POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA CIVILE
AMBIENTALE E TERRITORIALE

Effectiveness of a genetic algorithm to prioritize the rehabilitation plan of leaky water distribution networks

Tesi di Laurea Magistrale in
Civil Engineering - Ingegneria Civile

Author: **Luca Preite**

Student ID: 969490
Advisor: Prof. Stefano Malavasi
Co-advisors: Giacomo Ferrarese
Academic Year: 2022-23

Abstract

When dealing with leaky Water Distribution Networks (WDNs), the need of a proper rehabilitation plan is becoming more and more urgent due to recent climate changes and serious water scarcity. Often, a WDN manager is faced to several decision variables to schedule the interventions on the network but it is not always straightforward to assess the optimal intervention order.

In this work, several methodologies to define intervention schedules are proposed on the basis of engineering knowledge and expertise, a-priori considerations led by experience, and ex-ante known variables. These schedules are referred to as Ex-ante, or Engineering-Expertise, or Experience-led Schedules (ESs).

Then the study is focused on the development of a Genetic Algorithm (GA) able to find an optimal intervention schedule for a given network. This schedule is then compared to ESs in the analysis of the performance of both the algorithm and ESs, also considering the computational effort required for the elaboration.

Several water distribution networks are considered to study the behaviour of the methodologies applied to different networks. In particular, four WDNs with different size and properties are analysed: from little networks of 22 nodes to large networks of 2859 nodes.

Several objectives can be of interest for the optimization of the interventions scheduling; the proposed approach allows the network manager to customize the optimization problem following its necessities. Different objectives are studied and two fitness functions are proposed. The first fitness function tries to minimize the Water Loss Volume (WLV) throughout the entire network during the whole intervention period. The second fitness function brings the problem back to an economic point of view: it is a multi-objective (MO) function that tries to minimize the Intervention and Management Cost (IMC).

Finally, various consideration and conclusion could be appreciated: the solutions from the genetic algorithm had marginal improvements in comparison with ESs, but it allowed the analysis of the fitness functions in relation to different ESs and to different networks.

Keywords: leakages, water distribution networks, rehabilitation plan, genetic algorithm

Abstract in lingua italiana

Quando si ha a che fare con reti di distribuzione idriche degradate, la necessità di un adeguato piano di riabilitazione diventa sempre più urgente a causa dei recenti cambi climatici e conseguente scarsità d'acqua. Un gestore idrico si trova spesso ad affrontare diverse variabili di decisione e trovare l'ordine ottimale degli interventi non è banale.

In questo lavoro sono proposti diversi piani di intervento sulla base di conoscenze ingegneristiche, considerazioni preventive e guidate dall'esperienza e variabili conosciute a priori. Ci si riferisce a queste soluzioni come soluzioni ES.

In seguito, lo studio è focalizzato sullo sviluppo di un algoritmo genetico (GA) in grado di trovare una soluzione ottima data una rete. Questa soluzione è quindi comparata alle ES per analizzare la performance di entrambe le soluzioni delle ES e la soluzione dell'algoritmo, considerando anche il costo computazionale richiesto dall'algoritmo.

Molteplici reti idriche sono state considerate per studiare il comportamento delle metodologie applicate ai diversi tipi di reti. In particolare, quattro reti di diverse dimensioni e caratteristiche sono state analizzate: da piccole reti di 22 nodi fino a reti di 2859 nodi.

Nell'ottimizzazione, numerosi obiettivi possono essere di interesse per l'ottimizzazione della pianificazione degli interventi; ciò permette al gestore idrico di personalizzare il problema di ottimizzazione a seconda delle sue necessità. Differenti obiettivi sono stati studiati e due funzioni fitness sono proposte in questo lavoro. La prima tenta di minimizzare il volume di acqua perso (WLV) attraverso l'intera rete per tutta la durata dell'intervento. La seconda funzione fitness riporta il problema in termini economici: è una funzione multi-obiettivo che tenta di minimizzare il costo di intervento e gestione (IMC).

Infine, varie considerazioni e conclusioni possono essere dedotte: le soluzioni dell'algoritmo genetico hanno miglioramenti marginali se comparati con le ES, tuttavia l'algoritmo ha permesso l'analisi delle funzioni fitness in relazione alle differenti ES e alle differenti reti.

Parole chiave: perdite, reti di distribuzione idriche, piano di riabilitazione, algoritmo genetico

Contents

1 | Introduction

Water distribution networks (hereinafter named WDNs) are one of the most important infrastructures in modern cities. They consist of several artifacts that collect, store, and distribute water between supply sources and consumers [4]. In addition, the need to address water losses has recently led to a growing interest in studies of leakage modeling, detection, and repair.

Leakages in water distribution networks are often diffuse and scattered over several areas. For this reason, it is not always straightforward to plan the most efficient sequence of interventions to rehabilitate leaks. Often, water utilities have to plan the interventions based on their practical experience.

The goal of this work is to find new schedules that optimize the rehabilitation plan. New methodologies such as genetic algorithms are investigated. Also, the effectiveness of the new methodology is compared with practical engineering expertise.

The challenges faced in this work are similar to those faced during the 9th Battle competition of the joint WDSA/CCWI conference: the Battle of Intermittent Water Supply [24], held at the Universitat Politècnica de València in July 2022. A five-years rehabilitation plan of a WDN [12] was proposed for the battle.
The BIWS intensively inspired the present work in several ways. First, the input files for the networks were homologated and BIWS was used as an example.
Second, the intervention methodologies and their costs were used to repair the leakages as explained in chapter 2.5.

1.1. Problem

The task that is attempted to be accomplished is to find an optimal intervention schedule to plan the rehabilitation of a leaky water distribution network.
The problem is transformed into an optimization problem with an objective function. The objective function represents the requirements and goals of a network manager faced with the rehabilitation of the network.

In the paper, two objective functions are studied and the proposed methods were tested on several WDNs.

1.2. Material

The network properties and analyses are achieved using the Epanet 2.2 software [21]. The hydraulic simulations required to assess the hydraulic state of the network are performed utilizing the Open Water Analytics/Epanet-Matlab toolkit [7]. All developed algorithms and scripts are coded in Matlab.

Basically, two input files are needed for each network to be rehabilitated: an Epanet inp-file (without leakages) and a table containing for each leakage its magnitude and position. The latter is given by pipe and length from the initial node of the pipe where the leakage is located. This format is based on the material provided by BIWS.

In Epanet, leakage can be modeled as a pressure-driven phenomenon that follows the generalized form of the orifice equation, which is expressed as a power law as in equation 1.1 [9].

$$Q_{leak}(t) = k \cdot p(t)^{\gamma} \tag{1.1}$$

where the leaking flow $Q_{leak}[l/s]$ at a given time t is proportional to the leakage magnitude $k[l/s/m^{\gamma}]$ and the pressure head $p[m]$ raised to the pressure exponent γ.

The leakage magnitude k is also called the effective leakage area [10] because, when compared to the orifice equation 1.2 (also known as Torricelli's formula), it is equivalent to the leakage area A_{leak} times the discharge coefficient C_d and the square root of 2 times the gravitational constant g.

$$Q_{leak}(t) = A_{leak} C_d \sqrt{2gp(t)} \tag{1.2}$$

The value of the pressure exponent is equal to 0.5 for the orifice equation, which denotes only circular orifices/leakages. In real situations, its value depends on the pipe material and the causes of the opening and can vary in a wide range of values as low as 0.36 and even as high as 2.5 [10]. The emitter exponent is fixed for every leakage within the same network and its value is set to 0.5 if no information is provided by the source of the network.

2 | Benchmark networks

In order to study the proposed methodologies, diverse networks were required in the attempt to highlight different behaviours. An open-access repository of water distribution networks was found online: the Hydraulic Model Database from Ormsbee et al. [18]. It is a comprehensive table which summarizes several characteristics of several networks and it is subdivided into 8 sections. In addition, other networks found in the literature [1, 7, 24] were considered and attached in the last section labeled "09 added". The table can be found in Appendix A and includes all the networks that were taken into consideration for the analysis. Some of these networks were selected to be more suitable for the studies.

Initially, small benchmark networks were selected for the first phase of algorithm implementation. After this phase, it is realized that their simplicity did not allow a proper evolution of the GA. Afterwards, more complex networks were considered, but the computational cost of running the algorithm over several generations increased exponentially. Then, medium networks were used to compensate for these two obstacles.

Clearly, the presence of leakages was required. Among the considered networks of the Appendix A, only two networks have built-in leakages and those were one of the smallest (Araujo network) and one of the largest (BIWS network) with the complications already explained. Therefore, a script was created to insert a number of leakages at casual positions. This allowed to create networks that would be suitable for the studies.

In addition, in the selection of benchmark networks, particular attention was paid to the presence of tanks: these elements have been found to cause instability in Epanet simulations, so networks without tanks are preferred.

The benchmark networks for this work are listed in table 2.1. Mainly, they are Araujo network [1], BIWS network [24], Balerma network [20], MOD network [7]. They are presented and detailed in the following sections.

Table 2.1: Benchmark networks characteristics. The * indicates manually added leakages

Network		Araujo	BIWS	Balerma	MOD
Nodes	#	22	2 859	443	268
Pipes	#	37	3 231	454	317
NetworkLength	m	44 261	338 662	100 263	71 806
BaseDemand	m^3/d	12 960	17 519	95 377	35 160
Citizens	#	50 000	70 000	380 000	140 000
Reservoirs	#	3	6	4	4
Leakages	#	22	3 589	100*	100*
LeakFlow	m^3/d	192	14 475	2 735	2 381
maxDistance	m	8 592	12 534	1 152	6 037
medianDistance	m	4 302	3 090	328	2 301

The size of the network can be perceived looking at number of nodes, pipes and length of the network (which is the sum of the pipe lengths).
The amount of citizens supplied can be estimated by considering the daily base demand of the nodes in the network and an average water supply of 0.25 m^3/d per capita. This quantity is only an indication, no data were found in the literature.
Finally, the number of leakages is shown together with the daily leaking flow in the current network status (CNS). The * indicates networks where leakages were manually randomly added, while other networks already had the presence of leakages in literature.
In addition, the distance between the furthest leakages and the median distance between leakages are computed. they are indications for the vicinity cost that will be introduced later on.

2.1. Araujo network

The first WDN that has been analysed was the Araujo network. It was firstly proposed by Jowitt and Xu (1990) [13], but in this work is used the modified version of Araujo et al.(2006) [1].

It is a small benchmark network characterized by 22 nodes, 37 pipes for a total of 44 200 km, 3 sources. A total daily base demand of 13 000 m^3/d suggests a population of about 50 000 persons. It is affected by 22 leakages that result in a total daily water loss of 192 m^3/d (please refer to table 2.1). Incidentally, the pre-existing leaks are placed on nodes, while other WDNs have leakages located on pipes: to better simulate the leakages in the network and homologate the Araujo with the other WDNs, the built-in leakages have been moved from the node to a random connected pipe at a random distance. In this way, the Epanet inp-file deprived of emitters and the table containing the leakages and their characteristics have been created.

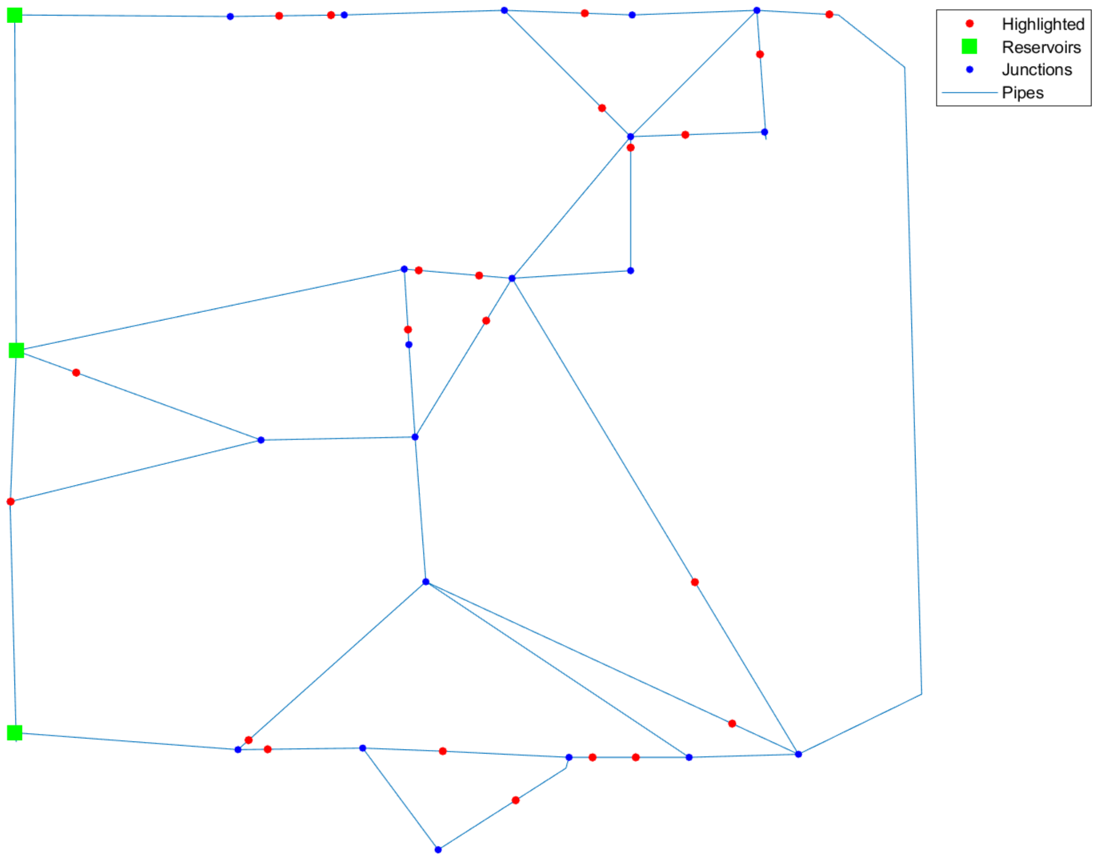

Figure 2.1: Sketch of the modified Araujo network. Leakages are highlighted.

This network was used mainly because of the already built-in leaks and its size. Such a simple network has a low computational cost and allows fast evaluations and simulations with the GA. However, its simplicity leads to a reduced search space and the GA had fewer opportunities to evolve.

2.2. BIWS network

The second WDN that has been analysed is the BIWS network. It was utilised in the competition of the Battle of Intermittent Water Supply 2022 [24].

It is a large network characterized by 2 859 nodes, 3 231 pipes for a total of 339 000 km, 6 sources, and 4 tanks. A total daily base demand of 17 500 m^3/d suggests a population of about 70 000 persons. It is affected by 3 589 leakages that result in a total daily water loss of 14 500 m^3/d (please refer to table 2.1).

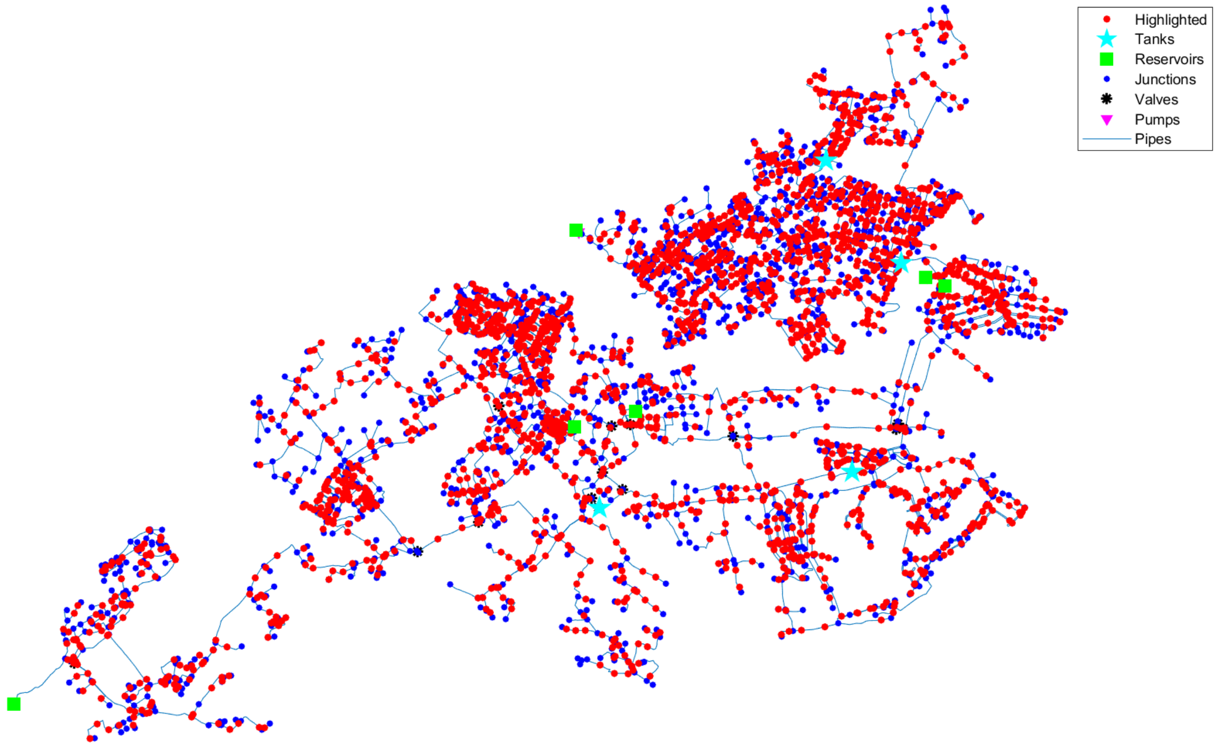

Figure 2.2: Sketch of the BIWS network. Leakages are highlighted.

This network was considered mainly because it either already had integrated leakage or because it had already been studied for participating in the Battle of Intermittent Water Supply [12]. It necessitated a rehabilitation plan that minimized the cost, but analysing the network different problems were observed: firstly, the network is in very poor condition, having a low quality of service, both in terms of water supply and pressure; pressure has a strong effect on leakage flow and becomes a problem with widespread negative pressures in the current network status (CNS). In fact, during the battle, in addition to repairing leakages, other measures were needed to improve the network.
Second, the BIWS network has a high computational cost due to the large number of leakages and the complexity of the network: it is affected by a number of leakages that is either more than 30 times the number of leaks introduced randomly in the Balerma and MOD networks, and almost 160 times the number of leakages affecting the Araujo network. The large number of leaks results in a large number of genes in each individual, which has a remarkable impact on the runtime per individual and the runtime per population.

Therefore, this network could not be suitable for the purposes of this work. By the way, an attempt was made to simplify the network with a clustering technique: instead of giving a bonus in the fitness function to prioritize the vicinity of leakages in subsequent interventions, the leaks were clustered to reduce the number of total leaks. However, the number of clusters was not sufficiently reduced to an appropriate number.

2.3. Balerma network

Another WDN was searched to satisfy the need for a more complex network then Araujo for a larger search space, but a less complex network then BIWS for an acceptable runtime. Hence, the Balerma network was selected within the database for its medium size but leakages are manually added. It was first presented by Reca and Martinez (2006) [20] and is based on the Sol-Poniente irrigation district.

Balerma network is characterized by 443 nodes, 454 pipes for a total of 100 000 km, and 4 sources. A total daily base demand of 95 000 m^3/d suggests a population of about 380 000 persons. A number of 100 random leakages is introduced in the network for a total daily water loss of 2 700 m^3/d (please refer to table 2.1).

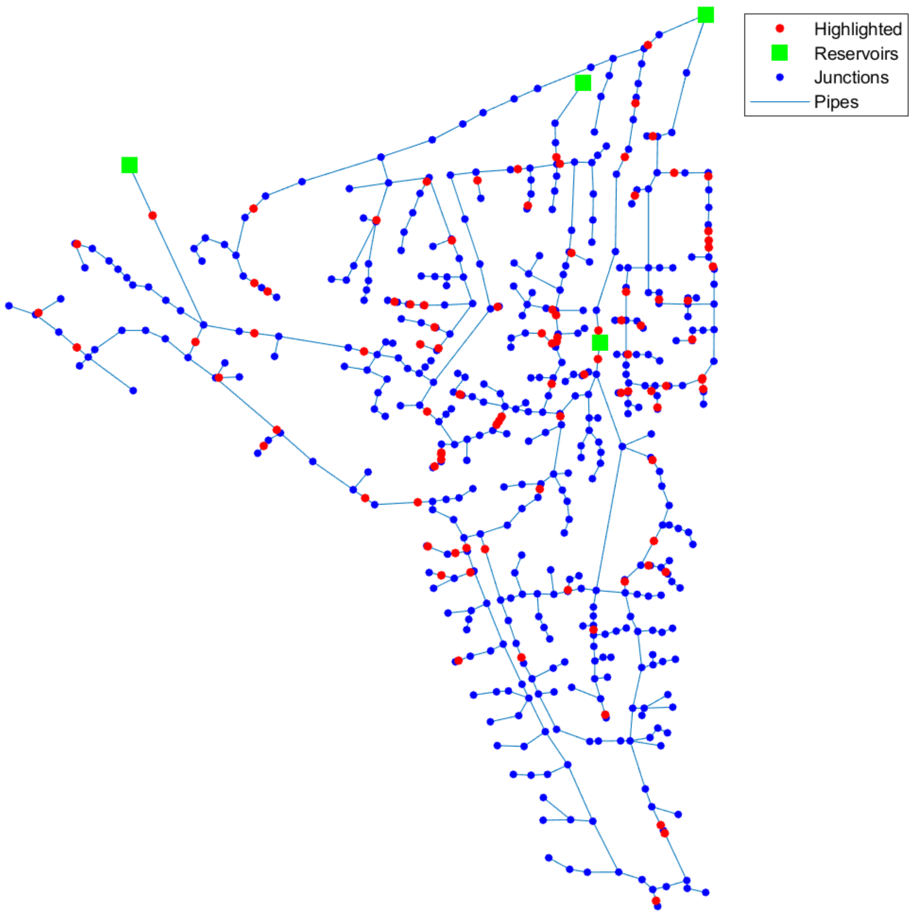

Figure 2.3: Sketch of the Balerma network. Added leakages are highlighted.

