
Samuel Pinto Ramos de Pina

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisor(s): Prof. Pedro Manuel Santos de Carvalho
Engª. Maria Inês Verdelho

Examination Committee

Chairperson: Profª. Célia Maria Santos Cardoso de Jesus
Supervisor: Prof. Pedro Manuel Santos de Carvalho
Member of the Committee: Prof. Rui Manuel Gameiro de Castro

June 2019
Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.
Dedicada à minha mãe e às minhas avós, que sempre me apoiam e ajudaram ao longo do meu percurso. Bem como ao Professor João Eira que me ensinou e me mostrou que às vezes é preciso pensar os problemas de uma forma diferente.
Acknowledgments

Quero agradecer a todos os que contribuíram de alguma forma para o desenvolvimento desta dissertação. Em particular ao Professor Pedro Carvalho por todo apoio, incentivo e confiança que me transmitiu durante todo o processo e pela paciência infinita durante as reuniões. Quero também agradecer à Engenheira Inês Verdelho da EDPD que me deu um apoio permanente e me forneceu todos os dados e informações necessários para o desenvolvimento da tese. Gostaria ainda de referir a Professora Mónica Aragüés Peñalba do CITCEA que me convidou para estagiá-lo no seu centro de investigação e que me introduziu ao GAMS, que acabou por ser uma ferramenta fundamental no trabalho desenvolvido.

A todos os professores, desde o básico ao Técnico, que ajudaram a moldar o meu interesse pela ciência e engenharia e me deram as ferramentas para poder levar este caminho até ao fim. Em particular à Glória, Profª. Sandra Martins e Clara Peixe e Prof. João Eira.

Não esquecendo também o agradecimento à Drª Maria José Ferrão e o Duarte Donas-Boto por me terem dado a oportunidade de integrar a equipa do NAPE, que acabou por ser uma parte essencial do meu percurso no Técnico. A todos os meus colegas e amigos tanto do NAPE, como de MEEC e MEFT e do secundário um obrigado pelo companheirismo e pelo incentivo e confiança nas minhas capacidades.

Quero também agradecer à Ana Luisa Carvalho, grande companheira e amiga, por toda a paciência, todo o apoio e incentivo e pela cumplicidade que não se pode escrever. Obrigado também pela ajuda essencial na revisão do texto e da escrita, sem a sua ajuda não teria sido possível obter este resultado final.

Por último, à minha família e em particular à minha Mãe, minha grande tutora e amiga, e às minhas avós Sá o Luísa agradeço o amor infinito, o esforço, o suporte e a confiança que sempre me deram de forma a que eu pudesse completar os meus estudos.
Resumo

O sistema de energia elétrica de um país é um componente fundamental para a sua soberania, sendo, portanto, essencial garantir o seu bom funcionamento. Os operadores de transporte e de distribuição são os responsáveis pela gestão da rede e manutenção dos respetivos ativos. Um dos principais ativos da rede de distribuição é o parque de transformadores de potência AT/MT e MT/MT. Os transformadores de potência (TP) sofrem com a idade uma degradação da sua condição, pelo que foi considerado importante desenvolver um modelo de envelhecimento que repercuta a idade do ativo na sua condição de saúde.

Nesta tese é desenvolvido e proposto um modelo estocástico de envelhecimento para simular a evolução dos níveis de saúde dos transformadores no tempo. O modelo é parametrizado com base nos dados disponíveis sobre os atuais índices de saúde do parque de TP. Posteriormente os resultados obtidos para a saúde foram mapeados nos valores expectáveis das taxas de avarias dos TP, de forma a permitir simular a correspondente degradação da sua fiabilidade.

O modelo de envelhecimento proposto é finalmente utilizado como ferramenta de suporte à decisão. É utilizado para simular o impacto de diferentes estratégias de investimento na renovação do parque de TP. Conhecendo a consequência da falha de cada um dos TP do parque no que respeita à energia não distribuída (END) e simulando o envelhecimento e correspondente evolução das taxas de avarias dos TP do parque, foi possível avaliar o risco e os custos de investimento correspondentes à implementação de diferentes estratégias de renovação. Foram avaliadas algumas estratégias tipo e comparados os seus resultados quanto à END e custos expectáveis.

**Palavras-chave:** Transformadores de Potência, Modelos de Envelhecimento, Fiabilidade de Sistemas de Energia, Diagnóstico da Degradação, Simulações de Monte Carlo, Análise da Energia Não Distribuída.
Abstract

A country’s electrical power system is a fundamental component of its sovereignty being therefore essential to ensure its availability and reliability. The transmission and distribution operators are responsible for the management of the system and the maintenance of its assets. Among the multitude of distribution assets, the HV/MV and MV/MV power transformers (PTs) are one of the most important. PTs suffer with age a degradation of their condition, reason why it was considered important to develop an aging model for this kind of asset that would map age onto health condition in a probabilistic way.

In this thesis, a stochastic aging model is proposed and developed in order to be used to simulate the health condition evolution of distribution PTs. The model is parameterized with the data available regarding the health indexes of the installed PTs. Health indexes are mapped onto expected values of PTs failure rates in order to allow simulating the reliability degradation together with the health condition evolution with age.

The developed aging model is used as a decision-making support tool. More specifically, it is used to simulate different grid renewal investment strategies. By understanding consequences of the failure of each PT, including the consequent energy not supplied (ENS) and simulating the reliability degradation of the stock of PTs, it was possible to evaluate the cost and corresponding risks of different stock renewal strategies. In this thesis, simple renewal strategies were evaluated and their outputs compared in terms of ENS and expected total costs.

Keywords: Power Transformers, Aging Models, Power System Reliability, Condition Assessment, Monte Carlo Simulation, ENS Analysis.
## Contents

Acknowledgments .................................................. vii
Resumo ............................................................... ix
Abstract ............................................................ xi
List of Tables ......................................................... xv
List of Figures ......................................................... xvii
List of Acronyms ................................................... xix

1 Introduction ....................................................... 1
   1.1 Research Motivation ......................................... 1
   1.2 Work Objectives ............................................. 2
   1.3 Dissertation Outline ......................................... 2
   1.4 Topic Overview and Background ............................ 3
      1.4.1 Equipment Reliability and Preventive Maintenance .. 3
      1.4.2 Power Transformers Condition Assessment ............ 4
      1.4.3 Power Transformers Aging Models ...................... 6
      1.4.4 Markov Model based aging model ...................... 9
      1.4.5 Monte Carlo Simulation Method ....................... 12
      1.4.6 Pareto Optimality Frontier ............................. 14

2 Aging Model for Power Transformers .......................... 17
   2.1 Time and Health discretization ............................ 17
      2.1.1 Age Discretization .................................... 17
      2.1.2 Partial Health Index Discretization .................. 18
      2.1.3 Partial Degradation Index (PDI) ...................... 19
      2.1.4 Age intervals merging ................................ 20
   2.2 PDI probability distributions per age group ............ 21
      2.2.1 PDI log-normal Distribution Fitting .................. 21
      2.2.2 Extreme Value Distribution Fitting ................... 24
      2.2.3 Evolution of PDI mean value and variance evolution with age .... 25
   2.3 Stationary PDI Probability Distribution Matrix ........... 27
      2.3.1 Matrix Computation Based on Real Data .............. 27
      2.3.2 Matrix computation based in the fitted probabilistic distributions ... 27
2.3.3 Corrections for parametrization feasibility ........................................ 28
2.4 PDI trajectories simulation .................................................................. 30
  2.4.1 Markov chains method for the degradation model ......................... 30
  2.4.2 Transition matrices computation method ...................................... 32
2.5 Failure probability estimation based on the PDI .................................... 35

3 Stock Health Simulation and Reliability Evaluation ................................. 39
  3.1 ENS Estimation due to Power Transformers Failures ......................... 39
  3.2 Global Stock Simulator Algorithm .................................................... 41
    3.2.1 Failure Simulation .................................................................... 41
    3.2.2 Outputs of the Simulator ......................................................... 41
  3.3 Monte Carlo Method Implementation ................................................ 43

4 Stock Renewal Strategies: Comparison and Result Analysis .................... 47
  4.1 Global Stock Renewal Strategies ....................................................... 47
    4.1.1 Zero Strategy Analysis ............................................................. 48
    4.1.2 Age based renewal strategy ....................................................... 57
    4.1.3 Top PDI and Top ENS renewal strategies .................................... 59
    4.1.4 RUL Based Renewal Strategy ................................................... 63
    4.1.5 Hybrid Renewal Strategy .......................................................... 66
  4.2 Results discussion .............................................................................. 69

5 Conclusions .............................................................................................. 71
  5.1 Achievements Summary ...................................................................... 71
  5.2 Practical Implications ......................................................................... 74
  5.3 Future Work ....................................................................................... 75

Bibliography ................................................................................................. 77

A Failure Rates, Peak Powers and Interruption Times for the Power Transformers (AT/MT and MT/MT) I

B PDI Distribution Tables for each age interval ........................................ VIII

C Markov Transition Matrices ................................................................... XIII

D Summary of the ENS and cost results of the simulation of different strategies XVII
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Age intervals discretization</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>Partial Health Index (PHI) levels discretization</td>
<td>19</td>
</tr>
<tr>
<td>2.3</td>
<td>Partial Degradation Index (PDI) levels discretization</td>
<td>20</td>
</tr>
<tr>
<td>2.4</td>
<td>Average PDI values for each age group before merging of intervals</td>
<td>21</td>
</tr>
<tr>
<td>2.5</td>
<td>Average PDI values for each age group after merging of intervals</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>PDI mean values and variances discretized by age groups</td>
<td>25</td>
</tr>
<tr>
<td>2.7</td>
<td>PDI probability matrix based in real data</td>
<td>27</td>
</tr>
<tr>
<td>2.8</td>
<td>PDI probability matrix based in the probability distributions obtained in</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>2.2.1</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>Final PDI probability matrix after feasibility constraint application</td>
<td>29</td>
</tr>
<tr>
<td>2.10</td>
<td>Comparison between the original PDI probability matrix and the final obtained</td>
<td>29</td>
</tr>
<tr>
<td>2.11</td>
<td>Chopped Markov transition Matrix for brand new PTs. transition from age 0 to</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>age interval 1 (1 to 5 years)</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>Installed PTs stock average failure rate and average failure index</td>
<td>36</td>
</tr>
<tr>
<td>4.1</td>
<td>ENS results compilation for the 35 years of simulation</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Zero Strategy cost analysis</td>
<td>55</td>
</tr>
<tr>
<td>4.3</td>
<td>Average number of failures per PT type in each time interval with the “Zero</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Strategy” renewal strategy</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>Analysis of the ENS and failures in non redundant PTs</td>
<td>55</td>
</tr>
<tr>
<td>4.5</td>
<td>Age Based Strategy ENS analysis</td>
<td>57</td>
</tr>
<tr>
<td>4.6</td>
<td>Age Based Strategy cost analysis</td>
<td>57</td>
</tr>
<tr>
<td>4.7</td>
<td>Top PDI and top ENS strategies ENS analysis</td>
<td>62</td>
</tr>
<tr>
<td>4.8</td>
<td>Top PDI and Top ENS strategies cost analysis</td>
<td>62</td>
</tr>
<tr>
<td>4.9</td>
<td>RUL based strategy ENS analysis</td>
<td>64</td>
</tr>
<tr>
<td>4.10</td>
<td>RUL based strategy cost analysis</td>
<td>64</td>
</tr>
<tr>
<td>4.11</td>
<td>Hybrid type strategy ENS analysis</td>
<td>67</td>
</tr>
<tr>
<td>4.12</td>
<td>Hybrid type strategy cost analysis</td>
<td>67</td>
</tr>
<tr>
<td>B.1</td>
<td>IDP distribution for TPs with age between 1 and 5 years</td>
<td>VIII</td>
</tr>
<tr>
<td>B.2</td>
<td>IDP distribution for TPs with age between 6 and 10 years</td>
<td>IX</td>
</tr>
<tr>
<td>B.3</td>
<td>IDP distribution for TPs with age between 11 and 15 years</td>
<td>IX</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Evolution of the maintenance engineering importance in the 20th century. Adapted from: Shenoy and Bhadury [2] .................................................. 3
1.2 The importance of different diagnostic methods to estimate PTs condition. Source: [10] . 5
1.3 Transformers failure causes. Adapted from: [10] ........................................ 6
1.4 Defective components of a transformer (LTC: load tap changer). Adapted from: [10] .... 6
1.5 Basin curve for failure rate of electronic equipment. Source: [18] .......................... 8
1.6 Offline Markov Model diagram. Source: [23] ........................................ 9
1.7 Modeling process of deterioration performance curve based on a Markov Model (MM). Adapted from: [24] .................................................. 10
1.8 Pareto Optimality Frontier example. Source: [34] ........................................ 14
1.9 Example of optimal point computation using Pareto Frontier. Source: [35] ............ 15
2.1 Log-normal Distribution fit over the PDI Data for PTs with ages between 1 and 5 year . 22
2.2 Log-normal distribution Fit over the PDI Data for PTs with ages between 6 and 10 year . 22
2.3 Log-normal distribution Fit over the PDI Data for PTs with ages between 11 and 15 year . 22
2.4 Log-normal distribution Fit over the PDI Data for PTs with ages between 16 and 25 year . 23
2.5 Log-normal distribution fit over the PDI Data for PTs with ages between 26 and 35 year . 23
2.6 Log-normal distribution Fit over the PDI Data for PTs with ages between 36 and 40 year . 23
2.7 Extreme Value Distribution fit over the PDI Data for PTs with ages between 41 and 45 year . 24
2.8 Extreme Value Distribution fit over the PDI Data for PTs with ages above 45 years old . 25
2.9 Data and fitting of the PTs PDI average mean value evolution with age .................. 26
2.10 Data and fitting of the PTs PDI Variance evolution with age ............................ 26
2.11 1st Markov transition matrix graphical example ............................................. 31
2.12 PDI to FI conversion schema ................................................................. 35
2.13 Failure probability as a function of the failure index ................................. 36
3.1 Computation order of the simulation outputs ............................................... 43
4.1 Evolution with time of the average PDI value of the installed PTs ...................... 49
4.2 Evolution with time of the average age of the installed PTs .......................... 49
4.3 Average failure index level and standard deviation evolution with time ............ 50
4.4 Average failure probability of the installed PTs evolution with time .................. 51
4.5 Evolution with time of the average amount of failures in the installed PTs .................. 52
4.6 Average number of new PTs installed and repaired in the grid per each time interval ... 53
4.7 Average ENS value and respective monetary valuation for each time interval .......... 54
4.8 Monte Carlo simulation failures distribution for different time intervals ............... 56
4.9 Main results regarding the simulation of the "over X" renewal strategies for different X age values ................................................................. 58
4.10 Main results regarding the simulation of the "Top X PDI level" renewal strategies for different X values ................................................................. 60
4.11 Main results regarding the simulation of the "Top X Possible ENS level" renewal strategies for different X values ................................................................. 61
4.12 Comparison between equivalent "Top X PDI level" and "Top X Possible ENS level" renewal strategies for different X equal to 20 and 60 .................................................. 62
4.13 Main results regarding the simulation of the RUL based renewal strategies in comparison with the "Top 60 possible ENS" and the "over 50 years" renewal strategies .......... 65
4.14 Main results regarding the simulation hybrid1 and hybrid2 strategies in comparison with "the over 60 years" renewal strategy .................................................. 68
4.15 ENS values as a function of the strategy cost for the simulated renewal strategies together with the respective Pareto Efficiency Frontier .............................................. 69
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg</td>
<td>Average</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>ENS</td>
<td>Energy Not Supplied</td>
</tr>
<tr>
<td>HV</td>
<td>High Voltage</td>
</tr>
<tr>
<td>MV</td>
<td>Medium Voltage</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PDI</td>
<td>Partial Degradation Index</td>
</tr>
<tr>
<td>PHI</td>
<td>Partial Health Index</td>
</tr>
<tr>
<td>PT</td>
<td>Power Transformer</td>
</tr>
<tr>
<td>TM</td>
<td>Transition Matrix</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Research Motivation

The energy distribution grid, managed and operated by EDP Distribution (EDPD), is a fundamental part of the national energy grid. It is composed by a variety of elements, one of them being the power transformers (PT), which are responsible for connecting grids with different voltage levels. There are hundreds of this type of transformers installed in the Portuguese distribution grid, which combine into a total of 17689 MVA of installed power, according to EDP [1]. These PTs can be divided into two groups. Ones responsible for the transformation between high voltage (HV) and medium voltage (MV) (E.G. 60/15kV, 60/10kV). The majority of the PTs installed are in this first group and are defined as HV/MV. The other ones make the transformation between different level of medium voltage (E.G. 30/10 kV) and are called MV/MV. These represent a minority of the installed stock of PTs.

The aforementioned PTs can have rated powers up to 40 MVA, therefore each time one of these transformers fails, there can be a considerable amount of energy not supplied to the consumers. Even though some substations can have more than one transformer working in parallel, in order to increase the redundancy, a PT failure can still have a major impact on the supplied energy. The value estimated for the average value of the energy not supplied (ENS) per year over the last three years is around 243 MWh.

EDPD owns a total of 769 PTs, with a current average age of 30.5 years. The expected lifetime defined by the manufacturers for the PTs is between 25 and 40 years, depending on the manufacturer. Each transformer can cost over 500 000 Euros, making it is crucial to develop, a strategic plan for the grid renewal, in order to guarantee that its implementation is technically sound and economically feasible.

The main motivation for this thesis is the development of this grid renewal strategy. The main contributions of this thesis will be the development of technical tools that may work as a support for the EDPD decision making process, regarding the strategic investment plan.
1.2 Work Objectives

This master’s thesis has three defined objectives. The first one, which was defined by EDPD, is the computation of an estimate for the current total amount of energy not supplied (ENS) associated with failures in the PTs of the national distribution grid, based on the data collected in recent years. The second objective is the development of an aging model which estimates the evolution of the PTs health condition degradation with time. Based on this, the final goal is to develop a simulator that allows to estimate, using the Monte Carlo method, the average evolution of the installed PTs condition and extract multiple output parameters, such as the average ENS for the upcoming years. This simulator should incorporate the option of testing different renewal strategies for the installed PTs stock, making it possible to test and estimate the impact of those strategies in the associated total costs and ENS values.

