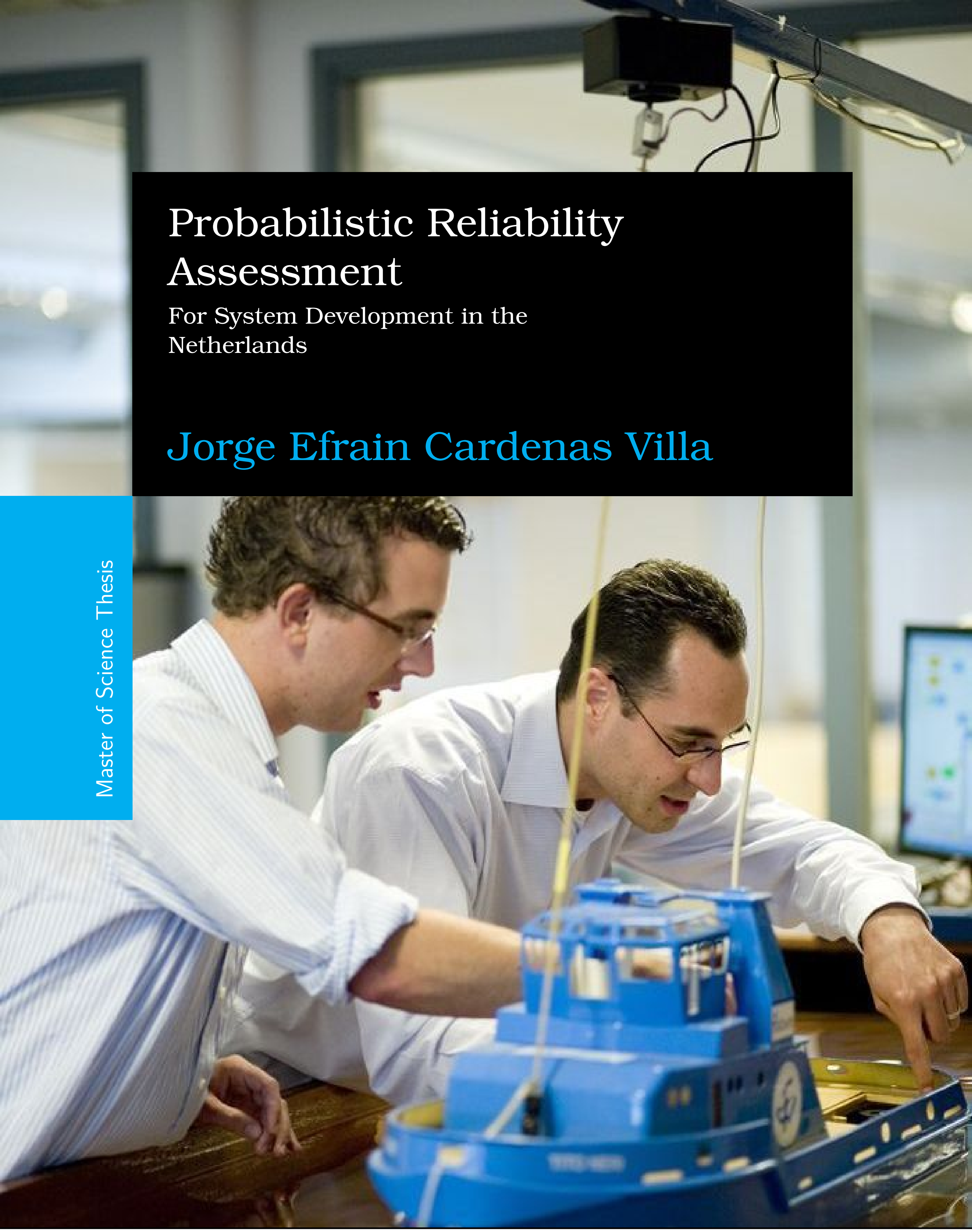
Probabilistic Reliability Assessment

For System Development in the Netherlands

Jorge Efrain Cardenas Villa

Master of Science Thesis



TU Delft Delft University of Technology

Electrical Engineering, Mathematics and Computer Science (EEMCS)

Probabilistic Reliability Assessment

For System Development in the Netherlands

Master of Science Thesis

For the degree of Master of Science in Electrical Engineering at Delft University of Technology

Jorge Efrain Cardenas Villa

defended publicly on Thursday October 10, 2019 at 15:00.

Student number:	4747550	
Thesis committee:	Dr. ir. Jose L. Rueda Torres,	TU Delft, supervisor
	Prof. ir. M.A.M.M. van der Meijden,	TU Delft
	J. (Jianning) Dong ,	TU Delft

Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) · Delft University of Technology

The work in this thesis was supported by TenneT TSO BV. Their cooperation is hereby gratefully acknowledged.

Abstract

Power systems have been subjected to several changes due to a constant evolution over the last years. The changes that come along with this evolution include several aspects such as the introduction of variable energy generation from renewable sources, changes in the network's topology, storage, new regulation policies, liberalization of markets, constant trading of electrical energy between countries among others. These changes, combined with a context of growing demand and aging network infrastructure can result in more stressed power systems, uncertainty and potentially more risks that could lead to blackouts. These new conditions have to be taken into account when assessing the reliability of the transmission system in order to have a complete insight of what is happening in the network and to identify potential causes of problems. This thesis provides a methodology using state enumeration to assess reliability in the transmission network in the context of long term planning using a probabilistic approach with the aim to identify potential improvements for current practices based on deterministic approaches. This methodology considers three main aspects: i) Failure probabilities of individual components (i.e transmission lines) based on historical data and operational conditions. The failure probability originated from operational conditions is obtained by combining expert's opinion and available data by applying Bayesian data analysis; ii) Contingency selection to obtain a representative list of contingencies to be analyzed. The screening of contingencies for the list is performed based on the failure probabilities and accepting a certain level of residual risk. The residual risk is calculated by assuming a lost load and finding the probability of the contingency that leads to that lost load; iii) Determination of the effect and the associated risk of contingencies. The purpose of this part is to quantify line overloadings in terms of money to obtain the total cost linked to a certain set of contingencies. The monetization of the effect is performed by using a value of lost load (VoLL). This monetization is first performed without the implementation of re-dispatch actions assuming that overload is lost load. Later on, re-dispatch actions for conventional generators are considered, the new cost of re-dispatch plus lost load is obtained. Furthermore, a comparison between the two costs is performed. Re-dispatch actions are implemented by using a linear programming approach with the selection of an objective function for the minimization of the amount of conventional generation. Risk of lost load is obtained in terms of money by the multiplication of the probability of the contingency times the cost. The proposed methodology obtains reliability in terms of system adequacy by investigating line overloads and lost load. Further on, this methodology is tested on a test model for the 380/220kV grid topology of the transmission grid of the Netherlands. The results show the variation of indicators of risk of lost load and guidelines to continue developing probabilistic approaches for reliability analysis. The computational burden and the time employed for this kind of approaches is still a challenge to be tackled.

Acknowledgements

I would like to thank everyone that contributed to the completion of this thesis through feedback, motivation, advice and support.

Especial gratitude to my supervisors Richard de Groot from TenneT and Jose Rueda from TU Delft for their feedback, discussions and the opportunity to work in such an interesting project.

Moreover, I would like to thank the whole asset management department in TenneT for their disposition to help and willingness to answer any questions related to this thesis.

Last but not least, I would like to thank my family and friends for their support throughout my Master's studies.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the objectives and intended structure of the work done for the project "Probabilistic Reliability Assessment for System Development in the Netherlands" with collaboration of the Dutch TSO TenneT B.V. (hereafter TenneT). The objectives and the intended course of work are presented in order to have a clear explanation of the goals and the necessary work to achieve them.

1-1 Background

In power systems, the transmission system operator (TSO) is the entity responsible for the operation, maintenance and development of the transmission system in a defined area, including its interconnection to other (foreign) networks. The TSO ensures that the transmission system is able to transport energy, meet future demands in a long term time frame and facilitate electricity market needs. On top of that, the TSO also manages power restoration after disruptions. In the Netherlands, the TSO is TenneT. TenneT is responsible for keeping a stable system operation maintaining the power balance and keeping voltage between limits. Also, TenneT has the duty of keeping the reliability of the transmission system at a high level and to ensure that enough transmission capacity is available in order to facilitate market needs and the power supply to loads. A reliable transmission system is immensely important, many activities and services require electricity. If there is a disruption on the electrical supply, then these services will be disrupted as well. TenneT as TSO, performs several activities to ensure reliability. These activities are divided in three categories that consider the creation, maintenance and operation of the transmission system over various time frames (long-term, mid-term, short-term). These categories are:

- Grid services: it concerns the organization and supervision of construction and refurbishment activities on the transmission grid.
- Asset management: it concerns the planning for the expansion of the transmission network to provide solutions for future operation. In addition, manages (maintenance

actions) the current equipment and infrastructure installed in the transmission system so they can be used in an efficient manner.

- System operation: deals with the operation of the current transmission system under current working scenarios. The aim is to maintain the system security at a high level.

This project will focus on the reliability assessment in the context of grid planning.

In today's transmission systems, uncertainty is mainly affected by the increased penetration of renewable energy sources (RES) with a variable nature, random failures of components, environmental conditions, demand side management, electrification of different sectors (transport, heating) and the increased difficulty of building bulky transmission assets due to societal opposition. Uncertainty might increase risk and a higher risk can result in a lower reliability of the power system.

Currently, to assess the reliability of the transmission system in the Netherlands, TenneT uses a deterministic approach (N-1, N-2 under maintenance) combined with a risk estimation using a predefined risk matrix. Some risk associated with factors such as the variability of RES, random behaviour of failures in contingency selection, re-dispatch actions and demand side management might be omitted by a purely deterministic assessment resulting in an inaccuracy in risk estimation.

A probabilistic reliability assessment (PRA) approach offers the opportunity to judge situations quantitatively and qualitatively considering the probability of events making possible the combination of contingencies to obtain more realistic results in risk estimation. The downside is that realistic results often mean more calculations resulting in a high computational burden.

TenneT's current procedure incorporates probabilistic aspects such as the variability of RES on scenario constructions and failure probabilities based on historical records. However, the operation and planning are still ruled by a deterministic approach.

By expanding the probabilistic approach in the reliability assessment performed by TenneT, the random failures of elements can be included for the contingency analysis making it possible to analyse a wider range of possible scenarios offering the potential to consider different variables that might give better insight of risks allowing for better cost benefit analysis.

Some work has been developed on probabilistic approaches, an example is the GARPUR methodology developed by a coalition of TSO's, universities and research providers with the support of the European union. The GARPUR methodology presents a framework called Reliability Management Approach and Criterion (RMAC) which has four main components which will be discussed in further chapters:

Different RMAC's have been designed to fit every context of the power system (system development, asset management , system operation) they all keep the basic elements as their base, the difference lays in the considerations and calculation methods employed for every context.

1-2 Current reliability assessment performed by TenneT

Every two years, TenneT is required to publish an investment plan report (IP report). This report addresses quality of service, presents the planning scenarios used in network calculations, the results of a reliability assessment and leads to an overview of planned investments. The reliability assessment shown in figure 1-1, is a simplified version of the actual reliability assessment from the IP report. This reliability assessment flux diagram was constructed based on the information acquired in a series of interviews with TenneT staff.

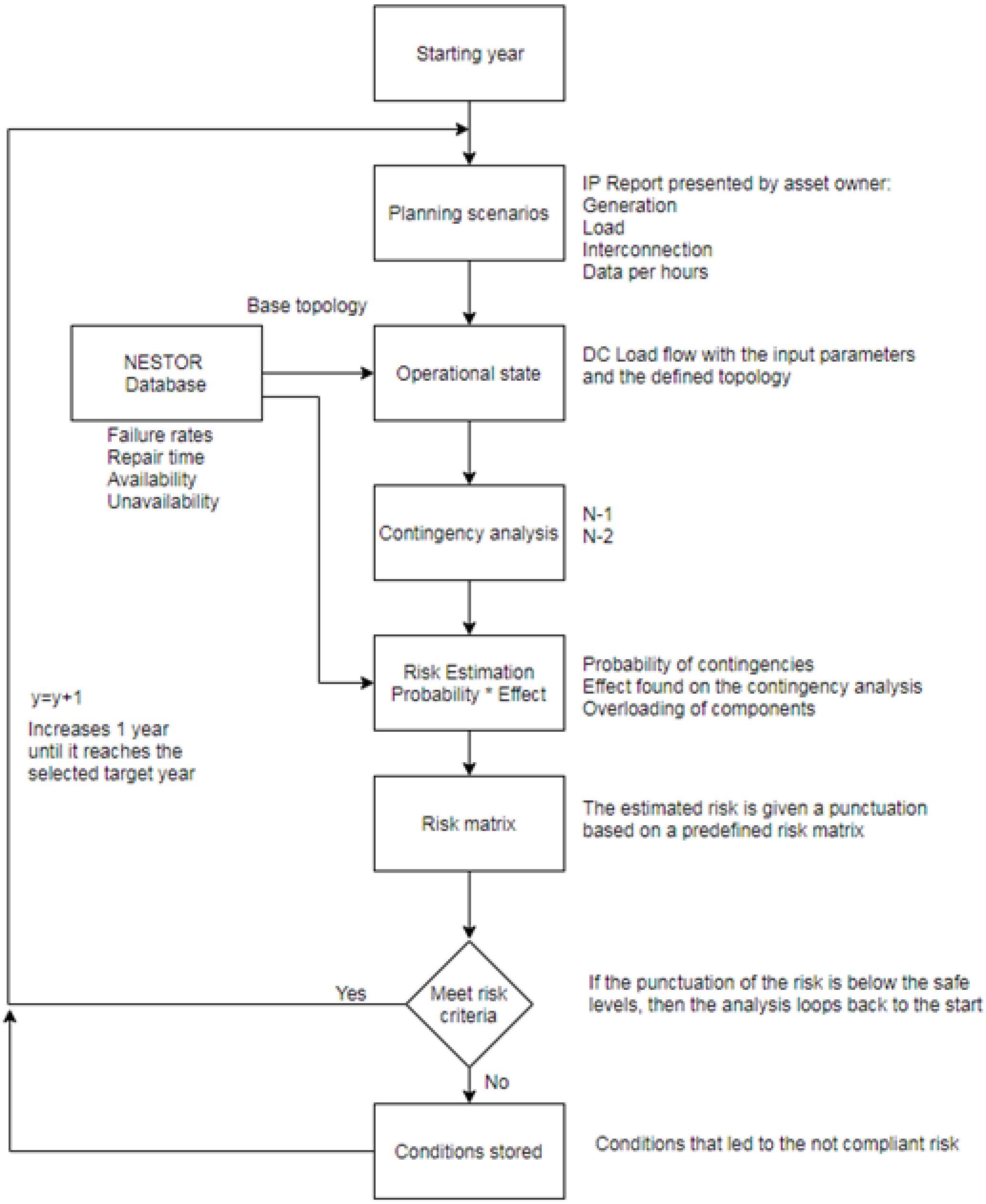


Figure 1-1: Simplified reliability assessment performed by TenneT

The description of the procedure for the reliability assessment of figure 1-1 is the following:

- The process starts with the selection of a starting year and an ending year.

- An hourly projection of generation (conventional units commitment and RES), load and interconnection between neighbours per year based on market, weather and expansion forecast is taken from the asset owner department.
- The base topology of the selected year is obtained from the state of components and the failure probabilities calculated from the NESTOR database.
- The operational state of the network is defined by performing a power flow analysis for every individual hour over the base topology.
- Once that the operational state has been determined, a deterministic (N-1) contingency analysis with a decoupled power flow is performed to determine any constraints in the grid. Since the method for the power flow calculation is the decoupled power flow, the constraints found are in the form of overloads. Other types of constraints such as voltage and stability issues are identified using separate studies that are not mentioned in figure 1-1 and are out of the scope of this report.
- The risk is estimated as the product of the probability of the contingency and the effect.
- The estimated risk is then placed in a predefined risk matrix that contains the acceptable levels of risk.
- If the risk is below the specified safe level, then the whole analysis is repeated for a new year.
- If the risk is over the specified safe level, then the identified conditions are stored for a later analysis and the whole analysis is iterated for a new year.

The reliability assessment ends when the target number of years has been reached and either safe and unsafe conditions have been identified. After the analysis is over, an investment proposal is developed based on a holistic view of all constraints.

1-2-1 Risk matrix

The risk matrix shown in figure 1-2 is a tool used by TenneT to estimate the risk associated with contingencies and future planning scenarios for network development. Several business values are defined namely safety, secure supply, financial, engage stakeholders, environment and compliance. The risk of an event is estimated by taking into consideration the effect and the frequency. If the risk is estimated to be in the green zone, then it is judged as acceptable and no further actions are needed. If the risk is placed outside the safe region, then something has to be done to either reduce the effect and/or the frequency. An event with a minor effect and low frequency will be placed in the safe region of the risk matrix, on the other hand, an event with an extreme effect and a high frequency will be placed in the not acceptable region of the matrix.

In the example of figure 1-2 the indicators for safety, compliance, engage stakeholders, security of supply and financial were placed in the risk matrix. As it can be seen compliance and safety are in the not acceptable zone. The frequency of the risks can be calculated by probabilistic methods and the effects can be determined by case studies and simulations.

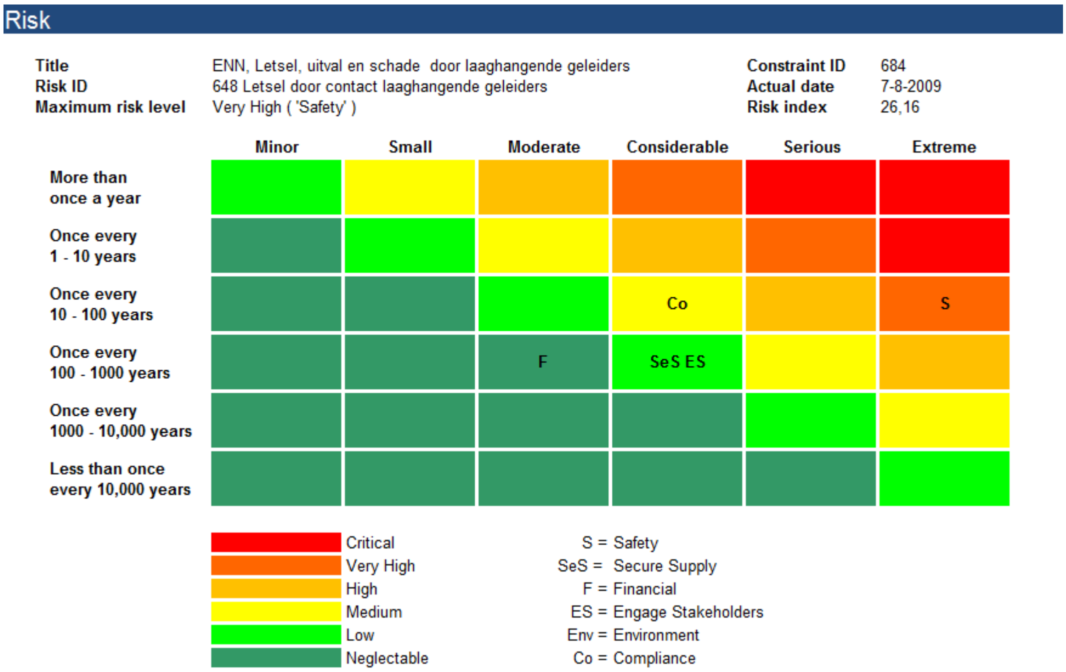


Figure 1-2: Risk matrix used by TenneT

1-2-2 Strengths and weaknesses of the analysis

This reliability analysis has a number of strengths and weaknesses. Some of its strengths are:

- It works well, resulting in an adequate transmission grid over these past years , analysts are used to it and has proven to be straightforward.
- The results are easy to interpret and it is easy to identify when to invest.
- The risk matrix is easy to apply.
- Overloaded elements can be easily identified.

Some of its weaknesses are:

- May lead to over-investments and overprotected systems. The deterministic nature of the contingency analysis might lead to pessimistic determination of the effect since the analysis covers worst case scenarios and not the most likely scenarios.
- The risk estimation may not be realistic, not all scenarios are investigated only a few scenarios with worst case conditions. There could be the case that another scenario (not the worst case) results in a higher risk for the transmission system.
- There is not a combination of contingency scenarios. Only one scenario is analyzed.

- There is not a determination of the socio-economic effect of a certain risk. The monetization of the effect is calculated based on the cost that TenneT would have to pay to reduce a certain risk, but there is not a calculation of the cost for society of that particular risk.

1-3 Upgrade opportunities for the reliability analysis

The identified opportunities for the upgrade of the current reliability analysis are presented in table 1-1.

Proposition	**Currently**
Development of clustering algorithms for the data provided by asset owner in a way that accuracy is not compromised.	Every hour is analyzed.
Failure analysis of components based on historical data and operational conditions of the components to obtain a more realistic availability and unavailability of individual components.	The failure frequencies and repair time are calculated based only on historical records.
Contingency probability based on the probability of failure of components. Combination of contingency scenarios.	N-1 and N-2 contingency analysis.
Establishment of an acceptable reliability target and acceptable residual risk levels.	The reliability target is being always N-1 secure. The risk is analyzed using the risk matrix.
Implementation of a discarding threshold and tuning of a relaxation principle for the analysis of contingencies.	N-1 and N-2 contingency analysis.
Determination of the effect and development of a proxy for the monetization of the effect.	The considered effect is overloading and the consequence of the effect is that overload is lost load.
Include re-dispatch options in the analysis and determine the effect of an ineffective re-dispatch action.	Re-dispatch is managed in the context of system operation. Re-dispatch bids are managed in day ahead and real time operation.

Table 1-1: Upgrade opportunities.

The aim is to work on the identification of hidden risk situations that may not be covered by the current method.

The potential development of a probabilistic approach for the reliability assessment performed by TenneT for the analysis of possible risks not considered by the current methodology leads to the following objective and research questions.

1-4 Objective and research questions

The main objective of this master's thesis is to investigate the potential improvements that can be brought to TenneT's reliability assessment in the context of system development by applying the principles proposed by the probabilistic reliability assessment approach. In order to fulfil this objective, the following research questions are proposed:

What is the feasibility of applying/combing a probabilistic methodology (GARPUR RMAC, others) for long-term transmission planning with the current method employed by TenneT?

This question is the base for the development of the project. The aim is to identify the requirements (assumptions, approximations), benefits and drawbacks implied for the application of a probabilistic approach into TenneT's current reliability assessment. The development of this question leads to the following sub-questions:

Which methods can be used for the determination of failure rates of components, contingency selection, effect determination and risk analysis as inputs for models in order to perform a probabilistic reliability assessment?

Probabilistic approach models rely on the quantity and quality of input parameters. This question seeks to investigate methods that can offer an upgrade to the practice currently implemented. The research will be directed towards the determination of the probability of failure of components (considering historical data and operating conditions), selection of suitable algorithms for a smart contingency screening, determination and monetization of effects resulting from the contingencies (line overload, load not supplied, re-dispatch) and risk analysis considering the probability of the contingency times the effect.

What is the computational performance resulting from the implementation of a probabilistic reliability assessment, how can it be reduced and under what conditions?

The implementation of a probabilistic reliability assessment supposes a high computational burden (great quantity of calculations) that implies long computational times. The time employed for an analysis is an important parameter and should be kept as low as possible, as an example, a maximum accepted time for long term analysis can be up to 48 hours. A reduction of computational time can result in an accuracy reduction, so the balance between time and accuracy should be kept in a way that both parameters remain acceptable. Computational time vs accuracy will be investigated.

What is the level of confidence and the robustness (in terms of sensitivity) that can be achieved with the implementation of a probabilistic reliability assessment?

The aim of this question is to determine how trustworthy the analysis is and what input parameters might affect the final result. A selection of contingencies implies that some situations might be discarded and some risk omitted. A level of confidence determines what the omitted risk is and its impact on the overall result. If the impact is negligible, then the level of confidence is high and the analysis is trustworthy. The sensitivity analysis determines how the variation of input parameters affect the result of the analysis.

1-5 Project approach

To achieve the objective of this project, the work will be divided into the stages described below.

Literature research

A literature research of methodologies and models related to probabilistic approaches will be conducted to explore feasibility, benefits and drawbacks for the implementation of a reliability assessment. It is important to note that every stage of the project will be combined with literature researches related to the current task.

Determination of failure probabilities of components

A research of models and methods for the calculation of failure probabilities of components not only based on historical data (NESTOR database), but also on operational conditions (loading, weather, external factors, human factors) will be conducted. Also, information needed to find correlations and dependencies of failures will be studied.

Determination of contingencies

A research for methodologies and algorithms for a smart selection of contingencies will be conducted. From the research, criteria for the selection will be developed in order to give importance ranks to contingencies and screen them. The selection of contingencies can be based either on the probability of occurrence or the associated effect. The selected methodology with its resulting contingency list will be evaluated in a later stage.

Determination of effect and associated risk

A research for methodologies and algorithms for the determination of the effect (finding the meaning of line overload) of certain scenarios will be developed. The calculated effect will be quantified in terms of money in order to obtain the total cost linked to the effect obtained from a set of contingencies. The risk associated to a certain effect and its cost will be calculated. First, the calculations will be performed without remedial actions. Once that the effect, its cost and the risk are defined, remedial actions will be taken into account for new calculations and a further comparison between options. For the development of this question, some assumptions will have to be made. These assumptions will be made in a way that they do not differ greatly from literature and real cases. To accomplish this, a research of the different processes involved will be conducted by series of interviews with TenneT's staff.

Development of a test model of the 380/220kV Dutch transmission grid

A test model for the 380/220 kV transmission system will be constructed using TenneT's available software tools (PowerFactory) for power flow calculations. The model will be coupled with a Python script for automatization of tasks and mathematical software (if needed) for further calculations required.

Study case development and testing

A proof of concept study case will be developed to obtain and analyse results. The methodologies and algorithms found for the previous stages will be tested over the developed test model in order to find differences and advantages of the implementation of probabilistic approaches into the reliability assessment. From the tests, computational burden and execution time will be determined. Alternatives to reduce the computational load will be proposed. The aim is to identify relevant situations worth of analysis and leave aside situations that do not contribute significantly to a result but contribute as computational load.

1-6 Outline

Next, a description of the chapters from this project will be provided.

Chapter 2: Theory and methods

Presents the theory needed for the completion of this project.

Chapter 3: Application of methods

Describes the application of the methods introduced in chapter 2.

Chapter 4: Case study and results

Presents the development of a case study and its results.

Chapter 5: Conclusions and recommendations

Presents a summary of the results and recommendations product of this project.

Chapter 2

Theory and methods

This chapter provides an explanation of the basic theory that is used for the development of this project. General notions of probabilistic reliability assessment are introduced as well as different approaches to assess reliability. The estimation of failure probabilities of transmission lines and Bayesian data analysis is also presented. Finally, effect determination and clustering techniques for dispatch scenarios are described.

2-1 General notions of probabilistic reliability assessment (PRA)

Power systems have been under constant changes over the last years. The power flows are more volatile. Some of the aspects that modify power flows are:

- The increased penetration of RES with a variable nature and stochastic behavior.
- The variable use of the electrical network with the introduction of power electronics and prosumers
- The increasing possibility of load management measures
- Increasing re-dispatch actions
- Constant liberalisation of the market
- Increased energy transfers via inter-connectors between countries
- Aging and random failures of components

The combination of all of these factors result in more uncertainties and operating conditions closer to the limits, making power grids more vulnerable [2] and stressed. These new conditions have to be taken into account when assessing the reliability of the transmission system in order to have a complete insight of what is happening in the network and to identify potential causes of power interruptions and blackouts.

2-1-1 Definition of Reliability and associated concepts

In power systems, reliability is traditionally defined as the ability to supply electrical energy to the loads over a defined time interval. If by any reason, the load is not supplied with electrical energy, then the power system is unreliable [1]. Another definition found in [3] states that reliability is the probability that a power system can fulfill a determined function under certain conditions for a given time frame.

The first definition seems outdated for modern power systems given that it implies that securing the supply of electrical energy to the loads depends on one single actor. The second definition tries to see reliability in a broader sense and is more fit for today's power systems. There are several entities that play a role in power systems, as an example we have producers, consumers, operators, traders and prosumers. The TSO is responsible of facilitating energy trading by securing enough transport capacity and connecting all customers (generation, loads). If consumers can not be connected or there are transport restrictions that limit the market, then the transmission system can be seen as unreliable. Reliability is not only about serving the loads but also allowing means of generation to be connected and provide enough transport capabilities for the market to develop without restrictions.

Reliability of power systems can be divided in two aspects: System adequacy and System security [1] [3].

System adequacy Takes into account the availability of resources in the power system to serve its customers, in other words looks in to the capacity of the power system to supply the requirements of customers. It is a static condition of power systems.

System security Takes into account the power system's ability of overcome disturbances and unplanned events, the dynamic behavior of the system is considered in system security.

This report focuses on system adequacy.

When defining reliability of power systems it is necessary to be aware of other associated concepts [1, 3] such as:

- **Scenarios:** Relates to a particular set of circumstances (values of generation and loads) and events within the power system.
- **Contingencies:** Which are unforeseen failures of any component of the power system (lines, generators, transformers). A contingency can also include the failure of multiple components at the same time. Contingencies can lead to outages.
- **Outages:** Can be defined as a state where a part of the system or the whole system is unable to perform its function.
- **Energy not supplied (ENS):** Is the estimated energy that would have been delivered to customers without an outage.
- **Value of lost load (VOLL):** Is the cost of the not supplied energy during a time-frame.

- **Risk:** Reliability is related to risk. If we take into account the probability of a contingency, as well as its effect, then risk is defined as the product of probability times effect. The risk takes the same units as the effect because probability has no dimension. This means that if the effect is expressed in money, the risk is expressed in money as well.

2-1-2 Reliability Assessment

In a reliability assessment of power systems, the goal is to determine if the power system is able to fulfill its functions if it is exposed to different expected situations. Reliability of the power system depends on the threats, vulnerabilities and consequences. The assessment tries to find what can go wrong and what are the consequences. [3] A reliability criterion is needed in order to perform a reliability assessment.

A reliability criterion is a set of rules which are used to judge whether the system is performing in an acceptable way or not, based on its response to certain contingencies. These rules are often seen as boundaries that indicate which operational conditions are acceptable and which ones are not.

2-1-3 Probabilistic approach

The probabilistic approach considers not only the effect, but also the probability of contingencies in order to determine if the system is secure enough. By using this approach, it is possible to consider a wide range of scenarios and contingencies in a single analysis. The obtained results can be delivered as a single reliability indicator (e.g. combined risk) or as several indicators (e.g. individual risks) weighted by the probability of the scenarios and contingencies used as inputs.

The consideration of the probabilities of scenarios and contingencies makes this approach more realistic. By using this approach, the variable nature of RES and the random failures of elements can be included in the study instead of analyzing only specific cases as is the case of the deterministic approach. These more realistic considerations may lead to the identification of potential risky situations that would have been omitted otherwise. On top of that, since it is possible to quantify a more realistic risk, the optimal size of assets (lines, transformers, protections) can be determined resulting in a potentially well designed power system and possibly more cost efficient in comparison to a power system designed with a deterministic approach.

There are a number of reliability indicators for probabilistic approaches. These indicators aim to estimate the probability of the system's adequacy and security. An extensive list of these indicators can be found in [1, 3]. Some of the indicators are:

- Probability of failure
- Expected frequency of failure
- Expected load curtailment
- Expected energy not supplied

- Expected duration of load curtailment
- Risk of lost load (MWh or Euro)

One of the challenges of probabilistic criteria is that it is difficult to define indicator values to the satisfaction of regulators, planners and operators. In this context, some of the questions raised are:

- What level of risk is acceptable and what level is desirable?
- Could it be that a value of security of supply of 0.999 is accepted and one of 0.998 not if some risk is implied?
- Is it economically effective to connect some remote facilities to the grid in a way that they are N-1 secure or is it acceptable to take some risk?
- Is it acceptable to take some risk under maintenance?
- How much power can be added to a grid without reinforcements?

There is still some work to do in order to define acceptable limits of indicators. In this sense, probabilistic approaches can work in parallel with deterministic approaches and not instead of. There is no doubt that probabilistic approaches have the potential of providing detailed insight of real risk in power systems but, the results still need to be presented in a clear way in order to take decisions based on the analysis.

The probabilistic approach still faces some challenges. The good thing is that given the potential of the approach there is a collective effort in academia and TSO's to overcome them. The most important challenges are:

- Low quantity and quality (sometimes nonexistent) of statistical information about failures and operation.
- Long computation time for probabilistic analysis of large power systems.
- Difficulty of establishing limits and interpreting results.

2-1-4 Application of the probabilistic approach

The basic steps for a generic probabilistic analysis are represented in the flux diagram of figure 2-1. Those steps are:

- **Step 1:** Create a base case corresponding to the scenario to be analyzed. The base case includes the expected topology, expected values of generation and load and forecasted operational conditions for the scenario.
- **Step 2:** Run a power flow for the base case to obtain the current state.

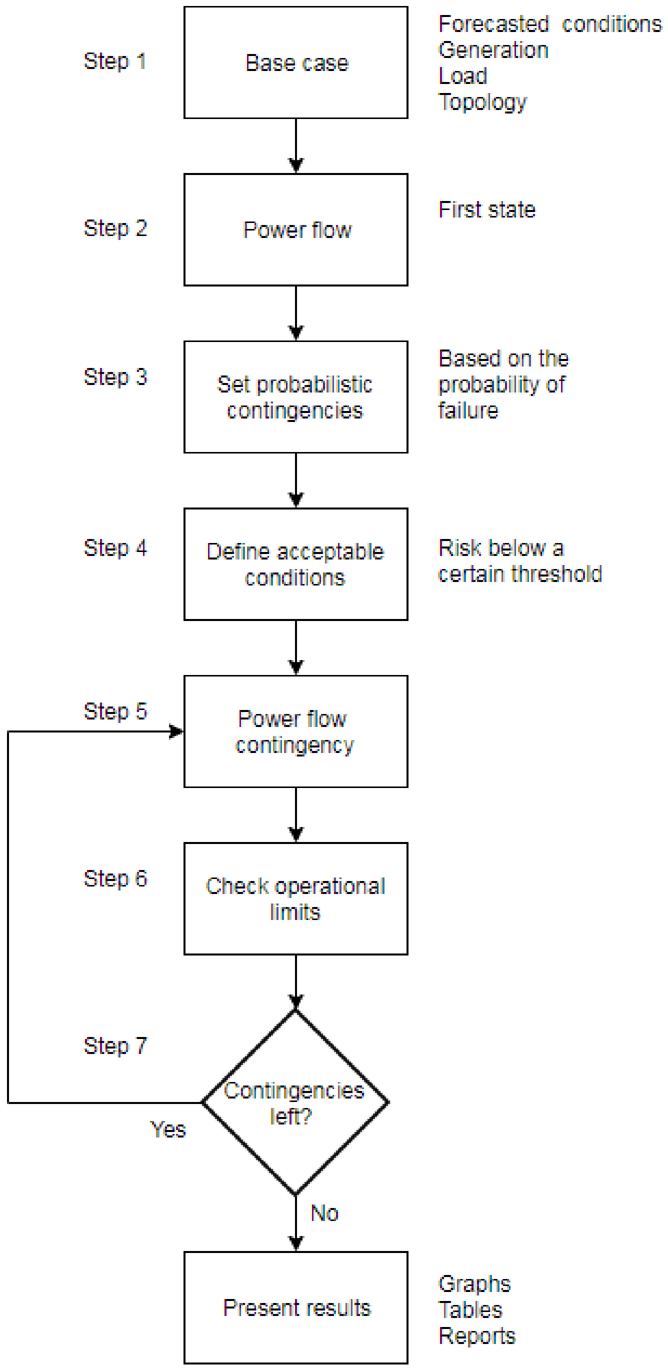


Figure 2-1: Basic flux diagram of a generic probabilistic analysis.

- **Step 3:** Create the set of contingencies based on statistical data of failure rates. Now the contingencies can be selected according to the probability of occurrence. Combination of contingencies can also be selected according to their probabilities.

- **Step 4:** Define the reliability indicator (risk) and the threshold level in which operation is acceptable or not.

- **Step 5:** Take the elements from the contingency set and run a power flow for each element from the set.
- **Step 6:** Check the indicators and identify the contingencies that lead to a violation of the limits (value below the established threshold).
- **Step 7:** Check if there are still contingencies left on the set. If there are, the analysis goes back to step 5. If there are not any contingencies left the analysis continues
- **Step 8:** Present results as information that can be understood and used by operators or planners

2-2 RMAC proposed by GARPUR.

The probabilistic reliability assessment shows an enormous potential with some innovative features to include uncertainty. Some concepts are defined but, there is still the need to standardize and give a common ground to develop assessment criteria to all actors from power systems.