2.4. MOD network

An alternative to Balerma is MOD network. It was found in the examples folder of the Epanet/Matlab toolkit [7], but no additional information was found there.

It is a medium network characterized by 268 nodes, 317 pipes for a total of 71 800 km, and 4 sources. A total daily base demand of 35 000 m^3/d suggests a population of about 140 000 persons. A number of 100 random leakages is introduced in the network for a total daily water loss of 2 400 m^3/h (please refer to table 2.1).

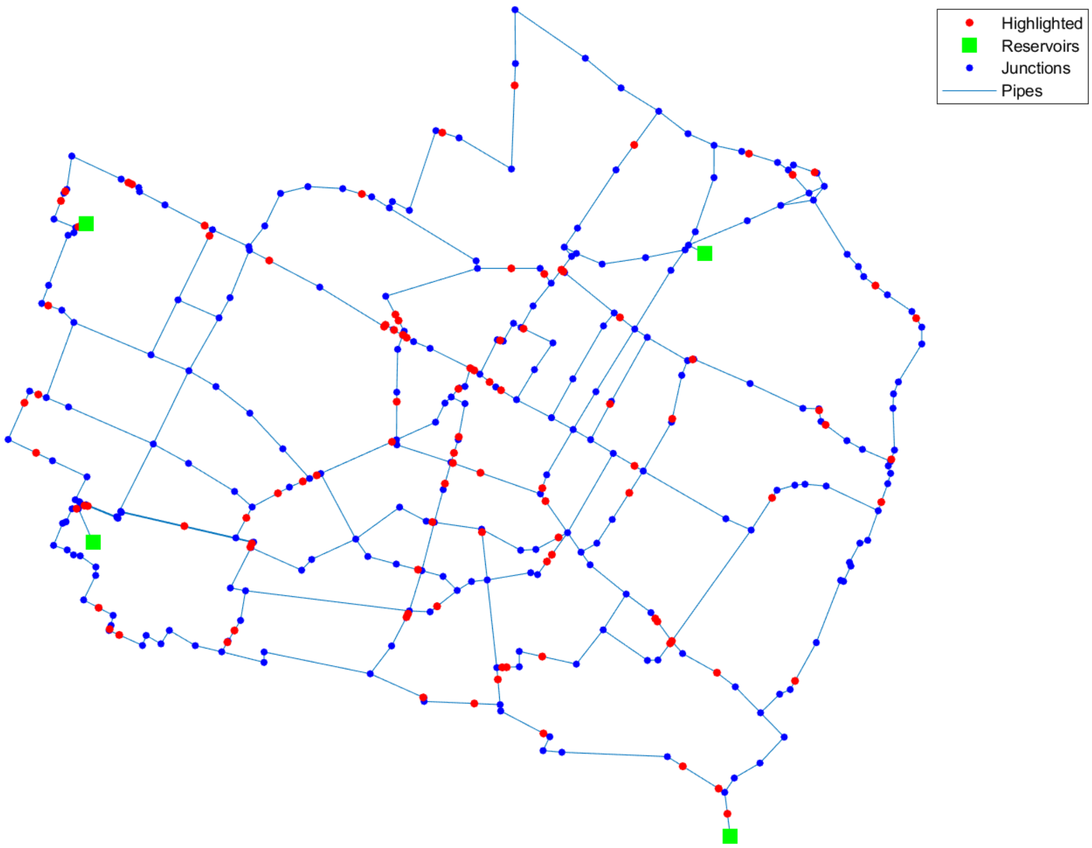

Figure 2.4: Sketch of the MOD network. Added leakages are highlighted.

2.5. Pre-processing and intervention cost assessment

Due to the lack of information concerning the leakages and their characteristics, two intervention modalities were considered, as was the case with the Battle of IWS [24]: the repair of the leak itself and the replacement of a portion of the pipe containing one or more leakages.

The costs associated with these interventions are the same proposed in [24], namely:

$$C_{repair}(D,k) = \left[94 - 0.3 \cdot D + 0.01 \cdot D^2\right] \cdot \left[1.5 + 0.11 \cdot \log_{10}(k)\right] \tag{2.1a}$$

$$C_{replace}(D,L) = \left[13 + 0.029 \cdot D + 0.0012 \cdot D^2\right] \cdot L \tag{2.1b}$$

where C_{repair} is the repair cost in € of a leakage of magnitude $k[l/s/m^\gamma]$ in a pipe of diameter $D[mm]$, while $C_{replace}$ is the replacing cost in € of a portion of pipe of length $L[m]$ and diameter $D[mm]$.

For each leakage, the intervention between the aforementioned two modalities is pre-determined with a pre-processing procedure based on the cost: for a given leakage, the cheaper intervention between repair and replacement is chosen and its cost is set. Moreover, when the replacement intervention is chosen to remove multiple leakages, they are integrated into a single emitter helping to reduce the number of leakages and genes, and hence the complexity of the problem, and hence the computational cost and runtime.

The network pre-processing script is schematized in Appendix B. It imports the leakage table and the input model, compares the replacing and repair costs, eventually groups the interventions, and then modify the input model to insert the emitter nodes.

A recursive function, schematized in Appendix C, was required in order to compare costs: for each pipe with multiple leakages, the cost of replacing portions of the pipe containing different leaks was compared to the repair costs of all the leaks in that portion of the pip. The least expensive configuration is selected, its intervention cost is evaluated, and possibly the interventions are grouped.

Afterwards, the input model is modified to insert the leakages in the exact position in the network: for each leakage, a new node has to be created and its coordinates are computed knowing end nodes and vertices of the pipe and the length from the initial node. Then, it is set as an emitter with the proper leakage magnitude and no base demand. The old pipe is then split into two pipes with the original properties. When every leakage is inserted into the network, the latter is saved into a new inp-file named "networkname_leak.inp" and it is ready to be analysed.

The pre-process is not much computationally expensive (few minutes for the largest BIWS network) and it has to be run just one time. Nevertheless, it helps to reduce the number of leakages if there are any convenient placed: as it will be explained in the next chapter, the computational cost of the proposed algorithm is drastically decreased if the number of leakages is reduced.

3 | Ex-ante Schedules (ESs)

This chapter presents several schedules that might prove to be good solutions to the problem. They are defined a-priori based on various parameters of the network and the leakage characteristics and configuration. This is intended to mimic the thought process of a network manager intervening in the network based on his engineering skills and experience knowledge to intervene in the network. These solutions are referred to as Ex-ante, or Engineering-Expertise, or Experience-led Schedules (ESs).

Several parameters can affect the scheduling such as:

- the entity of the leakage: its magnitude, shape and characteristics;
- the pressure the leakage is exposed to, which is tied to the leaked flow;
- the position of the leakage within the network: the diameter of the pipeline, the presence of other leakages in the vicinity;
- the location of the leakage outside the network: the accessibility of the pipeline;
- the intervention on the leakage: the type, the cost, the duration, the interruption of the supply to users.

Based on some of the listed parameters, several ESs can be defined, each prioritizing different aspects in the intervention plan. The considered ESs are explained in more detail below.

- ES-Ca, a schedule based on intervention cost in ascending order: the cheapest leakages are prioritised;
- ES-Cd, a schedule based on intervention cost in descending order: the most expensive leakages are prioritised;
- ES-kd, a schedule based on the water leakage magnitude in descending orders: the largest leakages are prioritised;
- ES-Pd, a schedule based on CNS pressures in descending order: the leakages located in high pressure areas are prioritised;

- ES-Qd, a schedule based on CNS leakage flow in descending order: the leakages with higher leak flowrate are prioritised;
- ES-Vk and ES-VQ, schedules based on leakage vicinity starting from the node with larger leakage magnitude (ES-Vk) or larger leakage flow (ES-VQ): the shortest path between leakages is prioritised.

The list of these schedules is presented in Table 3.1, where their prerequisites are highlighted.

ES name	Leakage position	Leakage entity	Pressure at node	Hydraulic simulations
ES-Ca	No	Yes	No	None
ES-Cd	No	Yes	No	None
ES-kd	No	Yes	No	None
ES-Pd	Yes	No	Yes	CNS
ES-Qd	Yes	Yes	Yes	CNS
ES-Vk	Yes	Yes	No	None
ES-VQ	Yes	Yes	Yes	CNS

Table 3.1: Ex-ante Schedules and their prerequisites.

Intervention cost ordered schedules (ES-Ca, ES-Cd)

The ES-Ca and ES-Cd are the first schedules considered: the network manager might choose to repair the cheapest leakages first for obvious reason, or he might choose to fix the most expensive leakages before they get worse.
These schedules do not require hydraulic simulation and can be easily estimated even in the absence of information about the network. Although it may appear that they do not take into account the characteristics of either the leaks or the network, they do in fact take into account the variables that affect costs (equations 2.1), such as the leakage magnitudes, the diameter of the pipes, and the spacing between leaks on the same pipe (i.e., the replacement lengths).

Leakage magnitude ordered schedule (ES-kd)

The ES-kd is one of the most intuitive ways to plan a rehabilitation schedule: leakages are repaired from the largest to the smallest entity.
This schedule does not require any hydraulic simulation and can be easily estimated

even in the absence of information about the network. However, it does not take into consideration the pressure trend at the emitter nodes and does not contemplate that a smaller leakage could have a larger flow if the hydraulic head is high.
It will come out that this schedule is good enough to optimize the water loss volume (chapter 5).

CNS pressure heads ordered schedule (ES-Pd)

The ES-Pd is the schedule that takes into consideration the pressure trend at the emitter nodes: the leakages are repaired from the largest to the smallest pressure head as computed in the CNS.

For this reason, this schedule requires only a hydraulic simulation along with all the information needed to launch the Epanet simulation and evaluate the pressure state of the network. However, it does not take into account the entities of leakages.

CNS leaking flow ordered schedule (ES-Qd)

The ES-Qd is the combination of ES-kd and ES-Pd: it fully considers the leakage entities and the pressure heads by scheduling the interventions from the largest to the smallest leaking flow, as computed in the CNS.

As for ES-Pd, this schedule requires only a hydraulic simulation to evaluate the hydraulic state of the network and the leaking flow of the emitters. But in general, the condition of the network can change drastically after an intervention, resulting in different pressures, different leakage flows, and different water loss volumes from the CNS.

However, it will turn out that this schedule is the best one to save the largest water volume from leakages (chapter 5): such drastic changes of the network state do not occur in the studied networks. This could be due to the simplicity of the networks, and it could actually be the optimal solution to the WLV problem.

Vicinity ordered schedules (ES-Vk, ES-VQ)

The ES-Vk and ES-VQ are the schedules proposed to mimic the manager's decision to proceed spatially from one leak to the closest leak.

This schedule requires only the positioning of the leakages and a starting point: two starting points are chosen to form two schedules, the first starts from the node with the larger leakage magnitude (ES-Vk), the second starts from the larger leakage flow in the CNS (ES-VQ). In the analyzed networks, the starting points are often the same node and both schedules are identical. A hydraulic simulation of the CNS is required to determine the maximum leakage flow for the latter schedule, instead any information about the network is required in the first case.

It will be found that these schedules are the best to optimize the vicinity cost, but the IMC fitness function takes into account various costs that affect the solution (chapter 6).

4 | The Genetic Algorithm (GA)

For the purpose to find better solutions to the problem, modern technologies such as better and high performance computers can be useful. However, the problem consists in finding schedules that give better results with respect to the considered objective function and there is not an analytical resolution: a way to assess the optimal solution could be the evaluation of every possible permutation of the intervention order and the global minima is then easily found. This is not a feasible way forward because there exist $N!$ permutations of a N-elements vector and a network could have hundreds and thousands of leakages to be repaired.

Heuristic techniques are developed to solve a problem when classical methods cannot find an exact solution, or to solve a problem faster when classical methods are too slow to find an approximate solution. This is achieved by trading optimality, completeness, accuracy, or precision for speed. A heuristic is a function that ranks alternatives based on available information to decide which branch to follow [19].

Nowadays, a number of optimization methods and algorithms are known and developed. The most important aspects to take into consideration when choosing an algorithm are the width of the search space, the computational cost, the velocity of convergence toward a solution, the risk of falling into local minima, the ability to escape from local minima, etc [23].

A genetic algorithm has been implemented for the study. With all its settings, it should be able to satisfy the requirements to find a new, better schedule. A genetic algorithm is a metaheuristic (a higher-level heuristic) inspired by the process of natural selection and belongs to the larger class of evolutionary algorithms (EA).
Genetic algorithms are widely used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection [16].

4.1. General structure of a GA

In a genetic algorithm, a population of candidate solutions (called individuals) is evolved into better solutions to the optimization problem. Each candidate solution consists of a set of genes that can be altered and mutated [25]. The evolution is an iterative process that usually starts from a population of randomly generated individuals. In each iteration, the population is called a generation and the fitness of each individual is evaluated. The fitness function is the objective function of the optimization problem to be solved. In each generation, the better individuals are stochastically selected from the current population to form a new generation by utilising some genetic operators such as crossover breeding and random mutation of their genes. The new generation of candidate solutions is then used in the next iteration of the algorithm. Usually, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level for the population has been reached.

In this work, the GA is stopped when a plateau is reached and the algorithm is not able to escape the local minima after a certain number of iterations and does not produce better results.

The figure 4.1 represents the flowchart of the genetic algorithm structure.

4.1.1. Individuals and genes

Traditionally, individuals are represented by binary genes as strings of zeros and ones, but other encodings are also possible [8]. For the purpose of the problem, the solutions are the schedules of the interventions on the leakages; in other words, the repair order of the leakages is desired. Therefore, the individuals are modeled as permutations of the leakage indices, with each gene containing a leakage index to be repaired: the intervention order is given by the sequence of the indices (genes) in the individual. The decision to work with permutation individuals has consequences for the genetic operators that normally work with binary genes [15], especially for the crossover and mutation operators.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

6	13	12	1	7	14	4	18	8	3	17	10	20	5	9	15	2	11	19	16

Figure 4.2: Two examples of schedules consisting in permutations of leakages indices

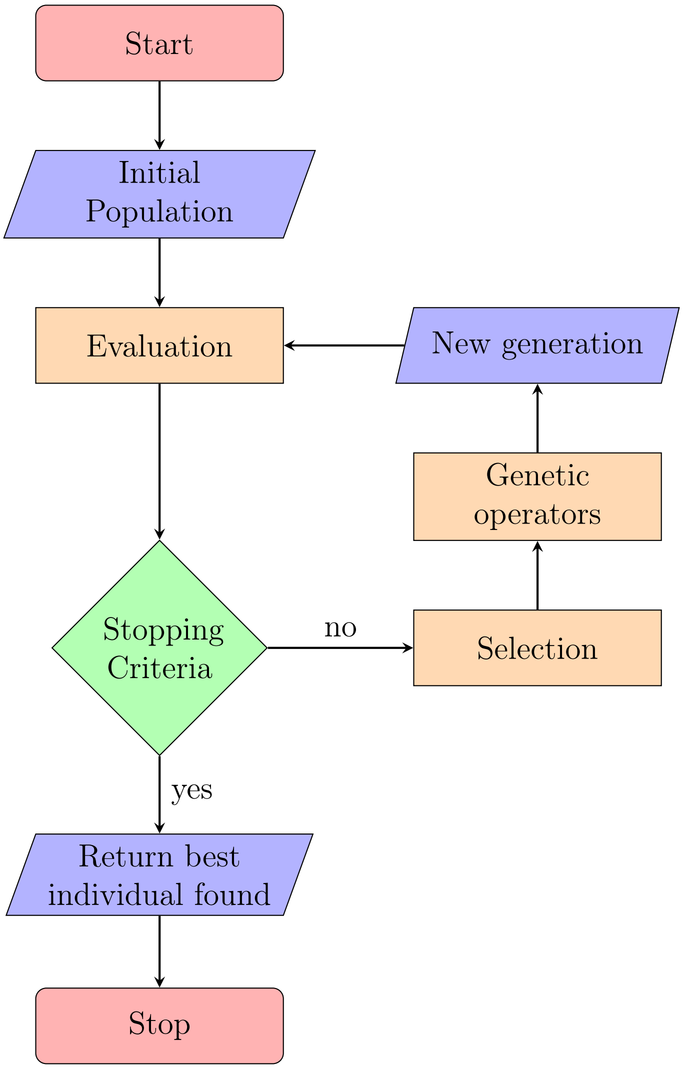

Figure 4.1: Flowchart of the general structure of a Genetic Algorithm.

A population consists of N_i individuals in total, and each individual consists of N_g genes, which is equal to the number of leakages in the network.

4.1.2. Evaluation

In each generation, each individual is evaluated against a specific fitness function. The goal of the fitness function is to summarize in a single value how close the individual is to the objectives of the optimization problem. Several fitness functions are considered in the study and they will be explained as needed.

The evaluation needs a hydraulic simulation of the network behavior using the Epanet-Matlab toolkit [7] after each intervention: for each l-th gene of a given individual, the leakages are repaired until the l-th position of the schedule, the hydraulic simulation is launched, and the l-th fitness value is assessed. The fitness of such an individual would be the sum of the fitness values of its genes.

4.1.3. Crossover breeding

Crossover, also called recombination, is a genetic operator analogous to the crossbreeding that occurs in biology during sexual reproduction. It allows the genotype of two parents to be combined to produce new offspring [5].
There exist several crossover techniques, and the choice also depends on the genetic representation used in the GA. When crossover operators for binary individuals, such as one- or n-point or uniform crossover, are used for permutation genomes, a child individual may contain some values twice and others may be missing. This can be remedied by genetic repair, e.g., by replacing the redundant genes with the missing ones.
In order to avoid the generation of invalid offspring, special crossover operators for permutations have been developed [15], which satisfy the basic requirements of such operators for permutations, namely that all elements of the original permutation are also present in the new one and only the order is changed. Some examples of the most commonly used crossover operators for permutations are Partially Mapped Crossover (PMX) [6], Order crossover (OX1) [17], Cycle crossover (CX) [17], Order-based crossover (OX2) [22], and Position-based crossover (POS) [22].

In this work, a novel crossover operator, called Head&Tail Crossover (HTX), is proposed. It is similar to the Order crossover (OX1) of Oliver et al. [17] with just one crossover point: fixed a crossover position in the middle of the individual, the crossover point halves two parent individuals. Let Head1 and Tail1 be the first and second half of the first parent respectively; similarly, let Head2 and Tail2 be the first and second half of the second parent respectively (please refer to figure 4.3). Consequently, the head of one individual consists of a portion of the permutation genotype and contains some of the genes required to compose an offspring; the remaining genes (Rem) that are missing in the genotype of the offspring are collected from the other parent in the order as they appear. Hence, four offsprings can be created by either combining Head of a parent and Rem of the another or combining Tail of a parent and Rem of another, and viceversa: a visual representation of this operator is shown in figure 4.3.

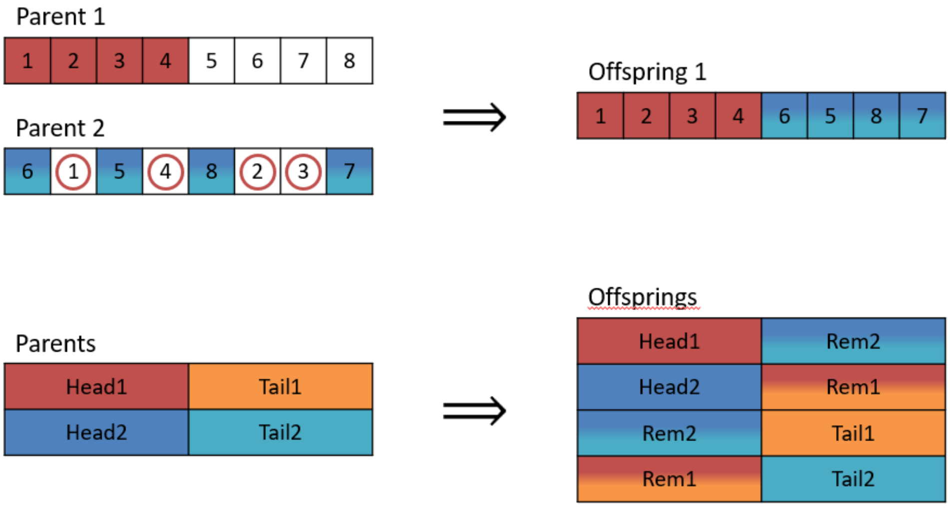

Figure 4.3: Visual representation of the Head&Tail crossover operator.