1.3 Dissertation Outline

This master thesis is divided into 5 chapters, considering the present one. On this first chapter, specifically on section 1.4 it is presented a topic overview on the thesis scope. Different aging models applied in multiple areas of science and engineering, that could possibly be applied to model the evolution of the PTs health condition. Also, the indexes used as a support for the health condition valuation are explained. Besides, some references regarding Markovian chains modeling are also presented. This will be used to model the evolution process of the PTs condition. In this chapter concepts which will be used as fundamental tools during the thesis are explained. For example, Monte Carlo tests, optimization problems solving techniques and Pareto Optimality Frontier.

On chapter two the designed model to represent the health condition evolution of the each individual PT is presented. In this chapter it is explained how to manipulate the time dimension, together with all the data from the health indexes from each transformer, in order to fulfill the objective of the model. It is also presented an hypothesis for a relationship between the age and the degradation, obtained through the fitting of different probability distributions to the PTs data. It is also explained how to use the Markov Chains theory in order to implement it to the specific problem.

The third chapter presents the global stock condition simulator. It is explained how this simulator is built and what kind of data is possible to extract from it. In particular, how it is possible, from the model defined in chapter three, to estimate variables from the PTs installed stock, such as: the average ENS; the average amount of PT failures, broken down by PT type; the total cost associated with each possible substitution/renewal strategy simulated.

In the fourth chapters the results extracted from the model, for different renewal strategies simulated, are presented together with a comparison between them. This analysis is performed in order to find the best trade-off between investment in the grid and cost associated with the degradation of the park and consequent non delivered energy.

The final chapter includes some main conclusions regarding the results obtained during the thesis work. Also some notes related with the technical implementation and possible improvements for the
simulator defined during the aging model development. At the end, there is a section regarding possible future work and general improvements, that can be done in order to hone the work accomplished during this masters thesis.

1.4 Topic Overview and Background

1.4.1 Equipment Reliability and Preventive Maintenance

With the second industrial revolution, new technological advancements initiated the emergence of large factories and organizational models of production as envisioned by Taylor and Ford. Since machinery and technical equipment are the main protagonists of production processes, a technical failure can have massive consequences. This is the reason for the large investment in studies and research associated with equipment reliability. This lead to the development of an engineering area, the so-called maintenance engineering, usually related with the mechanical and electrical engineering fields.

Although studies regarding the maintenance of equipment began on the 19th century, they had a major growth during the 20th century, in particular with the 2nd world war, and are still a major field of research and development nowadays. In the period before the war, maintenance was seen as an added cost to the final production cost, which, at the end, did not increased the product value. Since it was a costly option, maintenance was restricted to repairing the equipment only when there was a failure.

As presented in the diagram from Figure 1.1, after the 2nd world war, there has been an increase in the awareness for issues such as the environmental preservation, quality of product and services, which turned maintenance an important part in the success of any industry.

![Figure 1.1: Evolution of the maintenance engineering importance in the 20th century. Adapted from: Shenoy and Bhadury [2]](image)

The goal of these studies is to find mathematical procedures that can improve the assessment or the estimation of equipment reliability, in order to reduce the time of inoperability, as explained by [3]. Many different methods were developed as a support for this field of research. One of the major examples is the use of the Weibull and log-normal distributions applied to this scope [4, 5].
Replacing or repairing an equipment too often might drastically increase the production costs. Any decision-making policy that defines when and how to do maintenance of an equipment is defined as a maintenance strategy. If maintenance is performed before the equipment starts to malfunction it is called predictive maintenance. Based on this idea, a different concept named Strategic Replacement of Equipment was developed and introduced by P. C. I. Crooks in an article from the Journal of the Operational Research Society [6]. The goal of this concept is to try to anticipate, by predicting the degradation process of the equipment, the replacement and consequent investment in critical equipment before it fails, avoiding the consequences and costs associated with the failure. Most recently some new theories regarding the subject of preventive maintenance have been developed. These new theories incorporate concepts such as Bayesian failure-rate modeling, combined with gamma processes, to compute maintenance schedules and preventive maintenance optimization [7].

1.4.2 Power Transformers Condition Assessment

The predictive maintenance research can be applied to many different fields including the energy sector. One of the most important components of the energy distribution grids are the power transformers (PTs), which are responsible for changing the voltage of the power being transmitted throughout the grid.

Since the PTs have an important role on the energy distribution system, it is crucial to guarantee a good assessment of their condition. The health condition of a transformer is influenced by several physical parameters, which when combined can be quantified into a health index that describes the overall health condition of the asset [8]. This health index is computed through the combination of results obtained from operating observations, field inspections and laboratory testing. The health index uses data from gases, oil and furan analysis. The dissolved gases concentration analyzed are the hydrogen (H$_2$), methane (CH$_4$), ethane (C$_2$H$_6$), ethylene (C$_2$H$_4$), acetylene (C$_2$H$_2$), carbon dioxide (CO$_2$) and carbon monoxide (CO). The parameters of the oil that are analyzed are the Break Down Voltage (BDV), Interfacial Tension (IFT), acid and water content [9]. Besides these tests, there are some other complementary analysis that can be performed like the winding resistance, sweep frequency response analysis (SFRA) and vibration analysis [10]. On Figure 1.2 are presented the main tests that can be performed in order to assess the health condition of a transformer, organized from the most important to the less.

According to [10] the main causes for transformers failures include electrical breakdown, lightning, dielectric fault, loose connection, incorrect maintenance and excessive loading. The electrical breakdown is the most common failure cause, possibly being hasten by contamination, thermal aging, repetitive cycles of excessive voltage stress and mechanical deformation. Other common failure cause is the dielectric failure, which can have a profound effect on the useful life of the transformers. On Figure 1.3 is presented the distribution of the failures causes per occurrence rate. These type of failures can belong to two distinct categories: repairable and non-repairable. Depending on the category, there will either be a replacement or a repair time. The non-repairable failures are normally associated with age failures, as explained in [11]. On Figure 1.4 is presented the distribution of the components mainly responsible for the failures in PTs.
EDPD uses a custom made index named Partial Health Index (PHI) that is similar to the HI previously explained. PHI analyzes, in a global way, the same physical conditions aforementioned and assigns to each PT a value from zero to one hundred percent, where zero is total degradation and one hundred is perfect health condition. There is another index designated Global Health Index (GHI), which is computed based on the PHI affected by an age factor, defined internally by EDPD. This GHI can then be converted into a failure index (FI), which is a value computed based on the GHI and on an environmental factor. This factors represents the probability of external occurrences enhancing the probability of a transformer outage. External occurrences can comprehend nearby trees, possibility of floods, fire, among others. All the transformers installed on the grid are cataloged and have an associated PHI, GHI, EF and FI value.

These internal indexes are deeply connected with the health condition of the PTs. It is possible that there is a relationship between the age, the failure rate and the aforementioned indexes of the transformers. If these relationships can be defined, it might be possible to predict failure probabilities for all the transformers and take action, either on the maintenance or replacement, of the more degraded equipment.
1.4.3 Power Transformers Aging Models

It is logical that all equipment degrades with age, which is the reason why there is a large amount of research regarding aging models for equipment. This research studies and models the way the health of the machinery evolves with the using time [12]. Besides the technological application, this type of research is studied in other areas such as: medicine, biology and the demography fields. It is applied to study the lifespan evolution and general aging of living beings and the human demographic evolution [13–15].

The existing aging models that estimate the evolution of the condition of equipment can be based on two types of data [16]: the historical data retrieved from the assessment of the evolution of the equipment condition throughout its working lifetime; real time acquired data monitored with sensors. This last type of data is essential to the processes associated with the Industry 4.0, where live monitoring is a major concern.

The first aging models developed to estimate the aging behavior of electrical equipment used the
so-called Average Age Method [17]. This method uses a large amount of data regarding the mean life of the equipment to compute its average lifetime. Usually this was the method used to compute the value presented in the equipment descriptions and data sheets regarding the life expectancy of each equipment. This approach is usually not accurate, since it does not include real environmental and operational conditions of the equipment. Besides, this method is not adequate for power systems components, such as generators or transformers. This type of equipment has a long useful life, for an electrical equipment, therefore there is a short amount of end-of-life failure records. As an example, the current average failure rate for PTs in the Portuguese distribution grid is around 0.005 failures per PT per year. Besides, the operating conditions of each PT can be drastically different depending on the environmental conditions and efforts supported during its lifetime.

The average age method uses only data regarding the components that have died, which is a disadvantage of the method. With the condition assessment of the equipment it is possible to take a different approach to the development of the aging model. An alternative method is the probability-distribution-based methods, which take into consideration, not only the dead components, but the condition of the operating ones as well.

When developing an aging model for electrical system components, there are two concepts regarding age measurement: natural age and functional age. The natural age is the difference between the in-service date and the present date. The functional age is the natural age minus the time the equipment has been offline due to breakdown, or just because it was not in use. For the purpose of system management planning, a rough estimate is generally sufficient and the natural age can be used. During this thesis work the natural age will be the one analyzed, since this was the only data made available by EDPD.

The estimated mean life of an equipment, together with its current age, might work as an indicator of the risk of an end of life failure. However, the development of a more complete aging model is necessary to quantify the probability of a possible failure. There is a commonly accepted curve that generally represents the evolution of the probability of failure with the age of the equipment. It is the so-called Basin Curve, which is represented in Figure 1.5. This representation states that the failure rates in electronic equipment are usually higher when the equipment is installed, due to construction or installation failures. Then, the failure rate decreases during the normal operating lifetime. As the end of life approaches the failure rate of the equipment increases again.

There are two main approaches for the maintenance processes: corrective and preventive. The corrective maintenance as the name states, consists in repairing the equipment after a failure and can only be applied to repairable failures. The preventive maintenance, on the other hand, consists in performing maintenance procedures before the equipment fails in a scheduled time frame.

There are two ways for performing preventive maintenance: regular and predictive. The regular method is the most commonly used and consists in regularly performing maintenance procedures, based on the manufacturer's specifications or the experience obtained with time. This policy is simple to implement, but can result in unnecessary maintenance activities, increasing the costs associated with the equipment. It can also lead to not enough maintenance, increasing the risks of failure. The predictive
The method is based on condition assessment mixed with mathematical models and calculations in order to predict the maintenance needs and improve the equipment reliability. According to [17], the reliability centered maintenance should include:

- collecting statistical data such as operation history, failure records, aging status tests or assessments;

- estimating failure probabilities due to repairable and end-of-life failures of equipment;

- evaluating impacts of individual equipment failures on the system;

- quantifying the effects of maintenance activities on improving equipment failure frequencies/repair times and whole system reliability;

- applying economic or reliability criteria to determine the best maintenance scheme, which may be an optimal maintenance alternative (range and sequence in maintenance) or lowest-risk maintenance scheduling (timing) or most cost-efficient workforce planning.

The available data provided by EDPD for the installed PTs regard only the current health indexes of the transformers. There is no data regarding the historical health evolution, which is why there is a need to develop a custom made model, fitting the available data. The developed model should be able to give answer to all of the items above mentioned.

In the particular case of the power transformers, there are some physical tests that can be performed in order to analyze the current health condition of the PTs. This retrieved data can be compiled into a health index (PHI) like the ones aforementioned (PHI), which can then be the basis for the predictive aging model development. Although there are some predictive models already developed for electrical equipment, most of them require more data than what is available. However, there are some articles that can guide through the development of this thesis model.
1.4.4 Markov Model based aging model

Taking into account the available data, it will be proposed the development of a model based on Markov Models Theory (MMT). These type of models can be used to determine the evolution of the health condition based only on the current state data. The Markov Models are a probability based decision process, where the next state depends only upon the current state. Therefore this type of model was selected, since the only data available regards the current health condition of the PTs. MMT has been extensively applied in other areas such as civil engineering, where it is used to model the degradation of bridges, pavement, steel hydraulic structures or water piping components [19–22].

On this thesis work it is proposed the use of this type of processes to estimate the evolution of the degradation trajectory for the health condition of the PTs, based on the current health indexes. On [23], an algorithm is suggested for the implementation of a Markov predictive model. Although the proposed approach is not a complete fit for the problem at study, it is useful to understand how this kind of problems can be modeled. In Figure 1.6 it is presented a diagram that illustrates the different phases of building a Markov model. First the historical data is collected, then some statistical methods are applied, e.g. main components analysis and time series analysis are applied to extract important data items. After that, the HI is computed using the processed acquired data. Then the Markovian deterioration model is identified and parameterized based on historical observations of the equipment health data. This process allows to establish the long-run performance of the equipment, possibly predicting it [23].

![Offline Markov Model diagram](source: [23])

The method proposed by [23], includes some steps and requirements that will not be assured during the work developed in this thesis, making it necessary to modify some of the steps of the algorithm. In Figure 1.7 it is presented a modeling process for the Markov Process definition, adapted from [24]. The first step to build the model is to retrieve the information regarding the health indexes (HI) of the
transformers analyzed. Next, the average value for the HI, depending on the age, is determined and used to compute the Markov transition matrices (one per possible PT age interval). At the end it is possible to estimate, based on the Markov Chain algorithm, the evolution of the HI of the PTs.

The procedure implemented during the thesis will be slightly different from the one aforementioned. An example of a difference is the use of a degradation index (PDI) instead of a health index. Other major difference to what is done in [19–21, 23, 24] is the definition of age intervals larger than a year, due to the lack of data regarding the PTs in order to define a yearly based model. This procedure is explained in more detail in Section 2.1.1.

A Markov decision process is characterized as a memoryless process, which in this case predicts the future condition of equipment as a probabilistic estimate [22, 25, 26]. The Markov Chain process depends on the transition probabilities given as $P_{ij}$ [26]. Where $P_{ij}$ is the probability of the equipment evolving from state condition $i$ to $j$, in a specific time interval. A set of transition probabilities can be represented by a transition matrix, $P(t)$.

The authors from the consulted bibliography make some assumptions regarding the construction of the models. Some of them will also be implemented in the developed model. The main followed assumption states that the deterioration process of the transformers is considered to be a monotonous and irreversible process. This means that the health condition of the transformers either remains in its current condition level or moves to a more degraded state condition level. In order to develop the

Figure 1.7: Modeling process of deterioration performance curve based on a Markov Model (MM). Adapted from: [24]
homogeneity of the deterioration performance curve, the transition probabilities are determined based on the transformers age groups. This zoning technique is applied to avoid over/under estimations of the transformers conditions [19, 22]. The number of state levels should be defined depending on the data from the PTs analyzed. The last level of the transition matrix should be used for future prediction. For further simplification of the Markov Chain process, the higher state condition, $P_{ii}$, is set as 1 based on the assumption that all transformers will end up in a poor condition and will not evolve to any other state.

The consulted authors [19–21, 23, 24] assume that the health condition can only degrade, at maximum, one level on each transition. It is represented by transition matrices with all elements equal to zero, with the exception of the diagonal elements and the elements next to the them, as represented by

$$P = \begin{bmatrix} P_{11} & 1 - P_{11} & 0 & 0 \\ 0 & P_{22} & 1 - P_{22} & 0 \\ 0 & 0 & P_{33} & 1 - P_{33} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1.1)$$

In this work it is proposed that this last assumption should not be followed and it is acceptable that in each transition, the health condition state of a PT can degrade more than one level. Even though there will be technical constraints regarding the computation of the matrices that will oblige to implement some maximum degradation constraints, it will be a less strict constraint, which generate transition matrices of the form:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ 0 & P_{22} & P_{23} & P_{24} \\ 0 & 0 & P_{33} & 1 - P_{33} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1.2)$$

An adequate estimation of the transition matrices, $P(t)$, is essential, since it is the main element of the Markovian process. The transition probability matrices can be determined by heuristic or statistic methods [24]. In this study, a statistical method known as the nonlinear optimization technique will be implemented. The objective of this method is to identify the values of the matrix elements that would minimize the absolute differences between the computed and predicted health index data for each transformers age group [19, 22].

After computing the transition matrices for all the age intervals, $t$, it is possible to apply the Markov Chains algorithm, which allows to obtain the probability of the next health condition state for the transformers based on the current health state.

$$\pi_{n+1} = \pi_n \times P(n), \quad (1.3)$$

where $\pi_{n+1}$ is the probability of the next condition level for a PT, $\pi_n$ is the probability distribution of the current condition level for the PT and $P(n)$ is the transition matrix associated with the age interval of the PT.

The Markov Models are also able to predict the future condition state for a number of intervals, $t$, ...
from the initial state, \( \pi_0 \), and transition matrix, \( P \), which can be seen in Equation 1.4. In this work it will be considered that, all transformers at age level zero to be at the initial state where \( \pi_0 = [1 \ 0 \ ... \ 0] \) for zone one. Since the transformers condition measurements were performed every year, \( t \) was set to one.

\[
\pi_t = \pi_0 \times P^t
\]  

(1.4)

### 1.4.5 Monte Carlo Simulation Method

The Markov Chain models when combined with the Monte Carlo Simulation theory originate the so-called Markov Chain Monte Carlo method (MCMC). This is a powerful method to generate random samples from a probability distribution, which can be used in the computing of statistical estimates. This method is particularly useful in applications where one is forming an estimate based on multivariable probability distributions or density functions, that would not be possible to obtain analytically [27, 28].

The Monte Carlo simulations are effective when modeling situations that have a high level of uncertainty in the inputs, for example calculating risks in business. This is exactly the type of problem that will be addressed during this thesis. In a simplistic way, this method allows to compute a numerical estimation for processes or functions that depend on multiple random variables. These variables could arise organically as part of the modeling of a real-life system, such as a complex road network, the transport of neutrons, or the evolution of the stock market. In many cases, however, the random objects in Monte Carlo techniques are introduced artificially, in order to solve purely deterministic problems. In this case, the MCMC simply involves random sampling from certain known probability distributions. In either the natural or artificial setting of Monte Carlo techniques the idea is to repeat the experiment a large amount of times (or use a sufficiently long simulation run) to obtain large quantities of interest using the Law of Large Numbers [29]. At the end, the average values of the sample based outputs are computed.

As explained above, a Markov Chain Model will be constructed, to represent the health evolution of the PTs, which is a stochastic process. Each Markov Chain represents a probability distribution, which can be obtained by observing the chain after a sufficient number of samples. The higher the number of samples are ran, the more closely the distribution of the sample matches the actual desired distribution.

The common problem where MCMC is applied consists on the existence of a stochastic process that generates a random variable \( X \), and the goal is to compute the expected value of another process, \( f(X) \), that depends on \( X \), \( E[f(X)] \). Usually the value for the random variable cannot be modeled by an analytic function, making it impossible to obtain a deterministic value for the \( E[f(X)] \).