Based on this need, a European project was put in place in 2013 with the collaboration of 7 TSO's, 12 universities and 1 expert in innovation management under the coordination of SINTEF energy research to give birth to the GARPUR methodology [4].

GARPUR stands for "Generally Accepted Reliability Principle with Uncertainty modelling and through probabilistic Risk assessment" and its aim is to find new probabilistic criteria for reliability assessment. On the course of 4 years, GARPUR was able to design, develop and evaluate new criteria for a probabilistic reliability assessment that seeks to maximize social welfare. This section is entirely based on the GARPUR project deliverables [4, 3, 5]

2-2-1 Reliability Management Approach and Criteria (RMAC)

The Reliability Management Approach and Criteria (RMAC) is the methodology proposed by GARPUR to perform a reliability assessment and management. This methodology was designed to be:

- Flexible enough to fit within all time frames and activities of the power system
- Compatible with the resources of TSO's
- Usable and adaptable to the future
- Interpretable

The RMAC defines approximations to reduce the computational burden at the expense of accuracy. The basic functions of the RMAC are to specify what is reliable enough for the power system in a mathematical way and to describe the procedure to take decisions that will lead to an optimization of the socio-economic performance while keeping a certain level of reliability.

The RMAC is specified depending on the context of the power system, this means that for system operations the RMAC will provide the mathematical procedure to manage reliability in a short time frame with the conditions specified by this context, on the other hand, for system development, the RMAC will specify a long term procedure with different conditions. The candidate decisions and the goals change in every context of the power system and the RMAC is decomposed into a family of sub specifications focused on the different context of the power system.

General requirements for an RMAC

The RMACs of the different contexts were designed following the requirements presented below:

- **Relevant behavior of the system and uncertainty models:** Uncertainties are modelled depending on the context and the time frame. These uncertainties can be shared among contexts with the same time frame. The models that describe the behavior of the system and the cost functions of a certain time frame most likely have to be adapted to the decision to be evaluated in order to show the impact of the decision.
- **Computational tractability:** The RMAC has to be compatible with the current computational and data resources available at the TSO's. In order to accomplish this, some approximations are needed.
- **Sustainability:** The proposed RMACs should not be obsolete as soon as significantly more data and computational resources are available.
- **Interpretability:** The proposed RMAC should be able to be compared with current approaches such as the N-1 criterion.

Elements of a generic RMAC

The RMACs were formulated using the theoretical framework of stochastic programming. The stochastic programming allows to include the probability of parameters into an optimization problem that is formulated with an objective function weighted by the probabilities defined for the parameters. This framework is very general and allows to represent the system in a very realistic way, but it can lead to a very complex reliability management problem.

The RMAC is specified by 4 elements, the risk neutral socio-economic objective function, the reliability target, the uncertainty discarding principle and the relaxation principle.

Socio-economic objective function

The socio-economic objective function is a mathematical model of the socio-economic performance of the system over the considered time frame and context. This function can be specified as a cost function that evaluates the socio- economic performance along the proposed scenario, contingencies and candidate decisions to be implemented. This objective function balances the costs and benefits of reliability of all parties related to the power system.

Reliability target

The optimization of the socio-economic function may lead to unacceptable situations in terms of risk. The risk of interruptions may be too large in some of the optimal outcomes from the objective function. The reliability target is a specification of acceptable circumstances for the operation of the power system. It represents the introduction of cautiousness in the stochastic programming problem as a chance constraint with a small probability threshold of being outside the limits.

Discarding principle

The determination of the objective function and the reliability target may not be possible given that the power system and its uncertainties can be too complex. The discarding principle aims to simplify the formulation of the RMAC by a simplification of the uncertainty model, i.e. considering a smaller subset of scenarios and discarding a large space of scenarios. The challenge is to find the level in which the scenarios can be discarded. The idea is that the discarded scenarios would not lead to significant changes on the performance of the power system.

Relaxation principle

If after the implementation of the discarding principle the decision making problem is still not solvable, then a further simplification of the problem has to be implemented. This means discarding more scenarios starting from those that represent the lower risk. The threshold set by the discarding principle is moved to ignore the lower risk events. This threshold has to be relaxed as little as possible.

Each of the RMAC elements have to be specified for each reliability management context.

Proxy RMAC

RMACs have to be consistent along all the contexts of the power system, this means that a long term RMAC has to model the shorter term RMACs in a realistic and flexible way, different time frame and decision making RMACs are closely related one to another. In order to develop a long term RMAC it has to include shorter terms RMACs. For this matter GARPUR came up with the proxy RMAC.

A proxy is a technique that allows to model and determine the behavior of short term decision making stages in order to build a long term stage. An RMAC Proxy is a mathematical model of a reliability management of a certain context of a TSO that models a certain process that can be included in the reliability management of another context of another process of the same TSO or other TSO's.

An RMAC Proxy is general, does not give details but gives the constraints of a certain process to be used on a broader process of a broader reliability management context. If a longer term RMAC is considered, then the shorter term RMACs have to be modelled in an accurate way by developing proxies.

This project is not mainly based on the RMAC from GARPUR. However, several aspects are kept such as the notion of a proxy for the simulation of short term decisions over a long term analysis. The recommendations of computational tractability are followed in order to determine simulation time and the ways to decrease it.

2-3 Rationale of state enumeration analysis for reliability assessment

In large systems, the consideration of the combination of all the possible states of the components would be quite complicated given that the amount of states would be huge. The state enumeration method offers a solution for this problem by selecting the most likely system states. The selection has to be representative for the reliability of the whole system.

Figure 2-2 shows the general form of the method considering 4 transmission lines. In state enumeration, the states of the system are sorted according to the order of failure. The first order states have less impact after failures with a higher probability of occurrence. Therefore, this order is considered first. The higher orders have more impact after failures with a lower probability. Higher orders are considered in the analysis after the lower orders.

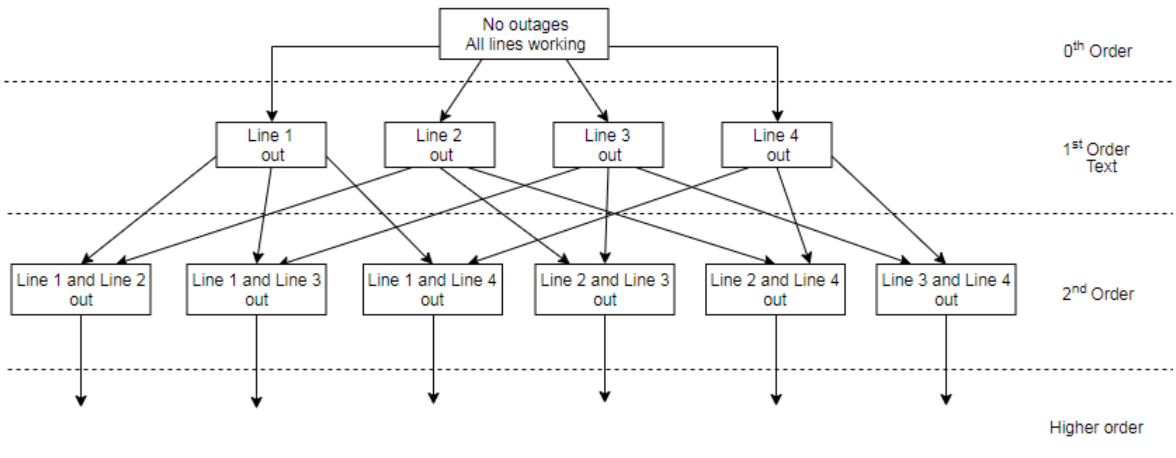


Figure 2-2: General form of state enumeration.

This method makes the assumption that the lowest order states are enough for the representation of the system's reliability.

This method is very similar to a deterministic contingency analysis. The difference is that the probabilities of the states and the effects are combined to calculate probabilistic reliability indicators. One of the most important indicators that can be calculated using this method are those related to the amount, cost, duration and frequency of load disconnection.

The state enumeration of large systems can be implemented with the aid of computer programs. The programs have to include means to calculate failure probabilities of the states, power flow simulations, means to simulate remedial actions (optional) and means to present results.

The generic implementation of state enumeration can be found in the flux diagram of figure 2-3. The algorithm iterates through every contingency and it is described by the following steps:

- **Step 1:** Contingencies are grouped in a contingency list according to their probability (first order, second order)
- **Step 2:** A snapshot of a certain scenario is considered with generation and load values.
- **Step 3:** Run power flow simulation to determine the current state of the system.
- **Step 4:** Check if there are any violations of the operational limits. If there are violations, remedial actions are implemented until the problems are fixed. Typical remedial actions are local re-dispatch and load shedding.
- **Step 5:** Check if all hours are considered. If not, the analysis jumps back to step 2, otherwise it continues.
- **Step 6:** Check if all contingencies are considered. If not, the analysis goes back to step 1. Otherwise it continues.
- **Step 7:** Collect information, calculate reliability indicators and present results.

Failure states are analyzed up to a certain order. Usually, single contingencies, dependent double contingencies and combination of those two are analyzed. An intelligent choice of states can reduce the number of states to be analyzed.

2-4 Estimation of Failure probabilities of transmission lines using historical records and operational conditions with Bayesian data analysis

The determination of failure rates, and repair times of components is an important step in order to perform a reliability analysis. Component's failures generate risks and higher risks results in lower reliability of the power system. Usually, as in TenneT's case, historical failure statistics are used to quantify risks and assess reliability. However, considering only historical failure statistics means that the actual operational conditions are not taken into account. The overall reliability of the system is closely related to the status and the operational conditions of the components. If the failure data of components is based only on historical records, then the impact of the operational conditions of the components is omitted in the risk estimation and the probabilistic reliability analysis might not be accurate. Combining historical failure

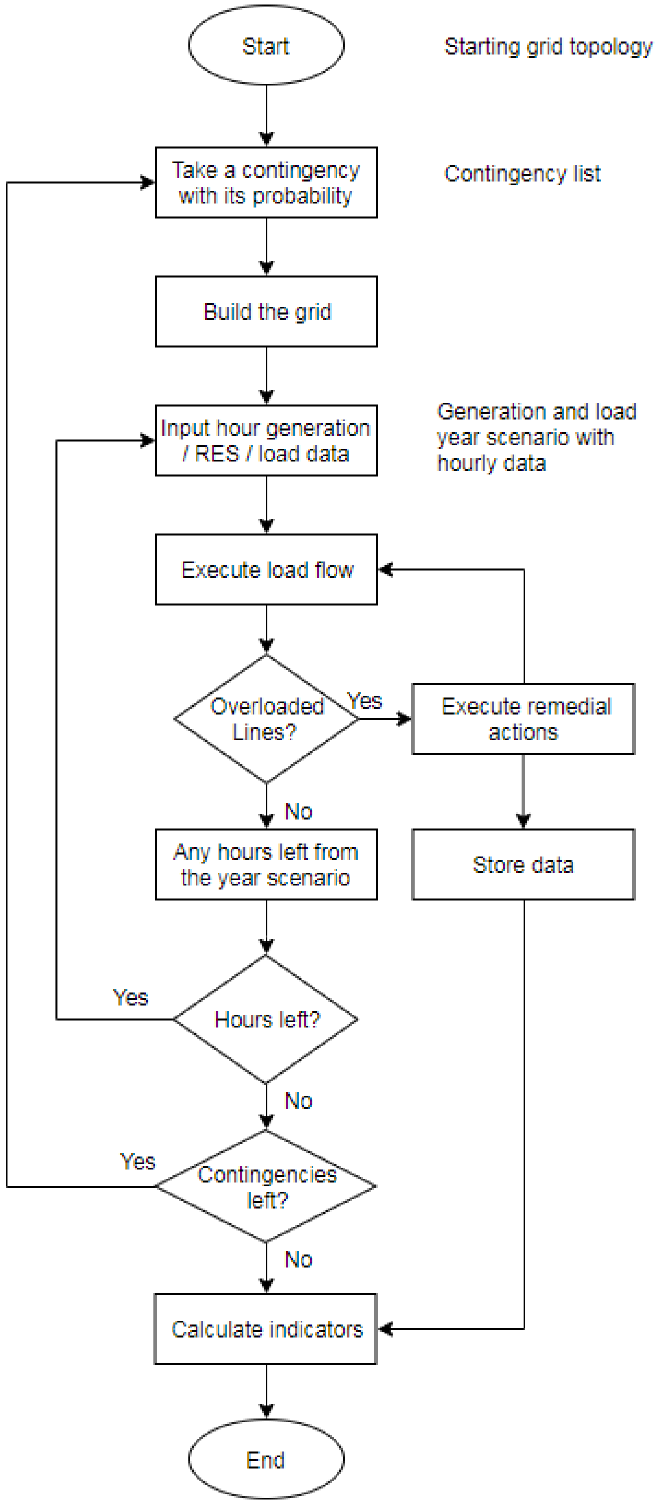


Figure 2-3: State enumeration generic algorithm.

statistics and operation conditions can increase the accuracy of risk assessments in every context of the power system from system operations to long-term planning. References [6] [7] provide detailed background about reliability assessment considering the operational condition of components.

2-4-1 Historical failure statistics of components

Before proceeding, it is important to define some terms regarding failures. These terms are [1]:

- **Failure:** The failure of a component is when the ability of performing a required function is partially or completely terminated.
- **Failure cause:** Is the cause that initiates the failure.
- **Failure mechanism:** Is the physical process of the failure.
- **Failure mode:** The appearance of the failure, what can be seen.
- **Fault:** In a power system a fault is an abnormal condition that interferes with the normal flow of current and affect the ability of the power system to comply with its function. It is important to note that a component failure does not necessarily end in a system fault and also a fault can occur without an element failure.

Historical failure statistics contain the information that can be derived from databases that record historical component failures and test results of the elements of the power system. These statistics are the main input information for a probabilistic reliability analysis. Some of the elements of failure statistics are failure causes, failure frequencies and repair times [1]. An accurate determination of the statistics results in accurate results. Nevertheless, sometimes, it can be quite challenging to collect information of individual components. In the Netherlands, information from failures of the transmission system is recorded in the NESTOR (Nederlandse Storingsregistratie) database. The NESTOR database holds the information of component, failure cause and repair time for every failure.

The activities associated to the recording of information into the NESTOR database are the following: [8]

- If an interruption or malfunction occurs on the national power system, the event is notified to the national control center and an alarm is displayed on the national Energy Management System (EMS).
- The issue is addressed by the operator on duty and the event is registered depending on its nature. The operator takes all the necessary steps for the registration and collects all the relevant information.
- The failure report is later processed in accordance with the NESTOR database procedures. If the information is relevant enough to be included in the NESTOR database, then the manager operator on duty proceeds to register the entry on NESTOR with all the information available regarding the event.

- The manager operator is responsible for handling the situation and solving the issue if necessary. On top of that, the manager is in charge of further investigation, communication with customer and further reporting.
- Over the next days, a investigation team is deployed. Details about grid impact and recovery times are gathered and also registered on the NESTOR database.

2-4-2 Failure frequencies and Repair times

Failure frequency means how often an element fails in a period of time, often a year, with a confidence interval for reference. It can be calculated from historical databases such as NESTOR. The confidence interval gives an indication of the accuracy of the failure frequency and depends on the quantity of data available. The more data collected, the narrower the confidence interval is.

The repair time is the amount of time needed to repair the failure. This repair time is reported per element. For some analyses an average repair time is used. This average time is not always a good indicator when there is not enough data available, because an extremely long repair time will dominate the average. Repair times can vary over a wide range. A component can have a short repair time if the damage is not complex. The same component may have an extremely long reparation time if the damage is very complex or if it needs to be replaced. This variation makes the estimation of the repair times a complicated task that can be solved by using statistical methods such as approximating and fitting the repair time data into distribution functions. [1] The determination of accurate repair times is a challenging task and it is even more challenging if there is not much information available. Expert's opinion gives a good estimate of repair times when there is not enough data.

2-4-3 Life cycle of components

The life cycle of a component consists of the aggregation of the periods when the component is working and the periods when the component is out of service. Figure 2-4 shows a graphical representation of the component life cycle with its parameters. It has to be noted that the figure shows the life cycle of a component that can be repaired.

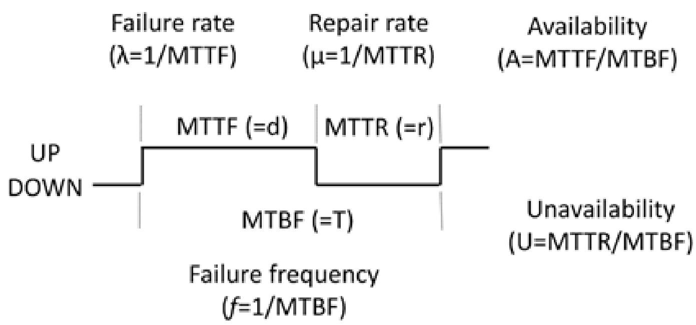


Figure 2-4: Life cycle of components. Adapted from [1]

The parameters of the life cycle are usually expressed in hours or years. These parameters are:

- MTTF (Mean time to fail): It is the mean time before a working component fails.
- MTTR (Mean time to repair): It is the mean time to repair a failed component
- MTBF (Mean time between failures): Is the mean time between failures of the component. The MTBF can be obtained by adding the MTTF and MTTR.
  MTBF = MTTF + MTTR
- Failure rate (λ): Is the rate in which a component fails. It is calculated as: $\lambda = 1\,/\,MTTF$
- Repair rate (μ): Is the rate in which the component is repaired. It is calculated as: $\mu = 1\,/\,MTTR$
- Failure frequency (f): It is the frequency at which a component fails. It is calculated as $f = 1\,/\,MTBF$
- Availability (A): Is the probability to find the component working at a random time It is calculated as: $A = MTTF\,/\,MTBF = f\,/\,\lambda$.
  Also as: $A = MTTF\,/\,(MTTF + MTTR) = (1\,/\,\lambda)/((1\,/\,\lambda) + (1\,/\,\mu)) = \mu\,/\,(\lambda + \mu)$
- Unavailability (U): Is the probability that the component is not working at a certain time. It is calculated as: $A = MTTR\,/\,MTBF = f * MTTR$
  Also as: $U = MTTR\,/\,(MTTF + MTTR) = (1\,/\,\mu)/((1\,/\,\lambda) + (1\,/\,\mu)) = \lambda\,/\,(\lambda + \mu)$

When calculating a parameter from another one, it is important to use the right time units. Sometimes parameters are given in hours and sometimes in years. More detailed information about failure rates of components can be found in reference [9].

2-4-4 Bathtub Curve

The failure of components over time follows a curve that resembles a bathtub as depicted in figure 2-5. The infant stage is a short time after the component has been installed. In this stage, the component is vulnerable to early failures such as construction and installation failures.

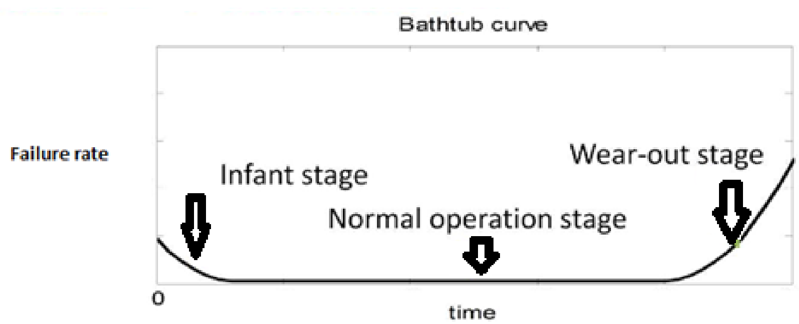


Figure 2-5: Bathtub curve. Adapted from [1]

The failure rate decreases as time passes and once the infant stage is finished, then the components reach the normal operation stage where the failure rate is constant. Failures in this stage are random. The last stage is the wear out stage. In this stage, the lifetime of the component has been reached and the failure rate increases. Failures are more likely to occur in this last stage.

2-4-5 Component failure rate based on operating conditions

The failure of components is mostly related to external factors such as environmental conditions (weather, temperature), operating conditions (current, voltage) and human factors (design, installation, maintenance, operation, accidents). A failure might be caused by one factor or several factors that can or cannot be related with each other. Extreme weather conditions, natural disasters and human errors are time independent and cannot be predicted, other factors such as loading and localized weather and temperature can be time dependent and can be described by a pattern[7]. Operation of the components close to the limits can increase failures in the power system, overloading can increase the failure rate of components and can produce cascade failures that could end in power system faults.

As an example considering transmission lines, operational conditions such as high temperature and overloading can produce heating in the material, this heating can result in temporary expansion and annealing of the material. As the material expands, lines tend to sag and get closer to trees or houses depending on their location and the annealing might result in temporary loss of strength in the material. Another effect that can appear due to operational conditions is dielectric failures. This effect can produce serious failures in components such as transformers.

The variations of the failure rates due to operational conditions imply a variation in the unavailability of components. These variations are time dependant and they are mainly linked to the instantaneous stress, weather and the physical conditions of the components. The total unavailability of a component can be expressed by equation 2-1:

$$U_b + U_t = U_T \tag{2-1}$$

Where U_T is the total failure rate, U_b is the historical unavailability and U_t is the time-operational conditions dependent unavailability.

2-4-6 Estimation of added failure probabilities using Bayesian data analysis

As mentioned before, operational conditions are key factors that can modify the failure behaviour of components. Perhaps the most important of these are weather conditions. Reference [10] states that since many of the components of the power system are highly affected by bad weather conditions, these conditions can affect reliability indicators. If reliability assessments are performed without taken into account weather, the results obtained can be highly optimistic and misleading.

Weather conditions are particularly important in the context of system operations. Reference [11] shows that, if failure probabilities are updated using forecasts of bad weather, then operators can use this information to take better decisions. However, in the context of long term planning, where the time frame of analysis is longer (years and beyond), weather conditions are not the main concern and other operational conditions take more importance such as the age of components and the location. Whatever the context is, it is important to acknowledge that operational conditions play an important role and failure probabilities should be updated.

Bayesian data analysis offers the option to update failure probabilities according to the operational conditions of the components. This section offers an introduction to Bayesian data analysis.

2-4-7 Bayesian Data Analysis (BDA)

BDA is very useful because it allows to solving problems using a very natural way of reasoning, i.e. by collecting evidence and experiences to build a general idea or belief and then updating this belief with further data that comes along. Extensive information about Bayesian analysis can be found in [12].

The prior beliefs can be addressed as priors. BDA can provide a statistical inference of how the prior changes with data. In other words, it allows the use of probability to represent the uncertainty of a statistical model. Bayesian analysis is based on Bayes theorem (equation 2-2) which states that the probability of A given B is equal to the probability of B given A times the probability of A divided by the probability of B.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{2-2}$$

Bayes theorem allows us to update our priors based on new information and it is really helpful when there is not much data available. The result obtained from the application of Bayes theorem is called the posterior and it contains information of both the prior and the collected data. The posterior result can be used as a new prior to be further updated with more new information. This represents an advantage because it allows integrating more and more data. Another representation of the Bayes theorem can be seen in equations 2-3 and 2-4. The posterior is equal to the likelihood times the prior and divided by a normalization factor.

$$P(A|B) = P(A) * \frac{P(B|A)}{P(B)} \tag{2-3}$$

$$Posterior = Prior * \frac{Likelihood}{Normalization} \tag{2-4}$$

Priors can be subjective and then adjusted to reality with real data. If someone does not agree with a certain prior, BDA allows to replace it in order to find a different posterior.

If there is limited data available (small quantity), the posterior will rely more on the prior. In contrast, if there is plenty of data available, the prior will not influence much the posterior. The posterior will rely on the actual data meaning that the result will be based on evidence and not on prior beliefs. Bayesian is widely used because it allows to integrate expert's opinion in calculations through the priors taking advantage of all the information at hand either from experience or real data.

There are two types of prior distributions for BDA. These are:

- Non-informative prior: does not contain any previous information
- Informative prior: contains previous information such as expert's opinion.

If the prior is not informative, then the posterior and the likelihood are the same as shown in figure 2-6. This means that the posterior is the same as the maximum likelihood.

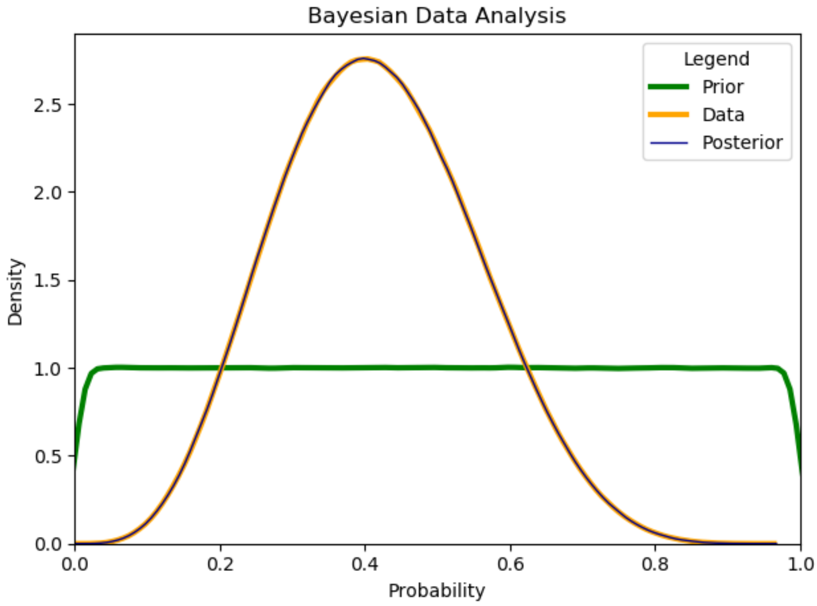


Figure 2-6: Non-informative prior

If the prior is informative and the number of samples is small, then the posterior resembles the prior because it is more influenced by it than by the data. This can be seen in figure 2-7.

If the sample size increases, the actual data will influence more the posterior than the prior as can be seen in figure 2-8.

The posterior distribution has the potential to be precise even if we have a small sample by using a good quality informative prior that can be obtained by expert's opinion. Bayesian analysis produces the result as a form of a distribution function that be used for later analysis. This distribution function represents the uncertainty on the statistical model implemented.

Bayesian data analysis is very useful for complex problems with large uncertainties and several sources of information that require to be integrated in one solution. On top of that, since it allows the integration of prior knowledge, it is tremendously useful when there is a small number of samples of real data available. The posterior offers as the result, the probability of an event and how uncertain that probability is.

Bayesian data analysis can be implemented with a number of methods. For this project, the method approximate Bayesian calculation (ABC) is the one chosen.

2-4-8 Approximate Bayesian calculation (ABC)

ABC offers an approach to implement BDA for simulation-based models [13]. In order to get the posterior distribution, the Bayes theorem of equation 2-2 has to solved. For this matter,

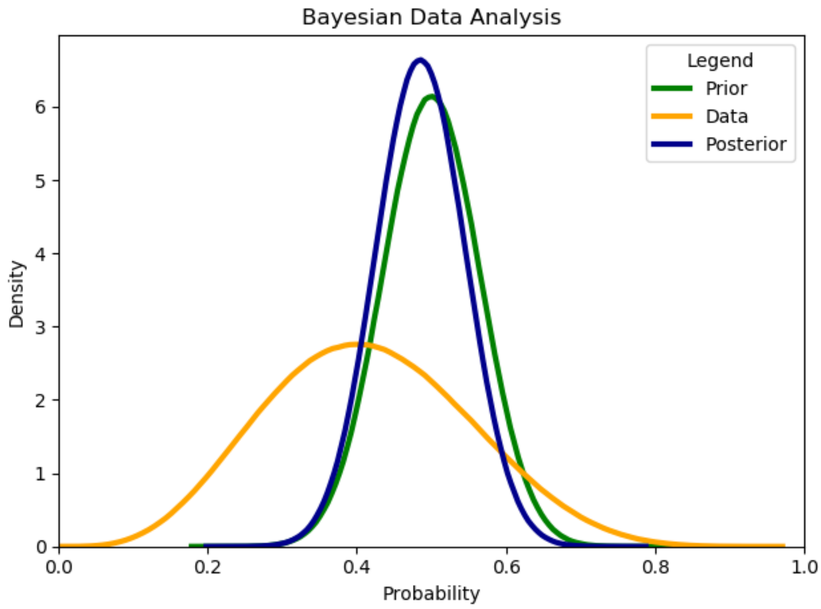


Figure 2-7: Informative prior with small number of samples

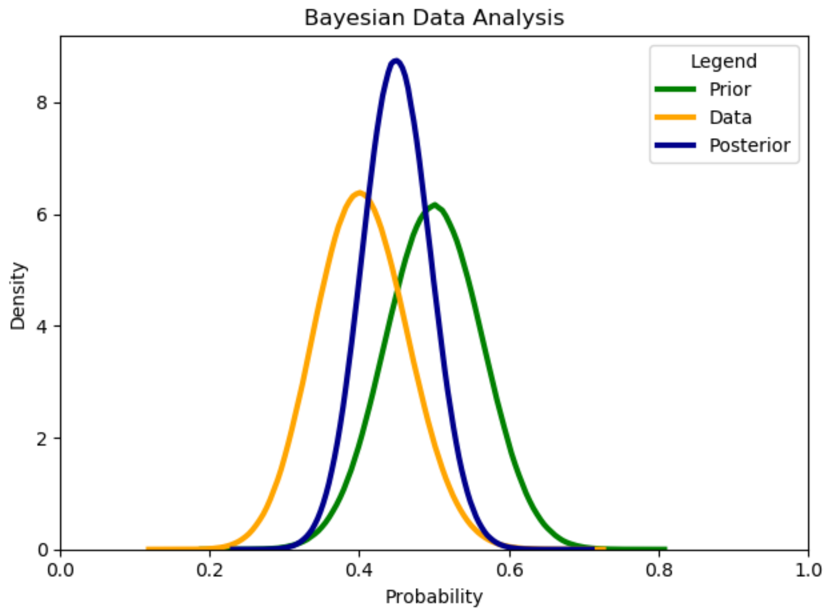


Figure 2-8: Informative prior with a larger number of samples

a prior distribution and the likelihood of the data are needed. In BDA, the parameters of distribution functions (i.e probability) and the data itself are considered as random variables.

Taking θ as the probability, the prior distribution would be written as $f(\theta)$ and the likelihood as $L(\theta|X)$ where X is the data. Rewriting Bayes theorem with these terms, equation 2-5 is obtained:

$$f(\theta|X) = \frac{L(X|\theta) * f(\theta)}{\int (L(\theta|X) * f(\theta))d\theta} \tag{2-5}$$

Where $f(\theta|X)$ is the posterior and $\int (L(\theta|X) * f(\theta))d\theta$ is the normalization factor introduced in equation 2-4.

Finding the posterior by solving analytically the Bayes theorem can be a complicated task. The difficulty lies in the integral of the denominator of equation 2-5 which can be unmanageable for some cases. In spite of this difficulty, the integral is only a complicated normalization factor which means that the posterior is proportional to the prior times the likelihood. ABC takes advantage of this proportionality and simplifies the determination of the posterior. Since the prior is already defined, the task that is left is to compute the likelihood.

ABC replaces the calculation of the likelihood with simulations. This means that artificial data is simulated by a generative model and then compared with the actual data. If the simulated data matches or is close enough to the actual observed data, then the parameters are kept. The ABC generic algorithm is the following:

- **Step 1:** A parameter θ is sampled from the prior distribution.
- **Step 2:** The parameter sampled is then used for the simulation of the artificial data using a generative model.
- **Step 3:** The simulated data is compared to the observed data. The distance between them is evaluated. If the distance is below a predefined tolerance, then the data is close enough and the parameter is stored. Otherwise the parameter is discarded.
- **Step 4:** The stored parameters are used for the approximation of the posterior. The posterior is the distribution function of the sampled parameters conditioned by the actual data.

ABC does not interfere with data collection. The parameters are treated as random quantities and the posterior is based on the PDF of the prior and the observed data. Complete information about ABC can be found in [13, 14].

Once the posterior is obtained, the variation on the probability can be studied. Credible intervals for the probability can be computed which are equivalent to the confidence intervals that can be obtained in classical statistics. If a single number for the probability is required, it can be obtained from the tendency analysis and indicators such as the mode, mean and median of the posterior distribution.

Alternative calculation: It is important to choose priors and generative models which are convenient from a mathematical point of view. This means to select PDF's and generative models that are easy to manipulate in order to obtain the posterior distribution. For this project, the beta function,described in appendix A, is chosen for priors given that its form can be easily manipulated by the proper selection of α and β. For the generative model, the binomial distribution (appendix A) is chosen. The two functions are mathematically convenient. The posterior that results from this selection is a conjugate prior which means that it is also a beta distribution. The advantage of choosing the distributions in this manner is presented with the next example:

Considering the beta prior of equation 2-6 for the failure probability of transmission lines. The parameter α_1 is the number of failures and the parameter β_1 is the total observations minus the number of failures.

$$Prior = Beta(\alpha_1, \beta_1) \tag{2-6}$$

The likelihood is obtained with the binomial distribution of equation 2-7 as the generative model for the simulation of the data. Where f is the number of failures and n the number of observations.

$$Likelihood = Binomial(f, n) \tag{2-7}$$

Since the posterior is the conjugate prior, its distribution function is the beta function of equation 2-8. And its parameters can be found by applying ABC.

$$Posterior = Beta(\alpha_2, \beta_2) \tag{2-8}$$

After applying ABC to find the posterior distribution, it can be seen that the parameters α_2 and β_2 correspond to equations 2-9 and 2-10 respectively:

$$\alpha_2 = f + \alpha_1 \tag{2-9}$$

$$\beta_2 = n - f + \beta_1 \tag{2-10}$$

The estimation of the posterior is simplified greatly by chosing the beta and binomial distribution functions.

Now that the methods for the estimation of failure probabilities have been introduced, their application for the estimation of failure probabilities of transmission lines will be presented in the next chapter.

2-5 Effect determination using linear programming techniques for re-dispatch options

Effect (impact) determination refers to the quantification of lines overloading in terms of money. In other words, the aim is to determine the associated cost linked to a determined dispatch scenario and contingencies to ultimately calculate risk. For this report, this cost will be in terms of redispatch cost, lost load cost and the addition of both. Before going any further, it is important to properly define dispatch scenario and redispatch:

- **Dispatch scenario:** Refers to the creation of a operational schedule for generation. This is the allocation of production levels for conventional generation plants and RES. Economical dispatch refers to scheduling generation in a way that the loads can be served with the least possible operating costs so there can be maximum profit for generation parties without the violation of any system constraints. Economic dispatch accounts for different fuel prices, availability of generating units and market behavior. Dispatch of RES is determined using weather forecasts to determine plant availability. Dispatch scenarios can also be determined with different objectives than the economical such as minimizing environmental impact. These other objectives are not in the scope of this report.

- **Re-dispatch** If a dispatch scenario is handed to the TSO, grid operators analyze it by means of load flow calculations. If the transmission grid is negatively affected by the dispatch scenario, then the scheduling can be altered by shifting generation in order to stabilize the system. This shift in the generation levels is the re-dispatch action.

2-5-1 Preliminary aspects

Power systems are complex structures that at some point can become unpredictable. Situations such as failure of components (contingencies), sudden variation of RES (mainly wind power plant WPP and solar energy PV) generation and sudden variation of loads can change the energy flow on lines leading to overloadings.

TSO's are responsible for the secure operation of the transmission system while offering the energy market enough capacity to operate normally. Part of the TSO duties is to be always prepared to solve any kind of power system constraints that may arise in the operations. In the case of transmission lines, the constraints that may appear are in the form of overloaded lines. Line overloading refers to the particular situation where the line has reached its maximum limits for safe operation which implies that any power transaction that involves the overloaded line can not be implemented and the market can be hindered.

In the case of the Dutch grid, TenneT has the task to ensure that the operations are secure [15] by keeping the components working below the maximum limits. TenneT is required by law to operate the power grid in a N-1 secure way. On top of that, the Dutch electricity market is a liberalized market. This market does not take into account the capacity constraints of the transmission grid [16]. Some of the energy transactions may overload lines. As a result, in order to be N-1 secure, re-dispatch actions can be required in order to keep the grid secure and comply with the law.

In TenneT, the system operations department is responsible for the secure operation of the grid. Operators run simulations over a dispatch scenario corresponding to the next 24 hours considering all N-1 contingencies. If an overloading is foreseen, then TenneT makes sure that an action is performed so the system can be compliant at every moment. This kind of actions are preventive actions because they alter the dispatch scenario on the base of what might happen in the next few hours. This kind of situation is not very uncommon especially because of the variation of RES generation due to a difference between forecasted weather conditions and actual conditions.

Almost all overloadings in the power systems are solved by applying any kind of actions. This in combination with the robustness of the Dutch transmission grid makes it very unlikely for a load shedding action to be implemented.

For this report the preventive actions will be modelled as re-dispatch actions. The price of re-dispatch will be assumed to be 90 euro per megawatt for either up and down re-dispatch actions.