The preservation of the first part of the parent individual gives importance to the first leakages to be repaired and shuffles the remaining ones, focusing the evolution of the second half of the individual. The same occurs with the preservation of the second half of the genes.

4.1.4. Mutation

Mutation is another genetic operator that corresponds to biological mutation. Its purpose is to preserve the genetic diversity of individuals in a population and to better explore the search space of the optimization problem to be solved. Mutation operators are used in an attempt to avoid local minima by preventing populations from becoming too similar, thus slowing or even stopping convergence to the global optimum [5]. This consideration also leads most GAs to avoid taking only the fittest of the population when generating the next generation.
There exist several mutation techniques, and their choice also depends on the genetic representation used in the GA. For individuals that are a permutation of a set, operators such as Rotation to the right [26] or Inversion [6] can be used.

In this work, a novel mutation operator, called Switch mutation, is proposed. It is similar to the Inversion mutation operator from Eiben and Smith [6]: for a given number of N_s switches per mutating individual, N_s genes are randomly selected to be exchanged with other N_s randomly selected genes within the same mutating individual: in total, $2 \cdot N_s$ genes are mutated. A visual representation of this operator is shown in Figure 4.4.

Figure 4.4: Visual representation of the Switch mutation operator.

Moreover, another strategy is proposed to increase the genetic diversity of the new population, i.e., the number of switches per mutation is not constant: during the mutation process, N_s increases proportionally to the number of mutated individuals, so that some mutated individuals have a low number of switches while others have a high number of switches.

4.1.5. Random addition

To complete the new population, a number of N_r individuals are added to the next generation. These are entirely new individuals, completely randomly generated and not derived from previous individuals. The purpose of this addition is the same as for mutation: to preserve even more the genetic diversity of the population and to search for new optimal solutions further in the search space towards the global minimum of the optimization problem being solved.

The number of random individuals per population is set to be: $N_r = 1/10 \cdot N_i$

For the same purpose, another operation is proposed: after the creation of the new population, the individuals are examined to check the presence of duplicate identical individuals; if there are any, they are replaced by completely randomly generated individuals.

4.2. Selection and Creation of the new generation

In the selection phase, individuals are selected from the current population to which the various operators are applied. In literature, various selection methods have been proposed and compared [11], such as Roulette Wheel Selection, Rank selection, Steady-state selection, Tournament selection, Elitist selection, and Boltzmann selection.

Each operator could have its own selection method and they are explained later on.

4.2.1. Elite selection

Elitism is a strategy in which a small portion of the current generation is carried over without changes to the next generation. This strategy guarantees that the quality of the solution achieved by the GA does not diminish from one generation to the next [2].

In this work, each generation contains exactly $N_e = 1$ elite individual from the previous generation, i.e. the fittest individual.

4.2.2. Crossover selection

In each generation, a total number of N_c individuals are allocated to be created by cross-breeding. Using the HTX crossover, two parents are needed to create four offsprings hence, a number of $N_c/2$ parents is needed to be crossbred. At the beginning, the selection of these $N_c/2$ individuals was elitist. However, elite parents, which were very similar among themselves, used to generate offsprings which were very similar to their parents (in double amount). This led to a portion of N_c individuals of the population to quickly converge toward a single individual. Hereinafter, half of the parents were the better $N_c/4$ individuals which would be crossbred with the worst $N_c/4$ individuals of the previous generation.

The adoption of HTX crossover require N_c to be a multiple of four. In this work, the number of crossbred individuals per population is set to be: $N_c = 4/10 \cdot N_i$.

4.2.3. Mutation selection

Finally, the new generation is completed with a number of $N_m = N_i - N_e - N_c - N_r$ individuals from the mutation operator. In this case, an elitist selection is used to provide two individuals to the mutation operator: the two fittest individuals are used as sample to be randomly mutated several times and to create N_m new individuals.

As consequence of the already set parameters, the number of mutated individuals per population results to be

$$N_m = N_i - 1 - \frac{4}{10}N_i - \frac{1}{10}N_i = \frac{5}{10}N_i - 1$$

.

4.2.4. New generation

In conclusion, the new generation is created and a visual representation is provided in Figure 4.5.

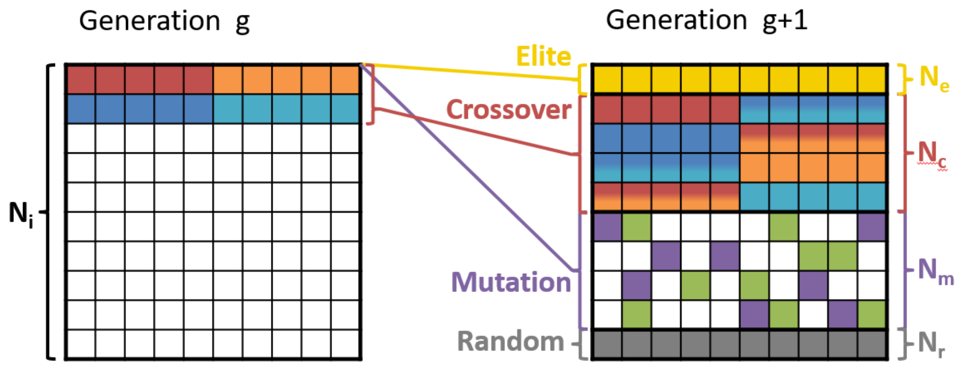

Figure 4.5: Visual representation of the new generation structure.

4.3. Explanation of GA results: plots, graphs and tables

In this paragraph, an explanation of some results is provided which will be useful later on in the thesis, starting from the next section.

The chart of the kind of figure 4.6 shows the evolution of the GA: several lines depict the different individuals of each generation; the generations increase on the x-axis. Individuals are coloured based on their fitness value from minimum (blue, best) to maximum (red, worst) among the current population. The lowest and thicker line represents the best individual in each generation and the red crosses are its improvement: each red cross is a new optimal solution. Finally, the lowest point of the lowest thick line represents the optimal solution for planning the intervention schedule.
The title lists respectively the Fitness function, the network, the number of genes which is equal to the number of the leakages within the network, the total runtime and the total number of Epanet simulations needed. The total number of generations and individuals is reported in the x-axis label, along with the composition of each generation.

The typology of the table 4.2 shows the results of the different schedules analysed: from left to right they are listed the fitness value of each schedule, the percentage improvement of each schedule over the worst solution, and the comparison of the optimal solution with respect to each schedule.

4.4. Initial generation

As mentioned earlier, GAs usually start from a population of randomly generated individuals. However, since this is a practical and engineering problem, the search could start from the well known candidate solutions presented in chapter 3. Therefore, some of the initial individuals are set to be exactly the ESs, and the GA can evolve starting from these solutions.

The possibility to use this strategy is proved by launching two runs of the same GA, on the same network, with same settings. The only change is that they start respectively from completely random and partially given initial population.
The results presented later on in the paragraph 5.2 regarding the WLV optimization on the Araujo network are hereunder repeated with the same GA, on the same network, and with the same setting but different initial population: the former (hereinafter PG) starts from a partially given initial population, while the latter (hereinafter CR) starts from a completely random generated population. The evolution of the CR run is presented in figure 4.6: the CR run evolved for 226 generations, starting from a completely random initial generation, until a plateau was reached for 50 successive generations (please refer to figure 4.6).

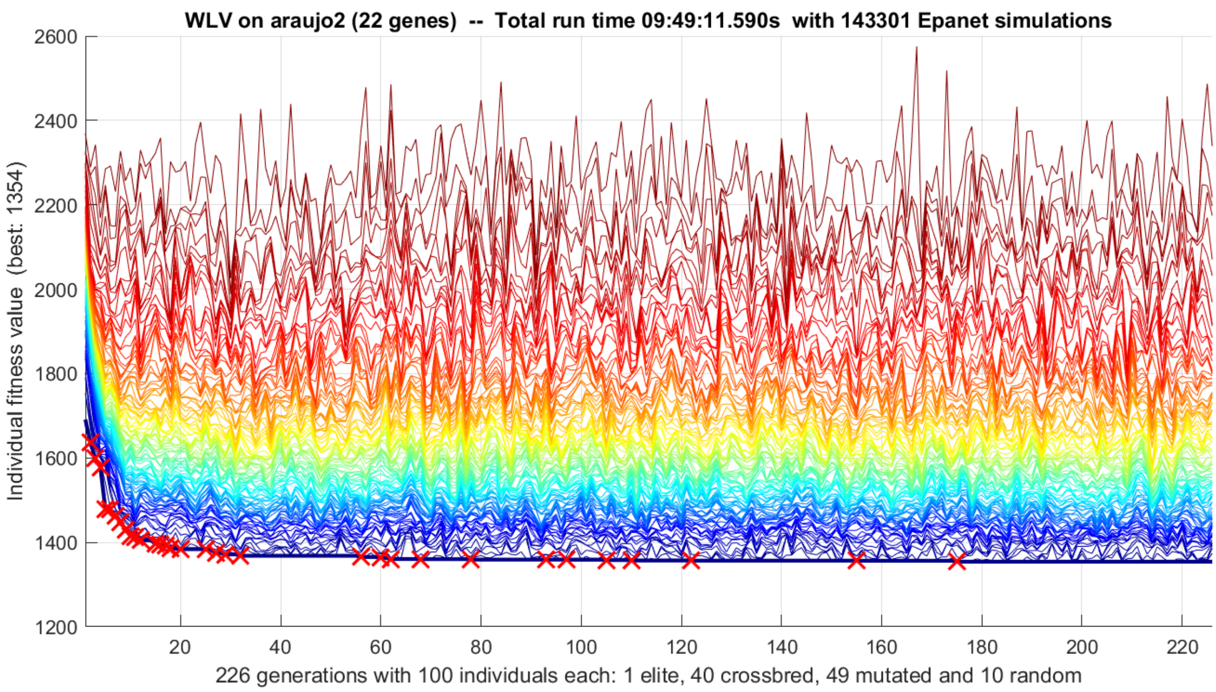

Figure 4.6: Evolution of the GA with completely random initial generation on the Araujo network.

In the figure 5.3, the GA did not evolve from the given ES-Qd during 101 generations, so it is assumed to be the optimal solution to the problem during the PG run. The CR solution converged towards ES-Qd and eventually it was almost reached: as it can be noted from table 4.1 which contains the indexes of the leakages to be repaired, one more switch between two successive genes is required to reach the ES-Qd solution.

ES-Qd	16	8	19	15	4	11	2	21	1	6	17	14	7	12	9	10	5	3	18	13	20	22
GA sol.	16	8	19	15	4	11	2	21	1	6	17	14	12	7	9	10	5	3	18	13	20	22

Table 4.1: Actual ES-Qd and GA solution schedules: ordered indices of leakages to be repaired

The table 4.2 shows the results of the CR run: the fitness value of the solution best GA (CR) differs from the solution ES-Qd (PG) by a small amount of 0.11 m^3 compared to the absolute value of about 1354 m^3. Eventually, the CR run would have achieved the same optimal solution as the PG run if it had developed more generations.

Schedule name	**WLV [m^3]**	**Improvement from worst [%]**	**Comparison of best [%]**
ES-Qd	1 354.06	-40.4	0.0
best GA	1 354.17	-40.4	0.0
ES-kd	1 365.42	-39.9	-0.8
ES-Cd	1 969.87	-13.3	-31.3
ES-Vk	1 981.96	-12.7	-31.7
ES-VQ	1 981.96	-12.7	-31.7
ES-Ca	2 120.88	-6.6	-36.2
ES-Pd	2 271.37	0.0	-40.4

Table 4.2: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

4.5. Code optimization

An attempt to reduce the computational cost and shorten the runtime of each individual is the research of similar individuals: instead of hydraulically simulating each gene of each individual, similar individuals are searched in the already evaluated populations of the previous generations.

Similar individuals until a certain l-th gene consist of the same first l repaired leakages regardless of the repair order hence, the l-th does not need a new hydraulic simulation. Anyway, the more generations are done the longer searching process could take: it has to search in a much larger database of individuals to try to find a similar individual. This process becomes much longer then a hydraulic simulation therefore, the research is discarded.

An actual optimization of the code has been implemented directly during the creation of the next generation: some of the new individuals created from the current generation are similar to the current individuals such as:

- elite individuals do not need to be simulated again;
- crossbreed individuals are similar to their parents until the crossover point;
- mutated individuals are similar to their original individuals until the first mutated gene and after the last mutated gene.

A pseudo-code of the developed GA is provided in Algorithm 4.1.

Algorithm 4.1 Pseudo-code of the Genetic Algorithm

```
%% Initialization
Input setting
Randomly initialize first generation
Introduce ESs in the initial population
%% Starting iterations
for fixed number of generations do
    for each individual do
        Fitness evaluation
    end for
    %% Formation of the next generation
    Individuals ranking and selection
    Elitists conservation
    Crossover breeding
    Mutation
    Random individual addition
end for
```

4.6. Training and settings

The initial studies on the development of the GA were conducted on the benchmark network firstly proposed by Araujo [1], a small WDN extensively discussed in chapter 2.1. The training phase allowed to study the evolution of the algorithm, its convergence, rate of finding new solutions, sensibility to the local minima and capacity to escape the local minima.

The completely random initial generation is used in the training of the GA to better study the evolution of the algorithm and to set its parameters. Thus, the settings of the GA have been chosen judging the behaviour of the GA evolution among several attempts.

A series of parameters values are listed which setting is been used to get the results later on.

A run of at least 100 generations is preferred to better observe the results; a stopover in terms of fitness improvement of 50 generations is chosen to end the evolution.

A population of 100 individuals is set for each iteration; the new generation is composed by 1 elite, 40 crossbred, 49 mutated, and 10 random individuals.

The number of switches for each mutation is determined by halving the number of genes (which depends on the number of leakages within the network).

5 | Water Loss Volume (WLV) optimization

The first objective function of the optimization problem is based on the Water Loss Volume: it is obtained considering the leaking flow through all the emitters in the whole network and a fixed time $t_{intervention} = 24h$ between repairing leakages. On that subject, reference is made to paragraph 5.6 of this chapter for the influence analysis of the intervention time on the results.

Hence, the individual fitness $F_i(s)$ of the schedule s is calculated as the sum of the l-th fitness value, which is the water leaked volume in the entire network in the l-th configuration, i.e. after repairing the l-th leakage.

$$F_i(s) = \sum_{l=1}^{N_{leaks}} F_l = \sum_{l=1}^{N_{leaks}} WLV_l$$

where the leaked water volume after each intervention is given by the sum of the leaking flow through all the emitters as follows:

$$\begin{aligned} WLV_l &= \left[\sum_{n=1}^{N_{nodes}} Q_{leak,n} \cdot t_{intervention} \right]_l = \\ &= \left[\sum_{n=1}^{N_{nodes}} (k_n \cdot {p_n}^{\gamma}) \cdot t_{intervention} \right]_l \end{aligned} \tag{5.1}$$

The GA has been run trying to optimize this fitness function starting from the ESs. Different network have been analysed. For further detail on how to interpret the resulting plots and graphs please refer to chapter 4.3.

5.1. Explanation of WLV results graphs

In addition to the results explained in chapter 4.3, another graph is depicted in figure 5.1. It shows the development of the network during the implementation of each intervention schedule: the lines are the cumulative fitness values after each intervention; the final point represents the fitness of the individual.
The black dashed line represents the case where no intervention is made: the leaking flow is constant and the total volume of water loss increases linearly.
The red dotted line represents the optimal GA solution.

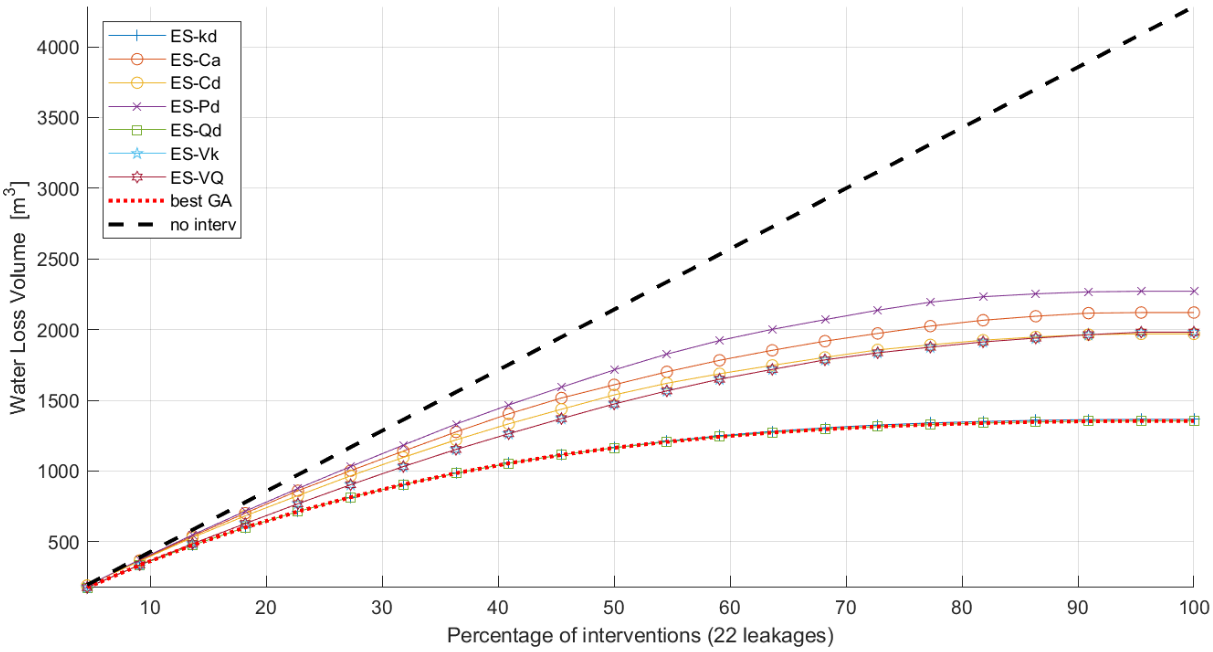

Figure 5.1: Exemplary comparison of ESs and GA solutions on the Araujo network (related to the run of figure 4.6)

5.2. WLV optimization on Araujo network

The first analysed WDN is the Araujo network [1]. As a first result, the comparison between the different ESs is shown in figure 5.2.

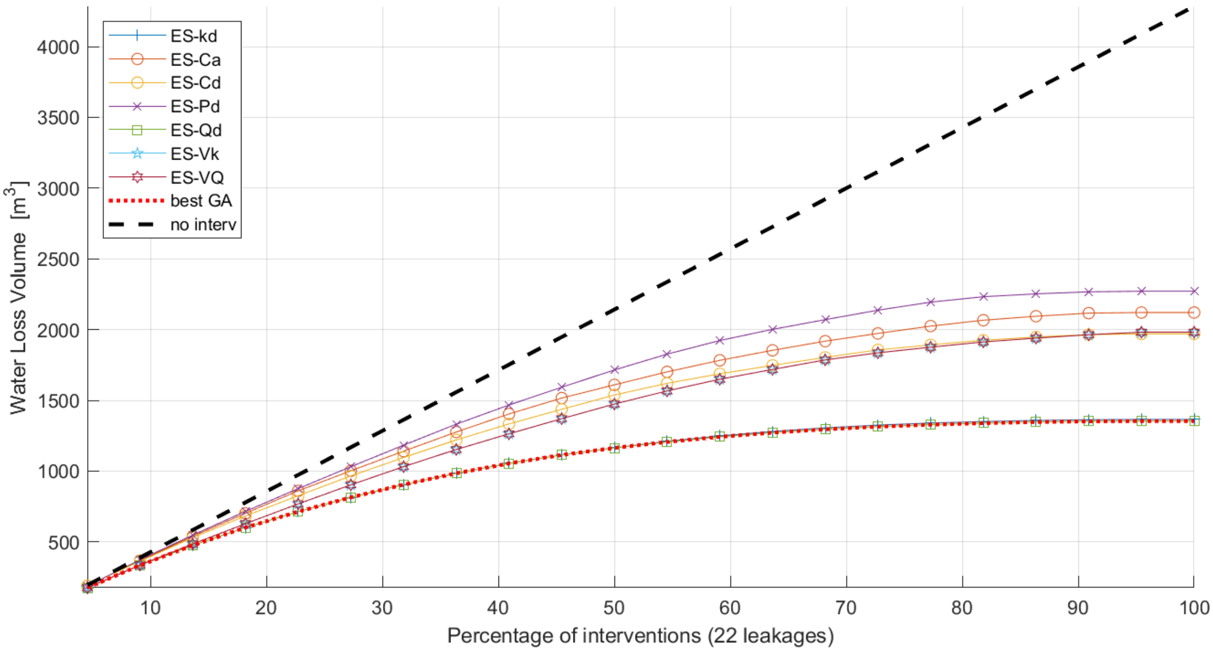

Figure 5.2: Comparison of ESs and GA solutions on the Araujo network.

