Technical University of Crete
Electrical & Computer Eng. Dept.

Operational Optimal Control of a Water Distribution Network Based on Genetic Algorithm

Diploma Thesis By

Asmargiannakis Petros

Supervisor: *Prof. Giorgos Stavrakakis*
Co-supervisor: *Dr. Eleftheria Sergaki*
Examining Committee: *Prof. Michael Zervakis, Dr. Eleftheria Sergaki*

March 2017, Chania

Acknowledgments

I would like to express my appreciation to my supervisor professor ***Giorgos Stavrakakis*** *for his continuous support, cooperation and guidance during the development of this thesis.*

My thankfulness also goes to ***Dr. Eleftheria Sergaki*** *for her enthusiastic assistance and encouragement, as well as her patience and invaluable advice throughout this research work, and professor* ***Michael Zervakis*** *for his helpful comments.*

I would also like to thank my family and friends for providing with their patience and support.

In particular, I would like to thank my good friend Thanos for his help and support during the development of this study.

Abstract

Despite wide research, design of water distribution networks is not realized using optimization techniques. The main reason for this assumption is that the design of water distribution networks is evaluated mostly, as a least cost optimization problem. Another parameter for preferring the traditional modeling practice is that, existing optimization algorithms are not presented to the user as friendly as it should be.

In this thesis we deal with the pump schedule optimization of water distribution systems. The aims and the possible applications of the presented method are significant and all of them are intended to solve a particular but realistic problem.

A genetic controller for a midscale water pumping station is developed in order to optimize the operation plan of the system for minimum energy consumption. A genetic algorithm, GA optimizer, is used to decide the operation plan of the pumps taking into account power consumption. Upper and lower limits of the water level in the tank is constrained by the GA.

The MATLAB Simscape blocks is used to build a full detailed model in a Simulink environment. The system consists of six pumps fed by a tank and deliver water to a buffer tank through the pipeline. This model is being processed many times to simulate, check and validate the behavior of the system. Chania Vlite's Pumping Station is taken as a case study. To optimize the operation plan of the system, the detailed model is used to design a preliminary control strategy which is used later to set the optimization problem. Fixed speed pump scheduling is found more efficient. The optimizer is tested and gave explainable results.

Finally, the case study is tested within 24 hour time horizon and an accepted operation plan has resulted. Detailed plots and charts are generated using MATLAB to explain what is going on through all the levels of work in this thesis. All detailed information about the operation of this pump station is from OAK AE.

Keywords: Genetic algorithms, pump schedule optimization, specific energy consumption, water distribution systems

Contents

1. Introduction

1.1. Introduction

A water distribution network is an essential infrastructure that conveys water from the source to the consumers. A typical water distribution network consists of pipes, pumps, tanks, reservoirs and valves. The system is mainly designed considering a demand pattern such as pressure limitations, velocity limitations, quality assurances and maintenance issues at minimum cost, which can be named as optimal design.

One of the most important consideration in designing and operating a water distribution system is to satisfy consumer demands under a range of quantity and quality considerations during the entire lifetime for the expected loading conditions. In addition a water distribution system must be able to accommodate abnormal conditions such as breaks in pipes, mechanical failure of pipes, valves and control systems, power outages, malfunction of storage facilities and inaccurate demand projections.

Traditionally, almost every water distribution network design is based on the proposed street plan and topography of the surrounding area. Using commercial software, the modeler simulates flows and pressures in the network and flows in and out to and from the tank for essential loadings.

Pumping in potable and waste systems absorbs a significant part of all the generated electricity. The ratio of energy consumption through water distribution is approximately 5% in the European countries. Therefore, reducing power consumptions in water distribution networks affects notably the amount of total consumed energy of a country, and has a big part in the development of sustainability and environment protection.

Optimization of a water distribution network is quite complicated due to nonlinear relationships between parameters. Recently, significant amount of research has been performed on the optimal operation of water distribution networks. Some of the first studies utilized linear programming and later studies applied nonlinear programming and genetic algorithm studies.

Significant amount of research on theories about optimization techniques for water networks designs and operation has been performed for years, but many of these theories cannot be modeled due to the complexity of methods and difficulty of technical application of these theories to real networks. Although nowadays as technology evolves it becomes more and more easy to model these theories, and as the cities grow, the importance of managing capital and maintenance costs of larger networks necessitates the use of optimization techniques.

A consequence of rising energy prices is to ration the available energy for various tasks. In water distribution systems, water often needs to be pumped to a higher elevation with adequate pressure. For example, hydraulic pumps transport water from a treatment plant into an elevated storage tank. From an elevated tank, waterfalls by gravity to reach with adequate pressure nodes in the water network where water is consumed.

Hydraulic pumps consume most of the energy required to operate a water distribution system. Therefore, optimizing pump operations may lead to significant reductions in energy expenditure. During the last decade many studies have proposed various automatic systems for the optimal scheduling of pump operations with the aim of saving energy and reducing operating and maintenance costs.

The present day energy requirements are higher: larger and more complex water networks need to be handled. Moreover, the problem formulation involves various constraints to maintain service levels besides minimization of operational and maintenance costs. Due to high variability and uncertainty involved with the prediction of demands, better results are expected to be obtained by using real-time optimization. For real-time optimization, the formulated model (including simulation model) and the chosen optimizer must be able to produce results in real-time. Therefore, despite the advances in computing and hydraulics, the problem seems as challenging today as it was a decade ago.

As an optimization problem, pump scheduling is difficult to solve because of its large search space, high computing requirements and the nonlinear and discontinuous nature of real world networks. As an engineering problem, automatic pump scheduling is an arduous task due to the diversity and complexity of real world water distribution systems. An engineer developing a system to automatically optimize pump operations in a real world network would need to address a multitude of problems: collect data about the physical network, build and calibrate according to this data a hydraulic model of the network and a demand forecast model, and define appropriate performance criteria and constraints in order to offer an adequate, reliable and competent service to customers.

All these tasks are prerequisites to develop the optimization system itself and they may well determine the design of the optimization system. As a research topic, automatic pump scheduling is challenging. Because of its strongly practical nature, it is difficult to provide a general and consistent formulation of the problem.

Moreover, each instance of the pump scheduling problem is motivated by a different application with unique peculiarities and requirements. Therefore, it is not surprising that each instance is often solved by a mostly specific approach that may not be suitable for a different instance of the problem. The end result is that most research on pump scheduling is never contrasted on multiple network instances, or compared against alternative techniques.

The goal of this thesis is to address the problem of pump scheduling in a water distribution network and find a cost efficient solution with the optimization technique of a genetic algorithm, having as the main parameter the water tank level adjustments

We focus on the cost and energy optimization of potable water systems. Our aim is to present novel methods which are capable of finding optimal solutions in cases of various types of water networks. For this goal, we present the basic concepts and definitions, which will be used in the latter parts of this thesis, giving a detailed overview on the modelling of the optimization problems from an engineering and mathematical point of view.

Reducing the energy consumption of water distribution networks has never had more significance than in the present day. The greatest energy savings can be obtained by careful scheduling of the operation of pumps. Moreover, the cost of operating pumps in a water distribution network represents a significant fraction of the total expenditure incurred in the operational management of water distribution networks worldwide.

Pumps consume large amounts of electrical energy for pumping water from source to storage tanks and/or demand nodes. In addition, they eventually need to be repaired and replaced, resulting in maintenance costs. Therefore, the goal of the pump scheduling problem is to minimize the total operational cost, which includes pumping cost and pump maintenance cost, while guaranteeing a competent network service.

In most cases, a competent network service is equivalent to supplying water to consumers at adequate pressures and achieving full recovery of tank levels by the end of operating period

Typical simplifications of a most realist model are also discussed. In the next part we will be focusing on a general literature overview of the optimization problems. Also, the basic concepts and definitions of the developed optimization techniques are shown.

Finally, the last chapter of this thesis describes the structure and the implementation of our optimization technique with the genetic algorithm.

The main aims of this thesis are to:

- Study the pump scheduling problem formulation in order to develop a general definition of its objectives and constraints.
- Describe various optimization techniques in order to determine why we chose genetic algorithm as our technique.
- Provide a sound and exhaustive analysis of the genetic algorithm technique applied to a water network in order to identify best settings, best representations.
- Based on the previous findings, to produce state-of-the-art results.

1.2. General literature and solver overview

As we discussed pumping of treated water represents the major fraction of the total operation cost in conventional water supply systems and even a small improvement in operational efficiency, causes enormous cost savings to the industry. Making the need of optimization mandatory, either in the design stages of waterworks or towards the operational level, achieving satisfying water demand with minimal energy consumption. Sophisticated operation can result in significant savings, even in small scale waterworks.

If sufficient storage capacity is available, the water demand can satisfied with a large number of pump schedules. As the energy consumption charge changes during the day or the specific energy consumption of the pumps are different, different overall energy charges correspond to pump schedules and thus it is beneficial to find and realize the most cost-effective one (Barán et al. 2005, Ormsbee & Lansey 1994, Coulbeck 1977, Tolnai et al. 1995, Fu˝zy 1991).

Although, human operators of water distribution networks usually use heuristic ideas of rules of thumb to minimize costs. Several researchers have been developing techniques for minimizing the operating costs associated with pumping systems of water supply.

Many researchers have been developing techniques for minimizing the operating costs associated with pumping systems of water supply. A state-of-the-art overview of the applied mathematical programming and spatial decomposition methods can be found in Mays (1999) and a detailed review is given in Ormsbee & Lansey (1994).

Among these techniques, soft computing methods and metaheuristics became more popular due to their robustness during the last decades, such as fuzzy logic (Angel et al. 1999, Vamvakeridou-Lyroudia et al. 2005), nonlinear heuristic optimisation (Ormsbee & Reddy 1995, Leon et al. 2000), genetic algorithms (Mackle et al. 1995, Savic & Walters 1997, Boulos et al. 2001, Labadie 2004, Tu et al. 2005), memetic algorithms (Zyl et al. 2004), particle swarm optimization techniques (Baltar & Fontane 2008), colony models (Ostfeld & Tubaltzev 2008), and genealogical decision trees (Ikonen et al. 2012). Although these techniques are robust and more or less insensitive for the modelling (e.g. for non-linearities), they suffer from the lack of reliability: they cannot guarantee reliable results, one even cannot be sure whether they are able to find a feasible solution (which satisfies all the constraints) for a single run.

Modeling of water consumptions has a notable literature background. Consumptions can be involved into the computations as deterministic or stochastic data. In the first case, where the water demands are modelled as deterministic from the optimization point of view (Certainty Equivalent Control, Bertsekas 2005), they are usually determined a priori from statistics or by any forecasting approach (Bárdossy *et al.* 2009, Alvisi *et al.* 2007, Adamowski 2008). When the stochastic behaviour is taken into ac- count, the consumptions can be described by a priori determined distributions (Ikonen & Bene 2010, 2011, Cervellera *et al.* 2006) or the forecast model can be included into the optimization technique itself. In this thesis, the consumption data are known, deterministic data, obtained from the industry.

Further literature overview will be given on the particular topics at the beginning of latter chapters.

2. Water distribution systems

A water distribution system is a hydraulic infrastructure that conveys water from the source to the consumers. The primary objective for the control of the system is to satisfy the residential and the industrial consumers with appropriate quality, quantity and pressure. A water distribution system consists of building elements such as pipes, valves, pumps, tanks and reservoirs, which can be classified into active and passive. The active elements, namely pumps and valves, control the flow and pressure in the system; and the decision variable consists of their operating states. Other elements are used to be called as passive.

2.1. Description of a WDS's most common elements

At first, in a water distribution system the water between the nodes of the system (pressure zones) is delivered by the pumps and pump groups. The operation of the pumps has a great influence on the overall energy consumption of the system thus their operating points are the most important variables of the problem. They are either on/off-type pumps, which can be only switched on or off or pumps with frequency converter, where the flow rate is a continuous variable. If two or more pumps are connected in parallel in an engine house, they build a pump group together.

Valves are usually used for controlling the flow of the reservoirs: they can set the reservoirs in filling, emptying, or closed state. They are mostly modelled as on/off type valves because a half-opened valve would cause an undesirable operation through the significant energy loss.

The pump groups consume electric energy, which is supplied by the power stations. The price of the energy can change during the optimization time horizon and our goal is to satisfy the consumer demands with the smallest operational cost. If the energy tariff is uniform, the cost optimization gives also the energy minimum. The total power of the pump groups which is connected to the power station must not exceed a given limit.

The amount of water which is fed into the network is obtained from water sources or wells. The exploitation of the wells must fulfill several technological requirements, e.g. the flow rate can be changed only few times a day, and the wells have lower and upper daily exploitation volume limits.

Reservoirs are the elements that represent the storage capacity of the network thus allowing the possibility of various controls. Also, due to e.g. fire safety issues, they form important constraints for the problem, which are the minimal and maximal water volumes (or sometimes formulated as water levels).

Pipelines, valves and bends serve as conveyors for the water. They cause energy loss which must be covered by the pumps.

Water demands exist in several nodes of the network and they are stochastic in the real world. When modeling, we can consider this fact by defining them with any kind of distribution functions, or we can approximate the reality by deterministic consumptions (by the expected value of the stochastic consumption).

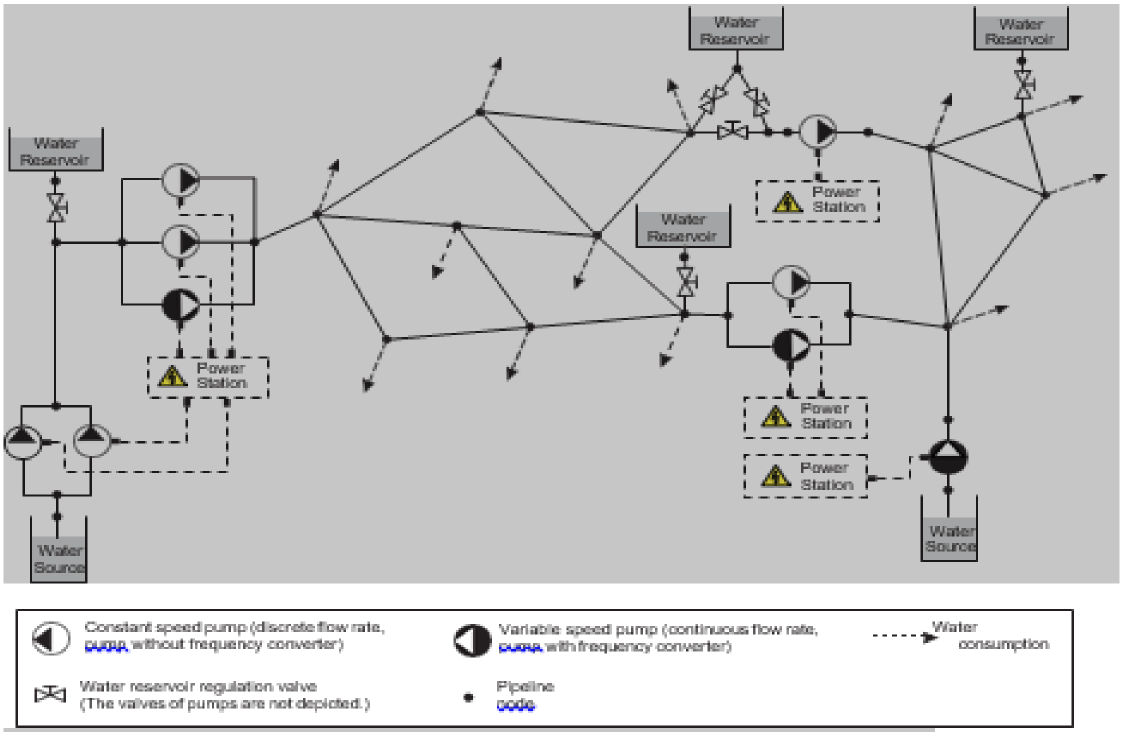


Fig 2.1 A model water distribution system of minimal size but full complexity.(Thesis JÓZSEF GERGELY BENE pump schedule optimization techniques for water distribution systems Oulu 2013)

2.2. Historical background

In order to have a better understanding of the role and aim of such complex systems as water distribution networks, we will try to explore their evolvements providing a short historical overview.

The water supply is coeval to the first civilizations, since all antique cities were founded next to riverbanks, such as the Ganges, the Sindhu, the Tigris, the Euphrates, or the Nile. The first successful examples to control water can be found in Mesopotamia and Egypt. About 5000 years ago, in the valley of the Sindhu river lots of houses were equipped with their own bath through ceramic pipelines, and bricked drainage system was established. The importance of the storage of water was firstly recognized in India. Note that nowadays the storage capacity enables the operational optimization in WDSs.

The ancient Rome is also well-known for its well organized water supply system. The water, originated from more sources, was conducted in collector reservoirs from which the water was delivered through open channels and pressurized lead or bronze pipes. Water was available for all the citizens. The first regulation on water supply was also born: in the case of thin runoff of the sources, first the houses of the citizens were disconnected from the system. On drier days, the water supply of the aristocrats was also terminated but the baths, public fountains were still served.

Another example for archaic water systems is Knossos in Greece, where pressurized pipelines were found. The beginning of the modern water supply is dated back to 1544 when the British Parliament expressed the right of London's inhabitants for clear water. The first water supply systems (Boston 1652; Shaeffer 1746) in this age still meant gravitational conduits and wooden water tanks from which the people were able to fill their buckets. The first pumps driven by steam-engines were used in Bathelehem, Pennsylvania, 1764.

Water supply developed significantly during the 19th century. More complex, looped iron-cast pipeline networks were established with newly developed stop-valves, check-valves, and other controlling elements. The electrification of the pumps began in the first half of the 20th century. These networks were already structurally the same as nowadays, however energy optimization was for a long time impossible due to the poor computational capacity of computers.

2.3. Types of water distribution systems

The first discussed aspect is the type of the water sources which clearly affects the possibilities of the water network operation. One possibility is using surface water like a river, e.g.

the Danube in Budapest. Water is usually mechanically filtered by the shingly riverbank itself. Usually there are no restrictions for the amount of exploited water.

Contrarily, in case of ground water wells, e.g. the city of Sopron in Hungary, there are strict constraints for pumping the water: the flow rate of the wells must be constant during longer periods, and there are capacity limits as well. They must be taken into account while scheduling. The suitable disinfection process depends mainly on the size of the pipeline: in case of smaller WDSs, a single ozone fertilizer is enough, but larger networks require protection also against re-infection: chlorine based methods are typical.

Let us consider the structure of the network: the main types are serial, branched, or looped. Tiny waterworks usually have a serial pipeline system, which connects the pump and the reservoir. For bigger networks, branched (i.e. tree-shaped) topology is used. Both network types have the advantage of having a hydraulic behavior which is easy to model.

Large WDSs usually have a looped structure. It means that the pipe network looks just like a net, the water can convey to a given demand point through different paths depending on the hydraulic state and the control (schedule of the pumps and valves). The biggest advantage of the system type resides in the more safety operation: consumers (and fire service) can be supplied also when a part of the pipeline is broken.

Besides, the structure helps avoid the stagnation of the water because the flow direction in the pipelines varies. The drawbacks are the most expensive investment and maintenance costs, moreover, the long running time of hydraulic simulations makes the optimization cumbersome when several thousand simulations must be performed.

2.3.1. Role of water reservoirs and geographical differences

Beyond their role of saving energy, water reservoirs are an essential part of water distribution systems also from technical point of view.

Water distribution systems are divided into pressure zones, which are connected with pump groups. Flat areas have only one zone but a WDS over a hilly terrain contains more. However, it is important to emphasize that even one pressure zone needs water storage in order to supply the consumes with pressurized water. Hence, the water is stored in high water towers usually built on plain areas, or in simple ’pools’ built on the top of a hill.

The question may arise why not to supply the consumers directly through pumps. In that case, we would not need to build or maintain these huge objects called reservoirs. The answer is very logical. Aside from the economical advantages, which a reservoir provides, it also means safety and trust on the system.

A WDS must work in any cases and it must supply the domestic consumers, hospitals, and fire service. If any catastrophe occurs, it often causes a power break while there is eager need for water in hospitals for the newly injured people. Besides, water storage is needed in everyday situations as well, such as during the replacement of an old pump or a broken pipeline. In case of small consumption zones, “towerless” districts are also possible. Then the demands must be satisfied through variable speed pumps and the fire service must be organized by particular thoughtfulness from other zones, i.e. there must be a possibility to fill tank cars in the neighborhood. Nevertheless, the bigger part of a WDS must always contain water reservoirs.

The size of the available water storage affects the operation habits of a WDS. For example in Tampere, Finland the storage capacity is almost half of the daily total water demand. In that case, an energetically optimal, constant flow rate can be set for the pumps (by local controllers) which satisfies the constraint that the daily sum covers the demand. The question may arise how to define the energetically optimal flow rate: it can be based e.g. on the maximization of the efficiency or on the minimization of the specific energy consumption. Also, the water level variation in the reservoirs can affect the optimal flow rate.

The central schedule optimization usually means ’playing with the storage capacity’. There will be a detailed explanation later on how to exploit this phenomenon.

2.3.2. Role of pumps

Pumps deliver water between different pressure zones. They either fill the reservoirs or supply the consumers with fresh water, or do both at the same time. In the past only directly driven pumps were used, but variable frequency drives have been spread due to their decreasing price in the last decades. Variable speed pumps make the control much more flexible, they allow to use the pumps economically even if there are big changes in the consumption habits.

Otherwise, in a huge network there is a high number of pumps. Equipping all of them with frequency inverter is impossible and, in addition, unnecessary. Usually more pumps run parallel within a pump group, and if one or two of them are VSP, it ensures the possibility of a very flexible control. Summarizing the above mentioned, one can conclude that mixed type WDSs (both FSPs and VSPs in the network) are the most common worldwide and they will remain so for a long time.

The possible control levels of the pumps must also be discussed in a nutshell. The two main possibilities are central regulation or the use of local controllers. In the first case one pump or an engine house with several parallel attached pumps can be controlled e.g. in order to keep the pressure or the flow as constant. On the other hand, central control can take into account more aspects at the same time and it is suitable e.g. for overall operational energy optimization.

2.4. Optimization of water distributions systems

The most important consideration in designing and operating a water distribution system is to satisfy consumer demands under a range of quantity and quality considerations during the entire lifetime for the expected loading conditions. It has to be able to accommodate abnormal conditions such as breaks in pipes, mechanical failure of pipes, valves, and control systems, power outages, malfunction of storage facilities and inaccurate demand projections.

The possibility of occurrence of each of these deficiencies should be examined to determine the overall performance and thereby the reliability of the system. In general, reliability is defined as the probability that the system performs successfully within specified limits for a given period of time in a specified environment. As it is defined above, reliability is the ability of a system to provide adequate level of service to the consumers, under both normal and abnormal conditions.

Traditionally, a water distribution network design is based on the proposed street plan and the topography. Like electric power lines, roads, and microwave radio networks, water systems may have a loop or branch network topology, or a combination of both. The piping networks are circular or rectangular. If any one section of water distribution main fails or needs repair, that section can be isolated without disrupting all users on the network.

Most systems are divided into zones; the extent or size of a zone is usually determined by hydraulics, telemetry systems, history, and population density. Sometimes systems are designed for a specific area then are modified to accommodate development. Terrain affects hydraulics and some forms of telemetry. While each zone may operate as a stand-alone system, there is usually some arrangement to interconnect zones in order to manage equipment failures or system failures.

Even a small network containing pipes at the order of thirty can require millions of combinations of pipes not including pumps, tanks and valves. It is scarcely possible that a modeler, using traditional modeling practices, finds the optimum solution even for a small network concerning a least cost design. That’s why, optimization techniques are applied for the design of water distribution networks.

Most of the optimization programs define the design problem basically as minimizing the pipe cost subjected to the satisfaction of the velocity and the pressure constraints and the satisfaction of nodal demands. However, modelers need to take into account, especially, reliability considerations and monetary limitations also.

Optimization of a water distribution system is quite complicated due to nonlinear relationships between parameters. Recently, significant amount of research has been performed on the optimal design of water distribution networks. Some of the first studies utilized linear programming (LP); later studies applied nonlinear programming (NP) and Genetic algorithm studies (GA).

Significant amount of research about optimization techniques for design of water distribution networks has been performed for years and there exist theories about optimization. But, many of these theories cannot be modeled due to complexity of methods and difficulty of technical application of these theories to real networks. Nowadays, it is easier to model these optimization theories by the help of computers. As the cities grow, the importance of managing capital and maintenance costs of larger networks necessitates using of optimization techniques.

There are various researches about water distribution system reliability based optimization. Reliability based optimization of water distribution systems requires combination of an optimization algorithm with a method for estimating reliability.

2.5. Modeling of water distribution systems

The hydraulic model of a WDS is an essential part of the daily operation. Waterworks need a well-maintained model in order to analyze the hydraulic behavior of the system, i.e. compute pressures, volume flow rates, chlorine concentrations, etc and perform optimization.

The computational accuracy and demand are always conflicting requirements; one has to balance how detailed hydraulic model one should use for the given purpose.

2.5.1. Use of a calibrated model

Note that the calibration of the hydraulic model is crucial for accuracy. Calibration means setting up the physical parameters of the system. Some of them are easy to identify, e.g. pipe length and diameters are known but others, e.g. pipe roughness, valve loss coefficients need a thorough calibration process when computations and measurements must be compared until sufficient accuracy is reached.

A well-calibrated hydraulic model can be used for analysis and design purposes. In the case of analysis, the hydraulic behavior of the system can be investigated. A design process can be traditional or optimality based. The traditional method means that the designer specifies for example pipe sizes based on own experience, then performs a hydraulic simulation, and checks whether the flow conditions are suitable for supplying the consumers. In the case of recovery, a hydraulic model can help answer what would happen if one of the pumps were changed to a new one with different characteristics.

Optimization methods do the same as traditional methods, but the parameters must be classified into two groups: some of them are fixed, which stem from model calibration or data sheets, and others can be varied, e.g. pipe diameters, pump types, and mean the free parameters of the system. An objective, the measure of goodness must also be defined, which expresses our aim, e.g. a combination of investment and operational costs.

2.5.2. The modelling of water consumptions

The first aim of a WDS is supplying the water demands of domestic purposes, e.g. cooking, drinking, cleaning, bathing, air conditioning; public use, e.g. swimming pools, parks, and hospitals; commercial consumers, e.g. hotels, restaurants, car washing; and manufacturers, Bhave & Gupta 2006.

The demands can be further classified into two groups (Máttyus 1987). Volumetric demands arise when the consumers need a given amount of water e.g. for filling a reservoir or drinking a glass of water. Another demand type is when the time of the consumption is fixed, for example by washing hands, having a shower, etc. In this latter case, the bigger the pressure in the network is, the bigger the consumption is. This phenomenon can result in significant wasting, however, it is paid by the consumers).

A water distribution system consumes a significant amount of water by itself, which is called self-consumption. This is used mainly for cleaning purposes, e.g. washing the reservoirs and pipelines, or consumed by the workers of the water company. Lost through leakage also means a remarkable amount of wasted money. Leakage is usually present in a system, but their portion can be decreased by regular maintenance. To detect leakages, a professional flow measurement system is needed, which also means the basis of an accurate billing system.

The forecast of these consumptions is a crucial part of any sising and scheduling processes. The forecast is highly challenging because of the stochastic nature of the consumptions as they depend on plenty of circumstances: the city type (suburban or city center), the season, the weather, etc. Human operators need to have a great knowledge and experience about the particular city where they work in order to satisfy all the demands. One has to keep in mind that the presence of optimization is just an extra possibility after the waterworks have satisfied all of their obligations.

