INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores

Future Railway Mobile Communication System Automated Planning

Artur Daniel Rocha Queirós

(Licenciado)

Dissertação para obtenção do Grau de Mestre
em Engenharia Electrónica e de Telecomunicações

Orientadores : Professor Nuno Cota
Professora Doutora Matilde Pós-de-Mina Pato

Júri:

Presidente: Professor Doutor Vitor Manuel de Oliveira Fialho

Vogais: Professor Doutor António João Nunes Serrador
Professor Nuno António Fraga Juliano Cota

Setembro, 2023

INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores

Future Railway Mobile Communication System Automated Planning

Artur Daniel Rocha Queirós

(Licenciado)

Dissertação para obtenção do Grau de Mestre
em Engenharia Electrónica e de Telecomunicações

Orientadores : Professor Nuno Cota
Professora Doutora Matilde Pós-de-Mina Pato

Júri:

Presidente: Professor Doutor Vitor Manuel de Oliveira Fialho

Vogais: Professor Doutor António João Nunes Serrador
Professor Nuno António Fraga Juliano Cota

Setembro, 2023

Acknowledgments

First I want to thank my professors, Nuno Cota and Matilde Pato for all the support given and all the hours spent helping and guiding me. I also want to thank Solvit for the data provided and the Instituto Superior de Engenharia de Lisboa and all my past professors for all the knowledge that has been given.

I want to thank my family, specially my brother João who has been the biggest help during all my thesis. My closest friends, who have a special place in my heart, I want to say thank you and sorry for all the plans canceled due to my thesis.

In last but most important I want to thank my girlfriend Raquel who was my greatest motivation and my biggest pillar, who has spend hours by my side supporting me and waiting for this chapter of my life to be closed.

Abstract

With the end of support for the Global System for Mobile Communications Railway (GSM-R) in sight, there needs to be a change in the technology used in railway communications. To achieve this, operators began the transition to Future Railway Mobile Communication System (FRMCS).

The work developed in this thesis demonstrates the application of the concept of genetic algorithms in a telecommunications network planning. Where the objective is to understand whether the results obtained are viable comparing to the currently practiced values. To achieve this, it is necessary to develop an algorithm that allows to obtaining the best possible solution for the placement of antennas along the line in the most efficient way, taking into account its coverage.

The work developed includes the construction of this same algorithm and all its phases. Using the Cascais line as a test subject and with the help of data made available by the company Solvit, it is possible to obtain different scenarios varying four parameters, the population size, the number of generations, the crossover probability and the mutation probability.

The final results prove that the use of genetic algorithms to optimize a railway telecommunications network can be a useful and powerful tool, as the results obtained presents an optimized value compared to the current solution used by public operators.

Keywords: Telecommunications; Railway Communications; Genetic Algorithms; Network Optimization; FRMCS

Resumo

Com o fim do suporte da rede Global System for Mobile Communications Railway (GSM-R) à vista, é necessário que ocorra uma mudança na tencologia usada em comunicações ferroviárias. Para isso os operadores começaram a transição para Future Railway Mobile Communication System (FRMCS).

O trabalho desenvolvido na presente tese demonstra a aplicação do conceito de algoritmos genéticos no planeamento de uma rede de telecomunicações. Onde o objetivo é perceber se os resultados obtidos são viáveis e de boa qualidade em comparação com os valores atualmente praticados. Para isso, é necessário o desenvolvimento de um algoritmo que, de forma eficiente permita obter a melhor solução possível para a colocação das antenas ao longo da linha, tendo em conta a cobertura da mesma.

O trabalho desenvolvido incluí a construção deste mesmo algorítmo e de todas as suas fases. Utilizando a linha de Cascais como sujeito de teste e com o auxílio de dados disponibilizados pela empresa Solvit é possível obter diversos cenários variando quatro parâmetros, a dimensão da população, o número de gerações, a probabilidade de cruzamento e a probabilidade de mutação.

Os resultados finais comprovam que o uso de algoritmos genéticos para a otimização de uma rede de telecomunicações em ferrovia pode ser uma ferramenta útil e poderosa, uma vez que os resultados obtidos apresentam um valor otimizados em comparação com o valor da solução atual com a parametrização usada.

Palavras-chave: Telecomunicações; Comunicações Ferroviárias; Algoritmos Genéticos; Otimização de redes; FRMCS.

Contents

List of Figures

List of Tables

Acronyms

3GPP	3rd Generation Partnership Project. 2
5G	Fifth Generation. 1, 2
6G	Sixth Generation. xi, 7, 8
BA-List	BCCH (Broadcast Control Channel) Allocation List. 9
ESD	Estimated Standard Deviation. 9
FRMCS	Future Railway Mobile Communication System. vii, ix, 1, 2
GA	Genetic Algorithms. 2, 9, 10
GSM	Global System for Mobile Communications. 9
GSM-R	Global System for Mobile Communications Railway. vii, ix, 1, 6, 9
kp	kilometric points. 11, 12, 14, 21, 23
MAPE	Mean Absolute Percentage Error. 5
ME	Medium Error. 9
ML	Machine Learning. 5, 7, 8
MLP	Multi-Layer Perceptron. 7
MSE	Mean Square Error. 7
QoS	Quality of Service. 2, 8

RMSE Root Mean Square Error. 5, 9

1

Introduction

Network planning holds vital importance in the telecommunications industry. Nevertheless, the human-driven methodologies can be both time-consuming and vulnerable to human-errors, often resulting in sub optimal configurations.

1.1 Motivation

With the evolution of the telecommunications network in railway infrastructure, operators are accepting a change in the technology used. With the end of market support for Global System for Mobile Communications Railway (GSM-R) technology scheduled for 2030 [1], a transition is necessary, with Future Railway Mobile Communication System (FRMCS) technology playing an important role. Utilizing FRMCS built upon the foundation of Fifth Generation (5G), it becomes feasible to provide highly dependable, low-latency communications essential for mission-critical operations [2].

The motivation for this thesis and the usage of FRMCS is, as mentioned before, the technology evolution. GSM-R is aging, becoming less flexible and with a lower interoperability than newer technologies [1]. For railway operators, this transition must be gradual and phased, as there are risks and disadvantages associated with postponing the transition of critical communication infrastructures that depend on GSM-R.

Regarding FRMCS, numerous features and benefits have been documented, as outlined in the literature [2], including:

- Simultaneously provide flexible support for both 3rd Generation Partnership Project (3GPP) and non-3GPP access technologies.
- Sustain optimal performance even at high speeds by incorporating features that effectively counter Doppler effects, guaranteeing reliable communication systems performance at speeds exceeding 500 km/h.
- Offer assured bit rates for mission-critical services through the implementation of flexible, flow-based Quality of Service (QoS), a 5G capability that segregates critical and non-critical applications and data.

Much like FRMCS, Genetic Algorithms (GA) are a topic that has witnessed significant growth in recent years. Applying this concept to the telecommunication sector GA can play a pivotal role in the planning of telecommunication networks, especially when dealing with advanced technologies as exemplified in this case with FRMCS. As a highly complex task, there are numerous variables to consider. These include coverage areas, capacity requirements, interference mitigation, and the optimal placement of base stations, among several other factors demanding careful evaluation.

These algorithms are well-suited for tackling complex optimization problems. They are a class of optimization techniques renowned for their ability to explore a vast array of potential network configurations. Their utility extends to seeking solutions that minimize interference, maximize coverage, optimize resource allocation, and fulfill various performance objectives. In the context of FRMCS, achieving efficient optimization is paramount to guaranteeing reliable and high-quality communication.

What makes GA particularly appealing is their potential to yield cost savings. By identifying network configurations that are resource-efficient, they can diminish the necessity for additional infrastructure or resources. This, in turn, leads to reductions in both capital and operational expenses.

Furthermore, GA can leverage parallel processing to accelerate the search for possible solutions. This capability is essential, particularly when dealing with the complexity of large-scale railway networks and the need to evaluate numerous potential network configurations.

In summary, the utilization of GA in the planning of telecommunication networks for FRMCS within railway systems can be advantageous. This advantage arises from the intricate nature of the task, the imperative need for optimization, the adaptability to ever-changing conditions, and the proficiency in handling constraints and scalability. These algorithms, by facilitating the deployment of advanced communication systems, can significantly contribute to the future of railway communication.

1.2 Objectives

The main objective of this dissertation is to explore the possibility of an automated planning telecommunications network within a railway infrastructure based on railway mobile communication system. This study involves the development and implementation of a specialized algorithm based on genetic algorithms, which utilizes four distinct variables to generate a diverse range of results. These results will then be systematically compared with the existing real-world data.

The chosen context for this study is the Cascais railway line, and pertinent data for this investigation has been provided by Solvit. This in-depth examination of the Cascais line will serve as a solid foundation for the research into the automation of telecommunications network planning within railway systems, offering valuable insights into the applicability of genetic algorithms in this domain. Through rigorous analysis and meticulous comparison, this research endeavor aims to contribute significantly to the ongoing research topic surrounding the optimization of telecommunications network.

1.3 Thesis structure

The structure of this thesis is divided into 5 chapters. The first chapter serving as an introduction to this document. The second chapter covers topics related to the theme developed in this thesis, such as the optimization of propagation models using genetic algorithms and the estimation of radio coverage using neural networks. Chapter 3 presents the development of the algorithm with all its steps explained, namely the explanation of the elements of an individual. It also shows how a solution to this problem is characterized, presenting some examples and finally how the evolution of an individual occurs during a test. Chapter 4 describes the results obtained from the experimental evaluation. By manipulating four different variables, different individuals' evolution are presented and their analysis. In the final part of this chapter, a comparison is made with the real values of the studied line to make it possible to understand the quality of the tests made. Finally, Chapter 5 concludes this document, summarizing the key findings of this thesis, as well as outlining the main directions for future research work

1.4 Contributions

To contribute to the scientific community, the algorithm used on this thesis is available on GitHub at:

https://github.com/arturqueiros/NetworkAutomatedPlanning_Thesis.git

2

Related work

This chapter explores previous research concerning railway communication systems. Due to a limited prior research available on the specific topic of this thesis, this chapter will focus on exploring the practical applications of machine learning algorithms and automated planning applied to various telecommunications areas.

2.1 Machine learning applied to telecommunications

Quoting El Naqa and Murphy [3] “Machine Learning (ML) is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment”. This section addresses, different cases on the telecommunication sector will be mentioned and their results presented.

2.1.1 Linear regression applied to point-to-point microwaves link.

A possible use case for ML is the application of linear regression [4] in point-to-point microwaves link problems [5].

In order to calibrate theoretical calculations using a model based on real received power data, the study [5] was made using linear regression. The model’s validation utilizes the following metrics: Root Mean Square Error (RMSE) [6], Mean Absolute Percentage Error (MAPE) [7] and R_2^a(Coefficient of determination) [8].