In long term planning, unlike systems operations, preventive actions are not taken into account for the calculations. If an overloading condition appears product of a expected dispatch scenario, then the assumption of overload equals lost load is accepted and further risk analysis is performed. However, as it was mentioned before, it is very likely for an overloading condition to be fixed by preventive actions. Not including those actions in long term planning implies that the analysis is not completely real and the results might be too pessimistic leading to a possibly oversized transmission grid. By including preventive actions in long term planning calculations, the results will potentially be more realistic and the risk analysis more precise.

In this report, preventive actions will be included as re-dispatch in long term planing calculations over a test grid with a 2025 scenario with large penetration of RES that will be introduced in the further chapters. The re-dispatch actions will be modelled using linear programming techniques. If a overloading problem can not be solved by executing re-dispatch, then the assumption of overload equals lost load will be accepted after all.

2-5-2 Re-dispatch actions using linear programming techniques

A redispatch action is supposed to alleviate any overloading condition presented on the grid. For these matter, distribution factors are used to calculate the maximum variation of flow that is acceptable for a line.

The distribution factors used for the calculations are the Power Transfer Distribution Factors (PTDFs). This distribution factor denotes the relative change in the power flow on a specific line due to a injection of power on one bus and the withdrawal of the injected power on another bus [17]. So, basically the PTDF aids in the determination of how much power from a particular generation unit flows through a particular line. The PTDFs are topology dependent. An example of a PTDF matrix for a defined topology is shown in figure 2-9. The value $PTDF_{L1,G1}$ is the distribution factor on line 1 due to generator 1, the value $PTDF_{L1,G2}$ is the distribution factor on line 1 due to generator 2 and so on. If the topology changes, so does the PTDF matrix. There will be as many PTDF matrices as contingencies.

$$
\begin{array}{cc}
 & \begin{array}{ccccc} \quad L1 \quad & \quad L2. \quad & & & \quad Ln \quad \end{array} \\
\begin{array}{c} G1 \\ G2 \\ : \\ : \\ Gn \end{array} &
\begin{pmatrix}
PTDF_{L1,G1} & PTDF_{L2,G1} & .. & .. & PTDF_{Ln,G1} \\
PTDF_{L1,G2} & PTDF_{L2,G2} & .. & .. & PTDF_{Ln,G2} \\
: & : & .. & .. & .. \\
: & : & .. & .. & .. \\
PTDF_{L1,GN} & : & .. & .. & PTDF_{Ln,Gn}
\end{pmatrix}
\end{array}
$$

Figure 2-9: PTDF matrix

Linear programming

Linear programming (LP) is an optimization technique that deals with maximizing or minimizing a linear objective function subjected to constraints [18, 19]. LP problems are presented in the following form:

Maximize or minimize:

$$Objective = c^T * x$$

Subjected to:

$$Ax \leq, =, \geq b$$

With:

$$x \geq 0$$

Where x is the vector of the control variables to be determined (generation dispatch), c^T is a transposed known vector, usually known as the cost vector (cost per generator), b is a known vector with limit quantities (maximum flow of lines) and A is a know matrix of coefficients (PTDFs).

This technique can be applied on large systems and it is flexible, fast and reliable [20]. LP is widely used for security constraint re-dispatch and optimal power flow calculations.

The proposed LP problem with its objective function and constraints can be seen in the following equations:

The objective function of equation 2-11 aims to minimize the quantity of conventional generation dispatch per unit (Xgk) needed to supply the demand in case of a line overloading.

$$\Delta D = \sum_{k=1}^{ng} Xgk \tag{2-11}$$

The constraints are the following:

Power balance (equation 2-12): generation and load have to be in balance at all times.

$$\sum_{k=1}^{ng} \Delta Pgk = 0 \tag{2-12}$$

Maximum rating of lines (equation 2-13): The flow in every line has to be less or equal to the maximum load of the line.

$$\sum_{k=1}^{ng} PTDF_{gk,Li} * \Delta Pgk \leq Flmax_{Li} - Fl_{Li} \tag{2-13}$$

Maximum rating of generators (equation 2-14): The generators cannot deliver more energy than its maximum.

$$Pgk \leq Pgk_{max} \tag{2-14}$$

Where:

- Ng = number of generators
- Xgk = dispatch of generator k
- Pgk = variation of dispatch in generator k
- PTDF = power transfer distribution factor in line i due to generator k

- FLmax = maximum flow on line i
- FL = current flow on line i
- Pgk = actual dispatch on generator k

Solving LP problems: Simplex and Revised simplex methods

Simplex method: The Simplex method is used to solve LP problems such as the one presented above. Basically the Simplex is a search method where a set of solutions is tested one at the time until the optimal solution is found [21]. The graphical representation of figure 2-10 helps to illustrate the method. The variables from the objective function are bounded by the constraints represented by the black lines. This boundaries create the gray feasible area, within this area remain the feasible solutions for the LP problem. The vertices of the areas, depicted with red dots, are basic solutions for the problem and at least one of this basic solutions is an optimal solution for the problem [18, 22]. The feasible area becomes a polytope for LP problems with many variables. The method consists in testing and comparing every one of the vertices of the feasible area until an optimal solution is found.

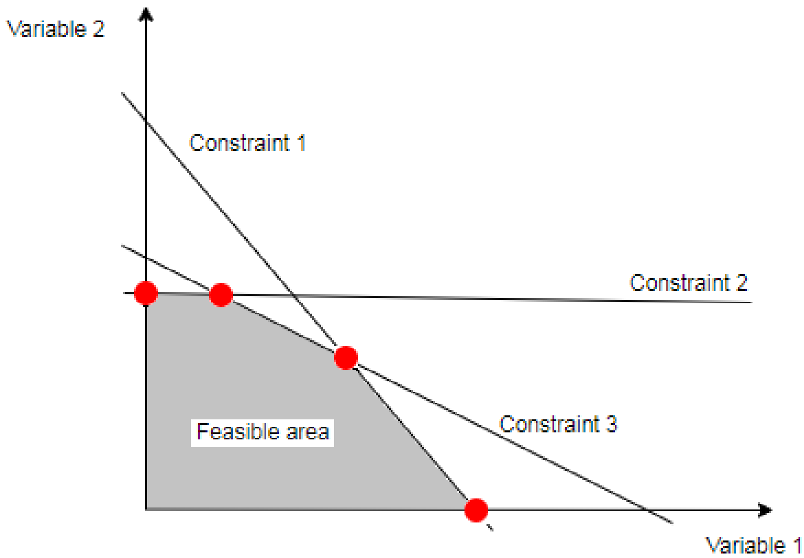


Figure 2-10: Graphical representation. Simplex method

Mathematically, in order to solve the LP problem by the Simplex method, the problem must be in the canonical form represented below:

Maximize or minimize:

$$f = c^T * x$$

Subjected to:

$$Ax = b$$

With:

$$x \geq 0$$

Where the variables of the vector x have to be positive and the inequalities have to be transformed in equalities. In order to do so, slack variables are introduced [18, 23].

LP problems have basic and not basic solutions. The basic solutions are found by setting one variable at the time to zero and solving $Ax = b$ [24]. The basic variables are those that are positive on the solution and the not basic variables are zero. The LP problem can be expressed as shown in equation 2-15 where A_N is the not basis matrix, A_B is the basis matrix, c is the cost vector, f is the objective function, x and x_s are the variables and slack variables and b is a known vector with limit quantities.

$$\begin{bmatrix} 0 & A_N & A_B \\ 1 & -c & 0 \end{bmatrix} \begin{bmatrix} f \\ x \\ x_s \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{2-15}$$

Once that the LP problem is on the canonical form, the simplex tableau of equation 2-16 can be established. This tableau allows to compute the solution by moving from one solution to another one [18, 25].

$$\begin{bmatrix} 0 & A_N & A_B & b \\ 1 & -c & 0 & 0 \end{bmatrix} \tag{2-16}$$

The solution is found by "pivoting" in the table. Pivoting is the operation used to test the solutions, this is to test the vertices of figure 2-10. Pivoting allows to move from one vertex to another by changing the set of basic variables using row operations [18]. Every solution is compared against each other. The best solution is chosen in accordance to what is needed either minimizing or maximizing the objective function.

Revised simplex method: Solving the Simplex method requires to store and update the full tableau. This represents a problem for big systems. The table can be so big that it might become impossible to manage. The Revised simplex method does not deal with the complete tableau. This method performs calculations and stores only information that is relevant to the current calculation employed to test and compare the solutions from the vertices [26].

Instead of dealing with the tableau, this method computes A_B^{-1} and the solution can be found with equation 2-17 [24].

$$x = A_B^{-1} b - A_B^{-1} A_N X_s \tag{2-17}$$

The Revised simplex is equivalent to the Simplex, the difference lies only in the implementation and the performance with dealing with large systems.

2-6 Clustering techniques for dispatch scenarios

Power system analysis, in the context of long term planning, can require of a great number of calculations for several scenarios over long periods of time (usually years). A large number of calculations implies long calculation times and the need of great computational power.

As an example, for a N-1 contingency analysis of transmission lines in a power grid of 100 lines (about the size of the 380/220 kV grid of the Netherlands) in a time frame of one year with an hourly dispatch scenario of generation and load (8760 hours), the number of calculations needed would be the number of hours times the number of lines resulting in 876,000 load flow calculations. If every load flow calculation would take one second to be completed, the total calculation time would be around 10 days. A more complex analysis would take even more time.

An option to reduce the number of calculations is to reduce the number of hours to be analyzed by means of clustering. The idea is to create clusters of similar dispatch hours and then analyze the centroid of every resulting cluster. Only the centroids will be analyzed and not every hour resulting in fewer calculations and off course less calculation time. The problem with this clustering alternative is that the accuracy of results will be reduced. A compromise between time saving and accuracy has to be established. The less time employed for the analysis the more errors and vice versa.

2-6-1 Clustering analysis

Clustering analysis refers to the statistical procedures designed to classify data sets. The goal of clustering is to organize data objects and group them into clusters in a way that the elements of one cluster share more properties with each other and less with elements of other clusters. This is maximize intra-class similarity and minimize inter-class similarity [27][28]. A graphical representation of clustering can be seen in figure 2-11. The group of data to the left is the base data. The clustered data can be found at the right side of the figure. An extensive description of clustering analysis is presented in [29].

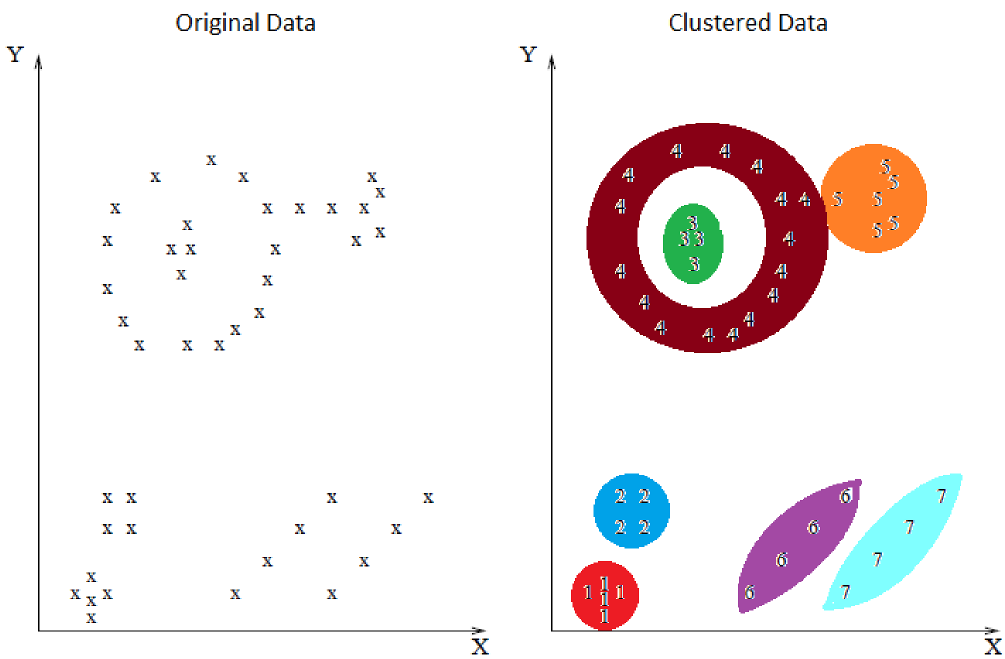


Figure 2-11: Clustering analysis

There are five main types of clustering techniques, they are:

- Partitioning techniques: Several partitions are built to be later evaluated by some criterion.
- Hierarchy techniques: The set of data undergoes a hierarchical decomposition following some criterion.
- Density based: They are based on connectivity and density functions.
- Grid based: Based on a granularity structure of multiple levels.
- Model based: A model is created for every cluster, then the models are fitted to each other.

This report will focus on Partitioning (K-means) and Hierarchy techniques.

Partitioning techniques

These type of techniques require the desired number of clusters k to be specified beforehand. The methodology is iterative. First, starting from the original data set, the elements are organized into k clusters where the centroids are chosen randomly. Then, the distance between the centroid to the elements is measured. This distance is minimized according to the grouping criterion and the centroids are updated. The distances are measured again and the process iterates until there are not changes in the clusters or the number of iterations is reached [30]. The partitioning algorithm considered in this report is the k-means clustering algorithm.

K-means clustering: Every cluster is represented by it's centroid. This algorithm works in the following steps:

- Step 1: With the number of clusters k already defined. The centroids are assigned randomly (centroids = k).
- Step 2: All the data points are assigned to a cluster depending on the distance between them and the centroid.
- Step 3: New centroids are calculated from the previous clusters.
- Step 4: The data points are reassigned according to the smallest distance between data point and centroid in order to create new clusters.
- Step 5: The centroids of these new clusters are recalculated.
- Step 6: Steps 4 and 5 are repeated until there are not more changes in the clusters or until a predefined number of iterations is achieved.

A graphical representation of k-means clustering can be seen in figure 2-12. From left to right, it is possible to see the base data, the clustered data and the centroids marked with the black dots.

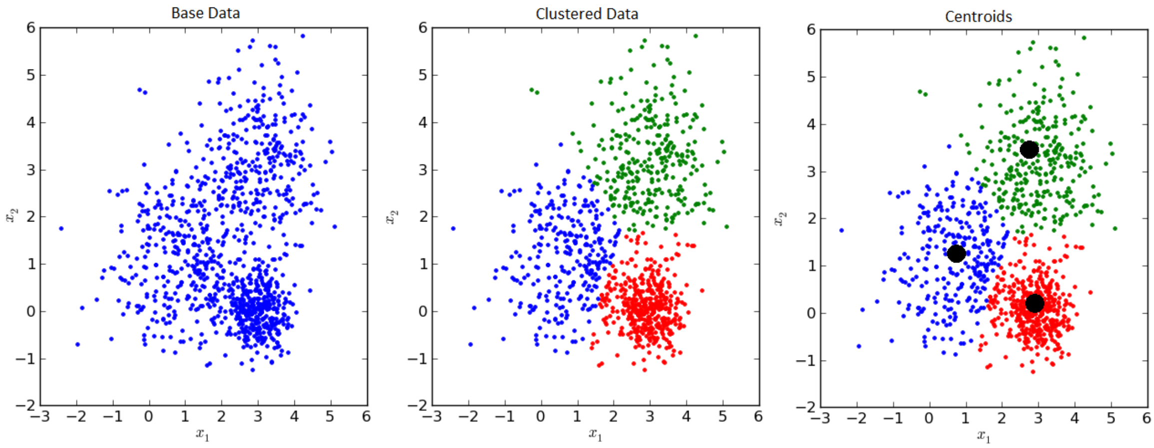


Figure 2-12: K-means clustering

Hierarchy techniques

Also known as connectivity techniques. It is based on the precept that nearby objects have more in common that objects far away. These algorithms build a hierarchy of clusters. The objects are connected to form clusters based on their distance. At different distances, a different number of clusters will be formed. These techniques are divided into agglomerative and divisive methods depending on how the hierarchy is formed [28] [30].

Agglomerative hierarchical clustering uses a bottom up strategy. First, all the objects from the data set form their own cluster. This clusters are later merged into larger objects until all objects form a single cluster or until a maximum distance is achieved.

Divisive hierarchical clustering follows a top down strategy. As opposite to agglomerative, this algorithm starts with all objects in one single cluster. This cluster is divided into smaller and smaller pieces until every element forms its own cluster or until a maximum distance is achieved.

The hierarchical clustering algorithm can alternatively be described using a tree structure called dendrogram as the one shown in figure 2-13. The algorithm works in the following way:

- Every data point forms its own cluster.
- The closest two clusters form a single cluster as can be seen with the 25 data points of figure 2-13. The height where two clusters are joined is the distance between the clusters in the original data set
- This procedure keeps repeating until only one single cluster is formed. The selection of the number of clusters depends on the maximum distance accepted between elements of the data set.

As a difference between hierarchical and k-means clustering, it can be noted that in hierarchical clustering, it is not required to know the number of clusters beforehand. The number

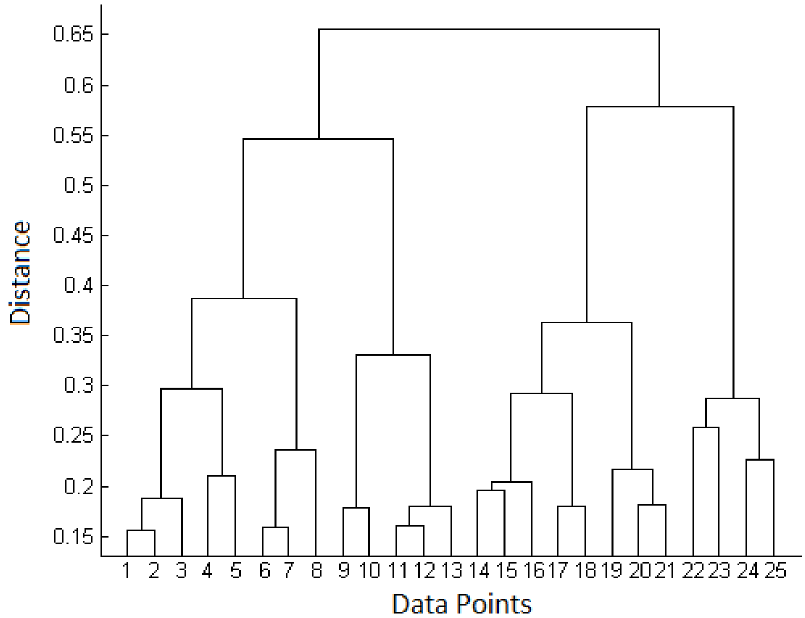


Figure 2-13: Dendrogram, hierarchical clustering

of clusters will be depending on the maximum distance accepted. Also, the results of hierarchical clustering are reproducible, in the case of k-means this is not always true since the first centroids are assigned randomly.

The application of the theory presented in this chapter will be presented next in chapter 3.

Chapter 3

Application of methods

This chapter describes the application of the methods introduced in the previous chapter for the implementation of the different stages of the current project. This chapter is organized as follows: the first section provide explanation about the development of the test grid model and the base scenario that are used through the course of the project. The second section talks about the application of Bayesian data analysis. The third section, describes the application of the methods for effect determination and the last section boards the application of clustering methods. All the algorithms presented were implemented using the programming language Python.

3-1 Development of test model

This section introduces the grid model that is used to test the application of the methods described in the previous chapter. This test model is a replica of the 380/220 kV transmission grid of the Netherlands planned for the year 2025. It is important to keep in mind that this test grid is not the actual grid, it serves only for demonstration purposes. The results that might come from this grid are not complete indicators of what is actually happening or might happen on the future on the Dutch transmission network. The whole transmission grid is not utilized because it is far to complex for the purposes of the present project. A referential image of part of test grid model can be found in figure 3-1.

The test grid model is composed of the following elements:

- 49 Extra High Voltage (EHV) buses 380/220 kV.
- 19 conventional power plants
- 30 wind power plants (WPP)
- 5 solar (PV) plants

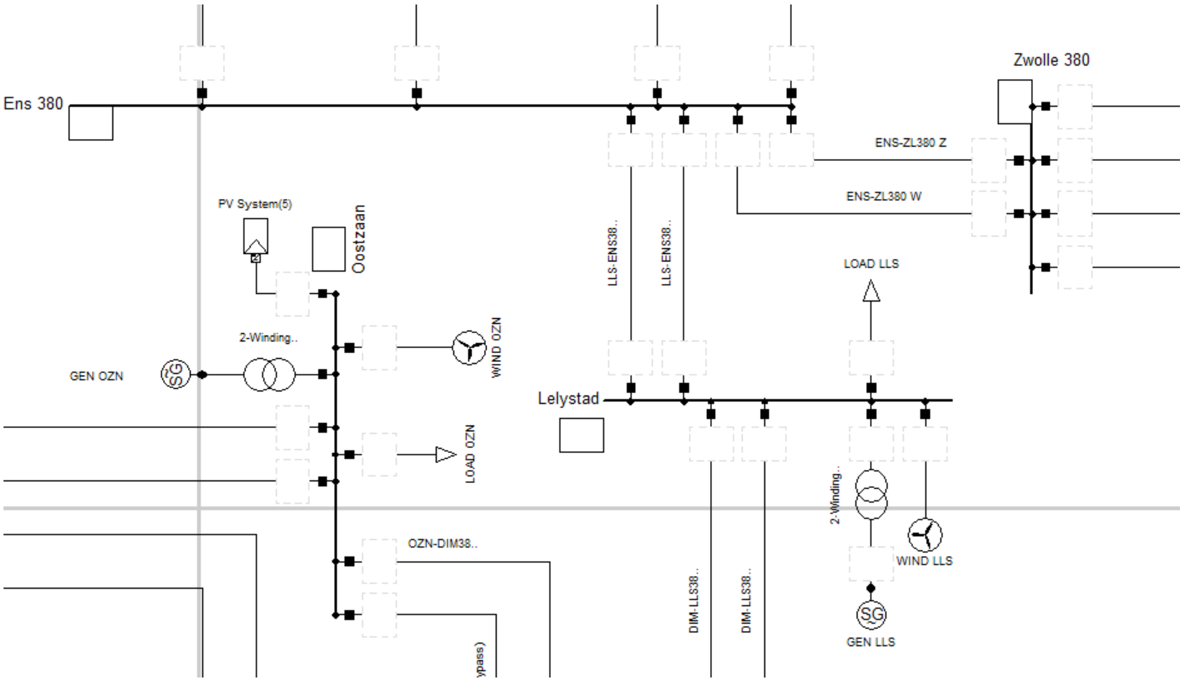


Figure 3-1: Referential image of the test model grid

- 19 2-winding transformers
- 15 3-winding transformers
- 100 single circuit transmission lines
- 40 general loads

3-1-1 General considerations

The test grid is a simplification of the complete Dutch transmission grid planned by TenneT for the year 2025. This test grid is implemented using the power system analysis software PowerFactory 2018 from DIgSILENT. Detailed information about PowerFactory can be found in [31, 32]. The considerations and simplifications taken for test grid are the following:

Simplifications of underlying grids: The Dutch transmission grid is formed by four voltage level grids (110, 150, 220, 380 kV) connected to each other. The test model considers only the 220/380 kV part of the transmission grid ignoring the 110/150kV grid. This simplification implies that all the generation is distributed to the loads only through the EHV grid and that the path from one EHV bus to another is only through the existing lines connecting them in the model.

Generation: For the test model, all generation from any source is connected directly to the 220/380kV grid. The instant dispatch of generators and WPP's along with its maximum limits are obtained by adding all the generation connected to a 220/380kV bus on the original grid model. The conventional power plants and WPP's are modelled from existing models. The PV plants are fictitious and they are used to represent an increase in generation from renewable sources.

Loads: The loads are treated in the same way that the generators. The loads are connected directly to the EHV grid and their instant value is the aggregation of all loads connected to the corresponding node on the original network.

Buses: From the existing 49 buses, 13 are on the 220kV level and 36 on the 380kV level. The location of the buses and the names are the same as in the original grid model.

Interconnections: The interconnections are represented as a load that takes either positive or negative values depending if there is power flowing to or from the interconnection. Interconnections are not taken into account for any kind of actions because they are not a modelled in a way that their behaviour can be represented.

Transmission lines: From the 100 transmission lines, 86 are taken into account for contingencies. The other 14 lines are for the interconnections and it is assumed that they are always working. The ratings and the length of the lines are the same as in the original grid model.

3-1-2 Test scenario

The test scenario consist in the dispatch values for generation and expected loads. The considerations for the test scenario are the following:

- The base for the test scenario is the dispatch and load predictions for the year 2025 provided by TenneT. The 2025 dispatch is given for the entire year with an hourly resolution with individual values for conventional generators, wind units, interconnections and loads. Information about the base scenario can be found in tables 3-1, 3-2 and 3-3.

Generation	Capacity	Maximum	Minimum	Average
Conventional	13,35 GW	13,25 GW	1,29 GW	6,37 GW
RES	22,2 GW	17,38 GW	1,76 GW	6,79 GW

Table 3-1: Generation. Base scenario

Load	Maximum	Minimum	Average
Amount	18,17 GW	9,05 GW	13,06 GW

Table 3-2: Load. Base scenario

Interconnections	Maximum	Minimum	Average
Input	7,05 GW	0.87MW	1,85 GW
Output	9,37 GW	1,52 MW	2,28 GW

Table 3-3: Interconnections. Base scenario

The base scenario has more power from RES than conventional sources. For the interconnections, input means that power enters to the test grid and is modelled as a negative load. Output means that the power is exiting the grid and is represented by a positive load.

- The test scenario is one with a high penetration of PV units. For this study, 6 GW of additional PV power is introduced in different parts of the grid on top of the base dispatch. The additional PV power is distributed over the following buses:
  - 2 GW Eemshaven Oudeschip 380
  - 1 GW Eemshaven 220
  - 1 GW Oude Haske 220
  - 1 GW Oostzaan380
  - 1 GW Meeden 380

  To keep power balance, 6 GW are also added to the loads. The added load is distributed by equally scaling the original loads.
- PV plants are assumed to be always active and delivering a constant amount of power.

The test scenario is fictitious. Nevertheless, the whole scenario was checked with TenneT's experts for consistency and to assure that it is not an absurd scenario.

3-2 Application of Bayesian data analysis for failure probabilities of transmission lines.

This section presents the application of the theory and methods presented in the previous chapter regarding the estimation of failure probabilities (Unavailability / availability) of transmission lines.

The estimation is performed as shown in figure 3-2. First, the values from historical data bases are used as base values. These base values are then combined with the added unavailability / availability that comes from operational conditions using a tree analysis.

The tree analysis structure can be seen in figure 3-3. It is important to select the factors to be included. It has to be kept in mind that the more factors included, the more complicated and large the tree analysis become. Every instance of the tree analysis can be represented as a "weight" that is a factor that multiplies the base unavailability. Figure 3-3 provides an example of the procedure. The following factors are assumed for demonstration purposes, actual factors can be obtained applying BDA as will be illustrated in the next section:

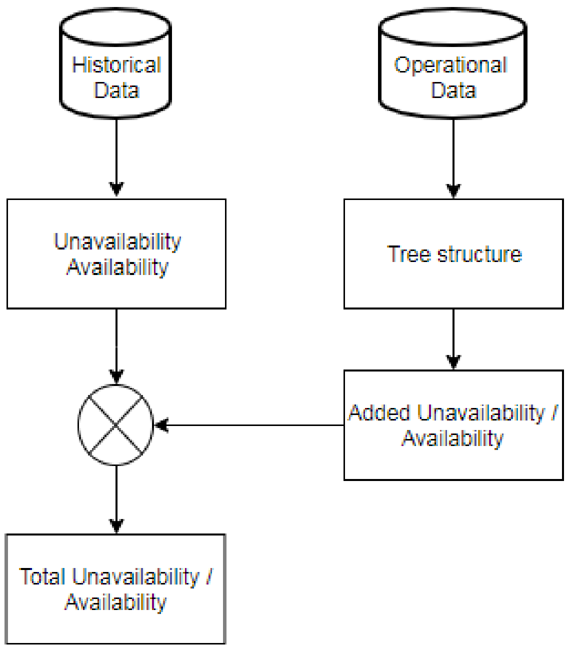


Figure 3-2: Determination of the total probability.

- Operation of the component after its lifetime is reached increases 75% the unavailability.
- Overloaded operation of the component increases 60% the unavailability.
- Operation of the component when it is snowing increase 20% the unavailability.

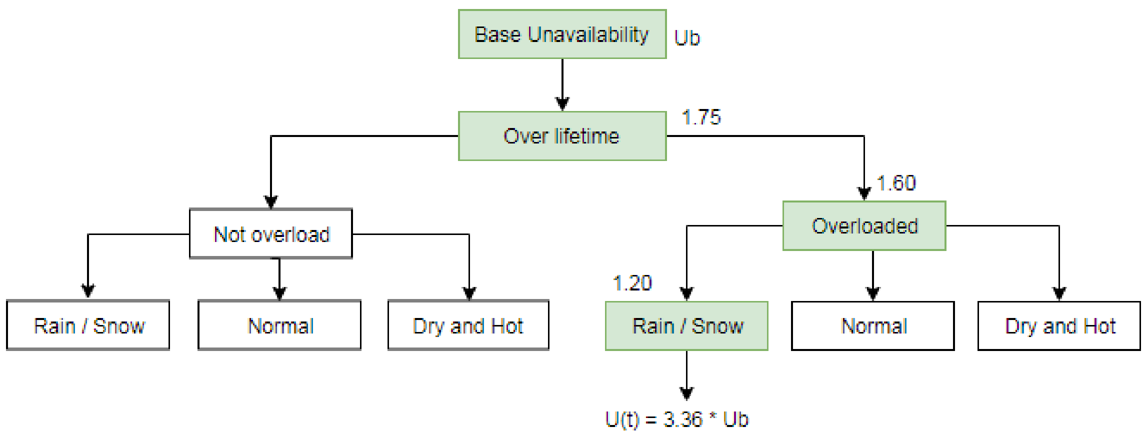


Figure 3-3: Tree structure.

The added unavailability for a component that is working under these conditions (over lifetime, overloaded, snow) is calculated using the tree structure. The resulting unavailability is:

$$U_T = 3.36 * U_b \tag{3-1}$$

The resulting unavailability is the combination of the base unavailability calculated from historical failure data and the added unavailability that comes from the particular conditions at the moment of the analysis. The added unavailability varies every time that the conditions vary. If there is an actual failure, the data will have to be fed back in order to update historical records.

3-3 Estimation of unavailability of overhead lines due to certain operational conditions using Bayesian data analysis

Using BDA, it is possible to consider any operational condition. However, not every operational condition is interesting for all contexts of the power system. To illustrate this, lets consider a momentary storm with heavy winds. Heavy wind conditions will momentarily increase the unavailability of lines on the affected area as long as the storm lasts. Once the heavy wind conditions are gone, the unavailability of lines will reduce to the normal values. In the context of system operations, where the considered time frame for analysis is short enough to include the variation of unavailability, these momentarily variations are important and analyzing them can result in better decisions that can improve overall reliability. On the other hand, in the context of long term planning where the analysis time frame is much longer, momentary conditions are not very interesting. The momentary variation of unavailability due to the storm is not entirely useful for long term planning.

Operational conditions that produce a more permanent variation on unavailability can be useful for long term planning. Conditions such as location and age can provide a permanent variation of unavailability of transmission lines that can be estimated using BDA. Last but not least, the impact of future climate change scenarios over the unavailability of transmission lines can also be estimated by means of BDA.

From this point forward this report will focus on the estimation of unavailability due to ageing of transmission lines. This section presents the procedure for the estimation of such unavailability using Bayesian data analysis.

3-3-1 Variation of unavailabilities of transmission lines due to ageing

Ageing on transmission lines refers to the process that modifies the physical or chemical characteristics of the line [33]. Transmission lines not only fail with the pass of time but also with the use. If a line is usually congested, then that line will age faster that other not congested lines. Actions such as maintenance can reduce the effect of ageing over the lines [34]. This report does not include the effect of maintenance.

In order to perform Bayesian data analysis, three elements are necessary and they are:

- **A prior:** information from expert's opinion about the unavailability of overhead lines due to age. This information is what we have before seeing any data.
- **Data:** actual collected data from the failure of lines due to ageing.

- **A generative model:** to create a distribution function of the data in order to perform the analysis. The purpose is to simulate data with certain parameters and then match the simulated data with our actual data. In this case, the generative model will be a binomial distribution.

Using the prior, data and the generative model, it is possible to determine what parameters can produce the observed data, this means going from the data to determine the probability. Exact information about failures of lines due to ageing is not available. Synthetic information and assumptions will be used for this proof of concept estimation.

The algorithm for the implementation of the ABC method, described in chapter 2, is presented in figure 3-4. First one random probability (r) is taken from the prior. Then, the r taken from the prior will be inserted into the binomial distribution to simulate data. If the simulated data is equal to the actual data, then the probability is kept to construct later the posterior distribution.

Prior distribution: Under the assumption that a line has reached its lifetime. An expert determines that the unavailability of the line increases around 75 % but no more than 80 %. This professional guess is formed based on the bathtub curve and the observations through time. This opinion is the prior belief and it is modelled by the beta distribution function depicted in the upper part of figure 3-5. The mean of the prior is 0.75 with a standard deviation of 0.05. The parameters of the resulting Beta distribution are $alpha = 155.5$ and $beta = 51.83$ ($Beta(155.5, 51.83)$). A description of the Beta distribution function and the calculation of its parameters is presented in appendix A. The resulting prior Beta distribution has a mean value of 75% with a 90% confidence interval from 70% to 80%.

Collected information: From historical records, it is found that from 55 failures, 48 failures occur due to lines reaching the end of its lifetime without maintenance.

Generative model: The generative model for the simulation required for the likelihood calculation is a binomial distribution function with parameters $Binomial(48, 55)$.

After applying the ABC algorithm, the posterior shown in the lower part of figure 3-5 is found: The posterior has a mean value of 77.6% with a 90% confidence interval from 73.1% to 81.7%.

This procedure can be applied to any distribution function given that it is based on simulations. Several number of repetitions might be required in order to create the posterior distribution.

Alternative calculation: Since the distribution functions selected for prior and the generative model were mathematically convenient, the estimation can be performed as indicated in section 2-4-8 of chapter 2. The result is:

$$Prior = Beta(155.5, 51.83) \tag{3-2}$$

$$Likelihood = Binomial(48, 55) \tag{3-3}$$

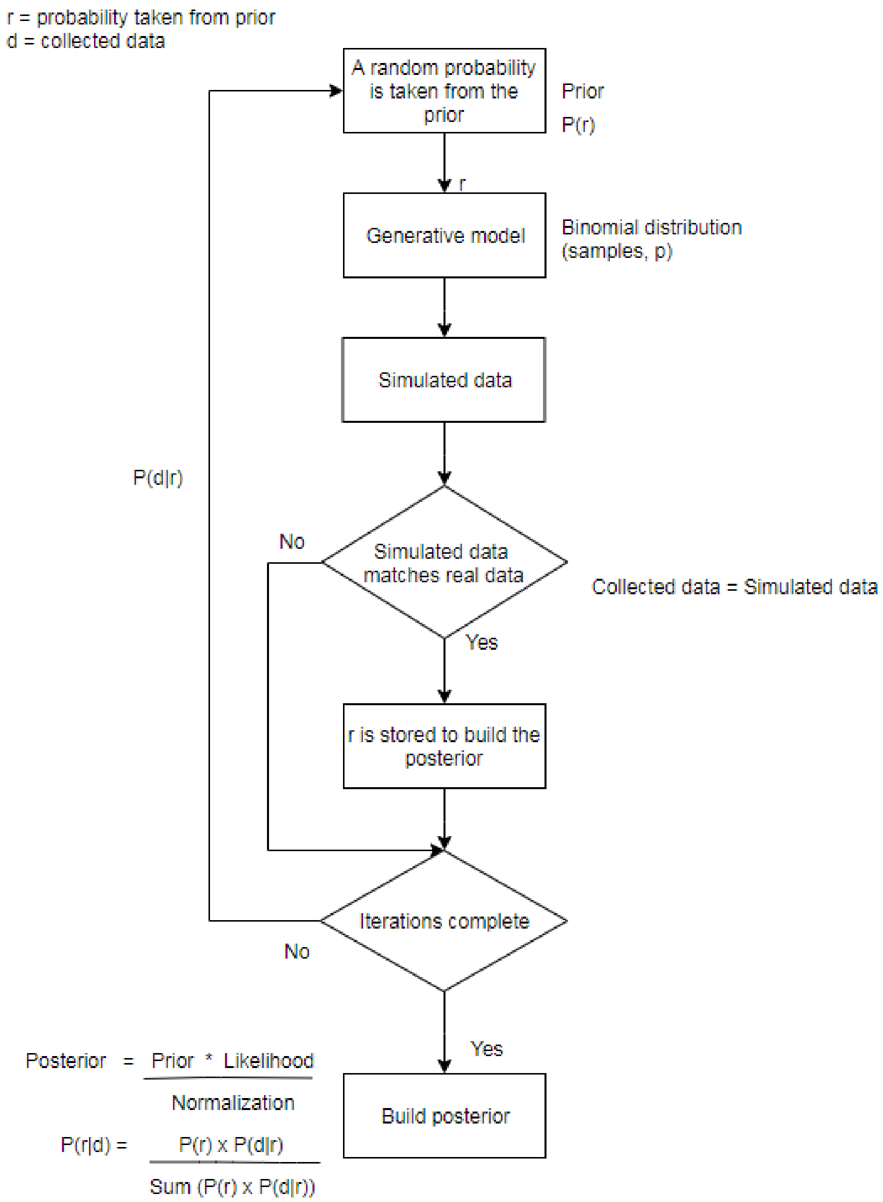


Figure 3-4: Flux diagram for the ABC algorithm.

$$Posterior = Beta(203.5, 58.83) \tag{3-4}$$

Leading to the posterior distribution function depicted in figure 3-6. The posterior distribution has the same mean value and confidence interval as calculated before using the ABC algorithm.