Let us assume, for example, that \( X \) is a continuous random vector with an associated probability density function. The problem can then be described as

\[
E[f(X)] = \int f(x)p(x)dx,
\]

(1.5)

where the integral is computed over the domain of \( X \). The same example can be applied to discrete random variables, as the health indexes that will be studied during this work. The equivalent to the \( f \) function for the thesis will be all the outputs computed based on the health index, such as: average
ENS, failure probability per PT, average number of failures, ENS costs, among others.

Since there is a large amount of estimates obtained, it is important to compute the expected value and the variance of the outputs. Let us consider the probability density of \( y \) as \( f(y) \), then the expected value for \( y \) is

\[
\mu_y = E(y) = \int_{-\infty}^{+\infty} y f(y) dy.
\]  

(1.6)

Considering the \( N \) trials ran, an estimator for \( \mu_y \) can be a simple mean, which can be demonstrated to have an expected error of zero, making it an unbiased estimator [30],

\[
\bar{y} = \frac{1}{N} \sum y_i.
\]  

(1.7)

Since the Monte Carlo simulation involves pseudo-random draws of the inputs, a different results will be obtained each time the probabilistic analysis is performed. That is, each time the Monte Carlo simulation ran, slightly different results for \( y \) will be obtained. In the following analysis, it is shown that the variability in the results of \( y \) (i.e., how much the mean estimate varies between Monte Carlo simulation runs) depends on \( N \), the number of trials in each Monte Carlo simulation.

The variance of the obtained estimates is obtained by

\[
V(\bar{y} - \mu_y) = \frac{\sigma_y^2}{N}.
\]  

(1.8)

By having an estimator for the output value and for its variance, it is possible to estimate the probability distribution of the error. For large values of \( N \), the central limit theorem can be applied to approximate the distribution of \( \bar{y} \). This theorem states that for a sufficiently large \( N \), the distribution of \( \bar{y} - \mu_y \) will approach a normal distribution with mean value of zero and variance obtained by \( \frac{\sigma_y^2}{N} \).

This way, it is possible to define confidence intervals, where the size of the interval depends on the number of \( N \) trials. For example, for a 95% confidence on the values of \( \mu_y \), and since the normal distribution has approximately 95% of its values within \( \pm 2 \) standard deviations of the mean, it is possible to state

\[
P\{-1.96 \frac{\sigma_y}{\sqrt{N}} \leq \bar{y} - \mu_y \leq 1.96 \frac{\sigma_y}{\sqrt{N}} \} \sim 0.95.
\]  

(1.9)

It is not possible to calculate the real value for the above error estimate, since it is depend on \( \sigma_y \), which is unknown. Although it is possible to use an estimate for its value through an unbiased estimator for \( \sigma_y^2 \):

\[
s_y^2 \equiv \hat{\sigma}_y^2 = \frac{1}{N - 1} \sum_{i=1}^{N} (y_i - \bar{y})^2.
\]  

(1.10)

The MCMC method will be vastly applied throughout the developed work, in order to obtain all of the output values of the simulator, which depend directly or indirectly on the evolution of the health index. This index will be modeled by the Markov Model defined and parametrized based on the available data.
regarding the condition of the PTs installed in the Portuguese grid.

### 1.4.6 Pareto Optimality Frontier

The simulator that will be built in order to simulate the evolution of the grid condition, should allow the testing of different renewal strategies for the installed equipment. Any renewal strategy is associated with a cost, which results from the sum of different parts, such as: the investment in new equipment, the cost of repairing damaged equipment and the valuation of the ENS. Each strategy will result in a different ENS value for the simulation time.

The decision about the best strategy to be implemented will, most likely, be defined mainly based on these two values: ENS and total strategy cost. The decision making process is usually complex, which is why it is important to present the data to support the decision in a clear and intuitive form. Even though the final goal is to find the best strategy to be implemented, it will not always be possible to achieve the global optimal solution. The decision makers might not be interested in the optimal solution, particularly if these solutions are sensitive to the variable perturbations which cannot be avoided in real life scenarios. In such cases, decision makers are interested in finding the so-called robust solutions, which are less sensitive to small changes in variables [31].

The problem being analyzed is a multi-objective optimization problem (MOO), where the objectives are the total strategy cost and the ENS output values. A particular set of optimal solutions is referred to as the Pareto optimal set or Pareto frontier. By definition, Pareto solutions are considered optimal, since there are no other designs that are superior in all objectives [32, 33]. The Pareto Optimality Frontier graph is a useful tool when comparing different solutions, in order to make a decision about which one is more efficient. If a solution is part of the frontier, it means that there is no other discrete solution that can provide the same output with the same value for one of the variables and a lower value for the other.

![Pareto Optimality Frontier example](34)

This type of data visual presentation is particularly useful, since it allows to represent solutions, which can be associated with complex procedures and algorithms, on a simple and intuitive form. This can be particularly useful when the decision making process is in the hand of personnel that was not directly
involved in the computation of the solutions, or do not have enough technical background to fully understand the process.

The Pareto Optimality Frontier can also be used as an optimization tool, allowing the computation of the exact optimal solution value. For that it is necessary to define an objective function to be minimized. The objective function must depend on the variables represented by the axes. On Figure 1.9, is presented an example from [35], where a Pareto Frontier is presented, together with the optimal solution found, based on the defined objective function. The optimal solution is represented by the point where the derivative of the objective function is tangent to the Pareto Optimality Frontier. An extensive survey regarding multiobjective optimization is presented in [35]. As this topic is not part of the main scope of this thesis and is well explained in the bibliography references, it will not be presented in a more extensive form.

![Pareto Frontier Diagram](image)

Figure 1.9: Example of optimal point computation using Pareto Frontier. Source: [35]
Chapter 2

Aging Model for Power Transformers

2.1 Time and Health discretization

As explained in the introduction chapter 1, one of the objectives of this thesis is the development of an aging model for the PTs, in order to simulate the evolution of the stock health condition throughout the upcoming years. The evolution process should have a so-called monotonous trajectory, meaning that the health condition of the transformers should never improve with time. In order to assure this requirement, it is necessary to develop a special stochastic process based on Markov chains. The Markov chains can only be applied when the evolution to a next state depends only upon the current state and those states are discrete and well defined, which is why it is necessary to define those discrete levels of health condition and age for the transformers that will be modeled.

The base for the parameterization of the model is the current state of the transformers stock. Using the health and age data of the installed PTs in the grid, it is possible to define some patterns. The main criteria used to make the discretization of the health and age of the transformers is the need to generate groups that have a sufficient level of significance in terms of statistic relevance. Therefore, groups that are too small cannot be defined and must be be merged with neighbor groups.

2.1.1 Age Discretization

Although the aging model can be develop to represent the PTs yearly health condition evolution, this approach will not be applied. The main problem with this approach is the scarcity of available data for each transformers age, which would render model meaningless. This is, there are not enough PTs within one year intervals, in order to develop a yearly model. As an alternative the defined model will make the analysis of the health condition in five-year intervals. This way, it is guaranteed that there is enough data to analyze in practically every age group, making it possible to model its behavior. The intervals that do not have enough data to analyze should be merged with a neighbor interval. There is an associated advantage in defining five-year interval, which is that the Distribution Grid Plan of Development and Investment (PDIRD) is also made for five-years periods. This way it is easier to plan the investments for the next 5 years if the model outputs predictions for the same time interval.
In the Table 2.1 the age intervals are presented together with its limits and the amount of transformers within each interval. It is also presented the average age of the PTs per interval and the probability of a PT being in each interval.

Table 2.1: Age intervals discretization

<table>
<thead>
<tr>
<th>Age Level</th>
<th>Minimum Age</th>
<th>Maximum Age</th>
<th>Average Age</th>
<th>number of PTs</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>3.11</td>
<td>31</td>
<td>0.045</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
<td>7.74</td>
<td>46</td>
<td>0.060</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>15</td>
<td>12.82</td>
<td>103</td>
<td>0.129</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>20</td>
<td>17.15</td>
<td>20</td>
<td>0.020</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>25</td>
<td>22.97</td>
<td>65</td>
<td>0.091</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
<td>30</td>
<td>27.75</td>
<td>51</td>
<td>0.080</td>
</tr>
<tr>
<td>7</td>
<td>31</td>
<td>35</td>
<td>33.18</td>
<td>74</td>
<td>0.077</td>
</tr>
<tr>
<td>8</td>
<td>36</td>
<td>40</td>
<td>37.47</td>
<td>174</td>
<td>0.257</td>
</tr>
<tr>
<td>9</td>
<td>41</td>
<td>45</td>
<td>42.75</td>
<td>77</td>
<td>0.100</td>
</tr>
<tr>
<td>10</td>
<td>46</td>
<td>50</td>
<td>46.91</td>
<td>47</td>
<td>0.066</td>
</tr>
<tr>
<td>11</td>
<td>51</td>
<td>55</td>
<td>52.18</td>
<td>35</td>
<td>0.052</td>
</tr>
<tr>
<td>12</td>
<td>56</td>
<td>62</td>
<td>57.20</td>
<td>16</td>
<td>0.023</td>
</tr>
</tbody>
</table>

2.1.2 Partial Health Index Discretization

In order to create a Markovian model for the evolution of the PTs health condition it is necessary to define discrete health levels, since this kind of processes cannot have a continuous range of values. The data that was made available by EDP Distribuição, had a partial health index (PHI) defined for each PT, which will be the starting point to define the health levels of the model. This health index evaluates the condition of the PTs based only upon physical tests performed on the equipment. This index will be used instead of the Global Health Index (GHI), also defined by EDPD for each PT, because this last one takes into consideration the age of the PTs. The PHI is the ideal index to develop the model because the objective is to find the relationship between the age and the health evolution of the transformers. The use of the GHI would be biased since the index is already affected by the age.

Each PT has an associated PHI which can go from 0%, meaning that the transformer is completely deteriorated, to 100%, which means that it is in perfect condition. It is intuitive that with aging, the health condition of the PTs gets deteriorated. Similarly to the age discretization presented in Table 2.1, the distribution of the number of PTs per PHI level can be consulted in the Table 2.2. It is also presented the limits of each interval, together with the average value and the probability of a PT being in each interval. The criteria used to define the intervals was obtained by considering the balance between the amount of transformers in each level, the extent of the interval and its statistical significance. Even though the interval distribution is not perfectly balanced, as can be seen from the first level where there are only 10 PTs with an PHI above 95, it is precise enough to fulfill the needs for the model development.
Table 2.2: Partial Health Index (PHI) levels discretization

<table>
<thead>
<tr>
<th>PHI Level</th>
<th>Maximum PHI</th>
<th>Minimum PHI</th>
<th>Average PDI</th>
<th>Amount of PTs</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0</td>
<td>95.0</td>
<td>98.78</td>
<td>10</td>
<td>0.013</td>
</tr>
<tr>
<td>2</td>
<td>95.0</td>
<td>89.0</td>
<td>89.83</td>
<td>65</td>
<td>0.085</td>
</tr>
<tr>
<td>3</td>
<td>89.0</td>
<td>85.0</td>
<td>86.66</td>
<td>65</td>
<td>0.085</td>
</tr>
<tr>
<td>4</td>
<td>85.0</td>
<td>81.0</td>
<td>83.48</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>5</td>
<td>81.0</td>
<td>77.0</td>
<td>78.52</td>
<td>64</td>
<td>0.083</td>
</tr>
<tr>
<td>6</td>
<td>77.0</td>
<td>74.0</td>
<td>74.58</td>
<td>69</td>
<td>0.090</td>
</tr>
<tr>
<td>7</td>
<td>74.0</td>
<td>71.5</td>
<td>72.49</td>
<td>58</td>
<td>0.075</td>
</tr>
<tr>
<td>8</td>
<td>71.5</td>
<td>70.0</td>
<td>71.21</td>
<td>61</td>
<td>0.079</td>
</tr>
<tr>
<td>9</td>
<td>70.0</td>
<td>68.5</td>
<td>69.04</td>
<td>39</td>
<td>0.051</td>
</tr>
<tr>
<td>10</td>
<td>68.5</td>
<td>67.0</td>
<td>68.02</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>11</td>
<td>67.0</td>
<td>65.5</td>
<td>66.03</td>
<td>38</td>
<td>0.049</td>
</tr>
<tr>
<td>12</td>
<td>65.5</td>
<td>64.0</td>
<td>64.86</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>13</td>
<td>64.0</td>
<td>61.0</td>
<td>62.69</td>
<td>42</td>
<td>0.055</td>
</tr>
<tr>
<td>14</td>
<td>61.0</td>
<td>58.0</td>
<td>59.61</td>
<td>54</td>
<td>0.070</td>
</tr>
<tr>
<td>15</td>
<td>58.0</td>
<td>0.0</td>
<td>53.94</td>
<td>48</td>
<td>0.062</td>
</tr>
</tbody>
</table>

2.1.3 Partial Degradation Index (PDI)

For convenience it is useful to define a new index called Partial Degradation Index (PDI), which is defined as

\[ PDI = 1 - PHI. \]  \hspace{1cm} (2.1)

This new index will be particularly useful, since although the PHI can never have values above 100%, the young PTs can have PHI values very close to 100%. This can lead to difficulties in the model developing, specifically during the simulation of the future PHI levels for the PTs. The simulation method, as will be explained later, resorts to random number generation, according to a probabilistic distributions associated with each PT, which can lead to obtaining PHI levels above 100%.

The PDI can never have a value below 0%. Throughout the values generation, it is easier to guarantee a PDI value above 0%, than a PHI value below 100%. This guarantee can be assured through the use of distributions that are only defined for positive values such as the log-normal, exponential, gamma or Weibull distributions.

The PDI levels discretization and the amount of PTs that are in each interval are presented in Table 2.3.
Table 2.3: Partial Degradation Index (PDI) levels discretization

<table>
<thead>
<tr>
<th>PDI Level</th>
<th>Minimum PDI</th>
<th>Maximum PDI</th>
<th>Average PDI</th>
<th>Amount of PTs</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>5.0</td>
<td>1.22</td>
<td>10</td>
<td>0.013</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>11.0</td>
<td>10.17</td>
<td>65</td>
<td>0.085</td>
</tr>
<tr>
<td>3</td>
<td>11.0</td>
<td>15.0</td>
<td>13.34</td>
<td>65</td>
<td>0.085</td>
</tr>
<tr>
<td>4</td>
<td>15.0</td>
<td>19.0</td>
<td>16.52</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>5</td>
<td>19.0</td>
<td>23.0</td>
<td>21.48</td>
<td>64</td>
<td>0.083</td>
</tr>
<tr>
<td>6</td>
<td>23.0</td>
<td>26.0</td>
<td>25.42</td>
<td>69</td>
<td>0.090</td>
</tr>
<tr>
<td>7</td>
<td>26.0</td>
<td>28.5</td>
<td>27.51</td>
<td>58</td>
<td>0.075</td>
</tr>
<tr>
<td>8</td>
<td>28.5</td>
<td>30.0</td>
<td>28.79</td>
<td>61</td>
<td>0.079</td>
</tr>
<tr>
<td>9</td>
<td>30.0</td>
<td>31.5</td>
<td>30.96</td>
<td>39</td>
<td>0.051</td>
</tr>
<tr>
<td>10</td>
<td>31.5</td>
<td>33.0</td>
<td>31.98</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>11</td>
<td>33.0</td>
<td>34.5</td>
<td>33.97</td>
<td>38</td>
<td>0.049</td>
</tr>
<tr>
<td>12</td>
<td>34.5</td>
<td>36.0</td>
<td>35.14</td>
<td>52</td>
<td>0.068</td>
</tr>
<tr>
<td>13</td>
<td>36.0</td>
<td>39.0</td>
<td>37.31</td>
<td>42</td>
<td>0.055</td>
</tr>
<tr>
<td>14</td>
<td>39.0</td>
<td>42.0</td>
<td>40.39</td>
<td>54</td>
<td>0.070</td>
</tr>
<tr>
<td>15</td>
<td>42.0</td>
<td>100.0</td>
<td>46.06</td>
<td>48</td>
<td>0.062</td>
</tr>
</tbody>
</table>

2.1.4 Age intervals merging

As previously discussed there are some age intervals which need to be merged with neighboring intervals for a variety of reasons, as it is explained below:

- PTs that have an age between 16 and 25 years will be considered to be in the same age group. The merging can be explained since there are only 13 transformers between the ages of 16 and 20 years, meaning the sample is not significant by herself.

- For the age interval 26-35 years, the merging is made in order to harmonize an outlier value. The interval between 26 and 30 years had an higher average PDI than the the interval 31 and 35 years, which is not logical. It is not expected that a group of younger PTs is more degraded, in average, than a group of older ones. The value corresponding to the 26-30 age interval stands out from the rest of the intervals, which is why it is considered to be an outlier interval. There are two options to solve an outlier point, the value can be ignored or it can be normalized to the rest of the data. Making the fusion between the 26-30 and the 31-35 age interval is a solution for the outlier problem. This option is possible to be applied since none of the intervals has an amount of PTs big enough that makes it too significant to be merged.

- The interval for transformers with ages between 46 and 62 was defined because for PTs with over 46 year it is not clear what is the behavior of the PDI distribution, so it is not possible to model it clearly, as will be demonstrated in the next section. This way, this interval was defined to contain all of the transformers that cannot be modeled in a feasible way.

It is interesting to compare the mean PDI values for each age interval before and after the merging.
to see how different those values are. The comparison is presented in Tables 2.4 and 2.5. Notice that the outlier values referenced before are highlighted in the tables.

