ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

INGENIERO INDUSTRIAL

TRANSMISSION EXPANSION PLANNING USING A GENETIC ALGORITHM

Autor:

Cristina Duro Guillén

Directores:

Sara Lumbreras Sancho

Andrés Ramos Galán

Madrid, Mayo 2014

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN ACCESO *ABIERTO* (*RESTRINGIDO*) DE DOCUMENTACIÓN

1º. Declaración de la autoría y acreditación de la misma.

El autor D. Cristina Duro Guillén, como alumna de la UNIVERSIDAD PONTIFICIA COMILLAS (COMILLAS), **DECLARA**

que es el titular de los derechos de propiedad intelectual, objeto de la presente cesión, en relación con la obra "Transmission Expansion Planning using a Genetic Algorithm" [1], que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual como titular único o cotitular de la obra.

En caso de ser cotitular, el autor (firmante) declara asimismo que cuenta con el consentimiento de los restantes titulares para hacer la presente cesión. En caso de previa cesión a terceros de derechos de explotación de la obra, el autor declara que tiene la oportuna autorización de dichos titulares de derechos a los fines de esta cesión o bien que retiene la facultad de ceder estos derechos en la forma prevista en la presente cesión y así lo acredita.

2º. Objeto y fines de la cesión.

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad y hacer posible su utilización de *forma libre y gratuita* (*con las limitaciones que más adelante se detallan*) por todos los usuarios del repositorio y del portal e-ciencia, el autor **CEDE** a la Universidad Pontificia Comillas de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución, de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra (a) del apartado siguiente.

3º. Condiciones de la cesión.

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia, el repositorio institucional podrá:

[1] Proyecto Fin de Carrera.

(a) Transformarla para adaptarla a cualquier tecnología susceptible de incorporarla a internet; realizar adaptaciones para hacer posible la utilización de la obra en formatos electrónicos, así como incorporar metadatos para realizar el registro de la obra e incorporar "marcas de agua" o cualquier otro sistema de seguridad o de protección.

(b) Reproducirla en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato. .

(c) Comunicarla y ponerla a disposición del público a través de un archivo abierto institucional, accesible de modo libre y gratuito a través de internet.[2]

(d) Distribuir copias electrónicas de la obra a los usuarios en un soporte digital. [3]

4º. Derechos del autor.

El autor, en tanto que titular de una obra que cede con carácter no exclusivo a la Universidad por medio de su registro en el Repositorio Institucional tiene derecho a:

a) A que la Universidad identifique claramente su nombre como el autor o propietario de los derechos del documento.

b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.

c) Solicitar la retirada de la obra del repositorio por causa justificada. A tal fin deberá ponerse en contacto con el vicerrector/a de investigación (curiarte@rec.upcomillas.es).

d) Autorizar expresamente a COMILLAS para, en su caso, realizar los trámites necesarios para la obtención del ISBN.

---

[2] En el supuesto de que el autor opte por el acceso restringido, este apartado quedaría redactado en los siguientes términos:

(c) Comunicarla y ponerla a disposición del público a través de un archivo institucional, accesible de modo restringido, en los términos previstos en el Reglamento del Repositorio Institucional

[3] En el supuesto de que el autor opte por el acceso restringido, este apartado quedaría eliminado.

d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5º. Deberes del autor.

El autor se compromete a:

a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.

b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.

c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.

d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6º. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

a) Deberes del repositorio Institucional:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.

- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.

- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.

b) Derechos que se reserva el Repositorio institucional respecto de las obras en él registradas:

- retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 28 de Mayo de 2014.

ACEPTA

Fdo……………………………………………………

Proyecto realizado por el alumno/a:

Cristina Duro Guillén

Fdo.: …………………… Fecha: 28/ Mayo/ 2014

Autorizada la entrega del proyecto cuya información no es de carácter confidencial

LOS DIRECTORES DEL PROYECTO

Sara Lumbreras Sancho

Fdo.: …………………… Fecha: 28/ Mayo/ 2014

Andrés Ramos Galán

Fdo.: …………………… Fecha: 28/ Mayo/ 2014

Vº Bº del Coordinador de Proyectos

Fernando de Cuadra García

Fdo.: …………………… Fecha: 28/ Mayo/ 2014

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

INGENIERO INDUSTRIAL

TRANSMISSION EXPANSION PLANNING USING A GENETIC ALGORITHM

Autor:

Cristina Duro Guillén

Directores:

Sara Lumbreras Sancho

Andrés Ramos Galán

Madrid, Mayo 2014

Agradecimientos

A mis padres.

A Carlos por su ánimo y a Carlota por acompañarme a lo largo de toda esta etapa de estudiante.

A todos aquellos compañeros del IIT que me han acogido y tendido su mano.

A Pablo por su interés y a los que han sido mis compañeros de planta en el IIT alegrándome cada día con una sonrisa.

Y sobre todo a Sara por guiarme en cada paso y contagiarme su entusiasmo y a Andrés por brindarme esta gran oportunidad.

RESUMEN

La red de transporte es la estructura básica responsable del suministro de electricidad a larga distancia. Interconecta los centros de generación con las subestaciones desde donde la red de distribución alimenta los núcleos de demanda. Por tanto la red de transporte tiene un papel esencial en el sistema de energía eléctrica.

Introducción

La Planificación de la Expansión de la red de Transporte (TEP) aborda el problema estocástico que determina las líneas óptimas y otros equipos a añadir en la red eléctrica para satisfacer la demanda estimada a largo plazo. El objetivo es definir cuándo y dónde deben instalar nuevos circuitos con un coste mínimo sujeto a un conjunto de restricciones.

La penetración de las renovables en la generación y la liberalización del mercado eléctrico introducen complicaciones. Por esto, TEP debe incorporar las incertidumbres inherentes a la generación y expansión de la red y anticipar los nuevos desarrollos.

Motivación

Los intentos de solucionar la cuestión de la expansión del transporte se han basado en modelos simplificados debido a la complejidad del problema. Para resolver TEP se emplean dos grandes grupos de métodos: clásicos y no clásicos. La Descomposición de Benders (BD) es uno de los métodos clásicos más importantes aplicados en el campo de la Optimización Estocástica. Permite resolver grandes problemas dividiéndolos en un problema maestro que propone nuevas soluciones y un subproblema que las evalúa y devuelve al maestro para las siguientes propuestas. Este proceso se resuelve iterativamente hasta convergencia. La resolución del maestro es lenta en el problema de TEP, debido a una cuestión de tamaño, condiciones de integrabilidad o por la adición de un gran número de cortes que complican la resolución del problema maestro [LUMB12].

La falta de adecuadas herramientas computacionales hace que investigar metaheurísticos sea atractivo para la generación de propuestas para TEP. Un metaheurístico es una técnica de resolución que proporciona una solución suficientemente buena para un problema de optimización, especialmente con capacidad de computación limitada. El Algoritmo Genético (GA) es un metaheurístico extendido en el campo de TEP.

Metodología

El primer objetivo de este proyecto es diseñar un GA adecuado para resolver TEP. El GA genera propuestas que son evaluadas por el subproblema, y usa la información de costes proporcionada por el subproblema para definir la función fitness. Después se aplican los operadores genéticos tradicionales sobre la población: cruce y mutación. Otros operadores añadidos son la conservación el mejor individuo (reina) y la introducción de inmigrantes. Para realizar el ajuste de los parámetros genéticos y así poder obtener mejores soluciones para TEP, el GA es incluido en TEPES (Transmission and Expansion Planning for an Electric System, un modelo desarrollado en el IIT que lleva a cabo TEP usando BD) y su correcto funcionamiento es comprobado en varios casos de estudio: 9BUS, 46BUS y un caso basado en el sistema español. Otro objetivo del proyecto es obtener soluciones óptimas para TEP. Finalmente, el GA es contrastado con la resolución con BD.

Resultados

Para resolver pequeños problemas el tamaño de población debe ser similar al número de nudos que configuran la red de transporte. El caso de estudio basado en el sistema español está compuesto por 1084 nudos y 294 plantas de generación. La red de transporte existente está formada por 1505 líneas y transformadores y la red candidata por 153 líneas. Se han construido 3 casos a partir de éste, considerando diferente número de líneas candidatas. Para estos 3 casos el mejor tamaño de la población es de 10 individuos. La tasa de mutación más adecuada es 1%. Las ejecuciones llevadas a cabo con el GA tardan menos de 20 minutos en alcanzar soluciones un error inferior al 0,3%. La introducción de inmigrantes no resulta ventajosa para obtener soluciones. La mejor forma de generar la población inicial es con una probabilidad de éxito en la distribución binomial ligeramente superior a la proporción de líneas instaladas en el plan óptimo.

BD alcanza la solución óptima o una muy similar con importantes ahorros de tiempo en los casos de estudio 1 y 2. En el caso más grande la resolución con BD queda se estancada con un error del 0,4% y el GA alcanza una solución con solo un error del 0.05% en un tiempo razonable.

Conclusiones

BD resuelve el subproblema una vez por iteración, pero el GA lo resuelve tantas veces como individuos forman la población, por lo que el número de individuos está directamente relacionado con el tiempo de ejecución. Para grandes problemas es adecuado usar un tamaño de población suficientemente bajo para reducir el tiempo de ejecución.

Los clusters de soluciones con costes similares están relacionados con la tasa de mutación. Una mutación suficientemente baja actúa como un filtro de clusters, con la población perdiendo diversidad rápidamente.

Una población inicial con buenas cualidades ayuda al GA a alcanzar mejores soluciones, pero la introducción de inmigrantes generados de la misma manera que la población inicial no aporta beneficios al algoritmo.

SUMMARY

The transmission grid is the basic infrastructure responsible for long-distance transferring electricity. It connects the generation plants with the substations from where the demand nucleuses are feed by the distribution grid. So transmission plays an essential role in the power systems of the future.

Introduction

Transmission and Expansion Planning (TEP) is a combinatorial stochastic problem that determines the optimal lines and other equipment to be added to a power network for supplying the forecasted demand in a long-term horizon. Its objective is to define when and where new circuits should be installed at minimum cost subject to a set of constraints.

The introduction of renewable generation and the liberalization of generation markets introduce further complications. Thus, TEP must incorporate the uncertainties inherent to generation expansion and anticipate new developments.

Motivation

Attempts to solve the transmission expansion problem have been made based on simplified models because of the complexity of the problem. Two types of methods have been used to solve TEP problem: classical and non-classical. Benders' decomposition (BD) is one of the main classical methods applied in the Stochastic Optimization domain. It allows solving large problems by dividing them into a master problem which proposes new solutions and a subproblem which evaluates them and sends feedback to the master for the next proposals. This process is solved iteratively until convergence. Master problem resolutions are slow. This can be a result of size, integrality conditions or the addition of a large number of cuts that complicate master problem resolution [LUMB12].

The lack of adequate computational tools becomes metaheuristics research an attractive issue for generating proposals for TEP. A metaheuristic is a solution technique that provides a sufficiently good solution to an optimization problem, especially with limited computation capacity. A extended metaheuristic in the TEP domain is the Genetic Algorithm (GA).

Methodology

The first objective of this project is to design a suitable GA for solving TEP. The GA generates proposals that are evaluated by the subproblem, and uses the cost information provided by the subproblem to define the fitness function. Then the traditional genetic operators are applied over the population: crossover and mutation. Other operators are added: saving the best individual and introducing immigrant individuals. In order to adjust the genetic parameters to obtain better results for TEP, the GA is included in Transmission and Expansion Planning for an Electric System (a model developed at the IIT that performs TEP using BD) and it is tested in several case studies: 9BUS, 46BUS and a Spain-based case. Another objective of this project is to obtain optimal solutions for TEP. Finally the GA is contrasted with BD resolution.

Results

For solving small problems the best population size is similar to the number of nodes that configured the transmission network. The Spain-based case study is composed of 1084 nodes and 294 power plants. The existing transmission network is configured by 1505 lines and transformers. The candidate transmission network consisted of 153 power lines. Three case studies have been constructed from the Spain-based case, considering different number of candidate lines. For the three case studies the best population size is 10 individuals. The most suitable mutation probability is 1%. The GA executions spent less than 20 minutes for reaching solutions with less than 0.3% error. The introduction of immigrants does not provide any advantage for obtaining solutions. The best way of generating the initial population is with a success probability in the binomial distribution slightly higher to the proportion of installed lines in the optimal plan.

BD reaches the best solution or a very similar one with important time savings in case studies 1 and 2. In the biggest case study BD gets stagnant with a 0.4% error and GA reaches a solution with only 0.05% error in a reasonable time.

Conclusions

BD solves the subproblem once per iteration, but GA solves it as many times as individuals that are in the population, so the number of individuals is directly related with the execution time. For large problems is suitable to use a population size low enough for reducing the execution time.

The mutation rate is related with the number of clusters. A mutation rate low enough acts as a filter of clusters. Higher mutation rates provide a faster response of the GA at the beginning of the iterative process, but contribute to stagnancy in the last iterations.

A good quality initial generation helps the GA for reaching better solutions, but the introduction of immigrants generated in the same way as the initial population is a useless operator for solving TEP.

Contents

Figures

Tables

1. INTRODUCTION

Transmission Expansion Planning (TEP) plays an important role in the power systems due to the electric energy industry restructuring that is being implemented worldwide. The introduction of renewable generation far from the existing transmission system and the liberalization of the markets are introducing further complications in TEP.

In nowadays globalized markets, since the competitiveness is dreadfully increasing supply chain design has been gaining attention. Companies have to, at least, keeps the same customer service level, while the market's competitiveness forces them to reduce the overall costs to maintain their profit margins. Transportation network design provides a remarkable potential to reduce the overall costs and also to improve the service level.

Consequently, it is necessary to develop methods to generate future power network proposals that contribute to an efficient transmission of the electric power, subject to electric, economic, social, environmental constraints and incorporating the uncertainties inherent to generation.

The transportation problem is considered as a NP-hard problem. The objective is to find the combination of candidate power lines that minimizes the total variable and fixed costs while satisfying the supply and demand requirements of each origin and destination. The problem has a dynamic nature so the requirements of transmission facilities should be defined over a time within a given horizon.

Because of the inherent complexity of the problem and the lack of adequate computational tools, attempts to solve TEP have been made based on two generic simplified models pertaining to whether the stochastic and dynamic aspects of the problem are considered or not [LUMB14].

Researchers have turned to heuristic algorithms because the methods are constrained by limits on computer time. Optimization approaches based on metaheuristcis have demonstrated the potential of finding high quality solutions.

1.1. State of the Art

1.1.1. *State of the Art: Transmission Expansion Planning*

Transmission Expansion Planning (TEP) is a combinatorial optimization problem. This strategic planning evaluates the future network needs, determining the optimal transmission lines and other equipment to be added to a power network in the long-term horizon. Its objective is to define when and where new circuits should be installed to supply the forecasted demand at minimum cost subject to a set of electrical, economic, financial and environmental constraints and with an adequate level of reliability.

Decisions related to TEP must be taken considering different stages, scenarios and assumptions as many factors have an impact. Attending to the **modeling assumptions** of the TEP problem, it could be classified as stated below.

Traditionally, power system planning has been studied considering the investment cost as the only objective taken into account in the optimization, *monocriterion*. Therefore, other factors are added to investment cost. In a centralized operation context this factors could be the operation cost for the scenarios considered, penalties that avoid solutions where expected loads cannot be satisfied, and the expected energy not supplied and other reliability indices. Social welfare and competition are the main factors considered in a generation market context. The *multicriteria* study includes several variables in the objective function such as environmental impact, flattening of nodal prices, geopolitical risks, financial resources required and renewable generation integration.

Regarding the dynamic complexity of the TEP problem, most studies aim for simplifying it, allowing the next classification. The majority of the research studies solve the problem according to the forecasted future at a particular moment in the time; this is known as *static* planning. The *sequential static* planning considers several future horizons ensuring that intermediate solutions are consistent with the long-term goal. This planning admits a forward approach or a backwards starting with the final year. Due to the difficulty and computational complexity of the *dynamic* issue, it has been applied to small case studies with direct resolution with classical methods. In order to incorporate the dynamic dimension into larger case studies, metaheuristic solution methods have been introduced.

Uncertainties can be classified in random and nonrandom. Random uncertainties arise from repeated deviations of parameters and are explained by a probability distribution and positive and negative observations compensate. Nonrandom uncertainties only happen a single time with no possibility of compensation and cannot be modeled exactly based-on past data. As far as the treatment of uncertainties is concerned the main techniques of resolution can be classified into three categories. *Robust* optimization minimizes the maximum regret focusing on a worst-case scenario analysis. *Fuzzy decision* analysis studies the outcomes of the different scenarios considered and the relevant importance of non-dominated solutions working with the decision maker. The *stochastic* paradigm incorporates random uncertainties directly in the decision process minimizing the expected cost over the average of futures.

In view of market considerations and regulatory implications most regulations use centralized planning: the regulator approves the plans proposed by the TSO (Transmission System Operator) or give transmission licenses. In some cases centralized planning is complemented with mixed planning in collaboration with market agents (i.e. proposals of users approved by the regulator) or coexists with market planning where market agents decide expansion on price signals. Most works carry out centralized TEP with centralized cost-based operation (to minimize operation cost); in a medium-term expansion some references consider centralized TEP with a generation market.

Referring to technical grid modeling options, simple models that only take into account Kirchoff´s first law reduce computational requirements of the optimization. Second Kirchoff´s law requires linearized DC power flows or hybrid models. More sophisticated grid modeling options incorporate the nonlinearities of the AC power flow when evaluating a transmission plan.

A simple **classified of solution techniques** appears below [LUMB14]. Two types of methods have been used to solve TEP problem [LUMB12]: classical and non-classical, such as the metaheuristics. Other non-classical method is the sensitivity analysis, which guides the local search of candidate solutions, or expert systems that apply complex rules from specialist knowledge in larger contexts.

An important range of classical techniques are those based on mathematical programming. Linear programming (LP) assumes important simplifications but it requires low computational effort, which makes it practical when large power systems are regarded. For settling this computational problem, this technique assumes some simplifications, as a DC power flow model, and ignores the discrete nature of investment variables. *Quadratic programming* (QP) has been applied so as to approximate losses from a DC power flow. *Mixed Integer Programming* (MIP) takes into care the discrete nature of investment decisions.

Also *Non Linear Programming* (NLP) and *Mixed Integer Non Linear Programming* (MINLP) have been used to shape TEP problem with the full AC power flow assumption.

Incorporating market considerations has promoted the appearance of *equilibrium formulations* of the problem.

<table>
<tr><th colspan="4">CLASSICAL METHODS</th><th colspan="2">NON- CLASSICAL METHODS</th></tr>
<tr><td colspan="2">MATHEMATICAL PROGRAMMING</td><td rowspan="4">STOCHASTIC DECOMPOSITION</td><td rowspan="4">EQUILIBRIUM FORMULATION</td><td rowspan="2">SENSITIVITY ANALYSIS</td><td rowspan="4">METAHEURISTICS</td></tr>
<tr><td>LP</td><td rowspan="2">NLP</td></tr>
<tr><td>QP</td><td rowspan="2">EXPERT SYSTEMS</td></tr>
<tr><td>MIP</td><td>MINLP</td></tr>
</table>

Table 1: TEP solution methods classification[LUMB14].

In order to solve TEP problem incorporating uncertainties directly in the decision process, stochastic decomposition techniques have earned special interest. Since Benders Decomposition (BD) has been applied in a wide variety of fields with notable success, it has earned a central part in the Stochastic Optimization domain. It has become one of the most important techniques for solving the TEP problem [LUMB12].

BD allows solving large problems by dividing them into two stages: a master problem which proposes new solutions and a subproblem which evaluates them and sends feedback to the master for the next proposals. The master problem represents the first stage plus some conditions, known as cuts, derived from the second stage. The subproblem represents the second stage for the solutions provided by the master problem. The process is repeated iteratively, alternating master and subproblem resolutions. This method is used in the TEPES (http://www.iit.upcomillas.es/aramos/TEPES.htm) model developed at the Institute for Research in Technology of ICAI (IIT). As master resolution requires expensive computational burden, this research project explores the possibilities of combining BD with metaheuristics.

1.1.1.1. <u>Transmission Expansion Planning for an Electric System (TEPES)</u>

As explained previously, TEP determines the investment plans for new facilities (lines and other network equipment) for supplying the forecasted demand at minimum cost. Tactical planning is concerned with time horizons of 10-20 years. Its objective is to evaluate the future network needs. The main results are the guidelines for future structure of the transmission network. The Institute for Research in Technology of ICAI (IIT) has developed a model for evaluating the future network needs in a tactical level, the Transmission Expansion Planning for an Electric System (TEPES), which has been carried out in a GAMS environment.

Long-term

TEPES model presents a decision support system for defining the transmission expansion plan of a large-scale electric system at a tactical level. A transmission expansion plan is defined as a set of network investment decisions for future years.

Two-stage stochastic optimization

The optimization method used is based on a functional decomposition between an automatic transmission plan generator (based on optimization) and an evaluator of these plans from different points of view (operation costs for several operating conditions, or reliability assessment for N-1 generation and transmission contingencies). The model is based on BD where the master problem proposes network investment decisions and the operation subproblem determines the operation cost for these investment decisions and the reliability subproblems determine the not served power for the generation and transmission contingencies given that investment decisions.

Objective function

The problem is modeled with and objective function which minimizes transmission investment and variable operation costs. The operation cost includes generation and emission costs. Also a reliability cost associated to N-1 generation and transmission contingencies is considered.

DCLF

The operation model (evaluator) is based on a DC load flow although a simpler transportation representation is allowed for some or all the lines. Network losses are disregarded in the case studies that are going to be studied in this research project, but the model has the option of also considering them.

Variables

By nature the transmission investment decisions are binary although can also be treated as continuous ones. The current network topology is considered as the starting point for the network expansion problem.

1.1.2. *State of the Art: The Genetic Algorithm*

A metaheuristic is a solution technique that provides a sufficiently good solution to an optimization problem, especially with limited computation capacity.

Metaheuristics are used for combinatorial optimization in which an optimal solution is sought over a search-space. These methods, as well as sensitivity analyses, have faster computational ability than mathematical programming [YOSH95]. Several metaheuristics have been proposed in the last decade to solve the TEP problem [YOSH95, ALTA12, SILV00]. Optimization approaches based on metaheuristics have demonstrated the potential of finding high quality solutions with a faster time-response.

Among metaheuristics we can bright light: Tabu Search, Swarm Intelligence and Evolutionary Algorithms [SILV10].

Tabu Search is an optimization technique that obtains a first solution from a global exploration mechanism. The algorithm uses a flexible memory of the previous states, performing neighbour research. It generates a list of useless information and this memory guides the process avoiding the saved elements until there are no movements that improve the current solution.

Swarm Intelligence is based on the interaction of organized agents with each other and with the environment. A global intelligence behavior is observed although there is no control structure guiding the agents. Ant Colony Optimization and Artificial Bee Colony algorithms are important examples of this category.

Finally *Evolutionary Algorithms* are a family of computational models inspired by biological evolution. They are robust optimization methods that work with a set of candidate solutions (individuals) named a population. In these algorithms, the population of individuals is modified along the evolution of generations by the application of operators such as: selection, recombination and mutation. The most extended evolutionary algorithm is the Genetic Algorithm (GA). Artificial Immune System (AIS) can be included in this category as they are based on the same ideas that lie behind the Evolutionary Algorithms, discarding the recombination operator. Considering TEP problem, several works can be found in the literature using AIS [ALTA12] and GA with good results [RUDN96, SILV00, DEB02, SILV06, MAGH09, SILV10, MAGH11, OTHM11].

<table>
<tr><th colspan="4">METAHEURISTICS</th></tr>
<tr><th colspan="3">INSPIRED BY EVOLUTION</th><th>TRAJECTORY</th></tr>
<tr><td>EP</td><td rowspan="2">AIS</td><td rowspan="2">SI</td><td rowspan="3">TABU SEARCH</td></tr>
<tr><td>GA</td></tr>
<tr><td>GP</td><td>ANN</td><td>SA</td></tr>
</table>

Table 2: Metaheuristics classification [LUMB14].

GA were first introduced by Holland in 1975. An implementation of the GA begins with an initial population (typically random). The population is a set of candidate solutions encoded on a chromosome data structure over which we apply selection, recombination and mutation. There are several possibilities of codifying the population. In a decimal one, the chromosome represents the number of circuits being added in each right-of-way. Although the decimal codification is being adopted in some studies with good results [SILV00, SILV06], the binary one is the most extended way of representing an individual [YOSH95, RUDN96, DEB02, ALTA12]. For multi-stage problems each member is represented by a matrix. In the particular case of considering a one-stage problem, the representation becomes a vector.

In most cases the initial population is generated randomly [ALTA12, MAGH11, SILV11], but some studies point out the importance of using a procedure to build initial good quality sequences for a better performance of the GA approach [RUDN96, SILV06, SILV00].

Once the initial generation is constructed, the population is modified by the application of the genetic operators. First, it is necessary to evaluate and assign a fitness value to each member that provides a measure of performance into an allocation of reproductive opportunities. This fitness value can be either directly [SILV06] or inverse [MAGH11] to the fitness value.

Recombination starts with the selection of the parents. Selection is applied to the current population to create an intermediate population. There are different methods of selecting this intermediate population. *Ranking Selection* consists on ranking the individuals according to their fitness value, and choosing randomly the best ones [RUDN96]. In the

Tournament Selection a random permutation of the current population is obtained. The members are then split up into a definite number of groups and the individual is selected randomly with a probability based on their fitness of each group [SILV00, SILV06]. In the *Roulette Wheel Selection* the members are selected by spinning a roulette wheel with slots sized according to the fitness function, so the probability of choosing a member is proportional to its fitness value [YOSH95, SILV11, GUPT12].