2.5.3. Commonly Used Water Distribution Networks Design Software

2.5.3.1. Matlab Simulink

Since some basic facts about pumps were introduced in the previous chapter, we are now able to deal with the presentation of our model and its implementation in Matlab Simulink. Firstly, we present our model, the equations that characterize it and the values of the pump's parameters. There is a brief description about the tool we use to implement our model which the Matlab Simulink® environment. Then an example is given and what follows is the implementation of our model in Matlab Simulink.

Our goal is to create a mathematical model of our system which describes the function of a water distribution network. In this procedure the processing mechanism of the system has to be understood and be described in the form of some mathematical equations. These equations constitute the mathematical model of the system. The mathematical model of the system should describe accurately the input/output behavior of the system and be simple enough. Simplicity is very important because it makes the control task easier to understand and implement and it is more reliable for practical purposes.

The tool used to implement the model presented is MATLAB Simulink. It is an environment developed by MathWorks used for multi domain simulation and Model-Based Design for dynamic and embedded systems. It can be used in many areas to design, simulate, implement, and test a variety of time-varying systems, including communications, controls, signal processing, water distribution networks, video processing, and image processing. Simulink provides an interactive graphical user interface environment and a customizable set of block libraries. It offers tight integration with the rest of the MATLAB environment and can either drive MATLAB or be scripted from it. Simulink can be used to explore the behavior of a wide range of real-world dynamic systems, including electrical circuits, shock absorbers, braking systems, and many other electrical, mechanical, and thermodynamic systems. Dynamic systems are systems whose outputs change over time and the way they are represented is by a set of differential equations in time. Simulating such a system with Simulink requires a user to create a block diagram using the

Simulink model editor that graphically depicts time-dependent mathematical relationships among the system's inputs, states, and outputs and then command Simulink to simulate the system represented by the model from a specified start time to a specified stop time.

2.5.3.2. Epanet

EPANET is a calculation program distributed by the USEPA, which uses a visual interface to model pressurized water distribution systems. Using this program, pipe networks (including pipes, nodes, pumps, valves, storage tanks and reservoirs) can be physically drawn in the user interface or imported through GPS output files, AutoCAD or Google Earth. EPANET calculates hydraulic head, pressure and water quality at every junction; the flow rate, velocity head loss and average water quality though every pipe; and the hydraulic head and water quality at every tank. Outputs include color-coded network maps, data tables and contour plots.

The consequences of Epanet make us consider using it. It should be used in combination with other software programs. Epanet is applicable but it is also very simplified and does not take into consideration the extra variables experienced with rural communities .Also it does not model surface water sources, it can only model with a storage tank and a flow control valve. It also assumes all pipes are pressurized and there is no air in the pipes, which is usually not the case in rural communities. Finally, cannot model break pressure tanks.

2.5.3.3. Goodwater

Goodwater is a decision support system which assists the user in site assessment, system design, project budgeting and scheduling, implementation and evaluation of gravity fed water systems. It is specifically tailored to the needs of rural water systems. It is a 5.17 MB file run through Visual Basic for Applications (through Microsoft Excel), developed by Stephen Good in his 2008 Master's Thesis from Michigan Technological University. The Solver add-in must be loaded, and macros enabled to use the program. The main tool of the system design module is the pipe diameter optimization tool (which optimizes for least capital cost), but it also includes other tools to assist with storage tank sizing, and locating possible air blocks and sedimentation areas. The program includes lists of suggested actions at each stage of the project to encourage sustainability. A help module includes an in-depth explanation of each suggested action and user input, as well as a list of references.

The limitations of that program are the reasons for not picking it. It does not allow for including components such as groundwater wells, pumps, valves, or rainwater harvesting systems. Also, a limited number of items are allowed to be input by the user. For example, only 10 possible pipe diameters, and 25 of each type of component (tanks, public tap stands, etc.), 25 materials per component. Furthermore, it does not provide recommendations for where and what type of valves to include. Afterwards, it is not always clear why a design does not work, though the program does indicate problems such as a minimum or maximum head problem. The Options module should be consulted for what constants are being used. If any user inputs are changed, the report for that module must be rerun.

2.5.3.4. Reent’s Spreadsheet

Reent’s spreadsheet aids in the design of gravity fed water distribution systems, from the hydraulic conduction line to the distribution network. It aids in the selection of the appropriate pipe diameters, tank volumes, flow rates and also has some useful tools for budget planning.

Its drawbacks are keeping us from using that software. Feasibility isn't taken into account. Also, there is no way to incorporate geospatial technologies. There are no means to optimize cost. Furthermore, it doesn't account for air blockages and it can only design for PVC or galvanized iron pipes. Then, there is a lack of information on the effectiveness of the design. Finally, there is not troubleshooting for issues with break pressure tanks.

2.5.3.5. NeatWork

NeatWork is used to design the transmission line from a storage tank to individual faucets in a gravity-fed water distribution system. In an attempt to lower costs and maintain acceptable flow at individual faucets, NeatWork compromises by regulating flow using friction in small diameter pipes. This results in a compromise where flow will vary, but costs will be acceptable. Because flow variability and construction cost are in direct conflict, it is left up to the user to decide what best meets the needs of the project.

There are plenty of reasons that make us not choosing that software. Some of them are that incorporating loops is very complicated; also it does not incorporate the conduction line to the storage tank used for the distribution system. Furthermore, it does not deal with air blocks in the distribution network. Finally it does not incorporate tank design or size and most significantly it does not have any plotting or graphing tools.

2.6. Economical aspects of water network

Water network optimization often means operational optimization. The most common aim is to optimize the total electric cost of a system. This is the most important from an economical point of view; however, saving money can result also in saving energy and saving water. These latter aspects have at least the same importance since they help not to waste the environmental resources. Maintenance costs can be also classified into the group of operational costs; one can for example minimize the total running time of pump groups which are proportional to their service costs. It also has to be mentioned that the parameter calibration, which is an essential part of any optimization, costs money. The optimization through operational scheduling is the main topic of this thesis thus it will be detailed later.

Water network design and renovation are also popular research topics. In these cases, the different planning horizons play significant roles. These are the economic life, physical life, period of analysis, and design period (Bhave 2003). Economic life means the time during the economic benefits of an element exceed the cost while physical life means the period when the item is able to function. Period of analysis means the duration of the performed economic analysis while the system can supply the demands in its design period (it needs maintenance, of course).

These definitions suggest that the economic settlement, renewal, and operation make a coherent, complex task, which is better to be handled simultaneously. However, if the system is given and the waterworks have no or limited budget, only the operational costs can be taken into account; but this is often rather a political question and not an engineering one.

If we are in the position that we can take into account the economic aspects during the design or renewal phase, we have to balance between the operational and investment costs. A pipeline with bigger diameter costs more but the flow loss is much lower through it. A more expensive pump has better efficiency and lower electric consumption. If the aim is feeding the network at various demands, a frequency converter can also cover its investment costs within months. Building a bigger reservoir can allow us to make more flexible pump schedules during the operation phase.

Smaller or bigger renewal of a WDS can be done for several reasons. It can be a simple replacement of an item, which cannot be repaired economically. Another option is a planned renewal process, which is intended to make the operation of the system more economical. Significant change in the consumer demands also requires rebuilding a system, such as the drastic decreasing of water demands in the post-socialist countries when the water tariff doubled abruptly.

2.6.1. Safety and infrastructural issues

The primary aim of the pump scheduling of a WDS is satisfying all the residential and industrial demands with high emergency reserve. If the water company could not serve potable water, it would lose its trustfulness. Moreover, the WDS must always supply e.g. hospitals and the fire departments. For this reason, the water levels of the reservoirs cannot decrease below a given limit. Human operators usually keep the water level much more above this limit, which has been already stated safety because they are afraid of causing emergency during their turn.

Note that not only scheduling affects the safety of a system, but bigger reservoirs mean bigger reserves as well. In the case of looped network, the consumers can be satisfied through different paths. It is also very important to understand that there is no such schedule optimization software which can substitute human operators. They can use a pump scheduler as help in order to find more economic operation, but they still have to have the possibility to act in any unexpected cases such as a broken pipe, a power break, or a pump malfunction.

One always has to keep in mind that the water suppliers are not just other companies. They are one of the most significant parts of the infrastructure and their operation influences the economy, industry, and everyday life. Although the presence of tap water seems natural, one must know that it is the result of a work of a finely aligned system and people. Thus one must be careful when to do any changes in the system in order to optimize something which is far less important than safety.

2.7. Cost representation of the optimization

2.7.1. Typical objectives

The objective functions of a water distribution system can be various. Investments and operational costs can be optimized in the planning or renovation phases as suggested e.g. by Clark et al. (2002) and Lauria (2004). The other possibility, which is the most common, is using only objectives of the water network, which are related to the operation. Then the system is considered as it is, meaning that topological changes cannot be performed.

The basic idea underlying pump (and valve) schedule optimization is that the water consumptions can be satisfied by several different schedules. The electric energy used by the pumps is the largest part of the total electricity bill of waterworks (Nitivattananon et al. 1996). Therefore, the total electric cost of the pumps over a finite time horizon is used as the most common objective function.

The number of switches of the pumps can be an alternative objective function. It describes how many times the pump operating points are changed during the optimi- sation time horizon (Kullmann 2004). The total operation time of the pumps can be also minimized (Cembrano et al. 2000). Both objective functions take into account the maintenance cost of the pumps: they are proportional to the deterioration of the pumps.

The maximum demand charge (McCormick & Powell 2003, Barán et al. 2005) is the cost of the maximum power peak billed by the electric company. The water level variation in the reservoir can also be minimized (Barán et al. 2005). Water quality properties can also serve as objective functions (Sakarya & Mays 2000).

The objective functions can also be used simultaneously by aggregations, i.e. multiplying them with weighting factors and summing them together, or by a real, multi- objective optimization like in Barán et al. (2005).

2.7.2. Possibilities to spare electric cost

The key questions of the optimizations are how to exploit the storage capacity of the reservoirs in order to decrease the electrical expenses and how to find an optimal schedule within reasonable time. Computational demand and time play a significant role since operators need to generate new schedules in minutes in real-life circumstances.

The most obvious possibility for decreasing the workload costs is filling up the reservoirs during the time periods when electricity is less expensive and covering the water demands from these reservoirs in the expensive tariff hours. The idea seems clear, but due to the large number of constraints (reservoir capacity, node pressure and power limits) and the mixed-integer type variables (constant and variable speed pumps) the problem becomes highly challenging from mathematical point of view.

The second possibility of decreasing the expenses is reducing the power consumption itself. This plays a big role especially if the energy tariff is uniform, as in Hungary and in Finland nowadays. In this case, the specific energy consumption of the pumps4 is a good quantity to describe the thrift of the system.

Energy can be saved by using the pumps which have lower specific consumption values or using the pumps close to their best-efficiency points, which are determined by the revolution number and the state of the whole system. In these cases, the storage capacity is also essential: it allows to store the spare water if the pumps deliver more water in their efficient operating points than needed.

2.8. Other requirements on the water network operation

Although the above mentioned rules of thumb seem obvious, determining the optimal schedule is a highly challenging task due to the constraints of the system. Some of the objective functions can be transformed into constraints e.g. the switching number of the pumps, the maximum power peak of pump groups, water level variations, and water quality properties.

Besides, the capacity limits of the reservoirs, the exploiting limits of wells, nodal pressure limitations make the optimization problem even more complex.

A precise description of the objective functions and constraints is given in the next sections.

3. Optimization techniques

3.1. Introduction

A significant problem in the modern industry is considered to be the design of reliable hydraulic networks. The optimum network layout is an important stage of network design with requirements such as pressure, power consumption and demands at different nodes and also to minimize cost while meeting a performance criterion.

This chapter introduces a brief review of optimization techniques of water distribution networks and water distribution system reliability is given. There are plenty of applications can be classified roughly into three classes. These classes are **calibration studies, operation studies and Design studies**. There are also two main branches that optimization techniques can be classified, **the deterministic optimization techniques** (linear programming, non-linear programming and dynamic programming) and **stochastic optimization techniques** (simulated annealing and genetic algorithms) as indicated in Fig 3.1.

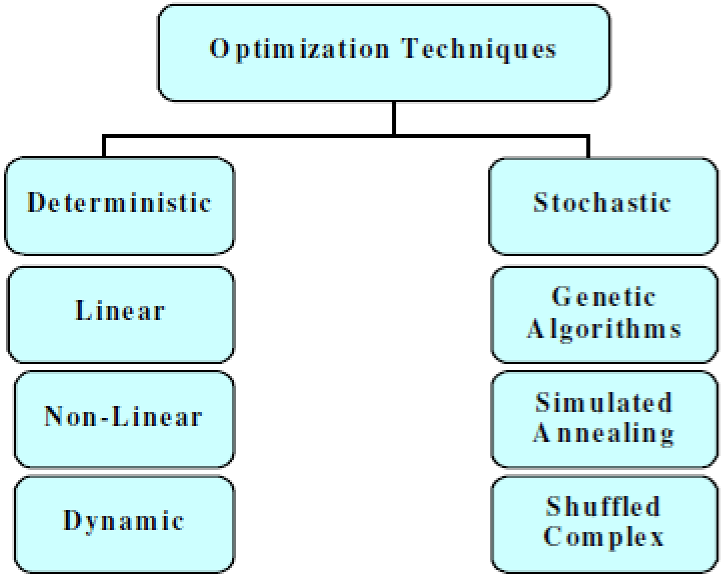


Fig. 3.1 Different methods of optimization in water distribution systems

An outline of some basic techniques involving deterministic algorithms for finding local minimum of multivariable functions whose influence are continuous and on which no restriction are imposed. For constrained problems, techniques are based on those for unconstrained problems. It should be considered that finding the global minimum is entirely different, and more challenging, problem which will not be addressed here. Basically stochastic methods are better suited at this time for large-scale global optimization.

3.2. Calibration Studies

Adjusting the selected parameters is essential during the construction of a calibrated hydraulic network by comparing their measured with the calculated values. If selected parameters of the network are pipe roughness and nodal demands, a procedure should be carried out in order to determine specific roughness values for the pipes and specific nodal demands for the nodes which will minimize the differences between measured and calculated values. The measured values are those which come up from the field, and the calculated are those from the hydraulic model. In the optimization-based models, the objective function which is the difference between the measured and calculated values is minimized while satisfying constraints, which describe the feasible solution (Ormsbee et al. 1989, Lansey and Basnet 1991).

Another issue in calibration is the decision of the sapling locations for making measurements of pressure. According to Walski(1983) , that measurements should be made near large demands and near the boundary of the pressure zone. Moreover, it was advised that sampling points should be away from sources. Nonetheless, it is difficult to decide for the exact location of sampling points, because in order to calibrate a hydraulic model, test data should have already been obtained, that means that sampling points' locations should be fixed beforehand. On the other side, calibration parameters can be obtained if the test location were determined. Admittedly, an iterative procedure should be realized, taking into account the sensitivity of the network.

The selection process of sampling points is accomplished by Bush and Uber at 1998, Pilled et al. 1999, Meier & Barkdoll 2000 among others.

3.3. Operation Studies

Energy costs from a large percentage of the total expenditure of the water utilities in general. It is essential to organize the operation of all the pumps to minimize energy consumption. Jowitt and Germanopoulos(1992) propose a linear programming model whereas Yu et al.(1994) and Percia et al.(1997) propose a nonlinear programming model. The most important benefit of linear programming is the possibility of using commercial software and finding global optimum; on the contrary, the loss of information through the linearization process is the most negative aspect.

3.4. Design Studies

In water distribution network design, the main problem aims in minimizing the whole network cost, since these systems are costly infrastructures. Nonetheless optimization of a water distribution system is quite complicated due to nonlinear relationships between parameters. Lately, significant amount of research has been performed on the optimal design of water distribution networks. One of the first computerized optimization study was accomplished by Schanke and Lai (1969), Walski (1985) reviewed approximately hundred studies concerding optimization since then. During the next fifteen years, another notable increase in this field was observed. Study of famous optimization problems such as New York tunnel problem (Schanke and Lai 1969), and various real systems (Jacobsen et al. 1998) were realized among others. The algorithms which were developed until then do not simulate the whole course of the design (Walski 2001). Reliability considerations developed so far were not applied realistically. Monetary constraints were not included and benefits were considered at all. Walski mentions that because of these reasons and unfriendly packaging of the related software, engineers continue to design using traditional tools.

A large amount of theories coupled with the availability of inexpensive powerful hardware, the development of theories in the last three decades has improved considerably the ability to simulate hydraulic behavior of large water distribution networks (Rossman et al.1993). An important role in layout, design and operation of water distribution systems is played by these models. A set of commercially available diameters is used to select the pipe diameters in order to form a water distribution network of least capital cost. The cost of maintenance and operation of a water distribution system may be noticeable, but still one of the main costs is that of the pipeline themselves. Lately, a number of optimization techniques have been developed primarily for the cost minimization aspect of network planning, although some reliability studies and stochastic modeling of demands have been tested.

Some of the first studies utilized which linear programming were performed by Alperovits and Shamir at 1977, Quindry et al. 1981, and Shamir and Howard 1985. Also, later studies applied nonlinear programming Su et al.1987, Lansey and Mays 1989, Xu and Goulter 1999 , or chance constrained approaches to the pipe network optimization problem. A large amount of recent literature has utilized genetic algorithms for the determination of low cost water distribution network design have been shown to have several advantages over more traditional optimization methods (Simpson et al.1994 , Savic and Walters 1997).

As for the determining of an optimal design for a water distribution network linear optimization methods have been widely studied. A method called linear programming gradient is used to design an optimal design of a water distribution system (Alperovits and Shamir 1977). Operation of the system under each of a set of demand loading is considered explicitly in the optimization.

3.5. Deterministic Techniques

The optimization tools in water distribution systems are considered to be the Linear Programming, Non-Linear Programming and Dynamic Programming which were first described by a lot of authors since 1966. The basic idea has been integrated into standard textbooks, and used by other authors in developing for applications.

Linear and Non-Linear programming refers to the area of applied mathematics dealing with the following problem: find numerical values for a given set of variables so that they are feasible and also a certain criterion called objective function, which depends on such variables , is optimized ,so that it is attains its minimum value among all the combination of feasible variables. On the other hand of classical problems of applied mathematics, most of which originate in physics, Linear and Non-Linear programming problems generally lack of solutions given by formulae, and be solved through numerical procedures, called algorithms, performed on computers. Plenty of features of the problem data call for different methods in order to achieve efficient solution to the problem. The analysis of the performance and the design of such algorithms are the backbone of the theoretical side of the area. Practically, the ability of these methods to solve very large problems has allowed for the modeling of highly realistic and detailed real life situations, so that nowadays these methods are routinely applied to the day-to-day execution of complex tasks in a wide range of activities.

3.6. Stochastic Techniques

A power search technique based on mechanics of population genetics is genetic algorithms. In chapter 4 Genetic Algorithms are explained in detail. Simulated Annealing combines between the steepest descent philosophy and the variable behavior within solution process, Cunha and Sousa 1999. Shuffled Complex Evolution deals with a set of population of points and searches in all directions within the feasible space based on objective function, Liong and Atiquzzaman 2004.

3.7. Enumeration Approach

The enumeration approach demands for all the possible system configurations and layouts to be explored, so that the optimum solution be found and met in the design of the piping networks with full consideration of all the elements that influence the system performance and cost. Thus, a global set of design components, including control devices, must be explored. The enumeration process involves the following steps:

- **i.** Development of all feasible designs for the given components.
- **ii.** Simulation of the performance of the different designs.
- **iii.** Evaluation of their performance.
- **iv.** Determination of costs for the feasible designs.
- **v.** Selecting the least-cost design.

3.8. Reliability of Water Distribution Systems

A water distribution system is defined as a one "including all water utility components for the distribution of finished or potable water by means of gravity storage feet or pumps through distribution-equalizing storage" by the American water works association in 1974. Cost of operation and cost of capital for first set up, maintenance and repair for the time water distribution network service to end users are large,and designers try to reduce total cost of system. On the other hard this is a very difficult process to obtain minimal cost solution for a water distribution system because of the large amount of parameters affecting cost. Designer must take some expected and unexpected loading conditions into consideration to ensure delivery of water to end user, during the optimization of a system. The most significant consideration in the design and operation of a water distribution system is to satisfy consumer needs under a range of desired quantity and quality during the systems' entire lifetime for the expected loading conditions. Furthermore, water distribution system must be able to accommodate abnormal conditions such as breaks in pipes, mechanical failure of pipes, valves, and control systems, power outages, malfunction of storage facilities and inaccurate demand projections. The possibility of occurrence of each of these deficiencies should be examined to determine the overall performance and thereby the reliability of the system. In general, reliability is defined as the probability that the system performs specified limits for a given period of time in a specified environment. As it is defined above reliability is ability of systems to provide adequate level of service to system consumers, under both normal and abnormal conditions. But there is still not convenient evaluation for water distribution system reliability as there are many measures of reliability. A review of the literature, Mays 1989 reveals that no universally acceptable definition or measure of the reliability of water distribution system is currently available.

3.9. Incorporation of Reliability in the Least-Cost Design of Looped Water Distribution Networks

Lately, considerable effort has been devoted to the development of optimization algorithms and models for the design of water distribution networks. A great amount of theories have the objective of minimizing the both capital and operating costs.(Alperovits and Shamir, 1977, Quindry et al. 1981, Shamir and Howard 1985, Lansey and Mays, 1989,Eiger et al. 1994,Simpson et al. 1994, Savic and Walters,1997) .Nevertheless the optimal design of a water distribution network is a complex multiple objective process involving trade-offs between the cost of the network and its reliability (Xu and Goulter 1999). Reliability incorporated optimization of water distribution systems requires combination of an optimization algorithm with a method for estimating reliability.

There isn't any well-defined meaning for the term "reliability" of water distribution networks. However, it is generally understood that reliability is concerned with the ability of the network to provide an adequate supply to the consumers, under both normal and abnormal operating conditions (Goulter 1995).

The first pronounced considerations of probabilistic issues in the reliability of water distribution networks were reported by Kettler and Goulter in 1983, who included the probability of pipe breakage as a constraint in an optimization model for the design of pipe networks. After three years they developed a quantitative approach to reliability measure in an optimized looped network. This approach starts by obtaining an “optimal” layout design through linear programming. Afterwards, approach addresses the probability of isolating a node through simultaneous failure of all links connected directly to that node. The probability of failure of individual links is modeled using the Poisson probability distribution.

3.10. Literature Review of Optimization techniques

3.10.1. Deterministic Optimization Techniques

The previous literatures involved in the branch of numerical optimization of piping networks using linear techniques (Fig. 3.2), details on the same figure will be deliberately mentioned in the following paragraphs.

The water distribution networks design has focused on mathematical approaches including linear, non-linear and dynamic programming (Bhave 1985, Alperovits and Shamir 1977, Quindry et al. 1981, Schaake and Lai 1969). On the other side, these deterministic methods cannot assure a global optimum solution. Furthermore, they necessitate that the functions satisfy certain limiting conditions that cannot be generally assured for a water distribution system.

Schaake and Lai focused in their studies on the mathematical approaches specifically in 1969. They used the linear programming in the design of the water distribution systems; nevertheless, these deterministic techniques cannot guarantee a global optimal solution as proven in past few years.

Author	Year	Optimization tool	Area of Concern
Schaake and Lai	1969	LP, NLP & DP	Design of WDS
Deb and Sarkar	1971	LP, NLP & DP	Design of WDS
Swamee and Khanna	1974	LP, NLP & DP	Design of WDS
Alperovits and Shamir	1977	LP, NLP & DP	Design of WDS
Quindry et al.	1981	LP, NLP & DP	Design of WDS
Featherstone and El-Jumaily	1983	NLP	Design of WDS
Bhave	1985	LP, NLP & DP	Design of WDS
Cenedese et al.	1987	Multi-objective reduced gradient	Closed hydraulic networks with pumping stations
Fujiwara and Dey	1987	LP	Two adjacent pipe diameters in water distribution system
Fujiwara et al.	1987	Modified LPG	Looped WDS
Fujiwara and Dey	1988	Lagrange multipliers	Water distribution network
Kessler and Shamir	1989	LPG	2-loop network
Lansey and Mays	1989	Generalized reduced gradient	Water distribution networks
Fujiwara and Khang	1990	2-phase decomposition	Water distribution networks
Bhave and Sonak	1992	LPG	Two and One loop networks
Eiger et al.	1994	Non smooth optimization & duality theory	2-loop, Hanoi and Real networks
Samani and Naeeni	1996	NLP	Water distribution networks
Taher and Labadi	1996	Franke-Wolfe algorithms using GIS	City of Greeley, Colo water distribution network
Berghout and Kuczera	1997	Iterative network NLP	Pressure reducing valve network
Sârbu and Borza	1997	LP	Looped water distribution networks
Sakarya and Mays	2000	Generalized reduced gradient	Water distribution pumps considering water quality
Djebedjian et al.	2000	NLP coupled with Newton-Raphson method	Design of WDS
Luong and Nagarur	2001	Semi-Markov process	WDS

LP- Linear Programming, NLP- Nonlinear Programming, DP- Dynamic-Programming, PSO- Particle Swarm Optimization and WDS- Water Distribution Systems.

Fig. 3.2 Literature review of WDS optimization in steady state

The same pipe diameter method for network optimization developed using the pressure surface profile capital cost functions for pipes, pumps and reservoirs (Deb and Sarkar 1971). It is proved that this method has two major disadvantages: first I lacks mathematical justification for cost equivalent pipes, and also a hydraulic pressure surface over the network must be artificially created (Swamee and Khanna 1974).

The majority of the work on the design of water distribution networks has focused on developing optimization procedures for the least cost pipe sizing problem. Optimization of the design for pipe networks is achieved by linear programming. Two principal approaches have been developed. The first one, has the ability to consider various components in a distribution network, nonetheless it is severely limited in the size of the system and the number of loads which it can handle (Alperovits and Shamir 1977).

An improvement of the method have been done by Quindry et al. 1981, that allows for a larger system to be considered, but difficulties arose when analyzing multiple loads. This method is limited because it considers only pipe portion, without taking into consideration any other component as pump, reservoir, etc.

A method that gets the minimum cost of a network by equating the first derivative of the total cost equation with zero presented for looped networks (Featherstone and El-Jumaily 1983).

A simple iterative procedure based on the identification of an efficient branched configuration applied by Bhave at 1985. The nodal heads for the branched configuration were treated as the design variables, and were initially taken so that each existing link needs strengthening by a new pipe parallel to the existing one, then given the maximum reduction in system cost.

Multi objective analysis is used for a mathematical approach to select the optimal solution for the hydraulic networks (Cenedese et al. 1987).

Each link at the optimal solution will not consist of at most two pipe segments with adjacent diameters as previously stated by other investigators, but a given set of pipe diameters is an optimal solution and a special case showed that the adjacency property holds if and only if pipe costs were a strictly convex function of a power of pipe diameters (Fujiwara and Dey 1987).

The linear programming gradient method developed by Alperovits and Shamir at 1977 modified (Fujiwara et al. 1987). A quasi-Newton search direction was used instead of the steepest descent direction, and the step size was determined by a backtracking line search method instead of a fixed step size. The results of a numerical example that the modified method applied shown an improved solution in comparison to the original linear programming gradient method.

Another method for design branched networks on flag terrain by using the Lagrange multipliers method to obtain optimal pipe size presented by Fujiwara and Dey in 1988. This method is limited to branched networks location flat terrain with a single source node and equal head for each end node.

Linear programming gradient used as an extension of the method by Alperovits and Shamir developed in 1977. It consists of two stages: Linear Programming is solved for a given flow distribution and then is conducted a search in the space of flow variables (Kessler and Shamir 1989). Afterwards, a two phase decomposition extending that of Alperovits and Shamir to non-linear modeling used by Fujiwara and Khang in 1990.