After comparing both power estimations [5], the dimensioning of the different link's margin will be possible. A regression model was developed to predict received power based on the characteristics of each microwave link (practical estimation). Both theoretical and practical estimates of received power were compared to calculate the link margin. This analysis provided a better understanding of the microwave link's attenuation margin. The results showed that in cases where theoretical models underestimated the link, a more accurate estimation of the attenuation margins was possible.

2.1.2 Global System for Mobile Communications-Railway radio coverage estimation using Neural Networks

To minimize installation expenses, the deployment of Global System for Mobile Communications Railway (GSM-R) systems on Portugal's railway lines requires an accurate computation of the signal. In similarity to [9] (with genetic algorithms), this study [10] proposes an alternative methodology by leveraging prior measurements and neural networks. Employing model parameters and topographical data, supervised training was employed to instruct the networks on the signal characteristics exhibited by railway lines. In the initial phase, a complex designed and meticulously implemented multi-layered neural network was employed to estimate radio coverage across various scenarios. The results demonstrated a significant reduction in estimation error, presenting a superior solution compared to previously explored methodologies.

Subsequently, a distinct approach was pursued in the second experiment, involving the training of a perceptron using a combination of Okumura-Hata model [11, 12] parameters and corrective factors. Despite yielding a higher degree of error, this experiment still afforded valid signal estimations, allowing for comparative analysis against the previous network.

In the final stage of the research, a neural network-based application was developed to provide a comprehensive solution applicable to all examined scenarios. A competitive neural network was trained to classify diverse samples into distinct environmental types. Subsequently, an optimized neural network was individually trained for each environment. Rigorous testing revealed a slight increase in error compared to the optimized networks within specific scenarios, but overall outperformed previous methods by showcasing reduced error, establishing itself as a more favorable solution.

The results demonstrated that the estimation of radio coverage in GSM-R through the implementation of neural networks presents a practical and advantageous alternative in both general and specific scenarios. When compared to the original theoretical

Okumura-Hata propagation model and the Okumura-Hata model optimized using genetic algorithms, the employment of neural networks exhibits significant superiority. This manifests in enhanced accuracy for GSM-R network planning, leading to reduced error margins, decreased base station requirements, and lowered system installation costs.

2.1.3 Artificial intelligence-based path loss models for cellular mobile networks.

Accurate estimation of path loss is crucial during the initial stages of wireless network deployment and cell design. Various path loss models, such as the Okumura-Hata model, are available. However, they are constrained by specific parameters. This study [13], proposes a path loss model based on Multi-Layer Perceptron (MLP) neural networks. The proposed model adopts an implementation network design and uses grid search-based hyperparameter tuning. It effectively approximates path losses for both mobile devices and base stations. To improve prediction accuracy, the study investigates the optimal number of neurons, learning rate and hidden layers.

The MLP models were created to train 14 data sets, and compared to predictions given by six existing empirical models. Two error measures were used to assess the created model: MSE and R-square. The presented MLP models outperformed the other six empirical models and showed the lowest error in various environments, which make them good fits for the measured data. Increasing the number of MLP input variables also increased the accuracy of the produced model, which can predict path loss measurements in wireless propagation situations. Therefore, the suggested MLP is effective in predicting path fading and is considered a generic data-driven model.

2.1.4 Distributed machine learning techniques for Sixth Generation (6G) networks.

The primary objective of this research paper [14] is to conduct an extensive survey of influential studies on distributed learning technologies in the context of 6G networks. The advent of 6G technology is expected to enable the development of an intelligent, highly scalable, dynamic, and programmable wireless communication network capable of catering to diverse wireless devices. ML techniques are anticipated to play a vital role in addressing complex networking challenges within the intelligent 6G network. This will involve the generation of substantial amounts of data through external sensors by various 6G nodes and devices, necessitating thorough data analysis.

Given the distributed and large nature of this data, combined with advancements in computing hardware, the utilization of distributed ML techniques becomes crucial in the 6G landscape. While these techniques offer several advantages over centralized ML approaches, their implementation in resource-constrained wireless environments presents notable challenges. Therefore, it is imperative to carefully select the most appropriate ML algorithm based on the wireless environment's characteristics and the learning process's resource requirements. This study aims to comprehensively survey recently introduced distributed ML techniques, analyzing their distinct characteristics and potential benefits, with a specific emphasis on the most influential research papers in this domain.

2.2 Genetic algorithms applied to telecommunications

As previously mentioned before, genetic algorithms are a topic that has witnessed significant growth in recent years, with that in mind this section addresses different cases of applications on telecommunications.

2.2.1 Cellular networks frequency plan optimization using genetic algorithms.

Continuous evolution of cellular networks is crucial to enhance Quality of Service (QoS) and ensure comprehensive coverage. The foundation of radio cellular systems lies in frequency allocation. The principle of frequency allocation involves selecting an optimal frequency plan that effectively addresses traffic demands, ensures communication quality, and minimizes radio interference. To address this issue, the paper [15] presents an original approach that employs genetic algorithms for optimizing frequency allocation while simultaneously minimizing both co-channel and adjacent channel interference. By leveraging the power of genetic algorithms, this proposed method offers an efficient solution to optimize frequency allocation in cellular networks.

Looking at the results, this study presents findings that highlight the limitations of planning software relying solely on environmental modeling and radio wave propagation models in achieving an optimal frequency plan. Furthermore, it emphasizes the necessity for an effective approach to minimize co-channel and adjacent channel interference in the network, thereby optimizing the frequency plan. To address this, the authors propose a novel optimization method, leveraging genetic algorithms in conjunction with the BCCH (Broadcast Control Channel) Allocation List (BA-List) table. This approach enables the identification of neighboring cells for each serving cell within the network. The effectiveness of the proposed method is validated through its application in the Global System for Mobile Communications (GSM) network of MOBILIS-Algeria operator to solve an optimization problem. The results obtained from this work are encouraging, as they demonstrate the ability of the developed approach to uncover improved solutions.

2.2.2 Optimization of propagation model using genetic algorithms.

The study [9] is an example of an application of Genetic Algorithms (GA), where the main goal focuses on developing solutions for the calibration of propagation models in order to make them more efficient when predicting radio coverage in railways using GAs. This project plans a GSM-R network and uses optimization techniques based on GAs.

Examining the results of this study, it is possible to observe that three tests were made in four different railways: Cascais, Sintra, Oeste and Évora.

During the initial test, random samples were chosen for each scenario, consisting of both a training sample and a testing sample containing a proportion of the total measurement points. The algorithm was applied to the training sample, resulting in optimized parameters that were subsequently utilized on the testing sample. Across all scenarios, the optimization algorithm effectively enhanced the accuracy of measurement predictions by reducing Medium Error (ME), RMSE and Estimated Standard Deviation (ESD) statistics, while also increasing correlation. Furthermore, the standard deviations obtained from the optimized parameters were significantly lower when compared to those used by public operators.

For the second test, a random training sample and testing sample were chosen for each scenario, ensuring an equal number of measurement points in both sets. The training samples were merged to form a resultant sample, from which optimized parameters were derived and applied to each respective testing sample. The enhancements observed in the second test were not as substantial as those witnessed in the previous test. With this test, it was possible to understand the difficulty of developing a model that can effectively adapt to various types of environments.

Considering the impact on signal propagation caused by water on the third and last test, the emphasis was placed on optimizing the points affected by this environment. This involved considering all points influenced by water, irrespective of the scenario. Subsequently, a random selection was made from these points to form both training and testing sample. The results from this test indicate a potential opportunity to group points with determined characteristics and identify model parameters that effectively describe them.

The usage of GA and the defined configuration was validated by the application of the techniques on the different scenarios. It was possible to reduce the margin typically used in cellular network planning by public operators. This reduction enables the possibility of lower number of base stations in the planning process, reducing the implementation costs.

3

Development and design

For the development of this thesis, it was necessary to have an algorithm that, given the railway map with its respective stations (and kilometer points) and the associated coverage levels, could produce an optimal solution. This optimal solution corresponds to the solution with the lowest value of the cost function, which will be explained further ahead. The fourth chapter will address topics such as the structure and parameters of the created algorithm.

3.1 Railway

Developing the necessary algorithm, proposed in this thesis, requires a fundamental consideration: ensuring its generality. To achieve this goal, it is essential to understand the elements and aspects of the line. For this problem, a railway line is composed by three elements, represented on Table 3.1.

Table 3.1: Line variables

Variable	Type
Stops	String
kilometric points (kp)	Float
Coverage	Float

A stop is by definition a point where the train passes or stops. It is characterized by a name, a type and a kilometric points (kp). The type is represented by six different classes, where each class has a priority level to have a base station[1]. The different classes are "anchor", "station", "halt", "level crossing", "signal" and "other" (the class other represents every stop type that isn't included in the previous classes). The priorities are represented by an integer and the lower the priorities the more favored are the stops to have a base station. Anchors are a special case where it is mandatory to have a base station on that location. The kp marks the point on the line where the stop is located.

The Coverage signal, mentioned on Table 3.1, is an important factor on any telecommunication problem. In this case, a minimum level is established as -95 dBm. There is also a low level signal (-85 dBm) where the signal is higher than the minimum, but a better coverage level is recommended. To calculate the level of signal in each solution, a file was provided containing all coverage along the line.

Input and output files

As mentioned before, the coverage values come from a file containing all coverage values. The file contains a column with all kp and a column for each different stop with the associated coverage. These values represent the coverage along the line if a single base station is present. When two or more stops are active, the algorithm will compare the coverage on each kp and chose the highest. If at any kp the coverage is zero, the file will have a blank cell.

As the goal of this algorithm is to be as generic as possible, another file is used as an input. This file contains 3 rows, the first with the name of every stop on the line, the second with the kp associated with each stop. The third and final column will have the type of each stop. This process is represented on Figure 3.1.

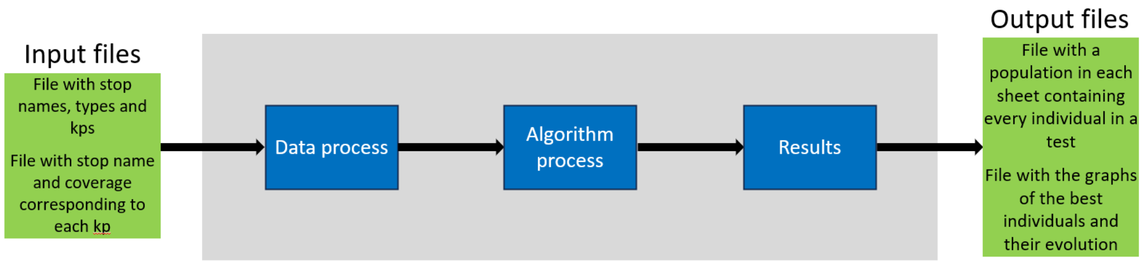


Figure 3.1: Algorithm creation workflow

For the output files, two files are generated with the results of each test. The first one contains every individual with their parameter value in a population. This file contains

[1]These classes were created in the context of this thesis

multiple sheets, each according to the number of generations. The second file has, on the first sheet, a graph with the evolution of the best individual in each generation, as shown on Figure 3.9. The second sheet contains the binary code for the best individual in each generation, as well as their coverage graph, as represented on Figure 3.6.