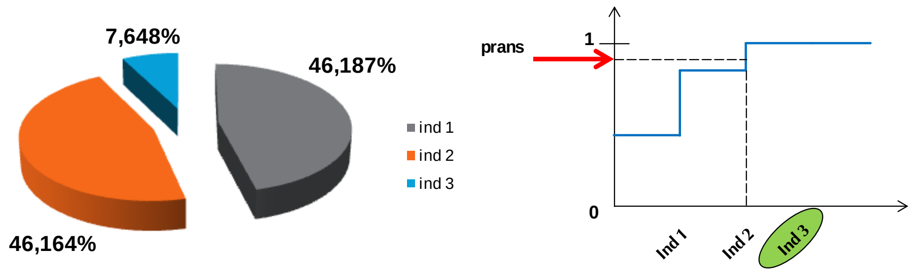


Figure 1: Roulette Wheel Selection.

The process continues with the crossover of the parents obtaining the offspring. The most extended recombination operator is the one-point crossover, so as each offspring has a piece of each parent [MAGH11, SILV11]. It is possible to use more than one recombination point, so the offspring is constructed with the segments of the parent strings between these points, but it has been proven that it does not provide improvements over the one-point crossover [WHIT94].

Node_1	.	Node_2	.	cc1	0	1	0	1
Node_1	.	Node_4	.	cc1	1	1	1	1
Node_2	.	Node_3	.	cc1	1	1	1	1
Node_2	.	Node_4	.	cc1	0	0	0	0
Node_2	.	Node_5	.	cc1	0	0	0	0
Node_2	.	Node_6	.	cc1	0	1	0	1
Node_3	.	Node_5	.	cc1	1	0	1	0
Node_3	.	Node_8	.	cc1	0	0	0	0
Node_4	.	Node_6	.	cc1	0	1	1	0
Node_5	.	Node_6	.	cc1	0	0	0	0
Node_5	.	Node_8	.	cc1	0	1	1	0
Node_6	.	Node_7	.	cc1	1	1	1	1
Node_6	.	Node_9	.	cc1	0	0	0	0
Node_7	.	Node_9	.	cc1	1	1	1	1

Figure 2: One-point crossover applied in TEPES_9BUS.

As a neighbour research strategy, mutation is then applied. It is an important process of any GA [MAGH11] that provides richness to the members' sequences, contributing to global optimality. The offspring is slightly modified with a low probability [RUDN96, DEB02]. The mutation can also be applied by the inversion of two randomly selected chromosomes from an offspring [OTHM11]. In reference [SILV00], a Simulated Annealing mechanism is developed for improving the mutation operator. It is based on the fact that nature performs an optimization of the total energy or a crystalline solid when it is annealed under a slow cooling schedule. It is a nice way to increase the mutation rate in order to enhance the local search around the optimum solution.

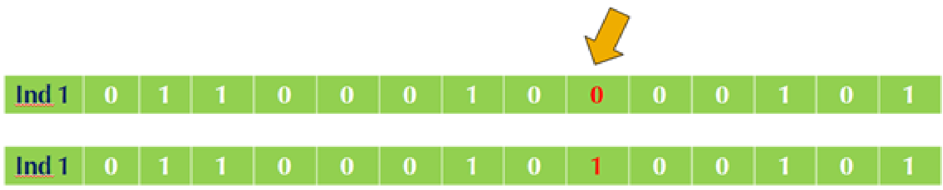
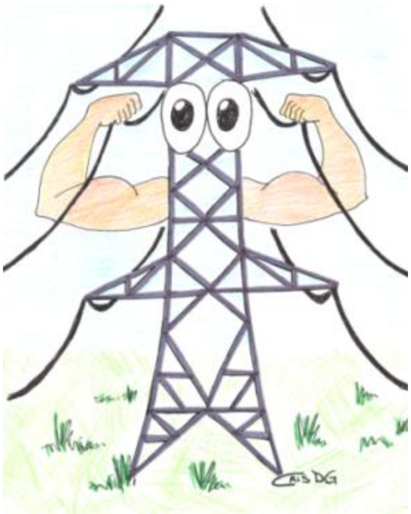


Figure 3: Mutation applied in TEPES_9BUS.

Some interesting improvements have been applied to the classic GA. In [YOSH95] it is observed that saving the best individual, named *the queen*, of each generation and preserving it along the evolution contributes to a faster search of near optimal solutions. Another way of avoiding the lack of diversity in the population is the introduction of immigrants that consists on creating an external population that can fight a duel with the indigenous population [DEB02]. Unlike the traditional GA, the genetic process developed in [SILV06] substitutes, in each step, only the worst member of the population, in such a way that all the population individuals may be different. This substitution preserves the best created topologies.

The process is repeated until a convergence criterion is reached. The most common and easiest criterion is to limit the total number of generations [YOSH95, SILV00] but other criteria, such as stopping the process when the best solution found does not improve after a specified number of iterations, has been implemented in [SILV06], or finishing when all the members have an identical genetic code [RUDN96].

1.2. Motivation



Despite the efficiency of BD resolution for solving TEP problem, a large size of the problem, integrality conditions or the addition of a large number of constraints (cuts) that complicate the master problem could cause slow master resolutions.

Owing to the simple implementation of heuristic optimization algorithms, using them for solving hard optimization problems like TEP should be suitable.

These algorithms have a better chance of finding better solutions than those obtained by mathematical approaches based on decomposition techniques for large problems. Results show that the GA is not only suitable, but a promising technique for solving such a problem, proposing new solutions to be evaluated by the subproblem [SILV00] and preserving the discrete nature of investment.

1.3. Objectives

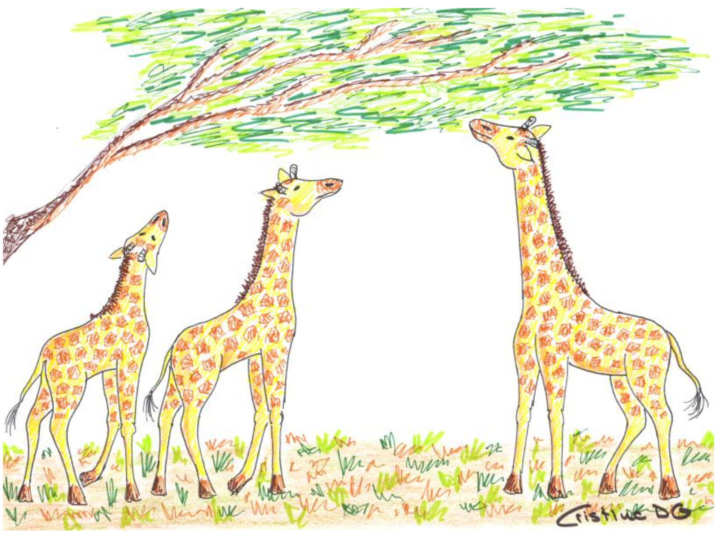
The aim of this final project is to introduce a GA based on the canonical one, in the TEPES model, as the master problem of BD to propose new candidate solutions to be evaluated by the subproblem. The first step in the implementation of a GA is to generate an initial population; in the canonical GA each member of this population will be a binary string. The fitness of each string is always defined with respect to other members of the current population. This canonical algorithm consists of three main operators: selection, crossover and mutation.

The most important objectives are to carry out an adequate experimental adjust of the genetic parameters in order to obtain optimal solutions for the TEP problem in Spain and to contrast the efficiency of the GA with BD in public case studies.

A field of improvement could be the modification of the canonical genetic algorithm. Saving the best individual of each generation could be interesting as a mechanism of leading the population while the best topologies are preserved so as to avoid losing the good qualified genetic information. Several studies discuss the importance of generating an initial good quality population [SILV10]. The introduction of immigrants in the population could be also interesting as a way of avoiding local optima [MAGH09].

2. The Genetic Algorithm

Evolution has lead to specialized living beings that in some way are optimum for their specific environments. These adaptation methods can be explored in order to develop new optimization techniques. This creative idea was first introduced by John Holland [HOLL75] for solving parameter optimization problems. GA has proven to be a robust method, it is able to provide good quality solutions in a wide range of problems and it can be easily implemented.



It all starts with an initial population, and then, generation after generation, the individuals, or members of the current population compete with each other. Darwinian selection is the main concept, fitter members of the population are more likely to mate with each other and have an offspring in the next generations. This crossover leads to a recombination of the genetic materials and to a sequence of generations that is successively fitter.

In addition to the crossover operator, nature introduces random modifications with a low probability in order to increase diversity among the population, avoiding homogeneity. This mechanism is known as mutation.

2.1. The problem encoding and the evaluation function

Usually there are only two main components of most genetic algorithms that are problem dependent: problem encoding and the evaluation function.

Problem encoding is necessary to translate each possible solution, namely individual or 'chromosome' in the GA literature, to a string of symbols. Each element of the string is called 'gene' and the symbol that represents it is known as 'allele'. Binary codification is frequently used in general problems because of its simplicity. This means that the variables are discretized in an a priori fashion, and that the range of the discretization corresponds to some power of 2. If the parameters are actually continuous then this discretization should provide enough resolution to make it possible to adjust the output with the desired level of precision.

The notion of evaluation and fitness are sometimes used interchangeably. However, it is useful to distinguish between the evaluation function and the fitness function used by a GA. The evaluation function provides a measure of performance with respect to a particular set of parameters and constraints. The fitness function transforms that measure of performance into an allocation of reproductive opportunities. The evaluation of a string is independent of the evaluation of any other string. The fitness value of that string, however, is defined with respect to the other members of the current population.

2.2. The Schema Theorem

The schema theorem [HOLL75] is the main mathematical support for the GA theory.

A schema is a template that identifies a subset of strings with similarities at certain string positions. Given an initial alphabet Ω and considering an extended alphabet Ω' constituted by the alphabet Ω plus a new element $'*'$, a schema is any of the strings formed by the extended alphabet.

The symbol $'*'$ represents any of the other symbols of the alphabet. The length of a schema is the distance of the last fixed character and the first one and its order is simply the number of defined bits.

The number of members of the population that belong to a given schema in each generation evolves according to its fitness. The fitness is defined with respect to the remaining individuals, its length and order. The GA biases future samples towards schemas that are estimated to have above-average fitness.

$$P(H, t+1) \geq P(H, t)\frac{f(H, t)}{m(t)}\left(1 - p_c\frac{d(H)}{l-1}\right) \cdot (1 - p_m)^{o(H)}$$

$P(H, t)$: the proportional representation of the schema H at time t.

$m(t)$: average fitness of the population at time t.

$f(H, t)$: fitness of members of the schema H at time t.

p_c: crossover probability.

p_m: mutation probability.

$d(H)$: defining length of the schema H associated with one-point crossover.

$o(H)$: the order of the schema H.

2.3. Selection methods

It is helpful to understand the execution of the genetic algorithm as a two stage process. It starts with the current population. Selection is applied to the current population to create an intermediate population. Recombination and mutation are applied to the intermediate population to create the next population. The process of going from the current population to the next population constitutes a generation.

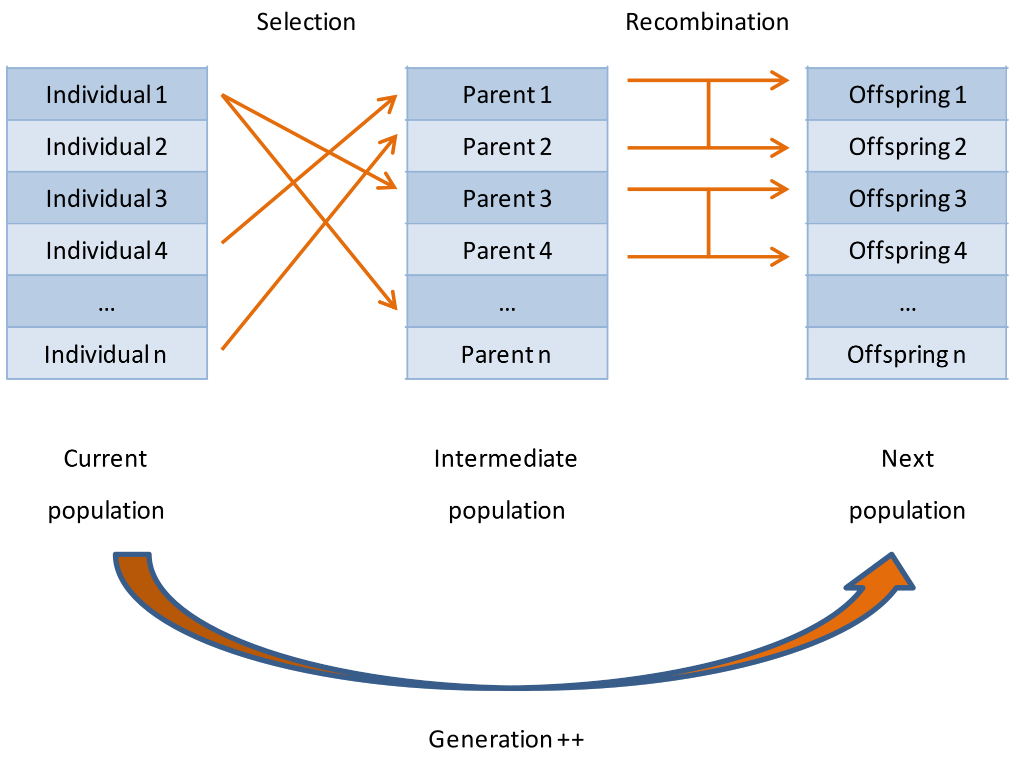


Figure 4: Current, intermediate, next population and generation concepts.

The fittest members of the population will be selected to be a part of the intermediate population, and therefore will have an offspring. There are two main methods to replicate this operator. The first one is called the Roulette Wheel Selection and is based on Inverse Transform Sampling. It consists on making a number of copies of each individual in the intermediate population proportional to its fitness. The selection process can be carried out by constructing a wheel with N slots, each of one corresponding to a member of the population and with an angular width proportional to its fitness. The wheel is spun as many times as the population size for selecting the individuals of the next generation.

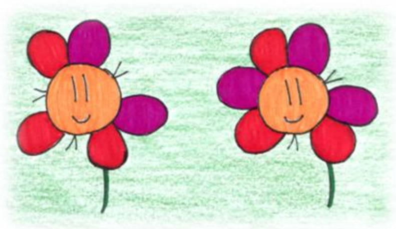
The second one is the Tournament Selection, which is a random permutation of the current individuals. They are then differentiated into a definitive number of groups and finally the fittest member of each group, with a random probability based on their fitness, is selected to generate the intermediate population.

2.4. Crossover operators

After selection has constructed the intermediate population, recombination can occur. Recombination generates the next population from the intermediate one; this operator returns two children out of a pair of parents by recombining their genetic information. The most common recombination technique is the one-point crossover which consists in interchanging genetic information of the two parents from a randomly chosen point within the string length.

2.5. Mutation

Mutation consists in introducing random changes in some members of the population with a very low probability. This mutation rate can be constant or follow a variation rule, such as a simulated annealing mechanism [SILV00].



Mutation, therefore acts as a background operator, occasionally changing bit values and allowing alternative alleles to be retested. The main contribution of this operator is that it reduces the risk of getting trapped in a local optimum. A variation of this mechanism includes immigration [DEB 02], which consists on the introduction of a definite number of new individuals in the population, replacing the worst ones.

3. BENDERS' DECOMPOSITION

Benders' Decomposition has been applied in a wide variety of fields where problem characteristics mean that an application of this method can result in considerable time savings.

BD has earned a central place as one of the most important techniques applied in the Stochastic Optimization domain. It has been used for solving problems as diverse as generation expansion planning, transmission expansion planning, distribution system design, hydrothermal coordination or the unit commitment problem.

BD is also known as *primal decomposition* because the master problem fixes variables in the subproblem, *L-shaped decomposition* because that is the shape of the constraint matrix or *recourse decomposition* because the master assigns directly the resource decisions to the subproblem.

The two-stage stochastic linear problem is defined in its complete form as follows:

$$\min_{x,y^{\omega}} c^T x + d^T y \tag{1}$$

$$Ax = b \tag{2}$$

$$Dy = e \tag{3}$$

$$Tx + Wy^{\omega} = h^{\omega} \tag{4}$$

$$x, y^{\omega} \geq 0 \tag{5}$$

where x, y^{ω}represent the first-stage and second-stage variables and $c^T, d^T$are their respective costs in the objective function. The equation (2) represents the first-stage constraints, the equation (3) the second-stage constraints and the equation (4) are the constraints that link both stage. The stochastic scenarios are referred to as index ω.

BD allows solving large problems by dividing them into two parts: a master problem which proposes new solutions and a subproblem which evaluates them and sends feedback to the master for the next proposals.

The master problem represents the first stage plus some conditions derived from the subproblem, known as cuts. The subproblem represents the second stage for fixed values of the first stage variables provided by the master resolution. The solution given by the subproblem provides an upper bound for the optimal value. In addition the information obtained by the subproblem is used to improve the description of the master problem by including a new cut. Both problems are solved iteratively until convergence is reached. Then the problem can be interpreted as:

The complete master problem

$$\min_{x} c^T x + \sum_{\omega \in \Omega} \theta^{\omega}(x) \tag{6}$$

$$Ax = b \tag{7}$$

$$x \geq 0 \tag{8}$$

where $\theta^{\omega}(x)$ represents the recourse function, that is, the second-stage objective function as a function of the first-stage decisions. This can be expressed as:

The relaxed master problem

$$\min_{x, y^{\omega}} c^T x + \sum_{\omega \in \Omega} \theta^{\omega} \tag{9}$$

$$Ax = b \tag{10}$$

$$\theta^{\omega} \geq f^{\omega l} + \pi^{\omega l} T(x^l - x) \qquad l = 1, \dots, j,\ \forall \omega \tag{11}$$

$$x, \geq 0 \tag{12}$$

where θ^{ω} represents the recourse variable, that is, the second-stage objective function as a function of the first-stage decisions, equation (11) is the expression of the cuts provided by the subproblem and $f^{\omega l}$ represents the optimal second-stage cost for the first-stage x^l proposed by the master problem evaluated in scenario ω. The optimal second-stage cost for iteration j is obtained by solving the subproblem:

The subproblem

$$f^{\omega j} = \min_{y^{\omega}} d^T y^{\omega} \tag{13}$$

$$Dy = e \tag{14}$$

$$Wy^{\omega} = h^{\omega} - Tx^j \qquad ; \pi^{\omega j} \tag{15}$$

$$y^{\omega} \geq 0 \tag{16}$$

where π^{ω} represent the dual variables of the constraints.

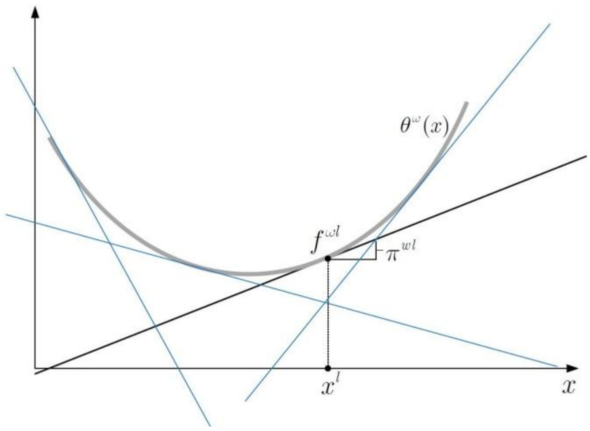


Figure 5: Graphical representation of Benders' Decomposition [LUMB14].

BD is likely to yield time savings in the resolution of a problem when the following conditions are fulfilled:

- First stage variables increase the difficulty of the problem. The problem could become considerably easier when fixing temporarily the first-stage variables. The number of first-stage variables should be smaller than the second-stage variables

- The master problem and the subproblem have a different nature. Decomposition allows the use of the most suitable technique for each problem, which results in improved efficiency. In the TEP problem, first-stage deals with investments that are integer variables. The second-stage holds power flows that have a linear nature.

4. A GENETIC ALGORITHM FOR THE TEPES

The aim of this research project is to design an evolutionary algorithm and explore the possibilities of combining the GA with BD in the TEPES model so as to solve the TEP problem in an efficient way.

As introduced in section 1, GAs are a family of computational models that imitate the mechanisms of evolution to solve complex optimization problems. The range of problems to which GAs have been applied is quite broad.

GAs have been applied to a wide range of problems and have proven their high capacity for obtaining quality good solutions in affordable computation times.

In a broader usage of the term, GA is any population-based model that uses selection and recombination operators to generate new sample points in a search space.

The GA developed in this final project is based on the canonical one. This section explains each of the operators used for the construction of this GA.

The goal of the project is to develop a GA for TEP and to compare the results with the solutions obtained with BD. The GA works as the master problem and constructs candidate solutions for the TEP problem. The GA sends this information to the subproblem. Then the subproblem calculates the cost of these candidate lines and sends feedback to the GA. The process is repeated until the convergence criterion is reached.

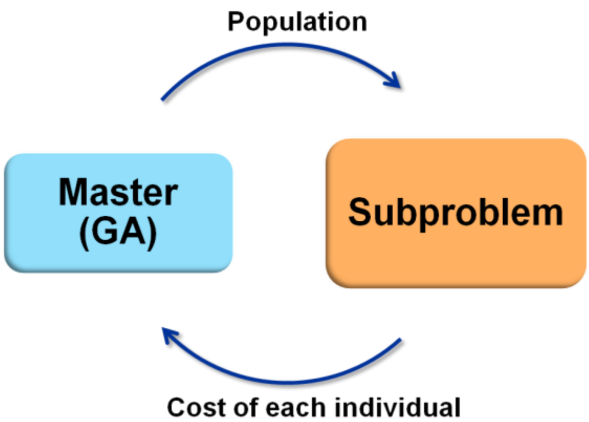


Figure 6: GA as the master problem of BD.

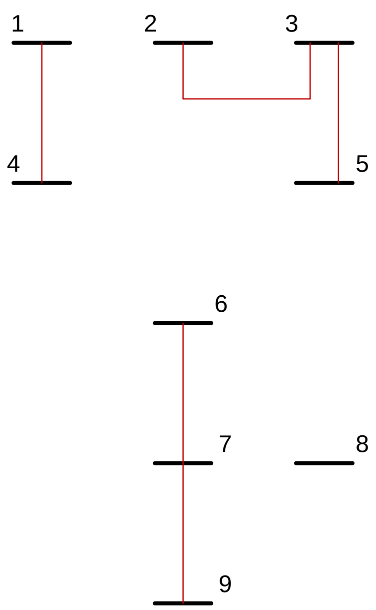
According to this approach, the first stage variables are the candidate solutions for the TEP problem, codified in the genetic population and the second stage variables are the costs of each individual. The subproblem calculates the operational and investment cost for each individual, so for one resolution of the master problem, the subproblem needs to be solved as many times as the number of individuals which form the population.

4.1. The Initial Generation

4.1.1. *Codification*

It all starts with the creation of the first generation, a set of new individuals. In this GA each member of the population will be a binary string of length l_c, which corresponds to the number of power lines. Installing a candidate line in the possible solution will be represented with a '1' and not considering connecting that line will be symbolized with a '0'.

This way of encoding the chromosomes or alleles of each member of the population results in a simple codification of the genetic operators.



Node_1	.	Node_2	.	cc1	0
Node_1	.	Node_4	.	cc1	1
Node_2	.	Node_3	.	cc1	1
Node_2	.	Node_4	.	cc1	0
Node_2	.	Node_5	.	cc1	0
Node_2	.	Node_6	.	cc1	0
Node_3	.	Node_5	.	cc1	1
Node_3	.	Node_8	.	cc1	0
Node_4	.	Node_6	.	cc1	0
Node_5	.	Node_6	.	cc1	0
Node_5	.	Node_8	.	cc1	0
Node_6	.	Node_7	.	cc1	1
Node_6	.	Node_9	.	cc1	0
Node_7	.	Node_9	.	cc1	1

Figure 7: Example of the population codification in TEPES9BUS.

4.1.2. *Initial generation*

The initial population is generated randomly following a binomial distribution. The fact that the optimal plan installs few candidate lines means that it could be interesting to connect less than a half of candidate lines at the beginning. Thus so the initial population will be closer to the optimal individual and it is reasonable to think that it will be necessary to carry out a lower number of iterations for obtaining the optimal solution. The algorithm has been tested with binomial success probabilities for the initial generation of 30, 40, 50 and 60%.

4.2. Selection

4.2.1. *Fitness construction*

The next genetic operators are applied over an intermediate population. The intermediate population is constructed from the current population. For constructing the intermediate set of members it is necessary to assign each individual a measure of its performance, a fitness value.

The fitness function is constructed from the evaluation function. In the TEPES case the evaluation function is the total cost provided by the subproblem. The subproblem gives the operation and investment cost of each planning and the fitness function is constructed to be inversely proportional to the total cost. Let it be l the index for iterations, referred as to generations in the genetic algorithm's literature. The individuals are referred to as index i.

$$Fitness(l,i) = \frac{1{,}1 \cdot \max_j\{totalCost(j)\} - totalCost(i)}{1{,}1 \cdot \max_j\{totalCost(j)\} - 0{,}9 \cdot \min_j\{totalCost(j)\}} \quad \forall\, i;\ \ j = 1, \dots n$$

$$FitnessNormalized(l,i) = \frac{Fitness(l,i)}{\sum_j Fitness(l,j)} \quad \forall\, i;\ \ j = 1, \dots n$$

A security band is defined above the highest cost of the generation, whose task is to prevent the best individual having the 100% probability of being chosen to fulfill the intermediate population. In addition another security band is defined under the lowest cost of the generation so as to avoid always eliminating the worst individual of each generation.