A chance constrained model for the minimum cost design of water distribution networks was presented by Lansey and Mays in 1989. Their methodology attempted to account for the uncertainties in required demands, required pressure heads, and pipe roughness coefficients. The problem of optimizing the system was formulated as a nonlinear programming model which was solved using a generalized reduced gradient method.

In 1992 Bhave and Sonak proved that the global optimal solution to a pipe network could be obtained by simplifying the network by canceling some pipes. Consequently, the network can be optimized using the linear programming approach. In that case, the solution has become similar to Alperovits and Shamir approach in 1977.

The Linear programming gradient method used as an extension of Alperovits and Shamir approach. This approach led to the determination of lengths of one or more segments in each link with discrete diameters (Eiger et al 1994).

A nonlinear optimization technique coupled with the Newton-Raphson method to minimize the design total cost with constraints in pipe diameters is proposed by Samani and Naeeni in 1996.

A prototype decision support system is presented by Taher and Labadie in 1966. This system (Water Distribution Optimization Program) is used to guide water distribution system design and analysis in response to changing water demands, timing, and use patters; and accommodation of new developments.

An iterative network linear programming algorithm for the hydraulic analysis of water networks is developed by Berghout and Kuczera in 1997. The new iterative technique used successive linear approximation to the nonlinear head loss equations. Primal dual and simplex nonlinear programming solvers in a hybrid scheme are used in order to reduce solution time to one-tenth of the solution time using the simplex nonlinear programming solver alone. A highly accurate solution was obtained using a smaller number of linear segments to represent each link content function.

Sarbu and Borza proposed a mathematical model and a numerical procedure for designing a pipe line network in 1997. The numerical procedure was developed for the purpose of computer simulation of natural gas line network based on the nonlinear system of equations.

In order to shape the optimal operation of water distribution system pumps and water quality considerations, Sakarya and Mays developed a new methodology in 2000. Based on a mathematical programming approach, the solution methodology resulted in a large-scale nonlinear programming problem that could not be solved using existing nonlinear codes

The optimization of water distribution systems in the steady state is investigated by Djebedjian et al. in 2000. They developed a mathematical model to design and evaluate the optimum network configuration in the steady state. Newton-Raphson method was used to solve the network equations in the nonlinear model.

A semi-Markov process to depict the behavior of the pipe, and replacement ages of the pipe in each of its deteriorating stages were taken as the decision variables by Luong and Nagarur in 2001. The model formulation converted to a linear problem by some simple transformations that resulted in the original nonlinear problem. Thus, numerical experiments were conducted in order to illustrate the applicability of the proposed model.

3.10.2. Genetic Algorithms

Methodologies for the application of genetic algorithms to the optimal design of water distribution network have been developed and published in 1990; The Fig3.3 shows the hierarchy researches developed using the GA technique.

Lately, the Genetic Algorith has become one of the most popular optimization choices for solving problems that are difficult for traditional deterministic optimization methods (Goldberg (1989); Simpson et al. (1994); and Dandy et al. (1996)). A huge advantage of GA is its ability to find the global optimum by using function values only.

In 1989, Goldberg applied Genetic Algorithms in the problem of pipe network optimization. In 1994, Simpson et al and Goldberg applied both simple genetic algorithm (SGA) and improved the genetic algorithm, with a variety of enhancements based on the nature of the problem, and reported talented solutions for problems from literature. Some problems associated with GA's are the doubt of termination of the search like all random search methods, the absence of guarantee for the global optimum.

Authors	Year	Optimization Method	Application
Goldberg	1989	GA	Pipe network optimization
Simpson et al.	1994	SGA	Pipe network optimization
Dandy et al.	1996	Improved GAs	New York network
Savic and Walters	1997	GAnet	2-loop, NYC and Hanoi networks
Reis et al.	1997	Simple GAs	Control valves in a WDS
Wu and Simpson	1997	Messy GAs	Two tanks, NYC & Moroccan networks
Castillo and González	1998	GAs	Examples
Murphy et al.	1998	GAs	Layout of Jamestown system
Abebe and Solomatine	1998	GA	2-loop and Hanoi networks
Gupta et al.	1999	GAs	2-loop network and network
Lippai et al.	1999	GA	New York tunnel problem
Wardlaw and Sharif	1999	GAs	Four and Ten reservoir system
Vairavamoorthy and Ali	2000	GAs	Hanoi and New York networks
Morley et al.	2001	GAnet	Open Net class hierarchy
Dandy and Engelhardt	2001	GAs	Rehabilitation of Water supply pipes in metropolitan Adelaide
Abdel-Gawad	2001	Improved GAs	NYC water supply tunnels
Wu and Simpson	2002	Messy GA	NYC water supply tunnels
van Vuuren	2002	GAs	Examples
van Zyl et al.	2004	Hybrid GA	Richmond WDS
Keedwell and Khu	2005	Hybrid GA	2-loop, A and B networks
Yu et al.	2005	GA and SA	Hypothetical WDS
Djebedjian et al.	2005b, 2006b	New adaptive penalty method for GA	WDS in steady state
Djebedjian et al.	2006a	GA	Design of WDS in steady state
Djebedjian et al.	2007	GA	WDS in steady state

Fig3.2 Previous Optimization studies using GA's

In 1996 an improved genetic algorithm formulation for pipe network optimization by using variable power scaling of the fitness function was developed by Dandy et al. The exponent introduced into the fitness function was increased in magnitude as the genetic algorithm computer run proceeds. In addition to the more commonly used bitwise mutation operator, an adjacency or creeping mutation operator was introduced. The application of the improved genetic algorithm formulation on the NYC tunnels indicated significantly better performance than the traditional GA formulation.

Savic and Walters (1997) developed a computer model GAnet which involved the application of standard genetic algorithm for the solution of nonlinear optimization problems. The application of the model to two networks, one for new design (two-loop network) and the other for rehabilitation of existing system (NYC water supply tunnels expansion problem) illustrated the potential of GAnet as a tool for water distribution network planning and management.

Different genetic-based search examples were compared by Wu and Simpson in 1997. Some of them were the standard genetic algorithm, the messy genetic algorithm, and the fast messy genetic algorithm for discrete optimization of pipeline networks. The application of the different genetic-based search techniques to optimization of pipeline problem indicated that the fast messy genetic algorithm was the most efficient algorithm among the genetic-based search paradigms. So this genetic algorithm provided a promising optimization algorithm for solving highly dimensional discrete optimization problems.

Reis et al. (1997) applied the genetic algorithm on a 37-pipe network to determine the appropriate location of three control valves and their settings to obtain maximum leakage reduction, as the objective function, for given nodal demands and reservoir levels. An optimal solution was reached in around 10 generations.

Genetic algorithm applied by Castillo and Conzalez in 1998, on a 16 nodes pipe network to find an optimal network features using a new problem-specific genetic operator. The final solution was feasible rather than optimal.

Murphy et al. (1998) applied the genetic algorithm to the Jamestown system expansion plan, Australia, to find alternative feasible combinations of new facilities given the constraints imposed on the system.

Abebe and Solomatine determined the optimal diameters of pipes in a network with a predetermined layout in 1998. The global optimization tool GLOBE with various random search algorithms and the network simulation model EPANET that can handle steady as well as dynamic loading conditions were used. The Hanoi network which contains 34 pipes, 31 demand nodes and a reservoir was used as a case study. GAs found the solution with lower 10% cost than Adaptive Cluster Covering with Local Search (ACCOL), but required 3 to 5 times more function evaluations (model runs).

The genetic algorithm was compared with the nonlinear programming technique based on interior penalty function with the Davidon- Fletcher-Powell method in the design of water distribution systems (Gupta et al., 1999).

A robust analysis and optimization on a water distribution network with four commercial optimizers was demonstrated by Lipai et al. in 1999. Intelligent search algorithms have very robust method for doing the optimization that retains the reality of the system. Intelligent search algorithms are robust in the sense that they can handle any kind of mathematical relationships, including lookup tables. They can also deal with discontinuities, nonconvexity, and other problems.

Wardlaw and Sharif (1999) applied the genetic algorithm for optimal four-reservoir system operation. They concluded that the most promising genetic algorithm approach comprises real-value coding, tournament selection, uniform crossover, and modified uniform mutation. For more complex reservoir systems, the genetic algorithm approach proved to be robust and easily applied than stochastic dynamic programming.

In 2000, Vairavamoorthy and Ali used genetic algorithm with strings coded by real values to avoid the problem of redundant states often found when using binary and Gray coding schemes, a fitness function which incorporated a variable penalty coefficient that depended on the degree of violation of the pressure constraints. This method did not demand solution of the nonlinear equations governing the flows and pressures in the distribution system for each individual member within the population. The application of the method on several networks showed a significant advantage compared with previously published techniques in terms of computational efficiency.

In 2001, Morley et al described an architecture for an integrated optimization application, GAnet, which included a GA application, a geographic information system GIS, and the EPANET hydraulic network solver.

The use of the genetic algorithm technique for finding a near optimal schedule for the replacement of the water supply pipes by minimizing the present value of capital, repair, and damage costs was established by Dandy and Engelhardt (2001). The application of the model on a case study in Adelaide, Australia showed that the genetic algorithm could be a powerful tool to help in planning the rehabilitation of water pipes.

In 2001, Abdel-Gawad tested many alternative formulations of genetic algorithm on the New York City water supply expansion. The results indicated that the most promising improved genetic algorithm approach for optimal design of pipe network problem comprises real-value coding, tournament selection, uniform crossover, and modified uniform mutation.

The self-adaptive boundary search strategy for selection of penalty factor within genetic algorithm optimization was introduced by Wu and Simpson in 2002 . The approach co-evolved and self-adapted the penalty factor such that the genetic algorithm search was guided towards and preserved around constraint boundaries. The strategy was tested on the New York City water tunnels problem and successfully found the least cost solution more effectively than a GA without the boundary search strategy. As a consequence, a reliable least cost solution was guaranteed for the GA optimization of a water distribution system.

A utility program (GAPOP) for demonstrating the application of GAs in the determination of the optimal pipe diameter in South Africa was developed by Van Vuuren (2002) . He concluded that GAs were potentially applied to hydrology and water resources assessment, network optimization, optimization of rehabilitation, extension and upgrading of distribution networks, and operation and maintenance scheduling of pumps and purification plants.

In 2004, Van Zyl et al based on the fact that genetic algorithm had good initial convergence attributes, but slow down considerably once the region of optimal solution had been identified, improved the efficiency of genetic algorithm operational optimization through a hybrid method which combined the GA method with a hill climber search strategy which complement GAs by being efficient in finding a local optimum. Two hill climber strategies were used, Hooke and Jeeves and Fibonacci methods. Upon an application on a hypothetical as well as an existing network in U.K., the hybrid method performed significantly better than the pure GA method, both in convergence speed and in the quality of the reached solutions.

In 2005, a novel method was proposed by Keedwell and Khu. This method is known as CANDA-GA, which used a heuristic-based, local representative cellular automata approach which provided a good initial population for genetic algorithm runs. The model performed well on large water distribution network problems, between 632 and 1277 pipes with no requirements for additional computations .Yu et al. (2005) introduced genetic algorithm combined with simulated annealing technology and self-adaptive crossover and mutation probabilities to deal with optimal allocation of water supply between pump-sources. Pump- station pressure head and initial tank water levels were considered as decision variables. The genetic algorithm was combined with the simulated annealing technology to overcome premature convergence of the genetic algorithm.

Djebedjian et al. (2005b, 2006b) A new adaptive penalty method for genetic algorithms was introduced. The model was applied to the two-loop network problem and the Hanoi network and the results showed that the least cost solution was obtained in a favorable number of function evaluations and was computationally much faster when compared to other studies.

Djebedjian et al. (2006a) applied the genetic algorithm along with the Newton method and the H-equations for hydraulic simulation to optimize pipe diameters of a large scale water distribution system. The numerical code was also capable of evaluating the network cost and was practically tested by application to the 389 pipe network of Suez City including 3 pumps and 3 reservoirs. The model showed attractive ability to handle the large scale pipe network optimization problems efficiently.

The evaluation of capacity reliability-based and uncertainty-based optimization in the water distribution systems was introduced from Djebedjian et al in 2007. The two approaches link the genetic algorithm (GA) as an optimization tool, Newton- Raphson technique as hydraulic simulation solver with the chance constraint in case of uncertainty-based or with Monte Carlo simulation in case of reliability- based optimization. For the first approach, optimal network design constrained by reliability for water distribution system analysis was formulated as an optimization problem under uncertainty. The optimal design problem was formulated as a chance constraint minimization problem restricted with a pre- specified level of uncertainty. The reliability of the system was then evaluated for the least-cost design of the network using the Monte Carlo simulation. For the second approach, network capacity reliability estimated using Monte Carlo simulation.

4. Genetic Algorithms

4.1. Introduction and Background

The genetic algorithm is an adaptive method which has a variety of uses in the solutions of search and optimization problems. Genetic algorithms are based on the genetic processes of biological organisms. Over many generations, natural populations evolve according to the principles of natural selection and “survival of the fittest”. By mimicking this process, genetic algorithms are able to evolve solutions to real world problems, if they have been suitably encoded. The basic principles of a genetic algorithm were first laid down rigorously in Holland in 1975.

As mentioned genetic algorithms work with a population of “individual”, each representing a possible solution to a given problem. Every individual is assigned a fitness score according to how good is the solution compare to the problem. Then the highly-fit individuals are given opportunities to reproduce, by cross breeding with other individuals in the population. This produces new individuals as offspring which share some features taken from each “parent”. Least fit members of the population are less likely to get selected for reproduction.

As a result, a whole new population of possible solutions is produced by selecting the best individuals from the current “generation”, and mating them to produce a new set of individuals. This new generation contains a higher proportion of the characteristics possessed by the good members of the previous generation. By this way, over many generations, good characteristics are spread throughout the population.

Also by favoring the mating of the more fit individuals, the most promising areas of the search space are explored. If the genetic algorithm has been designed well, the population will converge to an optimal solution to the problem. The Genetic Algorithm background was extracted from Busetti (2007).

4.2. The Method

4.2.1. Overview

First the evaluation function, or objective function, provides a measure of performance with respect to a particular set of parameters. The fitness function transforms that measure of performance into an allocation of reproductive opportunities. The evaluation of a string representing a set of parameters is independent of the evaluation of any other string. The fitness of that string, however, is always defined with respect to other members of the current population. In the genetic algorithm, fitness is defined by: f_i/f_A where f_i is the evaluation associated with string i and f_A is the average evaluation of all the strings in the population, Glodberg (1989).

The fitness can also be assigned based on a string's rank in the population or by sampling methods, such as tournament selection. The execution of the genetic algorithm is a two staged process. It starts with the current population. Then recombination and mutation are applied to the intermediate population to create the next population. In addition the process of going from the current population to the next population constitutes one generation in the execution of a genetic algorithm.

A characteristic example of a genetic algorithm representation is shown below:

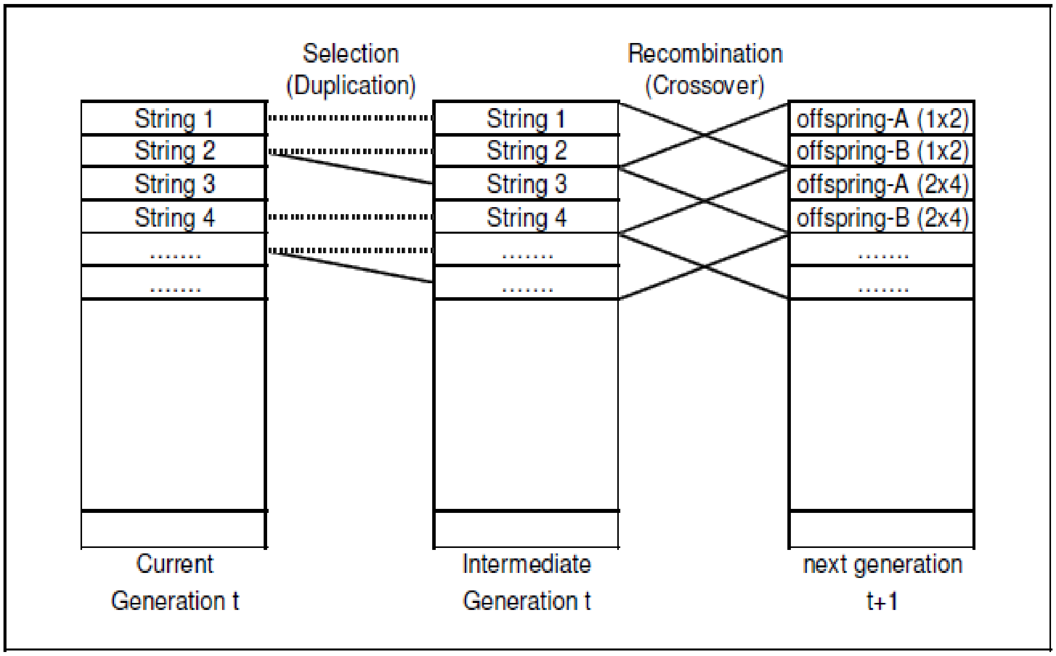


Fig 4.1. Standard genetic algorithm process schematic(Thesis* TEVFİK AKDOĞAN *April 2005)

As being presented in Fig 4.1., in the first generation the current population is also the initial population. After calculating f_i/f_A for all the strings in the current population, selection is carried out. The probability that strings in the current population are copied and placed in the intermediate generation is in proportion to their fitness.

Then individuals are chosen using "stochastic sampling with replacement" to fill the intermediate population. A selection process that will more closely match the expected fitness values is "remainder stochastic sampling." For each string I where f_i/f_A is greater than 1.0, the integer portion of this number indicates how many copies of that string are directly placed in the intermediate population. All strings (including those with f_i/f_A less than 1.0) then place additional copies in the intermediate population with a probability corresponding to the fractional portion of f_i/f_A. For example, a string with f_i/f_A = 1:36 places 1 copy in the intermediate population, and then receives a 0:36 chance of placing a second copy. A string with a fitness of f_i/f_A = 0:54 have a 0:54 chance of placing one string in the intermediate population. Remainder stochastic sampling is most efficiently implemented using a method known as stochastic universal sampling. Assume that the population is laid out in random order as in a pie graph, where each individual is assigned space on the pie graph in proportion to fitness. An outer roulette wheel is placed around the pie with N equally-spaced pointers. A single spin of the roulette wheel will now simultaneously pick all N members of the intermediate population.

So after selection has been carried out, the construction of the intermediate population is complete and the recombination can occur. This can be viewed as creating the next population from the intermediate population. Crossover is applied to randomly paired strings with a probability denoted p_c. The population should already be sufficiently shuffled by the random selection process. Then we pick a pair of strings. With probability p_c "recombine" these strings to form two new strings that are inserted into the next population.

For better understanding we present an example, consider the following binary string: 1101001100101101. The string would represent a possible solution to some parameter optimization problem. New sample points in the space are generated by recombining two parent strings. Consider this string 1101001100101101 and another binary string, yxyyxyxxyyyxyxxy, in which the values 0 and 1 are denoted by x and y. Using a single randomly chosen recombination point, 1 point crossover occurs as follows:

11010 \/ 01100101101

yxyyx /\ yxxyyyxyxxy

Swapping the fragments between the two parents produces the following offspring:

11010yxxyyyxyxxy and yxyyx01100101101

And then the recombination occurs and we can apply a mutation operator. Also for each bit in the population, mutate with some low probability p_m. Typically, the mutation rate is applied with 0.1% - 1% probability. After the process of selection, recombination and mutation is complete, the next population can be evaluated. The process of valuation, selection, recombination and mutation forms one generation in the execution of a genetic algorithm.

4.2.2. Coding of a Genetic Algorithm

Before being able to execute the genetic algorithm, a suitable coding must be invented. Another requirement also is the fitness function, which assigns a figure of merit to each coded solution. During the run, parents must be selected for reproduction, and recombined to generate offspring.

Every problem has a potential solution that is being represented as a set of parameters. These parameters known as genes are joined together to form a string of values which is referred as a “chromosome”.

For example, if our problem is to maximize a function of three variables, F(x; y; z), we might represent each variable by a 10-bit binary number (suitably scaled). Our chromosome would therefore contain three genes, and consist of 30 binary digits. The set of parameters represented by a particular chromosome is referred to as a genotype. The genotype contains the information required to construct an organism which is referred to as the phenotype. For example, in a bridge design task, the set of parameters specifying a particular design is the genotype, while the finished construction is the phenotype.

Also the fitness of an individual depends on the performance of the phenotype. This can be inferred from the genotype, i.e. it can be computed from the chromosome, using the fitness function. Assuming the interaction between parameters is nonlinear the size of the search space is related to the number of bits used in the problem encoding.

Genetic algorithms are often described as a global search method that does not use gradient information. Thus, no differentiable functions as well as functions with multiple local optima represent classes of problems to which genetic algorithms might be applied.

4.2.3. Fitness Function

In general a fitness function must be devised for each problem to be solved. Given a particular chromosome, the fitness function returns a single numerical "fitness," or "figure of merit," which is supposed to be proportional to the "utility" or "ability" of the individual which that chromosome represents. For many problems, particularly function optimization, the fitness function should simply measure the value of the function.

4.2.4. Reproduction

It is common that good individuals will probably be selected several times in a generation, though poor ones may not be at all. Having selected two parents, their chromosomes are recombined, typically using the mechanisms of crossover and mutation. The previous crossover example is known as single point crossover. Crossover is not usually applied to all pairs of individuals selected for mating. A random choice is made, where the likelihood of crossover being applied is typically between 0.6 and 1.0. If crossover is not applied, offspring are produced simply by duplicating the parents. This gives each individual a chance of passing on its genes without the disruption of crossover.

After crossover mutation is applied to each child individually. It randomly alters each gene with a small probability. In the diagram is being presented the fifth gene of a chromosome being mutated.

					Mutation Point ↓					
offspring	1	0	1	0	0	1	0	0	1	0
Mutated offspring	1	0	1	0	1	1	0	0	1	0

The traditional view is that crossover is the more important of the two techniques for rapidly exploring a search space. Mutation provides a small amount of random search, and helps ensure that no point in the search has a zero probability of being examined.

Table 4.1 genetic algorithm process example

Individual	Value	Fitness	Chromosome
Parent 1	0.08	0.05	00 01010010
Parent 2	0.73	0.000002	10 11101011
Offspring 1	0.23	0.47	00 11101011
Offspring 2	0.58	0.00007	10 01010010

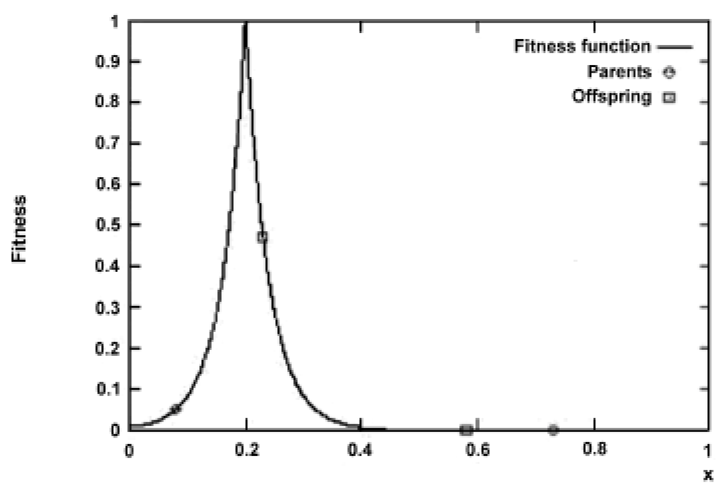


Fig 4.2. Standard genetic algorithm process curve

And so the fitness function is an exponential function of one variable, with a maximum at $x = 0.2$. It is coded as a 10-bit binary number. This illustrates how it is possible for crossover to recombine parts of the chromosomes of two individuals and give rise to offspring of higher fitness. (Crossover can also produce offspring of low fitness, but these will not be likely to get selected for reproduction in the next generation.)

4.2.5. Convergence

As we know the fitness of the best and the average individual in each generation increases towards a global optimum. Convergence is the progression towards increasing uniformity. A gene is said to have converged when 95% of the population share the same value. The population is said to have converged when all of the genes have converged. As the population converges, the average fitness will approach that of the best individual.

Eventually, a genetic algorithm will always be subject to stochastic errors. One such problem is that of genetic drift. Even in the absence of any selection pressure (i.e. a constant fitness function), members of the population will still converge to some point in the solution space. This happens simply because of the accumulation of stochastic errors. If, by chance, a gene becomes predominant in the population, then it is just as likely to become more predominant in the next generation as it is to become less predominant. If an increase in predominance is sustained over several successive generations, and the population is finite, then a gene can spread to all members of the population.

Once a gene has converged in this way, it is fixed crossover cannot introduce new gene values. This produces a ratchet effect, so that as generations go by, each gene eventually becomes fixed.

The rate of genetic drift therefore provides a lower bound on the rate at which a genetic algorithm can converge towards the correct solution. That is, if the genetic algorithm is to exploit gradient information in the fitness function, the fitness function must provide a slope sufficiently large to counteract any genetic drift. The rate of genetic drift can be reduced by increasing the mutation rate. However, if the mutation rate is too high, the search becomes effectively random, so once again gradient information in the fitness function is not exploited.

4.3. Comparisons Of a Genetic Algorithm

4.3.1. Advantages Of a Genetic Algorithm

Even though, the power of genetic algorithms comes from the fact that the technique is robust and can deal successfully with a wide range of difficult problems, they are not guaranteed to find the global optimum solution to a problem, but they are generally good at finding "acceptably good" solutions to problems "acceptably quickly". Where specialized techniques exist for solving particular problems, they are likely to outperform genetic algorithms in both speed and accuracy of the final result.

Even where existing techniques work well, improvements have been made by hybridizing them with a genetic algorithm. The basic mechanism of a genetic algorithm is so robust that, within fairly wide margins, parameter settings are not critical.

4.3.2. Disadvantages Of a Genetic Algorithm

An issue with genetic algorithms is that the genes from a few comparatively highly fit (but not optimal) individuals may rapidly come to dominate the population, causing it to converge on a local maximum. Once the population has converged, the ability of the genetic algorithm to continue to search for better solutions is effectively eliminated: crossover of almost identical chromosomes produces little that is new. Only mutation remains to explore entirely new ground, and this simply performs a slow, random search.

4.3.3. Comparison of Genetic Algorithm with other Methods

It is common for any efficient optimization algorithm to use two techniques to find a global maximum exploration to investigate new and unknown areas in the search space, and exploitation to make use of knowledge found at points previously visited to help find better points. These two requirements are contradictory, and a good search algorithm must find a tradeoff between the two.

Neural Nets

Genetic algorithms and neural nets are both adaptive, learn, can deal with highly nonlinear models and noisy data and are robust, "weak" random search methods. They do not need gradient information or smooth functions. In both cases their flexibility is also a drawback, since they have to be carefully structured and coded and are fairly application-specific. For practical purposes, they appear to work best in combination: neural nets can be used as the prime modeling tool, with GAs used to optimize the network parameters.

Random Search

In this method the brute force approach for difficult functions is a random, or an enumerated search. Points in the search space are selected randomly, or in some systematic way, and their fitness evaluated. This is a very unintelligent strategy, and is rarely used by itself.

Gradient Methods

Numerous different methods for optimizing well-behaved continuous functions have been developed which rely on using information about the gradient of the function to guide the direction of search. If the derivative of the function cannot be computed, because it is discontinuous, for example, these methods often fail. Such methods are generally referred to as hill climbing. They can perform well on functions with only one peak (unimodal functions). But on functions with many peaks, (multimodal functions), they suffer from the problem that the first peak found will be climbed, and this may not be the highest peak. Having reached the top of a local maximum, no further progress can be made.

