be pre-calibrated in order to simplify the presentation. The supply reservoirs feed chlorine with a constant concentration level of 0.5 ppm for the four objective functions.

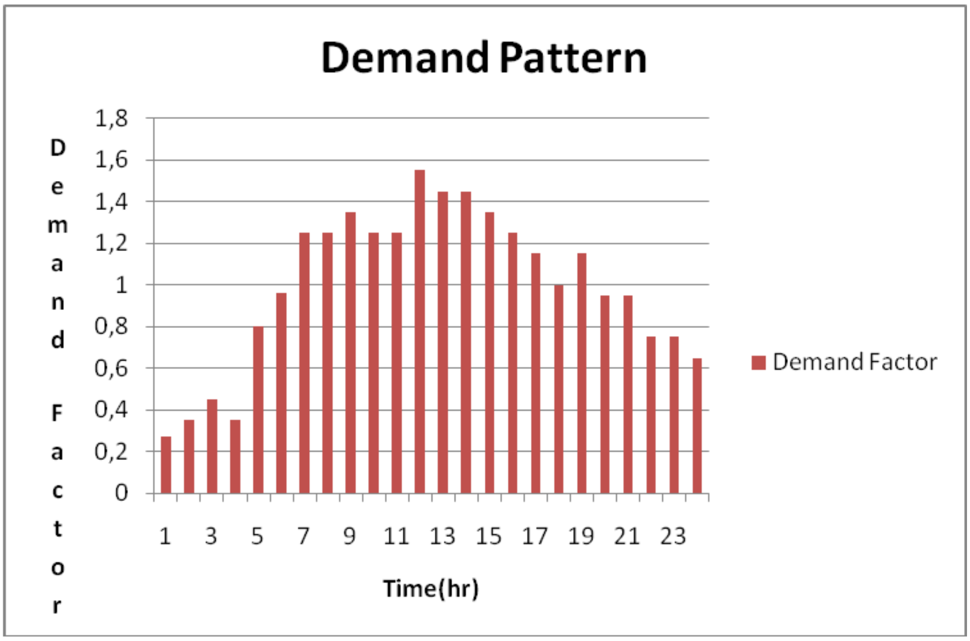


Figure 6.4 Demand Pattern for Test Network 2.

6.1.3. Zone 4 (Real Network)

A real network (Figure 6.5) has been designed from real water distribution system to exploit the optimization capabilities of BetterNet. This real network (Zone 4) is a part of “Konyaaltı” which is a district at the west part of “Antalya” [43]. This pilot project area “Konyaaltı” is divided into 20 sub-region (zone) in the context of TÜBİTAK-KAMAG project which is managed by Prof. Dr. Habib MUHAMMETOĞLU (see Appendix D).

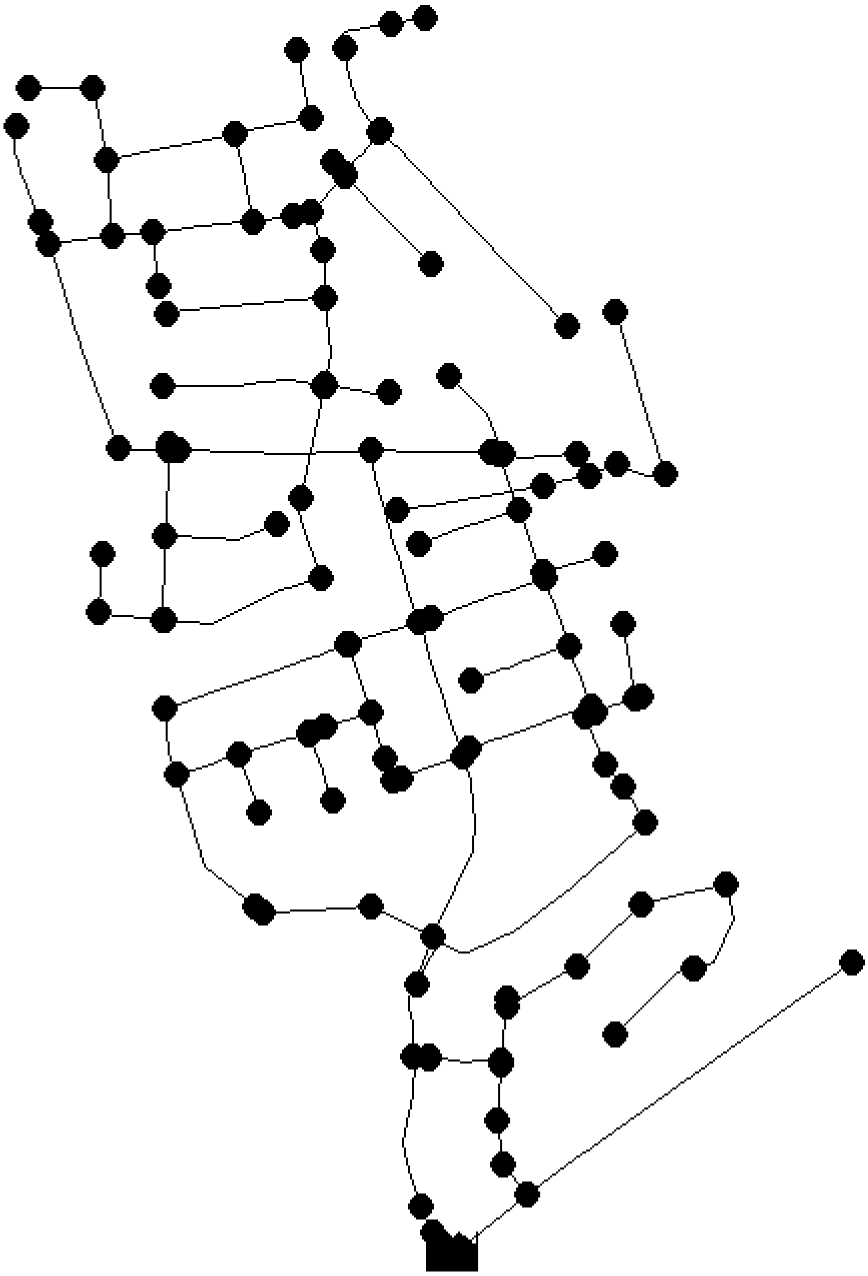


Figure 6.5 Zone 4 with 126 nodes and 1 reservoir.

Table 6.3 summarizes the major physical characteristics of the network under study. The designed network is considerably complex with 1 supply reservoirs with source chlorination and there is a pump in the system. There is a balancing reservoir. Total of pipe lengths is 85,8419 km.

Number of Nodes	Number of Source Reservoirs	Number of Tanks	Number of Pumps	Number of Pipes	Total of Pipe Lengths (km)	Total Base Nodal Demand (cmh)	Diameter Range (mm)
126	1		1	128	85,8419	57,93859	63-150

Table 6.3 The major physical characteristics of the Zone 4.

The assumed diurnal demand pattern is periodic with 48 hour intervals and summarized by Figure 6.6. The hydraulic and decay kinetics are calibrated. The supply reservoirs feed chlorine with a discrete concentration level for the four objective functions (see Figure 6.7).

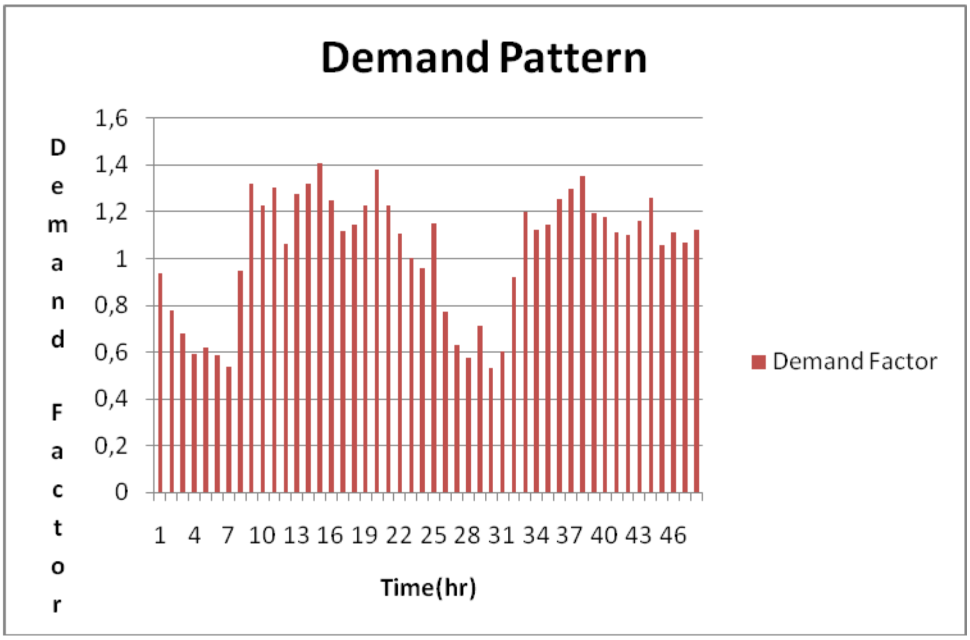


Figure 6.6 Demand pattern for Zone 4.