In terms of water loss volume, the best ES is ES-Qd, i.e., the schedule ordered by descending CNS leakage flow, with a fitness value of 1354.06 m^3 followed closely by ES-kd, i.e., the schedule ordered by descending leakage magnitude, with a fitness value of 1365.4 m^3. The Table 5.1 also shows the percentage improvement with respect to the worst solution (ES-Pd): it can be seen that these schedules reduce the fitness value down to 40.4% and 39.9% respectively.

Schedule name	**WLV [m^3]**	**Improvement from worst [%]**	**Comparison of best [%]**
ES-Qd	1 354.06	-40.4	0.0
best GA	1 354.06	-40.4	0.0
ES-kd	1 365.42	-39.9	-0.8
ES-Cd	1 969.87	-13.3	-31.3
ES-Vk	1 981.96	-12.7	-31.7
ES-VQ	1 981.96	-12.7	-31.7
ES-Ca	2 120.88	-6.6	-36.2
ES-Pd	2 271.37	0.0	-40.4

Table 5.1: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

The evolution of the GA is depicted in figure 5.3.

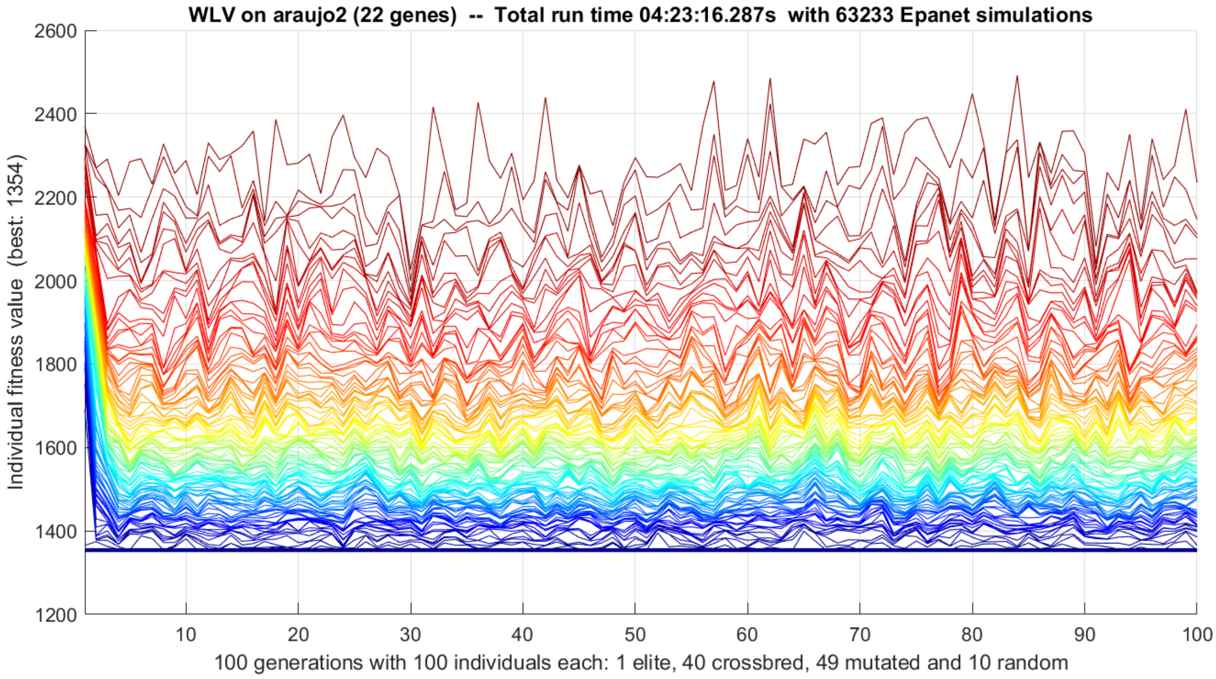

Figure 5.3: Evolution of the GA with WLV fitness function on the Araujo network.

It can be noticed that the lowest line is perfectly horizontal: the GA did not evolve from the given initial population and it was not able to find a better solution in 100 generations in a total runtime of 4.4 hours.

As a matter of fact, in Figure 5.2, the GA solution is overlapped with ES-Qd. Hence, the optimal solution to minimize the water volume loss during the whole intervention period is the schedule consisting in repairing leakages from highest to smallest flow during the CNS. The optimal schedule has a fitness value of 1354.06 m^3 of water volume leaked and a maximum percentage reduction of 40.4% with respect to ES-Pd.

This schedule allows to do just one hydraulic simulation with a computational time of few seconds if compared to the GA runtime. But, in general, the network state could change drastically after an intervention, leading to different pressures, different leakage flows and different water loss volumes.
However, such drastic changes in network state do not occur in the network under study. This could be due to the simplicity of the Araujo network, so it could actually be the optimal solution.

5.3. WLV optimization on Balerma network

In the attempt to find a more complex WDN than Araujo, the Balerma network is considered [20]. The comparison between the different ESs is shown in Figure 5.4.

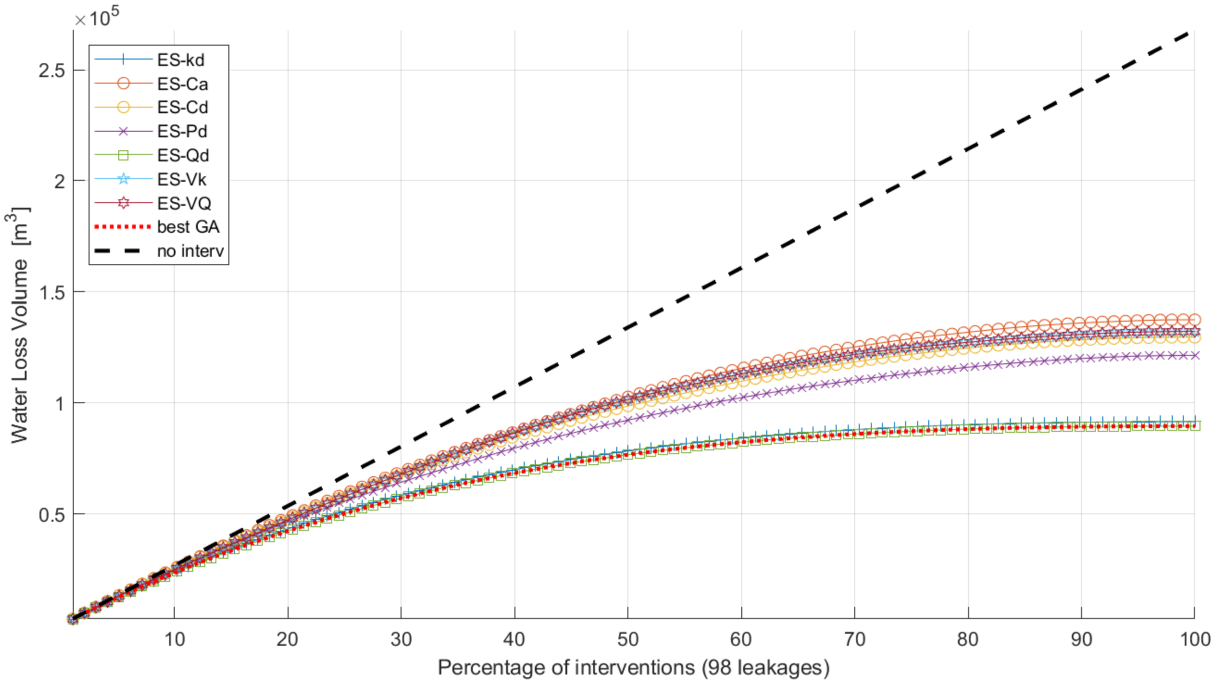

Figure 5.4: Comparison of ESs and GA solutions on the Balerma network.

In terms of water loss volume, the best ES is ES-Qd, i.e., the schedule ordered by descending CNS leakage flow, with a fitness value of 89 507.7 m^3 followed closely by ES-kd, i.e., the schedule ordered by descending leakage magnitude, with a fitness value of 91 536.7 m^3. The Table 5.2 also shows the percentage improvement with respect to the worst solution (ES-Ca): it can be seen that these schedules reduce the fitness value down to 34.9% and 33.4% respectively.

Schedule name	WLV [m^3]	Improvement from worst [%]	Comparison of best [%]
best GA	89 507.37	-34.9	0.0
ES-Qd	89 507.66	-34.9	0.0
ES-kd	91 536.72	-33.4	-2.2
ES-Pd	121 333.94	-11.7	-26.2
ES-Cd	129 526.66	-5.7	-30.9
ES-Vk	132 133.28	-3.8	-32.3
ES-VQ	132 133.28	-3.8	-32.3
ES-Ca	137 408.26	0.0	-34.9

Table 5.2: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

The evolution of the GA is depicted in Figure 5.5.

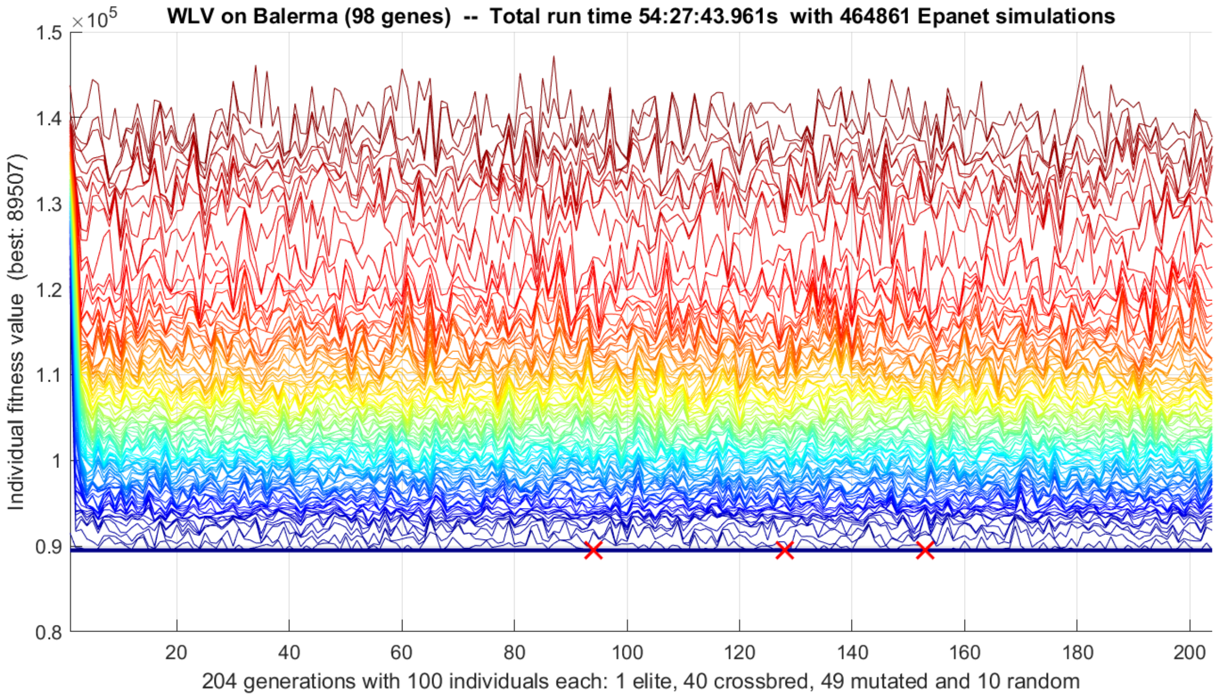

Figure 5.5: Evolution of the GA with WLV fitness function on the Balerma network

It can be noticed that the GA did run for 204 generations in a total runtime of 54.5 hours: it is because the algorithm found several better solution (the red crosses) and kept going until the stopping criteria were met. However, the lowest line seems perfectly horizontal: the solutions found had a slightly improvement of the fitness value but it was not so remarkable. The last optimal solution had a fitness value of 89 507.4 m^3 which was a reduction -0.0003% with respect to ES-Qd. Unfortunately, this little improvement is not enough to justify the huge computational cost, but it confirmed ES-Qd as the optimal solution.

5.4. WLV optimization on MOD network

A third network similar to Balerma, the MOD network, is considered [7]. The comparison between the different ESs is shown in Figure 5.6.

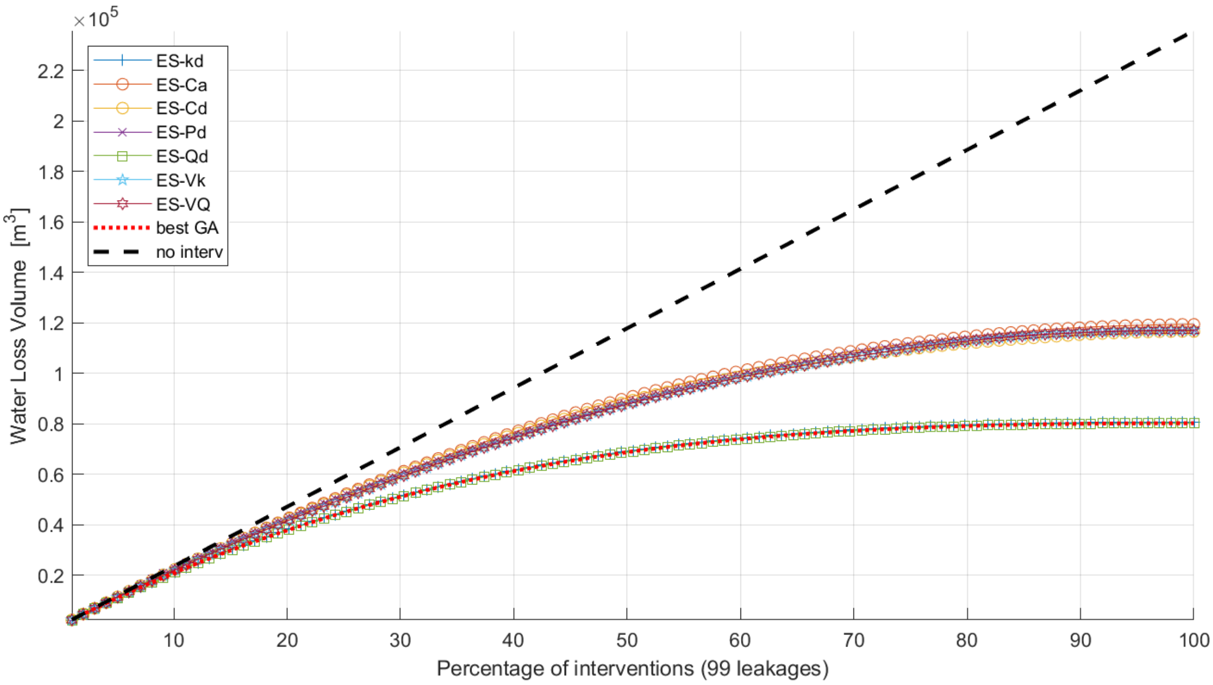

Figure 5.6: Comparison of ESs and GA solutions on the MOD network.

In terms of water loss volume, the best ES is ES-Qd, i.e., the schedule ordered by descending CNS leakage flow, with a fitness value of 80 237.0 m^3 followed closely by ES-kd, i.e., the schedule ordered by descending leakage magnitude, with a fitness value of 80 655.7 m^3. The Table 5.3 also shows the percentage improvement with respect to the worst solution (ES-Ca): it can be seen that these schedules reduce the fitness value down to 32.8% and 32.4% respectively.

Schedule name	WLV [m^3]	Improvement from worst [%]	Comparison of best [%]
ES-Qd	80 236.99	-32.8	0.0
best GA	80 236.99	-32.8	0.0
ES-kd	80 655.74	-32.4	-0.5
ES-Cd	116 609.56	-2.3	-31.2
ES-Pd	116 942.40	-2.0	-31.4
ES-Vk	117 037.40	-2.0	-31.4
ES-VQ	117 037.40	-2.0	-31.4
ES-Ca	119 381.18	0.0	-32.8

Table 5.3: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

The evolution of the GA is depicted in Figure 5.7.

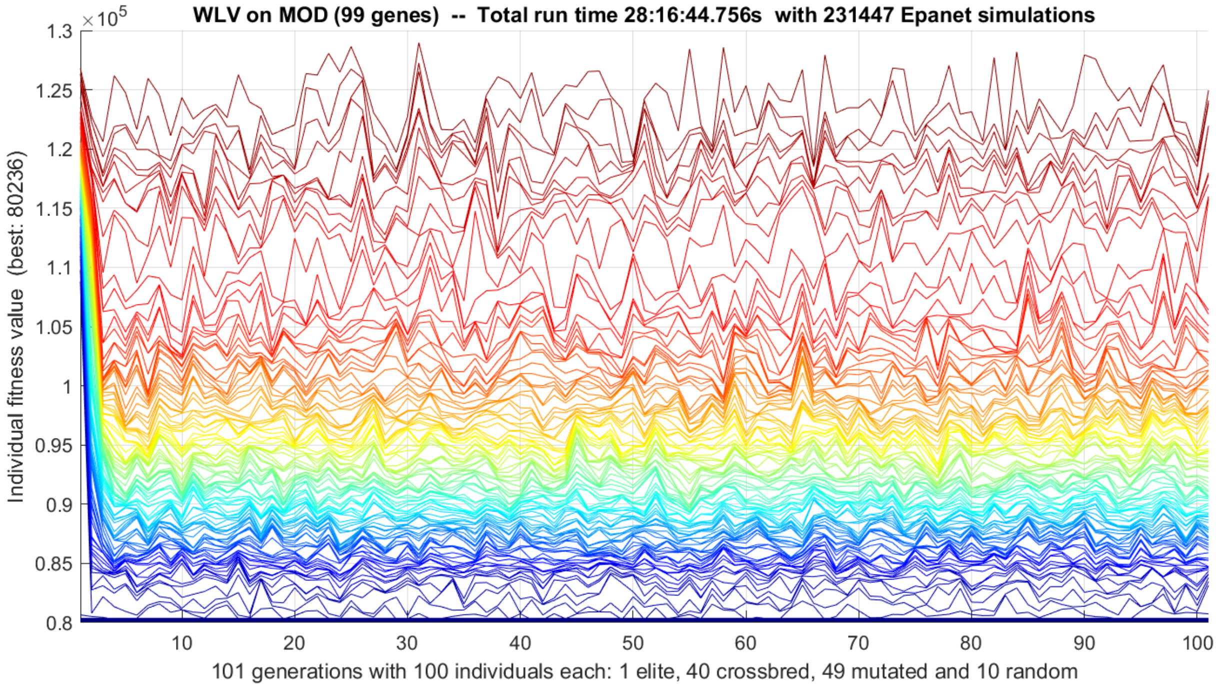

Figure 5.7: Evolution of the GA with WLV fitness function on the MOD network.

As in the case of the Araujo network, it can be noticed that the lowest line is perfectly horizontal: the GA did not evolve from the given initial population and it was not able to find a better solution in 101 generations in a total runtime of 28.3 hours.

As a matter of fact, in Figure 5.6, the GA solution is overlapped with ES-Qd. Hence, the optimal solution to minimize the water volume loss during the whole intervention period is the schedule consisting in repairing leakages from highest to smallest flow during the CNS. The optimal schedule has a fitness value of 80 237.0 m^3 of water volume leaked and a maximum percentage reduction of 32.8%.

5.5. WLV optimization on BIWS network

To be thorough, the GA is tried on the BIWS network. However, as explained in chapter 2.2, the BIWS is a large network affected by a very large number of leakages in comparison to the other networks (30 to 160 times the number of leakages of the other networks): the Figure 5.8 shows a run of the GA of just barely 2 generations that took up to 61 hours.