Iterated Search

This is a random and gradient search which may be combined to give an iterated hill climbing search. Once one peak has been located, the hill climb is started again, but with another, randomly chosen, starting point. This technique has the advantage of simplicity, and can perform well if the function does not have too many local maxima. However, since each random trial is carried out in isolation, no overall picture of the "shape" of the domain is obtained. As the random search progresses, it continues to allocate its trials evenly over the search space.

And this means that it will still evaluate just as many points in regions found to be of low fitness as in regions found to be of high fitness. A genetic algorithm, by comparison, starts with an initial random population, and allocates increasing trials to regions of the search space found to have high fitness. This is a disadvantage if the maximum is in a small region, surrounded on all sides by regions of low fitness. This kind of function is difficult to optimize by any method, and here the simplicity of the iterated search usually wins.

Simulated Annealing

A modified version of hill climbing. Starting from a random point in the search space, a random move is made. If this move takes us to a higher point, it is accepted. If it takes us to a lower point, it is accepted only with probability $p(t)$, where t is time. The function $p(t)$ begins close to 1, but gradually reduces towards zero, the analogy being with the cooling of a solid. Initially therefore, any moves are accepted, but as the "temperature" reduces, the probability of accepting a negative move is lowered.

Negative moves are essential sometimes if local maxima are to be escaped, but obviously too many negative moves will simply lead us away from the maximum. Like the random search, however, simulated annealing only deals with one candidate solution at a time, and so does not build up an overall picture of the search space. No information is saved from previous moves to guide the selection of new moves.

4.4. Suitability Of a Genetic Algorithm

Some of the most traditional genetic algorithm research has concentrated in the area of numerical function optimization. Genetic algorithms have been shown to be able to outperform conventional optimization techniques on difficult, discontinuous, multimodal, noisy functions.

These characteristics are typical of market data, so this technique appears well suited for our objective of market modeling and asset allocation. For asset allocation, combinatorial optimization requires solutions to problems involving arrangements of discrete objects. This is quite unlike function optimization, and different coding, recombination, and fitness function techniques are required.

They have been created many applications of genetic algorithms to learning systems, the usual paradigm being that of a classifier system. The genetic algorithm tries to evolve (i.e. learn) a set of "if : : : then" rules to deal with some particular situation. This has been applied to economic modeling and market trading, Deboeck (1994).

4.5. Practical Implementation of the Algorithm

4.5.1. Fitness Function

In addition to the coding scheme used, the fitness function is the most crucial aspect of any genetic algorithm. Ideally the fitness function should be smooth and regular, so that chromosomes with reasonable fitness are to chromosomes with slightly better fitness.

They should not have too many local maxima, or a very isolated global maximum. It should reflect the value of the chromosome in some "real" way, but unfortunately the "real" value of a chromosome is not always a useful quantity for guiding genetic search. In combinatorial optimization problems, where there are many constraints, most points in the search space often represent invalid chromosomes and hence have zero "real" value. Another approach which has been taken in this situation is to use a penalty function, which represents how poor the chromosome is, and construct the fitness as (constant-penalty).

Penalty functions which represent the amount by which the constraints are violated are better than those which are based simply on the number of constraints which are violated. Approximate function evaluation is a technique which can sometimes be used if the fitness function is excessively slow or complex to evaluate. A genetic algorithm is robust enough to be able to converge in the face of the noise represented by the approximation. Approximate fitness techniques have to be used in cases where the fitness function is stochastic.

4.5.2. Presentation of parent Selection Techniques

The task of allocating reproductive opportunities to each individual is parent selection. In principle, individuals from the population are copied to a "mating pool", with highly fit individuals being more likely to receive more than one copy, and unfit individuals being more likely to receive no copies. Under a strict generational replacement, the size of the mating pool is equal to the size of the population. After this, pairs of individuals are taken out of the mating pool at random, and mated. This is repeated until the mating pool is exhausted.

The behavior of the GA very much depends on how individuals are chosen to go into the mating pool. Ways of doing this can be divided into two methods:

- **Explicit Fitness Remapping**

For being able to keep the mating pool the same size as the original population, the average of the number of reproductive trials allocated per individual must be one. If each individual's fitness is remapped by dividing it by the average fitness of the population, this effect is achieved.

Each individual's remapped fitness will, in general, not be an integer. Since only an integral number of copies of each individual can be placed in the mating pool, we have to convert the number to an integer in a way that does not introduce bias. A widely used method is known as stochastic reminder sampling without replacement. A better method, stochastic universal sampling is elegantly simple and theoretically perfect. It is important not to confuse the sampling method with the parent selection method. Different parent selection methods may have advantages in different applications. But a good sampling method is always good, for all selection methods, in all applications.

- **Implicit Fitness Remapping**

In general implicit fitness remapping methods fill the mating pool without passing through the intermediate stage of remapping the fitness.

As an example, in a binary tournament selection, pairs of individuals are picked at random from the population. Whichever has the higher fitness is copied into a mating pool (and then both are replaced in the original population). This is repeated until the mating pool is full. Larger tournaments may also be used, where the best of *n* randomly chosen individuals is copied into the mating pool. Using larger tournaments has the effect of increasing the selection pressure, since below-average individuals are less likely to win a tournament and vice-versa.

4.5.3. Other Crossovers Techniques

4.5.3.1. Technique of two Point Crossover

The issue with adding additional crossover points is that building blocks are more likely to be disrupted. However, an advantage of having more crossover points is that the problem space may be searched more thoroughly. In 2-point crossover, (and multi-point crossover in general), rather than linear strings, chromosomes are regarded as loops formed by joining the ends together. To exchange a segment from one loop with that from another loop requires the selection of two cut points, as indicated in Fig. 4.3.

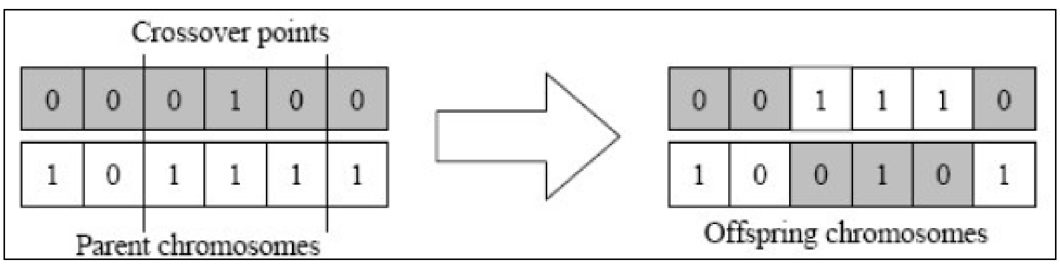


Fig. 4.3. Two point crossover

4.5.3.2. Technique of Uniform Crossover

The uniform crossover is radically different to 1-point crossover. Each gene in the offspring is created by copying the corresponding gene from one or the other parent, chosen according to a randomly generated crossover mask. Where there is a 1 in the crossover mask, the gene is copied from the first parent, and where there is a 0 in the mask, the gene is copied from the second parent, as shown below in Fig 4.4:

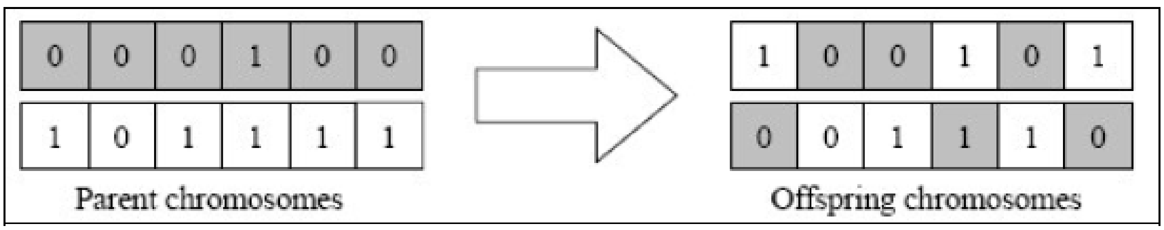


Fig. 4.4. Uniform crossover

4.5.4. Inversion and Reordering

Placement of genes on a chromosome is critical for the method to work effectively. Techniques for reordering the positions of genes in the chromosome during a run have been suggested. One such technique, inversion, works by reversing the order of genes between two randomly chosen positions within the chromosome. Reordering does nothing to lower epistasis (see 4.5.6), but greatly expands the search space. Not only is the genetic algorithm trying to find good sets of gene values, it is simultaneously trying to discover good gene orderings too.

4.5.5. Meaning of Epistasis

The meaning of epistasis is the interaction between different genes in a chromosome. It is he extent to which the "expression" (i.e. contribution to fitness) of one gene depends on the values of other genes. The degree of interaction will be different for each gene in a chromosome. If a small change is made to one gene we expect a resultant change in chromosome fitness. This resultant change may vary according to the values of other genes.

4.5.6. The Process of Mutation and Naïve Evolution

Even though mutation is traditionally seen as a "background" operator, responsible for reintroducing alleles or inadvertently lost gene values, preventing genetic drift and providing a small element of random search in the vicinity of the population when it has largely converged. It is generally held that crossover is the main force leading to a thorough search of the problem space. “Naive evolution" (just selection and mutation) performs a hill climb-like search which can be powerful without crossover.

However, mutation generally finds better solutions than a crossover only regime. Mutation becomes more productive, and crossover less productive, as the population converges. Despite its generally low probability of use, mutation is a very important operator.

4.5.7. Niche and Speciation Mechanisms

The process of speciation is where a single species differentiates into two (or more) different species occupying different niches. In a genetic algorithm, niches are analogous to maxima in the fitness function. Sometimes we have a fitness function which is known to be multimodal, and we may want to locate all the peaks. Unfortunately a traditional genetic algorithm will not do this the whole population will eventually converge on a single peak.

This is due to genetic drift. The two basic techniques to solve this problem are to maintain diversity, or to share the payoff associated with a niche. In pre-selection, offspring replace the parent only if the offspring's fitness exceeds that of the inferior parent. There is fierce competition between parents and children, so the payoff is not so much shared as fought over, and the winner takes all. This method helps to maintain diversity (since strings tend to replace others which are similar to themselves) and this helps prevent convergence on a single maximum.

4.5.8. Restricted Mating of Genes

The goal of restricted mating is to encourage speciation, and reduce the production of lethals. A lethal is a child of parents from two different niches. Although each parent may be highly fit, the combination of their chromosomes may be highly unfit if it falls in the valley between the two maxima. The general philosophy of restricted mating makes the assumption that if two similar parents (i.e. from the same niche) are mated, then the offspring will be similar.

However, this will very much depend on the coding scheme and low epistasis. Under conventional crossover and mutation operators, two parents with similar genotypes will always produce offspring with similar genotypes. However, in a highly epistatic chromosome, there is no guarantee that these offspring will not be of low fitness, i.e. lethals.

Total reward available in any niche is fixed, and is distributed using a bucket-brigade mechanism. In sharing, several individuals which occupy the same niche are made to share the fitness payoff among them. Once a niche has reached its "carrying capacity", it no longer appears rewarding in comparison with other, unfilled niches.

4.5.9. Diploids and Dominance

Diploidy is when chromosomes contain two sets of genes, rather than just one. Virtually all work on genetic algorithms concentrates on haploid chromosomes. This is primarily for simplicity, although use of diploid chromosomes might have benefits. Diploid chromosomes lend advantages to individuals where the environment may change over a period of time. Having two genes allows two different "solutions" to be remembered, and passed on to offspring. One of these will be dominant (that is, it will be expressed in the phenotype), while the other will be recessive. If environmental conditions change, the dominance can shift, so that the other gene is dominant.

This shift can take place much more quickly than would be possible if evolutionary mechanisms had to alter the gene. This mechanism is ideal if the environment regularly switches between two states.

4.6. Importance of Optimization in the Water Sector

Numerous optimization approaches, some general and others specific, have evolved in order to achieve economy of design, construction, operation and maintenance of these systems. However, it is enough to say that most of the pipeline optimization methodology is concerned with the optimization of systems under steady or nearly steady flow conditions.

Consideration of transients often takes place after assuming that the cost of controlling transients represents a small portion of the overall pipeline cost. There are important feedback mechanisms between the steady and transient portion of an optimized system. Optimal design of distribution systems has been approached from many angles and using a number of optimization tools.

The challenges in the water industry in developed countries and the world at large, together with the capital constraints and operational cost escalation, necessitate the evaluation of technical, economic and environmental parameters to reach an optimal solution.

In the water sector it has been indicated that large savings can be accomplished if optimal solutions are implemented, when new systems are designed or when existing systems are refurbished or extended. The need for the application of optimization techniques stems from the fact that the selection of system components to be evaluated in a water system is dependent on a number of inter-dependent variables.

For example, if an optimal diameter has to be determined it is known that by reducing the diameter the capital cost is reduced but the operating cost (pumping) will escalate and the possibility of pipe burst due to surge pressures associated with high-flow velocities will increase.

Genetic algorithms can be employed with great advantage in many sectors:

- Optimization of pipe diameters.
- Identification of pipe segments in a distribution network that should be rehabilitated to improve the performance of the system.
- Determining the application of a phased development of infrastructure for different development horizons.
- Efficient development of infrastructure for alternative service levels ensuring an affordable service.
- Optimization of reservoir sizes and determination of the required pump capacities.
- Optimization of operational scheduling.

The genetic algorithms are suited to solve problems that are not susceptible to attack by enumerative methods because the sheer number of potential solutions defies the possibility of testing them all. Such problems are typically multi constrained, that is the solution must be a balance of conflicting or synergistic properties.

When considering a problem with multiple dependencies you are normally forced to admit the possibility of isomeric solutions, i.e. solutions that give the same result using different processing routes. So for some problems there is no such thing as the “best solution”, but instead one looks or members of a fuzzy set of solutions that can be defined as “good enough”.

The solution to a problem can be the best, but can be lost in a vast result space of complex problems. If the solution space is limited then enumerative techniques can work. One of the great strengths of GAs is that they do not have to evaluate all the possible solutions.

This means that increasing the number of possible solutions has little impact on the running time of a genetic algorithm. Goldberg (1989) indicated that a GA differs from the traditional search methods in the following ways:

- Genetic algorithms work with coding of the parameter set, not the parameters themselves.
- A genetic algorithm can evaluate a population of points, not a single point.
- Genetic algorithms use objective function information, not derivatives or other auxiliary knowledge, to determine the fitness of the solution.
- Also genetic algorithms use probabilistic transition rules, not deterministic rules in the generation of the new populations.

The genetic algorithms use bit-strings to represent the state and characteristics of an object model. Changing the values of the bits in these bit-strings can be translated back into changes to the associated objects’ data.

Once the object has been converted (coded) into bit-strings, the genetic algorithm program (coder) can apply biologically analogous processes such as replication (or reproduction), crossover and mutation to the bit-strings, which can then be translated back to the objects themselves.

In this way the genetic algorithm coder can evolve the instance-state of the components within an object model to obtain a bit-string (solution) with a high fitness. Changes to the bit-string values can be accomplished through the process of reproduction, cross-over and mutation.

4.7. Summary

One of the most important advantages of genetic algorithms is their flexibility and robustness as a global search method. They can deal with highly nonlinear problems and no differentiable functions as well as functions with multiple local optima. They are also readily amenable to parallel implementation, which renders them usable in real-time.

Genetic algorithms have been applied on a number of different real problems and have resulted in exciting, but not always straightforward solutions. In complex water distribution systems, for instance, the alternative options when evaluating the extensions to water supply systems become numerous. They provide procedures for the evaluation of the optimal solutions in the solution space.

For the designing a competent GA, the objective is to develop a genetic algorithm that can solve problems with bounded difficulty and exhibit a polynomial (usually sub quadratic) scale-up with the problem size. Based on these goals, design decomposition has been proposed elsewhere, Sastry (2001).

5. Genetic Algorithm Based Optimization method

5.1. Introduction

The main focus of this chapter is to present a novel heuristic optimization technique which gives near optimal pump schedules for real size and complex water distribution systems. The only significant simplification of the method is that it considers the water consumptions as known, deterministic data (expected values are used). Any other effect can be taken into account: mixed-type variables, non-linearity through coupled hydraulic simulations, and large variety of constraints.

The basic idea behind the genetic algorithms is to simulate the natural selection, the survival of the fittest (Darwin 1859). Although these algorithms perform well on numerous problems, usually highly specialized novel methods are required to avoid the most common drawbacks of genetic algorithms, such as the too early convergence.

Every evolutionary method aims to find a good trade-off between the exploration and exploitation search phases in order to achieve good performance on optimization problems.

These phases can be directly linked to population diversity. During the exploration phase the diversity is high and the algorithm identifies the potential regions of the search space where the optimum can be located. In the exploitation phase a *local search* around a potential optimal solution candidate is performed resulting a gradual decline in population diversity. Maintaining balance between these two phases is crucial to reach good performance on search problems by evolutionary algorithms.

The objective functions of a water distribution system can be various. Investments and operational costs can be optimized in the planning or renovation phases. The other possibility, which is the most common, is using only objectives of the water network, which are related to the operation. Then the system is considered as it is, meaning that topological changes cannot be performed.

Pumping of treated water represents the major fraction of the total operation cost in conventional water supply systems and even a small improvement in operational efficiency can cause significant cost savings to the industry.

The need for optimization is twofold: it is required either at the design stage of waterworks or more frequently the demand focuses towards on operational level: having a given waterworks topology, one aims to achieve an optimal control of the active hydraulic elements (pumps, valves) satisfying water demand with minimal energy consumption. Sophisticated operation can result in significant savings, even in small scale waterworks.

If sufficient storage capacity is available, the water demand can be satisfied with a large number of pump schedules. As the energy consumption charge changes during the day or the specific energy consumption of the pumps are different, different overall energy charges correspond to pump schedules and thus it is beneficial to find and realize the most cost-effective one.

5.2. The Problem to be Solved

The purpose of a water distribution network is to deliver water to consumers with appropriate quality, quantity and pressure. A distribution system is used to describe collectively the facilities used to supply water from its source to the point of usage.

The developed solver is marked to solve problems on large scale, real systems. Hence, we performed a variation of tests through Simulink on the case study of Vlite water distribution network from which we obtained several significant conclusions for the applicability of our method. The image below depicts a sample of a water distribution network which helped us to establish the main idea behind our optimization technique.

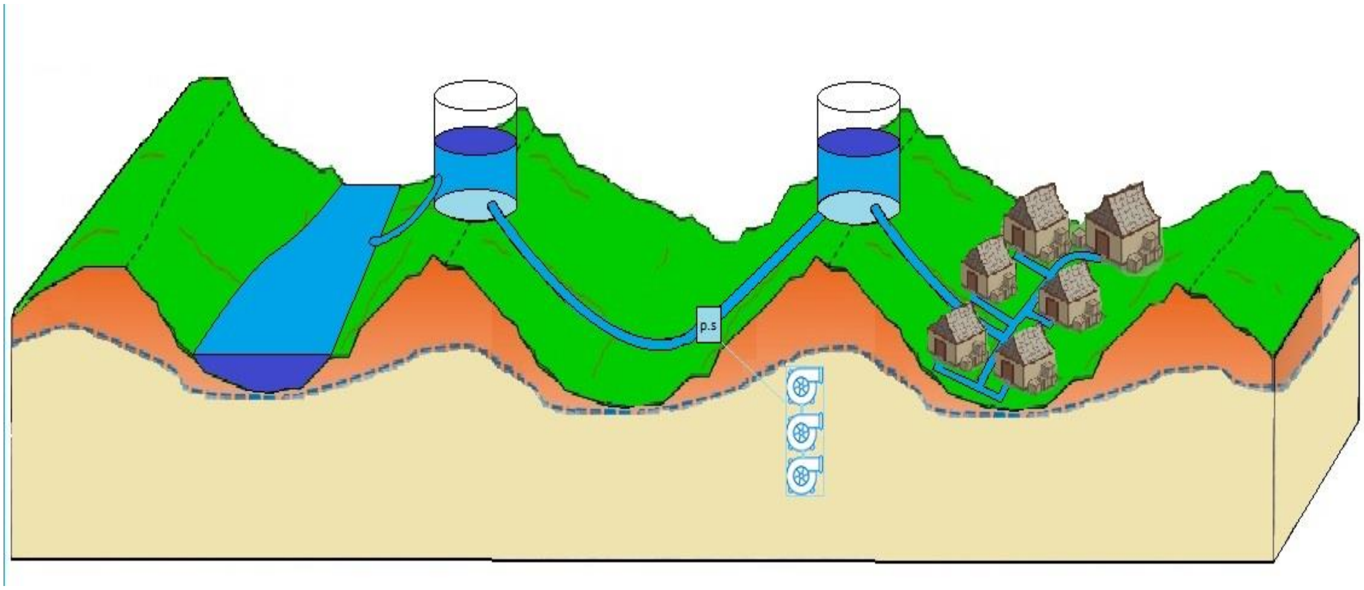


Fig 5.1. An example of a water distribution system.

At start the basic idea underlying pump (and valve) schedule optimization is that the water consumptions can be satisfied by several different schedules. The electric energy used by the pumps is the largest part of the total electricity bill of waterworks. Therefore, the total electric cost of the pumps over a finite time horizon is used as the most common objective function.

Our specific goal is to minimize the total electric cost of the pumps by optimizing the number of switches between the pumps so that they always operate at the minimum cost.

Some of the key questions of the optimizations are how to exploit the storage capacity of the reservoirs in order to decrease the electrical expenses and how to find an optimal schedule within reasonable time. Computational demand and time play a significant role since operators need to generate new schedules in minutes in real-life circumstances.

More specifically observing the picture above we can understand the basics behind the optimal operation of the pump stations, which as we described we will only operate on their cost efficient point.

For instance if the demand is higher than a single pump station's minimum cost operation point then and only then we will use a second pump station, and if the second pump station reaches also its cost efficient operation point we will use the third one. By this way we are able to preserve a great amount of energy and minimize the cost.

In this thesis we argue with the possibility of decreasing the expenses by reducing the power consumption itself. In this case, the specific energy consumption of the pumps is a good quantity to describe the thrift of the system. Energy can be saved by using the pumps which have lower specific consumption values or using the pumps close to their best efficiency points, which are determined by the revolution number and the state of the whole system. In these cases, the storage capacity is also essential: it allows to store the spare water if the pumps deliver more water in their efficient operating points than needed.

5.2.1. Requirements on the water network operation

Even though the above mentioned rules of thumb seem obvious, determining the optimal schedule is a highly challenging task due to the constraints of the system. Some of the objective functions can be transformed into constraints e.g. the switching number of the pumps, the maximum power peak of pump groups, water level variations, and water quality properties.

Besides, the capacity limits of the reservoirs, the exploiting limits of wells, nodal pressure limitations make the optimization problem even more complex.

5.2.2. Model definition

During optimization procedures, in order to yield reliable results, a mathematical model of the desired system must be established. The quality of the system modeling sets the bound for the performance of the optimization technique and the reliability of the results. Usually the modeling of physical systems is a process that has several limitations. Although, a system can be extensively modeled in order to include many of the physical laws, restrictions and randomness that appear within the system and that would give a complex system requiring sizeable computer effort in order to be compiled.

The standard technique is to model the basic laws and limitations of a system within specific bounds. Some admissions can be made in order to approximate several behaviors of the system which otherwise would be too complex to be calculated or even measured. Omitting parts of the system when their effect on the specific parameters the user is trying to examine is not rare at all. Also, sometimes grouping several parts of a system which display great complexity, in order to be replaced with a generic block that shows similar effects in the examined system, is accepted.

The current model definition started with a simple approach where there is a source, which is a tank, with enough water to supply a watering system. This source supplies another tank with the use of a pump station that in this simple version only has two pumps inside. There is one pump for each of the categories of pumps that are used in this study. The demand of the system is not limited so the pumps will deliver the flow at the operation point that those are calibrated. Each of the pumps has its specific pressure - volume curve stored inside the block and several blocks that help with its operation, measurement and calibration of the operation point.

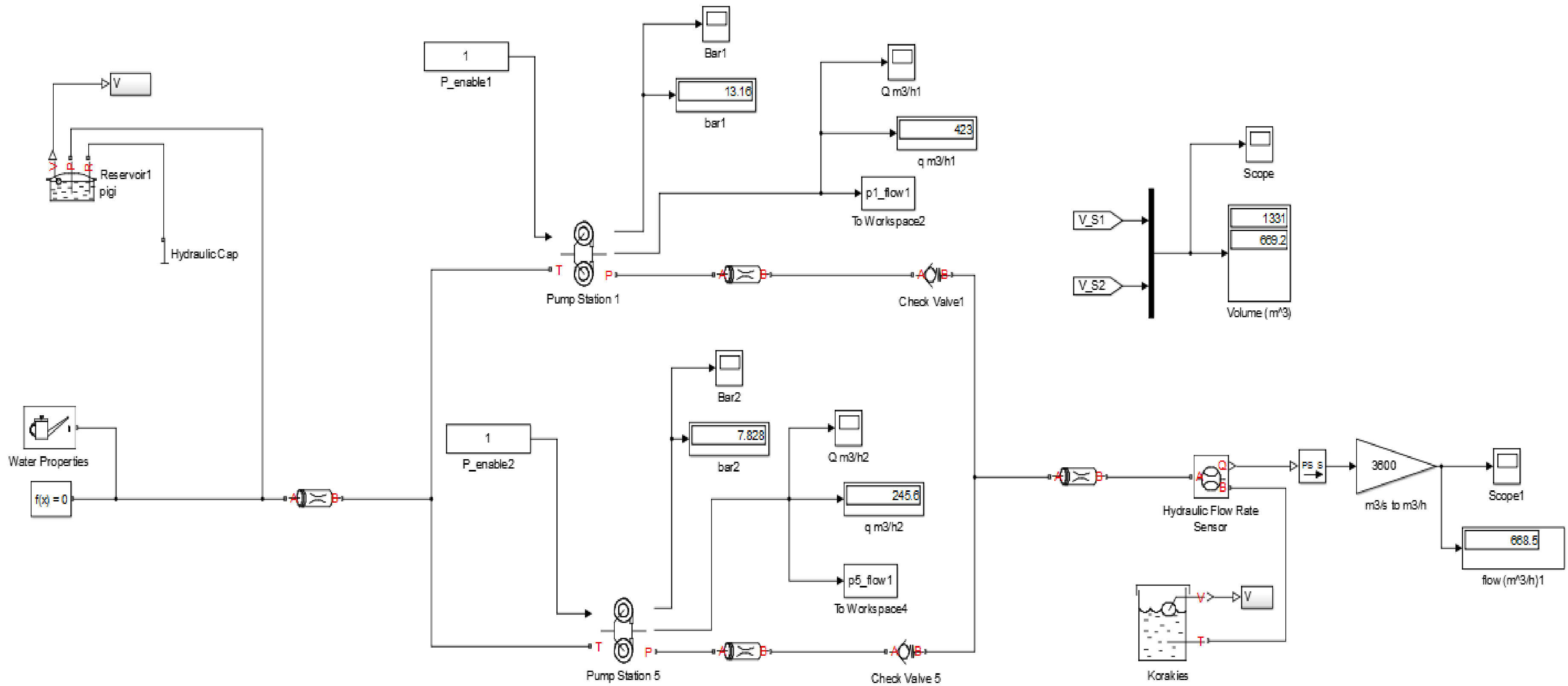


Fig 5.2. The Basic Simulation Model

The picture below shows the first version of the model. Some of the blocks used are necessary for the operation of the Matlab hydraulics simscape library. The water properties and the solver blocks are used with the settings for water and default solver properties. However in the Simulink configuration parameters menu the 15s stiff differential equation solver was selected and the maximum simulation step was set to 900 sec due to the large simulation times used here.