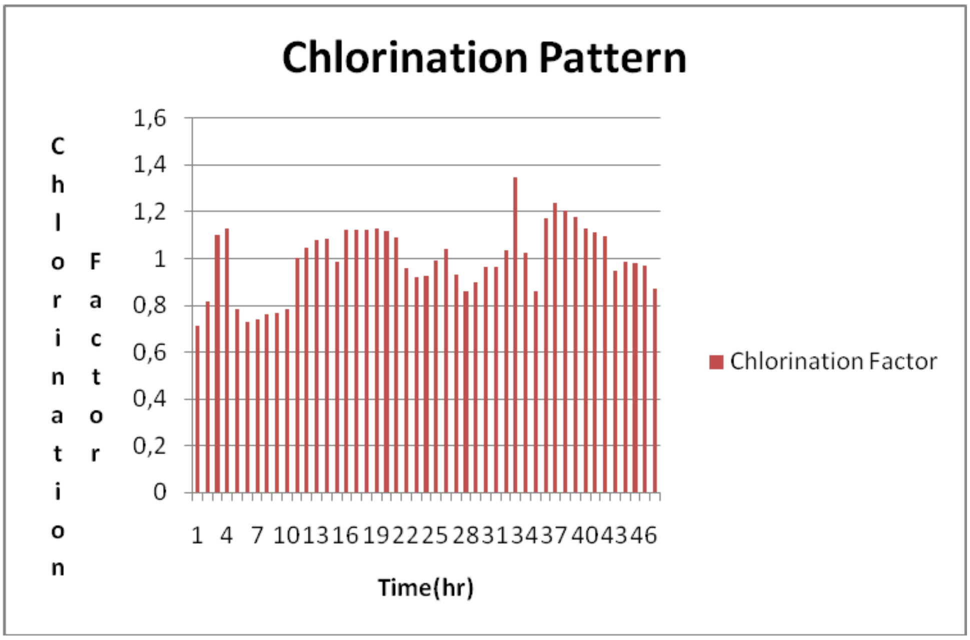


Figure 6.7 Chlorination pattern for Zone 4.

6.2. Input Variables and Genetic Algorithm Parameters

In order to see the improvement of new technique, 3 separate categories of optimization study, have been performed. Each study category corresponds to an individual model. Chlorine boosting range was set to 0.2-0.5 ppm. The number of boosters was selected to be 4, 6 and 8 for each study category. An incremental increase of 0.05 ppm was utilized in finding out the desirable boosting level. Minimum and maximum allowable FRC concentrations were set to be 0.2 and 0.5 ppm respectively. The transition stage was set as 24 hours. An incremental change of 1 % between subsequent iterations was used as stopping criteria for any search. Throughout the experiments, PS is set to 50 individuals and initial population is constituted, randomly. CP and MP values are fixed to 0.85 and 0.05. CP value was selected as 0.85, 0.90, 0.95 and EP is taken as 0.4, 0.5 and 0.6.

Model	TestNetwork1, TestNetwork2, Zone4
Method	GA, IGA
Objective Function	O1, O2, O3, O4
Crossover Probability	0.85, 0.90, 0.95
Elitism Percentage	0.4, 0.5, 0.6
Location Count	4, 6, 8
Population Size	50
Err	0.01
Mutation Probability	0.05
Concentration Range	0.2-0.5
Concentration Step	0.05
Chlorine Range	0.2-0.5
Transition Period	24

Table 6.4 Parameters for BetterNet runs.

6.3. Comparison of GA and IGA

The results of the optimizations are summarized in tables in which the values for the magnitudes of objective functions when 4, 6 or 8 boosters are employed under each objective function activation in three different WDS model (2 hypothetical, 1 real).

Table 6.5 shows the solutions for objective function 1 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 4 for the given runs. When we look at the given graph (Figure 6.8), we can see that IGA has a little improvement for testnetwork1 and zone4 but GA works beter for testnetwork2.

Model	GA	IGA
testnetwork1	22,903149	22,863474
testnetwork2	23,024158	30,044703
zone4	37,868840	37,248710

Table 6.5 Objective values of Objective 1 for CP=0.9, EP=0.6, 4 Nodes.

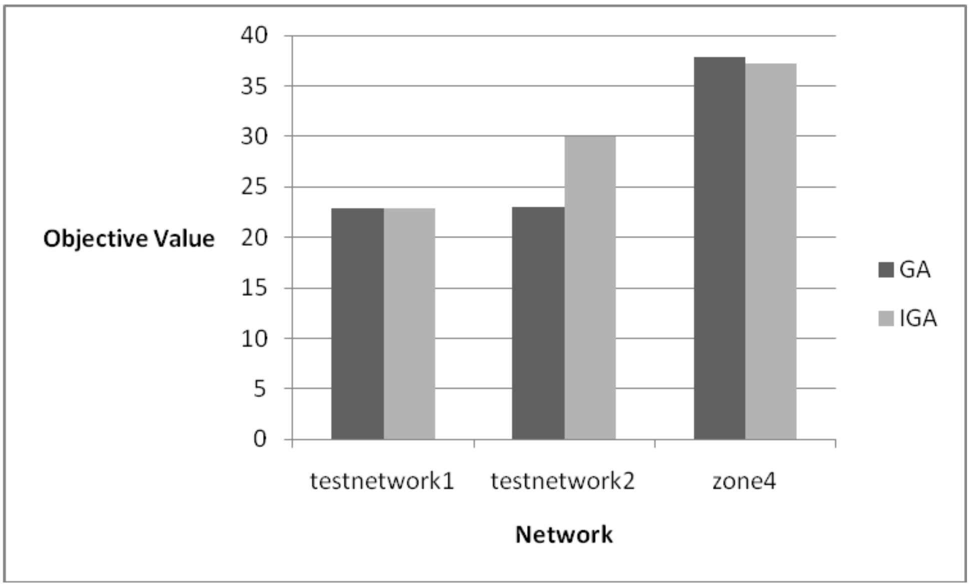


Figure 6.8 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 4 Nodes.

Table 6.6 shows the solutions for objective function 1 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 6 for the given runs. When we look at the given graph (Figure 6.9), we can see that IGA has a little improvement for testnetwork1 and zone4 but GA works beter for testnetwork2.

Model	GA	IGA
testnetwork1	17,229467	16,892658
testnetwork2	21,962868	25,931795
zone4	4,066068	3,725175

Table 6.6 Objective values of Objective 1 for CP=0.9, EP=0.6, 6 Nodes.

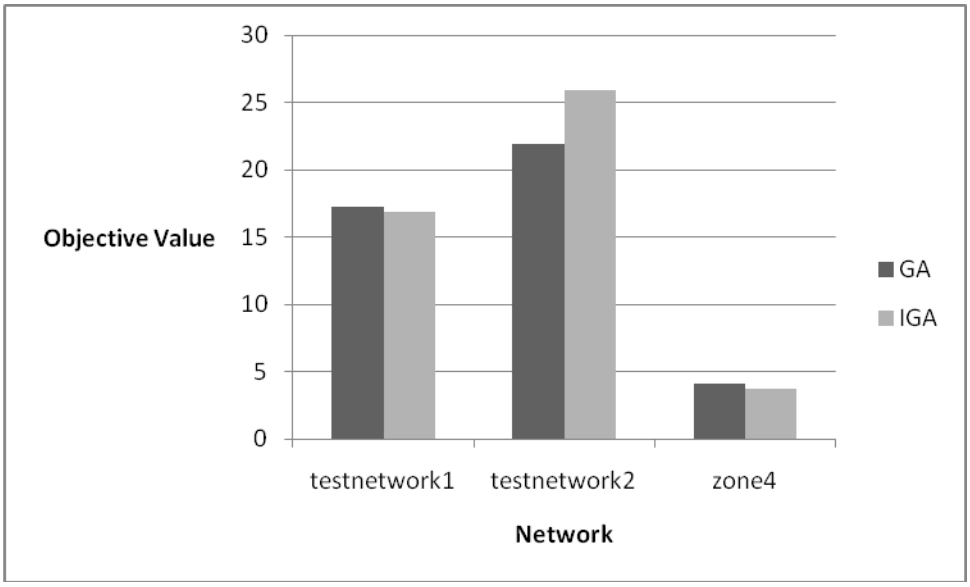


Figure 6.9 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 6 Nodes.

Table 6.7 shows the solutions for objective function 1 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 8 for the given runs. When we look at the given graph (Figure 6.10), we can see that IGA has a few unit improvements for testnetwork1 and testnetwork2, but GA and IGA finds closed solutions for zone4.

Model	GA	IGA
testnetwork1	16,563011	14,417663
testnetwork2	19,019035	17,250536
zone4	3,829691	3,908577

Table 6.7 Objective values of Objective 1 for CP=0.9, EP=0.6,8 Nodes.

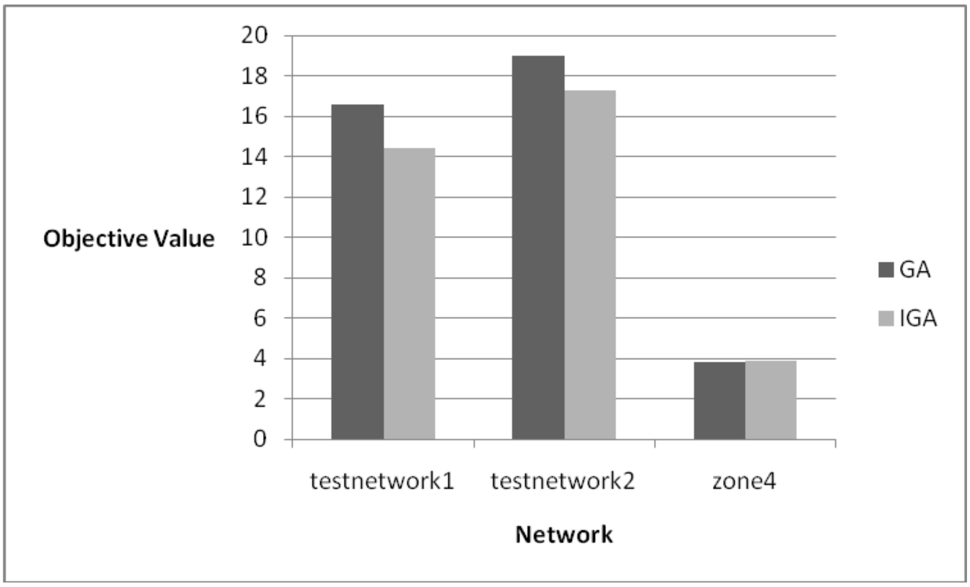


Figure 6.10 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 1, 8 Nodes.