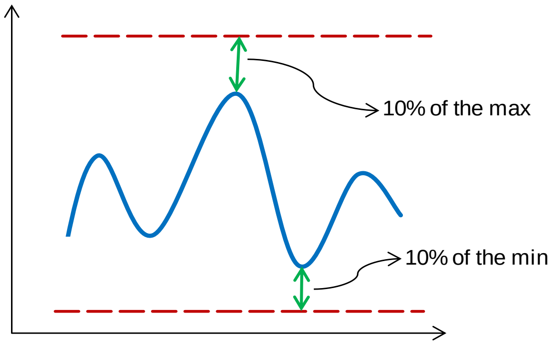


Figure 8: Fitness construction.

Once the fitness function is defined, then it is normalized with respect to the addition of all the fitness values in that generation so as to obtain the distribution fitness function. That is a necessary step to obtain measures of performance of each individual between 0 and 1, so the distribution fitness function represents directly the probability of each individual for having an offspring.

4.2.2. *Roulette Wheel Selection*

The technique used for selecting the intermediate population is the Inverse Laplace Transform. This method involves computing the cumulative distribution fitness function

In the GA literature it is known as the Roulette Wheel Selection, because the members are selected by spinning a roulette wheel with slots sized according to the fitness function, so the probability of choosing a member to create the intermediate population is proportional to its fitness value.

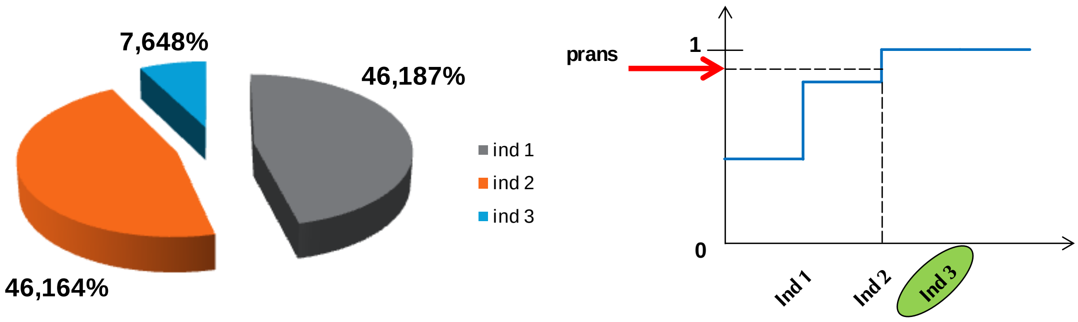


Figure 9: Example of Roulette Wheel Selection.

All the GA literature agrees in the suitability of using Roulette Wheel Selection. The Tournament Selection, although it has one more parameter to adjust, does not provide better results than the Inverse Transform Sampling.

4.3. Recombination: 1-Point Crossover

Once the intermediate population is constructed then the traditional genetic operators for generating the next generation are applied.

In order to maintain the population size constant along the algorithm evolution, two children are obtained from each pair of progenitors. The probability of crossover in this case is 100%.

The progenitors are two consecutive members of the intermediate population. Then the crossover point is generated randomly between 1 and the total number of candidate lines, or genes.

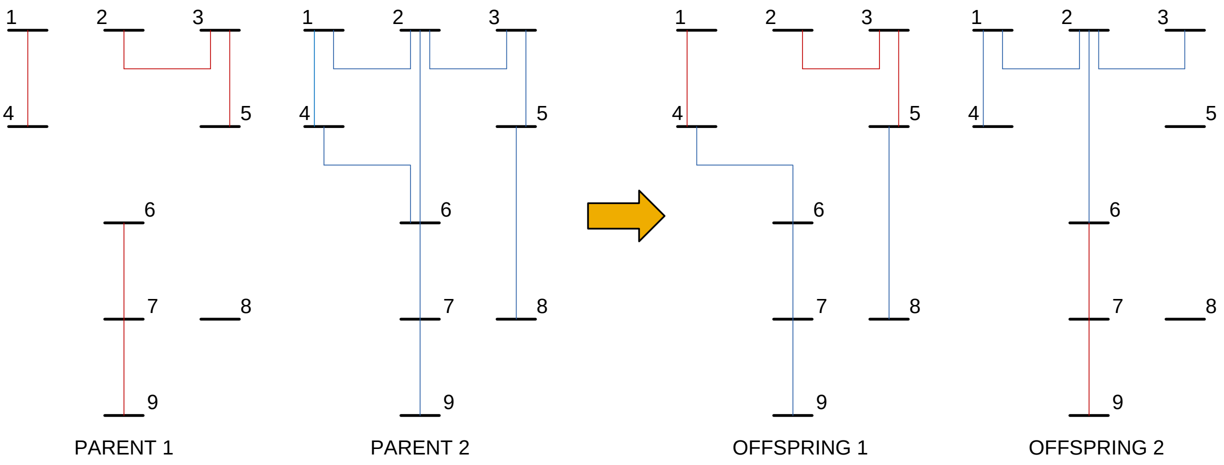


Figure 10: Example of 1 Point Crossover in IEEE 9BUS system.

4.4. Mutation

The motivation for using mutation is to prevent the permanent loss of any allele or bit. After several generations it is possible that selection and crossover drive all the bits in some position to a single value. This fact can lead to a premature.

Mutation acts as a background operator, occasionally changing some bit values, reintroducing missing information and contributing to global optimality.

The mutation operator developed in this project consists on disconnecting a candidate line if it is connected in the proposal solution, or connecting it if it is not considered to be installed in the candidate solution. A random number is assigned to each line. The numbers are generated by means of a uniform distribution between 0 and 1 and they are compared with the mutation probability. Mutation is applied independently to each one of the candidate lines or alleles of each individual of the population.

This exchange is carried out with a mutation rate defined at the beginning of the resolution and it stays constant during the generations. The adjustment of the mutation probability is carried out empirically; the optimal values for the TEP problem are between of 1% and 5%.

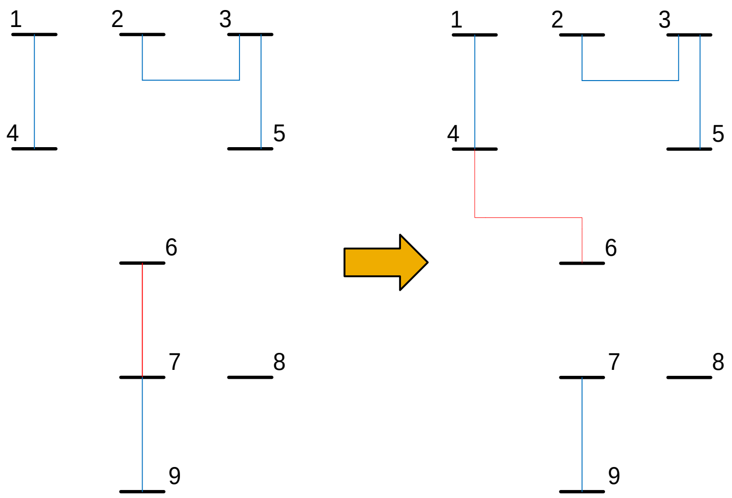


Figure 11: Example of mutation in TEPES 9BUS.

The previous figure shows how the mutation has been applied to an individual. The line between buses 4 and 6 has been connected and the line between nodes 6 and 7 has been disconnected.

4.5. The Queen

One extended strategy for not losing the best found topologies until this moment is to save the best individual in each generation and inserting it in the next population. This member is called *the queen*.

The strategy developed in this project for conserving the best individual is to insert it in the next generation. The queen individual will place the individual of the current generation which takes up the first position after the crossover and the mutation in the last population.

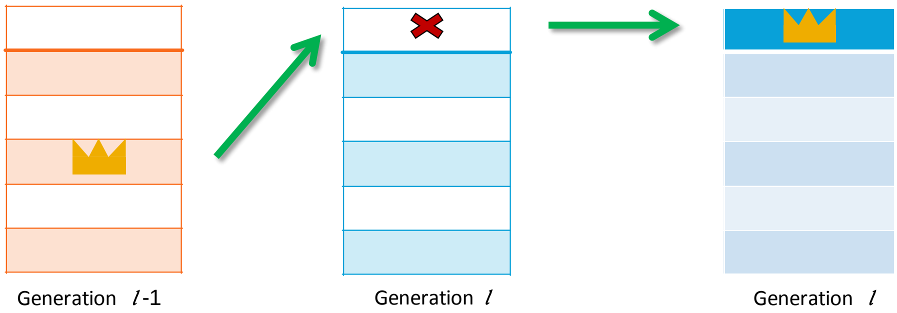


Figure 12: Queen individual.

This way of replacing any member of the population by the queen individual, instead of the worst one, do not contributes to an accused elitism.

4.6. Immigrants

Sometimes the use of the mutation is not enough to avoid premature convergence so it is arisen a new evolutionary operator, the introduction of immigrants in the next population.

The immigrants are generated randomly in the same way as the initial population and they are inserted in the next generation. These immigrant individuals are inserted in the population with a low fixed probability. The size of the immigrant population is a parameter defined at the beginning of the resolution; it should be a relatively low proportion of the population.

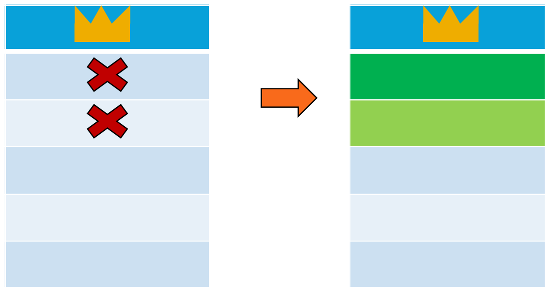


Figure 13: Introduction of immigrants.

4.7. Definitive Structure

To summarize, the genetic operators explained previously are applied over the population iteratively using the information provided by the subproblem of BD. With all the operators developed in the last section, the definitive structure for the problem resolution used stays as follows:

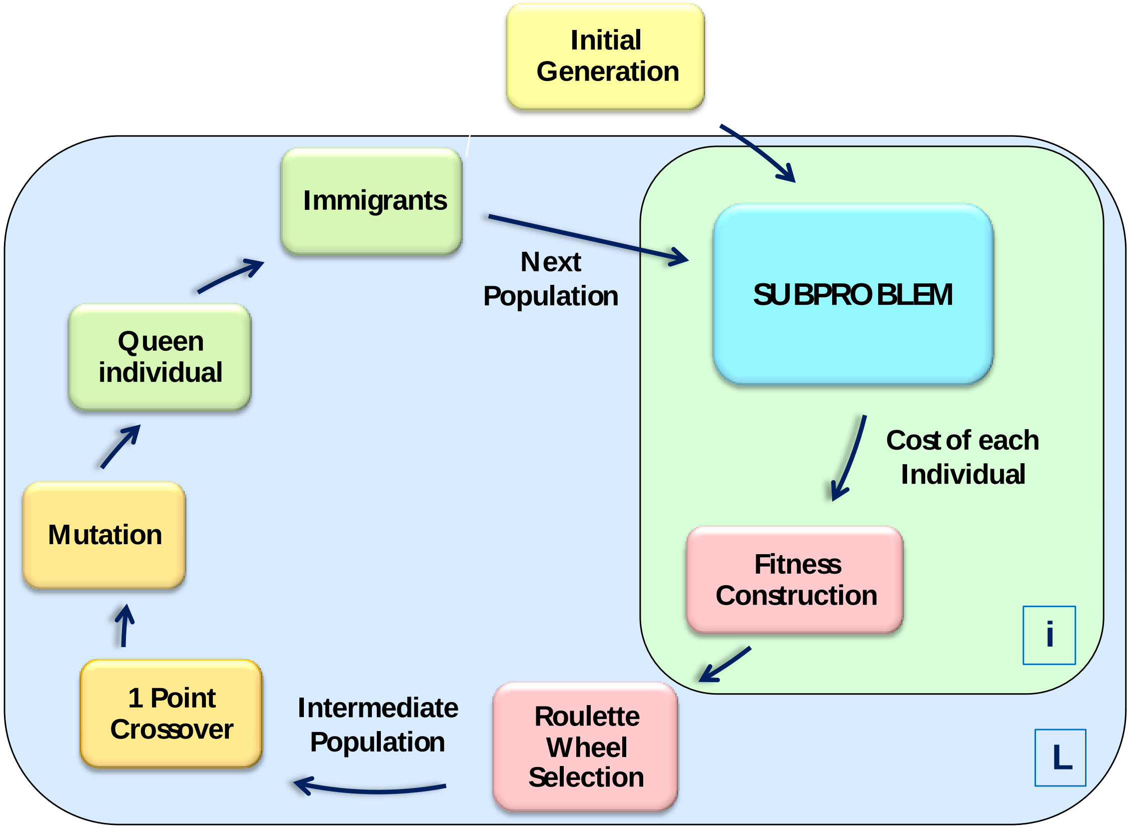


Figure 14: Definitive GA structure.

First of all, the initial generation is created randomly as explained in section 4.1. Then, this first population is sent to the subproblem, and the subproblem calculates the cost of each individual. The subproblem is run as many times as individuals are in the population.

The next step is to calculate the fitness from the total cost provided by the subproblem. Roulette Wheel Selection is applied in order to construct the intermediate population. Then, crossover and mutation are used to obtain the next population, and after that, the best individual (*the queen*) is included in the next population in the place of any other individual randomly chosen.

A predefined number of immigrant individuals are randomly generated, in the same way than the initial population, and they are included in that next population. Last, this group of individuals is sent to the subproblem in order to calculate the total cost of each individual. The process is repeated until the convergence criterion is reached or the maximum number of total generations is finished.

5. RESULTS

5.1. TEPES2020BUS

The purpose of this chapter is to test the Genetic Algorithm and adjust the parameters in order to achieve a suitable resolution for TEP. The case study is based on the Spanish power system. The transmission data was taken of 2008 from publicly available ENTSO-e REE E-SIOS cases [LUMB14]. The system is composed of 1084 nodes and 294 power plants (nuclear, coal, CCGTs, hydro, wind and solar). The existing transmission network is configured by 1505 lines and transformers. The candidate transmission network consisted of 153 power lines.
Three cases study have been constructed from the original one. The case study 1 considers 33 candidate power lines, the case study 2, 90 and the case study 3 takes into account all the candidate transmission lines of the case study based on Spain.

5.1.1. *Case study 1*

The candidate transmission network of the case study 1 is configured by 33 lines. In this case the optimal plan installs 16 candidate lines with a total cost of 19850.241 M€. This solution has been obtained solving the complete problem.

The following figure shows the distribution of the initial population. The Y axis represents the frequency and the X axis the total cost (the operation cost plus the investment cost) in M€. The graphic has been constructed with a sample size of 1000 individuals. Figure 15 represents the 99% of the members of the initial generation, avoiding outliers.

In this first graphic the initial population has been constructed connecting a half of the candidate lines. That corresponds with a binomial success probability of 50% in the initial generation.

The population distribution of this case study has three modal points, 35207 M€ with a frequency of 187; 53118 M€ with a frequency of 339; and 71217 M€ with a frequency of 85. This last point is related with energy not supplied of 4.9 MW which produces a higher operation cost.

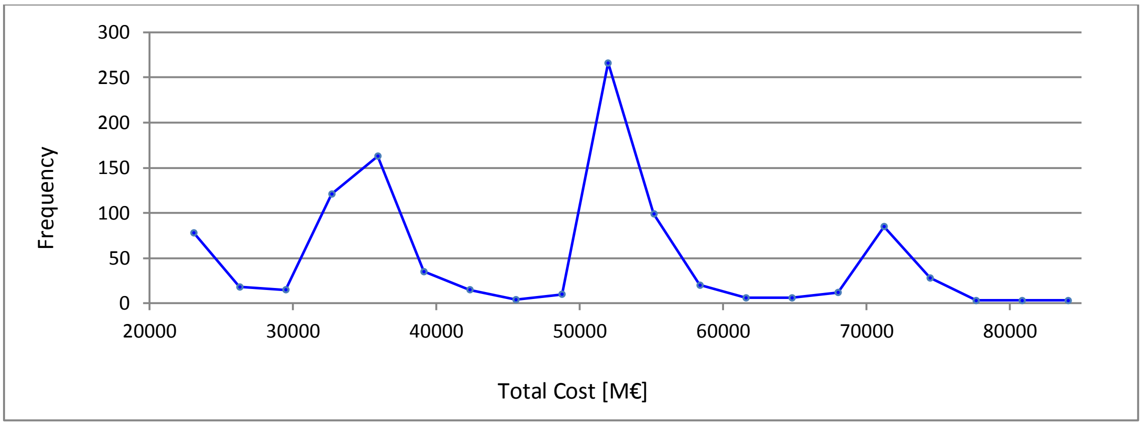


Figure 15: Candidate lines distribution in case study 1. Success probability of 50% in the initial population.

Figure 16 shows a comparison between generating the initial population with different success probabilities in the binomial distribution.

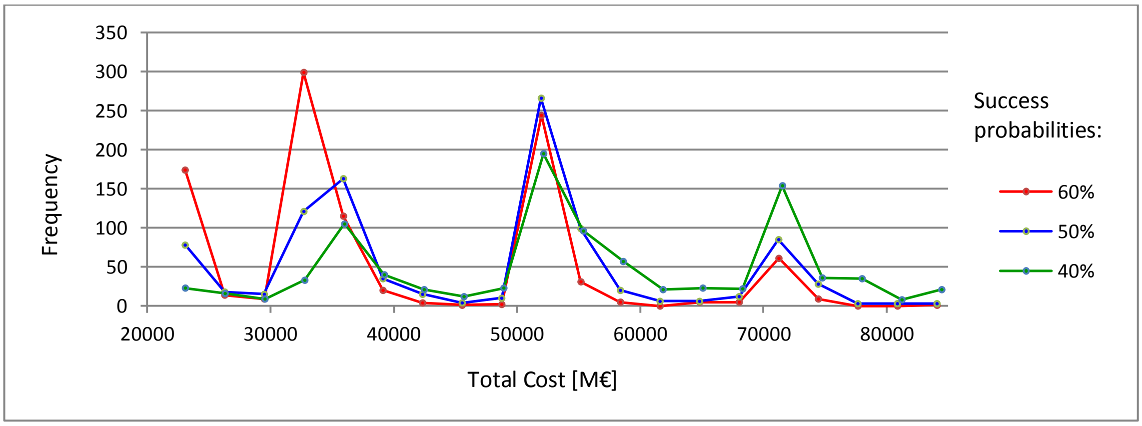


Figure 16: Candidate lines distribution in case study 1. Comparison with different success probabilities in the initial population.

The three functions have three modal points. Although the optimal plan installs 45% of the candidate lines, it can be seen that connecting 60% of the candidate lines in the initial generation has a better cost distribution, the costs are nearer to the optimal cost. This fact is due to the increment of operation cost related with the energy not supplied that appears if connecting less than a half of the candidate power lines.

5.1.2. *Case study 2*

The candidate transmission network of this case study is configured with 90 lines. The optimal solution, provided by the resolution of the complete problem, has a total cost of 19741.229 M€ and installs 19% of the candidate lines. The following graphic represents the density distribution of the initial generation with 50% probability of installing each candidate. 90% of the initial population belongs to the interval of costs from 19760M€ to 19780M€. The remaining 10% of the members has been omitted to eliminate outliers.

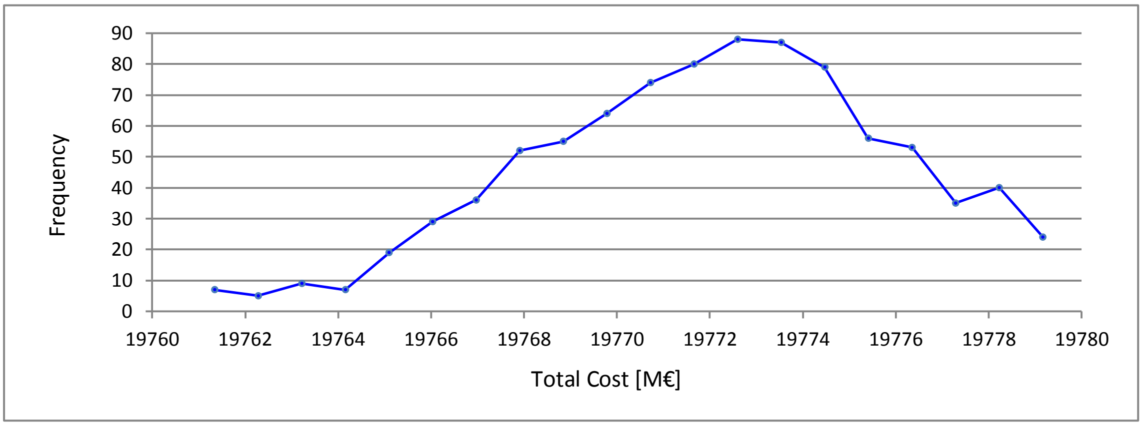


Figure 17: Candidate lines distribution in case study 1. Success probability of 50% in the initial population.

Figure 18 represents the initial generation distribution varying the probability of installing candidate lines. In the three first functions, it is represented the 90% percentile of the population. In the distribution with 30% of success probability the 35 percentile has been used for illustrating the relevant range of costs.

It can be observed that as the success probability decreases the distribution moves to a lower costs region. However a lower probability of installing candidate lines entails a more accused dispersion of costs, although the modal point is nearer to the optimal cost.

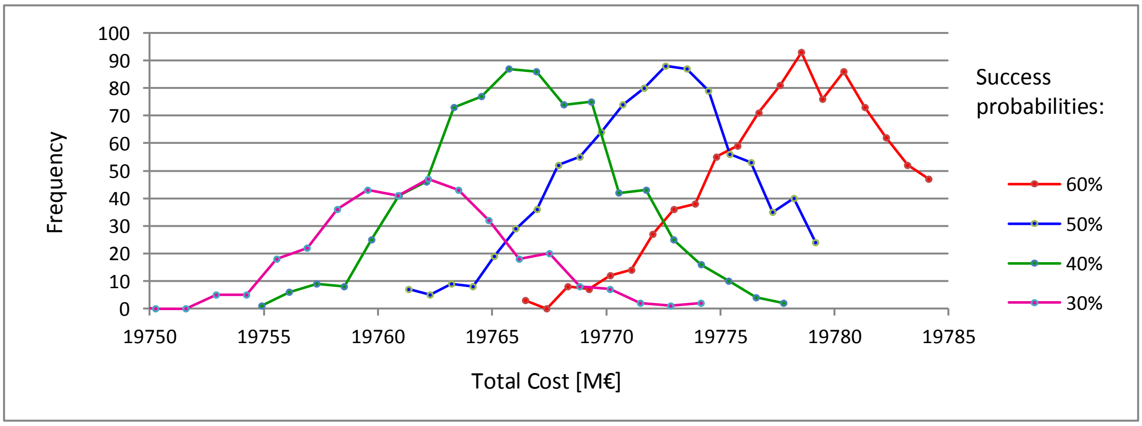


Figure 18: Candidate lines distribution in case study 2. Comparison with different success probabilities in the initial population.

5.1.3. *Case study 3*

The third case study considered has been constructed with 156 candidate power lines. The complete problem has been solved for obtaining the optimal plan. This optimal solution installs 45% of the candidate lines and has a total cost of 16853.906 M€.

In the following figure, where the Y axis represents the frequency and the X axis represents the total cost in M€, it can be observed that the density distribution of this problem is nearly triangular. The following figure represents the initial population distribution obtained installing a half of the candidate power lines. The mode is 17245M€ with a frequency of 95. This value is 2.32% higher than the optimal total cost.

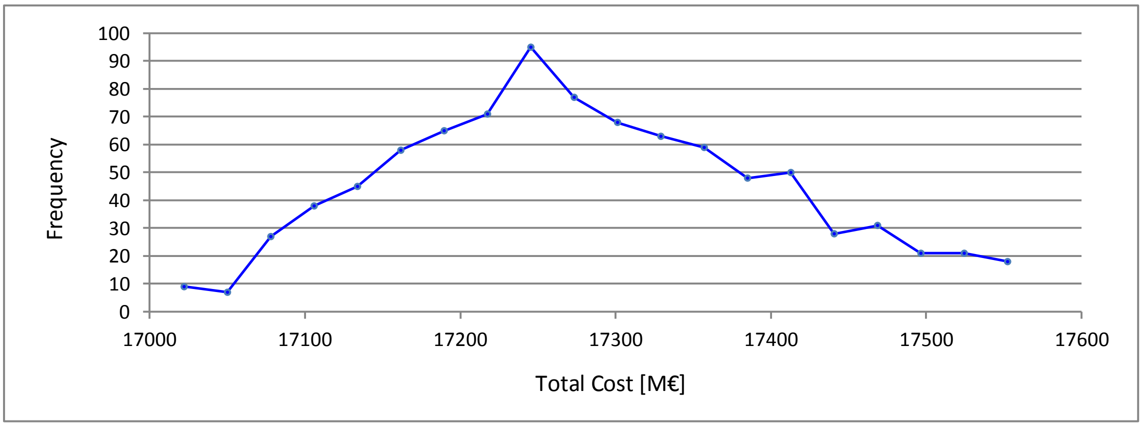


Figure 19: Candidate lines distribution in case study 2. Success probability of 50% in the initial population.

Figure 19 represents the 90% percentile. It can be observed that the interval of costs where the individuals are situated is quite narrow, 90% of the 1000 randomly generated individuals have total costs between 17000 M€ and 17600 M€.