3.2 Genetic Algorithm

The developed algorithm is based on genetic algorithms [16, 17], specifically on the well-known Traveling Salesman Problem. This problem's main focus is to find the optimal solution that corresponds to the shortest possible path between cities. In the case of this thesis, the main focus is to find a solution with the lowest possible cost function value. To achieve this, the algorithm follows the provided pseudocode:

Algorithm 1 Genetic Algorithm

```
Generate initial population
Calculate values and parameters associated with each individual
Extract the individual with the lowest cost function value
Apply selection method to extract parents
Generate new individuals and calculate their parameters values to keep the population size
Apply crossover method on parents to get children
Generate new individuals and calculate their parameters values to keep the population size
Apply mutation to the children
Add the extracted individual to the final population
Select elite from children
Add elite into the next generation
Repeat method until reaching number of generations desired
```

This pseudocode outlines the basic steps of the algorithm, including the initialization of the population, the evaluation of fitness, the reproduction process through crossover and mutation, the evaluation of offspring fitness, and the selection of individuals for the next generation using elitism. The process continues until certain termination criteria are met, and ultimately, the algorithm returns the best solution found. On the following subsections, further explanation will be given regarding each step.

3.2.1 Parameters

To better understand the algorithm, it is necessary to understand the parameters and concepts associated, which are:

- **Line**, the railway line is characterized by the stops, the associated kp and the coverage along the line.
- **Population**, will be characterized by individuals.
- **Individuals**, the individuals are solutions to the thesis's problem, they are different from each other within a population and are represented by a binary code (as shown on 3.1) that is directly associated with the number of stops in the railway, and it's active stops (an active stop means having a base station on that location)
- **Stop**, a stop is a point where the train passes. It can be a station, halt, crossing-level, signal, anchor, or other. This stop will be characterized by zeros and ones that represent the presence of a base station. In the case of anchor, a base station must be always present.
- **Coverage signal**, this is the most important factor on a telecommunication problem. In this case, a minimum level is established as -95 dBm. There is also a low level signal (-85 dBm) where the signal is higher than the minimum, but a better coverage level is recommended.

3.2.2 Generate Population

A population will be the foundation of every test. Initially, a fixed number of individuals will be created and every calculation to get the value of the cost function (calculation of the percentage of coverage level below the minimum and above the minimum but below the low level and the maximum extension between the two coverage level references) will be made. After obtaining the cost function value, the individuals will be sorted in ascending order of cost function value. Each individual code will be represented by 3.1.

$$individual = [a_1, a_2, a_3, a_4, a_5, a_6, a_7, ..., a_n] \tag{3.1}$$

Where a represents a binary value (0 or 1), this binary value will indicate the state of each stop (active or not). To generate this binary code, a randomizing process was used:

```
individual_binary_code = [random.randint(0, 1) for _ in totalstops]
```

This process generates an array with a sequence of binary values with the size of *totalstops*. Having the binary code generated, it is necessary to verify if the individual has an equal code to any individual in the same population. In case it is confirmed, a new binary codes is generated until the target size of population is reached, and all individuals have different codes.

Figure 3.2 represents the workflow of how a population is generated. The first step is to generate the individual binary code with the size corresponding to the line's total number of stops. Having the binary code generated it is necessary to verify if there is any individual in the same population with an equal binary code, if this is verified a new code is generated until this condition is false. The second step is to calculate each variable associated with an individual, after the calculation the individual will be inserted in the population. After these steps, the size of the population is calculated and, if the actual size corresponds to the pretended size, they will be sorted by ascending order of the cost function value. Otherwise, new individuals will be generated, and the process repeats until the desired population size is reached.

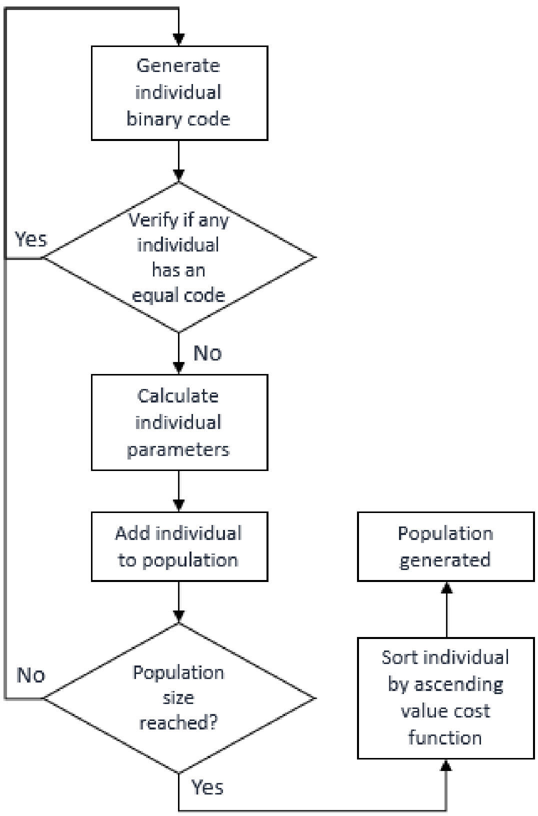


Figure 3.2: Population creation workflow

Each individual's associated variable, calculated on step two is represented on Table 3.2.

Table 3.2: Individual variables

Variable	Type
Binary code	Binary
Coverage below minimum	Float
Coverage below low reference	Float
Maximum extension in low reference	Float
Number of active stations	Integer
Number of active halts	Integer
Number of active anchors	Integer
Number of active level crossing	Integer
Number of active signals	Integer
Number of active others	Integer
Cost function value	Float

3.2.3 Selection method

Following the creation of the population and the individuals sorted, the selection method will be applied. There are numerous selection methods: Roulette [18], Rank [19], Tournament [20], etc. In the context of this thesis, two selection methods were used, one based on the roulette method and the other on the tournament method.

Roulette method

The Roulette method is structured as a concept where each individual has a percentage of getting selected, based on their fitness, Figure 3.3.

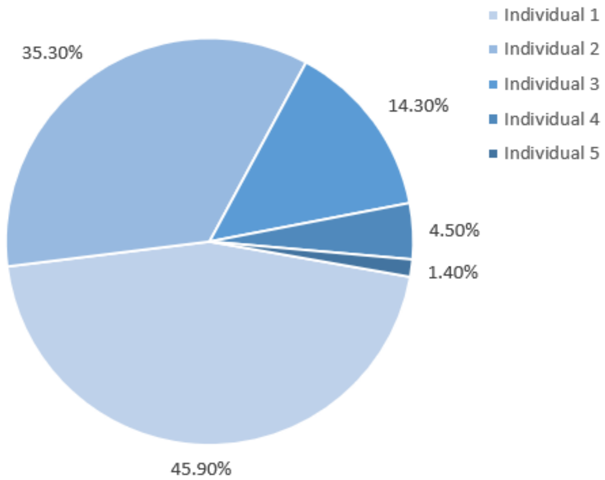


Figure 3.3: Illustrative example of a possible individual distribution in the roulette method

In this method, each individual has a calculated weight associated with the cost function value. This weight is obtained by inverting the cost function value and using cumulative weights. As all individuals have their respective weights calculated, it is necessary to select a group of individuals. The number of the selected individuals (parents) will be equal to 3.2.

$$\text{Number of parents} = \text{crossover probability} * \text{population size} \tag{3.2}$$

Individuals with lower cost function value have a higher percentage of being selected. To decide which individuals will be selected, a random number will be generated and compared to the individual's weights. The first individual that meets the requirement of having a value greater than the random number will be selected as one of the parents. This individual will be then excluded from the selection pool. This process is repeated until all necessary parents are selected. It is worth noticing that a random number is generated for each iteration.

Tournament method

The second method of selection is associated with a concept of a tournament, where a pair of randomly selected individuals engage in a comparison involving their respective cost function values. In this pairwise comparison, the winner is decided based on the individual with the lower cost function value. Following that, the chosen winner progresses to the subsequent phase of the process. Notably dissimilar from the roulette method, in this instance, unselected individuals are not removed and might participate in one or more comparisons.

Even if the selection method favors the individuals with lower cost function value, their selection isn't guaranteed. This way, elitism is applied before each selection method. This step will be explained in a following subsection. After the selection method is applied, all parents (selected individuals) go into the crossover phase.

3.2.4 Crossover method

The crossover method's objective is to create diversity among individuals. It uses two different parents from the output of the selection to obtain two new individuals, the children. There are multiple types of crossover: single-point crossover, two-point crossover, uniform crossover, etc. In this case, the method used is single-point crossover. On Figure 3.4, is possible to see an example of a single-point crossover where two new individuals (children) will be created.

Parent 1 [1, 1, 0, 1, 0, 0, 0, 1]

Parent 2 [1, 0, 0, 0, 1, 0, 0, 1]

Children 1 [1, 1, 0, 0, 1, 0, 0, 1]

Children 2 [1, 0, 0, 1, 0, 0, 0, 1]

Figure 3.4: Single-point crossover method

The crossover method starts by generating a random number to obtain the index where the crossover will happen (in this case, the index was 3. Using two parents and the index, the crossover method uses the first part of the *Parent*1 (in blue) and the second part of *Parent*2 (in blue) combining them into a new individual *children*1. The *Children*2 will be the outcome of the combination from the second part of the *Parent*1 (in red) and the first part of *Parent*2 (in red) as shown in Figure 3.4 Having new individuals is always needed to calculate all the parameters associated with an individual and sort them by cost function value.

3.2.5 Mutation method

The fourth step, and last method to be applied to a population, is the mutation method. The individuals will have a percentage (variable) where a mutation can occur. This mutation is represented by a single bit change, as shown in Figure 3.5.

Individual 1 [1, 1, 0, 0, 0, 1, 0, 0]

New Individual 1 [1, 1, 0, 1, 0, 1, 0, 0]

Figure 3.5: Single bit mutation method

Similar to the crossover method, in mutation, a random number is also generated to obtain the index (in this case 4) where the mutation will occur. The mutation causes the bit to invert its value. The only exception to this is the anchor's case. If the mutation index matches the anchor element's index, a new value is generated. This is explained by the fact that it is mandatory for the anchor element to be active in every solution. Until the mutation index ceases to match the index of the anchor, a new value is generated for the location of the mutation. After applying the mutation, the individuals have their parameters calculated and are sorted once again.