The complete distribution of the PTs per each PDI level discriminated by age range are presented in the tables from Appendix B. Notice that the first age intervals have a very high probability that a PT has a low PDI and the probability of having high PDI is lower for higher values. This illustrates the bigger probability of younger PTs to have a low degradation level. This behavior will lead to a very particular type of probability distribution which will model these intervals, as is explained in Subsection 2.2.1

### Table 2.4: Average PDI values for each age group before merging of intervals

<table>
<thead>
<tr>
<th>Minimum Age</th>
<th>Maximum Age</th>
<th>Frequency</th>
<th>PDI Average value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>5</td>
<td>36</td>
<td>2.170</td>
</tr>
<tr>
<td>6.0</td>
<td>10</td>
<td>48</td>
<td>3.440</td>
</tr>
<tr>
<td>11.0</td>
<td>15</td>
<td>103</td>
<td>5.630</td>
</tr>
<tr>
<td>16.0</td>
<td>20</td>
<td>20</td>
<td>5.200</td>
</tr>
<tr>
<td>21.0</td>
<td>25</td>
<td>65</td>
<td>7.980</td>
</tr>
<tr>
<td>26.0</td>
<td>30</td>
<td>53</td>
<td>9.580</td>
</tr>
<tr>
<td>31.0</td>
<td>35</td>
<td>74</td>
<td>8.010</td>
</tr>
<tr>
<td>36.0</td>
<td>40</td>
<td>182</td>
<td>9.040</td>
</tr>
<tr>
<td>41.0</td>
<td>45</td>
<td>85</td>
<td>9.540</td>
</tr>
<tr>
<td>46.0</td>
<td>50</td>
<td>50</td>
<td>11.580</td>
</tr>
<tr>
<td>51.0</td>
<td>55</td>
<td>37</td>
<td>10.190</td>
</tr>
<tr>
<td>56.0</td>
<td>60</td>
<td>16</td>
<td>9.380</td>
</tr>
</tbody>
</table>

1 = Outlier value
2 = Insufficient amount of PTs to make a significant interval

### Table 2.5: Average PDI values for each age group after merging of intervals

<table>
<thead>
<tr>
<th>Minimum Age</th>
<th>Maximum Age</th>
<th>Frequency</th>
<th>PDI Average value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>5</td>
<td>36</td>
<td>2.170</td>
</tr>
<tr>
<td>6.0</td>
<td>10</td>
<td>48</td>
<td>3.440</td>
</tr>
<tr>
<td>11.0</td>
<td>15</td>
<td>103</td>
<td>5.630</td>
</tr>
<tr>
<td>16.0</td>
<td>25</td>
<td>85</td>
<td>7.330</td>
</tr>
<tr>
<td>26.0</td>
<td>35</td>
<td>127</td>
<td>8.680</td>
</tr>
<tr>
<td>36.0</td>
<td>40</td>
<td>182</td>
<td>9.040</td>
</tr>
<tr>
<td>41.0</td>
<td>45</td>
<td>85</td>
<td>9.540</td>
</tr>
<tr>
<td>46.0</td>
<td>62</td>
<td>103</td>
<td>10.740</td>
</tr>
</tbody>
</table>

2.2 PDI probability distributions per age group

The transformers inserted in each age range have a specific probability of having one of the defined PDI levels. Modeling this probability distribution will be particularly useful for the simulation of the health/degradation evolution with time. At Table 2.1, 12 age intervals were defined for the PTs, meaning that there should exist 12 correspondent probability distributions for the PDI. Those distributions have a type, a mean and a variance, which will be the parameters used to simulate the evolution of the transformers health in time.

#### 2.2.1 PDI log-normal Distribution Fitting

For the transformers inserted in age intervals below 40 years, the log-normal distribution is a good fit for the PDI values of each age range. As can be seen by the PDF and CDF fittings over the data presented in Figures 2.1 up to 2.6. This fittings were made using the Matlab Distribution Fitter app, which is incorporated in the Statistics and Machine Learning Toolbox [36] [37].
Figure 2.1: Log-normal Distribution fit over the PDI Data for PTs with ages between 1 and 5 year

(a) PDF distribution fit  
(b) CDF distribution fit

Figure 2.2: Log-normal distribution Fit over the PDI Data for PTs with ages between 6 and 10 year

(a) PDF distribution fit  
(b) CDF distribution fit

Figure 2.3: Log-normal distribution Fit over the PDI Data for PTs with ages between 11 and 15 year

(a) PDF distribution fit  
(b) CDF distribution fit
Figure 2.4: Log-normal distribution Fit over the PDI Data for PTs with ages between 16 and 25 year

Figure 2.5: Log-normal distribution fit over the PDI Data for PTs with ages between 26 and 35 year

Figure 2.6: Log-normal distribution Fit over the PDI Data for PTs with ages between 36 and 40 year
2.2.2 Extreme Value Distribution Fitting

For the PTs inserted in higher age groups the log-normal distribution is not the best fit. The extreme value distribution is a more accurate fit for the PDI values of the two higher age groups. This difference to the distribution fittings of the lower age groups happens for two reasons:

- The peak of data, this is, the PDI level with the biggest frequency, is too high in in these groups to be modeled by the log-normal distribution. This can be explained by an increase in the service time of the PTs, therefore it is normal that the mean value of the PDI also gets higher, this is, it is more likely for a transformer to be degraded.

- For the PTs with an age higher than 45 years old the distribution of the correspondents PDI is not simple to model. For each of the five-year intervals above 45 years, the distribution of the PDI values have a behavior similar to white noise. This is why it was necessary to create a bigger age group from 46 to 62 years, as it was explained in Subsection 2.1.4. After taking into consideration this bigger age group, the data could be better fitted by an extreme value distribution, than by a log-normal distribution.

The result of the fitting over the data of the two higher age groups can be seen in Figures 2.7 and 2.8.

![PDF distribution fit](image1)
![CDF distribution fit](image2)

Figure 2.7: Extreme Value Distribution fit over the PDI Data for PTs with ages between 41 and 45 year
2.2.3 Evolution of PDI mean value and variance evolution with age

Based on the mean values and variances for the PDI values per age group, it is interesting to study if there is a correlation between the evolution of the index and the average age of the transformers per interval. To do so, it is necessary to plot the mean values and variances as a function of the mean age of the PTs in each interval. Using the age intervals defined in Table 2.5, the values used as data for the plotting are presented in Table 2.6.

Table 2.6: PDI mean values and variances discretized by age groups

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Minimum Age</th>
<th>Maximum Age</th>
<th>Amount of PTs</th>
<th>Average Age</th>
<th>PDI Mean Value</th>
<th>PDI Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>36</td>
<td>3.42</td>
<td>2.17</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
<td>48</td>
<td>8.17</td>
<td>3.44</td>
<td>2.72</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>15</td>
<td>103</td>
<td>13.22</td>
<td>5.63</td>
<td>7.44</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>25</td>
<td>85</td>
<td>21.98</td>
<td>7.33</td>
<td>11.44</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>35</td>
<td>127</td>
<td>31.28</td>
<td>8.68</td>
<td>15.88</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>40</td>
<td>182</td>
<td>37.68</td>
<td>9.04</td>
<td>13.58</td>
</tr>
<tr>
<td>7</td>
<td>41</td>
<td>45</td>
<td>85</td>
<td>42.28</td>
<td>9.54</td>
<td>15.49</td>
</tr>
<tr>
<td>8</td>
<td>46</td>
<td>62</td>
<td>103</td>
<td>50.74</td>
<td>10.74</td>
<td>12.08</td>
</tr>
</tbody>
</table>

Plotting the mean PDI value per interval as a function of the mean age of the same interval makes it simple to see that there is tendency for the increase of the average PDI with the age. This relationship can be modeled as a second degree exponential function

\[ PDI = ae^{bx} + ce^{dx}, \]  

where \( a = 6.951 \pm 3.2 \) \( b = 0.008407 \pm 0.0093 \) \( c = -6.912 \pm 3.081 \) \( d = -0.08172 \pm 0.0596 \) with 95% confidence bounds. This equation is a good fit for the data evolution, since the R-squared value of the fitting is 0.9954 and the adjusted R-squared is 0.9927. Graphically its also intuitive to see that there is a correlation between the equation obtained and the data, as can be seen in Figure 2.9. The plotting and fitting of the data it was used the Matlab Curve Fitting App [36] [37].
Figure 2.9: Data and fitting of the PTs PDI average mean value evolution with age

For the variance values, there is also a clear correlation between the average age of each interval and the variance of the PDI for the PTs within the same interval, as can be seen in Figure 2.10.

The evolution of the variance with the average of the interval can be defined as a second degree polynomial function

\[ \text{Var}(PDI) = ax^2 + bx + c, \]  

(2.3)

where \( a = -0.009022 \pm 0.004228 \), \( b = 0.7198 \pm 0.1775 \), and \( c = 0 \) with 95% confidence bounds. This fitting has a R-squared value of 0.9456 and an adjusted R-squared value of 0.9379, which are still acceptable values for a tendency fitting.

Figure 2.10: Data and fitting of the PTs PDI Variance evolution with age

The fittings obtained above will be particularly useful to create the probability matrix that will guide the generation of the PDI values for each PT according to the age group they are inserted in. This will allow, together with the Markov Chains theory, the generation of health evolution trajectories for each transformer, in order to make a prediction on how the average condition of the installed stock will evolve
in the years to come.

2.3 Stationary PDI Probability Distribution Matrix

In order to determine how the PDI levels distribute per each age interval, it is necessary to define the so-called PDI probability matrix. There are different approaches to the definition of this matrix, which are explained individually in each of the following subsections.

2.3.1 Matrix Computation Based on Real Data

Based in the data for the current PDI values of PTs installed in the grid it is possible to define the probability of a PT being in each PDI level depending on its age. All together those values generate the PDI probability matrix presented at Table 2.7. This probability distribution matrix for the PDI levels, which is based in the real data cannot be used as a base for the definition of the desired predictive model, since it has too many outliers. As can be seen in the table, there are two reason for a value to be considered an outlier. One is if there are two relative maximums of probability within the same age range. The second reason is if, for the same PDI level, there is an increase in probability while, for a younger age range, the probability was in decrease (e.g. the probability of having a level one PDI is zero for the age range three and four but, for the age range five, the probability increases to 0.008. This point is considered an outlier).

Table 2.7: PDI probability matrix based in real data

<table>
<thead>
<tr>
<th>Age</th>
<th>PDI 1</th>
<th>PDI 2</th>
<th>PDI 3</th>
<th>PDI 4</th>
<th>PDI 5</th>
<th>PDI 6</th>
<th>PDI 7</th>
<th>PDI 8</th>
<th>PDI 9</th>
<th>PDI 10</th>
<th>PDI 11</th>
<th>PDI 12</th>
<th>PDI 13</th>
<th>PDI 14</th>
<th>PDI 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22</td>
<td>0.33</td>
<td>0.17</td>
<td>0.09</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.31</td>
<td>0.29</td>
<td>0.13</td>
<td>0.10</td>
<td>0.06</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>0.12</td>
<td>0.16</td>
<td>0.09</td>
<td>0.19</td>
<td>0.17</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.07</td>
<td>0.09</td>
<td>0.02</td>
<td>0.12</td>
<td>0.18</td>
<td>0.09</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.13</td>
<td>0.12</td>
<td>0.04</td>
<td>0.06</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.08</td>
<td>0.08</td>
<td>0.06</td>
<td>0.07</td>
<td>0.09</td>
<td>0.07</td>
<td>0.11</td>
<td>0.05</td>
<td>0.12</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.08</td>
<td>0.06</td>
<td>0.11</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.06</td>
<td>0.18</td>
<td>0.07</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.05</td>
<td>0.06</td>
<td>0.11</td>
<td>0.04</td>
<td>0.08</td>
<td>0.06</td>
<td>0.07</td>
<td>0.12</td>
<td>0.10</td>
<td>0.12</td>
<td>0.17</td>
</tr>
</tbody>
</table>

2.3.2 Matrix computation based in the fitted probabilistic distributions

As a solution for avoiding so many outliers as in Table 2.7, it was developed a different PDI probability matrix (Table: 2.8), this time based in the probability distributions obtained in Subsections 2.2.1 and 2.2.2, and using as mean and variance parameters the values obtained in the fitting performed in 2.2.3. In this case instead of the real raw data, it is used the data generated from a probability distribution, where each age range has a specific and always growing mean value for the distribution together with a specific variance value. This process works as a filter that harmonizes the values between age levels, eliminating most of the outliers.

Notice that in this new table there are 10 age intervals instead of 8, each with 5 years. This is due to the fact that with the mean and variance values obtained from the fitting function from section 2.2.3 it
is possible to assume that the intervals below 40 years which were merged with each other also have a probability distributions with a specific mean value and variance. The two higher age intervals do not follow a log-normal distribution, but an extreme value distribution as was shown before, this is why there is still that one outlier that pops out in age range 9.

Despite the solving of the outliers issue, this is not yet the final probability matrix. Due to feasibility problems when calculating the transition matrices, the distributions need to be truncated. This procedure will originate a new probability matrix as, explained in Subsection 2.4.2.

Table 2.8: PDI probability matrix based in the probability distributions obtained in 2.2.1

<table>
<thead>
<tr>
<th>Age</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16</td>
<td>0.45</td>
<td>0.26</td>
<td>0.09</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.04</td>
<td>0.21</td>
<td>0.26</td>
<td>0.20</td>
<td>0.12</td>
<td>0.07</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
<td>0.11</td>
<td>0.19</td>
<td>0.19</td>
<td>0.15</td>
<td>0.11</td>
<td>0.08</td>
<td>0.05</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.07</td>
<td>0.14</td>
<td>0.16</td>
<td>0.15</td>
<td>0.12</td>
<td>0.10</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>0.05</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.12</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.04</td>
<td>0.09</td>
<td>0.12</td>
<td>0.13</td>
<td>0.12</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.03</td>
<td>0.07</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.02</td>
<td>0.06</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.10</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
<td>0.11</td>
<td>0.08</td>
</tr>
</tbody>
</table>

1 = Outlier value

2.3.3 Corrections for parametrization feasibility

In Subsection 2.4.2 it is shown that there is a technical constraint that needs to be assured in order to compute the PDI transition matrices. These matrices will define how the degradation index from each PT evolves along the time in order for the average PDI distribution to follow the defined probability matrix. This constraint defines that from one age interval to the next there can only exist two new PDI levels with probability different from zero. This means that within a five-year distance there cannot exist a top degradation higher than two level above the previous maximum. Although this assumption is an approximation, it is absolutely crucial in order to assure the feasibility of the problem.

In order to apply this constraint it was used the truncate tool from Matlab which, as the name indicates, truncates probability distributions within intervals defined by the user. The most useful characteristic of this tool is the preservation, if possible, of the mean and variance value, independently of the cut. The final PDI probability matrix after the appliance of the constraint is the one presented at Table 2.9, which does not greatly differentiate from the previous one.
modeling of the health degradation for PTs in higher age ranges, specially above 45 years. If a transformer is of age, it gets totally degraded and fails. This way it is possible to counter balance the not so accurate approach to model this phenomenon is to assume a rather drastic measure, which is to define the PDI years ago there were already PTs installed, those transformers all failed. This way, a more accurate Total degradation due to age limit

There is no data concerning transformers above the age of 62 years, which means that, since sixty five years ago there were already PTs installed, those transformers all failed. This way, a more accurate approach to model this phenomenon is to assume a rather drastic measure, which is to define the PDI for all the PTs above 65 years to be 100%. Meaning that when a PT overcomes the barrier of 65 years of age, it gets totally degraded and fails. This way it is possible to counter balance the not so accurate modeling of the health degradation for PTs in higher age ranges, specially above 45 years.

Comparison between the final used PDI probability matrix and the original based in the real data

It is possible to see that there is not a major difference between the original probability matrix obtained with the real PDI data and the one obtained through the fitted and later truncated probability distributions. On Table 2.10 the absolute differences between the corresponding values of the aforementioned tables are presented. The mean difference between the real data table and the final one is of 4% which is an acceptable error, specially when compared with all the possible problems that were solved through the applied method.

Notice again that the final probability matrix has 10 age levels and the real one only has 8, this is due to the merging of age levels done in the real data, which could be solved through the fitting of the probability distributions.

Table 2.10: Comparison between the original PDI probability matrix and the final obtained

<table>
<thead>
<tr>
<th>Age</th>
<th>PDI</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16</td>
<td>0.46</td>
<td>0.26</td>
<td>0.09</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>0.22</td>
<td>0.27</td>
<td>0.21</td>
<td>0.13</td>
<td>0.08</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
<td>0.12</td>
<td>0.20</td>
<td>0.20</td>
<td>0.16</td>
<td>0.12</td>
<td>0.08</td>
<td>0.06</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>0.08</td>
<td>0.15</td>
<td>0.17</td>
<td>0.16</td>
<td>0.19</td>
<td>0.10</td>
<td>0.08</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
<td>0.05</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.13</td>
<td>0.11</td>
<td>0.09</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>0.04</td>
<td>0.09</td>
<td>0.12</td>
<td>0.13</td>
<td>0.12</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>7</td>
<td>0.00</td>
<td>0.03</td>
<td>0.07</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.02</td>
<td>0.06</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>9</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.10</td>
<td>0.08</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.11</td>
<td>0.13</td>
<td>0.14</td>
<td>0.14</td>
<td>0.11</td>
<td>0.11</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Total degradation due to age limit

There is no data concerning transformers above the age of 62 years, which means that, since sixty five years ago there were already PTs installed, those transformers all failed. This way, a more accurate approach to model this phenomenon is to assume a rather drastic measure, which is to define the PDI for all the PTs above 65 years to be 100%. Meaning that when a PT overcomes the barrier of 65 years of age, it gets totally degraded and fails. This way it is possible to counter balance the not so accurate modeling of the health degradation for PTs in higher age ranges, specially above 45 years.
2.4 PDI trajectories simulation

The objective of the first part of the thesis is to develop an aging model that can represent the evolution of the health/degradation of each PT with time. This evolution is the so-called degradation trajectory. In the previous section it was defined what is the probability of a PT having a specific degradation level according with its age. In this section it will be presented the method used to guarantee that the model outputs a simulation that follows this probabilities constraint.

With the data of the PDI levels distribution according to each age that is presented in Section 2.3 it would be possible to generate random numbers that follow that distribution and use them as the degradation level of each PT along each time interval of the simulation. This way each PDI level evolution simulated for a PT would be independent of the level it had in the time period before the transition. This method is not valid for the needs of this problem because it could lead to a PT having a better PDI level from one year to the next, meaning that the degradation would not be monotonous, which is not acceptable. Even though, in real life there is the possibility of making interventions on the PTs, which could lead to an improve in the health status of the transformer, the frequency of this type of procedure is so low that it was not considered during this study.

2.4.1 Markov chains method for the degradation model

Markov chains method for the degradation model

In order to find an alternative that meets the need for a monotonous degradation of the PT, it was used the Markov Chains theory, which defines that the transition to a state level depends on its current state. In the end it was obtained a total of 12 matrices, 15 by 15, that represent the transition probabilities between the 15 PDI levels for the 12 different five-year age intervals, up to a maximum of 60 years.