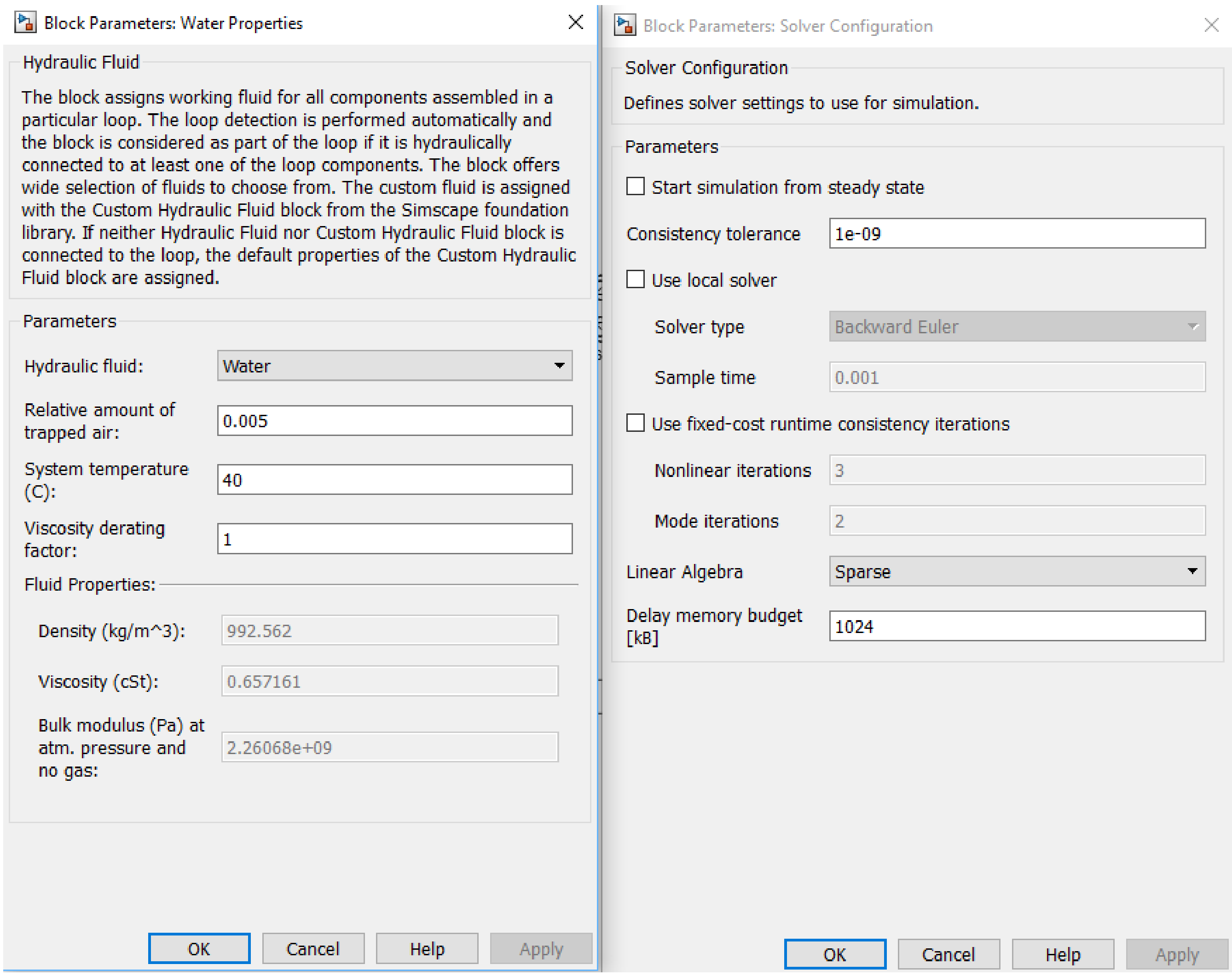


Fig 5.3. Model Parameters

The reservoir and variable head tank are similar blocks that are modeling tanks. In this topology both can be used. Both of these blocks have an output for the volume inside each tank which is driven into Mux and a scope for observation reasons. Using the reservoir block, as with all the hydraulic blocks require all its ports connected, so a hydraulic cap was connected in the return port of the block. The block that models the source was selected to have an initial volume of 2000m^3 while the other tank was considered empty. Inlet diameters were deliberately chosen to be high in order to avoid cavitation and overpressure problems in the inlets. The variable korakies_cs is a Matlab calculated variable based on the diameter of the tank.

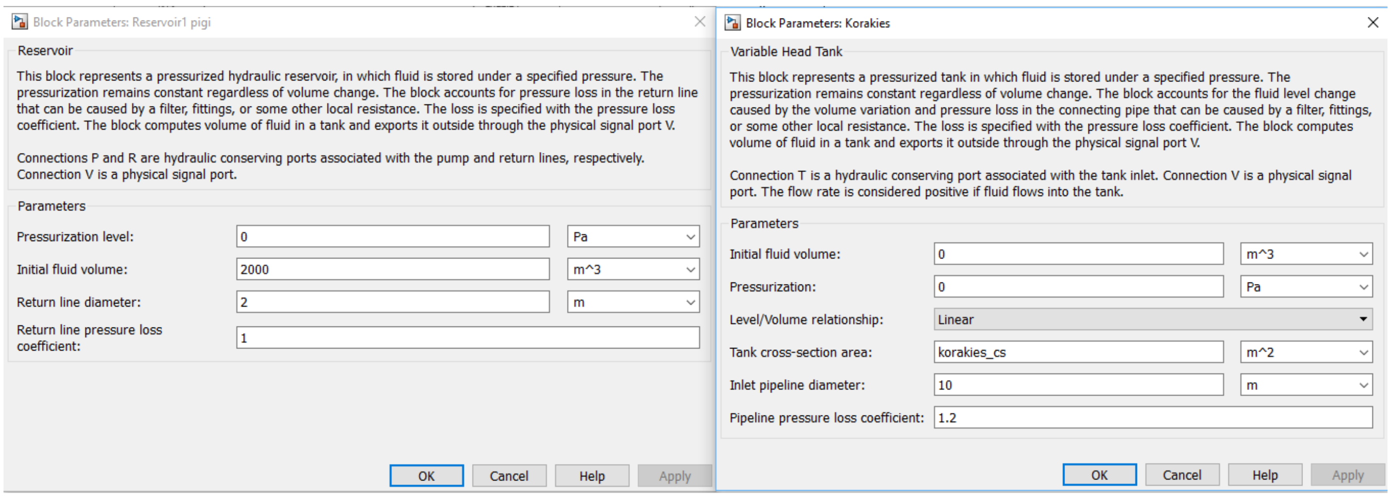


Fig 5.4 Model Parameters

The transportation of water from the source to the pumps and from the pumps to the second tank is implemented through pipes which are set with variables for shape, dimensions, length, friction coefficients and head. The water source is supposed to be in 212 m height the pump station is located in an altitude of 83 m and the second tank is also on 212 m.

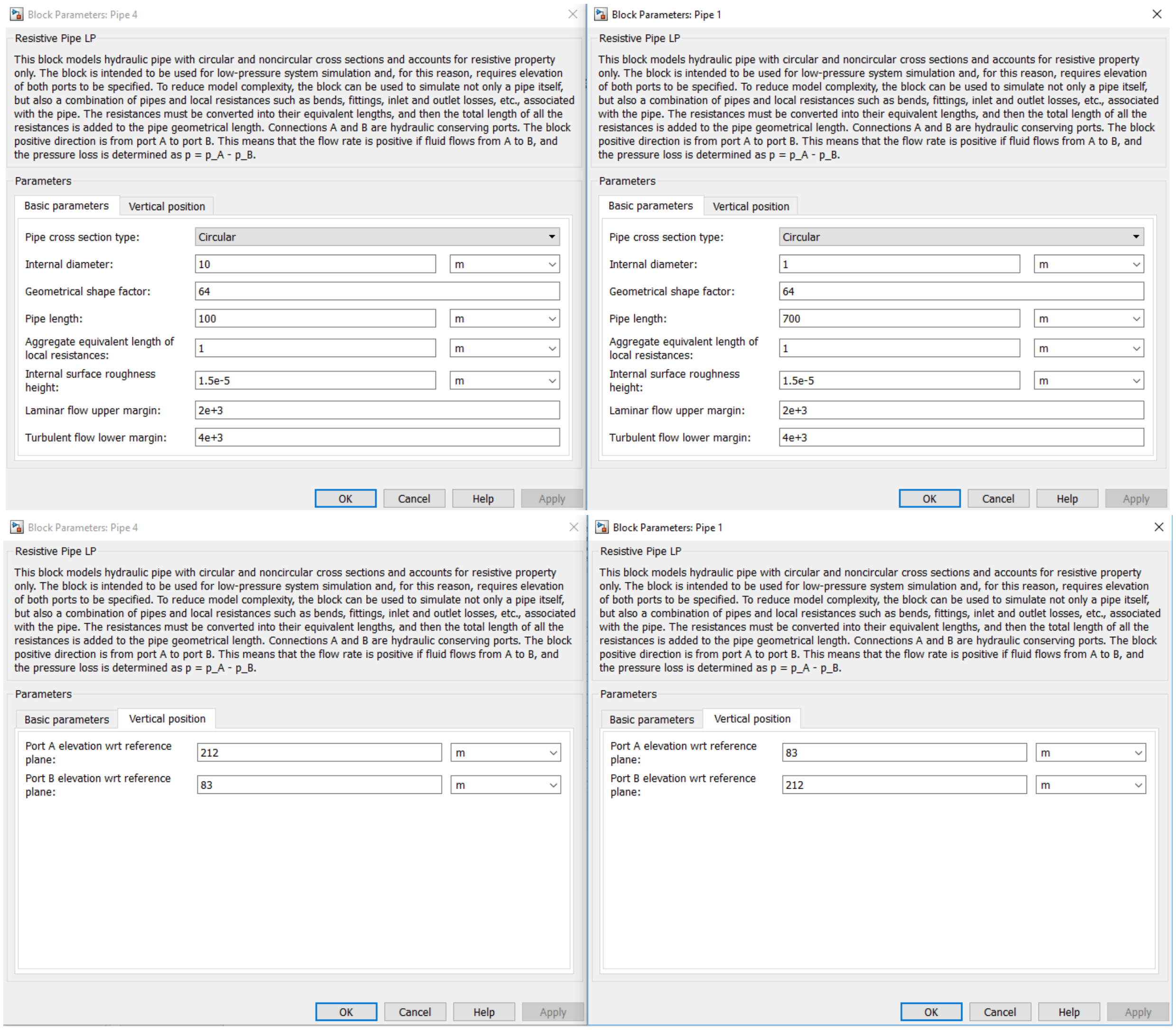


Fig 5.5. Model Parameters

The pumps were set up according to the pressure – flowrate curves and their efficiency curves. From the efficiency curves, the point of the optimum operation was located, that was the point in which the highest amount of hydraulic power was generated. After that, the specific flow rate was located in the pressure – flowrate curve in order to get the corresponding pressure. So, this point for the first pump was 13.4 Bar of pressure with 420 m^3/h flowrate and for the second pump 7.7 Bar of pressure with 253 m^3/h.

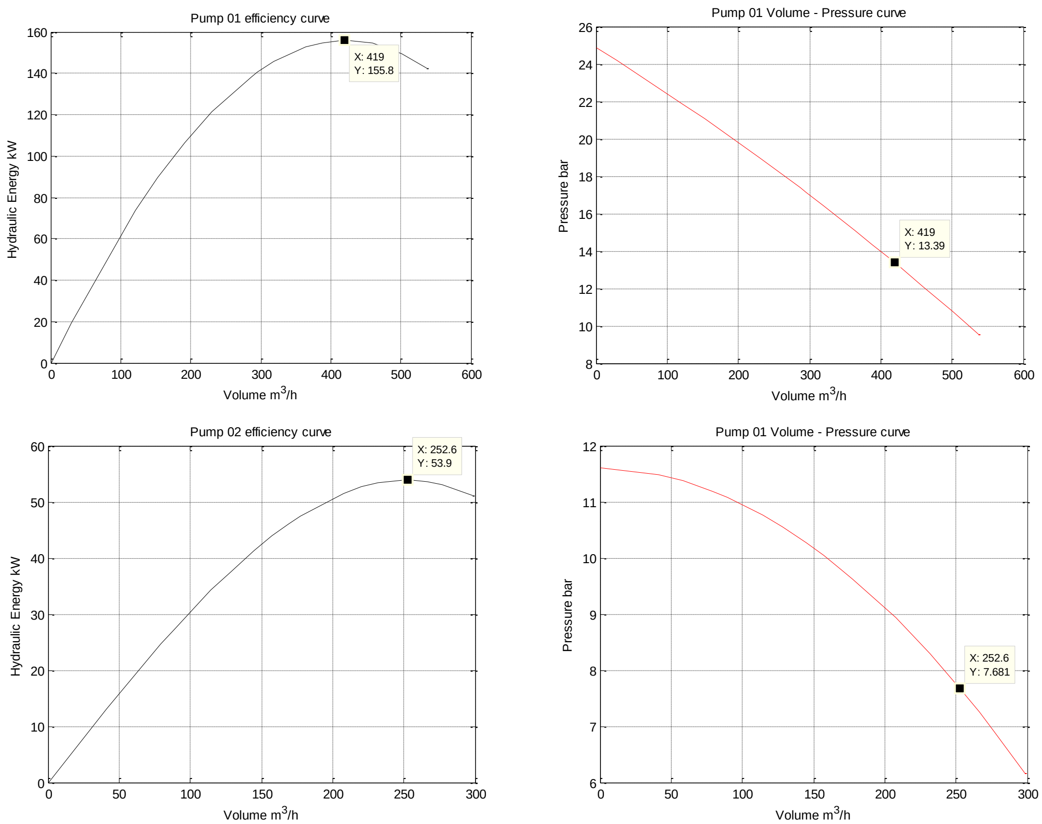


Fig 5.6. Operation Curves of the used pumps

The values of the curves were stored into Matlab variables the model_params.m file which is in the appendix. The P-Q and N-Q parametrization was selected from the three parametrization options available. The units for all the variables can change, however there was no m^3/H option for the flow rate, so instead m^3/sec was selected and the values were divided by 3600.

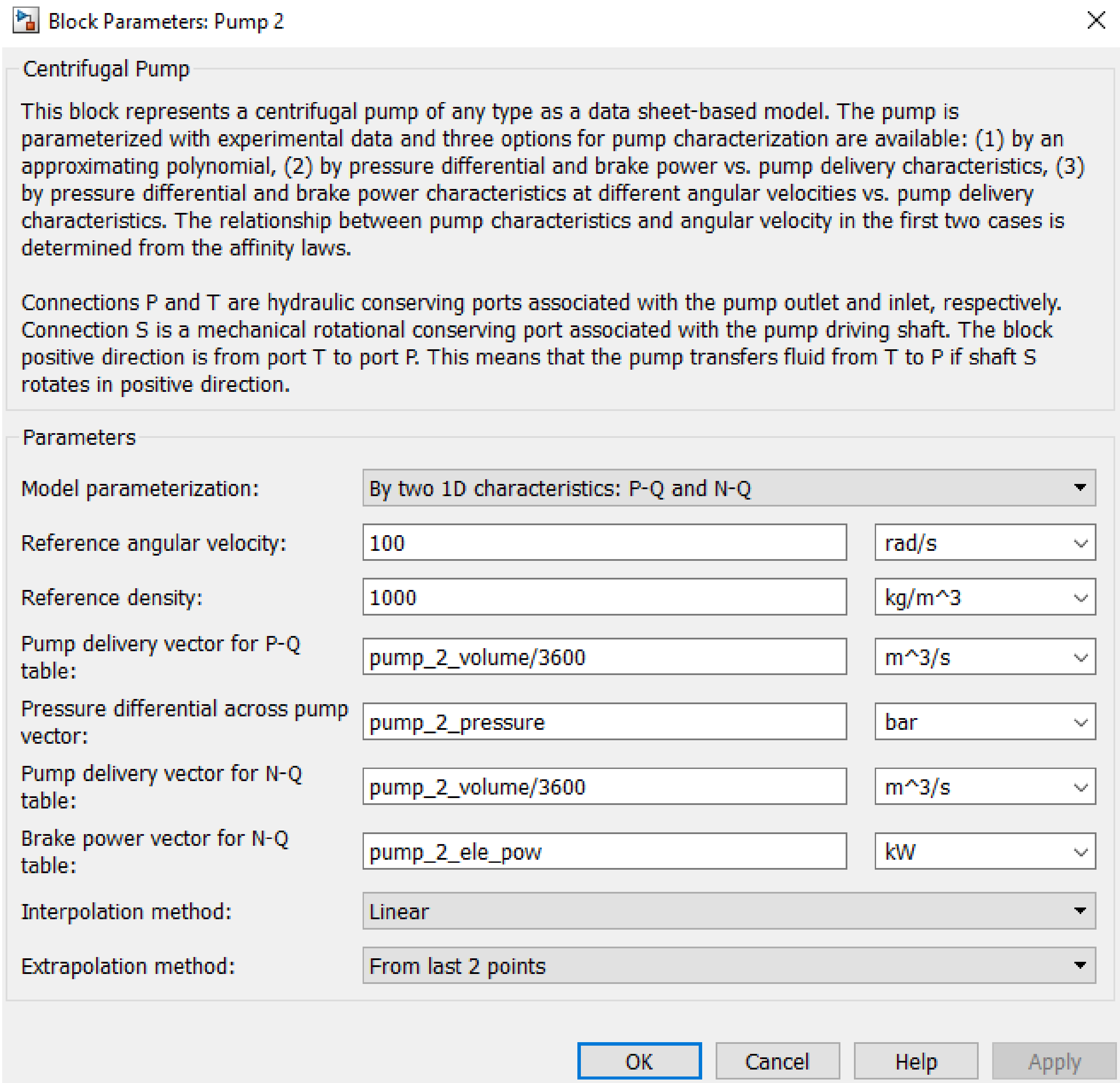


Fig 5.7. Model Parameters

Below, the model for the pump is depicted. Here a selection switch decides whether there will be 100 rad/sec or 0 rad/sec input in the prime mover of the pump, thus changing its condition between to states, on (input in the selector is 1) and off (input in the selector is 0). There are also pressure and flow rate gauges connected for monitoring. Gains are used for unit conversion.

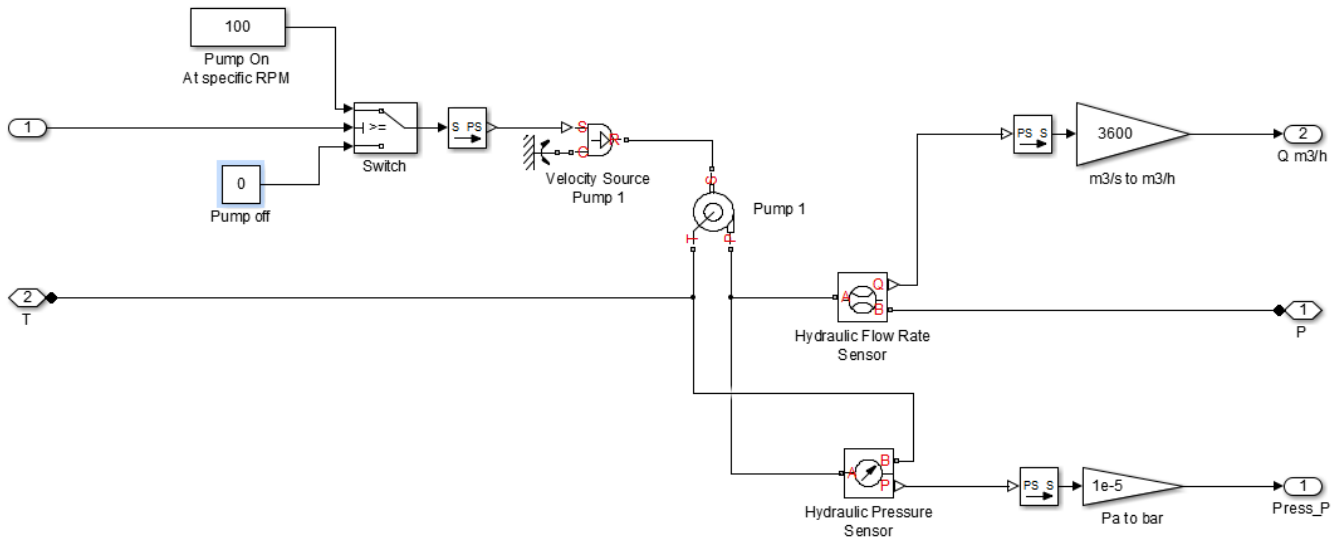


Fig 5.8. Pump Station 1

The pumps however use two additional blocks for improving its operation. A check valve is necessary to avoid water returns when pump is off or when the network pressure is higher than the pressure generated by the pump. Also a pipe block is used here in order to control the output of the pump by changing the cross section of the output (orifice). The general rule is that reducing the cross section increases pressure and reduces flowrate. These pipes' cross section was selected after experimentation so as the pressure and flowrate of the pump would be those of the optimum operation point selected from the efficiency curve in each machine.

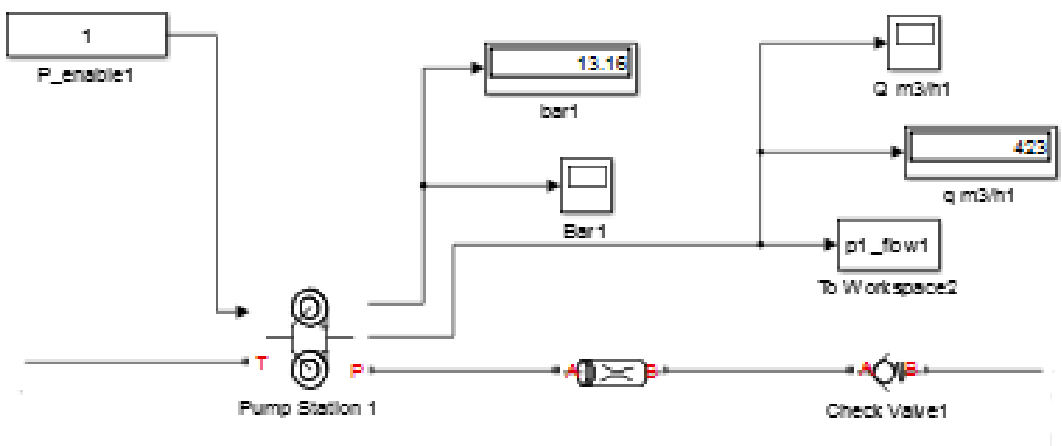


Fig 5.9. Pump protection and orifice control

The model described in this chapter was used for the understanding of the system operation. Inputs of the system can be considered the enable operators of the pumps and the output the volume of the second tank. However others can use the same model for experimenting with the head, pipe diameters and lengths and pump operation points.

The following table shows the simulated tank volume with all the combinations of pump operation for 3600 sec simulations.

Pump 1 enable	Pump 2 enable	Pump 1 flowrate m^3/H	Pump 1 pressure bar	Pump 2 flowrate m^3/H	Pump 2 pressure bar	Korakies tank volume m^3
0	0	0	0	0	0	0
0	1	0	0	246	7.814	246
1	0	423.2	13.15	0	0	423.2
1	1	423	13.16	245.6	7.828	668.5

Fig 5.10. Simulated Tank Volume for operation combinations

5.2.3. Simulation of the Case Study Vlite's Water Network

The design below shows a water supply system consisting of six pumping stations located at 83 m with respect to the reference plane, respectively. All six stations are expected to pump water in a tank located at 212 m with a 37 m diameter and 0.5 m minimum water level and 6 m maximum water level. All tanks are large enough to assume that the fluid level remains nearly constant. The initial volume of water in each tank is set to 0 m^3 except the Korakies tank whose volume is a simulation parameter, however it will always be between 538 and 6451 m^3 which is the volume of a round 37 m diameter in heights of 0.5 m and 6 m respectively. Each pumping station consists of a centrifugal pump and a prime mover rotating at 100 rad/sec. The pump characteristics are specified using lookup tables.

The objective of the simulation is to determine steady-state flow rates, pressures and volume. For this reason, all pipes are simulated with the Resistive Pipe LP block and no system dynamics are observed.

Below there is the version of the model that was used in the simulation. All six pumps were added in the pump station. The demand was added as an additional tank which is filled with natural flow from the Korakies tank since this new tank is considered to be in ground level which means there is 212 m head between the two. The pipe that feeds the demand is 1506.228 m long and has 700 mm diameter as it is stated by the Vlite site data. The flow rate of the demand is controlled with a variable hydraulic orifice block. Its input comes from a lookup table which correlates the hour of the day with requested flowrate.

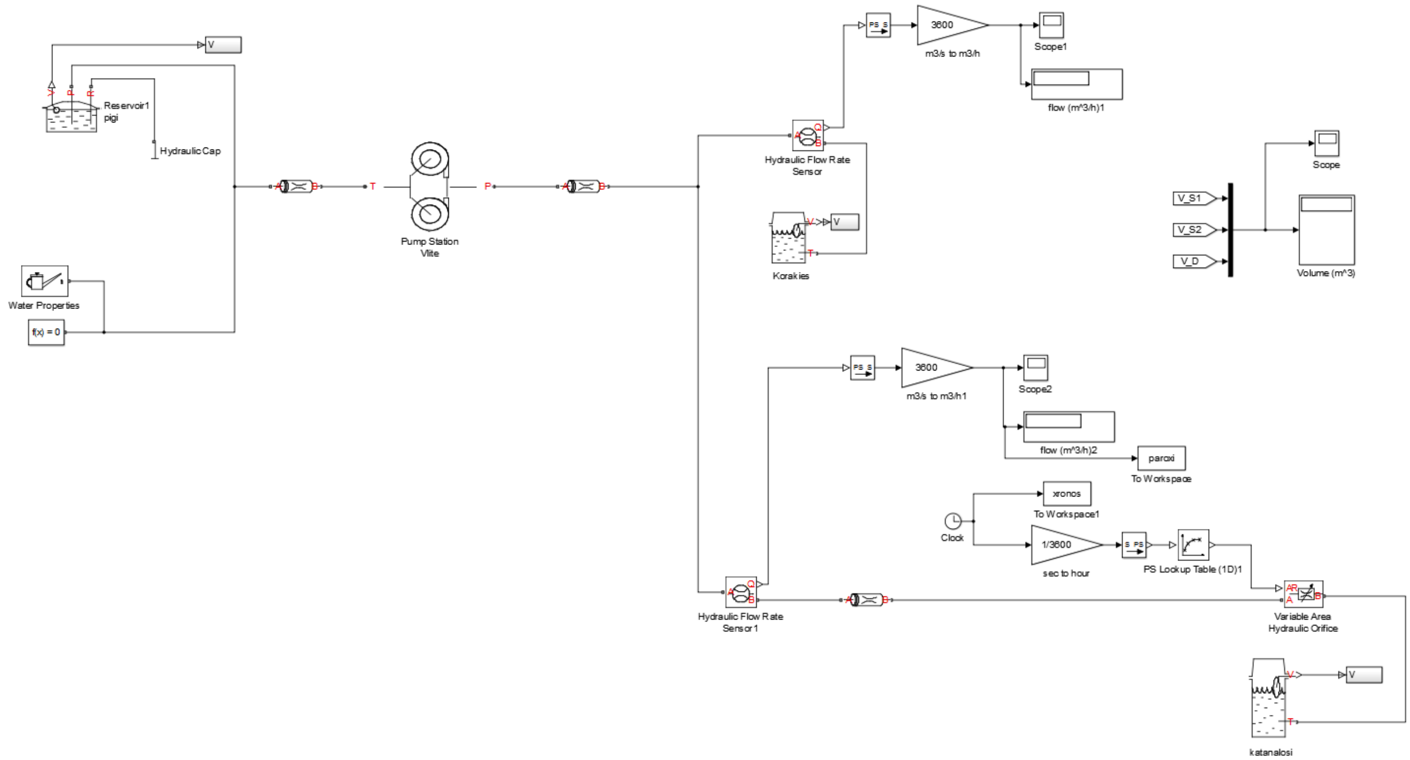


Fig. 5.11. Representation of the water network by Simulink

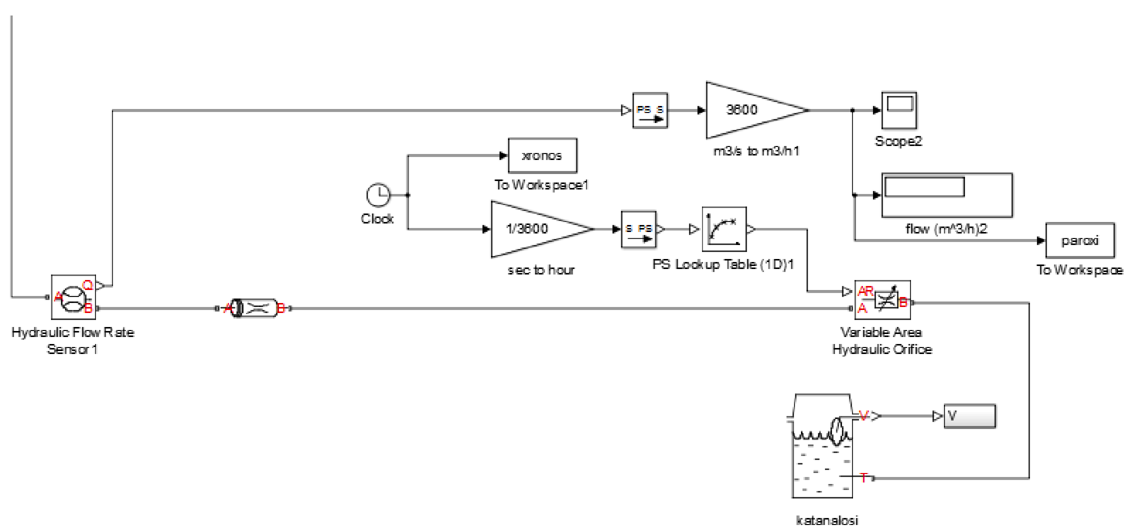


Fig. 5.12. Representation of water demand as a time function in Simulink

The look up table for the water demand that was used in this study derived from statistical data from the Vlite pump station. The daily demand profile has a similar form throughout the year, where there is a peak during noon and a secondary peak but with smaller magnitude during afternoon hours. Although the form remains the same the average consumption changes from month to month. In this study, two cases of demand will be examined a high demand day with average demand of 685m^3 and a medium consumption day with average demand of 520m^3.