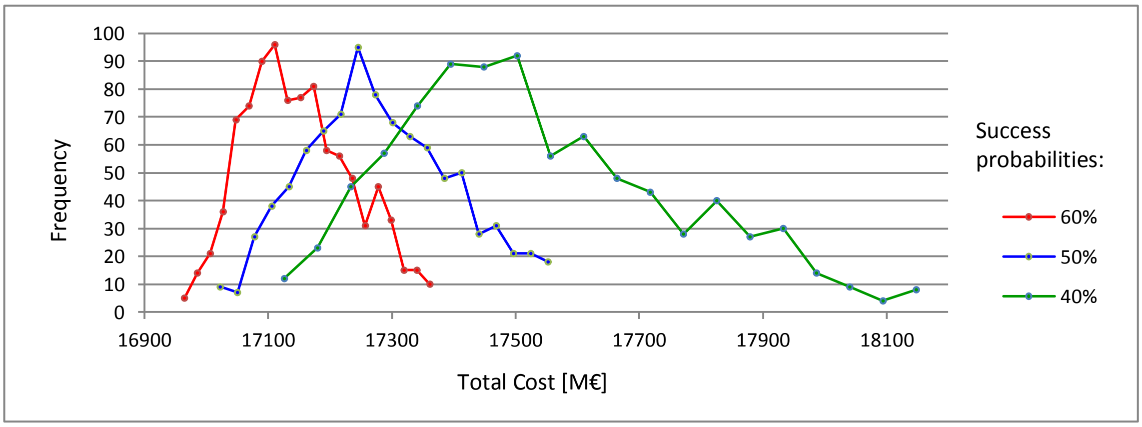


Figure 20: Candidate lines distribution in case study 2. Comparison with different success probabilities in the initial population.

The previous figure shows the variation of the initial distribution with different success probabilities. The distribution which installs 60% of the candidate lines represents the 95% of the initial generation and the distribution with 40% success probability represents the 85% of the initial population.

Just as the installation rate increases, the interval of costs in where the majority of the initial population is located decreases. As in case study 2, a lower probability of installing candidate lines means more dispersion in costs, although the modal point is nearer to the optimal cost.

5.2. Influence of the population size

The aim of this section is to find the most suitable value of the population size for the resolution of this case study and to study its influence for obtaining a good solution.

Several cases with different population sizes have been executed. The mutation probability was fixed at 1%. The convergence criterion consisted in fixing the total number of iterations at 150 or 200. In all these simulations the initial population has been generated with a success probability of 50%.

5.2.1. *Case study 1*

<u>*Population size of 100*</u>

This first execution has been carried out with a mutation probability of 1% and a population size of 100 individuals. Figure 21 shows all the individuals within the 150 iterations. With a red line the optimal plan cost is represented and the queen individual appears in blue.

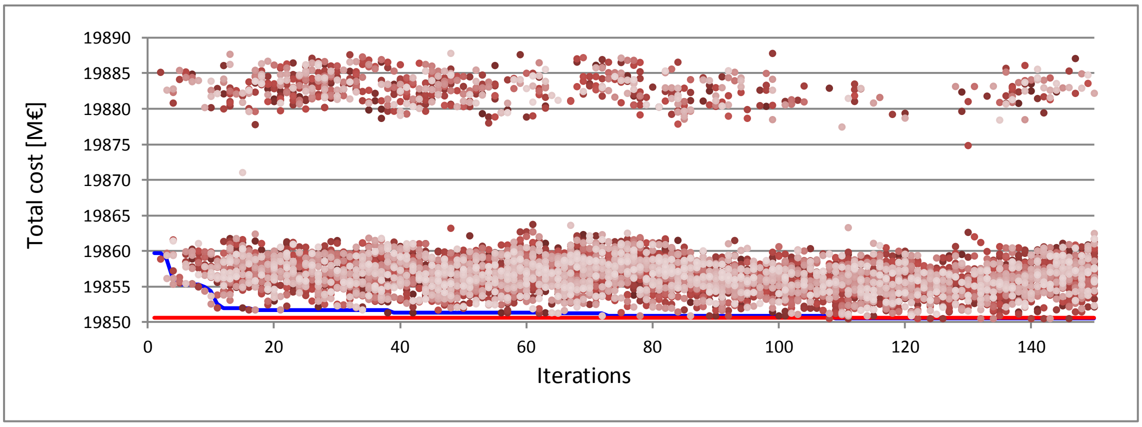


Figure 21: Evolution of case study 1 with 100 individuals.

Although the population stays reasonably homogeneous along the iterations, it can be observed that the best individual does improve.

It is interesting to observe that there are two clearly differentiable clusters, the one nearer the queen individual and the other one defined between 19888M€ and 19877M€.

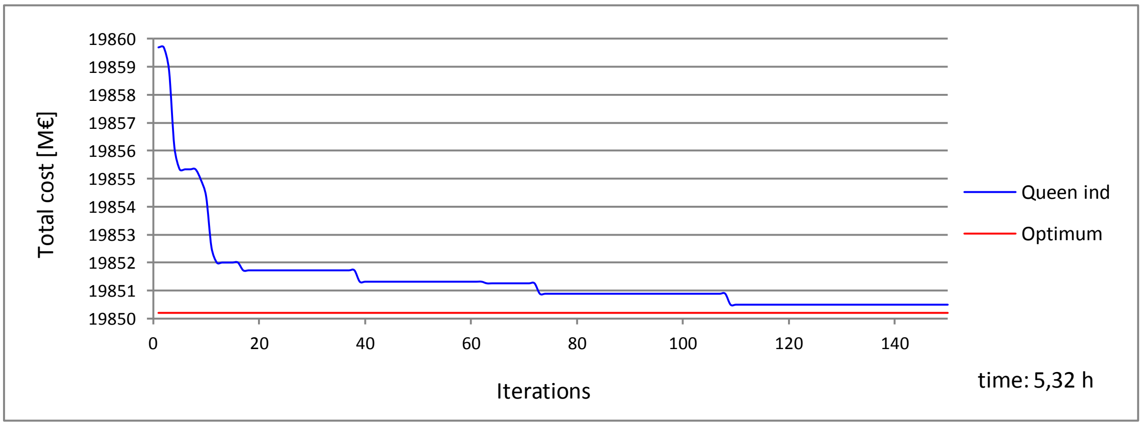


Figure 22: Queen individual in case study 1 with 100 members.

In Figure 22, the optimal solution calculated with the complete problem resolution and an optcr of 1e-4 is represented in red. The minimum cost found by the GA is 19850.5 M€, 0.001% higher than the optimal total cost. The GA achieves 0.046% for improvement since the beginning of the resolution process. This resolution has spent 5 hours and a half for running 150 iterations.

As the population does not actually improve in the last iterations, then it will be interesting to carry out the same simulation with a slightly higher mutation rate, in order to provide new genetic information.

Population size of 20

The next resolution has been carried out with the same parameters than in the previous case but with a population size of 20 individuals. All the population within the 200 iterations can be seen in the following figure.

The same cluster with costs between 19875 M€ and 19885 M€ that was shown in the case of using 100 individuals can be seen also in Figure 23.

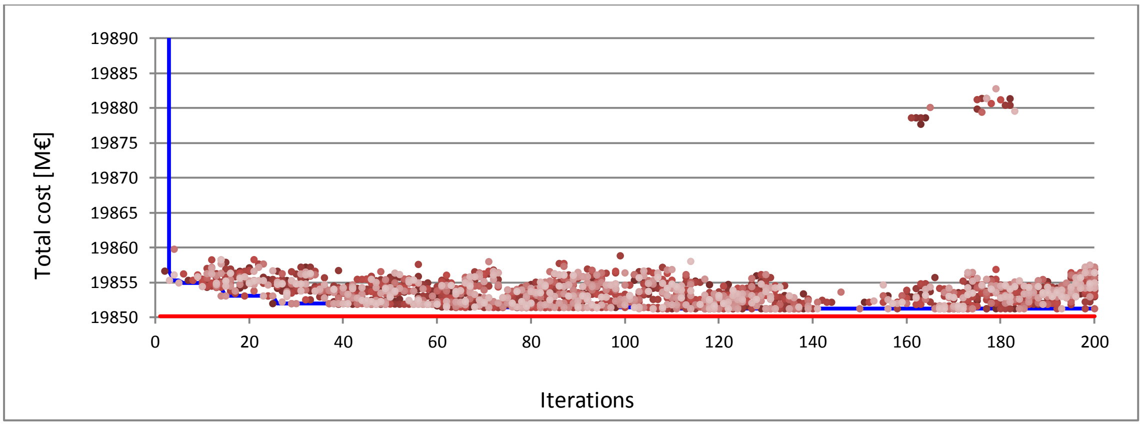


Figure 23: Evolution of case study 1 with 20 individuals.

The simulation spent 33 minutes for constructing 200 generations. At the beginning of the resolution the queen individual decreases with a steep slope. Since the generation 104 the problem gets stuck in 19851 M€. The percentage for improvement in this case is 10.78%.

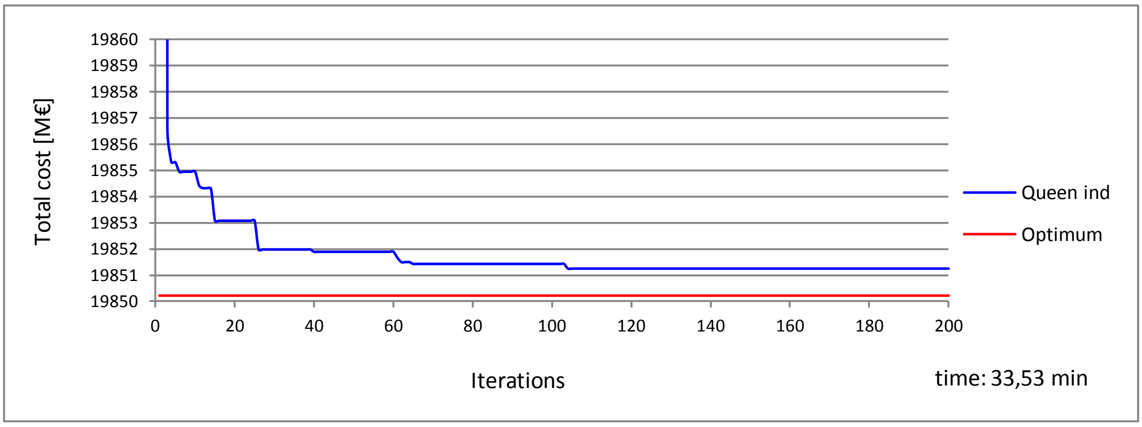


Figure 24: Queen individual in case study 1 with 20 members.

Population size of 10

The simulation with a fixed mutation rate of 1% was last tested with 10 individuals. Once more the cluster is manifested in Figure 24, where the simulation using a population size of 10 members is represented.

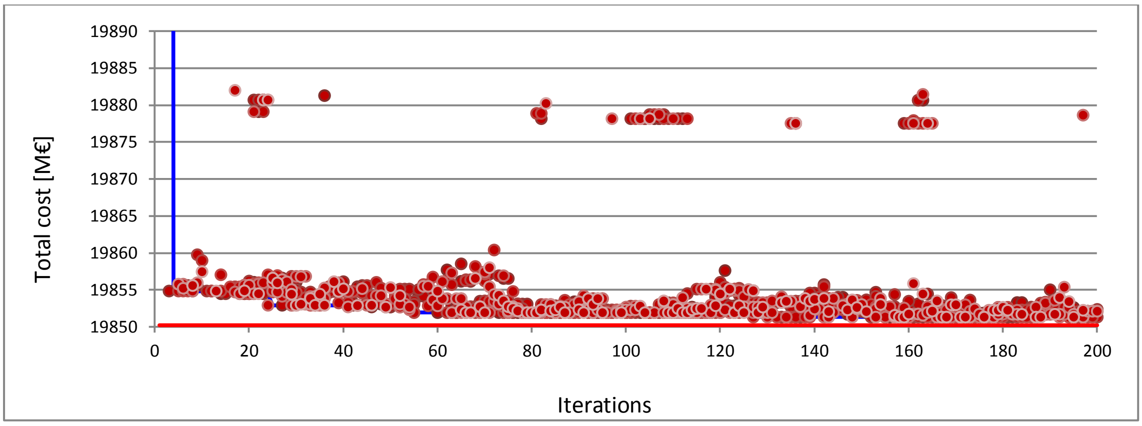


Figure 25: Evolution of case study 1 with 10 individuals.

The GA execution lasted for 31 minutes. Within the 200 generations the algorithm achieves an improvement of 10.78%. The best solution obtained installs one power line more than the optimal solution and its total cost is 0.005% higher than the cost of the optimal plan which was obtained solving the complete problem.

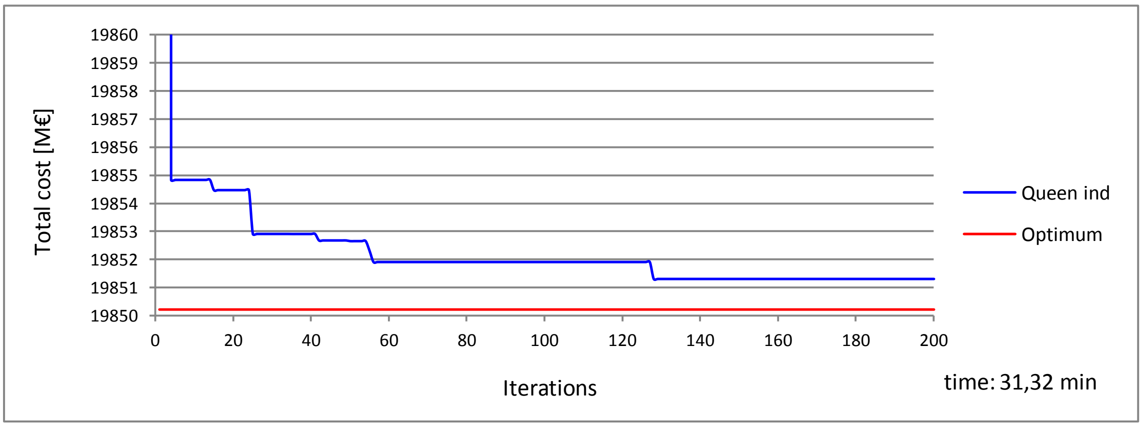


Figure 26: Queen individual in case study 1 with 20 members.

Figure 26 shows that the best solution obtained by the GA is found in the iteration 128. In the first 55 generations the cost evolution is more notable. Then the algorithm stagnates, and does not improve in the last 78 iterations.

Comparison

Figure 27 displays a comparison of the queen individuals obtained with different population sizes in the case study to test its influence in the problem resolution. The mutation probability is 1% in all the represented cases.

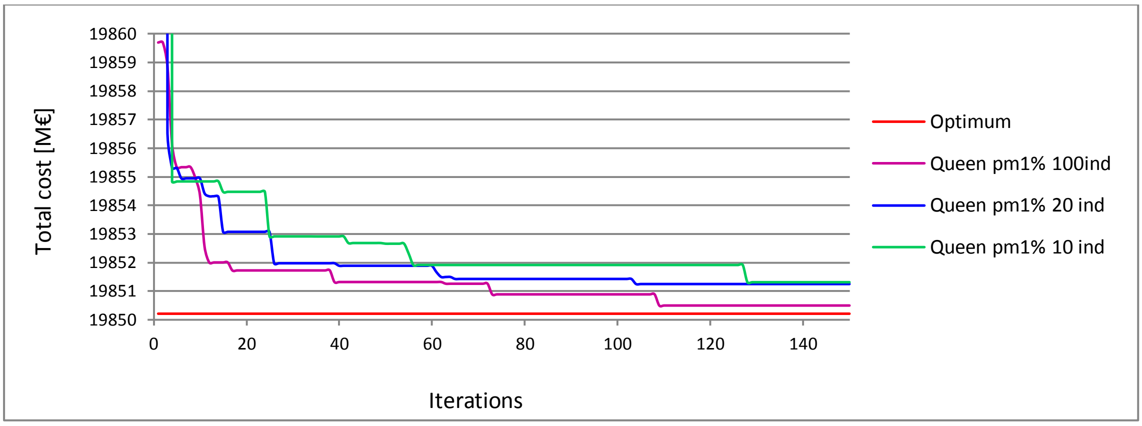


Figure 27: Comparison of the population size with mutation rate of 1% in case study 1.

The best solution is obtained with a population size of 100 individuals, even though in the case of using 10 individuals the solution reached is only 0.004% higher than with 100 individuals. Moreover, in the case of using 10 individuals the improvement achieved since the beginning of the simulation to the end is 10.73% higher than in the case of 100 members. In conclusion it is suitable to use a population size of 10 individuals.

The same comparison than in the previous graphic has been made in Figure 28, but with a mutation probability of 5% in all the represented resolutions.

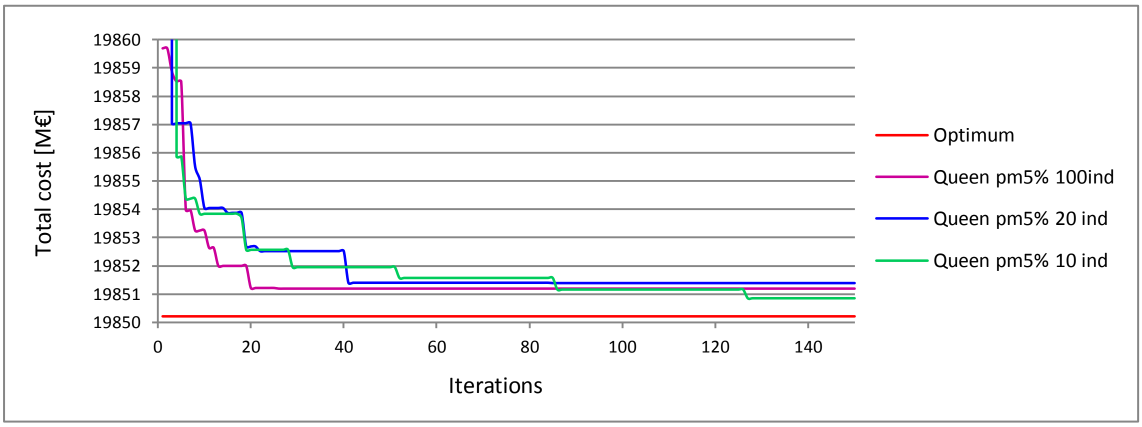


Figure 28: Comparison of the population size with mutation rate of 5% in case study 1.

In this case the best solution is achieved with a population size of 10 individuals. So with a mutation rate of 5% is clearly better to use a population size of 10 members than a higher one.

5.2.2. *Case study 2*

<u>*Population size of 100*</u>

The influence of the population size was then tested in case study 2. The mutation probability was also 1%. The number of iterations was limited to 150. The optimal solution in this case study, obtained with the complete problem resolution, installs 17 power lines with a total cost of 19740.97 M€.

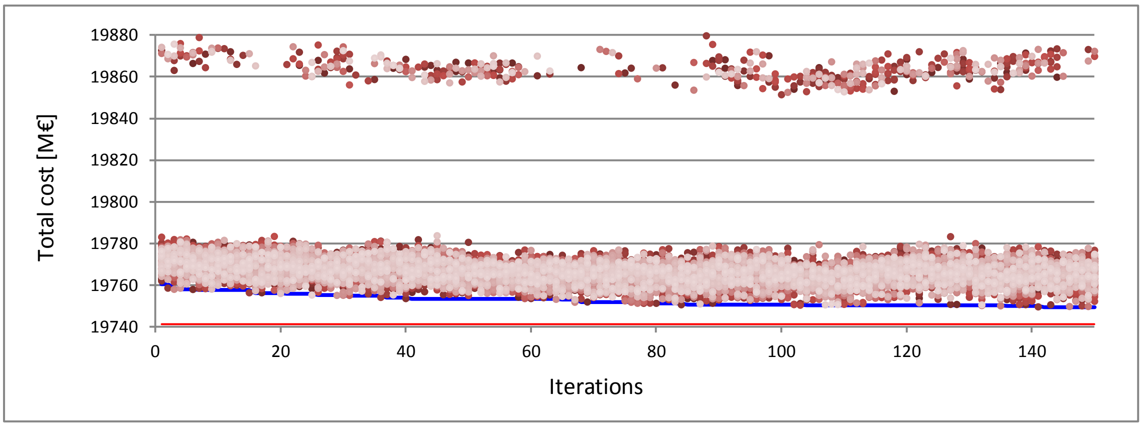


Figure 29: Evolution of the problem in case study2 with 100 individuals.

In Figure 29 a cluster appears between 19880 M€ and 19850M. Within the 150 iterations the population keeps relatively constant.

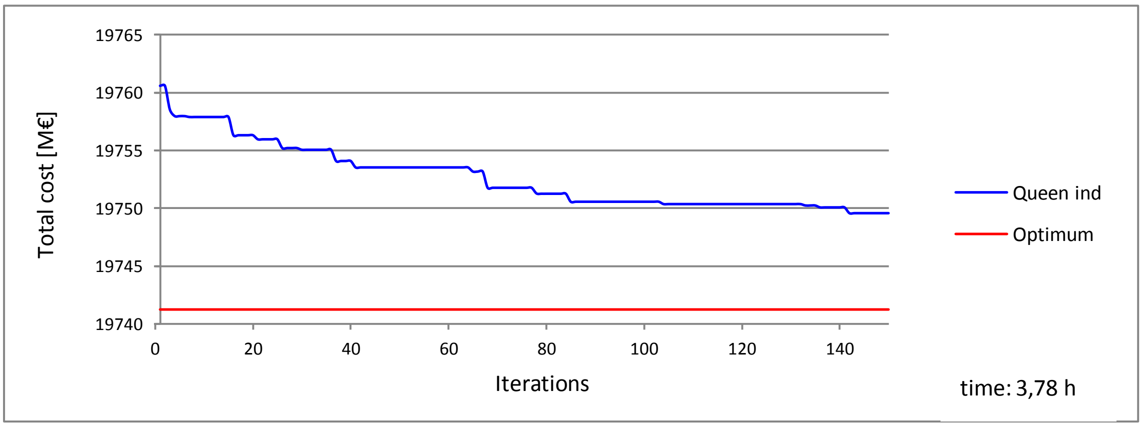


Figure 30: Queen individual in case study 2 with 100 members.

The queen individual experiments a gradual decrease, less accused in the lasts generations. After 3 hours and 47 minutes the queen individual has only reached 0.06 % for improvement with respect to the first generation. The final solution has a total cost of 19749.6 M€, with an error of 0.04 %.

Population size of 20

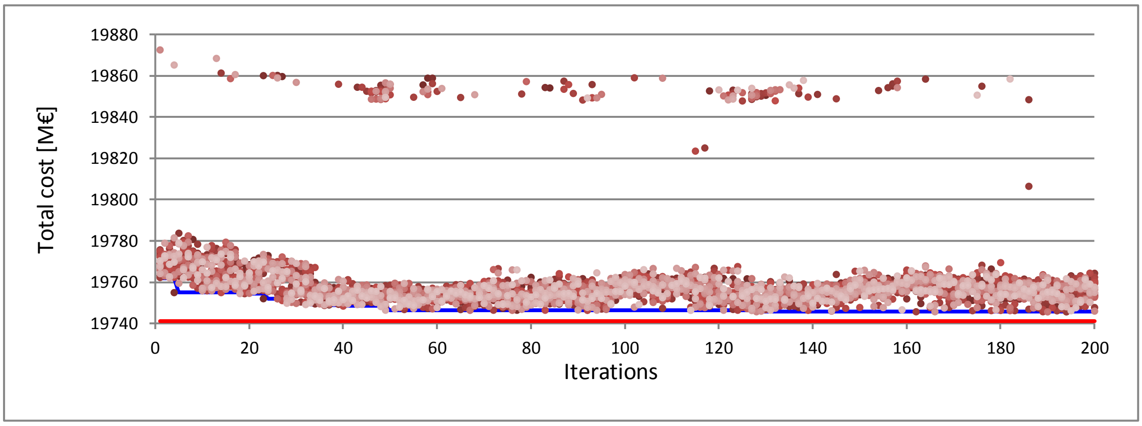


Figure 31: Evolution of the problem in case study 2 with 20 individuals.

Figure 31 is obtained solving case 2 with a population size of 20 individuals and a mutation rate of 1%. The cluster represented when fixing the population size at 100 individuals is also manifested in this figure. In the first 40 iterations a global decrease of the individuals' total costs is noticeable, following the queen individual trajectory. Then the population remains relatively constant.

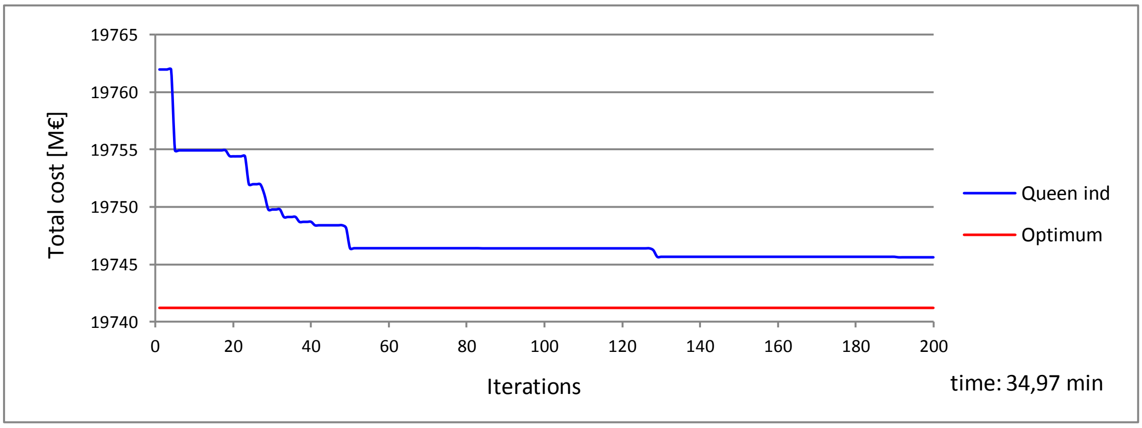


Figure 32: Queen individual in case study 2 with 20 members.

A notable decrease of the cost of the queen individual can be appreciated in the previous figure in the first 50 iterations, but this accused tendency is diminished in the next generations, getting stagnant in iteration 129. The execution spent 35 minutes for constructing the 200 generations, and the best solution obtained, with a total cost of 19745.6 M€, was 0.02 % higher than the optimal one. The improvement capacity reached in this simulation was 0.08 %, slightly better than with a population size of 100 individuals.