In Subsection 2.4.2 the method, the constraints applied and the software used in order to compute all the 12 matrices are explained in detail. All of the transition matrices are available in Appendix C. All the matrices are upper triangular, which means that the probability of evolving from an higher PDI level to a lower one is zero, as desired. Besides that, it is also possible to see that, at maximum, from one year to the next there are only two new PDI levels available. This is due to a constraint that had to be applied in order to guarantee the feasibility of the matrices computation. If more levels were to be introduced the problem would get too many free variables and would not be not solvable. This constraint is what originated the final probability matrix (Table 2.9) which will be the one applied.

As an example, in Table 2.11, it is shown the first transition matrix. This matrix represents the probabilities for the evolution of a brand new PT introduced in the grid. This corresponds to the PDI transition between the year zero and the age interval, which corresponds to the interval from one to five years. Although the matrix is a 15 by 15 matrix, since all the other values besides the diagonal are zero, they are not represented. It is possible to see that there is only the probability to jump from the PDI level 1 to the levels 1 to 5. This is because a brand new PT has an PDI level of 1, and from the data we can see that in the interval from 1 to 5 years the maximum PDI level is 5.
Table 2.11: Chopped Markov transition Matrix for brand new PTs. transition from age 0 to age interval 1 (1 to 5 years)

<table>
<thead>
<tr>
<th>Starting PDI Level</th>
<th>Final PDI Level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.158</td>
<td>0.457</td>
<td>0.264</td>
<td>0.093</td>
<td>0.028</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

In the Figure 2.11 it is shown a small example of the application of the 1st Markov transition matrix to the PTs which are at PDI level zero. The example shows how the next level depends on the current one, which is the 1st PDI level since the PTs are brand new. For the rest of the cases it would be the same process but for higher age ranges.

![Markov transition matrix graphical example](image)

Notice that for the probability matrix defined in Subsection 2.3.3 the last age range with information about the probability of each PDI correspond to the transformers with ages between 46 and 62 years. This way, the information about the PDI distribution for PTs in each of the intervals merged in that range is not defined. This occurs because, as was seen before, there is not enough information about those PTs. Besides, the information available does not present a tendency or clear distribution, which results in each of these intervals not being modeled individually. The adopted solution to solve this problem.
was to apply the last transition matrix for those three age intervals. This way the 11th and 12th transition matrices will be replicas of the 10th. This is a way of assuming the hypothesis that above 46 years of age the PTs maintain the same degradation pattern.

PDI trajectories simulation method

With the PDI transition matrices obtained in Subsection 2.4.1 and presented in Appendix C it is possible to simulate the health evolution trajectories for each PT. Take as an example a brand new PT, which has an age equal to zero and consequentially a PDI level equal 1. This will be used to simulate what will the PDI level on age range 1 be, meaning after a five-year period. It is necessary to generate a random number between 0 and 1, which will define according to the values in the first transition matrix, what will be the transition that should be applied and consequently the next PDI level of the PT. For the next age ranges PDI estimate the process is similar. For each new prediction it is generated a new random number using the Matlab rand function. This way it is possible to predict, based on all of the collected and processed data, what the evolution of the PDI levels will be like.

2.4.2 Transition matrices computation method

The transition matrices were computed using an optimization program named GAMS (General Algebraic Modeling System), which solves non linear problems with multiple constraints in a very intuitive and user friendly way. This program can be connect with Matlab and MS Excel in order to exchange and treat data in a simpler way. GAMS has different available methods that can be used to solve a wide spectrum of optimization problems. In this particular case, the used solver is called IPOPT (Interior Point Optimizer) which is an open-source solver for large-scale nonlinear programming (NLP). IPOPT implements an interior point line search filter method for nonlinear programming models, on which functions can be nonconvex, but should be twice continuously differentiable. The performance and robustness of IPOPT on larger models heavily relies on the used solver for sparse symmetric indefinite linear systems [38, 39]. To find more information about the solver itself it is recommended to check the manual [40]. The code of this solver has been written primarily by Andreas Wächter.

Each computed transition matrix needs to guarantee that, when multiplied by the PDI probabilities vector of the starting age interval, the result would be the desired PDI probabilities vector of the next age interval. This condition is the main constraint implemented in the optimization problem developed in order to compute these matrices. Besides this constraint, there were others applied in order to model the PDI evolution in a more realistic way. The applied constraints are explained next:

• **1st constraint** - The sum of each row of each probability matrix $P$ should be equal to one.

$$\forall i, \sum_{j=1}^{n} p_{ij} = 1 \quad (2.4)$$

This constraint assures that the sum of the probabilities of evolving to any other state or maintaining the condition is equal to 1.
• **2nd constraint** - Each transition matrix should guarantee the enforcement of the PDI probability matrix.

\[
[P_{PDI}]_{(n-1)} \times [TM]_n = [P_{PDI}]_n,
\]

(2.5)

where \([P_{PDI}]_{(n-1)}\) is the PDI probability distribution vector for the starting age interval \(n-1\), \(TM\) is the transition matrix associated with a particular age interval \(n\) and \([P_{PDI}]_n\) is the PDI probability distribution vector for the new age interval \(n\) of the PT.

Each computed matrix should guarantee that when multiplied by the PDI probabilities vector of the starting age interval the result is the desired PDI probabilities vector of the next age interval.

• **3rd constraint** - The probability of maintaining a PDI level is higher at a higher PDI levels when compared with the ones from lower PDI levels.

\[
\forall i, \quad p(i, i) > p(i - 1, i - 1)
\]

(2.6)

This constraint guarantees that if a PT is already at a high PDI level, the probability of degradation is lower than for a PT which has a lower PDI level. If a PT recently had an increase in the degradation index, it is less likely to degrade again than a PT which has not degraded in a larger time period.

• **4th constraint** - The probability of degrading slowly and incrementally should be higher than degrading drastically.

\[
\forall i, j, \quad p(i, j) > p(i, j + 1)
\]

(2.7)

This constraint can only be applied to the most recent PDI levels. For a PT with 50 years which still has a very low PDI level it is not probable that the degradation process will be smooth, for these cases this constraint should not be implemented.

• **5th constraint** - All matrices should be upper triangular. There can never occur a transition to a lower PDI level.

\[
\forall j < i \quad p_{ij} = 0
\]

(2.8)

This constraint is the one which guarantees that with time the PTs never have a decrease in the PDI level. There is always a degradation and never an improvement in the health status of the transformers.

• **6th constraint** - From one transition matrix to the next there should be a maximum of 2 new available PDI levels. This constraint needs to be implemented manually in each code for the computation of each transition matrix. This means that for example, in the first matrix a PT could degrade up to 5 PDI levels, the second matrix could only allow a PDI evolution up to the 7th PDI level and the third matrix allows degradation up to the 9th PDI level and so on until the 15th and
The final PDI level can be reached at the sixth transition matrix. This constraint had to be implemented in order to reduce the number of degrees of the problem and to make it solvable.

For the computation of each matrix, small adjustments had to be made in order to obtain a feasible solution. For the last matrices the problem began to have too many variables and specifications, which caused the need to loosen some of the constraints. In particular applying some of the constraints only to some particular rows of the the matrix, normally the higher age rows.

In addition to all of the applied constraints, there was the need to define an objective function which should be minimized during the computation. As there could be many solutions to the transition matrices which would fulfill the constraints this objective works as a criteria to guide to the type of desired solution. The defined objective is to maintain the similarities between matrices, together with the maximization of the probabilities of staying in a particular state from year to year. This approach assumes that the probabilities of degrading should be minimized and that with the evolution in time the degradation pattern should not change drastically.

\[
\text{obj} = \min \sum_{i} \sum_{j} (p_{ij}(n) - p_{ij}(n-1))^2 - p_{ii}(n) \tag{2.9}
\]

The optimization problem can then be summed up as:

minimize \[
\sum_{i} \sum_{j} (p_{ij}(n) - p_{ij}(n-1))^2 - p_{ii}(n)
\]

subject to \[
\sum_{j} p_{ij} = 1, \forall i,
\]
\[
[P_{PDI}(n-1) \times [TM]_n = [P_{PDI}(n)],
\]
\[
p(i, i) > p(i - 1, i - 1), \forall i,
\]
\[
p(i, j) > p(i, j + 1), \forall i, j,
\]
\[
p_{ij} = 0, \forall j < i,
\]
\[
(6^{th} \text{ constraint}).
\]

At the end, after computing all the matrices, it was applied a test to check if the PDI probabilities from each age range multiplied by the corresponding transition matrix originated the desired probabilities for the next age range. The results were then compared with the final probability matrix computed at Subsection 2.3.3, as is demonstrated by

\[
\forall n, \quad TM_{\text{error}} = [P_{PDI}(n-1) \times [TM]_n - [P_{PDI}(n)],
\]

where \( TM_{\text{error}} \) is the error between the initially desired value and the obtained through the transition matrix, \([P_{PDI}(n-1)]\) is the PDI probability distribution vector for a certain age interval \( n \), \( TM \) is the transition matrix associated with a particular age interval \( n \) and \([P_{PDI}(n)]\) is the value for the probability distribution vector obtained at Subsection 2.3.3. The average absolute error for the obtained transition matrices is around 0.02%. The sum of all the absolute errors for each element of the probability matrix
is 3%, which is an acceptable error, when taking into account all of the constraints and non linearity of the optimization problem.

### 2.5 Failure probability estimation based on the PDI

As explained before it is possible to obtain the failure index of each PT from the PDI level as is exemplified in Figure 2.12. All the intermediate indexes between the FI and the PDI were explained in the introduction chapter, together with the way they are defined.

Besides the already presented indexes, there is an extra index defined by EDPD named the criticality index. This index evaluates the impact of a PT failure multiplied by the associated impacts. In this work this index is not taken into account because the impacts defined by EDPD in that index take into account many variables that are not in the scope of what is pretended to be analyzed in this thesis, which is the ENS due to failure. As an example, one of the major impacts that the criticality index takes into account is the environmental impact of an oil spill if there is a failure in a transformer. In this work this type of problems are not taken into account.

![Figure 2.12: PDI to FI conversion schema](image)

Once the failure indexes of the transformers are computed based on the PDI levels generated through the aging model defined in Section 2.4, it is necessary to obtain a failure probability to each PT in order to simulate the failures in the transformers stock and the correspondent ENS.

Initially, based on the data from an experimental research project developed by EDPD, which proposed some values for the failure probability of each PT, a graph of the failure probability against the failure index was plotted. The goal of this plot was to to find a relationship between those two attributes that could lead to a function fitting. The result is presented in Figure 2.13 and shows that, even though there is a tendency of the failure probability to increase with the failure index, there is not a clear fit between both. The best fit is an exponential function which only has an R-Squared value of 0.04, which is not satisfactory in this context.

The second approach was to plot, based on the data from the historic average failure rate of the PTs per county in the years of 2012 to 2015, a graph of those failures against the corresponding failure index. The objective with this graph was to find a relationship between those two attributes. Although the results from this 2nd approach were better than the first one, they were not satisfying enough either. One of the reasons for this results can be the lack of data related with the failure rates of each individual PT, the available data is an average value per county instead of the data from each PT. Besides that, the
available data only regards a three-year interval. This amount of data is not enough, specially because
the amount of failures is very small, which makes the evaluation with a small amount of data particularly
difficult for this attributes.

As a 3rd alternative it is proposed a more simple method. It normalizes the average value of the
fail probability to the average value of the failure index. The idea is to assume that there is a linear
relationship between the average value of the FI and the historic average failure rate, as expressed by:

$$Pf_{(\text{per 5 year})} = FI \times \frac{FR_{(\text{fail PT \text{\_year}) \times 5}}}{FI},$$ (2.11)

where $Pf$ is the failure probability of each PT in a 5 year period, $FI$ is the failure index of each
PT, $FR$ is the yearly average failure rate of the installed PTs and $FI$ is the average failure index of the
installed PTs. Notice that in the developed model the analysis of the PDI, PHI, GHI and FI evolution is
made for five-year intervals, meaning that the failure probability should also be computed for five-year
intervals.

The values used as the average failure rate and average failure index are presented in Table 2.12
and were obtained from EDPD historic data (2012-2015). Although this approach might not be the most
complete one, it is an hypothesis that allows to continue the work in order to model the evolution of the
stock global health condition and ENS. A future improvement in the modeling of this relationship would
increase on a significant way the reliability of the global model.

Table 2.12: Installed PTs stock average failure rate and average failure index

<table>
<thead>
<tr>
<th>Stock average failure rate (fails/year/PT)</th>
<th>Stock average failure index</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0051</td>
<td>0.2824</td>
</tr>
</tbody>
</table>
Failure types

Once the failure probability is obtained it is possible to simulate, based on a random number generator, if a PT will fail in the next time interval. The idea is to generate a number between 0 and 1, if it is below the probability of failure it means that the PT will break down, otherwise it means that it will continue working.

According with EDPD, there are two different types of possible failures:

- **Critical Failure**: It is considered a critical failure every time the transformer cannot be repaired. If after the failure the PT becomes waste then it is inserted in this type of failure. Based on the experience and historic data every failure that affect a PT with over 50 years of age is considered to be a critical failure.

- **Non Critical Failure**: This type of failures correspond to half of the malfunctions that affect PTs under 50 years of age. Although this failure puts the transformer out of order, it can be repaired by the manufacturer, so it can become available to be re-installed in the grid again. This factory repairing of a PT has a cost of around 120K Euros. When a PT is re-installed in the grid, it will have nearly the same health indexes it had before the failure.
Chapter 3

Stock Health Simulation and Reliability Evaluation

In the previous chapter a model was developed to characterize the evolution of the PTs health condition along its lifetime. It was also presented a method to estimate the PTs associated probability of failure. In this third chapter it is presented an algorithm to estimate, based on a simulation method, the evolution of the average state of the global installed stock of PTs. This proposed simulator allows to extract information about several different parameters, such as the average age of the PTs stock or the average total ENS for the system. In order to do this simulation, it is applied the Monte Carlo Method, which, as is explained in the introduction chapter, relies on repeating random sampling in order to obtain a numeric value for a stochastic problem. This way it is possible to use randomness to solve problems that might be deterministic in principle. Initially in this chapter it is explained how the model for the global stock condition simulator works for one simulation. In the end of the chapter it is explained how the Monte Carlo method is applied.

3.1 ENS Estimation due to Power Transformers Failures

It is important to establish a method to compute the total amount of ENS due to failures in the PTs (HV/MV and MV/MV). The ENS value is one of the most important values that can be extracted as an output from the defined model. The ENS is computed based on the failure power, the stoppage time and the rate of failure associated with each transformer.

Whenever there is a failure in a PT, it takes time to reconfigure the grid schema in order to redistribute as much as possible the power load that was associated with the damaged transformer. After this first task is completed, the transformer replacement process begins. Usually, the time associated with the redistribution of the load from the damaged PT to others, that can handle the increase in their loads, is called reconfiguration time and in average corresponds to 15 minutes. The replacement of a failed transformer by a new one, either it is a backup PT or a mobile station, takes around 24 hours, according to EDP [41]. The two time values aforementioned are the ones considered for simulation purposes.

The total amount of energy not supplied associated with a damaged PT will depend on how much energy can be redistributed to others PTs and how much of it is in fact non distributed during the stoppage
time of the transformer. All of the data used for this computation was collected from the work delivered
to the regulator (ERSE - Entidade Reguladora dos Serviços Energéticos) within the scope of the 2015,
2016 and 2017 RARI (Regulamento de Acesso às Redes e Interligações do Setor Elétrico). This data
can be consulted in the Appendix A.

The data regarding the failure rates of the PTs was provided by EDP Distribuição [41]. The document
presented the failures rates discriminated for each one of the 278 Portuguese continental counties. This
leads to the necessity of crossing the information from the addresses of each transformer with the county
they are installed, in order to obtain the average failure rate associated with each PT. This method is
not the most precise, since it uses the same average failure rate of each county for all the PTs inside a
county, which might not correspond to the real failure rate of each PT. This is, however, the most exact
way to obtain an estimation for the average ENS given the data that was provided by EDP Distribuição.

A worst case scenario approach should be followed for the estimation of the ENS values per PT.
Instead of the average power carried out by each transformer, the value used for the estimation of the
ENS when there is a PT failure is the average peak power throughout the years of 2015, 2016 and
2017. There are two main parameters related with the load that have to be taken into account during
this computation:

- **Natural Load**: corresponds to the average peak power that a PT supplies and that is non delivered
during the period of reconfiguration, while the grid is being rearranged. Usually this period lasts
15 minutes, as was aforementioned.

- **Shut Down Load**: corresponds to the load that will be turned off until the PT is repaired or re-
placed. This load can vary from 0% to 100% of the natural load depending on how much of the
load can be redirected to others PTs. It is assumed that in average this load not supplied for a
period of 24 hours.

Based in the data previously obtained it is possible to obtain the average ENS for each PT due to
failures by multiplying the ENS power associated with each transformer by the failure rate per year and
by the time of interruption. This computation can be expressed by

\[
ENS = fr \times ((RT \times NL) + (RP \times SDL)),
\]

(3.1)

where \( ENS \) is the energy not supplied, \( fr \) is the failure rate, \( RT \) is the reconfiguration time, \( NL \) is
the natural load, \( RP \) is the replacement period and \( SDL \) is the shut down load.

The results for the ENS values associated with a failure in each of the 739 installed PTs can be
consulted in the Appendix A. Note that some of the values are negative, which only occurs in the PTs
that are connected to producers. The negative sign is for a matter of power flow direction convention.
To obtain the an estimation for the average total ENS value its necessary to apply the Equation 3.1 to all
the power transformers and sum the absolute value from all of them. It is important to sum the absolute
value in order to correct the negative sign that some PTs have for the ENS. This process is represented by:
\[ ENS_{\text{total}} = \sum_{n=1}^{739} |ENS_n|, \]  

(3.2)

where \(|ENS_n|\) is the absolute value of energy not supplied from each PT and \(n\) is the PT index. The estimation obtained for the current energy not supplied due to power transformers failure is 242.8 MVAh per year.

### 3.2 Global Stock Simulator Algorithm

#### 3.2.1 Failure Simulation

In Section 2.5 from the previous chapter it is defined how the failure probabilities for each PT are generated based on the failure index that is obtained through the simulation of the PDI evolution with time. The main goal of the developed model is to estimate the evolution of the degradation indexes of each PT for five-year intervals. This PDI level is directly converted into a PHI and then in a GHI, which then originates a failure index (FI). Finally this index is converted through a normalization function into a failure probability for each PT and in every time interval of the simulation horizon, as is explained by Figures 2.12 and 2.11.