3.2.6 Elitism

As mentioned in subsection 3.2.4, even if the selection method favors the individuals with lower cost function value, their selection isn't guaranteed. This means that a percentage (variable) of individuals automatically move to crossover. These individuals correspond to the best individuals in the current population (lower cost function values). Elitism will also occur after mutation when the same percentage of individuals will move on to the next generation.

3.3 Problem Solution

The purpose of this subsection is to clarify a show a possible solution Figure 3.6 without considering the cost function.

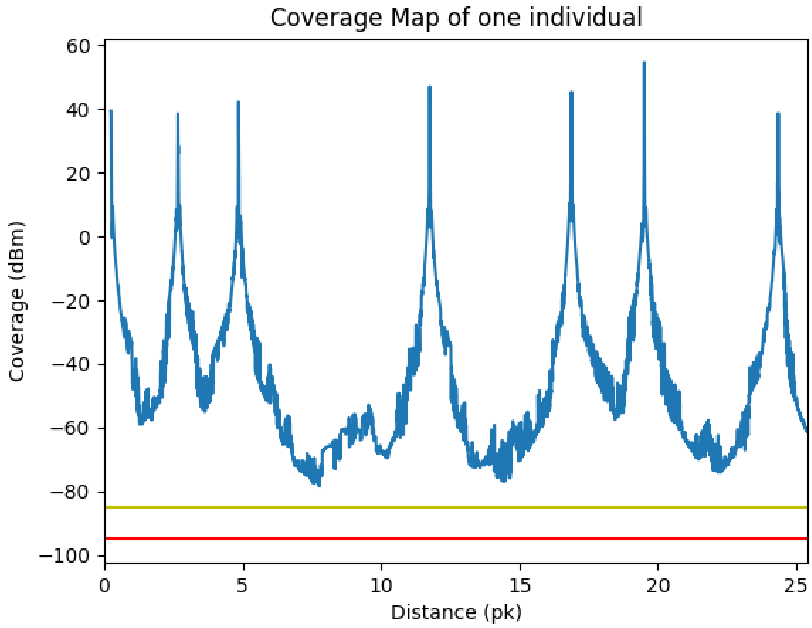


Figure 3.6: Example of one possible solution

As previously mentioned, Figure 3.6 represents a possible solution. It shows a graph with a blue line and seven peaks, this line represents the coverage signal along the railway and the peaks represent an active station. The yellow line represents the low level signal (-85 dBm) and the red line the minimum coverage level (-95 dBm).

This could possibly represent a good solution to the problem, in the next subsection the analysis on a good solution will be made just like the analysis on the formula for the cost function.

Cost Function Formula

To calculate each individual cost function, a formula was developed, tested and improved in order to obtain the best results. This formula contains the variables that will influence the quality of a solution and their associated weights. The variables and the weights are shown in the Table 3.3:

Table 3.3: Cost function formula variables and weights

Name of the parameter	Weight
Coverage below minimum	55
Coverage below low reference	20
Maximum extension in low reference [kp]	25
Number of active stations	2
Number of active halts	3
Number of active anchors	1
Number of active level crossing	5
Number of active signals	7
Number of active others	10

These were the chosen values as they demonstrated acceptable results after a trial and error process and a fine-tuning.

Utilizing the data from Table 3.3, a possible formula was obtained (also using the total number of stops within the railway) and it expresses as the sum of the multiplication between the variables and the weights. In this case of the number of each type of stops, the result of this multiplication will be divided by the total number of stops in the line. Summarizing, the formula is equal to:

1: Coverage below minimum * 55

2 :Coverage below low reference * 20

3: Maximum extension low reference * 25

4:

$$\frac{\text{Number active stations} * 2}{\text{total number of stops}}$$

5:

$$\frac{\text{Number active halts} * 3}{\text{total number of stops}}$$

6:

$$\frac{\text{Number active anchors} * 1}{\text{total number of stops}}$$

7:

$$\frac{\text{Number active level crossing} * 5}{\text{total number of stops}}$$

8:

$$\frac{\text{Number active signals} * 7}{\text{total number of stops}}$$

9:

$$\frac{\text{Number active others} * 10}{\text{total number of stops}}$$

$$\text{Cost Function} = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \tag{3.3}$$

Analyzing the equation (3.3) presented before, it's possible to verify that the higher the value of the variables, the higher the value of the cost function, making the solution worst. But from all the nine variables, three of them stand-out with higher weight, coverage below minimum, below reference and the maximum extension in low reference, meaning that these three variables will have a higher impact on the result depending on their value.

Analyzing possible solutions

The ideal solution would have all variables with a value of 0 (zero) and also no coverage below the minimum or the low reference. Such a solution would be impossible since, to obtain coverage, it's required to have active stops.

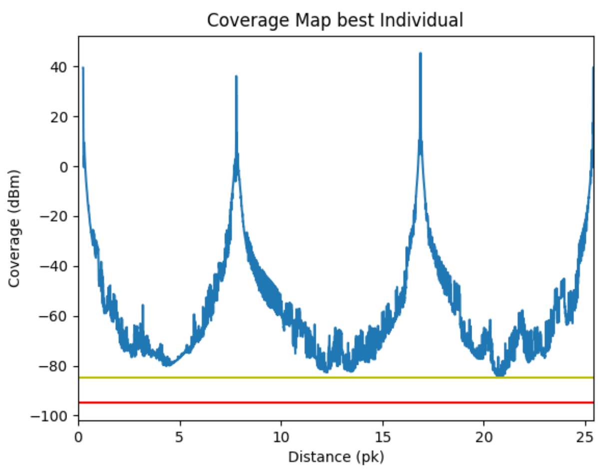
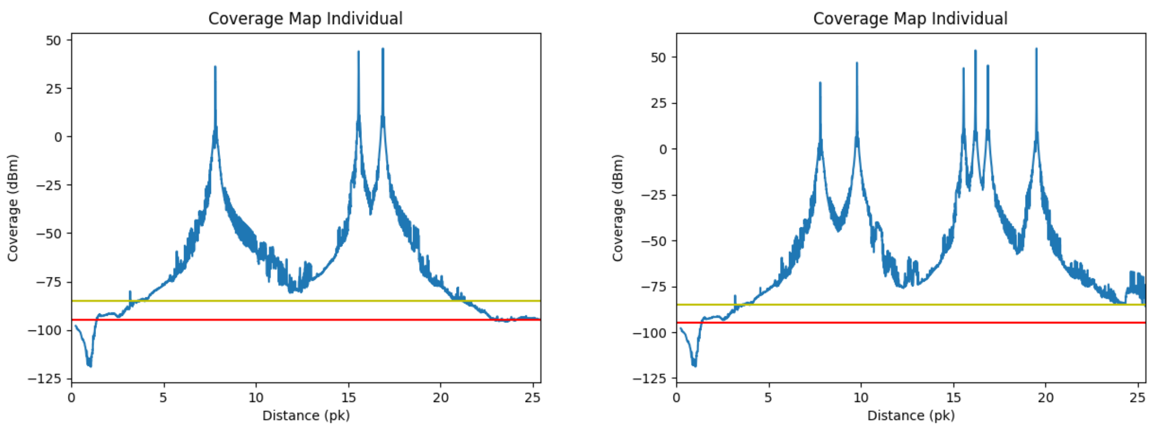


Figure 3.7: Coverage map of a possible solution with four active stops

The objective focuses on finding a solution with the minimum number of active stops, capable of maximizing the coverage of line extension, while minimizing the percentage of coverage below the low level (being above the low level will always be above the minimum level). The following analyzes presents three different possible solutions. Figure 3.7 represents one example of the best individual for this line with the lowest cost function value.

Observing Figure 3.7, it is possible to notice that it has four active stops, one anchor and three stations. The entire coverage along the line is above both minimum level and low level, and the cost function value is approximately 0.3889.

Other possible solution are represented on Figure 3.8(a) and (b):



(a) 3 active stops (b) 6 active stops

Figure 3.8: Examples of possible solutions

Observing Figure 3.8(a), it is possible to understand that it represents a bad example, due to the existence of coverage both below the low signal reference and the minimum signal, caused by the lack of active stops (only three active stops). With this two values below the minimum and adding the long extension of consecutive signal below low reference, the cost function value of this individual will be approximately 868.3506. Comparing this value with Figure 3.7 it is possible to conclude that this individual has a much higher cost function value making it a bad solution for this problem.

As previously stated, the lack of active stops can make an individual be considered as a bad solution. But this isn't the only factor that needs caution, as represented on Figure 3.8(b). In this case, there are six active stops, one anchor, two station and three halts, but even with a total of six active stops before kp 5 there is signal coverage below the low level reference and the minimum level. This was caused by a lack of dispersion of active stops.

3.4 Generations evolution

To obtain a possible solution, it is necessary to try different combinations of parameters. By varying the four different parameters, represented on Table 4.2, multiple solutions will be obtained. As a tool to understand the evolution of a solution, a plot is generated with each generation's final solution. This plot is represented on Figure 3.9.

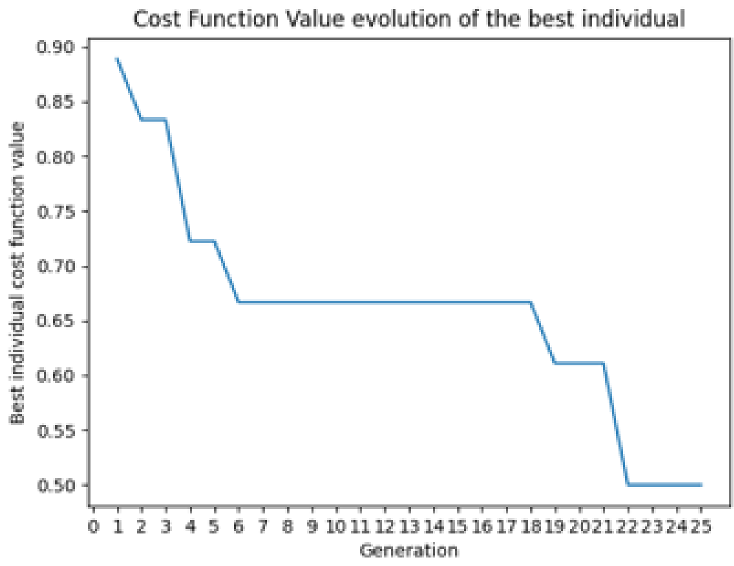


Figure 3.9: Best individual evolution over generation

In Figure 3.9 the test has 25 generations with 20 individuals in each generation. It also has a crossover probability of 0.4 and a mutation probability of 0.02. The individual cost function value fluctuates between each generation. Observing Figure 3.9 the best individual in the first generation has a cost function value of approximately 0.88, and it decreases on the second generation. This change means that the process previously mentioned created a more optimal individual in comparison to the first generation. On the third generation the best individual is the same as in the second generation that due to elitism passes on to the next generation. Between the sixth and eighteenth generation this combination of parameters (number of generations, generation size, crossover probability and mutation probability) isn't able to produce a better individual as the one produced by the sixth generation so, this value stays the same until the eighteenth generation. On the twenty-second generation the individual with the lowest cost function value is created. It stays the same until the last generation with a value of 0.5. On Chapter 4 more tests and analysis are made on how the four parameters represented on Table 4.2 can impact the quality of the tests.