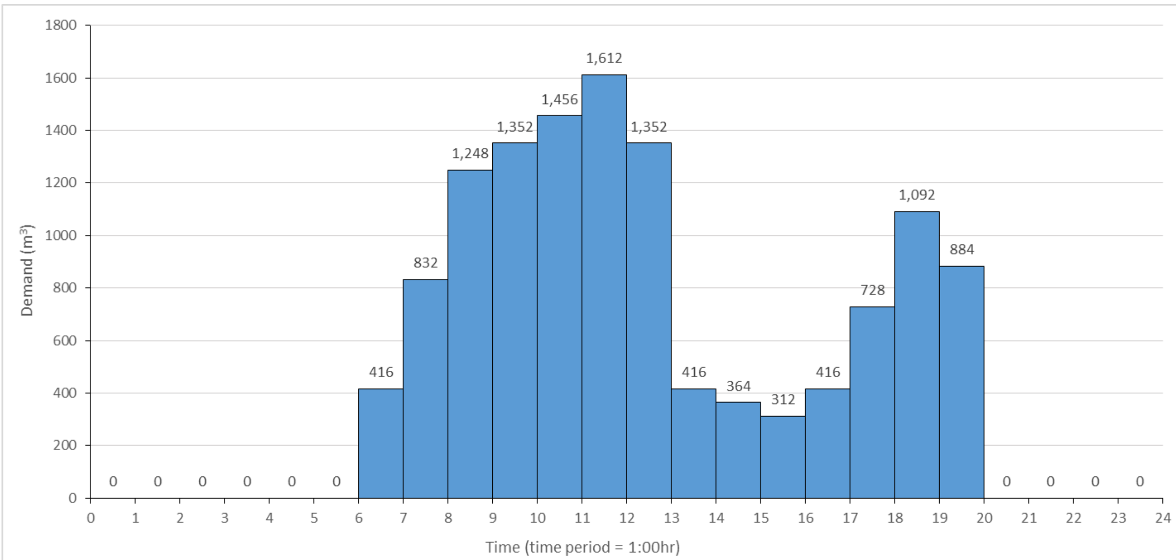


Fig. 5.13. Daily water demand curve for a medium demand day

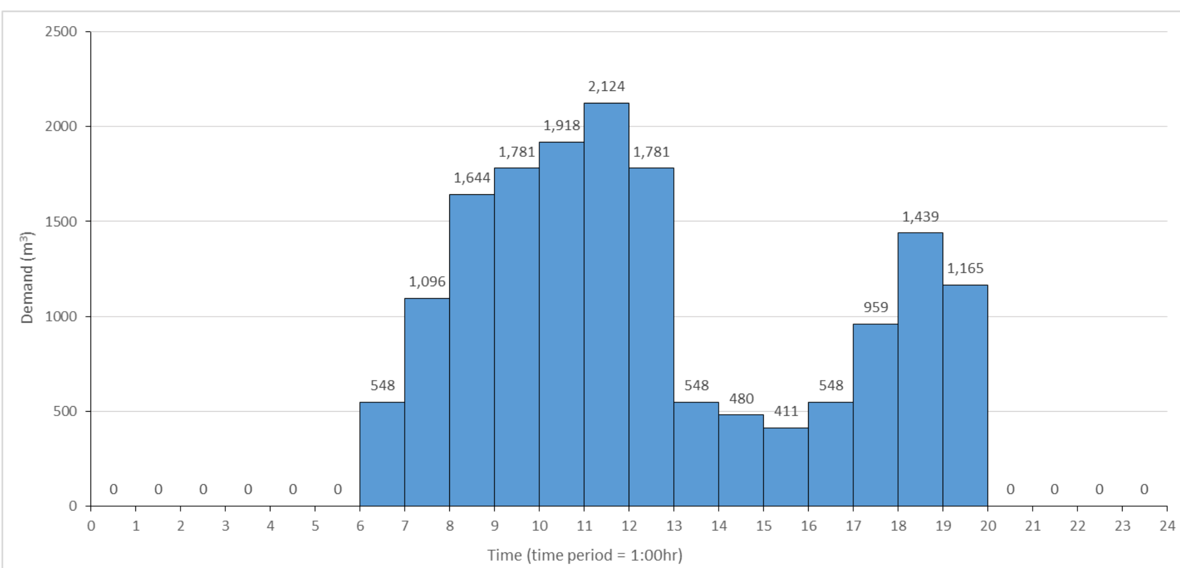


Fig. 5.14. Daily water demand curve for a high demand day

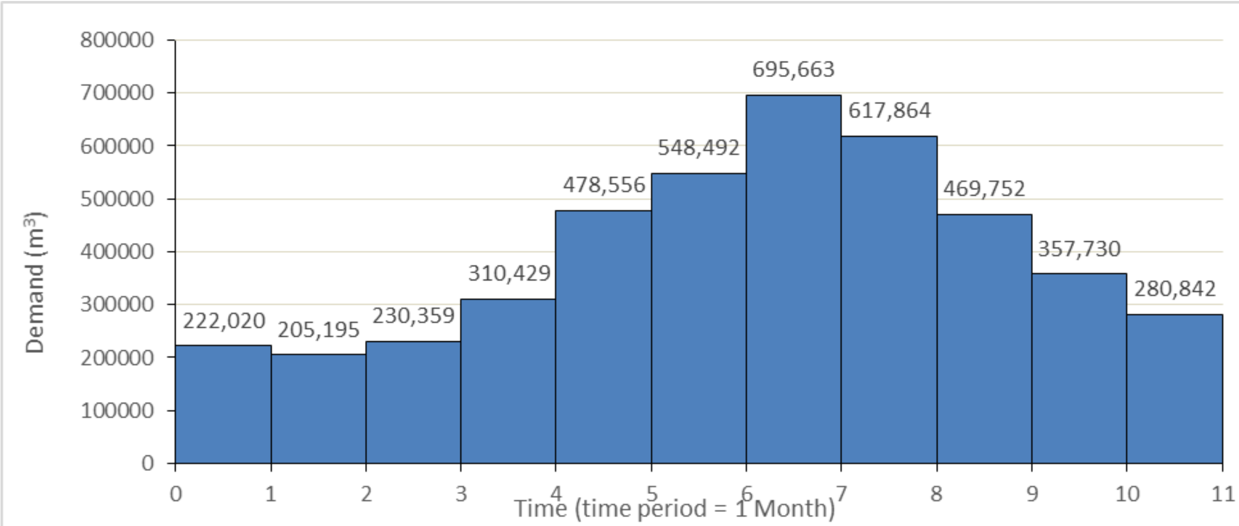


Fig. 5.15. Monthly water demand curve

The pump station of the model is composed by 6 pumps. Four of them are the same type and power and can give 350 m^3/h, while the other 2 are booster pumps with 120 m^3/h capacity. The pumps are all connected in a parallel topology.

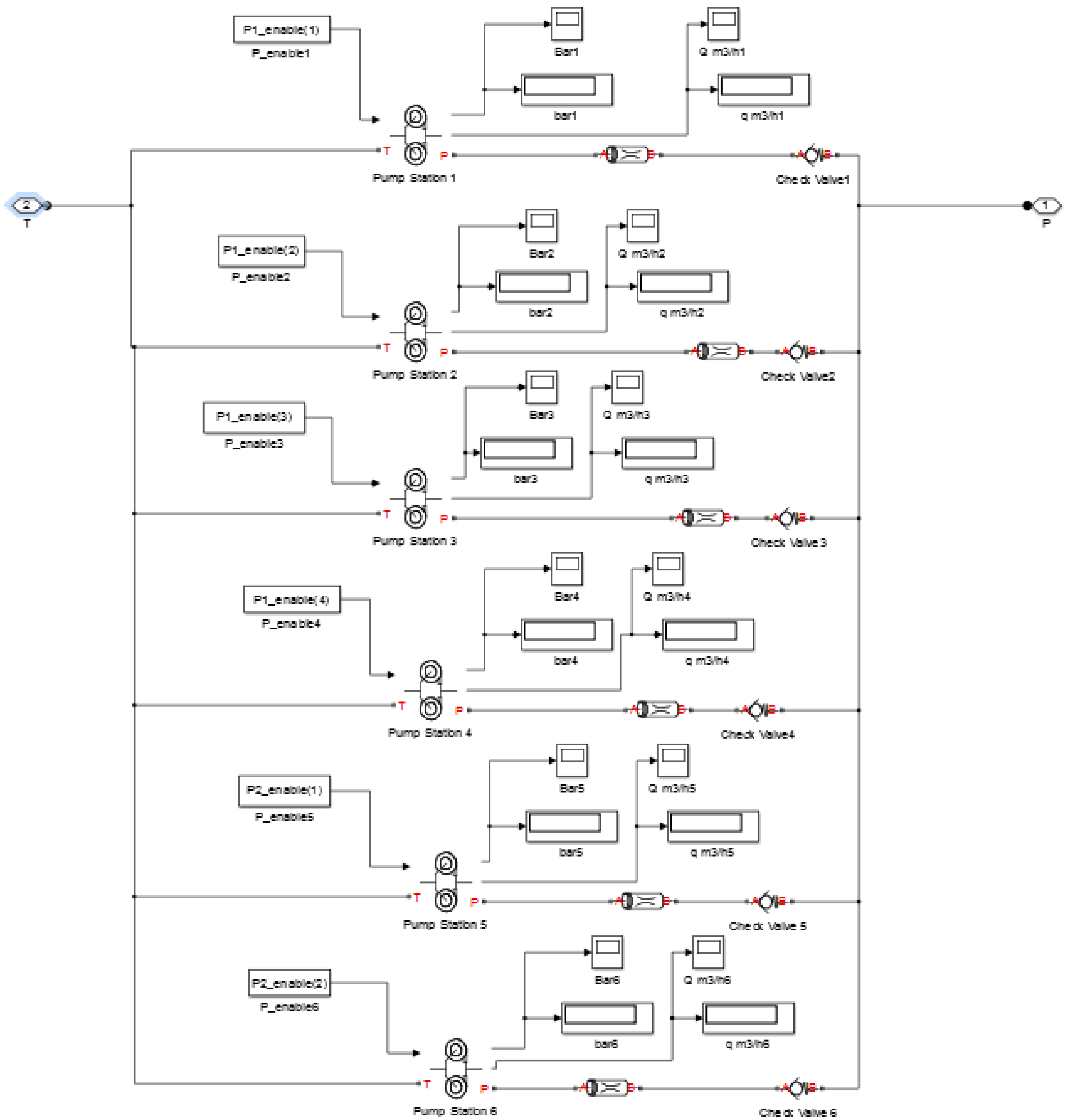


Fig 5.16. Implementation of the pump station

5.3. Mathematical Formulation and Physical Interpretation of the variables

5.3.1. Basic Definitions

Figure 5.2 depicts a modest sized water distribution network but it contains all the typical elements which must be modelled in order to determine the optimal control policy. A general overview of the possible modelling issues (objective functions, constraints) was given in the Introduction section, here will be introduced the exact mathematical formulation.

Since the pumps are operating on their cost efficient point, they are not allowed to be switched too often. Therefore, the optimization problem is divided into series of time intervals. Change of the control is only allowed at the beginning of these periods. A one-hour long time period is typically a good compromise between the accuracy and computational demand.

The criteria for optimization in every Genetic Algorithm is called cost function, as described before. Genetic Algorithms try to minimize the result of the cost function by appropriately manipulating the values of the inputs, using the experience they collect from previous solution evaluations. Altering the cost function in an identical problem can result to completely different results. It is crucial to clearly define the optimization parameters and to construct a mathematical formula that reduces the result of the cost function when desirable attitude is displayed while simultaneously punishing undesirable solutions by increasing the cost function result. It usually has as many terms as the condition the user wants to consider.

An important parameter to be considered while defining the cost function operation is the local minimum of the various fitness functions. Due to the nature of the Genetic Algorithms and mostly due to the fact that the fitness value is valid when compared to the fitness value of other solutions it is rather possible for a solution to be located around a local minimum which could be much higher than other local minimum solution or the global minimum. Due to this uncertainty the solution a GA provides, is always called suboptimum solution, since it cannot be assured (except certain linear and relatively simple problems) that the minimization of the cost function that is achieved is local or global. Special care should be taken in choosing the coefficients of the various terms of the cost function (scaling factors) since at problems where different physical magnitudes are correlated, the importance of its magnitude according to the specific application should be determined.

The aim of this study is to provide an operation schedule that will minimize the energy consumption of the pump station used to supply the Korakies tank with water while avoiding simultaneously the tank level fall below 0.5m or rise above 6m. This will ensure that water provisioning in the nearby area will be constant. Since the first aim is to minimize the energy consumption, the first term of the cost function will be the energy consumption of all six pumps during the simulation time.

$$J_E = \sum_{OPH=1}^{24} \left(P1_E(OPH) + P2_E(OPH) + P3_E(OPH) + P4_E(OPH) + P5_E(OPH) + P6_E(OPH)\right) \qquad \textbf{(5.1)}$$

Where:

- J_E is the cost component that depends on the energy consumption
- OPH is the operation hour of the simulation
- PX_E is the energy consumption of the X pump for every operation hour

A secondary term of the cost function is the need of reserving the water level in the Korakies tank within specific limits, as mentioned before. Since this cannot be physically quantified, the following equation is used.

Where:

$$J_L = \sum_{OPH=1}^{24} \begin{cases} if\ L > 6, PN_{OPH}100 \\ if\ L < 0.5, PN_{OPH}100 \end{cases} \tag{5.2}$$

- J_L is the cost component that depends on the tank level
- OPH is the operation hour of the simulation
- PN_{OPH} is a function that returns the value 1 if the water is out of the specified level and is calculated in hourly basis. The number 100 used in this formula is a scaling factor in order to correlate energy consumption and water level. It is actually the energy consumption of the system in a low demand day. So, even if the water level exceeds the bounds for one operation hour only the penalty will be a whole day's energy consumption.

There was also a practical addition to this cost function to avoid crash of the Matlab Simscape model. It was observed that in cases of tank levels out of the interval of 0 - 8 m level the simulation process reasonably produced an error. Therefore, another similar term was used in order to protect the model from those crashes.

$$J_p = \sum_{OPH=1}^{24} \begin{cases} if\ L > 8, 1000 \\ if\ L < 0,1000 \end{cases} \tag{5.3}$$

If this condition was met during the simulation procedure the rest of the simulation for this individual is skipped and it is punished through the cost function, as if it was outside water level boundaries for ten hours.

Eventually, the cost function calculates the score as:

$$J = J_E + J_L + J_P \tag{5.4}$$

5.3.2. Constraints of the Pump Scheduling Problem

In order to be useful in practice, feasible schedules must satisfy certain constraints. These constraints include hydraulic constraints, also called implicit system constraints, which define the hydraulic equilibrium state of the system, e.g., Conservation of Mass at each node and Conservation of Energy around each loop in the network.

On the other hand, implicit bound constraints represent system performance criteria. They include constraints on junction pressures, pipe flow rates or velocities, and tank water levels. Implicit bound constraints may also include constraints on pump operation switches. Frequent switching a pump on and off results in maintenance costs due to increasing wear on the pump, thus constraining the number of pump switches limits future maintenance costs.

Constraints on tanks water levels typically include minimum and maximum limits on tank levels, and balance between supply and demand from tanks. Minimum and maximum tank limits may be explicit constraints of the problem or can be implicitly enforced by a hydraulic simulator.

Balance between water supplied and consumed from tanks is achieved by ensuring that tanks recover their levels by the end of scheduling period. Balancing supply and demand also allows to apply a similar pump schedule to the next scheduling period, since consumer demands and network conditions are very similar in consecutive periods.

The meaning of the constraints of the $\Delta h^K(i)$ is obvious as it incorporates the minimum and maximum bounds of the reservoir volumes.

$$h_{min}^K(i) \leq \; h^K(i) \leq h_{max}^k(i) \qquad \textbf{(5.5)}$$

Note that $h_{min}^K(i)$ and $h_{max}^k(i)$ can vary during the optimization horizon. The usual case is that in the first T time period these values are constants.

The meaning of the constraints of the $u^K(i)$ is the minimum and maximum bounds of the water flow as we have already found them from our simulation process.

$$u_{min}^K(i) \leq \; u^K(i) \leq u_{max}^k(i) \qquad \textbf{(5.6)}$$

The above constraints are the most frequently used constraints. Additional constraints, such as limits on source flows or velocity constraints, may be incorporated to the problem formulation depending on particular requisites of a network and the optimization approach.

5.4. Genetic algorithm implementation

5.4.1. Introduction

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions.

At each step, the genetic algorithm randomly selects individuals from the current population and uses them as parents to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

You can apply the genetic algorithm to solve problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, no differentiable, stochastic, or highly nonlinear.

Also the genetic algorithm differs from a classical, derivative-based, optimization algorithm in two main ways, as summarized in the following table.

Classical Algorithm	**Genetic Algorithm**
Generates a single point at each iteration. The sequence of points approaches an optimal solution.	Generates a population of points at each iteration. The best point in the population approaches an optimal solution.
Selects the next point in the sequence by a deterministic computation.	Selects the next population by computation which uses random number generators.

In addition the computational cost of the hydraulic simulation was decreased heavily through simplifying the model and using a most efficient hydraulic solver. However, the genetic algorithm usually needs $10^5 - 10^6$ evaluations to find a ’good’ solution which entirely satisfies the constraints and has a ’low’ total electrical cost. Satisfying the water volume capacity limits in the reservoirs seems to be the highest challenge.

In order to ’help’ the genetic algorithm, a rule-of-thumb algorithm was implemented which helps to keep the reservoir levels in the prescribed range. If the algorithm observes that one of the reservoirs would be discharged, then along with the demand the suitable number of incoming pumps will be switched on, operating cost efficiently. This change is also set in the DNA of the individual, which helps the algorithm find feasible solutions also in the latter generations.

The GA solver handles linear constraints and bounds differently from nonlinear constraints. All the linear constraints and bounds are satisfied throughout the optimization. However, GA may not satisfy all the nonlinear constraints at every generation. If GA converges to a solution, the nonlinear constraints will be satisfied at that solution.

Ga uses the mutation and crossover functions to produce new individuals at every generation. The way the GA satisfies the linear and bound constraints is to use mutation and crossover functions that only generate feasible points. For example, in the previous call to GA, the default mutation function mutationgaussian will not satisfy the linear constraints and so the mutationadaptfeasible is used instead. If you provide a custom mutation function, this custom function must only generate points that are feasible with respect to the linear and bound constraints. All the crossover functions in the toolbox generate points that satisfy the linear constraints and bounds.

5.4.2. Genetic algorithm parameters selection

The computational software Matlab allows the use of GAs through the optimization toolbox. This toolbox has a preinstalled genetic algorithm and a set up wizard through which users can select the preferred values in various fields.

For the current problem to be solved, the following values were chosen for the various parameters of the GA.

Fitness function: The function that is executing one simulation of the model. This function is given in the appendix and includes the constraints, the constants of the system, the simulation and the values of every variable and finally the calculation of the fitness value which here will be called score.	@tank_obj
Number of variables: The number of values calculated by the GA for every individual of every generation.	144
Population type: Variables can be real numbers, integers or binary values. Here binary values have been chosen due to the nature of the problem.	Bit string
Population size: The number of individuals in every generation.	30
Scaling function: A function responsible for interpretating the raw fitness score to a value that the GA can understand. Here the score itself is not important but the relative score and the rank of the individual regarding its peers in the same generation.	rank
Selection function: The function responsible for choosing which solution are fit for reproduction and creating offspring for the next generation. Here, roulette is chosen which simulates the selection using a roulette wheel where every segment is proportional with the fitness of each individual	roulette
Elite count: The number of individual that have the best fitness and as such, they pass to the next generation bypassing the reproduction process.	2
Crossover fraction: This value shows the percentage of the population that is going to be originated from reproduction. Here only the 80% of each generation is going to be produced via reproduction.	0.8

Mutation function: The function used to spontaneously change parts of some solutions creating divergence into the system. Mutation is going to create the remaining 20% of the generation.	Uniform
Mutation Rate: The probability there is for each chromosome of every individual selected for mutation to actually mutate.	0.01
Crossover function: The function that defines the way the two parent individuals are combined to create a new individual (child) for the next generation. Here the parent individuals are placed as strings of binary values and this function decides integrals where the values for the child come from one parent and integrals where the values come from the other parent. This is also called multiple point crossover.	Scattered
Migration direction: In the main population, GA creates several subpopulations. Every so often individuals from one subpopulation can migrate from one population to another so that the best individuals of one subpopulation can take the place of the worst individuals of another population. The direction states that migration will occur from the first subpopulation to the second and so on until it wraps from the last to the first again.	Forward
Migration fraction: This is the percentage of the individuals that will be migrating from one population to another. This fraction refers to the smaller of the two subpopulations.	0.2
Migration interval: This value shows the frequency of the migration process. Here it occurs every 20 generations.	20
Maximum generations: This is one of the stopping criteria of the algorithm. It states that the process will end after 100 generations.	100
Maximum stall generations: This is the second stopping condition. It states that if after 10 consecutive generations there is no improvement in the fitness value the simulation will end.	10

Table 5.1. Genetic algorithm informational table

In order to introduce constraints to this simulation there were penalties chosen to punish undesired behavior. The main parameter which was necessary to be controlled was the level of the Korakies tank. So in order to punish the individual which gave values that created a tank level below 0.5 m or above 6 m, the fitness value was increased by 100. Actually, in order to help individuals that only stayed outside the for one or two hours every 24 the penalty was calculated as 100 times the sum of the hours the tank was out of bounds.

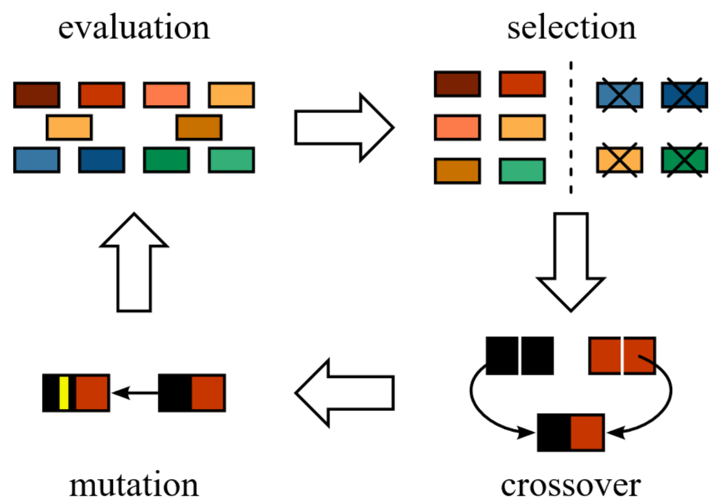


Fig 5.17. Flowchart of the basic operations of a genetic algorithm

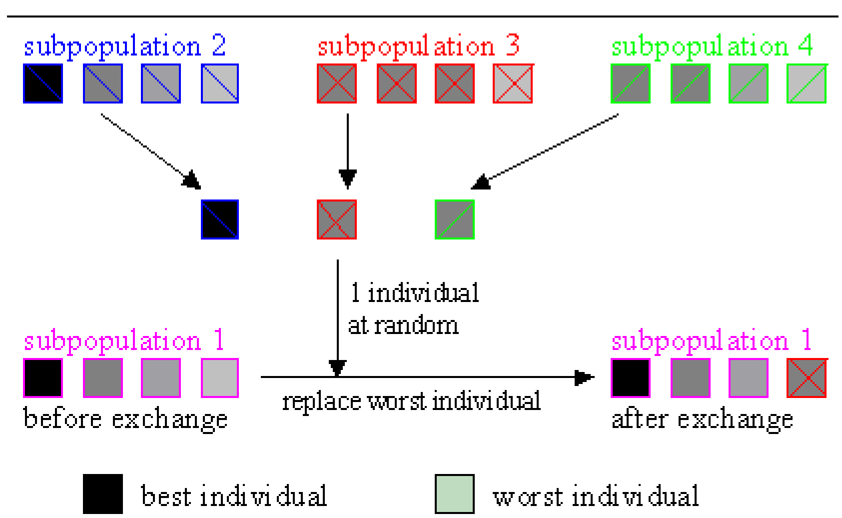


Fig 5.18. Migration of individuals across subpopulations example

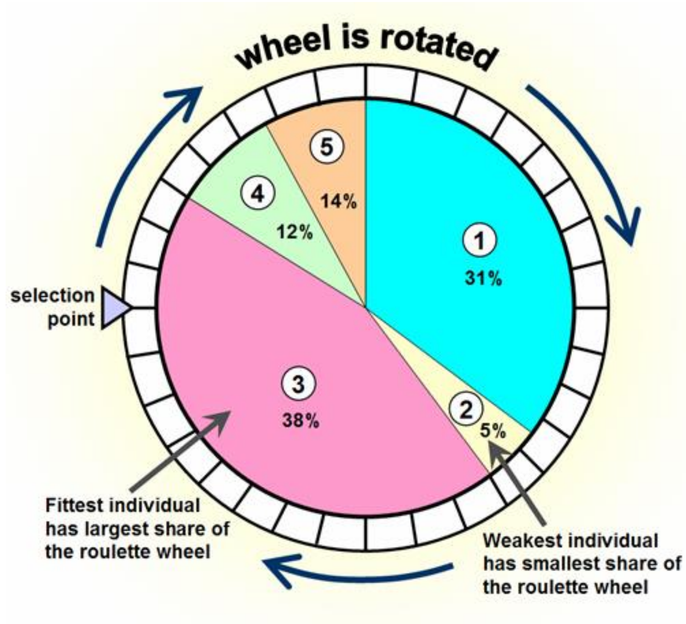


Fig 5.19. Roulette wheel selection

Some remarks are following below regarding the programming part of this study. For every individual of the GA, a simulation of the model is performed. That is a group of 24 hourly simulations where every hour the tank level at the end of the previous simulation becomes the initial level of the tank on the next simulation. In the Matlab Simscape model, when the Korakies tank level was lower than 0 or higher than 8 m, the result was crashing and stopping the GA. Consequently, in order to avoid this situation when the value for the tank level was below 0 or above 8 m, the simulation for this individual stops and a bad score is given to the individual producing this condition. This will generally force the individuals with this tendency out of the population formed.