Table 6.8 shows the solutions for objective function 2 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 4 for the given runs. When we look at the given graph (Figure 6.11), we can see that IGA improvements in solutions for testnetwork1, testnetwork2 and zone4.

Model	GA	IGA
testnetwork1	0,018348	0,017593
testnetwork2	0,033951	0,030838
zone4	0,001265	0,001232

Table 6.8 Objective values of Objective 2 for CP=0.9, EP=0.6, 4 Nodes.

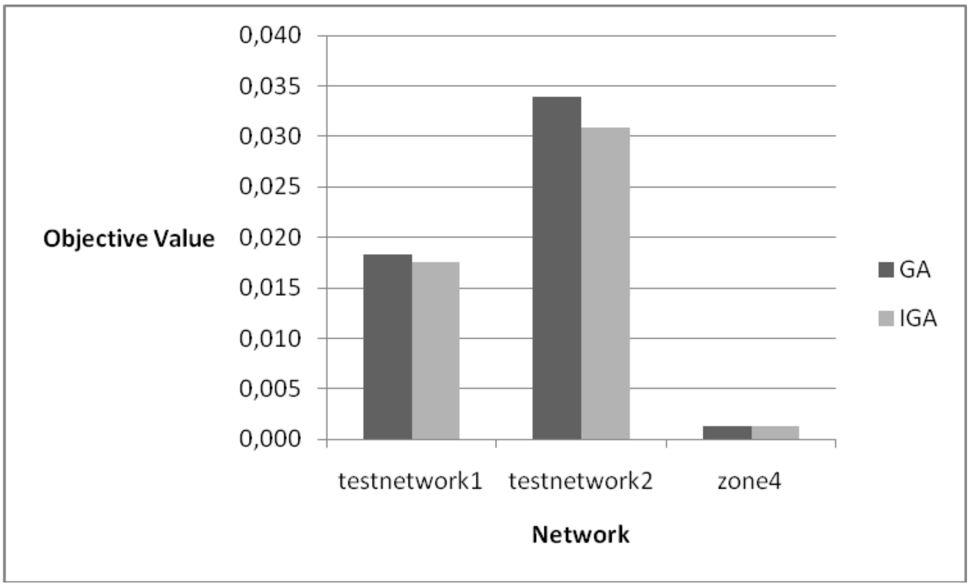


Figure 6.11 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 4 Nodes.

Table 6.9 shows the solutions for objective function 2 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 6 for the given runs. When we look at the given graph (Figure 6.12), we can see that IGA some improvements for testnetwork1 and testnetwork2, but GA and IGA find closed solutions for zone4.

Model	GA	IGA
testnetwork1	0,019852	0,014474
testnetwork2	0,022143	0,021738
zone4	0,001249	0,001265

Table 6.9 Objective values of Objective 2 for CP=0.9, EP=0.6, 6 Nodes.

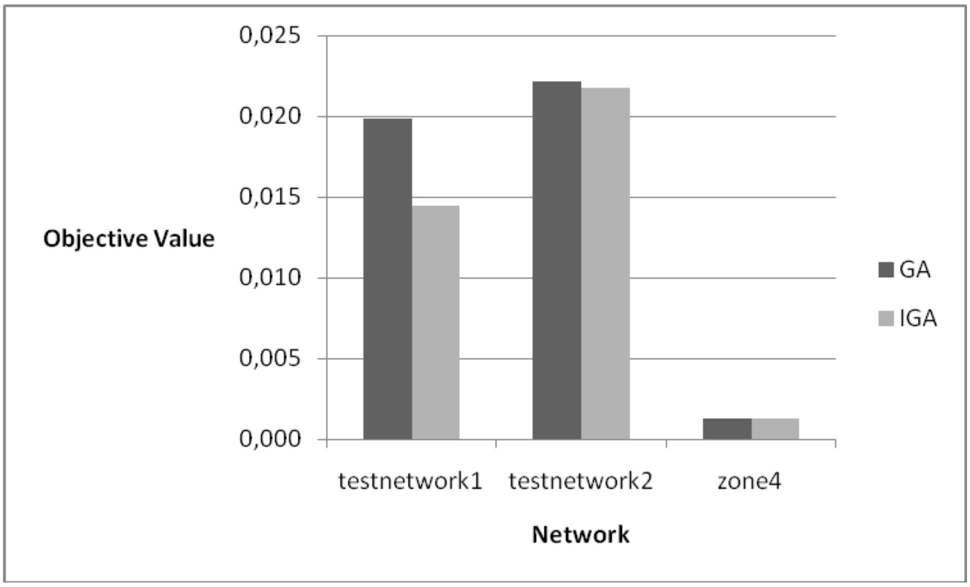


Figure 6.12 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 6 Nodes.

Table 6.10 shows the solutions for objective function 2 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 8 for the given runs. When we look at the given graph (Figure 6.13), we can see that IGA some improvements for testnetwork1 and testnetwork2, but GA and IGA find closed solutions for zone4.

Model	GA	IGA
testnetwork1	0,014345	0,010915
testnetwork2	0,021645	0,020023
zone4	0,001247	0,001323

Table 6.10 Objective values of Objective 2 for CP=0.9, EP=0.6, 8 Nodes.

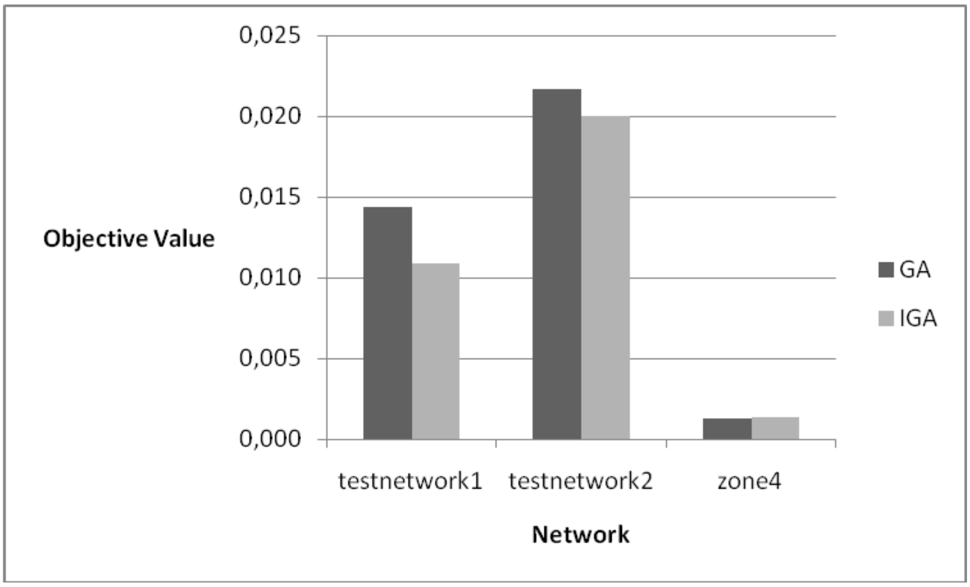


Figure 6.13 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 2, 8 Nodes.

Table 6.11 shows the solutions for objective function 3 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 4 for the given runs. When we look at the given graph(Figure 6.14), we can see that IGA and GA find 0 value (0 risk) for testnetwork1 and IGA finds 0 for testnetwork2, GA and IGA find 0 value for zone4 but in zone4 risk is already 0 without boosting chlorination.

Model	GA	IGA
testnetwork1	0,000000	0,000000
testnetwork2	0,045222	0,000000
zone4	0,000000	0,000000

Table 6.11 Objective values of Objective 3 for CP=0.9, EP=0.6, 4 Nodes.

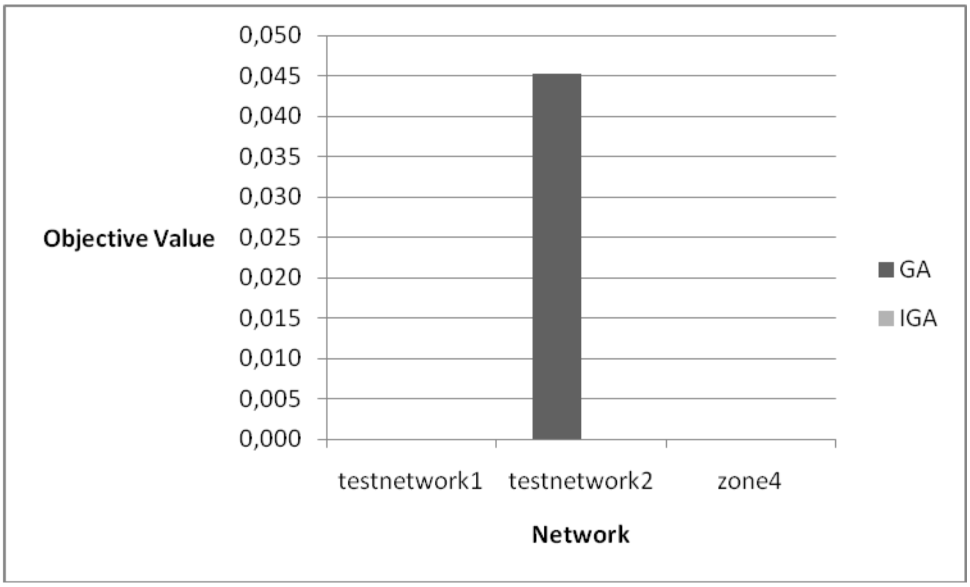


Figure 6.14 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 4 Nodes.

Table 6.12 shows the solutions for objective function 3 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 6 for the given runs. When we look at the given graph (Figure 6.15), we can see that IGA find 0-value (0-risk) for testnetwork1, testnetwork2 and zone4. In zone4, risk is already 0 without boosting chlorination. GA cannot find the 0-value.