Population size of 10 individuals

The resolution of the same case study has been executed fixing the population size at 10 members. Figure 33 shows a slight total cost decrease of the whole population within the 200 generations. The cluster previously mentioned in the simulations of case study 2 with population sizes of 100 and 20 can also be seen in the following figure with 10 members.

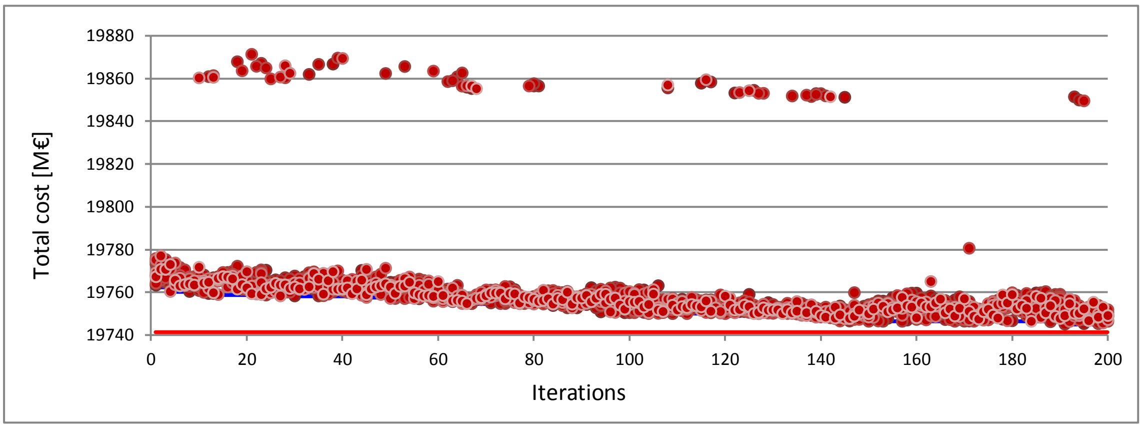


Figure 33: Evolution of the problem in case study 2 with 10 individuals.

The queen individual experiments a downspout trend within the 200 generations as it can be observed in Figure 34. In contrast with using population sizes of 100 and 20 members, in this case the queen individual does not get stagnant in the last generations. The GA obtained a solution with a cost of 19745.03M€ after 29 minutes, 0.02% more expensive than the optimal solution cost and 0.1% cheaper than the best cost found in the initial generation.

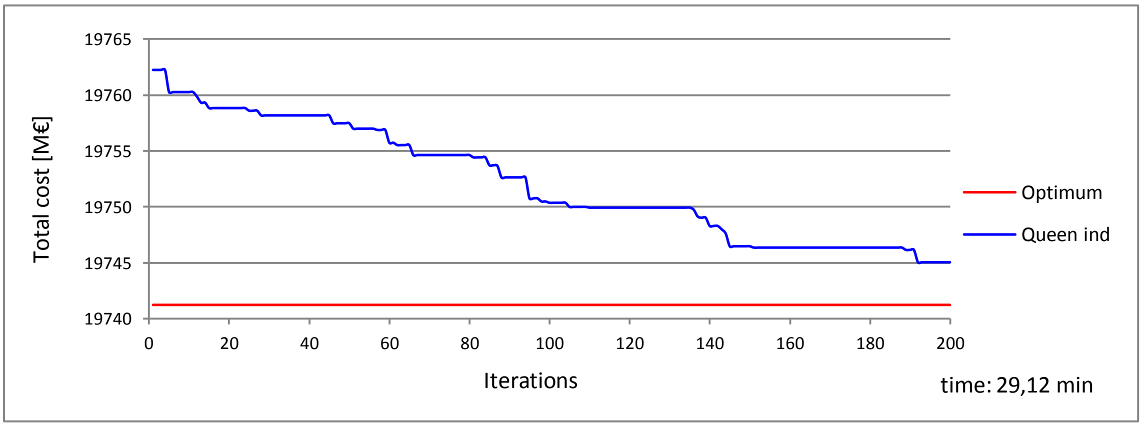


Figure 34: Queen individual in case study 2 with 10 members.

Comparison

As in the previous section, the following figures show a comparison of the use of different population sizes in case study 2. In Figure 35 the mutation probability has been defined as 1%.

The worst case is the one which considers 100 individuals and it is also the one which spent more time in the execution of the model. Not only its solution is the most expensive of all the represented but also is the case with worst improvement capacity, within the 200 iterations it only reaches 0.06% for improvement. On the other side using a population size of 10 or 20 individuals seems to be more suitable for this problem because they found chipper solutions with time savings and also they manifest a higher improvement capacity.

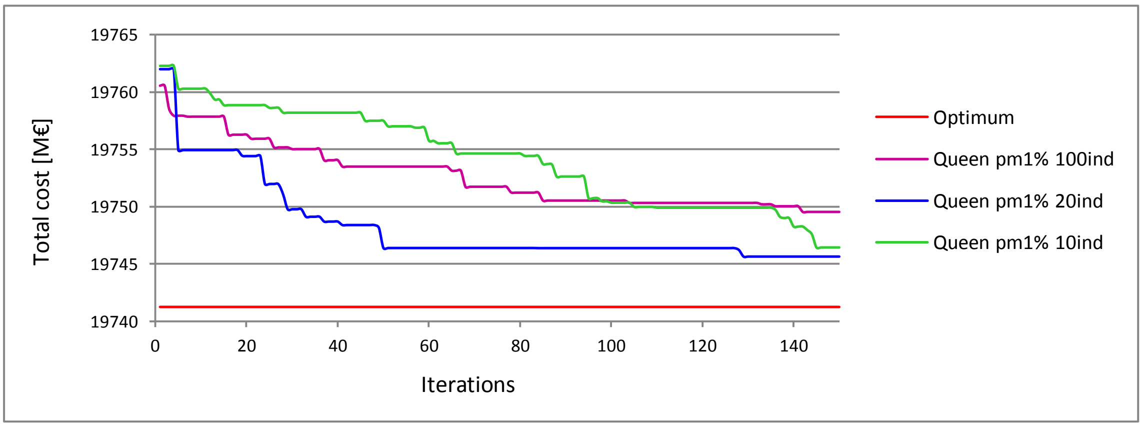


Figure 35: Comparison of the population size with mutation rate of 1% in case study 2.

In Figure 36, as in the previous graphic, a comparison between the queen individuals when using different population sizes has been represented but with a mutation probability of 5%. A great similarity between the three queen individuals can be seen in this figure. When using 100 members, the queen individual decreases in the first iterations for then staying slightly constant, contrasting with using 10 or 20 members. In these cases with a lower population size the queen individuals evolve more gradually to lower costs.

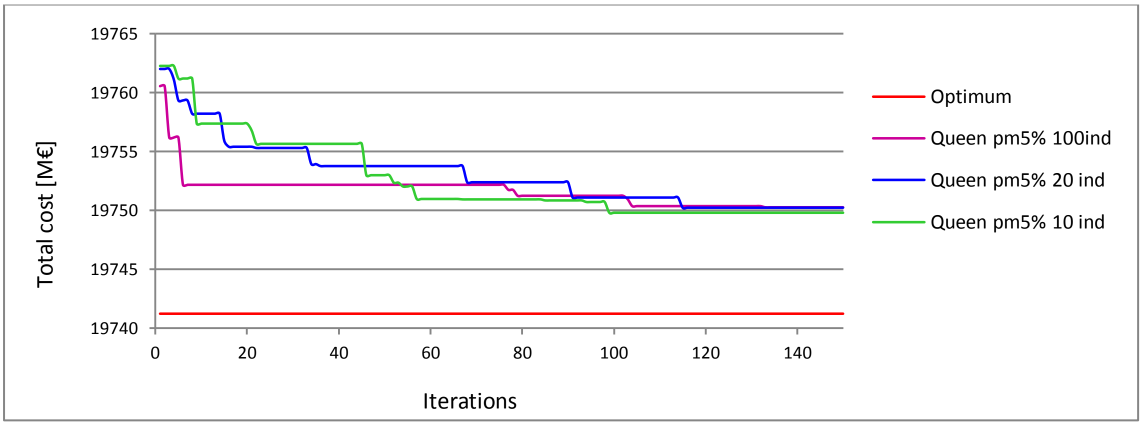


Figure 36: Comparison of the population size with mutation rate of 5% in case study 2.

5.2.3. *Case study 3*

Finally the same study that has been carried out in the previous case studies has been made with the third one. This case study composed of 1084 nodes and 294 power plants. The transmission network is configured by 1505 lines and 156 candidate lines. The mutation probability has been fixed at 1%.

Population size of 100 individuals

The Figure 37 has been constructed with a population size of 100 individuals. A global cost decrease tendency is observed in the population, following the queen individual trajectory.

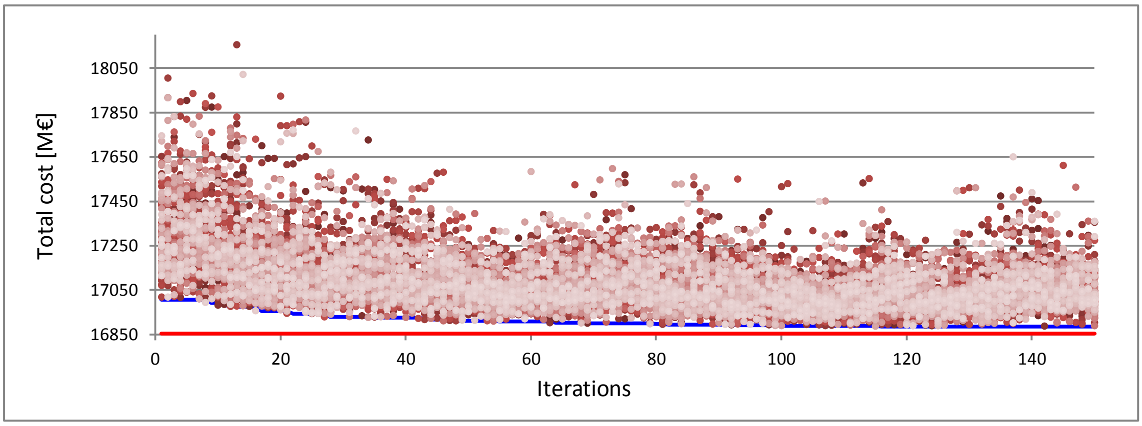


Figure 37: Evolution of the problem in case study 3 with 100 individuals.

In the following figure a gradually decrease, of the total cost of the best individual is shown. The solution obtained has a cost 0.2% more expensive than the optimal solution given by the resolution of the complete problem, which has a cost of 16853.9 M€. Running the simulation spent 7 hours with a convergence criterion of stopping at the iteration 200. The improvement achieved since the beginning to the end of the resolution amounts for 0.7%.

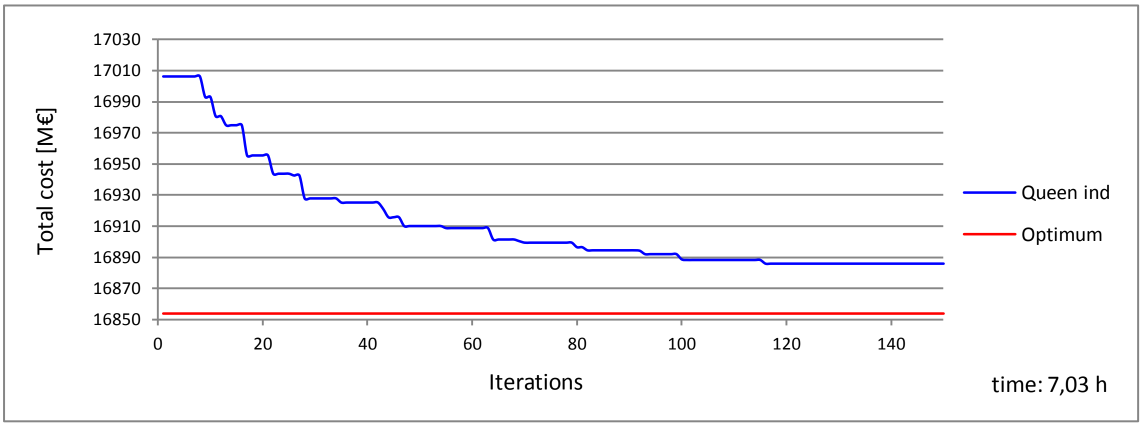


Figure 38: Queen individual with 100 members in case study 3.

Population size of 20 individuals

The evolution of the population fixing its size at 20 members is represented in Figure 39. The decrease of the whole population total cost can be observed, following the queen individual evolution.

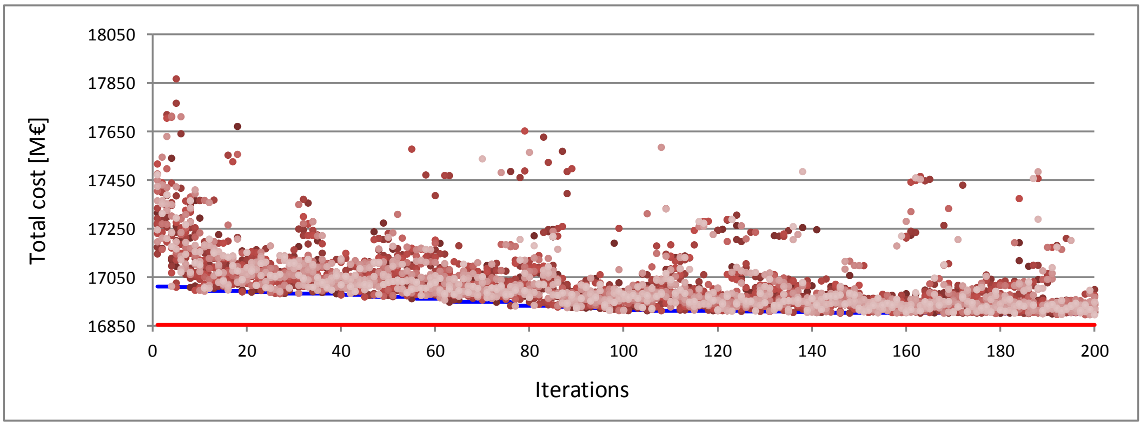


Figure 39: Evolution of the problem with 20 individuals in case study 3.

The gradual decrease of the queen individual total cost finishes with a solution 0.3% more expensive than the optimal one that provides the resolution of the complete problem. The capacity for improvement amounts to 0.7%, that means that the cost of the queen individual in the initial iteration is 16896.9 M€ and in the 200 generation is 17013.8 M€.

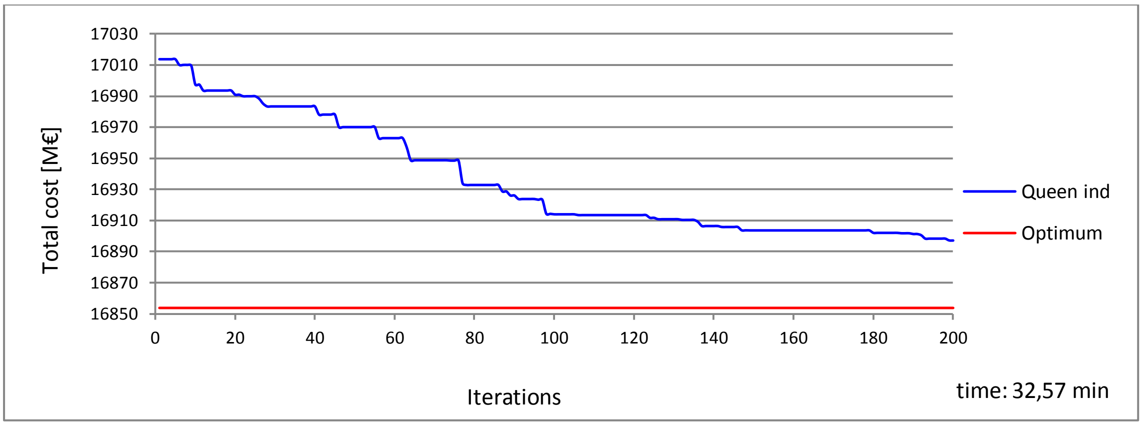


Figure 40: Queen individual with 20 members in case study 3.

Two parts in the queen individual evolution can be observed. Since the beginning to the generation 100 the queen individual cost decreases with a steeper slope than in the second part. The simulation of the 200 generations lasted for 33 minutes.

Population size of 10 individuals

For finishing the study of the influence of the population size in the case study that considers 156 candidate power lines, a population of 10 individuals has been defined for obtaining the simulation represented in the following figure.

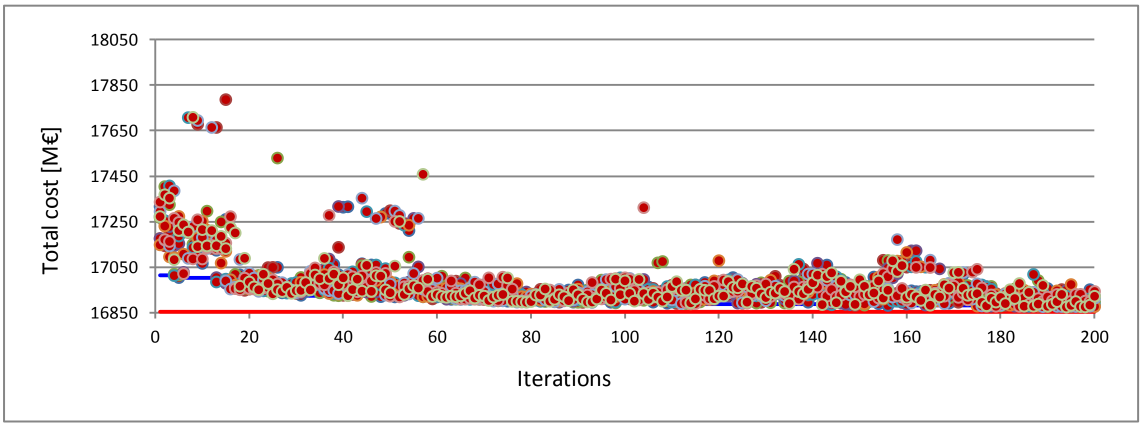


Figure 41: Evolution of the problem with 10 individuals in case study 3.

In the first generations a decrease of the total cost is observed, but since the 20 iteration this decrease tendency disappears. This outcome can be seen also in the queen individual cost evolution. In the first iterations the queen undergoes more accused changes than in the lasts generations.

The simulation lasted for 16 minutes and the solution cost amounts for 16873.8 M€, 0.1% higher than the optimal cost calculated by the complete problem. The improvement capacity in this case is 0.8%, higher than when using population sizes of 20 or 100 individuals.

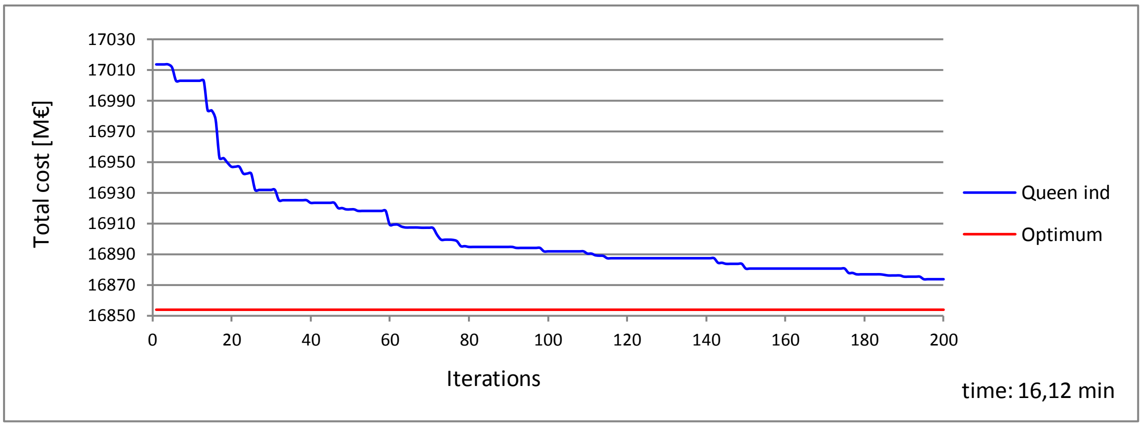


Figure 42: Queen individual with 10 members in case study 3.

<u>Comparison</u>

To conclude with the study of the influence of the population size for the solution obtaining a comparison of all the simulations carried out in the case study 3 is represented in the following figures. The first figure has been constructed with a mutation rate of 1%, and the second one defining a mutation probability of 5%.

The queen individual when using a population size of 20 members is significantly different from the other two, as it can be seen in Figure 43. The best cost obtained with the simulation of the GA with 20 members is 0.3% higher than the optimal cost.

The queen evolution and the solutions provided by the resolutions with 10 and 100 individuals are very similar, 0.17% more expensive than the optimal plan.

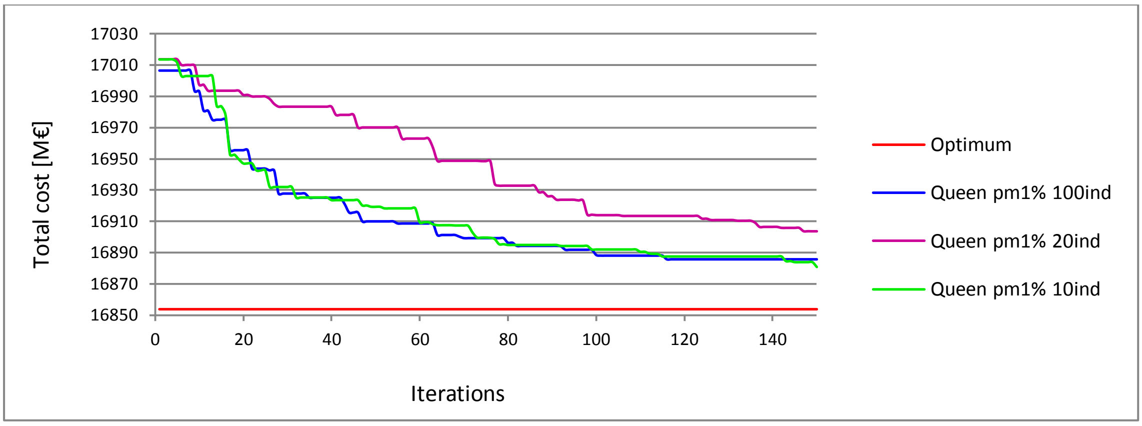


Figure 43: Comparison of the population size with mutation rate of 1% in case study 3.

In contrast with the gradual evolution of the costs in the previous figure, in this one irregular decrease of the costs is shown. In this case study, instead of what the mutation is thought to do, it causes cost stagnation in the queen individual evolution.

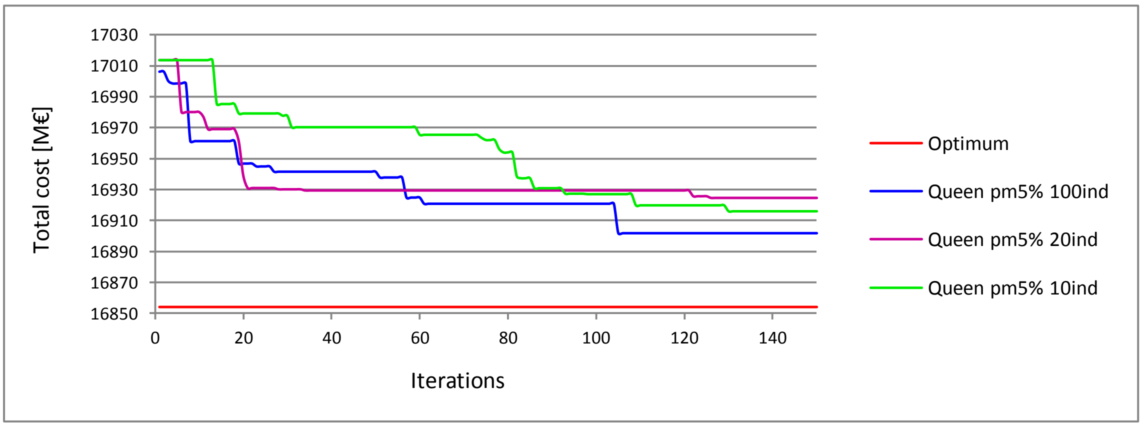


Figure 44: Comparison of the population size with mutation rate of 5% in case study 3.

5.3. Influence of the mutation rate

The aim of this section is to study the influence of the mutation rate in the GA simulation. For fulfilling this task the three case studies have been simulated with different mutation rates, keeping the population size constant. The initial population has been generated with a successful probability of 50% in the binomial distribution.

5.3.1. *Case study 1*

A population size has been fixed in the case study with 53 candidate power lines while the mutation rate has been modified in order to study its influence within the generations.

Population size of 100 individuals

First the GA has been tested with a population size of 100 individuals. In Figure 45 the evolution of the population within the generations is represented for different mutation rates, from top to the bottom: 1%, 5% and 10%.