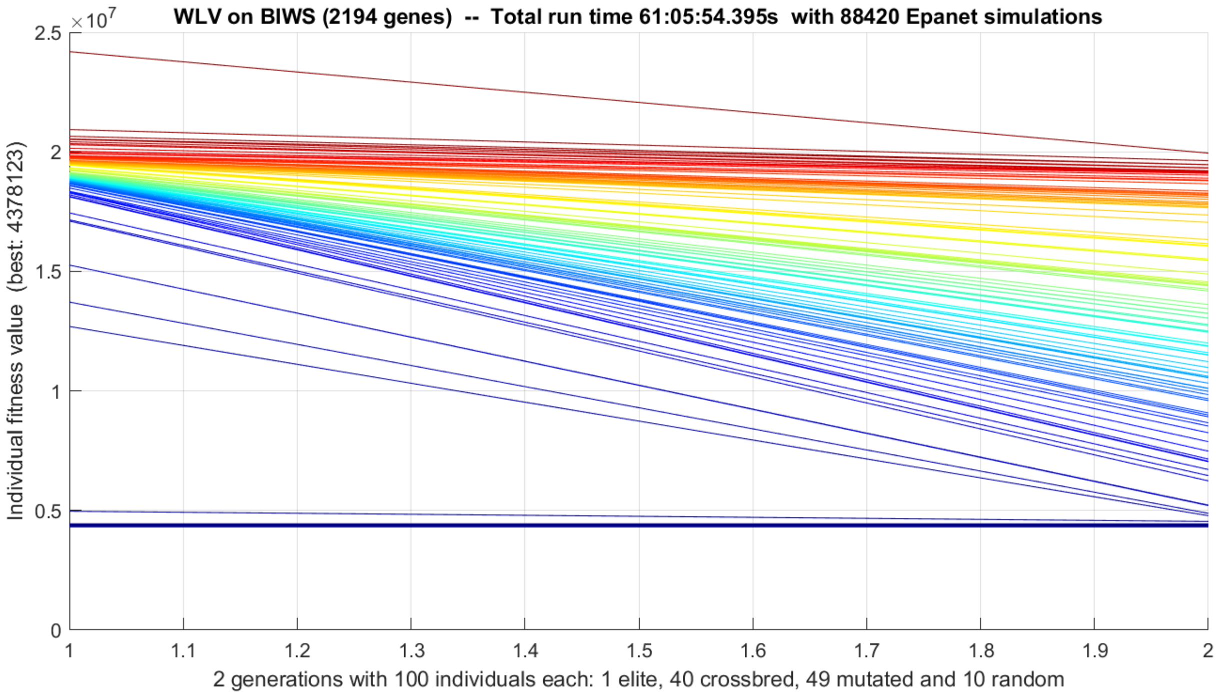

Figure 5.8: Evolution of the GA with WLV fitness function on the BIWS network.

A quick estimate of the expected total runtime is possible: first consider that when the successive populations are created, some of the fitness values are automatically assigned to the next generation, so the first population is the most computationally intensive. For this reason, it is observed in the other networks runs that the average runtime per generation is two-thirds of the first one. Therefore, an average runtime of 25 hours per generation can be assumed: 250 hours would be required for a total of 100 generations, i.e., more than 10 days of run without interruption. Such a long runtime is not appropriate for the purpose of the work and would not lead to relevant additional results.

Nonetheless, the comparison between the ESs solutions on the BIWS network can be observed in figure 5.9) and table 5.4.

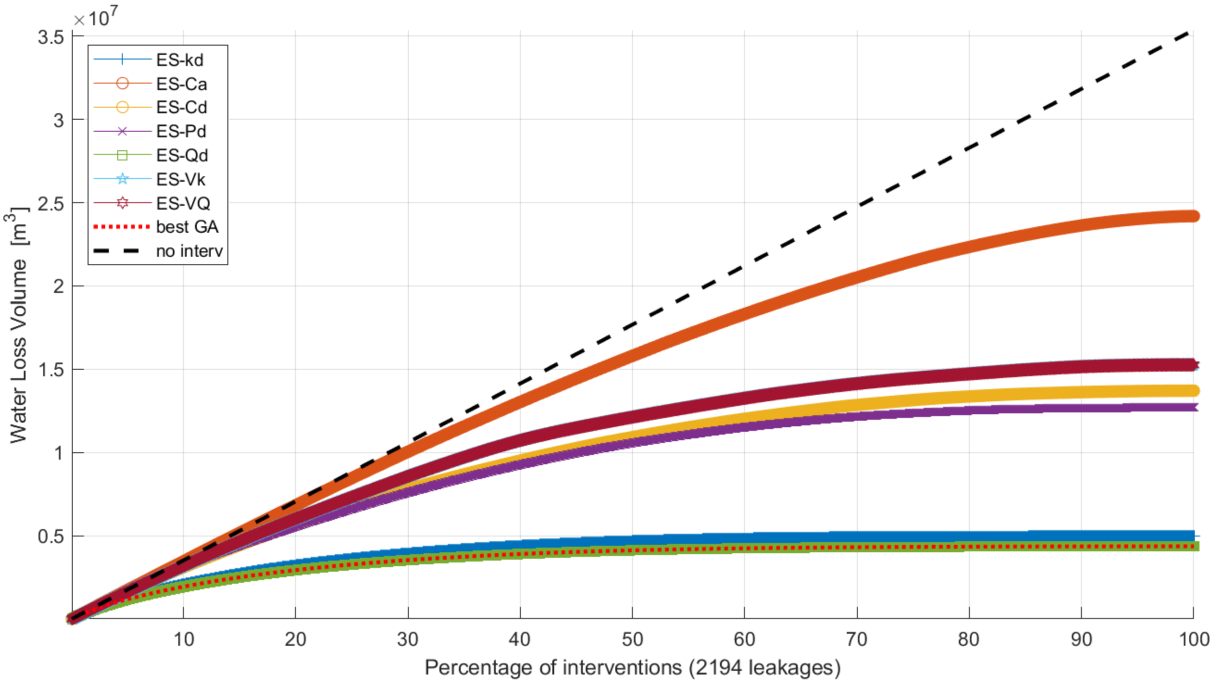

Figure 5.9: Comparison of ESs solutions on the BIWS network.

Schedule name	WLV [m^3]	Improvement from worst [%]	Comparison of best [%]
ES-Qd	4 378 123.50	-81.9	0.0
ES-kd	4 967 768.50	-79.5	-11.9
ES-Pd	12 694 875.00	-47.5	-65.5
ES-Cd	13 713 743.00	-43.3	-68.1
ES-Vk	15 262 087.00	-36.9	-71.3
ES-VQ	15 262 087.00	-36.9	-71.3
ES-Ca	24 187 384.00	0.0	-81.9

Table 5.4: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

In terms of water loss volume, the best ES is ES-Qd, i.e., the schedule ordered by descending CNS leakage flow, with a fitness value of 4 378 123.5 m^3 followed closely by ES-kd, i.e., the schedule ordered by descending leakage magnitude, with a fitness value of 4 967 768.5

m^3. It can be noticed that the percentage improvements with respect to the worst solution (ES-Ca) are 81.9% and 79.5% respectively.

5.6. Influence analysis of the intervention time

In order to assess the influence of $t_{intervention}$, another run of the GA has been studied. The intervention time has been set to one week instead of one day. The Araujo network is used to prove the irrelevance of the value of this parameter.

The evolution of the GA is depicted in figure 5.10, while the comparison between the different ESs is shown in figure 5.11 and in table 5.5.

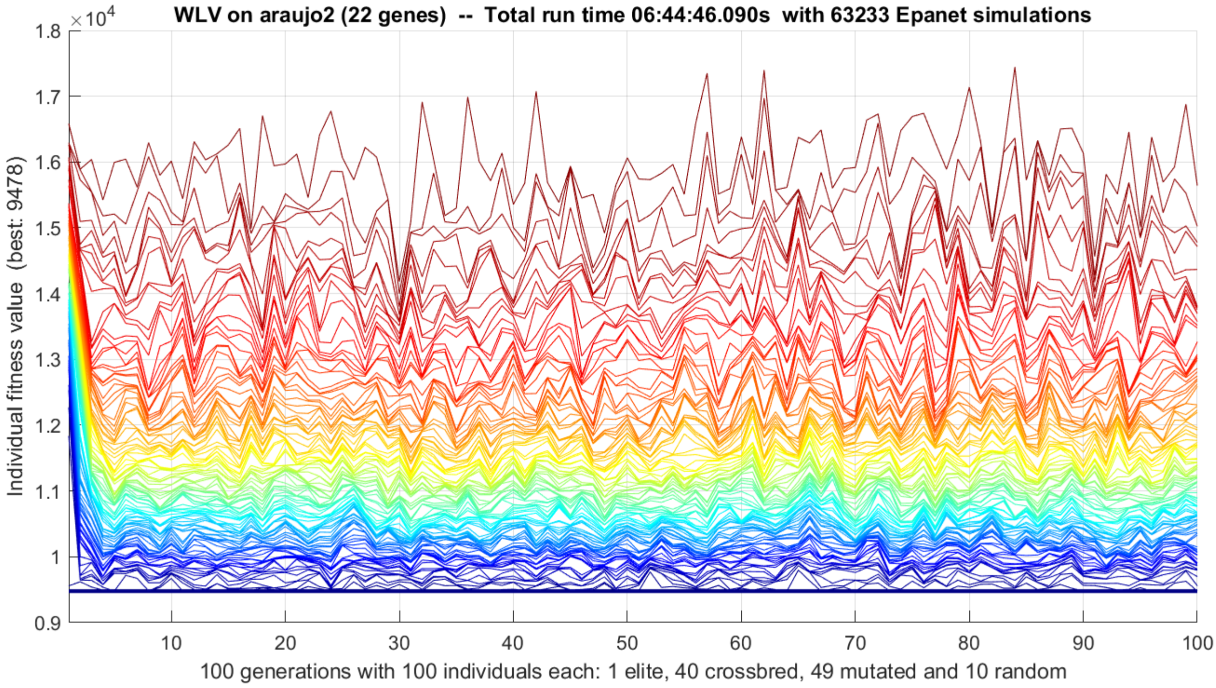

Figure 5.10: Evolution of the GA on the Araujo network with $t_{intervention} = 7d$

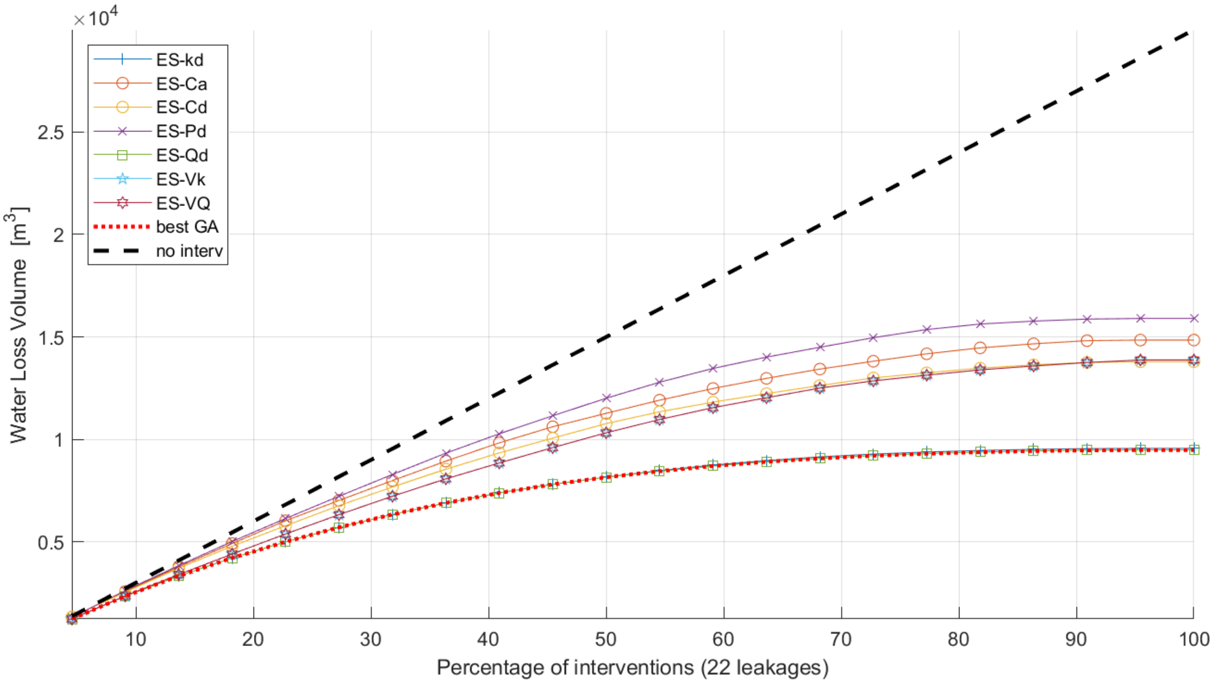

Figure 5.11: Comparison of ESs and GA solutions on the Araujo network with $t_{intervention} = 7d$

Schedule name	IMC [€]	Improvement from worst [%]	Comparison of best [%]
ES-Qd	9 478.43	-40.4	0.0
best GA	9 478.43	-40.4	0.0
ES-kd	9 557.95	-39.9	-0.8
ES-Cd	13 789.11	-13.3	-31.3
ES-Vk	13 873.72	-12.7	-31.7
ES-VQ	13 873.72	-12.7	-31.7
ES-Ca	14 846.15	-6.6	-36.2
ES-Pd	15 899.56	0.0	-40.4

Table 5.5: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

If respectively compared to figures 5.3 and 5.2 and to table 5.1, it can be noticed that every fitness values is simply multiplied by a factor of 7. In order to confirm it, the ratio between corresponding values in table 5.1 and table 5.5 is computed and the resulting value is a constant 7. The same could be done with each individual of the figure 5.3 and figure 5.10.

6 | Intervention and Management Costs (IMC) optimization

A more complex objective function is considered to take into account several issues a WDN manager is concerned with, most notably reducing the amount of the leaked water volume throughout the whole intervention period, the possibility to group very closely spaced leakages into a single pit, and the vicinity between successive interventions to minimize construction site transportation distances.
Consequently, a multi-objective (MO) function is needed. There are two general approaches to handle this type of function: one is to combine the different objective functions into a single composite function, and the other is to determine a Pareto-optimal solution set [14]. From an engineering point of view, it was considered appropriate for the problem to combine the singular functions expressing them as a cost.
The Intervention and Management Costs (IMC) fitness function is developed, that consider:

- the unsold water cost, directly tied to the WLV (chapter 5), which depends on the leakage magnitude and pressure evolution in the network;
- the repairing cost of the leakage itself, which is a-priori assessed and does not depend on the intervention order;
- the digging cost, which is function of the diameter of the pipe where the leakage is located;
- a vicinity cost which increases with the distance between subsequent interventions, it is the logistic cost associated to the relocation of the construction site.

6.1. IMC fitness function

The classical Weighted sum approach [14] is implemented in the GA to optimize the singular objective functions: a weight w_i is assigned to each objective function C_i so that the problem is converted to a single objective problem with a scalar objective function. Hence, the individual fitness $F_i(s)$ of the schedule s is calculated as follows:

$$\begin{aligned} F(s) = \sum_{l=1}^{N_{leaks}} F_l = \sum_{l=1}^{N_{leaks}} w_i \cdot C_i = \\ = \sum_{l=1}^{N_{leaks}} (w_{wlv} \cdot C_{wlv} + w_{int} \cdot C_{int} + w_{dig} \cdot C_{dig} + w_{vic} \cdot C_{vic})_l \end{aligned} \tag{6.1}$$

Water loss cost C_{wlv}

The water cost is associated to the economic loss of the unsold water thus, it is directly proportional to the leaked water volume calculated as in equation 5.1. The WLV is then multiplied by the selling price of the water c_{water} which is the charge for drinking water. The water price is set to be 1.35 €/m^3, which is the average charge for drinking water in Italy [3].

$$C_{wlv} = \sum_{l=1}^{N_{leaks}} c_{water} \cdot WLV_l \tag{6.2}$$

As with WLV, this cost is function of the leakages magnitude and the pressure at the nodes. It depends on the sequence of interventions.

Intervention cost C_{int}

The intervention cost is the actual amount to pay for the intervention operation. It is a-priori assessed through the pre-processing procedure explained in chapter 2.5 and for each leakage it can be either the cost to repair the leak or to replace the pipe.

$$C_{int} = \begin{cases} C_{repair}(D, k) & \text{if repairing intervention} \\ C_{replace}(D, L) & \text{if replacement intervention} \end{cases} \tag{6.3}$$

This cost is function of the diameter of the pipe and either the leakage magnitude or the length of the pipe replaced. It does not depend on the sequence of interventions.

Digging cost C_{dig}

This cost is introduced to incentivize solutions in which leakages that are very closely spaced are repaired in sequence. Namely, if two leaks to be repaired consecutively are very close to each other, there is a possibility to combine them in a single pit and reduce their individual excavation costs.

The digging cost is calculated by multiplying digging volume V_{dig} and digging price c_{dig}. The digging depth is set to $z_{dig} = 1.5$ meters, which is constant due to the lack of information concerning the ground altitude above the pipes; in addition, a square pit is assumed, and its width B_{dig} depends on the diameter D of the pipe to be revealed, as follows:

$$A_{dig} = {B_{dig}}^2 \qquad \text{where} \quad B_{dig} = \begin{cases} 0.60m & \text{if } D \leq 200mm \\ 1.00m & \text{if } D > 200mm \\ 1.25m & \text{if } D > 400mm \\ 1.50m & \text{if } D > 800mm \end{cases}$$

For each pair of consecutive leakages j and $k = j - 1$ during the schedule, it is checked if the relative digging areas do overlap: the distance L_{jk} between the consecutive leakages is computed and the pits overlap if L_{jk} is less than the sum of half the digging widths of the relative pits. In such an event, the area of the pit of the second leakage is reduced by the overlapping area between the pits as follows:

Algorithm 6.1 Overlapping condition of pits of two consecutive interventions

if $L_{jk} \geq B_{dig,j}/2 + B_{dig,k}/2$ **then**
  $A_{dig,j} = {B_{dig,j}}^2$
  $A_{dig,k} = {B_{dig,k}}^2$
else
  $A_{dig,j} = {B_{dig,j}}^2$
  $A_{dig,k} = {B_{dig,k}}^2 - A_{overlap}$
end if

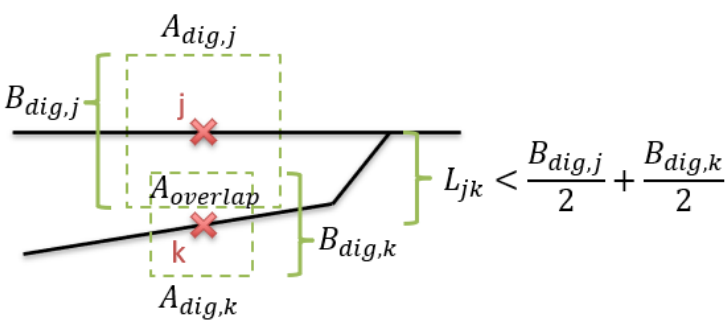

Figure 6.1: Visual representation of the excavation pits superimposition

A visual representation of the pits superimposition is depicted in figure 6.1, where:

$$A_{overlap} = [B_{dig,j}/2 + B_{dig,k}/2 - L_{jk}] \cdot \min\{B_{dig,j}; B_{dig,k}\}$$

For the digging price, reference has been made to the regional price list of the public works of Lombardy region regarding the excavation costs made by hand until a depth of 2 meters. The price for this operations is 150 €/m^3.

$$C_{dig} = \sum_{l=1}^{N_{leaks}} c_{dig} \cdot A_{dig,j} \cdot z_{dig} \tag{6.4}$$

Altogether, this cost is a function of the pipe diameter and it depends on the sequence of interventions.

Vicinity cost C_{vic}

This cost is introduced to encourage solutions which repair leakages by zone and to reduce path between interventions. This is meant to consider the logistic cost associated to the relocation of the construction site. Namely, the further the consecutive interventions are, the larger the vicinity cost is; though, an upper limit is set for the distance, because above a certain distance threshold L_{th} the movement of equipment and vehicles should not influence the solution anymore. The behaviour of this function is depicted in figure 6.2.

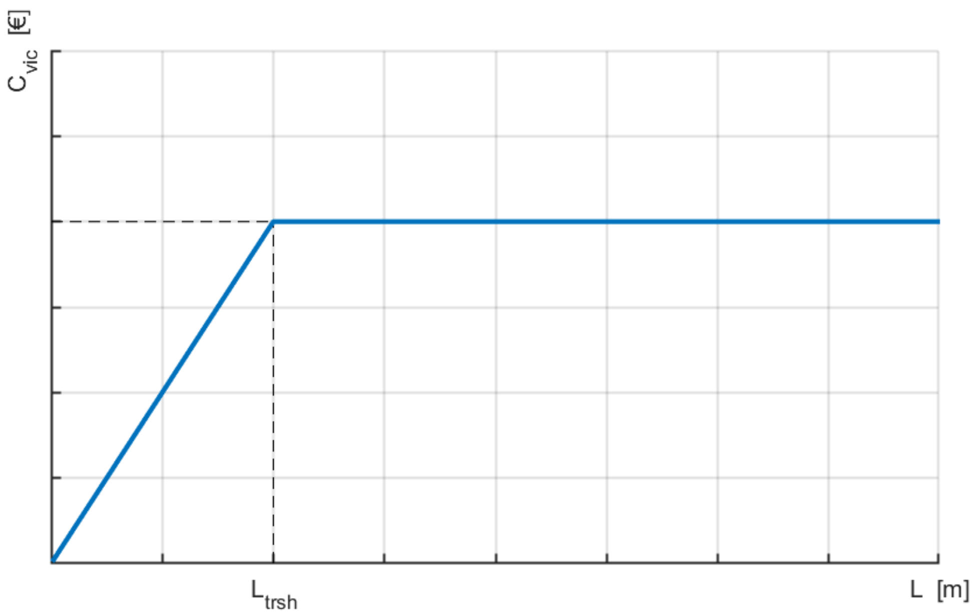

Figure 6.2: Vicinity cost function with respect to the distance of subsequent interventions

For each pair of consecutive leakages within the schedule, the distance L_l from the intervention l to the previous one is computed. The vicinity cost is calculated as follow:

$$C_{vic} = \sum_{l=1}^{N_{leaks}} c_{vic} \cdot \min\{L_l; L_{th}\} \tag{6.5}$$

Regarding the determination of the parameters c_{vic} and L_{th}, it was not trivial. The vicinity cost function should represent the cost of moving equipment, vehicles, and supplies from one intervention to the next: it is particularly difficult to find anything in this direction in the literature or to quantitatively assess its value.
For this reason, the parameter c_{vic} is left as a unit cost per length of construction site relocation, so that C_{vic} and its graphs also directly represent the path length between interventions.
Meanwhile, lower distance thresholds were found to have a much greater impact on the influence of the vicinity cost, especially the larger the network is: larger networks have much longer travel distances among interventions and they are cut off by L_{th}. A good value for this parameter is found around a tenth of the distance between the furthest leakages. But, for solidity, as a first attempt for the study, the threshold is set to $L_{th} = 600m$ for all the networks. The definition of these parameters is of interest for the utility and each management can set their own reference distance.