4 Case study

In this chapter, the chosen line to perform tests will be Cascais's railway. This is a well known railway line with numerous measurements, and the data used for this problem was provided by Solvit. This railway is represented by 18 stops presented on Table 4.1.

Table 4.1: Cascais's Railway Stops

Name	Type	kp
Cais do Sodre	Station	0.24
Santos	Halt	0.94
Alcântara - Mar	Station	2.66
Belém	Halt	4.86
Algés	Station	7.81
Cruz Quebrada	Halt	9.79
Caxias	Station	11.76
Paço de Arcos	Halt	13.20
Santo Amaro	Halt	15.57
Oeiras	Station	16.20
PK16.8	Anchor	16.87
Carcavelos	Station	17.80
Parede	Halt	19.51
São Pedro do Estoril	Station	21.06
São João do Estoril	Halt	22.51
Estoril	Station	23.66
Monte Estoril	Halt	24.35
Cascais	Station	25.41

As mentioned before, all stops with the type anchor need to have a base station. In this case, PK16.8 will always have a base station on that location.

4.1 Tests and results

To achieve the best solution, it is necessary to make different tests with different combinations by varying the four different parameters presented on Table 4.2:

Table 4.2: Tests parameters

Parameter	Values
Number of generations	[25, 50, 75, 100]
Population size	[25, 50, 75, 100]
Crossover probability	[0.2, 0.5, 0.8]
Mutation probability	[0.01, 0.2, 0.4, 0.8]

Using these four parameters and its different values and the two different selection methods 384 tests were made. The objective of these tests is to understand how can a possible solution evolve across generations and if it can achieve the best value. The following subsection will present some results on the tests made and some in-depth analysis.

4.1.1 Number of generations

As previously stated on Table 4.2 the number of generations will vary between those four values. To understand how does the different number of generations affect the evolution of a solution, the other three parameters will be fixated with the values presented on Table 4.3:

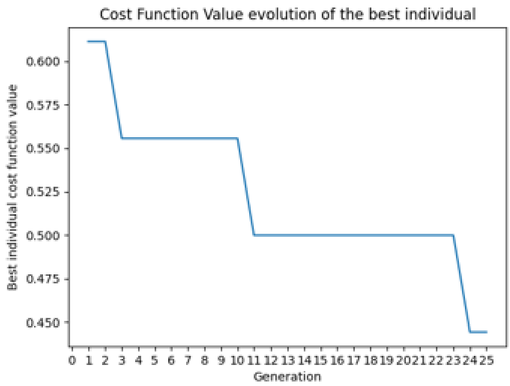
Table 4.3: Fixed values for variable number of generations

Parameter	Values
Population size	50
Crossover probability	0.5
Mutation probability	0.01

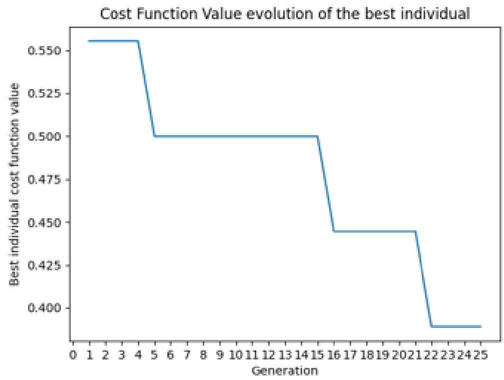
Values on Table 4.3 were randomly chosen, and don't follow a specific criteria.

25 Generations

Starting with 25 generations, the first result for each selection method obtained are present on Figure 4.1(a) and Figure 4.1(b)



(a) Using roulette selection method



(b) Using tournament selection method

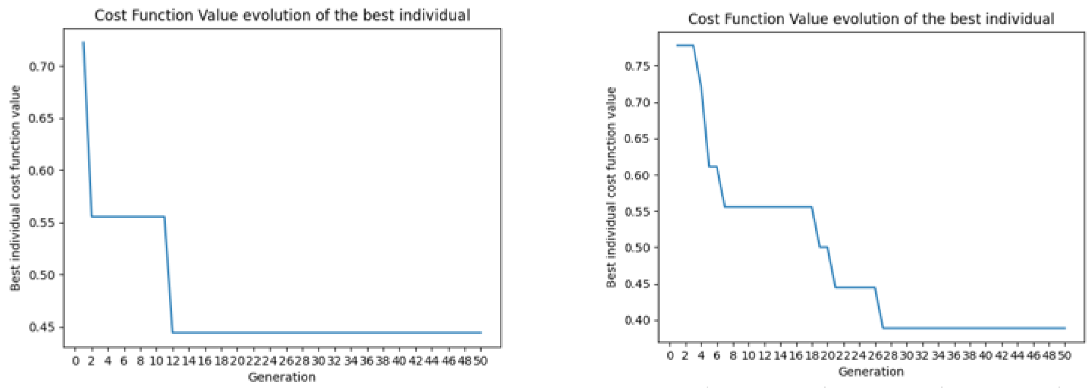
Figure 4.1: Test with 25 generations

Analyzing both figures, on (a) using the roulette selection method the cost function value of the best individual changes three times reaching a value of 0.44 on the 24^{th} generation making this the best value for this case. On (b) using the tournament selection method the cost function value of the best individual has the same behavior, changing three times but this time it reaches a value of 0.39 on the 22^{nd} generation. With this, using the tournament selection method produces better results for a test, with 25 generation achieving the best value possible.

50 Generations

The second test presented has 50 generations and the results are shown on Figure 4.2(a) and Figure 4.2(b)

Observing (a), the value of the cost function of the best individual changes its value twice, having a final value of 0.44 achieved on 12^{th} generation. This test does not reach the best value due to a lack of both individuals per generation and number of generations. On (b), the individual changes its cost function value six times and at the 27^{th} generation it stabilizes at a value of 0.39. This test has the same outcome as in Figure 4.1. Case (b) has a better result for a test with 50 generations.

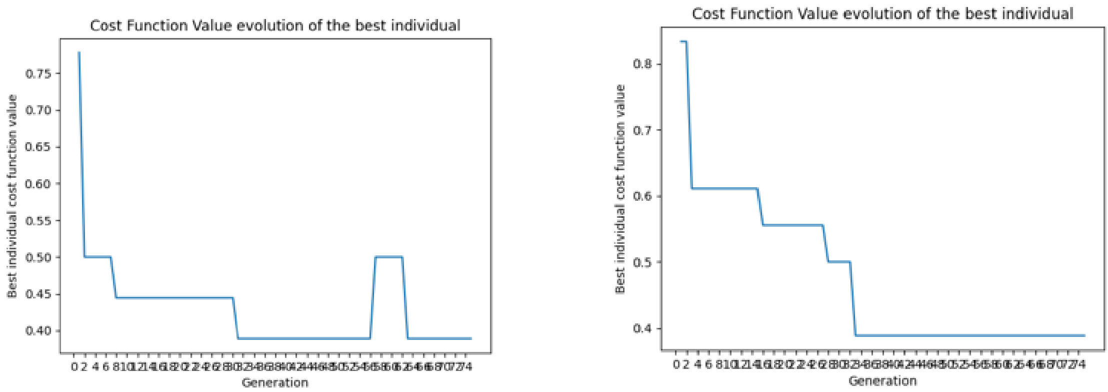


(a) Using roulette selection method (b) Using tournament selection method

Figure 4.2: Test with 50 generations

75 Generations

The third test has 75 generations and results are presented on Figure 4.3(a) and Figure 4.3(b)



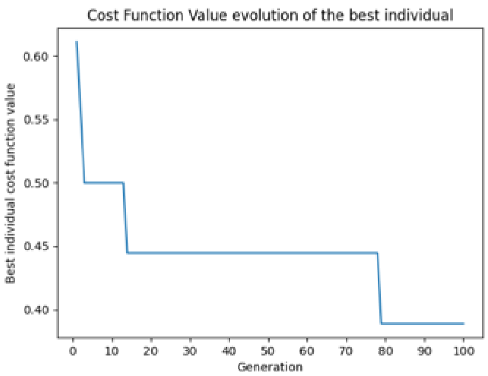
(a) Using roulette selection method (b) Using tournament selection method

Figure 4.3: Test with 75 generations

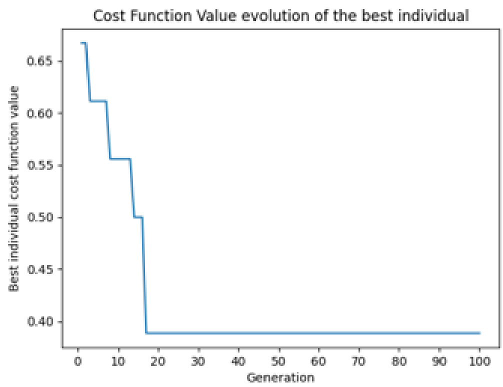
In this test, analyzing but Figure 4.3(a) and (b) both reach a final cost function value of 0.39. In this case, test (a) reaches this value before test (b), on the 31st and 33rd generation respectively. Observing test (a) there is a degradation of the value of the cost function on the 58th. This is caused by a mutation on the best individual, making its cost function value higher in comparison to the previous generation.

100 Generations

The fourth test varying the number of generations has 100 generations and results are presented on Figure 4.4(a) and Figure 4.4(b)



(a) Using roulette selection method



(b) Using tournament selection method

Figure 4.4: Test with 100 generations

In this case and starting by analyzing test (a) the final cost function value of the best individual for this test is 0.39. This value was achieved at the 79$^{\text{th}}$ generation and after three optimizations. For the test (b) the same cost function value was achieved but in this test it only took 17 generations, making the tournament selection a better selection method for this test.

4.1.2 Population size

For the second type of tests, instead of a variable number of generations, this value will be fixed as represented on Table 4.4. The size of each population will now be a variable parameter having four different values.