The same estimation procedure can be used for different operational conditions. The important thing is to realize that in order to implement BDA, the focus must be on the priors and the actual data recollection.

This alternative calculation works only if the prior is modelled as a Beta distribution and the generative model as a Binomial distribution. If the distribution functions are different, then

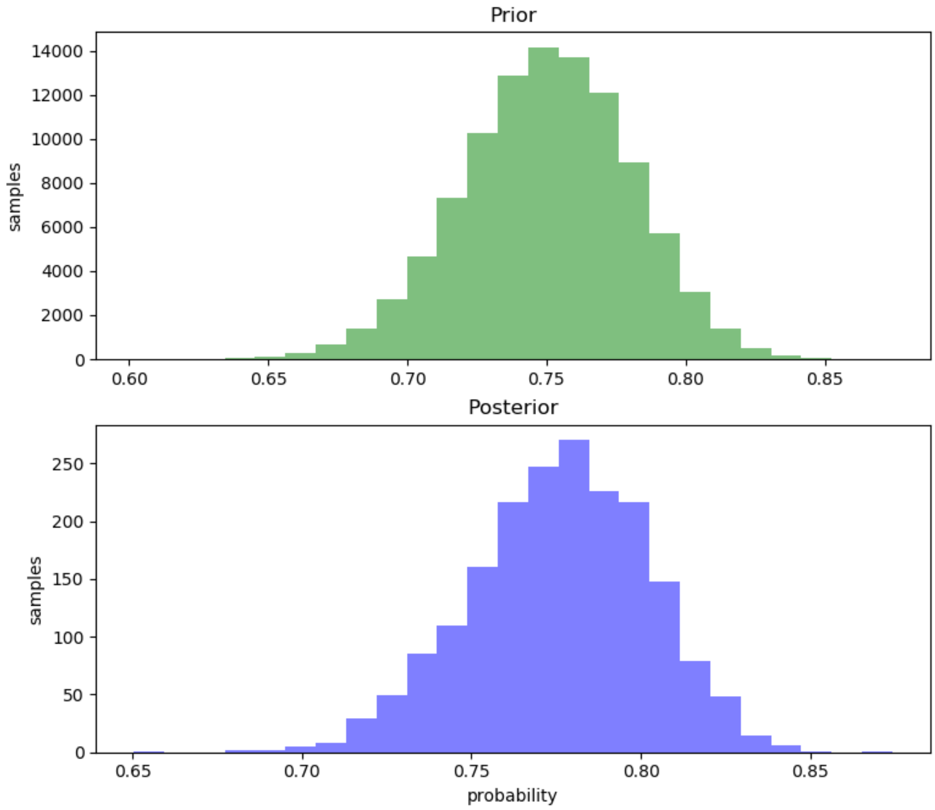


Figure 3-5: Probability estimation considering ageing.

the whole ABC simulation based algorithm has to be applied.

A proper estimation of unavailability cannot be performed if failure data is not available. For this purpose, it is important to record, update and predict failure data of components. TenneT records information of failures in the NESTOR database according to the procedure described in chapter two. However, sometimes information collection is incomplete and details about failures are lost.

A way to reduce this lost of information is to delegate teams with the purpose of collecting as much information as possible about failures. A proposition of the information that should be collected in order to allow a proper estimation of unavailability of components is presented in appendix B.

Bayesian data analysis has a great potential for the analysis of unavailability. Failures can be analyzed per individual components including conditions such as location, age and operational conditions. The important aspect that has to be kept in mind is that the priors and data collected have to be as accurate as possible. The estimated value for the unavailability found on this section will be used in a state determination analysis. The results of such analysis will be presented on the next chapter.

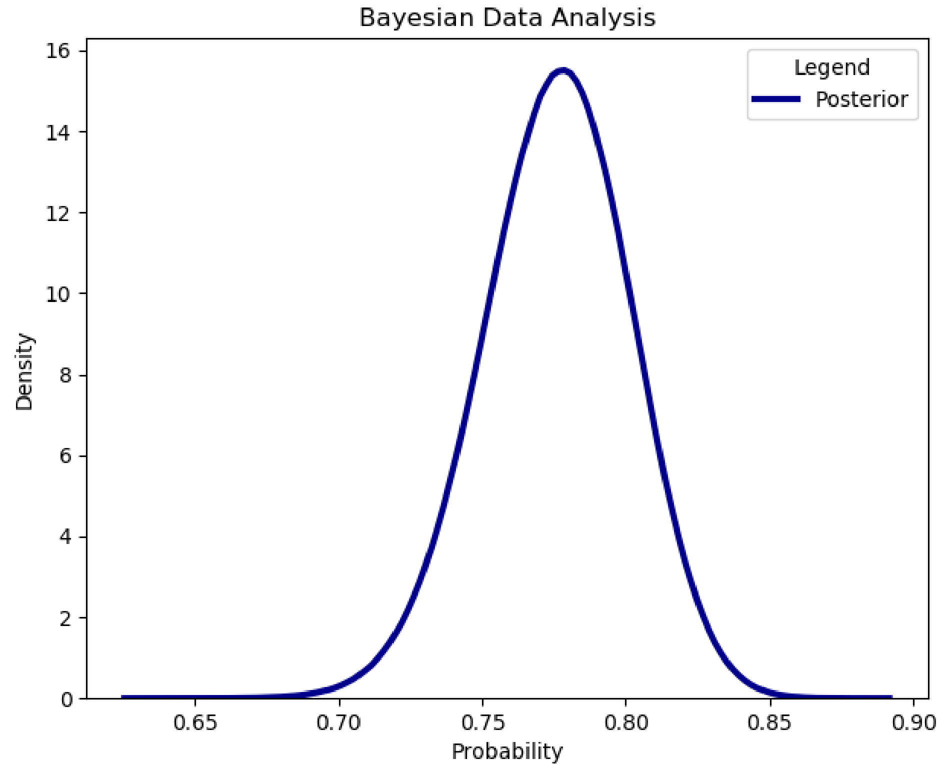


Figure 3-6: Posterior estimation of unavailability considering ageing, Alternative calculation.

3-4 Application of the methods for effect determination

This section describes the application of the theory and methodology presented in the previous chapter regarding the determination of the effect of contingencies in terms of money.

The algorithm works as depicted in the flowchart of figure 3-7. First, a load flow analysis is performed on the current grid topology to check if there is any overloads on the transmission lines. If there is at least one line overloaded, then a re-dispatch routine of only conventional generation is executed to try to solve it. If the overload is solved by the re-dispatch, then the effect is the cost of shifting up or down the generation. The cost for re-dispatch considered in this report is 90 euro/MWh for either shifting up or down. One important assumption is that every conventional generator connected to the grid is able to re-dispatch as long as it remains within its maximum limits, and the time needed for the shifting is short enough to be neglected. If the overload is partially solved or not solved at all, then the assumption that overload equals lost load is accepted and the price is calculated using 11000 euro / MWh as VoLL. The total cost is calculated adding re-dispatch costs and lost load costs. This routine is executed inside an iterative analysis.

3-4-1 Re-dispatch algorithm based on linear programming

The re-dispatch routine is executed only when there is an overloaded line. This routine works as described in the flux diagram of figure 3-8. First, the PTDF's of the current grid are

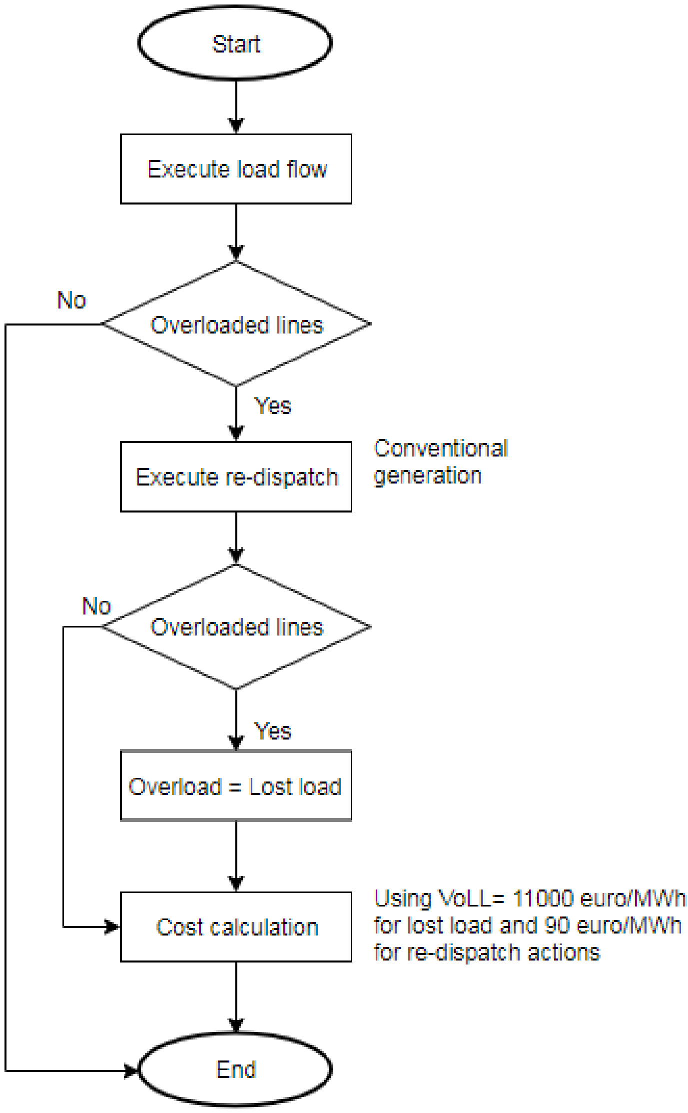


Figure 3-7: Effect determination.

calculated by injecting power on the buses were the generators are connected and withdrawing that power on a specific predefined bus. The calculated values are stored in a PTDF matrix.

The next step is to collect information about the generators and the lines. The collected information includes the current dispatch of the generators, its maximum limits, the current flow over the lines along with their maximum limits. The minimum limit for the dispatch of generators is assumed to be zero.

After the information is collected, it is used to construct the LP problem. The LP problem is constituted by control variables, an objective function and constraints. The LP problem is constructed in the following way:

Variables: The variables Xgk are positive and correspond to the dispatch of the generators. The variables are formulated in the following way:

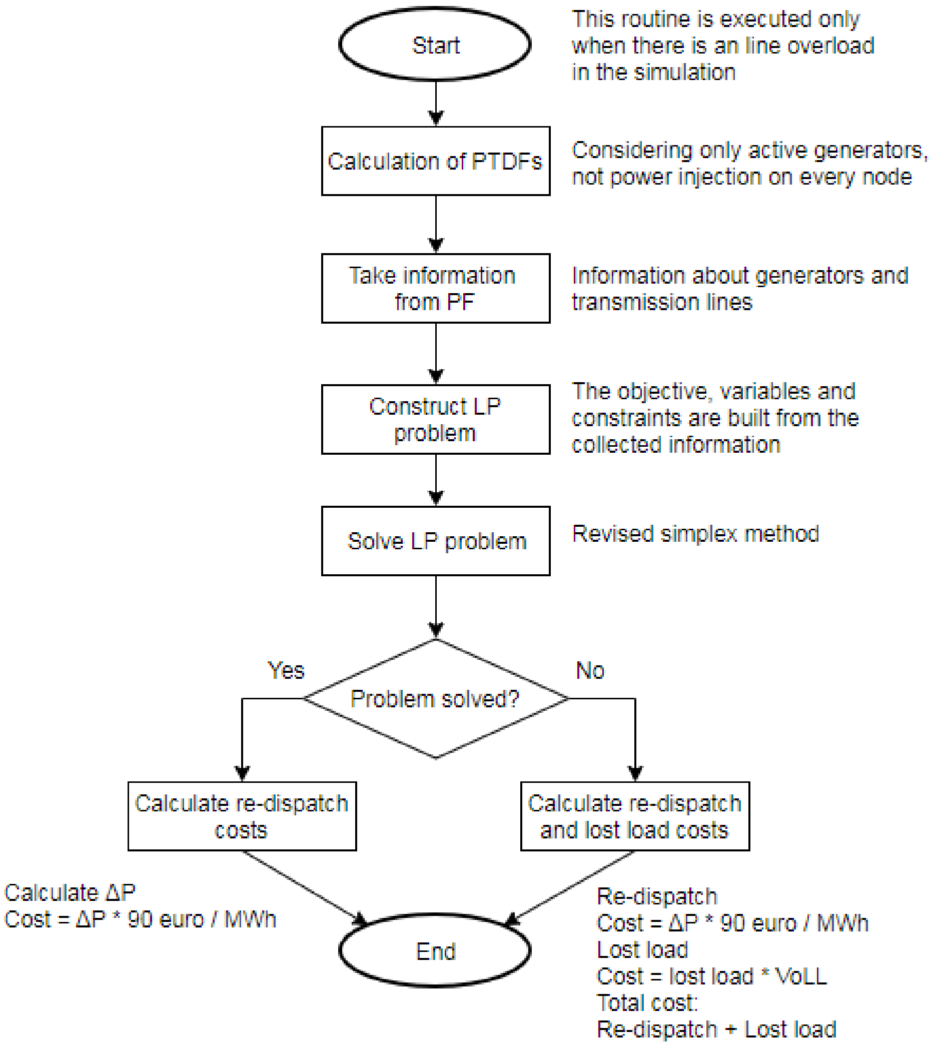


Figure 3-8: Re-dispatch algorithm.

- First the inequality of equation 3-5 is established where Pgk^{min} is the minimum limit of generator k, Pgk^0 is the actual dispatch of generator k, ΔPgk is the variation in the dispatch of generator k and Pgk^{max} is the maximum generation of generator k.

$$Pgk^{min} \leq Pgk^0 + \Delta Pgk \leq Pgk^{max} \tag{3-5}$$

- Working out the expression 3-5, the inequality presented in 3-6 is found:

$$Pgk^{min} - Pgk^0 \leq \Delta Pgk \leq Pgk^{max} - Pgk^0 \tag{3-6}$$

- Now, introducing Xgk, the expressions 3-7 and 3-8 can be found from the left and right sides of expression 3-6 respectively.

$$Xgk = \Delta Pgk - Pgk^{min} + Pgk^0 \tag{3-7}$$

$$Xgk = Pgk^{max} - Pgk^0 - \Delta Pgk \tag{3-8}$$

- Equalizing expressions 3-7 and 3-8, expression 3-9 can be found. The variable Xgk found is positive and bounded by the maximum and minimum limit of generator k. If the minimum limit is zero, then Xgk is bounded between zero and the maximum limit of generator k
$$0 \leq Xgk \leq Pgk^{max} - Pgk^{min} \tag{3-9}$$
- Finally, the variation in the dispatch of generator k is given by equation 3-10:
$$\Delta Pgk = Xgk - Pgk^{0} + Pgk^{min} \tag{3-10}$$

Objective function: With the variables defined, the objective function for the LP problem of equation 3-11 is formed. This function aims to find the minimum dispatch of the generators Xgk that is needed in order to solve the problem.

$$\Delta D = \sum_{k=1}^{ng} Xgk \tag{3-11}$$

The objective function of equation 3-11 is bounded by the following constraints

Constraints: The objective function is bounded by three conditions which are the power balance, the maximum loading limit of transmission lines and the maximum limit for the dispatch of the generators.

- **Power balance** The power balance condition states that the generation at any point of time has to match the load. After the re-dispatch is executed, this condition has to hold at every time. For this matter, if there is a generator that shifts up its dispatch, then another generator has to shift down its dispatch with the same amount. The overall variation on the dispatch of the generators involved on the re-dispatch action has to equal zero. This condition is presented with equation 3-12 where n represents the number of generators.
$$\sum_{k=1}^{n} \Delta Pgk = 0 \tag{3-12}$$
Equation 3-13 shows the condition expressed in terms of the variable Xgk
$$\sum_{k=1}^{n} [Xgk - Pgk^{0} + Pgk^{min}] = 0 \tag{3-13}$$
- **Maximum loading of transmission lines** If a line is overloaded, this means that its maximum loading limit was surpassed. This condition aims to keep the loading of transmission lines below the maximum limit in order to assure safe operation. This condition is represented by equation 3-14 where Zl_o is the current loading on line i, $PTDF_{gk,Li}$ is the distribution factor due to generator k over line i, ΔPgk is the variation on the dispatch of generator k and Zl_{max} is the maximum loading limit for line i.
$$Zl_0 + \sum_{k=1}^{n} (PTDF_{gk,Li} * \Delta Pgk) \leq Zl_{max} \tag{3-14}$$

The same condition expressed in terms of the variable Xgk is shown in equation 3-15

$$Zl_0 + \sum_{k=1}^{n} [PTDF_{gk,Li} * (Xgk - Pgk^0 + Pgk^{min})] \leq Zl_{max} \tag{3-15}$$

- **Maximum limit of generators** This condition is used to keep the dispatch of generators under their maximum limits. The condition is expressed as shown in equation 3-16.

$$Xgk \leq Pgk^{max} \tag{3-16}$$

After the LP problem has been stated, it is solved using the revised simplex method described in the previous chapter. If the problem has a solution that complies with the constraints, then the solution is labelled as optimal meaning that there are not any line overloadings in the system. On the other hand, if the problem cannot be solved, then the solution is labelled as unfeasible meaning that the overloads present on the system were not completely solved.

The last step is to calculate the cost of the re-dispatch action based on the variation of the dispatch of the generators ΔPgk and the lost load if any.

3-4-2 Cost of lost load calculation

The algorithm to calculate the cost of lost load is presented in the flux diagram of figure 3-9 .In order to determine the amount of lost load, it is necessary to first determine the bus from where the energy will be lost. The bus is determined by reading the direction of the energy flow over the overloaded line and the buses connected to the line. If the power is flowing towards a determined bus, then the lost load is assumed to be from that particular bus. After that, the quantity of lost load is determined by subtracting the current loading of the overloaded line minus its maximum loading limit. With the quantity of lost load and the bus defined, the cost is calculated by multiplying the quantity times the national value of lost load for the Netherlands (11000 euro / MWh) provided by TenneT.

The buses are sorted by provinces according to their location. Different VoLLs per province in the Netherlands can be considered. However, this report uses the national value and the determination of different VoLLs is out of the scope of this project.

3-5 Application of clustering

The clustering methods will be implemented over the base scenario of a 1008 representative hours for the year 2025. The two clustering methods to be implemented are the Hierarchical and the K-means methods.

3-5-1 Hierarchichal method

The hierarchical clustering algorithm is presented in the flux diagram of figure 3-10. First, the base data with the 1008 hours scenario is loaded. Every data point is now a cluster.

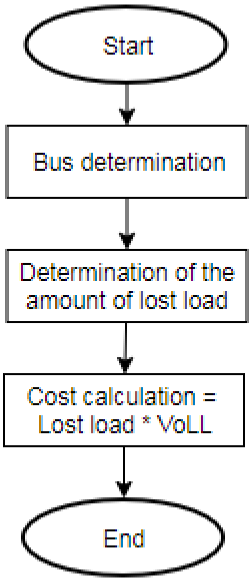


Figure 3-9: Cost of lost load.

The next step is to measure the distance between every cluster and build a matrix with those distances. The distance can be measured in any desired manner. For this report, the distance will be measured using the "Chebyshev" distance. The Chebyshev distance is the maximum distance between all the coordinates of a element of the cluster and all the coordinates of the centroid. This distance is defined as presented in equation 3-17 [35].

$$D_{Chebyshev}(e, c) = max(|e_i - c_i|) \tag{3-17}$$

Where e is the element of a cluster, c is the centroid of the same cluster, e_i is the current coordinate from the element and c_i is the current coordinate from the centroid.

After the distance is obtained, the elements are clustered by the hierarchical method described in the previous chapter until all the elements form one single cluster. The next step is to assign the data to the clusters by means of a maximum distance. The idea is that the original data points in each cluster have a maximum separation defined by the maximum distance.

After the clusters have been formed, the next step is to calculate the centroids. The centroids are calculated as the average of all the elements contained in the cluster. The information that results from this process is the number of clusters, the centroids and the number of elements per cluster.

The specified maximum distance between the centroid and every element of the cluster and the maximum distance between clusters is illustrated in figure 3-11. The centroids and the boundaries between clusters are represented by the blue dots and the data points of every cluster by the red dots. Every data point is located at a distance below the maximum distance allowed and the clusters are separated by twice this distance. For example, lets suppose that there are two clusters with data from a dispatch scenario. If the maximum distance specified in the clustering process is one hundred MW then all data points from one cluster will be located below this distance from the centroid. On the other hand, the distance between clusters will be of two hundred MW.

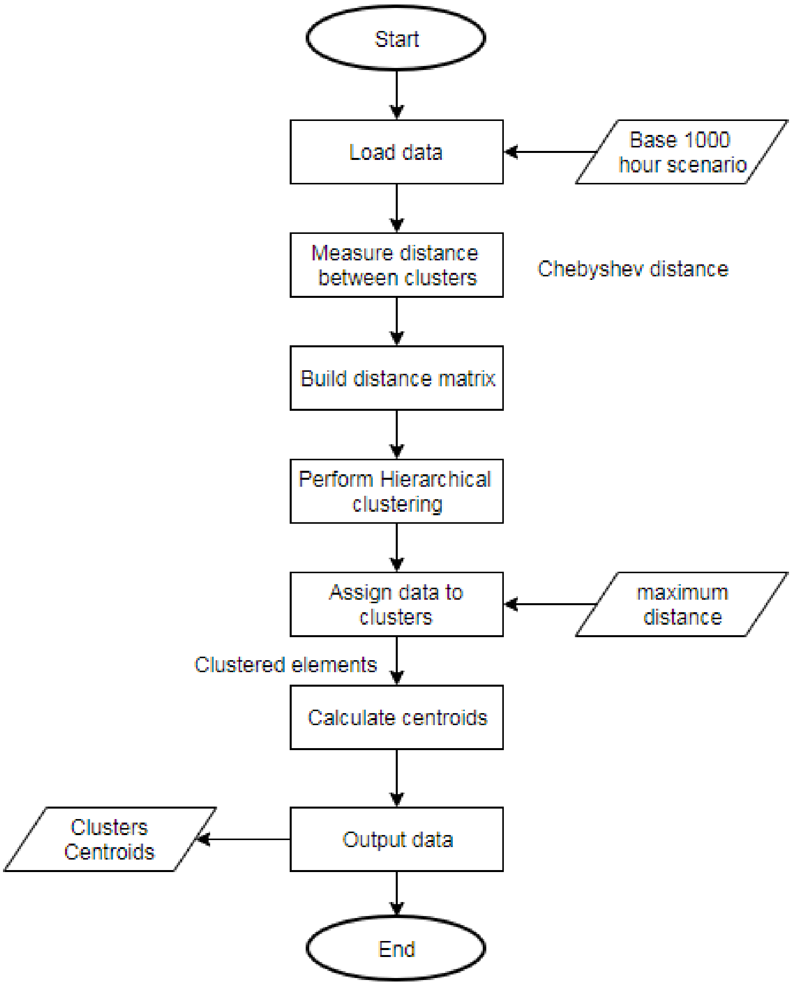


Figure 3-10: Hierarchical clustering algorithm.

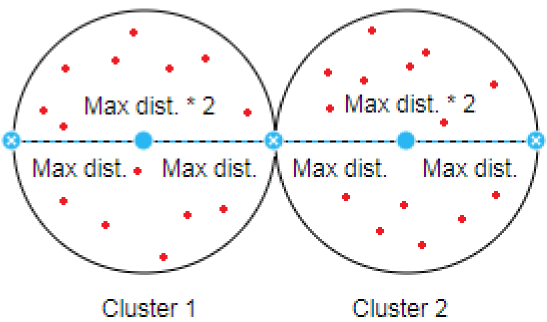


Figure 3-11: Maximum distance between clusters.

The maximum distance between centroid and the element represents in a way the maximum error allowed for the clusters. The bigger the distance is, the less clusters and centroids to analyze. However, the error in the analysis might increase because some elements of the cluster may be far from the centroid. If the distance is reduced, then there will be more centroids to analyze but the results might be better.

This clustering method offers the advantage to specify the maximum distance tolerated. In other words gives a sense of the clusters without specifying the number and without having much information about the variations and limitations of the data set. The obtained clusters and its performance will be evaluated in chapter 4.

3-5-2 K-means method

As oppose as to the hierarchical method, the K-means method needs the number of clusters to be specified which implies that there must be some previous knowledge about the data set in order to know how many clusters can be required. The K-means algorithm is depicted in the flux diagram of figure 3-12.

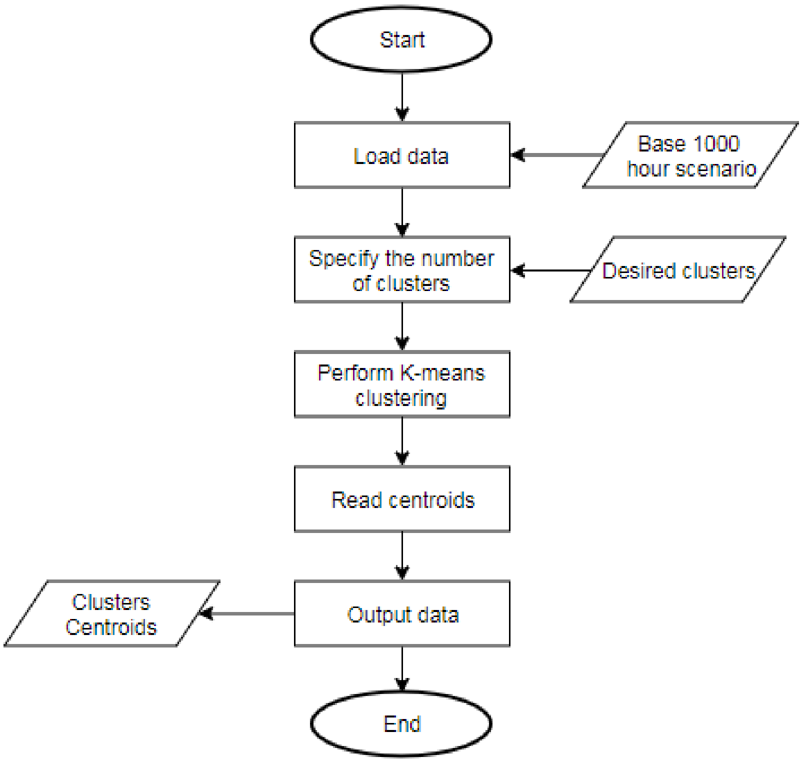


Figure 3-12: K-means clustering algorithm.

First, the same base data as the previous method is loaded. Then, the number of desired clusters is specified.

The next step is to perform the K-means clustering described in the previous chapter. After the clustering is performed, the centroids of the newly formed clusters are recorded. The information that results from this process is the number of elements per cluster and the centroids of every cluster.

After the clusters are formed, it is possible to calculate the maximum distance between the elements and the centroid of a certain cluster. In other words, only after the clustering is performed, it is possible to say something about the possible error introduced in the cluster. The K-means method is some sort of blind method until it is finished.

The performance of the two implemented clustering methods in comparison to the base case will give an idea whether it is feasible to cluster or not. In addition, an idea of which method perform better will be formed. Both methods have interesting features and downsides. The hierarchical method offers a more intuitive way to select clusters by specifying maximum distances. On the other hand, the K-means method is simpler to implement but the number of clusters have to be specified beforehand. The combination of the two methods can provide a good alternative. First the hierarchical can be used to get the number of clusters and then the K-means to get the centroids. Off course it all depends on the performance of every method. The performance and a comparison for the two methods will be presented on the next chapter.

Chapter 4

Case study and results

This chapter presents the study case developed and the results obtained from the application of the methods described in the previous chapters.

4-1 General considerations

This section describes the general considerations and the structure of the study case. The study case is developed for the test model and the test scenario introduced in chapter three. For the calculation, a representative set of data is used and the analysis is performed as a state enumeration analysis.

The general structure of the analysis is presented in the flux diagram of figure 4-1. The blocks represented by solid lines correspond to information, resources and procedures currently used and available at TenneT. The blocks represented by dashed lines correspond to information, resources and procedures not present in TenneT but proposed by this project. The analysis is described in the following way:

- First, starting on the left side of figure 4-1, the information about the dispatch and load scenario is provided by TenneT. This information will be used as input for a state enumeration analysis. The information is given with a resolution of one hour for a period of one year.
- The original set of data from the dispatch load scenario might be difficult or even impossible to analyze in a limited amount of time (i.e. within hours). In order to reduce computation time, a clustering procedure is applied to reduce the amount of data to be processed. The clustered data set serves as the input for the state enumeration analysis.
- Another input for the analysis is the topology of the grid and its parameters as planned for the year 2025. This topology is provided by TenneT.

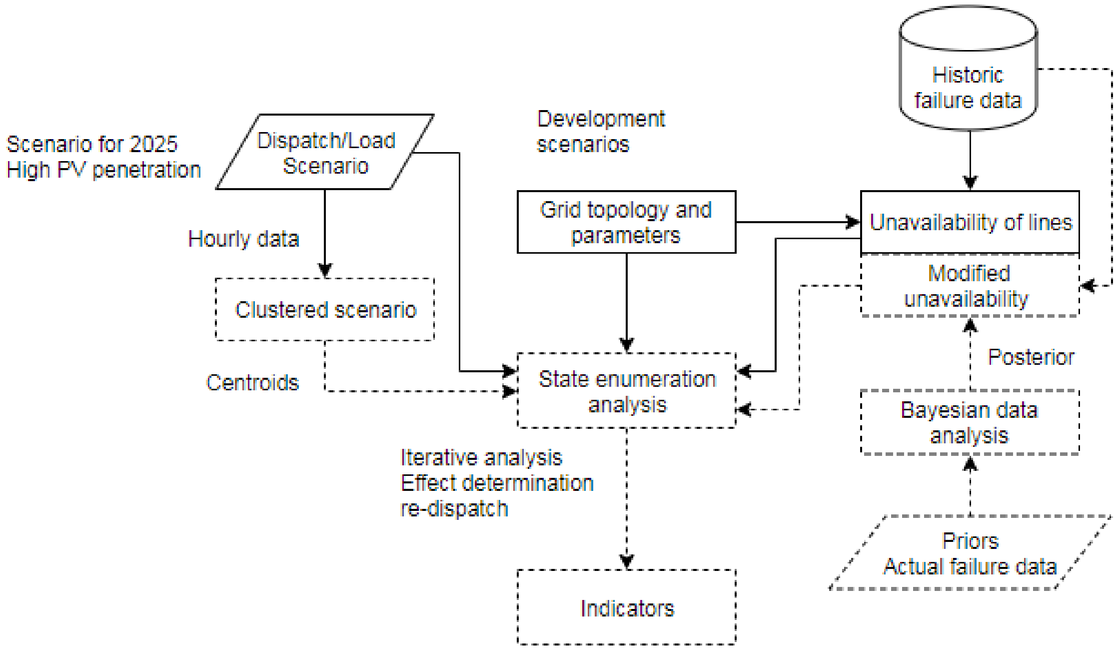


Figure 4-1: General structure probabilistic reliability analysis.

- On the right side of figure 4-1, starting from the upper part, there is the historic data of line failures. This information along with the parameters of the transmission lines are used for the calculation of line unavailability that serves as an input for the state analysis.
- Now, on the lower part of the right side of figure 4-1, there is the BDA framework that is employed to find the modified unavailability of lines. The input for the BDA is the prior knowledge from experts about failures and the actual failure data. The modified unavailability is also an input for the analysis.
- The state enumeration analysis covers up to first order contingencies. The result of the analysis are indicators about the performance of the grid. This indicators can be used for further analysis and decision making.

The next subsection introduces the representative set of data used for the analysis.

4-1-1 Representative year of a 1008 hours

The base operational scenario is based on the expected dispatch and load scenario for the year 2025 provided by TenneT. This scenario contains information for the whole 2025 year with an hourly resolution. In total there are 8736 hours (i.e 364 days) with generation and load data. The analysis of the whole set of available data requires a very long simulation time and high computational burden. For the purposes of a proof of concept case study, the complete set of information is reduced to a set of 1008 representative hours for the whole year. The hours were chosen in the following way:

- 252 hours per meteorological season (spring, summer, fall and winter) equivalent to 10 and a half days per season.
- In total there is information about 36 weekdays (Monday - Friday) and 19 weekend days (Saturday - Sunday).
- In Spring, information from 9 weekdays and 4 weekend days was chosen from the months of March, April and July.
- In Summer, information from 8 weekdays and 6 weekend days was chosen from the months of June, July and August.
- In Autumn, information from 10 weekdays and 4 weekend days was chosen from the months of September, October and November.
- In Winter, information from 9 weekdays and 5 weekend days was chosen from the months of December, January and February.

4-1-2 State enumeration analysis

The state enumeration analysis is implemented as shown in the flux diagrams of figures 4-2 and 4-3. Figure 4-2 presents the analysis for no contingencies and figure 4-3 presents the analysis for the first order contingencies equivalent to a N-1 contingency set.

The no contingency state enumeration analysis of figure 4-2 is described as follows:

- **Step 1:** the analysis requires as input the operational scenario with the hourly information of generation (conventional, RES) and loads. Also, for the analysis it is necessary to specify the starting grid topology with its parameters such as number of transmission lines, ratings, lengths, information about generators, wind units, PV units and loads.
- **Step 2:** the unavailability of lines is calculated using the input information. The calculated values of unavailability will be used for calculations of state probabilities and for the computation of indicators. The unavailability of every line is calculated using equation 4-1 where the failure frequency is given in years as well as the repair time.

$$Unavailability_{Linei} = Length_i * failurefrequency * repairtime \tag{4-1}$$

- **Step 3:** in this step, all indicators are initialized. As the analysis progresses the values of the indicators will be modified until the final value is calculated. The steps described until now are common for all the analysis. The indicators considered for this analysis are calculated before re-dispatch and after re-dispatch. The indicators before re-dispatch are:
  - Quantity and cost of lost load
  - Probability of lost load
  - Risk of lost load in terms of quantity and cost

  The indicators after re-dispatch are:

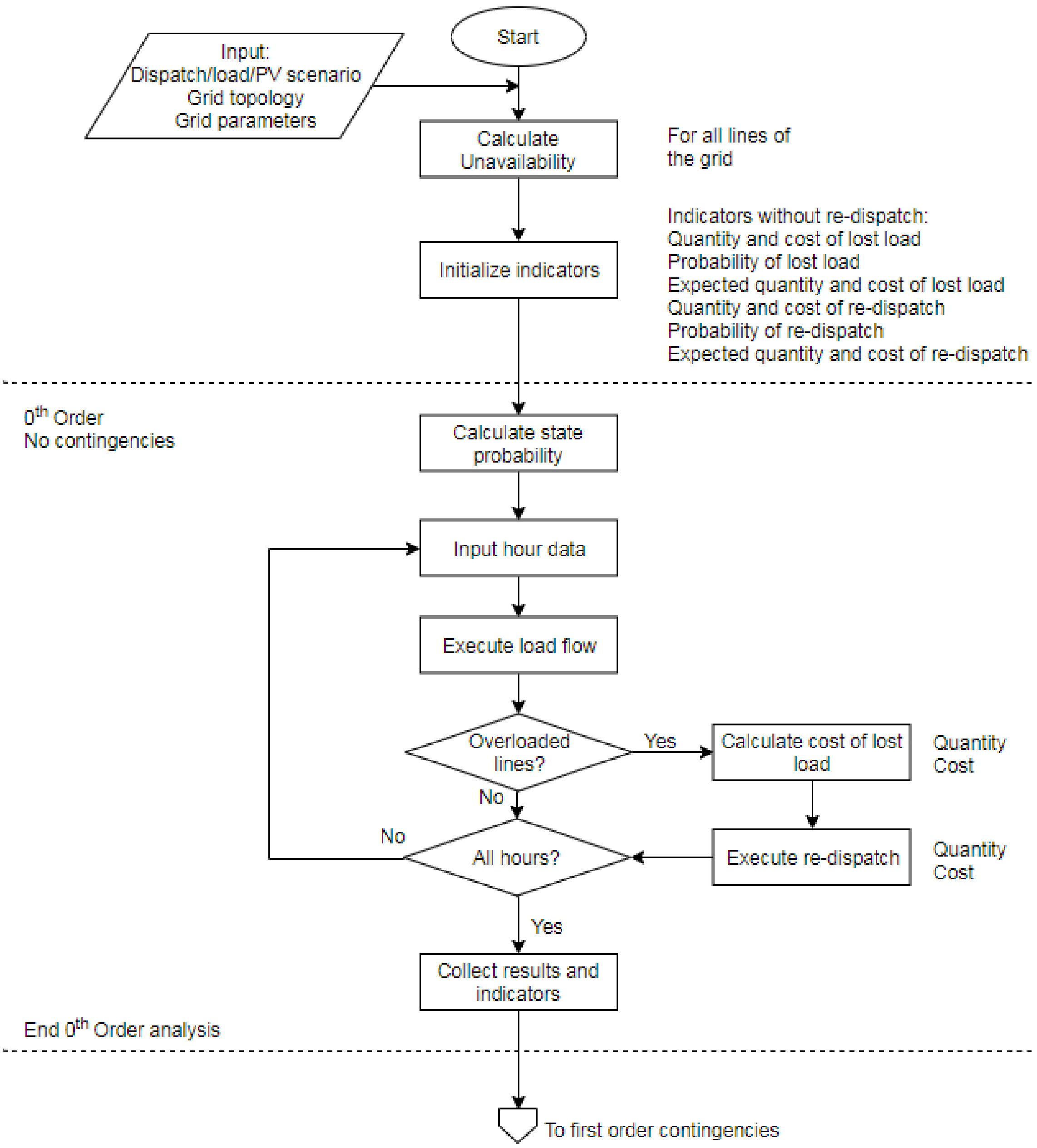


Figure 4-2: State enumeration no contingencies.