Model	GA	IGA
testnetwork1	0,006970	0,000000
testnetwork2	0,046026	0,000000
zone4	0,000000	0,000000

Table 6.12 Objective values of Objective 3 for CP=0.9, EP=0.6, 6 Nodes.

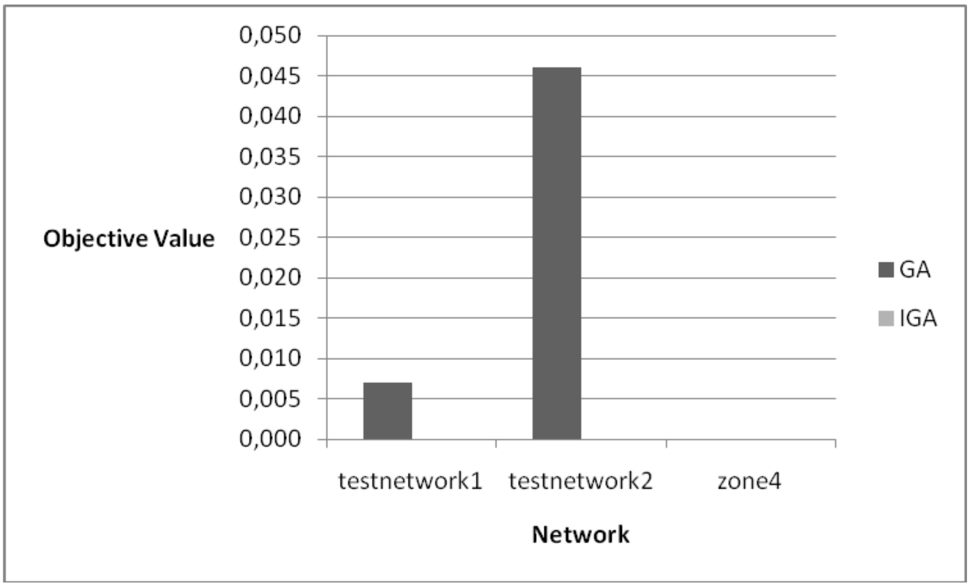


Figure 6.15 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 6 Nodes.

Table 6.13 shows the solutions for objective function 3 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 8 for the given runs. When we look at the given graph (Figure 6.16), we can see that IGA find 0-value (0-risk) for testnetwork1, and find a very little value for testnetwork2. In zone4, risk is already 0 without boosting chlorination. GA cannot find the 0-value.

Model	GA	IGA
testnetwork1	0,038255	0,000000
testnetwork2	0,048917	0,000304
zone4	0,000000	0,000000

Table 6.13 Objective values of Objective 3 for CP=0.9, EP=0.6, 8 Nodes.

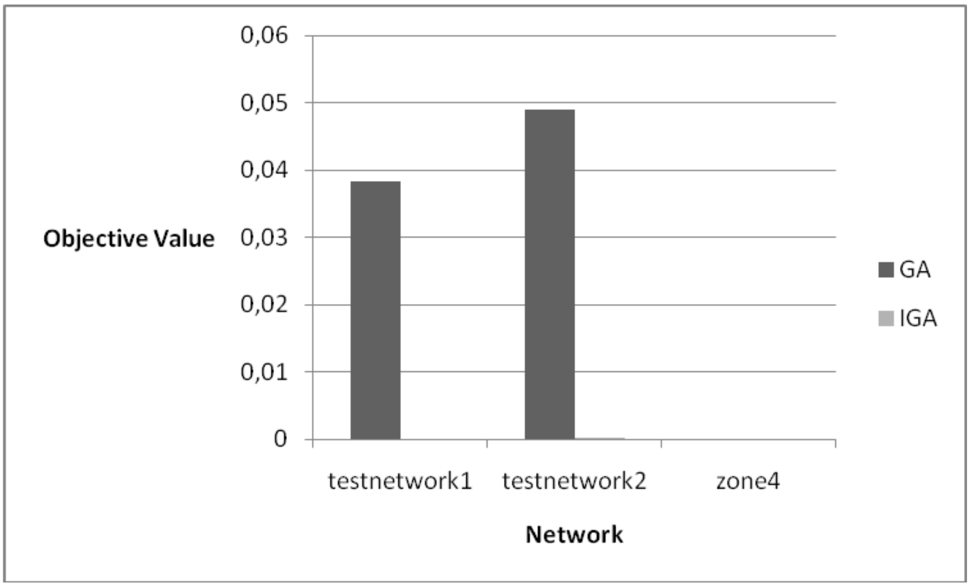


Figure 6.16 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 3, 8 Nodes.

Table 6.14 shows the solutions for objective function 4 which found by GA and IGA for three models. CP is 0.9, EP is 0.6, number of boosters is 4 for the given runs. When we look at the given graph (Figure 6.17), we can see that IGA finds better solution for testnetwork1, and GA finds better solution for testnetwork2 and also zone4.

Model	GA	IGA
testnetwork1	6,810511	5,580457
testnetwork2	5,790916	7,876226
zone4	2,367264	4,052771

Table 6.14 Objective values of Objective 4 for CP=0.9, EP=0.6, 4 Nodes.

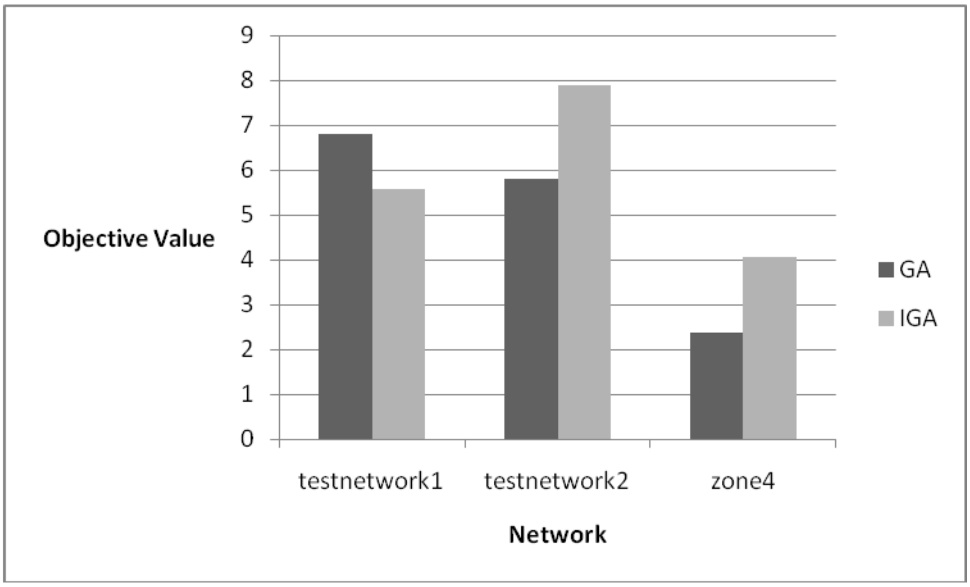


Figure 6.17 Comparison of GA & IGA for CP=0.9, EP=0.6, Objective 4, 4 Nodes.

Table 6.15 shows the solutions for objective function 4 which found by GA and IGA for three models. CP is 0.9, EP is 0.5, number of boosters is 6 for the given runs. When we look at the given graph (Figure 6.18), we can see that IGA finds better solution for all of three networks.

Model	GA	IGA
testnetwork1	7,539024	5,522160
testnetwork2	6,297567	4,869606
zone4	3,228049	0,229485

Table 6.15 Objective values of Objective 4 CP=0.9, EP=0.5, 6 Nodes.

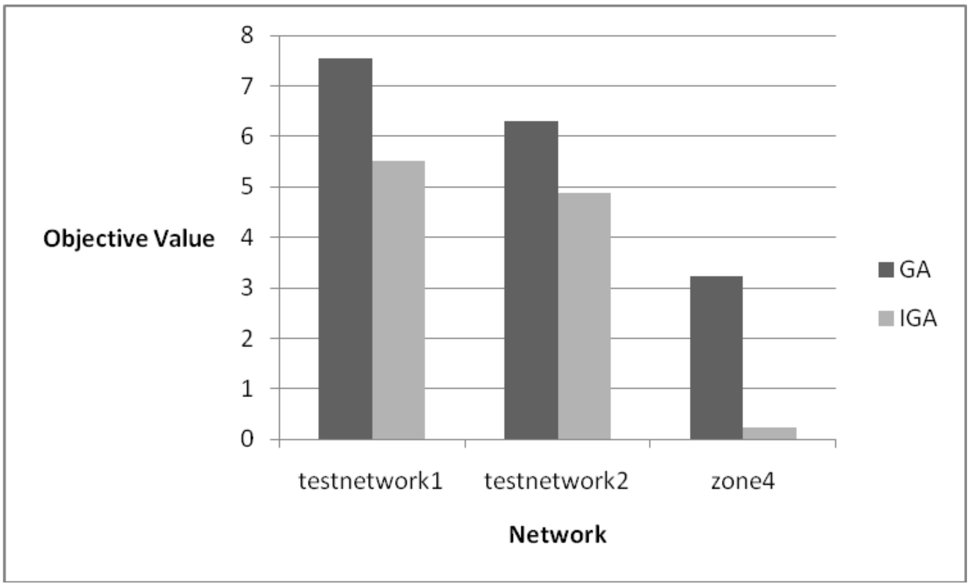


Figure 6.18 Comparison of GA & IGA for CP=0.9, EP=0.5, Objective 4, 6 Nodes.

Table … shows the solutions for objective function 4 which found by GA and IGA for three models. CP is 0.85, EP is 0.4, number of boosters is 8 for the given runs. When we look at the given graph (Figure 6.19), we can see that IGA finds better solution for all of three networks.

Model	GA	IGA
testnetwork1	6,632370	5,669595
testnetwork2	7,071061	6,133202
zone4	2,938426	0,449372

Table 6.16 Objective values of Objective 4 CP=0.85, EP=0.4, 8 Nodes.