Table 4.4: Fixed values for variable population size

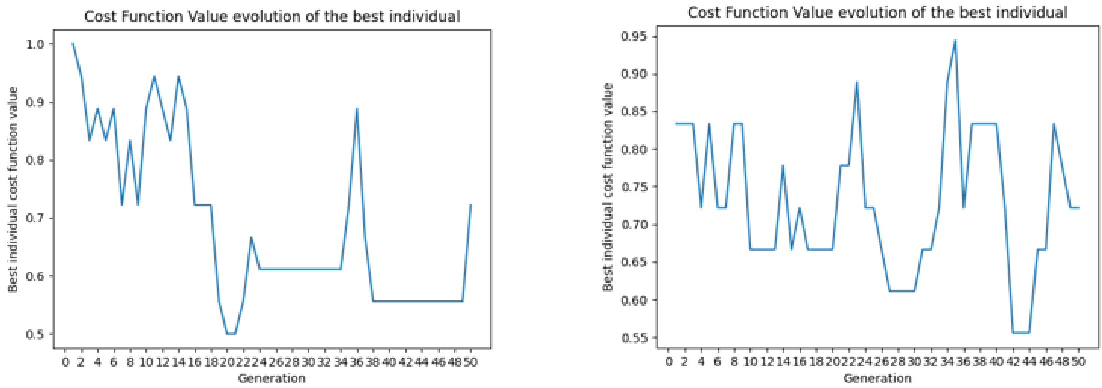
Parameter	Values
Number of generations	50
Crossover probability	0.8
Mutation probability	0.4

Values on Table 4.4 were randomly chosen, and don't follow a specific criteria.

Population size 25

The first value used for the population size will be 25 and the results are shown on Figure 4.5(a) and Figure 4.5(b).

With 0.4 as a value for mutation probability, is expected for the tests (a) and (b) to have some degradation. On Figure 4.5(a) the cost function value of the best individual



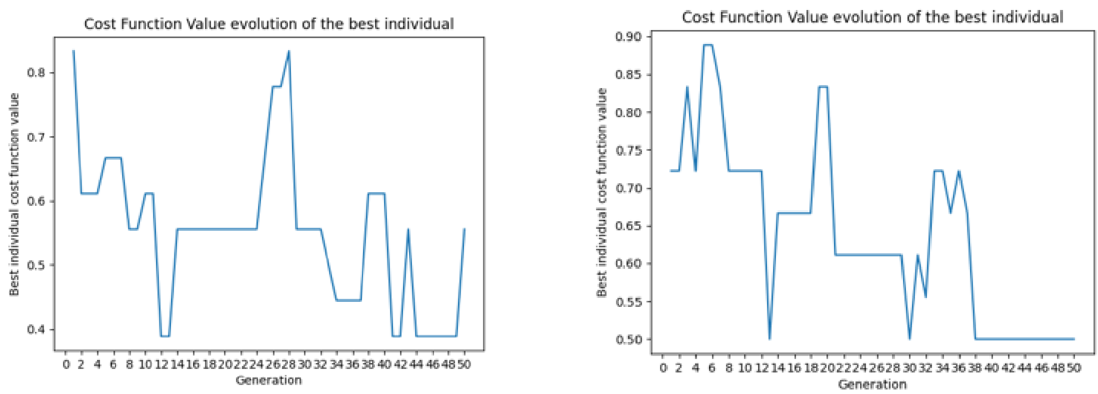
(a) Using roulette selection method (b) Using tournament selection method

Figure 4.5: Test with population size of 25

changes its value times 25 times and achieving a final value of 0.72 at the 50th generation. For Figure 4.5(b) the final value achieved is the same but in this case in the 49th generation. Both of this values are higher than the previous tests shown in section 4.1.1.

Population size 50

Increasing the population size, the following tests represented on Figure 4.6(a) and Figure 4.6(b) have a size of 50 individuals.



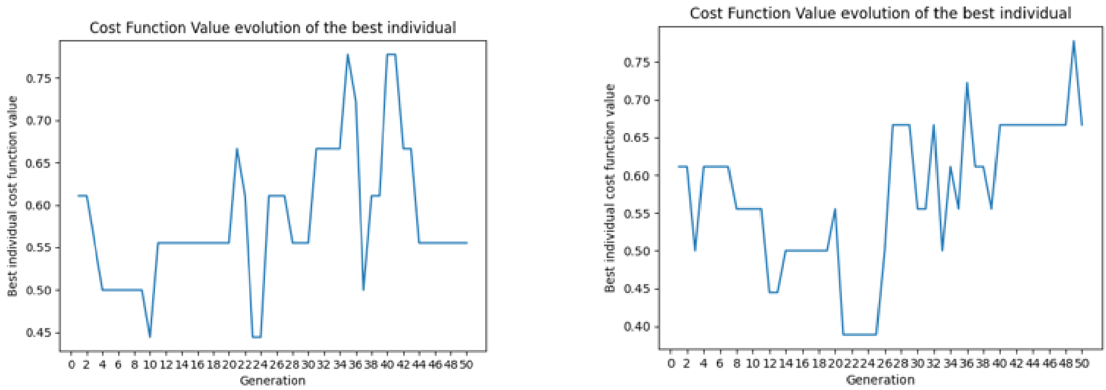
(a) Using roulette selection method (b) Using tournament selection method

Figure 4.6: Test with population size of 50

As observed on Figure 4.5 a high value of mutation probability will cause some degradation on the cost function value of the best individual. This also occurs on Figure 4.5(a) and Figure 4.5(b). Analyzing (a), the final value for the best individual cost function is 0.56. This value has seven degradation during the test. On test (b) the cost function of the best individual also degrades seven times, finishing with a value of 0.5.

Population size 75

The third test increasing the population size has a size value of 75 individuals and the results are shown on Figure 4.7(a) and Figure 4.7(b)



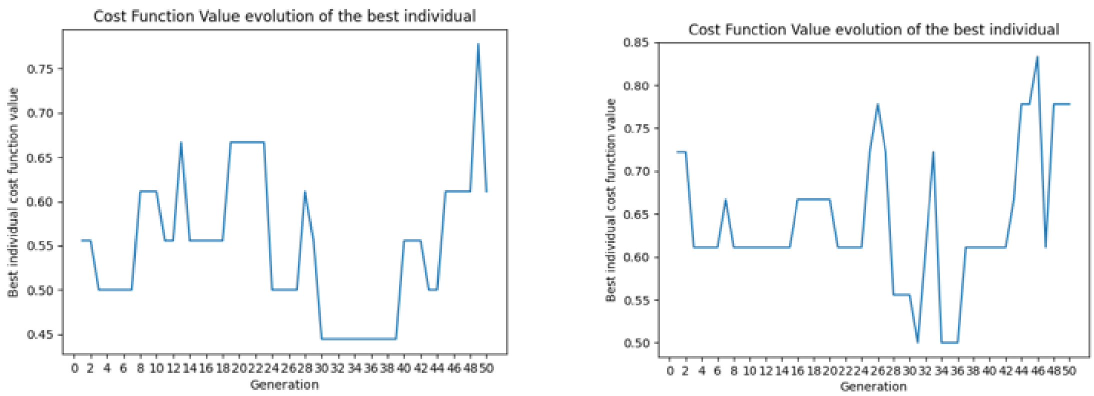
(a) Using roulette selection method (b) Using tournament selection method

Figure 4.7: Test with population size of 75

On Figure 4.7(a) and Figure 4.7(b), test (a) reaches a value of 0.56 changing its value 17 times stabilizing on the 45th generation. For test (b) the final value is 0.67 and this value was achieved after 19 changes.

Population size 100

A fourth and last test varying the population size was made using a value of 100 individuals, the results are represented on Figure 4.8.



(a) Using roulette selection method (b) Using tournament selection method

Figure 4.8: Test with population size of 100

As observed in the previous tests, the outcomes presented in both Figure 4.8(a) and Figure 4.8(b) are notably impacted by the mutation probability. In the case of test (a), it attains a value of 0.61, undergoing 15 alterations, with the final change occurring within the last generation. On the other hand, for test (b), the concluding value is 0.78, reached by the 48th generation and following 19 adjustments.

4.1.3 Crossover probability

Using a variable crossover probability and the other three parameters as fixed values as shown on Table 4.5, some tests were made, and the results will be presented.

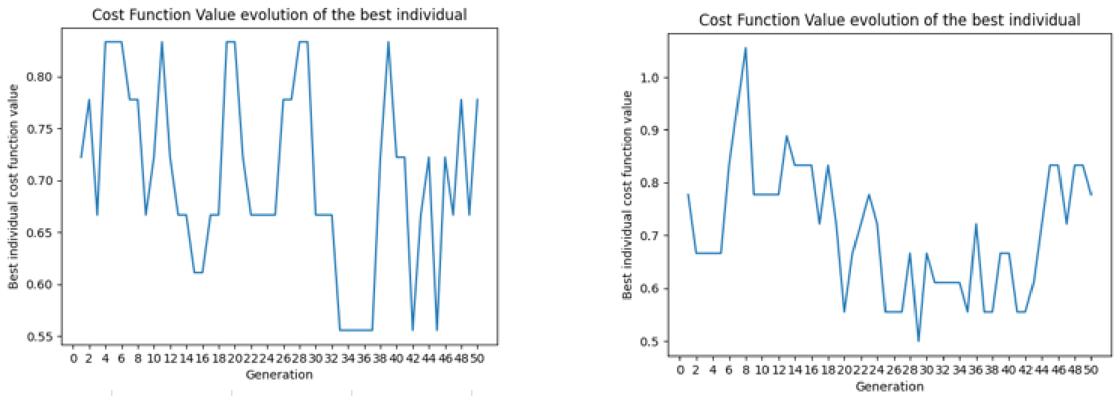
Table 4.5: Fixed values for variable crossover probability

Parameter	Values
Number of generations	50
Population size	50
Mutation probability	0.8

Values on Table 4.5 were randomly chosen, and don't follow a specific criteria.

Crossover probability 0.2

The first value assigned to the crossover probability is 0.2 and the results are presented on Figure 4.9.



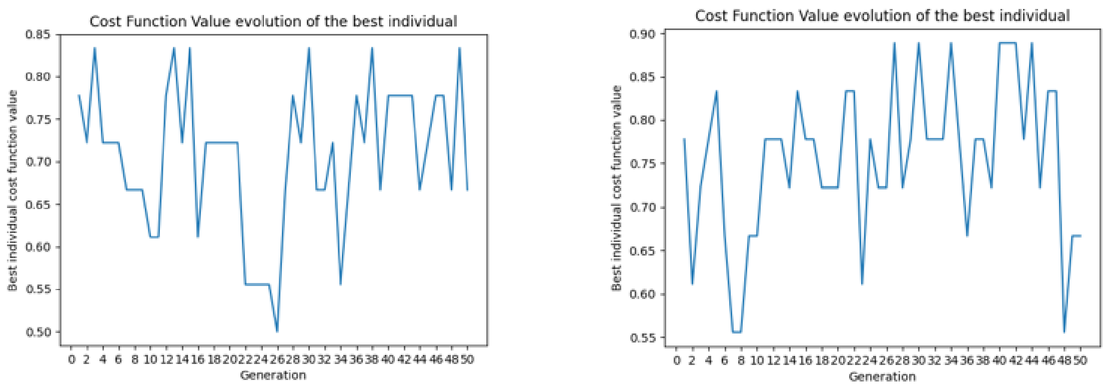
(a) Using roulette selection method (b) Using tournament selection method

Figure 4.9: Test with 0.2 crossover probability

Observing test (a) the final value obtained is 0.78. As observed this value suffers both optimization and degradation throughout the entire test changing its value 30 times. As for test (b) the final result is the same as (a), 0.78. This test suffers an inferior number of changes, changing 27 times.