- Quantity and cost of re-dispatch
- Probability of re-dispatch
- Risk of re-dispatch in terms of quantity and cost
- Quantity and cost of lost load
- Probability of lost load
- Risk of lost load in terms of quantity and cost

   - Total cost of lost load plus re-dispatch

- **Step 4:** the probability of finding the system on the starting state is calculated. The starting state of the grid is without any line outages. This probability is calculated using equation 4-2 where $Pstate$ is the probability of the state, n is the number of lines and A_i is the availability of line i. $Pstate$ can also be calculated using equation 4-3 where U_i is the unavailability of line i.

$$Pstate = \prod_{i=1}^{n} A_i \tag{4-2}$$

$$Pstate = \prod_{i=1}^{n} (1 - U_i) \tag{4-3}$$

- **Step 5:** the hour data for the dispatch and load scenario is loaded to the grid.
- **Step 6:** once that the information is loaded, a decoupled power flow is executed in order to obtain the power flowing in each line of the grid.
- **Step 7:** if there is an overloaded line, the amount and cost of lost load are calculated. After the calculation, the re-dispatch routine is executed. The values resulting from the routine are kept for later usage.
- **Step 8:** the analysis jumps back to the input data block (step 5) and it repeats until all hours are analyzed. Once all hours have been processed, the analysis carries on with the next step.
- **Step 9:** this step consist in collecting all the information generated and to calculate the zero order indicators corresponding to this part of the analysis. The values of all the indicators are kept since higher order indicators are built upon the first indicators.

The first order state determination analysis of figure 4-3 that continues the complete analysis is described as follows:

- **Step 1:** the same action as step 5 from the no contingency analysis is repeated.
- **Step 2:** the next step is to chose the first order contingency to be analyzed. The transmission lines of the grid are sorted by their names alphabetically. The first line of the list is selected as the first outage and so on. One line is selected at the time until the list is finished.
- **Step 3:** once that the line is out, The probability of finding the system on that particular state is calculated. This probability is calculated using equation 4-4 where U_k is the unavailability of the current line out. This probability is also calculated using equation 4-5.

$$Pstate = U_k \prod_{i=1}^{n} A_i \tag{4-4}$$

$$Pstate = U_k \prod_{i=1}^{n} (1 - U_i) \tag{4-5}$$

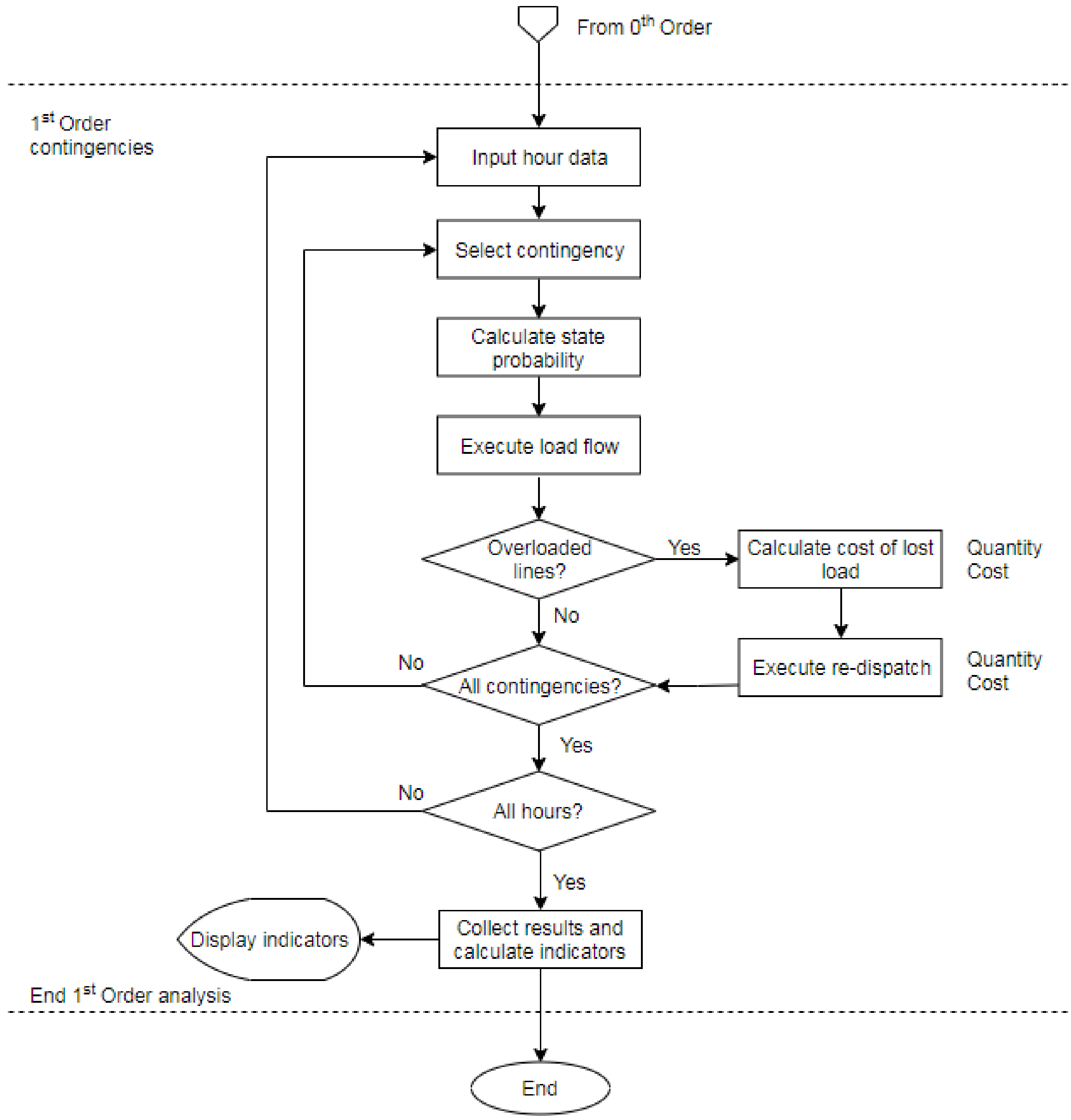


Figure 4-3: State enumeration first order contingencies.

- **Step 4:** after this calculation, step 6 from the no contingency analysis is repeated.
- **Step 5:** step 7 from the no contingency analysis is repeated. After these calculations are performed the analysis proceeds by checking if there is not any contingency left.
- **Step 6:** if there are still contingencies to analyze, the analysis jumps back to step 2 and repeats until there are not more contingencies left. Once that all contingencies are analyzed, the next step is to check whether all hours have been analyzed.
- **Step 7:** if there are still more hours to analyze, then the analysis goes back to step 1 and the procedure repeats until all hours are considered.

- **Step 8:** After all hours were analyzed, step 9 from the no contingency analysis is repeated, this time to calculate first order indicators. Once the calculations are performed, the indicators are displayed so they can be further analyzed. This step is the last for the state enumeration analysis.

This state enumeration analysis presented follows the logic hour - contingency which means that per every hour of the analysis, the whole set of contingencies is built implying that the grid structure is modified every hour. Another option is to follow the logic contingency - hour which means that every time that the grid is modified, every hour is analyzed. The performance of the implementation of one logic or the other depends on the selection of the simulation software and its performance at the moment of data writing or grid modification.

To illustrate this, lets suppose that there is a set of 10 hours to be analyzed over the test grid that contains 86 single lines. If the logic hour - contingency is followed, then the dispatch and load scenario input data will be written 10 times on the grid while the grid will be modified 860 times. On the other hand, if the logic contingency - hour is followed, the grid will be modified only 86 times but the data input will be executed 860 times.

If the software chosen can modify the grid faster than it can write data, then the logic hour - contingency might be the best option because there are less writing actions than grid modifications. In contrast, if the data writing is faster than the modification of the grid, then the logic contingency - hour might be the best option since there are less grid modification actions and more data input actions.

The next section, presents the results and the time performance of the analysis over the test model.

4-2 Results for state enumeration analysis over the test model

This section presents the results of the state enumeration analysis using the representative data from the test scenario and the test model. The unavailability of lines is calculated using 0.0020 [/cct*km*y] as failure frequency which corresponds to the average value calculated by TenneT for a kilometer of single 220/380kV circuits. Table 4-1 shows the results obtained from the analysis. The table presents the indicators calculated without re-dispatch and after executing re-dispatch. All cost indicators are given in Euros. The values obtained are used as base for further comparisons.

From the table, it is possible to see that the grid is not N-1 secure under the specified conditions. The left side of figure 4-4 shows the amount of lost load before and after re-dispatch and the amount of energy re-dispatched. Without re-dispatch, the amount of lost load is 5,94 TWh. Considering re-dispatch, the amount of lost load reduces in a 64% (3.84 TWh) to the value of 2.10 TWh. However, the amount of re-dispatched energy is 67 TWh, around 17 times the saved energy. In order to cut down the lost load, it is necessary to re-dispatch a considerable amount of energy.

The risk indicators in terms of lost load can be seen on the right side of figure 4-4. The risk indicators presented are the lost load before and after re-dispatch as well as the expected

Results for N-1	**1008 hours**
Not N-1 secure	
Quantity of lost load before re-dispatch	5,94 TWh
Cost of lost load before re-dispatch	65,5 Billion
Probability of lost load without re-dispatch	1,057exp-3
Risk of lost load without re-dispatch (Quantity)	474,5 MWh
Risk of lost load without re-dispatch (Cost)	5,21 Million
Quantity of re-dispatch	67 TWh
Cost of re-dispatch	6 Billion
Probability of re-dispatch	7,95exp-4
Expected amount of re-dispatch	5 GWh
Quantity of lost load after re-dispatch	2,1 TWh
Cost of lost load after re-dispatch	23,71 Billion
Probability of lost load with re-dispatch	2,62exp-4
Risk of lost load with re-dispatch (Quantity)	188,34 MWh
Risk of lost load with re-dispatch (Cost)	2,07 Million
Total cost after re-dispatch	**29,7 Billion**

Table 4-1: State enumeration using the base case

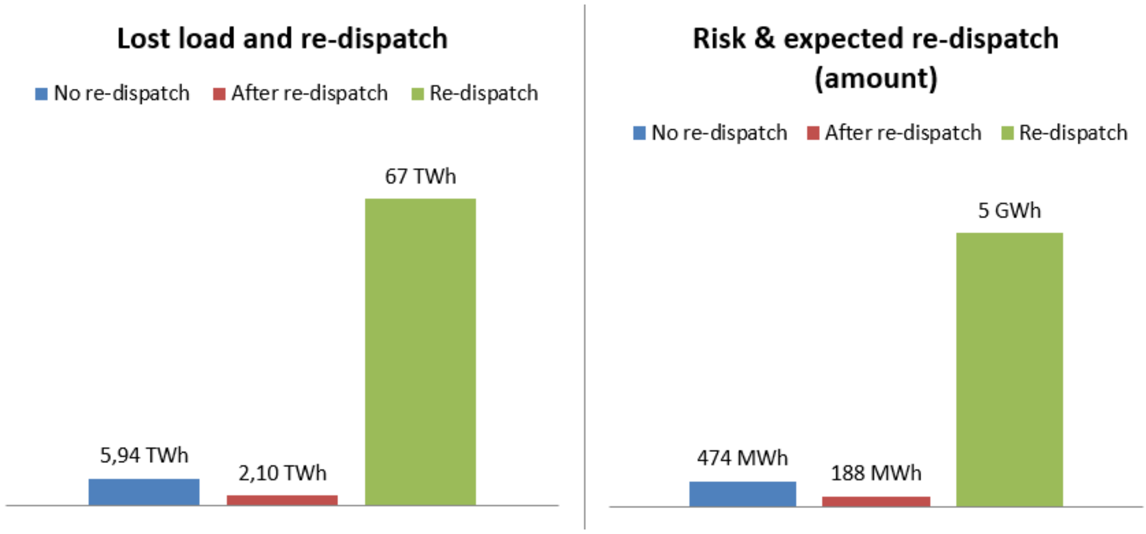


Figure 4-4: Lost load and re-dispatch.

amount of re-dispatch. The risk is calculated as the amount of lost load times the probability of the state that led to the lost load. The risk of lost load without re-dispatch is 474 MW, the risk of lost load with re-dispatch is 188 MW which correspond to around 40% of the risk without re-dispatch. Nevertheless, in order to achieve that reduction of risk, the expected amount of re-dispatched power is 5 GW.

The left side of figure 4-5 presents the costs in Euros linked to the amount of lost load and re-dispatch. The cost of the lost load before re-dispatch is 65,5 Billion. By implementing the re-dispatch routine, the lost load cost reduces to 23,7 Billion. The cost of executing re-dispatch is around 6 Billion and the total cost after re-dispatch (lost load after re-dispatch +

cost re-dispatch) amounts to 29,7 Billion. Executing re-dispatch reduces drastically the cost of lost load.

The right side of figure 4-5 shows the risk in terms of cost. The risk of lost load without re-dispatch is of 5,22 million Euros. By including re-dispatch, the risk reduces to 2,07 million Euros which is the 40% of the no re-dispatch risk. The expected cost for re-dispatch is 0,45 Million and the combination of risk after re-dispatch and the expected cost of re-dispatch is 2,52 Million Euro. The amount of lost load, cost and risks in this case are similar.

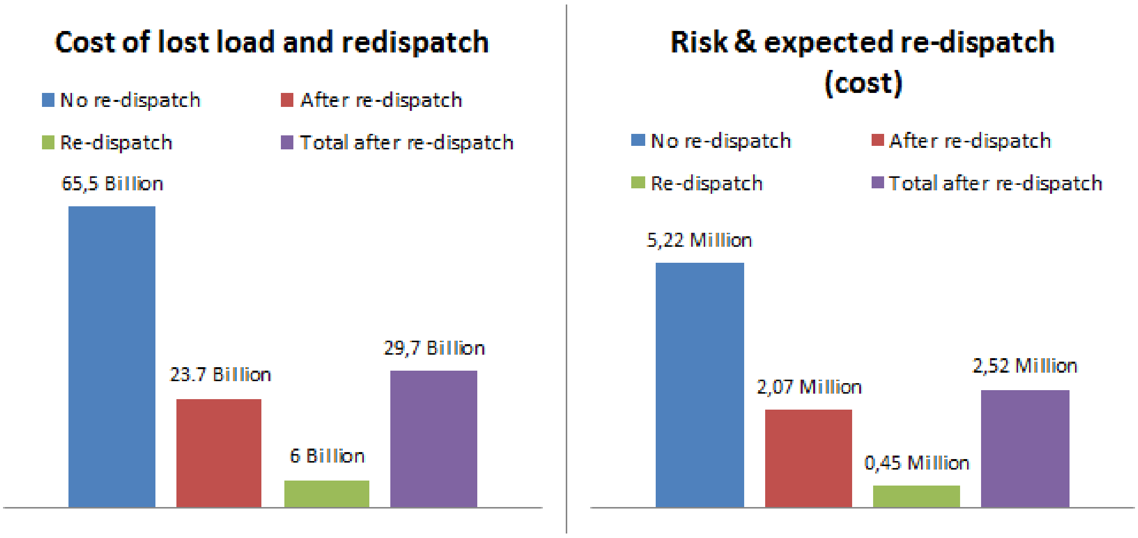


Figure 4-5: Cost of Lost load and re-dispatch.

Executing re-dispatch implies a great task. As it can be seen, it is necessary to shift 67 TWh in order to save 3,84 TWh of lost load. However, looking at the costs and risk behind the actions, re-dispatch costs 6 Billion and saves 41,8 Billion. Despite all the energy needed for re-dispatch, it makes sense to include it in the analysis. If decisions were to be made without including the effects of re-dispatch, the results might be too pessimistic and can lead to an over-dimensioned grid.

Mapping the amounts of lost load and the re-dispatch into TenneT's risk matrix, it is possible to see that the risk indicators are lowered by executing re-dispatch. Figure 4-6 shows the risk matrix before re-dispatch and figure 4-7 represents the risk matrix after re-dispatch. The risk index lowers from high to medium.

Both levels of risk are unacceptable according to the risk matrix and the N-1 criterion. This means that it is not possible to connect an additional 6 GW of renewable power without some modifications in the grid. Nevertheless, after the probabilistic analysis, an alternative way to assess this case with high penetration of RES is presented showing that it is possible to add that kind of power at the expense of some risk depicted by the risk indicators. The next question and forthcoming challenges lie in the determination of how much risk can be accepted and under which conditions. Perhaps a risk of lost load after re-dispatch of 2,52 million Euros is acceptable for the representative data set of 1008 hours with a probability of once every 3816 years. The terms of what is acceptable or not are out of the scope of this work but it is an important question that needs to be answered.

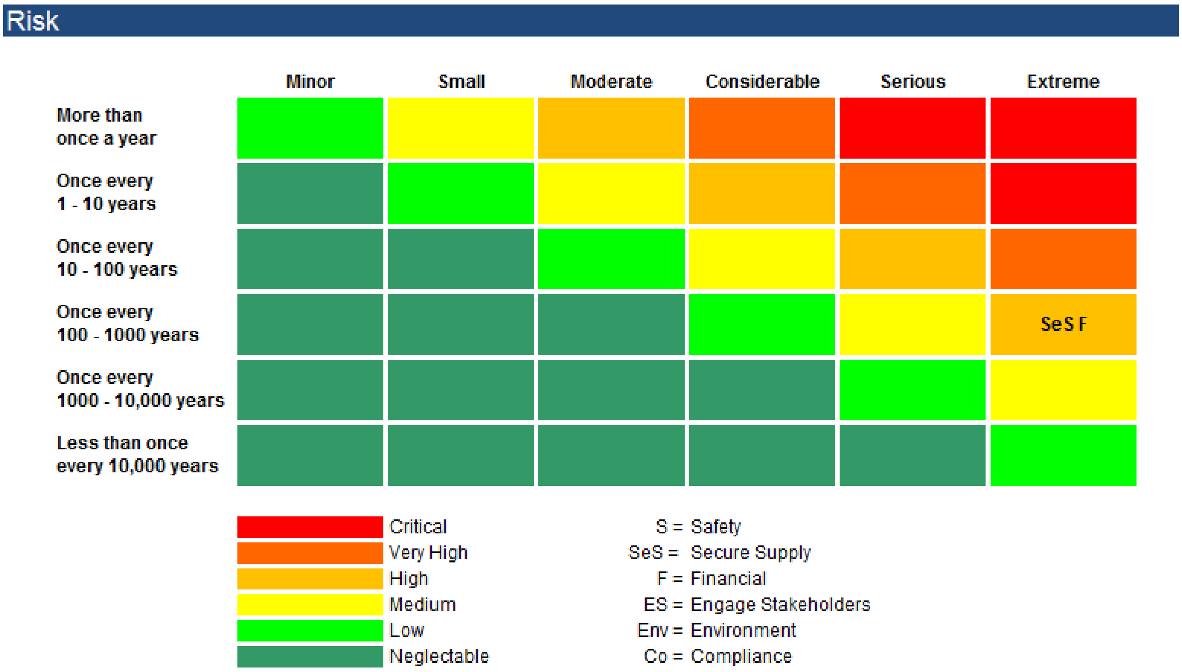


Figure 4-6: Risk matrix before re-dispatch.

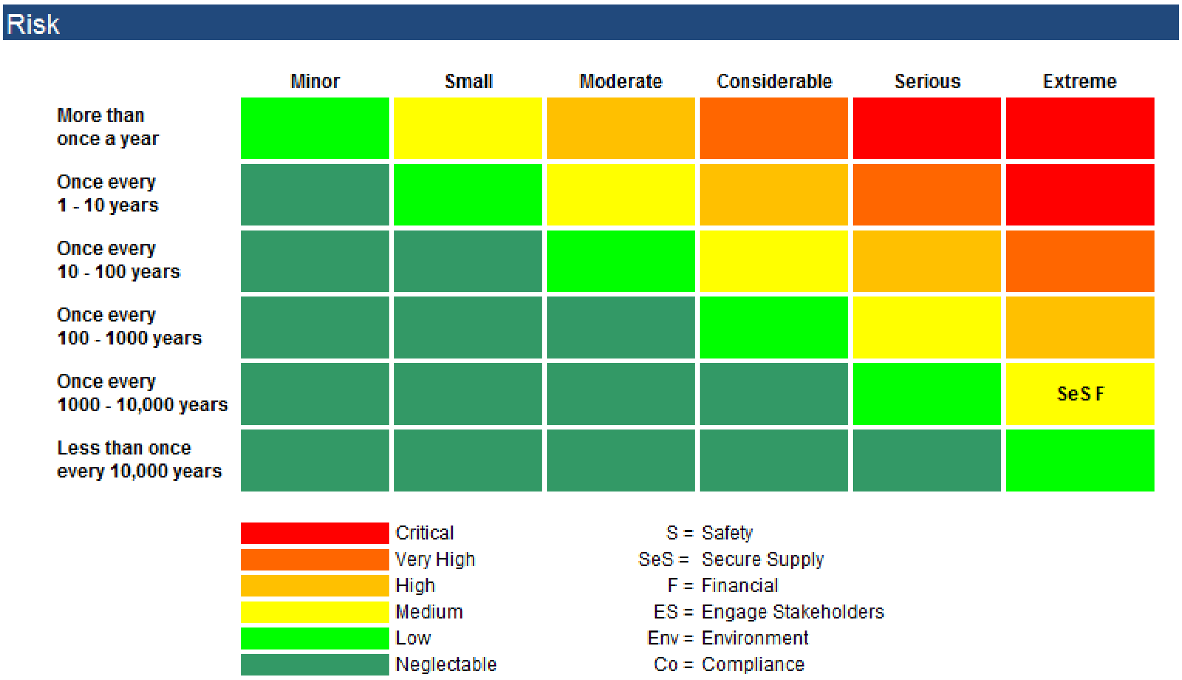


Figure 4-7: Risk matrix after re-dispatch.

4-2-1 Computational performance of the analysis

The analysis was mainly carried on two different computers with the following characteristics:

- **Computer 1:** Processor: Intel Xeon E3-1240V2 @ 3,40GHz, 4 cores. RAM memory: 16GB
- **Computer 2:** Processor: Intel Core i5-6300V @ 2,40GHz, 2 cores. RAM memory 8GB

The main structure of the simulation software is represented in figure 4-8. All the necessary software was coded in the programming language Python version 3.6 (Py). Python is coupled with the power system analysis software PowerFactory 2018 (PF).

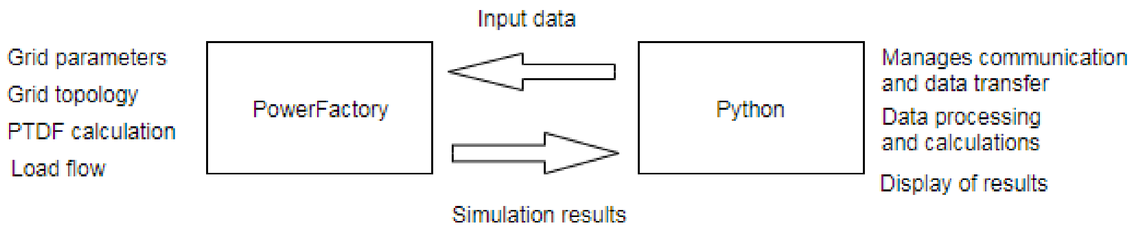


Figure 4-8: Python - PowerFactory structure.

PF manages all the aspects regarding the test grid. Py manages the connection and data communication with PF as well as the data preparation, processing, calculations and display of results. The input information is prepared and written into PF using Python. Once that PF receives the information, it manipulates the grid and performs load flow simulations. After that, the grid results are delivered back to Py. The received information is processed. If it is necessary, the processed information is written again into PF for further load flow simulations. The cycle repeats until no further simulations are required and the results are presented.

A generic representation of the Py-PF cycle is presented in figure 4-9. Steps one through three are executed by Py. Steps four through 6 are performed by PF. Steps from seven to nine are executed again by Py.

A single cycle from step one to nine without iterations takes the approximately the following time to be completed:

- **Computer 1:** For a first run, the cycle can take around 30 - 40 seconds distributed in the following way:
  - Steps 1 and 2: 5 to 10 seconds
  - Step 3: 25 to 35 seconds
  - Steps 4 to 9: 0,05 to 0,2 seconds

  If the cycle was launched before, then the execution time can take around 25 to 35 seconds distributed in the following way:
  - Steps 1 and 2: 0,05 to 0,15 seconds
  - Step 3: 25 to 35 seconds
  - Steps 4 to 9: 0,05 to 0,2 seconds

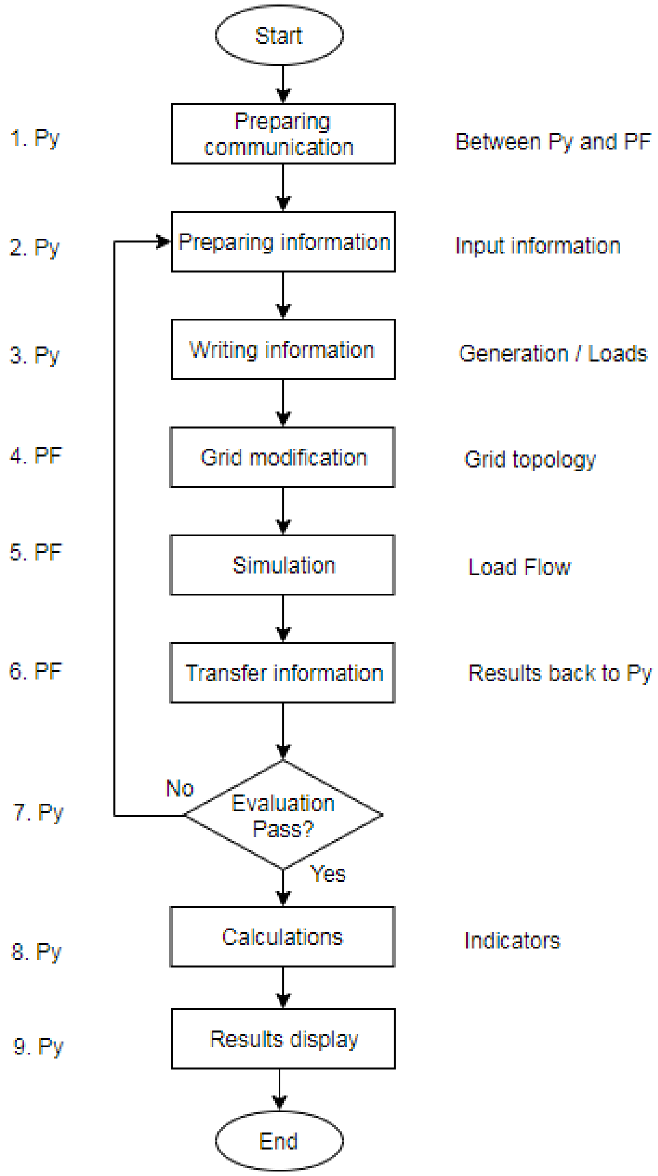


Figure 4-9: Generic Py-PF cycle.

- **Computer 2:** For a first run, the cycle can take around 60 - 70 seconds distributed in the following way:
  - Steps 1 and 2: 18 to 25 seconds
  - Step 3: 40 to 45 seconds
  - Steps 4 to 9: 0,5 to 1 seconds

  If the cycle was previously launched, the execution time can take around 40 to 45 seconds distributed in the following way:
  - Steps 1 and 2: 0,05 to 0,1 seconds

– Step 3: 40 to 45 seconds
– Steps 4 to 9: 0,5 to 1 seconds

The execution times are not absolute sometimes they vary. However this variation is in the range of a few seconds up or down.

In both computers, the execution time is strongly dominated by the writing information time (step 3). The rest of the cycle does not contribute greatly to the final execution time. This means that the modification of the grid and the power flow simulation are not very time demanding. As a consequence, the less writing actions executed, the better performance.

Performance logic contingency-hour

In this logic, there are more writing actions than modifications of the grid. In the implementation of the analysis, every time that a contingency is evaluated, a writing action is executed. The performance for the first order contingency analysis of one hour is presented in table 4-2: The execution time is quite high, assuming that the execution time remains

Hours	Computer 1	Computer 2
1	32 minutes	1,42 hours
20	11,23 hours	24,8 hours

Table 4-2: Performance contingency - hour

somewhat constant every hour, the estimated time for the simulation of the representative year of 1008 hours would be around 22 days for computer 1 and around 52 days for computer 2. The implementation of this logic is not feasible under the conditions of this analysis.

Performance logic hour-contingency

In this logic there are less writing actions than modifications of the grid. In an analysis using this logic, one writing action is performed and then the whole set of contingencies is evaluated. The performance of this logic for one hour analysis is presented in table 4-3:

Hours	Computer 1	Computer 2
1	6 minutes	12 minutes
100	9,16 hours	16,43 hours

Table 4-3: Performance hour - contingency

The execution time reduces drastically if compared to the previous logic. Assuming a rather constant execution time per hour, a projection for the execution time for the set of 1008 on computer 1 would be of around 4 days and around 7 days for computer 2. However, this assumption can be deceptive since the time employed per hour depends highly on the amount of hours that require a re-dispatch action. Every time that a re-dispatch action is executed, two writing actions are performed increasing in this way the simulation time.

For this project, the logic hour contingency is convenient because it takes less time to be completed. In other cases a different logic might be better. It all depends on the performance of the software in terms of data writing and grid management.

Performance of the analysis of the representative year

The workload for the analysis of the representative year was divided between the two computers. No parallel computing algorithms were implemented. As a result, the computation time for the analysis was 7 days and 14 hours. Even with the improvement of the simulation time provided by the logic employed, this execution time is quite enormous. If the analysis were to be performed using a set of data for one complete year, it could take around 60 days to be completed which is completely out of the question. The performance of the software selected under the simulation conditions is just not suitable for this kind of analysis. One further step out of the scope of this analysis would be to implement parallel computing, explore with different software and discussing the writing issue with the software vendor in order to rule out the writing issue as a bug.

A valid option to improve the simulation time is by executing clustering. The next section presents the performance of the application of clustering.

4-3 Clustering applied to the representative year set of data

Given the complication of the high simulation time, means of clustering are applied to the representative set of data in order to reduce the number of instances to be analyzed. The clustering is implemented by means of the Hierarchical and K-means methods previously described in chapters two and three. In order to evaluate the results, the centroids of the obtained clusters will serve as the input for the state enumerations analysis. After one centroid is analyzed, the result will be multiplied by the number of elements corresponding to that cluster. The unavailability has the same value as in the base case.

4-3-1 Performance of Hierarchical clustering

Figure 4-10 represents the performance of hierarchical clustering applied to the 1008 hours dispatch base scenario for generation and load. While the number of clusters increases, the maximum distance decreases (more clusters - less error). It is important to remember that the distance between clusters is twice the distance between the elements from a cluster and its centroid. As an example, according to the figure 4-10, if a reduction of 50% of clusters is desired then the maximum distance between clusters is around 483MW which implies that the maximum distance between centroid and data point is around 241,5MW.

The first maximum distance between centroids and the elements from the clusters is chosen aiming to achieve a reduction from 1008 hours to somewhere around 200. According to figure 4-10 a distance between clusters of about 1000MW might be an option. In that sense, the first distance chosen for the analysis is 500MW between centroid and elements or 1000MW between clusters.

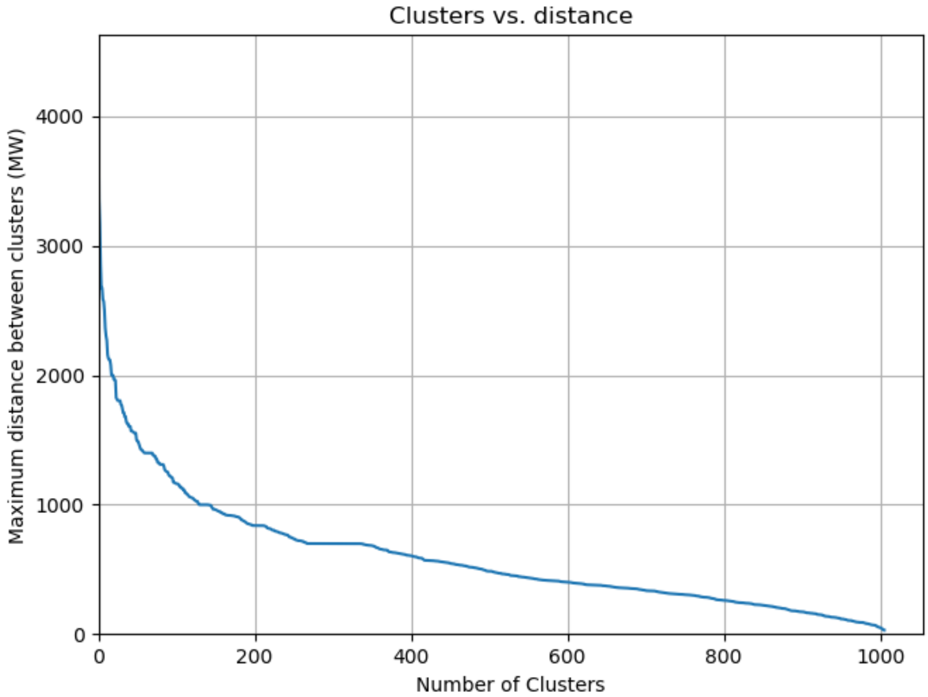


Figure 4-10: Number of clusters vs distance.

Clusters with 500 MW as maximum distance between centroids and elements (ClusterH500)

The specification of this maximum distance results in 141 clusters. The results obtained from the state enumeration analysis are presented in table 4-4. All costs from the subsequent tables and figures are given in Euros.

Figure 4-11 presents a comparison between the indicators for risk of lost load in MWh obtained from the analysis of the cluster and the indicators obtained from the base case. In all figures, the indicators from the cluster analysis are represented by the letter C after the name. The risk of lost load without re-dispatch from the cluster analysis is 6,5% smaller than the value from the base case. The risk of lost load after the re-dispatch obtained from the cluster analysis is 15,9% smaller than the value from the base case. As for the expected re-dispatch, the value from the cluster analysis is 5,8% larger than the base case.

Figure 4-12 presents the comparison of the risk indicators in terms of cost of lost load and expected cost of re-dispatch. The cost risk indicators are obtained from the risk indicators of figure 4-11. As expected, the difference between the risk before, after re-dispatch and the expected re-dispatch remains. The last two indicators at the right side of figure 4-12 represent the total cost risk after implementing re-dispatch. The total risk obtained from the clusters is 12% smaller than the value from the base case.