Thus, the two parameters L_{th} and c_{vic} could be set to a more appropriate value by the WDN manager who will use the algorithm.

All over, the vicinity cost depends only on the sequence of interventions.

Weights of the different components

The total fitness fuction is then computed as the weighted sum of its components as indicated in equation 6.1. The weights of the components could be tuned by the WDN manager who wants to rehabilitate the network in order to customize its priorities during the interventions.
Anyway, for the purpose of the work a set of weights has been assigned as follows:

$$w_{wlv} = 1$$

$$w_{int} = 1$$

$$w_{dig} = 1$$

$$w_{vic} = 3$$

where the vicinity cost has not unit weight. The vicinity weight is calibrated through a sensitivity analysis to ensure that the solution is affected by this fitness component and to compensate the lack of knowledge of the c_{vic}: it resulted in $w_{vic} = 3$.

6.2. Explanation of IMC results graphs

In addition to the results explained in chapter 4.3 and chapter 5.1, other charts are proposed in this chapter for the advanced IMC fitness function.

The figure 5.1 about the comparison of the different solutions is equivalent to the graphs in the typology of figure 6.4: the upper smaller four charts illustrate the different components of the IMC fitness function as in equations 6.2 to 6.5, while the last graph presents the overall IMC fitness function evolution of the network during the implementation of each intervention schedule. In all the graphs in this figure, the lines are the cumulative fitness values after each intervention. The final point represents the fitness of the individual. The red dotted line represents the optimal GA solution.

6.3. IMC optimization on Araujo network

The evolution of GA for this network is shown in figure 6.3. It ran for 301 generations and a total runtime of 22.7 hours until it was manually terminated due to time reasons. However, it could be leaved running but for the purpose of the thesis, the following comments would not be different.

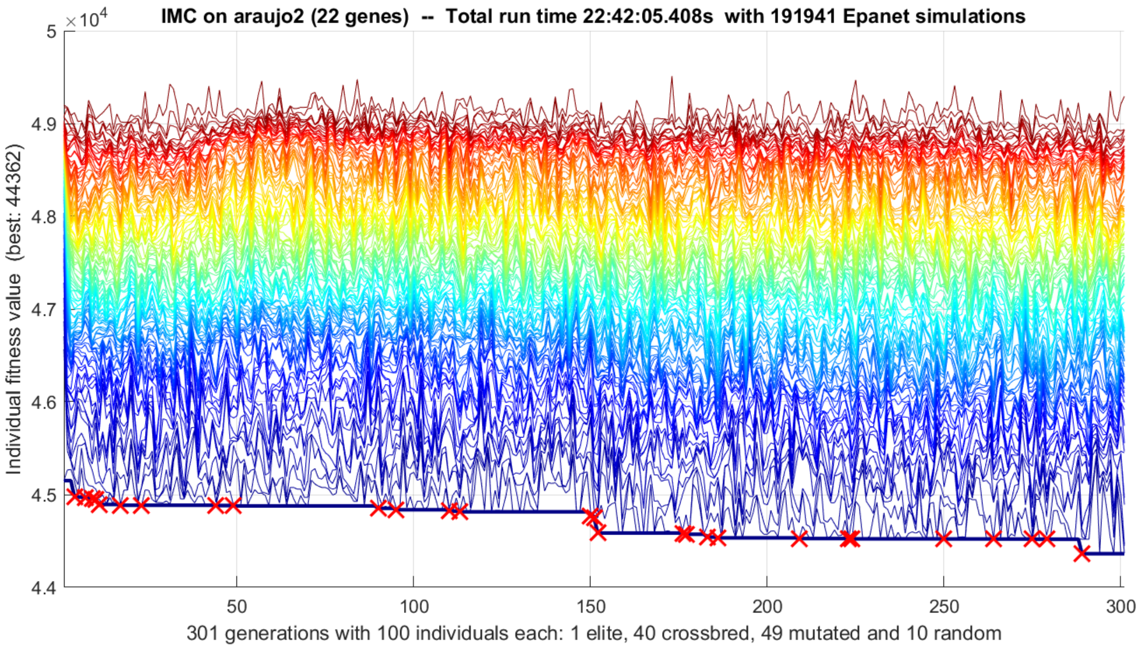

Figure 6.3: Evolution of the GA with IMC fitness function on the Araujo network.

As it can be noticed, the algorithm achieved several improved solutions (the red crosses) starting from ESs. The graphs in figure 6.4 depict the components of the optimal solution along with the different ESs: however, it can be seen that the GA solution does not have a very large improvement of the overall fitness function (last graph).

Moreover, the table 6.1 shows the comparison of fitness values among the different schedules: it can be seen that the GA solution is very close to schedules ES-Vk and ES-VQ (worse than GA of 1.7%), while the worst solutions for this network are ES-kd and ES-Qd.

Schedule name	**IMC [€]**	**Improvement from worst [%]**	**Comparison of best [%]**
best GA	44 362.95	-7.4	0.0
ES-Vk	45 150.21	-5.8	-1.7
ES-VQ	45 150.21	-5.8	-1.7
ES-Pd	46 707.63	-2.6	-5.0
ES-Cd	47 576.19	-0.7	-6.8
ES-Ca	47 783.07	-0.3	-7.2
ES-Qd	47 916.65	0.0	-7.4
ES-kd	47 932.21	0.0	-7.4

Table 6.1: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

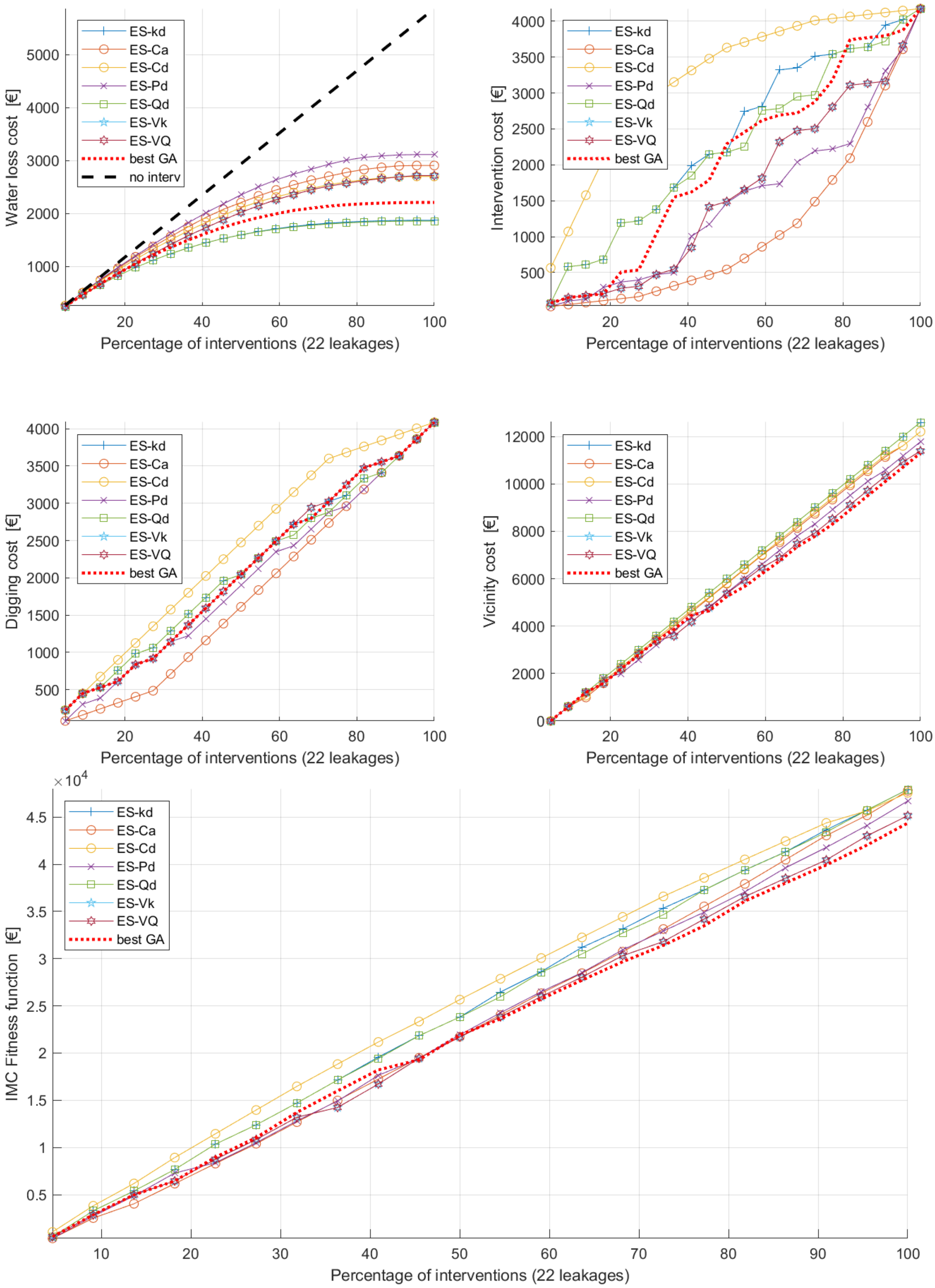

Figure 6.4: Fitness function comparison of ESs and GA on the Araujo network.

6.4. IMC optimization on Balerma network

The evolution of GA for this network is shown in figure 6.5. It ran for 202 generations and a total runtime of 54.5 hours until a plateau of 50 generations.

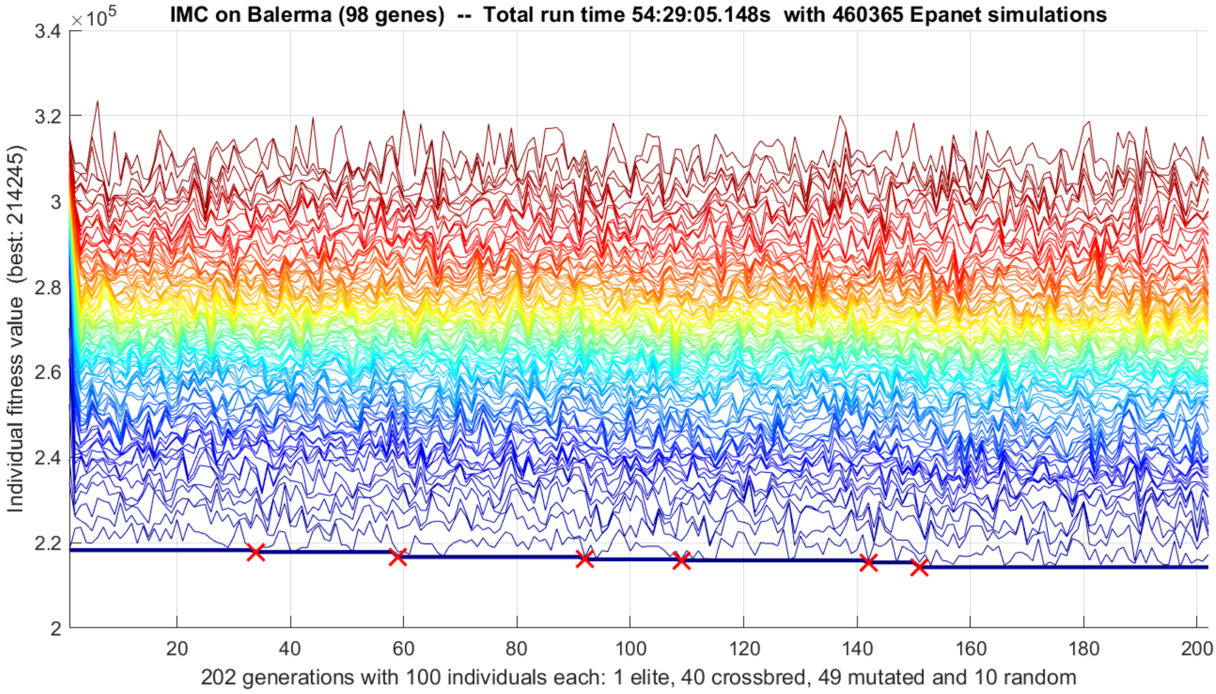

Figure 6.5: Evolution of the GA with IMC fitness function on the Balerma network.

As it can be noticed, the algorithm achieved several improved solutions (the red crosses) starting from ESs. The graphs in figure 6.6 depict the components of the optimal solution along with the different ESs: however, as for the case of Araujo network, it can be seen that the GA solution does not have a very large improvement of the overall fitness function (last graph).

Moreover, the table 6.2 shows the comparison of fitness values among the different schedules: it can be seen that the GA solution is very close to schedules ES-Vk and ES-VQ (worse than GA of 1.8%), while the worst solutions for this network are ES-Pd and ES-Ca.

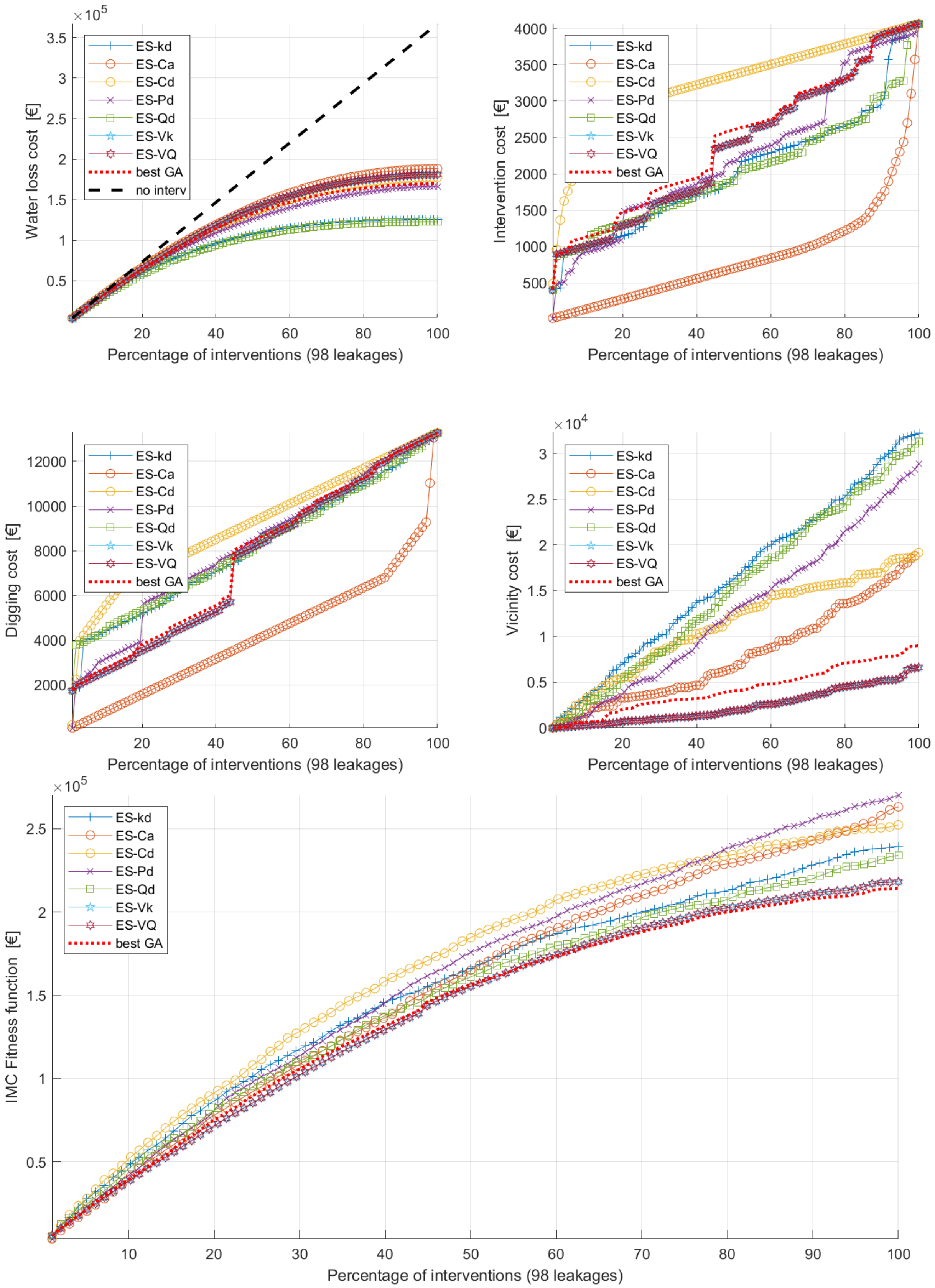

Figure 6.6: Fitness function comparison of ESs and GA on the Balerma network.

Schedule name	IMC [€]	Improvement from worst [%]	Comparison of best [%]
best GA	214 245.51	-20.7	0.0
ES-Vk	218 240.85	-19.2	-1.8
ES-VQ	218 240.85	-19.2	-1.8
ES-Qd	233 985.96	-13.4	-8.4
ES-kd	239 558.17	-11.4	-10.6
ES-Cd	252 328.82	-6.6	-15.1
ES-Ca	263 126.61	-2.6	-18.6
ES-Pd	270 252.51	0.0	-20.7

Table 6.2: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

6.5. IMC optimization on MOD network

The evolution of GA for this network is shown in figure 6.7. It ran for 146 generations and a total runtime of 38.7 hours until it reached the stopping criteria.

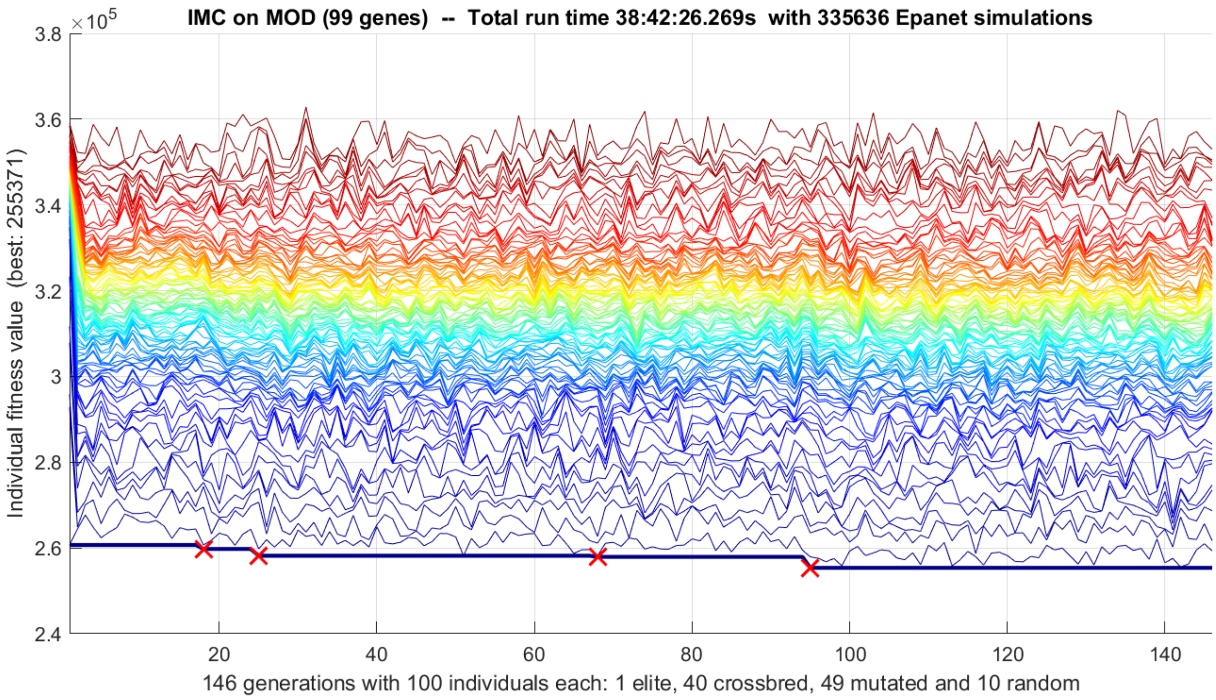

Figure 6.7: Evolution of the GA with IMC fitness function on the MOD network.