Crossover probability 0.5

Increasing the crossover probability to 0.5 the results of this test can be observed on Figure 4.10.



(a) Using roulette selection method (b) Using tournament selection method

Figure 4.10: Test with 0.5 crossover probability

These two tests, Figure 4.10(a) and Figure 4.10(b) show similar results to those previously presented on Figure 4.9 with both tests finishing with a cost value of 0.67. Comparing to the previous test, both (a) and (b) have higher genetic reuse. This implies that instead of generating entirely new individuals each generation, the process involves utilizing 50% of the individuals selected from the current generation to produce the new generation's individuals.

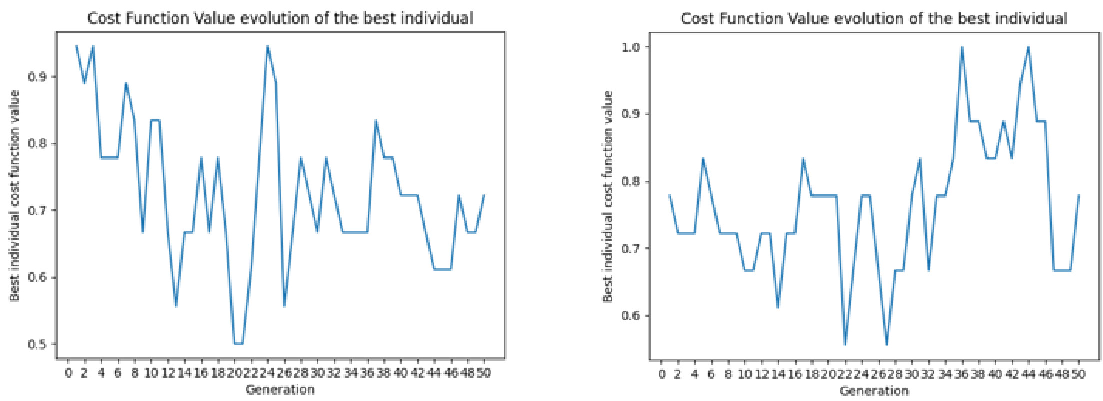
Crossover probability 0.8

The final tests varying the crossover probability has a value of 0.8, with the highest value tested the results are presented on Figure 4.11.

As previously mentioned, an increase in the crossover probability leads to a greater degree of genetic reuse. In this context, 80% of the individuals selected from the previous generation will participate in the crossover process to transition to the next generation. On Figure 4.11(a) the final value reached for the cost function is 0.72. On Figure 4.11(b) this value is slightly higher, 0.78.

4.1.4 Mutation probability

The last parameter to be subjected to variation is the mutation probability. As previously examined and observed, this parameter significantly influences the quality and



(a) Using roulette selection method (b) Using tournament selection method

Figure 4.11: Test with 0.8 crossover probability

outcome of each test. A greater probability corresponds to a higher likelihood of the cost function value undergoing fluctuations, which can lead to either optimization or degradation. The fixed values for the rest of the parameters are present on Table 4.6.

Table 4.6: Fixed values for variable mutation probability

Parameter	Values
Number of generations	100
Population size	50
Crossover probability	0.2

Values on Table 4.6 were randomly chosen, and don't follow a specific criteria.

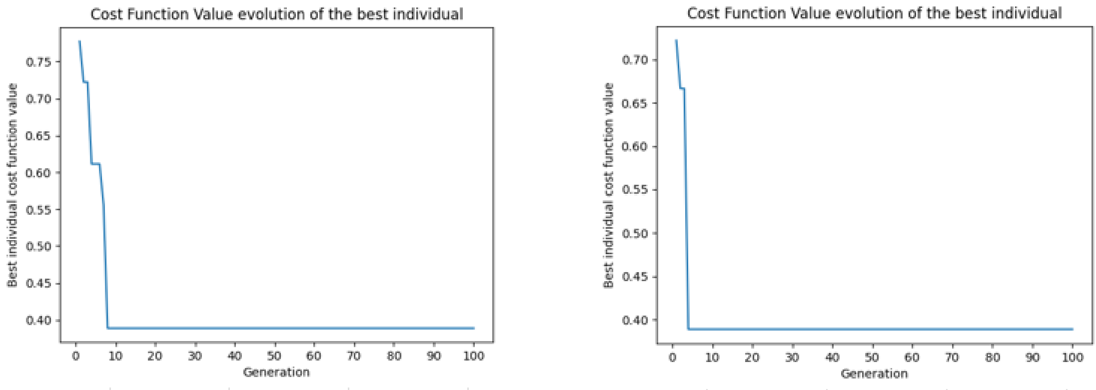
Mutation probability 0.01

The first value for the mutation probability is 0.01, this is the lowest value tested and the results are shown on Figure 4.12(a) and Figure 4.12(b).

With the lowest value for the mutation probability, both Figure 4.12(a) and Figure 4.12(b) show a low number of variations. With test (a) varying four times before reaching a final value of 0.39 on the 9th generation. On test (b) this variation occurs twice before achieving the same value of 0.39, but this time on the 4th generation.

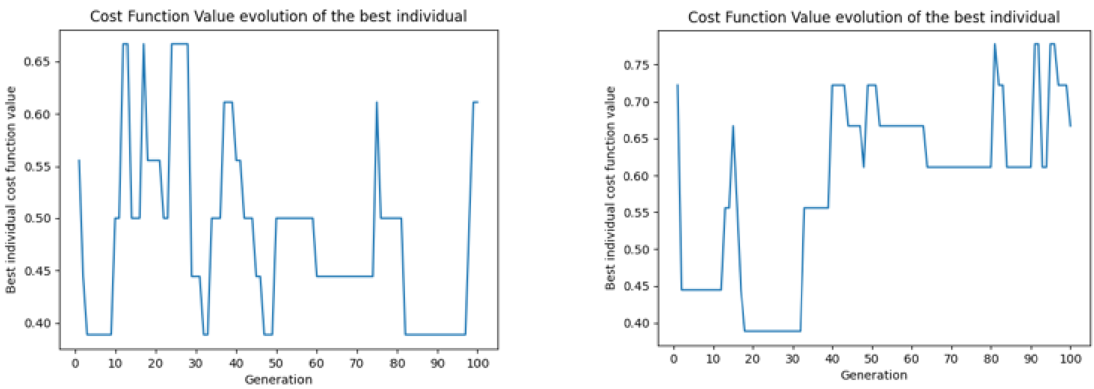
Mutation probability 0.2

By increasing the mutation probability, the new value assigned will be 0.2, and the results are displayed in Figure 4.13(a) and Figure 4.13(b).



(a) Using roulette selection method (b) Using tournament selection method

Figure 4.12: Test with 0.01 mutation probability



(a) Using roulette selection method (b) Using tournament selection method

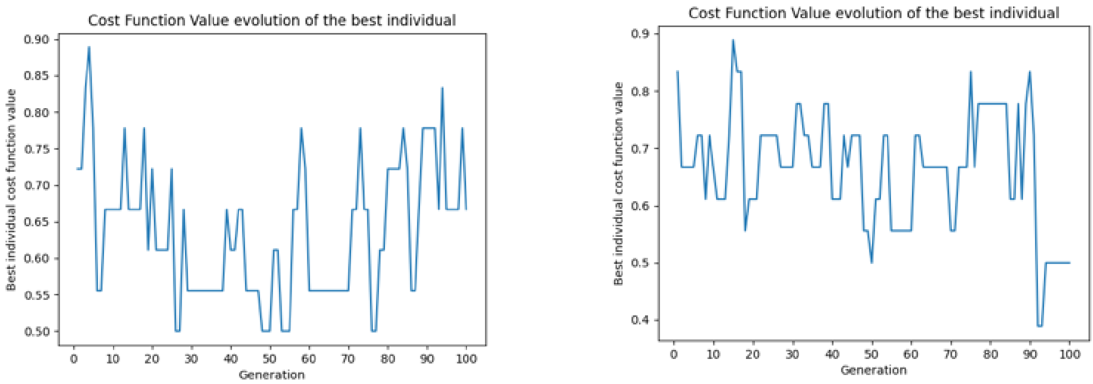
Figure 4.13: Test with 0.2 mutation probability

With an increased mutation probability, an increased number of variations is expected, and in Figure 4.13(a) and Figure 4.13(b) this can be observed. In both tests, this variation induced by an increased mutation probability led to a deterioration in the cost function value of the best individual. With test (a) finishing with a cost function value of 0.61 and test (b) with 0.67.

Mutation probability 0.4

In the third test, a mutation probability of 0.4 was set, and the outcomes are illustrated in Figure 4.14(a) and Figure 4.14(b).

In Figure 4.14(a) and Figure 4.14(b) it's possible to observe the results of a test with a mutation probability of 0.4, this caused both tests to have a high number of variation both optimizing and degrading the cost function value. Test (a) finishes with a value of 0.67 and test (b) with a value of 0.5.

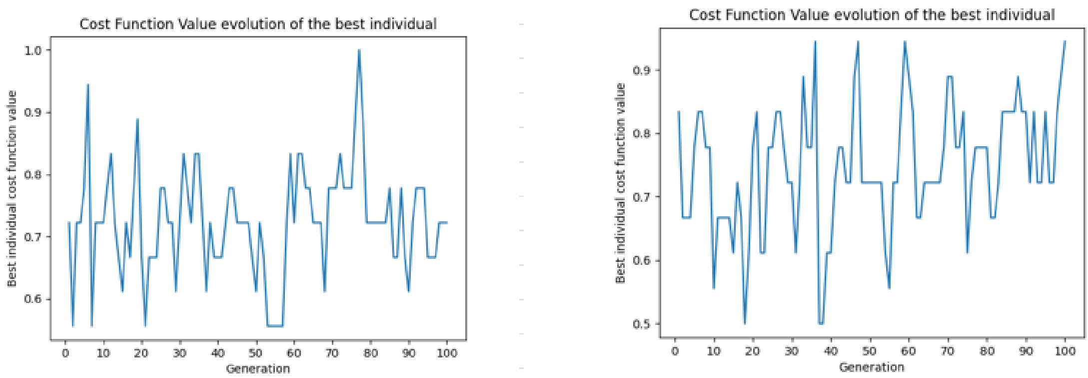


(a) Using roulette selection method (b) Using tournament selection method

Figure 4.14: Test with 0.4 mutation probability

Mutation probability 0.8

The fourth value chosen to vary the mutation probability is 0.8, and the final results are shown in Figure 4.15(a) and Figure 4.15(b).