Using 500 MW as the maximum distance between centroid and elements for the data grouping in clusters, produces the results presented above. The difference between the indicators obtained from the base case and the clusters is reasonable. The maximum difference is 15,9% and the minimum difference is 5,8%. The total execution time for the cluster analysis was

Results for Cluster H500	**141 Clusters**
Not N-1 secure	
Quantity of lost load before re-dispatch	5,59 TWh
Cost of lost load before re-dispatch	61,52 Billion
Probability of lost load without re-dispatch	1,06exp-3
Risk of lost load without re-dispatch (Quantity)	443,67 MWh
Risk of lost load without re-dispatch (Cost)	4,88 Million
Quantity of re-dispatch	69,71 TWh
Cost of re-dispatch	6,27 Billion
Probability of re-dispatch	8,27exp-4
Expected amount of re-dispatch	5,29 GWh
Quantity of lost load after re-dispatch	1,83 TWh
Cost of lost load after re-dispatch	20,18 Billion
Probability of lost load with re-dispatch	2,35exp-4
Risk of lost load with re-dispatch (Quantity)	158,17 MWh
Risk of lost load with re-dispatch (Cost)	1,74 Million
Total cost after re-dispatch	**26,46 Billion**
Execution Time	**50 hours 37 minutes**

Table 4-4: State enumeration ClusterH500

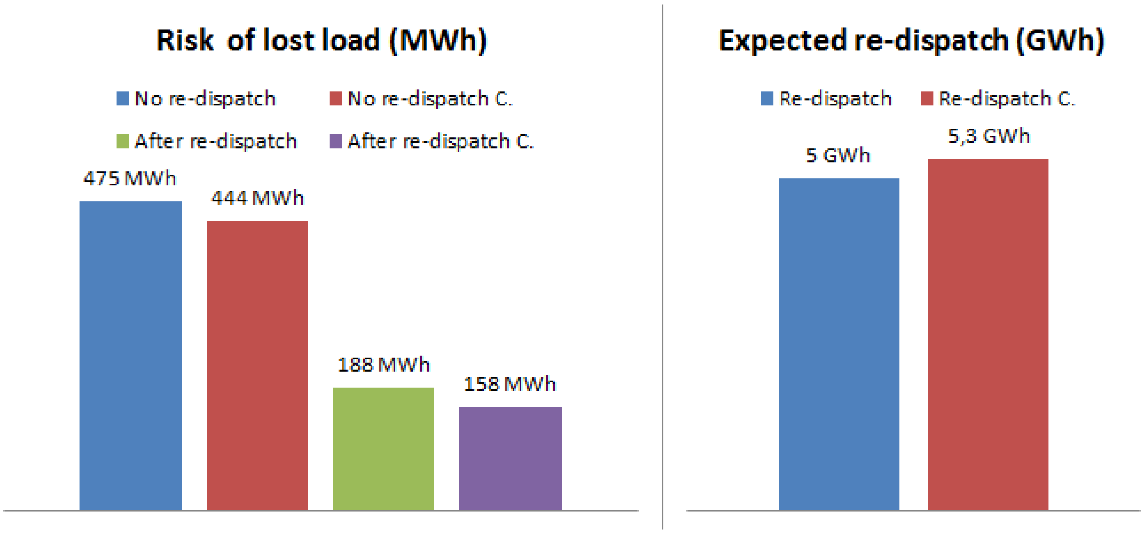


Figure 4-11: Comparison, risk of lost load and expected re-dispatch (MWh).

around 2 days and 2 hours. Compared to the execution time for the base case (7 days, 14 hours) there is a reduction of around 71% in the execution time.

The results suggest that the difference between the indicators produced with the centroid of a cluster and the elements of that cluster produce more or less the same results implying that the dispatch and load values grouped are quite similar between each other. The amount of time saved and the error found justify the application of clustering. If the distance employed for clustering reduces, then the number of clusters will increase. Better results, more similar to the base case would be expected. This means that there is always the possibility of analyze

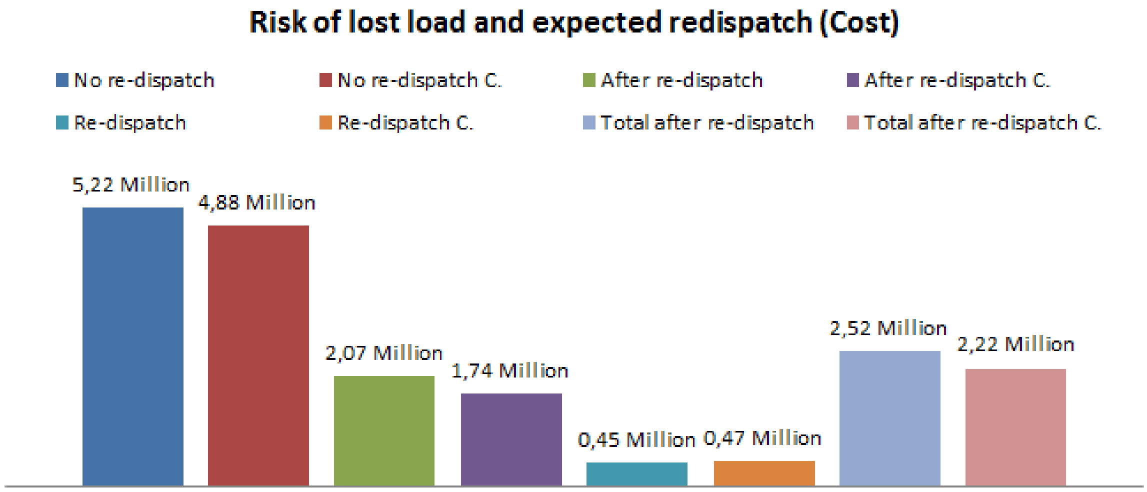


Figure 4-12: Comparison of cost of lost load and expected re-dispatch.

more clusters if better results are required. This work focuses in finding the results for further reductions in order to determine how far is it possible to cluster. The order to seek a further reduction on the hours to analyze, the next distance employed for the cluster formation is 750MW between centroid and the elements of a cluster.

Clusters with 750 MW as maximum distance between centroids and elements (ClusterH750)

This maximum distance results in 52 Clusters. The state enumeration results are presented in table 4-5.

Figure 4-13 shows the comparison between the risk indicators obtained using the clusters and the values obtained from the base case in MWh. The cluster risk indicators for lost load before and after re-dispatch are 14,23% and 20,8% respectively smaller than the values obtained from the base case. On the other hand, the expected re-dispatch is 5% larger using the clusters than the value from the base case.

Figure 4-14 shows the comparison of the risk indicators in terms of cost between the analysis of the clusters and the base case. The differences from the cost risk indicators obtained from the clusters and the cost risk indicators from the base case are 14,18% for risk of lost load before re-dispatch, 20,7% for lost load after re-dispatch, 4,9% for the expected re-dispatch and 16,3% for the total value after re-dispatch which is formed by the lost load after re-dispatch and the expected re-dispatch.

The differences between the results from the analysis of these clusters and the base case are bigger than in the previous case. The maximum difference is 20,8% and the minimum difference is 5%. A difference of 20% between the results from the cluster analysis and the base case is more representative, The error introduced is considerable. However, the results still give a clear idea of the general state of the grid.

As for the execution time, The total execution time needed for this analysis was 18 hour and 19 minutes. This time corresponds to a reduction of 89,9% of the execution time from the base

Results for ClusterH750	52 Clusters
Not N-1 secure	
Quantity of lost load before re-dispatch	5,16 TWh
Cost of lost load before re-dispatch	56,78 Billion
Probability of lost load without re-dispatch	1,052exp-3
Risk of lost load without re-dispatch (Quantity)	407,38 MWh
Risk of lost load without re-dispatch (Cost)	4,48 Million
Quantity of re-dispatch	68,07 TWh
Cost of re-dispatch	6,12 Billion
Probability of re-dispatch	7,92exp-4
Expected amount of re-dispatch	5,25 GWh
Quantity of lost load after re-dispatch	1,73 TWh
Cost of lost load after re-dispatch	19,03 Billion
Probability of lost load with re-dispatch	2,60exp-4
Risk of lost load with re-dispatch (Quantity)	149,27 MWh
Risk of lost load with re-dispatch (Cost)	1,64 Million
Total cost after re-dispatch	**25,1 Billion**
Execution Time	**18 hours 19 minutes**

Table 4-5: State enumeration ClusterH750

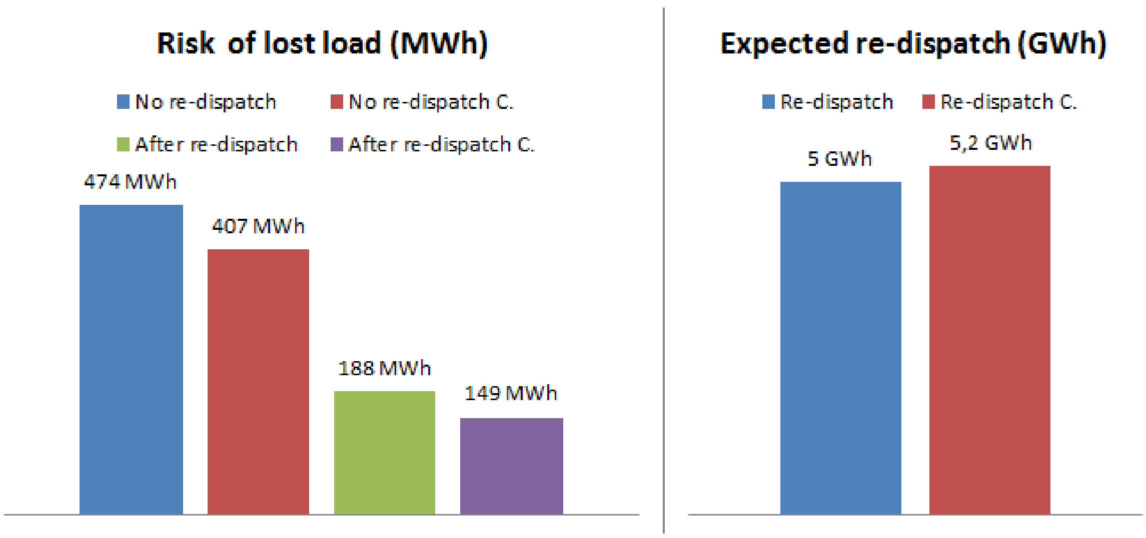


Figure 4-13: Comparison risk of lost load and expected re-dispatch.

case and a 35,24% reduction from the previous case. From the execution time perspective, this cluster is a good option for a quick analysis that can give insight about the performance of the system. Time reduces to a around 10% the original time at the expense of around 21% of maximum error. The next distance to be analyzed is 1000MW.

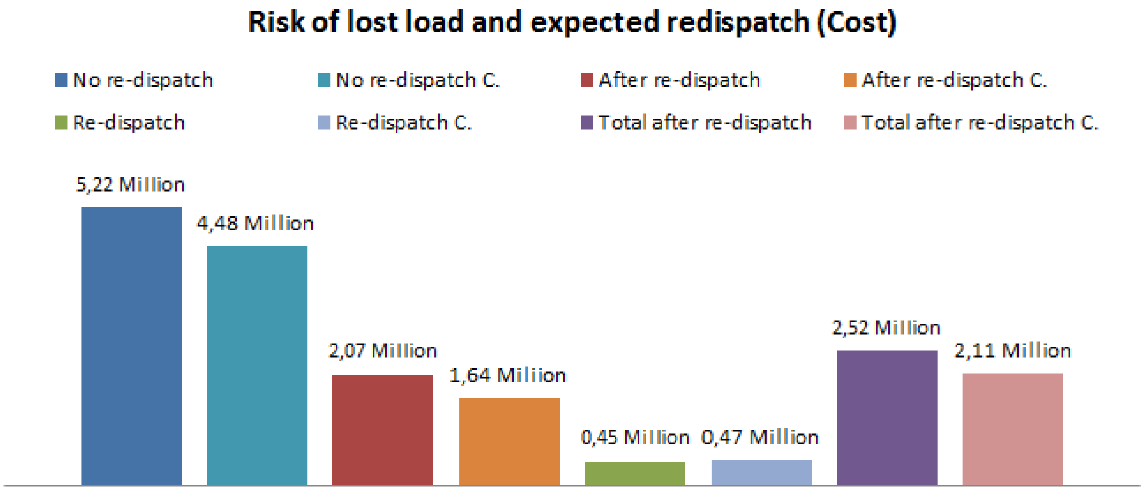


Figure 4-14: Comparison of risk of cost of lost load and expected re-dispatch (cost).

Clusters with 1000 MW as maximum distance between centroids and elements (ClusterH1000)

This maximum distance results in 22 Clusters. The corresponding state enumeration results are presented in table 4-6.

Results for Cluster H1000	**22 Clusters**
Not N-1 secure	
Quantity of lost load before re-dispatch	4,68 TWh
Cost of lost load before re-dispatch	51,47 Billion
Probability of lost load without re-dispatch	1,07exp-3
Risk of lost load without re-dispatch (Quantity)	368,6 MWh
Risk of lost load without re-dispatch (Cost)	4,06 Million
Quantity of re-dispatch	66,13 TWh
Cost of re-dispatch	5,95 Billion
Probability of re-dispatch	8,28exp-4
Expected amount of re-dispatch	4,73 GWh
Quantity of lost load after re-dispatch	1,34 TWh
Cost of lost load after re-dispatch	14,74 Billion
Probability of lost load with re-dispatch	2,40exp-4
Risk of lost load with re-dispatch (Quantity)	116,74 MWh
Risk of lost load with re-dispatch (Cost)	1,28 Million
Total cost after re-dispatch	**20,69 Billion**
Execution Time	**6 hours 30 minutes**

Table 4-6: State enumeration Cluster H1000

Figure 4-15 presents the comparison between the risk indicators for lost load in MWh from the base case and the indicators from the cluster analysis. As expected, the risk indicators resulting from the analysis of the clusters are more distant from the base values than in the previous cases. The lost load indicators from the analysis of the clusters before and after re-dispatch are 22,2% and 37,8% smaller respectively. The expected re-dispatch from the clusters is 5% smaller than the value from the base case.

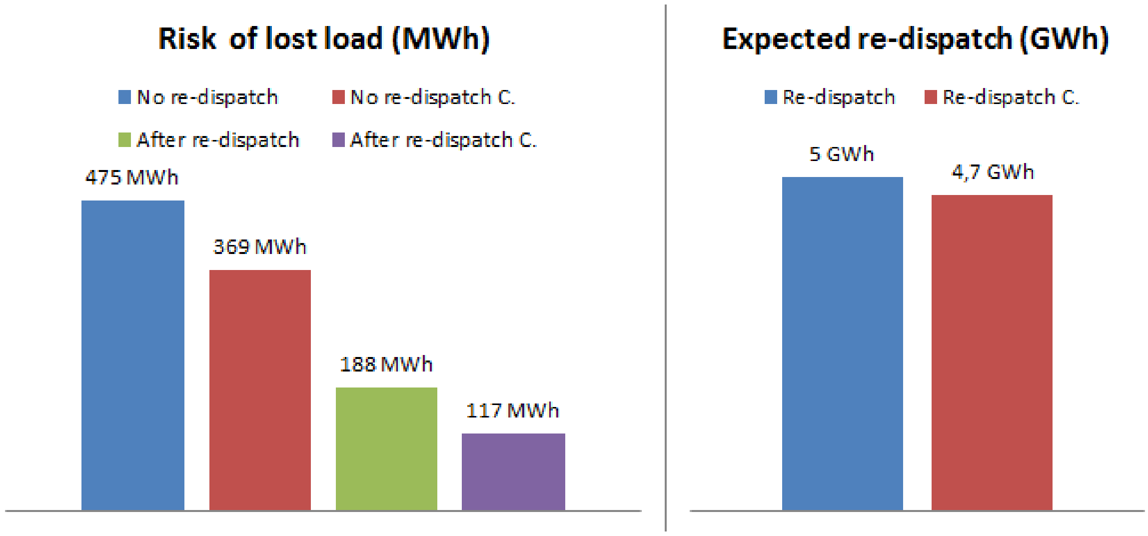


Figure 4-15: Comparison, risk of lost load and expected re-dispatch.

Figure 4-16 presents the risk indicators in terms of cost. The difference between the indicators from the cluster analysis are 22,2% for lost load before re-dispatch, 37,8% for lost load after re-dispatch, 5,4% for re-dispatch and finally for the total risk including lost load and re-dispatch the indicator is 32% smaller than that from the base case.

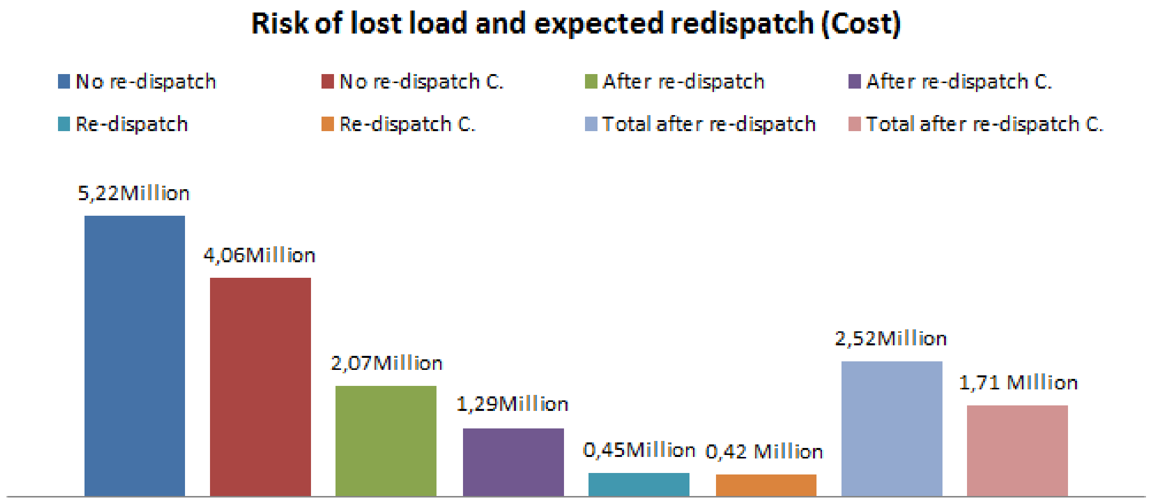


Figure 4-16: Comparison of risk of cost of lost load and expected re-dispatch (cost).

The performance of the analysis of these clusters shows that the risk indicator values are far from those from the base case. The maximum difference is around 38%. The error introduced is quite high. In terms of the execution time, the total time needed for the analysis of the clusters is 6 hours and 30 minutes. This time represents a reduction of 96,43% from the time needed for the base case. The analysis of clusters with this distance can provide fast results at the expense of considerable errors. This analysis can be used for a preliminary analysis dedicated to give an idea of what can be expected from the grid under certain conditions before conducting a more detailed analysis which might imply the analysis of more clusters or the whole set of data.

The poor performance of this analysis suggest that the elements present in the cluster have a considerable distance with the centroid and between each other. The analysis of a random element from one cluster might produce quite different results from the analysis of the centroid. This situation will be illustrated by the analysis of the centroid and a element from a cluster belonging to the set H1000.

Analysis of a individual cluster from ClusterH1000: The set ClusterH1000 has 22 elements. For the present analysis cluster 1 is selected. Cluster 1 has 29 elements. Table 4-7 shows the results from the analysis of the centroid and the sixth element from the cluster which corresponds to the hour 97 from the 1008 set of data. The analysis corresponds to whole set of first order contingencies for 1 hour. The percentage of difference is calculated taken the value from the centroid as reference.

Cluster 1 from ClusterH1000	**Centroid**	**Hour 97**	**Difference**
Not N-1 secure			
Quantity of lost load before re-dispatch	13,22 GWh	8,21 GWh	37,9%
Cost of lost load before re-dispatch	145,4 Million	90,38 Million	37,8%
Risk of lost load without re-dispatch (Quantity)	1,18 MWh	0,65 MWh	55%
Risk of lost load without re-dispatch (Cost)	12980	7150	44,9%
Quantity of re-dispatch	98,34 GWh	19,11 GWh	80,5%
Cost of re-dispatch	8,85 Million	1,72 Million	80,5%
Expected amount of re-dispatch	8,66 MWh	0,66 MWh	92,4%
Quantity of lost load after re-dispatch	8,94 GWh	7,18 GWh	19,7%
Cost of lost load after re-dispatch	98,4 Million	79 Million	19,7%
Risk of lost load with re-dispatch (Quantity)	0,66 MWh	0,62 MWh	6,1%
Risk of lost load with re-dispatch (Cost)	7260	6820	6,1%

Table 4-7: State enumeration cluster 1 from ClusterH1000

As it can be seen from the column "Difference" from table 4-7, the differences are quite high ranging from 6% to 92%. These results confirm the idea that the elements belonging to a particular cluster might produce different results than the centroid which explains the difference between the results from the analysis of the clusters and the base case.

The next section repeats with the performance analysis of clusters, but this time the performance of the K-means algorithm will be discussed.

4-3-2 Performance of k-means clustering

In order to cluster by using the k-means algorithm, it is necessary to specify the number of clusters required. The specification of the number of clusters require a prior knowledge of the data set, in this case the dispatch/load scenario. If there is not a previous reference, then it is difficult to estimate a number of clusters for a certain scenario. One option is the trial and error. First a number of clusters can be specified and the results from the cluster analysis can be later evaluated to determine if the number of clusters was good enough. This option has the disadvantage that it might require quite some time and effort. Also, the fact that, given the inner workings of the k-means algorithm that assigns the first centroid at random, the clusters are rarely the same when they are constructed again and this can complicate the selection for the number of clusters.

Since the Hierarchical clustering offers the possibility of choosing an error tolerance in the form of a maximum distance, the option for the specification of number of clusters adopted for this report is to use the number of clusters resulting from the Hierarchical clustering. Three specifications for the number of clusters will be analyzed, which are 141, 52 and 22 clusters.

Clusters K-means 141 elements (ClusterK141)

The state enumeration results for the 141 K-means clusters are presented in table 4-8. All cost values are given in Euros.

Results for Cluster K141	**141 Clusters**
Not N-1 secure	
Quantity of lost load before re-dispatch	5,67 TWh
Cost of lost load before re-dispatch	62,40 Billion
Probability of lost load without re-dispatch	1,03exp-3
Risk of lost load without re-dispatch (Quantity)	450,8 MWh
Risk of lost load without re-dispatch (Cost)	4,96 Million
Quantity of re-dispatch	67,74 TWh
Cost of re-dispatch	6,1 Billion
Probability of re-dispatch	7,75exp-4
Expected amount of re-dispatch	5,09 GWh
Quantity of lost load after re-dispatch	1,97 TWh
Cost of lost load after re-dispatch	21,67 Billion
Probability of lost load with re-dispatch	2,60exp-4
Risk of lost load with re-dispatch (Quantity)	165,1 MWh
Risk of lost load with re-dispatch (Cost)	1,81 Million
Total cost after re-dispatch	**27,77 Billion**
Execution Time	**45 hours 7 minutes**

Table 4-8: State enumeration Cluster K141

Figure 4-17 presents the comparison between the risk indicators for lost load and re-dispatch in MWh from the base case and the indicators obtained from the cluster analysis. As in the analysis of the previous section, the indicators from the cluster analysis are represented by a C. next to the name. The risk of lost load before re-dispatch from the cluster analysis is 5% smaller than the reference from the base case. The risk of lost load after re-dispatch from the cluster analysis is 12,3% smaller than the indicator from the base case and lastly, the expected re-dispatch from the clusters is 1,8% bigger than the value from the base case.

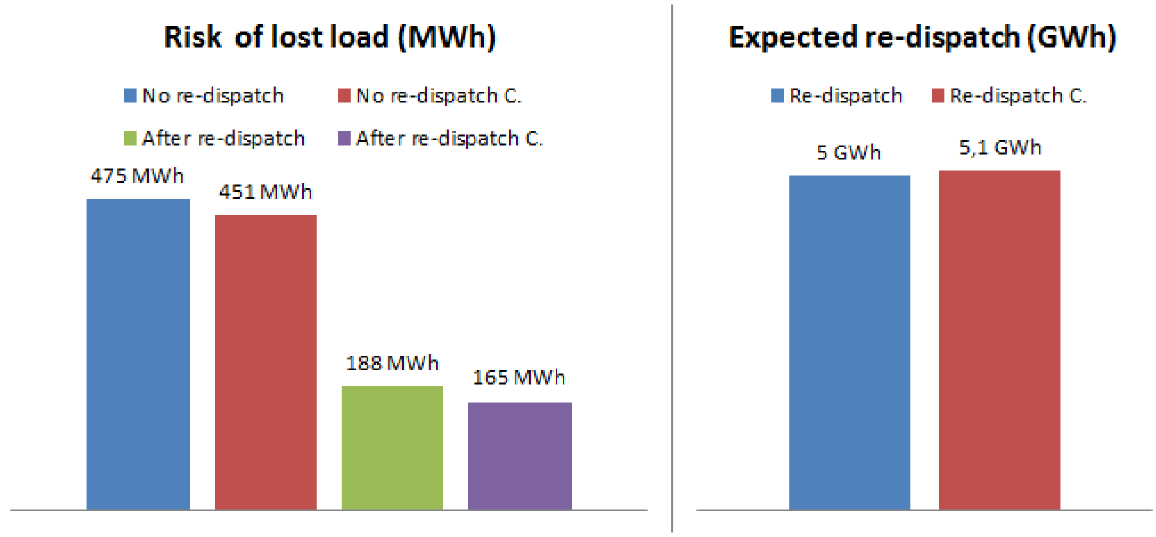


Figure 4-17: Comparison, risk of lost load and expected re-dispatch.

Figure 4-18 present the comparison between the risk indicators in terms of cost of lost load from the cluster analysis and from the base case. The differences from the cluster analysis are the same than before which is 5% smaller for lost load before re-dispatch, 12,3% smaller for lost load after re-dispatch and 1,8% bigger for the expected re-dispatch. The total cost after re-dispatch (lost load and re-dispatch) from the cluster analysis is 10% smaller than the value from the base case.

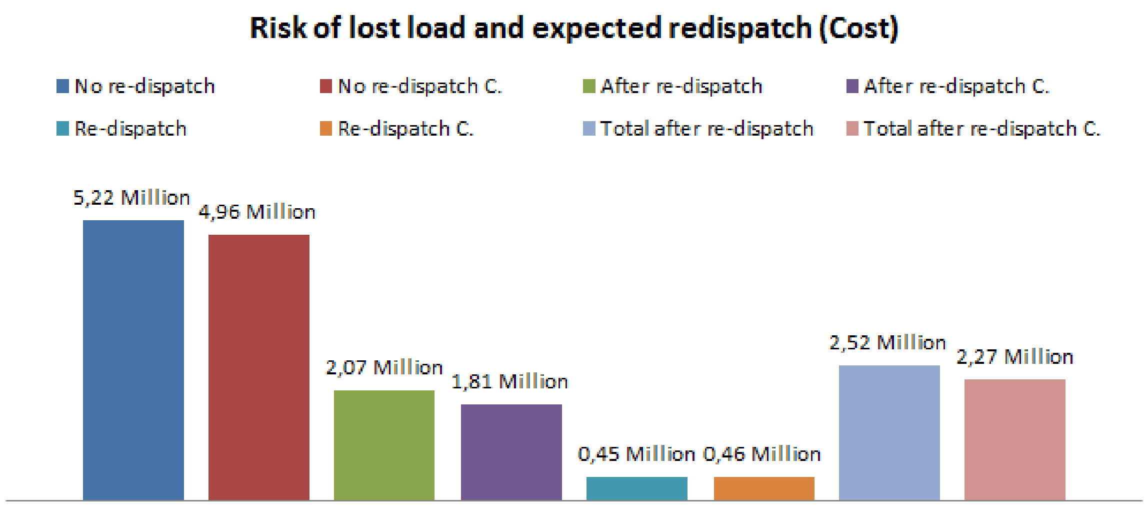


Figure 4-18: Comparison of risk of cost of lost load and expected re-dispatch (cost).

The results show that the maximum difference from the analysis of the clusters is 12,3% and the minimum difference is 1,8%. The total execution time for this analysis is 45 hours and 7 minutes. The execution time from the cluster analysis is reduced 75,2% from the execution time of the base case. This saving in time comes at the expense of a maximum error of 12,3%. The time saved and the error introduced makes this number of clusters a great option for the analysis. If more accuracy in the results is needed, more clusters can be employed. From the results, it is possible to infer that the elements and the centroid of every cluster are very similar. The elements are close to each other.

The next amount of clusters to be analyzed is 52 clusters.

Clusters K-means 52 elements (ClusterK52)

The results from the state enumeration analysis of these 52 clusters are presented in table 4-9.

Results for Cluster K52	**52 Clusters**
Not N-1 secure	
Quantity of lost load before re-dispatch	5,38 TWh
Cost of lost load before re-dispatch	59,22 Billion
Probability of lost load without re-dispatch	1,04exp-3
Risk of lost load without re-dispatch (Quantity)	427,8 MWh
Risk of lost load without re-dispatch (Cost)	4,71 Million
Quantity of re-dispatch	67,88 TWh
Cost of re-dispatch	6,11 Billion
Probability of re-dispatch	7,83exp-4
Expected amount of re-dispatch	5,08 GWh
Quantity of lost load after re-dispatch	1,71 TWh
Cost of lost load after re-dispatch	18,86 Billion
Probability of lost load with re-dispatch	2,59exp-4
Risk of lost load with re-dispatch (Quantity)	151,1 MWh
Risk of lost load with re-dispatch (Cost)	1,66 Million
Total cost after re-dispatch	**24,97 Billion**
Execution Time	**15 hours 55 minutes**

Table 4-9: State enumeration Cluster K52

Figure 4-19 presents the comparison between the risk indicators from the cluster analysis and the values obtained from the base case in MWh. The risk indicators obtained from the clusters for lost load before and after re-dispatch are 9,8% and 19,8% respectively smaller than the values obtained from the analysis of the base case. The expected re-dispatch from the clusters is 1,6% larger than the value from the base case.

Figure 4-20 shows the comparison of the risk indicators between the analysis of the clusters and the base case in terms of costs. The differences from the cost risk indicators obtained from the clusters and the cost risk indicators from the base case are the same as above. The total value after re-dispatch which is formed by the lost load after re-dispatch and the expected re-dispatch is 16% smaller for the analysis of the clusters.

As expected, the differences between the results from the analysis of the clusters and the base case increased from the previous case. The maximum difference is 19,8% and the minimum difference is 1,6%. A difference of 20% implies a considerable error. However, the results provide a solid idea about the general performance of the grid.

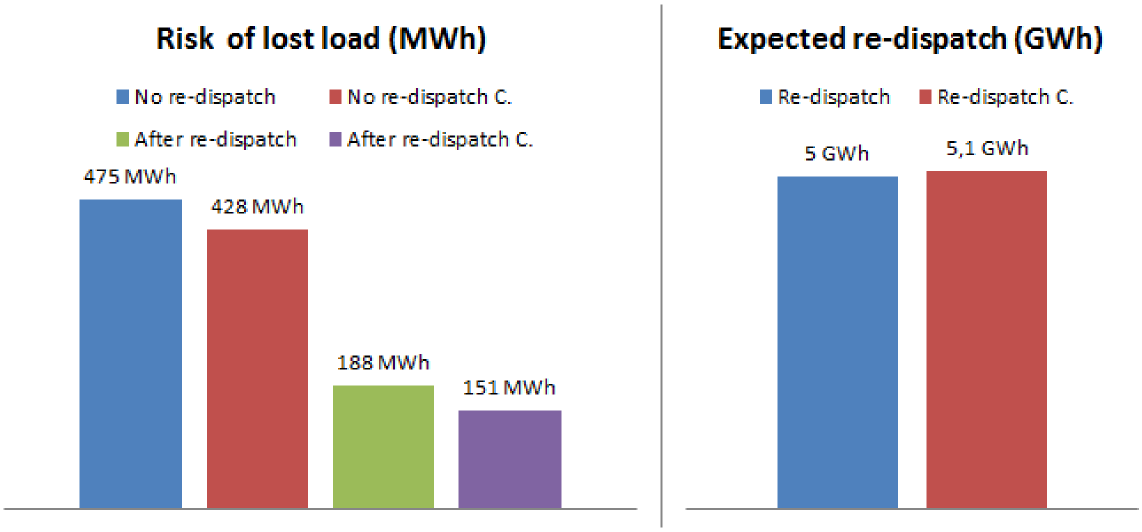


Figure 4-19: Comparison, risk of lost load and expected re-dispatch.

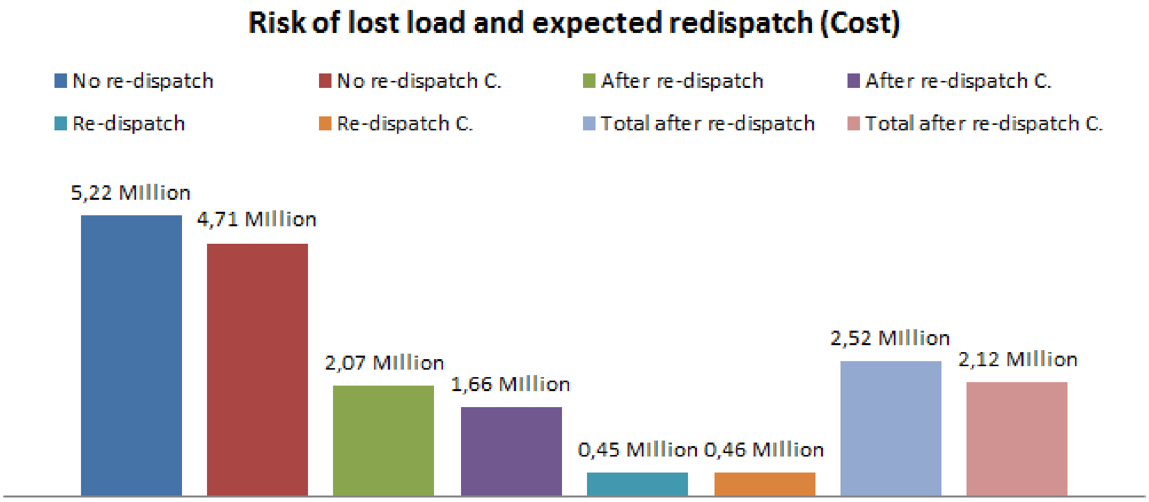


Figure 4-20: Comparison of risk of cost of lost load and expected re-dispatch (cost).

The total execution time needed for this analysis was 15 hour and 55 minutes. This time represents a reduction of 91,25% of the execution time from the base case. This reduction of the execution time turns this cluster into a good option for a fast analysis that can give insights about the performance of the system. The time reduction in achieved at the expense of around 20% of maximum error. The next amount of clusters to be analyzed is 22 clusters.

Clusters K-means 22 elements (ClusterK22)

The results from the state enumeration analysis of the 22 clusters are presented in table 4-10.

Figure 4-21 presents the comparison between the risk indicators for lost load in MWh from the cluster analysis and the ones from the base case. The risk indicators resulting from the

Results for Cluster K22	22 Clusters
Not N-1 secure	
Quantity of lost load before re-dispatch	5,09 TWh
Cost of lost load before re-dispatch	56,01 Billion
Probability of lost load without re-dispatch	1,04exp-3
Risk of lost load without re-dispatch (Quantity)	404,35 MWh
Risk of lost load without re-dispatch (Cost)	4,44 Million
Quantity of re-dispatch	68,03 TWh
Cost of re-dispatch	6,12 Billion
Probability of re-dispatch	8,24exp-4
Expected amount of re-dispatch	5,1 GWh
Quantity of lost load after re-dispatch	1,62 TWh
Cost of lost load after re-dispatch	17,76 Billion
Probability of lost load with re-dispatch	2,17exp-4
Risk of lost load with re-dispatch (Quantity)	139,83 MWh
Risk of lost load with re-dispatch (Cost)	1,54 Million
Total cost after re-dispatch	**23,89 Billion**
Execution Time	**4 hours 43 minutes**

Table 4-10: State enumeration Cluster K22

analysis of these set of clusters are more distant from the base values than in the previous cases. The lost load indicators from the analysis of the clusters before and after re-dispatch are 14,78% and 25,60% smaller respectively. The expected re-dispatch from the clusters is 2% larger than the value from the base case.