The penalty for this condition is set equal to 1000. Also, there is a common problem in these optimization problems due to the use of Matlab functions. Matlab functions are executed in their own domain so they use some internal variables that are not stored in the workspace and so Matlab itself does not have access to those. In this function this problem was resolved with the use of the “assignin” command. More specifically, during the execution of the function, in case that an internal variable X needed to be saved in the workspace (base) the following command was executed inside the function:

assisgnin('base','X',X); **(5.7)**

5.4.3. Formulation of input variables

A genetic algorithm, as described before, is a function that has several particularities. Some of those are the parameters for the population's size, the mutation probability etc, that were explained above. One of the most important conditions of GA however is the method though which the input of the objective function is formed into an individual, also called chromosome. It is a common practice to use a binary representation of the input variables stacked together, creating a string of zeros and ones (bit string) which has as many digits as the number of inputs times the bits per input. In the problem discussed here, the number of inputs is the operation value of each of the six pumps in one day using one hour intervals. So a bit string is created with 144 bits where the first six digits represent the operation values (on or off) for the six pumps in the first hour of the test day and so on. Below, the correlation between the chromosome and the working schedule of the pump station can be seen.

Typical chromosome of this GA, 144 bits long																																				
																																				…

Pump operation per hour schedule																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
P1							…																	
P2																								
P3																								
P4																								
P5																								
P6																								

Fig 5.20. Genetic Algorithm Chromosome Table

5.5. Results of Genetic Algorithm Implementation

The first scenario were the genetic algorithm was is a medium demand day with the daily that was mentioned before. The initial level of the Korakies tank was considered to be 3.25 which is the middle of the level margins. After 44 generations the following schedule prevailed with its cost function granting a score of 125.54.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
P1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0
P2	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
P3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
P4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	0
P5	0	1	1	1	0	1	0	0	0	1	0	1	0	1	0	1	1	1	0	0	0	0	0	0
P6	0	0	1	0	0	0	0	0	1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0

Fig 5.21. Genetic Algorithm solution for medium demand scenario

In addition, some bar graphs show the volume and level variation for supply demand and the middle tank.

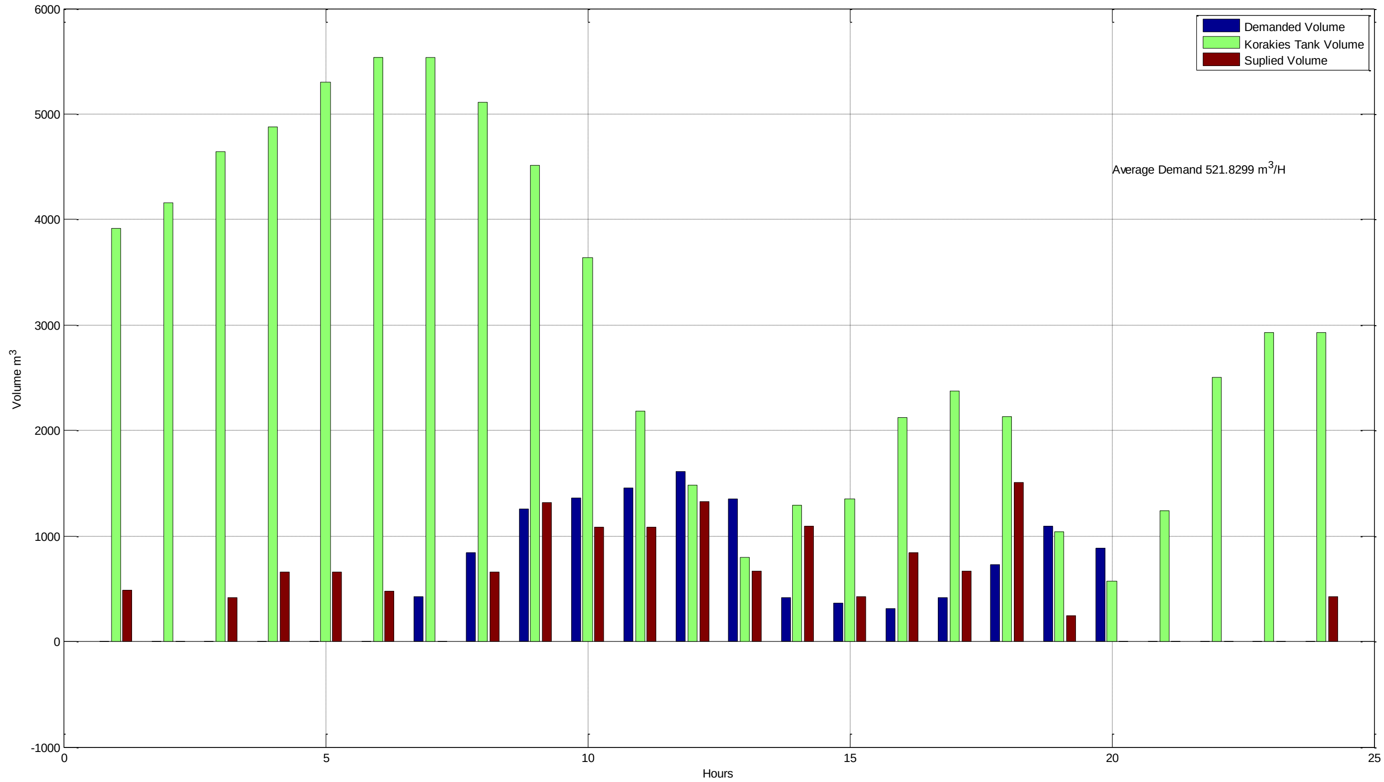


Fig 5.22. Water volume over time for medium demand scenario

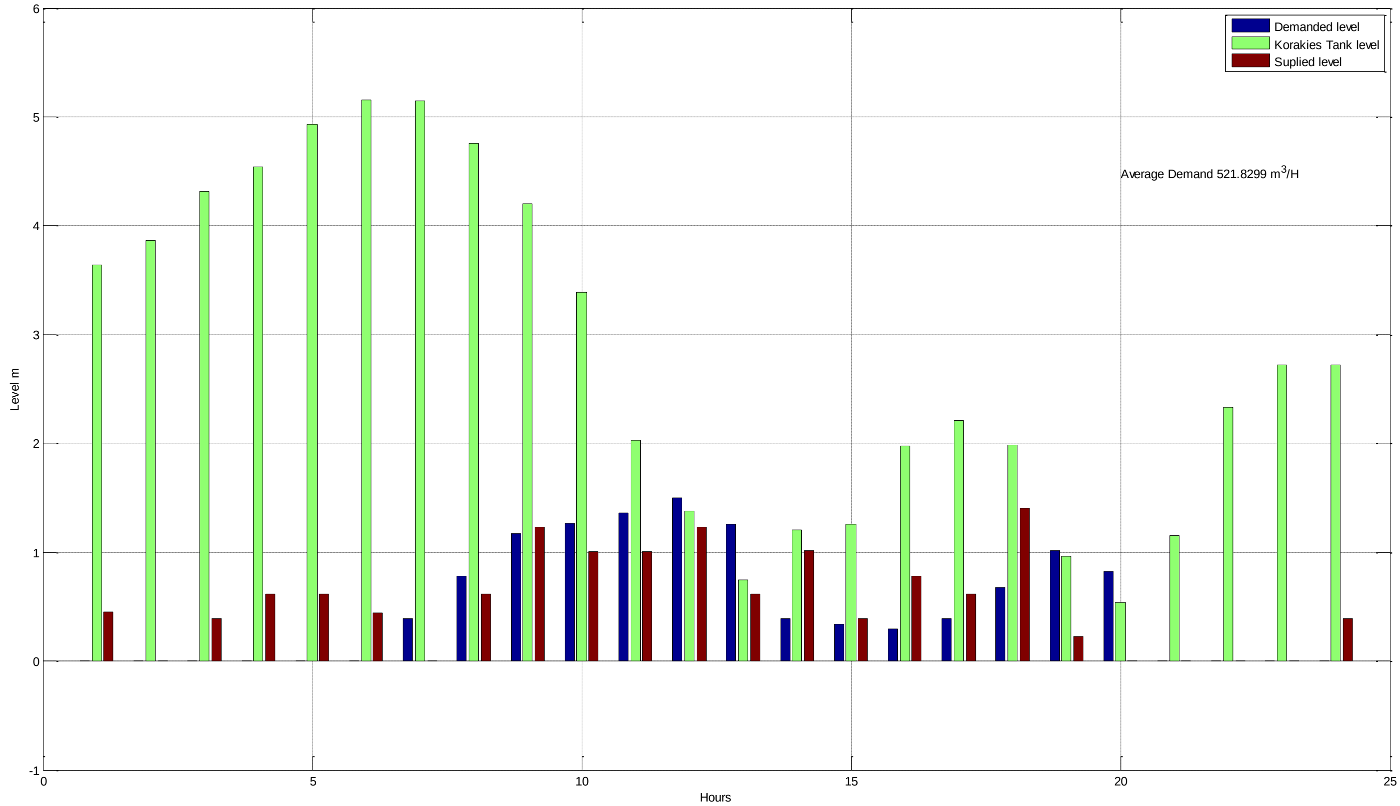


Fig 5.23. Water level over time for medium demand scenario

The second scenario is identical to the first, however it refers to a high demand with average consumption reaching 685 m^3. Here the GA took 62 generation to produce the following schedule which was evaluated and scored 130.27.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
P1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
P2	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0
P3	0	0	0	1	1	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
P4	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1
P5	1	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0
P6	1	0	0	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0

Fig 5.24. Genetic Algorithm solution for medium high scenario

In addition, some bar graphs show the volume and level variation for supply demand and the middle tank.

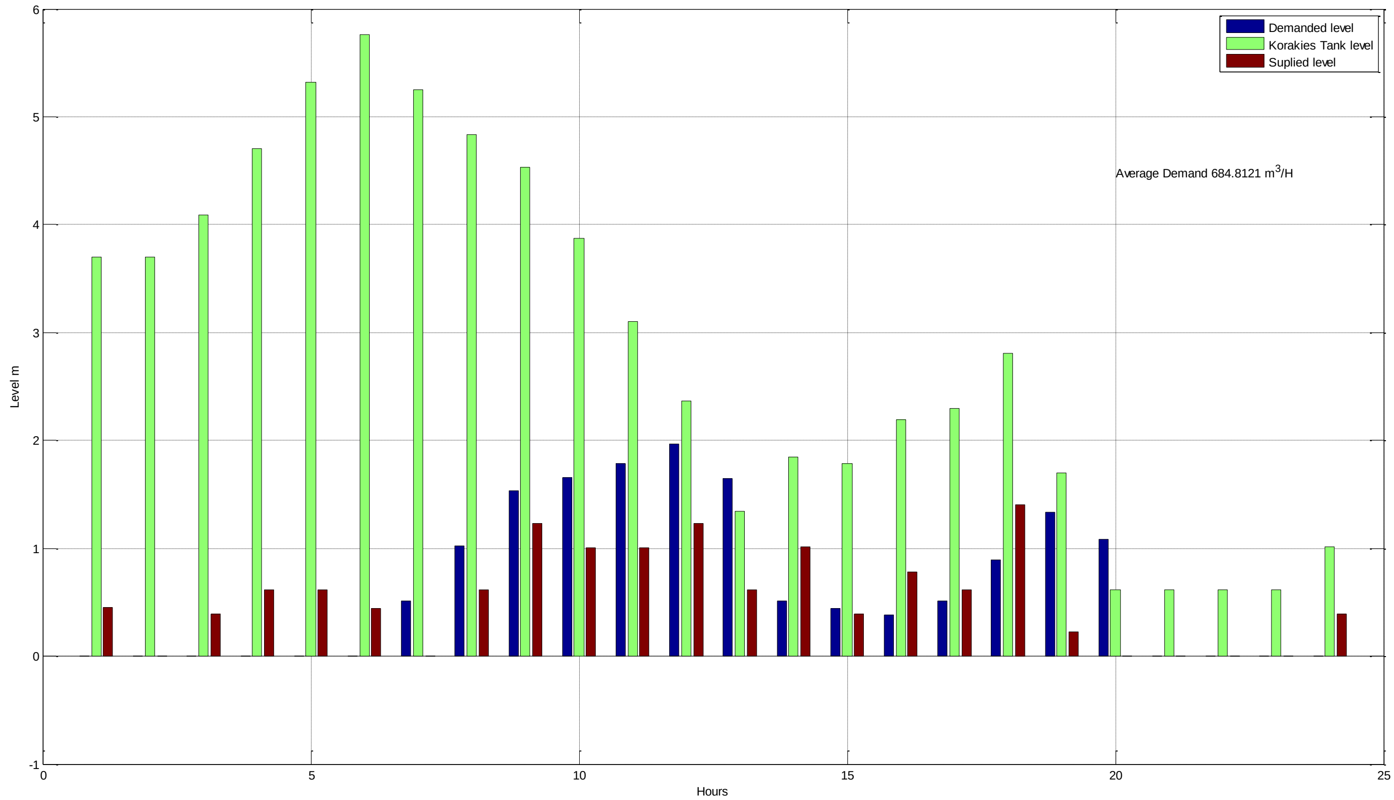


Fig 5.25. Water volume over time for high demand scenario

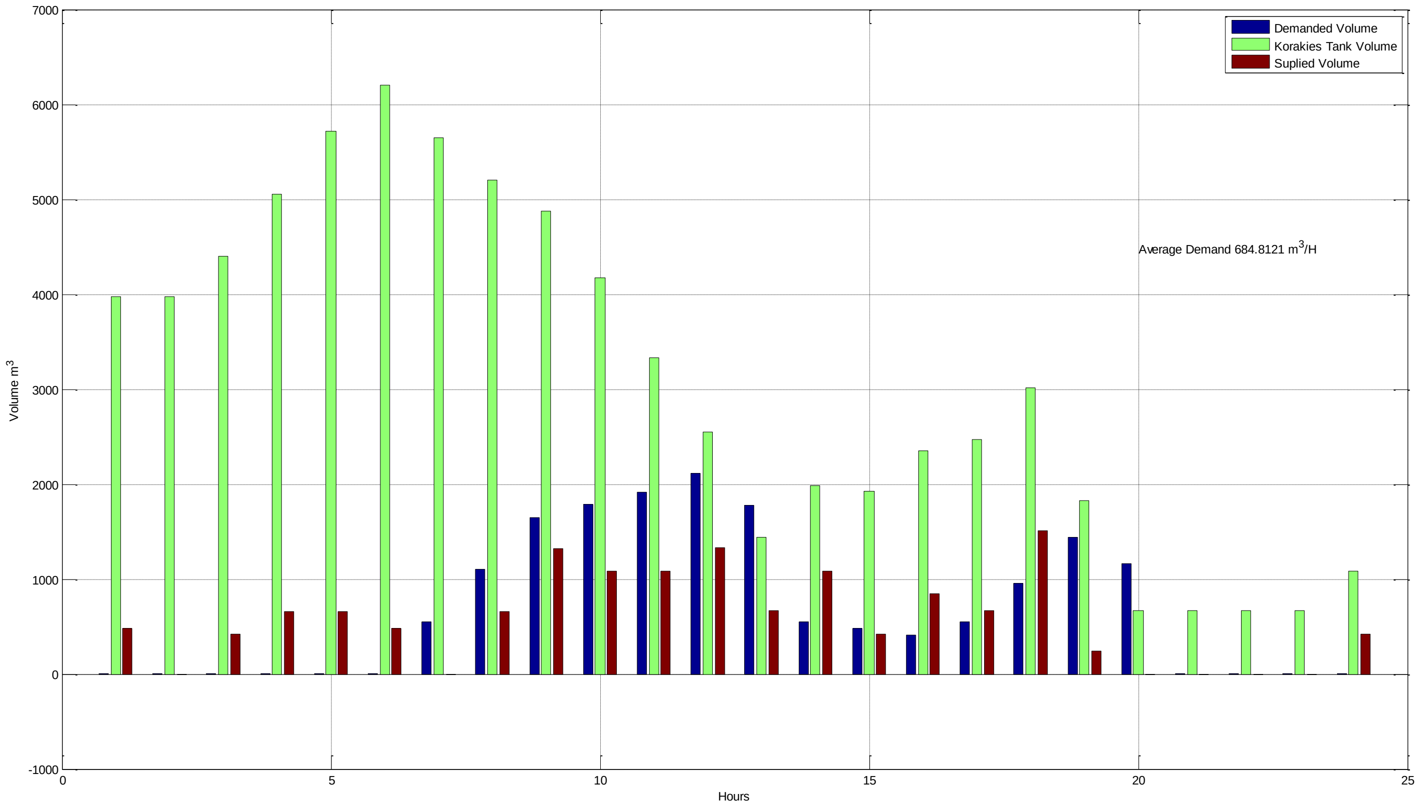


Fig 5.26. Water level over time for high demand scenario

As it shown here in both the operation scenarios the GA managed to find a solution so that the demand could be met. In addition there is no point where the level of the korakies tank fell under 0.5 m or rised above 6 m.

6. Conclusions

6.1. Summary

To sum up, the schedule optimization of regional water distribution systems covers a huge segment of water management problems. It means that it is almost impossible to deal with every aspect of this research field or to develop a superior optimizer which performs well in any circumstances in every kind of waterworks system.

Accordingly, the aim of this thesis was not to find the 'best schedule optimizer' but to present an improved optimization algorithm for solving different problems by different solving methods. In addition, more sections in this thesis focus not only on determining optimal or near optimal schedules, but also on understanding the optimality of the results.

Although the presented methods do not give the reader a detailed overview of the topic, I believe that they are good enough to provide an impression on how difficult and diversified the discussed research field is. The presented method can possibly serve as a basis for further methods. Besides, a relevant part of the presented technique already have industrial realizations which solve every day, real life problems in an elegant manner.

Objectives of scheduling problems could be various. The most used objective functions are obviously the total pumping energy or total pumping cost for the optimization period. The total number of pump switches can serve as an alternative objective function to be minimized.

The methods in the thesis involve the above mentioned objectives on deterministic, single objective, pump scheduling optimization systems. In the case of stochastic systems, disturbances can appear e.g. from the stochastic nature of the consumptions, and defining a proper objective function can also be a cumbersome problem.

The advantages of a novel developed method can also be various: they can provide 'better' or 'more reliable' results than former approaches. Its applicability can be deeper, it covers a wide range of the existing problems. It is interesting if seen from an industrial point of view or, obviously, from a scientific point of view.

In this thesis we present a novel genetic algorithm for solving a wide range of pump scheduling problems. Since evolutionary methods are very popular and they are said to be ’robust’, this method was intended to solve complex problems.

The technique is directly applicable to the daily optimization tasks of middle-sized and large systems, works well with a priori determined pump operation points and coupled hydraulic simulations as well.

The use of economical pump schedules plays a significant role in reducing the energy consumption of the mankind, thus, it helps to take into account the aims of the sustainable development. I hope, that my work would contribute to some degree to these aims.

6.2. Future directions

The water consumption of the next 24 hours is as an input data, which is obviously not known a priori but only estimated with the help of statistical techniques. Whatever the forecast technique is, from the point of view of the optimizer, the water consumption is a set of numbers which does not include any information on its uncertainty or probability, which can be considered as the bottleneck of the technique.

Also knowing that the real consumptions will obviously differ from the forecasted ones (Bárdossy *et al.* 2009, Alvisi *et al.* 2007), the computed optimal control will not be optimal for the real consumptions. While following the computed optimal pump schedule, a deviation between the real and predicted values (e.g. reservoir levels) will be experienced, which is mainly caused by the stochastic nature of water consumption.

A possible future research direction is to accelerate the hydraulic simulation of the optimizer. This is crucial from the point of view of computational speed since the evaluation of one single candidate solution requires 24 steady state hydraulic simulations, each hydraulic simulation consists of solving a non-linear equation system with a dimension of the number of branches plus the number of nodes. Almássy *et al.* (1981) solves an equation system of a size of the number of independent loops in the network while Todini & Pilati (1988) and Salgado *et al.*

(1988b) presented a method where the number of the equations are equal to the number of branches.

Although their method is widely applied and popular (e.g. the free software EPANET (2012) uses this approach), the last results were published in the late 80's thus a possible research gap appears in this research field. Others (Rao & Alvarruiz 2007, Jamieson *et al.* 2007) substitute the hydraulic solver with neural networks but also in that case the learning process requires a huge amount of simulations with the original simulation solver.

Another future research direction is that the basic model that represents our water network which is being used from the genetic algorithm for either one hour or one day, can be used to take a decision on which pumps should work the same time that our system is operating.

In addition the genetic algorithm criterion could be extended so that it takes in consideration the different costs of electric energy throughout specific hours of the day. For example the criterion could be the decrease of the workload costs by filling up the reservoirs during the time periods when electricity is less expensive and covering the water demands from these reservoirs in the expensive tariff hours. The idea seems clear, but due to the large number of constraints (reservoir capacity, node pressure and power limits) and the mixed integer type variables (constant and variable speed pumps) the problem becomes highly challenging from mathematical point of view.

The approach this thesis has taken allows the change of scope, so as other optimization parameters are taken into account. Changing the cost function for example could achieve several of the criterion mentioned above such as minimizing pump switching or energy cost reduction. Adding a constraint in this approach could for example calculate the schedule when one of the pumps is unavailable due to maintenance.

Finally, we have observed that most of the computation time is spent on the hydraulic simulation of potential schedules. Therefore, future algorithms should be developed with the goal of decreasing the number of simulations required to achieve a satisfactory solution. An alternative and complementary goal would be shortening the time required by the hydraulic simulations. Other techniques worth of consideration would be partial simulation of schedules and approximated evaluation of candidate schedules.

7. Appendix

7.1. Model Parameters File Coding

```
clc; clear all;
%% pump selection
P1_enable=[1, 1, 1, 1];
% all active
P2_enable=[1, 1];

%% General
wat_dens=1000; %kg/m^3
grav_acc=9.81; %m/sec^2


%% pump 01 curve 350 m^3/h
pump_1_volume=[0.0 29.2 121.0 153.0 191.3 230.4 287.0...
    295.2 319.0 364.4 388.2 419.0 459.2 501.0 536.5];
% m^3/h
pump_1_head=polyval([-0.0001,-0.2370,253.3063],pump_1_volume);
%m
pump_1_eff=polyval([-0.0005,0.4163,-1.9548],pump_1_volume);
% 100%
pump_1_int_diam=0.0485; %pressure / volume setting
%m


pump_1_pressure=pump_1_head*grav_acc/100;
%bar
pump_1_hydr_pow=pump_1_volume.*pump_1_head*grav_acc*wat_dens/360
0;  %watts
pump_1_hydr_pow=pump_1_hydr_pow/1000;
% kW
pump_1_ele_pow=pump_1_hydr_pow./pump_1_eff;
% kW


%% pump 02 curve 120 m^3/h
pump_2_volume=[0.0 41.0 58.2 79.2 89.4 105.9 114.5 ...
    127.9 144.8 157.5 169.3 177.6 207.6 220.4 231.8...
    252.6 266.3 276.8 298.5];
% m^3/h
pump_2_head=polyval([-0.0006,-0.0069,118.3256],pump_2_volume);
%m
```

```
pump_2_eff=polyval([-0.0016,0.6416,16.2563],pump_2_volume);
% 100%
pump_2_int_diam=0.0437; %pressure / volume setting
%m

pump_2_pressure=pump_2_head*grav_acc/100;
%bar
pump_2_hydr_pow=pump_2_volume.*pump_2_head*grav_acc*wat_dens/360
0;  %watts
pump_2_hydr_pow=pump_2_hydr_pow/1000;
% kW
pump_2_ele_pow=pump_2_hydr_pow./pump_2_eff;
% kW

%% demand
demand_multiplier=0.002;
Lookup_time=1:1:24;
Lookup_values=[0.00 0.00 0.00 0.00 0.00 0.00 0.80 1.60 2.40
2.60...
    2.80 3.10 2.60 0.80 0.70 0.60 0.80 1.40 2.10 1.70 0.00
0.00...
    0.00 0.00]*demand_multiplier;
%% plots

figure(1)
plot(pump_1_volume,pump_1_hydr_pow,'k','Linewidth',1);
grid on;
title('Pump 01 efficiency curve')
xlabel('Volume m^3/h')
ylabel('Hydraulic Energy kW')

figure(2)
plot(pump_1_volume,pump_1_pressure,'r','Linewidth',1);
grid on;
title('Pump 01 Volume - Pressure curve')
xlabel('Volume m^3/h')
ylabel('Pressure bar')

figure(3)
plot(pump_2_volume,pump_2_hydr_pow,'k','Linewidth',1);
grid on;
title('Pump 02 efficiency curve')
xlabel('Volume m^3/h')
ylabel('Hydraulic Energy kW')

figure(4)
plot(pump_2_volume,pump_2_pressure,'r','Linewidth',1);
```

```
grid on;
title('Pump 01 Volume - Pressure curve')
xlabel('Volume m^3/h')
ylabel('Pressure bar')
```

7.2. Tank File Coding

```
function [ score ] = tank_obj( X )
% TANK_OBJ minimizes the score of the system

% score depends on energy consumption of the pumps and the
% availlability of the water in korakies tank

% X is an 144X1 array that has the binary on off values for the
six pumps used
% for every operation hour




%% basikes parametroi
arxiki_sta8mi=3.25;
demand_multiplier=0.0032;






%%
sim_step=3600;  %simulation interval sec
assignin('base','sim_step',sim_step);
Lookup_values=[0.00 0.00 0.00 0.00 0.00 0.00 0.80 1.60 2.40
2.60...
    2.80 3.10 2.60 0.80 0.70 0.60 0.80 1.40 2.10 1.70 0.00
0.00...
    0.00 0.00]*demand_multiplier;
assignin('base','Lookup_values',Lookup_values);

score=0;

korakies_d=37;%m
korakies_cs=pi*(korakies_d^2)/4; %m^2
```

```
assignin('base','korakies_cs',korakies_cs);

volume_korakies=arxiki_sta8mi*korakies_cs; %m^3
assignin('base','volume_korakies',volume_korakies);

%% General
wat_dens=1000; %kg/m^3
grav_acc=9.81; %m/sec^2
%% pump 01 curve 350 m^3/h
pump_1_volume=[0.0 29.2 121.0 153.0 191.3 230.4 287.0...
    295.2 319.0 364.4 388.2 419.0 459.2 501.0 536.5];
% m^3/h
pump_1_head=polyval([-0.0001,-0.2370,253.3063],pump_1_volume);
%m
pump_1_eff=polyval([-0.0005,0.4163,-1.9548],pump_1_volume);
% 100%
pump_1_int_diam=0.0485; %pressure / volume setting
%m


pump_1_pressure=pump_1_head*grav_acc/100;
%bar
pump_1_hydr_pow=pump_1_volume.*pump_1_head*grav_acc*wat_dens/360
0;  %watts
pump_1_hydr_pow=pump_1_hydr_pow/1000;
% kW
pump_1_ele_pow=pump_1_hydr_pow./pump_1_eff;
% kW


%% pump 02 curve 120 m^3/h
pump_2_volume=[0.0 41.0 58.2 79.2 89.4 105.9 114.5 ...
    127.9 144.8 157.5 169.3 177.6 207.6 220.4 231.8...
    252.6 266.3 276.8 298.5];
% m^3/h
pump_2_head=polyval([-0.0006,-0.0069,118.3256],pump_2_volume);
%m
pump_2_eff=polyval([-0.0016,0.6416,16.2563],pump_2_volume);
% 100%
pump_2_int_diam=0.0437; %pressure / volume setting
%m

pump_2_pressure=pump_2_head*grav_acc/100;
%bar
pump_2_hydr_pow=pump_2_volume.*pump_2_head*grav_acc*wat_dens/360
0;  %watts
```

```
pump_2_hydr_pow=pump_2_hydr_pow/1000;
% kW
pump_2_ele_pow=pump_2_hydr_pow./pump_2_eff;
% kW

%%
op_hour=1;
while op_hour<=24,
    assignin('base','op_hour',op_hour);
    H_a=((op_hour-1)*6);
    P1_enable=X(H_a+1:H_a+4)
    assignin('base','P1_enable',P1_enable);
    P2_enable=X(H_a+5:H_a+6)
    assignin('base','P2_enable',P2_enable);
    init_water_vol=volume_korakies(end);
    assignin('base','init_water_vol',init_water_vol);
    init_water_lev=init_water_vol/korakies_cs;

    assignin('base','init_water_lev',init_water_lev);

    sim('Model_obj',sim_step)
    sta8mi=(volume_korakies(end)/korakies_cs)

    if sta8mi<=0 || sta8mi>=8,
        score=score+1000*(24-op_hour);
        assignin('base','score',score);
        op_hour=24;
    end

    if sta8mi<=0.5 || sta8mi>=6,
        punish=1;
    else
        punish=0;
    end

    p=polyfit(pump_1_volume,pump_1_ele_pow,2);
    p1_pow=polyval(p,mean(p1_flow));
    p2_pow=polyval(p,mean(p2_flow));
    p3_pow=polyval(p,mean(p3_flow));
    p4_pow=polyval(p,mean(p4_flow));
    p=polyfit(pump_2_volume,pump_2_ele_pow,2);
    p5_pow=polyval(p,mean(p5_flow));
    p6_pow=polyval(p,mean(p6_flow));
    score=score+p1_pow+p2_pow+p3_pow+p4_pow+p5_pow+p6_pow;
    score=score+ punish*100;

    op_hour=op_hour+1;
```

```
end
score
```

7.3. Topographies of Case Study Vlite Water Network

Below are depicted the topographies of Vlite's network system that we used for the development of our model which we acquired from OAK AE.