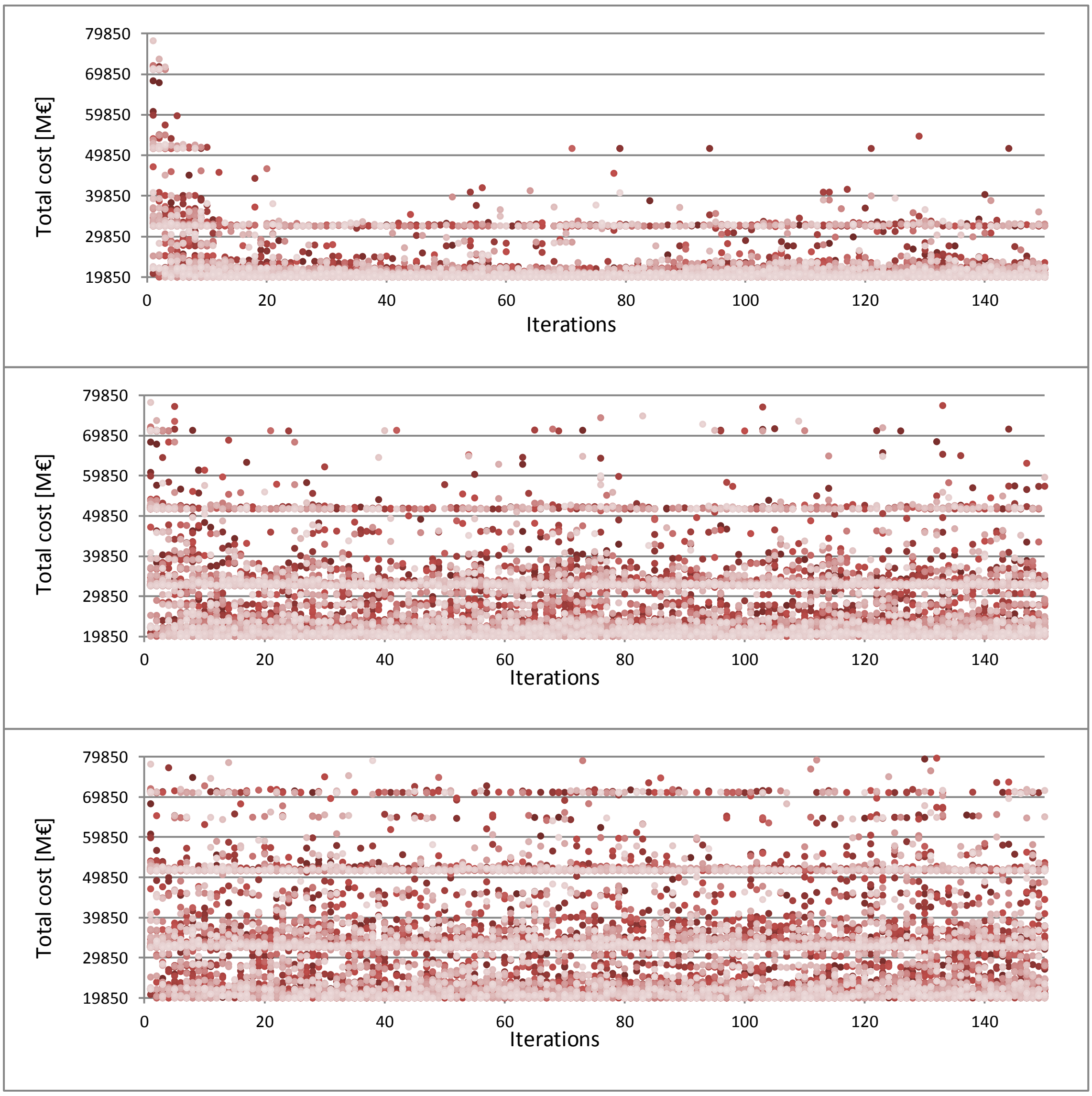


Figure 45: Evolution of the population in case study 1 with 100 members and mutation rates of 1,5,10%.

As the mutation rate increases new clusters appear. Moreover the clusters also acquire more density of population.

In the two last graphic four clusters are clearly identified, the one which is nearer to the optimal cost region, 2985 M €, 4985 M€ and the last one with a cost of 69850 M€. In the first representation of the generations the second cluster is not appreciable, because the mutation rate is not high enough.

All the clusters first appear in the initial generation regardless of whether mutation rate is defined. **Figure 46** shows the population evolution when using a mutation rate of 10% with the distribution function with a success probability of 50% in the initial population.

Then the higher ones disappear once the genetic process is being carried out and the mutation rate is low enough. In this way the mutation probability acts as a filter of clusters. A lower mutation rate diminishes the clusters apparition. So it seems that the clusters are not only related with the nature and topology of the case study but directly related with the mutation probability.

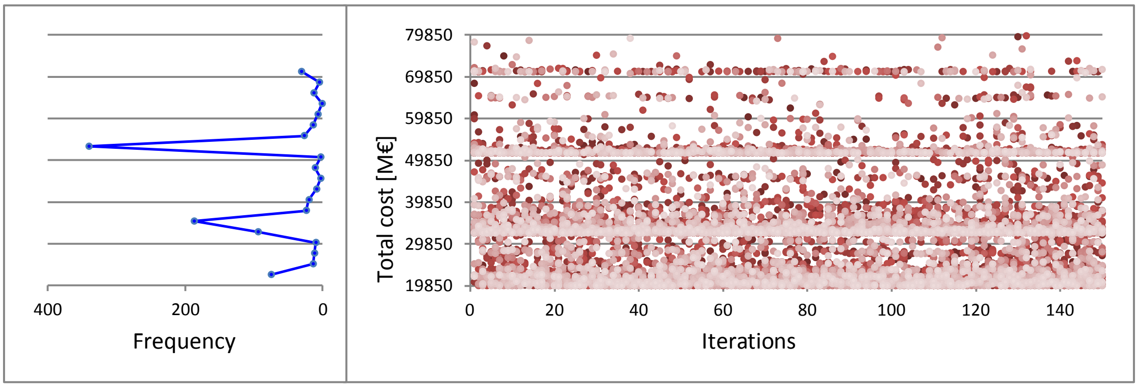


Figure 46: Clusters in case 1 with 100 members.

The next figure shows the rising of the total cost average of the population, which is another significant effect that comes with the mutation rate increase.

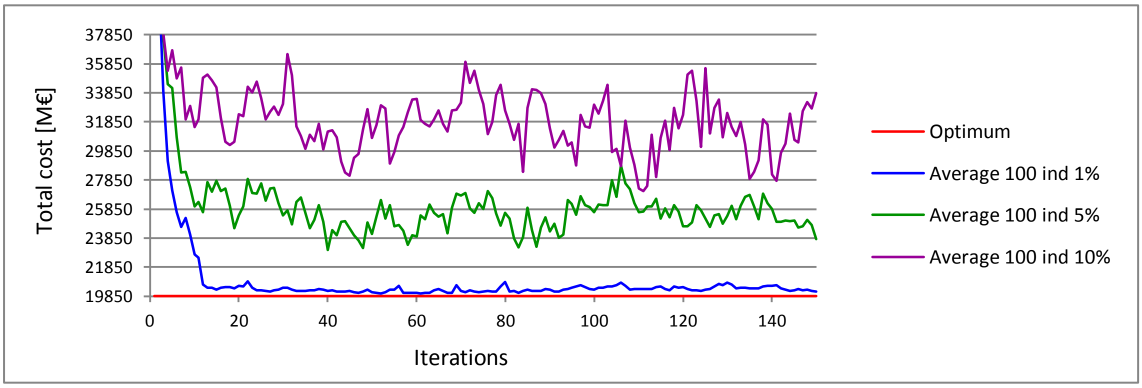


Figure 47: Cost average with a population of 100 individuals in case study 1.

When defining a mutation rate of 10% the cost population average amounts to a mean value of 32072.84 M€, while when using a mutation probability of 1% the main cost value is 21013.5 M€. Moreover the standard deviation of the costs average also decreases when the mutation rate do the same.

Figure 48 shows that when the mutation probability raises, the mean value of the costs average and its dispersion do the same.

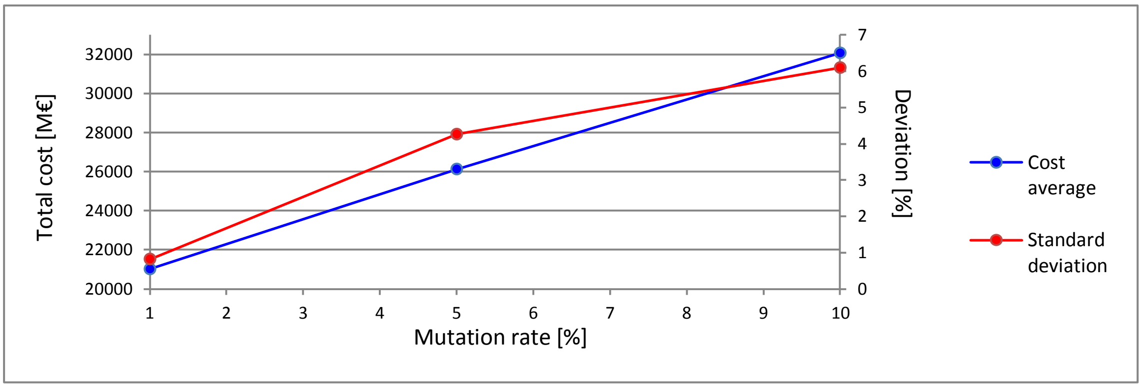
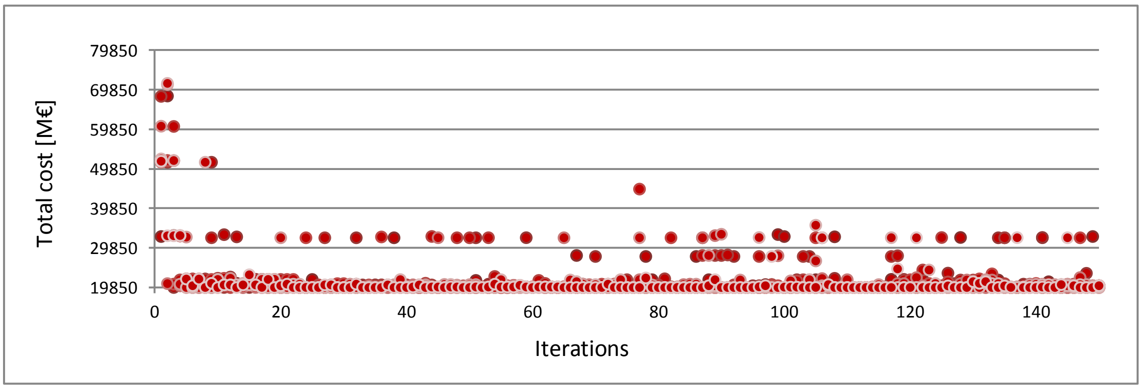


Figure 48: Relationship between the cost's mean average and the mean standard deviation with the mutation rate. Population size of 100 members in case study 1

Population size of 10 individuals

Figure 49 displays the same clusters that appeared in the resolution with a population size of 100 members. It is interesting to note that the number of clusters and its density are closely related to the mutation rate and not to the population size.



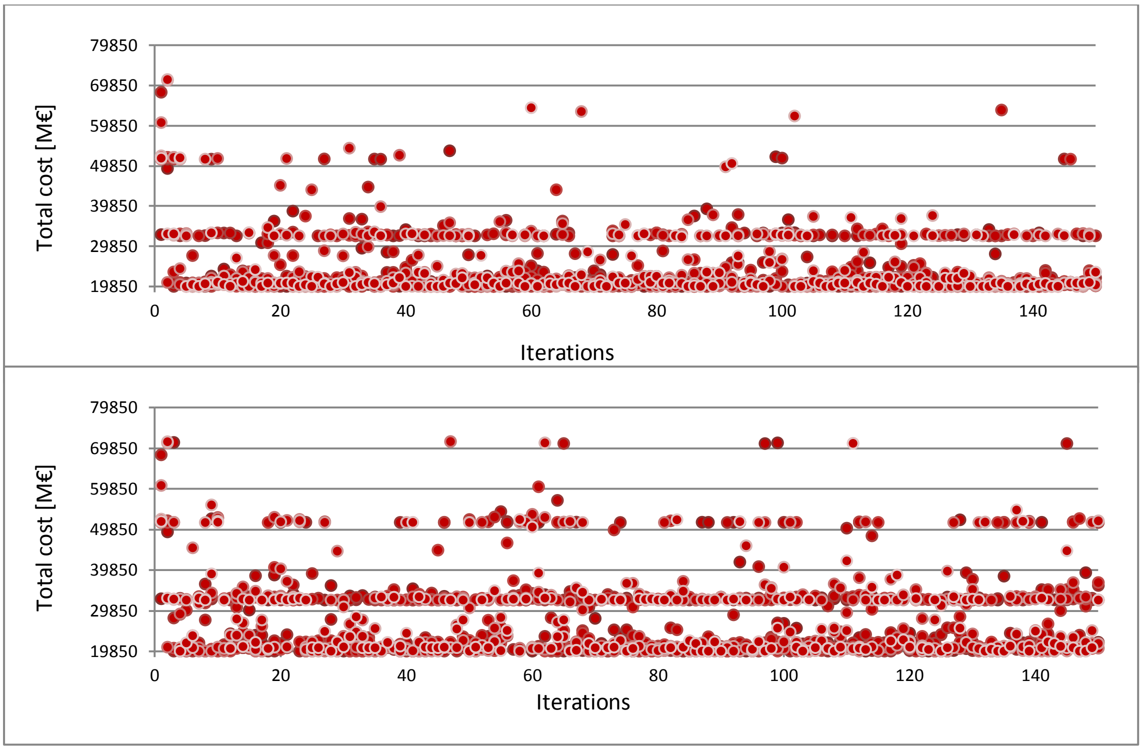


Figure 49: Evolution with 10 members and mutation rates of 1;5;10% in case study 1.

In this case the mean of the costs average of the population and its standard deviation also rise with the mutation rate increase.

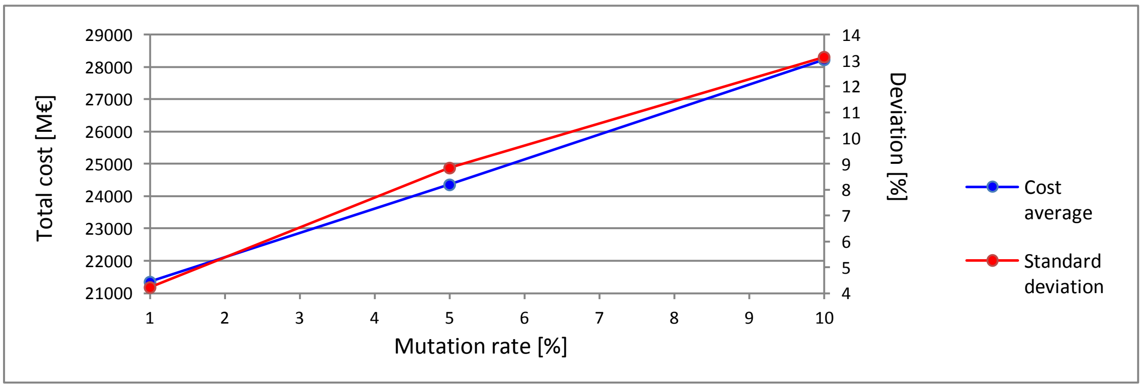


Figure 50: Relationship between the cost's mean average and the mean standard deviation with the mutation rate. Population size of 10 members in case study 1.

Comparison

To show how the mutation rate influences in the best solution obtained by the GA, a comparison varying the mutation rate while keeping the population size constant has been represented in Figure 51. The best solution is found using a mutation rate of 1%, which has been enough to avoid a premature convergence to a local optimum.

The queen individual in the last generation when using a mutation rate of 1% has a total cost of 19850.5 M€ only 0.001% higher than the optimal cost given by the resolution of the complete problem. The total costs found when using a mutation rate of 5% or 10% amounts to 19851M€. In the case of using a mutation rate of 10% at the beginning of the simulation the queen individual cost decreases more quickly than when using lower mutation rates, but unlikely to what it could be thought, the population gets stagnant since the 20 iteration.

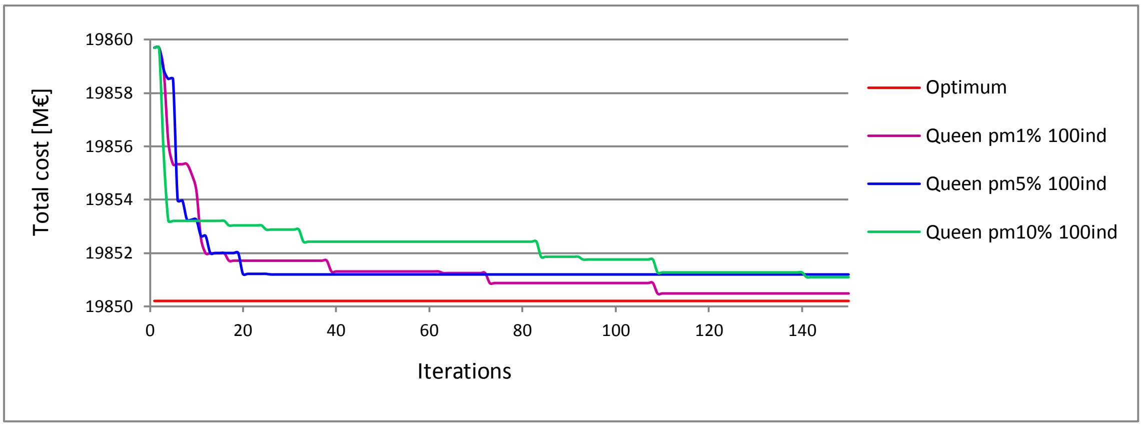


Figure 51: Comparison of the mutation rate with a population size of 100 in case study 1.

Figure 52 represents the same comparison but with a population size of 10 individuals. As in the previous figure, using a mutation rate of 10% is not suitable for this problem because the queen individual gets relatively stuck since the 10 generation. In this case using a mutation rate of 5% provides the best result when using a population size of 10 members. This solution is very similar to the best solution provided by the GA with 100 individuals and a mutation probability of 1%.

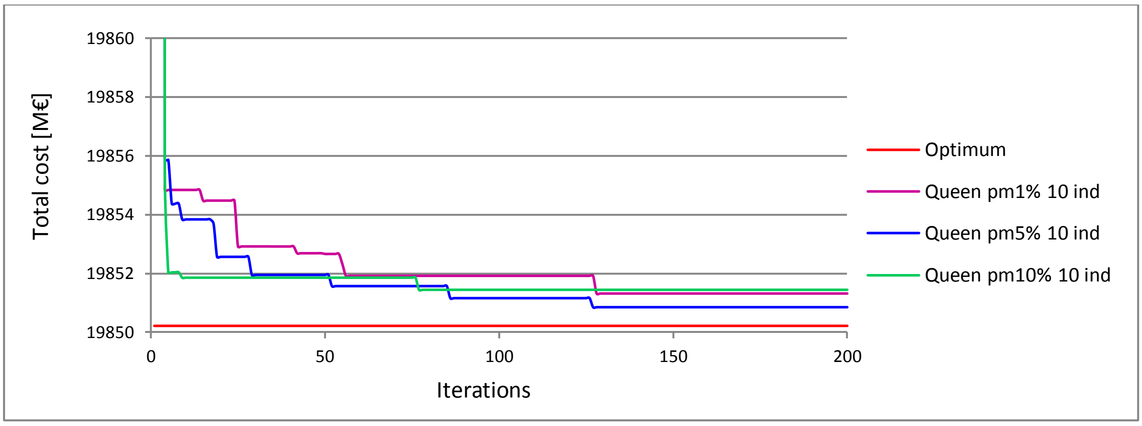


Figure 52: Comparison of the mutation rate with a population size of 10 in case study 1.

5.3.2. *Case study 2*

In order to test the mutation rate influence in the genetic resolution, the same analysis has been carried out in the second case study based on the spanish one and constituted by 90 candidate power lines.

Population size of 100 individuals

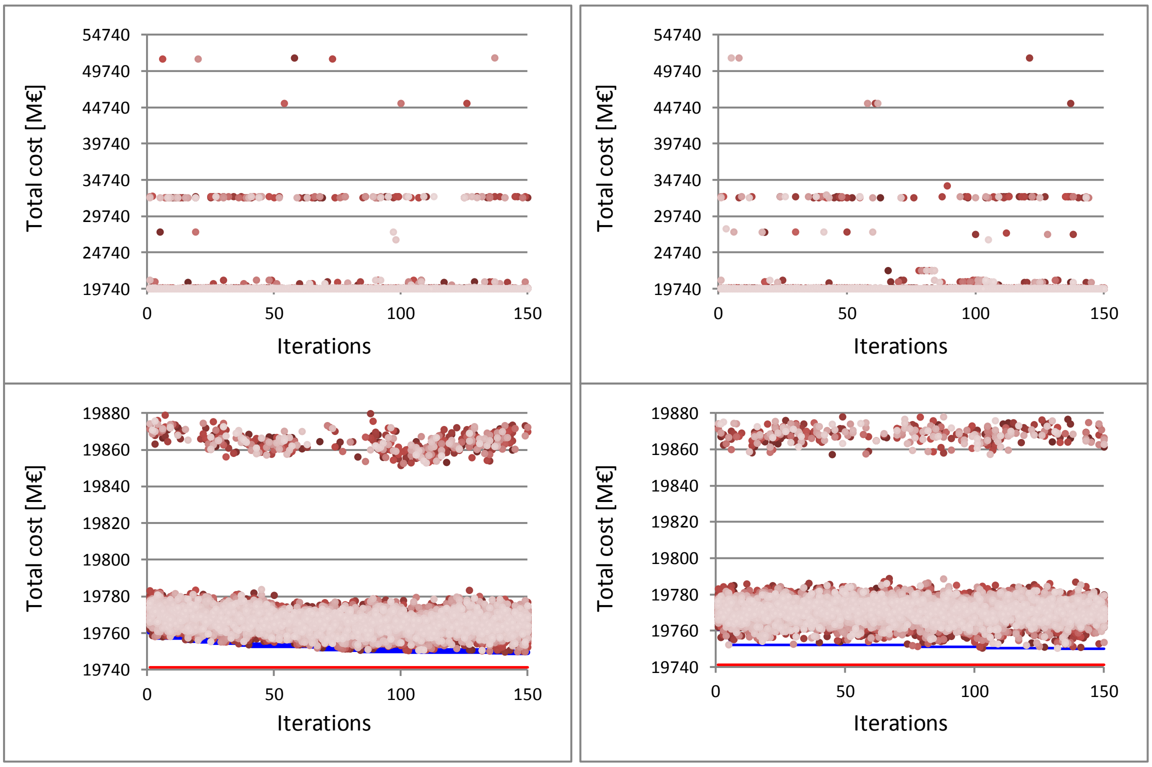


Figure 53: Evolution with 100 members and mutation rates of 1 and 5% in case study 1.

The previous figure confirms that the clusters become thicker with higher mutation rate. The figure on the left shows the resolution process with a mutation probability of 1%, and on the right with a mutation rate of 5%.

Figure 54 shows the population costs average with the queen individuals obtained in the resolution of case study 2. To the extent that the mutation rate grows the costs average does the same.

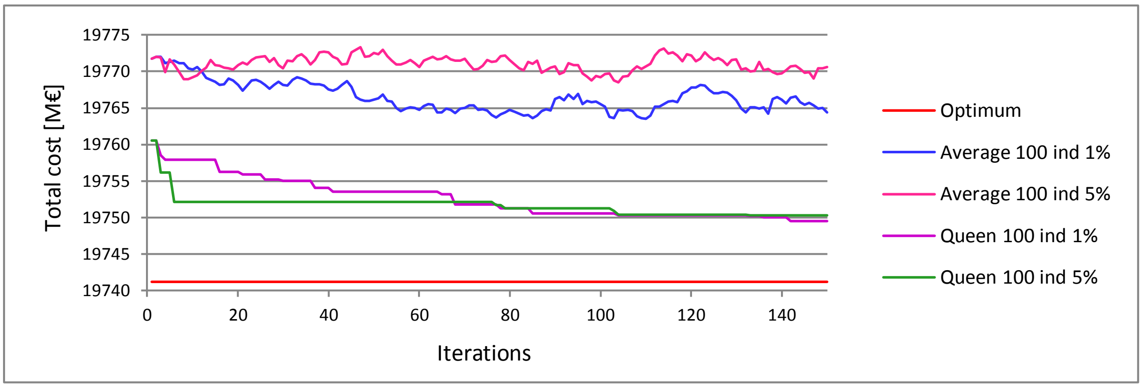


Figure 54: Total costs' average with a population size of 100 members and different mutation rates in case study 1.

Population size of 10 individuals

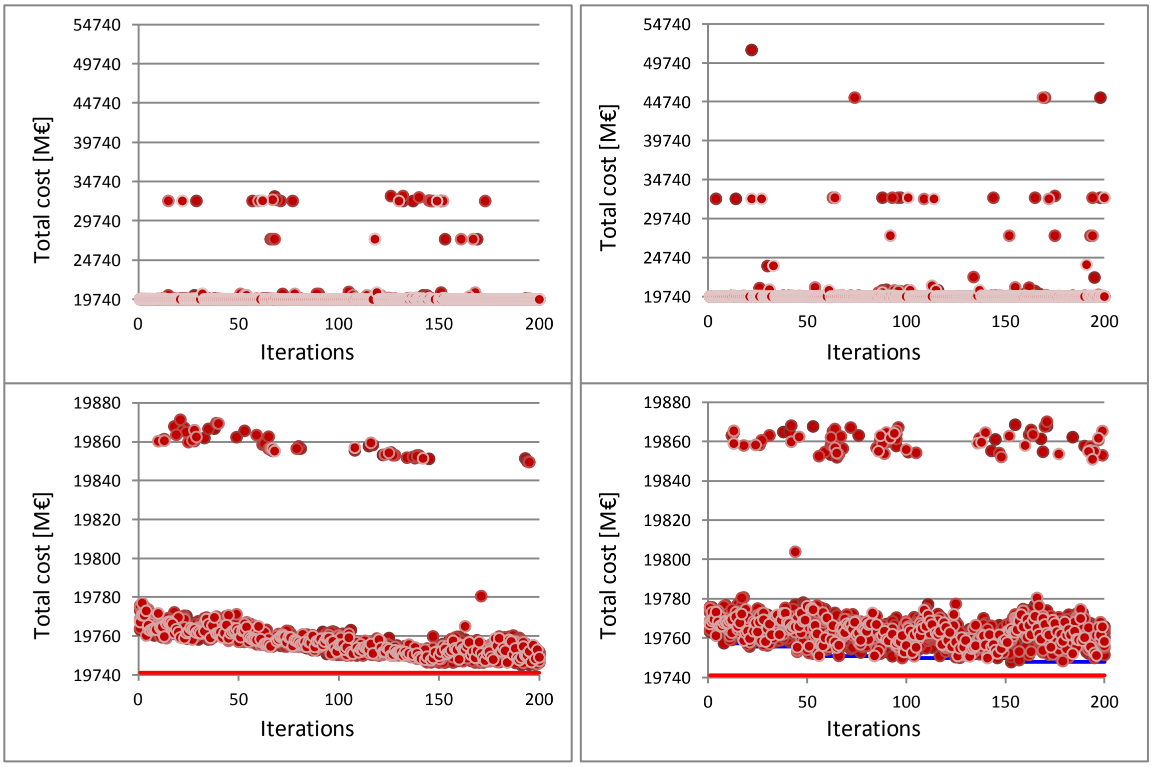


Figure 55: Evolution with 10 members and mutation rates of 1 and 5% in case study 2.