For every PT installed in the grid, it is generated, in each time interval simulated, a random number using the \(\text{Rand}\) function from Matlab. This function outputs a random number between zero and one. If this number is smaller than the failure probability associated with a PT in a specific time interval, it is considered that the PT has failed in that time period. This failure is registered and then, depending on if is a critical failure or not, a new PT can be installed as a replacement for the old one. When a new transformer is installed, it is considered to have, in the beginning of the next time interval of the simulation, zero years of age and a PDI level equal to one.

This process is applied to all the 738 PTs installed in the grid for a time horizon defined by the user. In this thesis it will be simulated a thirty five years interval, which corresponds to eight five-year intervals. The initial state is also considered an interval, since it represents the current state of the transformers stock. The failures simulation of the PTs is the base of this global stock simulator model, since it will be the factor that will define the ENS of the system.

#### 3.2.2 Outputs of the Simulator

The estimation of the PTs health condition degradation and consequent failure prediction is incorporated in the simulator in order to be possible to extract a large number of different outputs in each time period. A descriptive explanation of each of the obtainable parameters is presented below:

- **Average age of the installed PTs stock** - Outputs the average age of the transformers installed in the grid.

- **Average PDI value** - This parameter measures the average degradation level of the installed PTs based on the evolution model defined in Chapter 2.
• **Average Failure Index** - Average value of the failure index for the installed PTs, which is computed based on their PDI value.

• **Average failure probability** - Average value of the failure probability generated through the simulation of their PDI and failure evolution.

• **Total amount of failures** - This parameter indicates the total count of transformer failures in each five year interval. These failures can be of different types:
  
  – **PTs failed due to age over limit** - Presents the count of total transformers that failed because exceeded the age of sixty five years, as was defined in Subsection 2.3.3 as the limit for a PT to continue working.

  – **Critical failures** - Number of PT failures that were critical failures, which means that lead to the installation of a new PT.

  – **Non Critical Failures** - Number of PT failures which were not critical failures, meaning that the PTs can be recovered and installed later in the grid again.

  – **Failures in PTs with high loss percentage** - This values presents the total amount of failures in PTs which led to a delivery capacity loss of over 75%. Although the majority of the grid is connect in loop to avoid the total loss of energy supply when there is a failure, some parts of the grid are radial. This measure counts the fails in PTs that are usually at the leaves of these radial grids and that have a high percentage of loss in the supply capacity to the clients whenever there is a failure in the transformers.

• **Type of failed PTs** - This value represents the amount of failures in each time interval, discretized per different types of transformers. EDPD defined ten different groups of transformers according with the transformation ratio and connection types.

• **New PTs installed** - This parameter presents the amount of new transformers installed in the grid.
  
  – **New PTs installed due to critical failure** - Amount of brand new transformers installed in the grid due to critical failures.

  – **New PTs installed due to renewal strategy** - Amount of new transformers installed due to the renewal strategy applied during the simulation.

• **Energy Not Supplied** - This parameter represents the total ENS per five-year interval due to the transformers failures. The algorithm for the computation of this parameter is explained in Section 3.1.

• **Energy Not Supplied in PTs with high loss percentage** - ENS value associated with the **Failures in PTs with high loss percentage**, which was explained above in this list.

• **Total Cost per Time Interval** - This cost value for five-year periods is obtained by the sum of three other costs:
– **Cost of the ENS** - Cost associated with the total ENS in a time period. EDPD considers that every kWh has a cost of three Euros.

– **Cost of the New PTs Installed** - Cost associated with the new PTs installed in the grid in a time period. Each new transformer has a cost of 540 thousand Euros.

– **Cost of the Recovered PT** - Cost associated with the PTs that have a non-critical failure and can be recovered. Each recovery of a transformer has a cost of 120 thousand Euros.

- **Total Strategy Cost** - This parameter measures the total cost of the implemented strategy along the simulation period. It consists of the sum of the *Total Cost per Time Interval* measure of all the simulated time intervals updated to the year zero through a discount rate that was defined by EDPD as 6.75%. The computation of this value is in a way similar to a NPV (Net Present Value) computation.

\[
TSC = \sum_{t=0}^{n} \frac{TC_t}{(1 + r)^t},
\]  

where \( r \) is the interest rate.

In Figure 3.1 it is represented in a simplified graphical form the algorithm for the computation of the output values of the simulator.

![Figure 3.1: Computation order of the simulation outputs](image)

3.3 **Monte Carlo Method Implementation**

The Monte Carlo simulation method is applied in the developed global stock simulator in order to reduce the error and increase the confidence associated with the simulation. As explained in the topic overview Section, the Monte Carlo simulation is used to obtain a deterministic result based on a stochastic model, which in this case is the one defined in Chapter 2. The developed aging model simulates, based on random numbers, the evolution of the PDI level for each individual PT. This PDI level will lead to a failure probability, which will then lead to a possible ENS. This method is based on making a large amount of trials on each execution of the simulation in order to obtain more feasible results. This happens because, according to the law of large numbers, the average of the empiric results obtained through the
large repetition of the simulation will make the average value obtained for the computed variables tend to its real expected value.

The global stock condition simulator presented in this chapter does not analyze each PT individually, it analyses the global stock of PTs as a whole. This means that, if only one trial was preformed the obtained results could be drastically different each time the model was run. Even though the amount of failures, for example, could be the same between executions, the ENS could be drastically different because the PTs that failed would probably not be the same. Consequently the ENS would not be the same, which would generate a massive empiric error in the obtained outputs. The variance between results would be rather significant if the amount of repetitions was small.

The Monte Carlo simulation method allows to minimize the empiric error by making a large amount of samples. As previously stated, with the increase in the number of trials for each execution of the program and for the same confidence interval, the error will have a tendency to decrease. It is also expected that the errors associated with the simulated variables generate a normal distribution around their mean values.

It is necessary to develop a criteria to determine how many samples are needed in order to guarantee a small error for a specific confidence interval. In this process the desired confidence interval is 95% which corresponds to a standard deviation of $\pm 1.96$ from the mean value. Assuming that the errors distributions is normal around the mean, it can be defined as

\[
\text{relative error} = 1.96 \pm \frac{\sigma}{\sqrt{n}} \approx 1.96 \pm \frac{s_y}{\sqrt{n}}
\]

where $\sigma$ is the standard deviation, $s_y$ is the unbiased estimator of the standard deviation and $n$ is the amount of trials.

There are two different approaches to the simulation extension that can be implemented in order to determine the stopping criteria. The first one is the implementation of a dynamic stopping criteria which would stop the simulation when the relative error of a certain variable is smaller than a specified value. Although this approach guarantees that the relative error is smaller than a desired value, it has the downside of making the execution of the program more complex. Matlab is less efficient when working with dynamic length vectors, which would be required to apply this type of criteria, than it is when operating fixed length vectors. In order to obtain a relative error of 1% with a confidence interval of 95% for the simulated PDI average values it would take between 4500 and 5000 repetitions of the simulation. This amount of trials would take between 8 and 12 minutes for each execution, which is a large execution time, taking into account the amount of different strategies and corrections that had to be performed for obtaining the results to be analyzed.

A different approach is a fixed number of simulations for every execution. After many different values, the data shows that for one thousand repetitions, the maximum relative error with a 95% interval of confidence for the average PDI level of the global stock is of 2.5% and for most of the cases is below 2%, which is a very acceptable error margin. The approach with the fixed length simulation has an average execution time between 1 and 2 minutes, which is considerable less than the dynamic stopping
criteria.

The increment in the number of simulations reduces the relative error as it is expected. The computational cost to have a dynamic stopping criteria summed with the variable size vectors manipulation that is necessary to apply, makes it unfeasible to choose this approach, at least on a personal computer. As an alternative, the applied method used the fixed number of simulations criteria with n=1000. For each execution, the maximum PDI relative error is registered and, if it is too high, the amount of simulations should be increased and then the program should be executed again. Although this is a more rudimentary empiric method, it satisfies all the needs in this thesis context.
Chapter 4

Stock Renewal Strategies: Comparison and Result Analysis

In this chapter the results for the simulation of different grid renewal strategies are presented. At the end of the chapter it is made a comparison between the different output parameters from each strategy, in order to understand which are the most beneficial to be implemented.

4.1 Global Stock Renewal Strategies

The developed global stock simulator, as explained in Section 3.2.2, allows to estimate the average PDI evolution for the installed stock of PTs. This tool makes it is possible to simulate different renewal strategies for the global stock and analyze its output parameters in order to compare the strategies. This can be a useful tool when planning the investment strategy for the distribution grids renewal.

There is an infinite number of different strategies that can be implemented. In the next chapter the results from the applications of a few of those strategies are presented. Initially it is presented and analyzed the so-called “zero strategy”. This consists in not investing in the substitution of PTs, except in the case of failure. The outputs of this strategy will provide an idea of how the grid will evolve along the years to come and what the consequences and costs will be like if there is no investment in the grid.

Next the comparison between different strategies will be presented. The strategies are defined at the beginning of each time interval, but for computation reasons the substitutions are only applied at the end of each time interval. If a PT is defined to be substituted in a time interval, it is considered to have a PDI level equal to one from the beginning of that interval.

The analyzed strategies for renewal are the following:

- **Substitution Over X Years** - This strategy applies a renewal based exclusively on the age of the PTs. All the transformers with an age over X years are replaced.

- **Top N PDI Level** - This strategy is based on replacing the N number of PTs with the highest PDI value in each time interval.
• **TOP N Possible ENS** - This strategy analyzes what PTs have a higher value of ENS in case of failure, multiplied by the failure probability. The N number of PTs with higher value are replaced.

• **RUL Based Substitution** - This strategy is based on an EDPD report called RUL (Remaining Useful Life), which forecasts a year of failure for each transformer. When a transformers exceeds the fault age, it should be replaced according with this strategy.

• **Hybrid Strategy** - This kind of strategy is based on the mixing of any of the strategies mentioned above. If applied, it should be specified what are the sub strategies applied and respective parameters.

Due to the maximum extension of the dissertation, it is not possible to analyze the output of all proposed strategies with a depth similar to what is performed to the "zero strategy". Nonetheless the comparisons made are descriptive enough to give an indication of how each strategy affect the outputs and which ones are better.

### 4.1.1 Zero Strategy Analysis

The first simulation made is the so-called "zero strategy" strategy, which consists in not renewing any of the installed stock of PTs. This strategy, like all the others evaluated, will be simulated for the next seven five-year intervals, meaning a 35 years horizon. This will be useful to understand how the stock might evolve if there is not investment in the current grid.

#### Average PDI and Age Evolution with Time

By not implementing a renewal strategy, it is expected that there will be a general degradation of the health condition of the PTs installed in the grid, with the time. In Figure 4.1 it is presented the evolution of the average PDI value for the installed PTs confined by a one standard deviation band. The standard deviation value is about 1% of the average value, which is a very acceptable value.

The results show that in the next ten years there will be a marked deterioration of the installed PTs. From 10 years onward the average PDI level decreases, since the already degraded transformers fail and most of them are replaced by new PTs, which, since are brand new, have a low PDI value. From 25 to 30 years from now there is a steep descent in the average PDI value, meaning that there is a large amount of PTs being replaced in that time interval.

In Figure 4.2 it is presented the average age of the park together with the corresponding standard deviation band. It is possible to see that there is a correspondence between the general behavior of the average age and the average PDI level, as was expected, because the PDI level is related with the age of each PT. In the interval between the 25 and 30 years from now, there is also a major decay in the age average value, which corresponds to the major decrease in the PDI value as explained in the previous paragraph. Note that the standard deviations of the average age values is also, at maximum 1% of the mean value, which is an acceptable value.

It can also be interesting to analyze the evolution of the failure index, which is computed based on the PDI level together with the age and the environmental factor associated with each PT. As seen in
Figure 4.1: Evolution with time of the average PDI value of the installed PTs

Figure 4.2: Evolution with time of the average age of the installed PTs
the third chapter the failure probability of each transformer is generated through a normalization process applied to the failure index.

In Figure 4.3 it is possible to see the evolution of the average failure index. Note that the failure index has a graph which is very similar to the PDI and age graphs present in Figures 4.1 and 4.2. This is because the failure index of each transformer is computed based mainly on those two values. The standard deviation of the FI is once again around 1% of the mean value of the failure index. Notice that this average standard deviation, as well as all the others analyzed above, is computed based on the one thousand Monte Carlo Simulation made by the global stock condition simulator, it is not the standard deviation of the values from each installed PT.

**Average Failure Probability and Failures Count Evolution with Time**

The values obtained for the average failure probability of the PTs are presented in Figure 4.5 and the graph is quite different from the ones obtained for the average PDI, age and FI presented in Figures 4.1, 4.2 and 4.3. The values for the average failure probabilities are higher on the time intervals where there is a steep decline in the average PDI value. This happens because, for implementation reasons, the PDI average value is computed at the end of the time interval, while the fail probability is computed at the beginning of the time interval. This approach is used because the information about the failure probability at the beginning of the interval allows to prevent some failures during that same interval, by applying a preventive renewal strategy. On the other hand, information about the failure probabilities after the transformers already failed in the interval is not useful.
A significant rise in the average failure probability will, most of the times, result in a drop in the average PDI level, age and failure index of the stock at the end of the same interval. As the graphs demonstrate, whenever there is a high rise in the failure probability, there is a drop in the others, already analyzed, output variables. This increase in the failure probability, as explained before, is related with the increasing amount of transformers that are getting close to the failure age and see their failure probability increase. On the other hand, after the substitution of PTs in the 25-30 years from now time interval, the average failure probability decreases. This shows that the impact of the number of substitutions due to failure on that specific time interval will be quite significant.

In Figure 4.5 the obtained results for the total number of PTs failures per time interval are presented. It is possible to see that the plot is very similar to the failure probabilities plot. This was expected, taking into account that the number of failures should be directly proportional to the average failure probability. Also as expected, there is a major increase in the number of failures in the 25-30 years time frame, which is in line with the conclusions retrieved from the graphs analyzed above. This major increase in the number of failures is due to the degradation of the grid general health condition, caused by the aging of the transformers.

Notice that the width of the band correspondent to one standard deviation is quite larger in the fail count graph than in the failure probabilities one. The failure probabilities standard deviation has an average value around 2.3% of the corresponding mean value, while for the failures count it corresponds to 11%. This is an expected result because the total failure count has more variability, since it adds a bigger uncertainty to the process. The fact that a transformer fails or not is a probabilistic event, which,
Figure 4.5: Evolution with time of the average amount of failures in the installed PTs

as explained in the previous chapter, is defined based on the generation of a random number. The average failure probability, is a variable that is obtained based on the developed predictive aging model, which has already an intrinsic variability and uncertainty. For the failure simulation of each PT, it is introduced, with the random number generator another “source” of variability which explains this increase in the relative standard deviation.

Part of the PTs that fail are replaced by new ones, while a smaller part can be repaired by the manufacturer. The division generated by the two different approaches is presented in Figure 4.6. As it was explained before, all of the transformers above 50 years of age that fail are replaced by new ones, while the ones under 50 years have a 50% chance of being intervened and then re-installed in the grid. The major difference between the amount of repaired PTs and installed PTs is due to the fact that most of the PTs that fail are above 50 years, meaning that they are never repaired and always give origin to the installation of a brand new PT.
Costs associated with the ENS

Each PT that fails interrupts the energy distribution, therefore it has an associated ENS value as was explained in the previous chapter. For each of the 1000 trials ran during a simulation, it is computed an average global ENS per five-year interval. The average value for the 1000 Monte Carlo samples is computed and an average ENS value is obtained, together with a standard deviation. The same process is applied to obtain all of the output results presented above.

The ENS values obtained for the zero strategy renewal policy implementation are presented in Figure 4.7 together with the correspondent monetary appreciation. This value in millions of Euros is obtained based in the value defined by EDPD for each kVA not supplied, which is 3€/kWh (to simplify it was considered a power factor of 1). It is possible to see that the graph is similar to the graph obtained for the failure probability and failures count, which makes sense, since the ENS should be proportional to the amount of failures in the PTs. The peak of failures in the time frame of 25-30 years from now interval can possibly have a very high cost due to the ENS. As can be seen, the difference from the actual cost to the cost in that peak is from 3/4 M€ to around 30 M €, which is an increase of 1000%.

ENS Analysis

The output values regarding the ENS presented in the above graphs are compiled in Table 4.1. The way of displaying the results present in this table will be particularly useful when comparing different strategies, applied to the same time interval, since it synthesizes the results for the whole implementation of a strategy along a particular time period. Although this way of presenting the data is less specific, it
will allow to do a direct comparison between strategies without recurring to all of the data discriminated by each time period.

Table 4.1: ENS results compilation for the 35 years of simulation

<table>
<thead>
<tr>
<th>ENS (TVAh)</th>
<th>Yearly ENS (MVAh/year)</th>
<th>Number of failures (#)</th>
<th>Non redundant failures (#)</th>
<th>ENS due to non redundant failures (#)</th>
<th>New PTs (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28.9</td>
<td>826</td>
<td>503</td>
<td>43</td>
<td>5771</td>
<td>462</td>
</tr>
</tbody>
</table>

**Cost Analysis**

The total cost of this strategy, as defined in Section 3.3, can be computed through an interest rate applied to the the cost of the ENS appreciation, the investment in new PTs and the investment in the recovery of some others PTs in order to update those monetary values to the year zero. The results for the total strategy cost for the 35 year simulation and respective partial values are presented in Table 4.2. The interest rate value considered was 6.75% and was defined by EDPD.

Note that the values presented in this table, together with the values from Table 4.1, will be the ones used to compare strategies. This values compile some useful information, which allows to compare strategies outputs, but only if the comparison is done over the same time horizon.
Table 4.2: Zero Strategy cost analysis

<table>
<thead>
<tr>
<th>Global cost (M€)</th>
<th>ENS Cost (M€)</th>
<th>ENS Cost (%)</th>
<th>New PTs Investment (M€)</th>
<th>New PTs Investment (%)</th>
<th>Repaired PTs Cost (M€)</th>
<th>Repaired PTs Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>77.9</td>
<td>20.2</td>
<td>25.9%</td>
<td>55.9</td>
<td>71.7%</td>
<td>1.9</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

Other type of analysis

There are other analysis that can be performed based on the obtainable parameters of the simulation. For example, in Table 4.3 it is presented a discrimination of the average number of failures in each time period by PT type. The types of transformers were defined by EDPD and represent PTs with different transformation ratios, maximum power transformation limitations and connection typologies. This can be a useful information when planning and sizing the reserve PTs stock, in order to guarantee that there is enough backup transformers in stock to react to the estimated failures.