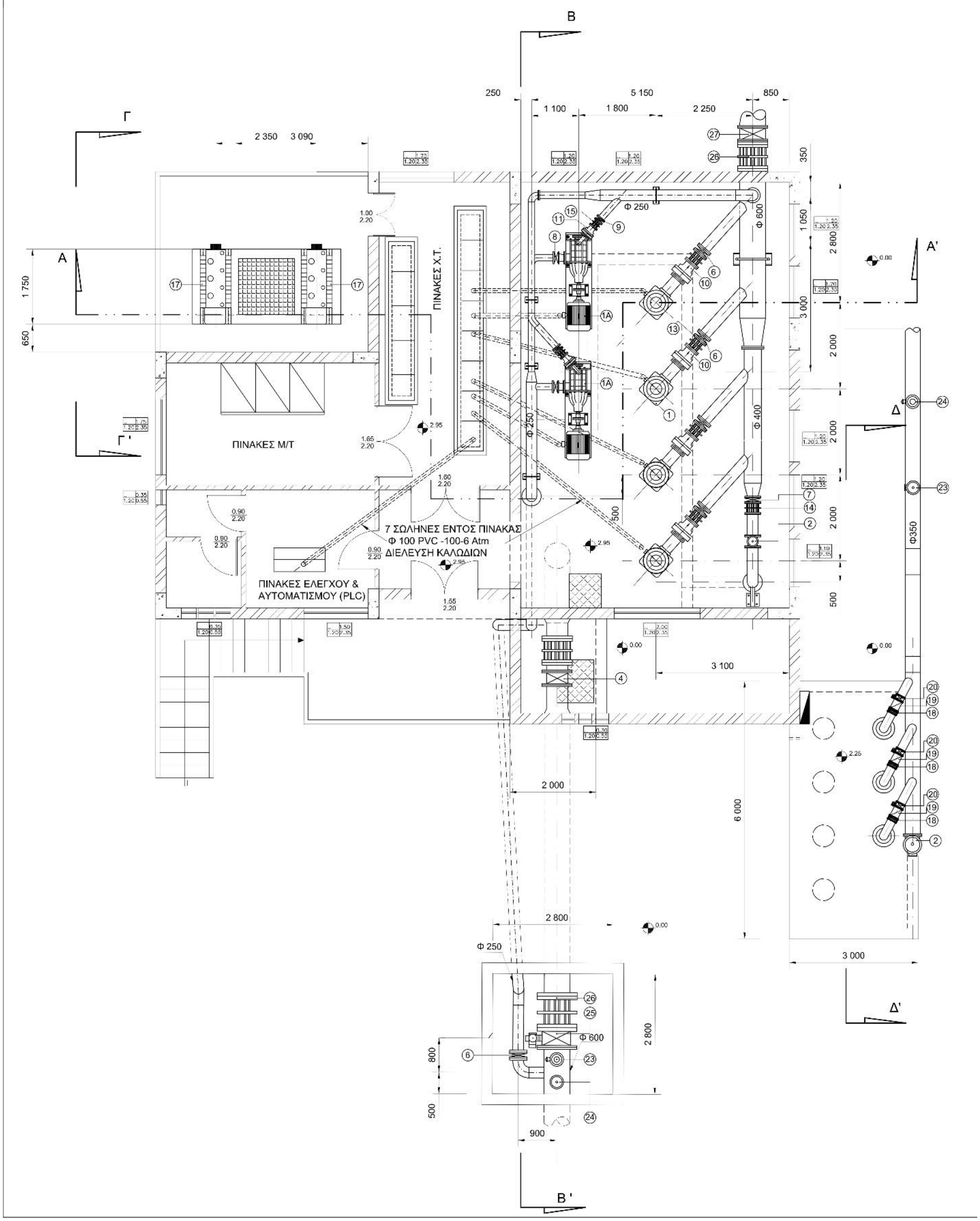
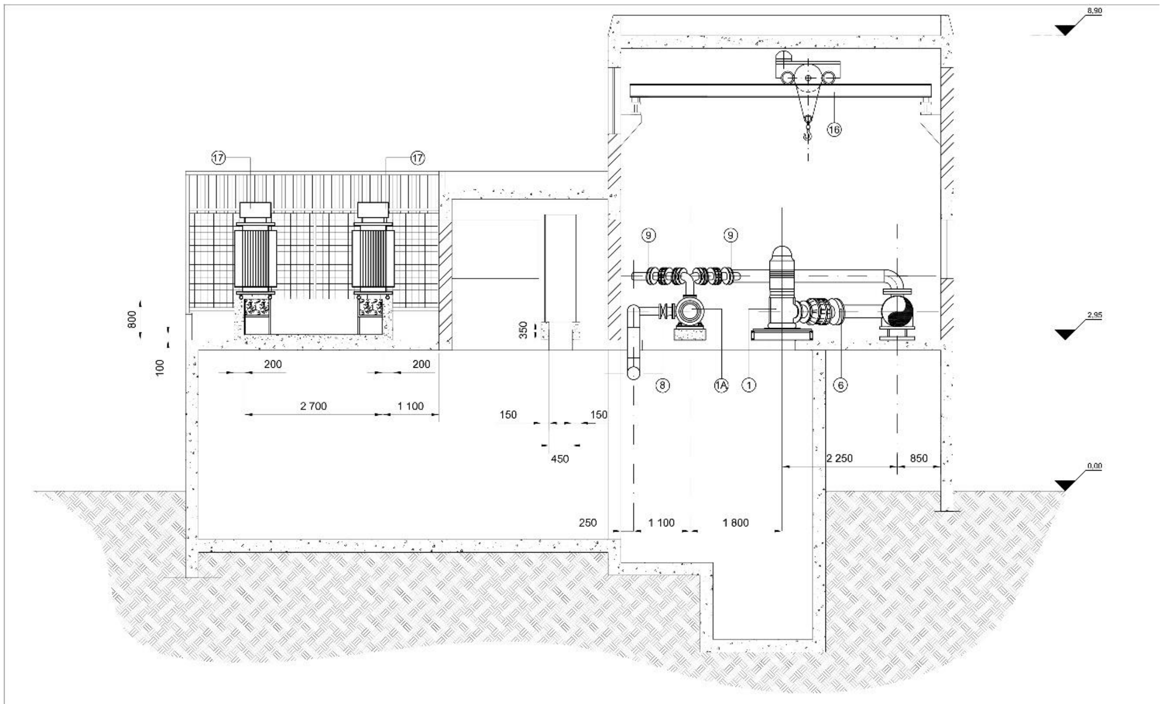
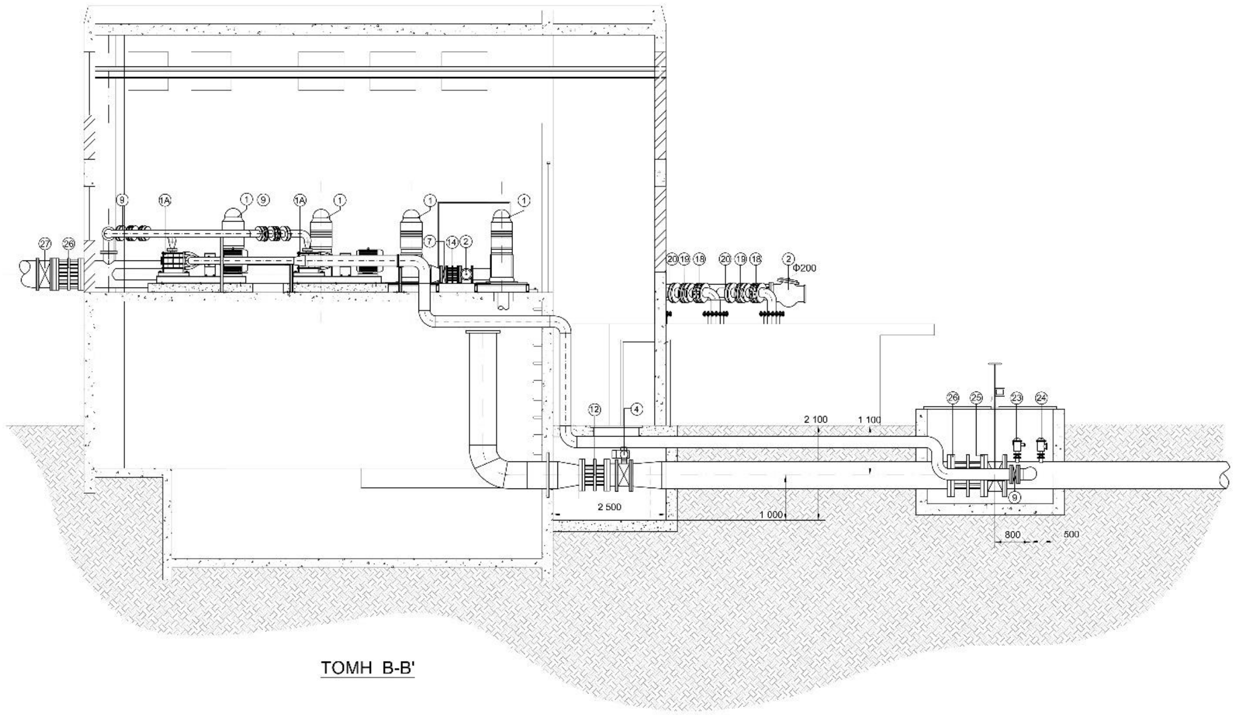


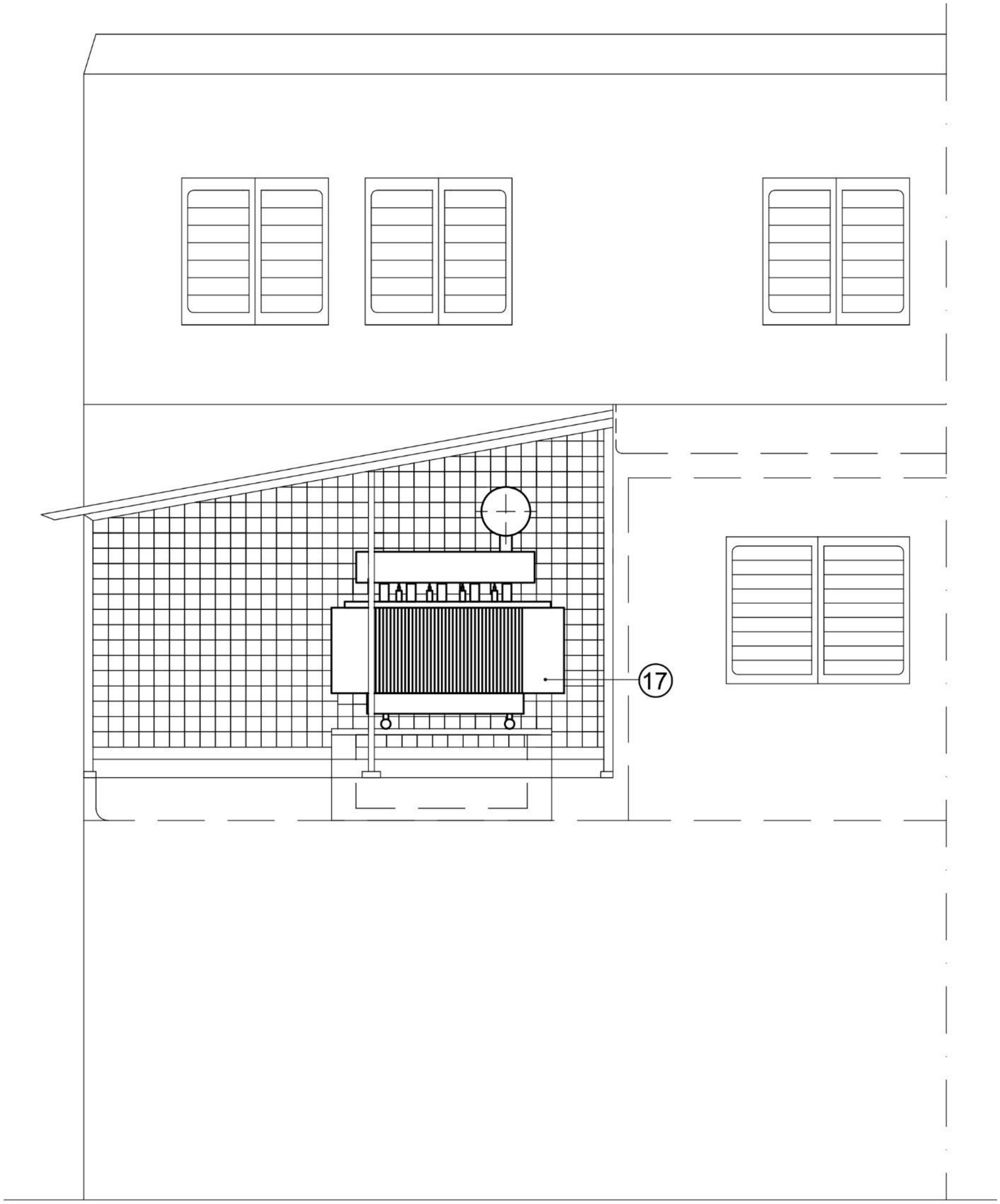
Fig 7.1. Topography of Vlite distribution system


8,90
2,95
0,00
800
100
350
200
200
2 700
1 100
150
150
450
250
1 100
1 800
2 250
850

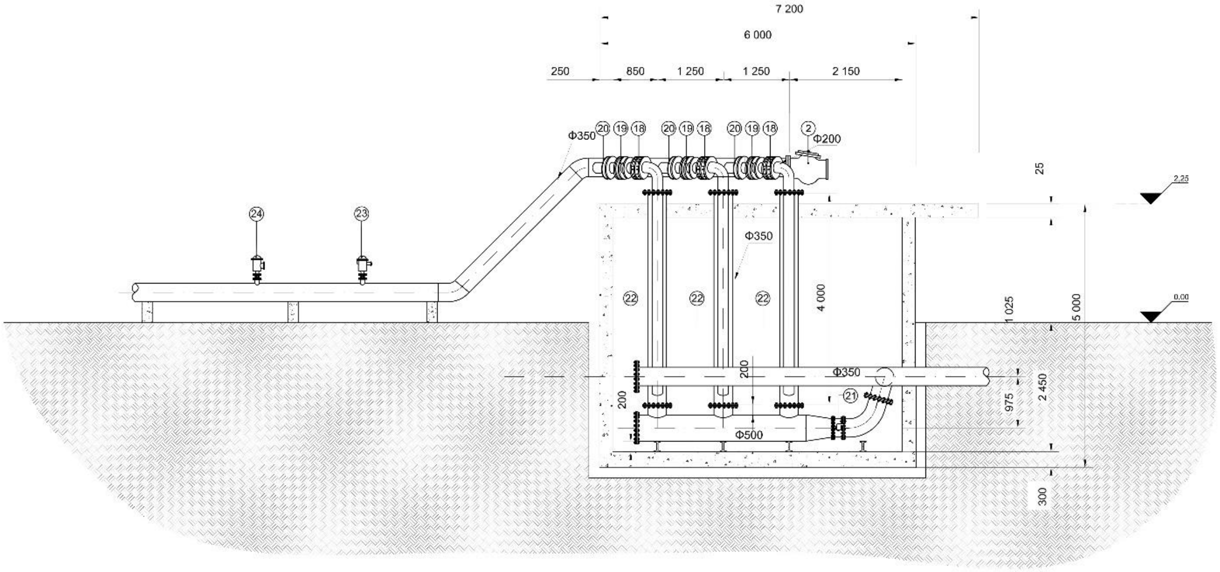
ΤΟΜΗ Α-Α'



ΤΟΜΗ Β-Β'



ΤΟΜΗ Γ-Γ'


7 200
6 000
250
850
1 250
1 250
2 150
Φ350
Φ200
Φ350
Φ350
Φ500
4 000
25
2.25
0.00
1 025
5 000
2 450
975
200
200
300

ΤΟΜΗ Δ-Δ'

8. References

1) Alperovits, E., and Shamir, U., “Design of Optimal Water Distribution System” Water Resources Research, 13: 885-900, 1977.
2) Abdel-Gawad, H.A.A., "Optimal Design of Pipe Networks by an Improved Genetic Algorithm," Proceedings of the Sixth International Water Technology Conference IWTC 2001, Alexandria, Egypt, March 23-25, 2001, pp. 155-163.
3) Abebe, A.J., and Solomatine, D.P., "Application of Global Optimization to the Design of Pipe Networks," Proc. 3rd International Conference on Hydroinformatics, Balkema, Rotterdam, Copenhagen, August 1998.
4) Alperovits, E., and Shamir, U., "Design of Optimal Water Distribution Systems," Water Resources Research, Vol. 13, No. 6, 1977, pp. 885-900.
5) Berghout, B.L., and Kuczera, G., "Network Linear Programming as Pipe Network Hydraulic Analysis Tool," Journal of Hydraulic Engineering, ASCE, Vol. 123, No. 6, 1997, pp. 549-559.
6) Bhave, P.R., "Optimal Expansion of Water Distribution Systems," Journal of the Environmental Engineering, ASCE, Vol. 111, No. 2, 1985, pp. 177-197.
7) Bhave, P.R., and Sonak, V.V., "A Critical Study of the Linear Programming Gradient Method for Optimal Design of Water Supply Networks," Water Resources Research, Vol. 28, No. 6, 1992, pp. 1577-1584.
8) Busetti, F., "Genetic Algorithms Overview," 2007, [Online document] Available at: http://www.geocities.com/francorbusetti/gaweb.pdf
9) AWWA, “Water-Distribution Research & Development Needs.” Journal of the American Water Works Association. pp. 385-390, June 1974.
10) Bush, C.A., and Uber, J.G., “Sampling Design Methods for Water Distribution Model Calibration”, J. of Water Res. Plan. and Man., ASCE, 124(6) ,334- 344, 1998.
11) Eiger, G., Shamir, U., and Be-Tal, A., “Optimal Design of Water Distribution Networks.” Water Resour. Res., 30(9), 2637-2646, 1994.
12) Goulter, I. C., Coals, A. V., “Quantitative Approaches to Reliability Assessment in Pipe Networks”, J. Transp. Engrg., ASCE, Vol. 112, No. 3 287-301, 1986.

13) Castillo, L., and González, A., "Distribution Network Optimization: Finding the Most Economic Solution by using Genetic Algorithms,"European Journal of Operational Research, Vol. 108, 1998, pp. 527-537. http://hera.ugr.es/doi/15018799.pdf

14) Cenedese, A., Gallerano, F., and Misiti, A., "Multiobjective Analysis in Optimal Solution of Hydraulic Networks," Journal of Hydraulic Engineering, ASCE, Vol. 113, No. 9, 1987, pp. 1133-1143.

15) Churchill, S.W., "Friction Factor Equations Spans All Fluid-Flow Regimes", Chemical Engineering, Vol. 84, 1977, pp. 91-92. Cunha, M.D.C., and Sousa, J., "Water Distribution Network Design Optimization: Simulated Annealing Approach," Journal of Water Resources Planning and Management, ASCE, Vol. 125, No. 4, 1999, pp. 215-221.

16) Dandy, G.C., and Engelhardt, M., "Optimal Scheduling of Water Pipe Replacement using Genetic Algorithms," Journal of Water Resources Planning and Management, ASCE, Vol. 127, No. 4, 2001, pp. 214-223.

17) Dandy, G.C., Simpson, A.R., and Murphy, L.J., "An Improved Genetic Algorithm for Pipe Network Optimization," Water Resources Research, Vol. 32, No. 2, 1996, pp. 449-458.

18) Deb, A.K., and Sarkar, A.K., "Optimization in Design of Hydraulic Networks," Journal of the Sanitary Engineering Division, ASCE, Vol. 97, No. SA2, 1971, pp. 141-159.

19) Deboeck, G. J., (Editor), Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets, Wiley, 1994.

20) Goulter, I. C. ‘‘Analytical and simulation models for reliability analysis in water distribution systems.’’ Improving efficiency and reliability in water distribution systems, E.Cabrera and A. F. Vela, eds., Kluwer Academic, London, 235–266, 1995.

21) Jacobsen, S., Dishary, Murphy, Frey, “Las Vegas Valley Water District Plans for Expansion Improvements using Genetic Algorithm Optimization” Proc. of the AWWA Information and Technology Conference, AWWA, Reno, Nevada,1998.

22) Jowitt, P.W., and Germanopoulos, G., “Optimal Pump Scheduling in Water Supply Networks”, J. of Water Res. Plan. and Man., ASCE, 118(4), 406-422, 1992.

23) Kettler, A., and Goulter, I., “Reliability Consideration in The Least Cost Design of Looped Water Distribution Networks.” Proc., 10th Int. Symp. On Urban Hydro., Hydr. and Sediment Control, University of Kentucky, Lexington, Ky., 305-312, 1983.

24) Lansey, K. E., and Mays, L. W. “Optimization Model for Water Distribution System Design.” J. Hydr. Engrg., ASCE, 115(10), 1401-1418, 1989.

25) Lansey, K. E., and Basnet, C., ”Parameter Estimation for Water Distribution Networks” J. of Water Res. Plan. and Man., ASCE 117(1), 126, 1991.

26) Mays, L. W., ed., “Reliability Analysis of Water Distribution Systems.”, American Society of Engineers, New York, 1989.

27) Meier, R.W., and Barkdoll, B.D., “Sampling Design for Network Model Calibration using Genetic Algorithms”, J. of Water Res. Plan. and Man., ASCE, 126(4), 245-250, 2000.

28) Ormsbee, L. E., Walski, T. M., Chase D. V., Sharp, W.,” Methodology for Improving Pump Operational Efficiency”, J. Water Res. Plan. and Man., ASCE, 15(2), 148, 1989.

29) Percia, C., Oron, G., and Mehrez, A., “Optimal Operation of Regional System with Diverse Water Quality Sources.” J. of Water Res. Plan. and Man., ASCE,123(2), 105-115, 1997.

30) Piller, O., Bremond, B., and Morel, P., “A Spatial Sampling Procedure for Physical Diagnosis in a Drinking Water Supply Network”, Proc. of Water Ind. Sys.: Modeling and Optimization Applications, Savic and Walters, eds., Vol.1, Exeter, United Kingdom, 1999.

31) Quindry, G., Brill, E. D., and Liebman, J. C. “Optimization of Looped Distribution Systems.” J. Environmental Engineering Division, ASCE, 107(4), 665-679, 1981.

32) Rossman, L. A, Boulos, P. F. and Altman, T., “Discrete Volume-Element Method for Network Water-Quality Models”, J. Water Resour. Plng. and Mgmt., Vol. 119, No. 5, 505-517, 1993.

33) Savic, D., and Walters, G., “Genetic Algorithms for Least-Cost Design of Water Distribution Networks.” J. Water Resour. Plng. and Mgmt., ASCE, 123(2), 67-77, 1997.

34) Schanke, J. C., and Lai, D. “Linear Programming and Dynamic Programming Applied to Water Distribution Network Design.”, MIT Hydrodynamics Lab Rep. 116, Cambridge, Massachusetts, 1969.

35) Shamir, U., and, Howard, C. D. D., “Reliability and Risk Assessment for Water Supply Systems”, Proc., ASCE Speciality Conference on Computer Applications in Water Resources, Buffalo, N.Y., pp. 1218-1228, 1985.

36) Simpson, A. R., Dandy, G. C., and Murphy, L. J. “Genetic Algorithms compared to other techniques for pipe optimization.”, J. Water Resour. Plng. and Mgmt., ASCE, 120(4), 423-443, 1994.

37) Su, Y., L., Mays, N. Duan, and K. Lansey, "Reliability Based Optimization for water Distribution Systems," ASCE Journal of Hydraulic Engineering,113: 589- 596, 1987.

38) Walski, T. M., "Technique for Calibrating Network models" J. of Water Res. and Plan. and Man., ASCE, 109(4), 360, 1983.

39) Walski, T. M., "State-of-the-Art: Pipe network optimization.", Computer Applications in Water Resources, Torno, H., ed., ASCE, New York, 1985.

40) Walski, T. M., Youshock, M., and Rhen, H., "Use of Modeling in Decision Making for Water Distribution Master Planning" Proc. of the ASCE EWRI conference, Minneapolis, Minnesota, 2000.

41) Walski, T. M., "The Wrong Paradigm - Why Water Distribution Optimization doesn't work", J. of Water Res. Plan. and Man., ASCE, 127(2), 203, 2001.

42) Xu, C., and Goulter, I., "Reliability-Based Optimal Design of Water Distribution Networks.", J. Water Resour. Plng. and Mgmt., ASCE, 352-362, 1999.

43) Yu, G., Powell, R.S., and Sterling, M.J.H., "Optimized Pump Scheduling in Water Distribution Systems ", J. of Opt., Theory and Appl., 83(3), 463, 1994.

44) http://www.jade-cheng.com/au/coalhmm/optimization/

45) http://www.pg.gda.pl/~mkwies/dyd/geadocu/algparal.html

46) http://stackoverflow.com/questions/23183862/genetic-programming-difference-between-roulette-rank-and-tournament-selection