As it can be noticed, the algorithm achieved several improved solutions (the red crosses) starting from ESs. The graphs in figure 6.8 depict the components of the optimal solution along with the different ESs: however, it can be seen that the GA solution does not have a very large improvement of the overall fitness function (last graph).

Moreover, the table 6.3 shows the comparison of fitness values among the different schedules: it can be seen that the GA solution is very close to schedules ES-Vk and ES-VQ (worse than GA of 2.0%), while the worst solutions for this network are ES-Pd and ES-Ca.

Schedule name	**IMC [€]**	**Improvement from worst [%]**	**Comparison of best [%]**
best GA	255 371.68	-21.0	0.0
ES-Vk	260 695.22	-19.4	-2.0
ES-VQ	260 695.22	-19.4	-2.0
ES-Qd	292 784.84	-9.4	-12.8
ES-kd	295 815.07	-8.5	-13.7
ES-Cd	307 869.32	-4.8	-17.1
ES-Ca	311 666.44	-3.6	-18.1
ES-Pd	323 260.36	0.0	-21.0

Table 6.3: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

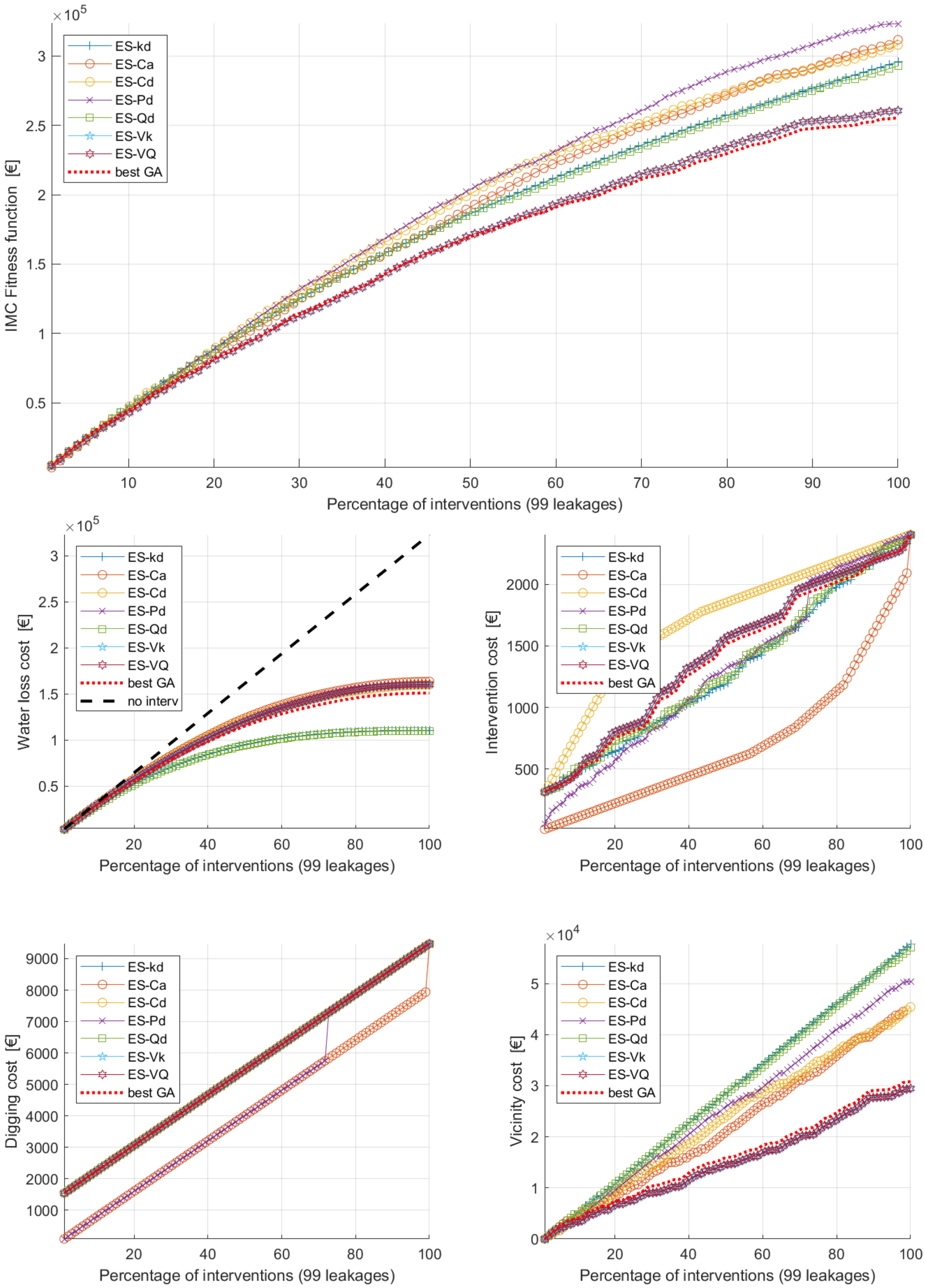

Figure 6.8: Fitness function comparison of ESs and GA on the MOD network.

6.6. IMC optimization on BIWS network

As for the previous fitness function, due to time constraints only the behaviour of ES in the BIWS network will be investigated. A runtime of 2.0 hours was needed to run just the seven ESs.
The graphs in figure 6.9 depict the components of the different ESs: the optimal solution is the ES-Qd, closely followed by ES-kd.

Moreover, the table 6.4 shows the comparison of fitness values among the different schedules: it can be seen that the best solution (ES-Qd) performs better than the worst (ES-Ca) with less than half of the fitness value (-53.3%).

Schedule name	IMC [€]	Improvement from worst [%]	Comparison of best [%]
ES-Qd	22 980 919.48	-53.3	0.0
ES-kd	23 772 971.65	-51.7	-3.3
ES-Pd	34 090 833.18	-30.7	-32.6
ES-Vk	34 624 204.95	-29.7	-33.6
ES-VQ	34 624 204.95	-29.7	-33.6
ES-Cd	34 876 137.26	-29.1	-34.1
ES-Ca	49 225 024.92	0.0	-53.3

Table 6.4: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

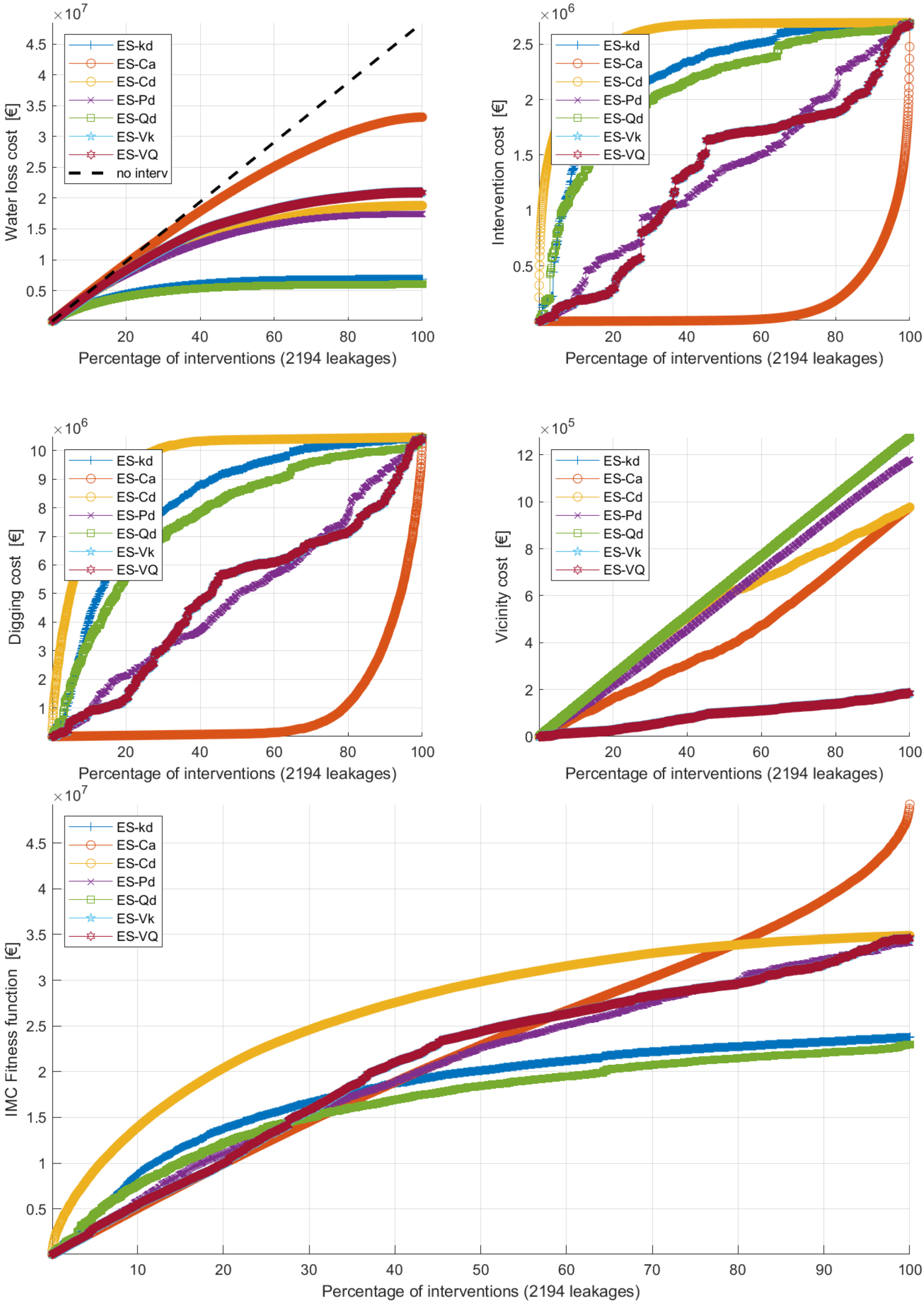

Figure 6.9: Fitness function comparison of ESs and GA on the BIWS network.

6.7. Threshold distance study

Several experiments have been performed to test the algorithm and this particular fitness function. The parameters have been changed several times to perform a sensitivity analysis.

An example is presented below to find a better threshold distance L_{th} for the vicinity cost component. The threshold in this run has been set to $L_{th} = 200m$ instead of $600m$. The evolution of GA for this run is shown in figure 6.10. It ran for 200 generations and a total runtime of 72.5 hours until it was manually terminated due to time constraint. However, the run in figure 6.5 also has the same amount of generations.

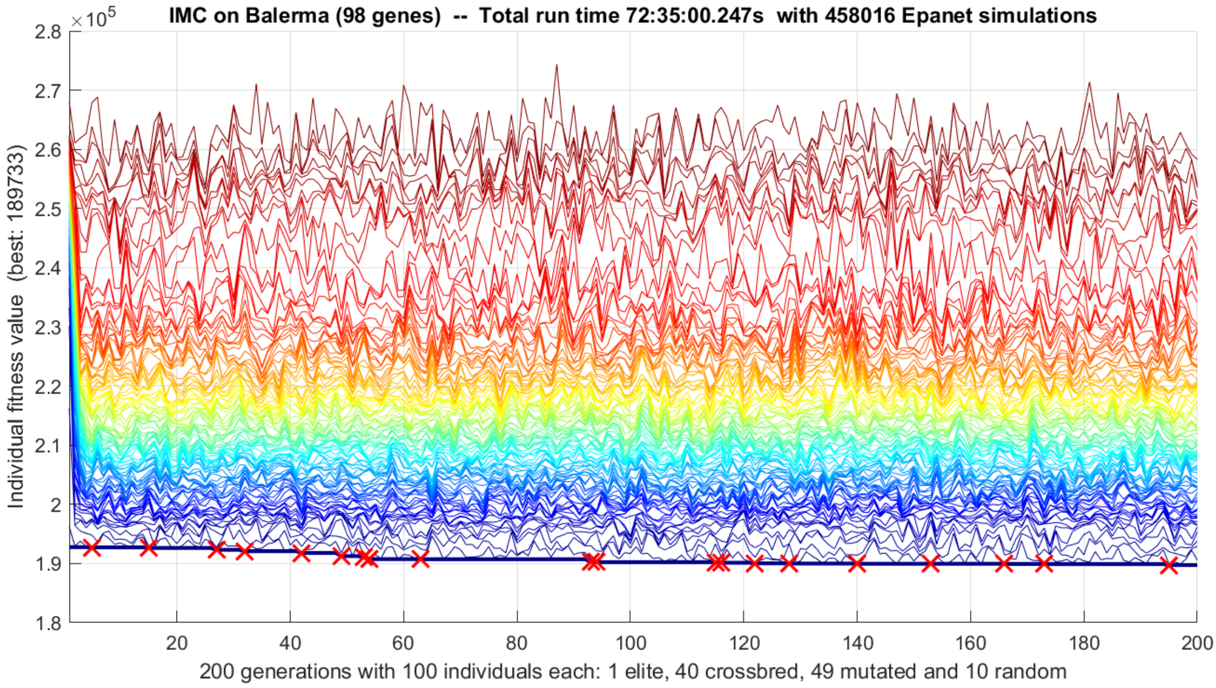

Figure 6.10: Evolution of the GA with on the Balerma network with $L_{th} = 200m$

As it can be noticed, the algorithm achieved several improved solutions (the red crosses) starting from ESs. They are much more in number and in frequency than figure 6.5. However, as for the case of $L_{th} = 600m$, it can be seen that the GA solution does not have a very large improvement of the overall fitness function. The graphs in figure 6.11 depict the components of the optimal solution along with the different ESs.

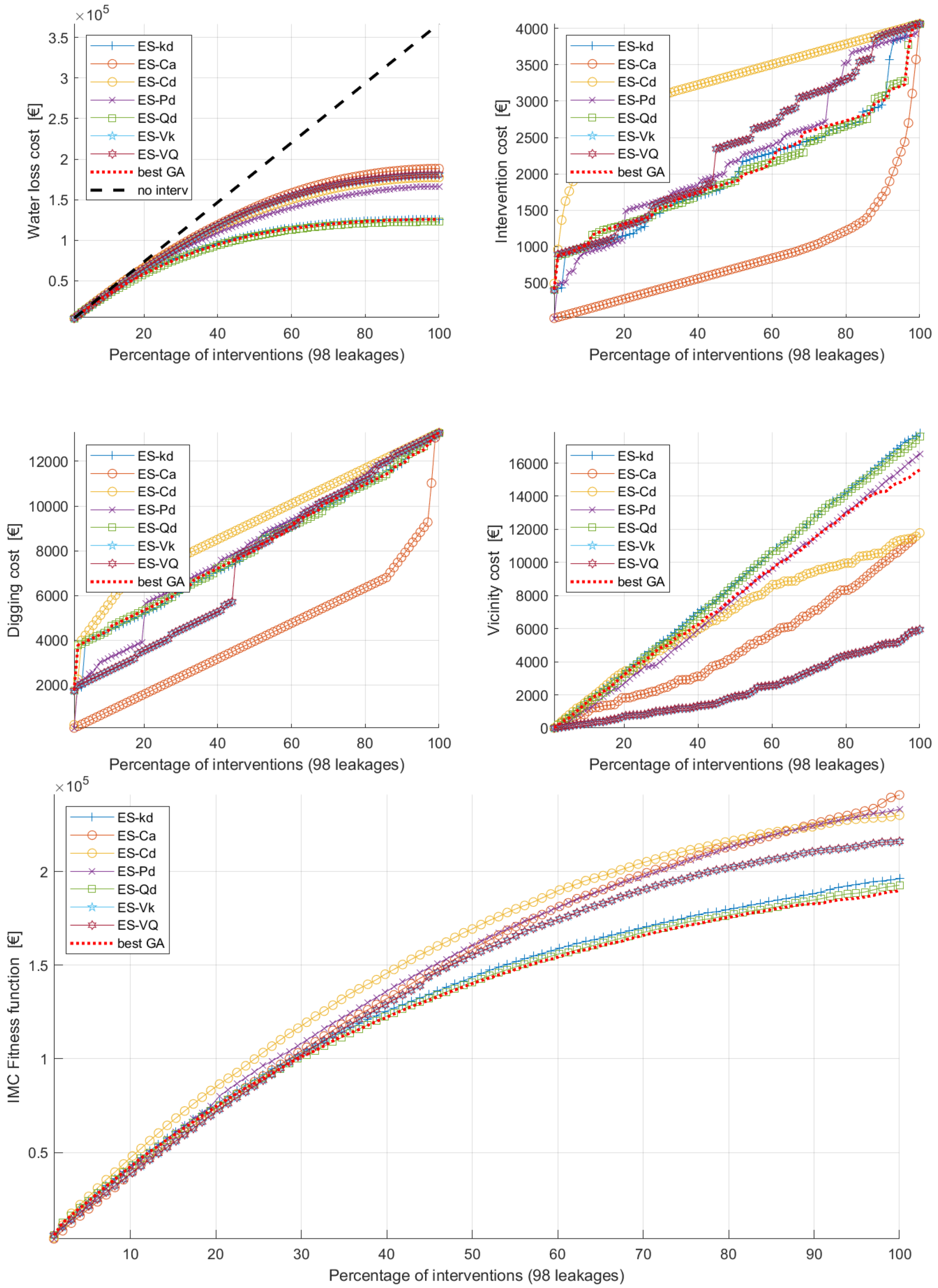

Figure 6.11: Comparison of ESs and GA solutions on the Balerma network with $L_{th} = 200m$

Moreover, the table 6.5 shows the comparison of fitness values among the different schedules: it can be seen that the GA solution is very close to schedules ES-Qd (worse than GA of 1.6%). This behaviour is different from the previous case where the best ES was ES-VQ: decreasing the limit beyond which the distances are truncated, the vicinity cost loses influence and the WLV cost became more important.

Schedule name	**IMC [€]**	**Improvement from worst [%]**	**Comparison of best [%]**
best GA	189 733.24	-21.3	0.0
ES-Qd	192 767.30	-20.0	-1.6
ES-kd	196 258.76	-18.5	-3.3
ES-Vk	216 212.80	-10.3	-12.2
ES-VQ	216 212.80	-10.3	-12.2
ES-Cd	230 140.58	-4.5	-17.6
ES-Pd	233 274.16	-3.2	-18.7
ES-Ca	240 938.37	0.0	-21.3

Table 6.5: Schedules results: fitness values, improvements from the worst schedule, and comparisons of the optimal solution with respect to other ESs

7 | Results and discussion

In the following sections, different aspects of the study are analyzed and compared. The main aspects that are going to be enlighten are:

- the performance of the intervention schedules,
- the results of the two fitness functions,
- the role of the diverse sizes and properties of the networks,
- the behaviour of the genetic algorithm.

7.1. Performance of schedules

In general, the costs ordered schedules (ES-Ca, ES-Cd) and the pressures ordered schedule (ES-Pd) resulted to be the worst schedules in every case.

The vicinity ordered schedules (ES-Vk, ES-VQ) better performed with the IMC fitness function because they optimize the vicinity cost.

The ES-Qd is always the better solution for the WLV fitness function: the water loss could be minimized just considering the leaking flow in the CNS. It was not proven that drastic changes to the pressures levels could happen during the intervention plan with the used networks and methodologies. This schedules allows the prediction of an excellent solution to the problem with just one simulation of the Current Network Status (CNS).

Another great solution to the WLV optimization problem was provided by ES-kd which achieved results as great as ES-Qd even if it does not consider the pressure influence. Moreover, any simulation is required for this schedule assessment.

7.2. Results of fitness functions

7.2.1. WLV optimization problem

The WLV fitness function attempted to optimize the water loss through the leakages. Inspecting the charts of the type of figure 5.4, the same trend of the solutions can be noted in every network and in every schedule: they are lines with decreasing slope and eventually reach horizontal asymptotes. This is because of how they are created and what they represent: each one is composed by the accumulated values of the Water Loss Volumes after the fixing of the leakages in the scheduled order hence, when the whole network is eventually repaired, the WLV is null and the lines stop increasing.

Comparing the results of the schedules with the case when no interventions are computed (black dashed line in figures such as 5.4), it can be noticed that the volume of water that would be lost without intervention is reduced by one half or two third by the diverse solutions.

7.2.2. IMC optimization problem

The IMC fitness function was designed to evaluate all the costs associated to the management and rehabilitation of the network, considering four components: the unsold water loss, the intervention cost, the digging cost, and the vicinity cost. The different components were different objectives to be optimized and the IMC fitness resulted in a weighted multi-objective sum optimization function (chapter 6.1).

Inspecting the charts of the type of figure 6.6, some general considerations on the trend of the different schedules can be expressed even if IMC problem is not as regular as WLV. In these figures, the four upper smaller charts represent the different components of the IMC fitness function while the bigger graph depicts the overall IMC fitness function.

The first chart has identical trend of the WLV fitness charts because they are directly proportional.

The second and third graphs show that intervention and digging costs do not depend on intervention order since every schedule has the same final fitness value; the trend of the ESs in the latter graphs are the same because both costs depend on the diameter of the pipe where the leakage is located.

The forth graph shows the vicinity cost. It has not a regular shape and varies among networks and schedules.

In the latter bigger graph, the GA solution definitely ends below ESs as it is an optimized solution.