Table 4.3: Average number of failures per PT type in each time interval with the “Zero Strategy” renewal strategy

<table>
<thead>
<tr>
<th>Age Interval</th>
<th>Type of PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Interval</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>0</td>
<td>5.0 3.5 2.3 2.1 1.4 0.7 0.6 0.5 0.3 2.0</td>
</tr>
<tr>
<td>1</td>
<td>5.6 3.8 2.5 3.2 1.6 0.8 0.7 0.5 0.3 2.2</td>
</tr>
<tr>
<td>2</td>
<td>5.9 9.8 3.5 2.5 1.6 0.8 0.8 0.6 0.3 9.7</td>
</tr>
<tr>
<td>3</td>
<td>15.1 11.4 8.1 5.3 5.3 0.9 0.9 2.4 0.4 3.9</td>
</tr>
<tr>
<td>4</td>
<td>9.6 8.4 7.6 12.9 4.2 0.9 1.7 0.7 2.0 14.1</td>
</tr>
<tr>
<td>5</td>
<td>19.1 20.1 17.8 8.1 7.6 1.0 0.9 0.7 0.3 11.2</td>
</tr>
<tr>
<td>6</td>
<td>34.5 43.8 12.0 27.2 16.6 1.1 6.0 1.6 2.0 20.9</td>
</tr>
<tr>
<td>7</td>
<td>19.6 16.0 12.6 7.4 2.0 1.2 2.6 6.4 3.6 6.3</td>
</tr>
</tbody>
</table>

Other interesting analysis that can be performed is the percentage of the ENS that correspond to PTs located in terminal branches. This is the same as saying transformers where, at least 75% of the clients fed by those substations lose the energy supply and the grid cannot be reconfigured in order to guarantee their supply. This analysis was made for the zero strategy discretized by time interval and is shown in Table 4.4, where it is possible to see the ENS value in terminal branches for the seven time intervals simulated, together with the amount fails and the monetary valuation of the ENS in those branches. It is also presented the weight of these values when compared to the total ENS, total amount of failures in PTs and total ENS cost valuation for the whole system for the same time intervals.

Table 4.4: Analysis of the ENS and failures in non redundant PTs

<table>
<thead>
<tr>
<th>Time interval</th>
<th>0 1 2 3 4 5 6 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENS (MWh/(5 years))</td>
<td>208 238 542 382 657 1333 1459 1166</td>
</tr>
<tr>
<td>Percentage of the total ENS</td>
<td>17.6% 17.9% 27.6% 12.9% 21.2% 31.4% 15.0% 21.2%</td>
</tr>
<tr>
<td>Amount of failures</td>
<td>7.4% 1.5% 3.5% 2.5% 8.4% 10.0% 12.2% 5.2%</td>
</tr>
<tr>
<td>Percentage of the total failures count</td>
<td>7.4% 7.1% 10.0% 4.7% 13.5% 11.5% 7.4% 6.7%</td>
</tr>
<tr>
<td>Cost of the ENS (M€)</td>
<td>0.8 0.7 1.6 1.1 2.0 4.0 4.4 3.5</td>
</tr>
<tr>
<td>Percentage of the total ENS Cost</td>
<td>7.4% 7.1% 10.0% 4.7% 13.5% 11.5% 7.4% 6.7%</td>
</tr>
</tbody>
</table>
Normal distribution around the mean output values

As mentioned in the topic overview section, the use of the MCMC method should result in a normal distribution for the values obtained for each simulated output, when taken into account the results obtained through the one thousand simulations ran. The results match this condition, and as an example it is shown below the evolution of the number of failures per time interval, and it is possible to see that there is a normal distribution around the mean values, which were already presented in the above subsections.

![Histogram of failures distribution for different time intervals](image)

Figure 4.8: Monte Carlo simulation failures distribution for different time intervals

On Figure 4.8 it is possible to see the expected effect. Although, the mean value of the failures is different according with the analyzed time interval, both of the results show a normal distribution around the mean value. This is relevant because it confirms the randomness of the simulation process. The same pattern can be seen in all of the results obtained from the simulator, independently of the strategy applied or the time interval analyzed.
4.1.2 Age based renewal strategy

One of the proposed renewal strategies is the so called "over X renewal strategy". This consists in the replacement of all of the PTs with an age over a specified "X" value, independently of their health condition. This can be a very conservative strategy, specially with low renewal ages because, as seen before, the PTs can maintain a low level of degradation up until an advanced age. Besides, the correlation between age and failures is not perfectly defined.

This strategy was tested with three different renewal ages: 40, 50 and 60 years. The main results are presented on Figure 4.9 and are all compiled in a ENS analysis and cost analysis in Tables 4.5 and 4.6.

From the graphs obtained it is possible to conclude that, with the decrease in the renewal age of the strategy, there is a major improvement in the ENS value together with a decrease in the average PDI value, average age and total amount of failures in the installed PTs. On the other hand, there is a major increase in the amount of new PTs installed, which has a great impact on the total cost of the strategy. From the ENS graph it is possible to observe that, with the decrease in the renewal age, there is an anticipation of the peak of substitutions in comparison with the zero strategy results. This happens because the same PTs, that were going to be replaced further along are being replaced earlier. The maximum value for the amount of replacements also decrease with the age of renewal, this is due to the fact that the amount of failures decreases, meaning that those transformers that do not get to fail, are replaced earlier. Therefore, the sum of substitutions by failure plus substitution by strategy decreases.

In the cost analysis it is possible to see that the total cost of the strategy implementation increases significantly with the decrease of the renewal age and that most of the cost is associated with the purchase of new PTs for the grid, while the savings associated with the ENS appreciation are not significant enough to justify the strategies with a much lower renewal age.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>ENS (TVAh)</th>
<th>Yearly ENS (MVAh/year)</th>
<th>Number of failures (#)</th>
<th>Non redundant failures (#)</th>
<th>ENS due to non redundant failures (TVAh)</th>
<th>New PTs (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over 40</td>
<td>6.41</td>
<td>183.00</td>
<td>96</td>
<td>6.5</td>
<td>1.23</td>
<td>687</td>
</tr>
<tr>
<td>Over 50</td>
<td>7.58</td>
<td>219.45</td>
<td>116</td>
<td>7.6</td>
<td>1.40</td>
<td>567</td>
</tr>
<tr>
<td>Over 60</td>
<td>8.83</td>
<td>252.37</td>
<td>135</td>
<td>9.0</td>
<td>1.62</td>
<td>462</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Global Cost (M€)</th>
<th>ENS Cost (M€)</th>
<th>ENS Cost (%)</th>
<th>New PTs Investment (M€)</th>
<th>New PTs Investment (%)</th>
<th>Repaired PTs Cost (M€)</th>
<th>Repaired PTs Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over 40</td>
<td>167.5</td>
<td>6.6</td>
<td>3.93%</td>
<td>159.0</td>
<td>94.90%</td>
<td>1.9</td>
<td>1.16%</td>
</tr>
<tr>
<td>Over 50</td>
<td>107.5</td>
<td>8.1</td>
<td>7.54%</td>
<td>97.3</td>
<td>90.50%</td>
<td>2.1</td>
<td>1.95%</td>
</tr>
<tr>
<td>Over 60</td>
<td>66.8</td>
<td>9.0</td>
<td>13.49%</td>
<td>55.9</td>
<td>83.68%</td>
<td>1.9</td>
<td>2.84%</td>
</tr>
</tbody>
</table>
Figure 4.9: Main results regarding the simulation of the “over X” renewal strategies for different X age values
4.1.3 Top PDI and Top ENS renewal strategies

The two renewal strategies addressed in this section are very similar in the way that are "Top values" based substitutions. The TOP X PDI level strategy replaces the X number of transformers with the worst PDI level in each time interval. On the other hand, the TOP X Possible ENS strategy takes into account the ENS associated with the failure of each PT multiplied by the probability of failure in the beginning of each time period. The graphical results obtained through the simulation of these two strategies for 3 different X values are presented in Figures 4.10 and 4.11.

For both strategies the results show that there is a decrease in the average PDI value, age, and failures, together with an increase in the number of planned replacements per time period, which is what was expected to occur. Also, as expected, there is a major peak of new PTs installed in the grid in the time interval between 25 and 30 years from now.

It is interesting to observe that with the Top X PDI strategy, the peak of installed PTs is higher for the X equal to 20, than when it is equal to 40 and 60. This means that the applied replacements for the previous intervals can atone the needs of the grid in that time interval, due to the aging of the installed PTs. On the other hand, with the TOP X Possible ENS strategy this does not verify. Since, even though the amount of planned substitutions defined by the strategy for each time interval is the same, the PTs replaced are not necessarily the ones with the worst health state, which on most case coincide with the older ones. This will lead to a major need for renewal due to age failure in the 25-30 years interval.

Another conclusion that can be withdrawn from the results, and that is presented in Figure 4.12, is that with the Top X PDI strategy the number of failures is considerable smaller than with the TOP X Possible ENS strategy, considering the same X. On the other hand, the ENS and respective associated cost is higher, which meets the expectations. It is logical that the PDI based renewal strategy will have a bigger impact on the failures, but not necessarily on the total ENS associated with those failures.

In Tables 4.7 and 4.8 is presented a summary of the ENS and cost related results along the whole simulation horizon (35 years). From there, it is possible to draw multiple conclusions, besides the ones already presented. First, it is possible to conclude that the Top Possible ENS strategies when compared with the equivalent Top PDI strategies have a lower value (around half) of ENS in terminal branches. Although the amount of failures is higher for the TOP Possible ENS, the difference is not of the same proportion. For the same comparison it is also possible to conclude that the number of new installed PTs is higher on the Top ENS strategies. This is associated with the higher number of total failures in this type of renewal strategy.

On the cost analysis comparison it is also possible to draw some interesting conclusions. First it is simple to see that, although, the difference in the total cost associated with each strategy is higher for the Top Possible ENS strategies, the difference is not significant. Besides, it is possible to conclude that the weight of the investment in new PTs in the global cost of the strategies is higher in the Top Possible ENS type of strategies. Although, while the weight of the ENS appreciation is higher for the Top PDI type of strategies, it is not a significant difference. This is related with the higher amount of failures in the ENS based strategy and the higher ENS value for the PDI based strategy.
Figure 4.10: Main results regarding the simulation of the "Top X PDI level" renewal strategies for different X values
Figure 4.11: Main results regarding the simulation of the "Top X Possible ENS level" renewal strategies for different X values

(a) Average PDI evolution with time

(b) Average PTs age evolution with time

(c) Evolution of the average amount of failures with time

(d) Average number of new PTs installed per each time interval

(e) Average ENS value and respective monetary appreciation per each time interval
Table 4.7: Top PDI and top ENS strategies ENS analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>ENS  (TVAh)</th>
<th>Yearly ENS (MVAh/year)</th>
<th>Number of failures (#)</th>
<th>Non redundant failures (#)</th>
<th>ENS due to non redundant failures (TVAh)</th>
<th>New PTs (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP 20 PDI</td>
<td>23.01</td>
<td>657.40</td>
<td>408</td>
<td>37.2</td>
<td>4809.50</td>
<td>506</td>
</tr>
<tr>
<td>TOP 20 ENS</td>
<td>15.92</td>
<td>454.89</td>
<td>450</td>
<td>33.0</td>
<td>2621.40</td>
<td>550</td>
</tr>
<tr>
<td>TOP 40 PDI</td>
<td>17.94</td>
<td>512.49</td>
<td>321</td>
<td>30.6</td>
<td>3902.20</td>
<td>557</td>
</tr>
<tr>
<td>TOP 40 ENS</td>
<td>10.99</td>
<td>313.89</td>
<td>414</td>
<td>26.8</td>
<td>1792.80</td>
<td>653</td>
</tr>
<tr>
<td>TOP 60 PDI</td>
<td>13.40</td>
<td>382.74</td>
<td>237</td>
<td>22.3</td>
<td>2941.80</td>
<td>612</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>8.10</td>
<td>231.46</td>
<td>383</td>
<td>23.0</td>
<td>1446.10</td>
<td>762</td>
</tr>
</tbody>
</table>

Table 4.8: Top PDI and Top ENS strategies cost analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Global Cost (M€)</th>
<th>ENS Cost (M€)</th>
<th>ENS Cost (%)</th>
<th>New PTs Investment (M€)</th>
<th>New PTs Investment (%)</th>
<th>Repaired PTs Cost (M€)</th>
<th>Repaired PTs Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOP 20 PDI</td>
<td>89.8</td>
<td>16.9</td>
<td>18.85%</td>
<td>71.1</td>
<td>79.10%</td>
<td>1.8</td>
<td>2.05%</td>
</tr>
<tr>
<td>TOP 20 ENS</td>
<td>91.0</td>
<td>13.1</td>
<td>14.40%</td>
<td>76.1</td>
<td>83.62%</td>
<td>1.8</td>
<td>1.98%</td>
</tr>
<tr>
<td>TOP 40 PDI</td>
<td>103.6</td>
<td>14.2</td>
<td>13.75%</td>
<td>87.5</td>
<td>84.47%</td>
<td>1.8</td>
<td>1.78%</td>
</tr>
<tr>
<td>TOP 40 ENS</td>
<td>108.6</td>
<td>9.7</td>
<td>8.90%</td>
<td>97.2</td>
<td>89.48%</td>
<td>1.8</td>
<td>1.52%</td>
</tr>
<tr>
<td>TOP 60 PDI</td>
<td>117.4</td>
<td>11.7</td>
<td>9.98%</td>
<td>103.9</td>
<td>88.46%</td>
<td>1.8</td>
<td>1.56%</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>128.2</td>
<td>7.7</td>
<td>5.98%</td>
<td>118.8</td>
<td>92.68%</td>
<td>1.7</td>
<td>1.34%</td>
</tr>
</tbody>
</table>

Figure 4.12: Comparison between equivalent "Top X PDI level" and "Top X Possible ENS level" renewal strategies for different X equal to 20 and 60
4.1.4 RUL Based Renewal Strategy

EDP has developed a study which forecasts the remaining useful life (RUL) for each of the installed PTs. Based on this info it is possible to develop a renewal strategy which substitutes the PTs, right before they fail, preventing the failure before it happens.

This type of strategy revealed to be very conservative and not very efficient, since it substitutes a large amount of transformers, while not having concordant gains. In Figure 4.13 the results for this strategy are presented, together with an example of the "Over X" and "Top Possible ENS" strategies for comparison reasons.

From the results it is possible to see that the RUL based strategy have a different evolution of the outputs parameters, from the ones obtained with the others strategies already analyzed. For the average PDI level of the installed PTs, it results on an abrupt improvement in the health condition indexes of the transformers in the first four time intervals. Meaning that there are a large amount of transformers with an useful life ending in the next 20 years. After this four times intervals, the strategy cannot maintain the PDI level and there is an increase in its average value. The average age of the installed stock has a behavior very similar to the PDI’s.

For the evolution of the total failure count, the result of the RUL based strategy is very similar to the "over 50 years" strategy, which has a very low amount of failures in the installed PTs. This makes sense, since the idea behind this strategy is to replace the transformers before they fail. In a way this strategy is similar in concept to the "over X" type of strategy, although using a different criteria.

For the installed PTs there is a major peak of new PTs in the grid associated with the RUL based strategy, which is higher than the peaks of the other two strategies. This strategy anticipates the peak in time by 15 years when compared with the "Top 60 possible ENS" strategy, for example. On the other hand, after this initial peak associated with the ending of the useful life of a large amount of PTs in the next 15 years, there is a low amount of installed PTs in the years after.

For the ENS estimated values, it is possible to see that the profiles of the analyzed strategies are truly different from each other. The profiles are similar to the inverse of the new PTs installed stock profiles already analyzed. On the first years the ENS levels increase slightly, while after the major amount of replacements, that occurs around 20 years from now, there is a drop in the ENS value, which makes sense, since there is a major amount of transformers replacements. When compared with the other strategies, the RUL based guarantees a more smooth profile than the "Top 60 possible ENS" strategy but is less robust than the "over 50 years" strategy.

On Table 4.9 the summary of the ENS general results for the time horizon of the simulation is presented. It is possible to conclude that after 35 years, the differences between the three compared strategies in terms of ENS is not significant. On the other hand, the total amount of failures is substantially higher on the "top 60 ENS" strategy than on the other two strategies. Analyzing the amount of failures in terminal branches, it is possible to see that the RUL based strategy has around 60% of the failures associated with the "Top 60 ENS" strategy and around the double of the failures associated with the "over 60 years" strategy. Even though it has less failures in terminal branches, the total amount of
ENS associated with those branches is higher with the RUL based strategy.

The total cost analysis associated with the strategies is presented on Table 4.10 and allows to conclude that the RUL based strategy is the most expensive one, mainly due to the high number of new PTs purchased in earlier years than on the others strategies. This is why, even though the amount of new installed PTs is higher on the "top 60 ENS" strategy, as they are bought later in time, the interest rate makes them less expensive when converted to the today currency value.

In summary the results allow to conclude that the RUL strategy is not the most efficient one and is not worthy of the financial cost, when compared with the ENS gains that come from the investment, specially when compared, for example, with the "over 50 years" strategy, which is less costly and has a similar ENS results.