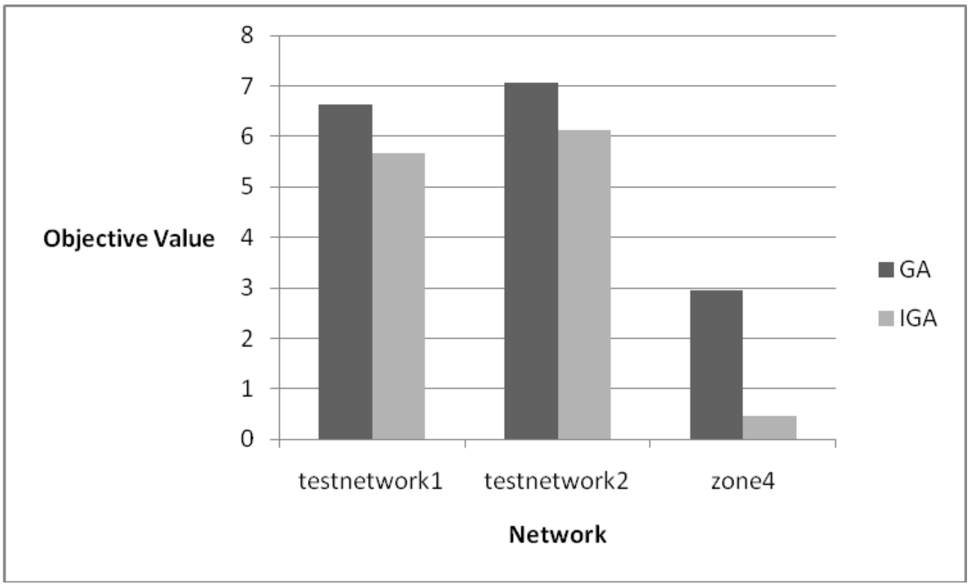


Figure 6.19 Comparison of GA & IGA for CP=0.85, EP=0.4, Objective 4, 8 Nodes.

After looking at each objective value under the given parameters, we can say that IGA improves the solution for many scenarios. In a few scenarios, GA optimizes solution more than IGA. Generally, IGA finds better solution or closed solution which is found by GA.

Risk of exposure is already zero for real network zone 4. Hence, there is no comparison for objective function 3 in zone 4. Zone 4 is a good network with good fitness without any boosting chlorination so genetic algorithm doesn’t do any more improvement on the model.

When we look at figure 6.7, 6.8 and 6.16, IGA didn’t produce better solution than classical GA. Initial population is produced randomly, thus GA may have better solutions at the beginning and GA may produce better solutions than IGA sometimes because of randomness. However, IGA is dominant to produce better solution when many comparisons are made between them.

Appendix E contains EPANET model with solution of GA and IGA for Test Network 1. Also, impacts on chlorine improvement of these methods are shown when booster stations are added to network.

CHAPTER 7

CONCLUSIONS

Decision support software for water distribution system chlorination is developed. The software can be used to answer basic domain questions about: the number, location and concentration for booster stations; keeping chlorine utilization at minimum levels in order to minimize system specific trihalomethane formation risk index and reducing the chlorine levels to lower values at the exit of main sources. As a case study, the software is used by domain expert on a hypothetical and real water distribution network, and remarkable results are taken. In practice, using BetterNet improves and speeds up WDS design process. More importantly, it reduces possible unnecessary expenditures due to wrong decisions about boosting station number, location and amount of chlorine to be used. Note that using BetterNet requires a domain expertise in water distribution network design and analysis. Besides from basic knowledge of GA based optimization, the user is still supposed to be able to do hydraulics modeling, calibration and water quality modeling.

GA does not guarantee a global optima and it's very possible to get stuck in local optima. The GA explores a very large search space (i.e., one that you could not possibly explore the entire thing) and it can lose some good solution with respect to GA parameters. Some of domain information can be added to GA and it can be improved for avoiding stuck in local minimum. In this thesis, risk value is used in order to avoid stuck in local minimum problem and better results than classical GA application are obtained.

With this thesis, a real GA application was implemented and the applicability of the method was proved in the field. A study was carried out professionally with real customers. In the context of software development, experience of professional tools using was gained from analysis to deployment. Collective research with researchers from different disciplines was experienced. Experience of working with public agency staff was gained.

REFERENCES

[1] Vasconcelos, J. J., Rossman, L. A., Grayman, W. M., Boulos, P. F., and Clark, R. M. (1997). ‘‘Kinetics of chlorine decay.’’ J. Am. Water Works Assoc., 89(7), 55–65.

[2] Tryby, M. E., Boccelli, D. L., Koechling, M. T., Uber, J. G., Summers, R. S., and Rossman, L. A. (1999). ‘‘Booster chlorination for managing disinfectant residuals.’’ J. Am. Water Works Assoc., 91(1), 95–108.

[3] Islam, M. R., Chaudhary, M. H., and Clark, R. M. (1997). ‘‘Inverse modeling of chlorine concentration in pipe networks under dynamic condition.’’ J. Environ. Eng., 123 (10), 1033 1040.

[4] Al-Omari, A. S., and Chaudhry, M. H. (2001). ‘‘Unsteady-state inverse chlorine modeling in pipe networks.’’ J. Hydraul. Eng., 127(8), 669–677.

[5] Rossman, L. A., and Boulos, P. F. (1996). ‘‘Numerical methods for water quality in distribution systems: A comparison.’’ J. Water Resour. Plan. Manage., 122(2), 137–146.

[6] Boccelli, D. L., Tryby, M. E., Uber, J. G., Rossman, L. A., Zierolf, M. L., Polycarpou, M. M. (1998). “Optimal scheduling of booster disinfection in water distribution systems.” Journal of Water Resources Planning and Management, 124(2), 99-111.

[7] Clark, R. M. (1998). ‘‘Chlorine demand and TTHM formation kinetics.’’ J. Environ. Eng., 124(1), 16–24.

[8] Clark, R. M., and Sivaganesan, M. (1998). ‘‘Predicting chlorine residuals and the formation of TTHMs in drinking water.’’ J. Environ. Eng., 124(12), 1203–1210.

[9] Clark, R. M., and Sivaganesan, M. (2002). ‘‘Predicting chlorine residuals in drinking water: Second order model.’’ J. Water Resour. Plan. Manage., 128(2), 152–161.

[10] Fang, H., West, J. R., Barker, R. A., and Forster, C. F. (1999). ‘‘Modelling of chlorine decay in municipal water supplies.’’ Water Res., 33(12), 2735–2746.

[11] Munavalli, G. R., and Mohan Kumar, M. S. (2002). ‘‘Water quality parameter estimation in a steady state distribution system.’’ J. WaterResour. Plan. Manage., in press.

[12] Savic, A., Walters, G. A. (1997). “Genetic algorithms for least-cost design of water distribution networks.” Journal of Water Resources Planning and Management, ASCE 123 (2), 67-77.

[13] Behzadian, K., Kapelan, Z., Savic, D., Ardeshir, A. (2009). “Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks.” Environmental Modelling & Software 24, 530-541.

[14] Chu, C. W., Lin, M. D., Liu, G. F., Sung, Y. H. (2008). “Application of immune algorithms on solving minimum-cost problem of water distribution network.” Mathematical and Computer Modelling, vol. 48, iss. 11-12, 1888-1900.

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

[16] Zecchin, A. C., Simpson, A. R., Maier, H. R., Leonard, M., Roberts, A. J., Berrisford, M. J. (2006). “Application of two ant colony optimization algorithms to water distribution system optimisation.” Mathematical and Computer Modelling, vol. 44, iss. 5-6, 451-468.

[17] Murphy, L. J., Simpson, A. R., and Dandy, G. C. (1993). ‘‘Design of network using genetic algorithm.’’ Water, 20, 40–42.

[18] Simpson, A. R., Dandy, G. C., and Murphy, L. J. (1994). ‘‘Genetic algorithms compared to other techniques for pipe optimization.’’ J. Water Resour. Plan. Manage., 120(4), 423–443.

[19] Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). ‘‘An improved genetic algorithm for pipe network optimization.’’ Water Resour. Res., 32(2), 449–458.

[20] Goldberg, D. E., and Kuo, C. H. (1987). ‘‘Genetic algorithms in pipeline optimization.’’ J. Comput. Civ. Eng., 1(2), 128–141.

[21] Balla, M. C., and Lingireddy, S. (2000). ‘‘Genetic algorithms in pipeline optimization.’’ J. Comput. Civ. Eng., 14(3), 199–205.

[22] Levi, Y., Mallevialle, J. (1995). “Global strategy and new tools for maintaining water quality in distribution system.” Proc. of the China Water Supply Association, Safety of Water Supply in Transmission and Distribution Systems, Shanghai, China.

[23] Nace, A., Harmant, P., Villon, P., (2001). “Optimization of location and chlorine dosage of the booster chlorination in water distribution network.” Proc of the World Water & Environmental Resource Congress, May 20-24, Orlando, Florida. In Phelps and G.Sehlke (eds), Bridging the Gap: Meeting the World’s Water and Environmental Resources Challenges.

[24] Rouhiainen, C. J., Tade, M. O., West, G. (2003). “Multi-objective genetic algorithm for optimal scheduling of chlorine dosing in water distribution systems.” In Advances in Water Supply Management, (ed. C. Maksimovic, D. Butler & F.A. Memon), Balkema, Rotterdam, (2003a), 459-469.

[25] Rouhiainen, C. J., Tade, M. O., West, G. (2003). “Genetic algorithms for optimal scheduling of chlorine dosing in water distribution systems.” Proc.

of the 20th Convention of the Australian Water Association, 6-10 April, 2003, Perth, Australia, (2003b). 1-16.