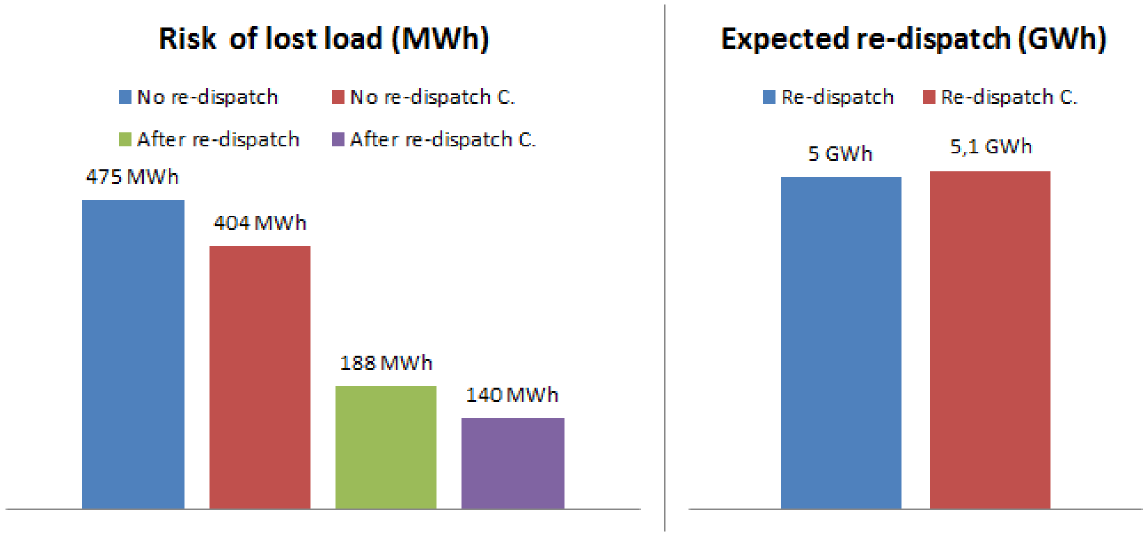


Figure 4-21: Comparison, risk of lost load and expected re-dispatch.

Figure 4-22 presents the risk indicators in terms of cost. The difference between the indicators from the cluster analysis are the same as the ones from figure 4-21.The total risk indicator including lost load and re-dispatch is 20,67% smaller than the value from the base case.

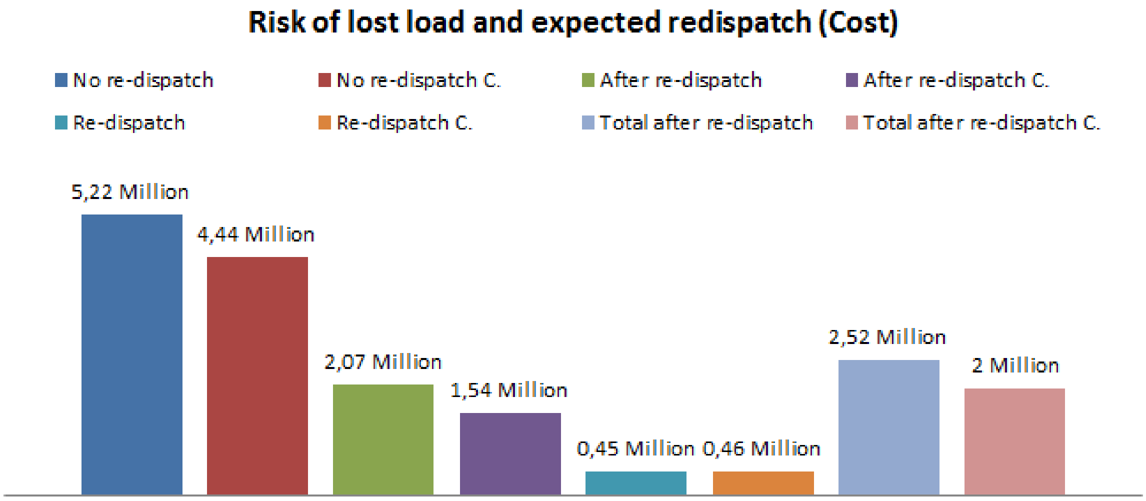


Figure 4-22: Comparison of risk of cost of lost load and expected re-dispatch (cost).

The maximum difference between the indicators from the clusters and those from the base case is around 26%. The error introduced is high. The performance of the analysis of these clusters shows that the risk indicator values are somewhat far from the ones obtained in the base case. In terms of the execution time, the total time needed for the analysis of the clusters is 4 hours and 43 minutes. This time represents a reduction of 97,4% from the time needed for the base case. However, it is important to keep in mind that this reduction comes at a maximum error of 26%. This analysis of 22 clusters offers fast results at the expense of considerable errors. This analysis can be used for a preliminary analysis for a quick overview of the state of the system and general behaviour.

The performance is rather poor in terms of errors introduced. From this, it is possible to deduce that the elements present in a cluster have a considerable distance with the centroid. Just as in the hierarchical case, the analysis of a random element from one cluster and the centroid might produce different results. The analysis of a individual cluster is presented next in order to understand more about the performance.

Analysis of a individual cluster from ClusterK22: For the present analysis cluster 17 is selected among the 22 clusters. Cluster 17 has 30 elements. Table 4-11 shows the results from the analysis of the centroid and the element number 23 from the cluster which corresponds to the hour 527 from the 1008 set of data. The analysis corresponds to whole set of first order contingencies for 1 hour. The percentage of difference is calculated taken the value from the centroid as reference.

From the column "Difference", it is possible to see that the differences are quite high. These differences range from 29,5% to 105,9%. For this cluster, the results from centroid indicate that there are not lost load after re-dispatch. However, the analysis of hour 23 indicates that there is some lost load after re-dispatch. The error related to the lost load after re-dispatch tends to infinity.

For all the clusters analyzed, the maximum difference can be found linked to the indicators

Cluster 1 from ClusterH1000	Centroid	Hour 97	Difference
Not N-1 secure			
Quantity of lost load before re-dispatch	3,34 GWh	5,19 GWh	55,4%
Cost of lost load before re-dispatch	36,8 Million	57,18 Million	55,4%
Risk of lost load without re-dispatch (Quantity)	0,17 MWh	0,35 MWh	55%
Risk of lost load without re-dispatch (Cost)	1870	3850	105,9%
Quantity of re-dispatch	67,9 GWh	47,87 GWh	29,5%
Cost of re-dispatch	6,1 Million	4,31 Million	29,5%
Expected amount of re-dispatch	4,94 MWh	2,33 MWh	52,8%
Quantity of lost load after re-dispatch	0 GWh	2,6 GWh	N/A
Cost of lost load after re-dispatch	0 Million	28,78 Million	N/A
Risk of lost load with re-dispatch (Quantity)	0 MWh	0,23 MWh	N/A
Risk of lost load with re-dispatch (Cost)	0	2530	N/A

Table 4-11: State enumeration cluster 17 from ClusterK22

of lost load after re-dispatch. The reason for that is that in some cases the centroid may yield no lost load after re-dispatch even if there is some lost load for some elements of the cluster.

Again, just as in Hierarchical clustering, the results reinforce the idea that the elements belonging to a particular cluster might produce different results than the centroid.

Until now, two methods for clustering were tested. The next section presents a comparison of the performance of the two methods.

4-3-3 Comparison of the Hierarchical and K-means methods for clustering

A comparison between the performance of the Hierarchical and the K-means cluster methods is presented in table 4-12. The comparison is performed taking as a reference the results from the base case. The maximum difference represents the largest deviation in the indicators found after the cluster analysis with respect to the indicators from the base case. The minimum difference represents the smallest deviation. The time reduction is calculated using the execution time of the base case as a reference.

Comparison of performance	H500	K141	H750	K52	H1000	K22
Number of Clusters	141	141	52	52	22	22
Maximum difference	15,9%	12,3%	20,8%	19,8%	37,8%	25,6%
Minimum difference	5,5%	1,8%	5%	1,6%	5%	2%
Execution Time	50 Hours 37 Min.	45 Hours 7 Min	18 Hours 19 Min	15 Hours 55 Min	6 Hours 30 Min.	4 Hours 43 Min.
Time reduction	71,6%	75,19%	89,9%	91,25%	96,43%	97,40%

Table 4-12: Comparison between cluster methods

Figure 4-23 illustrates both, the maximum distance for the hierarchical and the K-means method as well as the reduction in simulation time. According to the results, the K-means method is slightly superior than the Hierarchical method, especially with the cluster with

22 elements where the maximum difference goes from 37,8% in the Hierarchical method to 25,60% in the K-means method. The maximum error is quite similar in the other cases with almost no difference.

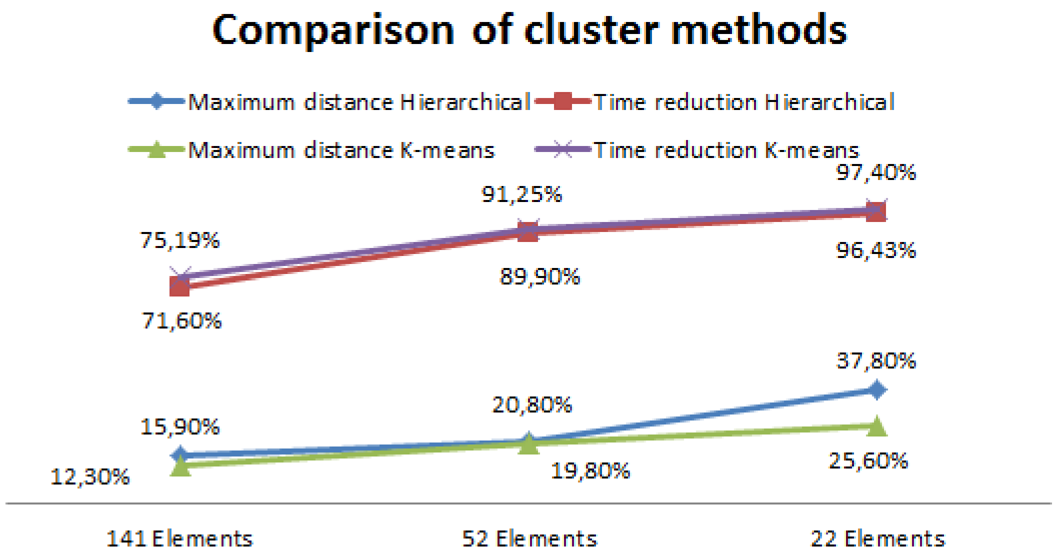


Figure 4-23: Comparison of the performance of cluster methods.

In terms of computational time, the K-means method is slightly faster although the difference is minimal. In general terms, the two methods can produce good results. The Hierarchical method can be used for the determination of the number of clusters and the K-means method can be used for the actual construction of the clusters. Regardless of the method used, clustering is a viable option that can produce quite accurate results while saving simulation time and computational resources.

The next section will present a decomposition of the lost load and risk per contingency after the state enumeration analysis.

4-4 Lost load and risk per contingency

The state enumeration analysis produces as a result a number of global indicators for lost load, re-dispatch, costs and the risk of lost load expressed as energy and costs. The global indicators describe the performance of the grid under certain conditions over a predefined time frame. A more detailed analysis of the indicators can provide information about which contingencies contribute the most to the risk indicators. This section presents an analysis of the contribution of the contingencies to the global indicators for lost load and risk of lost load. The aim of this analysis is to determine which line outages produce more risk. The analysis is performed using the information resulting from the state enumeration analysis of the K-means cluster with 141 elements.

Figure 4-24 presents a radial representation of the most representative contingencies that produce lost load before and after re-dispatch. From the figure, it is possible to see which

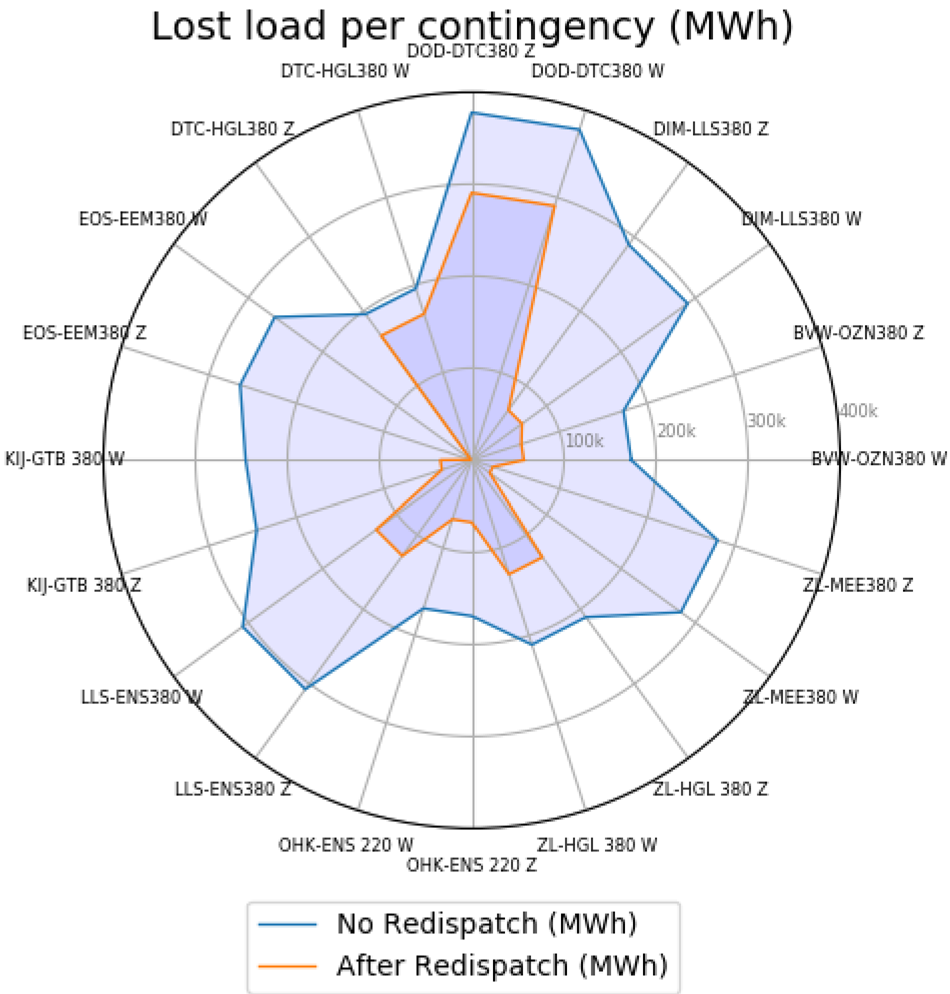


Figure 4-24: Lost load per contingency

lines contribute more to the lost load indicators. The blue line represents the lost load without re-dispatch and the orange line the lost load after re-dispatch. The lines DOD-DTC380 W/Z are the top contributing lines for the lost load before and after re-dispatch. The region delimited by the blue line serves as a visual representation of the lost load without considering re-dispatch. The region delimited by the orange line illustrates the lost load considering re-dispatch.

The region considering re-dispatch is clearly smaller than the no re-dispatch region. This indicates that executing re-dispatch reduces the amount of lost load produced by the contingencies. For some of the contingencies, this reduction is quite pronounced as is the case of the lines EOS-EEM380 W/Z. The effect of re-dispatch has a huge positive impact in the grid reducing the lost load in every line. For some lines the reduction is favorable going from a high value to a very low one.

By looking strictly from the lost load perspective, presented by figure 4-24, one logical conclusion would be that lines DOD-DTC380 W/Z are the most insecure ones and some action should be taken in order to improve the situation. This action could be a reinforcement for

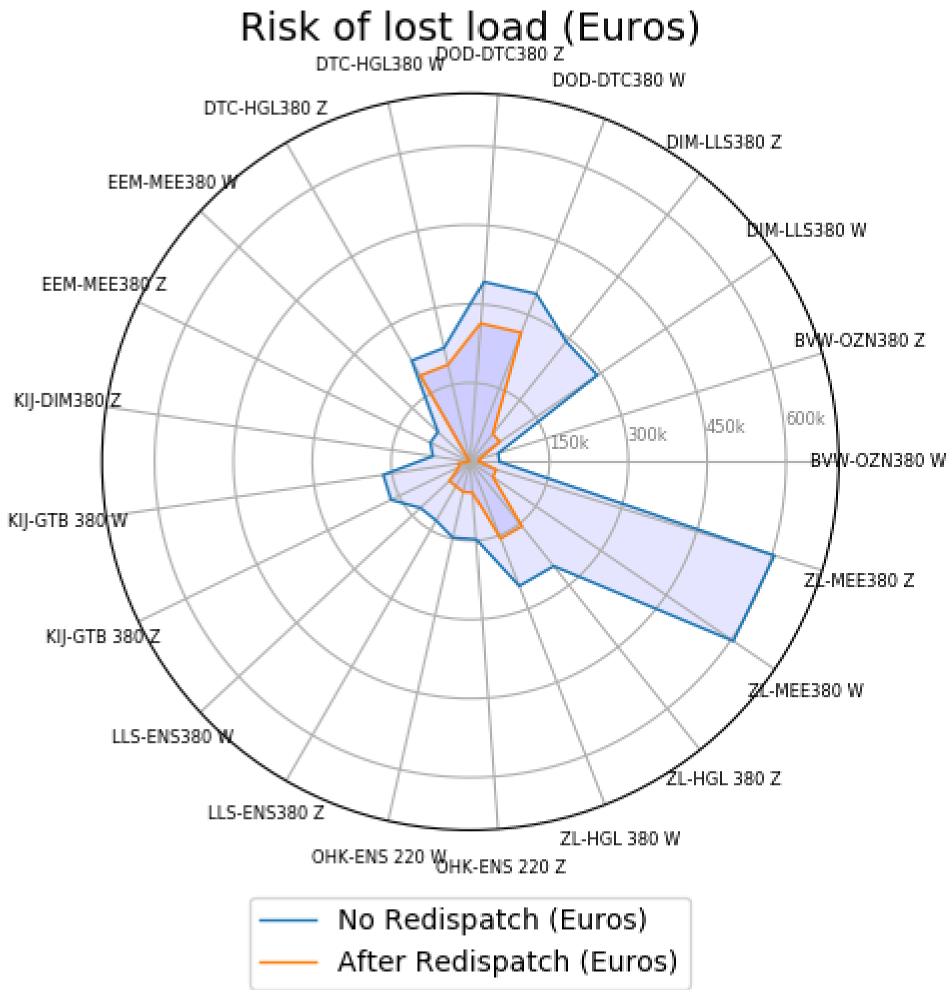


Figure 4-25: Risk of lost load per contingency

the lines or a new line to increase the transport capacity. In any way, the added capacity needed to solve the problem will differ greatly if re-dispatch is not considered in the analysis.

Figure 4-25 presents a visual representation of the risk (cost of lost load) contribution per line. The region delimited by the blue line represents the risk without the consideration of re-dispatch and the region delimited by the orange line represents the risk considering re-dispatch actions. Again, in this figure is possible to see that the risk after re-dispatch reduces significantly from the risk without re-dispatch. In some cases the risk becomes quite small.

According to the data displayed in figure 4-25, before re-dispatch, the lines that generate the greatest risk are the lines ZL-MEE380 W/Z. Looking back to figure 4-24 without re-dispatch, the lines DOD-DTC380 W/Z were the ones that produced the greatest lost load. This outcome suggest that not necessarily the contingency that results in the greatest lost load is the one that produces more risk. Since risk is calculated using the probability that the component is found unavailable at a determined time (unavailability), then the unavailability of lines ZL-MEE380 W/Z is bigger than that of lines DOD-DTC380 W/Z resulting in the outcome depicted in figure 4-25.

In contrast, looking at figure 4-24, if re-dispatch is taken into account, then the lost load of the lines ZL-MEE380 W/Z is almost entirely corrected. As a result, the risk after re-dispatch is minimum for those lines as is depicted in figure 4-25. Looking back at lines DOD-DTC380 W/Z in figure 4-24, the lost load indeed reduces but it is still high. As a result the risk of lost load for those lines is the highest.

This kind of risk analysis and the visual representations constitute tools that can aid in the process of making decisions. If a reinforcement decision was to be taken, without considering re-dispatch, the best option would be to reinforce lines ZL-MEE380 W/Z. However, if re-dispatch is considered, then the best option would be the lines DOD-DTC380 W/Z.

The numerical values values of the lost load per line can be found in table 4-13. The numerical values of the risk of lost load can be found in table 4-14. The data of lost load and risk from the tables is sorted from the lines with the highest lost load before re-dispatch to the lowest.

Contingencies before and after Re-dispatch		
Contingency	**Lost load (MWh) Before re-dispatch**	**Lost load (MWh) After re-dispatch**
DOD-DTC380 W/Z	377854,9	290561,4
LLS-ENS380 W/Z	307416,8	127983,7
DIM-LLS380 W/Z	289627,8	67144,7
ZL-MEE380 W/Z	280640,2	23780,8
EOS-EEM380 W/Z	264414,8	1133,7
KIJ-GTB 380 W/Z	245505,1	33917,6
ZL-HGL 380 W/Z	210511,9	130022,1
DTC-HGL380 W/Z	196079,8	167089,2
BVW-OZN380 W/Z	173170,2	56312,8
OHK-ENS 220 W/Z	169044,3	67211,2

Table 4-13: Lost load per contingencies MWh

Contingencies before and after Re-dispatch		
Contingency	**Risk (Euro) Before re-dispatch**	**Risk (Euro) After re-dispatch**
ZL-MEE380 W/Z	604,83 k	51,24 k
DOD-DTC380 W/Z	342,76 k	263,58 k
DIM-LLS380 W/Z	291,90 k	67,67 k
ZL-HGL 380 W/Z	253,68 k	156,68 k
DTC-HGL380 W/Z	221,95 k	189,13 k
KIJ-GTB 380 W/Z	166,88 k	23,05 k
OHK-ENS 220 W/Z	147,90 k	58,80 k
LLS-ENS380 W/Z	128,98 k	53,70 k
EEM-MEE380 W/Z	83,26 k	1,83 k
KIJ-DIM380 Z	71,50 k	2,98 k

Table 4-14: Risk per contingencies Euro

This type of risk analysis can be very useful at the moment of diagnosing the state of the grid.

However, it is not ready yet to serve as the main reliability criterion and to take decisions based only on its results. This type of analysis can be used along with actual methods. This method can provide information of which lines to reinforce or the priority of the lines to be intervened that along with current analysis such as the risk matrix can steer the decisions to take into a effective decision by attacking first the more risky situations. Additional radial representations for different situations can be found in appendix C.

The next section will discuss the residual risk calculation from the state enumeration analysis.

4-5 Residual risk assessment and contingency selection

This section will discuss the topic of contingency selection based on the probability of not considered states and residual risk calculation.

4-5-1 Residual risk and probability of not considered states

In the state enumeration analysis performed, it is assumed that the consideration of the first order contingencies is enough to get representative reliability indicators for the whole system. The first order contingencies are the ones with the higher probability of occurrence. Higher order contingencies are less probable to occur but their impact over the grid can be way more negative than the impact of the first order contingencies. The residual risk calculation provides a notion of the risk produced by the higher order contingencies that were not included in the state enumeration analysis.

The residual risk is calculated by the product of the probability of not considered states times the worst case effect. For this report the probability of the not considered states is calculated using the availability and the unavailability of lines. The worst case scenario for the impact is assumed to be a blackout and is calculated as the average of the sum of all loads connected to the grid through the year.

The probability of the not considered states is calculated as shown in equation 4-6, where $P_{not-considered}$ is the total probability of the not considered states and $P_{Considered}$ is the probability of the considered states.

$$P_{not-considered} = 1 - P_{considered} \tag{4-6}$$

If the analysis is up to first order contingencies, then the not considered probability is calculated as shown in equation 4-7, where $P(0^{th})$ is the probability of no contingencies and $P(1^{st})$ is the probability of all first order contingencies.

$$P_{not-considered} = 1 - P(0^{th}) - P(1^{st}) \tag{4-7}$$

The probability $P(0^{th})$ is calculated as the product of the availabilities of all the lines and the probability $P(1^{st})$ is calculated as the sum of all the state probabilities as shown in equation 4-8, where n is the total number of first order states and $P_i(State)$ is the probability of the state i.

$$P(1^{st}) = \sum_{i=1}^{n} P_i(State) \tag{4-8}$$

The probability of the states is calculated as shown in equation 4-9, where U_i is the unavailability of line i and A_k is the availability of the rest of the lines. The availabilities can be easily replaced by unavailabilities by the relationship $A = 1 - U$

$$P_i(State) = U_i * \prod_{k=1}^{n} A_k, n \neq k \tag{4-9}$$

4-5-2 Contingency selection based on the probability of not considered states

Table 4-15 presents the state probabilities and the probabilities of not considered states calculated up to second order contingencies. The unavailability used for the calculation is obtained using the same failure frequency as in the state determination analysis.

Probability of not considered states		
Contingencies	**State Probability**	**Probability of not considered states**
No contingencies	0,99443	0,005566
First order	5,30exp-3	0,00026435
Second order	1,954exp-5	0,00025046

Table 4-15: Probability of not considered states

The state probability, from table 4-15 indicates that the grid will be found without contingencies almost always. The probability of finding the grid with one line outage is 5,30 exp-3 or 1 in 189. The probability of finding the grid with two line outages is 1,95 exp-5 or 1 in 51177. The possibility of finding two lines down in the grid is extremely remote. This data suggest that the grid is quite strong and line outages in this EHV transmission grid may almost never happen.

If the state enumeration analysis includes up to the first order contingencies, then the probability of the rest of the states from second order onwards is 0.00026 or 1 in 3783. If the analysis includes up to second order contingencies, then the rest of the states have a probability of 0.00025 or 1 in 3992. This probabilities are again extremely remote which indicates that a contingency of second order and beyond may never ever happen during the lifetime of the lines.

The residual risk up to second order contingencies is presented in table 4-16. The value for the blackout is a lost load of 21,17GW and it is obtained from the average of the power demand through the 1008 dispatch / load scenario. If the analysis includes up to first order

Residual risk		
Contingencies	**Residual Risk (MW)**	**Residual Risk (Euro)**
First order	5,69	61,57 k
Second order	5,30	58,34 k

Table 4-16: Residual risk

contingencies, then the discarded risk is of 5,89 MW or 61577 Euro. This discarded risk includes all the high order contingencies assuming the worst case scenario of a blackout. If the analysis includes up to second order contingencies, then the discarded risk is of 5,30 MW or 58342 Euro. Events of low probability with high impact are included in the residual risk as well.

These values of residual risk succeed in giving an idea of what is ignored by choice at the moment of executing an analysis including up to first order contingencies. Figure 4-26 shows the residual risk mapped into the risk matrix.

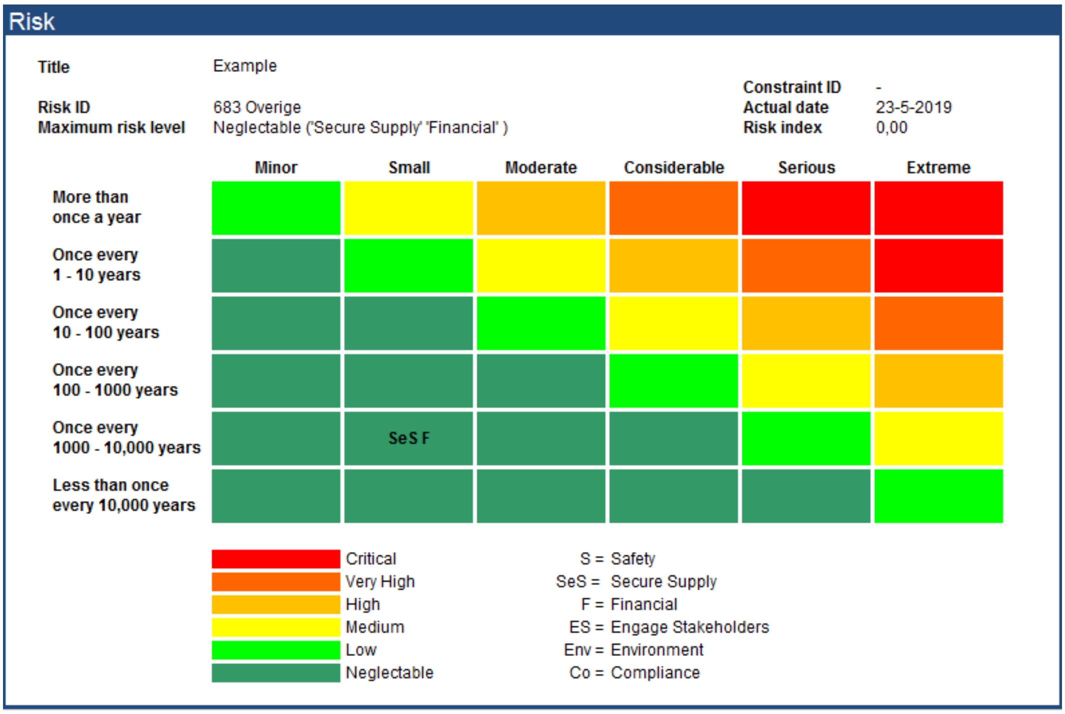


Figure 4-26: Risk matrix, residual risk up to first order

The risk is neglectable. Based on that, the decision of analysing up to first order contingencies is completely justified. There is no need to analyze further than first order contingencies. This does not mean that for specific situations, higher order contingencies might be analyzed anyway.

If the failure probabilities changes due to a change in the unavailability of lines, the residual risk might change. If the change is big enough, then the analysis of contingencies of higher order might or might not be required. The effect of the variation of unavailability over the residual risk is discussed next.

4-5-3 Variation of residual risk due to variation of line unavailability

Several situations can produce a change in the unavailability of transmission lines as was discussed in section 2-4 of chapter 2. The change in the unavailability is calculated by the application of the Bayesian data analysis discussed in sections 3-2 and 3-3 of chapter 3. The posterior distribution found is presented in equation 4-10. This posterior distribution is obtained from the assumption that the lines have already reached their functional lifetime. The distribution function has a mean value of 77.6% with a 90% confidence interval from 73.1% to 81.7%.

$$Posterior = Beta(203.5, 58.83) \tag{4-10}$$

The change in the unavailability is given by the mean value (77.6%) of the posterior distribution. The base unavailability calculated from the failure frequency provided by TenneT is increased by 77,6%. This modification of the unavailability is applied to all the lines.

Table 4-17 presents the state probabilities and the probabilities of not considered states calculated up to second order contingencies with the modified unavailability.

Probability of not considered states		
Contingencies	**State Probability**	**Probability of not considered states**
No contingencies	0,99014	0.00985951
First order	9,41exp-3	0.00048819
Second order	4,53exp-5	0.00044462

Table 4-17: Modified probability of not considered states

The change in the unavailability produces variations in both state probabilities and probability of not considered states. The probability of finding the grid without contingencies decreased in comparison with the value from table 4-15 from 0,99443 to 0,99014. The value is still high. However, the increase in the unavailability of lines produces a decrease in the likelihood of finding the grid without contingencies. The probabilities of finding one and two line outages increased. This means that it is more likely to find line outages of first and second order in the grid. This outcome makes sense since the increase in unavailability implies that the lines fail more.

The values of the probability of not considered states also vary. The values increase which means that the probability of higher order contingencies increase. If the state enumeration analysis include up to first order contingencies, the probability of states from second order onwards increase to 0.000488196 or 1 in 2028. In the same way, if the analysis includes up to second order contingencies the probability of the rest of states increase to 0.00044462 or 1 in 2249. Despite the increase on the probabilities of not considered states, the values are still extremely remote.

Table 4-18 presents the residual risk up to second order contingencies. The value for lost load due to a blackout is the same as in the previous calculation.

The analysis including up to first order contingencies, results in a discarded risk is of 10,34 MW or 113720 Euro. If the analysis includes up to second order contingencies, then the discarded risk is of 9,41 MW or 103570 Euro. The risk values increased from the previous case. Figure 4-27 shows the residual risk mapped into the risk matrix.

The risk is still neglectable. Again, the decision of analysing only up to first order contingencies is justified. In order to get these results, the unavailability of all lines was increased

Residual risk		
Contingencies	**Residual Risk (MW)**	**Residual Risk (Euro)**
First order	10,34	113,72 k
Second order	9,41	103,57 k

Table 4-18: Modified residual risk

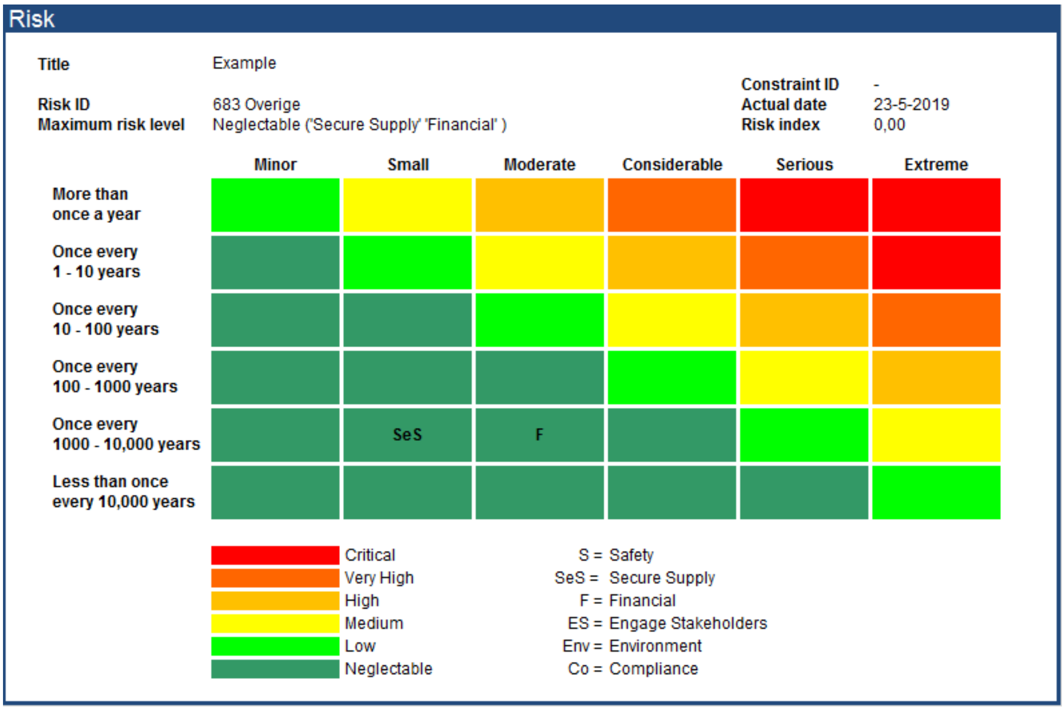


Figure 4-27: Modified risk matrix, residual risk up to first order

assuming that all lines reached their specified lifetime. If the unavailability is increased in one or two lines, the variation on the residual risk and the probabilities is imperceptible.

In any way, the variation is not very representative. However, it has to be kept in mind that the unavailability of lines is not constant, at contrary is dynamic depending on several factors suchs as the weather conditions, location and loading among others. If the variation of the unavailability of lines can be included in a PRA, then the results can be more realistic and eventually they can lead to the detection of risky situations.

If the unavailability changes, not only the residual risk will change, but the whole results from the analysis might change. The next section discusses the variation of the state enumeration results due to the variation of unavailability.

4-6 Variation of state enumeration results due to a variation of unavailability

The variation of state enumeration results is evaluated using the cluster K-means of 22 elements given that it can produce results with a relatively short simulation time. The variation of the unavailabilities is given by the posterior distribution employed on the previous section. The results from the state enumeration analysis are presented in table 4-19. It is possible to see that the indicators that are linked to any probability are the ones that change their value with respect to the results from the K22 cluster presented before using the base unavailability.

Results, modified unavailability	**22 Clusters**
Not N-1 secure	
Quantity of lost load before re-dispatch	5,09 TWh
Cost of lost load before re-dispatch	56,01 Billion
Probability of lost load without re-dispatch	1,84exp-3
Risk of lost load without re-dispatch (Quantity)	714,69 MWh
Risk of lost load without re-dispatch (Cost)	7,86 Million
Quantity of re-dispatch	68,03 TWh
Cost of re-dispatch	6,12 Billion
Probability of re-dispatch	1,46exp-3
Expected amount of re-dispatch	9,04 GWh
Quantity of lost load after re-dispatch	1,62 TWh
Cost of lost load after re-dispatch	17,76 Billion
Probability of lost load with re-dispatch	3,83exp-4
Risk of lost load with re-dispatch (Quantity)	247,15 MWh
Risk of lost load with re-dispatch (Cost)	2,71 Million
Total cost after re-dispatch	**23,89 Billion**
Execution Time	**9 hours 9 minutes**

Table 4-19: State enumeration results. Cluster K22 with modified unavailability

Figure 4-28 presents a comparison between the risk indicators for lost load in MWh obtained with the base unavailability and the indicators obtained with the modified unavailability. The risk indicators in terms of cost are presented in figure 4-29. The indicators obtained using the base unavailability are signed with the word base next to the name. The indicators obtained using the modified unavailability have the word mod next to them.

The risk indicators vary considerably with the modified unavailability. An increase of 77,6% in the unavailability results in a increase of 50,68% in the lost load before re-dispatch indicator, 31,15% increase in the lost load after re-dispatch indicator and a increase of 80,80% in the expected re-dispatch indicator. The higher the increase in the unavailability, the higher the risk. This result is logical, more risk is expected if the lines are more likely to fail.