The clusters density grows with the mutation probability. In the previous figure is shown another effect of the mutation rate. The population gets closer to the queen individual. This can be clearly seen in the figure on the bottom-left.

Comparison

Figure 56 represents the queen individuals when using different mutation rates. At the beginning of the iterative process, a higher mutation rate provides a faster response of the algorithm, nevertheless using a mutation rate of 1% gives a better solution at the end of the 150 generations.

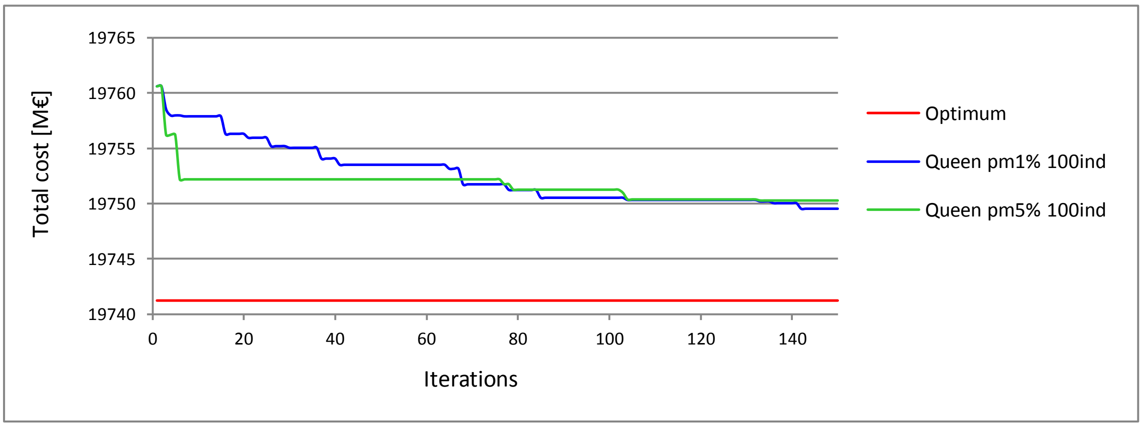


Figure 56: Comparison of the mutation rate with a population size of 100 in case study 2.

When using a population size of 20 individuals with a mutation rate of 1% there is a faster decrease of the queen individual total cost. It provides a better solution than when using a mutation rate of 5%, only 0.02% more expensive than the optimal planning cost.

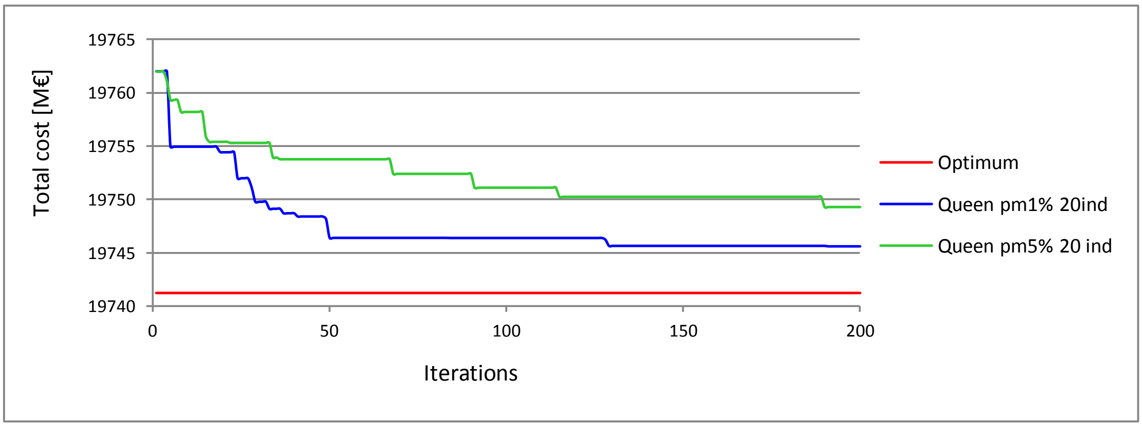


Figure 57: Comparison of the mutation rate with a population size of 20 in case study 2.

Figure 58 represents the same comparison as in the previous graphics but with a population size of 10 individuals. As in the case of using a population size of 100 members a faster response in the queen individual cost can be observed when using a mutation rate of 5%. However, in the last 45 iterations this tendency gets stagnant. The best cost is obtained with a mutation rate of 1%, which is 0.02% more expensive than the optimal cost provided by the resolution of the complete problem.

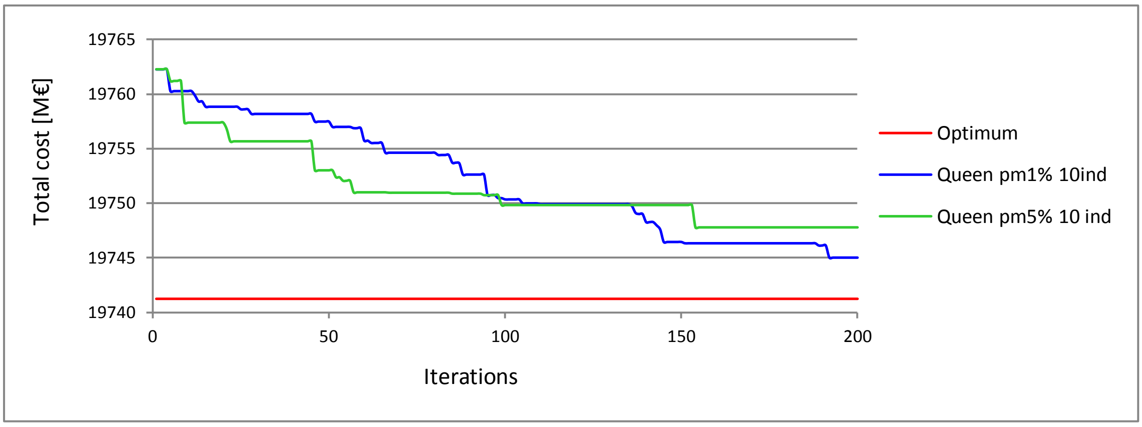


Figure 58: Comparison of the mutation rate with a population size of 10 in case study 2.

5.3.3. *Case study 3*

Finally to test the influence of the mutation rate in the GA resolution, case study 3 has been carried out keeping the population size constant.

The transmission network of this study case is configured with 156 candidate lines, and the plan obtained by the complete problem resolution installs 72 lines with a total cost of 16853.906 M€.

Population size of 100 individuals

The following figure shows the evolution of the 100 individuals within 150 iterations with mutation rates of 1%, on the left and 5% at the right. With both mutation rates two clusters can be observed. Nevertheless, when using a mutation rate of 5% the clusters are thicker, as it can be seen in the graphic on the top-right.

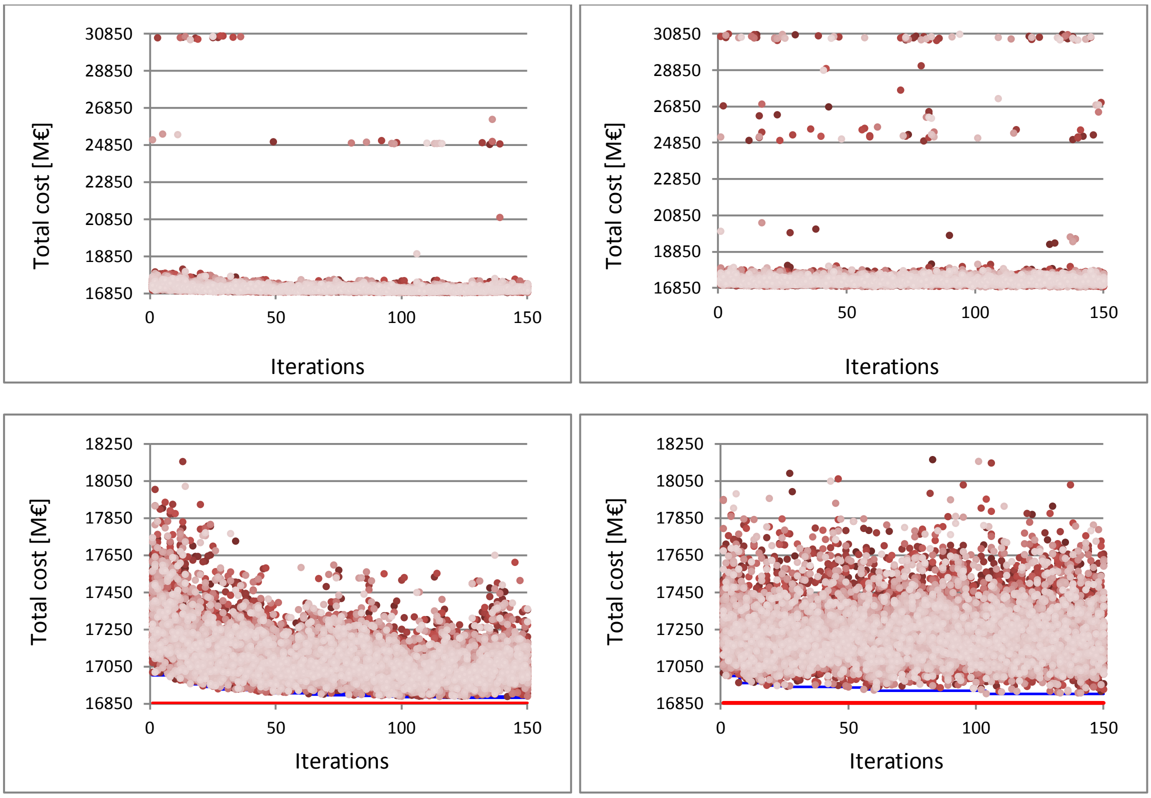


Figure 59: Evolution with 100 members and mutation rates of 1 and 5% in case study 3.

It is interesting to observe how the population closely follows the queen individual trajectory within the generations when using a mutation rate of 1% in contrast with the population costs in the case of using a mutation probability of 5%. That is shown in the figure on the bottom-right.

Figure 60 confirms this. The costs average of the population with a mutation probability of 1% is lower than with a mutation rate of 5%.

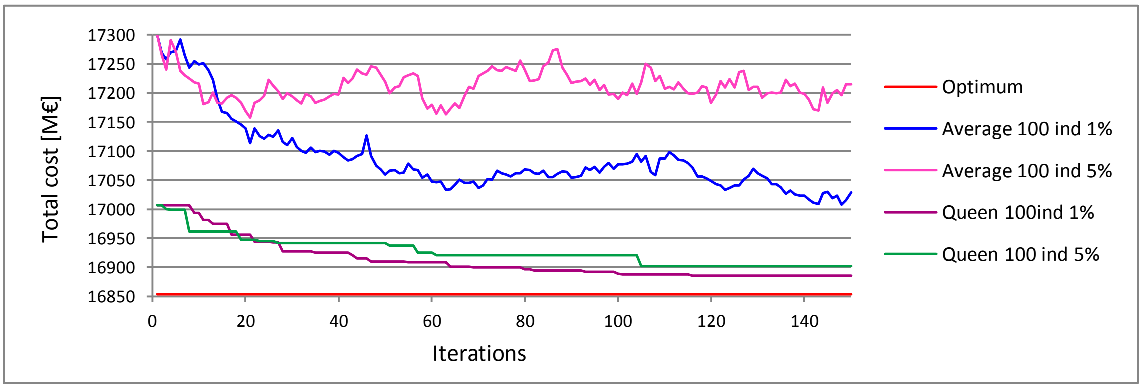


Figure 60: Total costs' average with a population size of 100 members and different mutation rates in case study 1.

Population size of 20 individuals

As in the simulation with 100 individuals, two clusters appear when using a mutation rate of 1% or 5% in the case of using 20 individuals, as it can be observed in the following figure, in the two pictures on the top.

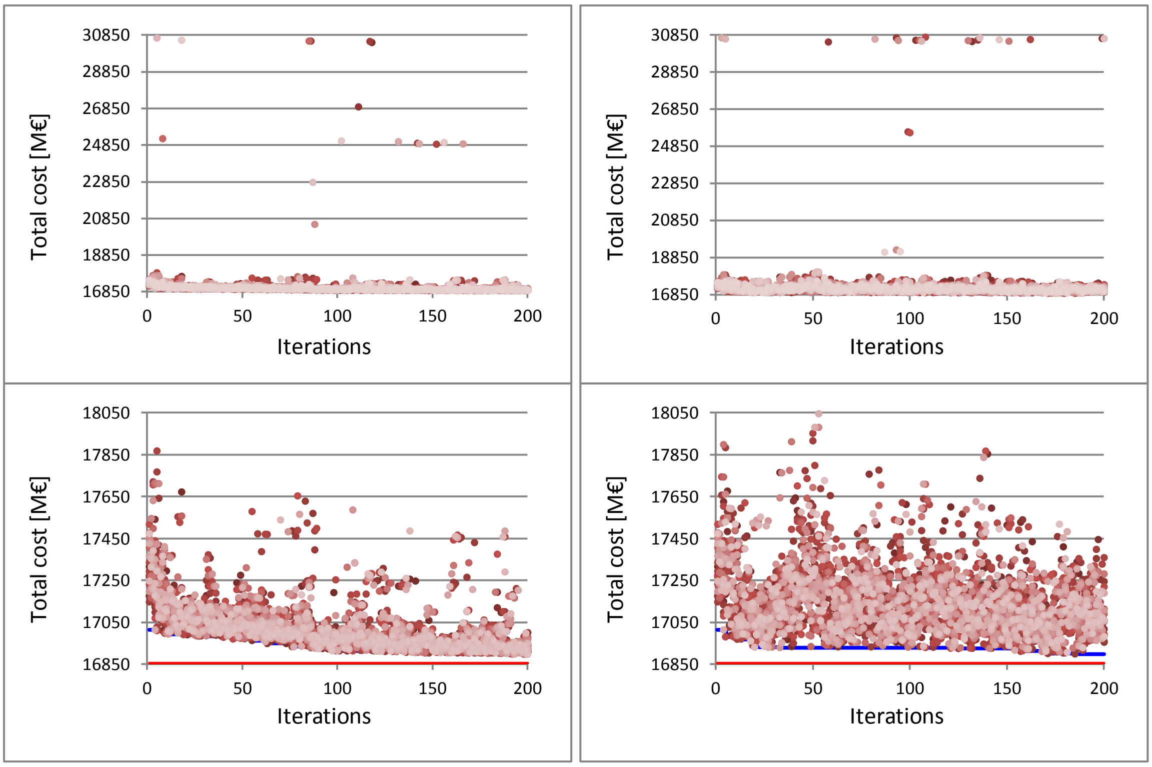


Figure 61: Evolution with 20 members and mutation rates of 1 and 5% in case study 3.

On the bottom-left, the genetic resolution with a mutation rate of 1% is showed and on the bottom-right with a mutation probability of 5%. The population clings to the queen individual evolution when using a mutation rate of 1%, unlike with a mutation probability of 5%, in which the population costs are more disperse.

Population size of 10 individuals

The evolution of the population with a mutation probability of 1% can be seen on the left in Figure 62. On the bottom-left it can be seen how the population follows the queen individual, a global decrease of the total cost of the population within the generations is reached and the whole population tightens the queen individual tendency.

On the right it is represented the evolution of the population for a mutation rate of 5%. It is interesting to appreciate the presence of a cluster in this case in contrast to the case of mutation probability of 1%. On the bottom-right, a dispersion of the population costs with respect to the queen individual cost can be observed.

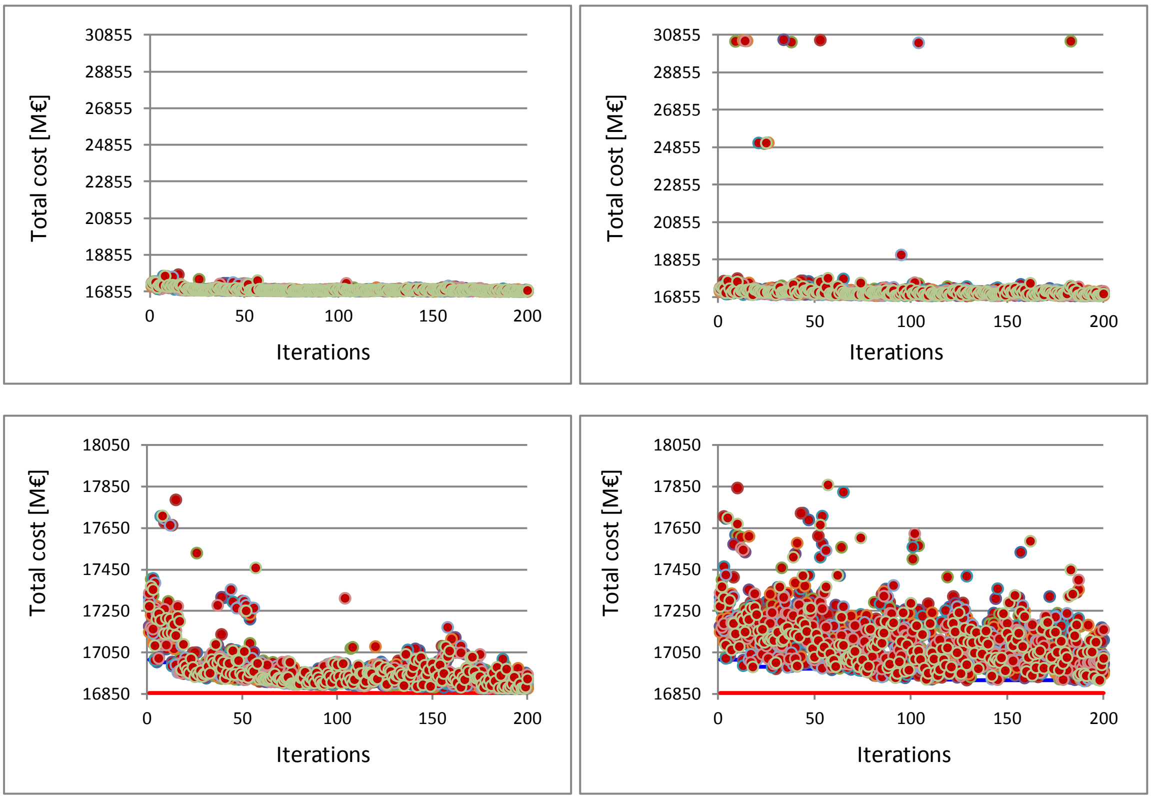


Figure 62: Evolution with 10 members and mutation rates of 1 and 5% in case study 3.

Figure 63 shows the other effect of the mutation, the population in the case of defining a mutation rate of 1% is nearer to the optimal plan than with a mutation of 5%.

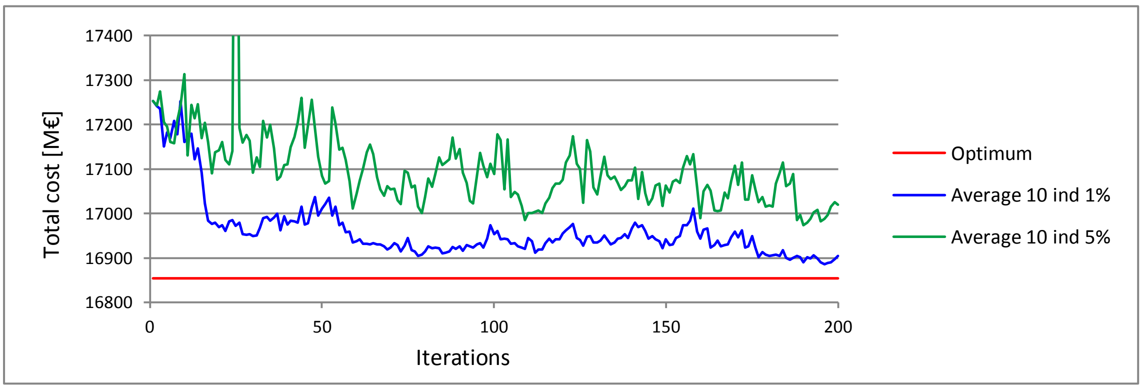


Figure 63: Total costs' average with a population size of 10 members and different mutation rates in case study 3.

Comparison

Figure 64 shows a comparison between queen individuals when using different mutation rates and a population size of 10 individuals in the case study 3. The electric system has 153 candidate lines, and the optimal plan provided by the complete problem resolution installs 72 power lines with a total cost of 16853.906 M€. The best solution is reached with a mutation probability of 1% for a population size of 10, but the same occurs with higher population sizes. The queen individual gets stagnant between iterations 40 and 60 and also at the end since iteration 130. A mutation rate of 5% is not able to introduce new valuable genetic information and slows the cost evolution of the queen individual.

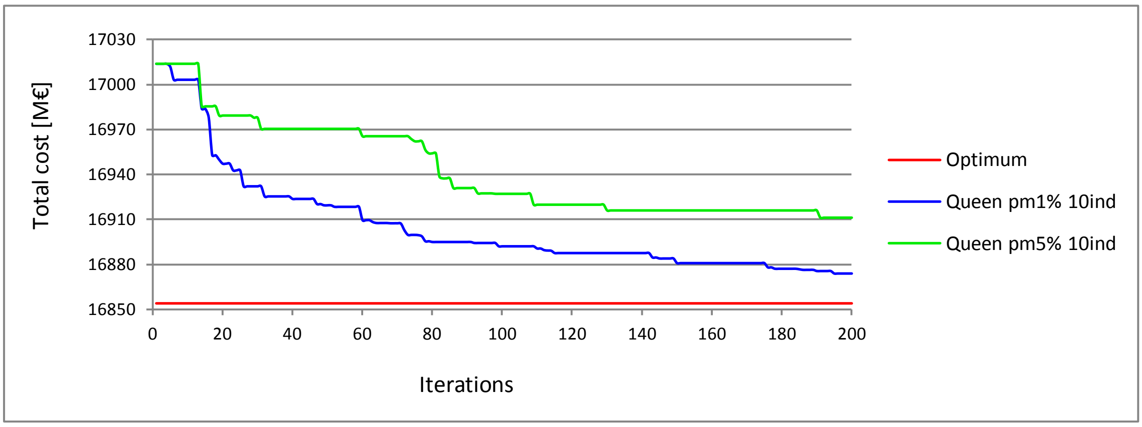


Figure 64: Queen's comparison with a population size of 10 and mutation rates of 1, 5 and 10% in case study 3.

5.4. Influence of the introduction of immigrants

This section explores the usefulness of introducing immigrant individuals for obtaining the solution for TEP. Starting from the best genetic parameters proposed in the previous sections, a predefined number of immigrants is introduced in a population of 10 individuals fixing the mutation rate at 1%. The objective is to test if the introduction of immigrants improves the best solution for each case study found by the GA.

The three following figures shows queen individuals when solving the case studies with immigrants. Resolutions with 2 and 4 immigrants have been carried out in order to test this operator efficiency. The best solutions for the three case studies were found without introducing immigrants.