[26] Ucaner, M. E., Ozdemir, O. N. (2003). “Application of genetic algorithms for booster chlorination in water supply Networks.” World Water & Environmental Resources Congress and Related Symposia.

[27] Issam, N., Lebdi, F. (2006). “Genetic algorithm for optimal choice of chlorine booster stations in drinking water networks - Algorithme Genetique (AG) Pour Le Choix Optimal Des Stations D’appoint De Chlorine Sur Les Reseaux D`eau Potable.” Revue des Sciences de l`eau, no:1,47-55.

[28] BetterNet User Manual By Soyupak S., Kılıç H., Pektürk M. K., Atılım University, Ankara, February 2010.

[29] Eiben, A. E. and Smith, J. E. (2003). “Introduction to Evolutionary Computing.” Springer, 1st edition.

[30] Beasley, D., Bull, D. R., and Martin, R. R. (1993). “An overview of genetic algorithms.” Part 1 & II, University Computing, University of Cardiff, UK, 2.

[31] EPANET 2 USERS MANUAL By Lewis A. Rossman, Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH 45268, EPA/600/R-00/057, September 2000.

[32] Castro, L. N. (2006). “Fundamentals of natural computing: Basic concepts, algorithms and applications.” Chapman & Hall/CRC, Taylor & Francis, Boca Raton, FL.

[33] Mitchell, M. (1998). “An Introduction to Genetic Algorithms.”, MIT Press, USA.

[34] Dorigo, M., Di Caro, G. (1999). “The Ant colony optimization metaheuristic.” In D.Corne, M.Dorigo, and F.Glover (eds.), New Ideas in Optimization, McGraw-Hill 13-49.

[35] Kennedy, J., Eberhart, R. (1995). “Particle swarm optimization.” Proc. of the IEEE International Conference on Neural Networks IV, 1942-1948.

[36] Farmer, J. D., Packard, N. H., Perelson, A. S. (1986). “The immune system, adaptation and machine learning.” Physica 22D, 187-204.

[37] Castro, L. N., Timmis, J. I. (2003). “Artificial immune systems as a novel soft computing paradigm.” Soft Computing, 7 (8), 526 - 544.

[38] Gibbs, M. S., Morgan, N., Maier, H. R., Dandy, G. C., Nixon, J. B., Holmes, M. (2006). “Investigation into the relationship between chlorine decay and water distribution parameters using data-driven methods.” Mathematical and Computer Methods, 44, 485-489.

[39] Goldberg, D. E. (1989). “Genetic algorithms in search, optimization and machine learning.” Addison-Wesley.

[40] Holland, J. H. (1975) “Adaptation in Natural and Artificial Systems.” University of Michigan Press, Ann Arbor.

[41] Back T., Fogel D. B., Michalewicz Z. (2000). “Evolutionary computation (Vol. 1. Basic algorithms and operators)”.

[42] EPANET Programmer’s Toolkit (2000) http://www.epa.gov/NRMRL/wswrd/dw/epanet/EN2toolkit.zip, last accessed in May, 2010.

[43] TÜBİTAK-KAMAG project, project no 107G088, “İçme suyu dağıtım şebekelerinde optimum klorlama uygulamalarının matematiksel modeller kullanılarak gerçekleştirilmesi ve dezenfeksiyon sistemlerinin yönetimi”, project finish date: 31 December, 2010.

APPENDIX

APPENDIX A - Generated Report by BetterNet

Tarih: Monday, February 01, 2010 4:50

Model: C:\Documents and Settings\selcuk soyupak\Desktop\testnetwork-2-eps.inp

----

**********Çözüm Parametreleri**********

(In Turkish)

Uygunluk Fonksiyonu: Konsantrasyonun Değişkenliği

Çözüm Methodu: Genetik Algoritma

Ek Dozlama İstasyonu Sayısı: 2

Dozlama Rejimi: Devamlı

Kaynakta Dozlama Mevcudiyeti: Mevcut

Ek Klorlama Aralığı: 0.3 - 0.4 mg/L

Ek Klorlama Adımı: 0.05 mg/L

Klor Konsantrasyonu Sınırları: 0.3 - 0.5 mg/L

Ortalama Klor Konsantrasyonu: 0.4 mg/L

Şebeke Kararlılığı Kontrolü: Mevcut

Kararlılık Periyodu: 24 saat

Popülasyon Büyüklüğü: 125

Çaprazlama Olasılığı: 0.85

Mutasyon Olasılığı: 0.01

Seçkinleştirme Yüzdesi: 0.4

```
Sonlandırma Modu: Nesil
Nesil Sayısı: 100
Epsilon Değeri: 2E-05
Populasyon Oluşturma Modu: Rastgele
--------------------------------------------------------
Sistemin ek klorlama yapılmaksızın uygunluk değeri: 0.004110267
--------------------------------------------------------
--------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 60.14507
Fark Kareleri Toplamı varyansı: 0.03299236
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003701052
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.3892544
Klor Azlığı Riski uygunluk değeri: 0.3548811
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.00929
Minimum Klor Kullanımı uygunluk değeri: 737.044

İşlem süresi: 00:00:10.9062500
0. nesildeki en iyi uygunluk değeri: 0.003701052
--------------------------------------------------------
--------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
```

```
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352

Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193

Klor Azlığı Riski uygunluk değeri: 0.3730903

Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859

Minimum Klor Kullanımı uygunluk değeri: 536.0321



İşlem süresi: 00:00:12.1093750

1. nesildeki en iyi uygunluk değeri: 0.003569352

---------------------------------------------------------

---------------------------------------------------------

Fark Kareleri Toplamı uygunluk değeri: 61.93249

Fark Kareleri Toplamı varyansı: 0.03397284

Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352

Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193

Klor Azlığı Riski uygunluk değeri: 0.3730903

Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859

Minimum Klor Kullanımı uygunluk değeri: 536.0321



İşlem süresi: 00:00:10.5781250

2. nesildeki en iyi uygunluk değeri: 0.003569352

---------------------------------------------------------

---------------------------------------------------------
```

```
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:09.7187500
3. nesildeki en iyi uygunluk değeri: 0.003569352
------------------------------------------------------
------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:09.7187500
4. nesildeki en iyi uygunluk değeri: 0.003569352
```

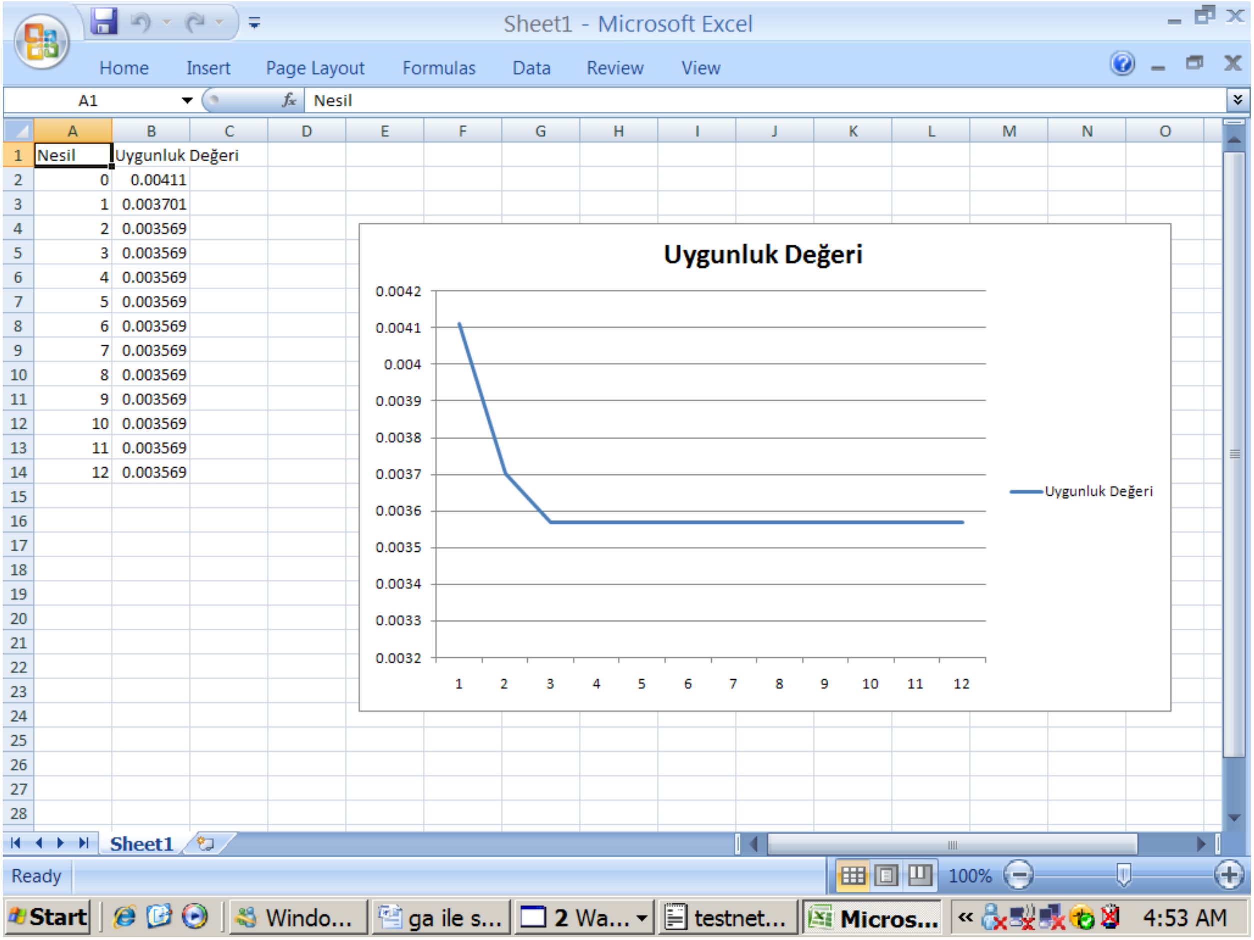
```
--------------------------------------------------------
--------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:09.7500000
5. nesildeki en iyi uygunluk değeri: 0.003569352
--------------------------------------------------------
--------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321
```

```
İşlem süresi: 00:00:09.9843750
6. nesildeki en iyi uygunluk değeri: 0.003569352
---------------------------------------------------------
---------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:12.1562500
7. nesildeki en iyi uygunluk değeri: 0.003569352
---------------------------------------------------------
---------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
```

```
Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:13.6718750

8. nesildeki en iyi uygunluk değeri: 0.003569352

---------------------------------------------------------

---------------------------------------------------------

Fark Kareleri Toplamı uygunluk değeri: 61.93249

Fark Kareleri Toplamı varyansı: 0.03397284

Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352

Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193

Klor Azlığı Riski uygunluk değeri: 0.3730903

Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859

Minimum Klor Kullanımı uygunluk değeri: 536.0321

İşlem süresi: 00:00:11.1093750

9. nesildeki en iyi uygunluk değeri: 0.003569352

---------------------------------------------------------

---------------------------------------------------------

Fark Kareleri Toplamı uygunluk değeri: 61.93249

Fark Kareleri Toplamı varyansı: 0.03397284

Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352

Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193

Klor Azlığı Riski uygunluk değeri: 0.3730903

Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
```

```
Minimum Klor Kullanımı uygunluk değeri: 536.0321
İşlem süresi: 00:00:11.2500000
10. nesildeki en iyi uygunluk değeri: 0.003569352
---------------------------------------------------------
---------------------------------------------------------
Fark Kareleri Toplamı uygunluk değeri: 61.93249
Fark Kareleri Toplamı varyansı: 0.03397284
Konsantrasyonun Değişkenliği uygunluk değeri: 0.003569352
Konsantrasyonun Değişkenliği kabul edilebilir aralıktaki konsantrasyonların genele oranı: 0.377193
Klor Azlığı Riski uygunluk değeri: 0.3730903
Fark Kareleri Toplamı(Min) uygunluk değeri: 28.93859
Minimum Klor Kullanımı uygunluk değeri: 536.0321
İşlem süresi: 00:00:11.4062500
11. nesildeki en iyi uygunluk değeri: 0.003569352
---------------------------------------------------------
Toplam işlem süresi: 00:02:12.6406250
Bulunan en iyi çözüm
Çözümün uygunluk değeri: 0.003569352
Düğüm noktası    Dozlama miktarı
-------------    ---------------
36               0.35
5                0.35
```