The indicators of expected re-dispatch present the maximum variation (80,0%). This indicates that if the lines are more prone to fail, then more re-dispatch actions are needed in order to solve the overloaded problems that may appear. The lost load before re-dispatch has a considerable increase as well (50,7%). This result is logical in the sense that if the lines are more unavailable, then the lost load will be higher.

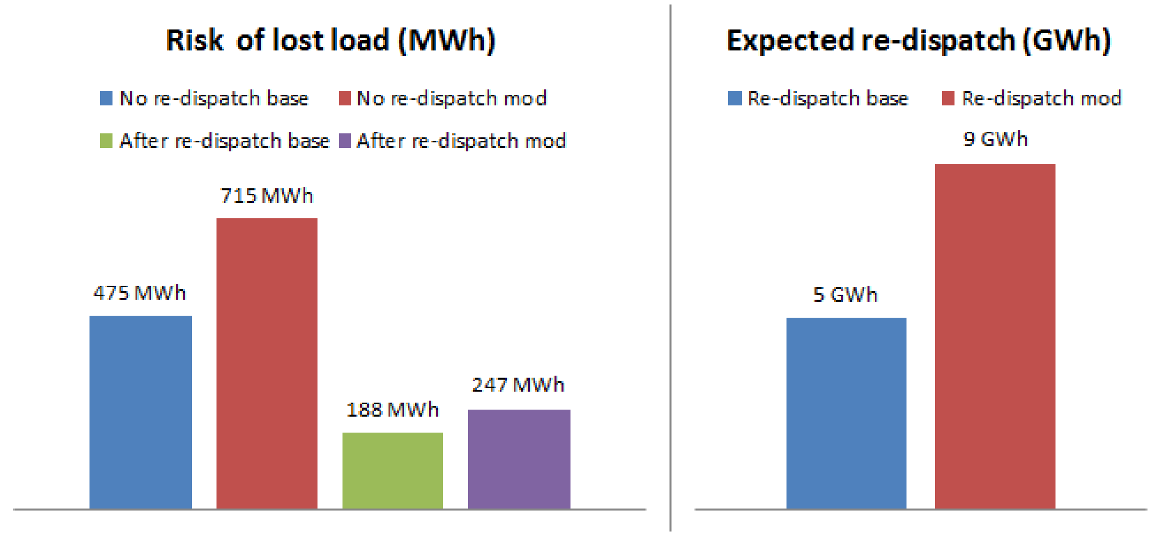


Figure 4-28: Risk of lost load and expected re-dispatch with modified unavailability

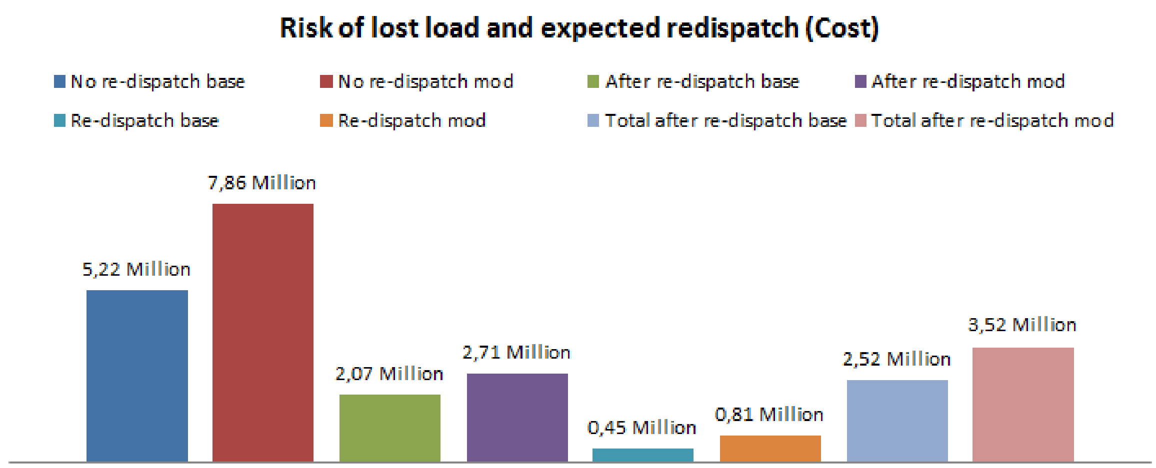


Figure 4-29: Comparison of risk of cost of lost load and expected re-dispatch (cost).

The increase on the unavailability of lines was product of the assumption that the lines reached the end of their lifetimes which is the period where more failure are likely to occur. On top of that, this increase was applied to all the lines of the grid. This assumption was taken in order to illustrate a extreme case of variation of unavailability.

According to the results, this increase of unavailability has more impact on the indicators that result from a state enumeration analysis than it does on the residual risk. The modification of unavailability does not influence much with the necessity of the analysis of high order contingencies, nevertheless, the influence over the risk indicators for lost load is noticeable.

The reach of the lifetime of the components is a rough condition to assume for all lines, in reality the increase might not be as drastic as in the example. In any case, it is important to include the modifications of failure probabilities of components into probabilistic analysis in order to obtain more realistic performance indicators.

The next chapter will present the conclusions and recommendation from this project.

Chapter 5

Conclusions and Recommendations

5-1 Conclusions

The main objective of this project was to investigate the potential improvements that can be brought to TenneT's reliability assessment in the context of system development by applying principles proposed by a general probabilistic approach.

In order to achieve that objective, this report presents a proof of concept Probabilistic Reliability Assessment (PRA) based on a state enumeration analysis whose indicators can provide an overview of the state of the system for a risk analysis. The analysis includes modelling of short time decisions as re-dispatch actions, clustering techniques and variation of lines unavailability. The proof of concept analysis aids to gain experience, knowledge and confidence with probabilistic approaches in order to understand future requirements for the analysis.

The conclusions obtained from the answers to the research questions are presented in the following paragraphs.

What is the feasibility of applying/combing a probabilistic methodology for long-term transmission planning with the current method employed by TenneT?

- The application of a probabilistic approach for long term planning with the resources of TenneT is feasible. The analysis is implemented using a test model for the 380/220 kV Dutch grid planned for the year 2025 and a test scenario based on the expected dispatch/load scenario expected for the same year, both described in section 3-1 of chapter 3.

- The analysis proves that risk of lost load indicators can be obtained from a state enumeration analysis which is described in section 4-1-2 from chapter 4. Risk variations and components are identified.

- A mapping from the obtained risk into the current risk matrix is shown in section 4-2 from chapter 4. In cases where the risk matrix states that some condition is unacceptable, the results from the analysis can provide more insight in a way that more information is available to understand better the variations of risk and its components.
- This project is not based on the GARPUR methodology. However, the notion of some elements and some recommendations are taken and implemented. The notion of a proxy from GARPUR to model operator's behaviour is implemented as re-dispatch actions. The notion of RMAC and its four elements described in section 2-2 from chapter 2 is kept in the logic of the project but not completely defined.
- A PRA can be beneficial in the sense that it provides insight of risk. This information can be used to take decisions or to make better assumptions. The drawback is that it requires high simulation time and that it is not a stand alone method. It can be applied in parallel with current methods.
- The introduction of renewables can result in the need for the construction of new assets given that the loads are not in the same place where power is generated. A PRA can be a useful option to assess reliability in a cost effective way.

The answers to the sub questions derived from the main research question are presented next.

Which methods can be used for the determination of failure rates of components, contingency selection, effect determination and risk analysis as inputs for models (simplified GARPUR RMAC, others) in order to perform a probabilistic reliability assessment?

Failure of components:

- As described in sections 2-4-5, 2-4-6, 2-4-7 and 2-4-8, operational conditions can alter the failure behaviour of components. A Bayesian data analysis (BDA) is simple to implement and provides a methodology to update the failure probabilities of transmission lines. On top of that, the BDA allows the inclusion of expert's opinion or professional guesses as prior information to make up for lack of data.
- As it is stated in sections 2-4-6 of chapter 2 and section 3-3 of chapter 3, the variation of component's unavailability due to dynamic operational conditions such as weather and loading levels is primarily important in the context of system operations. A BDA framework allows to update component's unavailability in order to provide information that can lead to better decisions. However, in the context of long term planning, dynamic conditions are not the main concern. Static operational conditions such as age and location are more interesting. A Bayesian framework can successfully provide the means to update failure probabilities according to operational conditions. The requirements for the application of BDA are presented in section 3-3 from chapter 3. One of the biggest challenges for estimation of component's failures is lack of data.
- Sections 4-5 and 4-6 from chapter 4 presents the impact of a high variation of unavailability over the results of a state determination analysis and over residual risk. The impact over state enumeration indicators is more critical than the impact over residual risk. Historical failure's databases and operational conditions contribute in the state of the system. Not analyzing all conditions might lead to incomplete failure estimation.

Effect determination and re-dispatch options

- The effect determination is discussed in sections 2-5 of chapter 2 and 3-4 of chapter 3. The effect is determined in terms of cost of lost load and cost of re-dispatch with the ultimately objective of calculate risk of lost load.
- Re-dispatch actions are introduced into analysis as an effort to model operators behavior when dealing with line overloadings. Load shedding actions are rarely implemented in systems operations. Line overloadings are solved by a number of resources implemented in this report as re-dispatch actions.
- Re-dispatch actions are implemented using linear programming techniques as presented in section 3-4 of chapter 3. Details about the objective function, variables and constraints can be found on the mentioned section.
- The impact on the results from the application of the re-dispatch algorithm is presented in sections 4-2, 4-3 and 4-4 of chapter 4. The results show that implementing re-dispatch makes a huge difference in the results obtained from a state determination analysis. The risk of lost load can have a significant variation that can lead to quite different outcomes. The no implementation of re-dispatch in the analysis might lead to an incorrect assessment of the situation.

Residual risk calculation and contingency selection

- Residual risk calculation and contingency selection is discussed in section 4-5 of chapter 4. The residual risk is calculated as the product of the probability of not considered states times the effect of a blackout. A justification for the analysis up to first order contingencies is provided in terms of residual risk. Results indicate that it is not necessary to analyze further than first order contingencies.

What is the computational performance resulting from the implementation of a probabilistic reliability assessment, how can it be reduced and under what conditions?

- The computational performance, described in section 4-2-1, of the state enumeration analysis with the selected software packages for this project is poor. Simulation time for a year round scenario of data is extremely high. Two options of implementation were discussed according to its computational performance. For this project, the best option is to follow the logic hour - contingency since it performs better in terms of simulation time. However, this is not always the case and it depends on the software employed for the simulations. The combination of the software selected for this project is not the best option. Altogether, commercial power system analysis software might not be the best choice for this kind of projects.
- Clustering of dispatch / load scenario constitute a viable option for reduction of simulation time as discussed in section 4-3 of chapter 4. Depending on the number of clusters to be analyzed, clustering can produce reasonable results while significantly reducing the amount of computation time which indicates that the relationship between the error introduced and the time saved is favorable in terms of time.

What is the level of sensitivity that can be achieved with the implementation of a probabilistic reliability assessment?

- Variations of residual risk and indicators from state enumeration analysis are discussed is sections 4-5-3 and 4-6 from chapter 4. The variation of lines unavailability does affect the results when it is applied to a large number of lines. Variation over single lines does not affect the results.
- The variation of risk depending of the contingencies analyzed is presented in section 4-4 from chapter 4. Not always the line that presents more lost load is the one that represents more risk. The variation of the parameters of lines such as length influence unavailability of lines and as consequence risk of lost load.

Final general conclusions

- To conclude, sections 4-2, 4-4, 4-5 and 4-6 show that PRA can provide the TSO with information about risky situations in order to assess, compare and ultimately optimize decisions in the context of long term planning. More information can also lead to better assumptions. Although, there is still a long way until a probabilistic approach can replace a well defined deterministic approach. A probabilistic approach can work along with current assessment methodologies such as the N-1 criteria and the risk matrix. In time, when evaluation criteria is well defined, PRA would be able to replace current methods.
- PRA has, without doubt, the potential to provide detailed insight of real risk in power systems. However there are some challenges that need to be sorted out. Perhaps, the greatest of all challenges is the definition of what is an acceptable risk under certain conditions. This acceptable risk should be defined in such a way that satisfies all actors in the power system including owners, regulators, operators and consumers.
- The definition of an acceptable risk will not be an easy task. The decision should include governments, universities, TSO's and customers in a regional level. Once that the clear levels of acceptable risk are defined, subsequent questions such as the ones raised in section 2-1-3 from chapter 2, will find a practical answer. This questions are:
  - What level of risk is acceptable and what level is desirable?
  - Could it be that a value of security of supply of 0.999 is accepted and one of 0.998 not if some risk is implied?
  - Is it economically effective to connect some remote facilities to the grid in a way that they are N-1 secure or is it acceptable to take some risk?
  - Is it acceptable to take some risk under maintenance?
  - How much power can be added to a grid without reinforcements?
- There is still a gap between the PRA features proposed by literature presented in sections 2-1 and 2-3 and the actual practical implementation of chapter 4. The implementation of projects with a probabilistic approach in TSO's such as TenneT will contribute to stretch this gap and hopefully close it in a near future.

5-2 Recommendations

The recommendations derived from the experience acquired after this project are presented in the following paragraphs.

General recommendations for the analysis

- Introduce the effect of protections in the analysis. Certain failures might produce line overloadings that could trigger protections and disconnection of lines. This actions can produce a chain reaction that might end in several lines disconnected and possibly a blackout.
- Maintenance actions can also be included in the analysis in order to assess the variation of risk during those actions.
- The existing interconnections between adjacent countries and the Dutch grid are modelled as a simple load with either a positive or negative value. Models for the behaviour of foreign grids must be used in the analysis in order to have a realistic assessment of the grid.
- This analysis focus on overloading conditions of transmission lines. The analysis can be expanded to assess voltage problems as well.
- The analysis assess the grid by means of a decoupled power flow analysis. After the identification of problems with decoupled power flow, a normal power flow can be executed to analyze further problems of the grid.

Failure of components and Bayesian data analysis

- A way to overcome the challenge imposed by the lack of data, is to start, join and maintain initiatives such as international TSO's cooperation for the establishment of joint failure databases.

Effect determination and re-dispatch actions

- The re-dispatch algorithm based on linear programming does not include fuel prices for the solution and assumes that every generator connected to the grid is capable to re-dispatch. This may not be the case. Specifications about available units, limits and prices per generator should be added to the re-dispatch algorithm.
- Another upgrade for the re-dispatch algorithm would be the addition of the possibility of curtailing energy from renewable sources.

Computational performance and clustering

- In order to reduce simulation time, it is necessary to develop tools designed with the solely purpose of implementing a PRA. Commercial power system simulation software offers functions that are not required for PRA and might slow down this kind of analysis. For this project, the greatest time consuming activity in the analysis is the writing action described in section 4-2-1 of chapter 4. A tool can be developed that allows fast data writing, modifies the grid, performs a load flow analysis, returns information and allows parallel working. There was an initiative involving several TSO's on the region where one of the premises was to work together in order to create specifications for a tool suitable for PRA. Initiatives such as this one are important to keep and develop.

- Clustering methodologies were applied over the representative set of data for a year constituted of 1008 hours with promising results. It is necessary to apply clustering to the whole set of hourly data for a year (around 8736 hours) in order to assess the performance of clustering techniques over complete sets of data.

- Explore parallel computing algorithms. Investigate if state enumeration analysis can be performed in parallel. Simulation time might be reduced if the work is parallelized.

- The writing action that slows down the analysis might be product of a bug in the simulation software. Investigate with the software vendor for any kind of bugs.

Appendix A

Probability Distributions

A-1 Binomial Distribution

This distribution function outlines the number of failures k found in a number of independent observations n. The binomial distribution complies with the following:

- The number of observations is known, positive, integer and different than zero.
- Every observation can end in one of two different outcomes: failure - operational, pass - fail, one - zero, etc.
- All results from observations are statistically independent of each other.
- The failure probability p is constant.

The binomial probability density function is given by equation A-1:

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, ..., n \tag{A-1}$$

With:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{A-2}$$

The mean and variance are given by equation A-3 and equation A-4 respectively:

$$E(X) = np \tag{A-3}$$

$$Var(X) = np(1-p) \tag{A-4}$$

A-2 Beta Distribution

This distribution function is shaped by two parameters α and β. The probability density function is given by equation A-5:

$$f(y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} \tag{A-5}$$

With the beta function B(α, β) defined by the equation A-6:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \tag{A-6}$$

The parameters α and β are positive. The variable y is bounded between $0 \le y \le 1$.

The mean and the variance are given by equation A-7 and equation A-8 respectively:

$$E(Y) = \frac{\alpha}{\alpha+\beta} \tag{A-7}$$

$$Var(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{A-8}$$

Appendix B

Proposition of information to be included in historical and operational failure databases

NESTOR contains information about historical failure statistics but not much information about operational conditions of the components at the time of the failure. The information that could be included is given for four components based on [1, 7, 36, 37]. The components are overhead lines, underground cables, transformers and capacitor banks.

B-1 Information for overhead lines

Overhead lines are completely exposed to the environment. Information such as weather conditions, external causes, loading, length and the operational age should be collected since those factors can modify the failure rates. The failure frequency of the line is assumed to increase linearly with the length [1].

The general information per line can be found in table B-1. The suggested content for the operational data is shown in tables B-2 and B-3. The basic information of the line includes data that does not change over the lifetime of the line. The operational data base holds the necessary information for the determination of the failure rate. The cause, age and repair time are detailed for future reference. The effect column details the information of the outcome from the failure and aids in the determination of dependent failures. The columns named season, weather and temperature can be used in correlation analysis to link failures to those operational conditions. With the analysis of the data contained in this database it is possible to determine the added unavailability.

Name	Location	Material	Length	Year of installation	AC/DC	Rating	Thermal rating	Lifetime
Line A	Forest	Al	15	2000	AC	2KA	200 C	60 Y

Table B-1: General information of the line.

Cause	Mechanism	Mode	Date Time	Age	Effect	Repair time	Season
Birds over the line	Sagging (weight)	Broken line	20/08/2010	10	Forced outage. 2 circuits disconnected	10 hours	Summer

Table B-2: Operational information of the line.

Weather	Temperature	Loading	Maintenance nearby	Human factor	External factors	Unavailability
Windy	30C	80%	Maintenance in the tower	People working on the line	Migration of animals	Calculated value

Table B-3: Operational information of the line. Continuation

B-2 Information for underground cables

Underground cables are mainly damaged by excavation works and component wear out by the soil. Weather conditions and external factors do not affect underground cables as they do on overhead lines. The length of the cable affect the failure rate in the same way as in overhead lines. Table B-4 contains the basic information of the cable and table B-5 contains the operational conditions. The information of the databases is used in the same way as previously explained. The columns temperature and soil from table B-5 can be used for correlation analysis. Temperature is included because an extreme high temperature can lead to insulation breakdown.

Name	Location	Material	Length	Depth	Type of soil
Cable B	Urban zone	Cu	100 Km	40m	Alkaline

Insulation type	Year of installation	AC/DC	Rating	Thermal rating	Lifetime
Paper and polyethylene	2000	DC	2KA	300C	40Y

Table B-4: General information of the cable

Cause	Mechanism	Mode	Date Time	Age	Effect	Repair time	Soil
Excavation	Insulation treeing	Insulation breakdown	02/06/2011	11	Planned outage. 1 circuit	300h	Type of soil at the failure

Weather	Loading	Temperature	Maintenance nearby	Human factor	External factor	Unavailability
Heavy rain	60%	80	No maintenance	Excavation works in the area	None	Calculated value

Table B-5: Operational information of the cable

B-3 Transformers with oil insulation

Transformers are primarily affected by the environment, external factors and loading. Most transmission system transformers use oil as a dielectric, the state of this oil depends on the overloading conditions and the lifetime of the transformer. Overloading can produce an increase in the oil temperature. A very high temperature might result in gas bubbles in the oil affecting its dielectric strength and producing failures. Table B-6 contains the basic information of the transformer and table B-7 holds the operational information of the transformer. The columns temperature, oil state, season and weather can be used to find correlations between the failures and the weather.

Name	Location	Material	Rating	Oil	Insulation class	Year of installation	Thermal rating	Lifetime
T A	Substation	Fe and Cu	3.5 KVA	Mineral 10GBNP	Insulation level	1990	100C	80Y

Table B-6: General information of the transformer

Cause	Mechanism	Mode	Date Time	Age	Effect	Repair time	Temperature	Oil state
High voltage	Insulation thermal degradation	Explosion	02/06/2005	15Y	Forced outage. Substation disconnected	200h	22C	Sediment and water

Oil state	Season	Weather	Loading	Maintenance nearby	Human factor	External factor	Unavailability
Sediment and water	Spring	Normal	90%	No maintenance around	People on the substation	Pollution	Calculated value

Table B-7: Operational information of the transformer

B-4 Capacitor banks

Capacitor banks are affected by the environment, external factors, loading and problems with the control system. Some capacitor banks contain oil and have the same problems with overloading and temperature as transformers. Fuses and controllers are key elements of the capacitor banks, if a fuse does not perform correctly, then failures might appear. Table B-4 contains the basic information of the bank and table B-9 records the operational information of the capacitor bank. The information of temperature, fuse state, season and weather from the historical failure database can be used for correlation analysis.

Name	Location	Material	Rating	Insulation	Series / parallel	Year of installation	Thermal rating	Lifetime	Control
Cap A	Substation	Material of the capacitor	Voltage rating	Insulation material	Series	2015	100C	80Y	Auto

Table B-8: General information for capacitor banks

Cause	Mechanism	Mode	Date Time	Age	Effect	Repair time	Temperature	Fuse state
Lightning strike	Thermal driven dielectric fail	Damaged units	02/10/2018	3Y	Planned outage.	100h	22ÂřC	Working

Season	Weather	Loading	Maintenance nearby	Human factor	External factor	Unavailability
Winter	Storm	90%	Maintenance around	People on the substation	Birds in the area	Calculated value

Table B-9: Operational information for capacitor banks

B-5 Repair times

Repair times are part of the failure database and together with the failure frequencies of elements are the input parameters needed in order to calculate unavailability. Usually, the average repair time is used to perform analysis but this is not the best option when there is not much data available. A better option is to manage the data statistically and fit it to distribution functions such as the exponential function [1]. The determination of repair times depends on the available data and the opinion of experts. The proposed information to be collected with the repair times is shown in table B-10. This information can be used for the determination of correlations among action, location, temperature, weather, resources and people available with the repair time. Later the information can be managed statistically to determine a single repair time.

Element	Action	Location	Temperature	Weather	Resources	Crew	Time
Line	Repair heavy damaged	Lake	20	Fog	Machinery	5 people	20h
Line	Repair	City	20	Normal	Machinery	10 people	8h
Transformer	Replacement	Substation	-5	Rain	Special tools	7 people	300h
Transformer	Replacement	Substation	20	Normal	Special tools	7 people	150h

Table B-10: Information for repair times

Appendix C

Lost load and risk per contingency graphs

The lost load and the risk per contingencies might change according to the scenario. This appendix presents different radial graphs for lost load and risk of lost load corresponding to different scenarios. The graphs were created from the result of the analysis executed using the cluster K22 (K-means of 22 elements).

C-1 Lost load and risk of lost load from the cluster K22

Figure C-1 presents the radial representation of the most representative contingencies that result in lost load before and after re-dispatch. Figure C-2 present the radial representation of the risk of lost load in euro from the most representative contingencies. From the two figures it is possible to see that the results are similar to those described in section 4-4 from chapter 4.

C-2 Lost load and risk of lost load with modified unavailability

Figure C-3 presents the radial representation of the most representative contingencies that result in lost load before and after re-dispatch from the case presented in section 4-6 of chapter 4 were the unavailability if lines was modified. Figure C-4 present the radial representation of the risk of lost load in euro from the most representative contingencies from the same case. As expected according to the results from section 4-6 of chapter 4, the modification of the unavailability results in a increase of the risk indicators per line.

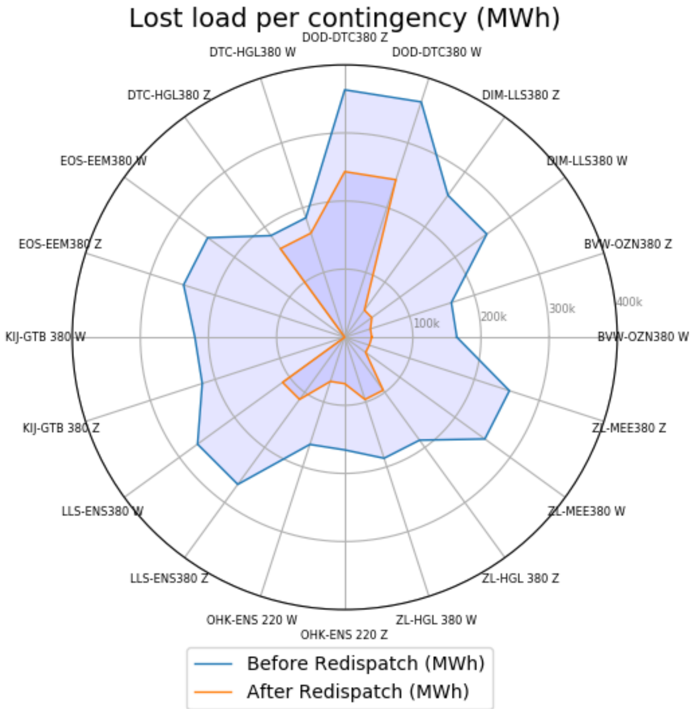


Figure C-1: Lost load per contingency cluster K22

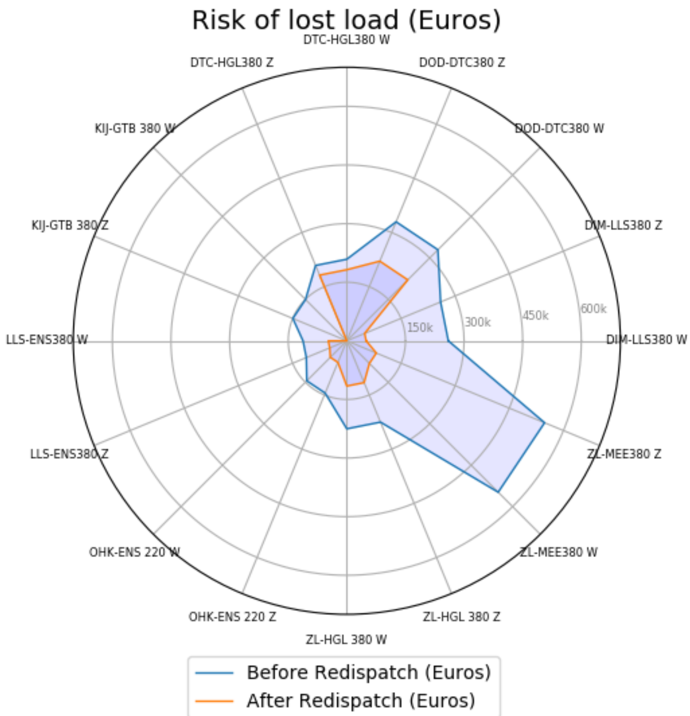


Figure C-2: Risk of lost load per contingency cluster K22

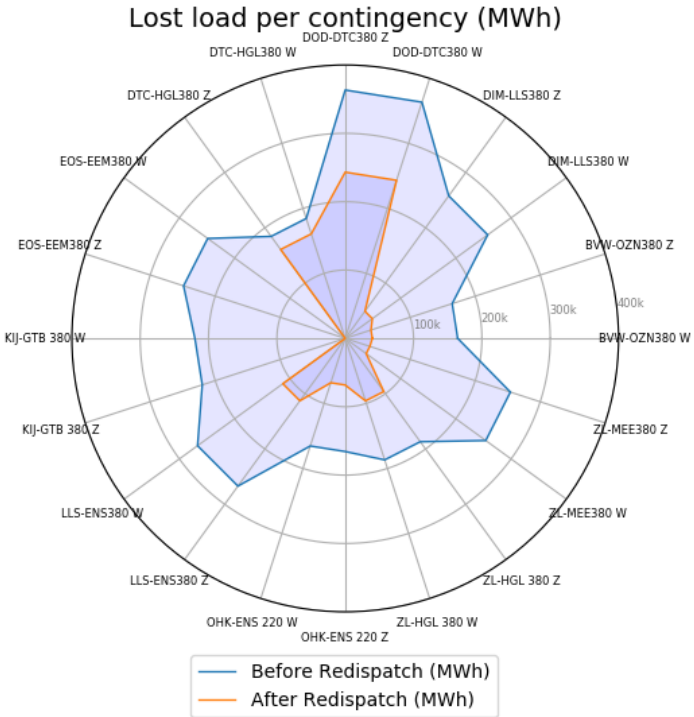


Figure C-3: Lost load per contingency, modified unavailability

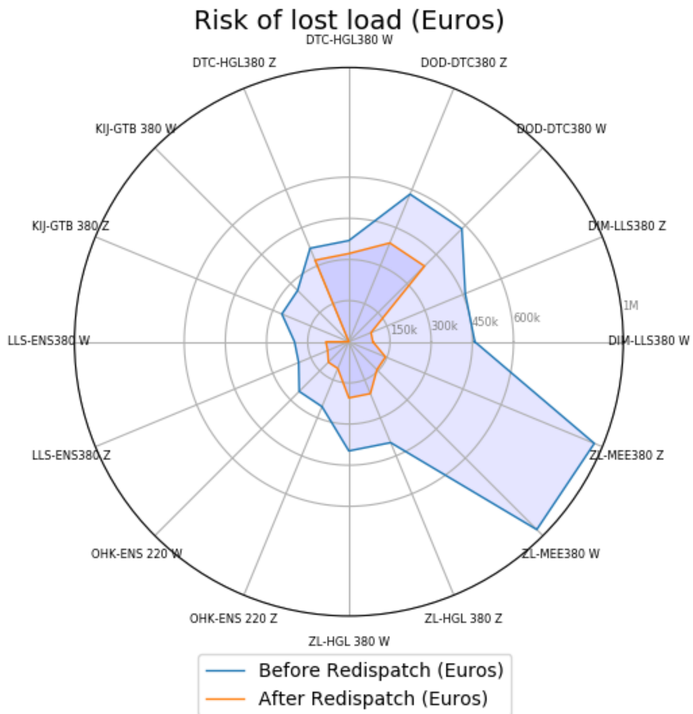


Figure C-4: Risk of lost load per contingency, modified unavailability

Bibliography

[1] B. W. Tuinema, "Reliability of sustainable power systems," July 2018. EE3065TU Course Reader, Version 03.

[2] I. T. Force, "Probabilistic security assessment for power system operations." IEEE transactions on power systems.

[3] GARPUR, "State of the art on reliability assessment in power systems. deliverable 1.1." https://www.sintef.no/globalassets/project/garpur/deliverables/garpur-d1.1-state-of-the-art-on-reliability-assessment-in-power-systems.pdf, 2014.

[4] GARPUR, "The garpur project results." https://www.sintef.no/globalassets/project/garpur/deliverables/garpur-brochure.pdf, 2017.

[5] GARPUR, "Guidelines for implementing the new reliability assessment and optimization methodology. deliverable 2.2." https://www.sintef.no/projectweb/garpur/deliverables/, 2016.

[6] L. Guo, Q. Qiu, J. Liu, Y. Zhou, and L. Jiang, "Power transmission risk assessment considering component condition," *Journal of Modern Power Systems and Clean Energy*, vol. 2, pp. 50–58, Mar 2014.

[7] Y. Sun, P. Wang, L. Cheng, and H. Liu, "Operational reliability assessment of power systems considering condition-dependent failure rate," *IET Generation, Transmission Distribution*, vol. 4, pp. 60–72, January 2010.

[8] TenneT TSO, "Kwaliteits- en capaciteitsdocument 2017," 2017.

[9] J. H. Jurgensen, *Individual Failure Rate Modelling and Exploratory Failure Data Analysis for Power System Components*. PhD thesis, KTH Skolan fÃűr elektroteknik och datavetenskap Elektroteknisk teori och konstruktion, 2018.

[10] R. Billinton and G. Singh, "Application of adverse and extreme adverse weather: modelling in transmission and distribution system reliability evaluation," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 153, pp. 115–120, Jan 2006.

[11] Ø. R. Solheim and G. Kjølle, "Wind dependent failure rates for overhead transmission lines using reanalysis data and a bayesian updating scheme," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–7, Oct 2016.

[12] D. Kelly and C. Smith, *Bayesian inference for probabilistic risk assessment. A practitioners guidebook.* Springer-Verlag London, 01 2011.

[13] B. M. Turner and T. V. Zandt, "A tutorial on approximate bayesian computation," *Journal of Mathematical Psychology*, vol. 56, no. 2, pp. 69 – 85, 2012.

[14] K. CsillÃľry, M. G. Blum, O. E. Gaggiotti, and O. FranÃğois, "Approximate bayesian computation (abc) in practice," *Trends in Ecology Evolution*, vol. 25, no. 7, pp. 410 – 418, 2010.

[15] TenneT TSO, "Product specifications, reserve power other purposes," 2019.

[16] M. van Blijswijk, "Applying basic system redispatch for managing congestion in the dutch transmission grid," *Delft University of Technology, Faculty of Technology, Policy and Management (TPM)*, 05 2011. https://repository.tudelft.nl/islandora/object/uuid:b212b14c-ded2-4117-9164-5f8548d7ad20/datastream/OBJ/download.

[17] C.-S. Song, C.-H. Park, M.-H. Yoon, and G. Jang, "Implementation of ptdfs and lodfs for power system security," *Journal of International Council on Electrical Engineering*, vol. 1, pp. 49–53, 01 2011.

[18] P. W. Corporation, "M3: Overview of linear programming." https://www.powerworld.com/files/M01ConceptsTrends.pdf.

[19] Wikipedia contributors, "Linear programming — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=Linear_programming&oldid=905853185, 2019. [Online; accessed 29-May-2019].

[20] B. Stott and J. L. Marinho, "Linear programming for power-system network security applications," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, pp. 837–848, May 1979.

[21] UTDallas, "The simplex method." https://personal.utdallas.edu/~scniu/OPRE-6201/documents/LP4-Simplex.html. [Online; accessed 07-July-2019].

[22] Carreira-PerpiÃśÃąn, Miguel ÃĄ. , "Simplex method.."

[23] HEC Montreal, "The steps of the simplex algorithm." https://www.hec.ca/en/cams/help/topics/The_steps_of_the_simplex_algorithm.pdf, 2019. [Online; accessed 07-July-2019].

[24] Marco Chiarandini - University of Southern Denmark, "Lecture 3, simplex method." https://imada.sdu.dk/~marco/DM559/Slides/dm545-lec3.pdf, 2018. [Online; accessed 27-July-2019].

[25] Laurel Nichols Gina Christofaro - Iowa State IMSE, "Explanation of simplex method." https://www.imse.iastate.edu/files/2015/08/Explanation-of-Simplex-Method.docx, 2015. [Online; accessed 27-July-2019].

[26] University of Babylon, "Lecture 11 - linear programming : The revised simplex method." http://www.uobabylon.edu.iq/eprints/publication_11_20693_31.pdf. [Online; accessed 27-July-2019].

[27] E. C. Bobric, G. Cartina, and G. Gheorghe, "Clustering techniques in load profile analysis for distribution stations," *Advances in Electrical and Computer Engineering*, vol. 9, 02 2009.

[28] N. P. K. N.V.V.Karunakar.Jureedi, K.M.Rosalina, "Clustering analysis and its application in electrical distribution system," *International Journal of Electrical, Electronics and Computer Systems (IJEECS)*, vol. 1, 2013.

[29] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, pp. 264–323, Sept. 1999.

[30] H. Teichgraeber and A. R. Brandt, "Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison," *Applied Energy*, vol. 239, pp. 1283 – 1293, 2019.

[31] DIgSILENT GmbH, "Digsilent | powerfactory." https://www.digsilent.de/en/powerfactory.html, 2019. [Online; accessed 29-June-2019].

[32] D. GmbH, "Powerfactory 2018 user manual," 2018.

[33] T. Lindquist, L. Bertling, and R. Eriksson, "A feasibility study for probabilistic modeling of aging in circuit breakers for maintenance optimization," in *2004 International Conference on Probabilistic Methods Applied to Power Systems*, pp. 1020–1025, Sep. 2004.

[34] T. Lindquist, "On reliability modelling of ageing equipment in electric power systems with regard to the effect of maintenance," 2005. QC 20101209.

[35] IEEE UAE Student Branch, "Distances in classification." http://www.ieee.ma/uaesb/pdf/distances-in-classification.pdf, 2016. [Online; accessed 13-August-2019].

[36] X. Zhang and E. Gockenbach, "Component reliability modeling of distribution systems based on the evaluation of failure statistics," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, pp. 1183–1191, October 2007.

[37] M. Lammintausta, R. Hirvonen, and M. Lehtonen, "Modelling of power system components for reliability analysis," in *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 4, pp. 6 pp. Vol.4–, June 2003.