Table 4.9: RUL based strategy ENS analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>ENS (TVAh)</th>
<th>Yearly ENS (MVAh/year)</th>
<th>Number of failures (#)</th>
<th>Non redundant failures (#)</th>
<th>ENS due to non redundant failures (TVAh)</th>
<th>New PTs (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUL</td>
<td>7.41</td>
<td>211.80</td>
<td>123</td>
<td>13.3</td>
<td>1.82</td>
<td>664</td>
</tr>
<tr>
<td>Over 50</td>
<td>7.68</td>
<td>219.45</td>
<td>116</td>
<td>7.6</td>
<td>1.40</td>
<td>567</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>8.10</td>
<td>231.46</td>
<td>383</td>
<td>23.0</td>
<td>1.45</td>
<td>762</td>
</tr>
</tbody>
</table>

Table 4.10: RUL based strategy cost analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Global Cost (M€)</th>
<th>ENS Cost (M€)</th>
<th>ENS Cost (%)</th>
<th>New PTs Investment (M€)</th>
<th>New PTs Investment (%)</th>
<th>Repaired PTs Cost (M€)</th>
<th>Repaired PTs Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUL</td>
<td>130.2</td>
<td>7.6</td>
<td>5.87%</td>
<td>120.5</td>
<td>92.54%</td>
<td>2.1</td>
<td>1.59%</td>
</tr>
<tr>
<td>Over 50</td>
<td>107.5</td>
<td>8.1</td>
<td>7.54%</td>
<td>97.3</td>
<td>90.50%</td>
<td>2.1</td>
<td>1.95%</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>128.2</td>
<td>7.7</td>
<td>5.98%</td>
<td>118.8</td>
<td>92.68%</td>
<td>1.7</td>
<td>1.34%</td>
</tr>
</tbody>
</table>
Figure 4.13: Main results regarding the simulation of the RUL based renewal strategies in comparison with the "Top 60 possible ENS" and the "over 50 years" renewal strategies.
4.1.5 Hybrid Renewal Strategy

It is possible to use the developed simulator to test hybrid renewal strategies for the stock of installed PTs. This type of strategies can be based in any of the above mentioned criteria or in any others that can be defined in the future. As an example, it is presented an hybrid strategy based on the "Over X" strategy mixed with the "Top possible ENS" strategy. The examples presented show the combination of the "over 55 years" together with the "top 20 possible ENS" which will be called hybrid1 strategy. A different strategy presented is the so called hybrid2, which corresponds to the "over 60 years" strategy together with "the top 40 possible ENS".

In the defined hybrid strategies the number of PTs replaced due to the "top X possible ENS" criteria will be the difference between the ones replaced by the "over X" criteria and the number defined for the "top possible ENS". For example for hybrid1 strategy, if in a time interval 15 PTs are replaced due to the "over 55 years" criteria, only the top 5 ENS will be also replaced. On the other hand if for a different time interval zero PTs are replaced due to the "over 55 years" criteria, there will be 20 PTs replaced due to the "top 20 ENS" criteria.

From the graph presented in Figure 4.14, it is possible to see that, in terms of the global installed stock, there is a major improvement in the parameters related with the average health of the installed PTs, namely the average PDI value, age and number of failures per time interval, when compared with the "over 60 years" strategy. On the other hand, there is a bigger overhead related with the replacement of transformers, which has a similar profile to the one from the "top 60 possible ENS".

On Tables 4.11 and 4.12 is presented the summary of the ENS and cost analysis of the global results along the time frame of the simulation. From there, it is possible to see that in terms of ENS, the results from the hybrid strategies are better than the ones from the "over 60 years" and the "Top 60 ENS" strategies. On the cost analysis it is clear that with the hybrid strategies the costs are lower than with the "top 60 ENS" strategy but higher than for the "over 60 years" one. Notice that the value of failures in terminal branches is slightly lower for the hybrid strategies but the ENS in those branches is significantly lower, which can be a significant criteria when evaluating the strategies to implement in the grid. Another interesting result is the weight that the ENS appreciation has in the cost of each strategy. Even though the general cost of the "over 60 years" strategy is significantly lower than others strategies, the weight of the ENS in the global cost is around the double than for the others analyzed strategies. This is an indicator of how the hybrid strategies, are more efficient in improving the ENS than the "over X years strategy".

The results obtained for these hybrid strategies prove that the most efficient strategies will be probably obtained through a strategy developed based in multiple criteria and parameters.
Table 4.11: Hybrid type strategy ENS analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>ENS (TVAh)</th>
<th>Yearly ENS (MVAh/year)</th>
<th>Number of failures (#)</th>
<th>Non redundant failures (#)</th>
<th>ENS due to non redundant failures (TVAh)</th>
<th>New PTs (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid1</td>
<td>6.27</td>
<td>179.21</td>
<td>119</td>
<td>7.0</td>
<td>1.02</td>
<td>596</td>
</tr>
<tr>
<td>Hybrid2</td>
<td>5.52</td>
<td>157.75</td>
<td>122</td>
<td>7.0</td>
<td>0.91</td>
<td>661</td>
</tr>
<tr>
<td>Over 60</td>
<td>8.83</td>
<td>252.37</td>
<td>135</td>
<td>9.0</td>
<td>1.62</td>
<td>462</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>8.10</td>
<td>231.46</td>
<td>383</td>
<td>23.0</td>
<td>1.45</td>
<td>762</td>
</tr>
</tbody>
</table>

Table 4.12: Hybrid type strategy cost analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Global Cost (M€)</th>
<th>ENS Cost (M€)</th>
<th>ENS Cost (%)</th>
<th>New PTs Investment (M€)</th>
<th>New PTs Investment (%)</th>
<th>Repaired PTs Cost (M€)</th>
<th>Repaired PTs Cost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid1</td>
<td>102.1</td>
<td>6.8</td>
<td>6.69%</td>
<td>93.3</td>
<td>91.42%</td>
<td>1.9</td>
<td>1.89%</td>
</tr>
<tr>
<td>Hybrid2</td>
<td>106.1</td>
<td>6.0</td>
<td>5.66%</td>
<td>98.3</td>
<td>92.65%</td>
<td>1.8</td>
<td>1.69%</td>
</tr>
<tr>
<td>Over 60</td>
<td>66.8</td>
<td>9.0</td>
<td>13.49%</td>
<td>55.9</td>
<td>83.68%</td>
<td>1.9</td>
<td>2.84%</td>
</tr>
<tr>
<td>TOP 60 ENS</td>
<td>128.2</td>
<td>7.7</td>
<td>5.98%</td>
<td>118.8</td>
<td>92.68%</td>
<td>1.7</td>
<td>1.34%</td>
</tr>
</tbody>
</table>
Figure 4.14: Main results regarding the simulation hybrid1 and hybrid2 strategies in comparison with “the over 60 years” renewal strategy.
4.2 Results discussion

The results obtained from the different strategies simulated, allow to withdraw some conclusions regarding the efficiency of the strategies. To do so in Figure 4.15 is presented a graph with the Pareto efficiency frontier obtained through the simulation of the above mentioned strategies.

![Figure 4.15: ENS values as a function of the strategy cost for the simulated renewal strategies together with the respective Pareto Efficiency Frontier](image)

The Pareto efficiency frontier shows that there are many simulated strategies that are not efficient, which are the ones to the right and above the blue line that represents the optimality frontier. The strategies on the frontier are the ones that are more efficient in terms of the binomial ENS-total cost. The “over 40 years” and the “zero strategy” have a very high value for one of those two parameters, which in this case are the total cost for the “over 40 years” strategy and the ENS for the “zero strategy”. On the other hand both hybrid strategies, together with the “over 60 years” strategy are the most balanced strategies, since there is an equilibrium between the ENS and the total cost associated with the strategies along the total simulation time frame.

Another interesting result obtained through the simulation is the fact that the amount of PTs that are repaired after failing is not significant, since the recovered transformers have an average weight permanently below 2% of the total strategy cost. This verifies in all the strategies simulated. This happens since the vast majority of the transformers fail after 50 years of life, which makes them unavailable for recovery, while, on the other hand, only half of the ones that fail before the 50 years are recovered. Making this result exactly what was expected to be obtained.
Chapter 5

Conclusions

5.1 Achievements Summary

At the introduction three main objectives were established for this thesis work: to compute the current energy not supplied (ENS) due to failures of the transformers (PTs) in the distribution grid; to develop a predictive model of the health condition evolution of the PTs with time; use this model to construct a simulator to test different strategies of grid renewal and obtain the corresponding ENS predictions.

The field of study related with the predictive aging models for the health of PTs has not been deeply explored from a technical implementation stand point. This was the motivation to develop a custom made model for the PTs aging. The model was parameterized based on the data provided by EDPD. Since the problem addressed in this thesis is applied to a particular equipment, the amount of bibliography regarding the development of this kind of models is limited. The bibliography analyzed turned out to be more useful for supporting the technical aspects of the model developed, such as: the Markov Chain Theory, the different types of probabilities distributions, that could be applied to the health behavior of the transformers, the Monte Carlo Simulation Theory and the data presentation theories, as for example the Pareto Optimality Frontier.

An aging model was developed to estimate the evolution of the health condition of each PT with time. The model was developed using the current condition data from over seven hundred PTs installed in the Portuguese distribution grid. The health condition is assessed through physical tests performed by EDPD and is represented by a health degradation index (PDI).

For the development of the aging model, it was necessary to define a stochastic process based on Markov Chains. These type of processes are particularly useful when there is a need to define the probability for the evolution of states, depending only upon the current state. In this case, the current state is a defined health index (PDI) associated with each PT. The next state is a PDI value associated with the same transformer, after a five year period. The use of a Markov Model was essential, since the only available data regarding the PTs was the static current health condition. This is, there was no information related with the historical dynamic evolution of the health condition of the transformers.

During the development of the aging model, there was a need to define discrete intervals for pa-
Parameters such as, PDI values and age intervals for the transformers. It was then possible to define, based on the empirical data, the probability of each PT to be in each PDI interval, depending on its age. The PDI probability distributions follow log-normal and maximum value distributions, depending on the corresponding age interval. There is a well defined correlation between the average PDI values and the age intervals of the PTs, represented by a second degree exponential function. The same phenomenon occurs with the variance of the PDI values and the PTs age. The average values together with the probability distributions for the PTs per age interval, allow the computation of the PDI probability matrix. This matrix represents, on a discrete way, the probability of each PT having a specific PDI value, depending on the age of the transformer.

The PDI probability matrix was the starting point for the computation of the PDI transition matrices. This matrices define the probability of a PT, from a certain age interval and on a certain PDI level, degrading to a higher PDI level. Eight transition matrices were defined, one per each one of the five year intervals of the simulation horizon. To parameterize the values of each transition matrix, there was a need to solve a non-linear multivariable optimization problem, for each of the matrices. To compute the solution of the defined problems it was necessary to use a specific optimization software, named GAMS (General Algebraic Modeling System).

The developed Markov Model allowed to simulate the so-called PDI trajectories, which are an estimate of the way the PDI value of each transformer evolves with time. The estimated PDI level can then be converted into a failure index, defined according to the EDPD criteria, and then normalized into a failure probability. The relationship between the failure index and the failure probability is obtained through a normalization between the current PTs average failure index and the average historical failure rate. The aging model allows the generation, for a certain time horizon (35 years), of an estimation for the PDI trajectories and the corresponding failure probability, for each of the installed PTs. Based on this failure probabilities and using Matlab random number generator it was possible to simulate, for each sample, if and when a transformer will fail during the 35 years simulated time frame.

The global stock condition simulator used the aforementioned failure probabilities. This simulator is an application developed to test the evolution of the average condition of the installed PTs. This program simulates the evolution of the average health condition of the installed PTs, based on the samples obtained from the Markov Model estimation. Besides the health condition of the PTs, the simulator allows to estimate other variables, such as: average ENS per year; total cost of the implementation of a renewal strategy; average number of failures per time interval, among others.

Every time a PTs fails it can either be replaced by a new one or repaired, with each option having different costs. This allows to estimate the investment needs of the grid throughout the years to come. With the use of the simulator it is possible to extract a larger number of outputs, complementing the analysis. Some of the extra outputs are: the total amount of new and repaired PTs; the amount of failures in non redundant PTs, meaning that the energy supply is lost to over 75% of the clients connected to them; the amount of failures per type of transformer.

The Monte Carlo Simulations Method was crucial in order to increase the reliability of the developed simulator. Each Monte Carlo simulation runs the global stock condition simulation one thousand times.
Each global stock simulation estimates the health condition trajectories for the over seven hundred PTs installed in the grid. At the end of each Monte Carlo simulation, there was a total of over seven hundred thousand trajectories simulated. An average value of the outputs of each simulation was then computed. The results of each general Monte Carlo Simulation are the average values and the Monte Carlo variance of each output variable. Taking into account that the modeled process is a stochastic process, the Monte Carlo Simulation Method allows to obtain the results based on statistical empirical data. The method uses randomness to solve a problem, which depends on a large amount of variables, that might be deterministic, but is not described by known equations. A validation test for the correct implementation of the Monte Carlo Simulation was the analysis of the frequency distribution of each output variable around its mean value. The values obtained for each of the output variables had a normal distribution around the mean value, which was the expected result.

Each installed transformer is associated with corresponding supplied clients, which have a specific energy consumption. It is also possible that some PTs are connected with small producers that supply energy, instead of consuming. Based in the energy consumption values and in the failure rate associated with each county, where the PTs are installed, it was possible to compute an estimation for the current total ENS. Then, based on the developed aging model, together with the current PDI data associated with each PT, it was possible to compute an estimation for the current failure probability of each individual PT and compute a value for its associated ENS. The value obtained for the ENS, based on the real failure rate per county, and the value based on the developed model turned out to be very similar, which corroborates the validity of the model.

Using the global simulator, multiple renewal strategies were tested. This simulator turned out to be particularly useful, since it allowed to test different grid renewal strategies and compare the respective outputs. The implemented renewal strategies had different base criteria, such as: the age of the transformers, the degradation level, the possible ENS associated with each PT failure or the RUL (remaining useful life), already defined by EDPD for each PT. A different kind of renewal strategies are the so called hybrid strategies, which use a combination of multiple criteria to define the grid renewal plan.

One of the outputs of the simulation is the total cost associated with the implementation of each strategy, for the whole simulation period. In order to enable the comparison of the total costs for each strategy, the partial costs are updated to today monetary values, through a defined interest rate. The updating process is similar to the Net Present Value (NPV) computation method, and has its algorithm defined by EDPD.

As a way to represent in a more intuitive way the results, the outputs are presented in a Pareto Optimality Frontier graph. The graph has on each axis the cost and the ENS associated with each strategy respectively. This is a good visual method for presenting the results, since it allows the direct comparison between strategies. The results state that from the tested strategies the hybrid ones are the most efficient. The results also suggest that it will be difficult to reduce the ENS, along the next 35 years, below the values already obtained for the hybrid strategies.

The three initially proposed objectives were achieved successfully. The developed aging model for each individual transformer models, on a feasible way, the evolution of the health state of the PTs. The
model can be improved in the future, if dynamic historical data regarding the PTs is made available. The global stock evolution simulator turned out to be versatile, allowing to test and simulate any desired renewal strategy and retrieving the corresponding estimation for the outputs values. This can be used by EDPD in the future as a tool to provide guidance when discussing and defining the renewal strategy to be implemented.

5.2 Practical Implications

The most relevant future implication drawn from this dissertation are the values obtained from the simulations regarding the evolution of the grid PTs health condition. At the end, from all of the simulations it is possible to conclude that there will be an extreme need for investment in the renewal of the stock of PTs installed in the grid, in order to avoid a possible collapse in the distribution service quality. ¹

There are over seven hundred PTs installed in the grid. It is a simple conclusion that, considering an optimistic lifespan of 55 years for each, the replacement rate in order to maintain the grid operational, without taking into account the early failures, would be of over sixty five transformers per five-year interval.

This work can indicate to EDPD the need to develop a long term strategic plan for the grid renewal, in order to avoid major problems in the service and reduce the associated investment effort. The main contribution of this work is the possibility to test the implementation of different renewal strategies. This way it is possible to develop less expensive and more efficient strategies, regarding the associated ENS based on the retrieved results.

The main differences in the results of the implemented renewal strategies are the time periods when the investment in the grid will have to be applied. Although there are also some differences in the way the investment is made, which consequently results in some differences in the financial burden associated with each strategy. At the end, the main conclusion is that there is no alternative to the investment in the grid renewal. The real question is how and when to do this investment, which, based on the performed simulations, should be based on a custom made hybrid strategy.

A possible obstacle for the strategies comparison is the valuation that is made for the ENS cost. The value assumed by EDP per KWh not supplied is too low to make it possible to truly differentiate strategies. The cost, for each simulated strategy, associated with the ENS is always a small fraction of the total implementation cost of the strategy. Without a higher valuation of the ENS it will be difficult to correctly asses the best renewal policy. With the current valuation, even the run to failure strategy (zero investment strategy) can be costs competitive, when compared with the other renewal strategies.

The simulations results also show that, if there is not an implementation of an adequate renewal strategy, it is likely that in the future there might occur a large amount of failures in the same time period. A major concern, besides the ENS and the cost associated with these possible failures, is the lack of capacity from the suppliers to provide the amount of new PTs needed to replace the ones that will fail in

¹phenomenon happening in Infraestruturas de Portugal, which is the Portuguese company responsible for managing the trains infrastructures, which saw a collapse in the service quality due to lack of investment.
a short time interval. It might be difficult to find companies capable of supplying over a hundred PTs in a five year period, which should be a concern to EDPD, when developing a grid renewal plan.

At the end, this work can be used as a support for the decision-making process regarding the grid renewal strategy, together with the creation and development of the long term strategic plan of investment in the renewal of the distribution grid PTs.

5.3 Future Work

The work developed for this thesis can be used as a base for future studies regarding the evolution of the distribution grid health condition.

Aging Model Improvements

In order to improve the reliability of the aging model developed for each PT individually there are some modifications that can be made.

If, in the future, it is possible to obtain data regarding the dynamic evolution of the PHI of each individual transformer it would be possible to improve the defined Markov Process by better defining the relationship between the age and the health level transitions for the transformers. Instead of defining this relationship based on the static data of the PHI values, it would be possible to analyze dynamic historical data and better define this process.

The estimation for the current ENS value can increase its reliability with the obtainment of data regarding the failure rates of each individual PT. This would be a major improvement to the current used data, regarding the average failure rates of the PTs per county. This new data would also be extremely important to more accurately define the relationship between the failure index of each transformer and its failure probability. This means that this new data would drastically improve the feasibility and reliability of the developed aging model.

In the developed model it is considered that all the malfunctions result in a transformer failure. If there was more data regarding the type of failures, and corresponding consequences for the transformer health condition, the aging model could be refined in order to reflect, in a more realistic way, the behavior of the transformers.

Global Installed Stock Analysis

In the future, the developed model can be used to study custom made renewal strategies in order to obtain a more efficient approach, than the ones presented in this thesis. Based on the developed Global Stock Condition Simulator, it is possible to define any desired criteria for the renewal strategies and test the influence on the output results. The optimal strategy might be difficult to find, since there are an infinite number of possible criteria that can be defined. It might be possible, although, to obtain new non-dominated points and with that redefine a new Pareto Optimality Frontier.

Another possible application for the developed simulator is the development of studies regarding the reserve stock of transformers. Based on the prediction of failures for the installed transformers, divided
by PT type, it is possible to study strategies for the sizing of the reserve PTs stock. There is a very large bibliography of studies and theories regarding this subject matter. Some examples include the articles in [42–44] and all the theory presented in the book [45].
Bibliography