APPENDIX B - Generation-Fitness Value Graph



Nesil	Uygunluk Değeri
0	0.00411
1	0.003701
2	0.003569
3	0.003569
4	0.003569
5	0.003569
6	0.003569
7	0.003569
8	0.003569
9	0.003569
10	0.003569
11	0.003569
12	0.003569

Figure B.1 Generation – fitness value graph for tracking progress.

APPENDIX C – Information about Real Network in Antalya

Antalya is located in the southwest of Anatolia. Surface area is 20.723 km², and it is 2.6% of Turkey's surface area [43] (It is translated from Turkish to English). The provincial border is formed by the Taurus Mountains.

According to data from the year 2008, Antalya has a population of 1,859,275 [43]. City economy focused on tourism, agriculture, commerce sectors. Antalya is one of Turkey's most important tourism city. Antalya is leading agriculture center in Turkey due to the great ecological properties. City's 20 percent of land area is agricultural land. The majority of the people of the region, such as 80% are engaged in agriculture. Climate is Mediterranean Climate. It has a karstic structure, when topographic structure is considered.

Figure C.1 Overview to Antalya [43].

Antalya's main drinking and industrial water needs are met from groundwater sources. 88% of the water supply network is met from the source Duraliler, 8% of the water supply network is met from the source Boğaçay and 4% of the water supply network is met from the source Gürkavak [43] (See Table C.1). Drawn from this source water is supplied to the city's drinking water network.

Resources	Capacity of Production (%)
Duraliler I, II	88
Boğaçay	8
Gürkavak	4

Table C.1 Production capacity of resources [43].

In Antalya, the length of the existing drinking water network is 3000 km. City drinking water supply is divided to 8 pressure zone for distributing water to the subscribers easily and facilitating the control of network.

Konyaaltı is a district at the west part of Antalya (Figure C.1). This pilot project area Konyaaltı is divided into 20 sub-region (zone) in the context of TÜBİTAK-KAMAG project which is managed by Prof. Dr. Habib MUHAMMETOĞLU (Figure C.2). Zone 4 is fed from the source Boğaçay. Zone 4 is located within first pressure zone.

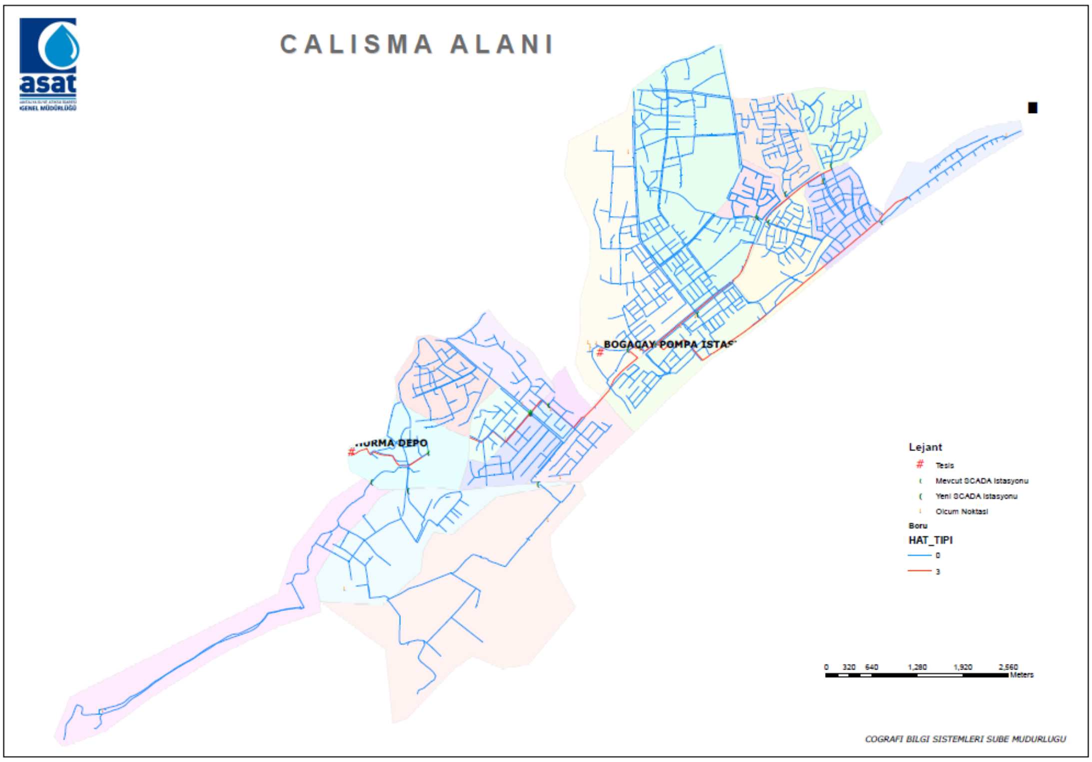


Figure C.2 Konyaaltı with 20 zones [43].

APPENDIX D – Comparison of GA & IGA Solution for Test Network 1

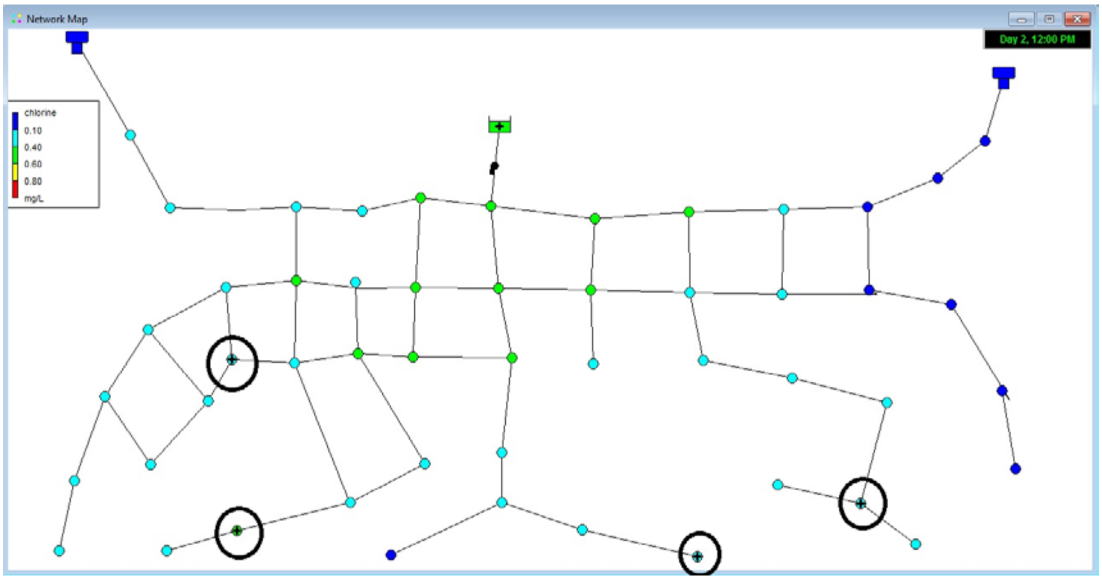


Figure D.1 Test Network 1 with 4 booster stations of GA solution.