(a) Using roulette selection method (b) Using tournament selection method

Figure 4.15: Test with 0.8 mutation probability

In this concluding test, utilizing the highest mutation probability value, both tests (a) and (b) reveal the most notable number of variations. In test (a), the cost function value remains equal from start to conclusion at 0.72. Conversely, in test (b), the cost function of the best individual experiences degradation, declining to 0.94 from its initial value.

4.2 Comparison with current real values

In order to ensure the quality of both tests and results previously presented, it is necessary to make a comparison with real values. In this case, with the current Cascais railway values. On Figure 4.16 the current line solution is presented. having four active stops (Alcântara-Mar, Cruz Quebrada, PK16.8 and São Pedro do Estoril).

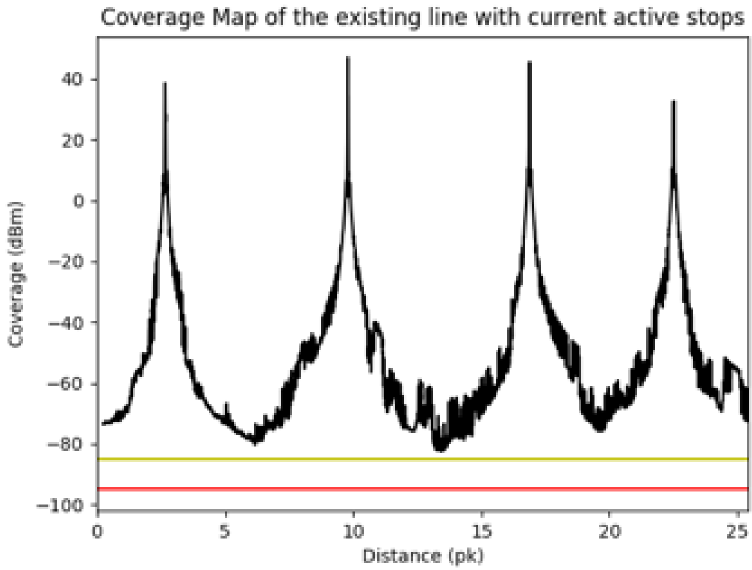
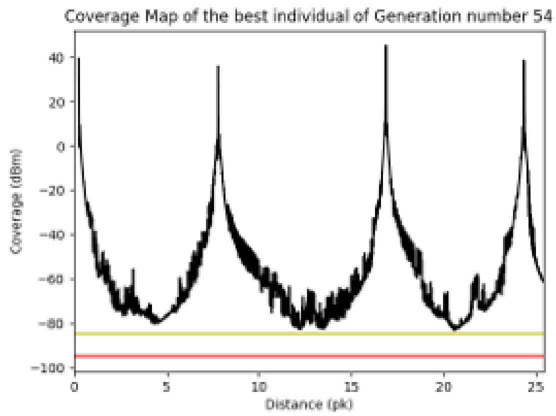


Figure 4.16: Coverage Map of the existing line with current active stops

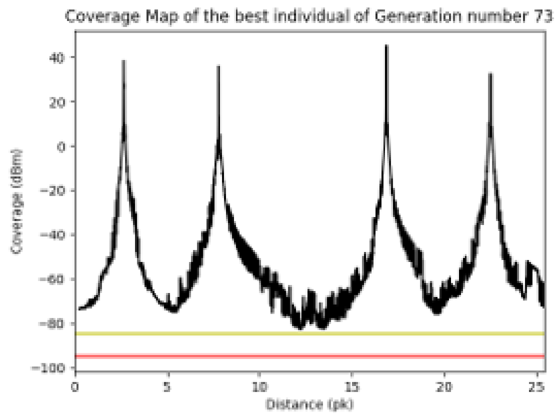
Observing Figure 4.16 and matching the name of each stop with its type shown on Table 4.1, this solution has two active stations, one active halt and one active anchor. Adding to the number of active stops and their type, this solution has the entire line coverage above both minimum coverage and low level signal. Calculating the cost function value using equation (3.3) the value for this solution is 0.44.

Comparing this solution with the results obtained on Figure 4.1(b), 4.2(b), 4.3(a) and (b), 4.4(a) and (b), and 4.12(b) it is possible to verify that these test have better solution than the one present on Figure 4.16. With this, the best solution for this problem using this algorithm and these parameters and weights presented on Table 3.3, can be represented by two different solutions with the same cost function value of approximately 0.39, as shown on Figure 4.17. Both Figure 4.17(a) and (b) have four active stops, divided in one anchor and three stations. The difference between Figure 4.17(a) and (b) is the location of each active station, with Figure 4.17(a) having Cais do Sodré, Algés and Estoril as the active stations and Figure 4.17(b) Alcântara-Mar, Algés and São Pedro do Estoril.

With solutions with the same cost function value, it is necessary to distinguish them and chose the best between them. Unfortunately, for this specific algorithm and with this specific parameters, it's not possible to determine which one is the best. So for the case shown on Figure 4.17 both solutions will be considered as optimal.



(a) Final solution 1



(b) Final solution 2

Figure 4.17: Two different final solutions to this case study

5

Conclusions

5.1 Work conclusions

The present thesis' objective is to understand whether, using a genetic algorithm, it would be possible to present an optimized solution for the distribution of antennas at railway stops. To carry out this analysis, an optimization algorithm based on genetic algorithms was developed, to make it possible to carry out several tests. Six steps were designed for each individual to go through each phase of the algorithm, these steps include the creation of the individual, the selection method, the crossover method, mutation, and elitism. This algorithm was designed so that it was possible to have the greatest variety of unique individuals. With the algorithm finished, it was necessary to choose a line and for this specific case, the Cascais line was used as a test subject.

As previously mentioned multiple tests were carried out with a number of parameters with a fixed value and four variable parameters, number of generations, generation size, crossover probability and mutation probability. Varying the number of generation and size of each generation with the values of 25, 50, 75 and 100. This will lead to an increase in the number of individuals per generation and per test, resulting in broad genetic diversity. With the variation of the crossover probability, new individuals will be created, increasing even further the genetic diversity. The same will occur with the probability of mutation, the higher the probability, the more mutations there will be. These mutations may not always have a positive impact on the tests, since a high probability may affect the best individual, causing a degradation in the value of its cost function. With the best individual degraded and the effect of the elitism a new individual with a different cost function value will carry onto the next generation possibly causing a degradation in the evolution of the best individual of each generation. This was observed in the tests presented in this document, tests with a mutation probability higher than 0.01 show worst results than tests with a low mutation probability. This conclusion can be made through an analysis of the figures because the final result is always higher than the optimal cost function value obtained on this study case (0.39).

Just by carrying out tests, it would not be possible to conclude whether this algorithm could create a better solution than the current solution, so it was necessary to make a comparison between both cases. Finally, the conclusions that were possible to draw were that the algorithm actually produced one or more tests whose cost function value was lower than that of the real solution with the parameters used.

5.2 Future work

For future work, it is proposed to increase variables such as the height and type of antennas, it is also proposed to ensure that all individuals have a unique genetic code, this point was not possible to achieve due to computational complexity.

Since only one line was tested, it is proposed that different railway lines be considered with the aim of testing and proving the quality of the results obtained.

References

[1] Fai Lam, *Why railways are already on the move to frmcs*, 2023. [Online]. Available: https://www.nokia.com/blog/why-railways-are-already-on-the-move-to-frmcs/ (visited on 09/28/2023).

[2] Nokia, "Future railway mobile communication system (frmcs)", 2023. [Online]. Available: https://www.nokia.com/networks/industries/railways/frmcs/?did=D00000006078&gclid=CjwKCAjwyNSoBhA9EiwA5aYlb2Ba69GBD2kIXaWxMy27Z0-z9AFaRDrbPwEC0ByECQv5NJWaB-KNvBoCFxwQAvD_BwE (visited on 09/28/2023).

[3] Issam El Naqa and Martin J. Murphy, "What is machine learning?", *Springer International Publishing*, pages 3–11, 2015.

[4] M.H. Kutner, C. Nachtsheim, and J. Neter, "Applied linear regression models", *McGraw-Hill/Irwin*, 2004.

[5] André da Silva Duarte, "Modelação de potência recebida e dimensionamento de margens de ligações micro-ondas ponto-a-ponto", *Instituto Superior de Engenharia de Lisboa*, 2021.

[6] "Root mean square error (rmse) or mean absolute error (mae)", *Geoscientific model development discussions*, vol. 7, no. 1, pages 1525–1534, 2014.

[7] Arnaud De Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi, "Using the mean absolute percentage error for regression models.", 2015.

[8] Nico JD Nagelkerke *et al.*, "A note on a general definition of the coefficient of determination", *Oxford University Press*, vol. 78, no. 3, pages 691–692, 1991.

[9] Ana Rita Beire, Helder Jorge Pinheiro Pita, and Nuno Cota, "Optimizing propagation models on railway communications using genetic algorithms", *Elsevier*, 2014.

[10] Tiago Correia, "Estimação de cobertura rádio em gsm-r através de redes neuronais", *Instituto Superior de Engenharia de Lisboa*, 2014.

[11] M. Hata, "Empirical formula for propagation loss in land mobile radio services", *IEEE Transactions on Vehicular Technology*, vol. VT-29, no. 3, 1980.

[12] Yoshihisa Okumura, "Field strength and its variability in vhf and uhf land-mobile radio service", *Review of the Electrical communication Laboratory*, vol. 16, no. 9, 1968.

[13] Moamen Alnatoor, Mohammed Omari, and Mohammed Kaddi, "Path loss models for cellular mobile networks using artificial intelligence technologies in different environments", *Applied Sciences*, vol. 12, no. 24, 2022.

[14] Eugenio Muscinelli, Swapnil Sadashiv Shinde, and Daniele Tarchi, "Overview of distributed machine learning techniques for 6g networks", *Algorithms*, vol. 15, no. 6, 2022.

[15] Hicham Megnafi, "Frequency plan optimization based on genetic algorithms for cellular networks", *Journal of Communications Software and Systems*, vol. 16, no. 3, pages 217–223, 2020.

[16] Oliver Kramer, "Genetic algorithms", *Springer International Publishing*, pages 11–19, 2017.

[17] Jonathan Shapiro, "Genetic algorithms in machine learning", *Springer Berlin Heidelberg*, pages 146–168, 2001.

[18] Seyedali Mirjalili, "Genetic algorithm", *Springer International Publishing*, pages 43–55, 2019.

[19] Rakesh Kumar and Jyotishree, "Blending roulette wheel selection & rank selection in genetic algorithms", *International Journal of Machine Learning and Computing*, page 365, Jan. 2012.

[20] Kumara Sastry, David Goldberg, and Graham Kendall, "Genetic algorithms", *Springer US*, pages 193–212, 2005.