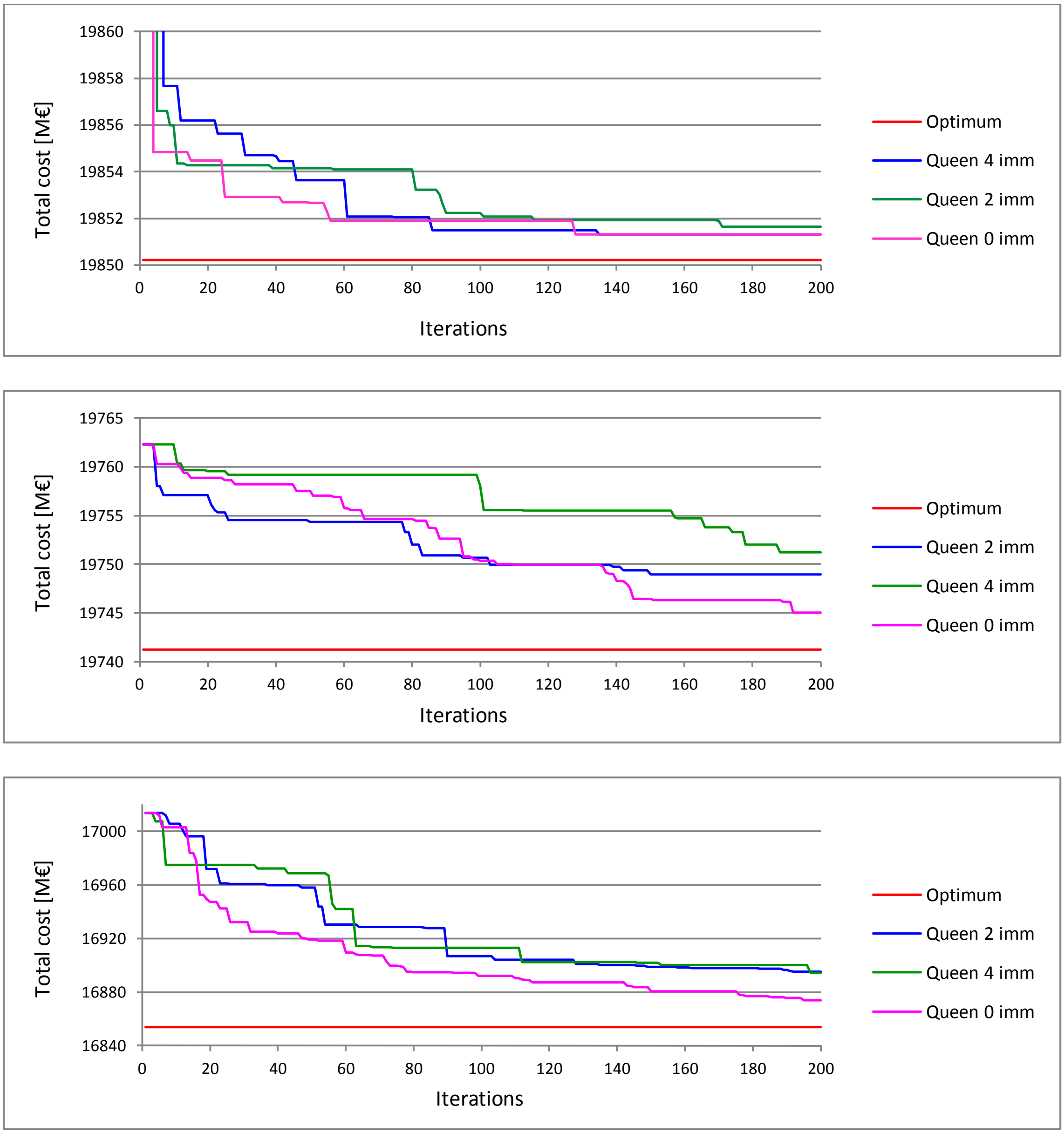


Figure 65: From the top to the bottom case studies 1, 2 and 3. Resolution with immigrants.

5.5. Contrasting GA efficiency with pure BD

5.5.1. *Case study 1*

BD spends 31.12 seconds and 334 iterations with a tolerance of 10e-9 for finding the optimal solution. This optimal plan installs 16 candidate power lines, out of the 53 that considers the case study 1, with a total cost of 19850.21M€.

For contrasting both resolution methods, the comparison has been made with a GA with a population size of 10 individuals, a mutation probability of 5% and without including immigrants. These parameters have been chosen because, as it was showed in the previous sections, they provide a good solution in a reasonable time.

Convergence criterion	Initial generation	Population size	Mutation probability
350 generations	60%	10 individuals	5%

Table 3: GA parameters for solving case study 1.

After 27 minutes and 127 generations the GA finds a solution with a total cost of 19850.6 M€ installing 17 candidate lines. On the basis of these results, pure BD resolution is clearly better than the GA one, BD provides the optimal solution and with important time savings. The following figure compares BD iterations with GA generations.

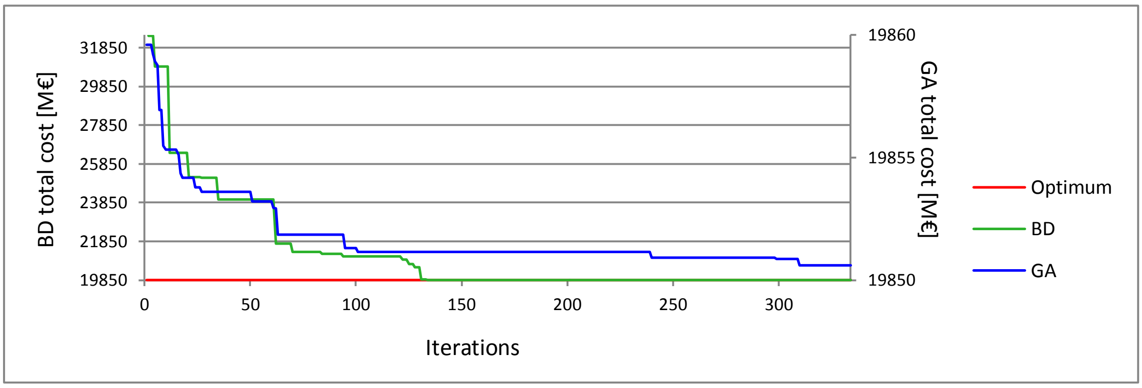


Figure 66: Contrasting GA generations with BD iterations in case study 1.

The GA solves the subproblem as many times as individuals that are in the population, that means that the GA solves 10 times the subproblem in each generation while BD only once per iteration. Therefore that is not the rigorously correct way of contrasting both techniques.

The key comparison is the execution time, and in the time that BD reaches the optimal solution, the GA only is able to carry out 3 generations, obtaining a solution that is 0.03% more expensive than the optimal one.

The comparison can be summed up in the following table where the main measures are contrasted.

METHOD	TIME [s]	ITERATIONS	LINES INSTALLED	TOTAL COST [M€]	ERROR [%]
BD	31.12	333	16	19850.214	0
GA	1757.51	309	17	19850,600	0.0020

Table 4: Contrasting GA efficiency with BD in case study 1.

5.5.2. *Case study 2*

In the problem that has 90 candidate power lines, the solution achieved by pure BD installs 16 lines with a total cost of 19741 M€. This solution is very similar to the optimal one except for the fact that it installs one less line, and installs one with a slightly higher fixed cost than the substitute proposed in the optimal plan. The resolution of the problem with BD lasted for 562 seconds with a tolerance of 10e-9.

The GA that provides the best result has carried out for contrasting its efficiency with BD, so the GA parameters have been chosen consistent with the previous parameter's study. The algorithm executed is defined with the following parameters:

Convergence criterion	Initial generation	Population size	Mutation probability
450 generations	40%	10 individuals	1%

Table 5: GA parameters for solving case study 2.

Figure 67 represents BD iterations with GA generations. The queen individual evolves gradually to finally getting stagnant with a total cost of 19742.295 M€. BD iterations suffer abrupt changes to finally converge with 0.003% error.

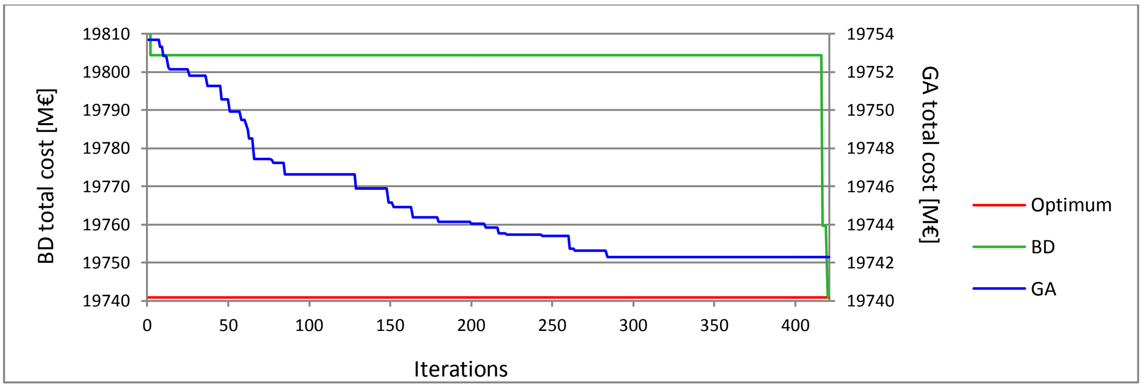


Figure 67: Contrasting GA generations with BD iterations in case study 2.

GA is able to construct 120 generations in the time that BD converges. After 120 generations the solution given by the GA is 0.03% higher than the optimal one. The following table illustrates the main measures of both resolutions processes.

METHOD	TIME [s]	ITERATION	LINES INSTALLED	TOTAL COST [M€]	ERROR [%]
BD	562.00	420	16	19741.032	0.0003
GA	2115.90	386	17	19742.295	0.0067

Table 6: Contrasting GA efficiency with BD in case study 2.

5.5.3. *Case study 3*

Finally the efficiency of the GA resolution is contrasted with BD in case study 3. The electric system is constituted by 1084 nodes and 294 power plants. The existing transmission network is configured by 1505 lines and transformers and there are 156 candidate lines in this case study.

The optimal plan provided by the complete problem resolution installs 72 candidate lines with a total cost of 16853.906 M€.

The GA implemented in this case study is defined by the following parameters:

Convergence criterion	Initial generation	Population size	Mutation probability
450 generations	60%	10 individuals	1%

Table 7: GA parameters for solving case study 3.

BD resolution is not able to converge with a tolerance lower than 4%, it gets stagnant with a cost of 16919.330 M€. This cost value is found in the second iteration after 37.02 seconds.

The GA method reaches a cheaper solution than BD. This solution has a cost of 16863.197 M€, only 0.0551% error, installing 13 more power lines than the optimal plan. The GA resolution spends 2392 seconds and 450 generations.

METHOD	TIME [s]	ITERATION	LINES INSTALLED	TOTAL COST [M€]	ERROR [%]
BD	37.02	2	156	16919.330	0.3882
GA	2414.70	446	85	16863.197	0.0551

Table 8: Contrasting GA efficiency with BD in case study 3.

6. CONCLUSIONS

6.1. Generating the initial population

The transmission network in case study 2 is configured by 90 candidate power lines. The optimal plan installs 17 lines, 19% of the total. In the light of this result it seems suitable to generate an initial population installing a similar percentage of candidate lines, to start the iterative process relatively close to the optimal solution. However the success probability in the binomial distribution for generating the initial population that better results achieved is 40%. Installing less than 40% of the candidate lines has resulted not to be a good idea. Despite of finding few individuals in low costs region, the remaining part of the population has high total costs, because of the power not supplied. Therefore, generating proposals with fewer lines installed in the initial generation involves an accused dispersion on the costs distribution function, not contributing to improve the GA resolution process.

In cases 1 and 3 the optimal plan provided by the complete problem resolution installs over 50% of the candidate power lines. However generating the initial population with a success probability of 60% in the binomial distribution has the best costs distribution and provides the best solution.

In conclusion the best success probability in the initial generation for case study 2 is 40% and for cases 1 and 3 is 60%, slightly higher than the proportion of lines installed in the optimal plan of each case study, to avoid a cost increment due to power not supplied.

6.2. Population size

As already stated in the number of individuals should be similar to the number of nodes that configured the transmission network for obtaining the best results in small problems. However this proportion gets stagnant with larger problems, so for the case study considered the most appropriate population size has resulted to be 10 individuals, independently of the number of candidate lines regarded.

A population size of 10 individuals has provided better solutions than defining population sizes of 20 or 100 members. Moreover the improvement capacity achieved by GA has also been higher in all the resolutions carried out when using 10 individuals and the time spent in the resolution process has been considerably lower. The population size is directly related with the time spent for reaching a solution. The subproblem is solved in each generation as many times as individuals are in the population. The genetic operators are very simple and they do not spend much time but the subproblem does. The population size should be high enough for not losing important genetic information along the algorithm resolution process, but low enough for no increasing the execution time unnecessarily.

6.3. Mutation probability

The mutation operator has demonstrated its importance in the resolution process of the GA. A mutation rate of 1% has provided the best results in the three case studies regarded, which is consistent with most of the literature. Higher mutation rates seem to be better at the beginning as they achieve lower costs in the first iterations, but then the population get stagnant, so the final solution provided is worst than with lower mutation probabilities.

The formation of clusters is a significant effect of the mutation rate. Clusters are high cost regions in the distribution function inherent to the case study. The number of clusters is directly related to the mutation probability, higher mutation rates involves more clusters. Furthermore the density of each cluster also grows as the mutation probability increases.

In case study 1 there are 4 clusters. At the beginning of the genetic resolution process we can see the 4 clusters, independently of the mutation rate defined, but after a few iterations some clusters disappear as consequence of a low mutation rate. Only 2 clusters are visible with a mutation rate of 1%. While the mutation probability grows, more clusters appear in cost increasing order and with a mutation rate of 10% all the clusters can be seen. Case studies 2 and 3 have two main clusters, but the same as in case study 1 happens. In this way the mutation rate can be thought as a filter of clusters, lower mutation rates block the apparition of higher cost clusters.

A higher mutation probability also causes a costs average increment in the population due to the costs dispersion caused by the creation of clusters. We can also state that the standard deviation of the costs average also grows with the mutation probability in the three case studies.

6.4. Introduction of immigrants

The introduction of few immigrants generated randomly as the same way than the initial population has resulted not to be a successful operator in the resolution process. The solutions obtained are all worse than the ones resulted of the genetic resolution without immigrants.

6.5. GA efficiency

The GA has been able to find good quality solutions in all the case studies, with errors under 0.3%. The selection of a suitable mutation rate is very important for finding better solutions. Three main operators have helped the GA to obtaining cheaper expansion plans: saving the queen individual, generating the initial population with a suitable binomial distribution, and an appropriate mutation. The introduction of immigrants has not been useful for solving the TEP problem in these case studies.

BD is a better resolution method for solving case studies 1 and 2, as it is illustrated in the previous section. It is able to provide the optimal solution or a very similar one with tremendous time savings.

However in case study 3, the GA seems to be a promising solution method. BD gets stagnant, while GA is able to find better solutions in a reasonable time.

6.6. Future Research

In the light of these results, we can conclude that the GA (with an appropriate exploitation of the information provided by the BD subproblem) is a promising method for solving large problems. A suitable selection of parameters contributes to reach a better solution with time savings. The optimal genetic parameters depend on the type of problem and initial data, so they have to be specifically chosen for each one.

The best final solution is provided by defining a mutation probability of 1% but in the first iterations a higher mutation rate reaches lower total costs, so it could be interesting to vary the mutation probability according to the population evolution phase. It would be useful to define the mutation rate by means of decreasing laws, such as a heat law of decreasing temperatures [SILV00].

Another possible ways to put into practice the mutation operator would be:

- To fix different mutation rates for the individuals of each cluster, so as to maximize the probability of being in a low-cost cluster.

- To define different mutation probabilities for each candidate line (allele) depending of its contribution to the individual fitness value in order to achieve a decrease in the total cost.

A good initial generation is very important in the development of the algorithm. To study the concrete characteristics of each area of the population distribution function would provide useful information to generate the initial set of individuals in a profitable way. Generating immigrants by means of binomial distribution does not provide any advantage. However we could also create immigrants using the information provided by the initial population distribution function.

Another promising technique could be to start the process of the solution research with GA, and when a certain tolerance is reached continue the resolution with pure BD. As we have seen in the section which contrasts GA efficiency with BD, in the first iterations the GA is able to reach good quality solutions, while BD starts its research process with very high costs. However the BD capacity for improvement seems to be better than the provided by the GA.

7. REFERENCES

[HOLL75] Holland. 1975.

[WHIT94] D. Whitley. “A Genetic Algorithm Tutorial.” 1994.

[YOSH95] K. Yoshimoto, K. Yasuda, R. Yokoyama. “*Transmission Expansion Planning Using Neuro-Computing Hybridized with Genetic Algorithm*.” Tokyo 1995.

[RUDN96] H. Rudnick, R. Palma, E. Cura, C. Silva. “*Economically adapted transmission systems in open access schemes - application of genetic algorithms*.” Santiago, Chile, 1996.

[MIRA97] V. Miranda, L. M. Proença. "Why risk analysis outperforms probabilistic choice as the effective decision support paradigm for power system planning." 1997.

[MIRA98] V. Miranda, L. M. Proença. "Probabilistic choice vs. risk analysis-conflicts and synthesis in power system planning." 1998.

[SILV00] E. L. da Silva, H. A. Gil, J. M. Areiza. "*Transmission Network Expansion Planning Under an Improved Genetic Algorithm.* " 2000.

[DEB02] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan. "A fast and multiobjective Genetic Algorithm: NSGA-II." 2002.

[SILV06] I. J. Silva M. J. Rider, R. Romero, C. A. Murari. "Genetic Algorithm of Chu and Beasley for static and multistage TEP." 2006.

[MAGH09] P. Maghouli, S. H. Hosseini, M. O. Buygi, M. Shahidehpour. *" A Multi-Objective Framework for Transmission Expansion Planning in Deregulated Environments.*" 2009.

[SILV10] A. M. L. da Silva, L. S. Rezende, L. A. F. Manso, G. J. Anders. " *Transmission Expansion Planning: A Discussion on Reliability and “N-1” Security Criteria.*" 2010.

[MAGH11] P. Maghouli, S. H. Hosseini, M. O. Buygi, M. Shahidehpour. " *A Scenario-Based Multi-Objective Model for Multi-Stage Transmission Expansion Planning.*" 2011.

[MOEI12] M. ;oeini_Aghtaie, A. Abbaspour, M. Fotuhi_Firuzabad. "Incorporating large-saclae distant wind farme in probabilistic transmission expansion planning." 2012.

[LUMB12] S. Lumbreras, A. Ramos. “*Improvements to Benders' decomposition. A Practical Evaluation Using a Transmission Expansion Planning Problem.*” Madrid 2012.

[SILV13] I. de J. Silva, M. J. Rider, R. Romero, C. A. Murari. "*Genetic Algorithm of Chu and Beasley for Static and Multistage Transmission Expansion Planning.*" 2013.

[LUMB14] S. Lumbreras Ph. D. Thesis."Decision Support Methods for large-scale Flexible Transmission Expansion Planning." Madrid, 2014 .

8. APPENDIXES

8.1. Codification of the TEPES-Genetic Algorithm in GAMS

```
SETS

l                   iterations

y                   year

nd                  node (bus)

cc                  circuit

lc (nd,nd,cc)       candidate lines
```

[...]

```
* GENETIC ALGORITHM SETS

i            individuals proposed by master problem based on optimization or on
             GA/ ind001*ind010 /

pp(i,i)      selector of individuals in the intermediate population

pqueens (i)  selector of the queen individual


PARAMETERS

pInstalCap_L  (l,y,nd,nd,cc)     first stage variables valuesin iteration l
```

[...]

```
* GENETIC ALGORITHM PARAMETERS

pqueen (l,i)                     queen selector parameter

pFitnessmax (l)                  max fitness

pm                               mutation probability            /0.01/

pCardlc                          cardinal of lc

pOrdlc (nd,nd,cc)                ordinal of lc

pInstalCapGA_L (l,i,y, nd,nd,cc) first stage variables values in iteration l
                                 solved by GA

pInstalCapGA_L_inter (l,i,y, nd,nd,cc) intermediate population

pInstalCapGA_L_cross (l,i,y, nd,nd,cc) population after cross
```

```
pCost (l,i)                 individuals' total cost(vTotalMCost)

pmaxcost (l)                max cost in iteration l

pmincost (l)                min cost in iteration l

pFitness (l,i)              population'sfitness

pFitnessa (l,i)             population'sprevious fitness

pRoulette (l,i)             cumulate fitness

ptotal (l)                  addition of the previous fitness

prans (l,i)                 random number for selection

ppmt (l,i, nd,nd,cc)        random number for mutation

ppc (l,i)                   cross point

ppm (l,i)                   mutation point

pGenInit                    success probability for generating the initial
                            population                               /0.6/

pNumImm                     number of immigrants                     /0/

pOptImm                     OPTION OF IMMIGRANTS (0 NO   1 YES)      /0/

pOptQueen                   OPTION OF QUEEN      (0 NO   1 YES)      /1/


VARIABLES

vInstalCap (y,nd,nd,cc)     indicator of cumulat candidate line   [0-1]

vTotalOCost (y,p)           total system variable         cost    [M€]

vTotalGCost (y,p)           total system gen reliability cost     [M€]

vTotalMCost                 total system master           cost    [M€]
```

[…TEPES MODEL...]

```
loop (l $[[abs(pC_Aux) > pTol or (pIteratOptcr>pBdTol and pInexactMaster)] and
pConverged = 0 and ord(l) <= pBdIter],
```

[…]

```
* GENETIC ALGORITHM CODE

   elseif pMasterOpm = 2,

       pCardlc=0 ;

      loop (lc,
```

```
            pCardlc     = pCardlc + 1 ;
            pOrdlc(lc) = pCardlc     ;
      );
        if (card(ll) = 0,
*            INITIALIZATION
            pFitness  (lll,iii) = 0 ;
            pFitnessa (lll,iii) = 0 ;
            pRoulette (lll,iii) = 0 ;
            pInstalCapGA_L_cross(lll,iii,y,lc) = 0 ;
        else
            pRoulette(l-1,i) $[ord(i)=1] = pFitness(l-1,i) $[ord(i)=1] ;
            loop (i $[ord(i)>1],
              pRoulette(l-1,i) = pRoulette(l-1,i-1) + pFitness(l-1,i) ;
        ) ;
***          SELECTION
           prans(l,i) = Uniform(0,1) ;
           pp(i,ii) = no ;
           pp(i,ii) $[pRoulette(l-1,ii) >= prans(l,i) and pRoulette(l-1,ii-1) <
prans(l,i)] = yes ;
             pInstalCapGA_L_inter(l-1,i,y,lc) = sum[pp(i,ii), pInstalCapGA_L(l-
1,ii,y,lc)] ;
***          CROSSOVER
        loop ((i,iii) ${(floor(ord(i)/2))*2 +1 = ord(i) and ord(iii) =
      ord(i)+1 and ord(iii) <= card(i)},
             ppc(l,i) = UniformInt(1,pCardlc) ;
             pInstalCapGA_L_cross(l,  i,y,lc) $[(pOrdlc(lc)<=ppc(l,i))] =
pInstalCapGA_L_inter(l-1,  i,y,lc) ;
             pInstalCapGA_L_cross(l,  i,y,lc) $[(pOrdlc(lc)> ppc(l,i))] =
pInstalCapGA_L_inter(l-1,iii,y,lc) ;
             pInstalCapGA_L_cross(l,iii,y,lc) $[(pOrdlc(lc)<=ppc(l,i))] =
pInstalCapGA_L_inter(l-1,iii,y,lc) ;
             pInstalCapGA_L_cross(l,iii,y,lc) $[(pOrdlc(lc)> ppc(l,i))] =
pInstalCapGA_L_inter(l-1,  i,y,lc) ;
             ) ;
***          MUTATION
```

```
            ppmt(l,i,lc)=Uniform(0,1);

            pInstalCapGA_L(l,i,y,lc) $[(ppmt(l,i,lc) <= pm) ] = 1 -
pInstalCapGA_L_cross(l,i,y,lc) ;

            pInstalCapGA_L(l,i,y,lc) $[(ppmt(l,i,lc) >  pm) ] =
pInstalCapGA_L_cross(l,i,y,lc) ;

***         QUEEN

             if(pOptQueen=1,

               pqueen(l,i)=0;

               pFitnessmax(l-1) = smax(i,pFitness(l-1,i));

               pqueen(l,i)$[pFitness(l-1,i)>=pFitnessmax(l-1)]=1;

               pqueens(i)=no;

               loop(i,

                  if(sum [ii$(ord(ii)<=ord(i)),pqueen(l,ii)]<=1,

                      pqueens(i)$[pqueen(l,i)=1]=yes;

                  );

               );

               pInstalCapGA_L(l,'ind001',y,lc) = 0 ;

pInstalCapGA_L(l,'ind001',y,lc) = sum[pqueens(i),pInstalCapGA_L(l-1,i,y,lc)];

            );

***         IMMIGRANTS

            if( pOptImm = 1,

               pInstalCapGA_L(l,iii,y,lc)$[(ord(iii) > 1) and (ord(iii) <
(pNumImm + 2))] = 0;

               pInstalCapGA_L(l,iii,y,lc)$[(ord(iii) > 1) and (ord(iii) <
(pNumImm + 2))] = UniformInt(0,1);

            );

   );

*        END OF GENETIC ALGORITHM CODE
```

[…]

```
   loop (i $[(ord(i) = 1 and pMasterOpm = 1) or pMasterOpm = 2],

*  GENETIC ALGORITHM INITIAL GENERATION

       pInstalCapGA_L('it0001',i,y,lc)$[pMasterOpm = 2] =
randBinomial(1,pGenInit);

*     Communication master and subproblem
```

```
      vInstalCap.l (  y,lc) $[pMasterOpm = 2] = pInstalCapGA_L(l,i,y,lc) ;

      pInstalCap_L (l,y,lc)                   = vInstalCap.l(y,lc) ;

      vInstalCap.fx(  y,lc) $[pOptCmplt = 0]  = vInstalCap.l(y,lc) ;

*     As the master problem is not solved, fixed costs are input separately

      vTotalFCost.fx $[pMasterOpm = 2] = alpha * sum[(y,lc), (card(y)-
ord(y)+1)*pFixedCost(lc)*[pInstalCap_L(l,y,lc) - pInstalCap_L(l,y-1,lc) ]] ;
```

[…SUBPROBLEM…]

```
* GENETIC ALGORITHM FITNESS

       pCost(l,i) $[pMasterOpm = 2] = vTotalMCost.l ;

   );

*  FITNESS CONSTRUCTION

   if (pMasterOpm = 2,

   option pCost:5;    display pCost;

      pmaxcost (l)    = smax(i, pCost(l,i))* 1.1 ;

      pmincost (l)    = smin(i, pCost(l,i))* 0.9 ;

      pFitnessa(l,i) = [pmaxcost(l)- pCost(l,i)]/[pmaxcost(l)- pmincost(l)] ;

      ptotal    (l   ) = sum[i, pFitnessa(l,i)] ;

      pFitness (l,i) = pFitnessa(l,i)/ptotal(l) ;

*     Show generation in console

      put CONSOLE   putclose'Population in iteration ' l.tl /

      put SUMMARYSL put'Population in iteration ' l.tl /

      loop (i,

         pAux = pCost(l,i);

         put CONSOLE   putclose'    Ind ' i.tl ' ' pAux:10:3

         put SUMMARYSL put'    Ind ' i.tl ' ' pAux:10:3

      ) ;

   ) ;
```

[…]