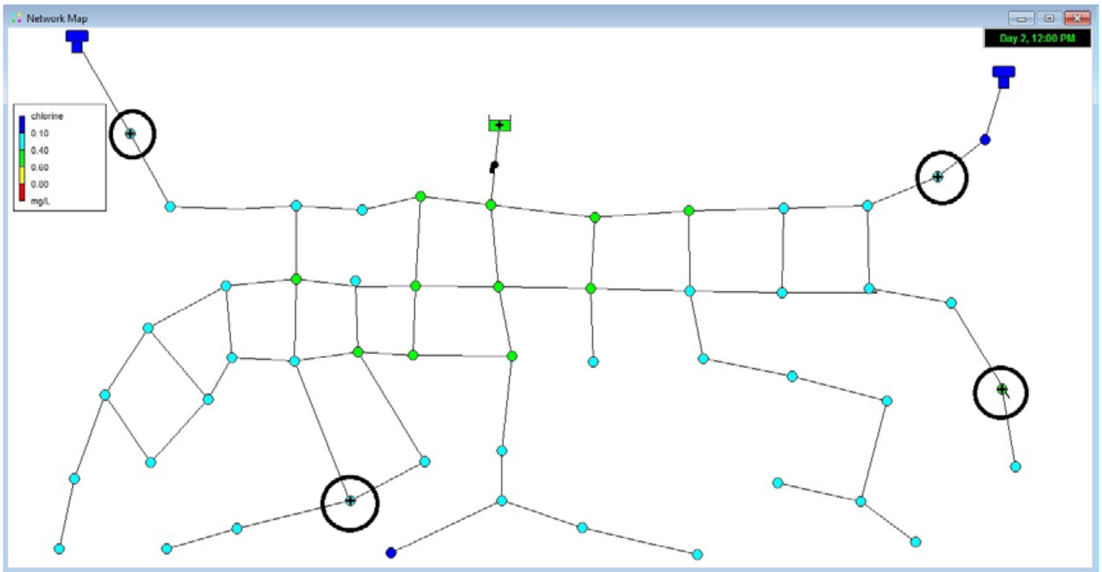


Figure D.2 Test Network 1 with 4 booster stations of IGA solution.

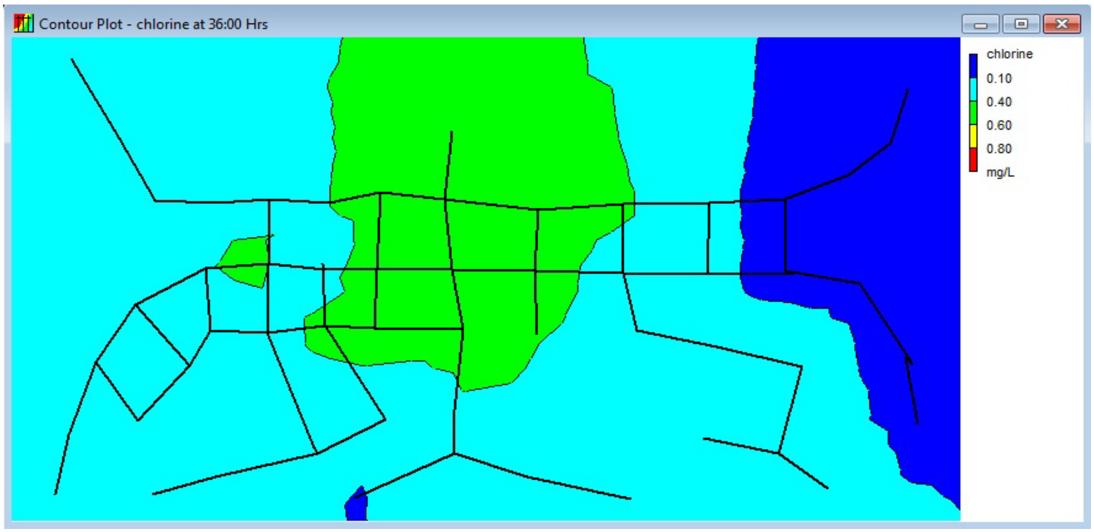


Figure D.3 Contour plot of chlorine at 36:00 hrs at GA solution.

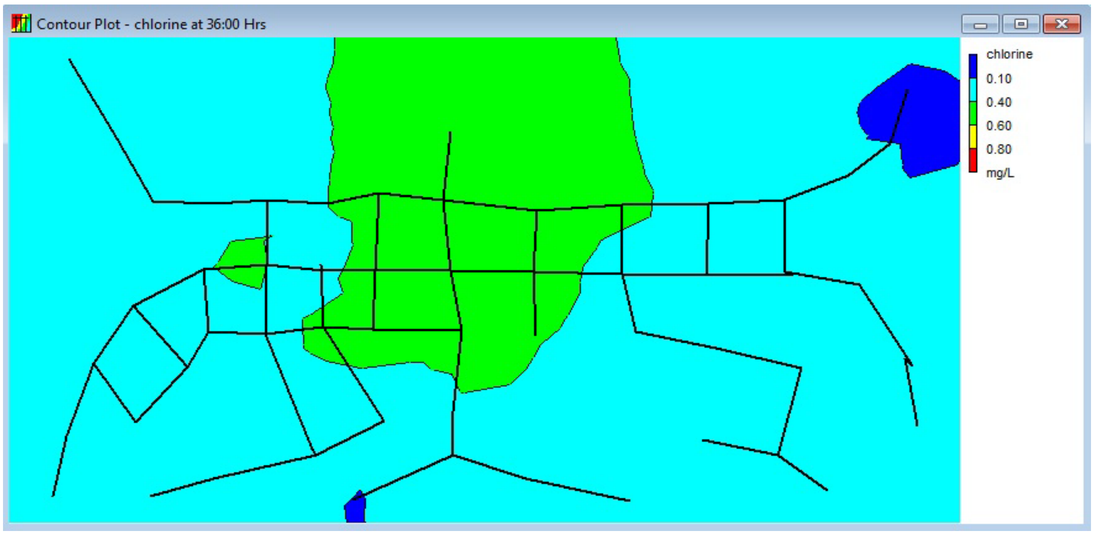


Figure D.4 Contour plot of chlorine at 36:00 hrs at IGA solution.

When we look at the Figure D.3, we see that there is a big problem at the right side of the network. At the Figure D.4, this problem is decreased by IGA.

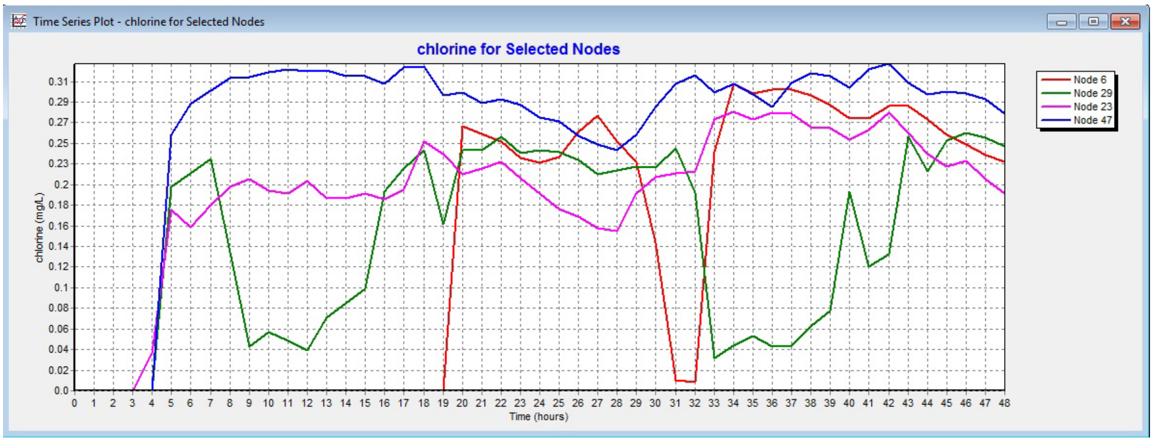


Figure D.5 Time Series plot of chlorine for node 6, 23, 29 and 47 at GA solution.

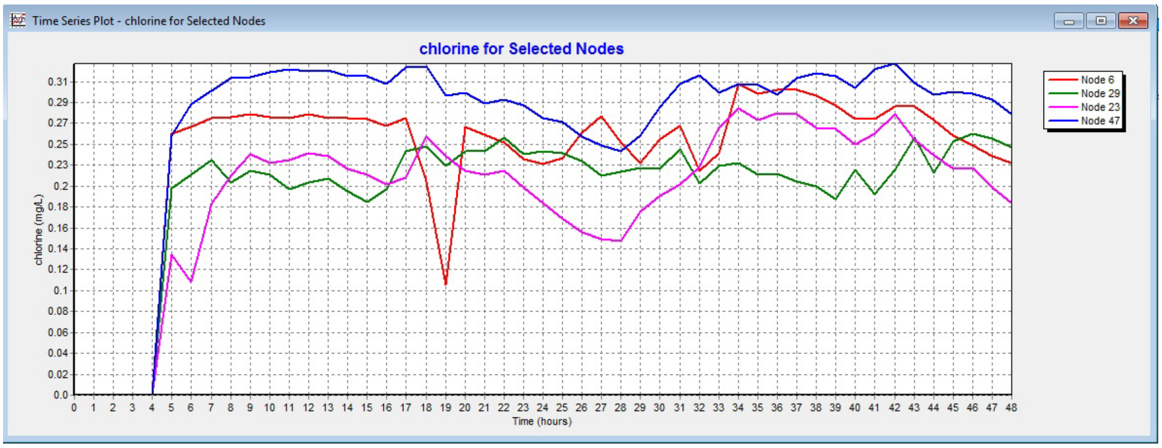


Figure D.6 Time Series plot of chlorine for node 6, 23, 29 and 47 at IGA solution.

When we look at Figure D.5, monitoring node 6 (red line) and monitoring node 29 (green line) have very big oscillation at GA solution. Low chlorine rates are seen at these nodes, 0.0 mg/L concentration is also seen for node6. When we look at Figure D.6, there is no chlorine concentration lower than 0.1 mg/L for monitoring nodes at IGA solution.