7.3. Role of network diversity

Clearly, it is observed that the network size and characteristics affect both the performance of the diverse schedules and the results of the different fitness functions.

It can be noticed from WLV fitness values in tables 5.1, 5.2, 5.3, and 5.4 that the water loss volume is strongly influenced by the number of leakages. For this reason, a network with a large number of leakages, such as BIWS network, has a large variability of the fitness values of the different solutions. Given this larger variability, the greater the number of leaks, the more important the choice of intervention order is.

The influence of the number of leakages upon the WLV behaviour has an effect also on the IMC fitness function: because of the water loss cost being a component of the IMC function, the greater the number of leaks, the greater influence of this component is in the overall fitness function with respect to the other components.

Therefore, in larger networks such as BIWS (please refer to figure 6.9), the most performing schedules are ES-Qd and ES-kd which, as stated before, are the schedules that optimize the WLV. Meanwhile in smaller networks such as Araujo (please refer to figure 6.4), the most performing schedules are ES-VQ and ES-Vk.

7.4. Behaviour of GA

Concerning the genetic algorithm proposed in this work, it is proven that it is able to find new better solutions whether the initial population is completely random or partially given (please refer to section 4.4).

It was found that one hundred individuals divided into 1 elite, 40 crossbred, 49 mutated, and 10 random individuals are reasonable numbers: more individuals lead to larger computational time, while less individuals do not to properly explore the search space of the problem.

Also the number of switches influence the evolution of the algorithm: initially, a fixed number of genes were switched in a individual mutation; eventually, it was found that a varying number of switches can help the genetic diversity of the new population.

However, not so large improvements in terms of fitness values were obtained by its solutions.

In the considered networks, the GA was not able to find a better solution for the WLV optimization problem, confirming that the water loss could be minimized just considering the leaking flow in the CNS, i.e. ES-Qd was the optimal schedule for that problem.

Whereas, with the IMC optimization problem, the GA have more space to evolve from ESs and better schedules were achieved. However, not so large improvements were obtained also with that fitness function.

Furthermore, it is noticed that the alteration of the parameters strongly influence the evolution capacity of the algorithm: if the parameters make one fitness component to be prevalent upon others, the GA does not evolve from the given initial population and it is not very successful in finding new better solutions.
An interesting note is remarked observing the study on the threshold distance (section 6.7) where this parameter has been changed. It is noted that lowering the threshold on the same network changed the fitness trend of ESs: ES-Qd gained the best placement in contrast to what happened in section 6.4 where ES-VQ was the best solution. It is remarked that decreasing the limit beyond which the distances are truncated, the vicinity cost loses influence and the WLV cost became more important.

8 | Conclusions and future developments

In this work, several intervention schedules are proposed to plan the interventions on a leaky water distribution network. They were tested on diverse networks and with diverse optimization problems. One of the aims of the thesis was to find the best intervention schedule among the proposed Ex-Ante Schedules and even a better one by developing a genetic algorithm able to find the optimal solution of the given optimization problem.

It can be stated that CNS leaking flow ordered schedule (ES-Qd) is the best schedule to minimize the Water Loss Volume. Moreover, the leakage magnitude ordered schedule (ES-kd) also achieves very good results, but it does not even require the hydraulic simulation of the CNS.

Whereas, concerning the IMC optimization problem, the fitness ranking of ESs varies with the network they are tested on: vicinity ordered schedules (ES-VQ, ES-Vk) better perform on smaller networks with fewer leakages, while CNS leaking flow ordered schedule (ES-Qd) and leakages magnitude ordered schedule (ES-kd) better perform on larger networks with more leakages.

Indeed, the networks diversity clearly affects the behaviour of either the different schedules and the two fitness functions. The larger the number of leakages, the more important the WLV optimization is: this is the cause of the variation of the IMC fitness raking of the schedules changing the network.

A future interesting study could be the analysis of the same network with different numbers of leakages end/or different magnitudes, exploiting their influence on the behaviour of the proposed fitness functions.

With regards to the developed genetic algorithm, it is a very powerful methodology in order to find optimal solutions for a optimization problem. However, it had a very expensive computational cost. A massive amount of just under a hundred of GA runs have been launched during this study. A grand total of about 355 hours of runtime (almost

15 days without interruption) have been spent just for the results presented in the thesis. Many other runs have not been reported in the work. Nonetheless, the genetic algorithm confirmed the goodness of the Ex-ante Schedules and their relation with the chosen fitness function.

Genetic algorithms are very functional solution but they have the drawback of an high computational time when the number of leakages to be fixed is high. Probably the code could be optimized to run faster. An attempt on this side, was to implement the parallel computing technique in order to run different individuals at the same time, but it was not an easy task due to the necessity to use the Epanet-Matlab Toolkit [7].

If the computational cost is reduced, more and larger networks can also be analysed. This could extend the study and help to generalize the behaviour of both ESs and genetic algorithm.

More fitness components could be added to the calculation to include even more aspects in the optimization. Moreover, as explained in chapter 6, different approaches could be taken to aggregate the fitness components into a single fitness function. In this work, it was thought of not normalizing the components to have a practical meaning of the fitness function (that is the economic aspect as Intervention and Management Cost): in future, the fitness function with the normalized components could be studied since the intervention cost and the digging cost weighted the comparisons even if they were constant costs and did not influence the solution. Additionally, a Pareto approach could be used to handle the multi-objective fitness function [14].

Finally, different designs of the GA could be future studies, in terms of different genetic operators or parameters. As explained, plenty of either crossover and mutation operators are designed in literature, each of them with its characteristics and qualities.
In addition, several implementation could be made to the operators, such as a random moving crossover point instead of a fixed point at the exact centre of the individual.

Furthermore, adaptive or dynamic parameter could be implemented, such as the number of switches or the composition of the next generation: they could be related to the trend of the fitness, exhorting the algorithm to enlarge its search space if it is approaching to a local minimum.

Bibliography

[1] L. S. Araujo, H. Ramos, and S. T. Coelho. Pressure control for leakage minimisation in water distribution systems management. *Water Resources Management*, 20:133–149, 2 2006. ISSN 0920-4741. doi: 10.1007/s11269-006-4635-3.

[2] S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithm. 1995.

[3] A. Barolini. Quanto costa, in media, l'acqua del rubinetto in italia, 7 2017. URL `https://www.lifegate.it/prezzi-acqua-italia-europa`.

[4] P. R. Bhave and R. Gupta. *Analysis of Water Distribution Networks*. Alpha Science International, 2006. ISBN 9781842653593. URL `https://books.google.it/books?id=7gxSAAAAMAAJ`.

[5] L. Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991. ISBN 9780442001735. URL `https://books.google.it/books?id=Kl7vAAAAMAAJ`.

[6] A. Eiben and J. Smith. *Introduction to Evolutionary Computing*. Springer Berlin Heidelberg, 2015. ISBN 978-3-662-44873-1. doi: 10.1007/978-3-662-44874-8.

[7] D. G. Eliades, M. Kyriakou, S. Vrachimis, and M. M. Polycarpou. Epanet-matlab toolkit: An open-source software for interfacing epanet with matlab. 9 2016.

[8] D. Fogel. Phenotypes, genotypes, and operators in evolutionary computation. page 193. IEEE, 1995. ISBN 0-7803-2759-4. doi: 10.1109/ICEC.1995.489143.

[9] S. Fox, R. Collins, and J. Boxall. Experimental study exploring the interaction of structural and leakage dynamics. *Journal of Hydraulic Engineering*, 143, 2 2017. ISSN 0733-9429. doi: 10.1061/(ASCE)HY.1943-7900.0001237.

[10] D. Fuchs-Hanusch, D. Steffelbauer, M. Günther, and D. Muschalla. Systematic material and crack type specific pipe burst outflow simulations by means of epanet2. *Urban Water Journal*, 13:108–118, 2 2016. ISSN 1573-062X. doi: 10.1080/1573062X.2014.994006.

[11] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. 1:69–93, 1 1991. ISSN 1081-6593. doi: 10.1016/B978-0-08-050684-5.50008-2.

[12] G. Guarino, L. Preite, G. Scaletti, G. Ferrarese, and S. Malavasi. Five-years rehabilitation plan of a water distribution network. pages 1–10, 2022.

[13] P. Jowitt and C. Xu. Optimal valve control in water-distribution networks. *Journal of Water Resources Planning and Management-asce - J WATER RESOUR PLAN MAN-ASCE*, 116, 7 1990. doi: 10.1061/(ASCE)0733-9496(1990)116:4(455).

[14] A. Konak, D. W. Coit, and A. E. Smith. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering System Safety*, 91:992–1007, 9 2006. ISSN 0951-8320. doi: 10.1016/J.RESS.2005.11.018.

[15] P. Larrañaga, C. Kuijpers, R. Murga, I. Inza, and S. Dizdarevic. Genetic algorithms for the travelling salesman problem: A review of representations and operators. *Artificial Intelligence Review*, 13:129–170, 1999. ISSN 02692821. doi: 10.1023/A:1006529012972.

[16] M. Mitchell. An introduction to genetic algorithms, 1996.

[17] I. M. Oliver, D. J. Smith, and J. R. C. Holland. A study of permutation crossover operators on the traveling salesman problem. pages 224–230. L. Erlbaum Associates Inc., 1987. ISBN 0805801588.

[18] L. Ormsbee, S. Hoagland, E. Hernandez, A. Hall, and A. Ostfeld. Hydraulic model database for applicated distribution systems research. *Journal of Water Resources Planning and Management*, 2022. doi: 10.13023/kwrri.wdsrd. URL `https://doi.org/10.13023/kwrri.wdsrd`.

[19] J. Pearl. *Heuristics: Intelligent search strategies for computer problem solving*. Addison-Wesley Pub. Co., Inc.,Reading, MA, 1984. URL `https://www.osti.gov/biblio/5127296`.

[20] J. Reca and J. Martínez. Genetic algorithms for the design of looped irrigation water distribution networks. *Water Resources Research*, 42, 5 2006. ISSN 00431397. doi: 10.1029/2005WR004383. Balerma.

[21] L. A. Rossman, H. Woo, M. Tryby, F. Shang, and R. Janke. Epanet 2.2 user manual.

[22] G. Syswerda. Schedule optimization using genetic algorithms. *Handbook of genetic algorithms*, 1991.

[23] M. Taherdangkoo, M. Paziresh, M. Yazdi, and M. Bagheri. An efficient algorithm for function optimization: modified stem cells algorithm. *Open Engineering*, 3:36–50, 3 2013. ISSN 2391-5439. doi: 10.2478/s13531-012-0047-8.

[24] UPV. Biws instructions. 2022.

[25] D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4, 6 1994. ISSN 0960-3174. doi: 10.1007/BF00175354.

[26] X. Yu and M. Gen. *Introduction to Evolutionary Algorithms*, volume 0. Springer London, 2010. ISBN 978-1-84996-128-8. doi: 10.1007/978-1-84996-129-5.

A | Appendix: Database of Water Distribution Networks

WDNname	Nodes #	Pipes #	Length m	Valves #	Pumps #	Reservoirs #	Tanks #	Users #	Demand m^3/d	Leakages #	Leak Flow m^3/d
1 Hystoric											
19 Pipe System	12	21	8 361	0	0	2	0	4	11 992	0	-
Modified 19 P.S.	12	21	10 058	0	0	2	0	4	11 992	0	-
FOWM	44	49	27 576	0	0	1	0	15	38 157	0	-
fourteenpipes	10	14	30 573	0	0	2	0	9	12 539	0	-
hanoi	31	34	39 420	0	0	1	0	31	478 561	0	-
new_york	19	42	222 992	0	0	1	0	19	4 935 964	0	-
3 Battles											
Anytown	19	40	35 174	0	1	3	0	16	34 886	0	-
CalibrationNetw.	388	429	56 724	4	11	1	7	0	0	0	-
LongTermImprov.	399	443	60 417	5	11	1	7	348	36 484	0	-
Sensor Network 1	126	168	37 559	8	2	1	2	79	5 150	0	-
4 Kentucky											
ky1	856	984	65 616	0	1	1	2	797	7 485	0	-
ky2	811	1 124	152 248	0	1	1	3	757	7 910	0	-
ky3	269	366	91 287	0	5	3	3	249	7 597	0	-
ky4	959	1 156	260 241	0	2	1	4	934	5 672	0	-
ky5	420	496	96 581	0	9	4	3	392	8 585	0	-
ky6	543	644	123 200	1	2	2	3	498	6 204	0	-
ky7	481	603	137 049	0	1	1	3	463	5 789	0	-
ky8	1 325	1 614	247 343	0	4	2	5	1 257	9 329	0	-
ky9	1 242	1 270	961 908	56	17	4	15	1 030	5 076	0	-
ky10	920	1 043	430 026	5	13	2	13	871	8 184	0	-

Table continues to next page

Table continued from previous page

WDNname	Nodes #	Pipes #	Length m	Valves #	Pumps #	Reservoirs #	Tanks #	Users #	Demand m^3/d	Leakages #	Leak Flow m^3/d
5 KY valved											
ky8_v	2 439	2 237	235 986	488	4	2	5	1 255	9 307	0	-
ky18_v	2 643	2 317	220 777	467	3	1	3	495	5 217	0	-
ky21_v	789	628	471 620	204	21	1	11	332	4 419	0	-
ky22_v	587	533	53 428	96	4	1	7	349	2 382	0	-
ky23_v	2 324	1 948	515 584	441	21	1	14	930	6 482	0	-
ky24_v	288	249	84 672	43	0	2	0	161	371	0	-
6 US units											
CA1	111	126	17 902	0	0	0	1	63	5	0	-
PA1	337	399	159 703	0	0	0	2	294	16 626	0	-
PA2	262	288	18 244	0	1	1	0	230	611	0	-
WA1	121	168	49 118	1	0	0	2	118	16 758	0	-
7 IS units											
Balerma	443	454	100 263	0	0	4	0	442	95 377	0	-
MarchiRural	379	476	1 288 420	0	0	2	0	66	8 363	0	-
Zhi Jiang	113	164	126 436	0	0	1	0	112	96 025	0	-
fossolo	36	58	8 406	0	0	1	0	36	2 930	0	-
modena	268	317	71 806	0	0	4	0	245	35 160	0	-
8 Synthetic											
Jilin	27	34	28 991	0	0	1	0	26	33 172	0	-
KL	935	1 274	252 498	0	0	1	0	623	29 086	0	-
MICROPOLIS	1 574	1 415	36 658	196	8	2	1	685	5 944	0	-
9 added											
Araujo	22	37	44 261	0	0	3	0	13	12 960	22	197
Araujo_noleak*	22	37	44 261	0	0	3	0	13	12 960	0	-
Araujo_leak*	44	59	44 261	0	0	3	0	13	12 960	22	192
BIWS	2 859	3 231	338 662	15	7	6	4	2 839	17 519	0	-
BIWS_leak	5 337	5 709	338 662	15	7	6	4	2 839	17 519	3 589	14 477
Balerma_leak*	541	552	100 263	0	0	4	0	442	95 377	100	2 735
MOD	268	317	71 806	0	0	4	0	245	35 160	0	-
MOD_leak*	367	416	71 806	0	0	4	0	245	35 160	100	2 381

* networks are modified for the purpouse of the work.

B | Appendix: Network pre-processing script

The following is a script for the pre-processing of a given network. Given a Epanet inp-file and a leakage table containing leakage magnitudes k and positions x_{leak} from initial node of the pipe, it can assess whether repair or replacement costs are convenient (please refer to section 2.5). Incidentally, the script can reduce the number of interventions by grouping the leakages if the replacement is cheaper.

Algorithm B.1 Pseudo-code of the network pre-processing script

Require: Costs setting and leakages table

%% Leakages importation
import table of leakages from file.xlsx
for each leakage l in the table **do**
  assignment of name ID
  assessment of diameter D of the pipe
  computing dig width $B_{dig}(D)$
end for
%% Grouping based on replacing/repair cost
for each pipe p *% with at least one leakage* **do**
  $N_{leak}(p)$ = number of leakages on the pipe p
  call **Grouping function** (please refer to Appendix C)
end for *% N_{group} are produced; also a single repaired leakage is considered a group*
for $g = 1$ to N_{group} **do**
  assignment of name ID
  assessment of $x_{leak}(g)$ *% as median of individual locations*
  computation of $k(g)$ *% as sum of individual magnitudes*
  retrieving of $Cost(g)$ *% by the previous for-loop*
  computation of dig width $B_{dig}(g)$
  setting of dig length $L_{dig}(g) = \max(x_{leak}) - \min x_{leak}$
end for *% group table is assembled*

Algorithm continues to next page

Algorithm continued from previous page

```
%% Inserting emitter nodes for each group in the .inp file
for g = 1 to N_group do
    retrieving of old pipe properties
    computation of coordinates* of the emitter
    creation of a new node
    setting of emitter magnitude
    removal of old pipe
    addition of two new pipes % upstream and downstream the novel node
    setting of properties of novel pipes % equal to old pipe
end for

Ensure: added nodes number and added links number equal to N_group
return group table and Epanet .inp file implemented with emitter nodes
```

* the computation of coordinates of the emitter is not straightforward when the pipe is segmented. They are computed knowing the coordinates of extremities and vertices of the pipe and the x_{leak} position of the group from first node of the pipe

C | Appendix: Grouping function

The following is a recursive function required for the pre-processing script. For a given pipe p and N_{leak} leakages on it to be repaired.

Algorithm C.1 Pseudo-code of the Grouping function

Require: diameter $D(p)$, leakage magnitudes k and positions x_{leak} on the pipe p

1: **for** $l = N_{leak}$ to 1 *% the leakages are scanned starting from the last to first* **do**
2: $L_{replace}(l)$ = length of pipe from the first to the l-th leak
3: $C_{replace}(l)$ = replacing cost of the portion of pipe of length $L_{replace}(l)$
4: $C_{repair}(l)$ = sum of the repair costs of the l leakages
5: **if** $C_{replace}(l) \leq C_{repair}(l)$ **then**
6: $done = l$ *% number of computed leakages*
7: $g = g + 1$ *% add another group*
8: **assignment** of leakages from 1 to l to the g-th group
9: $Cost(l) = C_{replace}$
10: **break** for loop
11: **end if**
12: **end for**
13: **if** $done == 0$ *% i.e., the condition in line 5 is never met* **then**
14: $g = N_{leak}$ *% a group for each leakage*
15: $Cost = C_{repair}$ *% for every leakage*
16: **else if** $done < N_{leak}$ *% i.e., there are uncomputed leakages* **then**
17: the Grouping function is recursively called for remaining leakages on the pipe p
18: **end if**
19: **return** $Cost$ and g

List of Figures

List of Tables

List of Algorithms

List of Symbols

Variable	Description	SI unit
Q_{leak}	Leak Flow	m^3/s
A_{leak}	Area of the leakage	m^2
C_d	Coefficient of discharge	-
g	Gravitational constant	m/s^2
k	Emitter leakage magnitude/Effective leakage area	$l/s/m^{\gamma}$
γ	Emitter pressure exponent	-
p	Pressure head in the leakage node	m
N_{leaks}	Number of leakages in the network	-
N_g	Number of genes ($= N_{leaks}$) in a individual	-
N_i	Number of individuals per generation	-
N_e	Number of elite individuals in a population	-
N_c	Number of crossbread individuals in a population	-
N_m	Number of mutated individuals in a population	-
N_s	Number of genes switches per mutation	-
N_r	Number of random individuals in a population	-
$t_{intervention}$	Time between interventions	s
F_i	Fitness value of the i-th individual	-
F_l	Fitness value of the l-th leakage/gene	-
WLV	Water Loss Volume	m^3
C_{wlv}	Water loss cost	€
c_{water}	Drinking water charge	€$/m^3$
w_{wlv}	Weight of C_{wlv}	-
C_{int}	Intervention cost	€
C_{repair}	Repair cost	€
$C_{replace}$	Replacement cost	€

Variable	Description	SI unit
w_{int}	Weight of C_{int}	-
C_{dig}	Digging cost	€
c_{dig}	Digging price	€$/m^3$
V_{dig}	Volume of the pit to excavate	m^3
A_{dig}	Area of the pit to excavate	m^2
z_{dig}	Depth of the pit to excavate	m
B_{dig}	Width of the pit to excavate	m
$A_{overlap}$	Overlapping area of subsequent interventions	m^2
w_{dig}	Weight of C_{dig}	-
C_{vic}	Vicinity cost	€
c_{vic}	Relocation price	€$/m$
L_{th}	Threshold distance	m
w_{vic}	Weight of C_{vic}	-

List of Acronyms

Acronym	Meaning
WDN	Water Distribution Network
CNS	Current Network Status
WDSA	Water Distribution Systems Analysis
CCWI	Computing and Control in the Water Industry
BIWS	Battle of Intermittent Water Supply (2022)
ES	Engineered Schedule
MO	Multi-Objective function
WLV	Water Loss Volume
IMC	Intervention and Management Cost
EA	Evolutionary Algorithm
GA	Genetic Algorithm
HTX	Head&Tail Crossover
PG	Partially Given inital generation
CR	Completely Random initial generation

Acknowledgements